SCIENCES SUP

Cours et exercices corrigés

L3 • Master • Écoles d'ingénieurs

PHYSIQUE DES SEMICONDUCTEURS ET DES COMPOSANTS ÉLECTRONIQUES

6e édition

Henry Mathieu
Hervé Fanet

DUNOD

PHYSIQUE DES SEMICONDUCTEURS ET DES COMPOSANTS ÉLECTRONIQUES

Cours et exercices corrigés

Consultez nos parutions sur dunod.com

PHYSIQUE DES SEMICONDUCTEURS ET DES COMPOSANTS ÉLECTRONIQUES

Cours et exercices corrigés

Henry Mathieu
Professeur à l'université Montpellier II

Hervé Fanet
Ingénieur de recherche au CEA LETI,
chargé de coopérations scientifiques et d'actions de formation
au sein du pôle MINIATEC (micro et nanotechnologies) de Grenoble

6e édition

DUNOD

Illustration de couverture : © *Fotolia*

Le pictogramme qui figure ci-contre mérite une explication. Son objet est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, particulièrement dans le domaine de l'édition technique et universitaire, le développement massif du photocopillage.
Le Code de la propriété intellectuelle du 1er juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique s'est généralisée dans les établissements d'enseignement supérieur, provoquant une baisse brutale des achats de livres et de revues, au point que la possibilité même pour les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée.
Nous rappelons donc que toute reproduction, partielle ou totale, de la présente publication est interdite sans autorisation de l'auteur, de son éditeur ou du Centre français d'exploitation du droit de copie (CFC, 20, rue des Grands-Augustins, 75006 Paris).

© Dunod, Paris, 2009
ISBN 978-2-10-054134-8

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2° et 3° a), d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4).
Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

PHYSIQUE DES SEMICONDUCTEURS ET DES COMPOSANTS ÉLECTRONIQUES

Cours et exercices corrigés

Consultez nos parutions sur dunod.com

PHYSIQUE DES SEMICONDUCTEURS ET DES COMPOSANTS ÉLECTRONIQUES

Cours et exercices corrigés

Henry Mathieu
Professeur à l'université Montpellier II

Hervé Fanet
Ingénieur de recherche au CEA LETI,
chargé de coopérations scientifiques et d'actions de formation
au sein du pôle MINIATEC (micro et nanotechnologies) de Grenoble

6e édition

DUNOD

Illustration de couverture : © *Fotolia*

Le pictogramme qui figure ci-contre mérite une explication. Son objet est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, particulièrement dans le domaine de l'édition technique et universitaire, le développement massif du photocopillage.
Le Code de la propriété intellectuelle du 1er juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique s'est généralisée dans les établissements d'enseignement supérieur, provoquant une baisse brutale des achats de livres et de revues, au point que la possibilité même pour les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée.
Nous rappelons donc que toute reproduction, partielle ou totale, de la présente publication est interdite sans autorisation de l'auteur, de son éditeur ou du Centre français d'exploitation du droit de copie (CFC, 20, rue des Grands-Augustins, 75006 Paris).

© Dunod, Paris, 2009
ISBN 978-2-10-051643-8

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2° et 3° a), d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4).
Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

Table des matières

AVANT-PROPOS XIII

1. NOTIONS FONDAMENTALES SUR LA PHYSIQUE DES SEMICONDUCTEURS 1

1.1. Structure cristalline 2

1.1.1. Géométrie du réseau cristallin 3
1.1.2. Opérations de symétrie 3
1.1.3. Systèmes cristallins 4
1.1.4. Plans réticulaires - Indices de Miller 4
1.1.5. Le système cubique 5
1.1.6. Réseau réciproque - Zones de Brillouin 7
1.1.7. Densité d'états dans l'espace réciproque 8
1.1.8. Théorème de Bloch 9

1.2. Etats électroniques dans les semiconducteurs Structure de bandes d'énergie 11

1.2.1. Des orbitales atomiques à la structure de bandes 11
1.2.2. Méthodes de calcul de la structure de bandes d'énergie 23
1.2.3. Caractéristiques de la structure de bandes d'énergie 28
1.2.4. Concept de masse effective 33
1.2.5. Densité d'états dans les bandes permises 40

1.3. Statistiques - Fonction de distribution des électrons 43

1.3.1. Statistique de Boltzmann 43
1.3.2. Statistiques quantiques 45
1.3.3. Signification physique des multiplicateurs de Lagrange 49
1.3.4. Fonction de Fermi 53

1.4. Le semiconducteur à l'équilibre thermodynamique 54

1.4.1. Semiconducteur non dégénéré 55
1.4.2. Semiconducteur dégénéré 57
1.4.3. Semiconducteur intrinsèque 60
1.4.4. Semiconducteur extrinsèque à la température ambiante 61
1.4.5. Evolution avec la température 65

1.5. Le semiconducteur hors équilibre 71

1.5.1. Courants dans le semiconducteur 71
1.5.2. Génération-Recombinaison. Durée de vie des porteurs 87
1.5.3. Equations d'évolution 95

1.6. Interface entre deux matériaux différents 108

1.6.1. Travail de sortie - Affinité électronique 108
1.6.2. Effet Schottky 112

1.6.3. Etats de surface et d'interface 114
1.6.4. Emission thermoélectronique dans les hétérostructures 116
1.6.5. Emission de champ 122

1.7. Les sources de bruit dans les semiconducteurs 124
1.7.1. Définitions générales pour le signal de bruit 124
1.7.2. Le bruit thermique 127
1.7.3. Le bruit de grenaille 128
1.7.4. Bruit de génération-recombinaison 131
EXERCICES DU CHAPITRE 1 133

2. JONCTION PN 137

2.1. Jonction abrupte à l'équilibre thermodynamique 138
2.1.1. Charge d'espace 139
2.1.2. Tension de diffusion 140
2.1.3. Potentiel et champ électriques dans la zone de charge d'espace 142
2.1.4. Largeur de la zone de charge d'espace 143
2.1.5. Signification de la longueur de Debye 144

2.2. Jonction abrupte polarisée 146
2.2.1. Distribution des porteurs 148
2.2.2. Courants de porteurs minoritaires 152
2.2.3. Densité de courant – caractéristique 154

2.3. Jonction à profil de dopage quelconque 158
2.3.1. Régions n et p très courtes 158
2.3.2. Jonction quelconque 161

2.4. Capacités de la jonction pn 161
2.4.1. Jonction polarisée en inverse - Capacité de transition 161
2.4.2. Jonction en régime alternatif - Capacité de diffusion 163

2.5. Mécanismes de génération-recombinaison dans la zone de charge d'espace 167
2.5.1. Polarisation inverse - Courant de génération 169
2.5.2. Polarisation directe - Courant de recombinaison 170

2.6. Jonction en régime transitoire - Temps de recouvrement 171

2.7. Claquage de la jonction polarisée en inverse 177

2.8. Jonction en régime de forte injection 179
2.8.1. Injection des porteurs 180
2.8.2. Densité de courant 182
2.8.3. Chute de tension dans les régions neutres de la diode 185
EXERCICES DU CHAPITRE 2 190

3. TRANSISTORS BIPOLAIRES 193

3.1. Effet transistor 194

3.2. Equations d'Ebers-Moll 197
3.2.1. Courants de porteurs minoritaires dans l'émetteur et le collecteur 197

3.2.2. Courant de porteurs minoritaires dans la base 198
3.2.3. Courants d'émetteur et de collecteur 200

3.3. Différents types de profil de dopage 202
3.3.1. Transistor à dopages homogènes 202
3.3.2. Transistor drift 204
3.3.3. Résumé du fonctionnement d'un transistor 206

3.4. Effet Early 207

3.5. Modèle dynamique et étages à transistor 210
3.5.1. Modèle dynamique 210
3.5.2. Montage base commune 212
3.5.3. Montage émetteur commun 214
3.5.4. Montage collecteur commun 216

3.6. Sources de bruit dans un transistor 217
3.6.1. Bruit de grenaille d'une diode 217
3.6.2. Bruit de grenaille du bipolaire 218
EXERCICES DU CHAPITRE 3 223

4. CONTACT METAL-SEMICONDUCTEUR DIODE SCHOTTKY 227

4.1. Diagramme des bandes d'énergie 228
4.1.1. Considérations générales sur les niveaux énergétiques 228
4.1.2. Premier cas : les travaux de sortie sont égaux, $\phi_m = \phi_s$ 229
4.1.3. Deuxième cas : travail de sortie du métal supérieur, $\phi_m \geq \phi_s$ 230
4.1.4. Troisième cas : travail de sortie du métal inférieur, $\phi_m \leq \phi_s$ 234

4.2. Zone de charge d'espace 240
4.2.1. Champ et potentiel électrique 240
4.2.2. Capacité 242

4.3. Caractéristique courant-tension 244
4.3.1. Courant d'émission thermoélectronique 245
4.3.2. Courant de diffusion 247
4.3.3. Combinaison des deux courants 251
EXERCICES DU CHAPITRE 4 255

5. STRUCTURE METAL-ISOLANT SEMICONDUCTEUR 257

5.1. Diagramme énergétique 258
5.1.1. Structure métal-vide-semiconducteur 258
5.1.2. Structure Métal-Isolant-Semiconducteur – MIS 263

5.2. Potentiels de contact 267

5.3. Le modèle électrique de base 269
5.3.1. Description phénoménologique 269
5.3.2. Modèle électrique 275

5.4. Le régime de forte inversion 281

5.5. Le régime de faible inversion 286
EXERCICES DU CHAPITRE 5 290

6. HETEROJONCTIONS ET TRANSISTOR A HETEROJONCTION 293

6.1. Diagramme de bandes d'énergie 294
6.1.1. Diagramme énergétique loin de la jonction 295
6.1.2. Etats d'interface 298
6.1.3. Diagramme énergétique au voisinage de la jonction 298

6.2. Hétérojonction à l'équilibre thermodynamique 304

6.3. Hétérojonction polarisée 308
6.3.1. Modèle d'émission thermoélectronique 309
6.3.2. Modèle de diffusion 323
6.3.3. Courant tunnel -courant de recombinaison 329

6.4. Transistor à hétérojonction – HBT 332
6.4.1. Principe de fonctionnement 332
6.4.2. Courants d'émetteur et de collecteur 339

7. TRANSISTORS A EFFET DE CHAMP JFET ET MESFET 343

7.1. Transistor à effet de champ à jonction – JFET 344
7.1.1. Structure et principe de fonctionnement 344
7.1.2. Equations fondamentales du JFET 347
7.1.3. Réseau de caractéristiques 353

7.2. Transistor à effet de champ à barrière de Schottky MESFET 358
7.2.1. Structure et spécificité 358
7.2.2. Courant de drain 360
7.2.3. Tension de saturation - Courant de saturation 365
7.2.4. Transconductance 367
7.2.5. Fréquence de coupure du transistor 369

7.3. Transistor MESFET-GaAs 371
7.3.1. Régime linéaire 373
7.3.2. Régime sous-linéaire 377
7.3.3. Régime de saturation 379
EXERCICES DU CHAPITRE 7 383

8. LE TRANSISTOR MOSFET ET SON EVOLUTION 385

8.1. Principe de base et historique 386

8.2. Comment la structure MOS se modifie 388

8.3. Le modèle canal long 394
8.3.1. Le régime de forte inversion 394
8.3.2. Le régime de faible inversion 400
8.3.3. La mobilité effective et ses effets 403
8.3.4. Le MOS canal p 404

8.4. Le modèle canal court 405
8.4.1. Effet de diminution de la longueur de canal 406
8.4.2. Effet de saturation de la vitesse 407
8.4.3. Effet de diminution du seuil effectif 416
8.4.4. Traitement analytique du canal court 417

8.5. Le fonctionnement dynamique du MOS 425
8.5.1. Régime quasi-statique 425
8.5.2. Régime dynamique 431

8.6. Les modèles du transistor MOSFET 433
8.6.1. Rappels du modèle statique 433
8.6.2. Modèles petits signaux : généralités 435
8.6.3. Le MOS interne en basse fréquence 435
8.6.4. Le MOS interne à fréquence moyenne 439
8.6.5. Le modèle du MOS complet aux fréquences moyennes 447
8.6.6. Un résumé du modèle du MOS 451

8.7. Le bruit du transistor MOS 455
8.7.1. Le bruit thermique du canal en forte inversion 455
8.7.2. Le bruit thermique du canal en faible inversion 459
8.7.3. Les autres sources de bruit 460

8.8. Applications du MOSFET 462
8.8.1. Amplification 462
8.8.2. Porte logique : cas de l'inverseur 464
EXERCICES DU CHAPITRE 8 477

9. COMPOSANTS OPTOELECTRONIQUES 479

9.1. Interaction rayonnement – semiconducteur 480
9.1.1. Photons et électrons 480
9.1.2. Interaction électron-photon - Transitions radiatives 484
9.1.3. Absorption - Emission spontanée - Emission stimulée 486
9.1.4. Recombinaison de porteurs en excès - Durée de vie 492
9.1.5. Création de porteurs en excès 494
9.1.6. Semiconducteurs pour l'optoélectronique 498

9.2. Photodétecteurs 504
9.2.1. Distribution des photoporteurs dans un semiconducteur photoexcité 504
9.2.2. Cellule photoconductrice 512
9.2.3. Cellule photovoltaïque – Photodiode 516
9.2.4. Cellule solaire – photopile 523

9.3. Emetteurs de rayonnement à semiconducteur 531
9.3.1. Diode électroluminescente – LED 531
9.3.2. Diode laser – LD 542

9.4. Applications des composants optroniques 566
9.4.1. Disques optiques - CD-DVD 566
9.4.2. Diagramme de chromaticité 569
EXERCICES DU CHAPITRE 9 574

10. EFFETS QUANTIQUES DANS LES COMPOSANTS, PUITS QUANTIQUES ET SUPERRESEAUX 577

10.1. Effets quantiques dans les hétérojonctions et les structures MIS 578

10.1.1. Structure de sous-bandes d'énergie 581
10.1.2. Energie potentielle des électrons 588
10.1.3. Méthodes de calcul dans l'approximation de Hartree 597

10.2. Puits quantiques 613

10.2.1. Spectre d'énergie 616
10.2.2. Multipuits quantiques 624
10.2.3. Puits quantiques couplés 625

10.3. Superréseaux 627

10.3.1. Structure de sous-bandes d'énergie 628
10.3.2. Modèle de Kronig-Penney 630

10.4. Transistor à effet de champ à gaz d'électrons bidimensionnel-TEGFET 639

10.4.1. Structure 639
10.4.2. Commande de grille 640
10.4.3. Polarisation de drain 648
10.4.4. Effet MESFET parasite 650

11. LA FABRICATION COLLECTIVE DES COMPOSANTS ET DES CIRCUITS INTEGRES 655

11.1. La lithographie 656

11.2. Les procédés de retrait de matériaux 660

11.2.1. La gravure humide 660
11.2.2. La gravure sèche 662

11.3. Les procédés d'apport de matériaux 663

11.3.1. Evaporation 664
11.3.2. Epitaxie par jets moléculaires (MBE) 665
11.3.3. Dépôt par « sputtering » 665
11.3.4. Dépôt par laser pulsé (PLD) 666
11.3.5. Le dépôt en phase vapeur (CVD) 667
11.3.6. Dépôt de type MOCVD 668
11.3.7. Dépôt de type CSD (Chemical Solution Deposition) 668
11.3.8. Croissance thermique 668
11.3.9. La croissance électrolytique 670
11.3.10. Implantation ionique 670
11.3.11. La diffusion thermique 672

11.4. Le flot simplifié pour la technologie CMOS 673

11.5. Les procédés alternatifs 676

11.5.1. La nano-impression 677
11.5.2. Techniques d'auto-assemblage 678

12. COMPOSANTS QUANTIQUES 679

12.1. Transport parallèle dans les structures quantiques 680

12.1.1. Spécificités 680
12.1.2. Mobilité 683
12.1.3. Effet Gunn bidimensionnel 686
12.1.4. Transfert dans l'espace réel – RST 687
12.1.5. Etalon de résistance 696

12.2. Transport perpendiculaire dans les structures quantiques 700

12.2.1. Oscillateur de Bloch 700
12.2.2. Effet Wannier-Stark 704
12.2.3. Transport dans un superréseau-RDN 705
12.2.4. Effet tunnel résonnant dans un puits quantique à double barrière 709
12.2.5. Effet tunnel résonnant dans un superréseau 717

12.3. Lasers à puits quantiques 718

12.3.1. Effets géométriques - Facteur de confinement 718
12.3.2. Effets quantiques 723
12.3.3. Laser à cavité verticale-VCSEL 725
12.3.4. Microcavité photonique - Laser sans seuil 730
12.3.5. Laser à cascade quantique 731
12.3.6. Laser à boites quantiques 734

12.4. Blocage de Coulomb et systèmes à peu d'électrons 735

12.5. Nanotubes et nanofils 744

12.5.1. La structure des nanotubes 744
12.5.2. Une nouvelle théorie de la conduction 746
12.5.3. Propriétés et fabrication des nanotubes 749
12.5.4. Applications des nanotubes 750
12.5.5. Les nanofils 751

13. SEMICONDUCTEURS A GRAND GAP 755

13.1. Solutions technologiques 756

13.1.1. Silicium sur isolant – SOI 756
13.1.2. Arséniures sur diamant 758

13.2. Semiconducteurs à grand gap 759

13.2.1. Diamant 762
13.2.2. Carbure de silicium – SiC 765
13.2.3. Nitrures III-N – BN, AlN, GaN, InN 771

13.3. Composants optoélectroniques de courte longueur d'onde 779

13.3.1. Détecteurs de rayonnement ultraviolet 780
13.3.2. Emetteurs bleus 784
13.3.3. Emetteurs ultraviolets 789
13.3.4. Emetteurs blancs 790

ANNEXES 793

A.1. Coefficient de transmission d'une barrière de potentiel — Méthode BKW 793

A.2. Linéarisation de la fonction de distribution de Fermi 795

A.3. Effet Hall 799

A.3.1. Modèle simple 799

A.3.2. Modèle plus élaboré 802

A.3.3. Magnétorésistance 806

A.3.4. Effet Hall quantique 808

A.4. Ancrage du niveau de Fermi par les états de surface 812

A.5. Calcul du courant dans une structure MOS 816

CONSTANTE PHYSIQUES 819

BIBLIOGRAPHIE 821

INDEX 823

Avant-propos

Cet ouvrage est principalement destiné aux étudiants de second et troisième cycles universitaires ainsi qu'aux élèves-ingénieurs, dans le domaine de la physique des semiconducteurs et des composants électroniques. C'est aussi un ouvrage de base pour les jeunes chercheurs.

L'étude du fonctionnement des différents types de composants électroniques passe par une maîtrise préalable des phénomènes physiques régissant les propriétés des électrons dans les semiconducteurs. En outre, les composants modernes faisant appel à des structures complexes de couches minces de matériaux différents, nous devons définir les grandeurs physiques qui, dans ces hétérostructures, permettent de caractériser le comportement des électrons aux interfaces. Ces études font l'objet du premier chapitre.

La jonction pn, qui résulte de la juxtaposition dans un même semiconducteur de deux régions de types différents, constitue une structure de base dont l'étude est développée dans le chapitre 2. Les transistors bipolaires qui mettent à profit les propriétés associées au couplage de deux ou plusieurs jonctions pn, sont étudiés dans le chapitre 3.

Les chapitres 4, 5 et 6 présentent les propriétés des hétérostructures de base que sont respectivement, le contact métal-semiconducteur, la structure métal-oxyde-semiconducteur et l'hétérojonction entre deux semiconducteurs différents. Les propriétés de ces hétérostructures sont mises à profit dans la réalisation des différents types de composants en particulier les transistors à effet de champ décrits dans le chapite 7.

Le chapitre 8 est entièrement consacré au transistor MOS qui est le composant de base de la microélectronique actuelle.

Le chapitre 9 traite de problèmes importants dans le domaine des télécommunications, en particulier des composants optoélectroniques. Les différents phénomènes qui régissent les interactions du rayonnement avec le

semiconducteur sont d'abord passés en revue. La deuxième partie du chapitre traite des photodétecteurs que sont les cellules photoconductrices et les photodiodes, ainsi que des convertisseurs d'énergie solaire que sont les photopiles. La troisième partie est consacrée à l'étude des photoémetteurs que sont les diodes électroluminescentes et les lasers à semiconducteur..

Le chapitre 10 présente les effets spécifiques de la nature bidimensionnelle du gaz d'électrons localisé aux interfaces, dans différents types d'hétérostructure. Ce chapitre constitue une introduction à l'étude de nouveaux phénomènes, qui permettent d'étendre les performances des composants. Ces phénomènes sont liés à la bidimensionalité associée aux puits quantiques dans lesquels évoluent les électrons.

Le chapitre 11 décrit de manière simplifiée les procédés de fabrication des composants à base de semiconducteurs.

Enfin, lorsque les dimensions de la zone active d'un composant atteignent l'échelle nanométrique, la réduction de dimensions devient une réduction de dimensionalité et la représentation corpusculaire de l'électron doit céder le pas à la représentation ondulatoire. Le confinement quantique résultant de la réduction de dimensionalité du mouvement de l'électron redistribue la structure de bandes et les densités d'états. En outre, certains effets spécifiques de la nature ondulatoire des électrons, peuvent être exploités et les fonctions d'onde électroniques produisent alors les mêmes sortes de phénomènes, interférences ou diffraction, que les fonctions d'ondes optiques. Comme pour les photons, le paramètre qui conditionne l'existence de ces effets *mésoscopiques* est la *longueur de cohérence de phase* qui correspond sensiblement au libre parcours moyen de l'électron-particule. Ces composants, que nous qualifierons de *quantiques,* et dont certains restent encore du domaine prospectif, font l'objet du chapitre 12.

Notons pour terminer que les composants électroniques traditionnels, à base de silicium, fonctionnent difficilement à des températures supérieures à 250 °C. Or de nombreuses utilisations nécessitent un fonctionnement à des températures bien supérieures. Citons par exemple les microprocesseurs embarqués, sur des véhicules terrestres, aériens ou spatiaux, ou dans l'industrie pétrolière. Notons aussi que les hautes températures peuvent certes être le fait d'un environnement hostile mais peuvent aussi résulter d'un échauffement interne du composant, c'est en particulier le cas dans un composant de puissance. Différentes voies, notamment technologiques, sont explorées pour prolonger le fonctionnement des composants traditionnels vers les hautes températures ou les fortes puissances, mais à mesure que les besoins évoluent vers des conditions de plus en plus extrêmes, les solutions technologiques butent de plus en plus sur la barrière infranchissable des propriétés intrinsèques des matériaux. Une solution alternative aux matériaux traditionnels devient ainsi progressivement nécessaire. Cette alternative n'existe que dans l'utilisation de matériaux de remplacement potentiellement plus performants, notamment dans les domaines des hautes températures et des fortes puissances.

Parmi ces matériaux les semiconducteurs à grand gap constituent évidemment des éléments de choix. Les spécificités de ces matériaux et les performances potentielles des composants qui en découlent sont développées dans le chapitre 13.

Des exercices et des questions sont ajoutés à la fin de la plupart des chapitres. Ils sont soit de simples applications numériques soit des questions sur les points les plus importants ou les plus délicats. Parfois, ils permettent de traiter des applications plus spécifiques (diode avalanche, diode impatt, thyristor, CCD…).

1. Notions fondamentales sur la physique des semiconducteurs

Les composants électroniques mettent à profit les propriétés des électrons dans les semiconducteurs. Il est par conséquent nécessaire, avant d'aborder l'étude des composants proprement dits, de préciser ces propriétés et de définir les grandeurs physiques dont les évolutions conditionnent les caractéristiques électriques ou optiques des composants. Les paramètres fondamentaux sont évidemment l'état de la population électronique à l'équilibre thermodynamique et l'évolution de cette population lorsque le semiconducteur est soumis à une perturbation extérieure telle qu'une tension électrique ou un rayonnement électromagnétique. En outre les composants modernes font de plus en plus appel à la juxtaposition de matériaux différents, semiconducteurs ou non. Il est donc nécessaire de préciser et de traduire par des grandeurs mesurables, les propriétés des interfaces entre les différents matériaux constituant ces hétérostructures.

L'objet de ce premier chapitre est une présentation aussi simple que possible des concepts de base permettant de comprendre et de traduire par des grandeurs mesurables, les propriétés des électrons dans les semiconducteurs. Il est organisé de la manière suivante :

1.1 . Structure cristalline
1.2. Etats électroniques et structure de bandes
1.3. Fonctions de distribution et statistiques
1.4. Le semiconducteur à l'équilibre thermodynamique
1.5. Le semiconducteur hors équilibre
1.6. Les interfaces entre deux matériaux différents
1.7. Les sources de bruit dans les semiconducteurs

1.1. Structure cristalline

Il existe deux types d'état solide, l'état dans lequel l'arrangement des atomes est aléatoire et celui dans lequel les atomes sont arrangés régulièrement aux noeuds d'un réseau.

Le premier état est dit *amorphe*, les matériaux qui se solidifient dans un état amorphe sont généralement appelés des *verres*. Cet état ne diffère de l'état liquide que par le taux de viscosité. Tout liquide dont la viscosité de cisaillement est supérieure à 10^{13} poises est appelé verre.

Le deuxième état, qui nous intéresse plus particulièrement ici, est l'état *cristallisé*, caractérisé par le fait que les atomes sont rangés aux noeuds d'un réseau périodique. Le résultat est un ensemble ordonné de noyaux et d'électrons liés entre eux par des forces essentiellement coulombiennes. Ces forces sont en principe connues mais leurs interactions sont tellement complexes que l'on ne peut pas les calculer en détail. En fait, on associe les électrons des couches internes des atomes avec leur noyau, ce qui représente un ion positif et on traite les électrons périphériques comme des particules quasi-libres dans le champ des ions. On distingue, à partir de ce type de représentation, essentiellement quatre familles de solides cristallisés : les cristaux ioniques, les cristaux covalents, les métaux et les cristaux moléculaires.

Les *cristaux ioniques* résultent de l'association d'un élément fortement électropositif et d'un élément fortement électronégatif. L'élément électropositif a généralement un seul électron périphérique (métaux alcalins : Li, Na, K, Rb, Cs), qu'il cède facilement pour devenir un ion positif avec une configuration électronique stable de couches saturées. L'élément électronégatif a généralement sept électrons périphériques (halogènes : F, Cl, Br, I), il accepte facilement un huitième électron pour devenir un ion négatif avec une configuration électronique stable. Les deux ions ainsi créés sont liés par attraction coulombienne, d'où le nom de cristaux ioniques que l'on donne aux cristaux tels que LiF, NaCl ou KBr. L'électron libéré par le métal alcalin est fortement fixé sur l'halogène de sorte qu'aucun électron n'est libéré dans le réseau du matériau, ces cristaux sont des isolants. En outre, l'énergie de liaison entre les atomes est très importante de sorte que ces cristaux sont généralement très durs.

Les *cristaux covalents* sont construits avec des éléments de la colonne IV du tableau périodique (C, Si, Ge, Sn). Ces éléments ont quatre électrons périphériques qu'ils mettent en commun avec quatre voisins pour établir des liaisons covalentes. Les électrons de valence sont liés mais leur énergie de liaison est beaucoup plus faible que dans les cristaux ioniques. Cette énergie de liaison est importante dans le carbone diamant, ce qui en fait un isolant, elle est nulle dans l'étain, ce qui en fait un conducteur. Dans le silicium et le germanium cette énergie a une valeur intermédiaire qui fait de ces matériaux des semiconducteurs.

Les *métaux* sont construits avec des éléments électropositifs, c'est-à-dire ayant un seul électron périphérique. Cet électron périphérique est libéré dans la

réalisation du cristal, ces matériaux sont très conducteurs. Les liaisons entre atomes sont plus faibles que dans les cristaux ioniques ou covalents, ces matériaux sont moins durs et fondent à basse température. On distingue les métaux alcalins Li, Na, K, Cs et les métaux nobles Cu, Ag, Au.

Les *cristaux moléculaires*, comme leur nom l'indique, sont bâtis sur une unité de base qui n'est plus l'atome mais la molécule. Les forces de liaison sont grandes à l'intérieur de la molécule mais du type Van der Waals entre molécules et par conséquent faibles. Ces matériaux sont peu résistants et fondent à basse température.

1.1.1. *Géométrie du réseau cristallin*

Un cristal est construit à partir d'un groupe de n particules qui constituent la *cellule de base*. Ce groupe est répété périodiquement, en un réseau défini par trois vecteurs de translation fondamentaux, $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, appelés *vecteurs primitifs*.

La position de chaque cellule du cristal est définie par un vecteur du réseau

$$\mathbf{t} = n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c} \tag{1-1}$$

où n_1, n_2 et n_3 sont des nombres entiers. L'ensemble des points définis par les vecteurs $\mathbf{t}$ pour toutes les valeurs de n_1, n_2, n_3 constitue le réseau. Le parallélépipède construit sur les trois vecteurs primitifs constitue la maille primitive ou *maille élémentaire*. On remplit tout l'espace en appliquant à la maille élémentaire les opérations de translation définies par les vecteurs $\mathbf{t}$. Le volume de la maille élémentaire est défini par le produit mixte. Le symbole $\times$ désigne le produit vectoriel et le symbole . désigne le produit scalaire.

$$v = \left|\mathbf{a} \times \mathbf{b}.\mathbf{c}\right| \tag{1-2}$$

Il existe une maille élémentaire par noeud du réseau, ce qui s'exprime en disant que la maille élémentaire contient un seul noeud. En fait, la maille élémentaire contient huit noeuds mis en commun avec huit mailles adjacentes.

1.1.2. *Opérations de symétrie*

Les opérations de symétrie sont les opérations qui laissent la structure cristalline invariante. Ces opérations sont d'une part des translations et d'autre part des opérations ponctuelles telles que des rotations autour d'un axe ou des inversions par rapport à un point. L'ensemble des opérations de symétrie du cristal constitue le *groupe de symétrie* du cristal.

En ne considérant que les opérations ponctuelles, rotations et inversions ou combinaisons des deux, on définit le *groupe ponctuel* du cristal. En considérant en outre les opérations de translation, on définit le *groupe spatial* du cristal.

Il existe 32 groupes ponctuels qui constituent les 32 classes de cristaux et 230 groupes spatiaux qui constituent 230 types de cristaux.

1.1.3. *Systèmes cristallins*

Tous les cristaux qui ont une maille primitive de même symétrie appartiennent au même *système cristallin.* La maille élémentaire est définie par les trois vecteurs primitifs **a** , **b** , **c** , c'est-à-dire par les longueurs *a,b,c* de ces vecteurs et les angles α, β, γ qu'ils font entre eux (Fig. 1-1). Suivant les valeurs relatives de ces six grandeurs, on définit sept systèmes cristallins.

Système cubique
$\alpha = \beta = \gamma = 90°$ $\quad a = b = c$
Système tétragonal
$\alpha = \beta = \gamma = 90°$ $\quad a = b \neq c$
Système orthorombique
$\alpha = \beta = \gamma = 90°$ $\quad a \neq b \neq c$
Système trigonal
$\alpha = \beta = \gamma \neq 90°$ $\quad a = b = c$
Système hexagonal
$\alpha = \beta = 90°, \gamma = 120°$ $\quad a = b \neq c$
Système monoclinique
$\alpha = \beta = 90° \neq \gamma$ $\quad a \neq b \neq c$
Système triclinique
$\alpha \neq \beta \neq \gamma$ $\quad a \neq b \neq c$

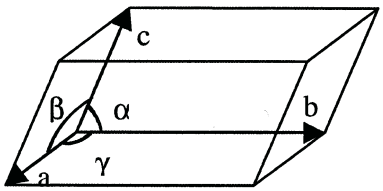


Figure 1-1 : La maille élémentaire

Toutes les propriétés physiques d'un cristal ont la symétrie du système cristallin auquel il appartient. Ainsi, par exemple, un cristal cubique est isotrope électriquement et optiquement.

En résumé, il existe donc 7 systèmes cristallins, 32 classes de cristaux ou groupes ponctuels et 230 types de cristaux ou groupes spatiaux. Le silicium par exemple appartient au système cubique, au groupe ponctuel 0_h et au groupe spatial 0_h^7.

1.1.4. *Plans réticulaires - Indices de Miller*

Pour définir un plan réticulaire du cristal, on utilise 3 indices déterminés de la manière suivante : soient *x*, *y*, *z* les coordonnées des intersections du plan à définir avec les axes sous-tendus par les vecteurs primitifs **a** , **b** , **c** (Fig. 1-2), soient 1/*x*, 1/*y* et 1/*z* leurs inverses, soient enfin *h*, *k*, *l* les trois entiers les plus petits possibles proportionnels à 1/*x*, 1/*y* et 1/*z*.

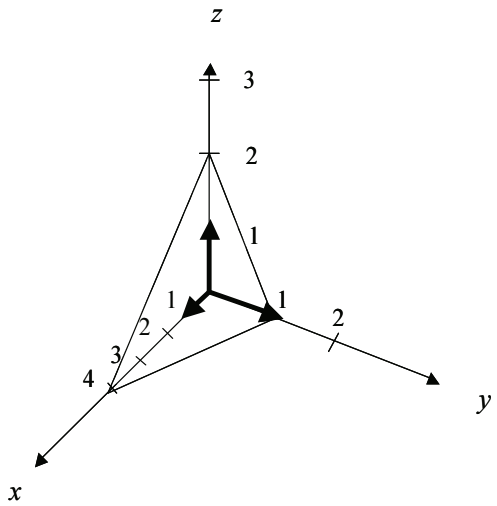


Figure 1-2 : Les indices de Miller

La figure (1-3) représente les plans (100), (110) et (111). L'ensemble des plans $(\bar{1}\,00)$, $(0\bar{1}\,0)$, $(00\bar{1})$, c'est-à-dire l'ensemble des faces du parallélépipède, s'écrit {100}. La direction perpendiculaire à un plan (*h k l*) s'écrit [*h k l*].

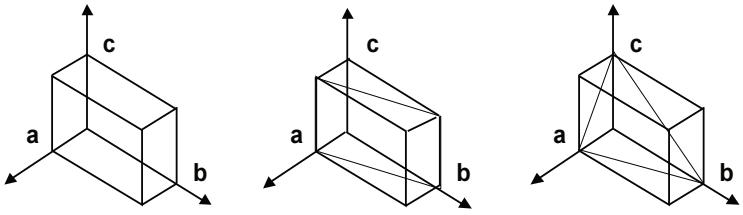


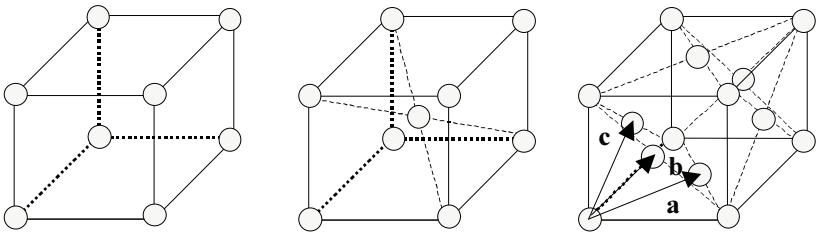
Figure 1.3 : Indices des principaux plans réticulaires

On définit le plan réticulaire par les indices (*h*, *k*, *l*) appelés *indices de Miller*. Sur la figure (1-2) par exemple, les valeurs de *x*, *y*, *z* sont respectivement 4, 1, 2, donc celles de 1/*x*, 1/*y*, 1/*z* sont 1/4, 1, 1/2 et par suite celles de *h*, *k*, *l* sont 1, 4, 2. Le plan représenté sur la figure est le plan (1, 4, 2).

1.1.5. *Le système cubique*

Le système cubique est le système dans lequel cristallisent la grande majorité des semiconducteurs et en particulier le silicium, le germanium, la plupart des composés III-V et certains composés II-VI. Il comprend trois réseaux différents

représentés sur la figure (1-4). Le réseau *cubique simple* (cs) est constitué par un noeud à chaque sommet du cube, ce réseau n'existe pas dans la nature. Le réseau *cubique centré* (cc) est constitué par un noeud à chaque sommet du cube et un noeud au centre du cube, c'est le réseau des métaux alcalins, avec un seul atome par cellule élémentaire. Le réseau *cubique faces centrées* (cfc) est constitué par un noeud à chaque sommet du cube et un noeud au centre de chaque face, c'est le réseau des métaux nobles et des gaz rares, avec un seul atome par cellule élémentaire. En ce qui concerne la maille élémentaire, nous avons vu qu'elle ne devait contenir qu'une seule cellule, c'est-à-dire qu'un seul noeud du réseau. Il en résulte que cette maille élémentaire n'est bâtie sur les arêtes du cube que pour le réseau cubique simple. Dans le réseau cubique faces centrées par exemple, la maille élémentaire est bâtie sur les vecteurs **a** , **b** , **c** représentés sur la figure(1-4).



a) Cubique simple b) Cubique centré c Cubique faces centrées

Figure 1-4 : Le système cubique

Le réseau du silicium est celui du diamant, il est constitué de deux réseaux cubiques faces centrées, imbriqués, décalés l'un de l'autre du quart de la diagonale principale (Fig. 1-5). Bien que le silicium soit un matériau monoatomique, la cellule élémentaire contient deux atomes de silicium, les atomes en position (0,0,0) et (1/4,1/4,1/4). Chaque atome a une coordination tétraédrique et établit des liaisons covalentes avec chacun de ses quatre voisins.

Le réseau de la blende ZnS est une variante du précédent, il est constitué de deux réseaux cubiques faces centrées, l'un de Zn l'autre de S, décalés du quart de la diagonale principale. C'est le réseau dans lequel cristallisent la plupart des composés III-V et certains composés II-VI. Chaque atome III a une coordination tétraédrique avec quatre atomes V et vice versa. Les atomes étant différents, les liaisons sont partiellement covalentes et partiellement ioniques. La cellule élémentaire contient un atome III et un atome V.

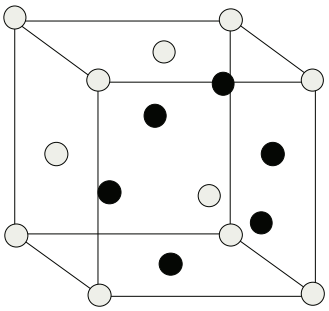


Figure 1-5 : Le réseau du silicium

1.1.6. *Réseau réciproque - Zones de Brillouin*

Le cristal étant périodique tous les paramètres internes du cristal, comme la densité d'électrons par exemple, ont la périodicité du réseau

$$\rho(\mathbf{r}+\mathbf{t})=\rho(\mathbf{r}) \tag{1-3}$$

Considérons pour simplifier le problème, un réseau à une dimension. La densité d'électrons s'écrit

$$\rho(x+t)=\rho(x)$$

avec $t=na$, où a est la périodicité.
La fonction $\rho(x)$ étant périodique, on peut la décomposer en série de Fourier

$$\rho(x)=\rho_o+\sum_{p>0}\left[C_p\cos(p\frac{2\pi}{a}x)+S_p\sin(p\frac{2\pi}{a}x)\right] \tag{1-4}$$

Le coefficient $2\pi/a$ dans l'argument des fonctions trigonométriques réalise la condition $\rho(x+na)=\rho(x)$. En effet :

$$\rho(x+na)=\rho_o+\sum_{p>0}\left[\ C_p\cos(p\frac{2\pi}{a}(x+na))+S_p\sin(p\frac{2\pi}{a}(x+na))\ \right]$$

$$=\rho_o+\sum_{p>0}\left[\ C_p\cos(p\frac{2\pi}{a}x)+S_p\sin(p\frac{2\pi}{a}x)\ \right]=\rho(x)$$

Le point d'abscisse $p2\pi/a$ est un point de l'espace de Fourier du cristal, que l'on appelle aussi *espace réciproque*, p étant un nombre entier cet espace est périodique de période $2\pi/a$. Ainsi, au réseau de périodicité a que l'on appellera désormais réseau direct ou *réseau de Bravais,* correspond un réseau de périodicité $A=2\pi/a$ que l'on appelle *réseau réciproque*. La zone comprise entre $-a/2$ et $+a/2$ dans le réseau direct constitue la maille élémentaire du réseau direct, la zone comprise entre $-\pi/a$ et $+\pi/a$ dans le réseau réciproque constitue la maille élémentaire du réseau réciproque, on l'appelle 1ère *zone de Brillouin.* Au même titre que le réseau direct est parfaitement défini par la maille élémentaire et la périodicité, le réseau réciproque est parfaitement défini par la 1ère zone de Brillouin et la périodicité.

Les grandeurs physiques d'un cristal sont périodiques dans l'espace direct de sorte qu'il suffit de les représenter dans la maille élémentaire pour les connaître dans tout le cristal. Il en est de même pour leurs images dans l'espace réciproque, il suffira de les représenter dans la 1ère zone de Brillouin. Toutes les propriétés des cristaux seront donc étudiées dans cette zone puis étendues à l'ensemble du cristal.

L'extension à trois dimensions de l'analyse de Fourier précédente, ne présente aucune difficulté.

$$\rho(\mathbf{r}) = \sum_k \rho_k e^{i\mathbf{kr}} \tag{1-5}$$

De même qu'à un réseau direct linéaire de périodicité a correspond un réseau réciproque linéaire de périodicité A tel que $A.a=2\pi$, à un réseau direct à trois dimensions de vecteurs de base $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ correspond un réseau réciproque à trois dimensions de vecteurs de base $\mathbf{A}, \mathbf{B}, \mathbf{C}$ tels que

$$\mathbf{A}\cdot\mathbf{a}=2\pi \quad \mathbf{B}\cdot\mathbf{b}=2\pi \quad \mathbf{C}\cdot\mathbf{c}=2\pi$$

Les vecteurs du réseau réciproque sont définis par les relations précédentes et de plus par les conditions $\mathbf{A}\cdot\mathbf{b}=\mathbf{A}\cdot\mathbf{c}=0$, $\mathbf{B}\cdot\mathbf{a}=\mathbf{B}\cdot\mathbf{c}=0$, $\mathbf{C}\cdot\mathbf{a}=\mathbf{C}\cdot\mathbf{b}=0$, soit

$$\mathbf{A}=2\pi\frac{\mathbf{b}\times\mathbf{c}}{\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}} \quad \mathbf{B}=2\pi\frac{\mathbf{c}\times\mathbf{a}}{\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}} \quad \mathbf{C}=2\pi\frac{\mathbf{a}\times\mathbf{b}}{\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}} \tag{1-6}$$

1.1.7. *Densité d'états dans l'espace réciproque*

Les vecteurs $\mathbf{a}, \mathbf{b}, \mathbf{c}$ et $\mathbf{A}, \mathbf{B}, \mathbf{C}$ étant respectivement les vecteurs du réseau direct et du réseau réciproque, les volumes des mailles élémentaires du réseau direct et du réseau réciproque sont donnés par

$$V_d=\mathbf{a}\cdot\mathbf{b}\times\mathbf{c} \tag{1-7}$$

$$V_r=\mathbf{A}\cdot\mathbf{B}\times\mathbf{C} \tag{1-8}$$

Il y a un point k par maille élémentaire de l'espace réciproque de sorte que la densité d'états dans l'espace réciproque est donnée par $\upsilon(\mathbf{k})=1/V_r$. Mais chaque maille élémentaire de l'espace réciproque correspond à une maille élémentaire de l'espace direct, c'est-à-dire à un volume V_d de cristal. Ainsi la densité d'états dans l'espace réciproque, par unité de volume du cristal, est donnée par

$$g(\mathbf{k})=\frac{\upsilon(\mathbf{k})}{V_d}=\frac{1}{V_d V_r}=\frac{1}{(\mathbf{a}\cdot\mathbf{b}\times\mathbf{c})(\mathbf{A}\cdot\mathbf{B}\times\mathbf{C})} \tag{1-9}$$

Ainsi en explicitant les vecteurs de base $\mathbf{A}, \mathbf{B}, \mathbf{C}$ et en prenant en compte la double dégénérescence de spin des états électroniques, la densité d'états dans l'espace des k par unité de volume de cristal est donnée par

$$g(\mathbf{k})=\frac{2}{(2\pi)^3} \tag{1-10}$$

D'une manière plus générale dans un système à n dimensions, la densité d'états dans l'espace réciproque est donnée par

$$g(\mathbf{k})=\frac{2}{(2\pi)^n} \tag{1-11}$$

1.1.8. *Théorème de Bloch*

En raison de la périodicité de la structure cristalline, le potentiel cristallin que voit un électron est périodique

$$V(\mathbf{r}+\mathbf{t})=V(\mathbf{r}) \tag{1-12}$$

Il en résulte que les propriétés des électrons sont invariantes dans une translation du réseau. Ainsi les fonctions d'onde $\psi(\mathbf{r})$ et $\psi(\mathbf{r}+\mathbf{t})$ représentent des états électroniques identiques. Ces fonctions d'onde ne peuvent par conséquent différer que d'un terme constant de module 1, soit

$$\psi(\mathbf{r}+\mathbf{t})=\lambda(\mathbf{t})\psi(\mathbf{r}) \tag{1-13}$$

Considérons par exemple la direction x, la condition s'écrit

$$\psi(x+na)=\lambda_n\ \psi(x) \tag{1-14}$$

soit $n = n'+n''$, on peut écrire

$$\psi(x+na)=\psi\left[x+(n'+n'')\,a\right]=\lambda_{n'+n''}\ \psi(x) \tag{1-15}$$

D'autre part

$$\begin{aligned}\psi\left[x+(n'+n'')a\right]&=\psi\left[(x+n'a)+n''a\right]\\&=\lambda_{n''}\,.\,\psi(x+n'a)\\&=\lambda_{n''}\,.\,\lambda_{n'}\,.\,\psi(x)\end{aligned}$$

Il en résulte que le facteur λ_n de module 1 doit vérifier la condition

$$\lambda_{n'+n''}\ =\lambda_{n'}\,.\,\lambda_{n''} \tag{1-16}$$

Cette relation est vérifiée par une fonction exponentielle, de sorte que la condition de périodicité de la fonction d'onde s'écrit

$$\psi(x+na)=e^{ik.na}\ \ \psi(x) \tag{1-17}$$

A trois dimensions cette relation s'écrit

$$\psi(\mathbf{r}+\mathbf{t})=e^{i\mathbf{k}\cdot\mathbf{t}}\psi(\mathbf{r}) \tag{1-18}$$

Dans le vide la fonction d'onde d'un électron est une onde plane caractérisée par une amplitude constante et un vecteur propagation $\mathbf{k}$

$$\psi(\mathbf{r})=Ae^{i\mathbf{k}\cdot\mathbf{r}} \tag{1-19}$$

Dans le réseau périodique, la fonction d'onde de l'électron n'a plus une amplitude constante en raison des variations de potentiel associées aux ions du réseau. Cette fonction d'onde peut s'écrire

$$\psi(\mathbf{r})=u(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}} \tag{1-20}$$

L'amplitude de la fonction d'onde s'écrit par conséquent

$$u(\mathbf{r})=\psi(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}} \tag{1-21}$$

Compte tenu de la relation (1-18), on peut écrire

$$\psi(\mathbf{r})=\psi(\mathbf{r}+\mathbf{t})e^{-i\mathbf{k}\cdot\mathbf{t}} \tag{1-22}$$

Ou, en portant dans l'expression (1-21)

$$u(\mathbf{r})=\psi(\mathbf{r}+\mathbf{t})e^{-i\mathbf{k}.(\mathbf{r}+\mathbf{t})}=u(\mathbf{r}+\mathbf{t})$$

Il en résulte que la fonction d'onde d'un électron dans un réseau périodique est de la forme

$$\psi(\mathbf{r})=u(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}} \tag{1-23}$$

avec $u(\mathbf{r}+\mathbf{t})=u(\mathbf{r})$. Cette fonction d'onde a été proposée par Bloch en 1928. On dit que la fonction d'onde d'un électron dans un réseau périodique est une *onde de Bloch*. Sa spécificité réside dans le fait que son amplitude a la périodicité du réseau. Les énergies E sont quantifiées quand on résout l'équation de Schrödinger $H\psi = E\psi$ pour décrire un état stationnaire d'un électron dans le cristal. Cette quantification est liée à la forme du potentiel auquel est soumis l'électron. Le vecteur d'onde **k** est lui aussi quantifié (les valeurs possibles sont en nombre élevé mais fini). Cette propriété est liée au fait que le cristal est fini et qu'il est nécessaire d'écrire des conditions périodiques pour la fonction d'onde. La résolution de Schrödinger conduit finalement à exprimer une relation entre l'énergie E et le vecteur d'onde **k.** Cette relation est appelée relation de dispersion et on peut écrire l'énergie comme une fonction du vecteur d'onde.

On considère dans l'étude des propriétés de conduction uniquement les électrons peu liés aux atomes, les électrons de valence. On suppose qu'ils sont décrits par des paquets d'ondes de Bloch correspondant à des valeurs significatives autour de la valeur $\mathbf{k_0}$ valeur moyenne du vecteur d'ondes. La fonction d'onde dépend alors du temps et s'écrit

$$\psi(\mathbf{r},t)=\int f(\mathbf{k}-\mathbf{k}_0)\cdot e^{i(\mathbf{k}\cdot\mathbf{r}-\omega(\mathbf{k})t)}d\mathbf{k} \tag{1-24}$$

Avec $E=\hbar\omega$

La fonction *f* n'a de valeur significative qu'autour de $\mathbf{k_0}$.

Les fonctions décrites en 1-23 représentent des états stationnaires, c'est-à-dire ayant une énergie bien définie et indépendante du temps. Dans ce cas, l'électron est distribué dans le réseau mais non mobile. Pour décrire un électron mobile, il faut former une superposition d'états stationnaires pour différentes énergies du type (1-24). La fonction d'onde a alors une amplitude qui dépend du temps. On peut montrer que la vitesse de la particule classique associée à ce paquet d'ondes est

$$\mathbf{v}_g=\left(\frac{dE}{d\mathbf{k}}\right)_{\mathbf{k}=\mathbf{k}_0}$$

1.2. Etats électroniques dans les semiconducteurs Structure de bandes d'énergie

1.2.1. *Des orbitales atomiques à la structure de bandes*

a. Orbitales atomiques

Considérons le plus simple de tous les atomes, l'atome d'hydrogène. Il est constitué d'un électron de charge *-e* gravitant autour d'un noyau de charge +e. L'Hamiltonien de l'électron dans son mouvement autour du noyau s'écrit

$$H=-\frac{\hbar^2}{2m}\nabla^2+V(\mathbf{r}) \tag{1-25}$$

où $-\hbar^2\nabla^2/2m$ représente son énergie cinétique et $V(\mathbf{r})$ son énergie potentielle dans le champ coulombien du noyau. L'opérateur ∇^2 est la somme des dérivées secondes partielles, il est aussi noté Δ et appelé Laplacien. L'équation aux valeurs propres s'écrit

$$H\,\phi_{nlm}(r,\theta,\phi)=E_{nl}\,\phi_{nlm}(r,\theta,\phi) \tag{1-26}$$

Les fonctions d'onde $\phi_{nlm}(r,\theta,\phi)$ sont appelées *orbitales atomiques* par analogie avec la notion classique d'orbite. En raison de la symétrie sphérique du potentiel du noyau, ces orbitales atomiques peuvent se mettre sous la forme d'un produit d'une fonction de la distance appelée *fonction radiale* $R_{nl}(r)$ par une fonction angulaire appelée *harmonique sphérique* $y_{lm}(\theta,\phi)$.

$$\phi_{nlm}(r,\theta,\phi)=R_{nl}(r)\,y_{lm}(\theta,\phi) \tag{1-27}$$

n est le nombre quantique principal, l est le nombre quantique azimutal, il représente le moment angulaire de l'électron dans son mouvement autour du noyau. m représente la projection du moment angulaire sur un axe choisi comme axe de quantification. Les valeurs possibles de n sont n= 1,2,..., pour chaque valeur de n il existe n valeurs de l , l = 0,1,..., n-1, pour chaque valeur de l il existe $2l$ +1 valeurs de m, m = -l , -(l -1),..., l -1, l . Dans la terminologie courante les états correspondant aux différentes valeurs de l portent des noms spécifiques.

Etat	s	l = 0
Etat	p	l = 1
Etat	d	l = 2
Etat	f	l = 3

On écrit les différents états électroniques sous la forme 1s ; 2s, 2p ; 3s, 3p, 3d ; ...

A chaque fonction propre ϕ_{nlm} correspond une valeur propre E_{nlm}. Dans le cas de l'atome isolé, en raison de la symétrie sphérique du potentiel les niveaux correspondant aux mêmes valeurs de n et de l sont tous confondus, $E_{nlm} = E_{nl}$ quel que soit m. En outre dans le cas de l'atome d'hydrogène, en raison de la nature en 1/r du potentiel coulombien, les niveaux correspondant aux mêmes valeurs de n sont tous confondus, $E_{nl} = E_n$ quel que soit l . Si le zéro de l'énergie correspond à l'électron complètement séparé du noyau sans énergie cinétique, les niveaux d'énergie E_n sont donnés par

$$E_n=-R\frac{1}{n^2} \tag{1-28}$$

où $R=me^4/2\hbar^2$ = 13,6 eV est la *constante de Rydberg*. Les orbitales s (l=0) et p (l=1) sont représentées sur la figure (1-6). Le rayon de Bohr de l'état fondamental est $a_0=\hbar^2/me^2$ = 0,529 Å.

Remarquons que les énergies sont négatives. En fait, les solutions générales de l'équation de Schrödinger d'une particule soumise à un potentiel s'annulant à l'infini (ce qui est le cas du potentiel électrique auquel est soumis l'électron avec les conditions aux limites généralement choisies) peuvent se classer en deux familles, celles correspondant aux valeurs propres négatives et appelées états liés et celles correspondant aux valeurs propres positives et appelées états de diffusion. Nous ne considérons que les états liés.

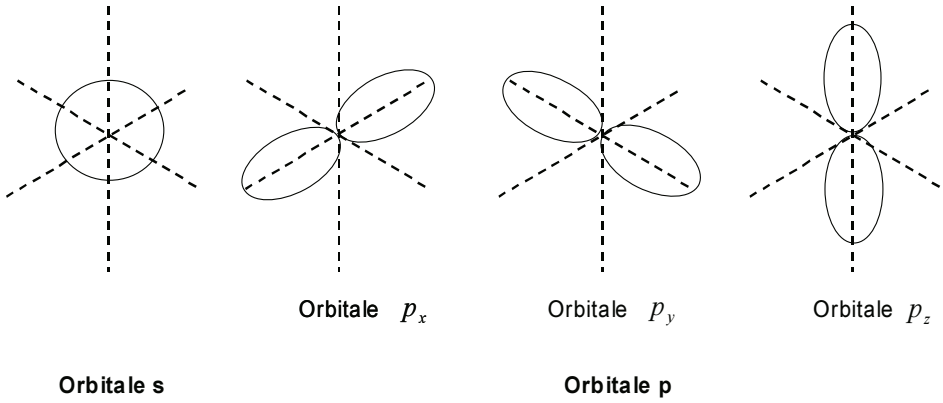


Figure 1.6 : Orbitales atomiques s et p

Dans le cas plus général, un atome de nombre atomique *Z* est constitué d'un noyau de charge +*Ze* et d'un nuage électronique de *Z* électrons. Les électrons de la couche externe sont appelés *électrons de valence*, les électrons des couches internes sont appelés *électrons du cœur*. Les électrons de valence jouent un rôle essentiel d'une part dans les propriétés chimiques des matériaux et d'autre part dans leurs propriétés électriques. Le potentiel que voit un électron de valence est le potentiel coulombien du noyau écranté, d'une part par les électrons du cœur et d'autre part par les autres électrons de valence. Le potentiel résultant peut encore être supposé central, c'est-à-dire à symétrie sphérique, mais il ne varie plus en 1/*r* comme dans le cas unique de l 'atome d'hydrogène. Il en résulte que les états correspondant aux différentes valeurs de *m* restent dégénérés, mais que les états correspondant aux différentes valeurs de *l* ont des énergies différentes. La différence d'énergie *np-ns* augmente à mesure que le potentiel s'éloigne d'une variation en 1/*r*, c'est-à-dire à mesure que le nombre d'électrons de valence augmente. En d'autres termes, sur une même ligne du tableau périodique, l'écart d'énergie np-ns augmente avec le nombre atomique. Sur la ligne du Lithium par exemple, les valeurs sont les suivantes (W.A. Harrison, 1980).

	Li	**Be**	**B**	**C**	**N**	**0**	**F**
E_{2p}- E_{2s}(eV)	-	4,04	5,88	8,52	11,52	15,04	18,84

En ce qui concerne les éléments de la colonne IV qui jouent un rôle essentiel dans les semiconducteurs, cette différence est (en eV) :

C	$E_{2p} - E_{2s}$	=	8,52
Si	$E_{3p} - E_{3s}$	=	7,04
Ge	$E_{4p} - E_{4s}$	=	8,04
Sn	$E_{5p} - E_{5s}$	=	6,56
Pb	$E_{6p} - E_{6s}$	=	6,32

b. Orbitales moléculaires

Considérons la molécule la plus simple, la molécule d'hydrogène. Elle est constituée de deux noyaux autour desquels gravitent deux électrons. Négligeons l'interaction électron-électron, chacun des électrons est dans un potentiel résultant des potentiels coulombiens de chacun des noyaux. L'Hamiltonien d'un électron s'écrit

$$H=-\frac{\hbar^2}{2m}\nabla^2+V(\mathbf{r}-\mathbf{R_1})+V(\mathbf{r}-\mathbf{R_2}) \tag{1-29}$$

où $\mathbf{R_1}$ et $\mathbf{R}_2$ représentent les positions des noyaux. La distance R_1R_2 est de 0,74 Å dans la molécule d'hydrogène. L'équation aux valeurs propres s'écrit

$$H\psi(\mathbf{r})= E\psi(\mathbf{r}) \tag{1-30}$$

Les fonctions d'onde $\psi(\mathbf{r})$ sont appelées *orbitales moléculaires*. Quand les deux atomes sont suffisamment éloignés les états électroniques sont représentés par les orbitales atomiques. Lorsque les deux atomes se rapprochent les états atomiques se couplent pour donner naissance aux états moléculaires qui deviennent les nouveaux états propres du système. On peut alors écrire les orbitales moléculaires sous la forme de combinaisons linéaires d'orbitales atomiques (LCAO : Linear Combinations of Atomic Orbitals).

$$\psi(\mathbf{r})=C_1\phi(\mathbf{r}-\mathbf{R_1})+C_2\phi(\mathbf{r}-\mathbf{R}_2) \tag{1-31}$$

En utilisant le formalisme de Dirac et en appelant $|1s\rangle$ l'état de valence de l'atome d'hydrogène, cette expression s'écrit

$$|\psi\rangle = C_1|1s\rangle_1+C_2|1s\rangle_2 \tag{1-32}$$

En portant cette expression de $|\psi\rangle$ dans l'équation de Schrödinger (1-30) et en multipliant à gauche successivement par ${}_1\langle 1s|$ et ${}_2\langle 1s|$ on obtient les deux équations

$$\begin{aligned}(E_{1s}-E)C_1-V_2C_2=0\\ -V_2C_1+(E_{1s}-E)C_2=0\end{aligned} \tag{1-33}$$

où $E_{1s} = {}_1\langle 1s|\, H|1s\rangle_1 = {}_2\langle 1s|H|1s\rangle_2$ est l'énergie de l'électron dans chacun des atomes isolés et $V_2 = -{}_1\langle 1s|H|1s\rangle_2 = -{}_2\langle 1s|H|1s\rangle_1$ représente l'énergie de couplage entre les deux atomes. Nous avons négligé l'intégrale de recouvrement ${}_1\langle 1s|1s\rangle_2$.

La solution du système est donnée par le déterminant séculaire

$$\begin{vmatrix} E_{1s} - E & -V_2 \\ -V_2 & E_{1s} - E \end{vmatrix} = 0 \qquad (1\text{-}34)$$

c'est-à-dire

$$E_l = E_{1s} - V_2 \qquad (1\text{-}35\text{-a})$$

$$E_a = E_{1s} + V_2 \qquad (1\text{-}35\text{-b})$$

En portant ces valeurs dans le système (1-33) on obtient les coefficients C_1 et C_2 pour les deux états moléculaires $C_{1l} = C_{2l} = 1/\sqrt{2}$, $C_{1a} = -C_{2a} = 1/\sqrt{2}$. Les orbitales moléculaires sont données en fonction des orbitales atomiques par les expressions

$$|\psi\rangle_l = \frac{1}{\sqrt{2}}\left[|1s\rangle_1 + |1s\rangle_2\right] \qquad (1\text{-}36\text{-a})$$

$$|\psi\rangle_a = \frac{1}{\sqrt{2}}\left[|1s\rangle_1 - |1s\rangle_2\right] \qquad (1\text{-}36\text{-b})$$

Les orbitales atomiques $|1s\rangle_1, |1s\rangle_2$ et moléculaires $|\psi\rangle_a, |\psi\rangle_l$ ainsi que les niveaux d'énergie atomiques et moléculaires sont représentés sur la figure (1-7). L'orbitale $|\psi\rangle_l$ est l'*orbitale liante*, l'orbitale $|\psi\rangle_a$ est l'*orbitale antiliante*.

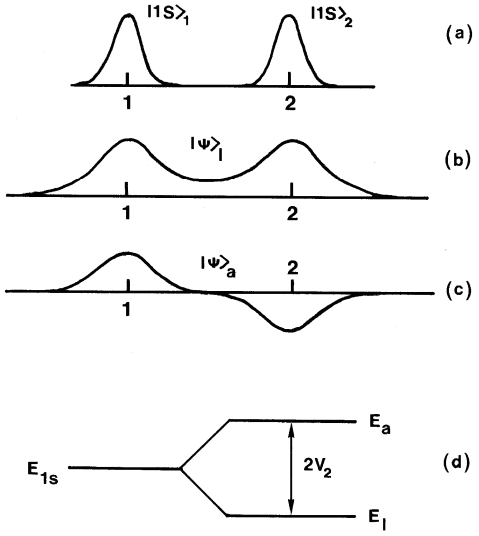


Figure 1.7 : Molécule d'hydrogène, orbitales atomiques et niveaux d'énergie

Il faut noter que dans l'association des deux atomes la densité d'états est conservée et qu'en outre, à l'équilibre thermodynamique, les électrons occupent les états de plus basse énergie. Ainsi les deux états atomiques $|1s\rangle$ dégénérés deux fois compte tenu du spin, donnent naissance à deux états moléculaires dégénérés deux fois. Les deux électrons occupent l'état liant $|\psi\rangle_l$, qui est l'état de plus basse énergie.

Le cas de la molécule d'hydrogène est très simple, mais malheureusement unique dans la mesure où le seul état atomique 1s conditionne la liaison de la molécule . Dans tous les autres cas les états *s* et *p* contribuent à la liaison. Considérons à titre d'exemple la molécule C_2 constituée de deux atomes de carbone. Les états de valence de chacun des atomes de carbone sont 2s et 2p, ils sont donc représentés par quatre orbitales atomiques, une orbitale *s* et trois orbitales

p. Les états moléculaires résultent des couplages de tous ces états, de sorte que de façon générale, et sans préjuger de la valeur des différents coefficients, les orbitales moléculaires peuvent s'écrire

$$\begin{aligned} |\psi\rangle_i = A_1^i |2s\rangle_1 + A_2^i |2p_x\rangle_1 + A_3^i |2p_y\rangle_1 + A_4^i |2p_z\rangle_1 \\ + B_1^i |2s\rangle_2 + B_2^i |2p_x\rangle_2 + B_3^i |2p_y\rangle_2 + B_4^i |2p_z\rangle_2 \end{aligned} \quad (1\text{-}37)$$

En fait si on appelle x l'axe qui porte les deux atomes de carbone, certaines orbitales atomiques ne se couplent pas pour des raisons de symétrie. Les orbitales p_y de chacun des atomes sont uniquement couplées entre elles pour donner naissance à des combinaisons liantes et des combinaisons antiliantes. Il en est de même des orbitales p_z. Les quatre orbitales moléculaires qui en résultent sont appelées *orbitales* π. Il y a deux orbitales liantes π_l et deux orbitales antiliantes π_a. Les orbitales s et p_x sont couplées pour donner naissance à des orbitales que l'on appelle *orbitales* σ. Il y a deux orbitales liantes σ_l et deux orbitales antiliantes σ_a formées de combinaisons linéaires des orbitales $|2s\rangle_1, |2s\rangle_2, |2p_x\rangle_1$ *et* $|2p_x\rangle_2$. On dit que dans la molécule, les états s et p sont *hybridés*. Dans la mesure où les orbitales s sont couplées avec une seule des composantes des orbitales p on dit que l'hybridation est du type sp^1 ou plus simplement sp. La figure (1-8) représente les différentes orbitales atomiques et les états d'énergie correspondants.

Considérons, comme dans la molécule d'hydrogène, la population électronique. Chaque atome de carbone est caractérisé par une couche de valence $2s^2\ 2p^2$, c'est-à-dire dispose, compte tenu de la dégénérescence du spin, de 8 états de valence peuplés par 4 électrons, 2 électrons s et 2 électrons p. La molécule C_2 dispose donc de 2×8 états, 4 états liants dégénérés deux fois et 4 états antiliants dégénérés deux fois, peuplés par 2×4 électrons. Lorsque la molécule est dans son état fondamental, les 8 électrons occupent les orbitales de plus basse énergie. Ces orbitales sont représentées en traits pleins sur la figure (1-8). L'état d'équilibre de la molécule de carbone correspond à une distance internucléaire de l'ordre de 1,3 Å, cette situation est représentée par une verticale en trait discontinu sur la figure (1-8-b). Ainsi dans la molécule C_2 les électrons occupent l'orbitale liante σ_l, les deux orbitales liantes π_l et l'orbitale antiliante σ_a. L'état de la molécule s'écrit $\sigma_l^{\,2}\,\sigma_a^{\,2}\,\pi_l^{\,4}$. En fait, comme le montre la figure, les états π_l et σ'_l sont très proches de sorte que la prise en considération des interactions électron-électron met la molécule dans l'état $\sigma_l^{\,2}\,\sigma_a\,\pi_l^{\,3}\,\sigma'^{\,2}_l$.

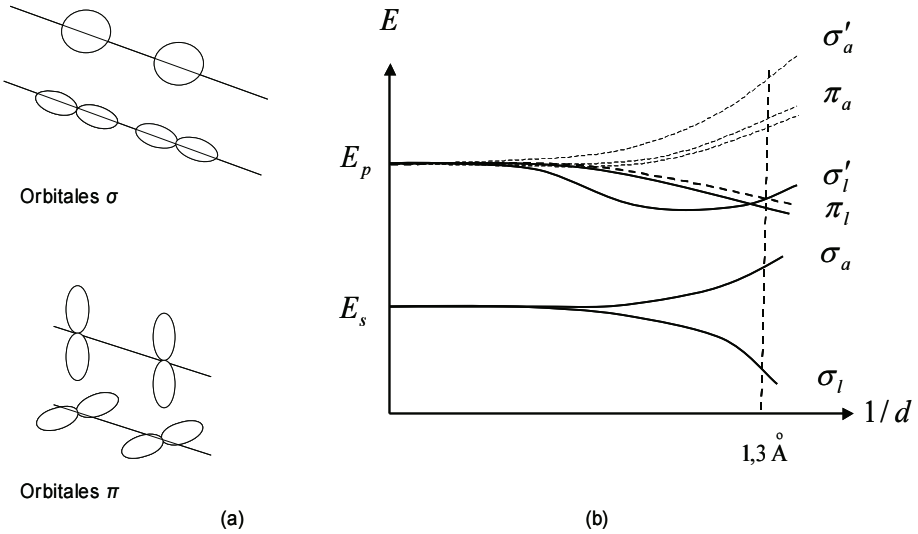


Figure 1.8 : Molécule de carbone. a) Couplage des orbitales. b) Energies en fonction de la distance interatomique

c. Cas du solide : Conducteur - Isolant – Semiconducteur

Par analogie avec la molécule de carbone, résultant de l'association de 2 atomes de carbone, considérons le carbone diamant constitué d'un réseau de N atomes de carbone. Le raisonnement développé pour la molécule peut être étendu au cas du solide avec maintenant une densité d'atomes de l'ordre de 10^{22} cm^{-3}.

Considérons des atomes de carbone arrangés aux noeuds d'un réseau périodique, mais avec une maille très grande. Tant que la distance entre atomes est grande l'état de valence de chaque atome est $2s^2$ $2p^2$, la structure ramenée à deux dimensions, pour des raisons évidentes de simplicité, est représentée sur la figure (1-9-a), les états d'énergie correspondants sont représentés par les énergies E_s et E_p sur la figure (1-9-c). Rapprochons homothétiquement les atomes les uns des autres. Les états de valence se couplent, et ceci d'autant plus que la distance interatomique diminue. La structure est représentée sur la figure (1-9-b), l'évolution des états d'énergie est schématisée sur la figure (1-9-c). Les états atomiques s'élargissent pour donner naissance non pas à 2×4 états liants et 2×4 états antiliants discrets, comme dans la molécule, mais à $N \times 4$ états liants et $N \times 4$ états antiliants répartis en deux ensembles, disjoints ou non suivant la distance internucléaire. Ces deux ensembles d'états constituent dans l'espace des énergies, des bandes permises pour les électrons. L'évolution des états liants et antiliants avec la distance atomique est représentée sur la figure (1-9-c).

Considérons maintenant la population électronique en élargissant le problème à tous les éléments de la colonne IV du tableau périodique. Tous ces éléments ont la même configuration électronique ns^2 np^2, leur spécificité est de disposer de 8 états occupés par 4 électrons. L'occupation des bandes d'énergie est schématisée par des grisées sur la figure (1-9-c).

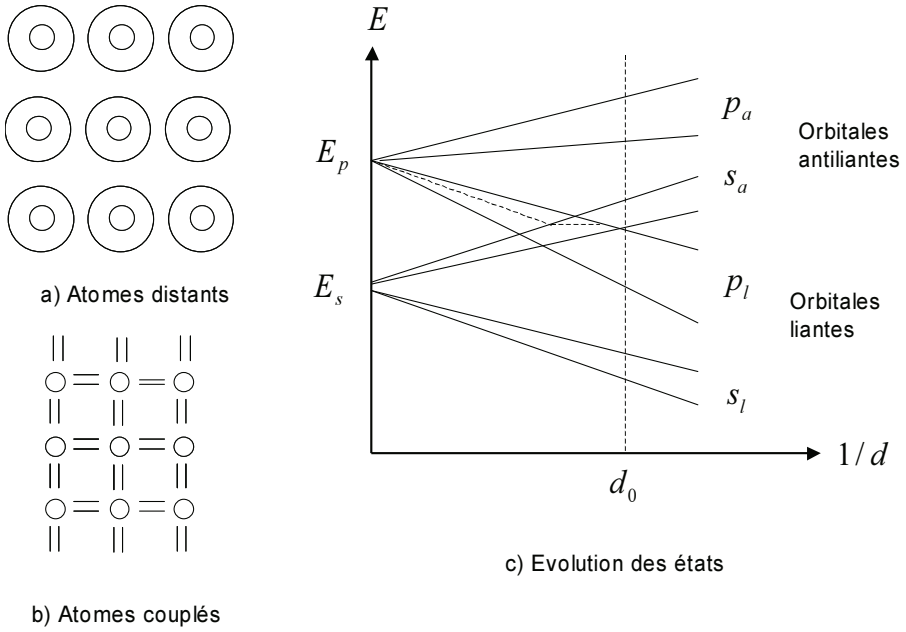


Figure 1-9 : Energies des états en fonction de la distance interatomique

Lorsque $1/d$ tend vers 0, les atomes sont dissociés, les états d'énergie E_s sont occupés par $2N$ électrons et ceux d'énergie E_p sont occupés par $2N$ électrons. Lorsque $1/d$ augmente les états *s* donnent naissance à N états liants occupés par N électrons et N états antiliants occupés par N électrons, les états *p* donnent naissance à $3N$ états liants occupés par $2N$ électrons et $3N$ états antiliants vides. Pour une distance d donnée la population des bandes permises est fonction de la différence entre l'éclatement *s-p* dans l'atome isolé et l'éclatement liant-antiliant dans la liaison chimique entre atomes du cristal.

Examinons maintenant les propriétés électriques du matériau, elles sont fonction des populations électroniques des différentes bandes permises. La conduction électrique résulte du déplacement des électrons à l'intérieur de chaque bande. Sous l'action du champ électrique appliqué au matériau, l'électron acquiert un gain d'énergie cinétique dans la direction du champ (en sens opposé) et se déplace dans le cristal en passant d'une orbitale liante (antiliante) sur une autre orbitale liante (antiliante). Ce transfert est rendu possible par le recouvrement des orbitales voisines et est d'autant plus facile que le recouvrement est important, c'est-

à-dire que les orbitales sont délocalisées. Considérons tout d'abord une bande vide, il est évident qu'elle ne participe pas au courant électrique en raison simplement de l'absence d'électrons. Considérons une bande pleine, elle ne participe pas davantage à la conduction électrique mais pour une toute autre raison. L'électron se déplace sous l'action du champ électrique en raison du gain d'énergie ΔE que lui communique le champ dans sa direction, mais ce déplacement n'est possible que dans la mesure où l'électron d'énergie E trouve, dans la bande, une place disponible à l'énergie $E + \Delta E$, or ce n'est pas le cas pour une bande pleine. Ainsi un matériau qui ne dispose que de bandes pleines et de bandes vides est un isolant. Un conducteur est un matériau dans lequel une bande est partiellement occupée. Il est évident que si la bande est presque vide la conduction est proportionnelle au nombre d'électrons. Si par contre la bande est presque pleine, la conduction est proportionnelle au nombre de places vides. La conduction électrique sera ainsi d'autant meilleure que la population de la bande se rapprochera de 50%.

Revenons sur la figure (1-9-c), dans la partie gauche de la figure, correspondant à des distances interatomiques grandes ($1/d < 1/d_o$), l'éclatement liant-antiliant est inférieur à l'éclatement atomique *s-p*. Dans ce cas les états s_l et s_a liants et antiliants sont pleins, les états p_l liants sont occupés aux 2/3. Cette bande incomplète est spécifique de la conduction métallique, le matériau est conducteur. Dans la partie droite de la figure ($1/d > 1/d_o$), la distance interatomique est faible, l'éclatement liant-antiliant est plus important que l'éclatement atomique *s-p*, et par suite la bande s_a a une énergie supérieure à la bande p_l. La bande s_a se vide alors de ses N électrons qui vont occuper les N places qui restent disponibles dans la bande p_l. Il en résulte que tous les états liants sont occupés et tous les états antiliants sont vides. Le matériau est isolant au zéro degré absolu. La dernière bande pleine est la *bande de valence,* la première bande vide est la *bande de conduction*. Il existe une bande interdite entre le sommet de la bande de valence et le bas de la bande de conduction, cette bande interdite est appelée le *gap* du matériau, sa valeur s'écrit E_g.

La différence entre l'isolant et le semiconducteur est moins nette que la différence entre l'isolant et le conducteur. Cette différence est essentiellement liée au rapport qui existe entre le gap du matériau et l'énergie d'agitation thermique des électrons à la température ambiante. Lorsque E_g est très grand, l'énergie thermique, à la température ambiante, n'est pas suffisante pour exciter un nombre conséquent d'électrons depuis la bande de valence vers la bande de conduction, le matériau est alors isolant ($E_g > 200\ kT$). Par contre lorsque le gap diminue ($E_g < 100\ kT$), un certain nombre d'électrons sont excités dans la bande de conduction par agitation thermique et le matériau présente une conductivité qui, sans être comparable à celle d'un métal, peut devenir appréciable. Le matériau est dit semiconducteur.

Il faut noter en outre que l'éclatement liant-antiliant représenté sur la figure (1-9-c), qui conditionne les largeurs des bandes de valence et de conduction et par suite le gap du matériau, est essentiellement fonction de la taille des atomes et de la distance interatomique. Or cette distance interatomique est différente suivant la direction considérée dans le cristal. Dans la structure tétraédrique du silicium par

exemple, la distance interatomique est plus faible dans la direction (111) que dans la direction (100). Ainsi la structure de bandes d'énergie du matériau sera représentée par les courbes $E_v(\mathbf{k})$ et $E_c(\mathbf{k})$ qui représentent les énergies des états liants et antiliants en fonction du vecteur de propagation $\mathbf{k}$.

La spécificité des semiconducteurs est d'avoir, au zéro degré absolu, une bande de valence pleine et une bande de conduction vide. Ceci résulte simplement du fait que dans la cellule élémentaire le nombre d'électrons périphériques est exactement égal à la moitié du nombre de places disponibles. Il est clair que cette condition peut être réalisée non seulement à partir des éléments de la colonne IV mais aussi avec des composés binaires du type III-V, II-VI ou I-VII, tels que GaAs, ZnSe ou CuBr, ou des composés ternaires du type I-III-VI_2 ou II-IV-V_2, tels que $AgGaS_2$ ou $ZnSiP_2$. Cette condition peut aussi être réalisée avec des alliages ternaires tels que $GaAs_xP_{1-x}$ ou $Cd_xHg_{1-x}Te$, ou des alliages quaternaires tels que $In_xGa_{1-x}As_yP_{1-y}$.

Le diagramme de la figure (1-9-c) est en fait très schématique et doit être modifié pour tenir compte de la nature réelle du cristal. On obtient la structure de bande du cristal en développant les fonctions d'ondes électroniques sur la base des orbitales atomique en tenant compte des deux propriétés essentielles de la structure cristalline, que sont la symétrie de la maille élémentaire et la symétrie de translation du réseau.

La première propriété nous amène à constater que dans les matériaux à structure cubique, du type diamant ou zinc-blende (ZnS), chaque atome établit des liaisons identiques avec chacun de ses quatre voisins. On simplifiera donc l'écriture du problème en tenant compte de cette symétrie dans la construction des fonctions de base. C'est la raison pour laquelle on développe les fonctions d'ondes électroniques du cristal non pas sur la base des orbitales s et p des différents atomes mais sur la base des orbitales de liaison qui résultent d'une hybridation sp^3 des orbitales atomiques. Les orbitales atomiques et les orbitales de liaison sont des représentations équivalentes, mais les secondes tiennent compte de la symétrie de la maille élémentaire. La coordination étant tétraédrique, la base naturelle est une base de quatre orbitales de liaison dans lesquelles la densité de charge est maximum dans la direction des quatre plus proches voisins. Les directions des quatre plus proches voisins sont les directions $(111),(1\,\bar{1}\,\bar{1}),(\bar{1}\,1\,\bar{1})$ et $(\bar{1}\,\bar{1}\,1)$ de sorte que les orbitales de liaison résultant de l'hybridation sp^3 des orbitales atomiques s'écrivent respectivement

$$|h_1\rangle = \frac{1}{2}\left[|s\rangle + |p_x\rangle + |p_y\rangle + |p_z\rangle\right] \qquad \text{(1-38-a)}$$

$$|h_2\rangle = \frac{1}{2}\left[|s\rangle + |p_x\rangle - |p_y\rangle - |p_z\rangle\right] \qquad \text{(1-38-b)}$$

$$|h_3\rangle = \frac{1}{2}\left[|s\rangle - |p_x\rangle + |p_y\rangle - |p_z\rangle\right] \qquad \text{(1-38-c)}$$

$$|h_4\rangle = \frac{1}{2}\left[|s\rangle - |p_x\rangle - |p_y\rangle + |p_z\rangle\right] \qquad \text{(1-38-d)}$$

L'indice 3 de l'écriture sp^3 signifie en quelque sorte que dans chacune des liaisons, l'électron passe 1/4 du temps sur l'état *s* et 3/4 du temps sur l'état *p*.

Pour chaque paire d'atomes voisins les orbitales hybrides se couplent pour donner naissance aux orbitales liantes et antiliantes avec 2 électrons par liaison (8 électrons pour les quatre liaisons du tétraèdre) qui vont se localiser sur les orbitales liantes. Ainsi pour chacune des liaisons on retrouve une situation analogue à la molécule d'hydrogène. Les orbitales liantes et antiliantes s'écrivent sous la forme de combinaisons linéaires des orbitales hybrides dans la direction des deux atomes considérés

$$|\psi\rangle = C|h\rangle + C'|h'\rangle \tag{1-39}$$

Comme pour la molécule d'hydrogène (Eq. 1-35 et 36), les orbitales liantes et antiliantes et les niveaux d'énergie correspondants sont donnés par

$$|\psi\rangle_l = \frac{1}{\sqrt{2}}\left[|h\rangle + |h'\rangle\right] \tag{1-40-a}$$

$$|\psi\rangle_a = \frac{1}{\sqrt{2}}\left[|h\rangle - |h'\rangle\right] \tag{1-40-b}$$

et

$$E_l = E_h + V_2 \tag{1-41-a}$$

$$E_a = E_h - V_2 \tag{1-41-b}$$

où $E_h = (E_s + 3E_p)/4$ est l'énergie de l'état hybride, qui n'a rien de réel mais constitue une étape intermédiaire dans le calcul et V_2 l'énergie de couplage entre les deux orbitales hybrides, c'est-à-dire l'élément de matrice

$$V_2 = \langle h|H|h'\rangle = \langle h'|H|h\rangle$$

A ce stade du calcul les électrons dans le cristal disposent de deux niveaux d'énergie E_l et E_a chacun étant dégénéré deux fois pour chaque liaison, c'est-à-dire au total $2N$ fois si N est le nombre d'atomes du cristal. Ces états d'énergie sont représentés sur la figure (1-10).

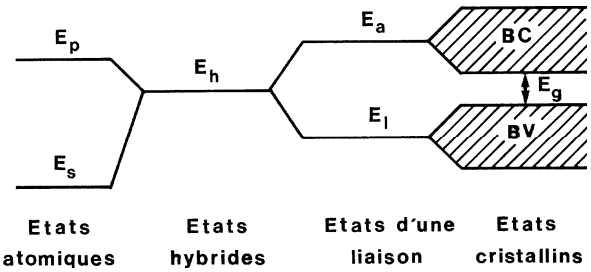


Figure 1.10 : Transformation des états atomiques en états cristallins

$$\left|\psi_k^l(\mathbf{r})\right\rangle = \sum_i e^{i\mathbf{k}\cdot\mathbf{r}_i}\left|\psi_\ell(\mathbf{r}-\mathbf{r}_i\right\rangle \qquad \text{(1-42-a)}$$

$$\left|\psi_k^a(\mathbf{r})\right\rangle = \sum_i e^{i\mathbf{k}\cdot\mathbf{r}_i}\left|\psi_a(\mathbf{r}-\mathbf{r}_i\right\rangle \qquad \text{(1-42-b)}$$

où i est sommé sur le réseau de Bravais du cristal.

Les termes $\left|\psi_\ell(\mathbf{r}-\mathbf{r}_i\right\rangle$ représentent les orbitales liantes et antiliantes de la paire d'atomes d'abscisse $\mathbf{r}_i$ et $\mathbf{k}$ est un vecteur de l'espace réciproque du cristal. Ces fonctions sont appelées *sommes de Bloch d'orbitales liantes et antiliantes.*
Les niveaux d'énergie sont calculés par les éléments de matrice du type

$$E_k^l = \left\langle\psi_k^l(\mathbf{r})\middle|\mathrm{H}\middle|\psi_k^l(\mathbf{r})\right\rangle \Big/ \left\langle\psi_k^l(\mathbf{r})\middle|\psi_k^l(\mathbf{r})\right\rangle \qquad \text{(1-43-a)}$$

$$E_k^a = \left\langle\psi_k^a(\mathbf{r})\middle|\mathrm{H}\middle|\psi_k^a(\mathbf{r})\right\rangle \Big/ \left\langle\psi_k^a(\mathbf{r})\middle|\psi_k^a(\mathbf{r})\right\rangle \qquad \text{(1-43-b)}$$

Ces niveaux d'énergie traduisent respectivement les élargissements des niveaux E_l et E_a en bande de valence et de conduction. Cette évolution est schématisée sur la figure (1-10).

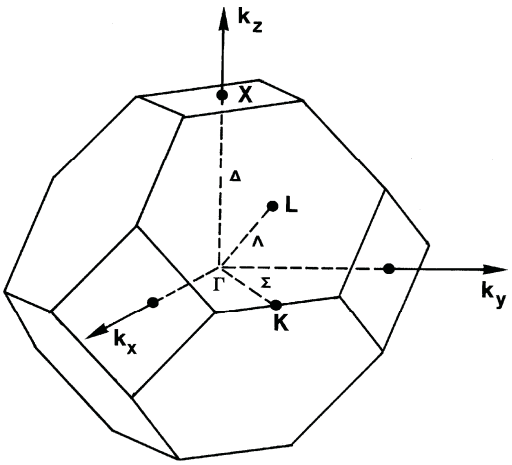


Figure 1.11 : Première zone de Brillouin d'un cristal cubique

Point Γ : centre de zone
Directions Δ : direction 100 et équivalentes
Directions Λ : direction 111 et équivalentes
Directions Σ : direction 110 et équivalentes
Points X : bord de zone dans la direction 100 et les directions équivalentes
Points L : bord de zone dans la direction 111 et les directions équivalentes
Points K : bord de zone dans la direction 110 et les directions équivalentes

Les expressions (1-42 et 43) montrent clairement que les fonctions d'onde électroniques et les niveaux d'énergie correspondants sont fonction du vecteur d'onde de l'électron. Ainsi la structure de bandes d'énergie du semiconducteur doit être représentée dans l'espace réciproque et dans les différentes directions du vecteur d'onde **k** . Dans la pratique on représente les énergies de la bande de valence et de la bande de conduction dans les directions de haute symétrie du vecteur **k** . Ces directions qui portent des noms bien définis sont représentées sur la figure (1-11) pour un matériau cubique.

Compte tenu des expressions (1-43, 42 et 40), les états d'énergie E_k^l et E_k^a sont essentiellement fonction des intégrales de recouvrement des orbitales atomiques entre les différents voisins. Dans la pratique, la prise en considération des premiers et des seconds voisins donne l'essentiel des largeurs des bandes de valence et de conduction. Si ces intégrales de recouvrement sont faibles, les bandes permises sont étroites et par suite le gap du matériau est grand. Le cristal est donc isolant ; c'est le cas du diamant. Au contraire, si les intégrales de recouvrement sont importantes, les bandes permises sont larges et par suite le gap est petit. Ces intégrales de recouvrement augmentent avec la taille des atomes de sorte que d'une manière générale, le gap des semiconducteurs diminue quand on descend le tableau périodique des éléments.

Considérons enfin les semiconducteurs construits sur une même ligne du tableau périodique, c'est-à-dire les semiconducteurs de type IV-IV, III-V, II-VI et I-VII. Le semiconducteur IV-VI est covalent puisque les atomes sont identiques, les autres matériaux sont partiellement ioniques avec une ionicité croissante du III-V vers le I-VII. A mesure que l'ionicité augmente les électrons de valence sont de plus en plus localisés et les intégrales de recouvrement diminuent, le gap du semiconducteur augmente.

1.2.2. *Méthodes de calcul de la structure de bandes d'énergie*

a. Les approximations de base

Un cristal est constitué d'un très grand nombre de particules en interaction, les électrons et les noyaux atomiques, de sorte que le calcul des états d'énergie du système passe nécessairement par un certain nombre d'hypothèses simplificatrices.

L'Hamiltonien total du système s'écrit

$$H = T_e + T_n + V_{ee} + V_{en} + V_{nn} \qquad (1\text{-}44)$$

où T_e et T_n représentent les énergies cinétiques des électrons et des noyaux et V_{ee}, V_{en} et V_{nn} les énergies d'interaction électron-électron, électron-noyau et noyau-noyau respectivement.

La première approximation consiste à limiter les interactions entre particules au terme le plus important que constitue l'interaction coulombienne. L'Hamiltonien du système s'écrit alors

$$H=-\sum_i \frac{\hbar^2}{2m}\nabla_i^2-\sum_I \frac{\hbar^2}{2M_I}\nabla_I^2+\sum_{i<j} \frac{e^2}{|\mathbf{r}_i-\mathbf{r}_j|}-\sum_{I,i} \frac{Z_I e^2}{|\mathbf{R}_I-\mathbf{r}_i|} +\sum_{I<J} \frac{Z_I Z_J e^2}{|\mathbf{R}_I-\mathbf{R}_J|} \tag{1-45}$$

où les indices i, j et I, J se rapportent respectivement aux électrons et aux noyaux, et Z représente la charge nucléaire.

Les états stationnaires d'énergie et les fonctions d'onde du système sont donnés par les solutions de l'équation de Schrödinger

$$H\psi(\mathbf{R},\mathbf{r})=E\psi(\mathbf{R},\mathbf{r}) \tag{1-46}$$

où **R** représente les coordonnées des noyaux et **r** celles des électrons.

Les électrons et les noyaux atomiques ont des masses très différentes de sorte que l'on peut utiliser l'approximation de Born-Oppenheimer pour séparer, comme dans l'étude de l'atome d'hydrogène, l'équation aux valeurs propres des noyaux de celle des électrons. Les états propres du système sont alors caractérisés par des fonctions d'onde produits d'une fonction d'onde électronique par une fonction d'onde nucléaire

$$\psi(\mathbf{R},\mathbf{r})=\psi_n(\mathbf{R})\psi_e(\mathbf{R},\mathbf{r}) \tag{1-47}$$

L'équation de Schrödinger s'écrit

$$\left[T_e+T_n+V_{ee}+V_{en}+V_{nn}\right]\psi_n(\mathbf{R})\psi_e(\mathbf{R},\mathbf{r})=E\psi_n(\mathbf{R})\psi_e(\mathbf{R},\mathbf{r}) \tag{1-48}$$

et se ramène aux deux équations interdépendantes

$$\left[T_e+V_{ee}+V_{en}\right]\psi_e(\mathbf{R},\mathbf{r})=E_e(\mathbf{R})\psi_e(\mathbf{R},\mathbf{r}) \tag{1-49-a}$$

$$\left[T_n+V_{nn}+E_e(R)\right]\psi_n(\mathbf{R})=E\psi_n(\mathbf{R}) \tag{1-49-b}$$

Ces équations constituent les équations de base de ce que l'on appelle *l'approximation adiabatique*. La première donne l'énergie électronique pour une valeur déterminée **R** des coordonnées nucléaires. Cette énergie électronique apparaît ensuite comme contribution potentielle dans l'équation aux valeurs propres du mouvement des noyaux.

Pour étudier les états d'énergie électroniques du cristal on n'utilise que l'équation (1-49-a), les noyaux étant supposés fixes à leur position d'équilibre. Mais cette équation traduit l'évolution d'un système à n corps et demeure un problème très difficile encore non résolu. Pour simplifier ce problème, on se place dans l'*approximation à un électron*. Cette approximation consiste à globaliser les

interactions individuelles électron-électron et à écrire que chaque électron évolue dans un potentiel moyen résultant de la présence de l'ensemble des autres électrons. L'équation de Schrödinger d'un électron s'écrit alors

$$H_{HF}\psi_e(\mathbf{r})=E_e\psi_e(\mathbf{r}) \tag{1-50}$$

où H_{HF} est l'*Hamiltonien de Hartree-Fock* donné par

$$H_{HF}=-\frac{\hbar^2}{2m}\nabla^2-\sum_I\frac{Z_Ie^2}{|\mathbf{r}-\mathbf{R}_I|}+V_{coul}+V_{exch} \tag{1-51}$$

Le premier terme d'énergie potentielle représente interaction électron-noyaux, les termes V_{coul} et V_{exch} globalisent les interactions électron-électron, ces termes sont donnés par

$$V_{coul}=e^2\sum_j\int\frac{|\psi_j(\mathbf{r}')|^2}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}' \tag{1-52}$$

$$V_{exch}\psi(\mathbf{r})=-e^2\sum_j\psi_j(\mathbf{r})\int\frac{\psi_j^*(\mathbf{r}')\psi(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}' \tag{1-53}$$

L'approximation de Hartree-Fock, qui ramène le problème à *n* corps au problème à 1 électron, est dans certains cas une mauvaise approximation comme nous le verrons par exemple dans l'étude des niveaux électroniques de la couche d'inversion d'une structure métal-isolant-semiconducteur. Néanmoins elle constitue la seule façon de résoudre le problème et donne de bons résultats dans le calcul de la structure de bandes d'énergie dans les semiconducteurs.

La résolution de l'équation de Hartree-Fock reste encore un problème mathématique très difficile dans le mesure où elle nécessite une approche auto-cohérente. En effet, les fonctions propres sont fonction de l'Hamiltonien qui est lui-même fonction des fonctions propres par ses composantes V_{coul} et V_{exch} . C'est la raison pour laquelle on simplifie encore le problème en représentant le potentiel que voient les électrons par un terme global $V_c(r)$ que l'on appelle *potentiel cristallin*. Ce potentiel cristallin est construit à partir des potentiels atomiques associés à chaque atome constituant le réseau cristallin, en tenant compte des propriétés de symétrie du cristal.

L'équation s'écrit alors

$$\left[-\frac{\hbar^2}{2m}\nabla^2+V_c(\mathbf{r})\right]\psi_n(\mathbf{k},\mathbf{r})=E_n(\mathbf{k})\psi_n(\mathbf{k},\mathbf{r}) \tag{1-54}$$

Il existe différentes méthodes de résolution de cette équation pour obtenir les états électroniques du cristal.

b. Méthode LCAO

La méthode LCAO (Linear Combinations of Atomic Orbitals), appelée aussi *méthode des liaisons fortes*, consiste à développer les fonctions d'onde du cristal sous forme de combinaisons linéaires d'orbitales atomiques, en tenant compte du théorème de Bloch auquel doivent satisfaire les fonctions d'onde du cristal. La fonction d'onde est écrite sous la forme d'une *somme de Bloch d'orbitales atomiques*. C'est la méthode évoquée dans le paragraphe précédent. Elle donne de bons résultats lorsque les orbitales atomiques sont très localisées autour des noyaux et changent peu lorsque les atomes sont rapprochés pour constituer le cristal. Ceci est vrai pour les états du cœur de l'atome mais l'est beaucoup moins pour les électrons de valence. En outre, en ce qui concerne les états de valence, la méthode est mieux adaptée au calcul des états associés aux combinaisons liantes, c'est-à-dire à la bande de valence, qu'aux états résultant des combinaisons antiliantes, c'est-à-dire à la bande de conduction.

En résumé, cette méthode est bien adaptée au calcul des bandes profondes étroites, résultant de l'élargissement des états du cœur, un peu moins adaptée au calcul de la bande de valence et peu adaptée au calcul de la bande de conduction.

c. Méthode OPW

La méthode LCAO procède de l'idée que les états électroniques dans le cristal sont essentiellement des états atomiques, plus ou moins perturbés par la nature périodique du cristal. La méthode OPW (Orthogonalized Plane Waves) procède de l'idée diamétralement opposée, que les états électroniques dans le cristal sont essentiellement les états de l'électron libre, plus ou moins perturbés par la nature périodique du cristal. Les fonctions d'onde des électrons dans le cristal sont alors développées sur la base des fonctions d'onde des électrons libres, c'est-à-dire d'ondes planes. Il en résulte que le domaine de validité de cette méthode est complémentaire du domaine de validité de la méthode précédente. Cette méthode est bien adaptée à l'étude de la bande de conduction, un peu moins à celle de la bande de valence et pas du tout à celle des états du cœur.

De simples considérations physiques montrent que le développement de la fonction d'onde en ondes planes, sans autres considérations, pose d'importants problèmes de convergence. Plus les états sont localisés, plus le nombre d'ondes planes pour les décrire est important. A la limite, pour décrire les états du cœur, il faut un nombre infini d'ondes planes. On tourne la difficulté en calculant les bandes profondes issues des états du cœur par la méthode LCAO par exemple. Les autres états du cristal, c'est-à-dire les états de valence et de conduction sont des états propres du cristal au même titre que les bandes profondes et par conséquent orthogonaux à ces états du cœur. Dans la mesure où la méthode OPW va servir à calculer uniquement les états de conduction et de valence, on utilise alors une base d'ondes planes constituée de combinaisons orthogonales aux états du cœur, d'où le nom d'*ondes planes orthogonalisées*.

La méthode OPW a beaucoup été utilisée pour calculer les bandes de valence et de conduction des semiconducteurs de type IV-IV et III-V. La principale difficulté

dans cette méthode est qu'elle nécessite la séparation entre les états du cœur d'une part, et les états de valence et de conduction de l'autre. Ceci suppose que les fonctions d'onde du cœur sont très localisées à l'intérieur de la cellule élémentaire et leurs niveaux d'énergie très séparés de ceux de la bande de valence. Ces conditions sont mal remplies dans le cas de matériaux constitués d'atomes ayant des couches d incomplètes. Les énergies des états atomiques d ne sont en fait pas très différentes de celles des états s et p de même nombre quantique principal, de sorte que ces états peuvent difficilement être inclus dans les états de cœur.

d. Méthode du pseudopotentiel

La méthode du pseudopotentiel, comme la méthode OPW, utilise les propriétés d'orthogonalité des états de valence et conduction avec les états du cœur. Mais dans le formalisme du pseudopotentiel l'effet de l'orthogonalité est inclus dans le potentiel sous la forme d'un potentiel équivalent appelé *pseudopotentiel*. L'effet d'orthogonalisation aux états de cœur revient à extraire du potentiel cristallin la contribution rapidement variable de la région du cœur. Le pseudopotentiel V_p est alors lentement variable et se prête bien à une approche du problème en terme de perturbation.

e. Autres méthodes

Les méthodes précédentes consistent à développer les fonctions d'onde du cristal sur une base complète de fonctions du type fonctions de Bloch et à calculer les coefficients du développement en écrivant que la fonction d'onde doit satisfaire à l'équation de Schrödinger. Un autre type de méthode consiste à développer les états du cristal sur une base complète de fonctions solutions de l'équation de Schrödinger dans la cellule unité du cristal. On calcule alors les coefficients du développement en écrivant que la fonction d'onde du cristal satisfait à des conditions aux limites appropriées.

La *méthode cellulaire* consiste à diviser la maille élémentaire en cellules contenant un seul atome. Le potentiel dans chaque cellule a alors une symétrie sphérique ce qui permet de calculer simplement les fonctions de base en séparant la partie radiale des harmoniques sphériques. La difficulté de la méthode réside essentiellement dans la maîtrise des conditions aux limites.

La *méthode APW* (Augmented Plane Waves) a été imaginée pour pallier le problème des conditions aux limites inhérentes à la méthode cellulaire. Le potentiel cristallin est supposé sphérique à l'intérieur de sphères de rayons r_s entourant les atomes, et constant à l'extérieur de ces sphères. Les fonctions d'ondes sont développées en ondes sphériques dans les régions où le potentiel est de type atomique et en ondes planes dans les régions où le potentiel est constant, ces fonctions d'onde sont appelées *ondes planes augmentées.* Elles sont continues en $r=r_s$ et ne présentent de ce fait aucun problème de conditions aux limites. Contrairement aux méthodes LCAO et OPW qui font appel à des techniques numériques relativement simples, la diagonalisation de matrices à coefficients constants, la méthode APW nécessite des techniques numériques plus sophistiquées

et par suite des machines plus performantes. Une variante de cette méthode consiste à utiliser le formalisme de la fonction de Green.

1.2.3. *Caractéristiques de la structure de bandes d'énergie*

a. Semiconducteurs univallée et multivallée

Les différents types de structure de bande sont représentés sur la figure (1-12), suivant les directions de plus haute symétrie de l'espace réciproque, c'est-à-dire suivant les directions Δ (001) et Λ (111).

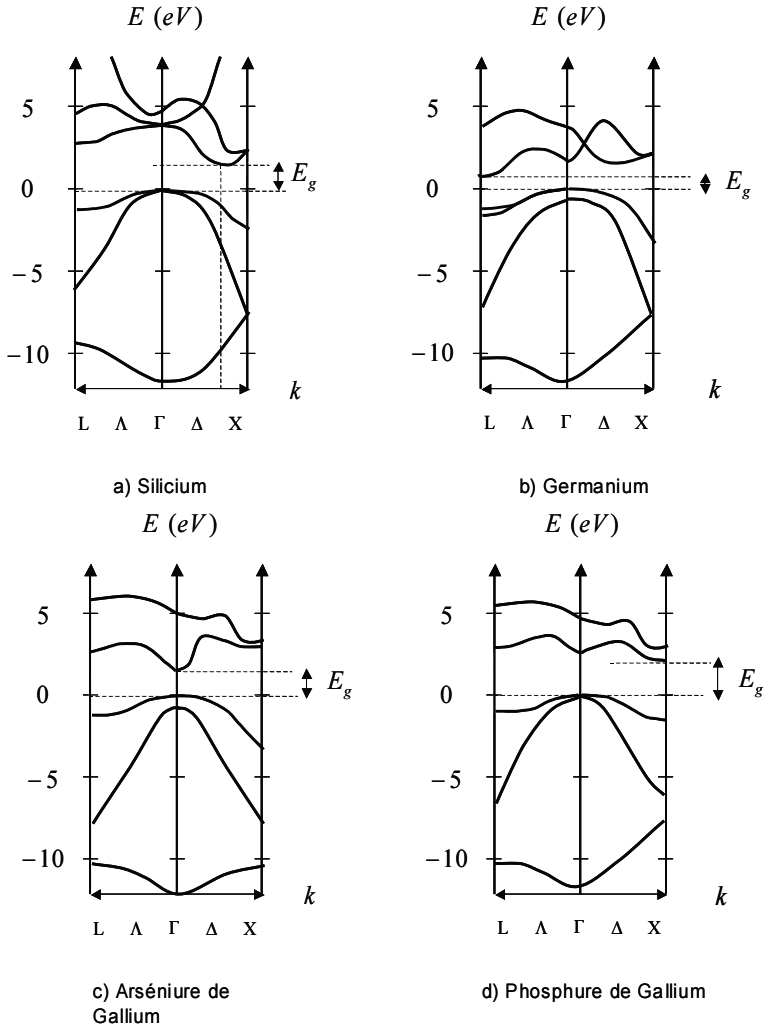


Figure 1-12 : Structures de bandes du silicium, du Germanium du Phosphure de Gallium et de l'Arséniure de Gallium dans les directions de haute symétrie

La bande interdite est grisée, les bandes supérieures sont les bandes de conduction, les bandes inférieures sont les bandes de valence. Au zéro degré absolu la bande de valence est pleine d'électrons, la bande de conduction est vide. A la température ambiante, à laquelle fonctionnent la plupart des composants électroniques, certains électrons sont thermiquement excités depuis la bande de valence et occupent la bande de conduction. Il est évident que les électrons les plus susceptibles de passer de la bande de valence vers la bande de conduction sont les électrons de la bande de valence supérieure. En outre, ces électrons occuperont la bande de conduction inférieure. De plus, parmi les états de ces bandes, ceux qui jouent le rôle essentiel sont respectivement ceux du sommet de la bande de valence et du minimum de la bande de conduction.

Considérons tout d'abord le sommet de la bande de valence, il est caractérisé par deux propriétés essentielles. Il est situé au centre de la zone de Brillouin, c'est-à-dire en k=0, ce point est appelé point Γ. Ce maximum est constitué de la convergence de deux bandes qui sont dégénérées au sommet. Ces deux propriétés sont communes à tous les semiconducteurs à structure cubique. Compte tenu de l'unicité du point Γ dans la première zone de Brillouin, le maximum de la bande de valence est unique.

Considérons maintenant le minimum de la bande de conduction, la figure (1-12) montre que la situation n'est pas unique.

Cas du silicium : le minimum de la bande de conduction est situé dans la direction (001) appelée direction Δ, au point d'abscisse $(00k_o)$ avec k_o=0,85 K_x où K_x représente l'abscisse du point X, limite de la première zone de Brillouin dans la direction Δ. Compte tenu de la structure cubique du silicium, il existe 6 directions équivalentes qui sont $(100), (\bar{1}00), (010), (0\bar{1}0), (001)$ *et* $(00\bar{1})$. La bande de conduction présente donc six minima équivalents, on dit que le silicium est un *semiconducteur multivallée* à 6 vallées Δ. La variation $E(k)$ de l'énergie de la bande de conduction au voisinage du minimum n'est pas isotrope, elle est plus rapide dans le plan perpendiculaire à l'axe considéré que suivant cet axe. Les surfaces d'énergie constante sont des ellipsoïdes de révolution autour de chacun des axes équivalents. Ces ellipsoïdes sont schématisés sur la figure (1-13). Le cas du silicium est unique sur l'ensemble des semiconducteurs utilisés dans les composants électroniques.

Cas du germanium : le minimum de la bande de conduction est situé dans la direction (111) appelée direction Λ, à l'extrémité de la zone de Brillouin au point appelé point L. Il existe 8 directions Λ équivalentes, les directions $(\overset{\pm\pm\pm}{111})$. Dans la mesure où le minimum est situé en bord de zone de Brillouin, les surfaces d'énergie constante sont 8 demi-ellipsoïdes de révolution autour de chacun des axes $(\overset{\pm\pm\pm}{111})$. Les points extrêmes de la zone de Brillouin étant équivalents, on peut représenter ces 8 demi-ellipsoïdes par 4 ellipsoïdes centrés respectivement aux points $(111), (\bar{1}11), (1\bar{1}1)$ *et* $(11\bar{1})$. On dit que le germanium est un *semiconducteur multivallée* à 4 vallées L. Au même titre que le silicium est le seul semiconducteur à 6 vallées Δ, le germanium est le seul semiconducteur à 4 vallées L (Fig. 1-13).

Cas du GaP : Le GaP est un composé III-V, sa bande de conduction présente 6 minima aux points X, c'est-à-dire en bord de zone dans les directions Δ. Les surfaces d'énergie constante sont constituées de 6 demi-ellipsoïdes que l'on peut ramener à trois ellipsoïdes pour les mêmes raisons que le germanium. On dit que le GaP est un *semiconducteur multivallée* à 3 vallées X. Les composés III-V AlP, AlAs et AlSb ont une bande de conduction de ce type (Fig. 1-13).

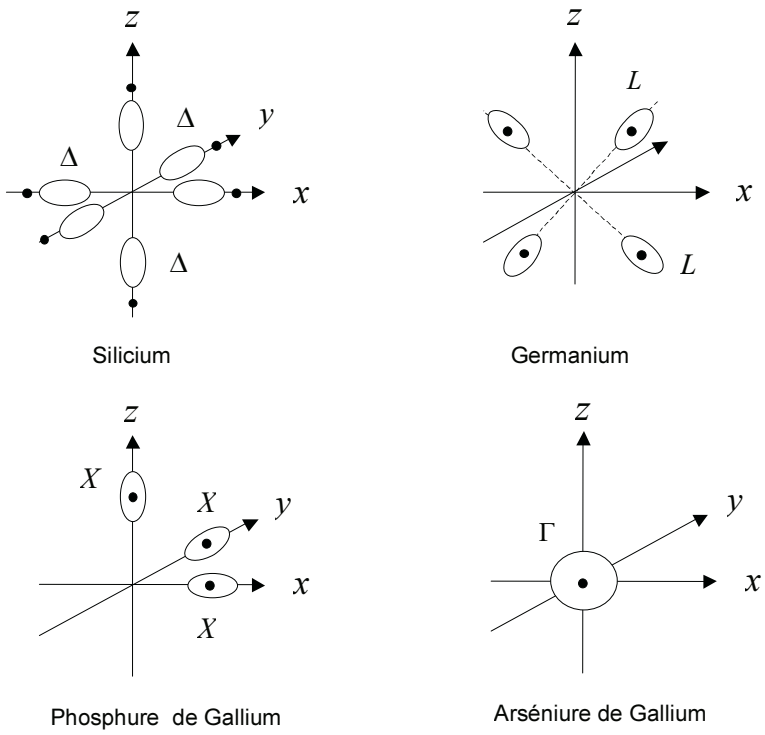


Figure 1-13 : Représentation schématique des surfaces d'énergie constante des bandes de conduction de Si, Ge, GaP, GaAs.

Les minima de la bande correspondent aux centres des ellipsoïdes. Dans hacun des minima les électrons de vecteur k dirigé suivant l'axe de l'ellipsoïde, c'est-à-dire se propageant suivant la direction correspondante, sont caractérisés par la masse effective longitudinale m_l. Les électrons de vecteur k perpendiculaire à cet axe sont caractérisés par la masse effective transverse m_t.

Cas du GaAs : Le minimum de la bande de conduction est situé en k = 0 au point Γ. Il est par conséquent unique. Les surfaces d'énergie constante au voisinage du minimum sont des sphères centrées au point Γ. On dit que le GaAs est un *semiconducteur univallée*.

b. Gap direct - Gap indirect

Considérons le gap des différents semiconducteurs. Le gap est par définition la largeur de la bande interdite, c'est-à-dire la différence d'énergie entre le minimum

absolu de la bande de conduction et le maximum absolu de la bande de valence. Les structures de bandes représentées sur la figure (1-12) font apparaître deux types fondamentaux de semiconducteur. Les semiconducteurs dans lesquels le minimum de la bande de conduction et le maximum de la bande de valence sont situés en des points différents de l'espace des *k*, et les semiconducteurs pour lesquels ces extrema sont situés au même point de l'espace des *k*. Les seconds sont dits à *gap direct*, le prototype en est le GaAs, les premiers sont dits à *gap indirect.*

TABLEAU 1.1. : Energie du gap des différents semiconducteurs

Semiconducteur		**gap** (eV) 4 K	300 K	**Nature du gap**	**Constante diélectrique relative** $\varepsilon_r = \varepsilon / \varepsilon_o$
C	(c)	5,48	5,45	indirect	5,57
Si	(c)	1,169	1,12	"	12
SiC	(6H)	-	2,86	"	9,7
Ge	(c)	0,747	0,66	"	16
AlP	(c)	2,52	2,45	"	9,8
AlAs	(c)	2,24	2,16	"	10,1
AlSb	(c)	1,63	1,60	"	10,3
GaP	(c)	2,35	2,25	"	8,4
AlN	(H)	-	6,28	direct	$\bar{\varepsilon} = 9,14$
GaN	(H)	-	3,39	"	$\varepsilon_{//} = 10,4 \quad \varepsilon_{\perp} = 9,5$
GaAs	(c)	1,52	1,43	"	11,5
GaSb	(c)	0,81	0,68	"	14,8
InN	(H)	-	1,95	"	-
InP	(c)	1,42	1,27	"	12,1
InAs	(c)	0,42	0,36	"	12,5
InSb	(c)	0,237	0,17	"	15,9
ZnO	(H)	3,40	-	"	$\varepsilon_{//} = 8,7 \quad \varepsilon_{\perp} = 7,8$
ZnS	(H)	3,80	3,68	"	$\bar{\varepsilon} = 9,6$
ZnSe	(c)	2,82	2,67	"	9,1
ZnTe	(c)	2,39	2,26	"	8,7
CdS	(H)	2,56	2,42	"	$\bar{\varepsilon} = 9,4$
CdSe	(H)	1,84	1,7	"	$\bar{\varepsilon} = 10$
CdTe	(c)	1,60	1,44	"	9,6
SiO_2		8,8	-		3,9

c : structure cubique *H : structure hexagonale*

Si on se limite aux principaux semiconducteurs que sont les éléments du groupe IV et les composés binaires III-V et II-VI, les matériaux à gap indirect sont Si, Ge, AlP, AlAs, AlSb et GaP, tous les autres ont un gap direct. La nature du gap joue un rôle fondamental dans l'interaction du semiconducteur avec un rayonnement électromagnétique et par suite dans le fonctionnement des composants optoélectroniques (Chap. 9). Les gaps de différents semiconducteurs sont portés dans le tableau (1-1) et sur la figure (1-13').

On remarque, comme nous l'avons déjà précisé, que le gap augmente quand on passe de l'élément IV aux composés III-V et II-VI sur une même ligne du tableau périodique et diminue quand on descend le tableau.

c. Notion de trou

Comme nous l'avons précisé plus haut, lorsque la température est différente de zéro, un certain nombre d'électrons de la bande de valence sont excités dans la bande de conduction. Compte tenu de la nature antiliante des fonctions d'onde électroniques dans la bande de conduction, ces électrons sont des particules quasi-libres dans le semiconducteur.

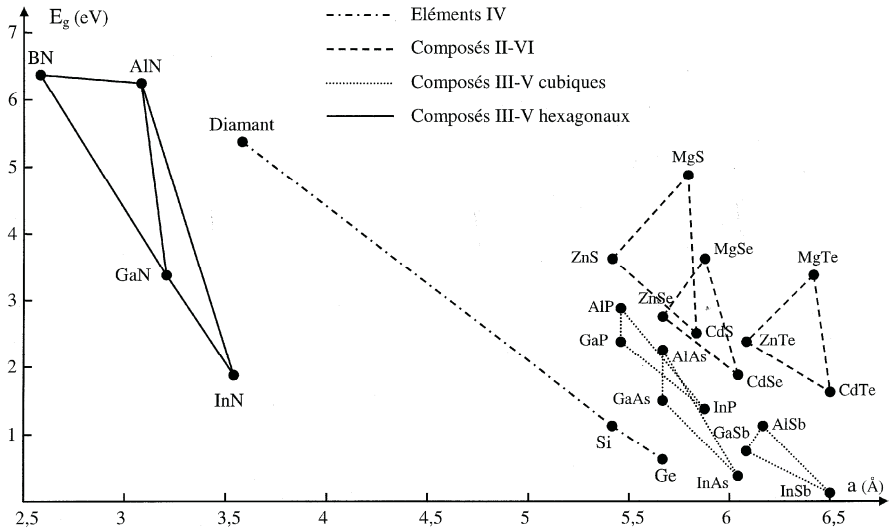


Figure 1.13' : Gaps et paramétres de maille

Nous verrons dans le paragraphe suivant comment représenter ces particules quasi-libres par des quasi-particules libres, en leur affectant une masse effective différente de la masse de l'électron libre. Mais revenons pour l'instant sur les électrons de la bande de valence. Dans la mesure où la bande de valence est incomplète, ces électrons peuvent aussi se déplacer et par suite transporter du courant électrique. Lorsqu'un électron d'une orbitale liante occupée passe sur une orbitale liante voisine, vide, la place vide se déplace dans l'autre sens, cette place vide est appelée un *trou*. Dans la mesure où la bande de valence est toujours quasi-pleine, l'étude du mouvement des électrons de cette bande est un problème à N corps. On ramène alors le problème à une particule en affectant une identité au trou et en étudiant le mouvement des trous qui, en raison de leur densité relativement faible, peuvent être traités comme des quasi-particules élémentaires indépendantes. Dans la mesure où le trou se déplace dans le sens opposé à l'électron, on lui affecte une charge positive égale à $+e$. D'autre part, le mouvement du trou résulte du mouvement de l'ensemble des électrons occupant les orbitales liantes du cristal, on lui affecte comme à l'électron une masse effective.

1.2.4. *Concept de masse effective*

a Masse effective des électrons

Un électron dans la bande de conduction est caractérisé par une fonction d'onde qui est une somme de Bloch d'orbitales antiliantes. En termes corpusculaires, c'est une particule dans un potentiel cristallin. On représente cette particule quasi-libre de charge *-e* et de masse m_o, par une quasi-particule libre de charge *-e* et de masse m_e qu'on appelle *masse effective* de l'électron. Cette représentation est loin d'être évidente et sa justification peut se trouver dans un certain nombre d'ouvrages théoriques (voir les ouvrages de Landau et Nozieres). Une description plus précise de la conduction conduit à représenter les états comme les états de pseudo particules (électrons et trous) au voisinage du niveau de Fermi.

Considérons un cristal soumis à une différence de potentiel. Un électron de conduction du cristal est soumis d'une part à une force interne $\mathbf{F}_i$ résultant du champ cristallin, et d'autre part à une force d'origine externe $\mathbf{F}_e$ résultant du champ électrique appliqué au cristal. L'équation de la dynamique s'écrit pour cet électron

$$m_o \frac{d\mathbf{v}}{dt} = \mathbf{F}_i + \mathbf{F}_e \tag{1-55}$$

On écrit que l'électron dans le cristal répond à la sollicitation de la force externe $\mathbf{F_e}$, comme une quasi-particule de masse m_e dans le vide

$$m_e \frac{d\mathbf{v}}{dt} = \mathbf{F}_e \tag{1-56}$$

La masse effective m_e contient en quelque sorte l'inertie additionnelle que donne à l'électron le potentiel cristallin, c'est-à-dire contient l'effet global du potentiel cristallin sur l'électron. Dans le réseau cristallin les fonctions propres électroniques sont des ondes de Bloch de la forme

$$\psi(\mathbf{r},t) = u(r) e^{i\mathbf{k}\cdot\mathbf{r}} e^{-i\omega_k t} \tag{1-57}$$

avec $E_k = \hbar\omega_k$

L'électron dans un état $\mathbf{k}$ est représenté par un paquet d'onde centré sur la pulsation ω_k. La vitesse de cet électron est égale à la vitesse de groupe du paquet d'ondes.

$$\mathbf{v} = \frac{d\omega}{d\mathbf{k}} = \frac{1}{\hbar}\frac{dE}{d\mathbf{k}} \tag{1-58}$$

Il est alors possible de montrer que le vecteur d'onde associé à l'électron obéit à l'équation suivante :

$$\hbar \frac{d\mathbf{k}}{dt} = -e\left[\mathbf{E} + \frac{1}{c}\mathbf{v} \times \mathbf{H}\right] \tag{1-59}$$

Cette équation est identique à l'équation classique d'un électron dans un champ électromagnétique. Le vecteur d'onde considéré est le vecteur moyen du paquet d'ondes. La justification rigoureuse de cette relation est délicate car le potentiel électrique créé par le réseau ne peut être considéré comme constant dans une dimension de l'ordre de la longueur d'onde. L'approximation classique doit être remplacée par une approximation semi-classique. En pratique, il faut admettre que le paquet d'ondes s'étale sur plusieurs mailles du réseau direct. L'accélération de cet électron est alors donnée par

$$\boldsymbol{\gamma} = \frac{d\mathbf{v}}{dt} = \frac{1}{\hbar}\frac{d}{dt}\frac{dE}{d\mathbf{k}} = \frac{1}{\hbar}\frac{d}{d\mathbf{k}}\frac{dE}{dt}$$

En mécanique classique, si une particule est soumise à une force $\mathbf{F}$ pendant un intervalle de temps dt, la variation de son énergie cinétique est donnée par

$$dE = \mathbf{F}.\mathbf{v}dt \quad ou \quad \frac{dE}{dt} = \mathbf{F}.\mathbf{v} \tag{1-60}$$

En portant l'expression (1-60) dans l'expression de l'accélération on obtient compte tenu de (1-58)

$$\boldsymbol{\gamma} = \frac{1}{\hbar}\frac{d}{d\mathbf{k}}(\mathbf{F}.\mathbf{v}) = \mathbf{F}\frac{1}{\hbar^2}\frac{d^2E}{d\mathbf{k}^2} \tag{1-61}$$

ce qui s'écrit $\mathbf{F} = m_e\boldsymbol{\gamma}$ avec

$$m_e = \frac{\hbar^2}{d^2E/d\mathbf{k}^2} \tag{1-62}$$

La masse effective des électrons apparaît donc comme inversement proportionnelle à la dérivée seconde de la courbe de dispersion $E(k)$, c'est-à-dire à la courbure des bandes d'énergie dans l'espace des $\mathbf{k}$. Au voisinage d'un minimum de la bande de conduction, c'est-à-dire dans la région du diagramme énergétique où sont localisés les électrons de conduction, on peut développer la fonction $E(\mathbf{k})$ en série de Taylor. Dans la mesure où il s'agit d'un minimum, la dérivée première est nulle, de sorte qu'en limitant le développement au deuxième ordre, l'énergie s'écrit

$$E(\mathbf{k})=E(\mathbf{k}_0)+\frac{1}{2}\frac{\partial^2 E}{\partial k_x^2}(k_x-k_{ox})^2+\frac{1}{2}\frac{\partial^2 E}{\partial k_y^2}(k_y-k_{oy})^2+\frac{1}{2}\frac{\partial^2 E}{\partial k_z^2}(k_z-k_{oz})^2 \qquad (1\text{-}63)$$

Les surfaces d'énergie constante dans l'espace des $\mathbf{k}$, sont définies par la condition $E(\mathbf{k})$=Cte, c'est-à-dire à partir de (1-63) par l'équation

$$\frac{(k_x-k_{ox})^2}{A_x}+\frac{(k_y-k_{oy})^2}{A_y}+\frac{(k_z-k_{oz})^2}{A_z}=1 \qquad (1\text{-}64)$$

Avec $$A_i=\left[E(\mathbf{k})-E(\mathbf{k}_0)\right]\Big/\left[\frac{1}{2}\partial^2 E/\partial k_i^2\right] \quad i=x,y,z$$

Ces surfaces sont par conséquent des quadriques, dont la forme dépend des signes et des valeurs relatives des coefficients A_x, A_y, A_z. Dans la mesure où il s'agit d'un minimum, l'énergie augmente dans chacune des directions de l'espace des $\mathbf{k}$, de sorte que tous les coefficients sont positifs. Il en résulte que les surfaces d'énergie constante sont des ellipsoïdes. Si deux coefficients sont égaux, les ellipsoïdes sont de révolution, si les trois coefficients sont égaux les surfaces sont des sphères.

Considérons tout d'abord le cas d'un semiconducteur à gap direct. Pour simplifier on se place dans un espace à une dimension. La bande de conduction est univallée, centrée en $k_0 = 0$ et isotrope au voisinage de k_0. Tous les coefficients A_i sont égaux, de sorte que si on appelle E_c l'énergie du minimum, l'expression (1-63) s'écrit

$$E(k)=E_c+\frac{1}{2}\frac{\partial^2 E}{\partial k^2}k^2 \qquad (1\text{-}65)$$

ou, compte tenu de la définition de la masse effective (Eq. 1-62)

$$E(k)=E_c+\frac{\hbar^2 k^2}{2m_e} \qquad (1\text{-}66)$$

Ainsi l'électron au voisinage du minimum de la bande de conduction se comporte comme un électron libre de masse m_e. Comme l'énergie est un minimum local par rapport au vecteur d'onde, la vitesse de groupe de la particule classique associée y est nulle et donc son énergie cinétique. Dans la mesure où la courbure de la bande de conduction varie peu au voisinage du minimum, la masse effective est constante et par suite l'énergie $E(k)$ varie quadratiquement avec le vecteur d'onde $\mathbf{k}$. Cette loi de variation constitue ce que l'on appelle *l'approximation des bandes paraboliques*. Lorsque l'énergie cinétique des électrons devient très importante l'électron s'éloigne de E_c dans l'espace des énergies, sa masse effective varie, et l'approximation parabolique n'est plus justifiée. On peut montrer que la

variation de courbure des bandes est d'autant plus importante que le gap du matériau est petit. Il en résulte que l'approximation des bandes paraboliques est d'autant plus justifiée que le gap du semiconducteur est grand.

Revenons maintenant au cas réel à trois dimensions. Si le semiconducteur est à gap indirect, la bande de conduction est multivallée et anisotrope, avec plusieurs minima équivalents situés en différents points de la zone de Brillouin. Considérons l'un quelconque de ces minima, centré en $k_0 \neq 0$. Les surfaces d'énergie constante au voisinage de k_0 sont des ellipsoïdes qui, pour des raisons de symétrie, sont de révolution autour de l'axe portant le point k_0. Dans la mesure où l'on se place dans le système d'axes principaux de l'ellipsoïde, l'axe z portant le point k_0, les coefficients A_i sont tels que $A_x = A_y \neq A_z$. L'expression (1-63) s'écrit alors

$$E(\mathbf{k}) = E_c + \frac{1}{2}\frac{\partial^2 E}{\partial k_{//}^2}(k_{//} - k_o)^2 + \frac{1}{2}\frac{\partial^2 E}{\partial k_{\perp}^2}k_{\perp}^2 \qquad (1\text{-}67)$$

où $k_{//} = k_z$ est la composante de $\mathbf{k}$ portée par l'axe de révolution de l'ellipsoïde, et $k_{\perp} = \sqrt{k_x^2 + k_y^2}$ est la composante de $\mathbf{k}$ dans le plan perpendiculaire. Compte tenu de la définition de la masse effective (Eq. 1-62), on pose

$$m_{\ell} = \frac{\hbar^2}{\partial^2 E / \partial k_{//}^2} \qquad (1\text{-}68\text{-a})$$

$$m_t = \frac{\hbar^2}{\partial^2 E / \partial k_{\perp}^2} \qquad (1\text{-}68\text{-b})$$

m_{ℓ} est la masse effective de l'électron dans son mouvement suivant l'axe de révolution de l'ellipsoïde, c'est-à-dire suivant l'axe portant le point k_0 où se situe le minimum considéré. m_t est la masse effective de l'électron dans son mouvement dans le plan perpendiculaire à cet axe. m_{ℓ} est appelée *masse effective longitudinale*, m_t est appelée *masse effective transverse*. L'expression (1-67) s'écrit alors

$$E(\mathbf{k}) = E_c + \frac{\hbar^2}{2m_{\ell}}(k_{//} - k_o)^2 + \frac{\hbar^2}{2m_t}k_{\perp}^2 \qquad (1\text{-}69)$$

Pour des raisons fondamentales liées aux couplages interatomiques, le gap du matériau et la courbure de la bande de conduction en centre de zone sont directement liés. Lorsque le gap augmente la courbure de la bande diminue et par suite la masse effective des électrons augmente. La figure (1-14) représente la tendance générale de cette variation pour les semiconducteurs à gap direct.

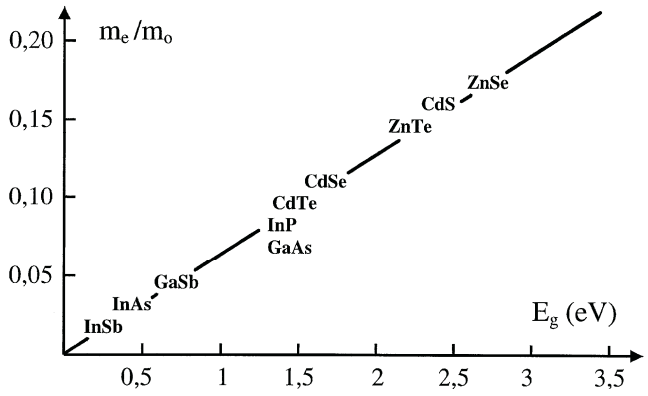


Figure 1.14 : Allure de la masse effective des électrons pour les semiconducteurs à gap direct

b. Masse effective des trous

Considérons les trous de la bande de valence, leur masse effective est définie comme pour les électrons, par l'inverse de la dérivée seconde de la courbe de dispersion. Mais la bande de valence des semiconducteurs cubiques est, comme le montre la figure (1-12), composée de deux branches dégénérées en k=0. Cette structure est dilatée sur la figure (1-15). La figure (1-15-a) représente les courbes de dispersion au voisinage de k = 0, les énergies des trous sont comptées positivement vers le bas. La bande de plus grande courbure, bande inférieure, correspond à des trous de masse effective inférieure. On appelle ces trous des trous légers lh (light holes), et la bande correspondante, *bande de trous légers*. La bande de plus faible courbure, bande supérieure, est appelée *bande trous lourds* hh (heavy holes).

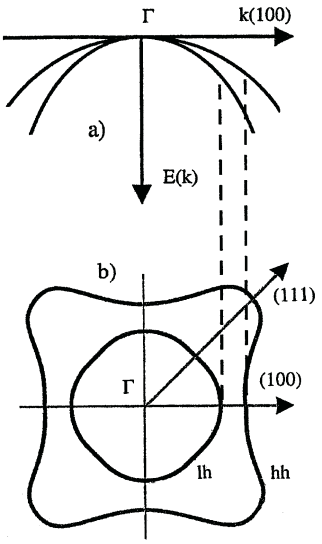


Figure 1-15 : Trous lourds et trous légers

En outre, en raison de la dégénérescence qui se produit en k = 0, ces bandes ne sont pas isotropes. Pour chacune d'elles, les surfaces d'énergie constante peuvent être représentées par des "sphères gauchies" centrées en k = 0. Ces surfaces sont schématisées à

deux dimensions par les courbes de la figure (1-15-b). Les surfaces d'énergie constante des trous légers présentent des protubérances dans les directions de haute symétrie [100] et équivalentes, et des creux dans les directions de haute symétrie [111] et équivalentes. La déformation des sphères représentant les énergies des trous lourds est symétrique de la précédente.
Il existe donc au voisinage du sommet de la bande de valence deux types de trous dont les énergies en fonction du vecteur d'onde sont données par des expressions plus compliquées que dans le cas des électrons. Ces expressions sont de la forme

$$E(k)=E_v+\frac{\hbar^2}{2m_o}\left[Ak^2\pm\sqrt{B^2k^4+C^2(k_x^2k_y{}^2+k_y^2k_z^2+k_z^2k_x^2)}\right] \tag{1-70}$$

où le signe + correspond aux trous légers et le signe - aux trous lourds.

On peut néanmoins, avec une approximation satisfaisante, remplacer les surfaces réelles par les sphères représentées en pointillés sur la figure (1-15-b). Dans ce cas on définit des masses effectives isotropes pour les trous lourds et les trous légers. En posant

$$\gamma_1=A \quad \gamma_2=B/2 \quad \gamma_3=\frac{1}{2\sqrt{3}}\sqrt{C^2+3B^2} \quad \bar{\gamma}=\sqrt{2\gamma_2^2+2\gamma_3^2} \tag{1-71}$$

les masses effectives des trous lourds et légers sont données par

$$m_{hh}=\frac{m_o}{\gamma_1-\bar{\gamma}} \tag{1-72-a}$$

$$m_{lh}=\frac{m_o}{\gamma_1+\bar{\gamma}} \tag{1-72-b}$$

Les paramètres $\gamma_1,\gamma_2,\gamma_3$ sont appelés *paramètres de bandes de valence* du semiconducteur. Leurs valeurs ont été calculées pour tous les semiconducteurs cubiques et mesurées pour la plupart d'entre eux.

On peut alors développer les bandes de valence dans l'approximation parabolique par des lois semblables à celles utilisées pour la bande de conduction. En comptant l'énergie positivement vers le bas, ces lois s'écrivent

$$E_{hh}(k)=E_v+\frac{\hbar^2k^2}{2m_{hh}} \tag{1-73-a}$$

$$E_{lh}(k)=E_v+\frac{\hbar^2k^2}{2m_{lh}} \tag{1-73-b}$$

Comme les électrons dans la bande de conduction, les trous dans les bandes de valence se comportent comme des particules libres de masses respectives m_{hh} et m_{lh}.

Les masses effectives des électrons, les paramètres de bandes de valence et les masses effectives des trous sont portés dans le tableau (1-2) pour la plupart des semiconducteurs IV-IV, III-V et II-V.

TABLEAU 1.2. : Masses effectives des électrons et des trous aux extrema des bandes. Paramètres caractéristiques de la bande de valence

Semi conducteur	Masses effectives des électrons m_e / m_o		Paramètres de la bande de valence			Masses effectives des trous m_h / m_o	
	m_l / m_o	m_t / m_o	γ_1	γ_2	γ_3	m_{hh} / m_o	m_{lh} / m_o
C (Diamant)	1,4	0,36	-	-	-	2,18	0,7
Si	0,916	0,191	4,22	0,39	1,44	0,53	0,16
Ge	1,588	0,082	13,35	4,25	5,69	0,35	0,043
SiC	0,3 //c	0,4 ⊥c	-	-	-	0,84	0,79
AlN	0,3	-	-	-	-	1	0,47
AlP	3	0,2	3,47	0,06	1,15	0,63	0,20
AlAs	1,1	0,19	4,04	0,78	1,57	0,76	0,15
AlSb	1,61	0,23	4,15	1,01	1,75	0,94	0,14
GaN	0,2	-	-	-	-	0,8	0,26
GaP	1,1	0,22	4,20	0,98	1,66	0,79	0,14
GaAs	0,067		7,65	2,41	3,28	0,62	0,074
GaSb	0,043		11,8	4,03	5,26	0,49	0,046
InN	0,1		-	-	-	0,5	-
InP	0,073		6,28	2,08	2,76	0,85	0,089
InAs	0,023		19,67	8,37	9,29	0,60	0,027
InSb	0,012		35,08	15,64	16,91	0,47	0,015
Zn0	0,24		-	-	-	2,27	0,79
ZnS	0,28		2,54	0,75	1,09	1,76	0,23
ZnSe	0,17		3,77	1,24	1,67	1,44	0,149
ZnTe	0,15		3,74	1,07	1,64	1,27	0,154
CdS	0,16		-	-	-	5,0	0,70
CdSe	0,11		-	-	-	2,5	0,45
CdTe	0,09		5,29	1,89	2,46	1,38	0,103
HgS	-		-41,28	-21,0	-20,73	2,78	-0,012
HgSe	0,019		-25,96	-13,69	-13,20	1,36	-0,019
HgTe	0,027		-18,68	-10,19	-9,56	1,12	-0,026

1.2.5. *Densité d'états dans les bandes permises*

Dans un cristal à trois dimensions, la densité d'états dans l'espace des **k** est donnée par l'expression (1-10). La densité d'états dans l'espace des énergies $g(E)$ est obtenue en explicitant k en fonction de l'énergie à partir des courbes de dispersion. Au voisinage de chacun des extrema, on peut utiliser l'approximation parabolique.

a. Bande de conduction

- *Semiconducteur à gap direct*

La bande de conduction présente un minimum unique, et isotrope dans un matériau cubique. Dans l'approximation parabolique, l'énergie des électrons est alors donnée par l'expression (1-66), de sorte que le vecteur d'onde k est donné par

$$k=\left(\frac{2m_e}{\hbar^2}(E-E_c)\right)^{1/2} \tag{1-74}$$

La densité d'états étant constante dans l'espace des k, le nombre d'états contenus dans la sphère de rayon k est donné par

$$N=g(k)V_k=\frac{2}{(2\pi)^3}\frac{4}{3}\pi k^3 \tag{1-75}$$

ou en explicitant le vecteur k en fonction de l'énergie

$$N=\frac{2}{(2\pi)^3}\frac{4}{3}\pi\left(\frac{2m_e}{\hbar^2}(E-E_c)\right)^{3/2} \tag{1-76}$$

La densité d'états par unité d'énergie est donnée par la dérivée dN/dE, soit

$$N_c(E)=\frac{1}{2\pi^2}\left(\frac{2m_e}{\hbar^2}\right)^{3/2}(E-E_c)^{1/2} \tag{1-77}$$

- *Semiconducteur à gap indirect*

La bande de conduction est multivallée, dans chacun des minima l'énergie des électrons est donnée par l'expression (1-69). Les surfaces d'énergie constante sont des ellipsoïdes de révolution, on ramène ces surfaces à des sphères en dilatant les axes perpendiculaires à l'axe de révolution de l'ellipsoïde. En d'autres termes, l'équation (1-69) peut s'écrire

$$E=E_c+\frac{\hbar^2}{2m_\ell}\left((k_{//}-k_o)^2+\frac{m_\ell}{m_t}k_\perp^2\right) \tag{1-78}$$

Posons $k_\perp^{'}=\left(\frac{m_\ell}{m_t}\right)^{1/2}k_\perp \quad k_{//}^{'}=k_{//} \quad k_o^{'}=k_o$

Dans la mesure où la seule composante non nulle du vecteur $\mathbf{k}_o$ est la composante k_0 suivant l'axe de l'ellipsoïde, l'expression (1-78) s'écrit

$$E=E_c+\frac{\hbar^2}{2m_\ell}(\mathbf{k}^{'}-\mathbf{k}_o^{'})^2 \tag{1-79}$$

Parallèlement, en raison de la dilatation des axes k_x et k_y perpendiculaires à l'axe k_z de l'ellipsoïde, la densité d'états dans l'espace des $k^{'}$ s'écrit

$$g(k')=g(k)\frac{k_x}{k_x^{'}}\frac{k_y}{k_y^{'}}\frac{k_z}{k_z^{'}}=g(k)\frac{m_t}{m_\ell}$$

$$g(k')=\frac{2}{(2\pi)^3}\frac{m_t}{m_\ell} \tag{1-80}$$

Le nombre d'états contenus dans la sphère centrée en $k_o^{'}$ et de rayon $k'-k_o^{'}$ est donnée par

$$N_1=\frac{2}{(2\pi)^3}\frac{m_t}{m_\ell}\frac{4}{3}\pi(k^{'}-k_o^{'})^3 \tag{1-81}$$

où en explicitant $(k^{'}-k_o^{'})$ à partir de l'expression (1-79)

$$N_1=\frac{2}{(2\pi)^3}\frac{m_t}{m_\ell}\frac{4}{3}\pi\left(\frac{2m_\ell}{\hbar^2}(E-E_c)\right)^{3/2}=\frac{1}{\pi^2}\frac{1}{3}\left(\frac{2(m_\ell^{1/2}m_t)^{2/3}}{\hbar^2}(E-E_c)\right)^{3/2}$$

En tenant compte du nombre *n* d'ellipsoïdes équivalents, le nombre total d'états d'énergie *E* est donné par $N=nN_1$, soit

$$N=\frac{1}{\pi^2}\frac{1}{3}\left(\frac{2(nm_\ell^{1/2}m_t)^{2/3}}{\hbar^2}(E-E_c)\right)^{3/2} \tag{1-82}$$

La densité d'états dans l'espace des énergies est donnée comme précédemment par dN/dE et s'écrit

$$N_c(E)=\frac{1}{2\pi^2}\left(\frac{2m_c}{\hbar^2}\right)^{3/2}(E-E_c)^{1/2} \tag{1-83}$$

$$m_c=(nm_\ell^{1/2}m_t)^{2/3} \tag{1-84}$$

Dans l'expression de la densité d'états, m_c joue dans le semiconducteur multivallée le même rôle que m_e dans le semiconducteur univallée, m_c est appelée *masse effective de densité d'états*.

	Si	**Ge**	**GaP**
n	6	4	3
m_c/m_0	1,06	0,55	0,78

b. Bande de valence

La densité d'états dans la bande de valence est la somme des densités d'états de trous lourds et de trous légers. Dans l'approximation isotrope et parabolique chacune d'elles se met sous une forme analogue à l'expression (1-77). La densité d'états globale est donnée par

$$N_v(E)=\frac{1}{2\pi^2}\left(\frac{2}{\hbar^2}\right)^{3/2}\left(m_{hh}^{3/2}+m_{lh}^{3/2}\right)(E_v-E)^{1/2} \tag{1-85}$$

en posant $m_v=(m_{hh}^{3/2}+m_{lh}^{3/2})^{2/3}$ la densité d'états s'écrit

$$N_v(E)=\frac{1}{2\pi^2}\left(\frac{2m_v}{\hbar^2}\right)^{3/2}(E_v-E)^{1/2} \tag{1-86}$$

m_v est la masse effective de densité d'états de la bande de valence.

	Si	**Ge**	**GaP**	**GaAs**	**InP**
m_v/m_0	0,59	0,36	0,83	0,64	0,87

En résumé, quel que soit le type de semiconducteur, on peut écrire la densité d'états au voisinage de l'extremum d'une bande sous la forme générale

$$N(E)=\frac{1}{2\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{3/2}(E-E_o)^{1/2} \tag{1-87}$$

L'énergie est comptée positivement vers le haut pour les électrons et vers le bas pour les trous. E_0 est l'énergie du bord de bande et m^* la masse effective de densité d'états correspondante.

1.3. Statistiques - Fonction de distribution des électrons

Le but de ce chapitre est de déterminer les états effectivement occupés. Le chapitre précédent a permis de quantifier les états possibles dans un cristal. Les électrons de valence étant en nombre fini, tous les états possibles ne seront pas occupés. Si le solide était à température nulle, ce sont les états correspondant aux plus basses énergies qui seraient occupés puisque le système cherche à se placer dans un état d'énergie minimum. Quand la température n'est pas nulle ce n'est plus le cas et la répartition des états occupés dépend de la température. Déterminer cette répartition est l'objet de ce chapitre.

Considérons un ensemble de particules identiques indépendantes, caractérisé par un nombre de particules déterminé n, possédant une énergie interne constante E, enfermées dans un volume constant V. L'état d'une particule i à un instant donné est défini par six paramètres, les trois coordonnées de position x_i, y_i, z_i, et les trois coordonnées de vitesse v_{xi}, v_{yi}, v_{zi} . Définissons un espace symbolique à 6 dimensions appelé *espace des phases*, dans lequel l'état instantané d'une particule est représenté par un point P de coordonnées $x_i, y_i, z_i, p_{xi}, p_{yi}, p_{zi}$ où $\mathbf{p}=m\mathbf{v}$ représente le vecteur quantité de mouvement ou moment de la particule.

L'état de l'ensemble est entièrement défini si on connaît dans l'espace des phases les points P_i représentatifs de toutes les particules. De toute évidence ceci n'est physiquement pas possible lorsque le nombre de particules est important. On définit alors l'état d'équilibre le plus probable de cet ensemble, c'est le but des statistiques.

Divisons l'espace des phases en cellules élémentaires telles que toutes les particules d'une même cellule aient les mêmes coordonnées $x_i, y_i, z_i, p_{xi}, p_{yi}, p_{zi}$ L'état de l'ensemble de particules est décrit par la connaissance des six coordonnées des différentes cellules et les nombres de particules par cellule. Dans un traitement quantique les différentes cellules sont tout simplement les différents états quantiques.

Avant d'étudier les différents types de statistique, on peut définir deux conditions auxquelles doit satisfaire toute répartition : le nombre total de particules et l'énergie interne du système sont constants

$$\sum_i n_i = n \qquad \sum_i n_i E_i = E \tag{1-88}$$

1.3.1. *Statistique de Boltzmann*

La statistique classique de Boltzmann est bâtie sur les hypothèses suivantes:

- Les particules sont discernables, c'est-à-dire en quelque sorte portent un numéro. On définit alors ce que l'on appelle une *complexion*, par la connaissance des coordonnées de chaque cellule, des nombres de particules par cellule et des identités des particules des différentes cellules.

- Toutes les complexions sont également probables.
- L'état d'équilibre statistique du système, c'est-à-dire son état le plus probable, est l'état qui est représenté par le plus grand nombre de complexions.
- Enfin, chaque cellule de l'espace des phases peut contenir un grand nombre de particules.

Considérons un état d'équilibre du système correspondant à n_1 particules dans la cellule n° 1, ..., n_i particules dans la cellule n° i... Le nombre de complexions qui permettent de réaliser cet état est le nombre de permutations de n particules entre les différentes cellules. Ce nombre est obtenu en divisant le nombre total de permutations des n particules, c'est-à-dire $n!$, par le nombre de permutations à l'intérieur d'une même cellule, soit $n_1!$ dans la cellule n° 1,..., $n_i!$ dans la cellule n° i... etc. Ainsi le nombre total de complexions différentes s'écrit

$$W = \frac{n!}{n_1!\,n_2!\ldots n_i!\ldots} = \frac{n!}{\prod_i n_i!} \tag{1-89}$$

L'état d'équilibre statistique du système est celui qui correspond au maximum de complexions, c'est-à-dire au maximum de W ou ce qui est équivalent, au maximum de $Ln\ W$.

$$Ln\,W = Ln\,n! - \sum_i Ln\,n_i! \tag{1-90}$$

Dans la mesure où chaque cellule peut contenir un grand nombre de particules, on peut approximer $Ln\ x!$ par la formule de Stirling $Ln\ x! = x\ Ln\ x - x$. Ainsi, compte tenu de la condition $\sum_i n_i = n$, $Ln\ W$ s'écrit

$$Ln\,W = n\,Ln\,n - \sum_i n_i\,Ln\,n_i \tag{1-91}$$

Le maximum de W correspond au zéro de sa différentielle. Compte tenu du double fait que tous les n_i sont variables et que n est constant, on obtient

$$d(Ln\,W) = -\sum_i Ln\,n_i\,dn_i \tag{1-92}$$

W est maximum pour $d(Ln\,W) = 0$, mais les n_i ne sont pas indépendants, ils doivent en outre satisfaire aux équations (1-88) qui donnent, en différentiant,

$$\sum_i dn_i = 0 \qquad \sum_i E_i\,dn_i = 0$$

On obtient donc la configuration la plus probable en écrivant que les trois conditions suivantes sont satisfaites

$$\sum_i Ln\, n_i\, dn_i = 0 \qquad \sum_i dn_i = 0 \qquad \sum_i E_i\, dn_i = 0 \tag{1-93}$$

En utilisant la méthode des multiplicateurs de Lagrange cette triple condition s'écrit

$$\sum_i (Ln\, n_i + \alpha + \beta E_i)\, dn_i = 0 \tag{1-94}$$

α et β sont les multiplicateurs de Lagrange dont il faudra déterminer les valeurs. Les variations dn_i étant indépendantes les unes des autres, pour que la somme sur i soit nulle il faut que chacun de ses termes soit nul, c'est-à-dire, $Ln\, n_i = -\alpha - \beta E_i$, ou

$$n_i = e^{-\alpha - \beta E_i} \tag{1-95}$$

n_i représente le nombre de particules contenues dans la cellule i caractérisée par l'énergie E_i. Si plusieurs cellules correspondent au même état d'énergie E_i, on dit que le niveau d'énergie E_i est dégénéré. Le niveau est g_i fois dégénéré si g_i cellules correspondent au même état d'énergie E_i. Le nombre de particules d'énergie E_i est alors donné par

$$n_i = g_i\, e^{-\alpha - \beta E_i}$$

La probabilité pour qu'une particule soit sur l'état d'énergie E_i est alors donnée par le rapport du nombre de particules au nombre de places, soit

$$\overline{n}_i = \frac{n_i}{g_i} = e^{-\alpha - \beta E_i} \tag{1-96}$$

$\overline{n}_i$ est la fonction de distribution de Boltzmann.

1.3.2. *Statistiques quantiques*

La notion principale qui distingue les statistiques quantiques de la statistique classique de Boltzmann est la notion d'indiscernabilité des particules identiques, en d'autres termes les particules identiques ne sont pas numérotables. En outre, le principe d'incertitude d'Heisenberg postule qu'il est impossible de connaître simultanément, avec une précision infinie, une coordonnée de position x et son moment conjugué p_x. Le produit des incertitudes est supérieur à $\hbar$, $\Delta x.\Delta p_x \geq \hbar$. Il en résulte que l'espace des phases est discontinu avec des cellules élémentaires de dimension $\hbar^3$.

Considérons un état macroscopique du système de n particules caractérisé par le fait que chaque état d'énergie E_i, g_i fois dégénéré, est occupé par n_i particules. On peut représenter chaque état d'énergie E_i par une couronne contenant g_i cases et n_i particules. Dans l'exemple de la figure (1-16) $g_i = 6, n_i = 10$. Il s'agit de trouver le nombre de façons différentes de répartir les n_i billes dans les g_i cases de chaque état d'énergie E_i. Ce nombre dépend évidemment du fait que l'on puisse ou non mettre plusieurs billes par case. Ceci dépend de la nature des particules étudiées et se traduit par l'existence de deux types de statistique quantique. Si les particules ont un spin nul ou entier on peut en mettre un nombre quelconque par état quantique, leur distribution obéit à la statistique de Bose-Einstein, ces particules sont des *bosons*, c'est le cas des photons et des phonons en particulier. Si par contre les particules ont un spin demi-entier, le principe d'exclusion de Pauli interdit d'en mettre plus d'une par état quantique. Ces particules obéissent à la statistique de Fermi-Dirac, on les appelle des *fermions*, c'est le cas des électrons en particulier.

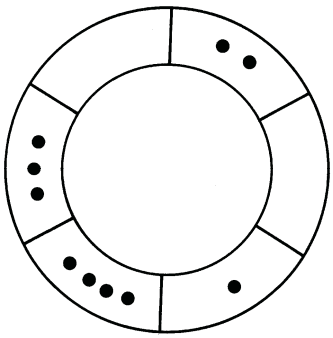


Figure 1-16

a. Statistique de Bose-Einstein

Il s'agit de répartir n_i billes dans g_i cases, avec un nombre quelconque de billes par case. Ceci revient dans la couronne de la figure (1-16), à permuter les n_i billes et les g_i barres de séparation des cases. Le nombre total de permutations entre ces $(n_i + g_i)$ objets est $(n_i + g_i)!$. Mais toute permutation entre deux billes ne donne pas une nouvelle distribution, car les billes sont indiscernables, il y en a $n_i!$. De même, toute permutation entre deux barres est inefficace, il y en a $g_i!$. Ainsi, le nombre de permutations qui modifient l'état microscopique de la répartition sur le niveau i est donné par

$$W_i = \frac{(n_i + g_i)!}{n_i!\, g_i!} \tag{1-97}$$

W_i représente le nombre de possibilités de peupler le niveau E_i avec n_i particules. Le nombre total de possibilités sur tous les états d'énergie E_i est donné par le produit des W_i

$$W = \prod_i W_i = \prod_i \frac{(n_i + g_i)!}{n_i!\, g_i!} \tag{1-98}$$

L'état le plus probable étant celui qui correspond au maximum de W, on obtient la probabilité d'un état macroscopique en écrivant que $d(Ln\,W)=0$ avec les deux conditions (1-88)

$$LnW=\sum_i \{Ln(n_i+g_i)!-Lnn_i!-Lng_i!\} \tag{1-99}$$

ou en utilisant la formule de Stirling

$$LnW=\sum_i \{n_i[Ln(n_i+g_i)-Lnn_i]+g_i[Ln(n_i+g_i)-Lng_i]\}$$

La différentielle de *Ln W* s'écrit

$$d(LnW)=\sum_i dn_i Ln\frac{n_i+g_i}{n_i} \tag{1-100}$$

La répartition la plus probable est donnée par les trois conditions

$$\sum_i dn_i\ Ln\frac{n_i+g_i}{n_i}=0 \qquad \sum_i dn_i=0 \qquad \sum_i E_i\,dn_i=0$$

ou en utilisant la méthode des multiplicateurs de Lagrange

$$\sum_i dn_i\,(Ln\frac{n_i+g_i}{n_i}-\alpha-\beta E_i)=0 \tag{1-101}$$

Les variations dn_i étant indépendantes les unes des autres, chaque terme de la somme sur i doit satisfaire à la condition, soit

$$Ln\frac{n_i+g_i}{n_i}-\alpha-\beta E_i=0$$

ou

$$n_i=\frac{g_i}{e^{\alpha+\beta E_i}-1} \tag{1-102}$$

g_i étant le degré de dégénérescence de l'état d'énergie E_i, la probabilité pour qu'une particule occupe l'état d'énergie E_i est donnée par le rapport n_i/g_i.

$$\overline{n_i}=\frac{n_i}{g_i}=\frac{1}{e^{\alpha+\beta E_i}-1} \tag{1-103}$$

$\overline{n_i}$ est la fonction de **distribution de Bose-Einstein**.

b. Statistique de Fermi-Dirac

Compte tenu du principe d'exclusion de Pauli, il s'agit ici de répartir les n_i billes de la figure (1-16) dans les g_i cases avec 0 ou 1 bille par case. Ainsi, sur les g_i cases de l'état d'énergie E_i, il y en a n_i pleines et $(g_i - n_i)$ vides. Le nombre de façons de peupler l'état d'énergie E_i avec n_i billes est donc tout simplement le nombre de permutations possibles entre n_i cases pleines et $(g_i - n_i)$ cases vides.

Le nombre de permutations entre g_i cases est $g_i!$, mais les permutations des cases pleines entre elles (il y en a $n_i!$) et des cases vides entre elles (il y en a $(g_i - n_i)!$) ne modifient pas l'état microscopique du peuplement du niveau d'énergie E_i. Ainsi le nombre de possibilités de ranger n_i billes dans g_i cases avec une seule bille par case est tout simplement

$$W_i = \frac{g_i!}{n_i!(g_i - n_i)!} \tag{1-104}$$

La probabilité d'un état macroscopique est alors donnée par

$$W = \prod_i W_i = \prod_i \frac{g_i!}{n_i!(g_i - n_i)!} \tag{1-105}$$

comme précédemment, on obtient l'état macroscopique le plus probable en écrivant que W est maximum compte tenu des conditions (1-88)

$$Ln\, W = \sum_i Ln\, g_i! - Ln\, n_i! - Ln(g_i - n_i)! \tag{1-106}$$

ou en utilisant la formule de Stirling

$$Ln\, W = \sum_i \left[g_i\, Ln\, g_i - n_i\, Ln\, n_i - (g_i - n_i)\, Ln(g_i - n_i) \right]$$

La différentielle de $Ln\, W$ s'écrit

$$d(Ln\, W) = \sum_i dn_i\, Ln \frac{g_i - n_i}{n_i} \tag{1-107}$$

Comme précédemment, en utilisant la méthode des multiplicateurs de Lagrange, la répartition la plus probable est donnée par

$$\sum_i dn_i \left(Ln \frac{g_i - n_i}{n_i} - \alpha - \beta E_i \right) = 0 \tag{1-108}$$

Ainsi le nombre de particules n_i occupant un niveau d'énergie E_i est donné par

$$Ln\frac{g_i - n_i}{n_i} - \alpha - \beta E_i = 0$$

ou

$$n_i = \frac{g_i}{1 + e^{\alpha + \beta E_i}} \tag{1-109}$$

La probabilité pour qu'une particule soit dans l'état d'énergie E_i est par conséquent donnée par le rapport n_i / g_i

$$\overline{n_i} = \frac{n_i}{g_i} = \frac{1}{1 + e^{\alpha + \beta E_i}} \tag{1-110}$$

$\overline{n_i}$ est la fonction de **distribution de Fermi-Dirac**.

c. Comparaison des différents types de statistique

Considérons un système dilué, c'est-à-dire un système dans lequel le nombre de places disponibles à chaque niveau d'énergie E_i est grand devant le nombre de particules susceptibles d'occuper ces places, $n_i << g_i$. La condition $\overline{n_i} << 1$ est alors vérifiée, ce qui correspond au fait que dans les expressions (1-103 et 110), le dénominateur est beaucoup plus grand que le numérateur, c'est-à-dire à la condition $e^{\alpha + \beta E_i} >> 1$. On peut alors négliger 1 devant l'exponentielle au dénominateur, de sorte que les fonctions de distribution de Bose-Einstein et de Fermi-Dirac deviennent équivalentes et se ramènent à la fonction de distribution de Boltzmann

$$\overline{n_i} = e^{-\alpha - \beta E_i} \tag{1-111}$$

Cette équivalence entre les deux types de statistique quantique traduit simplement le fait que lorsque la condition $n_i << g_i$ est réalisée la probabilité de trouver deux particules dans la même case devient négligeable et par suite le principe d'exclusion qui différencie les deux types de statistique est sans effet.

1.3.3. *Signification physique des multiplicateurs de Lagrange*

Les équations (1-101 et 108) montrent que le paramètre α traduit la conservation du nombre total de particules du système et le paramètre β la conservation de son énergie totale. Ces paramètres sont donc indépendants de la nature des particules.

a. Paramètre β

Considérons la population d'électrons libres dans un semiconducteur non dégénéré. Nous verrons que le nombre de ces électrons est faible devant le nombre de places disponibles de sorte qu'ils constituent un système dilué qui, par

conséquent, se comporte comme un gaz parfait. L'énergie moyenne par électron est donnée par

$$< E >= \frac{3}{2} kT \tag{1-112}$$

où k est la constante de Boltzmann et T la température absolue.

D'un autre côté cette énergie moyenne par électron est donnée par le rapport de l'énergie totale du système au nombre d'électrons

$$< E >= \frac{E}{n} = \frac{\sum_i n_i E_i}{\sum_i n_i} \tag{1-113}$$

Le nombre n_i d'électrons d'énergie E_i est simplement donné par $n_i = g_i \overline{n}_i$ où $\overline{n}_i$ est la probabilité d'occupation de l'état E_i et g_i la densité d'états d'énergie E_i. Dans un semiconducteur cette densité d'états est donnée par l'expression (1-87) que nous pouvons écrire sous une forme simplifiée $g_i = CE_i^{1/2}$ en prenant l'origine des énergies au bas de la bande de conduction. Le nombre d'électrons d'énergie E_i s'écrit donc $n_i = CE_i^{1/2}\overline{n}_i$

Dans un système dilué $\overline{n}_i$ est donné par l'expression (1-111) de sorte que n_i est donné par

$$n_i = C\, E_i^{1/2}\, e^{-\alpha-\beta E_i} \tag{1-114}$$

L'énergie moyenne par électron s'écrit alors

$$< E >= \frac{\sum_i C\, E_i^{3/2}\, e^{-\alpha-\beta E_i}}{\sum_i C\, E_i^{1/2}\, e^{-\alpha-\beta E_i}} = \frac{\sum_i E_i^{3/2}\, e^{-\beta E_i}}{\sum_i E_i^{1/2}\, e^{-\beta E_i}}$$

En remplaçant les sommes discrètes par des intégrales sur l'énergie on obtient la relation

$$E = \frac{\int_0^\infty E^{3/2} e^{-\beta E}\, dE}{\int_0^\infty E^{1/2} e^{-\beta E}\, dE} \tag{1-115}$$

On calcule ces intégrales en utilisant les propriétés de la fonction Γ définie par

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt$$

avec $\Gamma(1/2) = \sqrt{\pi}$ $\quad \Gamma(x+1) = x\Gamma(x)$ $\quad \Gamma(n+1) = n!$ pour n entier.

En posant $\beta E = t$, l'expression (1-115) s'écrit

$$<E>=\frac{1}{\beta}\frac{\int_0^\infty t^{3/2}e^{-t}dt}{\int_0^\infty t^{1/2}e^{-t}dt}=\frac{1}{\beta}\frac{\Gamma(5/2)}{\Gamma(3/2)}=\frac{3}{2\beta} \tag{1-116}$$

La comparaison des expressions (1-116 et 112) donne

$$\beta=\frac{1}{kT} \tag{1-117}$$

b. Paramètre α

Nous avons appelé W le nombre de complexions du système de particules, et calculé l'état le plus probable en écrivant que $Ln\,W$ était maximum. A partir de cette grandeur statistique on définit une grandeur thermodynamique S par la relation

$$S = k\,Ln\,W \tag{1-118}$$

La quantité S est appelée *entropie* du système, c'est une mesure logarithmique du nombre de complexions W. L'état d'équilibre statistique du système correspond au maximum de W c'est-à-dire au maximum de S. L'entropie mesure donc le degré de désordre du système. La définition (1-118) montre que S a les dimensions de la constante de Boltzmann, donc TS a les dimensions d'une énergie, c'est l'*énergie de désordre* du système.

La différence entre *l'énergie interne* E du système et son *énergie de désordre* TS est *l'énergie libre* F définie par

$$F = E - TS \tag{1-119}$$

En explicitant E et S cette expression s'écrit

$$F = \sum_i n_i E_i - kT \sum_i Ln\,W_i \tag{1-120}$$

Supposons maintenant que le nombre de particules du système varie de dn, le paramètre α varie de $d\alpha$ et l'énergie libre varie de dF, soit

$$\frac{dn}{d\alpha}=\sum_i \frac{dn_i}{d\alpha} \tag{1-121}$$

et

$$\frac{dF}{d\alpha} = \sum_i E_i \frac{dn_i}{d\alpha} - kT\sum_i \frac{d(LnW_i)}{d\alpha} = \sum_i \left(E_i - kT\frac{d(LnW_i)}{dn_i}\right)\frac{dn_i}{d\alpha} \tag{1-122}$$

L'expression (1-107) permet d'écrire

$$\frac{d(Ln\,W_i)}{dn_i} = Ln\frac{g_i - n_i}{n_i} \tag{1-123}$$

D'autre part l'expression de n_i (1-109) permet d'écrire $(g_i - n_i)/n_i = e^{\alpha+\beta E_i}$, soit

$$Ln\frac{g_i - n_i}{n_i} = \alpha + \beta E_i \tag{1-124}$$

L'expression (1-122) s'écrit alors

$$\frac{dF}{d\alpha} = \sum_i \left(E_i - kT(\alpha + \beta E_i)\right)\frac{dn_i}{d\alpha}$$

En explicitant β donné par l'expression (1-117), cette expression devient

$$\frac{dF}{d\alpha} = \sum_i -\alpha kT\frac{dn_i}{d\alpha} = -\alpha kT\frac{dn}{d\alpha}$$

d'où l'expression du paramètre α

$$\alpha = -\frac{dF/dn}{kT} \tag{1-125}$$

La quantité dF/dn représente la variation de l'énergie libre du système en fonction du nombre de particules. On appelle cette quantité *potentiel chimique* du système. Dans un système de fermions on a coutume d'appeler ce potentiel chimique, *énergie de Fermi* ou *niveau de Fermi* E_F du système. Le paramètre α s'écrit alors

$$\alpha = -E_F/kT \tag{1-126}$$

Remarquons que nous avons développé le calcul des paramètres α et β pour un système de fermions, mais ces paramètres sont indépendants de la nature des particules. On obtiendrait les mêmes expressions pour un système de bosons. Il suffit de reprendre le même calcul avec les expressions (1-97 et 102) pour W_i et n_i respectivement.

Le potentiel chimique appelé parfois abusivement énergie de Fermi ne doit pas être confondu avec l'énergie de Fermi définie à température nulle comme l'énergie la plus élevée des électrons dans un solide. A température non nulle et pour les métaux les deux valeurs sont légèrement différentes. Dans le cas des semiconducteurs les valeurs sont très différentes.

Dans l'analyse précédente, le niveau de Fermi est commun aux électrons et aux trous. Cela est vrai pour des semiconducteurs en équilibre thermodynamique. Quand on applique une tension extérieure cette hypothèse doit être abandonnée et on considère généralement que les populations d'électrons et de trous s'équilibrent séparément avec deux niveaux de Fermi distincts pour chaque population. Ils sont alors appelés ***pseudo-niveaux de Fermi***.

Une propriété importante du niveau de Fermi ou du pseudo-niveau de Fermi sera largement utilisée dans la suite de cet ouvrage. Elle relie le niveau de Fermi au potentiel appliqué à un solide. Considérons deux morceaux de solide en contact à même température et reliés à un générateur de tension extérieur. Ils échangent des électrons. Supposons que chaque morceau est en équilibre thermodynamique. On peut alors définir deux niveaux de Fermi, E_{F1} et E_{F2}. Faisons ensuite passer adiabatiquement un électron du morceau 1 au morceau 2.
La variation d'énergie libre du système total, puisque le nombre d'électrons de chaque sous-système a varié de un (on applique la relation 1-125), est donc

$$\Delta F = E_{F2} - E_{F1}$$

D'autre part, en appliquant cette fois la relation 1-119

$$\Delta F = -e\left(V_2 - V_1\right)$$

En effet, la variation d'entropie étant nulle, la variation d'énergie libre du système des deux solides est égale à la variation d'énergie mais est aussi égale à l'énergie électrique apportée par le générateur extérieur. On en déduit donc

$$e\left(V_1 - V_2\right) = E_{F2} - E_{F1} \qquad (1\text{-}127)$$

La même relation peut s'écrire pour les pseudo-niveaux de Fermi des électrons et des trous.

1.3.4. *Fonction de Fermi*

En explicitant α et β dans l'expression (1-110), on obtient la probabilité d'occupation d'un niveau d'énergie E par un électron, à la température T.

$$f(E) = \frac{1}{1 + e^{(E - E_F)/kT}} \qquad (128)$$

$f(E)$ est appelée *fonction de Fermi*. Elle est représentée sur la figure (1-17). On vérifie immédiatement que $f(E_F) = 1/2$ quel que soit T.

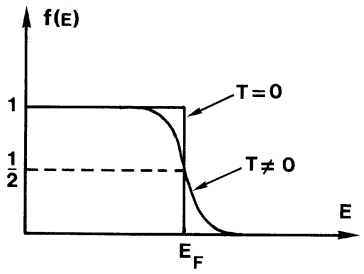


Figure 1-17 : La fonction de Fermi

D'autre part si $T = 0$, $f(E) = 1$ pour $E < E_F$ et $f(E) = 0$ pour $E > E_F$.
Ainsi au zéro absolu tous les états d'énergie situés au-dessous du niveau de Fermi sont occupés et tous les états d'énergie situés au-dessus sont vides. Le niveau de Fermi délimite donc, dans le système, les états occupés et les états vides.

Enfin, dès que $E - E_F$ devient de l'ordre de quelques kT, on peut négliger 1 devant l'exponentielle au dénominateur et la fonction de Fermi se ramène à une distribution de Bolztmann.

$$f(E) \approx e^{-(E-E_F)/kT} \tag{1-129}$$

La fonction de distribution $f(E)$ représente la probabilité pour qu'un électron occupe un état d'énergie E. Il en résulte que la probabilité pour que cet état ne soit pas occupé par un électron est $1 - f(E)$. Cette fonction représente par conséquent la fonction de distribution des trous dans le semiconducteur.
On écrit

$$f_p(E) = 1 - f(E) = \frac{1}{1 + e^{-(E-E_F)/kT}} \tag{1-130}$$

1.4. Le semiconducteur à l'équilibre thermodynamique

Le nombre d'électrons ou de trous d'énergie E dans le semiconducteur est donné par le produit de la densité d'états par la fonction de distribution. Ainsi les densités totales d'électrons dans la bande de conduction et de trous dans la bande de valence sont données par

$$n = \int_{E_c}^{E_{c\max}} N_c(E) f_n(E) dE$$

$$p = \int_{E_{v\min}}^{E_v} N_v(E) f_p(E) dE$$

Les densités d'états $N_c(E)$ et $N_v(E)$ et les fonctions de distribution $f_n(E)$ et $f_p(E)$ sont données respectivement par les expressions (1-77 ou 83 et 86) d'une part et (1-128 et 130) d'autre part. En outre étant donné les largeurs des bandes de conduction et de valence et la variation rapide de la fonction de Fermi, on peut

intégrer les expressions (1-130) de E_c à l'infini dans la bande de conduction et de moins l'infini à E_v dans la bande de valence. Les densités de porteurs libres sont représentées sur la figure (1-18).

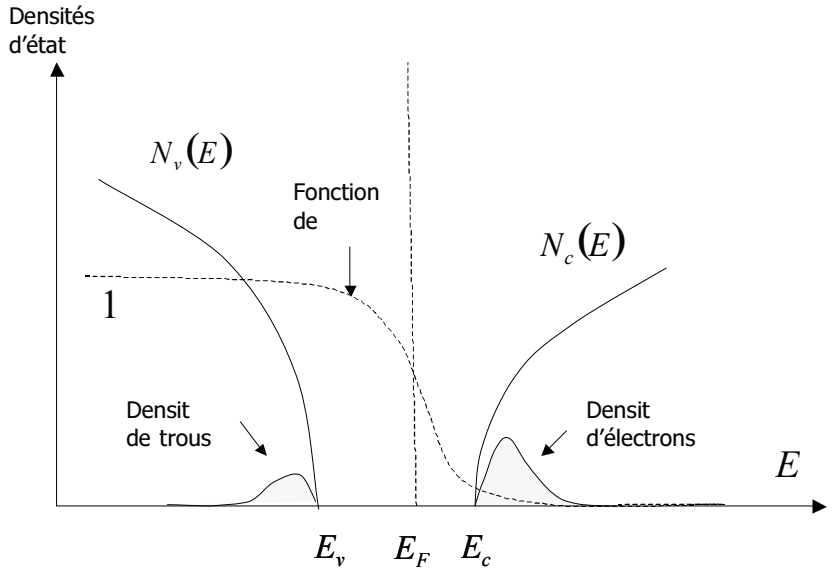


Figure 1-18 : Densités de porteurs dans les bandes permises. Les surfaces grisées représentent les intégrales

1.4.1. *Semiconducteur non dégénéré*

On dit qu'un semiconducteur est dégénéré lorsque le niveau de Fermi est situé dans une bande permise, bande de valence ou bande de conduction. Dans le cas contraire, qui correspond au fonctionnement de la majorité des composants, le niveau de Fermi est situé dans la bande interdite. S'il est alors distant des extrêma des bandes permises de plusieurs kT, typiquement si $E_c - E_F > 2kT$ et $E_F - E_v > 2kT$, les fonctions de distribution des électrons et des trous se ramènent à des distributions de Boltzmann correspondant à un système dilué

$$f_n(E) \approx e^{-(E-E_F)/kT} \qquad f_p(E) \approx e^{(E-E_F)/kT}$$

Les expressions (1-130) s'écrivent alors

$$n = \int_{E_c}^{\infty} N_c(E) e^{-(E-E_F)/kT} dE \qquad \text{(1-131-a)}$$

$$p = \int_{-\infty}^{E_v} N_v(E) e^{(E-E_F)/kT} dE \qquad \text{(1-131-b)}$$

On modifie la forme de ces intégrales en faisant apparaître les énergies du minimum E_c de la bande de conduction et du maximum E_v de la bande de valence

$$n=\int_{E_c}^{\infty} N_c(E)e^{-(E-E_c)/kT}\,dE.e^{-(E_c-E_F)/kT}$$
$$p=\int_{-\infty}^{E_v} N_v(E)e^{(E-E_v)/kT}\,dE.e^{(E_v-E_F)/kT}$$

Ce qui permet d'écrire les densités de porteurs libres sous la forme

$$n=N_c e^{-(E_c-E_F)/kT} \tag{1-132-a}$$

$$p=N_v e^{(E_v-E_F)/kT} \tag{1-132-b}$$

avec

$$N_c=\int_{E_c}^{\infty} N_c(E)e^{-(E-E_c)/kT}\,dE \tag{1-133-a}$$
$$N_v=\int_{-\infty}^{E_v} N_v(E)e^{(E-E_v)/kT}\,dE \tag{1-133-b}$$

Les quantités N_c et N_v sont caractéristiques des densités d'états dans les bandes de conduction et de valence respectivement. N_c par exemple, représente une somme pondérée de tous les états de la bande de conduction. Le coefficient de pondération est le facteur de Boltzmann qui traduit le fait que plus les états sont hauts dans la bande de conduction plus la probabilité pour qu'un électron occupe ces états est faible. N_c représente en quelque sorte le nombre d'états utiles, à la température T, dans la bande de conduction. Il en est de même pour N_v dans la bande de valence. N_c et N_v sont appelés *densités équivalentes d'états* dans les bandes de conduction et de valence. On obtient leurs expressions en explicitant $N_c(E)$ et $N_v(E)$ dans les expressions (1-33) à partir des expressions (1-83 et 86). Calculons N_c par exemple

$$N_c=\frac{1}{2\pi^2}\left(\frac{2m_c}{\hbar^2}\right)^{3/2}\int_{E_c}^{\infty}(E-E_c)^{1/2}e^{-(E-E_c)/kT}\,dE \tag{1-134}$$

En posant $x=(E-E_c)/kT$ l'expression s'écrit

$$N_c=\frac{1}{2\pi^2}\left(\frac{2m_c kT}{\hbar^2}\right)^{3/2}\int_0^{\infty}x^{1/2}e^{-x}\,dx \tag{1-135}$$

L'intégrale sur *x* est égale à $\sqrt{\pi}/2$, de sorte que N_c, et par suite N_v, s'écrivent

$$N_c=2\left(\frac{2\pi m_c kT}{h^2}\right)^{3/2} \quad (1\text{-}136\text{-a})$$

$$N_v=2\left(\frac{2\pi m_v kT}{h^2}\right)^{3/2} \quad (1\text{-}136\text{-b})$$

En explicitant les constantes, ces expressions se mettent sous une forme simple

$$N_c=2{,}5.10^{19}\left(m_c/m_0\right)^{3/2}\left(T/300\right)^{3/2}cm^{-3} \quad (1\text{-}137\text{-a})$$

$$N_v=2{,}5.10^{19}\left(m_v/m_0\right)^{3/2}\left(T/300\right)^{3/2}cm^{-3} \quad (1\text{-}137\text{-b})$$

Si le semiconducteur est univallée m_c est simplement la masse effective m_e des électrons. Considérons par exemple les cas de Si, Ge, GaP, GaAs et InP à la température ambiante.

	m_C/m_O	m_V/m_O	N_C (10^{19} cm^{-3})	N_V (10^{19} cm^{-3})
Si	1,06	0,59	2,7	1,1
Ge	0,55	0,36	1	0,5
GaP	0,78	0,83	1,7	1,9
GaAs	0,067	0,64	0,04	1,3
InP	0,073	0,87	0,05	2

On constate que les densités équivalentes d'états sont de l'ordre de 10^{19} cm^{-3} dans la bande de valence pour tous les semiconducteurs. En ce qui concerne la bande de conduction, on retrouve une différence importante entre les semiconducteurs à gap indirect, tels que Si, Ge et GaP, et les semiconducteurs à gap direct, tels que GaAs et InP. Dans les premiers la densité équivalente d'états de la bande de conduction est de l'ordre de 10^{19} cm^{-3}, dans les seconds cette densité est de l'ordre de 10^{18} cm^{-3}.

Les expressions (1-132 et 136) montrent que les densités de porteurs libres varient doublement avec la température, d'une part à travers les densités équivalentes d'états et d'autre part explicitement de manière exponentielle. En fait N_c et N_v varient comme $T^{3/2}$ de sorte que les variations des densités de porteurs en fonction de la température sont essentiellement exponentielles.

Les expressions (1-132) permettent de montrer que le produit np est constant à une température donnée

$$np=N_c N_v e^{-E_g/kT} \quad (1\text{-}138)$$

1.4.2. *Semiconducteur dégénéré*

Le semiconducteur est dégénéré lorsque le niveau de Fermi est situé dans une bande permise. Les densités de porteurs correspondant à l'apparition du régime de dégénérescence sont obtenues simplement à partir des expressions (1-132). Si on

fait $E_F = E_c$ dans l'expression (1-132-a) par exemple, on obtient $n = N_c$. De même la dégénérescence de la bande de valence se produit lorsque $p = N_v$. En fait, ces grandeurs ne sont qu'approximatives dans la mesure où en régime de dégénérescence les expressions (1-132), qui ont été établies dans l'approximation de Boltzmann, ne sont plus valables. Ces expressions permettent néanmoins de connaître, à une température donnée (N_c et N_v sont fonction de la température), l'ordre de grandeur de la densité de porteurs libres correspondant au régime de dégénérescence.

En régime dégénéré on ne peut donc plus approximer la distribution de Fermi par une distribution de Boltzmann et les densités de porteurs libres sont données par les expressions (1-130). En explicitant les densités d'états et les fonctions de distribution la densité, d'électrons par exemple, s'écrit

$$n=\frac{1}{2\pi^2}\left(\frac{2m_e}{\hbar^2}\right)^{3/2}\int_{E_c}^{\infty}\frac{(E-E_c)^{1/2}}{1+e^{(E-E_F)/kT}}dE \tag{1-139}$$

En faisant apparaître la densité équivalente d'états déjà définie par l'expression (1-136-a), la densité d'électrons s'écrit

$$n=N_c F_{1/2}(\eta) \tag{1-140}$$

avec

$$F_{1/2}(\eta)=\frac{2}{\sqrt{\pi}}\int_0^{\infty}\frac{\varepsilon^{1/2}}{1+e^{\varepsilon-\eta}}d\varepsilon \tag{1-141}$$

où $\eta=(E_F-E_c)/kT$ et $\varepsilon=(E-E_c)/kT$

De la même manière, en posant $\varepsilon_g = E_g / kT$, la densité de trous est donnée par

$$p=N_v F_{1/2}(-\eta-\varepsilon_g) \tag{1-142}$$

Les fonctions intégrales $F_{1/2}(x)$ ne sont pas d'un emploi très facile et il est parfois utile, dans l'étude de certains composants, d'étendre aux semiconducteurs dégénérés la nature analytique des expressions des densités de porteurs utilisées pour les semiconducteurs non dégénérés (Approximation de Boltzmann).

Si le semiconducteur n'est que faiblement dégénéré, de type *n* par exemple, on peut prolonger l'approximation analytique de Boltzmann (Eq. 1-132) moyennant un facteur correctif. En effet pour $x<2$ on peut montrer mathématiquement que l'intégrale $F_{1/2}(x)$ se ramène à l'expression

$$F_{1/2}(x) \approx e^x \Big/ \left(1+e^x/4\right)$$

Ainsi, pour $E_F < E_c+2kT$ la densité d'électrons peut s'écrire sous la forme analytique

$$n \approx N_c e^{-(E_c-E_F)/kT} \frac{4}{4+e^{-(E_c-E_F)/kT}} \tag{1-143}$$

Lorsque x devient inférieur à -2, c'est-à-dire $E_F < E_c-2kT$, l'exponentielle devient négligeable devant 4 au dénominateur de l'expression (1-143) de sorte que l'on retrouve l'approximation de Boltzmann, correspondant au semiconducteur non dégénéré.

Si par contre le semiconducteur est fortement dégénéré, $x>5$, la fonction $F_{1/2}(x)$ se ramène à

$$F_{1/2}(x) \approx \frac{4}{3\sqrt{\pi}} x^{3/2}$$

Ainsi pour $E_F > E_c+5kT$, la densité d'électrons s'écrit

$$n \approx N_c \frac{4}{3\sqrt{\pi}} \left(\frac{E_F - E_c}{kT} \right)^{3/2} \tag{1-144}$$

Il faut noter que N_c varie comme $(kT)^{3/2}$, de sorte que la densité d'électrons devient indépendante de la température. Le semiconducteur présente alors un comportement métallique. Si on explicite N_c (Eq. 1-136) dans l'expression (1-144), la densité d'électrons s'écrit, comme dans un métal

$$n \approx \frac{8\pi}{3h^3} (2m_c)^{3/2} (E_F - E_c)^{3/2}$$

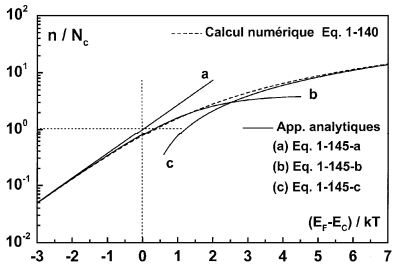


Figure 1.19 : Semiconducteur dégénéré

Ainsi en résumé, avec $\eta = (E_F - E_c)/kT$ la densité d'électrons s'écrit

$\eta < -2$ $\quad n \approx N_c e^{\eta}$ (1-145-a)

$\eta < 2$ $\quad n \approx N_c e^{\eta} \left(1/\left(1+e^{\eta}/4\right)\right)$ (1-145-b)

$\eta > 5$ $\quad n \approx \left(4/3\sqrt{\pi}\right) N_c \eta^{3/2}$ (1-145-c)

Ces différentes approximations sont comparées au résultat du calcul numérique sur la figure (1-19) La courbe en trait discontinu représente la valeur exacte issue du calcul numérique (Eq. 1-140). Les courbes en traits pleins représentent les approximations analytiques pour des semiconducteurs présentant divers taux de dégénérescence (Eq. 1-145).

1.4.3. *Semiconducteur intrinsèque*

Un semiconducteur intrinsèque est un semiconducteur dépourvu de toute impureté susceptible de modifier la densité de porteurs. Les électrons de la bande de conduction ne peuvent résulter que de l'excitation thermique d'électrons liés de la bande de valence. Il en résulte que les électrons et les trous existent nécessairement par paires et $n = p = n_i$. n_i est appelé *densité de porteurs intrinsèques,* c'est une caractéristique du semiconducteur à une température donnée.

a. Densité de porteurs

Un semiconducteur intrinsèque n'est jamais dégénéré de sorte que le produit $np = n_i^2$ est donné par l'expression (1-138). Il en résulte que la densité de porteurs intrinsèques s'écrit

$$n_i = \left(N_c N_v\right)^{1/2} e^{-E_g/2kT} \tag{1-146}$$

Cette densité de porteurs intrinsèques est une fonction exponentielle du gap du matériau et de la température. A la température ambiante on obtient

	Si	**Ge**	**GaAs**	**InP**
n_i **(cm^{-3})**	10^{10}	2.10^{13}	3.10^{6}	3.10^{7}

b. Niveau de Fermi

On obtient la position du niveau de Fermi en écrivant que $n = p$, soit

$$N_c\, e^{-(E_c - E_{Fi})/kT} = N_v\, e^{(E_v - E_{Fi})/kT}$$

Soit

$$E_{Fi} = \frac{E_c + E_v}{2} + \frac{1}{2} kT\, Ln \frac{N_v}{N_c} \tag{1-147}$$

En explicitant le rapport N_v / N_c en fonction des masses effectives, on obtient

$$E_{Fi} = \frac{E_c + E_v}{2} + \frac{3}{4} kT\, Ln \frac{m_h}{m_e} \tag{1-148}$$

Le rapport des masses effectives de densité d'états est de l'ordre de 1 dans les semiconducteurs à gap indirect et de l'ordre de 10 dans les semiconducteurs à gap direct. ***Il en résulte que le niveau de Fermi d'un semiconducteur intrinsèque est toujours très voisin du milieu du gap à la température ambiante***,

$$E_{Fi} \approx \left(E_c + E_v\right)/2 \tag{1-149}$$

1.4.4. *Semiconducteur extrinsèque à la température ambiante*

a. Donneurs et accepteurs

On modifie considérablement les propriétés électriques d'un semiconducteur en le dopant de manière contrôlée avec des atomes spécifiques.

Considérons par exemple un semiconducteur de la colonne IV comme le silicium, dans lequel on introduit un atome de la colonne V tel que l'arsenic. Dans le réseau cristallin, l'atome d'arsenic remplace un atome de silicium et établit des liaisons avec ses quatre voisins. L'atome d'arsenic est alors environné de 9 électrons dont 8 saturent les orbitales liantes du cristal. Le 9ème électron occupe alors une orbitale beaucoup plus délocalisée dans le champ de l'ion positif As^+. L'énergie de liaison $E_D=E_c-E_d$ de cet électron est de l'ordre de quelques meV dans les semiconducteurs à gap direct et quelques dizaines de meV dans les semiconducteurs à gap indirect. A la température ambiante cet électron est donc libéré dans le réseau par agitation thermique et occupe un état de la bande de conduction. Parallèlement l'arsenic a une charge positive excédentaire et devient un ion As^+. On dit que l'arsenic dans le silicium est un *donneur* en raison du fait qu'il donne un électron de conduction au réseau.

Remplaçons un atome de silicium par un atome de la colonne III, tel que le gallium. Dans la coordination tétraédrique, il apparaît maintenant un déficit d'électron. L'électron manquant est alors facilement remplacé par un électron provenant d'une liaison voisine. Le passage d'un électron depuis une liaison intrinsèque du cristal vers l'atome de gallium entraîne d'une part la création d'un trou dans la bande de valence et d'autre part l'apparition d'une charge négative excédentaire au voisinage de l'atome de gallium. L'atome de gallium, qui est appelé *accepteur*, devient alors un ion négatif Ga^-. Par analogie avec l'électron sur le donneur, on peut considérer que le trou qui apparaît dans la bande de valence était piégé sur l'accepteur. L'énergie qu'il faut pour transférer un électron de la bande de valence sur l'accepteur est alors appelée énergie de liaison $E_A=E_a-E_v$ du trou sur l'accepteur. Cette énergie de liaison est de l'ordre de quelques dizaines de meV.

b. Densité de porteurs et niveau de Fermi

Considérons un semiconducteur contenant une densité N_d de donneurs et une densité N_a d'accepteurs. Soit N_d^+ le nombre de donneurs ionisés, c'est-à-dire ayant libéré leur électron supplémentaire. Soit N_a^- le nombre d'accepteurs ionisés, c'est-à-dire ayant accepté un électron supplémentaire. Le matériau étant neutre, l'ensemble des charges positives est égal à l'ensemble des charges négatives. L'équation de neutralité électrique du matériau s'écrit

$$n + N_a^- = p + N_d^+ \tag{1-150}$$

où n et p représentent les densités de porteurs libres et N_a^- et N_d^+ les densités d'accepteurs et de donneurs ionisés.

Etudions tout d'abord le système à la température ambiante. En raison du fait que l'énergie thermique kT est du même ordre de grandeur que les énergies de liaison de l'électron sur le donneur et du trou sur l'accepteur, tous les donneurs et accepteurs sont ionisés. L'équation de neutralité électrique se réduit donc à

$$n + N_a = p + N_d \tag{1-151}$$

où N_a et N_d sont les densités d'accepteurs et de donneurs.

Notons que la condition $N_a = N_d = 0$ n'est jamais réalisée en raison du fait que lors de la cristallogénèse, le semiconducteur est toujours contaminé par des impuretés. En d'autres termes le semiconducteur intrinsèque n'existe pas, il existe en fait trois types de semiconducteur.

- *Semiconducteur compensé - Semi-Isolant*

Il existe toujours dans un semiconducteur des impuretés incontrôlées que l'on appelle *impuretés résiduelles*. Pour compenser ces impuretés, on dope le semiconducteur de manière à se rapprocher le plus possible de la condition $N_a = N_d$. Supposons que la condition soit réalisée, l'équation (1-151) s'écrit alors $n = p$ et par conséquent $n = p = n_i$. la densité de porteurs libres dans le semiconducteur est la même que s'il était intrinsèque. La densité intrinsèque n_i qui est fonction du gap du semiconducteur, est généralement très faible à la température ambiante. Le semiconducteur est alors, en ce qui concerne la conductivité, plus proche de l'isolant que du conducteur, on l'appelle alors *semi-isolant*. Le niveau de Fermi, dont la position est réglée par la population de porteurs libres, est donc situé, comme pour un semiconducteur intrinsèque, au voisinage du milieu du gap.

- *Semiconducteur de type n*

Considérons un semiconducteur dopé avec une densité de donneurs supérieure à la densité d'accepteurs, $N_d > N_a$. Il en résulte, d'après l'équation (1-151), $n > p$, on dit alors que le semiconducteur est de *type n*. Les électrons sont appelés *porteurs majoritaires*, les trous, *porteurs minoritaires*.

L'équation (1-151) associée à la relation $np = n_i^2$ permet d'écrire l'équation

$$n^2 - (N_d - N_a)n - n_i^2 = 0$$

La racine positive de cette équation donne le nombre d'électrons, la valeur absolue de sa racine négative donne le nombre de trous.

$$n = \frac{1}{2}\left[N_d - N_a + \left((N_d - N_a)^2 + 4n_i^2\right)^{1/2}\right]$$
$$p = -\frac{1}{2}\left[N_d - N_a - \left((N_d - N_a)^2 + 4n_i^2\right)^{1/2}\right]$$

Dans la pratique N_d, N_a et $N_d - N_a$ sont toujours très supérieurs à n_i, de sorte que les densités d'électrons et de trous s'écrivent respectivement

$$n \approx N_d - N_a \qquad \text{(1-152-a)}$$

$$p \approx n_i^2 / (N_d - N_a) \qquad \text{(1-152-b)}$$

- *Semiconducteur de type p*

Si le dopage est tel que $N_a > N_d$, les porteurs majoritaires sont les trous, le semiconducteur est de type p. Les densités de porteurs sont données par

$$p \approx N_a - N_d \qquad \text{(1-153-a)}$$

$$n \approx n_i^2 / (N_a - N_d) \qquad \text{(1-153-b)}$$

- *Niveau de Fermi*

Dans la mesure où le semiconducteur n'est pas dégénéré, les densités de porteurs sont données par les expressions (1-132). Ainsi dans un semiconducteur de type n, où $n = N_d - N_a$ et dans un semiconducteur de type p où $p = N_a - N_d$ les expressions (1-132) s'écrivent respectivement

$$N_d - N_a = N_c e^{-(E_c - E_{Fn})/kT} \qquad \text{(1-154-a)}$$
$$N_a - N_d = N_v e^{(E_v - E_{Fp})/kT} \qquad \text{(1-154-b)}$$

D'où les expressions du niveau de Fermi dans chaque type de semiconducteur

$$E_{Fn} = E_c - kTLn\left(N_c / (N_d - N_a)\right) \qquad \text{(1-155-a)}$$
$$E_{Fp} = E_v + kTLn\left(N_v / (N_a - N_d)\right) \qquad \text{(1-155-b)}$$

Dans chaque type de semiconducteur, le niveau de Fermi se rapproche d'autant plus de la bande de porteurs majoritaires que le dopage est important. En particulier, $E_{Fn} = E_c$ pour $N_d - N_a = N_c$ et $E_{Fp} = E_v$ pour $N_a - N_d = N_v$. En d'autres termes le semiconducteur devient dégénéré de type n quand la densité de donneurs excédentaires $N_d - N_a$ devient égale à la densité équivalente d'états de la bande de conduction. Il en est de même pour le semiconducteur de type p. En fait

les expressions (1-155) ont été établies dans le cadre de l'approximation de Boltzmann et ne sont plus valables lorsque le niveau de Fermi atteint une bande permise. Néanmoins ces expressions permettent d'obtenir une estimation du dopage qu'il est nécessaire de réaliser pour dégénérer de type n (ou de type p), un semiconducteur, $N_d - N_a > N_c$ (ou $N_a - N_d > N_v$). Le type du semiconducteur et la densité de porteurs majoritaires sont obtenus expérimentalement par la mesure du coefficient de Hall.

Il est souvent utile de définir la position du niveau de Fermi par sa distance au niveau de Fermi intrinsèque E_{Fi} du matériau (Fig 1-20).

Dans un semiconducteur intrinsèque

$$n=p=n_i=N_c e^{-(E_c-E_{Fi})/kT}=N_v e^{(E_v-E_{Fi})/kT}$$

Dans un semiconducteur dopé

$$n=N_c e^{-(E_c-E_F)/kT} \quad p=N_v e^{(E_v-E_F)/kT}$$

En posant $e\phi_{Fi} = E_F - E_{Fi}$, les expressions précédentes permettent d'écrire

$$n=n_i e^{+e\phi_{Fi}/kT} \qquad \text{(1-156-a)}$$

$$p=n_i e^{-e\phi_{Fi}/kT} \qquad \text{(1-156-b)}$$

$e\phi_{Fi} > 0$ entraîne $n > n_i$ et $p < n_i$, le semiconducteur est de type n

$e\phi_{Fi} < 0$ entraîne $p > n_i$ et $n < n_i$, le semiconducteur est de type p

Dans certains cas (Le transistor MOS par exemple), on choisit la convention inverse et $e\phi_{Fi}=-E_F+E_{Fi}$ ce qui conduit à ϕ_{Fi} positif pour du silicium p.

Les expressions (1-156) donnent

$$e\phi_{Fi}=kT\,Ln(n/n_i)=-kT\,Ln(p/n_i)$$

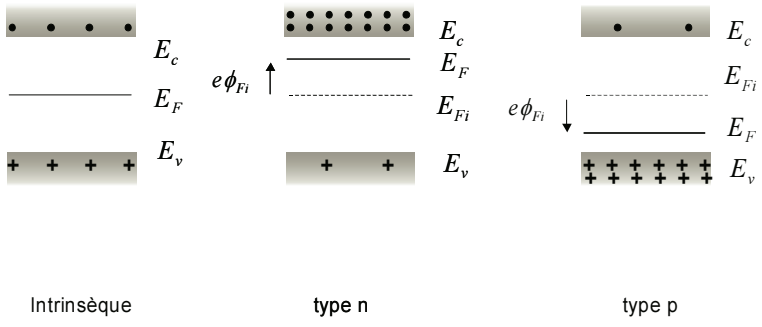


Figure 1-20 : Position du niveau de Fermi dans les différents types de semiconducteur

Dans un semiconducteur de type n, $n = N_d - N_a$, dans un semiconducteur de type p, $p = N_a - N_d$, soit

SC type n $$e\phi_{Fi} = kTLn\left((N_d - N_a)/n_i\right) \quad \text{(1-157-a)}$$

SC type p $$e\phi_{Fi} = -kTLn\left((N_a - N_d)/n_i\right) \quad \text{(1-157-b)}$$

1.4.5. *Evolution avec la température*

Lorsque la température du semiconducteur est abaissée, les densités de donneurs ou d'accepteurs ionisés diminuent progressivement de même que les densités de porteurs libres.

La densité de donneurs neutres N_d^0 est donnée par le produit de la densité de donneurs N_d par la probabilité d'occupation de l'état d'énergie associé au donneur.

$$N_d^o = N_d \frac{1}{1 + (1/2)e^{(E_d - E_F)/kT}} \quad \text{(1-158)}$$

où E_d représente l'énergie de l'état associé au donneur. Le facteur 1/2 qui apparaît au dénominateur de l'expression (1-158) résulte de la spécificité de l'état donneur. Le donneur est analogue à l'atome d'hydrogène, son état fondamental est un état 1s deux fois dégénéré, mais la répulsion coulombienne électron-électron interdit que cet état soit occupé par deux électrons.

La densité de donneurs ionisés est donnée par

$$N_d^+ = N_d - N_d^o = N_d \frac{1}{1 + 2e^{(E_F - E_d)/kT}} \quad \text{(1-159)}$$

De la même manière la densité d'accepteurs ionisés, c'est-à-dire occupés par un électron s'écrit

$$N_a^- = N_a \frac{1}{1 + (1/4)e^{(E_a - E_F)/kT}} \quad \text{(1-160)}$$

En raison de la dégénérescence du sommet de la bande de valence, le degré de dégénérescence de l'état fondamental de l'accepteur est 4 dans les semiconducteurs cubiques, mais cet état ne peut être occupé que par un seul électron.

Considérons le cas d'un semiconducteur de type n, non dégénéré et dépourvu d'accepteurs. L'équation de neutralité électrique (1-150) s'écrit

$$n - p = N_d^+$$

En explicitant N_d^+ à partir de l 'expression (1-159) et n et p à partir des expressions (1-132), cette condition s'écrit

$$N_c\ e^{-(E_c-E_F)/kT} - N_v\ e^{(E_v-E_F)/kT} = N_d\ \frac{1}{1+2\ e^{(E_F-E_d)/kT}} \tag{1-161}$$

Il faut résoudre cette équation pour calculer d'abord $E_F(T)$ et ensuite $n(T)$ et $p(T)$.

a. Niveau de Fermi

Afin de simplifier les calculs et de disposer d'expressions analytiques simples, nous pouvons définir différents régimes de température. A basse température la densité d'électrons est conditionnée par la population du niveau donneur, la densité de trous est négligeable. Lorsque la température augmente les donneurs s'ionisent progressivement, la densité d'électrons devient de l'ordre de N_d, la densité de trous devient de l'ordre de n_i^2 / N_d. Il en résulte que la densité de trous reste négligeable tant que n_i n'est pas de l'ordre de N_d. Or n_i est donné par l'expression (1-146) de sorte que la condition $n_i << N_d$ se traduit par

$$kT << \frac{E_g}{2\ Ln\left[(N_c\ N_v)^{1/2} / N_d\right]}$$

Pour $N_c \approx N_v \approx 10^{19}\,cm^{-3}$ et $N_d \approx 10^{17}\,cm^{-3}$, n_i devient de l'ordre de N_d pour $kT \approx E_g / 10$. Nous fixerons à cette valeur la limite de température au-dessous de laquelle on peut négliger la densité de trous. L'équation (1-161) s'écrit alors

$$N_c\ e^{-(E_c-E_F)/kT} - N_d\ \frac{1}{1+2\ e^{(E_F-E_d)/kT}} = 0 \tag{1-162}$$

soit

$$N_c e^{(E_F-E_d)/kT} e^{(E_d-E_c)/kT} + 2N_c e^{2(E_F-E_d)/kT} e^{(E_d-E_c)/kT} - N_d = 0 \tag{1-163}$$

En multipliant par $\frac{1}{2N_c} e^{(E_c-E_d)/kT}$, on obtient une équation du second degré de la forme

$$\left(e^{(E_F-E_d)/kT}\right)^2 + \frac{1}{2}\left(e^{(E_F-E_d)/kT}\right) - \frac{1}{2}\frac{N_d}{N_c} e^{(E_c-E_d)/kT} = 0 \tag{1-164}$$

La racine positive de cette expression ($e^x > 0$) s'écrit

$$e^{(E_F-E_d)/kT}=\frac{1}{4}\left[-1+\left(1+8\frac{N_d}{N_c}e^{(E_c-E_d)/kT}\right)^{1/2}\right] \tag{1-165}$$

d'où l'expression du niveau de Fermi en fonction de la température

$$E_F = E_d + kT\,Ln\left\{\frac{1}{4}\left[-1+\left(1+8\frac{N_d}{N_c}e^{(E_c-E_d)/kT}\right)^{1/2}\right]\right\} \tag{1-166}$$

Pour étudier la variation de E_F dans cette gamme de température définie par $kT < E_g/10$ on peut considérer deux régions différentes :

- cas $kT \prec\prec E_C - E_d$

On peut négliger 1 sous la racine et -1 devant la racine. On obtient alors une expression de E_F valable aux très basses températures

$$E_F \approx \frac{E_c+E_d}{2}+\frac{kT}{2}Ln(N_d/2N_c) \tag{1-167}$$

Cette expression est comparable à l'expression (1-147) en remplaçant E_v par E_d et N_v par $N_d/2$, le niveau donneur joue le rôle que jouait la bande de valence dans le matériau intrinsèque.

Quant la température tend vers zéro, $kT \to 0$ *et* $Ln(N_d/2N_c) \to \infty$ car N_c varie comme $T^{3/2}$. Mais la puissance l'emporte sur le logarithme de sorte que le deuxième terme de l'expression (1-167) tend vers zéro. Il en résulte qu'au zéro absolu le niveau de Fermi est donné par

$$E_F(T=0)=\frac{E_c+E_d}{2} \tag{1-168}$$

Le niveau de Fermi est situé au milieu entre la bande de conduction et le niveau donneur. On retrouve l'analogie citée plus haut avec le semiconducteur intrinsèque dans lequel au zéro degré absolu le niveau de Fermi est situé au milieu du gap.

Quand la température augmente à partir de $T = 0$, le niveau de Fermi s'élève tant que N_c, qui varie comme $T^{3/2}$ (Eq. 1-136-a), est inférieur à N_d. Il passe par un maximum pour $N_c = N_d$ puis descend.

- cas $kT \succ\succ E_C - E_d$

Dans cette gamme de température, on peut utiliser un développement limité de la racine dans l'expression (1-166) et on obtient

$$E_F = E_c - kT\ Ln(N_c/N_d) \qquad (1\text{-}169)$$

Dans cette gamme de température, qui contient la température ambiante pour la plupart des semiconducteurs, la densité équivalente d'états N_c est supérieure à la densité de donneurs N_d de sorte que $Ln(N_c/N_d)$ est positif. Il en résulte que le niveau de Fermi descend suivant une loi en température pratiquement linéaire. Lorsque la température devient supérieure à $E_g/10$, l'approximation précédente n'est plus justifiée et on ne peut plus négliger la densité de trous. La densité de porteurs intrinsèques n_i devient d'abord comparable puis supérieure à la densité de donneurs. Lorsque la condition $n_i >> N_d$ est satisfaite, on tend vers un régime dans lequel les porteurs intrinsèques jouent le rôle essentiel, les donneurs deviennent négligeables. Ce régime est appelé *régime intrinsèque*. Le niveau de Fermi est alors donné par l'expression (1-147). Il monte avec la température si $N_v > N_c$, il descend dans le cas contraire. A partir des expressions (1-167, 169 et 147), on peut tracer l'allure de la courbe de variations de E_F avec la température (Fig. 1-21).

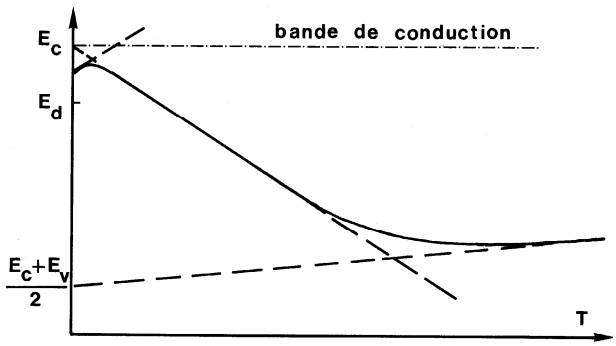


Figure 1-21 : Allure de variation du niveau de Fermi avec la température dans un semiconducteur de type n avec $N_V > N_C$

b. Densité de porteurs

A partir de l'évolution du niveau de Fermi, on obtient les évolutions des densités de porteurs par les expressions (1-132). Le semiconducteur étant de type n nous allons considérer l'évolution de la population électronique.

Dans la gamme des températures élevées, c'est-à-dire dans le régime intrinsèque, les donneurs ne jouent plus aucun rôle significatif, le niveau de Fermi est donné par l'expression (1-147) et la densité de porteurs $n = p = n_i$ par l'expression (1-146). En échelle logarithmique, la densité de porteurs varie linéairement en fonction de $1/T$ avec une pente égale à $E_g/2$.

Dans la gamme de plus basses températures, le niveau de Fermi est d'abord donné par l'expression (1-169) qui peut s'écrire

$$\frac{E_c - E_F}{kT} = Ln\,\frac{N_d}{N_c} \tag{1-170}$$

En portant cette expression dans l'expression (1-132-a) on obtient

$$n = N_d \tag{1-171}$$

Dans cette gamme de température, la densité d'électrons est constante et égale à la densité de donneurs. L'énergie thermique est suffisante pour ioniser tous les donneurs mais insuffisante pour créer un nombre conséquent de porteurs intrinsèques. Ce régime est appelé *régime d'épuisement des donneurs*.

Enfin, dans la gamme des très basses températures, le niveau de Fermi est donné par l'expression (1-167) qui permet d'écrire

$$\frac{E_F - E_c}{kT} = -\frac{E_c - E_d}{2kT} + Ln\,(N_d/2N_c)^{1/2} \tag{1-172}$$

En portant cette expression dans l'expression (1-132-a), on obtient

$$n = \left(\frac{1}{2}N_d N_c\right)^{1/2} e^{-(E_c - E_d)/2kT} \tag{1-173}$$

On retrouve un régime comparable au régime intrinsèque dans lequel $N_d/2$ joue le même rôle que N_v et $E_c - E_d$ le même rôle que $E_g = E_c - E_v$. En échelle logarithmique la densité d'électrons varie linéairement en fonction de $1/T$ avec une pente égale à $-(E_c - E_d)/2$. Lorsque T tend vers zéro tous les donneurs deviennent neutres, $n \to 0$, c'est le *régime de gel* des électrons.

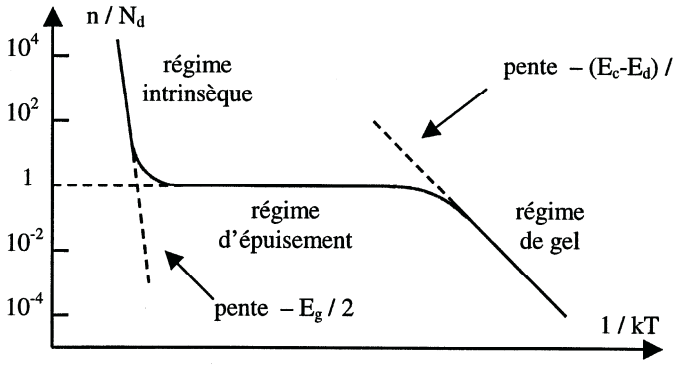


Figure 1-22 : Allure de variation de la densité de porteurs libres avec la température

A partir des expressions (1-146, 171 et 173) on peut tracer l'allure de la courbe de variation de la densité d'électrons en fonction de la température. Cette courbe est représentée en échelle logarithmique en fonction de $1/T$ sur la figure (1-22). Les trois régimes de fonctionnement sont schématisés sur la figure (1-23).

Nous avons considéré le cas d'un semiconducteur ne contenant que des donneurs. Dans la pratique les deux types d'impuretés sont présents, avec dans un semiconducteur de type n la condition $N_d > N_a$. Les niveaux accepteurs étant à plus basse énergie que les niveaux donneurs, N_a donneurs s'ionisent pour peupler les niveaux accepteurs et les $N_d - N_a$ donneurs restants sont susceptibles de peupler la bande de conduction. On dit qu'il y a *compensation partielle* des donneurs par les accepteurs.

L'équation de neutralité électrique est l'équation (1-150) dans laquelle n, p, N_d^+ et N_a^- sont donnés respectivement par les expressions (1-132, 159 et 160). On simplifie considérablement le problème en supposant qu'aux températures pratiques les conditions et $p << n$ et $N_a^- = N_a$ sont réalisées. En d'autres termes tous les accepteurs sont ionisés par les donneurs. L'équation de neutralité qu'il faut résoudre pour obtenir E_F et par suite n, se réduit donc à $n - N_d^+ + N_a = 0$, soit

$$N_c\, e^{-(E_c - E_F)/kT} - N_d \frac{1}{1 + 2\, e^{(E_F - E_d)/kT}} = -N_a \qquad \text{(1-174)}$$

La différence essentielle avec le cas N_a=0 est qu'à T=0 K le niveau de Fermi ne converge pas entre E_d et E_c (Fig.1-21), mais sur le niveau E_d, partiellement occupé en raison de la compensation partielle des N_d donneurs par les N_a accepteurs.

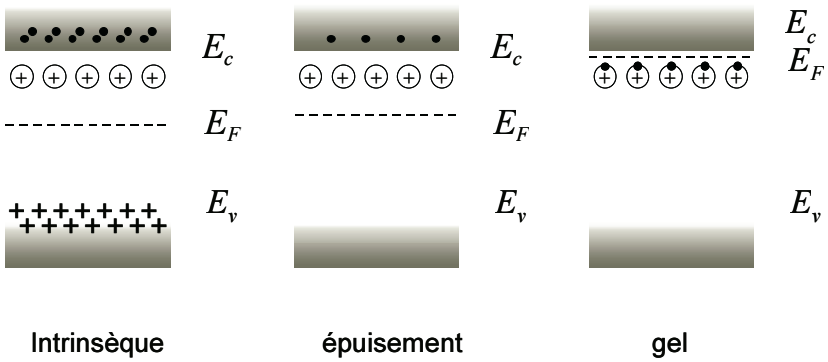


Figure 1-23 : Les régimes de fonctionnement

1.5. Le semiconducteur hors équilibre

1.5.1. *Courants dans le semiconducteur*

a. Rappels sur le courant

La notion de courant est généralement connue mais certaines précisions peuvent être apportées. Le courant est défini de manière générale en mécanique classique comme le flux de particules chargées de charge q à travers une surface comme le montre la figure 1-24. En calculant la charge qui traverse la surface ds par unité de temps, ce qui constitue la définition du courant, on en déduit facilement l'expression du vecteur *densité de courant* $\mathbf{J}$ défini par la relation.

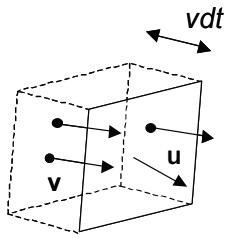


Figure 1-24 : La densité de courant

$$\frac{dQ}{dt} = \mathbf{J} \cdot \mathbf{u} \cdot ds$$

Dans cette expression $\mathbf{u}$ est le vecteur unitaire normal à la surface.

On généralise facilement à une surface quelconque par intégration.

Dans le cas d'une surface infinitésimale, on suppose que toutes les charges ont la même vitesse $\mathbf{v}$ et il est facile de constater que :

$$\mathbf{J} = qn \cdot \mathbf{v} \qquad (1\text{-}175)$$

Dans cette expression n est la densité de charge au niveau de la surface considérée. Dans le cas des électrons la charge est négative et

$$\mathbf{J} = -en \cdot \mathbf{v}$$

Quand la surface est traversée par des particules de vitesses différentes on somme les contributions correspondant à chaque vitesse.

$$\mathbf{J} = \sum_i eq_i \cdot \mathbf{v}_i \qquad (1\text{-}176)$$

Les courants dans le semiconducteur résultent du déplacement des porteurs de charge, électrons et trous, sous l'action d'une force. L'origine de la force peut être un champ électrique ou un gradient de concentration. Dans le premier cas, le courant est dit de *conduction*, dans le second il est dit de *diffusion*.

En l'absence de champ électrique les porteurs libres dans le semiconducteur sont animés, comme les molécules d'un gaz dans une enceinte, d'un mouvement brownien. Entre deux collisions le mouvement est rectiligne uniforme, caractérisé

par une vitesse que l'on appelle *vitesse thermique* du porteur. On obtient cette vitesse en écrivant que l'énergie cinétique du porteur est égale à l'énergie thermique

$$\frac{1}{2}m^* v_{th}^2 = \frac{3}{2}kT$$

de sorte que

$$v_{th} = \sqrt{3kT/m^*}$$

Dans le silicium, $m_e \approx m_o$ et $m_h \approx 0{,}5m_o$, à la température ambiante la vitesse thermique des porteurs est de l'ordre de 10^7 cm/s.

Le temps entre deux collisions est appelé *temps de collision* τ_c, il est fonction de la pureté du matériau et est de l'ordre de 10^{-13} à 10^{-12} s.
La distance parcourue entre deux collisions est donnée par

$$\ell = v_{th}\tau_c \tag{1-177}$$

ℓ est le *libre parcours moyen* du porteur, il est de l 'ordre de 100 à 1000 Å. Les collisions étant isotropes, toutes les directions du vecteur vitesse à l'issue d'un choc sont équiprobables. Il en résulte que la vitesse moyenne du porteur est nulle.

En présence d'un champ électrique $\mathbf{E}$ appliqué à l'instant $t=t_o$, un porteur de charge q est soumis à une force $\mathbf{F}=q\mathbf{E}$. La composante de vitesse instantanée $v_i(t)$ du porteur, dans la direction du champ, est alors donnée par l'équation de la dynamique

$$qE = m^* \frac{dv_i(t)}{dt} \tag{1-178}$$

En intégrant cette équation de t_o à $t_o + t$ on obtient

$$v_i(t) = \frac{qE}{m^*}t \tag{1-179}$$

Cette composante de vitesse augmente donc linéairement entre deux chocs, mais dans la mesure où ces derniers sont isotropes elle s'annule statistiquement à chaque choc (Fig. 1-25).

La variation de $v_i(t)$ est périodique. Si on appelle τ_c la valeur moyenne du temps de collision, la valeur moyenne de la vitesse est donnée par

$$v = \frac{1}{\tau_c}\int_0^{\tau_c} v_i(t)dt = \frac{1}{\tau_c}\frac{qE}{m^*}\frac{\tau_c^2}{2}$$

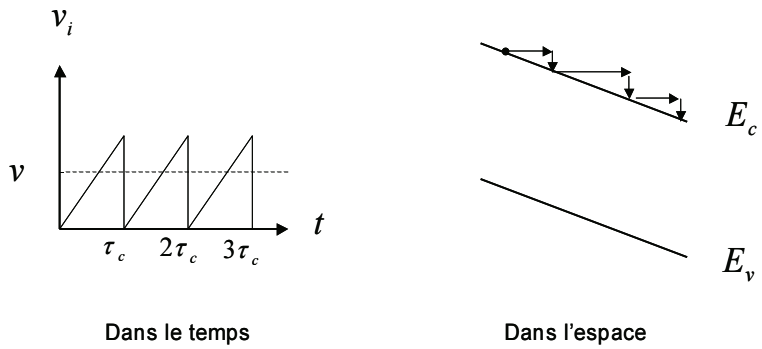


Figure 1.25 : Mouvement d'une charge dans un solide

En posant $\tau = \tau_c / 2$, l'expression s'écrit

$$v = \pm \mu E \tag{1-180}$$

avec

$$\mu = \left| q\tau / m^* \right| \tag{1-181}$$

La grandeur μ est appelée *mobilité des porteurs* et τ le *temps de relaxation*. La vitesse **v** est appelée *vitesse de dérive* ou *vitesse d'entraînement* ou *vitesse drift*.

La mobilité est par définition une grandeur positive, elle mesure l'aptitude des porteurs à se déplacer dans le réseau cristallin et s'exprime en $cm^2V^{-1}s^{-1}$. Elle est d'autant plus grande que le cristal est pur et que la masse effective des porteurs est faible. C'est un paramètre fondamental qui conditionne le fonctionnement des composants en haute fréquence. Les vitesses des électrons et des trous s'écrivent

$$\mathbf{v}_n = -\mu_n \mathbf{E} \qquad \mathbf{v}_p = +\mu_p \mathbf{E}$$

Si on exclut pour l'instant le domaine des champs forts, qui correspond au fonctionnement des composants haute fréquence, la mobilité est une constante et par suite la vitesse moyenne des porteurs est proportionnelle au champ électrique.

Revenons maintenant au courant en considérant une surface traversées par des charges mobiles. Le mouvement des charges est la résultante de l'agitation thermique qui est isotrope et d'un mouvement global dans le sens du champ appliqué. Le calcul rigoureux du courant fait appel à l'équation de Boltzmann mais peut se simplifier en séparant la partie due à l'agitation thermique (courant de diffusion) et la partie due au mouvement moyen d'ensemble sous l'effet du champ (courant de dérive). Etudions dans un premier temps le courant de dérive.

Au déplacement des charges correspond un courant dont la densité est définie comme la quantité de charge qui traverse l'unité de surface pendant l'unité de temps, soit pour chaque type de porteurs

$$\mathbf{J}_n = -ne\mathbf{v}_n = ne\mu_n \mathbf{E} \qquad (1\text{-}182\text{-a})$$

$$\mathbf{J}_p = +pe\mathbf{v}_p = pe\mu_p \mathbf{E} \qquad (1\text{-}182\text{-b})$$

Le courant résultant du déplacement des électrons et des trous sous l'action du champ électrique, s'écrit

$$\mathbf{J}_c = \mathbf{J}_n + \mathbf{J}_p = \sigma \mathbf{E} \qquad (1\text{-}183)$$

avec

$$\sigma = ne\mu_n + pe\mu_p \qquad (1\text{-}184)$$

$\mathbf{J}_c$ est appelé *courant de conduction*, σ est la *conductivité* du matériau. Si le semiconducteur est intrinsèque $n=p=n_i$

$$\sigma_i = n_i e(\mu_n + \mu_p) \qquad (1\text{-}185)$$

Il faut noter qu'en raison de la nature antiliante des fonctions d'ondes électroniques et de la nature liante des fonctions d'onde des trous, le temps de relaxation des électrons est beaucoup plus important que celui des trous. En outre dans la plupart des semiconducteurs la masse effective des électrons est beaucoup plus faible que celle des trous (Tab. 1-2). Il en résulte que la mobilité des électrons est toujours beaucoup plus importante que celle des trous. En outre les mobilités augmentent beaucoup quand la température diminue car le temps de collision augmente. Le temps de collision est essentiellement conditionné par les chocs d'une part sur les atomes du cristal et d'autre part sur les impuretés ionisées. Lorsque la température diminue les vibrations thermiques des atomes autour de leur position d'équilibre diminuent, les atomes occupent alors une section efficace de collision plus faible et le temps de collision des porteurs augmente. A basse température T<100K, la diffusion des porteurs par le réseau cristallin devient négligeable et la mobilité est conditionnée par les interactions coulombiennes des porteurs avec les impuretés ionisées. A la température ambiante, les mobilités des porteurs dans les semiconducteurs les plus utilisés dans la réalisation de composants sont les suivantes en $cm^2\ V^{-1}\ s^{-1}$

	Si	**Ge**	**GaP**	**GaAs**	**GaSb**	**InP**	**InAs**	**InSb**
μ_n	1 350	3 600	300	8 000	5 000	4 500	30 000	80 000
μ_p	480	1 800	150	300	1 000	100	450	450

A partir des grandeurs locales j_c et σ que nous venons de définir, on retrouve aisément les lois d'Ohm et de Joule macroscopiques.

- *Loi d'Ohm*

La relation $j=\sigma E$, qui s'écrit aussi $j=\rho j$ où $\rho=1/\sigma$ est la résistivité du matériau, n'est autre que l'écriture microscopique de la loi d'Ohm. En l'intégrant sur un barreau semiconducteur homogène de longueur finie ℓ et de section finie s, on obtient la loi

$$V=RI \qquad \text{où} \quad V=E\ell \qquad I=js \qquad R=\rho\ell/s$$

- *Loi de Joule*

Considérons le point de vue énergétique. Entre deux chocs, l'électron acquiert de la part du champ électrique un excès d'énergie ΔE qu'il cède au réseau au moment du choc, cette énergie est alors dissipée sous forme thermique. Excès d'énergie acquis par l'électron entre deux chocs s'écrit

$$\Delta E=\int_0^{\tau_c} \mathbf{F}.d\mathbf{r} = \int_0^{\tau_c} F v_i(t)dt$$

où $v_i(t)$ représente la composante de vitesse instantanée de l'électron dans la direction du champ électrique. En explicitant F et $v_i(t)$ l'expression précédente s'écrit

$$\Delta E=\int_o^{\tau_c} qE\frac{qE}{m^*}t\,dt=\frac{(qE)^2}{m^*}\frac{\tau_c^{\,2}}{2}$$

Ainsi pour chaque électron, la puissance moyenne fournie par le champ électrique et dissipée par le matériau est donnée par

$$\Delta w=\frac{\Delta E}{\tau_c}=\frac{(qE)^2}{m^*}\frac{\tau_c}{2}=\frac{(qE)^2}{m^*}\tau$$

Si la densité d'électrons dans le semiconducteur est n, la puissance moyenne dissipée par l'unité de volume du matériau s'écrit

$$w=n\Delta w=n\frac{(qE)^2}{m^*}\tau$$

Cette expression s'écrit sous la forme

$$w = nq\frac{q\tau}{m^*}E^2 = nq\mu E^2 = \sigma E^2 = \frac{\sigma^2 E^2}{\sigma} = \frac{j^2}{\sigma} = \rho j^2$$

La relation $w = \rho j^2$ n'est autre que l'écriture microscopique de la loi de Joule. En l'intégrant sur un barreau semiconducteur homogène de longueur finie ℓ et de section finie *s*, on obtient la loi

$$W = RI^2 \qquad \text{avec } W = ws\ell \quad R = \rho\ell / s \quad I = js$$

Etudions maintenant le courant de diffusion ou plus exactement la composante de diffusion du courant global. On suppose le champ appliqué nul. Si la densité est constante, le nombre de porteurs traversant la surface dans un sens est égal au nombre de porteurs traversant la surface dans le sens opposé et le courant total est nul. Quand la densité varie de manière significative de part et d'autre de la surface cela n'est plus vrai. Un modèle grossier à une dimension aide à dimensionner le phénomène.

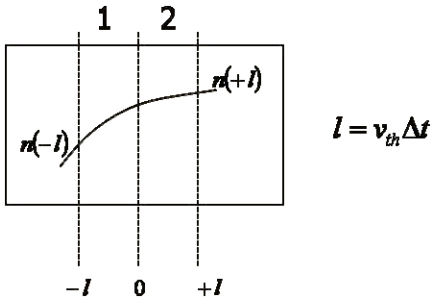


Figure 1-26 : Le courant de diffusion

Dans le cas de la diffusion (figure 1-26), la densité varie rapidement de part et d'autre de la surface. Les électrons supposés soumis à une forte agitation thermique traversent la surface dans les deux sens avec des vitesses variables. Pour simplifier considérons un cas à une dimension et donc deux flux de sens opposés. On considére que tous les électrons ont la même vitesse en valeur abolue v_{th}. Les deux flux pendant un intervalle de temps Δt correspondant aux charges comprises entre $-v_{th}\Delta t$ et $v_{th}\Delta t$ sont donc, l'origine des distances étant au niveau de la surface

$$I_1 = \frac{1}{2}n(-v_{th}\Delta t)\cdot v_{th}\cdot e$$
$$I_2 = \frac{1}{2}n(v_{th}\Delta t)\cdot v_{th}\cdot e$$

La densité de courant traversant la surface est donc

$$J = \frac{1}{2} n(v_{th}\Delta t)\cdot v_{th}\cdot e - \frac{1}{2} n(-v_{th}\Delta t)\cdot v_{th}\cdot e$$

Un simple développement limité conduit à

$$J = \frac{1}{2}\frac{dn}{dx} 2 v_{th}\Delta t \cdot e v_{th}$$

Le choix de Δt peut se faire en se limitant aux électrons qui n'ont pas subi de collisions. C'est donc le temps de collision τ_c. On obtient alors

$$J = eD_n \frac{dn}{dx} \qquad \text{(1-186)}$$

Avec

$$D_n = v_{th}^{\ 2}\tau_c$$

Ce résultat se généralise à trois dimensions pour donner la relation de Fick

Lorsque les porteurs libres ne sont pas distribués uniformément dans le semiconducteur il sont soumis au processus général de diffusion. Leur mouvement s'effectue dans un sens qui tend à uniformiser leur distribution spatiale. Le flux de porteurs est, conformément à la première loi de Fick, proportionnel à leur gradient de concentration. Si on appelle n_x^d le nombre d'électrons qui diffusent par seconde dans la direction $o\vec{x}$, à travers l'unité de surface perpendiculaire à $o\vec{x}$, la loi de Fick s'écrit

$$n_x^d = -D_n\, dn/dx \qquad \text{(1-187a)}$$
$$p_x^d = -D_p\, dp/dx \qquad \text{(1-187b)}$$

Les constantes D_n *et* D_p sont appelées *constantes de diffusion* des électrons et des trous respectivement. Le signe moins traduit le fait que les porteurs diffusent dans la direction de plus faible concentration. A trois dimensions les flux de porteurs s'écrivent

$$\mathbf{n}^d = -D_n \mathbf{grad} n \qquad \text{(1-188-a)}$$
$$\mathbf{p}^d = -D_p \mathbf{grad} p \qquad \text{(1-188-b)}$$

Aux déplacements des porteurs de charge correspondent des courants appelés *courants de diffusion*

$$\mathbf{J}_{dn} = eD_n \mathbf{grad} n \qquad \text{(1-189-a)}$$
$$\mathbf{J}_{dp} = -eD_p \mathbf{grad} p \qquad \text{(1-189-b)}$$

Le courant total de diffusion s'écrit donc

$$\mathbf{J}_d = eD_n \mathbf{grad} n - eD_p \mathbf{grad} p \tag{1-190}$$

On obtient donc pour le courant total

$$\mathbf{J}_n = ne\mu_n \mathbf{E} + eD_n \mathbf{grad} n \tag{1-191-a}$$

$$\mathbf{J}_p = pe\mu_p \mathbf{E} - eD_p \mathbf{grad} p \tag{1-191-b}$$

Ces relations seront largement utilisées dans la suite.

b. Masses effectives de conductivité

L'expression (1-181) montre que la mobilité des porteurs est proportionnelle au rapport τ/m^*. Nous pouvons supposer isotrope le temps de relaxation τ, mais pas la masse effective m^*. Cette dernière n'est un simple scalaire que pour les électrons de conduction d'un semiconducteur univallée. Dans ce cas $m^*=m_e$ et la mobilité des électrons s'écrit

$$\mu_n = e\tau/m_e \tag{1-192}$$

Dans un semiconducteur multivallée, l'expression précédente est modifiée par le fait que les électrons de chaque vallée sont caractérisés par une masse effective longitudinale m_l et une masse effective transverse m_t. De même, en ce qui concerne la bande de valence, la présence de deux types de trous, les trous lourds de masse effective m_{hh} et les trous légers de masse effective m_{lh}, modifie l'expression (1-186). Toutefois, la mobilité macroscopique étant nécessairement un scalaire dans un semiconducteur cubique, on peut l'écrire dans tous les cas sous une forme analogue à l'expression (1-192) en définissant des masses effectives équivalentes que l'on appelle *masses effectives de conductivité*.

Les masses effectives de conductivité conditionnent les phénomènes de conduction au même titre que les masses effectives de densité d'états conditionnent les populations de porteurs libres.

- *Bande de conduction* et *semiconducteur multivallée*

Considérons tout d'abord le cas du silicium polarisé dans une direction de haute symétrie. La bande de conduction du silicium est constituée de 6 vallées équivalentes portées par les axes $(\overset{\pm}{1}00)$, $(0\overset{\pm}{1}0)$, $(00\overset{\pm}{1})$, que nous appellerons x,y,z afin de simplifier l'écriture. Nous supposerons le champ électrique parallèle à la direction x.

Le courant électronique macroscopique résulte de la somme des courants transportés par les électrons de chacune des 6 vallées. Mais dans leur mouvement

suivant la direction x (direction du champ électrique), les électrons des 2 vallées x sont caractérisés par leur masse effective longitudinale m_l , alors que les électrons des 4 vallées y et z sont caractérisés par leur masse effective transverse m_t . Ainsi, en appelant j_{nx}, j_{ny} et j_{nz} les courants transportés respectivement par les électrons des 2 vallées x, des 2 vallées y et des 2 vallées z , le courant total s'écrit

$$j_n = j_{nx} + j_{ny} + j_{nz}$$

avec

$$j_{nx} = \sigma_{nx} E = 2\frac{n}{6} e \mu_{nx} E = n e^2 \tau_n \frac{1}{3m_l} E \qquad j_{ny} = j_{nz} = n e^2 \tau_n \frac{1}{3m_t} E$$

Le courant s'écrit donc

$$j_n = ne^2 \tau_n \frac{1}{3}\left(\frac{1}{m_\ell} + \frac{2}{m_t}\right) E$$

ou $j_n = n e \mu_n E$ avec $\mu_n = \dfrac{e\,\tau_n}{m_{ec}}$ et

$$\frac{1}{m_{ec}} = \frac{1}{3}\left(\frac{1}{m_l} + \frac{2}{m_t}\right) \tag{1-193}$$

m_{ec} est appelé *masse effective de conductivité* des électrons.

Ce calcul simple développé pour un champ électrique dirigé suivant l'axe principal d'un ellipsoïde, peut être étendu sans difficulté au cas d'un champ de direction quelconque, le résultat est le même dans la mesure où le semiconducteur est cubique.

Considérons le cas du germanium, c'est un semiconducteur multivallée dont la bande de conduction présente 4 minima équivalents situés en bord de zone de Brillouin dans les directions (111), $(\bar{1}11)$, $(\bar{1}\,\bar{1}\,1)$, $(11\bar{1})$ et $(1\,1\,\bar{1})$. Le courant électronique macroscopique est la somme des courants transportés par les électrons de chacune des 4 vallées. Les axes principaux des ellipsoïdes des énergies associés à chaque vallée sont

$[1\ 1\ 1]$	$\vec{i}_1 = [1\,\bar{1}\,0]$	$\vec{j}_1 = [1\,1\,\bar{2}]$	$\vec{k}_1 = [1\,1\,1]$
$[\bar{1}\ 1\ 1]$	$\vec{i}_2 = [1\,1\,0]$	$\vec{j}_2 = [\bar{1}\,1\,\bar{2}]$	$\vec{k}_2 = [\bar{1}\,1\,1]$
$[\bar{1}\,\bar{1}\ 1]$	$\vec{i}_3 = [\bar{1}\,1\,0]$	$\vec{j}_3 = [\bar{1}\,\bar{1}\,\bar{2}]$	$\vec{k}_3 = [\bar{1}\,\bar{1}\,1]$
$[1\ \bar{1}\ 1]$	$\vec{i}_4 = [\bar{1}\,\bar{1}\,0]$	$\vec{j}_4 = [1\,\bar{1}\,\bar{2}]$	$\vec{k}_4 = [1\,\bar{1}\,1]$

Les matrices de rotation permettent de passer du système d'axes cristallographiques $\vec{x}=[100]$, $\vec{y}=[010]$, $\vec{z}=[001]$ au système d'axes principaux de chacun des ellipsoïdes, et inversement

$$\begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = [R_\alpha] \begin{Bmatrix} i_\alpha \\ j_\alpha \\ k_\alpha \end{Bmatrix}$$

$$\begin{Bmatrix} i_\alpha \\ j_\alpha \\ k_\alpha \end{Bmatrix} = [R_\alpha]^{-1} \begin{Bmatrix} x \\ y \\ z \end{Bmatrix}$$

avec α=1, 2, 3, 4. Les matrices de rotation s'écrivent

$$[R_1] = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ -1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 0 & -\sqrt{2/3} & 1/\sqrt{3} \end{bmatrix} \qquad [R_1]^{-1} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -\sqrt{2/3} \\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix}$$

$$[R_2] = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{6} & -1/\sqrt{3} \\ 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 0 & -\sqrt{2/3} & 1/\sqrt{3} \end{bmatrix} \qquad [R_2]^{-1} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{6} & 1/\sqrt{6} & -\sqrt{2/3} \\ -1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix}$$

$$[R_3] = \begin{bmatrix} -1/\sqrt{2} & -1/\sqrt{6} & -1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & -1/\sqrt{3} \\ 0 & -\sqrt{2/3} & 1/\sqrt{3} \end{bmatrix} \qquad [R_3]^{-1} = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{6} & -1/\sqrt{6} & -\sqrt{2/3} \\ -1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix}$$

$$[R_4] = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ -1/\sqrt{2} & -1/\sqrt{6} & -1/\sqrt{3} \\ 0 & -\sqrt{2/3} & 1/\sqrt{3} \end{bmatrix} \qquad [R_4]^{-1} = \begin{bmatrix} -1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & -1/\sqrt{6} & -\sqrt{2/3} \\ 1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix}$$

Supposons le champ électrique parallèle à la direction *z*. Les composantes du champ électrique dans le système d'axes cristallographiques sont $0, 0, E$. Dans le système d'axes principaux de chacun des ellipsoïdes, ces composantes s'écrivent, en vertu des expressions des matrices de rotation.

$$E_{i1} = E_{i2} = E_{i3} = E_{i4} = 0$$

$$E_{j1} = E_{j2} = E_{j3} = E_{j4} = E_\perp = -\sqrt{2/3}\ E$$

$$E_{k1} = E_{k2} = E_{k3} = E_{k4} = E_{//} = \sqrt{1/3}\ E$$

A ces composantes du champ électrique correspondent, dans le système d'axes de chaque ellipsoïde, des composantes de courant électronique données par

$$\begin{aligned} j_{i1} = j_{i2} &= j_{i3} = j_{i4} = 0 \\ j_{j1} = j_{j2} &= j_{j3} = j_{j4} = -\sqrt{2/3}\ \sigma_{n\perp}\ E \\ j_{k1} = j_{k2} &= j_{k3} = j_{k4} = \sqrt{1/3}\ \sigma_{n//}\ E \end{aligned}$$

avec $\sigma_{n\perp} = n\, e\, \mu_{n\perp}/4$ et $\sigma_{n//} = n e \mu_{n//}/4$ où n représente le nombre total d'électrons de conduction, $\mu_{n\perp} = e\,\tau_n / m_t$ *et* $\mu_{n//} = e\,\tau_n / m_l$ où m_t *et* m_l sont les masses effectives transverse et longitudinale des électrons dans chacune des vallées.

On obtient les composantes dans le système d'axes cristallographiques, des contributions de chaque vallée, à partir des expressions (1-190) et de la relation (1-189-a), soit

$$\begin{aligned} j_{1x} = -j_{2x} = -j_{3x} = j_{4x} &= \frac{1}{3}(\sigma_{n//} - \sigma_{n\perp})E \\ j_{1y} = j_{2y} = -j_{3y} = -j_{4y} &= \frac{1}{3}(\sigma_{n//} - \sigma_{n\perp})E \\ j_{1z} = j_{2z} = j_{3z} = j_{4z} &= \frac{1}{3}(\sigma_{n//} + 2\sigma_{n\perp})E \end{aligned}$$

Il faut noter que si on considère chacune des vallées séparément, les composantes $j_{\alpha x}$ *et* $j_{\alpha y}$ du courant ne sont pas nulles bien que les composantes E_x et E_y du champ électrique le soient. Ainsi à un vecteur champ électrique $\mathbf{E}$ correspond un vecteur densité de courant $\mathbf{j_\alpha}$ qui n'est pas parallèle à $\mathbf{E}$. Ceci résulte du fait que la conductivité associée aux électrons de chaque vallée prise isolément n'est pas un scalaire mais un tenseur. Dans le système d'axes principaux de l'ellipsoïde des énergies d'une vallée, ce tenseur est diagonal, ses composantes sont $\sigma_\perp, \sigma_\perp, \sigma_{//}$. La nature scalaire de la conductivité électronique globale du matériau est restaurée lorsque l'on calcule les composantes de la densité totale du courant électronique, résultant de la contribution simultanée des 4 vallées.

$$\begin{aligned} j_x &= \sum_\alpha j_{\alpha x} = 0 \\ j_y &= \sum_\alpha j_{\alpha y} = 0 \\ j_z &= \sum_\alpha j_{\alpha z} = \frac{4}{3}(\sigma_{//} + 2\sigma_\perp)E \end{aligned}$$

On retrouve l'isotropie du matériau cubique, la conductivité est un scalaire, le vecteur densité de courant est colinéaire au vecteur champ électrique. En explicitant $\sigma_{//}$ et $\sigma_\perp$, la densité de courant s'écrit

$$j = n\,e\,\mu_n\,E \qquad \text{avec } \mu_n = \frac{e\,\tau_n}{m_{ec}}$$

et

$$\frac{1}{m_{ec}} = \frac{1}{3}\left[\frac{1}{m_\ell} + \frac{2}{m_t}\right] \tag{1-194}$$

On retrouve le même résultat que pour le silicium, il en est de même pour n'importe quel semiconducteur multivallée cubique.

- *Bande de valence et masses effectives de conductivité*

Si on appelle p_h et p_l les densités respectives de trous lourds et légers, et j_{ph} et j_{pl} les courants correspondants, on peut écrire

$$j_{ph} = \sigma_{ph}\;E = p_h\;e\;\mu_{ph}\;E = e^2\;\tau_p\;\frac{p_h}{m_{hh}}\;E$$

$$j_{pl} = e^2\;\tau_p\;\frac{p_l}{m_{lh}}\;E$$

$$p = p_l + p_h$$

Le courant total de trous s'écrit

$$j_p = e^2\,\tau_p\left(\frac{p_h}{m_{hh}} + \frac{p_l}{m_{lh}}\right)E$$

ou

$$j_p = \frac{e^2\;\tau_p\;p}{m_h}\;E = p\;e\;\mu_p E$$

avec

$$\mu_p = \frac{e\,\tau_p}{m_h}$$

en posant

$$\frac{1}{m_h} = \frac{1}{p}\left(\frac{p_h}{m_{hh}} + \frac{p_l}{m_{lh}}\right) \tag{1-195}$$

Les densités de trous lourds et légers sont proportionnelles aux masses effectives à la puissance 3/2 (1-136), de sorte que

$$\frac{p_h}{p} = \frac{m_{hh}^{3/2}}{m_{hh}^{3/2} + m_{lh}^{3/2}} \qquad \frac{p_l}{p} = \frac{m_{lh}^{3/2}}{m_{hh}^{3/2} + m_{lh}^{3/2}}$$

ce qui donne

$$\frac{1}{m_h} = \frac{m_{hh}^{1/2} + m_{lh}^{1/2}}{m_{hh}^{3/2} + m_{lh}^{3/2}} \tag{1-196}$$

m_h est appelée *masse effective de conductivité des trous.*

c. Relations d'Einstein

La constante diffusion D et la mobilité μ d'un type de porteur traduisent l'aptitude des porteurs à se déplacer dans le réseau cristallin sous l'action d'une force. Dans le premier cas la force résulte d'un gradient de concentration, dans le second elle résulte d'un gradient de potentiel. Ces deux quantités ne sont donc pas indépendantes.

Considérons un barreau semiconducteur isolé à l'équilibre thermodynamique le niveau de Fermi est horizontal. En effet si aucune tension n'est appliquée, les notions développées dans le paragraphe 1.3.3 amènent à considérer que le niveau de Fermi ne varie pas en fonction de la position. Si la densité d'électrons est homogène, elle est donnée par l'expression (1-132-a), si elle n'est pas homogène suivant la direction x, elle est donnée par

$$n(x)=N_c e^{-(E_c(x)-E_F)/kT} \tag{1-197}$$

L'inhomogénéité de la densité d'électrons entraîne l'existence d'une force de diffusion et d'un champ électrique, qui créent un courant de conduction et un courant de diffusion. Le barreau étant isolé le courant résultant est nul, en d'autres termes le courant de diffusion est équilibré par le courant de conduction

$$n(x)\, e\, \mu_n\, E(x) + e\, D_n \frac{dn(x)}{dx} = 0$$

Le champ $E(x)$ résulte du gradient de concentration, on peut donc le calculer en fonction de $dn(x)/dx$. Si on appelle $V(x)$ le potentiel au point x, on peut écrire

$$E(x) = -\frac{dV(x)}{dx} = \frac{1}{e}\frac{dE_c(x)}{dx} \tag{1-198}$$

Cette relation amène quelques précisions. Les électrons situés en bas de la bande de conduction ont une énergie cinétique nulle. En effet, la vitesse définie par $\mathbf{v} = \frac{1}{\hbar}\frac{\partial E}{\partial \mathbf{k}}$ est nulle par définition du bas de la bande de conduction. L'énergie des électrons est donc strictement potentielle ce qui permet d'écrire la relation 1-198.

$$\frac{dE_c(x)}{dx} = \frac{dE_c(x)}{dn(x)}\frac{dn(x)}{dx}$$

En dérivant l'expression (1-197) on obtient

$$\frac{dn(x)}{dE_c(x)} = -\frac{1}{kT}n(x)$$

Le champ électrique s'écrit donc

$$E(x) = -\frac{kT}{e}\frac{1}{n(x)}\frac{dn(x)}{dx}$$

En explicitant $E(x)$ dans l'expression du courant total on obtient

$$-\mu_n kT + eD_n = 0$$

On obtiendrait une relation équivalente avec les trous, c'est la relation d'Einstein qui s'écrit

$$\frac{D_n}{\mu_n} = \frac{D_p}{\mu_p} = \frac{kT}{e} \tag{1-199}$$

Compte tenu de cette relation, les courants d'électrons et de trous peuvent s'écrire sous l'une ou l'autre des formes suivantes

$$\mathbf{J}_n = e\mu_n\left(n\mathbf{E} + \frac{kT}{e}\mathbf{grad}n\right) = eD_n\left(n\frac{e}{kT}\mathbf{E} + \mathbf{grad}n\right) \tag{1-200a}$$

$$\mathbf{J}_p = e\mu_p\left(p\mathbf{E} - \frac{kT}{e}\mathbf{grad}p\right) = eD_p\left(p\frac{e}{kT}\mathbf{E} - \mathbf{grad}p\right) \tag{1-200b}$$

Il faut noter que la relation d'Einstein a été établie à partir de l'expression (1-197) qui résulte de l'approximation de Boltzmann, justifiée uniquement dans un semiconducteur non dégénéré. Il en résulte que cette relation n'est plus valable dans un semiconducteur dégénéré.

Quand le niveau de Fermi n'est plus défini mais quand les populations de trous et d'électrons sont séparément en équilibre définis par des pseudo-niveaux de Fermi, on peut obtenir une relation entre le courant et le gradient du pseudo-niveau de Fermi. Il suffit de reprendre les expresions des densités de porteurs en remplaçant les niveaux de Fermi par les pseudo-niveaux de Fermi.

$$n(x) = N_c e^{-(E_c(x) - E_{Fn}(x))/kT}$$

$$p(x) = N_v e^{(E_v(x) - E_{Fp}(x))/kT}$$

En dérivant ces expressions on obtient par exemple pour les électrons

$$\frac{dn}{dx} = \frac{n(x)}{kT}\left(\frac{dE_{Fn}}{dx} - \frac{dE_c}{dx}\right)$$

Comme $\frac{dV}{dx} = -\frac{1}{e}\frac{dE_c}{dx}$ on obtient

$$\frac{dn}{dx} = \frac{n(x)}{kT}\left(\frac{dE_{Fn}}{dx} + e\frac{dV}{dx}\right)$$

Exprimons alors le courant en intégrant les relations d'Einstein

$$J_n = D_n\left(-n(x)\frac{e}{kT}\frac{dV}{dx} + \frac{dn}{dx}\right)$$

Il reste

$$J_n = \mu_n n(x)\frac{dE_{Fn}}{dx}$$

En trois dimensions on obtient pour les électrons et les trous

$$\mathbf{J_n} = \mu_n n(x)\mathbf{grad}E_{Fn} \qquad \text{(1-201-a)}$$

$$\mathbf{J}_p = \mu_p p(x)\mathbf{grad}E_{Fp} \qquad \text{(1-201-b)}$$

La variation du pseudo-niveau de Fermi est donc directement liée au courant traversant le dispositif. Notons que le pseudo-niveau de Fermi depend de la position dans le solide.

d. Courant de déplacement

Outre les courants de conduction et de diffusion, il existe en régime alternatif, un courant de déplacement donné par

$$\mathbf{J}_D = \frac{\partial \mathbf{D}}{\partial t}$$

où $\mathbf{D}=\varepsilon\mathbf{E}$ est le vecteur déplacement et ε la constante diélectrique du semiconducteur. L'introduction du courant de déplacement peut se faire à partir des équations de base de l'électromagnétisme rappelées ci-dessous.

Dans une première étape, il est possible à partir de la loi de Coulomb de définir un potentiel électrique défini par la relation :

$$\mathbf{E} = -\mathbf{grad}V \qquad \text{(1-202)}$$

Cette définition est possible car la force de Coulomb varie inversement proportionnellement au carré de la distance séparant les deux charges. Pour d'autres forces il est impossible de définir un potentiel. Il est également assez facile de

prouver le théorème de Gauss qui donne une expression simple du flux du champ électrique sur une surface entourant un ensemble de charges.

$$\int_S \mathbf{E} \cdot \mathbf{n} \cdot dS = \frac{Q}{\varepsilon} \tag{1-203}$$

Quand on choisit un volume infinitésimal, cette relation s'écrit

$$div\mathbf{E} = \frac{\rho}{\varepsilon} \tag{1-204}$$

La divergence étant définie par

$$div\mathbf{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} \tag{1-205}$$

Cette équation s'écrit également à partir de la relation (1-202)

$$\Delta V + \frac{\rho}{\varepsilon} = 0 \tag{1-206}$$

Cette relation permet de calculer le potentiel quand la densité de charge est connue. Il faut aussi savoir que le théorème de Gauss reste vrai quand les charges sont en mouvement. Ce n'est pas le cas pour la force de Coulomb. Les relations précédentes qui en dérivent (1-204 et 1-206) restent également valables quand les charges sont en mouvement.
Il est possible d'écrire le bilan des charges entrant et sortant d'un volume infinitésimal. On obtient alors

$$div\mathbf{J} + \frac{\partial \rho}{\partial t} = 0 \tag{1-207}$$

On en déduit donc

$$div\ \mathbf{J} + div\ \varepsilon \frac{\partial \mathbf{E}}{\partial t} = 0$$

Cette relation exprime que la divergence de la somme $\mathbf{J} + \varepsilon \frac{\partial \mathbf{E}}{\partial t}$ est nulle. Ce vecteur de divergence nulle a donc la propriété d'avoir un flux nul sur une surface fermée, ce qui exprime la conservation du courant. Le terme additionnel $\varepsilon \frac{\partial \mathbf{E}}{\partial t}$ est le courant de déplacement. Il est à considérer dans le modèle dynamique des composants. La conservation du courant est vérifiée à condition de considérer les deux termes, le courant de conduction et le courant de déplacement. La figure 1-27 exprime cette conservation dans le cas d'un fil conducteur cylindrique.

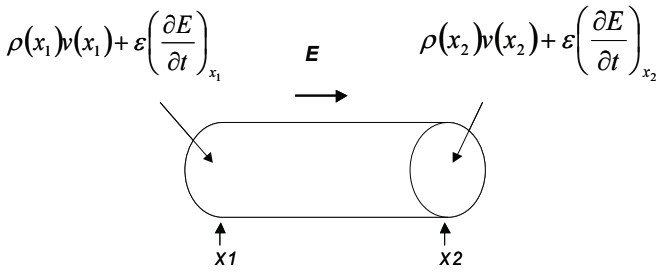


Figure 1-27 : Courant de déplacement pour un fil cylindrique.

Par symétrie cylindrique, le flux du champ est nul le long du cylindre puique le champ est perpendiculaire au vecteur normal. On déduit donc :

$$\rho(x_1)v(x_1)+\varepsilon\left(\frac{\partial E}{\partial t}\right)_{x_1} = \rho(x_2)v(x_2)+\varepsilon\left(\frac{\partial E}{\partial t}\right)_{x_2}$$

Si on néglige les variations temporelles du champ, on retrouve la conservation du courant au sens continu du terme.

Si le champ électrique est de la forme $E=E_o e^{j\omega t}$, le courant de déplacement s'écrit

$$J_D=j\varepsilon\omega E_o e^{j\omega t}$$

Dans un semiconducteur la constante diélectrique relative ε_r est de l'ordre de 10 à 15 de sorte que ε est de l'ordre de 10^{-10} Fm^{-1}. Il en résulte que le courant de déplacement ne présente une amplitude appréciable que pour des fréquences de supérieures au GH_3.

1.5.2. *Génération-Recombinaison. Durée de vie des porteurs*

On caractérise la création de porteurs dans le semiconducteur par un paramètre qui mesure le nombre de porteurs créés par unité de volume et unité de temps, g' ($cm^{-3}s^{-1}$). On caractérise la recombinaison des porteurs par un paramètre qui mesure le nombre de porteurs qui disparaissent par unité de volume et unité de temps, r' ($cm^{-3}s^{-1}$).

Le paramètre g' résulte de la contribution de deux types de génération de porteurs. Il existe d'une part les générations spontanées dues à l'agitation thermique, que l'on caractérisera par un paramètre g_{th} que nous appellerons *taux de génération thermique.* Il existe d'autre part les générations résultant de l'excitation du semiconducteur par une source extérieure. Ces générations peuvent résulter d'une

excitation optique, d'une irradiation par des particules, d'un champ électrique intense, d'une injection électrique....Ce type de génération sera caractérisé par un paramètre *g*, qui est spécifique du processus mis en jeu.

Le paramètre *r'* est fonction des processus qui dans le semiconducteur, régissent la recombinaison des porteurs excédentaires. C'est un paramètre qui est propre au matériau.

La variation du nombre de porteurs par unité de volume et unité de temps, due aux processus de génération-recombinaison, s'écrit

$$\left(\frac{dn}{dt}\right)_{gr} = g' - r' = g + g_{th} - r'$$

Parmi les trois paramètres *g*, g_{th} et *r'*, le premier est fonction de l'excitation éventuelle du matériau, les deux autres sont spécifiques du matériau à une température donnée. On groupe ces derniers dans un même terme qui représente leur différence en posant $r = r' - g_{th}$, r représente donc le bilan entre les recombinaisons et les générations thermiques.

On écrira donc la variation du nombre de porteurs par unité de volume et unité de temps, résultant des phénomènes de génération-recombinaison, sous la forme

$$\left(\frac{dn}{dt}\right)_{gr} = g - r$$

g et *r* sont respectivement appelés *taux de génération* et *taux de recombinaison* des porteurs. Le taux de génération est spécifique du processus mis en jeu dans l'excitation du semiconducteur, nous l'expliciterons en temps utile. Nous allons traiter ici plus particulièrement du taux de recombinaison, qui est un paramètre spécifique du matériau.

La recombinaison d'un électron avec un trou, dans le semiconducteur, peut se produire soit directement par la rencontre des deux particules, soit indirectement par l'intermédiaire d'une impureté qui joue en quelque sorte le rôle d'agent de liaison.

a. Recombinaison directe électron-trou

Le nombre de recombinaisons directes électron-trou est proportionnel d'une part au nombre d'électrons et d'autre part au nombre de trous. De plus, le nombre d'électrons qui se recombinent est égal, et pour cause, au nombre de trous qui se recombinent, soit

$$r_n' = r_p' = r' = k\, n\, p \qquad (1\text{-}208)$$

Le taux de recombinaison s'écrit donc

$$r_n = r_p = r = r' - g_{th} = k\, n\, p - g_{th}$$

A l'équilibre thermodynamique, c'est-à-dire en l'absence de toute excitation extérieure, $n=n_o$, $p=p_o$ et $r=0$. Ainsi $k\, n_o\, p_o - g_{th} = 0$ et par suite $g_{th} = k\, n_i^2$. Le taux de recombinaison s'écrit donc

$$r = k\,(np - n_i^2)$$

En régime hors équilibre $n = n_o + \Delta n,\ p = p_o + \Delta p$ avec, en raison de la condition de neutralité électrique, $\Delta n = \Delta p$. L'expression précédente s'écrit alors

$$\begin{aligned} r &= k\left((n_o+\Delta n)(p_o+\Delta p) - n_i^2\right) \\ &= k\left(n_o \Delta p + p_o \Delta n + \Delta n \Delta p\right) \\ &= k\left(n_o + p_o + \Delta n\right)\Delta n = k\left(n_o + p_o + \Delta p\right)\Delta p \end{aligned}$$

On écrit cette expression sous la forme

$$r = \frac{\Delta n}{\tau(\Delta n)} = \frac{\Delta p}{\tau(\Delta p)}$$

avec

$$\tau(\Delta n) = \frac{1}{k\,(n_o + p_o + \Delta n)} \qquad (1\text{-}209)$$

$\tau(\Delta n)$ est appelé *durée de vie des porteurs* dans le semiconducteur excité. Ce paramètre n'est pas spécifique du semiconducteur dans la mesure où il est fonction de l'excitation par le terme Δn. ***En fait, sauf cas spécifiques que nous discuterons en temps utile, les composants fonctionnent dans un régime que l'on qualifie de faible injection. Ce régime est défini par la condition*** $\Delta n = \Delta p << n_o$ ***ou*** p_0***. En d'autres termes un excès de porteurs par rapport au régime d'équilibre est faible devant la densité de porteurs majoritaires du semiconducteur.***

Considérons un matériau de type p en régime de faible injection, les conditions sont $p_o >> n_o$, $\Delta n = \Delta p << p_o$, $\tau \approx 1/kp_0$ et par suite

$$p = p_o + \Delta p \approx p_o \ , \quad n = n_o + \Delta n \ , \quad r \approx \Delta n/\tau$$

Considérons un matériau de type n en régime de faible injection, les conditions sont $n_o >> p_o$, $\Delta n = \Delta p << n_o$, $\tau \approx 1/kn_0$ et par suite

$$p = p_o + \Delta p \ , \ n = n_o + \Delta n \approx n_o \ , \ r \approx \Delta p/\tau$$

En résumé dans un matériau dopé et en régime de faible injection, on suppose constante la densité de porteurs majoritaires et on écrit le taux de recombinaison des porteurs minoritaires sous la forme

sc de type p $r_n = \Delta n / \tau_n$ (1-210-a)

sc de type n $r_p = \Delta p / \tau_p$ (1-210-b)

$\tau_n = 1/kp_o$ ($\tau_p = 1/kn_o$) est appelé *durée de vie des porteurs minoritaires*.

b. Recombinaison assistée par centres de recombinaison

Lorsque le semiconducteur est peu dopé les densités de porteurs libres sont faibles de sorte que la probabilité pour qu'un électron et un trou se rencontrent est faible. La durée de vie des porteurs devrait alors être considérable (3 heures dans le silicium à 300K). En fait il n'en est pas ainsi car la présence d'impuretés incontrôlées joue alors un rôle non négligeable dans le processus de recombinaison. Une impureté piège un électron (trou) qui par attraction coulombienne attire un trou (électron), ce qui provoque la recombinaison des deux particules.

En ce qui concerne le rôle joué dans ce domaine par les impuretés on distingue deux types de centre. Si le défaut qui a capturé un électron a une plus grande probabilité de capturer ainsi un trou que de réémettre cet électron vers la bande de conduction, il capture le trou et provoque de ce fait la recombinaison de la paire électron-trou. Ce centre porte le nom de *centre de recombinaison*. Si au contraire le défaut qui a capturé un électron a une plus grande probabilité de réémettre cet électron vers la bande de conduction que de capturer un trou, il a simplement piégé momentanément un électron, ce centre porte alors le nom de *piège à électron*.

Les pièges à électron, ou à trou, jouent un rôle important sur le courant dans le semiconducteur, dans la mesure où ils diminuent momentanément la densité de porteurs libres, mais ne jouent pas de rôle spécifique dans la recombinaison des porteurs. Nous considérerons ici les centres de recombinaison. Le calcul du taux de recombinaison associé à ces centres fait l'objet de la théorie de Shockley-Read.

Soit N_R la densité de centres de recombinaison, E_R le niveau d'énergie qu'ils introduisent dans le gap et f_R la probabilité d'occupation de ce niveau, c'est la fonction de Fermi. Soit C_n C_p E_n et E_p les coefficients de capture et d'émission des électrons et des trous par ces centres (Fig. 1-28).
Les densités de centres occupés N_{R0} et de centres vides N_{RV} sont respectivement données par $N_{RO} = N_R f_R$, $N_{RV} = N_R(1-f_R)$ avec

$$f_R = \frac{1}{1+e^{(E_R-E_F)/kT}} \qquad (1\text{-}211)$$

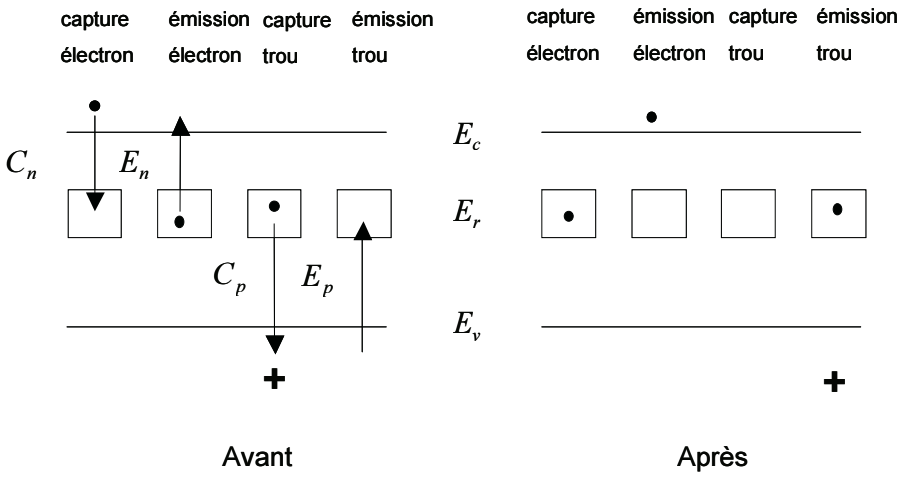


Figure 1-28 : Centres de recombinaisons

Les taux de recombinaison des électrons et des trous s'écrivent

$$r_n = C_n\, n\, N_{RV} - E_n\, N_{RO}$$
$$r_p = C_p\, p\, N_{RO} - E_p\, N_{RV}$$

ou

$$r_n = N_R\left(C_n n(1-f_R) - E_n f_R\right)$$
$$r_p = N_R\left(C_p p f_R - E_p(1-f_R)\right)$$

A l'équilibre thermodynamique $n=n_o$, $p=p_o$, $r_n=r_p=0$ et par suite

$$E_n = C_n n_o\left(1-f_R\right)/f_R$$

$$E_p = C_p p_o f_R/\left(1-f_R\right)$$

En explicitant f_R à partir de l'expression (1-211) et n_o et p_o à partir des expressions (1-156), les taux d'émission s'écrivent

$$E_n = C_n n_i e^{(E_R - E_{Fi})/kT}$$
$$E_p = C_p n_i e^{-(E_R - E_{Fi})/kT}$$

Les taux de recombinaison sont par conséquent donnés par les expressions

$$r_n = C_n N_R\left(n(1-f_R) - f_R n_i e^{(E_R - E_{Fi})/kT}\right)$$

$$r_p = C_p N_R \left(p f_R - (1 - f_R) n_i e^{-(E_R - E_{Fi})/kT} \right)$$

Le centre de recombinaison provoque la recombinaison d'une paire électron-trou de sorte que $r_n = r_p$. En écrivant cette égalité on obtient la probabilité d'occupation du centre

$$f_R = \frac{C_n n + C_p n_i e^{-(E_R - E_{Fi})/kT}}{C_n \left(n + n_i e^{(E_R - E_{Fi})/kT} \right) + C_p \left(p + n_i e^{-(E_R - E_{Fi})/kT} \right)}$$

En explicitant f_R on obtient $r = r_n = r_p$

$$r = C_n C_p N_R \frac{pn - n_i^2}{C_n (n + n_i e^{(E_R - E_{Fi})/kT}) + C_p (p + n_i e^{-(E_R - E_{Fi})/kT})}$$

En posant $\tau_{po} = 1 / C_p N_R \quad et \quad \tau_{no} = 1 / C_n N_R$, le taux de recombinaison s'écrit

$$r = \frac{pn - n_i^2}{\tau_{po} (n + n_i e^{(E_R - E_{Fi})/kT}) + \tau_{no} (p + n_i e^{-(E_R - E_{Fi})/kT})} \qquad (1\text{-}212)$$

Dans la pratique les centres qui jouent un rôle important dans les processus de recombinaison introduisent des niveaux d'énergie voisins du milieu du gap du semiconducteur et de plus ont des coefficients de capture tels que $C_n \approx C_p$. Les centres caractérisés par des coefficients de capture d'électrons et de trous très différents jouent davantage le rôle de piège à électron ($C_n >> C_p$) ou à trous ($C_p >> C_n$). On peut donc écrire l'expression précédente sous une forme simplifiée en posant

$$E_R = E_{Fi} \quad , \quad C_n = C_p = C \quad , \quad \tau_{po} = \tau_{no} = \tau_m = 1 / CN_R$$

soit

$$r = \frac{1}{\tau_m} \frac{pn - n_i^2}{2n_i + p + n} \qquad (1\text{-}213)$$

Considérons deux cas limites qui sont respectivement le cas d'un matériau dopé, c'est-à-dire dans lequel un type de porteur est nettement majoritaire, et le cas d'une zone dépeuplée dans un semiconducteur, comme par exemple la zone de charge d'espace d'une jonction pn.

- *Semiconducteur de type n*

On peut écrire $n_o >> n_i >> p_o$ et, sous faible excitation, $n = n_o + \Delta n \approx n_o$ et $p = p_o + \Delta p << n_o$. L'expression (1-213) se simplifie et s'écrit alors

$$r=\frac{1}{\tau_m}\frac{pn-n_i^2}{n}=\frac{p-n_i^2/n}{\tau_m}=\frac{p-p_o}{\tau_m}=\frac{\Delta p}{\tau_m}$$

On obtient alors une expression du taux de recombinaison des porteurs minoritaires semblable à l'expression **(1-210-b),** associée aux recombinaisons directes électron-trou. En raison même de leur définition les taux de recombinaison sont additifs de sorte que le taux de recombinaison résultant de la présence des deux processus s'écrit

$$r=\frac{\Delta p}{\tau} \qquad (1\text{-}214)$$

$$avec \quad \frac{1}{\tau}=\frac{1}{\tau_d}+\frac{1}{\tau_m} \qquad \text{où} \quad \tau_d=1/kn_0 \quad et \quad \tau_m=1/CN_R$$

Nous avons considéré le cas d'un seul centre de recombinaison. Dans la pratique il existe effectivement dans chaque type de semiconducteur un centre prédominant, mais si plusieurs centres de recombinaison existent avec des densités et des coefficients de capture comparables, la durée de vie des porteurs minoritaires s'écrit

$$\frac{1}{\tau}=\frac{1}{\tau_d}+\sum_i \frac{1}{\tau_{mi}} \quad avec \quad \tau_{mi}=1/C_i\,N_{Ri}$$

- *Zone dépeuplée*

Dans ce cas les recombinaisons directes sont pratiquement inexistantes. D'autre part dans l'expression (1-213) on peut négliger n et p devant n_i et np devant n_i^2. Le taux de recombinaison des porteurs dans une zone dépeuplée s'écrit alors

$$r=-\frac{n_i}{2\,\tau_m} \qquad (1\text{-}215)$$

Le taux de recombinaison est négatif, tout simplement parce que dans la mesure où les densités de porteurs sont inférieures aux densités d'équilibre, définies par $np=n_i^2$, le nombre de porteurs créés thermiquement est plus important que le nombre de porteurs qui se recombinent. Un taux de recombinaison négatif correspond à une génération thermique de porteurs. L'expression (1-213) montre que r est positif si $np>n_i^2$ et négatif dans le cas contraire.

En résumé, pour tenir compte des phénomènes de génération-recombinaison nous utiliserons l'expression (1-213) pour un semiconducteur peu dopé,

l'expression (1-214) pour un semiconducteur dopé et l'expression (1-215) pour une région dépeuplée.

c. Retour à l'équilibre d'un matériau excité

Considérons un semiconducteur isolé, excité de manière homogène, par un éclairement par exemple (Fig. 1-29). Les variations des densités de porteurs par unité de temps sont simplement données par la différence entre les taux de génération et les taux de recombinaison

$$\frac{dn}{dt} = g_n - r_n \qquad \frac{dp}{dt} = g_p - r_p$$

Considérons le cas d'un semiconducteur de type p en régime de faible excitation, $p = p_o + \Delta p \approx p_o$, $n = n_0 + \Delta n$. Le taux de recombinaison des électrons s'écrit sous la forme (1-214). L'évolution de la densité d'électrons est régie par l'équation

$$\frac{dn}{dt} = g_n - \frac{n - n_o}{\tau_n} \quad \text{ou} \quad \frac{d\Delta n}{dt} = g_n - \frac{\Delta n}{\tau_n} \qquad (1\text{-}216)$$

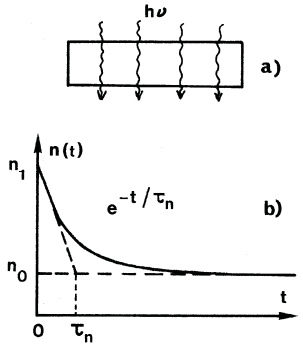


Figure 1-29 : Éclairement d'un semiconducteur

En régime stationnaire, c'est-à-dire sous éclairement permanent, $d\Delta n/dt = 0$, $\Delta n = \Delta n_0 = \text{Cte}$, l'équation précédente donne

$$\Delta n_0 = g_n \, \tau_n \qquad (1\text{-}217)$$

L'excédent d'électrons par rapport à l'équilibre est donné par le produit du taux de génération par la durée de vie des porteurs.

En régime transitoire, résultant de la suppression de l'excitation, $g_n = 0$ et l'équation (1-216) s'écrit

$$\frac{d\Delta n}{dt} = -\frac{\Delta n}{\tau_n} \qquad (1\text{-}218)$$

En intégrant avec la condition initiale $\Delta n = \Delta n_0$ à $t = 0$, la solution s'écrit

$$\Delta n = \Delta n_0 \, e^{-t/\tau_n} \qquad (1\text{-}219)$$

L'excédent d'électrons décroît exponentiellement avec une constante de temps τ_n (Fig. 1-29).

1.5.3. *Équations d'évolution*

a. Équations de continuité

Les équations de continuité régissent la condition d'équilibre dynamique des porteurs dans le semiconducteur. Dans un barreau semiconducteur excité et parcouru par un courant dans la direction ox, considérons un élément de volume de section unitaire s=1 dans le plan perpendiculaire à ox, et d'épaisseur dx (Fig.1-30).

La variation du nombre de porteurs par unité de temps dans cet élément de volume est la somme algébrique des nombres de porteurs qui entrent, j_x/q, qui sortent, j_{x+dx}/q, qui se créent, $gsdx$=gdx et qui se recombinent, rdx. En ramenant à l'unité de volume, l'expression s'écrit pour les électrons

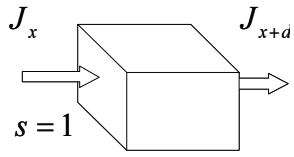


Figure 1-30 : Équation de continuité

$$\frac{\partial n}{\partial t}=\frac{1}{dx}\left[\left(\frac{1}{e}j_x\right)-\left(\frac{1}{e}\left(j_x+\frac{dj_x}{dx}dx\right)\right)\right]+g_n-r_n$$

Soit, pour les électrons et les trous

$$\frac{\partial n}{\partial t}=+\frac{1}{e}\frac{\partial j_{nx}}{\partial x}+g_n-r_n \qquad \text{(1-220-a)}$$

$$\frac{\partial p}{\partial t}=-\frac{1}{e}\frac{\partial j_{px}}{\partial x}+g_p-r_p \qquad \text{(1-220-b)}$$

ou, à trois dimensions

$$\frac{\partial n}{\partial t}=+\frac{1}{e}div\mathbf{J}_n+g_n-r_n \qquad \text{(1-221-a)}$$

$$\frac{\partial p}{\partial t}=-\frac{1}{e}div\mathbf{J}_p+g_p-r_p \qquad \text{(1-221-b)}$$

Dans un matériau dopé en régime de faible injection, on pourra écrire les taux de recombinaison sous la forme (1-214), si d'autre part, on explicite les courants à partir des expressions (1-191), les équations s'écrivent, à une dimension

$$\frac{\partial n}{\partial t}=+n\mu_n\frac{\partial E}{\partial x}+\mu_n E\frac{\partial n}{\partial x}+D_n\frac{\partial^2 n}{\partial x^2}+g_n-\frac{n-n_o}{\tau_n} \qquad \text{(1-222-a)}$$

$$\frac{\partial p}{\partial t}=-p\mu_p\frac{\partial E}{\partial x}-\mu_p E\frac{\partial p}{\partial x}+D_p\frac{\partial^2 p}{\partial x^2}+g_p-\frac{p-p_o}{\tau_p} \qquad \text{(1-222-b)}$$

b. Longueur de diffusion

Considérons un barreau semi-conducteur de type p, excité en surface par un rayonnement peu pénétrant qui crée à la surface un excès Δn_1 de paires électron-trou (Fig. 1-31). Les électrons diffusent à l'intérieur du barreau et créent un courant de diffusion donné par

$$j_{nx} = e\, D_n \frac{\partial n}{\partial x}$$

L'évolution de la densité d'électrons en un point quelconque du barreau est donnée par l'équation (1-220-a) qui s'écrit en explicitant j_{nx}

$$\frac{\partial n}{\partial t} = D_n \frac{\partial^2 n}{\partial x^2} + g - \frac{n - n_o}{\tau_n} \qquad (1\text{-}223)$$

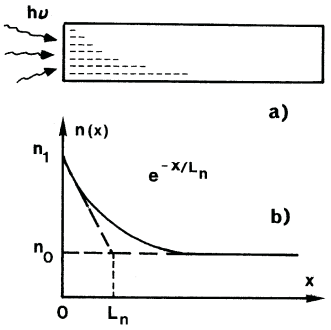


Figure 1-31 : Éclairement d'un semiconducteur

Dans la mesure où le rayonnement est peu pénétrant g_n=0 dans le volume du semiconducteur. Si en outre l'éclairement est permanent, le régime est stationnaire et $\partial n/\partial t = 0$. L'équation (1-223) s'écrit

$$\frac{d^2(\Delta n)}{dx^2} - \frac{(\Delta n)}{L_n^2} = 0 \qquad (1\text{-}224)$$

avec $L_n^2 = D_n\, \tau_n$

L'intégration de l'équation (1-244) avec les conditions aux limites $\Delta n(x=0)=\Delta n_0$ et $\Delta n(x \to \infty)=0$, donne

$$\Delta n = \Delta n_0\, e^{-x/L_n} \qquad (1\text{-}225)$$

La densité de porteurs en excès décroît exponentiellement avec une constante de temps L_n (Fig. 1-31-b). L_n est appelé *longueur de diffusion* des électrons dans le matériau de type p. On définit la même grandeur pour les trous dans un matériau de type n

$$L_n = \sqrt{D_n\, \tau_n} \qquad L_p = \sqrt{D_p\, \tau_p}$$

c. Charge d'espace - Équation de Poisson

Toute charge d'espace $\rho(x,y,z)$ est accompagnée d'un champ électrique donné par la loi de Poisson

$$div\,\mathbf{E}=\frac{\rho(x,y,z)}{\varepsilon}$$

où ε est la constante diélectrique du semiconducteur. Le champ électrique est d'autre part relié au potentiel par la relation $\mathbf{E}=-\mathbf{grad}V$ ce qui donne en explicitant le champ $\mathbf{E}$

$$\Delta V=-\frac{\rho(x,y,z)}{\varepsilon} \qquad (1\text{-}226)$$

C'est l'équation de Laplace, dont l'intégration permet de calculer la variation du potentiel dans un semiconducteur à partir de la charge d'espace.

La charge d'espace est calculée en tenant compte de toutes les charges qui existent en un point du semiconducteur, c'est-à-dire d'une part des charges mobiles que sont les électrons et les trous, et d'autre part des charges fixes qui peuvent être localisées sur des donneurs ou accepteurs ionisés ou sur des centres profonds. En l'absence de centres profonds ionisés cette charge d'espace est donnée par

$$\rho = e\left(N_d^+ - N_a^- + p - n\right) \qquad (1\text{-}227\text{-a})$$

À la température ambiante tous les donneurs et accepteurs sont ionisés de sorte que la charge d'espace s'écrit simplement

$$\rho = e\left(N_d - N_a + p - n\right) \qquad (1\text{-}227\text{-b})$$

1.5.4. *Notion de neutralité électrique*

L'état électrique d'un semiconducteur et au-delà d'un composant, est défini par la trilogie *charge, champ, potentiel*, ces trois quantités étant reliées entre elles par l'équation de Laplace (1-226) et la relation $\mathbf{E}=-\mathbf{grad}V$.

La connaissance de cet état électrique passe tout simplement par l'intégration de l'équation de Laplace. Mais la difficulté réside dans le fait que cette intégration présente rarement des solutions analytiques et nécessite la plupart du temps l'utilisation d'hypothèses simplificatrices. Les expressions (1-227) montrent en effet que de manière générale la charge d'espace est constituée de charges fixes, les ions donneurs et accepteurs, et de charges libres, les électrons et les trous. Or si la

distribution des premières est prédéterminée par le profil de dopage, celle des secondes est fonction de la distribution du potentiel, c'est-à-dire de la solution du problème. Il en résulte que la résolution de l'équation nécessite généralement un calcul de type auto-cohérent. Deux cas limites présentent néanmoins des solutions analytiques simples et constituent bien souvent des hypothèses simplificatrices permettant de résoudre des problèmes réels. Le premier consiste à supposer la charge d'espace vide de porteurs libres c'est-à-dire constituée uniquement des charges fixes que sont les ions donneurs et accepteurs. La distribution spatiale de la charge d'espace est alors indépendante du potentiel et l'équation de Poisson a des solutions analytiques généralement simples. Le second cas consiste à supposer que la distribution des porteurs libres n-p est exactement calquée sur celle des donneurs et accepteurs N_d-N_a, c'est-à- dire en d'autres termes que la charge d'espace est nulle, c'est l'hypothèse de *neutralité électrique*. Dans ce dernier cas la solution de l'équation de Poisson est évidente, le champ électrique est constant ou nul, le potentiel est linéaire ou constant. Ces simplifications sont souvent utilisées dans l'étude des composants, la première dans les zones de déplétion, la seconde dans les régions conductrices. Mais ces approximations, si elles sont très utiles, ne présentent un réel intérêt que si elles sont justifiées, c'est la question que nous allons évoquer ici.

En fait une charge d'espace n'existe, dans un matériau conducteur ou semiconducteur, que dans certaines conditions très particulières associées soit à des situations transitoires, comme l'injection instantanée de porteurs par exemple, soit à des inhomogénéités très spécifiques comme la jonction brutale entre deux matériaux différents, jonction pn abrupte par exemple. Ceci résulte simplement du fait que si dans une région conductrice une charge d'espace existe, le potentiel électrostatique qui en résulte crée au niveau de chaque porteur libre une énergie potentielle considérablement plus grande que l'énergie thermique moyenne. Les porteurs libres de signe opposé à la charge d'espace sont rapidement attirés par celle-ci afin de restaurer la neutralité électrique.

Considérons à titre d'exemple un semiconducteur dans lequel un excès de donneurs existe dans une région donnée. Au zéro degré absolu chaque atome donneur est neutre, le cinquième électron périphérique est lié au donneur et gravite autour de celui-ci sur une orbitale dont le rayon moyen est typiquement de l'ordre de quelques dizaines d'Angströms, le semiconducteur est neutre à l'échelle de l'atome, c'est-à-dire à l'échelle nanométrique.

Considérons maintenant ce même semiconducteur à la température ambiante. L'énergie thermique est alors supérieure à l'énergie de liaison de l'électron, ce dernier est libéré dans le cristal donnant naissance à une charge négative délocalisée et un ion positif fixe, la neutralité électrique n'existe plus à l'échelle nanométrique. Qu'en est-il à l'échelle micrométrique ? Il existe localement une distribution d'ions donneurs fixes et une population d'électrons libres dont la distribution spatiale résulte du bilan entre les forces électrostatiques qui retiennent ces électrons au voisinage des ions, et les forces de diffusion qui diluent ces électrons dans les régions de moindre concentration. L'état de charge du matériau est alors conditionné par le gradient de la

distribution de donneurs. Si ce gradient est nul (matériau homogène) les forces de diffusion sont nulles et les électrons bien que libérés dans la bande de conduction restent au voisinage des ions donneurs, le matériau est neutre et les populations de porteurs sont données par $n\text{-}p = N_d\text{-}N_a$. Si le gradient est non nul mais faible, il en est de même des forces de diffusion et par suite de la séparation spatiale des charges positives et négatives, le matériau est quasiment neutre à l'échelle du micron. Si par contre le gradient de concentration est important, la séparation spatiale des charges peut atteindre plusieurs microns, une charge d'espace existe alors à l'échelle micrométrique. En d'autres termes cela signifie que si le μm^3 est pris comme unité de volume, un volume unitaire du semiconducteur n'a pas forcément le même état de charge que son voisin.

Nous allons essayer d'approcher de manière un peu plus quantitative cette frontière assez floue entre les régimes de *charge d'espace* et de *neutralité électrique.* Nous examinerons successivement le problème dans l'espace et dans le temps.

a. Neutralité électrique dans l'espace - Longueur de Debye

Notons tout d'abord qu'une charge d'espace ne peut exister que dans un matériau inhomogène. En effet, un cristal est constitué d'atomes dont chacun, donneur et accepteur compris, est électriquement neutre, il est par conséquent globalement neutre. Si le matériau est homogène la charge d'espace est donc à la fois uniforme et globalement nulle, elle est donc nulle partout.

Si le semiconducteur n'est pas homogène, le cristal reste globablement neutre mais il peut exister des charges positives dans une région et négatives dans une autre. Pour étudier le comportement de ce semiconducteur non homogène, il faut résoudre l'équation de Laplace qui s'écrit, en supposant tous les donneurs et accepteurs ionisés

$$\Delta V = -\frac{e}{\varepsilon}(N_d - N_a + p - n)$$

À l'équilibre thermodynamique le niveau de Fermi étant horizontal, on peut exprimer la variation du potentiel et les densités de porteurs libres en fonction de la distance du niveau de Fermi au niveau de Fermi intrinsèque $e\phi_{Fi} = E_F - E_{Fi}$. Les densités de porteurs *n* et *p* sont données par les expressions (1-156).

$$n = n_i e^{+e\phi_{Fi}/kT}$$
$$p = n_i e^{-e\phi_{Fi}/kT}$$

Les considérations du paragraphe 1-5-1 ont montré que les niveaux E_c et E_v varient comme le potentiel électrique *V*. Rappelons que le potentiel considéré ici est un potentiel moyen qui ne tient pas compte du potentiel local créé par la structure du cristal variant à l'échelle de la maille élémentaire. Le potentiel de Fermi peut être remplacé par le pseudo-potentiel de Fermi quand électrons et trous sont en équilibre séparément. Le niveau de Fermi intrinsèque ϕ_{Fi} est approximativement au milieu de la bande interdite. Le potentiel ϕ_{Fi} varie donc simplement comme *V* puisque ϕ_{Fi} est défini à partir de E_c et E_v. Ce potentiel est dû aux charges des zones chargées du

semiconducteur et au champ éventuellement appliqué par un potentiel extérieur. L'équation de Laplace s'écrit alors

$$\Delta\phi_{Fi} = -\frac{e}{\varepsilon}\left(N_d - N_a - 2n_i sh\left(e\phi_{Fi}/kT\right)\right)$$

ou, en posant $u = e\phi_{Fi}/kT$

$$\Delta u = \frac{1}{L_{Di}^2}\left(sh(u) - \frac{N_d - N_a}{2n_i}\right) \qquad (1\text{-}228)$$

$L_{Di} = (\varepsilon kT/2e^2 n_i)^{1/2}$ est la *longueur de Debye* du matériau intrinsèque. Pour le silicium, $n_i \approx 10^{10}\ cm^{-3}$ et $\varepsilon \approx 12\,\varepsilon_o$, à la température ambiante $L_{Di} \approx 30\ \mu m$.

- *Distribution de porteurs associée à un profil de dopage donné*

À partir d'un profil de dopage (N_d-N_a) donné, l'équation (1-228) permet de calculer u, donc le profil de potentiel ϕ_{Fi}, et par suite la distribution des porteurs libres n et p. On peut alors en déduire le profil de la charge d'espace ρ. En fait l'équation (1-228) ne présente généralement pas de solution analytique et des hypothèses simplificatrices doivent être utilisées pour résoudre la plupart des cas réels. Le traitement numérique de cette équation permet de préciser de manière quantitative les conditions d'utilisation de ces hypothèses. Dans la plupart des composants le profil de dopage varie suivant une seule direction. Supposons cette variation linéaire, ce profil peut alors s'écrire sous la forme

$$N_d - N_a = 2n_i \frac{x}{L}$$

où L est une longueur caractéristique du gradient de dopage, qui en fait correspond ici à la longueur sur laquelle le dopage varie de deux fois la densité de porteurs intrinsèques. L'équation de Laplace s'écrit alors

$$\frac{d^2u}{dx^2} = \frac{1}{L_{Di}^2}\left(sh(u) - \frac{x}{L}\right)$$

L'intégration numérique de cette équation montre que l'état électrique du barreau est fonction du gradient de dopage à travers le rapport L/L_{Di}. Lorsque ce rapport est faible ($L/L_{Di} < 10^{-3}$), il existe une zone d'épaisseur comparable à L_{Di} dans laquelle la charge d'espace est pratiquement égale à N_d-N_a, la densité de porteurs libres y est négligeable. Par contre lorsque L/L_{Di} est supérieur à 1 la densité de charge reste partout très inférieure à N_d-N_a, (sauf en x=0 où N_d-N_a, n-p et ρ sont tous nuls). Dans le premier cas qui correspond à un gradient de dopage important, il existe une zone de déplétion vide de porteurs dans laquelle la charge d'espace est sensiblement

donnée par $\rho = N_d\text{-}N_a$. Dans le second cas qui correspond à un gradient de dopage faible, la charge d'espace est négligeable, le semiconducteur est en régime de neutralité électrique avec $n\text{-}p = N_d\text{-}N_a$.

- *Profil de dopage associé à une distribution de porteurs donnée*

L'équation (1-228) qu'il faut résoudre pour déterminer les distributions de porteurs libres résultant d'un profil de dopage donné, n'a généralement pas de solution analytique. Par contre le problème inverse, qui consiste à déterminer le profil de dopage à partir d'une distribution donnée de porteurs libres, a quant à lui, des solutions analytiques. Considérons un barreau semiconducteur dans lequel existe une distribution inhomogène d'électrons et de trous, quel profil de dopage $N_d\text{-}N_a$ est à l'origine de cette distribution de porteurs libres ?

À l'équilibre thermodynamique les courants d'électrons et de trous sont nuls. Considérons par exemple le courant d'électrons

$$j_n = ne\,\mu_n E + eD_n \frac{dn}{dx} = 0$$

Compte tenu de la relation d'Einstein $D/\mu = kT/e$, le champ électrique existant dans le barreau s'écrit donc

$$E = -\frac{kT}{e}\frac{1}{n}\frac{dn}{dx} \equiv \frac{kT}{e}\frac{1}{p}\frac{dp}{dx} \qquad (1\text{-}229)$$

Ce champ électrique résulte des forces de diffusion qui ont tendance à déplacer les porteurs libres dans le sens des gradients de population négatifs. Ce déplacement des porteurs par rapport aux ions dont ils sont originaires crée autant de dipoles électriques qui entraînent l'existence du champ E qui retient ces porteurs au voisinage des ions. Ce champ électrique a donc tendance à restaurer la neutralité électrique.

La charge d'espace est donnée par la loi de Poisson

$$\rho = \varepsilon\frac{dE}{dx} = -\frac{\varepsilon kT}{e}\frac{d}{dx}\left(\frac{1}{n}\frac{dn}{dx}\right) = -\frac{\varepsilon kT}{e}\frac{d^2}{dx^2}\left(Ln(n)\right)$$

L'équation (1-227-b) permet alors d'obtenir le profil de dopage

$$N_d - N_a = n - p + \frac{\rho}{e} = n - \frac{n_i^2}{n} - \frac{\varepsilon kT}{e^2}\frac{d}{dx}\left(\frac{1}{n}\frac{dn}{dx}\right)$$
$$= n_i\left(\frac{n}{n_i} - \frac{n_i}{n} - 2L_{Di}^2\frac{d}{dx}\left(\frac{1}{n}\frac{dn}{dx}\right)\right)$$

où L_{Di} est la *longueur de Debye* du matériau intrinsèque.

Considérons tout d'abord le cas très spécifique d'un semiconducteur dans lequel existe une distribution de porteurs exponentielle de la forme

$$n = n_o e^{-ax}$$

Le champ électrique, donné par l'expression (1-229) s'écrit

$$E = \frac{kT}{e} a$$

Ce champ est constant, de sorte que dE/dx=0 et par suite ρ=0. La charge d'espace est donc nulle malgré la distribution inhomogène de porteurs. Le profil de dopage est par conséquent donné par

$$N_d - N_a = n - p = n_o e^{-ax} - \frac{n_i^2}{n_o} e^{ax}$$

La distribution spatiale des électrons et des trous n-p est exactement donnée par le profil de dopage N_d-N_a. En fait cette égalité parfaite est spécifique de la loi de distribution exponentielle. Pour tout autre type de profil de dopage cette identité rigoureuse entre les distributions d'atomes dopants et de porteurs libres n'existe pas.

Considérons par exemple un semiconducteur de type n avec une distribution d'électrons linéaire de la forme

$$n = n_o (1 + \frac{x}{L}) \qquad avec \quad n_o >> n_i$$

Le champ électrique donné par l'expression (1-229) s'écrit

$$E = -\frac{kT}{e} \frac{1/L}{1 + x/L}$$

Ce champ n'est pas constant, sa variation résulte de la présence d'une densité de charge donnée par la loi de Poisson.

$$\rho = \varepsilon \frac{dE}{dx} = \frac{\varepsilon kT}{e} \left(\frac{1/L}{1 + x/L} \right)^2$$

Il existe donc une densité de charge non nulle et le profil de dopage s'écrit, en négligeant la densité de trous

$$N_d - N_a = n + \frac{\rho}{e} = n_o(1 + \frac{x}{L}) + \frac{\varepsilon kT}{e^2}\left(\frac{1/L}{1 + x/L}\right)^2$$

Le semiconducteur n'est donc pas en régime de neutralité électrique. Il pourra cependant être supposé tel si dans le membre de droite de l'expression précédente le deuxième terme est partout négligeable devant le premier. Cette condition s'écrit

$$\frac{\varepsilon kT}{e^2}\frac{1}{L^2} << n_o \qquad \text{ou} \qquad \frac{\varepsilon kT}{e^2 n_o} << L^2$$

c'est-à-dire $L / L_D >> 1$, où $L_D = (\varepsilon kT / 2e^2 n_o)^{1/2}$ est la longueur de Debye du matériau en x=0. Le profil de dopage est alors donné par

$$N_d - N_a \approx n = n_o(1 + x / L)$$

Le semiconducteur pourra donc être considéré en régime de neutralité électrique si la condition $L/L_D>>1$ est réalisée. Il faut noter que plus le matériau est conducteur, plus n_o est grand, plus L_D est petit, et par suite plus le gradient doit être important pour qu'une charge d'espace existe. En d'autres termes la condition de neutralité électrique est d'autant mieux justifiée que le matériau est conducteur. Pour n_o de l'ordre de $10^{18}\,cm^{-3}$ par exemple, L_D est de l'ordre de 30 Å, et la valeur frontière $L=L_D$ correspond à un gradient de dopage de l'ordre de $3.10^{16}\,cm^{-3}$/Å, ce qui est considérable.

Il en résulte donc que, sauf dans des régions très spécifiques comme par exemple la jonction abrupte entre deux semiconducteurs de types différents, les régions conductrices d'un semiconducteur pourront être supposées en régime de neutralité électrique.

b. Neutralité électrique dans le temps - Temps de relaxation diélectrique

Nous avons étudié, (Par.1.5.2.c), comment évoluait dans le semiconducteur, un excès de porteurs minoritaires. Cette évolution est conditionnée par un paramètre qui est la durée de vie de ces porteurs. Nous allons étudier maintenant comment évolue un excès de porteurs majoritaires.

Considérons un semiconducteur homogène de type n, avec $n_o>>n_i>>p_o$ dans lequel on crée un excès d'électrons $\Delta n<<n_o$. En l'absence d'excitation le semiconducteur est neutre de sorte que sous excitation la seule charge d'espace est

$$\rho = -e\Delta n$$

Les porteurs créés étant des porteurs majoritaires, ils ne peuvent se recombiner pour la simple raison qu'ils n'ont pas suffisamment de trous à leur disposition.

Prenons un exemple avec du silicium, si $n_o = 10^{16} cm^{-3}$ et $\Delta n = 10^{10} cm^{-3}$ avec $p_o = n_i^2 / n_o = 10^4 cm^{-3}$ on constate que $\Delta n >> p_o$.

Lorsque la source d'excitation est supprimée, la relaxation des charges est obtenue à partir de l'équation de continuité (1-221-a) qui s'écrit, dans la mesure où $g_n = 0$ et $\tau_n = \infty$

$$\frac{\partial n}{\partial t} = \frac{1}{e} div \mathbf{J}_n \quad ou \quad \frac{\partial \rho}{\partial t} = -div \mathbf{J}_n \tag{1-230}$$

La densité de courant est donnée par l'expression (1-200-a) qui s'écrit compte tenu du fait que $n = n_0 + \Delta n \approx n_0$

$$\mathbf{J}_n = \sigma \cdot \mathbf{E} + eD_n \mathbf{grad} n$$

où $\sigma = n_o \, e \, \mu_n$ est la conductivité du semiconducteur.

Le champ électrique $\mathbf{E}$ qui est dû à la charge d'espace ρ, est lié à cette dernière par la loi de Poisson, de sorte que $div \, \mathbf{J}_n$ s'écrit

$$\begin{aligned} div \mathbf{J}_n &= \sigma \, div \mathbf{E} + eD_n \Delta n \\ &= \frac{\sigma}{\varepsilon} \rho - D_n \Delta \rho \end{aligned}$$

Compte tenu de l'expression (1-230), l'équation s'écrit

$$\frac{\partial \rho}{\partial t} = -\frac{\sigma}{\varepsilon} \rho + D_n \Delta \rho \tag{1-231}$$

Considérons le cas simplifié où la charge d'espace est répartie uniformément dans le semiconducteur, $\Delta \rho = 0$. L'équation (1-231) s'écrit

$$\frac{d\rho}{\rho} = -\frac{dt}{\tau} \text{ avec } \tau = \frac{\varepsilon}{\sigma}$$

la solution est simplement

$$\rho = \rho_o e^{-t/\tau} \tag{1-232}$$

Si la distribution des électrons n'est pas uniforme, la solution de l'équation (1-231) est plus complexe, mais τ mesure toujours le temps de mise en équilibre du semiconducteur.

La charge d'espace dans le semiconducteur s'annule exponentiellement avec la constante de temps $\tau = \varepsilon / \sigma$. τ est appelé *temps de relaxation diélectrique*. Les

électrons créés en excès se répartissent à la surface du semiconducteur pour maintenir la neutralité électrique à l'intérieur du matériau, comme dans un métal. Considérons par exemple du silicium de type n avec une résistivité $\rho = 1\ \Omega cm$, la constante diélectrique relative du silicium est $\varepsilon_r = 12$, on obtient $\tau = 10^{-12} s$. La neutralité électrique est donc restaurée en une picoseconde, c'est la raison pour laquelle, dans l'étude des composants, on supposera que la condition de neutralité électrique est toujours réalisée dans les régions conductrices.

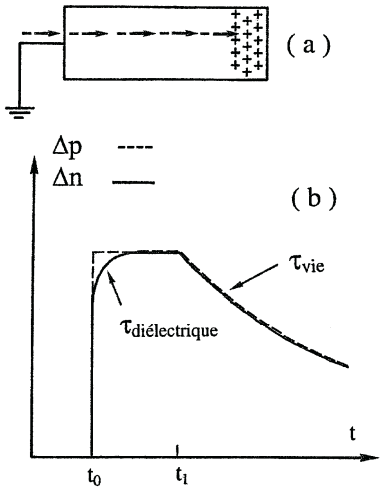


Figure 1-32 : Relaxation des charges

Pour résumer ce qui précède et schématiser l'évolution des charges mobiles dans le semiconducteur, considérons un barreau semiconducteur de type *n* dans lequel un certain nombre de trous sont injectés à une extrémité, l'autre extrémité du barreau étant reliée à la masse (Fig. 1-32-a). L'injection de trous commence à $t=t_o$ et est arrêtée à $t=t_1$. Considérons successivement les évolutions des populations de trous et d'électrons dans le barreau.

Pour $t_o<t<t_1$, le nombre de trous dans le barreau est établi à une valeur d'équilibre régie par l'égalité des taux de génération et de recombinaison des trous. Pour $t>t_1$, le taux de génération est nul, les trous excédentaires se recombinent en un temps conditionné par leur durée de vie τ_p. L'excès de trous dans le barreau est représenté par la courbe en trait discontinu sur la figure (1-32-b).

Considérons maintenant la population d'électrons. Lorsque les trous sont générés à l'extrémité du barreau, un champ électrique est créé dans le barreau par ces trous. Sous l'action de ce champ, les électrons se déplacent vers la région où sont créés les trous, un nombre d'électrons égal à l'excès de trous est fourni par la masse, la neutralité électrique s'établit et le champ s'annule. L'augmentation du nombre d'électrons dans le barreau est régie par la constante de temps diélectrique. Pour $t>t_1$, les électrons et les trous excédentaires se recombinent, et ceci à la même vitesse. L'évolution de la population d'électrons dans le barreau est représentée en traits pleins sur la figure (1-32-b).

Les constantes de temps d'établissement de la neutralité électrique et du retour à l'équilibre sont respectivement la constante de temps diélectrique et la durée de vie des porteurs. La première est de l 'ordre de 10^{-12} s, la seconde est supérieure à 10^{-9} s. Ainsi tant que la fréquence de fonctionnement des composants est inférieure

à $1/\tau_{diel}$, c'est-à-dire à 10^{12} Hz on peut considérer que la relaxation des charges est instantanée, c'est l'hypothèse dans laquelle nous nous placerons dans tout ce qui suit.

Notons pour terminer, que si dans l'expérience précédente le barreau n'est pas relié à la masse, le phénomène de relaxation se traduit par un déplacement des électrons vers les trous. Ces électrons laissent en surface des charges positives qui créent un champ nul à l'intérieur du matériau. Le matériau est alors chargé.

1.5.5. *Résumé des relations les plus utiles*

Il n'est pas inutile à ce stade de résumer les principales relations nécessaires au calcul des composants à semiconducteurs. Les conditions d'application seront données pour chaque relation.

Les densités de porteurs

$n=n_i e^{+e\phi_{Fi}/kT}$ vrai uniquement en régime d'équilibre

$p=n_i e^{-e\phi_{Fi}/kT}$ vrai uniquement en régime d'équilibre

avec $e\phi_{Fi}=E_F-E_{Fi}$

Il faut noter que le niveau de Fermi E_F est supposé constant spatialement dans une région en équilibre mais le niveau E_{Fi} varie avec le potentiel électrique donc avec la position.

$n=n_i e^{+e\phi_{Fi}/kT}$

encore vrai hors équilibre avec la définition suivante

$p=n_i e^{-e\phi_{Fi}/kT}$

$e\phi_{Fi}=E_{Fn}-E_{Fi}$ dans laquelle E_{Fn} est le pseudo-niveau de Fermi des électrons

$e\phi_{Fi}=E_{Fp}-E_{Fi}$ dans laquelle E_{Fp} est le pseudo-niveau de Fermi des trous

Les pseudo-niveaux de Fermi ne sont plus constants avec la position mais varient en fonction du courant circulant dans le dispositif selon les relations

$\mathbf{J_n} = \mu_n\, n\, \mathbf{grad} E_{Fn}$ vrai dans un semiconducteur

$\mathbf{J_p} = \mu_p\, p\, \mathbf{grad} E_{Fp}$ vrai dans un semiconducteur

Ils sont reliés aux différences de potentiel appliquées par

$E_{Fn1}-E_{Fn2} = -e(V_1 - V_2)$ toujours vrai

$E_{Fp1} - E_{Fp2} = -e\left(V_1 - V_2\right)$ toujours vrai

Les équations dynamiques

$\frac{\partial n}{\partial t} = +\frac{1}{e} div\mathbf{J}_n + g_n - r_n$ toujours vrai

$r_n = \frac{n - n_0}{\tau_n}$ vrai uniquement pour les porteurs minoritaires

$\frac{\partial p}{\partial t} = -\frac{1}{e} div\mathbf{J}_p + g_p - r_p$ toujours vrai

$r_p = \frac{p - p_0}{\tau_p}$ vrai uniquement pour les porteurs minoritaires

Les courants continus

$\mathbf{J}_n = ne\mu_n \mathbf{E} + eD_n \mathbf{grad} n$ toujours vrai dans un semiconducteur avec correction pour les champs élevés (régime de saturation)

$\mathbf{J}_p = pe\mu_p \mathbf{E} - eD_p \mathbf{grad} p$ toujours vrai dans un semiconducteur avec correction pour les champs élevés (régime de saturation)

$\frac{D_n}{\mu_n} = \frac{D_p}{\mu_p} = \frac{kT}{e}$ toujours vrai dans le cas non dégénéré

Dans certains cas (jonction pn et bipolaire), on considère que les deux composantes (diffusion et dérive) sont peu différentes et négligeables devant le courant total.

Le potentiel électrique

$\Delta V + \frac{1}{\varepsilon} e\left(N_d^+ - N_a^- + p - n\right) = 0$ toujours vrai

Remarques

La neutralité électrique n'est pas toujours assurée (cas des zones de charge d'espace). Le théorème de Gauss est souvent utile pour écrire des relations globales entre charges.

Ces relations relativement peu nombreuses seront les plus utilisées par la suite et permettront de modéliser la plupart des composants. Les effets quantiques significatifs pour les très petites dimensions seront pris en compte par d'autres relations (chapitres 10 et 12).

1.6. Interface entre deux matériaux différents

Les composants électroniques utilisent de plus en plus les propriétés d'hétérostructures réalisées par la juxtaposition de matériaux différents. Nous allons définir ici les grandeurs qui caractérisent les interfaces et conditionnent les transferts de charges entre les matériaux.

1.6.1. *Travail de sortie - Affinité électronique*

Le problème essentiel dans l'étude des hétérostructures est la détermination de la barrière de potentiel qui existe aux différentes interfaces. Cette barrière conditionne le passage d'un électron ou d'un trou d'un matériau à l'autre. On la détermine à partir de la barrière que dans chacun des matériaux, l'électron doit franchir pour sortir du matériau dans le vide.

Considérons tout d'abord un métal, un électron de conduction est soumis de la part de tous les ions constituant le métal à un ensemble de forces dont la résultante est nulle. Il en résulte que cet électron est libre de se déplacer, sous l'action d'un champ appliqué par exemple. Lorsqu'un électron atteint la surface du métal, la compensation des forces dues aux ions n'est plus totale, de sorte que l'électron est retenu à l'intérieur du métal. Pour extraire cet électron il faut lui fournir de l'énergie. Au zéro degré absolu, tous les électrons libres étant situés au-dessous du niveau de Fermi, l'énergie minimum qu'il faut fournir pour extraire un électron du métal est l'énergie nécessaire à l'extraction d'un électron du niveau de Fermi. Cette quantité est appelée *Travail de sortie* du métal, on l'appellera $e\phi_m$, elle est caractéristique du métal (Fig. 1-33). Le tableau (1-3) donne le travail de sortie de certains métaux utilisés en électronique ou en optoélectronique. Le niveau appelé NV sur la figure (1-31) est le *niveau du vide* , c'est l'énergie d'un électron extrait du métal sans vitesse initiale, c'est-à-dire l'énergie potentielle de l'électron dans le vide au voisinage du métal.

Il est utile de détailler un peu cette définition qui peut amener un certain nombre de questions. Pour cela il faut revenir au problème de base, la description des états des électrons libres dans une structure périodique mais en prenant en compte cette fois le fait que le volume soit fini. Dans cette description les conditions aux limites se traduisent uniquement par les relations de Born von Karman conduisant à quantifier le vecteur d'onde. Cette description amène à calculer les énergies comme solution de l'équation de Schrödinger.

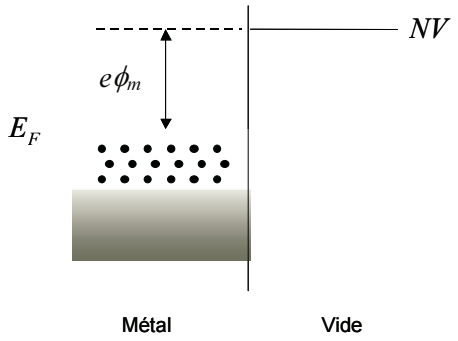


Figure 1-33 : Travail d'extraction dans le cas d'un métal

En fait, la prise en compte de la dimension finie du cristal conduit à supposer que le potentiel auquel est soumis l'électron est beaucoup plus élevé à l'extérieur qu'à l'intérieur ce qui confine les électrons dans le cristal. Une modélisation simplifiée du problème amène donc à considérer un potentiel périodique comme le montre la figure 1-34. Rappelons que le potentiel est défini à une constante additive près. Dans le cas des électrons libres cette constante est choisie pour que le potentiel soit nul à l'infini.

Le potentiel électrique s'écrit donc de manière générale

$$V = \sum_i \frac{1}{4\pi\varepsilon} \cdot \frac{q_i}{r_i} + C$$

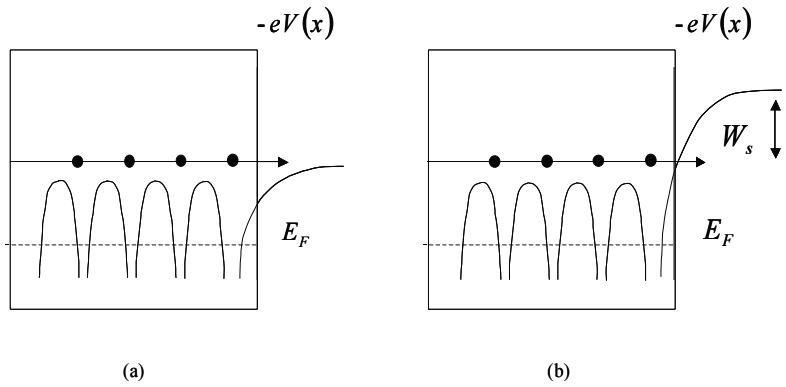


Figure 1-34: Potentiel en fonction des interfaces

On somme sur tous les atomes du cristal. En pratique, on associe électrons liés et noyaux . Dans la description semi-classique du mouvement des porteurs, on oublie les variations du potentiel à l'échelle du réseau pour ne considérer que sa valeur moyenne qui varie peu à cette échelle.

L'énergie potentielle est de manière générale

$$-eV = -\sum_i \frac{1}{4\pi\varepsilon} \cdot \frac{eq_i}{r_i} - Ce$$

La figure 1-34a représente à une dimension les variations de l'énergie potentielle quand le cristal est considéré infini. Les énergies solutions de l'équation de Schrödinger sont négatives. Si le potentiel n'était pas déformé en sortie, le travail nécessaire pour extraire un électron serait simplement E_F niveau de Fermi des électrons dans le cristal. En fait, le potentiel est très déformé à la surface à cause de la présence d'une couche dipolaire et il faut fournir un travail beaucoup plus important comme le montre la figure 1-34(b). On choisit alors la constante additive pour que le potentiel à l'intérieur du solide soit identique à celui qui serait calculé dans le cas 1-34(a). On suppose que le travail de sortie pour extraire un électron au niveau de Fermi et l'amener dans le vide à proximité de son lieu d'extraction et sans vitesse initiale dans le vide est une caractéristique du métal.

Considérons maintenant un semiconducteur ou un isolant, le travail de sortie est défini de la même manière. Cependant, si le travail de sortie est un paramètre spécifique d'un métal, il n'en est pas de même pour un semiconducteur en raison du fait que le niveau de Fermi de ce dernier est essentiellement fluctuant en fonction du dopage. Notons en outre que, sauf pour les semiconducteurs dégénérés, il n'y a pas d'électron au niveau de Fermi. On caractérise alors le semiconducteur par un autre paramètre qui est l'énergie qu'il faut fournir à un électron situé au bas de la bande de conduction, pour l'extraire du semiconducteur et l'amener dans le vide sans vitesse initiale, c'est l'*affinité électronique,* on l'appellera $e\chi$. On définit la même quantité pour un isolant. Ce paramètre est une grandeur spécifique du semiconducteur ou de l'isolant. Le tableau (1-3) donne les affinités électroniques de quelques semiconducteurs et isolants utilisés en microélectronique.

Ainsi, pour extraire un électron de conduction du métal et l'amener dans le vide, il faut lui fournir une énergie $e\phi_m$; pour extraire un électron de conduction du semiconducteur et l'amener dans le vide il faut lui fournir une énergie $e\chi$. Imaginons que le métal et le semiconducteur soient séparés par un intervalle très faible que l'on fait tendre vers une distance interatomique. Il faut fournir l'énergie $e\phi_m$ pour extraire l'électron du métal. Cet électron restitue l'énergie $e\chi$ en entrant dans le semiconducteur.

TABLEAU 1-3 : TRAVAIL DE SORTIE DES MÉTAUX USUELS ET AFFINITÉ DES SEMICONDUCTEURS

	Travail de sortie $e\Phi_m$ (eV)	**Affinité $e\chi$ (eV)**	**Travail de sortie maximum (eV)**
Li	2,3		
Na	2,3		
K	2,2		
Rb	2,2		
Cs	1,8		
Fe	4,4		
Ni	4,5		
Al	4,1		
Cu	4,4		
Ag	4,3		
Au	4,8		
Pt	5,3		
Si		4,01	5,13
Ge		4,13	4,79
AlP		3 ,44	5,89
AlAs		3,5	5,66
AlSb		3,6	5,20
GaP		4,3	6,55
GaAs		4,07	5,50
GaSb		4,06	4,74
InP		4,38	5,65
InAs		4,90	5,26
ZnS		3,9	7,48
ZnSe		4,09	6,76
ZnTe		3,5	5,76
CdTe		4,28	5,76
SiO2		1,1	

Il en résulte qu'au niveau de l'interface, la barrière de potentiel que doit franchir l'électron pour passer du métal dans le semiconducteur est donnée par (Fig. 1-35)

$$E_b = e\phi_m - e\chi \tag{1-233}$$

De la même manière, lorsque deux semiconducteurs différents sont au contact l'un de l'autre (Fig. 1-35), il existe à l'interface une barrière de potentiel donnée par

$$E_b = e\chi_1 - e\chi_2 \tag{1-234}$$

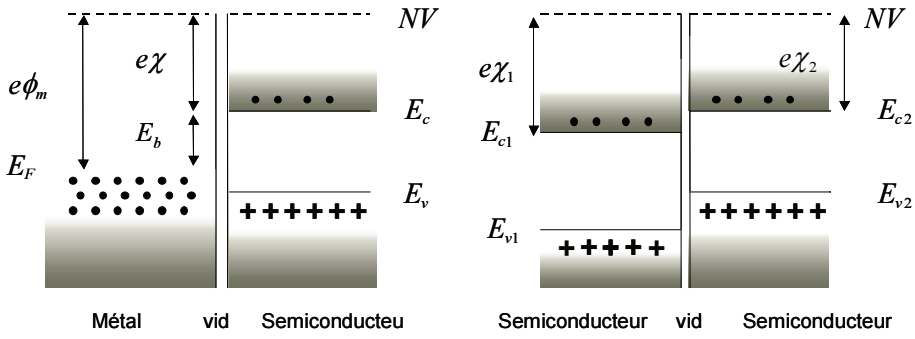


Figure 1-35 : Travail de sortie et affinité électronique

1.6.2. *Effet Schottky*

Dans ce qui précède nous avons implicitement supposé que l'électron extrait du matériau n'avait plus aucune interaction avec celui-ci. Nous avons négligé le fait que lorsque l'électron est émis par le matériau, il polarise celui-ci. Il en résulte une force de rétention de l'électron par le matériau, c'est *l'effet Schottky*. On montre que la force exercée par le matériau sur l'électron émis à la distance x de ce dernier est la même que celle qu'exercerait une charge $+e$ située à la distance $-x$ dans le matériau. Cette charge est appelée *charge image*, la force qui en résulte est la *force image* et le potentiel dont dérive cette force est le *potentiel image*. La force image est donnée par

$$F = -\frac{1}{4\pi\varepsilon_o}\frac{e^2}{(2x)^2}$$

Le travail nécessaire pour amener l'électron à l'infini, depuis sa position à la distance x du matériau est donné par

$$\int_x^\infty F\,dx = \frac{e^2}{16\pi\varepsilon_o x}$$

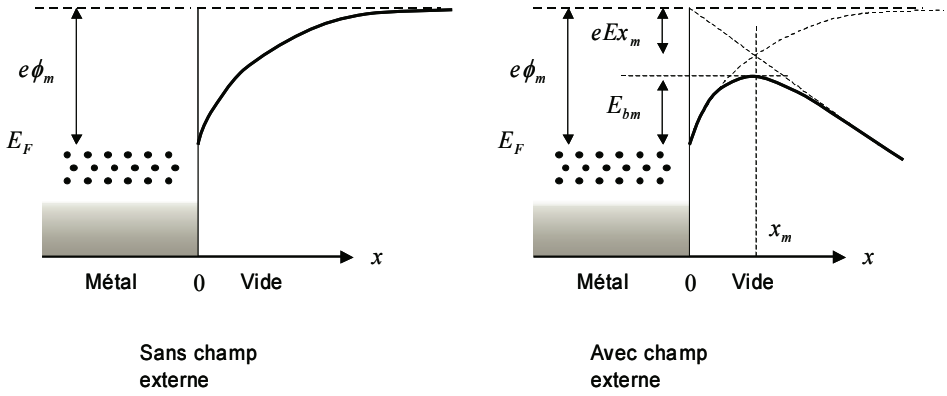


Figure 1-36 : Effet Schottky avec et sans champ externe

L'électron extrait du matériau ne passe donc pas brutalement à une énergie potentielle nulle, mais évolue progressivement vers celle-ci à mesure que son interaction avec le matériau diminue, c'est-à-dire à mesure qu'il s'éloigne de celui-ci. Son énergie ne devient rigoureusement nulle que lorsqu'il est rejeté à l'infini. A la distance x du matériau son énergie potentielle est égale au travail, changé de signe, qu'il faut fournir pour l'amener à l'infini.

$$W(x) = -\frac{e^2}{16\pi\varepsilon_o x} \tag{1-235}$$

Pour des raisons de continuité cette énergie potentielle tend vers E_F à la surface du matériau. Ceci résulte du fait que lorsque l'électron est dans le matériau la force image disparaît (Fig. 1-36).

La barrière de potentiel n'est donc pas une fonction du type Heaviside de la forme $E_b(x) = e\phi_m H(x)$ avec $H(x) = 0$ pour $x < 0$ et $H(x) = 1$ pour $x > 0$, mais une fonction de la forme

$$E_b(x) = e\phi_m - \frac{e^2}{16\pi\varepsilon_0 x} \tag{1-236}$$

Cette expression montre que le travail de sortie est l'énergie nécessaire pour éloigner définitivement l'électron du matériau. La barrière que doit franchir l'électron conserve la même hauteur $e\phi$ mais elle est étalée dans l'espace.

Ceci n'est plus vrai s'il existe un champ électrique à l'extérieur du matériau. En effet, dans ce cas l'énergie de l'électron à l'extérieur du matériau s'écrit

$$E_b(x) = e\phi_m - \frac{e^2}{16\pi\varepsilon_o x} - eV(x) \tag{1-237}$$

avec $V(x)=-\int_O^x E(x)dx$. Considérons un champ uniforme, dirigé vers le matériau $E(x)=-E$, la barrière s'écrit

$$E_b(x)=e\phi_m-\frac{e}{16\pi\varepsilon_o x}-eEx \tag{1-238}$$

La barrière de potentiel présente une valeur maximum E_{bm} en un point x_m (Fig. 1-36) donnés respectivement par

$$E_{bm}=e\phi_m-2eEx_m \tag{1-239-a}$$
$$x_m=\sqrt{e/16\pi\varepsilon_0 E} \tag{1-239-b}$$

Les actions conjuguées du potentiel image et du champ électrique, entraînent donc un abaissement de la barrière de potentiel.

Le même phénomène existe, dans une hétérostructure au niveau de l'interface. Il suffit de remplacer le travail de sortie $e\phi$, par la hauteur de barrière E_b et la constante diélectrique du vide ε_o par une combinaison des constantes diélectriques des deux matériaux

$$E_{bm}=E_b-2eEx_m \tag{1-240-a}$$
$$x_m=\sqrt{e/16\pi\varepsilon E} \tag{1-240-b}$$

Il faut noter que dans le cas d'une hétérostructure le champ électrique peut résulter de la charge d'espace éventuelle qui peut exister au niveau de l'interface. Notons toutefois qu'en prenant pour la constante diélectrique une valeur moyenne de l'ordre de $10\varepsilon_o$, on obtient $E_{bm}=E_b-10^{-5}E^{1/2}$ eV. Ainsi avec un champ électrique $E=10^6$ V/m la diminution de la hauteur de barrière est de l'ordre de 10 meV, c'est-à-dire nettement inférieure à kT à la température ambiante.

1.6.3. *Etats de surface et d'interface*

Les états électroniques dans le volume du semiconducteur sont d'une part les bandes de valence et de conduction résultant de la périodicité du réseau cristallin, et d'autre part les états discrets associés aux donneurs et accepteurs ou aux centres profonds. A la surface du semiconducteur les états électroniques sont modifiés en raison d'une part d'un phénomène intrinsèque et d'autre part de phénomènes extrinsèques. Le phénomène intrinsèque résulte de la rupture de périodicité du réseau. Dans le volume, chaque atome établit des liaisons avec chacun de ses voisins, en surface l'atome n'établit de liaison que dans un demi-plan, il reste côté vide ce que l'on a coutume d'appeler des *liaisons pendantes*. Cette rupture de périodicité entraîne l'existence d'états électroniques différents de ceux existant dans le volume.

A ce phénomène intrinsèque il faut ajouter des phénomènes extrinsèques résultant de l'adsorption à la surface d'atomes étrangers dont les plus courants sont les atomes d'oxygène qui entraînent une oxydation de la surface du semiconducteur. La présence d'une part d'atomes étrangers et d'autre part de distorsions du réseau, résultant de la différence de maille entre le semiconducteur et son oxyde, entraîne l'existence d'états de surface extrinsèques.

Enfin, si on considère l'interface entre deux matériaux au niveau d'une hétérostructure, le réseau passe sur une distance de quelques angströms de la périodicité d'un matériau à celle de l'autre. Il en résulte des états électroniques différents de ceux de chacun des matériaux, ce sont des *états d'interface*. Ces états jouent des rôles différents suivant que les niveaux d'énergie qui leur sont associés sont situés dans le gap du semiconducteur ou dans une bande permise. Dans le fonctionnement des hétérostructures, les états jouant un rôle important sont ceux qui introduisent des niveaux d'énergie permis dans le gap, car ils piègent des porteurs libres et par conséquent modifient les populations des bandes permises. Il en résulte qu'au voisinage de la surface, la distance du niveau de Fermi aux bandes permises est modifiée. Mais le niveau de Fermi étant horizontal à l'équilibre thermodynamique, les bandes se courbent vers le bas ou le haut suivant que la population électronique augmente ou diminue. Cette courbure traduit l'existence d'une différence de potentiel entre la surface du semiconducteur et son volume. Si l'énergie du bas de la bande de conduction est E_{cv} dans le volume et E_{cs} à la surface, la différence de potentiel entre la surface et le volume est donnée par

$$V_s = -\left(E_{cs} - E_{cv}\right)/e \qquad (1\text{-}241)$$

On appelle cette quantité le *potentiel de surface*. C'est la barrière que doit franchir un électron de volume du semiconducteur pour atteindre la surface. Cette quantité s'ajoute à l'affinité électronique pour extraire du semiconducteur un électron de volume, situé dans la bande de conduction.

La figure (1-37) représente le diagramme énergétique d'un semiconducteur de type n, en l'absence (a) et en présence (b,c) d'états de surface. La figure (1-37-b) représente le cas d'un semiconducteur dans lequel les états de surface créent des niveaux accepteurs en densité relativement faible devant la densité de donneurs, c'est-à-dire devant la densité de porteurs majoritaires à la température ambiante. Il en résulte que ces états de surface, qui piègent les électrons, sont saturés sans trop affecter la densité de porteurs majoritaires, c'est-à-dire la nature du semiconducteur. Au voisinage de la surface le semiconducteur présente une légère courbure de bandes mais reste de type *n*. Dans la figure (1-37-c) les états de surface créent des niveaux accepteurs en densité surfacique importante, il en résulte qu'au voisinage de la surface le semiconducteur devient de type *p*. Il existe une *couche d'inversion* à la surface du semiconducteur. La densité d'états de surface est souvent telle qu'elle entraîne un *ancrage du niveau de Fermi* (Annexe A-4).

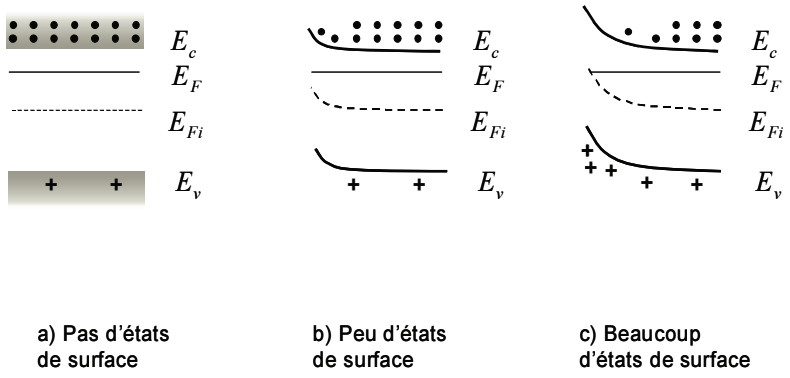


Figure 1-37 : Evolution de la courbure de bandes d'un semiconducteur de type n avec la densité d'états accepteurs de surface. E_{Fi} représente le niveau de Fermi intrinséque.

1.6.4. *Emission thermoélectronique dans les hétérostructures*

A l'interface entre deux matériaux, la barrière de potentiel constitue un obstacle au passage des électrons d'un matériau dans l'autre. Au zéro absolu aucun électron ne peut passer par dessus la barrière. Lorsque la température est différente de zéro l'énergie thermique permet à certains électrons de franchir la barrière, c'est *l'émission thermoélectronique*. Ce phénomène existe à une interface entre deux matériaux au même titre qu'à une interface métal-vide et se calcule de la même manière.

a . Métal-Vide

Considérons un métal caractérisé par son travail de sortie $e\phi_m$ (Fig. 1-38).

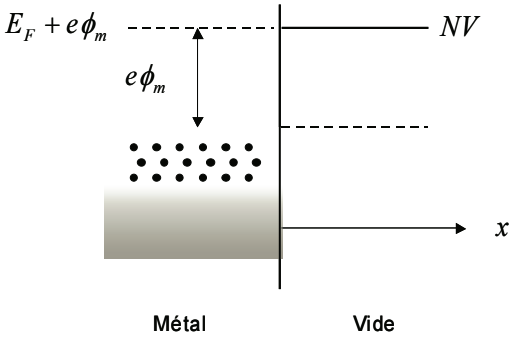


Figure 1-38 : Émission thermoélectronique d'un métal

Les électrons susceptibles de sortir du métal dans la direction x sont ceux dont l'énergie cinétique dans le métal E est telle que

$$E>E_F+e\phi_m \tag{1-242}$$

Pour calculer le nombre d'électrons susceptibles de sortir, calculons d'abord dans le vide le nombre d'électrons d'énergie E , ayant une composante de vitesse $v_{x.}$ L'énergie d'un électron dans le vide s'écrit (figure 1-34)

$$E=\frac{\hbar^2k^2}{2m}-eV=\frac{\hbar^2}{2m}\left(k_x^2+k_y^2+k_z^2\right)+W_s \tag{1-243}$$

Dans cette formule W_s est le travail à fournir pour faire passer un électron à travers la couche électrique de surface. La figure 1-34 montre que:

$$e\phi_m = W_s - E_F$$

La fonction de distribution des électrons dans le vide s'écrit

$$f(E)=\frac{1}{1+e^{\frac{E-E_F}{kT}}}$$

Le niveau de Fermi est en effet le même que dans le métal.

D'autre part, dans le vide on peut écrire : $\mathbf{p}=m\mathbf{v}=\hbar\mathbf{k}$. Ainsi les électrons d'énergie E, ayant une composante de vitesse suivant x comprise entre v_x et v_x+dv_x , ont une composante de vecteur d'onde suivant x comprise entre k_x et k_x+dk_x . Leur nombre est obtenu en sommant sur toutes les valeurs possibles des composantes k_y et k_z

$$dn(v_x)=dn(k_x)=g(k)\,dk_x\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(E)dk_y\,dk_z$$

où $g(k)$ = $2/(2\pi)^3$ représente la densité d'états dans l'espace des k et $f(E)$ la probabilité pour que l'électron ait l'énergie E, c'est-à-dire la fonction de Fermi.
Dans la mesure où les électrons que l'on considère ici sont tels que $E\text{-}E_F\geq e\phi_m$ avec $e\phi_m>>kT$, on peut ramener la fonction de Fermi à une distribution de Boltzmann

$$f(E)=e^{-(E-E_F)/kT}=e^{-\left(\hbar^2(k_x^2+k_y^2+k_z^2)/2m\ +e\phi_m\right)/kT} \tag{1-244}$$

En explicitant $f(E)$ on obtient

$$dn(k_x)=\frac{2}{(2\pi)^3}dk_x e^{-\left(\hbar^2 k_x^2/2m+e\phi_m\right)/kT}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}e^{-\hbar^2\left(k_y^2+k_z^2\right)/2mkT}dk_y dk_z$$

En posant $\alpha = \hbar^2/2mkT$ et l'intégrale double s'écrit

$$I=\left(\int_{-\infty}^{+\infty}e^{-\alpha u^2}du\right)^2=\frac{\pi}{\alpha}=\frac{2\pi mkT}{\hbar^2}$$

ainsi

$$dn(k_x)=\frac{2mkT}{h^2}dk_x e^{-\left(\hbar^2 k_x^2/2m+e\phi_m\right)/kT}$$

$dn(k_x)$ étant le nombre d'électrons dont la composante de vitesse dans la direction x est comprise entre v_x et v_x+dv_x, la densité de courant transportée par ces électrons est donnée par

$$dJ_x=-e\,dn(k_x)\,v_x \qquad (1\text{-}245)$$

La densité totale de courant dans la direction x s'écrit

$$j_x=-e\int_0^{\infty}v_x dn(k_x) \qquad (1\text{-}246)$$

Dans le but d'utiliser k_x comme variable d'intégration, explicitons v_x en fonction de k_x

$$v_x=\frac{\hbar}{m}k_x$$

L'expression (1-246) s'écrit alors

$$j_x=-e\frac{kT}{\pi h}e^{-e\phi_m/kT}\int_0^{\infty}e^{-\hbar^2k_x^2/2mkT}k_x dk_x$$

En posant $\alpha=\hbar/2mkT$ et $u=\alpha k_x^2$, $du=2\alpha k_x\,dk_x$ et l'intégrale s'écrit

$$I=\frac{1}{2\alpha}\int_0^{\infty}e^{-u}du=\frac{1}{2\alpha}$$

ou en intégrant et en explicitant α, la densité de courant s'écrit alors

$$j_x=-A\,T^2\,e^{-e\phi_m/kT} \qquad (1\text{-}247)$$

Avec $$A=\frac{4\pi emk^2}{h^3}=120\,A\,cm^{-2}K^{-2}$$

A est la *constante de Richardson.* Le signe moins traduit simplement le fait que lorsqu'un électron sort du métal dans la direction des *x* positifs, il crée un courant en sens inverse.

b . Métal-Semiconducteur

La figure (1-39) représente les états d'énergie dans le métal et dans le semiconducteur à l'interface. $e\phi'_F$ représente dans le semiconducteur, la distance bande de conduction-niveau de Fermi en x=0. Il faut noter qu'en raison de la présence éventuelle de charges d'espace et par suite de courbure des bandes, la distance $e\phi'_F$ à l'interface peut être différente de la distance $e\phi_F$ dans la région neutre du semiconducteur, loin de l'interface.

La barrière de potentiel que doivent franchir les électrons pour passer du métal dans le semiconducteur est $E_b=e\phi_m-e\chi$.

Dans le métal, les électrons sont caractérisés par une masse effective m_1 et des composantes de vitesse v_{1x}, v_{1y}, v_{1z}, lorsqu'ils passent dans le semiconducteur, leur masse effective change et une partie de leur énergie cinétique se transforme en énergie potentielle, leurs paramètres caractéristiques deviennent m_2, v_{2x}, v_{2y}, v_{2z}.

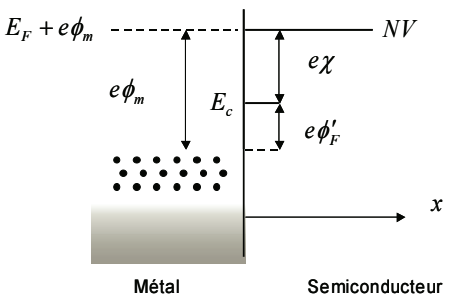


Figure 1-39 : Courant thermoélectrique d'une jonction métal-semiconducteur

Les électrons susceptibles d'entrer dans le semiconducteur sont ceux dont la composante de vitesse est $v_{2x}\geq 0$. Ces électrons ont une énergie totale donnée par

$$E=E_C+\frac{1}{2}m_2v^2=E_F+E_b+\frac{1}{2}m_2v^2 \qquad (1\text{-}248)$$

En reprenant le même calcul que pour l'émission thermoélectronique dans le vide, on obtient pour le nombre d'électrons d'énergie *E*, ayant dans le semiconducteur une vitesse de composante en *x* comprise entre v_{2x} et $v_{2x}+dv_{2x}$, l'expression,

$$dn(v_{2x})=dn(k_{2x})=\frac{2m_2kT}{h^2}dk_{2x}e^{-(E_b+\hbar^2k_2^2/2m_2)/kT} \qquad (1\text{-}249)$$

La densité de courant correspondante est donnée par $dJ_x = -e\,dn(v_{2x})\,v_{2x}$, et $j_x = -e\int_{v_{2x}=0}^{\infty} v_{2x}\,dn(v_{2x})$. En explicitant comme dans le paragraphe précédent v_{2x} en fonction de k_{2x} on obtient l'expression

$$j_x = -\frac{ekT}{\pi h} e^{-E_b/kT} \int_0^{\infty} e^{-\hbar^2 k_{2x}^2/2m_2 kT}\, k_{2x}\, dk_{2x} \tag{1-250}$$

Ce qui donne en intégrant

$$j_x = -A^* T^2\, e^{-E_b/kT} \tag{1-251}$$

avec $$A^* = A\frac{m_2}{m} = 120\frac{m_2}{m}\ A\,cm^{-2}\,K^{-2}$$

c . Semiconducteur-Semiconducteur

La figure (1-40) représente le diagramme énergétique à l'interface de deux semiconducteurs. La barrière de potentiel est donnée par $E_b = e\chi_1 - e\chi_2$. Ici encore, il faut préciser que $e\phi'_{F1}$ et $e\phi'_{F2}$ représentent les distances bande de conduction-niveau de Fermi, à l'interface, et peuvent être différents de $e\phi_{F1}$ et $e\phi_{F2}$ dans les régions neutres de chacun des semiconducteurs.

Comme précédemment, le courant thermoélectronique est le courant transporté par les électrons qui dans le semiconducteur 2 ont une composante $v_{2x} > 0$.

Ces électrons ont une énergie totale donnée par

$$E = E_{c2} + \frac{1}{2}m_2 v_2^2 = E_F + e\phi'_{F2} + \frac{1}{2}m_2 v_2^2 \tag{1-252}$$

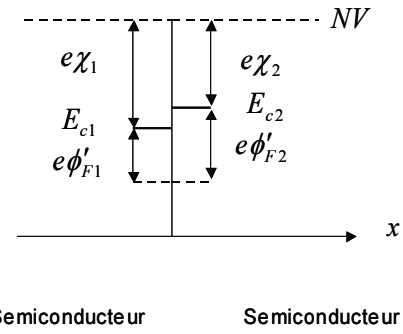


Figure 1-40 : Courant thermoélectrique d'une jonction semiconducteur semiconducteur

Le nombre d'électrons d'énergie E ayant dans le semiconducteur 2 une vitesse dont la composante en x est comprise entre v_{2x} et $v_{2x} + dv_{2x}$ est donné par

$$dn(v_{2x}) = dn(k_{2x}) = \frac{2m_2 kT}{h^2} dk_{2x} e^{-\left(e\phi'_{F2} + \hbar^2 k_{2x}^2 / 2m_2\right)/kT} \tag{1-253}$$

La densité de courant résultant du franchissement de la barrière par les électrons qui ont une composante $v_{2x} > 0$ est donnée par $j_x = -e\int_a^\infty v_{2x}\, dn(v_{2x})$ ou en utilisant la variable k_{2x}

$$j_x = -\frac{ekT}{\pi h} e^{-e\phi'_{F2}/kT} \int_0^\infty e^{-\hbar^2 k_{2x}^2 / 2m_2 kT} k_{2x} dk_{2x} \tag{1-254}$$

soit

$$j_x = -A^* T^2 e^{-e\phi'_{F2}/kT} \tag{1-255}$$

avec $A^* = Am_2/m$.

Nous obtenons la même expression que pour la structure métal-semiconducteur. La seule différence est que dans ce dernier cas le paramètre à l'interface est imposé par la différence entre le travail de sortie du métal et l'affinité électronique du semiconducteur $(e\phi'_F = e\phi_m - e\chi = E_b)$, de sorte que $e\phi'_F = E_b$ est une caractéristique du couple métal-semiconducteur considéré. Dans la structure semiconducteur-semiconducteur, le terme $e\phi'_{F2}$ est différent de $E_b = e\chi_1 - e\chi_2$. Ce n'est plus un paramètre spécifique du couple de semiconducteurs, mais il dépend d'une part des dopages respectifs de chaque semiconducteur et d'autre part des éventuelles charges d'espace qui peuvent se développer à l'interface.

La figure (1-40) montre que l'on peut écrire $e\phi'_{F2} = e\phi'_{F1} + E_b$. On peut alors écrire la densité de courant (1-255) sous la forme

$$j_x = -A^* T^2 e^{-\left(e\phi'_{F1} + E_b\right)/kT} \tag{1-256}$$

ou plus simplement

$$j_x = -A_1^* T^2 e^{-E_b/kT} \tag{1-257}$$

avec

$$A_1^* = A^* e^{-e\phi'_{F1}/kT} = A\frac{m_2}{m_1} e^{-e\phi'_{F1}/kT}$$

L'expression (1-257) est tout à fait comparable à l'expression (1-251), la seule différence est un coefficient multiplicatif entre A_1^* et A^*. Ce coefficient multiplicatif, $\exp\left(-e\phi'_{F1}/kT\right)$, explicite le fait que le nombre d'électrons susceptibles de passer du semiconducteur 1 vers le semiconducteur 2 est proportionnel au nombre d'électrons présents dans le semiconducteur 1 au

voisinage de l'interface. Ce coefficient représente tout simplement la probabilité pour que l'électron soit présent dans la bande de conduction du semiconducteur 1 au moment de franchir la barrière, il est évidemment égal à 1 dans le cas du contact métal-semiconducteur.

Nous avons supposé dans tout ce qui précède que tous les électrons ayant une composante de vitesse suivant x vérifiant la condition de franchissement de la barrière, passaient effectivement du milieu 1 dans le milieu 2. En fait, ceci est vrai en mécanique classique, mais ne l'est plus tout à fait en mécanique ondulatoire. En effet, à la traversée de l'interface, l'électron conserve son énergie totale mais en raison de la discontinuité de milieu, change de masse effective et de vitesse, c'est-à-dire de quantité de mouvement. Il en résulte que l'onde associée à l'électron change de vecteur d'onde et par suite de longueur d'onde. En d'autres termes les milieux 1 et 2 présentent, pour cette onde, des indices n_1 et n_2 différents. Il en résulte une réflexion partielle de l'onde, caractérisée par un coefficient de réflexion donné par $R=\left((n_1-n_2)/(n_1+n_2)\right)^2$ et par suite un coefficient de transmission donné par $T=1-R=4n_1n_2/(n_1+n_2)^2$. Les électrons ne vérifiant pas la condition de vitesse ont une composante de vitesse v_{2x} nulle ou négative dans le milieu 2, ce qui correspond à un indice de réfraction n tendant vers l'infini ou purement imaginaire. Il en résulte $R=1$ et $T=0$, c'est-à-dire une réflexion totale de l'onde, on retrouve la limite de la mécanique classique. En ce qui concerne les électrons vérifiant la condition classique de vitesse $v_{2x}>0$, il faut noter que dès que leur énergie cinétique $m_2v_2^2/2$ atteint une valeur de l'ordre de kT à la température ambiante, leur longueur d'onde est de l'ordre de quelques centaines d'angströms. A des longueurs d'onde aussi courtes les différents milieux ont des indices comparables et voisins de 1 de sorte que le coefficient de transmission devient très vite égal à 1. Seuls les électrons ayant l'énergie minimum pour franchir la barrière subiront une réflexion partielle. On peut donc avec une bonne approximation supposer que le coefficient de transmission est nul pour $v_{2x}<0$ et égal à 1 pour $v_{2x}>0$.

1.6.5. *Emission de champ*

L'émission thermoélectronique n'est pas le seul effet régissant le transfert d'électrons entre deux matériaux. Cet effet traduit le franchissement de la barrière de potentiel d'interface par saut, thermiquement activé, au-dessus de la barrière. Un autre effet, typiquement quantique, permet le transfert d'électrons, c'est l'*effet tunnel*. Si la barrière est relativement étroite, quelques nanomètres, la nature ondulatoire de l'électron permet à ce dernier de passer à travers celle-ci. Or à l'interface entre deux matériaux, on peut réduire artificiellement la largeur de la barrière d'interface par l'action d'un champ électrique, interne ou appliqué. Lorsque le champ électrique présent à l'interface entre deux matériaux atteint des valeurs de l'ordre de 10^7 V/cm, la barrière de potentiel devient suffisamment étroite pour autoriser un transfert par effet tunnel. Une cathode peut ainsi émettre des électrons même à la température ambiante. Cette émission *froide* porte le nom d'*émission de champ*, elle résulte de la nature ondulatoire de l'électron.

a. Emission de champ dans le vide

La figure (1-41) représente la barrière de potentiel à l'interface métal-vide, en présence d'un champ électrique E_0 à la surface du métal. L'effet de la force image est négligé. Le taux d'émission d'électrons, c'est-à-dire le coefficient de transmission de la barrière est donné en annexe A1. Les bornes d'intégration sont ici x_1=0 et x_2=w. La forme de la barrière et sa largeur pour un électron d'énergie E, sont ici données par

$$U(x) = E_F + e\phi_m - eE_0 x \qquad \text{(1-258a)}$$

$$w = \left(E_F + e\phi_m - E\right)/eE_0 \qquad \text{(1-258b)}$$

Ainsi le coefficient de transmission de la barrière de surface s'écrit

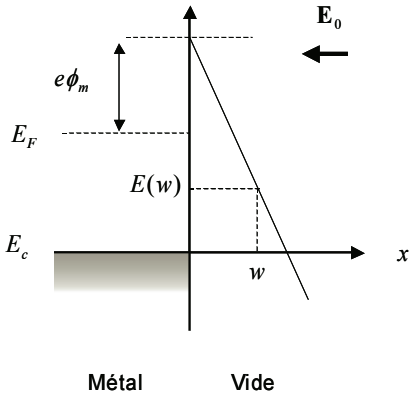


Figure 1-41 : Emission de champ dans le vide

$$T = \exp\left(-2\left(\frac{2m}{\hbar^2}\right)^{1/2}\int_0^w \left(E_F + e\phi_m - E - eE_0 x\right)^{1/2} dx\right) \qquad \text{(1-259)}$$

En calculant l'intégrale et en explicitant w , on obtient

$$T = \exp\left(-\frac{4}{3}\left(\frac{2m}{\hbar^2}\right)^{1/2}\frac{\left(E_F + e\phi_m - E\right)^{3/2}}{eE_0}\right) \qquad \text{(1-260)}$$

b. Emission de champ interne

La même expression donne le taux d'émission à une interface métal-semiconducteur. Il suffit de remplacer la masse m de l'électron dans le vide par sa masse effective m_e dans le semiconducteur et le travail de sortie $e\phi_m$ du métal par la barrière de Schottky E_b=$e\phi_m$-$e\chi$ du contact métal-semiconducteur. L'expression de T s'écrit alors

$$T = \exp\left(-\frac{4}{3}\left(\frac{2m_e}{\hbar^2}\right)^{1/2}\frac{\left(E_F + E_b - E\right)^{3/2}}{eE_0}\right) \qquad \text{(1-261)}$$

c. Courant d'émission de champ

Les électrons d'énergie *E,* qui ont la probabilité *T(E)* de traverser la barrière dans la direction x , sont ceux dont l'énergie est $E = m^* v^2 \big/ 2 = p^2 \big/ 2m^*$, où p est le moment et m^* la masse effective des électrons dans le métal. La densité de tels électrons est $dn(p) = 2 \big/ h^3 . dp_x dp_y dp_z$. La densité de courant transportée par ces électrons s'écrit

$$dJ = -e\, v_x\, T(v_x)\, dn(v_x) = -e \frac{p_x}{m^*} T(p_x)\, dn(p_x) \qquad (1\text{-}262)$$

L'intégration de cette expression sur tous les électrons de la bande de conduction du métal donne la densité totale du courant d'émission de champ

$$J = -\frac{2e}{m^* h^3} \iiint p_x T(p_x) dp_x dp_y dp_z \qquad (1\text{-}263)$$

On obtient

$$J = -C_1\, E_0^2\, e^{-C_2/E_0} \qquad (1\text{-}264)$$

Pour l'émission de champ dans le vide et l'émission de champ interne, dans un semiconducteur, les constantes C_1 et C_2 sont respectivement

$$C_1 = \frac{m^*}{m} \frac{e^2}{8\pi h \phi_m} \quad C_2 = \frac{4}{3} \left(\frac{2em}{\hbar^2} \right)^{1/2} \phi_m^{3/2}$$

$$C_1 = \frac{m^*}{m_e} \frac{e^3}{8\pi h E_b} \quad C_2 = \frac{4}{3} \left(\frac{2m_e}{e^2 \hbar^2} \right)^{1/2} E_b^{3/2}$$

1.7. Les sources de bruit dans les semiconducteurs

1.7.1. *Définitions générales pour le signal de bruit*

Les considérations précédentes ont permis de calculer le courant dans un semiconducteur. En fait elles ont conduit à calculer la valeur moyenne du courant au sens statistique du terme sans se soucier des fluctuations autour de cette valeur moyenne. Les fluctuations sont dues à la nature corpusculaire du courant qui est la somme des effets créés par le déplacement de chaque charge élémentaire, électron ou trou. Le but de ce dernier paragraphe est de quantifier ces fluctuations appelées bruit.

Le bruit a une grande importance en électronique car il affecte la précision d'une mesure et introduit des aléas dans le fonctionnement des dispositifs logiques. La figure 1-42 montre un signal en sortie d'une voie électronique suffisamment amplifié pour que les fluctuations soient visibles. A la valeur moyenne du signal, constante dans ce cas, s'ajoute un signal variable dans le temps de plus faible amplitude qui est le signal de bruit.

On admet généralement que le signal de bruit est formé de composantes de différentes fréquences, la gamme est continue et assez large (plusieurs GHz pour les composants actuels). Cette décomposition est assez naturelle si on pense à la décomposition du signal en intégrale de Fourier. Quand le signal de bruit est observé avec une bande passante plus faible les fréquences élevées sont éliminées et le signal observé est de plus faible amplitude comme le montre la figure 1-42.

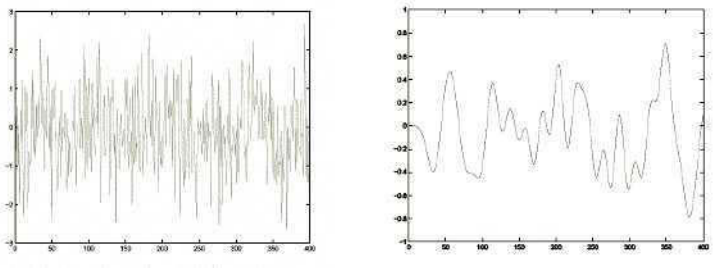


Figure 1-42 : Signal de bruit en sortie d'une voie électronique.

Pour comprendre la formation du signal de bruit, il faut revenir à l'équation générale (1-221) exprimant le courant dans un semiconducteur.

$$\frac{\partial n}{\partial t} = \frac{1}{e} div \mathbf{J}_n + g_n - r_n$$

Exprimons le courant comme la somme du courant de dérive et du courant de diffusion.

$$\frac{\partial n}{\partial t} = \frac{1}{e} div\left(\mathbf{J}_{nc} + \mathbf{J}_{nd}\right) + g_n - r_n \qquad (1\text{-}265)$$

- Les fluctuations sur le courant de conduction $\mathbf{J}_{nc}$ donnent lieu au bruit de grenaille ou bruit « shot ».
- Les fluctuations sur le courant de diffusion $\mathbf{J}_{nd}$ donnent lieu au bruit thermique.

- Les fluctuations sur les taux de génération et recombinaison donnent lieu à un bruit basse fréquence.
- Enfin, les variations de mobilité conduisent à des fluctuations sur la valeur du courant de conduction et créent le bruit dit en *1/f* .

Pour caractériser le bruit on définit une grandeur qui est liée à la moyenne des écarts quadratiques par rapport à la valeur moyenne ce qui permet de mesurer la puissance du bruit au sens électrique du terme. Cette grandeur est la *densité spectrale* ou DSP. La DSP d'un signal aléatoire *x(t)* est définie de manière générale par

$$S_x(f) = \lim_{T\to\infty} \frac{1}{2T} E\left|X_T(f)\right|^2 \tag{1-266}$$

Dans cette relation, le terme $X_T(f)$ désigne la transformée de Fourier du signal $x_T(t)$ égal au signal de bruit de $-T$ à $+T$ et nul ailleurs. Cet artifice est nécessaire pour pouvoir définir la transformée de Fourier d'un signal qui ne converge pas vers zéro de manière absolue ce qui est le cas du signal de bruit. On calcule l'espérance mathématique (moyenne sur toutes les probabilités de réalisations possibles) pour tenir compte de la nature aléatoire du signal de bruit. Quand le paramètre T tend vers l'infini, le bruit et le signal $x_T(t)$ se confondent.

On utilisera par la suite deux propriétés importantes.

- La densité spectrale de bruit en sortie d'un système linéaire est le produit de la densité spectrale en entrée multipliée par le module au carré de la fonction de transfert

$$S_y(f) = S_x(f)\cdot\left|H(f)\right|^2 \tag{1-267}$$

- La densité spectrale de puissance est la transformée de Fourier du coefficient de corrélation défini par

$$r_x(\tau) = E\left[x(t)x(t+\tau)\right] \tag{1-268}$$

On suppose dans cette définition que le coefficient de corrélation ne dépend pas de la valeur de t mais uniquement du délai τ ce qui limite la définition aux signaux dits stationnaires au sens large. Cette importante relation constitue le théorème de Wiener Khintchine. Les bruits rencontrés dans les composants sont de ce type. Calculée en zéro, on obtient alors la formule

$$r_x(0) = E\left[x(t)x(t)\right] = \int_{-\infty}^{+\infty} S_x(f)df \tag{1-269}$$

Cette relation permet de calculer la puissance d'un signal de bruit et par la suite le rapport signal sur bruit défini comme le rapport entre la puissance du signal et celle du bruit. Le bruit souvent considéré de valeur moyenne nulle (il suffit de retrancher au signal total sa valeur moyenne) se caractérise par la moyenne temporelle ou statistique du carré de l'amplitude.

Il est souvent utile de calculer la densité spectrale de signaux composés. La densité spectrale de la somme de signaux non corrélés deux à deux est par exemple la somme des densités spectrales

$$S_{x+y} = S_x + S_y \tag{1-270}$$

De même, si a est une grandeur certaine

$$S_{ax} = a^2 \cdot S_x \tag{1-271}$$

Ces relations se démontrent facilement à partir de la définition (1-266).

1.7.2. Le bruit thermique

On s'intéresse au bruit lié à la composante de diffusion. On peut prouver par différentes méthodes utilisant la physique statistique que le bruit lié aux fluctuations d'un courant traversant un morceau de semiconducteur de résistance R peut se représenter par une source de courant en parallèle sur la résistance dont la densité spectrale de puissance est donnée par

$$S_i = \frac{4kT}{R} \tag{1-272}$$

La DSP est calculée pour le bruit associé au courant. Il est facile de montrer en utilisant le théoème de Thévenin de la théorie des circuits que le bruit peut également se représenter sous la forme d'une source de tension en série avec la résistance. La DSP associée à cette tension s'écrit alors

$$S_v = 4kTR \tag{1-273}$$

Quelques questions peuvent se poser :

- Que devient la DSP pour une impédance complexe ?
- Cette formule est–elle valable dans tout le domaine de fréquences ?
- Que devient cette formule quand la relation entre courant et tension n'est pas linéaire ?

Pour traiter le cas d'une impédance complexe, il suffit de prendre la partie réelle de l'impédance totale.

La relation (1-273) peut s'étendre à un domaine de fréquences plus large et s'écrit alors

$$S_v = 4R\left[\frac{1}{2}hf + \frac{hf}{e^{\frac{hf}{kT}} - 1}\right] \tag{1-274}$$

Quand la fréquence tend vers zéro, on retrouve la relation (1-273).

Le cas non linéaire est plus complexe. La relation entre la tension aux bornes du dispositif et le courant traversant le dispositif n'est plus linéaire comme dans le cas d'une résistance. Ce problème a été traité par Gunn en 1968 et par Gupta en 1978. Sous certaines hypothèses assez générales, la DSP devient

$$S_v = 4kT\left(\frac{dV}{dI} + I\frac{d^2V}{dI^2}\right)_{I=E(I_{DC})} \tag{1-275}$$

Dans cette relation, les valeurs des dérivées sont calculées pour la valeur moyenne au sens statistique du terme du courant I_{DC} traversant le dispositif. Le cas d'une diode peut se traiter de cette manière.

1.7.3. *Le bruit de grenaille*

On s'intéresse maintenant à la composante de conduction dans le courant total (voir relation 1-265). Dans la direction du champ on suppose donc que les porteurs franchissent la zone dans laquelle il y a un champ électrique. Cette région peut être supposée de faible largeur dans un premier temps. On oublie dans cette analyse les composantes de diffusion de la vitesse qui ont été étudiées dans le paragraphe précédent pour s'intéresser uniquement à la valeur moyenne de la vitesse de dérive. Les régions dans lesquelles il y a un champ important sont les zones de charge d'espace dans un dispositif semiconducteur comme il a été expliqué dans le paragraphe 1-5-4.

Le courant total traversant une surface est la somme des courants dûs à chaque charge élémentaire (électron ou trou) traversant le dispositif. Le courant élémentaire créé par un électron traversant la surface au temps t' est donc :

$$i(t) = e\delta(t - t')$$

Dans cette relation la fonction $\delta(t - t')$ est la fonction de Dirac centrée en t', grandeur aléatoire. Observons l'effet de ce courant en sortie d'une chaîne linéaire sous forme d'une tension par exemple comme le montre la figure 1-43.

$$v(t) = e \cdot h(t - t')$$

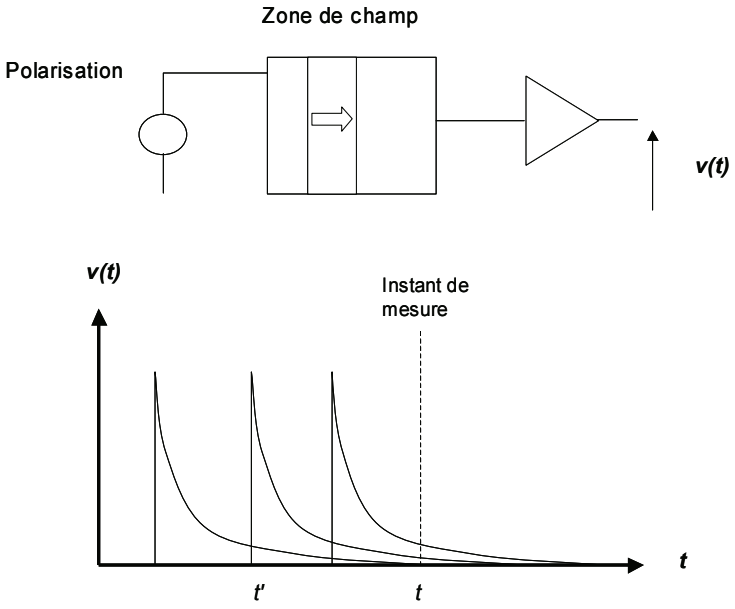


Figure 1.43 : Bruit de grenaille

La fonction *h* est la réponse impulsionnelle de la chaîne linéaire. Le signal observé total est la somme de tous les signaux créés par les électrons franchissant la zone de champ avant le temps *t* considéré.

$$v(t) = \int_{-\infty}^{t} ndt'e \cdot h(t-t')dt'$$

Danc cette relation nt' est le nombre d'électrons traversant la zone de champ en t'. On admet que ce nombre aléatoire infinitésimal est indépendant de t'.
Il est alors possible de calculer espérance et variance de la variable aléatoire *v*

$$E(v(t)) = \int_{-\infty}^{t} E(ndt')e \cdot h(t-t')dt'$$

L'espérance de ndt' peut s'écrire $\overline{n}dt'$ avec les hypothèse précédentes. Les deux valeurs possibles sont 1 et 0 avec les probabilités $\overline{n}dt'$ et $1-\overline{n}dt'$. La variance de ndt' est donc par définition

$$\text{var}(ndt') = (1-\overline{n}dt')(0-\overline{n}dt')^2 + \overline{n}dt'(1-\overline{n}dt')^2 = \overline{n}dt'$$

Si on admet que les franchissements de la zone de champ par chaque électron sont des événements non corrélés, la variance du signal total observé est

$$\mathrm{var}(v(t)) = \int_{-\infty}^{t} \mathrm{var}(ndt')e^2 \cdot h(t-t')^2\,dt' = e^2\overline{n}\int_{-\infty}^{t} h(t-t')^2\,dt'$$

Un simple changement de variable conduit à

$$\mathrm{var}(v(t) = e^2\overline{n}\int_{-\infty}^{+\infty} h(u)^2\,du \tag{1-276}$$

On peut exprimer le paramètre $\overline{n}$ en fonction du courant moyen *I* qui traverse la zone de champ en écrivant

$$I = e\overline{n}$$

$$\mathrm{var}(v(t) = eI\int_{-\infty}^{+\infty} h(u)^2\,du$$

Il est facile d'exprimer l'intégrale en fonction de la fonction de transfert *H(f)* de la chaîne, la fonction de transfert étant la transformée de Fourier de la réponse impulsionnelle.

$$\mathrm{var}(v(t) = eI\int_{-\infty}^{+\infty} \mid H(f) \mid^2 df$$

On utilise le fait que le module de la fonction de transfert est une fonction paire pour écrire

$$\mathrm{var}(v(t) = 2eI\int_{0}^{+\infty} \mid H(f) \mid^2 df$$

La comparaison avec la relation 1-267 montre que la DSP du bruit de grenaille du courant I est donc

$$S_i = 2eI \tag{1-277}$$

Cette relation importante sera reprise dans la suite.
Quand le temps de passage τ à travers la zone de champ n'est plus supposé nul, la densité spectrale de bruit devient

$$S_i(f) = 2eI\left(\frac{\sin 2\pi f\frac{\tau}{2}}{2\pi f\frac{\tau}{2}}\right)^2 \tag{1-278}$$

La démonstration de ce résultat est proposée en exercice.

1.7.4. *Bruit de génération-recombinaison*

Ce bruit est associé aux fluctuations des générations et recombinaisons des porteurs. Etudions dans un premier temps le cas d'un semiconducteur comportant N porteurs libres de durée de vie τ. On peut alors écrire l'équation de continuité pour la variation $\Delta N(t)$.

$$\frac{d}{dt}(\Delta N(t)) = -\frac{\Delta N}{\tau} + H(t) \qquad (1\text{-}279)$$

La fonction $H(t)$ représente la source des variations. On suppose que sa densité spectrale est constante avec la fréquence. En prenant la transformée de Fourier de cette équation, on obtient

$$i2\pi f \Delta N(f) = -\frac{\Delta N(f)}{\tau} + H(f)$$

On en déduit facilement

$$S_{\Delta N}(f) = S_H(0)\frac{\tau^2}{1+4\pi^2 f^2 \tau^2}$$

Par ailleurs le théorème de Wiener Khintchine appliqué à la variable aléatoire $\Delta N(t)$ permet d'écrire

$$E(\Delta N^2) = \int_0^\infty S_{\Delta N}(f)\frac{\tau^2}{1+4\pi^2 f^2 \tau^2} df = S_H(0)\,\tau \int_0^\infty \frac{du}{1+u^2} = S_H(0)\,\tau\,\frac{1}{4}$$

On en déduit la relation

$$S_{\Delta N}(f) = 4E(\Delta N^2)\frac{\tau}{1+4\pi^2 f^2 \tau^2} \qquad (1\text{-}280)$$

Dans cette relation on reconnaît la variance de ΔN. Cette relation permet de calculer la densité spectrale de la fluctuation de porteurs en fonction de la variance sur le nombre de porteurs. Un calcul assez complexe (Van der Ziel) permet de calculer la variance du nombre de porteurs en fonction des taux g et r de génération et recombinaison pour la valeur moyenne N_0 et de leurs dérivées g' et r'.

$$E(\Delta N^2) = \frac{1}{2}\frac{g(N_0)+r(N_0)}{r'(N_0)-g'(N_0)}$$

Les phénomènes de génération-recombinaison conduisent à la création d'un bruit dont la densité spectrale varie inversement proportionnellement à la fréquence ce qui justifie l'appellation « bruit en *1/f* ».

Pour prouver cette dépendance, il suffit de considérer que *N* centres de génération-recombinaison sont présents. La densité spectrale s'écrit alors

$$S_{\Delta N} = \sum_i \frac{4\tau_i A_i}{1+\omega^2 \tau_i^{\,2}}$$

Il est facile de passer au continu en supposant une grande quantité de centres différents.

$$g(\tau_i) = \frac{A_i}{\sum_i A_i}$$

$$S_{\Delta N} = 4 \operatorname{var} \Delta N \int_0^\infty \frac{\tau \cdot g(\tau) d\tau}{1+\omega^2 \tau^{\,2}}$$

En admettant que la densité de centres varie en $1/\tau$ dans un intervalle $(\tau_1\ \tau_2)$ on obtient

$$S_{\Delta N} = \frac{2 \operatorname{var} \Delta N}{\pi\ f \ln\left(\frac{\tau_1}{\tau_0}\right)} \left[\tan^{-1}(\omega\tau_1) - \tan^{-1}(\omega\tau_0)\right]$$

On obtient donc bien une variation en 1*/f*.

EXERCICES DU CHAPITRE 1

- **Exercice 1 : Relations de commutation entre opérateurs**

Les opérateurs position $\hat{x}$ et impulsion $\hat{p}$ sont définis par : $\hat{x}\psi(x) = x\psi(x)$ et $\hat{p}\psi(x) = \frac{\hbar}{i}\frac{d\psi}{dx}$. Calculer les opérateurs $\hat{x}\hat{p}$, $\hat{p}\hat{x}$, $\hat{x}\hat{p} - \hat{p}\hat{x}$

- **Exercice 2 : Réseau réciproque d'un réseau cubique centré**

Montrez que les vecteurs fondamentaux du réseau cubique centré sont $\mathbf{a} = \frac{a}{2}(\mathbf{x}+\mathbf{y}-\mathbf{z})$ $\mathbf{b} = \frac{a}{2}(-\mathbf{x}+\mathbf{y}+\mathbf{z})$ $\mathbf{c} = \frac{a}{2}(\mathbf{x}-\mathbf{y}+\mathbf{z})$
Calculez les vecteurs fondamentaux du réseau réciproque. En déduire que c'est un réseau cubique face centré.

- **Exercice 3 : Électrons libres**

Ecrire l'équation de Schrödinger à une dimension d'un électron libre dans un solide de dimension *L*. Résoudre cette équation avec la condition aux limites $\psi(0) = \psi(L)$. Si *N* électrons sont disponibles calculez l'énergie maximale d'un électron en fonction de *N/L*. Application numérique avec *N/L* =0,4 électron par angström.
Calculer l'énergie totale des *N* électrons et la densité d'états.
Reprendre le calcul pour un système à trois dimensions en fonction de *N/V.* Application numérique pour *N/V* =10^{22} électrons par cm^3.

- **Exercice 4 : Électron dans un puits de potentiel**

Dans un système à une dimension, le potentiel vaut 0 de $-a$ à $+a$ et vaut V_0 ailleurs. Résoudre l'équation de Schröndinger dans les trois régions. En écrivant la continuité de la fonction d'onde et de sa dérivée et en tenant compte de la symétrie du problème montrez que les solutions sont ou symétriques ou antisymétriques et de la forme :

$-\infty,-a \quad \psi(x) = Be^{Kx}$ $\qquad -a,+a \quad \psi(x) = A\cos Kx$ $\qquad a,+\infty \quad \psi(x) = Be^{-Kx}$

$-\infty,-a \quad \psi(x) = -De^{Kx}$ $\qquad -a,+a \quad \psi(x) = C\sin Kx$ $\qquad a,+\infty \quad \psi(x) = De^{-Kx}$

Montrez que

$$k^2a^2 + K^2a^2 = \frac{2mV_0a^2}{\hbar^2}$$

En déduire les solutions possibles.
Quelle est la condition pour qu'il y ait un seul état lié ?

- **Exercice 5 : Étalement du paquet d'ondes**

Soit une particule de masse *m* dans un potentiel *V(x).*
On notera $\chi = \langle x^2 \rangle - \langle x \rangle^2$ dans laquelle *x* est l'opérateur position et $\theta = \langle p^2 \rangle - \langle p \rangle^2$ dans laquelle *p* est l'opérateur impulsion. Le centre du paquet d'onde est la moyenne de l'opérateur position *x*. Cette définition est quasi identique à l'approximation classique à la valeur donnée par la stationnarité de la phase.
On exprimera le potentiel en série : $V(x) = V(\langle x \rangle) + (x - \langle x \rangle) V'(\langle x \rangle) + ...$

Montrez que

$$\frac{d}{dt}\chi = \frac{1}{m}\left[\langle pq + qp \rangle - 2\langle p \rangle \langle q \rangle\right] \quad \text{et} \quad \frac{d^2}{dt^2}\chi = \frac{2\theta}{m^2} - \frac{1}{m}\left[\langle V'x + xV' \rangle - 2\langle x \rangle \langle V' \rangle\right]$$

Montrez que les développements en série conduisent à la relation approchée

$$\frac{d^2}{dt^2}\chi = \frac{2}{m^2}\left[\theta - mV''(\langle x \rangle)\chi\right]$$

Dans le cas d'une particule libre, le potentiel est nul ; montrez que

$$\chi = \chi_0 + \chi_0' t + \frac{\theta_0}{m^2} t^2$$

En conclure que le paquet d'ondes s'étale indéfiniment.

- **Exercice 6 : Vitesse de l'électron dans un état de Bloch**

Montrez que la valeur moyenne de la vitesse d'un électron dans un état de Bloch est

$$\mathbf{v} = \frac{1}{\hbar}\frac{\partial E_{nk}}{\partial \mathbf{k}}$$

On peut utiliser la théorie des perturbations au premier ordre.
Comparez cette approche à celle du paquet d'ondes.

- **Exercice 7 : Densité d'occupation des trous**

Montrer que la probabilité d'occupation d'un état *E* par un trou dans la bande de valence est 1-*F(E)* dans laquelle *F* est la fonction de Fermi.

- **Exercice 9 : Dopage et énergies**

Un échantillon de silicium est dopé en accepteurs à 10^{15} atomes par cm^3. Quel est le dopage à apporter pour qu'il devienne de type n avec un niveau de Fermi à 300K situé à 0,2 eV de la bande de conduction ?

- **Exercice 10 : Libre parcours moyen**

Calculez le libre parcours moyen à 300K d'un électron dans le silicium. On suppose que la masse effective est 0,26 la masse et que la vitesse est thermique.

- **Exercice 11 : Résistivité**

Calculer la résistivité du silicium dopé à 10^{15} atomes par cm^3.
Calculez celle du germanium pour le même dopage.

- **Exercice 11 : Diffusion**

Un champ de 50V/cm est appliqué à un échantillon de silicium. La mobilité des trous est de 200 cm^2V-s. Calculer vitesse de dérive et coefficient de diffusion des trous.

- **Exercice 12 : Champ dans un semiconducteur en équilibre**

Un semiconducteur est dopé de manière non uniforme $N_D(x)$. Montrer que le champ est donné par : $E(x) = \frac{-kT}{e} \frac{1}{N_D(x)} \frac{dN_D}{dx}$

- **Exercice 12 : Bruit de grenaille**

Quand le temps de transit dans la zone de champ est τ, montrez que la DSP du bruit de quantification sur le courant est

$$DSP_{Ic}(f) = 2eI \left(\frac{\sin 2\pi f \frac{\tau}{2}}{2\pi f \frac{\tau}{2}} \right)^2$$

- **Exercice 13 : Bruit aux bornes d'un circuit RC**

Calculez le bruit aux bornes d'un circuit RC parallèle en supposant que la bande passante de mesure n'est limitée que par le circuit lui-même.

- **Exercice 14 : Densité spectrale**

Démontrez les relations 1-267, 1-270 et 1-271

- **Exercice 15 : Recombinaisons**

Calculez *n* et *p* dans un semiconducteur dopé n à $10^{15}/cm^3$ et éclairé avec un taux de réaction de $10^{16}cm^{-3}s^{-1}$. Les durées de vie des électrons et des trous sont de 10 microsecondes.

- **Exercice 16 : Effet tunnel**

Calculez le coefficient de transmission d'une barrière de 6 eV pour un électron de 2 eV avec des épaisseurs de barrière de 0,1 et 1 nm.

- **Exercice 17 : Niveaux de Fermi**

Calculez les niveaux de Fermi du silicium n dopé de 10^{15} à 10^{19} dopants/cm^3 et justifiez l'hypothèse *n* peu différent de N_D

- **Exercice 18 : Variations des niveaux énergétiques**

Comment varient les niveaux *Ec*, *Ev*, et *Ei* dans un semiconducteur dans lequel il y a un champ électrique. Comparez à un métal.

2. Jonction pn

Une jonction pn est constituée par la juxtaposition de deux régions de types différents d'un même monocristal de semiconducteur. La différence des densités de donneurs et d'accepteurs, N_d-N_a, passe d'une valeur négative dans la région de type p à une valeur positive dans la région de type n. La loi de variation de cette grandeur dépend essentiellement de la technique de fabrication. Différents modèles peuvent être utilisés pour étudier théoriquement les propriétés de la jonction, jonction abrupte, linéaire… Le modèle de la jonction abrupte donne des résultats en très bon accord avec le comportement de la jonction. C'est le modèle que nous allons développer. Nous verrons ensuite comment généraliser les résultats à une jonction à profil de dopage quelconque.

La jonction pn est une structure de base dans les composants. Les composants étant formés de semiconducteurs dopés de manière différente, les jonctions pn ou np sont présentes aux interfaces. Il est donc indispensable de bien comprendre les phénomènes physiques qui s'y manifestent. La jonction pn est également un composant en soit. La fonction de ce composant est de laisser passer le courant dans un seul sens. Elle permet donc de transformer un signal alternatif en un signal unipolaire. Cette fonction est largement utilisée dans les systèmes électroniques d'alimentation ou de détection radiofréquence. Le chapitre est organisé comme suit.

2.1 Jonction abrupte à l'équilibre thermodynamique
2.2. Jonction abrupte polarisée
2.3. Jonction à profil de dopage quelconque
2.4. Capacités de la jonction pn
2.5. Génération-recombinaison dans la zone de charge d'espace
2.6. Jonction en régime transitoire et temps de recouvrement
2.7. Claquage de la jonction polarisée en inverse
2.8. Régime de forte injection

2.1. Jonction abrupte à l'équilibre thermodynamique

Dans le modèle de la jonction abrupte, la différence N_d-N_a passe brutalement dans le plan $x = 0$, d'une valeur négative dans la région de type p à une valeur positive dans la région de type n (Fig. 2-2-a).

Quand deux semiconducteurs de type p et de type n sont mis en contact (au sens métallurgique du terme), les trous, majoritaires dans la région de type p, diffusent vers la région de type n. Il en est de même pour les électrons, dans l'autre sens (figure 2-1). La diffusion des porteurs libres de part et d'autre de la jonction fait apparaître une *charge d'espace* résultant de la présence des donneurs et accepteurs ionisés, dont les charges ne sont plus intégralement compensées par celles des porteurs libres.

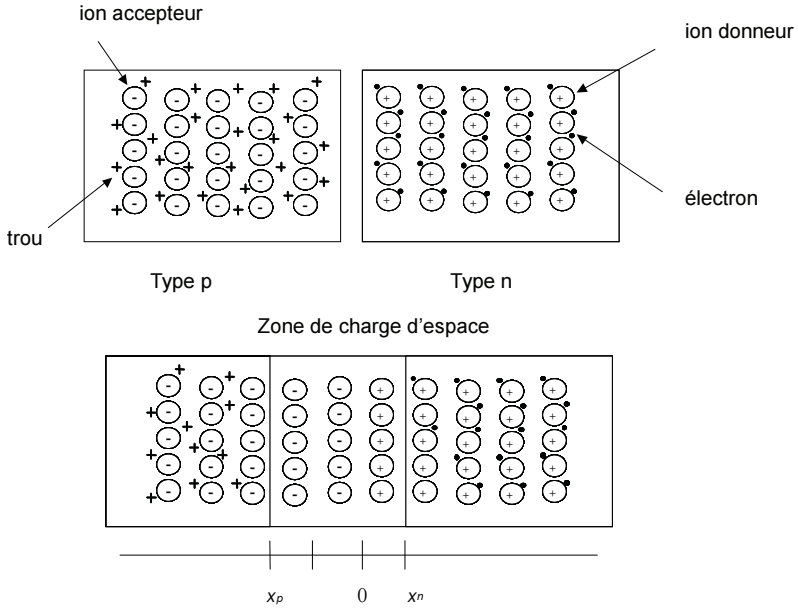


Figure 2-1 : La jonction pn

Il s'établit alors, au voisinage de la jonction métallurgique, un champ électrique qui s'oppose à la diffusion des porteurs majoritaires. L'équilibre thermodynamique est établi lorsque la force électrique, résultant de l'apparition du champ, équilibre la force de diffusion associée aux gradients de concentration de porteurs libres

2.1.1. *Charge d'espace*

Nous supposerons que dans chacune des régions la conductivité est de nature extrinsèque, c'est-à-dire que les conditions $(N_d - N_a)_n >> n_i$ et $(N_a - N_d)_p >> n_i$ sont remplies dans chacune des régions. En outre, dans le but de simplifier l'écriture, nous appellerons N_d l'excédent de donneurs dans la région de type n, $N_d = (N_d - N_a)_n$ et N_a l'excédent d'accepteurs dans la région de type p, $N_a = (N_a - N_d)_p$.

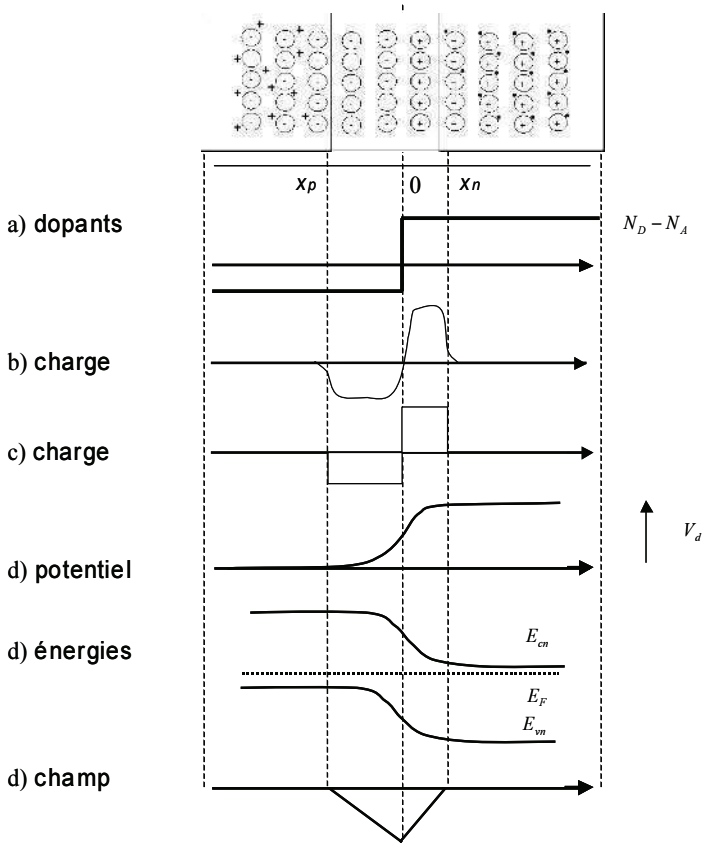


Figure 2-2 : Jonction pn à l'équilibre thermodynamique.

Les densités de porteurs libres dans chacune des régions sont alors données par

$$n_n = N_d \qquad p_n = n_i^2 / N_d \qquad \text{(2-1-a)}$$

$$n_p = n_i^2 / N_a \qquad p_p = N_a \qquad \text{(2-1-b)}$$

En supposant tous les donneurs et accepteurs ionisés, la charge d'espace dans chacune des régions de la jonction s'écrit

$$\rho(x) = e\left[N_d - N_a + p(x) - n(x)\right] \tag{2-2}$$

En raison de la présence du champ électrique, la diffusion-recombinaison des porteurs est limitée au voisinage de la jonction métallurgique. Loin de la jonction, les densités de porteurs libres sont données par les expressions (2-1) et le semiconducteur est neutre, $\rho(x) = 0$. La charge d'espace est représentée sur la figure (2-2-b) : côté n la charge est positive car elle résulte de la présence des donneurs ionisés N_d ; côté p la charge résulte de la présence des accepteurs ionisés N_a, elle est négative. Il faut noter que, la charge d'espace résultant de recombinaisons de paires électron-trou, les quantités totales de charges développées dans chacune des régions sont égales.

En raison de la présence, dans la zone de charge d'espace, d'un champ électrique intense, la densité de porteurs libres dans cette région est négligeable. En outre les frontières entre la zone dépeuplée et les zones neutres de la jonction sont très abruptes. Nous modéliserons la charge d'espace réelle de la figure (2-2-b) par la distribution représentée sur la figure (2-2-c). En d'autres termes, nous supposerons que la zone de charge d'espace est entièrement dépeuplée de porteurs libres et limitée par des frontières abruptes d'abscisses x_p et x_n. La densité de charge s'écrit dans cette hypothèse

$$\rho(x) = 0 \qquad \text{pour} \quad x < x_p \ \text{ et } \ x > x_n \tag{2-3a}$$
$$\rho(x) = -\,e\,N_a \qquad \text{pour} \quad x_p < x < 0 \tag{2-3b}$$
$$\rho(x) = e\,N_d \qquad \text{pour} \quad 0 < x < x_n \tag{2-3c}$$

2.1.2. *Tension de diffusion*

La présence d'une charge d'espace entraîne l'existence d'un champ électrique et d'une variation de potentiel. Le potentiel varie d'une valeur V_p dans la région neutre de type p, à une valeur V_n dans la région neutre de type n (Fig. 2-1-d). La différence de potentiel entre ces deux régions constitue une barrière de potentiel que l'on appelle *tension de diffusion*, en raison du fait que c'est la barrière qui équilibre les forces de diffusion

$$V_d = V_n - V_p$$

On peut calculer la tension V_d de deux manières en écrivant simplement que la structure est en équilibre thermodynamique. La première consiste à écrire que le niveau de Fermi est horizontal dans toute la structure, la seconde consiste à écrire que les transferts d'énergie sont nuls, c'est-à-dire que les courants d'électrons et de trous sont nuls.

Écrivons que le niveau de Fermi est le même dans toute la structure. Les densités d'électron dans chacune des régions s'écrivent

$$n_n = N_c e^{-(E_{cn}-E_F)/kT} \tag{2-4a}$$

$$n_p = N_c e^{-(E_{cp}-E_F)/kT} \tag{2-4b}$$

ce qui permet d'écrire

$$E_{cp} - E_{cn} = kT\ Ln \frac{n_n}{n_p} = kT\ \ Ln\ \frac{N_d\, N_a}{n_i^2} \tag{2-5}$$

La différence d'énergie E_{cp}-E_{cn} n'est autre que la différence d'énergie potentielle des électrons de conduction entre la région de type p et la région de type n. En effet, les électrons et les trous participant à la conduction étant situés aux extrémités des bandes, leur énergie cinétique est nulle puisque la vitesse de groupe est nulle (extremum de la fonction énergie) et toute leur énergie est potentielle.

$$V_d = V_n - V_p = (E_{cp} - E_{cn})/e$$

La tension de diffusion est par conséquent donnée par l'expression

$$V_d = \frac{kT}{e}\ Ln\ \frac{N_d\, N_a}{n_i^2} \tag{2-6}$$

Pour les dopages utilisés en pratique et compte tenu des valeurs de n_i à la température ambiante, les valeurs de V_d sont respectivement de l'ordre de 0,7 V et 0,35 V dans les jonctions au silicium et au germanium.

Écrivons maintenant que le courant de chaque type de porteur est nul

$$j_n = \mu_n \left(enE + kT \frac{dn}{dx} \right) \equiv 0 \tag{2-7-a}$$

$$j_p = \mu_p \left(epE - kT \frac{dp}{dx} \right) \equiv 0 \tag{2-7-b}$$

Ce qui permet d'écrire les équations différentielles

$$\frac{dn}{n} = -\frac{e}{kT} E\, dx = \frac{e}{kT}\, dV \tag{2-8-a}$$

$$\frac{dp}{p} = \frac{e}{kT} E\, dx = -\frac{e}{kT}\, dV \tag{2-8-b}$$

En intégrant l'équation (2-8-a) dans toute la zone de charge d'espace, c'est-à-dire de x_p à x_n on obtient

$$Ln\frac{n(x_n)}{n(x_p)}=\frac{e}{kT}(V_n-V_p)=\frac{e}{kT}V_d$$

En explicitant les densités de porteurs $n(x_n) = N_d$ et $n(x_p) = n_i^2/N_a$, on obtient pour la tension de diffusion V_d la même expression que précédemment (Eq. 2-6).

2.1.3. *Potentiel et champ électriques dans la zone de charge d'espace*

Il suffit, pour obtenir le potentiel et le champ électriques, d'intégrer l'équation de Poisson avec la densité de charge donnée par les équations (2-3).

$$\frac{d^2V}{dx^2}=-\frac{\rho(x)}{\varepsilon} \tag{2-9}$$

Pour $x_p < x < 0$ l'équation de Poisson s'écrit

$$\frac{d^2V}{dx^2}=\frac{eN_a}{\varepsilon} \tag{2-10}$$

En intégrant deux fois avec les conditions $E = 0$ et $V = V_p$ en $x = x_p$ on obtient

$$\frac{dV}{dx}=\frac{eN_a}{\varepsilon}(x-x_p) \tag{2-11}$$

$$V=\frac{eN_a}{2\varepsilon}(x-x_p)^2+V_p \tag{2-12}$$

Pour $0 < x < x_n$ l'équation de Poisson s'écrit

$$\frac{d^2V}{dx^2}=-\frac{eN_d}{\varepsilon} \tag{2-13}$$

En intégrant deux fois avec les conditions $E = 0$ et $V = V_n$ en $x = x_n$ on obtient

$$\frac{dV}{dx}=-\frac{eN_d}{\varepsilon}(x-x_n) \tag{2-14}$$

$$V=-\frac{eN_d}{2\varepsilon}(x-x_n)^2+V_n \tag{2-15}$$

Le champ électrique est dirigé suivant x et donné par $E = -dV/dx$, soit

$$\text{pour } x_p < x < 0 \qquad E = -\frac{eN_a}{\varepsilon}(x - x_p) \qquad \text{(2-16-a)}$$

$$\text{pour } 0 < x < x_n \qquad E = \frac{eN_d}{\varepsilon}(x - x_n) \qquad \text{(2-16-b)}$$

Les variations du potentiel et du champ électriques sont représentées sur les figures (2-2-d et f). La figure (2-2-e) représente l'évolution des bandes de valence et de conduction dans toute la structure.

2.1.4. *Largeur de la zone de charge d'espace*

La continuité en $x = 0$, de la composante normale du vecteur déplacement $\mathbf{D}=\varepsilon\mathbf{E}$, permet d'établir une relation entre x_n et x_p. Écrivons $\varepsilon E_{0^-} = \varepsilon E_{0^+}$, soit

$$e\,N_a\,x_p = -\,e\,N_d\,x_n$$

En posant $W_p = |x_p| = -x_p$ *et* $W_n = |x_n| = x_n$, la relation s'écrit

$$N_a W_p = N_d W_n \qquad \text{(2-17)}$$

Cette relation traduit simplement le fait que les surfaces grisées sur la figure (2-1-c) sont égales, c'est-à-dire que le total des charges négatives développées dans la région de type p est égal au total des charges positives développées dans la région de type n. L'expression (2-17) montre ainsi que la charge d'espace s'étend principalement dans la région la moins dopée. Cette propriété sera mise à profit en particulier dans les transistors à effet de champ.

On obtient l'expression de la largeur de la zone de charge d'espace en écrivant la continuité du potentiel en $x = 0$.

$$\frac{eN_a}{2\varepsilon}x_p^2 + V_p = -\frac{eN_d}{2\varepsilon}x_n^2 + V_n \qquad \text{(2-18)}$$

soit

$$V_d = V_n - V_p = \frac{e}{2\varepsilon}(N_d W_n^2 + N_a W_p^2) \qquad \text{(2-19)}$$

En utilisant la relation (2-17), cette expression s'écrit sous l'une ou l'autre des formes suivantes

$$V_d = \frac{eN_d}{2\varepsilon}W_n^2\left(1 + \frac{N_d}{N_a}\right) = \frac{eN_a}{2\varepsilon}W_p^2\left(1 + \frac{N_a}{N_d}\right)$$

Ce qui donne pour W_n et W_p les expressions

$$W_n^2 = \frac{2\varepsilon}{eN_d}\ \frac{1}{1+N_d/N_a}V_d \qquad \text{(2-20-a)}$$

$$W_p^2 = \frac{2\varepsilon}{eN_a}\ \frac{1}{1+N_a/N_d}V_d \qquad \text{(2-20-b)}$$

En explicitant la tension de diffusion à partir de l'expression (2-6) on obtient

$$W_n = 2L_{Dn}\left(\frac{1}{1+N_d/N_a}Ln\frac{N_aN_d}{n_i^2}\right)^{1/2} \qquad \text{(2-21-a)}$$

$$W_p = 2L_{Dp}\left(\frac{1}{1+N_a/N_d}Ln\frac{N_dN_a}{n_i^2}\right)^{1/2} \qquad \text{(2-21-b)}$$

avec

$$L_{Dn} = \left(\frac{\varepsilon kT}{2e^2N_d}\right)^{1/2} \qquad L_{Dp} = \left(\frac{\varepsilon kT}{2e^2N_a}\right)^{1/2} \qquad \text{(2-22)}$$

L_{Dn} *et* L_{Dp} sont les *longueurs de Debye* dans les régions de type n et de type p.

La largeur de la zone de charge d'espace est $W = W_n + W_p$. Si la jonction est très dissymétrique, avec par exemple $N_a >> N_d$, la zone de charge d'espace se développe essentiellement dans la région la moins dopée, sa largeur est donnée par

$$W \approx W_n \approx 2L_{Dn}\left(Ln\frac{N_aN_d}{n_i^2}\right)^{1/2} \qquad \text{(2-23)}$$

Prenons l'exemple d'une jonction pn au silicium dopée avec $N_a = 10^{18}$ cm^{-3} et $N_d = 10^{17}$ cm^{-3}, on obtient $W = 12\,L_{Dn}$. La largeur de la zone de charge d'espace est donc d'environ un ordre de grandeur supérieure à la longueur de Debye de la région la moins dopée. Suivant le profil et l'amplitude du dopage, la zone de charge d'espace d'une jonction pn varie entre 0,1 et 1 µm.

2.1.5. *Signification de la longueur de Debye*

Nous avons calculé le potentiel $V(x)$ dans la zone de charge d'espace dans l'hypothèse de la zone dépeuplée de porteurs. En fait, si le champ électrique est important dans la majeure partie de la zone de charge d'espace, il diminue au voisinage des frontières de cette zone pour s'annuler en $x = x_n$ et $x = x_p$. Il en

résulte qu'au voisinage de $x = x_n$ certains électrons pénètrent dans la zone de charge d'espace, de même que certains trous au voisinage de $x = x_p$.

En supposant que les porteurs libres qui pénètrent dans la zone de charge d'espace perturbent peu le potentiel, on peut calculer la densité d'électrons qui pénètrent dans cette zone au voisinage de $x = x_n$. Prenons l'origine des potentiels dans la région de type n, le potentiel au voisinage de $x = x_n$ est donné par l'expression (2-15) avec $V_n = 0$.

$$V(x) = -\frac{eN_d}{2\varepsilon}(x - x_n)^2 \tag{2-24}$$

En un point *x* quelconque, la densité d'électrons est donnée par

$$n(x) = N_c e^{-(E_c(x) - E_F)/kT} \tag{2-25}$$

Dans la partie neutre de la région de type n, le potentiel est $V_n = 0$, l'énergie de la bande de conduction est $E_c(x) = E_{cn}$, la densité d'électrons est donnée par

$$n_n = N_d = N_c e^{-(E_{cn} - E_F)/kT} \tag{2-26}$$

Dans la partie chargée de la région de type n, le potentiel est $V(x)$ donné par l'expression (2-24), l'énergie de la bande de conduction est $E_c(x) = E_{cn} - eV(x)$, et par la suite la densité d'électrons s'écrit

$$n(x) = N_c e^{-(E_{cn} - eV(x) - E_F)/kT} \tag{2-27}$$

Le rapport des expressions (2-27 et 26) donne

$$n(x) = N_d e^{eV(x)/kT} \tag{2-28}$$

En explicitant $V(x)$ dans cette expression on obtient

$$n(x) = N_d e^{-\frac{e^2 N_d}{2\varepsilon kT}(x - x_n)^2} \tag{2-29}$$

ou

$$n(x) = N_d e^{-\left(\frac{x - x_n}{2L_{Dn}}\right)^2} \tag{2-30}$$

Cette expression montre que $2L_{Dn}$ est la longueur sur laquelle la densité d'électrons libres passe de N_d à N_d/e . ***En d'autres termes cette longueur mesure la profondeur de pénétration des électrons dans la zone de charge d'espace. La grandeur*** $2L_{Dp}$ ***mesure la même grandeur pour les trous dans la région de type p.***
L'expression (2-23) permet d'expliciter le rapport $W_n / 2L_{Dn}$

$$\frac{W_n}{2L_{Dn}} \approx \left(Ln \frac{N_a N_d}{n_i^2} \right)^{1/2} \tag{2-31}$$

W_n représente la largeur dans la région n, de la zone de charge d'espace, et $2\,L_{Dn}$ la profondeur de pénétration des électrons dans cette région. L'expression (2-31) montre que ce rapport est d'autant plus grand que la jonction est plus dopée. Il en résulte que l'hypothèse de la zone de charge d'espace dépeuplée est d'autant plus justifiée que la jonction est plus dopée, c'est-à-dire que le gradient de dopage est plus grand. On retrouve les conditions discutées au chapitre 1 (Par. 1.5.4.a).

Revenons sur l'expression (2-28). Cette expression représente la densité d'électrons $n(x)$ en un point x où le potentiel est $V(x)$, en fonction de la densité d'électrons N_d en un point où le potentiel est $V = 0$, en l'occurrence la partie neutre de la région de type n de la jonction.

Nous utiliserons cette relation dans d'autres circonstances, elle peut être généralisée sous la forme

$$n(V_o + V) = n(V_o)\, e^{eV/kT} \tag{2-32}$$

où $n(V_0 + V)$ représente la densité d'électrons en un point du semiconducteur où le potentiel est $V_0 + V$, en fonction de $n(V_0)$ la densité d'électrons en un point du semiconducteur où le potentiel est V_0.

2.2. Jonction abrupte polarisée

Lorsque l'on polarise la jonction, on modifie la barrière de potentiel et par suite la diffusion des porteurs d'une région vers l'autre. De manière plus précise, les résultats du chapitre 1 montrent que les niveaux de Fermi ne sont plus alignés mais décalés de la tension appliquée. On considère généralement que les électrons et les trous ont le même niveau de Fermi en dehors de la zone de charge d'espace mais ont des niveaux différents (un pour les trous et un pour les électrons) dans la zone de charge d'espace qui n'est plus à l'équilibre thermodynamique. Relions à la masse la région de type n de la jonction et portons au potentiel V_a la région de type p (Fig. 2-3-a). ***Il faut distinguer la tension de polarisation V_a de la diode que constitue la structure, de la tension de polarisation V_j de la jonction proprement dite***. La tension V_a s'écrit $V_a = V_j + V_s$, où V_j représente la tension appliquée aux bornes de la zone de charge d'espace ($x_p < x < x_n$) et V_s la somme des chutes ohmiques dans les résistances série que constituent les régions neutres ($x < x_p$ et $x > x_n$). En raison de la nature conductrice des régions neutres (dopées) et de la nature isolante de la zone de charge d'espace (vide de porteurs) les conditions $V_j >> V_s$ et par suite $V_a \approx V_j$ sont vérifiées toutes les fois que la zone de charge d'espace existe. ***Dans la suite nous appellerons V la tension V_j afin d'alléger l'écriture.***

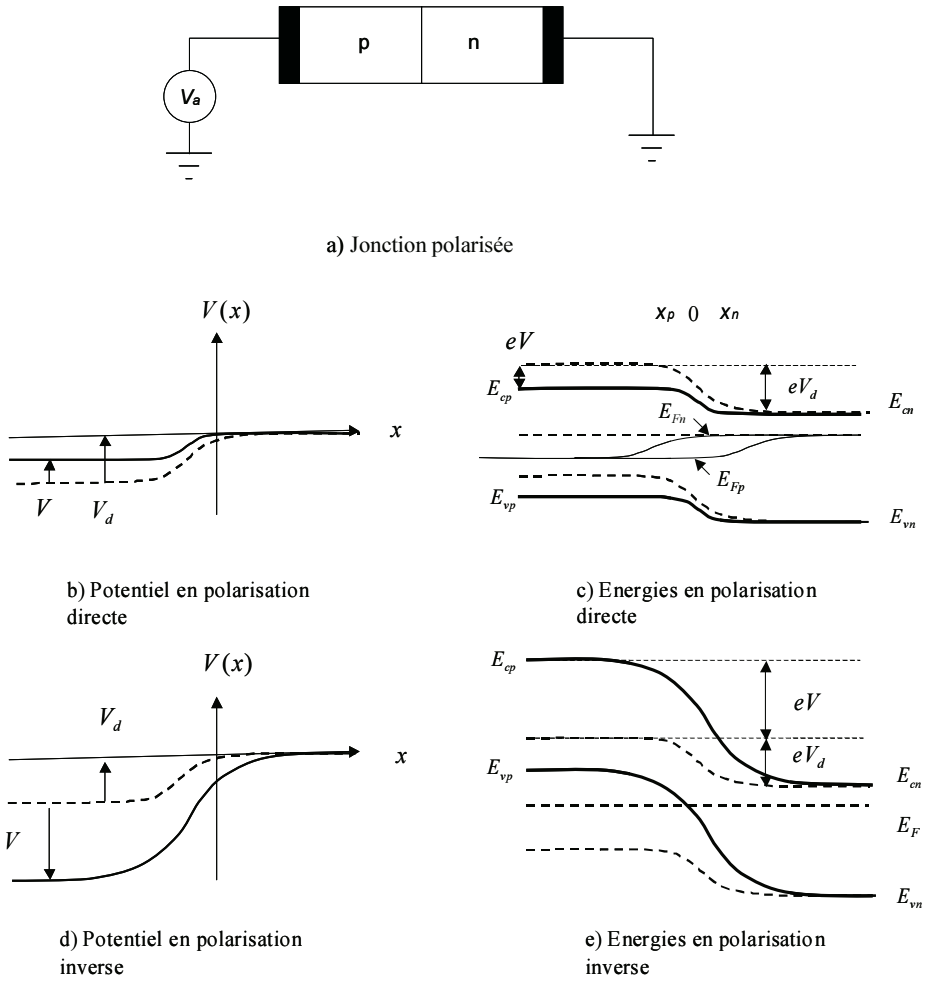


Figure 2-3 : Jonction pn polarisée

Lorsque la tension de polarisation V est positive, la différence de potentiel entre les régions n et p au niveau de la jonction, devient $V_n - V_p = V_d - V$ (Fig. 2-3-b) et la hauteur de la barrière de potentiel devient $e(V_d - V)$ (Fig. 2-3-c). Il suffit d'observer le décalage des niveaux de Fermi dans les régions neutres pour expliquer cette modification. La différence des niveaux de Fermi dans les régions neutres est la différence de potentiel appliquée. Bande de valence et bande de conduction se positionnent dans les régions neutres par rapport au niveau de Fermi en fonction du dopage. La barrière de potentiel n'est plus suffisante pour arrêter la

diffusion des porteurs, les électrons diffusent de la région n vers la région p, et les trous de la région p vers la région n. La diode est polarisée dans le sens direct, le courant direct circule de la région p vers la région n.

Lorsque la tension de polarisation V est négative, la différence de potentiel aux bornes de la zone de charge d'espace est augmentée, $V_n - V_p = V_d + |V|$ (Fig. 2-3-d), et la hauteur de la barrière de potentiel devient $e(V_d + |V|)$ (Fig. 2-3-e). L'équilibre entre le courant de conduction et le courant de diffusion est rompu mais cette fois au profit du courant de conduction. La diffusion des porteurs majoritaires est bloquée, seuls les porteurs minoritaires qui atteignent la zone de charge d'espace passent dans la région opposée, propulsés par le champ électrique. La diode est polarisée dans le sens inverse, le courant inverse circule de la région n vers la région p. En raison des différences de densités entre les porteurs majoritaires et les porteurs minoritaires, le courant inverse est considérablement plus faible que le courant direct.

Pour calculer le courant traversant la jonction en fonction de la tension de polarisation, nous nous placerons dans un premier temps, dans l'hypothèse de faible niveau d'injection : dans chacune des régions, les porteurs majoritaires à l'équilibre thermodynamique restent majoritaires lorsque la jonction est polarisée. En d'autres termes la densité de porteurs majoritaires dans chacune des régions n'est pas affectée par la polarisation.

2.2.1. *Distribution des porteurs*

a. Densités de porteurs aux limites de la zone de charge d'espace

À l'équilibre thermodynamique, l'égalité des courants de conduction et de diffusion est représentée par les équations (2-7). Lorsque la jonction est polarisée, le champ électrique à l'intérieur de la zone de charge d'espace est modifié, les courants d'électrons et de trous s'écrivent en notant le champ E' pour ne pas confondre avec l'énergie

$$j_n = \mu_n \left(neE' + kT \frac{dn}{dx} \right) \tag{2-33-a}$$

$$j_p = \mu_p \left(peE' - kT \frac{dp}{dx} \right) \tag{2-33-b}$$

En régime de faible injection, on peut montrer que les courants j_n et j_p sont beaucoup plus faibles que chacune de leurs composantes. Calculons par exemple l'ordre de grandeur du courant de diffusion de trous. Au niveau de la zone de charge d'espace d'épaisseur W on peut écrire le gradient de concentration des trous sous la forme approchée,

$$\frac{dp}{dx} \approx \frac{\Delta p}{\Delta x} = \frac{p_p - p_n}{W} \approx \frac{p_p}{W}$$

Le courant de diffusion des trous s'écrit alors

$$j_{pd} = kT\,\mu_p \frac{dp}{dx} \approx kT\,\mu_p \frac{p_p}{W} = \frac{kT}{e}\frac{\sigma_p}{W}$$

où σ_p est la conductivité de la région de type p de la jonction. Pour une jonction réalisée avec une région de type p telle que $\sigma_p = 100\ \Omega^{-1}cm^{-1}$ et $W = 1\ \mu m$ on obtient $j_{pd} = 10^4\ A\,cm^{-2}$, soit un courant de 1 A pour une jonction de 10^4 μm^2 de section. Ce courant est considérablement plus grand que le courant direct d'une diode de 10^4 μm^2 de section, qui est de l'ordre de quelques dizaines de mA.

On peut donc négliger les courants j_n et j_p devant chacune de leurs composantes, les équations (2-33) s'écrivent alors

$$ne\,E' + kT\frac{dn}{dx} = 0 \qquad \text{(2-34-a)}$$

$$pe\,E' - kT\frac{dp}{dx} = 0 \qquad \text{(2-34-b)}$$

Comme à l'équilibre thermodynamique, ces équations s'écrivent

$$\frac{dn}{n} = -\frac{e}{kT}E'dx = \frac{e}{kT}dV \qquad \text{(2-35-a)}$$

$$\frac{dp}{p} = \frac{e}{kT}E'dx = -\frac{e}{kT}dV \qquad \text{(2-35-b)}$$

La densité d'électrons en $x = x_n$ est N_d, pour calculer la densité d'électrons en $x = x_p$ on intègre l'équation (2-35-a) de x_n à x_p. De même pour calculer la densité de trous en $x = x_n$ on intègre de x_p à x_n. Considérons par exemple les électrons

$$Ln\frac{n(x_n)}{n(x_p)} = \frac{e}{kT}(V_n - V_p) \qquad \text{(2-36)}$$

En l'absence de polarisation, $V_n - V_p = V_d$, $n(x_n) = n_n$ et $n(x_p) = n_p$ l'expression (2-36) s'écrit

$$Ln\frac{n_n}{n_p} = \frac{eV_d}{kT} \qquad \text{(2-37)}$$

En présence d'une polarisation V, $V_n - V_p = V_d - V$, $n(x_n) = n_n$, $n(x_p) = n_p^{'}$ l'expression (2-36) s'écrit

$$Ln \frac{n_n}{n_p^{'}} = \frac{e(V_d - V)}{kT} \tag{2-38}$$

La différence membre à membre entre les expressions (2-37 et 38) donne

$$Ln \frac{n_p^{'}}{n_p} = \frac{eV}{kT} \qquad \Rightarrow \; n_p^{'} = n_p \, e^{eV/kT}$$

Avec une expression semblable pour les trous.

Ainsi en régime de faible injection, les densités de porteurs majoritaires ne sont pas affectées, les densités de porteurs minoritaires aux frontières de la zone de charge d'espace sont données par

$$n'(x = x_p) = n_p^{'} = n_p \, e^{eV/kT} = \frac{n_i^2}{N_a} e^{eV/kT} \tag{2-39-a}$$

$$p'(x = x_n) = p_n^{'} = p_n \, e^{eV/kT} = \frac{n_i^2}{N_d} e^{eV/kT} \tag{2-39-b}$$

Le produit *np* en chacun des points x_n et x_p s'écrit donc

$$n_p^{'} \, p_p^{'} = n_n^{'} \, p_n^{'} = n_i^2 \, e^{eV/kT} \tag{2-40}$$

– En polarisation directe, *V* est positif, on obtient $n_p^{'} > n_p$ et $p_n^{'} > p_n$, des électrons sont injectés dans la région de type p et des trous dans la région de type n.
– En polarisation inverse, $n_p^{'} < n_p$ et $p_n^{'} < p_n$. Des électrons sont extraits de la région de type p et des trous de la région de type n.

b. Distribution des porteurs dans les régions neutres

Les porteurs majoritaires ne sont pas affectés. Les distributions des porteurs minoritaires sont régies par les équations de continuité (1-220). Considérons les trous dans la région de type n, l'équation s'écrit

$$\frac{\partial p}{\partial t} = -\frac{1}{e} \frac{\partial j_p(x)}{\partial x} - \frac{p - p_n}{\tau_p} \tag{2-41}$$

où τ_p représente la durée de vie des trous dans la région de type n. Les régions neutres sont dopées et par suite relativement conductrices, de sorte que la tension

aux bornes de ces régions est négligeable. Le champ électrique dans ces régions est donc faible alors que parallèlement le gradient de la concentration de porteurs minoritaires y est important en raison du phénomène d'injection. Il en résulte que le courant $j_p(x)$ est essentiellement un courant de diffusion

$$j_p(x) = -e\,D_p \frac{\partial p}{\partial x} \tag{2-42}$$

L'équation (2-41) s'écrit alors

$$\frac{\partial p}{\partial t} = D_p \frac{\partial^2 p}{\partial x^2} - \frac{p - p_n}{\tau_p} \tag{2-43}$$

Si la polarisation de la jonction est constante, le régime est stationnaire de sorte que $\partial p/\partial t = 0$, l'équation s'écrit

$$\frac{d^2 p}{dx^2} - \frac{p - p_n}{L_p^2} = 0 \tag{2-44}$$

où $L_p = \sqrt{D_p \tau_p}$ est la *longueur de diffusion* des trous dans la région de type n de la jonction. La solution de l'équation (2-44) est de la forme générale

$$p - p_n = A\,e^{-x/L_p} + B\,e^{x/L_p} \tag{2-45}$$

Les constantes *A* et *B* sont déterminées par les conditions aux limites, d'une part en $x = x_n$ à la frontière de la zone de charge d'espace, et d'autre part en $x = x_c$ au contact avec l'électrode métallique à l'extrémité de la région de type n. En $x = x_n$ la densité de trous est donnée par l'expression (2-39-b). En $x = x_c$ en raison de l'importante perturbation créée dans le réseau cristallin par la soudure du contact métallique, la vitesse de recombinaison des porteurs excédentaires est très grande. En d'autres termes leur durée de vie est très faible, de sorte que la densité de trous en $x = x_c$ est la densité d'équilibre $p(x_c) = p_n = n_i^2 / N_d$. En portant ces deux conditions aux limites dans l'expression (2-45), on calcule les constantes *A* et *B*. La distribution de trous dans la région de type n, et par suite la distribution d'électrons dans la région de type p, sont données par

$$p - p_n = \frac{p_n}{sh(d_n / L_p)} \left(e^{eV/kT} - 1\right) sh\left((x_c - x)/L_p\right) \tag{2-46-a}$$

$$n - n_p = \frac{n_p}{sh(d_p / L_n)} \left(e^{eV/kT} - 1\right) sh\left((x - x_c')/L_n\right) \tag{2-46-b}$$

où $d_n = x_c - x_n$ et $d_p = x_p - x_c'$ représentent les longueurs des régions neutres de type n et de type p respectivement. Ces expressions se simplifient dans les cas

limites où les longueurs d_n et d_p sont beaucoup plus grandes ou beaucoup plus petites que les longueurs de diffusion des porteurs minoritaires correspondants.

– Pour $d_n >> L_p$ et $d_p >> L_n$ les expressions (2-46) se simplifient en développant les fonctions hyperboliques, on obtient

$$p - p_n \approx p_n \left(e^{eV/kT} - 1\right) e^{(x_n - x)/L_p} \qquad \text{(2-47-a)}$$

$$n - n_p \approx n_p \left(e^{eV/kT} - 1\right) e^{(x - x_p)/L_n} \qquad \text{(2-47-b)}$$

– Pour $d_n << L_p$ et $d_p << L_n$, les arguments des sinus hyperboliques sont très petits et on peut écrire $sh\,\varepsilon \approx \varepsilon$. Les expressions (2-46) s'écrivent alors

$$p - p_n = \frac{p_n}{d_n}\left(e^{eV/kT} - 1\right)(x_c - x) \qquad \text{(2-48-a)}$$

$$n - n_p = \frac{n_p}{d_p}\left(e^{eV/kT} - 1\right)(x - x_c^{'}) \qquad \text{(2-48-b)}$$

Les distributions des porteurs dans les régions neutres sont représentées sur la figure (2-4). Dans l'hypothèse des faibles injections que nous avons utilisée, les densités de porteurs majoritaires sont constantes. Les densités de porteurs minoritaires décroissent dans chacune des régions à partir de la zone de charge d'espace, suivant des lois qui sont fonction des rapports d_n / L_p et d_p / L_n. Ces rapports peuvent être différents d'une région à l'autre.

2.2.2. *Courants de porteurs minoritaires*

Les courants de trous dans la région de type n et d'électrons dans la région de type p sont des courants de diffusion. Ils sont par conséquent donnés par

$$j_p(x) = -e\,D_p \frac{dp}{dx} \qquad j_n(x) = e\,D_n \frac{dn}{dx}$$

En explicitant $n_p = n_i^2 / N_a$ *et* $p_n = n_i^2 / N_d$ dans les expressions (2-46, 47 et 48), on obtient par simple dérivation

– Pour d_n/L_p et d_p/L_n quelconques

$$j_p = \frac{e n_i^2 D_p}{N_d L_p sh(d_n / L_p)}\left(e^{eV/kT} - 1\right) ch\left((x_c - x)/L_p\right) \qquad \text{(2-49-a)}$$

$$j_n = \frac{en_i^2 D_n}{N_a L_n sh(d_p / L_n)} \left(e^{eV/kT} - 1\right) ch\left((x - x'_c)/L_n\right) \qquad \text{(2-49-b)}$$

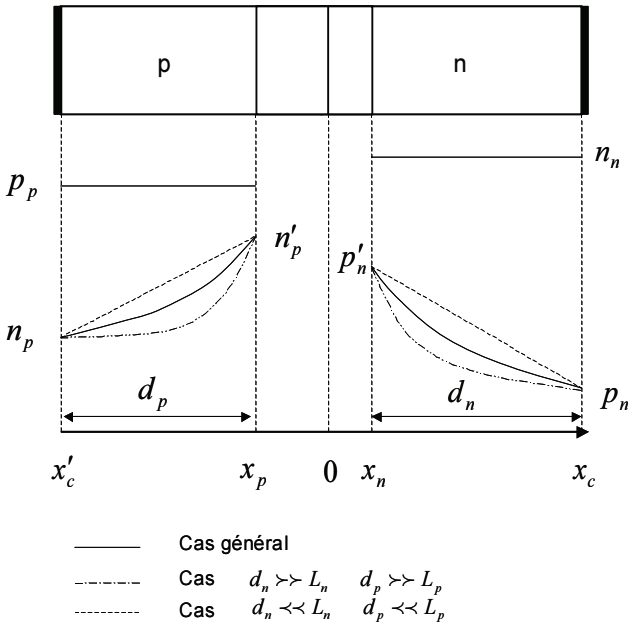


Figure 2-4 : Distribution des porteurs

– Pour $d_n >> L_p$ et $d_p >> L_n$

$$j_p = \frac{en_i^2 D_p}{N_d L_p} \left(e^{eV/kT} - 1\right) e^{(x_n - x)/L_p} \qquad \text{(2-50-a)}$$

$$j_n = \frac{en_i^2 D_n}{N_a L_n} \left(e^{eV/kT} - 1\right) e^{(x - x_p)/L_n} \qquad \text{(2-50-b)}$$

– Pour $d_n << L_p$ et $d_p << L_n$

$$j_p = \frac{en_i^2 D_p}{N_d d_n} \left(e^{eV/kT} - 1\right) \qquad \text{(2-51-a)}$$

$$j_n = \frac{en_i^2 D_n}{N_a d_p} \left(e^{eV/kT} - 1\right) \qquad \text{(2-51-b)}$$

Ces densités de courant sont représentées sur la figure (2-5). Dans le cas où la longueur d'une région est très inférieure à la longueur de diffusion des porteurs minoritaires dans cette région, la distribution linéaire des porteurs excédentaires

crée un courant de diffusion constant dans toute la région. Dans le cas contraire les courants de porteurs minoritaires ont disparu avant d'atteindre le contact correspondant.

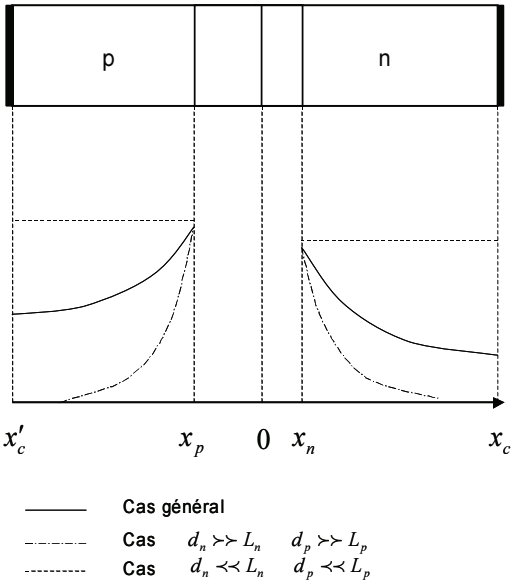


Figure 2-5 : Courants de porteurs minoritaires

2.2.3. *Densité de courant - caractéristique*

Les densités de courant d'électrons et de trous que nous avons calculées sont relatives à des régions différentes de la jonction. Le courant total traversant la jonction est la somme des courants d'électrons et de trous en un même point. Pour calculer ce courant ***supposons dans un premier temps que la zone de charge d'espace n'est le siège d'aucun mécanisme de génération-recombinaison***. Les courants j_n et j_p sont alors constants à la traversée de cette zone et on peut écrire les égalités

$$j_n(x_p) = j_n(x_n) \qquad \text{(2-52-a)}$$

$$j_p(x_n) = j_p(x_p) \qquad \text{(2-52-b)}$$

Les expressions établies au paragraphe précédent permettent de calculer $j_n(x_p)$ d'une part et $j_p(x_n)$ d'autre part. On peut donc compte tenu des relations (2-52) connaître les courants d'électrons et de trous en un même point, x_n ou x_p. Le courant total qui est conservatif, c'est-à-dire indépendant de *x.* Pour prouver

rigoureusement cette propriété, il suffit de tenir compte du fait que la divergence du courant total est nul puis de négliger le courant de déplacement (nul en régime établi). Appliquée à un volume de infinitésimal normal à la section de la jonction cette propriété permet de montrer que le courant de conduction est constant en fonction de *x*.

$$J = j_n(x_p) + j_p(x_p) = j_n(x_n) + j_p(x_n) = j_n(x_p) + j_p(x_n) \tag{2-53}$$

En faisant $x = x_n$ dans les expressions de $j_p(x)$, et $x = x_p$ dans les expressions de $j_n(x)$, on obtient

$$J = J_s \left(e^{eV/kT} - 1 \right) \tag{2-54}$$

avec

$$J_s = \frac{e n_i^2 D_p}{N_d L_p th(d_n / L_p)} + \frac{e n_i^2 D_n}{N_a L_n th(d_p / L_n)} \tag{2-55-a}$$

pour $d_i >> L_j$, $th(d_i / L_j) \approx 1$ $$J_s = \frac{e n_i^2 D_p}{N_d L_p} + \frac{e n_i^2 D_n}{N_a L_n} \tag{2-55-b}$$

pour $d_i << L_j$, $th(d_i / L_j) \approx d_i / L_j$ $$J_s = \frac{e n_i^2 D_p}{N_d d_n} + \frac{e n_i^2 D_n}{N_a d_p} \tag{2-55-c}$$

La caractéristique de la diode (Eq. 2-54) est représentée sur la figure (2-6-a).

On notera V_a la tension appliquée au lieu de V comme dans les relations précédentes. En polarisation inverse, $V_a < 0$, dès que $-V_a > kT/e$, l'exponentielle négative devient négligeable devant 1 et le courant atteint une valeur constante $J = -J_s$. J_s est appelé *courant de saturation* de la diode. En polarisation directe, $V_a > 0$. Dès que $V_a > kT/e$, l'exponentielle positive devient très supérieure à 1 et le courant s'écrit

$$J = J_s e^{eV_a/kT}$$

Lorsque V_a devient comparable à V_d, l'hypothèse de faible injection utilisée dans le calcul devient injustifiée car dans l'une ou l'autre des régions la densité de porteurs minoritaires injectés devient comparable à la densité de porteurs majoritaires. On peut montrer (Par 2.8), que le courant devient alors proportionnel à $e^{eV_a/2kT}$. Enfin lorsque V_a devient supérieur à V_d, la barrière de potentiel s'annule et le courant n'est limité que par les résistances série des régions neutres, le régime devient ohmique $J = r_s(V_a - V_d)$. r_s est la *résistance série* de la diode.

La figure (2-6-b) représente, pour une diode symétrique ($N_d = N_a$), la variation du potentiel dans toute la diode. Pour $V_a < V_d$ la variation du potentiel est localisée dans la zone de charge d'espace dont la largeur varie comme $(V_d - V_a)^{1/2}$ (courbe en pointillés), les zones neutres sont quasiment des équipotentielles, le courant est essentiellement de diffusion. Pour $V_a > V_d$, la barrière de potentiel est nulle à la fois en hauteur et en largeur, l'excès de tension $V_a - V_d$ est absorbé par la résistance série. Le courant est alors de conduction et le régime est ohmique. Notons que pour une jonction symétrique la variation de tension est plus faible dans la région de type n que dans la région de type p car sa résistivité est plus faible ($N_d = N_a$, $\mu_n >> \mu_p$). Enfin dans chacune des régions la variation de tension, linéaire loin de la jonction, n'est pas linéaire au voisinage de celle-ci car l'injection de porteurs minoritaires, assortie de l'augmentation de la densité de porteurs majoritaires assurant la neutralité électrique, augmente la conductivité du matériau sur une longueur de l'ordre de grandeur de la longueur de diffusion de ces porteurs.

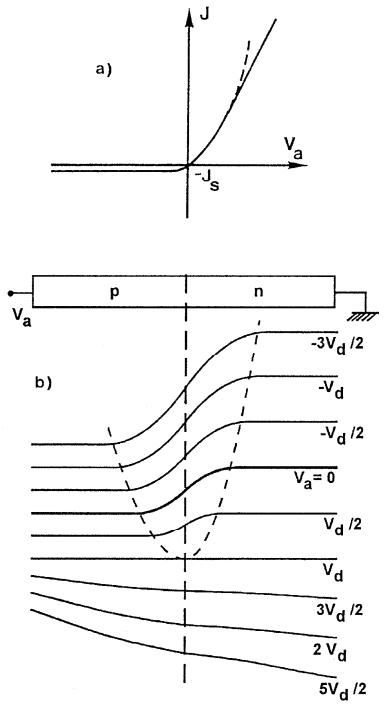


Figure 2.6 : a) caractéristique et b) distribution du potentiel

Nous avons considéré dans tous les calculs qui précèdent les courants de porteurs minoritaires. On peut en déduire facilement les courants de porteurs majoritaires en écrivant que le courant total est conservatif. La densité de courant total est indépendante de x et s'écrit

$$J = j_n(x) + j_p(x)$$

Dans la région de type n par exemple nous avons calculé $j_p(x)$, on en déduit $j_n(x)$, c'est-à-dire la densité de courant de porteurs majoritaires, par différence. Les densités de courant J, $j_n(x)$ et $j_p(x)$ sont représentées sur la figure (2-7).

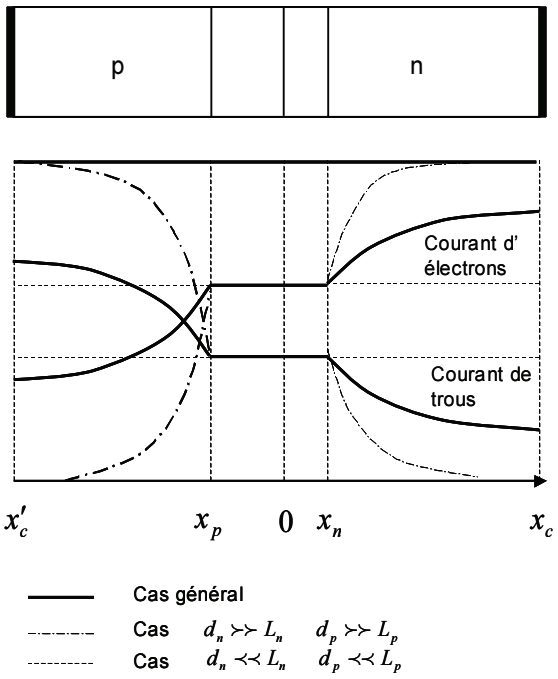


Figure 2-7 : Densités de courant

Le courant de porteurs majoritaires résulte de l'injection de porteurs minoritaires. Dans chacune des régions, lorsque les porteurs minoritaires sont injectés, un nombre égal de porteurs majoritaires est fourni par l'électrode métallique. Ces porteurs se déplacent dans la région considérée, vers les porteurs minoritaires, pour assurer la neutralité électrique. Ce déplacement est régi par la constante de temps diélectrique et est de ce fait pratiquement instantané, ~ 10^{-12} s (Par. 1.5.4.b). Les porteurs majoritaires et minoritaires se recombinent ensuite en un intervalle de temps régi par la durée de vie des porteurs minoritaires. Le régime est maintenu stationnaire par la polarisation constante qui crée une injection permanente de porteurs minoritaires.

2.3. Jonction à profil de dopage quelconque

Dans la jonction abrupte, les régions *n* et *p* sont dopées de manière homogène et le semiconducteur change brutalement de type en $x = 0$. Nous allons considérer le cas où le profil de dopage est quelconque, avec toutefois une variation suffisamment brutale à l'interface pour assurer la formation d'une zone de charge d'espace (Par. 1.5.4.a). Pour simplifier les calculs nous supposerons dans un premier temps que les longueurs d_n et d_p des régions de types *n* et *p* sont très faibles devant les longueurs de diffusion L_p et L_n des porteurs minoritaires.

2.3.1. *Régions n et p très courtes*

L'hypothèse $d_n << L_p$ et $d_p << L_n$ permet de supposer, comme on l'a vu dans le cas de la jonction abrupte, que les porteurs minoritaires traversent les régions neutres sans se recombiner et par suite que leurs distributions dans ces régions sont linéaires. En conséquence les courants j_n et j_p peuvent être supposés constants.

Considérons la région de type n de la jonction, l'hypothèse de faible injection implique $n(x) = N_d(x)$ et $p(x) << n(x)$. Les densités de courant dans cette région s'écrivent

$$j_p(x) = eD_p\left(p(x)\frac{e}{kT}E'(x) - \frac{dp(x)}{dx}\right) \qquad \text{(2-56-a)}$$

$$j_n(x) = eD_n\left(n(x)\frac{e}{kT}E'(x) + \frac{dn(x)}{dx}\right) \qquad \text{(2-56-b)}$$

En remplaçant $n(x)$ par $N_d(x)$ et en éliminant E' entre les deux équations, on obtient

$$\frac{j_p(x)}{eD_p} - \frac{p(x)}{n(x)}\,\frac{j_n(x)}{eD_n} = -\frac{p(x)}{N_d(x)}\,\frac{dN_d(x)}{dx} - \frac{dp(x)}{dx} \qquad \text{(2-57)}$$

Compte tenu de la relation $p(x) << n(x)$ on peut négliger le deuxième terme du premier membre de l'expression précédente et écrire

$$\frac{j_p(x)}{eD_p} = -\frac{p(x)}{N_d(x)}\,\frac{dN_d(x)}{dx} - \frac{dp(x)}{dx} \qquad \text{(2-58)}$$

Dans le cas particulier de la jonction abrupte, le dopage de chaque région est homogène et $dN_d(x)/dx = 0$. On retrouve la relation $j_p(x) = -eD_p\,dp(x)/dx$ et le courant de trous dans la région de type n est alors uniquement de diffusion. Dans le cas de la jonction quelconque, s'ajoute un courant de conduction associé au

champ électrique résultant de la distribution non homogène d'ions donneurs $N_d(x)$.

Dans la mesure où la condition $d_n << L_p$ est réalisée, le courant de trous dans la région de type n est constant, $j_p(x) = j_p$. L'équation (2-58) est alors une équation différentielle du premier ordre en $p(x)$.

$$\frac{dp(x)}{dx} + \frac{1}{N_d(x)} \frac{dN_d(x)}{dx} \quad p(x) = -\frac{j_p}{eD_p} \tag{2-59}$$

En posant $y(x) = p(x)$ et $B = -j_p / eD_p$, cette expression est de la forme

$$y'(x) + A(x)\, y(x) = B \qquad \text{avec} \quad A(x) = \frac{1}{N_d(x)} \frac{dN_d(x)}{dx} = \frac{d}{dx}\left(Ln N_d(x)\right)$$

La solution est de la forme

$$y = e^{-\int A(x)dx} \left(K + B \int e^{\int A(x)dx} dx \right)$$

Compte tenu de l'expression de $A(x)$, $e^{\int A(x)dx} = N_d(x)$, de sorte que la distribution de porteurs s'écrit

$$p(x) = \frac{1}{N_d(x)} \left(K - \frac{j_p}{eD_p} \int_{x_n}^{x} N_d(x')dx' \right) \tag{2-60}$$

où K est une constante d'intégration déterminée par les conditions aux limites.

Si on appelle respectivement $p(x_n)$ et $N_d(x_n)$ la densité de trous et le dopage en $x = x_n$ à la frontière de la zone de charge d'espace, l'expression (2-60) donne

$$p(x_n) = \frac{K}{N_d(x_n)} \qquad \Rightarrow \qquad K = p(x_n)\, N_d(x_n)$$

Si on appelle $p(x_c)$ et $N_d(x_c)$ la densité de trous et le dopage en $x = x_c$ au contact ohmique à l'extrémité de la région de type n, l'expression (2-60) s'écrit

$$p(x_c) = \frac{1}{N_d(x_c)} \left(p(x_n) N_d(x_n) - \frac{j_p}{eD_p} \int_{x_n}^{x_c} N_d(x)dx \right)$$

Il en résulte que le courant de trous s'écrit

$$j_p = eD_p \frac{p(x_n)N_d(x_n) - p(x_c)N_d(x_c)}{\int_{x_n}^{x_c} N_d(x)dx} \tag{2-61}$$

$p(x_n)$ est la densité de trous injectés en $x = x_n$ dans la région de type *n*. Cette densité est donnée par l'expression (2-39-b). Soit en explicitant $p_n = n_i^2 / N_d(x_n)$

$$p(x_n)\, N_d(x_n) = n_i^2 e^{eV/kT} \tag{2-62}$$

$p(x_c)$ est la densité de trous en $x = x_c$ au contact ohmique. ***En raison de l'alliage créé par la soudure du contact métallique, la durée de vie des porteurs excédentaires en ce point est très faible, de sorte que p(x_c) est égal à la densité de trous à l'équilibre*** $p(x_c) = n_i^2 / N_d(x_c)$ ***et par suite***

$$p(x_c)\, N_d(x_c) = n_i^2 \tag{2-63}$$

Compte tenu des relations (2-62 et 63), l'expression (2-61) s'écrit

$$j_p = \frac{en_i^2 D_p}{\int_{x_n}^{x_c} N_d(x)dx}\left(e^{eV/kT} - 1\right) \tag{2-64}$$

Le même calcul, pour le courant d'électrons dans la région de type p, donne

$$j_n = \frac{en_i^2 D_n}{\int_{x'_c}^{x_p} N_a(x)dx}\left(e^{eV/kT} - 1\right) \tag{2-65}$$

Les conditions $d_n << L_p$ et $d_p << L_n$ entraînent que les courants j_n et j_p sont indépendants de x. Si en outre on néglige les phénomènes de génération-recombinaison dans la zone de charge d'espace, ces courants sont constants dans toute la structure de sorte que le courant total est donné par $j_n + j_p$, soit

$$J = J_s\left(e^{eV/kT} - 1\right) \tag{2-66}$$

avec

$$J_s = \frac{en_i^2 D_p}{\int_{x_n}^{x_c} N_d(x)dx} + \frac{en_i^2 D_n}{\int_{x'_c}^{x_p} N_a(x)dx} \tag{2-67}$$

2.3.2. *Jonction quelconque*

À partir de l'expression (2-67) établie dans le cas d'une jonction à profil de dopage quelconque mais avec des régions n et p très courtes, et des expressions (2-55-c) puis (2-55-b) et (2-55-a), on peut établir, par analogie, l'expression du courant de saturation dans une jonction quelconque.

Dans les conditions $d_n << L_p$ et $d_p << L_n$, l'expression (2-67) donne le courant de saturation d'une jonction à dopage quelconque et l'expression (2-55-c) donne le courant de saturation de la jonction abrupte. Les produits $N_d\, d_n$ et $N_a\, d_p$ qui apparaissent dans l'expression (2-55-c) sont tout simplement les intégrales qui apparaissent dans l'expression (2-67), avec $N_d(x) = N_d$ et $N_a(x) = N_a$

Dans la jonction abrupte, on passe du cas $d_n << L_p$ et $d_p << L_n$ au cas $d_n >> L_p$ et $d_p >> L_n$ en remplaçant dans l'expression du courant de saturation, d_n par L_p et d_p par L_n. Par analogie, dans le cas de la jonction à dopage quelconque il suffit de modifier les bornes des intégrales qui deviennent respectivement x_n et $x_n + L_p$ au lieu de x_n d'une part et x_c, et x_p- L_n et x_p au lieu de x'_c et x_p d'autre part.

On peut enfin écrire l'expression du courant de saturation dans le cas d'une jonction quelconque par le même type d'analogie avec l'expression (2-55-a), soit

$$J_s = \frac{en_i^2 D_p}{\int_{x_n}^{x_n + L_p\, th(d_n / L_p)} N_d(x)dx} + \frac{en_i^2 D_n}{\int_{x_p - L_n\, th(d_p / L_n)}^{x_p} N_a(x)dx} \qquad (2\text{-}68)$$

Les cas limites correspondant aux valeurs extrêmes des rapports d_n / L_p et d_p / L_n s'obtiennent simplement à partir de l'expression (2-68) en développant la tangente hyperbolique

$$d >> L \;\Rightarrow\; th(d / L) \approx 1 \qquad d << L \;\Rightarrow\; th(d / L) \approx d / L$$

2.4. Capacités de la jonction pn

2.4.1. *Jonction polarisée en inverse - Capacité de transition*

Nous avons calculé la largeur de la zone de charge d'espace de la jonction en écrivant la continuité du potentiel en x=0, et établi l'expression (2-19)

$$V_n - V_p = \frac{e}{2\varepsilon}\left(N_d W_n^2 + N_a W_p^2\right)$$

Si la jonction n'est pas polarisée, $V_n - V_p = V_d$. Si la jonction est polarisée par une tension V, $V_n - V_p = V_d - V$ où V est positif pour une polarisation directe et négatif pour une polarisation inverse. Considérons une polarisation inverse et appelons V_i la valeur absolue de la tension appliquée. Dans ce cas $V_n - V_p = V_d + V_i$ et l'expression précédente s'écrit

$$V_d + V_i = \frac{e}{2\varepsilon}\left(N_d W_n^2 + N_a W_p^2\right) \tag{2-69}$$

Considérons le cas d'une jonction très dissymétrique avec par exemple $N_a >> N_d$, il en résulte d'après la relation (2-17) $W_p << W_n$ et par suite $W = W_p + W_n \approx W_n$. L'expression (2-69) s'écrit

$$V_d + V_i = \frac{eN_d}{2\varepsilon} W_n^2 \left(1 + \frac{N_d}{N_a}\right) \approx \frac{eN_d}{2\varepsilon} W_n^2 \tag{2-70}$$

et par suite

$$W_n = \left(\frac{2\varepsilon}{eN_d}\left(V_d + V_i\right)\right)^{1/2} \tag{2-71}$$

Une variation dV_i de la tension de polarisation V_i entraîne une variation dW_n de la largeur de la zone de charge d'espace dans la région de type n, et par suite une variation dQ de la charge d'espace développée dans cette région (Fig 2-8). Ces variations sont données par

$$dW_n = \frac{1}{2}\left(\frac{2\varepsilon}{eN_d}\right)^{1/2} \left(V_d + V_i\right)^{-1/2} dV_i \tag{2-72}$$

$$dQ = eN_d \, S \, dW_n \tag{2-73}$$

où S représente la section de la jonction. Compte tenu des expressions (2-72 et 71), l'expression (2-73) s'écrit

$$dQ = \frac{\varepsilon S}{W_n} dV_i \tag{2-74}$$

Ainsi une variation dV_i de la tension inverse produit une variation dQ de la charge d'espace, proportionnelle à la variation de tension. La jonction pn polarisée en inverse se comporte donc comme un condensateur de capacité dynamique

$$C_t = \varepsilon S / W_n \tag{2-75}$$

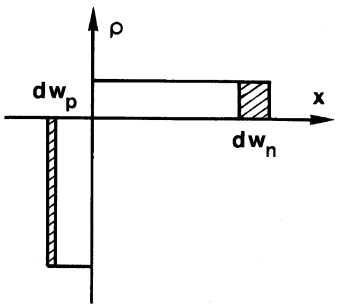


Figure 2-8

La capacité différentielle C_t est appelée *capacité de transition* de la jonction. En explicitant W_n à partir de l'expression (2-71), cette capacité s'écrit

$$C_t = S(\varepsilon e N_d/2)^{1/2} \Big/ \sqrt{V_d + V_i} \qquad (2\text{-}76)$$

Cette relation montre que la capacité de transition varie avec la tension inverse comme $1/\sqrt{V_d + V_i}$. Cette propriété de la jonction pn est mise à profit dans les *varactors*.

2.4.2. *Jonction en régime alternatif - Capacité de diffusion*

Considérons une jonction pn polarisée dans le sens direct par une tension statique V et soumise à un signal alternatif de faible amplitude v et de fréquence $f = \omega / 2\pi$. La tension appliquée à la jonction s'écrit

$$V(t) = V + v e^{j\omega t} \qquad (2\text{-}77)$$

La densité de trous injectés en x_n est toujours donnée par l'expression (2-39-b), et la diffusion de ces trous dans la région de type n est toujours régie par l'équation (2-43) mais avec ici $V(t)$ au lieu de V et $dp/dt \neq 0$.

Considérons tout d'abord la densité de trous injectés en $x = x_n$. Cette densité s'écrit

$$p(x_n,t) = p_n e^{eV(t)/kT} = p_n e^{e(V+ve^{j\omega t})/kT} \qquad (2\text{-}78)$$

Dans la mesure où l'on peut considérer $v << kT/e$, on peut développer la deuxième exponentielle en se limitant au terme du premier ordre

$$p(x_n,t) = p_n e^{eV/kT}\left(1 + \frac{ev}{kT} e^{j\omega t}\right) \qquad (2\text{-}79)$$

Le premier terme correspond à la composante continue, le deuxième à la composante alternative. Cette expression s'écrit

$$p(x_n,t) = P(x_n) + p(x_n)\, e^{j\omega t} \qquad (2\text{-}80)$$

avec

$$P(x_n) = p_n\, e^{eV/kT} = p_n'$$

$$p(x_n) = p_n\, e^{eV/kT}\ \frac{ev}{kT} = p_n'\ \frac{ev}{kT}$$

Considérons maintenant la diffusion des trous dans la région de type n. En $x = x_n$ la densité de trous est la somme de deux contributions, l'une continue l'autre alternative. Il en est de même pour tout x dans la région de type n

$$p(x,t) = P(x) + p(x)\, e^{j\omega t} \tag{2-81}$$

On calcule les dérivées $\partial p(x,t)/\partial t$ et $\partial^2 p(x,t)/\partial x^2$ et on les porte dans l'équation de diffusion (2-43). En séparant les termes constants des termes fonction du temps, on obtient alors les deux équations

$$\frac{d^2 P(x)}{dx^2} - \frac{P(x) - p_n}{L_p^2} = 0 \tag{2-82-a}$$

$$\frac{d^2 p(x)}{dx^2} - p(x)\frac{1 + j\omega\tau_p}{L_p^2} = 0 \tag{2-82-b}$$

L'équation (2-82-a) est analogue à l'équation (2-44) dont la solution, dans le cas particulier où $d_n >> L_p$ par exemple, est donnée par l'expression (2-47-b) soit

$$P(x) - p_n = p_n \left(e^{eV/kT} - 1\right) e^{(x_n - x)/L_p} \tag{2-83}$$

C'est la réponse statique à la polarisation continue V.

En ce qui concerne l'équation (2-82-b) on peut l'écrire

$$\frac{d^2 p(x)}{dx^2} - \gamma_p^2\, p(x) = 0 \tag{2-84}$$

avec

$$\gamma_p^2 = (1 + j\omega\tau_p) \big/ L_p^2 \tag{2-85}$$

La solution est de la forme générale

$$p(x) = A\, e^{-\gamma_p x} + B\, e^{\gamma_p x}$$

En supposant $d_n >> L_p$, les conditions aux limites sont analogues à celles du régime statique

$$\text{en } x = x_n \qquad p(x_n) = p_n e^{eV/kT} \frac{ev}{kT}$$
$$\text{en } x \to \infty \qquad p(x \to \infty) = 0$$

La deuxième condition donne B=0, la première permet alors de calculer A. L'expression de la composante alternative de la densité de porteurs s'écrit

$$p(x) = p_n e^{eV/kT} \frac{ev}{kT} e^{\gamma_p (x_n - x)} \tag{2-86}$$

En portant les expressions (2-83 et 86) dans l'expression (2-81) on obtient l'expression de la densité de trous en un point x de la région de type n en fonction du temps

$$p(x,t) = p_n \left(1 + \left(e^{eV/kT} - 1 \right) e^{(x_n - x)/L_p} + e^{eV/kT} \frac{ev}{kT} e^{\gamma_p (x_n - x)} e^{j\omega t} \right) \tag{2-87}$$

Le courant de trous dans la région de type n est un courant de diffusion, il est donné par $j_p(x,t) = -e D_p \partial p(x,t) / \partial x$. Ainsi l'amplitude de la composante alternative de ce courant est donnée par

$$j_{p\approx}(x) = -e D_p \frac{dp(x)}{dx}$$

À partir de l'expression (2-86) et avec une expression analogue pour les électrons dans la région de type p, on obtient

$$j_{p\approx} = eD_p \gamma_p p_n e^{eV/kT} \frac{ev}{kT} e^{\gamma_p (x_n - x)} \tag{2-88-a}$$

$$j_{n\approx} = eD_n \gamma_n n_p e^{eV/kT} \frac{ev}{kT} e^{\gamma_n (x - x_p)} \tag{2-88-b}$$

avec

$$\gamma_i^2 = (1 + j\omega\tau_i) / L_i^2 \qquad i = n, p \tag{2-89}$$

Pour obtenir le courant alternatif total on ajoute, comme en régime statique, $j_{p\approx}(x_n)$ *et* $j_{n\approx}(x_p)$ soit

$$j_{\approx} = e\left(p_n D_p \gamma_p + n_p D_n \gamma_n \right) \frac{ev}{kT} e^{eV/kT} \tag{2-90}$$

L'admittance de la jonction en courant alternatif est donnée par le rapport $y = j_{\approx} / v$

$$y=\frac{e^2}{kT}e^{eV/kT}\left(p_n D_p \gamma_p + n_p D_n \gamma_n\right) \tag{2-91}$$

- Dans le domaine des fréquences relativement basses telles que $\omega\tau<<1$, les expressions de γ_p et γ_n peuvent être simplifiées en développant la racine,

$$\gamma_i = (1+ j\omega\tau_i/2)/L_i \qquad i=n,p$$

Ce qui permet d'écrire l'admittance y sous la forme

$$y=\frac{e^2}{kT}e^{eV/kT}\left(\frac{p_n D_p}{L_p}+\frac{n_p D_n}{L_n}+\frac{1}{2}j(p_n L_p + n_p L_n)\omega\right) \tag{2-92}$$

L'admittance de la jonction est donc la somme d'une conductance et d'une capacité

$$y = g_d + j C_d \omega$$

avec

$$g_d=\frac{e^2}{kT}\left(\frac{p_n D_p}{L_p}+\frac{n_p D_n}{L_n}\right)e^{eV/kT} \tag{2-93-a}$$

$$C_d=\frac{e^2}{2kT}\left(p_n L_p + n_p L_n\right)e^{eV/kT} \tag{2-93-b}$$

g_d est la *conductance de diffusion*, C_d est la *capacité de diffusion*.

Compte tenu de l'expression du courant direct de la jonction qui s'écrit pour $V >> kT/e$, $J = J_s\, e^{eV/kT}$ où J_s est donné par l'expression (2-55-b), la conductance g_d s'écrit

$$g_d = \frac{e}{kT} J \tag{2-94}$$

La conductance en courant alternatif est donc proportionnelle au courant direct de polarisation.

Considérons la capacité de diffusion, si la jonction est fortement dissymétrique, avec par exemple $N_a >> N_d$, l'expression (2-93-b) se réduit à

$$C_d \approx \frac{e^2}{2kT}\, p_n L_p\, e^{eV/kT} = \frac{e}{2kT}\ \frac{e\, p_n D_p}{L_p}\ \tau_p\, e^{eV/kT}$$

soit

$$C_d \approx \frac{e}{2kT}\,\tau_p\,J \tag{2-95}$$

La capacité de diffusion est proportionnelle au courant direct de polarisation et à la durée de vie des porteurs minoritaires. Cette capacité n'a pas un sens physique comme la capacité de transition mais traduit simplement le déphasage, c'est-à-dire le retard apporté au signal, par les phénomènes de diffusion.

- Dans le domaine des hautes fréquences telles que $\omega\tau >> 1$, les expressions de γ_n et γ_p se simplifient sous une autre forme

$$\gamma_i^2 = j\frac{\omega\,\tau_i}{L_i^2} = j\,\frac{\omega}{D_i} = \frac{\omega}{D_i}e^{j\pi/2} \qquad i = n, p$$

de sorte que

$$\gamma_i = \sqrt{\omega/D_i}\;e^{j\pi/4} = 1/\sqrt{2}\cdot\sqrt{\omega/D_i}\cdot(1+j) \tag{2-96}$$

L'admittance *y* s'écrit alors sous la forme

$$y = \frac{e^2}{kT}e^{eV/kT}\sqrt{\omega/2}\cdot\left(p_n\sqrt{D_p} + n_p\sqrt{D_n}\right)(1+j) \tag{2-97}$$

Cette admittance est la somme d'une conductance et d'une capacité, toutes deux fonction de la fréquence

$$y = g_d(\omega) + j\,C_d(\omega)\,\omega \tag{2-98}$$

avec

$$g_d(\omega) = \frac{e^2}{kT\sqrt{2}}\left(p_n\sqrt{D_p} + n_p\sqrt{D_n}\right)e^{eV/kT}\sqrt{\omega} \tag{2-99-a}$$

$$C_d(\omega) = \frac{e^2}{kT\sqrt{2}}\left(p_n\sqrt{D_p} + n_p\sqrt{D_n}\right)e^{eV/kT}\frac{1}{\sqrt{\omega}} \tag{2-99-b}$$

2.5. Mécanismes de génération-recombinaison dans la zone de charge d'espace

Nous avons calculé le courant traversant la jonction en négligeant les générations-recombinaisons dans la zone de charge d'espace et en écrivant les égalités $j_n(x_n) = j_n(x_p)$ et $j_p(x_p) = j_p(x_n)$. En fait, la zone de charge d'espace de

la jonction est le siège de générations thermiques et de recombinaisons, qui rendent injustifiées les égalités précédentes.

La densité de porteurs dans la zone de charge d'espace étant faible, le taux de recombinaison est donné par la formule de Shockley-Read (1-213)

$$r = \frac{1}{\tau_m} \frac{pn - n_i^2}{2n_i + p + n}$$

Lorsque la jonction est polarisée par une tension V le produit np est donné par l'expression (2-40), $np = n_i^2 \, e^{eV/kT}$, de sorte que le taux de recombinaison s'écrit

$$r = \frac{1}{\tau_m} \frac{n_i^2 \left(e^{eV/kT} - 1 \right)}{2n_i + p + n} \qquad (2\text{-}100)$$

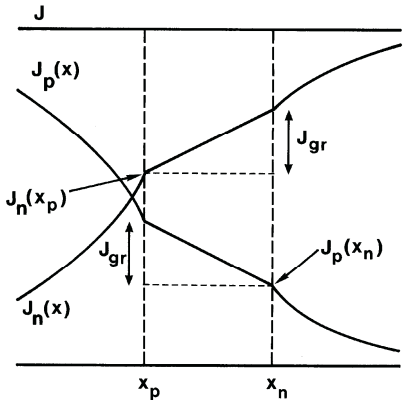


Figure 2-9 : Courants compte tenu des générations recombinaisons

En polarisation directe la tension V est positive, l'exponentielle est donc supérieure à 1, et par suite r est positif ; il y a recombinaison de porteurs dans la zone de charge d'espace. Cette recombinaison diminue le courant direct de la jonction.

En polarisation inverse la tension V est négative et par suite r est négatif ; il y a génération thermique de porteurs dans la zone de charge d'espace. Cette génération augmente le courant inverse de la jonction.

Ainsi, en raison des phénomènes de génération-recombinaison , $j_n(x)$, $j_p(x)$ ne sont pas constants à la traversée de la zone de charge d'espace et le courant total s'écrit (Fig. 2-9)

$$\begin{aligned} J &= j_n(x_n) + j_p(x_n) \equiv j_n(x_p) + j_p(x_p) \\ &= \left(j_n(x_p) + j_{gr}\right) + j_p(x_n) \equiv j_n(x_p) + \left(j_p(x_n) + j_{gr}\right) \\ &= j_n(x_p) + j_p(x_n) + j_{gr} \end{aligned}$$

soit

$$J = J_s\left(e^{eV/kT} - 1\right) + j_{gr} \tag{2-101}$$

La composante j_{gr} est positive ou négative suivant que le taux de recombinaison r est positif ou négatif.

Dans la mesure où il n'y a pas d'excitation externe autre que la polarisation (g=0), et où le régime est stationnaire $\left(\partial n / \partial t = 0\right)$, l'équation de continuité des électrons (1-239-a) s'écrit

$$0 = \frac{1}{e}\frac{dj_n(x)}{dx} - r \qquad \Rightarrow \qquad dj_n(x) = er\,dx$$

On obtiendrait la même équation pour le courant de trous. Ainsi la variation des courants d'électrons et de trous à la traversée de la zone de charge d'espace, appelée *courant de génération-recombinaison* s'écrit

$$j_{gr} = \Delta j_n = \Delta j_p = e\int_{x_p}^{x_n} r\,dx \tag{2-102}$$

2.5.1. *Polarisation inverse - Courant de génération*

Considérons une polarisation inverse, c'est-à-dire $V < 0$ avec $-V >> kT/e$, l'expression du taux de recombinaison (2-100) se simplifie dans la mesure où d'une part le terme exponentiel est négligeable devant 1, et où d'autre part les densités de porteurs n et p sont très inférieures à n_i, soit

$$r \approx -\frac{n_i}{2\tau_m} \tag{2-103}$$

r est négatif et correspond à une génération thermique de porteurs. En l'explicitant dans l'expression (2-102), le *courant de génération* s'écrit

$$j_{gr} = -\frac{en_i}{2\tau_m}W \tag{2-104}$$

L'expression (2-101) s'écrit alors, compte tenu de la condition $-V >> kT/e$

$$J_i = -J_s + j_{gr} = -J_s - \frac{en_i}{2\tau_m} W \qquad (2\text{-}105)$$

Le courant de génération s'ajoute au courant inverse. Remarquons que j_{gr} est, par l'intermédiaire de W, fonction de la tension appliquée. Dans une jonction p^+n par exemple, $W \approx W_n$ est donné par l'expression (2-71) où $V_i = |V|$. Le courant inverse de la jonction s'écrit donc

$$J_i = -J_s - J_g \sqrt{1 + V_i / V_d} \quad avec\ J_g = \frac{n_i}{\tau_m} \left(\frac{e\varepsilon}{2N_d} V_d \right)^{1/2} \qquad (2\text{-}106)$$

Il faut noter que J_s étant très faible, c'est j_{gr} qui constitue l'essentiel du courant inverse de la diode. Ce dernier varie par conséquent comme $\sqrt{V_d + V_i}$.

2.5.2. *Polarisation directe - Courant de recombinaison*

En polarisation directe la tension V est positive, un grand nombre de porteurs traversent la zone de charge d'espace et il n'est plus possible de simplifier l'expression du taux de recombinaison. De plus, r varie beaucoup à l'intérieur de la zone de charge d'espace. En particulier r passe par un maximum au point où la somme $n + p$ est minimum. Le produit np étant constant, la somme $n + p$ est minimum au point où $n = p = n_i\, e^{eV/2kT}$. En ce point le taux de recombinaison est maximum et donné par

$$r_m = \frac{1}{\tau_m} \frac{n_i^2 \left(e^{eV/kT} - 1 \right)}{2n_i \left(1 + e^{eV/2kT} \right)} \qquad (2\text{-}107)$$

Avec la condition $V >> kT/e$, l'expression précédente se simplifie

$$r_m \approx \frac{n_i}{2\tau_m}\, e^{eV/2kT} \qquad (2\text{-}108)$$

On peut ainsi définir une *largeur effective* W_{eff} de la zone de charge d'espace (Fig. 2-10), par la relation

$$\int_{x_p}^{x_n} r\, dx = r_m W_{eff} \qquad (2\text{-}109)$$

Le *courant de recombinaison* s'écrit donc

$$j_{gr} = e r_m W_{eff} = \frac{en_i}{2\tau_m} W_{eff}\, e^{eV/2kT}$$

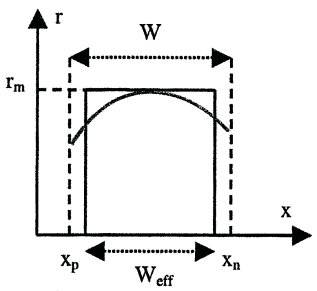


Figure 2-10 : Polarisation directe et largeur effective

Notons que W_{eff}, qui est une fraction de la largeur W de la zone de charge d'espace, varie avec la tension de polarisation V. En fait cette variation, qui est en $\sqrt{V}$, est négligeable devant la variation exponentielle.

Le courant direct à travers la jonction s'écrit alors

$$J_d = J_s\, e^{eV/kT} + J_r\, e^{eV/2kT} \quad \text{avec} \quad J_r = \frac{en_i}{2\tau_m} W_{eff} \tag{2-110}$$

2.6. Jonction en régime transitoire - Temps de recouvrement

Dans une jonction polarisée par une tension statique V, le courant J traversant la jonction est constant, les distributions des porteurs minoritaires dans chacune des régions de la jonction sont données par les équations (2-46, 47 et 48).

Considérons, pour simplifier le raisonnement, le cas d'une jonction p^+n dans laquelle la longueur de la région de type n est beaucoup plus grande que la longueur de diffusion des trous $d_n >> L_p$. La distribution de trous dans la région de type n est alors donnée par l'expression (2-47-a). En outre, en raison de la nature p^+n de la jonction, on peut négliger l'injection d'électrons et attribuer le courant au niveau de la zone de charge d'espace à la seule injection de trous $J \approx j_p(x_n)$.

Lorsque la jonction est polarisée dans le sens direct par une tension statique V^+, telle que $V^+ >> kT/e$, le courant traversant la jonction est J_d, la distribution de trous dans la région de type n s'écrit (Eq. 2-47-a).

$$p(x) \approx p_n + p_n e^{eV^+/kT} e^{-(x-x_n)/L_p} \tag{2-111}$$

Lorsque la jonction est polarisée dans le sens bloqué par une tension statique V^-, telle que $-V^- >> kT/e$, le courant traversant la jonction est $-J_s$, la distribution de trous dans la région de type n s'écrit

$$p(x) \approx p_n - p_n e^{-(x-x_n)/L_p} \tag{2-112}$$

Les distributions de trous dans chacun des régimes de polarisation sont représentées sur la figure (2-11). L'expression (2-112) montre en particulier que $p(x_n) = 0$ en polarisation inverse.

Supposons maintenant que la tension de polarisation passe brutalement de $V = V^+$ à $V = V^-$ à l'instant t=0. Le courant doit évoluer de $J = J_d$ à $J = -J_s$ et la densité de trous doit évoluer de la distribution (a) à la distribution (b) (Fig. 2-12). Cette évolution est conditionnée par la recombinaison et par la diffusion des porteurs minoritaires que sont les trous dans la région de type n. Ces processus ne sont pas instantanés de sorte que le basculement de la jonction du régime direct au régime inverse, c'est-à-dire de l'état passant à l'état bloqué, nécessite un intervalle de temps fini appelé *temps de recouvrement* de la jonction.

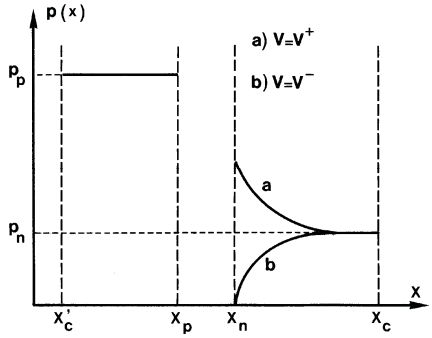


Figure 2-11 : Jonction p+n et distribution des trous en polarisation directe (a) et inverse (b).

À l'instant t=0 où la tension de polarisation change de signe en passant de V^+ à V^-, le courant change aussi de signe et passe de $J = J_d$ à une valeur négative $J = -J_i$. Mais l'amplitude du courant inverse J_i est proportionnelle à la densité de porteurs minoritaires. Ainsi à l'instant t=0 où la tension de polarisation est inversée, dans la mesure où la densité de porteurs minoritaires dans la région de type n est alors représentée par la courbe (a) sur la figure (2-12), le courant inverse est très important, $J_i >> J_s$. Ce courant évolue vers J_s au fur et à mesure que la distribution des trous évolue de la courbe (a) à la courbe (b).

La densité de trous en $x = x_n$ est donnée en fonction de la tension appliquée aux bornes de la zone de charge d'espace V_j par l'expression

$$p(x_n) = p_n e^{eV_j / kT} \tag{2-113}$$

En régime statique, $V_j = V^-$ en polarisation inverse et $V_j \approx V_d$ (à kT/e près), en polarisation directe. Mais pendant le régime transitoire, la tension V_j évolue avec la densité de trous en x=x_n. Compte tenu de (2-113), cette évolution est donnée par

$$V_j(t) = \frac{kT}{e} Ln \frac{p(x_n,t)}{p_n} \tag{2-114}$$

Tant que la condition $p(x_n,t) > p_n$ est réalisée la tension $V_j(t)$ varie peu, puisqu'elle évolue de V_d à zéro. Dans cette première phase, que nous limiterons à l'instant $t = t_1$ où $p(x_n,t_1) = p_n$, la tension V^- appliquée à la structure se retrouve donc en presque totalité aux bornes de la résistance série r_s de la diode. Le courant est alors donné par $J = V^- / r_s = Cte$. Il faut alors remarquer que, r_s étant petit, ce courant peut atteindre des valeurs très importantes s'il n'est pas limité par une résistance extérieure dans le circuit de polarisation. Ainsi pendant l'intervalle de temps t_1, la tension aux bornes de la zone de charge d'espace de la jonction reste

sensiblement constante, de l'ordre de *kT/e*, et le courant inverse qui s'établit instantanément à travers la jonction reste constant à une valeur importante $-J_i$. L'intervalle de temps t_1 est appelé *temps de stockage* de la diode. Quand t augmente au-delà de t_1, $p(x_n,t)$ devient inférieur à p_n, la tension V_j devient négative et s'établit progressivement à V^-. Le courant inverse diminue et s'établit progressivement à $-J_s$. Les variations de la tension V_j et du courant J sont représentées sur la figure (2-12).

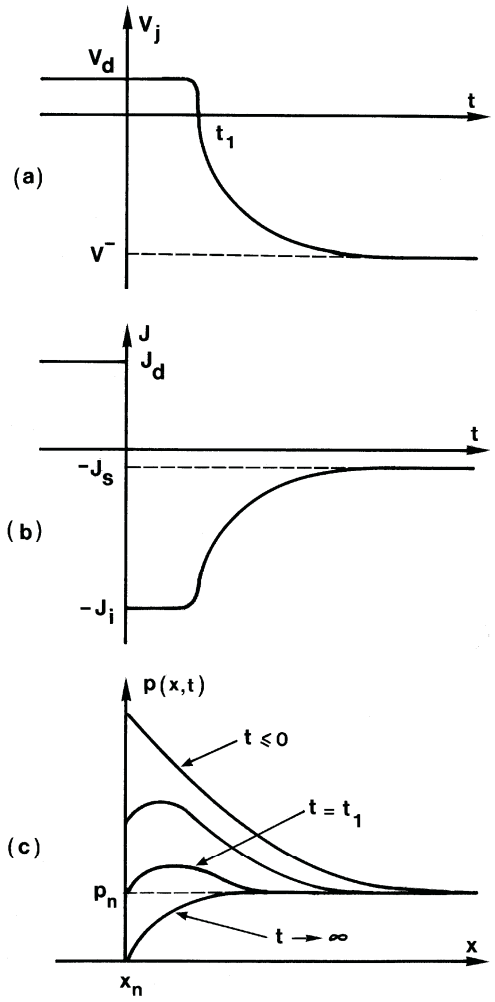


Figure 2-12 : Évolution dans le temps des distributions de trous. a) Tension aux bornes de la zone de charge d'espace. b) Courant à travers la jonction. c) Distribution des trous dans la région de type n

La densité de trous dans la région de type n, $p(x,t)$ évolue de la distribution de régime direct $t<0$, à la distribution de régime inverse $t\to\infty$. Les allures des courbes de distribution sont représentées sur la figure (2-12-c). Le régime inverse est établi au bout d'un temps τ_r appelé *temps de recouvrement* de la diode. On obtient l'évolution dans le temps de la distribution de trous en intégrant l'équation de continuité. Dans la mesure où le courant de trous dans la région de type n est un courant de diffusion, cette équation s'écrit

$$\frac{\partial p(x,t)}{\partial t} = D_p \frac{\partial^2 p(x,t)}{\partial x^2} - \frac{p(x,t)-p_n}{\tau_p} \tag{2-115}$$

Les conditions aux limites sont données par les courbes (a) et (b) de la figure (2-10), qui représentent respectivement $p(x,0)$ et $p(x,\infty)$.

Dans l'utilisation de la diode en régime transitoire, comme interrupteur rapide en particulier, le paramètre caractéristique important est le temps de recouvrement τ_r. On peut en obtenir un ordre de grandeur sans intégrer l'équation (2-115).

Pendant le régime transitoire, le courant qui traverse la diode est dû à la diffusion et à la recombinaison des trous excédentaires de la région de type n. Si on appelle $Q_p(t)$ la densité superficielle de charge due à ces trous, le courant s'écrit

$$J(t) = d\,Q_p(t)/dt$$

et par suite

$$Q_p(t=\infty) - Q_p(t=0) = \int_0^\infty J(t)dt \tag{2-116}$$

En raison de la nature des phénomènes qui régissent la disparition de ces trous, diffusion et recombinaisons, le courant $J(t)$ présente une variation que l'on peut représenter par une loi exponentielle. Compte tenu de la figure (2-11-b), on peut l'écrire sous la forme

$$\begin{aligned} &0\le t\le t_1 \quad J(t) = -J_i \\ &t\ge t_1 \quad J(t) = -J_i e^{-(t-t_1)/\tau} \end{aligned} \tag{2-117}$$

$Q_p(t=0)$ représente la densité superficielle de charge associée à l'excès de trous existant dans la région de type n en polarisation directe, on posera $Q_p(t=0) = Q_{pd}$. $Q_p(t=\infty)$ représente la densité superficielle de charge associée au défaut de trous existant en polarisation inverse. Il est évident que l'on peut négliger ce deuxième terme devant le premier. L'équation (2-116) s'écrit alors compte tenu de (2-117)

$$Q_{pd} = J_i\left(t_1 + \int_{t_1}^{\infty} e^{-(t-t_1)/\tau}\, dt\right) = J_i(t_1 + \tau) \tag{2-118}$$

La densité superficielle de charge Q_{pd} en régime de polarisation directe est d'autre part donnée par

$$Q_{pd} = e\int_{x_n}^{x_c} \left(p(x,0) - p_n\right) dx \tag{2-119}$$

Considérons une jonction de type p^+n, le courant total qui traverse la jonction est donné par $J = j_p(x_n) + j_n(x_p) \approx j_p(x_n)$. En polarisation directe en particulier

$$J_d = j_p(x_n) \tag{2-120}$$

Nous allons calculer Q_{pd} dans les deux cas limites correspondant respectivement à $d_n << L_p$ et $d_n >> L_p$.

– Pour $d_n << L_p$, la distribution et le courant de trous dans la région de type n sont donnés par les expressions (2-48-a et 51-a)

$$p(x) - p_n = p_n\left(e^{eV/kT} - 1\right)\frac{x_c - x}{d_n} \tag{2-121}$$

$$j_p(x) = j_p = \frac{en_i^2 D_p}{N_d d_n}\left(e^{eV/kT} - 1\right) \tag{2-122}$$

Ainsi à partir des expressions (2-120, 121 et 122), on peut écrire l'expression de la distribution $p(x,0)$ des trous à $t=0$, qui correspond à la distribution existant en polarisation directe

$$p(x,0) - p_n = \frac{J_d}{e\, D_p}(x_c - x) \tag{2-123}$$

En portant cette expression dans l'intégrale (2-119), on obtient

$$Q_{dp} = \frac{J_p}{2\, D_p} d_n^2 \tag{2-124}$$

- Pour $d_n >> L_p$, la distribution et le courant de trous sont donnés par les expressions (2-47-a et 50-a)

$$p(x) - p_n = p_n\left(e^{eV/kT} - 1\right)e^{-(x - x_n)/L_p} \tag{2-125}$$

$$j_p(x)=\frac{en_i^2 D_p}{N_d L_p}\left(e^{eV/kT}-1\right)e^{-(x-x_n)/L_p} \qquad (2\text{-}126)$$

Compte tenu des expressions (2-125, 126 et 120), la distribution de trous $p(x,0)$ à t=0 s'écrit

$$p(x,0)-p_n=J_d\frac{L_p}{eD_p}e^{-(x-x_n)/L_p} \qquad (2\text{-}127)$$

En portant cette expression dans l'intégrale (2-119) on obtient

$$Q_{dp}=J_d\frac{L_p^2}{D_p}\left(1-e^{-d_n/L_p}\right) \qquad (2\text{-}128)$$

Dans la mesure où la condition $d_n >> L_p$ est réalisée et où $L_p^2 = D_p\,\tau_p$, cette expression s'écrit

$$Q_{dp} = J_d\,\tau_p \qquad (2\text{-}129)$$

Si on appelle $\tau_r = t_1 + \tau$ le temps de recouvrement, on obtient à partir de l'expression (2-118) et des expressions (2-124 et 129)

$$\tau_r=\frac{J_d}{J_i}\frac{d_n^2}{2D_p} \qquad pour\ \ d_n << L_p$$

$$\tau_r=\frac{J_d}{J_i}\tau_p \qquad pour\ \ d_n >> L_p$$

Dans l'expression $L_p^2 = D_p\tau_p$, la durée de vie τ_p représente en quelque sorte le temps de transit des trous sur une longueur égale à la longueur de diffusion L_p . Par analogie, on peut écrire $d_n^2 = D_p\tau_t$ où τ_t représente le temps de transit des trous sur une longueur égale à la longueur de la région de type n de la diode. Dans chacun des cas les forces provoquant le transit des trous sont les forces de diffusion.
On peut donc écrire le temps de recouvrement τ_r sous la forme

$$\tau_r=\frac{J_d}{J_i}\tau_{ro} \qquad (2\text{-}130)$$

où τ_{r0} représente le temps de recouvrement dans la diode pour $J_i = J_d$

$$\tau_{ro} = \tau_t/2 \qquad pour\ \ d_n << L_p$$

$$\tau_{ro} = \tau_p \qquad pour\ \ d_n >> L_p$$

Dans le premier cas, le temps de recouvrement de la diode est indépendant de la durée de vie des porteurs minoritaires et proportionnel au temps de transit des porteurs à travers la région neutre de la diode. Ceci résulte simplement du fait que dans ce cas tous les porteurs se recombinent au contact métallique, le temps de recouvrement correspond donc au temps mis pour atteindre ce contact. Dans le deuxième cas, le temps de recouvrement est proportionnel à la durée de vie des porteurs minoritaires. Pour le minimiser il faut donc diminuer au maximum cette durée de vie. On y parvient en introduisant dans le semiconducteur des centres de recombinaison, comme par exemple l'or dans le silicium. Toutefois, on ne peut augmenter inconsidérément le taux de centres de recombinaison sans augmenter fortement le courant de génération de la jonction et par suite le courant inverse de la diode. La durée de vie des porteurs minoritaires est beaucoup plus faible dans les semiconducteurs à gap direct que dans les semiconducteurs à gap indirect, en raison des processus de recombinaison radiatifs. Il en résulte que le GaAs par exemple est mieux adapté que le silicium à la réalisation de diodes rapides.

2.7. Claquage de la jonction polarisée en inverse

Lorsque la jonction est polarisée en inverse par une tension de module V_i , la largeur de la zone de charge d'espace augmente comme $\sqrt{V_i}$ alors que la tension à ses bornes augmente comme V_i . Il en résulte que le champ électrique à l'intérieur de cette zone augmente. Dans le cas de la jonction abrupte dissymétrique p^+n par exemple, les expressions (2-16 et 71) montrent que le champ est maximum à la jonction métallurgique, en $x = 0$, et varie comme

$$E_o = \frac{eN_d}{\varepsilon} W_n \quad avec\ W_n = \left(\frac{2\varepsilon}{eN_d} (V_d + V_i) \right)^{1/2}$$

Ce qui donne pour $V_i >> V_d$

$$E_o = (2\, e\, N_d / \varepsilon)^{1/2} \sqrt{V_i} \tag{2-131}$$

La tension V_i ne peut augmenter indéfiniment car il existe une limite à la valeur du champ électrique E_0 . En effet, lorsque le champ électrique augmente, la force électrique $\mathbf{F} = -e\mathbf{E}_o$ exercée sur les électrons liés augmente. Lorsque cette force est supérieure à la force de liaison des électrons de valence sur les noyaux, ces électrons sont libérés, le matériau devient conducteur et la tension V_i n'augmente plus. En d'autres termes, le champ électrique maximum que l'on peut établir dans un cristal semiconducteur est celui qui provoque l'excitation directe d'un électron lié de la bande de valence vers un état libre de la bande de conduction, c'est-à-dire l'ionisation du matériau. Dans le silicium ce champ maximum est de l'ordre de 10^6 V/cm.

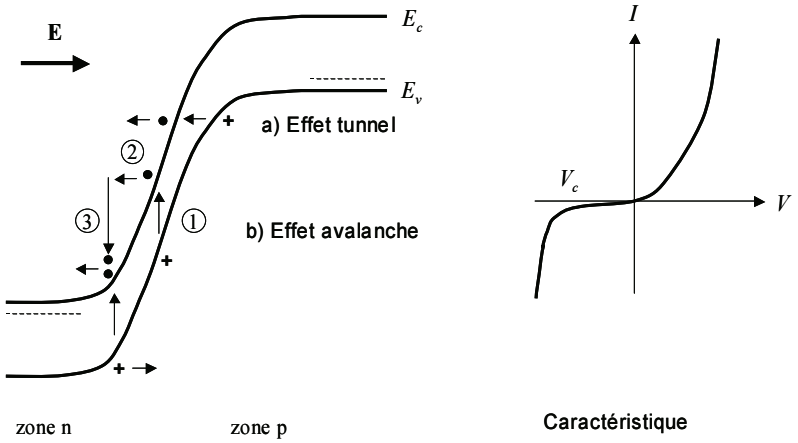


Figure 2-13 : Claquage par avalanche

Dans la jonction pn la tension limite correspondante est fonction de la largeur de la zone de charge d'espace et par suite des dopages du semiconducteur. L'expression (2-131) donne $V_i \approx 3$ V pour $N_d = 10^{18}$ cm^{-3}. Notons que la largeur de la zone de charge d'espace est alors de l'ordre de 700 Å (Eq. 2-71). La tension limite est appelée *tension Zener* V_z , l'effet de claquage de la jonction qui en résulte est appelé *effet Zener*. La paire électron-trou créée par l'ionisation est évacuée par le champ électrique et donne naissance à un courant inverse important. Le processus est représenté sur la figure (2-13-a). L'électron est directement émis, par effet tunnel à travers la zone de charge d'espace, de la bande de valence de la région de type p vers la bande de conduction de la région de type n. La tension inverse ne peut augmenter au-delà de la tension de claquage $V_c = V_z$, la caractéristique $I(V)$ est représentée sur la figure (2-13-b).

Dans la pratique cet effet n'est observable que dans les jonctions pn fortement dopées, dans lesquelles la zone de charge d'espace est très étroite et de l'ordre de 500 Å. Lorsque la zone de charge d'espace n'est pas particulièrement étroite, $W > 1000$ Å, un autre phénomène entraîne le claquage de la jonction, avant l'apparition de l'effet Zener, c'est *l'effet d'avalanche*.

Dès que le champ électrique est de l'ordre de 10^5 V/cm, c'est-à-dire pour une valeur environ dix fois inférieure au seuil d'effet Zener, l'accélération acquise par les quelques porteurs, essentiellement d'origine thermique, qui transportent le courant inverse, est suffisante pour leur permettre de générer des paires électron-trou par ionisation par choc des atomes du cristal. Ces paires sont à leur tour accélérées et peuvent créer d'autres paires, il en résulte un processus en chaîne tout à fait comparable à celui qui régit la décharge dans les gaz, c'est *l'effet d'avalanche*. Ce processus est représenté sur la figure (2-13-a) : la phase 1 correspond à la création thermique d'une paire électron-trou. Dans la phase 2 l'électron est accéléré par le champ électrique et se trouve de ce fait de plus en

plus haut dans la bande de conduction, il devient un porteur chaud. La phase 3 correspond au moment où son énergie cinétique est suffisante pour créer par choc une autre paire électron-trou, à l'issue du choc l'électron se retrouve au bas de la bande de conduction et une deuxième paire électron-trou est créée. Le processus peut se poursuivre si la largeur de la zone de charge d'espace le permet. Le processus identique existe pour le trou. La caractéristique $I(V)$ qui résulte de ce phénomène d'avalanche est en tout point semblable à la caractérisation $I(V)$ qui résulte de l'effet Zener (Fig. 2-13-b). La tension de claquage V_c a simplement une origine différente.

Le processus d'avalanche nécessite non seulement un champ élevé mais aussi une distance suffisante permettant l'accélération des porteurs, c'est la raison pour laquelle dans les jonctions très dopées l'effet Zener précède l'effet d'avalanche, car la zone de charge d'espace n'a pas une largeur suffisante, par contre dans la plupart des jonctions l'effet d'avalanche précède l'effet Zener car son champ de seuil est plus faible. Les deux effets peuvent être distingués expérimentalement par la variation avec la température de la tension de claquage V_c. Lorsque le claquage résulte de l'effet Zener la tension de claquage présente un coefficient de température négatif, au contraire lorsque le claquage résulte de l'effet d'avalanche le coefficient de température de V_c est positif. D'une manière générale, les tensions de claquage inférieures à 5 volts résultent de l'effet Zener, les tensions de claquage supérieures à 7 volts résultent de l'effet d'avalanche. Pour des tensions intermédiaires les deux effets se combinent. Ces phénomènes sont mis à profit dans les diodes régulatrices de tension.

2.8. Jonction en régime de forte injection

L'étude que nous avons développée de la jonction pn polarisée dans le sens direct a été faite dans l'hypothèse dite de *faible niveau d'injection* dont le fondement est de supposer que l'injection de porteurs minoritaires dans chacune des régions n'affecte pas la densité de porteurs majoritaires. Cette hypothèse, qui reste justifiée quand la tension de polarisation est nettement inférieure à la tension de diffusion, doit faire l'objet d'une attention toute particulière lorsque la polarisation devient importante.

Rappelons qu'une tension de polarisation V_a de la diode se distribue entre la zone de charge d'espace et les régions neutres, $V_a=V_j+V_s$. Tant que V_a est inférieur à la tension de diffusion V_d on peut écrire $V_j=V_a$, le courant est conditionné par la tension aux bornes de la zone de charge d'espace. Lorsque V_a devient égal puis supérieur à V_d la barrière de potentiel n'existe plus, le courant est alors conditionné par la tension V_s aux bornes des régions neutres de la diode.

Nous allons étudier successivement comment varient d'une part l'injection de porteurs minoritaires et d'autre part le courant, lorsque la tension de polarisation V_a atteint et dépasse la tension de diffusion V_d.

2.8.1. *Injection des porteurs*

Les rapports de populations de porteurs aux frontières de la zone de charge d'espace sont donnés par l'expression (2-36). Rappelons que cette équation vient du fait que l'on considère que les composantes de diffusion et de dérive sont presque égales et très supérieures au courant total somme de ces deux composantes.

$$Ln\frac{n(x_n)}{n(x_p)} = Ln\frac{p(x_p)}{p(x_n)} = \frac{e}{kT}(V_n - V_p) \tag{2-132}$$

En l'absence de polarisation

$$V_n - V_p = V_d$$

$$n(x_n) = n_n = N_d \qquad n(x_p) = n_p = n_i^2 / N_a$$

$$p(x_n) = p_n = n_i^2 / N_d \qquad p(x_p) = p_p = N_a$$

L'équation (2-132) s'écrit

$$Ln\frac{n_n}{n_p} = Ln\frac{p_p}{p_n} = Ln\frac{N_d N_a}{n_i^2} = \frac{e}{kT}V_d \tag{2-133}$$

En présence d'une polarisation V_a de la diode, la zone de charge d'espace de la jonction est polarisée par la tension $V_j = V_a - V_s$. Ainsi

$$V_n - V_p = V_d - V_j$$

$$n(x_n) = n_n + \Delta n_n \qquad n(x_p) = n_p + \Delta n_p$$

$$p(x_n) = p_n + \Delta p_n \qquad p(x_p) = p_p + \Delta p_p$$

L'équation (2-132) s'écrit

$$Ln\frac{n_n + \Delta n_n}{n_p + \Delta n_p} = Ln\frac{p_p + \Delta p_p}{p_n + \Delta p_n} = \frac{e}{kT}(V_d - V_j) \tag{2-133}$$

La condition de neutralité électrique dans les régions conductrices fait que l'injection de porteurs minoritaires entraîne une augmentation équivalente de la densité de porteurs majoritaires

$$\Delta n_n = \Delta p_n \qquad \Delta p_p = \Delta n_p \tag{2-135}$$

La relation (2-210) s'écrit alors en explicitant les densités de porteurs minoritaires injectés

$$Ln\frac{n_n+\Delta p_n}{n_p+\Delta n_p}=Ln\frac{p_p+\Delta n_p}{p_n+\Delta p_n}=\frac{e}{kT}(V_d-V_j) \qquad (2\text{-}136)$$

En posant $\alpha=e(V_d-V_j)/kT$ les densités de porteurs minoritaires injectés dans chacune des régions de la jonction sont ainsi régies par le système de deux équations suivant

$$p_p+\Delta n_p=(p_n+\Delta p_n)e^{\alpha} \qquad (2\text{-}137\text{-a})$$
$$n_p+\Delta n_p=(n_n+\Delta p_n)e^{-\alpha} \qquad (2\text{-}137\text{-b})$$

La résolution de ce système donne

$$\Delta n_p=\frac{n_n-p_n+p_pe^{-\alpha}-n_pe^{\alpha}}{e^{\alpha}-e^{-\alpha}} \qquad (2\text{-}138\text{-a})$$
$$\Delta p_n=\frac{p_p-n_p+n_ne^{-\alpha}-p_ne^{\alpha}}{e^{\alpha}-e^{-\alpha}} \qquad (2\text{-}138\text{-b})$$

En explicitant d'une part le paramètre α et d'autre part les densités de porteurs, ces expressions s'écrivent

$$\Delta n_p=\frac{n_i^2}{N_a}(e^{eV_j/kT}-1)\,\frac{1+\frac{N_a}{N_d}e^{-e(V_d-V_j)/kT}}{1-e^{-2e(V_d-V_j)/kT}} \qquad (2\text{-}139\text{-a})$$

$$\Delta p_n=\frac{n_i^2}{N_d}(e^{eV_j/kT}-1)\,\frac{1+\frac{N_d}{N_a}e^{-e(V_d-V_j)/kT}}{1-e^{-2e(V_d-V_j)/kT}} \qquad (2\text{-}139\text{-b})$$

Chacune des expressions précédentes est le produit de deux termes. Le premier correspond au régime de faible injection, le second est un facteur correctif qui tend vers 1 en régime de faible injection. L'analyse du terme correctif montre qu'il n'est vraiment différent de 1 que lorsque V_d-V_j devient inférieur à kT/e (26 mV à la température ambiante). En effet, ce terme qui est maximum pour une jonction symétrique $(N_a=N_d)$ prend la valeur $(1+e^{-1})/(1-e^{-2})\approx 1{,}6$ pour $V_d-V_j=kT/e$. Pour une jonction dissymétrique p^+n, avec $N_a/N_d=100$ par exemple, le terme correctif qui entre dans l'expression de Δp_n est d'environ 1,2 . Δn_p est dans ce cas négligeable. Il faut noter que les expressions (2-139) divergent pour $V_j=V_d$ car le dénominateur tend vers zéro. Ceci résulte de la condition de neutralité électrique qui se comporte ici comme une boucle de réaction positive : l'injection de porteurs minoritaires dans une région est proportionnelle à la densité de porteurs majoritaires dans l'autre, mais cette dernière augmente au gré de l'injection de porteurs minoritaires afin d'assurer la neutralité électrique. En fait cette divergence n'existe pas car le système est alors régulé par la chute de tension dans les régions neutres de la diode de sorte que la tension V_j reste toujours inférieure à la tension V_d. En raison

de l'énergie thermique des électrons la barrière effective s'annule non pas pour $V_j = V_d$ mais pour $V_j \approx V_d - kT/e$.

Il résulte de ce qui précède, qu'en régime de forte injection les densités de porteurs minoritaires aux frontières de la zone de charge d'espace restent données par des expressions semblables aux expressions (2-39) à la condition de bien préciser que V_j représente la tension appliquée aux bornes de la zone de charge d'espace.

$$\Delta n_p \approx \frac{n_i^2}{N_a}(e^{eV_j/kT} - 1) \qquad \text{(2-140-a)}$$

$$\Delta p_n \approx \frac{n_i^2}{N_d}(e^{eV_j/kT} - 1) \qquad \text{(2-140-b)}$$

2.8.2. *Densité de courant*

Considérons par exemple le cas d'une jonction de type p^+n, l'injection d'électrons est négligeable de sorte que le courant au niveau de la jonction est essentiellement un courant de trous. On peut alors écrire les deux équations

$$j_n(x_p) = j_n(x_n) = (ne\mu_n E + eD_n \frac{dn}{dx})_{x_n} \approx 0 \qquad \text{(2-141)}$$

$$J = j_n(x_p) + j_p(x_n) \approx j_p(x_n) = (pe\mu_p E - eD_p \frac{dp}{dx})_{x_n} \qquad \text{(2-142)}$$

L'équation (2-141) permet d'expliciter le champ électrique $E(x_n)$ qui, compte tenu de la condition de neutralité électrique qui entraîne $dn/dx = dp/dx$, s'écrit

$$E(x_n) = -\frac{D_n}{\mu_n}\frac{1}{n}\frac{dp}{dx}\Big|_{x_n} \qquad \text{(2-143)}$$

En tenant compte de la relation $D_n/\mu_n = D_p/\mu_p$, la densité de courant (2-142) s'écrit alors

$$J = -eD_p^* \frac{dp}{dx}\Big|_{x_n} \qquad \text{(2-144)}$$

où $D_p^* = D_p(1 + p/n)_{x_n}$ représente un *coefficient de diffusion effectif*. La quantité dp/dx en $x = x_n$ est obtenue à partir de l'équation de continuité des trous (1-220-b) qui s'écrit ici avec $j_p = J$ donné par l'expression (2-144)

$$D_p^* \frac{d^2p}{dx^2} - \frac{\Delta p}{\tau_p^*} = 0 \qquad \text{(2-145)}$$

où τ_p^* est la durée de vie des trous dans la région de type n en régime de forte injection. L'équation (2-145) s'écrit

$$\frac{d^2\Delta p}{dx^2} - \frac{\Delta p}{L_p^{*2}} = 0 \qquad (2\text{-}146)$$

où $L_p^* = \sqrt{D_p^* \tau_p^*}$ représente la longueur de diffusion effective des trous dans la région de type n.

L'équation (2-146) est analogue à l'équation (2-44) établie en régime de faible inversion, les solutions sont semblables c'est-à-dire fonction du rapport de la longueur d_n de la région de type n à la longueur de diffusion effective L_p^* des trous dans cette région.

Pour $d_n >> L_p^*$ la solution est une exponentielle de la forme

$$\Delta p = \Delta p(x_n) e^{-(x-x_n)/L_p^*} \qquad (2\text{-}147\text{-a})$$

et par suite

$$\frac{d\Delta p}{dx}\Big|_{x_n} = -\frac{\Delta p(x_n)}{L_p^*} = -\frac{\Delta p_n}{L_p^*} \qquad (2\text{-}147\text{-b})$$

Pour $d_n << L_p^*$ la solution est linéaire de la forme

$$\Delta p = \Delta p(x_n) \frac{x_c - x}{d_n} \qquad (2\text{-}148\text{-a})$$

et par suite

$$\frac{d\Delta p}{dx}\Big|_{x_n} = -\frac{\Delta p(x_n)}{d_n} = -\frac{\Delta p_n}{d_n} \qquad (2\text{-}148\text{-b})$$

De manière générale, quel que soit le rapport d_n / L_p^*, $d\Delta p / dx$ en $x = x_n$ est de la forme

$$\frac{d\Delta p}{dx}\Big|_{x_n} = -\frac{\Delta p_n}{L_p^* th(d_n / L_p^*)} \qquad (2\text{-}149)$$

En explicitant Δp_n à partir de l'expression (2-140-b) et en portant dans l'expression (2-144) on obtient l'expression du courant en régime de forte injection

$$J = J_s^*(e^{eV_j/kT} - 1) \qquad (2\text{-}150\text{-a})$$

avec

$$J_s^* = \frac{en_i^2 D_p^*}{N_d L_p^* th(d_n / L_p^*)} \qquad (2\text{-}150\text{-b})$$

Les expressions (2-150) sont respectivement analogues aux expressions (2-54 et 55-a) établies dans l'hypothèse de faible injection. Nous n'avions alors fait aucune hypothèse sur une éventuelle dissymétrie de la diode, d'où l'existence de deux composantes dans l'expression (2-55-a). D'autre part nous avions appelé V la tension V_j dans le but d'alléger l'écriture.

Afin de comparer plus facilement les résultats obtenus dans chacun des deux régimes, les expressions (2-54 et 55-a) sont rappelées ci-dessous pour une jonction dissymétrique p^+n en rétablissant l'écriture V_j définissant la tension appliquée aux bornes de la zone de charge d'espace

$$J = J_s(e^{eV_j/kT} - 1) \qquad (2\text{-}151\text{-a})$$

$$J_s = \frac{en_i^2 D_p}{N_d L_p th(d_n / L_p)} \qquad (2\text{-}151\text{-b})$$

En régime de faible injection, $p(x_n) << n(x_n) = N_d$, le rapport *p/n* en $x = x_n$ est très inférieur à 1 de sorte $D_p^* = D_p$, en outre $\tau_p^* = \tau_p$ et par suite $L_p^* = L_p$, les expressions (2-150) se ramènent respectivement aux expressions (2-151).

En régime de forte injection, $p(x_n) > N_d$, le rapport *p/n* en $x = x_n$ tend vers 1 en raison de la condition de neutralité électrique. Ceci entraîne $D_p^* \approx 2D_p$, $\tau_p^* \approx \tau_p N_d / n \approx \tau_p N_d / N_a$ et par suite $L_p^* \approx L_p \sqrt{2N_d / N_a}$. La durée de vie des trous diminue car la densité d'électrons qui régit leur taux de recombinaison passe de $n = N_d$ à faible injection à $n \approx p$ à forte injection et tend vers $n \approx p \approx N_a$ en régime de très forte injection. Il en résulte que le courant de saturation devient $J_s^* \approx 2J_s$ si $d_n << L_p^*$ et $J_s^* \approx J_s \sqrt{2N_a / N_d}$ si $d_n >> L_p^*$.

La différence entre les deux régimes de fonctionnement n'est pas uniquement une augmentation d'amplitude à travers J_s^*. À cette augmentation s'ajoute une modification de la loi de variation avec la tension de polarisation V_a de la diode. En régime de faible injection la tension appliquée à la diode V_a se retrouve presque intégralement aux bornes de la zone de charge d'espace de sorte que $V_a = V_j + V_s \approx V_j$ et les lois $J(V_j)$ et $J(V_a)$ sont identiques. La caractéristique de la diode est donnée par l'expression (2-151-a) dans laquelle $V_j = V_a$. Soit

$$J = J_s(e^{eV_a/kT} - 1) \approx J_s e^{eV_a/kT} \qquad (2\text{-}152)$$

En régime de forte injection on ne peut plus négliger V_s devant V_j de sorte que les lois $J(V_j)$ et $J(V_a)$ sont différentes. L'expression (2-150-a) traduit la relation qui lie le courant à la tension appliquée à la zone de charge d'espace de la diode mais ne

traduit plus la caractéristique $J(V_a)$ de la diode. Pour atteindre cette dernière, il faut expliciter la relation qui lie V_j, V_s et V_a.

2.8.3. *Chute de tension dans les régions neutres de la diode*

Nous allons maintenant calculer la chute de tension dans les régions neutres de la diode, $V_s = (V_{x_c'} - V_{x_p}) + (V_{x_n} - V_{x_c}) = V_{sp} + V_{sn}$ (Fig. 2-4).

La jonction étant de type p^+n l'injection d'électrons dans la région de type p est négligeable de sorte que la conductivité de cette région est constante et par suite la chute de tension V_{sp} est purement ohmique. Elle s'écrit simplement

$$V_{sp} = r_{sp} J \tag{2-153}$$

avec $r_{sp} = d_p / \sigma_p$. Le paramètre r_{sp} représente la résistance par unité de surface de la région de type p, σ_p sa conductivité et d_p sa longueur. En explicitant σ_p, la chute ohmique dans la région de type p s'écrit donc

$$V_{sp} = J d_p / p_p e \mu_p = J d_p / N_a e \mu_p \tag{2-154}$$

Dans la région de type n l'injection d'une densité de trous comparable puis supérieure à la densité d'électrons à l'équilibre, modifie considérablement les distributions de porteurs de sorte que la chute de tension V_{sn} n'est pas purement ohmique. Les courants d'électrons et de trous dans cette région s'écrivent

$$j_n = ne\mu_n E + eD_n \frac{dn}{dx} \tag{2-155-a}$$

$$j_p = pe\mu_p E - eD_p \frac{dp}{dx} \tag{2-155-b}$$

La condition de neutralité électrique entraîne $n = N_d + p$ et $dn/dx = dp/dx$, de sorte que compte tenu de la relation $D/\mu = kT/e$, le courant total s'écrit

$$J = j_n + j_p = e\left(p\mu_p + (N_d + p)\mu_n\right)E + kT(\mu_n - \mu_p)\frac{dp}{dx} \tag{2-156}$$

Le champ électrique présent dans cette région, à la suite de l'injection des porteurs, s'écrit donc

$$E = \frac{J}{e}\frac{1}{p\mu_p + (N_d + p)\mu_n} - \frac{kT}{e}\frac{\mu_n - \mu_p}{p\mu_p + (N_d + p)\mu_n}\frac{dp}{dx} \tag{2-157}$$

La chute de potentiel dans cette région, de longueur $d_n = x_c - x_n$, s'écrit donc

$$V_{sn} = V_{xn} - V_{xc} = \int_{x_n}^{x_c} E dx$$

$$V_{sn} = \frac{J}{e}\int_{x_n}^{x_c} \frac{dx}{p\mu_p + (N_d + p)\mu_n} - \frac{kT}{e}(\mu_n - \mu_p)\int_{x_n}^{x_c} \frac{dp/dx}{p\mu_p + (N_d + p)\mu_n} dx \qquad (2\text{-}158)$$

La chute de tension V_{sn} apparaît donc comme la somme algébrique de deux composantes. La première qui est proportionnelle au courant J peut-être qualifiée de *résistive*, même si le coefficient de proportionnalité n'est pas constant. La seconde qui s'annule si $\mu_n = \mu_p$, résulte de la mobilité différentielle des électrons et des trous, on la qualifiera de *diffusive.*

La composante résistive représente tout simplement la chute de tension dans la résistance série de la région de type n. Cette composante de la forme $r_{sn}J$, est l'analogue pour cette région, de l'expression (2-153) établie pour la région de type p. La différence entre r_{sn} et r_{sp} résulte de la forte injection de trous qui rend r_{sn} fonction du courant alors que r_{sp} est constant. La résistance r_{sn} est donnée par

$$r_{sn} = \frac{1}{e}\int_{x_n}^{x_c} \frac{dx}{p\mu_p + (N_d + p)\mu_n} \qquad (2\text{-}159)$$

L'expression (2-158) s'écrit

$$V_{sn} = r_{sn}J - \frac{kT}{e}(\mu_n - \mu_p)\int_{x_n}^{x_c} \frac{dp/dx}{p\mu_p + (N_d + p)\mu_n} dx \qquad (2\text{-}160)$$

Un changement de variable permet d'intégrer le deuxième terme sur la variable p, l'expression s'écrit alors

$$V_{sn} = r_{sn}J - \frac{kT}{e}\frac{\mu_n - \mu_p}{\mu_n + \mu_p}\int_{p(x_n)}^{p(x_c)} \frac{dp}{p + N_d\mu_n/(\mu_n + \mu_p)} \qquad (2\text{-}161)$$

Compte tenu du fait que la densité de trous au contact x_c est la densité d'équilibre $p(x_c) = p_n << N_d$, l'intégration de l'expression (2-161) donne

$$V_{sn} = r_{sn}J + \frac{kT}{e}\frac{\mu_n - \mu_p}{\mu_n + \mu_p} Ln\left(\frac{p(x_n)(\mu_n + \mu_p) + N_d\mu_n}{N_d\mu_n}\right) \qquad (2\text{-}162)$$

En régime de faible injection $p(x_n) << N_d$, l'argument du logarithme tend vers 1 de sorte que le deuxième terme s'annule. D'autre part, la densité de trous dans la région de type n est alors partout très inférieure à la densité d'électrons $n = N_d$ et la résistance r_{sn} prend la valeur r_{sno} donnée par

$$r_{sno} = \frac{1}{e}\int_{x_n}^{x_c} \frac{dx}{N_d \mu_n} = \frac{d_n}{N_d e \mu_n} = \frac{d_n}{\sigma_n} \tag{2-163}$$

où d_n et σ_n sont respectivement la longueur et la conductivité au repos de la région de type n. r_{sno} , analogue à r_{sp}, représente la résistance série de la région de type n au repos. La chute de tension dans les régions neutres de la diode est alors purement ohmique et s'écrit simplement

$$V_s = V_{sp} + V_{sn} = (r_{sp} + r_{sno})J = r_s J \tag{2-164}$$

En régime de forte injection, $p(x_n) > N_d$, la chute de tension V_{sn} est conditionnée par le rapport d_n / L_p.

Si la longueur d_n de la région de type n est beaucoup plus grande que la longueur de diffusion L_p des trous dans cette région, l'injection de trous ne perturbe cette région que sur une faible fraction de sa longueur et la chute de tension reste essentiellement conditionnée par la chute ohmique due à r_{sno} . L'expression (2-164) constitue alors une bonne approximation.

Si par contre la longueur de la région de type n est très faible, la chute de tension V_{sn} n'est plus du tout ohmique. La recombinaison des trous injectés dans la région de type n se produit alors essentiellement au contact x_c, de sorte que la distribution des trous est linéaire et le courant de trous sensiblement constant. En d'autres termes, on peut écrire les égalités

$$j_p = j_p(x_n) \qquad dp/dx = (dp/dx)_{x_n}$$

Le courant traversant la diode s'écrit alors (Eq 2-144 avec $p/n = 1$)

$$J = j_p(x_n) = j_p = -2eD_p \frac{dp}{dx}\Big|_{x_n} = -2eD_p \frac{dp}{dx}$$

ou

$$J = -2kT\mu_p \frac{dp}{dx} \tag{2-165}$$

La composante résistive de la chute de tension V_{sn} s'écrit donc en explicitant r_{sn} (Eq. 2-159) et J (Eq. 2-165)

$$r_{sn} J = -2\frac{kT}{e}\mu_p \int_{x_n}^{x_c} \frac{dp/dx}{p\mu_p + (N_d + p)\mu_n} dx$$

L'intégrale sur x a déjà été calculée dans l'équation (2-161), on obtient ici

$$r_{sn}J = \frac{kT}{e}\frac{2\mu_p}{\mu_n+\mu_p}Ln\left(\frac{p(x_n)(\mu_n+\mu_p)+N_d\mu_n}{N_d\mu_n}\right) \tag{2-166}$$

La chute de tension V_{sn} s'écrit donc

$$V_{sn} = \left(\frac{2\mu_p}{\mu_n+\mu_p}+\frac{\mu_n-\mu_p}{\mu_n+\mu_p}\right)\frac{kT}{e}Ln\left(\frac{p(x_n)(\mu_n+\mu_p)+N_d\mu_n}{N_d\mu_n}\right) \tag{2-167}$$

La somme des deux termes présents dans les premières parenthèses de l'équation (2-I67) est évidemment égale à 1, mais chacun d'eux pondère la contribution des effets *résistifs* et *diffusifs*. Si $\mu_n = \mu_p$, le deuxième terme est nul et le premier est égal à 1, la chute de tension V_{sn} est d'origine exclusivement résistive. Si par contre $\mu_n >> \mu_p$, le premier terme est négligeable et le second voisin de 1, la chute de tension est d'origine diffusive. Dans le silicium par exemple avec $\mu_n \approx 1200\ cm^2V^{-1}s^{-1}$ et $\mu_p \approx 300\ cm^2V^{-1}s^{-1}$, la chute de tension V_{sn} est environ 40% résistive et 60% diffusive.

L'expression (2-167) s'écrit

$$V_{sn} = \frac{kT}{e}Ln\left(\frac{p(x_n)(\mu_n+\mu_p)+N_d\mu_n}{N_d\mu_n}\right) \tag{2-168}$$

En régime de forte injection, l'inégalité $p(x_n)(\mu_n+\mu_p) >> N_d\mu_n$ permet d'écrire V_{sn} sous la forme

$$V_{sn} = \frac{kT}{e}Ln\left(\frac{\mu_n+\mu_p}{\mu_n}\frac{p(x_n)}{N_d}\right) \tag{2-169}$$

En explicitant $p(x_n) = \Delta p_n$ donné par l'expression (2-140-b) on obtient

$$V_{sn} = \frac{kT}{e}Ln\left(\frac{\mu_n+\mu_p}{\mu_n}\frac{n_i^2}{N_d^2}e^{eV_j/kT}\right) \tag{2-170}$$

que l'on peut écrire

$$V_{sn} = V_j - V_{N_d} \tag{2-171}$$

avec

$$V_{N_d} = \frac{kT}{e}Ln\left(\frac{\mu_n}{\mu_n+\mu_p}\frac{N_d^2}{n_i^2}\right) \approx 2\frac{kT}{e}Ln\left(\frac{N_d}{n_i}\right) \tag{2-172}$$

V_{N_d} est lié à la présence des ions donneurs ionisés. Dans une jonction p⁺n au silicium avec $N_a = 10^{18}\,cm^{-3}$ et $N_d = 10^{16}\,cm^{-3}$, la tension $V_{N_d} \approx 700\,mV$ alors que la tension de diffusion $V_d \approx 800\,mV$.

Rappelons que la tension de polarisation V_a de la diode se distribue entre la zone de charge d'espace et les régions neutres

$$V_a = V_j + V_s = V_j + V_{sp} + V_{sn} \tag{2-173}$$

En explicitant V_{sn} donné par l'expression (2-171) on obtient

$$V_a = 2V_j + V_{sp} - V_{N_d} \tag{2-174}$$

La tension appliquée à la zone de charge d'espace s'écrit donc

$$V_j = (V_a - V_{sp} + V_{N_d})/2 \tag{2-175}$$

En portant cette expression de V_j dans l'expression (2-150-a), on obtient la loi de variation du courant direct J avec la tension de polarisation V_a de la diode. Pour une jonction p⁺n on peut négliger la chute ohmique V_{sp} et le courant s'écrit

$$J = J_s^{**} e^{eV_a/2kT} \tag{2-176}$$

avec

$$J_s^{**} = J_s^* e^{eV_{Nd}/2kT} = J_s^* \frac{N_d}{n_i} = \frac{en_i D_p^*}{L_p^* th(d_n / L_p^*)} \tag{2-177-a}$$

Si la longueur de la région de type n est très inférieure à la longueur de diffusion des trous J_s^{**} s'écrit

$$J_s^{**} = \frac{en_i D_p^*}{d_n} \approx 2\frac{en_i D_p}{d_n} \tag{2-177-b}$$

Avant de résumer ce qui précède rappelons qu'à très faible niveau d'injection la loi $J(V_a)$ présente déjà une composante en $e^{eV_a/2kT}$ conditionnée par les phénomènes de recombinaison dans la zone de charge d'espace (Eq. 2-110).

Ainsi donc sous l'action d'une tension croissante de polarisation directe V_a***, le courant traversant une diode à jonction pn varie successivement comme*** $e^{eV_a/2kT}$***,*** $e^{eV_a/kT}$***,*** $e^{eV_a/2kT}$ ***et*** $(V_a - V_d)/r_s$ ***où*** r_s ***représente une résistance série dont les origines peuvent être diverses.*** La loi de variation du courant est donc de la forme générale $e^{eV_a/\alpha kT}$ où α varie entre 1 et 2 avec la tension appliquée et ceci de façon différente suivant les caractéristiques du semiconducteur, du profil de dopage, des dimensions géométriques, etc.

EXERCICES DU CHAPITRE 2

- **Exercice 1 : Les dimensions d'une jonction**

Calculez les longueurs de Debye et l'étendue de la zone de charge d'espace pour une jonction pn dopée avec $Na=10^{19}/cm^3$ et $Nd=10^{17}/cm^3$

- **Exercice 2 : Calcul du courant inverse**

$Na=5\ 10^{16}/cm^3$ et $Nd=10^{16}/cm^3$
La section est de $2\ 10^{-4}\ cm^2$
Calculez le courant inverse de saturation.

- **Exercice 3 : Le courant dans une jonction abrupte**

Expliquez pourquoi un courant peut traverser une jonction pn alors que la zone de charge d'espace ne comporte pas de porteurs libres.

- **Exercice 4** : **Faible et forte injection**

Rappelez les hypothèses faites dans le calcul du courant d'une jonction pn en régime de faible injection puis en régime de forte injection.

- **Exercice 5 : Capacités d'une jonction**

Calculez la relation donnant la capacité par cm^2 d'une jonction p+n et abrupte en fonction de la tension appliquée. Calculez les coefficients pour des dopages $Nd=10^{15}/cm^3$ $Nd=10^{16}/cm^3$ $Nd=10^{17}/cm^3$.

- **Exercice 6 : Champ maximum**

Pour $Na=10^{19}/cm^3$ et $Nd=10^{16}/cm^3$ calculez le champ maximum dans la jonction.

- **Exercice 7 : Capacités en fonction de la tension**

$Nd= 8\ .10^{15}/cm^3$ $Na=2\ .10^{19}/cm^3$
Calculez la capacité par unité de surface à 0V et à 4V en inverse.

- **Exercice 8 : Diode à avalanche**

Il s'agit d'étudier le courant d'une diode dans laquelle le champ est si élevé que les porteurs se multiplient par création de paires électron-trou.
$g_n(x)=\alpha_n(x)n(x)|v(x)|$ g est le taux de génération de paires, n est la densité d'électrons et v est la vitesse. Il y a une formule équivalente pour les trous. Le

coefficient α est donc le nombre de paires créées par électron et par unité de longueur. Les valeurs pour les électrons et les trous sont données par

$\alpha_n = A_n e^{-\frac{b_n}{E}}$ et $\alpha_p = A_p e^{-\frac{b_p}{E}}$

Ecrire l'équation de continuité pour les électrons et les trous en régime stationnaire. Les densités de courant des électrons et des trous sont j_n et j_p.

Montrez que

$$\frac{dj_n}{dx} - \left[\alpha_n - \alpha_p\right] j_n = \alpha_p J \text{ et } \frac{dj_p}{dx} - \left[\alpha_n - \alpha_p\right] j_p = -\alpha_n J \text{ avec } J = j_n + j_p$$

Montrez que

$$J = M_p j_p(x_n) + M_n j_n(x_p)$$

Avec $M_p = \dfrac{2}{1 - G(x_n)\left[F(x_n) - 1\right]}$ et $M_n = \dfrac{2G(x_n)}{1 - G(x_n)\left[F(x_n) - 1\right]}$

$$F(x_n) = \int_{x_p}^{x_n} \left(\alpha_n + \alpha_p\right) e^{-\int_{x_p}^{x_n} (\alpha_n - \alpha_p) dx'} dx \text{ et } G(x_n) = e^{\int_{x_p}^{x_n} (\alpha_n - \alpha_p) dx'}$$

Que deviennent ces valeur si α_n et α_p sont égaux.

En déduire la condition du régime d'avalanche (*M* infini).

- **Exercice 9 : Diode IMPATT**

C'est une diode à avalanche particulière de type p+n i n+ .

La région p+ est suivie d'une région dopée n puis d'une région intrinsèque et enfin d'une zone n+ .

Une polarisation inverse place la diode juste en dessous du claquage par avalanche. Si une tension alternative est superposée à la valeur continue, il y a avalanche pendant l'alternance positive puis augmentation des porteurs qui transitent ensuite dans la zone intrinsèque pendant un temps τ. Si ce temps est supérieur au quart de la période du signal alternatif, le courant augmente quand la tension diminue ce qui est équivalent à une résistance négative.

Représentez le champ électrique dans ce dispositif.

Dans un premier temps on calcule l'impédance de la zone d'avalanche p+n.

Écrire les équations de continuité dépendant du temps.

On suppose α_n et α_p égaux et on néglige les générations thermiques. On néglige le courant de diffusion.

En supposant que la zone d'avalanche s'étend de x_p à x_n montrez que le courant J suit la loi : $\dfrac{\partial J}{\partial t} = -2\dfrac{J}{\tau_a}\left(1 - \int_{x_p}^{x_n} \alpha(x) dx\right) + 2\dfrac{J_s}{\tau_a}$

Le temps τ_a est le temps de traversée de la zone d'avalanche, on suppose que la vitesse des porteurs (électrons et trous) y est constante. J_s est le courant de saturation de la diode.

Pour simplifier le problème on définit un taux d'avalanche moyen $\overline{\alpha}$ dans la zone d'avalanche. Montrez que

$\frac{\partial J}{\partial t} = 2\frac{J}{\tau_a}\left(\overline{\alpha} L_a - 1\right)$ avec L_a égal à x_n-x_p. On néglige la courant de saturation.

On applique maintenant un champ alternatif supplémentaire $E_a e^{j\omega t}$.

Montrez que le courant alternatif total somme du courant et du courant de déplacement s'écrit

$$J = j\left(-\frac{2\alpha' v_s J_c}{\omega} + \omega\varepsilon\right)E_a$$

On écrit $E = E_c + E_a e^{j\omega t}$ et $J = J_c + J_a e^{j\omega t}$

On suppose $\frac{d\overline{\alpha}}{dE} = \alpha'$ constant.

Montrez que la jonction peut entrer en résonance. Quelle est la fréquence de résonance ?

Montrez que l'impédance de la jonction s'écrit

$$Z_a = \frac{L_a}{j\omega\varepsilon}\frac{1}{1-\left(\frac{\omega_a}{\omega}\right)^2} \qquad \text{avec} \qquad \omega_a = \sqrt{\frac{2\alpha' v_s J_c}{\varepsilon}}$$

Calculez maintenant l'impédance Z_i de la zone intrinsèque de longueur L_i en supposant que les porteurs gardent la vitesse limite v_s.

Montrez que

$$Z_i = \frac{L_i}{j\omega\varepsilon}\left[1 - \frac{{\omega_a}^2}{{\omega_a}^2 - \omega^2}\frac{1-\cos\theta + j\sin\theta}{j\theta}\right]$$

$$\theta = \frac{\omega L_i}{v_s}$$

Le dispositif a également une résistance de contact r_s.

Montrez que l'impédance totale a une partie réelle qui peut s'annuler.

Expliquez la fonction oscillateur du dispositif.

3. Transistors bipolaires

Le transistor bipolaire a été inventé par Bardeen et Brattain en 1948, la théorie en a été élaborée par Shockley en 1949, le premier transistor à jonction a été fabriqué en 1951. C'est historiquement le premier composant actif à semiconducteur, son influence dans l'industrie électronique a été considérable. Le transistor bipolaire peut être considéré du point de vue électronique comme une source de courant commandée par une tension. Ce courant peut alors créer une différence de potentiel aux bornes d'une charge (une résistance par exemple) plus importante que la tension de commande. On dit alors que le dispositif a un gain en tension. Il est facile de comprendre alors qu'il est possible de l'utiliser pour réaliser des amplificateurs électroniques. Quand la variation de tension appliquée en entrée permet de faire passer le transistor de l'état non conducteur à l'état conducteur, on peut l'utiliser pour réaliser des fonctions binaires. Il a donc été aussi un composant de base de l'électronique numérique. Les transistors bipolaires ont cependant été de moins en moins utilisés à partir des années 80 pour être remplacés par les transistors de type MOSFET offrant des avantages en terme de consommation électrique. Les transistors bipolaires sont encore utilisés pour des fonctions exigeantes en terme de rapport signal sur bruit ou dans des applications d'électronique de puissance. Le terme bipolaire vient du fait que les deux types de porteurs (électrons et trous) participent à la conduction du dispositif.
Les points suivants seront traités.

3.1. Effet transistor
3.2. Equations d'Ebers-Moll
3.3. Etude pour différents types de profil de dopage
3.4. Effet Early
3.5. Equations dynamiques et étages à transistors
3.6. Sources de bruit dans un tranistor bipolaire

3.1. Effet transistor

Un transistor bipolaire est constitué comme le montre la figure 3.1 de trois régions différemment dopées pnp ou npn, d'un même monocristal semiconducteur. Il s'agit donc de deux jonctions pn tête-bêche présentant une région commune. Les trois régions portent respectivement les noms d' *Emetteur*, *Base* et *Collecteur*. C'est l'interaction étroite entre la jonction émetteur-base (E-B) et la jonction base-collecteur (B-C), favorisée par une faible épaisseur de base, qui est à l'origine de l'*effet transistor*.

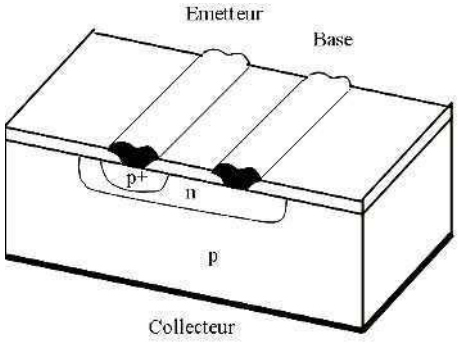


Figure 3-1 : Le transistor bipolaire

Nous avons vu dans le chapitre précédent que le courant inverse dans une jonction pn était proportionnel aux densités de porteurs minoritaires. ***Le principe de l'effet transistor consiste à moduler le courant inverse de la jonction base-collecteur polarisée en inverse, par une injection de porteurs minoritaires dans la base à partir de la jonction émetteur-base polarisée dans le sens direct.*** Une autre manière de comprendre le fonctionnement est de considérer les trous injectés dans la base à travers la jonction émetteur-base polarisée en direct. Ils sont ensuite soumis au champ intense de la jonction base collecteur polarisée en inverse et dérivent vers le collecteur.

Le bon fonctionnement du transistor nécessite que les porteurs minoritaires, injectés dans la base depuis l'émetteur, atteignent la jonction base-collecteur. Il est donc impératif que ces porteurs ne se recombinent pas à la traversée de la base, il faut par conséquent que l'épaisseur de la base soit très inférieure à la longueur de diffusion des porteurs minoritaires.

Ainsi, dans la mesure où la base est suffisamment étroite, une forte proportion du courant direct de la jonction émetteur-base constitue le courant inverse de la jonction base-collecteur. A partir d'un signal d'entrée (E-B) constitué d'un fort courant (courant direct E-B) et d'une faible tension (tension directe E-B), on recueille donc à la sortie (C-B) un signal constitué d'un fort courant (courant

inverse C-B ≈ courant direct E-B) et d'une forte tension (tension inverse C-B). Le dispositif est par conséquent amplificateur de puissance, c'est un composant actif.

Un raisonnement phénoménologique permet d'écrire simplement les équations qui régissent le fonctionnement du transistor à partir de l'équation caractéristique de la jonction pn.

Considérons la figure (3-2) et superposons deux états d'équilibre caractérisés successivement par les jonctions C-B puis E-B, en court-circuit (Fig 3-2-b et c).

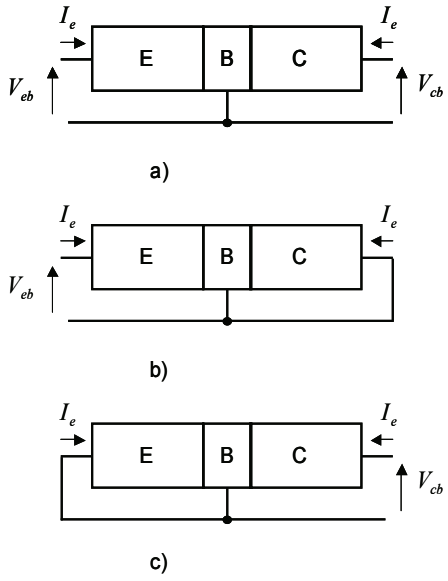


Figure 3-2 : Principe de superposition appliqué au bipolaire

- $V_{eb} \neq 0, V_{cb} = 0$

Le courant d'émetteur I_e est le courant d'une jonction pn, la jonction émetteur-base, polarisée par une tension V_{eb}, soit (Eq. 2-54)

$$I_e = I_{s1}\left(e^{eV_{eb}/kT} - 1\right) \qquad (3\text{-}1\text{-a})$$

où I_{s1} est le courant de saturation de la jonction émetteur-base.

Le courant I_e se distribue entre l'électrode de base et l'électrode de collecteur. Si la base est très étroite et peu dopée, la plupart des porteurs minoritaires injectés dans la base atteignent la zone de charge d'espace de la jonction base-collecteur. Ils sont alors happés par le champ favorable qui règne dans cette zone, et constituent le

courant collecteur I_c. Une faible proportion de ces porteurs se recombinent dans la base et créent un courant de recombinaison qui contribue au courant base I_b.

Le courant collecteur est donc une fraction importante du courant d'émetteur, compte tenu de la convention de signe portée sur la figure (3-2), ce courant s'écrit

$$I_c = -\alpha I_e = -\alpha I_{s1}\left(e^{eV_{eb}/kT} - 1\right) \qquad \text{(3-1-b)}$$

α représente le *gain courant* du transistor à tension collecteur-base nulle, V_{cb}=0.

- $\boldsymbol{V_{eb} = 0,\ V_{cb} \neq 0}$

C'est le régime inverse du précédent, le fonctionnement de la structure est formellement semblable au précédent en inversant les indices e et c.

Le courant I_c est le courant de la jonction collecteur-base polarisée par une tension V_{cb}

$$I_c = I_{s2}\left(e^{eV_{cb}/kT} - 1\right) \qquad \text{(3-2-a)}$$

où I_{s2} représente le courant de saturation de la jonction base-collecteur.

Le courant collecteur I_c se distribue entre l'électrode d'émetteur et celle de base, si on appelle α_i la proportion de ce courant qui constitue le courant d'émetteur, ce dernier s'écrit

$$I_e = -\alpha_i I_c = -\alpha_i I_{s2}\left(e^{eV_{cb}/kT} - 1\right) \qquad \text{(3-2-b)}$$

α_i représente le gain en courant inverse du transistor à tension émetteur-base nulle Il faut noter que pour des raisons de dissymétrie, à la fois dans le dopage et dans la géométrie, les coefficients α et α_i sont différents, $\alpha > \alpha_i$.

Si on superpose les deux états d'équilibre précédents, les courants I_e et I_c sont respectivement donnés par les sommes des expressions (3-1-a) et (3-2-b) d'une part et (3-1-b) et (3-2-a) d'autre part

$$I_e = I_{s1}\left(e^{eV_{eb}/kT} - 1\right) - \alpha_i I_{s2}\left(e^{eV_{cb}/kT} - 1\right) \qquad \text{(3-3-a)}$$

$$I_c = -\alpha I_{s1}\left(e^{eV_{eb}/kT} - 1\right) + I_{s2}\left(e^{eV_{cb}/kT} - 1\right) \qquad \text{(3-3-b)}$$

Ces équations de fonctionnement du transistor bipolaire portent le nom d'*équations d'Ebers-Moll*. Nous les avons déduites de raisonnements qualitatifs, nous allons les établir quantitativement.

3.2. Equations d'Ebers-Moll

L'étude du transistor est fondée sur les hypothèses déjà formulées dans l'étude de la jonction pn. ***La concentration des porteurs minoritaires injectés dans l'une quelconque des régions reste faible devant celle des majoritaires, c'est l'hypothèse de faible injection. Les phénomènes de génération-recombinaison dans les zones de charge d'espace sont négligés. En outre, une hypothèse simplificatrice est ajoutée en ce qui concerne la région de base. Celle-ci est de faible épaisseur et généralement peu dopée***. Il en résulte que les porteurs minoritaires ont dans cette région un temps de transit réduit alors que leur durée de vie est importante. En d'autres termes la longueur de diffusion de ces porteurs est beaucoup plus grande que la largeur de la base. Ceci permet de négliger les phénomènes de recombinaison dans cette région. Nous considérerons le cas d'un transistor pnp.

3.2.1. *Courants de porteurs minoritaires dans l'émetteur et le collecteur*

La figure 3-3 représente les charges et les courants d'un transistor pnp polarisé.

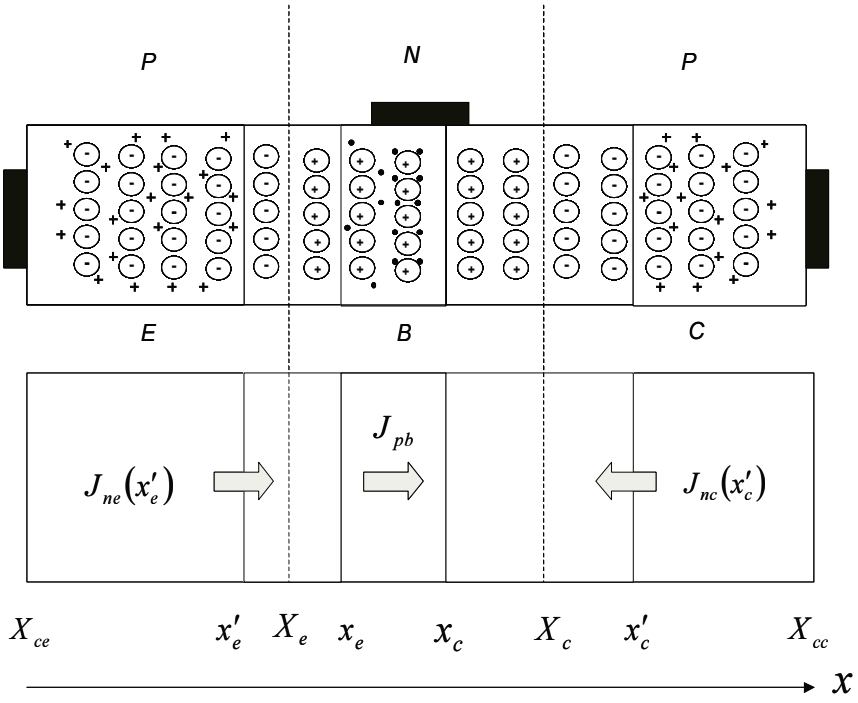


Figure 3-3 : Transistor pnp polarisé

Les régions d'émetteur et de collecteur sont en tout point semblables aux régions neutres d'une jonction pn. Dans chacune de ces régions, des porteurs minoritaires sont injectés à la frontière de la zone de charge d'espace par la polarisation de la jonction. Ces porteurs diffusent vers le contact ohmique situé à l'extrémité de la région. Leur distribution est fonction des valeurs relatives de leur longueur de diffusion et de la longueur de chacune des régions. La diffusion de ces porteurs est à l'origine du courant de porteurs minoritaires dans les régions d'émetteur et de collecteur. Compte tenu des expressions (2-66 et 68), les courants d'électrons en $x=x'_e$ dans l'émetteur, et en $x=x'_c$ dans le collecteur (Fig. 3-3) s'écrivent

$$j_{ne}(x'_e)=\frac{en_i^2 D_{ne}}{\int_E N_a(x)dx}\left(e^{eV_{eb}/kT}-1\right) \tag{3-4-a}$$

$$j_{nc}(x'_c)=\frac{en_i^2 D_{nc}}{\int_C N_a(x)dx}\left(e^{eV_{cb}/kT}-1\right) \tag{3-4-b}$$

où les intégrales sur l'émetteur et le collecteur sont respectivement de la forme

$$\int_E=\int_{x'_e-L_{ne}\,th(W_e/L_{ne})}^{x'_e} \tag{3-5-a}$$

$$\int_C=\int_{x'_c}^{x'_c+L_{nc}\,th(W_c/L_{nc})} \tag{3-5-b}$$

L_{ne} et L_{nc} représentent les longueurs de diffusion des électrons dans l'émetteur et le collecteur respectivement. $W_e=x'_e-X_{ce}$ soit environ X_e-X_{ce} et $W_c=X_{CC}-x'_c$ soit environ $X_{cc}-X_c$, représentent les longueurs des régions neutres d'émetteur et de collecteur. L'émetteur et le collecteur étant beaucoup plus dopés que la base, la zone de charge d'espace de chacune des jonctions s'étale principalement dans la base. On peut donc confondre x'_e et X_e d'une part et x'_c et X_c d'autre part. A partir des ordres de grandeur relatifs des paramètres W_e et L_{ne} d'une part et W_c et L_{nc} d'autre part, les bornes des intégrales peuvent se simplifier, comme pour la jonction pn, à partir des approximations $th\,(W/L)\approx 1$ pour $W>>L$, $th(W/L)\approx W/L$ pour $W<<L$.

3.2.2. *Courant de porteurs minoritaires dans la base*

Dans la mesure où on néglige les phénomènes de recombinaison dans la base, le calcul de la distribution des porteurs minoritaires est le même que celui que nous avons développé dans le cas d'une jonction pn à profil de dopage quelconque, avec des régions n et p très courtes (Par. 2-3-1). Seules changent les conditions aux limites. La distribution des trous dans la base est par conséquent donnée par l'expression (2-60) qui s'écrit ici avec les notations utilisées

$$p(x)=\frac{1}{N_d(x)}\left(K-\frac{j_{pb}}{eD_{pb}}\int_{x_e}^{x} N_d(x)dx \right) \tag{3-6}$$

où K est une constante d'intégration déterminée par les conditions aux limites, D_{pb} la constante de diffusion des trous dans la base, $N_d(x)$ la distribution de donneurs dans la base, et x_e et x_c les frontières de la région neutre. j_{pb} représente le courant de trous dans la base.

Comme pour la jonction pn, si on appelle p_e et N_{de} les densités de trous et de donneurs en $x=x_e$ d'une part, et p_c et N_{dc} les densités de trous et de donneurs en $x=x_c$ d'autre part, le courant de trous j_{pb} est donné par l'expression (2-61) qui s'écrit ici

$$j_{pb}=eD_{pb}\frac{p_eN_{de}-p_cN_{dc}}{\int_{x_e}^{x_c} N_d(x)dx} \tag{3-7}$$

j_{pb} est indépendant de x car nous avons négligé les phénomènes de recombinaison dans la base. Dans l'approximation des faibles injections, les densités de porteurs minoritaires injectés aux frontières de la zone de charge d'espace d'une jonction pn, sont données par les expressions (2-39), p'_n et V dans ces expressions correspondent respectivement à p_e et p_c d'une part et V_{eb} et V_{cb} d'autre part. Ainsi p_e et p_c sont donnés par

$$p_e=\frac{n_i^2}{N_{de}}e^{eV_{eb}/kT} \tag{3-7-a}$$

$$p_c=\frac{n_i^2}{N_{dc}}e^{eV_{cb}/kT} \tag{3-7-b}$$

On obtient donc l'expression du courant de porteurs minoritaires dans la base, en fonction des tensions émetteur-base et collecteur-base, en portant les expressions (3-7-a et b) dans l'expression (3-7), soit

$$j_{pb}=\frac{en_i^2D_{pb}}{\int_{x_e}^{x_c} N_d(x)dx}\left(e^{eV_{eb}/kT}-e^{eV_{cb}/kT}\right) \tag{3-8}$$

ou

$$j_{pb}=\frac{en_i^2D_{pb}}{\int_{x_e}^{x_c} N_d(x)dx}\left[\left(e^{eV_{eb}/kT}-1\right)-\left(e^{eV_{cb}/kT}-1\right)\right] \tag{3-9}$$

Le courant j_{pb} apparaît comme la superposition de deux contributions, la première correspondant au régime de fonctionnement défini par $V_{eb}\neq 0$ et $V_{cb}=0$, la seconde correspondant au régime défini par $V_{eb}=0$ et $V_{cb}\neq 0$.

3.2.3. *Courants d'émetteur et de collecteur*

Dans la mesure où l'on néglige les phénomènes de génération-recombinaison dans les zones de charge d'espace des jonctions, on peut écrire les égalités

$$j_{ne}(x_e^{'}) = j_{nb}(x_e)$$
$$j_{nc}(x_c^{'}) = j_{nb}(x_c)$$

On obtient alors le courant d'émetteur J_e en effectuant la somme algébrique des courants $j_{nb}(x_e)$ et j_{pb}, et le courant de collecteur en effectuant la somme algébrique des courants $j_{nb}(x_c)$ et j_{pb}. Compte tenu des conventions de signe portées sur la figure (3-2), ces courants s'écrivent

$$j_e = j_{ne}(x_e^{'}) + j_{pb}$$
$$j_c = j_{nc}(x_c^{'}) - j_{pb}$$

En explicitant les diverses composantes à partir des expressions (3-4 et 9) on obtient les équations d'Ebers-Moll

$$j_e = J_{s1}\left(e^{eV_{eb}/kT} - 1\right) - \alpha_i J_{s2}\left(e^{eV_{cb}/kT} - 1\right) \tag{3-10-a}$$

$$j_c = -\alpha J_{s1}\left(e^{eV_{eb}/kT} - 1\right) + J_{s2}\left(e^{eV_{cb}/kT} - 1\right) \tag{3-10-b}$$

avec

$$J_{s1} = en_i^2\left(\frac{D_{ne}}{\int_E N_a(x)dx} + \frac{D_{pb}}{\int_{x_e}^{x_c} N_d(x)dx}\right) \tag{3-11-a}$$

$$J_{s2} = en_i^2\left(\frac{D_{nc}}{\int_C N_a(x)dx} + \frac{D_{pb}}{\int_{x_e}^{x_c} N_d(x)dx}\right) \tag{3-11-b}$$

$$\frac{1}{\alpha_i} = 1 + \frac{D_{nc}}{D_{pb}}\frac{\int_{x_e}^{x_c} N_d(x)dx}{\int_C N_a(x)dx} \tag{3-11-c}$$

$$\frac{1}{\alpha} = 1 + \frac{D_{ne}}{D_{pb}}\frac{\int_{x_e}^{x_c} N_d(x)dx}{\int_E N_a(x)dx} \tag{3-11-d}$$

On vérifie facilement l'égalité

$$\alpha_i J_{s2}=\alpha J_{s1}=en_i^2 D_{pb}\Big/\int_{x_e}^{x_c} N_d(x)dx \qquad (3\text{-}12)$$

Si $V_{cb}=0$, les expressions (3-10) se réduisent à

$$J_e=J_{s1}\left(e^{eV_{eb}/kT}-1\right) \qquad J_c=-\alpha J_e$$

J_{s1} représente le courant de saturation de la jonction émetteur-base et α le gain en courant du transistor, en régime statique à $V_{cb}=0$. En fait si $V_{cb} \neq 0$, dans la mesure où en fonctionnement normal $V_{cb}<0$ avec $-V_{cb}>>kT/e$, le deuxième terme de l'expression (3-10-b) se réduit à J_{s2} qui est négligeable devant le premier terme. α est donc le gain statique du transistor.

En posant

$$J_1=J_{s1}\left(e^{eV_{eb}/kT}-1\right)$$
$$J_2=J_{s2}\left(e^{eV_{cb}/kT}-1\right)$$

les équations d'Ebers-Moll s'écrivent simplement

$$J_e = J_1 - \alpha_i J_2 \qquad (3\text{-}13\text{-a})$$

$$J_c = -\alpha J_1 + J_2 \qquad (3\text{-}13\text{-b})$$

Ces équations permettent d'établir le schéma équivalent du transistor représenté sur la figure (3-4-a), c'est le schéma d'Ebers-Moll.

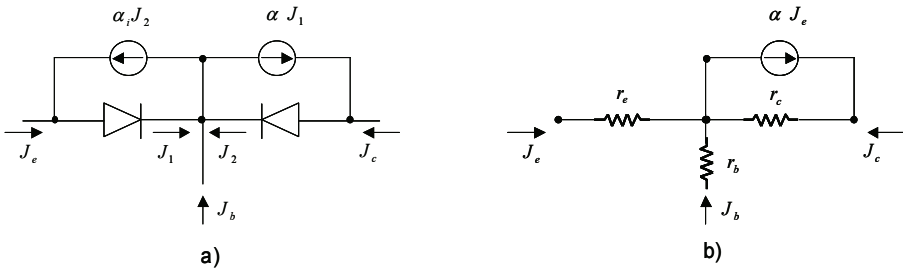


Figure 3-4 : Schéma équivalent d'Ebers-Moll

En régime de fonctionnement normal la jonction émetteur-base est polarisée dans le sens passant et la jonction collecteur-base dans le sens bloqué. La première se comporte donc comme une faible résistance r_e appelée *résistance d'émetteur* et la seconde comme une grande résistance r_c appelée *résistance de collecteur*. En outre le courant J_1 est important tandis que le courant J_2 est faible. Le schéma d'Ebers-Moll peut alors être représenté sous la forme simplifiée de la figure (3-4-b), sur laquelle est rajoutée la faible *résistance de base* r_b, ignorée dans le modèle d'Ebers-Moll.

3.3. Différents types de profil de dopage

3.3.1. *Transistor à dopages homogènes*

Les différentes régions du transistor sont dopées de manière homogène avec $N_a(x)=N_{ae}$ et N_{ac} dans l'émetteur et le collecteur respectivement, et $N_d(x)=N_{db}$ dans la base. Le profil de dopage et la structure de bandes du transistor sont représentés sur la figure (3-5). A partir de ce profil de dopage on peut calculer les différentes intégrales apparaissant dans les expressions (3-11) et par suite les paramètres du transistor. Généralement l'émetteur et le collecteur sont beaucoup plus dopés que la base, il en résulte que les zones de charge d'espace des jonctions s'étalent essentiellement dans la base, on supposera donc $x'_e \approx X_e$ et $x'_c \approx X_c$, de sorte que les différentes intégrales s'écrivent

$$\int_E N_a(x)dx=\int_{X_e-L_{ne}\,th(W_e/L_{ne})}^{X_e} N_{ae}\,dx=N_{ae}L_{ne}th\left(W_e/L_{ne}\right) \tag{3-14-a}$$

$$\int_C N_a(x)dx=\int_{X_c}^{X_c+L_{nc}\,th(W_c/L_{nc})} N_{ac}dx=N_{ac}L_{nc}th(W_c/L_{nc}) \tag{3-14-b}$$

$$\int_{x_e}^{x_c} N_d(x)dx=N_{db}.(x_c-x_e)=N_{db}W_{be} \tag{3-14-c}$$

où W_{be} représente la largeur de la région neutre de la base, on appelle W_{be} *largeur effective* de la base.

Il suffit d'utiliser les expressions (3-14) pour calculer les divers paramètres physiques du transistor. Considérons plus particulièrement le gain en courant

$$\alpha=1\Big/\left(1+\frac{D_{ne}}{D_{pb}}\frac{N_{db}W_{be}}{N_{ae}L_{ne}th(W_e/L_{ne})}\right) \tag{3-15}$$

Compte tenu de la relation d'Einstein $D/\mu=kT/e$, cette expression se met sous la forme

$$\alpha=1\Big/\left(1+\frac{\mu_{ne}}{\mu_{pb}}\frac{N_{db}}{N_{ae}}\frac{W_{be}}{L_{ne}th(W_e/L_{ne})}\right) \tag{3-16}$$

Si la longueur de diffusion des électrons dans l'émetteur L_{ne} et la longueur de l'émetteur W_e sont très différentes, on peut utiliser les expressions approchées de $th(x)$, $th(x)\approx 1$ pour $x>>1$ et $th(x)\approx x$ pour $x<<1$. α s'écrit alors

$$\alpha = 1 \Big/ \left(1 + \frac{\mu_{ne}}{\mu_{pb}} \frac{N_{db}}{N_{ae}} \frac{W_{be}}{Min(W_e, L_{ne})} \right) \qquad (3\text{-}17\text{-a})$$

où $Min(W_e, L_{ne})$ représente la plus petite des grandeurs W_e et L_{ne}.

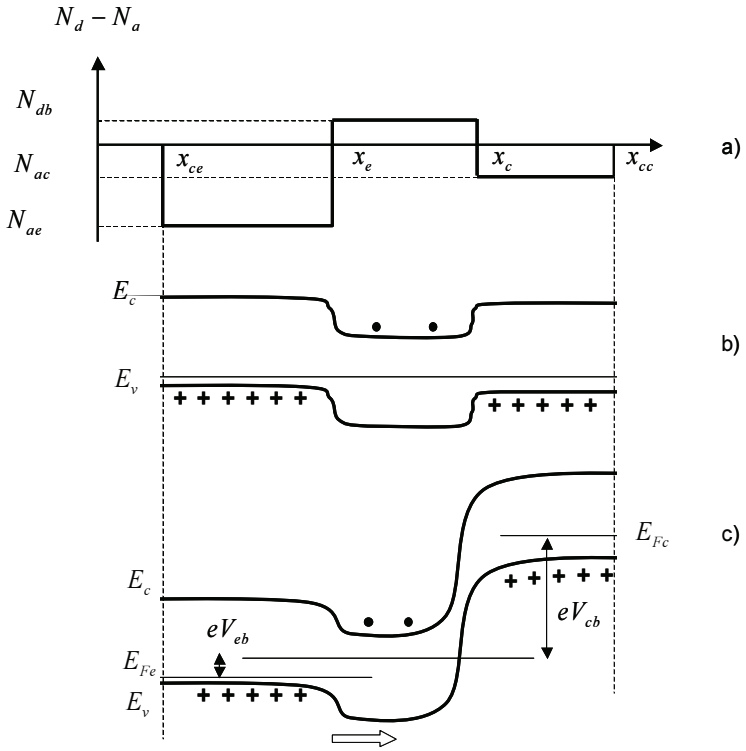


Figure 3-5 : Transistor pnp à dopages homogènes. a) Profil de dopage. b) Diagramme énergétique à l'équilibre thermodynamique. c) Diagramme énergétique sous polarisation en régime normal de fonctionnement

Pour un transistor de type npn on obtient de la même manière

$$\alpha = 1 \Big/ \left(1 + \frac{\mu_{pe}}{\mu_{nb}} \frac{N_{ab}}{N_{de}} \frac{W_{be}}{Min(W_e, L_{pe})} \right) \qquad (3\text{-}17\text{-b})$$

Pour obtenir un bon transistor bipolaire, il faut que α soit le plus proche de 1 possible, de sorte qu'il faut réaliser les conditions $\mu_b >> \mu_e$, $N_e >> N_b$, $Min(W_e, L_e) >> W_{be}$. On réalisera donc un transistor de type npn pour que la mobilité des porteurs minoritaires dans la base soit beaucoup plus importante que la mobilité des porteurs minoritaires dans l'émetteur. On réalisera une base beaucoup moins dopée que l'émetteur. Pour satisfaire à la troisième condition on réalisera une base très étroite

et on utilisera du matériau de bonne qualité pour avoir une longueur de diffusion aussi grande que possible.

3.3.2. *Transistor drift*

La base du transistor est dopée de manière non homogène pour créer un champ interne dans le but d'accélérer les porteurs minoritaires et par suite de diminuer leur recombinaison.

Le courant de porteurs majoritaires dans la base d'un transistor pnp s'écrit

$$j_n = eD_n\left(n(x)\frac{e}{kT}E(x) + \frac{dn(x)}{dx} \right) \tag{3-18}$$

A l'équilibre thermodynamique, c'est-à-dire en l'absence de polarisation, j_n=0, et d'autre part $n(x)=N_{db}(x)$. Il en résulte que

$$E(x) = -\frac{kT}{e}\frac{1}{N_{db}(x)}\frac{dN_{db}(x)}{dx} = -\frac{kT}{e}\frac{d}{dx}\left(LnN_{db}(x)\right) \tag{3-19}$$

Si la base est dopée de manière homogène, $N_{db}(x)=N_{db}$=Cte et par suite $E(x)$=0. Un cas particulièrement intéressant est celui dans lequel le dopage de la base varie suivant une loi exponentielle. Dans ce cas, il existe un champ électrique dans la base et ce champ est constant. Considérons un transistor pnp dans lequel la concentration de donneurs dans la base $N_{db}(x)$ varie exponentiellement suivant la loi

$$N_{db}(x) = N_{dbe}\, e^{-a(x-X_e)/W_b} \tag{3-20}$$

N_{dbe} représente la concentration de donneurs dans la base côté émetteur en $x=X_e$. $W_b=X_c-X_e$ est la largeur métallurgique de la base. Enfin, a représente la quantité $Ln\,(N_{dbe}/N_{dbc})$, en effet en $x=X_c$ l'expression (3-20) s'écrit

$$N_{db}(x_c) = N_{dbc} = N_{dbe}\, e^{-a(X_c-X_e)/W_b} = N_{dbe}\, e^{-a}$$

L'expression (3-20) permet d'écrire

$$LnN_{db}(x) = LnN_{dbe} - a(x-X_e)/W_b$$

et par suite

$$\frac{d}{dx}(Ln\,N_{db}(x)) = -a/W_b = -1/W_b \cdot Ln\,(N_{dbe}/N_{dbc})$$

En portant cette expression dans l'expression (3-19) on obtient le champ électrique dans la base à dopage exponentiel

$$E = \frac{kT}{eW_b} Ln\left(N_{dbe}/N_{dbc}\right) \tag{3-21}$$

On constate que le champ est constant, d'autre part si on veut accélérer les porteurs minoritaires c'est-à-dire les trous, il faut réaliser $E>0$, c'est-à-dire $N_{dbe}>N_{dbc}$. Le profil de dopage du transistor et la structure de bandes correspondante sont représentés sur la figure (3-6).

Les intégrales sur l'émetteur et le collecteur sont toujours données par les expressions (3-14), l'intégrale sur la base s'écrit

$$\int_{x_e}^{x_c} N_{db}(x)dx = -\frac{W_b}{a} N_{dbe}\left[e^{-a(x-X_e)/W_b}\right]_{x_e}^{x_c} = \frac{W_b}{a}\left(N_{db}(x_e) - N_{db}(x_c)\right)$$

soit

$$\int_{x_e}^{x_c} N_{db}(x)dx = \frac{W_b}{a} N_{db}(x_e)\left(1 - N_{db}(x_c)/N_{db}(x_e)\right) \tag{3-22}$$

A partir des relations (3-14 et 22) l'expression (3-11-d) s'écrit, compte tenu de la relation d'Einstein,

$$\alpha = 1\Big/\left(1 + \frac{1}{a}\frac{\mu_{ne}}{\mu_{pb}}\frac{N_{db}(x_e)}{N_{ae}}\frac{W_b}{L_{ne}\,th(W_e/L_{ne})}\left(1 - N_{db}(x_c)/N_{db}(x_e)\right)\right) \tag{3-23}$$

En régime de fonctionnement normal, la jonction émetteur-base est polarisée dans le sens direct, la zone de charge d'espace de cette jonction est donc réduite, de sorte que $x_e = X_e$ et $N_{db}(x_e) = N_{db}(X_e) = N_{dbe}$.

D'autre part, si le dopage de la base est fortement dissymétrique on peut supposer $N_{db}(x_c) << N_{db}(x_e)$. En explicitant le paramètre a le gain en courant du transistor s'écrit alors

$$\alpha = 1\Big/\left(1 + \frac{1}{Ln(N_{dbe}/N_{dbc})}\frac{\mu_{ne}}{\mu_{pb}}\frac{N_{dbe}}{N_{ae}}\frac{W_b}{L_{ne}\,th(W_e/L_{ne})}\right) \tag{3-24}$$

Comme précédemment si la longueur de l'émetteur W_e et la longueur de diffusion des électrons dans l'émetteur sont très différentes, on peut utiliser une expression approchée de $th(W_e/L_{ne})$ et écrire

$$\alpha = 1 \Big/ \left(1 + \frac{1}{Ln(N_{dbe}/N_{dbc})} \frac{\mu_{ne}}{\mu_{pb}} \frac{N_{dbe}}{N_{ae}} \frac{W_b}{Min(W_e, L_{ne})} \right) \qquad (3\text{-}25)$$

Pour augmenter α on a donc intérêt à réaliser un dopage de la base tel que $N_{dbe} >> N_{dbc}$.

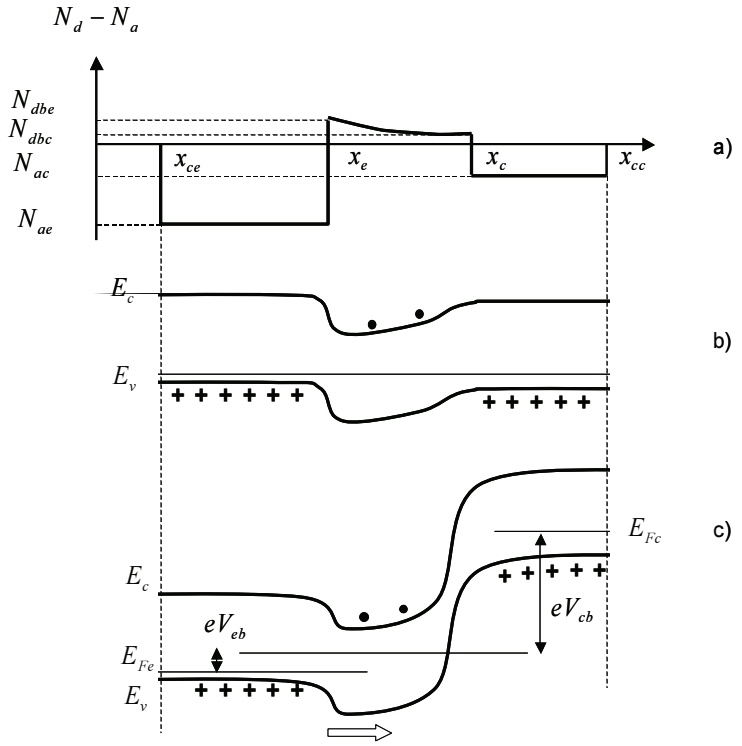


Figure 3-6 : Transistor pnp à dopage de base exponentiel. a) Profil de dopage. b) Diagramme énergétique à l'équilibre thermodynamique. c) Diagramme énergétique sous polarisation

3.3.3. *Résumé du fonctionnement d'un transistor*

La figure 3-7 représente les densités de porteurs n et p dans un transistor pnp pour les différents régimes de fonctionnement : actif, bloqué, saturé et inversé.

- Dans le régime actif, la jonction émetteur-base est polarisée en mode direct et la jonction base-collecteur en mode inverse.
- Dans le mode bloqué, les deux jonctions sont polarisées en mode inverse.
- Dans le mode saturé, les deux jonctions sont polarisées en direct.
- Dans le mode inversé, la jonction émetteur-base est polarisée en inverse et la jonction base-collecteur en direct.

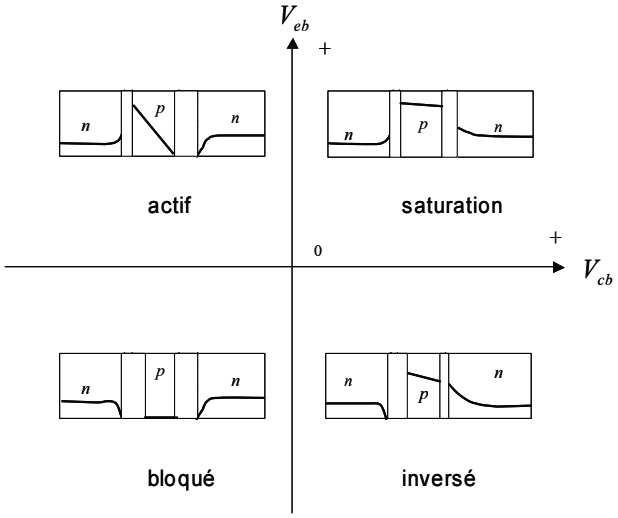


Figure 3-7 : Les divers modes de fonctionnement d'un bipolaire pnp

La figure représente les densités des minoritaires en fonction des tensions appliquées aux deux jonctions. On peut en déduire les courants de diffusion et de dérive Bien que le mode actif soit le plus utilisé, il est utile de connaître les autres modes qui peuvent apparaître quand le transistor est utilisé en commutation.

3.4. Effet Early

Les équations d'Ebers-Moll traduisent les relations courants-tensions dans le transistor, mais les différentes tensions apparaissent à la fois explicitement et implicitement. Les tensions émetteur-base V_{eb} et collecteur-base V_{cb} apparaissent explicitement dans les termes exponentiels des équations (3-10). Mais ces mêmes tensions apparaissent aussi dans les différents coefficients J_{s1}, J_{s2}, α et α_i, dans la mesure où ces coefficients sont fonction d'intégrales sur les différentes régions neutres du transistor (Eq. 3-11). Les bornes de ces intégrales x'_e, x_e, x_c et x'_c sont fonction des tensions de polarisation V_{eb} et V_{cb}.

En fait, l'émetteur et le collecteur étant beaucoup plus dopés que la base, les zones de charge d'espace des jonctions et leurs variations en fonction des tensions de polarisation, s'étendent essentiellement dans la base. On peut donc supposer constants $x'_e \approx X_e$ et $x'_c \approx X_c$. Les tensions V_{eb} et V_{cb} modifient respectivement x_e et x_c et par suite la valeur de la largeur effective $W_{be}=x_c-x_e$ de la base. Cette modulation de la largeur effective de la base constitue l'*effet Early*. Il faut toutefois noter que la jonction émetteur-base étant polarisée dans le sens direct et la jonction collecteur-base dans le sens inverse, la valeur de x_e est beaucoup moins affectée par une

variation ΔV_{eb} de la tension V_{eb}, que ne l'est la valeur de x_c par une variation ΔV_{cb} de la tension V_{cb}.

Ainsi en régime de fonctionnement normal, c'est-à-dire lorsque la jonction émetteur-base est polarisée dans le sens direct et la jonction collecteur-base dans le sens inverse, les coefficients J_{s1}, J_{s2}, α et α_i sont fonction des tensions de polarisation V_{eb} et V_{cb}. Néanmoins, pour toutes variations ΔV_{eb} et ΔV_{cb} autour du point de polarisation statique, on peut supposer que les différents paramètres sont indépendants de ΔV_{eb}, et sensibles uniquement à ΔV_{cb} par l'intermédiaire de Δx_c.

Dans le but de traduire explicitement les variations des différents paramètres, et de faire apparaître des coefficients constants dans les équations, on définit les valeurs de repos des différents coefficients, à V_{eb}=0 et V_{cb}=0. On traduit les effets des tensions de polarisations statiques V_{eb} et V_{cb} et des variations ΔV_{cb} de V_{cb} par des termes correctifs.

En l'absence de polarisation, c'est-à-dire pour V_{eb}=0 et V_{cb}=0 les différents coefficients s'écrivent

$$J_{s10}=en_i^2\left(\frac{D_{ne}}{\int_E N_a(x)dx}+\frac{D_{pb}}{\int_{W_{be0}} N_d(x)dx}\right) \tag{3-26-a}$$

$$J_{s20}=en_i^2\left(\frac{D_{nc}}{\int_C N_a(x)dx}+\frac{D_{pb}}{\int_{W_{be0}} N_d(x)dx}\right) \tag{3-26-b}$$

$$\frac{1}{\alpha_{i0}}=1+\frac{D_{nc}}{D_{pb}}\frac{\int_{W_{be0}} N_d(x)dx}{\int_C N_a(x)dx} \tag{3-26-c}$$

$$\frac{1}{\alpha_0}=1+\frac{D_{ne}}{D_{pb}}\frac{\int_{W_{be0}} N_d(x)dx}{\int_E N_a(x)dx} \tag{3-26-d}$$

La relation (3-12) s'écrit

$$\alpha_{i0}J_{s20}=\alpha_0 J_{s10}=en_i^2 D_{pb}\Big/\int_{W_{be0}} N_d(x)dx \tag{3-27}$$

où $W_{beo}=x_{co}-x_{eo}$ représente la largeur effective de la base en l'absence de polarisation.

En définissant la quantité $J_s^{'}$ par l'expression

$$J_s^{'}=en_i^2 D_{pb}\left(\frac{1}{\int_{x_e}^{x_c} N_d(x)dx}-\frac{1}{\int_{W_{be0}} N_d(x)dx}\right) \tag{3-28}$$

on montre simplement que les différents coefficients qui apparaissent dans les équations d'Ebers-Moll pour des polarisations non nulles s'écrivent

$$J_{s1}=J_{s10}+J_s^{'} \tag{3-29-a}$$
$$J_{s2}=J_{s20}+J_s^{'} \tag{3-29-b}$$
$$\alpha J_{s1}=\alpha_0 J_{s10}+J_s^{'} \tag{3-29-c}$$
$$\alpha_i J_{s2}=\alpha_{i0} J_{s20}+J_s^{'} \tag{3-29-d}$$

Les équations (3-10) s'écrivent alors sous la forme

$$J_e=J_{s10}\left(e^{eV_{eb}/kT}-1\right)-\alpha_{i0} J_{s20}\left(e^{eV_{cb}/kT}-1\right)+J_s^{'}\left(e^{eV_{eb}/kT}-e^{eV_{cb}/kT}\right) \tag{3-30-a}$$

$$J_c=-\alpha_0 J_{s10}\left(e^{eV_{eb}/kT}-1\right)+J_{s20}\left(e^{eV_{cb}/kT}-1\right)-J_s^{'}\left(e^{eV_{eb}/kT}-e^{eV_{cb}/kT}\right) \tag{3-30-b}$$

Le paramètre $J_s^{'}$ est fonction des tensions de polarisation V_{eb} et V_{cb} par l'intermédiaire des bornes x_e et x_c de l'intégrale qui apparaît dans l'expression (3-28), mais pour toute variation de tension autour du point de polarisation, la variation de $J_s^{'}$ n'est fonction que des variations ΔV_{cb} de la tension V_{cb}. On peut expliciter cette variation et la linéariser. Pour toute variation dV_{cb} de la tension collecteur-base autour du point de polarisation, la variation de $J_s^{'}$ peut s'écrire

$$dJ_s^{'}=\frac{dJ_s^{'}}{dx_c}\frac{dx_c}{dV_{cb}}dV_{cb} \tag{3-31}$$

En dérivant l'expression (3-28) par rapport à x_c on obtient

$$\frac{dJ_s^{'}}{dx_c}=-en_i^2 D_{pb} N_{dc}\Big/\left(\int_{x_e}^{x_c} N_d(x)dx\right)^2$$

Compte tenu de l'expression (3-12), cette dérivée peut s'écrire sous la forme

$$\frac{dJ_s^{'}}{dx_c}=-\alpha J_{s1} N_{dc}\Big/\int_{x_e}^{x_c} N_d(x)dx$$

L'expression (3-31) s'écrit alors

$$dJ_s^{'} = -\alpha J_{s1} \frac{N_{dc}}{\int_{x_e}^{x_c} N_d(x)dx} \frac{dx_c}{dV_{cb}} dV_{cb}$$

ou plus simplement

$$dJ_s^{'} = -\frac{e}{kT} \alpha J_{s1} \mu dV_{cb} \tag{3-32}$$

avec

$$\mu = \frac{kT}{e} \frac{N_{dc}}{\int_{x_e}^{x_c} N_d(x)dx} \frac{dx_c}{dV_{cb}} \tag{3-33}$$

Le paramètre μ est appelé facteur de réaction d'Early du transistor. Il traduit l'effet sur $J_s^{'}$ c'est-à-dire sur J_e et J_c, de la modulation de la largeur effective de la base par la variation dV_{cb} de la tension collecteur-base.

3.5. Modèle dynamique et étages à transistor

3.5.1. *Modéle dynamique*

En régime dynamique le transistor est soumis à des tensions de polarisation que nous appellerons V_{ebo} et V_{cbo}, sur lesquelles est superposé le signal sous forme de petites variations, $\Delta V_{eb} = v_{eb}$ et $\Delta V_{cb} = v_{cb}$, soit

$$V_{eb} = V_{ebo} + v_{eb} \qquad I_e = I_{eo} + i_e$$

$$V_{cb} = V_{cbo} + v_{cb} \qquad I_c = I_{co} + i_c$$

En régime de fonctionnement normal, la jonction émetteur-base est polarisée dans le sens direct $V_{ebo} >> kT/e$, et la jonction collecteur-base est polarisée dans le sens inverse, $V_{cbo} < 0$ avec $-V_{cbo} >> kT/e$. Dans ce cas, les équations (3-30) se simplifient et s'écrivent après intégration sur la section du transistor

$$I_e = I_{s10} e^{eV_{eb}/kT} + \alpha_{i0} I_{s20} + I_s^{'} e^{eV_{eb}/kT} \tag{3-34-a}$$

$$I_c = -\alpha_0 I_{s10} e^{eV_{eb}/kT} - I_{s20} - I_s^{'} e^{eV_{eb}/kT} \tag{3-34-b}$$

Le fonctionnement dynamique du transistor, c'est-à-dire sa réponse aux variations des tensions V_{eb} et V_{cb} autour des valeurs statiques V_{ebo} et V_{cbo}, est donné par la différentielle des expressions précédentes, soit

$$dI_e = \frac{e}{kT} I_{s10}\, e^{eV_{eb0}/kT}\, dV_{eb} + \frac{e}{kT} I_s'\, e^{eV_{eb0}/kT}\, dV_{eb} + \frac{dI_s'}{dV_{cb}} e^{eV_{eb0}/kT}\, dV_{cb}$$

$$dI_c = -\frac{e}{kT} \alpha_0\, I_{s10}\, e^{eV_{ebO}/kT}\, dV_{eb} - \frac{e}{kT} I_s'\, e^{eV_{ebO}/kT}\, dV_{eb} - \frac{dI_s'}{dV_{cb}} e^{eV_{eb0}/kT}\, dV_{cb}$$

En groupant les coefficients de dV_{eb}, et compte tenu d'une part des expressions (3-29-a et c) et de la relation (3-32), ces expressions s'écrivent

$$dI_e = \frac{e}{kT} I_{s1}\, e^{eV_{eb0}/kT}\, dV_{eb} - \frac{e}{kT}\, \alpha I_{s1}\, \mu e^{eV_{eb0}/kT}\, dV_{cb} \qquad (3\text{-}35\text{-a})$$

$$dI_c = -\frac{e}{kT}\, \alpha I_{s1}\, e^{eV_{eb0}/kT}\, dV_{eb} + \frac{e}{kT}\, \alpha I_{s1}\, \mu e^{eV_{eb0}/kT}\, dV_{cb} \qquad (3\text{-}35\text{-b})$$

En représentant par des minuscules les variations des grandeurs statiques on obtient

$$i_e = y_{11b} v_{eb} + y_{12b} v_{cb} \qquad (3\text{-}36\text{-a})$$

$$i_c = y_{21b} v_{eb} + y_{22b} v_{cb} \qquad (3\text{-}36\text{-b})$$

les paramètres y_{ijb} sont les paramètres de la *matrice admittance* en montage base commune, ils sont donnés par

$$y_{11b} = \frac{e}{kT} I_{s1} e^{eV_{eb0}/kT} \qquad y_{12b} = -\alpha\mu y_{11b}$$

$$y_{21b} = -\alpha y_{11b} \qquad y_{22b} = -y_{12b}$$

Pour obtenir un schéma complet, il suffit d'ajouter aux équations 3-10 les relations de conservation du courant et les relations d'addition des potentiels.

$$I_e = I_{s1}\left(e^{eV_{eb}/kT} - 1\right) - \alpha_i I_{s2}\left(e^{eV_{cb}/kT} - 1\right) \qquad (3\text{-}37\text{-a})$$

$$I_c = -\alpha I_{s1}\left(e^{eV_{eb}/kT} - 1\right) + I_{s2}\left(e^{eV_{cb}/kT} - 1\right) \qquad (3\text{-}37\text{-b})$$

$$V_{eb} + V_{bc} + V_{ce} = 0 \qquad (3\text{-}37\text{-c})$$

$$I_e + I_b + I_c = 0 \qquad (3\text{-}37\text{-d})$$

Il suffit maintenant de particulariser ces équations pour les trois montages de base, base commune, émetteur commun et collecteur commun pour obtenir les équations utiles dans le calcul des circuits.

3.5.2. *Montage base commune*

Dans le montage base commune par exemple, la tension de base est supposée constante, la variation de cette tension par rapport au référentiel du circuit (la masse) est donc nulle. Les tensions et leurs variations sont alors exprimées par rapport à la tension de la base. On retient ainsi V_{eb} et v_{eb} pour l'entrée et V_{cb} et v_{cb} pour la sortie. Les quatre équations (3-37) permettent alors d'exprimer les courants (I_e et I_c) en fonction des grandeurs d'entrée (V_{eb} et V_{cb}). Le système comportait six inconnues mais on considère maintenant les deux tensions comme des données. De même, il est possible d'exprimer les petites variations pour obtenir le modèle dynamique. Pour être précis, il vaudrait mieux utiliser le terme « modèle petits signaux » car le modèle dynamique intègre en plus les éléments dépendant du temps. Il faut alors faire intervenir les capacités des jonctions et les temps de transit des charges dans le dispositif. A plus haute fréquence les éléments parasite correspondant aux connexions (résistances et inductances) contribuent également à la formation du signal de sortie.

Etudions dans un premier temps les caractéristiques d'entrée. Pour cela on fixe la tension collecteur base de sortie à une valeur constante. A V_{cb}=Cte l'équation (3-37-a) s'écrit

$$I_e = I_{s1}\left(e^{eV_{eb}/kT} - 1\right) + C \quad avec \quad C = -\alpha_i I_{s2}\left(e^{eV_{cb}/kT} - 1\right)$$

En régime de fonctionnement normal, la jonction collecteur-base est polarisée en inverse avec $-V_{cb} >> kT/e$ de sorte que la constante C se réduit à $\alpha_i I_{s2}$. Le courant d'émetteur s'écrit

$$I_e = I_{s1}\left(e^{eV_{eb}/kT} - 1\right) + \alpha_i I_{s2} \tag{3-38}$$

Le courant d'émetteur résulte donc de la somme du courant direct de la jonction émetteur-base polarisée dans le sens passant, et du terme $\alpha_i I_{s2}$ beaucoup plus faible que le précédent. Le réseau de caractéristiques d'entrée se ramène donc à une seule caractéristique 3.8. Le calcul de l'admittance d'entrée $\Delta I_e / \Delta V_{eb}$ montre que cette valeur est approximativement $e/kT \cdot I_e$. Elle est donc élevée ce qui veut dire que ***l'impédance d'entrée de l'étage base commune est faible***.

Les caractéristiques de sortie représentent les variations du courant collecteur en fonction de la tension collecteur-base, à courant d'émetteur constant, $I_c(V_{cb})_{Ie}$.

En multipliant par α l'équation (3-37-a) et en ajoutant membre à membre à l'équation (3-37-b), on obtient la relation

$$I_c = (1 - \alpha\alpha_i) I_{s2}\left(e^{eV_{cb}/kT} - 1\right) - \alpha I_e \tag{3-39}$$

Ainsi pour I_e=Cte, le courant I_c apparaît comme la somme de deux composantes. L'une est variable et relativement faible en fonctionnement normal dans la mesure où V_{cb}<0, l'autre est constante. Dès que $-V_{cb}>>kT/e$ le courant collecteur devient constant et égal à

$$I_c = (\alpha\alpha_i - 1) I_{s2} - \alpha I_e \approx -\alpha I_e \qquad (3\text{-}40)$$

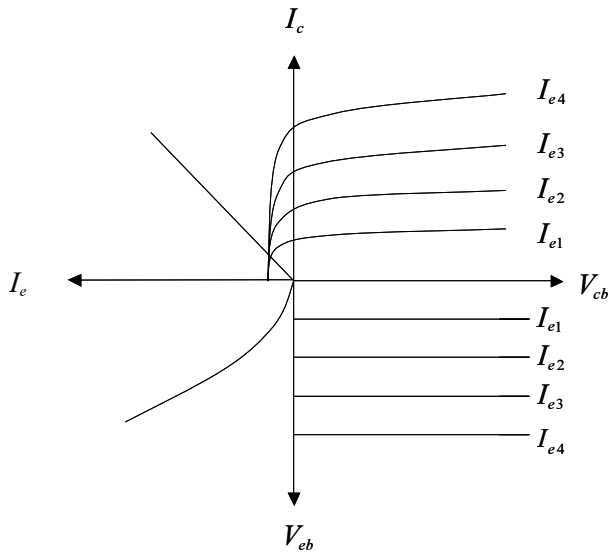


Figure 3-8 : Réseaux de courbes du bipolaire en base commune

Le réseau de caractéristiques de sortie est représenté dans le premier quadrant de la figure (3-8). Le courant collecteur étant constant en fonction de la tension collecteur-base cela veut dire que l'impédance de sortie de l'étage base commune est élevée. ***Le transistor est équivalent à une source de courant.***

Les caractéristiques de transfert en courant représentent $I_c(I_e)$ à V_{cb}=Cte, elles sont données par l'expression (3-39). A V_{cb}=Cte, le courant collecteur apparaît comme la somme de deux composantes dont l'une est importante et varie proportionnellement à I_e , l'autre est constante et faible en régime de fonctionnement normal. Pour $-V_{cb}>>kT/e$ le courant I_c est donné par l'expression (3-40). La courbe représentative est une droite de pente $-\alpha$ représentée dans le deuxième quadrant de la figure (3-8). Le gain en courant $-\alpha=I_c/I_e$ est voisin de -1.
Le montage base commune a donc un gain en courant peu différent de un.

Les caractéristiques de réaction en tension représentent $V_{eb}(V_{cb})$ à I_e=Cte. La relation qui lie V_{eb} à V_{cb} et I_e est contenue implicitement dans l'équation (3-37-a) que l'on peut écrire sous la forme explicite

$$V_{eb}=\frac{kT}{e}Ln\left(\alpha_i\frac{I_{s2}}{I_{s1}}\left(e^{eV_{cb}/kT}-1\right)+1+\frac{I_e}{I_{s1}}\right) \tag{3-41}$$

En régime de fonctionnement normal V_{cb} est négatif avec $-V_{cb}>>kT/e$ de sorte que V_{eb} devient pratiquement indépendant de V_{cb} et reste lié à I_e par une expression équivalente à la relation (3-38). Le réseau de caractéristiques est représenté dans le quatrième quadrant de la figure (3-8).

En résumé le montage base commune est utilisé pour obtenir une faible impédance d'entrée et une forte impédance de sortie. Le gain en courant est voisin de un. Le transistor est équivalent à une source de courant.

3.5.3. *Montage émetteur commun*

Il faut ici faire apparaître explicitement à partir des expressions (3-37) les lois de variation des courants I_b et I_c en fonction des tensions V_{be} et V_{ce}. Il suffit d'expliciter dans les relations (3-37-a et b) le courant I_e et les tensions V_{eb} et V_{cb} à partir des relations

$$\begin{aligned} I_e &= -\left(I_b+I_c\right) \\ V_{eb} &= -V_{be} \\ V_{cb} &= V_{ce}-V_{be} \end{aligned}$$

Les calculs se simplifient beaucoup en utilisant des expressions approchées des équations d'Ebers-Moll, correspondant au régime de fonctionnement normal caractérisé par $V_{eb}>>kT/e$ et $-V_{cb}>>kT/e$,

$$I_e \approx I_{s1}\, e^{eV_{eb}/kT} \tag{3-42-a}$$

$$I_c \approx -\alpha\, I_e \tag{3-42-b}$$

$$I_b \approx -\left(1-\alpha\right)I_{s1}\, e^{-eV_{be}/kT} \tag{3-43-a}$$

$$I_c \approx \beta I_b \tag{3-43-b}$$

Avec

$$\beta=\alpha/(1-\alpha) \tag{3-44}$$

Les caractéristiques d'entrée représentent $I_b(V_{be})$ à V_{ce}=Cte. L'expression (3-43-a), montre que I_b est indépendant de V_{ce} de sorte que, comme dans le montage base commune, la caractéristique d'entrée est unique. L'expression (3-43-a) est comparable à l'expression (3-38). La différence essentielle qui existe entre les deux types de montage réside dans l'ordre de grandeur du courant d'entrée. En raison du fait que α est voisin de 1, le courant I_b est très inférieur au courant I_c. ***L'impédance d'entrée du transistor est donc beaucoup plus importante en montage émetteur commun qu'en montage base commune***. La caractéristique d'entrée est représentée dans le troisième quadrant de la figure (3-9).

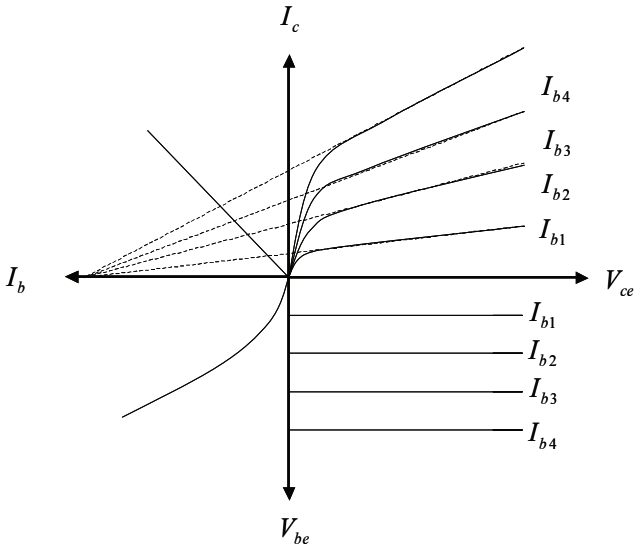


Figure 3-9 : Réseaux de caractéristiques du transistor bipolaire en montage émetteur commun

Les caractéristiques de sortie sont données par $I_c(V_{ce})$ à I_b=Cte, or l'expression (3-43-b) montre que I_c paraît indépendant de V_{ce} de sorte que les caractéristiques devraient être horizontales. Toutefois l'expression (3-44), associée au fait que $\alpha \approx 1$, montre que le coefficient β est très sensible aux faibles variations de α. Or α est fonction de V_{ce} par l'intermédiaire de l'effet Early. Il en résulte que le courant collecteur I_c est fonction de la tension V_{ce} à travers l'effet Early.

Les expressions (3-29-a et c) permettent d'écrire la relation

$$\alpha\left(J_{s10}+J_s^{'}\right)=\alpha_0 J_{s10}+J_s^{'}$$

soit

$$\alpha=\alpha_0\frac{J_{s10}}{J_{s10}+J_s^{'}}+\frac{J_s^{'}}{J_{s10}+J_s^{'}}$$

β s'écrit, en remplaçant les densités de courant J par les courants I

$$\beta=\frac{\alpha}{1-\alpha}=\frac{\alpha_0}{1-\alpha_0}+\frac{I_s^{'}/I_{s10}}{1-\alpha_0} \tag{3-45}$$

L'équation (3-43-b) s'écrit alors

$$I_c=\frac{\alpha_0}{1-\alpha_0}I_b+\frac{I_b/I_{s10}}{1-\alpha_0}I_s^{'} \tag{3-46}$$

Ainsi pour I_b=Cte, le premier terme de l'expression (3-46) est constant et le deuxième terme varie linéairement avec $I_s^{'}$. Or $I_s^{'}$ est fonction de la largeur effective de la base (Eq. 3-28), qui est elle-même fonction de V_{ce} à travers x_c. De plus, dans la mesure où le coefficient $\alpha J_{s1}\mu$ varie peu, l'intégration de l'expression (3-32) montre que le courant $I_s^{'}$ varie linéairement avec la tension V_{cb} et par suite avec la tension V_{ce}. Il en résulte que l'expression (3-46) peut s'écrire sous la forme

$$I_c = \beta_0 I_b + K I_b V_{ce} \tag{3-47}$$

Le courant de sortie I_c varie donc linéairement avec la tension de sortie V_{ce} avec une pente proportionnelle au courant d'entrée I_b. Le réseau de caractéristiques est représenté dans le premier quadrant de la figure (3-9). Pour des valeurs faibles de la tension V_{ce} les approximations à partir desquelles sont établies les expressions (3-43) ne sont plus valables. L'expression (3-47) montre que les extrapolations des parties linéaires des caractéristiques convergent, quel que soit I_b , en un même point d'abscisse $V_a = -\beta_0 / K$ de l'axe des tensions, V_a est la *tension d'Early.*(Fig. 3-9). Les caractéristiques de transfert sont données par la relation (3-43-b). Compte tenu de l'expression (3-44) avec $\alpha \approx 1$, ***le gain en courant β du montage émetteur commun est beaucoup plus important que le gain en courant -α du montage base commune.***

En outre, en raison de l'effet Early, β est donné par l'expression (3-45) qui s'écrit dans la mesure où $I_s^{'}$ est proportionnel V_{ce}.

$$\beta = \beta_0 + K V_{ce} \tag{3-48}$$

Le gain en courant augmente linéairement avec la tension de polarisation V_{ce}. La caractéristique de transfert en courant est représentée pour une valeur donnée de V_{ce}, dans le deuxième quadrant de la figure (3-9).

Comme dans le cas du montage base commune, la tension V_{be} est peu sensible aux variations de la tension V_{ce} les caractéristiques sont pratiquement horizontales (Fig. 3-9).

3.5.4. *Montage collecteur commun*

Il faut faire apparaître explicitement les lois de variation des courants I_b et I_e en fonction des tensions V_{bc} et V_{ec} à partir des équations (3-37). A partir des expressions simplifiées correspondant au régime de fonctionnement normal, on obtient les relations

$$I_b = -(1-\alpha) e^{eV_{ec}/kT} \; e^{-eV_{bc}/kT} \tag{3-49-a}$$

$$I_e = -\frac{1}{1-\alpha} I_b = -(\beta+1) I_b \approx -\beta I_b \tag{3-49-b}$$

A partir de ces relations on peut représenter les différents réseaux de caractéristiques. Nous remarquerons simplement que le gain en courant qui est donné ici par -(β + 1) est sensiblement égal, et de signe opposé, au gain en courant en montage émetteur commun dans la mesure où $\alpha \approx 1$. En outre, en raison du facteur (1 - α) dans l'expression (3-49-a), le courant d'entrée est faible et par suite l'impédance d'entrée est importante. ***En résumé, le montage collecteur commun présente une forte impédance d'entrée et une faible impédance de sortie. Le gain en tension est peu différent de un. Ce montage est utilisé comme étage de sortie dans un systéme électronique pour pouvoir commander d'autres étages***.

3.6. Sources de bruit dans un transistor

Le transistor est vu comme deux jonctions pn en série. Les courants traversant ces jonctions seront donc soumis à des fluctuations comme il a été expliqué dans le chapitre 1. Quand il s'agit d'amplifier des signaux de faibles valeurs (signal en provenance d'une antenne ou d'un capteur), ces sources de bruit peuvent avoir des valeurs proches du signal. Il est alors indispensable de les mimimiser. Cela explique l'importance de ce paragraphe consacré au bruit dans un transistor bipolaire. Avant de calculer le bruit d'un transistor, il faut établir le bruit de grenaille associé au courant traversant une jonction.

3.6.1. *Bruit de grenaille d'une diode*

Dans un premier temps exprimons le bruit de grenaille d'un courant traversant une jonction pn. Le courant s'écrit

$$I = I_0\left[\exp^{\frac{eV}{kT}} - 1\right]$$

Exprimons la densité spectrale associée notée S en appliquant les résultats du chapitre 1-7, en particulier l'équation 1-277. On suppose les deux termes non corrélés.

$$S_I = 2e\left[I_0 \exp^{\frac{eV}{kT}} + I_0\right] \tag{3-50}$$

Soit

$$S_I = 2e\left[I + 2I_0\right]$$

Comme le terme I_0 est faible devant I, il est souvent négligé.

3.6.2. *Bruit de grenaille du bipolaire*

Pour donner les expressions des densités spectrales de bruit du transistor, il faut revenir à la composition des courants d'émetteur et de collecteur en diverses contributions, chacune d'entre elles étant liée à un type de porteurs particulier. Dans le cas d'un transistor pnp, les courants sont principalement des courants de trous. On écrit donc à partir des relations 3-37.

$$I_e = I_{s1}\left(e^{eV_{eb}/kT} - 1\right) - \alpha_i I_{s2}\left(e^{eV_{cb}/kT} - 1\right)$$

$$I_c = -\alpha I_{s1}\left(e^{eV_{eb}/kT} - 1\right) + I_{s2}\left(e^{eV_{cb}/kT} - 1\right)$$

Ces relations peuvent s'écrire également en supposant que la tension entre base et collecteur est très supérieure à 25 mV.

$$I_e = I_{s1}\left(e^{eV_{eb}/kT} - 1\right) + \alpha_i I_{s2} = I_{s1} e^{eV_{eb}/kT} + \alpha_i I_{s2} - I_{s1}$$

$$I_c = -\alpha I_{s1}\left(e^{eV_{cb}/kT} - 1\right) - I_{s2} = -\alpha I_{s1} e^{eV_{eb}/kT} + \alpha I_{s1} - I_{s2}$$

On écrira les courants sous une forme plus simple et de plus on orientera le courant de collecteur vers l'extérieur. Les résultats du chapitre 2 permettent d'écrire les courants traversant le dispositif comme la somme de courants de porteurs minoritaires donc des trous pour la zone contrale de type n.

$$I_e = I_{s1} e^{\frac{eV_{eb}}{kT}} - I_{BE}$$

$$I_c = \alpha\, I_{s1} e^{\frac{eV_{eb}}{kT}} + I_{BC}$$

$$I_b = I_e - I_c$$

On a remplacé $-\alpha_i I_{s2} + I_{s1}$ par I_{BE} et $\alpha I_{s1} - I_{s2}$ par -I_{BC}.
Rappelons que les courants I_{s2} et I_{s1} sont les courants de saturation des deux jonctions et peuvent se calculer par les courants de minoritaires (équation 2-67).
Ces courants s'écrivent aussi en posant

$$I_{C0} = \alpha\, I_{BE} + I_{BC}$$

$$I_c = \alpha\, I_e + I_{C0}$$

Dans la suite on utilisera les relations suivantes pour les densités spectrales (faciles à démontrer à partir de la définition). Rappelons que la densité spectrale d'intercorrélation de deux signaux *x* et *y* est définie par

$$S_{x,y} = \lim_{T \to \infty} \frac{1}{2T} E\left[X_T(f) Y_T^*(f)\right]$$

On notera $S_{x,x}$ simplement S_x

$$S_{x+y,z+w} = S_{x,y} + S_{x,z} + S_{y,z} + S_{y,w}$$
$$S_{x,ay} = a\, S_{x,y}$$

En particulier

$$S_{x,ax+y} = a\, S_x + S_{x,y}$$
$$S_{ax+by} = a^2\, S_x + b^2 S_y + ab\left(S_{x,y} + S_{y,x}\right)$$

Comme $S_{x,y}$ et $S_{y,x}$ sont complexes conjugés

$$S_{ax+by} = a^2\, S_x + b^2 S_y + 2ab\,\mathrm{Re}\left(S_{x,y}\right)$$

En appliquant ces régles de calcul on obtient

$$S_{i_e}(f) = 2e\left[I_{s1} \exp^{\frac{eV_{eb}}{kT}} + I_{BE} \right] = 2e\left[I_e + 2I_{BE}\right]$$

$$S_{i_c}(f) = 2e\left[\alpha\ I_{s1} \exp^{\frac{eV_{eb}}{kT}} + I_{BC} \right] = 2e\left[I_c\right]$$

On peut également calculer la densité spectrale d'intercorrélation entre les bruits des courants d'émetteur et de collecteur. Ces deux courants sont en effet formés à partir d'un courant commun $\alpha\ I_{ES} \exp^{\frac{eV_{EB}}{kT}}$, les termes additionnels sont supposés non corrélés.

$$S_{i_e i_c}(f) = 2e\left[\alpha\ I_{s1} e^{\frac{eV_{eb}}{kT}} \right]$$
$$= 2e\left[I_c - I_{bc}\right]$$

Les résultats obtenus sont cependant difficilement utilisables en pratique car les deux bruits sont corrélés. Il est d'usage de représenter le bruit d'un composant par deux sources de bruit non corrélées, une source de courant et une source de

tension. Le calcul qui suit montre cette transformation. On se place dans le cas du montage base commune mais les résultats sont transposables aux autres montages.

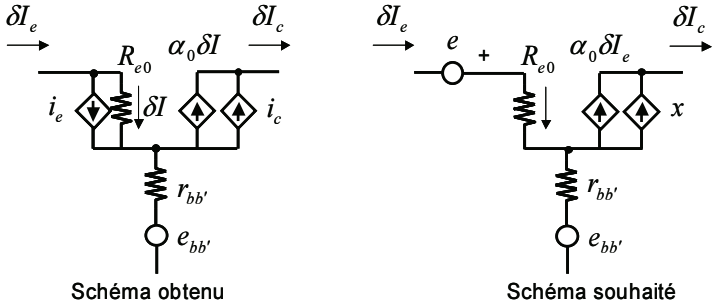


Figure 3-10 : Schémas de représentation du bruit pour un montage base commune

La figure 10 montre le schéma que nous avons obtenu et celui que nous souhaitons obtenir. Les variations des courants sont notés δ et les courants de bruit i_e et i_c .

La résistance d'accès à la base $r_{bb'}$ et son bruit thermique sont ajoutés au schéma de base. Cet effet dû à la technologie est d'une grande importance.
Le coefficient α_0 est le gain dynamique en courant, voisin du terme α. Le terme R_{e0} est l'impédance d'entrée dynamique du montage. L'équivalence des deux montages permet d'écrire.

$$\begin{aligned} \alpha_0 \delta I_e + x &= \alpha_0 \delta I + i_c \\ i_e + \delta I &= \delta I_E \\ -e + R_{e0} \delta I_e &= R_{e0} \delta I \end{aligned}$$

Soit

$$\begin{aligned} x &= -\alpha_0 i_e + i_c \\ e &= R_{e0} i_e \end{aligned} \qquad (3\text{-}51)$$

Il suffit maintenant de calculer les densités spectrales en supposant α et α_0 égaux.

$$\begin{aligned} S_{-\alpha_0 i_e + i_c} &= S_{i_c} + \alpha^2 S_{ie} - 2\alpha \ \mathrm{Re}\left(S_{i_e i_c}\right) \\ &= 2e\left[\alpha\, I_e + I_{C0}\right] + 2e\alpha^2\left[I_e + 2I_{BE}\right] - 4e\alpha^2\left[I_e + I_{BE}\right] \\ &= 2e\left[\alpha\, I_e - \alpha^2 I_e + I_{C0}\right] \\ &= 2e\left[\alpha\,(1-\alpha)\, I_e + I_{C0}\right] \approx 2eI_{C0} \end{aligned} \qquad (3\text{-}52)$$

On obtient

$$S_e = R_{e0}{}^2 S_{ie} = \frac{kT}{e}\frac{1}{I_e + I_{BE}} R_{e0} 2e\left(I_e + 2I_{BE}\right) = 2kTR_{e0}\frac{I_e + 2I_{BE}}{I_e + I_{BE}} \qquad (3\text{-}53)$$
$$\approx 2kTR_{e0}$$

Il faut maintenant vérifier que ces deux sources sont non corrélées.

$$S_{R_{e0}i_e \,-\alpha\, i_e + i_c} = R_{e0}\left[S_{ie\,ic} - \alpha\, S_{i_e}\right] = R_{e0}\left[2e\left(I_C - I_{BC}\right) - 2e\alpha\,\left(I_E + 2I_{BE}\right)\right]$$
$$= R_{e0} 2e\left[I_c - I_{BC} - \alpha\, I_e - 2\alpha\, I_{BE}\right]$$
$$\approx R_{e0} 2e\left[I_{CO} - I_{BC} - \alpha\, I_{BE} - I_{BE}\right]$$
$$= -R_{e0} 2eI_{BE}$$
$$= -2kT\frac{I_{BE}}{I_{BE} + I_E} \approx 0$$

En résumé le transistor base commune génère une tension de bruit et un courant de bruit non corrélés et représentés sur la figure 3-11.

La différence entre le gain dynamique et le terme α peut être mis en évidence à partir des relations suivantes.

$$I_c = \alpha\, I_c + I_{C0}$$
$$\alpha_0 = \frac{\partial I_c}{\partial I_e} = \alpha\ + I_E \frac{\partial \alpha}{\partial I_e} \approx \alpha$$

Il est maintenant possible de représenter le bruit par des sources en entrée du transistor en plaçant cette fois la source de bruit en entrée. Cela est possible puisque le gain en courant est environ un. Il est donc indifférent de placer la source de courant en entrée ou en sortie. Le schéma final 3-12 est classique.

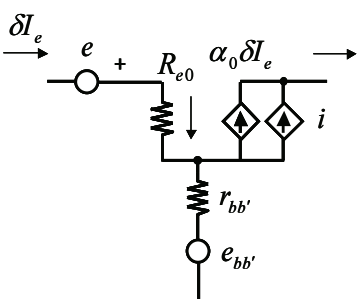


Figure 3-11 : Sources de bruit d'un bipolaire

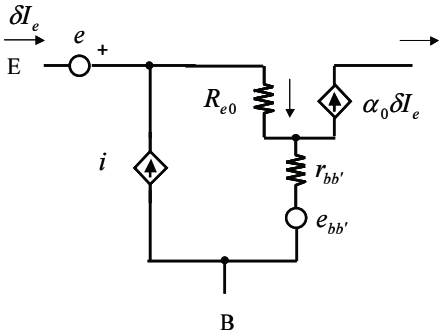


Figure 3-12 : Sources de bruit d'un bipolaire

Les densités spectrales des deux sources sont respectivement

- Source de tension : $S_e = 2kTR_{e0}$
- Source de courant : $S_i \approx 2eI_b$
- Résistance de base : $S_{ebb} = 4kTr_{bb'}$

On peut prouver que cette représentation est également valable pour les montages émetteur commun et base commune. Il est ensuite possible de calculer le bruit en sortie d'un schéma quelconque en traitant ces sources de manière séparée et en multipliant la densité spectrale par le carré du module de la fonction de transfert.

EXERCICES DU CHAPITRE 3

- **Exercice 1 : Calcul des courants**

Un transistor pnp a les dopages suivants : 10^{19}, 10^{17} et 5.10^{15} /cm^3. Les temps de vie sont respectivement 1 /100, 1/10 et 1 microseconde. La section est de 0.05 mm^2. Les coefficients de diffusion sont de 1 cm^2/s dans l'émetteur, 10 cm^2/s dans la base et 2 cm^2/s dans le collecteur. L'épaisseur de la base est de 0,5 micron. La section est de 0,05 mm^2. La tension émetteur base est choisie à 0,6 V. La jonction base collecteur est fortement polarisée. Calculez les courants de trous et d'électrons dans l'émetteur ainsi que dans le collecteur. En déduire le courant de base et le gain en courant.

- **Exercice 2 : Calcul du gain**

Montrez qu'une valeur simplifiée du gain en courant est $\alpha = 1 - \frac{w^2}{2L_p{}^2}$ dans lequel w est l'épaisseur de la base et $D_p = \sqrt{D_p \tau_p}$ dans la base.

- **Exercice 3 : Montage collecteur commun**

Calculez le gain en courant et en tension du montage collecteur commun. Calculez les impédances d'entrée et de sortie.

- **Exercice 4 : Transistor npn**

Reprendre les calculs de courants de l'exercice 1 pour un transistor npn.

- **Exercice 5 : Bruit d'un transistor**

Montrez que les sources de bruit ramenées en entrée (tension et courant) d'un transistor monté en émetteur commun sont approximativement les mêmes que celles du montage base commune.

- **Exercice 6 : Bruit d'un transistor**

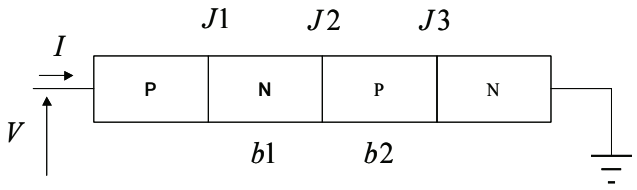
On suppose que le gain dynamique α_0 et le gain statique α sont différents. Calculez la densité spectrale d'intercorrélation.

$$S_{R_e 0 i_e \; -\alpha_0 i_e + i_c} = -2kT \frac{\alpha \; I_{s1}}{I_e + I_{BE}} + 2kT(\alpha \; - \alpha_0) \frac{I_e + 2I_{BE}}{I_e + I_{BE}}$$

En déduire que les sources restent non corrélées.

- **Exercice 7 : Diode de Shockley**

C'est une structure pnpn. Elle est donc formée de trois jonctions 1,2 et 3 et la tension *V* est appliquée entre la première région p et la dernière région n.



On suppose que la jonction 3 est court-circuitée.

Ecrire les équations de Ebers-Moll et posez $K_{12} = \frac{I_{s2}}{I_{s1}} \frac{1-\alpha_{1i}}{1-\alpha_1}$

Montrez que le courant *I* traversant le dispositif s'écrit

$$I = \alpha_1 I_{s1} \left(\frac{1+K_{12}}{1+K_{12} e^{-\frac{eV}{kT}}} - 1 \right) - I_{s2} \left(\frac{1+K_{12}}{1+K_{12} e^{-\frac{eV}{kT}}} e^{-\frac{eV}{kT}} - 1 \right)$$

En déduire les caractéristiques pour les valeurs positives et négatives de la tension appliquée.

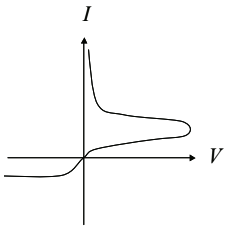
On suppose maintenant que la jonction 1 est court-circuitée.
Calculez le courant traversant le dispositif. Comparez au cas précédent.

On considère finalement le système sans court-circuit.
Le courant de saturation de la jonction 2 est séparé en un courant de trous I_{j2p} et un courant d'électrons I_{j2n}. On note α_1 et α_2 les gains des deux transistors pnp et npn et α_1' la fractions du courant de trous de la jonction 2 injectés de la base du transistor 2 dans la base du transistor 1 atteignant la jonction J1 et α_2' la fraction du courant d'électrons injecté de la base du transistor 1 dans la base du transistor 2 et atteignant la jonction J3. Ce ne sont pas exactement les gains inverses.
Montrez que

$$I = \frac{1-\alpha_1\alpha_1'}{1-(\alpha_1+\alpha_2)} I_{j2p} + \frac{1-\alpha_2\alpha_2'}{1-(\alpha_1+\alpha_2)} I_{j2n}$$

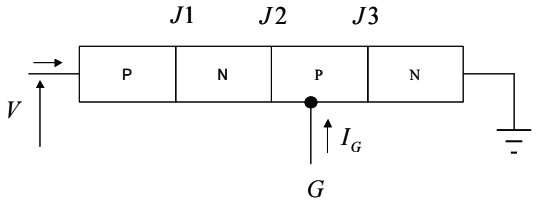
On suppose que les deux composantes I_{s2p} et I_{s2n} sont égales et que les taux d'injection des électrons et des trous dans J2 sont égaux.
Etudiez *I* en fonction de la valeur $\alpha_1+\alpha_2$ devant 1.

Montrez comment on passe de l'état direct-bloqué (J2 bloquée) à l'état direct-passant (J2 passant) et expliquez la caractéristique ci-dessous



- **Exercice 8 : Le thyristor**

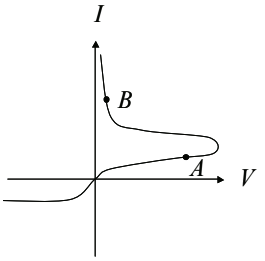
C'est une diode de Schockley mais avec une polarisation appliquée sur la zone p qui est dans le dispositif. Cette zone est la grille de commande G



Montrez que le courant entrant dans la première zone p de la structure pnpn est

$$I = \frac{1-\left(\alpha_1\alpha_1' + \alpha_2\alpha_2'\right)/2}{1-\left(\alpha_1 + \alpha_2 + \alpha_2 \dfrac{I_G}{I}\right)} I_{j2}$$

Une impulsion de courant sur G peut donc changer le signe du dénominateur et faire passer le dispositif dans l'état direct-passant (passage de A à B).



Calculez le courant de déclenchement et montrez que le dispositif peut permetre de commander des courants très élevés (des centaines d'ampères).

4. Contact métal-semiconducteur Diode Schottky

Ce chapitre présente les phénomènes physiques à l'interface métal-semiconducteur. Cette étude est fondamentale car les dispositifs semiconducteurs sont en relation avec l'extérieur par des contacts métalliques déposés sur le semiconducteur. Ces contacts jouent souvent un rôle important dans le fonctionnement global du composant. Dans certains cas (diode Schottky) ils constituent la région active du composant.

Le chapitre est formé de trois parties permettant d'expliquer la structure des bandes d'énergie dans le dispositif et de calculer les courants traversant le dispositif quand une tension extérieure est appliquée. Les méthodes permettant de calculer les bandes à l'interface de deux matériaux différents seront présentées dans la première partie de ce chapitre. Elles sont d'une importance fondamentale dans l'étude des dispositifs solides.

Les notions développées dans le paragraphe 1-6 du chapitre 1 seront largement mises en application.

4.1. Diagramme de bandes
4.4. Zone de charge d'espace
4.3. Caractéristiques courant-tension

4.1. Diagramme des bandes d'énergie

4.1.1. *Considérations générales sur les niveaux énergétiques*

Nous abordons ici l'étude des hétérostructures, qui résultent de l'association de matériaux différents : métal-semiconducteur, semiconducteur1-semiconducteur2, métal-isolant-semiconducteur,... Dans l'étude des structures réalisées dans un matériau unique nous avons vu qu'un principe intangible permettait d'établir le diagramme énergétique à l'équilibre thermodynamique : le niveau de Fermi est horizontal dans toute la structure. Ce principe reste évidemment valable dans les hétérostructures. Néanmoins pour comprendre simplement le processus de mise en équilibre de la structure lorsque l'on associe les différents matériaux nous devons établir le diagramme énergétique de la *prè-structure*, préalablement à l'association des matériaux. ***A ce stade, les matériaux étant isolés les uns des autres, le niveau de Fermi est horizontal dans chacun d'eux mais ne constitue pas une référence commune. Ces niveaux de Fermi n'ont à priori aucune relation entre eux***. En fait si les matériaux sont électriquement neutres, l'énergie potentielle d'un électron dans le vide c'est-à-dire à l'extérieur des matériaux, est la même quelle que soit sa position par rapport aux matériaux (nous négligeons évidemment l'effet Schottky dû au potentiel image traduisant la polarisation du matériau par l'électron). En d'autres termes, si nous appelons *NV* (Niveau du Vide) l'énergie potentielle d'un électron dans le vide au voisinage des différents matériaux, *NV* est la même pour tous les matériaux et peut servir de référence. Rappelons que l'énergie potentielle est définie à une constante près. En général, elle est choisie nulle loin de tous les matériaux. Quand le matériau comporte des zones chargées (zones de charge d'espace) elles peuvent modifier le niveau du vide à proximité. Cela explique que le niveau du vide peut varier dans l'espace. En outre, pour chacun des matériaux l'énergie potentielle d'un électron dans le vide au voisinage du matériau est située à $e\phi$ (travail de sortie) au-dessus du niveau de Fermi. On peut donc, dans la *pré-structure,* positionner les différents niveaux de Fermi les uns par rapport aux autres.

Ainsi donc, à l'équilibre thermodynamique nous disposerons de deux références énergétiques :

- Avant association, le niveau du vide *NV* est horizontal, les niveaux de Fermi E_{Fj} horizontaux dans chaque matériau, se distribuent suivant les travaux de sortie de chacun.
- Après association, le niveau de Fermi E_F est horizontal dans toute la structure, les niveaux du vide NV_j se distribuent en fonction des travaux de sortie. Notons que dans chacun des matériaux NV_j ne sera pas forcément horizontal car le travail de sortie évoluera au gré de la redistribution électronique, notamment au voisinage des interfaces.

Lorsqu'un métal et un semiconducteur sont au contact, il existe à l'interface une barrière de potentiel donnée par l'expression (1-233)

$$E_b = e\phi_m - e\chi$$

où $e\phi_m$ représente le travail de sortie du métal et $e\chi$ l'affinité électronique du semiconducteur. La structure de bandes au voisinage de l'interface est conditionnée par la différence éventuelle des travaux de sortie du métal et du semiconducteur.

La structure de bande de matériaux reliés entre eux par des sources de potentiel se détermine comme suit :

- Les niveaux de Fermi se positionnent relativement en fonction des différences de potentiel appliquées.
- Les bandes de conduction et de valence se positionnent en profondeur dans le semiconducteur en fonction des dopages et en fonction du gap pour les isolants. Les électrons atteignent le niveau de Fermi pour les métaux.
- Les valeurs des travaux de sortie et des affinités (données pour chaque matériau) permettent de placer le niveau du vide aux interfaces par rapport au niveau de Fermi (cas des métaux) et par rapport au niveau de la bande de conduction (cas des semiconducteurs et des isolants). On peut alors par continuité tracer le niveau du vide au voisinage des interfaces.
- Les bandes de conduction et de valence se déforment alors pour suivre le niveau du vide. Ces courbures correspondent aux zones de charge d'espace aux voisinage des interfaces.

Rappelons que les électrons passent d'un matériau à un autre si le travail de sortie du premier est inférieur à celui du deuxième.

4.1.2. *Premier cas* : *les travaux de sortie sont égaux*, $\phi_m = \phi_s$

Si on prend le niveau de l'électron dans le vide NV comme niveau de référence, le diagramme énergétique dans chacun des matériaux est représenté sur la figure (4-1-a). Lorsque le métal et le semiconducteur sont mis au contact, ils peuvent échanger de l'énergie et constituent un seul système thermodynamique. La distribution statistique des électrons dans ce système est alors représentée par un niveau de Fermi unique, les niveaux E_{Fm} et E_{Fs} s'alignent. En fait dans la mesure où ici $\phi_m = \phi_s$ ces niveaux sont alignés en l'absence de contact, de sorte que l'équilibre thermodynamique est réalisé sans aucun échange d'électron. Le diagramme énergétique est représenté sur la figure (4-1-b). La barrière de potentiel E_b s'établit au niveau de l'interface. Ce diagramme énergétique est valable quel que soit le type (p ou n) du semiconducteur, dans la mesure où la condition $\phi_m = \phi_s$ reste respectée. Les bandes sont horizontales, on dit que le système est en régime de *bandes plates*.

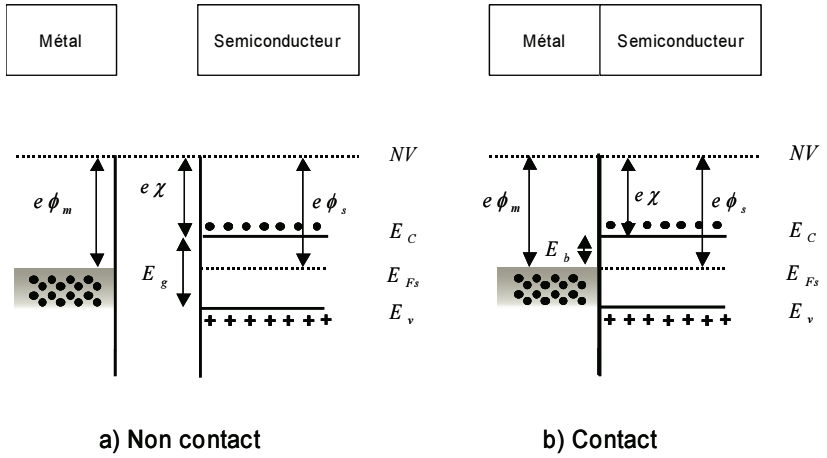


Figure 4-1 : Contact métal-semiconducteur avec $\phi_m=\phi_s$

4.1.3. *Deuxième cas : travail de sortie du métal supérieur, $\phi_m > \phi_s$*

Le diagramme énergétique dans chacun des matériaux est représenté sur la figure (4-2). Lorsque les deux matériaux sont mis au contact, le travail de sortie du semiconducteur étant inférieur à celui du métal, les électrons passent du semiconducteur dans le métal. Le système se stabilise à un régime d'équilibre défini par l'alignement des niveaux de Fermi. Le diagramme énergétique résultant est différent suivant le type de semiconducteur.

a. Semiconducteur de type n

Les électrons qui passent du semiconducteur vers le métal, entraînent des modifications énergétiques dans chacun des matériaux. Dans le semiconducteur, une *zone de déplétion* se crée, les ions donneurs ionisés N_d^+ ne sont plus compensés par les électrons, il apparaît une charge d'espace positive. D'autre part la distance bande de conduction-niveau de Fermi, qui traduit la population électronique, est plus grande au voisinage de l'interface que dans la région neutre du semiconducteur $\phi'_F>\phi_F$. Le niveau de Fermi étant horizontal, il en résulte une courbure des bandes vers le haut (Fig. 4-3). Dans le métal, il apparaît une accumulation d'électrons à l'interface. Le niveau du vide varie à cause de ces zones chargées. A cette double charge d'espace sont associés un champ électrique et une tension de diffusion V_d qui, comme dans le cas de la jonction pn, équilibrent les forces de diffusion et déterminent l'état d'équilibre.

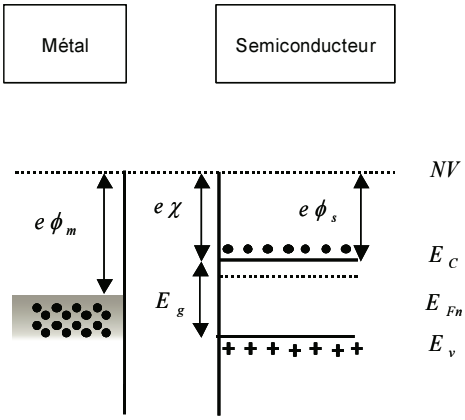


Figure 4-2 : Travaux de sortie différents

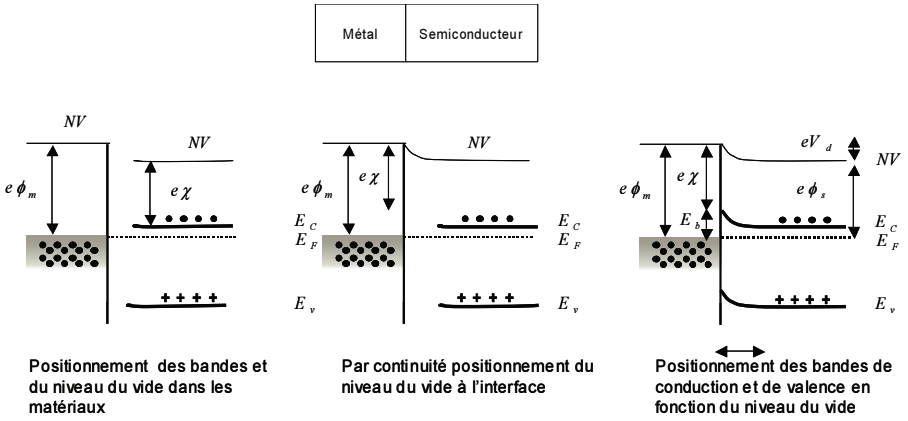


Figure 4-3 : Contact métal-semiconducteur(n) avec $\phi_m > \phi_s$ à l'équilibre thermodynamique

Notons que le nombre de charges positives développées dans le semiconducteur est égal au nombre de charges négatives développées dans le métal. Ces dernières sont des charges d'accumulation, la densité d'états dans le métal étant de l'ordre de 10^{22} cm^{-3}, ces charges se développent à la surface du métal. Dans le semiconducteur, ces charges sont des charges de déplétion dues aux ions donneurs, la densité de ces donneurs étant typiquement de l'ordre de 10^{16} à 10^{18} cm^{-3}, cette charge d'espace est relativement étalée à l'intérieur du semiconducteur, ce qui entraîne une extension des courbures de bandes.

Polarisons la structure par une tension semiconducteur-métal V négative (Fig. 4-4-a). La bande de conduction du semiconducteur s'élève de eV, la courbure diminue. Ainsi la barrière semiconducteur→métal diminue alors que la barrière métal→semiconducteur reste inchangée. L'équilibre est rompu, les électrons

diffusent du semiconducteur vers le métal et créent un courant *I* du métal vers le semiconducteur. Si on augmente encore la tension de polarisation on atteint le régime de bandes plates lorsque $V=V_d$

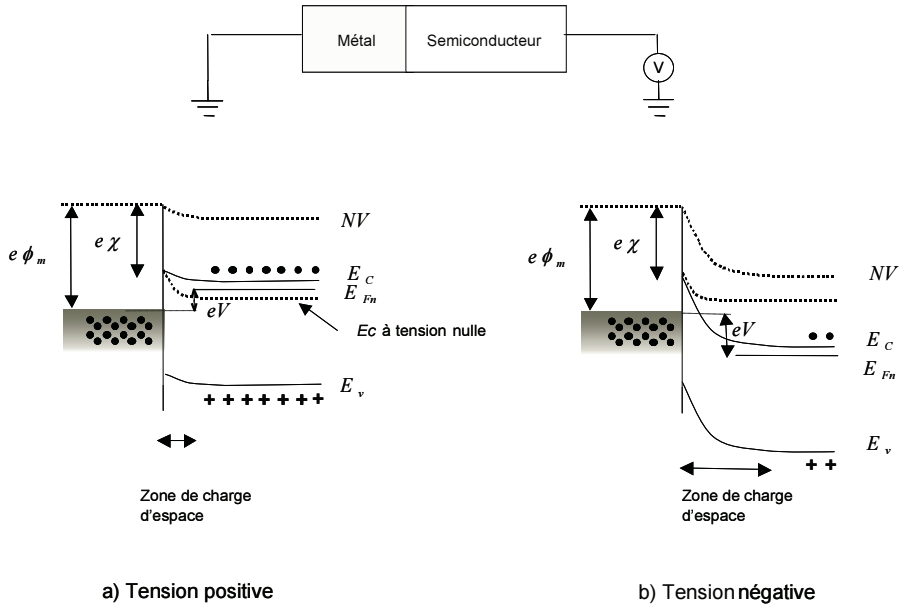


Figure 4-4 : Contact métal-semiconducteur(n) avec $\phi_m > \phi_s$ sous polarisation. a) $V_{sc}-V_m < 0$. b) $V_{sc}-V_m > 0$

Polarisons la structure par une tension semiconducteur-métal positive (Fig. 4-4-b). La bande de conduction du semiconducteur est abaissée, ce qui augmente la hauteur de la barrière qui s'opposait à la diffusion des électrons. La structure est polarisée en inverse. La structure métal-semiconducteur (n) avec $\phi_m > \phi_s$ constitue donc un contact redresseur. C'est une *diode Schottky*. Notons que la différence des niveaux de Fermi est la différence de potentiel appliquée.

b. Semiconducteur de type p

Lorsque les deux matériaux sont mis au contact les électrons diffusent du semiconducteur vers le métal jusqu'à alignement des niveaux de Fermi. Le diagramme énergétique résultant est représenté sur la figure (4-5). Il apparaît une zone de charge d'espace négative dans le métal, positive dans le semiconducteur. Comme précédemment cette charge d'espace est accompagnée d'une courbure vers le haut des bandes de valence et de conduction.

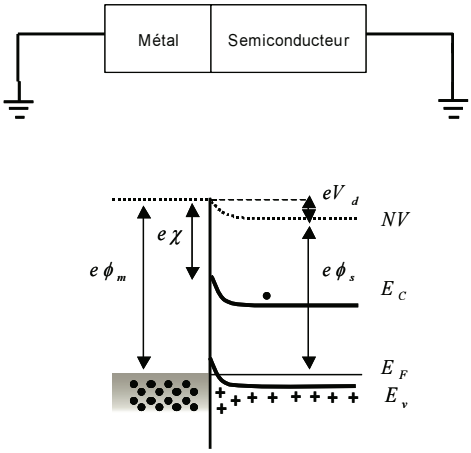


Figure 4-5 : Contact métal-semiconducteur(p) avec $\phi_m > \phi_s$ à l'équilibre thermodynamique

La charge d'espace dans le métal est due à une accumulation d'électrons à la surface. Considérons la charge d'espace positive, développée dans le semiconducteur. La courbure des bandes vers le haut montre que la distance niveau de Fermi-bande de valence diminue près de la surface, ce qui traduit une accumulation de trous. Comme les électrons dans le métal, les trous dans le semiconducteur sont des charges libres qui vont s'accumuler à la surface du semiconducteur. En fait si la densité d'états électroniques dans la bande de conduction du métal est de l'ordre de 10^{22} cm^{-3}, la densité équivalente d'états de la bande de valence du semiconducteur à la température ambiante, est de l'ordre de 10^{19} cm^{-3}. Il en résulte un étalement de la charge d'espace dans le semiconducteur beaucoup plus important que dans le métal. Toutefois cet étalement reste beaucoup plus faible que dans la structure précédente où les charges positives étaient des charges fixes dues aux ions donneurs en densité de l'ordre de 10^{16} à 10^{18} cm^{-3}.

Notons ici que dans l'étude des hétérostructures il faut toujours avoir à l'esprit les ordres de grandeur de trois paramètres, densité d'états électroniques dans la bande de conduction du métal ($\approx 10^{22}$ cm^{-3}), densité équivalente d'états dans les bandes de valence et de conduction d'un semiconducteur à la température ambiante ($\approx 10^{19}$ cm^{-3}), densité de donneurs ou accepteurs dans un semiconducteur dopé ($\approx 10^{16}$ - 10^{18} cm^{-3}). Remarquons cependant que la densité équivalente d'états dans une bande permise du semiconducteur varie avec la température comme $T^{3/2}$. Il en résulte que cette densité est réduite d'un facteur 8 quand on passe de la température ambiante à celle de l'azote liquide, et d'un facteur 600 quand la température devient celle de l'hélium liquide.

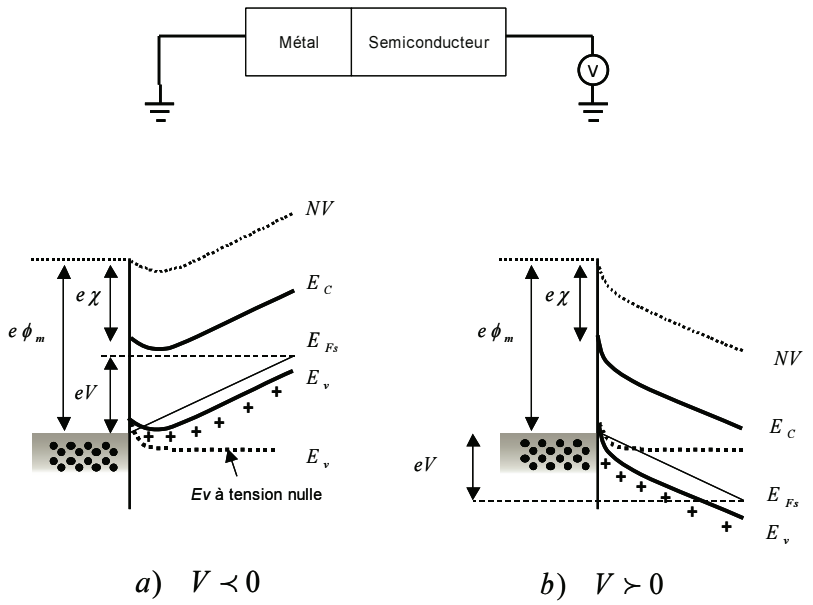


Figure 4-6 : Contact métal-semiconducteur(p) avec $\phi_m > \phi_S$ sous polarisation.
a) $V_{SC}-V_m<0$. b) $V_{SC}-V_m>0$

La différence essentielle entre cette structure métal-semiconducteur (p), et la structure précédente métal-semiconducteur(n), réside dans le fait que la charge d'espace dans le semiconducteur correspond à un régime d'accumulation et non de déplétion. Il en résulte qu'il n'existe pas de zone vide de porteurs, donc isolante, à l'interface. Ainsi lorsque l'on polarise cette structure (Fig. 4-6-a, b), la tension appliquée n'est plus localisée dans la zone de charge d'espace du semiconducteur, comme dans le cas précédent, mais distribuée dans tout le semiconducteur, plus résistif que le reste de la structure. Au niveau de l'interface, l'arrivée ou le départ d'un trou dans le semiconducteur est immédiatement compensée par l'arrivée, ou le départ, d'un électron dans le métal. La constante de temps mise en jeu est la constante de temps diélectrique, l'équilibre est donc pratiquement instantané. Il en résulte que le courant circule librement dans les deux sens au niveau du contact. Le contact est *ohmique*.

4.1.4. *Troisième cas : travail de sortie du métal inférieur, $\phi_m < \phi_s$*

Avant le contact le diagramme énergétique est représenté sur la figure (4-7). Lorsque les deux matériaux sont mis au contact, le travail de sortie du métal étant inférieur à celui du semiconducteur, les électrons sortent du métal pour entrer dans le semiconducteur. Le système évolue jusqu'à alignement des niveaux de Fermi. Le diagramme énergétique est différent suivant le type de semiconducteur.

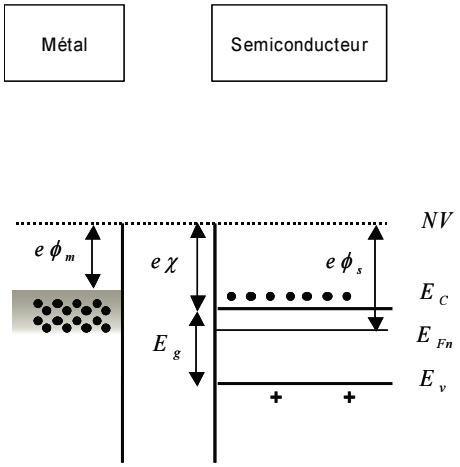


Figure 4-7 : Contact métal-semiconducteur(p) avec $\phi_m<\phi_s$ à l'équilibre thermodynamique

a. Semiconducteur de type n

Les électrons qui passent du métal dans le semiconducteur font apparaître dans le métal un déficit d'électrons localisé à la surface, et dans le semiconducteur une zone d'accumulation très peu étalée. Il en résulte une courbure vers le bas des bandes de valence et de conduction (Fig.4-8).

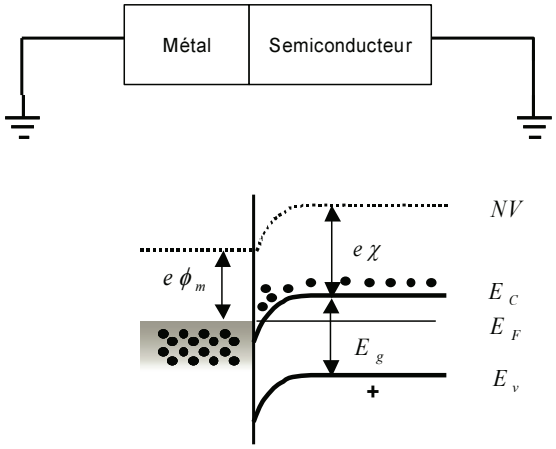


Figure 4-8 : Contact métal-semiconducteur(n) avec $\phi_m<\phi_s$ à l'équilibre thermodynamique

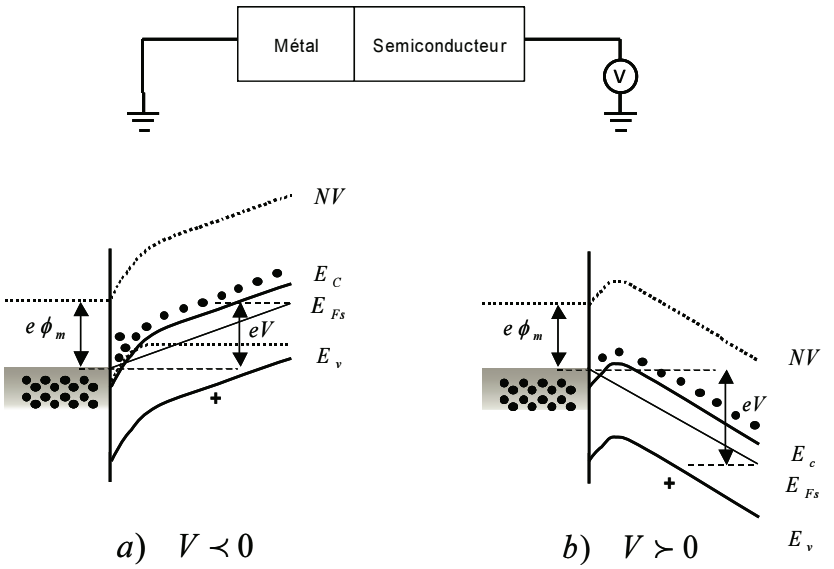


Figure 4-9 : Contact métal-semiconducteur(n) avec $\phi_m < \phi_s$ sous polarisation.
a) $V_{sc}-V_m < 0$. b) $V_{sc}-V_m > 0$

Comme dans le cas précédent, si on polarise la structure (Fig. 4-9-a, b) la tension de polarisation est distribuée dans tout le semiconducteur. Tout électron qui arrive à l'interface dans le semiconducteur passe librement dans le métal et vice versa. Le contact est *ohmique*.

b. Semiconducteur de type p

Les électrons passent du métal dans le semiconducteur. Il apparaît un déficit d'électrons à la surface du métal. Dans le semiconducteur, les électrons qui viennent du métal se recombinent avec les trous créant une zone de déplétion due à la présence des ions N_a^- qui ne sont plus compensés par les trous. Il apparaît ainsi une zone de charge d'espace étalée dans le semiconducteur. Le système évolue jusqu'au moment où le champ et la tension de diffusion résultants, arrêtent la diffusion des électrons (Fig. 4-10). La hauteur de la barrière d'interface, que voient les trous pour passer du semiconducteur au métal, est alors donnée par

$$E_b = E_g + e\chi - e\phi_m$$

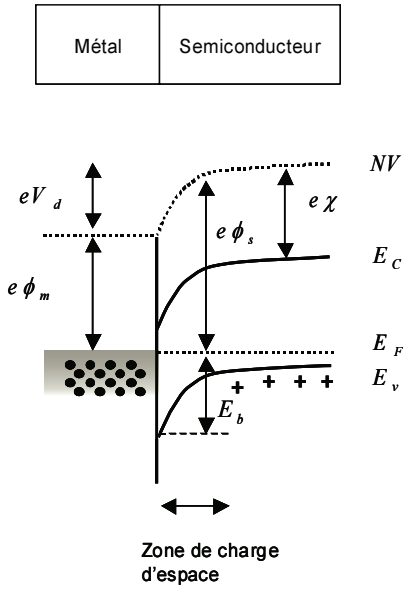


Figure 4-10 : Contact métal-semiconducteur(p) avec $\phi_m<\phi_s$ à l'équilibre thermodynamique

Polarisons la structure, la tension de polarisation se localise au niveau de la zone de déplétion, isolante. Si la tension semiconducteur-métal est négative, les bandes de conduction et de valence s'élèvent, la courbure des bandes augmente.

La barrière de potentiel est augmentée, le courant ne circule pas, la structure est polarisée en inverse (Fig. 4-11-a). Si cette tension est positive, les bandes sont abaissées et la barrière de potentiel que doivent franchir les trous pour passer dans le métal est réduite, le courant circule librement, la structure est polarisée dans le sens passant (Fig. 4-11-b). La structure métal-semiconducteur (p) avec $\phi_m<\phi_s$ constitue donc un contact redresseur, c'est une *diode Schottky.*
En résumé, le contact métal-semiconducteur est ohmique ou redresseur suivant la différence des travaux de sortie et le type du semiconducteur.

Avec $\phi_m>\phi_s$:	***le contact métal-semiconducteur n est redresseur***
	le contact métal-semiconducteur p est ohmique
Avec $\phi_m<\phi_s$:	***le contact métal-semiconducteur n est ohmique***
	le contact métal-semiconducteur p est redresseur

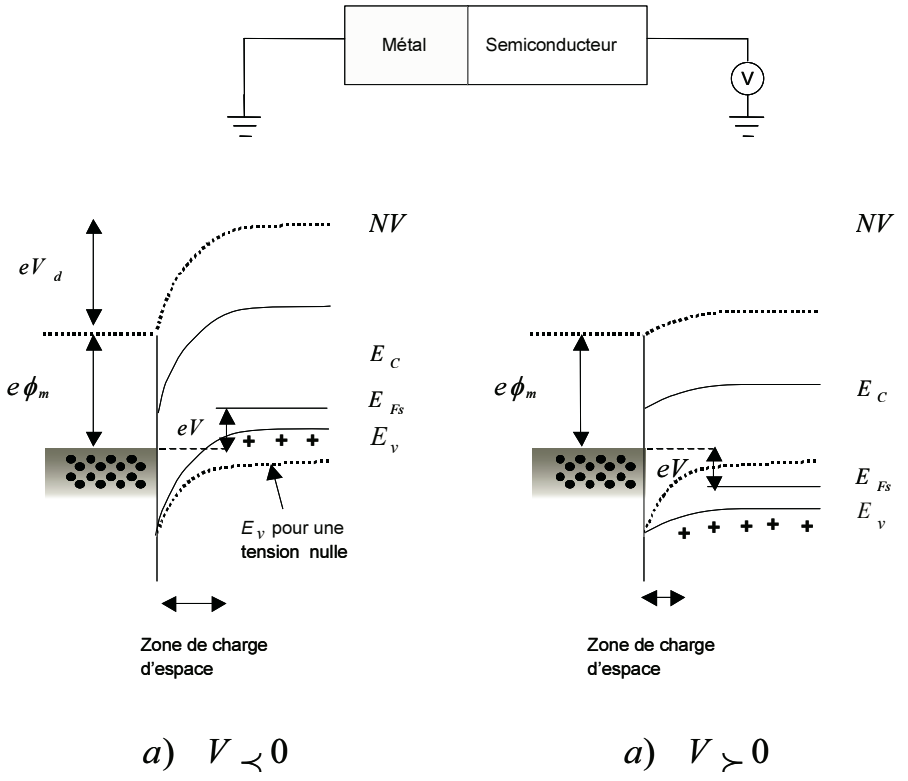


Figure 4-11 : Contact métal-semiconducteur(p) avec $\phi_m<\phi_s$ sous polarisation.

Dans ce qui précède nous avons ignoré la présence éventuelle d'états d'interface. En fait ces états peuvent jouer un rôle très important si les niveaux d'énergie qui leur sont associés sont situés dans le gap du semiconducteur (Par. 1.6.3). En effet, lors de la mise en équilibre de l'interface ces états participent à l'échange de porteurs qui se manifeste entre le métal et le semiconducteur. Il en résulte que le diagramme énergétique de la structure, et par suite la hauteur de la barrière de Schottky, sont conditionnés par la densité et la population de ces états.

Pour prendre en considération le rôle joué par la charge localisée sur les états d'interface on peut écrire les hauteurs de barrière des structures Métal-SC(n) et Métal-SC(p) sous les formes respectives (Cowley A.M. - 1965).

$$E_{bn} = \gamma\left(e\phi_m - e\chi\right) + \left(1-\gamma\right)\left(E_g - e\phi_0\right)$$
$$E_{bp} = \gamma\left(E_g + e\chi - e\phi_m\right) + \left(1-\gamma\right)e\phi_0$$

Le paramètre $\gamma = \varepsilon_i / \left(\varepsilon_i + e^2 \delta_i D_i\right)$, spécifique du couple considéré, pondère les effets respectifs des phénomènes intrinsèques et des états d'interface. Il est fonction de la densité D_i de ces états, supposés répartis uniformément dans le gap du

semiconducteur, ainsi que de l'épaisseur δ_i et de la constante diélectrique ε_i de la couche interfaciale sur laquelle ils sont distribués. Le terme $e\phi_0$ traduit la population initiale de ces états. Il représente la distance du niveau de Fermi au sommet de la bande de valence, à la surface du semiconducteur avant la mise en équilibre de la structure. Dans les expressions précédentes, l'abaissement de la barrière associé au potentiel image est négligé.

- Si la densité D_i d'états d'interface est négligeable, le paramètre γ est voisin de 1 et on retrouve les expressions intrinsèques de la barrière de potentiel.

$$E_{bn} \approx e\phi_m - e\chi$$
$$E_{bp} \approx E_g + e\chi - e\phi_m$$

- Si par contre la densité d'états d'interface est très importante, le paramètre γ devient très faible et les hauteurs de barrière s'écrivent

$$E_{bn} \approx E_g - e\phi_0$$
$$E_{bp} \approx e\phi_0$$

Les hauteurs de barrière deviennent alors indépendantes du travail de sortie du métal et ne sont conditionnées que par la population initiale des états d'interface à travers le paramètre $e\phi_0$. Ceci traduit le fait que l'échange de porteurs se fait dans ce cas entre le métal et les états d'interface. La densité de ces derniers étant très importante, cet échange peut concerner un grand nombre de porteurs de charge sans pour autant entraîner une variation conséquente de la position du niveau de Fermi. Il se produit alors un ancrage du niveau de Fermi à la position $e\phi_0$.

Notons que pour réaliser un contact ohmique, c'est-à-dire satisfaire à la condition $E_b \leq 0$, il faut alors réaliser la condition $e\phi_0 \geq E_g$ dans un semiconducteur de type *n*, ou $e\phi_0 \leq 0$ dans un semiconducteur de type p. En d'autres termes il faut que les états d'interface soient entièrement saturés par les porteurs majoritaires, c'est-à-dire entièrement occupés dans un semiconducteur de type n, et totalement vides dans un semiconducteur de type p. ***Dans la pratique, pour réaliser un contact ohmique sur un semiconducteur de type*** **n** ***(ou*** **p*****) on dépose le métal sur une région préalablement surdopée*** **n^+** ***(ou*** **p^+*****).***

Le paramètre γ mesure la sensibilité de la barrière de Schottky au travail de sortie du métal, $\gamma = \partial E_b / \partial e\,\phi_m$, on peut donc le considérer comme un *indice de comportement de l'interface*. γ=1 lorsque le travail de sortie du métal conditionne entièrement la hauteur de la barrière de Schottky, γ=0 lorsque ce rôle est joué par les états d'interface. La figure (4-12) représente la variation de cet indice avec l'ionicité du semiconducteur. Cette ionicité, qui peut être définie de différentes manières, est ici mesurée par la différence d'électronégativité des constituants du matériau. Pour les semiconducteurs covalents comme le silicium et le germanium,

ou pseudo-covalents comme la plupart des composés III-V, l'indice γ est très faible. La barrière de potentiel E_b est alors peu sensible au travail de sortie du métal car le niveau de Fermi est ancré par les états d'interface. Par contre pour les matériaux plus ioniques, dans lesquels la différence d'électronégativité des constituants est supérieure à 1 eV, le travail de sortie du métal joue un rôle majeur. C'est le cas de la plupart des composes II-VI, et des composés III-V du haut du tableau de Mendéléev. Cette variation de γ avec l'ionicité du matériau est reliée à la localisation des fonctions d'onde des états de valence. La perturbation de ces d'états par l'interface est plus grande lorsque les fonctions d'onde sont plus délocalisées. Ainsi la densité d'états d'interface varie en raison inverse de l'ionicité du matériau. Ceci explique l'absence d'ancrage du niveau de Fermi dans les matériaux ioniques.

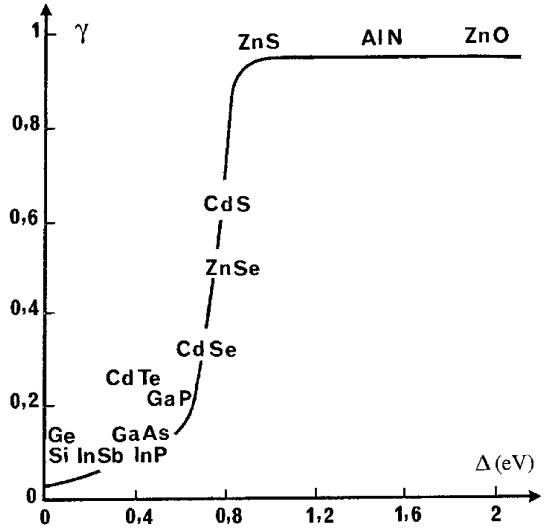


Figure 4-12 : Variation du paramètre γ avec la différence d'électronégativité Δ des constituants du semiconducteur

4.2. Zone de charge d'espace

4.2.1. *Champ et potentiel électrique*

Considérons la structure métal-semiconducteur de type n, avec $\phi_m > \phi_s$. On obtient la distribution du potentiel dans la zone de charge d'espace en intégrant l'équation de Poisson. Nous supposerons que le semiconducteur est homogène, avec une densité de donneurs excédentaires $(N_d\text{-}N_a)_n$ que nous appellerons N_d pour alléger l'écriture. Nous admettrons que tous les donneurs sont ionisés à la température ambiante et que la densité d'états d'interface est négligeable. Nous ferons l'hypothèse de la zone de charge d'espace vide de porteurs et nous appellerons W la largeur de cette zone. Ainsi la densité de charge dans le semiconducteur (Fig. 4-13-a) s'écrit

$0 < x < W$ $\rho(x) = e\,N_d$

$x > W$ $\rho(x) = 0$

L'équation Poisson s'écrit

$$\frac{d^2V(x)}{dx^2} = -\frac{e\,N_d}{\varepsilon_s}$$

En intégrant une première fois avec la condition E=0 pour x>W on obtient

$$\frac{dV(x)}{dx} = -E(x) = -\frac{eN_d}{\varepsilon_s}(x - W) \tag{4-1}$$

Le champ électrique est négatif et varie linéairement dans la zone de charge d'espace (Fig. 4-13-b), sa valeur à l'interface est

$$E_s = -\frac{eN_d}{\varepsilon_s}W \tag{4-2}$$

En intégrant une deuxième fois en prenant l'origine des potentiels à l'interface, on obtient (Fig. 4-13-c)

$$V(x) = -\frac{eN_d}{\varepsilon_s}\left(\frac{x^2}{2} - Wx\right) \tag{4-3}$$

La tension de diffusion résulte de la différence des travaux de sortie du métal et du semiconducteur $V_d = \phi_m - \phi_s$. Cette tension correspond à la différence de potentiel entre la surface du semiconducteur et son volume, c'est-à-dire aux bornes de la zone de charge d'espace

$$V_d = V(x=W) - V(x=0) = -\frac{eN_d}{\varepsilon_s}\left(\frac{W^2}{2} - W^2\right) = \frac{eN_d}{2\varepsilon_s}W^2$$

d'où l'expression de la largeur de la zone de charge d'espace à l'équilibre

$$W = \left(\frac{2\varepsilon_s}{eN_d}V_d\right)^{1/2} = \left(\frac{2\varepsilon_s}{eN_d}(\phi_m - \phi_s)\right)^{1/2} \tag{4-4}$$

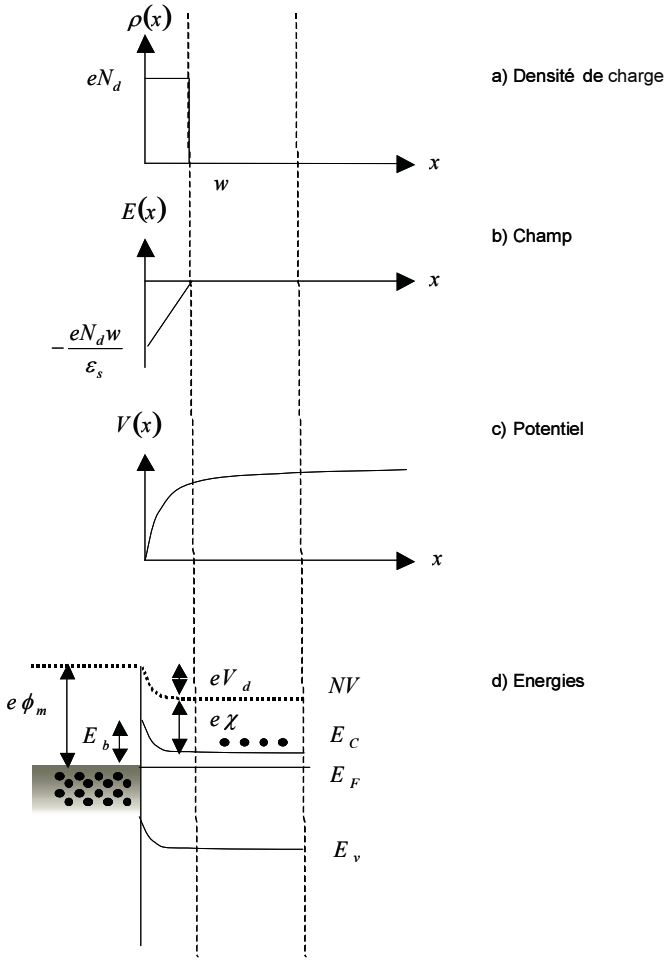


Figure 4-13 : Diode Schottky à l'équilibre thermodynamique.

4.2.2. *Capacité*

Si la structure est polarisée par une tension V, supposée positive dans le sens direct c'est-à-dire quand le métal est polarisé positivement par rapport au semiconducteur, la barrière de potentiel devient $V_d - V$ et la largeur de la zone de charge d'espace devient

$$W(V) = \left(\frac{2\varepsilon_s}{eN_d} (V_d - V) \right)^{1/2} \tag{4-5}$$

Comme dans la jonction pn toute variation de V entraîne une modulation de $W(V)$ et par suite une modulation de la charge totale développée dans le semiconducteur. Il en résulte que la structure présente une capacité différentielle. La charge d'espace est donnée par

$$Q_{sc} = -Q_m = eN_d W = \left(2\varepsilon_s eN_d (V_d - V)\right)^{1/2} \tag{4-6}$$

La capacité différentielle est donnée par

$$C(V) = \left|\frac{dQ}{dV}\right| = \left(\frac{\varepsilon_s eN_d}{2}\right)^{1/2} (V_d - V)^{-1/2} = \frac{\varepsilon_s}{W} \tag{4-7}$$

Cette capacité est équivalente à celle d'un condensateur plan d'épaisseur W. On peut écrire l'expression de $C(V)$ sous la forme

$$C^{-2}(V) = \frac{2}{\varepsilon_s eN_d}(V_d - V) \tag{4-8}$$

La courbe représentant $C^{-2}(V)$ est une droite dont la pente permet de déterminer la densité de donneurs N_d , et dont l'abscisse à l'origine permet de déterminer la hauteur de barrière V_d (Fig. 4-14). Si le dopage du semiconducteur n'est pas homogène, la courbe représentant $C^{-2}(V)$ n'est plus une droite.
Nous avons supposé dans le calcul précédent qu'il n'existait pas d'états d'interface dans la structure. Si une certaine densité de tels états existe, une partie des charges fixes développées dans le semiconducteur se localise sur ces états, l'autre partie constitue la zone de déplétion. L'égalité des charges développées dans le métal et dans le semiconducteur s'écrit alors

$$Q_m = -Q_{sc} = -\left(Q_{dep} + Q_s\right) \tag{4-9}$$

où Q_{dep} représente la charge de déplétion et Q_s la charge de surface.

En raison de la présence de charges d'interface, la différence de potentiel entre le métal et le semiconducteur s'établit en partie dans la zone de déplétion du semiconducteur, en partie dans la zone d'interface où V_d représente comme précédemment la tension existant aux bornes de la zone de charge d'espace et ΔV_i la chute de tension dans l'interface.

$$V_{sc} - V_m = V_d + \Delta V_i$$

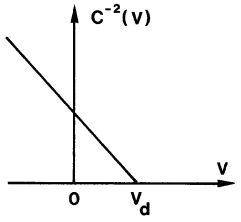


Figure 4.14 : Variation de la capacité avec la tension

Il en résulte que la barrière que doivent franchir les électrons pour passer du semiconducteur dans le métal est donnée par $e(V_d+\Delta V_i)$. En fait la largeur de la zone d'interface étant de quelques Angströms les électrons ne franchissent pas la barrière $e\Delta V_i$ mais passent à travers par effet tunnel. La hauteur de barrière effective est par conséquent égale à eV_d.

Le calcul développé précédemment reste correct dans la mesure où nous avons choisi comme origine des potentiels le potentiel à la surface du semiconducteur. En l'absence d'états d'interface ce potentiel est celui du métal (Fig 4-15-a), en présence d'états d'interface la chute de potentiel dans la zone d'interface amène le potentiel du métal à $-\Delta V_i$ (Fig. 4-15-b). Le point anguleux qui apparaît en x=0, correspond à la discontinuité du champ électrique résultant de la discontinuité de contante diélectrique entre la couche d'interface et le semiconducteur.

En ce qui concerne la capacité de la structure, il existe deux capacités en série, celle de la couche d'interface et celle de la zone de déplétion. La capacité de la couche d'interface est donnée par ε_i/δ si on appelle ε_i la constante diélectrique de la couche et δ son épaisseur. La capacité de la zone de déplétion en l'absence de polarisation, est donnée par la même expression que précédemment. En fait la couche d'interface étant très étroite, (δ~5Å), la capacité de la structure se réduit à celle de la zone de déplétion.

Lorsque la structure est polarisée, la tension de polarisation se répartit entre la couche d'interface et la zone de charge d'espace. Si V est la tension appliquée, $V_{dep}=kV$ et $\Delta V_i=(1-k)V$. L'expression de $C^{-2}(V)$ s'écrit alors

$$C^{-2}(V)=\frac{2}{\varepsilon_s e N_d}(V_d-kV)) \qquad (4\text{-}11)$$

4.3. Caractéristique courant-tension

Le courant de porteurs minoritaires étant négligeable, le courant dans la structure est essentiellement dû aux porteurs majoritaires. Ce courant est conditionné par des phénomènes physiques différents dans les différentes régions de la structure. A l'interface il est conditionné par l'émission thermoélectronique par-dessus la barrière de potentiel. Dans la zone de charge d'espace du semiconducteur, il est régi par les phénomènes de diffusion. Le courant de porteurs majoritaires étant le seul courant existant, il est nécessairement conservatif. On peut donc le calculer dans l'une ou l'autre des régions précédentes. Nous allons considérer successivement chacune de ces régions, dans une structure métal-semiconducteur de type n, avec $\phi_m>\phi_s$.

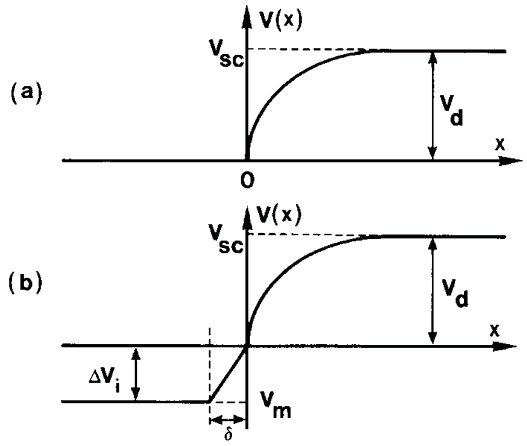


Figure 4-15 : Variation du potentiel en l'absence (a) et en présence (b) d'états d'interface

4.3.1. *Courant d'émission thermoélectronique*

Nous avons établi (Par. 1.6.4.b) l'expression du courant thermoélectronique existant au niveau d'une hétérostructure métal-semiconducteur, en fonction de la barrière de potentiel existant à l'interface. En l'absence de toute polarisation, le courant résultant est nul de sorte que chacun des courants thermoélectroniques s'écrit (1-247)

$$J_{m\to sc} = J_{sc\to m} = A^* T^2 e^{-E_b / kT} \tag{4-12}$$

où A^* est la constante de Richardson donnée par $A^* = 4\pi e m_e k^2 / h^3$ et E_b est la barrière de potentiel donnée par (Fig 4-16-a)

$$E_b = e\phi_m - e\chi = e\phi_b + e\phi_F \tag{4-13}$$

où $e\phi_F = E_c - E_F$. En explicitant E_b chacune des composantes du courant s'écrit

$$J = A^* T^2 e^{-e\phi_F / kT} e^{-e\phi_b / kT} \tag{4-14}$$

Mais la densité d'électrons dans le semiconducteur, qui est égale à N_d si on suppose tous les donneurs et accepteurs ionisés, s'écrit (1-132-a)

$$n = N_d = N_c e^{-e\phi_F / kT} \tag{4-15}$$

où N_c représente la densité équivalente d'états de la bande de conduction donnée par l'expression (1-136-a)

$$N_c = 2\left(2\pi m_e kT / h^2\right)^{3/2} \tag{4-16}$$

Compte tenu des relations (4-14, 15 et 16), l'expression du courant devient

$$J = A^* T^2 \frac{N_d}{N_c} e^{-e\phi_b / kT} \tag{4-17}$$

En explicitant A^* et N_c et en précisant le fait qu'en l'absence de polarisation $\phi_b = V_d$ où V_d représente la tension de diffusion de la structure, les composantes du courant thermoélectronique s'écrivent

$$J_{m\to sc} = J_{sc\to m} = eN_d (kT/2\pi m_e)^{1/2} e^{-eV_d / kT} \tag{4-18}$$

Polarisons la structure par une tension V_m-V_{sc}=V. V est donc compté positivement lorsqu'on abaisse le potentiel du semiconducteur par rapport à celui du métal. La figure (4-16-b) représente le diagramme énergétique pour V>0.

- La barrière métal-semiconducteur est inchangée, le courant thermoélectronique $J_{m\to sc}$ est donc inchangé

- La barrière semiconducteur-métal devient $e\phi'_b = e\phi_b - eV = e(V_d - V)$, le courant semiconducteur-métal devient

$$J_{sc\to m} = eN_d (kT/2\pi m_e)^{1/2} e^{-e(V_d - V)/kT} \tag{4-19}$$

Le courant résultant est donné par $J = J_{sc\to m} - J_{m\to sc}$

$$J = eN_d (kT/2\pi m_e)^{1/2} \left(e^{-e(V_d - V)/kT} - e^{-eV_d / kT}\right)$$

soit

$$J = J_{se}\left(e^{eV/kT} - 1\right) \tag{4-20-a}$$

$$J_{se} = e N_d (kT/2\pi m_e)^{1/2} e^{-eV_d/kT} \tag{4-20-b}$$

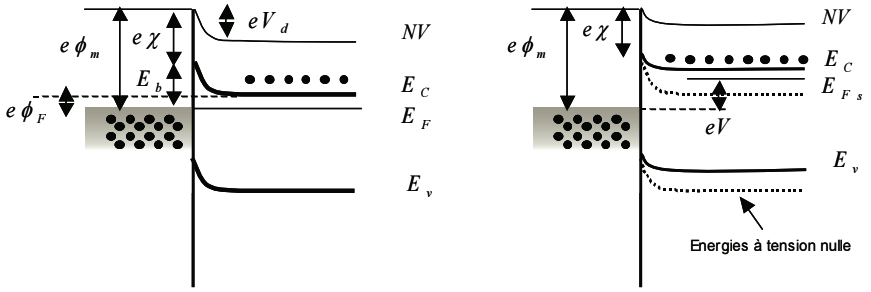


Figure 4-16 : Diode Schottky.

Pour $V>>kT/e$, le terme exponentiel devient très vite beaucoup plus grand que 1, le courant augmente exponentiellement avec la tension appliquée. Le sens passant correspond donc à une polarisation positive du métal par rapport au semiconducteur. Pour $V<0$ au contraire le terme exponentiel devient négligeable et $J=-J_{se}$, J_{se} est le courant de saturation (Fig. 4-17).

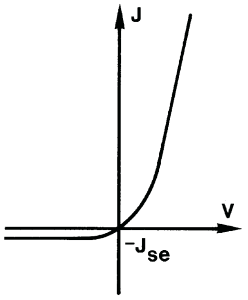


Figure 4-17 : Relation courant-tension.

Le cas d'une structure métal-semiconducteur de type p avec $e\phi_m<e\phi_s$ se calcule de la même manière. Le courant à travers la barrière est alors transporté par les trous et la polarisation directe correspond à une tension semiconducteur-métal positive. Rappelons que les structures M-SC(n) avec $e\phi_m<e\phi_s$ sont des contacts ohmiques

4.3.2. *Courant de diffusion*

Le calcul consiste à résoudre l'équation de diffusion des porteurs dans la zone de charge d'espace du semiconducteur, compte tenu de la présence dans cette zone, d'un gradient de concentration et d'un champ électrique. Le courant d'électrons s'écrit, en supposant constante la mobilité des porteurs,

$$j_n = e\left(n(x)\mu_n E(x) + D_n \frac{dn(x)}{dx} \right) \tag{4-21}$$

Compte tenu de la relation d'Einstein $D/\mu=kT/e$, et du fait que $E(x)=-dV(x)/dx$, le courant s'écrit

$$j_n = eD_n\left(-\frac{e}{kT}n(x)\frac{dV(x)}{dx}+\frac{dn(x)}{dx}\right) \tag{4-22}$$

On multiplie à gauche et à droite par $e^{-eV(x)/kT}$

$$j_n e^{-eV(x)/kT} = eD_n\left(-\frac{e}{kT}n(x)\frac{dV(x)}{dx}e^{-eV(x)/kT}+\frac{dn(x)}{dx}e^{-eV(x)/kT}\right)$$

soit

$$j_n e^{-eV(x)/kT}\,dx = eD_n\,d\left(n(x)e^{-eV(x)/kT}\right) \tag{4-23}$$

En régime stationnaire, et dans la mesure où l'on néglige le courant de porteurs minoritaires $j_n=J$ et par suite est indépendant de x. En intégrant l'expression (4-23) entre $x=0$ et $x=W$ on obtient

$$J\int_0^W e^{-eV(x)/kT}\,dx = eD_n\left[n(x)e^{-eV(x)/kT}\right]_0^W \tag{4-24}$$

Les conditions aux limites sont définies de la manière suivante

- En $x = 0$ $\quad n(x=0)=N_c e^{-(E_c(x=0)-E_F)/kT}=N_c e^{-E_b/kT}$
  $\quad V(x=0) = 0$

- En $x = W$ $\quad n(x=W) = N_d = N_c e^{-e\phi_F/kT}$
  $\quad V(x=W) = V_{sc} = V_d - V$

Comme dans le calcul du courant d'émission thermoélectronique, on choisit l'origine des potentiels à la surface du semiconducteur et on définit la polarisation par la tension $V_m - V_{sc} = V$. Ainsi $V>0$ entraîne un abaissement de la barrière semiconducteur-métal. En explicitant les conditions aux limites on obtient

$$J\int_0^W e^{-eV(x)/kT}\,dx = eD_n\left(N_d e^{-e(V_d-V)/kT}-N_c e^{-E_b/kT}\right) \tag{4-25}$$

or $N_d = n = N_c e^{-e\phi_F/kT}$ donc $N_c = N_d e^{e\phi_F/kT}$. En portant cette valeur de N_c dans l'expression (4-25) on obtient

$$J\int_0^W e^{-eV(x)/kT}\,dx = eD_n\left(N_d e^{-e(V_d-V)/kT}-N_d e^{-(E_b-e\phi_F)/kT}\right) \tag{4-26}$$

Mais $E_b - e\phi_F = eV_d$ ce qui donne pour J l'expression

$$J = \frac{eD_n N_d}{\int_0^W e^{-eV(x)/kT}\,dx}e^{-eV_d/kT}\left(e^{-eV/kT}-1\right) \tag{4-27}$$

Pour calculer l'intégrale du dénominateur il faut expliciter le potentiel $V(x)$ dans la zone de charge d'espace à partir de l'équation (4-3)

$$V(x)=-\frac{eN_d}{\varepsilon_s}\left(\frac{x^2}{2}-Wx\right)=\frac{eN_d}{2\varepsilon_s}W^2-\frac{eN_d}{2\varepsilon_s}(x-W)^2$$

or $W^2=2\varepsilon_s/eN_d\cdot(V_d-V)$ ce qui donne

$$V(x)=V_d-V-\frac{eN_d}{2\varepsilon_s}(x-W)^2 \qquad (4\text{-}28)$$

L'intégrale s'écrit alors

$$I=\int_0^W e^{-eV(x)/kT}dx=e^{-e(V_d-V)/kT}\int_0^W e^{\frac{e^2N_d}{2\varepsilon_s kT}(x-W)^2}dx$$

on pose

$$u^2=\frac{e^2N_d}{2\varepsilon_s kT}(x-W)^2,\ 2udu=\frac{e^2N_d}{2\varepsilon_s kT}2(x-W)dx\ ,\ \text{soit}\quad dx=\left(\frac{2\varepsilon_s kT}{e^2N_d}\right)^{1/2}du$$

En ce qui concerne les bornes d'intégration

- Pour $x=0$, $\quad u^2=\frac{e^2N_d}{2\varepsilon_s kT}W^2=\frac{e}{kT}(V_d-V)$
- Pour $x=W$, $\quad u^2=0$

soit en posant $z=\left(e(V_d-V)/kT\right)^{1/2}$

$$I=\left(\frac{2\varepsilon_s kT}{e^2N_d}\right)^{1/2}e^{-z^2}\int_0^z e^{u^2}du$$

La valeur asymptotique de $e^{-z^2}\int_0^z e^{u^2}du$ pour $z>>1$ est $1/2z$. Or $z>>1$ correspond à $|V_d-V|>>kT/e$, ce qui est largement vérifié (sauf au voisinage de $V=V_d$) compte tenu de la faible valeur de kT/e=26 mV à la température ambiante. Ainsi, en prenant la valeur asymptotique de l'intégrale et en explicitant z , l'expression de I s'écrit

$$I=\left(\frac{2\varepsilon_s kT}{e^2N_d}\right)^{1/2}\frac{1}{2}\left(\frac{kT}{e(V_d-V)}\right)^{1/2}$$

ou en faisant apparaître W

$$I=\left(\frac{2\varepsilon_s kT}{e^2 N_d}\right)^{1/2}\frac{1}{2}\left(\frac{kT}{e(V_d-V)}\right)^{1/2}\frac{1}{W}\left(\frac{2\varepsilon_s(V_d-V)}{eN_d}\right)^{1/2}$$

soit

$$I=\frac{\varepsilon_s kT}{e^2 N_d W} \tag{4-29}$$

En explicitant I dans l'expression du courant on obtient

$$J=J_{sd}(e^{eV/kT}-1) \tag{4-30}$$

avec

$$J_{sd}=\frac{e^3 N_d^2 D_n}{\varepsilon_s kT} W e^{-eV_d/kT} \tag{4-31}$$

L'expression (4-31) peut s'écrire sous la forme

$$J_{sd}=eN_d\frac{eD_n}{kT}\frac{eN_d W}{\varepsilon_s}e^{-eV_d/kT} \tag{4-32}$$

Mais $eD_n/kT=\mu_n$ la mobilité des porteurs, et $eN_dW/\varepsilon_s=E_s$ le champ électrique à l'interface. Ainsi en appelant $v_d=\mu_n E_s$ la vitesse des porteurs à l'interface, le courant de saturation s'écrit

$$J_{sd}=eN_d v_d e^{-eV_d/kT} \tag{4-33}$$

On obtient la variation du courant de saturation avec la tension appliquée en explicitant W dans l'expression (4-31)

$$J_{sd}=\frac{e^3 N_d^2 D_n}{\varepsilon_s kT}\left(\frac{2\varepsilon_s}{eN_d}\right)^{1/2}(V_d-V)^{1/2}e^{-eV_d/kT}$$

$$=eN_d\frac{eD_n}{kT}\frac{eN_d}{\varepsilon_s}\left(\frac{2\varepsilon_s}{eN_d}\right)^{1/2}(V_d-V)^{1/2}e^{-eV_d/kT}$$

soit

$$J_{sd}=eN_d\mu_n\left(\frac{2eN_d}{\varepsilon_s}\right)^{1/2}(V_d-V)^{1/2}e^{-eV_d/kT} \tag{4-34}$$

4.3.3. *Combinaison des deux courants*

On obtient des expressions de J très voisines d'un calcul à l'autre : le courant varie comme $(e^{eV/kT}-1)$ en fonction de la tension appliquée, le courant de saturation varie exponentiellement avec la température et la différence des travaux de sortie du métal et du semiconducteur, $J_s \sim e^{-eV_d/kT}$ avec $V_d = \phi_m - \phi_s$. Mais dans la mesure où le courant est conservatif, ces expressions devraient être identiques

Comparons les expressions du courant de saturation obtenues dans chacun des modèles

$$J_{se} = eN_d \left(\frac{kT}{2\pi m_e} \right)^{1/2} e^{-eV_d/kT} \qquad \text{(4-35-a)}$$

$$J_{sd} = eN_d v_d e^{-eV_d/kT} \qquad \text{(4-35-b)}$$

Ces expressions sont de la forme générale $J = nev$, où $n = N_d\, e^{-eV_d/kT}$ représente le nombre de porteurs à l'interface en l'absence de polarisation, et v leur vitesse. Dans le modèle de la diffusion la vitesse à l'interface est $v_d = \mu_n E_s$. Dans le modèle de l'émission thermoélectronique cette vitesse est donnée par

$$v_e = (kT/2\pi m_e)^{1/2} \qquad \text{(4-36)}$$

Cette vitesse décrit l'action de collection des électrons par le métal, elle joue le même rôle qu'une *vitesse de recombinaison* à l'interface. Or la vitesse thermique des porteurs est donnée par $v_{th} = (3kT/m_e)^{1/2}$, ainsi $v_e = v_{th}/\sqrt{6\pi} \approx v_{th}/4$

On peut donc écrire le courant de saturation dans chacun des modèles sous une même forme

$$J_{se} = eN_d v_e e^{-eV_d/kT} \quad J_{sd} = eN_d v_d e^{-eV_d/kT}$$

La vitesse v_d varie proportionnellement au champ en surface alors que la vitesse v_e varie avec la température comme la vitesse thermique, en $T^{1/2}$.

On combine les deux modèles en écrivant que les courants de diffusion et d'émission sont en série, le courant étant d'émission à l'interface et de diffusion dans la zone de charge d'espace du semiconducteur. La condition de raccordement consiste à écrire que ces courants sont égaux. Mais dans le calcul de chacun d'eux nous avons supposé que la densité de porteurs à l'interface était donnée par le régime de pseudo-équilibre $n(x=0) = N_d\, e^{-E_b/kT}$, c'est-à-dire en d'autres termes que cette densité de porteurs était indépendante du flux de porteurs résultant du passage du courant. La condition de raccordement consiste à écrire que la densité de porteurs dans le semiconducteur à l'interface est, dans le modèle de la diffusion,

conditionnée par la présence du courant d'émission thermoélectronique, et vice versa. Dans le modèle de la diffusion cette condition à l'interface est équivalente à l'introduction d'une *vitesse de recombinaison à l'interface* $s=v_e$ analogue à une *vitesse de recombinaison en surface.*

Ainsi, sous une polarisation V nous écrirons que $n(x=0)$ n'est plus une constante mais une fonction de V, soit $n(x=0) = n_s(V)$.

Le courant de diffusion est toujours donné par l'équation (4-24), mais la modification d'une condition aux limites modifie l'équation (4-25) qui devient

$$J_d \int_0^W e^{-eV(x)/kT} dx = eD_n \left(N_d e^{-e(V_d - V)/kT} - n_s(V) \right) \tag{4-37}$$

L'intégrale du premier membre n'est pas modifiée de sorte que le courant s'écrit

$$J_d = eN_d v_d e^{-e(V_d - V)/kT} - ev_d n_s(V) \tag{4-38}$$

Le courant d'émission thermoélectronique s'écrivait dans le régime de pseudo-équilibre, c'est-à-dire en supposant $n(x=0)$ indépendant du courant

$$J_e = eN_d v_e e^{-eV_d/kT} \left(e^{eV/kT} - 1 \right) \tag{4-39}$$

c'est-à-dire pour une polarisation directe $V >> kT/e$

$$J_e = ev_e N_d e^{-e(V_d - V)/kT} \tag{4-40}$$

Or $N_d e^{-e(V_d - V)/kT} = n(x=0)$ représente le nombre de porteurs dans le semiconducteur à l'interface, en régime de speudo-équilibre. Il suffit donc de remplacer ce terme par le nombre de porteurs en régime dynamique $n_s(V)$. Ainsi le courant d'émission thermoélectronique s'écrit

$$J_e = e\, n_s(V)\, v_e \tag{4-41}$$

Il suffit donc d'écrire $J = J_d = J_e$ pour obtenir $n_s(V)$. En portant le résultat obtenu dans l'expression de J_e on obtient l'expression du courant résultant des effets combinés de la diffusion et de l'émission thermoélectronique. Soit à partir des expressions (4-38 et 41)

$$ev_d N_d e^{-e(V_d - V)/kT} - ev_d n_s(V) = ev_e n_s(V)$$

$$n_s(V) = N_d \frac{v_d}{v_d + v_e} e^{-e(V_d - V)/kT} \tag{4-42}$$

En portant cette expression de $n_s(V)$ dans l'expression (4-41) on obtient

$$J = e N_d \frac{v_d v_e}{v_d + v_e} e^{-e(V_d - V)/kT} \tag{4-43}$$

ou

$$J = e N_d v^* e^{-eV_d/kT} e^{eV/kT} \tag{4-44}$$

avec

$$\frac{1}{v^*} = \frac{1}{v_d} + \frac{1}{v_e} \tag{4-45}$$

Pour $v_d << v_e$, $v^* = v_d$ le courant dans la structure est conditionné par la diffusion des porteurs dans la zone de charge d'espace du semiconducteur. Par contre pour $v_d >> v_e$, $v^* = v_e$, le courant est conditionné par l'émission thermoélectronique à l'interface.

Les vitesses v_d et v_e sont fonction de paramètres différents, $v_d = \mu_n E_s$ et $v_e \approx v_{th}/4$ ainsi la condition $v_d >> v_e$ s'écrit $\mu_n E_s >> v_{th}/4$ ou $\mu_n v_{th} E_s >> v_{th}^2/4$. En explicitant $\mu_n = e\tau/m_e$, où τ représente le temps de relaxation des porteurs et $v_{th}^2 = 3kT/m_e$, la condition s'écrit

$$e v_{th} \tau E_s >> 3kT/4$$

En posant $\ell = v_{th}\tau$ où ℓ représente le libre parcours moyen des porteurs, on obtient

$$E_s >> 3kT/4e\ell$$

En explicitant E_s en fonction de la polarisation à partir des équations (4-2 et 5), la condition s'écrit

$$\left(\frac{2eN_d}{\varepsilon_s}\right)^{1/2} (V_d - V)^{1/2} >> 3kT/4e\ell$$

soit

$$V_d - V >> \frac{\varepsilon_s}{4eN_d \ell^2} \left(\frac{kT}{e}\right)^2 \tag{4-46}$$

Considérons du silicium de type n dopé avec $N_d = 10^{17}$ cm^{-3}. A la température ambiante $kT/e = 26$ mV, $v_{th} = 10^7$ cms^{-1}, $\tau = 10^{-12}$ s ce qui donne $\ell = 1000$ Å. D'autre part $\varepsilon_s = 12\varepsilon_o$. La condition s'écrit

$$V_d - V >> 1mV \qquad \text{c'est-à-dire} \qquad V < V_d$$

Ainsi, tant que la tension de polarisation est inférieure à la tension de diffusion, le courant est conditionné par l'émission thermoélectronique. Au-delà la structure est ohmique.

Nous n'avons pas tenu compte dans les calculs précédents de l'effet Schottky. La combinaison du potentiel image et du champ électrique dans la zone de charge d'espace du semiconducteur, entraîne un abaissement de la barrière de potentiel et un déplacement de son maximum vers l'intérieur du semiconducteur. Il en résulte une modification de la tension de diffusion V_d et de la vitesse de diffusion v_d dans la zone de charge d'espace. Dans la mesure où le courant dans la structure est essentiellement conditionné par l'émission thermoélectronique, nous pouvons avec une bonne approximation négliger la variation de v_d. En ce qui concerne la hauteur de barrière, elle devient (chapitre 1-6-5)

$$E_{bm}=E_b-\left(\frac{e^3 E}{4\pi\varepsilon_s}\right)^{1/2} \tag{4-47}$$

de sorte que la tension de diffusion devient

$$V_d^{'}=V_d-\left(\frac{eE}{4\pi\varepsilon_s}\right)^{1/2} \tag{4-48}$$

L'expression du courant s'écrit alors, en polarisation directe

$$J=eN_d\frac{v_d v_e}{v_d+v_e}e^{-eV_d^{'}/kT}e^{eV/kT} \tag{4-49}$$

Enfin un autre phénomène modifie la caractéristique $I(V)$ de la structure, aussi bien dans le sens direct que dans le sens inverse, c'est l'effet tunnel à travers la barrière de potentiel. Cet effet ne joue un rôle appréciable que pour des barrières très étroites, c'est-à-dire dans le cas limite de semiconducteurs dégénérés. Dans l'autre cas limite de semiconducteurs peu dopés, l'injection des porteurs minoritaires peut aussi jouer un rôle mais toujours négligeable.

EXERCICES DU CHAPITRE 4

- **Exercice 1 : Hauteur de barrière**

Calculez la hauteur de la barrière d'un contact métal-semiconducteur n avec dopage $10^{16}/cm^3$, affinité 4,01 eV et travail de sortie du métal 4,1 eV.

- **Exercice 2 : Zone de charge d'espace**

La structure précédente est polarisée en inverse à 1V. Calculez la largeur de la zone de charge d'espace.

- **Exercice 3 : Niveau du vide**

Expliquez pourquoi le niveau du vide n'est pas constant quand on met en contact des matériaux différents.

- **Exercice 4 : Structure de bandes**

Un semiconducteur de type n dopé à $10^{16}/cm^3$ est en contact avec de l'aluminium, du cuivre et de l'or. Tracez la structure de bandes dans chaque cas. Reprendre le problème avec du silicium p.

- **Exercice 5 : Calcul du courant**

Expliquez pourquoi le calcul du courant dans une jonction métal-semiconducteur effectué en appliquant la formule de l'émission thermoélectronique et le calcul effectué en sommant courant de diffusion et courant de dérive ne donnent pas le même résultat.

- **Exercice 7 : Courant Schottky**

En reprenant les conditions du problème 1, calculez la densité de courant à température ambiante pour une tension de 0,5 V en direct et de 5 V en inverse.

- **Exercice 8 : Zone de charge d'espace**

Dans un contact métal-semiconducteur expliquez pourquoi la zone de charge d'espace peut s'étendre sur des dimensions très différentes.

- **Exercice 9 : Contact ohmique**

Expliquez pouquoi un contact métal sur une zone n+ d'un semiconducteur n est en général ohmique.

- **Exercice 10 : Passage des électrons**

Quel est le critère qui permet de dire que dans un contact métal-semiconducteur, les électrons passent du semiconducteur au métal ?
Que dire pour les trous ?

5. Structure métal-isolant semiconducteur

Cette structure est fondamentale car elle est la base du composant le plus utilisé en électronique : le transistor à effet de champ ou MOSFET. De plus, la surface des semiconducteurs est en général recouverte d'un oxyde après fabrication aussi mince soit-il. Les interfaces métalliques sont donc en réalité des structures métal-oxyde-semiconducteur. Les notions fondamentales de potentiel de contact, de régime d'inversion ou d'accumulation seront également détaillées. La compréhension de la structure de bandes du système métal-isolant-semiconducteur est indispensable pour la suite de l'ouvrage. Le chapitre est organisé de la manière suivante :

5.1. Diagramme énergétique
5.2. Les potentiels de contact
5.3. Le modèle électrique de base
5.4. Le régime de forte inversion
5.5. Le régime de faible inversion

5.1. Diagramme énergétique

5.1.1. *Structure métal-vide-semiconducteur*

Considérons un métal et un semiconducteur dans le vide. Le métal est caractérisé par son travail de sortie $e\phi_m$. Le semiconducteur est caractérisé par son travail de sortie $e\phi_s$ et son affinité électronique $e\chi_s$. Ces deux systèmes sont indépendants, les niveaux de Fermi dans chacun d'eux sont respectivement à la distance $e\phi_m$ et $e\phi_s$ du niveau du vide *NV*. Le diagramme énergétique est représenté sur la figure (5-1-a).

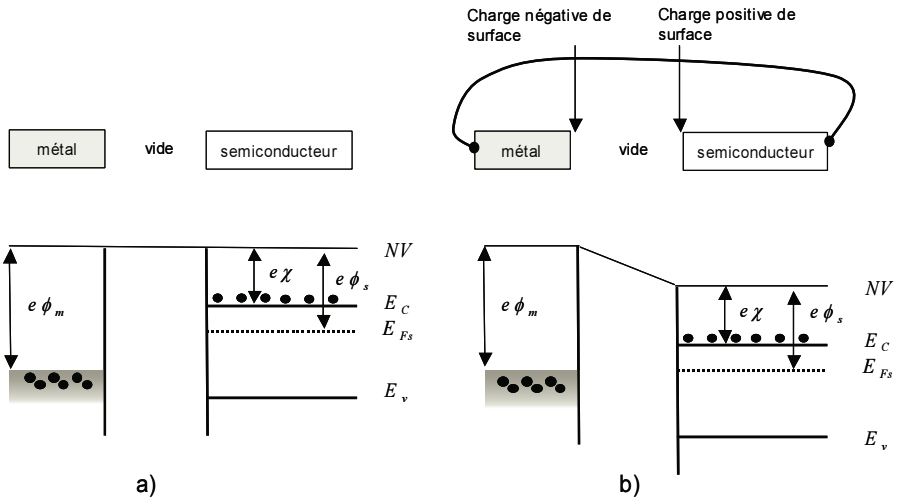


Figure 5.1 : Structure métal-vide-semiconducteur

Relions ces deux systèmes par un fil conducteur, ils échangent de l'énergie et constituent un seul système thermodynamique, les niveaux de Fermi s'alignent. Il en résulte une différence de potentiel de contact, analogue à la tension de diffusion de la jonction pn ou du contact métal-semiconducteur, donnée par

$$V_d = \phi_m - \phi_s \qquad \text{ou} \qquad \phi_{ms} \tag{5-1}$$

A cette différence de potentiel sont associés un champ électrique et une charge d'espace par les relations

$$E = -\frac{dV}{dx} \qquad \frac{d^2V}{dx^2} = -\frac{\rho(x)}{\varepsilon}$$

L'amplitude du champ électrique et la densité de la charge d'espace sont d'autant plus importantes que le gradient du potentiel est grand, c'est-à-dire puisque ΔV est fixé par la différence des travaux de sortie, que la distance entre le métal et le semiconducteur est faible. Tant que cette distance est importante le champ et la charge d'espace sont négligeables, le diagramme énergétique est représenté sur la figure (5-1-b).

Rapprochons le métal du semiconducteur, des charges apparaissent au voisinage de la surface du métal et au voisinage de la surface du semiconducteur. Le système est analogue à un condensateur plan dont la tension entre les armatures est $\Delta V=V_d$ $Q=C\Delta V$. ΔV étant constant, lorsque la distance entre le métal et le semiconducteur diminue, la capacité et par suite la charge, augmentent. La charge totale développée dans le métal est égale et opposée à la charge totale développée dans le semiconducteur. La densité d'états étant beaucoup plus grande dans le métal que dans le semiconducteur, ces charges sont superficielles dans le métal et s'étendent davantage dans le semiconducteur. La charge d'espace dans le métal résulte de la variation de la densité d'électrons au voisinage de la surface. La charge d'espace dans le semiconducteur résulte de la variation de la densité de porteurs libres, électrons et trous, au voisinage de la surface. La variation de la densité de porteurs libres est associée à la variation de la distance bande permise-niveau de Fermi. Dans la mesure où le niveau de Fermi est fixé par l'équilibre thermodynamique, il en résulte une courbure des bandes de valence et de conduction vers le bas ou le haut, suivant que la densité d'électrons augmente ou diminue. La nature de la charge d'espace et la courbure des bandes sont fonction d'une part du type du semiconducteur et d'autre part de la différence des travaux de sortie $e\phi_m$-$e\phi_s$.

Considérons un semiconducteur de type n et différentes valeurs relatives des travaux dortie du métal et du semiconducteur.

- Premier cas : $\phi_m<\phi_s$

Dans ce cas, V_s-V_m est négatif, des charges négatives se développent dans le semiconducteur et des charges positives se développent dans le métal. Les charges positives dans le métal résultent d'un départ d'électrons de la surface. Les charges négatives dans le semiconducteur résultent d'une accumulation d'électrons vers la surface, la bande de valence et la bande de conduction se courbent vers le bas . Le semiconducteur est dit en *régime d'accumulation*, le diagramme énergétique est représenté sur la figure (5-2-a).

- Deuxième cas : $\phi_m=\phi_s$

Dans ce cas, V_s-V_m=0, la tension de diffusion est nulle, aucune charge n'apparaît, les bandes restent horizontales, le semiconducteur est dit en *régime de bandes plates.* Le diagramme énergétique est représenté sur la figure (5-2-b).

- Troisième cas : $\phi_m>\phi_s$

Dans ce cas, V_s-V_m est positif de sorte que des charges positives se développent dans le semiconducteur et des charges négatives se développent dans le métal. Les charges négatives dans le métal résultent d'une accumulation d'électrons à la surface. Les charges positives dans le semiconducteur résultent du départ d'électrons et proviennent d'une part de la présence d'ions donneurs non compensés par la charge électronique, et d'autre part de l'augmentation correspondante du nombre de trous résultant de la condition *np=Cte*. A la diminution de la densité électronique est associée une courbure des bandes vers le haut.

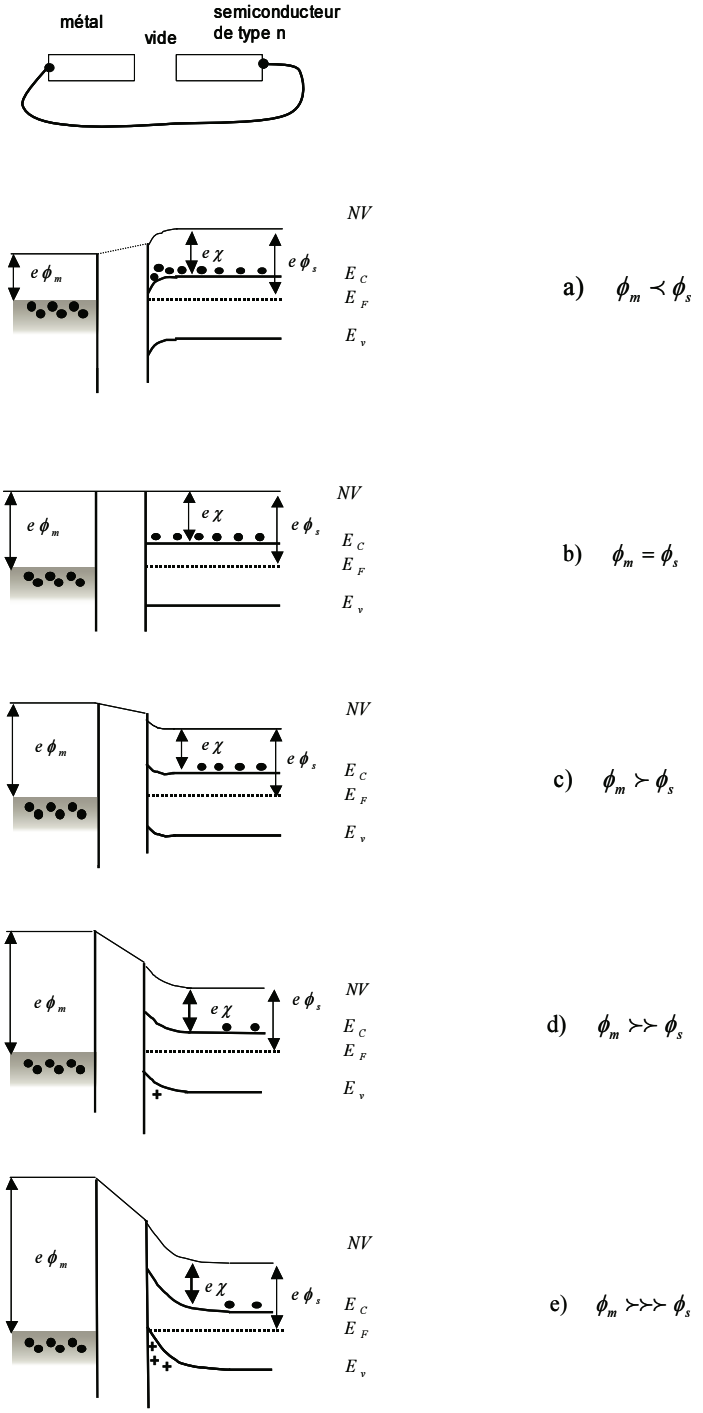


Figure 5-2 : Métal-semiconducteur de type n

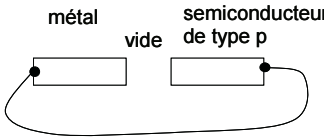
Si le gradient de potentiel est relativement faible, la densité de trous reste inférieure à n_i ($p<n_i<$n) , les ions donneurs constituent alors l'essentiel de la charge d'espace, le semiconducteur est dit en *régime de déplétion*. Le diagramme énergétique est représenté sur la figure (5-2-c).

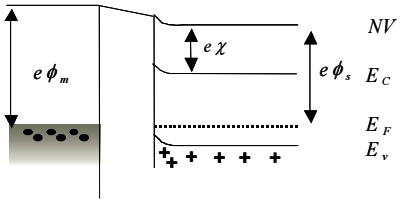
Si le gradient de potentiel est plus important, le déficit d'électrons, et par suite la densité de trous, augmentent. Lorsque la densité de trous devient supérieure à n_i ($n< n_i<p$) le semiconducteur devient de type p au voisinage de la surface, on dit que le semiconducteur est en *régime d'inversion*. Mais si la densité de trous reste inférieure à la densité de donneurs ($n<n_i<p<N_d$) l'essentiel de la charge d'espace reste due aux ions donneurs, le semiconducteur est dit en *régime de faible inversion*. Le diagramme énergétique est représenté sur la figure (5-2-d).

Enfin, si le gradient de potentiel est tel que $n<n_i<N_d<p$ la charge d'espace devient conditionnée par la présence des trous libres au voisinage de la surface, le semiconducteur est dit en *régime de forte inversion.* Le diagramme énergétique est représenté sur la figure (5-2-e).

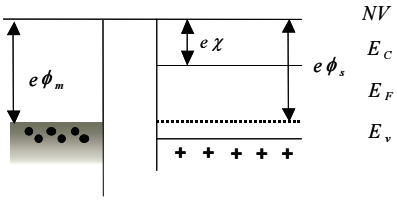
Ces différents régimes existent pour un semiconducteur de type p, les diagrammes énergétiques correspondants sont représentés sur la figure (5-3).

Rappelons qu'en raison des ordres de grandeur très différents de la densité d'états électronique de la bande de conduction du métal d'une part (~10^{22}cm^{-3}), des densités équivalentes d'états des bandes de conduction et de valence du semiconducteur d'autre part (~10^{19}cm^{-3} à la température ambiante) et des densités de donneurs ou d'accepteurs (~10^{16} - 10^{18}cm^{-3}), ***l'extension spatiale de la charge d'espace est très différente suivant sa nature. Dans le métal la charge reste localisée sur une fraction de couche atomique, c'est-à-dire rigoureusement à la surface. Dans le semiconducteur les charges d'inversion ou d'accumulation s'étendent sur quelques nanomètres, les charges de déplétion s'étendent sur quelques centaines de nanomètres suivant l'importance du dopage.***

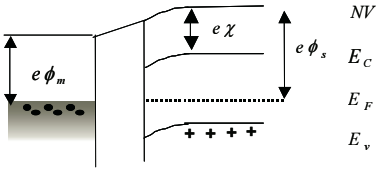




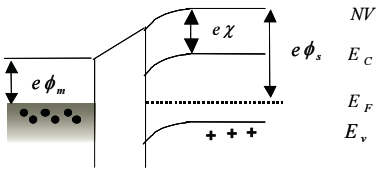
a) $\phi_m \succ \phi_s$



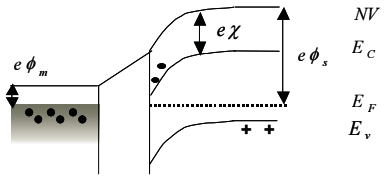
b) $\phi_m = \phi_s$



c) $\phi_m \prec \phi_s$



d) $\phi_m \prec\prec \phi_s$



e) $\phi_m \prec\prec\prec \phi_s$

Figure 5-3 : Métal-semiconducteur de type p

5.1.2. *Structure Métal-Isolant-Semiconducteur – MIS*

Dans la structure MIS (Métal Isolant Semiconducteur) l'intervalle entre le métal et le semiconducteur est rempli par un isolant. En technologie silicium cet isolant est l'oxyde de silicium SiO_2 d'où le nom plus communément utilisé de structure MOS (Métal Oxyde Semiconducteur).

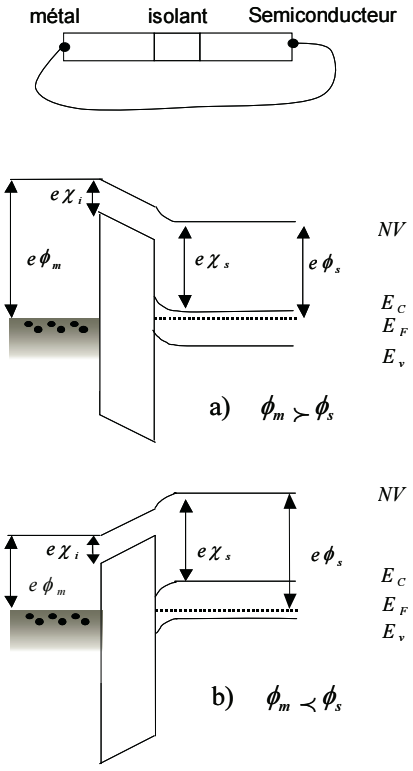


Figure 5-4 : Structure MIS à l'équilibre thermodynamique.
a) Semiconducteur(n) et $\phi_S < \phi_M$. b) Semiconducteur(p) et $\phi_S > \phi_M$

Dans la mesure où l'isolant peut être supposé parfait, cette structure présente les mêmes diagrammes énergétiques que ceux représentés sur les figures (5-2 et 3). L'isolant est caractérisé par son gap E_{gi} et par son affinité électronique $e\chi_i$. Les diagrammes énergétiques en régime de déplétion sont représentés sur la figure 5-4.

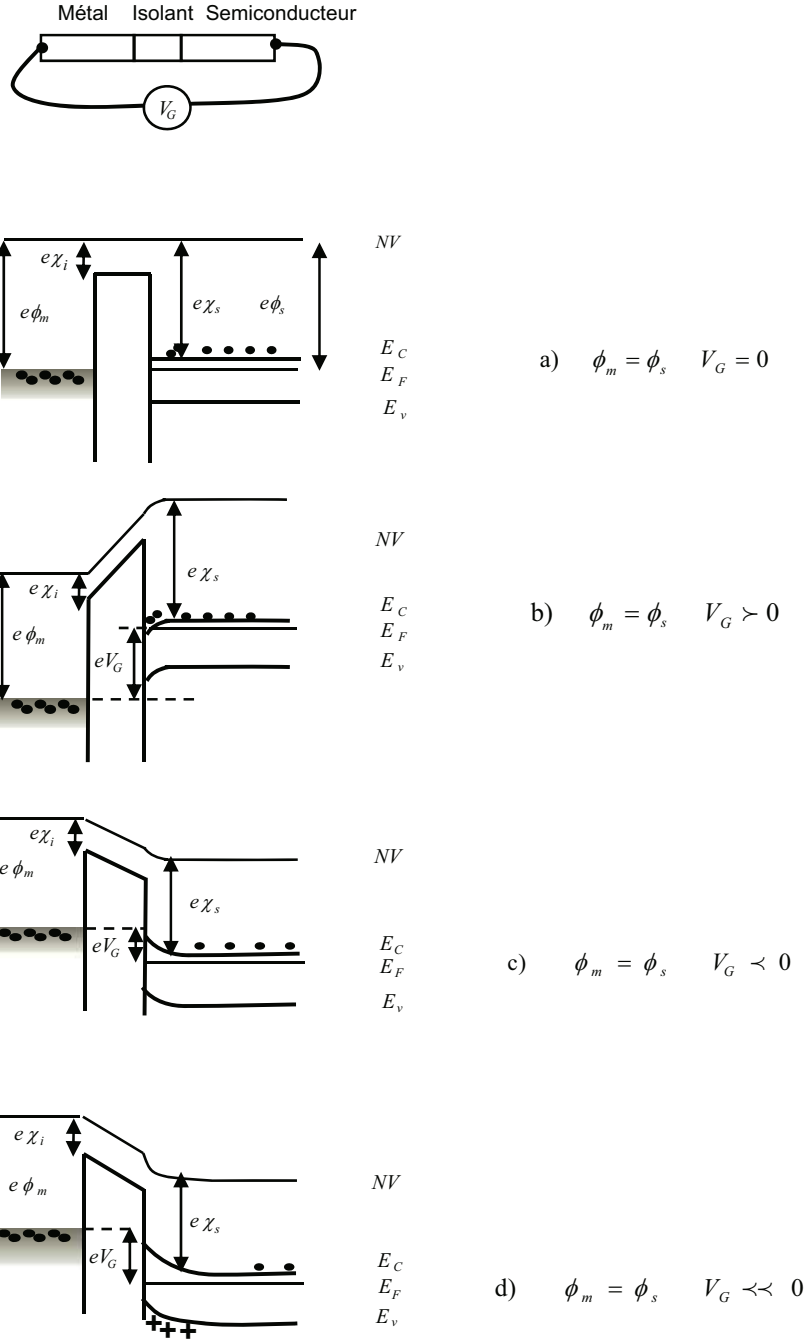


Figure 5-5 : Structure MIS(n) polarisée dans le cas particulier $\phi_m = \phi_s$

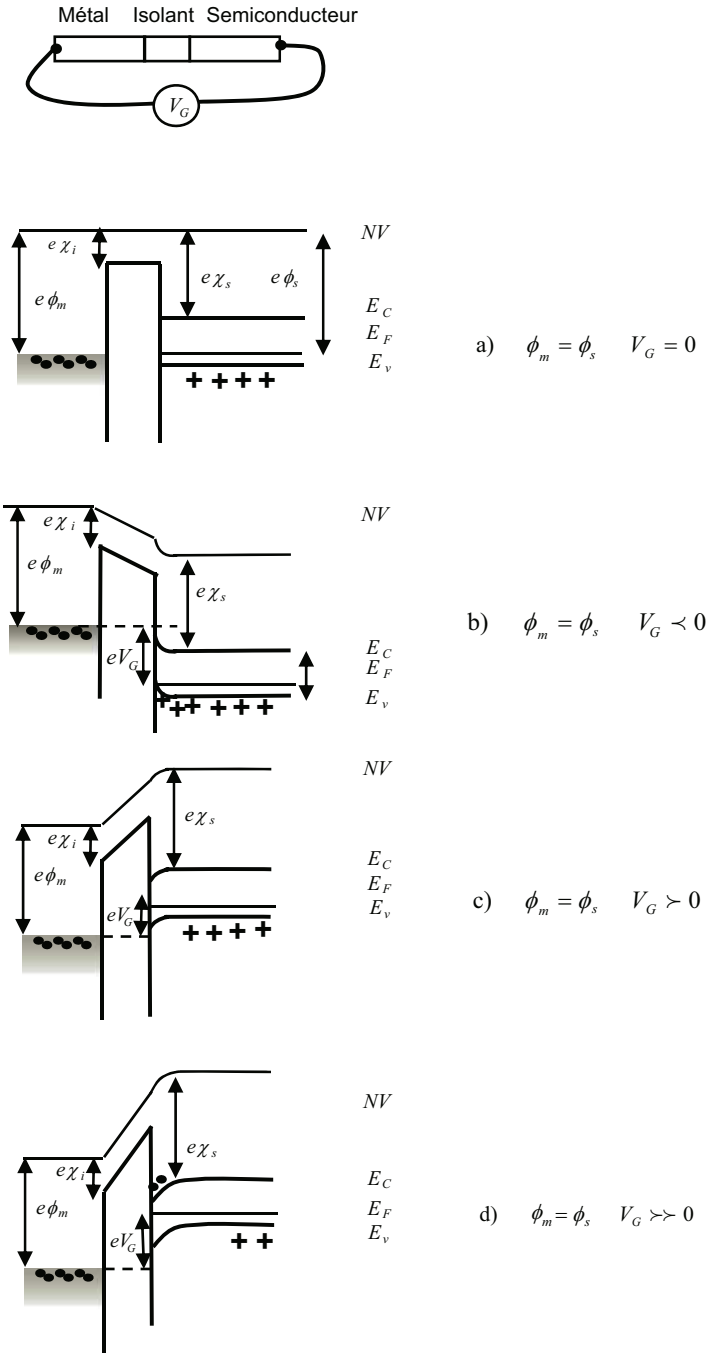


Figure 5-6 : Structure MIS(p) polarisée dans le cas particulier $\phi_m = \phi_s$

La hauteur de barrière entre le métal et le semiconducteur est toujours donnée par la différence des travaux de sortie du métal et du semiconducteur $V_d=\phi_m-\phi_s$. Cette barrière s'étend en partie dans le semiconducteur, en partie dans l'isolant. La forme de la barrière dans le semiconducteur est fonction de la distribution de la charge d'espace. Dans l'isolant supposé parfait, l'absence de charge entraîne $d^2V/dx^2=0$ et par suite une variation linéaire de V. Les bandes de conduction et de valence de l'isolant varient donc linéairement et celles du semiconducteur présentent une variation fonction de la distribution de la charge d'espace.

Les différents régimes de fonctionnement représentés sur les diagrammes des figures (5-2 et 3) se retrouvent dans le cas de la structure MIS. Ces diagrammes sont fonction du gradient de potentiel qui existe entre le métal et le semiconducteur. Ce gradient de potentiel a deux origines qui sont, d'une part la différence des travaux de sortie entre les deux matériaux, et d'autre part la différence de potentiel résultant de la polarisation éventuelle d'un matériau par rapport à l'autre. Les effets sont additifs, les figures (5-2 et 3) représentent l'effet de la différence des travaux de sortie à polarisation nulle, les figures (5-5 et 6) représentent l'effet de la polarisation en supposant la structure en régime de bandes plates ($\phi_m=\phi_s$) à polarisation nulle. Sous l'action de la polarisation la structure évolue d'un régime d'accumulation à un régime d'inversion en passant par les régimes de bandes plates et de déplétion. Dans le cas général ($\phi_m\neq\phi_s$) le régime de bandes plates n'est pas obtenu pour $V_G = 0$ mais pour $V_G = V_{FB} = \phi_m-\phi_s$.

Outre la différence des travaux de sortie et la polarisation extérieure, un autre phénomène modifie la barrière de potentiel et par suite les différents régimes de fonctionnement, la présence de charges localisées à l'interface isolant-semiconducteur.

Nous avons vu (Par.1.6.3) que les états de surface et d'interface pouvaient jouer un rôle important sur la courbure des bandes du semiconducteur et entraîner, indépendamment des phénomènes précédents, la présence d'une couche d'inversion à l'interface. Nous avons discuté le rôle effectif joué par ces états d'interface dans l'étude du contact métal-semiconducteur (Par. 4.1.3).

Dans le cas de la structure Si-SiO_2 l'expérience montre que la densité d'états localisés à l'interface isolant-semiconducteur est de l'ordre de quelques $10^{10}cm^{-2}$ quels que soient le type de silicium et la technique d'oxydation, et que ces charges d'interface sont toujours positives. Ces charges d'interface par unité de surface Q'_{0_s} induisent dans le semiconducteur une charge équivalente et de signe opposé $Q'_c=-Q'_0$. Il existe donc entre le métal et le semiconducteur une différence de potentiel additionnelle résultant de la présence de ces charges, donnée par

$$\Delta V=V_m-V_s=Q'_m/C'_{OX}=-Q'_c/C'_{OX}=Q'_0/C'_{OX} \qquad (5\text{-}2)$$

où Q'_0 représente la densité de charges d'interface et C_{OX} la capacité de l'isolant par unité de surface. $C_{OX} = \varepsilon_i / d$ où ε_i est la constante diélectrique de l'isolant et d son épaisseur.

Ainsi, en prenant en considération d'une part la différence des travaux de sortie et d'autre part la présence des charges d'interface, la tension de polarisation nécessaire à l'établissement du régime de bandes plates s'écrit

$$V_{FB} = (\phi_m - \phi_s) - Q'_0 / C'_{OX} \tag{5-3}$$

V_{FB} est appelée *tension de bandes plates* (Flat Band). Dans la mesure où ϕ_m est inférieur à ϕ_s et Q'_0 toujours positif, la tension de bandes plates est négative. Considérons par exemple une structure de type A1-SiO_2-Si:

Le travail de sortie de l'aluminium est $e\phi_m$=4,3 eV. Avec un dopage de 10^{15}cm^{-3}, le travail de sortie du silicium est $e\phi_{sn}$=4,6 eV pour du type n et $e\phi_{sp}$=5,2 eV pour du type p. Avec une épaisseur d'oxyde d=1000 Å , la capacité de l'oxyde est $C'_{OX} \approx 35.10^{-9}$ Fcm^{-2}. Enfin, si on suppose une densité de pièges d'interface Q'_0/e de l'ordre de 5.10^{10} cm^{-2} , on obtient V_{FB} = - 0,5 V pour du type n et V_{FB} = - 1,1 V pour du type p.

5.2. Potentiels de contact

Le problème des contacts entre matériaux différents est d'une grande importance dans les dispositifs micro-électroniques. Il a été introduit dans les paragraphes précédents en étudiant la structure de bandes. Il est aussi possible d'en donner une représentation simple en définissant le potentiel de contact. Cette approche est équivalente à la représentation des bandes d'énergie mais elle est plus facile à utiliser. Si on considère deux matériaux différents 1 et 2 mis en contact, isolants ou semi-conducteurs, il y a diffusion des porteurs des régions de forte concentration vers les régions de faible concentration tout comme dans la jonction pn. Il se forme alors de la même manière une zone de charge d'espace et une différence de potentiel apparaît naturellement sans tension extérieure appliquée. On notera cette tension V_{12} égale à V_1-V_2 .

Quand le matériau 2 est du silicium intrinsèque, le potentiel V_2 est noté V_I. La différence de potentiel entre le matériau 1 et le silicium intrinsèque est notée ϕ . Quand le matériau 1 est un semi-conducteur elle est appelée potentiel de Fermi et est notée $-\phi_F$. Le potentiel de Fermi est défini comme l'inverse du potentiel de contact. Cette définition n'a pas d'autre justification que la tradition des notations de la micro-électronique.

$$\phi_F = -(V_1 - V_I)$$

L'étude de la jonction pn a montré que si le matériau 1 est un semiconducteur de type p dopé avec une concentration N_A d'atomes accepteurs alors,

$$V_1 - V_I = -\phi_t \ln \frac{N_A}{n_i}$$

Les trous diffusant de la région p vers la région intrinsèque, la différence de potentiel entre la région p et la région intrinsèque est bien négative. Dans le cas d'un semiconducteur de type n, on obtient :

$$V_1 - V_I = \phi_t \ln \frac{N_D}{n_i}$$

Si maintenant on met en contact deux matériaux quelconques et si on veut calculer le potentiel de contact en fonction des potentiels de contact de chacun d'eux relativement au silicium intrinsèque, il suffit de considérer l'assemblage de la figure 5.7.

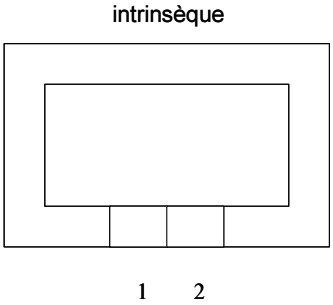


Figure 5. 7 : Calcul du potentiel de contact entre deux matériaux

Pour calculer le potentiel de contact entre les matériaux 1 et 2, il suffit d'écrire

$$V_1 - V_2 = V_1 - V_I + V_I - V_2$$

$$V_1 - V_2 = \phi_1 - \phi_2$$

De même, si on considère plusieurs matériaux en série comme dans la figure 5.8, on peut écrire

$$V_1 - V_n = V_1 - V_2 + V_2 - V_3 + ... + V_{n-1} - V_n$$

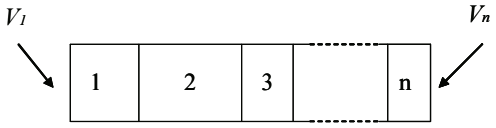


Figure 5.8 : Matériaux en série

Pour terminer de paragraphe, il reste à donner quelques valeurs de potentiel de contact relativement au silicium intrinsèque pour différents métaux et pour du silicium dopé.

Matériau	Potentiel de contact
Argent	-0.4 V
Or	-0.3 V
Cuivre	0 V
Nickel	0.15 V
Aluminium	0.6 V
Silicium *p*	$-\phi_t \ln N_A / n_i$
Silicium *n*	$\phi_t \ln N_D / n_i$
Silicium intrinsèque	0 V

On notera $\phi_F = \phi_t \ln \frac{N_A}{n_i}$ dans le cas d'un semiconducteur dopé p et $\phi_F = -\phi_t \ln \frac{N_D}{n_i}$ dans le cas d'un semi-conducteur dopé *n*. Le lien entre le potentiel de contact et le travail de sortie est facile à faire. Le contact métal-semiconducteur intrinséque permet d'écrire si ϕ_M est le potentiel de contact du métal , ϕ_m le travail de sortie du métal, E_g le gap du semiconducteur et χ l'affinité

$$\phi_M = \chi - \phi_m + \frac{E_g}{2}$$

5.3. Le modèle électrique de base

5.3.1. *Description phénoménologique*

Dans le chapitre 1, nous avons présenté un dispositif dans lequel un courant peut être commandé par une tension appliquée non pas directement sur le matériau mais à travers une mince couche d'oxyde. Rappelons les choix de ce matériau. Si le dispositif était isolant, le courant ne pourrait circuler. Si le matériau était un métal, le champ électrique ne pourrait pénétrer à l'intérieur et dans ce cas l'action de commande par la tension serait impossible. Le choix se porte alors naturellement vers les semiconducteurs.

La première étape est de comprendre comment une tension appliquée à travers un oxyde peut contrôler la charge stockée. C'est l'objectif de ce paragraphe. Le chapitre 8 montrera ensuite comment ce modèle évolue quand deux réservoirs de

charge sont placés de part et d'autre de ce dispositif et comment un courant peut circuler de l'un à l'autre.

Le dispositif de base est donc un empilement de trois matériaux : un semiconducteur dopé de type *p* ou *n*, un isolant de faible épaisseur (quelques nanomètres pour les technologies les plus avancées) et une couche métallique appelée grille. La technologie des circuits intégrés a conduit pendant de longues années à choisir du silicium fortement dopé à la place du métal pour des raisons qui seront expliquées ultérieurement mais le comportement est équivalent à celui d'un métal. Il faut ajouter à cela une couche métallique en face arrière pour prendre le contact électrique, en général de l'aluminium. Ce dispositif est représenté figure 5.9.

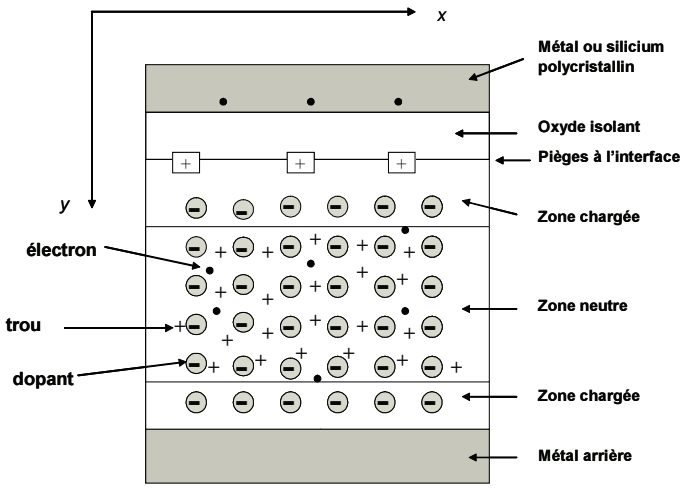


Figure 5.9 : Le dispositif Métal-Oxyde-Semiconducteur

Quelques commentaires sont nécessaires pour expliquer cette figure. Le silicium est de type *p* mais un dispositif équivalent peut être imaginé avec du silicium de type *n*. Le dopage du silicium est d'environ 10^{17} dopants par cm^3 soit 10^5 par μ^3. L'épaisseur de l'oxyde est d'environ 2 nm pour les technologies actuelles soit 2/1000 de micron ce qui représente quelques couches atomiques. L'épaisseur de silicium est relativement très importante et ne joue aucun rôle dans le fonctionnement électrique. La dimension longitudinale que nous noterons *x* est supposée très supérieure à la dimension transverse de la zone chargée correspondant à l'épaisseur du dispositif. La troisième dimension, non représentée, est également très supérieure à l'épaisseur. En pratique, un tel dispositif s'étend sur quelques microns de côté.

À l'interface entre le semiconducteur et l'oxyde, apparaît une fine couche de charges inhérente au processus de fabrication. Ces charges sont soit des impuretés soit des atomes de silicium ayant des liaisons manquantes. La densité est de l'ordre de 0.001 fC par μ^2 et leur charge est souvent positive. Le semiconducteur de base

est formé d'un grand nombre d'atomes de silicium non représentés sur la figure et neutres. Un faible nombre de dopants (des atomes de Bore pour le silicium *p*) sont ajoutés pendant la fabrication du dispositif. Ces atomes sont fixes, chargés négativement et associés chacun à un trou mobile symbolisé par une croix. Quelques paires électron-trou peuvent être créées thermiquement ce qui explique la présence de quelques électrons dans le silicium représentés par des billes noire.

Etudions maintenant le comportement de ce dispositif si on relie le métal au semiconducteur par un simple fil conducteur donc sans tension extérieure appliquée. Les niveaux de Fermi s'alignent alors. Des échanges de porteurs ont lieu entre les différentes régions dans lesquelles les concentrations sont différentes. La zone de silicium en contact avec l'oxyde est concernée à cause des charges présentes en surface qui créent un champ. Il en est de même pour la zone en contact avec le métal en face arrière car les électrons peuvent diffuser du métal vers le silicium. Comme dans la jonction *pn,* les mouvements de charge par diffusion conduisent à la création d'un champ électrique qui s'oppose à ce mouvement et les courants de diffusion et de dérive se compensent. Le courant total traversant le dispositif est nul mais des échanges ont lieu dans les deux zones blanchies sur le dessin dans lesquelles le champ électrique n'est pas nul. Dans la zone centrale grisée, la densité locale de charge est nulle et le champ électrique également ce qui veut dire que le potentiel est constant.

Détaillons les potentiels formés aux contacts en considérant cette fois un système complet incluant des électrodes capables d'appliquer une tension extérieure comme le montre le schéma 5.10. On suppose que les électrodes sont fabriquées avec le même métal. On appellera V_{bulk} le potentiel dans la zone neutre du dispositif. Il sera pris comme référence.

Dans une première étape, exprimons la différence de potentiel extérieure V_{ext} en fonction des différents potentiels présents dans le dispositif.

$$V_{ext} = V_A - V_G + V_G - V_{bulk} + V_{bulk} - V_M + V_M - V_B$$

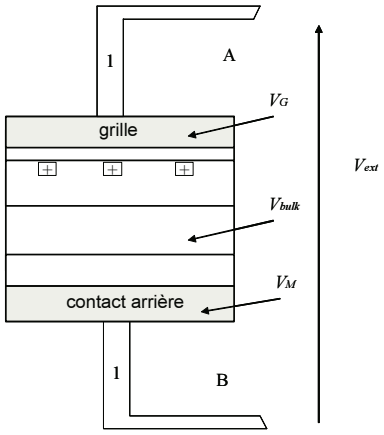


Figure 5.10 : Les potentiels de contact

Exprimons les potentiels par rapport au silicium intrinsèque comme il a été vu dans le paragraphe 2.

$$V_{ext} = \phi_1 - \phi_G + V_G - V_{bulk} + \phi_{si} - \phi_M + \phi_M - \phi_1$$

Les potentiels ϕ figurant dans cette formule sont les potentiels de contact par rapport à l'intrinsèque et sont donnés dans les tables. Ils sont notés ϕ_1 pour le métal de l'électrode, ϕ_{si} pour le semiconducteur, ϕ_M pour le métal face arrière et ϕ_G pour le matériau de grille. Il reste

$$V_{ext} = -\phi_G + V_G - V_{bulk} + \phi_{si}$$

On notera alors

$$\phi_{MS} = \phi_{si} - \phi_G$$

Cette notation est traditionnellement utilisée dans le monde des semi-conducteurs. Elle reste valable quand la grille n'est pas un métal mais du silicium polycristallin fortement dopé. Rappelons que ϕ_{MS} peut aussi s'exprimer en fonction des travaux de sortie. Il est facile de prouver avec ϕ_{ms} égal à $\phi_m - \phi_s$.

$$\phi_{MS} = \phi_{ms} - \phi_F$$

On obtient donc finalement

$$V_{ext} = V_G - V_{bulk} + \phi_{MS} \tag{5.4}$$

La tension présente dans le dispositif V_G-V_{bulk} dépend de la tension appliquée et des potentiels de Fermi du semiconducteur et du matériau de grille et absolument pas des autres matériaux utilisés dans le dispositif, contact arrière et matériaux des fils de contact. Il est utile de donner quelques valeurs numériques pour ϕ_{MS} en appliquant les formules du tableau du paragraphe 2.

Pour du silicium dopé p de dopage 10^5 par μ^3 en contact avec de l'aluminium, on trouve

$$\phi_{si} = -0.4 \text{ V}$$
$$\phi_G = 0.6 \text{ V}$$
$$\phi_{MS} = -0.4 \text{ V} - 0.6 \text{ V} = -1 \text{ V}$$

Si le métal de grille est remplacé par du silicium polycristallin fortement dopé, son potentiel de Fermi est environ 0.56 V, la valeur de ϕ_{MS} est alors peu différente

$$\phi_{MS} = -0.4 \text{ V} - 0.56 \text{ V} = -0.96 \text{ V}$$

Si le silicium de base est dopé *n* on peut calculer les potentiels de la même manière mais le potentiel de contact du semiconducteur est dans ce cas de signe positif. Appliquons maintenant une tension nulle entre grille et contact arrière comme le montre la figure 5.11. Les trous de la zone *p* sont repoussés vers l'intérieur par les ions en surface et une zone de charge d'espace chargée négativement apparaît. De même, une zone chargée négativement apparaît au niveau du contact arrière car des électrons du métal diffusent et se recombinent avec les trous.

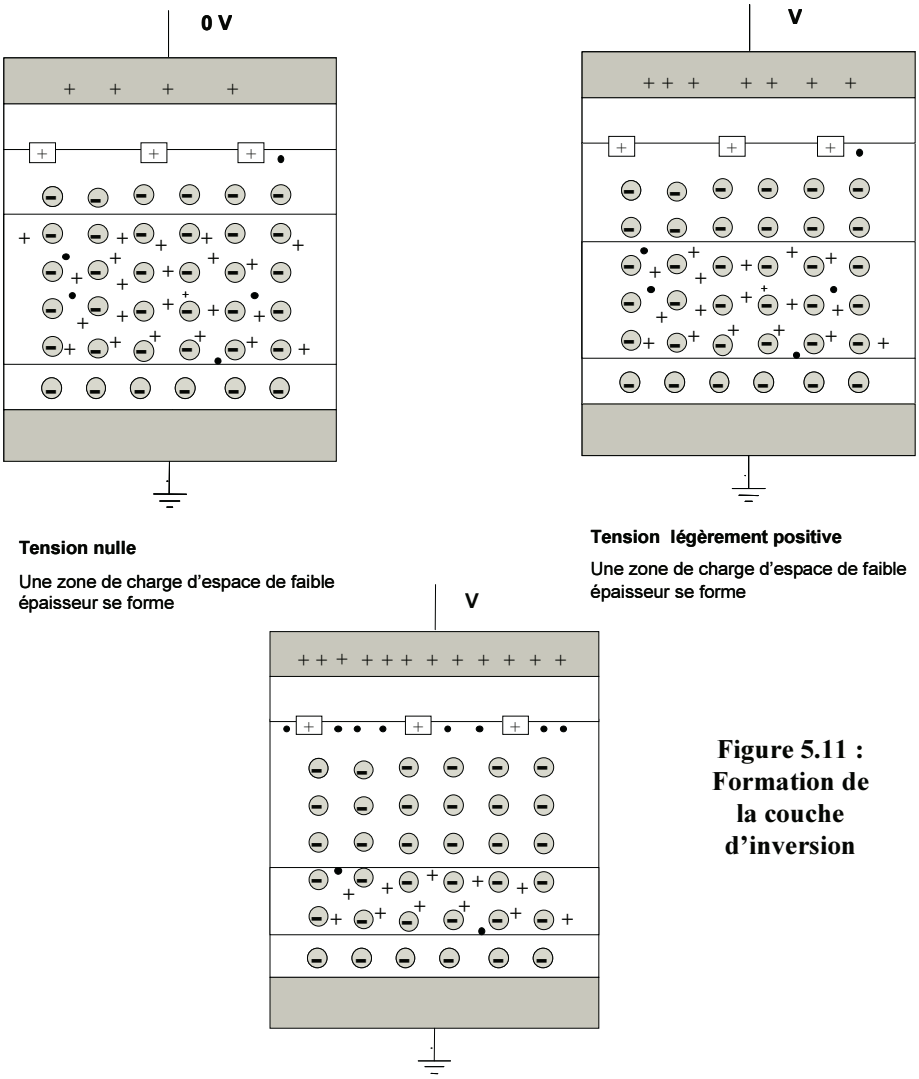


Figure 5.11 : Formation de la couche d'inversion

Si on augmente encore la tension sur la grille, la zone de charge d'espace au niveau de l'oxyde s'étend d'avantage. Celle créée au niveau du contact arrière reste inchangée car le potentiel de la grille n'a pas d'effet dans cette région. La figure 5.11 ne respecte pas les échelles car l'épaisseur de silicium neutre, représenté par la zone grisée, est en fait beaucoup plus importante que les zones de charge d'espace. Il y a un facteur mille environ.

Quand la tension augmente encore, elle finit par être assez importante pour attirer les quelques électrons présents dans le dispositif qui vont alors s'accumuler à l'interface oxyde-semiconducteur. Cet effet est appelé l'*inversion* du semi-conducteur puisqu'il se forme une couche de porteurs minoritaires. C'est le phénomène de base du fonctionnement du transistor. La figure 5.11 illustre la formation de la couche d'inversion.

Quand la tension augmente encore, cette couche d'électrons forme une sorte d'écran électrostatique entre la grille et le semiconducteur et l'excédent de tension est quasiment appliqué intégralement de part et d'autre de l'oxyde. Au fur et à mesure que la charge négative augmente dans le silicium, il se forme sur la grille une charge positive de la même manière que les deux armatures d'un condensateur se chargent avec des polarités opposées.

Il est possible de donner la profondeur de la zone de charge d'espace en dessous de l'oxyde de grille. Il suffit d'appliquer la loi de Poisson dans le silicium.

$$\varepsilon_s \frac{dE}{dy} = -eN_A$$

La densité est supposée égale à celle des dopants en suivant une approximation très classique.

En fonction du potentiel, on obtient

$$\varepsilon_s \frac{d^2V}{dy^2} = eN_A$$

On en déduit facilement la profondeur de la zone dans laquelle règne un champ électrique en fonction de la différence de potentiel ΔV entre la surface et la partie profonde du silicium dans laquelle le champ est nul.

$$y_b = \sqrt{\frac{2\varepsilon_s \Delta V}{eN_A}}$$

Quelques valeurs typiques donnent des ordres de grandeur.

$$\Delta V = 1V$$
$$\varepsilon_s = 0.104\ fF/\mu$$
$$e = 1.6\ 10^{-4}\ fC$$
$$N_A = 10^5/\mu^3$$

On calcule alors : $l \approx 0.1$ micron

5.3.2. *Modèle électrique*

L'objectif de ce paragraphe est d'exprimer la charge d'inversion en fonction de la tension appliquée. En effet, les propriétés de conduction des MOSFETs dépendent de la valeur de cette charge d'inversion responsable de la conduction. Le dispositif de base est représenté figure 5.12. La profondeur est notée *y* et l'origine est prise à l'interface oxyde- semiconducteur. Le potentiel représenté est la différence de potentiel entre un point à une profondeur *y* et un point en profondeur dans le matériau, région dans laquelle le champ est nul. Cette différence de potentiel est représentée sur la figure 5.12.

On suppose que le système est en équilibre thermodynamique. Le potentiel chimique du gaz d'électrons est alors égal à celui des trous. Cette hypothèse ne sera plus valable dans le transistor qui échange des charges avec l'extérieur et il sera nécessaire de définir des *pseudo-potentiels chimiques* différents pour les électrons et les trous. On part alors des relations classiques données dans le chapitre 1 pour exprimer les densités d'électrons et de trous.

$$n(y) = n_i e^{\frac{E_F - E_{Fi}(y)}{k\ T}} \qquad p(y) = n_i e^{\frac{E_{Fi}(y) - E_F}{k_B T}}$$

L'énergie E_{Fi} est environ au milieu du gap et dépend de la profondeur à cause du potentiel électrique qui n'est pas constant dans le dispositif. Les bandes de conduction et de valence varient en fonction du potentiel mais le gap reste constant.

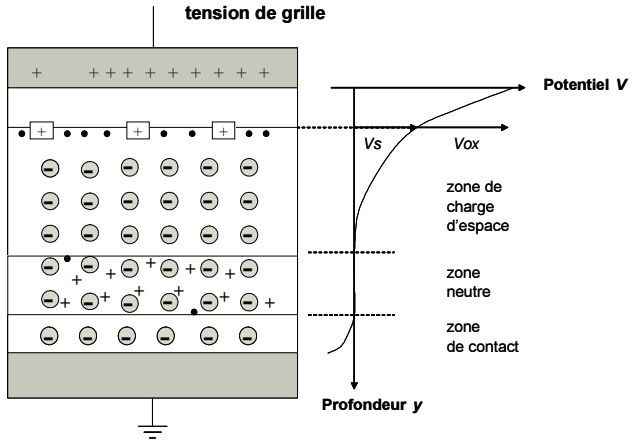


Figure 5.12 :

Le modèle électrique du dispositif MOS

Le potentiel chimique est lui constant par définition de l'équilibre thermodynamique. La valeur de n_i est à température ambiante de 0.015 par μ^3. On peut exprimer les valeurs de la concentration d'électrons n_s à l'interface oxyde-semi-conducteur et également la concentration n_b d'électrons en profondeur. En profondeur ou dans le bulk veut dire dans la région grisée sur la figure. Dans cette région, le champ est nul.

$$n_s = n_i e^{\frac{E_F - E_{Fis}}{k\,T}} \qquad n_b = n_i e^{\frac{E_F - E_{Fib}}{k\,T}}$$

De même, les densités de trous p_s à l'interface et en profondeur p_b sont

$$p_s = n_i e^{\frac{E_{Fis} - E_F}{k\,T}} \qquad p_b = n_i e^{\frac{E_{Fib} - E_F}{k\,T}}$$

De plus, il est admis que la densité de trous en profondeur est

$$p_b = N_A$$

On en déduit alors facilement

$$\frac{n_b}{N_A} = e^{\frac{2(E_F - E_{Fib})}{kT}} \qquad \frac{n_s}{n_b} = e^{\frac{E_{Fib} - E_{Fis}}{kT}}$$

Si on appelle V_s la différence entre le potentiel à l'interface et le potentiel en profondeur alors

$$E_{Fis} - E_{Fib} = -eV_s$$

Définissons maintenant le *potentiel de Fermi* ϕ_F du substrat par la relation

$$\phi_F = \frac{E_F - E_{ib}}{-e}$$

Le potentiel ainsi défini a un signe opposé à celui qui a été défini pour la diode et le transistor. Ce n'est qu'une simple convention qui permet d'avoir une valeur positive de ϕ_F pour un semiconducteur de type p utilisé dans le transistor MOS canal n. On rappelle que la quantité *kT/e* est notée ϕ_t .

La densité d'électrons à l'interface s'écrit alors

$$n_s = N_A e^{\frac{V_s - 2\phi_F}{\phi_t}} \tag{5.5}$$

Il reste à exprimer le potentiel de Fermi en fonction du dopage et la relation est établie entre la concentration des charges d'inversion et le potentiel de surface. La relation exprimant la concentration de trous dans le substrat permet d'écrire

$$N_A = n_i e^{\frac{\phi_F}{\phi_t}}$$

On en déduit :

$$\phi_F = \phi_t \ln \frac{N_A}{n_i}$$

Le potentiel de Fermi est positif pour un semiconducteur de type *p*, il est négatif pour un semi-conducteur de type *n* et donné par la relation

$$\phi_F = -\phi_t \ln \frac{N_D}{n_i}$$

On retrouve bien les valeurs données dans l'étude des potentiels de contact.

En pratique, le potentiel de Fermi varie assez peu avec le dopage étant donné la relation logarithmique. Il est compris entre 0.2 V et 0.5 V. La courbe 5.13 indique les valeurs possibles.

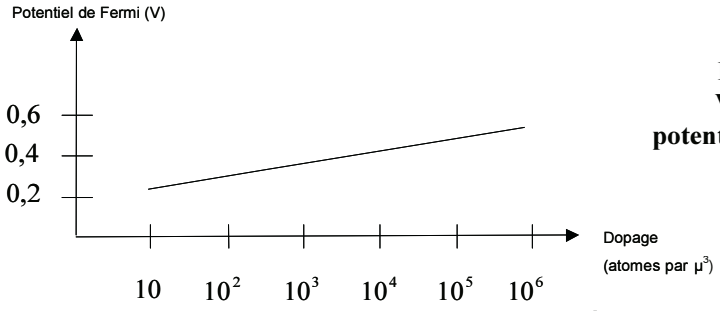


Figure 5 13 : Variation du potentiel de Fermi

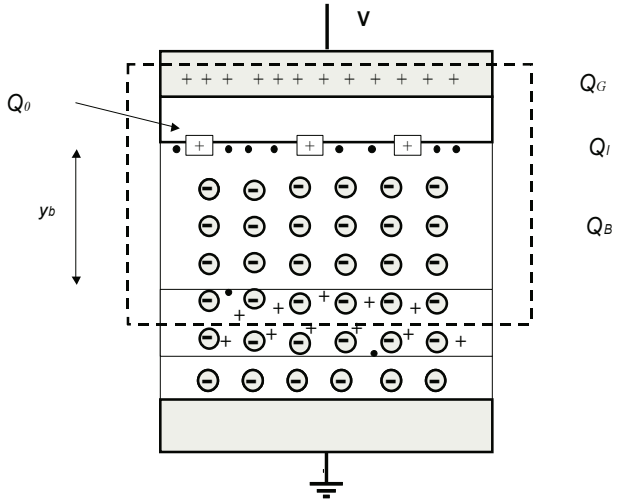


Figure 5.14 : Les charges du problème

Les expressions précédentes font appel aux résultats de la physique des solides établis dans le chapitre 1.

Pour exprimer la densité des électrons d'inversion en fonction du potentiel appliqué il est nécessaire de considérer les différentes charges présentes dans le système en s'appuyant sur la figure 5.14.

Les différents types de charge sont les suivants :

- Les charges positives Q_G accumulées à la surface de la grille à l'interface avec l'oxyde.

L'électromagnétisme montre qu'il n'y a pas de charge dans les métaux. Dans le cas d'une grille en silicium polycristallin fortement dopé, on fera la même hypothèse.

- Les électrons minoritaires d'inversion Q_I .

Ils forment une couche à l'interface entre le semiconducteur et l'oxyde et l'épaisseur de cette couche est infiniment mince.

- Les charges positives Q_0, fixes, à l'interface entre l'oxyde et le semiconducteur.
- Les ions dopants privés de leurs trous Q_B dans la zone de charge d'espace.

Elle s'étend sur une profondeur y_b de moins d'un micron.

- La densité des électrons mobiles dans le semiconducteur notée *n(y).*

Les électrons de la couche d'inversion correspondent à *n(0)* .

- La densité des trous mobiles notée *p(y).*

On ne prend pas en compte la charge au niveau du contact arrière car on exprime le potentiel entre la grille et l'intérieur du silicium et non pas entre la grille et le contact arrière. Cette charge apparaîtrait comme une constante dans les développements qui vont suivre.

Pour éviter toute confusion, reprenons l'expression 5.4 exprimant le potentiel appliqué.

$$V_{ext} = V_G - V_{bulk} + \phi_{MS}$$

Dans la suite du calcul, les potentiels sont mesurés relativement au potentiel en profondeur dans le silicium, appelé V_{bulk} , comme il est indiqué sur la figure 5 .12. On notera en particulier

$$V_s = V(0) - V_{bulk}$$
$$V_{ox} = V_G - V(0)$$

On écrit donc

$$V_{ext} = V_G - V_{bulk} + \phi_{MS} = V_{ox} + V_s + \phi_{MS}$$

La résolution de ce système suppose connus la densité des charges de surface $Q_{0,}$ le dopage du semiconducteur N_A et la tension appliquée à la grille V_{ext}.

Les inconnues du problème sont les charges (Q_I, Q_B, Q_G), les densités de charge (*n(y)* et *p(y)*), les potentiels ($V(y)$, V_s, V_{ox}) ainsi que la profondeur de la zone de charge d'espace y_b .

Les densités d'électrons *nb* et de trous *pb* sont également inconnues. Il y a au total 11 inconnues. Il faut donc écrire 11 équations indépendantes pour résoudre le problème. Elles seront notées de 5.6 à 5.16. La première est déjà établie.

$$V_{ext} = V_{ox} + V_s + \phi_{MS} \tag{5.6}$$

Les considérations précédentes ont montré que ϕ_{MS} était donné par des tables.

La deuxième équation exprime la densité de charge *n(y)* . Le calcul est équivalent à celui effectué précédemment pour calculer la densité de charge à l'interface**.**

$$\frac{n(y)}{n_b} = e^{\frac{V(y)}{\phi_t}} \tag{5.7}$$

De même,

$$\frac{p(y)}{p_b} = e^{-\frac{V(y)}{\phi_t}} \tag{5.8}$$

Dans la zone neutre

$$p_b - n_b = N_A \tag{5.9}$$

De plus

$$n_b p_b = n_i^2 \tag{5.10}$$

La charge sur la grille est reliée à la différence de potentiel V_{ox} par la relation des condensateurs plans. On raisonnera plus facilement sur les charges par unité de surface notée prime pour ne pas les confondre avec les charges globales. La charge de la grille par unité de surface s'écrit donc

$$Q'_G = C'_{ox} V_{ox} \tag{5.11}$$

La valeur de C'_{ox}, capacité par unité de surface, est donnée par la relation classique des condensateurs plans et ne dépend que des dimensions du dispositif et de la constante diélectrique de l'oxyde, toutes grandeurs connues.

Il est possible d'écrire la loi de Poisson reliant densité et potentiel.

$$\frac{d^2V}{dy^2} + e\left(\frac{p(y) - n(y) - N_A}{\varepsilon_s}\right) = 0 \tag{5.12}$$

Il est également possible de calculer la charge d'inversion par unité de surface en fonction de la densité d'électrons, il suffit d'intégrer dans toute la zone de charge d'espace

$$Q_I' = \int_0^{y_b} -en(y)dy \tag{5.13}$$

La profondeur de la zone de charge d'espace est la solution de l'équation

$$\left(\frac{dV}{dy}\right)_{y=y_b} = 0 \tag{5.14}$$

La charge Q'_b par unité de surface peut s'écrire

$$Q_B' = \int_0^{y_b} -eN_A dy \tag{5.15}$$

Enfin, il est possible d'écrire une règle de conservation de la charge. Cette règle est toujours délicate à appliquer car il n'est pas évident d'identifier les charges qui réellement se conservent. La manière la plus sûre est d'appliquer le théorème de Gauss en trouvant une surface au travers de laquelle le flux du champ est nul. La surface identifiée figure 5.14 est intéressante car le champ est ou nul ou parallèle aux faces considérées.

On peut donc écrire

$$Q_G' + Q_0' + Q_I' + Q_B' = 0 \tag{5.16}$$

On a négligé dans cette formule la charge des trous restant dans la zone de charge d'espace. On suppose qu'ils sont tous repoussés ou recombinés. Pour résoudre ce système on procède de la manière suivante

On part de l'équation de Poisson.

$$\frac{d^2V}{dy^2} + e\left(\frac{p(y) - n(y) - N_A}{\varepsilon_s}\right) = 0$$

$$\frac{n(y)}{n_b} = e^{\frac{V(y)}{\phi_t}} \quad \text{et} \quad \frac{p(y)}{p_b} = e^{-\frac{V(y)}{\phi_t}}$$

De plus,

$$p_b - n_b = N_A$$

On obtient alors

$$\frac{d^2V}{dy^2}+\frac{e}{\varepsilon_s}\left(p_b\left(e^{-\frac{V(y)}{\phi_t}}-1\right)-n_b\left(e^{\frac{V(y)}{\phi_t}}-1\right)\right)=0$$

Si on suppose N_A largement supérieur à n_b, on obtient

$$\frac{d^2V}{dy^2}+\frac{eN_A}{\varepsilon_s}\left(e^{-\frac{V(y)}{\phi_t}}-1-e^{-\frac{2\phi_F}{\phi_t}}\left(e^{\frac{V(y)}{\phi_t}}-1\right)\right)=0$$

Si on multiplie cette équation par $2\frac{dV}{dy}$, on obtient alors

$$\frac{d}{dy}\left(\frac{dV}{dy}\right)^2+\frac{2eN_A}{\varepsilon_s}\left(\frac{dV}{dy}e^{-\frac{V(y)}{\phi_t}}-\frac{dV}{dy}-\frac{dV}{dy}e^{-\frac{2\phi_F}{\phi_t}}\left(e^{\frac{V(y)}{\phi_t}}-1\right)\right)=0$$

Cette équation s'intègre entre un point y et un point en profondeur dans le silicium pour lequel le champ $\frac{dV}{dy}$ est nul ainsi que le potentiel $V(y)$.

$$\left(\frac{dV}{dy}\right)^2=\frac{2eN_A}{\varepsilon_s}\left(\phi_t e^{-\frac{V(y)}{\phi_t}}+V(y)-\phi_t+e^{-\frac{2\phi_F}{\phi_t}}\left(\phi_t e^{\frac{V(y)}{\phi_t}}-V(y)-\phi_t\right)\right) \quad (5.17)$$

Cette équation ne peut se résoudre analytiquement qu'en effectuant d'autres approximations.

5.4. Le régime de forte inversion

Le principe du calcul est de calculer le champ électrique à l'interface soit pour y égal à 0. L'équation générale établie précédemment et le théorème de Gauss permettront alors de calculer facilement la charge d'inversion. La figure 5.15 illustre l'application du théorème de Gauss au calcul du champ à l'interface.

La surface délimitée sur la figure comprend la charge d'inversion mais pas les charges positives à l'interface oxyde-semiconducteur. Elle passe dans la partie neutre du dispositif. Le flux à travers cette surface est donc égal à la somme des charges à l'intérieur ce qui s'écrit

$$\left(-\frac{dV}{dy}\right)_{y=0}=\frac{Q'_I+Q'_B}{\varepsilon_s}$$

Reprenons maintenant l'équation générale en l'appliquant à y égal à zéro et en supposant que le potentiel V_s est très supérieur à ϕ_t soit 26 mV à température ambiante.

$$\left(\frac{dV}{dy}\right)^2_{y=0} = \frac{2eN_A}{\varepsilon_s}\left[\phi_t e^{-\frac{V_s}{\phi_t}} + V_s - \phi_t + e^{-2\frac{\phi_F}{\phi_t}}\left(\phi_t e^{\frac{V_s}{\phi_t}} - V_s - \phi_t\right)\right]$$

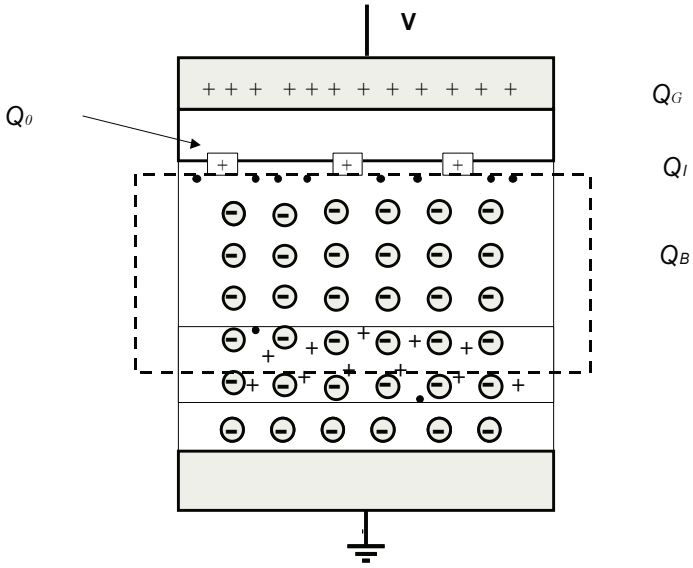


Figure 5.15 : Le calcul du champ à l'interface

De plus, le quotient ϕ_F / ϕ_t est largement supérieur à l'unité comme le montre le graphique 5.13. La formule se simplifie, il reste

$$\left(\frac{dV}{dy}\right)^2_{y=0} = \frac{2eN_A}{\varepsilon_s}\left[V_s + \phi_t e^{\frac{V_s - 2\phi_F}{\phi_t}}\right]$$

On en déduit donc

$$\left(\frac{dV}{dy}\right)_{y=0} = \pm\left(\frac{2eN_A}{\varepsilon_s}\left[V_s + \phi_t e^{\frac{V_s - 2\phi_F}{\phi_t}}\right]\right)^{\frac{1}{2}}$$

ensuite

$$Q'_I + Q'_B = \pm\varepsilon_s\sqrt{\frac{2eN_A}{\varepsilon_s}\left(V_s + \phi_t \exp^{\frac{V_s - 2\phi_F}{\phi_t}}\right)}$$

On choisira pour des raisons physiques la valeur négative. La charge qui nous intéresse est uniquement Q_I, il faut donc maintenant calculer Q_B. Pour cela, il suffit de supposer que la charge d'inversion est infiniment mince. La charge Q_B est alors le produit de la densité de dopant par le volume de la zone de charge d'espace de profondeur y_b. Ce calcul s'effectue simplement en écrivant l'équation de Poisson dans le semiconducteur.

$$\varepsilon_s \frac{d^2V}{dy^2} = eN_A$$

On en déduit que V varie avec le carré de y,

$$\varepsilon_s V = \frac{eN_A}{2}(y - y_b)^2$$

On en déduit la valeur de y_b en fonction de V_s

$$\varepsilon_s V_s = eN_A \frac{y_b^{\,2}}{2}$$

Le calcul de Q'_B est alors simple ainsi que celui de Q'_I.

$$Q'_B = -\sqrt{2eN_A\varepsilon_s}\sqrt{V_s}$$

La *charge d'inversion* est donc

$$Q'_I = -\sqrt{2eN_A\varepsilon_s}\left[\sqrt{V_s + \phi_t \exp^{\frac{V_s - 2\phi_F}{\phi_t}}} - \sqrt{V_s}\right] \qquad (5.18)$$

Cette relation est fondamentale et mérite d'être un peu commentée. La charge d'inversion est proportionnelle à la surface du dispositif et au dopage. Elle est nulle quand le potentiel de surface est faible devant le potentiel de Fermi puis augmente exponentiellement au-delà de cette valeur.

Rappelons que le potentiel de surface n'est pas le potentiel appliqué sur la grille. Il reste encore à établir la relation entre ces deux potentiels. Pour cela, il faut revenir aux équations générales du dispositif complétées par la relation établie précédemment. Le potentiel extérieur appliqué V_{ext} sera noté V_G.

$$Q'_G + Q'_0 + Q'_I + Q'_B = 0$$

$$V_G = V_{ox} + V_s + \phi_{MS}$$

$$Q'_G = C'_{ox}V_{ox}$$

$$Q'_I + Q'_B = -\varepsilon_s \sqrt{\frac{2eN_A}{\varepsilon_s}\left(V_s + \phi_t e^{\frac{V_s - 2\phi_F}{\phi_t}}\right)}$$

Ce système de quatre équations permet de calculer les charges et potentiels inconnus : $Q'_I+Q'_B$, Q'_G, V_s et V_{ox}. On en déduit facilement

$$V_G = \phi_{MS} - \frac{Q'_0}{C'_{ox}} + V_s + \frac{1}{C'_{ox}}\sqrt{2eN_A\varepsilon_s}\sqrt{V_s + \phi_t e^{\frac{V_s - 2\phi_F}{\phi_t}}} \qquad (5.19)$$

Cette relation permet de calculer le potentiel de surface en fonction de la tension extérieure appliquée entre grille et contact arrière. Il n'y a pas de solution analytique. L'hypothèse que nous allons faire maintenant est uniquement valable en forte inversion, c'est-à-dire pour des tensions appliquées suffisamment élevées. On suppose que le potentiel de surface est constant. Cette hypothèse semble peu fondée. Il suffit cependant d'examiner les résultats fournis par la résolution numérique de l'équation donnant le potentiel de surface pour s'apercevoir qu'il varie assez peu quand la tension appliquée sur la grille augmente. La figure 5.16 illustre ce comportement.

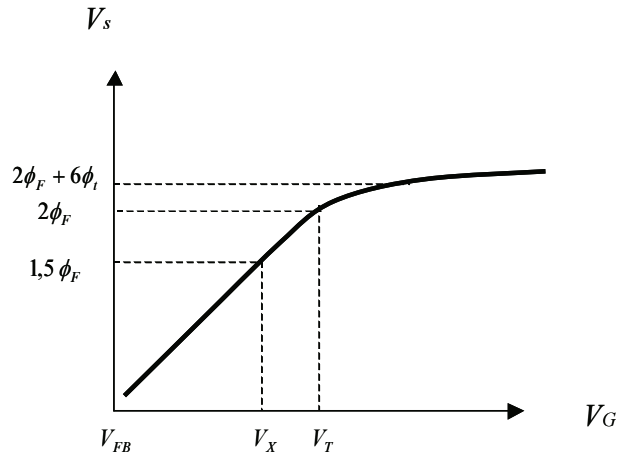


Figure 5.16 : Variation du potentiel de surface

Tout se passe comme si, au-delà d'une certaine valeur, toute la tension était appliquée aux bornes de la couche d'oxyde. Si on considère l'écran électrique formé par la couche d'inversion, ce phénomène apparaît peu surprenant. La suite du calcul consiste à prendre une valeur bien choisie pour cette valeur constante. Elle est notée ϕ_B. Certains auteurs choisissent alors $2\phi_F$ en se basant sur le terme

exponentiel de la formule. D'autres choisiront une valeur légèrement supérieure en prenant par exemple $2\phi_F + 6\phi_t$.

Reprenons maintenant certaines des équations générales du système pour calculer la charge d'inversion.

$$\begin{aligned} Q'_G + Q'_0 + Q'_I + Q'_B &= 0 \\ V_G &= V_{ox} + V_s + \phi_{MS} \\ Q'_G &= C'_{ox} V_{ox} \\ Q'_B &= -\sqrt{2eN_A\varepsilon_s}\sqrt{V_s} \end{aligned}$$

À partir de ces quatre équations et en prenant une valeur donnée ϕ_B du potentiel de surface, on obtient facilement

$$Q'_I = -C'_{ox}\left(V_G - V_T\right) \tag{5.20.a}$$

Dans cette formule le paramètre V_T est appelé *tension de seuil*. Il est défini par

$$V_T = \phi_{MS} - \frac{Q'_0}{C'_{ox}} + \phi_B + \frac{1}{C'_{ox}}\sqrt{2eN_A\varepsilon_s}\sqrt{\phi_B} \tag{5.20.b}$$

Ces deux formules sont fondamentales dans l'étude des MOS. Elles quantifient la formation de la couche d'inversion. La tension de seuil dépend du paramètre ϕ_{MS} variant lui-même avec la nature du matériau de grille, des charges de surface et d'un terme en racine de la tension. Ce dernier terme est appelé effet « body » dans la littérature. Le coefficient $\sqrt{2eN_A\varepsilon_s}/C'_{ox}$ est appelé coefficient γ.

	0.80 micron	**0.18 micron**	**0.045 micron**
Epaisseur de l'oxyde en nm	14	4	1.5
***C'ox* en *fF*/µ²**	2	4	25
N_A en atomes/cm³	10^{14}	10^{15}	10^{16}
Zone de charge d'espace en micron	0.8	0.10	0.05
Charges de surface en *FC*/µ²	0.1	0.08	0.001
ϕ_B en V	0.75	0.80	0.85
γ en V$^{1/2}$	0.4	0.2	0.15
Tension de seuil en V	0.26	0.18	0.13

Pour donner à cette formule un sens physique, on peut dire que le potentiel doit franchir un certain nombre de barrières pour créer une charge d'inversion : le potentiel de contact ϕ_{MS} , puis la barrière des charges de surface, puis la barrière du silicium et enfin la barrière de la charge de la zone de charge d'espace. Pour terminer cette description du dispositif MOS, il est possible de donner quelques valeurs numériques des grandeurs physiques pour différentes technologies.

Les valeurs de ce tableau sont soit données par la technologie soit calculées par les formules des paragraphes précédents. C'est le cas de la profondeur de la zone de charge d'espace, du coefficient γ et de la tension de seuil. La valeur de la capacité de l'oxyde par unité de surface est prise à 0.035 fF par micron d'épaisseur pour du dioxyde de silicium. On constate une diminution de la tension de seuil quand la technologie est plus fine, ce qui est un phénomène très important dans l'industrie du semiconducteur.

5.5. Le régime de faible inversion

Dans le régime de *faible inversion*, la tension appliquée sur la grille est plus faible mais le potentiel de surface est encore supposé très supérieur à ϕ_t. Par contre, il est inférieur à $2\phi_F$. Les approximations de la formule générale ne sont plus valables. Ce régime était exceptionnel dans les anciennes technologies mais, pour les technologies avancées et en particulier pour les circuits faible consommation, ce régime de fonctionnement est relativement standard.

Reprenons l'équation générale 5.18 en supposant cette fois que le terme $\phi_t e^{\frac{V_s-2\phi_F}{\phi_t}}$ noté alors ξ est largement inférieur à V_s. L'équation s'écrit rappelons le

$$Q'_I = -\sqrt{2eN_A\varepsilon_s}\left[\sqrt{V_s+\phi_t e^{\frac{V_s-2\phi_F}{\phi_t}}}-\sqrt{V_s}\right]$$

soit

$$Q'_I = -\sqrt{2eN_A\varepsilon_s}\left[\sqrt{V_s+\xi}-\sqrt{V_s}\right]$$

Un développement limité de la racine carrée permet d'écrire

$$Q'_I = -\sqrt{2eN_A\varepsilon_s}\,\frac{\xi}{2\sqrt{V_s}}$$

Reprenons alors les équations du système.

$$Q'_G + Q'_0 + Q'_I + Q'_B = 0$$

$$V_G = V_{ox} + V_s + \phi_{MS}$$

$$Q'_G = C'_{ox} V_{ox}$$

$$Q'_B = -\sqrt{2eN_A\varepsilon_s}\sqrt{V_s}$$

$$Q'_I = -\sqrt{2eN_A\varepsilon_s}\frac{\xi}{2\sqrt{V_s}}$$

Pour simplifier le calcul, on néglige la charge d'inversion Q'_I devant la charge Q'_B ce qui permet d'écrire à partir des quatre premières équations

$$V_G = \phi_{MS} - \frac{Q'_0}{C'_{ox}} + V_s + \gamma\sqrt{V_s}$$

Cette équation du deuxième degré en $\sqrt{V_s}$ permet de calculer le potentiel de surface en fonction du potentiel extérieur appliqué. Une seule des racines a un sens physique. On peut ensuite calculer la charge d'inversion avec cette valeur.

Il est cependant possible de simplifier encore en considérant que le potentiel de surface a une valeur donnée égale à $1.5\,\phi_F$ dans le terme en racine carrée. C'est une valeur intermédiaire entre 0 et $2\,\phi_F$, valeur à partir de laquelle commence le régime d'inversion forte. On exprime alors la pente de la fonction donnant la valeur de V_G au voisinage de $1.5\,\phi_F$ par

$$n_0 = \frac{dV_G}{dV_s} = 1 + \frac{\gamma}{2\sqrt{1.5\,\phi_F}}$$

On peut alors calculer le potentiel de surface en fonction de la tension de grille en faisant une approximation linéaire.

$$V_S = \frac{1}{n_0}\left(V_G - \phi_{MS} + \frac{Q'_0}{C'_{ox}}\right)$$

Appelons V_X la valeur de la tension de grille pour laquelle le potentiel de surface est égal à $1.5\,\phi_F$.

$$V_S - 1.5\,\phi_F = \frac{V_G - V_X}{n_0}$$

$$V_X = \phi_{MS} - \frac{Q'_0}{C'_{OX}} + 1.5\phi_F + \gamma\sqrt{1.5\,\phi_F}$$

Cette tension V_X est donc équivalente à la tension de seuil en régime de forte inversion.

La relation donnée au début de ce paragraphe permet de calculer la charge d'inversion

$$Q'_I = -\sqrt{2e\varepsilon_s N_A}\,\frac{\phi_t e^{\frac{-0.5\phi_F}{\phi_t}}}{2\sqrt{1.5\,\phi_F}}\,e^{\frac{V_G - \phi_{MS} + \frac{Q_0}{C_{ox}} - 1.5\phi_F - \gamma\sqrt{1.5\,\phi_F}}{\left(1+\frac{\gamma}{2\sqrt{1.5\,\phi_F}}\right)\phi_t}}$$

Il est maintenant possible d'exprimer cette charge en fonction de la tension V_X. On obtient alors :

$$Q'_I = Q'_{I0}\,e^{\frac{V_G - V_X}{n_0\phi_t}} \tag{5.21}$$

Dans cette formule, plus lisible, les paramètres sont définis par

$$Q'_{I0} = -\sqrt{2e\varepsilon_s N_A}\,\frac{\phi_t e^{\frac{-0.5\phi_F}{\phi_t}}}{2\sqrt{1.5\,\phi_F}}$$

$$n_0 = 1 + \frac{\gamma}{2\sqrt{1.5\,\phi_F}}$$

Il est également intéressant d'exprimer le facteur *n* appelé « *ideality* » en fonction des capacités du dispositif. Il est défini à partir de l'équation obtenue précédemment. Il dépend de la tension de surface et donc de la tension appliquée sur la grille.

$$V_G = \phi_{MS} - \frac{Q'_0}{C'_{ox}} + V_s + \gamma\sqrt{V_s}$$

$$n = \frac{dV_G}{dV_s} = 1 + \frac{\gamma}{2\sqrt{V_S}}$$

Comme le coefficient γ est donné par $\sqrt{2e\varepsilon_s N_A}\,/\,C'_{ox}$, on obtient

$$\frac{\gamma}{2\sqrt{V_S}} = \frac{\sqrt{2e\varepsilon_s N_A}}{2C'_{ox}\sqrt{V_S}}$$

soit,

$$\frac{\gamma}{2\sqrt{V_S}} = \frac{C'_{si}}{C'_{ox}}$$

Dans cette formule, on reconnaît la capacité de la zone de charge d'espace et la capacité de la couche d'oxyde.

$$n = \frac{dV_G}{dV_s} = 1 + \frac{C'_{si}}{C'_{ox}} \tag{5.22}$$

Cette relation importante sera reprise dans l'étude du MOSFET. La valeur de n est idéalement un car dans ce cas la grille contrôle totalement le potentiel dans le silicium. En pratique, elle est comprise entre 1 et 1.6. On comprend l'intérêt d'augmenter la capacité de l'oxyde et de diminuer la capacité de la zone de charge d'espace dans le silicium. Le paramètre n_0 est la valeur particulière de n pour V_S égal à 1.5 ϕ_F .

EXERCICES DU CHAPITRE 5

- **Exercice 1 : Structure de bandes**

Représentez les bandes d'une structure aluminium-silice-silicium n pour différents dopages du silicium : 10^{14}, 10^{15}, $10^{16}/cm^3$. Même exercice avec du titane. Les travaux de sortie sont tableau 1-3.

- **Exercice 2 : Les porteurs minoritaires**

Expliquez d'où viennent les électrons de la charge d'inversion dans une structure MOS en régime d'inversion.

- **Exercice 3 : Tension de seuil**

En supposant nulles les charges de surface, calculez la tension de seuil d'une capacité MOS formée d'un contact d'aluminium, d'une couche de silice de 4 nm et d'un silicium p dopé à $10^{15}/cm^3$.

- **Exercice 4 : Capacité de la sructure MOS**

Expliquez comment varie la capacité mesurée entre métal et face arrière quand la tension varie à partir de 0V.

- **Exercice 5 : Régime de faible inversion**

Comparez les tensions de seuil définies en faible et en forte inversion.

- **Exercice 6 : NMOS**

Reprendre les calculs de l'exercice 1 mais pour un silicium dopé p.

- **Exercice 7 : Calcul de la capacité MOS**

Avec les données de l'exercice 3 calculez la capacité par unité de surface de la structure pour les tensions suivantes : 0 V, 0,5 V, 2 V, 10 V.

- **Exercice 8 : Zone de charge d'espace**

Pour un dopage p de $5\ 10^{16}/cm^3$ calculez la valeur maximale de la largeur de la zone de charge d'espace.

- **Exercice 9 : Les charges piégées**

Les charges dans l'oxyde ont une densité de $5\ 10^{11}/cm^2$. Elles forment une couche fine dans un oxyde de 10 nm d'épaisseur et sont au milieu (à 5 nm). Calculez la

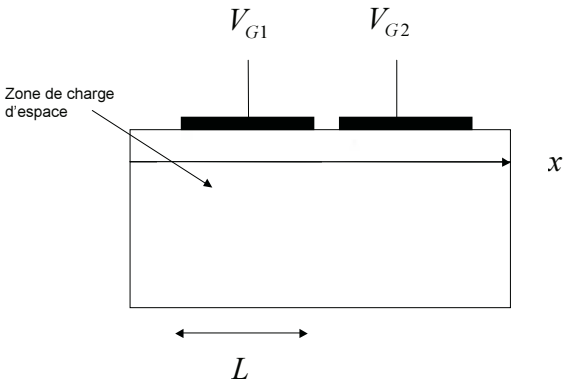
variation de la tension de seuil du dispositif. Expliquez comment cet effet peut être utilisé pour créer une cellule mémoire.

- **Exercice 10 : Potentiel de contact**

Etablir la relation entre le potentiel de contact, le travail de sortie et l'affinité.

- **Exercice 11 : CCD**

C'est un ensemble de capacités MOS les unes à côté des autres. On cherche à faire passer la charge d'inversion d'une cellule à l'autre de manière unidirectionnelle. La charge d'inversion dans la première cellule est créée par exemple par effet photoélectrique.
Montrez que ce transfert est possible si les zones de charge d'espace se recouvrent en partie.



Montrez que dans une cellule on peut écrire avec les notations classiques

$$V_s = V_G + \frac{Q_I}{C_{OX}}$$

La grille de la première cellule est portée à V_0, valeur assez basse, puis à V_1 pendant un temps t_1. Ensuite la tension de la grille de la première cellule retourne linéairement vers V_0 alors que la grille de la deuxième cellule passe de V_0 à V_1.
Montrez que le transfert est favorisé par un champ longitudinal qui se forme pendant le transfert.
Montrez que l'équation donnant la densité d'électrons en surface s'écrit

$$\frac{\partial n(x,t)}{\partial t} = \frac{\partial}{\partial x}\left[\left(\frac{\mu_n e}{C'_{OX}} n(x,t) + D_n\right)\frac{\partial n(x,t)}{\partial x}\right]$$

Au début du transfert le processus dominant est le champ induit. Montrez que la solution est du type

$$n(x,t) = g(t)h(x)$$
$$avec \;\; g(t) = g_0 \frac{t_0}{t - t_0}$$

A la fin du transfert, la diffusion thermique domine. Montrez qu'une solution approchée s'écrit

$$n(x,t) = a_0 e^{-\frac{t}{\tau_d}} \cos\frac{\pi x}{2L}$$

avec

$$\tau_d = \frac{4L^2}{\pi^2 D_n}$$

L est la longueur de la cellule selon l'axe x.

Montrez qu'en utilisant trois signaux de commande de grille décalés dans le temps, on peut déplacer la charge initiale de cellule à cellule avec une bonne efficacité de transfert.
Expliquez comment on peut stocker une image.

6. Hétérojonctions et transistor à hétérojonction

Quand deux semiconducteurs différents sont mis en contact, l'interface présente des propriétés particulières qui peuvent être mises en application pour certains dispositifs. On appelle alors ce dispositif *hétérojonction*. Les composants optroniques en particulier mettent à profit ces propriétés . Ce chapitre reprend donc les trois phases de l'analyse des interfaces : l'étude des bandes, l'étude de l'hérérojonction à l'équilibre thermodynamique c'est-à-dire non polarisée par une source externe et enfin l'étude de l'hétérojonction polarisée. Enfin, le transistor à hétérojonction sera présenté comme une application de ce chapitre.

6.1. Diagramme de bandes d'énergie
6.2. Hétérojonction à l'équilibre thermodynamique
6.3. Hétérojonction polarisée
6.4. Transistor à hétérojonction

6.1. Diagramme de bandes d'énergie

Lorsque deux semiconducteurs différents sont au contact, il apparaît une barrière de potentiel à l'interface, donnée par (voir chapitre 1).

$$E_b = e(\chi_1 - \chi_2) \tag{6-1}$$

où $e\chi_1$ et $e\chi_2$ représentent les affinités électroniques des semiconducteurs. La forme de cette barrière et le diagramme énergétique correspondant, sont en outre fonction des gaps des semiconducteurs et de leurs dopages respectifs.

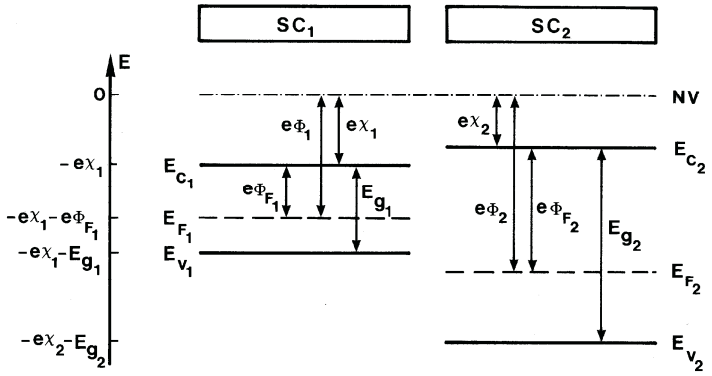


Figure 6-1 : Diagramme énergétique dans chacun des semiconducteurs

Considérons deux semiconducteurs caractérisés par les affinités électroniques $e\chi_1$ et $e\chi_2$, les gaps E_{g1} et E_{g2} et les dopages représentés par les distances $e\phi_{F1}$ et $e\phi_{F2}$ du niveau de Fermi à la bande de conduction. En l'absence de tout contact les diagrammes énergétiques dans chacun des semiconducteurs sont représentés sur la figure (6-1).

L'axe des énergies est orienté positivement vers le haut, l'axe des potentiels est orienté positivement vers le bas. On a choisi comme origine des énergies, l'énergie potentielle d'un électron extrait du semiconducteur 1 au voisinage de celui-ci. Les deux semiconducteurs étant supposés non chargés, ce zéro est le même pour les deux. Le même diagramme représente les énergies potentielles E des électrons et le potentiel électrique V.

Compte tenu de la définition de l'affinité électronique, la bande de conduction de chaque semiconducteur est située à la distance $e\chi_j$ (j=1,2) au-dessous du zéro de l'énergie, c'est-à-dire à l'abscisse $-e\chi_j$ sur l'axe des énergies. De même les niveaux de Fermi et les bandes de valence sont situés aux abscisses $-e\chi_j - e\phi_{Fj}$ et $-e\chi_j - E_{gj}$ respectivement. Les deux semiconducteurs étant indépendants, les distributions

électroniques sont indépendantes et caractérisées par des niveaux de Fermi différents.

Amenons les deux semiconducteurs au contact, et étudions le diagramme énergétique d'une part loin de la jonction et d'autre part au niveau de la jonction.

6.1.1. *Diagramme énergétique loin de la jonction*

Lorsque les deux semiconducteurs sont mis au contact, ils échangent des électrons de manière à aligner leurs niveaux de Fermi. Cet échange se fait au voisinage de la jonction et fait apparaître, comme dans la jonction pn, une charge d'espace à laquelle est associée une barrière de potentiel (la tension de diffusion V_d) qui arrête la diffusion des porteurs et définit l'état d'équilibre. Nous reviendrons plus loin sur cette région, considérons tout d'abord le diagramme énergétique loin de la jonction, dans la région neutre de chacun des semiconducteurs. Ce diagramme est représenté sur la figure (6-2). Nous avons choisi comme origine des énergies l'énergie potentielle de l'électron dans le vide au voisinage du semiconducteur 1, soit NV_1=0.

A partir de cette origine, le niveau de Fermi de la structure est fixé à la distance $e\phi_1$ au-dessous, $e\phi_1$ représente le travail de sortie du semiconducteur 1. A partir de ce niveau on peut positionner E_{c1} , E_{v1} , E_{c2} , et E_{v2}. Le niveau NV_2 de l'électron dans le vide au voisinage du semiconducteur 2 est situé au-dessus de E_F à la distance $e\phi_2$.

La différence d'énergie potentielle entre l'électron dans le vide au voisinage du semiconducteur 1, et l'électron dans le vide au voisinage du semiconducteur 2 est

$$NV_2 - NV_1 = e(\phi_2 - \phi_1) \tag{6-2}$$

Il en résulte que la différence de potentiel entre les deux semiconducteurs, c'est-à-dire la tension de diffusion, est donnée par

$$V_d = V_2 - V_1 = \phi_1 - \phi_2 \tag{6-3}$$

La condition $\phi_2 > \phi_1$ entraîne $V_d < 0$, le potentiel du semiconducteur 2 est inférieur à celui du semiconducteur 1. Il s'établit une différence de potentiel positive entre le semiconducteur à faible travail de sortie et le semiconducteur à fort travail de sortie.

Les différences de densité d'états et de dopage des semiconducteurs entraînent des valeurs différentes des paramètres $e\phi_{F1}$ et $e\phi_{F2}$, c'est-à-dire des valeurs différentes des énergies des bandes de conduction des régions neutres des deux semiconducteurs. Appelons ΔE_{cn} la différence d'énergie entre les bandes de conduction des régions neutres

$$\Delta E_{cn} = E_{c2} - E_{c1} = e(\phi_{F2} - \phi_{F1}) \tag{6-4}$$

Rappelons que comme dans le MOS le potentiel de Fermi ϕ_F est défini par

$$\phi_F = \frac{E_F - E_{F_I}}{-e}$$

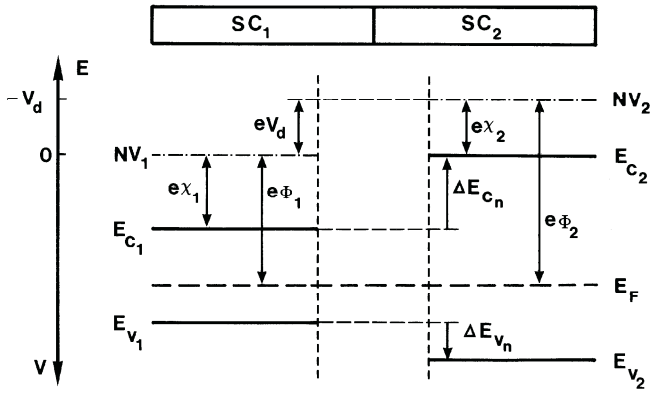


Figure 6-2 : Diagramme énergétique loin de l'interface

Enfin si la différence des gaps des deux semiconducteurs est différente de ΔE_{cn} le complément se traduit par une différence d'énergie des bandes de valence avec la condition

$$\Delta E_g = E_{g2} - E_{g1} = (E_{c2} - E_{v2}) - (E_{c1} - E_{v1}) = (E_{c2} - E_{c1}) - (E_{v2} - E_{v1})$$

soit

$$\Delta E_g = \Delta E_{cn} - \Delta E_{vn} \tag{6-5}$$

La différence d'énergie des bandes de valence est par conséquent donnée par

$$\Delta E_{vn} = e(\phi_{F2} - \phi_{F1}) - \Delta E_g \tag{6-6}$$

Sur la figure (6-2) les conditions sont $\Delta E_{cn} > 0$ *et* $\Delta E_{vn} < 0$.

Il en résulte que le signe de ΔE_{vn} est fonction à la fois des signes et des amplitudes de $e(\phi_{F2}\text{-}\phi_{FI})$ et ΔE_g. Dans chacun des semiconducteurs on peut écrire

$$n_1 = N_{c1}\, e^{-(E_{c1} - E_F)/kT} \qquad p_1 = N_{v1}\, e^{(E_{v1} - E_F)/kT}$$

$$n_2 = N_{c2}\, e^{-(E_{c2} - E_F)/kT} \qquad p_2 = N_{v2}\, e^{(E_{v2} - E_F)/kT}$$

soit

$$\frac{n_1}{n_2}=\frac{N_{c1}}{N_{c2}}e^{(E_{c2}-E_{c1})/kT} \qquad \frac{p_1}{p_2}=\frac{N_{v1}}{N_{v2}}e^{(E_{v1}-E_{v2})/kT}$$

ainsi

$$E_{c2}-E_{c1}=kT\ Ln\frac{n_1\,N_{c2}}{n_2\,N_{c1}} \qquad E_{v2}-E_{v1}=-kT\ Ln\frac{p_1\,N_{v2}}{p_2\,N_{v1}}$$

On obtient donc les relations

$$\Delta E_{cn}\ =\ kT\ Ln\frac{n_1\,N_{c2}}{n_2\,N_{c1}} \tag{6-7-a}$$

$$\Delta E_{vn}\ =\ -\,kT\ Ln\frac{p_1\,N_{v2}}{p_2\,N_{v1}} \tag{6-7-b}$$

Dans une homojonction $N_{c1}=N_{c2}$ et $N_{v1}=N_{v2}$ de sorte que ΔE_{cn} et ΔE_{vn} s'écrivent

$$\Delta E_{cn}=kT\ Ln\ \frac{n_1}{n_2} \qquad \Delta E_{vn}=-\,kT\ \ Ln\ \frac{p_1}{p_2}$$

$$\Delta E_{cn}-\Delta E_{vn}=kT\ Ln\frac{n_1\,p_1}{n_2p_2}=kT\ Ln\frac{n_i^2}{n_i^2}=0$$

La différence entre l'hétérojonction et l'homojonction réside dans le fait que dans une homojonction ΔE_{cn} et ΔE_{vn} , qui sont directement donnés par la tension de diffusion, ne sont fonction que des dopages respectifs des deux parties du semiconducteur. Dans une hétérojonction ΔE_{cn} et ΔE_{vn} sont fonction, d'une part des dopages respectifs et d'autre part des paramètres intrinsèques de chacun des matériaux. Afin d'expliciter ces deux types de contribution revenons sur les expressions (6-4 et 6)

$$\begin{aligned}\Delta E_{cn}&=e(\phi_{F2}-\phi_{F1})=e\big((\phi_2-\chi_2)-(\phi_1-\chi_1)\big)\\&=e\big((\phi_2-\phi_1)-(\chi_2-\chi_1)\big)=-eV_d-e(\chi_2-\chi_1)\end{aligned}$$

$$\begin{aligned}\Delta E_{vn}&=\Delta E_{cn}-\Delta E_g=-eV_d-e(\chi_2-\chi_1)-\Delta E_g\\&=-eV_d-\big(e(\chi_2-\chi_1)+\Delta E_g\big)\end{aligned}$$

Ainsi ΔE_{cn} et ΔE_{vn} , les différences des énergies des bandes de conduction et de valence des régions neutres des semiconducteurs, sont composés de deux termes dont l'un est spécifique des propriétés intrinsèques des matériaux et l'autre fonction de leurs dopages respectifs.
Posons

$$\Delta E_{ci}=-\,e(\chi_2-\chi_1) \tag{6-8-a}$$

$$\Delta E_{vi} = -\left(\Delta E_g + e(\chi_2 - \chi_1)\right) \tag{6-8-b}$$

On obtient les relations

$$\Delta E_{cn} = \Delta E_{ci} - eV_d \tag{6-9-a}$$
$$\Delta E_{vn} = \Delta E_{vi} - eV_d \tag{6-9-b}$$

et

$$\Delta E_{ci} - \Delta E_{vi} = \Delta E_{cn} - \Delta E_{vn} = \Delta E_g \tag{6-10}$$

6.1.2. *Etats d'interface*

La *règle de l'affinité électronique* selon laquelle la discontinuité des bandes de conduction est directement donnée par la différence des affinités électroniques doit être utilisée avec circonspection. En fait la discontinuité des bandes induit la présence d'états d'interface dans chacun des semiconducteurs. Ces états sont chargés et créent des dipôles dont le potentiel réduit la discontinuité des bandes. Cet écrantage électrostatique de la discontinuité peut être très important et devenir le facteur dominant dans l'établissement du diagramme énergétique (J. Tersoff, 1984).

Notons en outre dans une approche plus phénoménologique, que la relaxation importante des positions atomiques à l'interface peut entraîner l'existence de couches d'interface dont les affinités électroniques sont différentes de celles des matériaux isolés.

D'une part parce que le problème du calcul exact de la discontinuité des bandes reste ouvert et d'autre part parce que ce problème sort largement du cadre de cet ouvrage, nous supposerons valable dans ce qui suit, la règle de l'affinité électronique.

6.1.3. *Diagramme énergétique au voisinage de la jonction*

En raison de la différence des travaux de sortie, les électrons diffusent du semiconducteur à plus faible travail de sortie vers l'autre. Cette diffusion entraîne l'apparition d'une zone de charge d'espace, positive dans le semiconducteur à faible travail de sortie, négative dans l'autre. Comme dans l'homojonction, la tension de diffusion augmente et s'établit à la valeur qui arrête la diffusion et définit l'état d'équilibre.

Il faut noter que le potentiel est continu à la jonction entre les deux semiconducteurs, et qu'il varie de façon monotone si les deux semiconducteurs sont dopés de manière homogène.

Nous avons défini des paramètres qui caractérisent le diagramme énergétique loin de la jonction. Ces paramètres, ΔE_{cn} et ΔE_{vn} représentent les différences d'énergie des bandes de conduction et de valence des régions neutres des semiconducteurs en fonction d'une part des paramètres intrinsèques ΔE_{ci} et ΔE_{vi} et

d'autre part de la différence de potentiel qui existe entre ces deux régions $V_d=V_2-V_1$. Les expressions (6-9-a,b) peuvent être généralisées en remplaçant V_d par V_2-V_1.

$$\Delta E_{cn}=\Delta E_{ci}-e\left(V_2-V_1\right) \quad (6\text{-}11\text{-a})$$
$$\Delta E_{vn}=\Delta E_{vi}-e\left(V_2-V_1\right) \quad (6\text{-}11\text{-b})$$

On peut ainsi écrire les différences d'énergie des bandes de conduction et de valence dans deux cas spécifiques. le premier concerne la structure polarisée. La différence de potentiel entre les deux régions extrêmes de la structure est alors donnée par $V_2-V_1=V_d-V$ où V est la tension de polarisation du semiconducteur 1 par rapport au semiconducteur 2.

Les expressions (6-11-a,b) s'écrivent alors

$$\Delta E_{cn}(V)=\Delta E_{ci}-e\left(V_d-V\right) \quad (6\text{-}12\text{-a})$$
$$\Delta E_{vn}(V)=\Delta E_{vi}-e\left(V_d-V\right) \quad (6\text{-}12\text{-b})$$

Le deuxième cas, qui nous intéresse plus particulièrement pour le moment, concerne les différences d'énergie des bandes de conduction et de valence, à l'interface. En raison de la continuité du potentiel, à l'interface $V_2=V_1$ les expressions (6-11-a,b) s'écrivent

$$\Delta E_c\left(x=0\right)=\Delta E_{co}=\Delta E_{ci} \quad (6\text{-}13\text{-a})$$
$$\Delta E_v\left(x=0\right)=\Delta E_{vo}=\Delta E_{vi} \quad (6\text{-}13\text{-b})$$

Au voisinage de l'interface, le diagramme énergétique varie en fonction de la nature des semiconducteurs, c'est-à-dire des valeurs de ΔE_{ci} et ΔE_{vi} et en fonction de leurs dopages, c'est-à-dire de la différence de leurs travaux de sortie.

a. premier cas : $\phi_1 = \phi_2$

En l'absence de polarisation

$$V_2-V_1=V_d=-\frac{1}{e}\left(e\phi_2-e\phi_1\right)=\phi_1-\phi_2$$

Les travaux de sortie des semiconducteurs étant égaux, la différence de potentiel V_2-V_1 est nulle, de sorte que

$$\Delta E_{cn}=\Delta E_{co}=\Delta E_{ci}$$

$$\Delta E_{vn}=\Delta E_{vo}=\Delta E_{vi}$$

Les différences d'énergie des bandes de conduction et de valence sont les mêmes à l'interface et loin de l'interface, la structure est en *régime de bandes plates*.

La condition $\Delta E_g \neq 0$ entraîne l'existence de quatre cas de figure suivant que ΔE_{ci} et ΔE_{vi} sont positifs ou négatifs, c'est-à-dire suivant les valeurs relatives des affinités électroniques et des gaps des semiconducteurs

La condition	$\chi_1 - \chi_2 > 0$	entraîne	$\Delta E_c > 0$
La condition	$\chi_1 - \chi_2 > \Delta E_g / e$	entraîne	$\Delta E_v > 0$

Les différents cas sont représentés sur la figure (6-3). Notons que dans chacun d'eux, on peut représenter différents types de dopage des semiconducteurs en déplaçant verticalement le niveau de Fermi. On peut ainsi réaliser des hétérostructures de types nn, np, pp ou pn.

b. deuxième cas : $\phi_1 \neq \phi_2$

Si les travaux de sortie sont différents, les électrons diffusent du semiconducteur à faible travail de sortie vers l'autre. Une zone de charge d'espace s'établit au voisinage de l'interface. Supposons par exemple le travail de sortie du semiconducteur 2 inférieur au travail de sortie du semiconducteur 1,($\phi_2 < \phi_1$). Les électrons diffusent du semiconducteur 2 vers le semiconducteur 1, et vice-versa pour les trous. Il apparaît une charge d'espace, positive dans le semiconducteur 2 et négative dans le semiconducteur 1. Cette charge d'espace est de nature et d'extension spatiale tout à fait différentes suivant le type de chaque semiconducteur.

Dans le semiconducteur 1 la charge d'espace est négative. Si ce semiconducteur est de type n, cette charge d'espace est due à une augmentation de la densité d'électrons au voisinage de l'interface. Le semiconducteur 1 est en régime d'accumulation. Ces électrons sont distribués dans la bande de conduction dont la densité d'états est relativement importante, de sorte qu'ils sont localisés au voisinage de l'interface sur une distance inférieure à 100 Å. Supposons par exemple une densité d'électrons par unité de surface de 10^{12} cm^{-2} (ce qui correspond à une densité volumique de 10^{18} cm^{-3}). Avec une densité équivalente d'états de la bande de conduction à la température ambiante $N_c = 10^{19}$ cm^{-3} la distance sur laquelle s'étalent ces électrons est de l'ordre de 10^{-7} cm, soit 10 Å. En fait l'utilisation de la densité équivalente d'états résulte de l'approximation de Boltzmann, valable pour des semiconducteurs non dégénérés. Or pour des densités de porteurs de l'ordre de 10^{18} cm^{-3} la plupart des semiconducteurs sont dégénérés de sorte que le calcul simple fait précédemment n'est qu'approximatif, toutefois il respecte l'ordre de grandeur. En résumé, si le semiconducteur 1 est de type n, la charge d'espace est une charge d'accumulation et elle est localisée au voisinage immédiat de l'interface.

Si le semiconducteur 1 est de type p les électrons qui diffusent depuis le semiconducteur 2 se recombinent avec les trous à leur entrée dans le semiconducteur 1.

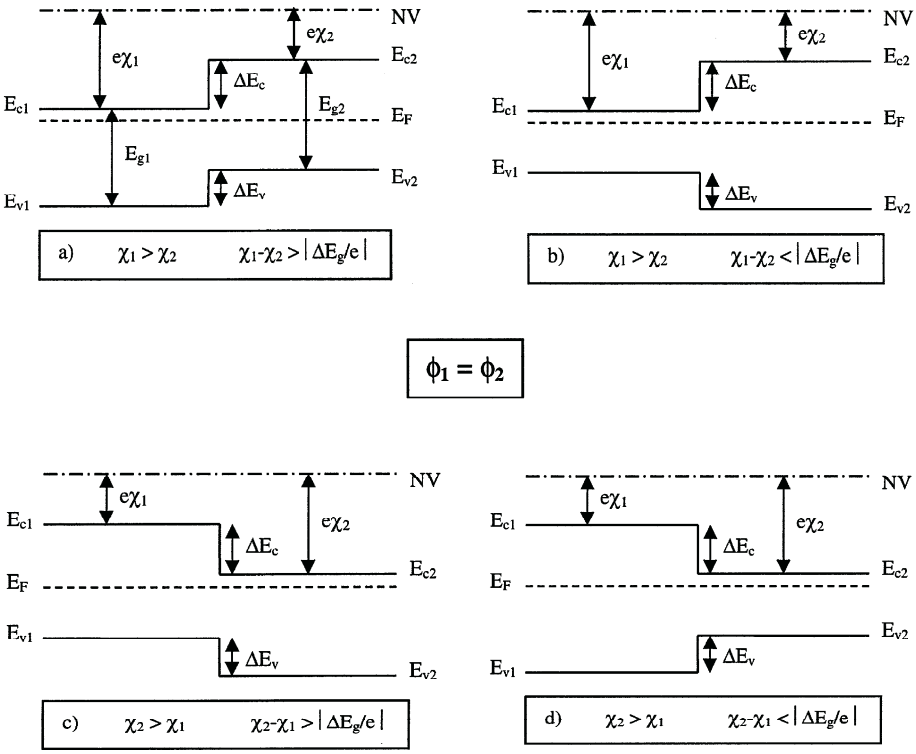


Figure 6-3 : Diagramme énergétique d'une hétérojonction entre deux semiconducteurs différents ayant même travail de sortie $\phi_1=\phi_2$

Ils font apparaître une charge d'espace résultant des ions accepteurs non compensés par les trous. La charge d'espace résulte de la disparition des trous, le semiconducteur 1 est en régime de déplétion. Ce régime est celui qui correspond au fonctionnement de l'homojonction pn.

Contrairement à la charge d'accumulation qui en raison de la grande densité d'états de la bande de conduction était localisée à l'interface, la charge de déplétion, en raison de la relativement faible densité d'accepteurs, est étendue sur une distance de l'ordre de 1000 à 10 000 Å suivant le dopage. Reprenons par exemple la densité superficielle de charges de 10^{12} cm^{-2}, avec une densité d'accepteurs N_a=10^{16} cm^{-3}, la distance sur laquelle s'étend cette charge de déplétion est de l'ordre de 10^{-4} cm, soit 1 μm.

Dans le semiconducteur 2, d'où partent les électrons, la charge d'espace est au contraire positive. Si ce semiconducteur est de type n il s'établit, au voisinage de l'interface, un régime de déplétion avec une certaine extension spatiale de la densité de charge. Si le semiconducteur est de type p, il s'établit un régime d'accumulation.

Compte tenu des différentes valeurs possibles des paramètres χ_1-χ_2 et ΔE_g les divers cas possibles sont représentés sur la figure (6-4). L'allure générale de chacun des cas reste la même quel que soit le type du semiconducteur considéré, seule l'extension de la charge d'espace varie, suivant sa nature.

Les différents cas de figure présentent des caractéristiques particulières. Dans le cas de la figure (6-4-a) le régime d'équilibre s'établit par diffusion des électrons du semiconducteur 2 vers le semiconducteur 1 et vice-versa pour les trous. Dans le cas de la figure (6-4-b) les électrons diffusent du semiconducteur 2 vers le semiconducteur 1, mais en raison du signe de ΔE_{vo} les trous ne peuvent pas diffuser du semiconducteur 1 vers le semiconducteur 2. La distribution de trous à l'interface se fait par équilibre thermodynamique avec la distribution d'électrons qui diminue dans le semiconducteur 2, et augmente dans le semiconducteur 1. Le cas de la figure (6-4-d) est inverse de celui de la figure (6-4-b), seuls les trous diffusent. Le cas de la figure (6-4-c) est encore différent. Compte tenu des barrières ΔE_{co} et ΔE_{vo} ni les électrons ni les trous ne peuvent diffuser. Quelques porteurs peuvent toutefois diffuser, soit en franchissant la barrière de potentiel par agitation thermique, soit en passant au travers par effet tunnel. Mais en fait la mise en équilibre de la structure se réalise ici dans chacun des semiconducteurs par le transfert de porteurs de l'interface vers le volume ou inversement. Dans le semiconducteur 2 les électrons sont repoussés de l'interface et les trous sont attirés. Dans le semiconducteur 2 il se produit le phénomène inverse.

Dans chacun des cas, la tension de diffusion V_d s'établit en partie dans chacun des semiconducteurs.

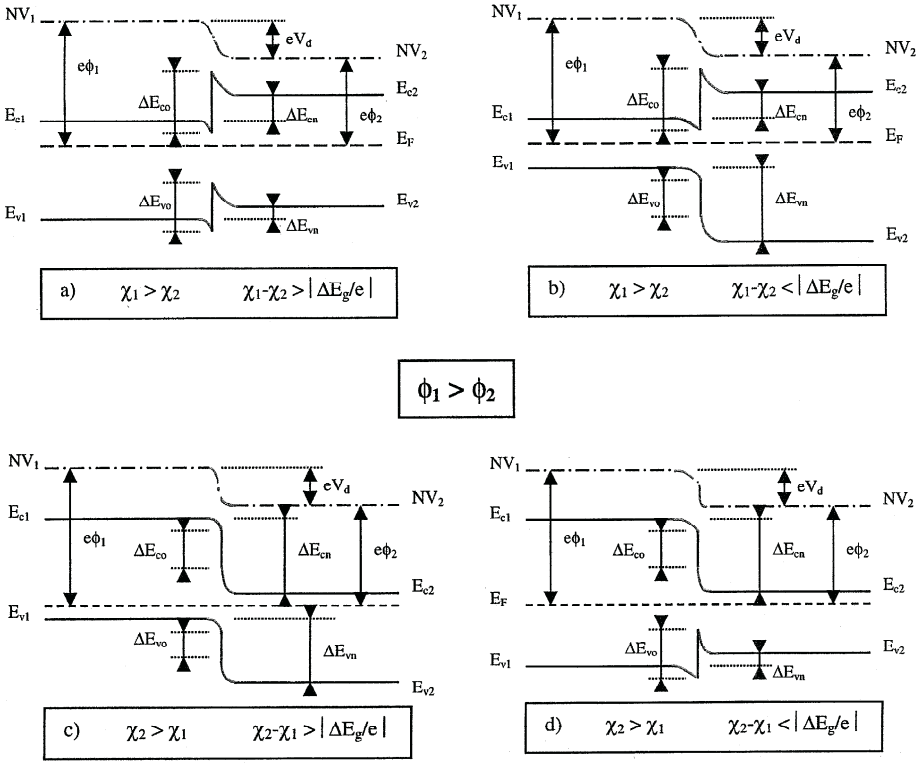


Figure 6-4 : Diagramme énergétique d'une hétérojonction entre deux semiconducteurs différents avec $\phi_2 < \phi_1$

6.2. Hétérojonction à l'équilibre thermodynamique

Le calcul de la distribution du potentiel au voisinage de l'interface se fait comme dans le cas de l'homojonction pn par intégration de l'équation de Poisson. En fait, ce calcul est relativement compliqué dans le cas où la charge d'espace résulte d'un régime d'accumulation pour plusieurs raisons. D'abord parce que le potentiel est fonction de la distribution de charges, mais les charges étant des porteurs libres, leur distribution est fonction du potentiel. Il en résulte que le calcul doit être auto-cohérent et ne présente pas de solution analytique. Ensuite, étant donné la grande densité de porteurs dans la couche d'accumulation, il faut tenir compte des corrélations électron-électron. Enfin, en raison du fait que cette charge est localisée très près de l'interface, il faut tenir compte du potentiel image associé à ces charges.

Le calcul peut être développé en régime de déplétion en supposant, comme dans le cas de l'homojonction pn, que la zone de déplétion est vide de porteurs libres. Pour que la zone de charge d'espace soit de déplétion dans chacun des semiconducteurs, il faut que ces derniers soient de types différents.

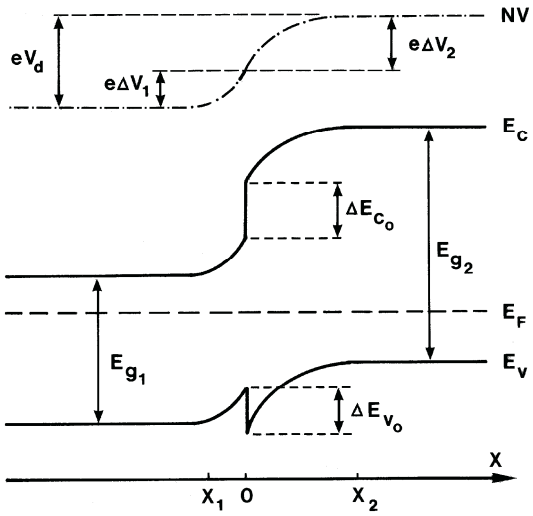


Figure 6-5 : Hétérojonction Ge(n)-GaAs(p)

Considérons l'hétérojonction Ge(n)-GaAs(p) représentée sur la figure (6-5). Nous supposerons les semiconducteurs dopés de manière homogène. Nous appellerons N_{d1} l'excès de donneurs N_d-N_a dans le semiconducteur 1, et N_{a2} l'excès d'accepteurs N_a-N_d dans le semiconducteur 2. L'équation de Poisson s'écrit

$$\frac{d^2V(x)}{dx^2} = -\frac{\rho(x)}{\varepsilon} \tag{6-14}$$

Dans le semiconducteur 1, $\rho(x)= e\,N_{d1}$

$$\frac{d^2 V(x)}{dx^2} = -\frac{e\,N_{d1}}{\varepsilon_1} \tag{6-15}$$

En intégrant une fois avec la condition E=0 en x=x_1 on obtient

$$\frac{dV(x)}{dx} = -E(x) = -\frac{eN_{d1}}{\varepsilon_1}\left(x - x_1\right) \tag{6-16}$$

en x=0

$$E_{s1} = -\frac{e\,N_{d1}}{\varepsilon_1}\,x_1 \tag{6-17}$$

En intégrant une deuxième fois et en appelant V_1 le potentiel de la région neutre du semiconducteur 1, on obtient

$$V(x) = -\frac{eN_{d1}}{2\varepsilon_1}\left(x - x_1\right)^2 + V_1 \tag{6-18}$$

Dans le semiconducteur 2, $\rho(x)$=-$e\,N_{a2}$

$$\frac{d^2 V(x)}{dx^2} = \frac{e\,N_{a2}}{\varepsilon_2} \tag{6-19}$$

$$\frac{dV(x)}{dx} = -E(x) = \frac{eN_{a2}}{\varepsilon_2}\left(x - x_2\right) \tag{6-20}$$

en x=0

$$E_{s2} = \frac{e\,N_{a2}}{\varepsilon_2}\,x_2 \tag{6-21}$$

$$V(x) = \frac{eN_{a2}}{2\varepsilon_2}\left(x - x_2\right)^2 + V_2 \tag{6-22}$$

Le champ et le potentiel électriques sont représentés sur la figure (6-6). La continuité du vecteur déplacement à l'interface s'écrit

$$\varepsilon_1\,E_{s1} = \varepsilon_2\,E_{s2} \tag{6-23}$$

$$-e\,N_{d1}\,x_1 = e\,N_{a2}\,x_2 \tag{6-24}$$

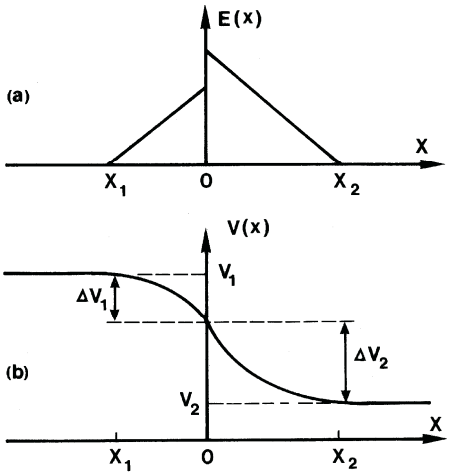


Figure 6-6 : Champ et potentiel électriques à l'interface d'une hétérojonction Ge(n)-GaAs(p)

ou, en posant $W_1 = |x_1|$ et $W_2 = |x_2|$

$$N_{d1} W_1 = N_{a2} W_2 \tag{6-25}$$

La continuité du potentiel en x=0 s'écrit

$$-\frac{e N_{d1}}{2\varepsilon_1} W_i^2 + V_1 = \frac{e N_{a2}}{2\varepsilon_2} W_2^2 + V_2 \tag{6-26}$$

$$V_1 - V_2 = \frac{e N_{d1}}{2\varepsilon_1} W_1^2 + \frac{e N_{a2}}{2\varepsilon_2} W_2^2 \tag{6-27}$$

ou en utilisant la relation (6-25)

$$V_1 - V_2 = \frac{e N_{d1}}{2\varepsilon_1} W_1^2 \left(\frac{\varepsilon_1 N_{d1} + \varepsilon_2 N_{a2}}{\varepsilon_2 N_{a2}} \right) = \frac{e N_{a2}}{2\varepsilon_2} W_2^2 \left(\frac{\varepsilon_1 N_{d1} + \varepsilon_2 N_{a2}}{\varepsilon_1 N_{d1}} \right) \tag{6-28}$$

d'où les expressions de la largeur de la zone de charge d'espace dans chacun des semiconducteurs

$$W_1 = \left(\frac{2 N_{a2}}{e N_{d1}} \frac{\varepsilon_1 \varepsilon_2}{\varepsilon_1 N_{d1} + \varepsilon_2 N_{a2}} \right)^{1/2} (V_1 - V_2)^{1/2} \tag{6-29-a}$$

$$W_2=\left(\frac{2}{e}\frac{N_{d1}}{N_{a2}}\frac{\varepsilon_1\varepsilon_2}{\varepsilon_1 N_{d1}+\varepsilon_2 N_{a2}}\right)^{1/2}\left(V_1-V_2\right)^{1/2} \tag{6-29-b}$$

La largeur totale de la zone de déplétion est donnée par $W=W_1+W_2$

$$W=\left(\frac{2\varepsilon_1\varepsilon_2\left(N_{d1}+N_{a2}\right)^2}{eN_{d1}N_{a2}\left(\varepsilon_1 N_{d1}+\varepsilon_2 N_{a2}\right)}\right)^{1/2}\left(V_1-V_2\right)^{1/2} \tag{6-30}$$

La différence de potentiel V_1-V_2 s'établit en partie dans chacun des semiconducteurs, le rapport des chutes de potentiel correspondantes est donné par

$$\frac{\Delta V_1}{\Delta V_2}=\frac{\left(V(x=0)-V_1\right)_{sc1}}{\left(V(x=0)-V_2\right)_{sc2}}=\frac{eN_{d1}W_1^2/2\varepsilon_1}{eN_{a2}W_2^2/2\varepsilon_2}=\frac{\varepsilon_2 N_{d1}W_1^2}{\varepsilon_1 N_{a2}W_2^2} \tag{6-31}$$

Soit en utilisant les relations (6-29-a,b)

$$\frac{\Delta V_1}{\Delta V_2}=\frac{\varepsilon_2 N_{a2}}{\varepsilon_1 N_{d1}} \tag{6-32}$$

En l'absence de polarisation extérieure la différence de potentiel V_1-V_2 correspond à la tension de diffusion. En présence d'une polarisation V du semiconducteur 2 par rapport au semiconducteur 1 , cette différence devient V_1-V_2=V_d-V. La tension de diffusion V_d est distribuée entre les deux semiconducteurs dans le rapport

$$\frac{\Delta V_{d1}}{\Delta V_{d2}}=\frac{\varepsilon_2 N_{a2}}{\varepsilon_1 N_{d1}} \tag{6-33}$$

Comme dans le cas de l'homojonction pn ou de la diode Schottky, toute variation de V entraîne une modulation de la largeur W de la zone de charge d'espace et par suite une modulation de la charge développée dans chacun des semiconducteurs. Il en résulte que la structure présente une capacité différentielle. La charge d'espace est donnée par

$$Q_{sc1}=-Q_{sc2}=e\,N_{d1}W_1=e\,N_{a2}W_2 \tag{6-34}$$

Soit en explicitant W_1 donné par l'expression (6-29-a), et V_1-V_2 par V_d-V

$$Q=\left(\frac{2e\varepsilon_1\varepsilon_2 N_{d1}N_{a2}}{\varepsilon_1 N_{d1}+\varepsilon_2 N_{a2}}\right)^{1/2}\left(V_d-V\right)^{1/2} \tag{6-35}$$

La capacité différentielle est donnée par

$$C(V)=\left|\frac{dQ}{dV}\right|=\left(\frac{e\varepsilon_1\varepsilon_2 N_{d1}N_{a2}}{2(\varepsilon_1 N_{d1}+\varepsilon_2 N_{a2})}\right)^{1/2}(V_d-V)^{-1/2} \tag{6-36}$$

Cette expression se simplifie dans le cas où l'un des semiconducteurs est beaucoup plus dopé que l'autre. Supposons par exemple $N_{a2} >> N_{d1}$

$$C(V)=\left(\frac{e\varepsilon_1\varepsilon_2 N_{d1}N_{a2}}{2\varepsilon_2 N_{a2}(1+\varepsilon_1 N_{d1}/\varepsilon_2 N_{a2})}\right)^{1/2}(V_d-V)^{-1/2} \tag{6-37}$$

Dans la mesure où les constantes diélectriques sont du même ordre de grandeur pour tous les semiconducteurs, si $N_{a2}>>N_{d1}$ l'expression de $C(V)$ s'écrit

$$C(V)\approx\left(\frac{e\varepsilon_1 N_{d1}}{2}\right)^{1/2}(V_d-V)^{-1/2} \tag{6-38}$$

On obtient une expression semblable à celles obtenues dans le cas de l'homojonction pn ou de la diode Schottky.

6.3. Hétérojonction polarisée

Les différents diagrammes énergétiques représentés sur la figure (6-4) montrent qu'il existe, pour les bandes de conduction et de valence, deux types de discontinuité.

Le premier correspond au cas où la tension de diffusion s'ajoute à la différence d'énergie des bandes considérées. Dans ce cas la variation d'énergie de la bande est monotone, nous appellerons cette discontinuité une *pseudo-continuité*. C'est le cas par exemple de la bande de valence de la figure (6-4-b). Le deuxième type de discontinuité correspond au cas où la tension de diffusion et la discontinuité des gaps jouent des rôles opposés, c'est le cas de la bande de conduction de la figure (6-4-b). Nous qualifierons de *forte discontinuité* ce cas de figure.

En ce qui concerne les courants d'électrons et de trous, il apparaît clairement que chacun d'eux ne peut être important que lorsque la bande mise en jeu présente une *forte discontinuité*. En effet, dans ce cas la tension de diffusion et la hauteur de barrière résultant de la différence des gaps, se compensent partiellement. Considérons par exemple le courant d'électrons. Ce courant ne peut être important, en fonction de la polarisation, que si la bande de conduction de la structure présente une forte discontinuité, c'est le cas de la figure (6-4-b) par exemple. Dans le cas de la figure (6-4-d) au contraire, la bande de conduction est *pseudo-continue* de sorte que dans le sens $SC_2 \rightarrow SC_1$ la barrière de potentiel est très importante, le courant d'électrons dans ce sens restera nécessairement très faible. Dans le sens $SC_1 \rightarrow SC_2$ il n'existe aucune barrière au passage des électrons, mais compte tenu de l'alignement des niveaux de Fermi à polarisation nulle , la distance bande de conduction moins

niveau de Fermi dans le semiconducteur 1 est nécessairement importante. Il en résulte que la densité d'électrons dans le semiconducteur 1 est nécessairement faible. Ainsi le courant d'électrons $SC_1 \rightarrow SC_2$ reste faible, quelle que soit la polarisation.

Par contre au niveau d'une *forte discontinuité* comme c'est le cas pour la bande de conduction de la figure (6-4-b) par exemple, la tension de diffusion réduit la hauteur de barrière et, suivant le signe de la tension de polarisation, le courant d'électrons peut être important. Le même raisonnement est valable au niveau de la bande de valence pour le courant de trous.

Passons en revue les différents cas représentés sur la figure (6-4). Dans le cas (a) les courants d'électrons et de trous peuvent jouer un rôle important suivant le type de chaque semiconducteur. Dans le cas (b) (ou d), seul le courant d'électrons (ou de trous) peut être important. Enfin dans le cas (c) aucun des courants ne peut être important, sauf évidemment si la discontinuité des gaps devient très faible, ce qui ramène la structure à une homojonction.

Ainsi en résumé, les porteurs responsables du courant au niveau d'une hétérojonction sont ceux qui voient une barrière à forte discontinuité. Nous allons calculer successivement le courant d'électrons et le courant de trous.

Comme dans le cas du contact métal-semiconducteur, le courant est conditionné, à l'interface par le mécanisme d'émission thermique des porteurs, et dans chacun des semiconducteurs par la diffusion de ces porteurs. Ces deux processus constituent deux formes différentes d'un même courant et sont par conséquent conditionnés l'un par l'autre. On peut toutefois calculer le courant dans chacun des modèles indépendamment. Si les valeurs ainsi obtenues sont très différentes, c'est le phénomène de plus faible amplitude qui impose sa loi de variation au courant résultant.

6.3.1. *Modèle d'émission thermoélectronique*

Nous avons établi (Par.1.6.4.c) l'expression du courant thermoélectronique existant à l'interface d'une hétérojonction entre deux semiconducteurs. Les expressions (1-255 et 256) s'écrivent en remplaçant E_b par ΔE_{co} qui représente la même quantité

$$j_n = A_e^* T^2 e^{-e\phi'_{F2}/kT} = A_e^* T^2 e^{-(e\phi'_{F1} + \Delta E_{co})/kT} \qquad (6\text{-}39)$$

où $A_e^* = 4\pi e m_{e2} k^2 / h^3$ est la *constante de Richardson* de la structure.

Les différents paramètres apparaissant dans l'expression sont définis dans la figure (1-40).

- l'indice 1 se rapporte au semiconducteur de plus grande affinité électronique

$\chi_1 > \chi_2$

- m_{e2} représente la masse effective des électrons dans le semiconducteur de plus faible affinité électronique, c'est-à-dire dans le semiconducteur ayant la bande de conduction la plus haute.

- $\Delta E_{co} = E_b = -e(\chi_2 - \chi_1)$ représente la discontinuité des bandes de conduction à l'interface. Dans la mesure où la condition $\chi_1 > \chi_2$ est réalisée cette quantité est positive.

- $e\phi'_{F1}$ et $e\phi'_{F2}$ représentent les distances bande de conduction-niveau de Fermi à l'interface, dans chacun des semiconducteurs. Il faut noter que compte tenu de la courbure des bandes résultant de la charge d'espace, ces paramètres sont différents à l'interface et dans les régions neutres des semiconducteurs.

$$\phi'_{F1} \neq \phi_{F1} \qquad \phi'_{F2} \neq \phi_{F2} \tag{6-40}$$

L'expression (6-39), qui donne le courant thermique d'électrons à l'interface, peut être étendue au courant thermique de trous en redéfinissant les paramètres

$$j_p = A_h^* T^2 e^{-e\phi'_{F2}/kT} = A_h^* T^2 e^{-(e\phi'_{F1} - \Delta E_{vo})/kT} \tag{6-41}$$

avec $A_h^* = 4\pi e m_{h2} k^2 / h^3$

- L'indice 1 se rapporte au semiconducteur de plus basse bande de valence, c'est-à-dire en quelque sorte au semiconducteur ayant "l'affinité de trous" la plus grande, $\Delta E_g - e(\chi_2 - \chi_1) > 0$.

- m_{h2} représente la masse effective des trous dans le semiconducteur dont l'affinité de trous est la plus faible.

- $\Delta E_{vo} = \Delta E_g - e(\chi_2 - \chi_1)$ représente la discontinuité des bandes de valence à l'interface. Cette quantité est négative.

- $e\phi'_{F1}$ et $e\phi'_{F2}$ représentent les distances bande de valence-niveau de Fermi à l'interface. Comme pour le courant électronique ces quantités sont différentes à l'interface et dans le volume.

a . Courant d'émission thermique d'électrons

Généralités

Dans la mesure où nous avons affecté l'indice 1 au semiconducteur de plus grande affinité électronique $\Delta E_{co} = -e(\chi_2 - \chi_1) > 0$. Ainsi il existe une forte discontinuité de la bande de conduction uniquement pour $\phi_2 < \phi_1$.

En l'absence de polarisation, le diagramme énergétique est représenté sur la figure (6-7). Les paramètres ϕ'_{F1} et ϕ'_{F2} qui conditionnent le courant, sont donnés par

$$\phi'_{F1} = \phi_{F1} - \phi_{b1} \tag{6-42-a}$$
$$\phi'_{F2} = \phi_{F2} - \phi_{b2} \tag{6-42-b}$$

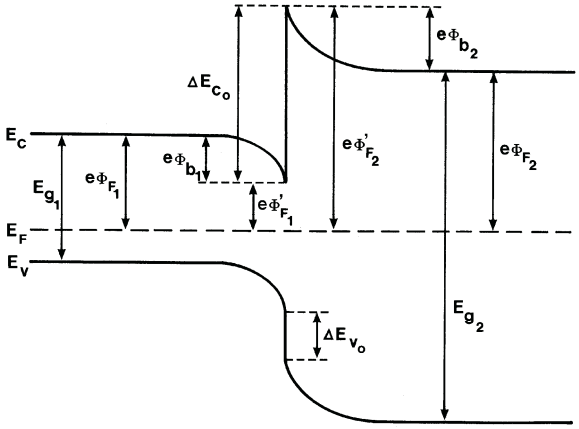


Figure 6-7 : Modèle émission thermoélectronique

Le courant s'écrit

$$j_n = A_e^* T^2 e^{-(e\phi_{F2}+e\phi_{b2})/kT} = A_e^* T^2 e^{-(e\phi_{F1}-e\phi_{b1}+\Delta E_{co})/kT}$$

$$j_n = A_e^* T^2 e^{-e\phi_{F2}/kT}\, e^{-e\phi_{b2}/kT} = A_e^* T^2 e^{-e\phi_{F1}/kT}\, e^{-(\Delta E_{co}-e\phi_{b1})/kT}$$

$$j_n = A_{e1}^* e^{-e\phi_{b2}/kT} = A_{e2}^* e^{-(\Delta E_{co}-e\phi_{b1})/kT} \qquad (6\text{-}43)$$

où les constantes A_{e1}^* et A_{e2}^* sont fonction des dopages de chaque semiconducteur

$$A_{e1}^* = A_e^* T^2 e^{-e\phi_{F2}/kT} \qquad A_{e2}^* = A_e^* T^2 e^{-e\phi_{F1}/kT} \qquad (6\text{-}44)$$

Ces constantes peuvent être explicitées en fonction des densités électroniques dans chaque semiconducteur et de la vitesse thermique, comme dans le cas de la diode Schottky

$$A_e^* T^2 = \frac{4\pi e m_{e2} k^2 T^2}{h^3}$$

$$n_1 = N_{c1} e^{-e\phi_{F1}/kT} \qquad n_2 = N_{c2} e^{-e\phi_{F2}/kT}$$

avec $$N_{c1} = 2\left(\frac{2\pi m_{e1} kT}{h^2}\right)^{3/2} \qquad N_{c2} = 2\left(\frac{2\pi m_{e2} kT}{h^2}\right)^{3/2}$$

soit $$e^{-e\phi_{F1}/kT} = \frac{n_1}{N_{c1}} = \frac{n_1}{2}\left(\frac{h^2}{2\pi m_{e1} kT}\right)^{3/2} \qquad e^{-e\phi_{F2}/kT} = \frac{n_2}{2}\left(\frac{h^2}{2\pi m_{e2} kT}\right)^{3/2}$$

ce qui donne pour les paramètres A^*_{e1} et A^*_{e2}

$$A^*_{e1}=en_2\left(\frac{kT}{2\pi m_{e2}}\right)^{1/2} \qquad A^*_{e2}=\frac{m_{e2}}{m_{e1}}en_1\left(\frac{kT}{2\pi m_{e1}}\right)^{1/2} \tag{6-45}$$

comme dans la diode Schottky on pose

$$v_{e1}=\left(\frac{kT}{2\pi m_{e1}}\right)^{1/2} \qquad v_{e2}=\left(\frac{kT}{2\pi m_{e2}}\right)^{1/2} \tag{6-46}$$

$$v_{e1}=v_{the1}/\sqrt{6\pi}\approx v_{the1}/4 \qquad v_{e2}=v_{the2}/\sqrt{6\pi}\approx v_{the2}/4$$

où v_{the1} et v_{the2} représentent les vitesses thermiques des électrons dans chacun des semiconducteurs. Les constantes A^*_{e1} et A^*_{e2} s'écrivent donc

$$A^*_{e1}=e\,n_2\,v_{e2} \qquad A^*_{e2}=\frac{m_{e2}}{m_{e1}}\,e\,n_1\,v_{e1} \tag{6-47}$$

En l'absence de polarisation le courant d'émission thermoélectronique est le même dans les deux sens, le courant résultant est nul

$$j_n=A^*_{e1}e^{-e\phi_{b2}/kT}-A^*_{e2}e^{-(\Delta E_{co}-e\phi_{b1})/kT}\equiv 0 \tag{6-48}$$

Polarisons le semiconducteur 1 par rapport au semiconducteur 2 par une tension $V=V_{sc1}-V_{sc2}$. Cette tension de polarisation se répartit entre les deux zones de charge d'espace, proportionnellement à la résistance de chacune, $V_1=\alpha V$ et $V_2=(1-\alpha)V$ avec $V=V_1+V_2$. La valeur du paramètre α est fonction de la nature et de la largeur de la zone de charge d'espace, c'est-à-dire du type de chaque semiconducteur. Nous reviendrons sur la valeur de α dans chaque cas particulier.

Les barrières de potentiel $e\phi_{b2}$ et $\Delta E_{co}-e\phi_{b1}$ deviennent respectivement

$$e\phi'_{b2}=e\phi_{b2}-eV_2 \tag{6-49-a}$$

$$\Delta E_{co}-e\phi'_{b1}=\Delta E_{co}-(e\phi_{b1}-eV_1)=\Delta E_{co}-e\phi_{b1}+eV_1 \tag{6-49-b}$$

Les diagrammes énergétiques résultants sont représentés sur la figure (6-8) pour des polarisations V positive et négative.

- Premier cas : $V > 0$

La barrière $SC_1 \rightarrow SC_2$ augmente et la barrière $SC_2 \rightarrow SC_1$ diminue. L'équilibre des émissions thermoélectroniques est rompu, un flux net d'électrons passe du

semiconducteur 2 vers le semiconducteur 1. Il en résulte un courant électrique dans l'autre sens.

- Deuxième cas : $V<0$

La barrière $SC_1 \rightarrow SC_2$ diminue, la barrière $SC_2 \rightarrow SC_1$ augmente, le courant circule dans l'autre sens.

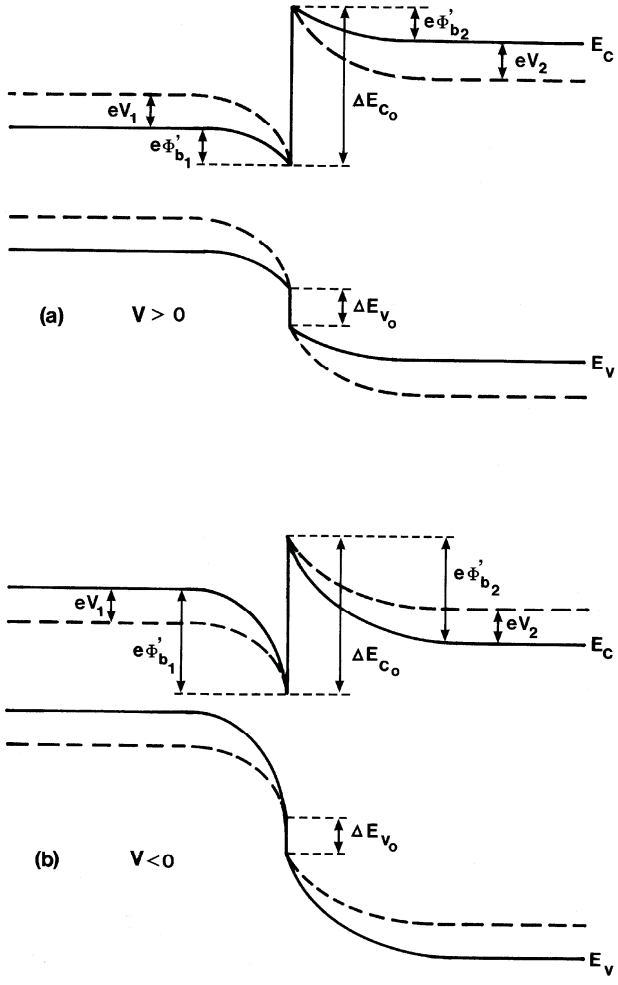


Figure 6-8 : Hétérojonction polarisée

L'amplitude du courant s'écrit

$$j_n = A_{e1}^* e^{-e(\phi_{b2}-V_2)/kT} - A_{e2}^* e^{-(\Delta E_{co}-e(\phi_{b1}-V_1))/kT}$$

soit

$$j_n = A_{e1}^* e^{-e\phi_{b2}/kT}\, e^{eV_2/kT} - A_{e2}^* e^{-(\Delta E_{co}-e\phi_{b1})/kT}\, e^{-eV_1/kT} \tag{6-50}$$

Etant donné la condition d'équilibre en l'absence de polarisation (Eq. 6-48), les coefficients devant les exponentielles de V_1 et V_2 sont égaux, de sorte que le courant s'écrit

$$j_n = A_{e1}^* e^{-e\phi_{b2}/kT}\left(e^{eV_2/kT} - e^{-eV_1/kT}\right) \tag{6-51}$$

ou en explicitant V_1 et V_2 en fonction de V

$$j_n = A_{e1}^* e^{-e\phi_{b2}/kT}\left(e^{(1-\alpha)eV/kT} - e^{-\alpha eV/kT}\right) \tag{6-52}$$

soit

$$j_n = j_{ns} e^{-\alpha eV/kt}\left(e^{eV/kT} - 1\right) \tag{6-53}$$

avec

$$j_{ns} = A_{e1}^* e^{-e\phi_{b2}/kT} = e n_2 v_{e2} e^{-e\phi_{b2}/kT} \tag{6-54}$$

Mais ϕ_{b2} est la fraction de la tension de diffusion développée dans le semiconducteur 2, de sorte que $\phi_{b2}=(1-\alpha)V_d$ d'où l'expression de j_{ns}

$$j_{ns} = e n_2 v_{e2} e^{-(1-\alpha)eV_d/kT} \tag{6-55}$$

La variation du courant avec la tension de polarisation est conditionnée par le paramètre α. Ce paramètre est fonction de la nature de chaque zone de charge d'espace, c'est-à-dire du type de chaque semiconducteur. Considérons les différents cas possibles.

Semiconducteurs isotypes

Notons tout d'abord que si les deux semiconducteurs sont de type p, le courant d'électrons, que nous calculons ici, est négligeable. Le courant est dû à l'émission thermique de trous, nous le calculerons au paragraphe suivant.

Considérons le cas de deux semiconducteurs de type n. La charge d'espace est alors une charge d'accumulation dans le semiconducteur 1 et une charge de déplétion dans le semiconducteur 2. Il en résulte que la zone de charge d'espace du semiconducteur 1 est d'une part très étroite et d'autre part conductrice, alors que celle du semiconducteur 2 est large et isolante. En conséquence la tension de

polarisations s'établit uniquement dans la zone de charge d'espace du semiconducteur 2, soit $\alpha \approx 0$, $V_1=0$, $V_2=V$. Le diagramme énergétique est représenté sur la figure (6-9).

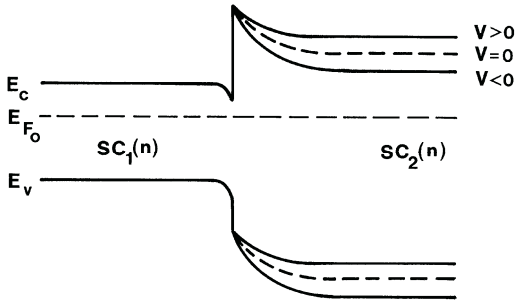


Figure 6-9 : Hétérojonction $SC_1(n)$-$SC_2(n)$ polarisée

Le courant est donné, en faisant $\alpha=0$ et $n_2=N_{d2}$ dans les expressions (6-53 et 55) par

$$j_n = j_{ns}\left(e^{eV/kT} - 1\right) \tag{6-56}$$

avec

$$j_{ns} = e\, N_{d2}\, v_{e2}\, e^{-eV_d/kT} \tag{6-57}$$

- Pour $V>0$, le courant augmente exponentiellement avec la tension, la structure est polarisée dans le sens passant.
- Pour $V<0$, $j_n=-j_{ns}$, le courant est limité au courant de saturation, la structure est polarisée en inverse.

Les deux semiconducteurs étant de type n le courant d'électrons représente le courant total. D'où la caractéristique $I(V)$ (Fig. 6-10).

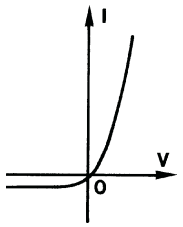


Figure 6-10 : Courbe courant-tension

Semiconducteurs de types différents

- Premier cas : Semiconducteur 1 (p) - Semiconducteur 2 (n)

La charge d'espace est une charge de déplétion dans chacun des semiconducteurs. La tension V s'établit dans chacune des zones de charge d'espace proportionnellement à leur résistance, c'est-à-dire proportionnellement à leur largeur. La tension de polarisation s'établit dans le même rapport que la tension de

diffusion, donné par l'équation (6-32). Compte tenu du type de chaque semiconducteur ce rapport s'écrit

$$V_1/V_2 = \varepsilon_2 N_{d2}/\varepsilon_1 N_{a1} \tag{6-58}$$

Ainsi le paramètre α est donné par

$$\alpha = \varepsilon_2 N_{d2}/(\varepsilon_1 N_{a1} + \varepsilon_2 N_{d2}) \tag{6-59}$$

La caractéristique courant-tension est alors fonction des rapports de dopage entre les deux semiconducteurs. Considérons les deux cas limites d'un semiconducteur beaucoup plus dopé que l'autre.

$N_{a1} << N_{d2}$

La charge d'espace se développe essentiellement dans le semiconducteur 1. Il en est de même de la tension de la diffusion et de la tension de polarisation, et par suite $\alpha \approx 1$ (Eq. 6-59). L'expression du courant devient

$$j_n = j_{ns} e^{-eV/kT}\left(e^{eV/kT} - 1\right)$$

soit

$$j_n = -j_{ns}\left(e^{-eV/kT} - 1\right) \tag{6-60}$$

avec

$$j_{ns} = e N_{d2} v_{e2} e^{-eV_d/kT} \tag{6-61}$$

La caractéristique $I(V)$ est la même que dans le cas des deux semiconducteurs de type n en changeant les signes de J et V. C'est la tension négative qui correspond au sens passant (Fig. 6-11).

$N_{a1} >> N_{d2}$

La charge d'espace se développe essentiellement dans le semiconducteur 2 ainsi que la tension de polarisation

$$\alpha \approx \varepsilon_2 N_{d2}/\varepsilon_1 N_{a1} << 1 \qquad 1 - \alpha \approx 1$$

Le courant d'électrons s'écrit

$$j_n \approx j_{ns}\left(e^{eV/kT} - e^{-\alpha eV/kT}\right) \tag{6-62}$$

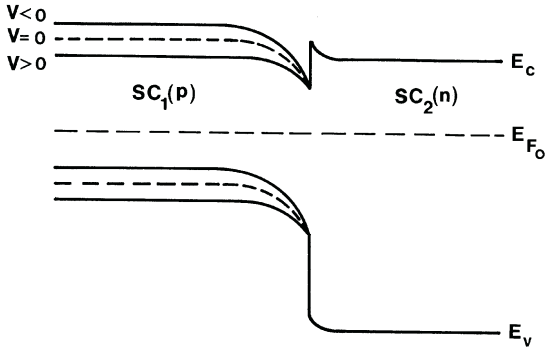


Figure 6-11 : Semiconducteurs de types différents

Pour une polarisation positive, le deuxième terme devient négligeable et le courant augmente exponentiellement avec la tension. Pour une polarisation négative, le premier terme devient négligeable. Le deuxième terme présente une variation exponentielle mais avec un exposant pondéré par le coefficient $\alpha << 1$. L'allure de la caractéristique est représentée sur la figure (6-12). La partie négative de la caractéristique est semblable à la partie positive en changeant l'échelle des tensions dans un rapport $1/\alpha$.

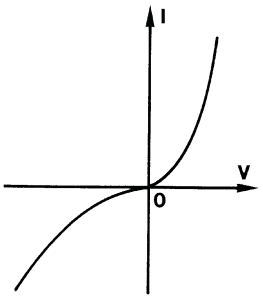


Figure 6-12 : Courbe courant-tension

- Deuxième cas : Semiconducteur 1(n) - Semiconducteur 2(p)

Dans ce cas la charge d'espace est une charge d'accumulation dans chacun des semiconducteurs. Il n'y a par conséquent pas de zone vide de porteurs de sorte que toute polarisation s'établit dans les régions neutres des semiconducteurs. Le contact entre les deux semiconducteurs est ohmique.

b. Courant d'émission thermique de trous

Généralités

Il n'existe une forte discontinuité de la bande de valence que dans le cas $\phi_2 > \phi_1$

En l'absence de polarisation le diagramme énergétique est représenté sur la figure (6-13). Ce cas de figure correspond à la structure Ge-GaAs . Les paramètres ϕ'_{F1} et ϕ'_{F2} sont donnés par

$$\phi'_{F1}=\phi_{F1}-\phi_{b1} \tag{6-63-a}$$

$$\phi'_{F2}=\phi_{F2}-\phi_{b2} \tag{6-63-b}$$

Le courant thermique de trous s'écrit

$$j_p=A_h^* T^2 e^{-(e\phi_{F2}+e\phi_{b2})/kT}=A_h^* T^2 e^{-(e\phi_{F1}-e\phi_{b1}-\Delta E_{vo})/kT}$$

soit

$$j_p=A_{h1}^* e^{-e\phi_{b2}/kT}=A_{h_2}^* e^{-(-\Delta E_{vo}-e\phi_{b1})/kT} \tag{6-64}$$

avec

$$A_{h1}^*=A_h^* T^2 e^{-e\phi_{F2}/kT} \qquad A_{h2}^*=A_h^* T^2 e^{-e\phi_{F1}/kT} \tag{6-65}$$

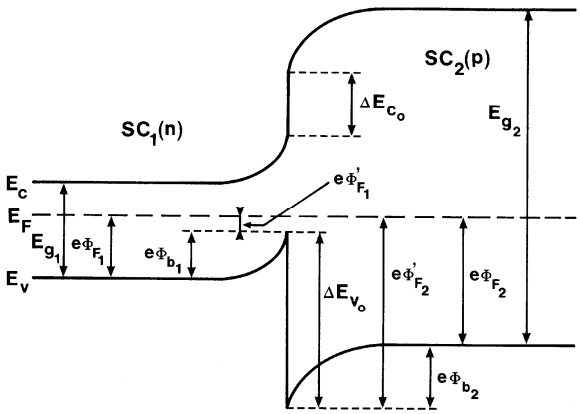


Figure 6.13 : Hétérojonction Ge-GaAs

Comme pour le courant électronique, les constantes A_{h1}^* et A_{h2}^* peuvent être explicitées en fonction des densités de trous dans chaque semiconducteur et de la vitesse thermique. Le même calcul que celui développé dans le cas des électrons donne

$$A_{h1}^*=e\,p_2\,v_{h2} \qquad A_{h2}^*=\frac{m_{h2}}{m_{h1}}e\,p_1\,v_{h1} \tag{6-66}$$

avec

$$v_{h2}=\left(\frac{kT}{2\pi m_{h2}}\right)^{1/2} \qquad v_{h1}=\left(\frac{kT}{2\pi m_{h1}}\right)^{1/2}$$

$$v_{h1} = v_{thh1} / \sqrt{6\pi} \approx v_{thh1} / 4 \qquad v_{h2} = v_{thh2} / \sqrt{6\pi} \approx v_{thh2} / 4$$

où v_{thh1} et v_{thh2} représentent les vitesses thermiques des trous dans chaque semiconducteur.

En l'absence de polarisation, le courant d'émission thermique de trous est le même dans les deux sens, le courant résultant est nul

$$j_p = -A^*_{h1} e^{-e\phi_{b2}/kT} + A^*_{h2} e^{-(-\Delta E_{vo} - e\phi_{b1})/kT} \equiv 0 \tag{6-67}$$

Si on polarise le semiconducteur 1 par rapport au semiconducteur 2 par une tension $V=V_{sc1}-V_{sc2}$ la tension V se répartit entre les deux zones de charge d'espace proportionnellement à leurs résistances

$$V_1 = \alpha V \quad V_2 = (1-\alpha)V$$

Les barrières de potentiel $e\phi_{b2}$ et $-\Delta E_{vo}-e\phi_{b1}$ deviennent respectivement

$$e\phi'_{b2} = e\phi_{b2} + eV_2 \tag{6-68-a}$$

$$-\Delta E_{vo} - e\phi'_{b1} = -\Delta E_{vo} - e(\phi_{b1} + V_1) \tag{6-68-b}$$

Les diagrammes énergétiques sont représentés sur la figure (6-14) pour des polarisations V positive et négative.

Pour $V>0$ la barrière que doivent franchir les trous diminue dans le sens $SC_1 \rightarrow SC_2$ et augmente dans le sens $SC_2 \rightarrow SC_1$. L'équilibre est rompu, un courant circule dans le sens $SC_1 \rightarrow SC_2$. Le phénomène est inversé pour V négatif.

L'amplitude du courant de trous s'écrit

$$j_p = -A^*_{h1} e^{-e(\phi_{b2}+eV_2)/kT} + A^*_{h2} e^{-(-\Delta E_{vo} - e(\phi_{b1}+V_1))/kT}$$

ou

$$j_p = -A^*_{h1} e^{-e\phi_{b2}/kT} e^{-eV_2/kT} + A^*_{h2} e^{-(-\Delta E_{vo} - e\phi_{b1})/kT}\ e^{eV_1/kT} \tag{6-69}$$

Compte tenu de la relation d'équilibre (6-67), le courant s'écrit

$$j_p = A^*_{h1} e^{-e\phi_{b2}/kT} \left(e^{eV_1/kT} - e^{-eV_2/kT} \right) \tag{6-70}$$

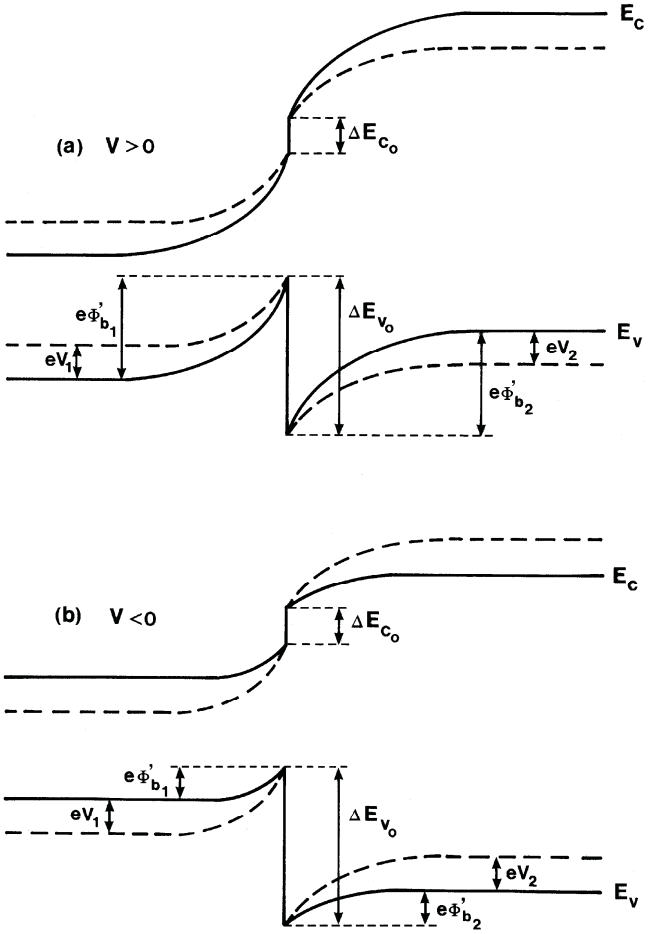


Figure 6-14 : Hétérojonction Ge-GaAs polarisée

En explicitant V_1 et V_2 en fonction de la tension de polarisation V on obtient

$$j_p = -j_{ps} e^{\alpha e V / kT} \left(e^{-eV/kT} - 1 \right) \tag{6-71}$$

avec

$$j_{ps} = A^*_{h1} e^{-e\phi_{b2}/kT} = e\, p_2\, v_{h2}\, e^{-e\phi_{b2}/kT}$$

ou

$$j_{ps} = e p_2 v_{h2} e^{-(1-\alpha) e V_d / kT} \tag{6-72}$$

La variation du courant de trous avec la tension de polarisation est conditionnée par la valeur du paramètre α, c'est-à-dire par la nature des zones de charge d'espace de chaque semiconducteur, et par conséquent du type de chaque semiconducteur.

Semiconducteurs isotypes

A l'inverse du courant d'électrons, le courant de trous ne peut être important que si les deux semiconducteurs sont de type p. La charge d'espace est alors une charge d'accumulation dans le semiconducteur 1 et une charge de déplétion dans le semiconducteur 2. En conséquence la tension de polarisation s'établit uniquement dans la zone de charge d'espace du semiconducteur 2, soit $\alpha = 0$, $V_1 = 0$ et $V_2 = V$.

Le diagramme énergétique est représenté sur la figure (6-15), le courant est donné par

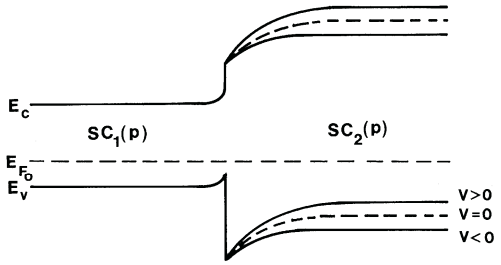
$$j_p = -j_{ps}\left(e^{-eV/kT} - 1\right) \tag{6-73}$$

avec

$$j_{ps} = e\,N_{a2}\,v_{h2}\,e^{-e\phi_{b2}/kT} \tag{6-74}$$

Pour $V > 0$, $j_p = j_{ps}$ le courant est limité au courant de saturation, la structure est polarisée en inverse. Pour $V < 0$, $j_p = -j_{ps}e^{-eV/kT}$, le courant augmente exponentiellement avec la tension, la structure est polarisée dans le sens passant.

Dans la mesure où les deux semiconducteurs sont de type p seul le courant de trous joue un rôle important. De sorte que $J = j_p$ la caractéristique $I(V)$ est représentée sur la figure (6-16).



Figures 6-15 :
Diagramme énergétique

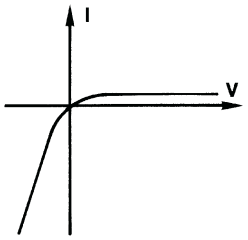


Figure 6-16 :
Caractéristiques

Semiconducteurs de types différents

- Premier cas : Semiconducteur 1(n) - Semiconducteur 2 (p)

La charge d'espace est une charge de déplétion dans chacun des semiconducteurs. La tension de polarisation s'établit partiellement dans chacune des zones de charge d'espace, dans le rapport

$$V_1/V_2 = \varepsilon_2 \, N_{a2} \big/ \varepsilon_1 \;\; N_{d1} \tag{6-75}$$

Le paramètre α est donné par

$$\alpha = \varepsilon_2 \, N_{a2} / (\varepsilon_1 \, N_{d1} + \varepsilon_2 \, N_{a2}) \tag{6-76}$$

Comme pour le courant d'électrons, le courant de trous est fonction des dopages respectifs de chaque semiconducteur.

$$N_{d1} << N_{a2}$$

La charge d'espace se développe dans le semiconducteur 1 de même que la tension de polarisation, $\alpha \approx 1$ (Fig. 6-17).

$$j_p \approx -j_{ps} e^{eV/kT} \left(e^{-eV/kT} - 1 \right)$$

soit

$$j_p \approx j_{ps} \left(e^{eV/kT} - 1 \right) \tag{6-77}$$

La caractéristique $I(V)$ est la même que pour deux semiconducteurs de type p en changeant les signes de I et de V. Le sens passant correspond à une tension positive.

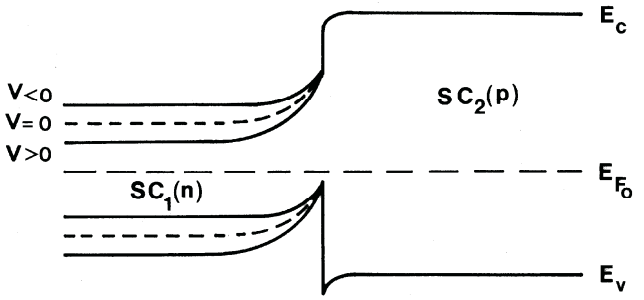


Figure 6-17 : Semiconducteurs de types différents

$N_{d1} >> N_{a2}$

La charge d'espace se développe essentiellement dans le semiconducteur 2 ainsi que la tension de polarisation

$$\alpha \approx \varepsilon_2 N_{a2} / \varepsilon_1 N_{d1} \quad 1 - \alpha \approx 1$$

Le courant de trous s'écrit

$$j_p \approx j_{ps} \left(e^{eV/kT} - e^{-\alpha eV/kT} \right) \tag{6-78}$$

On obtient une caractéristique semblable à celle obtenue pour le courant électronique dans le paragraphe précédent. Le courant est la différence de deux exponentielles dont la première varie beaucoup plus vite ($\alpha << 1$) que l'autre. Le sens passant correspond à $V > 0$ mais le courant inverse tout en restant beaucoup plus faible que le courant direct augmente exponentiellement avec la tension inverse (Fig. 6-12).

- Deuxième cas : Semiconducteur 1 (p) - Semiconducteur 2 (n)

La charge d'espace est d'accumulation dans chacun des semiconducteurs, toute polarisation s'établit dans les régions neutres des semiconducteurs, le contact est ohmique.

6.3.2. *Modèle de diffusion*

Généralités

Le calcul du courant dans le modèle de la diffusion est fonction des types respectifs de chacun des semiconducteurs. Si l'hétérostructure est réalisée avec deux semiconducteurs isotypes, l'un d'eux est nécessairement en régime d'accumulation et l'autre en régime de déplétion. La tension de polarisation s'établit alors essentiellement dans la zone de déplétion, c'est dans cette zone que l'on calculera la relation $I(V)$. Mais les deux semiconducteurs étant de même type, les porteurs qui diffusent dans la zone de charge d'espace sont des porteurs majoritaires. On peut donc négliger le courant de porteurs minoritaires de sorte que le courant de porteurs majoritaires constitue le courant total. Il est donc constant et indépendant de x. C'est l'hypothèse que nous avons utilisée pour calculer le courant de diffusion dans le contact métal-semiconducteur. Dans cette hypothèse on développe pour l'hétérojonction le même calcul que pour le contact métal-semiconducteur.

Si au contraire l'hétérojonction est de type pn, les porteurs qui sont émis thermiquement à l'interface à partir d'un semiconducteur, deviennent dans l'autre

des porteurs minoritaires. Il en résulte qu'ils ne constituent plus la totalité du courant et que par voie de conséquence on ne peut plus écrire que le courant qu'ils transportent est indépendant de *x.* Le calcul développé dans le cas du contact métal-semiconducteur ne s'applique plus. Il faut alors développer un calcul analogue à celui développé pour l'homojonction pn.

Semiconducteurs isotypes

Considérons l'hétérojonction nn représentée sur la figure (6-18). Le semiconducteur 1 est en régime d'accumulation, le semiconducteur 2 est en régime de déplétion.

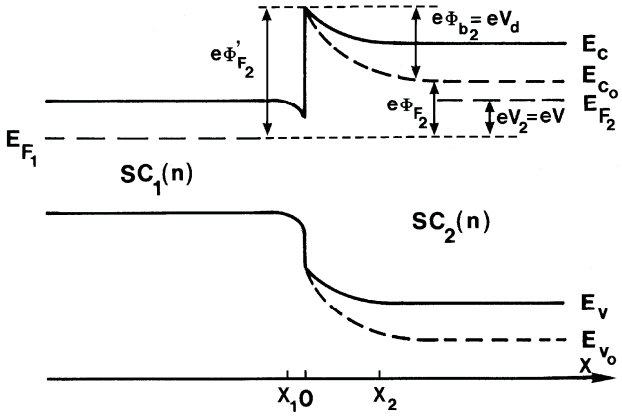


Figure 6-18 : Semiconducteurs isotypes

En négligeant le courant de porteurs minoritaires on peut écrire que $j_n(x)=J$=Cte. On calcule alors ce courant de la même manière que pour le contact métal-semiconducteur, en intégrant l'équation de diffusion des électrons dans la zone de charge d'espace du semiconducteur 2 avec j_n=Cte

$$J=j_n=e\left(n(x)\mu_2 E(x)+D_2\frac{dn(x)}{dx}\right) \tag{6-79}$$

où μ_2 et D_2 représentent la mobilité et la constante de diffusion des électrons dans le semiconducteur 2. Compte tenu de la relation d'Einstein et du fait que $E(x)=-dV(x)/dx$, l'équation s'écrit

$$J=eD_2\left(-\frac{e}{kT}n(x)\frac{dV(x)}{dx}+\frac{dn(x)}{dx}\right) \tag{6-80}$$

ou, en multipliant à gauche et à droite par $e^{-eV(x)/kT}$

$$J e^{-eV(x)/kT}\,dx = eD_2\,d\left(n(x)e^{-eV(x)/kT}\right) \tag{6-81}$$

En intégrant l'équation entre 0 et x_2

$$J\int_0^{x_2} e^{-eV(x)/kT}\,dx = eD_2\left(n(x)e^{-eV(x)/kT}\right)_0^{x_2} \tag{6-82}$$

Les conditions aux limites sont les suivantes

- Condition en x=0

$$n(x=0) = N_{c2}e^{-(E_{c2}(x=0)-E_F)/kT} = N_{c2}e^{-e\phi'_{F2}/kT} \tag{6-83}$$

or $\phi'_{F2} = \phi_{F2} + \phi_{b2}$ de sorte que

$$n(x=0) = N_{c2}e^{-e(\phi_{F2}+\phi_{b2})/kT} \tag{6-84}$$

Mais $N_{c2}\,e^{-e\phi_{F2}/kT} = n_2 = N_{d2}$, si on appelle $N_{d2} = (N_d - N_a)_2$ l'excédent de donneurs dans le semiconducteur 2. D'autre part ϕ_{b2} est la fraction de la tension de diffusion développée dans le semiconducteur 2. Dans la mesure où le semiconducteur 1 est en régime d'accumulation et le semiconducteur 2 en régime de déplétion, la totalité de la tension de diffusion se développe dans le semiconducteur 2, en d'autres termes, pour reprendre l'écriture précédente $\phi_{b1} = \alpha V_d$ et $\phi_{b2} = (1-\alpha)V_d$ avec ici $\alpha \approx 0$ soit $\phi_{b2} \approx V_d$. Ainsi la densité d'électrons en x=0 s'écrit

$$n(x=0) = N_{d2}e^{-eV_d/kT} \tag{6-85}$$

En ce qui concerne le potentiel, on prend l'origine en x=0 de sorte que $V(x=0)=0$.

- Condition en $x=x_2$

Le nombre de porteurs est donné par

$$n(x=x_2) = n_2 = N_{d2} \tag{6-86}$$

En l'absence de polarisation, et compte tenu de l'origine des potentiels, le potentiel en $x=x_2$ est donné par $V(x=x_2)=V_d$. En présence d'une polarisation V du semiconducteur 1 par rapport au semiconducteur 2, dans la mesure où α=0, cette tension de polarisation s'établit en totalité dans la zone de charge d'espace du semiconducteur 2, de sorte que le potentiel en $x=x_2$ s'écrit

$$V(x=x_2) = V_d - V \tag{6-87}$$

En explicitant ces conditions aux limites dans l'expression (6-82), on obtient

$$J\int_0^{x_2} e^{-eV(x)/kT}\,dx = eD_2\left(N_{d2}e^{-e(V_d-V)/kT} - N_{d2}e^{-eV_d/kT}\right)$$

soit

$$J = \frac{eD_2 N_{d2}}{\int_0^{x_2} e^{-eV(x)/kT}\,dx} e^{-eV_d/kT}\left(e^{eV/kT}-1\right) \tag{6-88}$$

Pour calculer l'intégrale apparaissant au dénominateur, il faut expliciter l'expression de $V(x)$ dans la zone de charge d'espace du semiconducteur 2. Cette expression est analogue à (6-18) et s'écrit ici

$$V(x) = -\frac{eN_{d2}}{2\varepsilon_2}(x-x_2)^2 + V_d - V \tag{6-89}$$

Cette expression est semblable à celle obtenue dans le cas du contact métal-semiconducteur (4-28) en remplaçant N_d par N_{d2} , ε_s par ε_2 et W par x_2. L'intégrale a donc la même expression que (4-29), soit

$$I = \frac{\varepsilon_2 kT}{e^2 N_{d2} x_2} \tag{6-90}$$

Ainsi le courant est donné par

$$J = J_s\left(e^{eV/kT}-1\right) \tag{6-91}$$

avec

$$J_s = e N_{d2} \frac{e D_2}{kT} \frac{e N_{d2} x_2}{\varepsilon_2} e^{-eV_d/kT} \tag{6-92}$$

Mais $eD_2 / kT = \mu_2$ la mobilité des électrons dans le semiconducteur 2, et $eN_{d2} x_2 / \varepsilon_2 = E_{s2}$ le champ électrique à l'interface dans le semiconducteur 2. Il en résulte que

$$J_s = e N_{d2} v_{d2} e^{-eV_d/kT} \tag{6-93}$$

où $v_{d2} = \mu_2 E_{s2}$ représente la vitesse des porteurs à l'interface. On obtient donc une expression du courant tout à fait semblable à celle obtenue dans le modèle de l'émission thermique (6-57) en remplaçant v_{e2} par v_{d2}.

Semiconducteurs de types différents

Considérons l'hétérojonction pn représentée sur la figure (6-19). Les deux semiconducteurs sont en régime de déplétion.

En raison du fait que la forte discontinuité apparaît sur la bande de conduction et que la bande de valence est pseudo-continue, la hauteur de barrière au niveau de la bande de valence réduit à zéro la diffusion des trous du semiconducteur 1 vers le semiconducteur 2. Seuls les électrons peuvent diffuser du semiconducteur 2 vers le semiconducteur 1. Il en résulte que, dans l'hypothèse où l'on néglige les porteurs minoritaires d'équilibre, le courant dans le semiconducteur 2 est exclusivement un courant d'électrons, le courant dans le semiconducteur 1 résulte d'un courant d'électrons et d'un courant de trous. Ainsi la conservation du courant s'écrit : dans le semiconducteur 2, $j_n(x)=J=$Cte; dans le semiconducteur 1, $j_n(x)+j_p(x)=J=$Cte.

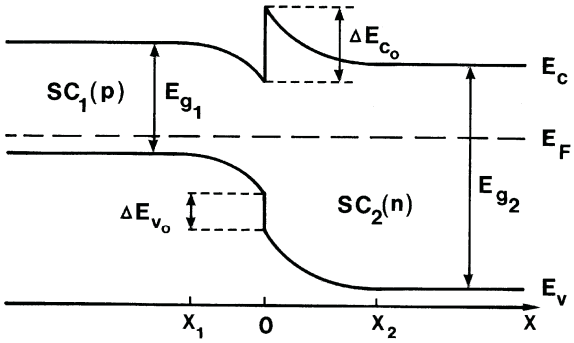


Figure 6-19 : Semiconducteurs de types différents

Dans la mesure où le courant est limité par la diffusion d'électrons du semiconducteur 2 vers le semiconducteur 1, ce courant est donné par $j_n(x_2)$. Si on néglige la recombinaison des électrons dans la zone de charge d'espace on peut écrire $j_n(x_2)=j_n(x_1)$. Ainsi le courant est donné par

$$J=j_n(x_1)$$

Ce courant se calcule de la même manière que le courant $j_n(x_p)$ dans une homojonction pn. Supposons pour simplifier les calculs $N_{d2}>>N_{a1}$. Ceci entraîne que la zone de charge d'espace s'établit essentiellement dans le semiconducteur 1 de sorte que la tension appliquée se retrouve en totalité dans cette zone, en d'autres termes $\alpha = 1$.

Revenons sur le résultat établi pour l'homojonction pn, nous avons obtenu l'expression (2-50-b) pour le courant d'électrons dans la région de type p beaucoup plus longue que la longueur de diffusion des porteurs

$$j_n(x)=\frac{en_i^2 D_n}{N_a L_n}\left(e^{eV/kT}-1\right)e^{(x-x_p)/L_n} \tag{6-94}$$

ce qui donne en $x=x_p$

$$j_n(x_p)=e\frac{n_i^2}{N_a}\frac{D_n}{L_n}\left(e^{eV/kT}-1\right) \tag{6-95}$$

où n_i^2/N_a représente la densité d'électrons en $x=x_p$ en l'absence de polarisation soit $n(x_p)$. Ainsi

$$j_n(x_p)=en(x_p)\frac{D_n}{L_n}\left(e^{eV/kT}-1\right) \tag{6-96}$$

D'autre part la tension de diffusion de la jonction est donnée par (Eq. 2-37)

$$V_d=\frac{kT}{e}\,Ln\frac{n(x_n)}{n(x_p)}$$

Ainsi

$$n(x_p)=n(x_n)\,e^{-eV_d/kT}=N_d\,e^{-eV_d/kT}$$

Il en résulte que le courant d'électrons en $x=x_p$ s'écrit

$$j_n(x_p)=eN_d\frac{D_n}{L_n}e^{-eV_d/kT}\left(e^{eV/kT}-1\right) \tag{6-97}$$

Par analogie avec cette expression, le courant d'électrons en $x=x_1$ dans l'hétérojonction s'écrit

$$J=j_n(x_1)=eN_{d2}\frac{D_{n2}}{L_{n2}}e^{-eV_d/kT}\left(e^{eV/kT}-1\right) \tag{6-98}$$

ou en posant $v_{d2}=D_{n2}/L_{n2}=\sqrt{D_{n2}/\tau_{n2}}$

$$J=J_s\left(e^{eV/kT}-1\right) \tag{6-99}$$

avec

$$J_s=e\,N_{d2}\,v_{d2}\,e^{-eV_d/kT} \tag{6-100}$$

on obtient une expression semblable à celle que nous avons établie dans le cas de l'hétérojonction de semiconducteurs isotypes, avec une expression différente de la vitesse de diffusion. Cette expression se généralise à une valeur quelconque du rapport N_{d2}/N_{a1}. Dans ce cas α est donné par l'expression (6-59) et le courant s'écrit

$$J=J_s\left(e^{(1-\alpha)eV/kT}-e^{-\alpha eV/kT}\right) \tag{6-101}$$

En résumé, on obtient la même expression du courant dans le modèle de l'émission thermique des porteurs et dans le modèle de la diffusion, avec des valeurs différentes des vitesses des porteurs. Comme dans le cas du contact métal-semiconducteur la combinaison des deux modèles se traduit par l'existence d'une vitesse équivalente, ce qui donne pour l'expression générale du courant d'électrons

$$j_n=j_s\left(e^{(1-\alpha)eV/kT}-e^{-\alpha eV/kT}\right) \tag{6-102}$$

où

$$j_s=envv^*e^{-eV_d/kT} \tag{6-103}$$

avec

$$v^*=\frac{v_{d2}\,v_{e2}}{v_{d2}+v_{e2}} \tag{6-104}$$

α représente le rapport des tensions aux frontières des zones de charge d'espace de chacun des semiconducteurs.

6.3.3. *Courant tunnel -courant de recombinaison*

Dans les processus précédents, conditionnés par les mécanismes de diffusion dans la zone de charge d'espace ou d'émission à l'interface, le courant résulte du franchissement de la barrière de potentiel par les porteurs ayant une énergie thermique suffisante. Ce courant est schématisé sur la figure (6-20). Il est toujours beaucoup plus important au niveau d'une *forte discontinuité* qu'au niveau d'une *pseudo-continuité*.

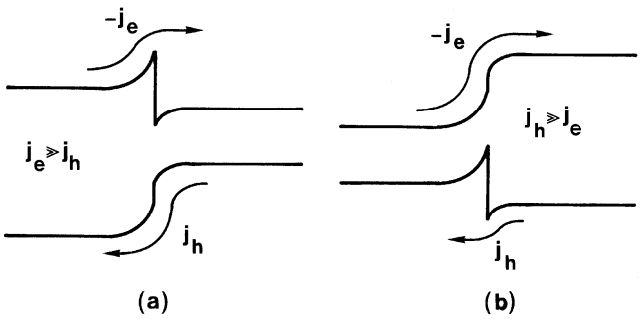


Figure 6-20 : Courants d'émission par dessus les barrières

Au niveau d'une *forte discontinuité* la barrière de potentiel est relativement étroite en particulier au voisinage du sommet, de sorte que certains porteurs peuvent passer d'un semiconducteur dans l'autre par effet tunnel. Ce processus est un processus que l'on peut qualifier d'intrabande dans la mesure où un seul type de bande est concerné. Il est représenté sur la figure (6-21).

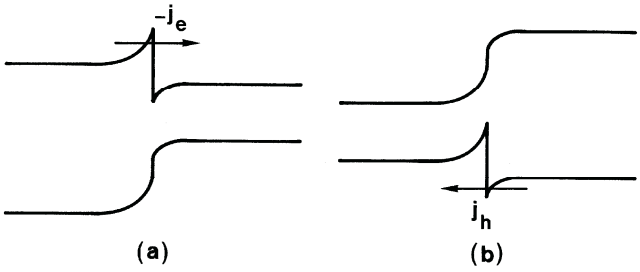


Figure 6-21 : Courant tunnel intrabande

En raison de la discontinuité de la bande considérée et de la variation quadratique du potentiel dans la zone de charge d'espace (6-19 et 22), la barrière est dissymétrique. La barrière s'élargit de manière importante quand on s'éloigne du sommet de sorte que l'effet tunnel ne joue un rôle important que pour les électrons ayant une énergie thermique voisine du sommet. Un calcul numérique développé dans l'approximation BKW montre que la fraction de barrière autorisant un effet tunnel reste limitée à environ 0,1 eV du sommet. Ce courant tunnel est donc assisté thermiquement, il obéit de ce fait à la même loi de température que le courant d'émission.

En polarisation inverse le courant tunnel joue un rôle plus important et se manifeste, comme dans une homojonction, entre la bande de valence d'un semiconducteur et la bande de conduction de l'autre. Il s'agit d'un processus interbande, il est représenté sur la figure (6-22).

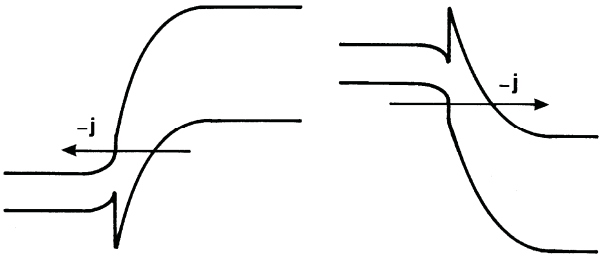


Figure 6.22 : Courant tunnel interbande

Un autre type de processus peut jouer un rôle important sur le courant, c'est comme dans l'homojonction pn, le processus de recombinaison dans la zone de charge d'espace. Dans l'hétérojonction, la présence d'états d'interface augmente la vitesse de recombinaison à l'interface et localise pratiquement à ce niveau l'essentiel des recombinaisons. Ces états d'interface jouent un rôle sur la recombinaison des porteurs mis en jeu dans le courant d'émission, et créent un courant de recombinaison, localisé à l'interface. Ce processus est représenté sur la figure (6-23).

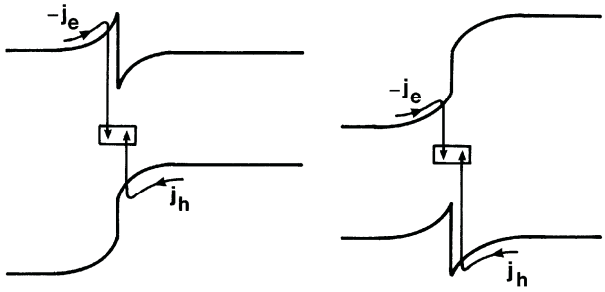


Figure 6-23 : Courant de recombinaison

Enfin les états d'interface et l'effet tunnel peuvent se combiner pour donner naissance à un processus plus complexe. Il se produit un effet tunnel entre une bande et un état d'interface suivi d'une recombinaison entre l'autre type de bande et cet état d'interface. Le processus est représenté sur la figure (6-24).

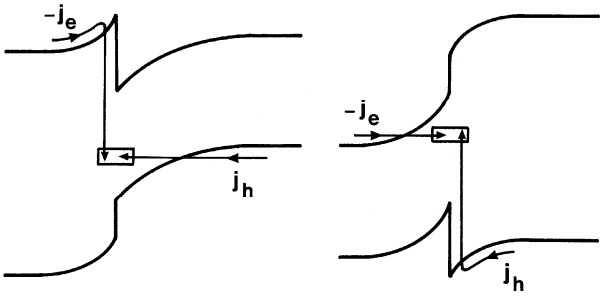


Figure 6-24 : Courant tunnel- recombinaison

6.4. Transistor à hétérojonction - HBT

Dans le domaine des hyperfréquences, le paramètre qui limite les performances des transistors à effet de champ, qui sont des structures planaires, est la longueur de grille. Cette longueur est limitée par la résolution des techniques de photolithographie, qui permettent difficilement de réduire la distance drain-source en deçà de 22 nm. La structure du transistor HBT (Heterojonction Bipolar Transistor) est verticale, comme celle du transistor bipolaire, mais réalisée à partir d'hétérojonctions. Dans ce type de structure, les techniques modernes d'épitaxie que sont la MOCVD (Metal Organic Chemical Vapor Deposition) et la MBE (Molecular Beam Epitaxy), permettent de réaliser des distances émetteur-collecteur nettement inférieures à 0,1 µm. Il en résulte des performances en fréquence très élevées. Notons aussi que rien n'interdit la réalisation de surfaces d'émetteur et de collecteur relativement importantes, ce qui autorise des courants et par suite des puissances plus élevées que dans les transistors à effet de champ.

6.4.1. *Principe de fonctionnement*

Considérons le transistor *npn* schématisé sur la figure (6-25). Les courants résultent des doubles injections d'électrons et de trous dans chacune des jonctions. Si on néglige les recombinaisons dans les zones de charge d'espace, les courants d'émetteur et de collecteur s'écrivent, compte tenu des conventions de signe adoptées sur la figure,

$$j_e = -j_n(x_e) - j_p(x_e') \tag{6-105-a}$$

$$j_c = j_n(x_c) + j_p(x_c') \tag{6-105-b}$$

En régime de fonctionnement normal, les jonctions émetteur-base et collecteur-base sont respectivement polarisées en direct et en inverse. Les composantes du courant collecteur j_c sont alors dues aux transferts des porteurs minoritaires à travers la jonction collecteur-base. Il en résulte que la composante $j_p(x_c')$, qui est proportionnelle à la densité de trous présents dans le collecteur, est négligeable. Seule la composante $j_n(x_c)$, qui résulte du transfert d'électrons de la base vers le collecteur est importante en raison de la présence dans la base des électrons injectés depuis l'émetteur (c'est l'effet transistor). Le courant collecteur peut donc être limité à sa seule composante électronique $j_n(x_c)$.
En raison de recombinaisons dans la base, tous les électrons injectés en x_e n'atteignent pas x_c de sorte que si on appelle α_n le *coefficient de transfert* des électrons à travers la base, la composante $j_n(x_c)$ s'écrit

$$j_n(x_c) = \alpha_n j_n(x_e) \tag{6-106}$$

En l'absence de recombinaisons le coefficient de transfert α_n est égal à 1, en présence de recombinaisons ce coefficient, inférieur à 1, est donné par

$$\alpha_n = \frac{j_n(x_c)}{j_n(x_e)} = \frac{eD_n(d\Delta n/dx)_{xc}}{eD_n(d\Delta n/dx)_{xe}} = \frac{(d\Delta n/dx)_{xc}}{(d\Delta n/dx)_{xe}} \qquad (6\text{-}107)$$

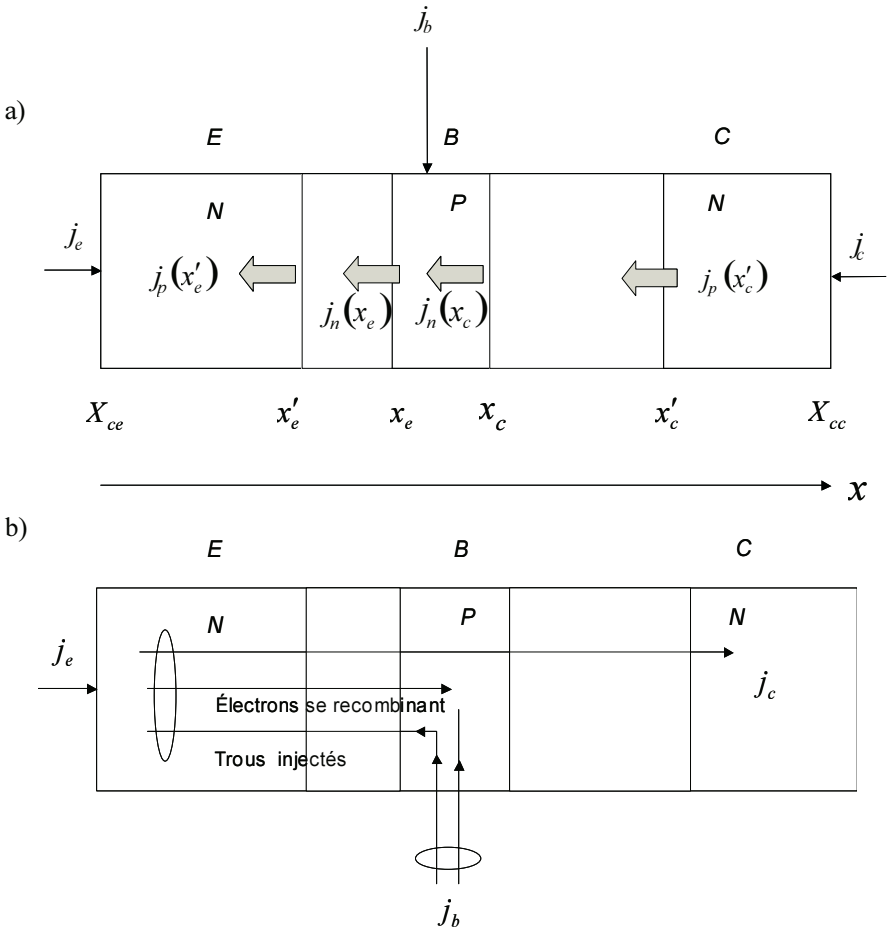


Figure 6-25 a et b : Transistor à hétérojonction

La distribution des électrons dans la base, obtenue par intégration de l'équation de continuité, est de la forme

$$\Delta n = A e^{-x/L_n} + B e^{x/L_n} \qquad (6\text{-}108)$$

Les constantes d'intégration A et B sont déterminées par les conditions aux limites de la base. En $x = x_e$, $\Delta n = \Delta n(x_e)$, en $x = x_c$, $\Delta n = 0$ en raison de la présence du champ électrique présent dans la zone de charge d'espace de la jonction collecteur-base qui évacue instantanément les porteurs excédentaires. La distribution d'électrons s'écrit alors

$$\Delta n = \frac{\Delta n(x_e)}{sh(W_{be}/L_n)} sh\left(\frac{x_c - x}{L_n}\right) \tag{6-109}$$

où $W_{be} = x_c - x_e$ est la longueur effective de la base. Ainsi

$$\frac{d\Delta n}{dx} = -\frac{\Delta n(x_e)}{L_n sh(W_{be}/L_n)} ch\left(\frac{x_c - x}{L_n}\right) \tag{6-110}$$

Les relations (6-107 et 110) permettent d'établir l'expression du coefficient α_n.

$$\alpha_n = 1/ch(W_{be}/L_n) \tag{6-111}$$

Les différents courants (Fig. 6-25) s'écrivent donc

$$j_e = -j_n(x_e) - j_p(x_e') \tag{6-112-a}$$
$$j_b = j_p(x_e') + (1-\alpha_n) j_n(x_e) \tag{6-112-b}$$
$$j_c = \alpha_n j_n(x_e) \tag{6-112-c}$$

Le courant de base résulte d'une part de l'injection de trous de la base vers l'émetteur et d'autre part de la recombinaison partielle des électrons dans la base (Fig. 6-25).

En fait la base est réalisée très étroite de sorte que la condition $W_{be} << L_n$ rend le coefficient de transfert α_n très voisin de 1 (pour $W_{be}=L_n/10$ l'expression (6-111) donne $\alpha_n=0{,}995$). Les équations (6-112) peuvent alors se simplifier et s'écrivent

$$j_e = -j_n(x_e) - j_p(x_e') \tag{6-113-a}$$
$$j_b = j_p(x_e') \tag{6-113-b}$$
$$j_c = j_n(x_e) \tag{6-113-c}$$

Le gain en courant α du transistor s'écrit donc

$$\alpha = -j_c/j_e = 1/\left(1 + j_p(x_e')/j_n(x_e)\right) \tag{6-114}$$

Le gain est par conséquent conditionné par le rapport des taux d'injection d'électrons et de trous à la jonction émetteur-base. Pour optimiser ce gain, il faut augmenter le rapport $j_n(x_e)/j_p(x_e')$ c'est-à-dire favoriser l'injection émetteur $\rightarrow$ base d'électrons au détriment de l'injection base $\rightarrow$ émetteur de trous.

La composante de trous du courant d'émetteur s'écrit, compte tenu des conventions de signe adoptées sur la figure (6-25)

$$j_p(x_e^{'}) = e\,D_p\,(dp/dx)_{x_e^{'}} \tag{6-115}$$

Si la longueur d_e de l'émetteur est très supérieure à la longueur de diffusion L_p des trous, la distribution de trous dans l'émetteur est exponentielle (Eq. 2-47) et s'écrit

$$\Delta p(x) = \Delta p(x_e^{'})\,e^{(x-x_e^{'})/L_p} \tag{6-116}$$

Le courant de trous dans l'émetteur s'écrit donc

$$j_p(x) = \frac{eD_p}{L_p}\,\Delta p(x_e^{'})\,e^{(x-x_e^{'})/L_p} \tag{6-117}$$

et en $x = x_e^{'}$

$$j_p(x_e^{'}) = \frac{eD_p}{L_p}\,\Delta p(x_e^{'}) \tag{6-118}$$

La composante électronique du courant d'émetteur s'écrit, compte tenu des conventions de signe,

$$j_n(x_e) = -e\,D_n\,(dn/dx)_{x_e} \tag{6-119}$$

Si la longueur W_b de la base est très inférieure à la longueur de diffusion des électrons on peut négliger les recombinaisons. La distribution des électrons dans la base est alors linéaire (Eq. 2-48) et le courant s'écrit

$$j_n(x_e) = -e\,D_n\,\frac{\Delta n(x_c) - \Delta n(x_e)}{W_{be}} \tag{6-120}$$

où $W_{be} = x_c - x_e$ représente la longueur effective de la base.

Lorsque les électrons arrivent en $x = x_c$, ils sont happés par le champ électrique qui règne dans la zone de charge d'espace de la jonction collecteur-base, de sorte que $\Delta n(x_c) \approx 0$. Le courant d'électrons s'écrit donc

$$j_n(x_e) = \frac{eD_n}{W_{be}}\,\Delta n(x_e) \tag{6-121}$$

En portant les expressions (6-118 et 121) dans l'expression du gain, α s'écrit

$$\alpha = 1 \Big/ \left(1 + \frac{D_p}{D_n}\,\frac{W_{be}}{L_p}\,\frac{\Delta p(x_e^{'})}{\Delta n(x_e)}\right) \tag{6-122}$$

Nous avons écrit $\Delta n(x_c) \approx 0$. Ceci revient à supposer que les électrons qui atteignent le collecteur en x_c sont évacués instantanément, ou en d'autres termes que leur vitesse d'entraînement en x_c est infinie. En fait cette vitesse n'est pas infinie mais limitée à la vitesse de saturation v_s qu'atteignent les électrons lorsque le champ électrique créé dans la zone de charge d'espace de la jonction collecteur-base, par la tension de polarisation inverse, atteint et dépasse la valeur critique E_{c2}.

Pour prendre en considération cette vitesse de saturation il suffit d'écrire la conservation du courant collecteur, c'est-à-dire d'écrire que ce courant a la même valeur sous chacune de ses deux formes consécutives : diffusion dans la base, vitesse de saturation dans la zone de charge d'espace.

Dans la base les équations (6-113 et 120) permettent d'écrire

$$j_c = -eD_n \left(\Delta n(x_c) - \Delta n(x_e)\right) / W_{be} \tag{6-123}$$

Dans la zone de charge d'espace

$$j_c = -e\,\Delta n(x_c)\, v_s \tag{6-124}$$

Le système d'équations (6-123 et 124) permet de calculer à la fois $\Delta n(x_c)$ et j_c, on obtient

$$j_c = e\, D_n\, \Delta n(x_e) / (W_{be} + D_n / v_s) \tag{6-125}$$

Le gain α est toujours donné par l'expression (6-122) dans laquelle W_{be} est remplacé par une longueur effective W_{be} donnée par

$$W_{beff} = W_{be} + D_n / v_s \tag{6-126}$$

Ainsi, compte tenu de la relation d'Einstein $D/\mu = kT/e$, le gain α s'écrit

$$\alpha = 1 \Big/ \left(1 + \frac{\mu_p}{\mu_n} \frac{W_{beff}}{L_p} \frac{\Delta p(x_e')}{\Delta n(x_e)}\right) \tag{6-127}$$

Revenons sur le transistor à homojonctions. Les densités d'électrons et de trous injectés au niveau de la jonction émetteur-base sont données par (Eq. 2-39)

$$\Delta n(x_e) = \frac{n_i^2}{N_{ab}} \left(e^{eV_{eb}/kT} - 1\right) \tag{6-128-a}$$

$$\Delta p(x_e') = \frac{n_i^2}{N_{de}} \left(e^{eV_{eb}/kT} - 1\right) \tag{6-128-b}$$

Ainsi le rapport des taux d'injection s'écrit

$$\frac{\Delta p(x_e^{'})}{\Delta n(x_e)} = \frac{N_{ab}}{N_{de}} \tag{6-129}$$

où $N_{ab} et\ N_{de}$ représentent respectivement les dopages de la base et de l'émetteur. Le gain α s'écrit alors

$$\alpha = 1 \Big/ \left(1 + \frac{\mu_p}{\mu_n} \frac{W_{beff}}{L_p} \frac{N_{ab}}{N_{de}} \right) \tag{6-130}$$

On retrouve une expression analogue à l'expression (3-17) que nous avons établie, pour un transistor pnp, à partir du modèle d'Ebers-Moll. Pour optimiser le gain il faut, entre autre, réaliser la condition $N_{de} >> N_{ab}$, c'est-à-dire doper la base beaucoup moins que l'émetteur.

Considérons maintenant le transistor à hétérojonction. Les figures (6-26-a et b) représentent le diagramme énergétique du transistor à l'équilibre thermodynamique et sous polarisation en régime normal avec $V_{eb} < 0$ de l'ordre de V_d et $V_{cb} >> 0$. Le diagramme de la figure (6-26-b) montre clairement que la discontinuité ΔE_v des bandes de valence au niveau de la jonction émetteur base, réduit fortement l'injection de trous depuis la base vers l'émetteur, sans pour autant que la discontinuité ΔE_c des bandes de conduction limite l'injection d'électrons depuis l'émetteur vers la base. Ainsi le rapport $\Delta p(x_e^{'}) / \Delta n(x_e)$ et considérablement diminué et par suite le gain α augmenté. C'est la spécificité du transistor HBT.
Mais il faut aussi noter que la présence de ΔE_v, en réduisant considérablement l'injection de trous, autorise un dopage de base beaucoup plus important que dans le transistor à homojonctions. En particulier des transistors de type np^+n, avec une base très dopée ($N_{ab} \approx 10^{19}\, cm^{-3}$), peuvent être réalisés avec un gain important. Le fait de pouvoir doper la base de manière importante autorise la réalisation de bases très étroites, inférieures à 0,1 µm, sans pour autant que les polarisations entraînent un recouvrement des zones de charge d'espace. *L'intérêt de la discontinuité ΔE_v des bandes de valence est donc multiple : le fort dopage de la base réduit la résistance série de base ; la très faible épaisseur de base réduit le temps de transit des porteurs et par suite augmente la fréquence de coupure du transistor*. Des transistors HBT présentant des fréquences de coupure supérieures à 100 GHz ont été réalisés avec le couple GaAlAs/GaAs. Ces transistors sont à l'heure actuelle plus performants que les transistors à effet de champ en ce qui concerne d'une part la puissance dissipée et d'autre part la fréquence de coupure. Il faut noter que, lorsque la technologie le permettra, le couple GaAlAs/GaAs pourra être avantageusement remplacé par les couples InAlAs/InGaAs ou Si/SiGe, qui présentent des discontinuités de bandes de valence ΔE_v plus importantes. Le deuxième couple présente en outre l'avantage d'être compatible avec la technologie silicium, il a toutefois un inconvénient majeur lié au fort désaccord de maille qui

existe entre le silicium et le germanium, ceci entraîne la présence de contraintes internes qui limitent fortement la stabilité thermique du dispositif.

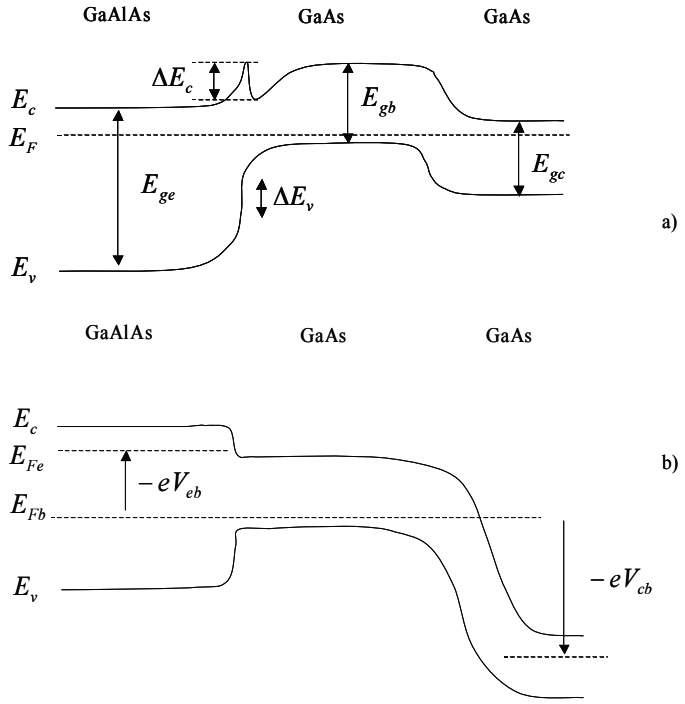


Figure 6-26 : Diagramme énergétique a) A l'équilibre thermodynamique b) Sous polarisation

La discontinuité ΔE_c des bandes de conduction ne joue pas un rôle aussi crucial que ΔE_v. Néanmoins cette discontinuité joue un double rôle à la fois positif et négatif. Son action négative est associée au fait que l'injection d'électrons de l'émetteur vers la base n'est plus de type purement diffusif, comme dans le transistor à homojonctions, mais présente de fortes composantes de types émission thermoélectronique par dessus la barrière et effet tunnel à travers la barrière. Ceci perturbe l'efficacité d'injection de l'émetteur et par suite réduit le courant collecteur. L'action positive résultant de la présence de ΔE_c est associée au fait que cette discontinuité joue le rôle de rampe de lancement pour l'injection des électrons de l'émetteur vers la base (Fig. 6-26-b). En effet les électrons issus de l'émetteur pénètrent dans la base avec un excès d'énergie ΔE_c par rapport au bas de la bande de conduction, c'est-à-dire avec une énergie cinétique égale à ΔE_c, considérable. Ceci entraîne la présence d'électrons chauds et de transport à travers la base de type balistique qui, en réduisant le temps de transit, augmente la fréquence de coupure du transistor.

6.4.2. *Courants d'émetteur et de collecteur*

Le calcul du courant d'émetteur doit prendre en considération d'une part, au niveau de la barrière, l'émission thermoélectronique éventuellement assistée d'effet tunnel, et d'autre part, au-delà de la barrière, les phénomènes de diffusion.

La présence de la discontinuité ΔE_v des bandes de valence permet de négliger l'injection de trous depuis la base vers l'émetteur, de sorte que le courant d'émetteur, qui se limite alors au courant d'électrons, s'écrit au niveau de la barrière

$$j_e = -\, j_n(X_e) \tag{6-131}$$

A ce niveau le courant d'électrons est de type émission thermoélectronique. Ce courant résulte de la somme algébrique des courants d'émission émetteur → base et base → émetteur. En négligeant l'effet tunnel à travers la barrière, ce courant s'écrit donc

$$j_n(X_e) = e v_{th} \left(n(X_e^-) - n(X_e^+) e^{-\Delta E_c / kT} \right) \tag{6-132}$$

où $v_{th} = (8kT / \pi m_e)^{1/2}$ est la vitesse thermique moyenne des électrons. A l'équilibre thermodynamique, les densités d'électrons côté émetteur et côté base, au niveau de la jonction métallurgique sont données par

$$n_o(X_e^-) = n(x_e^{'})\, e^{-\gamma e V_d / kT} \tag{6-133-a}$$
$$n_o(X_e^+) = n(x_e)\, e^{(1-\gamma) e V_d / kT} \tag{6-133-b}$$

où $V_d = \phi_b - \phi_e$ est la tension de diffusion de la jonction, $e\phi_b$ et $e\phi_e$ étant les travaux de sortie des semiconducteurs constituant respectivement la base et l'émetteur. Le paramètre $\gamma = \varepsilon_b N_{ab} / (\varepsilon_b N_{ab} + \varepsilon_e N_{de})$ représente la distribution de cette tension entre les deux régions (Eq. 6-33).

Sous polarisation V_{eb} de la jonction émetteur-base, la tension aux bornes de la jonction devient $V_d + V_{eb}$, V_{eb} étant négatif en polarisation directe. Les densités de porteurs s'écrivent alors

$$n(X_e^-) = n(x_e^{'})\; e^{-\gamma e (V_d + V_{eb}) / kT} \tag{6-134-a}$$
$$n(X_e^+) = n(x_e)\; e^{(1-\gamma) e (V_d + V_{eb}) / kT} \tag{6-134-b}$$

$n(x_e^{'}) = N_{de}$, où N_{de} est le dopage de l'émetteur. En explicitant les densités de porteurs, l'expression (6-132) s'écrit

$$j_n(X_e) = e v_{th} \left(N_{de}\, e^{-\gamma e (V_d + V_{eb}) / kT} - n(x_e) e^{\left((1-\gamma) e (V_d + V_{eb}) - \Delta E_c\right) / kT} \right) \tag{6-135}$$

L'expression (6-135) est le courant d'électrons à la jonction métallurgique. A ce niveau ce courant est de nature thermoélectronique. Si on néglige les phénomènes

de recombinaison dans la zone de charge d'espace de la jonction ce courant se conserve à la traversée de cette zone et on peut écrire l'égalité

$$j_n(x_e) = j_n(X_e) \tag{6-136}$$

Dans la base, c'est-à-dire au-delà de x_e, le courant devient de diffusion et on peut par conséquent l'écrire sous la forme

$$j_n(x_e) = -e\,D_n\,(dn/dx)_{x_e} \tag{6-137}$$

Si on néglige les recombinaisons dans la base, la distribution des électrons est linéaire et l'expression (6-137) s'écrit

$$j_n(x_e) = -eD_n\left(n(x_c) - n(x_e)\right)/W_{be} \tag{6-138}$$

Enfin au collecteur, si on néglige les recombinaisons dans la zone de charge d'espace de la jonction collecteur-base, le courant collecteur s'écrit

$$j_c = j_n(x_c) = j_n(x_e) = -eD_n\left(n(x_c) - n(x_e)\right)/W_{be} \tag{6-139}$$

Comme précédemment (Eq. 6-124), dans la zone de charge d'espace de la jonction collecteur-base le courant est limité par la vitesse de saturation v_s des électrons et s'écrit

$$j_c = -e\,n(x_c)\,v_s \tag{6-140}$$

Les équations (6-139 et 140) permettent d'éliminer $n(x_c)$ et d'écrire le courant collecteur sous la forme

$$j_c = e\,D_n\,n(x_e)/W_{beff} \tag{6-141}$$

avec

$$W_{beff} = W_{be} + D_n/v_s$$

Les différentes relations permettent d'écrire $j_n(X_e) = j_c$, de sorte qu'en explicitant ces courants à partir des expressions (6-135 et 141) on obtient l'expression de $n(x_e)$

$$n(x_e) = ev_{th}N_{de}\frac{e^{-\gamma e(V_d+V_{eb})/kT}}{eD_n/W_{beff} + ev_{th}e^{\left((1-\gamma)e(V_d+V_{eb})-\Delta E_c\right)/kT}} \tag{6-142}$$

En explicitant $n(x_e)$ dans l'expression (6-141) on obtient l'expression du courant d'émetteur et de collecteur

$$j_c = -j_e = ev_{th} N_{de} \frac{e^{-\gamma e(V_d+V_{eb})/kT}}{1+\frac{v_{th} W_{beff}}{D_n} e^{((1-\gamma)e(V_d+V_{eb})-\Delta E_c)/kT}} \qquad (6\text{-}143)$$

Considérons un transistor de type np^+n réalisé avec le couple $Ga_{0.7}Al_{0.3}As/GaAs$. La base étant beaucoup plus dopée que l'émetteur, la condition $N_{ab} >> N_{de}$ entraîne $\gamma \approx 1$. D'autre part, pour ce couple de semiconducteurs la différence de gaps est $\Delta E_g \approx 450$ meV. Cette différence se distribue à raison de 67 % pour ΔE_c et 33 % pour ΔE_v. Il en résulte que $\Delta E_c \approx 300 meV$ est très supérieur à kT. Le dénominateur de l'expression (3-143) se réduit donc à 1 et le courant présente une loi de variation exponentielle donnée par

$$j_c = -j_e \approx ev_{th} N_{de} e^{-e(V_d+V_{eb})/kT} \qquad (6\text{-}144)$$

Lorsque la tension de polarisation négative V_{eb} atteint la valeur de la tension de diffusion, $V_{eb} \approx -V_d$, la jonction émetteur-base est quasiment en régime de bandes plates (Fig. 6-26-b) et le courant atteint la valeur limite

$$j_c = -j_e \approx e v_{th} N_{de} \qquad (6\text{-}145)$$

Tous les électrons, N_{de}, présents dans l'émetteur sont injectés dans la base et diffusent jusqu'au collecteur. La vitesse thermique des électrons dans GaAs est de l'ordre de 10^7 cm/s de sorte que pour un dopage $N_{de}=10^{17}$ cm^{-3}, la densité de courant est $j_c \approx 1,6\ A/cm^2$. Pour une surface d'émetteur de 5x5 μm^2 la valeur du courant collecteur est donc $I_c \approx 40\ mA$.

7. Transistors à effet de champ JFET et MESFET

Les transistors à effet de champ fonctionnent sur un principe totalement différent de celui du transistor à jonctions. Ils consistent essentiellement en un barreau conducteur appelé *canal*, dont les deux extrémités portent des électrodes appelées respectivement *source* et *drain*. Lorsque le barreau est polarisé longitudinalement par une tension drain-source V_{ds} , un courant appelé *courant de drain* I_d circule dans le canal Les transistors à effet de champ sont de trois types : JFET, MESFET et MOSFET. Le MOSFET a pris une telle importance dans la micro-élecronique qu'un chapitre entier lui sera consacré. Les JFET et les MESFET sont réservés pour certaines applications faible bruit ou haute fréquence.
Le chapitre est organisé de la manière suivante.

7.1. Transistor de type JFET
7.2. Transistor de type MESFET
7.3. Transistor MESFET à canal court

7.1. Transistor à effet de champ à jonction - JFET

7.1.1. *Structure et principe de fonctionnement*

Lorsque le barreau est polarisé longitudinalement par une tension drain-source V_{ds} , un courant appelé *courant de drain* I_d circule dans le canal. L'intensité de ce courant est proportionnelle à la conductance du canal,

$$I_d = g\, V_{ds}$$

Le courant I_d est modulé par un signal transversal qui, appliqué au barreau par l'intermédiaire d'une électrode latérale appelée *grille*, modifie la conductance du canal. La conductance du canal est donnée par $g=\sigma S/L$ où $\sigma=ne\mu_n+pe\mu_p$ représente la conductivité du matériau, S et L représentent respectivement la section et la longueur du canal. La variation de g peut être obtenue par modulation de la section du canal ou par modulation de la densité de porteurs. Les divers types de transistors à effet de champ diffèrent par la nature du paramètre modulé et par le processus mis en œuvre dans cette modulation.

Dans le transistor à effet de champ à jonction, JFET (Junction Field Effect Transistor), l'électrode de commande est constituée par une jonction pn latérale, polarisée en inverse. La variation de la section conductrice du barreau est obtenue par modulation de la largeur de la zone de déplétion de la jonction résultant de la variation de la tension de polarisation. Dans le transistor à effet de champ à barrière de Schottky, MESFET (MEtal Semiconductor Field Effect Transistor), le processus mis en jeu est le même que précédemment mais la zone de déplétion est obtenue par polarisation inverse d'une diode Schottky. Dans un transistor à effet de champ à grille isolée, IGFET (Insulated Gate Field Effect Transistor) ou MOSFET (Metal Oxide Semiconductor Field Effect Transistor) la conductance du canal est modulée par variation de la densité de porteurs. La capacité MOS est polarisée en régime de forte inversion, la densité de porteurs dans le canal que constitue la couche d'inversion, est modulée par le signal appliqué à la grille.

Dans tous ces transistors le courant est transporté par un seul type de porteur. C'est la raison pour laquelle on les qualifie de *transistors unipolaires* pour les différencier du transistor à jonctions, qualifié de *bipolaire* dans la mesure où les deux types de porteurs sont mis en jeu.

On supposera dans ce qui suit, afin de simplifier le formalisme, que la source est portée à la masse, V_s=0. Les tensions de polarisation drain-source et grille-source s'écriront simplement $V_{ds}=V_d-V_s=V_d$, $V_{gs}=V_g-V_s=V_g$.

La structure du transistor est représentée sur la figure (7-1). Le canal conducteur est constitué par une couche épitaxiée de type n, sur un substrat semi-isolant ou de type p. La zone de charge d'espace à l'interface isole électriquement la couche épitaxiée, du substrat. Une diffusion de type p^+ à la surface de la couche réalise l'électrode de grille. Le dopage est réalisé p^+ pour que la zone de charge d'espace de la jonction grille-canal se développe essentiellement dans le canal. Deux diffusions n^+, aux extrémités du canal, permettent d'assurer les contacts ohmiques de source et de drain. On appellera Z et L la largeur et la longueur du canal , et a son épaisseur.

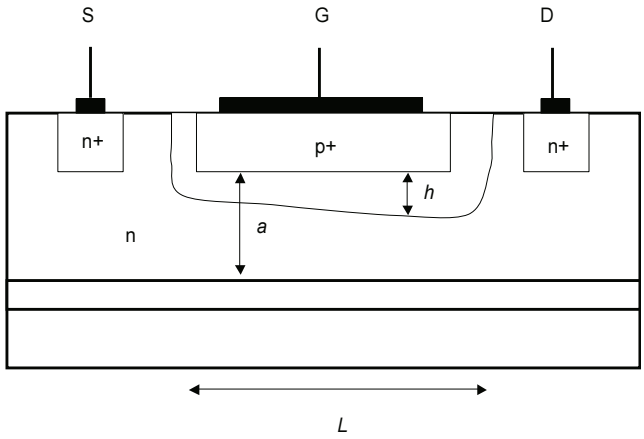


Figure 7-1 : Schéma de principe du transistor à effet de champ à jonction

La grille est polarisée par une tension V_g négative, qui polarise en inverse la jonction p^+n et développe dans le canal une zone vide de porteurs de profondeur h. La section conductrice du canal est alors a-h. Le drain est polarisé par une tension V_d positive créant dans le canal un courant de drain I_d. En raison de la polarisation du drain, la jonction grille-canal est plus polarisée côté drain que côté source, ce qui entraîne une variation de h, c'est-à-dire de la section conductrice, tout le long du canal.

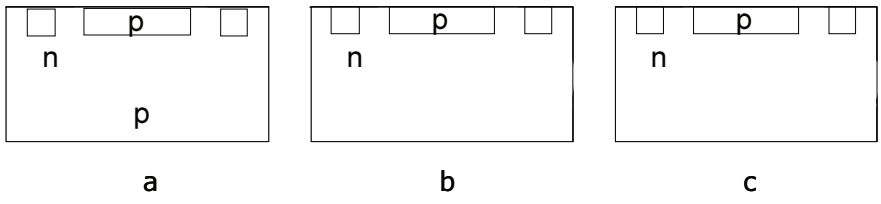


Figure 7-2 : Régimes de fonctionnement du JFET

Le courant de drain I_d est par conséquent fonction de la tension grille V_g et de la tension drain V_d , les réseaux de caractéristiques du transistor sont donnés par la loi $I_d = f(V_g, V_d)$.
Le principe de fonctionnement est schématisé sur la figure (7-2), qui représente la zone active du transistor sous différents régimes de polarisation.

- $V_d << V_{ds\ t}$: La tension drain-source est faible, un courant circule dans le canal entre la source et le drain La variation relative de la section du canal $\Delta(a-h)/a$ est négligeable, la conductance du canal reste sensiblement constante, le courant de drain varie proportionnellement à la tension drain-source. C'est le *régime linéaire*.

- $V_d \leq V_{dsat}$: Quand la tension drain-source augmente la déformation du canal devient significative et la conductance du canal diminue. Le courant présente alors une variation sous-linéaire avec la tension V_d et amorce une saturation. Lorsque les zones de charge d'espace se recouvrent, la largeur conductrice du canal devient nulle côté drain. C'est le *régime de pincement*, la tension drain-source correspondante est appelée *tension de saturation* V_{dsat} , le courant correspondant est appelé *courant de saturation* I_{dsat}.

- $V_d > V_{dsa\ t}$: La tension drain-source entraîne une distribution de potentiel sur toute la longueur du canal, avec une valeur locale qui varie de $V_{loc}=V_s=0$ coté source à $V_{loc}=V_d$ coté drain. En un point donné du canal ce potentiel local entraîne une polarisation locale de la jonction grille-canal $V(h)=(V_n-V_p)_{loc}=V_{di}+V_{loc}-V_g$ qui crée dans cette dernière une zone de déplétion de largeur h (V_{di} est la tension de diffusion de la jonction). Le pincement du canal se produit au point P où le potentiel local du canal V_{loc} prend la valeur critique V_o qui entraîne $h=a$. En d'autres termes, au point de pincement, quelle que soit sa position, le potentiel du canal est toujours égal à V_o. Lorsque $V_d=V_{dsat}$,le pincement se produit au drain, de sorte que V_o n'est autre que la tension de saturation, $V_o=V_{dsat}$. Lorsque V_d augment au delà de V_{dsat} ,la distribution du potentiel le long du canal évolue et le point du canal où $V_{loc}=V_{dsat}$ se déplace vers la source. Le point de pincement P, qui suit ce dernier, se déplace donc vers la source. L'excédent de tension de drain V_d-V_{dsat} se retrouve aux bornes d'une zone de déplétion qui s'établit alors longitudinalement à l'extrémité du canal entre le point P et le drain. Le canal conducteur est ainsi raccourci, mais si sa variation relative de longueur reste faible sa conductance reste constante. D'autre part la tension à ses bornes reste aussi constante, égale à V_{dsat}. Il en résulte donc que le courant de drain reste constant, $I_d=I_{dsat}$, c'est le *régime de saturation*. Le courant est transporté par les porteurs majoritaires qui circulent dans le canal entre la source et le point de pincement. Ces porteurs sont ensuite injectés dans la zone de charge d'espace longitudinale où ils sont soumis à un champ favorable qui les propulse vers l'électrode de drain. Dans cette zone, les porteurs se comportent comme dans la zone de charge d'espace de la jonction base-collecteur d'un transistor bipolaire.

Lorsque la tension drain-source varie, le régime de pincement est atteint d'autant plus rapidement que la zone de charge d'espace est importante à $V_d=0$,

c'est-à-dire que V_g est important. Le réseau de caractéristiques est représenté sur la figure (7-4).

Lorsque la tension de polarisation grille-source augmente, la largeur conductrice du canal à $V_d = 0$ diminue. A partir d'une certaine valeur de V_g, le canal est obturé quelle que soit la valeur de V_d. Le transistor est bloqué, la tension grille correspondante est appelée *tension de cut-off* V_{gco}. Dans l'autre sens, la tension de polarisation ne peut guère aller au-delà de zéro car la jonction grille-canal se trouve alors polarisée dans le sens direct et devient un court-circuit. Seules des polarisations de grille négatives sont compatibles avec le fonctionnement du transistor. Le transistor précédent est réalisé à canal n, l'équivalent existe à canal p, les polarités des tensions sont alors inversées.

Le signal à amplifier est appliqué à la grille à travers laquelle ne circule que le courant inverse de la jonction grille-canal. Ainsi la spécificité de ce transistor, qui le différencie du transistor bipolaire, est la valeur élevée de son impédance d'entrée. Notons que cette impédance d'entrée est d'autant plus importante que le courant de saturation de la jonction grille-canal est faible, c'est-à-dire que la densité de porteurs minoritaires est faible (grille très dopée), et que les dimensions géométriques du transistor sont faibles (Z et L petits).

7.1.2. *Equations fondamentales du JFET*

Considérons la structure précédente dont la zone active est schématisée sur la figure (7-3). Les axes $o\vec{x}$ et $o\vec{y}$ représentent les axes longitudinal et transversal de la structure, l'origine est prise côté source. La largeur totale du canal est a, supposée constante. La largeur de la zone de déplétion en un point d'abscisse x est représentée par le paramètre h.

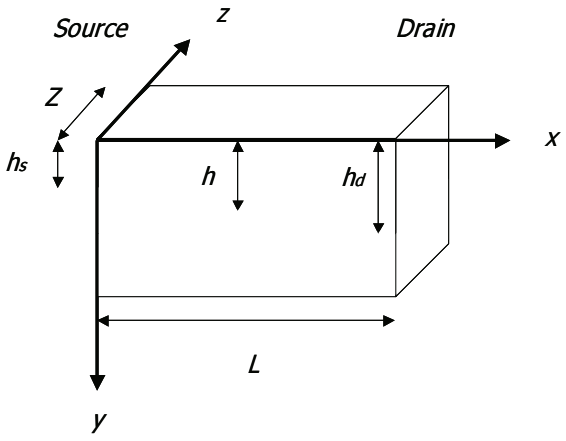


Figure 7.3 : Zone active du FET

Nous nous placerons dans l'hypothèse du canal graduel, le potentiel le long du canal et par suite la largeur h de la zone de charge d'espace varient graduellement entre la source et le drain. Cette largeur est h_s côté source et h_d côté drain.

a. Polarisation transversale - Tension de pincement

Le potentiel dans la zone de charge d'espace du canal est donné par l'intégration de l'équation de Poisson

$$\Delta V = \frac{d^2V}{dx^2} + \frac{d^2V}{dy^2} + \frac{d^2V}{dz^2} = -\frac{\rho(x,y,z)}{\varepsilon} \tag{7-1}$$

Si le dopage du canal est homogène, la densité de charge est constante. En fait pour des raisons liées aux techniques de fabrication, ce dopage est homogène dans le plan de la structure mais peut varier dans la direction perpendiculaire. Nous écrirons cette densité de charge sous la forme $\rho(y)$. Pour un canal n par exemple, $\rho(y) = e\,N_d(y)$, pour un canal p , $\rho(y) = -e\,N_a(y)$.

En ce qui concerne le potentiel, nous pouvons écrire pour des raisons de symétrie, qu'il est constant dans la direction z de sorte que $d^2V/dz^2=0$. D'autre part nous ferons l'hypothèse du canal graduel. Le canal étant conducteur et sa longueur étant beaucoup plus importante que la largeur h de la zone de charge d'espace, la variation du champ électrique est plus importante dans la direction perpendiculaire à la structure que dans la direction longitudinale. On peut donc ramener l'équation de Poisson à une dimension

$$\frac{d^2V}{dy^2} = -\frac{dE_y}{dy} = -\frac{\rho(y)}{\varepsilon} \tag{7-2}$$

En intégrant l'équation (7-2) de h à y , avec la condition $E_y=0$ en $y=h$, limite de la zone de charge d'espace, on obtient

$$\frac{dV}{dy} = -E_y = \frac{1}{\varepsilon}\left(\int_0^h \rho(y)\,dy - \int_0^y \rho(y)dy\right) \tag{7-3}$$

Les intégrales qui apparaissent dans l'expression précédente, représentent chacune le nombre total de charges, $Q(h)$ ou $Q(y)$, contenues dans un volume d'épaisseur h ou y, et de surface unité

$$Q(h) = \int_0^h \rho(y)dy \qquad Q(y) = \int_0^y \rho(y)dy \tag{7-4}$$

Ainsi

$$\frac{dV}{dy} = \frac{1}{\varepsilon}\big(Q(h) - Q(y)\big) \tag{7-5}$$

En intégrant une deuxième fois entre $y = 0$ et $y= h$ on obtient la tension aux bornes de la zone de déplétion, $V(h) = V(y = h) - V(y = 0)$.

$$V(h)=\frac{1}{\varepsilon}\int_0^h \big(Q(h)-Q(y)\big)dy$$

$$V(h) =\frac{1}{\varepsilon}\left(Q(h)\int_0^h dy-\int_0^h Q(y)dy \right)$$

soit

$$V(h)=\frac{1}{\varepsilon}\int_0^h y\,\rho(y)dy \qquad (7\text{-}6)$$

En particulier si le canal est de type n dopé de manière homogène avec une densité N_d de donneurs, la charge d'espace *est* $\rho(y) = \rho = e\,N_d$ et l'expression (7-6) s'écrit simplement

$$V(h)=\frac{1}{\varepsilon}eN_d\int_0^h ydy=\frac{eN_d}{2\varepsilon}h^2$$

on retrouve la relation bien connue qui existe entre la largeur de la zone de charge d'espace et la tension aux bornes de cette zone, dans la zone de transition d'une jonction pn abrupte (Eq. 2-20).

La valeur maximum de h est a. Lorsque h atteint cette valeur en un point du canal, celui-ci est obturé et le régime de pincement est atteint. La tension $V(h=a)$ correspondante est appelée *tension de pincement* V_p

$$V_p=\frac{1}{\varepsilon}\int_0^a y\,\rho(y)\,dy \qquad (7\text{-}7)$$

La tension $V(h)$ représente la différence de potentiel $V_n(x)$-$V_p(x)$ entre les régions n et p de la jonction grille-canal au point d'abscisse x.

- *En l'absence de toute polarisation* (V_g=0, V_d=0), la différence $V_n(x)$-$V_p(x)$ correspond tout simplement à la tension de diffusion $V_{di}(x)$ de la jonction grille-canal au point d'abscisse x. Si le dopage du canal est uniforme suivant x, cette tension est indépendante de x, de sorte que $V(h)$ s'écrit

$$V(h)_0 = V_{di} \qquad (7\text{-}8)$$

- *Sous polarisation* ($V_g\neq0$, $V_d\neq0$), la tension grille-source V_g porte la région de type p de la jonction au potentiel V_g, la tension drain-source V_d se distribue le long du canal et porte le point d'abscisse x de la région de type n à un potentiel $V(x)$. Ainsi, au point d'abscisse x, la jonction grille-canal est polarisée par la tension $V=V_g$-$V(x)$. La tension $V(h)$ devient alors $V(h)=V_{di}$ -V et s'écrit

$$V(h) = V_{di} - V_g + V(x) \tag{7-9}$$

Coté source: $x = 0$ $V(x) = V_s = 0$ et $h = h_s \rightarrow V(h_s) = V_{di} - V_g$
Coté drain: $x = L$ $V(x) = V_d$ et $h = h_d \rightarrow V(h_d) = V_{di} - V_g + V_d$

b. Polarisation longitudinale - Courant de drain

La composante suivant l'axe x de la densité de courant en un point de coordonnées x,y,z du canal conducteur est simplement donnée par la loi d'Ohm

$$J_x(x,y,z) = \sigma(x,y,z) E_x(x,y,z) = -\sigma(x,y,z) \frac{dV(x,y,z)}{dx} \tag{7-10}$$

La quantité $\rho(y)$ représente à la fois la densité de charges fixes dans la zone de déplétion ($0<y<h$), et la densité de charges mobiles dans la zone conductrice ($h<y<a$). Pour un canal n par exemple $\rho(y)$ représente à la fois $eN_d(y)$ pour $0<y<h$ et $en(y)$ pour $h<y<a$. Ainsi, en supposant constante la mobilité, la conductance s'écrit

$$\sigma(x,y,z) = \mu\rho(y) \tag{7-11}$$

D'autre part, le canal conducteur n'est polarisé que longitudinalement. Dans la direction z il n'existe pas de polarisation, et dans la direction y la tension de polarisation (V_g) est limitée à la zone de déplétion. Les plans yz sont des plans équipotentiels de sorte que $V(x,y,z)=V(x)$, où $V(x)$ résulte de la tension drain-source.

La densité de courant de drain J_d , comptée par convention positivement dans le sens drain→source ($J_d = -J_x$), s'écrit donc

$$J_d(x,y) = \mu\rho(y) \frac{dV(x)}{dx} \tag{7-12}$$

Le courant de drain I_d est obtenu en intégrant la densité de courant $J_d(x,y)$ sur toute la section conductrice du canal, soit

$$I_d = \int_0^Z \int_h^a J_d(x,y)\, dz\, dy = Z \int_h^a J_d(x,y)\, dy = Z\mu \frac{dV(x)}{dx} \int_h^a \rho(y)\, dy \tag{7-13}$$

Le potentiel $V(x)$ est lié à la tension $V(h)$ par l'expression (7-9), de sorte que $dV(x)/dx = dV(h)/dx$, cette dérivée peut s'écrire $dV(h)/dh.dh/dx$. En outre, la dérivée de l'expression (7-6) donne $dV(h)/dh = h\,\rho(h)/\varepsilon$. L'expression (7-13) s'écrit donc

$$I_d = \frac{Z\mu}{\varepsilon} h\rho(h) \frac{dh}{dx} \int_h^a \rho(y) dy \tag{7-14}$$

soit

$$I_d\,dx=\frac{Z\mu}{\varepsilon}\big(Q(a)-Q(h)\big)h\rho(h)dh \tag{7-15}$$

On obtient le courant de drain en intégrant sur tout le barreau, c'est-à-dire de x=0 à x=L, et de h=h_s à h=h_d. Le courant étant conservatif I_d est constant. D'autre part nous avons supposé constante la mobilité μ des porteurs. Nous verrons dans l'étude du transistor à effet de champ à barrière de Schottky les modifications apportées par la prise en considération des variations éventuelles de mobilité. L'intégrale de l'expression précédente donne

$$I_d=\frac{Z\mu}{\varepsilon L}\int_{h_s}^{h_d}\big(Q(a)-Q(h)\big)h\rho(h)dh \tag{7-16}$$

c. Transconductance et conductance de drain

L'expression de I_d permet de calculer les deux paramètres fondamentaux du transistor que sont la *transconductance* ou pente g_m et la conductance du canal appelée plus communément *conductance de drain* g_d.

Lorsque le transistor est polarisé à un point de fonctionnement donné par les tensions statiques V_g et V_d , le courant de drain est I_d. La variation du courant de drain résultant des variations des tensions de grille et de drain s'écrit

$$dI_d=\frac{\partial I_d}{\partial V_g}dV_g+\frac{\partial I_d}{\partial V_d}dV_d \tag{7-17}$$

ce qui permet de linéariser le fonctionnement du transistor autour d'un point de polarisation par l'expression

$$i_d=g_m v_g+g_d v_d \tag{7-18}$$

où i_d , v_g et v_d représentent respectivement les variations du courant de drain et des tensions grille-source et drain-source.

La transconductance est donnée par

$$g_m=\frac{\partial I_d}{\partial V_g}=\frac{\partial I_d}{\partial h_s}\,\frac{\partial h_s}{\partial V_g}+\frac{\partial I_d}{\partial h_d}\,\frac{\partial h_d}{\partial V_g} \tag{7-19}$$

Les dérivées partielles sont obtenues à partir des expressions de I_d (7-16) et de $V(h)$ (7-6), en remplaçant dans cette dernière h par h_s et h_d successivement.

Rappelons que

$$\frac{d}{da}\int_a^b f(x)dx=\frac{d}{da}\big(F(b)-F(a)\big)=-\frac{d}{da}F(a)=-f(a)$$
$$\frac{d}{db}\int_a^b f(x)dx=\frac{d}{db}\big(F(b)-F(a)\big)=\frac{d}{db}F(b)=f(b)$$

Ainsi

$$\frac{\partial I_d}{\partial h_s}=-\frac{Z\mu}{\varepsilon L}\big(Q(a)-Q(h_s)\big)h_s\rho(h_s) \tag{7-20}$$

$$\frac{\partial I_d}{\partial h_d}=\frac{Z\mu}{\varepsilon L}\big(Q(a)-Q(h_d)\big)h_d\rho(h_d) \tag{7-21}$$

D'autre part h_s est relié à V_g par la relation (7-9). Coté sourse, en x=0, cette relation donne

$$V_g=V_{di}-V(h_s)=V_{di}-\frac{1}{\varepsilon}\int_0^{h_s} y\rho(y)dy \tag{7-22}$$

La dérivée de (7-22) s'écrit

$$\frac{dV_g}{dh_s}=-\frac{1}{\varepsilon}h_s\rho(h_s) \tag{7-23}$$

Enfin h_d est aussi relié à V_g par la relation (7-9). Coté drain, en x=L, cette relation donne

$$V_g=V_{di}+V_d-V(h_d)=V_{di}+V_d-\frac{1}{\varepsilon}\int_0^{h_d} y\rho(y)dy \tag{7-24}$$

Ainsi, dans la mesure où V_d est constant dans le calcul de g_m

$$\frac{\partial V_g}{\partial h_d}=-\frac{1}{\varepsilon}h_d\rho(h_d) \tag{7-25}$$

Compte tenu des expressions (7-20, 21, 23 et 25) l'expression (7-19) s'écrit

$$g_m=\frac{Z\mu}{L}\big(Q(h_d)-Q(h_s)\big) \tag{7-26}$$

La conductance du canal est donnée par

$$g_d=\frac{\partial I_d}{\partial V_d}=\frac{\partial I_d}{\partial h_s}\frac{\partial h_s}{\partial V_d}+\frac{\partial I_d}{\partial h_d}\frac{dh_d}{dV_d} \tag{7-27}$$

La largeur de la zone de charge d'espace côté source, est indépendante de la tension de drain, $\partial h_s/\partial V_d$=0. D'autre part h_d est relié à V_d par la relation (7-9). Coté drain cette relation donne

$$V_d = V_g - V_{di} + V(h_d) = V_g - V_{di} + \frac{1}{\varepsilon}\int_0^{h_d} y\rho(y)dy \tag{7-28}$$

Dans la mesure où V_g est constant dans le calcul de g_d, la dérivée de l'expression (7-28) s'écrit

$$\frac{\partial V_d}{\partial h_d} = \frac{1}{\varepsilon} h_d \rho(h_d) \tag{7-29}$$

La conductance de drain est alors donnée par

$$g_d = \frac{Z\mu}{L}\left(Q(a) - Q(h_d)\right) \tag{7-30}$$

Il est intéressant de comparer les expressions de g_m et de g_d.

Si $V_d = 0$, $h_d = h_s$ $\Rightarrow$ $g_m = g_{mo} = 0$ $g_d = g_{do} = \frac{Z\mu}{L}\left(Q(a) - Q(h_s)\right)$

Si $V_d > V_{dsat}$, $h_d = a$ $\Rightarrow$ $g_d = g_{ds} = 0$ $g_m = g_{ms} = \frac{Z\mu}{L}\left(Q(a) - Q(h_s)\right)$

On peut faire apparaitre une quantité g_o , spécifique de la structure du transistor et indépendante de toute polarisation, qui n'est autre que la conductance maximum que présenterait le canal si sa largeur conductrice était a.

$$g_o = Z\,\mu Q(a) / L \tag{7-31}$$

Ainsi

$$g_{do} = g_{ms} = g_o\left(1 - Q(h_s)/Q(a)\right) \tag{7-32}$$

Lorsque $V_d \geq V_{dsat}$, $g_d = g_{ds} = 0$, la conductance différentielle de drain devient nulle, le courant devient indépendant de la tension de drain, c'est le *régime de saturation.*

7.1.3. Réseau de caractéristiques

Pour obtenir le réseau de caractéristiques, c'est-à-dire la forme explicite de la loi $I_d = f(V_g, V_d)$, il suffit de calculer l'expression (7-16) en explicitant la distribution de charge $\rho(y)$. Cette distribution est étroitement liée à la technique d'élaboration du transistor. Considérons un canal n , à dopage homogène, avec une densité N_d de

donneurs excédentaires par rapport aux accepteurs résiduels. Il en résulte que dans le modèle de la déplétion totale, la densité de charge est constante dans la zone de charge d'espace et donnée par $\rho(y) = eN_d$. Ainsi $Q(h) = eN_d\, h$, $Q(a) = eN_d\, a$, $Q(h_s) = eN_d\, h_s$ et $Q(h_d) = eN_d\, h_d$. L'expression (7-16) s'écrit alors

$$I_d = \frac{Z\mu}{\varepsilon L} e^2 N_d^2 \int_{h_s}^{h_d} (a-h)h\,dh$$

soit

$$I_d = \frac{Z\mu e^2 N_d^2}{\varepsilon L}\left(\frac{a}{2}(h_d^2 - h_s^2) - \frac{1}{3}(h_d^3 - h_s^3)\right) \qquad (7\text{-}33)$$

On obtient h_d et h_s en intégrant l'équation (7-6) avec $\rho(y)=eN_d$

$$V(h) = \frac{eN_d}{2\varepsilon} h^2$$

Soit

$$h = \left(\frac{2\varepsilon}{eN_d} V(h)\right)^{1/2} \qquad (7\text{-}34)$$

V(h) est donné par l'expression (7-9)

En $x = 0$ $\quad h = h_s$ et $\quad V(h_s) = V_{di} - V_g$
En $x = L$ $\quad h = h_d$ et $\quad V(h_d) = V_{di} - V_g + V_d$

l'expression (7-34) permet donc d'écrire

$$h_s = \left(\frac{2\varepsilon}{eN_d}(V_{di} - V_g)\right)^{1/2} \qquad (7\text{-}35\text{-a})$$

$$h_d = \left(\frac{2\varepsilon}{eN_d}(V_{di} - V_g + V_d)\right)^{1/2} \qquad (7\text{-}35\text{-b})$$

En portant les expressions (7-35-a et b) dans l'expression (7-33) on obtient

$$I_d = \frac{Z\mu eN_d a}{L}\left(V_d - \frac{2}{3a}\sqrt{\frac{2\varepsilon}{eN_d}}\left[(V_{di} - V_g + V_d)^{3/2} - (V_{di} - V_g)^{3/2}\right]\right) \qquad (7\text{-}36)$$

Nous avons précédemment défini les quantités g_o (Eq. 7-31) et V_p (Eq. 7-7), en explicitant la densité de charge ces quantités s'écrivent

$$g_o = \frac{Z\mu e N_d a}{L} = \mu e N_d \frac{Za}{L} = \sigma \frac{S_0}{L} \qquad (7\text{-}37)$$

$$V_p = \frac{e N_d a^2}{2\varepsilon} \tag{7-38}$$

Ce qui permet d'écrire le courant de drain sous la forme

$$I_d = g_o \left(V_d - \frac{2}{3\sqrt{V_p}} (V_{di} - V_g + V_d)^{3/2} + \frac{2}{3\sqrt{V_p}} (V_{di} - V_g)^{3/2} \right) \tag{7-39}$$

Le paramètre g_o représente la valeur limite de la conductance de drain correspondant à $h_s = h_d = 0$. L'expression (7-39) permet de tracer le réseau de caractéristiques du transistor.

Pour de faibles valeurs de V_d un développement limité permet de linéariser l'expression (7-39), on obtient

$$I_d \approx g_o \left(1 - \sqrt{\frac{V_{di} - V_g}{V_p}} \right) V_d \tag{7-40}$$

c'est le *régime linéaire*. Lorsque $V_g = V_{di} - V_p$, $I_d = 0$ quelle que soit la valeur de V_d, la tension $V_{gco} = V_{di} - V_p$ est la *tension de cut-off*. Si par contre la grille est polarisée positivement à sa valeur limite $V_g = V_{di}$, la zone de charge d'espace de la jonction grille-canal disparait et la conductance du canal atteint sa valeur maximum g_o.

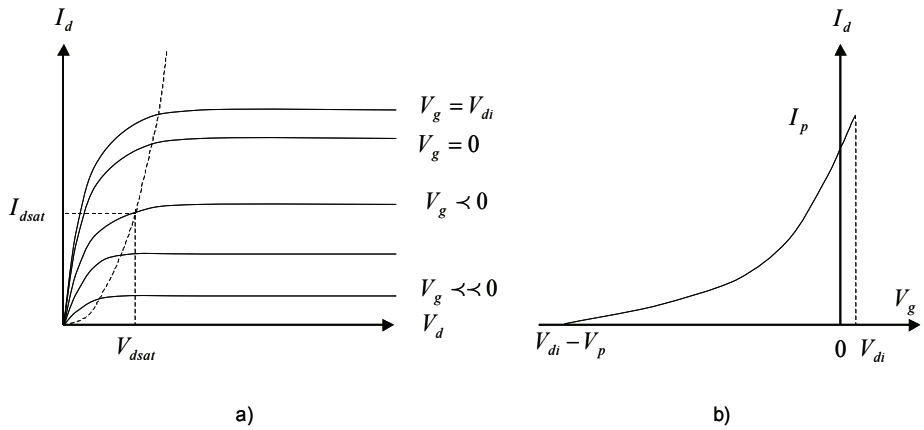


Figure 7-4 : Réseau de caractéristiques

Lorsque V_d devient plus important le courant augmente sous-linéairement et tend vers une valeur I_{dsat} correspondant au régime de saturation. Le réseau de caractéristiques est représenté sur la figure (7-4-a).

Le courant de saturation est atteint lorsque, pour une tension V_g donnée, le pincement du canal est réalisé au voisinage du drain, c'est-à-dire quand $h_d = a$. Il

suffit donc de faire $h_d = a$ dans l'expression (7-33) pour obtenir l'expression du courant de saturation

$$I_{dsat} = \frac{Z\mu e^2 N_d^2}{\varepsilon L}\left(\frac{a}{2}(a^2 - h_s^2) - \frac{1}{3}(a^3 - h_s^3)\right)$$

soit

$$I_{dsat} = \frac{Z\mu e^2 N_d^2 a^3}{6\varepsilon L}\left(1 - 3\frac{h_s^2}{a^2} + 2\frac{h_s^3}{a^3}\right) \qquad (7\text{-}41)$$

En définissant le *courant de pincement* par

$$I_p = \frac{Z\mu e^2 N_d^2 a^3}{6\varepsilon L} = \frac{1}{3} g_o V_p \qquad (7\text{-}42)$$

Le courant de saturation s'écrit

$$I_{dsat} = I_p\left(1 - 3\frac{h_s^2}{a^2} + 2\frac{h_s^3}{a^3}\right) \qquad (7\text{-}43)$$

En explicitant h_s (Eq.7-35-a) et a (Eq.7-38), on obtient

$$I_{dsat} = I_p\left(1 - 3\left(\frac{V_{di} - V_g}{V_p}\right) + 2\left(\frac{V_{di} - V_g}{V_p}\right)^{3/2}\right) \qquad (7\text{-}44)$$

La caractéristique de transfert en régime de saturation, donnée par l'expression (7-44), est représentée sur la figure (7-4-b). Pour $V_g = V_{di} - V_p$, c'est-à-dire $V_g = V_{gco}$, le courant s'annule, le canal est obturé. Le courant de saturation atteindrait sa valeur maximum I_p pour $V_g = V_{di}$, en fait V_g doit rester négatif ou nul pour assurer une polarisation inverse de la jonction grille-canal, de sorte que I_{dsat} reste inférieur à I_p.

On obtient la tension de saturation V_{dsat} en écrivant dans l'expression (7-35-b) que $h_d = a$ pour $V_d = V_{dsat}$

$$a = \left(\frac{2\varepsilon}{eN_d}(V_{di} - V_g + V_{dsat})\right)^{1/2}$$

soit

$$V_{dsat} = \frac{e N_d a^2}{2\varepsilon} - V_{di} + V_g \qquad (7\text{-}45)$$

ou

$$V_{dsat} = V_p - V_{di} + V_g \qquad (7\text{-}46)$$

En remplaçant V_{di}-V_g par V_p-V_{dsat} dans l'expression (7-44), on obtient la loi de variation du courant de saturation en fonction de la tension de saturation

$$I_{dsat}=I_p\left(1-3\left(1-V_{dsat}/V_p\right)+2\left(1-V_{dsat}/V_p\right)^{3/2}\right) \tag{7-47}$$

Cette loi de variation est représentée par la courbe en traits discontinus sur la figure (7-4-a). Les expressions de I_{dsat} et V_{dsat} montrent que lorsque la polarisation V_g (négative pour un canal *n*) augmente, le courant de saturation et la tension de saturation diminuent.

La transconductance g_m et la conductance de drain g_d du transistor sont obtenues en explicitant $Q(h_s)$, $Q(h_d)$ et $Q(a)$ dans les expressions (7-26 et 30)

$$g_m=\frac{Z\mu e N_d}{L}(h_d-h_s) \qquad g_d=\frac{Z\mu e N_d}{L}(a-h_d)$$

Soit, compte tenu de (7-35-a et b)

$$g_m=\frac{Z\mu}{L}(2\varepsilon eN_d)^{1/2}\left((V_{di}-V_g+V_d)^{1/2}-(V_{di}-V_g)^{1/2}\right) \tag{7-48}$$

$$g_d=\frac{Z\mu}{L}(2\varepsilon eN_d)^{1/2}\left(V_p^{1/2}-(V_{di}-V_g+V_d)^{1/2}\right) \tag{7-49}$$

En régime de saturation, g_d=0 et la transconductance est donnée par

$$g_{ms}=\frac{Z\mu}{L}\left(2\varepsilon eN_d\right)^{1/2}\left(V_p^{1/2}-(V_{di}-V_g)^{1/2}\right) \tag{7-50}$$

Nous avons considéré le cas d'une distribution homogène de densité de charge. Les calculs développés avec d'autres types de distribution donnent des résultats tout à fait comparables et montrent que les caractéristiques du transistor sont assez peu sensibles à cette distribution.

Enfin nous avons supposé que la mobilité des porteurs était constante. En fait lorsque le champ longitudinal dans le canal devient important, la mobilité n'est plus constante avec la tension de drain, mais diminue quand V_d augmente. Ce phénomène doit être pris en considération dans l'étude des transistors à effet de champ à canal court. Nous le traiterons dans l'étude du transistor à effet de champ à barrière de Schottky.

7.2. Transistor à effet de champ à barrière de Schottky MESFET

7.2.1. *Structure et spécificité*

La structure du transistor à effet de champ à barrière de Schottky est représentée sur la figure (7-5). Cette structure est tout à fait comparable à celle du transistor à effet de champ à jonction, la jonction p^+n est remplacée par une barrière de Schottky métal-semiconducteur.

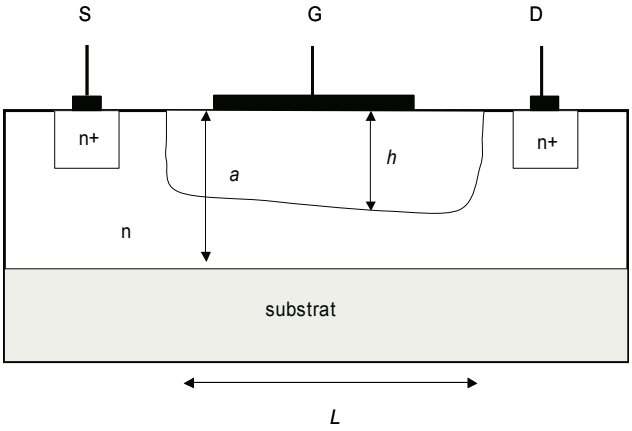


Figure 7-5: Schéma de principe du MESFET

Le principe de fonctionnement est tout à fait comparable à celui du JFET. La largeur du canal conducteur est modulée par la variation de la largeur *h* de la zone de charge d'espace de la diode Schottky.

En fait, le MESFET a été réalisé pour mettre à profit la rapidité de réponse de la diode Schottky résultant de l'absence de stockage de porteurs minoritaires. Des transistors à effet de champ à barrière de Schottky présentant une fréquence de coupure très élevée ont été réalisés. Pour mettre à profit dans les meilleures conditions la vitesse de réponse de la diode Schottky, la longueur du canal est en outre considérablement réduite par rapport au transistor à effet de champ à jonction, cette longueur *L* est inférieure au micron. Il en résulte que pour des tensions de polarisation drain-source comparables, le champ électrique longitudinal dans le canal est considérablement plus important dans le MESFET que dans le JFET. En conséquence l'hypothèse que nous avons utilisée dans l'étude du JFET qui consiste à supposer constante la mobilité des porteurs dans le canal, n'est plus justifiée dans l'étude du MESFET.

Dans le domaine des faibles champs électriques, les porteurs libres sont en équilibre thermodynamique avec le réseau et leur vitesse moyenne est

proportionnelle au champ électrique. En d'autres termes la mobilité des porteurs est indépendante du champ électrique et la vitesse de dérive s'écrit simplement

$$\mathbf{v}=\pm\mu\mathbf{E} \tag{7-51}$$

Lorsque le champ devient important, les interactions des porteurs avec les vibrations de réseau, phonons acoustiques d'abord phonons optiques ensuite, entraînent une diminution de mobilité des porteurs. Cette diminution de mobilité se traduit par une variation sous-linéaire de la vitesse de dérive des porteurs

$$\mathbf{v}=\pm\mu(E)\mathbf{E} \tag{7-52}$$

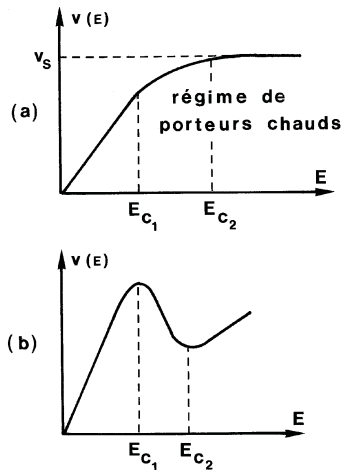


Figure 7-6 : Vitesse de dérive a) semiconducteur multivallé ; b) semiconducteur univallée

La loi de variation de la mobilité avec le champ électrique varie d'un matériau à l'autre, en fonction de la nature de la structure de bandes du semiconducteur. Dans les semiconducteurs multivallées comme le germanium et le silicium, les diffusions intervallées jouent un rôle important et on peut schématiquement définir deux valeurs critiques E_{c1} et E_{c2} du champ électrique. Pour $E<E_{c1}$, la mobilité des électrons est constante, pour $E_{c1}<E<E_{c2}$, on peut approximer la loi de variation de la mobilité par une expression de la forme

$$\mu(E)=\mu_o\left(E_{c1}/E\right)^{1/2} \tag{7-53}$$

où μ_o représente la mobilité à faible champ. La vitesse de dérive des porteurs s'écrit alors

$$v=\pm\mu_o\,(E_{c1}\,E)^{1/2} \tag{7-54}$$

Enfin au-delà de E_{c2}, l'expérience montre que, dans le silicium en particulier, la mobilité décroît linéairement avec l'inverse du champ électrique

$$\mu(E)=\mu_o\,E_{c2}/E \tag{7-55-a}$$

Il en résulte que la vitesse de dérive devient

$$v=\pm\mu_o\,E_{c2}=v_s \tag{7-55-b}$$

La vitesse de dérive des porteurs sature à une valeur v_s . La courbe représentant la vitesse de dérive des porteurs en fonction du champ électrique est schématisée sur la figure (7-6-a).

Dans le silicium, les valeurs des champs critiques sont, en V/cm

$E_{c1}\approx$2 500 V/cm et $E_{c2}\approx$15 000 V/cm pour les électrons
$E_{c1}\approx$7 000 V/cm et $E_{c2}\approx$20 000 V/cm pour les trous

Dans un canal de longueur L=10 μ, on obtient un champ E=10^4 V/cm pour une tension de polarisation drain-source de 10 Volts. Ainsi dans les transistors à canal court, la valeur de E_{c2} est souvent atteinte et dépassée.

Dans les semiconducteurs à gap direct comme InP et GaAs, la loi de variation de la vitesse de dérive avec le champ électrique est différente. La courbe de variation est schématisée sur la figure (7-6-b). On observe une variation linéaire pour des champs électriques inférieurs à une valeur critique E_{c1}, suivie d'une vitesse différentielle négative pour $E_{c1}<E<E_{c2}$. Ensuite la vitesse augmente à nouveau.

Dans InP, E_{c1}=5 000 V/cm et E_{c2}=11 000 V/cm . Dans GaAs, les valeurs sont légèrement plus faibles.

Nous allons étudier le fonctionnement du transistor en prenant en considération la saturation de vitesse.

7.2.2. *Courant de drain*

Le schéma de principe est tout à fait semblable à celui du transistor à effet de champ à jonction, la diode Schottky remplaçant la jonction pn. Considérons le diagramme représenté sur la figure (7-7).

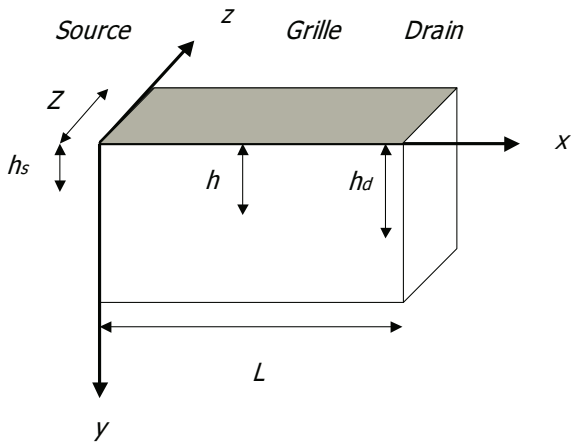


Figure 7-7 : Zone active du MESFET

La densité de courant en un point du canal conducteur s'écrit

$$J_x = \sigma(x,y,z)\, E_x \tag{7-56}$$

avec

$$\sigma(x,y,z) = \sigma(y) = \rho(y)\, \mu(E_x) \tag{7-57}$$

soit

$$J_x = \rho(y)\, \mu(E_x).E_x = \pm \rho(y).v_x(E_x) \tag{7-58}$$

Le signe + correspond à un canal p et le signe - à un canal n. Considérons un transistor à canal n, le courant de drain compté positivement dans le sens drain-source s'obtient en intégrant ($-j_x$) sur la section conductrice du canal.

$$I_d = -\int_s J_x\, ds = -\int_0^z \int_h^a J_x\, dz dy = Z v_x(E_x) \int_h^a \rho(y) dy$$

soit

$$I_d = Z v_x(E_x)\big(Q(a) - Q(h)\big) \tag{7-59}$$

Considérons une structure à dopage homogène caractérisée par une densité N_d de donneurs. Les quantités $Q(a)$ et $Q(h)$ et par suite le courant I_d s'écrivent respectivement

$$Q(a) = e\, N_d\, a \qquad\qquad Q(h) = e\, N_d\, h \tag{7-60}$$

$$I_d = Z\, e\, N_d v_x(E_x) \cdot (a - h) \tag{7-61}$$

Si on appelle μ_o la mobilité des électrons à faible champ et v_s leur vitesse de saturation à fort champ, on peut écrire la vitesse v_x en fonction du champ sous la forme

$$v_x(E_x) = \frac{-\mu_o E_x}{1 - \mu_o E_x / v_s} \tag{7-62}$$

Dans le domaine des faibles champs, $\mu_o E_x / v_s << 1$ et $v_x \approx v_o = -\mu_o E_x$. Lorsque le champ devient important $\mu_o E_x / v_s >> 1$ et $v_x \approx v_s$.

Le courant de drain s 'écrit

$$I_d = Z\, e\, N_d \frac{-\mu_o E_x}{1 - \mu_o E_x / v_s} (a - h) \tag{7-63}$$

La largeur de la zone de charge d'espace de la diode Schottky est donnée par la même loi que celle de la zone de charge d'espace de la jonction pn, ainsi les

quantités h_s et h_d sont données par les mêmes expressions que dans le JFET (7-35-a et b).

$$h_s=\left(\frac{2\varepsilon}{eN_d}(V_{di}-V_g)\right)^{1/2} \qquad h_d=\left(\frac{2\varepsilon}{eN_d}(V_{di}-V_g+V_d)\right)^{1/2}$$

En un point d'abscisse x, la largeur de la zone de déplétion est h, donné dans l'approximation du canal graduel par

$$h=\left(\frac{2\varepsilon}{eN_d}\left(V_{di}-V_g+V(x)\right)\right)^{1/2} \tag{7-64}$$

où $V(x)$ représente le potentiel du canal conducteur au point considéré. L'expression du courant de drain s'écrit

$$I_d=ZeN_d\frac{-\mu_o E_x}{1-\mu_o E_x/v_s}\left(a-\left(\frac{2\varepsilon}{eN_d}\left(V_{di}-V_g+V(x)\right)\right)^{1/2}\right) \tag{7-65}$$

En mettant a en facteur, on fait apparaître la tension de pincement V_p définie comme dans le JFET par

$$V_p=\frac{e\,N_d\,a^2}{2\,\varepsilon} \tag{7-66}$$

d'autre part le champ électrique est donné par $E_x=-dV(x)/dx$, ce qui permet d'écrire I_d sous la forme

$$I_d=ZeN_d\,a\frac{\mu_o\,dV(x)/dx}{1+\dfrac{\mu_o}{v_s}dV(x)/dx}\left(1-\left(\frac{V_{di}-V_g+V(x)}{V_p}\right)^{1/2}\right) \tag{7-67}$$

Posons $y^2=(V_{di}-V_g+V(x))/V_p \quad \Rightarrow \quad 2y\dfrac{dy}{dx}=\dfrac{1}{V_p}\dfrac{dV(x)}{dx} \quad \Rightarrow \quad \dfrac{dV(x)}{dx}=V_p\,2yy'$

En mettant v_s en facteur, I_d s'écrit

$$I_d=Z\,eN_d\,a\,v_s\frac{\dfrac{\mu_o}{v_s}V_p\,2yy'}{1+\dfrac{\mu_o}{v_s}V_p\,2yy'}(1-y) \tag{7-68}$$

Posons $z^* = \dfrac{\mu_o V_p}{v_s}$ *et* $I_o = Z\,eN_d\,a\,v_s$, le courant I_d s'écrit

$$I_d = I_o\,(1-y)\,2z^*\,yy'/(1+2z^*\,yy') \tag{7-69}$$

$$I_d\,(1+2z^*\,yy') = I_o\,(1-y)\,2z^*\,yy'$$

$$\frac{I_d}{I_o} = -\frac{I_d}{I_o}2z^*\,yy' + (1-y)\,2z^*\,yy'$$

$$= 2z^*\left(\left(1-\frac{I_d}{I_o}\right)y\frac{dy}{dx} - y^2\frac{dy}{dx}\right)$$

soit

$$\frac{I_d}{I_o}dx = 2z^*\left(\left(1-\frac{I_d}{I_o}\right)y - y^2\right)dy \tag{7-70}$$

On intègre de x=0 à x quelconque

$$\frac{I_d}{I_o}x = 2z^*\left(\frac{1}{2}\left(1-\frac{I_d}{I_o}\right)\left|y^2\right|_0^x - \left|\frac{1}{3}y^3\right|_0^x\right) \tag{7-71}$$

Soit t^2 et u^2 les valeurs de y^2 côté source et côté drain respectivement

$$t^2 = y^2\,(x=0) = (V_{di} - V_g)/V_p \tag{7-72-a}$$

$$u^2 = y^2\,(x=L) = (V_{di} - V_g + V_d)/V_p \tag{7-72-b}$$

L'expression (7-71) s'écrit

$$\frac{I_d}{I_o}x = \frac{1}{3}z^*\left(3\left(1-\frac{I_d}{I_o}\right)(y^2 - t^2) - 2(y^3 - t^3)\right) \tag{7-73}$$

côté drain c'est-à-dire en x=L, cette expression s'écrit

$$\frac{I_d}{I_o}L = \frac{1}{3}z^*\left(3\left(1-\frac{I_d}{I_o}\right)(u^2 - t^2) - 2(u^3 - t^3)\right) \tag{7-74}$$

ou, en posant $z = z^*/L = \mu_0 V_p / v_x\,L$

$$\frac{I_d}{I_o} = \frac{1}{3}z\left(3\left(1-\frac{I_d}{I_o}\right)(u^2 - t^2) - 2(u^3 - t^3)\right)$$

$$\frac{I_d}{I_o}\left(1+z(u^2-t^2)\right)=\frac{z}{3}\left(3(u^2-t^2)-2(u^3-t^3)\right) \tag{7-75}$$

Soit en posant $I_p=zI_o/3$

$$I_d = I_p \frac{3(u^2-t^2)-2(u^3-t^3)}{1+z(u^2-t^2)} \tag{7-76}$$

On peut maintenant expliciter les différents termes de l'expression (7-76)

$$I_p = \frac{1}{3} z I_o = \frac{1}{3} \frac{\mu_o V_p}{v_s L} Z e N_d a v_s = \frac{Z e N_d a \mu_o V_p}{3L} \tag{7-77}$$

$$u^2 - t^2 = V_d / V_p \tag{7-78}$$

$$u^3 - t^3 = \left(\frac{V_{di} - V_g + V_d}{V_p} \right)^{3/2} - \left(\frac{V_{di} - V_g}{V_p} \right)^{3/2} \tag{7-79}$$

Soit

$$I_d = \frac{Z e N_d a \mu_o V_p}{3L} \frac{3\frac{V_d}{V_p} - 2\left(\frac{V_{di} - V_g + V_d}{V_p} \right)^{3/2} + 2\left(\frac{V_{di} - V_g}{V_p} \right)^{3/2}}{1 + \frac{\mu_o V_p}{v_s L} \frac{V_d}{V_p}} \tag{7-80}$$

ou

$$I_d = \frac{Z e N_d a \mu_o}{L + \frac{\mu_o}{v_s} V_d} \left(V_d - \frac{2}{3\sqrt{V_p}} \left(V_{di} - V_g + V_d \right)^{3/2} + \frac{2}{3\sqrt{V_p}} \left(V_{di} - V_g \right)^{3/2} \right) \tag{7-81}$$

Posons

$$L^* = L + \frac{\mu_o}{v_s} V_d = L\left(1 + \frac{\mu_o E}{v_s} \right) = L\left(1 + \frac{v_o}{v_s} \right) \tag{7-82}$$

v_o représente la vitesse des électrons extrapolée linéairement en supposant la mobilité constante. Le paramètre g_o, défini dans l'étude du JFET (7-37) pour un transistor à canal long s'écrit ici

$$g_o^* = \frac{ZeN_d a\mu_o}{L^*} = g_o / \left(1 + v_0 / v_s\right) \tag{7-83}$$

L'expression (7-81) se met alors sous une forme comparable à l'expression (7-39) établie pour le JFET

$$I_d = g_o^* \left(V_d - \frac{2}{3\sqrt{V_p}} (V_{di} - V_g + V_d)^{3/2} + \frac{2}{3\sqrt{V_p}} (V_{di} - V_g)^{3/2} \right) \tag{7-84}$$

Si la vitesse de saturation n'existe pas $v_s \to \infty$, $L^* = L$ et $g_0^* = g_o$, on retrouve l'expression (7-39) établie pour le JFET à canal long.

7.2.3. *Tension de saturation - Courant de saturation*

Le courant de saturation correspond au maximum de I_d en fonction de V_d c'est-à-dire en fonction de u. Ainsi la tension de saturation V_{dsat} correspond à la valeur u_s de u qui annule la dérivée dI_d/du. Cette dérivée est obtenue à partir de l'expression (7-76)

$$\frac{1}{I_p} \frac{dI_d}{du} = \frac{\left(1 + z(u^2 - t^2)\right)\left(6u - 6u^2\right) - 2zu\left(3(u^2 - t^2) - 2(u^3 - t^3)\right)}{\left(1 + z(u^2 - t^2)\right)^2} \tag{7-85}$$

Le numérateur de cette expression s'écrit

$$N = 2u(3 - 3u - zu^3 + 3zut^2 - 2zt^3) \tag{7-86}$$

et s'annule pour $u = u_s$ tel que

$$u_s^3 + 3u_s\left(\frac{1}{z} - t^2\right) + 2t^3 - \frac{3}{z} = 0 \tag{7-87}$$

La tension de saturation est obtenue en portant la valeur de u_s dans l'expression (7-72-b). Il n'y a pas d'expression analytique de V_{dsat}.

Le courant de saturation est obtenu en fonction de la tension de saturation en portant l'expression de t^3 tirée de (7-87) dans l'expression (7-76). Soit

$$2t^3 = -u_s^3 - 3u_s\left(\frac{1}{z} - t^2\right) + \frac{3}{z} \tag{7-88}$$

et

$$I_{dsat} = I_p \frac{3u_s^2 - 3t^2 - 2u_s^3 - u_s^3 - 3u_s / z + 3u_s t^2 + 3/z}{1 + z(u^2 - t^2)} \tag{7-89}$$

Le développement de cette expression donne

$$I_{dsat} = I_p \frac{3(1-u_s)}{z} \tag{7-90}$$

ou, dans la mesure où $I_p = zI_0/3$

$$I_{dsat} = I_0 (I - u_s) \tag{7-91}$$

En explicitant I_o et u_s à partir de leurs expressions on obtient

$$I_{dsat} = ZeN_d av_s \left(1 - \left(\frac{V_{di} - V_g + V_{dsat}}{V_p} \right)^{1/2} \right) \tag{7-92}$$

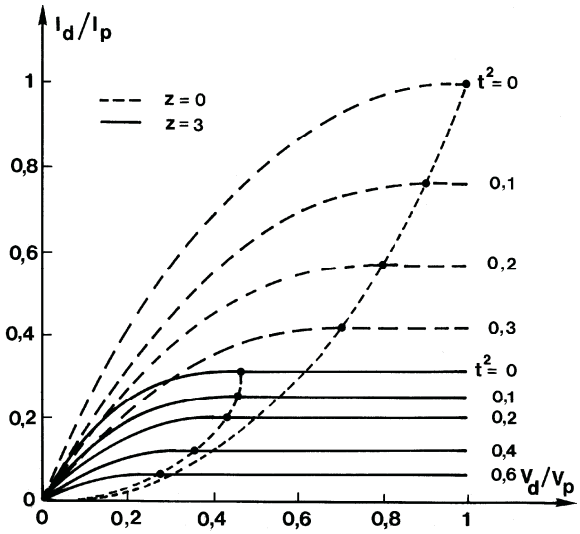


Figure 7-8 : Réseau de caractéristiques d'un FET AsGa

Cette expression donne le courant de saturation quand celui-ci est conditionné par la vitesse limite des porteurs. Afin de comparer ce résultat à celui obtenu dans le cas du JFET (avec μ=Cte), appelons I_{pin} le courant de saturation conditionné par le pincement du canal (7-47) et I_{lim} le courant de saturation conditionné par la vitesse limite des porteurs, soit

$$I_{pin}=I_p\left(1-3\left(1-\frac{V_{dsat}}{V_p}\right)-2\left(1-\frac{V_{dsat}}{V_p}\right)^{3/2}\right) \tag{7-93}$$

$$I_{\lim}=\frac{3}{z}I_p\left(1-\left(\frac{V_{di}-V_g+V_{dsat}}{V_p}\right)^{3/2}\right) \tag{7-94}$$

Rappelons que $z = \mu_o V_p/v_s L = (\mu_o V_p/L)/v_s = \mu_o E_p/v_s = v_o/v_s$ où v_o représente la vitesse qu'auraient les porteurs en régime de pincement si leur mobilité μ était constante et égale à μ_o.

Si la vitesse limite des porteurs est très élevée ou si la longueur du canal est importante, z est petit et $I_{lim}>>I_{pin}$, c'est alors le pincement du canal qui impose la saturation du courant de drain, ce dernier est donné par l'expression (7-93). Au contraire si le canal est très court c'est la vitesse limite des porteurs qui impose la saturation du courant de drain, celui-ci est alors donné par l'expression (7-94). La figure (7-8) représente le réseau de caractéristiques en unités réduites, dans le cas où la saturation du courant résulte du pincement du canal (z=0), et dans le cas où cette saturation résulte de la vitesse limite des porteurs (z=3). La figure montre clairement que l'existence d'une vitesse limite entraîne une réduction du courant de saturation et une tension de saturation inférieure à celle qui résulterait du pincement du canal.

7.2.4. *Transconductance*

La transconductance en régime de saturation est donnée par $g_{ms} = \partial I_{dsat}/\partial V_g$ ou

$$g_{ms}=\frac{\partial I_{dsat}}{\partial u_s}\,\frac{\partial u_s}{\partial t^2}\,\frac{\partial t^2}{\partial V_g} \tag{7-95}$$

Les expressions (7-91 et 72-a) donnent respectivement

$$\frac{\partial I_{dsat}}{\partial u_s}=-I_o \tag{7-96}$$

$$\frac{\partial t^2}{\partial V_g}=-\frac{1}{V_p} \tag{7-97}$$

D'autre part l'équation (7-87) s'écrit

$$u_s^3+3u_s\left(\frac{1}{z}-t^2\right)+2\,(t^2)^{3/2}-\frac{3}{z}=0$$

ce qui donne en différentiant

$$3u_s^2\,du_s+3\left(\frac{1}{z}-t^2\right)du_s-3u_s\,dt^2+2\frac{3}{2}t\,dt^2=0$$

soit

$$\frac{du_s}{dt^2}=\frac{u_s-t}{u_s^2-t^2+1/z} \tag{7-98}$$

En portant (7-96, 97 et 98) dans (7-95) on obtient

$$g_{ms}=\frac{I_o}{V_p}\frac{u_s-t}{u_s^2-t^2+1/z} \tag{7-99}$$

ou

$$g_{ms}=g_o\frac{u_s-t}{z(u_s^2-t^2)+1} \tag{7-100}$$

avec

$$g_o=\frac{I_o z}{V_p}=\frac{Z\,eN_d\,a\,v_s}{V_p}\frac{\mu_o V_p}{v_s L}=\frac{Z\,eN_d a\,\mu_o}{L} \tag{7-101}$$

Cette expression de g_o est identique à celle que nous avons établie dans le cas du JFET en supposant une mobilité constante (7-37). Dans le cas du canal long, z=0, d'autre part le régime de saturation résulte du pincement du canal, c'est-à-dire de la condition V_{di}-V_g+V_d=V_p ou en d'autres termes u_s=1. On retrouve la transconductance dans le cas d'un transistor à canal long en faisant z=0 et u_s=1 dans l'expression (7-100), soit

$$g_{ms}=g_o(1-t) \tag{7-102}$$

En explicitant g_o et t l'expression de g_{ms} s'écrit

$$g_{ms}=\frac{ZeN_d a\mu_o}{L}\left(1-\left(\frac{V_{di}-V_g}{V_p}\right)^{1/2}\right) \tag{7-103}$$

En mettant $V_p^{-1/2}$ en facteur et en l'explicitant à partir de l'expression (7-66), on obtient

$$g_{ms}=\frac{ZeN_d a\mu_o}{L}\left(\frac{2\varepsilon}{eN_d a^2}\right)^{1/2}\left(V_p^{1/2}-(V_{di}-V_g)^{1/2}\right)$$

soit

$$g_{ms}=\frac{Z\mu_o}{L}(2\varepsilon eN_d)^{1/2}\left(V_p^{1/2}-(V_{di}-V_g)^{1/2}\right) \tag{7-104}$$

On retrouve l'expression (7-50) établie pour un transistor à canal long en supposant une mobilité constante.

7.2.5. *Fréquence de coupure du transistor*

Le transistor à effet de champ à barrière de Schottky étant un composant haute fréquence, la fréquence de coupure est un paramètre caractéristique important. Cette fréquence est définie par la relation

$$f_c = \frac{g_m}{2\pi C_g} \tag{7-105}$$

où C_g est la capacité de grille du transistor donnée par $C_g = \left| dQ_{dep}/\mathrm{d}V_g \right|$ où Q_{dep} représente la charge de déplétion de la barrière de Schottky.

$$Q_{dep} = -ZeN_d \int_0^L h\,dx \tag{7-106}$$

h est la largeur de la zone de déplétion qui est donnée côté source, côté drain et en un point d'abscisse *x* quelconque, par les expressions respectives

$$h_s = \left(\frac{2\varepsilon}{eN_d}(V_{di} - V_g) \right)^{1/2} \tag{7-107-a}$$

$$h_d = \left(\frac{2\varepsilon}{eN_d}(V_{di} - V_g + V_d) \right)^{1/2} \tag{7-107-b}$$

$$h = \left(\frac{2\varepsilon}{eN_d}\left(V_{di} - V_g + V(x)\right) \right)^{1/2} \tag{7-107-c}$$

En utilisant la variable *y* déjà définie, $y = \left(\left(V_{di} - V_e + V(x)\right) / V_p \right)^{1/2}$ la variable *h* s'écrit $h = \left(2\varepsilon / eN_d\right)^{1/2} V_p^{1/2} y$, ce qui donne en explicitant V_p

$$h = ay \tag{7-108}$$

D'autre part $dx = \frac{dx}{dy} dy$. Pour *x*=0, *y*=*t* et pour *x*=*L*, *y*=*u*. La quantité de charge s'écrit

$$Q_{dep} = -ZeN_d a \int_t^u y \frac{dx}{dy} dy \tag{7-109}$$

La relation entre la variable *x* et la fonction *y* est donnée par l'expression (7-73) qui s'écrit en explicitant $z^* = zL$

$$x=L\frac{I_o}{I_d}\frac{z}{3}\left(3\left(1-\frac{I_d}{I_o}\right)\left(y^2-t^2\right)-2\left(y^3-t^3\right)\right) \tag{7-110}$$

de sorte que

$$\frac{dx}{dy}=L\frac{I_o}{I_d}\frac{z}{3}\left(6y\left(1-\frac{I_d}{I_o}\right)-6y^2\right) \tag{7-111}$$

Ainsi

$$Q_{dep}=-ZeN_d aL\frac{I_o}{I_d}\frac{z}{3}\left(6\left(1-\frac{I_d}{I_o}\right)\int_t^u y^2\,dy-6\int_t^u y^3\,dy\right)$$

soit

$$Q_{dep}=-ZeN_d aL\frac{I_o}{I_d}\frac{z}{3}\left(2\left(1-\frac{I_d}{I_o}\right)(u^3-t^3)-\frac{3}{2}(u^4-t^4)\right) \tag{7-112}$$

En régime de saturation $u=u_s$ et le courant de drain est donné par $I_d=I_{dsat}=I_o(1-u_s)$, de sorte que $I_d/I_o=1-u_s$. La charge de déplétion s'écrit alors

$$Q_{dep}=-ZeN_d aL\frac{z}{6}\frac{1}{1-u_s}\left(4u_s\left(u_s^3-t^3\right)-3\left(u_s^4-t^4\right)\right)$$

ou

$$Q_{dep}=-ZeN_d aL\frac{z}{6}\frac{1}{1-u_s}\left(u_s^4-4u_s t^3+3t^4\right) \tag{7-113}$$

La capacité de la grille est donnée par

$$C_g=\left|\frac{d\,Q_{dep}}{d\,V_g}\right|=\left|\frac{d\,Q_{dep}}{dt^2}\;\frac{d\,t^2}{d\,V_g}\right| \tag{7-114}$$

L'expression (7-72-a) donne $dt^2/dV_g = -1/V_p = -2\varepsilon/eN_d a^2$, d'où l'expression de C_g

$$C_g=\frac{Z\varepsilon L}{3a}z\frac{d}{dt^2}\left(\frac{u_s^4-4u_s t^2+3t^4}{1-u_s}\right) \tag{7-115}$$

Compte tenu de la relation (7-98), on obtient

$$C_g=\frac{Z\varepsilon L}{3a}z\frac{u_s-t}{1-u_s}\left(\frac{4(u_s^3-t^3)-3(u_s^4-t^4)}{(1-u_s)(u_s^2-t^2+1/z)}-6t\right) \tag{7-116}$$

Soit
$$C_g = C_{go}\frac{z}{3}\frac{u_s - t}{1-u_s}\left(\frac{4(u_s^3 - t^3) - 3(u_s^4 - t^4)}{(1-u_s)(u_s^2 - t^2 + 1/z)} - 6t\right) \qquad (7\text{-}117)$$

où $C_{go} = \varepsilon ZL/a$ représente la valeur de la capacité de grille lorsque la zone de charge d'espace occupe tout le canal.

La fréquence de coupure du transistor est donnée par l'expression (7-105) où g_m est donné par l'expression (7-100) et C_g par l'expression (7-116). On obtient en explicitant les différents paramètres

$$f_c = \frac{3}{\pi}\frac{v_s}{Lz}(1-u_s)^2 \Big/ \Big(4(u_s^3 - t^3) - 3(u_s^4 - t^4) - 6t(1-u_s)(u_s^2 - t^2 + 1/z)\Big) \qquad (7\text{-}118)$$

Le paramètre qui conditionne le fonctionnement du transistor est le paramètre $z = \mu_o V_p / v_s L$. La fréquence de coupure du transistor est d'autant plus élevée que z est petit c'est-à-dire que la vitesse de saturation est élevée. Dans cette voie les matériaux binaires GaAs, InP et les matériaux ternaires InAsP et InGaAs sont très prometteurs.

7.3. Transistor MESFET-GaAs

Dans l'étude du transistor MESFET que nous avons développée (Par. 7.2) nous avons vu comment la saturation de la vitesse des porteurs dans le canal modifiait les caractéristiques du transistor (Fig. 7-8). Mais ce modèle ne permet pas de rendre compte du fonctionnement du MESFET-GaAs à canal court et ceci pour deux raisons essentielles. La première raison est que pour calculer la variation de largeur du canal sous la double polarisation de grille et de drain nous avons intégré l'équation de Poisson à une dimension, la dimension transversale. Cette procédure n'est plus justifiée quand le canal devient très court en raison de la présence d'un gradient de potentiel très important le long du canal conducteur. L'équation de Poisson doit être intégrée à deux dimensions. La seconde raison concerne la loi de variation de vitesse que nous avons utilisée (Eq. 7-62). Cette loi traduit l'évolution monotone de la vitesse des électrons vers un régime de saturation qui se manifeste dans un semiconducteur à gap indirect comme le silicium. Cette loi ne rend pas compte du régime de survitesse qui se manifeste dans certains semiconducteurs à gap direct comme GaAs. Une analyse réaliste du fonctionnement du MESFET-GaAs à canal court doit intégrer ces spécificités (Chang C.S.-1989, Shih K.M.-1992).

La figure (7-9-a) représente la structure de la zone active du MESFET-GaAs. Le canal conducteur peut être divisé en plusieurs zones à l'intérieur desquelles le champ électrique longitudinal, et par suite la vitesse des porteurs, sont différents. Le champ électrique et la vitesse des électrons dans chacune des régions du canal sont représentés sur les figures (7-9-b et c).

La région L_a est une région à faible champ dans laquelle le champ électrique longitudinal est inférieur au seuil du régime non-linéaire $E<E_O$. La vitesse des porteurs dans cette région du canal est simplement donnée par $v=-\mu_O E$ où μ_O est la mobilité à faible champ.

Dans la région L_b le régime de vitesse est non-linéaire mais le champ électrique est inférieur au champ de seuil E_m , $E_O<E<E_m$, la vitesse des électrons atteint le régime de survitesse v_m.

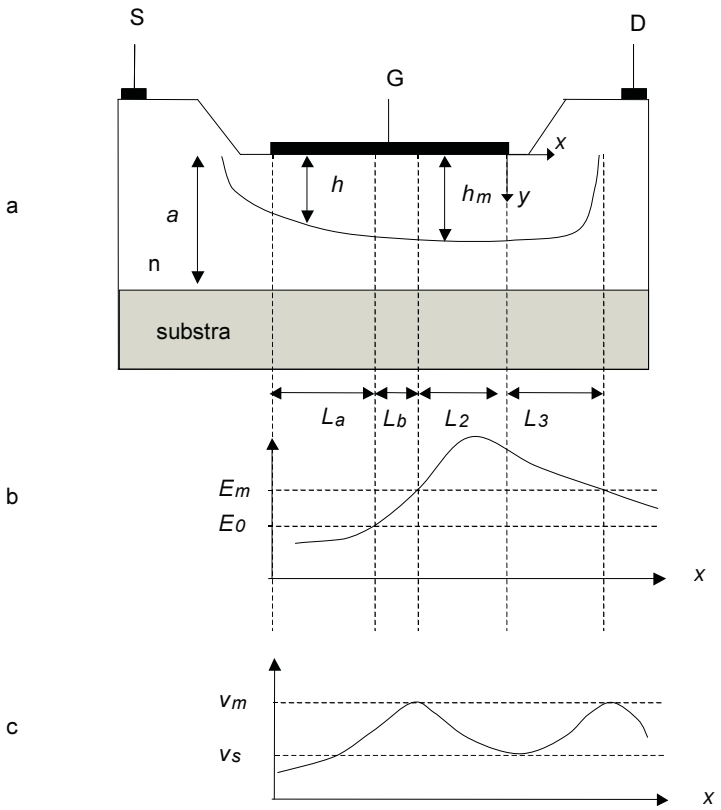


Figure 7-9 : MESFET GaAs

Les régions L_2 et L_3 sont des régions à fort champ où le champ électrique est supérieur au champ de seuil E_m. La région L_2 est située sous la grille alors que la région L_3 est située entre la grille et le drain. Dans ces régions, la section du canal conducteur est relativement constante. Ceci résulte de la compensation entre deux effets antagonistes. Dans la région L_2, le potentiel augmente suivant x de sorte que la polarisation du canal augmente suivant x et par suite la zone de charge d'espace devrait s'élargir progressivement. Mais en parallèle, le champ électrique lui aussi augmente suivant x, or ce champ étant ici supérieur à E_m, la vitesse des électrons diminue quand E augmente. Il en résulte une accumulation d'électrons qui repousse

vers la grille la zone de charge d'espace. Dans la région L_3, c'est la compensation entre d'un côté l'augmentation du potentiel du canal, et de l'autre l'éloignement progressif de la grille, qui maintient sensiblement constante la section du canal.

Suivant les valeurs des tensions de polarisation drain-source V_d et grille-source V_g, les régions L_a, L_b et L_2 occupent plus ou moins de place dans le canal.

7.3.1. *Régime linéaire*

Lorsque la tension de drain V_d est faible, le champ électrique dans le canal conducteur est partout inférieur à E_O, la région L_a occupe alors tout le canal et la vitesse des électrons est donnée par $v=-\mu_O E$. Le canal se comporte comme une résistance pure, l'hypothèse du canal graduel est justifiée, le champ électrique dans la zone de déplétion est alors purement transversal, c'est-à-dire suivant y. La relation entre la polarisation transversale V_g du transistor et la largeur h de la zone de charge d'espace, et par suite la largeur a-h du canal, est obtenue par intégration de l'équation de Poisson à une dimension

$$\frac{d^2V}{dy^2} = -\frac{eN_d(y)}{\varepsilon} \tag{7-119}$$

où V est le potentiel électrostatique, $N_d(y)$ le dopage du canal et ε la constante diélectrique de GaAs.

Il suffit d'intégrer deux fois en suivant la même procédure que pour le JFET (Par. 7.1.2.). Une première intégration entre h et y avec la condition $E_y=0$ en $y=h$ donne

$$\frac{dV}{dy} = \frac{1}{\varepsilon}(Q(h) - Q(y)) \tag{7-120}$$

avec

$$Q(h) = e\int_0^h N_d(y')dy' \quad Q(y) = e\int_0^y N_d(y')dy' \tag{7-121}$$

Une deuxième intégration entre $y=0$ et $y=h$ donne le potentiel V à la frontière de la zone de charge d'espace ($y=h$) en fonction de la polarisation transversale V_g du transistor

$$V = V_g - V_{di} + \frac{1}{\varepsilon}\int_o^h \big(Q(h) - Q(y)\big)dy \tag{7-122}$$

où V_{di} est la tension de diffusion de la diode Schottky. En intégrant par parties, l'expression (7-122) s'écrit

$$V = V_g - V_{di} + \frac{e}{\varepsilon}\int_o^h y N_d(y)\,dy \tag{7-123}$$

En fonction de la polarisation longitudinale du canal, la densité de courant en un point quelconque du canal est donnée par la loi d'Ohm

$$j_x = \sigma E_x = -\sigma \frac{dV}{dx} \tag{7-124}$$

La région L_a occupant toute la longueur du canal, la mobilité des électrons est indépendante de *x*, elle peut cependant être fonction de *y* dans la mesure où le dopage varie suivant *y*. Le courant s'écrit donc

$$j_x = -e\,N_d(y)\,\mu_0(y)\frac{dV}{dx} \tag{7-125}$$

Le courant de drain I_d , compté positivement dans le sens drain-source, est alors obtenue en intégrant $-j_x$ sur toute la section conductrice du canal.

$$I_d = -Z\int_h^a j_x\,dy \tag{7-126}$$

où Z est la largeur du canal conducteur dans la direction *z* , et *a-h* sa largeur dans la direction *y*. Soit

$$I_d = Z\,e\frac{dV}{dx}\int_h^a N_d(y)\;\mu_o(y)\,dy \tag{7-127}$$

En développant dV/dx sous la forme $dV/dh.\ dh/dx$, l'expression s'écrit

$$I_d dx = Z\,e\frac{dV}{dh}dh\int_h^a N_d(y)\;\mu_o(y)\,dy \tag{7-128}$$

En posant

$$N(h) = \int_o^h N_d(y)\;\mu_o(y)\;dy \tag{7-129-a}$$

$$N(a) = \int_o^a N_d(y)\;\mu_o(y)\;dy \tag{7-129-b}$$

L'expression (7-128) s'écrit

$$I_d dx = Ze\big(N(a) - N(h)\big)\frac{dV}{dh}dh \tag{7-130}$$

La dérivée de l'expression (7-123) donne $dV/dh=ehN_d(h)/\varepsilon$, ce qui permet d'écrire l'équation (7-130) sous la forme

$$I_d\,dx=\frac{Ze^2}{\varepsilon}\big(N(a)-N(h)\big)hN_d(h)dh \tag{7-131}$$

On obtient le courant de drain en intégrant sur toute la longueur du canal, c'est-à-dire de $x=0$ à $x=L$ sur la variable x , et de $h=h_S$ (en $x=0$) à $h=h_d$ (en $x=L$) sur la variable h. Le courant I_d est conservatif, d'autre part la région L_a occupant tout le canal la mobilité des électrons est indépendante de x. On obtient

$$I_d=\frac{Ze^2}{\varepsilon L}\int_{h_s}^{h_d}\big(N(a)-N(h)\big)hN_d(h)dh \tag{7-132}$$

Les largeurs h_S et h_d de la zone de déplétion côtés source et drain, sont obtenues à partir de l'équation (7-123) avec $V=r_SI_d$ côté source, et $V=V_d-r_dI_d$ côté drain. r_S et r_d sont les résistances de source et de drain incluant les résistances de contact et les résistances source-canal et canal-drain. Soit

$$\int_o^{h_s} yN_d(y)dy=\frac{\varepsilon}{e}\big(r_sI_d-V_g+V_{di}\big) \tag{7-133-a}$$

$$\int_o^{h_d} yN_d(y)dy=\frac{\varepsilon}{e}\big(V_d-r_dI_d-V_g+V_{di}\big) \tag{7-133-b}$$

Le régime linéaire existe tant que la région L_a occupe tout le canal, c'est-à-dire tant que le champ électrique dans le canal est inférieur à E_O. La limite est obtenue quand $E=E_O$ en $x=L$. La largeur de la zone de déplétion atteint alors une valeur maximum pour le régime linéaire h_a. Les limites du régime linéaire correspondent à un courant de drain maximum I_{da} donné par l'expression (7-127) avec $dV/dx=E_O$ et $h=h_a$, et une tension de drain maximum V_{da} donnée par l'expression (7-133-b) avec $h=h_a$ et $I_d=I_{da}$, soit

$$I_{da}=ZeE_o\int_{h_a}^{a}N_d(y)\mu_o(y)dy \tag{7-134}$$

$$V_{da}=V_g-V_{di}+r_dI_{da}+\frac{e}{\varepsilon}\int_o^{h_a}yN_d(y)dy \tag{7-135}$$

Si le dopage du canal est uniforme, $N_d(y)=N_d$ et par suite $\mu_O(y)=\mu_O$ sont constants. Le courant de drain (Eq. 7-132), s'écrit alors

$$I_d=\frac{Ze^2\mu_oN_d^2a^3}{6\varepsilon L}\left(\frac{3}{a^2}(h_d^2-h_s^2)-\frac{2}{a^3}(h_d^3-h_s^3)\right) \tag{7-136}$$

où h_s et h_d sont donnés par les expressions (7-133-a et b) qui s'écrivent pour un dopage uniforme

$$h_s=\left(\frac{2\varepsilon}{eN_d}(r_s I_d - V_g + V_{di})\right)^{1/2} \tag{7-137-a}$$

$$h_d=\left(\frac{2\varepsilon}{eN_d}(r_s I_d - V_g + V_{di})\right)^{1/2} \tag{7-137-b}$$

En faisant apparaître la tension de pincement V_p, le courant de saturation I_O et le courant de pincement I_p, déjà définis

$$V_p = e\,N_d\,a^2 \big/ 2\,\varepsilon \tag{7-138}$$

$$I_o=ZeN_d a v_s=ZeN_d a\mu_o E_c \tag{7-139}$$

$$I_p = V_p\,I_o \big/ 3\,L\,E_c \tag{7-140}$$

où v_S est la vitesse de saturation des électrons et E_C le champ critique, le courant de drain s'écrit

$$I_d=I_p\left(\frac{3}{a^2}(h_d^2-h_s^2)-\frac{2}{a^3}(h_d^3-h_s^3)\right) \tag{7-141}$$

avec

$$h_s=a\left(\frac{V_{di}-V_g+r_s I_d}{V_p}\right)^{1/2} \tag{7-142-a}$$

$$h_d=a\left(\frac{V_{di}-V_g+V_d-r_d I_d}{V_p}\right)^{1/2} \tag{7-142-b}$$

Le courant maximum du régime linéaire (Eq. 7-134) est alors donné par

$$I_{da} = S\,\sigma_o\,E_o \tag{7-143}$$

où $S=Z(a-h_a)$ est la section conductrice du canal en $x=L$, $\sigma_O=eN_d\mu_O$ est la conductivité du canal à faible champ et E_O est le champ électrique dans le canal en $x=L$.

La tension de drain maximum du régime linéaire (Eq. 7-135) est donnée par

$$V_{da}=V_g - V_{di} + r_d\,S\,\sigma_o\,E_o + \frac{eN_d}{2\varepsilon}h_a \tag{7-144}$$

7.3.2. *Régime sous-linéaire*

Quand la tension de drain augmente, le champ électrique dans le canal augmente au-delà de E_O. Le canal sous la grille présente alors deux régions, l'une de longueur L_a dans laquelle le champ est inférieur à E_O et où la vitesse des électrons est donnée par $v=-\mu_O E$, l'autre de longueur $L_b=L-L_a$ dans laquelle le régime de vitesse est sous-linéaire et où le champ est supérieur au champ E_O mais inférieur au champ de seuil E_m.

La longueur de la région L_a est donnée par l'expression (7-132) où h_d est remplacé par $h_a=h(x=L_a)$ déterminé par l'expression (7-135), soit

$$L_a=\frac{Ze^2}{\varepsilon I_d}\int_{h_s}^{h_a}\big(N(a)-N(h)\big)hN_d(h)dh \tag{7-145}$$

Dans la région L_b le courant de drain est donné par l'expression (7-127) où $\mu_O dV/dx$ est remplacé avec $u(E-E_O)=1$.

$$I_d=-Ze\frac{E}{\sqrt{1+\big((E-E_o)/E_c\big)^2}}\int_h^a N_d(y)\,\mu_o(y)\,dy \tag{7-146}$$

L'expression (7-146) est l'intégrale sur la section conductrice du canal de l'équation générique (8-31). Dans la région L_b où le champ électrique vérifie la condition $E_O<E<E_m$ cette équation se met sous la forme (8-32-b), soit

$$\frac{dV}{dx}=\frac{E_o(1+c^2)}{1+\sqrt{(1+c^2)\beta^2-c^2}} \tag{7-147}$$

où $c=E_c/E_O$ et β donné ici par

$$\beta=\frac{ZeE_c}{I_d}\int_h^a N_d(y)\,\mu_o(y)\,dy \tag{7-148}$$

dV/dx est donné par $dV/dx=dV/dh\,.\,dh/dx$ où dV/dh est dérivée de l'expression (7-123), soit

$$\frac{dV}{dx}=\frac{e\,h\,N_d(h)}{\varepsilon}\,\frac{dh}{dx} \tag{7-149}$$

L'équation (7-147) s'écrit alors

$$dx=\frac{e}{\varepsilon E_o(1+c^2)}\left(1+\sqrt{(1+c^2)\beta^2-c^2}\right)hN_d(h)\,dh \tag{7-150}$$

En intégrant sur x de $x=L_a$ à $x=L$ et sur h de $h=h_a$ à $h=h_L$, on obtient l'expression de la longueur de la région L_b,

$$L_b=L-L_a=\frac{e}{\varepsilon E_o(1+c^2)}\int_{h_a}^{h_L}\left(1+\sqrt{(1+c^2)\beta^2-c^2}\right)hN_d(h)dh \qquad (7\text{-}151)$$

En écrivant la somme $L=L_a+L_b$ (Eq. 7-145 et 151), on obtient la relation implicite courant-tension dans le régime sous-linéaire

$$L=\frac{Ze^2}{\varepsilon I_d}\int_{h_s}^{h_a}\left(N(a)-N(h)\right)hN_d(h)dh$$

$$+\frac{e}{\varepsilon E_0(1+c^2)}\int_{h_a}^{h_L}\left(1+\sqrt{(1+c^2)\beta^2-c^2}\right)hN_d(h)dh \qquad (7\text{-}152)$$

où h_s est donné par l'expression (7-133-a) et h_a par l'expression (7-135).

h_L est la longueur de la zone de déplétion en $x=L$. Le champ électrique dans la région L_b est inférieur au champ de seuil E_m et à la limite atteint E_m en $x=L$ où h atteint alors la valeur h_m. La vitesse des électrons est alors maximum et donnée par

$$v=\mu_o E_o\sqrt{1+c^2} \qquad (7\text{-}153)$$

La largeur maximum h_m de la zone de déplétion est alors donnée par l'expression (7-127) en remplaçant $\mu_0 dV/dx$ par l'expression de la vitesse (Eq. 7-153), soit

$$I_d=ZeE_o\sqrt{1+c^2}\int_{h_m}^{a}N_d(y)\mu_o(y)dy \qquad (7\text{-}154)$$

La tension maximum de drain correspondante est alors obtenue en faisant $h=h_m$ dans l'expression (7-123).

Si le canal est dopé de manière uniforme, $N_d(y)=N_d$ et $\mu_O(y)=\mu_O$, l'expression de L_a (Eq. 7-145) s'écrit

$$L_a=\frac{Ze^2\mu_o N_d^2 a^3}{6\varepsilon I_d}\left(\frac{3}{a^2}(h_a^2-h_s^2)-\frac{2}{a^3}(h_a^3-h_s^3)\right) \qquad (7\text{-}155)$$

ou, en faisant apparaître I_p (Eq. 7-140)

$$L_a=L\frac{I_p}{I_d}\left(\frac{3}{a^2}(h_a^2-h_s^2)-\frac{2}{a^3}(h_a^3-h_s^3)\right) \qquad (7\text{-}156)$$

L'expression de L_b (Eq. 7-151) s'écrit

$$L_b=\frac{eN_d}{\varepsilon E_o(1+c^2)}\left(\frac{1}{2}(h_L^2-h_a^2)+\int_{h_a}^{h_L}\sqrt{(1+c^2)\beta^2-c^2}\,dh\right) \tag{7-157}$$

où β est donné par l'expression (7-148) qui s'écrit

$$\beta=\frac{Ze\mu_o N_d E_c}{I_d}(a-h) \tag{7-158}$$

En faisant apparaître V_p (Eq. 7-138) et I_O (Eq. 7-139) l'expression de L_b s'écrit

$$L_b=\frac{V_p}{E_o(1+c^2)}\left(\frac{h_L^2}{a^2}-\frac{h_a^2}{a^2}+\frac{2}{a^2}\int_{h_a}^{h_L}\sqrt{(1+c^2)\beta^2-c^2}\,dh\right) \tag{7-159}$$

avec

$$\beta=\frac{I_o}{I_d}\frac{a-h}{a} \tag{7-160}$$

7.3.3. *Régime de saturation*

Quand la tension de drain augmente encore, le champ électrique dans le canal conducteur devient supérieur au champ de seuil E_m et un domaine à fort champ se forme. Le canal sous la grille peut alors être divisé en trois régions. La région $0<x<L_a$ de longueur L_a où le champ électrique est inférieur à E_O, la région $L_a<x<L_a+L_b$ de longueur L_b où $E_O<E<E_m$ et la région $L_a+L_b<x<L$ de longueur L_2 où $E>E_m$. Une région à fort champ de longueur L_3 existe aussi entre l'extrémité du canal et le drain.

Dans la région $0<x<L_a$ la relation $I(V)$ reste donnée par l'expression implicite (7-145), soit

$$L_a=\frac{Ze^2}{\varepsilon I_d}\int_{h_s}^{h_a}\big(N(a)-N(h)\big)hN_d(h)dh \tag{7-161}$$

Dans la région $L_a<x<L_a+L_b$, la relation $I(V)$ est donnée par l'expression (7-151) où h_L est remplacé par h_m. Cette région est limitée au point $x=L_a+L_b$ où $E=E_m$ et $h=h_m$ déterminé par l'expression (7-154)

$$L_b=\frac{e}{\varepsilon E_o(1+c^2)}\int_{h_a}^{h_m}\left(1+\sqrt{(1+c^2)\beta^2-c^2}\right)hN_d(h)dh \tag{7-162}$$

Dans la région à fort champ $x>L_a+L_b$, le calcul de l'effet de la polarisation transversale sur la largeur de la zone de déplétion nécessite l'intégration de l'équation de Poisson à deux dimensions car le champ dans la zone de déplétion n'est pas purement transversal. Pour intégrer cette équation nous utiliserons le système de coordonnées XY dont l'origine est l'extrémité du canal côté drain (Fig. 7-9), $X=x-L$, $Y=y$. L'équation de Poisson s'écrit

$$\frac{\partial^2 V}{\partial X^2}+\frac{\partial^2 V}{\partial Y^2}=-\frac{eN_d(Y)}{\varepsilon} \tag{7-163}$$

Les conditions aux limites sont les suivantes

$$V_2(-L_2,Y)=V_g-V_{di}+E_D Y-\frac{1}{\varepsilon}\int_o^Y Q(Y')dY' \tag{7-164-a}$$

$$V_2(X,0)=V_g-V_{di} \tag{7-164-b}$$

$$\frac{\partial V_2}{\partial Y}(X,h_m)=\frac{\partial V_3}{\partial Y}(X,h_m)=0 \tag{7-164-c}$$

$$\frac{\partial V_2}{\partial X}(-L_2,h_m)=\frac{\partial V_3}{\partial X}(L_3,h_m)=E_m \tag{7-164-d}$$

$$V_3(0,Y)=V_2(0,Y) \tag{7-164-e}$$

$$\frac{\partial V_3}{\partial X}(0,h_m)=\frac{\partial V_2}{\partial X}(0,h_m) \tag{7-164-f}$$

$$\frac{\partial V_3}{\partial X}(0,0)=\frac{\partial V_3}{\partial Y}(0,0)+E_{Fc} \tag{7-164-g}$$

La condition (7-164-a) traduit la continuité du potentiel à l'interface entre la région à faible champ (L_b) et la région à fort champ (L_2), c'est-à-dire en $X=-L_2$. Dans la région à faible champ l'intégration de l'équation de Poisson à une dimension donne une expression analogue à l'équation (7-122)

$$V(-L_2,Y)=V_g-V_{di}+\frac{1}{\varepsilon}\int_o^Y \big(Q(h_m)-Q(Y')\big)dY' \tag{7-165}$$

avec

$$Q(h_m)=e\int_o^{h_m} N_d(Y')dY' \tag{7-166-a}$$

$$Q(Y)=e\int_o^{Y} N_d(Y')dY' \tag{7-166-b}$$

En posant $E_D=Q(h_m)/\varepsilon$, on obtient la condition (7-164-a).

La condition (7-164-b) traduit le fait que la surface sous la grille est une équipotentielle fixée par la polarisation V_g de la grille et la tension de diffusion V_{di} de la diode Schottky.

La condition (7-164-c) traduit le fait que le champ électrique dans le canal est longitudinal, pour la simple raison que le courant de drain est longitudinal. Il en résulte que la composante transversale du champ dans la zone de déplétion s'annule en $X=h_m$.

La condition (7-164-d) suppose que le champ électrique longitudinal dans le canal conducteur, est égal au champ de seuil E_m à chacune des extrémités de la région à fort champ.

Les conditions (7-164-e et f) traduisent les continuités du potentiel et de la composante longitudinale du champ à l'interface entre les régions L_2 et L_3.

Enfin la condition (7-164-g) traduit la présence de charges d'interface entre le métal de la grille et la couche d'oxyde de passivation.

En raison de la présence de charges qui existent à la surface de la grille, les composantes E_x et E_y du champ électrique ne sont pas continues en X=0 et Y=0. Dans le métal E_x=0 et E_y=0, dans la zone de déplétion du semiconducteur $E_x\neq 0$, et $E_y\neq 0$, la discontinuité du flux de champ électrique est égale à la densité superficielle de charges. Deux types de charges existent à l'interface grille-semiconducteur, les premières sont induites par le champ électrique et se localisent à la surface du métal, les secondes résultent d'états d'interface et se localisent dans l'isolant. Si les premières existaient seules la discontinuité du champ électrique serait la même suivant X et Y de sorte que la condition s'écrirait : $\partial V/\partial X\,(X=0,Y=0)=\partial V/\partial Y\,(X=0,Y=0)$. La couche de passivation n'existe que suivant la direction X de sorte que les charges fixes P_{Fc} dans l'oxyde n'affectent que la composante E_x du champ électrique. La relation entre les composantes E_x et E_y est alors donnée par la condition (7-164-g) où $E_{Fc}=\rho_{Fc}/\varepsilon$ est la discontinuité de E_x résultant des charges localisées dans l'isolant de passivation.

L'intégration de l'équation de Poisson (7-163) avec les conditions aux limites (7-164-a...g) donne (Chang C.S.-1989)

$$V(L_3)-V(-L_2)=\frac{1}{3}E_m L_3\left(2e^{\lambda_2}+1\right)+\left(\frac{2}{\pi}h_m+\frac{1}{3}L_3\right)E_m e^{-\lambda_3}\,sh(\lambda_2) \qquad (7\text{-}167)$$

avec $\lambda_2=\pi L_2/2h_m$ *et* $\lambda_3=\pi L_3/2h_m$. Les potentiels en X=-L_2 (Eq. 7-122 avec $h=h_m$) et en $X=L_3$ sont donnés par

$$V(-L_2)=V_g-V_{di}+\frac{h_m}{\varepsilon}Q(h_m)-\frac{1}{\varepsilon}\int_o^{h_m}Q(y)dy \qquad (7\text{-}168)$$

$$V(L_3)=V_d-r_d I_d \qquad (7\text{-}169)$$

La longueur L_3 qui apparaît dans l'expression (7-167) est obtenue à partir de la condition (7-164-d)

$$L_3^2 = h_m^2 \, E_m \frac{e^{\lambda_2} - 1 - e^{-\lambda_3} sh(\lambda_2)}{E_d + E_{Fc} - E_m ch(\lambda_2)} \tag{7-170}$$

En combinant les équations (7-161, 162 et 170) avec l'équation (7-167) et la relation $L=L_a+L_b+L_3$ on obtient une expression implicite du courant de drain I_d en fonction des tensions de grille et de drain V_g et V_d.

EXERCICES DU CHAPITRE 7

- **Exercice 1 : Bipolaire et JFET**

Expliquez la différence de fonctionnement entre un transistor bipolaire et un transistor JFET. Pourquoi le bruit du JFET est-il plus faible que celui du bipolaire.

- **Exercice 2 : Fonctionnement du JFET**

Expliquez pourquoi un courant peut circuler dans la canal d'un JFET au-delà de la tesnion de pincement, tout le volume du dispositif côté drain étant alors dépourvu de porteurs.

- **Exercice 3 : Utilisation du JFET**

Imaginez comment on peut réaliser un amplificateur avec un JFET et une résistance. Calculez le gain en tension.

- **Exercice 4 : Transistor MESFET**

Un transistor MESFET arséniure de gallium a une électrode de grille en or. La barrière est de 0, 89 V et le dopage n est de $2\ 10^{15}/\text{cm}^3$. Le canal a une épaisseur de 0,6 micron. Calculez la tension de pincement et la tension V_{di}.

- **Exercice 5 : Transistor MESFET**

Un transistor GaAs a un canal n dopé à $2\ 10^{15}/\text{cm}^3$. Le canal a un micron de long et 0,5 micron de profondeur. La largeur est 50 microns. La barrière métal-semiconducteur est de 0,8 V. La mobilité des électrons est 4500 cm2/V-s.
Calculez la tension de pincement et le courant de saturation pour une tension de grille nulle. Calculez la tension de seuil.

- **Exercice 6 : Caractéristiques du MESFET**

Comparez à courant égal la transconductance du JFET et du transistor bipolaire.
Que peut-on en déduire pour le bruit d'un étage amplificateur ?

8. Le transistor MOSFET et son évolution

Le but de ce chapitre est d'expliquer le fonctionnement du dispositif de base de la micro-électronique actuelle : le transistor MOS. La très grande majorité des circuits intégrés logiques et analogiques sont fabriqués en combinant des transistors de type MOS (*Metal Oxyde Semiconductor*) et des composants passifs (principalement des résistances et des condensateurs). Le fonctionnement du transistor MOS est simple dans le principe mais assez complexe quand on rentre dans le détail et quand on s'intéresse à son optimisation. Les résultats de ce chapitre permettent d'établir des modèles électriques qui seront utilisés dans la conception des circuits. La fabrication du transistor MOS et des circuits intégrés à base de MOS est décrite dans le chapitre 11 de cet ouvrage. Le chapitre est organisé de la manière suivante.

8.1. Principe de base et brève histoire
8.2. Comment la structure MOS se modifie
8.3. Le modèle canal long
8.4. Le modèle canal court
8.5. Le fonctionnement dynamique du MOS
8.6. Les modèles du transistor MOS
8.7. Le bruit du MOSFET
8.7. Applications du MOSFET

8.1. Principe de base et historique

L'idée initiale fut de contrôler la conduction d'un fil par une grille de la même manière qu'une grille contrôle le courant émis par le filament d'une triode et qu'un robinet contrôle le débit de l'eau. Un fil isolant ne pouvait convenir car le courant traversant un isolant est nul. Un fil conducteur semblait peu utilisable car le champ ne pénètre pas dans un conducteur. Il était donc difficile d'imaginer dans ce cas un mode de contrôle. Le semiconducteur apparut alors comme le bon matériau puisqu'il offrait les deux propriétés de base : possibilité de laisser passer un courant et pénétration du champ à l'intérieur. Le premier dispositif a été imaginé et breveté par Lilienfield en 1933. Ces dispositifs étaient du type MOS à appauvrissement. Le principe du MOS à appauvrissement est de rejeter les porteurs hors du canal de conduction en appliquant une tension. Ce n'est qu'en 1948 que Bardeen eut l'idée de reprendre ce principe mais en créant une couche d'inversion, c'est-à-dire formée par les porteurs minoritaires et cette fois en augmentant la densité de porteurs par application d'une tension.

Etudions, dans un premier temps, le fonctionnement du MOSFET à appauvrissement. Le dispositif est un morceau de semi-conducteur appelé « *bulk* » dans lequel sont créées deux régions fortement dopées de type *n* qui jouent le rôle de réservoirs d'électrons. Ce sont la *source* et le *drain*. Dans le MOS à appauvrissement une zone supplémentaire de type *n* est créée entre source et drain. On suppose, pour simplifier, que la face arrière du semi-conducteur et la source sont reliées électriquement. Une tension positive V_{DS} entre drain et source a pour effet de faire passer un courant de conduction (courant de dérive) à la condition qu'il y ait des électrons dans la zone de type *n*. Quand on applique une tension nulle sur la *grille*, les électrons sont en grand nombre puisqu'ils proviennent de la zone dopée *n*. Un courant circule de la source vers le drain. Quand une tension négative est appliquée sur la grille, elle attire les trous du semi-conducteur qui se recombinent aux électrons du canal si bien que la densité d'électrons de conduction diminue. En conséquence, le courant diminue. Remarquons que les jonctions du dispositif sont soit polarisées en inverse (jonction bulk-drain) soit non polarisées (jonction source-bulk). Les courants de jonction sont très faibles. Le courant de drain est uniquement dû à la conduction dans le canal. Il est contrôlé par la tension de grille.

Dans le MOS à enrichissement, il n'y a plus de zone dopée servant de canal de conduction. Les trous du matériau de base ne peuvent donner lieu à un courant puisque les deux jonctions source-bulk et bulk-drain sont respectivement non polarisée et polarisée en inverse. Seuls, les électrons peuvent créer un courant dans ce type de dispositif. Quand une tension nulle est appliquée sur la grille, les électrons ne sont pas injectés dans le semi-conducteur et aucun courant ne circule de la source vers le drain. Quand une tension positive est appliquée sur la grille, elle attire des électrons fournis par la source et le drain et un courant peut alors s'établir.

Les transistors MOS à appauvrissement ont été progressivement abandonnés et les transistors à enrichissement se sont imposés dans l'industrie microélectronique. La fabrication est en effet plus simple. De plus, les MOS à enrichissement permettent de réaliser des circuits consommant très peu ce qui a donné lieu à la technologie CMOS avec laquelle on réalise aujourd'hui la majorité des circuits intégrés. Les dispositifs à canal *n* ne sont pas les seuls à être fabriqués. Des dispositifs équivalents peuvent être réalisés en jouant sur la conduction des trous. Le canal est alors de type *p*. Une tension négative de grille est dans ce cas appliquée pour enrichir le canal. Ces deux types de transistors à enrichissement, MOS canal *n* et MOS canal *p*, sont les deux briques de base de la technologie CMOS.

Etudions maintenant le fonctionnement du transistor à enrichissement. La figure 8.1 reprend le fonctionnement du MOS à enrichissement et explique comment la conduction varie avec les tensions appliquées.

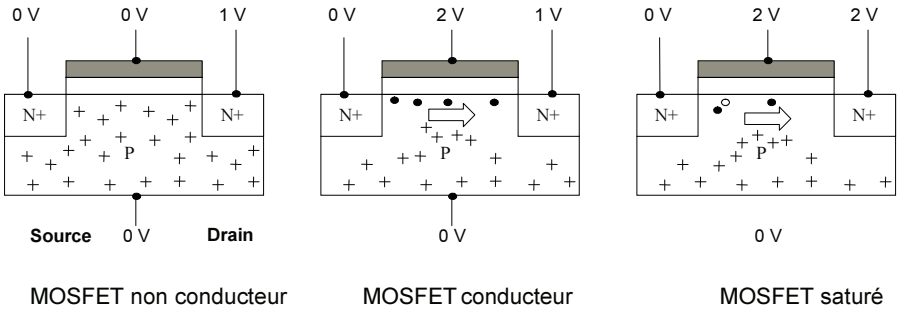


Figure 8-1 : Fonctionnement du FET canal n à enrichissement

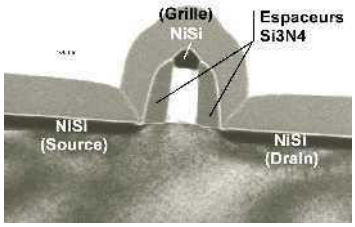
Dans une première étape, la grille est à tension nulle et le drain est polarisé positivement. Comme il a été vu dans le chapitre 5, aucun électron n'est confiné à l'interface semiconducteur isolant et de ce fait il n'y a pas de courant. Le MOSFET est non conducteur.

Quand une tension positive est appliquée sur la grille, une couche d'inversion se forme et on passe du régime de *faible inversion* au régime de *forte inversion* au fur et à mesure que la tension de grille augmente. Ce phénomène apparaît quand la tension de grille est supérieure à une tension dite de seuil de l'ordre de 0,4 V. Les électrons sont fournis par la source et un courant peut circuler de la source vers le drain sous l'effet du champ électrique présent dans le dispositif. Le MOSFET est conducteur.

Si maintenant, la tension de grille restant constante, on augmente la tension du drain, la différence de potentiel entre la grille et la zone du canal proche du

drain peut devenir inférieure à la tension de seuil. La couche d'inversion est alors nulle en bout de canal. Ce dernier régime est appelé régime de saturation. On peut considérer le canal comme la mise en série d'une zone de conduction faiblement résistive et d'une jonction polarisée en inverse. Toute augmentation supplémentaire de la tension de drain se traduit par une augmentation de la tension aux bornes de la jonction *pn* en bout de canal et aucune augmentation de tension ne peut alors se manifester aux bornes du canal de conduction. Le courant de drain reste donc constant.

La géométrie décrite précédemment est en fait très simplifiée. La géométrie réelle est représentée figure 8.2. La longueur du canal est très faible (moins de 90 nm pour les technologies les plus avancées), l'épaisseur de l'oxyde est de quelques nanomètres pour que l'influence de la grille soit maximale. La largeur du transistor est définie par les concepteurs mais inférieure au micron pour les technologies numériques.



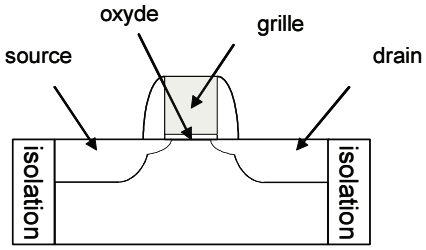


Figure 8.2 : Le MOSFET réel dans une technologie avancée

Remarquons la forme particulière des régions de drain et de source au niveau du canal. Cette forme adoucie du dopage permet de limiter le champ électrique dans ces régions.

8.2. Comment la structure MOS se modifie

Etudions plus en détail le fonctionnement du transistor. Le fonctionnement du transistor est compris quand on peut exprimer le courant sortant du drain en

fonction des tensions appliquées. On distingue trois régimes : le *régime statique*, le *régime quasi-statique* et le *régime dynamique*.

Dans le *régime statique*, les électrons sortant du drain sont exactement compensés par ceux qui entrent dans la source. Les charges sont en mouvement dans le canal mais tout se passe comme si elles étaient fixes pour le calcul des potentiels puisque la charge totale des électrons du canal est constante. Ce régime est également appelé régime continu. Le courant dans le canal est constant en fonction de la position le long du canal.

Dans le *régime quasi-statique*, les tensions appliquées varient mais suffisamment lentement pour que les électrons sortant soient compensés par les électrons entrants. La relation donnant le courant en fonction des tensions appliquées est donc la même que dans le cas précédent mais les tensions sont des fonctions du temps.

Dans le *régime dynamique*, les charges ne sont pas intégralement compensées et le courant ne peut plus être considéré comme constant dans le canal de conduction. Il faut alors écrire une relation locale de conservation du courant et résoudre le problème. Ce cas plus difficile sera traité dans le paragraphe 5.

Le cas statique peut sembler identique au système MOS décrit dans le chapitre 5. Il y a cependant une différence fondamentale dans l'origine des électrons qui constituent le canal de conduction. Dans la structure MOS, les électrons viennent du substrat et sont attirés par la grille. Dans le transistor MOS, les électrons sont fournis par la source. Cette différence se traduit également par une différence dans la valeur des potentiels chimiques ce qui conduit à des expressions différentes des concentrations de porteurs. La figure 8.3 représente les deux structures : la structure MOS étudiée dans le chapitre 5 et le transistor MOS étudié dans ce chapitre.

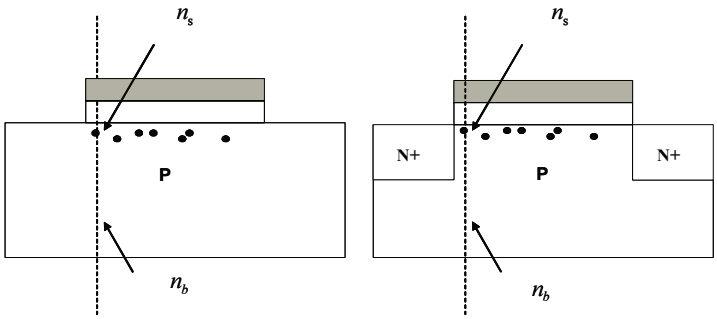


Figure 8-3 : Modification des concentrations

Si on examine la concentration des électrons au voisinage de la source (ligne en pointillés sur la figure) on peut écrire dans le cas de la simple structure MOS

$$n_s = n_i e^{\frac{E_F - E_{is}}{k T}}$$

Dans cette expression, le potentiel chimique E_F est supposé constant dans le dispositif car il n'y a pas transfert de charges entre régions. La concentration n_b des électrons en profondeur dans la région dopée p s'écrit

$$n_b = n_i e^{\frac{E_F - E_{ib}}{k T}}$$

Dans le dispositif MOSFET le potentiel chimique ne peut être considéré comme constant pour les électrons car il y a transfert d'électrons de la source au canal et formation d'un courant. On écrit alors en appelant E_{Fn} le potentiel chimique pour les électrons maintenant fonction de la position x dans le canal

$$n_s = n_i e^{\frac{E_{Fn}(x) - E_{is}(x)}{k T}}$$

En fait, la grandeur E_{Fn} n'est plus un potentiel chimique pour l'ensemble des porteurs du dispositif, électrons et trous. On admet que chaque population est en équilibre avec son propre potentiel chimique. La grandeur E_{Fn} est un pseudo potentiel chimique défini pour les électrons. En profondeur, on suppose l'équilibre et le potentiel chimique des électrons est égal à celui des trous. Il est noté E_F. Le niveau E_{is} dépend également de la position puisque le potentiel et donc l'énergie potentielle peut varier le long du canal.

$$n_b = n_i e^{\frac{E_F - E_{ib}(x)}{kT}}$$

Fort heureusement, il y a une relation simple entre le potentiel chimique dans le canal au niveau de la source et le potentiel chimique en profondeur dans le dispositif comme il a été vu dans le chapitre 1.

$$E_{F_n}(0) - E_F = -eV_{SB}$$

V_{SB} est la tension extérieure appliquée entre la source et le substrat.

Le potentiel chimique dans le canal au niveau de la source est en effet égal à celui qui règne dans la source car il y a échange d'électrons entre source et canal avec la formation d'un état d'équilibre local. On écrit alors

$$\frac{n_s}{n_b} = e^{\left(E_{Fn}(x) - E_F - E_{is}(x) + E_{ib}(x) \right) / k T}$$

La différence d'énergie E_{ib}-E_{is} est la différence de potentiel V_s entre la zone profonde et le canal, multipliée par l'inverse de la charge de l'électron. On obtient donc

$$\frac{n_s}{n_b} = e^{\frac{eV_s(x) - eV_{CB}(x)}{k\,T}} \tag{8-1}$$

On a défini $V_{CB}(x)$ par

$$V_{CB}(x) = -\frac{E_F - E_{FM}(x)}{e} \tag{8-2}$$

Dans la structure MOS, la relation s'écrivait simplement

$$\frac{n_s}{n_b} = e^{\frac{+eV_s}{k\,T}}$$

Les relations sont inchangées pour les trous car il n'y a pas d'échange avec la source. Ces considérations s'appliquent également pour exprimer la densité d'électrons non pas en surface mais en profondeur dans le silicium (annexe 5).

La figure 5-9 montre que la variable *y* indique la profondeur et la variable *x* la position longitudinale par rapport à la source.

Pour calculer les densités de charge au niveau de la source, il suffit donc de remplacer *V(y)* par *V(0,y)* $-V_{SB}$ dans l'équation générale.

$$\left(\frac{dV}{dy}\right)^2 = \frac{2eN_A}{\varepsilon_s}\left(\varphi_t e^{-\frac{V(0,y)}{\varphi_t}} + V(0,y) - \varphi_t + e^{-\frac{2\varphi_F}{\varphi_t}}\left(\varphi_t e^{\frac{V(0,y)-V_{SB}}{\varphi_t}} - V(0,y) - \varphi_t\right)\right) \tag{8-3}$$

A cette différence près, toutes les relations établies dans le chapitre 5 restent valables. On peut ainsi écrire la relation donnant le potentiel de surface (*y=0*) au niveau de la source

$$V_{GB} = \phi_{MS} - \frac{Q'_0}{C'_{OX}} + V_s(0) + \frac{1}{C'_{OX}}\sqrt{2eN_A\varepsilon_s}\sqrt{V_s(0) + \phi_t e^{\frac{V_s(0) - 2\phi_F - V_{SB}}{\phi_t}}} \tag{8-4}$$

Si on fait le même raisonnement au niveau du drain, le potentiel chimique $E_{Fn}(L)$ des électrons du canal au niveau du drain est alors :

$$E_{Fn}(L) - E_F = -eV_{DB}$$

Remarquons que le potentiel chimique des électrons au niveau du drain a une valeur différente de celui au niveau de la source. Il varie en fait tout le long du canal. Cela n'est pas étonnant si on reprend les équations générales donnant concentrations et courant d'une population d'électrons. Il ne faut pas confondre l'énergie E_i et le champ électrique E .

$$n(x) = n_i e^{\frac{E_{Fn}(x) - E_i(x)}{k\,T}}$$

$$J_n = e\mu_n n E_x + eD_n \frac{dn}{dx}$$

On écrit alors

$$\frac{dn}{dx} = \frac{n}{k\,T}\left(\frac{dE_{Fn}}{dx} - \frac{dE_i}{dx}\right)$$

$$E_x = \frac{1}{e}\frac{dE_i}{dx}$$

On en déduit

$$J_n = \mu_n\, n \frac{dE_{Fn}}{dx} \tag{8-5}$$

On vérifie bien que le passage d'un courant s'accompagne d'une variation du pseudo-potentiel chimique de la population d'électrons. Cette propriété explique aussi que le pseudo-potentiel chimique des électrons ne varie pas avec la profondeur y puisque l'on considère que le courant transverse est nul. Une relation équivalente peut s'écrire pour une population de trous. La relation donnant le potentiel de surface s'écrit donc au niveau du drain

$$V_{GB} = \phi_{MS} - \frac{Q'_0}{C'_{OX}} + V_s(L) + \frac{1}{C'_{OX}}\sqrt{2eN_A\varepsilon_s}\sqrt{V_s(L) + \phi_t e^{\frac{V_s(L) - 2\phi_F - V_{DB}}{\phi_t}}} \tag{8-6}$$

Nous allons maintenant chercher à exprimer le courant fourni par le transistor en régime continu en fonction des tensions appliquées. Le point de départ est l'écriture de la densité de courant dans un semi-conducteur comme elle a été calculée dans le chapitre 1.

$$J_n = e\mu_n n E_x + eD_n \frac{dn}{dx} \tag{8-7}$$

On ne considère que la partie longitudinale du courant correspondant à l'axe des *x.* Dans les autres dimensions on admet que le courant est nul. Cette hypothèse n'est vraie que pour un canal long devant les dimensions transverses. On considère donc que le courant transverse de diffusion compense exactement le courant transverse de dérive comme le montre la figure 8-4.

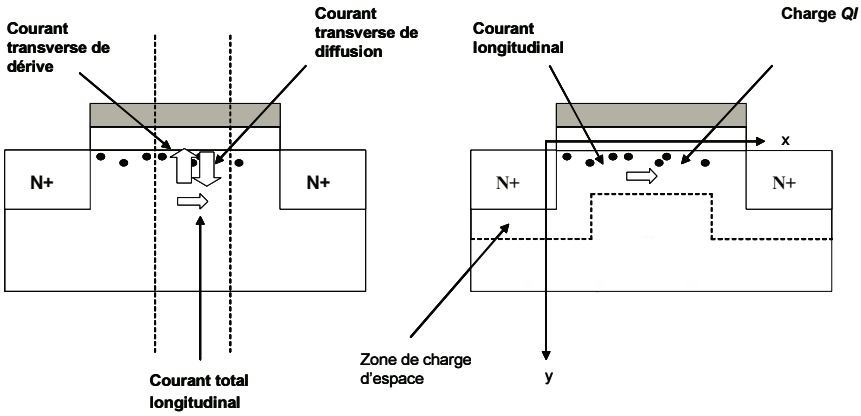


Figure 8-4 : Le courant dans un transistor

Si on appelle Q'_I la charge des électrons du canal par unité de surface, W et L la largeur du dispositif et la longueur du canal comme le montre la figure 8.5, le courant traversant le dispositif s'écrit alors

$$I_D = \mu_n W\left(-Q'_I\right)\frac{dV_s}{dx} + \mu_n W\,\phi_t\,\frac{dQ'_I}{dx} \tag{8-8}$$

Le passage de J_D (densité de courant) à I_D (courant) n'est pas tout à fait évident. Il amène à remplacer n par WQ'_I comme le lecteur pourra le vérifier.

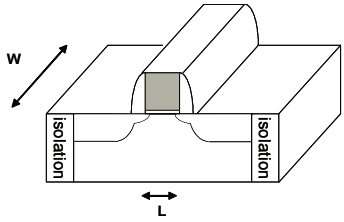


Figure 8.5 : Le transistor 3D et ses dimensions

On se place maintenant en régime stationnaire et on suppose que le courant est constant le long du canal. On peut appliquer la règle de conservation du flux de

courant à travers le canal et ignorer le courant de déplacement. Comme le courant est supposé constant le long de l'axe des *x*, cette équation s'intègre très facilement de la source (x=0) au drain (x=L). On obtient alors

$$\int_0^L I_D dx = \mu_n W \int_{V_s(0)}^{V_s(L)} \left(-Q'_I\right) \frac{dV_s}{dx} dx + \mu_n W \phi_t \int_{Q'_I(0)}^{Q'_I(L)} \frac{dQ'_I}{dx} dx = I_D \, L \tag{8.9}$$

Pour aller plus loin dans le calcul, il faut exprimer la charge par unité de surface et le potentiel de surface en fonction des tensions appliquées. Ce calcul sera effectué dans le paragraphe suivant dans deux cas : le régime de forte inversion et le régime de faible inversion.

8.3. Le modèle canal long

8.3.1. *Le régime de forte inversion*

Reprenons l'équation 8.1 en explicitant les valeurs de la charge Q'_I et du potentiel de surface $V_s(x)$. Les résultats sont tirés du paragraphe 2 et du chapitre 5.

$$Q'_G + Q'_0 + Q'_I + Q'_B = 0$$

$$V_{GB} = V_{OX}(x) + V_s(x) + \phi_{MS}$$

$$Q'_G = C'_{OX} V_{OX}(x)$$

$$Q'_B = -\sqrt{2eN_A\varepsilon_s}\sqrt{V_s(x)}$$

On écrit facilement à partir de ces équations

$$Q'_I = -C'_{OX}\left(V_{GB} - V_{FB} - V_S - \gamma\sqrt{V_s}\right) \tag{8-10}$$

En appelant V_{FB} la tension de bandes plates

$$V_{FB} = \phi_{MS} - \frac{Q'_0}{C'_{ox}} \tag{8-11}$$

$$\gamma = \frac{\sqrt{2eN_A\varepsilon_s}}{C'_{ox}} \tag{8-12}$$

L'équation 8.1 s'intègre alors sous la forme

$$I_D L = \mu_n W \int_{V_s(0)}^{V_s(L)} C'_{OX}\left(V_{GB} - V_{FB} - V_s - \gamma\sqrt{V_s}\right)\frac{dV_s}{dx}dx + \mu_n W \phi_t \int_{Q'_I(0)}^{Q'_I(L)} \frac{dQ'_I}{dx}dx$$

soit,

$$I_D L = \mu_n W \int_{V_s(0)}^{V_s(L)} C'_{OX}\left(V_{GB} - V_{FB} - V_s - \gamma\sqrt{V_s}\right)dV_s + \mu_n W \phi_t \int_{Q'_I(0)}^{Q'_I(L)} dQ'_I$$

Le courant de drain s'exprime alors facilement en fonction du potentiel de surface.

$$I_D L = \mu_n W C'_{OX}\left[\left(V_{GB} - V_{FB}\right)\left(V_s(L) - V_s(0)\right) - \frac{1}{2}\left(V_s^2(L) - V_s^2(0)\right) - \frac{2}{3}\gamma\left(V_s^{\frac{3}{2}}(L) - V_s^{\frac{3}{2}}(0)\right)\right]$$
$$+\mu_n W C'_{OX}\phi_t\left[V_s(L) - V_s(0) + \gamma\left(V_s^{\frac{1}{2}}(L) - V_s^{\frac{1}{2}}(0)\right)\right]$$

En fonction des considérations du paragraphe 8.2, le potentiel de surface s'écrit sous la forme

$$V_S(x) = \phi_B + V_{CB}(x)$$

V_{CB} est égale à V_{SB} au niveau de la source et égale à V_{DB} au niveau du drain. Le terme Φ_B est classiquement choisi égal à $2\Phi_F$ (voir chapitre 5). Cette relation est une conséquence du régime de forte inversion et généralise les résultats du chapitre 5. Le potentiel de surface tend vers une valeur constante quand la tension de grille augmente. On peut montrer facilement que le terme $V_{CB}(x)$ est la différence entre le pseudo niveau de Fermi des électrons de conduction et celui des trous du bulk (annexe 5).

On peut alors calculer le courant de drain. Après quelques manipulations algébriques, on obtient la relation 8-13.

$$I_D = \frac{W}{L}\mu_n C'_{OX}\left[\left(V_{GS} - V_{FB} - 2\phi_F\right)V_{DS} - \frac{1}{2}{V_{DS}}^2 - \frac{2}{3}\gamma\left[\left(2\phi_F + V_{SB} + V_{DS}\right)^{\frac{3}{2}} - \left(2\phi_F + V_{SB}\right)^{\frac{3}{2}}\right]\right]$$

Il est maintenant possible de tracer le courant en fonction de V_{DS} les autres paramètres étant fixés, en particulier la tension de grille.

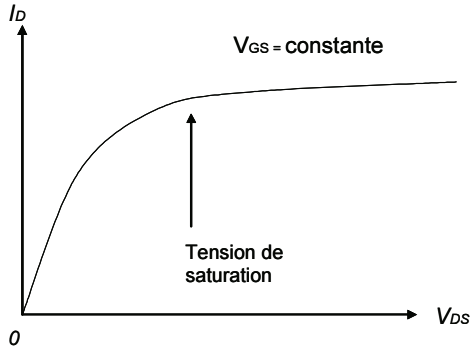


Figure 8-6 : Variation du courant de drain en forte inversion

On constate que la pente de cette courbe diminue quand la tension augmente. Elle s'annule pour une certaine valeur de V_{DS} . Cette valeur est la solution de l'équation

$$\frac{dI_D}{dV_{DS}} = 0$$

On trouve alors

$$V_{DS\,sat} = V_{GS} - 2\phi_F - V_{FB} + \frac{\gamma^2}{2} - \gamma\sqrt{V_{GS} - V_{FB} + V_{SB} + \frac{\gamma^2}{4}} \qquad (8\text{-}14)$$

Rappelons que le paramètre γ a la dimension d'une racine carrée de tension et que sa valeur est d'environ 0,5 $V^{1/2}$. Au-delà de cette valeur, le transistor n'est plus en régime de forte inversion au niveau du drain et la formule n'est plus valable. Si elle l'était, le courant de drain diminuerait ce qui est contraire au sens physique. Il reste à prouver que cette valeur de la tension de drain correspond à la valeur qui annule la charge d'inversion au niveau du drain.

Pour cela on utilise la relation générale 5-18 donnant la charge d'inversion au niveau du drain comme il a été expliqué dans le chapitre 5 mais en remplaçant V_s par V_s -V_{DB}

$$Q'_I = -\sqrt{2e\varepsilon_s N_A}\left[\sqrt{V_s + \varphi_t \exp^{\frac{V_s - V_{DB} - 2\phi_F}{\varphi_T}}} - \sqrt{V_s}\right] \qquad (8\text{-}15)$$

Cette charge s'annule pour

$$V_s = V_{DB} + 2\phi_F$$

En supposant nulle la charge d'inversion, les équations du dispositif au niveau du drain s'écrivent

$$Q'_G + Q'_0 + Q'_B = 0$$

$$V_{GB} = V_{OX} + V_s + \phi_{MS}$$

$$Q'_G = C'_{OX} V_{OX}$$

$$Q'_B = -\sqrt{2eN_A\varepsilon_s}\sqrt{V_s}$$

On calcule alors V_s à partir de ces équations :

$$V_{GB} = V_{FB} + V_s + \gamma\sqrt{V_s}$$

Cette équation du second degré par rapport à $\sqrt{V_s}$ a comme solution

$$V_s = \left(-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + V_{GB} - V_{FB}} \right)^2$$

On en déduit V_{DB} puis V_{DS} . On retrouve bien la valeur de $V_{DS\,sat}$ obtenue en 8-14.

Ce calcul un peu long montre la cohérence de la représentation. ***La valeur de saturation correspond à l'annulation de la charge d'inversion au niveau du drain***. En réalité, le calcul du courant de drain n'est pas valable en régime de saturation puisque le dispositif n'est plus en régime de forte inversion au niveau du drain. On supposera cependant que l'erreur introduite par cette approximation est négligeable ce qui est vérifié dans la pratique. Quand la tension de drain continue d'augmenter, l'expérience montre que le courant de drain reste constant. En effet, la région au niveau du drain a une charge d'inversion nulle, si bien que la zone canal-drain peut être considérée comme une jonction *pn* polarisée en inverse et donc électriquement équivalente à une résistance de valeur élevée. Toute augmentation de tension appliquée à ce dispositif comprenant en série la partie conductrice du canal et la jonction polarisée en inverse au niveau du drain est donc intégralement appliquée à cette jonction. La tension appliquée aux bornes de la partie conductrice du canal n'augmentant pas, le courant de drain n'augmente pas également. Il ne faudrait pas penser que la jonction *pn* polarisée en inverse au niveau du drain est un obstacle au passage d'un courant de la source vers le drain. Le champ appliqué a en effet la bonne orientation pour pousser les électrons du canal conducteur vers le drain. Cette zone en bout de canal dans laquelle la charge d'inversion est nulle est appelée région de pincement ou *pinch-off*.

La figure 8-7 représente les variations du courant de drain en fonction de la tension appliquée entre drain et source pour différentes valeurs de la tension appliquée entre grille et source (1V, 1.5V et 2V). Le transistor représenté a un canal de 0,18 micron de longueur et une largeur *W* de 10 microns.

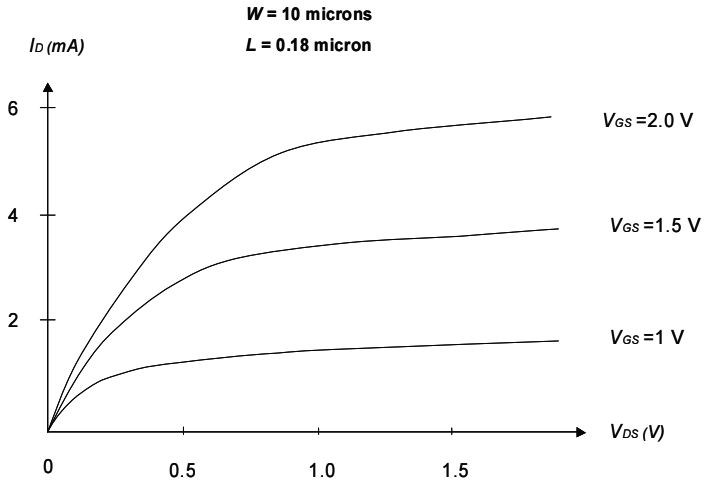


Figure 8-7 : Courbes courant de drain / tension drain source pour un transistor de 10 microns de large dans une technologie 0,18 micron.

Il est maintenant possible de simplifier l'expression 8.13 du courant de drain moyennant quelques approximations supplémentaires. Pour cela on reprend les équations de base rappelées en début de ce chapitre.

$$Q_G + Q'_0 + Q'_I + Q'_B = 0$$

$$V_{GS} = V_{ox} + V_s + \phi_{MS}$$

$$Q'_G = C'_{OX} V_{OX}$$

$$Q'_B = -\sqrt{2eN_A\varepsilon_s}\sqrt{V_s}$$

Le courant s'écrit

$$I_D L = \mu_n W \int_{V_s(0)}^{V_s(L)} C'_{OX}\left(V_{GB} - V_{FB} - V_S - \gamma\sqrt{V_s}\right) dV_s$$

Le dernier terme de l'intégrale sera approximé par une fonction linéaire

$$\gamma\sqrt{V_s} \approx \gamma\sqrt{V_{SB}+2\phi_F} + \delta\cdot\left(V_{CB}(x)-V_{SB}\right)$$

La valeur de δ sera discutée ultérieurement. Le calcul du courant conduit alors au résultat suivant

$$I_D = \frac{W}{L}\mu_n C'_{OX}\left[\left(V_{GS}-V_T\right)V_{DS} - \frac{1}{2}\left(1+\delta\right)V_{DS}{}^2\right] \tag{8-16}$$

Dans cette relation, le seuil V_T est défini par

$$V_T = V_{T_0} + \gamma\left(\sqrt{2\phi_F + V_{SB}} - \sqrt{2\phi_F}\right) \tag{8-17}$$

avec

$$V_{T_0} = V_{FB} + 2\phi_F + \sqrt{2\phi_F} \tag{8-18}$$

Cette relation est valable tant que V_{DS} est inférieure à $V_{DS\,sat}$ c'est à dire avant le régime de saturation. La valeur $V_{DS\,max}$ est obtenue en annulant la dérivée du courant par rapport à la tension de drain, on obtient alors

$$V_{DSsat} = \frac{V_{GS}-V_T}{1+\delta}$$

Le courant de saturation est alors

$$I_{Dsat} = \frac{W}{L}\mu_n C'_{OX}\frac{\left(V_{GS}-V_T\right)^2}{2\left(1+\delta\right)} \tag{8-19}$$

Cette valeur est évidemment indépendante de V_{DS} .

Le calcul précédent a également mis en relief la dépendance de la tension de seuil V_T avec la polarisation de la source $V_{SB.}$ Cet effet important est appelé effet *body* dans la littérature. Il est possible d'en donner une interprétation physique simple. Quand la source est polarisée positivement par rapport au substrat, elle attire les électrons de la couche d'inversion. Il faut donc compenser cet effet en augmentant le potentiel de grille ce qui est équivalent à une augmentation de la tension de seuil.

Pour terminer cet important paragraphe, il faut indiquer les valeurs de δ les plus utilisées. La valeur la plus immédiate est la dérivée de la fonction pour la valeur $V_{SB} + 2\phi_F$. On obtient alors

$$\delta = \frac{\gamma}{2\sqrt{2\phi_F + V_{SB}}}$$

Comme cette valeur surestime le calcul, on utilise aussi les valeurs suivantes

$$\delta = k\frac{\gamma}{2\sqrt{2\phi_F + V_{SB}}} \quad \text{ou} \quad \delta = \frac{\gamma}{2\sqrt{2\phi_F + V_{SB} + \phi_3}}$$

dans lesquels k et ϕ_3 sont des paramètres d'ajustement.

Pour terminer, il faut également noter les relations approchées très souvent utilisées dans la littérature qui correspondent à δ égal à zéro

$$I_D = \frac{W}{L}\mu_n C'_{OX}\left[(V_{GS} - V_T)V_{DS} - \frac{1}{2}{V_{DS}}^2\right] \tag{8-20}$$

$$I_{Dsat} = \frac{W}{L}\mu_n C'_{OX}\frac{(V_{GS} - V_T)^2}{2} \tag{8-21}$$

Ces relations sont très utilisées mais introduisent des erreurs importantes.

8.3.2. *Le régime de faible inversion*

Nous allons maintenant reprendre le calcul du courant pour des tensions de grille faibles. Les deux hypothèses du régime de forte inversion (courant de diffusion négligeable et potentiel de surface ayant atteint une valeur de saturation) seront abandonnées et remplacées par les deux hypothèses suivantes :

– Le courant de dérive est négligeable car la charge d'inversion a une valeur très faible dans tout le dispositif.
– Le potentiel de surface est de faible valeur.

À partir de la relation donnant le courant et négligeant cette fois le courant de dérive on écrit

$$I_D L = \mu_n W \phi_t \int_{Q'_I(0)}^{Q'_I(L)} dQ'_i = \mu_n W \phi_t \left[Q'_I(L) - Q'_I(0)\right]$$

Il suffit alors de reprendre l'expression de la charge d'inversion calculée dans le chapitre 5 en régime de faible inversion en prenant bien garde de remplacer dans le terme exponentiel V_S par V_s -V_{SB} au niveau de la source et V_s par V_s -V_{DB} au niveau du drain.

$$Q'_I(0) = -\frac{\gamma\, C'_{OX}}{2}\frac{\phi_t}{\sqrt{V_s(0)}}e^{\frac{V_s(0)-V_{SB}-2\phi_F}{\phi_t}}$$

$$Q'_I(L) = -\frac{\gamma\, C'_{OX}}{2}\frac{\phi_t}{\sqrt{V_s(L)}}e^{\frac{V_s(L)-V_{DB}-2\phi_F}{\phi_t}}$$

Le calcul exposé dans le paragraphe précédent montre que quand la charge d'inversion est faible, le potentiel de surface dépend uniquement de la tension de grille et s'exprime par

$$V_s(0) = V_s(L) = \left(-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + V_{GB} - V_{FB}}\right)^2$$

Le potentiel de surface étant constant tout le long du canal, le champ de dérive est nul et le courant de dérive également ce qui est cohérent avec l'hypothèse initiale.

On calcule alors le courant de drain :

$$I_D L = \mu_n W \phi_t \left[Q'_I(L) - Q'_I(0)\right] = -\mu_n W \phi_t Q'_I(0)\left[1 - \frac{Q'_I(L)}{Q'_I(0)}\right]$$

Comme le potentiel de surface est constant le long du canal, on obtient

$$I_D = -\frac{W}{L}\mu_n \phi_t Q'_I(0)\left(1 - e^{-\frac{V_{DS}}{\phi_t}}\right)$$

On peut alors reprendre la formule simplifiée 5.21 du chapitre 5 pour exprimer $Q'_I(0)$.

$$Q'_I(0) = -\sqrt{2e\varepsilon_s N_A}\;\frac{1}{2}\frac{\phi_t\, e^{\frac{-0,5\phi_F}{\phi_t}}}{\sqrt{1.5\,\phi_F + V_{SB}}}e^{\frac{V_{GS} - \phi_{MS} + \frac{Q_0}{C_{OX}} - 1.5\phi_F - \gamma\sqrt{1,5\,\phi_F + V_{SB}}}{\left(1 + \frac{\gamma}{2\sqrt{1,5\,\phi_F + V_{SB}}}\right)\phi_t}}$$

Cette relation un peu compliquée met en évidence la variation du courant avec le seuil défini par le terme V_X égal à la valeur suivante

$$V_X = \phi_{MS} - \frac{Q'_0}{C'_{OX}} + 1{,}5\,\phi_F + \gamma\sqrt{1.5\phi_F + V_{SB}}$$

Il est possible d'écrire la relation donnant le courant sous une forme plus lisible.

$$I_D = I_S \exp^{\frac{V_{GS}-V_X}{n_0\phi_t}}\left(1-\exp^{-\frac{V_{DS}}{\phi_t}}\right) \tag{8-22}$$

Dans laquelle,

$$I_S = \frac{W}{L}\mu_n \phi_t^{\,2}\sqrt{2e\varepsilon_s N_A}\,\frac{1}{2}\frac{e^{-\frac{0,5\phi_F}{\phi_t}}}{\sqrt{1.5\,\phi_F + V_{SB}}} \tag{8-23}$$

Rappelons que d'après les résultats du chapitre 3, le coefficient « *ideality* » *n* s'exprime par

$$n = \frac{dV_G}{dV_s} = 1 + \frac{\gamma}{2\sqrt{V_S}} \tag{8-24}$$

$$n = \frac{dV_G}{dV_s} = 1 + \frac{C'_{si}}{C'_{ox}} \tag{8.25}$$

Le courant d'inversion s'exprime en fonction de la valeur n_0 de n pour une valeur du potentiel de surface égal à $1.5\,\phi_F$. On peut montrer que le courant s'exprime approximativement sous la forme suivante souvent utilisée dans la littérature

$$I_D = \frac{W}{L}(n_0 - 1)\mu_n C'_{OX}\cdot\phi_t^{\,2}\cdot e^{\frac{V_{GS}-V_X}{n_0\phi_t}}\left(1-e^{-\frac{V_{DS}}{\phi_t}}\right) \tag{8-26}$$

Il est intéressant de comparer les deux seuils que nous avons définis : le seuil V_T en régime de forte inversion et le seuil V_X en régime de faible inversion.

$$V_X = \phi_{MS} - \frac{Q'_0}{C'_{OX}} + 1{,}5\phi_F + \gamma\sqrt{1.5\,\phi_F + V_{SB}} \tag{8-27}$$

$$V_T = \phi_{MS} - \frac{Q'_0}{C'_{ox}} + 2\phi_F + \gamma\left(\sqrt{2\phi_F + V_{SB}} - \sqrt{2\phi_F}\right) \quad (8\text{-}28)$$

Ces deux valeurs sont différentes ce qui n'est pas toujours pris en compte dans la littérature. On peut constater que ces valeurs sont très proches pour V_{SB} égal à zéro.

La formule 8.22 met en évidence un des phénomènes les plus importants dans le fonctionnement du MOSFET, l'augmentation de la consommation statique avec la diminution du seuil. Quand le MOSFET canal *n* est bloqué, la tension grille-source est alors nulle. Le courant de drain qui traverse le transistor est faible mais non nul. Il varie selon la formule 8-22 en $\exp(-V_X/n\Phi_t)$ ce qui montre bien son augmentation rapide quand la tension de seuil diminue. Au fur et à mesure que les technologies microélectroniques progressent, les tensions diminuent car les dimensions diminuent. La tension de seuil suit également cette loi ce qui entraîne une augmentation du courant de non conduction du transistor. Cet effet est d'autant plus important que le facteur *n* est élevé. Un objectif de la technologie est donc de réduire ce coefficient lié aux capacités du système comme le montre la formule 8.25 rappelée dans ce paragraphe.

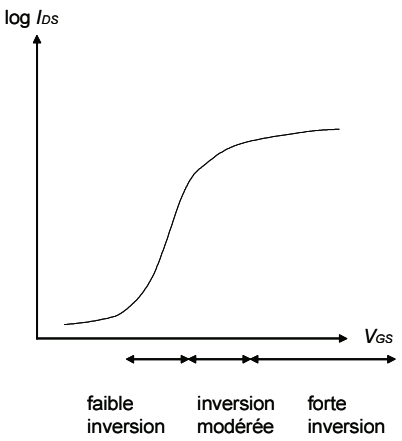


Figure 8-8 : les trois régimes de fonctionnement du MOSFET

Pour terminer cet important paragraphe, rappelons les différents régimes de fonctionnement du MOSFET sur la figure 8.8. L'échelle logarithmique met en relief le régime de faible inversion.

8.3.3 *La mobilité effective et ses effets*

La mobilité des porteurs dans le canal de conduction est différente de la mobilité dans le semi-conducteur en profondeur. Elle est plus faible car le champ électrique transverse attire les électrons ou les trous vers l'interface avec l'oxyde et favorise des interactions avec les impuretés présentes à cette interface. La prise en compte de ces effets dans le calcul du courant de drain est très complexe car la mobilité en surface est reliée au champ électrique transverse dans le dispositif par une relation de la forme

$$\mu = \frac{\mu_0}{1 + \alpha E_y}$$

Dans cette relation, μ_0 est une valeur qui dépend de la température. Elle est égale à environ la moitié de la mobilité mesurée dans le volume d'un semi-conducteur. La constante α a une valeur d'environ 0.025 µm/V à la température ambiante. Il est assez difficile de calculer analytiquement le courant de drain avec cette hypothèse. Il est en général admis que les relations données dans le paragraphe précédent restent valables à la condition de remplacer la mobilité μ par une mobilité effective donnée en première approximation par :

$$\mu_{eff} = \frac{\mu}{1 + \theta(V_{GS} - V_T) + \theta_B V_{SB}}$$

Les paramètres θ et θ_B sont respectivement $0.002/d_{OX}$ et quelques centièmes de V^{-1} avec d_{ox} exprimé en micron.

8.3.4. *Le MOS canal p*

La littérature étudie généralement le MOS canal *n* et ne fait que mentionner le MOS canal *p* en supposant qu'il est identique au premier à la condition d'inverser le signe des tensions. Cela est vrai pour l'essentiel mais le MOS canal *p* présente des particularités qu'il est bon de connaître et cela d'autant plus qu'il y a autant de MOS canal *p* que de MOS canal *n* dans les circuits intégrés. Reprenons donc le schéma de fonctionnement de base.

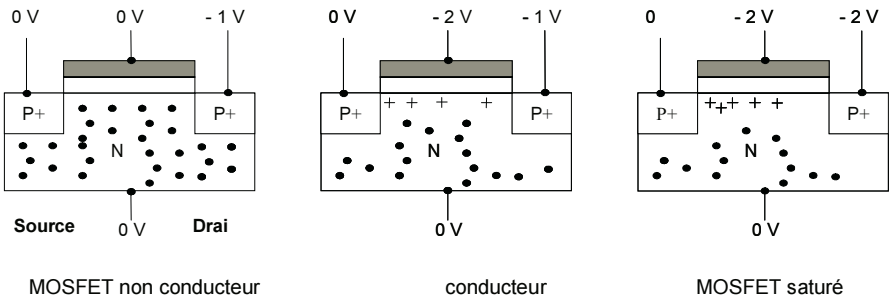


Figure 8-9 : Le fonctionnement du MOS canal p

L'application sur la grille d'une tension négative attire à l'interface avec l'oxyde les trous provenant de la source fortement dopée. Un champ électrique longitudinal créé entre la source et le drain et convenablement orienté (tension de drain plus négative que celle de la source) permet de créer un courant de dérive dans le dispositif. Les relations 8-16 et 8-17 exprimant le courant et la tension de seuil deviennent

$$I_D = -\frac{W}{L}\mu_p\ C'_{OX}\left[\left(V_{GS} - V_T\right)\ V_{DS} - \frac{1}{2}\left(1+\delta\right)\ {V_{DS}}^2\right]$$

$$V_T = V_{T_0} - \gamma\left(\sqrt{-2\phi_F - V_{SB}} - \sqrt{-2\phi_F}\right)$$

Il faut noter que la valeur de la mobilité à prendre en compte est celle des trous. Elle est environ trois fois plus faible que celle des électrons. Les tensions $V_{GS,}\ V_{DS,}\ \phi_F$ sont négatives dans ces formules. À dimensions égales et pour des tensions appliquées égales en valeur absolue, le courant de drain d'un MOS canal *p* est donc environ trois fois plus faible que celui d'un MOS canal *n*. La tension de seuil est également négative.

Donnons quelques précisions sur les signes car c'est souvent une source de confusion. En règle générale, on cherche à éviter les valeurs négatives dans les formules. On exprimera donc le courant de drain mesuré de la source vers le drain et non pas du drain vers la source comme dans le NMOS. Les tensions sont inversées par rapport à celles du NMOS. On exprime V_{SG} au lieu de V_{GS} et V_{SD} au lieu de V_{DS}. La grandeur ϕ_F est négative puisque le substrat est de type *n*.
Avec ces conventions, la formule du PMOS est la même que celle du NMOS.

$$I_D = \frac{W}{L}\mu_p\ C'_{OX}\left[\left(V_{SG} - \left|\ V_T\ \right|\right)\ V_{SD} - \frac{1}{2}\left(1+\delta\right)\ {V_{SD}}^2\right]$$

$$\left|\ V_T\ \right| = \left|\ V_{T_0} - \gamma\left(\sqrt{-2\phi_F - V_{SB}} - \sqrt{-2\phi_F}\right)\right|$$

8.4. Le modèle canal court

En réalité et pour les technologies actuelles, le transistor MOS ne peut être considéré comme un dispositif de longueur élevée par rapport aux autres dimensions. Les effets de source et de drain ne peuvent être négligés et une véritable description à deux dimensions serait en théorie nécessaire. Pour simplifier, le modèle à une dimension est conservé avec quelques ajouts :

- prise en compte de la zone de charge d'espace au niveau du drain et diminution de la longueur effective du canal,
- prise en compte de l'effet de saturation de la vitesse des électrons dans le canal,
- introduction du seuil effectif et de l'effet d'abaissement de ce seuil avec la tension de drain.

8.4.1. *Effet de diminution de la longueur de canal*

L'examen des courbes représentant le courant de drain en fonction de la tension drain-source fait apparaître que le courant n'est pas strictement indépendant de la tension drain-source en régime de saturation mais augmente légèrement quand celle-ci augmente. Cet effet peut s'expliquer par la modulation de la longueur du canal avec la tension de drain. Pour la valeur $V_{DS_{sat}}$ de la tension de drain, le transistor entre en régime de saturation et la densité d'électrons au niveau du drain devient nulle. Toute augmentation de tension supplémentaire se traduit par l'apparition d'une zone de d'espace dans la jonction pn qui se forme au niveau du drain comme le montre la figure 8.10. La profondeur de la zone de charge d'espace augmente quand la tension de drain augmente et la longueur du canal de conduction (région dans laquelle on trouve des électrons) diminue de ΔL ce qui a pour effet d'augmenter le courant de drain qui varie inversement proportionnellement à la longueur du canal.

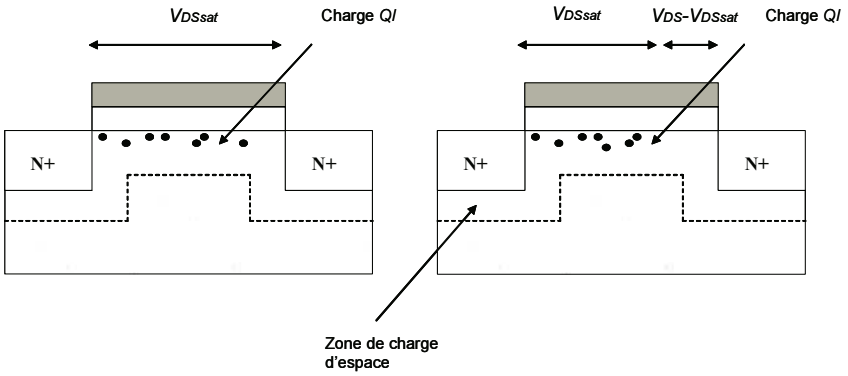


Figure 8-10 : Modulation de la longueur de canal

La prise en compte de ces effets conduit à écrire de manière approximative la valeur du courant de drain en régime de saturation sous la forme

$$I_D = I_{D_{sat}}\left(1+\frac{V_{DS}-V_{DS_{sat}}}{V_A}\right) \tag{8-29}$$

avec

$$V_A = kL\sqrt{N_A}$$

Le paramètre k est une constante égale à environ 0.15 V $\mu m^{1/2}$ et N_A est le dopage du substrat.

8.4.2. *Effet de saturation de la vitesse*

Dans le calcul précédent, la vitesse des électrons était supposée proportionnelle au champ électrique longitudinal (parallèle à la direction du canal), la constante de proportionnalité étant la mobilité. En fait, à partir d'une certaine valeur du champ E_C cette loi n'est plus vérifiée et la vitesse tend vers une valeur limite indépendante du champ et notée v_s.

Le formalisme développé jusqu'ici pour calculer le courant de conduction dans un semiconducteur converge vers la loi d'Ohm. C'est la conséquence directe du fait que la vitesse d'entraînement des électrons est proportionnelle au champ électrique. Ce formalisme, justifié dans le domaine des champs faibles, cesse d'être opérationnel quand le champ électrique dépasse typiquement 10^4 V/cm. Or si on considère, dans un composant, un canal conducteur de 1 µm de long, il y règne un champ de 10^4 V/cm pour une polarisation de seulement 1 V. Il en résulte que la loi d'Ohm n'est plus de mise dans un composant submicronique.

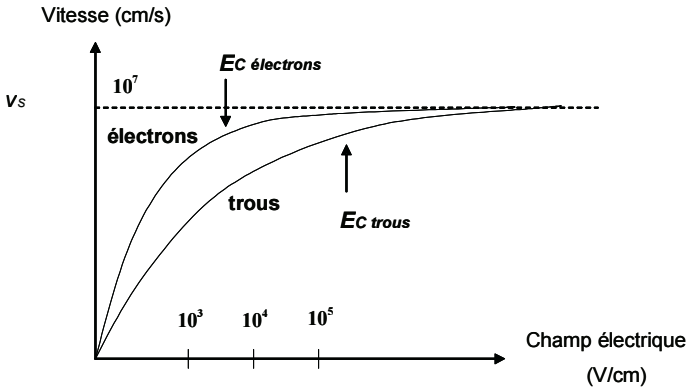


Figure 8-11 : Vitesse de saturation

a. Porteurs chauds

Lorsque le champ électrique créé dans le matériau augmente, l'énergie cinétique des porteurs libres augmente. Tant que le taux d'augmentation d'énergie est inférieur au taux de perte par interaction avec le réseau, les porteurs se thermalisent aux extrema des bandes. Ils sont alors en équilibre thermodynamique avec le réseau et leur distribution est régie par la fonction de Fermi.

Lorsque le champ devient important, le taux de gain d'énergie augmente plus vite que le taux de perte. Quand le premier atteint la valeur du second un nouveau régime de pseudo-équilibre s'établit. Les porteurs ne sont plus en équilibre thermodynamique avec le réseau, cependant leurs fortes interactions mutuelles leur permettent de se mettre en équilibre entre eux. Il s'agit alors d'un régime de pseudoéquilibre thermodynamique, leur distribution énergétique est toujours régie par une fonction de distribution mais avec une température effective propre au gaz de porteurs, différente de celle du réseau. On qualifie ces porteurs de *porteurs chauds* et la température qui régit leur distribution énergétique de *température électronique* $T_e > T$ où T est la température du réseau.

Lorsque l'énergie cinétique des porteurs devient telle que leurs interactions mutuelles sont faibles, le régime de pseudo-équilibre lui-même n'existe plus. Ces porteurs sont toujours qualifiés de porteurs chauds, mais il n'est plus possible de définir une température électronique car leur distribution énergétique n'est plus régie par une loi thermodynamique. En un point de l'espace et à un instant donné, la distribution énergétique des électrons est alors de type Dirac, c'est-à-dire un pic très localisé centré autour d'une valeur E_b. Ces porteurs chauds sont alors qualifiés de *porteurs balistiques*. Ils sont caractérisés par une *vitesse balistique* v_b donnée par $E_b = mv_b^2 / 2$.

b. Vitesse de saturation

En régime de porteurs chauds les propriétés de transport ne sont plus les mêmes qu'en régime d'équilibre thermodynamique. En particulier la vitesse d'entraînement des porteurs n'est plus proportionnelle au champ électrique. Dans le silicium en particulier, pour des champs inférieurs à 10^3 V/cm, la vitesse d'entraînement augmente linéairement avec le champ et le courant obéit à la loi d'Ohm. Au-delà, la vitesse présente d'abord une loi de variation sous-linéaire puis un régime de saturation. Lorsque les porteurs atteignent leur *vitesse de saturation*, le courant dans le composant devient indépendant de la polarisation.

La relaxation des électrons est conditionnée d'une part par les interactions électron-électron et d'autre part par les interactions électron-phonon. Le premier phénomène dépend beaucoup de la population électronique et ne joue vraiment un rôle significatif que dans les matériaux dégénérés. Le temps de relaxation associé aux interactions électron-phonon est de l'ordre de 10^{-13} seconde avec les phonons optiques et environ 100 fois plus grand avec les phonons acoustiques. Il en résulte que le processus de relaxation mettant en jeu les phonons optiques est de loin, le plus efficace.

On obtient la vitesse de saturation des porteurs à partir des équations de conservation du moment et de l'énergie.

Formellement, le taux net de variation du module du moment des électrons peut s'écrire sous la forme

$$\frac{dp_e}{dt} = \left.\frac{dp_e}{dt}\right|_E + \left.\frac{dp_e}{dt}\right|_r$$

Les deux termes du deuxième membre représentent les variations du moment p_e des électrons résultant respectivement du champ électrique et des processus de relaxation. Le premier est simplement donné par l'équation de la dynamique

$$F = m_e\gamma \;\Rightarrow\; eE = m_e\frac{dv}{dt} = \frac{dp_e}{dt}$$

soit

$$\left.\frac{dp_e}{dt}\right|_E = eE$$

Le deuxième terme est régi par les processus de relaxation que l'on peut globaliser par un temps de relaxation τ_m

$$\left.\frac{dp_e}{dt}\right|_r = -\frac{p_e}{\tau_m}$$

L'intégration de cette expression donne $p_e = p_o\, e^{-t/\tau_m}$, de sorte que τ_m, appelé *temps de relaxation du moment*, représente le temps nécessaire aux processus de relaxation pour ramener le moment de l'électron à sa valeur initiale. On obtient alors

$$\frac{dp_e}{dt} = eE - \frac{p_e}{\tau_m}$$

Le taux net de variation de l'énergie des électrons se met sous une forme analogue

$$\frac{dE_e}{dt} = \left.\frac{dE_e}{dt}\right|_E + \left.\frac{dE_e}{dt}\right|_r$$

Si la vitesse des électrons dans la direction du champ électrique est v, leur gain d'énergie par unité de temps est

$$\left.\frac{dE_e}{dt}\right|_E = eEv$$

La principale source de relaxation des électrons étant les phonons optiques, l'échange d'énergie électron-phonon est quantifié, le quantum d'énergie étant

l'énergie du phonon optique $\hbar\Omega$. Ainsi, dans chaque processus de relaxation par un phonon, la perte d'énergie d'un électron est limitée à $\hbar\Omega$. Le deuxième terme du second membre de l'expression s'écrit donc

$$\left.\frac{dE_e}{dt}\right|_r = -\frac{\hbar\Omega}{\tau_e}$$

où $\hbar\Omega$ est l'énergie du phonon optique et τ_e *le temps de relaxation de l'énergie.* L'équation d'évolution de l'énergie s'écrit donc

$$\frac{dE_e}{dt} = eEv - \frac{\hbar\Omega}{\tau_e}$$

En régime stationnaire $dE_e/dt = 0$ et $dp_e/dt = 0$ de sorte que les expressions précédentes s'écrivent

$$eE - \frac{m_e v}{\tau_m} = 0$$

$$eEv - \frac{\hbar\Omega}{\tau_e} = 0$$

En éliminant le champ électrique entre les deux équations on obtient l'expression de la vitesse de saturation

$$v_s = \sqrt{\frac{\hbar\Omega}{m_e}\frac{\tau_m}{\tau_e}} \qquad (8\text{-}30)$$

Les temps de relaxation du moment τ_m et de l'énergie τ_e sont en général différents, tout simplement par le fait que parmi les collisions électron-phonon, certaines sont élastiques d'autres non. Or les premières relaxent le moment, par changement de direction de la vitesse de l'électron, sans relaxer l'énergie. Ainsi le rapport τ_e / τ_m >1 mesure le rapport des probabilités de collisions élastiques et inélastiques.

Si par souci de simplification on suppose égaux ces deux temps de relaxation, la vitesse de saturation des porteurs s'écrit

$$v_s \approx \sqrt{\frac{h\Omega}{m_e}}$$

L'énergie des phonons optiques est du même ordre de grandeur dans tous les semiconducteurs, $\hbar\Omega \approx 40$ meV. Il en est de même pour la masse effective des électrons de haute énergie, qui est comparable à celle de l'électron libre, $m_e \approx m_O$. Il

en résulte que la vitesse de saturation des électrons est sensiblement la même dans tous les semiconducteurs, avec une valeur typique de l'ordre de 10^7 cm/s.

c. *Régime de survitesse*

La loi de variation de la vitesse d'entraînement des électrons avec le champ électrique, est différente suivant le type de matériau. Dans les semiconducteurs multivallées comme le silicium, la vitesse augmente d'abord linéairement (μ=Cte), puis sous-linéairement, pour enfin saturer à la valeur limite v_S.

Dans certains semiconducteurs à gap direct comme GaAs, la loi de variation de la vitesse présente un maximum que l'on qualifie de *régime de survitesse,* suivi d'une pente négative. Cette pente négative qui correspond à une mobilité différentielle négative, résulte de l'effet RWH, c'est-à-dire du transfert d'électrons depuis le minimum de la bande de conduction vers des minima satellites situés à quelques centaines de meV au-dessus.

Cette loi de variation très spécifique, qui joue un rôle important dans le fonctionnement de certains composants au GaAs, peut être modélisée par une expression empirique relativement simple. Dans la région du régime de survitesse, la variation de la vitesse d'entraînement des électrons avec le champ peut se mettre sous la forme (Chang C.S.-1986)

$$v=\frac{-\mu_o E}{\left(1+u\left(E-E_o\right)\left(\frac{E-E_o}{E_c}\right)^2\right)^{1/2}} \tag{8-31}$$

où $u(E\text{-}E_O)$ est la fonction de Heaviside, $u(E\text{-}E_O)$=0 pour $E < E_O$ et $u(E\text{-}E_O)$=1 pour $E > E_O$. μ_O est la mobilité à faible champ. E_C=-v_S/μ_O est le *champ critique* pour lequel la vitesse en régime linéaire est égale à la vitesse de saturation. Le paramètre E_O est donné par

$$E_o=\frac{1}{2}\left(E_m+\left(E_m^2-4E_c^2\right)^{1/2}\right)$$

où E_m est le *champ de seuil*, correspondant au maximum du régime de survitesse. E_m est donné par dv/dE = 0. Pour $E > E_m$ la mobilité différentielle devient négative.

L'expression contient trois paramètres caractéristiques du semiconducteur, la mobilité à faible champ μ_O , la vitesse de saturation v_S , et le champ de seuil E_m . Pour GaAs, $E_m \approx 4000$ V/cm , $v_S \approx 10^7$ cm/s et $\mu_O \approx 5000$ cm^2/Vs pour n=10^{17} cm^{-3}.

Dans le canal conducteur d'un transistor à effet de champ au GaAs,la densité de courant donnée par $J=\rho v$ où $\rho=-ne$ est la densité de charge électronique, s'écrit donc

$$J=-\frac{\rho\mu_o E}{\left(1+u\left(E-E_o\right)\left(\frac{E-E_o}{E_c}\right)^2\right)^{1/2}}$$

Compte tenu de la relation $E=-dV/dx$, la relation est l'équation différentielle courant-tension dans le canal du transistor. Cette équation s'explicite sous différentes formes opérationnelles correspondant aux différents régimes de fonctionnement. En régimes linéaire ($E<E_O$) , sous-linéaire ($E_O<E<E_m$) , et à mobilité différentielle négative ($E>E_m$) , cette équation s'écrit (Chang C.S.-1986)

$$E\le E_o \qquad \frac{dv}{dx}=\frac{E_c}{\beta} \tag{8-32-a}$$

$$E_o\le E\le E_m \qquad \frac{dv}{dx}=\frac{E_o\left(1+c^2\right)}{1+\sqrt{\left(1+c^2\right)\beta^2-c^2}} \tag{8-32-b}$$

$$E_m\le E \qquad \frac{dv}{dx}=\frac{E_o\left(1+c^2\right)}{1-\sqrt{\left(1+c^2\right)\beta^2-c^2}} \tag{8-32-c}$$

où $c=E_c/E_o$ *et* $\beta=\rho\mu_o E_c/J$. Le numérateur des expressions est la valeur du champ de seuil correspondant à la vitesse maximum, $E_m=E_O(1+c^2)$.

d. Retour au MOSFET Silicium

La relation entre vitesse et champ peut s'écrire de manière simplifiée

$$v(x)=v_S\frac{\frac{1}{E_C}\frac{dV_S}{dx}}{1+\frac{1}{E_C}\frac{dV_S}{dx}}$$

La relation de base exprimant le courant était

$$I_D=\mu_n W\left(-Q_I'\right)\frac{dV_s}{dx}+\mu_n W\phi_t\frac{dQ_I'}{dx}$$

soit en négligeant le courant de diffusion

$$I_D = \mu_n W\left(-Q_I'\right)\frac{dV_s}{dx}$$

Intégrons dans cette expression la formule donnant la vitesse

$$I_D = W\left(-Q_I'\right)\ v_S \frac{\dfrac{1}{E_C}\dfrac{dV_S}{dx}}{1+\dfrac{1}{E_C}\dfrac{dV_S}{dx}}$$

On obtient donc

$$I_D\left(1+\frac{1}{E_C}\frac{dV_S}{dx}\right) = W\left(-Q_I'\right)\ v_S \frac{1}{E_C}\frac{dV_S}{dx}$$

En intégrant cette équation de $x=0$ à $x=L$, on obtient alors

$$I_D\left(L+\frac{V_{DB}-V_{SB}}{E_C}\right) = W\mu_n \int_{V_{SB}}^{V_{DB}} \left(-Q_I'\right) dV_S$$

$$I_D = \frac{W}{L}\ \frac{\mu_n}{1+\dfrac{V_{DS}}{LE_C}} \int_{V_{SB}}^{V_{DS}} \left(-Q_I'\right) dV_S$$

Si on note I_{D0} le courant calculé en ne tenant pas compte de l'effet de saturation de la vitesse, on peut écrire

$$I_D = \frac{1}{1+\dfrac{V_{DS}}{LE_C}} I_{D0}$$

Si on exprime dans cette formule le courant calculé sans effet de saturation par sa valeur 8-16, on obtient

$$I_D = \frac{W}{L}\left(\frac{1}{1+\frac{V_{DS}}{LE_C}}\right)\mu_n C'_{OX}\left[\left(V_{GS}-V_T\right)V_{DS}-\frac{1}{2}\left(1+\delta\right)V_{DS}{}^2\right] \qquad (8\text{-}33)$$

Cette relation est valable en régime non saturé, puis au-delà d'une tension V_{DSsat} le courant reste constant quand la tension de drain augmente. Cette valeur V_{DSsat} est obtenue quand la dérivée du courant par rapport à la tension V_{DS} est nulle. On obtient facilement

$$V_{DS_{sat}} = LE_c\left[\sqrt{1+\frac{2\left(V_{GS}-V_T\right)}{\left(1+\delta\right)LE_C}}-1\right] \qquad (8\text{-}34)$$

Cette relation montre que le régime de saturation est atteint pour des valeurs plus faibles de la tension de drain. En effet, la valeur de la tension de saturation sans effet de saturation de la vitesse est

$$V_{DSsat} = \frac{V_{GS}-V_T}{1+\delta} \qquad (8\text{-}35)$$

Il ne faut pas confondre la saturation de la vitesse et le régime de saturation du transistor. Les deux valeurs de la tension de saturation sont égales pour un champ critique infini. En pratique, le champ critique est de l'ordre de 10^4 V/cm. Le produit LE_C est donc de 0.1 V pour une longueur de 100 nm et de 0.8 V pour une longueur de 0.8 micron. La tension de saturation est donc environ de 50 mV pour une longueur de 100 nm.

Que deviennent ces expressions quand le canal est très petit ce qui est le cas des technologies modernes ? Il est maintenant possible de calculer la valeur du courant de saturation à partir des relations 8.33 et 8.34 en faisant l'hypothèse que le terme V_{DS}/LE_C est très supérieur à 1.

$$I_{D\,sat} = WE_C\mu_n C'_{OX}\left[\left(V_{GS}-V_T\right)-\frac{\left(1+\delta\right)}{2}V_{DSsat}\right] \qquad (8.36)$$

La tension de saturation est donnée par la formule 8.34. Cette expression est parfois simplifiée en négligeant la tension de saturation.

$$I_{D\,sat} = W\mu_n C'_{OX}\left[V_{GS}-V_T\right]E_C \qquad (8.37)$$

La valeur du champ critique E_C est de 1V/µ à 3V/µ pour les électrons et de 2V/µ à 10V/µ pour les trous. On met ainsi en évidence la variation linéaire du courant avec la tension de grille. Rappelons que dans le cas d'un canal long, la dépendance était quadratique. Le courant est donc dans le cas du canal court indépendant de la longueur du canal.

Pour donner un sens physique à ce résultat qui peut sembler paradoxal, il suffit de considérer la dépendance de la charge et du temps de transit en fonction de la longueur du canal. Le temps de transit à vitesse constante est proportionnel à la longueur du canal. La charge stockée dans le canal est également proportionnelle à la longueur du canal. Le courant qui est le rapport des deux est donc indépendant de cette longueur. Notons également que les vitesses de saturation des électrons et des trous sont voisines (environ 100 microns par ns). Les performances en vitesse des NMOS et des PMOS sont donc voisines en régime de canal court. Le tableau suivant illustre les deux régimes de fonctionnement.

Courant de drain pour un canal long en régime de saturation	**Courant de drain pour un canal court en régime de saturation**
$I_{Dsat} = \frac{W}{L}\mu_n C'_{OX}\frac{(V_{GS}-V_T)^2}{2(1+\delta)}$	$I_{D\,sat} = W\mu_n C'_{OX}\left[V_{GS}-V_T-\frac{1+\delta}{2}V_{DSsat}\right]E_C$

Il est possible d'illustrer ces conclusions sur deux courbes représentant les variations du courant de drain en fonction de la tension de grille, la première pour un canal de 20 microns et la seconde pour un canal de 45 nm. Pour un transistor à canal court le courant de drain varie proportionnellement à la tension de grille.

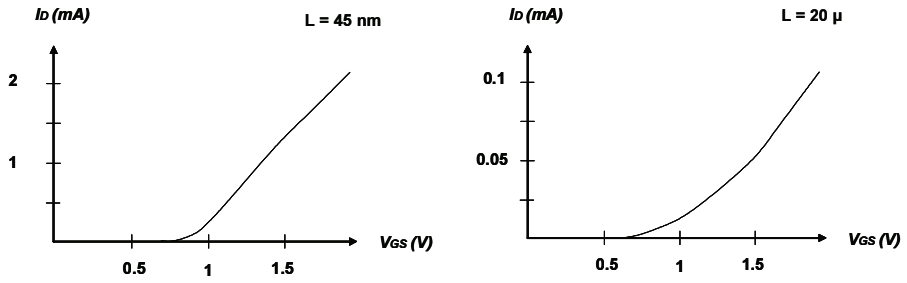


Figure 8-12 : Variation du courant de drain pour L= 45 nm et pour L=20 microns

8.4.3. *Effet de diminution du seuil effectif*

Les effets de canal court se manifestent également par une déformation des lignes de champ dans la zone de charge d'espace du dispositif comme le montre la figure 8-13

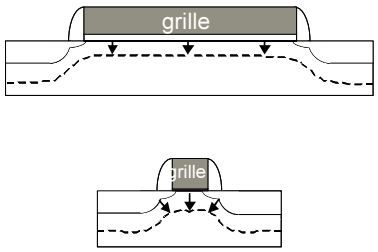


Figure 8-13 : Modification de la zone de charge d'espace

Quand le dispositif devient plus court, pour une même tension de grille la zone de charge d'espace s'étend en profondeur (effets de bord dus aux zones de diffusion de la source et du drain). La charge Q_B de la zone de charge d'espace est alors plus importante que si le dispositif était idéal. Le potentiel de surface augmente donc et en conséquence le courant de drain. À tension de grille égale, le courant de drain est plus important dans un dispositif à canal court. Cet effet peut également se représenter par une diminution de la tension de seuil. Il est important puisqu'il peut atteindre plusieurs dixièmes de volts. Des calculs approchés et basés sur des considérations purement géométriques de la déformation de la zone de charge d'espace conduisent à l'expression suivante de la diminution de la tension de seuil.

$$\Delta V_T = 2\beta \frac{\varepsilon_s}{\varepsilon_{ox}} \frac{d_{ox}}{L} \left(2\phi_F + V_{SB}\right)$$

Dans cette expression on notera l'effet des permittivités du silicium et de l'oxyde de grille (ε_s et ε_{ox}) ainsi que l'effet de l'épaisseur de l'oxyde (d_{ox}). Le coefficient β est peu différent de 1.

Dans les considérations précédentes, la tension entre drain et source n'a pas été prise en compte. Elle a pourtant un rôle important exprimé dans le paramètre DIBL (*Drain Induced Barrier Lowering*). Il exprime la diminution de la tension de seuil en fonction de la tension de drain. Quand la tension de drain augmente, les électrons sous l'oxyde de grille sont attirés par le drain ce qui augmente le courant dans le canal. On peut exprimer cette augmentation par une diminution de la tension de seuil. Cet effet est d'autant plus important que le canal est court car l'influence électrostatique du potentiel de drain dépend de la distance. Ce terme

qui conduit à faire varier la tension de seuil doit être le plus faible possible car les variations de tension de seuil peuvent conduire à dégrader les performances des circuits logiques réalisés avec des MOSFETs. La relation précédente devient donc

$$\Delta V_T = 2\beta \frac{\varepsilon_s}{\varepsilon_{ox}} \frac{d_{ox}}{L} \left[\left(2\phi_F + V_{SB}\right) + \chi V_{DS} \right] \tag{8-38}$$

Le coefficient χ est proportionnel à l'inverse de la longueur du canal. La figure ci-dessous illustre comment la tension de seuil varie avec la tension de drain dans diverses technologies.

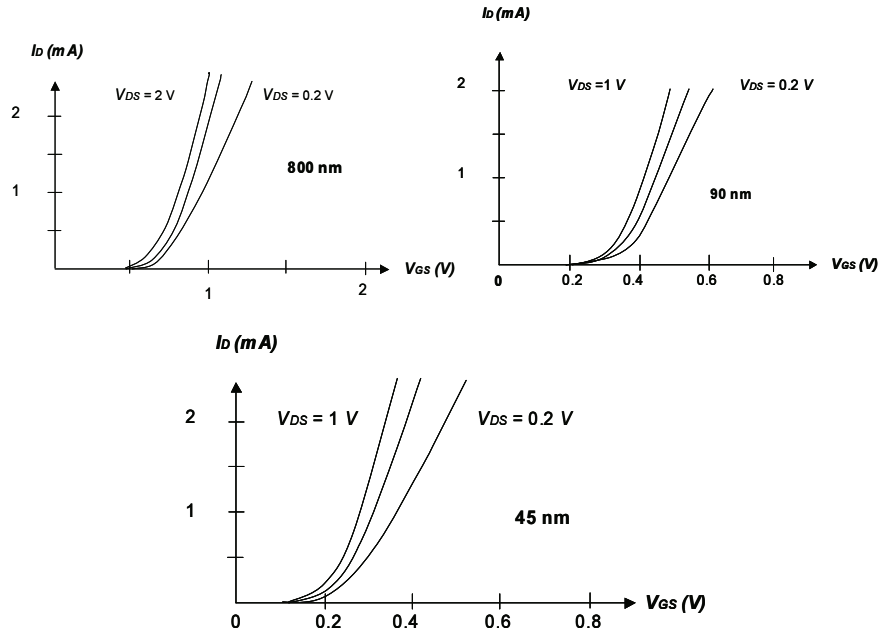


Figure 8.14 : Diminution de la tension de seuil avec la tension de drain

8.4.4. *Traitement analytique du canal court*

Les développements qui suivent montrent comment on peut traiter de manière rigoureuse le cas du canal court. Ils permettrons se retrouver les résultats qualitatifs exposés dans le paragraphe précédent. Le calcul est basé sur le théorème de superposition et fait intervenir les trois états représentés figure 8-15. On part d'un transistor polarisé. Le principe de superposition conduit à définir soit le potentiel soit les charges dans les différentes régions du système. Les états sont choisis pour retrouver le cas du canal long (état 2 au centre de la figure) et

conduire par addition des charges et des potentiels à l'état initial (à gauche de la figure). Les potentiels imposés sont représentés avec une flèche pleine. Les charges prises en compte sont les dopants considérés comme des conducteurs ponctuels. Elles sont annulées dans la zone de drain et de déplétion du canal dans l'état 1, dans la source et le drain dans l'état 2, dans la source et la zone de déplétion pour l'état 3.

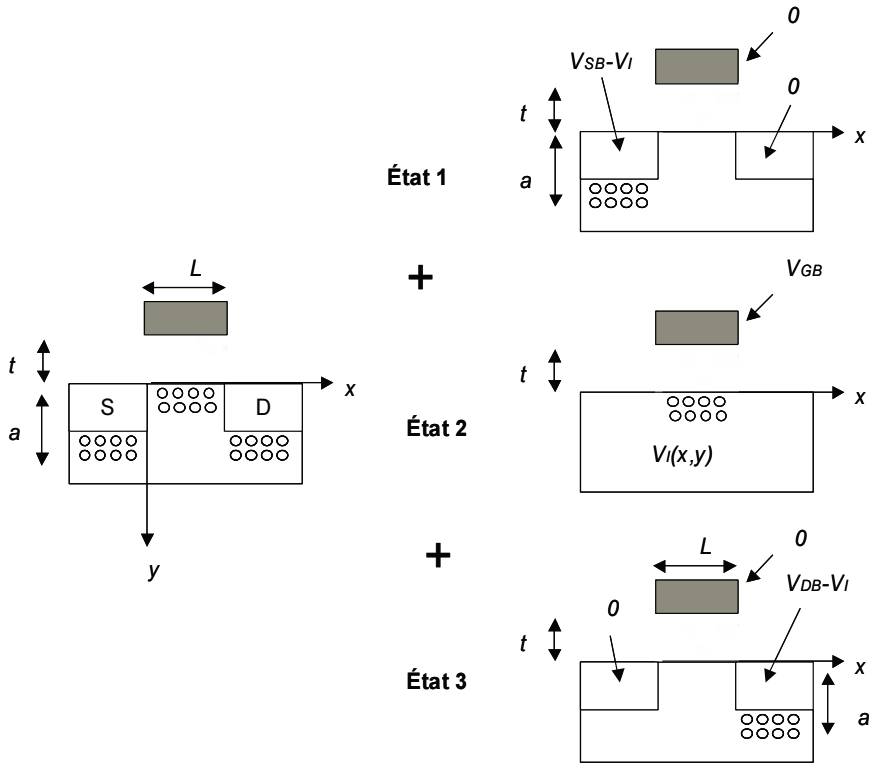


Figure 8-15 : Calcul en 2D

Dans l'état 2, les dopants de source et de drain sont de charge nulle et une tension V_{GB} est appliquée sur la grille. La zone de charge d'espace est dopée. Le potentiel $V_I(x,y)$ est alors celui calculé en régime de canal long.

Dans le premier état, les dopants sont placés à charge nulle dans la zone de charge d'espace mais la tension appliquée dans la zone de drain est nulle. Une tension nulle est appliquée sur la grille. Le potentiel dépend alors de *x* et de *y* dans le silicium et dans l'oxyde. Une tension $V_{SB} - V_I$ est appliquée dans la zone de source. Cette valeur permet de retrouver par somme des états la valeur du dispositif

initial. On se place dans la suite du calcul en régime de faible inversion car le but est d'étudier la modification de la tension de seuil. Le potentiel de l'état deux en régime de faible inversion est alors indépendant de *x*. Dans le troisième état ce sont les charges de la zone de source qui sont annulées. Une tension $V_{DB} - V_l$ est appliquée dans la zone de drain et la tension appliquée dans la zone de source est nulle. Une tension nulle est appliquée sur la grille. Il est facile de vérifier que la superposition des trois états conduit à l'état initial par addition des charges et des potentiels.

Pour calculer le potentiel dans les états un et trois on applique la technique de décomposition en série valable pour la résolution de l'équation de Laplace quand la charge d'espace est nulle.

$$V_1(x,y) = \int_0^\infty a(k) sh[k(x-L)] \sin[k(y-a)] dk$$

Cette solution intègre les conditions aux limites

$$V_1(L,y) = 0 \qquad V_1(x,a) = 0$$

On peut écrire de même dans l'oxyde

$$V_{01}(L,y) = 0 \qquad V_{01}(x,-t) = 0$$

$$V_{01}(x,y) = \int_0^\infty a_0(k) sh[k(x-L)] \sin[k(y+t)] dk$$

La continuité du potentiel à l'interface oxyde-semiconducteur permet d'écrire

$$a(k)\sin(ka) + a_0(k)\sin(kt) = 0$$

La continuité de $\varepsilon \mathbf{E}$ à l'interface conduit à écrire

$$\varepsilon_s a(k)\cos(ka) - \varepsilon_{ox} a_0(k)\cos(kt) = 0$$

On en déduit donc

$$\varepsilon_s \cos(ka)\sin(kt) + \varepsilon_{ox} \sin(ka)\cos(kt) = 0 \qquad (8\text{-}39)$$

Cette équation permet de montrer que *k* ne prend que des valeurs discrètes comme le montre la figure 8-16 qui représente la fonction précédente en fonction de *k*.

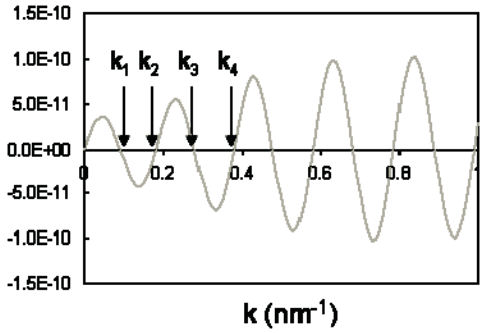


Figure 8-16 : Valeurs possibles de *k*

L'étape suivant du calcul consiste à limiter le développement en série à un seul terme celui correspondant à la plus petite valeur de *k*. Le potentiel s'écrit alors

$$V(x,y) = a_1 sh[k_1(x-L)]\sin[k_1(y-a)]$$

La valeur en $x=0$ et $y=0$ permet de calculer la constante a_1.

$$V_{SB} - V_I(x,y) + V_d = a_1 sh[k_1 L]\sin[k_1 a]$$

On reconnaît dans cette formule la différence de potentiel V_d qui se forme dans une jonction non polarisée. Il faut calculer le potentiel aux bornes de la jonction np. Le potentiel est calculé par rapport au bulk et le potentiel calculé en $x=0$ et $y=0$ est la différence de potentiel aux bornes de la jonction np et non pas le potentiel appliqué. Cela explique V_d dans la formule. On obtient alors

$$V(x,y) = \frac{V_{SB} + V_d - V_I(x,y)}{sh[k_1 L]\sin[k_1 a]} sh[k_1(x-L)]\sin[k_1(y-a)]$$

Le même calcul au niveau du drain conduit à une formule équivalente.
Le calcul du potentiel dans l'état 2 est celui effectué dans le cas du canal long. On obtient au final.

$$V(x,y) = V_I(x,y) + \frac{V_{SB} + V_d - V_I(x,y)}{sh[k_1 L]\sin[k_1 a]} sh[k_1(x-L)]\sin[k_1(y-a)]$$
$$+ \frac{V_{DB} + V_d - V_I(x,y)}{sh[k_1 L]\sin[k_1 a]} sh[k_1 x]\sin[k_1(y-a)]$$

Dans la suite on cherche à calculer la valeur de la tension de seuil en canal court. On se place donc en régime de faible inversion. Le potentiel en canal long de l'état 2 est alors indépendant de *x*.
Il est possible de calculer la valeur minimale du potentiel global en fonction de la position *x.* on peut monter que ce minimum est atteint quand la tension V_{DS} est faible pour *x =L/2.*

On définit alors le seuil du transistor par la valeur de la tension de grille qui permet à la valeur minimale du potentiel de surface d'atteindre la valeur $2\phi_F + V_{SB}$. Pour justifier cette définition, il faut revenir à la définition de la tension de seuil (la charge d'inversion devient non nulle) et utiliser la relation 8-4.

On peut alors prouver la relation suivante pour des tensions V_{DS} faibles.

$$V_{GS} = V_{Tlong} + n_0\left(1 + \frac{1}{ch\left(k_1\frac{L}{2}\right)-1}\right)\left(V_{\min} - (2\phi_F + V_{SB})\right) - \frac{n_0}{ch\left(k_1\frac{L}{2}\right)-1}\left(-2\phi_F + V_d + \frac{V_{DS}}{2}\right)$$

Dans cette relation V_{Tlong} est la tension de seuil en canal long et n_0 est défini par

$$n_0 = 1 + \frac{\gamma}{2\sqrt{1{,}5\phi_F + V_{SB}}}$$

$$V_T = V_{FB} + 1{,}5\phi_F + V_{BS} + \gamma\sqrt{1.5\phi_F + V_{BS}}$$

On se place en régime de faible inversion.
Pour la valeur minimale du potentiel de surface on obtient donc

$$V_T = V_{Tlong} - \frac{n_0}{ch\left(k_1\frac{L}{2}\right)-1}\left(V_d - 2\phi_F + \frac{V_{DS}}{2}\right)$$

On en déduit alors facilement les deux composantes de la diminution de la tension de seuil, le *roll-off* et le *DIBL*.
Le roll-off est le terme indépendant de la tension V_{DS}. Il s'écrit

$$V_{Tlong} - V_T = \frac{n_0(V_d - 2\phi_F)}{ch(k_1 L/2)-1} \qquad (8\text{-}40)$$

Le DIBL dépend de la tension drain-source et s'exprime par

$$-\left(\frac{dV_T}{dV_{DS}}\right)_{V_{DS}=0} = \frac{n_0}{2}\frac{1}{ch(k_1 L/2)-1} \tag{8-41}$$

Il reste à exprimer le coefficient k_1 en fonction des paramètres physiques du transistor. Reprenons la relation 8-39.

$$\varepsilon_s \cos(ka)\sin(kt) + \varepsilon_{ox} \sin(ka)\cos(kt) = 0$$

On en déduit

$$0 = \varepsilon_s \tan(k_1 t) + \varepsilon_{ox} \tan(k_1 a)$$

Il est plus facile d'interpréter les effets de canal court si on exprime de manière plus simple le paramètre k_1 dans l'équation 8-39. Un développement limité de cette équation permet d'écrire quand l'épaisseur d'oxyde est mince

$$\frac{\pi}{k_1} = a + \frac{\varepsilon_s}{\varepsilon_{ox}} t - \frac{\pi^2}{3}\frac{\varepsilon_s}{\varepsilon_{ox}}\left(\frac{\varepsilon_s^{\,2}}{\varepsilon_{ox}^{\,2}} - 1\right)\frac{t^2}{a^2} t$$

Cette équation se simplifie pour les faibles valeurs de l'épaisseur d'oxyde

$$\frac{\pi}{k_1} = a + \frac{\varepsilon_s}{\varepsilon_{ox}} t \tag{8-42}$$

Les effets de variation de la tension de seuil sont d'autant plus faibles que k_1 est élevé. Il est donc nécessaire de diminuer le plus possible l'épaisseur de l'oxyde et l'étendue de la zone de charge d'espace. Une autre solution, très utilisée, est d'augmenter la constante diélectrique de l'oxyde ce qui explique l'intérêt des oxydes à haute permittivité en micro-électronique.

Pour terminer ce paragraphe, il est possible de noter les évolutions attendues pour les futures générations de transistors. Le but est de miniaturiser à l'extrême pour réduire le coût des circuits intégrés. Cette miniaturisation conduit cependant à une dégradation des propriétés du transistor : diminution du ratio entre le courant de conduction et le courant à l'état bloqué, augmentation de l'effet de la tension de drain sur la valeur du seuil, augmentation des courants de fuite et enfin augmentation des dispersions. La règle de conception appliquée dans les premiers temps de la micro-électronique était de maintenir le champ électrique constant dans le canal du transistor au fur et à mesure que la taille diminuait. Cette règle est difficilement applicable avec les technologies actuelles.

Les principales grandeurs physiques intervenant dans le fonctionnement du transistor sont : la longueur physique du canal (L), l'épaisseur de l'oxyde de grille (d_{OX}), le dopage du caisson (N_A), la profondeur de la zone de charge d'espace (y_b) des jonctions source-caisson et drain-caisson, la largeur (l) des contacts de source et de drain. Il est important de noter que la largeur (W) du transistor est considérée comme une variable indépendante et peut être choisie uniquement en fonction des impératifs du design, principalement en fonction de la valeur du courant de commutation . Il est assez naturel de penser que si la dimension du canal diminue d'un facteur α, la tension d'alimentation doit diminuer du même facteur pour garder le champ constant dans le canal. Le facteur α de réduction est égal à 1.37 quand on passe d'une génération à un autre, en pratique tous les 18 mois. Cette règle est appelée la loi de Moore. Elle n'a rien d'une loi car elle ne repose sur aucun principe physique. C'est une sorte de compromis entre la résolution des problèmes techniques liés à la réduction de taille et la nécessité économique de proposer des produits plus performants ou moins chers.

Si on admet que les contraintes de design imposent de réduire la largeur W du transistor de la même valeur pour conserver le rapport W/L constant, on en arrive donc à un dispositif dont la surface a été réduite d'un facteur α^2. Remarquons que les dimensions des contacts de source et de drain et donc l'encombrement de l'interconnexion doivent être réduits du même facteur. On peut donc réduire la surface d'un circuit d'un facteur α^2 en changeant de génération ce qui correspond à un facteur 2 environ. Il faut maintenant expliquer pourquoi il est nécessaire de réduire l'épaisseur de l'oxyde (d_{OX}), ce qui pose par ailleurs des difficultés technologiques considérables. Cette dimension influe directement sur la capacité par unité de surface C'_{OX} et sur la valeur du courant traversant le transistor en régime de saturation.

$$I_{D\,sat} = W\mu_n\, C'_{OX}\left[V_{GS} - V_T\right]E_C$$

Cette relation exprime le courant d'un transistor saturé en régime de canal court. La tension de saturation est négligée dans cette formule.

Les tensions diminuant d'un facteur α ainsi que la dimension W, le courant diminuerait en α^2 si l'épaisseur de l'oxyde était constante. Si on admet que les contraintes du design imposent que la diminution du courant de conduction soit au maximum d'un facteur α, il est donc nécessaire d'augmenter C'_{OX} d'un facteur α. Un argument possible pour justifier cette règle est de considérer la charge d'une capacité par le courant d'un transistor. La capacité diminuant d'un facteur α à cause de la réduction de taille, on peut tolérer que le courant de charge diminue du même facteur. Dans ce cas, la vitesse de commutation n'est pas réduite. En résumé, il est nécessaire d'augmenter la capacité C'_{OX} d'un facteur α ce qui conduit à diminuer l'épaisseur de l'oxyde du même facteur.

Expliquons l'évolution du dopage N_A du caisson du transistor. La profondeur de la zone de charge d'espace est donnée par la relation suivante

$$y_b = \sqrt{\frac{2\varepsilon_S}{eN_A}\Delta V}$$

La variation de potentiel est la différence entre le potentiel en surface et la partie profonde du silicium. Pour réduire cette dimension d'un facteur α et maintenir les mêmes conditions de fonctionnement du point de vue de l'électrostatique, il faut donc augmenter le dopage d'un facteur α puisque la variation de tension diminue d'un facteur α.

En pratique, la règle consistant à maintenir le champ constant dans le canal n'est plus strictement appliquée car elle conduirait à réduire les tensions en dessous d'un seuil acceptable pour le bon fonctionnement des circuits. Une autre raison pour modifier cette règle est l'augmentation du courant sous le seuil quand la tension de seuil diminue trop. Cet effet que nous avons déjà noté apparaît dans la relation suivante.

$$I_D = \frac{W}{L} k \exp^{-\frac{V_{GS}-V_t}{n\phi_t}} \left(1 - \exp^{-\frac{V_{DS}}{\phi_T}}\right)$$

La règle pratique appliquée est donc de diminuer la tension non pas en divisant par α mais en divisant par α/ε avec ε supérieur à un. La conséquence est une augmentation du champ électrique d'un facteur ε.

Les conséquences de la réduction de taille du transistor sont résumées dans le tableau ci-dessous qui exprime la variation des principaux paramètres physiques et électriques en fonction des coefficients α et ε. Rappelons que les deux coefficients sont supérieurs à 1.

	Règle standard	**Règle modifiée**
Longueur du canal	$1/\alpha$	$1/\alpha$
Épaisseur de l'oxyde de grille	$1/\alpha$	$1/\alpha$
Largeur du transistor (*W*)	$1/\alpha$	$1/\alpha$
Champ électrique	*1*	ε
Tension d'alimentation	$1/\alpha$	ε/α
Courant de conduction	$1/\alpha$	ε/α
Dopage du silicium	α	$\varepsilon\alpha$
Surface du transistor	$1/\alpha^2$	$1/\alpha^2$
Capacités du dispositif	$1/\alpha$	$1/\alpha$
Retard intrinsèque	$1/\alpha$	$1/\alpha$
Dissipation de puissance	$1/\alpha^2$	ε^2/α^2
Dissipation de puissance par unité de surface	*1*	ε^2

Les paramètres électriques comme le retard ou la puissance dissipée sont estimés au premier ordre de la manière suivante. Le retard intrinsèque est le temps de transit dans le canal c'est à dire le rapport entre la longueur du canal et la vitesse de saturation. La puissance consommée est la puissance dynamique, produit de la capacité du dispositif par le carré de la tension et par la fréquence de fonctionnement. La capacité varie en $1/\alpha$ mais la fréquence de fonctionnement varie en α.

Il est maintenant possible de donner les valeurs absolues de ces paramètres en se basant sur les prévisions de l'ITRS, organisation internationale regroupant tous les fabricants de semiconducteurs. Ce tableau donne à la fois des information techniques et économiques relatives à l'évolution des transistors et des circuits intégrés. On notera que la longueur du canal est plus faible que le nœud de la technologie, défini comme le demi pas minimum entre deux pistes conductrices. Le tableau suivant indique les prévisions faites par l'industrie microélectronique.

	2007	**2010**	**2013**	**2016**	**2019**
Nœud de la technologie (nm)	*65*	*45*	*32*	*22*	*16*
Longueur physique du canal (nm)	25	18	13	9	6
Tension d'alimentation (V)	1.1	1.0	0.9	0.8	0.7
Epaisseur électrique équivalente (nm)	1.8	1.1	1.0	0.9	0.9
Tension de seuil (V)	0.165	0.167	0.185	0.195	0.205
Courant de fuite (µA/µm)	0.20	0.22	0.11	0.11	0.11
Courant à l'état passant (µA/µm)	1200	1815	2220	2713	2744
Retard du NMOS – CV/I (ps)	0.640	0.400	0.250	0.150	0.100

8.5. Le fonctionnement dynamique du MOS

8.5.1 *Régime quasi-statique*

Comme il a été expliqué dans l'introduction, la première étape est de décrire le régime quasi-statique du transistor MOS. L'hypothèse principale est de supposer le courant de conduction constant dans le canal. Les expressions précédentes obtenues pour le calcul des charges seront donc utilisées à la différence près avec les paragraphes précédents qu'elles sont maintenant des fonctions du temps. Les raisonnements seront effectués en considérant la partie active du transistor, c'est-à-dire la zone de canal. Source, drain, grille et les paramètres électriques associés ne seront pas pris en compte dans cette analyse.

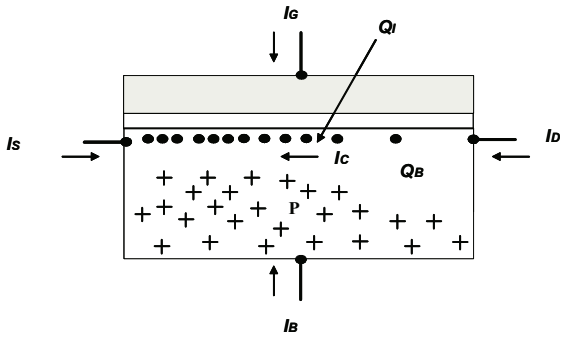


Figure 8-17 : Schéma de base pour l'analyse quasi-statique

On écrit les équations de conservation du courant en faisant la différence entre le courant de conduction I_C (diffusion et dérive des charges) et les courants transitoires I_{DV} et I_{SV} créés par les variations de charges dans le dispositif.

$$I_D(t) = I_C(t) + I_{DV}(t)$$

$$I_S(t) = -I_C(t) + I_{SV}(t)$$

Le régime quasi-statique suppose que la différence de courant est due à la variation de charge du canal. Les autres charges n'ont pas le temps de changer de manière significative. On écrit alors

$$I_{DV}(t) + I_{SV}(t) = \frac{dQ_I}{dt}$$

Il nous faut maintenant calculer ces deux composantes du courant. Pour faire ce calcul de manière rigoureuse, oublions les hypothèses du régime quasi-statique et considérerons le problème de manière générale en prenant une partie du canal de conduction de longueur *dx* et de largeur *W* comme le montre la figure 8-18. On néglige les courants ayant des directions différentes de celle du canal comme il été expliqué précédemment.

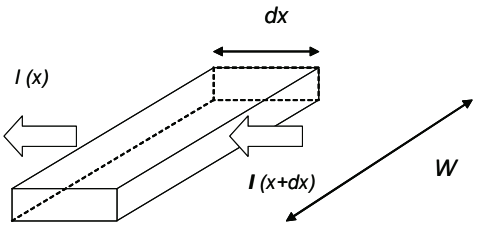


Figure 8-18 : Équation de continuité

Si on applique le principe de conservation de la charge, on peut écrire en fonction de la charge par unité de surface dans le canal

$$I(x+dx)-I(x)=\frac{\partial}{\partial t}\left(Q_I' W dx\right)$$

On en déduit donc,

$$\frac{\partial I(x,t)}{\partial x}=W\frac{\partial Q_I'(x,t)}{\partial t} \tag{8-43}$$

On peut également écrire dans le canal en négligeant le courant de diffusion, cas de la forte inversion.

$$I(x,t)=-\mu_n W Q_I'(x,t)\frac{\partial V_{CB}(x,t)}{\partial x}$$

Rappelons que le potentiel $V_{CB}(x,t)$ est la partie variable du potentiel de surface dans le canal (voir annexe 5).

$$V_s(x,t)=\phi_B+V_{CB}(x,t)$$

La résolution de l'équation (8-20) permet d'écrire

$$I(x,t)-I(0,t)=W\int_0^x\frac{\partial Q_I'(x',t)}{\partial t}dx'$$

Le courant $I(0,t)$ n'est autre que le courant $-I_S(t)$. La relation précédente peut donc s'écrire

$$I_S(t)=W\int_0^x\frac{\partial Q_I'(x',t)}{\partial t}dx'+\mu_n W Q_I'(x,t)\frac{\partial V_{CB}(x,t)}{\partial x}$$

On intègre alors cette équation de $x=0$ à $x=L$ et on obtient

$$I_S(t)L=W\int_0^L\int_0^x\frac{\partial Q_I'(x',t)}{\partial t}\,dx'dx+\mu_n W\int_0^L Q_I'(x,t)\,\frac{\partial V_{CB}(x,t)}{\partial x}dx$$

Cette équation s'écrit aussi

$$I_S(t)L=W\frac{\partial}{\partial t}\int_0^L\int_0^x Q_I'(x',t)\,dx'dx+\mu_n W\int_0^L Q_I'(x,t)\,\frac{\partial V_{CB}(x,t)}{\partial x}dx$$

En intégrant par parties la première expression de la forme $\int_0^L F(x)dx$, dans laquelle la fonction *F(x)* est définie par $\int_0^x Q'_I dx'$ on obtient

$$I_S(t)L = W\frac{\partial}{\partial t}\int_0^L (L-x)Q'_I(x,t)dx + \mu_n W\int_0^L Q'_I(x,t)\frac{\partial V_{CB}(x,t)}{\partial x}dx$$

soit,

$$I_S(t) = W\frac{\partial}{\partial t}\int_0^L \left(1-\frac{x}{L}\right)Q'_I(x,t)dx + \mu_n \frac{W}{L}\int_0^L Q'_I(x,t)\frac{\partial V_{CB}(x,t)}{\partial x}dx$$

Cette équation peut se comparer à l'équation établie en début de ce paragraphe pour le courant au niveau de la source. On obtient donc pour le courant au niveau de la source

$$I_S(t) = -I_C(t) + I_{SV}(t)$$

avec,

$$I_{SV}(t) = W\frac{\partial}{\partial t}\int_0^L \left(1-\frac{x}{L}\right)Q'_I(x,t)dx \qquad (8\text{-}44)$$

Ce courant est donc la partie additionnelle au courant de conduction vue au niveau de la source. En régime quasi-statique, la formule de la charge par unité de surface est supposée être la même qu'en régime purement statique. Elle est donc donnée par les relations établies dans les paragraphes précédents. En régime statique, ce terme additionnel est nul, les charges entrant dans le canal étant compensées par celles qui sortent.

De la même manière, on peut calculer le courant au niveau du drain, il s'écrit

$$I_D(t) = I_C(t) + I_{DV}(t)$$

avec,

$$I_{DV}(t) = W\frac{\partial}{\partial t}\int_0^L \frac{x}{L}Q'_I(x,t)dx \qquad (8\text{-}45)$$

Il reste à exprimer la charge d'inversion en fonction des tensions appliquées pour pouvoir écrire les courants transitoires calculés précédemment. En régime de forte inversion et de canal long, on obtient donc

$$Q'_I = \mu_n C'_{OX}\left(V_{GS} - V_{FB} - V_S - \gamma\sqrt{V_s}\right)$$

Rappelons que V_{FB} est défini par

$$V_{FB} = \phi_{MS} - \frac{Q'_0}{C'_{OX}}$$

On reprend l'approximation classique

$$\gamma\sqrt{V_s} \approx \gamma\sqrt{V_{SB} + 2\phi_F} + \delta\left(V_{CB}(x) - V_{SB}\right)$$

La valeur de la tension de saturation est alors

$$V_{DSsat} = \frac{V_{GS} - V_T}{1+\delta}$$

Tous ces résultats ont été établis dans le chapitre 5 et seuls les résultats utiles sont repris.
Si maintenant on introduit le paramètre α qui exprime la position de la tension de drain par rapport à la tension de saturation, on écrit

$$\alpha = 1 - \frac{V_{DS}}{V_{DSsat}} \tag{8-46}$$

Il est égal à 0 au début du régime saturé et égal à 1 quand le point de fonctionnement est éloigné du régime de saturation. Plus généralement il est une fonction de la tension entre drain et source.

Il est alors possible de calculer la charge d'inversion par unité de surface Q'_I puis la charge d'inversion totale Q_I en fonction du paramètre α. La charge d'inversion totale est donnée par la relation suivante.

$$Q_I = \int_0^L Q'_I(x) W dx$$

Comme la dépendance de la charge d'inversion en fonction de la position n'est pas connue, il est préférable d'exprimer les charges en fonction de la variation de potentiel de surface. Pour cela on utilise la relation donnant le courant de

conduction en supposant qu'il est exclusivement créé par la dérive des électrons en régime de forte inversion.

$$I_D dx = -\mu_n W Q'_I dV_{CB}(x)$$

On écrit alors.

$$Q_I = -\frac{\mu_n W^2}{I_D}\int_{V_{SB}}^{V_{DB}} Q'^2_I dV_{CB}$$

Le calcul est un peu long et n'est pas détaillé dans cet ouvrage

$$Q_I = -WLC'_{OX}(V_{GS} - V_T)\frac{2}{3}\frac{1+\alpha+\alpha^2}{1+\alpha} \tag{8-47}$$

On peut calculer de la même manière les autres charges du dispositif.

$$Q_B = -WLC'_{OX}\left(\gamma\sqrt{\phi_B + V_{SB}} + \frac{\delta}{1+\delta}(V_{GS} - V_T)\left(1 - \frac{2}{3}\frac{1+\alpha+\alpha^2}{1+\alpha}\right)\right) \tag{8-48}$$

On en déduit la charge Q_G en fonction de la charge Q_0 de l'interface.

On peut également calculer Q_S et Q_D puis les expressions des courants de transition I_{SV} et I_{DV}.

$$I_{SV}(t) = W\frac{\partial}{\partial t}\int_0^L\left(1 - \frac{x}{L}\right)Q'_I(x,t)dx$$

On obtient après des calculs assez longs non détaillés dans cet ouvrage.

$$I_{SV}(t) = -\frac{\partial}{\partial t}WLC'_{OX}(V_{GS} - V_T)\frac{6+12\alpha+8\alpha^2+4\alpha^3}{15(1+\alpha)^2}$$

$$I_{DV}(t) = -\frac{\partial}{\partial t}WLC'_{OX}(V_{GS} - V_T)\frac{4+8\alpha+12\alpha^2+6\alpha^3}{15(1+\alpha)^2}$$

Résumons les résultats de ce paragraphe. Tout se passe comme si les courants de transition étaient dus à des charges « virtuelles », Q_S et Q_D variant dans le temps. Ces charges sont respectivement

$$Q_S(t) = W\int_0^L \left(1 - \frac{x}{L}\right) Q_I'(x,t)dx \tag{8-49}$$

$$Q_D(t) = W\int_0^L \frac{x}{L} Q_I'(x,t)dx \tag{8-50}$$

Les calculs analytiques pour obtenir les expressions des courants en fonction du temps sont complexes. En pratique, on se base sur les résultats donnés par les simulateurs électriques. Il est cependant utile de pouvoir fixer quelques ordres de grandeur et surtout d'être conscient des hypothèses de calcul.

En régime de faible inversion, les calculs sont plus simples puisque le potentiel de surface est constant et indépendant de la position.

$$V_s(0) = V_s(L) = \left(-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + V_{GB} - V_{FB}}\right)^2$$

Les résultats du chapitre 3.2 permettent de calculer la charge d'inversion par unité de surface. On constate qu'elle varie quasi-linéairement entre la source et le drain. Les charges virtuelles de source et de drain se calculent facilement

$$Q_S = WL\left(\frac{Q'_{I\,source}}{3} + \frac{Q'_{I\,drain}}{6}\right)$$

$$Q_D = WL\left(\frac{Q'_{I\,source}}{6} + \frac{Q'_{I\,drain}}{3}\right)$$

En pratique, ces charges sont négligeables en régime de faible inversion devant les charges extérieures au canal et elles ne seront pas prises en compte dans la modélisation dynamique.

8.5.2. *Régime dynamique*

Quelles sont les conditions de validité du régime quasi-statique ? Il faut que les distributions de charges s'établissent plus vite que les variations de tension. En pratique, on donne souvent la condition suivante : le temps de montée de la tension d'entrée (en général, la tension de grille) doit être au moins vingt fois plus important que le temps de transit des charges dans le canal. Le temps de transit τ est défini comme suit

$$\tau = \frac{|Q_I|}{I_D}$$

En régime de forte inversion et en régime saturé, on trouve alors

$$Q_I = -WLC'_{OX}\left(V_{GS} - V_T\right)\frac{2}{3}$$

$$I_{Dsat} = \frac{W}{L}\mu_n C'_{OX}\frac{\left(V_{GS} - V_T\right)^2}{2\left(1+\delta\right)}$$

On en déduit

$$\tau = \left(1+\delta\right)\frac{L^2}{\mu_n\left(V_{GS} - V_T\right)}\cdot\frac{4}{3} \qquad (8\text{-}51)$$

Quelques ordres de grandeur peuvent être donnés.

– Pour une technologie ancienne :
L = 0.8 µm, V_{GS}-V_T = 3 V, μ_n = 64 µm^2/V. ns alors τ = 3 ps
– Pour une technologie moderne :
L =0.05 µm, V_{GS}-V_T = 1 V, μ_n = 64 µm^2/V. ns alors τ = 0.04 ps

Ces ordres de grandeur montrent que le domaine d'application du régime quasi-statique est très large. Revenons cependant à la manière la plus générale de traiter le comportement dynamique d'un MOSFET en reprenant les équations de base

$$\frac{\partial I\left(x,t\right)}{\partial x} = W\frac{\partial Q'_I\left(x,t\right)}{\partial t}$$

$$I\left(x,t\right) = -\mu_n W Q'_I\left(x,t\right)\frac{\partial V_{CB}\left(x,t\right)}{\partial x}$$

$$Q'_I = \mu_n C'_{OX}\left(V_{GB} - V_{FB} - V_S - \gamma\sqrt{V_s}\right)$$

$$V_s = \phi_B + V_{CB}\left(x,t\right)$$

Ce système de quatre équations permet en théorie de trouver les quatre inconnues du problème : $I(x,t)$,$Q'(x,t)$, $V_{CB}(x,t)$ et $V_s(x,t)$. Il est nécessaire pour résoudre ces équations de disposer d'un jeu de conditions initiales. Une étude plus détaillée des équations de base pourrait montrer que la grandeur $V_{CB}(x,t)$ est en fait la différence entre le pseudo-potentiel chimique des électrons dans le canal et le pseudo-potentiel chimique des trous en profondeur dans le matériau (annexe 5).

8.6. Les modèles du transistor MOSFET

Ce paragraphe est une sorte de résumé des paragraphes précédents. Il donne en plus les bases pour constituer un ensemble d'équations utilisables par un logiciel de simulation. Cet ensemble d'équations est un modèle. Les modèles sont définis en fonction de leur domaine d'application. On distingue classiquement les trois domaines suivants :

- Le domaine statique : les courants sont exprimés en fonction des tensions appliquées.
- Le domaine dynamique dit petits signaux : ce ne sont plus les courants qui sont exprimés mais les variations de ces courants pour des variations faibles des tensions. Les courants considérés sont les courants qui entrent dans le dispositif : courant de source, courant de grille, courant de drain et courant de bulk.
- Le domaine dynamique dit petits signaux et haute fréquence. Les effets supplémentaires à prendre en compte sont d'une part les éléments parasites négligés dans les représentations précédentes (capacités et self-inductances) et d'autre part les effets liés à la représentation spatiale des phénomènes électriques. Les dispositifs sont alors représentés comme des ensembles comportant des éléments répartis dans l'espace et la dimension spatiale compte. Notons que la modélisation du comportement du canal de conduction s'inscrit dans ce mode de représentation

8.6.1. *Rappels du modèle statique*

Le tableau ci-dessous illustre les différents cas possibles traités dans les paragraphes précédents.

	Faible inversion	**Forte inversion non saturé**	**Forte inversion saturé**
Canal long	A	B	C
Canal court	D	E	F
Haute fréquence	G	H	I

Reprenons dans chaque cas l'expression du courant de drain en fonction des tensions appliquées.

- **Cas A** : Le courant de drain s'écrit comme suit

$$I_D = -\frac{W}{L}\mu_n \phi_T Q'_I(0)\left(1 - e^{-\frac{V_{DS}}{\phi_T}}\right)$$

$$I_D = I_S \exp^{\frac{V_{GS}-V_T}{n_0\phi_t}}\left(1-\exp^{-\frac{V_{DS}}{\phi_t}}\right)$$

- **Cas B** : Le courant de drain s'écrit donc

$$I_D = \frac{W}{L}\mu_n C'_{OX}\left[\left(V_{GS}-V_T\right)V_{DS}-\frac{1}{2}\left(1+\delta\right)V_{DS}{}^2\right]$$

$$\delta = \frac{\gamma}{2\sqrt{2\phi_F + V_{SB}}}$$

- **Cas C** : Le courant de drain s'écrit

$$I_{Dsat} = \frac{W}{L}\mu_n C'_{OX}\frac{\left(V_{GS}-V_T\right)^2}{2\left(1+\delta\right)}$$

- **Cas D** : Le courant de drain est alors

$$I_D = -\frac{W}{L}\mu_n\phi_t Q'_I(0)\left(1-\exp^{-\frac{V_{DS}}{\phi_t}}\right)$$

$$I_D = I_S \exp^{\frac{V_{GS}-V_T}{n_0\phi_t}}\left(1-\exp^{-\frac{V_{DS}}{\phi_t}}\right)$$

- **Cas E** : Le courant de drain s'écrit comme suit

$$I_D = \frac{W}{L}\left(\frac{1}{1+\frac{V_{DS}}{LE_C}}\right)\mu_n\, C'_{OX}\left[\left(V_{GS}-V_T\right)V_{DS}-\frac{1}{2}\left(1+\delta\right)V_{DS}{}^2\right]$$

- **Cas F** : L'expression du courant de drain se simplifie

$$I_{D\,sat} = W\mu_n C'_{OX}\left[V_{GS}-V_T-\frac{1+\delta}{2}V_{DSsat}\right]E_C$$

Les autres cas, cas de la haute fréquence, ne donnent pas de valeurs différentes du courant.

8.6.2. *Modèles petits signaux : généralités*

On s'intéresse aux variations des grandeurs électriques autour d'un point de polarisation donné. La méthode est adaptée à la description des circuits analogiques qui manipulent en général des signaux de faibles amplitudes. Elle est par extension également utilisée en numérique bien que les variations soient importantes. Le principe est simple : il suffit de dériver le courant par rapport aux tensions de commande pour avoir des générateurs de courant équivalents. Prenons l'exemple du courant de drain dans le mode non saturé.

$$I_D = \frac{W}{L}\mu_n C'_{OX}\left[\left(V_{GS} - V_T\right)V_{DS} - \frac{1}{2}\left(1+\delta\right){V_{DS}}^2\right]$$

Si on superpose à la tension V_{GS} une variation de faible amplitude v_{gs}, on obtient

$$I_D + i_d = \frac{W}{L}\mu_n C'_{OX}\left[\left(V_{GS} + v_{gs} - V_T\right)V_{DS} - \frac{1}{2}\left(1+\delta\right){V_{DS}}^2\right]$$

La valeur de i_d s'obtient par dérivation

$$i_d = \left(\frac{\partial I_D}{\partial V_{GS}}\right)_{\substack{V_{GS}=V_{GS0} \\ V_{DS}=V_{DS0}}} \cdot v_{gs}$$

Ces données ne suffisent pas pour obtenir un schéma électrique. Il faut également tenir compte des courants de transition que nous avons introduits dans le modèle quasi-statique du transistor. Enfin, il faut tenir compte des éléments parasites (résistances, condensateurs et inductances). L'établissement d'un modèle est donc une opération assez complexe et particulièrement si on impose de passer d'un régime de fonctionnement à un autre sans discontinuité.

Nous procéderons de manière itérative, tout d'abord en considérant la partie centrale du transistor à basse fréquence. Ensuite, nous étendrons cette description aux fréquences intermédiaires, ensuite nous introduirons les éléments parasites et finalement nous donnerons quelques éléments sur la modélisation à haute fréquence.

8.6.3. *Le MOS interne en basse fréquence*

Les équations conduisent naturellement au schéma suivant. Les courants de transition sont supposés nuls.

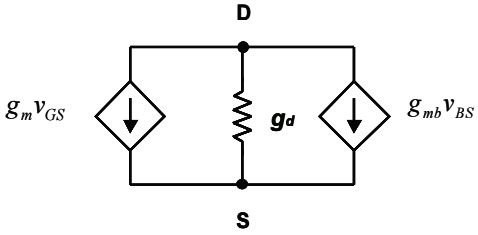


Figure 8-19 : Le modèle basse fréquence interne

Les coefficients du modèle sont obtenus par simple dérivation

$$g_m = \left(\frac{\partial I_D}{\partial V_{GS}} \right)_{V_{BS}, V_{DS}}$$

La notation indique que les autres différences de potentiel (V_{DS} et V_{BS}) restent constantes. On obtient alors en fonction des régimes de fonctionnement.

- **Cas A et D : faible inversion et canal long ou court**

$$g_m = \frac{1}{n} \frac{I_D}{\phi_t} \tag{8-52}$$

On obtient de même pour le paramètre g_{mb}

$$g_{mb} = \left(\frac{\partial I_{DS}}{\partial V_{BS}} \right)_{V_{GS}, V_{DS}}$$

On obtient après quelques calculs

$$g_{mb} \approx (n-1) g_m \approx \frac{\varepsilon_s}{\varepsilon_{ox}} \frac{d_{ox}}{y_B} g_m \tag{8-53}$$

Dans cette relation, y_B est la profondeur moyenne de la zone de charge d'espace.

On peut également calculer la conductance de sortie.

$$g_d = \left(\frac{\partial I_{DS}}{\partial V_{DS}} \right)_{V_{GS}, V_{BS}}$$

$$g_d = \frac{\exp^{-\frac{V_{DS}}{\phi_t}}}{1-\exp^{-\frac{V_{DS}}{\phi_t}}} \frac{I_D}{\phi_t} \tag{8-54}$$

La conductance de sortie décroît rapidement quand la tension de drain augmente.

- **Cas B et C : forte inversion, canal long, régime non saturé et saturé**

Les mêmes calculs sont repris avec une expression différente pour le courant de drain.

$$g_m = \frac{W}{L}\mu_n C'_{OX} V_{DS} \qquad \text{si} \qquad V_{DS} \leq V_{DSsat}$$
$$g_m = \frac{W}{L}\mu_n C'_{OX} V_{DSsat} \qquad \text{si} \qquad V_{DS} \succ V_{DSsat}$$

En régime saturé et uniquement dans ce cas, on peut écrire

$$g_m = \frac{W}{L}\mu_n\, C'_{OX} \frac{V_{GS}-V_T}{1+\delta} \qquad \text{si} \quad V_{DS} \succ V_{DSsat}$$

$$g_m = \frac{2I_D}{V_{GS}-V_T} \qquad \text{si } V_{DS} \succ V_{DSsat} \tag{8-55}$$

Cette dernière formule est d'un usage très courant car elle relie transconductance et courant consommé. On peut également écrire en exprimant la différence de tension V_{GS}-V_T en fonction du courant de drain

$$g_m{}^2 = \frac{W}{L}\mu_n \frac{C'_{OX}}{1+\delta} 2I_D$$

On note

$$\beta_n = \frac{W}{L}\mu_n C'_{OX}$$

alors

$$g_m = \sqrt{2\beta_n \frac{I_D}{1+\delta}} \qquad \text{si } V_{DS} \succ V_{DSsat} \tag{8-56}$$

Cette formule est très utilisée, souvent en négligeant δ **.** Rappelons qu'elle est valable uniquement en régime saturé et pour un MOS à canal long.

Le paramètre g_{mb} est plus difficile à calculer de manière rigoureuse. Nous nous contenterons de donner le résultat.

$$\frac{g_{mb}}{g_m} = \gamma \frac{\sqrt{V_{DS} + V_{SB} + \phi_B} - \sqrt{V_{SB} + \phi_B}}{V_{DS}} \qquad \text{si} \quad V_{DS} \leq V_{DSsat}$$

$$\frac{g_{mb}}{g_m} = \gamma \frac{\sqrt{V_{DSsat} + V_{SB} + \phi_B} - \sqrt{V_{SB} + \phi_B}}{V_{DSsat}} \qquad \text{si} \quad V_{DS} \succ V_{DSsat}$$

Calculons maintenant la conductance g_d. C'est la dérivée du courant de sortie par rapport à la tension de drain, la tension de grille étant donnée.

La formule du courant en régime non saturé conduit à

$$g_d = \frac{W}{L} \mu C'_{OX} \left[V_{GS} - V_T - (1+\delta) V_{DS} \right]$$

Quand le régime de saturation est atteint le courant reste constant et la transconductance est nulle. Il faut alors revenir aux résultats du paragraphe 4.1 pour expliquer l'origine de la conductance de sortie et pour en déduire sa valeur.

$$I_D = I_{D_{sat}} \left(1 + \frac{V_{DS} - V_{DS_{sat}}}{V_A} \right)$$

On en déduit,

$$g_d = \frac{I_{Dsat}}{V_A} \tag{8-57}$$

Rappelons que le paramètre V_A est égal à $kL\sqrt{N_A}$, formule dans laquelle k vaut environ 0.15 V.µm^2.

- **Cas E et F : canal court, forte inversion, régime saturé et non saturé**

En régime saturé, on obtient simplement :

$$g_m = W C'_{OX} \mu_n E_C \tag{8-58}$$

La formule approchée est particulièrement simple. La transconductance ne dépend plus que de la largeur du MOS au premier ordre. On néglige dans ce calcul tiré de la relation 8.36 la variation de la tension de saturation avec la tension de grille.

La conductance de sortie est due à la variation de la longueur effective de canal et est donc donnée par la formule établie dans les cas B et C.

$$g_d = \frac{I_{Dsat}}{V_A}$$

On peut également montrer que dans le régime de canal court la relation suivante lie transconductance et conductance de sortie.

$$\frac{g_d}{g_m} \approx 0.5 \frac{\varepsilon_s}{\varepsilon_{ox}} \frac{d_{OX}}{L} \tag{8-59}$$

Cette relation est d'un grand intérêt car elle illustre la compétition entre la grille et le drain pour commander le transistor. Les influences des permittivités et des dimensions sont évidentes. Pour réaliser un dispositif commandé par la grille, il faut minimiser ce ratio et donc réduire l'épaisseur de l'oxyde ou bien augmenter la permittivité de l'oxyde de grille. Cette remarque explique les efforts accomplis actuellement pour développer de nouveaux oxydes de permittivité plus élevée que celle de la silice.

8.6.4. *Le MOS interne à fréquence moyenne*

Un schéma électrique équivalent est une représentation des équations donnant les courants aux quatre bornes de sortie du transistor. Ces courants sont en fait les petites variations des courants globaux, ils sont notés avec des minuscules pour bien faire la différence. La méthode consiste à exprimer chaque variation par rapport aux tensions appliquées puis à sommer les contributions. Si nous reprenons maintenant les courants de source et de drain, rappelons les relations suivantes

$$I_D(t) = I_C(t) + I_{DV}(t) \tag{8-60-a}$$

$$I_S(t) = -I_C(t) + I_{SV}(t) \tag{8-60-b}$$

Rappelons que le courant I_c est le courant de conduction, somme du courant de diffusion et du courant de dérive. Les courants sont composés d'une partie continue et d'une partie variable autour de cette valeur continue et de faible amplitude appelée composante petit signal. On peut donc écrire

$$I_D + i_d(t) = I_C + i_c(t) + i_{dv}(t) \tag{8-61-a}$$

$$I_S + i_s(t) = -I_C - i_c(t) + i_{sv}(t) \tag{8-61-b}$$

Comme les courants I_{DV} et I_{SV} ne comportent pas de composantes continues puisqu'ils sont des courants liés à des variations de charge, ils seront notés i_{dv} et i_{sv} dans la suite de l'analyse. L'écriture des variations demande quelques précisions. On superpose aux valeurs continues des tensions appliquées au dispositif des valeurs variables dans le temps et de faibles valeurs. Ces valeurs sont notées avec des minuscules comme par exemple v_g et on écrit

$$V_G(t) = V_G + v_g(t)$$

La fonction $v_g(t)$ peut être une fonction sinusoïdale par exemple. On écrit de manière évidente

$$\frac{dV_G}{dt} = \frac{dv_g}{dt}$$

Le dispositif est représenté figure 8.17. Il ne prend en compte que le canal et assimile source et drain à de simples contacts ponctuels.

On exprime alors les variations de courant de drain en séparant les parties conduction et transition. Pour compléter l'analyse du paragraphe 8-5 qui s'intéressait aux courants transitoires i_{dv} et i_{sv}, nous allons également calculer les courants transitoires de grille et de bulk.

Nous allons dans une première étape nous intéresser à un modèle complet. Un modèle complet est l'expression des variations des quatre courants de transition entrant dans le dispositif en fonction des variations des quatre tensions de commande ce qui conduit à 16 capacités de couplage. Ce modèle global s'écrit

$$i_{dv}(t) = C_{dd}\frac{dv_d}{dt} - C_{dg}\frac{dv_g}{dt} - C_{db}\frac{dv_b}{dt} - C_{ds}\frac{dv_s}{dt}$$

$$i_g(t) = -C_{gd}\frac{dv_d}{dt} + C_{gg}\frac{dv_g}{dt} - C_{gb}\frac{dv_b}{dt} - C_{gs}\frac{dv_s}{dt}$$

$$i_b(t) = -C_{bd}\frac{dv_d}{dt} - C_{bg}\frac{dv_g}{dt} + C_{bb}\frac{dv_b}{dt} - C_{bs}\frac{dv_s}{dt}$$

$$i_{sv}(t) = -C_{sd}\frac{dv_d}{dt} - C_{sg}\frac{dv_g}{dt} - C_{sb}\frac{dv_b}{dt} + C_{bs}\frac{dv_s}{dt} \qquad (8\text{-}62)$$

Le courant i_c est la partie variable du courant de conduction. Il est également à considérer.
Il se calcule à partir de l'expression du courant continu par dérivation par rapport aux variations des tensions de grille et de bulk.

En fait, neuf capacités sont indépendantes. Il y a en effet quatre relations du type suivant :

$$C_{gg} = C_{gs} + C_{gd} + C_{gb}$$

Cette relation se démontre à partir de l'expression générale du courant de grille en considérant le système particulier dans lequel les tensions varient mais sont toutes égales. Le courant transitoire de grille, dans ce cas, est nécessairement nul ce qui établit la relation. On obtient les trois autres équations en considérant les trois autres courants.

$$C_{dd} = C_{dg} + C_{db} + C_{ds}$$

$$C_{bb} = C_{bd} + C_{bg} + C_{bs}$$

$$C_{ss} = C_{sd} + C_{sg} + C_{sb}$$

On peut également obtenir quatre relations en écrivant que la somme des courants de transition entrant dans le dispositif est nulle. En effet, la somme des courants globaux est nulle et la somme des courants de conduction est également nulle. Dans le cas particulier où toutes les tensions sont fixes sauf la tension de drain, on obtient à partir des équations générales

$$\frac{dv_d}{dt}\left(C_{dd} - C_{gd} - C_{bd} - C_{sd}\right) = 0$$

Comme cette relation est vérifiée pour toute valeur de la tension de drain, on écrit

$$C_{dd} - C_{gd} - C_{bd} - C_{sd} = 0$$

De même, on obtient :

$$C_{gg} - C_{dg} - C_{bg} - C_{sg} = 0$$

$$C_{bb} - C_{db} - C_{gb} - C_{sb} = 0$$

$$C_{ss} - C_{ds} - C_{gs} - C_{bs} = 0$$

On peut aussi noter que la connaissance de trois courants suffit puisque le quatrième s'en déduit par la condition de nullité de la somme. De plus, les relations intéressantes font intervenir non pas les tensions absolues v_g, v_d, v_s, v_b mais les tensions mesurées relativement à la source soient v_{ds}, v_{gs} et v_{bs}. Par exemple, la tension v_{ds} s'exprime par

$$v_{ds} = v_d - v_s$$

On en arrive facilement au jeu d'équations suivantes

$$\begin{aligned} i_{dv}(t) &= C_{dd}\frac{dv_{ds}}{dt} - C_{dg}\frac{dv_{gs}}{dt} - C_{db}\frac{dv_{bs}}{dt} \\ i_{g}(t) &= -C_{gd}\frac{dv_{ds}}{dt} + C_{gg}\frac{dv_{gs}}{dt} - C_{gb}\frac{dv_{bs}}{dt} \\ i_{b}(t) &= -C_{bd}\frac{dv_{ds}}{dt} - C_{bg}\frac{dv_{gs}}{dt} + C_{bb}\frac{dv_{bs}}{dt} \end{aligned} \tag{8-63}$$

Ce jeu d'équations suffit à résoudre notre problème puisque le courant $i_{sv}(t)$ s'exprime en fonction des trois autres. Il est maintenant possible de faire correspondre à ces équations un schéma équivalent. L'écriture des équations de Kirchhoff aux nœuds de ce schéma doit alors conduire au même jeu d'équations. C'est le critère de validité du modèle. Cette manière de procéder est rigoureuse mais conduit à un schéma complexe peu utilisé en pratique. On lui préfère un schéma plus simple.

En effet, on peut négliger un certain nombre de termes et aboutir au schéma classique de la figure 8-20. Les hypothèses sont les suivantes.

$$\begin{aligned} C_{gd} &= C_{dg} \\ C_{db} &= C_{bd} \\ C_{bg} &= C_{gb} \\ C_{sd} &= 0 \end{aligned} \tag{8-64}$$

On obtient alors à partir de 8-63 en intégrant les relations entre coefficients

$$\begin{aligned} i_{dv}(t) &= C_{gd}\frac{dv_{dg}}{dt} + C_{bd}\frac{dv_{db}}{dt} \\ i_{g}(t) &= C_{gd}\frac{dv_{gd}}{dt} + C_{gb}\frac{dv_{gb}}{dt} + C_{gs}\frac{dv_{gs}}{dt} \\ i_{b}(t) &= C_{bd}\frac{dv_{bd}}{dt} + C_{gb}\frac{dv_{bg}}{dt} + C_{bs}\frac{dv_{bs}}{dt} \end{aligned} \tag{8-65}$$

Ces équations sont équivalentes au schéma 8-20. Rappelons qu'un schéma équivalent n'est que la représentation d'un système d'équations à la condition d'appliquer les lois de Kirchhoff sur la conservation du courant. Il faut ajouter le courant i_c et g_d pour représenter complètement le dispositif. Les courants symbolisés par les sources liées ($g_m v_{gs}$ *et* $g_{mb} v_{bs}$) sont les parties variables du courant de conduction.

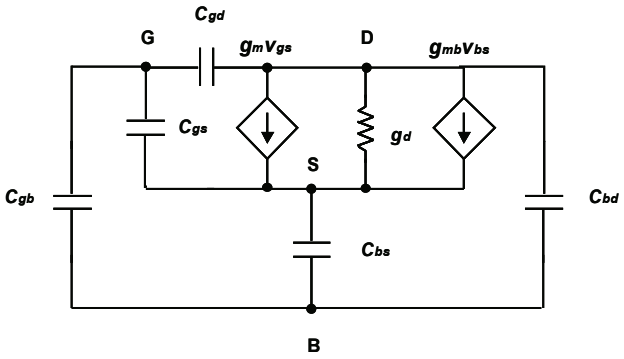


Figure 8-20 : Schéma équivalent petits signaux

Pour calculer les capacités du modèle, il faut maintenant exprimer des relations entre les courants transitoires et les charges. Revenons au schéma global du transistor MOS en oubliant provisoirement source et drain comme le montre la figure 8-17. Ce schéma a pour but d'expliquer les variations des charges stockées dans le dispositif, à savoir la charge de grille Q_G et la charge de la zone de charge d'espace Q_B. Ces charges varient quand les tensions appliquées sont modifiées pendant *dt*. Prenons l'exemple de la charge accumulée sur la grille. Quand elle varie sous l'effet d'une variation de l'une des tensions appliquées au système, un courant transitoire $i_g(t)$ circule dans le fil de grille. Pour s'en convaincre, on applique le principe de conservation de la charge à la grille.

$$i_g(t) = \frac{dQ_G}{dt}$$

On peut alors considérer les variations induites par les variations de V_S,V_B,V_G et V_D sur la charge de grille. Les tensions sont mesurées à partir d'une référence quelconque.

$$i_g(t) = \frac{\partial Q_G}{\partial V_S}\frac{dv_S}{dt} + \frac{\partial Q_G}{\partial V_B}\frac{dv_B}{dt} + \frac{\partial Q_G}{\partial V_D}\frac{dv_D}{dt} + \frac{\partial Q_G}{\partial V_G}\frac{dv_G}{dt}$$

On pose alors

$$C_{gg} = +\left(\frac{\partial Q_G}{\partial V_G}\right)_{V_S,V_D,V_B}$$

$$C_{gs} = -\left(\frac{\partial Q_G}{\partial V_S}\right)_{V_G,V_D,V_B}$$

$$C_{gd} = -\left(\frac{\partial Q_G}{\partial V_D}\right)_{V_S,V_G,V_B}$$

$$C_{gb} = -\left(\frac{\partial Q_G}{\partial V_B}\right)_{V_S,V_G,V_D}$$

Le même calcul peut se faire en considérant la zone de bulk incluant la zone de charge d'espace de charge Q_B mais excluant la zone de canal. Le seul courant entrant dans cette zone est le courant transitoire de bulk. Rappelons une nouvelle fois que les courants de diffusion et de dérive sont répartis tout le long du canal et perpendiculaires à celui-ci. Ils se compensent et n'interviennent donc pas dans l'expression du courant total. On en déduit donc les capacités équivalentes

$$C_{bb} = +\left(\frac{\partial Q_B}{\partial V_B}\right)_{V_G,V_D,V_S}$$

$$C_{bs} = -\left(\frac{\partial Q_B}{\partial V_S}\right)_{V_G,V_D,V_B}$$

$$C_{bd} = -\left(\frac{\partial Q_B}{\partial V_D}\right)_{V_G,V_S,V_B}$$

$$C_{bg} = -\left(\frac{\partial Q_B}{\partial V_G}\right)_{V_D,V_S,V_B}$$

Cette dernière capacité n'est pas nécessairement égale à C_{gb}. Cette remarque importante illustre bien le fait que les capacités introduites sont des capacités équivalentes exprimant les influences électriques dans le dispositif et ne sont pas les capacités de condensateurs géométriquement identifiés. Pour établir le schéma 8-20, il a cependant été nécessaire de supposer C_{gb}=C_{bg}.

Au niveau du drain et de la source, les relations entre courants transitoires et charges sont beaucoup moins évidentes et ont conduit à définir les charges équivalentes de source et de drain comme il a été expliqué dans le paragraphe 8.5.

Le calcul de ces capacités équivalentes peut se faire en fonction des expressions analytiques des charges dans les différents régimes de fonctionnement. Rappelons que l'hypothèse quasi-statique conduit à supposer que les expressions établies en régime continu sont encore valables. Pour obtenir la valeur des capacités équivalentes, il faut calculer les charges totales en fonction des charges par unité de surface. Prenons l'exemple de la charge de grille en inversion forte.

$$Q_G = W\int_0^L Q'_G dx$$

Comme,

$$I_D = \mu_n W\left(-Q'_I\right)\frac{dV_{CB}(x)}{dx}$$

On exprime la variation spatiale en fonction de la variation de potentiel.

$$dx = -\frac{\mu_n W}{I_D} Q'_I \; dV_{CB}(x)$$

On écrit alors

$$Q_G = -\frac{\mu_n W^2}{I_D}\int_{V_{SB}}^{V_{DB}} Q'_G Q'_I dV_{CB}$$

On calcule de même Q_B et Q_I. Les formules approchées du régime de forte inversion conduisent aux valeurs suivantes

$$Q_B = -WLC'_{OX}\left[\gamma\sqrt{\phi_B + V_{SB}} + \frac{\delta}{1+\delta}\left(V_{GS} - V_T\right)\left(1 - \frac{2}{3}\frac{1+\alpha+\alpha^2}{1+\alpha}\right)\right]$$

$$Q_I = -WLC'_{OX}\left(V_{GS} - V_T\right)\frac{2}{3}\frac{1+\alpha+\alpha^2}{1+\alpha} \qquad (8\text{-}66)$$

$$Q_G = WLC'_{OX}\left[\gamma\sqrt{\phi_B + V_{SB}} + \frac{V_{GS} - V_T}{1+\delta}\left(\delta + \frac{2}{3}\frac{1+\alpha+\alpha^2}{1+\alpha}\right)\right] - Q_0$$

On en déduit donc les capacités équivalentes en négligeant les dérivées de δ par rapport aux tensions V_S et V_B. Les calculs sont assez fastidieux et ne sont pas détaillés.

$$C_{gs} = -\left(\frac{\partial Q_G}{\partial V_S}\right)_{V_G,V_D,V_B} = \frac{2}{3}C'_{OX}WL\frac{1+2\alpha}{(1+\alpha)^2}$$

$$C_{bs} = -\left(\frac{\partial Q_B}{\partial V_S}\right)_{V_G,V_D,V_B} = \delta\frac{2}{3}C'_{OX}WL\frac{1+2\alpha}{(1+\alpha)^2}$$

$$C_{gd} = -\left(\frac{\partial Q_G}{\partial V_D}\right)_{V_G,V_S,V_B} = \frac{2}{3}C'_{OX}WL\frac{\alpha^2+2\alpha}{(1+\alpha)^2} \qquad (8\text{-}67)$$

$$C_{bd} = -\left(\frac{\partial Q_B}{\partial V_D}\right)_{V_G,V_S,V_B} = \delta\frac{2}{3}C'_{OX}WL\frac{\alpha^2+2\alpha}{(1+\alpha)^2}$$

$$C_{gb} = -\left(\frac{\partial Q_G}{\partial V_B}\right)_{V_G,V_S,V_B} = \frac{\delta}{3(1+\delta)}C'_{OX}WL\left(\frac{1-\alpha}{1+\alpha}\right)^2$$

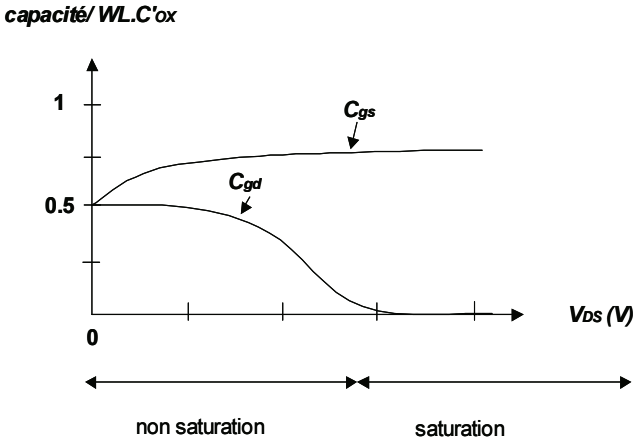


Figure 8-21 : Capacités équivalentes en fonction de la tension de drain

Rappelons qu'en régime saturé, le paramètre α est nul.

Il est maintenant possible de représenter les variations de ces capacités en fonction de la tension de drain et donc du régime de fonctionnement. La figure 8.21 montre les variations des capacités équivalentes les plus importantes pour un transistor en régime de conduction. Les autres capacités ayant des valeurs plus faibles ne sont pas représentées.

Il est intéressant de noter que la capacité grille-source passe de $1/2WLC'_{ox}$ à $2/3WLC'_{ox}$ quand on passe du régime non saturé au régime saturé. Les capacités grille-drain et bulk-drain tendent vers 0 en régime saturé. La tension de drain n'a en effet plus d'action sur la valeur des charges du dispositif. La variation des capacités équivalentes peut également se représenter en fonction de la tension de grille pour une tension de drain donnée.

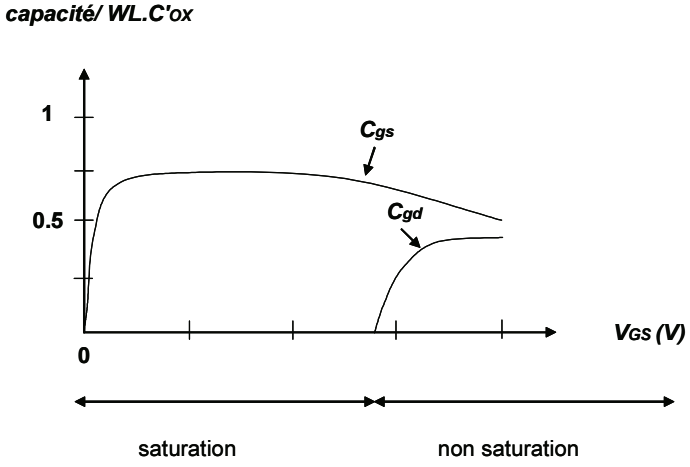


Figure 8.22 : Capacités équivalentes en fonction de la tension de grille

8.6.5. Le modèle du MOS complet aux fréquences moyennes

Une première amélioration du modèle est de revenir sur les approximations (8-64) pour obtenir un schéma amélioré. Ce schéma est donné figure 8-23 à partir du schéma 8-20.

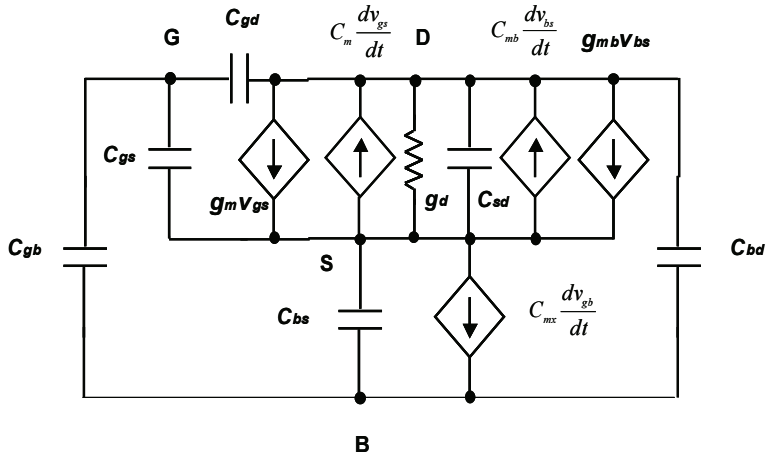


Figure 8-23 : Schéma fréquences moyennes

Les modifications introduites par ce schéma sont l'introduction de trois sources de courant supplémentaires. On posera

$$\begin{aligned} C_{dg} - C_{gd} &= C_m \\ C_{db} - C_{bd} &= C_{mb} \\ C_{bg} - C_{gb} &= C_{mx} \\ C_{sd} &\neq 0 \end{aligned} \tag{8-68}$$

Il faut ensuite partir des équations de base

$$i_{dv}(t) = C_{gd}\frac{dv_{dg}}{dt} + C_{bd}\frac{dv_{db}}{dt}$$

$$i_g(t) = C_{gd}\frac{dv_{gd}}{dt} + C_{gb}\frac{dv_{gb}}{dt} + C_{gs}\frac{dv_{gs}}{dt}$$

$$i_b(t) = C_{bd}\frac{dv_{bd}}{dt} + C_{gb}\frac{dv_{bg}}{dt} + C_{bs}\frac{dv_{bs}}{dt}$$

Elles sont ensuite modifiées de la manière suivante

$$i_{dv}(t) = C_{gd}\frac{dv_{dg}}{dt} + C_{sd}\frac{dv_{ds}}{dt} + C_{bd}\frac{dv_{db}}{dt} - C_m\frac{dv_{gs}}{dt} - C_{mb}\frac{dv_{bs}}{dt}$$

$$i_g(t) = C_{gd}\frac{dv_{gd}}{dt} + C_{gb}\frac{dv_{gb}}{dt} + C_{gs}\frac{dv_{gs}}{dt}$$

$$i_b(t) = C_{bd}\frac{dv_{bd}}{dt} + C_{gb}\frac{dv_{bg}}{dt} - C_{mx}\frac{dv_{gd}}{dt} + C_{bs}\frac{dv_{bs}}{dt}$$

Il est alors possible de reprendre tous les coefficients du schéma équivalent

$$C_{gs} = -\left(\frac{\partial Q_G}{\partial V_S}\right)_{V_G,V_D,V_B} = \frac{2}{3}C'_{OX}WL\frac{1+2\alpha}{(1+\alpha)^2}$$

$$C_{bs} = -\left(\frac{\partial Q_B}{\partial V_S}\right)_{V_G,V_D,V_B} = \delta\frac{2}{3}C'_{OX}WL\frac{1+2\alpha}{(1+\alpha)^2}$$

$$C_{gd} = -\left(\frac{\partial Q_G}{\partial V_D}\right)_{V_G,V_S,V_B} = \frac{2}{3}C'_{OX}WL\frac{\alpha^2+2\alpha}{(1+\alpha)^2}$$

$$C_{bd} = -\left(\frac{\partial Q_B}{\partial V_D}\right)_{V_G,V_S,V_B} = \delta\frac{2}{3}C'_{OX}WL\frac{\alpha^2+2\alpha}{(1+\alpha)^2}$$

$$C_{gb} = -\left(\frac{\partial Q_G}{\partial V_B}\right)_{V_G,V_S,V_B} = \frac{\delta}{3(1+\delta)}C'_{OX}WL\left(\frac{1-\alpha}{1+\alpha}\right)^2 \qquad (8\text{-}69)$$

$$C_{dg} = -\left(\frac{\partial Q_D}{\partial V_G}\right)_{V_S,V_D,V_B} = C_{OX}\frac{4+28\alpha+22\alpha^2+6\alpha^3}{15(1+\alpha)^3}$$

$$C_{db} = -\left(\frac{\partial Q_D}{\partial V_B}\right)_{V_G,V_D,V_B} = \delta\cdot C_{dg}$$

$$C_{bg} = -\left(\frac{\partial Q_B}{\partial V_G}\right)_{V_S,V_D,V_B} = \frac{\delta}{3(1+\delta)}C_{OX}\left(\frac{1-\alpha}{1+\alpha}\right)^2$$

$$C_{sd} = -\left(\frac{\partial Q_s}{\partial V_D}\right)_{V_S,V_G,V_B} = -\frac{4}{15}C_{OX}(1+\delta)\frac{\alpha+3\alpha^2+\alpha^3}{(1+\alpha)^3}$$

Ces coefficients peuvent se calculer dans les deux cas extrêmes du régime saturé et du régime linéaire. Les résultats sont reportés dans un tableau récapitulatif.

	$\alpha = 0$	$\alpha = 1$
C_{sg}	2/5 C_{OX}	1/2 C_{OX}
C_{gs}	2/3 C_{OX}	1/2 C_{OX}
C_{bs}	2/3 $C_{OX}\,\delta$	1/2 $C_{OX}\,\delta$
C_{gd}	0	1/2 C_{OX}
C_{dg}	4/15 C_{OX}	1/2 C_{OX}
C_{gb}	1/3 $C_{OX}\,\delta/\delta$+1	0
C_{bg}	1/3 $C_{OX}\,\delta/\delta$+1	0
C_{sd}	0	–1/6 $C_{OX}\,(1+\delta)$
C_{ds}	–4/15 $C_{OX}\,(1+\delta)$	–1/6 $C_{OX}\,(1+\delta)$
C_m	4/15 C_{OX}	0
C_{mb}	4/15 $C_{OX}\,\delta$	0
C_{mx}	0	0

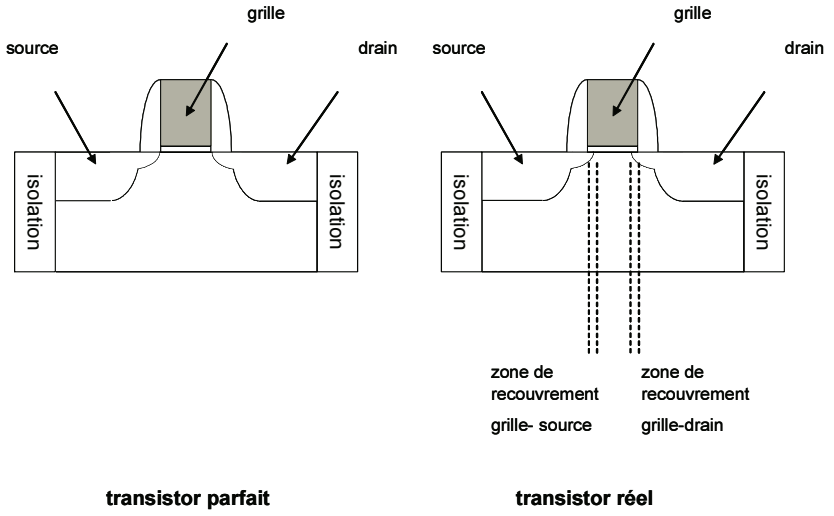


Figure 8-24 : Capacités parasites de recouvrement

Le schéma précédent n'est pas suffisant et doit être complété par les éléments suivants :

- prise en compte du recouvrement entre les extrémités et la grille,
- prise en compte des capacités des jonctions substrat-source et drain-source,
- prise en compte des résistances d'accès aux électrodes.

Le premier effet est du aux capacités parasites de couplage entre la grille et les zones de source et de drain comme il est illustré figure 8.24. La technique d'auto-alignement de la source et du drain à la fabrication n'est pas parfaite et il subsiste une zone de recouvrement.
Si la longueur de la zone de recouvrement est Lr, les deux capacités parasites ont pour valeurs $WL_rC'_{OX}$. Elles sont notées respectivement C_{gsr} et C_{gdr}. Les effets sont très différents pour le fonctionnement du transistor car la capacité C_{gsr} est une simple augmentation de la capacité C_{gs} que nous avons calculé précédemment et relativement faible. La capacité C_{gdr} a par contre un effet majeur sur le fonctionnement du MOS en saturation car la capacité électrique équivalente C_{gd} est nulle en régime de saturation si on ne tient pas compte du recouvrement .
Cela apparaît figure 8-24. Cette capacité de recouvrement injecte en entrée une partie de la tension de sortie ce qui introduit une contre-réaction dans le fonctionnement électrique du MOS. Cet effet conduit à une diminution importante de la bande passante du transistor.

8.6.6. *Un résumé du modèle du MOS*

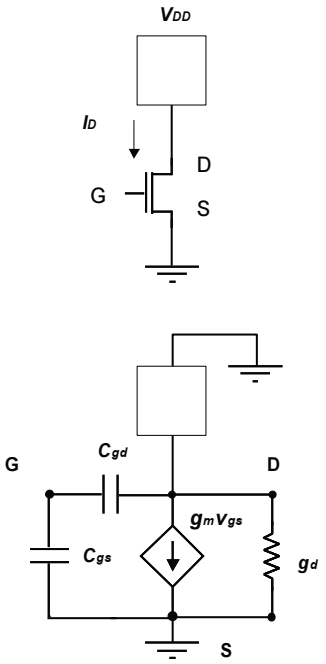


Figure 8.25 : Modèle du NMOS

CANAL LONG

	g_m	g_d	C_{gd}	C_{gs}
off	0	0	$WL_{gdr}C'_{OX}$	$WL_{gsr}C'_{OX}$
Faible inversion	$\frac{I_D}{n_0\phi_t}$	$\frac{\exp\frac{V_{gs}}{n_0\phi_t}\frac{I_{ds}}{\phi_t}}{1-\exp\frac{-V_{ds}}{\phi_t}}$	$\frac{1}{2}WLC'_{OX}$	$\frac{1}{2}WLC'_{OX}$
Triode α=0	$\frac{W}{L}\mu_n C'_{OX}V_{DS}$	$\beta_n[V_{GS}-V_T-(1+\delta)V_{DS}]$	$\frac{1}{2}WLC'_{OX}$	$\frac{1}{2}WLC'_{OX}$
Saturé α=1	$\sqrt{2\beta_n I_D}$	$\frac{I_{Dsat}}{V_A}$	$WL_{gdr}C'_{OX}$	$\frac{2}{3}WLC'_{OX}$

CANAL COURT

	g_m	g_d	C_{gd}	C_{gs}
off	0	0	$WL_rC'_{OX}$	$WL_{gsr}C'_{OX}$
Faible inversion	$\frac{I_D}{n_0\phi_t}$	$\frac{\exp\frac{V_{gs}}{n_0\phi_t}\frac{I_{ds}}{\phi_t}}{1-\exp\frac{-V_{ds}}{\phi_t}}$	$\frac{1}{2}WLC'_{OX}$	$\frac{1}{2}WLC'_{OX}$
Triode α=0	$\frac{WC'_{OX}\mu_n\frac{V_{DS}}{L}}{1+(V_{DS}/LE_{Cn})}$	$\frac{\beta_n[V_{GS}-V_T-(1+\delta)V_{DS}]}{1+(V_{DS}/LE_{Cn})}$	$\frac{1}{2}WLC'_{OX}$	$\frac{1}{2}WLC'_{OX}$
Saturé α=1	$W\mu_n C'_{OX}E_{Cn}$	$\frac{I_{Dsat}}{V_A}$	$WL_{gdr}C'_{OX}$	$\frac{2}{3}WLC'_{OX}$

Nous allons maintenant donner un résumé des résultats précédents en simplifiant au maximum la représentation du transistor afin de pouvoir prévoir au premier ordre le comportement des circuits intégrés. Il est ensuite possible et souvent nécessaire d'utiliser les modèles élaborés pour obtenir des résultats précis. Le premier tableau est relatif au transistor NMOS dans la configuration classique. La grille et le substrat sont reliés et mis à la masse du circuit. Les effets « body » sont alors annulés.

Rappelons les termes de ces formules. Le courant de drain s'exprime de quatre manières en fonction du régime :

- Faible inversion

$$I_D = I_S e^{\frac{V_{GS}-V_T}{n_0\phi_t}}\left(1-e^{-\frac{V_{DS}}{\phi_t}}\right)$$

- Canal long, triode et forte inversion

$$I_D = \frac{W}{L}\mu_n C'_{OX}\left[\left(V_{GS}-V_T\right)V_{DS}-\frac{1}{2}\left(1+\delta\right)V_{DS}{}^2\right]$$

- Canal long, saturation et forte inversion

$$I_D = \frac{W}{L}\mu_n C'_{OX}\frac{(V_{GS}-V_T)^2}{2(1+\delta)}\left(1+\frac{V_{DS}-V_{DS_{sat}}}{V_A}\right)$$

- Canal court, triode, forte inversion

$$I_D = \frac{\frac{W}{L}\mu_n C'_{OX}}{1+\frac{V_{DS}}{LE_C}}\left[(V_{GS}-V_T)V_{DS}-\frac{1}{2}(1+\delta){V_{DS}}^2\right]$$

- Canal court, saturation, forte inversion

$$I_D = WC'_{OX}v_{nsat}\left[V_{GS}-V_T-\frac{1+\delta}{2}V_{DSsat}\right]\left(1+\frac{V_{DS}-V_{DS_{sat}}}{V_A}\right)$$

Dans ces relations les paramètres sont définis comme suit

- W : largeur du transistor
- L : longueur du canal
- C'_{OX}: capacité de l'oxyde par unité de surface (égale à ε_{ox}/t_{ox} rapport de la constante diélectrique de l'oxyde sur l'épaisseur de l'oxyde)
- μ_n : mobilité effective des électrons à l'interface, différente de la mobilité dans le matériau.
- V_T : tension de seuil du NMOS, dépend de paramètres technologiques et de l'effet « body » caractérisé par le paramètre γ défini par $\gamma = \sqrt{2eN_A\varepsilon_s}/C'_{ox}$ mais aussi de la tension de drain de manière plus indirecte.
- δ : paramètre de la modélisation égal à $\delta = \gamma/2\sqrt{2\phi_F + V_{SB}}$
- N_A : dopage du silicium *p*
- ε_s : permittivité du silicium
- ϕ_F **:** potentiel de Fermi du silicium *p*
- ϕ_t **:** potentiel égal à kT/e
- n_0 : égal à $1+\gamma/2\sqrt{1.5\,\phi_F + V_{SB}}$
- V_{DSsat} **:** tension de saturation

Les tensions de saturation sont respectivement

$$V_{DSsat} = \frac{V_{GS}-V_T}{1+\delta} \qquad \text{pour un MOS canal long}$$

$$V_{DS_{sat}} = LE_c\left[\sqrt{1+\frac{2(V_{GS}-V_T)}{(1+\delta)LE_C}}-1\right] \qquad \text{pour un MOS canal court}$$

- E_c : champ critique en relation avec v_{sat} par la relation : $v_{sat}=\mu_n E_c$
- V_A : paramètre pour la modulation de canal égal à $kL\sqrt{N_A}$ formule dans laquelle le coefficient k vaut environ 0.15 V $\mu m^{1/2}$
- β_n est défini par : $\beta_n=\mu_n C'_{OX}.W/L$
- L_{gdr} et L_{gsr} sont les longueurs,de recouvrement entre grille et drain et grille et source dues aux imperfections de la technologie.
- Le courant I_S exprimé dans la formule de la faible inversion est une formule complexe donnée dans le paragraphe 8.2.

Il faut ajouter à ce tableau un certain nombre de remarques. Le paramètre $\mu_n C'_{OX}$ n'a pas de signification pour un MOS canal court. De la même manière, la vitesse de saturation est utilisée uniquement pour décrire un MOS canal court. Il faut noter une diminution de la résistance de sortie pour le MOS canal court.

Un modèle de type équivalent peut être donné pour le PMOS. On suppose que le transistor est de manière classique connecté entre l'alimentation positive V_{DD} et le reste du circuit comme l'indique la figure 8.26. La source est à la tension la plus positive et le caisson du transistor est relié lui aussi à ce potentiel. Tous les montages électriques ne conduisent pas à cette configuration mais elle est très souvent utilisée aussi bien en logique qu'en analogique.

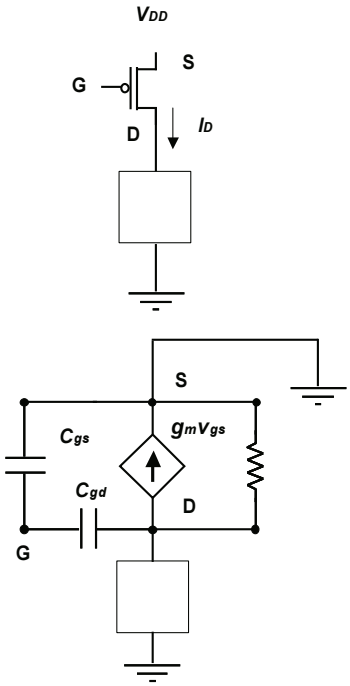


Figure 8-26 : Modèle du PMOS

CANAL LONG

	g_m	g_d	C_{gd}	C_{gs}
off	0	0	$WL_{gdr}C'_{OX}$	$WL_{gsr}C'_{OX}$
Faible inversion	$\frac{I_D}{n_0\phi_t}$	$g_d=\frac{\exp\frac{-V_{ds}}{\phi_t}}{1-\exp\frac{-V_{ds}}{\phi_t}}\frac{I_D}{\phi_t}$	$\frac{1}{2}WLC'_{OX}$	$\frac{1}{2}WLC'_{OX}$
Triode α=0	$\frac{W}{L}\mu_p C'_{OX}V_{SD}$	$\beta_p\left(V_{sg}-V_T-(1+\delta)V_{sd}\right)$	$\frac{1}{2}WLC'_{OX}$	$\frac{1}{2}WLC'_{OX}$
Saturé α=1	$\sqrt{2\beta_p I_D}$	$\frac{I_{Dsat}}{V_A}$	$WL_{gdr}C'_{OX}$	$\frac{2}{3}WLC'_{OX}$

CANAL COURT

	g_m	g_d	C_{gd}	C_{gs}
off	0	0	$WL_r C'_{OX}$	$WL_{gsr}C'_{OX}$
Faible inversion	$\frac{I_D}{n_0\phi_t}$	$\frac{\exp\frac{-V_{ds}}{\phi_t}}{1-\exp\frac{-V_{ds}}{\phi_t}}\frac{I_D}{\phi_t}$	$\frac{1}{2}WLC'_{OX}$	$\frac{1}{2}WLC'_{OX}$
Triode α=0	$\frac{WC'_{OX}\mu_p\frac{V_{DS}}{L}}{1+\left(V_{SD}/LE_{Cp}\right)}$	$\frac{\beta_p\left(V_{SG}-V_T-(1+\delta)V_{SD}\right)}{1+\left(V_{SD}/LE_{Cp}\right)}$	$\frac{1}{2}WLC'_{OX}$	$\frac{1}{2}WLC'_{OX}$
Saturé α=1	$W\mu_p C'_{OX}E_{Cp}$	$\frac{I_{Dsat}}{V_A}$	$WL_{gdr}C'_{OX}$	$\frac{2}{3}WLC'_{OX}$

Le lecteur doit prendre garde à l'apparente inversion du sens de la source de courant. Ce n'est que l'effet de la représentation du schéma inversé par rapport à

celui du NMOS puisque la source est maintenant en haut de la figure et non pas en bas.

On rappelle les équations donnant le courant de drain en fonction des tensions. Ce sont les mêmes que celle du NMOS à la condition de remplacer V_{GS} par V_{SG} et V_{DS} par V_{SD}. La tension de seuil du PMOS est, rappelons-le, négative quand elle est définie pour la tension entre la grille et la source. Elle est donc positive quand elle est relative à la tension entre source et grille. Ces conventions conduisent donc à des **valeurs toutes positives** pour le PMOS. Ce point est souvent source de confusion. On obtient alors

- Faible inversion

$$I_D = I_S e^{\frac{V_{SG}-V_T}{n_0\phi_t}}\left(1-e^{-\frac{V_{SD}}{\phi_t}}\right)$$

- Canal long, triode et forte inversion

$$I_D = \frac{W}{L}\mu_p C'_{OX}\left[\left(V_{SG}-V_T\right)V_{SD}-\frac{1}{2}\left(1+\delta\right){V_{SD}}^2\right]$$

- Canal long, saturation et forte inversion

$$I_D = \frac{W}{L}\mu_p C'_{OX}\frac{\left(V_{SG}-V_T\right)^2}{2\left(1+\delta\right)}\left(1+\frac{V_{SD}-V_{SD_{sat}}}{V_A}\right)$$

- Canal court, triode, forte inversion

$$I_D = \frac{\frac{W}{L}\mu_p C'_{OX}}{1+\frac{V_{SD}}{LE_{Cp}}}\left[\left(V_{SG}-V_T\right)V_{SD}-\frac{1}{2}\left(1+\delta\right){V_{SD}}^2\right]$$

- Canal court, saturation, forte inversion

$$I_D = WC'_{OX}v_{psat}\left[V_{SG}-V_T-\frac{1+\delta}{2}V_{SDsat}\right]\left(1+\frac{V_{SD}-V_{SD_{sat}}}{V_A}\right)$$

La différence fondamentale entre NMOS et PMOS est la mobilité des porteurs. La mobilité des trous est trois fois plus faible que celle des électrons en régime non saturé au sens de la saturation de la vitesse. On ne retrouve pas cette différence pour la vitesse de saturation qui est environ la même pour les électrons et les trous. On observe cependant une différence équivalente pour les valeurs des

champs critiques. Le champ critique du PMOS est environ trois fois plus élevé que celui du NMOS. En résumé, que ce soit en régime de canal court ou en régime de canal long, le PMOS doit être plus large que le NMOS pour délivrer le même courant. Cette règle fondamentale pour le design conduit à réaliser en général des PMOS trois fois plus larges que les NMOS dans les technologies de canal long et deux fois plus larges dans les technologies de canal court.

8.7. Le bruit du transistor MOS

Les sources de bruit du transistor MOS sont le bruit thermique du canal de conduction, le bruit de grenaille associé au courant sous le seuil et éventuellement aux courants tunnel et enfin le bruit basse fréquence dit en un sur *f*.

Le bruit thermique du canal est le plus classique et avait un rôle dominant pour les anciennes technologies MOS. Les technologies modernes conduisent à une augmentation significative du bruit de grenaille et du bruit basse fréquence.

8.7.1. *Le bruit thermique du canal en forte inversion*

On considère le canal comme une série de résistances de valeurs différentes, chaque tronçon générant une tension de bruit dont la valeur est donnée par la relation 1-273. Le schéma électrique est représenté figure 8-27.

La valeur *r(x)* est la résistance par unité de longueur. On appelle *dV* la variation du potentiel et ΔV la tension de bruit associée à l'élément *r(x)dx*. Le courant qui traverse le canal est constant le long du canal et noté I_D. $\Delta I_D(x)$ est la composante due au bruit.

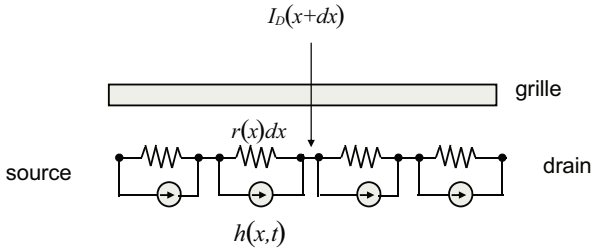


Figure 8-27 : Le bruit du canal

$$dV = r(x)dxI_D$$
$$V(x) = V_0(x) + \Delta V(x)$$
$$I_D = I_{D0} + \Delta I_D(x)$$

La valeur moyenne du courant est notée I_{D0}.

$$I_{D0} = \frac{1}{r(x)}\frac{dV_0}{dx} = g(x)\frac{dV_0}{dx}$$

Le bruit est représenté par une source de courant *h(x,t).* On peut donc écrire aux bornes d'un élément résistif.

$$I_D(x) + \Delta I_D(x) = \left[g\left(V_0(x)\right) + \frac{dg}{dV}\Delta V(x) \right]\left[\frac{dV_0}{dx} + \frac{d}{dx}\Delta V(x) \right] + h(x,t)$$

$$I_D(x) + \Delta I_D(x) = g\left(V_0\right)\frac{dV_0}{dx} + \frac{dg}{dV}\Delta V\frac{dV_0}{dx} + g\left(V_0\right)\frac{d\Delta V}{dx} + h(x,t)$$

On en déduit

$$\Delta I_D(x) = \frac{d}{dx}\left[g(V_0)\Delta V \right] + h(x,t)$$

Le courant étant constant le long du canal, on obtient

$$\Delta I_D \cdot L = \left[g(V_0)\Delta V \right]_0^L + \int_0^L h(x,t)dx$$

La tension étant imposée à chaque extrémité du canal par la polarisation du dispositif, la relation se réduit à

$$\Delta I_D = \frac{1}{L}\int_0^L h(x,t)dx$$

Cette relation exprime que le bruit associé au canal est la moyenne des contributions de chaque élément. On peut maintenant calculer la densité spectrale en appliquant le théorème de Wiener Khintchine.

$$E\left[\Delta I_D(t)\Delta I_D(t+s) \right] = \frac{1}{L^2}\int_0^L\int_0^L E\left[h(x,t)h(x',t+s) \right] dxdx'$$

$$S_{\Delta I_D} = \frac{1}{L^2}\int_{-\infty}^{+\infty} ds \int_0^L\int_0^L E\left[h(x,t)h(x',t+s) \right] e^{-2\pi jfs} dxdx'$$

Les sources de bruit étant supposées non corrélées.

$$E\left[h(x,t)h(x',t+s) \right] = E\left[h(x,t)h(x',t+s) \right]\delta(x-x')$$

$$S_{\Delta I_D} = \frac{1}{L^2} \int_{-\infty}^{+\infty} ds \int_0^L E\left[h(x,t)\,h(x,t+s)\right] e^{-2\pi jfs} dx$$

L'expression de la densité spectrale du bruit thermique permet d'écrire

$$\int_{-\infty}^{+\infty} E\left[h(x,t)\,h(x,t+s)\right] e^{-2\pi js} ds = 4kTg(V_0)$$

Comme,

$$I_{D0} = g(V_0)\frac{dV_0}{dx}$$

On obtient en définitive.

$$S_{\Delta I_D} = \frac{1}{L^2} \int_0^L 4kTg^2(V_0)\frac{1}{I_{D0}} dV_0$$

Il suffit maintenant de remplacer la conductance par unité de longueur par sa valeur.

$$I_{D0} = \frac{1}{r(x)}\frac{dV_0}{dx} = g(x)\frac{dV_0}{dx}$$

$$S_{\Delta I_D} = \frac{4kT}{L^2}\frac{1}{I_{D0}} \int_{V_0(0)}^{V_0(L)} g^2(V_0) dV_0 \qquad (8\text{-}70)$$

On peut écrire

$$g(x) = Q_I' \cdot W\mu$$

En régime de forte inversion

$$Q_I'(x) = -C_{OX}'\left[V_G - V_0(x) - V_T\right]$$

$$V_0(x) = V_{CB}(x) + \phi_B$$

$$V_T = \Phi_{MS} - \frac{Q_0'}{C_{OX}'} + \gamma\sqrt{V_0(x)}$$

Dans un souci de simplification, étudions le cas $\gamma = 0$

$$V_0(0) = V_{SB} + \phi_B$$
$$V_0(L) = V_{DB} + \phi_B$$

Un calcul assez long conduit à

$$\int_{V_0(0)}^{V_0(L)} g^2(V_0) dV_0 = \mu^2 W^2 C'^2_{OX} \left[V_D (V_G - V_T)^2 + \frac{1}{3} v^3 (V_G - V_T)^3 - v^2 (V_G - V_T)^3 \right] \qquad (8\text{-}71)$$

On a posé

$$v = \frac{V_D}{V_G - V_T}$$
$$V_D = V_{DS}$$
$$V_G = V_{GS}$$

Reprenons la forme classique du courant de drain en forte inversion.

$$I_{D0} = \frac{\mu W}{L} C'_{OX} \left[(V_G - V_T) V_D - \frac{V_D^2}{2} \right] = \frac{\mu W}{L} C'_{OX} (V_G - V_T)^2 \left(v - \frac{v^2}{2} \right)$$

Le cas *v*=1 correspond à la saturation. Définissons alors les deux conductances

$$g_{d0} = \left(\frac{\partial I_{D0}}{\partial V_D} \right)_{V_D = 0} = \mu \frac{W}{L} C'_{OX} (V_G - V_T) \qquad (8\text{-}72)$$

$$g_0 = g_{x=0} = W \mu C'_{OX} (V_G - V_T) \qquad (8\text{-}73)$$

On obtient alors

$$S_{\Delta I_D} = 4kT g_{d0} \eta$$
$$\eta = \frac{1}{g_0 I_D L} \int_{\phi_B}^{\phi_B + V_D} g^2(V_0) dV_0 \qquad (8\text{-}74)$$

Le résultat de la formule 8-71 permet d'écrire

$$\eta = \frac{v + \frac{1}{3} v^3 - v^2}{v - \frac{v^2}{2}} = \frac{1 - v + \frac{1}{3} v^2}{1 - \frac{v}{2}}$$

On obtient alors les deux valeurs extrêmes de η : 1 en régime linéaire et 2/3 en régime saturé.

Quand on abandonne l'hypothèse $\gamma=0$, les calculs sont plus complexes. Le résultat est le suivant.

$$S_{\Delta I_D} = 4kTg_{d0}\eta_{sat}\frac{g_{d0}}{g_{\max}}g_{\max}$$

$$S_{\Delta I_D} = 4kT\alpha_{sat}g_{\max}$$

Le terme g_{max} est défini par :

$$g_{\max} = \left(\frac{\partial I_D}{\partial V_G}\right)_{V_D=V_{D_{sat}}}$$

Le coefficient $\frac{g_{d0}}{g_{\max}}$ est représenté figure 8-28 en fonction de γ^2.

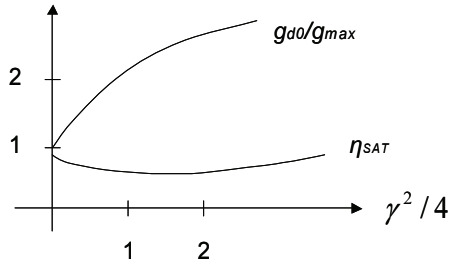


Figure 8-28 : Effet de γ sur le bruit thermique

Il y donc intérêt pour diminuer γ à diminuer le dopage et à augmenter la capacité de l'oxyde par unité de surface.

8.7.2. *Le bruit thermique du canal en faible inversion*

On reprend la formule générale 8-70 mais en modifiant la formule du courant de drain.

$$g(V_0) = \frac{\mu W C'_D(x)}{\phi_t} e^{-\frac{V_0}{\phi_t}} e^{\left(\frac{V_G - V_X}{n\phi_t} - \frac{1}{2}\frac{\phi_F}{\phi_t}\right)}$$

On suppose que la capacité $C'_D(x)$ de déplétion varie peu avec la position dans le canal.

$$V_X = \phi_{MS} - \frac{Q'_{OX}}{C'_{OX}} + \frac{3}{2}\phi_F + \left(\frac{Q'_B}{C'_{OX}}\right)_{V_0=1.5\phi_F}$$

$$I_{D0} = \frac{1}{L}\int_{V_0(0)}^{V_0(L)} g(V_0)dV_0 = I_{D\,sat}\left(1 - \exp^{-\frac{V_D}{\phi_t}}\right)$$

Avec

$$I_{D\,sat} = \frac{\mu W C_D(1{,}5\phi_F)}{\phi_t L} e^{\left(-\frac{V_G - V_X}{n(1{,}5\phi_F)} - \frac{1}{2}\frac{\phi_F}{\phi_t}\right)}$$

Un calcul non détaillé dans cet ouvrage conduit à la relation

$$S_{\Delta I_D} = 2qI_{D\,sat}\left[1 + \exp^{-\frac{V_D}{\phi_t}}\right] \tag{8-75}$$

Cette relation montre que la densité spectrale de bruit en régime de faible inversion peut être considérée comme la densité de bruit d'un bruit de grenaille correspondant au courant de saturation. Elle augmente d'un facteur 2 quand la tension de drain est élevée. Cette relation confirme le fait que bruit thermique et bruit de grenaille ont la même origine physique : la granularité de la composition du courant.

8.7.3. *Les autres sources de bruit*

Les autres sources de bruit sont le bruit induit par la grille qui traduit en fait le couplage entre la grille et le canal de conduction, le bruit de grenaille associé aux courants de fuite du transistor, courants inverses des diodes source-substrat et drain-substrat et enfin le bruit de grenaille associé aux courants tunnel grille-substrat et source-drain.

En haute fréquence l'impédance entre la grille et le canal de conduction n'est pas purement capacitive et une certaine impédance réelle peut être calculée donc source de bruit thermique. L'impédance vue de la grille peut s'écrire en faisant un calcul approché non détaillé dans cet ouvrage.

$$Y_G = j\omega C_G + g_G$$

avec

$$C_G = \frac{2}{3} C'_{OX} WL$$
$$g_g = \frac{1}{5} \frac{\omega^2 C_g^{\ 2}}{g_{d0}} \qquad (8\text{-}76)$$
$$g_{d0} = \left(\frac{\delta I_D}{\delta V_D} \right)_{V_D = 0}$$

Ces valeurs conduisent à une densité spectrale de bruit.

$$S_{\Delta I_G} = \beta 4kTg_G$$
$$avec \ \ \beta = 4/3 \ \ pour \ un \ MOSFET \ saturé$$

Ce bruit est corrélé au bruit thermique du drain.

$$S_{\Delta I_G \Delta I_D} = 1/6 \cdot j\omega C_G \cdot 4kT$$
$$pour \ un \ MOSFET \ saturé$$

Les bruits associés aux courants inverses ont des densités spectrales proportionnelles aux valeurs de ces courants.

Le bruit en un sur *f* est lié aux phénomènes de génération et recombinaisons de charges dûs aux défauts et impuretés principalement à l'interface oxyde semiconducteur. Il est difficile de calculer précisément la valeur de la densité spectrale de ce bruit car il dépend beaucoup des procédés technologiques mis en œuvre. La figure 8-29 montre qu'elle augmente quand la longueur de canal diminue. De même, elle diminue à courant constant quand la largeur du transistor augmente. ***En conclusion, il faut réaliser des transistors de grande taille pour minimiser le bruit ce qui est contraire à l'évolution naturelle de la technologie microélectronique.***

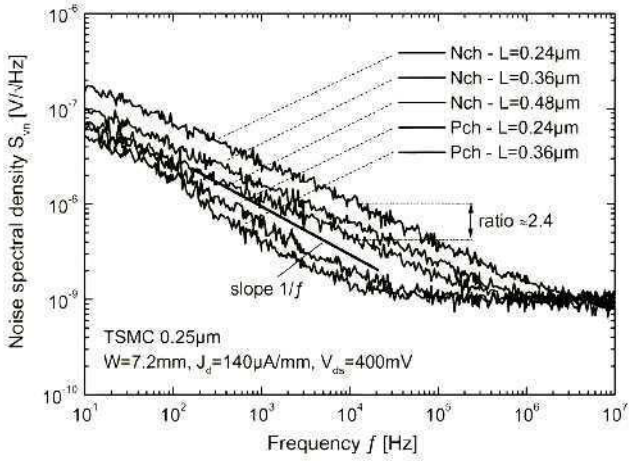


Figure 8-29 : Bruit en *1/f* en fonction de la longueur de canal

8.8 Applications du MOSFET

8.8.1 Amplification

Le transistor MOS étant un dispositif permettant de commander le courant de drain par la tension de grille, le premier schéma venant à l'esprit pour réaliser un amplificateur de tension est celui de la figure 8.30.

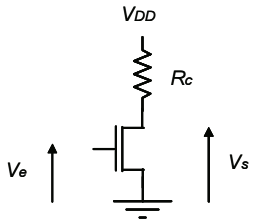


Figure 8-30 : Amplificateur à MOSFET

Ce schéma très simple conduit à écrire les équations suivantes

$$V_s = V_{DD} - R_C I_{DS}$$
$$I_{DS} = k\frac{W}{2L}(V_e - V_T)^2(1 + \lambda V_{DS})$$

Dans cette dernière relation valable en régime de saturation, on a négligé la tension de saturation. Les paramètres λ et k sont définis par les formules

$$k = \mu_n C'_{OX}$$
$$\lambda = \frac{1}{V_A}$$

On exprime alors facilement le gain en tension

$$A_V = \frac{dV_s}{dV_e} \approx -k\frac{W}{L}(V_e - V_T)R_c$$

On a négligé le terme en λ. Le transistor est supposé saturé. Dans une technologie 0.8 micron, la valeurs de k est environ 110 µA par V^2. Si le transistor a un facteur de forme de 10 (valeur de W/L), on obtient pour une valeur typique de $(V_e - V_T)$ égale à 1 V.

$$A_V \approx 10^{-3} R_C$$

Un gain significatif peut être obtenu pour des résistances de charge supérieures à 10^4 Ω. Comme il est difficile de réaliser des valeurs de résistance élevée, il faut trouver une autre solution.

L'idée est d'utiliser un transistor saturé comme charge. En effet, un transistor saturé est voisin d'une source de courant donc d'une résistance de valeur élevée. On en arrive naturellement au schéma 8.31

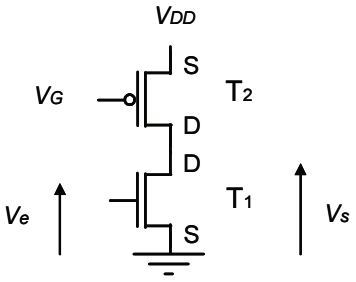


Figure 8-31 : Amplificateur à charge active

Le transistor T2 sert de charge au transistor T1. C'est un PMOS afin de pouvoir fonctionner avec des tensions uniquement positives. La tension *VG* fixe, appliquée sur la grille, est négative par rapport à la source reliée au *VDD.* Le caisson est relié à la tension d'alimentation *VDD* et le transistor est supposé en régime de saturation. Il est équivalent à une source de courant soit une impédance de forte valeur.

En fonction des résultats établis, on peut tracer le schéma équivalent en petits signaux correspondant au schéma 8-32.. La tension de grille d'entrée du transistor T2 étant constante, la source de tension en petits signaux est nulle ainsi que la source de courant commandée correspondante. Le nœud correspondant à la tension d'alimentation *VDD* est considéré à la masse dans la représentation petits signaux.

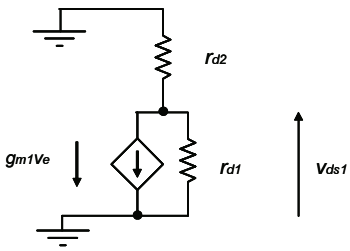


Figure 8-32 : Schéma équivalent en petits signaux

Pour déterminer le gain de cet étage, il est nécessaire d'écrire la relation entre le courant traversant le transistor T2 et la tension drain-source de ce transistor. On utilisera le schéma équivalent petits signaux du transistor en régime saturé.
Ce schéma est valable uniquement dans la région où les deux transistors sont saturés.
Un calcul élémentaire conduit à :

$$v_{ds1} = -g_{m1} v_e \frac{1}{\frac{1}{r_{d1}} + \frac{1}{r_{d2}}}$$

En exprimant les résistances en fonction du courant défini dans la formule 8.57, on obtient

$$g_d = \frac{I_{Dsat}}{V_A}$$

$$v_{ds1} = -g_{m1} v_e \frac{1}{g_{d1} + g_{d2}}$$

Le gain en tension est donc élevé et varie avec la transconductance du transistor T1. Elle est proportionnelle à la racine carrée du produit W_1/L_1 par le courant de drain en régime de canal long. En effet, la relation 8-56 nous apprend qu'en régime de canal long

$$g_{m1} = \sqrt{2\beta_n \frac{I_{D1}}{1+\delta}} = \sqrt{2\mu_n C'_{OX} \frac{W_1}{L_1} \frac{I_{D1}}{1+\delta}}$$

Les termes g_{d1} et g_{d2} varient proportionnellement aux courants de saturation des transistors. Ces courants sont peu différents du courant de drain traversant les deux transistors en série. Le gain en tension varie donc au total de la manière suivante

$$A_V \approx \sqrt{\frac{W_1}{L_1} \frac{1}{I_D}}$$

Ce résultat montre bien l'impact du choix des dimensions sur les performances.

8.8.2 Porte logique : cas de l'inverseur

a. Le MOS en commutation

Pour étudier les circuits logiques, il serait possible de calculer les circuits correspondant aux différentes portes ainsi que les assemblages de portes, comme cela est fait pour les fonctions analogiques. Cette méthode a rapidement des limites

si on pense aux circuits complexes de la logique formés de milliers de portes. Il est donc nécessaire de définir les propriétés les plus importantes du transistor MOS utilisé en commutateur. Les simulateurs logiques feront de même mais avec le niveau d'automatisation nécessaire dans la conception de circuits complexes.

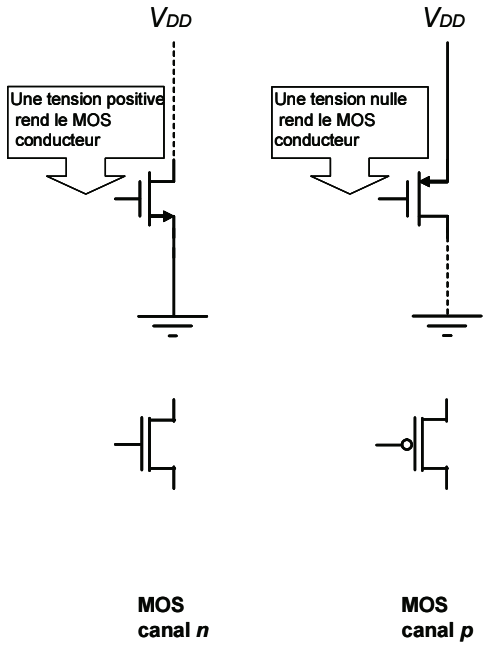


Figure 8-33 : Les schémas des MOS en logique

Le principe général de la logique MOS est de commuter la tension d'alimentation V_{DD} ou la tension nulle de référence en jouant sur la valeur de la tension de grille d'un MOSFET. En logique, on adopte une représentation simple pour distinguer les MOS canal *n* et canal *p*. La figure 8.33 rappelle les deux schémas.

Rappelons que le MOS canal *p* fonctionne généralement en polarisant la source et le puits à la tension positive d'alimentation V_{DD}. Les potentiels de grille et de drain sont donc négatifs par rapport au potentiel de la source.

Le schéma 8.34 illustre le principe de fonctionnement d'une cellule logique de base. On ignore pour l'instant la charge du MOS et on considère seulement une capacité de charge due par exemple à la piste de connexion et à une autre cellule logique reliée à celle que nous étudions.

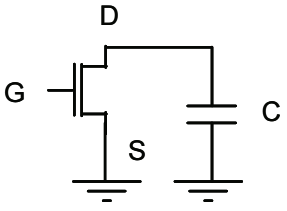


Figure 8-34 : Cellule logique de base

L'examen des caractéristiques du transistor permet d'observer ce qui se passe quand la tension de grille passe de 0 V à V_{DD}.

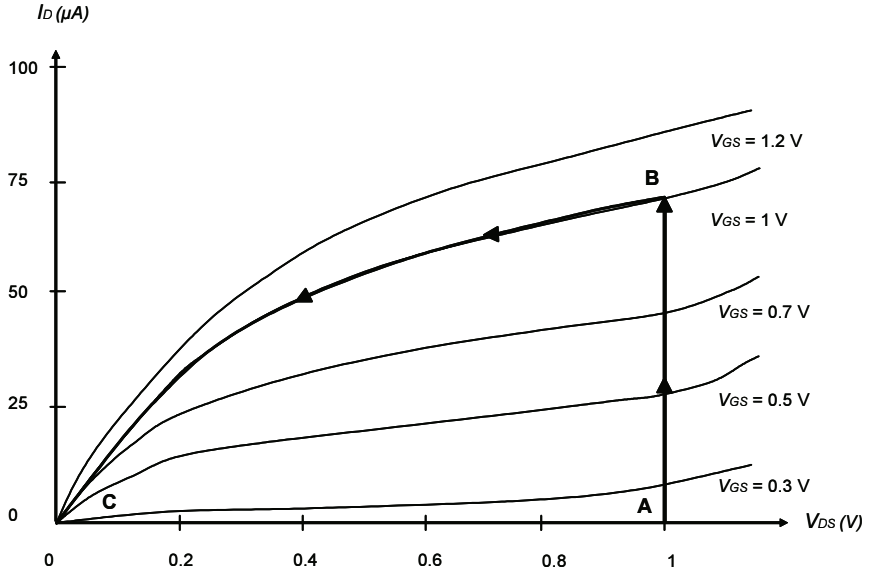


Figure 8-35 : Commutation d'un transistor

La figure 8.35 illustre la commutation d'un MOS dans une technologie avancée de 50 nm. Initialement, le transistor est non passant car une tension nulle est appliquée sur sa grille. Il est alors relié à la tension d'alimentation, 1V dans le cas présent. Ensuite, sa grille est portée au potentiel 1V et le MOS devient conducteur. Le point de fonctionnement passe de B à C. Pendant la transition, il y a un courant d'environ 70 µA puis le régime s'établit et le courant revient à zéro en suivant la courbe V_{GS} = 1 V. Si on mesure la pente de la droite BC, on obtient une valeur moyenne de la résistance du MOS pendant la commutation. Dans le cas présent, elle est de 14 kΩ. Il faut noter que les courbes sont relatives à un NMOS de petite taille. Le ratio *W/L* est de 2. Cette méthode est grossière mais elle donne pourtant les ordres de grandeur des caractéristiques électriques. Rappelons les formules donnant le courant en régime de saturation. Dans le cas d'un canal court c'est-à-dire pour les technologies numériques actuelles

$$I_D = W\mu C'_{OX}\left[V_{GS} - V_T - V_{DSsat}\right]E_C$$

Pour les technologies plus anciennes, le modèle canal long s'applique et

$$I_D = \frac{W}{L}\mu C'_{OX}\frac{\left(V_{GS} - V_T\right)^2}{2\left(1+\delta\right)}$$

Rappelons que V_T est la tension de seuil et que les paramètres physiques sont approximativement :

- C'_{ox} = 1.75 fF /µm^2 pour un NMOS dans une technologie 0.8 µ
- C'_{ox} = 25 fF /µm^2 pour un NMOS dans une technologie 45 nm
- μ est de 600 cm^2/ V. s pour les électrons et 200 cm^2 /V. s pour les trous
- E_C est de l'ordre de 1V / µm

Si on exprime ces valeurs dans le calcul précédent on obtient la valeur de la résistance moyenne du MOS pendant la commutation en fonction de sa largeur, pour des valeurs typiques de tension d'alimentation de 3 V et 1 V.

	Technologie 0.8 micron	**Technologie 45 nm**
NMOS	15 kΩ L /W	30 kΩ /W
PMOS	45 kΩ L/W	70 kΩ /W

Les dimensions L et W sont exprimées en unités relatives pour simplifier l'utilisation des formules, c'est-à-dire en prenant comme unité la grandeur caractéristique de la technologie. Dans une technologie 45 nm, les valeurs W et L d'un transistor de largeur 450 nm et dont le canal mesure 45 nm de long seront simplement W=10 et L=1.

Le schéma du MOS en commutation prend donc la forme très simple indiquée figure 8-36. Ce modèle grossier donne cependant des résultats assez proches de ceux que l'on peut obtenir à l'aide d'un simulateur de type SPICE.

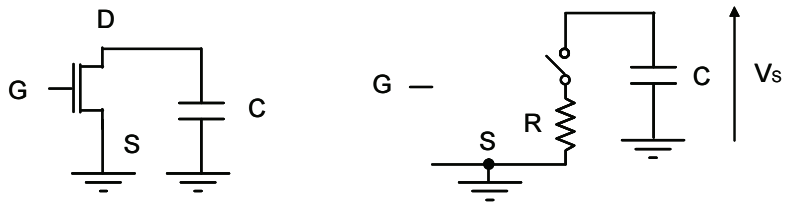


Figure 8-36 : Modèle grossier du MOS en commutation

Il est naturel de s'intéresser à l'aspect temporel de la commutation car la vitesse de fonctionnement des circuits logiques est une propriété fondamentale. D'autres caractéristiques devront être également étudiées : la consommation électrique et la tolérance au bruit et aux dispersions.

Le comportement temporel d'un étage logique peut être caractérisé par deux paramètres temporels : le retard et le temps de montée. Ces deux valeurs sont indiquées figure 8-37 en comparant la tension appliquée à l'entrée et la tension en sortie.

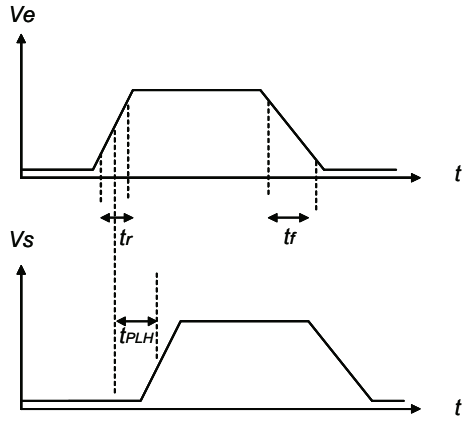


Figure 8-37 : Propriétés temporelles

Les impulsions logiques ne sont pas parfaites. Elles sont caractérisées par un temps de montée t_r et un temps de descente t_f . Les valeurs sont mesurées en prenant le temps correspondant à 10% et 90% du signal. Si l'impulsion d'entrée est définie par son temps de montée et par son temps de descente, l'impulsion de sortie est caractérisée de la même manière par les temps de montée (t_{LH}) et de descente (t_{HL}). Ils peuvent éventuellement être plus courts que ceux d'entrée. Pensons à un étage logique ayant un gain en tension élevé. Si la tension d'entrée sature l'étage, la pente de la tension de sortie peut être plus forte.

Le retard apparaît alors comme le temps correspondant à une variation de 50% de l'amplitude soit 0.7 *RC* et le temps de montée comme le temps correspondant à 90% de l'amplitude soit 2.2 *RC*. On suppose que la porte située en aval commute quand son entrée est à un potentiel supérieur à $V_{DD}/2$. Rappelons que seule la capacité de charge du MOS a été prise en compte dans ce modèle simpliste. Il faut pour affiner le modèle tenir compte des capacités du MOSFET lui-même.

Les résultats établis en 8.6.4 montrent qu'en régime triode les capacités entre grille et drain et entre grille et source s'expriment de la manière suivante.

$$C_t = \frac{1}{2} C'_{OX} WL$$

La valeur C'_{OX} est la capacité par unité de surface.

Le schéma de la figure 8.36 se modifie alors comme il est indiqué figure 8.38. Ce résultat n'est pas tout à fait immédiat et demande quelques précisions. Les capacités sont valables en régime triode uniquement. La capacité drain-source est donc surestimée notablement.

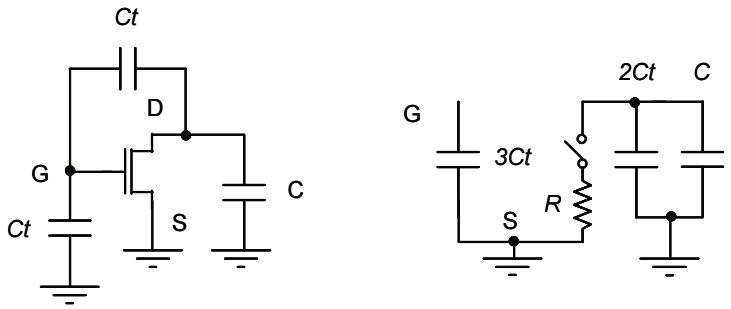


Figure 8-38 : Le MOS en commutation

Pour passer du schéma de gauche à celui de droite, calculons par exemple le courant d'entrée dans la commutation. La grille passe de 0 à V_{DD} pendant Δt et le drain de V_{DD} à 0. Le courant qui circule dans la capacité drain-grille est donc $2C_tV_{DD}/\Delta t$ alors que le courant qui circule dans la capacité grille-source est $C_tV_{DD}/\Delta t$. Au total, un courant $3C_tV_{DD}/\Delta t$ circule dans l'entrée pendant la commutation. De même, on peut calculer le courant qui circule en sortie. C'est $2C_tV_{DD}/\Delta t$. Le schéma équivalent de droite est donc justifié.

L'interrupteur est fermé quand le MOS passe de l'état bloqué à l'état conducteur et cela pour une tension qui est appelée tension de seuil de l'étage. Elle ne doit pas être confondue avec la tension de seuil du transistor qui est un paramètre physique. Dans le paragraphe suivant nous verrons comment calculer la tension de seuil de l'étage.

On constate alors que les valeurs de retard et de temps de montée deviennent :

Retard : $0.7\ R\ (C+2C_t)$
Temps de montée : $2.2\ R\ (C+2C_t)$

La commutation d'un PMOS se calcule de la même manière en s'appuyant sur le schéma 8.39. Dans ce cas, comme il sera vu par la suite, le transistor est relié par sa source à la tension d'alimentation V_{DD} si bien que toutes les tensions sont négatives par rapport à la source.

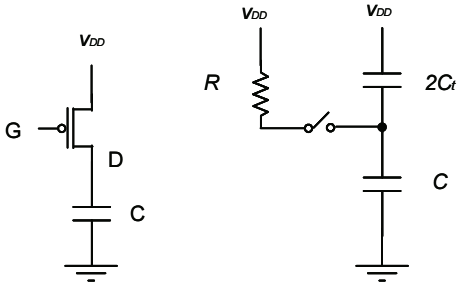


Figure 8-39 : Commutation du PMOS

Les constantes de temps de commutation s'expriment de la même manière que pour la commutation du NMOS en fonction de la valeur $R(2C_t+C)$.

On peut alors exprimer la capacité typique du transistor *Ct* en fonction des dimensions relatives du transistor pour les deux technologies envisagées.

Technologie	***R (kΩ)***	***C_t***
0.8 µ (canal long) NMOS	15 *L/W*	0.9 *W.L* (fF)
0.8 µ (canal long) PMOS	45 *L/W*	0.9 *W .L* (fF)
45 nm (canal court) NMOS	30 */W*	30 *W .L* (aF)
45 nm (canal court) PMOS	70 */W*	30 *W.L* (aF)

Rappelons que les dimensions *W* et *L* sont en unité relative. Un transistor de 450 nm de large dans la technologie 45 nm a par exemple une largeur relative de 10. Ce tableau ne fait que donner des ordres de grandeur et il est facile d'interpoler les valeurs pour d'autres technologies en se basant sur les résultats de ce chapitre.

La constante de temps qui caractérise le retard et le temps de montée doit intégrer la capacité du circuit de charge du transistor qui commute. Cette capacité, notée *C,* est un général très supérieure à la capacité C_t du transistor lui-même. Elle est due aux transistors interconnectés et surtout aux liaisons entre transistors. Une valeur typique de 50 *fF* permet de donner des ordres de grandeur.

b. L'étage inverseur

C'est véritablement la brique de base de la logique CMOS et toutes les autres fonctions logiques sont dérivées de cette cellule élémentaire. La fonction logique est le complément : un état « 0 » devient un état « 1 » et réciproquement. Elle est en pratique réalisée par la mise en série d'un PMOS et d'un NMOS comme le montre la figure 8-40.

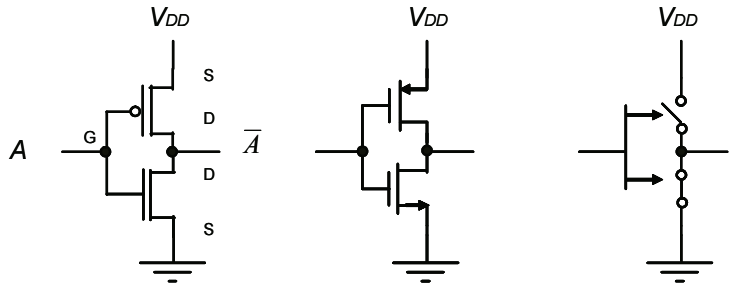


Figure 8-40 : L'inverseur CMOS

La figure 8.40 donne les trois représentations de l'inverseur : la représentation logique, électrique et fonctionnelle. Quand une tension positive, de l'ordre de V_{DD} est appliquée sur la grille, le transistor NMOS conduit et le transistor PMOS est bloqué. La connexion de sortie est reliée à la masse et isolée de la tension d'alimentation. Elle est nulle. Il y donc bien changement d'état logique. Quand une tension nulle est appliquée sur l'entrée, c'est l'inverse. Le transistor NMOS est bloqué et le transistor PMOS conduit. On remarque que, dans les deux états, un transistor est conducteur et l'autre est bloqué. Comme les deux transistors sont en série, le courant traversant l'ensemble est nul, en fait très faible dans la réalité. La notion de transistor conducteur peut sembler inadaptée dans ce raisonnement mais il faut imaginer le fonctionnement en dynamique. Les capacités du schéma se chargent et se déchargent à travers les transistors MOS.

Il est intéressant de chercher la tension pour laquelle la tension d'entrée est égale à celle de sortie. C'est la tension de commutation de l'étage inverseur qui ne doit pas être confondue avec la tension de seuil du transistor. Ecrivons le courant du NMOS.

$$I_D = \left(\frac{W}{L}\right)_n \mu_n C'_{OX} \frac{\left(V - V_{Tn}\right)^2}{2\left(1+\delta\right)}$$

soit

$$I_D = \frac{1}{2}\beta_n \left(V - V_{Tn}\right)^2$$

De même, on écrit pour le PMOS

$$I_D = \frac{1}{2}\beta_p \left(V_{DD} - V - V_{Tp}\right)^2$$

Dans cette formule, toutes les grandeurs sont positives. On obtient donc

$$V = \frac{\sqrt{\frac{\beta_n}{\beta_p}} V_{T_n} + \left(V_{DD} - V_{Tp}\right)}{1 + \sqrt{\frac{\beta_n}{\beta_p}}}$$

Si on souhaite que la tension de commutation soit égale à la moitié de la tension d'alimentation, ce qui a pour effet de symétriser le système, on obtient pour des valeurs numériques typiques la condition suivante

$$\frac{\beta_n}{\beta_p} \approx 1$$

Comme la mobilité des trous est environ trois fois plus faible que celle des électrons, on en déduit

$$\left(\frac{W}{L}\right)_p \approx 3\left(\frac{W}{L}\right)_n$$

A longueur de grille égale, le PMOS sera donc trois fois plus large que le NMOS. On comprend que cette condition est avantageuse pour assurer une garantie maximum de fonctionnement vis-à-vis du bruit. En effet, si la tension de commutation de l'étage était proche de V_{DD} ou de 0 V, une variation de la tension d'alimentation ou une variation du potentiel local de référence pourrait plus facilement être à l'origine d'un changement d'état non lié à un changement de l'état logique en entrée.

Dans une technologie canal court, la relation correspondant à cette même condition devient en négligeant la tension de saturation :

$$I_D = W_n \mu_n C'_{OX} \left[V - V_{Tn}\right] E_{Cn}$$

$$I_D = W_p \mu_p C'_{OX} \left[V_{DD} - V - V_{Tp}\right] E_{Cp}$$

On en déduit donc :

$$V = \frac{\frac{W_p \mu_p E_{Cp}}{W_n \mu_n E_{Cn}} \left(V_{DD} - V_{Tp}\right) + V_{Tn}}{\frac{W_p \mu_p E_{Cp}}{W_n \mu_n E_{cn}} + 1}$$

De la même manière, un choix des dimensions tel que $W_p \mu_p E_{Cp} / W_n \mu_n E_{Cn}$ soit égal à un conduit à une valeur de V égale à $V_{DD}/2$ ce qui est le but recherché. On obtient donc une condition géométrique en régime de canal court assez proche de celle obtenue en régime de canal long. Il faut tenir compte de la différence de valeur de la mobilité et du champ critique. Finalement, on choisit un rapport de deux entre le PMOS et le NMOS.

Il est maintenant possible de calculer le temps de commutation de l'inverseur en reprenant la méthode et les schémas du paragraphe précédent. La figure 8.41 représente le modèle électrique simplifié de l'inverseur incluant les capacités dans les deux états possibles.

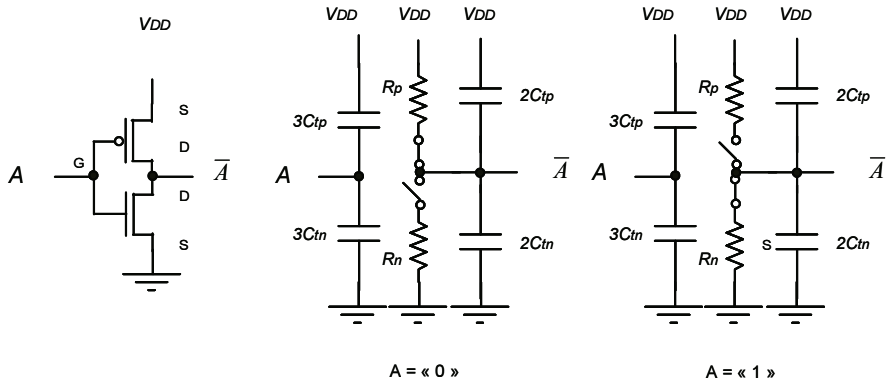


Figure 8-41 : Schéma électrique de l'inverseur

Quand l'entrée passe de « 1 » à « 0 », on suppose que le NMOS devient instantanément non conducteur et que le PMOS devient instantanément conducteur. Le schéma correspondant à A= « 0 » s'applique. La constante de temps caractéristique de la montée de la tension de sortie est alors $R_p(2C_{tp}+2C_{tn})$. Le temps de propagation de l'inverseur est donc

$$t_{PLH} = 0.7\ R_p(2C_{tp}+2C_{tn})$$

De manière symétrique, quand l'entrée passe de « 1 » à « 0 », le retard s'écrit

$$t_{PHL} = 0.7\ R_n(2C_{tp}+2C_{tn})$$

Si l'inverseur commande un autre étage équivalent à une capacité C, il faut ajouter cette valeur aux capacités des formules précédentes. Quelques exemples numériques illustrent les possibilités des technologies du point de vue de la rapidité. Un inverseur dans une technologie 45 nm réalisé avec un NMOS ayant un W/L de 10/1 et un PMOS ayant un W/L de 20/1 se caractérise donc par les temps suivants

$$t_{PHL} = 0.7\ R_p\,(2C_{tp}+2C_{tn}) = 0.7 \cdot 3.4\ \text{k} \cdot (0.6+1.2)\ \text{fF} = 4.5\ \text{ps}$$

Une simulation plus précise conduit cependant à des valeurs notablement supérieures. À cette échelle de temps, d'autres phénomènes doivent être pris en compte. Si la capacité de charge de l'étage est de 50 fF, les temps de propagation deviennent

$$t_{PLH} = 0.7\ R_pC = 0.7 \cdot 3.4\ \text{k} \cdot 50\ \text{fF} = 120\ \text{ps}$$

$$t_{PHL} = 0.7\ R_nC = 0.7 \cdot 3.4\ \text{k} \cdot 50\ \text{fF} = 120\ \text{ps}$$

Le dernier calcul à effectuer est celui de la consommation. Quand les états sont établis la consommation statique est nulle car le courant traversant l'inverseur est nul. En fait, il y a un courant résiduel appelé courant sous le seuil et un courant tunnel entre grille et drain. Ces valeurs ne sont plus négligeables dans les technologies avancées au-delà du 90 nm. Ce phénomène est d'autant plus important qu'il est permanent : un inverseur qui ne commute pas consomme en permanence de l'énergie statique. Il est maintenant possible de calculer l'énergie dynamique c'est-à-dire l'énergie dissipée dans un changement d'état. Nous avons vu en effet que lors d'un changement d'état les deux transistors sont simultanément conducteurs pendant la commutation. Ce calcul fondamental peut se faire dans un cadre assez général et est illustré figure 8.42.

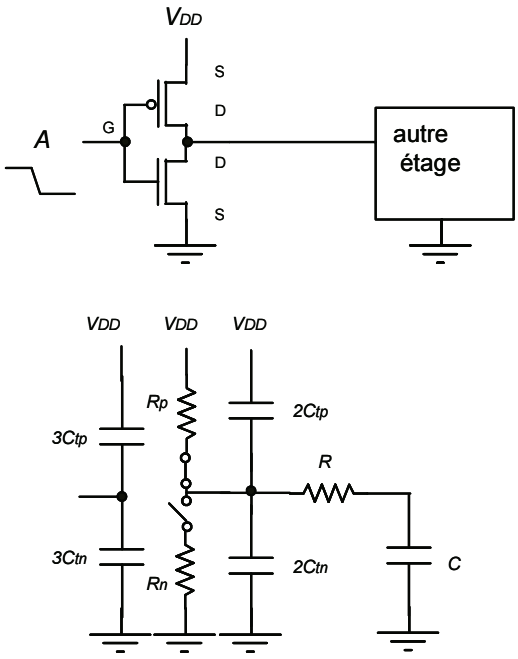


Figure 8.42 : Deux circuits logiques interconnectés

Le circuit de charge est simplement représenté par une capacité. La piste de connexion entre l'inverseur et le circuit de charge est modélisée par une résistance *R* et une capacité dont la valeur est intégrée à la capacité d'entrée *C* du circuit de charge. On suppose que la transition d'entrée fait conduire le PMOS et bloque le NMOS. Ce circuit peut se représenter sous une forme très simple illustrée figure 8.43. Le schéma peut se simplifier pour aboutir au schéma de droite.

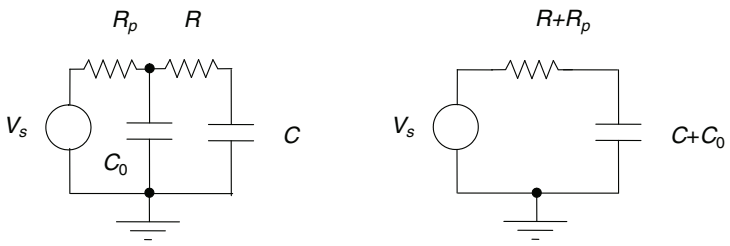


Figure 8-43 : Schéma simplifié de l'inverseur

Il est maintenant possible de calculer l'énergie dissipée lors de la commutation. L'énergie est dissipée dans les résistances. La source de tension est supposée être un échelon commutant de la valeur 0 V à la valeur V_{DD}. Le principe de ce calcul est de déterminer le courant qui circule dans ce dispositif puis de mesurer l'énergie dissipée par effet joule dans la résistance.

$$E = (R + R_p)\int_0^\infty i(t)^2 dt$$

Dans le formalisme de Laplace, on écrit étant donné la forme de la tension d'entrée

$$\frac{V_{DD}}{s} = \left(R + R_p + \frac{1}{(C + C_0)s} \right) i(s)$$

On en déduit facilement

$$i(s) = \frac{V_{DD}}{(R + R_p)} \frac{1}{s + \dfrac{1}{(R + R_p)(C + C_0)}}$$

soit,

$$i(t) = \frac{V_{DD}}{(R + R_p)} \gamma(t) \cdot \exp^{-\frac{t}{(R+R_p)(C+C_0)}}$$

On en déduit donc

$$E = (R + R_p)\int_0^\infty \frac{{V_{DD}}^2}{(R + R_p)^2} \cdot \exp^{-\frac{2t}{(R+R_p)(C+C_0)}} dt$$

Un calcul simple conduit à

$$E = \frac{1}{2}\left(C + C_0\right)\ V_{DD}{}^2$$

Il est remarquable que la puissance dissipée dans les résistances ne dépende pas de la valeur de ces résistances. Il est possible de justifier physiquement ce résultat. C'est en fait l'énergie fournie par l'alimentation qui se dissipe pour la moitié dans les résistances et pour l'autre moitié est stockée dans la capacité $C+C_0$.

L'énergie dissipée en chaleur dans la commutation est proportionnelle à la capacité totale vue en sortie et au carré de la tension d'alimentation. Quand la tension de sortie revient à 0, il y une dissipation de même valeur et quand ces deux transitions se produisent à une fréquence *f*, l'énergie dissipée par unité de temps c'est-à-dire la puissance devient

$$P = f\left(C + C_0\right)\ V_{DD}{}^2 \qquad (8\text{-}77)$$

On comprend immédiatement à la vue de cette formule l'intérêt de réduire la tension d'alimentation et l'impact de la fréquence de fonctionnement sur la consommation d'un circuit.

EXERCICES DU CHAPITRE 8

- **Exercice 1 : Courants inverses**

Expliquez pourquoi les courants inverses des diodes du MOS sont faibles. Pouvez vous identifier les transistors parasites du dispositif. Comment peut-on contrôler leurs courants ?

- **Exercice 2 : Potentiel dans un MOS**

Comparez les équations donnant le potentiel dans le cas de la capacité MOS et dans le cas du transistor MOS. Qu'en déduire pour les variations du potentiel.

- **Exercice 3 : Le courant de canal**

Démontrez 8-8

- **Exercice 4 : Tension de bandes plates**

Montrez que si la tension V_{FB} est appliquée sur la grille, alors les bandes ne dépendent pas de la profondeur.

- **Exercice 5 : Le courant en forte inversion**

Etablir la relation 8-13

- **Exercice 6 : Potentiel au niveau du drain**

Démontrez la relation donnant le potentiel au niveau du drain en régime de saturation à partir de l'expression du courant.

- **Exercice 7 : Faible et forte inversion**

Comparez les variations du potentiel en fonction de la position en faible et en forte inversion.

- **Exercice 8 : Tension de saturation**

Le dopage p est de $10^{17}/\text{cm}^3$, l'oxyde a une épaisseur de 8 nm. Calculez la tension de saturation.

- **Exercice 9 : Tensions de seuil**

Comparez les tensions de seuil en faible et en forte inversion.

- **Exercice 10 : Calcul de seuils**

Pour un dopage p de $10^{17}/cm^3$, une épaisseur d'oxyde de 8nm et un contact aluminium, calculez les seuils en faible et forte inversion.

- **Exercice 11 : Modification du seuil**

Pour un dopage p de $10^{17}/cm^3$, une épaisseur d'oxyde de 5nm et un contact aluminium, les charges de surface ont une densité de $5\ 10^{11}\ /cm^2$. Calculez la tension de seuil. Quelle dose de Bore faut-il implanter pour faire passer le seuil à 0,6 V.

- **Exercice 12 : Effet du substrat**

Pour le MOS du problème 11, quel est l'effet sur le seuil d'une polarisation de 2V de la source.

- **Exercice 13 : Expliquez la différence entre roll-off et DIBL**

- **Exercice 14 : Epaisseur limite**

Quelles sont les limites à la réduction de l'épaisseur de l'oxyde de grille. Pourquoi les oxydes à haute permittivité sont-ils si intéressants ?

- **Exercice 15 : Courant en faible inversion**

Démontrez 8-26.

- **Exercice 16 : Canal court**

Démontrez que la valeur minimale du potentiel est atteinte pour $x=L/2$.

- **Exercice 17 : Le bruit**

Démontrez 8-71.

- **Exercice 18 : Bruit de grenaille**

Démontrez 8-75.

9. Composants optoélectroniques

Le but de ce chapitre est de décrire les composants d'émission et de réception des photons. Ces composants sont très utilisés dans les produits électroniques. On peut citer les photodiodes, les diodes laser en particulier dans les lecteurs optiques, les diodes LED amenées à jouer un rôle de plus en plus important pour l'éclairage, les cellules photovoltaïques pour capter l'énergie solaire.

Dans tous les cas, il s'agit d'étudier l'interaction des photons avec le semiconducteur en distingant les semiconducteurs à gap direct et à gap indirect. Le domaine des composants optroniques est très vaste. Nous ne présenterons dans ce chapitre que les principes de base et les principaux composants. Le chapitre est organisé de la manière suivante.

9.1. Interaction rayonnement-semiconducteur
9.2. Photodétecteurs
9.3. Emetteur de rayonnement à semiconducteur
9.4. Applications des composants optroniques

9.1. Interaction rayonnement - semiconducteur

9.1.1. *Photons et électrons*

a. Photons

La nature ondulatoire du rayonnement électromagnétique est représentée par la combinaison d'un champ électrique $\mathbf{E}$ et d'un champ magnétique $\mathbf{H}$ satisfaisant aux équations de Maxwell qui s'écrivent dans un milieu non magnétique, isotrope et non chargé

$$\mathbf{rot}\,\mathbf{E} = -\mu_o \frac{\partial \mathbf{H}}{\partial t} \tag{9-1-a}$$

$$\mathbf{rot}\,\mathbf{H} = \varepsilon \frac{\partial \mathbf{E}}{\partial t} \tag{9-1-b}$$

$$div\mathbf{E}=0 \tag{9-1-c}$$

$$div\mathbf{H}=0 \tag{9-1-d}$$

où $\varepsilon=\varepsilon_r\varepsilon_o$ est la constante diélectrique du milieu. Les champs électrique et magnétique vibrent perpendiculairement l'un à l'autre et à la direction de propagation. La similarité de leurs comportements permet de ramener la représentation du rayonnement à la seule expression du champ électrique qui s'écrit

$$\mathbf{E} = \mathbf{E}_0 e^{j(\omega t - \mathbf{k}.\mathbf{r})} \tag{9-2}$$

où $\mathbf{E}_o$ représente l'amplitude et la polarisation, $\omega = 2\pi\upsilon$ la pulsation et $\mathbf{k}$ le vecteur d'onde. L'expression (9-2) est solution des équations (9-1), elle représente une onde plane monochromatique de pulsation ω se propageant dans la direction $\mathbf{r}$. En portant cette expression dans les équations (9-1) on obtient facilement la relation

$$\omega = \frac{c}{\sqrt{\varepsilon_r}} k = \frac{c}{n} k = v\,k \tag{9-3}$$

où $c = \sqrt{\varepsilon_o \mu_o}$ et $v=c/n$ représentent respectivement les vitesses de propagation de l'onde dans le vide et dans un milieu de constante diélectrique relative ε_r, c'est-à-dire d'indice de réfraction $n = \sqrt{\varepsilon_r}$.

La période de l'onde est $T=1/\upsilon$, sa longueur d'onde est $\lambda_o=cT$ dans le vide et $\lambda = cT/n$ dans un milieu d'indice n. Ainsi $\omega = 2\pi\upsilon = 2\pi/T$ *et* $k = \omega/v = 2\pi/\lambda$, de sorte que l'expression (9-2) peut s'écrire sous la forme

$$\mathbf{E}=\mathbf{E}_0 e^{j2\pi\left(\frac{t}{T}-\frac{r}{\lambda}\right)} \tag{9-4}$$

La période T représente la périodicité dans le temps et la longueur d'onde λ la périodicité dans l'espace.

Revenons sur l'expression (9-2), le terme ($\omega t-\mathbf{k.r}$) est la phase de l'onde. A un instant t donné, l'ensemble des points de l'espace correspondant à une même phase constitue une *surface d'onde*. Lorsque le temps s'écoule, les surfaces d'onde se déplacent avec la vitesse $v=\omega/k$ que l'on appelle *vitesse de phase*. L'expression (9-2) représente une onde parfaitement monochromatique qu'il est en pratique impossible de produire. Un rayonnement réel de pulsation ω correspond en fait à un groupe d'ondes monochromatiques de pulsations voisines de ω. Les interférences entre ces ondes font que leur résultante n'est différente de zéro qu'en certaines régions de l'espace où elle constitue des paquets d'ondes. Il en résulte que, contrairement à l'onde parfaitement monochromatique, qui est étendue dans tout l'espace, l'onde réelle est constituée de paquets d'ondes qui se propagent avec une vitesse $v_g = d\omega / dk$ que l'on appelle *vitesse de groupe*.

La représentation ondulatoire du rayonnement, que nous venons de rappeler brièvement, est bien adaptée à l'étude des phénomènes tels que les interférences et la diffraction, qui mettent en jeu l'interaction rayonnement-rayonnement. En ce qui concerne l'étude des interactions rayonnement-matière, et plus particulièrement lorsqu'il y a échange d'énergie, comme c'est le cas dans les composants optoélectroniques, la représentation corpusculaire du rayonnement est mieux adaptée. Einstein a suggéré que l'énergie du rayonnement n'était pas étalée dans tout l'espace mais concentrée dans certaines régions se propageant comme des particules qu'il a appelées des *photons*. L'énergie du photon est donnée par

$$E=h\upsilon=\hbar\omega \tag{9-5}$$

où h est la constante de Planck et $\hbar=h/2\pi$.

Compte tenu de l'expression (9-3), la relation de dispersion du photon, qui relie l'énergie au vecteur d'onde, s'écrit

$$E=\frac{\hbar c}{\sqrt{\varepsilon_r}}k \tag{9-6}$$

Cette relation est représentée sur la figure (9-1).

Notons en outre que, si le rayonnement est communément caractérisé par sa longueur d'onde dans le vide (ou dans l'air) mesurée en microns, le semiconducteur est quant à lui communément caractérisé par son gap mesuré en électron-Volt. Dans l'étude des composants optoélectroniques, qui mettent en jeu l'interaction rayonnement-semiconducteur, il est utile d'avoir en permanence à l'esprit la relation

énergie-longueur d'onde, pour traduire en eV la caractéristique d'un rayonnement définie en μm.

$$E = h\upsilon = \frac{h}{T} = \frac{hc}{\lambda}$$

soit

$$E(eV) = \frac{1{,}24}{\lambda(\mu m)} \qquad (9\text{-}7)$$

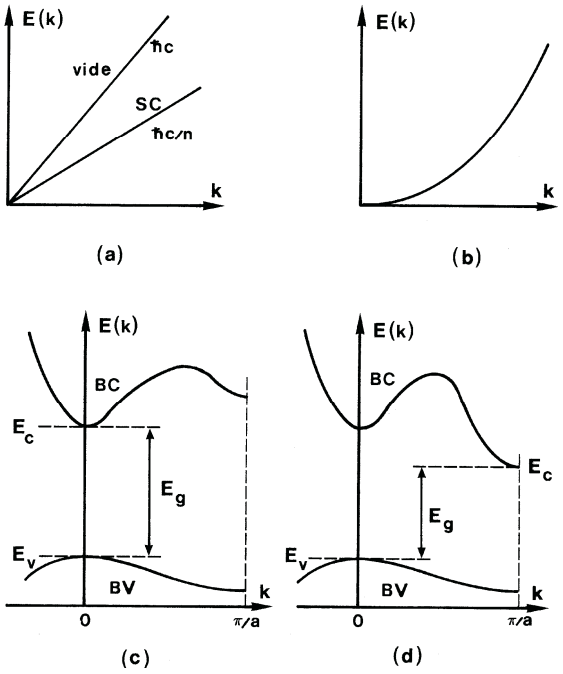


Figure 9-1 : Courbes de dispersion. a) Photons dans le vide et dans le semiconducteur. b) Electrons dans le vide. c) Electrons dans un semiconducteur à gap direct. d) Electrons dans un semiconducteur à gap indirect.

b. Electrons

Considérons tout d'abord un électron libre dans le vide, sa quantité de mouvement et son énergie cinétique sont données par

$$\mathbf{p} = m\mathbf{v} \quad E = mv^2/2 = p^2/2m \qquad (9\text{-}8)$$

De même qu'Einstein a proposé qu'à une onde plane de fréquence ν et de longueur d'onde λ, on devait associer une particule, le photon d'énergie $E=h\nu$ et de quantité de mouvement $p=h/\lambda$, Louis de Broglie a proposé, en 1924, qu'à une particule matérielle d'énergie E et de quantité de mouvement p, on devait associer une onde de fréquence $\nu=E/h$ et de longueur d'onde $\lambda=h/p$. Ainsi la quantité de mouvement de la particule et le vecteur d'onde de l'onde qui lui est associée sont liés par la relation

$$p=\frac{h}{\lambda}=\frac{h}{2\pi}\frac{2\pi}{\lambda}=\hbar k$$

ou, en représentation vectorielle

$$\mathbf{p}=\hbar\mathbf{k} \tag{9-9}$$

Il en résulte, compte tenu des relations (9-9 et 8), que la relation de dispersion de l'électron libre s'écrit

$$E=\frac{\hbar^2 k^2}{2m} \tag{9-10}$$

Cette relation est représentée sur la figure (9-1).

Considérons maintenant l'électron dans le semiconducteur. Nous avons vu (Chap. 1) qu'en raison de la périodicité du réseau cristallin, les fonctions d'onde des électrons ne sont plus des ondes planes mais des ondes de Bloch, périodiques dans l'espace. Leur énergie ne varie plus de manière continue mais présente une structure de bandes permises séparées par des bandes interdites. Les courbes de dispersion typiques sont représentées sur la figure (1-12). La caractéristique essentielle de cette structure de bande, qui va conditionner l'interaction électron-photon, est la nature du gap du semiconducteur. Nous savons qu'il existe des semiconducteurs à gap direct et des semiconducteurs à gap indirect, suivant que les extrema des bandes de valence et de conduction sont situés en un même point ou en des points différents de la zone de Brillouin, c'est-à-dire de l'espace des vecteurs d'onde k. Les courbes de dispersion d'un semiconducteur à gap direct et d'un semiconducteur à gap indirect sont schématisées sur la figure (9-1). Il faut noter que lorsque le semiconducteur est indirect, le minimum de la bande de conduction est situé en bord de zone de Brillouin (ou à son voisinage dans le cas du silicium), c'est-à-dire pour $k \approx \pi / a$ où a représente la maille du réseau cristallin.

Au voisinage de chaque extremum la courbe de dispersion présente une loi parabolique analogue à celle de l'électron dans le vide, mais avec une coubure caractérisée par la masse effective de l'électron (Eq. 1-66 et 69). Pour des semiconducteurs univallée et multivallée la loi de dispersion s'écrit respectivement

$$E = E_c + \frac{\hbar^2 k^2}{2m_e} \tag{9-11}$$

$$E = E_c + \frac{\hbar^2}{2m_\ell}\left(k_{//} - k_o\right)^2 + \frac{\hbar^2}{2m_t}k_\perp^2 \tag{9-12}$$

où k_o est voisin de π / a.

9.1.2. *Interaction électron-photon - Transitions radiatives*

L'interaction du rayonnement avec les électrons du semiconducteur se manifeste selon trois processus distincts

- un photon peut induire le saut d'un électron, d'un état occupé de la bande de valence vers un état libre de la bande de conduction, c'est l'*absorption fondamentale.* Ce processus sera mis à profit dans les capteurs de rayonnement.
- un électron de la bande de conduction peut retomber spontanément sur un état vide de la bande de valence avec émission d'un photon, c'est l'*émission spontanée*. Ce processus sera mis à profit dans les émetteurs de rayonnements tels que les diodes électroluminescentes.
- un photon présent dans le semiconducteur peut induire la transition d'un électron de la bande de conduction vers un état vide de la bande de valence, avec émission d'un deuxième photon de même énergie, c'est l'*émission stimulée*. Ce processus sera mis à profit dans les lasers à semiconducteur.

Ces différents processus sont conditionnés par les règles qui régissent les chocs élastiques entre deux particules, ici le photon et l'électron, la conservation de l'énergie et la conservation de la quantité de mouvement $\mathbf{p}$, c'est-à-dire, compte tenu de la relation (9-9), du vecteur d'onde $\mathbf{k}$. Si on repère par les indices *i* et *f*, les états initial et final de l'électron, et par l'indice *p* l'état du photon, les règles de conservation s'écrivent

$$E_f - E_i = \pm\, E_p \tag{9-13-a}$$

$$\mathbf{k}_f - \mathbf{k}_i = \pm \mathbf{k}_\mathbf{p} \tag{9-13-b}$$

où le signe + correspond à l'absorption du photon et le signe - à l'émission.

Comparons les ordres de grandeur des vecteurs d'onde des photons et des électrons. Compte tenu du fait que le gap des différents semiconducteurs est de l'ordre de 1 eV, les rayonnements mis en jeu dans les composants optoélectroniques sont caractérisés par des longueurs d'onde de l'ordre du micron (Eq. 9-7). Il en résulte que le vecteur d'onde des photons $k = 2\pi / \lambda$, est de l'ordre de 10^{-3}Å^{-1}. En ce qui concerne les électrons, leur vecteur d'onde varie de zéro au centre de la zone de

Brillouin à $k_0 = \pi / a$ en bord de zone. La maille d'un semiconducteur étant de l'ordre de quelques Å, le vecteur d'onde des électrons de bord de zone est de l'ordre de 1 $Å^{-1}$. Ainsi, à l'échelle des figures (9-1-c et d), la courbe de dispersion du photon, représentée sur la figure (9-1-a), est pratiquement verticale. En d'autres termes, si on exclut une petite zone très étroite autour de k=0, le vecteur d'onde du photon est toujours négligeable devant celui de l'électron. Il en résulte que la condition de conservation du vecteur d'onde (9-13-b) s'écrit simplement $k_f \approx k_i$. La transition d'un électron entre les bandes de valence et de conduction, se fait donc avec conservation du vecteur d'onde. On dit que les *transitions radiatives*, c'est-à-dire accompagnées de l'absorption ou de l'émission d'un photon, sont verticales dans l'espace des k.

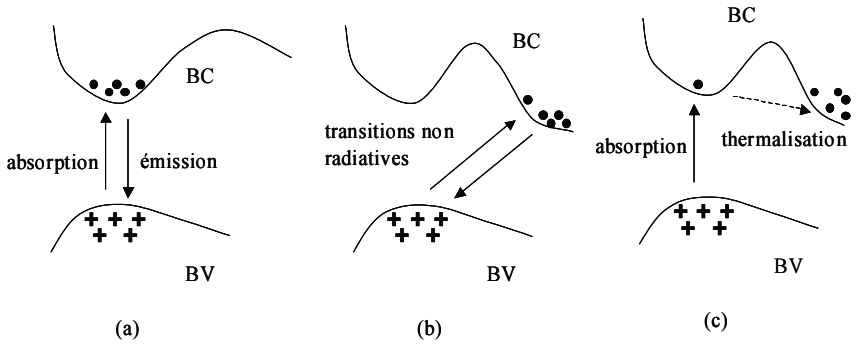


Figure 9-2 : Transitions électroniques entre les extrema des bandes de valence et de conduction. a) Semiconducteur à gap direct, les transitions sont verticales et par suite radiatives. b) Semiconducteur à gap indirect, les transitions sont obliques et non radiatives au premier ordre. c) Absorption directe de photons dans un semiconducteur à gap indirect

C'est la raison pour laquelle les processus d'absorption ou d'émission de photons, au voisinage du gap fondamental, sont considérablement plus importants dans les matériaux à gap direct que dans les matériaux à gap indirect. La figure (9-2) représente les différents types de transitions. Dans le semiconducteur à gap direct (Fig. 9-2-a), les transitions électroniques entre les extrema des bandes de valence et de conduction sont verticales, elles obéissent à la règle de conservation des k, et par suite sont radiatives. Dans le semiconducteur à gap indirect (Fig. 9-2-b), les transitions électroniques entre les extrema des bandes sont obliques et de ce fait non radiatives au premier ordre.

Notons toutefois que dans un semiconducteur à gap indirect, on peut exciter verticalement, c'est-à-dire optiquement, des électrons du sommet de la bande de valence vers le minimum central de la bande de conduction. Les électrons ainsi excités, se thermalisent ensuite dans le minimum absolu de la bande de conduction et peuvent participer aux phénomènes de conduction (Fig. 9-2-c).

Notons enfin que les règles de conservation (9-13) régissent les transitions radiatives entre les bandes de valence et de conduction, c'est-à-dire entre les états de Bloch du cristal parfait. La présence d'impuretés dans le semiconducteur pondère beaucoup la règle de conservation des k. En densité relativement faible, les impuretés créent dans la bande interdite des états discrets dont les fonctions d'onde sont des combinaisons de fonctions de Bloch, la combinaison étant d'autant plus étendue que l'électron est localisé sur l'atome d'impureté. Il en résulte que suivant la nature de l'impureté, la transition niveau discret-bande permise peut satisfaire à la règle de conservation des k, même dans un semiconducteur à gap indirect. Le centre azote dans GaP par exemple, crée des états d'énergie voisins du bas de la bande de conduction, dont le rendement radiatif particulièrement important est mis à profit dans certaines diodes électroluminescentes. Lorsque la densité d'impuretés est importante ($N_i > N_c$ ou N_v), les états discrets s'élargissent en bandes d'énergie qui atteignent les bandes permises et créent, aux extrema de celles-ci, des queues de densité d'états. Les transitions radiatives entre ces queues de densité d'états ne sont soumises à aucune règle de conservation des k la variation de quantité de mouvement de l'électron au cours de la transition étant transmise à un atome d'impureté.

9.1.3. *Absorption - Emission spontanée - Emission stimulée*

a. Semiconducteur hors équilibre - Notion de pseudo-niveau de Fermi

A l'équilibre thermodynamique les densités d'électrons et de trous n_0 et p_0 dans le semiconducteur, sont caractérisées par un seul paramètre, le niveau de Fermi. Lorsqu'une excitation extérieure modifie les densités de porteurs $n \neq n_0$ et $p \neq p_0$, le niveau de Fermi, qui est un paramètre d'équilibre thermodynamique, n'est plus défini. Les porteurs injectés dans le semiconducteur ont des énergies souvent très différentes des énergies E_c et E_v des extrema des bandes permises. Ces porteurs, par leur interaction avec le réseau cristallin, perdent de l'énergie et se thermalisent dans les extrema des bandes permises. La durée de cette thermalisation est conditionnée par le temps de collision des porteurs avec le réseau, c'est-à-dire par le temps de relaxation qui est de l'ordre de 10^{-13} à 10^{-12} s. D'un autre côté la durée de vie des porteurs dans une bande est de l'ordre de 10^{-9} à 10^{-3} s. Il en résulte que les électrons et les trous se thermalisent, respectivement dans les bandes de conduction et de valence, avant de se recombiner. Il s'établit alors dans le semiconducteur excité, un régime de pseudo-équilibre dans chacune des bandes. Les populations d'électrons et de trous sont chacune en équilibre thermodynamique, mais indépendamment l'une de l'autre. Ce régime de pseudo-équilibre est défini par des pseudo-niveaux de Fermi E_{Fn} et E_{Fp} caractérisant les densités et les distributions des électrons dans la bande de conduction et des trous dans la bande de valence. Ainsi, dans la mesure où la durée de vie des porteurs dans une bande permise est très supérieure à leur temps de relaxation, leurs distributions transitoires sont données par

$$f_{nBC}(E)=f_c(E)=\frac{1}{1+e^{(E-E_{Fn})/kT}} \qquad (9\text{-}14\text{-a})$$

$$f_{nBV}(E)=f_v(E)=\frac{1}{1+e^{(E-E_{Fp})/kT}} \qquad (9\text{-}14\text{-b})$$

A l'équilibre thermodynamique $E_{Fn}=E_{Fp}=E_F$.

b. Taux net d'émission

Le taux net d'émission de photons d'énergie E par le semiconducteur est le nombre de photons d'énergie E (de fréquence $\nu=E/h$) émis par seconde par unité de volume du cristal. C'est la différence entre le nombre de photons émis et le nombre de photons absorbés. Si le taux net d'émission est positif le matériau émet un rayonnement d'énergie E, si le taux net d'émission est négatif le matériau absorbe le rayonnement d'énergie E

$$r(E)=r_{sp}(E)+N(E)r_{stim}(E)-N(E)r_{abs}(E) \qquad (9\text{-}15)$$

$N(E)$ représente la densité de photons d'énergie E, donnée à l'équilibre thermodynamique par la formule de Planck

$$N_0(E)=1/(e^{E/kT}-1) \qquad (9\text{-}16)$$

Les deux premiers termes de l'expression (9-15) correspondent aux émissions spontanée et stimulée de rayonnement, le troisième correspond à l'absorption. En groupant les termes fonction de la densité de rayonnement présent dans le semiconducteur, l'expression (9-15) s'écrit

$$r(E)=r_{sp}(E)+N(E)r_{st}(E) \qquad (9\text{-}17)$$

où $r_{st}(E)=r_{stim}(E)$ - $r_{abs}(E)$ représente le taux net d'émission stimulée, c'est-à-dire le bilan des transitions stimulées par le rayonnement, ou en d'autres termes la différence entre les transitions stimulées de haut en bas (émission) et de bas en haut (absorption). Si le taux net d'émission stimulée est positif, tout rayonnement d'énergie E dans le matériau induit davantage de transitions de haut en bas que de bas en haut, et par suite le matériau est amplificateur de rayonnement. Dans le cas contraire le matériau est absorbant.

Considérons un système quelconque, le taux net d'émission spontanée de photons d'énergie E s'écrit

$$r_{sp}(E)=\sum_{i,j} A_{ij} g_i f_i g_j (1-f_j) \quad avec E_i-E_j=E \qquad (9\text{-}18)$$

A_{ij} représente la probabilité de transition radiative entre l'état i et l'état j , f_i la probabilité pour que l'état i soit occupé, $(1-f_j)$ la probabilité pour que l'état j soit vide, g_i et g_j sont les densités d'états. La somme est étendue à tous les états i et j tels que $E_i-E_j=E$. Le taux d'émission stimulée s'écrit

$$r_{st}(E)=\sum_{i,j} A_{ij}\, g_i\, g_j \left[f_i(1-f_j)-f_j(1-f_i) \right]$$

le premier terme représente l'émission stimulée, le deuxième l'absorption. L'expression s'écrit

$$r_{st}(E)=\sum_{i,j} A_{ij}\, g_i\, g_j \left[f_i - f_j \right] \qquad (9\text{-}19)$$

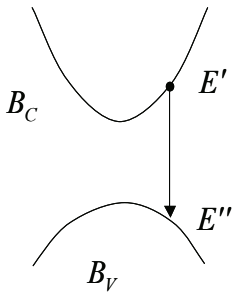


Figure 9-3 : Émission de photon

Considérons maintenant le cas d'un semiconducteur (Fig. 9-3). L'émission d'un photon d'énergie E résulte de transitions électroniques entre les états d'énergie E' dans la bande de conduction, et des états d'énergie E'' dans la bande de valence, se correspondant verticalement et tels que $E'-E''=E$. La densité de photons émis spontanément dans ces transitions est proportionnelle au nombre d'électrons dans l'état E', au nombre de places dans l'état E'', et à la probabilité de transition radiative $E' \rightarrow E''$ soit

$$r_{sp}(E',E'') = B(E',E'')\, N_c(E')\, f_c(E')\, N_v(E'')\left[1-f_v(E'')\right]$$

On se ramène à une seule variable en utilisant la règle de conservation de l'énergie $E''=E'-E$.

$$r_{sp}(E',E) = B(E',E)\, N_c(E')\, f_c(E')\, N_v(E'-E)\left[1-f_v(E'-E)\right]$$

On obtient le taux d'émission spontanée en intégrant sur toutes les valeurs de E'

$$r_{sp}(E)=\int_{E'} B(E',E) N_c(E') f_c(E') N_v(E'-E)\left[1-f_v(E'-E)\right] dE' \qquad (9\text{-}20)$$

De même le taux net d'émission stimulée s'écrit

$$r_{st}(E)=\int_{E'} B(E',E)\{N_c(E')f_c(E')N_v(E'-E)[1-f_v(E'-E)] \\ -N_c(E')[1-f_c(E')]N_v(E'-E)f_v(E'-E)\}dE'$$

soit

$$r_{st}(E)=\int_{E'} B(E',E)N_c(E')N_v(E'-E)[f_c(E')-f_v(E'-E)]dE' \qquad (9\text{-}21)$$

Avec

$$f_c(E')=\frac{1}{1+e^{(E'-E_{Fc})/kT}} \qquad f_v(E'-E)=\frac{1}{1+e^{(E'-E-E_{Fv})/kT}}$$

Dans la pratique, les matériaux utilisés en optoélectronique sont souvent très dopés de sorte que les extrema des bandes sont perturbés et la règle de conservation des *k* disparaît. Il en résulte que la probabilité de transition *B(E',E)* est sensiblement constante. On écrit

$$r_{sp}(E)=B\int_{E'} N_c(E')N_v(E'-E)f_c(E')[1-f_v(E'-E)]dE' \qquad (9\text{-}22\text{-a})$$

$$r_{st}(E)=B\int_{E'} N_c(E')N_v(E'-E)[f_c(E')-f_v(E'-E)]dE' \qquad (9\text{-}22\text{-b})$$

La constante *B* a été calculée pour la plupart des semiconducteurs à 300K

Semiconducteur	Type de gap	B ($cm^3\ s^{-1}$)
Si	indirect	$1{,}8\ 10^{-15}$
Ge	"	$5{,}3\ 10^{-14}$
GaP	"	$5{,}4\ 10^{-14}$
GaAs	direct	$7{,}2\ 10^{-10}$
GaSb	"	$2{,}4\ 10^{-10}$
InP	"	$1{,}3\ 10^{-9}$
InAs	"	$8{,}5\ 10^{-11}$
InSb	"	$4{,}6\ 10^{-11}$

c. Emission spontanée

L'expression (9-22-a) représente le taux d'émission spontanée de photons d'énergie *E*. Le taux d'émission spontanée global, c'est-à-dire le nombre total de photons émis par seconde par l'unité de volume du semiconducteur, est donné par

$$R_{sp}=\int_0^\infty r_{sp}(E)dE=\int_{E_g}^\infty r_{sp}(E)dE \qquad (9\text{-}23)$$

En explicitant $r_{sp}(E)$ et en intégrant sur *E* avec *B* constant, on obtient

$$R_{sp}=Bnp \tag{9-24}$$

d. Emission stimulée

A partir des expressions (9-22-a et b) et en explicitant les fonctions de distribution, on peut exprimer le taux d'émission stimulée $r_{st}(E)$ en fonction du taux d'émission spontanée $r_{sp}(E)$

$$r_{st}(E) = r_{sp}(E)\left(1 - e^{(E-\Delta F)/kT}\right) \tag{9-25}$$

avec $\Delta F=E_{Fc}-E_{Fv}$

L'émission stimulée existe réellement lorsque le taux net d'émission stimulée est positif, ($r_{st}(E)>0$). La condition d'amplification de rayonnement par le semiconducteur s'écrit donc compte tenu de (9-25)

$$\Delta F > E \tag{9-26}$$

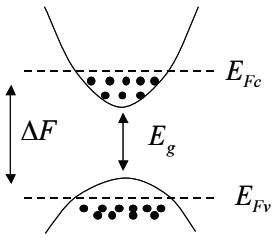


Figure 9-4 : Émission stimulée

Compte tenu de la règle de conservation de l'énergie, l'énergie E du rayonnement émis par le semiconducteur ne peut être inférieure au gap E_g. La condition (9-26) s'écrit alors

$$E_{Fc} - E_{Fv} > E_g \tag{9-27}$$

Il en résulte que le pseudo-niveau de Fermi des électrons E_{Fc} est dans la bande de conduction et le pseudo-niveau des trous E_{Fv} dans la bande de valence (Fig. 9-4). La condition (9-27) traduit une *inversion de population* entre le sommet de la bande de valence et le bas de la bande de conduction.

e. Régimes de fonctionnement

On peut écrire la densité de photons d'énergie E dans le matériau sous la forme

$$N(E)=N_0(E)+\Delta N(E)$$

où $N_0(E)$ représente la densité de photons d'énergie E à l'équilibre thermodynamique, donnée par l'expression (9-16), et $\Delta N(E)$ l'écart par rapport à l'équilibre.

Compte tenu de l'expression (9-25), le taux net d'émission (Eq. 9-17) s'écrit en explicitant $N_0(E)$ (Eq. 9-16)

$$r(E)=r_{sp}(E)\left(1+\left[1+\frac{\Delta N(E)}{N_o(E)}\right]\left[\frac{1-e^{(E-\Delta F)/kT}}{e^{E/kT}-1}\right]\right) \tag{9-28}$$

- A l'équilibre thermodynamique, $\Delta N(E)=0$, $\Delta F=0$, l'expression (9-28) s'écrit

$$r(E)=r_{sp}(E)\left(1-\frac{1-e^{E/kT}}{1-e^{E/kT}}\right)=0 \tag{9-29}$$

Le taux net d'émission est nul, les émissions spontanée et stimulée sont compensées par l'absorption.

- Lorsque le semiconducteur est faiblement excité par un rayonnement d'énergie E supérieure à E_g , $\Delta N(E) \neq 0$ avec $\Delta F \approx 0$. L'expression (9-28) s'écrit

$$r(E)\approx -r_{sp}(E)\frac{\Delta N(E)}{N_o(E)}<0 \tag{9-30}$$

Le taux net d'émission de rayonnement d'énergie E est négatif, le semiconducteur absorbe le rayonnement.

- Lorsque le semiconducteur est fortement excité par un rayonnement d'énergie E supérieure à E_g , $\Delta N(E) \neq 0$ avec $\Delta F > 0$. Compte tenu des relations $E > E_g$ et $E_g >> kT/e$ on peut négliger 1 devant l'exponentielle au dénominateur de l'expression (9-28) qui s'écrit

$$r(E)=r_{sp}(E)\left(1+\left(1+\frac{\Delta N(E)}{N_o(E)}\right)\left(e^{-E/kT}-e^{-\Delta F/kT}\right)\right)>0 \tag{9-31}$$

E et ΔF sont positifs, de sorte que les deux exponentielles sont inférieures à 1, $r(E)$ est positif, le semiconducteur émet un rayonnement d'énergie E.

9.1.4. *Recombinaison de porteurs en excès - Durée de vie*

a. Durée de vie radiative

L'émission de photons résulte de la recombinaison radiative de porteurs, ainsi le taux global d'émission de photons R correspond au taux de recombinaisons radiatives de porteurs

$$\begin{aligned} R &= \int_E \left(r_{sp}(E) + N(E) r_{st}(E) \right) dE \\ &= \int_E r_{sp}(E) dE + \int_E N(E) r_{st}(E) dE \\ &= R_{sp} + R_{st} \end{aligned}$$

Compte tenu de (9-24)

$$R = Bnp + R_{st} \tag{9-32}$$

A l'équilibre thermodynamique $n = n_0$, $p = p_0$ et $R = R_0 = B\, n_0 p_0 + R_{sto} \equiv 0$. En présence d'un excès de porteurs $n = n_0 + \Delta n$, $p = p_0 + \Delta p$ et

$$R = R_0 + \Delta R = B(n_0 + \Delta n)(p_0 + \Delta p) + R_{sto} + \Delta R_{st} \tag{9-33}$$

Si Δn et Δp sont faibles par rapport au régime d'inversion de population, $\Delta R_{st} \approx 0$ et l'expression (9-33) s'écrit

$$R_0 + \Delta R = Bn_0 p_0 + R_{sto} + B(n_0 \Delta p + p_0 \Delta n + \Delta n\, \Delta p) \tag{9-34}$$

Soit en négligeant le terme du deuxième ordre

$$\Delta R = Bn_0 \Delta p + Bp_0 \Delta n \tag{9-35}$$

ou

$$\Delta R = \frac{\Delta p}{\tau_{rp}} + \frac{\Delta n}{\tau_{rn}} \tag{9-36}$$

avec

$$\tau_{rp} = 1 / Bn_o \qquad \tau_{rn} = 1 / Bp_o$$

ΔR représente le taux de recombinaisons radiatives des porteurs, τ_{rn} et τ_{rp} représentent les durées de vie radiatives. Dans un matériau de type p par exemple, le taux de recombinaisons radiatives des porteurs minoritaires que sont les électrons s'écrit $R = \Delta n / \tau_{rn}$ avec $\tau_{rn} = 1/Bp_0$.

b. Durée de vie non radiative

Outre les processus radiatifs, d'autres processus tels que transfert sur centre de recombinaison, émission de phonons ou effet Auger, peuvent entraîner la recombinaison de porteurs. Ces autres processus conditionnent la durée de vie des porteurs par une contribution τ_{nr}, que l'on appelle *durée de vie non radiative*.

Compte tenu de leur définition, les taux de recombinaison sont additifs, ces taux varient en $1/\tau$, de sorte que la durée de vie résultante des porteurs est donnée par

$$\frac{1}{\tau}=\frac{1}{\tau_r}+\frac{1}{\tau_{nr}} \tag{9-37}$$

Si $\tau_r << \tau_{nr}$, $\tau \approx \tau_r$ la durée de vie des porteurs est conditionnée par les processus radiatifs, les recombinaisons de porteurs se font avec émission de photons, le matériau a un bon rendement radiatif. Au contraire si $\tau_r >> \tau_{nr}$, $\tau \approx \tau_{nr}$, les porteurs excédentaires se recombinent par d'autres voies que l'émission de photons, le matériau a un mauvais rendement radiatif.

Le rendement radiatif d'un semiconducteur est mesuré par le rapport

$$\eta=\frac{1/\tau_r}{1/\tau}=\frac{\tau_{nr}}{\tau_r+\tau_{nr}} \tag{9-38}$$

si $\tau_r << \tau_{nr}$, $\eta \approx 1$

c. Vitesse de recombinaison en surface

Nous avons vu (Par. 1.6.3) qu'en raison de phénomènes intrinsèques (disparition de la périodicité du réseau) et extrinsèques (adsorption d'atomes étrangers), la surface du semiconducteur est le siège d'états spécifiques, appelés *états de surface*, dont les niveaux d'énergie peuvent se situer dans le gap. Certains de ces états jouent le rôle de *centre de recombinaison*. La durée de vie des porteurs en surface est de ce fait toujours inférieure à leur durée de vie en volume. Il en résulte que dans un semiconducteur excité, la densité de porteurs excédentaires en surface est toujours inférieure à sa valeur en volume. Ce gradient de concentration donne naissance à un courant de diffusion qui s'écrit

$$j=\pm e\,D_n\frac{d\,\Delta n}{dx} \tag{9-39}$$

Le signe ± est fonction de l'orientation de l'axe *x* normal à la surface et précise que le flux de porteurs est toujours dirigé vers la surface du semiconducteur.

Afin de définir un paramètre pour caractériser la surface du semiconducteur, on exprime le courant *j* sous une autre forme en écrivant que le courant en surface est proportionnel à la densité de porteurs excédentaires en surface et à leur vitesse

$$j = e\ \Delta n\ s \tag{9-40}$$

Compte tenu des expressions (9-39 et -40) le paramètre s est donné par

$$s = \pm \frac{D_n}{\Delta n} \frac{d\Delta n}{dx} \tag{9-41}$$

s est appelé *vitesse de recombinaison en surface*, Δn et $d\Delta n/dx$ sont respectivement la densité de porteurs excédentaires et son gradient, à la surface.

La même expression peut être définie au niveau d'un contact ohmique entre un semiconducteur et un métal, où la soudure perturbe beaucoup le réseau cristallin. On peut ainsi caractériser le contact par une vitesse de recombinaison au contact. En fait, la perturbation au niveau d'une soudure est telle que cette vitesse de recombinaison est très importante. On traduit alors la présence du contact en écrivant que la durée de vie des porteurs excédentaires est nulle, c'est-à-dire que leur densité est nulle.

9.1.5. *Création de porteurs en excès*

Plusieurs processus peuvent être utilisés pour créer un excès de porteurs dans un semiconducteur.

a. Photoexcitation

L'excès de porteurs par rapport à l'équilibre est créé à l'aide d'un rayonnement (Fig. 9-5). Soit $\phi_0(E)$ le flux de photons incidents d'énergie E, c'est-à-dire de longueur d'onde $\lambda(\mu m)$=1,24 / E(eV). C'est le nombre de photons d'énergie E qui frappent l'unité de surface du semiconducteur par seconde, il s'exprime en ph/s/cm^2.

Si $R(E)$ est le coefficient de réflexion du semiconducteur pour le rayonnement d'énergie E, le flux de photons $\phi_t(E)$ transmis à la surface, c'est-à-dire entrant effectivement dans le semiconducteur, s'écrit

$$\phi_t(E) = (1 - R(E))\ \phi_0(E) \tag{9-42}$$

Les photons sont absorbés au cours de leur propagation dans le matériau. On définit le *coefficient d'absorption* α du matériau comme la variation relative de la densité de rayonnement par unité de longueur. Ainsi, si la densité de rayonnement varie de $d\phi$ sur une longueur dx, α est donné par

$$\alpha(E,x) = -\frac{1}{dx} \frac{d\phi(E,x)}{\phi(E,x)} \tag{9-43}$$

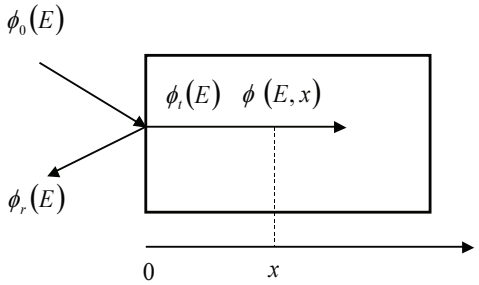


Figure 9-5 : Photoexcitation

α est positif lorsque $d\phi$ est négatif, c'est-à-dire lorsqu'il y a absorption du rayonnement. Si α est constant dans tout le matériau, $\alpha(E,x)=\alpha(E)$ et l'intégration de l'expression (9-43) donne

$$\frac{d\phi(E,x)}{\phi(E,x)} = -\alpha(E)dx \quad Ln(\phi(E,x)) = Ln(\phi_t(E)) - \alpha(E)x$$

soit

$$\phi(E,x) = \phi_t(E)\, e^{-\alpha(E)x} \tag{9-44}$$

Le flux de photons d'énergie E à l'abscisse x à l'intérieur du semiconducteur s'écrit donc en fonction du flux incident

$$\phi(E,x) = (1-R(E))\phi_0(E)e^{-\alpha(E)x} \tag{9-45}$$

Si $\alpha(E)$ est nul, le rayonnement d'énergie E traverse le matériau sans atténuation, le matériau est transparent à ce rayonnement. C'est le cas des semiconducteurs intrinsèques pour des rayonnements d'énergie E tels que $E<E_g$. Si par contre $\alpha(E)$ n'est pas nul, le matériau absorbe le rayonnement d'énergie E , qui s'atténue alors exponentiellement au cours de sa propagation. Cette absorption se traduit par la création de paires électron-trou. Chaque photon absorbé crée une paire électron-trou, de sorte qu'en un point d'abscisse x le nombre de paires créées par seconde est égal au nombre de photons disparus. Le taux de génération de paires électron-trou et donc égal au taux de disparition de photons,

$$g(E,x) = -\frac{d\phi(E,x)}{dx} \tag{9-46}$$

Soit, compte tenu de (9-45)

$$g(E,x) = (1-R(E))\ \phi_o(E)\ \alpha(E)\ e^{-\alpha(E)x} \tag{9-47}$$

Si le rayonnement n'est pas monochromatique, le taux global de génération de paires électron-trou au point d'abscisse x est obtenu en intégrant l'expression précédente sur tout le spectre, soit

$$g(x) = \int_E g(E,x)dE \tag{9-48}$$

Si $\alpha(E)$ est petit, le rayonnement est peu absorbé, il pénètre alors profondément dans le matériau. Il crée peu de porteurs par unité de volume mais il en crée dans un grand volume. Au contraire si $\alpha(E)$ est grand le rayonnement est très absorbé, les porteurs sont créés dans un petit volume sous la surface.

L'intégration de l'expression (9-48) nécessite la connaissance des lois de variation du coefficient de réflexion *R(E)* et du coefficient d'absorption $\alpha(E)$ du semiconducteur avec l'énergie.

Le coefficient de réflexion $R(E)$ est fonction de la nature du semiconducteur, mais en règle générale varie peu avec l'énergie pour des rayonnements dont l'énergie est voisine du gap du matériau, on peut écrire $R(E)=R$. Par contre sa valeur est très sensible à l'angle d'incidence du rayonnement, elle est minimum en incidence normale où elle est donnée par

$$R=\left(\frac{n-1}{n+1}\right)^2=\left(\frac{\sqrt{\varepsilon_r}-1}{\sqrt{\varepsilon_r}+1}\right)^2 \tag{9-49}$$

n est l'indice de réfraction du matériau et ε_r sa constante diélectrique relative. Pour tous les semiconducteurs n varie entre 3 et 4 de sorte que R varie entre 0,25 et 0,35. Ainsi, dans les meilleures conditions c'est-à-dire en incidence normale, environ 30% du rayonnement excitateur est réfléchi par le semiconducteur.

La loi de variation de $\alpha(E)$, est sensiblement la même pour tous les semiconducteurs à gap direct, elle est représentée sur la figure (9-6) pour le GaAs à la température ambiante. Ses caractéristiques essentielles, communes à tous les semiconducteurs à gap direct, sont d'une part une variation sensiblement en marche d'escalier, $\alpha(E)\approx 0$ pour $E<E_g$ et $\alpha(E)\approx\alpha$=Cte pour $E>E_g$, et d'autre part l'ordre de grandeur, $\alpha\approx 10^4$ cm^{-1}. Compte tenu de la loi exponentielle (9-47), il faut noter que pour $x = 1/\alpha \approx 1\ \mu m$, $g(x) = g(x=0)/e \approx g(x=0)/3$. En d'autres termes les photoporteurs sont créés sur une épaisseur de quelques microns à partir de la surface. Il en résulte que la vitesse de recombinaison en surface joue un rôle fondamental dans le rendement des composants optoélectroniques.

Dans la mesure où le coefficient α(E) est sensiblement constant pour $E>E_g$ tant que E n'est pas trop important, on peut écrire $\alpha(E)=\alpha$ et intégrer l'expression (9-48)

$$g(x)=(1-R)\alpha e^{-\alpha x}\int_{E_g}^{\infty}\phi_o(E)dE \tag{9-50}$$

En appelant ϕ_0 , la densité totale de rayonnement d'énergie supérieure au gap du semiconducteur, l'expression (9-50) s'écrit

$$g(x) = (1-R)\,\phi_o\,\alpha e^{-\alpha x} \tag{9-51}$$

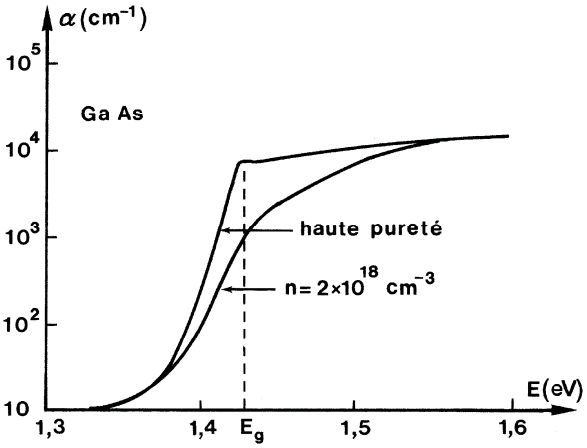


Figure 9-6 : Coefficient d'absorption de GaAs à la température ambiante (Casey H.C. 1975)

b. Injection électrique

Lorsque l'on polarise une jonction pn dans le sens direct, les électrons majoritaires dans la région de type n, sont injectés dans la région de type p où ils deviennent excédentaires par rapport à l'équilibre thermodynamique et diffusent en se recombinant. Si les recombinaisons sont radiatives, cette partie de la jonction pn est le siège d'une émission de rayonnement. Il en est de même de la région de type n avec l'injection de trous.

Les densités de porteurs minoritaires injectés dans chacune des régions de la jonction sont données par les expressions (2-46-a et b), soit

$$\Delta p(x)=\frac{p_n}{sh(d_n/L_p)}\left(e^{eV/kT}-1\right)sh\left((x_c-x)/L_p\right) \qquad (9\text{-}52\text{-a})$$

$$\Delta n(x)=\frac{n_p}{sh(d_p/L_n)}\left(e^{eV/kT}-1\right)sh\left((x-x'_c)/L_n\right) \qquad (9\text{-}52\text{-b})$$

où $d_n = x_c - x_n$ et $d_p = x_p - x'_c$, représentent les longueurs des régions neutres n et p de la diode. L_n et L_p sont les longueurs de diffusion des porteurs minoritaires.

c. Autres processus

D'autres processus physiques permettent de créer des porteurs excédentaires dans un semiconducteur. Notons *l'effet d'avalanche* dans une jonction pn polarisée en inverse, ou *l'effet tunnel* dans une diode tunnel ou une jonction Schottky. La *cathodoexcitation*, qui consiste à bombarder le semiconducteur par un faisceau d'électrons, permet aussi par ionisation par choc de créer des paires électron-trou. Notons enfin *l'effet Destriau* , utilisé dans les panneaux électroluminescents. Une poudre de semiconducteur, généralement un composé II-VI, enrobée dans un liant isolant, est prise en sandwich entre deux armatures dont l'une est transparente. L'application d'une tension aux plaques du condensateur permet une injection d'électrons dans la bande de conduction, par choc ou par effet tunnel, à partir de niveaux accepteurs.

9.1.6. *Semiconducteurs pour l'optoélectronique*

Le choix d'un semiconducteur pour la réalisation d'un composant optoélectronique est évidemment fonction du type d'utilisation du composant. Ainsi, la courbe de sensibilité de l'œil conditionne le choix des matériaux destinés à la réalisation de systèmes d'affichage, de même le spectre solaire pour les convertisseurs d'énergie, les fenêtres de transparence des fibres optiques pour les télécommunications ou la nature infrarouge des applications militaires. Enfin la lecture des disques optiques, CD (Compact Disk) et des vidéo disques numériques DVD (Digital Versatile ou Video Disk), nécessite des émissions laser de longueur d'onde λ de plus en plus courte, la densité de stockage étant proportionnelle à λ^{-2}.

a. Spectre de sensibilité de l'œil

A la suite de mesures effectuées sur un grand nombre d'individus, une commission internationale a établi la courbe représentée sur la figure (9-7). Le maximum de sensibilité de l'œil se situe dans le jaune (λ=0,555 μ) en éclairement normal, et dans le vert (λ=0,513 μ) en éclairement atténué.

b. Spectre solaire

Le rayonnement émis par le soleil correspond à celui du corps noir à la température de 6000°. L'intensité du rayonnement au-dessus de l'atmosphère est de 1,35 kW/m^2, avec un spectre centré au voisinage de λ=0,48 μ.

A la surface du sol la densité de puissance n'est plus que de 0,9 kW/m^2, en raison de l'absorption essentiellement par l'ozone, l'eau et le gaz carbonique. En outre, le spectre n'est plus continu mais présente des bandes d'absorption. Pour mesurer l'effet de l'atmosphère on utilise l'*Air Masse*, défini par AM=1/cos α où α représente l'angle que fait la direction du soleil avec la verticale. AM1 correspond au soleil au zenith (α=0) et AM4 à l'horizon (α=75°). AM1,5 (α=48,19°) est la référence généralemant utilisée pour comparer les performances des piles solaires.

AM0 est utilisé pour préciser les conditions au-dessus de l'atmosphère. Le spectre solaire est représenté sur la figure (9-8).

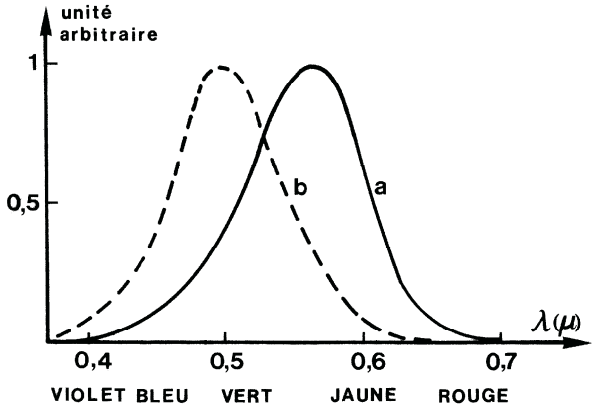


Figure 9-7 : Spectre de sensibilité de l'œil. a) Eclairement moyen. b) Eclairement atténué

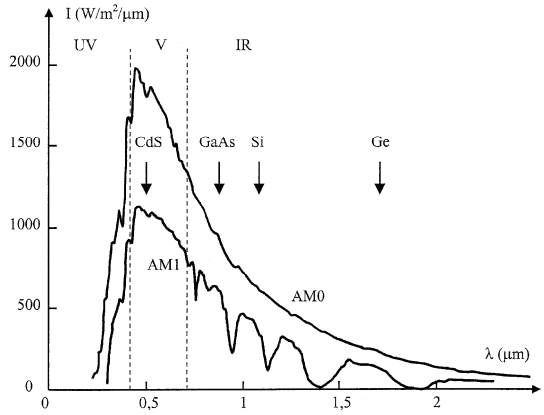


Figure 9-8 : Spectre solaire

c. Spectre de transparence des fibres optiques

Deux phénomènes conditionnent la transmission d'un signal lumineux à travers une fibre optique, *l'atténuation* et la *dispersion*.

L'atténuation a deux causes principales, la *diffusion* et *l'absorption* du rayonnement. Les fibres optiques étant réalisées à partir de verres, qui sont des structures désordonnées, la *diffusion* résulte des défauts de structure et de composition. Ces défauts créent des fluctuations d'indice qui sont autant de centres diffusants. Cette diffusion Rayleigh se traduit par un coefficient d'absorption

effectif, qui varie comme λ^{-4}. *L'absorption* du rayonnement est essentiellement due, dans le visible et le proche-infrarouge, à la présence d'impuretés à l'état de traces et dans l'infrarouge aux vibrations de réseau. Si P_1 est la puissance du rayonnement à l'entrée de la fibre, et P_2 sa puissance après un parcours de L km, *l'atténuation* est mesurée par la quantité $1/L.Ln(P_1/P_2)$ *dB*/km. La courbe d'atténuation typique d'une fibre de silice est représentée sur la figure (9-9). Les plus faibles atténuations sont obtenues à l'heure actuelle avec des fibres de SiO_2 dopées GeO_2, qui présentent un minimum d'atténuation de 0,2 dB/km à λ = 1,55 μ. Cette atténuation minimum résulte de la diffusion Rayleight, elle pourra être abaissée dans la mesure où l'on repoussera vers l'infrarouge la queue d'absorption due aux vibrations de réseau.

La *dispersion* du rayonnement a des causes multiples. Lorsque plusieurs modes se propagent dans le guide d'onde que constitue la fibre, ces modes se propagent avec des vitesses différentes, même si le rayonnement est monochromatique, ce qui donne naissance à une *dispersion intermode*. Ainsi lorsqu'une impulsion de rayonnement incident excite plusieurs modes, la différence des vitesses de propagation des différents modes entraîne un élargissement de l'impulsion dans le temps. On peut toutefois éviter ce type de dispersion en utilisant des fibres dont le diamètre est suffisamment petit pour n'autoriser la transmission que d'un seul mode. Cependant d'autres sources de dispersion existent si le rayonnement n'est pas purement monochromatique : la dispersion due au guide d'onde par le fait que la vitesse de propagation d'un mode est fonction de sa longueur d'onde, et la dispersion propre au matériau.

La dispersion propre au matériau est due à la variation de l'indice de réfraction avec la longueur d'onde. Une impulsion de rayonnement est un paquet d'ondes dont la vitesse de groupe est donnée par $v_g=d\omega/dk$. Puisque $\omega=2\pi\upsilon$ et $k=2\pi/\lambda$ cette vitesse s'écrit

$$v_g=\frac{d\upsilon}{d(1/\lambda)}=-\lambda^2\frac{d\upsilon}{d\lambda} \tag{9-53}$$

compte tenu de la relation $\upsilon=c/n\lambda$, v_g s'écrit

$$v_g=-c\lambda^2\left(-\frac{1}{n\lambda^2}-\frac{1}{n^2\lambda}\frac{dn}{d\lambda}\right)=\frac{c}{n}\left(1-\frac{\lambda}{n}\frac{dn}{d\lambda}\right) \tag{9-54}$$

Si la largeur spectrale du rayonnement est $\Delta\lambda$, l'étalement de la vitesse de groupe est $\Delta v_g = dv_g/d\lambda\cdot\Delta\lambda$, soit

$$\Delta v_g=\frac{c\lambda}{n^2}\left(\frac{d^2n}{d\lambda^2}-\frac{2}{n}\left(\frac{dn}{d\lambda}\right)^2\right)\Delta\lambda \tag{9-55}$$

Il en résulte que l'étalement dans le temps d'une impulsion brève, après un trajet d'une distance L dans le matériau, est donnée par

$$\Delta\tau=\left|\frac{L\,\Delta v_g}{v_g^2}\right|=\left|\frac{L\lambda}{c}\frac{d^2n}{d\lambda^2}\Delta\lambda\right| \tag{9-56}$$

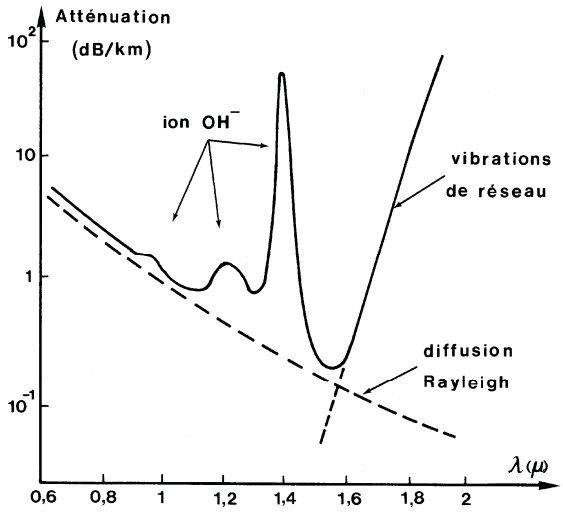


Figure 9-9 : Courbe d'atténuation d'une fibre de silice (Wilson J.1983)

La courbe représentant les variations de $d^2n/d\lambda^2$ en fonction de la longueur d'onde, est portée sur la figure (9-10) pour la silice pure. Si on considère une fibre de silice de 1km de longueur, excitée par une diode électroluminescente au GaAs émettant un rayonnement de longueur d'onde λ=8500 Å et de largeur $\Delta\lambda$=500 Å, avec $d^2n/d\lambda^2=3.10^{10}$ m^{-2}, l'expression (9-56) donne $\Delta\tau=4.10^{-9}$ s.

La figure (9-10) montre que $d^2n/d\lambda^2$ et par suite la dispersion du matériau, s'annule pour λ=1,3μm et change de signe au-delà. Il existe donc une longueur d'onde pour laquelle la dispersion due au matériau peut compenser les autres causes de dispersion, ce qui permet d'annuler la dispersion totale. Le calcul exact de cette longueur d'onde est assez complexe mais on peut montrer (Gambling 1979) qu'elle est comprise entre 1,3 et 1,8 μm. Compte tenu des minima de la courbe d'atténuation (Fig. 9-9), il en résulte que les longueurs d'onde les mieux adaptées à la transmission par fibre optique de silice sont de 1,3 et 1,55 μm.

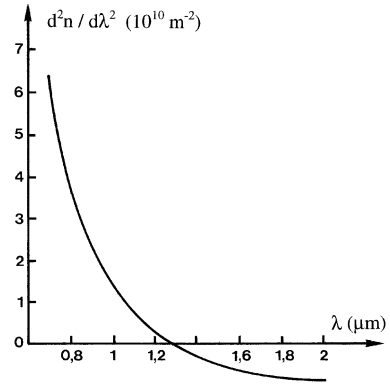


Figure 9-10 : Variation de $d^2n/d\lambda^2$ pour une fibre de silice pure

d. Disques optiques

Les CD-ROM et DVD-ROM (Read Only Memory) utilisent pour leur lecture des diodes lasers qui détectent, par réflectivité différentielle, les creux imprimés à

la fabrication dans un sillon gravé en spirale dans une couche de polycarbonate déposée sur la surface du disque. Une puissance de 5 mW est suffisante pour cette lecture. Les CD-RW (ReWritable) et DVD-RAM utilisent des lasers plus puissants (~50 mW) qui lisent de la même manière mais permettent aussi d'effacer et de réécrire. Ces disques mettent à profit la différence de réflectivité de régions amorphes et polycristallines d'un matériau tel que AgInSbTe. Le matériau initialement polycristallin est réfléchissant. L'écriture est réalisée par le spot laser qui élève localement la température à une valeur supérieure à la température de fusion (500-700 °C), une trempe rapide crée alors une région amorphe peu réfléchissante. L'effacement est réalisé avec le même laser, les régions amorphes sont chauffées vers 200 °C et recuites ce qui permet la recristallisation du matériau.

La taille minimum du spot est limitée à la longueur d'onde du laser par la diffraction du faisceau optique, il en résulte que la densité de stockage varie comme $1/\lambda^2$. Les CD actuels utilisent des diodes lasers infrarouges à AlGaAs dont la longueur d'onde d'émission est λ=780 nm, le diamètre du spot est de 1,6 µm, ces disques ont une capacité de stockage de 650 Mo. Les DVD utilisent des diodes lasers rouges à AlGaInP émettant à 650 nm, le diamètre du spot est de 0,74 µm, la capacité de stockage est de 4,7 Go par face. Enfin la nouvelle génération de DVD à haute densité (HD-DVD) augmente la densité de stockage jusqu'à 15 Mo par face par l'utilisation de lasers bleus à GaN qui en émettant à λ=410 nm permettent de réduire le diamètre du spot jusqu'à 0,24 µm. La courbe *$\lambda(E)$* (Eq. 9-7) et les coefficients de réflexion et d'absorption de quelques semiconducteurs sont donnés sur la figure (9-11) et le tableau (9-1).

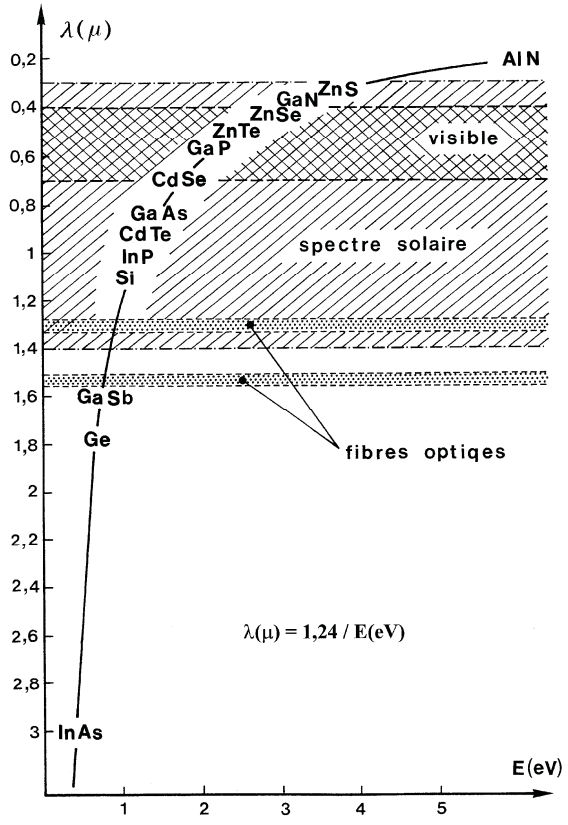


Figure 9-11 : Rayonnements et semiconducteurs

Tableau 9.1.a.: Coefficient de réflexion *R* de quelques semiconducteurs, dans le visible et le proche ultraviolet. (*Aspnes D.E. 1983*)

E(eV)	λ(μ)	Si	Ge	GaP	GaAs	GaSb	InP	InAs	InSb
1,6	0,78	0,33	0,43	0,28	0,33	0,41	0,31	0,34	0,42
1,8	0,69	0,34	0,45	0,28	0,34	0,44	0,31	0,35	0,47
2,0	0,62	0,35	0,50	0,29	0,35	0,49	0,32	0,37	0,44
2,2	0,56	0,36	0,52	0,30	0,36	0,46	0,33	0,39	0,45
2,4	0,52	0,38	0,51	0,31	0,38	0,47	0,34	0,43	0,46
2,6	0,48	0,40	0,48	0,33	0,41	0,47	0,36	0,44	0,43
2,8	0,44	0,43	0,46	0,35	0,46	0,45	0,39	0,45	0,42
3,0	0,41	0,46	0,46	0,37	0,47	0,44	0,43	0,41	0,42
3,2	0,39	0,52	0,48	0,39	0,47	0,46	0,46	0,39	0,43
3,4	0,37	0,59	0,50	0,43	0,43	0,48	0,42	0,38	0,46
3,6	0,34	0,56	0,51	0,50	0,42	0,50	0,39	0,37	0,49
3,8	0,33	0,57	0,53	0,50	0,41	0,53	0,38	0,37	0,54
4,0	0,31	0,59	0,56	0,45	0,42	0,58	0,38	0,39	0,61
4,2	0,30	0,65	0,61	0,44	0,44	0,66	0,39	0,44	0,63
4,4	0,28	0,73	0,71	0,45	0,49	0,67	0,42	0,53	0,61
4,6	0,27	0,74	0,70	0,48	0,54	0,64	0,49	0,59	0,59
4,8	0,26	0,71	0,68	0,54	0,60	0,61	0,58	0,62	0,56
5,0	0,25	0,68	0,65	0,58	0,67	0,59	0,61	0,58	0,54
5,2	0,24	0,65	0,62	0,65	0,66	0,57	0,60	0,56	0,54
5,4	0,23	0,66	0,60	0,69	0,63	0,58	0,55	0,54	0,56
5,6	0,22	0,68	0,60	0,66	0,60	0,60	0,53	0,50	0,56
5,8	0,21	0,67	0,63	0,62	0,57	0,60	0,50	0,46	0,56
6,0	0,21	0,68	0,65	0,58	0,55	0,61	0,46	0,45	0,57

Tableau 9.1.b.: Coefficient d'absorption α (10^3 cm^{-1}) de quelques semiconducteurs, dans le visible et le proche ultraviolet. (*Aspnes D.E. 1983*)

E(eV)	λ(μ)	Si	Ge	GaP	GaAs	GaSb	InP	InAs	InSb
1,6	0,78	1,25	55,9	0,00	14,8	67,5	35,3	75,1	121
1,8	0,69	2,38	91,2	0,00	27,4	111	49,3	96,7	254
2,0	0,62	4,47	189	0,00	42,7	279	64,3	128	359
2,2	0,56	7,17	456	0,04	61,4	389	84,6	183	412
2,4	0,52	14,7	597	0,86	90,3	477	111	312	556
2,6	0,48	23,8	608	2,94	142	600	152	496	565
2,8	0,44	46,2	619	33,4	281	612	223	626	577
3,0	0,41	81,7	652	68,2	592	641	379	618	606
3,2	0,39	204	758	109	742	716	695	613	678
3,4	0,37	932	888	195	715	837	709	616	816
3,6	0,34	1089	1006	499	717	981	683	630	987
3,8	0,33	1225	1150	986	735	1182	682	661	1234
4,0	0,31	1454	1352	880	778	1477	701	729	1497
4,2	0,30	1974	1706	878	880	1758	750	893	1461
4,4	0,28	2383	2082	932	1143	1618	868	1356	1353
4,6	0,27	2181	1846	1105	1477	1495	1229	1669	1294
4,8	0,26	1936	1707	1506	1834	1424	1711	1629	1253
5,0	0,25	1806	1620	1839	2069	1394	1771	1455	1237
5,2	0,24	1767	1567	2197	1836	1394	1589	1394	1288
5,4	0,23	1842	1562	2128	1685	1452	1451	1340	1310
5,6	0,22	1820	1615	1871	1597	1476	1410	1275	1291
5,8	0,21	1789	1689	1724	1543	1457	1340	1235	1299
6,0	0,21	1769	1686	1635	1503	1469	1285	1284	1300

9.2. Photodétecteurs

9.2.1. *Distribution des photoporteurs dans un semiconducteur photoexcité*

a. Equation de continuité

Lorsque le semiconducteur est photoexcité, les porteurs excédentaires créés diffusent et se recombinent. L'évolution de ce régime hors équilibre est régie par les équations de continuité (1-222-a et b), dans lesquelles les taux de génération g_n et g_p sont le taux de génération optique donné par l'expression (9-51). Pour simplifier l'écriture, on posera

$$g_n = g_p = \phi \alpha e^{-\alpha x} \tag{9-57}$$

où $\phi=(1-R)\phi_0$ représente le flux de photons dans le semiconducteur, immédiatement sous la surface. Les équations de continuité s'écrivent donc, à une dimension

$$\frac{\partial n}{\partial t} = n\mu_n \frac{\partial E}{\partial x} + \mu_n E \frac{\partial n}{\partial x} + D_n \frac{\partial^2 n}{\partial x^2} + \phi \alpha e^{-\alpha x} - \frac{n - n_o}{\tau_n} \tag{9-58-a}$$

$$\frac{\partial p}{\partial t} = -p\ \mu_p \frac{\partial E}{\partial x} - \mu_p E \frac{\partial p}{\partial x} + D_p \frac{\partial^2 p}{\partial x^2} + \phi \alpha e^{-\alpha x} - \frac{p - p_o}{\tau_p} \tag{9-58-b}$$

La composante du champ électrique dans la direction x, est la résultante de deux contributions, $E=E_a+E_i$. E_a représente le champ appliqué à l'échantillon par une éventuelle tension de polarisation. E_i représente un champ interne qui résulte de la différence de mobilité entre les électrons et les trous.

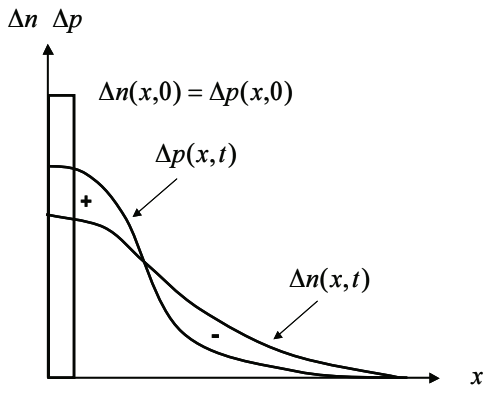


Figure 9-12 : Densités de porteurs après excitation

Si à l'instant t=0 où le semiconducteur est photoexcité, les densités de photoporteurs $\Delta n(x,0)$ et $\Delta p(x,0)$ sont égales, il n'en est plus de même à un instant $t \neq 0$ en raison du fait que les électrons diffusent plus rapidement que les trous (Fig. 9-12). Il en résulte l'existence d'une charge d'espace et par suite d'un champ électrique interne. Ce champ E_i est tel qu'il accélère les porteurs les plus lents et freine les porteurs les plus rapides.

Ainsi le système d'équations (9-58-a et b) a trois inconnues, n, p et E_i. Sa résolution passe donc par l'utilisation d'une troisième équation qui n'est autre que l'équation de Poisson, qui relie le potentiel donc le champ électrique, à la distribution de charges

$$-\Delta V = \frac{\rho}{\varepsilon} = \frac{dE}{dx} = \frac{e}{\varepsilon}\left(p + N_d^+ - n - N_a^-\right) \tag{9-59}$$

La résolution de ce système de trois équations est relativement complexe, mais nous avons vu (Par. 1.5.4), qu'en raison du rapport important qui existe entre la constante de temps diélectrique $\tau_{diel} \sim 10^{-12}$ s, et la durée de vie des porteurs $\tau \sim 10^{-9}$ à 10^{-3}s, la neutralité électrique du matériau est rétablie en un temps considérablement plus bref que son retour à l'équilibre. On peut donc supposer avec une bonne approximation que la neutralité électrique du semiconducteur est conservée sous éclairement, en d'autres termes $\Delta n(x,t) \approx \Delta p(x,t)$. Les électrons et les trous excédentaires créés par la photoexcitation diffusent ensemble, c'est la *diffusion ambipolaire*.

Notons d'autre part que, les taux de recombinaison des électrons et des trous étant nécessairement égaux , puisque les porteurs se recombinent par paires , si $\Delta n = \Delta p$ les durées de vies des porteurs sont égales $\tau_n = \tau_p = \tau$. Ainsi dans la mesure où $n = n_o + \Delta n$ et $p = p_o + \Delta$p , et où le champ appliqué E_a est à divergence nulle, les équations (9-58-a et b) s'écrivent

$$\frac{\partial \Delta n}{\partial t} = n\,\mu_n \frac{\partial E_i}{\partial x} + \mu_n (E_a + E_i)\frac{\partial \Delta n}{\partial x} + D_n \frac{\partial^2 \Delta n}{\partial x^2} + \phi\,\alpha\, e^{-\alpha x} - \frac{\Delta n}{\tau} \tag{9-60-a}$$

$$\frac{\partial \Delta n}{\partial t} = -\,p\;\mu_p \frac{\partial E_i}{\partial x} - \mu_p (E_a + E_i)\frac{\partial \Delta n}{\partial x} + D_p \frac{\partial^2 \Delta n}{\partial x^2} + \phi\,\alpha\, e^{-\alpha x} - \frac{\Delta n}{\tau} \tag{9-60-b}$$

Nous montrerons (Par. 9.2.1.c) que le terme $\mu_n E_i\, \partial \Delta n / \partial x$ est négligeable devant le terme $D_n\, \partial^2 \Delta n / \partial x^2$. Ainsi, en multipliant la première équation par $p\mu_p$, la deuxième par $n\mu_n$ et en effectuant la somme membre à membre, on obtient compte tenu de la relation d'Einstein $D/\mu = kT/e$,

$$\frac{\partial \Delta n}{\partial t} = \mu^* E_a \frac{\partial \Delta n}{\partial x} + D^* \frac{\partial^2 \Delta n}{\partial x^2} - \frac{\Delta n}{\tau} + \phi\alpha\, e^{-\alpha x} \tag{9-61}$$

avec

$$\mu^* = \frac{(p-n)\,\mu_n \mu_p}{p\mu_p + n\mu_n} \tag{9-62-a}$$

$$D^* = \frac{(p+n)\,D_n D_p}{pD_p + nD_n} \tag{9-62-b}$$

L'équation (9-61) est *l'équation de diffusion ambipolaire*, les coefficients μ^* et D^* sont appelés *mobilité* et *constante de diffusion ambipolaires*. Cette équation se simplifie dans la plupart des cas pratiques. Dans un matériau intrinsèque ou compensé $n_0 = p_0$ et par suite $n = p$ de sorte que $\mu^* = 0$ et $D^* = 2D_n D_p / (D_n + D_p)$, l'équation (9-61) s'écrit

$$\frac{\partial \Delta n}{\partial t} = D^* \frac{\partial^2 \Delta n}{\partial x^2} - \frac{\Delta n}{\tau} + \phi \alpha e^{-\alpha x} \tag{9-63}$$

Dans un matériau de type n avec $n_0 >> p_0$, en régime de faible excitation la condition $n>>p$ est vérifiée, de sorte que $n\mu_n >> p\mu_p$ et $nD_n >> pD_p$. Il en résulte que $\mu^* = \mu_p$ et $D^* = D_p$, l'équation s'écrit

$$\frac{\partial \Delta p}{\partial t} = -\mu_p E_a \frac{\partial \Delta p}{\partial x} + D_p \frac{\partial^2 \Delta p}{\partial x^2} - \frac{\Delta p}{\tau} + \phi \alpha e^{-\alpha x} \tag{9-64}$$

De même dans un matériau de type *p*

$$\frac{\partial \Delta n}{\partial t} = \mu_n E_a \frac{\partial \Delta n}{\partial x} + D_n \frac{\partial^2 \Delta n}{\partial x^2} - \frac{\Delta n}{\tau} + \phi \alpha e^{-\alpha x} \tag{9-65}$$

Dans chacun des cas l'équation régit la distribution des porteurs minoritaires. La densité de porteurs majoritaires n'est pas affectée par le rayonnement, sauf sous très forte excitation entraînant $\Delta n = \Delta p \sim n_o$ dans du matériau de type *n* ou $\Delta n = \Delta p \sim p_o$ dans du matériau de type *p*.

b. Distribution des photoporteurs

Considérons un échantillon semiconducteur d'épaisseur *d*, soumis à aucune tension extérieure (E_a=0), et éclairé uniformément sur l'une de ses faces par un rayonnement permanent d'énergie supérieure au gap (Fig. 9-13). Le régime étant stationnaire, dn/dt=0, l'équation de diffusion ambipolaire s'écrit

$$D\frac{d^2 \Delta n}{dx^2} - \frac{\Delta n}{\tau} + \phi \alpha e^{-\alpha x} = 0$$

où suivant la nature du matériau, D représente le coefficient de diffusion ambipolaire ou le coefficient de diffusion des porteurs minoritaires. En introduisant la longueur de diffusion des porteurs $L = \sqrt{D\tau}$, l'équation s'écrit

$$\frac{d^2\Delta n}{dx^2} - \frac{\Delta n}{L^2} = -\frac{\phi\alpha\tau}{L^2} e^{-\alpha x} \tag{9-66}$$

La solution de cette équation différentielle est la somme de la solution générale de l'équation sans second membre et d'une solution particulière de l'équation avec second membre, soit

$$\Delta n = A\, e^{-x/L} + B\, e^{x/L} + \frac{\phi\alpha\tau}{1-\alpha^2 L^2} e^{-\alpha x} \tag{9-67}$$

Les constantes A et B sont déterminées par les conditions aux limites, qui sont données par les vitesses de recombinaison en surface en x=0 et en x=d, soit

$$s_o = \frac{D}{\Delta n_o}\left(\frac{d\Delta n}{dx}\right)_o \qquad s_d = -\frac{D}{\Delta n_d}\left(\frac{d\Delta n}{dx}\right)_d \tag{9-68}$$

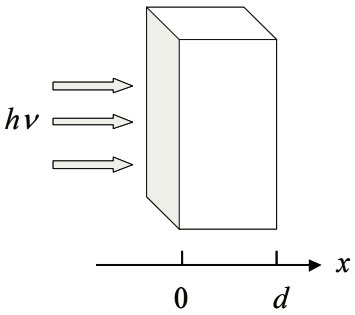


Figure 9-13 : Éclairement d'un semiconducteur

L'écriture de la solution générale est un peu lourde, considérons les deux cas limites de l'échantillon épais et de l'échantillon mince.

- E*chantillon épais, d >>L et 1/α*

Dans ces conditions, les porteurs excédentaires n'atteignent pas la face arrière de l'échantillon qui peut de ce fait être rejetée à l'infini. Ainsi quand $x \to \infty$, $\Delta n \to 0$ ce qui entraîne B=0. La constante A est alors déterminée par la vitesse de recombinaison sur la face frontale de l'échantillon s_o=s,

$$s\Delta n_o = D\left(\frac{d\Delta n}{dx}\right)_o$$

$$\Delta n_o = A + \frac{\phi\alpha\tau}{1-\alpha^2 L^2} \quad \left(\frac{d\Delta n}{dx}\right)_o = -\frac{A}{L} - \frac{\phi\alpha\tau}{1-\alpha^2 L^2}\alpha$$

ce qui donne

$$A = -\frac{\phi\alpha\tau}{1-\alpha^2 L^2}\frac{s/D+\alpha}{s/D+1/L} \tag{9-69}$$

L'expression (9-67) s'écrit

$$\Delta n = \frac{\phi\alpha\tau}{1-\alpha^2 L^2}\left(e^{-\alpha x} - \frac{s/D+\alpha}{s/D+1/L}e^{-x/L}\right) \tag{9-70}$$

La courbe de distribution est fonction des valeurs relatives de s/D, α et L. Notons que dans la pratique α et $1/L$ sont du même ordre de grandeur ($\alpha \approx 10^4$ cm^{-1} et $L \approx 1$ à 10 μm). Cette courbe est représentée sur la figure (9-14-a). En surface (x=0), la densité de photoporteurs et sa dérivée sont respectivement données par

$$\Delta n_o = \frac{\phi\alpha\tau}{1-\alpha^2 L^2}\left(1 - \frac{s/D+\alpha}{s/D+1/L}\right)$$

$$\left(\frac{d\Delta n}{dx}\right)_o = \frac{\phi\alpha\tau}{1-\alpha^2 L^2}\left(-\alpha + \frac{1}{L}\frac{s/D+\alpha}{s/D+1/L}\right)$$

Ainsi en surface, pour s=0 la dérivée est nulle et la densité de photoporteurs est maximum, pour s très grand la densité est nulle et sa dérivée est maximum.

- *Echantillon mince, $d \ll L$ et $1/\alpha$*

Supposons en outre que les vitesses de recombinaison en surface s_o et s_d sont négligeables. La condition $d \ll 1/\alpha$ entraîne $x < d \ll 1/\alpha$ et par suite $\alpha x \ll 1$ et $e^{-\alpha x} \approx 1$. Il en résulte que le deuxième membre de l'équation (9-66) se réduit à une constante et l'équation s'écrit

$$\frac{d^2\Delta n}{dx^2} - \frac{\Delta n}{L^2} = -\frac{\phi\alpha\tau}{L^2} \tag{9-71}$$

La solution de l'équation sans second membre est inchangée, la solution de l'équation avec second membre s'écrit $\Delta n = C$ avec $C = \phi\alpha\tau$, d'où la solution générale

$$\Delta n = A\,e^{-x/L} + B\,e^{x/L} + \phi\alpha\tau \tag{9-72}$$

Avec $s_o=s_d=0$ les conditions (9-68) s'écrivent simplement $d\Delta n/dx=0$ en $x=0$ et en $x=d$, ce qui donne compte tenu de (9-72), les relations

$$A=B \quad \text{et} \quad A=B\, e^{2d/L}$$

Ces deux relations ne sont compatibles que si $A=B=0$, l'expression (9-72) s'écrit donc

$$\Delta n = \phi\alpha\tau \tag{9-73}$$

Δn est constant dans tout l'échantillon (Fig. 9-14-b). Si les vitesses de recombinaison en surface ne sont pas négligeables, la densité de photoporteurs décroît au voisinage des surfaces.

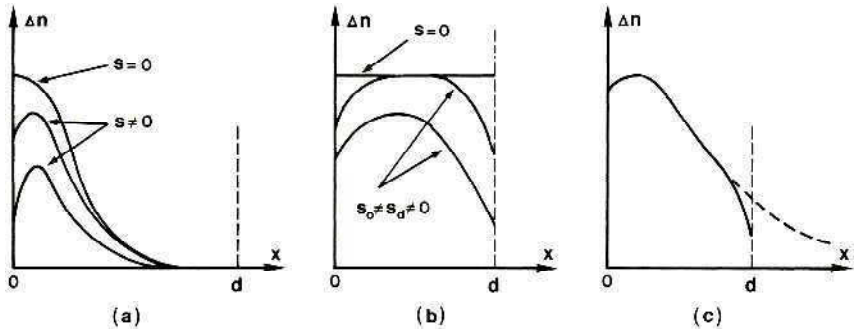


Figure 9-14 : Courbe de distribution des photoporteurs dans un semiconducteur. a) Echantillon épais. b) Echantillon mince. c) Echantillon quelconque

c. Champ interne

Le champ interne E_i résulte de la différence des longueurs de diffusion des électrons et des trous (Fig. 9-12). On peut donc, pour estimer l'ordre de grandeur de ce champ, s'affranchir des paramètres tels que la vitesse de recombinaison en surface s, le coefficient d'absorption α ou le champ appliqué E_a qui, ici, ne jouent pas un rôle fondamental. Considérons donc un échantillon épais, non polarisé, dont la vitesse de recombinaison en surface est négligeable et dont le coefficient d'absorption est très important devant $1/L_i$ (i=*n*,*p*). Dans le modèle de la diffusion ambipolaire, la densité de photoporteurs est donnée par l'expression (9-70) qui s'écrit, compte tenu des conditions simplificatrices précédentes.

$$\Delta n = \Delta p = \frac{\phi\tau}{L} e^{-x/L}$$

où L représente la longueur de diffusion ambipolaire

Dans la limite ambipolaire, $\Delta n(x)$ et $\Delta p(x)$ sont rigoureusement égaux et le champ électrique interne E_i est nul. A l'inverse, on obtiendra une limite supérieure de E_i en supposant que les porteurs diffusent indépendamment les uns des autres. Dans ce cas limite, purement théorique, les distributions des électrons et des trous, représentées schématiquement sur la figure (9-12), sont données par

$$\Delta n = \frac{\phi\tau}{L_n} e^{-x/L_n} \tag{9-73-a}$$

$$\Delta p = \frac{\phi\tau}{L_p} e^{-x/L_p} \tag{9-73-b}$$

Le champ électrique E_i qui résulte de cette double distribution de charges est donné par le théorème de Gauss

$$\frac{dE_i}{dx} = \frac{\rho(x)}{\varepsilon} = \frac{e}{\varepsilon}\left(\Delta p(x) - \Delta n(x)\right)$$

soit

$$\frac{dE_i}{dx} = \frac{\phi\tau e}{\varepsilon}\left(\frac{1}{L_p} e^{-x/L_p} - \frac{1}{L_n} e^{-x/L_n}\right) \tag{9-73-c}$$

On obtient l'expression de E_i en intégrant l'équation précédente avec la condition $E_i=0$ quand $x\to\infty$

$$E_i = \frac{\phi\tau e}{\varepsilon}\left(e^{-x/L_n} - e^{-x/L_p}\right) \tag{9-73-d}$$

Le champ E_i est nul en $x=0$ et pour $x\to\infty$, il passe par un maximum au point où $dE_i/dx = 0$, c'est-à-dire au point où $\Delta n(x) = \Delta p(x)$, soit

$$x_m = \frac{L_n L_p}{L_n - L_p} Ln \frac{L_n}{L_p} \tag{9-73-e}$$

On obtient la valeur maximum du champ E_i en portant l'expression de x_m dans l'expression (9-73-d), soit

$$E_{im} = \frac{\phi\tau e}{\varepsilon}\left(\left(\frac{L_n}{L_p}\right)^{-L_p/(L_n-L_p)} - \left(\frac{L_n}{L_p}\right)^{-L_n/(L_n-L_p)}\right) \tag{9-73-f}$$

Dans le cas particulier défini par $L_n=L_p$ on obtient évidemment $E_i(x)=0$ quel que soit x et par suite $E_{im}=0$. Les porteurs diffusent ensemble et la charge d'espace est nulle. La diffusion est, de fait, rigoureusement ambipolaire avec $L=L_n=L_p$.

En fait les longueurs de diffusion des électrons et des trous sont généralement différentes, de sorte que, dans l'hypothèse limite de diffusions indépendantes, le champ interne $E_i(x)$ est différent de zéro et son maximum est donné par l'expression (9-73-f). Ce champ maximum E_{im} est d'autant plus important que le rapport L_n/L_p est grand. Dans les conditions extrêmes définies par $L_n>>L_p$, l'expression (9-73-f) se simplifie et s'écrit

$$E_{im} \approx \phi\tau e/\varepsilon \qquad \text{(9-73-g)}$$

Cette expression de E_{im} constitue rappelons-le, la valeur maximum du champ interne E_i , calculée dans les conditions extrêmes ($L_n>>L_p$) d'un cas limite (diffusions indépendantes des porteurs). En fait, en raison du regroupement des porteurs résultant de leurs attractions coulombiennes, le champ réel E_i est, de plusieurs ordres de grandeurs, inférieur à cette valeur

$$E_i << \frac{\phi\tau e}{\varepsilon} \qquad \text{(9-73-h)}$$

Afin de calculer l'ordre de grandeur de cette limite supérieure, attribuons aux différents paramètres des valeurs typiques. La durée de vie des porteurs minoritaires dans un semiconducteur à gap direct est de l'ordre de 10 ns. La constante diélectrique est de l'ordre de 12 ε_0. Supposons le semiconducteur exposé au rayonnement solaire, dont l'intensité I est de l'ordre de 1 kW/m^2 et dont le spectre est centré au voisinage de λ=0,5 µ. Le flux de photons incident $\phi_0=I/h\upsilon$ est de l'ordre de 6.10^{17} photons/s/cm^2, compte tenu de la valeur élevée de la constante diélectrique le coefficient de réflexion du matériau est voisin de 30%, de sorte que le flux de photons entrant dans le semiconducteur est de l'ordre de $\phi\approx 4.10^{17}$ photons/s/cm^2. A partir de ces différentes valeurs, l'expression (9-73-h) donne $E_i<<5$ V/cm.

Comparons maintenant les ordres de grandeur respectifs des termes $\mu_n E_i \partial\Delta n/\partial x$ et $D_n \partial^2\Delta n/\partial x^2$ qui apparaissent dans l'équation différentielle (9-60-a). En explicitant les dérivées successives de Δn à partir de l'expression (9-73-a), les amplitudes de ces termes s'écrivent

$$\left|\mu_n E_i \frac{\partial\Delta n}{\partial x}\right| = \mu_n E_i \frac{\Delta n}{L_n} \qquad \left|D_n \frac{\partial^2\Delta n}{\partial x^2}\right| = \frac{D_n}{L_n}\frac{\Delta n}{L_n}$$

La comparaison de ces termes se ramène donc à celle de $\mu_n E_i$ et D_n/L_n ou, compte tenu de la relation d'Einstein, $D/\mu=kT/e$, à la comparaison des termes E_i et kT/eL_n. A la température ambiante kT/e=26 mV, la longueur de diffusion des porteurs minoritaires est de l'ordre du micron, soit $kT/eL_n\approx$250 V/cm. Ainsi dans la mesure où E_i est très inférieur à E_{im} , la condition $E_i<<kT/eL_n$ est amplement vérifiée, ce qui justifie l'approximation faite dans la résolution du système d'équations (9-60-a et b).

9.2.2. *Cellule photoconductrice*

Une cellule photoconductrice exploite l'augmentation de conductivité électrique du semiconducteur, résultant de la création de porteurs sous éclairement (Fig. 9-15). Une source de tension débite un courant I dans le semiconducteur par l'intermédiaire de deux contacts ohmiques. La variation du nombre de porteurs $\Delta n = \Delta p$, entraîne une augmentation de conductivité du matériau, donc de la conductance du barreau, et par suite du courant I et de la tension V_s.

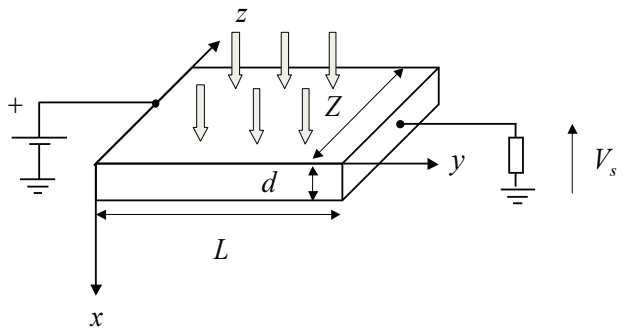


Figure 9-15 : Principe de la cellule photoconductrice

a. Variation de la conductance

En l'absence d'éclairement la conductivité du matériau s'écrit

$$\sigma_o = n e \mu_n + p e \mu_p$$

Sous éclairement, cette conductivité devient $\sigma = \sigma_0 + \Delta\sigma$ avec

$$\Delta\sigma = e(\mu_n \Delta n + \mu_p \Delta p) = e\mu_n (1 + \mu_p / \mu_n) \Delta n$$

Dans la mesure où la mobilité des électrons est très supérieure à celle des trous la variation de conductivité est donnée par

$$\Delta\sigma = e \mu_n \Delta n \tag{9-74}$$

La variation de la conductance du barreau s'écrit

$$\Delta g = \int_V \Delta\sigma \frac{ds}{dy} = \int_V \Delta\sigma \frac{dxdz}{dy} \tag{9-75}$$

Dans la mesure où l'éclairement est uniforme dans le plan yz, $\Delta\sigma(x,y,z)$ s'écrit $\Delta\sigma(x)$ et l'intégrale de volume se ramène à une intégrale sur x

$$\Delta g = \frac{Z}{\ell}\int_0^d \Delta\sigma(x)dx = \frac{Z\mu_n e}{\ell}\int_0^d \Delta n(x)dx \qquad (9\text{-}76)$$

La variation de la densité de photoporteurs est donnée par l'expression générale (9-67), qui se simplifie dans les cas limites de la cellule épaisse ou de la cellule mince.

- *Cellule épaisse, $d >> L$ et $1/\alpha$*

La densité de photoporteurs est donnée par l'expression (9-70), qui donne en portant dans l'expression (9-76)

$$\Delta g = \frac{Z\mu_n e}{\ell}\frac{\phi\alpha\tau}{1-\alpha^2 L^2}\left(\int_0^d e^{-\alpha x}dx - \frac{s/D+\alpha}{s/D+1/L}\int_0^d e^{-x/L}dx\right) \qquad (9\text{-}77)$$

Les conditions $d >> L$ et $1/\alpha$ entraînent $\alpha d >> 1$ et $d/L >> 1$, le calcul de l'expression précédente donne

$$\Delta g = \frac{Z\mu_n e}{\ell}\frac{\phi\tau}{1-\alpha^2 L^2}\left(1 - \frac{s/D+\alpha}{s/D+1/L}\alpha L\right) \qquad (9\text{-}78)$$

Si la vitesse de recombinaison en surface est négligeable l'expression (9-78) se simplifie

$$\Delta g = \frac{Z\mu_n e}{\ell}\phi\tau \qquad (9\text{-}79)$$

- *Cellule mince $d << L$ et $1/\alpha$*

Si d'autre part la vitesse de recombinaison en surface peut être négligée, Δn est donné par l'expression (9-70). La densité de photoporteurs est constante dans tout le volume de la cellule, l'expression (9-76) donne

$$\Delta g = \frac{Z\mu_n e}{\ell}\phi\tau\,\alpha d \qquad (9\text{-}80)$$

Dans cette expression αd est inférieur à 1, de sorte que la comparaison des expressions (9-78 et 80) montre que la variation absolue de photoconductance est plus importante dans la cellule épaisse que dans la cellule mince. Cependant la conductance au repos est aussi plus importante dans la cellule épaisse que dans la cellule mince, de sorte qu'un choix judicieux de l'épaisseur permet d'optimiser la variation relative de conductance $\Delta g/g$. Cette variation relative s'écrit

Cellule épaisse: $\frac{\Delta g}{g} = \frac{\phi \tau}{n} \frac{1}{d}$ Cellule mince: $\frac{\Delta g}{g} = \frac{\phi \tau}{n} \alpha$

Ces expressions montrent que dans la cellule épaisse, qui absorbe la totalité du rayonnement, $\Delta g/g$ augmente comme $1/d$. Mais lorsque $1/d$ devient supérieur à α la cellule devient mince et $\Delta g/g$ sature à la valeur de la cellule mince. Pour optimiser $\Delta g/g$ on utilisera donc un matériau dans lequel τ et α sont grands et n petit, et on réalisera d de l'ordre de 2 à 3 fois $1/\alpha$ afin d'absorber la totalité du rayonnement.

b. Gain de la cellule

Le paramètre caractéristique d'une cellule photoconductrice est son gain, défini comme le rapport du *nombre de charges débitées* par seconde, au *nombre de photons absorbés* par seconde.

Si la variation de conductance est Δg le photocourant est $\Delta I = \Delta g V$, où V est la tension appliquée. Le nombre de charges débitées par seconde est par conséquent $\Delta I/e = \Delta g V/e$.

D'autre part si le flux de photons incidents est ϕ_o le flux de photons entrant dans la cellule est $\phi=(1-R)\phi_o$. Si la cellule est épaisse tous les photons sont absorbés de sorte que le nombre de photons absorbés par seconde est égal à $\phi Z \ell$, où $Z\ell$ représente la surface sensible de la cellule. Le gain de la cellule s'écrit donc

$$G=\frac{\Delta g V}{e\phi Z \ell} \tag{9-81}$$

Si la vitesse de recombinaison en surface est négligeable Δg est donné par l'expression (9-79), ce qui donne en portant dans l'expression (9-81)

$$G=\frac{\mu_n V}{\ell^2}\tau=\frac{\mu_n V/\ell}{\ell}\tau=\frac{\mu_n E}{\ell}\tau=\frac{v_n}{\ell}\tau=\frac{\tau}{\ell/v_n}$$

En posant $\tau_t=\ell/v_n$ où τ_t représente le temps de transit des électrons dans la cellule, le gain s 'écrit

$$G=\tau/\tau_t \tag{9-82}$$

Le gain de la cellule est directement donné par le rapport de la durée de vie des photoporteurs, à leur temps de transit à travers la cellule. Si la cellule est mince l'expression précédente est multipliée par le produit αd.

c. Réponse de la cellule

Le paramètre qui mesure *in fine* les performances d'une cellule photoconductrice est sa *réponse*, définie comme le rapport de la variation du courant débité à la puissance incidente, $\Re = \Delta I/P_i$.

Si le rayonnement incident est monochromatique, la puissance incidente s'écrit $P_i = \phi_0 Z \ell h\nu$, où le flux incident ϕ_0 est relié au flux ϕ entrant dans la cellule, par $\phi = (1-R)\phi_0$. D'autre part la variation du courant débité est donnée par $\Delta I = \Delta g\ V$. En explicitant Δg et en faisant apparaître le gain G, la réponse s'écrit

$$\Re = (1-R)G\frac{e}{h\nu}$$

Si on exprime $h\nu$ en eV, ou λ en μm, la réponse en A/W est donnée par

$$\Re(A/W) = (1-R)\,G/h\nu(eV) = (1-R)\,G\,\lambda(\mu m)/1{,}24 \qquad (9\text{-}83)$$

d. Structure d'une cellule photoconductrice

La géométrie type d'une cellule photoconductrice est représentée sur la figure (9-16). La partie active de la cellule est constituée d'une couche mince de semiconducteur déposée sur un substrat isolant. Les électrodes métalliques sont réalisées par évaporation d'un métal tel que l'or, à travers un masque protégeant une fraction de la surface qui deviendra la surface sensible de la cellule. La géométrie de la cellule permet de réaliser Z grand et ℓ petit, afin d'augmenter Δg (Eq. 9-80).
Les semiconducteurs couramment utilisés pour réaliser des cellules photoconductrices sont les composés binaires CdS et CdSe dans le spectre visible, le sulfure de plomb PbS dans le proche infrarouge, et les composés ternaires $Hg_xCd_{1-x}Te$ dans la gamme 5-15 μm.

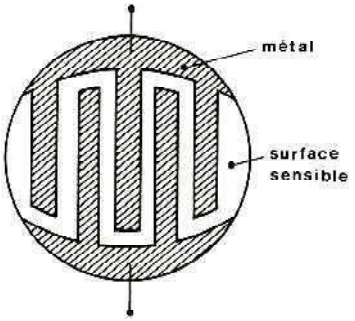


Figure 9-16 : Géométrie d'une cellule

9.2.3. *Cellule photovoltaïque - Photodiode*

a. Principe de fonctionnement

Le courant inverse d'une jonction *pn* est fonction d'une part des densités de porteurs minoritaires dans les régions neutres de la diode, et d'autre part de la génération de paires électron-trou dans la zone de charge d'espace. Dans une photodiode, le rayonnement augmente le courant inverse par la création de porteurs minoritaires dans les régions neutres et la génération de paires électrons-trous dans la zone de charge d'espace.

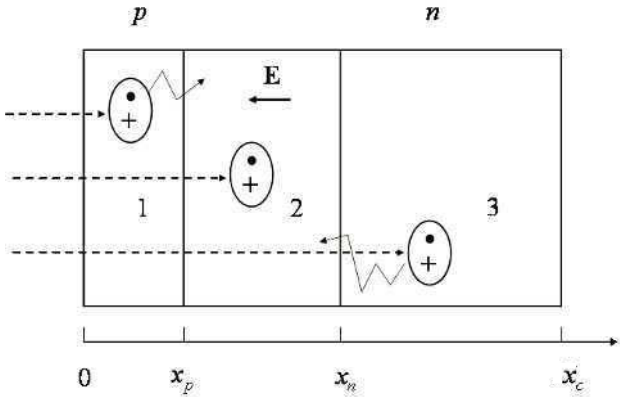


Figure 9-17 : Principe de la photodiode

Le principe de fonctionnement d'une cellule photovoltaïque est illustré sur la figure (9-17). Les photons incidents créent des porteurs dans chacune des régions 1, 2 et 3. Le comportement de ces porteurs libres diffère suivant le lieu de leur création. Dans les régions électriquement neutres *p* et *n*, les photoporteurs minoritaires diffusent, ceux qui atteignent la zone de charge d'espace sont propulsés par le champ électrique vers la région où ils deviennent majoritaires. Ces photoporteurs contribuent donc au courant par leur diffusion, ils créent un *photocourant de diffusion*. Dans la zone de charge d'espace, les paires électrons-trous créées par les photons sont dissociées par le champ électrique, l'électron est propulsé vers la région de type *n* et le trou vers la région de type *p*. Ces porteurs donnent naissance à un *photocourant de génération*. Ces différentes contributions s'ajoutent pour créer un photocourant résultant I_{ph} qui contribue au courant inverse de la diode

$$I = I_s\left(e^{eV/kT} - 1\right) - I_{ph} \tag{9-84}$$

La caractéristique de la photodiode est représentée sur la figure (9-18). Le photocourant est pratiquement indépendant de la tension de polarisation. Dans la pratique, on mesure soit le photocourant débité par la diode, soit le photovoltage qui apparaît aux bornes de la diode.

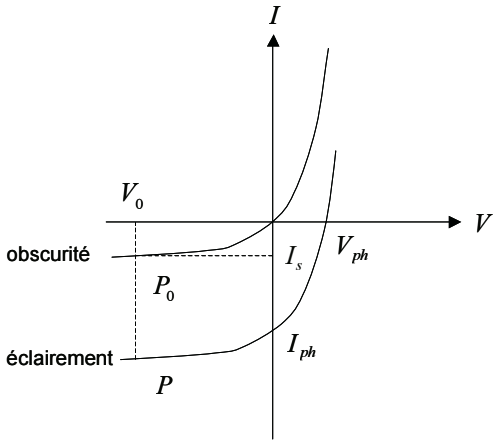


Figure 9-18 : Courant d'une photodiode

Dans le premier cas, la diode est polarisée en inverse par une tension négative V_o. Dans la mesure où $-V_o >> kT/e$, l'expression (9-84) s'écrit

$$I = -(I_s + I_{ph}) \tag{9-85}$$

Dans la pratique I_s est très inférieur à I_{ph} de sorte que le courant mesuré est égal au photocourant et par suite proportionnel au rayonnement incident.

Dans le mode photovoltaïque la diode est connectée aux bornes d'un voltmètre, le courant est alors nul et $V=V_{\text{ph}}$ le photovoltage. L'expression (9-84) donne alors

$$V_{ph} = \frac{kT}{e} Ln\left(I_{ph} / I_s + 1\right) \tag{9-86}$$

Le photovoltage varie donc logarithmiquement avec le photocourant, et par voie de conséquence avec l'intensité du rayonnement.

b. Calcul du photocourant

Le photocourant résultant est la somme de trois composantes, le courant de diffusion des photoélectrons de la région de type *p*, le courant de photogénération dans la zone de charge d'espace et le courant de diffusion des phototrous de la région de type *n*. On obtient le photocourant total en ajoutant ces trois composantes calculées en un même point, au point x_n par exemple (Fig. 9-19), soit

$$J_{ph} = j_{ndiff}\left(x=x_n\right) + j_g\left(x=x_n\right) + j_{pdiff}\left(x=x_n\right) \tag{9-87}$$

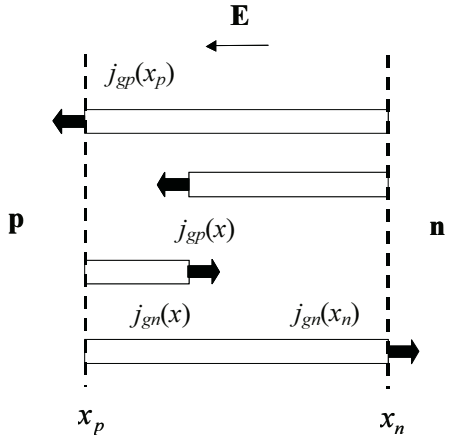


Figure 9-19 : Les courants dans une photodiode

Considérons tout d'abord le courant de diffusion d'électrons. On ne peut le calculer que dans la région *p*, mais si on néglige les recombinaisons dans la zone de charge d'espace tous les électrons qui arrivent en x_p se retrouvent en x_n. On peut donc écrire

$$j_{ndiff}(x_n) = j_{ndiff}(x_p) \tag{9-88}$$

Considérons le courant de génération dans la zone de charge d'espace (Fig. 9-19) .

- En un point quelconque x de cette zone le courant résulte à la fois des photoélectrons créés entre x_p et x et des phototrous créés entre x et x_n. Il s'écrit $j_g(x) = j_{gn}(x) + j_{gp}(x)$

- En $x=x_p$ ce courant n'est dû qu'aux trous, qui arrivent en x_p et qui ont été créés tout au long de la zone de charge d'espace, depuis x_p jusqu'à x_n, soit $j_g(x_p) = j_{gp}(x_p)$.

- En $x=x_n$, le courant de génération est dû uniquement aux électrons, qui ont été créés depuis x_p jusqu'à x_n, soit

$$j_g(x_n) = j_{gn}(x_n) \tag{9-89}$$

Compte tenu des expressions (9-88 et 89), l'expression (9-87) s'écrit sous la forme

$$J_{ph} = j_{ndiff}(x_p) + j_{gn}(x_n) + j_{pdiff}(x_n) \tag{9-90}$$

Nous allons calculer ces différentes composantes du photocourant.

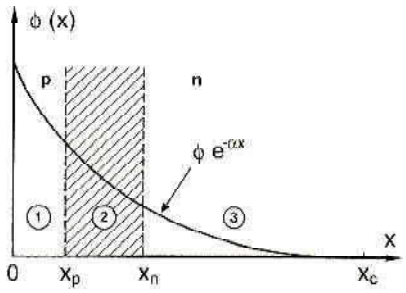


Figure 9-20 : Répartition du flux

- Courant de génération, région 2

En régime stationnaire ($\partial n/\partial t$=0) et en négligeant les recombinaisons dans la zone de charge d'espace (r_n=0), l'équation de continuité (Eq. 1-220-a) s'écrit

$$0=\frac{1}{e}\frac{dj_n}{dx}+g \qquad (9\text{-}91)$$

Le courant de génération d'électrons en $x=x_n$ est par conséquent donné par

$$j_{gn}(x_n)-j_{gn}(x_p)=j_{gn}(x_n)=-e\int_{x_p}^{x_n} g\,dx \qquad (9\text{-}92)$$

Le taux de génération de photoporteurs est donné par l'expression (9-51) qui s'écrit $g=\phi\alpha e^{-\alpha x}$ où ϕ=(1-R)ϕ_0 . En portant cette expression de g dans l'équation (9-92) et en intégrant, on obtient

$$j_{gn}(x_n)=-e\phi(e^{-\alpha x_p}-e^{-\alpha x_n}) \qquad (9\text{-}93)$$

ou

$$j_{gn}(x_n)=-e\phi e^{-\alpha x_p}(1-e^{-\alpha w}) \qquad (9\text{-}94)$$

où $w=x_n-x_p$ est la largeur de la zone de charge d'espace. Le courant est négatif en raison de l'orientation de l'axe x sur la figure (9-20).

- *Courant de diffusion de trous en $x=x_n$, région 3*

Le courant de trous dans la région 3 de la structure est un courant de diffusion

$$j_{pdiff}=-e\,D_p\frac{d\,\Delta p}{dx} \qquad (9\text{-}95)$$

La distribution des photoporteurs dans cette région est donnée par l'expression générale (9-67). Les constantes A et B sont déterminées par les conditions aux limites, qui sont ici les densités de phototrous en $x=x_n$ et en $x=x_c$. En $x=x_n$, Δp=0 car les phototrous qui atteignent ce point sont propulsés par le champ électrique à travers la zone de charge d'espace. La durée de vie des trous en ce point est pratiquement nulle, non pas parce qu'ils se recombinent mais parce qu'ils sont évacués. De même en $x=x_c$ au contact ohmique, la perturbation du cristal est telle que la durée de vie des trous est nulle, $\Delta p(x=x_c)$=0. Pour simplifier l'expression de Δp on peut toutefois supposer que la région arrière de la jonction (région 3) est

beaucoup plus épaisse que la longueur de diffusion des trous et que $1/\alpha$. On peut alors rejeter x_c à l'infini et la deuxième condition aux limites se ramène à $\Delta p=0$ pour $x\rightarrow\infty$ ce qui entraîne $B=0$. On peut alors calculer simplement la constante A à l'aide de la première condition qui s'écrit

$$0 = A\,e^{-x_n/L_p} + \frac{\phi\alpha\tau_p}{1-\alpha^2 L_p^2} e^{-\alpha x_n} \tag{9-96}$$

soit

$$A = -\frac{\phi\alpha\tau_p}{1-\alpha^2 L_p^2} e^{-\alpha x_n + x_n/L_p} \tag{9-97}$$

La densité de phototrous s'écrit donc

$$\Delta p = \frac{\phi\alpha\tau_p}{1-\alpha^2 L_p^2} e^{-\alpha x_n} \left(e^{-\alpha(x-x_n)} - e^{-(x-x_n)/L_p} \right) \tag{9-98}$$

En portant cette expression dans l'équation (9-95) on obtient

$$j_{pdiff} = -eD_p \frac{\phi\alpha\tau_p}{1-\alpha^2 L_p^2} e^{-\alpha x_n} \left(-\alpha e^{-\alpha(x-x_n)} + \frac{1}{L_p} e^{-(x-x_n)/L_p} \right) \tag{9-99}$$

en $x=x_n$

$$j_{pdiff}(x_n) = -e\,\phi \frac{\alpha L_p}{1+\alpha L_p} e^{-\alpha x_n} \tag{9-100}$$

- *Courant de diffusion d'électrons en $x=x_p$, région 1*

La distribution de photoélectrons dans la région de type *p* est de la même forme que la distribution de phototrous dans la région de type *n*

$$\Delta n = A\,e^{-x/L_n} + B\,e^{x/L_n} + \frac{\phi\alpha\tau_n}{1-\alpha^2 L_n^2} e^{-\alpha x} \tag{9-101}$$

Les conditions aux limites, permettant de calculer les constantes A et B, sont ici définies en $x=x_p$ et en $x=0$ à la surface de l'échantillon. En $x=x_p$, $\Delta n(x=x_p)=0$, car les électrons qui atteignent ce point sont propulsés à travers la zone de charge d'espace. En $x=0$, la densité de photoélectrons est conditionnée par la vitesse de recombinaison en surface (Eq. 9-41). On peut donc déterminer les constantes A et B. Le calcul ne présente aucune difficulté mais les expressions de ces constantes sont assez lourdes, nous les appellerons simplement A_n et B_n. On obtient alors le courant de diffusion d'électrons en $x=x_p$ en dérivant l'expression (9-101), et en faisant $x=x_p$, soit

$$j_{ndiff}(x_p)=eD_n\left(-\frac{A_n}{L_n}e^{-x_p/L_n}+\frac{B_n}{L_n}e^{x_p/L_n}-\frac{\phi\alpha^2\tau_n}{1-\alpha^2L_n^2}e^{-\alpha x_p}\right) \qquad (9\text{-}102)$$

On obtient le photocourant j_{ph} en portant les expressions (9-94, 100 et 102) dans l'expression (9-90). On simplifie considérablement l'expression de J_{ph} en se plaçant dans le cas pratique où la région frontale (région 1) de la photodiode, est d'épaisseur beaucoup plus faible que $1/\alpha$. Cette condition doit être remplie en pratique si l'on veut que le nombre de photoporteurs créés dans la zone de charge d'espace soit important. Dans ce cas, le courant de diffusion d'électrons est négligeable et d'autre part on peut faire les approximations $x_p \approx 0$ et $x_n \approx w$ dans les expressions des deux autres courants, qui s'écrivent alors

$$j_g(x_n)\approx -e\phi(1-e^{-\alpha w}) \qquad (9\text{-}103)$$

$$j_{pdiff}(x_n)\approx -e\phi\frac{\alpha L_p}{1+\alpha L_p}e^{-\alpha w} \qquad (9\text{-}104)$$

Le photocourant résultant s'écrit par conséquent

$$j_{ph}\approx -e\phi\left(1-\frac{1}{1+\alpha L_p}e^{-\alpha w}\right) \qquad (9\text{-}105)$$

Le signe - montre tout simplement, compte tenu de l'orientation de l'axe x sur la figure (9-20), que le photocourant à travers la jonction est dirigé de la région de type n vers la région de type p, c'est-à-dire que c'est un courant inverse. Pour obtenir un photocourant important on a intérêt à réaliser la condition $\alpha w >> 1$, dans ce cas le photocourant est maximum et simplement donné par

$$j_{ph}\approx -e\phi \qquad (9\text{-}106)$$

Cette expression traduit tous simplement le fait que dans ces conditions optima le flux d'électrons débités par la photodiode j_{ph}/e , est égal au flux ϕ de photons d'énergie supérieure au gap du semiconducteur, qui pénètrent dans le détecteur. Le rendement de la photodiode est maximum.

c. Constante de temps de la photodiode

Compte tenu de la double origine du photocourant, le temps de réponse de la photodiode est conditionné d'une part par la diffusion des photoporteurs minoritaires des régions neutres vers la zone de charge d'espace, et d'autre part par le temps de transit des porteurs à travers la zone de charge d'espace. Le premier phénomène est relativement lent et se traduit par une constante de temps de l'ordre de 10^{-8} à 10^{-9} s, le second phénomène peut être très rapide si la tension de polarisation inverse de la diode est importante. Dans ce cas, les porteurs traversent

la zone de charge d'espace avec leur vitesse de saturation v_s et le temps de transit est donné par w/v_s de l'ordre de 10^{-11} à 10^{-10} s.

Pour diminuer au maximum la constante de temps de la photodiode, on a donc intérêt à ce que le rayonnement soit essentiellement absorbé dans la zone de charge d'espace de la jonction, où il crée le courant de génération. On réalisera donc une photodiode avec une zone frontale aussi mince que possible et une zone de charge d'espace w suffisamment épaisse pour absorber la majeure partie du rayonnement. On limitera toutefois w à une valeur qui n'augmente pas trop le rapport w/v_s. On choisira $w \sim 1/\alpha$.

d. Différents types de cellules photovoltaïques

- *Photodiodes pin*

Nous venons de voir l'intérêt qu'il y a à donner à la zone de charge d'espace de la diode une valeur suffisante pour que le photocourant soit essentiellement dû à la photogénération de porteurs dans cette zone. On augmente artificiellement la valeur de w en intercalant une région intrinsèque entre les régions de type n et de type p (Fig. 9-21). Si la polarisation inverse de la structure est suffisante, un champ électrique important existe dans toute la zone intrinsèque et les photoporteurs atteignent très vite leur vitesse limite v_s. On obtient ainsi des photodiodes rapides et très sensibles. La structure *pin* est la photodiode la plus commune.

- *Photodiode à avalanche*

Lorsque la polarisation inverse de la photodiode est voisine de la tension de claquage, les photoporteurs créés dans la zone de charge d'espace sont multipliés par effet d'avalanche. On obtient ainsi une multiplication interne du photocourant, la photodiode est alors l'analogue solide du photomultiplicateur. Le gain ainsi obtenu est facilement supérieur à 100, toutefois certaines précautions doivent être prises dans le circuit de polarisation car ce gain est très sensible à la tension de polarisation et à la température. Les photodiodes à avalanche sont utilisées dans les systèmes de télécommunication par fibres optiques.

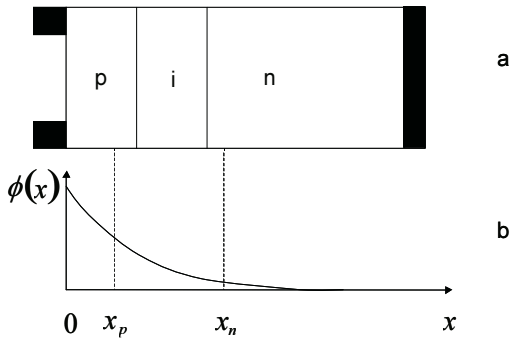


Figure 9-21
Photodiode pin

- *Phototransistor*

Un phototransistor est un transistor bipolaire dont la base est accessible au rayonnement à détecter. Cette base est flottante de sorte que le courant de base est nul, et s'écrit en l'absence d'éclairement

$$I_b = -I_e - I_c = -(1-\alpha) I_e + I_{co} \equiv 0 \qquad (9\text{-}107)$$

où I_{co} représente le courant de saturation de la jonction collecteur-base. Lorsque la base est soumise à un rayonnement, les photoporteurs créent un courant de génération et le courant de base s'écrit

$$I_b = -(1-\alpha) I_e + I_{co} + I_{ph} \equiv 0 \qquad (9\text{-}108)$$

Ainsi le courant d'émetteur, égal au courant de collecteur, s'écrit

$$I_e = (1+\beta)\ (I_{co} + I_{ph}) \approx \beta I_{ph} \qquad (9\text{-}109)$$

Le photocourant est ainsi multiplié par le gain β du transistor. Ce dernier est cependant faible car la base doit être assez large pour absorber un maximum de photons. La sensibilité du phototransistor ~6 A/W, se situe entre celles de la photodiode *pin* ~0,5 A/W, et de la photodiode à avalanche ~60 A/W. Son principal inconvénient est inhérent à la nature du photocourant, qui résulte de la diffusion des photoporteurs à travers la base car la constante de temps est très importante, $\sim 10^{-5}$ s.

- *Photodiode Schottky*

Une photodiode Schottky est constituée d'un substrat de silicium de type *n*, sur lequel est déposée une couche mince métallique, généralement de l'or. On réalise ainsi une barrière de Schottky. Lorsque le rayonnement crée des paires électron-trou dans la zone de charge d'espace du semiconducteur, la diode est le siège d'un photocourant de génération analogue à celui de la photodiode à jonction *pn*. L'avantage de la photodiode Schottky réside dans le fait que la couche métallique, si elle est suffisamment mince, est transparente au rayonnement dans le domaine du proche ultraviolet.

9.2.4. *Cellule solaire - photopile*

La photopile n'est autre qu'une photodiode qui fonctionne sans polarisation extérieure et débite son photocourant dans une charge. Sous éclairement la caractéristique $I(V)$ de la diode ne passe plus par l'origine des coordonnées, il existe une région dans laquelle le produit VI est négatif (Fig. 9-18), la diode fournit de

l'énergie. Si on se limite à cette région active et si on compte positivement le courant inverse, la figure (9-18) se ramène à la figure (9-22).

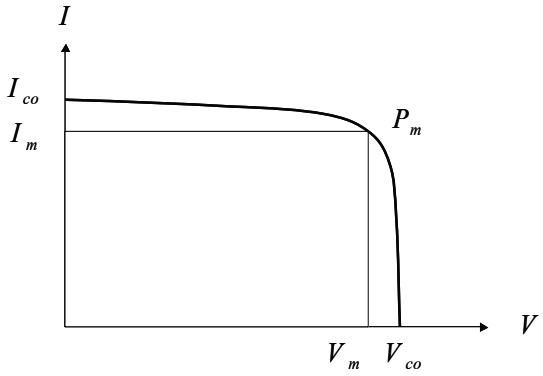


Figure 9-22 : Courbe d'une cellule solaire

L'équation (9-84) s'écrit alors

$$I = I_{ph} - I_s\left(e^{eV/kT} - 1\right) \tag{9-110}$$

Le premier terme de l'expression (9-110) est le photocourant, le second est un courant direct qui résulte de la polarisation de la diode dans le sens passant par la tension V qui apparait aux bornes de la résistance de charge.

Le *courant électromoteur* de la pile est le courant de court-circuit, sa *force électromotrice* est la tension de circuit ouvert, l'expression (9-110) donne

$$I_{cc} = I_{ph} \tag{9-111}$$

$$V_{co} = \frac{kT}{e} Ln\left(I_{ph} / I_s + 1\right) \tag{9-112}$$

a. Puissance débitée

La puissance fournie par la pile est donnée par le produit VI

$$P = VI = V\left(I_{ph} - I_s(e^{eV/kT} - 1)\right) \tag{9-113}$$

Cette puissance est maximum au point P_m (Fig. 9-22), défini par dP/dV=0, soit

$$I_{ph} - I_s\,(e^{eV/kT} - 1) - I_s \frac{eV}{kT} e^{eV/kT} = 0 \tag{9-114}$$

La tension V_m et le courant I_m au point P_m sont donnés par

$$\left(1+\frac{eV_m}{kT}\right)e^{eV_m/kT}=1+\frac{I_{ph}}{I_s} \tag{9-115}$$

$$I_m = I_s \frac{eV_m}{kT} e^{eV_m/kT} \tag{9-116}$$

La puissance débitée est alors donnée par le produit $V_m I_m$ qui s'écrit

$$P_m = V_m I_m = FF\, V_{co}\, I_{cc} \tag{9-117}$$

Le paramètre *FF* est le *facteur de remplissage* ou *facteur de forme*, il mesure le caractère rectangulaire de la courbe *I*(*V*). Il varie de 0,25 pour une cellule à faible rendement à 0,9 pour une cellule idéale.

b. Facteur de forme

Si V_j est la tension aux bornes de la jonction et r_s la résistance série de la diode, le courant et la tension de sortie sont donnés par

$$I = I_{ph} - I_s\ (e^{eV_j/kT} - 1) \tag{9-118}$$

$$V = V_j - r_s\, I \tag{9-119}$$

La caractéristique *I*(*V*) de la photopile est alors donnée par

$$I = I_{ph} - I_s\left(e^{e(V+r_s I)/kT} - 1\right) \tag{9-120}$$

Les allures de la caractéristique *I*(*V*) pour plusieurs valeurs de r_s sont représentées sur la figure (9-23). Ces caractéristiques montrent les effets dramatiques de r_s sur le facteur de forme (*FF*>0,8 pour r_s=0 et *FF*=0,25 pour r_s>1Ω). Le facteur de forme augmente avec V_{co} c'est-à-dire avec le gap du semiconducteur, pour r_s=0, *il* passe de 0,83 dans Si où V_{co}~0,6 V à 0,87 dans GaAs où V_{co}~0,9 V.

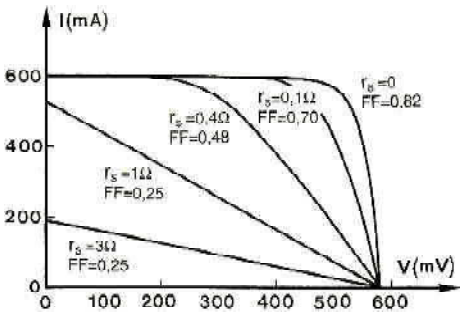


Figure 9-23: Caractéristique I(V) pour plusieurs valeurs de la résistance série

c. Rendement

Le rendement de la photopile est donné par le rapport de la puissance maximum disponible à la puissance du rayonnement incident

$$\eta = P_m \,/\, P_{solaire} = FF\, V_{co}\, I_{cc} \,/\, P_{solaire} \qquad (9\text{-}121)$$

Avec $P_{solaire}$~1 kW/m^2, V_{co}~0,7 V, I_{cc}~30 mA/cm^2 et FF~0,8 , on obtient η~20%.

L'expression (9-121) montre que les performances d'une photopile résultent directement des valeurs des trois paramètres I_{cc} , V_{co} et FF. Ces paramètres sont fonction d'une part de propriétés spécifiques du matériau telles que le gap, les coefficients d'absorption et de réflexion, la longueur de diffusion des porteurs ou la vitesse de recombinaison en surface, et d'autre part de paramètres technologiques tels que la profondeur de la jonction, la largeur de la zone de charge d'espace ou la présence de résistances parasites. Ainsi par exemple la nature et la valeur du gap jouent un rôle majeur. Un matériau à gap direct comme GaAs a un coefficient d'absorption bien supérieur à un matériau à gap indirect comme Si. Dans le premier les photoporteurs sont donc créés sur une plus faible profondeur (1 à 2 μm) que dans le second (quelques dizaines de μm). Il en résulte que dans le premier la vitesse de recombinaison en surface et la profondeur de jonction jouent un rôle plus important que dans le second. En contre partie les recombinaisons sur le contact arrière jouent un rôle plus important dans le second que dans le premier. La valeur du gap quant à elle conditionne la tension de circuit ouvert V_{co} à travers la hauteur de barrière de la jonction, et le courant de court-circuit I_{cc} à travers le taux de recouvrement du spectre solaire et de la réponse spectrale de la diode. V_{co} est d'autant plus grand et I_{cc} d'autant plus petit que le gap est grand. Enfin le facteur de forme FF est fonction de la tension en circuit ouvert V_{co} et des résistances série r_s et shunt r_{sh} de la diode. FF est d'autant plus grand que V_{co} et r_{sh} sont grands et que r_s est petit. Un bon facteur de forme nécessite r_s<1 Ωcm^2 et r_{sh} >10^4 Ω. Une quatrième figure de mérite, considérée notamment dans les utilisations spatiales, est la résistance aux rayonnements cosmiques. Ce paramètre est mesuré par le rapport des rendements de la cellule en début de vie BOL (Beginning Of Life) et en fin de vie EOL (End Of Life) du satellite.

La valeur limite η_o du rendement de conversion d'une photopile peut être calculée dans les conditions idéales où tous les photons d'énergie supérieure au gap créent des paires électron-trou (R=0) et où la diode est idéale (r_s~0 , r_{sh}~∞). La tension V_{co} est alors donnée par l'expression (9-112) et le courant I_{cc} (=I_{ph}) par l'expression (9-106) dans laquelle $\phi = (1\text{-}R)\,\phi_o \equiv \phi_o$. Le rendement s'écrit alors

$$\eta_0 = \frac{FF\,\phi_0\, kT\, Ln\left(e\phi_0 / J_s + 1\right)}{\int_0^\infty \phi_0(E)\, E\, dE} \qquad (9\text{-}122)$$

où $\phi_o(E)$ est le flux de photons d'énergie E et $\phi_0 = \int_{E_g}^{\infty} \phi_0(E)dE$ l'intégrale de ce flux sur tous les photons d'énergie supérieure au gap du matériau. En supposant tous les autres paramètres indépendants du gap, la variation de η_o avec le gap du matériau est représentée sur la figure (9-24). En prenant les paramètres du silicium de résistivité 1 Ωcm, le maximum de η_o se situe au voisinage de 1,5 eV avec une valeur de l'ordre de 22 %. Lorsque le gap du matériau s'éloigne de part et d'autre de la valeur optimum de 1,5 eV le rendement diminue. Du côté des hautes énergies cette diminution résulte du fait que le matériau devient transparent à une fraction croissante du spectre solaire, ϕ_o diminue (Eq. 9-122). Du côté des basses énergies, certes ϕ_o augmente, mais le rendement diminue car de plus en plus de photons absorbés ont des énergies bien supérieures au gap. Or la transformation directe d'énergie optique en énergie électrique est un phénomène quantique, *un* photon crée *une* paire électron-trou. Ainsi, même si le rendement quantique est égal à 1, le rendement énergétique n'est pas pour autant égal à 1. Les photons de haute énergie $h\nu >> E_g$, transportent plus d'énergie optique que les photons de basse énergie $h\nu \sim E_g$ mais ne créent pas davantage de paires électron-trou que ces derniers. Ils créent des porteurs chauds, qui participent aux phénomènes de conduction électrique après thermalisation aux extrema de bande. Cette thermalisation correspond à la perte de l'excès d'énergie optique du photon qui se retrouve dans le cristal sous forme thermique. On réduit cet effet en empilant des cellules de gaps différents, qui échantillonnent le spectre solaire (cellules multicolores).

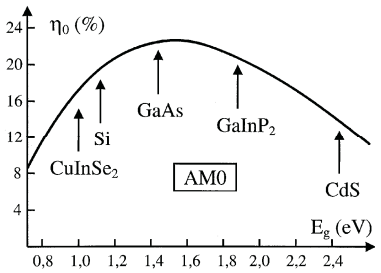


Figure 9-24 : Rendement d'une photopile

Notons enfin que l'effet du coefficient de réflexion $R = (n\text{-}1)^2 / (n\text{+}1)^2 \sim 30\%$ est minimisé par le dépôt sur la face frontale, d'une couche antiréfléchissante.

d. Différents types de cellule

Il existe deux approches de base dans la réalisation des cellules solaires. La première consiste à réaliser la cellules en couches minces, généralement amorphes ou polycristallines, déposées sur un substrat peu coûteux comme le verre ou une feuille métallique. Leurs principaux intérêts résident d'une part dans leur stabilité et d'autre part dans la possibilité de réaliser des cellules de grande surface sensible (jusqu'à 0,5 m^2). En outre une grande variété de techniques de dépôt étant disponible il est possible d'optimiser le coût de fabrication. Les matériaux les plus utilisés dans cette approche sont les semiconducteurs II-VI et surtout les composés de type I-III-VI_2 tel que $CuInSe_2$. Ce dernier en particulier a un coefficient d'absorption parmi les plus importants de tous les semiconducteurs ($\sim 10^5$ cm^{-1}), on

peut de ce fait réaliser des zones actives très minces, ce qui rend moins critique la longueur de diffusion des photoporteurs. La valeur du gap de $CuInSe_2$ (1,02 eV) est toutefois assez éloignée de la valeur optimum (Fig. 9-24), on l'augmente en utilisant l'alliage quaternaire $CuIn_xGa_{1-x}Se_2$.

La seconde approche consiste à réaliser la cellule avec du matériau monocristallin de haute qualité. Cette approche est dominée par les cellules à base de silicium et de composés III-V, son objectif est le haut rendement.

- *Cellules au silicium*

Compte tenu des caractéristiques intrinsèques du silicium (gap indirect, inférieur à la valeur optimum) les cellules au silicium ne sont pas les plus performantes. Toutefois, en raison de leur coût elles demeurent compétitives au moins dans les applications terrestres. L'amélioration de leurs performances est obtenue par les optimisations des collectes d'une part des photons incidents et d'autre part des porteurs photogénérés. La première est obtenue en minimisant la surface frontale des contacts, en optimisant les couches antiréfléchissantes et éventuellement en créant un passage multiple du flux de photons avec une face arrière réfléchissante. La seconde optimisation, qui concerne la collecte des photoporteurs, doit intégrer le fait qu'en raison de la nature indirecte de son gap le silicium a un coefficient d'absorption relativement faible. La zone active dépasse très largement la zone de charge d'espace de sorte que le photocourant résulte essentiellement de la diffusion des photoporteurs créés dans les régions neutres. Pour optimiser cette diffusion on fait en sorte que l'essentiel du rayonnement soit absorbé dans la région de type *p*, car les porteurs minoritaires y sont les électrons plus mobiles que les trous, en outre on positionne cette région à l'arrière de la jonction afin d'éviter les recombinaisons en surface, enfin pour éviter la perte des photoélectrons sur le contact arrière on élabore une région p^+ qui crée une barrière de potentiel et renvoie les photoélectrons vers la jonction. La structure type d'une cellule au silicium est représentée sur la figure (9-25). La région frontale de type *n*, appelée *émetteur,* est très mince afin de laisser passer le rayonnement. La région de type *p*, appelée *absorbeur*, est de l'ordre d'une centaine de μm.

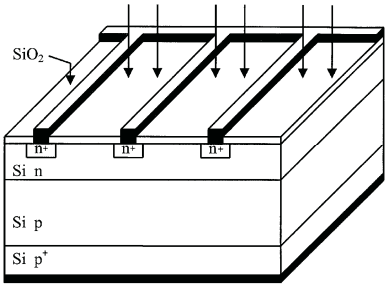


Figure 9-25 : Cellule silicium

Au stade de la production les cellules au silicium ont un rendement voisin de 17 % pour le silicium mono-cristallin et 8 % pour le silicium poly-cristallin. Au stade du développement ces mêmes cellules atteignent des rendements respectifs de 24 % et 18 % .

- *Cellules multicolores à composés III-V*

L'objectif des cellules multicolores, ou multijonctions, est la réduction des deux principales causes de perte de rendement des cellules unijonction, les pertes résultant de la non absorption des photons d'énergie $h\nu<E_g$ et les pertes thermiques associées à la thermalisation des photoporteurs chauds créés par les photons d'énergie $h\nu>E_g$. Dans cette optique les semiconducteurs III-V tels que GaAs et InP, ainsi que leurs alliages, sont potentiellement plus performants que le silicium pour deux raisons évidentes, leur gap est direct et sa valeur est voisine de la valeur optimum. Le principal frein dans les applications terrestres reste leur coût, mais leur domaine de prédilection est incontestablement le domaine spatial où le marché est moins sensible au coût qu'aux performances. Les figures de mérite primordiales sont ici d'une part le rapport puissance/poids et d'autre part la résistance aux rayonnements cosmiques. Cette dernière conditionne en particulier le rapport P_{EOL}/P_{BOL} des puissances débitées en fin et début de vie du satellite (*la durée de vie d'un satellite de télécommunication est comprise entre 5 et 15 ans*). Parce que leur gap est plus grand, les composés III-V résistent mieux que le silicium aux rayonnements cosmiques et aux températures de fonctionnement en orbite (~50 °C). Les puissances de sortie en EOL de photopiles à base de composés III-V sont 40 à 60 % supérieures à celles de photopiles à base de silicium.

Le concept des piles multicolores est simple: plusieurs cellules avec des gaps différents sont empilées de manière à ce que le rayonnement incident interagisse séquentiellement avec les cellules de gaps décroissants. Les cellules supérieures, à grands gaps, absorbent et convertissent les photons de haute énergie, et transmettent les

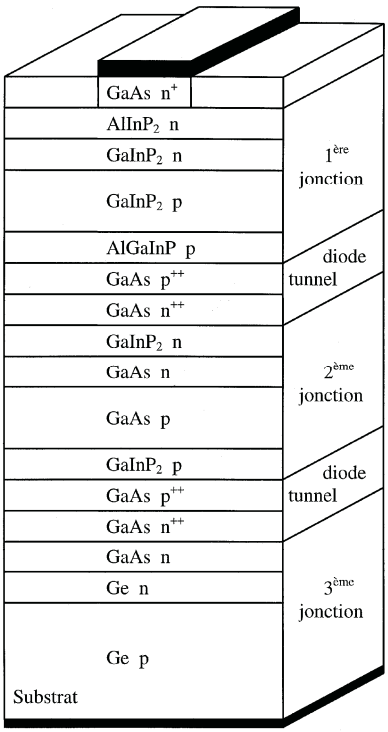


Figure 9-26 : Cellule multicolore

photons de basse énergie aux cellules sous-jacentes de gaps plus petits. Ces dernières absorbent alors et convertissent les photons de plus basse énergie. Les pertes par thermalisation sont ainsi minimisées car chaque cellule convertit des photons d'énergie $h\nu \sim E_g$. Deux paramètres majeurs doivent être maîtrisés dans ce type de structure. Le premier est la réalisation, entre chacune des jonctions, d'une région conductrice avec contact ohmique de part et d'autre, transparente à la gamme spectrale exploitée par les cellules sous-jacentes. Ces liaisons sont assurées par des diodes tunnel. Le deuxième paramètre à maîtriser est l'accord des photocourants. En effet, alors que les tensions de circuit ouvert V_{co} s'ajoutent, les courants de court-circuit I_{cc} doivent être ajustés car le plus faible d'entre eux impose sa valeur au courant résultant. Ainsi les gaps et les épaisseurs des cellules supérieures doivent être ajustés pour que ces dernières n'absorbent ni trop ni trop peu de rayonnement. En effet dans le premier cas les courants débités par les cellules inférieures sont trop faibles, dans le deuxième cas ce sont les courants débités par les cellules supérieures qui sont trop faibles. La figure (9-26) représente la structure d'une cellule tricolore épitaxiée en accord de maille sur un substrat de germanium. La figure (9-27) représente l'échantillonnage que cette structure opère dans le spectre solaire. Le rendement maximum théorique de ces photopiles est de 35% .

A titre de comparaison les meilleures cellules au silicium monocristallin ont un rendement de 17 % en BOL et 12 % en EOL, les cellules unijonction GaAs, bijonction $GaInP_2$/GaAs et trijonction $GaInP_2$/GaAs/Ge ont respectivement des rendements de 19, 21,6 et 23 % en BOL et 16, 18 et 20 % en EOL. Des cellules bicolores $GaInP_2$/GaAs ont été réalisées avec V_{co}~2,4 V, I_{cc}~14 mA/cm^2, FF~89 % et un rendement de l'ordre de 29 %. Enfin toujours à titre de comparaison, avec des panneaux classiques de 70 m^2 les cellules à base de GaAs débitent 15 kW en EOL alors que les mêmes panneaux à cellules silicium ne débitent que 6 kW.

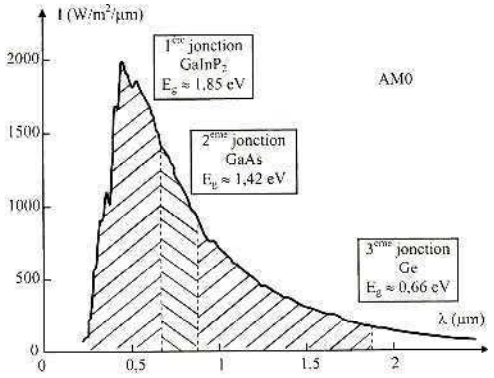


Figure 9-27 : Echantillonnage du spectre

9.3. Emetteurs de rayonnement à semiconducteur

9.3.1. *Diode électroluminescente - LED*

a. Principe de fonctionnement

Lorsqu'une jonction *pn* est polarisée dans le sens direct, les électrons, qui sont majoritaires dans la région de type *n*, sont injectés dans la région de type *p* où ils se recombinent avec les trous. Inversement pour les trous. La structure de base d'une diode électroluminescente, LED (Light Emitting Diode), est une jonction *pn* réalisée à partir de semiconducteurs dans lesquels les recombinaisons des porteurs excédentaires sont essentiellement radiatives. La structure type d'une diode électroluminescente et son principe de fonctionnement sont représentés sur la figure (9-28).

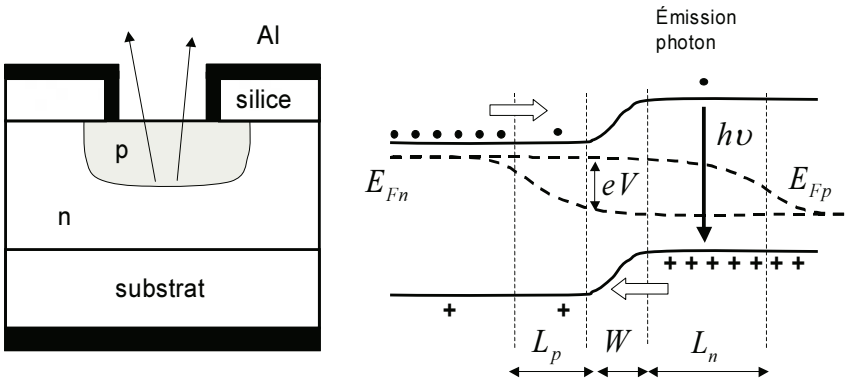


Figure 9-28 : Diode électroluminescente

Une tension de polarisation directe *V* fixe la séparation des niveaux de Fermi. Les distributions des électrons dans la région de type *p* et des trous dans la région de type *n*, sont données par les expressions (9-52-a et b). Les recombinaisons des porteurs excédentaires se manifestent dans trois régions différentes qui sont d'une part la zone de charge d'espace, et d'autre part les régions neutres *n* et *p*. Dans chacune de ces dernières, la zone émettrice est limitée à la longueur de diffusion des porteurs minoritaires. La zone de charge d'espace joue quant à elle un rôle mineur dans la mesure où elle est pratiquement inexistante en raison du fait que la jonction est fortement polarisée dans le sens direct.

Les taux d'injection de porteurs minoritaires dans chacune des régions de la jonction, sont définis par

$$\gamma_n = \frac{j_n(x_p)}{J} \qquad \gamma_p = \frac{j_p(x_n)}{J}$$

Dans la mesure où les longueurs des régions *n* et *p* sont plus importantes que les longueurs de diffusion des porteurs minoritaires, les courants $j_n(x_p)$ et $j_p(x_n)$ sont donnés par (Eq. 2-50-a et b)

$$j_n(x_p)=\frac{en_i^2 D_n}{N_a L_n}\left(e^{eV/kT}-1\right) \tag{9-123}$$

$$j_p(x_n)=\frac{en_i^2 D_p}{N_d L_p}\left(e^{eV/kT}-1\right) \tag{9-124}$$

Le rapport des taux d'injection d'électrons et de trous s'écrit donc

$$\frac{\gamma_n}{\gamma_p}=\frac{D_n L_p N_d}{D_p L_n N_a}=\sqrt{\frac{\mu_n \tau_p}{\mu_p \tau_n}}\frac{N_d}{N_a}\approx\sqrt{\frac{\mu_n N_d}{\mu_p N_a}}=\sqrt{\frac{\sigma_n}{\sigma_p}} \tag{9-125}$$

La mobilité des électrons étant beaucoup plus grande que celle des trous, le taux d'injection d'électrons dans la région de type *p* est plus important que le taux d'injection de trous dans la région de type *n*. Il en résulte que la région la plus radiative est la région de type *p*. C'est la raison pour laquelle cette région constitue la face émettrice dans la structure de la figure (9-28). Il faut ajouter que pour des raisons d'intensité d'émission, les régions *n* et *p* de la diode sont très dopées. Ces dopages importants se traduisent par une diminution du gap, dont on peut montrer qu'elle est plus importante dans la région *p* que dans la région *n*. Cette différence de gap favorise encore l'injection d'électrons par rapport à celle de trous.

b. Spectre d'émission

Le spectre, c'est-à-dire la couleur, du rayonnement émis par une diode électroluminescente est évidemment conditionné par le gap du matériau de type *p*, dans lequel se produit l'essentiel des recombinaisons radiatives. Dans la mesure où certaines transitions mettent en jeu des niveaux d'impureté, le spectre d'émission est aussi conditionné par le type de dopant.

La plupart des semiconducteurs de type III-V sont miscibles entre eux en toute proportion, de sorte que la réalisation d'alliages ternaires ou quaternaires permet à l'heure actuelle de couvrir tout le spectre visible. Les alliages de type GaAlAs, GaAsP ou InGaAlP, permettent de couvrir le domaine des grandes longueurs d'onde du spectre visible par la seule variation de composition. En ce qui concerne le domaine des courtes longueurs d'onde, les composés II-VI et III-V à grand gap sont énergétiquement bien adaptés (Fig. 9-11), mais la difficulté de maîtriser leur dopage a longtemps compliqué la réalisation de jonctions *pn*. La première démonstration d'émission bleu-vert a été faite en 1991 avec le composé II-VI ZnSe. Les émissions verte et bleue sont désormais obtenues avec l'alliage III-V InGaN. La figure (9-29) représente les énergies des raies d'émission que l'on peut obtenir avec différents alliages.

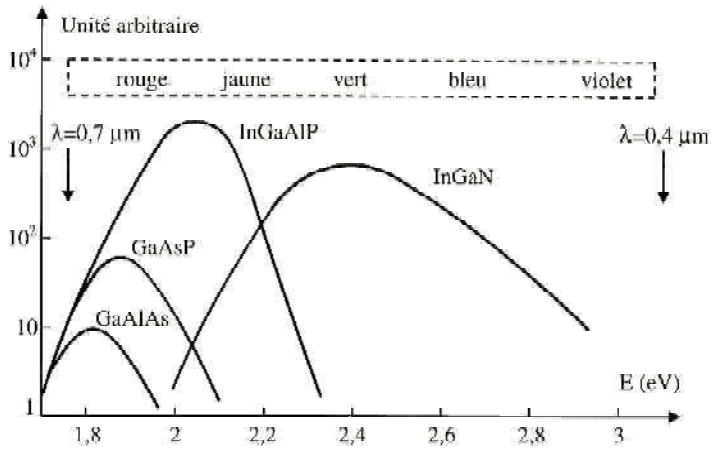


Figure 9-29 : Spectre d'émission de différents alliages

c. Rendement

Outre la longueur d'onde du rayonnement émis, un paramètre essentiel dans le fonctionnement d'une diode électroluminescente est son rendement. On définit plusieurs rendements.

- *Rendement quantique interne* η_i

Considérons le processus physique interne, les porteurs excédentaires injectés dans chacune des régions, se recombinent avec les porteurs majoritaires. Ces recombinaisons mettent en jeu divers processus dont certains sont radiatifs, d'autres non. On définit le rendement quantique interne η_i , comme le rapport du nombre de photons créés à la jonction, au nombre de porteurs qui traversent cette jonction.

Tous les porteurs qui traversent la jonction se recombinent, les uns radiativement, les autres non, de sorte que η_i n'est autre que le rapport du taux de recombinaisons radiatives au taux global de recombinaisons.

$$\eta_i = \frac{r_r}{r} = \frac{r_r}{r_r + r_{nr}} \tag{9-126}$$

Les taux de recombinaison sont exprimés en fonction des durées de vie par $r_r = -\Delta n / \tau_r$, $r_{nr} = -\Delta n / \tau_{nr}$ où τ_r et τ_{nr} sont les durées de vie radiative et non radiative. Le rendement interne s'écrit donc

$$\eta_i = \frac{\tau_{nr}}{\tau_r + \tau_{nr}} \tag{9-127}$$

Cette expression est identique à l'expression (9-38), η_i n'est autre que le rendement radiatif du matériau. Si $\tau_r << \tau_{nr}$, $\eta_i \approx 1$. La durée de vie radiative est faible dans les matériaux à gap direct, où le couplage électron-photon est grand en raison de la possibilité de satisfaire simultanément aux règles de conservation de l'énergie et du moment. C'est la raison pour laquelle les matériaux à gap direct sont les mieux adaptés à la réalisation de diodes électroluminescentes. Il faut toutefois noter l'exception de GaP qui a un gap indirect, mais qui dopé azote, présente un bon rendement radiatif. L'importante luminescence du centre azote dans GaP-N est due à la nature du potentiel associé à l'azote, en substitution du phosphore.

Les matériaux utilisés dans la réalisation de diodes électroluminescentes sont caractérisés par $\eta_i \approx 100$ %.

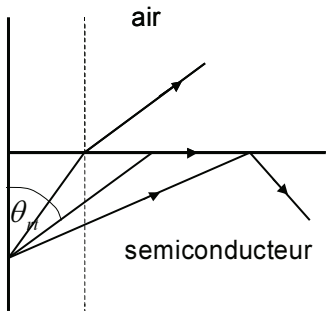


Figure 9-30 : Rendement optique

- *Rendement optique* η_o

Tous les photons créés à la jonction ne sortent pas de la diode. Une fraction de ces photons est réabsorbée, souvent après réflexion à la surface du matériau. Le rendement optique η_O, est défini comme le rapport du nombre de photons émis à l'extérieur de la diode, au nombre de photons créés à la jonction.

Le paramètre qui conditionne le rendement optique, est l'indice de réfraction du matériau. Compte tenu de la valeur élevée de cet indice, $n \approx 3{,}5$, deux phénomènes limitent l'émission des photons à l'extérieur du matériau: la valeur élevée du coefficient de réflexion à l'interface air-semiconducteur et l'existence d'un angle de réflexion totale relativement faible (Fig. 9-30).

Le coefficient de réflexion en incidence normale est donné par

$$R = \left(\frac{n-1}{n+1}\right)^2 \approx \left(\frac{3{,}5-1}{3{,}5+1}\right)^2 \approx 0{,}3$$

L'angle de réflexion totale est donné par la loi de Descartes, $n_1 \sin \theta_1 = n_2 \sin \theta_2$, soit pour $n_1 = n$, $n_2 = 1$ et $\theta_2 = \pi/2$

$$n \sin \theta_{rt} = \sin(\pi/2) = 1 \quad \Rightarrow \quad \sin \theta_{rt} = 1/n \quad \Rightarrow \quad \theta_{rt} \approx 16°$$

Les photons qui atteignent la surface de sortie de la diode sous une incidence supérieure à 16° sont totalement réfléchis à l'intérieur du semiconducteur. Le coefficient de transmission $T=1-R$ varie donc de 0,7 à 0 quand l'angle d'incidence varie de 0 à 16°. Seuls les photons émis à la jonction dans un cône de 16°, *cône d'extraction*, sortent de la diode. L'angle solide couvrant tout l'espace est (Fig. 9-31)

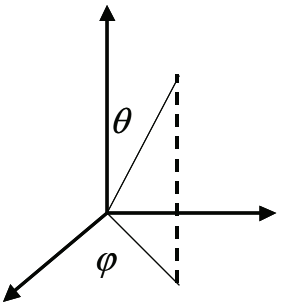


Figure 9-31 : Coordonnées sphériques

$$\Omega_o = \int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta d\theta = 4\pi \qquad (9\text{-}128)$$

L'angle solide sous-tendu par l'angle θ_{rt} est donné par

$$\Omega_{rt} = \int_0^{2\pi} d\phi \int_0^{\theta_{rt}} \sin\theta d\theta = 2\pi\left(1 - \cos\theta_{rt}\right) \qquad (9\text{-}129)$$

Dans la mesure où θ_{rt} est petit (<16°), $\cos\theta_{rt} \approx 1 - \theta_{rt}^2/2$ et l'angle solide du cône d'extraction s'écrit

$$\Omega_{rt} \approx 2\pi\left(1 - \left(1 - \theta_{rt}^2/2\right)\right) = \pi\theta_{rt}^2 \qquad (9\text{-}130)$$

Le rayonnement étant émis de manière isotrope dans tout l'espace au niveau de la jonction, le rendement de sortie de ce rayonnement est donné par

$$\eta_o = \frac{\Omega_{rt}}{\Omega_o}\overline{T} = \frac{\theta_{rt}^2}{4}\overline{T} = \frac{\theta_{rt}^2}{4}(1 - \overline{R}) \qquad (9\text{-}131)$$

où $\overline{T}$ et $\overline{R}$ représentent les coefficients moyens de transmission et de réflexion pour des angles d'incidence à la surface compris entre 0 et θ_{rt}. En supposant R constant égal à sa valeur en incidence normale, l'expression (9-131) s'écrit

$$\eta_o \approx \frac{\theta_{rt}^2}{4}\left(1 - \left(\frac{n-1}{n+1}\right)^2\right) \qquad (9\text{-}132)$$

Enfin dans la mesure où θ_{rt} est petit, $\theta_{rt} \approx \sin\theta_{rt} = 1/n$, le rendement optique s'écrit

$$\eta_o \approx \frac{1}{4n^2}\left(1-\left(\frac{n-1}{n+1}\right)^2\right) \approx \frac{1}{n(n+1)^2} \qquad (9\text{-}133)$$

Avec $n = 3{,}5$, on obtient $\eta_o \approx 1$ % . On améliore le rendement optique en recouvrant la diode d'un matériau plastique transparent, dont l'indice n_p est supérieur à celui de l'air, $n_p \approx 1{,}5$. L'expression (9-133) s'écrit alors

$$\eta_o \approx \frac{1}{4}\left(\frac{n_p}{n}\right)^2\left(1-\left(\frac{n-n_p}{n+n_p}\right)^2\right) \qquad (9\text{-}134)$$

Avec $n = 3{,}5$ et $n_p = 1{,}5$, le rendement optique à l'interface semiconducteur-plastique est $\eta_o \approx 4\%$. Il existe quelques pertes à l'interface plastique-air, mais on limite ces pertes en réalisant le plastique sous forme de dôme hémisphérique pour que le rayonnement sorte en incidence normale (Fig. 9-33). Le coefficient de transmission à cet interface est donné par $T' = 1-((n_p-1)/(n_p+1))^2 = 96$ %.

- *Rendement quantique externe* η_e

Le rendement quantique externe est défini comme le rapport du nombre de photons émis par la diode, au nombre de porteurs traversant la jonction. En d'autres termes ce rendement est donné par

$$\eta_e = \eta_i \ \ \eta_o \qquad (9\text{-}135)$$

- *Rendement global* η

Le rendement global est défini comme le rapport de la puissance lumineuse émise, à la puissance électrique absorbée

$$\eta = W_{op} / W_{el} \qquad (9\text{-}136)$$

Si le nombre de photons émis par seconde est N_{ph} et si l'énergie de chacun de ces photons supposés identiques, est $h\upsilon$, la puissance lumineuse émise est $W_{op}=N_{ph}h\upsilon$.

La puissance électrique absorbée est donnée par le produit VI, où V est la tension appliquée à la diode et I le courant débité. I est la quantité de charges qui traversent la jonction par seconde, $I = N_{el}\,e$, ainsi $W_{el} = N_{el}\,eV$.

Le rendement global s'écrit donc

$$\eta = \frac{N_{ph}\ \ h\upsilon}{N_{el}\ \ eV} = \eta_e \frac{h\upsilon}{eV} \qquad (9\text{-}137)$$

L'énergie du rayonnement émis est voisine du gap du semiconducteur, $h\upsilon \approx E_g$. D'autre part la tension appliquée V est en partie perdue dans la résistance série de la diode, de sorte que le rendement s'écrit

$$\eta = \eta_e \frac{E_g / e}{r_s I + V_d} \tag{9-138}$$

où r_s représente la résistance série et V_d la tension de diffusion.

La cause essentielle du faible rendement des LED est la réflexion totale, à l'interface de sortie du semiconducteur, de la majeure partie du rayonnement créé dans la zone active. Ce rayonnement est émis dans un angle solide de 4π stéradians alors que seuls sortent de la diode les photons émis dans le cône d'extraction dont l'angle solide est inférieur à $\pi/10$ stéradian (Eq. 9-130). Afin d'augmenter la densité de photons dans le cône extraction différentes approches sont explorées.

Une technique consiste à dépolir la surface de sortie du semiconducteur, la rugosité d'interface augmente alors considérablement la probabilité de sortie des photons. Des rendements externes pouvant atteindre 30% ont été démontrés, mais la complexité des procédés de fabrication hypothèque beaucoup le développement de cette technique.

La technique la plus prometteuse consiste à insérer la région active de la diode entre les deux miroirs d'une cavité Pérot-Fabry, qui joue le rôle de résonateur. La LED porte alors le nom de RCLED (Resonant Cavity LED) ou LED à microcavité. La cavité, dont les miroirs peuvent être des miroirs de Bragg distribués (chapitre 12) ou de simples couches métalliques, est accordée sur le spectre d'émission spontanée de la région active. Par rapport à une diode conventionnelle l'effet de microcavité améliore à la fois le rendement externe et la directivité de la diode en augmentant la densité de photons dans le cône d'extraction, et la pureté spectrale du rayonnement par la résonance sur les modes de la cavité.

d. Temps de réponse - Fréquence de coupure

Un des intérêts majeurs de la diode électroluminescente dans le domaine des émetteurs de rayonnement, est sa simplicité de modulation par variation du courant injecté dans la diode. La fréquence de cette modulation est limitée par les capacités de transition et de diffusion de la jonction (chapitre 2). La diode étant polarisée dans le sens direct seule la capacité de diffusion joue un rôle appréciable.

Considérons une diode électroluminescente avec $d_p >> L_n$, polarisée par une tension continue V_0, sur laquelle est superposée une faible composante alternative v_1 de pulsation ω. En régime de faible injection la durée de vie des porteurs peut être supposée constante et la distribution des électrons dans la région de type p est régie par l'équation

$$\frac{\partial \Delta n(x,t)}{\partial t} = D_n \frac{\partial^2 \Delta n(x,t)}{\partial x^2} - \frac{\Delta n(x,t)}{\tau_n} \tag{9-139}$$

Sous la double action, de la tension de polarisation statique V_0 et de la tension de modulation $v_1 e^{j\omega t} t$, la densité d'électrons peut se mettre sous la forme

$$\Delta n(x,t) = \Delta n_o(x) + \Delta n_1(x) e^{j\omega t} \tag{9-140}$$

En portant cette expression dans l'équation (9-139) on obtient

$$\frac{d^2 \Delta n_o(x)}{dx^2} - \frac{\Delta n_o(x)}{L_n^2} = 0 \tag{9-141-a}$$

$$\frac{d^2 \Delta n_1(x)}{dx^2} - \frac{\Delta n_1(x)}{L_n^{*2}} = 0 \tag{9-141-b}$$

avec

$$L_n = \sqrt{D_n \tau_n} \qquad L_n^* = \sqrt{\frac{D_n \tau_n}{1 + j\omega\tau_n}}$$

Les solutions sont de la forme générale

$$\Delta n = A e^{-(x-x_p)/L} + B e^{(x-x_p)/L} \tag{9-142}$$

Avec $L=L_n$ pour la composante statique Δn_o, et $L=L_n^*$ pour la composante alternative Δn_1 ; x_p est la frontière de la zone de charge d'espace de la jonction dans la région de type *p* (Fig. 9-28). Dans la mesure où la condition $d_p >> L_n$ est vérifiée la constante *B* est nulle. En appelant $\Delta n_i(x_p)$ la valeur de chaque composante (i=0,1) en $x=x_p$, l'expression (9-142) s'écrit

$$\Delta n_o(x) = \Delta n_o(x_p) e^{-(x-x_p)/L_n} \tag{9-143-a}$$

$$\Delta n_1(x) = \Delta n_1(x_p) e^{-(x-x_p)/L_n^*} \tag{9-143-b}$$

Dans la mesure où l'injection d'électrons est beaucoup plus importante que l'injection de trous, le courant qui traverse la jonction s'écrit

$$J = j_n(x_p) + j_p(x_n) \approx j_n(x_p) \tag{9-144}$$

soit

$$J = eD_n \left(\frac{d\Delta n_o}{dx} \right)_{xp} + eD_n \left(\frac{d\Delta n_1}{dx} \right)_{xp} e^{j\omega t} = J_o + J(\omega) e^{j\omega t}$$

Compte tenu de (9-143-b), l'amplitude $J(\omega)$ de la composante alternative du courant s'écrit donc

$$J(\omega)=\frac{e\,D_n}{L_n^*}\Delta n_1(x_p) \tag{9-145}$$

Si η_e est le rendement quantique externe, le taux d'émission de photons est η_e fois le taux de recombinaison d'électrons. Ainsi le nombre de photons émis sur toute la profondeur de la jonction supposée beaucoup plus grande que la longueur de diffusion des porteurs minoritaires, est donné par unité de surface par

$$N=\eta_e\int_{x_p}^{\infty}\frac{\Delta n_o}{\tau_n}dx+\eta_e\int_{x_p}^{\infty}\frac{\Delta n_1}{\tau_n}dx\,e^{j\omega t}=N_o+N(\omega)e^{j\omega t}$$

Le nombre de photons émis est donc modulé à la fréquence $f=\omega/2\pi$. L'amplitude de cette modulation est donnée par

$$N(\omega)=\eta_e\int_{x_p}^{\infty}\frac{\Delta n_1}{\tau_n}dx=\frac{\eta_e}{\tau_n}\Delta n_1(x_p)\int_{x_p}^{\infty}e^{-(x-x_p)/L_n^*}dx$$

soit

$$N(\omega)=\eta_e\frac{L_n^*}{\tau_n}\Delta n_1(x_p) \tag{9-146}$$

Ainsi, l'efficacité de modulation définie comme le rapport du taux de modulation du rayonnement émis au taux de modulation du courant excitateur s'écrit, compte tenu des relations (9-146 et 145)

$$R(\omega)=\frac{N(\omega)}{J(\omega)}=\frac{\eta_e L_n^{*2}}{eD_n\tau_n}=\frac{\eta_e}{e}\left(\frac{L_n^*}{L_n}\right)^2$$

ou en explicitant L_n et L_n^*

$$R(\omega)=\frac{\eta_e}{e}\;\frac{1}{1+j\omega\tau_n} \tag{9-147}$$

Si on appelle $R_0=\eta_e/e$, l'efficacité de modulation en basse fréquence, l'amplitude du rapport du taux de modulation du rayonnement émis au taux de modulation du courant excitateur s'écrit, en posant $\omega_c=1/\tau_n$

$$R=\frac{R_o}{\sqrt{1+\omega^2/\omega_c^2}} \tag{9-148}$$

La fréquence de coupure $f_c=\omega_c/2\pi$ de la diode est d'autant plus élevée que la durée de vie des électrons dans la région de type p est faible. Or $\tau_n=1/Bp_0$ (Par. 9.1.4.a), on augmente donc la fréquence de coupure par un dopage important de la région de type *p*. Cependant, lorsqu'un semiconducteur est dopé au voisinage de la

limite de solubilité des agents dopants, des centres non radiatifs apparaissent dans le matériau. Dans GaAs, pour des concentrations de dopage supérieures à 10^{18} cm^{-3} le rendement quantique externe commence à diminuer en raison de la diminution du rendement radiatif. Pour ce type de dopage la durée de vie des électrons dans la région de type p est $\tau_n = (Bp_o)^{-1} = (7{,}2.10^{-10}.10^{18})^{-1} = 1{,}4.10^{-9}\, s$, la fréquence de coupure de la diode $f = 1/2\pi\tau_n$ est par conséquent de l'ordre de 100 MHz.

Il faut noter qu'en raison du phénomène d'émission stimulée, la durée de vie radiative est plus faible dans les diodes lasers que dans les diodes électroluminescentes. La fréquence de coupure des lasers est de ce fait plus élevée que celle des diodes électroluminescentes (Par. 9.3.2.f).

e. Structure de la diode

La structure d'une diode électroluminescente est conditionnée par le but recherché. Les trois principales utilisations des LED sont l'affichage, les photocoupleurs et les télécommunications par fibres optiques.

Dans le premier cas, le but recherché est évidemment une émission visible de couleur déterminée, qui fixe le gap du matériau utilisé, et une surface émettrice de dimension suffisante, qui conditionne le mode d'encapsulation.

Dans un photocoupleur, la diode électroluminescente est montée en regard d'un phototransistor au silicium dans un boîtier unique. Cette structure qui présente généralement une isolation électrique de l'ordre de 2000 volts permet de transmettre des signaux variables entre deux circuits entièrement isolés l'un de l'autre. Ces photocoupleurs sont essentiellement utilisés pour la transmission à grande vitesse de signaux logiques. Dans ces systèmes, la caractéristique essentielle de la diode est évidemment un spectre d'émission compatible avec la réponse spectrale du silicium, on utilise généralement des diodes au GaAsP.

Dans les télécommunications par fibres optiques, le but recherché est une émission centrée dans les fenêtres de transparence et de faible dispersion des fibres, λ=1,3 μm et λ=1,55 μm, une grande brillance c'est-à-dire une forte luminosité à partir d'une faible surface, compatible avec les dimensions des fibres, et enfin une bonne possibilité de modulation.

Certains paramètres sont communs à tous les types d'utilisation, c'est le cas en particulier du rendement quantique interne. Un bon rendement interne η_i passe par une grande durée de vie non radiative. Ceci nécessite la réalisation de matériaux de très bonnes qualités chimique et cristallographique, ce qui impose pratiquement la réalisation de couches épitaxiées. Pour augmenter le rendement optique d'autre part, on réalise la couche émettrice de type p en surface, avec une épaisseur optimum $d_p \approx 3\, L_n$, où L_n est la longueur de diffusion des électrons.

La structure interne d'une diode électroluminescente à émission orange est représentée sur la figure (9-32).

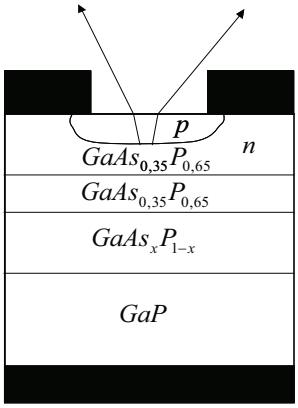


Figure 9-32 : Diode électroluminescente orange

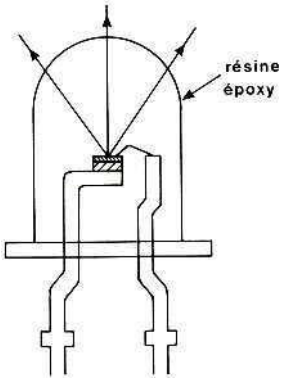


Figure 9-33 : Packaging d'une LED

Elle est constituée d'un substrat de GaP de type *n*, qui sert de support à la couche active et assure de bonnes conductivités électrique et thermique, vers l'électrode métallique. Une couche de $GaAs_xP_{1-x}$ où x varie de 0 à 0,35 sert d'interface en réalisant l'accord de maille entre le substrat et la couche active. Une couche de $GaAs_{0,35}P_{0,65}$ parfait cette interface. La couche active est réalisée par une jonction *pn* au $GaAs_{0,35}P_{0,65}$, dont la région de type *p*, de quelques microns d'épaisseur, constitue la face de sortie de la diode. Le contact métallique supérieur doit avoir une surface et une géométrie assurant à la fois une bonne homogénéité du courant et une surface émettrice suffisante.

La structure ainsi réalisée est ensuite encapsulée dans un dôme de résine transparente ou translucide. Ce dôme a trois fonctions essentielles, il protège la structure et surtout le fil de connexion, il augmente le rendement optique en réduisant la discontinuité d'indice entre le semiconducteur et l'air extérieur, il conditionne la distribution spatiale du rayonnement émis. La résine utilisée est une résine époxy à deux composants, les diodes sont encapsulées à partir de barrettes d'une vingtaine d'éléments, la forme du dôme est fixée par un moule de polypropylène. L'encapsulation standard consiste en un cylindre terminé par un dôme hémisphérique jouant le rôle de lentille (Fig. 9-33). La distribution spatiale du rayonnement émis est fonction de la hauteur du dôme par rapport à la surface active, la diode est d'autant plus directive que le dôme est éloigné.

f. Brillance de la diode et distribution spatiale du rayonnement

La brillance B d'une source est définie comme la puissance émise par unité de surface de la source et unité d'angle solide, B(W/sr/m^2). Si la source a une surface émettrice S homogène et si le rayonnement est émis de manière isotrope dans tout l'espace, le flux total d'énergie émis est $\phi_i = 4\pi SB$.

La diode électroluminescente a une surface émettrice plane. Elle n'émet de ce fait que dans un demi-plan, et de plus de manière non isotrope. La brillance de la diode est fonction de la direction d'émission par rapport à la normale à la surface active. Cette brillance obéit sensiblement à la loi de Lambert et s'écrit

$$B = B_o \cos\alpha \tag{9-149}$$

où B_0 est la brillance dans la direction normale à la surface et α l'angle d'émission par rapport à cette direction. Un exemple de distribution spatiale du rayonnement est représenté sur la figure (9-34).

Si la surface émettrice S de la diode est supposée homogène, le flux total d'énergie lumineuse émise est donné par

$$\phi_t = S\int_\Omega B d\Omega \tag{9-150}$$

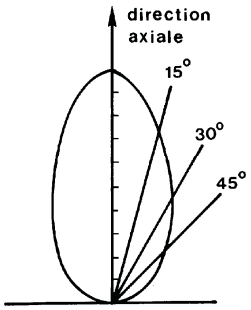


Figure 9-34 : Répartition de l'émission

L'angle solide Ω sous-tendu par l'angle α est donné par $\Omega = 2\pi(1\text{-}\cos\alpha)$ (Eq. 9-129), et par suite

$$d\Omega = 2\pi \sin\alpha \, d\alpha \tag{9-151}$$

Le flux émis par la diode dans tout le demi-plan s'écrit donc

$$\phi_t = 2\pi S\int_0^{\pi/2} B\sin\alpha d\alpha \tag{9-152}$$

ou, en explicitant B à partir de l'expression (9-149)

$$\phi_t = 2\pi S B_o\int_0^{\pi/2} \sin\alpha\cos\alpha d\alpha = 2\pi S B_o\int_0^1 \sin\alpha \, d(\sin\alpha)$$

soit

$$\phi_t = \pi S B_o \tag{9-153}$$

9.3.2. *Diode laser - LD*

Historiquement l'aventure du laser commence en 1917 quand Einstein découvre l'existence du processus d'émission stimulée, c'est-à-dire l'émission d'un photon commandée par un autre photon. En 1951 Weber et Twones aux Etats-Unis, et en 1954 Basov et Prokhorov en Union Soviétique, proposent d'utiliser l'émission stimulée pour amplifier les hyperfréquences. En 1954 Twones réalise le premier MASER (Microwave Amplification by Stimulated Emission of Radiation) en utilisant les propriétés d'inversion de la molécule d'ammoniac. En 1958 Shalow et Twones démontrent la possibilité d'étendre le MASER aux longueurs d'onde visibles, et en juillet 1960 Maiman, à la Hugues Aircraft Company, réalise le premier LASER (Light Amplification by Stimulated Emission of Radiation) en

utilisant les niveaux de l'ion C_r^{3+} dans Al_2O_3 , c'est le laser à rubis. Dès lors de nombreuses recherches se développent et différents types de laser, solides, à gaz, à colorants, sont réalisés. En ce qui concerne les lasers à semiconducteur, dès 1958 Aigrain émet l'hypothèse de l'utilisation des semiconducteurs pour obtenir l'effet laser, le premier laser est réalisé à GaAs, en 1962.

a. Principe de fonctionnement

Le principe du laser est simple, la figure (9-35) montre le principe de fonctionnement du laser à quatre niveaux.

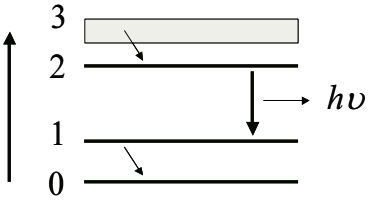


Figure 9-35 : Principe du laser

Le système est *pompé* de l'état fondamental (0) dans l'état (3), par une excitation extérieure. Le niveau (3) se vide alors dans le niveau (2) et le niveau (1) se vide dans le niveau (0). Si la durée de vie dans l'état (2) est très supérieure aux durées de vie associées aux transitions (3)→(2) et (1)→(0), la population de l'état (2) augmente et celle de l'état (1) diminue. Quand *l'inversion de population* est réalisée, c'est-à-dire quand l'état (2) est plus peuplé que l'état (1), $N_2 > N_1$, tout rayonnement d'énergie $h\upsilon = E_2 - E_1$ induit dans le matériau, davantage de transitions de haut en bas que de bas en haut, le milieu amplifie le rayonnement d'énergie $h\upsilon$, la condition d'effet laser est réalisée.

Dans le cas des semiconducteurs, le problème est sensiblement différent. Alors que dans les lasers classiques les états électroniques sont localisés et de spectre discret, dans les semiconducteurs les niveaux d'énergie sont groupés dans des bandes permises où leur répartition est quasi-continue. Cette spécificité du semiconducteur entraîne deux conséquences au niveau du laser. La première est que la condition classique d'inversion de population entre deux niveaux discrets, ($N_2 > N_1$) doit s'exprimer ici dans un formalisme adapté à la structure de bandes d'énergie. L'effet laser se produit ici entre les états du bas de la bande de conduction, où se thermalisent les électrons injectés dans cette bande, et les états du sommet de la bande de valence où se thermalisent les trous créés dans cette bande. La condition d'inversion de population entre ces deux ensembles d'états s'écrit (Eq. 9.27) $E_{Fc} - E_{Fv} > E_g$, c'est-à-dire que les pseudo-niveaux de Fermi des électrons et des trous dans le matériau excité, sont respectivement dans les bandes de conduction et de valence. Ceci se traduit par le diagramme énergétique représenté sur la figure (9-36).

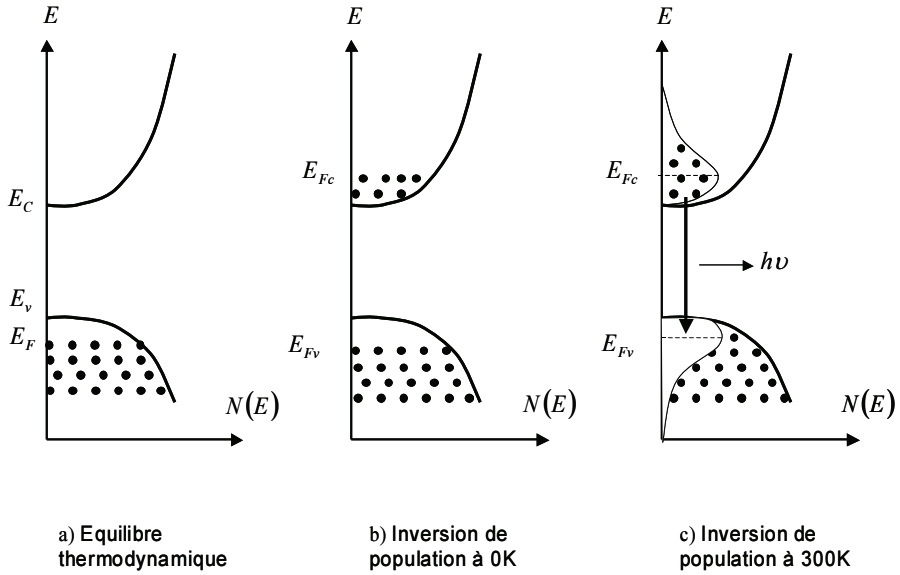


Figure 9-36 : Population des bandes d'énergie.

La deuxième conséquence résultant de la nature pseudo-continue des états dans les bandes permises, est que le rayonnement amplifié est en partie réabsorbé par des transitions intrabandes.

Pour réaliser l'inversion de population, il faut donc créer beaucoup de paires électron-trou dans le matériau. Divers processus tels que le bombardement électronique ou l'excitation optique, peuvent être utilisés. Dans les lasers à semiconducteur, on injecte beaucoup d'électrons dans une région de type p, au moyen d'une jonction pn. La structure du laser est celle d'une diode électroluminescente, mais dont les régions de types n et p sont dégénérées. La région de type p est très dopée pour qu'à l'équilibre, le niveau de Fermi soit dans la bande de valence. La région de type n est très dopée pour que la densité d'électrons injectés dans la région de type p sous l'action de la tension de polarisation, soit telle que le speudo-niveau de Fermi des électrons E_{Fc} soit dans la bande de conduction. En raison du processus d'excitation mis en jeu, les lasers à semiconducteurs sont appelés *lasers à injection* ou *diodes lasers*. La configuration de la jonction est représentée sur la figure (9-37).

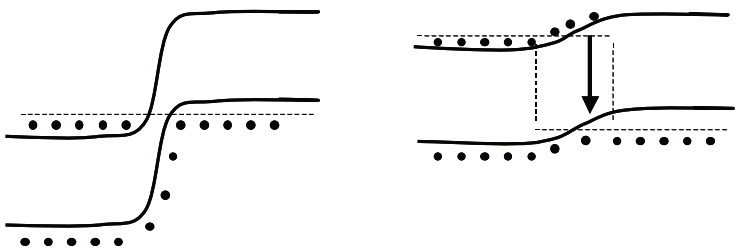


Figure 9-37 : Diagramme énergétique

Il faut noter que, les régions n et p étant très dopées, les extrema des bandes de valence et de conduction sont perturbés par l'étalement des états donneurs et accepteurs et par l'effet d'écran associé à la grande densité de porteurs libres. Il en résulte une renormalisation du gap du matériau qui est alors caractérisé par un gap effectif $E'_g < E_g$ (Par. 2.8.2.b). La longueur d'onde du rayonnement émis par le laser est conditionnée par la valeur du gap effectif E'_g.

Avant d'entrer dans le détail de l'étude du laser à injection, on peut déjà préciser les caractéristiques qui le différencient des lasers conventionnels. Tout d'abord la taille, les dimensions d'un laser à injection se chiffrent en microns alors que celles des lasers conventionnels se chiffrent en décimètres et en mètres. En liaison avec ces faibles dimensions, la puissance émise et les cohérences spatiale et temporelle sont beaucoup plus faibles que dans des lasers conventionnels. Le laser à injection est l'exemple type de conversion directe d'énergie électrique en énergie optique, le rendement est bien meilleur que dans les lasers conventionnels. Enfin, compte tenu du type d'excitation mis en jeu, les lasers à injection se caractérisent par une grande facilité de modulation, ce qui les rend particulièrement bien adaptés aux télécommunications par fibres optiques. Ajoutons en ce qui concerne la longueur d'onde du rayonnement émis, que pratiquement tout le spectre visible et proche infrarouge peut être couvert par la réalisation d'alliages de composés III-V et II-VI.

b. Gain

Le gain, ou coefficient d'amplification, du matériau est défini, comme le coefficient d'absorption (Eq. 9-43), par la variation relative de la densité de rayonnement par unité de longueur, soit

$$g(E) = \frac{1}{\phi(E)} \frac{d\phi(E)}{dx} \qquad (9\text{-}154)$$

La variation du flux de photons au cours d'un trajet de longueur x dans le matériau est obtenue en intégrant l'expression (9-154), soit

$$\phi(E) = \phi_o(E)\, e^{g(E).x} \tag{9-155}$$

$g(E)$ représente indifféremment le coefficient d'amplification et le coefficient d'absorption du matériau. Si $g(E)$ est positif, $d\phi(E)/dx$ est positif, le flux de photons d'énergie E augmente en se propageant dans le matériau . Si $g(E)$ est négatif, $d\phi(E)/dx$ est négatif, le flux de photons d'énergie E diminue en se propageant dans le matériau. Dans ce cas on pose $\alpha(E)=-g(E)$, où $\alpha(E)$ positif représente le coefficient d'absorption (Eq. 9-43).

Le gain $g(E)$ est directement lié au taux d'émission stimulée $r_{st}(E)$. (Par. 9.1.3.b). $r_{st}(E)$ représente le nombre de photons créés par émission stimulée, par unité de volume et unité de temps, de sorte que si $\phi(E)$ représente le flux de photons dans le matériau, $r_{st}(E)$ est donné par

$$r_{st}(E) = \frac{d\phi(E)}{dx}$$

Ainsi l'expression (9-154) s'écrit

$$g(E) = \frac{r_{st}(E)}{\phi(E)} \tag{9-156}$$

Le gain est directement proportionnel au taux d'émission stimulée, de sorte que $g(E)$ est positif si la condition d'inversion de population $\Delta F>E$ est réalisée. Mais cette condition nécessaire n'est pas suffisante pour obtenir une réelle amplification du rayonnement. Il faut que $g(E)$ soit non seulement positif mais de plus, supérieur aux pertes. Ces pertes, qui n'existent pas dans les lasers conventionnels, sont dues au fait qu'un électron de la bande de conduction peut absorber le rayonnement d'énergie $h\upsilon$ pour sauter sur un état d'énergie $E_2=E_1+h\upsilon$ dans la bande (Fig. 9-38).

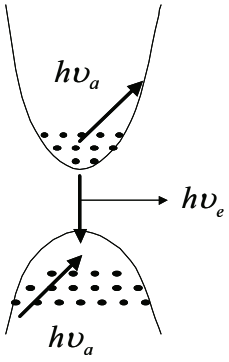


Figure 9-38 : Effet Auger

Ce phénomène qui correspond à une absorption par les porteurs libres, est appelé effet Auger. Il se manifeste aussi dans la bande de valence. On le mesure par un coefficient $\alpha_p(E)$, que l'on appelle *coefficient d'absorption par porteurs libres*. Compte tenu de la présence d'un grand nombre de porteurs libres, nécessaire à la double dégénérescence du matériau, ce coefficient, loin d'être négligeable, est de l'ordre de quelques dizaines de cm^{-1}.

On définit le coefficient net d'amplification du milieu par la différence

$$A(E)=g(E)-\alpha_p(E) \qquad (9\text{-}157)$$

La condition d'émission stimulée s'écrit alors $A(E)>0$. Lorsque cette condition est réalisée, le milieu acquiert de réelles propriétés amplificatrices pour le rayonnement de fréquence $\upsilon=E/h$. Le moindre rayonnement spontané, qui existe au voisinage de la jonction, est alors amplifié pour donner naissance à une émission stimulée. Le seuil d'émission stimulée se traduit par une augmentation brutale du signal lumineux émis par la diode, et l'apparition d'une certaine directivité dans la direction de la plus grande longueur du milieu amplificateur, on dit que la diode est *superradiante*.

En raison de la valeur élevée de l'indice de *réfraction* du semiconducteur ($n\approx 3,5$), nous avons vu que le coefficient de réflexion en incidence normale à l'interface air-semiconducteur était de l'ordre de 30 %. Ainsi, une fraction de l'énergie rayonnée au niveau de la jonction est renvoyée à l'intérieur du matériau. La diode amplificatrice de rayonnement, avec ses deux faces latérales semi-réfléchissantes, joue alors le rôle de cavité résonnante. Le système entre en résonance si l'amplification du rayonnement au cours d'une traversée de la cavité est supérieure aux pertes de cette cavité. La condition de seuil du résonateur consiste à écrire que le coefficient net d'amplification, $A(E)$, est supérieur aux pertes de la cavité. La diode se présente sous la forme d'un parallélépipède rectangle dont deux faces clivées perpendiculaires au plan de la jonction constituent les faces semi-réfléchissantes du résonateur (Fig. 9-39). La cavité se présente donc sous la forme d'un Pérot-Fabry. La longueur L de la diode est typiquement de l'ordre de 300 μm. L'épaisseur de la zone active est conditionnée par la longueur de diffusion des électrons dans la région de type p. En régime d'émission stimulée, la durée de vie radiative des électrons étant réduite par les recombinaisons stimulées, leur longueur de diffusion est faible, l'épaisseur de la zone active est de l'ordre de 0,2 micron.

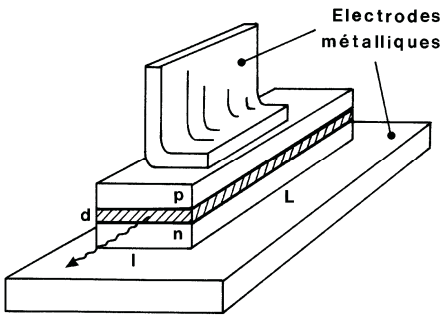


Figure 9-39 : Diode laser

Soit L la longueur de la cavité, $R_{1,2}$ les coefficients de réflexion sur les faces de sortie et $A(E)$ le coefficient net d'amplification de la zone active (Fig. 9-40).

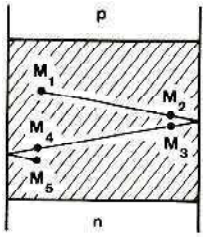


Figure 9-40 : Cavité Laser

Si $\phi(E)$ est le flux de photons au point M_1, il est $\phi(E)\, e^{A(E).L}$ au point M_2, $R_1\phi(E)\, e^{A(E).L}$ au point M_3, $R_1\phi(E)\, e^{2A(E).L}$ au point M_4 et enfin $R_1R_2\phi(E)\, e^{2A(E).L}$ au point M_5, équivalent à M_1. Si le flux en M_5 est égal ou supérieur au flux en M_1, la cavité entre en résonance. La condition de résonance s'écrit donc

$$R_1\, R_2\, e^{2.A(E).L} > 1 \tag{9-158}$$

ou

$$A(E) > \frac{1}{2L}\, Ln\, \frac{1}{R_1\, R_2} \tag{9-159}$$

En explicitant le coefficient net d'amplification et en supposant $R_1=R_2=R$, cette condition s'écrit

$$g(E) > \alpha_p(E) + \frac{1}{L}\, Ln\, \frac{1}{R} \tag{9-160}$$

Supposons d'une part $\alpha_p(E) \approx 70$ cm^{-1}, d'autre part avec $n = 3{,}6$, $R \approx 0{,}3$,de sorte que si $L = 300$ μm les pertes de la cavité sont $1/L\ Ln\ (1/R) \approx 30$ cm^{-1}. Il en résulte que la condition d'oscillation de la cavité est $g(E) > 100$ cm^{-1}.

Pour une injection donnée, le taux d'émission stimulée $r_{st}(E)$ et par suite le gain $g(E)$, sont fonction de l'énergie E du rayonnement considéré. Le fonctionnement de la diode suivant la gamme spectrale considérée est résumé sur la figure (9-41), qui représente le gain $g(E)$, le coefficient d'absorption par porteurs libres α_p qui est sensiblement indépendant de E, et les pertes de la cavité.

- Pour $E < E'_g$, l'énergie du rayonnement est inférieure au gap effectif du matériau, la diode n'émet aucun rayonnement.

- Pour $E > E_0$, g(E) est négatif, la condition d'inversion de population n'est pas satisfaite, la diode émet, à cette énergie, un rayonnement spontané.

- Pour $E'_g < E < E_0$, $g(E)$ est positif, la condition d'inversion de population, est satisfaite. Dans les gammes d'énergie $E < E_1$ et $E > E'_1$, le gain est inférieur aux pertes dues à l'absorption par porteurs libres ($g(E)<\alpha_p$), le gain net est négatif, de sorte que le rayonnement émis est encore spontané. Dans les gammes d'énergie $E_1<E<E_2$ et $E'_2<E<E'_1$, le gain est supérieur aux pertes par absorption ($g(E)>\alpha_p$), le milieu est amplificateur de ce type de rayonnement, la diode émet un rayonnement stimulé caractérisé par une certaine directivité. Enfin dans la gamme $E_2<E<E'_2$, le gain est supérieur aux pertes de la cavité ($g(E)>\alpha_p+1/L\ Ln(1/R)$), la condition d'oscillation est remplie. Le rayonnement émis présente une structure de raies, caractéristique de la cavité Pérot-Fabry.

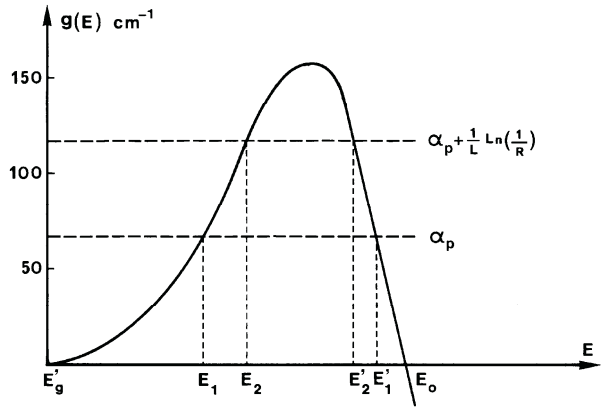


Figure 9-41 : Distributions spectrales du gain et des pertes

c. Distribution spectrale du rayonnement

Le spectre du rayonnement émis est conditionné par le spectre du gain $g(E)$. La figure (9-41) montre que le spectre de l'émission stimulée est beaucoup plus étroit que le spectre de l'émission spontanée, qui s'étend depuis E'_g jusqu'à des énergies beaucoup plus importantes que E_0.

Le taux d'émission stimulée $r_{st}(E)$, est donné en fonction du taux d'émission spontanée $r_{sp}(E)$, par l'expression (9-25)

$$r_{st}(E)=r_{sp}(E)\left(1-e^{(E-\Delta F)/kT}\right) \qquad (9\text{-}161)$$

On obtient la position énergétique de la raie d'émission stimulée par rapport au spectre d'émission spontanée en dérivant cette expression par rapport à l'énergie

$$\frac{dr_{st}(E)}{dE}=\frac{dr_{sp}(E)}{dE}\left(1-e^{(E-\Delta F)/kT}\right)-\frac{1}{kT}r_{sp}(E)e^{(E-\Delta F)/kT} \qquad (9\text{-}162)$$

La position du maximum de la raie d'émission stimulée correspond à $dr_{st}(E)/dE = 0$. L'équation (9-162) donne alors

$$\left(\frac{dr_{sp}(E)}{dE}\right)_{rst\max} = \frac{1}{kT}r_{sp}(E)\frac{1}{e^{(\Delta F-E)/kT}-1} \qquad (9\text{-}163)$$

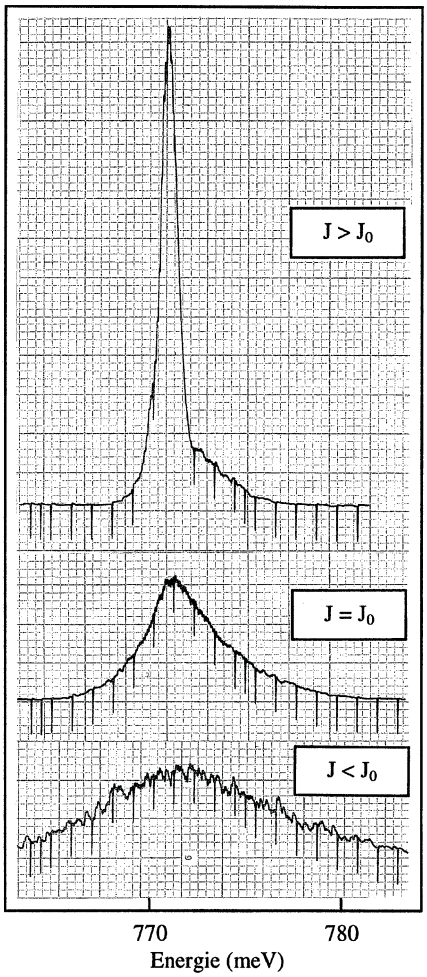


Figure 9-42 : Evolution avec d'injection du spectre d'émission d'un laser au GaSb *(Mathieu H. - 1969)*

L'émission stimulée ne se produit qu'en régime d'inversion, c'est-à-dire lorsque la condition $\Delta F=E_{Fc}-E_{Fv}>E$ est vérifiée. L'expression (9-163) montre alors que $(dr_{sp}(E)/dE)_{rstmax}$ est positif. En d'autres termes, la raie d'émission stimulée se situe sur le flanc basse énergie du spectre d'émission spontanée. D'autre part, quand ΔF augmente c'est-à-dire quand l'excitation augmente, l'exponentielle augmente au dénominateur de l'expression (9-163) et $(dr_{sp}(E)/dE)_{rstmax}$ tend vers zéro. Autrement dit la raie d'émission stimulée se rapproche du sommet du spectre d'émission spontanée. L'évolution du spectre d'émission avec l'excitation est représentée sur la figure (9-42).

La figure (9-43) représente les spectres des taux d'émissions stimulée et spontanée, calculés à différentes températures et pour différents niveaux d'injection, pour une diode au GaSb dont la région de type *p* est dopée avec une concentration d'accepteurs de 10^{19} cm^{-3}. En raison de l'importance du dopage, le gap effectif E'_g est inférieur au gap nominal E_g et la densité d'états au voisinage des extrema des bandes présente une loi de variation différente de $(E-E_g)^{1/2}$, cette loi est sensiblement

exponentielle. Ceci explique pourquoi les taux d'émission sont différents de zéro pour des valeurs négatives de E-E_g . On constate que $r_{st}(E)$ se rapproche d'autant plus de $r_{sp}(E)$ que l'injection est importante et que la température est faible. Ces spectres montrent en particulier qu'à basse température les taux d'émissions spontanée et stimulée sont pratiquement confondus, on réduit donc le seuil d'émission stimulée de la diode en abaissant sa température. A la température ambiante le spectre d'émission stimulée est considérablement plus étroit que le spectre d'émission spontanée, la diode laser émet donc un rayonnement de largeur spectrale beaucoup plus faible que la diode électroluminescente. Le taux d'émission stimulée coupe l'axe des abscisses ($r_{st}(E)$=0) au point $E = \Delta F = E_{Fc}$-E_{Fv}.

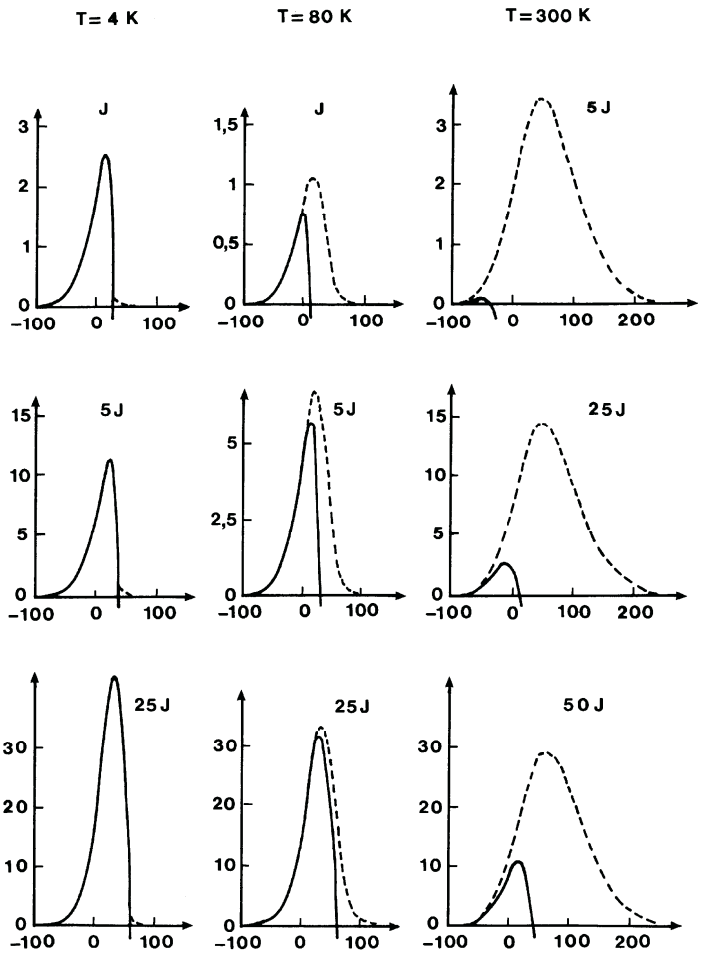


Figure 9-43 : Spectres des taux d'émissions spontanée (traits discontinus) et stimulée (traits continus), d'un laser au GaSb pour différents courants d'excitation à différentes températures. Echelle horizontale E-E_g(meV), où E_g représente le gap nominal. Echelle verticale arbitraire.

Lorsque la condition d'oscillation est remplie dans un intervalle d'énergie à l'intérieur de la raie d'émission stimulée, la cavité sélectionne un certain nombre de modes de résonance définis par la relation

$$2nL=k\lambda \qquad (9\text{-}164)$$

où L est la longueur de la cavité, n l'indice du milieu et k l'ordre d'interférence. L'intervalle entre deux modes successifs est donné par la différentielle de l'expression (9-164)

$$dk=\frac{Ld\lambda}{\lambda^2}2n\left(\frac{\lambda}{n}\frac{dn}{d\lambda}-1\right) \qquad (9\text{-}165)$$

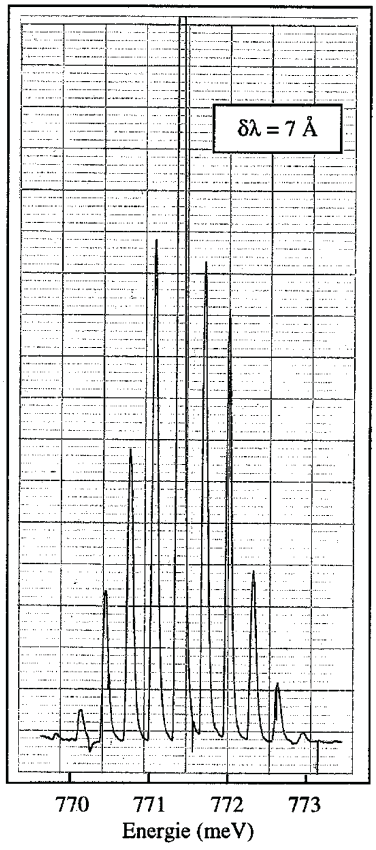


Figure 9-44 : Structure fine de la raie d'émission d'un laser au GaSb. La largeur des modes n'est pas réelle, elle est déterminée ici par la résolution du spectromètre, elle est en fait nettement inférieure à 1Å *(Mathieu H.-1969)*

Entre deux modes successifs dk=-1, de sorte que la distance intermode est donnée par

$$\delta\lambda=\frac{\lambda^2}{2L}\left(n-\lambda\frac{dn}{d\lambda}\right)^{-1} \qquad (9\text{-}166)$$

La structure fine du spectre d'émission d'une diode laser est représentée sur la figure (9-44). L'enveloppe du spectre est la raie d'émission stimulée, représentée sur la figure (9-42). Dans certaines conditions et pour une excitation légèrement supérieure au seuil, la largeur de la raie d'émission stimulée peut n'autoriser qu'un seul mode de résonance de la cavité, l'émission de la diode laser est alors *monomode*. De manière générale pour des excitations supérieures au seuil de la diode cette émission est *multimode*.

Le spectre de la figure (9-44) concerne un laser au GaSb. Ce spectre est centré à la longueur d'onde λ=1,6 μm, il présente une distance intermode $\delta\lambda$=7Å avec une largeur réelle de mode très inférieure à 1Å. Dans le cas d'un laser au GaAs, l'indice de réfraction est n=3,6 et sa dispersion au voisinage du gap est

$dn/d\lambda$=-1 μm^{-1}, la longueur d'onde d'émission est centrée autour de λ=0,8 μm, pour une diode de longueur L=300 μm la distance intermode est $\delta\lambda \approx 2$ Å.

d. Distribution spatiale du rayonnement

La structure de base de la diode est représentée sur la figure (9-45), les dimensions typiques sont L=300 μm, l=10 μm, d<1 μm. La zone active, dont l'épaisseur d est fixée par la longueur de diffusion des électrons dans la région de type p est grisée sur la figure.

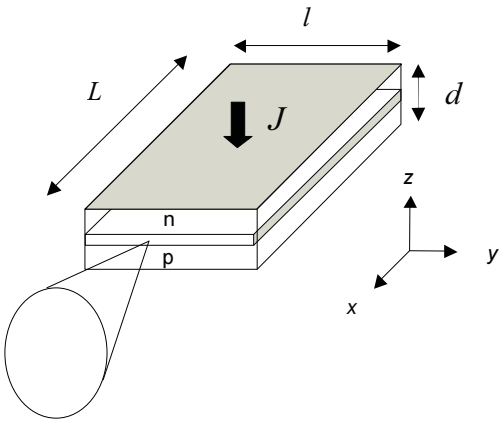


Figure 9-45 : Structure de base de la diode laser

Quand la diode est polarisée au-dessus du seuil, la forte injection d'électrons se traduit par le fait que dans la zone active le gain est supérieur aux pertes et l'indice de réfraction a une valeur légèrement supérieure à sa valeur nominale. Les variations du gain et de l'indice dans la direction normale au plan de la jonction, sont représentées sur la figure (9-46). Le gain, supérieur aux pertes dans la zone active, entraîne une amplification du rayonnement qui est maximum dans la direction x correspondant à la plus grande dimension de la diode. La variation d'indice joue un rôle de guide d'onde et crée un confinement partiel des photons dans la zone active, en réduisant leur étalement dans les régions de pertes.

Sur la face de sortie de la diode, la surface émettrice est donnée par $S=ld$, dans la mesure où l'amplification est homogène dans tout le plan de la jonction. Compte tenu des faibles valeurs de l et d , qui sont comparables à la longueur d'onde du rayonnement émis, l'ouverture du faisceau est conditionnée par la diffraction au niveau de l'ouverture de surface S , que constitue la trace de la zone active sur la face de sortie. L'angle de diffraction θ par une ouverture de largeur e est donné en radian par $\theta=\lambda/e$. Dans la mesure où $\lambda \approx 1\mu m$, $d \approx 1\mu m$ et $l \approx 10\mu m$, l'ouverture du faisceau émis par la diode est de l'ordre de 6° dans le plan de la jonction et 60° dans

le plan perpendiculaire. La distribution spatiale du faisceau est représentée sur la figure (9-46).

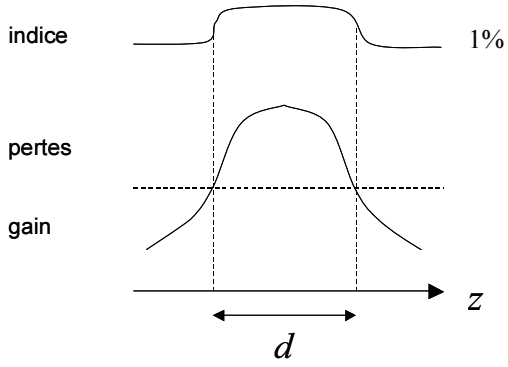


Figure 9-46 : Variations de gain et d'indice

e. Courant de seuil

L'intensité globale du rayonnement émis par la diode est fonction du courant excitateur. L'allure de la courbe de variation de cette intensité est représentée en échelle linéaire sur la figure (9-47). A faible niveau d'injection, l'inversion de population n'est pas réalisée ou est insuffisante pour compenser les pertes par porteurs libres, elle est telle que $g(E)<\alpha_p$ pour toutes les valeurs de l'énergie. L'émission est alors spontanée, le rendement radiatif de la diode est sensiblement constant, l'intensité du rayonnement émis est proportionnelle à la densité de porteurs injectés c'est-à-dire au courant traversant la diode (région *a* de la courbe).

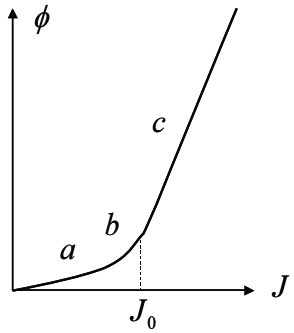


Figure 9-47 : Emission de la diode laser

Quand le courant devient supérieur à la valeur pour laquelle la condition $g(E)>\alpha_p$ est vérifiée pour une énergie E donnée, le gain à cette énergie devient alors supérieur aux pertes par porteurs libres et la diode amplifie le rayonnement

spontané d'énergie E. La courbe d'émission globale présente alors une superlinéarité traduisant la présene d'émission stimulée (région *b* de la courbe).

Enfin quand le courant devient égal à la valeur J_0 pour laquelle $g(E)$ devient supérieur aux pertes globales de la cavité, $g(E)>\alpha_p+1/L\ Ln(1/R)$, la diode oscille, d'abord sur un mode, ensuite sur plusieurs. J_0 est appelé *courant de seuil* de la diode laser. L'émission sur ces modes devient alors prépondérante. La durée de vie des porteurs dans la région d'oscillation diminue à mesure que la densité de photons augmente et toute augmentation de courant est canalisée dans cette région. Tout nouvel électron injecté donne naissance à un photon sur ce ou ces modes, il s'établit un nouveau régime linéaire correspondant à l'oscillation de la cavité (région *c* de la courbe).

Considérons une diode laser excitée par une densité de courant J. En raison du phénomène d'émission stimulée, les populations d'électrons et de photons dans la zone active sont interdépendantes et leurs évolutions sont régies par des équations couplées. Les électrons injectés en $x=x_p$ dans la région de type *p* de la jonction diffusent et se recombinent avec les trous sur une profondeur d qui constitue l'épaisseur de la zone active. L'évolution de la densité n de ces électrons excédentaires résulte du bilan entre les taux d'injection et de recombinaison de ces porteurs. Elle est régie par l'équation de continuité, dans laquelle $g_n=0$. Nous supposerons que le taux d'injection d'électrons est beaucoup plus grand que celui de trous $\gamma_n>>\gamma_p$, et que la zone active est homogène et occupe tout le volume de la cavité. La condition $\gamma_n>>\gamma_p$, qui traduit la relation $j_n(x_p)>>j_p(x_n)$, permet de ramener la densité de courant à l'expression

$$J = j_n(x_p) + j_p(x_n) \approx j_n(x_p)$$

L'hypothèse de la zone active homogène permet de remplacer la variable locale $\partial j_n/\partial x$ par la valeur moyenne $\Delta j_n/d$. D'autre part, si on admet que tous les électrons injectés en $x=x_p$ se recombinent dans la zone active, le courant d'électrons est nul en $x = x_p+d$. La variable locale $\partial j_n/\partial x$ s'écrit donc simplement

$$\frac{\partial j_n}{\partial x} = \frac{\Delta j_n}{d} = \frac{j_n(x_p) - j_n(x_p + d)}{d} = \frac{j_n(x_p)}{d} = \frac{J}{d}$$

L'équation de continuité s'écrit donc

$$\frac{dn}{dt} = \frac{J}{ed} - r \tag{9-167}$$

J est la densité de courant traversant la jonction et d l'épaisseur de la zone active. Le taux de recombinaisons r s'écrit

$$r = R_{sp} + R_{st} = R_{sp} + \sum_i R_{sti}$$

où R_{sp} représente indifféremment le taux de recombinaisons spontanées d'électrons et le taux d'émissions spontanées de photons. Ce taux de recombinaisons s'écrit sous la forme

$$R_{sp} = \frac{n}{\tau_n}$$

où τ_n représente la durée de vie des électrons en régime d'émission spontanée, régime qui régit le fonctionnement d'une diode électroluminescente.

R_{st} représente le taux de recombinaisons stimulées d'électrons. Ce taux résulte d'une somme de termes tels que R_{sti} qui représente indifféremment le taux de recombinaisons d'électrons stimulées par le mode d'oscillation i de la cavité et le taux d'émissions stimulées de photons sur ce mode i. Ce taux de recombinaisons est proportionnel à la densité N_i de photons présents sur le mode i

$$R_{sti} = A_i(n)\, N_i$$

La fonction $A_i(n)$ est proportionnelle au gain du laser qui est évidemment fonction de la densité d'électrons excédentaires dans la zone active. Plusieurs lois de variation de $A_i(n)$ avec n ont été proposées, notamment des lois linéaires, logarithmiques ou de forme plus élaborée. La loi de la forme $A_i(n) = A_i n^{\ell}$ est bien adaptée à la majorité des structures réelles. A basse température (~100K) en particulier, les résultats expérimentaux et les calculs théoriques convergent vers une valeur de l'exposant ℓ de l'ordre de 1. A la température ambiante, la valeur de l'exposant varie entre $\ell \approx 3$ et $\ell \approx 1$ suivant la nature de la structure. Pour des raisons de simplicité nous prendrons $\ell \approx 1$, ce qui permet d'écrire le taux de recombinaisons stimulées sous la forme

$$R_{sti} = A_i\, n\, N_i$$

où A_i(cm^3 s^{-1}) est un coefficient de proportionnalité.

Ainsi l'évolution de la densité n d'électrons excédentaires dans la zone active s'écrit

$$\frac{dn}{dt} = \frac{J}{ed} - \sum_i A_i\, n\, N_i - \frac{n}{\tau_n} \qquad (9\text{-}168)$$

L'évolution de la densité N_i de photons sur le mode d'oscillation i de la cavité résulte du bilan entre les taux de génération et de "recombinaison" des photons

$$\frac{dN_i}{dt} = g_i - r_i$$

Le taux de génération g_i de photons sur le mode *i* résulte de la somme du taux de génération stimulée et de la fraction γ_i sur le mode *i* , du taux global de générations spontanées,

$$g_i = A_i\, n\, N_i + \gamma_i\, \frac{n}{\tau_n}$$

Le paramètre γ_i représente la probabilité pour qu'un photon émis spontanément soit précisément sur le mode *i*.

Le taux de "recombinaison" des photons peut s'écrire sous une forme analogue à celui des électrons

$$r_i = \frac{N_i}{\tau_N}$$

où τ_N représente la durée de vie des photons dans la cavité. Cette durée de vie est conditionnée d'une part par l'absorption de ces photons et d'autre part par leur émission à l'extérieur de la cavité. En d'autres termes, l'inverse de la durée de vie est donnée par le produit de la vitesse des photons dans la cavité par les pertes globales de cette cavité, soit pour une cavité Pérot-Fabry,

$$\frac{1}{\tau_N} = \frac{c}{n}\left(\alpha_p + \frac{1}{L} Ln \frac{1}{R}\right) \qquad (9\text{-}169)$$

où α_p représente le coefficient d'absorption par les porteurs libres et $1/L.Ln1/R$ les pertes propres de la cavité résultant de l'émission des photons à l'extérieur(Eq. 9-160).

L'évolution de la densité de photons sur le mode *i* de la cavité s'écrit donc

$$\frac{dN_i}{dt} = A_i\, n\, N_i + \gamma_i \frac{n}{\tau_n} - \frac{N_i}{\tau_N} \qquad (9\text{-}170)$$

Si la cavité oscille sur *m* modes, on obtient ainsi un système de *m*+1 équations couplées.

$$\frac{dn}{dt} = \frac{J}{ed} - \sum_{i=1}^{m} A_i\, n\, N_i - \frac{n}{\tau_n} \qquad (9\text{-}171\text{-a})$$

$$\frac{dN_i}{dt} = A_i\, n N_i + \gamma_i \frac{n}{\tau_n} - \frac{N_i}{\tau_N} \quad (i = 1\ à\ m) \qquad (9\text{-}171\text{-b})$$

Pour simplifier le calcul, nous supposerons que la diode est monomode, *m*=1. D'autre part, compte tenu des largeurs relatives des spectres d'émission spontanée

(plusieurs centaines d'Angströms) et d'émission stimulée sur un mode (<1 Å), le paramètre γ est de l'ordre de 10^{-5} , on supposera γ=0. Le système de m+1 équations se réduit alors à un système de 2 équations,

$$\frac{dn}{dt} = \frac{J}{ed} - A\,n\,N - \frac{n}{\tau_n} \tag{9-172-a}$$

$$\frac{dN}{dt} = A\,n\,N - \frac{N}{\tau_N} \tag{9-172-b}$$

En régime stationnaire, le courant de polarisation est constant et par suite n et N sont constants, le système précédent s'écrit

$$\frac{J}{ed} - A\,n\,N - \frac{n}{\tau_n} = 0 \tag{9-173-a}$$

$$A\,n\,N - \frac{N}{\tau_N} = 0 \tag{9-173-b}$$

d'où la relation

$$\frac{n}{\tau_n} + \frac{N}{\tau_N} = \frac{J}{ed} \tag{9-173-c}$$

- Si la diode est polarisée par un courant constant inférieur au courant de seuil, $J < J_0$, elle n'émet qu'un rayonnement spontané, la densité de photons par mode est négligeable, l'équation (9-173-c) donne

$$n = \frac{J\,\tau_n}{ed} \quad \rightarrow \quad R_{sp} = \frac{n}{\tau_n} = \frac{J}{ed} \tag{9-174}$$

La durée de vie des porteurs excédentaires est constante, c'est la durée de vie spontanée, la densité de ces porteurs augmente linéairement avec le courant, ainsi que le taux d'émission spontanée (Fig. 9-47-a).

- Lorsque la diode est polarisée par un courant égal au courant de seuil, J=J_0, la densité d'électrons excédentaires est donnée par

$$n_o = \frac{J_o\,\tau_n}{ed} \tag{9-175-a}$$

D'autre part le gain dans la région active compense les pertes globales de la cavité et le nombre de photons sur le mode d'oscillation devient différent de zéro, l'équation (9-173-b) donne

$$n_o = \frac{1}{A\,\tau_N} \tag{9-175-b}$$

Les expressions (9-175-a et b) permettent de calculer le courant de seuil

$$J_o = \frac{ed}{A\tau_n\tau_N} = \frac{ed}{A\tau_n}\frac{c}{n}\left(\alpha_p + \frac{1}{L}Ln\frac{1}{R}\right) \tag{9-176}$$

le courant de seuil de la diode est donc inversement proportionnel aux durées de vie des électrons et des photons dans la cavité.

- Lorsque la diode est polarisée par un courant supérieur au courant de seuil, $J > J_0$, la diode oscille sur le mode sélectionné par la cavité, la densité N de photons sur ce mode devient importante et la durée de vie des électrons excédentaires diminue en raison du phénomène d'émission stimulée. La densité d'électrons excédentaires n'augmente plus et reste égale à sa valeur de seuil n_0. La densité de photons sur le mode d'oscillation est alors donnée par l'équation (9-173-a) dans laquelle $n=n_0$. On obtient, compte tenu de l'expression (9-175-a)

$$N = \frac{1}{A\tau_n}\left(\frac{J}{J_o} - 1\right) = \frac{\tau_N}{ed}\left(J - J_o\right) \tag{9-177}$$

La densité de photons sur le mode d'oscillation, et par suite l'intensité du rayonnement émis sur ce mode, augmentent donc proportionnellement à l'excès de courant par rapport au courant de seuil. On retrouve un nouveau régime linéaire (Fig. 9-47-c).

Il faut toutefois noter que dans la pratique, plusieurs phénomènes perturbent ce type de fonctionnement idéal. Le régime monomode n'est généralement obtenu que pour des courants de polarisation voisins du courant de seuil J_0 , au-delà la diode laser fonctionne très vite en régime multimode. En outre nous avons supposé la zone active homogène et occupant tout le volume de la cavité, ceci est pratiquement irréalisable. La zone d'inversion est généralement de type filamentaire, les différents filaments opérant indépendamment les uns des autres.

f. Fréquence de coupure

La méthode la plus directe pour imprimer une information sur le rayonnement émis par une diode, électroluminescente ou laser, est de moduler l'amplitude de ce rayonnement par la variation du courant excitateur. Cette méthode de modulation analogique est particulièrement bien adaptée aux diodes lasers pour deux raisons. La première est liée au fait que le rayonnement émis par la diode laser est proportionnel au courant injecté dans la diode sur une gamme étendue de courant ; la conversion des variations de courant en modulation du rayonnement est linéaire.

La deuxième raison réside dans le fait que la fréquence de coupure de la diode est conditionnée par la durée de vie des porteurs minoritaires injectés dans la zone active de la jonction. ***Dans une diode électroluminescente cette durée de vie est régie par les recombinaisons spontanées et est de l'ordre de 1ns. Dans une diode laser, le régime d'inversion, en favorisant l'émission stimulée de photons c'est-à-dire la recombinaison stimulée d'électrons, réduit considérablement la durée de vie des porteurs. Ainsi la fréquence de coupure de la modulation d'amplitude d'une diode laser peut être très élevée.***

Supposons la diode laser polarisée par un courant continu J supérieur au courant de seuil J_0 sur lequel est superposée une faible composante alternative ΔJ de pulsation ω. Dans la mesure où l'amplitude de modulation ΔJ est suffisamment faible, $\Delta J << J-J_0$, la diode reste en régime d'inversion et les densités d'électrons et de photons peuvent être linéarisées sous la forme

$$n' = n + \Delta n e^{j\omega t}$$
$$N' = N + \Delta N e^{j\omega t}$$

où Δn et ΔN sont respectivement les amplitudes complexes des densités d'électrons excédentaires et de photons présents dans la zone active.

En portant ces expressions dans les équations (9-172-a et b), en négligeant le terme du deuxième ordre, et compte tenu des relations (9-173-a et b) on obtient le système d'équations

$$\left(\frac{1}{\tau_n} + AN + j\omega\right)\Delta n + An\Delta N = \frac{\Delta J}{ed} \qquad \text{(9-178-a)}$$

$$AN\Delta n - \left(\frac{1}{\tau_N} - An + j\omega\right)\Delta N = 0 \qquad \text{(9-178-b)}$$

ce qui permet d'écrire la relation

$$\left(\frac{1}{\tau_n} + j\omega\right)\Delta n + \left(\frac{1}{\tau_N} + j\omega\right)\Delta N = \frac{\Delta J}{ed} \qquad \text{(9-178-c)}$$

Les équations (9-178-a et b) permettent de calculer l'expression de ΔN en fonction de ΔJ à la fréquence de modulation $f = \omega/2\pi$, et par suite la profondeur de modulation $\Delta N/N$.

En explicitant N à partir de l'expression (9-177) et $n=n_0$ à partir de l'expression (9-175-b), les équations (9-178-a et b) s'écrivent, en multipliant la deuxième par $1/\tau_N$

$$\left(\frac{1}{\tau_n}\frac{J}{J_o}+j\omega\right)\Delta n+\frac{\Delta N}{\tau_N}=\frac{\Delta J}{ed} \tag{9-179-a}$$

$$\frac{1}{\tau_n\tau_N}\left(\frac{J}{J_o}-1\right)\Delta n-j\omega\frac{\Delta N}{\tau_N}=0 \tag{9-179-b}$$

En posant

$$\omega_o=\left(\frac{1}{\tau_n\tau_N}\left(\frac{J}{J_o}-1\right)\right)^{1/2} \qquad \beta=\frac{1}{\tau_n}\frac{J}{J_o} \tag{9-180}$$

Le système (9-179) s'écrit

$$(\beta+j\omega)\Delta n+\frac{\Delta N}{\tau_N}=\frac{\Delta J}{ed} \tag{9-181-a}$$

$$\omega_o^2\,\Delta n-j\omega\frac{\Delta N}{\tau_N}=0 \tag{9-181-b}$$

La résolution de ce système donne les amplitudes complexes des modulations Δn et ΔN des densités d'électrons et de photons à la pulsation ω.

$$\Delta n=\frac{\Delta J}{ed}\,\frac{1}{\omega_o}\,\frac{1}{\beta/\omega_o+j(\omega/\omega_o-\omega_o/\omega)} \tag{9-182-a}$$

$$\Delta N=\frac{\Delta J}{ed}\,\omega_o^2\;\tau_N\,\frac{1}{\omega_o^2-\omega^2+j\beta\omega} \tag{9-182-b}$$

Compte tenu des expressions respectives de la densité de photons à l'équilibre N (Eq. 9-177), de la pulsation ω_o (Eq. 9-180) et du courant de seuil J_0 (Eq. 9-176), la profondeur de modulation $\Delta N/N$ s'écrit

$$\frac{\Delta N}{N}=\frac{1}{\tau_n\,\tau_N}\,\frac{\Delta J/J_o}{\omega_o^2-\omega^2+j\beta\omega} \tag{9-183-a}$$

ou

$$\frac{\Delta N}{N}=\frac{1}{\tau_n\,\tau_N}\,\frac{\Delta J/J_o}{\omega_o^2}\,\frac{1}{1-\dfrac{\omega^2}{\omega_o^2}+j\dfrac{\beta}{\omega_o}\dfrac{\omega}{\omega_o}} \tag{9-183-b}$$

En posant $\gamma=\dfrac{\beta}{\omega_o}=\left(\dfrac{\tau_N\,(J/J_o)^2}{\tau_n\;J/J_o-1}\right)^{1/2}$ et $A_o=\dfrac{1}{\tau_n\tau_N}\dfrac{\Delta J/J_o}{\omega_o^2}=\dfrac{\Delta J}{J-J_o}$

l'amplitude de la profondeur de modulation $A=\left|\Delta N/N\right|$ s'écrit

$$A=\frac{A_o}{\left((1-\frac{\omega^2}{\omega_o^2})^2+\gamma^2\frac{\omega^2}{\omega_o^2}\right)^{1/2}} \tag{9-184}$$

Cette profondeur de modulation varie donc de $A\approx A_0$ pour $\omega<<\omega_0$ à $A = A_0\omega_0^2/\omega^2$ pour $\omega>>\omega_0$ en passant par $A=A_0/\gamma$ pour $\omega=\omega_0$. Il existe une fréquence de résonance voisine de $\omega_0/2\pi$ pour laquelle A passe par un maximum $A_m\approx A_0/\gamma$ et au-delà de laquelle A décroît comme ω^{-2}. La fréquence de résonance est donnée par

$$f_o\approx\frac{1}{2\pi}\left(\frac{1}{\tau_n\tau_N}(\frac{J}{J_o}-1)\right)^{1/2} \tag{9-185}$$

Cette fréquence de résonance qui résulte de l'interdépendance des populations d'électrons et de photons dans la zone active est d'autant plus élevée que le courant de polarisation est supérieur au courant de seuil et que les durées de vie des électrons et des photons sont faibles. Dans une diode laser la durée de vie des électrons en régime spontané τ_n est de l'ordre de 10^{-9}s. La durée de vie des photons τ_N est donnée par l'expression (9-169) dans laquelle $\alpha_p \approx 60$ cm^{-1}, $R \approx 30\%$ et $L \approx 300$ μm. Il en résulte un coefficient global de pertes de l'ordre de 100 cm^{-1} et par suite une durée de vie des photons $\tau_N \approx 10^{-12}$ s. Ainsi pour un courant de polarisation J supérieur de 10% au courant de seuil J_0, on obtient une fréquence de résonance $f_0 \approx 1{,}6$ GHz. En outre, $A_0 = 10\,\Delta J/J$, $\gamma = 1/10$ et $A_m = 10A_0$.

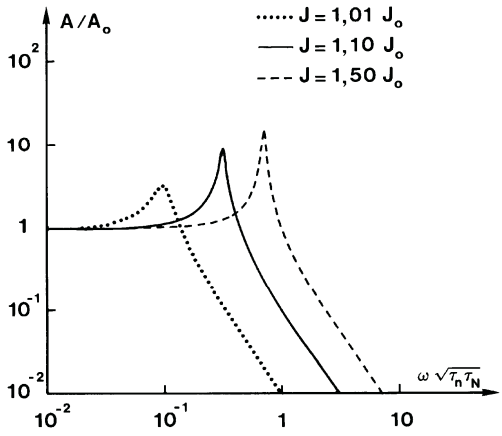


Figure 9-48 : Profondeur de modulation en fonction de la pulsation pour des courants supérieurs au courant de seuil

Les courbes représentant les variations du rapport A/A_0 en fonction de $\omega\sqrt{\tau_n \tau_N}$ pour différentes valeurs du courant de polarisation sont portées sur la figure (9-48). Les fréquences de résonance correspondantes sont respectivement $f_0 = 0{,}5$, 1,6 et 3,5 GHz pour $J = 1{,}01$, 1,1 et 1,5 J_0. Il faut noter que la fréquence de coupure de la diode laser est bien supérieure à celle d'une diode électroluminescente (Par. 9.3.1.d). Ceci résulte à travers le phénomène d'émission stimulée, de la faible durée de vie des électrons dans la cavité.

g. Structure d'une diode laser

La structure réelle d'une diode laser est plus compliquée qu'une simple jonction pn, le but poursuivi étant de minimiser le courant de seuil. Deux voies permettent de progresser dans ce sens, confiner le courant sur une faible section, confiner les photons dans un petit volume.

On optimise ces paramètres en réalisant une double hétérojonction (DH), dont la structure est représentée sur la figure (9-49) pour une diode laser au GaAs. La couche active de type p, dont l'épaisseur est de l'ordre de 0,1 à 0,3 μm, est prise en sandwich entre deux couches de $Ga_{0,7}Al_{0,3}As$, dopées respectivement n et p. Ce sandwich confine dans la région active, à la fois les électrons et les photons.

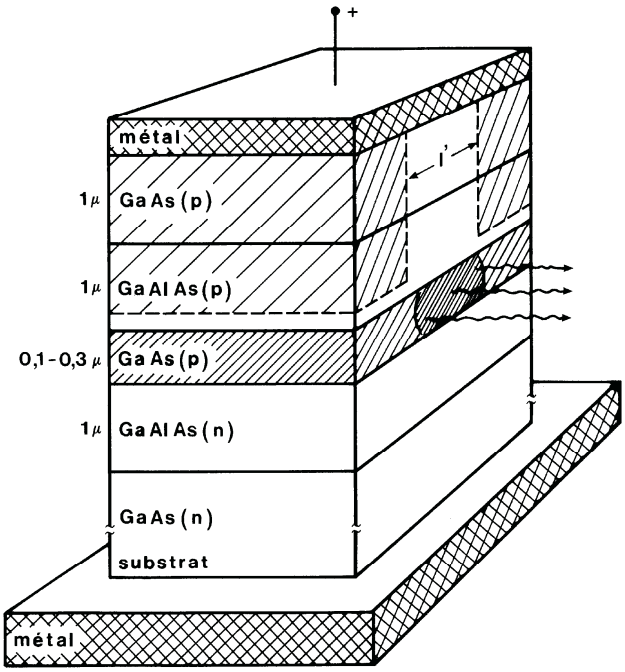


Figure 9-49 : Structure d'une diode laser au GaAs

Les photons sont confinés dans la couche active de GaAs (p) par la différence d'indice qui existe entre GaAs et GaAlAs. Les réflexions totales à l'interface des deux matériaux évitent l'étalement des photons dans le GaAlAs. L'efficacité du confinement est plus grande que dans une homojonction en raison du fait que l'écart d'indice est plus important.

Les électrons sont confinés dans la couche active en raison de la différence de gap qui existe entre $Ga_{0,7}Al_{0,3}As$ ($E_g \approx 1{,}9$ eV) et GaAs ($E_g \approx 1{,}4$ eV). Les deux couches de GaAlAs constituent des barrières de potentiel qui empêchent les électrons et les trous de diffuser au-delà du GaAs.

Les variations de l'indice de réfraction et de la structure de bande dans la diode polarisée, sont représentées sur la figure (9-50).

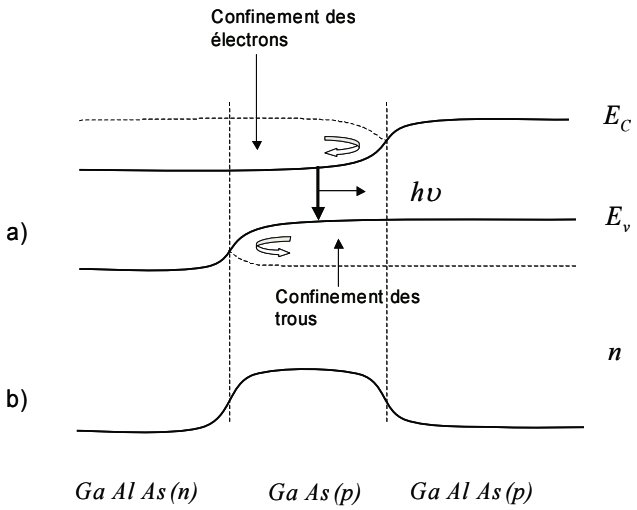


Figure 9-50 : Diagramme énergétique et variation d'indice

On diminue encore le courant de seuil, en réduisant la section conductrice de la diode à un canal de section rectangulaire de longueur L et de largeur $l'<1$, en créant une zone de forte résistivité dans les couches de GaAs (p) et GaAlAs (p), par bombardement avec des protons. Le bombardement par les protons, en détruisant partiellement le réseau cristallin, crée une zone à faible mobilité et par suite à faible conductivité. La zone protégée reste très conductrice.

Le système GaAs-GaAlAs a été jusqu'ici le plus étudié et le plus utilisé pour réaliser des diodes lasers, pour différentes raisons. Le GaAs est un semiconducteur à gap direct que l'on peut doper n et p facilement. Le composé ternaire GaAlAs peut être fabriqué sur une gamme étendue de composition et présente avec GaAs un désaccord de maille très faible (~ 0,1 %) pour toutes les valeurs de x. L'accord de maille entre les constituants d'une structure multicouche est un paramètre important car il conditionne l'absence de contrainte au niveau des interfaces, ces contraintes

créant des centres de recombinaison non radiative. Enfin les valeurs relatives des gaps et des indices de GaAs et GaAlAs créent un bon confinement des électrons et des photons.

La diode laser du type GaAs - GaAlAs émet un rayonnement centré à λ=0,9 μ, or les télécommunications par fibres optiques nécessitent l'utilisation de diodes lasers émettant dans la gamme spectrale 1,1-1,6 μ, où les fibres optiques actuelles présentent un minimum d'absorption et de dispersion (Par. 9.1.6.c). Ces diodes lasers sont réalisées à partir du composé quaternaire $Ga_xIn_{1-x}P_yAs_{1-y}$. Un choix judicieux des paramètres x et y permet de réaliser un émetteur à 1,1-1,3 μ, avec un accord de maille parfait sur un substrat de InP, et par suite d'obtenir une hétérostructure libre de toute contrainte.

h. Laser à puits quantiques

La réduction de volume de la zone d'inversion est obtenue par la réalisation de lasers à puits quantiques. Le puits quantique présente le double intérêt d'une part de réduire l'extension spatiale des électrons et des trous, et d'autre part de confiner leurs distributions énergétiques par la nature bidimensionnelle de la densité d'états.

Le paramètre qui joue un rôle essentiel dans ces structures à puits quantiques est le facteur de confinement Γ qui mesure le taux de recouvrement des distributions spatiales des porteurs et des photons, c'est-à-dire le taux de recouvrement de la zone active et du rayonnement.

Dans les lasers à double hétérostructure (DH), ce facteur est voisin de 1 car les mêmes barrières, barrières de potentiel pour les uns et d'indice pour les autres, confinent à la fois les porteurs et les photons. En fait, dans ces structures la distance entre les barrières est typiquement de l'ordre de 0,2 μm, c'est-à-dire voisine des longueurs d'ondes optiques mais très supérieure aux longueurs d'ondes électroniques. Il en résulte un bon confinement des photons mais pas un confinement optimisé des porteurs.

Dans une structure à puits quantique la situation est inversée. Le confinement quantique des porteurs nécessitant des largeurs de puits de quelques dizaines de nanomètres, la largeur de la zone active est alors comparable aux longueurs d'ondes électroniques mais très inférieure aux longueurs d'ondes optiques. Il en résulte un confinement optimisé des porteurs, mais pas des photons. Le facteur de confinement est alors faible.

On combine le double confinement des porteurs et des photons par la réalisation de structures à confinements séparés SCH (Separate Confinement Heterostructure). Les porteurs sont confinés dans un puits de potentiel et les photons sont confinés par des barrières de variation d'indice. Le confinement optique est ici optimisé par la réalisation de guides d'ondes à indice graduel GRIN (GRaded INdex), de part et d'autre du puits de potentiel. Ces structures portent le nom de structures GRIN-SCH (Fig. 9-51).

Il faut en outre noter que le rayonnement émis par un laser à puits quantique est déplacé vers les courtes longueurs d'onde par rapport au rayonnement émis par un laser équivalent à double hétérostructure. Ce déplacement est dû aux énergies de confinement des porteurs (Fig. 9-51). Enfin le simple puits quantique peut être remplacé par des multipuits quantiques dans le but d'augmenter le facteur de confinement (Par. 12.6).

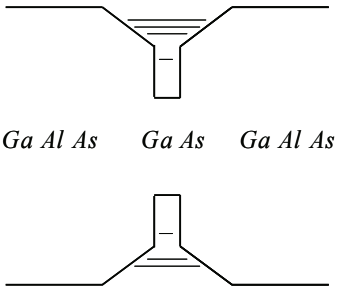


Figure 9-51 : Laser à puits quantique

9.4. Applications des composants optroniques

9.4.1. *Disques optiques - CD-DVD*

Le stockage optique de l'information s'effectue sur des disques de dimension relativement réduite dont l'appellation est désormais consacrée, il s'agit des CD (Compact Disks) et des DVD (Digital Video Disks ou Digital Versatile Disks). Le principe de ce type de stockage consiste à exploiter la réponse optique contenue dans la composante réfléchie d'un faisceau laser focalisé sur une zone du disque. Le paramètre qui définit l'état binaire de la zone explorée est, suivant l'effet physique exploité, l'intensité du faisceau réfléchi ou sa polarisation. Dans le premier cas le disque est tout-optique, dans le second il est magnéto-optique.

Le point commun aux CD et aux DVD est que chacun d'eux présente deux variantes, une variante *préenregistrée* ROM (Read Only Memory) utilisable en lecture uniquement, il s'agit des CD-audio, CD-ROM, DVD-ROM, DVD-video …, et une variante *inscriptible* R (Recordable) ou *réinscriptible* RW (ReWritable) sur lesquels l'utilisateur peut stocker lui-même ses informations, il s'agit des CD-R, CD-RW, DVD-RAM, … Dans la première variante l'information est stockée une fois pour toutes lors de la fabrication du disque, sous forme d'une série d'alvéoles imprimées dans une sillon gravé en spirale sur une couche réfléchissante d'aluminium déposée sur un substrat plastique de polycarbonate. Une couche acrylique de protection moule l'information et reçoit le label du disque. La lecture est effectuée avec un laser de quelques mW qui parcourt le sillon et détecte la différence de réflectivité des diverses zones. La seconde variante permet non

seulement de lire mais aussi de stocker des informations à l'aide du rayonnement, la puissance du laser est alors de l'ordre de quelques dizaines de mW.

La différence la plus importante entre les CD et les DVD se situe au niveau de la densité de stockage, qui est directement liée à la longueur d'onde du rayonnement émis par le laser utilisé. La densité de stockage est en effet déterminée par la taille du spot lumineux sur le disque, or cette taille est limitée du coté des faibles dimensions par les phénomènes de diffraction qui sont eux-mêmes conditionnés par la longueur d'onde du rayonnement. Il en résulte que la densité de stockage sur un disque varie comme $1/\lambda^2$ où λ est la longueur d'onde du faisceau laser. Les CD utilisent l'émission infrarouge (λ=780 nm) de lasers à AlGaAs alors que les DVD utilisent l'émission rouge (λ=650 nm) de lasers à AlGaInP, la nouvelle génération de DVD à haute densité HD-DVD utilise l'émission bleue (λ=410 nm) de lasers à GaN. Le pas du sillon gravé en spirale sur le disque est de 0,74 μm dans un DVD contre 1,6 μm dans un CD, quant à la largeur minimum du sillon elle est de 0,4 μm dans un DVD comparée à 0,8 μm dans un CD. Dans un HD-DVD ces paramètres sont encore diminués par l'utilisation de lasers bleus, la largeur des sillons est réduite à 0,24 μm. Notons que les DVD ont en outre une plus grande vitesse de rotation et utilisent un procédé de compression des données binaires. Enfin les CD sont des disques simple face alors que les DVD sont simple ou double face. Les CD-ROM ont une capacité de stockage de 650 Mo alors que les DVD-ROM ont une capacité de 4,7 Go par face, les HD-DVD ont une capacité de 15 Go par face. Notons que de nouvelles technologies envisagent d'augmenter encore la capacité en superposant plusieurs couches actives par face et en stockant les données non seulement dans le sillon mais aussi dans l'espace entre sillons.

Le principe du dispositif de lecture d'un disque optique est représenté sur la figure (9-52).

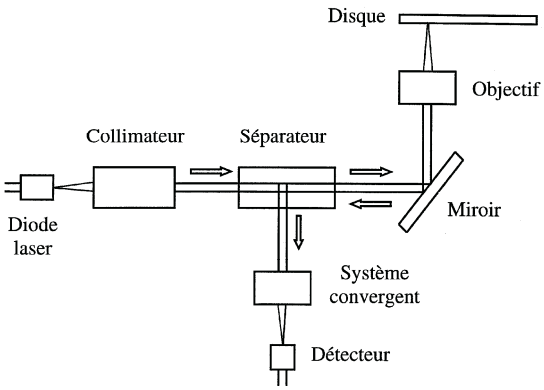


Figure 9-52 : CD et DVD

Ce dispositif est constitué d'un laser à semiconducteur dont le faisceau traverse un collimateur puis un séparateur avant d'être focalisé sur le disque. Le faisceau réfléchi est renvoyé vers un détecteur par le séparateur. Lorsque le faisceau laser éclaire la surface du disque en suivant le sillon il reçoit des réponses spécifiques des zones explorées. Le lecteur enregistre ces spécificités, dont la nature dépend du processus d'enregistrement, comme données binaires. L'information est contenue dans l'intensité du faisceau réfléchi ou dans sa polarisation suivant que le disque est tout-optique ou magnéto-optique.

Le dispositif d'écriture est différent suivant qu'il s'agit des CD-R ou des CD-RW. Dans les CD-R, sur lesquels on ne peut écrire qu'une seule fois, une couche de colorant est excitée par le laser d'écriture qui imprime de manière irréversible des zones de faible réflectivité. Ces disques servent essentiellement à un archivage, comme par exemple le digitalisation de photographies ou de tableaux. Sur les CD-RW la couche active est constituée d'un matériau qui peut basculer de manière réversible d'un état stable à un autre sous l'action du faisceau laser, ces deux états pouvant être distingués au moyen de ce même faisceau laser. Cette couche active est insérée entre deux couches diélectriques de quelques dizaines de nanomètres qui jouent un double rôle de protection et d'ajustement des conditions optiques. Deux types de couche mettant à profit des propriétés physiques différentes sont utilisées, les couches magnéto-optiques et les couches à changement de phase.

Dans l'enregistrement magnéto-optique les matériaux utilisés sont des alliages amorphes Terre Rare - Métal de Transition RE-TM (Rare Earth-Transition Metal) tels que TbFeCo. Ces matériaux présentent une aimantation perpendiculaire aux couches qu'il est possible d'orienter par application d'un champ magnétique à haute température. A la température ambiante le champ coercitif de ces matériaux est très important de sorte qu'ils conservent indéfiniment leur aimantation. Lorsqu'on élève leur température au voisinage de la température de Curie le champ coercitif tend vers zéro, un faible champ magnétique est alors suffisant pour réorienter les spins électroniques. Le dispositif d'écriture associe un laser et une tête magnétique en regard l'un de l'autre par rapport à la surface du disque. Le laser chauffe la zone à imprimer vers 300 °C et la tête magnétique fixe le sens de l'aimantation par un champ de quelques centaines de Gauss. La lecture se fait avec le même laser que l'écriture, à plus faible puissance, en exploitant l'effet Kerr magnéto-optique. Les différentes zones du disque font tourner la polarisation du faisceau réfléchi dans un sens ou dans l'autre suivant leur aimantation. Un système différentiel permet ensuite de détecter le signe de cette rotation.

Dans l'enregistrement par changement de phase la couche active est constituée d'un matériau à changement de phase à base de tellure comme GeSbTe ou AgInSbTe. La couche est initialement polycristalline et par suite réfléchissante. Le processus d'écriture consiste à chauffer la zone à imprimer au-dessus du point de fusion avec le spot laser (environ 500-700 °C), ceci nécessite une puissance de laser de l'ordre de quelques dizaines de mW. Un refroidissement rapide laisse alors cette zone dans un état amorphe dont le pouvoir réflecteur est très inférieur à celui des zones cristallines. Cette trempe rapide est obtenue grace à une couche métallique

qui joue le rôle de puits thermique, la vitesse de refroidis-sement est voisine de 1 °C/ns. Pour la réécriture le laser chauffe préalable-ment les zones amorphes vers 200 °C pendant un temps suffisant pour permettre la recristallisation du matériau. La lecture se fait avec le même laser, à plus faible puissance, et exploite les différences de réflectivité des zones cristallines et amorphes. La structure de principe est représentée sur la figure (9-53).

La technique magnéto-optique est plus performante que la technique du changement de phase en ce qui concerne le débit d'écriture, ceci résulte du fait qu'elle met à profit un processus purement électronique, le renversement des spins, alors que la seconde nécessite un changement structural. Par contre en ce qui concerne la capacité de stockage, l'évolution nécessaire vers des lasers de courte longueur d'onde rend plus performante la technique du changement de phase. En effet cette dernière ne présente aucune limitation spécifique alors que la technique magnéto-optique est limitée par le fait que l'angle de rotation Kerr des alliages utilisés diminue très vite avec la longueur d'onde. Cette rotation devient difficilement détectable dans le vert et le bleu. La solution dans le cadre de cette technologie passe par l'utilisation d'autres matériaux magnétiques, plus efficaces aux courtes longueurs d'onde.

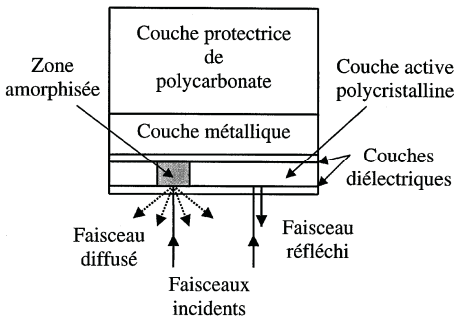


Figure 9-53 : CD à changement de phase

9.4.2. *Diagramme de chromaticité*

La réponse que nous faisons à une excitation lumineuse est le fruit d'un processus neurophotochimique complexe qui fait intervenir la physique au niveau du rayonnement, la physiologie au niveau de l'œil et le psychique au niveau de notre interprétation du signal perçu. Alors que la gamme spectrale visible du rayonnement électromagnétique est bien définie (380-780 nm) la couleur d'un rayonnement est quant à elle beaucoup plus difficile à appréhender car ce qui se passe au niveau de la rétine, et surtout après, nous est en grande partie inconnu. Si la longueur d'onde d'un rayonnement monochromatique se révèle être un paramètre en nette corrélation avec la sensation de couleur, il faut néanmoins remarquer qu'il

n'existe aucun rayonnement monochromatique qui soit rose, marron ou tout simplement blanc. D'autre part aucun raisonnement de physique ne permet de comprendre pourquoi lorsque notre oreille est soumise à deux émissions audibles de fréquences différentes elle perçoit deux sons, alors que lorsque notre œil reçoit simultanément des lumières monochromatiques rouge et verte il perçoit une couleur unique, en l'occurrence le jaune. La mesure de la couleur, qui fait intervenir des disciplines aussi différentes que la physique, la physiologie et le psychique, porte le nom de *colorimétrie*.

C'est au niveau de la rétine que se manifeste la conversion d'énergie lumineuse en signal nerveux transmis au cerveau par les récepteurs que sont les *bâtonnets* et les *cônes*. Les bâtonnets sont le siège d'un pigment organique, la rhodopsine, et sont les artisans de la vision scotopique ou nocturne, où l'impression de couleur est pratiquement absente. La sensation de couleur est due aux cônes, qui sont les artisans de la vision photopique ou diurne. Ces derniers renferment trois sortes de pigments organiques, le chlorolabe, l'érythrolabe et le cyanolabe, qui présentent des courbes de sensibilité centrées respectivement à 440, 530 et 570 nm correspondant aux couleurs bleue, verte et rouge. L'existence de ces trois types de pigment dans les photorécepteurs des cônes sert de base physiologique au *modèle trichromatique*, ou *trivariance visuelle*, de la perception des couleurs. Depuis fort longtemps les teinturiers et les coloristes ont su reproduire une grande variété de nuances par le mélange de teintes de base convenablement choisies. Les teintes de base universellement utilisées dans l'industrie graphique sont le cyan (bleu-vert), le jaune et le magenta (rouge-pourpre). Dans le domaine de l'affichage, les écrans de visualisation reproduisent tout le spectre des couleurs par les émissions bleues, vertes et rouges de trois types de luminophore. Ces trois couleurs de base, dont on voit qu'elles ne sont pas universelles mais relatives à un processus déterminé, sont appelées *couleurs primaires*. L'association française de normalisation (AFNOR) a défini le principe de trivariance visuelle de la manière suivante : *Un rayonnement de couleur quelconque peut être reproduit visuellement à l'identique par le mélange algébrique, en proportions définies de manière unique, des flux lumineux de trois rayonnements qui peuvent être arbitrairement choisis, sous réserve qu'aucun d'eux ne puisse être reproduit par un mélange des deux autres*. Remarquons que ce principe de trivariance n'est en aucun cas une propriété physique du rayonnement électromagnétique mais traduit en fait une propriété physiologique de notre œil.

Si on définit trois stimulus de référence, de couleurs C_1, C_2, C_3, la décomposition dans cette base de tout stimulus de couleur s'écrit formellement

$$C = \alpha C_1 + \beta C_2 + \gamma C_3$$

Les paramètres α, β, γ représentent les quantités des trois couleurs de référence nécessaires à la constitution de la couleur C. Ils définissent les *composantes trichromatiques* de C dans l'espace colorimétrique engendré par C_1, C_2, C_3

Triangle des couleurs

La recherche d'une métrique permettant d'exprimer une *mesure* de la couleur conduit à la conception d'un espace vectoriel tridimensionnel engendré par la base ($\vec{C}_1, \vec{C}_2, \vec{C}_3$), la couleur *C* est représentée par un vecteur $\vec{C}$ donné par

$$\vec{C} = \alpha \vec{C}_1 + \beta \vec{C}_2 + \gamma \vec{C}_3$$

L'extrémité du vecteur $\vec{C}$ détermine un *point couleur*, son orientation définit la *chromaticité*, sa longueur la *luminosité* (Fig. 9-54). La chromaticité ne dépend donc que des valeurs relatives des composantes α, β, γ de sorte que les unités de longueur des trois axes $\vec{C}_1, \vec{C}_2, \vec{C}_3$ peuvent être choisies de manière arbitraire. On choisit ces unités pour que la couleur reproduite soit le blanc lorsque l'on donne la valeur 1 à chacune des composantes

$$\vec{B} = \vec{C}_1 + \vec{C}_2 + \vec{C}_3$$

Le vecteur $\vec{B}$ coupe le plan unitaire $\alpha+\beta+\gamma=1$ au point *b*. Tous les vecteurs $\vec{C}$ coupent le plan unitaire en un point *c* et un seul. Ce point, appelé *point chromatique*, représente tous les vecteurs couleur de même teinte, la longueur du vecteur représente l'intensité du rayonnement considéré. Le triangle unitaire est appelé *triangle des couleurs*, toutes les couleurs sont représentées par un point et un seul dans le triangle. Le triangle des couleurs permet de ramener à une représentation plane la représentation vectorielle peu pratique.

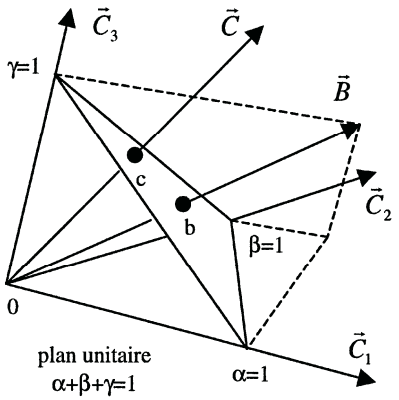


Figure 9-54 : Triangle des couleurs

Système de référence CIE

Compte tenu des caractéristiques physiologiques de notre œil les couleurs primaires qui semblent s'imposer tout naturellement sont le rouge, le vert et le bleu,

le système basé sur cet ensemble de primaires est appelé système RGB (Red Green Blue). Ce système est défini par des primaires de longueurs d'onde respectives 700,0 , 541,6 et 435,8 nm. Si on travaille en unités photométriques les luminences relatives des primaires RGB, ou *composantes trichromatiques spectrales*, qui sont nécessaires pour reproduire un spectre blanc sont dans les proportions 1 , 4,5907 , 0,0601. Ceci signifie que si le flux lumineux de la primaire R vaut 1 lumen, ceux des primaires G et B doivent être respectivement 4,5907 et 0,0601 lumens. En unités énergétiques quant à elles les proportions sont 71,0962 , 1,3791 et 1, ce qui signifie que si le flux énergétique de R vaut 71,0962 watts, ceux de G et B valent respectivement 1,3791 et 1 watt. Ces valeurs très inégales, tant en flux énergétique qu'en flux lumineux, confinent une bonne partie des couleurs autour du point rouge pour les flux énergétiques ou autour de la droite rouge-vert pour les flux lumineux. Ces unités ne sont donc pas d'utilisation pratique. On fixe alors arbitrairement les unités de telle sorte que les coordonnées trichromatiques du blanc soient toutes les trois égales à 1/3, plaçant ainsi le blanc au centre de gravité du triangle des couleurs. Cette évaluation des primaires, ni en flux énergétique ni en flux lumineux, est dite en *flux chromatique*. On peut s'interroger sur la raison pour laquelle les unités lumineuses ne donnent pas de résultat convenable.

L'utilisation des primaires réelles R, G et B conduit en outre à d'autres inconvénients pratiques tels par exemple qu'il est parfois nécessaire d'introduire des composantes trichromatiques négatives pour égaliser certaines couleurs réelles pures ou très saturées. Sans entrer dans les détails de ce genre de problème qui sort du cadre de nos propos, disons simplement que pour éviter ces valeurs négatives, peu commodes dans les calculs colorimétriques, la Commission Internationale de l'Eclairage (CIE) a défini un autre système colorimétrique, appelé système XYZ, sur la base de trois primaires virtuelles. Ces stimulus de référence n'existant pas physiquement il n'est évidemment pas possible d'obtenir expérimentalement des égalisations visuelles de couleur avec ces trois primaires, comme c'était le cas avec les primaires RGB. Les fonctions colorimétriques associées au système XYZ sont obtenues par un changement de base à partir des fonctions colorimétriques du système RGB.

Diagramme de chromaticité

Dans la pratique une couleur est identifiée non pas par ses composantes trichromatiques mais par trois nouvelles quantités qui permettent de représenter dans un plan tous les stimulus de couleur, ces nouvelles quantités portent le nom de *coordonnées trichromatiques*. La représentation graphique qui en découle est appelée *diagramme de chromaticité CIE*. Dans le triangle des couleurs du système RGB les primaires choisies sont situées aux sommets d'un triangle équilatéral alors que dans le diagramme de chromaticité du système XYZ les primaires ont été choisies de façon à former un triangle rectangle. Par convention tout stimulus de couleur est représenté par un point chromatique dont les coordonnées trichromatiques vérifient les relations, $x \leq 1$, $y \leq 1$, $z \leq 1$, $x + y + z = 1$. La connaissance de deux des trois coordonnées suffit donc à positionner le point chromatique dans le diagramme de chromaticité. Par convention le triangle est

rectangle en *Z* et seules les coordonnées *x* et *y* sont utilisées. Les coordonnées des primaires X, Y et Z sont respectivement X(1,0,0), Y(010), Z(001), les coordonnées de tous les stimulus monochromatiques du spectre visible sont tabulées (Trouvé A.-1991). Si on représente ces stimulus entre 380 et 780 nm on obtient une courbe ouverte aux deux extrémités appelée *lieu spectral* (Fig. 9-55). La droite qui referme la courbe en joignant les points à 380 et 780 nm est appelée *droite des pourpres purs*, car elle représente des couleurs obtenues par mélange de rouge et de violet situés aux extrémités du spectre. Un rayonnement monochromatique est représenté par un point sur le lieu spectral, un rayonnement quelconque est représenté par un point à l'intérieur de la surface définie par le lieu spectral et la droite des pourpres purs. Par exemple la couleur du rayonnement monochromatique de longueur d'onde 560 nm est représentée par le point de coordonnées x=0,373, y=0,625 (et par suite z=0,002). La lumière blanche est représentée par le point d'égales coordonnées trichromatiques x=0,333, y=0,333. La lumière du jour est représentée par le point de coordonnées x=0,3101, y=0,3162. On peut remarquer qu'il existe à l'intérieur du triangle XYZ des zones ne correspondant à aucune couleur réelle.

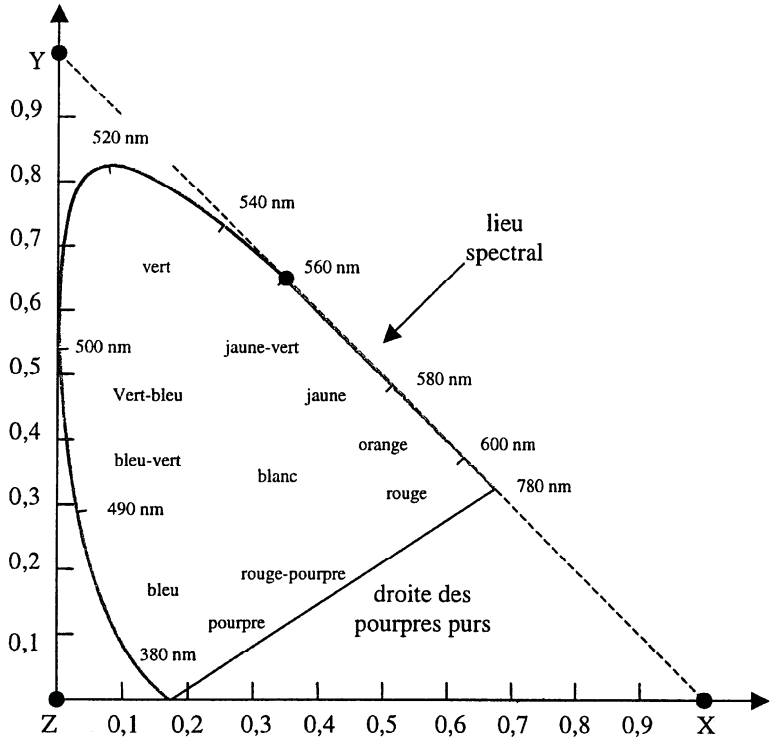


Figure 9-55 : Système CIE

EXERCICES DU CHAPITRE 9

- **Exercice 1 : Énergie absorbée**

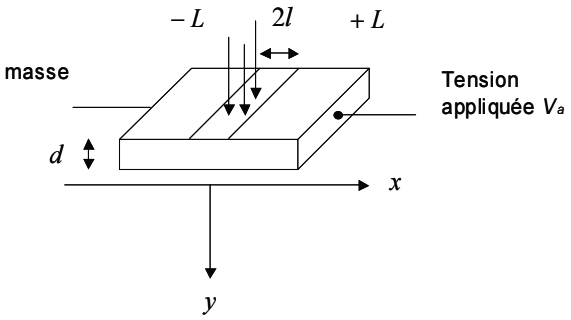
Calculez l'énergie absorbée par seconde par un morceau de silicium de 0,25 micron d'épaisseur soumis à un flux de 10 mW de photons de 3 eV. Calculez l'énergie par seconde dissipée en chaleur.

- **Exercice 2 : Absorption du GaAs**

Un morceau de Ga As est illuminé à 0,6 micron. La puissance incidente est 15 mW. On suppose qu'un tiers de la puissance est réfléchie et qu' un autre tiers sort de l'échantillon. Calculez l'épaisseur et l'énergie thermique dissipée par seconde.

- **Exercice 3 : Eclairement localisé**

Un dispositif semiconducteur est éclairé à travers un masque de largeur *2l* comme le montre la figure.



On suppose que L est très supéreur à la longueur de diffusion et que αd est petit devant 1, α étant le coefficient d'absorption.

En éclairement permanent montrez que Δn est de la forme $Ae^{-\frac{x}{L_1}} + Be^{\frac{x}{L_2}}$ avec $L_1 = -\frac{1}{r_+}$ et $L_2 = \frac{1}{r_-}$ où r_+ et r_- sont solutions de $L_n^2 r^2 + L_E r - 1 = 0$ avec $L_E = \mu_n \tau_n E_a$ où Ea est le champ constant dans le dispositif.

Dans le cas V_a nul montrez que Δn est maximum en $x=0$ et donné par

$$\phi\alpha\tau_n\left(1 - e^{-\frac{l}{L_n}}\right)$$

Dans le cas V_a non nul, montrez *que* Δn est maximum pour $x = l\ \frac{1-\frac{L_E^{\ 2}}{L_n^{\ 2}}}{1-\frac{L_E^{\ 2}}{L_n^{\ 2}}}$.

- **Exercice 4 : Eclairement variable dans le temps**

On reprend les données du problème 3 mais l'éclairement dure de $t=0$ à $t=\tau_0$.
On suppose $2l=\delta(x)$ et $\tau_0=\delta(t)$.

Montrez que Δn vérifie

$$\frac{\partial \Delta n}{\partial t} = \mu_n E_a \frac{\partial \Delta n}{\partial x} + D_n \frac{\partial^2 \Delta n}{\partial x^2} - \frac{\Delta n}{\tau_n}$$

Résoudre l'équation en prenant la transformée de Fourier et montrez que la solution est

$$\Delta n(x,t) = \frac{\phi\alpha}{2\sqrt{\pi D_n t}} e^{-\frac{t}{\tau_n} - \frac{(x-\mu_n E_a t)^2}{4 D_n t}}$$

Etudiez les cas V_a nul et non nul.

- **Exercice 5 : Courant de seuil d'une diode laser**

Calculez le courant de seuil de la diode laser arséniure de gallium ayant les caractéristiques suivantes : reflexions avant et arrière 0,4 et 0,99, longueur et largeur de la cavité 300 et 5 microns, coefficient d'absorption 100 cm^{-1} . Le terme $edc / nA\tau_n$ sera supposé égal à 1/0,1 $cm^3 A^{-1}$.

- **Exercice 6 : Cellule solaire**

Une cellule solaire silicium de surface 2 cm^2 a comme dopants N_a=2 10^{16} cm^{-3} et N_d=5 10^{19} cm^{-3} . Le photocourant est 100 mA. Calculez la caractéristique courant-tension de la diode ainsi que la tension en circuit ouvert et la puissance maximale fournie.

10. Effets quantiques dans les composants, puits quantiques et Superréseaux

Le but de ce chapitre est de montrer les effets quantiques qui se manifestent dans des régions de dimension comparable à la longueur d'onde de l'électron. Ces effets entraînent des modifications dans le comportement des composants semiconduscteurs miniaturisés et doivent être pris en compte. Ils permettent également de construire des dispositifs nouveaux comme les puits quantiques.
Ce chapitre montre comment il est possible d'inclure les effets quantiques liés à l'existence d'une variation brutale du potentiel électrique dans le calcul des énergies et des états des électrons mobiles dans un semiconducteur. Deux exemples importants sont traités : le canal de conduction d'un transistor MOS inversé et le dispositif formé par une série de couches alternées de deux semiconducteurs de types différents (superréseau). Le dispositif plus simple formé par une série de trois semiconducteurs (SC_2-SC_1-SC_2) est appelé puits quantique. Ces structures sont utilisées pour fabriquer des composants optroniques (émetteurs et détecteurs) qui seront détaillées dans le chapitre 12.

Le chapitre est organisé de la manière suivante.

10.1. Effets quantiques dans les hétérojonctions et les structures MIS
10.2. Puits quantiques
10.3. Superréseaux
10.4. Le TEGFET

10.1. Effets quantiques dans les hétérojonctions et les structures MIS

L'étude des hétérostructures a montré que la région voisine de l'interface entre deux matériaux, pouvait être le siège d'une grande densité de porteurs libres, fonction de la nature des matériaux et de la polarisation de la structure. Ces porteurs libres sont localisés au voisinage de l'interface, dans un puits de potentiel très étroit, représenté dans le diagramme énergétique par la courbure des bandes de valence et de conduction. Considérons plus particulièrement la structure MIS, l'intégration classique de l'équation de Poisson a permis de calculer la nature et la densité des charges, en fonction de la polarisation. Le calcul classique est développé en considérant d'une part que tous les donneurs et accepteurs sont ionisés, et d'autre part que les porteurs libres obéissent à la statistique de Boltzmann. La première hypothèse est peu justifiée en régime d'accumulation et à basse température, car les impuretés majoritaires sont alors en partie neutralisées par les porteurs. La deuxième hypothèse est mise en défaut en régimes d'accumulation et d'inversion, en raison du fait que le semiconducteur est alors dégénéré.

Néanmoins, les résultats du chapitre 5 donnent avec une bonne approximation l'ordre de grandeur des densités superficielles de charge mises en jeu dans les différents régimes de fonctionnement de la structure. La densité surfacique de charge de déplétion est de l'ordre de 10^{-8} C.cm^{-2}, ce qui résulte d'une densité d'accepteurs ionisés de l'ordre de 10^{11} cm^{-2}. La densité surfacique de charge d'accumulation, ou d'inversion, varie entre 10^{-8} et 10^{-5} C.cm^{-2}, ce qui correspond à une densité de trous en régime d'accumulation, et d'électrons en régime d'inversion, variant entre 10^{10} et 10^{13} cm^{-2}. Considérons par exemple, une situation moyenne dans le régime de forte inversion (Fig. 10-1), la charge d'espace est constituée de charges de déplétion, en densité superficielle de l'ordre de 10^{11}cm^{-2}, et de charges d'inversion en densité surfacique de l'ordre de 10^{12} cm^{-2}. Les charges de déplétion sont dues aux accepteurs, de sorte que si on suppose le dopage homogène avec N_a=4.10^{15} cm^{-3}, la largeur de la zone déplétion est de l'ordre de 10^{11}/4.10^{15}, soit $z_{dep}\approx 1\mu m$. Les charges d'inversion sont les électrons de la bande de conduction, en prenant le modèle le plus simple avec une couche d'inversion supposée homogène et une distribution de Boltzmann, la densité équivalente d'états de la bande de conduction étant de l'ordre de 10^{19} cm^{-3}, la largeur moyenne de la couche d'inversion est de l'ordre de $10^{12}/10^{19}$ soit $z_{inv}\approx$10Å. Ce calcul est évidemment très grossier dans la mesure où la densité d'électrons est loin d'être homogène dans la couche d'inversion et où l'approximation de Boltzmann n'est plus justifiée quand le semiconducteur est dégénéré. Néanmoins, l'ordre de grandeur est respecté et la couche d'inversion s'étend, à partir de la surface, sur une distance de quelques dizaines d'Angströms seulement.

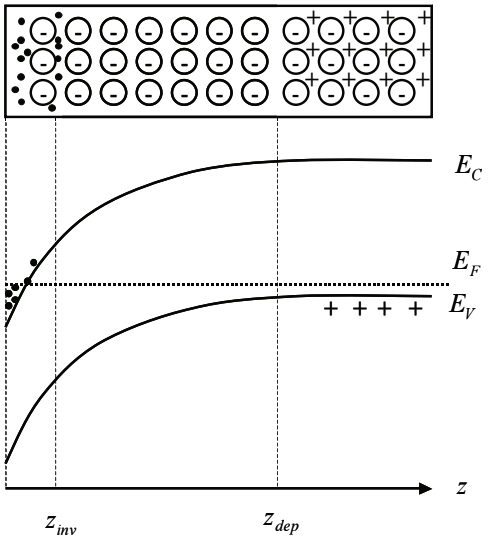


Figure 10-1 : Distributions spatiale et énergétique de la charge d'espace dans une structure MIS en forte inversion

Examinons maintenant le mouvement des électrons dans ce puits de potentiel de quelques dizaines d'Angströms d'épaisseur.

Considérons tout d'abord la représentation corpusculaire des électrons dans le semiconducteur. Ces électrons sont des particules de masse effective m_e, animées d'un mouvement brownien caractérisé par un libre parcours moyen $\ell = v_{th}\tau_c$ où v_{th} est la vitesse thermique des porteurs et τ_c leur temps de collision (Par. 1.5.1.a). A la température ambiante $v_{th}\approx 10^7$ cm/s et $\tau_c\approx 10^{-13}$ à 10^{-12} s, ce qui donne ℓ=100 à 1000 Å. Comparons z_{inv} et ℓ. La largeur du puits de potentiel est inférieure au libre parcours moyen des électrons. Ces électrons se comportent alors comme des particules libres dans un puits de potentiel, leur mouvement est conditionné par les réflexions sur les faces du puits, leurs états d'énergie sont quantifiés.

Considérons maintenant la représentation ondulatoire. Les électrons dans le semiconducteur occupent des états d'énergie donnés par

$$E=E_c+\frac{\hbar^2 k^2}{2m_e} \tag{10-1}$$

La longueur d'onde de de Broglie associée à ces électrons est donnée par

$$\lambda = \frac{2\pi}{k} = \frac{h}{\sqrt{2m_e(E-E_c)}} \tag{10-2}$$

Si on suppose le niveau de Fermi situé à une centaine de meV dans la bande de conduction, en prenant $m_e\approx m_o$, la longueur d'onde associée aux électrons du niveau

de Fermi est $\lambda_F \approx 40$ Å. Comparons z_{inv} et λ_F. On constate que les électrons sont confinés dans un puits de potentiel dont la largeur est du même ordre de grandeur que leur longueur d'onde. En d'autres termes, la longueur de l'onde associée à l'électron est du même ordre de grandeur que la cavité dans laquelle est enfermée cette onde. Il en résulte l'existence d'ondes stationnaires, données par $2z_{inv} = k\lambda$. Dans la mesure où λ est du même ordre de grandeur que z_{inv}, l'ordre d'interférence k est petit, et par suite la différence $\delta\lambda$ entre deux ondes stationnaires, c'est-à-dire la différence d'énergie entre les états correspondants, est grande. Au contraire, si z_{inv} devient beaucoup plus grand que λ, $\delta\lambda$ devient petit et la quantification disparaît.

Considérons enfin le principe d'incertitude. Dans la direction z, la relation d'Heisenberg s'écrit $\Delta z.\Delta p_z > \hbar/2$ où $h = 2\pi\hbar$ est la constante de Planck et où Δz et Δp_z représentent respectivement les incertitudes de position et d'impulsion suivant z, définie par

$$\Delta z = \left(\left\langle z^2 \right\rangle - \left\langle z \right\rangle^2\right)^{1/2} \qquad \Delta p_z = \left(\left\langle p_z^2 \right\rangle - \left\langle p_z \right\rangle^2\right)^{1/2}$$

Dans le puits de largeur z_{inv} l'incertitude de position dans la direction z est nécessairement inférieure à z_{inv} ($\Delta z < z_{inv}$). Il en résulte que l'incertitude d'impulsion Δp_z est nécessairement supérieure à $\hbar/2\, z_{inv}$. Soit

$$\Delta p_z = \left(\left\langle p_z^2 \right\rangle - \left\langle p_z \right\rangle^2\right)^{1/2} > \hbar/2z_{inv}$$

D'autre part, l'électron étant confiné dans le puits, il ne peut se propager dans la direction z, son impulsion moyenne suivant z est par conséquent nulle, $\left\langle p_z \right\rangle = 0$. La relation précédente s'écrit donc:

$$\left\langle p_z^2 \right\rangle^{1/2} > \hbar/2z_{inv} \rightarrow \left\langle p_z^2 \right\rangle > \hbar^2/4z_{inv}^2$$

L'énergie de l'électron dans son mouvement suivant z, qui est donnée par $E = \left\langle p_z^2 \right\rangle / 2m_e$, obéit donc à la relation

$$E > \hbar^2/8m_e z_{inv}^2$$

Ainsi l'électron ne peut pas être au repos au fond du puits, son énergie est quantifiée avec une valeur minimum de confinement. Dans le cas du GaAs par exemple, m_e=0,067 m_o, avec z_{inv}=10 Å, on obtient E>150 meV.

En conséquence, les électrons (ou les trous) confinés dans une couche d'inversion ou d'accumulation, se comportent comme un gaz d'électrons (ou de trous) à deux dimensions. Leur mouvement est libre dans le plan de la structure et quantifié dans la direction perpendiculaire. Il en résulte que le calcul classique de la

distribution de ces porteurs libres n'est qu'approché, et qu'une étude détaillée passe par un traitement quantique du problème.

Notons que la profondeur du puits de potentiel, $e\phi_d$, qui varie avec la polarisation, n'évolue plus beaucoup quand le semiconducteur devient dégénéré. Il en résulte que l'énergie de confinement est très différente dans une couche d'accumulation et dans une couche d'inversion (Fig. 10-2).

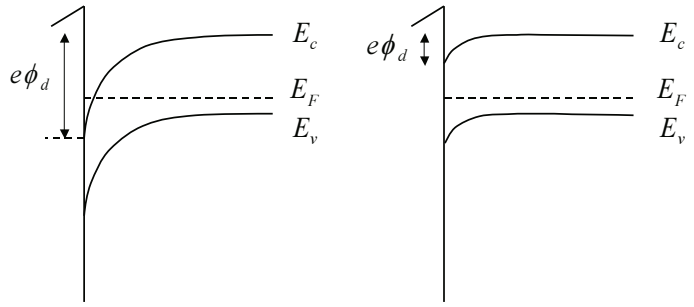


Figure 10-2 : Puits de potentiel. a) Couche d'inversion dans un semiconducteur de type p. b) Couche d'accumulation dans un semiconducteur de type n

10.1.1. *Structure de sous-bandes d'énergie*

Il s'agit de calculer les niveaux d'énergie et les fonctions d'onde des électrons confinés dans une couche d'inversion, ou d'accumulation, au voisinage de la surface du semiconducteur. Il suffit pour cela d'intégrer l'équation de Schrödinger

$$\left(T+V(\mathbf{r})\right)\psi(\mathbf{r})=E\ \psi(\mathbf{r}) \tag{10-3}$$

où T représente l'opérateur énergie cinétique des électrons, et $V(\boldsymbol{r})$ l'opérateur énergie potentielle. $\psi(\boldsymbol{r})$ et E représentent respectivement la fonction d'onde et l'énergie associées à l'état propre . Dans la couche d'inversion, ou d'accumulation, les électrons voient un potentiel relativement complexe. D'un côté de l'interface, ils voient le potentiel périodique du semiconducteur sur lequel est superposé un potentiel plus lentement variable résultant, de la différence des travaux de sortie du métal et du semiconducteur, de la présence de la charge d'espace et de la polarisation éventuelle de la structure. De l'autre côté de l'interface, ils voient le potentiel de l'isolant. En outre une couche de transition, avec des propriétés différentes de celles de l'isolant et de celles du semiconducteur, peut exister à l'interface.

La résolution de l'équation (10-3) passe préalablement par le calcul du terme d'énergie potentielle $V(\boldsymbol{r})$. Près de l'interface isolant-semiconducteur ce terme est relativement complexe, et résulte de plusieurs contributions. Avant de le calculer, deux simplifications peuvent être faites.

a. Hauteur de barrière côté isolant

Côté isolant, les électrons voient une barrière verticale dont la hauteur résulte de la différence d'énergie, à l'interface, des bandes de conduction de l'isolant et du semiconducteur. Cette différence n'est autre que la différence des affinités électroniques de l'isolant et du semiconducteur $e\chi_s$-$e\chi_i$. Dans la structure SiO_2-Si, $e\chi_s$=4 eV et $e\chi_i$=0,85 eV, la hauteur de barrière est donc de l'ordre de 3 eV. Compte tenu de cette valeur, nous supposerons la hauteur de barrière infinie côté isolant, aucun électron ne passe du semiconducteur dans l'isolant. En d'autres termes, la fonction d'onde des électrons s'annule à l'interface isolant-semiconducteur, que nous prendrons comme origine des coordonnées. Cette hypothèse est pleinement justifiée dans les structures MIS réalisées avec une couche d'isolant relativement épaisse, d>100 Å. Dans une structure MIS dont l'isolant est relativement mince, l'effet tunnel joue un rôle important et les fonctions d'onde des électrons ne s'annulent plus à l'interface. Dans les hétérojonctions, cette hypothèse est assez peu justifiée dans la mesure où la différence des affinités électroniques $e\chi_{s1}$-$e\chi_{s2}$, peut-être relativement faible. La hauteur de barrière ΔE_c qui en résulte est alors peu importante.

b. Approximation de la masse effective

Parmi les différentes contributions au potentiel que voient les électrons, le potentiel périodique associé aux ions du cristal, c'est-à-dire le potentiel cristallin, joue évidemment un rôle important. Les états propres et les fonctions propres correspondants sont évidemment connus, il s'agit pour les premiers de la bande de conduction et pour les secondes des fonctions de Bloch. On peut alors s'affranchir aisément de ce terme en utilisant l'approximation de la masse effective. L'électron de masse m_0 dans le potentiel cristallin, se comporte comme un électron de masse effective m_e et dont la fonction d'onde est une onde de Bloch, cet électron se déplace dans un réseau périodique supposé vide d'ions. En d'autres termes, le potentiel cristallin est représenté par le fait que l'électron a une masse effective m_e différente de la masse m_0 de l'électron libre. La périodicité de la structure cristalline entraîne une périodicité de la fonction d'onde associée à l'électron qui doit satisfaire au théorème de Bloch. L'approximation de la masse effective est utilisée avec succès dans l'étude de tous les composants, cependant elle atteint ici une limite de validité dans la mesure où les électrons sont confinés dans une épaisseur de l'ordre de quelques couches atomiques. Toutefois l'ignorance de certains paramètres physiques, comme par exemple la structure réelle de la couche de transition isolant-semiconducteur où le rôle exact joué par les interactions électron-électron, perturbent beaucoup plus le problème que l'approximation de la masse effective. Nous nous placerons dans cette approximation.

La version la plus simple de l'approximation de la masse effective consiste à affecter à l'électron la masse effective de la bande de conduction correspondante, supposée isotrope et parabolique. L'opérateur énergie cinétique s'écrit alors

$$T=-\frac{\hbar^2}{2m_e}\Delta \tag{10-4}$$

L'opérateur énergie potentielle s'écrit

$$V(z)=-e\,\phi(z) \tag{10-5}$$

où $\phi(z)$ est le potentiel électrostatique qui, compte tenu de la géométrie de la structure, ne varie que dans la direction z perpendiculaire à la structure. Ce potentiel est fonction de la densité de charge par l'équation de Poisson

$$\frac{d^2\phi(z)}{dz^2}=-\frac{\rho(z)}{\varepsilon_s} \tag{10-6}$$

$\rho(z)$ est la densité totale de charge, charge de déplétion et charge d'inversion ou d'accumulation, ε_s est la constante diélectrique du semiconducteur. Ce potentiel satisfait en outre aux conditions aux limites suivantes

$$\phi(z\to\infty)=0 \tag{10-7-a}$$

$$\varepsilon_i\left(d\phi/dz\right)_{0^-}=\varepsilon_s\left(d\phi/dz\right)_{0^+} \tag{10-7-b}$$

En l'absence de charge d'espace la fonction d'onde de l'électron est une fonction propre du potentiel cristallin, c'est-à-dire une fonction de Bloch. L'état propre de l'électron est un état de la bande de conduction d'énergie $E=E_c+\hbar^2k^2/2m_e$. En présence d'une charge d'espace, qui est toujours lentement variable devant le potentiel cristallin, la fonction d'onde s'écrit sous la forme du produit d'une onde de Bloch par une fonction enveloppe. L'énergie de l'électron est la somme de son énergie dans la bande de conduction et de l'énergie de son mouvement associé à la présence de la charge d'espace. La fonction enveloppe et l'énergie additionnelle sont respectivement fonction propre et sa valeur propre de l'équation de la masse effective. Mathématiquement on cherhe la solution sous forme d'un produit de deux fonctions $\xi_i(z)e^{\left(ik_x x+ik_y y\right)}\varphi(x,y)$. On montre alors que la valeur propre totale (énergie totale) est la somme des énergies E_i et $E_c\dfrac{\hbar^2k_{//}^2}{2m_e}$ respectivement valeurs propres pour les fonctions $\xi_i(z)$ et $e^{\left(ik_x x+ik_y y\right)}\varphi(x,y)$.

L'énergie potentielle de l'électron n'étant fonction que de la variable z, le mouvement de l'électron n'est perturbé que dans la direction z. Il reste libre dans le

plan *xy*. Il en résulte que la fonction enveloppe peut s'écrire sous la forme d'un produit d'une fonction de *z* par une onde plane dans le plan *xy.* La fonction d'onde de l'électron s'écrit alors

$$\psi_i(r)=\xi_i(z)e^{(ik_x x+ik_y y)}\varphi(x,y) \tag{10-8}$$

où $\varphi(x,y)$ est la fonction de Bloch, et $\xi_i(z)$ est une solution de l'équation de la masse effective. L'indice *i* est relatif à la quantification des solutions. On en dèduit alors

$$\frac{\hbar^2}{2m_e}\frac{d^2\xi_i(z)}{dz^2}+\left[E_i-V(z)\right]\xi_i(z)=0 \tag{10-9}$$

avec les conditions aux limites

$$\xi_i(z=0)=0 \qquad \xi_i(z\to\infty)=0 \tag{10-10}$$

Les énergies E_i sont les énergies des électrons dans leur mouvement dans la direction *z.* Elles décrivent la quantification des états électroniques dans la direction perpendiculaire à l'interface. Dans le plan de l'interface, le mouvement des électrons n'est pas modifié. Il en résulte une structure de sous-bandes d'énergie avec une quantification discrète suivant k_z et une variation pseudo-continue suivant $k_{//}$ avec $k_{//}^2 = k_x^2 + k_y^2$

$$E=E_c+E_i+\frac{\hbar^2 k_{//}^2}{2m_e} \tag{10-11}$$

Les courbes de dispersion sont représentées sur la figure (10-3).

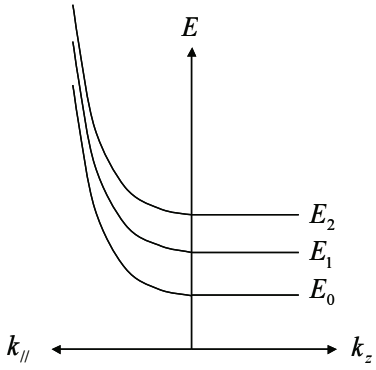


Figure 10-3 : Sous bandes

Le mouvement des électrons est donc limité à deux dimensions. Dans un système à *n* dimensions, la densité d'états dans l'espace des *k* est donnée, compte tenu du spin, par (Eq. 1-11).

$$g(k)=\frac{2}{(2\pi)^n} \tag{10-12}$$

Le système étant ici réduit à deux dimensions, la densité d'états s'écrit

$$g(k)=\frac{2}{(2\pi)^2} \tag{10-13}$$

Ainsi sur le disque de rayon $k_{//}$ le nombre d'états est donné par

$$N(k_{//})=\frac{2}{(2\pi)^2}\pi k_{//}^2 \tag{10-14}$$

Si on prend dans chacune des sous-bandes, l'origine des énergies au bas de la sous-bande, l'énergie s'écrit

$$E=\frac{\hbar^2 k_{//}^2}{2m_e} \tag{10-15}$$

de sorte que

$$k_{//}^2=\frac{2m_e}{\hbar^2}E \tag{10-16}$$

En explicitant $k_{//}$ dans l'expression (10-14) on obtient

$$N(E)=\frac{2}{(2\pi)^2}\pi\frac{2m_e}{\hbar^2}E \tag{10-17}$$

Ainsi la densité d'états dans l'espace des énergies s'écrit pour une sous-bande

$$g_i(E)=\frac{dN(E)}{dE}=\frac{m_e}{\pi\hbar^2} \tag{10-18}$$

C'est une caractéristique essentielle d'un système à deux dimensions, la densité d'états est constante en fonction de l'énergie. En tenant compte de toutes les sous-bandes la densité d'états s'écrit

$$g(E)=\frac{m_e}{\pi\hbar^2}\sum_i H(E-E_i) \tag{10-19}$$

où $H(x)$ est la fonction de Heaviside , $H(x)=0$ pour $x<0$, $H(x)=1$ pour $x>0$. La courbe représentant la densité d'états est représentée sur la figure (10-4)

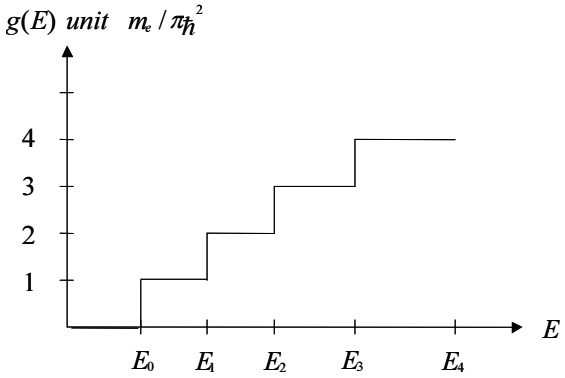


Figure 10-4 : Densité d'états à deux dimensions

A la température *T*, la population de chaque sous-bande est donnée par

$$n_i = \int_{E_i}^{\infty} g_i(E) f(E) dE \tag{10-20}$$

soit

$$n_i = \frac{m_e}{\pi\hbar^2} \int_{E_i}^{\infty} \frac{dE}{1+e^{(E-E_F)/kT}} \tag{10-21}$$

L'intégrale de Fermi est donnée par

$$\int_{E_i}^{\infty} \frac{dE}{1+e^{(E-E_F)/kT}} = kT\, Ln\left(1+e^{(E_F-E_i)/kT}\right) \tag{10-22}$$

Ainsi la population de la sous-bande *i* s'écrit

$$n_i = \frac{m_e}{\pi\hbar^2} kT\, Ln\left(1+e^{(E_F-E_i)/kT}\right) \tag{10-23}$$

Sauf pour $|E_F - E_i| \approx kT$, on peut établir deux expressions simplifiées de n_i valables respectivement pour une sous-bande située au-dessous du niveau de Fermi et pour une sous-bande située au-dessus du niveau de Fermi

- Pour $E_F - E_i >> kT$ $\qquad Ln\left(1+e^{(E_F-E_i)/kT}\right) \approx \frac{E_F - E_i}{kT}$

$$n_i \approx \frac{m_e}{\pi\hbar^2}(E_F - E_i) \tag{10-24}$$

- Pour $E_i - E_F >> kT$ $\qquad Ln\left(1+e^{(E_F-E_i)/kT}\right) \approx e^{(E_F-E_i)/kT}$

$$n_i \approx \frac{m_e}{\pi\hbar^2} kT e^{(E_F-E_i)/kT} \tag{10-25}$$

La densité totale d'électrons dans la couche d'inversion est donnée par $n = \sum_i n_i$, soit

$$n = \frac{m_e}{\pi\hbar^2} kT \sum_i Ln\left(1+e^{(E_F-E_i)/kT}\right) \tag{10-26}$$

En outre, le carré du module de la fonction enveloppe, $|\xi_i(z)|^2$, représente la probabilité de présence, au point d'abscisse z, d'un électron de la sous-bande i. Il en résulte que dans la mesure où le nombre d'électrons de la sous-bande i est n_i , la distribution spatiale de ces électrons est donnée par

$$n_i(z) = n_i \left|\xi_i(z)\right|^2 \tag{10-27}$$

La distribution spatiale de l'ensemble des électrons de la couche d'inversion s'écrit

$$n(z) = \sum_i n_i \left|\xi_i(z)\right|^2 \tag{10-28}$$

Il en résulte que la charge d'espace associée à la couche d'inversion est donnée par

$$\rho_{inv}(z) = -e \sum_i n_i \left|\xi_i(z)\right|^2 \tag{10-29}$$

ou en explicitant n_i

$$\rho_{inv}(z) = -\frac{em_e}{\pi\hbar^2} kT \sum_i Ln\left(1+e^{(E_F-E_i)/kT}\right) \left|\xi_i(z)\right|^2 \tag{10-30}$$

On constate en particulier que la charge d'espace, résultant de la présence de la couche d'inversion, est directement liée à la fonction d'onde électronique $\xi_i(r)$. Mais $\xi_i(r)$ est une solution de l'équation de Schrödinger, donc fonction de $V(z)$, qui est lui-même fonction de la densité de charge. Il en résulte que le calcul des états électroniques doit se faire de manière auto-cohérente. Avant d'aborder les méthodes de résolution de l'équation de Schrödinger, nous devons étudier les différentes contributions à l'énergie potentielle et expliciter $V(z)$.

10.1.2. *Energie potentielle des électrons*

Dans l'approximation de la masse effective, les fonctions de base étant les fonctions de Bloch, le potentiel cristallin disparaît. Néanmoins la résolution de l'équation (10-9) nécessite la connaissance du terme d'énergie potentielle $V(z)$ spécifique de la structure. Cette énergie potentielle $V(z) = -e\phi(z)$ est associée au potentiel électrique $\phi(z)$, qui résulte de la présence au voisinage de l'interface, des charges de déplétion et d'inversion, et de l'existence d'un potentiel image dû à la discontinuité de constante diélectrique. L'énergie potentielle $V(z)$ résulte donc de la contribution de trois termes et peut s'écrire sous la forme

$$V(z) = V_{dep}(z) + V_{inv}(z) + V_{im}(z) \tag{10-31}$$

Nous allons étudier successivement chacun de ces termes.

a. Potentiel de déplétion

$V_{dep}(z) = -e\phi_{dep}(z)$, où $\phi_{dep}(z)$ est le potentiel associé à la présence des charges de déplétion. Pour calculer $\phi_{dep}(z)$ nous supposerons que le dopage du semiconducteur est homogène, que la zone de déplétion est vide de porteurs sur une profondeur z_d, et enfin que toutes les impuretés sont ionisées.

En appelant $N_a = (N_a - N_d)$, l'excédent d'accepteurs dans le semiconducteur de type p, la charge d'espace de déplétion s'écrit $\rho_{dep} = -eN_a$ et l'équation de Poisson s'écrit

$$\frac{d^2\phi_{dep}(z)}{dz^2} = -\frac{\rho_{dep}(z)}{\varepsilon_s} = \frac{eN_a}{\varepsilon_s} \tag{10-32}$$

En intégrant deux fois, avec la condition $E(z=z_d)=0$ et $\phi(z=0)=0$ on obtient

$$\frac{d\phi_{dep}(z)}{dz} = \frac{eN_a}{\varepsilon_s}(z - z_d) \tag{10-33}$$

et

$$\phi_{dep}(z) = -\frac{eN_a}{\varepsilon_s} z_d \;\; z\left(1 - \frac{z}{2z_d}\right) \tag{10-34}$$

Soit ϕ_d la courbure globale du potentiel, c'est-à-dire la différence entre le potentiel de surface et le potentiel de volume du semiconducteur (Fig. 10-5).

$$\phi_d = \phi_{dep}(z=0) - \phi_{dep}(z=z_d) \tag{10-35}$$

Soit

$$\phi_d = \frac{eN_a}{2\varepsilon_s} z_d^2 \tag{10-36}$$

Ainsi en fonction de ϕ_d, la profondeur de la zone de déplétion est donnée par

$$z_d = \left(\frac{2\varepsilon_s \phi_d}{eN_a} \right)^{1/2} \tag{10-37}$$

D'autre part en faisant apparaître le nombre total de charges de déplétion, donné par $N_{dep}=N_a z_d$, le potentiel s'écrit

$$\phi_{dep}(z) = -\frac{eN_{dep}}{\varepsilon_s} z \left(1 - \frac{z}{2z_d} \right) \tag{10-38}$$

L'énergie potentielle associée à la charge de déplétion est alors donnée par

$$V_{dep}(z) = \frac{e^2 N_{dep}}{\varepsilon_s} z \left(1 - \frac{z}{2z_d} \right) \tag{10-39}$$

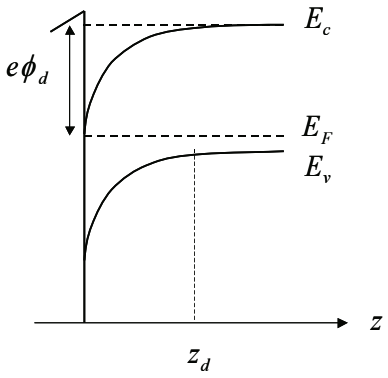


Figure 10-5 : Structure de bandes

Si on augmente ϕ_d , en polarisant la structure, z_d et N_{dep} augmentent jusqu'à ce que le régime de forte inversion soit atteint. A partir de cet instant, l'augmentation de la tension de polarisation se traduit par une augmentation exponentielle de la densité de charges d'inversion, et les trois quantités ϕ_d , z_d et N_{dep} saturent. Le régime de forte inversion est atteint quand le bas de la bande de conduction atteint le niveau de Fermi de la structure (Fig. 10-5). Or dans le matériau de type p le niveau de Fermi est proche de la bande de valence, de sorte que, lorsque le régime de forte inversion est atteint, on peut écrire avec une bonne approximation

$$\phi_d \approx \frac{E_c - E_v}{e} = \frac{E_g}{e} \tag{10-40}$$

Si on considère une structure SiO_2-Si avec N_a=10^{15} cm^{-3}, ε_s=12 ε_o et E_g=1,15 eV. On obtient ϕ_d=1,15 V, z_d=1,2 μm et N_{dep}=1,2.10^{11} cm^{-2}.

Ce potentiel de déplétion résulte de la présence de charges fixes associées aux ions donneurs et accepteurs, N_a-N_d. Un potentiel équivalent existe en régime d'accumulation à basse température, bien qu'aucune zone de déplétion n'existe. En régime d'accumulation dans un semiconducteur de type *n*, une fraction des électrons accumulés à l'interface, neutralisent les ions donneurs (Fig. 10-6). Il en résulte une distribution homogène d'ions accepteurs, de densité N_a , qui créent un potentiel semblable au potentiel de déplétion. En déplétion ou inversion dans un semiconducteur de type *p*, ce potentiel est proportionnel à la densité N_a-N_d , en accumulation dans un semiconducteur de type *n* ce potentiel est proportionnel à la densité N_a.

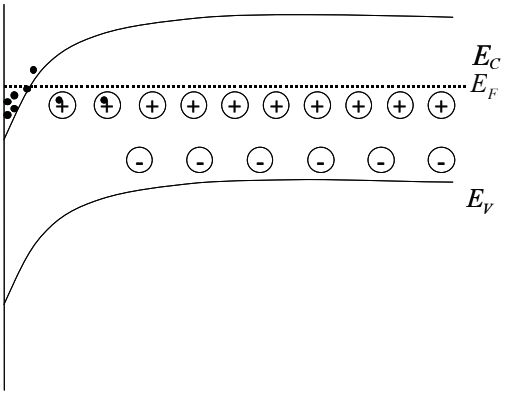


Figure 10-6 : Régime d'accumulation

b. Potentiel d'inversion

Compte tenu de la grande densité d'électrons présents dans la couche d'inversion ou d'accumulation, il n'est pas possible de négliger les interactions électron-électron. Cependant, l'étude de ces interactions constitue un problème à n corps dont le traitement rigoureux passe par la résolution d'un système de n équations couplées. On simplifie le problème en utilisant l'approximation de Hartree.

Si l'énergie cinétique moyenne des électrons est supérieure à l'énergie d'interaction électron-électron, on peut représenter l'interaction électron-électron par un potentiel moyen $\phi_{inv}(z)$, qui représente le potentiel créé au point z où se trouve l'électron, par l'ensemble des autres électrons. Ce potentiel moyen est appelé *potentiel de Hartree*. Or l'énergie cinétique moyenne des électrons est fonction de la position du niveau de Fermi et augmente avec la densité d'électrons, d'autre part l'interaction électron-électron est directement liée à l'extension spatiale de la fonction d'onde électronique.

Soit a_o^* le rayon de Bohr effectif d'un électron de conduction dans le semiconducteur. C'est le rayon de Bohr de l'atome d'hydrogène pondéré par la masse effective des électrons et la constante diélectrique du semiconducteur.

$$a_o^* = \frac{\varepsilon_s \hbar^2}{e^2 m_e} \tag{10-41}$$

Si la densité d'électrons est telle que le nombre d'électrons contenus dans la sphère de rayon a_o^* est important, on peut représenter l'interaction de ces électrons avec l'électron *i*, par un effet moyen représenté par le potentiel de Hartree. Si *n* est la densité d'électrons, le rayon de la sphère contenant un électron est donné par

$$r = \left(\frac{3}{4\pi n} \right)^{1/3} \tag{10-42}$$

de sorte que la condition de validité de l'approximation du potentiel moyen s'écrit

$$r / a_o^* << 1 \tag{10-43}$$

Ainsi, l'approximation de Hartree est d'autant plus justifiée que la densité d'électrons est importante. Pour une densité d'électrons donnée, la condition est d'autant mieux satisfaite que ε_s est grand et que m_e est petit. Il en résulte que cette condition est plus facilement satisfaite dans les semiconducteurs à gap direct et petit, car à un gap direct et petit sont associées une masse effective faible et une constante diélectrique élevée. Par contre dans les semiconducteurs à gap indirect comme le silicium, les effets d'échange et corrélation électron-électron jouent un rôle important.

Nous représenterons l'interaction électron-électron par le potentiel moyen $\phi_{inv}(z)$, nous verrons ensuite comment on modifie ce potentiel, pour tenir compte du fait que l'approximation qu'il représente peut éventuellement n'être pas justifiée. Ce potentiel est donné par l'intégration de l'équation de Poisson dans laquelle la densité de charge est donnée par

$$\rho_{inv}(z) = -e \sum_i n_i \left| \xi_i(z) \right|^2 = -e\, n(z) \tag{10-44}$$

soit

$$\frac{d^2 \phi_{inv}(z)}{dz^2} = \frac{e}{\varepsilon_s} n(z) \tag{10-45}$$

En multipliant par *z* à gauche et à droite, cette équation s'écrit

$$z \frac{d^2 \phi_{inv}(z)}{dz^2} = \frac{e}{\varepsilon_s} z\, n(z) \tag{10-46}$$

ou

$$\frac{d}{dz}\left(z\frac{d\phi_{inv}(z)}{dz}\right)-\frac{d\phi(z)}{dz}=\frac{e}{\varepsilon_s}zn(z) \tag{10-47}$$

Intégrons tout d'abord cette équation de z=0 à z=∞, en prenant l'origine des potentiels à la surface du semiconducteur $\phi_{inv}(z=0)=0$. $z\to\infty$ correspond à la région neutre du semiconducteur de sorte que d'une part $\phi(z\to\infty)=\phi_{sc}$ où ϕ_{sc} représente le potentiel du semiconducteur, et d'autre part $E(z\to\infty)=0$ c'est-à-dire $(d\phi/dz)_{z\to\infty}=0$. Ainsi l'intégration de l'équation donne

$$\left[z\frac{d\phi_{inv}(z)}{dz}\right]_0^{\infty}-\left[\phi(z)\right]_0^{\infty}=\frac{e}{\varepsilon_s}\int_0^{\infty}zn(z)dz \tag{10-48}$$

Le potentiel du semiconducteur $\phi_{sc}=\phi(z\to\infty)$ est par conséquent donné par

$$\phi_{sc}=-\frac{e}{\varepsilon_s}\int_0^{\infty}zn(z)dz \tag{10-49}$$

ou, en explicitant $n(z)$

$$\phi_{sc}=-\frac{e}{\varepsilon_s}\sum_i n_i\int_0^{\infty}z\left|\xi_i(z)\right|^2dz \tag{10-50}$$

On peut écrire ce potentiel sous la forme

$$\phi_{sc}=-\frac{e}{\varepsilon_s}\sum_i n_i\,z_i=-\frac{e}{\varepsilon_s}n_s\,z_m \tag{10-51}$$

avec

$$z_i=\int_0^{\infty}z\left|\xi_i(z)\right|^2dz \tag{10-52}$$

et

$$n_s=\sum_i n_i \tag{10-53}$$

En rappelant que $\left|\xi_i(z)\right|^2$ représente la probabilité de présence à l'abscisse z, des électrons du niveau E_i, on constate que z_i représente l'abscisse moyenne des électrons du niveau E_i et z_m représente l'abscisse moyenne de l'ensemble des électrons de la couche d'inversion.

Revenons sur l'équation différentielle, pour calculer le potentiel en un point d'abscisse z on utilise z' comme variable d'intégration et on intègre l'équation de z'=0 à z'=z avec toujours la condition $\phi(z'=0)=0$. Soit

$$\left[z' \frac{d\phi_{inv}(z')}{dz'} \right]_0^z - \left[\phi_{inv}(z') \right]_0^z = \frac{e}{\varepsilon_s} \int_0^z z' n(z') dz' \tag{10-54}$$

$$z \frac{d\phi_{inv}(z)}{dz} - \phi_{inv}(z) = \frac{e}{\varepsilon_s} \int_0^z z' n(z') dz' \tag{10-55}$$

$$\phi_{inv}(z) = z \frac{d\phi_{inv}(z)}{dz} - \frac{e}{\varepsilon_s} \int_0^z z' n(z') dz' \tag{10-56}$$

Il faut calculer le terme $d\phi_{inv}(z)/dz$. Pour ce faire, il suffit d'intégrer une fois l'équation de Poisson de $z'=0$ à $z'=z$ soit

$$\left[\frac{d\phi_{inv}(z')}{dz'} \right]_0^z = \frac{e}{\varepsilon_s} \int_0^z n(z') dz' \tag{10-57}$$

soit

$$\frac{d\phi_{inv}(z)}{dz} = \left(\frac{d\phi_{inv}(z)}{dz} \right)_{z=0} + \frac{e}{\varepsilon_s} \int_o^z n(z') dz' \tag{10-58}$$

Pour calculer la dérivée du potentiel à l'interface on utilise le théorème de Gauss, en prenant comme surface de Gauss un cylindre suffisamment long avec une base à l'interface et l'autre dans la région neutre du semiconducteur (Fig. 10-7). Sur la première base

$$E_s = E(z=0) = -\left(\frac{d\phi_{inv}(z)}{dz} \right)_{z=0} \tag{10-59}$$

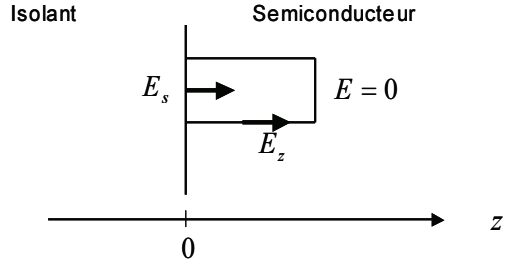


Figure 10-7 : Théorème de Gauss

Sur la deuxième base $E=0$. Enfin sur les faces latérales du cylindre le champ est parallèle à la surface. Le théorème de Gauss s'écrit donc

$$\int_s \mathbf{E}.d\mathbf{s} = \frac{\rho}{\varepsilon_s} \Rightarrow -E_s = -\frac{e}{\varepsilon_s} \sum_i n_i \tag{10-60}$$

Ainsi

$$\left(\frac{d\phi_{inv}(z)}{dz}\right)_{z=0} = -E_s = -\frac{e}{\varepsilon_s}\sum_i n_i \tag{10-61}$$

En portant cette valeur dans l'expression (10-58) et le résultat obtenu dans l'expression (10-56), on obtient

$$\phi_{inv}(z) = -\frac{e}{\varepsilon_s}\sum_i n_i z + \frac{e}{\varepsilon_s} z\int_0^z n(z')dz' - \frac{e}{\varepsilon_s}\int_o^z z' n(z')dz'$$

ou

$$\phi_{inv}(z) = -\frac{e}{\varepsilon_s}\sum_i n_i z + \frac{e}{\varepsilon_s}\int_0^z (z - z')n(z')dz' \tag{10-62}$$

En explicitant $n(z')$ on obtient

$$\phi_{inv}(z) = -\frac{e}{\varepsilon_s}\sum_i n_i\left(z + \int_0^z (z'-z)\left|\xi_i(z')\right|^2 dz'\right) \tag{10-63}$$

Ainsi l'énergie potentielle associée à la charge d'inversion $V_{inv}(z)=-e\phi_{inv}(z)$ est donnée par

$$V_{inv}(z) = \frac{e^2}{\varepsilon_s}\sum_i n_i\left(z + \int_0^z (z'-z)\left|\xi_i(z')\right|^2 dz'\right) \tag{10-64}$$

c. Potentiel image

Le potentiel image résulte de la discontinuité de la constante diélectrique à l'interface isolant-semiconducteur. Il joue un rôle non négligeable en régime d'inversion ou d'accumulation dans la mesure où les charges sont localisées au voisinage de l'interface.

Considérons un électron de charge $q = -e$, en un point M d'abscisse d dans le semiconducteur (Fig. 10-8). Cette charge q crée un champ électrique qui polarise l'isolant. L'isolant polarisé est équivalent à une assemblée de dipôles qui crée en M où se trouve la charge q, un champ et un potentiel. Ce potentiel $\phi_{im}(z)$ est appelé *potentiel image* de la charge q. Compte tenu de la symétrie du système, le champ créé en M par l'isolant polarisé, est dirigé suivant $o\vec{z}$. Ainsi, ce champ et le potentiel qui lui est associé, sont équivalents à ceux que créerait en M une charge ponctuelle q', située sur l'axe $o\vec{z}$. D'autre part, ce potentiel est proportionnel à la polarisation de l'isolant, donc au champ créé par la charge q, donc à la charge q, soit $q'=kq$. Pour calculer le potentiel image, on peut donc remplacer l'isolant par une charge fictive ponctuelle de valeur $q'=kq$ portée par la droite $o\vec{z}$. On peut situer cette charge à une abscisse quelconque, le coefficient k prendra une valeur en

conséquence. Par raison de symétrie, on situe cette charge à l'abscisse $-d$, elle est appelée *charge image* de la charge q. Le potentiel image est par conséquent le potentiel créé en M , par la charge image q' , soit

$$\phi_{im}(M) = \frac{1}{4\pi\varepsilon_s}\ \frac{kq}{2d} \tag{10-65}$$

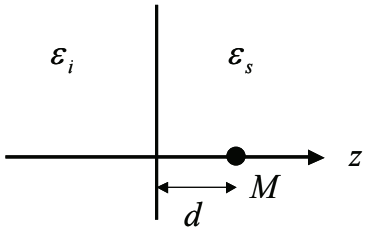


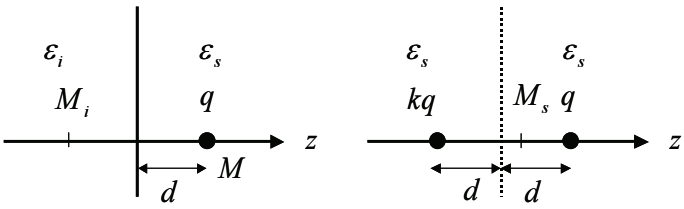
Figure 10-8 : Potentiel

Il s'agit maintenant de calculer le coefficient k. Pour ce faire, il suffit d'écrire que le potentiel et la composante normale du vecteur déplacement sont continus à l'interface isolant-semiconducteur, soit

$$\phi_i(z \to 0) = \phi_{sc}(z \to 0) \tag{10-66-a}$$

$$\varepsilon_i \left(\frac{d\phi_i}{dz} \right)_{z \to 0} = \varepsilon_s \left(\frac{d\phi_{sc}}{dz} \right)_{z \to 0} \tag{10-66-b}$$

Considérons successivement, le système réel, constitué par la structure isolant-semiconducteur, c'est-à-dire par deux milieux de constantes diélectriques ε_i et ε_s , avec une charge q située à l'abscisse d dans le semiconducteur (Fig. 10-9-a), et le système équivalent, constitué d'un milieu unique de constante diélectrique ε_s avec deux charges q et kq, situées respectivement aux abscisses d et $-d$ (Fig. 10-9-b).



a) Système réel

a) Système équivalent

Figure 10-9 : Potentiel image

Dans le système réel, le potentiel en un point M_i d'abscisse z dans l'isolant est proportionnel à la charge q et peut s'écrire

$$\phi_{Mi} = \frac{1}{4\pi\varepsilon_i} \frac{k'q}{d-z} \tag{10-67}$$

z étant l'abscisse du point M_i est négatif. Le coefficient k' traduit la discontinuité du milieu (pour un milieu continu k'=1). Dans le système équivalent, le potentiel créé en un point M_s d'abscisse z dans le semiconducteur résulte des charges q et kq, ce potentiel est donné par

$$\phi_{Ms} = \frac{1}{4\pi\varepsilon_s}\left(\frac{q}{d-z} + \frac{kq}{d+z}\right) \tag{10-68}$$

z est ici positif.

Ecrivons les conditions de continuité en faisant z=0, la condition (10-66-a) s'écrit

$$\frac{1}{4\pi\varepsilon_i}\frac{k'q}{d} = \frac{1}{4\pi\varepsilon_s}\left(\frac{q}{d} + \frac{kq}{d}\right) \tag{10-69}$$

ce qui donne la relation

$$\frac{k'}{\varepsilon_i} = \frac{1+k}{\varepsilon_s} \tag{10-70}$$

Pour écrire la condition (10-66-b), il faut dériver par rapport à z les expressions du potentiel, remarquons que $d/dz=d/d(d+z)=-d/d(d-z)$, ce qui donne

$$4\pi\varepsilon_i \frac{d\phi_{Mi}}{dz} = \frac{k'q}{(d-z)^2} \qquad 4\pi\varepsilon_s \frac{d\phi_{Ms}}{dz} = \frac{q}{(d-z)^2} - \frac{kq}{(d+z)^2} \tag{10-71}$$

En z=0

$$4\pi\varepsilon_i \frac{d\phi_{Mi}}{dz} = \frac{k'q}{d^2} \qquad 4\pi\varepsilon_s \frac{d\phi_{Ms}}{dz} = \frac{q}{d^2} - \frac{kq}{d^2} \tag{10-72}$$

Ainsi la condition de continuité donne

$$\frac{k'q}{d^2} = \frac{q}{d^2} - \frac{kq}{d^2}$$

soit

$$k' = 1-k \tag{10-73}$$

Les relations (10-70 et 73) entraînent

$$k = \frac{\varepsilon_s - \varepsilon_i}{\varepsilon_s + \varepsilon_i} \qquad k' = \frac{2\varepsilon_i}{\varepsilon_s + \varepsilon_i} \tag{10-74}$$

Le potentiel image est alors donné par

$$\phi_{im}(z) = \frac{1}{4\pi\varepsilon_s}\frac{\varepsilon_s - \varepsilon_i}{\varepsilon_s + \varepsilon_i}\frac{q}{2z} = -\frac{1}{4\pi\varepsilon_s}\frac{\varepsilon_s - \varepsilon_i}{\varepsilon_s + \varepsilon_i}\frac{e}{2z} \qquad (10\text{-}75)$$

L'énergie potentielle correspondante est donc

$$V_{im}(z) = \frac{1}{4\pi\varepsilon_s}\frac{\varepsilon_s - \varepsilon_i}{\varepsilon_s + \varepsilon_i}\frac{e^2}{2z} \qquad (10\text{-}76)$$

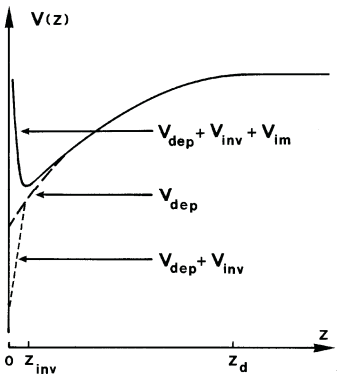


Figure 10-10 : Energie potentielle

Il faut noter que la constante diélectrique du semiconducteur est toujours supérieure à celle de l'isolant, le potentiel image est donc négatif, c'est-à-dire répulsif pour les électrons. En conséquence ce potentiel a tendance à éloigner les électrons de l'interface. L'allure de la variation avec z de l'énergie potentielle $V(z)$ est représentée sur la figure (10-10) avec la contribution de chacun des termes.

10.1.3. *Méthodes de calcul dans l'approximation de Hartree*

L'approximation de Hartree consiste à traiter le problème à un électron, c'est-à-dire à ramener l'interaction électron-électron à l'action sur un électron, du potentiel moyen résultant de la présence de l'ensemble des autres électrons. La contribution de cette interaction à l'énergie potentielle de l'électron est alors représentée par le terme $V_{im}(z)$ calculé au paragraphe précédent. La structure des états électroniques dans la couche d'inversion est alors donnée par la résolution de l'équation de Schrödinger (10-9), dans laquelle $V(z)$ est explicité à partir des résultats du paragraphe précédent.

a. Approximation du potentiel triangulaire

Si on réduit l'énergie potentielle $V(z)$ à $V_{dep}(z)+V_{inv}(z)$ en négligeant le potentiel image, on constate que le potentiel varie linéairement au voisinage de l'interface ($z\rightarrow 0$) et se courbe quand z augmente, pour tendre vers une valeur constante dans la région neutre du semiconducteur. Le potentiel d'inversion devient constant quand z devient de l'ordre de quelques z_m , le potentiel de déplétion devient constant pour $z=z_d$. Les ordres de grandeur de z_m et z_d sont respectivement de 30 Å et 1 µm, de sorte que, dans la région où se situent les charges d'inversion, les valeurs de z sont telles que $z \approx z_m << z_d$. On peut donc simplifier l'expression de $V_{dep}(z)$ que voient les électrons

$$V_{dep}(z) = \frac{e^2 N_{dep}}{\varepsilon_s} z \left(1 - \frac{z}{2 z_d}\right) \approx \frac{e^2 N_{dep}}{\varepsilon_s} z \qquad (10\text{-}77)$$

L'énergie potentielle résultant des charges de déplétion, varie linéairement avec z.

En ce qui concerne le potentiel associé aux charges d'inversion $\phi_{inv}(z)$, le problème est moins simple car ce potentiel varie dans la zone d'inversion. Toutefois si la charge d'inversion est faible devant la charge de déplétion on peut négliger la contribution de $\phi_{inv}(z)$ et écrire l'énergie potentielle des électrons sous la forme

$$V(z) = e\,E\,z \qquad (10\text{-}78)$$

où $E = e\,N_{dep} / \varepsilon_s$ représente le champ électrique dans le semiconducteur. Ce champ est constant au voisinage de l'interface dans la mesure où le potentiel varie linéairement.

Si la charge d'inversion n'est pas négligeable devant la charge de déplétion, on écrit l'énergie potentielle sous la forme approchée

$$V(z) = e\,E_{eff}\,z \qquad (10\text{-}79)$$

avec

$$E_{eff} = \frac{e(N_{dep} + f\,n)}{\varepsilon_s} \qquad (10\text{-}80)$$

E_{eff} représente le champ électrique effectif et f est un coefficient numérique pondérant la participation moyenne des charges d'inversion. $f = 0$ correspond à la seule contribution des charges de déplétion , $f = 1$ donne à E_{eff} la valeur du champ électrique à l'interface, $f = 1/2$ donne à E_{eff} une valeur moyenne à l'intérieur de la zone d'inversion.

La forme correspondante du puits de potentiel est représentée sur la figure (10-11). Il s'agit d'un puits triangulaire dissymétrique, avec une barrière verticale infinie en $z = 0$, et une barrière variant linéairement pour $z > 0$, c'est

l'approximation du puits triangulaire. Cette approximation est d'autant plus justifiée que le semiconducteur est peu dopé, ce qui entraîne z_d grand, et que la masse effective des électrons est plus grande, ce qui entraîne une faible extension spatiale des fonctions d'onde électroniques et par suite une faible valeur de z_m.

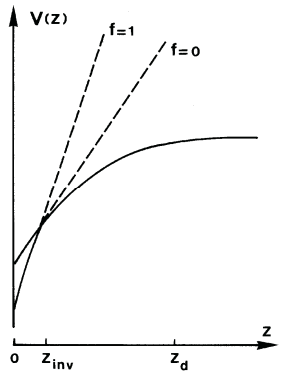


Figure 10-11 : Potentiel triangulaire

Dans ce modèle on peut résoudre l'équation (10-9) avec les conditions aux limites sur la fonction d'onde

$$\xi_i(z=0)=0 \qquad \xi_i(z\to\infty)=0 \tag{10-81}$$

L'équation s'écrit

$$\frac{d^2\xi_i(z)}{dz^2}-\frac{2m_e}{\hbar^2}\left(ezE_{eff}-E_i\right)\xi_i(z)=0 \tag{10-82}$$

on pose

$$\frac{2m_e}{\hbar^2}(eE_{eff}z-E_i)=\left(\frac{2m_e eE_{eff}}{\hbar^2}\right)^{2/3}u \tag{10-83}$$

Ainsi

$$\frac{d}{dz}=\frac{d}{du}\frac{du}{dz} \quad \frac{d^2}{dz^2}=\frac{du}{dz}\frac{d}{du}\left(\frac{du}{dz}\frac{d}{du}\right)=\left(\frac{du}{dz}\right)^2\frac{d^2}{du^2}$$

Soit

$$\frac{d^2}{dz^2}=\left(\frac{2m_e eE_{eff}}{\hbar^2}\right)^{2/3}\frac{d^2}{du^2} \tag{10-84}$$

En portant les expressions (10-83 et 84) dans l'équation (10-82) on obtient

$$\frac{d^2\xi_i(u)}{du^2}-u\,\xi_i(u)=0 \tag{10-85}$$

Les solutions de cette équation sont des fonctions d'Airy

$$\xi_i(u)=Ai(u) \quad avec \quad u=\left(\frac{2m_e eE_{eff}}{\hbar^2}\right)^{1/3}\left(z-\frac{E_i}{eE_{eff}}\right)$$

soit

$$\xi_i(z)=Ai\left(\left(\frac{2m_e eE_{eff}}{\hbar^2}\right)^{1/3}\left(z-\frac{E_i}{eE_{eff}}\right)\right) \tag{10-86}$$

Les valeurs propres sont de la forme

$$E_i \approx \left(\frac{\hbar^2}{2m_e}\right)^{1/3}\left(\frac{3\pi e E_{eff}}{2}(i+\frac{3}{4})\right)^{2/3} \qquad (10\text{-}87)$$

$i = 0, 1, 2, ...$

Les solutions exactes sont obtenues en remplaçant $(i + 3/4)$ par 0,7587, 1,7540 et 2,7575 pour $i = 0,1$ et 2 respectivement. Pour $E_{eff} = 10^5$ V/cm on obtient

$$E_i(meV) = 44\left(\frac{m_o}{m_e}\right)^{1/3}\left(i+\frac{3}{4}\right)^{2/3} \qquad (10\text{-}88)$$

Pour GaAs,	m_e=0,067 m_o	E_0=88 meV	E_1=155 meV	E_2=210 meV
Pour InSb,	m_e=0,013 m_o	E_0=154 meV	E_1=272 meV	E_2=367 meV

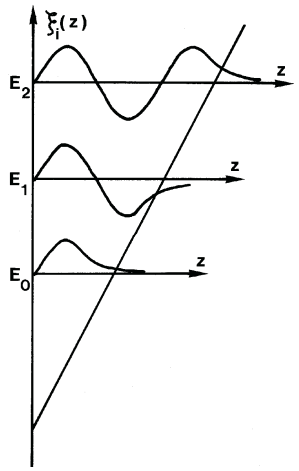


Figure 10-12 : Fonctions d'onde

On constate que l'intervalle entre les différentes sous-bandes d'énergie varie beaucoup d'un semiconducteur à l'autre en raison des différences importantes de masse effective. Les fonctions d'onde sont des fonctions d'Airy, qui s'annulent à l'interface et sont évanescentes dans le semiconducteur. L'indice i correspond au nombre de noeuds de la fonction d'onde. Les fonctions d'onde correspondant aux trois premières sous-bandes sont représentées sur la figure (10-12).

Le cas du silicium, comme celui du germanium, est très spécifique quant à la structure de sous-bandes d'énergie en raison de la nature multivallée de la bande de conduction. Rappelons que le silicium est un semiconducteur à structure de bande indirecte, le minimum de la bande de conduction n'est pas situé, dans l'espace des $\dot{k}$, au même point que le sommet de la bande de valence. Celui-ci est situé en k=0 alors que celui-là se situe sur l'axe [001] en k_m=0,85 k_o où k_o représente la longueur de la première zone de Brillouin dans la direction [001]. Dans la mesure où, en raison de la symétrie cubique du réseau du silicium, il existe 6 directions équivalentes, $[100], [\bar{1}00], [010]$,..., il existe 6 minima équivalents de la bande de conduction. Ces minima sont communément appelés des vallées, on dit que le silicium est un *semiconducteur multivallée* (Par. 1.2.3.a).

Dans un semiconducteur à gap direct au contraire le minimum de la bande de conduction se situe en $k = 0$, il est de ce fait unique, le semiconducteur est dit *univallée*. Dans ce cas, les surfaces d'énergie constante au voisinage du minimum sont des sphères centrées en k_m=0, l'énergie d'un électron de la bande de conduction est donnée par

$$E=E_c+\frac{\hbar^2 k^2}{2m_e} \quad avec\, k^2=k_x^2+k_y^2+k_z^2 \tag{10-89}$$

Dans le silicium, les surfaces d'énergie constante au voisinage de chacun des minima sont, pour des raisons de symétrie, des ellipsoïdes de révolution autour de l'axe considéré, centrés en $\mathbf{k}_m$. L'énergie d'un électron de la vallée centrée en $(0,0,k_m)$ par exemple, est donnée par (Eq. 1-69)

$$E=E_c+\frac{\hbar^2}{2m_t}(k_x^2+k_y^2)+\frac{\hbar^2}{2m_\ell}(k_z-k_m)^2 \tag{10-90}$$

avec $\quad m_\ell=\dfrac{\hbar^2}{\partial^2 E/dk_z^2} \qquad m_t=\dfrac{\hbar^2}{\partial^2 E/dk_x^2}=\dfrac{\hbar^2}{\partial^2 E/dk_y^2}$

m_l et m_t sont respectivement appelées *masse effective longitudinale* et . Leurs valeurs mesurées expérimentalement sont, dans le silicium

$$m_l=0{,}916\ m_o \qquad m_t=0{,}191\ m_o$$

Ces valeurs très différentes vont entraîner l'existence de deux séries de sous-bandes d'énergie. Considérons en effet une structure MIS dont le plan est perpendiculaire à la direction cristallographique [001] (Fig. 10-13). Dans leur mouvement perpendiculaire à l'interface, mouvement décrit par l'équation (10-82), les électrons des vallées $[0,0,k_m]$ sont caractérisés par leur masse effective longitudinale m_l . Par contre les électrons des vallées $[k_m,0,0]$ et $[0,k_m,0]$ sont caractérisés par leur masse effective transverse m_t. Ainsi, dans l'approximation de la masse effective, il existe deux équations distinctes et par suite deux séries distinctes de sous-bandes d'énergie, E_0, E_1, E_2... et E'_0, E'_1, E'_2,..., données par

$$E_i=\left(\frac{\hbar^2}{2m_\ell}\right)^{1/3}\left(\frac{3\pi e E_{eff}}{2}(i+\frac{3}{4})\right)^{2/3} \tag{10-91-a}$$

$$E'_i=\left(\frac{\hbar^2}{2m_t}\right)^{1/3}\left(\frac{3\pi e E_{eff}}{2}(i+\frac{3}{4})\right)^{2/3} \tag{10-91-b}$$

En raison de la différence importante qui existe entre m_l et m_t, les sous-bandes d'énergie E'_i ont des énergies beaucoup plus importantes que les sous-bandes d'énergie E_i. Pour E_{eff}= 10^5 V/cm, on obtient en meV

E_0=38 E_1=66 E_2=89 E'_0=63 E'_1=111 E'_2=150

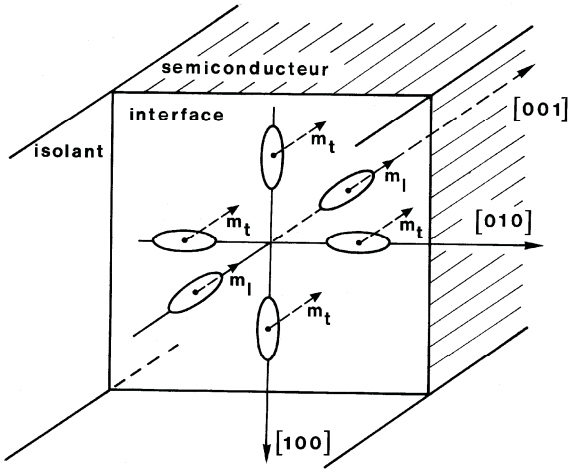


Figure 10-13 : Structure MIS dans le cas du silicium

Il faut remarquer que si la structure opère à basse température en particulier, tous les électrons se situent dans la sous-bande de plus basse énergie, c'est-à-dire dans la sous-bande E_0. Ces électrons sont caractérisés par leur masse longitudinale dans leur mouvement perpendiculaire à l'interface. Parallèlement, leur mouvement dans le plan de la structure est spécifique de leur masse transverse, et par suite leur mobilité $\mu=e\ \tau/m$ dans ce mouvement est importante. Or le mouvement dans le plan de la structure est précisément celui qui est mis à profit dans le fonctionnement des transistors MOSFET et des CCD.
Ainsi en abaissant la température, on localise les électrons dans la première sous-bande d'énergie où ils ont une mobilité, dans le plan, très importante.

Il faut noter que cette double série de sous-bandes existe aussi dans le germanium ou le GaP, qui sont des semiconducteurs indirects, mais n'existe pas dans les semiconducteurs à gap direct, tels que GaAs, InAs ou InP.

b. Calcul auto-cohérent

Dans l'approximation du potentiel triangulaire, nous avons linéarisé l'équation de Schrödinger en ramenant le potentiel de Hartree, associé aux charges d'inversion, à un potentiel proportionnel à la densité de charges d'inversion mais indépendant de la distribution spatiale de ces charges. Il s'agit là d'une

approximation relativement osée dans la mesure où l'expression même de $V_{inv}(z)$ fait apparaître une variation explicite avec la fonction d'onde $\xi_i(z)$. L'introduction de l'expression correcte de $V_{inv}(z)$ dans l'équation de Schrödinger, se traduit par une équation non linéaire qui n'admet pas de solution analytique, il faut alors traiter le problème numériquement.

On résout l'équation de manière auto-cohérente par un processus itératif. On choisit un potentiel d'essai, on obtient alors les fonctions d'ondes électroniques à partir desquelles on modifie le potentiel d'essai et on recommence le calcul. La solution numérique du problème est obtenue quand le potentiel calculé est suffisamment proche du potentiel d'essai. La méthode la plus simple consiste à prendre comme potentiel d'essai pour résoudre la $(n+1)^{ème}$ itération, une combinaison linéaire du potentiel d'essai et du potentiel calculé de l'itération précédente, soit

$$V_{ess}^{n+1}(z)=V_{ess}^{n}(z)+k\left(V_{cal}^{n}(z)-V_{ess}^{n}(z)\right) \qquad (10\text{-}92)$$

Il existe différentes méthodes empiriques ou systèmatiques, permettant de choisir le coefficient k. Précisons simplement que k=1 consiste à prendre comme potentiel d'essai de la $(n+1)^{ème}$ itération, le potentiel calculé à l'issue de la $n^{ème}$ itération. La méthode est rapide mais peut poser des problèmes de convergence. Inversement une faible valeur de k est plus sûre, mais entraîne un nombre d'itérations plus important et par suite un calcul beaucoup plus long.

Les résultats de calculs auto-cohérents concernant la couche d'inversion dans du silicium de type p, sont représentés sur les figures (10-14). Ces résultats ont été obtenus en négligeant le potentiel image.

La figure (10-14-a) représente le diagramme énergétique et la distribution spatiale des électrons, à basse température, pour une densité superficielle de charge de déplétion N_{dep}=5.10^{10} cm^{-2}, et une densité superficielle de charge d'inversion n = 10^{12} cm^{-2}. L'origine des énergies est prise sur le niveau E_0, qui représente le bas de la première sous-bande de conduction. La variation de l'opérateur énergie potentielle des électrons, est représentée par la courbe E_c qui représenterait le bas de la bande de conduction en l'absence de quantification c'est-à-dire dans le modèle classique. La courbe $\xi_0^2(z)$ représente la distribution spatiale des électrons qui, à basse température, sont tous dans la première sous-bande. La position moyenne de la couche d'inversion est représentée par z_m, qui est ici de l'ordre de 30 Å. Lorsque la température est plus élevée, les électrons peuplent davantage les bandes supérieures E_1, E_2,..., où leur fonction d'onde est plus étendue. Ceci se traduit par une augmentation de leur distance moyenne à l'interface, z_m. La figure (10-14-c) représente les variations avec la température, de la position moyenne des électrons de la première sous-bande z_{mo}. z_{mo} reste sensiblement constant alors que z_m passe de 30 Å à basse température à 80 Å à la température ambiante.

La figure (10-14-b) présente une comparaison entre les distributions spatiales des électrons, obtenues dans le modèle quantique et dans le modèle classique à la température de 150K. On constate en particulier que les ordres de grandeur des extensions spatiales sont comparables, néanmoins cette extension spatiale est plus

grande dans le calcul quantique et son maximum est déplacé vers l'intérieur du semiconducteur.

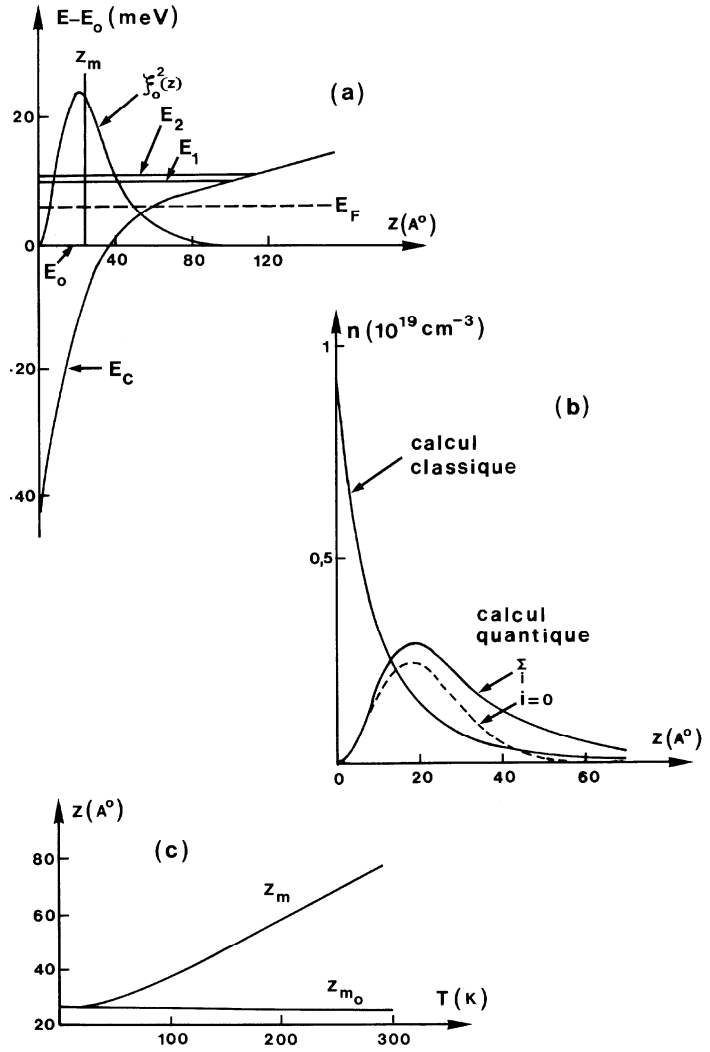


Figure 10-14 : Résultat d'un calcul auto-cohérent dans le silicium de type p. a) Diagramme énergétique et distribution spatiale des électron s. b) Comparaison du modèle quantique et du calcul classique. c) Position moyenne des électrons. ***(F. Stern 1967,1972)***

c. Fonction d'onde variationnelle

Les approximations précédentes permettent d'exprimer les fonctions d'onde électroniques, sous forme numérique en ce qui concerne le calcul auto-cohérent, et sous forme de fonctions d'Airy en ce qui concerne l'approximation du potentiel triangulaire. Ces formes sont peu pratiques lorsqu'il s'agit d'exploiter les résultats du calcul dans le but d'étudier les propriétés des électrons de la couche d'inversion, et plus particulièrement leur réponse à une perturbation extérieure. Il est alors utile de disposer de fonctions analytiques simples, représentant les fonctions d'onde des électrons de façon approximative mais d'emploi pratique.

Les plus simples de ces fonctions, en ce qui concerne la première sous-bande d'énergie, sont de la forme

$$\xi_o(z)=\left(\frac{b^3}{2}\right)^{1/2} z e^{-bz/2} \tag{10-93-a}$$

ou

$$\xi_o(z)=\left(\frac{3b^3}{2}\right)^{1/2} z e^{-(bz)^{3/2}/2} \tag{10-93-b}$$

Le paramètre variationnel b, est obtenu en minimisant l'énergie totale du système pour des valeurs données des densités de charge de déplétion et d'inversion. Pour des raisons de simplicité de calcul nous utiliserons la fonction variationnelle (10-93-a). La pénétration moyenne des électrons dans le semiconducteur est donnée par

$$z_{mo}=\int_0^\infty z\left|\xi_o(z)\right|^2 dz=\frac{b^3}{2}\int_0^\infty z^3 e^{-bz}\,dz \tag{10-94}$$

Dans la suite de ce paragraphe nous aurons à calculer plusieurs fois ce type d'intégrale avec différents exposants pour z, une intégration par parties montre sans difficulté que ce type d'intégrale donne

$$\int_0^\infty z^n e^{-\alpha z}\,dz=\frac{n!}{\alpha^{n+1}} \tag{10-95}$$

Ce qui donne pour l'expression (10-94)

$$z_{mo}=\frac{b^3}{2}\frac{3!}{b^4}=\frac{3}{b} \tag{10-96}$$

Ainsi, le paramètre $3/b$ représente la pénétration moyenne des électrons de la couche d'inversion dans le semiconducteur. Nous allons calculer le paramètre b en minimisant l'énergie totale du système d'électrons.

En raison de la simplicité de la fonction d'onde, on peut calculer aisément les éléments de matrice de chacun des termes de l'Hamiltonien, c'est-à-dire les valeurs propres de l'énergie cinétique $\langle T \rangle$ et de l'énergie potentielle $\langle V \rangle$. L'énergie totale de l'électron est alors donnée par $E = \langle T \rangle + \langle V \rangle$ avec

$$\langle T \rangle = \left\langle \xi_o(z) \left| \frac{-\hbar^2}{2m_e} \frac{d^2}{dz^2} \right| \xi_o(z) \right\rangle \tag{10-97-a}$$

$$\langle V \rangle = \langle \xi_o(z) | V | \xi_o(z) \rangle \tag{10-97-b}$$

Dans la mesure où $\xi_o(z)$ est réelle, et nulle dans l'isolant, c'est-à-dire pour z négatif, ces quantités s'écrivent

$$\langle T \rangle = \frac{\hbar^2}{2m_e} \int_0^\infty \xi_o(z) \xi''_o(z) dz \tag{10-98-a}$$

$$\langle V \rangle = \int_0^\infty \left(V_{dep}(z) + V_{inv}(z) + V_{im}(z) \right) \xi_o^2(z) dz \tag{10-98-b}$$

- *Energie cinétique*

Il faut calculer la dérivée seconde de la fonction d'onde, soit

$$\xi'_o(z) = \left(\frac{b^3}{2} \right)^{1/2} \left(1 - \frac{bz}{2} \right) e^{-bz/2} \tag{10-99}$$

$$\xi''_o(z) = \left(\frac{b^5}{2} \right)^{1/2} \left(-1 + \frac{bz}{4} \right) e^{-bz/2} \tag{10-100}$$

Ainsi en explicitant le produit $\xi_o(z)\, \xi''_o(z)$, l'énergie cinétique s'écrit

$$\langle T \rangle = \frac{\hbar^2}{2m_e} \frac{b^4}{2} \int_0^\infty z \left(1 - \frac{bz}{4} \right) e^{-bz} dz$$

$$= \frac{\hbar^2 b^4}{4m_e} \left(\int_0^\infty z e^{-bz} dz - \frac{b}{4} \int_0^\infty z^2 e^{-bz} dz \right) \tag{10-101}$$

Compte tenu de (10-95) la première intégrale de (10-101) donne $1/b^2$ et la deuxième $2/b^3$, de sorte que $\langle T \rangle$ s'écrit

$$\langle T \rangle = \frac{\hbar^2 b^2}{8m_e} \tag{10-102}$$

- *Energie potentielle*

L'énergie potentielle résulte de plusieurs contributions que l'on va calculer successivement

i) calcul du terme $\langle V_{dep} \rangle$

$$\langle V_{dep} \rangle = \frac{b^3 e^2 N_{dep}}{2 \varepsilon_s} \int_0^\infty z\left(1 - \frac{z}{2 z_d}\right) z^2 e^{-bz} dz$$

$$\langle V_{dep} \rangle = \frac{b^3 e^2 N_{dep}}{2 \varepsilon_s} \left(\int_0^\infty z^3 e^{-bz} dz - \frac{1}{2 z_d} \int_0^\infty z^4 e^{-bz} dz \right) \qquad (10\text{-}103)$$

La première intégrale est égale à $6/b^4$ et la seconde à $24/b^5$, ce qui donne

$$\langle V_{dep} \rangle = \frac{3 e^2 N_{dep}}{\varepsilon_s b} \left(1 - \frac{2}{b z_d} \right) \qquad (10\text{-}104)$$

Rappelons que $b = 3/z_m$ de sorte que

$$\langle V_{dep} \rangle = \frac{3 e^2 N_{dep}}{\varepsilon_s b} \left(1 - \frac{2}{3} \frac{z_{mo}}{z_d} \right) \qquad (10\text{-}105)$$

Dans la mesure où z_{mo} est de l'ordre de quelques dizaines d'Angströms, et z_d de l'ordre du micron, on peut négliger le deuxième terme dans l'expression de $\langle V_{dep} \rangle$ et écrire avec une bonne approximation

$$\langle V_{dep} \rangle = \frac{3 e^2 N_{dep}}{\varepsilon_s b} \qquad (10\text{-}106)$$

ii) calcul du terme $\langle V_{inv} \rangle$

$$V_{inv} = \frac{b^3 e^2}{2 \varepsilon_s} \int_0^\infty \sum_i n_i \left(z + \int_0^z (z' - z) \left| \xi_i(z') \right|^2 dz' \right) z^2 e^{-bz} dz \qquad (10\text{-}107)$$

Dans la mesure où on limite le calcul à la première sous-bande, et dans la mesure où c'est la seule peuplée, la somme sur i se limite à cette sous-bande et $\sum_i n_i = n_0 = n_s$. L'expression s'écrit en explicitant $\xi_0(z')$

$$V_{inv} = \frac{b^3 e^2 n_s}{2 \varepsilon_s} \int_0^\infty \left(z + \frac{b^3}{2} \int_0^z (z' - z) z'^2 e^{-bz'} dz' \right) z^2 e^{-bz} dz \qquad (10\text{-}108)$$

Calculons tout d'abord l'intégrale sur z'

$$I=\int_0^z z'^3 e^{-bz'}dz'-z\int_0^z z'^2 e^{-bz'}dz'$$

$$=-\frac{1}{b}\left[z'^3 e^{-bz'}\right]_0^z+\frac{3}{b}\int_0^z z'^2 e^{-bz'}dz'-z\int_0^z z'^2 e^{-bz'}dz'$$

$$=-\frac{1}{b}z^3 e^{-bz}+\left(\frac{3}{b}-z\right)\int_0^z z'^2 e^{-bz'}dz'$$

$$I=-\frac{1}{b}z^3 e^{-bz}+\left(\frac{3}{b}-z\right)\left(-\frac{1}{b}\left[z'^2 e^{-bz'}\right]_0^z+\frac{2}{b}\int_0^z z'e^{-bz'}dz'\right)$$

$$=-\frac{1}{b}z^3 e^{-bz}+\left(\frac{3}{b}-z\right)\left(-\frac{1}{b}z^2 e^{-bz}+\frac{2}{b}\left(-\frac{1}{b}\left[z'e^{-bz'}\right]_0^z+\frac{1}{b}\int_0^z e^{-bz'}dz'\right)\right)$$

$$=-\frac{1}{b}z^3 e^{-bz}+\left(\frac{3}{b}-z\right)\left(-\frac{1}{b}z^2 e^{-bz}-\frac{2}{b^2}ze^{-bz}+\frac{2}{b^2}\left[-\frac{1}{b}e^{-bz}\right]_0^z\right)$$

$$I=\left(-\frac{z^2}{b^2}-\frac{4z}{b^3}-\frac{6}{b^4}\right)e^{-bz}-\frac{2z}{b^3}+\frac{6}{b^4} \qquad (10\text{-}109)$$

ce qui donne, en portant dans l'expression (10-108)

$$\langle V_{inv}\rangle=\frac{b^3 e^2 n_s}{2\ \varepsilon_s}\int_0^\infty\left(z+\frac{b^3}{2}\left(\left(-\frac{z^2}{b^2}-\frac{4z}{b^3}-\frac{6}{b^4}\right)e^{-bz}-\frac{2z}{b^3}+\frac{6}{b^4}\right)\right)z^2e^{-bz}dz$$

$$\langle V_{inv}\rangle=\frac{b^3e^2n_s}{2\varepsilon_s}\left(-\frac{b}{2}\int_0^\infty z^4e^{-2bz}dz-2\int_0^\infty z^3e^{-2bz}dz-\frac{3}{b}\int_0^\infty z^2e^{-2bz}dz+\frac{3}{b}\int_0^\infty z^2e^{-bz}dz\right) \qquad (10\text{-}110)$$

Compte tenu de (10-95), on obtient

$$\langle V_{inv}\rangle=\frac{b^3e^2n_s}{2\varepsilon_s}\left(-\frac{b}{2}\frac{24}{(2b)^5}-2\frac{6}{(2b)^4}-\frac{3}{b}\frac{2}{(2b)^3}+\frac{3}{b}\frac{2}{b^3}\right)$$

$$=\frac{b^3e^2n_s}{2\varepsilon_s}\frac{33}{8b^4}=\frac{33e^2n_s}{16\ \varepsilon_s b}$$

Soit

$$\langle V_{inv}\rangle\approx\frac{2e^2n_s}{\varepsilon_s b} \qquad (10\text{-}111)$$

iii) calcul du terme $\langle V_{im} \rangle$

$$\langle V_{im} \rangle = \frac{1}{4\pi\varepsilon_s} \frac{\varepsilon_s - \varepsilon_i}{\varepsilon_s + \varepsilon_i} \frac{e^2}{2} \frac{b^3}{2} \int_0^\infty z e^{-bz} dz \tag{10-112}$$

Soit

$$\langle V_{im} \rangle = \frac{e^2 b}{16\pi\varepsilon_s} \frac{\varepsilon_s - \varepsilon_i}{\varepsilon_s + \varepsilon_i} \tag{10-113}$$

- *Energie et fonction d'onde de la première sous-bande*

L'énergie totale d'un électron de la première sous-bande est donnée par la somme des termes précédents. Il faut minimiser l'énergie totale du système pour déterminer le paramètre variationnel b. Rappelons que le terme $\langle V_{inv} \rangle$ représente l'énergie d'interaction électron-électron, il faut donc le diviser par deux dans le calcul de l'énergie par électron, pour ne pas le compter deux fois. Ainsi l'énergie totale du système par électron est donnée par

$$E = \langle T \rangle + \langle V_{dep} \rangle + \frac{1}{2} \langle V_{inv} \rangle + \langle V_{im} \rangle \tag{10-114}$$

Si on néglige le potentiel image cette expression s'écrit en explicitant chacun des termes à partir de (10-102, 106 et 111)

$$E = \frac{\hbar^2 b^2}{8m_e} + \frac{3e^2 N_{dep}}{\varepsilon_s b} + \frac{e^2 n_s}{\varepsilon_s b}$$

$$E = \frac{\hbar^2 b^2}{8m_e} + \frac{3e^2}{\varepsilon_s b} \left(N_{dep} + n_s / 3 \right) \tag{10-115}$$

La dérivée de l'énergie s'écrit

$$\frac{dE}{db} = \frac{\hbar^2 b}{4m_e} - \frac{3e^2}{\varepsilon_s b^2} \left(N_{dep} + n_s / 3 \right) \tag{10-116}$$

Cette dérivée s'annule pour

$$\frac{\hbar^2 b}{4m_e} = \frac{3e^2}{\varepsilon_s b^2} \left(N_{dep} + n_s / 3 \right) \tag{10-117}$$

Soit

$$b = \left(\frac{12 e^2 m_e}{\hbar^2 \varepsilon_s} \left(N_{dep} + n_s / 3 \right) \right)^{1/3} \tag{10-118}$$

Ainsi, pour des valeurs données des densités superficielles de charges de déplétion et d'inversion, on peut calculer b. On en déduit l'extension moyenne z_{mo} des électrons dans le semiconducteur, l'énergie E_0 du bas de la première sous-bande de conduction et on trace la couche représentative de $\xi_0(z)$.

L'extension moyenne des électrons dans le semiconducteur est donnée par

$$z_{mo}=\frac{3}{b}=\left(\frac{27\hbar^2\varepsilon_s}{12m_e e^2\left(N_{dep}+n_s/3\right)}\right)^{1/3} \tag{10-119}$$

Pour des densités de charge $N_{dep}=10^{11}$ cm^{-2} et $n_s=10^{12}$ cm^{-2}, on obtient dans le cas du silicium $z_{mo}\approx 30$ Å. Ce résultat correspond à quelques pour-cent près, à ce que donne un calcul auto-cohérent.

L'énergie du bas de la première sous-bande de conduction est donnée en négligeant le potentiel image par

$$E_o=\langle T\rangle+\langle V_{dep}\rangle+\langle V_{inv}\rangle \tag{10-120}$$

En explicitant chacun des termes on obtient

$$E_o=\frac{\hbar^2}{8m_e}\left(\frac{12m_e e^2}{\hbar\varepsilon_s}\left(N_{dep}+n_s/3\right)\right)^{2/3}+\frac{3e^2\left(N_{dep}+2n_s/3\right)}{\varepsilon_s}\left(\frac{\hbar^2\varepsilon_s}{12m_e e^2\left(N_{dep}+n_s/3\right)}\right)^{1/3} \tag{10-121}$$

Cette expression se simplifie dans les cas limites où les densités superficielles de charges de déplétion et d'inversion sont très différentes

Pour $n_s<<N_{dep}$ $$E_o\approx 2\left(\frac{\hbar e^2}{\varepsilon_s m_e^{1/2}}N_{dep}\right)^{2/3} \tag{10-122-a}$$

Pour $n_s>>N_{dep}$ $$E_o\approx\left(\frac{2\hbar e^2}{\varepsilon_s m_e^{1/2}}n_s\right)^{2/3} \tag{10-122-b}$$

Dans le cas du silicium, avec $n_s<< N_{dep}=10^{11}$ cm^{-2} on obtient $E_0\approx 11$ meV, avec $N_{dep}<< n_s=10^{12}$ cm^{-2}, on obtient $E_0\approx 40$ meV.

Nous avons négligé la contribution du potentiel image, si on en tient compte la minimisation de l'énergie doit être faite numériquement, mais sans aucune difficulté. De manière générale la fonction d'onde variationnelle (10-93-a) donne des résultats en accord avec le calcul auto-cohérent à environ 7% près, la fonction d'onde (10-93-b), un peu plus élaborée, donne des résultats pratiquement équivalents à ceux résultant d'un calcul auto-cohérent.

d. Effets d'échange et corrélation

Dans ce qui précède nous avons étudié la structure de sous-bandes d'énergie de la couche d'inversion dans le modèle à un électron, c'est-à-dire en globalisant les interactions électron-électron par un potentiel de Hartree. Cette approximation est d'autant plus justifiée, que d'une part la densité d'électrons dans la couche d'inversion est élevée, et que d'autre part la masse effective de ces électrons est faible. L'approximation de Hartree est souvent justifiée dans les semiconducteurs à gap direct en raison de la faible valeur de la masse effective électronique. Dans InSb et InAs par exemple, m_e=0,012 m_0 et 0,023 m_0 respectivement, et la condition $r/a_0^* << 1$ est vérifiée pour des densités d'électrons de l'ordre de 10^{12} cm^{-2}. En ce qui concerne les semiconducteurs à gap indirect, comme le silicium, la masse effective longitudinale des électrons est par contre de l'ordre de m_0, de sorte que le paramètre r/a_0^* est plutôt supérieur à 1, pour des valeurs typiques de la densité d'électrons. En conséquence, l'interaction électron-électron ne peut plus être globalisée et les effets d'échange et corrélation jouent un rôle important. L'amplitude de ces effets devient comparable à la séparation des sous-bandes d'énergie issues d'un calcul de Hartree.

L'interaction coulombienne entre deux électrons localisés respectivement en (r,z) et $(r'z')$ où $r^2=x^2+y^2$ s'écrit

$$V(r-r',z,z') = \frac{e}{4\pi\varepsilon_s}\left((r-r')^2+(z-z')^2\right)^{-1/2} + \frac{e^2}{4\pi\varepsilon_s}\frac{\varepsilon_s-\varepsilon_i}{\varepsilon_s+\varepsilon_i}\cdot\left((r-r')^2+(z-z')^2\right)^{-1/2} \tag{10-123}$$

Le premier terme décrit l'interaction directe entre les deux électrons, le deuxième terme décrit l'interaction d'un électron avec la charge image de l'autre. Les effets d'échange et corrélation résultent de la sommation de ce terme d'interaction sur l'ensemble des électrons. Une façon de prendre en considération ces effets consiste à les traiter en perturbation sur la base des états résultant d'un calcul de Hartree. L'inconvénient du calcul de perturbation est évidemment qu'il n'est plus applicable lorsque la perturbation est du même ordre de grandeur que la distance inter-sous-bandes non perturbée. Rappelons que dans l'approximation de Hartree, les fonctions d'ondes des électrons des sous-bandes inférieures sont localisées près de l'interface isolant-semiconducteur. Celles des électrons des sous-bandes supérieures sont plus étalées vers l'intérieur du semiconducteur. Les effets d'échange et corrélation, en couplant les différents états, mélangent partiellement leurs fonctions d'onde. Ainsi, les électrons des sous-bandes inférieures se trouvent déplacés vers l'intérieur du semiconducteur par la présence dans leur fonction d'onde, d'une fraction des fonctions d'onde associées aux sous-bandes supérieures. A contrario, les électrons des sous-bandes supérieures sont ramenés vers l'interface isolant-semiconducteur. Il en résulte donc une contraction spatiale de la population électronique. Ceci provient du fait que l'approximation de Hartree surestime les

forces coulombiennes de répulsion électron-électron. Il faut noter que le potentiel image, qui a tendance à éloigner les électrons de l'interface, a un effet contraire et entraîne un étalement des fonctions d'onde électroniques vers l'intérieur du semiconducteur. Ainsi les effets d'échange et corrélation et du potentiel image se compensent partiellement, ceci justifie le fait que, dans un calcul de Hartree, on a intérêt à ne pas prendre en compte le potentiel image dans la mesure où les effets d'échange et corrélation sont négligés.

En ce qui concerne les énergies des différentes sous-bandes de conduction, il est bien connu que des états couplés se repoussent, dans le diagramme énergétique. Dans la mesure où les effets d'échange et corrélation couplent entre elles les différentes sous-bandes, ces sous-bandes vont se séparer davantage et les plus basses, tout particulièrement la première, vont descendre en énergie.

Ainsi, les effets d'échange et corrélation modifient les résultats issus d'un calcul de Hartree, par une contraction de la distribution spatiale des électrons, une dilatation du diagramme énergétique et un abaissement important des énergies des premières sous-bandes.

Une autre façon de prendre en considération les effets d'échange et corrélation, consiste à traiter le problème à un électron en ajoutant au potentiel de Hartree un potentiel additionnel V_{ec} , représentant un effet équivalent. La difficulté ici est d'expliciter ce potentiel d'échange et corrélation. La figure (10-15) permet de comparer les résultats d'un calcul auto-cohérent à un électron, avec un potentiel de Hartree (traits discontinus) et un potentiel de Hartree effectif.

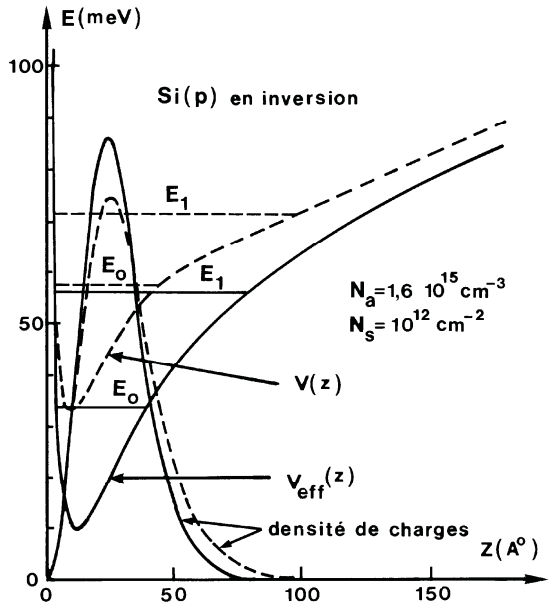


Figure 10-15 : Structure de sous-bandes d'énergie compte-tenu des effets d'échange et corrélation. *(T. Ando 1975)*

10.2. Puits quantiques

Les techniques modernes de croissance de cristaux semiconducteurs, que sont l'épitaxie par jet moléculaire MBE (Molecular Beam Epitaxy) et le dépôt en phase vapeur à partir d'organo-métalliques MOCVD (Metal-Organic Chemical Vapour Deposition), permettent de réaliser des couches monocristallines avec une maîtrise exceptionnelle de la composition chimique, des qualités cristallographiques et de l'épaisseur. Ces techniques permettent, à partir des éléments constitutifs d'un matériau, de déposer ce matériau sur un substrat convenablement choisi, pratiquement couche atomique par couche atomique. Les très faibles vitesses de croissance que l'on peut mettre en oeuvre permettent de réaliser des hétérostructures constituées par la juxtaposition de couches de matériaux différents et/ou différemment dopés, avec des profils de composition et/ou de dopage extrêmement abrupts. La discontinuité d'une couche à l'autre s'étend typiquement sur une épaisseur de l'ordre de la mono-couche atomique.

La condition nécessaire à la réalisation d'une bonne hétéro-épitaxie d'un semiconducteur SC_1 sur un semiconducteur SC_2 est évidemment que les deux matériaux aient d'une part la même structure cristalline et d'autre part des paramètres de maille voisins. Si les structures cristallines sont différentes, le dépôt est généralement amorphe ou dans le meilleur des cas polycristallin. Si les paramètres de maille sont différents, le matériau constituant la couche de plus grande épaisseur impose sa maille à l'autre, au moins au voisinage de l'interface. Ceci entraîne l'existence, dans le matériau de faible épaisseur, d'une contrainte biaxiale dans le plan de la couche. Si les couches sont d'épaisseurs comparables, la contrainte est répartie dans les deux matériaux. Dans l'un comme l'autre des cas précédents, on dit que l'hétérostructure est contrainte. Si les paramètres de maille sont trop différents, les contraintes sont telles que la densité de dislocations dans la structure devient rédhibitoire.

Différents types de semiconducteurs, tels que les composés III-V ou II-VI et leurs alliages respectifs, peuvent être épitaxiés les uns sur les autres pour former différents types d'hétérostructures. Les hétérostructures les plus simples sont d'une part les hétérojonctions entre deux semiconducteurs différents, et d'autre part les structures Schottky ou MIS dont les spécificités sont étudiées dans les paragraphes précédents. Nous allons étudier des hétérostructures plus élaborées.

Considérons une hétérostructure constituée d'une couche de semiconducteur SC_1 d'épaisseur L_1 en sandwich entre deux couches d'un semiconducteur SC_2 tel que $E_{g1}<E_{g2}$. La différence de gap ΔE_g est distribuée entre les bandes de conduction et de valence de deux manières différentes suivant la différence des affinités électroniques des semiconducteurs. Le diagramme énergétique de la structure est représenté sur la figure (10-16). Dans le cas de la figure (10-16-a), les extrema des bandes de valence et de conduction sont situés dans le même matériau, c'est-à-dire dans la même région de l'espace ; ces hétérostructures sont dites de

type I. Dans le cas de la figure (10-16-b), ces extrema sont spatialement séparés, ces hétérostructures sont dites de *type II.* Il est évident que ces deux types d'hétérostructure présentent des propriétés différentes. Si des porteurs sont injectés dans la structure, ces porteurs sont confinés dans les puits de potentiel que constituent les extrema des bandes. Dans l'hétérostructure de type I les électrons et les trous sont piégés dans le même semiconducteur, ici SC_1. Dans l'hétérostructure de type II les électrons et les trous sont spatialement séparés. Dans le premier cas leurs recombinaisons seront importantes, dans le deuxième cas ces recombinaisons seront beaucoup moins probables.

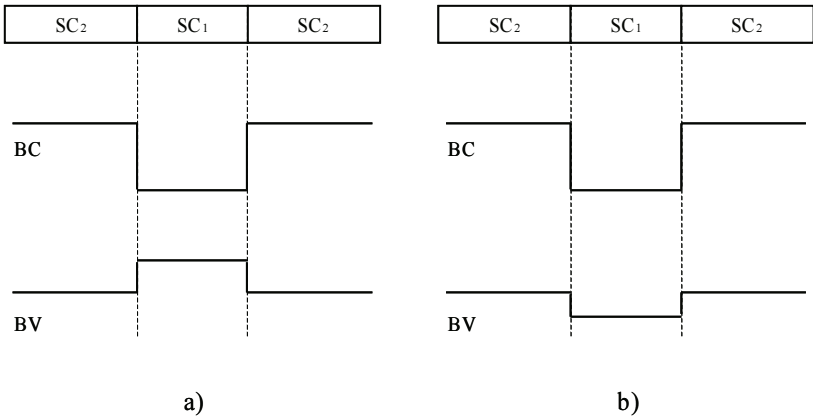


Figure 10-16 : Configuration des bandes de conduction et de valence dans une hétérostructure de type I (a) et une hétérostructure de type II (b)

Considérons par exemple l'hétérostructure du type SC_2-SC_1-SC_2 avec SC_2=$Al_xGa_{1-x}As$ et SC_1=GaAs. On peut traduire la variation de la différence des gaps par l'expression

$$\Delta E_g = E_{g2} - E_{g1} \approx 1{,}36x + 0{,}22x^2$$

On peut d'autre part linéariser la variation d'énergie du sommet de la bande de valence par l'expression $\Delta E_v = E_{v2} - E_{v1} = \Delta E_{vm}\, x$ où $\Delta E_{vm} = E_v(AlAs) - E_v(GaAs)$.

En prenant le zéro de l'énergie au niveau du vide, on peut écrire E_v=-$e\chi$-E_g ce qui donne à partir des tableaux (1-1 et 4)

$$E_v(AlAs) = -3{,}5 - 2{,}24 = -5{,}74\ eV$$

$$E_v(GaAs) = -4{,}07 - 1{,}52 = -5{,}59\ eV$$

soit

$$\Delta E_{vm} = -0{,}15\ eV$$

Pour la composition correspondant à x=0,3 on obtient

$$\Delta E_g = 0,428\ eV \quad \Delta E_c = 0,383\ eV \quad \Delta E_v = -0,045\ eV$$

ΔE_c étant positif et E_v négatif, l'hétérostructure est de type I. Elle correspond au diagramme énergétique de la figure (10-16-a).

Un exemple d'hétérostructure de type II est donné par la structure SC_2-SC_1-SC_2 avec SC_2=GaSb et SC_1=InAs. Les tableaux (1-1 et 4) donnent respectivement

$$E_{g2} - E_{g1} = 0,81 - 0,42 = 0,39 \text{ eV}$$
$$\Delta(e\chi) = e\chi_2 - e\chi_1 = 4,06 - 4,90 = -0,84 \text{ eV}$$

on obtient donc

$$\Delta E_g = 0,39 \text{ eV} \qquad \Delta E_c = 0,84 \text{ eV} \qquad \Delta E_v = 0,45 \text{ eV}$$

ΔE_c et ΔE_v sont de même signe, l'hétérostructure est de type II. Mais cette hétérostructure présente en outre une spécificité toute particulière résultant du fait que la bande de conduction dans InAs a une énergie inférieure à la bande de valence dans GaSb (Fig. 10-17). Il en résulte que les électrons liés de la bande de valence de GaSb se libèrent dans la bande de conduction de InAs. La structure présente un comportement semimétallique.

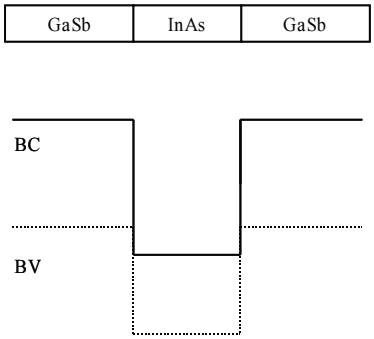


Figure 10-17 : Hétérostructure GaSb-InAs-GaSb

Revenons sur l'hétérostructure de type I et supposons que l'on dope avec des donneurs le semiconducteur à grande bande interdite SC_2. En raison de la différence d'énergie des bandes de conduction, les électrons fournis par les atomes donneurs, quittent la bande de conduction du semiconducteur SC_2 pour s'accumuler dans le puits de potentiel que constitue la bande de conduction du semiconducteur SC_1. L'intérêt évident de ce type de structure est de disposer d'un matériau, le semiconducteur SC_1, fortement dégénéré sans pour autant être dopé. La séparation spatiale des porteurs libres et des centres diffusants, que sont les donneurs ionisés, diminue considérablement les processus de diffusion et par suite permet d'obtenir

de grandes mobilités. Les conditions n et μ grands sont définies simultanément. Des mobilités aussi importantes que 2.10^6 $cm^2V^{-1}s^{-1}$ ont été obtenues dans ce type de structure à basse température. La conductivité $\sigma_{//}$ dans le plan de la structure est alors très importante.

En raison du dopage du semiconducteur SC_1, la double hétérojonction que constitue la structure présente une double courbure de bandes dont chacune est semblable à celle décrite sur la figure (6-4) pour une hétérojonction unique. Le diagramme énergétique du puits de potentiel est représenté sur la figure (10-18).

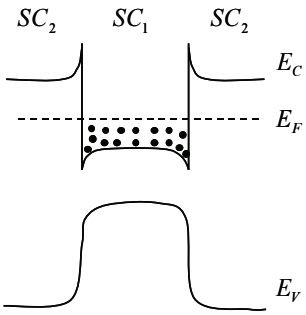


Figure 10-18 : Hétérostructure de type I à dopage sélectif

10.2.1. *Spectre d'énergie*

Si l'épaisseur L_1 du semiconducteur SC_1 est faible, typiquement L_1<200 Å, les états électroniques ne correspondent plus au bas de la bande de conduction, mais sont quantifiés en structure de sous-bandes d'énergie analogue à celle que nous avons étudiée dans le cas de la couche d'inversion d'une structure MOS. Le mouvement des électrons est quasi-libre dans le plan de la structure et quantifié dans la direction perpendiculaire. Comme dans le paragraphe (10.1.1), on peut séparer le mouvement dans le plan de la structure du mouvement dans la direction perpendiculaire et écrire la fonction d'onde des électrons sous la forme (Eq. 10-8)

$$\psi(r) = \xi(z)\, e^{i(k_x x + k_y y)}\, \varphi(x,y) \tag{10-124}$$

où $\varphi(x,y)$ est la fonction de Bloch et $\xi(z)$ une fonction enveloppe qui décrit la quantification du mouvement suivant z. Dans l'approximation de la masse effective le mouvement suivant z est régi par l'équation de Schrödinger

$$\frac{\hbar^2}{2m_e}\frac{d^2\xi(z)}{dz^2} + \big(E - V(z)\big)\xi(z) = 0 \tag{10-125}$$

L'énergie potentielle $V(z)$ est celle du puits carré à une dimension, défini, en prenant l'origine des énergies au bas de la bande de conduction du semiconducteur SC_1, par (Fig. 10-19)

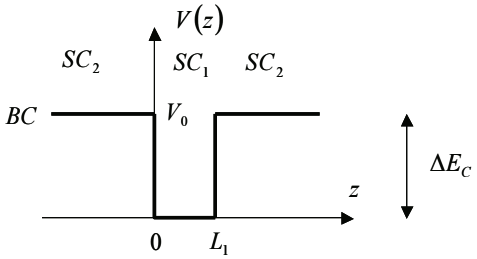


Figure 10-19 : Puits quantique

$$V(z) = \Delta E_c \qquad \textbf{pour} \quad z < 0 \quad \textbf{et} \quad z > L_1$$
$$V(z) = 0 \qquad \textbf{pour} \quad 0 < z < L_1$$

Les valeurs de l'énergie E décrivent la quantification des états électroniques dans la direction perpendiculaire à la structure. Dans le plan de la structure le mouvement des électrons n'est pas affecté. Il en résulte une structure de sous-bandes d'énergie avec une quantification discrète suivant k_z et une variation pseudo-continue suivant $k_{//} = \sqrt{k_x^2 + k_y^2}$. L'énergie totale d'un électron s'écrit

$$E(k) = E_c + E + \frac{\hbar^2 k_{//}^2}{2m_e} \tag{10-126}$$

Les énergies E des minima des différentes sous-bandes sont évidemment fonction de la profondeur et de la largeur du puits de potentiel.

a. Puits de profondeur infinie

Si ΔE_c est très important on peut supposer en première approximation que les électrons sont confinés dans l'espace $0 < z < L_1$ par des murs de potentiel de hauteur infinie. Le potentiel s'écrit alors

$$V(z) = \infty \qquad \text{pour } z < 0 \text{ et } z > L_1$$
$$V(z) = 0 \qquad \text{pour } 0 < z < L_1$$

Les conditions aux limites, définissant les constantes d'intégration de l'équation (10-125), sont par conséquent

$$\xi(z = 0) = 0 \qquad \xi(z = L_1) = 0$$

Dans le puits de potentiel $V(z)$=0, l'équation (10-125) s'écrit

$$\frac{d^2\xi(z)}{dz^2} + K^2\xi(z) = 0 \qquad (10\text{-}127)$$

avec $K = \sqrt{2m_e E}/\hbar$, où m_e représente la masse effective des électrons dans le semiconducteur SC_1. Les solutions de cette équation sont des sinusoïdes de la forme

$$\xi(z) = A\sin(Kz + \varphi) \qquad (10\text{-}128)$$

Les conditions aux limites permettent d'écrire

$$\xi(z=0) = 0 \qquad donc \quad \varphi = 0$$
$$\xi(z=L_1) = 0 \qquad donc \quad K = n\pi / L_1$$

Compte tenu de la définition de K, l'énergie est donnée par $E = \hbar^2 K^2/2m_e$ soit

$$E_n = n^2 \frac{\pi^2\hbar^2}{2m_e L_1^2} \qquad (10\text{-}129)$$

L'énergie totale des électrons dans le puits de potentiel s'écrit donc, compte tenu de (10-129 et 126),

$$E(k) = E_c + n^2 \frac{\pi^2\hbar^2}{2m_e L_1^2} + \frac{\hbar^2 k_{//}^2}{2m_e} \qquad (10\text{-}130)$$

Les énergies des minima des différentes sous-bandes de conduction varient comme n^2, avec n entier. Les courbes de dispersion sont représentées sur la figure (10-20). La quantification des niveaux varie comme $1/L_1^2$, c'est-à-dire en raison inverse du carré de la largeur du puits.

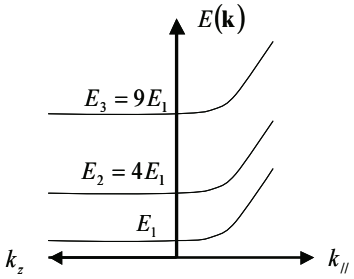


Figure 10-20 : Puits de profondeur infinie

Considérons par exemple la structure $Al_xGa_{1-x}As$-GaAs-$Al_xGa_{1-x}As$, avec x=0,3 et L_1=110 Å. La masse effective des électrons dans GaAs est m_e= 0,067 m_o , on obtient E_n= 48 n^2 (meV), soit

E_1=48 meV $\quad$ E_2=192 meV $\quad$ E_3=432 meV

Il est évident que l'approximation du puits infini est d'autant moins justifiée que n est grand, et que m_e et L_1 sont petits.

Considérons maintenant les fonctions d'onde des électrons dans les différentes sous-bandes de conduction. Dans la mesure où $\varphi=0$ et $K=n\pi/L_1$, l'expression (10-128) s'écrit

$$\xi_n(z) = A\sin\frac{n\pi}{L_1}z \qquad (10\text{-}131)$$

$\xi(z)$ étant nul à l'extérieur de l'intervalle $(0,L_1)$, la condition de normalisation s'écrit

$$\int_0^{L_1}\left|\xi_n(z)\right|^2 dz=1 \quad \text{soit} \quad A^2\int_0^{L_1}\sin^2\left(\frac{n\pi}{L_1}z\right)dz=1$$

ou

$$\frac{A^2}{2}\int_0^{L_1}\left(1-\cos\frac{2n\pi}{L_1}z\right)dz=1$$

Ainsi, $A=\sqrt{2/L_1}$ et la fonction enveloppe $\xi(z)$ s'écrit

$$\xi_n(z)=\sqrt{2/L_1}\sin\frac{n\pi}{L_1}z \qquad (10\text{-}132)$$

Ces fonctions d'onde sont représentées sur la figure (10-21) pour les états n=1, 2 et 3. Notons enfin que, le mouvement des électrons étant limité à deux dimensions, la densité d'états dans la bande de conduction est donnée par l'expression (10-19). Elle est représentée sur la figure (10-22-a).

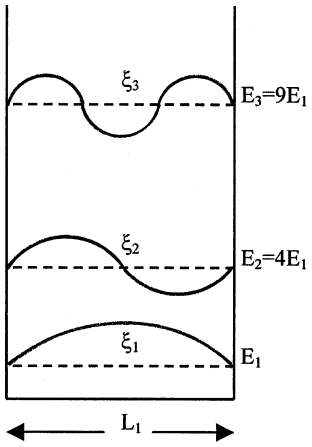


Figure 10-21 : Fonctions d'onde dans un puits de profondeur infinie

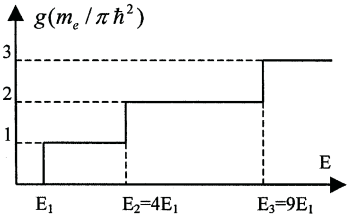


Figure 10-22-a : Densité d'états bidimensionelle

b. Puits de profondeur finie

Lorsque l'énergie de confinement des électrons n'est plus négligeable devant la hauteur de la barrière de potentiel, soit parce que ΔE_c n'est pas très important, soit parce que L_1 est très petit, les résultats précédents sont modifiés par la prise en considération de la profondeur finie du puits de potentiel. En posant $\Delta E_c = V_0$, le potentiel s'écrit (Fig. 10-19)

$$V(z) = V_0 \quad \text{pour} \quad z < 0 \text{ et } z > L_1$$
$$V(z) = 0 \quad \text{pour} \quad 0 < z < L_1$$

Le mouvement des électrons d'énergie $E<V_0$ n'est plus borné en $z=0$ et en $z=L_1$, l'électron a une probabilité non nulle de se trouver à l'extérieur du puits. Le potentiel carré délimite trois régions

- *Région II*$_1$ $z < 0$

Le potentiel est V_0, l'équation de Schrödinger (10-125) s'écrit

$$\frac{d^2\xi(z)}{dz^2} + \frac{2m_2}{\hbar^2}(E - V_0)\xi(z) = 0 \qquad (10\text{-}133)$$

où m_2 représente la masse effective des électrons dans le semiconducteur SC_2. Dans la mesure où on étudie les états liés du puits de potentiel, l'énergie des électrons est inférieure à V_0, de sorte que (V_0-E) est positif et l'équation s'écrit

$$\frac{d^2\xi(z)}{dz^2} - K_2^2\,\xi(z) = 0 \qquad (10\text{-}134)$$

avec $K_2 = \sqrt{2m_2(V_0 - E)}/\hbar$. Compte tenu de la condition $\xi_{II1}(z \to -\infty)=0$, la solution de l'équation (10-134) est de la forme

$$\xi_{II1}(z) = A\,e^{K_2 z} \qquad (10\text{-}135)$$

- *Région I,* $0 < z < L_1$

Le potentiel est nul, l'équation (10-125) s'écrit

$$\frac{d^2\xi(z)}{dz^2} + \frac{2m_1}{\hbar^2}E\xi(z) = 0 \qquad (10\text{-}136)$$

où m_1 représente la masse effective des électrons dans le semiconducteur SC_1. En posant $K_1 = \sqrt{2m_1E}/\hbar$, l'équation (10-136) s'écrit

$$\frac{d^2\xi(z)}{dz^2} + K_1^2\ \xi(z) = 0 \qquad (10\text{-}137)$$

La solution est de la forme

$$\xi_I(z) = B\sin(K_1 z + \varphi) \qquad (10\text{-}138)$$

- *Région II*$_2$, $z > L_1$

Le potentiel est à nouveau V_0, compte tenu de la condition $\xi_{II2}(z\to\infty)=0$, la fonction d'onde s'écrit

$$\xi_{II2}(z) = C\,e^{-K_2 z} \qquad (10\text{-}139)$$

Les constantes d'intégration A, B, C et φ sont déterminées par les conditions aux limites. Ces conditions sont les continuités de la fonction d'onde ξ et du courant de probabilité $(1/m)(d\xi/dz)$ aux interfaces. A partir des expressions (10-135, 138 et 139) on obtient les relations

- En $z = 0$, les conditions $\xi_{II1}(0) = \xi_I(0)$ et $\frac{1}{m_2}\xi'_{II1}(0) = \frac{1}{m_1}\xi'_I(0)$ entraînent

$$A = B\sin\varphi \qquad (10\text{-}140\text{-a})$$

$$A\frac{K_2}{m_2} = B\frac{K_1}{m_1}\cos\varphi \qquad (10\text{-}140\text{-b})$$

- En $z = L_1$, les conditions $\xi_I(L_1) = \xi_{II2}(L_1)$ et $\frac{1}{m_1}\xi'_1(L_1) = \frac{1}{m_2}\xi'_{II2}(L_1)$ entraînent

$$B\sin(K_1 L_1 + \varphi) = Ce^{-K_2 L_1} \qquad (10\text{-}140\text{-c})$$

$$B\frac{K_1}{m_1}\cos(K_1 L_1 + \varphi) = -C\frac{K_2}{m_2}e^{-K_2 L_1} \qquad (10\text{-}140\text{-d})$$

En divisant membre à membre, d'une part les équations (10-140-a et b) et d'autre part les équations (10-140-c et d), on obtient respectivement

$$tg\,\varphi = \frac{K_1 m_2}{K_2 m_1} \qquad (10\text{-}141\text{-a})$$

$$tg(K_1 L_1 + \varphi) = -\frac{K_1 m_2}{K_2 m_1} \qquad (10\text{-}141\text{-b})$$

Compte tenu de la relation $\sin^2\alpha = 1/\left(1+1/tg^2\alpha\right)$ l'expression (10-141-a) donne

$$\sin\varphi = \frac{K_1 / m_1}{\sqrt{K_1^2 / m_1^2 + K_2^2 / m_2^2}} \tag{10-142}$$

D'autre part la relation $tg\,\varphi = -tg(K_1 L_1 + \varphi)$ entraîne $\varphi = -(K_1 L_1 + \varphi) + n\pi$ où n est un nombre entier. Cette relation s'écrit

$$K_1 L_1 = n\pi - 2\varphi \tag{10-143}$$

Soit, en explicitant φ (Eq. 10-142)

$$K_1 L_1 = n\pi - 2\,Arc\sin\frac{K_1 / m_1}{\sqrt{K_1^2 / m_1^2 + K_2^2 / m_2^2}} \tag{10-144}$$

En explicitant K_1 et K_2 à partir de leurs définitions, on obtient la relation

$$\frac{\sqrt{2m_1 E}}{\hbar} L_1 = n\pi - 2\,Arc\sin\sqrt{\frac{E}{V_0\, m_1 / m_2 + E(1 - m_1/m_2)}} \tag{10-145}$$

L'expression (10-145) définit les valeurs possibles de l'énergie E correspondant aux différentes valeurs de n. Si le puits est infini, $V_0 \to \infty$, Arc sin(0)=0, on retrouve la quantification définie par l'expression (10-129).

c. Puits de profondeur finie - Solution analytique

L'énergie de confinement E_1, c'est-à-dire l'énergie du niveau fondamental n=1, peut être obtenue de façon analytique avec une très bonne approximation par un développement limité de l'équation (10-145). Supposons les masses effectives identiques dans les matériaux puits et barrières, $m_1 = m_2 = m_e$, l'équation (10-145) s'écrit

$$\frac{\sqrt{2m_e E_1}}{\hbar} L_1 = \pi - 2Arc\sin\sqrt{\frac{E_1}{V_0}} \qquad \text{ou} \qquad \sqrt{\frac{E_1}{V_0}} = \sin\frac{\pi}{2}\left(1 - \sqrt{\frac{E_1}{E_1^\infty}}\right)$$

où $E_1^\infty = \pi^2\hbar^2 / 2m_e L_1^2$ est l'énergie de confinement dans un puits de même largeur mais avec des barrières de hauteur infinie (Eq. 10-129). Compte tenu de la relation $E_1 < E_1^\infty$, un développement limité de la fonction *sinus* permet d'écrire l'équation sous la forme

$$\sqrt{\frac{E_1}{V_0}} \approx \frac{\pi}{2}\left(1 - \sqrt{\frac{E_1}{E_1^\infty}}\right) - \frac{1}{3!}\left(\frac{\pi}{2}\left(1 - \sqrt{\frac{E_1}{E_1^\infty}}\right)\right)^3 + \frac{1}{5!}\left(\frac{\pi}{2}\left(1 - \sqrt{\frac{E_1}{E_1^\infty}}\right)\right)^5 - \frac{1}{7!}\left(\frac{\pi}{2}\left(1 - \sqrt{\frac{E_1}{E_1^\infty}}\right)\right)^7$$

Un développement au deuxième ordre donne l'équation du second degré en $\sqrt{E_1}$ suivante

$$E_1 + \left(\frac{B}{C}\sqrt{E_1^\infty} - \frac{1}{C}\frac{E_1^\infty}{\sqrt{V_0}} \right)\sqrt{E_1} + \frac{A}{C}E_1^\infty \approx 0$$

avec A=0,9998≈1, B=0,00147≈0, C=-1,2393.

L'énergie de confinement E_1 est donnée par le carré de la solution positive de cette équation, soit compte tenu des valeurs des coefficients

$$E_1 \approx E_1^\infty \left(\frac{\sqrt{E_1^\infty / V_0}}{2C} + \sqrt{\frac{E_1^\infty / V_0}{4C^2} - \frac{1}{C}} \right)^2 \qquad \text{(10-145-a)}$$

Pour la structure $Al_xGa_{1-x}As$-GaAs-$Al_xGa_{1-x}As$ avec x=0,3 et L_1=110 Å, dans laquelle V_o=ΔE_c=383 meV et m_e=0,067 m_o, $E_1^\infty = 48$ meV, on obtient E_1≈28 meV.

La figure (10-22-b) permet de comparer pour différentes largeurs et profondeurs de puits, les résultats de la solution approchée (Eq. 10-145-a) et de la résolution numérique de l'équation exacte (10-145).

Notons que la structure de sous-bandes d'énergie, résultant de la différence ΔE_c des bandes de conduction, existe de la même manière dans le puits de potentiel résultant de la différence ΔE_v des bandes de valence. Les niveaux d'énergie et les fonctions d'onde sont représentés sur la figure (10-23). Les trous sont confinés dans les puits de potentiel associés aux bandes de valence de plus haute énergie.

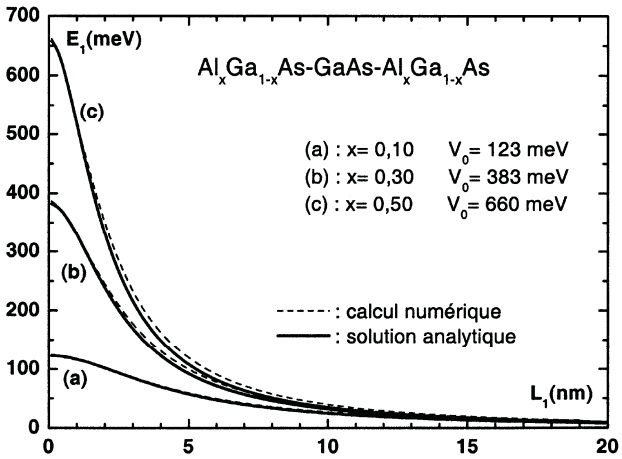


Figure 10-22-b: Energie de confinement des électrons dans un puits de type $Al_xGa_{1-x}As$-GaAs-$Al_xGa_{1-x}As$. Variations avec la largeur pour différentes profondeurs (Mathieu H. - 1992)

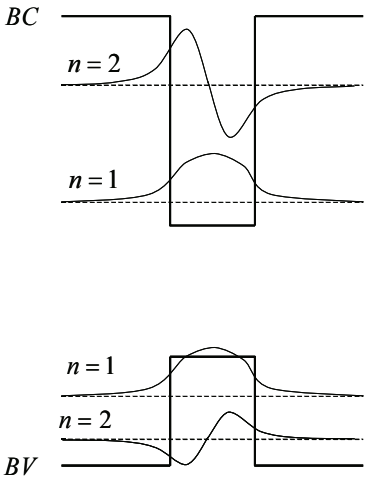


Figure 10-23 : Puits de profondeur finie. Niveaux d'énergie et fonctions d'onde

10.2.2. *Multipuits quantiques*

Considérons une structure résultant de la juxtaposition d'une série de couches alternées d'un semiconducteur SC_1 et d'un semiconducteur SC_2. Il existe une succession de puits dont le diagramme énergétique est représenté sur la figure (10-24). Si les couches de semiconducteur SC_1 sont de faible épaisseur (L_1~100 Å) les états électroniques dans chacun des puits sont quantifiés et présentent une structure de sous-bandes d'énergie. Si en outre les couches de semiconducteur SC_2 sont relativement épaisses (L_2>200 Å) la probabilité pour qu'un électron passe d'un puits dans l'autre par effet tunnel à travers la barrière est faible, les puits sont indépendants les uns des autres. Dans chacun des puits la structure de sous-bandes d'énergie est analogue à celle du puits unique, le mouvement des électrons dans la structure est bidimensionnel. Cette série de puits quantiques indépendants porte le nom de *structure à multipuits quantiques*.

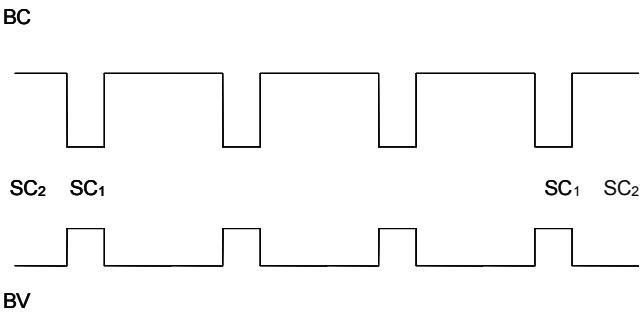


Figure 10-24 : Structure à multipuits quantiques de type I

10.2.3. *Puits quantiques couplés*

Considérons la structure du type SC_2-SC_1-SC_2-SC_1-SC_2 , représentée sur la figure (10-25). Les bandes de conduction des couches SC_1 constituent des puits de potentiel de profondeur ΔE_c et de largeur L_1. Considérons, pour simplifier l'étude, la première sous-bande de conduction. La figure (10-25-b) représente, pour chacun des puits, l'énergie du bas de la première sous-bande et la fonction d'onde des électrons dans leur mouvement suivant z. En l'absence de couplage ces fonctions d'onde sont données par les expressions (10-135, 138 et 139) que l'on écrira, suivant les régions de l'espace, sous la forme suivante

$$z<z_1 \qquad \xi_1(z)=Ae^{K_2(z-z_1)} \qquad \text{(10-146-a)}$$

$$z_1<z<z_1+L_1 \qquad \xi_1(z)=B\sin\left(K_1(z-z_1)+\varphi\right) \qquad \text{(10-146-b)}$$

$$z_1+L_1<z \qquad \xi_1(z)=Ce^{-K_2(z-z_1)} \qquad \text{(10-146-c)}$$

$$z<z_2 \qquad \xi_2(z)=Ae^{K_2(z-z_2)} \qquad \text{(10-147-a)}$$

$$z_2<z<z_2+L_1 \qquad \xi_2(z)=B\sin\left(K_1(z-z_2)+\varphi\right) \qquad \text{(10-147-b)}$$

$$z_2+L_1<z \qquad \xi_2(z)=Ce^{-K_2(z-z_2)} \qquad \text{(10-147-c)}$$

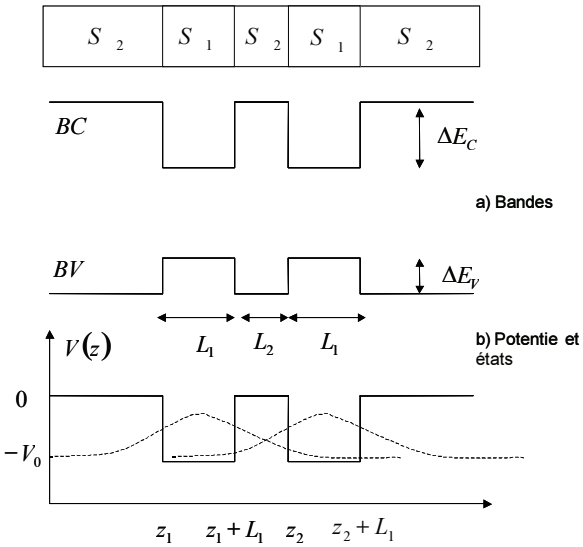


Figure 10-25 : Structure à multipuits quantiques couplés

Si l'épaisseur L_2 du semiconducteur SC_2 est grande, les fonctions d'onde associées à chacun des puits, qui décroissent exponentiellement à l'extérieur du puits, sont nulles dans la région de l'espace occupée par l'autre puits. Les deux puits sont alors indépendants, leurs états propres sont donnés par l'expression (10-145), ces états sont doublement dégénérés.

Si l'épaisseur L_2 diminue, $\xi_1(z)$ est alors différent de zéro en $z>z_2$, la probabilité de présence de l'électron associé au puits 1, en $z>z_2$ est non nulle ; cet électron *voit* le puits 2. La réciproque est vraie pour l'électron associé au puits 2. Les états électroniques du puits 1 et du puits 2 sont couplés par effet tunnel. Ces états ne sont plus états propres du système constitué par les deux puits.

L'Hamiltonien des électrons s'écrit

$$H=-\frac{\hbar^2}{2m_e}\frac{d^2}{dz^2}+V^{(1)}(z)+V^{(2)}(z) \qquad (10\text{-}148)$$

Avec

$$V^j(z)=0 \quad pour\ z<z_j\ et\ z>z_j+L_1$$
$$=-V_o \quad pour\ z_j<z<z_j+L_1 \quad (j=1,2)$$

Dans la mesure où l'on connaît les états propres et les fonctions propres dans chacun des puits supposés indépendants, on peut développer les fonctions d'onde du système couplé sur la base des fonctions d'onde des puits découplés. Pour chacun des électrons, et dans la mesure où l'on néglige les couplages avec les états de plus grande énergie, la fonction d'onde s'écrit donc sous la forme d'une combinaison linéaire des fonctions d'onde $\xi_1(z)$ et $\xi_2(z)$, soit, avec i =1,2,

$$\xi^{(i)}(z)=a_i\,\xi_1(z)+b_i\,\xi_2(z) \qquad (10\text{-}149)$$

Les états propres sont donnés par la résolution du système d'équations de Schrödinger relatives à chacun des électrons

$$H\xi^{(i)}(z)=E\xi^{(i)}(z) \qquad (10\text{-}150)$$

En portant les expressions (10-149) dans les expressions (10-150), en multipliant à gauche par $\xi^{(i)*}(z)$ et en intégrant dans tout l'espace, on obtient le déterminant

$$\begin{vmatrix} E_1+s-E & (E_1-E)r+t \\ (E_1-E)r+t & E_1+s-E \end{vmatrix} = 0 \qquad (10\text{-}151)$$

$E_1=\left\langle \xi^{(i)}(z)\right| -\frac{\hbar^2}{2m_e}\frac{d^2}{dz^2}+V^{(j)}(z)\left|\xi^{(i)}(z)\right\rangle$, représente l'énergie des électrons dans chacun des puits en l'absence de couplage.

$r=\left\langle \xi_1(z)\right|\xi_2(z)\rangle=\left\langle \xi_2(z)\right|\xi_1(z)\rangle$, représente les intégrales de recouvrement des fonctions d'onde.

$s = \langle \xi_1(z) | V^{(2)}(z) | \xi_1(z) \rangle = \langle \xi_2(z) | V^{(1)}(z) | \xi_2(z) \rangle$, représente les intégrales de dérive, qui traduisent la dérive des niveaux d'énergie de chacun des puits résultant de la présence de l'autre puits.

$t = \langle \xi_1(z) | V^{(1)}(z) | \xi_2(z) \rangle = \langle \xi_2(z) | V^{(2)}(z) | \xi_1(z) \rangle$, représente les intégrales de couplage qui traduisent l'interaction entre les niveaux associés à chacun des puits. Ce couplage lève la dégénérescence des niveaux.

Si on néglige les intégrales de recouvrement, le déterminant (10-151) se simplifie et s'écrit

$$\begin{vmatrix} E_1 + s - E & t \\ t & E_1 + s - E \end{vmatrix} = 0 \qquad (10\text{-}152)$$

Les solutions sont de la forme

$$E^{(1)} = E_1 + s + t$$
$$E^{(2)} = E_1 + s - t$$

Le diagramme énergétique est représenté sur la figure (10-26). Les électrons sont spatialement répartis dans les deux puits et à l'extérieur de ces deux puits, c'est-à-dire dans toute la structure SC_2-SC_1-SC_2-SC_1-SC_2. La probabilité de présence des électrons dans la première couche SC_1 et à son voisinage immédiat est donnée par a_i^2 , leur probabilité de présence dans la deuxième couche SC_1 est donnée par b_i^2. Le niveau de plus basse énergie correspond à la combinaison symétrique (a_i et b_i de même signe) des fonctions d'onde $\xi_1(z)$ et $\xi_2(z)$. Le niveau de plus haute énergie correspond à la combinaison antisymétrique.

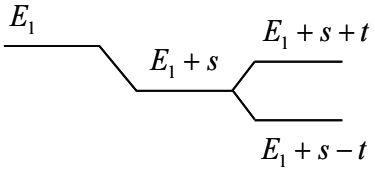


Figure 10-26 : Diagramme énergétique

10.3. Superréseaux

Les états résultant du couplage de deux puits quantiques sont des combinaisons symétriques et antisymétriques des états associés à chacun des puits, isolé. Ce système de deux puits couplés est à rapprocher de la molécule diatomique, dans laquelle les orbitales moléculaires associées aux combinaisons liantes et antiliantes

des orbitales atomiques, résultent du couplage entre les deux atomes (Par. 1.2.1.b). La séquence orbitales atomiques → orbitales moléculaire → états cristallins, résultant des couplages d'un grand nombre d'atomes (Par. 1.2.1.c), peut donc être reproduite ici sous la forme analogue : puits quantique isolé → double-puits quantiques couplés → multipuits quantiques couplés. On obtient ainsi un réseau à une dimension de puits quantiques couplés, analogue au réseau d'atomes dans le cristal. Ce réseau de puits quantiques a évidemment une maille plus importante que celle du réseau atomique et se superpose à ce dernier dans une direction déterminée. Ce système de multipuits quantiques couplés porte le nom de *superréseau*.

La figure (10-27) représente une succession de puits quantiques couplés, de largeur L_1, distants de L_2, résultant de la juxtaposition de multicouches alternées de semiconducteurs SC_1 et SC_2, d'épaisseurs respectives L_1 et L_2. La maille du superréseau est $L=L_1+L_2$.

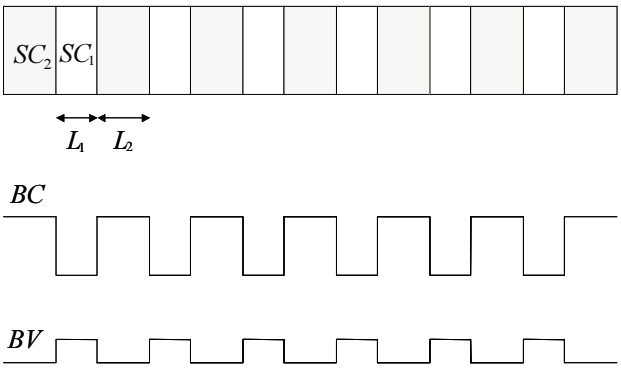


Figure 10-27 : Superréseau de type I

10.3.1. *Structure de sous-bandes d'énergie*

Alors que dans la structure à multipuits quantiques isolés, comme dans le puits unique, le mouvement des électrons est bidimensionnel, dans le superréseau le couplage entre les puits successifs ramène le mouvement à trois dimensions. La structure périodique du superréseau, superposée à celle du cristal, entraîne l'existence d'une *superstructure de bandes* caractérisée par une zone de Brillouin de longueur, $2\pi/a$ dans les directions x et y (plan de la structure) et $2\pi/L$ dans la direction z (axe de la structure), où a et L sont respectivement les mailles du réseau cristallin et du superréseau. Le paramètre L étant beaucoup plus grand que le paramètre a, la *super-zone de Brillouin* est considérablement réduite dans la direction z.

Considérons le semiconducteur SC_1 par exemple (Fig. 10-27). S'il était seul, la courbe de dispersion de l'énergie des électrons de conduction dans la direction k_z serait donnée par la courbe (a) dans la figure (10-28). Dans le superréseau, la

première zone de Brillouin dans la direction k_z est réduite de la longueur $2\pi/a$, à la longueur $2\pi/L$, les bandes d'énergie sont alors repliées dans l'intervalle $(-\pi/L,+\pi/L)$ (Fig.10-28 courbes en pointillés). En outre, la discontinuité du potentiel aux interfaces SC_1-SC_2 entraîne l'ouverture de bandes interdites aux points $k_z = 0$ et $k_z = \pm\pi/L$ de la zone de Brillouin. La superstructure de bandes est représentée en traits pleins sur la figure (10-28).

Au voisinage du bas de la bande de conduction, l'approximation parabolique permet d'écrire la variation d'énergie de la bande de conduction sous la forme

$$E(k)=E_c+\frac{\hbar^2 k_z^2}{2m_e}$$

Dans la mesure où, au bord de la super-zone de Brillouin, $k_z=\pi/L$, la différence d'énergie entre le bas et le sommet de la première sous bande de conduction, c'est-à-dire la largeur ΔE de cette sous-bande, ne peut excéder la valeur ΔE_{max} donnée par

$$\Delta E_{\max}=\frac{\hbar^2}{2m_e}\left(\frac{\pi}{L}\right)^2$$

Considérons à titre d'exemple un superréseau avec un pas L=100 Å, on obtient suivant le type de semiconducteur constituant les couches SC_1

SC_1	m_e/m_0	ΔE_{max} (meV)
GaAs	0,067	57
GaSb	0,043	87
InP	0,073	52
InAs	0,023	164
InSb	0,012	313

On obtient donc une largeur de sous-bandes de conduction de quelques dizaines de meV. Rappelons pour mémoire que dans un semiconducteur homogène les largeurs de bandes sont de l'ordre de 10 à 15 eV. La figure (10-20) représente les courbes de dispersion correspondant au puits quantique unique. Ces courbes varient comme $\hbar^2 k_{//}^2/2m_e$ dans le plan de la structure et sont non dispersives dans la direction perpendiculaire. Dans le superréseau la périodicité dans l'axe de la structure se traduit par les courbes de dispersion de la figure (10-28). Les résultats sont résumés sur la figure (10-29).

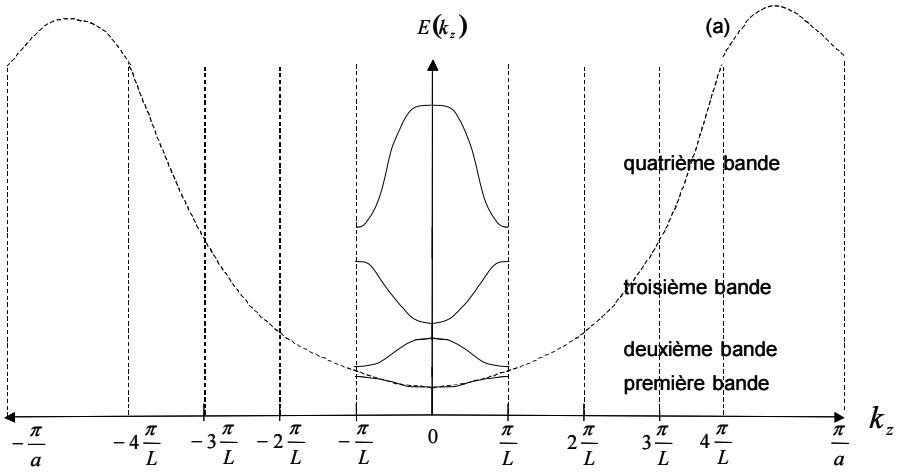


Figure 10-28 : Evolution de la courbe de dispersion résultant de la réduction de la zone de Brillouin dans un superréseau de période L. La courbe (a) représente la courbe de dispersion dans le cristal homogène

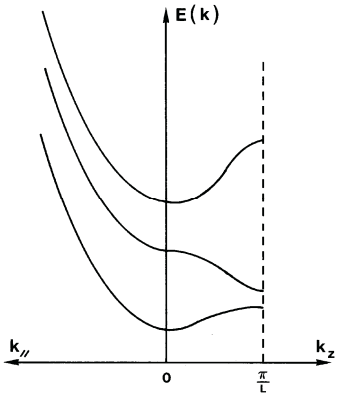


Figure 10-29 : Courbes de dispersion

10.3.2. *Modèle de Kronig-Penney*

La bande de conduction de la figure (10-27) se présente, dans la direction du superréseau, comme une succession de puits quantiques de profondeur $V_0=\Delta E_c$, de largeur L_1 et distants de L_2. Ce potentiel périodique, de période $L=L_1+L_2$, correspond tout à fait au modèle de Kronig-Penney.

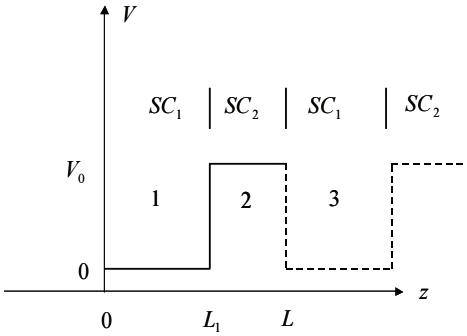


Figure 10-30 : Maille élémentaire du superréseau. Les épaisseurs de couche des semiconducteurs SC_1 et SC_2 sont L_1 et L_2

Considérons la maille élémentaire du superréseau représentée sur la figure (10-30). L'équation de Schrödinger des électrons est, dans l'approximation de la masse effective, de la forme

$$\left(-\frac{\hbar^2}{2m_e}\frac{d^2}{dz^2}+V(z)\right)\psi(z)=E\psi(z) \qquad (10\text{-}153)$$

avec $V(z)=0$ et $m_e=m_1$ dans la région I correspondant au semiconducteur SC_1 et $V(z)=V_0$ et $m_e=m_2$ dans la région II correspondant au semiconducteur SC_2. Dans chacune des régions, l'équation (10-153) s'écrit respectivement

Région I $$\frac{d^2\psi(z)}{dz^2}+K_1^2\,\psi(z)=0 \qquad (10\text{-}154\text{-a})$$

Région II $$\frac{d^2\psi(z)}{dz^2}-K_2^2\,\psi(z)=0 \qquad (10\text{-}154\text{-b})$$

avec dans la mesure où on ne s'intéresse qu'aux états d'énergie $E<V_0$

$$K_1=\sqrt{2m_1E}\,/\hbar \qquad K_2=\sqrt{2m_2(V_o-E)}\,/\hbar$$

Les solutions des équations (10-154-a et b) sont respectivement de la forme

$$\psi_1(z)=A_1\cos K_1z+B_1\sin K_1z \qquad (10\text{-}155)$$

$$\psi_2(z)=A_2\,ch\,K_2(z-L_1)+B_2\,sh\,K_2(z-L_1) \qquad (10\text{-}156)$$

En outre en raison de la périodicité du superréseau, les fonctions d'onde électroniques et leurs dérivées doivent satisfaire au théorème de Bloch (1-17). En d'autres termes, les fonctions d'onde des électrons dans la région 3 ne différent des fonctions d'onde dans la région 1 que par le terme de phase e^{ikL}

$$\psi_3(z)=e^{ikL}\psi_1(z-L)$$

Soit

$$\psi_3(z)=e^{ikL}\left(A_1\cos K_1(z-L)+B_1\sin K_1(z-L)\right) \qquad (10\text{-}157)$$

Il en est de même pour les dérivées qui s'écrivent dans chacune des régions

$$\psi_1'(z)=-A_1K_1\sin K_1z+B_1K_1\cos K_1z \qquad (10\text{-}158)$$

$$\psi_2'(z)=A_2K_2\,sh\,K_2(z-L_1)+B_2K_2\,ch\,K_2(z-L_1) \qquad (10\text{-}159)$$

$$\psi_3'(z)=e^{ikL}\left(-A_1K_1\sin K_1(z-L)+B_1K_1\cos K_1(z-L)\right) \qquad (10\text{-}160)$$

Les conditions de continuité de la fonction d'onde ψ et du courant de probabilité $1/m.\,d\psi/dz$ en $z=L_1$ et en $z=L$ s'écrivent

$$\psi_1(z=L_1)=\psi_2(z=L_1) \qquad (10\text{-}161)$$

$$\frac{1}{m_1}\psi_1'(z=L_1)=\frac{1}{m_2}\psi_2'(z=L_1) \qquad (10\text{-}162)$$

$$\psi_2(z=L)=\psi_3(z=L) \qquad (10\text{-}163)$$

$$\frac{1}{m_2}\psi_2'(z=L)=\frac{1}{m_1}\psi_3'(z=L) \qquad (10\text{-}164)$$

En explicitant ψ et ψ' et compte tenu de la relation $L=L_1+L_2$, ces conditions s'écrivent

$$A_1\cos K_1L_1+B_1\sin K_1L_1-A_2=0 \qquad (10\text{-}165)$$

$$-A_1\frac{K_1}{m_1}\sin K_1L_1+B_1\frac{K_1}{m_1}\cos K_1L_1-B_2\frac{K_2}{m_2}=0 \qquad (10\text{-}166)$$

$$e^{ikL}A_1-A_2\,ch\,K_2L_2-B_2\,sh\,K_2L_2=0 \qquad (10\text{-}167)$$

$$e^{ikL}B_1\frac{K_1}{m_1}-A_2\frac{K_2}{m_2}sh\,K_2L_2-B_2\frac{K_2}{m_2}ch\,K_2L_2=0 \qquad (10\text{-}168)$$

On obtient quatre équations homogènes relatives aux quatre coefficients A_1, B_1, A_2, B_2, qui s'écrivent sous forme matricielle,

$$\begin{bmatrix} \cos K_1 L_1 & \sin K_1 L_1 & -1 & 0 \\ -\frac{K_1}{m_1}\sin K_1 L_1 & \frac{K_1}{m_1}\cos K_1 L_1 & 0 & -\frac{K_2}{m_2} \\ e^{ikL} & 0 & -chK_2L_2 & -shK_2L_2 \\ 0 & e^{ikL}\frac{K_1}{m_1} & -\frac{K_2}{m_2}shK_2L_2 & -\frac{K_2}{m_2}chK_2L_2 \end{bmatrix} \begin{bmatrix} A_1 \\ B_1 \\ A_2 \\ B_2 \end{bmatrix} = 0$$

Les coefficients ne sont différents de zéro que si le déterminant de la matrice est nul, ce qui donne la relation

$$\cos kL = chK_2L_2 \cdot \cos K_1L_1 + \frac{1}{2}\left(\frac{K_2m_1}{K_1m_2} - \frac{K_1m_2}{K_2m_1}\right) shK_2L_2 \cdot \sin K_1L_1 \qquad (10\text{-}170)$$

Compte tenu des expressions de K_1 et K_2, l'expression (10-170) est de la forme

$$f(V_0, L_1, L_2, m_1, m_2, E) = \cos kL \qquad (10\text{-}171)$$

Dans la mesure où le cosinus est toujours compris entre -1 et +1, l'expression (10-171) détermine les valeurs permises de l'énergie, par la simple condition $|f(V_0, L_1, L_2, m_1, m_2, E)| \leq 1$. Cette condition délimite les bandes permises dans lesquelles k est réel . Au contraire, les bandes d'énergie correspondant à $|f(V_0, L_1, L_2, m_1, m_2, E)| > 1$ sont des bandes interdites dans lesquelles k est imaginaire.

Nous allons étudier comment varient les bandes d'énergie quand la structure évolue d'une configuration à couches épaisses à une configuration à couches minces, c'est-à-dire quand le système évolue d'une structure de multipuits quantiques découplés (L_2 grand) à une structure de superréseau (puits couplés, L_2 petit).

a. Puits découplés, $L_2 >> L_1$

Supposons V_0 et L_1 finis, si L_2 devient très grand le cosinus et le sinus hyperboliques tendent vers la même valeur

$$ch\,K_2L_2 \approx sh\,K_2L_2 \approx \frac{1}{2}e^{K_2L_2}$$

L'expression (10-170) s'écrit alors

$$2e^{-K_2L_2}\cos kL = \cos K_1L_1\left(1 + \frac{1}{2}\left(\frac{K_2m_1}{K_1m_2} - \frac{K_1m_2}{K_2m_1}\right) tg\,K_1L_1\right) \qquad (10\text{-}172)$$

Lorsque L_2 devient très grand le premier membre de l'expression précédente tend vers zéro, de sorte que la relation est satisfaite pour

$$tg\,K_1L_1=-2\Big/\left(\frac{K_2m_1}{K_1m_2}-\frac{K_1m_2}{K_2m_1}\right) \tag{10-173}$$

Compte tenu de la relation trigonométrique $tg(x)=2tg(x/2)\big/\big(1-tg^2(x/2)\big)$, l'expression (10-173) s'écrit

$$\frac{2tg(K_1L_1/2)}{1-tg^2(K_1L_1/2)}=-2\Big/\left(\frac{K_2m_1}{K_1m_2}-\frac{K_1m_2}{K_2m_1}\right)$$

ou
$$tg^2(K_1L_1/2)-\left(\frac{K_2m_1}{K_1m_2}-\frac{K_1m_2}{K_2m_1}\right)tg(K_1L_1/2)-1=0$$

Les solutions de cette équation du second degré en tg(K_1L_1/2) sont de la forme

$$tg(K_1L_1/2)=\frac{1}{2}\left(\frac{K_2m_1}{K_1m_2}-\frac{K_1m_2}{K_2m_1}\right)\pm\frac{1}{2}\left(\frac{K_2m_1}{K_1m_2}-\frac{K_1m_2}{K_2m_1}\right)$$

c'est-à-dire
$$tg\frac{K_1L_1}{2}=-\frac{K_1m_2}{K_2m_1} \tag{10-174-a}$$

$$cotg\frac{K_1L_1}{2}=\frac{K_1m_2}{K_2m_1} \tag{10-174-b}$$

On retrouve les niveaux d'énergie du puits isolé. Ceci résulte du fait que lorsque L_2 est important, les puits de potentiel sont découplés les uns des autres. En particulier dans la limite du puits infini, $k_2\to\infty$ et les expressions (10-174-a et b) se réduisent à

$$tg\frac{K_1L_1}{2}=0 \qquad\qquad cotg\frac{K_1L_1}{2}=0$$

soit
$$\frac{K_1L_1}{2}=n\frac{\pi}{2}$$

Ce qui donne, compte tenu de l'expression de K_1

$$E_n=n^2\frac{\hbar^2\pi^2}{2m_1L_1^2}$$

on retrouve l'expression (10-129).

b. Puits couplés, $L_2 \approx L_1$

En explicitant K_2 en fonction de K_1, l'expression (10-172) correspondant à L_2 grand, peut s'écrire sous la forme

$$2\,e^{-K_2 L_2} \cos kL = f(K_1, L_1) \tag{10-175}$$

Dans la limite des puits très séparés ($L_2 \to \infty$), nous venons de voir que l'expression précédente permettait de déterminer les niveaux d'énergie par la relation $f(k_1, L_1)=0$. Lorsque L_2 diminue, les puits se rapprochent, le premier membre de l'expression (10-175) devient non nul et on peut développer le second en se limitant au terme du premier ordre tant que L_2 n'est pas trop petit

$$f(K_1, L_1)=f(K_{1n},L_1)+(K_1-K_{1n})\left(\frac{df(K_1, L_1)}{dK_1}\right)_{K_{1n}} \tag{10-176}$$

où $f(K_{1n}, L_1)=0$, et correspond au cas du puits isolé. L'équation (10-175) s'écrit alors

$$2e^{-K_2 L_2} \cos kL = f'(K_{1n}, L_1).(K_1 - K_{1n}) \tag{10-177}$$

Soit

$$K_1 = K_{1n} + 2\,e^{-K_2 L_2}\ \frac{\cos kL}{f'(K_{1n}, L_1)} \tag{10-178}$$

En explicitant K, l'expression s'écrit

$$\frac{\sqrt{2m_1 E}}{\hbar} = \frac{\sqrt{2m_1 E_n}}{\hbar} + 2e^{-K_2 L_2}\ \frac{\cos kL}{f'(K_{1n}, L_1)} \tag{10-179}$$

En élevant au carré et en ne conservant que le terme de premier ordre, l'expression de l'énergie s'écrit

$$E = E_n\left(1 + W_n(k)\right) \tag{10-180}$$

avec

$$W_n(k) = 4\,e^{-K_2 L_2}\ \frac{\cos kL}{K_{1n}\ \ f'(K_{1n}, L_1)}$$

Lorsque L_2 diminue, les niveaux d'énergie discrets E_n, correspondant aux puits isolés ($L_2 >> L_1$), s'élargissent en bandes d'énergie. Dans chacune des bandes, cos kL varie de -1 à +1, en d'autres termes k varie de $-\pi/L$ à $+\pi/L$. Les largeurs des bandes d'énergie sont obtenues en faisant cos $kL = \pm 1$, soit

$$\Delta E_n = 8\,E_n \frac{e^{-K_2 L_2}}{K_{1n}\left|f'(K_{1n}, L_1)\right|} \tag{10-181}$$

Le bas de la bande est situé en k=0 si $f'(K_{1n}, L_1) < 0$ et en $k = \pm\pi/L$ si $f'(K_{1n}, L_1) > 0$

c. Puits très couplés, $L_2 << L_1$

Supposons maintenant L_2 faible devant L_1, mais conservons constant le produit $V_0\,L_2$ pour ne pas réduire à zéro la barrière de potentiel qui sépare chacun des puits. Les fonctions hyperboliques deviennent respectivement

$$ch\,K_2\,L_2 \approx 1 \qquad\qquad sh\,K_2 L_2 \approx K_2 L_2$$

L'équation (10-170) s'écrit alors

$$\cos kL = \cos K_1 L_1 + \frac{1}{2}\left(\frac{K_2 m_1}{K_1 m_2} - \frac{K_1 m_2}{K_2 m_1}\right) K_2 L_2 \sin K_1 L_1 \tag{10-182}$$

Compte tenu des définitions de K_1 et K_2, le coefficient de sin $K_1 L_1$ s'écrit

$$\frac{1}{2}\left(\frac{K_2 m_1}{K_1 m_2} - \frac{K_1 m_2}{K_2 m_1}\right) K_2 L_2 = \frac{1}{2\hbar}\sqrt{\frac{2m_1}{E}}\,V_o L_2 - \frac{1}{2\hbar}\sqrt{2m_1 E}\left(1 + \frac{m_2}{m_1}\right) L_2 \tag{10-183}$$

Dans la mesure où le produit $V_0\,L_2$ reste constant quand L_2 tend vers zéro, ce coefficient se réduit à son premier terme et l'expression (10-182) s'écrit

$$\cos kL = \cos K_1 L_1 + \frac{1}{2\hbar}\sqrt{\frac{2m_1}{E}}\,V_o L_2 \sin K_1 L_1 \tag{10-184}$$

En posant $A = m_1 V_0 L_1 L_2 / \hbar^2$, et compte tenu du fait que $L = L_1 + L_2 \approx L_1$, l'expression précédente s'écrit

$$\cos kL = \cos K_1 L + A\,\frac{\sin K_1 L}{K_1 L} = f(K_1 L) \tag{10-185}$$

Les bandes permises sont alors délimitées par la condition

$$-1 < f(K_1 L) < 1$$

La fonction $f(K_1 L) = \cos K_1 L + A \sin K_1 L / K_1 L$ est représentée sur la figure (10-31) en fonction de $K_1 L$. Les intersections avec les horizontales d'ordonnées ± 1

définissent les extrémités des bandes permises. Les régions correspondantes de l'axe des abscisses définissent les valeurs permises de K_1L, c'est-à-dire de l'énergie.

d. Courbes de dispersion E(k)

La figure (10-31) montre qu'à une valeur de $f(K_1L)$, c'est-à-dire de cos *kL* (horizontale sur la figure) correspondent plusieurs valeurs de K_1, c'est-à-dire de l'énergie.

Ces différentes valeurs de l'énergie appartiennent aux différentes bandes permises. Mais d'autre part à une valeur de cos *kL* correspond une infinité de valeurs de *k* qui sont

$$k = \pm\frac{Arc\cos kL}{L} + 2n\frac{\pi}{L}$$

Ainsi à une valeur de *k* correspond une valeur de *E* dans chaque bande permise.

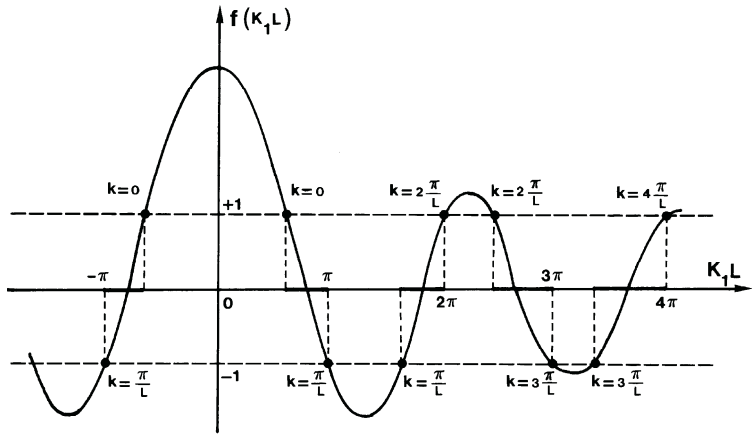


Figure 10-31 : Résolution graphique de l'équation (10-185). Les intersections de la courbe avec les horizontales +1 et -1 définissent les frontières entre les bandes permises et les bandes interdites. Les valeurs permises de l'énergie sont données par les valeurs permises de K_1L. Ces valeurs sont représentées par des traits renforcés sur l'axe des abscisses

La relation de dispersion *E*(*k*) se déduit de l'expression (10-185). Les courbes représentatives sont des portions de parabole à l'intérieur de chaque bande permise avec des déformations aux voisinages des extrema correspondant à $k=\pm n\pi/L$. On obtient l'évolution des courbes *E*(*k*) aux extrema des bandes en différentiant l'expression (10-185) au voisinage de $k=n\pi/L$

$$\frac{d}{dk}(\cos kL) = \frac{d}{dk} f(K_1L) \qquad (10\text{-}186)$$

soit

$$-L\sin kL=\frac{d}{dK_1}f(K_1L)\ \frac{dK_1}{dk}$$

d'où l'on tire

$$\frac{dK_1}{dk}=-\frac{L\sin kL}{\frac{d}{dK_1}f(K_1L)} \qquad (10\text{-}187)$$

Compte tenu de la définition de K_1, l'énergie est donnée par $E=\frac{\hbar^2K_1^2}{2m_1}$ de sorte que

$$\frac{dE}{dk}=\frac{\hbar^2}{m_1}K_1\frac{dK_1}{dk}$$

Soit, compte tenu de la relation (10-187)

$$\frac{dE}{dk}=-\frac{\hbar^2K_1L}{m_1}\frac{\sin kL}{\frac{d}{dK_1}f(K_1L)} \qquad (10\text{-}188)$$

En explicitant $f(K_1L)$ à partir de (10-185) on obtient

$$\frac{d}{dK_1}f(K_1L)=-\left(L+\frac{A}{K_1^2L}\right)\sin K_1L+\frac{A}{K_1}\cos K_1L \qquad (10\text{-}189)$$

Compte tenu de (10-188 et 189), la dérivée de la courbe $E(k)$ s'écrit donc

$$\frac{dE}{dk}=\frac{\hbar^2K_1L}{m_1}\frac{\sin kL}{\left(L+\frac{A}{K_1^2L}\right)\sin K_1L-\frac{A}{K_1}\cos K_1L} \qquad (10\text{-}190)$$

Aux extrema des bandes permises $k=n\pi/L$ de sorte que $kL=n\pi$ et par suite $\sin kL=0$. Il en résulte que $dE/dk=0$. Les courbes $E(k)$ sont représentées sur la figure (10-32). A l'intérieur d'une bande permise l'énergie est une fonction périodique de k avec la périodicité $2\pi/L$ du réseau réciproque. Cette périodicité permet de limiter la représentation à la première zone de Brillouin, c'est-à-dire dans l'intervalle $(-\pi/L,+\pi/L)$.

Notons enfin pour terminer, l'existence de superréseaux à modulation de dopage du type SC(n)-SC(i)-SC(p)-SC(i)-SC(n)... ou plus simplement du type SC(n)-SC(p)-SC(n)...Dans ces structures qui portent le nom de *superréseaux nipi*, les électrons libérés par les donneurs des couches dopées n, compensent les accepteurs situés dans les couches dopées p. Il apparaît ainsi un superpotentiel dû aux charges N_a^- et N_d^+, qui se superpose au potentiel cristallin, avec une périodicité plus faible et qui module les énergies électroniques. Suivant l'importance des

dopages de ces superréseaux, la bande de conduction de la région dopée n peut se situer au-dessus ou au-dessous de la bande de valence de la région dopée p. Dans le premier cas, le superréseau est de type semiconducteur, dans le second il est de type semimétallique

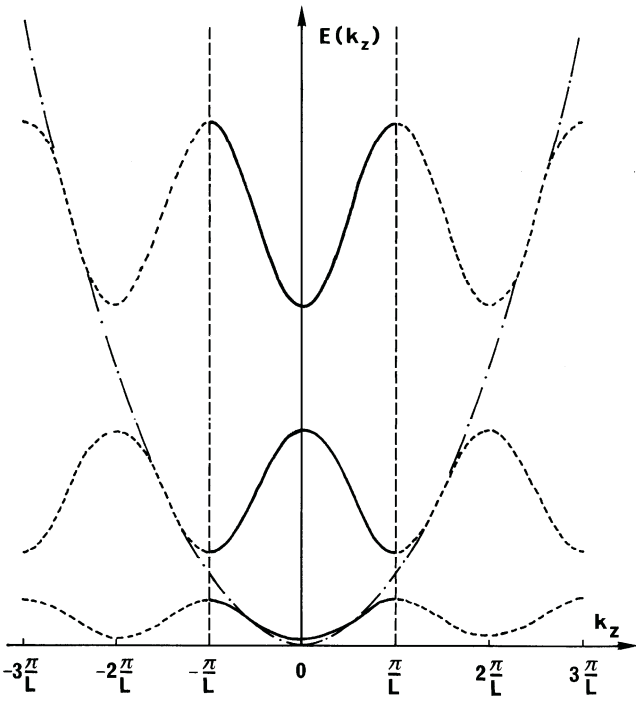


Figure 10-32 : Courbes de dispersion des électrons dans la direction parallèle à l'axe du superréseau

10.4. Transistor à effet de champ à gaz d'électrons bidimensionnel-TEGFET

10.4.1. *Structure*

Le principe de base du TEGFET (Two-dimensional Electron Gas Field Effect Transistor), également connu sous le nom de HEMT (High Electron Mobility Transistor), ou MODFET (MOdulation Doped Field Effect Transistor), consiste à

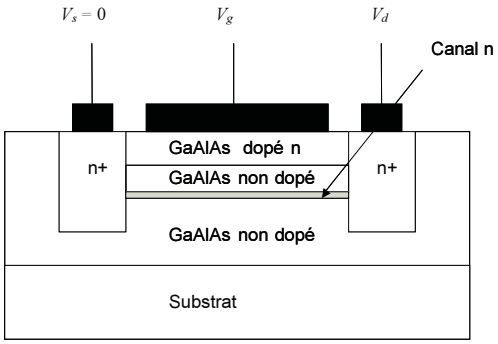


Figure 10-33 : TEGFET

mettre à profit les propriétés de haute mobilité d'un gaz bidimensionnel d'électrons formé à l'interface d'une hétérojonction. L'objectif est de séparer spatialement les électrons libres, des donneurs ionisés dont ils proviennent. La structure du TEGFET GaAs-GaAlAs peut être comparée à celle du MESFET-GaAs dans lequel une couche mince de GaAlAs est intercalée entre l'électrode métallique et le GaAs ou à celle du MOSFET-GaAs dans lequel la couche d'oxyde est remplacée par une couche de GaAlAs. La structure du transistor est représentée sur la figure (10-33).

La couche de GaAlAs, dont le gap est supérieur à celui de GaAs, est dopée de type n. L'électrode de contrôle est de type Schottky, sa tension de diffusion et sa tension de polarisation mettent la couche de GaAlAs en déplétion totale, les électrons libérés par les donneurs de cette couche sont transférés dans le GaAs où ils constituent un gaz d'électrons bidimensionnel à l'interface de l'hétérojonction. Dans le GaAs non dopé le gaz d'électrons constitue le canal du transistor. Ces électrons ont une grande mobilité en raison de l'absence d'ions donneurs dont on connaît l'importance dans les phénomènes de diffusion. Le gaz bidimensionnel d'électrons étant localisé à l'interface de l'hétérojonction, ces derniers peuvent subir l'interaction coulombienne des donneurs ionisés présents dans le GaAlAs. Pour minimiser cet effet on intercale entre les donneurs et le canal une couche intrinsèque de GaAlAs appelée *spacer*. La couche de GaAlAs n'est donc pas dopée de manière uniforme, d'où le nom de MODFET donné parfois au transistor. La conductance du canal ainsi constitué est contrôlée par la grille métallique à travers la couche isolante de GaAlAs, l'analogie est totale avec le MOSFET.

10.4.2. *Commande de grille*

Sous la triple action des différences des travaux de sortie métal-GaAlAs, et GaAlAs-GaAs et de la tension de polarisation V_g du métal, la diode Schottky métal-GaAlAs et l'hétérojonction GaAlAs-GaAs sont polarisées. L'allure de la structure de bande du dispositif est représentée sur la figure (10-34). Les électrons libérés par les donneurs de GaAlAs sont transférés dans GaAs à l'interface de l'hétérojonction. La

charge d'espace Q_2 développée dans GaAlAs correspond à un régime de déplétion, la charge d'espace Q_1 développée dans GaAs peut être qualifiée d'accumulation ou d'inversion suivant que le GaAs non dopé est résiduellement n^- ou résiduellement p^-, dans la suite nous la qualifierons d'inversion.

L'épaisseur d de GaAlAs et la tension de polarisation V_g du métal sont telles que les zones de déplétion de la diode Schottky et de l'hétérojonction se recouvrent. Nous dirons que le GaAlAs est en déplétion totale. La couche de GaAlAs est alors isolante ce qui permet, comme dans une capacité MOS, de commander la population n d'électrons de la couche d'inversion de GaAs par la tension de grille V_g , Les niveaux de Fermi sont horizontaux, d'une part dans le métal et d'autre part dans les semiconducteurs, décalés l'un de l'autre de l'énergie $-eV_g$ résultant de la tension de polarisation V_g de la grille.

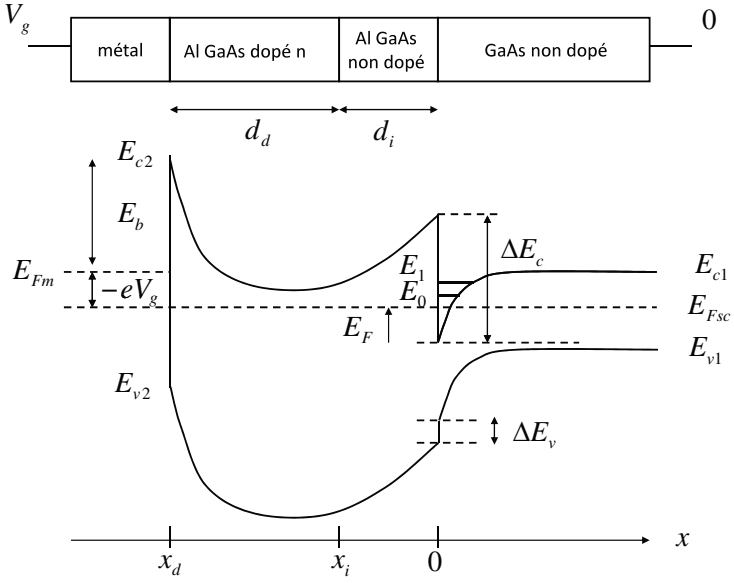


Figure 10-34 : Diagramme énergétique

Nous allons établir la relation $n(V_g)$ qui lie la densité superficielle d'électrons dans le canal du transistor, c'est-à-dire dans la couche d'inversion de GaAs, à la tension de commande V_g appliquée à la grille. Une première relation du type $n(V_g,E_F)$ peut être établie en intégrant l'équation de Poisson dans le GaAlAs. Une deuxième relation du type $n(E_F)$ peut être établie dans la couche d'inversion. La résolution du système de deux équations ainsi obtenu permet alors d'atteindre $n(V_g)$. Nous affecterons l'indice 1 aux paramètres associés au GaAs(SC_1) et l'indice 2 aux paramètres associés au GaAlAs(SC_2).

a. Relation $n(V_g,E_F)$

La continuité du vecteur déplacement à l'interface GaAlAs-GaAs s'écrit

$$\varepsilon_1 E_1(o) = \varepsilon_2 E_2(o) \tag{10-191}$$

où ε_1 et ε_2 sont respectivement les contantes diélectriques de GaAs et GaAlAs.

Pour atteindre $E_2(o)$ il suffit d'intégrer l'équation de Poisson dans le GaAlAs, pour $x_i < x < O$ et $x_d < x < x_i$ successivement :

- Dans le *spacer* non dopé, c'est-à-dire par $x_i < x < o$, la charge d'espace est nulle de sorte que le champ électrique est constant et en particulier égal à $E_2(o)$.

$$\frac{d^2V}{dx^2} = o \quad \Rightarrow \quad \frac{dV}{dx} = Cte = -E_2(o) \tag{10-192}$$

En appelant V_o et V_{xi} le potentiel en $x = o$ et $x = x_i$ respectivement, et d_i l'épaisseur de la région intrinsèque, c'est-à-dire du *spacer*, une deuxième intégration donne

$$V_{xi} - V_o = \int_0^{x_i} -E_2(o)dx = -E_2(o)x_i = E_2(o)d_i \tag{10-193}$$

- Dans la région dopée de GaAlAs, c'est-à-dire pour $x_d < x < x_i$, et dans la mesure où la déplétion est totale et le dopage homogène, la charge d'espace est eN_d. Ainsi

$$\frac{d^2V}{dx^2} = -\frac{e\,N_d}{\varepsilon_2}$$

$$\frac{dV}{dx} = -\frac{e\,N_d}{\varepsilon_2}x + Cte = -\frac{e\,N_d}{\varepsilon_2}(x - x_i) - E_{xi} \tag{10-194}$$

La continuité du vecteur déplacement en $x = x_i$ permet d'écrire $E_{xi} = E_2(o)$. Une deuxième intégration entre x_d et x_i donne

$$V_{xd} - V_{xi} = -\frac{e\,N_d}{2\,\varepsilon_2}d_d^2 + E_2(o)\,d_d \tag{10-195}$$

Les expressions (10-193 et 194) permettent d'écrire l'expression de la différence de potentiel existant aux bornes du GaAlAs d'épaisseur $d = d_d + d_i$

$$\Delta V = V_{xd} - V_o = -\frac{e\,N_d}{2\,\varepsilon_2}\,d_d^{\,2} + E_2(o)\,d \tag{10-196}$$

Le diagramme énergétique de la figure (10-34) permet de relier cette différence de potentiel à la tension de polarisation V_g

$$\Delta V = V_{xd} - V_o = -\frac{1}{e}\left(E_{c2}(x_d) - E_{c2}(o)\right) \tag{10-197}$$

En appelant E_b la hauteur de la barrière de Schottky métal-GaAlAs, ΔE_c la discontinuité de bandes de conduction de l'hétérojonction GaAlAs-GaAs et E_F l'énergie de dégénérescence du puits d'interface de GaAs, on peut écrire les relations

$$E_{c2}(x_d) = E_{c1}(o) + E_F - eV_g + E_b \tag{10-198-a}$$

$$E_{c2}(o) = E_{c1}(o) + \Delta E_c \tag{10-198-b}$$

De sorte que ΔV s'écrit

$$\Delta V = -\frac{E_F}{e} + V_g - \frac{E_b}{e} + \frac{\Delta E_c}{e} \tag{10-199}$$

Les relations (10-196 et 199) donnent

$$E_2(o) = \frac{1}{d}\left(V_g - \frac{E_F}{e} - V_t\right) \tag{10-200}$$

où la tension V_t est donnée par

$$V_t = \frac{E_b}{e} - \frac{\Delta E_c}{e} - \frac{eN_d}{2\varepsilon_2} d_d^2 \tag{10-201}$$

Nous devons maintenant calculer $E_1(o)$ dans GaAs. Le GaAs étant peu ou pas dopé, la charge d'espace présente est essentiellement constituée par les électrons de la couche d'inversion. Le théorème de Gauss appliqué à un cylindre d'axe x et de base unité dans le GaAs permet d'écrire

$$E_1(o) = -\frac{Q_1}{\varepsilon_1} = \frac{en}{\varepsilon_1} \tag{10-202}$$

La continuité du vecteur déplacement à l'interface GaAlAs-GaAs (Eq. 10-191) et les relations (10-200 et 202) permettent d'établir la relation $n(V_g, E_F)$

$$n = \frac{\varepsilon_2}{ed}\left(V_g - \frac{E_F}{e} - V_t\right) \tag{10-203}$$

Remarquons que les quantités n et $E_2(o)$ sont nulles simultanément pour $V_g = V_{th} = V_t + E_F / e$. Les relations $E_2(o) = o$ et $n = o$ traduisent respectivement le régime de bandes plates de l'hétérojonction et l'absence de charges d'inversion. La tension V_{th} est la tension de seuil du transistor.

b. Relation n (E_F) dans GaAs

Dans la couche d'inversion de l'hétérojonction, les électrons sont confinés, dans GaAs, dans un puits de potentiel très étroit dont l'épaisseur typique est de l'ordre de quelques nanomètres. Ces porteurs libres se comportent alors comme un gaz d'électrons bidimensionnel dont le mouvement est libre dans le plan (yz) de la structure et quantifié dans la direction (x) perpendiculaire. Il en résulte que le calcul classique de la distribution de ces porteurs libres en termes de structure de bandes et densités d'états tridimensionnelles est en défaut. Une étude détaillée du comportement de ces porteurs passe par un traitement quantique du problème.

La résolution de l'équation de Schrödinger de ces électrons bidimensionnels montre que les états électroniques sont distribués dans des sous-bandes d'énergie données par

$$E = E_{c1}(o) + E_i + \frac{\hbar^2}{2m_e}(k_y^2 + k_z^2) \tag{10-204}$$

dans lesquelles la densité d'états est constante et donnée par

$$g(E) = \frac{m_e}{\pi\hbar^2} \tag{10-205}$$

L'énergie E_i du bas de chaque sous-bande est donnée, dans l'approximation du potentiel triangulaire, par

$$E_i = \left(\frac{\hbar^2}{2m_e}\right)^{1/3}\left(\frac{3}{2}\pi e E_{eff}\left(i + \frac{3}{4}\right)\right)^{2/3} \tag{10-206}$$

avec $i = 0,1,2...$; E_{eff} représente le champ électrique effectif présent dans la zone de charge d'espace. Sa valeur moyenne est donnée par

$$E_{eff} = \frac{e(N_{dep} + n/2)}{\varepsilon_1} \tag{10-207}$$

En régime d'inversion et dans la mesure où le GaAs est peu dopé, on peut négliger la charge de déplétion N_{dep} devant n de sorte qu'en explicitant E_{eff} dans (10-206), E_i s'écrit

$$E_i = \gamma_i n^{2/3} \tag{10-208}$$

avec

$$\gamma_i = \left(\frac{\hbar^2}{2m_e}\right)^{1/3}\left(\frac{3\pi e^2}{4\varepsilon_1}\left(i + \frac{3}{4}\right)\right)^{2/3} \tag{10-209}$$

La population électronique de chaque sous-bande est donnée par l'intégrale

$$n_i = \int_{E_i}^{\infty} g(E) f(E) dE$$

Soit

$$n_i = \frac{m_e}{\pi \hbar^2} kT \, Ln\left(1 + e^{(E_F - E_i)/kT}\right) \qquad (10\text{-}210)$$

La densité superficielle totale d'électrons dans la couche d'inversion s'écrit donc

$$n = \frac{m_e}{\pi \hbar^2} kT \sum_i Ln\left(1 + e^{(E_F - E_i)/kT}\right) \qquad (10\text{-}211)$$

Dans la pratique, la densité d'électrons est telle que seule les deux premières sous-bandes (i=0 et 1)sont peuplées.

c. Relation $n(V_g)$

En limitant la population électronique de la couche d'inversion aux deux premières sous-bandes, les équations (10-203 et 211) permettent d'écrire le système

$$n = \frac{\varepsilon_2}{ed}(V_g - \frac{E_F}{e} - V_t) \qquad (10\text{-}212\text{-a})$$

$$n = \frac{m_e}{\pi \hbar^2} kT \, Ln\left((1 + e^{(E_F - E_o)/kT})(1 + e^{(E_F - E_1)/kT})\right) \qquad (10\text{-}212\text{-b})$$

avec

$$V_t = \frac{E_b}{e} - \frac{\Delta E_c}{e} - \frac{eN_d}{2\varepsilon_2} d_d^2$$

$$E_o = \gamma_o n^{2/3} \qquad E_1 = \gamma_1 n^{2/3}$$

$$\gamma_o = (\frac{\hbar^2}{2m_e})^{1/3} (\frac{9}{16} \frac{\pi e^2}{\varepsilon_1})^{2/3}$$

$$\gamma_1 = (\frac{\hbar^2}{2m_e})^{1/3} (\frac{21}{16} \frac{\pi e^2}{\varepsilon_1})^{2/3}$$

Le système (10-212) ne présente pas de solution analytique exacte et nécessite une résolution auto-cohérente. Néanmoins on peut obtenir des solutions approchées dans deux gammes de polarisation spécifiques correspondant à deux gammes de densités électroniques dans la couche d'inversion. La première, que nous qualifierons de faible inversion, correspond à une densité d'électrons associée à un niveau de Fermi situé au-dessous de la première sous-bande E_o, la deuxième que nous qualifierons de forte inversion, correspond à une densité d'électrons associée à un niveau de Fermi situé au-dessus de la deuxième sous-bande E_1.

- *Régime de faible inversion*

Pour des faibles densités d'électrons d'interface le niveau de Fermi reste situé au-dessous de la première sous-bande. Les exposants qui apparaissent dans l'équation (10-212-b) sont alors négatifs de sorte qu'un développement limité ($Ln(1+\varepsilon) \approx \varepsilon$) permet d'écrire cette équation

$$n = \frac{m_e}{\pi\hbar^2} kT \left(e^{(E_F - E_o)/kT} + e^{(E_F - E_1)/kT}\right) \tag{10-213}$$

Soit

$$n = \frac{m_e}{\pi\hbar^2} kT\, e^{E_F/kT} \left(e^{-E_o/kT} + e^{-E_1/kT}\right) \tag{10-214}$$

E_o et E_1 sont proportionnels à $n^{2/3}$ de sorte qu'en régime de faible inversion ces quantités sont inférieures à kT et l'expression (10-214) se réduit à

$$n = 2\frac{m_e}{\pi\hbar^2} kT\, e^{E_F/kT} \tag{10-215}$$

ainsi l'énergie de Fermi E_F est liée à la densité électronique par la relation

$$E_F = kT\, Ln\left(\frac{n}{2m_e kT/\pi\hbar^2}\right) \tag{10-216}$$

En portant cette expression de E_F dans l'équation (10-212-a) on obtient la relation

$$n + \frac{\varepsilon_2}{ed}\frac{kT}{e} Ln\left(\frac{n}{2m_e kT/\pi\hbar^2}\right) = \frac{\varepsilon_2}{ed}(V_g - V_t) \tag{10-217}$$

En régime de faible inversion, le niveau de Fermi étant situé au-dessous de la première sous-bande le taux d'occupation des sous-bandes est faible. En d'autres termes la densité d'électrons n est faible devant la densité d'états $2m_e/\pi\hbar^2$ associée aux deux premières sous-bandes. Il en résulte que l'argument du terme logarithmique est très inférieur à 1 de sorte que ce terme devient prépondérant dans le premier membre de l'expression (10-217), qui s'écrit

$$Ln\left(\frac{n}{2m_e kT/\pi\hbar^2}\right) = \frac{e}{kT}(V_g - V_t) \tag{10-218}$$

La densité d'électrons est donc reliée à la tension de grille par la loi exponentielle

$$n = 2\frac{m_e}{\pi\hbar^2} kT\, e^{e(V_g - V_t)/kT} \tag{10-219}$$

Il faut noter qu'en régime de faible inversion le puits de potentiel est à la fois peu profond et large, il en résulte que la quantification des états électroniques est faible. Le gaz d'électrons présente alors un caractère plus tridimensionnel que bidimensionnel, les différentes sous-bandes d'énergie E_o, E_1, ... sont très proches les unes des autres et tendent à se confondre avec la bande de conduction tridimensionnelle. Il est ainsi peu justifié de ne considérer que les deux premières sous-bandes. Une résolution classique du problème est dans ce cas mieux adaptée. Cette résolution a été développée dans le cas de la structure MOS. Il faut noter que la loi de variation ainsi obtenue est qualitativement comparable à l'expression (10-219).

- *Régime de forte inversion*

Lorsque la population électronique devient suffisamment importante, le niveau de Fermi passe au-dessus du bas de la deuxième sous-bande E_1. Les termes exponentiels dans l'équation (10-212-b) deviennent alors très supérieurs à 1 et cette équation s'écrit

$$n = \frac{m_e}{\pi\hbar^2}(2E_F - E_o - E_1)$$

L'énergie de Fermi est donc donnée par

$$E_F = \frac{E_o + E_1}{2} + \frac{\pi\hbar^2}{2m_e}n$$

En explicitant E_o et E_1 en fonction n, E_F s'écrit

$$E_F = \frac{\gamma_o + \gamma_1}{2}n^{2/3} + \frac{\pi\hbar^2}{2m_e}n$$

En portant cette expression dans l'équation (10-212-a) on obtient la relation

$$n(1 + \frac{\varepsilon_2\pi\hbar^2}{2e^2dm_e}) + \frac{\varepsilon_2(\gamma_o + \gamma_1)}{2e^2d}n^{2/3} = \frac{\varepsilon_2}{ed}(V_g - V_t)$$

En régime de forte inversion le terme linéaire en n est prépondérant de sorte que la densité d'électrons suit une loi sensiblement linéaire donnée par

$$n = \beta(V_g - V_t) \tag{10-220}$$

avec

$$\beta = \frac{2\varepsilon_2 em_e}{2de^2m_e + \varepsilon_2\pi\hbar^2}$$

10.4.3. *Polarisation de drain*

Lorsque le drain du transistor est polarisé par une tension drain-source V_d, la polarisation de la structure est distribuée longitudinalement (Fig. 10-35). Il en est de même de la densité d'électrons qui varie alors le long du canal et devient $n(y)$ donné, en régime de forte inversion, par

$$n = \beta(V_g - V_t - V) \tag{10-221}$$

où V, qui est fonction de y, varie de $V = 0$ côté source à $V = V_d$ côté drain.

Le courant de drain, compté positivement dans le sens drain-source, s'écrit

$$I_d = Z e n v \tag{10-222}$$

où Z représente la largeur du canal et v la vitesse d'entraînement des porteurs donnée par $v = -\mu E = \mu dV / dy$. Comme dans le cas du transistor MOSFET, le courant s'écrit donc

$$I_d = Z e \mu n \frac{dV}{dy} \tag{10-223}$$

En explicitant la densité d'électrons, cette expression s'écrit

$$I_d = Z e \mu \beta (V_g - V_t - V) \frac{dV}{dy} \tag{10-224}$$

Le courant I_d étant conservatif, il suffit d'intégrer cette expression sur toute la longueur du canal pour obtenir la loi de variation $I_d(V_g, V_d)$.

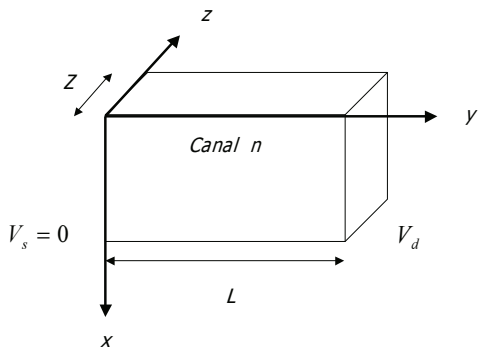


Figure 10-35 : Polarisation du canal

Lorsque V_d est faible la mobilité des porteurs est constante et l'intégration de l'équation (10-224) de 0 à L sur la variable y et de 0 à V_d sur la variable V donne

$$I_d = \frac{Ze\mu\beta}{L}\left((V_g - V_t)V_d - V_d^2/2\right) \tag{10-225}$$

Lorsque V_d devient important la mobilité des porteurs diminue et la vitesse de ces derniers atteint une valeur de saturation v_s pour $V_d = V_{dsat}$. Comme dans le transistor MESFET le courant de drain sature alors à une valeur I_{dsat}, tout simplement parce que la vitesse des porteurs n'augmente plus avec la tension de polarisation. Revenons sur l'expression (10-224), lorsque l'on parcourt le canal de la source vers le drain, V augmente de zéro à V_d et par suite la quantité $(V_g - V_t - V)$ diminue de $(V_g - V_t)$ à $(V_g - V_t - V_d)$. Le courant I_d étant conservatif, il en résulte que le champ électrique dV/dy augmente de la source vers le drain. Ce champ prend donc sa valeur maximum au drain. Ainsi le seuil de saturation se produit lorsque

$$\left(\frac{dV}{dy}\right)_{y=L} = E_s \tag{10-226}$$

où E_s est le champ critique entraînant $v = v_s$.
Côté drain $(y = L)$ le courant I_d s'écrit

$$I_d = Z e \beta (V_g - V_t - V_d)\left(\mu \frac{dV}{dy}\right)_{y=L} \tag{10-227}$$

Quand $V_d = V_{dsat}$, $E_{y=L} = E_s$ *et* $\left(\mu \frac{dV}{dy}\right)_{y=L} = v_s$. Le courant de saturation s'écrit donc

$$I_{dsat} = Z e \beta (V_g - V_t - V_{dsat})\, v_s \tag{10-228}$$

D'autre part pour $V_d = V_{dsat}$ l'équation (10-225) s'écrit, en remplaçant μ par v_s / E_s

$$I_{dsat} = \frac{Ze\beta}{L}\frac{v_s}{E_s}\left((V_g - V_t)V_{dsat} - V_{dsat}^2/2\right) \tag{10-229}$$

En égalant les expressions (10-228 et 229) on obtient une équation du second degré permettant de calculer la tension de saturation V_{dsat}.

$$V_{dsat}{}^2 - 2(V_g - V_t + V_s)V_{dsat} + 2V_s(V_g - V_t) = 0 \tag{10-230}$$

où $V_s = L E_s$ correspond à la tension drain-source qui établirait le champ critique E_s sur toute la longueur du canal.

$$V_{dsat} = V_g - V_t + V_s - \left((V_g - V_t)^2 + V_s{}^2\right)^{1/2} \tag{10-231}$$

On obtient le courant de saturation en explicitant V_{dsat} dans l'expression (10-229).

$$I_{dsat}=g_o\left((V_g-V_t)^2+V_s^2\right)^{1/2}-V_s\Big) \tag{10-232}$$

où $g_o=Ze\beta v_s$ a les dimensions d'une conductance.

Dans un transistor à canal long, $V_s=L\,E_s$ est très important de sorte que les expressions (10-231 et 232) se simplifient

$$V_{dsat}\approx V_g-V_t \tag{10-233}$$

$$I_{dsat}\approx g_o\,(V_g-V_t)^2\big/2V_s \tag{10-234}$$

Au contraire dans un transistor à canal court, V_s est très inférieur à V_g-V_t et

$$V_{dsat}\approx V_s \tag{10-235}$$

$$I_{dsat}\approx g_o\,(V_g-V_t) \tag{10-236}$$

La vitesse des électrons est alors en régime de saturation sur toute la longueur du canal.

La transconductance du transistor est donnée par $g_m=\partial I_d/\partial V_g$. En régime de saturation l'expression (10-232) donne

$$g_m=g_o\frac{V_g-V_t}{\left((V_g-V_t)^2+V_s{}^2\right)^{1/2}} \tag{10-237}$$

Pour un transistor à canal court g_m prend une valeur maximum donnée par $g_{m\max}\approx g_o$

10.4.4. *Effet MESFET parasite*

En régime normal les zones de déplétion de la diode Schottky métal-GaAlAs et de l'hétérojonction GaAlAs-GaAs se recouvrent, le GaAlAs est donc en déplétion totale. La tension de polarisation V_g de la grille commande alors la densité d'électrons dans le canal conducteur situé dans GaAs. Mais lorsque la tension de polarisation de la grille V_g augmente positivement la barrière de potentiel GaAlAs-métal est abaissée, $E_{c2}(x_d)$ diminue, et la zone de déplétion de la diode Schottky diminue. Lorsque V_g atteint et dépasse une valeur de seuil V_m les deux zones de charge d'espace ne se recouvrent plus et un canal conducteur de largeur w_m (de section Zw_m) peuplé d'une densité N_d d'électrons, apparaît dans GaAlAs (Fig. 10-37)

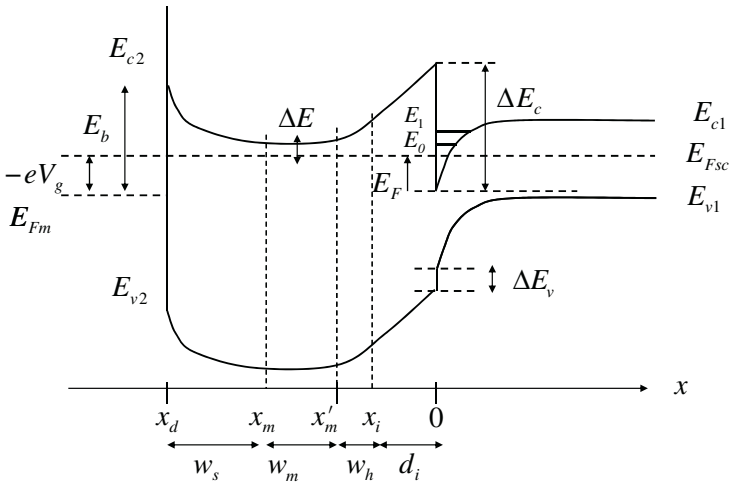


Figure 10-36 : Diagramme énergétique

Dans ces conditions la tension de polarisation $V_g > V_m$ ne commande plus la conductivité du canal du TEGFET, dont la densité électronique sature à une valeur n_m, mais commande la largeur w_m du canal conducteur qui se manifeste dans GaAlAs. En parallèle avec le canal du TEGFET localisé dans GaAs il apparaît donc un deuxième canal conducteur qui n'est autre que le canal du transistor MESFET constitué par l'électrode métallique et la couche de GaAlAs dopée n. Le dispositif se comporte alors comme deux transistors en parallèle (Fig. 10-37) dont seul le MESFET est sensible à la tension de grille $V_g > V_m$.

La largeur w_s de la zone de déplétion de la diode Schottky est donnée par l'intégration de l'équation de Poisson entre x_d et x_m (Fig. 10-36).

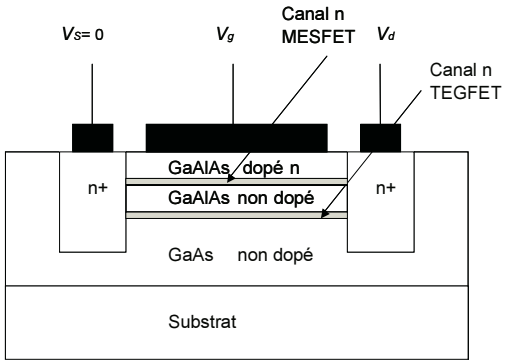


Figure 10-37 : MESFET parasite

Le champ électrique étant nul en $x = x_m$ une première intégration donne

$$\frac{dV}{dx} = -\frac{eN_d}{\varepsilon_2}(x - x_m) \tag{10-238}$$

Une deuxième intégration entre x_d et x_m donne

$$V_{xd} - V_{xm} = -\frac{eN_d}{2\varepsilon_2} w_s^2 \tag{10-239}$$

Le diagramme énergétique de la figure (10-36) permet d'autre part de relier la différence de potentiel $V_{xd} - V_{xm}$ à la tension de grille V_g

$$V_{xd} - V_{xm} = -\frac{1}{e}(E_{c2}(x_d) - E_{c2}(x_m)) = -\frac{1}{e}(E_b - eV_g - \Delta E) \tag{10-240}$$

Ainsi

$$w_s = \left(\frac{2\varepsilon_2}{eN_d}\left(\frac{E_b}{e} - V_g - \frac{\Delta E}{e}\right) \right)^{1/2} \tag{10-241}$$

ΔE qui représente dans le GaAlAs la distance de la bande de conduction au niveau de Fermi est directement lié au dopage N_d par la relation

$$\Delta E = kT\ Ln\frac{N_c}{N_d} \tag{10-242}$$

où N_c est la densité équivalente d'états de la bande de conduction de GaAlAs. Ainsi la largeur w_s de la zone de déplétion de la diode Schottky s'écrit

$$w_s = \left(\frac{2\varepsilon_2}{eN_d}\left(\frac{E_b}{e} - V_g - \frac{kT}{e} Ln\frac{N_c}{N_d}\right) \right)^{1/2} \tag{10-243}$$

La largeur w_h de la zone de déplétion de l'hétérojonction s'obtient simplement en écrivant l'égalité des modules des charges développées dans GaAlAs, $Q_2 = e\,N_d\,w_h$, et dans GaAs, $Q_1 = -en$. Soit

$$w_h = n / N_d$$

La largeur w_m du canal du transistor MESFET s'écrit donc

$$w_m = d - d_i - w_h - w_s$$

$$w_m = d - d_i - \frac{n}{N_d} - \left(\frac{2\varepsilon_2}{eN_d} \left(\frac{E_b}{e} - V_g - \frac{kT}{e} Ln \frac{N_c}{N_d} \right) \right)^{1/2}$$

La tension de seuil V_m à partir de laquelle l'effet parasite du transistor MESFET se manifeste est la tension V_g pour laquelle $w_m = 0$. Soit

$$V_m = \frac{E_b}{e} - \frac{kT}{e} Ln \frac{N_c}{N_d} - \frac{eN_d}{2\varepsilon_2} \left(d - d_i - \frac{n_m}{N_d} \right)^2 \qquad (10\text{-}244)$$

La densité d'électrons dans la couche d'inversion de GaAs est alors égale à n_m et reste égale à cette valeur pour $V_g > V_m$. L'expression (10-220) donne alors $n_m = \beta(V_m - V_t)$. En portant cette expression dans l'équation (10-244) on obtient une équation du second degré en V_m.

$$V_m = \frac{E_b}{e} - \frac{kT}{e} Ln \frac{N_c}{N_d} - \frac{eN_d}{2\varepsilon_2} \left(d - d_i - \frac{\beta}{N_d} (V_m - V_t) \right)^2$$

La tension de seuil de l'effet MESFET est donnée par

$$V_m = V_1 - (V_1^2 - V_2^2)^{1/2} \qquad (10\text{-}245)$$

avec

$$V_1 = V_t + \frac{(d - d_i) N_d}{\beta} - \frac{\varepsilon_2 N_d}{e\beta^2}$$

$$V_2 = \left(\left(V_t + \frac{(d - d_i) N_d}{\beta} \right)^2 - \left(\frac{E_b}{e} - \frac{kT}{e} Ln \frac{N_c}{N_d} \right) \frac{2\varepsilon_2 N_d}{e\beta^2} \right)^{1/2}$$

11. La fabrication collective des composants et des circuits intégrés

Le but de ce chapitre est d'expliquer de manière très simplifiée les procédés utilisés classiquement pour fabriquer les composants semiconducteurs et d'indiquer les évolutions possibles. L'industrie de la micro-électronique est arrivée à un degré de maturité élevé avec la fabrication de circuits intégrés comportant des millions de transistors de tailles submicrométriques tout en garantissant des rendements de fabrication supérieurs à 80%. Des ouvrages entiers sont consacrés à la description des procédés de fabrication qui sont très nombreux et mettent en jeu un grand nombre de paramètres. Ce chapitre est une simple introduction permettant de classer et de définir les procédés principaux. Le flot de fabrication de la technologie CMOS sera plus particulièrement détaillé.

La réduction de taille du transistor avec des longueurs de grille attendues très inférieures au micron conduit à une complexité croissante des équipements de fabrication et à des exigences draconiennes en terme de propreté et de contrôle des impuretés. Ces contraintes entraînent une croissance exponentielle du coût des équipements. De nouvelles techniques sont donc investiguées pour limiter cette explosion des coûts comme les techniques de nanolithographie et les techniques d'auto-assemblage. Le chapitre est organisé de la manière suivante :

11.1. La lithographie
11.2. Les procédés de retrait de matériaux
11.3. Les procédés d'apport de matériaux
11.4. Le flot simplifié pour la technologie CMOS
11.5. Les procédés alternatifs

11.1. La lithographie

Le principe général de fabrication collective des composants consiste à créer sur une tranche de semiconducteur les zones dopées, les zones d'oxyde et les zones de contact nécessaires pour faire un composant donné. Un grand nombre de composants peuvent être fabriqués en même temps puisque la surface d'un composant élémentaire est très petite devant la surface de la tranche de semiconducteur. Cette tranche épaisse de 1 mm environ mais de diamètre pouvant atteindre 300 mm est appelée *wafer*. On peut donc fabriquer simultanément des composants élémentaires (diodes, transistors, diode laser…). Il est également possible de fabriquer des ensembles de composants interconnectés appelés circuits. En pratique l'industrie fabrique sur un wafer de manière collective des milliers de circuits, chaque circuit pouvant comporter des millions de composants élémentaires.

Pour fabriquer des composants à des endroits précis sur le wafer, il faut construire des régions présentant des propriétés différentes dans le semiconducteur. Certaines sont fortement dopées, d'autres faiblement dopées et d'autres isolantes. De plus, il faudra réaliser des connexions conductrices entre zones. Ces régions seront réalisées sur le support silicium de base en ajoutant des dopants, en déposant du métal (aluminium ou cuivre) ou de l'oxyde. Dans tous les cas l'opération devra être strictement limitée aux régions choisies sur le wafer. La lithographie permet de faire ce choix. C'est la technique de base de la fabrication collective des composants.

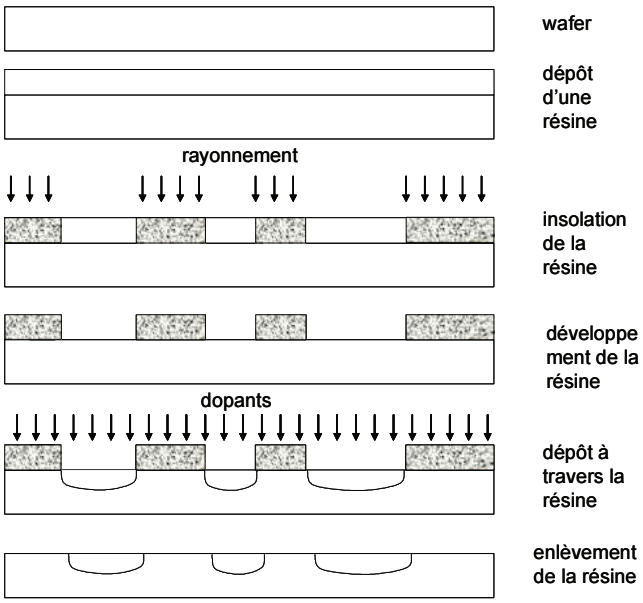


Figure 11-1 : Principe de la fabrication par lithographie

Imaginons le cas simple de la fabrication d'une zone dopée *n* sur un wafer dopé *p*. La figure 11.1 représente les étapes. Reprenons les étapes du procédé. Dans une première phase, une résine est déposée sur le wafer. Dans une seconde étape, un rayonnement est envoyé sur la résine avec une modulation spatiale représentant le motif à réaliser. La façon de réaliser cette modulation sera vue dans la suite. Dans les procédés industriels, le rayonnement est généralement de la lumière et principalement des ultraviolets. Les régions exposées de la résine voient une modification de leurs propriétés chimiques. Elles seront par exemple plus résistantes à l'action d'un solvant. La quatrième étape est la dissolution des régions non exposées de la résine. L'inverse est également possible et dans ce cas, les régions insolées sont dissoutes plus facilement. Dans une cinquième étape, on fait diffuser les ions devant être inclus pour doper le silicium (du phosphore pour un dopage de type *n*) dans les régions non recouvertes de résine. Les techniques permettant cette opération seront étudiées ultérieurement. Enfin, et c'est la sixième étape, on enlève la résine pour obtenir le wafer structuré demandé. Les régions grisées sont dopées *n* et peuvent, par exemple, constituer les puits des transistors PMOS.

Il nous faut maintenant étudier comment moduler la lumière. Deux méthodes sont utilisées : les méthodes parallèles et les méthodes séquentielles. Le principe des méthodes séquentielles est de graver tous les motifs en même temps en se servant d'un masque. Le masque est un objet interposé entre la source de lumière et le wafer qui laisse passer la lumière de manière sélective. Le principe des méthodes séquentielles est d'inscrire point par point le motif à reproduire en guidant le faisceau lumineux. Les deux méthodes sont illustrées figure 11 .2. Quand le faisceau est un faisceau d'électrons et non pas de lumière, la lithographie est dite à faisceau d'électrons et est appelée lithographie « e-beam ».

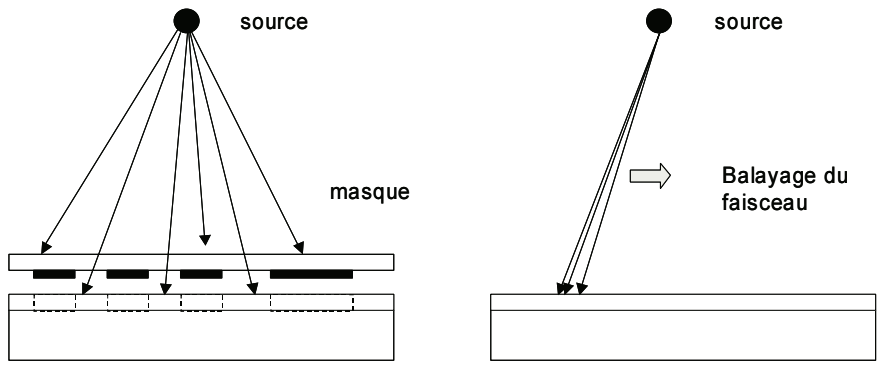


Figure 11-2 : Méthodes parallèle et séquentielle

Dans la lithographie parallèle, un masque interposé entre la source de lumière et le wafer arrête les rayons lumineux aux endroits définis par le design. Il est constitué d'une plaque de quartz transparente aux photons recouverte d'une couche de chrome gravée en fonction des motifs à réaliser. Ce masque très précis est en général fabriqué par une méthode de gravure faisant usage de la lithographie à faisceau d'électrons permettant des précisions bien inférieures au micron. Cela explique cependant le coût élevé d'un jeu de masques.
Dans sa réalisation la plus simple, la lithographie par contact consiste à placer le masque au contact avec le wafer comme le montre la figure 11.3.

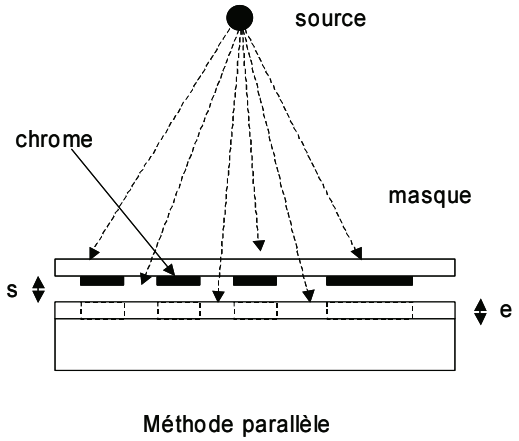


Figure 11.3 : Lithographie par contact

Cette technique très simple a été utilisée jusqu'au milieu des années 70. Elle est limitée par la diffraction de Fresnel qui prédit que le motif le plus fin (a_{min}) réalisable est donné par :

$$a_{\min} = \frac{3}{2}\sqrt{\lambda\left(s + \frac{e}{2}\right)} \qquad (11\text{-}1)$$

Dans cette relation, λ est la longueur d'onde de la lumière utilisée pour insoler et e est l'épaisseur le la résine. La distance s entre le wafer et le masque est minimisée mais non nulle.
Avec une longueur d'onde de 400 nm et une distance de 10 microns entre wafer et masque, on atteint une résolution de 3 microns environ. Cette technique est limitée fondamentalement par la contrainte de planéité du wafer ce qui fixe la distance s minimale à 10 microns. La lithographie par contact n'est plus utilisée de nos jours de manière importante car sa résolution est insuffisante.

La lithographie par projection s'est imposée au fil du temps pour résoudre les problèmes de planéité évoqués précédemment. Elle consiste à interposer un objectif photographique entre le masque et le wafer comme le montre la figure 11-4.

Elle est très contraignante pour l'optique qui doit être de grande qualité. On considère alors le masque comme une source étendue. L'image de cette source est formée sur le wafer par l'optique avec un facteur de réduction dépendant de la distance focale de l'optique. La séparation minimale de deux objets est donnée par la relation classique.

$$a_{\min} = \frac{k\lambda}{n \sin i} \tag{11-4}$$

Dans cette relation de base, λ est la longueur d'onde de la lumière. L'indice du milieu est n et i est l'angle maximum de collection de la lumière défini sur la figure 11.4. Le coefficient k a une valeur théorique de 0.61 mais est en pratique égal à 0.8 pour des installations classiques. Le produit ($n.sini$), appelé ouverture numérique, est passé de 0.3 à 0.9 avec des progrès constants dans la conception des optiques. Pour améliorer la précision de la lithographie, il est donc possible de diminuer la longueur d'onde. C'est le sens de la lithographie UV et à plus long terme de la lithographie à base de rayons X. Il est également possible avec une source UV d'utiliser des techniques plus sophistiquées comme les masques à contraste de phase ou la lithographie à immersion.

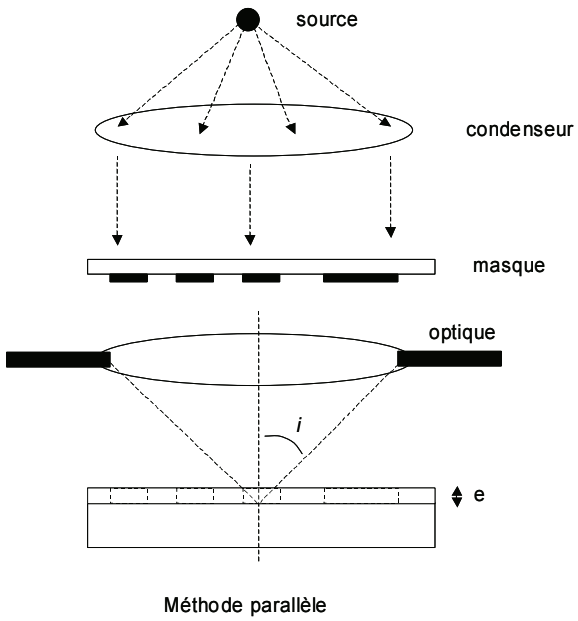


Figure 11.4 : Lithographie par projection

Il ne faudrait pas croire qu'un wafer entier puisse être insolé de cette manière. La zone utile appelée champ est bien plus petite que la surface du wafer. La solution généralement mise en œuvre est d'insoler une zone du wafer

correspondant au champ puis de déplacer le wafer pour insoler une autre zone et ainsi de suite. Cette technique permet d'appliquer un facteur de grandissement G variant entre 5 et 20. Le motif sur le masque peut être G fois plus grand que le motif gravé sur le wafer. Le masque est donc plus facile à fabriquer que s'il était à la même échelle. Si le facteur de grandissement est élevé, il faut déplacer le wafer un grand nombre de fois pour une insolation totale du wafer. Le temps d'insolation sera donc plus important. Il y a un choix optimal à faire en fonction de la résolution nécessaire et du coût de l'opération.

En réalité, quand on fabrique un circuit on utilise non pas un masque mais un jeu de masques car les opérations à effectuer sont nombreuses. Le coût d'un jeu de masques dans une technologie avancée est un véritable problème pour l'industrie micro-électronique. Dans une technologie 50 nm des coûts de 4 millions d'euros sont annoncés pour réaliser un circuit intégré.

La précision de l'alignement des masques est un facteur décisif pour la fabrication collective de composants miniaturisés. La figure 11-5 montre les erreurs possibles en cas de mauvais alignement. En pratique les masques sont positionnés les uns par rapport aux autres à l'aide de motifs spéciaux.

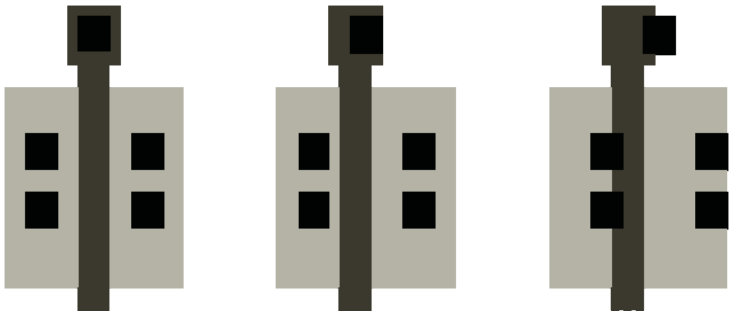


Figure 11.5 : Alignement des masques

11.2. Les procédés de retrait de matériaux

Ils permettent d'enlever de la matière dans des zones définies par la lithographie. Trois procédés sont possibles : la gravure humide, la gravure sèche et la gravure ionique réactive

11.2.1. La gravure humide

Prenons l'exemple de la gravure d'une tranchée dans le silicium. Les autres zones sont protégées par la résine. Le substrat est trempé dans un bain chimique attaquant le silicium mais pas la résine. Une zone gravée se forme d'autant plus profonde que le temps d'immersion est élevé. Des vitesses de gravure de plusieurs microns par minute sont possibles en fonction de la concentration de la solution d'attaque. La figure 11.6 illustre ce procédé de base. La gravure est un procédé

simple mais présente l'inconvénient de la sous-gravure c'est-à-dire la gravure sous la résine. Ce procédé est donc peu adapté à l'obtention de motifs fins et à l'obtention de flancs de gravure raides.

L'orientation du réseau cristallin change les vitesses de gravure. Les plans les plus denses du silicium sont par exemple gravés beaucoup plus lentement que les autres régions. Il est donc possible en orientant convenablement le wafer de réaliser des gravures non isotropes comme le montre la figure 11.6.

On peut graver le silicium mais aussi les oxydes et les métaux comme le montre le tableau ci-dessous.

Matériau à graver	Agent de gravure
Silicium	Soude (KOH)
Dioxyde de silicium	Acide fluoridrique (HF)
résine	$H_2SO_4 + H_2O_2$

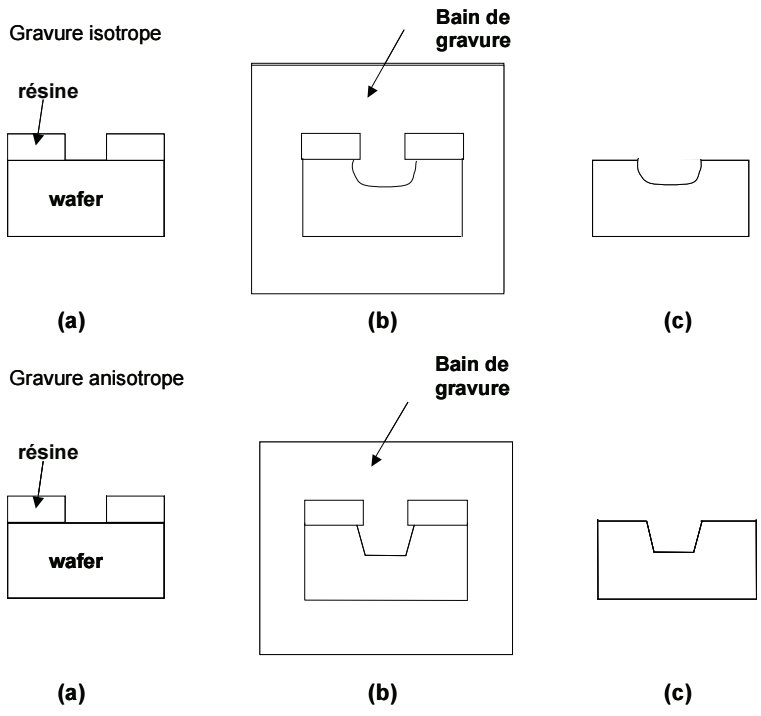


Figure 11.6 : Les procédés de gravure humide

11.2.2. *La gravure sèche*

Le principe n'est plus de contrôler une réaction chimique dans un bain mais de bombarder les zones non protégées de la surface avec des ions. L'opération se passe dans une enceinte sous vide comme le montre la figure 11.7. Trois techniques sont alors possibles : la gravure dite par « sputtering » , la gravure par plasma, la gravure ionique réactive (RIE).

Ces trois techniques font usage de deux principes physiques : le premier est l'arrachement d'un atome de la surface par collision élastique avec un ion incident (*sputtering*), le second est l'activité chimique en surface du wafer quand un plasma est créé dans l'enceinte à vide. Ce plasma est créé par un champ radiofréquence à 13,56 MHz.

La gravure par *sputtering* utilise le premier effet ; elle est purement mécanique. La gravure par plasma utilise le deuxième effet ; elle est purement chimique. La gravure ionique réactive combine les deux effets et permet ainsi de décaper la surface avec une grande efficacité. Elle est donc très largement mise en œuvre dans les procédés actuels de fabrication. Elle permet également d'atteindre d'excellentes résolutions de gravure (quelques dizaines de nm) et de réaliser des flancs quasi-verticaux. Les trois méthodes sont illustrées figure 11.7.

Ces méthodes peuvent être améliorées de la manière suivante.

- Utilisation de triodes RF : Deux sources RF indépendantes permettent de dissocier le contrôle de la densité du plasma et de l'énergie des ions sur la plaquette
- Méthode MERIE (Magnetically Enhanced Reactive Ion Etching) : Ajout d'un champ magnétique qui permet d'augmenter la densité de plasma jusqu'à 5.1010 cm^{-3}
- Réacteurs à décharge inductive (ICP ou TCP) : Le plasma est entretenu par couplage inductif RF, et non plus capacitif, ce qui permet d'atteindre des densités de plasma très élevées 1012 cm-3 avec des pressions de quelques mTorrs
- Réacteurs ECR (Electron Cyclotron Resonance) : L'énergie est fournie aux électrons du plasma par une micro-onde dont la fréquence est égale à la fréquence de Larmor des électrons sous champ magnétique permanent (2.45 GHz pour un champ magnétique de 875 Gauss)

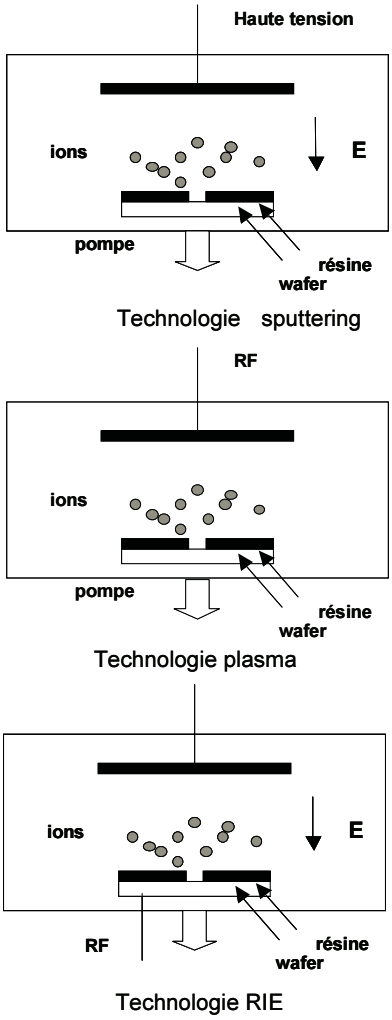


Figure 11-7 : Les trois méthodes de gravure sèche

Certains réacteurs sont configurés en mode « remote plasma ». Le plasma est généré à une distance assez grande de la plaquette. La gravure est alors essentiellement chimique, par les radicaux réactifs qui diffusent jusqu'à la plaquette

11.3. Les procédés d'apport de matériaux

Il s'agit non plus d'enlever un matériau existant mais d'amener dans des régions définies par la lithographie les matériaux nécessaires à la fabrication des composants. Les matériaux sont du silicium, des isolants (dioxyde de silicium ou nitrure de silicium) et des métaux (aluminium, cuivre, cobalt, titane et tungstène).

Ce sont aussi des ions (Bore, arsenic et phosphore) qui sont ajoutés au silicium pour modifier le dopage.

Les procédés sont assez nombreux : implantation ionique, PVD, CVD, croissance thermique, croissance électrolytique. Ils sont classés dans la figure 11.8 selon trois types d'apport. Les procédés de dépôt consistent à amener sur le wafer des matériaux existants. Les dépôts sont alors des couches minces (épaisseur inférieure au micron). Les procédés de croissance consistent à déclencher la croissance du matériau à partir de ses constituants chimiques. Enfin, les procédés d'apport en profondeur consistent à introduire des atomes dans le wafer pour changer les propriétés électriques.

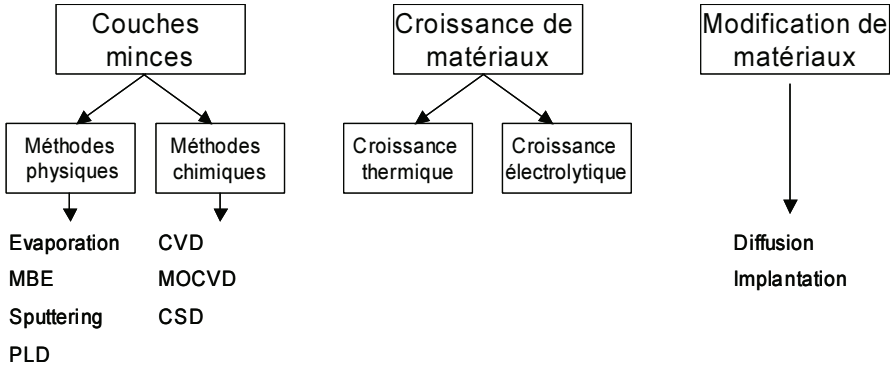


Figure 11.8 : Les procédés d'apport de matériaux

11.3.1. *Evaporation*

C'est un procédé très simple qui consiste à chauffer un matériau dans un creuset, à le transformer en phase vapeur puis à provoquer la croissance du même matériau en phase solide sur les zones non protégées du wafer. Ce sont les métaux qui sont principalement déposés par cette technique car le point de fusion est relativement bas. Le matériau peut être chauffé par une résistance mais aussi par un faisceau d'électrons quand il est nécessaire d'atteindre des températures plus importantes. Les taux de déposition sont élevés, jusqu'à 5000 nm par minute. Ce procédé assez simple est mis en œuvre dans de nombreux laboratoires de recherche.

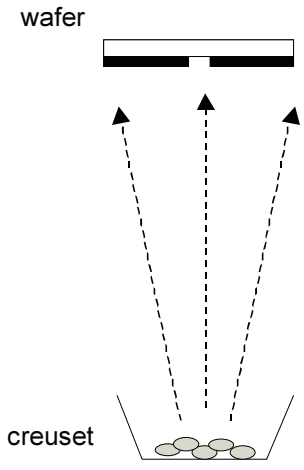


Figure 11.9 : Évaporation

11.3.2. *Epitaxie par jets moléculaires (MBE)*

C'est une amélioration de la technique d'évaporation qui bénéficie des techniques d'ultra vide. Différentes sources sont présentes dans l'enceinte sous vide. Des obturateurs rapides sont placés devant chaque source et des équipements de caractérisation sont en général ajoutés pour contrôler les dépôts. Un équipement type est symbolisé figure 11.10.

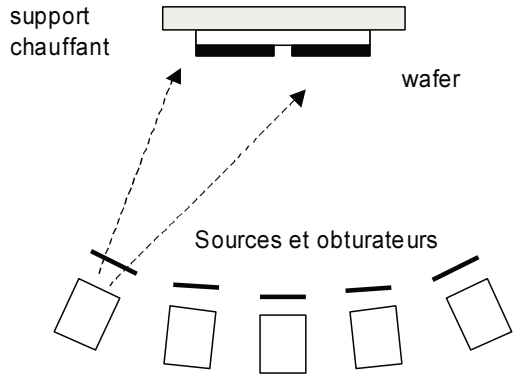


Figure 11.10 : Épitaxie par jets moléculaires

Cette technique permet en particulier de réaliser des dépôts multicouches. Des couches très fines (quelques couches atomiques) et monocristallines peuvent être fabriquées par cette technique. La méthode est donc toute indiquée pour réaliser des structures multicouches. Le procédé est assez long (un micron par heure environ) et est principalement utilisé dans les laboratoires de recherche. Des systèmes chauffant ou des canons à électrons peuvent être utilisés pour réaliser les dépôts.

11.3.3. *Dépôt par « sputtering »*

Cette technique est principalement utilisée pour déposer des contacts métalliques. Elle a remplacé d'année en année la méthode par évaporation car les taux de contamination peuvent être très faibles. L'équipement est le même que celui de la gravure par « sputtering » à la différence près que l'anode est le wafer et que la cathode est une cible faite dans le matériau qu'il faut déposer. Un gaz inerte comme l'argon est ionisé et les ions sont accélérés vers la cible. Des atomes de la cible sont éjectés et vont se déposer sur le wafer. Le *sputtering* RF peut également être utilisé pour déposer des isolants.

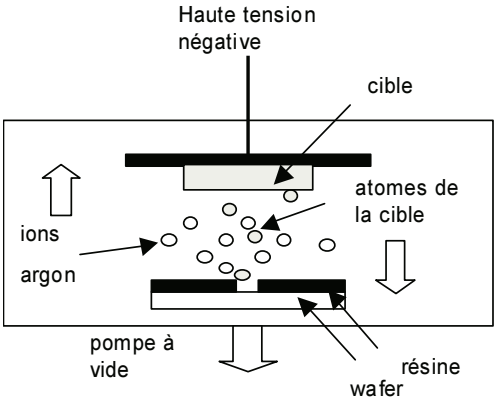


Figure 11.11 : Dépôt par sputtering

11.3.4. *Dépôt par laser pulsé (PLD)*

Cette technique plus récente est mise en œuvre pour déposer des isolants en multicouches. Un laser pulsé de puissance (un Joule par impulsion environ) forme un plasma au niveau de la cible.

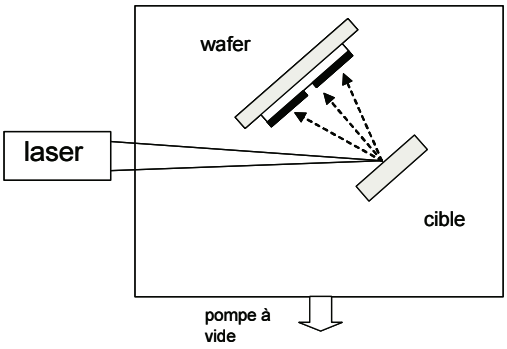


Figure 11.12 : Dépôt par ablation laser

Ce plasma contient des atomes, des ions et des molécules de la cible. Ces composés se déposent sur le wafer. Le dispositif est représenté figure 11.12.

11.3.5. *Le dépôt en phase vapeur (CVD)*

Cette méthode est très utilisée pour fabriquer des composants et des circuits puisqu'elle permet de dépose des semiconducteurs, des oxydes et des métaux. Le principe est de faire croître sur un substrat une couche relativement mince à partir de composants en phase vapeur appelés précurseurs. Le substrat est chauffé dans un dispositif comme celui de la figure 14.

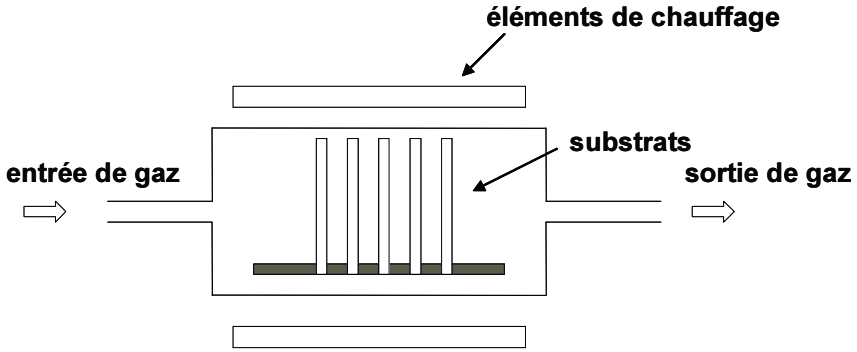


Figure 11.13 : Réacteur CVD

Différentes techniques sont possibles :
- pression atmosphérique : APCVD,
- basse pression : LPCVD,
- haute pression : HPCVD.

Le dépôt à basse pression se fait à plus haute température. Le dépôt à haute pression peut se faire à plus basse température. Il est possible de déposer des semi-conducteurs et des isolants en exploitant les réactions chimiques suivantes :

- dépôt de silicium polycristallin :
  $SiH_4 \rightarrow Si + 2H_2$ 580-650 °C pour une pression de 1 mbar

- dépôt de nitrure de silicium :
  $3\ S_iH_4 + 4NH_3 \rightarrow Si_3N_4 + 12\ H_2$ 700-900 °C à la pression atmosphérique

- dépôt de silice :
  $SiH_4 + O_2 \rightarrow SiO_2 + 2H_2$ 450 °C

Ce sont les réactions les plus classiques. Elles utilisent un gaz appelé silane (SiH_4). Le dioxyde de silicium produit par CVD ne peut cependant pas remplacer celui formé par oxydation thermique dans un four soumis à un flux d'oxygène ou de vapeur d'eau. L'oxyde ainsi formé a d'excellentes propriétés électriques. La CVD

est également utilisée pour déposer des métaux comme le tungstène, l'aluminium et le titane. Le cuivre pourrait aussi être déposé par cette méthode mais une méthode spécifique appelée « damascène » est généralement mise en œuvre.

11.3.6. *Dépôt de type MOCVD*

Le dépôt MOCVD (*Metal Organic Chimical Vapour Deposition*) est une évolution de la méthode CVD adaptée au dépôt de composés métalliques. Les précurseurs contiennent le métal à déposer et des composés organiques. Cette technique permet en particulier de déposer les matériaux à haute permittivité nécessaires à la fabrication des DRAMs et autres composants de la micro-électronique. Prenons comme exemple le dépôt de PZT (composé à base de plomb, de titane et de zirconium). Les précurseurs sont les liquides suivants : $Pb(C_2H_5)_4$, $Ti(OC_3H_7)_4$, $Zr(OC_4H_9)_4$.

Comme exemple de réaction, prenons la formation de l'arséniure de gallium à partit du trimemthylgallium.

$$AsH_3 + Ga(CH_3)_3 \rightarrow GaAs + 3CH_4$$

11.3.7. *Dépôt de type CSD (Chemical Solution Deposition)*

Ce sont des méthodes chimiques qui consistent à partir de précurseurs généralement en phase liquide pour arriver à un film polycristallin ou cristallin en passant par une phase amorphe. Il faut également citer les méthodes de type Langmuir-Blodgett pour déposer des composés organiques. Des molécules organiques présentant une extrémité hydrophobe et une autre extrémité hydrophile peuvent être déposées sur un volume d'eau de la même manière qu'un film d'huile se forme au dessus d'un volume liquide. Le film mince ainsi formé peut alors être transféré sur un substrat. Cette technique simple permet en particulier la fabrication des diodes électroluminescentes organiques et laisse espérer le développement d'une électronique grande surface.

11.3.8. *Croissance thermique*

Ce sont les méthodes qui permettent en particulier de fabriquer le wafer lui-même. Leur principe a été donné dans le paragraphe 2.3.5 mais la méthode CVD est limitée à la fabrication d'une couche mince. Dans ce paragraphe, on étudie comment réaliser une tranche de 1 mm d'épaisseur. Les wafers sont découpées dans un lingot de silicium. Un lingot est un cylindre de diamètre élevé (300 mm dans les fabrications les plus avancées) qui présente la propriété importante d'être monocristallin. D'autres semiconducteurs peuvent être fabriqués comme le germanium, l'arséniure de gallium ou le tellurure de cadmium. Le silicium a cependant pris une place largement majoritaire. L'obtention de cristaux monocristallins d'arséniure de gallium ou de tellurure de cadmium reste une

opération difficile. Le procédé de croissance du lingot est le procédé Czochralski. Il est représenté de manière simplifiée figure 11.14.

La méthode utilise le matériau EGS silicium polycristallin fondu de haute pureté. Ce matériau est obtenu à partir de sable de haute pureté après un certain nombre de réactions chimiques. A partir d'un germe de silicium monocristallin, le silicium fondu apporté dans le creuset se solidifie autour du germe au fur et à mesure que le solide ainsi formé se déplace selon un mouvement combinant rotation autour de l'axe du cylindre et translation vers le haut. Le silicium dans le creuset est formé à partir de la décomposition de la silice, matériau très abondant dans la nature et d'un processus de purification permettant de contrôler les impuretés. Les déplacements du lingot (rotation et translation) sont très lents si bien que la fabrication de lingots de haute pureté avec un contrôle précis de la structure cristallographique reste une opération complexe et lente justifiant le prix relativement élevé des wafers. Le lingot de silicium est ensuite découpé en tranches de 1 mm d'épaisseur environ pour assurer un minimum de rigidité mécanique. Du Bore et du Phosphore peuvent être apportés pendant la croissance pour obtenir un lingot de résistivité donnée. On peut également fabriquer par cette méthode des lingots d'Arséniure de Gallium ou de Tellurure de Cadmium. Pour ces deux matériaux la méthode « Bridgman » utilisant deux zone de cristallisation est également mise en ouvre.

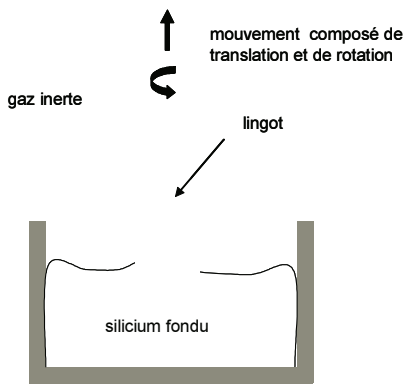


Figure 11.14 : Croissance thermique du silicium

Après la croissance, des tranches d'épaisseur inférieure au mm sont coupées dans le lingot ce qui permet d'obtenir les « wafers ». Il est ensuite nécessaire de polir les faces pour garantir un minimum de planéité puis d'éliminer les zones externes comportant des défauts suite au polissage. Une attaque chimique est donc une étape indispensable. A la fin du processus, il est nécessaire de polir les faces pour être compatible avec les très faibles dimensions des composants.

Les tranches obtenues ne sont pas des cristaux parfaits. Des impuretés peuvent occuper des sites substitutionnels (ils remplacent un atome du cristal) ou interstitiels (ils sont insérés dans le cristal). Enfin la régularité du réseau cristallin peut être perturbée localement, on parle alors de dislocations. Ces défauts auront une grande influence sur les propriétés électriques des composants fabriqués ultérieurement et un grand défi de la technologie est de réduire au maximum la densité de défauts par wafer.

11.3.9. *La croissance électrolytique*

Cette technique permet d'obtenir des couches relativement épaisses de métal sur un substrat. Prenons l'exemple du dépôt de nickel représenté figure 11.15.

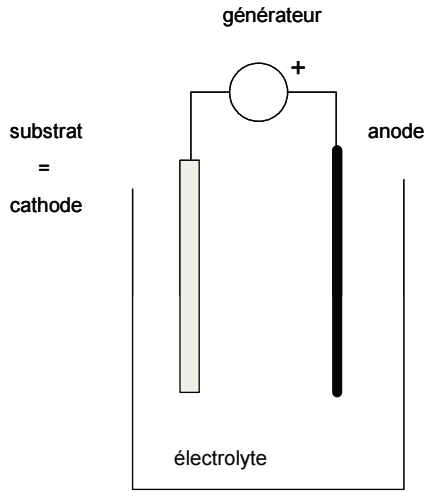


Figure 11.15 : Dépôt par électrolyse

La solution est dans ce cas du chlorure de nickel mélangé à du chlorure de potassium. À la cathode, il y a une réaction de réduction :

$$Ni^{2+} + 2e^- \rightarrow Ni$$

Au niveau de l'anode, il y a une réaction d'oxydation :

$$2Cl^- \rightarrow Cl^2 + 2e^-$$

Il y a donc dépôt de nickel sur le wafer jouant le rôle de cathode. La réaction peut se faire de manière sélective au travers d'ouvertures gravées dans la couche de résine déposée sur le substrat. Si le substrat n'est pas un bon conducteur, il faut déposer par une autre méthode une mince couche de métal qui sert de base de déclenchement de la réaction électrochimique.

11.3.10. *Implantation ionique*

Nous abordons maintenant les techniques qui permettent de modifier des matériaux en profondeur. La notion de profondeur est toute relative puisque les régions sont créées en général à moins d'un micron de la surface du wafer. La zone

active du dispositif est donc située dans une région d'épaisseur négligeable devant l'épaisseur du wafer.

La première technique étudiée est l'implantation ionique. Ce n'est pas réellement une technique permettant de fabriquer en profondeur un matériau mais une technique qui permet de changer les propriétés électriques du matériau de base. Le principe est d'envoyer perpendiculairement à la surface du wafer un flux d'ions de haute énergie. L'énergie de ces ions est suffisamment élevée pour qu'ils pénètrent à l'intérieur de la matière avant d'être arrêtés sous l'effet des interactions avec les électrons des atomes de silicium du wafer. Le principe d'un implanteur ionique est illustré figure 11.16.

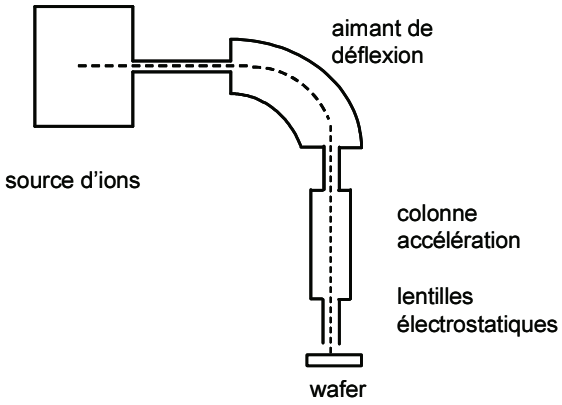


Figure 11.16 : Implantation ionique

Des atomes sont ionisés dans la source puis extraits vers l'aimant de déflexion. Cet aimant est utilisé en spectromètre de masse. La valeur du champ magnétique créé permet de dévier les ions en fonction de leur masse. Les ions choisis seront les seuls à pouvoir être extraits de cette zone de champ magnétique pour être ensuite accélérés par la colonne d'accélération. L'énergie des ions est donc réglable ce qui permet d'ajuster la profondeur de pénétration dans la matière. Des dispositifs de focalisation et de balayage complètent le dispositif. En résumé, l'implanteur permet les réglages suivants :

- choix des dopants par le réglage du champ magnétique,
- choix de la pénétration des ions par le réglage de la tension d'accélération,
- choix de la dose implantée par le réglage de l'intensité du faisceau et du temps d'implantation.

Les énergies d'implantation sont comprises entre quelques keV et quelques MeV. Si on examine le wafer dans sa profondeur après implantation, on obtient une répartition des ions implantés du type de la courbe 11.17. La profondeur

d'implantation ne peut être définie qu'en moyenne car tous les ions ne subissent pas le même nombre de collisions avec les électrons du wafer. La profondeur moyenne dépend de l'énergie des ions et est inférieure au micron. Pour les faibles énergies, les ions sont implantés au voisinage de la surface du wafer.

Il ne suffit pas d'implanter des ions pour créer en profondeur des zones de dopage donné, il faut que les dopants (Bore, Arsenic, Phosphore) occupent des positions substitutionnelles dans le réseau cristallin comme il a été expliqué dans le chapitre 1. C'est à cette condition que les dopants sont électriquement actifs. De plus, l'implantation créée de nombreux défauts dans le réseau cristallin en déplaçant les atomes. Il est donc nécessaire pour ces deux raisons de chauffer le substrat implanté pour guérir les défauts et pour rendre les dopants électriquement actifs et modifier les propriétés électriques. Cette opération s'appelle le recuit et se pratique vers 600 °C.

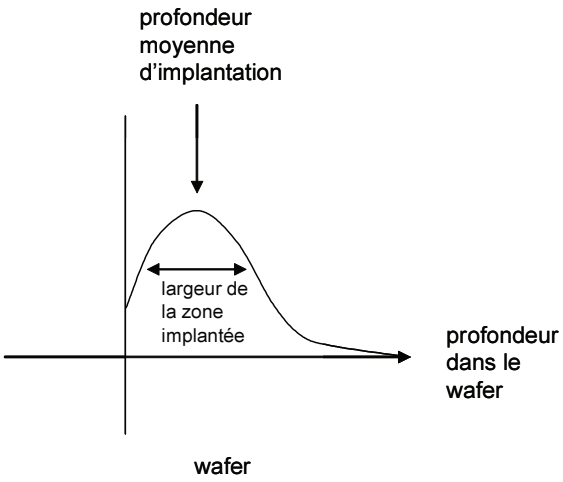


Figure 11.17 : Répartition des dopants après implantation

11.3.11. *La diffusion thermique*

La diffusion thermique est une opération complémentaire de l'implantation ionique qui permet de distribuer des dopants dans un substrat. Si le matériau est chauffé à haute température (au dessus de la température du recuit), les dopants peuvent avoir assez d'énergie pour se déplacer dans la matière en quittant leurs sites initiaux. Ces déplacements dépendent de la température et du temps pendant lequel la haute température est appliquée. La diffusion élargit la distribution initialement créée par l'implantation ionique. Il est également possible de faire diffuser dans le wafer des atomes déposés en surface.

11.4. Le flot simplifié pour la technologie CMOS

Ce paragraphe est fondamental. Il montre comment sont réalisés les circuits CMOS. La description est simplifiée mais est conforme à un procédé industriel réel. Les étapes seront décrites par une série de dessins montrant comment il est possible de réaliser sur un wafer un PMOS, un NMOS et les interconnexions. Dans la pratique, tous les transistors du circuit sont réalisés en même temps. Les dernières étapes permettent de réaliser les connexions entre transistors ce qui permet de définir les fonctions. Le NMOS sera implanté à droite du dessin et le PMOS à gauche. Dans un premier temps, montrons les grandes étapes de la fabrication sans trop se soucier des modes de réalisation comme cela est représenté sur la figure 11.18.

Dans une première étape représentée figure 11.18 (a), le substrat de type *p* est découpé puis poli et traité chimiquement pour être prêt à recevoir les traitements de la micro-électronique. Des tranchées d'isolation entre NMOS et PMOS sont gravées dans le matériau puis remplies d'un oxyde de haute qualité. Le but est d'éviter que des courants de fuite puissent circuler. Il y a en fait quatre technologies possibles : substrat de type *p* et puits de type *n* , substrat de type *n* et puits de type *p*, substrat de type *p* et deux puits *n* et *p*, technologie triple puits. On décrira ici la technologie avec un puits de type *n*.

Dans une seconde étape, les oxydes de grille et les grilles en silicium polycristallin sont réalisés. Notons cependant que des procédés avancés introduisent des grilles métalliques afin de réduire les résistances d'accès.

Dans une troisième étape, les zones de grille et de drain sont réalisées par implantation ionique à travers le silicium polycristallin de grille jouant le rôle de masque. Le procédé est dit auto-aligné et son avantage est de réduire les capacités parasites drain-grille qui ont un effet négatif sur la vitesse de fonctionnement des transistors.

Dans une quatrième étape, on réalise les interconnexions entre transistors ainsi que les contacts traversants pour passer d'un niveau d'interconnexion à un autre. Le schéma 11.18 (c) montre deux niveaux d'interconnexion mais dans les technologies avancées, on peut trouver jusqu'à 7 niveaux. Il ne faut pas oublier la fabrication de l'oxyde entre pistes et l'oxyde de passivation qui protège le circuit des effets chimiques venant de l'extérieur. Certaines zones de cet oxyde de passivation sont ouvertes pour permettre de relier par « bonding » une sortie du circuit à une broche du boîtier. Dans les technologies avancées, les derniers niveaux de métallisation sont utilisés pour les connexions à longue distance et pour réaliser des pseudo plans de masse. Ils sont constitués d'alliages de cuivre ce qui permet de réduire les résistances des pistes. Cette évolution de la technologie CMOS s'est imposée difficilement à cause des effets chimiques introduits par l'électromigration du cuivre.

Rappelons que tous les transistors d'un circuit et que tous les circuits identiques sont réalisés en même temps sur le wafer ce qui permet d'envisager la fabrication simultanée de milliers de circuits intégrés comportant chacun des millions de transistors.

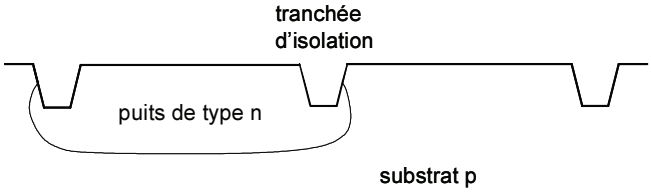


Figure 11.18 (a) : puits et tranchées

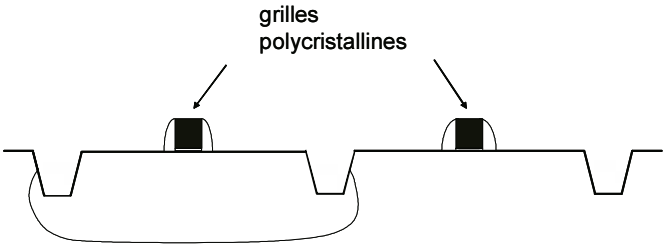


Figure 11.18 (b) : Isolants de grille et grilles

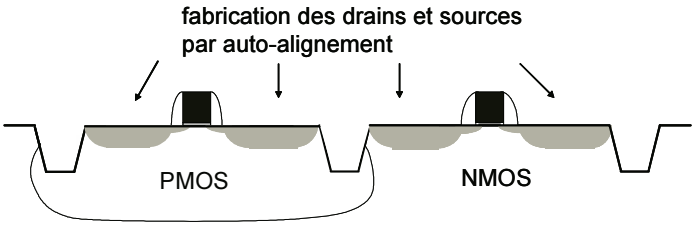


Figure 11.18 (c) : Fabrication des sources et des drains

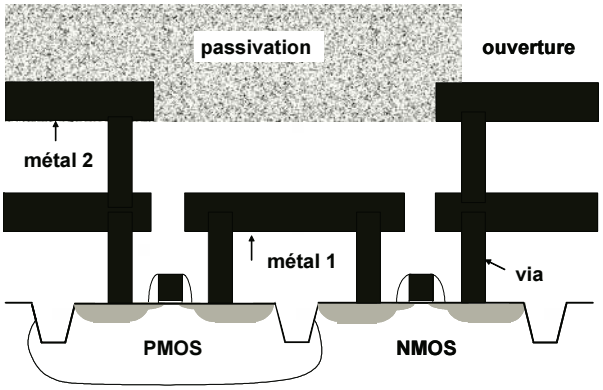


Figure 11.18 (d) : Métallisations et passivation

La technologie CMOS a connu de profondes modifications ces dernières années avec deux évolutions majeures actuellement mises en œuvre dans les technologies les plus avancées.

La première est l'introduction du cuivre pour améliorer les propriétés électriques des interconnexions. La miniaturisation du transistor et des liaisons entre transistors conduit en effet à une augmentation de la constante *RC* d'une ligne par simple effet géométrique. Il faut donc compenser en diminuant la résistivité du métal.
Parallèlement il est également nécessaire de diminuer la capacité par unité de longueur des lignes de connexion ce qui conduit à utiliser des diélectriques poreux de faible permittivité.

La seconde évolution est une modification fondamentale de l'oxyde de grille. Le dioxyde de silicium a été le matériau de choix pour réaliser la capacité MIS du transistor mais la miniaturisation conduit à une réduction excessive de l'épaisseur de l'oxyde et à l'augmentation des fuites par effet tunnel. Il a donc été nécessaire d'introduire des oxydes de plus haute permittivité pour conserver une épaisseur raisonnable tout en assurant un contrôle électrostatique minimum. Cette profonde modification s'est accompagnée par le remplacement du silicium polycristallin de la grille par des matériaux métalliques.

La photographie 11.19 réalisée au microscope électronique montre l'empilement des interconnexions dans un circuit intégré.

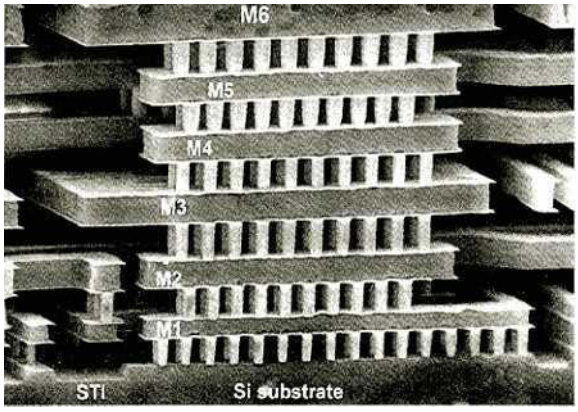


Figure 11.19 : Technologie à six niveaux de métal

La figure 11-20 illustre l'utilisation d'un jeu de masques pour réaliser un circuit intégré.

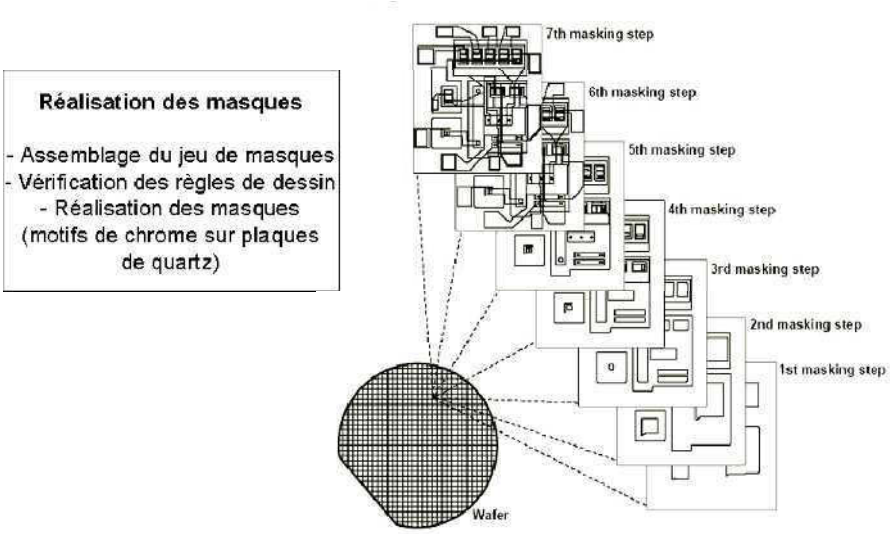


Figure 11-20 : Réalisation d'un jeu de masques

11.5. Les procédés alternatifs

La lithographie a fait la force de la micro-électronique en permettant la fabrication collective des circuits. Elle est cependant la principale source de difficultés pour l'avenir. En effet, les dispositifs de lithographie sont d'une complexité croissante au fur et à mesure que la résolution demandée s'affine. Le coût des équipements devient très élevé. Un dispositif d'insolation avec photorépétition est acheté autour de 10 millions d'euros et un jeu de masques pour réaliser un circuit complexe peut coûter quelques millions d'euros. Les masques sont fabriqués à l'aide d'un appareil générant un faisceau d'électrons capable de balayer la surface du masque avec une précision extrême. Le procédé n'est plus limité par la longueur d'onde puisque la longueur d'onde associée à l'électron est très faible. La résolution possible de gravure est de quelques nm mais le temps d'insolation est très long à cause du caractère séquentiel de l'opération ce qui explique en partie le coût du jeu de masques fabriqué par cette méthode.

On pourrait imaginer des procédés de fabrication permettant de se passer des masques. On pourrait par exemple utiliser la lithographie par faisceau d'électrons non pas pour fabriquer les masques mais pour produire directement les circuits intégrés. Cela est effectivement pratiqué pour réaliser des prototypes. La

lithographie par faisceau d'électrons ne peut cependant être mise en œuvre dans un procédé industriel car le temps d'insolation est long par principe puisqu'il est séquentiel. Il est donc naturel que la micro-électronique investigue des techniques alternatives à la lithographie optique.

11.5.1. *La nano-impression*

Ce procédé a été proposé par S.Y. Chou en 1995. Il est devenu une méthode de référence pour fabriquer des nano dispositifs car très simple de mise en œuvre. Son principe est hérité des technologies de l'impression. Il est illustré figure 11.21.

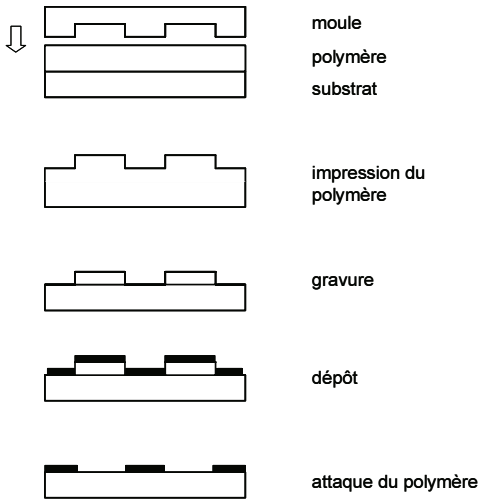


Figure 11.21 : Nano-impression

Le procédé s'explique simplement, Un moule réalisé par lithographie électronique sur un support de silicium écrase un polymère déposé sur le substrat. La pression est de quelques dizaines de bars. La gravure ionique réactive permet ensuite de graver le polymère en atteignant le substrat. Un film mince est ensuite déposé sur le polymère résiduel, un film métallique par exemple. L'étape suivante appelée « lift-off » consiste à enlever le polymère par dissolution chimique. On obtient alors le motif choisi imprimé dans le film déposé. Ce film peut par la suite servir de masque de gravure. Cette technique permet de réaliser des motifs avec une résolution de l'ordre de 10 nm. Le moule peut en effet être fabriqué par une technique mettant en œuvre la lithographie e-beam puisqu'il est fabriqué une seule fois. La nanoimpression est appliquée pour étudier des dispositifs de très faible dimension et pour réaliser des matériaux magnétiques ou optiques nanostructurés.

Des techniques dérivées sont également étudiées comme la nanoimpression sous irradiation, la nanocompression et la lithographie molle. La nanoimpression sous irradiation permet de travailler à température ambiante. La nanocompression utilise le polymère sous forme de granulés. La lithographie molle est basée sur l'utilisation d'un élastomère pour réaliser une sorte de tampon encreur capable de déposer des molécules organiques. On peut également citer la lithographie en champ proche, technique de lithographie par contact mais qui utilise un masque souple en contact parfait avec le substrat. Toutes ces techniques sont encore au stade de la recherche et ne sont pas encore appliquées dans l'industrie.

11.5.2. *Techniques d'auto-assemblage*

Pour éviter les coûts de la lithographie, il est séduisant d'imaginer des structures organisées qui seraient réalisées par des procédés purement chimiques. Il est difficile d'imaginer que de telles méthodes soient capables de générer des systèmes non réguliers comme les unités de calcul mais il est envisageable de les appliquer à la fabrication de systèmes réguliers tels que les mémoires.

La fabrication atome par atome est possible en appliquant les principes de la microscopie en champ proche mais elle n'est pas applicable pour la réalisation de systèmes complexes. Il est donc nécessaire d'imaginer des méthodes collectives conduisant à la création de structures régulières de nano-objets. Ces méthodes sont très nombreuses mais peuvent se répartir en deux grandes familles. La première s'appuie sur les propriétés cristallographiques du support et la seconde intègre différentes techniques issues de la synthèse chimique.

Il est intéressant d'envisager les méthodes de croissance de nano-objets sur des surfaces pré-structurées. Prenons l'exemple de la surface obtenue par section d'un cristal. La surface de coupe peut, si la coupe ne se fait pas dans la direction d'un plan cristallin, présenter une succession de marches à l'échelle nanométrique. Ce support structuré peut alors être la base pour faire croître des nanofils le long des marches.

Trois grandes techniques sont mises en œuvre pour former les surfaces pré-structurées :

- profiter des propriétés intrinsèques des surfaces (réseaux de défauts, réseau de marches, …),
- structurer la surface par des techniques de gravure particulières (géométrie de la surface ou création de réseaux de dislocations),
- coupler les deux méthodes précédentes.

Les méthodes de synthèse chimique s'inspirent des assemblages moléculaires qui ont conduit aux organismes vivants. Les assemblages peuvent alors se faire en volume ou en surface. Ces méthodes permettent en particulier de créer des réseaux de nano particules magnétiques mais aussi des monocouches organisées sur des surfaces.

12. Composants quantiques

Lorsque les dimensions de la zone active d'un composant atteignent l'échelle nanométrique, la réduction de dimensions devient une réduction de dimensionalité et la représentation corpusculaire de l'électron doit céder le pas à la représentation ondulatoire. Le confinement quantique résultant de la réduction de dimensionalité du mouvement de l'électron, redistribue la structure de bandes et les densités d'états (Chap. 10). Ce sont alors les effets quantiques spécifiques de ces nouvelles dimensionalités qui sont mis à profit dans les composants que nous qualifierons de *quantiques*.

12.1. Transport parallèle dans les structures quantiques
12.2. Transport perpendiculaire dans les structures quantiques
12.3. Lasers à puits quantiques
12.3. Blocage de Coulomb et systèmes a peu d'électrons
12.4. Nanofils et nanotubes

12.1. Transport parallèle dans les structures quantiques

12.1.1. *Spécificités*

Dans certains cas, on peut encore utiliser une approche *semi-classique* de type corpusculaire, notamment quand le transport s'effectue à l'intérieur d'une bande permise. C'est par exemple le cas pour le TEGFET dont nous avons décrit le fonctionnement de manière semi-classique avec l'introduction d'une faible dose de nouveaux concepts comme la *densité d'états bidimensionnelle*.

Lorsque les phénomènes de transport mettent en jeu des transferts à travers des gaps d'énergie interdite, les modèles semi-classiques atteignent leur limite, et on doit faire appel à des concepts radicalement différents. Le transport dans ces structures peut alors être décrit semi-quantitativement en utilisant des concepts relativement simples de la mécanique quantique, tel que l 'effet tunnel.

Enfin lorsque la taille et la qualité des composants le permettent, certains phénomènes spécifiques de la nature ondulatoire des électrons peuvent être exploités. Dans un composant macroscopique les processus individuels de diffusion des porteurs sont aléatoires et non corrélés. Dans ce cas, le traitement ondulatoire du transport consiste à sommer les carrés des amplitudes des fonctions d'onde individuelles pour obtenir l'intensité du signal en un point de la structure. Lorsque les dimensions du composant deviennent très faibles, l'information de phase de la fonction d'onde peut être conservée tout le long de la structure. Dans ce cas l'intensité du signal en un point de la structure s'obtient en élevant au carré la somme des amplitudes et non en sommant leurs carrés. Les résultats sont totalement différents car les fonctions d'onde électroniques produisent alors les mêmes sortes de phénomènes, interférences ou diffraction, que les fonctions d'ondes optiques. Comme pour les photons, le paramètre qui conditionne l'existence de ces phénomènes est la *longueur de cohérence de phase* qui correspond sensiblement au libre parcours moyen de l'électron-particule. L'observation du phénomène dépend donc des qualités du matériau et des interfaces, mais aussi de la longueur du canal actif du composant. Les techniques modernes d'épitaxie et de lithographie permettent de réaliser des structures dont les qualités et les dimensions permettent aux fonctions d'onde électroniques de conserver une certaine cohérence de phase d'un contact à l'autre. Ces structures sont qualifiées de *mésoscopiques.*

Dans les structures mésoscopiques, la cohérence de phase étant conservée au cours du transport, les moyennes macroscopiques utilisées précédemment pour définir la mobilité ou la conductivité n'ont plus de sens. Des effets spectaculaires, spécifiques de cette cohérence peuvent se manifester sur la conductivité par exemple. Dans les phénomènes d'interférences optiques, si la longueur d'onde du rayonnement est modulée, l'intensité en un point de l'espace est modifiée. Ce concept peut être étendu aux interférences électroniques dans un système

mésoscopique. Prenons l'exemple d'une structure bidimensionnelle: la longueur d'onde électronique au niveau de Fermi est donnée à partir des expressions (10-2 et 24) par $\lambda_F \approx \sqrt{2\pi / n_s}$ où n_s est la densité superficielle de charges. Dans un TEGFET, cette densité n_s peut être modulée par le potentiel de grille. On peut donc concevoir une nouvelle génération de composants rapides, tels que des portes logiques, basés sur la modulation des phénomènes d'interférence par la tension grille via n_s et λ_F. Contrairement aux interrupteurs FET classiques, qui nécessitent une variation importante de n_s dans le canal, un tel composant ne requiert qu'une faible variation de n_s car une variation même faible de λ_F entraîne une variation drastique du phénomène d'interférence.

Certes l'exploitation des phénomènes mésoscopiques est encore un peu futuriste, mais la vitesse des progrès technologiques, notamment dans les domaines de l'hétéroépitaxie et de la lithographie, permet de penser que l'échéance est proche. La réalisation d'hétérostructures à l'échelle nanométrique a longtemps souffert des températures de fabrication qui, trop élevées, entraînaient l'interdiffusion des espèces chimiques, interdisant ainsi la réalisation d'interfaces abruptes. Les techniques modernes d'hétéroépitaxie basse température, comme l'épitaxie par jet moléculaire EJM ou MBE (Molecular Beam Epitaxy) et le dépôt en phase vapeur par pyrolyse d'organométalliques MOCVD (Metal-Organic-Chemical-Vapor-Deposition), permettent une véritable *ingénierie de bandes* par la juxtaposition de matériaux binaires, ternaires ou même quaternaires. Ces techniques permettent à la fois de maîtriser de faibles épaisseurs de couches grâce à un faible taux de croissance, et de réaliser des interfaces abruptes grâce à un faible taux d'interdiffusion lié aux basses températures de croissance. Il est donc possible aujourd'hui de réaliser des hétérostructures de haute qualité à l'échelle de la monocouche atomique, avec des profils de potentiel et/ou de dopage prédéterminés, et d'accéder ainsi au régime mésoscopique.

Nous avons vu comment le confinement quantique des porteurs dans une hétérostructure modifiait à la fois la distribution énergétique, la fonction d'onde, et la densité d'états des porteurs (Chap. 10). Le mouvement des porteurs étant considérablement modifié, le transport électrique résultant de l'action d'un champ électrique, est très anisotrope. Lorsque la direction du champ électrique correspond à un degré de liberté des porteurs, on qualifie le transport de *parallèle.* C'est le cas dans une hétérostructure bidimensionnelle, comme une hétérojonction, un puits quantique ou un superréseau, lorsque le champ électrique est dans le plan des couches. C'est aussi le cas dans une structure unidimensionnelle comme un fil quantique, lorsque le champ électrique est dirigé suivant l'axe du fil. Lorsque le champ électrique est dirigé suivant l'axe, ou dans le plan, de quantification de la structure, le transport est qualifié de *perpendiculaire.* C'est aussi le cas de la conduction dans le canal d'un transistor MOS.

Conceptuellement le transport parallèle est peu affecté par le confinement quantique. Les méthodes semi-classiques d'approche restent opérationnelles moyennant quelques aménagements liés essentiellement à la densité d'états et aux

phénomènes de diffusion. La plus importante application du transport parallèle est le transistor TEGFET. Le transport perpendiculaire, par contre, nécessite une approche totalement originale. Des concepts purement quantiques sont nécessaires à l'interprétation des effets physiques spécifiques de ce genre de transport et à la modélisation des composants qui exploitent ces effets.

Dans le transport parallèle, les processus dont les modifications par le confinement quantique affectent le plus les propriétés de transport, sont sans conteste les processus de diffusion.

Ces processus de diffusion sont tout d'abord affectés par la nature de la fonction d'onde électronique (Eq. 10-8). Les intégrales permettant de calculer les éléments de la matrice de diffusion sont évidemment fonction du caractère très anisotrope du mouvement des porteurs.

La redistribution en sous-bandes du diagramme énergétique est à l'origine d'un autre effet important sur la diffusion des porteurs. En effet, à faible énergie, c'est-à-dire à faible champ électrique et basse température, les électrons d'une sous-bande ne peuvent diffuser qu'à l'intérieur de celle-ci. Par contre lorsque l'électron acquiert des énergies plus importantes il peut diffuser d'une sous-bande dans une autre. Dans le premier cas, la diffusion est de type *intra-sous-bande*, dans le second elle est de type *inter-sous-bande*. Les densités d'états, bidimensionnelle ou unidimensionnelle, étant très différentes de la densité d'états tridimensionnelle, les processus de diffusion intra-sous-bande sont très différents dans les systèmes à dimensionalité réduite et dans les matériaux massifs . Par contre, la densité d'états α-dimensionnelle (α<3) intégrée sur plusieurs sous-bandes tend vers la densité d'états tridimensionnelle. Ainsi à fort champ, la diffusion étant de type inter-sous-bande, le processus de diffusion seront assez peu différents de leurs homologues à 3-dimensions.

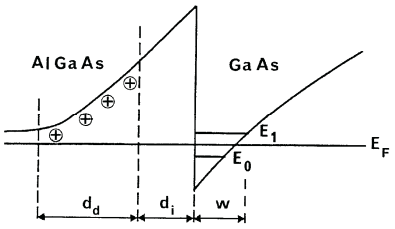


Figure 12-1 : Couche tampon

Une autre contribution importante aux processus de diffusion est la présence d'impuretés ionisées. Or dans les structures à dimensionalité réduite, il est possible de séparer spatialement les électrons des ions donneurs. En effet, dans la structure à dopage modulé, les atomes donneurs sont distribués dans les barrières alors que les électrons sont thermalisés dans les puits de potentiel. Cette séparation spatiale des électrons et des ions réduit considérablement ce type de diffusion. Cette réduction peut encore être optimisée par la réalisation d'une couche tampon non dopée qui éloigne encore davantage les ions donneurs des électrons (Fig. 12-1). Une couche tampon d'épaisseur $d_i \approx 100$ Å permet d'éliminer presque totalement la diffusion par les ions donneurs.

Un autre paramètre spécifique des hétérostructures, et inexistant dans les matériaux massifs, est la rugosité des interfaces. Dans les hétérostructures réelles les interfaces entre les différents matériaux ne sont pas parfaites mais présentent certaines rugosités, dont l'amplitude est de l'ordre de la monocouche atomique, distribuées aléatoirement sur l'interface. Ces rugosités peuvent jouer un rôle sur les phénomènes de diffusion si les porteurs se déplacent près des interfaces. En fait, en raison du potentiel image (Par. 10.1.2.c), les électrons présents dans le puits de potentiel d'une hétérojonction sont éloignés de l'interface. Les fonctions d'onde électroniques présentent un maximum localisé à quelques dizaines d'Angströms de l'interface (Fig. 10-15). Ce qui réduit considérablement les effets des rugosités.

Une autre source de diffusion est liée aux défauts d'alliage. Si les matériaux constituant le puits, et/ou les barrières, sont des alliages ternaires ou quaternaires, les fluctuations de composition constituent autant de centres diffuseurs.

Enfin comme dans un matériau massif, les phonons jouent un rôle essentiel dans les processus de diffusion quand la température augmente. Comme celui des électrons, le spectre des phonons est modifié par la dimensionalité réduite de l'hétérostructure. Certains modes sont étalés dans toute la structure alors que d'autres sont localisés aux interfaces. En fait, tant que les puits de potentiel ont une largeur supérieure à une cinquantaine d'Angströms, les processus de diffusion par les phonons spécifiques de la structure sont peu différents des processus associés aux phonons de volume.

Notons enfin que dans une structure à fil quantique, le mouvement des électrons est unidimensionnel. Il en résulte que la diffusion intra-sous-bande des électrons est sévèrement réduite. En effet, lors d'un processus de diffusion élastique, un électron dans un état k ne peut diffuser que vers un état $-k$. Des mobilités considérables, de l'ordre de 10^7 cm^2/Vs, peuvent alors être espérées.

12.1.2. *Mobilité*

Dans le transport parallèle à l'intérieur d'une hétérostructure, l'amélioration des performances par rapport au matériau massif est incontestablement liée à la mobilité. Cette mobilité est calculée en moyennant le temps de relaxation du moment des électrons, sur l'ensemble des différents processus de diffusion.

A faible champ électrique, les diffusions inter-sous-bandes sont négligeables et parmi tous les processus de diffusion, trois d'entre eux jouent un rôle majeur avec des importances relatives différentes suivant la température.

A basse température le processus le plus actif est la diffusion par les impuretés ionisées. Or dans une structure à dopage modulé, où les ions donneurs sont distribués dans les barrières et isolés du puits par une couche tampon non dopée, cet effet peut être minimisé. Il en résulte que si en outre la diffusion par les rugosités d'interface est négligeable, la mobilité d'un gaz d'électrons bidimensionnel peut approcher ses limites théoriques. Des mobilités supérieures à 10^6 cm^2/Vs ont été obtenues à basse température dans des hétérostructures de type AlGaAs/GaAs. Dans un matériau massif la diffusion par les impuretés ionisées entraîne, à basse température, une diminution de la mobilité avec la température. Dans les structures à dopage modulé la séparation spatiale des électrons et des ions évite cette

diminution et la mobilité tend vers une valeur de saturation. L'expression qui rend compte de la variation expérimentale de mobilité associée à la diffusion par les impuretés ionisées éloignées de l'interface, est de la forme (Lee K.-1983).

$$\mu_I = \frac{64\pi\hbar^3 \varepsilon S_o^2 (2\pi n_s)^{3/2}}{e^3 m_e} \left(\frac{1}{(d_i + z_w)^2} - \frac{1}{(d_i + z_w + d_d)^2} \right)^{-1} \qquad (12\text{-}1)$$

où d_i est la largeur de la couche tampon non dopée, d_d la largeur de la zone de charge d'espace de la barrière et z_w la largeur effective du puits de potentiel. S_o est l'inverse de la longueur d'écran bidimensionnelle due aux n_s porteurs libres du puits. Dans les limites dégénérées et non dégénérées cette quantité est donnée respectivement par

$$S_o = e^2 m_e / 2\pi\varepsilon\hbar^2 \qquad (12\text{-}2\text{-a})$$

$$S_o = e^2 n_s / 2\,\varepsilon kT \qquad (12\text{-}2\text{-b})$$

Quand la température augmente la mobilité dépend de la diffusion par les phonons acoustiques à travers le potentiel associé à la déformation du réseau et, pour les matériaux non centro-symétriques, du champ piézo-électrique. L'effet piézo-électrique ne joue un rôle significatif sur la diffusion des électrons que dans le domaine des faibles températures. Dans les températures intermédiaires il joue un rôle négligeable. Une expression simple donnant la variation de la mobilité d'un gaz d'électrons bidimensionnel, associée à la diffusion par les phonons acoustiques, est de la forme (Arora V.K.-1985).

$$\mu_{ac} = \frac{2e\hbar^3 \rho u^2 d}{3 m_e^2 E_1^2 kT} \qquad (12\text{-}3)$$

où ρ est la densité du matériau, u la vitesse du son dans le plan de la structure, E_1 le potentiel de déformation, et d la largeur effective du puits. Dans le puits triangulaire d'une hétérojonction la largeur effective d est directement reliée au paramètre variationnel b caractérisant l'extension spatiale de la fonction d'onde électronique dans le semiconducteur (Eq. 10-96). En prenant pour le potentiel de déformation dans GaAs la valeur E_1=7 eV communément admise, la mobilité associée à la diffusion par les phonons acoustiques s'écrit

$$\mu_{ac} \approx 3.10^6 \frac{d(nm)}{T} \quad cm^2/Vs \qquad (12\text{-}4)$$

A T=100 K, la composante de mobilité associée à la diffusion par les phonons acoustiques, est de l'ordre de 3.10^5 cm²/Vs pour les électrons bidimensionnels d'un puits de 100 Å de large.

Enfin à haute température la mobilité des porteurs est limitée par la diffusion par les phonons optiques. Dans les hétérostructures réalisées avec des

matériaux polaires, c'est-à-dire autres que les éléments de la colonne IV comme le silicium, la diffusion par les phonons optiques non polaires est négligeable devant celle due aux phonons optiques polaires. Pour des températures supérieures à 100 K, la mobilité associée à la diffusion par les phonons optiques polaires est de la forme (Ridley B.K.-1982 ; Arora V.K.-1985)

$$\mu_{op} = \frac{4\pi\varepsilon_p \hbar^2}{e\Omega m_e^2 d}\left(e^{\hbar\Omega/kT} - 1\right) \qquad (12\text{-}5)$$

avec $\varepsilon_p^{-1} = \varepsilon_\infty^{-1} - \varepsilon_s^{-1}$, où ε_∞ est la constante diélectrique du semiconducteur à haute fréquence et ε_s sa constante diélectrique statique. d est la largeur du puits et $\hbar\Omega$ l'énergie des phonons optiques. Pour T=300 K et d=100 Å, la mobilité du gaz d'électrons bidimensionnel d'un puits de GaAs, associée à la diffusion par les phonons optiques, est de l'ordre de 10^5 cm^2/Vs.

L'allure de variation de la mobilité d'un gaz d'électrons bidimensionnel est représentée sur la figure (12-2) avec ses trois principales composantes.

A fort champ électrique, les électrons acquièrent une énergie suffisante pour sauter d'une sous-bande dans une autre de sorte que la diffusion inter-sous-bande joue dans ce domaine de champ un rôle non négligeable sur la mobilité. Le comportement des électrons est alors comparable à celui obtenu dans un matériau massif, avec un régime linéaire à faible champ, suivi d'un régime de survitesse et enfin d'un régime de saturation. Comparativement au cas du matériau massif, la mobilité à faible champ est plus importante.

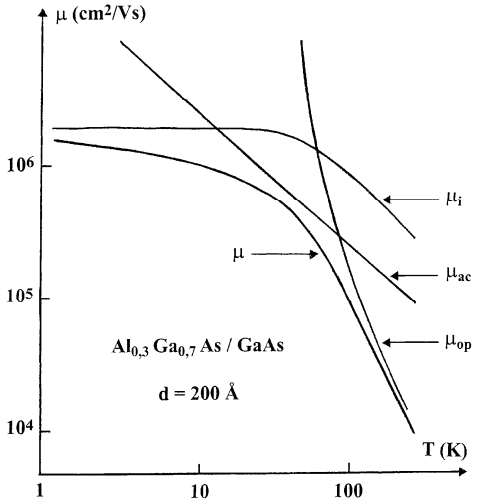


Figure 12-2 : Principales contributions à la mobilité d'un gaz d'électrons bidimentionnel dans une hétérojonction de type AlGaAs/GaAs. μ_i : diffusion par les impuretés ionisées. μ_{ac} : diffusion par les phonons acoustiques. μ_{op} : diffusion par les phonons optiques. *(Walukiewiez W. – 1984)*

12.1.3. *Effet Gunn bidimensionnel*

Le transfert électronique inter-vallée entraînait l'apparition d'une région à pente négative dans la relation vitesse-champ électrique. Ceci se produit lorsque les électrons de faible masse effective et grande mobilité de la vallée Γ de la bande de conduction, sont transférés, sous l'action du champ électrique, dans les vallées satellites L où leur masse effective est plus grande et leur mobilité plus faible.

Cet effet peut être reproduit avantageusement dans le transport parallèle le long d'une hétérojonction de type AlGaAs/GaAs.

A faible champ électrique les électrons du puits quantique de GaAs sont thermalisés dans la première sous-bande d'énergie. La diffusion inter-sous-bande est négligeable, la mobilité des électrons est constante, leur vitesse augmente linéairement avec le champ électrique, la loi $I(V)$ est ohmique. Quand le champ électrique augmente, la population électronique de la première sous-bande diminue, d'abord au profit des sous-bandes supérieures de la vallée Γ, ensuite par transfert dans les sous-bandes des vallées satellites L et X, où leur mobilité est plus faible. Dans le GaAs massif, les vallées satellites qui participent au transfert sont uniquement les vallées L, les vallées X étant situées à plus haute énergie. Dans le puits quantique AlGaAs/GaAs les deux types de vallées participent au transfert car le confinement quantique rapproche les sous-bandes L des sous-bandes X. En effet, le puits de potentiel AlGaAs(L)/GaAs(L) présente une barrière plus haute que le puits de potentiel AlGaAs(X)/GaAs(X). En outre les couches épitaxiales étant des plans [001], la masse effective électronique mise en jeu dans la quantification des états de symétrie X est, pour la première sous-bande, la masse longitudinale m_{Xl}, alors que pour les états de symétrie L la masse effective concernée est la masse moyenne $m_L^{-1} = (m_{Ll}^{-1} + 2m_{Lt}^{-1})/3$ (Eq. 1-192). m_L étant très inférieure à m_{Xl}, l'énergie de confinement est plus importante dans les vallées L que dans les vallées X.

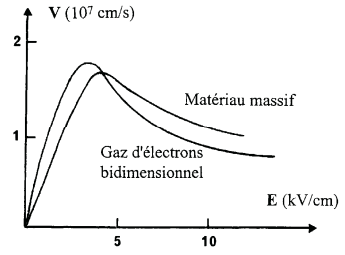


Figure 12-3 : Lois de variation v(E) dans un gaz d'électrons bidimen-sionnel et dans un matériau massif *(Yokoyama K. - 1986)*

Le calcul dans l'hétérostructure, du transport longitudinal à fort champ, n'est pas trivial. Il concerne un transport multi-sous-bande, multi-vallée. En outre il doit intégrer tous les phénomènes de diffusion et notamment intra- et inter-sous-bande dans la vallée Γ, et inter-vallées Γ-X-L, et ceci compte tenu des différents confinements quantiques. Il n'existe pas de solution analytique réaliste. Une simulation numérique, utilisant la méthode Monte-Carlo, permet d'obtenir la loi de variation de la vitesse $v(E)$, représentée sur la figure (12-3) (Yokoyama K.-1986). Elle est comparée au résultat d'un calcul

tridimensionnel dans le GaAs massif. La pente à faible champ est plus importante et le champ de seuil E_m, correspondant au maximum de survitesse, est plus faible. A fort champ les lois de variation sont pratiquement identiques.
L'augmentation de la pente à faible champ et la diminution du champ de seuil E_m résultent essentiellement de la plus grande mobilité des électrons à faible champ. Cette grande mobilité, spécifique du transport bidimensionnel, permet d'atteindre plus tôt le régime d'électrons chauds sous l'action du champ électrique. Les effets non-linéaires apparaissent typiquement pour des champs électriques inférieurs à 1 V/cm.

En outre, en raison des différences des masses effectives, les énergies de quantification des états L et X sont beaucoup plus faibles que celle des états Γ. Les distances énergétiques Γ-L et Γ-X sont réduites d'autant.

Enfin, la similitude de comportement à fort champ, des gaz d'électrons bi- et tridimensionnels résulte simplement du fait que la plupart des électrons sont alors transférés dans les vallées satellites L et X. La quantification de ces vallées étant très compacte, en raison des masses effectives, le comportement de leurs électrons est quasi-tridimensionnel.

12.1.4. *Transfert dans l'espace réel - RST*

Dans l'effet Gunn à trois ou deux dimensions, la résistance différentielle négative résulte d'une propriété intrinsèque du semiconducteur et correspond à un transfert électronique dans l'espace des moments, c'est l'effet RWH. Dans une hétérostructure, un effet comparable peut être obtenu par un transfert électronique dans l'espace réel, RST (Real Space Transfer).

Considérons l'hétérostructure de la figure (12-4) constituée d'une série de puits quantiques de GaAs de largeur d_p, séparés par des barrières de AlGaAs de hauteur ΔE_c et d'épaisseur d_b. Les barrières sont dopées avec des donneurs et les électrons sont thermalisés dans les puits de GaAs. Le transport parallèle des électrons est conditionné par le champ électrique longitudinal E. A faible champ les électrons restent thermalisés dans la première sous-bande de chaque puits de GaAs où leur mobilité est constante. La loi $I(V)$ est ohmique. Quand le champ électrique augmente, l'énergie cinétique des électrons augmente jusqu'à atteindre et dépasser la hauteur ΔE_c des barrières de potentiel. Ces électrons chauds se propagent alors longitudinalement dans les puits de GaAs avec une énergie supérieure au minimum de la bande de conduction de AlGaAs. Il suffit par conséquent qu'un processus de diffusion relaxe leur moment à 90° pour que certains de ces électrons soient transférés dans les barrières. Leur énergie cinétique est alors transformée en énergie potentielle. Le champ longitudinal réaccélère ces électrons proportionnellement à leur nouvelle mobilité qui, dans les barrières de AlGaAs, est très inférieure à celle qu'ils avaient dans les puits de GaAs. Un équilibre dynamique s'établit entre les populations des puits et des barrières et la mobilité moyenne de la population électronique s'écrit

$$\bar{\mu} = \frac{n_{sp}\mu_p + n_{sb}\mu_b}{n_{sp} + n_{sb}} \tag{12-6}$$

où n_{sp} et n_{sb} sont respectivement les densités d'électrons par unité de surface dans les puits et les barrières. Ces densités surfaciques sont reliées aux densités volumiques n_p et n_b par les relations $n_{sp}=n_p d_p$, $n_{sb}=n_b d_b$.

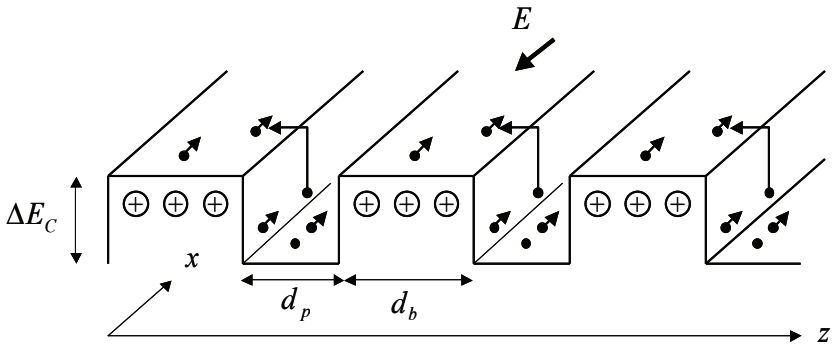


Figure 12-4 : Transfert dans une hétérostructure

Sous l'action du champ électrique, les électrons sont ainsi transférés d'une région à grande mobilité vers une région à faible mobilité. Ce transfert dans l'espace réel produit sur la relation vitesse-champ électrique, les mêmes effets que le transfert $\Gamma \rightarrow L$ dans l'espace des moments. On obtient, comme dans l'effet Gunn, un régime à *résistance différentielle négative*. Si la hauteur de barrière ΔE_c est inférieure à la différence d'énergie inter-vallée $\Delta_{\Gamma L}$, le champ de seuil E_{mr} du transfert dans l'espace réel est inférieur au champ de seuil E_{mp} du transfert $\Gamma \rightarrow L$ dans l'espace des moments.

L'avantage de l'effet RTS sur l'effet RWH est que les principaux paramètres peuvent être modulés. Le champ de seuil en particulier, peut être contrôlé par la hauteur de barrière ΔE_c , c'est-à-dire par la composition de l'alliage. Le rapport pic/vallée peut être contrôlé par les rapports des largeurs de puits et de barrière d_p/d_b et des mobilités μ_p/μ_b. La mobilité μ_b des électrons dans les barrières est très inférieure à celle des électrons dans le puits en raison des masses effectives plus importantes, des dopages des barrières et des effets d'alliage. La mobilité, à faible champ, des électrons dans les puits de GaAs est optimisée par un dopage modulé, des mobilités supérieures à 5000 cm^2/Vs peuvent être obtenues à la température ambiante. Dans les barrières, la mobilité des électrons peut être contrôlée par un dopage compensé. La densité d'électrons est donnée par la différence N_d-N_a mais la mobilité est conditionnée par la somme N_d+N_a . Ainsi dans AlGaAs fortement compensé, la mobilité varie de 4000 à 50 cm^2/Vs quand le taux de dopage varie de 10^{17} à 10^{20} cm^{-3}.

a. Approche analytique

Le transfert électronique entre les puits et les barrières et la résistance différentielle négative qui en résulte, peuvent être modélisés sur la base des courants d'émission thermoélectronique transversaux. Les densités d'électrons dans les puits et les barrières sont obtenues en équilibrant ces courants. La distribution énergétique des électrons est supposée régie par la statistique de Boltzmann en caractérisant les électrons chauds dans le puits de GaAs par une température électronique T_e supérieure à celle du réseau. Les électrons dans les barrières peuvent être supposés à la température T du réseau.

Les densités volumiques d'électrons dans les puits et les barrières sont données par les lois de Boltzmann.

$$n_p = N_{cp}\, e^{-\frac{E_{cp}-E_{Fp}}{kT_e}} \tag{12-7-a}$$

$$n_b = N_{cb}\, e^{-\frac{E_{cb}-E_{Fb}}{kT}} \tag{12-7-b}$$

où N_{cp} et N_{cb} représentent les densités équivalentes d'états supposées tridimensionnelles. T_e est la température électronique dans le puits, supérieure à la température T du réseau.

Les courants thermoélectroniques puits→barrières et barrières→puits sont donnés par la loi de Richardson (chapitre 1)

$$j_{p\to b} = A^* T^2\, e^{-(E_{cp}-E_{Fp}+\Delta E_c)/kT_e} \tag{12-8-a}$$

$$j_{b\to p} = A^* T^2\, {}^{-(E_{cb}-E_{Fb})/kT} \tag{12-8-b}$$

où $A^* = 4\pi e m_e k^2/h^3$ est la constante de Richardson. La masse effective m_{ep} des électrons chauds dans le puits est comparable à la masse effective m_{eb} des électrons thermalisés dans les barrières.

- *Régime stationnaire*

En régime stationnaire ces deux courants sont égaux. En écrivant $j_{p\to b} = j_{b\to p}$ on obtient la relation

$$e^{-(E_{cp}-E_{Fp})/kT_e}\, e^{-\Delta E_c/kT_e} = e^{-(E_{cb}-E_{Fb})/kT} \tag{12-9}$$

Les relations (12-7-a,b) et (12-9) permettent d'établir le rapport des densités de population des puits et des barrières

$$\frac{n_b}{n_p} = \frac{N_{cb}}{N_{cp}}\, e^{-\Delta E_c/kT_e} \tag{12-10}$$

Les masses effectives des électrons chauds des puits et des électrons thermalisés des barrières étant comparables cette expression s'écrit

$$\frac{n_b}{n_p} = e^{-\Delta E_c / kT_e} \tag{12-11}$$

A faible champ électrique T_e=T et kT est très inférieur à ΔE_c de sorte que n_b≈0 et n_p≈n. Quand le champ électrique devient important, la température électronique devient importante et le rapport des populations augmente. Le rapport des densités surfaciques dans les puits et les barrières est par conséquent donné par

$$\frac{n_{sb}}{n_{sp}} = \frac{d_b}{d_p} e^{-\Delta E_c / kT_e} \tag{12-12}$$

Pour calculer la température T_e des électrons chauds dans le puits, nous supposerons la structure à T=100 K. La diffusion est alors conditionnée essentiellement par les phonons acoustiques et la température électronique peut être reliée au champ électrique par la relation (Van der Ziel A.-1976)

$$T_e = 0{,}927 \frac{E}{E_o} T \tag{12-13}$$

où T est la température du réseau et E_o une valeur critique du champ donnée par

$$E_o = 1{,}514 \frac{u}{\mu_p} \tag{12-14}$$

où μ_p est la mobilité des électrons du puits à faible champ et u la vitesse du son dans le matériau.

La mobilité moyenne des électrons est donnée par l'expression (12-6) qui s'écrit compte tenu de l'expression (12-12)

$$\overline{\mu} = \mu_p \frac{1 + \mu_b d_b / \mu_p d_p . e^{-\Delta E_c / kT_e}}{1 + d_b / d_p . e^{-\Delta E_c / kT_e}} \tag{12-15}$$

La vitesse moyenne des électrons est donnée par $v = \overline{\mu} E$.

La figure (12-5) représente la loi de variation de cette vitesse avec le champ électrique pour différentes valeurs du rapport d_b/d_p, pour T=100 K, μ_p=8000 cm²/Vs, μ_b=50 cm²/Vs et u=5000 cm/s. Il faut noter que les densités d'états dans les puits et les barrières étant comparables, l'apparition d'une région à pente négative nécessite la réalisation d'un rapport d_b/d_p relativement important.

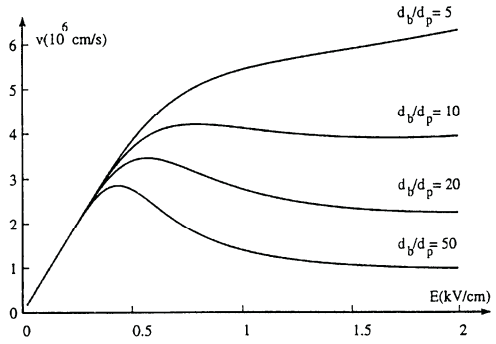


Figure 12-5 : Vitesse en fonction du champ

- *Régime dynamique*

La dynamique du transfert électronique à une interface est régie par l'équation de continuité (Eq. 1-239-a) qui s'écrit ici en négligeant les générations et les recombinaisons

$$\frac{\partial n}{\partial t} = \frac{1}{e}\frac{\partial j_n}{\partial z} \tag{12-16}$$

où z est la direction transversale de la structure. Le courant d'émission thermoélectronique, transversal, s'établit sur une certaine longueur comparable au libre parcours moyen des électrons. On peut ainsi écrire le deuxième terme de l'expression (12-16) sous la forme approchée

$$\frac{\partial j_n}{\partial z} \approx \frac{J_{z_o+l} - j_{z_o}}{l} \approx -\frac{j_{z_o}}{l} \tag{12-17}$$

où z_o est l'abscisse de l'interface considérée, dans la direction transversale de la structure, et l le libre parcours moyen des électrons dans le matériau où ils sont injectés.

Considérons tout d'abord *l'établissement* du régime de pente négative. Il résulte des transferts puits→barrières , le courant j_{zo} est alors le courant $j_{p\to b}$ donné par l'expression (12-8-a). L'expression (12-16) s'écrit alors

$$\frac{\partial n_p}{\partial t} = -\frac{j_{p\to b}}{e l_b} = -\frac{A^* T^2}{e l_b} e^{-\left(E_{cp} - E_{Fp}\right)/kT_e} e^{-\Delta E_c / kT_e} \tag{12-18}$$

Compte tenu de l'expression (12-7-a), cette relation s'écrit

$$\frac{\partial n_p}{\partial t}=-\frac{A^*T^2 n_p}{eN_{cp}l_b}e^{-\Delta E_c/kT_e} \tag{12-19}$$

Soit

$$\frac{\partial n_p}{\partial t}=-\frac{n_p}{\tau_e} \tag{12-20}$$

avec

$$\tau_e=\frac{eN_{cp}l_b}{A^*T^2}e^{\Delta E_c/kT_e} \tag{12-21}$$

En explicitant la constante de Richardson et la densité équivalente d'états (Eq. 1-136-a), la constante de temps τ_e s'écrit

$$\tau_e=\tau_o e^{\Delta E_c/kT_e} \tag{12-22}$$

avec

$$\tau_o=\left(\frac{2\pi m_{ep}l_b^2}{kT}\right)^{1/2} \tag{12-23}$$

L'établissement du régime de pente négative est régi par les lois de variation des densités n_p et n_b d'électrons dans les puits et les barrières. Avec $n_p(t=0)=n$ l'intégration de l'équation (12-20) donne la loi de variation de n_p. Parallèlement, la loi de variation de n_b est donnée par $n_b=n-n_p$, soit

$$n_p=ne^{-t/\tau_e} \tag{12-24}$$

$$n_b=n\left(1-e^{-t/\tau_e}\right) \tag{12-25}$$

Considérons maintenant le *retour à l'équilibre* du système excité. Ce retour résulte du transfert électronique barrières→puits, le courant j_{zo} de l'expression (12-17) est alors le courant $j_{b\to p}$ donné par l'expression (12-8-b). L'expression (12-16) s'écrit alors

$$\frac{\partial n_b}{\partial t}=-\frac{j_{b-p}}{el_p}=-\frac{A^*T^2}{el_p}e^{-(E_{cb}-E_{Fb})/kT} \tag{12-26}$$

ou, compte tenu de l'expression (12-7-b)

$$\frac{\partial n_b}{\partial t}=-\frac{n_b}{\tau_r} \tag{12-27}$$

avec

$$\tau_r = \frac{eN_{cb}l_p}{A^*T^2} = \left(\frac{2\pi m_{eb}l_p^2}{kT}\right)^{1/2} \tag{12-28}$$

Le retour des électrons, depuis les barrières vers les puits est régi par la loi

$$n_b = n_{bo}e^{-t/\tau_r} \tag{12-29}$$

Ainsi dans la mesure où l'on suppose comparables les masses effectives et les libres parcours moyens des électrons dans les puits et les barrières, les temps d'*établissement* du régime de pente négative et du *retour à l'équilibre* sont donnés par

$$\tau_e = \tau_o e^{\Delta E_c / kT_e} \tag{12-30-a}$$

$$\tau_r = \tau_o \tag{12-30-b}$$

avec

$$\tau_o \approx \left(\frac{2\pi m_e l^2}{kT}\right)^{1/2} \tag{12-31}$$

Avec m_e=0,07 m_o et l=1000 Å, on obtient à la température ambiante τ_o=1 ps. Le temps de retour à l'équilibre τ_r est donc très bref. Par contre le temps d'établissement du régime de pente négative τ_e est une fonction exponentielle de la hauteur de la barrière et de la température électronique. Lorsque kT_e est comparable à ΔE_c , τ_e est comparable à τ_r.

b. Simulation numérique

Le modèle analytique, basé sur le calcul des taux d'occupation des puits et des barrières par l'équilibre des courants d'émission thermoélectronique transversaux, permet une approche relativement simple des effets qualitatifs du transfert électronique mais comporte trop d'hypothèses simplificatrices pour traduire ces effets quantitativement.

En premier lieu, nous avons supposé la distribution énergétique des électrons régie par une loi de Boltzmann associée à une température électronique T_e. Ceci est justifié dans les barrières avec $T_e \approx T$, mais pas dans les puits lorsque le champ électrique devient intense. En effet, quand l'énergie cinétique des électrons devient importante leurs interactions mutuelles diminuent et il n'existe plus de régime de pseudo-équilibre. Il n'est alors plus possible de définir une température électronique, la distribution n'est plus régie par une loi thermodynamique.

En outre nous avons traité indépendamment les régions de puits et de barrière, en fait il existe un échange permanent d'électrons chauds entre les deux types de région.

Nous avons supposé constantes les mobilités des électrons, μ_p dans les puits et μ_b dans les barrières. Or, en régime d'électrons chauds, il existe un transfert électronique non seulement dans l'espace réel mais aussi, en particulier dans les puits de GaAs, dans l'espace des moments (effet RWH). Une loi de variation réaliste $\mu_p(E)$ doit donc être utilisée.

Nous avons totalement ignoré la présence éventuelle d'un champ électrique transversal. Or ce champ transversal existe en raison du dopage modulé de la structure, les ions donneurs sont distribués dans les barrières alors que les électrons, au moins à faible champ longitudinal, sont distribués dans les puits. Il en résulte un champ transversal important, qui en outre varie avec la distribution électronique, c'est-à-dire avec le champ longitudinal. En effet, en raison du transfert électronique, qui réduit la charge d'espace, ce champ transversal diminue quand le champ longitudinal augmente. Ce champ électrique transversal exerce sur les électrons chauds une force qui favorise leur transfert dans les barrières et qui s'oppose à leur retour dans les puits. Le calcul de ce champ, qui est à la fois cause et effet de la distribution électronique, nécessite donc une approche auto-cohérente.

Nous avons calculé les distributions électroniques sur la base de densités d'états tridimensionnelles, ceci reste justifié dans les barrières mais dans les puits, lorsque la largeur de ces derniers devient inférieure à quelques dizaines de nanomètres, la quantification des états en sous-bandes d'énergie entraîne une densité d'états de type bidimensionnel.

Enfin, nous n'avons pris en compte qu'un seul processus de diffusion, en l'occurrence les phonons acoustiques, un traitement quantitatif doit évidemment inclure tous les processus de diffusion jouant un rôle significatif à la température considérée.

Une simulation numérique, par la méthode Monte-Carlo, prenant en considération les différents paramètres évoqués ci-dessus, donne à la température ambiante les résultats représentés sur la figure (12-6). La structure étudiée est constituée de puits de GaAs de largeur d_p=400 Å , séparés par des barrières de AlGaAs de hauteur ΔE_c=176 meV et de largeur d_b=4000 Å. Les mobilités à faible champ dans les puits et les barrières sont respectivement μ_p=8000 et μ_b=500 cm^2/Vs. La largeur relativement importante des puits autorise l'utilisation de densités d'états tridimensionnelles. Les courbes (p) et (b) représentent, en fonction du champ électrique longitudinal, des lois de variation de la vitesse des électrons dans les puits et les barrières respectivement. Dans les puits de GaAs (courbe p) la loi de variation présente un pic de survitesse avec un champ de seuil $E_{mp}\approx$5 kV/cm, correspondant à l'effet RWH, c'est-à-dire au transfert électronique $T\rightarrow L$ dans l'espace des moments. Dans les barrières de AlGaAs (courbe b) la loi de variation est linéaire jusqu'à un régime de saturation qui apparaît pour un champ $E_{sb}\approx$15 kV/cm. La courbe (t) représente la loi de variation de la vitesse moyenne de l'ensemble des électrons. Le seuil $E_{mr}\approx$3 kV/cm associé à l'effet RST précède le seuil E_{mp} associé à l'effet RWH, la survitesse est de l'ordre de 10^7 cm/s.

La figure (12-7) représente les taux d'occupation des puits et des barrières. La présence du champ transversal entraîne un taux d'occupation des barrières non nul à champ longitudinal nul. Pour E>6 kV/cm, la distribution des électrons ne varie pratiquement plus, 20 % environ d'entre eux demeurent dans les puits. Il faut

noter que le rapport des populations des barrières et des puits conditionne, avec le rapport des mobilités, le rapport pic/vallée de la caractéristique *I*(*V*) du composant.

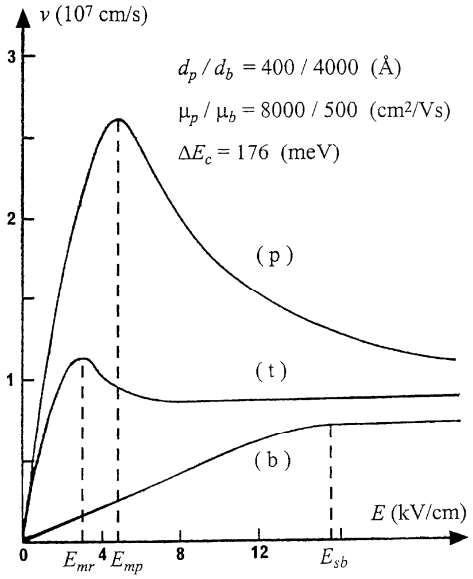


Figure 12-6 : Caractéristique vitesse-champ dans un dispositif RTS de type GaAlAs/GaAs. p) Puits de GaAs seuls. b) Barrières de GaAlAs seules. t) Dispositif total. ***(Littlejohn M.A. - 1983)***

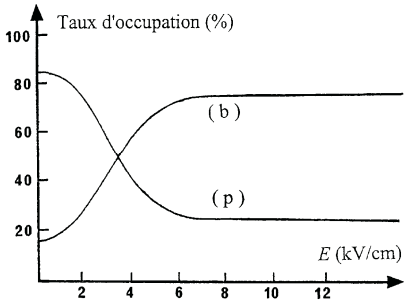


Figure 12-7 : Taux d'occupation des puits (p) et des barrières (b) en fonction du champ électrique (Littlejohn M.A. - 1983).

12.1.5. *Etalon de résistance*

Nous avons vu (Chap. 10) comment la réduction à deux dimensions du mouvement des électrons modifiait la fonction d'onde (Eq. 10-8), les états énergétiques (Eq. 10-11) et la densité d'états (Eq. 10-19) électroniques dans un *puits quantique*.

Si le mouvement de l'électron est réduit à une seule dimension, ce dernier est alors confiné dans un *fil quantique* et sa fonction d'onde s'écrit sous la forme

$$\psi_i(r)=\xi_i(x,y)e^{ik_z z}\varphi(z) \tag{12-32}$$

où $\varphi(z)$ est la fonction de Bloch et ξ_i (x,y) une fonction enveloppe solution de l'équation de Schrödinger dans l'approximation de la masse effective

$$\left(\frac{\hbar^2}{2m_e}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)+E_i-V(x,y)\right)\xi_i(x,y)=0 \tag{12-33}$$

Les énergies E_i sont les énergies de l'électron dans son mouvement dans le plan xy. Elles décrivent la quantification des états électroniques dans le plan perpendiculaire à l'axe du fil quantique. Dans la direction longitudinale z du fil, le mouvement de l'électron est libre. Il en résulte une structure de sous-bandes d'énergie avec une quantification discrète dans le plan xy et une variation pseudo-continue suivant z. En prenant l'origine des énergies au bas de la bande de conduction non confinée, l'énergie de l'électron s'écrit

$$E=E_i+\frac{\hbar^2 k_z^2}{2m_e} \tag{12-34}$$

Le mouvement de l'électron est donc limité à *une* dimension, la densité d'états dans l'espace des k par unité de longueur du fil est par conséquent donnée, compte tenu de la dégénérescence de spin, par (Eq. 1-11)

$$g(k_z)=\frac{2}{2\pi} \tag{12-35}$$

Ainsi sur la longueur de l'espace des k comprise entre $-k_z$ et $+k_z$ le nombre d'états est donné par

$$N=\frac{2}{2\pi}2k_z=\frac{2}{\pi}k_z \tag{12-36}$$

Soit en explicitant k_z en fonction de l'énergie (Eq. 12-34)

$$N=\frac{2}{\pi\hbar}\sqrt{2m_e(E-E_i)} \tag{12-37}$$

La densité d'états, dans la sous-bande *i*, par unité de longueur et unité d'énergie, est donnée par *dN/dE*, soit

$$g_i(E)=\frac{1}{\pi\hbar}\sqrt{\frac{2m_e}{E-E_i}} \tag{12-38}$$

C'est une spécificité d'un système à *une* dimension, la densité d'états présente un maximum et même une singularité, au bas de chaque sous-bande.

La population électronique de chaque sous-bande d'énergie est donnée par

$$n_i=\int_{E_i}^{\infty} g_i(E)f(E)dE \tag{12-39}$$

où $f(E)$ est la fonction de distribution de Fermi.

Supposons une température très faible, à la limite T=0, la fonction de Fermi est alors égale à 1 pour $E<E_F$ et zéro pour $E>E_F$, l'expression (12-39) s'écrit

$$n_i=\int_{E_i}^{E_F} g_i(E)dE \tag{12-40}$$

Soit

$$n_i=\frac{2}{\pi\hbar}\sqrt{2m_e(E_F-E_i)} \tag{12-41}$$

Si le fil est polarisé par une tension V, il est parcouru par un courant I résultant du déplacement des électrons.

Dans un système macroscopique à trois dimensions, les électrons sont distribués au minimum de la bande de conduction où ils ont tous la même vitesse. Le courant est donné par $I=\iint j_z\,ds \; avec \; j_z=-nev_z$ où v_z est la vitesse d'entraînement des électrons donnée par v_z=-μE_z.

Dans le fil quantique, les états électroniques sont distribués en sous-bandes d'énergie de sorte que les électrons ne constituent pas une population homogène, en outre la notion de densité de courant n'a plus de sens. Le courant transporté par *un* électron est donné par

$$I_1=-e\,v_z \tag{12-42}$$

où v_z est, dans la direction longitudinale z du fil, la vitesse de groupe du paquet d'onde associé à l'électron, donnée par (Eq. 1-58)

$$v_z = \frac{1}{\hbar}\frac{dE}{dk_z} \tag{12-43}$$

Le courant I_1 s'écrit donc

$$I_1 = -\frac{e}{\hbar}\frac{dE}{dk_z} \tag{12-44}$$

En explicitant dE/dk_z à partir de l'expression (12-34), ce courant s'écrit

$$I_1 = -\frac{e\hbar}{m_e}k_z \tag{12-45}$$

Le courant total est obtenu en sommant l'expression (12-45) sur l'ensemble des électrons.

A l'équilibre thermodynamique la somme dans l'espace des k_z est étendue de $-k_F$ à $+k_F$, où k_F est le vecteur d'onde du niveau de Fermi. Cette intégrale est évidemment nulle, la contribution d'un électron de vecteur d'onde k_z est compensée par celle d'un électron de vecteur d'onde $-k_z$.

En présence d'une polarisation V, tous les vecteurs d'onde varient de δk_z, la somme dans l'espace des k_z est alors étendue de $-k_F+\delta k_z$ à $+k_F+\delta k_z$. La contribution d'une sous bande i au courant total, s'écrit alors

$$I_i = \int_{-k_F+\delta k_z}^{k_F+\delta k_z} I_1 n_i(k_z)\,dk_z \tag{12-46}$$

A température nulle $n_i(k_z)$ est donné par $g(k_z)$, en explicitant I_1 (Eq. 12-45) et $g(k_z)$ (Eq. 12-35), le courant s'écrit

$$I_i = -\frac{e\hbar}{m_e}\frac{2}{2\pi}\int_{-k_F+\delta k_z}^{k_F+\delta k_z} k_z\,dk_z \tag{12-47}$$

En effectuant le changement de variable $k_z \rightarrow E$, les bornes de l'intégrale deviennent E_F et $E_F + \delta E = E_F - eV$ et la dérivée de l'expression (12-34) donne $k_z dk_z = m_e/\hbar^2 .dE$. L'expression (12-47) s'écrit alors

$$I_i = -\frac{e\hbar}{m_e}\frac{2}{2\pi}\frac{m_e}{\hbar^2}\int_{E_F}^{E_F-eV} dE \tag{12-48}$$

Soit

$$I_i = g_i V \tag{12-49}$$

avec

$$g_i = \frac{e^2}{\pi\hbar} = \frac{2e^2}{h} \tag{12-50}$$

g_i est la contribution à la conductance du fil quantique, de tous les électrons de la sous-bande i. Si N sous-bandes sont peuplées la conductance du fil s'écrit

$$g = N\frac{2e^2}{h} \tag{12-51}$$

Lorsque la population électronique augmente progressivement, le niveau de Fermi se déplace vers les hautes énergies et les différentes sous-bandes se peuplent successivement. Lorsqu'une nouvelle sous-bande est peuplée la conductance du fil quantique augmente brutalement d'une quantité finie, par saut de $2e^2/h$. La conductance du fil est donc quantifiée, avec un pas qui présente une universalité remarquable indépendante de tout paramètre physique ou géométrique.

Cette universalité existe aussi sur la résistance de Hall en régime quantique (Annexe A3). Cette dernière présente en fonction du champ magnétique, des paliers de valeurs $1/N.h/e^2$, où N est le nombre de niveaux de Landau peuplés. La constante $R_K=h/e^2$ porte le nom de *constante de von Klitzing* (prix Nobel 1985). Le facteur 2 qui différencie les deux phénomènes résulte simplement du fait que sur les états quantiques mis en jeu, la dégénérescence de spin est levée par le champ magnétique (effet Hall) alors qu'elle ne l'est pas par le champ électrostatique (fil quantique). Cette résistance de Hall, qui est mesurée avec une incertitude inférieure à 10^{-9}, est depuis 1990 utilisée comme *étalon de résistance*, sa valeur est

$$R_K = \frac{h}{e^2} = 25\,812{,}807\,\Omega$$

Ces variations universelles de conductance peuvent être observées dans différentes expériences, à très basse température, dans lesquelles un paramètre permet de moduler la distance du niveau de Fermi aux sous-bandes d'énergie. Ce paramètre peut faire varier la distribution des sous-bandes d'énergie à population électronique constante, ou inversement, la population électronique à sous-bande constantes. Le premier cas de figure correspond à l'effet Hall quantique, le second peut être obtenu dans le canal d'un TEGFET par la modulation de la tension de grille. La conductance du canal fait alors apparaître, en fonction de la tension grille, une succession de plateaux avec des sauts de $2e^2/h$ (Fig. 12-8).

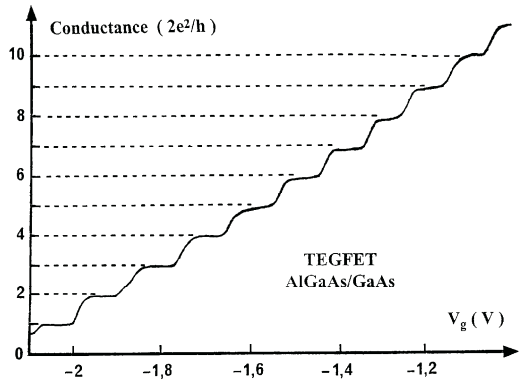


Figure 12-8 : Variation de conductance mesurée sur le canal d'un TEGFET AlGaAs/GaAs en fonction de la tension grille V_g *(Van Weels B.J. - 1988)*

12.2. Transport perpendiculaire dans les structures quantiques

12.2.1. *Oscillateur de Bloch*

Considérons un électron dans la bande de conduction d'un semiconducteur de maille *a,* et supposons que cet électron ne soit soumis à aucun phénomène de relaxation par des défauts ou vibrations thermiques du réseau. Les états d'énergie de cet électron sont donnés par la loi de dispersion $E(k)$ représentée schématiquement, dans la première zone de Brillouin, pour la direction k_z , sur la figure (12-9).

La vitesse de l'électron dans un état de vecteur d'onde $\boldsymbol{k}$ (Fig. 12-9) est donnée par la vitesse de groupe du paquet d'onde centré sur le vecteur $\boldsymbol{k}$ (Eq. 1-58).

$$\mathbf{v}(k)=\frac{1}{\hbar}\frac{dE}{d\mathbf{k}} \tag{12-52}$$

Appliquons un champ électrique dans la direction z, d'amplitude E_z suffisamment faible pour ne pas induire de transitions interbandes. L'électron est soumis à une force $F_z=-eE_z$. Les variations, parallèles, du bas et du sommet de la bande de conduction dans l'espace réel sont représentées sur la figure (12-9) pour $E_z<0$.

Considérons un électron du bas de la bande de conduction, c'est à dire en $k=0$ à l'instant $t=0$. Son énergie est E_c , sa vitesse est nulle (Fig. 12-9-a). L'équation de son mouvement s'écrit simplement

$$F_z = m\gamma_z = m\frac{dv_z}{dt} = \frac{d(mv_z)}{dt} = \frac{dp_z}{dt}$$

En explicitant $p_z=\hbar k_z$ et $F_z = -e\,E_z$ on obtient l'équation bien connue

$$\frac{dk_z}{dt}=-\frac{1}{\hbar}eE_z \tag{12-53}$$

L'intégration de cette équation avec la condition initiale $k(t=0)=0$, donne

$$k_z(t)=-\frac{1}{\hbar}eE_z t \tag{12-54}$$

Ainsi l'électron se déplace dans le sens opposé au champ électrique et son vecteur d'onde varie linéairement avec le temps. La position, l'énergie et la vitesse de l'électron sont représentées sur la figure (12-9) à différents instants. De $t=0$ à $t=t_1$ la vitesse de l'électron augmente et passe par un maximum à l'instant t_1 (Fig. 12-9-b). Ensuite elle diminue et s'annule à l'instant t_2 pour $k_z=\pi/a$. L'électron est alors en bord de zone de Brillouin, au sommet de la bande de conduction (Fig. 12-9-c). Il

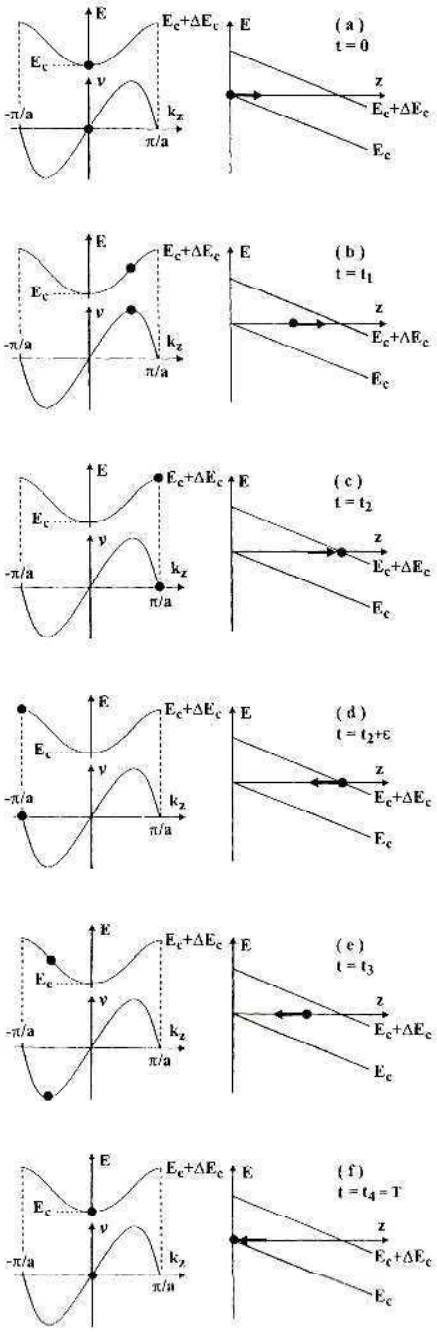


Figure 12-9 : Oscillateur de Bloch

subit ici une réflexion de Bragg, son vecteur d'onde change de signe et il se retrouve au point équivalent $-\pi/a$ de la zone de Brillouin avec une vitesse nulle (Fig. 12-9-d). Sa vitesse réaugmente alors par valeurs négatives (Fig. 12-9-e), puis diminue et tend vers zéro. A l'instant $t=t_4=T$ l'électron se retrouve au point de départ (Fig. 12-9-f).

Sous les actions combinées du champ périodique cristallin et du champ statique appliqué, l'électron oscille donc dans le temps et dans l'espace. Ce phénomène porte le nom *d'oscillations de Bloch*. Il faut remarquer que dans ces conditions, le champ statique appliqué ne produit aucun courant continu.

La période du mouvement correspond au temps mis par l'électron pour que son vecteur d'onde varie de $2\pi/a$. L'expression (12-54) donne $2\pi/a = eE_zT/\hbar$, soit

$$T=\frac{h}{eE_z a} \qquad (12\text{-}55)$$

Le mouvement de l'électron dans l'espace réel est donné par l'intégration de l'équation (12-52) qui s'écrit suivant z

$$\frac{dz}{dt}=\frac{1}{\hbar}\frac{dE}{dk_z} \qquad (12\text{-}56)$$

Soit

$$z=z_o+\frac{1}{\hbar}\int_o^t \frac{dE}{dk_z}dt \tag{12-57}$$

L'équation (12-53) permet d'expliciter dt en fonction de dk_z

$$dt=-\frac{\hbar}{eE_z}dk_z \tag{12-58}$$

L'expression (12-57) s'écrit alors

$$z=z_o-\frac{1}{eE_z}\int_{k_{zo}}^{k_z} \frac{dE}{dk_z}dk_z \tag{12-59}$$

ou

$$z=z_o-\frac{E(k_z)-E(k_{zo})}{eE_z} \tag{12-60}$$

Au cours de son mouvement périodique, l'amplitude de la variation d'énergie de l'électron est $E(k_z=\pi/a)-E(k_z=0)=\Delta E_c$ où ΔE_c représente la largeur de la bande de conduction. Ainsi l'amplitude du mouvement dans l'espace réel est donnée par

$$\Delta z=\frac{\Delta E_c}{eE_z} \tag{12-61}$$

En fait le modèle développé ci-dessus est idéalisé car dans le cristal réel les électrons sont soumis à des collisions. Si la fréquence de collision est importante, le libre parcours moyen des électrons est inférieur à Δz , leur mouvement n'est alors plus régi par les réflexions de Bragg mais correspond au schéma représenté sur la figure (12-10). Les oscillations de Bloch n'existent pas et le courant présente une composante continue non nulle.

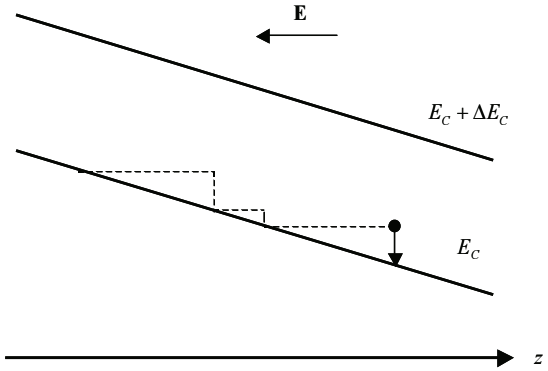


Figure 12-10 : Conduction dans un solide

La maille a d'un cristal semiconducteur est de l'ordre de quelques Angströms et le temps de relaxation des électrons est compris entre 10^{-13} et 10^{-12} seconde. Avec a=3,5 Å et $\tau=10^{-13}$ s on obtient $E_z>10^6$ V/cm . En outre la largeur ΔE_c de la bande de conduction étant supérieure à 10 eV, l'amplitude du mouvement spatial de l'électron est alors $\Delta z \approx \Delta E_c / e\,E_z \approx 1000$ Å. Notons d'une part que l'amplitude Δz est supérieure au libre parcours moyen des électrons et d'autre part que la valeur du champ nécessaire à l'observation des oscillations est considérable. Cette valeur est comparable au champ d'ionisation du matériau c'est-à-dire au champ électrique suffisant pour induire des transitions électroniques de la bande de valence vers la bande de conduction. De fait, les oscillations de Bloch n'ont jamais été observées dans des cristaux semiconducteurs massifs, et l'idée de générer des ondes ultra courtes à partir d'un oscillateur de Bloch a vite été abandonnée.

L'espoir de réaliser des oscillateurs de Bloch a retrouvé une seconde jeunesse avec l'avènement des superréseaux. L'empilement des structures de multicouches alternées de semiconducteurs différents permet de créer une surstructure, le *superréseau* (Par. 10-3), dont on peut maîtriser, dans une certaine gamme, la périodicité. Dans ce cas, la périodicité spatiale de la structure dans la direction perpendiculaire aux couches du superréseau est beaucoup plus importante que la maille du cristal, il en résulte que la période du mouvement des électrons est beaucoup plus faible. En outre la largeur des minibandes du superréseau étant très inférieure à celle de la bande de conduction du semiconducteur homogène, l'amplitude des oscillations est réduite d'autant.

Considérons un superréseau, à base de GaAs , de 100 Å de période (à comparer à 3,5 Å dans un semiconducteur homogène) : la condition d'observation des oscillations de Bloch est un champ $E_z>4.10^4$ V/cm. La largeur des minibandes est $\Delta E_c \approx 60$ meV (à comparer à plusieurs eV dans le semiconducteur homogène). L'amplitude spatiale des oscillations est alors $\Delta z \approx 60.10^{-3} / 4.10^4 \approx 150$ Å. On constate aisément que les conditions d'observation des oscillations sont beaucoup plus favorables.

Néanmoins ces oscillations n'ont jamais été observées, et la raison n'est pas simplement d'ordre expérimental. En effet, dans des champs électriques aussi élevés que ceux que nous venons d'obtenir, le modèle semi-classique, utilisé ci-dessus pour décrire le mouvement de l'électron, atteint ses limites car il suppose implicitement que la structure de bandes d'énergie et les relations de dispersion sont conservées en présence du champ électrique. En d'autres termes, le concept d'oscillation de Bloch résulte d'une extrapolation, au cas perturbé par un fort champ électrique, de résultats rigoureux obtenus en l'absence de champ électrique. Cette approche semi-classique est sujette à caution pour plusieurs raisons dont la plus évidente est que la périodicité de l'Hamiltonien, qui existe en l'absence de champ électrique, disparaît en présence du champ. Le champ électrique détruit l'invariance par translation et par suite la nature étendue des états d'énergie. La base des fonctions de Bloch ne reflète donc plus la nouvelle symétrie du problème. Les fonctions d'onde électroniques, initialement étendues à tout le cristal, ou le

superréseau, se trouvent modifiées et localisées sur un nombre de périodes, d'autant plus faible que le champ est important. Cette idée, apparemment paradoxale, de localisation des électrons par un champ électrique statique, avait déjà été émise par Wannier en 1959, sans beaucoup de succès à l'époque car invérifiable sur les cristaux massifs.

12.2.2. *Effet Wannier-Stark*

Dans un superréseau, le traitement semi-classique de la réponse d'un électron de conduction à un champ électrique statique, prévoit l'existence d'oscillations de Bloch. Toutefois, ce type de traitement apparaît limité compte tenu de la valeur élevée du champ critique nécessaire à l'observation de ce régime d'oscillation.

Le traitement quantique du problème consiste à résoudre l'équation de Schrödinger du mouvement de l'électron dans le potentiel périodique du superréseau $V_c(z)$, soumis à un champ électrique stationnaire E_z dans la direction du superréseau.

$$\left(-\frac{\hbar^2}{2m_e}\frac{\partial^2}{\partial z^2}+V_c(z)+eE_z z\right)\psi(z)=E\psi(z) \qquad (12\text{-}63)$$

Le potentiel $V_c(z)$ étant périodique de période a dans la direction z, $V_c(z+a)=V_c(z)$, les propriétés de translation de l'Hamiltonien font clairement apparaître que si E est une valeur propre de l'état $\psi(z)$, $E+neE_z a$ est une valeur propre de l'état $\psi(z-na)$ où n est un nombre entier. Ainsi d'une cellule à l'autre du superréseau, c'est-à-dire d'un puits à son voisin, les états d'énergie varient par sauts de valeur $\Delta E=eE_z a$. L'ensemble de ces états constitue ce que l'on appelle l'*échelle de Wannier-Stark* des niveaux d'énergie.

La relation énergie-pulsation $\Delta E=\hbar\omega$, permet de relier l'écart énergétique entre deux niveaux Stark à une période d'oscillation

$$\omega=\Delta E/\hbar = eE_z a/\hbar \Rightarrow \nu=\omega/2\pi=eE_z a/h \Rightarrow T=1/\nu=h/eE_z a$$

On retrouve la période T des oscillations de Bloch (Eq. 12-55). Le régime d'états stationnaires sur l'échelle de Wannier-Stark est le pendant dans le domaine énergétique des oscillations de Bloch dans le domaine temporel.

Considérons les électrons de conduction du superréseau. En l'absence de champ électrique les fonctions d'onde électroniques s'étendent dans tout le superréseau, les niveaux d'énergie constituent un quasi-continuum dans chacune des sous-bandes permises (Fig. 12-11-a). Considérons la première sous-bande. La présence du champ électrique découple partiellement les puits adjacents et réduit la cohérence spatiale des fonctions d'onde. L'extension spatiale de la fonction d'onde électronique diminue quand le champ électrique augmente et l'électron se localise sur un nombre décroissant de puits. Dans la limite des champs forts, la localisation

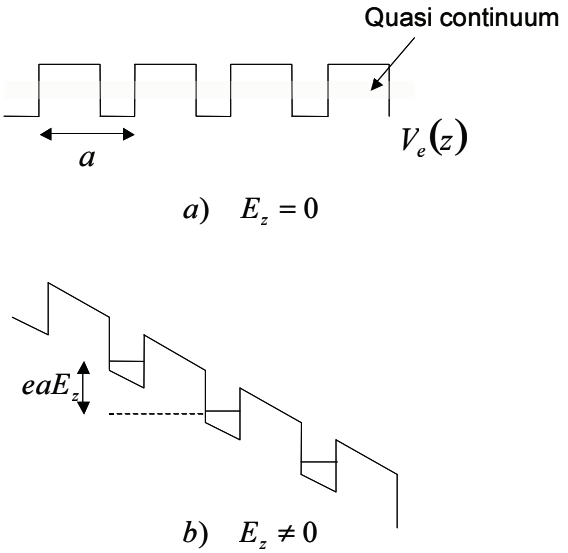


Figure 12-11 : Effet Wannier-Stark

est totale, l'électron est confiné dans un seul puits du superréseau. Le superréseau devient alors un ensemble de multipuits quantiques indépendants. Parallèlement le quasi-continuum d'états de la sous-bande de conduction se scinde en un ensemble de niveaux discrets équidistants en énergie d'une quantité $\Delta E=eE_z a$ (Fig. 12-11-b). Dans la limite des champs forts correspondant à la localisation totale de l'électron, ce dernier est piégé dans un puits et on retrouve l'absence totale de courant électrique malgré la présence du champ, comme dans le régime d'oscillations de Bloch.

12.2.3. *Transport dans un superréseau-RDN*

Nous venons de voir comment un électron placé dans un réseau périodique de période *a,* et soumis à un champ électrique uniforme E_z , peut, dans les limites du modèle, effectuer un mouvement périodique dans la direction du champ électrique avec une fréquence caractéristique $f=eE_z a/h$, (oscillations de Bloch). Ceci résulte du fait que le moment de cet électron augmente linéairement avec le temps alors que son énergie et sa vitesse sont des fonctions périodiques du moment cristallin. La distance parcourue par l'électron pendant un cycle est de l'ordre de $\Delta E_c/eE_z$ où ΔE_c est la largeur de la minibande dans laquelle évolue l'électron.

En langage de la mécanique quantique, la présence d'un champ électrique scinde la bande d'énergie électronique en un ensemble de niveaux discrets séparés par un intervalle énergétique constant $eE_z a$, (échelle de Wannier-Stark). Chacun des

niveaux correspond à une fonction d'onde centrée sur un site du réseau avec une étendue de l'ordre de $\Delta E_c/eE_z$, dans la direction du champ. Ceci résulte simplement du fait que dans le superréseau, une translation d'une période change l'Hamiltonien du système d'une valeur constante $eE_z a$.

Il apparaît donc possible, en principe, d'induire dans un solide périodique un courant alternatif avec un champ continu. En fait, en raison des phénomènes de relaxation, l'électron redescend au bas de la sous-bande bien avant d'avoir effectué une période d'oscillation, de sorte que les oscillations de Bloch n'ont jusqu'ici pas été observées. Néanmoins sous l'action d'un champ suffisant et malgré la présence des phénomènes de relaxation, l'électron peut monter suffisamment haut dans la sous-bande de conduction pour atteindre des régions à masse effective négative.

Dans ce cas, sous l'action du champ électrique l'énergie et le vecteur d'onde d'un électron augmentent de façon monotone, mais son accélération change de signe. Au-delà d'une certaine valeur du champ, l'électron ralentit quand le champ augmente, ce qui doit se traduire par une caractéristique $I(V)$ à pente négative, c'est-à-dire par une résistance différentielle négative (RDN). La figure (12-9) n'est que schématique mais montre néanmoins que le sommet de la courbe de vitesse apparaît pour une valeur relativement importante k_s du vecteur d'onde, k_s étant supérieur à $\pi/2a$. Dans un cristal homogène, a est de l'ordre de quelques Angströms, de sorte que k_s est grand, il en résulte que le champ nécessaire est trop important pour que le phénomène soit accessible expérimentalement. Dans un superréseau par contre, la périodicité a se chiffre en centaines d'Angströms et les champs électriques nécessaires sont beaucoup plus accessibles.

Considérons un superréseau dont la courbe de dispersion présente dans la direction z, une loi simplifiée de la forme

$$E(k)=E_c+\frac{\Delta E_c}{2}\left(1-\cos(ak_z)\right) \tag{12-64}$$

a est la périodicité du superréseau, E_c l'énergie du minimum de la 1ère sous-bande de conduction et ΔE_c sa largeur. Ce superréseau est soumis à un champ électrique E_z dans la direction perpendiculaire aux couches.

L'électron étant soumis à des phénomènes de relaxation, si on appelle $P(t)$ la probabilité pour qu'il n'ait subi aucune collision à l'instant t, sa vitesse moyenne d'entraînement dans la direction z s'écrit

$$v_n=\int_o^\infty P(t)dv \tag{12-65}$$

Si le temps entre deux collisions est τ, la probabilité $P(t)$ peut s'écrire sous la forme

$$P(t)=e^{-t/\tau} \tag{12-66}$$

L'équation (12-52) permet d'écrire

$$\frac{dv}{dt}=\frac{1}{\hbar}\frac{d}{dt}\frac{dE}{dk_z}=\frac{1}{\hbar}\frac{d^2E}{dk_z^2}\frac{dk_z}{dt} \tag{12-67}$$

Soit en explicitant dk_z/dt (Eq. 12-53)

$$\frac{dv}{dt}=-\frac{eE_z}{\hbar^2}\frac{d^2E}{dk_z^2}dt \tag{12-68}$$

ou

$$dv=-\frac{eE_z}{\hbar^2}\frac{d^2E}{dk_z^2}dt \tag{12-69}$$

Ainsi en explicitant $P(t)$ (Eq. 12-66) et dv (Eq. 12-69) , l'expression (12-65), s'écrit

$$v_n=-\frac{eE_z}{\hbar^2}\int_o^\infty e^{-t/\tau}\frac{d^2E}{dk_z^2}dt \tag{12-70}$$

ou, en explicitant d^2E/dk_z^2 à partir de l'expression (12-64)

$$v_n=-\frac{eE_z\Delta E_c a^2}{2\hbar^2}\int_o^\infty e^{-t/\tau}\cos(ak_z)dt \tag{12-71}$$

En explicitant k_z à partir de l'expression (12-54), on obtient compte tenu de la parité du cosinus

$$v_n=-\frac{eE_z\Delta E_c a^2}{2\hbar^2}\int_o^\infty e^{-t/\tau}\cos(\frac{eE_z a}{\hbar}t)dt \tag{12-72}$$

Soit

$$v_n=-\frac{e\Delta E_c a^2}{2\hbar^2}\frac{\tau}{1+\left(aeE_z\tau/\hbar\right)^2}E_z \tag{12-73}$$

La masse effective de l'électron dans la direction z du superréseau est donnée par

$$\frac{1}{m_e(k_z)}=\frac{1}{\hbar^2}\frac{d^2E}{dk_z^2} \tag{12-74}$$

Soit, compte tenu de l'expression (12-64)

$$\frac{1}{m_e(k_z)}=\frac{\Delta E_c a^2}{2\hbar^2}\cos(ak_z) \tag{12-75}$$

Au bas de la bande de conduction, en k_z=0, la masse effective m_e de l'électron s'écrit

$$\frac{1}{m_e}=\frac{\Delta E_c a^2}{2\hbar^2} \tag{12-76}$$

La vitesse moyenne de l'électron s'écrit donc

$$v_n=-\frac{e}{m_e}\frac{\tau}{1+\left(aeE_z\tau/\hbar\right)^2}E_z \tag{12-77}$$

Si n est la densité d'électrons de la sous-bande, le courant traversant le superréseau s'écrit

$$j_n=-nev_n=\frac{ne^2}{m_e}\frac{\tau}{1+\left(aeE_z\tau/\hbar\right)^2}E_z \tag{12-78}$$

Lorsque le champ électrique est faible , le dénominateur de l'expression (12-78) se réduit à 1 et l'expression s'écrit simplement

$$j_n=\frac{ne^2\tau}{m_e}E_z=ne\mu_nE_z=\sigma E_z \tag{12-79}$$

avec $\mu_n=e\tau/m_e$. On retrouve l'expression classique (Eq. 1-83) où σ est la conductivité à faible champ. Le courant varie linéairement, la caractéristique $I(V)$ est ohmique.

Lorsque le champ devient important, l'expression (12-78) présente d'abord une variation sous-linéaire, puis le courant j_n passe par un maximum pour $E_z=E_{zmax}$ donné par

$$E_{z\max}=\frac{\hbar}{ea\tau} \tag{12-80}$$

Ensuite j_n diminue quand E_z augmente, la caractéristique $I(V)$ présente alors une pente négative. Le dipôle ainsi polarisé au-delà de E_{zmax} présente donc une résistance différentielle négative (RDN).

Il suffit de comparer les expressions (12-80 et 12-62) pour constater que le champ requis pour accéder à cette résistance différentielle négative est 2π fois inférieur au champ nécessaire à l'établissement des oscillations de Bloch.

12.2.4. *Effet tunnel résonnant dans un puits quantique à double barrière*

L'effet tunnel à travers une barrière de potentiel est l'un des phénomènes les plus étudiés en mécanique quantique. Il joue un rôle déterminant dans certains composants électroniques, comme par exemple la diode tunnel qui exploite les propriétés de l'effet tunnel des électrons à travers la zone de charge d'espace d'une jonction *pn* dégénérée.

L'effet tunnel résonnant se manifeste lorsqu'une particule doit traverser successivement deux barrières de potentiel, "pontées" par un ou plusieurs états discrets permis. Considérons l'hétérostructure constituée d'un puits quantique pris en sandwich entre deux barrières (Fig. 12-12-a). C'est le système dual du double puits quantique (Par. 10.2.3). Le puits est constitué d'un semiconducteur SC1 faiblement dopé, de gap E_{g1} et d'épaisseur L_1, entre deux couches de semiconducteur SC2 non dopé, de gap $E_{g2}>E_{g1}$ et d'épaisseur commune L_2. Cette structure à trois couches est bordée à ses extrémités par des couches de semiconducteur SC1 dégénéré de type *n*. Ces deux extrémités constituent l'*émetteur* et le *collecteur* du dipôle. Les figures (12-12-b à d) présentent le diagramme énergétique de la bande de conduction de la structure et son évolution sous une polarisation $V>0$. Nous supposerons que la hauteur des barrières et la largeur du puits sont telles qu'une seule sous-bande de conduction d'énergie E_1 est présente dans le puits (Par. 10.2.1). Les barrières étant isolantes et le puits faiblement dopé, on peut supposer que la chute de potentiel entre l'émetteur et le collecteur se distribue essentiellement dans les barrières, en d'autres termes le fond du puits reste sensiblement horizontal.

Sous l'action de la polarisation positive V du collecteur, l'énergie E_1 du bas de la sous-bande de conduction du puits diminue. Pour une certaine valeur de V , cette sous-bande se présente en regard des états occupés de la bande de conduction de l'émetteur. Les électrons de ce dernier peuvent alors traverser la première barrière par effet tunnel pour occuper la sous-bande E_1 du puits. De là, les électrons peuvent passer, par effet tunnel à travers la deuxième barrière, sur les états vides de la bande de conduction du collecteur. Ils se thermalisent alors dans la mer de Fermi de ce dernier. Le courant tunnel, qui est nul en l'absence de polarisation, augmente à mesure que le nombre d'états en regard, occupés dans la bande de conduction de l'émetteur et vides dans la sous-bande de conduction du puits, augmente. La sous-bande E_1 sert de pont pour assister l'effet tunnel entre l'émetteur et le collecteur à travers les deux barrières. Cet effet est évidemment maximum à la résonance. Lorsque la chute de potentiel dans la première barrière atteint la valeur E_1/e le bas de la sous-bande 1 s'aligne avec le bas de la bande de conduction de l'émetteur, le courant est alors maximum. Au-delà, le niveau E_1 passe au-dessous de la bande de

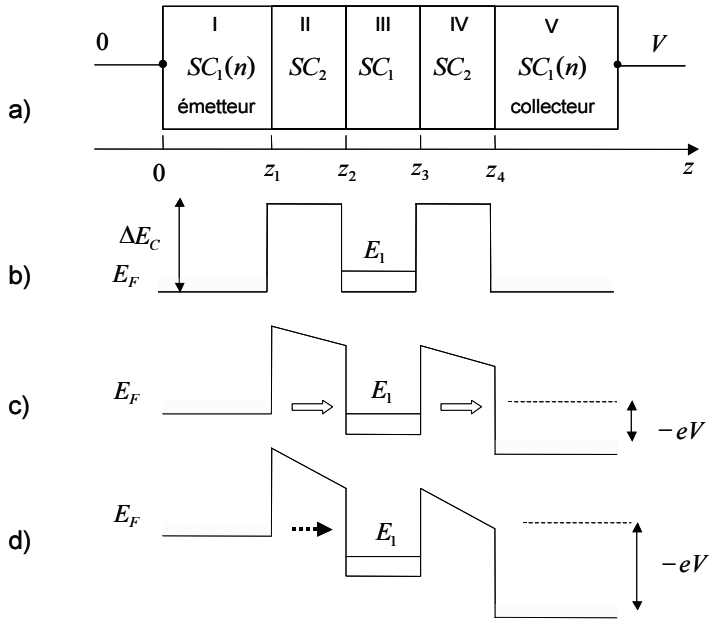


Figure 12-12 : Effet tunnel résonnant

conduction de l'émetteur, la résonance disparaît et le courant tunnel diminue rapidement puis s'annule. Notons que, compte tenu d'une part de la symétrie de la structure (barrières d'égales largeurs L_2) et d'autre part du fait que la variation du potentiel se distribue dans les barrières, la chute de potentiel est la même dans chaque barrière. Cette chute atteint donc la valeur E_1/e pour une tension de polarisation $V=2E_1/e$. La caractéristique $I(V)$ du dipôle est représentée sur la figure (12-13).

Pour une tension de polarisation supérieure à $2E_1/e$, elle présente une pente négative qui fait de ce dipôle une résistance différentielle négative. Lorsque la tension continue d'augmenter, le courant augmente à nouveau par émission thermique au-dessus des barrières. Cette émission est éventuellement assistée par effet tunnel, soit au sommet triangulaire de la barrière polarisée, soit par résonance sur des états excités du puits. Comme la diode tunnel (Par. 2.8), le dipôle est caractérisé par un courant de pic et un courant de vallée.

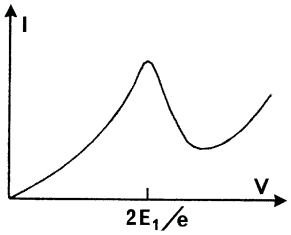


Figure 12-13 : Caractéristique du dipôle

Divers modèles théoriques plus ou moins sophistiqués permettent de décrire quantitativement le phénomène d'effet tunnel résonnant à travers une double barrière. Le plus simple est basé sur l'utilisation de la *matrice de transfert* (Tsu R.-1973).

Considérons un électron de vecteur d'onde k dans l'émetteur, il doit traverser les barrières par effet tunnel pour atteindre le collecteur. La structure n'étant pas invariante par translation suivant z, la composante k_z du vecteur d'onde n'est pas conservée, par contre, la composante $k_{//} = (k_x^2 + k_y^2)^{1/2}$ est conservée. Dans l'approximation de la masse effective on peut donc séparer le mouvement de l'électron suivant z de son mouvement dans le plan. Dans l'émetteur, le puits, et le collecteur, l'électron se comporte comme une particule libre de masse effective m_1. Dans le plan xy sa fonction d'onde est une onde plane, son énergie est donnée par $E_{//}=\hbar^2 k_{//}^2 / 2m_1$. Dans la direction z son mouvement est régi par l'équation de Schrödinger unidimensionnelle

$$\left[-\frac{\hbar^2}{2m}\frac{d^2}{dz^2}+V(z)-E_z\right]\psi(z)=O \tag{12-81}$$

L'énergie potentielle $V(z)$ étant constante dans chacune des régions, la solution générale de l'équation est de la forme

$$\psi(z)= A\, e^{ik_z z} + B\, e^{-ik_z z} \tag{12-82}$$

Avec

$$\frac{\hbar^2 k_z^2}{2m}=E_z-V \tag{12-83}$$

où V représente la valeur, constante, de l'énergie potentielle dans chacune des régions considérées. Dans les régions où E_z est supérieur à V, k_z^2 est positif de sorte que k_z est réel et par suite la fonction d'onde est une onde plane. Dans les régions où E_z est inférieur à V, k_z^2 est négatif de sorte que k_z est imaginaire et par suite la fonction d'onde est une exponentielle croissante ou décroissante. Nous nous intéressons ici aux électrons qui traversent les barrières par effet tunnel, c'est-à-dire aux électrons dont l'énergie est inférieure à la hauteur des barrières. Il en résulte que le premier cas (E_z-V>0) correspond à l'émetteur, au puits, et au collecteur, et le second cas (E_z-V<0) aux barrières.

Les constantes d'intégration A et B sont déterminées par les conditions habituelles de continuité de la fonction d'onde et du courant de probabilité aux interfaces. La masse effective des électrons est m_1 dans le semiconducteur SC1 et m_2 dans le semiconducteur SC2. Ces conditions s'écrivent

$$\psi(z_j^-)=\psi(z_j^+) \tag{12-84-a}$$

$$\frac{1}{m}\frac{\partial\psi}{\partial z}\bigg|_{z_j^-}=\frac{1}{m}\frac{\partial\psi}{\partial z}\bigg|_{z_j^+} \tag{12-84-b}$$

avec j=1,2,3,4 (Fig. 12-12).

Au niveau de la première interface, ces conditions donnent la relation matricielle

$$\begin{Bmatrix} A_I \\ B_I \end{Bmatrix} = M_1 \begin{Bmatrix} A_{II} \\ B_{II} \end{Bmatrix} \tag{12-85}$$

où M_1 qui porte le nom de *matrice de transfert* est donné par

$$M_1 = \frac{1}{2}\begin{bmatrix} \left(1+\frac{k_2 m_1}{k_1 m_2}\right) e^{i(k_2-k_1)z_1} & \left(1-\frac{k_2 m_1}{k_1 m_2}\right) e^{-i(k_2+k_1)z_1} \\ \left(1-\frac{k_2 m_1}{k_1 m_2}\right) e^{i(k_2+k_1)z_1} & \left(1+\frac{k_2 m_1}{k_1 m_2}\right) e^{-i(k_2-k_1)z_1} \end{bmatrix} \tag{12-86}$$

La composante k_{zj} a été écrite k_j pour simplifier le formalisme. La généralisation aux quatre interfaces de la structure s'écrit

$$\begin{Bmatrix} A_I \\ B_I \end{Bmatrix} = M_1 M_2 M_3 M_4 \begin{Bmatrix} A_V \\ B_V \end{Bmatrix} \tag{12-87}$$

avec

$$M_j = \frac{1}{2}\begin{bmatrix} \left(1+\frac{k_{j+1} m_j}{k_j m_{j+1}}\right) e^{i(k_{j+1}-k_j)z_j} & \left(1-\frac{k_{j+1} m_j}{k_j m_{j+1}}\right) e^{-i(k_{j+1}+k_j)z_j} \\ \left(1-\frac{k_{j+1} m_j}{k_j m_{j+1}}\right) e^{i(k_{j+1}+k_j)z_j} & \left(1+\frac{k_{j+1} m_j}{k_j m_{j+1}}\right) e^{-i(k_{j+1}-k_j)z_j} \end{bmatrix} \tag{12-88}$$

ou k_j est donné par la relation (12-83)

$$k_j = \frac{1}{\hbar}\sqrt{2m_j(E_z - V_j)} \tag{12-89}$$

A partir de ces résultats, on peut calculer simplement le coefficient de transmission à travers une structure donnée. La probabilité de présence dans l'émetteur, d'un électron incident se propageant dans la direction z^+ est A_I^2. La probabilité de présence dans une région j, d'un électron transmis se propageant dans la direction z^+ est A_J^2. Le coefficient de transmission d'un électron d'énergie E_z est par conséquent donné par

$$T(E_z)=\left|\frac{A_J}{A_I}\right|^2 \tag{12-90}$$

Ainsi le coefficient de transmission à travers une simple barrière, la première de notre structure par exemple, est donné en fonction de son énergie E_z dans la direction z, par

$$T(E_z)=\left|\frac{A_{III}}{A_I}\right|^2 \tag{12-91}$$

avec

$$\begin{Bmatrix} A_I \\ B_I \end{Bmatrix} = M_1 M_2 \begin{Bmatrix} A_{III} \\ B_{III} \end{Bmatrix} \tag{12-92}$$

La hauteur de la barrière est $\Delta E_c=E_{c2}-E_{c1}$, sa largeur est L_2, le produit des matrices M_1M_2 donne

$$T(E_z)=\frac{1}{1+\frac{\Delta E_c^2}{4E_z(\Delta E_c-E_z)}sh^2\left(\frac{L_2}{\hbar}\sqrt{2m_1(\Delta E_c-E_z)}\right)} \tag{12-93}$$

Ce coefficient de transmission est évidemment inférieur à 1. Il est nul pour E_z=0. Il tend vers zéro quand la hauteur ΔE_c ou la largeur L_2 de la barrière tend vers l'infini. Si la barrière tend vers une fonction de Dirac, c'est-à-dire $L_2\to 0$ et $\Delta E_c\to\infty$ avec $L_2\Delta E_c$=C, le coefficient de transmission s'écrit

$$T(E_z)=\frac{1}{1+\frac{1}{2}\frac{m_1C^2}{\hbar^2E_z}} \tag{12-94}$$

Revenons sur la double barrière, le coefficient de transmission de la structure est donnée par

$$T(E_z)=\left|\frac{A_V}{A_I}\right|^2 \tag{12-95}$$

La figure (12-14) représente la variation de $T(E_z)$ pour une double barrière de hauteur ΔE_c=1,2 eV et de largeur L_2=26 Å, bordant un puits de largeur L_1=50 Å. Compte tenu des caractéristiques du puits, plusieurs sous-bandes existent dans le puits, les pics de transmission correspondent à l'effet tunnel résonnant sur chacune

des deux premières sous-bandes, dont les énergies sont données dans l'approximation du puits infini, par (Eq. 10-29).

$$E_1 \approx E_c + \frac{\pi^2 \hbar^2}{2 m_1 L_1^2} \quad E_2 \approx E_c + 4 \frac{\pi^2 \hbar^2}{2 m_1 L_1^2} \tag{12-96}$$

Il faut noter que le potentiel, qui est constant dans chacune des régions en l'absence de polarisation, devient variable lorsque la structure est polarisée. La méthode de la matrice de transfert reste opérationnelle, il suffit de modéliser le potentiel lentement variable par une succession de marches de longueurs infinitésimales, à l'intérieur desquelles le potentiel est constant.

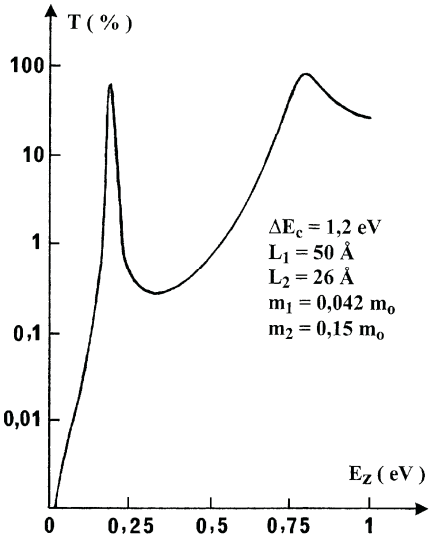


Figure 12-14 : Coefficient de transmission des électrons à travers une double barrière en fonction de leur énergie longitudinale *(Singh J. - 1993)*

La loi de variation du coefficient de transmission des électrons à travers la structure en fonction de leur énergie longitudinale E_z , pilote la loi de variation du courant à travers le dipôle en fonction de sa polarisation. Ce courant est obtenu en sommant la probabilité d'effet tunnel sur toute la distribution des électrons dans l'émetteur.

La densité d'électrons d'énergie E_z dans l'émetteur est donnée par

$$n_e(E_z) = \int_0^\infty g(E_{//}) f(E) \, dE_{//} \tag{12-97}$$

où $f(E)$ est la fonction de distribution de Fermi et $g(E_{//})$ la densité d'états bidimensionnelle dans le plan xy. Cette densité d'états est donnée par (Eq. 10-18)

$$g(E_{//}) = \frac{m_1}{\pi \hbar^2} \tag{12-98}$$

Soit

$$n_e(E_z) = \frac{m_1}{\pi \hbar^2} \int_{E_z}^\infty \frac{dE_{//}}{1 + e^{(E_z + E_{//} - E_{Fe})/kT}} \tag{12-99}$$

En explicitant l'intégrale de Fermi (Eq. 10-22), on obtient

$$n_e(E_z)=\frac{m_1 kT}{\pi\hbar^2}Ln\left(1+e^{(E_{Fe}-E_z)/kT}\right) \tag{12-100}$$

où E_{Fe} est l'énergie du niveau de Fermi dans l'émetteur.

La probabilité de transfert émetteur → collecteur, d'un électron d'énergie E_z , est donnée par le coefficient de transmission $T(E_z)$, de sorte que la densité d'électrons transmis est donnée par

$$n_{et}(E_z)=T(E_z)\ n_e(E_z) \tag{12-101}$$

Soit en explicitant $n_e(E_z)$

$$n_{et}(E_z)=\frac{m_1 kT}{\pi\hbar^2}T(E_z)Ln\left(1+e^{(E_{Fe}-E_z)/kT}\right) \tag{12-102}$$

Le courant transporté par ces électrons s'écrit

$$j_{e\to c}(E_z)=n_{et}(E_z)\ e\ v(E_z) \tag{12-103}$$

où la vitesse suivant z est donnée par (Eq. 1-58)

$$v(E_z)=\frac{1}{\hbar}\frac{dE_z}{dk_z} \tag{12-104}$$

De sorte que le courant transporté par les électrons d'énergie E_z dans l'émetteur s'écrit

$$j_{e\to c}(E_z)=\frac{e}{\hbar}n_{et}(E_z)\frac{dE_z}{dk_z} \tag{12-105}$$

Le courant transporté par tous les électrons de l'émetteur s'écrit alors

$$j_{e\to c}=\frac{e}{\hbar}\sum_{k_z>o}n_{et}(E_z)\frac{dE_z}{dk_z} \tag{12-106}$$

En remplaçant la somme discrète par une intégrale sur k_z

$$j_{e\to c}=\frac{e}{2\pi\hbar}\int_0^\infty n_{et}(E_z)\frac{dE_z}{dk_z}dk_z \tag{12-107}$$

En intégrant non pas sur la variable k_z mais sur la variable E_z on obtient

$$j_{e\to c}=\frac{e}{2\pi\hbar}\int_{E_{ce}}^{\infty} n_{et}(E_z)\,dE_z \qquad (12\text{-}108)$$

où E_{ce} est le bas de la bande de conduction de l'émetteur. En explicitant $n_{et}(E_z)$ (Eq. 12-102), le courant s'écrit

$$j_{e\to c}=\frac{em_1kT}{2\pi^2\hbar^3}\int_{E_{ce}}^{\infty} T(E_z)\,Ln\left(1+e^{(E_{Fe}-E_z)/kT}\right)dE_z \qquad (12\text{-}109)$$

Le même type de développement permet d'obtenir l'expression du courant collecteur → émetteur

$$j_{c\to e}=\frac{em_1kT}{2\pi^2\hbar^3}\int_{E_{ce}}^{\infty} T(E_z)\,Ln\left(1+e^{(E_{Fc}-E_z)/kT}\right)dE_z \qquad (12\text{-}110)$$

Il faut noter que la borne inférieure de l'intégrale reste, comme dans l'expression (12-109), le bas de la bande de conduction de l'émetteur E_{ce}. Aucun électron ne peut passer par effet tunnel du collecteur vers l'émetteur si son énergie suivant z n'est pas supérieure à E_{ce}. $T(E_z)$ est nul pour $E_z<E_{ce}$.

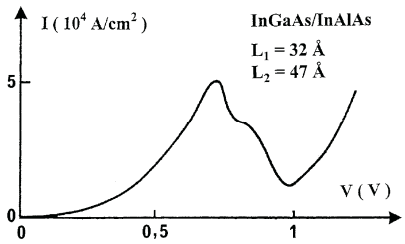


Figure 12-15 : Caractéristique d'une structure à double barrière
(Sugiyama Y. - 1988)

Le courant résultant est donné par la différence de ces deux courants, qui s'écrit compte-tenu du fait que la différence des niveaux de Fermi est égale à la polarisation de la structure, $E_{Fc}=E_{Fe}$-eV

$$J=\frac{em_1kT}{2\pi^2\hbar^3}\int_{E_{ce}}^{\infty} T(E_z)\,Ln\left(\frac{1+e^{(E_{Fe}-E_z)/kT}}{1+e^{(E_{Fe}-eV-E_z)/kT}}\right)dE_z \qquad (12\text{-}111)$$

La caractéristique $J(V)$ du dipôle présente une allure de variation conditionnée en majeure partie par le coefficient de transmission $T(E_z)$. Elle présente un courant de pic, un courant de vallée, et une région à pente négative. La

figure (12-15) représente la caractéristique d'une structure à double barrière avec un puits de InGaAs et des barrières de InAlAs.

Quand la tension de polarisation augmente, la bande de conduction du collecteur descend très vite. Même à la température ambiante, peu d'électrons de la bande de conduction du collecteur ont une énergie supérieure au bas de la bande de conduction de l'émetteur. Il en résulte que le courant tunnel collecteur→émetteur devient négligeable. En termes quantitatifs, E_z étant nécessairement supérieur à E_{ce} dans l'expression (12-111), lorsque la condition $V >> (E_{Fe}-E_{ce}) / e$ est vérifiée, le terme exponentiel du dénominateur de l'expression (12-111) devient négligeable, et l'expression s'écrit

$$J \approx \frac{e m_1 kT}{2\pi^2 \hbar^3} \int_{E_{ce}}^{\infty} T(E_z) Ln\left(1+e^{(E_{Fe}-E_z)/kT}\right) dE_z \qquad (12\text{-}112)$$

On constate alors que le courant J n'est plus explicitement fonction de la tension de polarisation V. Il n'est fonction de celle-ci qu'à travers le facteur de transmission $T(E_z)$. Il en résulte que la caractéristique $J(V)$ est alors entièrement conditionnée par le facteur de transmission.

Le modèle simple développé ci-dessus donne un accord semi-quantitatif avec le comportement réel du dipôle. Des calculs plus élaborés, notamment en traitant de manière auto-cohérente les équations de Poisson et de Schrödinger, et en prenant en considération les phénomènes de diffusion, donnent des résultats plus quantitatifs.

12.2.5. *Effet tunnel résonnant dans un superréseau*

Le processus d'effet tunnel résonnant dans un puits quantique à double barrière peut être reproduit d'une manière sensiblement différente dans un superréseau. Le diagramme énergétique d'un superréseau polarisé est représenté sur la figure (12-16-a). Sous l'action d'un champ électrique élevé, les sous-bandes de conduction du superréseau se scindent en niveaux discrets (échelle de Wannier-Stark) et les électrons sont localisés dans les puits du superréseau, qui se retrouvent découplés par le champ électrique. Mais lorsque le champ électrique est tel que $eE_z a = E_j - E_1$, l'état fondamental E_1 du $n^{\text{ème}}$ puits est dégénéré avec l'état excité j (j=1,2,...) du $(n+1)^{\text{ème}}$ (Fig. 12-16-a). Dans ces conditions, un électron peut passer d'un puits au suivant par effet tunnel , avec dans ce dernier une relaxation du niveau j vers le niveau fondamental. L'électron peut donc se propager à travers le superréseau par le processus séquentiel effet tunnel résonnant-relaxation. La caractéristique $I(V)$ du dipôle présente ainsi un ou plusieurs pics de courant correspondant aux résonances successives (Fig. 12-16-b).

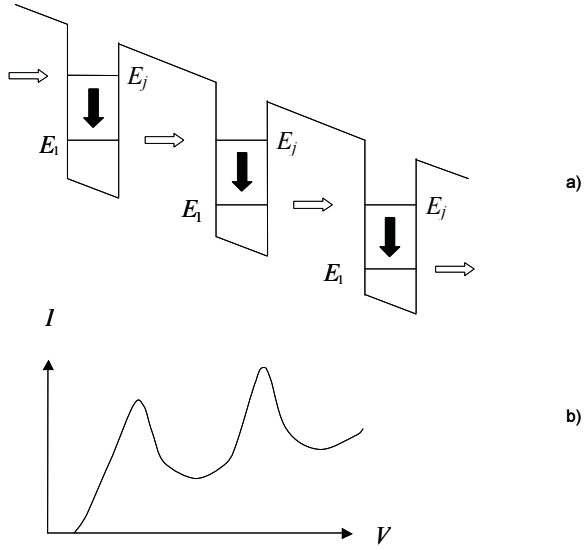


Figure 12-16 : Effet tunnel résonnant dans un superréseau

La difficulté essentielle qui se manifeste dans la réalisation de ce genre de dipôle, provient de la non uniformité du champ électrique à travers tout le superréseau. Le passage du courant est en effet conditionné par une résonance simultanée de tous les puits

12.3. Lasers à puits quantiques

Les lasers à puits quantiques mettent à profit les effets spécifiques du confinement quantique à la fois, sur les fonctions d'onde électroniques, sur la distribution énergétique, et sur la nature bidimensionnelle de la densité d'états. Le premier effet est d'ordre géométrique, il engendre un confinement des porteurs dans un faible volume ce qui, pour une injection donnée, permet d'augmenter la densité de porteurs et par suite d'accéder plus facilement à l'inversion de population. Les deux autres effets sont typiquement quantiques, ils se combinent pour améliorer les performances du laser notamment en ce qui concerne le courant de seuil, l'ajustement de la longueur d'onde d'émission, la sensibilité du courant de seuil à la température...

12.3.1. *Effets géométriques - Facteur de confinement*

Le courant de seuil dans une diode laser est proportionnel à l'épaisseur d de la zone active de la structure (Eq. 9-176). La diminution de ce courant de seuil passe donc par une réduction du volume actif.

Dans un laser à simple jonction *pn* la zone active, c'est-à-dire la zone d'inversion de population où le gain est supérieur aux pertes, est fixée par les longueurs de diffusion des porteurs minoritaires. Elle est typiquement de l'ordre de 1 μm.

La réalisation de lasers à double hétérojonction (DH) permet de réduire notablement le volume de la zone active en confinant les porteurs et les photons dans le matériau à petit gap. Les porteurs sont confinés par les barrières de potentiel associées aux discontinuités de gap, les photons sont confinés par les barrières optiques associées aux sauts d'indice de réfraction. L'indice de réfraction est en effet plus grand dans un semiconducteur à petit gap que dans un matériau à grand gap. Dans une structure DH l'épaisseur de la zone active est typiquement de l'ordre de 0,2 μm.

La valeur de 0,2 μm, caractéristique de la structure DH, est comparable aux longueurs d'ondes optiques et par conséquent permet un confinement optimisé des photons. Par contre, cette valeur est très supérieure aux longueurs d'ondes électroniques, de sorte que le confinement des électrons peut considérablement être amélioré. C'est le rôle du puits quantique. Mais lorsque les porteurs sont confinés dans un puits quantique, typiquement de l'ordre de quelques nanomètres, l'épaisseur de la zone active est très inférieure à la longueur d'onde optique. Il en résulte que les photons occupent alors un volume qui s'étend bien au-delà de la zone active. On peut donc facilement perdre en recouvrement porteurs-photons ce que l'on gagne en confinement de porteurs. Le résultat net est fonction de ce recouvrement. Ce dernier est mesuré par un paramètre appelé *facteur de confinement* des photons.

On chiffre l'efficacité du recouvrement porteurs-photons par le *facteur de confinement* Γ, qui mesure la proportion de la densité de rayonnement effectivement en interaction avec le milieu actif. Si L est la largeur du puits de confinement des porteurs, le facteur de confinement des photons s'écrit

$$\Gamma = \int_{-L/2}^{L/2} E^2(z)dz \Big/ \int_{-\infty}^{+\infty} E^2(z)dz \qquad (12\text{-}113)$$

où $E(z)$ représente la variation suivant z de l'amplitude du champ électrique du rayonnement.

La condition nécessaire au régime d'oscillation de la cavité laser, c'est-à-dire la condition de seuil (Eq. 9-160), s'écrit alors

$$\Gamma g \geq \alpha + \frac{1}{L} Ln\left(\frac{1}{R}\right) \qquad (12\text{-}114)$$

où g est le gain de la zone active, α le coefficient de pertes par absorption, et $1/L\ Ln(1/R)$ les pertes par émission à l'extérieur de la cavité. Le paramètre Γg est appelé *gain modal*.

L'expression (12-114) peut être généralisée pour tenir compte d'une part du fait que le coefficient de réflexion peut être différent d'une face à l'autre de la cavité, et d'autre part du fait que le coefficient d'absorption α n'est pas le même à l'intérieur et à l'extérieur de la zone active

$$\Gamma g \geq \Gamma \alpha_i + (1-\Gamma)\alpha_e + \frac{1}{2L} Ln\left(\frac{1}{R_1 R_2}\right) \qquad (12\text{-}115)$$

Pour une diode conventionnelle avec $\alpha \approx 10$ cm^{-1}, $L \approx 300$ µm et $R \approx 0{,}3$, le gain modal de seuil Γg est de l'ordre de 50 cm^{-1}.

Le calcul du facteur de confinement Γ n'est généralement pas très facile et nécessite une approche numérique. Toutefois une expression analytique simple permet d'obtenir Γ avec une très bonne approximation (Botez D.-1981)

$$\Gamma = \frac{D^2}{2 + D^2} \qquad (12\text{-}116)$$

D représente l'épaisseur normalisée de la zone active, donnée par

$$D = \frac{2\pi}{\lambda} (n_i^2 - n_e^2)^{1/2} d$$

où λ est la longueur d'onde du rayonnement dans le vide et d l'épaisseur de la zone active. n_i et n_e sont les indices de réfraction respectivement à l'intérieur et à l'extérieur de la zone active.

Dans une double hétérostructure de type $Al_{0,5}Ga_{0,5}As/GaAs$ dont la zone active est de 0,1 µm (Fig. 12-17-a), l'expression (12-116) donne $\Gamma \approx 0{,}4$. Ce qui nécessite, pour avoir un gain modal de 50 cm^{-1}, un gain volumique net $g \approx 10^2$ cm^{-1}.

Dans une structure à simple puits quantique, la largeur d de la zone active est nettement inférieure à la longueur d'onde λ du rayonnement. L'épaisseur normalisée D du guide d'onde est alors très inférieure à 1. L'expression (12-116) se réduit alors à $\Gamma \approx D^2/2$, soit

$$\Gamma_{sp} \approx \frac{2\pi^2}{\lambda^2} (n_i^2 - n_e^2) d^2 \qquad (12\text{-}117)$$

Pour un puits quantique de type $Al_{0.5}Ga_{0.5}As/GaAs/Al_{0.5}Ga_{0.5}As$ d'épaisseur d=250 Å (Fig. 12-17-b), on obtient $\Gamma \approx 0{,}04$. Ce qui nécessite un gain volumique net $g \approx 10^3$ cm^{-1}. Le gain de seuil, et par suite le courant de seuil, sont donc d'un ordre de grandeur supérieurs à ceux d'une double hétérostructure.

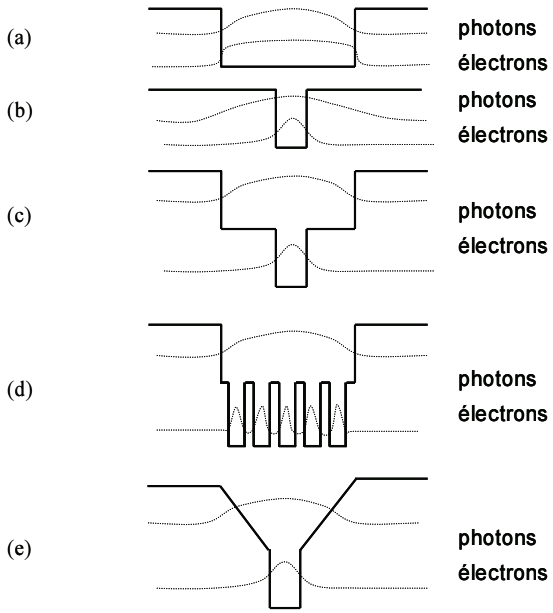


Figure 12-17 : Effets de confinement

Afin d'augmenter le facteur de confinement, on réalise des structures à confinements séparés SCH (Separate Confinement Heterojonction). Les porteurs sont confinés dans un puits quantique d'épaisseur d_p ,les photons sont confinés dans un guide d'onde de largeur d_{ph} (Fig. 12-17-c).

On améliore encore le facteur de confinement en remplaçant le puits unique par une structure à multipuits quantiques (Fig. 12-17-d). Si on appelle N_p et N_b les nombres de puits et de barrières, d_p et d_b leurs épaisseurs respectives, et enfin n_p et n_b les indices respectifs des matériaux puits et barrière, on peut définir l'indice moyen de la région d'épaisseur $\overline{d} = N_p d_p + N_b d_b$, contenant la structure de multipuits, par la relation (Streifer W.-1979)

$$\overline{n} = \frac{N_p d_p n_p + N_b d_b n_b}{N_p d_p + N_b d_b} \tag{12-118}$$

Le facteur de confinement $\overline{\Gamma}$ du mode optique dans la région d'épaisseur $\overline{d}$ couvrant l'ensemble des multipuits, est alors donné par une expression semblable à l'expression (12-116)

$$\overline{\Gamma} = \frac{\overline{D}^2}{2 + \overline{D}^2} \tag{12-119}$$

où $$\overline{D} = \frac{2\pi}{\lambda} (\overline{n}^2 - n_e^2)^{1/2} \ \overline{d}$$

Le facteur de confinement dans les véritables régions actives, que constitue chacun des puits de la structure, est obtenu en multipliant $\overline{\Gamma}$ par le rapport de la largeur active réelle $N_d d_p$ à la largeur totale $\overline{d}$ du système de multipuits

$$\Gamma_{mp} = \overline{\Gamma} \frac{N_p d_p}{N_p d_p + N_b d_b} \tag{12-120}$$

En première approximation Γ_{mp} est de l'ordre de N_p fois le facteur de confinement du simple puits quantique Γ_{sp}.

Par rapport au puits unique, les multipuits quantiques augmentent certes le facteur de confinement Γ, mais en parallèle diminuent le gain g, notamment à faible injection. En effet, les porteurs sont alors distribués dans N_p puits au lieu d'un seul. Il en résulte qu'à injections comparables, la densité de porteurs dans la zone active est N_p fois plus faible dans la structure à multipuits quantiques que dans la structure à puits quantique unique. Il est par conséquent souhaitable de conserver le puits unique en améliorant toutefois le facteur de confinement. On y parvient en remplaçant le guide optique à saut d'indice par un guide à gradient d'indice (Fig. 12-17-e). La composition, et par suite l'indice et le gap, de l'alliage constituant le guide optique, varient graduellement. Ces structures portent le nom de structures GRIN-SCH (GRaded INdex-SCH).

L'augmentation progressive de la hauteur des barrières de potentiel, dans le guide optique, a deux effets bénéfiques liés au confinement des porteurs. Le premier, associé aux fonctions d'ondes, réduit l'étalement des porteurs à l'extérieur de la zone active. Le deuxième effet est lié à la densité d'états. Dans le guide optique à saut d'indice la largeur du guide est telle que les électrons ont, dans ce guide, un mouvement tridimensionnel avec une densité d'états importante au fond du guide, c'est-à-dire à proximité immédiate du puits quantique. Dans le guide à gradient de composition, l'augmentation progressive de la largeur du puits de potentiel entraîne un confinement électronique qui réduit considérablement la densité d'états électroniques au fond du guide optique. Il en résulte que les porteurs restent plus facilement dans le puits. Cet effet réduit en particulier la sensibilité à la température.

12.3.2. *Effets quantiques*

a. Gain

Dans une structure à puits quantique la nature bidimensionnelle de la densité d'états modifie la courbe de gain (Weisbuch C. 1991-1995).

A trois dimensions, la densité d'états augmente avec l'énergie comme $\sqrt{E}$. Il en résulte que lorsque l'injection augmente le maximum de la courbe de gain se déplace vers les hautes énergies (Fig. 12-18-a). Par contre dans un puits quantique, la densité d'états est constante dans chacune des sous-bandes. Ainsi lorsque le pseudo-niveau de Fermi s'élève sous l'effet de l'injection, le sommet de la courbe de gain reste fixé à l'énergie du bas de la sous-bande (Fig. 12-18-b). Tant que la première sous-bande est la seule concernée, le sommet de la courbe de gain reste donc fixe. Il en résulte que l'augmentation de l'injection augmente plus efficacement le gain pour des énergies correspondant au bas de la sous-bande.

La contre partie de la nature bidimensionnelle de la densité d'états est que le gain de la région active sature lorsque les premières sous-bandes des électrons et des trous sont totalement inversées. Si le gain alors obtenu est insuffisant pour compenser les pertes, le seuil d'oscillation n'est pas atteint et les deuxièmes sous-bandes doivent être mises à contribution. La raie d'émission du laser correspond alors à la transition e_2-h_2 entre les deuxièmes sous-bandes des électrons et des trous.

b. Longueur d'onde d'émission

Les propriétés de symétrie des fonctions d'onde électroniques dans un puits quantique entraînent l'existence de règles de sélection sur les transitions optiques. Ces règles de sélection, qui sont strictes dans un puits de hauteur infinie, sont amenuisées dans un puits de hauteur finie par la sortie des porteurs dans les barrières. Néammoins même dans un puits de hauteur finie, l'intégrale de recouvrement des fonctions enveloppes des électrons et des trous ne présente de valeur appréciable que pour $\Delta n=0$. $\Delta n=n_e-n_h$ où n_e et n_h sont respectivement les nombres quantiques des sous-bandes d'électrons et de trous.

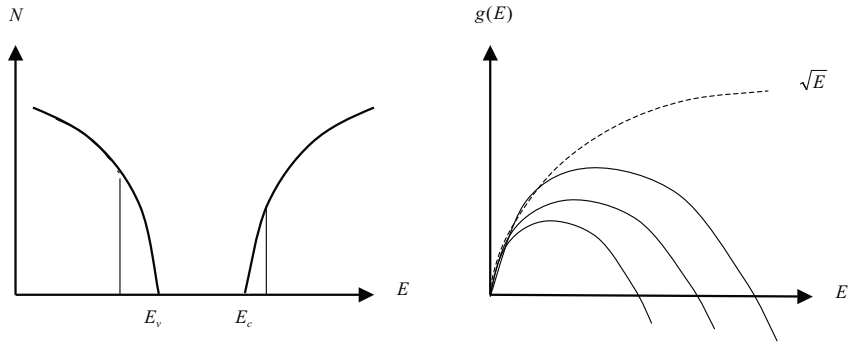
Si l'inversion de population des sous-bandes fondamentales d'électrons et de trous e_1 et h_1, permet de créer un gain supérieur aux pertes, c'est-à-dire d'atteindre le seuil d'oscillations de la cavité, la raie d'émission du laser est donnée par

$$\hbar\omega=E_g+E_{e1}+E_{h1} \qquad (12\text{-}121)$$

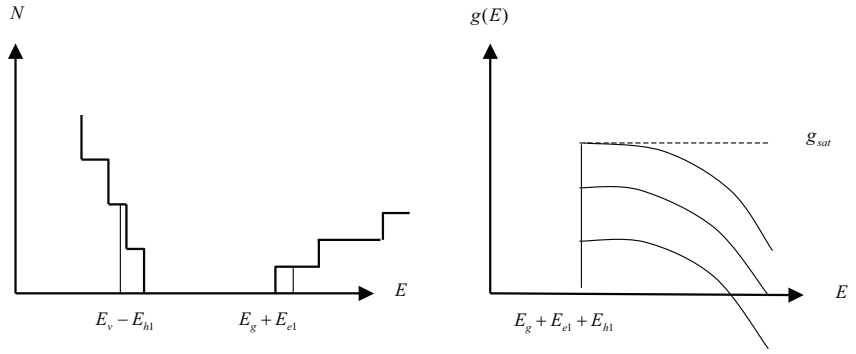
où E_g est le gap du matériau constituant le puits de potentiel et E_{e1}, E_{h1} les énergies de confinement des électrons et des trous dans ce puits.

Il est donc facile de maîtriser, dans une certaine gamme, la longueur d'onde d'émission du laser en jouant sur la largeur du puits. Dans une double

hétérostructure cette possibilité n'existe pas, seule la composition des matériaux permet d'ajuster la longueur d'onde.



a) Structure tridimensionnelle



b) Structure bidimensionnelle

Figure 12-18 : courbes de gain

c. Sensibilité à la température

La nature bidimensionnelle de la densité d'états entraîne que les pseudo-niveaux de Fermi des porteurs varient peu avec la température. Il en résulte que le courant de seuil d'un laser à puits quantique est beaucoup moins sensible à la température que celui d'une double hétérostructure.

La variation du courant de seuil avec la température peut être décrite par l'expression simple suivante (De Cremoux B-1987)

$$I_{s2} = I_{s1}\, e^{(T_2 - T_1)/T_o} \qquad (12\text{-}122)$$

où I_{s1} et I_{s2} sont respectivement les valeurs du courant de seuil aux températures T_1 et T_2 . T_0 est un paramètre qui caractérise la sensibilité du courant de seuil.

L'augmentation de la température entraîne un étalement énergétique de la distribution des porteurs. Comme le gain du seuil d'oscillation dépend de la densité de porteurs par unité d'énergie, le courant de seuil augmente avec la température. Dans un laser à puits quantique, la densité d'états constante réduit cet étalement et par suite rend le courant de seuil moins sensible à la température.

12.3.3. *Laser à cavité verticale-VCSEL*

a. Structure

La structure du laser à cavité verticale VCSEL (Vertical Cavity Surface Emitting Laser) est représentée sur la figure (12-19). Les faces réfléchissantes de la cavité ne sont plus des faces clivées perpendiculaires au plan de la jonction, comme dans une diode laser à émission latérale, mais sont réalisées par des miroirs de Bragg, DBR (Distributed Bragg Reflector), situés au-dessus et au-dessous de la région active. Le rayonnement est émis perpendiculairement aux couches épitaxiées, dans un faisceau peu divergent (~ 8°), de section circulaire.

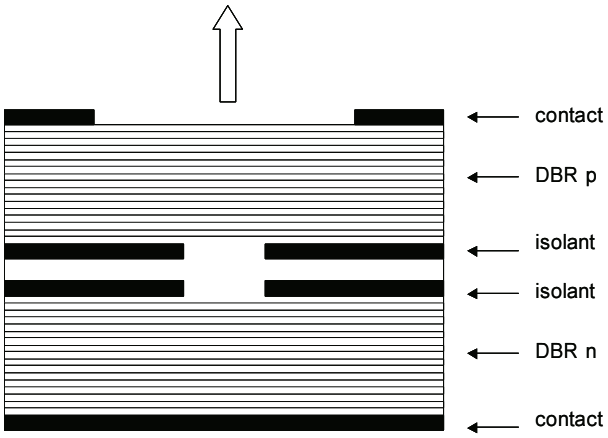


Figure 12-19 : Structure du laser à cavité verticale - VCSEL

Ce type de structure permet de réaliser un volume actif minimum car la longueur de la cavité Perot-Fabry, qui est ici conditionnée par l'épaisseur des couches, peut être réalisée de l'ordre de grandeur de la longueur d'onde du rayonnement. C'est donc potentiellement une structure à faible seuil. Mais la faible longueur de la cavité entraîne une augmentation de ses pertes. Il est donc nécessaire, pour compenser cet effet, de réaliser des faces réfléchissantes de fort pouvoir réflecteur, ce qui impose la réalisation de miroirs de Bragg.

La structure ne comporte pas moins d'une centaine de couches, mais ne nécessite que des processus standards de technologie de fabrication de circuits intégrés. A titre d'exemple de la complexité du processus d'épitaxie, le laser visible VCSEL utilise une cavité résonnante à base de phosphore de type AlInP/GaInP avec des miroirs de Bragg à base d'arsenic de type AlGaAs/GaAs, ce qui implique

un changement total de l'élément du groupe V pendant le processus d'épitaxie. Malgré le nombre important de couches, la technologie de fabrication est en fait plus simple que celle des lasers à émission latérale qui nécessitent la réalisation de faces clivées.

Notons que la géométrie verticale simplifie le couplage avec les fibres optiques, puisque le rayonnement est émis perpendiculairement à la surface d'épitaxie. En outre, cette géométrie permet la réalisation de réseaux à deux dimensions.

b.Gain de seuil

Pour calculer le gain de seuil du laser VCSEL, il suffit de reprendre la procédure schématisée sur la figure (9-39). Elle consiste à écrire que le gain de boucle de la cavité est égal à 1, mais il faut ici tenir compte des spécificités de la structure. La première spécificité, commune à tous les lasers à puits quantiques, est liée au fait que le recouvrement entre la zone active et le rayonnement n'est pas total. Ce recouvrement est caractérisé par le facteur de confinement Γ. La seconde, typique du laser VCSEL, est que, contrairement aux lasers à émission latérale, le milieu n'est pas amplificateur sur toute la longueur L de la cavité. Ce milieu n'est actif que sur une longueur effective $L_{eff}=N_p d_p$ où N_p et d_p représentent respectivement le nombre de puits quantiques et leur largeur. Ainsi l'équation (9-158), caractéristique du laser à émission latérale, s'écrit pour le laser VCSEL

$$R_1 R_2 \, e^{2\left(\Gamma g L_{eff} - \alpha_i L_{eff} - \alpha_e (L - L_{eff})\right)} \geq 1 \qquad \text{(12-123)}$$

où α_i et α_e représentent respectivement les pertes à l'intérieur et à l'extérieur des zones actives. En supposant $\alpha_i \approx \alpha_e = \alpha$ et en explicitant la longueur effective L_{eff}, le gain de seuil s'écrit donc

$$\Gamma g_o N_p d_p \approx \alpha L + \frac{1}{2} Ln\left(\frac{1}{R_1 R_2}\right) \qquad \text{(12-124)}$$

Dans la mesure où les coefficients de réflexion sur les miroirs de Bragg sont voisins de 1, le développement limité du logarithme permet d'écrire l'expression sous la forme

$$\Gamma g_o \approx \alpha \frac{L}{N_p d_p} + \frac{1}{2 N_p d_p}(1 - R_1 R_2) \qquad \text{(12-125)}$$

Le facteur de confinement Γ dépend de la position des puits quantiques par rapport à la distribution de l'onde stationnaire dans la cavité optique. Les puits situés sur les nœuds de vibration participent peu à Γ, alors que les puits situés sur les ventres de vibration ont une participation maximum. On optimise donc Γ en réalisant un milieu à *gain périodique résonnant* RPG (Resonant Periodic Gain). La

technique consiste à réaliser une structure à multipuits quantiques avec des puits étroits, distants les uns des autres d'une demi-longueur d'onde optique. En localisant les puits sur les ventres de vibration de l'onde optique stationnaire on optimise Γ. La structure RPG est réalisable dans les lasers VCSEL en raison de la faible valeur de la longueur L de la cavité optique. Cette longueur permet d'abord de ne sélectionner qu'un seul mode optique dans la largeur spectrale du gain. Ensuite, elle entraîne une période des ondes stationnaires suffisamment importante pour autoriser la réalisation de multipuits quantiques accordés.

Considérons le cas le plus simple d'un laser VCSEL avec un seul puits de largeur d_p=100 Å et supposons $\Gamma \approx 1$ et $\alpha \approx 0$. L'expression (12-125) montre que si on veut limiter le gain de seuil à g_0=1000 cm^{-1}, les pouvoirs réflecteurs des faces de la cavité doivent être R_1R_2=0,998. Si les deux miroirs sont identiques, cela donne R_1=R_2=R=99,9%. Des valeurs aussi élevées de pouvoir réflecteur ne peuvent être obtenues avec des miroirs métalliques. Il est nécessaire d'utiliser des miroirs de Bragg à faibles pertes diélectriques.

b. Miroir de Bragg – DBR

Dans une diode laser à émission latérale, la cavité optique est un Perot-Fabry horizontal dont la longueur est typiquement de quelques centaines de microns (L≈200 μm). Cette valeur étant très grande devant la longueur d'onde du rayonnement, l'intervalle entre les modes de la cavité est très faible $(\Delta k = 2\pi / L)$. Il en résulte qu'un nombre important de modes sont en recouvrement avec la courbe spectrale de gain, et peuvent de ce fait créer de l'émission stimulée. En d'autres termes, la sélectivité de la cavité est relativement faible. Quand l'excitation du laser augmente, plusieurs modes, qui se situent au voisinage du sommet de la courbe de gain, entrent très vite en résonance. La réalisation de cavités verticales permet de réduire considérablement la longueur du Perot-Fabry, et par suite d'améliorer d'autant la sélectivité. Il en résulte une meilleure monochromaticité du laser.

Lorsqu'un matériau présente une périodicité d dans l'espace réel, une structure de bandes apparaît dans l'espace réciproque, avec des bords de bandes correspondant aux valeurs π/d, $2\pi/d$,... du vecteur d'onde. Lorsque $k=\pi/d$ l'onde subit une réflexion de Bragg, lorsque $k\neq\pi/d$ l'onde n'est pas réfléchie. Si la périodicité concerne le potentiel électrique, les êtres physiques concernés sont les porteurs de charge, le résultat est la structure de bandes d'énergie électronique des matériaux cristallins. Si la périodicité concerne l'indice de réfraction, les êtres concernés sont les photons, la structure est alors un miroir de Bragg présentant une bande permise pour la réflexion de l'onde électromagnétique.

Le miroir de Bragg consiste donc en une succession de couches de matériaux à haut et faible indice de réfraction, avec un chemin optique dans chaque couche égal au quart de la longueur d'onde, $n_1d_1=n_2d_2=\lambda/4$. Si le rapport des indices entre les deux matériaux est important, on obtient un fort pouvoir réflecteur avec seulement quelques périodes. C'est le cas du couple hybride semiconducteur-isolant

ZnSe-CaF_2 , pour lequel $n_2/n_1 \approx 1,7$. L'inconvénient de ce type de miroir, dans une structure laser, est que les couches isolantes interdisent l'injection du courant. Les miroirs monolithiques de type SC_1-SC_2, permettent l'injection de courant à travers le miroir, mais présentent un rapport d'indice plus faible ($n_2/n_1 \approx 1,2$), ce qui nécessite un nombre de périodes de l'ordre de 20.

Les performances du miroir sont conditionnées par le nombre de périodes et la maîtrise des épaisseurs et des indices des différentes couches. La figure (12-20) représente l'effet du nombre de périodes sur le pouvoir réflecteur d'un miroir de Bragg idéal centré à λ=1,55 µm, élaboré avec des semiconducteurs présentant un rapport d'indice typique de 1,2. Les effets d'un désaccord de 10% sur le trajet optique dans un type de couche sont représentés, pour un miroir de 20 périodes, sur la figure (12-21). Ce désaccord peut résulter d'un désaccord d'indice (Fig. 12-21-b) et/ou d'un désaccord d'épaisseur (Fig. 12-21-c). On constate à la fois une diminution et un décalage du maximum de réflectivité. Dans la pratique la méconnaissance des indices de réfraction des alliages semiconducteurs nécessite souvent une approche empirique. Néanmoins on atteint assez facilement des pouvoirs réflecteur supérieurs à 99%.

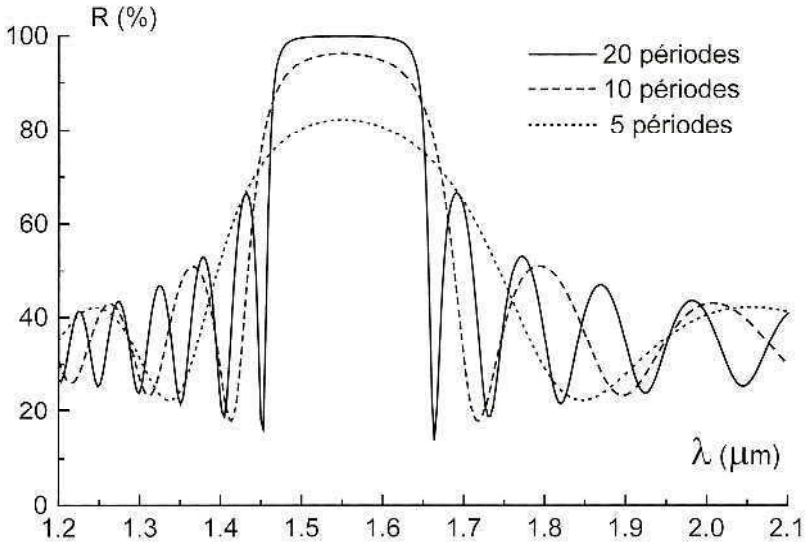


Figure 12-20 : Pouvoir réflecteur d'un miroir de Bragg idéal, centré à λ = 1,55 µm, en fonction du nombre de périodes *(Malzac J.P. - 1996)*

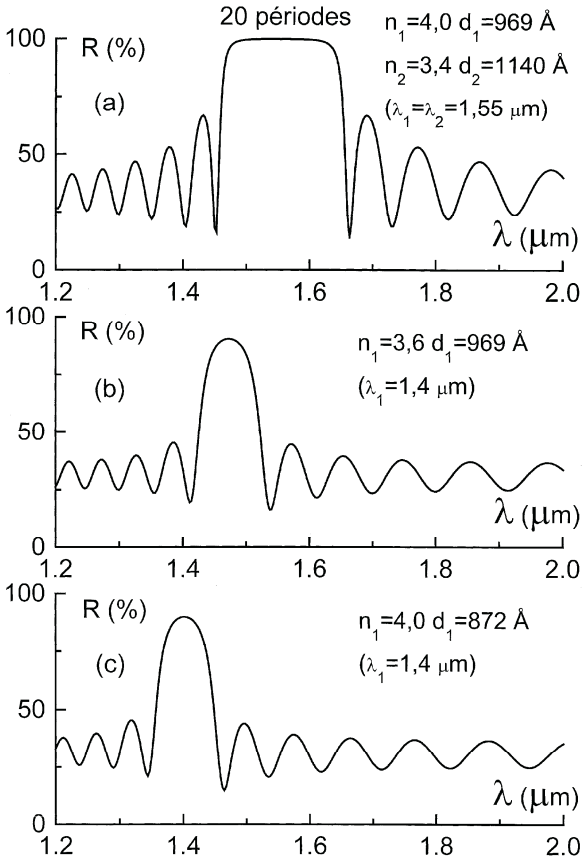


Figure 12-21 : Pouvoir réflecteur d'un miroir de Bragg

c. Performances

La région active du laser, qui contient un ou plusieurs puits quantiques, est située au centre de la cavité optique afin d'optimiser le facteur de confinement. La faible épaisseur de la cavité optique entraîne une grande séparation des modes de résonance de sorte que généralement un seul mode est en recouvrement avec la courbe de gain (Fig. 12-22). Les performances du laser sont de ce fait conditionnées par la position spectrale de ce mode par rapport au sommet de la courbe de gain. Or, sous l'action de la température liée à

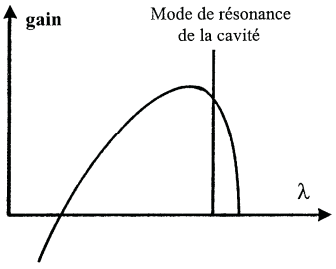


Figure 12-22 : Courbe de gain

l'injection, la courbe de gain se déplace vers les grandes longueurs d'onde plus vite que le mode de résonance de la cavité. Pour compenser cet effet le mode de la cavité est alors volontairement décalé vers les grandes longueurs d'onde par rapport au sommet de la courbe de gain (Fig. 12-22). Ainsi, aux températures de fonctionnement ce mode résonne avec le maximum de la courbe.

La caractéristique essentielle du laser VCSEL est le faible volume de la zone active. C'est donc potentiellement une structure à très faible seuil. Des lasers VCSEL avec des courants de seuil inférieurs à 10 μA, des tensions de seuil inférieures à 1,3 V et des rendements supérieurs à 50% ont été réalisés. En outre, la combinaison d'un faible volume actif avec une grande densité de photons permet une grande vitesse de modulation, des fréquences de modulation supérieures à 15 GHz ont été obtenues.

La contre partie du faible volume actif est que la puissance émise par un laser VCSEL est inférieure à celle d'un laser à émission latérale. Néanmoins des puissances d'émission de 100 mW avec un laser unique, et 1 W avec un réseau bidimensionnel, ont été obtenues sous excitation pulsée.

12.3.4. *Microcavité photonique - Laser sans seuil*

Dans la structure représentée sur la figure (12-19), la cavité optique est calculée pour sélectionner un seul mode de résonance. Mais ce mode est tridimensionnel dans la mesure où il est étalé sur toute la section de la cavité. Si les dimensions latérales de la structure sont réduites, à des valeurs comparables à la longueur d'onde du rayonnement, on réalise une *microcavité* optique, qui confine les photons au même titre qu'une boîte quantique confine les électrons. Le mode de résonance de la cavité, sélectionné longitudinalement, est alors discrètement espacé transversalement. Le confinement optique associé à la microcavité entraîne une double redistribution, à la fois spectrale et spatiale, de la densité optique. La redistribution spatiale en particulier, qui devient non uniforme, favorise la contribution de l'émission spontanée au mode de résonance de la cavité.

On décrit l'effet de microcavité par le *facteur d'émission spontanée* γ, qui mesure le rapport de l'angle solide soutendu par le mode de résonance de la cavité, à l'angle solide, couvrant tout l'espace, dans lequel est émis le rayonnement spontané, $\gamma = \Omega/4\pi$. Dans un laser conventionnel (Par.9.3.2.e), γ est très petit (10^{-6} à 10^{-5}), la majeure partie du rayonnement spontané sort de la cavité sans participer au mode de résonance. La contribution de l'émission spontanée au mode de la cavité est de ce fait très faible, et la densité de photons sur ce mode n'augmente vraiment qu'en régime d'émission stimulée. Le seuil d'émission stimulée apparaît alors très clairement sur la caractéristique rayonnement-courant du laser. Quand γ augmente, le poids d'émission spontanée dans le mode de la cavité augmente, et le seuil d'émission stimulée devient de moins en moins marqué. A la limite, si l'unique mode de résonance occupe tout le volume actif de la cavité, tout photon émis (qu'il soit stimulé ou spontané) est alors émis sur ce mode. Dans ce cas $\gamma = 1$ et la caractéristique rayonnement-courant ne présente aucun seuil de superlinéarité. Le laser est alors qualifié de *laser sans seuil*. Sa caractéristique reste partout linéaire,

l'émission change simplement de nature lorsque le régime d'inversion de population est atteint. Le rayonnement, spontané à faible injection, devient stimulé à forte injection, il est toujours émis sur l'unique mode de la cavité :

- A faible injection, la durée de vie τ_n des porteurs est constante, et la densité Δn de porteurs excédentaires augmente linéairement avec le courant J. Le taux d'émission $\Delta n(J)/\tau_n$, qui est alors spontané , augmente linéairemant avec le courant.

- Quand l'injection atteint et dépasse le seuil d'inversion de population, l'émission devient stimulée. La durée de vie des porteurs diminue alors proportionnellement à la densité de photons. La densité de porteurs excédentaires atteint un régime de saturation Δn_s, et le taux d'émission $\Delta n_s/\tau_n(J)$, continue d'augmenter linéairement. Aucun seuil ne se manifeste sur la caractéristique rayonnement-courant. Simplement, le rayonnement émis sur l'unique mode de résonance de la cavité change de nature, de spontané il devient stimulé.

En fait, ce processus reste idéal, et l'apparition du régime d'émission stimulée se manifeste toujours par un changement de pente sur la caractéristique rayonnement-courant. Ceci peut simplement résulter de la réabsorption partielle du rayonnement spontané par transitions interbandes.

12.3.5. *Laser à cascade quantique*

Dans le laser à jonction *pn* l'amplification optique résulte de recombinaisons radiatives entre les électrons de la bande de conduction et les trous de la bande de valence, on pourrait à cet égard le qualifier de *bipolaire*. Deux des caractéristiques de ce système sont d'une part que la longueur d'onde d'émission est déterminée par le gap du matériau et d'autre part que le spectre de gain est relativement large car l'inversion de population est distribuée entre deux bandes ayant des courbures de signes opposés, la bande de conduction et la bande de valence.

Le laser à cascade quantique diffère fondamentalement du laser à jonction en ce sens qu'il met en jeu un seul type de porteur, on pourrait de ce fait le qualifier d'*unipolaire*. Il exploite non pas les transitions interbandes comme le précédent mais les transitions électroniques entre sous-bandes résultant de la quantification de la bande de conduction dans une hétérostructure. La longueur d'onde d'émission n'est plus déterminée par le gap du matériau mais par le confinement quantique. En outre sa courbe de gain a un spectre étroit car la densité conjointe d'états présente l'allure d'une fonction de Dirac du fait que les sous-bandes mises en jeu ont des courbures de même signe. Ces sous-bandes sont parallèles contrairement aux bandes de conduction et de valence qui sont antiparallèles.

Le principe de fonctionnement du laser à cascade quantique est illustré sur la figure (12-23) qui représente le diagramme énergétique du minimum de la bande de conduction de la structure sous polarisation. La structure est périodique, chaque étage étant constitué d'une région d'injection à gap graduel et d'une région active.

La région d'injection est réalisée avec un pseudo-alliage ou un superréseau de composition variable dans lequel le gap graduel augmente de la gauche vers la

droite de la figure. En l'absence de polarisation la bande de conduction de cette région augmente donc de la gauche vers la droite de sorte que le diagramme énergétique présente une enveloppe en dent de scie. Lorsque la structure est polarisée par une tension de polarité et d'amplitude convenables le pseudo-champ électrique interne associé au gradient de gap est compensé par le champ appliqué et le diagramme énergétique présente l'allure de marches d'escalier de la figure (12-23). La bande de conduction de la région d'injection, représentée en pointillés sur la figure, est horizontale.

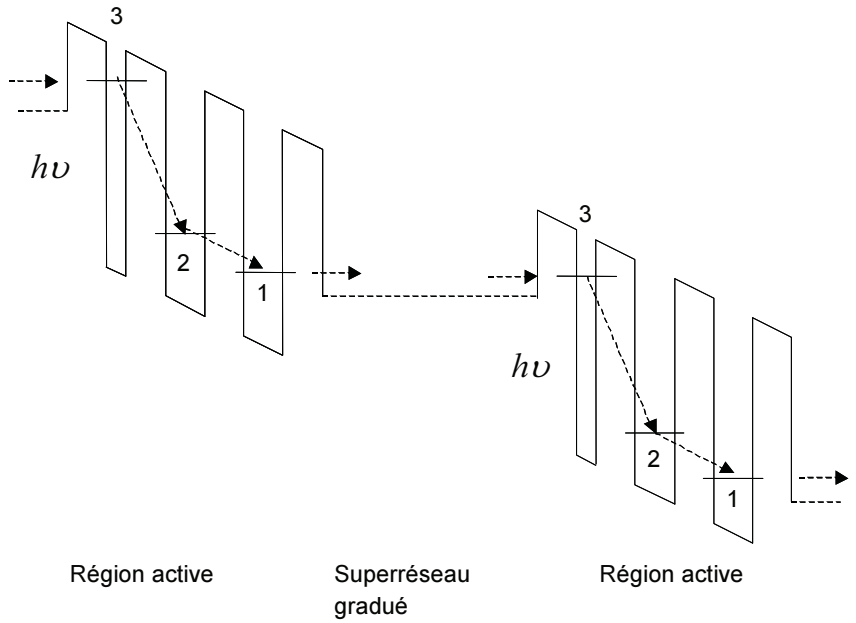


Figure 12-23 : Diagramme énergétique de la bande de conduction d'un étage.

La région active est constituée de puits quantiques couplés, elle se comporte comme un laser à 4 niveaux dans lequel l'inversion de population est réalisée entre les deux sous-bandes excitées n=3 et n=2. La sous-bande 3 est peuplée par injection tunnel depuis la région à gap graduel qui constitue le niveau 4. La sous-bande 2 est vidée par relaxation inélastique, assistée par phonons optiques, dans la sous-bande 1 qui se vide elle-même par effet tunnel dans la région à gap graduel adjacente. L'amplification optique résulte de transitions radiatives entre les sous-bandes 3 et 2. Le motif étant répété un certain nombre de fois dans la structure, un même électron injecté à l'entrée peut créer plusieurs photons en cascade, le rendement quantique peut de ce fait être supérieur à 1.

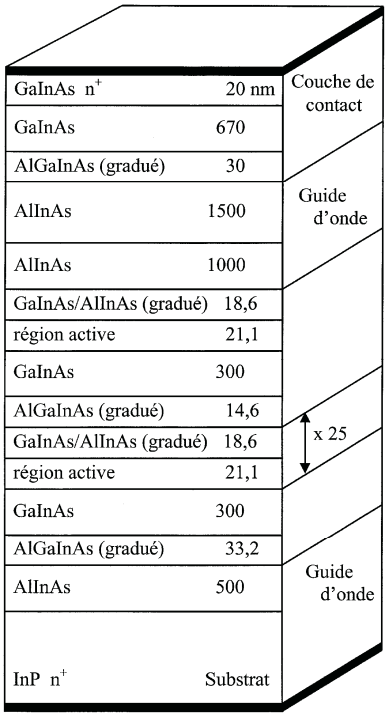


Figure 12-24 : Laser à cascade quantique

Un exemple de réalisation d'une telle structure est représenté sur la figure (12-24). Elle est élaborée par MBE en épitaxie pseudomorphique d'hétérostructures de type AlInAs/GaInAs en accord de maille sur un substrat de InP (Faist J.-1994). La structure comporte 25 étages avec un total de 500 couches. Chaque étage consiste en une région active constituée de 3 puits quantiques couplés non dopés et d'une région d'injection à gap graduel constituée d'un superréseau de type GaInAs/AlInAs dopé *n* au silicium afin de minimiser les effets de charge d'espace associés à l'injection. Les horizontales en pointillés sur la figure (12-23) représentent le bas de la bande de conduction du superréseau. Les électrons sont injectés dans la sous-bande 3 de la région active à travers une barrière de 4,5 nm de AlInAs. La région active est constituée de 3 puits de GaInAs de 0,8 , 3,5 et 2,8 nm séparés par des barrières de AlInAs de 3,5 et 3 nm d'épaisseur. Les électrons sortent de la région active par effet tunnel à travers une barrière de AlInAs de 3 nm. Les distances énergétiques des sous-bandes 3 et 2 d'une part et 2 et 1 d'autre part sont respectivement de 295 et 30 meV. La valeur de 295 meV fixe la longueur d'onde d'émission. La valeur de 30 meV est en résonance avec le phonon optique de GaAs et assure un vidage efficace du niveau 2 dans le niveau 1.

Des couches additionnelles de part et d'autre de la séquence des 25 étages constituent un guide d'onde qui confine le rayonnement dans la région active, le facteur de confinement Γ est de l'ordre de 0,5. La cavité résonnante de longueur L=500 μm est réalisée par clivage perpendiculaire au plan des couches.

A basse température le laser a une densité de courant de seuil de l'ordre de 15000 A/cm^2 et émet en régime pulsé une puissance supérieure à 8 mW à une longueur d'onde λ=4,2 μm.

Ce type de laser fonctionne actuellement à la température ambiante avec des longueurs d'onde d'émission de l'ordre de 11 μm. L'intérêt de ce principe réside essentiellement dans la possibilité d'obtenir des grandes longueurs d'onde d'émission sans avoir recours à des matériaux à petit gap, qui sont d'une part

sensibles à la température et d'autre part souvent difficiles à *processer*. Ajoutons que la longueur d'onde peut facilement être maîtrisée, par les profondeur et largeur des puits et des barrières, sur une gamme relativement étendue entre le proche infrarouge et les ondes submillimétriques.

12.3.6. *Laser à boites quantiques*

L'utilisation de boîtes quantiques dans la zone active d'un laser présente plusieurs intérêts potentiels. Le premier est lié à la densité d'états, cette dernière est en effet redistribuée sur des niveaux discrets ce qui élimine la dispersion en k et augmente d'autant la sélectivité du gain. L'intégrale de la courbe de gain est comparable à celle du puits quantique mais la valeur de pic est beaucoup plus importante. Le courant de seuil d'émission stimulée est considérablement réduit, d'un facteur 10 à 100. En d'autres termes le laser à boîtes quantiques constitue la limite extrême de sophistication du laser à semiconducteur en mettant en jeu des états discrets de type atomique, toutes les recombinaisons de paires électron-trou se produisent à la même énergie. Un autre intérêt des boîtes quantiques réside dans le fait que, piégés dans des boites, les porteurs sont moins sollicités par les centres de recombinaison. Leur durée de vie non radiative s'en trouve augmentée. En outre, au moins tant que les porteurs restent piégés dans les boîtes, le fonctionnement du système et notamment le seuil d'émission stimulée, est pratiquement insensible à la température. Enfin le confinement quantique, conditionné par la taille et éventuellement la composition du matériau constituant les boites, apporte un degré de liberté supplémentaire dans le choix de la longueur d'onde d'émission.

La réalisation de boîtes quantiques est obtenue par croissance épitaxiale, par MBE ou MOCVD, d'un matériau présentant un très fort désaccord de maille avec le substrat. L'épaisseur critique de la couche est alors très faible et la croissance se développe sous forme d'îlots de forme pyramidale et de taille nanométrique. Les boites quantiques mises à profit actuellement dans la réalisation de lasers sont des îlots contraints de InAs épitaxiés sur des couches de GaAs.

Des lasers ont été élaborés sur la base de ces boîtes quantiques (Huffaker D.L. - 1998). La zone active est constituée de 11 monocouches réalisées chacune par

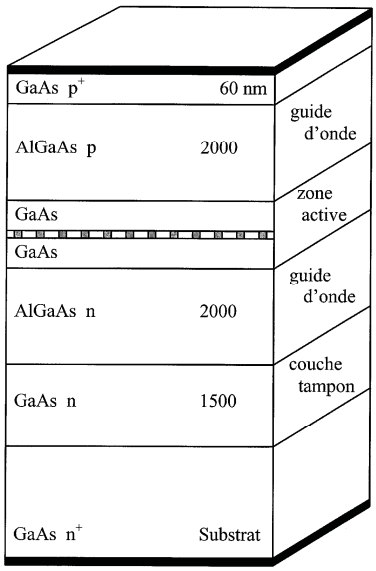


Figure 12-25 : Laser à boîtes quantiques

un dépôt de 25% d'indium, 25% de gallium et 50% d'arsenic. Chaque monocouche contient des boîtes quantiques de 50 nm de diamètre avec une densité surfacique de l'ordre de 10^{10} cm^{-2}. Cette zone active est implantée au centre d'une couche de GaAs non dopé de 0,23 μm d'épaisseur, elle même prise en sandwich entre deux couches de $Al_{0,7}Ga_{0,3}As$ de 2 μm d'épaisseur, l'une dopée *n* l'autre dopée *p*. Les taux de dopage sont relativement faibles afin de minimiser l'absorption par porteurs libres. Ces couches d'alliage jouent le rôle de guide d'onde et confinent les photons dans la zone active. Ces lasers ont un courant de seuil de l'ordre de 11 A/cm^2 à basse température et 270 A/cm^2 à la température ambiante. Le courant de seuil est pratiquement indépendant de la température jusqu'à environ 200 K et augmente exponentiellement au delà. Ce comportement correspond probablement au dépiégeage des porteurs à l'extérieur des boites de InAs, ces porteurs diffusent alors dans la couche de GaAs et se recombinent non radiativement aux interfaces AlGaAs/GaAs. La qualité de ces interfaces conditionne de ce fait le courant de seuil à la température ambiante. Ces lasers émettent un rayonnement dont la longueur d'onde est λ=1,23 μm à la température de l'azote liquide et λ=1,31 μm à la température ambiante.

Il faut enfin noter que des boites quantiques non intentionnelles jouent probablement un rôle réel dans le fonctionnement de certains lasers comme les émetteurs bleus à GaN par exemple. L'expérience montre en effet que les émetteurs à GaN ont un rendement multiplié par 10 lorsque des traces d'indium sont introduites dans la zone active. Il semble que ces traces d'indium en introduisant des îlots de InGaN, de gap inférieur à celui de GaN, jouent un rôle majeur dans le rendement quantique du dispositif.

12.4. Blocage de Coulomb et systèmes à peu d'électrons

Le blocage de Coulomb est un effet physique qui traduit l'interaction qui se manifeste entre un contenant et son contenu dans les conditions limites où la *taille* du quantum de ce dernier est commensurable avec la *capacité* du premier.

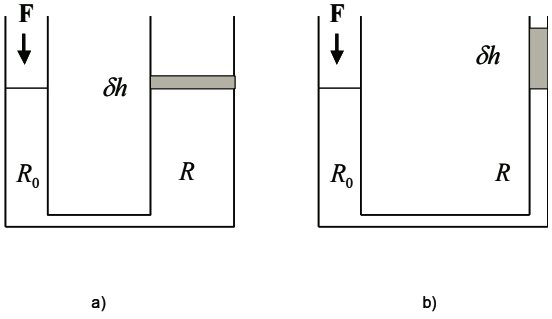


Figure 12-26 : Système hydraulique

Considérons par exemple le système hydraulique représenté sur la figure (12-26) qui consiste à alimenter un réservoir R en exerçant une force F à la surface d'un liquide contenu dans un autre réservoir R_0. Supposons que le quantum de liquide, c'est-à-dire son volume élémentaire insécable, soit la goutte. Lorsque, sous l'action de la force F une goutte de liquide passe du réservoir R_0 dans le réservoir R le niveau de liquide dans ce dernier s'élève de δh. L'énergie nécessaire pour effectuer ce transfert est $W=\rho g \delta h$ où ρ est la densité du liquide et g la constante de gravitation. Le volume de la goutte est représenté par la zone ombrée. Si la capacité du réservoir R est grande devant le volume de la goutte (Fig. 12-26-a) δh est infinitésimal de sorte que le transfert du liquide varie continûment avec W depuis W=0. Supposons maintenant que le réservoir R soit un tube capillaire de très faible section. Le passage d'une goutte de liquide entraîne une élévation δh macroscopique du niveau. L'énergie $W=\rho g \delta h$ nécessaire au passage de la goutte devient finie. Si la plus petite quantité d'eau qui peut être transférée est la goutte il existe donc un seuil d'énergie W_t qu'il faut atteindre avant que le liquide commence à être transféré. En d'autres termes le transfert est *bloqué* tant que l'énergie développée est inférieure au seuil W_t.

Considérons maintenant un condensateur de capacité C polarisé par une tension V, il est chargé par une charge $Q=CV$, rappelons que l'énergie électrostatique emmagasinée s'écrit $W = QV/2 \equiv CV^2/2 \equiv Q^2/2C$ de sorte que si la charge Q varie de dQ l'énergie électrostatique W varie de dW donné par

$$dW = \frac{Q}{C} dQ \qquad (12\text{-}126)$$

Considérons le condensateur plan de la figure (12-27-a) polarisé par une tension $V>0$, il apparaît à la surface des armatures les charges $+Q$ et $-Q$ données par $Q=CV$.

Notons que cette charge varie de façon continue depuis zéro par la simple déformation des nuages électroniques, les fonctions d'ondes électroniques se rapprochent de la surface dans l'électrode (1) et s'en éloignent dans l'électrode (2). Le diagramme énergétique est représenté sur la figure (12-27-b), le niveau de Fermi de l'électrode (2) est abaissé de $\Delta E=-eV$ par rapport à celui de l'électrode (1) soit $E_{F1} - E_{F2} = eV$.

Supposons l'épaisseur de l'isolant suffisamment faible, la capacité est par exemple une diode tunnel, la position relative des niveaux de Fermi autorise alors le passage d'un électron de l'électrode (1) vers l'électrode (2) par effet tunnel. La variation d'énergie électrostatique qui en résulte est donnée par l'intégrale de 0 à e de l'expression (12-126), soit

$$\Delta W = \int_0^e \frac{Q}{C} dQ = \frac{e^2}{2C} \qquad (12\text{-}127)$$

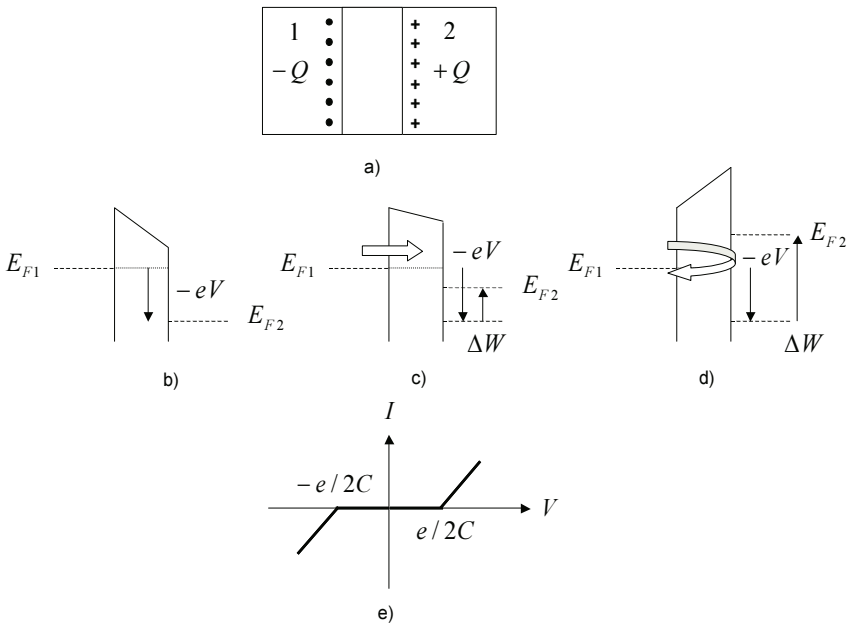


Figure 12-27 : Condensateur et blocage de Coulomb

Le niveau de Fermi E_{F2} remonte donc par rapport au niveau de Fermi E_{F1} de la quantité $\Delta W=e^2/2C$, la différence des niveaux de Fermi devient alors $E_{F1}-E_{F2} = eV-e^2/2C$.

- Si la condition $eV > \Delta W$ est réalisée la différence $E_{F1}-E_{F2}$ reste positive après le transfert, le niveau E_{F2} reste au-dessous du niveau E_{F1}, l'électron est transféré dans l'électrode (2) (Fig. 12-27-c). Il est ensuite évacué par le contact de sortie et revient dans l'électrode (1) à travers le générateur, le courant tunnel circule.

- Si par contre la condition $eV<\Delta W$ est réalisée la différence $E_{F1}-E_{F2}$ devient négative, le niveau E_{F2} se retrouve, après le transfert de l'électron, au-dessus du niveau E_{F1}, l'électron est alors refoulé dans l'électrode (1) (Fig. 12-27-d). En d'autres termes, compte tenu de l'énergie de l'état final, le transfert (1)→(2) correspondrait à un transfert tunnel entre un état *occupé* de l'électrode (1) et un état *occupé* lui aussi dans l'électrode (2), ce qui est impossible. Le courant tunnel est donc nul.

Ainsi lorsque la tension de polarisation V augmente le courant tunnel ne devient différent de zéro que lorsque la condition $eV>\Delta W$, c'est-à-dire $eV > e^2/2C$ ou $V > e/2C$ est réalisée (Fig. 12-27-e), c'est le *blocage de Coulomb*.

Le Blocage de Coulomb se manifeste lorsque la variation d'énergie capacitive $e^2/2C$ associée au passage d'un électron par effet tunnel est supérieure à l'énergie de polarisation eV. Il faut donc pour que l'effet soit

observable que les fluctuations thermiques de charge aux bornes de la capacité $<Q^2>=C\ kT$ soient négligeables devant $e^2/2$. En d'autres termes, les niveaux de Fermi E_{F1} et E_{F2} ne délimitent des frontières rigoureuses entre les états vides et occupés qu'au zéro absolu. A température finie, dans un intervalle de quelques kT autour de E_F les niveaux d'énergie sont partiellement occupés de sorte que des transferts peuvent se produire depuis l'électrode (1) vers des états situés au-dessous du niveau de Fermi dans l'électrode (2) puisque certains d'entre eux sont vides. Le blocage est donc inhibé si la quantité $e^2/2C$ n'est pas très supérieure à kT. L'observation du blocage de Coulomb est donc soumise à la condition $e^2/2C >> kT$ c'est-à-dire à une relation capacité-température de la forme

$$C << \frac{e^2}{2\,kT} \qquad (12\text{-}128)$$

En explicitant les constantes cette relation s'écrit quantitativement

$$C(Farad) << \frac{10^{-15}}{T(Kelvin)}$$

Pour conserver des valeurs de C accessibles expérimentalement il faut donc travailler à très basse température. Dans un cryostat à dilution, qui permet de descendre à des températures de quelques dizaines de mK, la capacité doit être inférieure à 10^{-14} F.

Alimentation en tension

Si la jonction tunnel est polarisée par un générateur idéal de tension (résistance interne nulle), la tension entre les deux électrodes, donc aux bornes de la résistance tunnel R_t est égale à la tension de polarisation V quoi qu'il arrive. La charge de la capacité reste donc toujours égale à $Q=CV$, qu'un électron traverse par effet tunnel ou non. La différence entre les deux niveaux de Fermi E_{F1}-E_{F2} reste en permanence imposée à la valeur eV par le générateur. Il en résulte que les électrons peuvent passer par effet tunnel dès que la condition $V\neq 0$ est vérifiée. Il n'y a pas blocage de l'effet tunnel. Ceci signifie que lorsqu'un électron passe par effet tunnel de l'électrode (1) vers l'électrode (2) il ne remonte pas de $e^2/2C$ le niveau de Fermi E_{F2} mais est instantanément recyclé par le générateur, dont la résistance interne est nulle, et revient sur l'électrode (1) par le circuit extérieur. Il n'y a pas blocage de Coulomb, la caractéristique $I(V)$ est ohmique avec une pente $1/R_t$ où R_t est la résistance tunnel de la diode (Fig. 12-29).

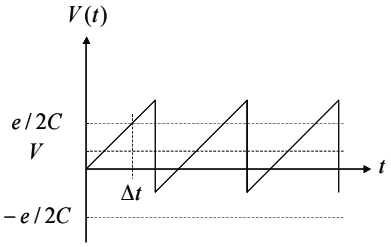


Figure 12-28 : Charge et décharge

Alimentation en courant

Si la jonction est alimentée par un générateur idéal de courant (résistance interne infinie) le courant est toujours imposé à la valeur I par le générateur. Dans un premier temps ce courant correspond à la charge de la capacité, les électrons se rapprochent de la surface de l'électrode (1) qui se charge négativement et s'éloignent de la surface de l'électrode (2) qui se charge positivement. La tension entre les deux électrodes augmente, les niveaux de Fermi s'écartent de la quantité E_{F1}-$E_{F2} = eV$. Si un électron passe par effet tunnel de l'électrode (1) vers l'électrode (2) le niveau de Fermi de l'électrode (1) descend et celui de l'électrode (2) monte, la différence E_{F1}-E_{F2} diminue de la quantité $e^2/2C$. Ainsi tant que la condition $eV < e^2/2C$ est vérifiée, c'est-à-dire $V < e/2C$, le résultat du transfert est E_{F1}-$E_{F2} < 0$, l'électron revient sur l'électrode d'origine, l'effet tunnel est bloqué, c'est le blocage de Coulomb. Le courant qui circule alors dans le circuit, qui est en permanence imposé par le générateur, est le seul courant de charge de la capacité, le courant tunnel est nul. La charge instantanée $Q(t)$ de la capacité, qui s'effectue à courant constant I, et la tension instantanée $V(t)$ qui en résulte, s'écrivent

$$Q(t) = \int I\,dt = I\int dt = It$$

$$V(t) = \frac{Q}{C} = \frac{I}{C}t \qquad (12\text{-}129)$$

La charge et la tension augmentent linéairement avec le temps (Fig. 12-28).

Lorsque la charge de la capacité atteint et dépasse la valeur $e/2$, c'est-à-dire quand la tension aux bornes de la capacité atteint et dépasse la valeur $e/2C$, un électron peut passer par effet tunnel de l'électrode (1) vers l'électrode (2), ce qui ramène à $Q(t)$-e la charge et à $V(t)$-e/C la tension, l'effet tunnel est à nouveau bloqué et le processus de charge recommence. La variation de la tension entre les électrodes est représentée sur la figure (12-28). Cette tension a une valeur moyenne V non nulle donnée par $V = \frac{1}{T}\int_0^T V(t)\,dt$ où $V(t)$ est donné par l'expression (12-129), soit

$$V = \frac{I}{C}\Delta t$$

où Δt représente la valeur moyenne du temps de transfert d'un électron par effet tunnel. Ce temps moyen Δt n'est pas constant mais diminue quand le courant de commande augmente, pour I petit ($I << e/2R_tC$) on montre que V varie comme $I^{1/2}$. Lorsque le courant devient important le blocage de Coulomb disparaît et la caractéristique devient linéaire avec une loi de la forme

$$V \approx \frac{e}{2C} + R_t I$$

Les caractéristiques de la diode alimentée en tension ou en courant sont représentées sur la figure (12-29). L'observation expérimentale du blocage de Coulomb à travers une jonction unique n'est donc possible que sous une alimentation en courant.

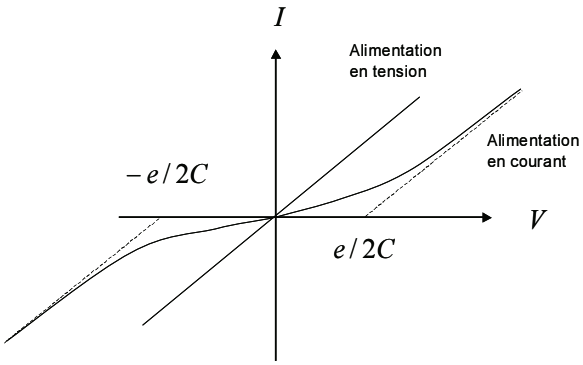


Figure 12-29 : Courbe courant tension

Dispositif à deux jonctions

Considérons le dispositif de la figure (12-30-a) constitué de deux jonctions tunnel en série, polarisé par un générateur idéal de tension. La tension aux bornes de l'ensemble, imposée à la valeur V par le générateur, se distribue entre les deux jonctions en fonction de leurs capacités et d'une éventuelle charge δQ injectée dans l'électrode centrale.

En l'absence d'effet tunnel $Q = C_1V_1 = C_2V_2$ avec $V = V_1+V_2$ de sorte que les tensions V_1 et V_2 s'écrivent

$$V_1 = \frac{C_2}{C_1 + C_2} V \qquad \text{(12-130-a)}$$

$$V_2 = \frac{C_1}{C_1 + C_2} V \qquad \text{(12-130-b)}$$

Le diagramme énergétique est représenté sur la figure (12-30-b). Compte tenu des positions relatives des niveaux de Fermi, les électrons peuvent passer de la gauche vers la droite par effet tunnel à travers chaque jonction. Si on appelle n_1 et n_2 les nombres d'électrons qui traversent respectivement les jonctions (1) et (2) la variation de charge de l'îlot central associée à ces transferts s'écrit $\delta Q = -e(n_1-n_2)$. Les tensions V_1 et V_2 dont la somme reste égale à V s'écrivent alors

$$V_1 = \frac{C_2}{C_1 + C_2} V + \frac{\delta Q}{C_1 + C_2} = \frac{C_2}{C_1 + C_2} V - e \frac{n_1 - n_2}{C_1 + C_2} \qquad \text{(12-131-a)}$$

$$V_2 = \frac{C_1}{C_1 + C_2} V - \frac{\delta Q}{C_1 + C_2} = \frac{C_1}{C_1 + C_2} V + e \frac{n_1 - n_2}{C_1 + C_2} \qquad \text{(12-131-b)}$$

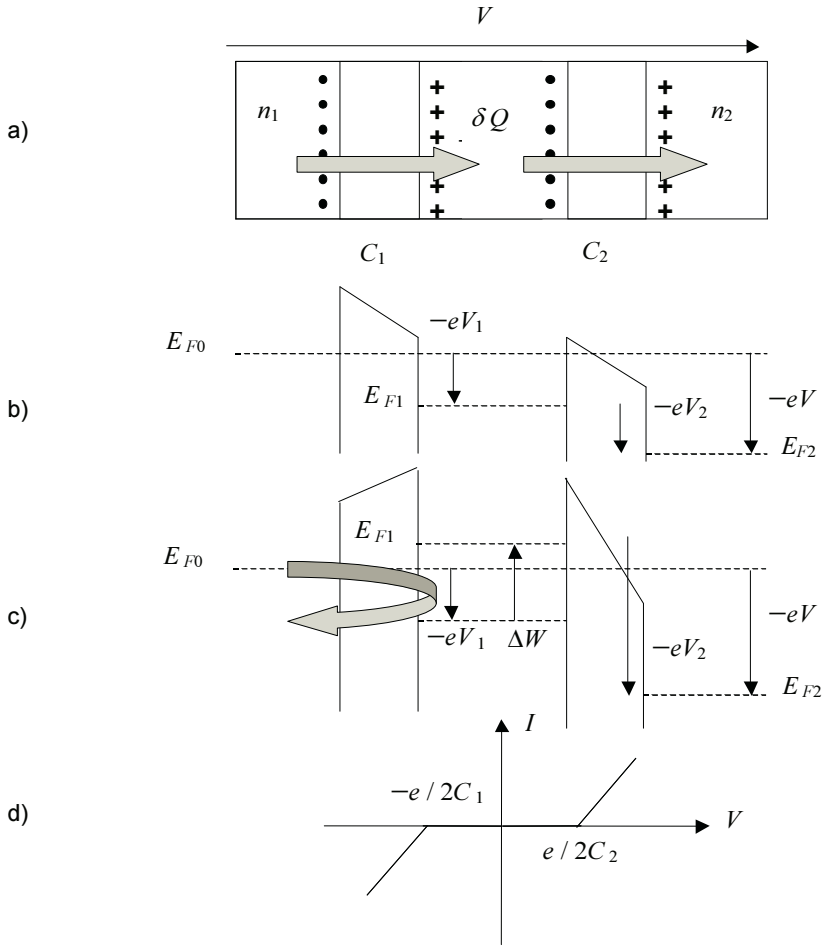


Figure 12-30 : Dispositif à deux jonctions

Le passage d'*un* électron à travers la jonction (1), c'est-à-dire le cas de figure n_1=1 , n_2=0, entraîne donc une diminution de V_1 et une augmentation de V_2 de la quantité $e/(C_1+C_2)$. Ainsi, comme dans le cas de la jonction unique le niveau de Fermi de l'îlot central remonte de $\Delta W=e^2/2(C_1+C_2)$. Si la tension V_1 , avant le

transfert, est inférieure à $e/2(C_1+C_2)$, le transfert de l'électron entraîne une remontée du niveau de Fermi de l'îlot central E_{F1} au-dessus de E_{F0} , l'électron repasse alors dans l'électrode de gauche, l'effet tunnel est inhibé (Fig. 12-30-c). La condition de déblocage du courant est donc $V_1 > e/2(C_1+C_2)$ ce qui entraîne d'après la relation (12-130-a) $V > e/2C_2$. Si la polarisation est inversée le processus se produit dans l'autre sens avec une tension de seuil égale à $e/2C_1$. La caractéristique est représentée sur la figure (12-30-d).

Ce dispositif à deux jonctions permet donc d'observer le blocage de Coulomb sous une polarisation en tension alors que le dispositif à jonction unique n'autorise son observation que sous une alimentation en courant. Il est de ce fait plus facile de mettre expérimentalement en évidence le phénomène avec ce dispositif.

Supposons maintenant le système très dissymétrique avec la résistance tunnel de la première jonction très inférieure à celle de la seconde $R_{t1} \ll R_{t2}$. En d'autres termes le taux de transfert par effet tunnel est beaucoup plus grand à la première jonction qu'à la seconde. Les électrons entrent donc dans l'îlot central plus facilement qu'ils n'en sortent, ils ont donc tendance à s'y accumuler. Mais à mesure que le nombre d'électrons dans cet îlot augmente le niveau de Fermi E_{F1} monte. L'îlot central va donc se remplir de n électrons si le $(n+1)^{\text{ème}}$ électron fait passer E_{F1} au-dessus de E_{F0}. Un nouvel électron ne pourra entrer dans l'îlot que lorsqu'un électron en sera sorti à travers la capacité C_2. Les électrons passent donc de manière corrélée à travers le dispositif qui se comporte comme une écluse laissant passer les électrons l'un après l'autre, à un rythme commandé par R_{t2}.

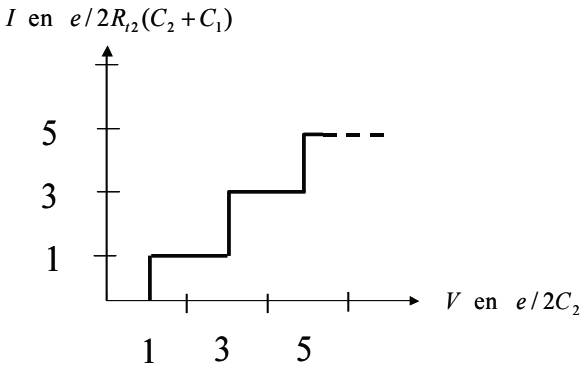


Figure 12-31 : Courant d'une double jonction

Pour augmenter le courant I il faut augmenter le nombre d'électrons susceptibles de passer par effet tunnel à travers la jonction (2) c'est-à-dire le nombre d'électrons dans l'îlot central. Pour augmenter ce nombre d'*une* unité il faut augmenter V d'une quantité ΔV suffisante pour laisser V_1 positif après le passage de l'électron à travers la jonction (1). Or une augmentation ΔV de la tension de polarisation entraîne une augmentation de la tension V_1 de la quantité

$\Delta V_1=\Delta V\ C_2/(C_1+C_2)$ (Eq. 12-131-a) et le passage d'*un* électron dans l'îlot central diminue V_1 de la quantité $\Delta V_1=e/(C_1+C_2)$ (Eq. 12-131-a), la condition de passage d'*un* électron s'écrit donc $\Delta V\ C_2/(C_1+C_2) > e/(C_1+C_2)$ soit $\Delta V > e/C_2$. Mais d'un autre côté chaque nouvel électron arrivant dans l'îlot central augmente la tension V_2 de la quantité $\Delta V_2=e/(C_1+C_2)$ (Eq. 12-131-b), il en résulte que le passage subséquent de cet électron dans la résistance tunnel R_{t2} de la jonction (2) entraîne une variation de courant donnée par $\Delta I=\Delta V_2/R_{t2}$ soit $\Delta I=e/R_{t2}(C_1+C_2)$. Ainsi donc la caractéristique courant-tension du dispositif présente une allure de marches d'escalier de hauteurs $\Delta I=e/R_{t2}(C_1+C_2)$ et distantes horizontalement de $\Delta V=e/C_2$ (Fig. 12-31).

Il faut noter que la hauteur de la première marche est moitié de celle des autres car le seuil de passage du premier électron est $V=e/2C_2$ (Fig.12-30-d) alors que la distance entre marches est $\Delta V=e/C_2$. Enfin lorsque la tension V devient importante les marches s'estompent progressivement car l'énergie eV_2 devient importante ce qui accélère le transfert tunnel à travers la jonction (2). La condition $R_{t1}<<R_{t2}$ s'amenuise alors car les taux de transfert à travers chacune des jonctions deviennent comparables, les électrons traversent simultanément les deux jonctions sans saturer l'îlot central car la jonction (2) ne joue plus son rôle temporisateur.

Transistor à un électron – SET

Supposons maintenant qu'une électrode latérale permette, par effet capacitif par exemple, de faire varier de δq la charge de l'îlot central, on réalise ainsi un transistor (Fig. 12-32-a). La charge additionnelle δq, qui peut varier de façon continue, ne modifie en rien la hauteur ΔI des marches de courant, car cette dernière résulte du passage tunnel d'*un* électron à travers le dispositif, mais modifie leurs positions à travers la valeur de la tension V_1 aux bornes de la jonction (1). Ainsi donc par l'intermédiaire de la charge δq l'électrode latérale permet de commander le courant à travers le dispositif. Le transistor se comporte donc comme un transistor à effet de champ dans lequel le courant drain-source est contrôlé électron par électron, c'est le *transistor à un électron* SET (Single Electron Transistor). La caractéristique $I_d(V_d)$ à tension grille constante présente une zone de blocage comprise entre $-e/(2C_t+C_g)$ et $+e/(2C_t+C_g)$ suivie d'un régime linéaire de pente $1/2R_t$, où C_t et R_t sont la capacité et la résistance des jonctions tunnel et C_g la capacité de grille (Fig. 12-32-b). La caractéristique de transfert $I_d(V_g)$ à tension drain constante présente des variations périodiques de conductance de période e/C_g.

Notons pour terminer que les capacités doivent être très faibles avec toutefois des épaisseurs d'isolant très faibles pour autoriser l'effet tunnel. Il en résulte que l'îlot central est nécessairement de taille très réduite. Il se comporte donc à la limite comme un fil ou une boîte, quantique. Il faut alors prendre en considération les effets de basse dimensionalité, et en particulier le fait que les états d'énergie dans l'îlot deviennent discontinus et de plus varient avec la population de l'îlot.

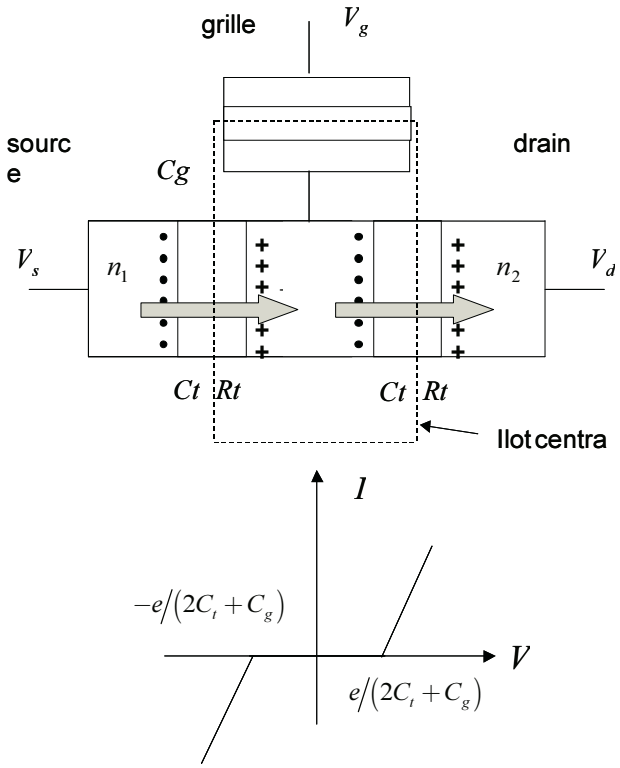


Figure 12-32 : Transistor à un électron

12.5. Nanotubes et nanofils

12.5.1. La structure des nanotubes

Ce sont actuellement les composants qui font l'objet de la recherche la plus intense. Une des raisons est sans doute la relative facilité avec laquelle il est possible de les fabriquer. Ils ont été découverts en 1991 par Sumio Iijma dans une expérience dont le but était d'étudier les filaments de carbone. Ce sont des plans de graphène qui s'enroulent sous forme de cylindre comportant une ou plusieurs feuilles.

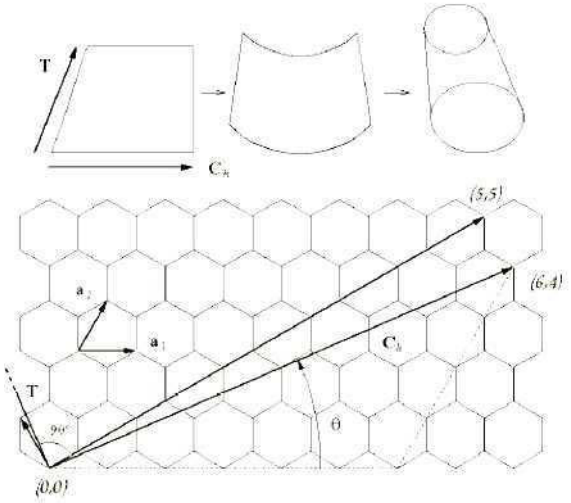


Figure 12.33 : Formation des nanotubes

Ces objets ont la propriété remarquable de présenter un rapport entre la longueur et le diamètre exceptionnellement élevé puisque le diamètre est compris entre 1 et 10 nm et puisque la longueur peut atteindre des dizaines de microns. La figure 12.33 représente deux types possibles de nanotubes monoparoi.

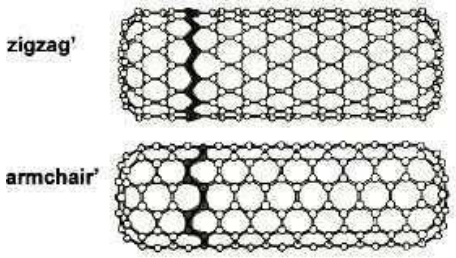


Figure 12.34 : Les nanotubes de carbone

La conduction dans ces dispositifs est totalement différente de celle habituellement observée dans les systèmes de plus grande taille. Il est donc nécessaire de donner quelques principes de base. Le chapitre 1 a permis de comprendre comment les densités d'états énergétiques pouvaient se former dans des structures à deux, à une et à zéro dimension. Il faut maintenant partir de ces résultats pour en déduire le courant électrique traversant de tels dispositifs.

Le résultat sera très différent de la loi d'ohm, uniquement valable quand les électrons subissent de nombreuses collisions dans le dispositif. Dans les dispositifs de taille nanométrique (au moins dans une dimension), la distance entre deux collisions peut être grande devant la dimension du dispositif. De manière plus précise, les systèmes de petite taille peuvent être caractérisés par trois longueurs :

- La longueur d'onde de Fermi est la longueur d'onde correspondant à l'énergie de Fermi des électrons soit :

$$\lambda_F = \frac{2\pi}{k_F} \text{ et } E_F = \frac{\hbar^2 {k_F}^2}{2m^*}$$

  Elle vaut 10 nm pour les semi-conducteurs et environ 1 nm pour les métaux.

- Le libre parcours moyen est la distance entre deux collisions. Il est environ de 1 micron pour des semi-conducteurs de bonne qualité.

- La longueur de cohérence de la phase est la longueur que parcourt l'électron avant que sa phase initiale ne soit modifiée. Elle est plus grande que le libre parcours moyen.

Quand les dispositifs ont des dimensions plus faibles que le libre parcours moyen, le régime est dit balistique et la loi d'ohm ne s'applique plus puisqu'elle est due aux collisions. Quand les dimensions du dispositif sont faibles devant la longueur de cohérence, des interférences entre les différentes trajectoires possibles peuvent avoir lieu et produire des effets de nature purement quantique. Les collisions inélastiques de l'électron avec le milieu détruisent la cohérence de phase et le comportement classique de la conduction est à nouveau vérifié. Pour expliquer la conduction dans les systèmes nanométriques, il faut faire appel à la théorie de la conduction due à Landauer. Cette théorie est assez difficile et nous ne donnerons ici qu'une interprétation très simplifiée.

12.5.2. *Une nouvelle théorie de la conduction*

On considère deux réservoirs d'électrons de grande taille, caractérisés par leurs potentiels chimiques, en liaison par un système unidimensionnel comme le montre la figure 12.35. Dans ce système unidimensionnel, les fonctions d'onde des électrons se propagent et on définit un canal de propagation pour un mode de propagation donnée.

Ce modèle peut être comparé à un mode de propagation dans un guide d'onde. Le fil unidimensionnel est traité comme un milieu de propagation idéal présentant un centre de diffusion pour la fonction d'onde de l'électron. On suppose de plus que la fonction d'onde se transmet en partie et se réfléchit en partie avec des paramètres T et R qui expriment respectivement le coefficient de transmission et le coefficient de réflexion.

Dans un premier temps, on considère un seul mode de propagation. Le vecteur d'onde de la fonction d'onde est quantifié quand on écrit les conditions périodiques dans la dimension transverse. Si A est la section du fil, il y a environ $A/\lambda_F{}^2$ valeurs possibles du vecteur d'onde.
Le paramètre λ_F est la longueur d'onde d'un électron dont l'énergie est celle de Fermi. Un mode de propagation, ou un canal, correspond à une valeur particulière du vecteur d'onde.

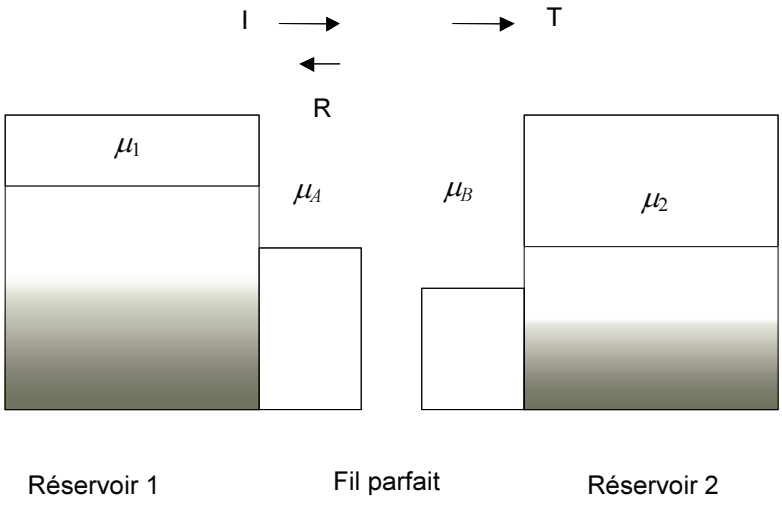


Figure 12.35. : Modèle de Landauer de la conduction

Pour les énergies en dessous de μ_2, il y a autant d'électrons qui passent de gauche à droite que l'inverse. Le courant est donc nul. Le courant est donc dû aux n électrons d'énergies entre μ_2 et μ_1. La valeur de ce courant, si v_F est la vitesse au niveau de l'énergie de Fermi, est alors

$$I = ev_F nT$$

Le nombre d'électrons impliqués n est

$$n = 2(\mu_1 - \mu_2)\frac{dn}{dE}$$

Pour un système à une dimension, le chapitre 1 a permis d'établir que

$$\frac{dn}{dE} = \frac{2}{hv_F}$$

On obtient donc

$$I = \frac{2e}{h} T\left(\mu_1 - \mu_2\right)$$

Si la tension entre les deux réservoirs est V, la relation classique établie dans le chapitre 2 s'applique

$$eV = \mu_1 - \mu_2$$

et donc,

$$I = \frac{2e^2}{h} TV$$

Il faut noter que même pour un fil parfait, transmission totale et T égal à un, la résistance n'est pas nulle. La valeur la plus faible de la résistance ou quantum de résistance est donc

$$R = \frac{h}{2e^2}$$

La valeur de ce quantum de résistance est 12.9 kΩ. C'est la valeur minimale de la résistance d'un tel dispositif. Cette propriété quantique est bien différente de ce que notre intuition pourrait prévoir.

Si maintenant nous mesurons la différence de potentiel V' au niveau des fils on peut écrire

$$V' = e\left(\mu_A - \mu_B\right)$$

Les potentiels μ_A et μ_B sont les potentiels chimiques des électrons dans le fil de part et d'autre de la barrière. En raisonnant sur les échanges d'électrons entre réservoirs et fils, on obtient

$$I = \frac{2e^2}{h} \frac{T}{R} V'$$

Cette fois, quand le dispositif est parfait (T=1 et R=0), la résistance est bien nulle.

La formule de Landauer peut se généraliser à plusieurs canaux. On définit alors une matrice de transition entre les canaux du fil de gauche et ceux du fil de droite. Les coefficients T_{ij} expriment les amplitudes de probabilité pour qu'un électron dans un

mode *i* à gauche soit transmis dans un mode *j* à droite. On montre alors que la conductance s'écrit

$$G = \frac{I}{V} = \sum_i \frac{2e^2}{h} T_i$$

avec,

$$T_i = \sum_j T_{ij}$$

La mesure de la conductance fait donc apparaître une variation marquée par des sauts au fur et à mesure que les canaux s'ouvrent comme le montre la figure 12.8. La figure 12.8 est un résultat expérimental illustrant les variations de la conductance entre deux gaz d'électrons situés à 250 nm, en fonction de la tension appliquée.

12.5.3. *Propriétés et fabrication des nanotubes*

Les propriétés des nanotubes découlent directement de leur filiation avec le graphite. De plus, des propriétés spécifiques sont apportées par l'enroulement et la faible dimension obtenue pour leur diamètre. En fonction de la direction de l'enroulement, on peut obtenir un comportement métallique ou semi-conducteur. Quand aucune précaution n'est prise la résistance d'un nanotube en contact avec deux conducteurs métalliques est de l'ordre du MΩ mais avec une évaporation appropriée on peut approcher la valeur théorique donnée précédemment. Il faut aussi tenir compte du fait qu'il y a deux sous-bandes au niveau de Fermi.

De plus, les nanotubes sont capables de transporter des densités de courant considérables de l'ordre de 10^{10} A/cm^2 soit 100 fois plus que les métaux. Cette propriété justifie l'intérêt présenté par ces dispositifs pour réaliser des fils de connexion dans les circuits intégrés ou pour réaliser le canal de conduction d'un transistor. Dans un autre registre, les nanotubes présentent également des propriétés mécaniques exceptionnelles. Ils peuvent se courber et se tordre très facilement.

Il est possible de dire quelques mots sur les techniques de fabrication. Deux procédés sont utilisés : la synthèse à haute température et la synthèse à moyenne température.

À haute température, l'opération se fait en deux étapes : évaporation du graphite puis condensation dans une enceinte dans laquelle il y a de fortes variations de températures. Il faut dépasser 3200 ° C et pour cela on peut faire usage d'un arc électrique ou d'un laser de puissance. On obtient en général des nanotubes multiparois. Pour obtenir des nanotubes monoparoi, qui présentent des propriétés plus séduisantes, il faut ajouter un catalyseur métallique à la poudre de graphite.

Dans le procédé à moyenne température, un gaz carboné est décomposé entre 500 et 1000°C en contact avec des particules de catalyseur métallique. Le carbone libéré par la décomposition du gaz donne naissance au niveau des particules métalliques à des nanotubes. Cette voie semble plus compatible avec un procédé micro-électronique. La figure 12.36 montre des nanotubes obtenus par ces techniques

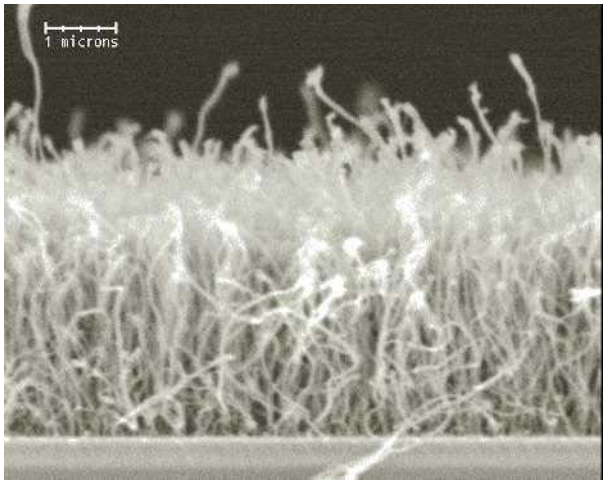


Figure 12.36 : Nanotubes obtenus par synthèse

12.5.4. *Applications des nanotubes*

La première application envisagée est d'utiliser les nanotubes comme conducteurs dans les futurs circuits intégrés. Ils permettent en effet de transporter des courants avec des densités remarquables de 10^{10} A/cm^2 ce qui est largement supérieur aux meilleurs conducteurs actuels, le cuivre par exemple. Il est cependant nécessaire de savoir réaliser des contacts de faible résistance avec d'autres matériaux utilisés en micro-électronique. Quand aucune précaution n'est prise, des contacts de plusieurs MΩ sont obtenus. Il a été démontré cependant qu'il était possible de s'approcher de la résistance théorique.

En fonction des conditions de fabrication, des propriétés métalliques ou semi-conductrices peuvent être obtenues ce qui ouvre le champ à la fabrication de dispositifs actifs comme des diodes ou même des transistors. Les propriétés de conduction dépendent fondamentalement de la manière dont la feuille de graphène s'enroule sur elle-même pour former un tube et le contrôle de cette propriété à la

fabrication reste un défi pour les années futures. Aujourd'hui la seule possibilité est de les trier après fabrication.

La forme tout à fait particulière des nanotubes laisse envisager des applications intéressantes en émission. Les nanotubes sont des émetteurs d'électrons à faible tension. En effet, le champ électrique intense à l'extrémité d'un nanotube peut favoriser l'émission d'électrons par effet de pointe.

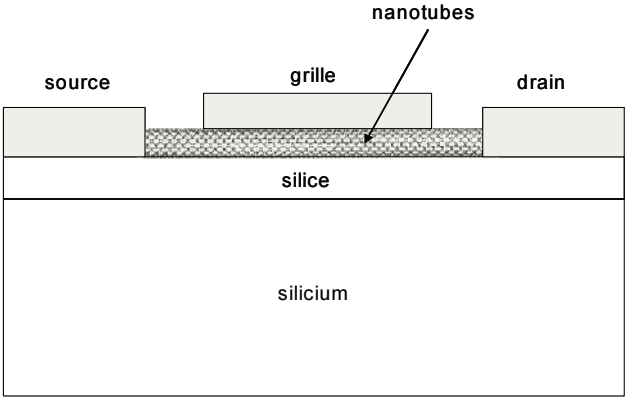


Figure 12.37 : Transistor à nanotubes

Enfin, les nanotubes peuvent être utilisés comme éléments de conduction dans un transistor dont le canal de conduction serait alors remplacé par ces éléments cylindriques. La figure 12.37 représente un tel dispositif.

Ce dispositif est très semblable à un transistor MOS classique. La seule différence est dans la constitution de la zone de canal. Différentes géométries ont été expérimentées en plaçant la grille soit dessus soit dessous. Le fonctionnement de ces dispositifs n'est pas encore parfaitement compris mais il est admis que l'effet transistor est du à la modulation de la barrière de potentiel au niveau des contacts et non pas à la modulation de la conductance dans les nanotubes eux-mêmes. Les contacts entre les nanotubes et les électrodes de drain et de source doivent être considérés comme des contacts Schottky entre un métal et un semi-conducteur. Les performances électriques obtenues pour la transconductance sont remarquables ce qui pousse l'industrie électronique à s'intéresser à des dispositifs avancés.

12.5.5. *Les nanofils*

Ce sont des dispositifs nanométriques assez proches des nanotubes de carbone. Ils sont également candidats pour réaliser de nouveaux dispositifs dans le futur. On peut les définir comme des objets dont le rapport longueur sur largeur est supérieur à 10 avec une largeur de quelques dizaines de nanomètres. Leur facteur

de forme est sans commune mesure avec celui des nanotubes de carbone. Ils sont en général fabriqués dans un matériau semiconducteur ce qui laisse envisager une possible intégration à une technologie micro-électronique. Leur intérêt est dans leur faible dimension. Les méthodes de fabrication appartiennent à deux familles : les méthodes dites « top-down » empruntées à la micro-électronique et les méthodes dites « bottom-up » inspirées des procédés chimiques.

L'approche « top-down » fait usage de la lithographie extrême et doit être considérée comme une première approche pour fabriquer des nanofils. Ce n'est pas une méthode envisageable pour une production industrielle. Un objectif majeur est en effet de réaliser des dispositifs à base de fils en se passant le plus possible de la lithographie de très grande précision. Des techniques sophistiquées comme la microscopie en champ proche permettent également en manipulant les atomes de réaliser des motifs nanométriques. Il faut également citer la nano-impression qui, à l'image des procédés de l'imprimerie, permet de reproduire un motif fabriqué à l'aide d'une lithographie de haute précision.

L'approche « bottom-up » a pour ambition d'être applicable dans le futur à l'échelle industrielle. L'objectif est alors de réaliser sur une surface un grand nombre de dispositifs identiques en utilisant des procédés chimiques. On parle d'auto-assemblage quand la surface présente une structure périodique régulière et peut servir de support à la croissance d'un matériau donné. La technique VLS (Vapor Liquid Solid) est souvent mise en œuvre. Le principe est de créer une goutte en fusion du semiconducteur choisi dans un réacteur à haute température (1000 degrés environ) puis de faire croître les fils en s'appuyant sur les irrégularités précédemment décrites.

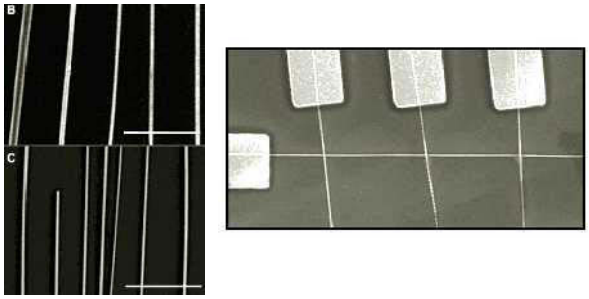


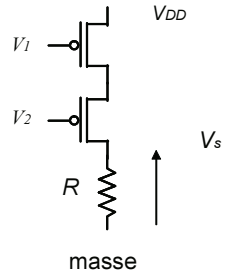
Figure 12. 38 : Réseau de nanofils

Les semiconducteurs habituels (silicium, arseniure de gallium, …) sont expérimentés dans les laboratoires. Le dopage peut être contrôlé par la composition de la phase vapeur d'entrée. Des fils semi-conducteurs de quelques microns de long pour une dizaine de nm de marge sont ainsi obtenus. La figure 12.38 montre des réalisations de laboratoire.

Les résultats présentés sont en grande partie issus des travaux de Charles Lieber et de ses collaborateurs. Un nanofil seul n'a pas de fonctionnalité particulière mais quand on réalise un croisement de nanofils, de nombreuses possibilités sont offertes.

La première est de réaliser une simple connexion. Dans une version plus sophistiquée, il est possible de rendre cette connexion programmable par exemple en faisant jouer les forces électrostatiques capables de coller les deux fils quand les valeurs des tensions appliquées sont convenablement choisies. Si les deux fils sont dopés différemment alors le croisement des deux fils matérialise une jonction *pn* . Il est donc facile d'imaginer une logique à diodes utilisant de tels dispositifs.

Faisons maintenant croître un oxyde à la surface de l'un des fils de type semi-conducteur et appliquons un autre fil supposé conducteur et placé perpendiculairement. Le dispositif réalisé est de type transistor. Si le fil semi-conducteur est de type *p*, une tension positive appliquée sur le fil conducteur a pour effet de repousser les trous du fil de part et d'autre de la zone de croisement. La conduction est alors impossible dans le fil considéré et son comportement est analogue à celui d'un transistor bloqué.



V_1 (V)	V_2(V)	V_S(V)
0	0	4.85
5	0	0.02
0	5	0.06
5	5	0.01

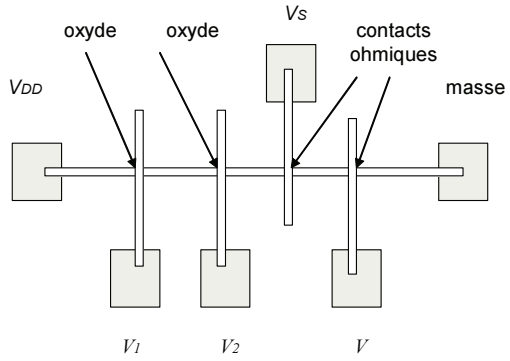


Figure 12. 39 : Exemple de fonction logique à base de nanofils

Ces différents principes ont permis de réaliser les principales fonctions logiques de la micro-électronique. L'exemple du NOR est donné figure 12.39 et illustre les travaux de Charles Lieber. Les caractéristiques de cette nouvelle logique semblent séduisantes. Il faut cependant être conscient des difficultés de la fabrication collective de tels dispositifs et être attentif aux difficultés de positionnement des fils entre eux. Une autre application intéressante est le dopage des fils dans le sens de la longueur.

13. Semiconducteurs à grand gap Composants haute température

Plus de 80% des composants électroniques sont aujourd'hui réalisés à base de silicium, mais la plupart d'entre eux sont qualifiés pour fonctionner à des températures inférieures à 70 °C. Cette limite devient 125 °C dans les applications militaires et peut être portée à 250 °C moyennant une encapsulation appropriée mettant en jeu des matériaux et des techniques spécifiques. Or de nombreuses utilisations nécessitent un fonctionnement à des températures bien supérieures. Citons par exemple les microprocesseurs embarqués, sur des véhicules terrestres, aériens ou spatiaux, qui dans l'environnement de moteurs peuvent avoir à supporter des températures avoisinant 650 °C. Citons aussi l'industrie pétrolière qui a besoin, dans les têtes de forage, de capteurs accompagnés de leur électronique de contrôle fonctionnant à haute température et haute pression. Notons enfin que les hautes températures peuvent certes être le fait d'un environnement hostile mais peuvent aussi résulter d'un échauffement interne du composant, c'est particulièrement le cas dans un composant de puissance. Les figures de mérite qui conditionnent les performances de ce type de composant sont, en régime passant sa faculté à évacuer la chaleur, en régime bloqué sa faculté à supporter des tensions élevées.

13.1. Solutions technologiques
13.2. Semiconducteurs à grand gap
13.3. Composants optroniques de courte longueur d'onde

13.1. Solutions technologiques

Le fonctionnement des composants à haute température est limité par différentes contraintes dont chacune est souvent rédhibitoire. Une contrainte importante est par exemple la densité de porteurs intrinsèques, qui, conditionnée par le gap du matériau, varie exponentiellement avec la température (Eq. 1-146). Ces porteurs thermiques augmentent les courants de fuite, en outre lorsque leur densité devient comparable aux concentrations de dopants le contrôle de la charge d'espace dans le composant devient impossible. La densité de porteurs intrinsèques du silicium par exemple, passe de 10^{10} cm^{-3} à la température ambiante à 10^{17} cm^{-3} à 600 °C. D'autres types de problème associés aux hautes températures résultent par exemple de la diffusion des espèces chimiques constituant les dopages. Cette diffusion détruit à terme la spécificité du composant. D'autres problèmes sont liés à la stabilité des contacts ohmiques. D'autres contraintes enfin sont liées à la dilatation thermique différentielle des différents constituants du composant.

Différentes voies sont explorées pour prolonger le fonctionnement des composants vers les hautes températures ou les fortes puissances. La première consiste à trouver des solutions technologiques pour optimiser toujours davantage les performances des semiconducteurs classiques, silicium ou GaAs. Mais à mesure que les besoins évoluent vers des conditions de plus en plus extrêmes, les solutions technologiques butent de plus en plus sur la barrière infranchissable des propriétés intrinsèques des matériaux. De fait, les limites théoriques des performances de bon nombre de composants sont aujourd'hui pratiquement atteintes. Une solution alternative aux matériaux traditionnels devient ainsi progressivement nécessaire. Cette alternative n'existe que dans l'utilisation de matériaux de remplacement potentiellement plus performants, notamment dans les domaines des hautes températures et des fortes puissances. Parmi ces matériaux les semiconducteurs à grand gap constituent évidemment des éléments de choix.

13.1.1. *Silicium sur isolant – SOI*

Les régions actives des composants au silicium peuvent potentiellement fonctionner à des températures bien supérieures à leur limite pratique, cette dernière est en effet essentiellement fixée par la trop faible résistivité des substrats. La résistivité des substrats silicium est dans le meilleur des cas de l'ordre de quelques 10^3 Ωcm, alors que celle des substrats GaAs ou InP se situe dans la gamme de 10^6 à 10^8 Ωcm. Cette insuffisante résistivité des substrats silicium entraîne d'une part l'existence de courants de fuite importants qui augmentent inutilement la consommation énergétique des circuits intégrés, et d'autre part la présence de capacités parasites de couplage avec le substrat, ce qui réduit les performances des circuits radiofréquence. L'utilisation de substrats plus isolants permet de réduire notablement la consommation d'énergie notamment par la réduction des surfaces de jonctions qui peuvent alors être directement intégrées au substrat. Dans les domaines à la fois des grandes vitesses et des radiofréquences la technologie silicium peut devenir plus compétitive que la technologie GaAs par l'utilisation de

substrats silicium aussi résistifs que leurs concurrents GaAs. Ces substrats sont réalisés et utilisés dans le cadre de la technologie *Silicium-sur-Isolant* , SOI (Silicon On Insulator). Deux méthodes de base permettent de les réaliser.

a. Technologie SIMOX

La première technique permettant de réaliser des substrats SOI est la technique d'*Isolation par Implantation d'Oxygène* SIMOX (Separation by IMplanted OXygen). La technique consiste à créer, par implantation ionique, une couche d'oxygène enterrée sous la surface d'une plaquette de silicium. La plaquette est ensuite recuite pendant environ 6 heures à 1300 °C. Ce recuit régénère d'une part les qualités cristallines du silicium traversé, et entraîne d'autre part la formation d'une couche enterrée de silice SiO_2, dont l'épaisseur est fonction de la dose d'implantation. La structure type d'un substrat SIMOX est constituée d'une couche d'oxyde de 400 nm d'épaisseur enterrée à 200-400 nm sous la surface du silicium. Elle est obtenue par implantation d'ions O^+ de haute énergie (100 mA à 200 keV), avec une dose de l'ordre de 2.10^{18} cm^{-2}.

b. Technologie BESOI

La seconde technique utilisée dite BESOI (Bond Etch back SOI), consiste à solidariser deux plaquettes de silicium sur lesquelles une couche de silice est préalablement réalisée par oxydation thermique. Les plaquettes sont solidarisées par leurs faces oxydées en utilisant les forces de Van der Walls. Un recuit permet ensuite de renforcer mécaniquement la liaison des deux plaquettes, les deux couches d'oxyde forment ainsi un isolant enterré. L'une des plaquettes est enfin amincie, par voie sèche ou humide. On obtient ainsi une couche de silicium dont l'épaisseur peut varier de quelques dizaines de nanomètres à quelques dizaines de microns, isolée de la plaquette par une couche d'oxyde enterrée dont l'épaisseur peut varier dans les mêmes proportions. Ce procédé qui nécessite deux plaquettes de départ est évidemment plus onéreux que le procédé SIMOX, en contre partie il présente plus de latitude sur les épaisseurs des différentes couches.

c. Procédé Smart-cut

Une mince couche de silicium est fabriquée après une implantation d'hydrogène ce qui permet de créer une zone de clivage dans une plaquette de silicium. Elle est ensuite rapportée sur une couche préalablement oxydée comme dans le procédé BESOI.

d. Composants suspendus

L'utilisation de substrats SOI réduit de façon drastique les courants de fuite, par la présence de la couche isolante enterrée, mais a relativement peu d'incidence sur les performances haute fréquence car ces dernières restent dégradées par le couplage radiofréquence entre le composant et le substrat à travers la couche

d'oxyde. Une étape supplémentaire est franchie dans l'isolation radiofréquence par l'élimination locale du substrat de silicium sous l'oxyde enterré supportant le composant. On transforme ainsi des portions du substrat de silicium en régions à haute résistivité et on obtient une structure dans laquelle chaque composant est pratiquement situé sur un pont d'oxyde. L'érosion du substrat de silicium est réalisée après passivation du composant, à travers de petites ouvertures réalisées sous la couche d'oxyde. Une érosion très sélective entre Si et SiO_2 est obtenue avec SF_6, de sorte que de larges cavités, sur des profondeurs de l'ordre de 100 μm, peuvent être formées sous chacun des composants. On améliore de la sorte les performances des composants radiofréquence.

13.1.2. *Arséniures sur diamant*

Un avantage incontestable de la technologie GaAs sur la technologie silicium est la possibilité de disposer de substrats semi-isolants (10^6 Ωcm). Malheureusement la résistivité de ces substrats de GaAs se dégrade à haute température, entraînant une augmentation des courants de fuite qui affectent les performances des FET. Néanmoins des transistors MESFET au GaAs fonctionnent à des températures voisines de 350 °C. On améliore les performances du substrat en intercalant une couche tampon de AlGaAs. Des capteurs à effet Hall réalisés suivant cette technologie, sont utilisés dans l'industrie automobile, à des températures avoisinant 500 °C.

Mais outre la dégradation de leur résistivité à haute température, les substrats de GaAs présentent un autre inconvénient, ils sont notoirement mauvais conducteurs thermiques. Ceci est particulièrement préjudiciable aux transistors bipolaires de puissance, qui sont traversés par de fortes densités de courant. Leurs performances sont de ce fait limitées par des effets thermiques liés aux températures de jonction. L'augmentation trop importante des températures de jonction entraîne, à forte puissance, l'apparition de résistances différentielles négatives sur les réseaux de caractéristiques de sortie (Fig. 13-1). Cet effet est accompagné d'une réduction du gain en courant β quand la tension collecteur-émetteur augmente.

Or les récents progrès réalisés dans la manipulation des couches minces permettent maintenant de séparer la couche active d'un composant, du substrat sur lequel il a été élaboré, et de le déposer ensuite sur un autre substrat, possédant par exemple de meilleures propriétés thermiques. Le diamant dans ce domaine est particulièrement apprécié car sa conductivité thermique, de l'ordre de 20 W/cmK, est 40 fois supérieure à celle de GaAs. Ainsi des transistors bipolaires à hétérojonction, de type GaAs/AlGaAs, sont séparés de leur substrat initial par une érosion sélective de ce dernier, puis transférés sur un substrat de diamant où ils sont fixés par une couche d'interface d'indium ou de palladium. L'érosion du substrat de GaAs est réalisée par la technique TSR (Total Substrate Removal), le substrat de diamant peut être un monocristal ou une couche polycristalline obtenue par dépôt en phase vapeur CVD (Chemical Vapour Deposition).

La figure (13-1) représente les réseaux de caractéristiques d'un transistor HBT GaAs/AlGaAs dans trois types de configuration: sur substrat GaAs, en l'absence de substrat, et sur substrat diamant. Il est clair que l'absence de substrat, en minimisant la dissipation thermique, détériore notoirement l'allure des caractéristiques. Par contre le substrat diamant, en évacuant mieux la chaleur que le substrat GaAs, améliore les performances du transistor. Il faut cependant noter que l'essentiel du gradient de température se situe au voisinage immédiat de la source de chaleur. En d'autres termes le flux de chaleur rencontre la plus grande résistance thermique dans la couche de GaAs, de 1 à 2 microns d'épaisseur, constituant le collecteur. Le diamant ne peut donc jouer pleinement son rôle dissipateur. Ainsi, bien que sa conductivité thermique soit 40 fois supérieure à celle de GaAs, la résistance thermique du composant sur substrat diamant n'est que 10 à 20 fois inférieure à celle du composant sur substrat GaAs. Néanmoins, par rapport au transistor sur substrat GaAs, le transistor équivalent sur substrat diamant peut dissiper une puissance environ trois fois supérieure.

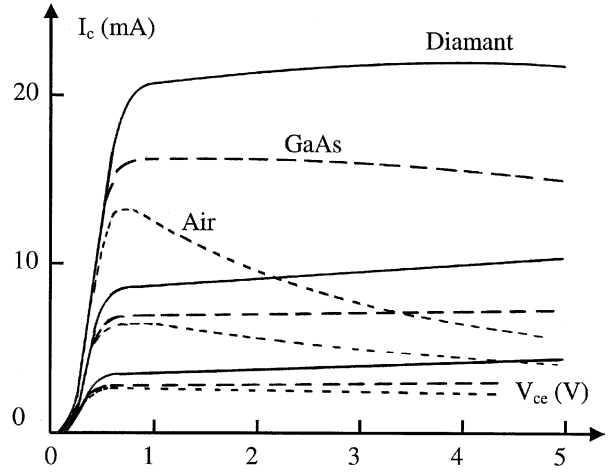


Figure 13-1 : Réseaux de caractéristiques de sortie d'un transistor HBT GaAs/AlGaAs dans trois types de configuration (Arbet-Engels V. - 1995). a) Sur substrat GaAs. b) Sans substrat. c) Sur substrat diamant *(Arbet-Engels V. - 1995)*

13.2. Semiconducteurs à grand gap

Les semiconducteurs à grand gap sont les matériaux de choix pour la réalisation de composants haute température. Non seulement ils présentent des densités de porteurs intrinsèques qui restent toujours faibles jusqu'à 1000 °C (Fig. 13-2-a), mais en outre leurs propriétés physiques remarquables, comme l'énergie de

cohésion, leur permettent de supporter sans dommage des températures très élevées. Plusieurs domaines des composants sont concernés par l'utilisation de ce type de semiconducteur.

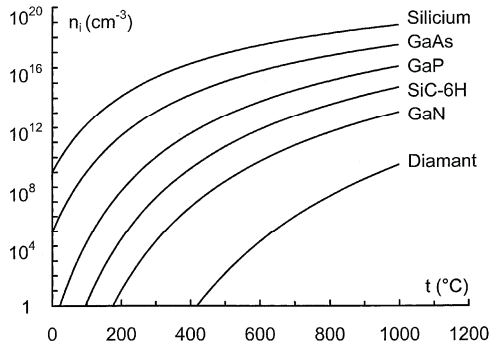


Figure 13-2-a : Densités de porteurs intrinsèques en fonction de la température dans différents semiconducteurs.

Les semiconducteurs à grand gap intéressent tout d'abord le domaine de la puissance. Les raisons essentielles sont d'une part leur bonne conductivité thermique qui permet une dissipation efficace de la chaleur en régime passant, de l'autre leur tension de claquage élevée qui assure une bonne tenue en tension inverse en régime bloqué.

Le domaine des hautes fréquences, de puissance et/ou à haute température, comme les radars embarqués par exemple, est aussi un utilisateur de semiconducteurs à grand gap. Ces derniers autorisent une double réduction de taille et de poids car leur faculté de fonctionner avec des densités de puissance élevées permet de réduire d'autant les dispositifs de refroidissement. Ces deux paramètres, taille et poids, sont capitaux dans les matériels embarqués et ceci dans tout type de véhicule mais surtout évidemment dans le domaine spatial. Notons en outre que dans les semiconducteurs à grand gap les électrons conservent une vitesse de saturation élevée qui conditionne les performances en haute fréquence, même à haute température. Les systèmes de communication mis en œuvre notamment dans les avions sans pilote et les réseaux de satellites, sont demandeurs d'émetteurs de forte puissance en bande X, fonctionnant à des températures de l'ordre de 500 °C.

Les semiconducteurs à grand gap ont aussi des applications, et non des moindres, dans le domaine de l'optoélectronique à la fois en détection et en émission. En ce qui concerne les détecteurs, leur gap les rend à la fois performants dans le spectre ultraviolet et aveugles au rayonnement visible. Ceci permet de faire de l'imagerie UV en présence du soleil en évitant tout phénomène d'éblouissement. Les émetteurs à grand gap sont quant à eux particulièrement concernés par les mémoires optiques. En effet, la densité de stockage sur disque optique est limitée par la longueur d'onde du rayonnement émis par les diodes laser utilisées pour lire

l'information. Cette densité varie comme λ^{-2}. En technologie GaAs les lasers utilisés émettent un rayonnement infrarouge dont la longueur d'onde est de l'ordre de 0,8 μm. L'utilisation de lasers bleus, ou *a fortiori* ultraviolets, permet d'augmenter considérablement la capacité de stockage. En outre, ces lasers à grand gap peuvent travailler à des températures relativement élevées sans que leurs performances soient affectées de manière significative.

Les hétérostructures aussi ont un profit certain à tirer de l'utilisation de semiconducteurs à grands gaps. En effet les discontinuités de bandes peuvent être considérables ce qui permet dans les gaz d'électrons bidimensionnels, d'atteindre des densités de porteurs très importantes ($>10^{13}$ cm^{-2}). Par voie de conséquence les transistors à effet de champ élaborés à partir de ces matériaux présentent des conductances de drain et des transconductances plus élevées.

Notons enfin que les semiconducteurs à grand gap ont la particularité de présenter une affinité électronique faible qui peut même sous certaines conditions, de symétrie et de traitement de surface, devenir négative. Cette propriété, très appréciée dans le domaine de la photoémission, débouche sur la réalisation de cathodes froides et de dispositifs d'affichage à écrans plats.

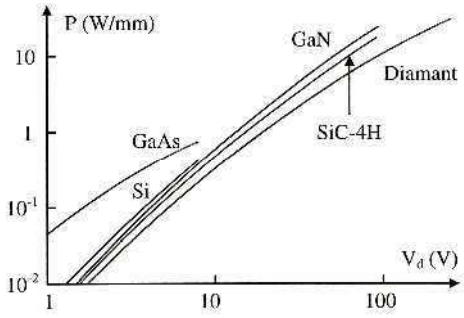


Figure 13-2-b : Densités de puissance P(W/mm) en fonction de la tension de drain, pour des transistors MESFET à Si, GaAs, GaN, SiC et diamant
(Weitzel C.E. - 1996)

Dans la panoplie des semiconducteurs à grand gap, les plus prometteurs sont le *diamant* (C), le *carbure de silicium* (SiC), et les *composés III-N* (BN, AlN, GaN, InN). Certains semiconducteurs II-VI, comme ZnS, ZnSe et leurs alliages ZnSSe, sont des matériaux à grand gap, mais ils présentent généralement des énergies de cohésion trop faibles pour supporter sans dommage les hautes températures. A titre d'illustration, la figure (13-2-b) présente une comparaison des performances modélisées de composants à base de Silicium, GaAs, GaN, SiC et diamant dans le domaine des radiofréquences. Le composant sélectionné est un transistor MESFET, en raison du fait que sa modélisation est parmi les plus simples, avec un canal de longueur L=3 μm et de largeur Z=1 mm, dopé n avec N_d=10^{17} cm^{-3}. Les densités de

puissance accessibles sont respectivement de l'ordre de 17, 22 et 32 W/mm avec SiC, GaN et le diamant, alors qu'elles sont limitées à 0,5 et 0,8 W/mm pour les transistors à base de Si et GaAs. Ces performances sont liées aux tensions de claquage, qui sont respectivement de l'ordre de 0,3 et 0,4 MV/cm pour Si et GaAs et atteignent 2, 3,3 et 5,6 MV/cm pour SiC, GaN et le diamant.

Les valeurs élevées des énergies de cohésion, qui conditionnent la tenue du composant en température, sont en contrepartie des sources de difficultés majeures dans la réalisation des dispositifs car elles compliquent d'autant toutes les étapes du processus de fabrication, polissages mécaniques, érosions chimiques, dopages, réalisation de contacts ohmiques, etc. En outre, les matériaux à grand gap ont généralement des températures de fusion très élevées et nécessitent pour leurs épitaxies des températures de substrats élevées. Or non seulement il est difficile d'atteindre et de contrôler des températures élevées, mais surtout, à ces températures les matériaux utilisés dans les systèmes d'élaboration dégazent des impuretés qui sont susceptibles de contaminer le cristal. Il est de ce fait difficile d'obtenir des substrats d'homoépitaxie de taille supérieure à quelques millimètres, alors que parallèlement les substrats silicium atteignent 30 cm de diamètre. La difficulté d'obtenir des substrats d'homoépitaxie de qualité se traduit par le fait que les semiconducteurs à grand gap sont souvent déposés sous forme d'hétéroépitaxies, avec d'importants désaccords paramétriques couche-substrat. Il en résulte des couches de mauvaises qualités cristallines contenant beaucoup de dislocations.

13.2.1. *Diamant*

Le diamant présente des caractéristiques très attractives qui en font potentiellement le semiconducteur idéal dans différents domaines des composants électroniques. Plusieurs problèmes technologiques restent toutefois à résoudre. Le principal d'entre eux est le dopage *n*. Le dopage *p* est actuellement réalisé avec le bore. Il pose néanmoins un problème lui aussi car les niveaux accepteurs associés au bore sont très profonds. Ces accepteurs sont loin d'être tous ionisés à la température ambiante. Enfin la réalisation de substrats opérationnels peu coûteux est aussi un réel problème.

a. Elaboration

Le diamant peut être élaboré en couches minces par dépôt en phase vapeur CVD (Chemical Vapour Deposition). L'homoépitaxie sur substrat diamant donne des couches monocristallines avec des mobilités de trous de l'ordre de 1500 cm^2/Vs, l'hétéroépitaxie sur substrat silicium donne des couches polycristallines avec des mobilités de trous de l'ordre de 70 cm^2/Vs. Le diamant intrinsèque constitue un substrat fortement isolant, sa résistivité, supérieure à 10^{16} Ωcm, est de 8 à 10 ordres de grandeur supérieure à celle du GaAs semi-isolant. En outre sa grande conductivité thermique en fait un substrat de choix lorsqu'il est nécessaire d'évacuer beaucoup de chaleur. Les couches peuvent alors être polycristallines, non dopées, et par suite réalisées sur substrat relativement peu coûteux. La figure (13-1)

représente les réseaux de caractéristiques de sortie d'un transistor HBT GaAs/AlGaAs sur deux types de substrat. Il est clair que le substrat diamant, en évacuant mieux la chaleur que le substrat GaAs, améliore les performances du transistor.

b. Dopage

Le diamant peut être dopé de type *p* , avec le bore, mais il est actuellement difficile de le doper de type *n.* Le bore donne un niveau accepteur situé à 365 meV au-dessus de la bande de valence, cette valeur élevée de l'énergie d'activation des accepteurs signifie que ces derniers ne seront totalement ionisés que pour des températures typiquement supérieures à 500 °C. A cette température la mobilité des trous chute approximativement à 125 cm^2/Vs.

La difficulté actuelle de le doper *n* limite l'éventail des composants réalisables à des hétérojonctions, des diodes Schottky, et certains transistors à effet de champ. D'autre part la valeur élevée de l'énergie d'activation des accepteurs, limite son utilisation à des températures supérieures à 500 °C. La réalisation de dopage *n* et l'utilisation de dopants *p* moins profonds s'avèrent donc nécessaires.

c. Caractéristiques

Outre les avantages inhérents à la valeur élevée de son gap, qui sont communs à tous les semiconducteurs à grand gap, le diamant présente des caractéristiques qui lui sont propres, notamment dans le domaine thermique et dans celui de l'émission de champ. Ces caractéristiques spécifiques élargissent ses potentialités bien au-delà du simple champ d'application des semiconducteurs à grand gap.

Le diamant naturel a un gap indirect de 5,45 eV et un gap direct de 7,5 eV. Les mobilités des porteurs y sont les meilleures de tous les semiconducteurs à grand gap, à la température ambiante les mobilités des électrons et des trous sont respectivement de l'ordre de 2200 et 1600 cm^2/Vs. Les vitesses de saturation des porteurs sont supérieures à 2.10^7 cm/s. Son champ de claquage, le plus grand de tous les semiconducteurs, est de l'ordre de 6.10^6 V/cm. Sa conductivité thermique est exceptionnelle, 20 W/cmK, 13 fois supérieure à celle du silicium et 40 fois supérieure à celle de GaAs. Son coefficient de dilatation thermique est parmi les plus faibles, 10^{-6} K^{-1}. Enfin, et cette caractéristique est loin d'être dénuée d'intérêt, il peut présenter une affinité électronique négative. C'est notamment le cas pour les faces (111) hydrogénées, qui présentent une affinité électronique χ=-0,7 eV. Le bas de la bande de conduction est alors situé au-dessus du niveau du vide, de sorte que les électrons, éventuellement injectés dans la bande de conduction, sortent librement du matériau. Cette propriété trouve une application immédiate dans la réalisation de cathodes froides et de dispositifs d'affichage à émission de champ, car le matériau peut émettre des électrons de surface à la température ambiante et à partir de champs relativement faibles.

L'absence de dopage *n* interdit évidemment la réalisation de jonctions *pn* au diamant, mais la possibilité de dopage *p* autorise la réalisation d'hétérojonctions. Une hétérojonction de type *diamant(p)-GaAs(n)* par exemple, peut êtrte obtenue de la manière suivante. Une couche mince de 2 µm d'épaisseur de diamant(p) dopé au bore est épitaxiée sur un substrat de diamant isolant, par dépôt en phase vapeur. Le contact ohmique sur ce diamant est réalisé par évaporation d'un alliage or/titane. En parallèle, une couche mince de 1 µm de GaAs(n) dopée au sélénium est épitaxiée sur une couche mince non dopée de AlAs déposée sur un substrat GaAs. Un contact ohmique est réalisé sur la couche de GaAs(n) par évaporation suivie d'un recuit, d'un alliage or/germanium. Une dissolution sélective de l'interface de AlAs permet ensuite de séparer la couche de GaAs de son substrat. La couche détachée est enfin fixée sur le diamant(p) avec lequel elle est solidarisée par des forces de Van der Walls. On obtient ainsi une hétérojonction *diamant(p)-GaAs(n)* (Sugino T. - 1996). Le diagramme énergétique de la structure est représenté sur la figure (13-3), les gaps du diamant et de GaAs sont respectivement de 5,45 et 1,42 eV, les affinités électroniques respectives sont de 2,7 et 4,1 eV, les discontinuités de bandes sont estimées à ΔE_c=1,77 et ΔE_v=2,31 eV. La tension de diffusion de la diode est de l'ordre de 2,5 volts, la caractéristique I(V) est représentée sur la figure (13-3). La barrière que voient les électrons de GaAs est beaucoup plus importante que la barrière que voient les trous du diamant. En conséquence le courant direct de la jonction est essentiellement un courant de trous. Pour injecter des électrons dans le diamant il faudrait utiliser un semiconducteur de gap beaucoup plus important, comme AlN par exemple.

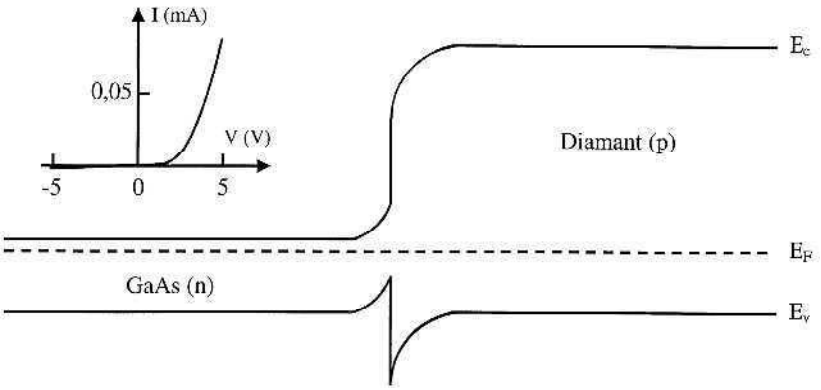


Figure 13-3 : Diagramme énergétique de l'hétérojonction diamant(p)-GaAs(n). Caractéristique I(V) de la diode.

Des transistors MISFET fonctionnant à des températures atteignant 500 °C ont été réalisés à base de structures *métal-diamant isolant-diamant semiconducteur* et *métal-isolant-diamant semiconducteur*. Ces structures présentent des densités d'états d'interface relativement importantes, 10^{12} $cm^{-2}eV^{-1}$, de 2 ordres de grandeur

supérieures à celles obtenues dans les MOS silicium. Les transistors ont, à 500 °C, des transconductances de 1,3 mS/mm avec des courants de drain de 10 mA.

Le diamant est aussi utilisé pour réaliser des filtres à ondes acoustiques de surface. Son intérêt réside ici dans sa grande vitesse de phase, 10 km/s, comparée à 4,5 km/s pour le niobate de lithium ($LiNbO_3$) généralement utilisé dans ce type de composant. Cette vitesse de phase permet de réaliser des filtres interdigités à 3,5 GHz, avec des peignes de 0,75 µm d'intervalle au lieu de 0,25 µm pour des filtres équivalents à base de niobate de lithium.

Dans le domaine de l'optoélectronique le diamant présente aussi un intérêt certain. La valeur très élevée de son gap en fait un matériau de choix pour la réalisation de détecteurs, dans la gamme des rayonnements UV et ionisants. Des cellules photoconductrices de type *métal-diamant-métal* et des photodiodes Schottky à base de diamant de type *p* ont été réalisées. Les cellules photoconductrices présentent un pic de sensibilité maximum à 200 nm avec une sensibilité 10^6 fois plus grande à cette longueur d'onde que dans le spectre visible. Cette caractéristique, unique, a de nombreuses applications car elle permet de réaliser des détecteurs UV insensibles à l'éblouissement solaire.

Enfin la possibilité de l'obtenir avec une affinité électronique négative ouvre un champ d'application dans le domaine des écrans plats par la réalisation de dispositifs à émission de champ. Le diamant est alors utilisé comme cathode. Une anode située à quelques microns de distance, polarisée positivement, crée un champ électrique qui extrait les électrons de surface de la cathode. Ces électrons émis excitent alors un écran fluorescent. Le problème à résoudre est la nécessité d'une alimentation en électrons afin d'assurer le maintien du courant d'émission. Il faut donc injecter des électrons par la face arrière de la cathode de diamant. Dans le diamant de type *p* l'injection d'électrons pourrait être réalisée à partir d'une hétérojonction constituée de diamant(p) et d'un nitrure(n) de gap plus élevé, comme par exemple le nitrure de bore cubique BN (E_g=6,4 eV) ou le nitrure d'aluminium AlN (E_g=6,28 eV).

13.2.2. *Carbure de silicium - SiC*

Parmi les semiconducteurs à grand gap le carbure de silicium (SiC) occupe une place particulière car il est actuellement le seul disponible à la fois sous forme de monocristaux massifs et sous forme de couches épitaxiées de bonnes qualités. En outre il peut facilement être dopé *n* ou *p*. Ses possibilités dans le domaine des hautes températures sont théoriquement bien supérieures à celles du silicium. Les composants à base de SiC sont capables de fonctionner à des températures atteignant au moins 600 à 700 °C. En raison de sa composition chimique, un bon nombre de ses propriétés sont intermédiaires entre celles du diamant et du silicium. C'est par exemple le cas pour le gap, la constante diélectrique, la conductivité thermique, le champ de claquage, ou la dureté. C'est un matériau relativement dur, dans un autre champ d'application que celui de l'électronique, celui des abrasifs, il porte le nom de *carborundum*.

a. Polytypisme

L'une des propriétés les plus remarquables de SiC est sa faculté d'exister sous de très nombreux *polytypes*. Ces différents polytypes, dont près de 200 ont été répertoriés, résultent d'un polymorphisme unidimensionnel favorisé par une faible énergie de formation des fautes d'empilement. Cette particularité nécessite un contrôle sévère des conditions thermodynamiques et cinétiques de croissance car si les conditions sont mal optimisées différents polytypes coexistent dans le même cristal. Les divers polytypes sont différenciés par la séquence d'empilement des bicouches de Si-C. Les longueurs des liaisons sont pratiquement identiques et la symétrie du cristal est déterminée par la périodicité de l'empilement. Trois types de symétrie existent : cubique (C), hexagonale (H) et rhomboédrique (R). Dans le réseau, chaque bicouche de SiC peut s'établir suivant trois positions différentes appelées arbitrairement A, B, C. Si la séquence d'empilement est du type ABC..., la structure cristallographique est de symétrie cubique, de type blende de zinc. Le polytype correspondant est le SiC-3C, ou β-SiC. Le chiffre 3 traduit le nombre de couches nécessaire à la périodicité. Si la séquence d'empilement est du type AB.., la structure cristallographique est de symétrie hexagonale, de type wurtzite, le polytype correspondant est le SiC-2H. Tous les autres polytypes sont des mélanges de ces derniers. Le polytype SiC-4H est constitué de nombres égaux de liaisons cubiques et hexagonales, la structure est du type ABAC...., la symétrie résultante est hexagonale. Le polytype SiC-6H est constitué de deux tiers de liaisons cubiques et un tiers de liaisons hexagonales, la structure est du type ABCACB....., la symétrie résultante est encore hexagonale. Les polytypes hexagonaux sont groupés sous le terme générique de α-SiC. La structure rhomboédrique la plus courante est du type SiC-15R, la cellule élémentaire comporte une séquence de 15 bicouches le long de l'axe c du cristal. L'essentiel des travaux concerne la structure cubique SiC-3C et les structures hexagonales SiC-4H, et -6H.

Ce polytypisme entraîne des propriétés spécifiques que seul présente le SiC. Le nombre d'atomes par cellule élémentaire par exemple, varie considérablement d'un polytype à l'autre, ce qui affecte d'autant le nombre de bandes d'énergie électronique et le nombre de branches de vibration de phonons. Cette diversité des structures de bandes d'énergie électronique et vibrationnelle affecte certaines propriétés des différents polytypes, notamment les propriétés optiques. Une autre spécificité liée au polytypisme est la diversité des sites de réseau non équivalents. Les deux polytypes qui présentent le plus d'intérêt à l'heure actuelle, SiC-4H et SiC-6H, ont respectivement 2 et 3 sites non équivalents. Il en résulte que le dopage de ces polytypes peut entraîner l'existence de 2 types de donneurs et accepteurs dans SiC-4H et 3 types de donneurs et accepteurs dans SiC-6H.

b. Elaboration

Les polytypes 4H et 6H peuvent être élaborés sous forme massive et par suite servir de substrats pour les homoépitaxies. Les autres polytypes présentent des problèmes d'élaboration et d'utilisation qui restent à résoudre. Le polytype 2H par exemple, peut difficilement servir de substrat car d'une part il est rare, et d'autre

part il est très instable. Il se transforme en un mélange de polytypes 6H et 3C aux températures d'épitaxie.

La structure microscopique des cristaux de polytypes 4H et 6H présente souvent des défauts très spécifiques qu'il est nécessaire d'éliminer pour réaliser des substrats de surface importante. Ces défauts sont des *microtubes* de section hexagonale, dont le diamètre varie entre 0,1 et 5 μm, et dont la densité varie entre 10 et 10^3 cm^{-2}. Ces microtubes se présentent sous la forme de trous qui traversent tout le cristal le long de l'axe de croissance. En outre, ils ont une fâcheuse tendance à se propager dans les couches épitaxiées. Les mécanismes, d'ordre thermodynamique, cinétique ou technologique, qui sont à l'origine de la formation de ces microtubes sont encore mal identifiés.

Des substrats conducteurs avec des résistivités de l'ordre de 10^{-3} Ωcm, comparables à celle du silicium, ainsi que des substrats semi-isolants avec des résistivités de l'ordre de 10^9 Ωcm à 500 K, sont actuellement disponibles. Les premiers sont utilisés dans la réalisation de composants de puissance verticaux, les second sont nécessaires à la réalisation de composants haute fréquence. Les diamètres des plaquettes de SiC-6H orienté suivant l'axe *c* et de SiC-4H désorienté de 8°, sont de l'ordre de 5 cm. Des plaquettes de SiC-4H de 7,5 cm de diamètre sont aussi réalisées. A titre de comparaison notons que le GaAs et le silicium sont respectivement disponibles en plaquettes de 10 et 30 cm de diamètre.

En ce qui concerne les couches épitaxiées, plusieurs méthodes et plusieurs types de substrat sont utilisés. Les méthodes les plus communes sont les dépôts en phase vapeur (CVD) sous pression atmosphérique ou sous basse pression. Le substrat le plus commun est le SiC massif. Les homoépitaxies sur substrats SiC-6H et SiC-4H sont réalisées sur des faces désorientées de quelques degrés par rapport à l'axe de symétrie du cristal. Les températures d'épitaxie sont le plus souvent de l'ordre de 1400 à 1600 °C. Les couches de symétrie 4H présentent des propriétés intrinsèques meilleures que leurs homologues de symétrie 6H. Néanmoins, la migration des défauts, microtubes et dislocations, depuis le substrat vers la couche épitaxiée, reste un problème majeur. La minimisation de ces défauts est une condition nécessaire au développement des composants de puissance et des circuits intégrés à base de SiC.

c. Dopage

Contrairement au silicium et autres composés III-V, le SiC peut difficilement être dopé par diffusion sélective. La raison en est simplement que les énergies de cohésion des atomes sont trop importantes. Les températures nécessaires à la diffusion des agents dopants, supérieures à 1800 °C, sont de ce fait beaucoup trop élevées. Des dégradations rédhibitoires de la surface du matériau et des couches de masquage sont inévitables.

Le processus de dopage sélectif généralement utilisé est l'implantation ionique. Mais ici encore, toujours en raison de l'importance des énergies de liaison, les températures d'implantation et surtout de recuit, sont très importantes (500 à 1700 °C). L'implantation à haute température est souvent préférée à l'implantation à température ambiante car cette dernière entraîne une importante amorphisation du matériau. Les dopages *n* et *p* du polytype 6H peuvent être réalisés respectivement par implantations d'azote à 700 °C et d'aluminium à 850 °C, suivies d'un recuit thermique de 30 minutes à 1650 °C. Des couches d'isolation de grande résistance sont sélectivement réalisées par implantation de vanadium, qui donne des centres profonds.

En ce qui concerne les couches épitaxiées, le dopage peut être effectué *in situ* pendant la croissance de chaque couche, en injectant les dopants spécifiques *n* ou *p* dans le réacteur d'épitaxie. Les sources de dopage *p* les plus communes sont le triméthyl-aluminium pour le dopage à l'aluminium et le diborane pour le dopage au bore. Les couches de type *n* sont réalisées par dopage à l'azote.

L'implantation ionique et le dopage en cours d'épitaxie sont les deux techniques qui permettent la réalisation de composants. Le dopage contrôlé pendant l'épitaxie présente l'avantage de ne pas introduire de défauts de réseau, qu'il est nécessaire de réduire ensuite par des recuits à haute température. En contre partie l'implantation ionique est une technique plus versatile qui autorise, par exemple, des dopages sélectifs.

Le choix du site occupé par un agent dopant est en grande partie déterminé par les tailles relatives de l'atome dopant et de l'atome substitué, silicium ou carbone. Les rayons de covalence du silicium et du carbone sont respectivement de 1,17 et 0,77 Å , ceux de l'aluminium et de l'azote sont respectivement de 1,26 et 0,74 Å. Il en résulte que le donneur azote occupe préférentiellement les sites carbone alors que l'accepteur aluminium occupe les sites silicium. Le bore qui a un rayon de covalence de 0,82 Å, peut occuper les sites carbone ou silicium. Le contrôle de l'incorporation de dopants peut être maîtrisé par la variation du rapport Si/C dans le réacteur. Par exemple le passage de 0,1 à 0,5 du rapport Si/C multiplie par 100 l'incorporation d'azote. Les gammes de dopage des couches épitaxiées de SiC-6H varient entre 10^{14} et 10^{20} cm^{-3} pour le dopage *n* et entre 10^{16} et 10^{18} cm^{-3} pour le dopage *p*.

Le dopage *n* est donc réalisé avec l'azote qui se met en site carbone. Mais dans un polytype donné tous les sites carbone ne sont pas équivalents. Dans SiC-4H, il existe dans la maille élémentaire, un site carbone à symétries cubique et un site carbone à symétrie hexagonale. Dans SiC-6H, la cellule élémentaire comprend trois sites carbone différents, deux à symétrie cubique et un à symétrie hexagonale. Les sites non équivalents ayant des environnements différents, les nivaux donneurs associés à l'azote sont différents suivant le carbone substitué. Dans SiC-4H on observe deux niveaux donneurs dont les énergies d'ionisation sont 102 meV pour le site cubique et 59 meV pour le site hexagonal. Dans SiC-6H, on observe trois

niveaux donneurs dont les énergies d'ionisation sont respectivement 137, et 142 meV pour les sites cubiques, et 83 meV pour le site hexagonal.

Le dopage *p* est réalisé avec l'aluminium ou le bore qui peuvent être introduits pendant la croissance ou après, par implantation ionique. L'aluminium se met en site silicium et crée un niveau accepteur relativement peu profond dont l'énergie d'ionisation, 211 meV dans le polytype 6H, et 225 meV dans le polytype 4H, varie peu avec la symétrie du site. Rappelons que le bore peut se mettre en site silicium ou carbone, en site silicium il donne un niveau accepteur relativement profond situé à 320 meV au-dessus de la bande de valence, en site carbone il donne un centre profond situé à 730 meV à l'intérieur du gap. Ces caractéristiques en font un agent dopant moins intéressant que l'aluminium.

Notons que les valeurs relativement élevées des énergies d'ionisation des agents dopants, donneurs et accepteurs, font qu'une partie des porteurs sont gelés à la température ambiante.

d. Contacts ohmiques

La réalisation de contacts ohmiques sur un semiconducteur à grand gap est toujours un problème difficile, or c'est une étape incontournable dans la fabrication d'un composant. En outre les domaines des hautes températures et des fortes puissances nécessitent que ces contacts restent stables dans les conditions extrêmes. La nature d'un contact métal-semiconducteur (Chap. 4) est conditionnée par la différence $e\phi_m$-$e\phi_s$ des travaux de sortie du métal et du semiconducteur. Dans SiC de type *n*, le niveau de Fermi étant voisin de la bande de conduction, le travail de sortie $e\phi_s$ est voisin de l'affinité électronique eχ, c'est-à-dire de l'ordre de 4,2 eV. Il est alors relativement facile de trouver un métal de travail de sortie comparable (Tab. 1-3), afin d'annuler la barrière de Schottky $E_{bn}=e\phi_m-e\chi$. Dans SiC de type *p* par contre, le niveau de Fermi étant voisin de la bande de valence, le travail de sortie $e\phi_s$ est voisin de la somme $E_g+e\chi$, c'est-à-dire de l'ordre de 7 eV. Il n'existe pas de métaux ayant des travaux de sortie comparables, et par conséquent le contact métal-SiC(*p*) se comportera plutôt comme une diode Schottky. Dans une diode Schottky de type métal-semiconducteur(*p*), la hauteur de la barrière est $E_{bp}=E_g+e\chi-e\phi_m$, et sa largeur décroît comme $N_a^{-1/2}$ où N_a représente la densité d'accepteurs ionisés. En fait, si en régime d'émission thermoélectronique par dessus la barrière le contact est redresseur, en régime de transfert par effet tunnel à travers la barrière il est par contre ohmique. Sa résistance varie alors comme $r_c \propto e^{E_{bp}/\sqrt{N_a}}$. On favorisera donc la réalisation de contacts ohmiques sur le matériau de type *p* en dopant fortement le semiconducteur au voisinage du contact, de manière à réduire la largeur de barrière, et de ce fait favoriser le transfert des porteurs majoritaires par effet tunnel à travers celle-ci. Enfin, notons que la différence d'électronégativité entre le silicium et le carbone est de l'ordre de 0,7 eV. L'ionicité de SiC situe donc ce matériau dans la gamme des semiconducteurs pour lesquels l'*indice de comportement* γ du contact métal-semiconducteur est voisin de 0,4. En d'autres

termes la densité d'états d'interface est relativement importante de sorte que le comportement du contact est conditionné de façon non négligeable par l'état de charge de ces états, qui amenuise très sérieusement la sensibilité du contact au travail de sortie du métal. Les propriétés de ce contact sont donc très sensibles au traitement de surface préalable à la réalisation du contact.

Les contacts ohmiques sur le type *p* sont réalisés avec un alliage Aluminium-Titane. Sur SiC dopé aluminium avec $N_a \approx 10^{19}$ cm^{-3} , la résistance de contact est typiquement de l'ordre de 10^{-5} Ωcm^2 . Sur le matériau de type *n* les contacts ohmiques les plus performants sont réalisés avec le nickel. Le siliciure de nickel (Ni_2Si) qui est formé après un court recuit à haute température (~1 mn à 1000 °C), présente une faible résistance de contact et une excellente stabilité lors des recuits ultérieurs à des températures supérieures à 500 °C. La résistance de contact est inférieure à 10^{-5} Ωcm^2 sur du matériau modérément dopé. Des contacts performants sont aussi réalisés avec des alliages de type nickel-chrome (80/20), ou nickel-titane (50/50). Ce dernier met a profit les fortes affinités du nickel pour le silicium et du titane pour le carbone, pour former un contact thermiquement stable de type nitrure de silicium / carbure de titane.

e. Caractéristiques

Les cristaux massifs de SiC non intentionnellement dopés, présentent une conductivité de type *n* en raison de la présence d'azote, qui contamine toujours la croissance.

Tous les polytypes ont une structure de bandes indirecte, avec un gap qui varie de E_g=2,4 eV pour la structure cubique SiC-3C, à E_g=3,35 eV pour la structure hexagonale SiC-2H. Les autres polytypes, qui sont des mélanges de ces deux là, ont des gaps intermédiaires, E_g=3,25 et 2,86 eV pour , SiC-4H, et SiC-6H, respectivement. Les différents polytypes ayant des gaps indirects, ils sont de mauvais candidats pour l'optoélectronique mais de bons candidats pour l'électronique de puissance.

Les mobilités des porteurs dépendent fortement du polytype, de la température, du dopage, et de la compensation. Dans les polytypes 3C, 4H, et 6H, la mobilité des électrons à la température ambiante est respectivement de 800, 1000 et 400 cm^2/Vs. Elle chute d'un facteur 10 quand la densité de donneurs varie de 10^{16} à 10^{19} cm^{-3} , ou quand la température passe de 300 à 600 K. En contre partie, dans le polytype 4H qui est le plus performant, elle atteint une valeur supérieure à 10^4 cm^2/Vs à 50 K. Il faut en outre noter qu'en raison de la symétrie du matériau la mobilité des porteurs est anisotrope. Cette anisotropie est évidemment nulle dans le polytype cubique, elle est faible dans le 4H et plus importante dans le 6H. La mobilité des trous est de l'ordre de 40, 115, et 100 cm^2/Vs dans les polytypes 3C, 4H, et 6H respectivement.

La masse effective des électrons est anisotrope. Dans SiC-4H les masses effectives parallèle et perpendiculaire à l'axe *c* sont respectivement $m_{//} = 0{,}29\ m_o$, $m_{\perp} = 0{,}42\ m_o$. La vitesse de saturation des électrons est pratiquement la même dans tous les polytypes, de l'ordre de 2.10^7 cm/s. Le champ de claquage est de 1,2 , 2 , et $3{,}8.10^6$ V/cm pour SiC-3C, -4H, et -6H respectivement, soit environ 10 fois celui du silicium. Enfin le coefficient de conductivité thermique de SiC, qui varie entre 3 et 5 W/cmK d'un polytype à l'autre, est supérieur à celui du cuivre et d'un ordre de grandeur supérieur à celui de GaAs.

Notons enfin qu'il n'existe pas à l'heure actuelle de produit permettant de réaliser un nettoyage chimique de la surface de SiC par voie humide. La technique utilisée est l'érosion ionique réactive RIE (Reactive Ion Etching). Un taux d'érosion de l'ordre de 100 à 1000 Å/mn est obtenu avec un plasma fluoré de type $CF_4/H_2/Ar$. La nécessité d'utiliser la technique RIE n'est pas une limitation à l'utilisation de SiC car la réduction de taille des composants entraîne une préférence généralisée des procédés secs par rapport aux procédés humides.

f. Oxydation et passivation

SiC est le seul semiconducteur composé sur lequel on peut faire croître thermiquement des films de SiO_2 de très bonne qualité. Lorsqu'une couche d'oxyde d'épaisseur *e* se développe à la surface, une épaisseur *e*/2 de SiC est consommée, le silicium forme avec l'oxygène l'oxyde SiO_2, le carbone s'échappe sous forme de CO. Le taux de croissance de la couche d'oxyde est beaucoup plus faible sur SiC que sur le silicium. En atmosphère humide à 1200 °C, le taux de croissance de SiO_2 en Å/mn$^{1/2}$, est de l'ordre de $1200\sqrt{t}$ sur Si, alors que sur SiC-6h il est de l'ordre de $220\sqrt{t}$ sur les faces carbone et $60\sqrt{t}$ sur les faces silicium. Les taux d'oxydation du SiC-4H sont comparables à ceux de SiC-6H.

Sur SiC de type *n* peu dopé, la densité d'états d'interface est de l'ordre de 10^{11} cm^{-2}eV^{-1}. Sur SiC de type *p*, les conditions sont moins favorables, en raison probablement de la présence d'aluminium à l'interface SiO_2/SiC.

Les applications potentielles immédiates de SiC concernent les domaines des hautes températures, hautes puissances et hautes fréquences, ainsi que tous les profits qui peuvent être tirés de la synergie entre ces différents domaines et celui des détecteurs optoélectroniques de courte longueur d'onde. Les grandes mobilités de ses porteurs, leur faible anisotropie, et la possibilité de réaliser des substrats de taille importante, font du polytype 4H le meilleur candidat à la réalisation de composants.

13.2.3. *Nitrures III-N – BN, AlN, GaN, InN*

Les nitrures III-N peuvent cristalliser sous les formes cubique (β-III-N) ou hexagonale (α-III-N). La phase hexagonale est la phase stable obtenue dans les conditions habituelles de croissance. La phase cubique de GaN par exemple, peut

être obtenue dans des conditions spécifiques d'homoépitaxie sur substrat GaN(001) ou d'hétéroépitaxie sur substrat GaAs(111). Le diagramme représentant les gaps des phases hexagonales en fonction des mailles est porté sur la figure (13-4). Les différents substrats d'épitaxie possibles sont aussi portés sur la figure. Le prototype des matériaux III-N est le GaN. AlN est principalement utilisé comme couche tampon dans les hétéroépitaxies où il assure une bonne nucléation de la croissance, et dans les alliages où il assure un ajustement du gap. L'alliage ternaire AlGaN est très utilisé, notamment dans les hétérostructures, car il est pratiquement en accord de maille avec GaN. BN d'une part et InN de l'autre ne sont pas très utilisés car ils présentent de forts désaccords de maille avec GaN. En outre InN qui n'est pas un semiconducteur à grand gap, est fortement concurrencé par les autres composés III-V, comme les arséniures ou les phosphures. Néanmoins les alliages de type InGaN à forte teneur en gallium, dont le gap se situe dans la gamme vert-bleu-violet du spectre visible, sont très utilisés dans les composants optoélectroniques.

a. Elaboration

Le GaN massif est très difficile à obtenir car son élaboration nécessite des températures de l'ordre de 1700 °C sous des pressions partielles d'azote de l'ordre de 20 kbars. Malgré ces conditions extrêmes, des plaquettes de GaN cristallin de 100 mm^2 présentant un taux de dislocations inférieur à 10^5 cm^{-2} sont réalisées. En raison de leur trop faible taille ces cristaux massifs ne sont pour l'instant pas utilisés industriellement comme substrats d'homoépitaxie, le GaN est généralement élaboré en couches minces par hétéroépitaxie en phase vapeur, sur différents types de substrats dont les mailles sont portées sur la figure (13-4).

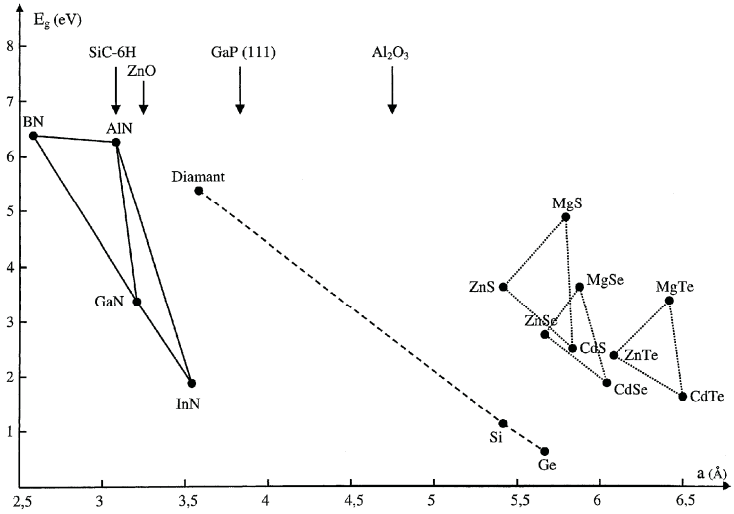


Figure 13-4 : Gaps et paramètres de maille des nitrures III-N et des composés II-VI à grand gap. Mailles de certains substrats utilisés dans l'épitaxie des nitrures

L'épitaxie peut être réalisée dans un réacteur HVPE (Hybride Vapour Phase Epitaxy), sur un substrat porté à 1000 °C, à partir d'un mélange gazeux $GaCl_3/NH_3$ comme sources de gallium et d'azote respectivement. Mais la vitesse de croissance élevée (~ plusieurs μm/mn) rend difficile le contrôle des épaisseurs de couches. En contre partie, cette vitesse permet la réalisation de couches épaisses de plusieurs centaines de microns. Cette technique permet ainsi de réaliser des couches tampon d'excellentes qualités dans les hétéroépitaxies.

La méthode MBE (Molecular Beam Epitaxy), la plus performante dans l'élaboration de couches minces de la plupart des composés semiconducteurs, est difficile à mettre en œuvre dans l'élaboration des nitrures comme GaN. La difficulté de générer suffisamment de radicaux azote pour réagir avec la source de gallium entraîne une vitesse de croissance très faible. Cette méthode présente néanmoins l'intérêt de travailler à basse température (~700 °C), ce qui permet de mieux maîtriser les phénomènes d'interdiffusion dans l'élaboration des hétérostructures. En outre elle autorise un meilleur contrôle des couches *in situ*. La méthode MBE standard, qui consiste à évaporer simultanément le gallium et l'azote, donne souvent naissance à des structures colonnaires caractérisées par une très mauvaise morphologie de surface. La méthode MEE (Migration Enhanced Epitaxy) qui consiste à alterner les évaporations de gallium et d'azote sur le substrat permet une meilleure migration des atomes de gallium sur la surface ce qui augmente le taux de croissance latérale. Cette technique réduit certes la vitesse de croissance mais donne de meilleures morphologies de surface que la méthode MBE standard. En outre elle facilite l'incorporation d'agents dopants ou d'indium pour la réalisation du composé ternaire InGaN.

La méthode industrielle la plus généralement utilisée est la méthode MOCVD (MetalOrganic Chemical Vapor Deposition), les sources de gallium et d'azote sont respectivement le triéthyl ou triméthylgallium ($(CH_2)_3Ga$, $(CH_3)_3Ga$) et l'ammoniac (NH_3), le substrat est porté à 1000 °C, la vitesse de croissance est de quelques μm/h. En raison du fort désaccord de maille qui existe entre le substrat et la couche cette dernière est le siège d'un taux important de dislocations. Une technique d'épitaxie sélective, la technique ELOG (Epitaxial Lateral OverGrowth) permet de réduire considérablement ce taux de dislocations. Cette technique, appelée aussi LEO (Lateral Epitaxial Overgrowth), consiste à déposer la couche sur un substrat sur lequel un masque, constitué par des bandes de SiO_2 séparées par des fenêtres, a préalablement été déposé. La croissance sur un tel substrat se développe d'abord verticalement sous forme de rubans au-dessus des fenêtres, ensuite ces rubans coalescent par croissance latérale au-dessus des bandes de SiO_2 pour former une couche continue. L'expérience montre que les régions qui poussent latéralement au-dessus des masques de SiO_2 présentent, comparativement aux autres régions, un taux de dislocations inférieur de plusieurs ordres de grandeur ($\sim 10^6$ cm^{-2}).

En ce qui concerne les substrats, plus d'une dizaine d'options ont été envisagées. Il va de soi que le GaN massif sera le mieux approprié quand des plaquettes de taille suffisante seront élaborées. En attendant, les substrats les plus

utilisés sont d'une part le saphir (Al_2O_3) qui est facilement disponible, peu cher, et de bonne qualité, et d'autre part le carbure de silicium (SiC) qui est mieux adapté en maille que le saphir, mais qui pour le moment est relativement cher.

Le désaccord de maille entre le plan de base de Al_2O_3 et celui de GaN est supérieur à 30%. En fait l'épitaxie se développe sur le sous-réseau constitué par les atomes d'aluminium, dont la maille élémentaire est tournée de 30° par rapport celle du saphir. Dans ces conditions le désaccord de maille est ramené à 16% environ. Malgré ce désaccord encore considérable, et une différence de coefficient de dilatation thermique de l'ordre de 35%, le saphir est néanmoins le substrat traditionnellement utilisé à l'heure actuelle. Le désaccord de maille entraîne cependant la présence d'une grande densité de dislocations ($\sim 10^{10}$ cm^{-2}).

En ce qui concerne le substrat SiC, les conditions d'épitaxie sont plus favorables. Le désaccord de maille n'est que de 3,5% avec GaN, et inférieur à 1% avec AlN. En outre, le SiC comme les nitrures, présente des coordinations tétraédriques et des surfaces polaires, de type silicium ou carbone. Les couches épitaxiales sont plus stables sur les surfaces silicium que sur les surfaces carbone. Les couches de GaN élaborées sur SiC, par l'intermédiaire d'une couche tampon de AlN, sont de meilleures qualités que les couches élaborées sur saphir, les mobilités des porteurs par exemple y sont deux fois supérieures. Les inconvénients actuels des substrats SiC sont d'une part leur coût, et d'autre part la difficulté de trouver des mélanges chimiques permettant de nettoyer correctement leur surface.

Le GaN est aussi élaboré sur divers autres substrats tels que Si ou GaAs, qui sont largement disponibles, ou ZnO avec lequel il présente un désaccord de maille de 2%. Divers oxydes sont aussi utilisés comme $MgAl_2O_4$ ou $LiGaO_2$. Ce dernier présente avec le GaN hexagonal un faible désaccord de maille (0,9%), mais les couches présentent des problèmes d'adhérence. Notons enfin que les homoépitaxies de GaN sur GaN, présentent des taux de dislocations inférieurs à 10^6 cm^{-2}.

b. Dopage

Le GaN non dopé est naturellement de type *n*. Ce dopage résiduel *n* a souvent été attribué à des vacances d'azote. En fait l'énergie de formation de ces vacances est très importante ce qui rend difficile la formation spontanée de ce type de défaut au cours de la croissance. Il semblerait que ce dopage *n* résulte plutôt d'une contamination, probablement à l'oxygène ou au silicium qui donnent des niveaux donneurs peu profonds. Des couches de GaN de forte résistivité, 10^4 Ωcm, ont été obtenues par la technique MBE. Le GaN est alors déposé à 800 °C sur une couche tampon de AlN préalablement déposée sur le substrat porté à la température de 650 °C.

Le dopage standard est réalisé par implantation ionique et recuit à haute température (>1000 °C). Il faut noter que les températures de recuit généralement mises en œuvre dans le dopage des semiconducteurs classiques, comme Si ou GaAs par exemple, correspondent en général aux 2/3 de leur température de fusion. Or

1000 °C correspond seulement à la moitié de la température de fusion de GaN. Ainsi, bien que l'activation des agents dopants soit obtenue avec des températures de recuit de 1100 °C, il est possible que des températures voisines de 1700 °C soient plus efficaces.

Les éléments du groupe IV sont potentiellement les dopants de choix pour les composés III-V, en site cation ils se comportent en donneurs, en site anion ils se comportent en accepteurs. Le carbone s'incorpore en site azote, alors que le silicium et le germanium s'incorporent en site gallium. L'inconvénient de ces impuretés amphotères est que les taux de dopage sont souvent limités par un phénomène d'autocompensation, à fortes doses certaines de ces impuretés s'incorporent à la fois en sites cation et anion.

Le dopage *n* est réalisé avec le silicium qui donne un niveau donneur peu profond dont l'énergie d'ionisation est de l'ordre de 60 meV. L'implantation ionique est réalisée avec des tensions d'accélération de l'ordre de 100 keV, et des doses d'implantation variant entre 10^{13} et 10^{16} cm^{-2}. Le silicium se met en site gallium à relativement basse température et devient électriquement actif après un recuit à haute température, qui annihile les antisites créés lors de l'implantation. Il faut noter que la solubilité du silicium dans GaN est importante. D'autre part le seuil d'amorphisation de GaN sous implantation ionique est élevé (2.10^{16} cm^{-2} à comparer à 4.10^{13} cm^{-2} pour GaAs). Ces caractéristiques permettent d'atteindre des taux de dopage *n* très importants.

Le dopage *p* pose généralement un problème dans les semiconducteurs à grand gap, car les énergies d'ionisation des accepteurs sont toujours relativement élevées. Ceci limite le nombre d'accepteurs ionisés, et par suite le nombre de trous, à la température ambiante. L'accepteur le moins profond dans GaN est le magnésium, son énergie d'ionisation est néanmoins comprise entre 150 et 170 meV. En outre la solubilité de Mg dans GaN et AlGaN est limitée. Ceci entraîne des concentrations de trous limitées, qui sont généralement insuffisantes pour assurer des contacts ohmiques de faible résistance. Il n'en demeure pas moins, qu'en l'état actuel des connaissances, le magnésium semble être le dopant *p* le plus performant.

c. Contacts ohmiques

La différence d'électronégativité entre le gallium et l'azote est de l'ordre de 1,9 eV. L'ionicité de GaN le situe donc dans la gamme des semiconducteurs pour lesquels l'*indice de comportement* γ du contact métal-semiconducteur est voisin de 1 (Fig. 4-12). En d'autres termes la densité d'états d'interface est relativement faible de sorte que le comportement du contact est conditionné par la différence des travaux de sortie du métal et du semiconducteur. Ainsi pour réaliser un contact ohmique sur du matériau de type *n* il suffit d'utiliser un métal dont le travail de sortie est inférieur à celui de GaN(*n*) (~4 eV), ce qui est relativement facile. Par contre la réalisation d'un contact ohmique sur GaN(*p*) nécessite un métal dont le

travail de sortie est supérieur à celui du semiconducteur, c'est-à-dire très important, ce qui est plus difficile à trouver.

Les contacts ohmiques sur le matériau de type *n* sont généralement réalisés par une métallisation de type Ti-Al, Ti-Ag, Ti-Au, ou même Ti-Al/Ni-Au, assortie d'un recuit thermique rapide à 900 °C pendant quelques secondes. Pendant le recuit, l'azote d'interface est extrait de GaN et réagit avec Ti pour former TiN. Cette migration crée parallèlement des vacances d'azote qui surdopent *n* la région sous le contact. Cette technique donne des contacts de faible résistance mais thermiquement peu stables. Les résistances de contact varient typiquement de quelques 10^{-3} Ωcm^2 à 10^{-5} Ωcm^2, pour des dopages variant de 10^{17} à 10^{19} cm^{-3}. Sur GaN dopé à quelques 10^{16} cm^{-3} les contacts sont difficilement ohmiques.

Sur le type *p* les contacts ohmiques sont plus difficiles à réaliser essentiellement en raison du fait que le travail de sortie de GaN de type *p* est considérable. Il est de ce fait difficile de trouver un métal approprié. Différents alliages sont utilisés, comme par exemple l'alliage Ni-Au recuit à 600 °C pendant 10 mn. Les plus faibles résistances de contact sont obtenues avec l'alliage Cr-Au recuit à 500 °C pendant 1 mn. Sur du GaN de type *p* avec une concentration de porteurs de l'ordre de 10^{20} cm^{-3}, la résistance de contact est de l'ordre de 10^{-4} Ωcm^2.

Sur du matériau modérément dopé (~10^{17} à 10^{18} cm^{-3}), aussi bien *n* que *p*, des contacts ohmiques de résistance raisonnable (~10^{-1} à 10^{-4} Ωcm^2) peuvent être réalisés avec l'alliage Cr-Ni-Au.

d. Caractéristiques

Les nitrures sont des composés robustes mécaniquement, et presque inertes chimiquement, ce qui les rend particulièrement stables à haute température et dans des environnements hostiles. En contre partie, cette stabilité complique d'une part leurs dopages et d'autre part les différentes étapes des processus d'élaboration des composants. Notons que les différents nitrures ne sont pas au même stade d'évolution. Si l'élaboration et le dopage de GaN sont relativement maîtrisés, il n'en est pas de même pour les autres nitrures. InN non intentionnellement dopé par exemple, présente une densité d'électrons de l'ordre de 10^{18} à 10^{19} cm^{-3}. En contre partie, AlN même dopé reste isolant avec une résistivité de l'ordre de 10^{10} Ωcm.

Le gap des nitrures est direct, sa valeur est respectivement 1,95, 3,39, 6,28, et 6,4 eV, pour InN, GaN, AlN, et BN. A la température ambiante, le composé quaternaire AlGaInN présente un gap direct dont la valeur varie, en fonction de la composition, de 1,95 à 6,28 eV, ce qui en fait un matériau de choix pour les composants optoélectroniques dans le domaine des courtes longueurs d'onde. La variation du gap du composé ternaire $Al_xGa_{1-x}N$ est donnée en fonction du paramètre de composition *x* par la loi classique

$$E_g = E_g^{GaN}(1-x) + E_g^{AlN}x - bx(1-x) \qquad (13\text{-}1)$$

avec un paramètre de non linéarité b=1,3 eV. Au niveau du gap du matériau le coefficient d'absorption optique est sensiblement constant avec la composition, il est de l'ordre de 7.10^4 cm^{-1} (Angerer H.-1997).

Les masses effectives des électrons sont de l'ordre de 0,1 , 0,2 , et 0,3 m_o dans InN, GaN, et AlN respectivement. Les masses effectives des trous sont de l'ordre de 0,5 , 0,8 et 1 m_o.

Le dopage n de GaN est réalisé avec le silicium, qui donne un niveau donneur situé à environ 60 meV au-dessous de la bande de conduction. La mobilité des électrons est de l'ordre de 1000 cm^2/Vs. Le dopage p est réalisé avec le magnésium, qui donne un niveau accepteur situé à environ 160 meV au-dessus de la bande de valence. La mobilité des trous à la température ambiante est de l'ordre de 50 cm^2/Vs.

Dans GaN, la densité de porteurs intrinsèques à la température ambiante est de l'ordre de 2.10^{-10} cm^{-3}, la vitesse de saturation des électrons est voisine de $2,5.10^7$ cm/s , le champ de claquage est de l'ordre de 5.10^6 V/cm. Le coefficient de conductivité thermique est comparable à celui du silicium, il est de 1,5 W/cmK. Enfin les coefficients de dilatation thermique sont de l'ordre de 5,6 et $7,7.10^{-6}$ K^{-1} suivant les axes a et c respectivement.

L'expérience montre que les performances radiatives des composants optoélectroniques à base d'arséniures ou de phosphures, comme GaAs, AlGaAs ou GaAsP, sont très sensibles à la densité de dislocations. Cette dernière devient gênante à partir de 10^4 cm^{-2} car la durée de vie des porteurs devient alors essentiellement non-radiative. Or dans les couches de nitrures, comme GaN, des densités de dislocations de l'ordre de 10^8 à 10^{10} cm^{-2} semblent peu affecter les performances optoélectroniques. Cette propriété pourrait être liée à la faible longueur de diffusion des porteurs compte tenu de la structure fine des couches, qui se présente sous la forme de colonnes verticales de monocristaux dépourvus de dislocations. La faible longueur de diffusion des porteurs cantonne ces derniers à l'intérieur des colonnes où le rendement quantique est excellent.

Tous les composés III-V, par le fait qu'ils sont non-centrosymétriques, sont piézoélectriques. Mais en ce qui concerne les nitrures cette piézoélectricité est très importante car la liaison III-N est fortement polarisée, les électrons sont essentiellement localisés sur l'atome d'azote,. Les constantes piézoélectriques des nitrures sont près de dix fois supérieures à celles des composés III-V et II-VI conventionnels. AlN par exemple a les constantes piézoélectriques les plus élevées de tous les composés à liaisons tétraédriques, elles sont seulement trois fois inférieures à celles des pérovskites ferroélectriques. En régime linéaire la polarisation piézoélectrique P est reliée au tenseur de contrainte ε par la relation $P_i = e_{ij}\ \varepsilon_j$ avec sommation sur les indices répétés. Les composantes du tenseur de piézoélectricité des différents composés III-V sont données dans le tableau (13-1).

Afin de comparer les constantes piézoélectriques des nitrures, dont la structure est hexagonale, avec celles des autres composés III-V, dont la structure est cubique, le tenseur piézoélectrique des composés cubiques a été transformé par une rotation du système de coordonnées amenant l'axe z suivant la direction cristallographique (111). Ainsi, les composantes e_{33} et e_{31} de ces matériaux sont reliées à leur constante piézoélectrique unique e_{14} par les relations $e_{33} = 2e_{14}\sqrt{3}$ et $e_{31} = -e_{14}\sqrt{3}$.

TABLEAU 13.1 : Composantes du tenseur piézoélectrique des composés III-V. (en C/m^2). Pour les matériaux cubiques $e_{14} = e_{33}/2\sqrt{3}$**.** *(Bernardini F. 1997)*

	AlN	GaN	InN	AlP	GaP	InP	AlAs	GaAs	InAs	AlSb	GaSb	InSb
e_{33}	1,46	0,73	0,97	0,04	-0,07	0,04	-0,01	-0,12	-0,03	-0,04	-0,12	-0,06
e_{31}	-0,60	-0,49	-0,57	-0,02	0,03	-0,02	0,01	0,06	0,01	0,02	0,06	0,03

Cette forte piézoélectricité peut être exploitée dans la réalisation de modulateurs acousto-optiques. En déposant des peignes interdigités sur la surface d'une couche de GaN par exemple, il est possible de générer des ondes acoustiques stationnaires de surface. Ces ondes acoustiques consistent en une alternance de régions dilatées et comprimées. Les régions comprimées ont un indice de réfraction plus élevé que les régions dilatées. Ces régions forment donc un réseau qui peut réfracter un faisceau laser traversant la couche de GaN. Il suffit alors de changer la fréquence des ondes acoustiques pour moduler l'angle de réfraction du faisceau.

Notons enfin en ce qui concerne AlN que, si les faces terminées par l'azote et passivées à l'hydrogène ont une affinité électronique faible mais néanmoins positive, les faces terminées par l'aluminium ont quant à elles, une affinité électronique négative. Cette dernière propriété ouvre un champ d'application potentiel dans le domaine des écrans plats, par la réalisation par exemple de dispositifs à émission de champ.

e. Interfaces entre les différents nitrures

GaN et AlN ont des mailles voisines, mais InN présente avec ces derniers des désaccords paramétriques conséquents. Les hétérostructures réalisées avec ces matériaux sont donc soumises à des contraintes d'interface importantes qui conditionnent partiellement la distribution des discontinuités de gap entre les bandes de valence et de conduction. Compte tenu de ces effets de contrainte, les diagrammes énergétiques calculés, des hétérojonctions AlN/GaN, AlN/InN, et GaN/InN réalisées en couches pseudomorphiques sur substrat AlN sont représentées sur la figure (13-5). Toutes ces hétérojonctions sont de type I.

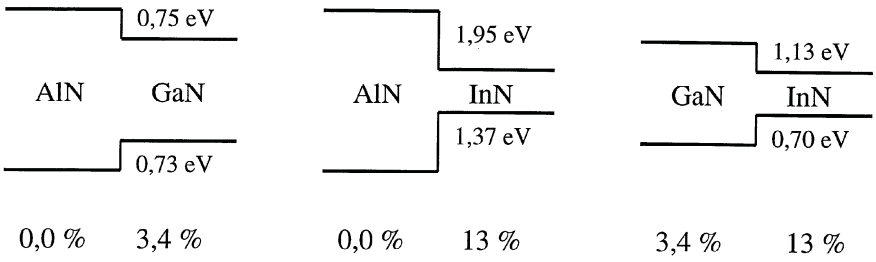


Figure 13-5 : Diagrammes énergétiques calculés de différentes hétérojonctions déposées en épitaxie pseudomorphique sur substrat AlN. Les désaccords paramétriques avec le substrat de AlN sont précisés. (Bernholc J. - 1996)

Le différents paramètres spécifiques des semiconducteurs à grand gap sont résumés sur le tableau (13-2) où ils sont comparés à ceux du silicium et de GaAs.

TABLEAU 13.2 : Paramètres spécifiques des semiconducteurs à grand gap. Gap E_g (d=direct, i=indirect), constante diélectrique ε_r, mobilités μ, champ de claquage E_c, vitesse de saturation des électrons v_s, conductivité thermique K, dilatation thermique α.

	E_g (eV)	ε_r	μ_n (cm^2/Vs)	μ_p (cm^2/Vs)	E_c (10^6 V/cm)	v_s (10^7 cm/s)	K (W/cmK)	α (10^{-6} K^{-1})
Si	1,15 (i)	11,8	1350	450	0,3	1	1,54	2,6
GaAs	1,42 (d)	12,8	8500	400	0,4	1	0,5	5,8
SiC-3C	2,40 (i)	9,6	800	40	1,2	2	3,2	
SiC-4H	3,26 (i)	10	1000/650	115	2	2	3,7	
SiC-6H	2,86 (i)	9,7	400/50	100	3,8	2	4,9	2,9
AlN	6,28 (d)	8,5		15			2	4,2/5,3
GaN	3,39 (d)	9	1000	50	5	2,5	1,5	5,6/7,7
Diamant	5,45 (i)	5,5	2200	1600	5,6	2,7	20	1

13.3. Composants optoélectroniques de courte longueur d'onde

Dans le domaine des composants optoélectroniques de courte longueur d'onde, certes la valeur de son gap fait de SiC un matériau relativement bien adapté, mais il est, par rapport aux nitrures, fortement pénalisé par toute une série de propriétés évidentes. Il est à gap indirect donc à, faible rendement. Les nitrures par contre ont un gap direct et par conséquent un bon rendement radiatif, leur coefficient d'absorption est de l'ordre de 4.10^4 cm^{-1} pour des rayonnements de longueur d'onde inférieure à 360 nm. En outre, la possibilité de réaliser des alliages ternaires ou quaternaires élargit considérablement le champ d'application des

nitrures. La simple modulation de la concentration d'aluminium permet par exemple d'ajuster la longueur d'onde de coupure de l'alliage AlGaN à une valeur comprise entre 360 et 200 nm. Enfin les nitrures offrent plus de versatilité par la réalisation possible d'hétérojonctions et au-delà d'hétérostructures notamment à confinement quantique. Il résulte de tout ceci que la grande majorité des composants optoélectroniques de courte longueur d'onde concerne essentiellement les nitrures.

13.3.1. *Détecteurs de rayonnement ultraviolet*

La détection des rayonnements ultraviolets (λ<400 nm) présente de nombreuses applications tant dans les domaines civils que militaires. La plupart de ces applications nécessitent des détecteurs sensibles au rayonnement UV mais en même temps insensibles aux rayonnements visible et infrarouge qui peuvent constituer un fond important dans lequel se trouve la composante UV à détecter sélectivement. C'est par exemple le cas de la détection d'un signal UV en présence d'un important rayonnement solaire, le détecteur doit être sensible au rayonnement UV sans pour autant être aveuglé par le soleil.

Les détecteurs UV traditionnels sont des tubes photomultiplicateurs dont le principe consiste à multiplier par émissions secondaires à l'aide d'une cascade d'anodes, le courant de photoélectrons émis par une photocathode sensible au rayonnement à détecter. Les qualités essentielles des photomultiplicateurs sont leur grande sensibilité associée à un faible bruit, en outre leur insensibilité au rayonnement solaire peut facilement être obtenue par un choix judicieux du matériau constituant la photocathode. Ce matériau est généralement un métal alcalin, ou une association de métaux alcalins, choisis en raison de la faible valeur de leur travail de sortie. Les défauts majeurs des photomultiplicateurs, notamment dans certaines applications militaires, sont liés au principe même de leur fonctionnement, sous vide et sous fortes tensions d'accélération. Ils nécessitent la réalisation de tubes à vide qui sont à la fois fragiles et encombrants, en outre ils sont sensibles aux champs magnétiques, enfin ils sont souvent chers. L'approche alternative a été d'utiliser des photodiodes *pin* ou Schottky au silicium. Leurs avantages sont ceux de tous les composants solides par rapport aux tubes à vide, ils sont de petite taille et nécessitent des faibles tensions de polarisation. Leurs inconvénients sont essentiellement liés aux spécificités du silicium. Le silicium est un semiconducteur à gap indirect, de sorte que son rendement radiatif est faible, ce qui entraîne une faible sensibilité du détecteur. Mais surtout, le maximum de sensibilité du silicium se situe au voisinage de 700 nm, ce qui le rend sensible au rayonnement solaire. L'utilisation de ces détecteurs UV en présence du soleil nécessite donc l'adjonction de filtres bloquant les composantes visible et infrarouge. Le dispositif devient alors à la fois complexe, volumineux et cher. Le détecteur idéal doit donc allier la compacité du détecteur solide et l'insensibilité au rayonnement solaire du photomultiplicateur. Les semiconducteurs à grand gap constituent dans ce domaine les matériaux de choix.

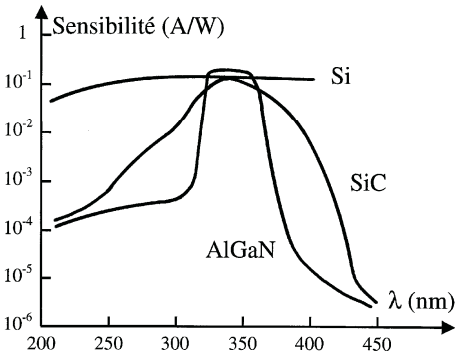


Figure 13-6 : Réponses spectrales comparées

Une nouvelle génération de détecteurs UV est basée sur l'utilisation des semiconducteurs à grand gap (E_g>3 eV). Contrairement au silicium, qui avec son gap de 1,15 eV est naturellement sensible aux composantes visibles et infrarouges du rayonnement solaire, les semiconducteurs à gap supérieur à 3 eV, qui ne peuvent détecter que des rayonnements de longueur d'onde inférieure à 400 nm, sont naturellement protégés du soleil. Les deux types de matériau susceptibles d'être opérationnels dans ce genre d'application sont d'une par le SiC dont le gap varie entre 2,4 et 3,26 eV selon le polytype, et d'autre part les nitrures de type AlGaN dont le gap varie entre 3,39 et 6,28 eV suivant la concentration d'aluminium.

Pour les raisons que nous avons déjà évoquées, l'essentiel des détecteurs UV concerne les nitrures. Néanmoins des photodiodes à jonction *pn* et à diodes Schottky, sensibles dans la gamme 200-400 nm, ont été réalisées avec SiC-6H. Les réponses spectrales typiques de photodétecteurs UV à base de silicium, de nitrures et de SiC sont représentées sur la figure (13-6). Le maximum de sensibilité est de l'ordre de 0,1 à 0,2 A/W dans tous les cas, mais les nitrures sont les seuls à présenter un taux de réjection important du rayonnement solaire (~10^4).

a. Cellules photoconductrices à base de nitrures III-N

Le principe de fonctionnement de la cellule photoconductrice est détaillé dans le chapitre 9 (Par. 9.2.3). La gamme de sensibilité de la cellule est définie par le gap du matériau, ses performances sont conditionnées par différents paramètres dont la conductance d'obscurité liée au dopage résiduel, le coefficient d'absorption optique et la durée de vie des photoporteurs.

Diverses cellules photoconductrices à base de nitrures ont été réalisées, par épitaxie MOCVD ou MBE, sur substrats saphir, SiC-6H ou Si(111), avec des performances variables. Dans

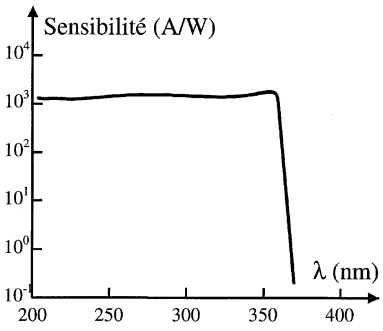


Figure 13-7 : Cellule GaN

le dépôt par MOCVD sur substrat saphir par exemple, l'élément actif est une couche de GaN de haute résistivité, de 0,8 µm d'épaisseur ($\alpha d \approx 3$), épitaxiée sur le saphir par l'intermédiaire d'une couche tampon de AlN de 0,1 µm d'épaisseur. La structure est de type interdigitée et correspond à une longueur équivalente de 10 µm sur une largeur effective de 50 cm. La réponse spectrale typique de ce genre de cellule est représentée sur la figure (13-7). Elle est pratiquement constante entre 200 et 365 nm, avec une valeur de l'ordre de 2000 A/W. Elle chute brutalement de plus de trois ordres de grandeur aux environs de 365 nm, ce qui rend la cellule insensible au rayonnement visible. Le temps de réponse de la cellule est de l'ordre de la milliseconde. Il faut noter que ce temps de réponse est anormalement important compte tenu de la faible valeur de la durée de vie des porteurs dans le matériau, qui rappelons-le a un gap direct. Les cellules réalisées par MBE avec du GaN semi-isolant ont des temps de réponse beaucoup plus courts (~20 ns), mais avec une sensibilité plus faible (~100 A/W).

L'utilisation de l'alliage AlGaN permet de moduler la réponse spectrale du détecteur et plus particulièrement son seuil de coupure. Les réponses spectrales de divers détecteurs réalisés avec différentes concentrations d'aluminium sont portées sur la figure (13-8). La longueur d'onde de coupure passe de 365 nm pour le GaN à 260 nm pour un alliage à 50% d'aluminium. La sensibilité de ces cellules reste limitée (~0,1 A/W) par rapport aux cellules à GaN. Ceci résulte probablement du fait que le dopage résiduel relativement important de l'alliage (~10^{18} cm^{-3}) diminue fortement la durée de vie des photoporteurs.

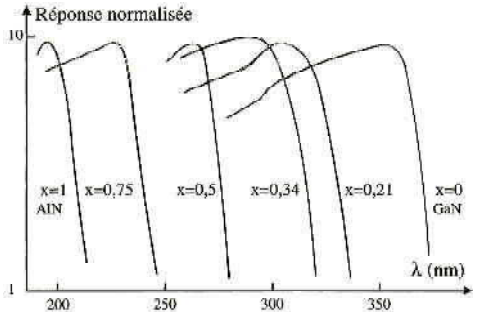


Figure 13-8 : Réponses des cellules $Al_xGa_{1-x}N$

b. Cellules photovoltaïques à base de nitrures III-N

Des cellules photovoltaïques à base de diode Schottky et de jonction *pn* ont été réalisées. La figure (13-9) représente la réponse spectrale d'une photodiode Schottky de type Pd/GaN(n^-). La cellule est excitée à travers la couche métallique semi-transpa-

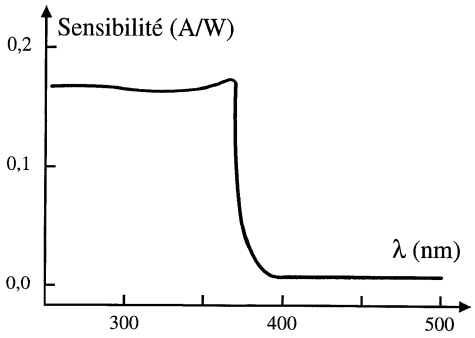


Figure 13-9 : Photodiode Schottky

rente, le photo-courant est de type génération dans la zone de charge d'espace (Par. 9.2.4). La couche de GaN reçoit le contact ohmique à travers une interface dopée n^+. La réponse spectrale est pratiquement constante entre 250 et 365 nm, avec une valeur de l'ordre de 0,17 A/W. Elle chute brutalement aux environs de 365 nm ce qui rend le détecteur insensible au rayonnement visible. Le temps de réponse de la photodiode est de l'ordre de 100 ns.

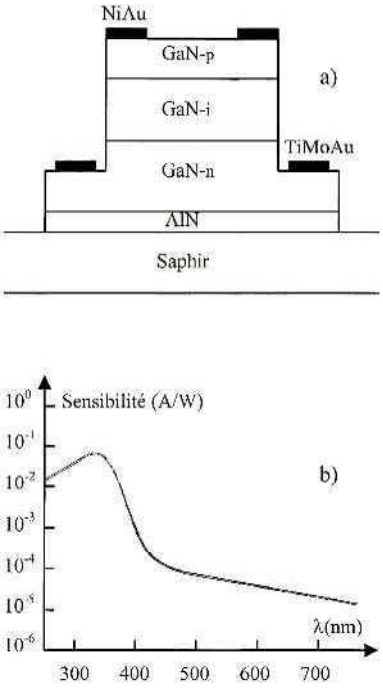


Figure 13-10 : Photodiode PIN

Des photodiodes *pin* à GaN ont été réalisées par épitaxie sur saphir. La structure de la photodiode et sa réponse spectrale sont représentées sur la figure (13-10).

La structure consiste en une couche de GaN de 1 µm d'épaisseur, dopée *n* au silicium avec un taux de 10^{18} cm^{-3}, déposée par MBE sur un substrat de saphir à travers une couche tampon de AlN. Cette couche est suivie d'une couche non intentionnellement dopée avec une densité résiduelle d'électrons de l'ordre de 10^{16} cm^{-3}. La structure est complétée par une couche de GaN de 3000 Å, dopée *p* au magnésium avec un taux de dopage de 5.10^{17} cm^{-3}. La réponse spectrale fait apparaître un pic de sensibilité de 0,1 A/W à 360 nm avec un taux de réjection du rayonnement visible compris entre 10^3 et 10^4. La réponse en puissance est linéaire entre quelque 10^{-10} et 10^{-6} W.

c. Phototransistor HFET GaN-AlGaN

La structure du phototransistor et sa réponse spectrale sont représentées sur la figure (13-11).

La couche active de GaN, non intentionnellement dopée, est déposée sur le substrat de saphir par l'intermédiaire d'une

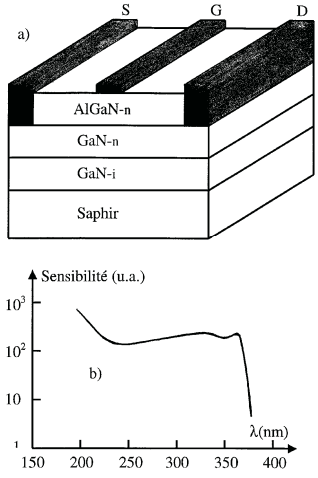


Figure 13-11 : Phototransistor HFET GaN/AlGaN. a) Structure. b) Réponse spectrale *(Asif Khan.-1995)*

couche tampon de GaN semi-isolant. L'épaisseur et le dopage de la barrière de AlGaN sont respectivement de l'ordre de 250 Å et 4.10^{18} cm^{-3}. La longueur de la grille est de 0,2 μm. Le phototransistor est exposé au rayonnement à travers le substrat de saphir. Les photoporteurs sont créés dans le GaN, à la fois dans la couche tampon semi-isolante et dans la couche active dopée *n*. Les électrons se thermalisent dans le canal du transistor où ils sont évacués vers le drain par la tension drain-source. La réponse spectrale (Fig. 13-11-b) présente un maximum de l'ordre de 3000 A/W, au voisinage de 325 nm, pour une tension drain-source de 10 V et une tension grille de 1 V. Elle chute brutalement de deux ordres de grandeur aux environs de 365 nm ce qui rend le phototransistor insensible au rayonnement visible.

13.3.2. *Emetteurs bleus*

Les émetteurs de rayonnement à semiconducteurs sont d'une part les diodes électroluminescentes LED (Light Emitting Diode) et d'autre part les diodes laser LD (Laser Diode). Pour la réalisation de tels émetteurs dans la gamme du bleu, trois types de semiconducteur ont a priori un gap dont la valeur est adaptée, les composés II-VI comme ZnS et ZnSe, les différents polytypes de SiC et enfin les nitrures III-N dont le prototype est le GaN. Les matériaux II-VI présentent un inconvénient majeur qui résulte de leur faible énergie de cohésion et du fait qu'ils sont le siège d'une densité importante de défauts, les émetteurs réalisés jusqu'ici ont une trop faible durée de vie. Le SiC quant à lui présente l'inconvénient rédhibitoire d'avoir un gap indirect, son rendement radiatif est de ce fait beaucoup trop faible. Néanmoins, bien que leur rendement ne soit que de l'ordre de 10^{-3} %, des LED bleues à base de SiC-6H ont depuis longtemps été réalisées et commercialisées jusqu'à ces dernières années car il n'existait pas d'alternative. Mais la maîtrise de la fabrication et du dopage des nitrures III-N, associée à la versatilité qu'offrent les alliages ternaires et quaternaires de ces matériaux, a complètement changé les données du problème. Les émetteurs de courte longueur d'onde sont désormais réalisés à base de nitrures. Le gap du composé quaternaire AlGaInN est d'une part direct, et d'autre part varie de 1,95 à 6,28 eV ce qui permet la réalisation d'émetteurs jaunes, verts, bleus et même ultraviolets.

a. Diodes électroluminescentes –LED

La maîtrise de diodes électroluminescentes bleues, vertes et rouges permet aujourd'hui de réaliser des dispositifs émettant toutes les couleurs du spectre visible, y compris des émetteurs de lumière blanche ayant de meilleurs rendements que les lampes à incandescence. Les émetteurs de lumière blanche sont réalisés soit par mixage des émissions bleue et verte de diodes à InGaN et de l'émission rouge de diodes à AlGaAs, soit par excitation de luminophores par une diode bleue.

Les premières LED bleues ont été réalisées sur la base de doubles hétérojonctions (DH) de type InGaN/AlGaN (Nakamura S.-1994). La zone active est la couche de InGaN doublement dopée Si et Zn. Le spectre d'émission de ces diodes est représenté sur la figure (13-12) pour plusieurs valeurs du courant

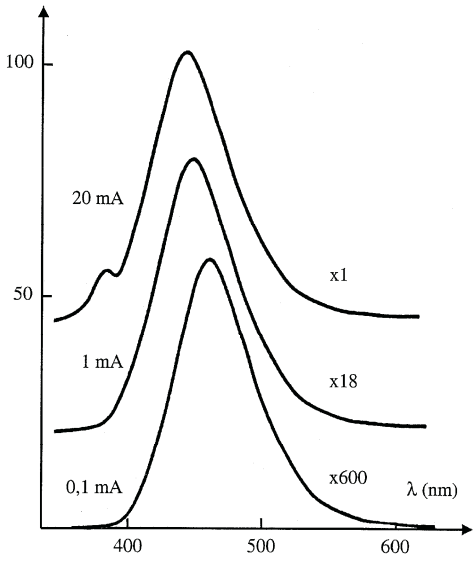


Figure 13-12 : Spectre d'émission de diodes électroluminescentes bleues à double hétéro-jonction InGaN/AlGaN, pour différentes valeurs du courant d'excitation *(Nakamura S.-1998)*

d'excitation. Quand l'excitation augmente, la puissance lumineuse augmente proportionnellement et le spectre se déplace vers les courtes longueurs d'onde. Pour un courant de 20 mA la raie d'émission est centrée à 447 nm avec une largeur à mi-hauteur de l'ordre de 70 nm, la puissance lumineuse émise est de 3 mW. Le rendement externe est de l'ordre de 5,4 %.

Des diodes électroluminescentes bleues, vertes et jaunes à simple puits quantique de InGaN ont été réalisées sur la base de la structure type représentée sur la figure (13-13). Pour la diode verte, représentée sur la figure, la couche active est un puits de $In_{0,45}Ga_{0,55}N$ non dopé, de 30 Å d'épaisseur, en sandwich entre une barrière de $Al_{0,2}Ga_{0,8}N$ de 1000 Å d'épaisseur dopée *p* au magnésium, et une barrière de GaN de 4 µm dopée *n* au silicium. La diode est épitaxiée sur un substrat de saphir à travers une couche tampon de GaN de 300 Å d'épaisseur.

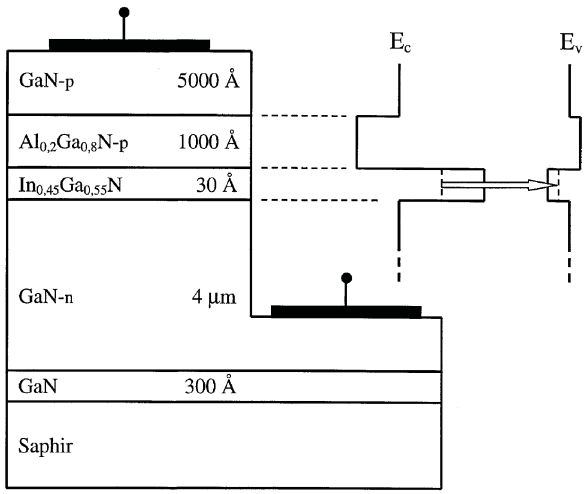


Figure 13-13 : Structure et diagramme énergétique d'une diode électroluminescente verte à simple puits quantique de InGaN *(Nakamura S.-1998)*

Les spectres d'émission de ces diodes sont représentés sur la figure (13-14) pour un courant d'excitation de 20 mA. Ces spectres, qui sont obtenus par le seul changement de la fraction molaire d'indium dans le matériau puits, sont centrés à 450 (bleu), 525 (vert) et 600 nm (jaune). Quand le taux d'indium augmente la largeur du spectre augmente, probablement en raison de la plus grande inhomogénéité de la couche.

L'origine de l'émission, dans ces structures à base d'alliage ternaire InGaN, n'est pour l'instant pas totalement élucidée. Cependant, on constate que le spectre d'émission est nettement décalé vers les grandes longueurs d'onde par rapport au gap de l'alliage. On peut donc légitimement penser que cette émission résulte de la recombinaison de porteurs, ou d'excitons, piégés dans des minima de potentiel associés à des états localisés dans le plan du puits quantique. La nature de ces états est probablement liée à des fluctuations d'alliage, qui jouent le rôle de boîtes quantiques à zéro dimension et qui de ce fait augmentent la force d'oscillateur des transitions radiatives.

L'inconvénient des LED réalisées sur substrat saphir réside dans la nature isolante du substrat. L'alimentation électrique de la diode nécessite un usinage spécifique (Fig. 13-13) avec deux contacts de surface et deux fils de connexion. Pour pallier cet inconvénient des LED à base de nitrures sont aussi réalisées sur substrat SiC conducteur. Outre la conductivité thermique de SiC, 10 fois supérieure à celle du saphir, l'intérêt du substrat conducteur réside dans le fait que la diode présente une configuration verticale. La structure est comparable à celle des diodes à base de GaAs ou de GaP, le substrat sert d'électrode (Fig. 13-15).

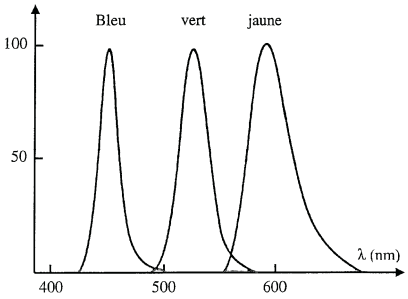


Figure 13-14 : Spectres d'émission de diodes électroluminescentes bleue, verte et jaune à simple puits quantique de InGaN *(Nakamura S.-1998)*

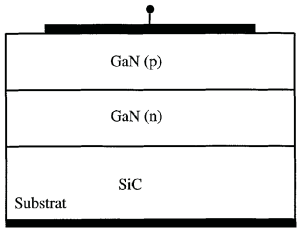


Figure 13-15 : Diode électrolumi-nescente à GaN sur substrat SiC

b. Diodes laser – LD

Les premiers lasers bleus ont été réalisés sur la base de structures à multipuits quantiques, sur substrat saphir, en utilisant le composé ternaire InGaN comme couche active (Nakamura S.-1996). Leur durée de vie réelle dépasse actuellement 4000 heures, avec une émission de 2 mW à la température ambiante. Des tests de fonctionnement de plus de 1100 heures à 50 °C permettent d'estimer leur durée de vie en régime normal à plus de 10 000 heures. Les performances actuelles de ces lasers bleus sont caractérisées par un courant de seuil de 1500 A/cm^2 sous une tension de fonctionnement de 4 à 5 V, la puissance de sortie est de l'ordre de 50 mW. Leur durée de vie est suffisante pour justifier leur commercialisation, leur puissance est suffisante pour permettre la lecture des disques optiques DVD.

La structure de ces diodes laser est du type à confinements séparés SCH (Fig. 12-17-d). Les photons sont confinés dans un guide d'onde de GaN par des barrières d'indice constituées de superréseaux AlGaN/GaN. Les porteurs sont confinés dans des multipuits quantiques de type $In_xGa_{1-x}N/In_yGa_{1-y}N$. Cette structure et son diagramme énergétique sont représentés sur la figure (13-16). Le laser est épitaxié sur un substrat de saphir à travers une couche tampon de GaN de 300 Å suivie d'une couche de GaN(n) de 10 µm d'épaisseur, élaborée par croissance latérale ELOG (Epitaxial Laterally OverGrown) au-dessus d'un masque de SiO_2. Le rôle de cette couche est de réduire le taux de dislocations afin d'augmenter la durée de vie du laser.

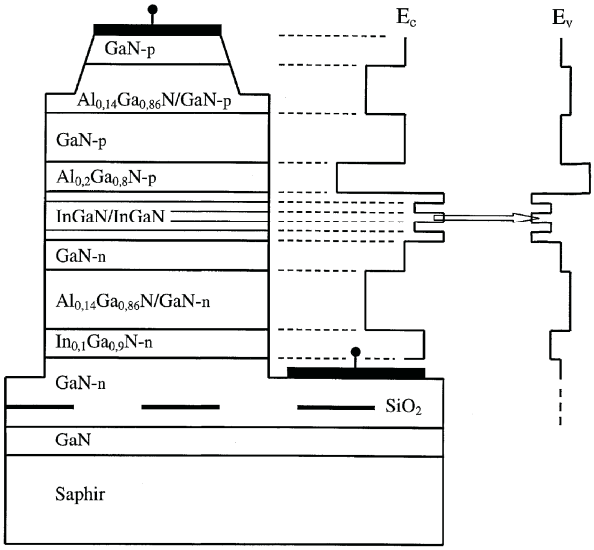


Figure 13-16 : Structure et diagramme énergétique du laser bleu à multipuits quantiques de InGaN *(Nakamura S.-1997)*

La région active du laser est une structure à multipuits quantiques, constituée de 4 puits de 35 Å de $In_{0,15}Ga_{0,85}N$ non dopé, séparés par des barrières de 70 Å de $In_{0,02}Ga_{0,98}N$ non dopé. Cette structure est insérée dans un puits de GaN d'épaisseur totale 0,23 μm. Ce puits est dopé *n* au silicium sur 0,1 μm au-dessous des multipuits, et *p* au magnésium sur 0,1 μm au-dessus. Il est bordé par des barrières d'indice constituées par des superréseaux contraints de type $Al_{0,14}Ga_{0,86}N$/GaN à modulation de dopage. Ces superréseaux contraints, avec des épaisseurs de couches inférieures aux épaisseurs critiques, s'accordent mieux à la maille du GaN avec lequel ils sont en contact, que l'alliage AlGaN d'indice équivalent. Leur dopage permet en outre de réduire la chute ohmique aux bornes de la diode. Une couche de 200 Å de $Al_{0,2}Ga_{0,8}N$ dopée *p* au magnésium, sépare les multipuits quantiques de la couche de GaN(*p*). La structure à multipuits quantiques assure le confinement quantique des porteurs, le puits de GaN assure le confinement des photons. La longueur de la cavité est de 550 μm, les faces réfléchissantes sont réalisées par des multicouches diélectriques quart-d'onde TiO_2/SiO_2.

Le courant de seuil J_o du laser est de l'ordre de 100 mA, ce qui correspond a une densité de courant de l'ordre de 4000 A/cm^2. Le spectre d'émission est représenté sur la figure (13-17) pour deux valeurs du courant d'excitation. Pour $J=J_o$ le spectre d'émission est constitué d'une série de raies distantes de 0,42 Å ($\Delta E \approx 0,3$ meV), sélectionnées par la cavité Pérot-Fabry. Les différentes raies apparaissent groupées en sous-bandes, séparées d'environ 2,5 Å ($\Delta E \approx 2$ meV). L'origine de ces sous-bandes n'est pas clairement identifiée, mais résulte très probablement du confinement quantique des états électroniques. Pour $J=1,2\ J_o$ le laser devient monomode et émet environ 2 mW au voisinage de 406 nm.

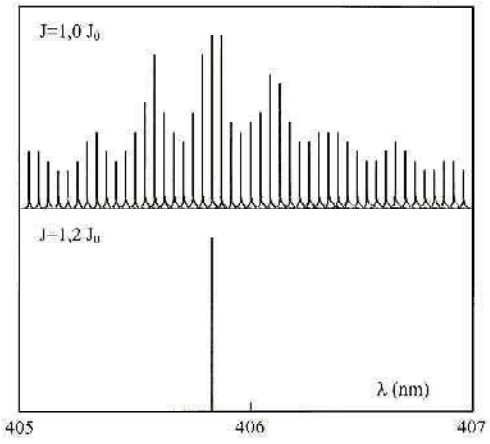


Figure 13-17 : Spectre d'émission du laser bleu à multipuits quantiques de InGaN, en régime continu à la température ambiante, pour deux valeurs du courant d'excitation
(Nakamura S.-1998)

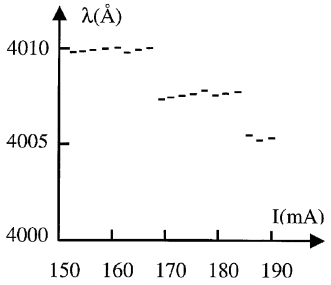


Figure 13-18 : Déplacement de la raie d'émission du laser bleu à multipuits quantiques de InGaN, en fonction du courant d'excitation
(Nakamura S.-1998)

La figure (13-18) représente l'évolution du pic d'émission avec le courant d'excitation. Lorsque le courant augmente, le pic se déplace d'une part de façon continue vers les grandes longueurs d'onde, et d'autre part par sauts discrets vers les courtes longueurs d'onde. L'évolution vers les grandes longueurs d'onde résulte simplement de la diminution des gaps des semiconducteurs avec l'élévation de température résultant de l'excitation. L'évolution par sauts traduit la variation de la courbe de gain qui résulte des participations successives des différentes minibandes de quantification. Les sauts de 2,5 Å résultent des participations des minibandes successives. Des sauts de plus faible amplitude (~0,4 Å), correspondent au changement de mode de la cavité, à l'intérieur d'une minibande.

13.3.3. *Emetteurs ultraviolets*

Des diodes électroluminescentes émettant dans l'ultraviolet ont été réalisées sur la base de doubles hétérojonctions InGaN/AlGaN. La structure de ces diodes est représentée sur la figure (13-19).

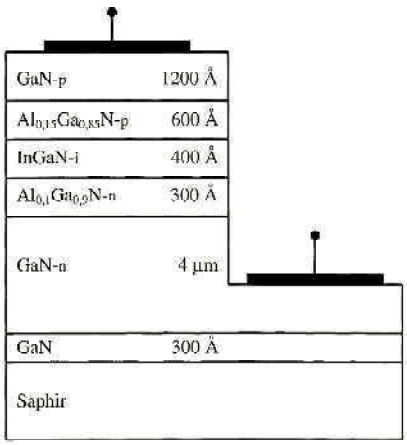


Figure 13-19 : Structure d'une diode électroluminescente UV à double hétéro-jonction InGaN/AlGaN
(Mukai T.-1997)

La diode est épitaxiée sur un substrat de saphir à travers une couche tampon de GaN de 300 Å, déposée à basse température. La zone active de la diode est une couche de InGaN non dopée, très pauvre en indium, dont l'épaisseur est variable. Le meilleur rendement est obtenu pour une épaisseur de 400 Å, la diode émet alors une puissance de 5 mW à 371 nm avec une largeur de raie de 8,6 nm. Les diodes réalisées avec une couche active de GaN, c'est-à-dire avec une absence totale d'indium, émettent un rayonnement centré à 368 nm mais avec une puissance dix fois inférieure. Il est clair que les traces d'indium augmentent légèrement la longueur d'onde d'émission mais surtout jouent un rôle majeur dans le rendement quantique de la diode. Ceci résulte du fait que dans le composé InGaN les fluctuations d'alliage dues à la faible solubilité de l'indium dans GaN (<6% à la température normale de croissance, 700-800 °C) créent des îlots riches en indium, peut-être même des îlots de InN, de taille nanométrique. Ces îlots jouent le rôle de boîtes quantiques qui confinent les excitons. Ce confinement améliore le rendement radiatif par le double fait qu'il diminue la durée de vie radiative des excitons en augmentant l'intégrale de recouvrement électron-trou, et surtout qu'il augmente leur durée de vie non radiative en les localisant spatialement, ce qui les rend moins sensibles aux défauts cristallins qui par ailleurs sont en grande densité. La longueur

d'onde d'émission est alors conditionnée, à travers la composition et la taille des boîtes et la présence éventuelle d'un champ piézoélectrique interne, par le confinement quantique des excitons ; elle augmente avec la concentration nominale d'indium. Ceci résulte de l'augmentation du taux d'indium dans les boîtes et/ou de l'augmentation de la taille de ces dernières. Dans le premier cas c'est le gap du matériau constituant les îlots qui diminue, dans le second c'est le confinement quantique qui diminue. Ces deux effets ne sont évidemment pas exclusifs et coexistent probablement mais dans des proportions qu'il est difficile de déterminer.

Des diodes lasers émettant une puissance supérieure à 80 mW en régime monomode à 400 nm ont été réalisées. Les premiers lasers *violets* commercialisés (janvier 1999) délivrent une puissance de 5 mW à 400 nm avec un courant d'excitation de 40 mA sous une tension de polarisation de 5 V, leur durée de vie est estimée à 10 000 heures à la température ambiante. Ces lasers sont suffisants pour permettre la lecture de disques optiques.

13.3.4. *Emetteurs blancs*

Dans la mesure où l'on dispose de diodes électroluminescentes rouges, vertes et bleues, il suffit de les associer pour obtenir n'importe qu'elle couleur du spectre visible, y compris le blanc. En fait pour obtenir le blanc une méthode plus simple, calquée sur le fonctionnement des tubes fluorescents, est utilisée. Cette méthode consiste à exciter avec une LED bleue la luminescence d'un matériau fluorescent.

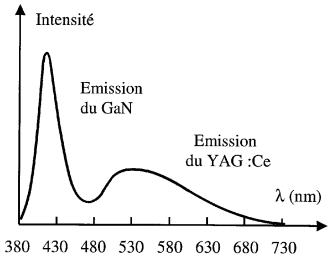


Figure 13-20 : Spectre d'émission d'une diode blanche GaN-YAG :Ce

La LED bleue est une diode à GaN dont le spectre d'émission est centré à 430 nm, ou une diode à InGaN dont le spectre est centré à 470 nm. Le matériau luminescent est un grenat d'yttrium et d'aluminium ($Y_3Al_5O_{12}$), communément appelé YAG, dopé au cérium pour une association avec une diode GaN ou au gadolinium pour une association avec une diode InGaN. Ce matériau peut être synthétisé avec une grande pureté et présente de bonnes résistances thermique, chimique et à la corrosion. Il se présente sous la forme d'une poudre dont les grains ont une taille typique de 20 μm, que l'on mélange avec une résine plastique. La réalisation de la diode blanche consiste à déposer une couche mince du mélange concentré résine-YAG sur la diode, avant ou après encapsulation de cette dernière. L'épaisseur et la concentration de la couche de résine sont très critiques car elles conditionnent la chromaticité de la diode. En effet, le YAG émet un spectre large centré dans le jaune de sorte que c'est le mélange de cette émission avec le rayonnement bleu émis par la diode à travers le YAG et non absorbé par ce dernier, qui crée la lumière blanche. Le spectre d'émission de la diode blanche constituée par le couple GaN-YAG :Ce est représenté sur la figure (13-20). Dans les applications où une grande brillance est nécessaire le couple InGaN-YAG :Gd

est préféré au précédent car il bénéficie du meilleur rendement des diodes InGaN par rapport aux diodes GaN.

Dans le diagramme de chromaticité CIE de la Commission Internationale de l'Eclairage , les coordonnées du point blanc idéal, avec une égale distribution de toutes les couleurs, sont x=0,33 ; y=0,33 et par suite z=0,33. L'émission jaune du YAG :Ce correspond à x=0,46 ; y=0,52 et l'émission bleue de la diode GaN à x=0,15 ; y=0,09. La chromaticité de la diode blanche GaN-YAG :Ce correspond aux coordonnées x=0,29 ; y=0,31 , ce qui correspond à une température de 7500 K. En modifiant la concentration du YAG dans la résine différentes nuances peuvent être obtenues.

Une nouvelle étape sera franchie par l'utilisation de diodes UV comme source primaire. L'association de ces émetteurs avec un matériau luminescent absorbant 100 % de leur rayonnement permettra la réalisation de LED blanches dont le spectre sera uniquement conditionné par le spectre d'émission du matériau luminescent, comme dans les tubes fluorescents classiques.

Une autre technique, utilisant l'addition de couleurs à l'intérieur d'une couche active déposée sur silicium a récemment été proposée (Damilano B.-1999). Cette technique consiste à déposer sur un substrat de silicium(111) une couche de AlN dans laquelle on fait croître des îlots de GaN de différentes tailles. Ces îlots jouent le rôle de boites quantiques qui émettent une luminescence dont la couleur varie avec la taille des boîtes. Il est bien connu que les désaccords de maille entre le substrat et les couches épitaxiées induisent dans les hétérostructures à base de nitrures wurtzite des champs électriques internes très intenses (~5 MV/cm), il en résulte que les niveaux d'énergie sont conditionnés à la fois par le confinement quantique et par l'effet Stark confiné (Par. 12.4.3). Sur les îlots de GaN/AlN une différence d'épaisseur d'une monocouche entraîne un déplacement énergétique du spectre d'émission de 150 mV. Compte tenu de la valeur du gap de GaN tout le spectre visible est ainsi parcouru en réalisant des îlots dont les épaisseurs varient d'une dizaine de monocouches. Les spectres *a*, *b*, *c* de la figure (13-21) représentent chacun l'émission d'un plan d'îlots dont l'épaisseur est 7, 10, 12 monocouches. Ces spectres correspondent aux couleurs (bleu, vert et orange). Leurs coordonnées (x ;

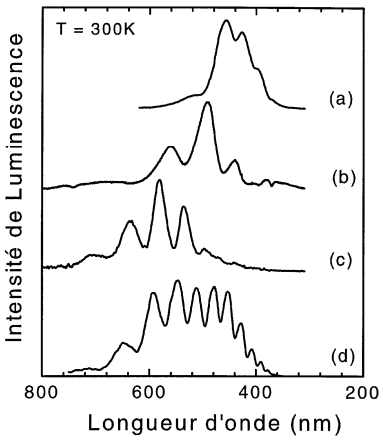


Figure 13-21 : Spectres d'émission de boites quantiques GaN/AlN. a, b, c coorrespondent à un seul plan d'îlots d'épaisseur 7, 10, 12 monocouches. Le spectre d correspond à l'émission d'un empilement de 4 plans d'îlots de tailles différentes *(Damilano B. – 1999).*

y) dans le diagramme de chromaticité sont (0,16 ; 0,10), (0,25 ; 0,38), (0,43 ; 0,46). Une structure contenant un empilement de quatre plans d'îlots de tailles convenablement choisies permet ainsi de produire de la lumière blanche par addition de couleurs à l'intérieur même de la couche de AlN. Le spectre obtenu est représenté par la courbe *d* sur la figure (13-21). Malgré le faible rendement de l'émission rouge ce spectre est pratiquement blanc, ses coordonnées dans le diagramme de chromaticité sont x=0,31 et y=0,35, très voisines des coordonnées idéales de la lumière blanche. Les oscillations présentes sur tous les spectres résultent de phénomènes d'interférences dans la couche.

Le développement des diodes blanches devrait prendre une ampleur considérable dans tous les domaines, automobile, feux de signalisation, éclairage, affichage, …, compte tenu du fait que leur rendement et leur durée de vie sont respectivement 3 fois et 10 fois supérieurs à ceux des lampes à incandescence. Leur seul défaut actuel est leur coût.

Annexes

A.1. Coefficient de transmission d'une barrière de potentiel — Méthode BKW

Pour calculer la probabilité de transfert d'une particule à travers une barrière de potentiel, on utilise généralement la méthode BKW (Brillouin Kramers Wentzel).

Considérons un électron d'énergie E se présentant devant une barrière d'énergie potentielle $U(x)=-eV(x)$ (Fig. A-1). L'équation de Schrodinger de cet électron s'écrit

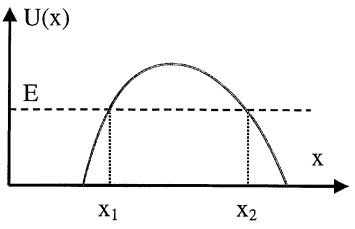


Figure A-1

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + U(x)\psi = E\psi \qquad \text{(A-1)}$$

ou

$$\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + \big(E - U(x)\big)\psi = 0 \qquad \text{(A-2)}$$

soit, en posant $K(x) = \sqrt{\frac{2m}{\hbar^2}\big(U(x) - E\big)}$

$$\frac{d^2\psi}{dx^2} - K^2(x)\,\psi = 0 \qquad \text{(A-3)}$$

La solution de cette équation différentielle est de la forme $\psi = e^{\alpha(x)}$, où $\alpha(x)$ est négatif car si l'électron se déplace de la gauche vers la droite sa fonction d'onde diminue nécessairement suivant x. Les dérivées successives de ψ s'écrivent donc

$$\frac{d\psi}{dx} = e^{\alpha(x)}\,\frac{d\alpha(x)}{dx} \qquad \text{(A-4-a)}$$

$$\frac{d^2\psi}{dx^2} = e^{\alpha(x)}\ \frac{d^2\alpha(x)}{dx^2} + e^{\alpha(x)}\left(\frac{d\alpha(x)}{dx}\right)^2 \qquad \text{(A-4-b)}$$

L'équation s'écrit donc

$$\frac{d^2\alpha(x)}{dx^2} + \left(\frac{d\alpha(x)}{dx}\right)^2 - K^2(x) = 0 \qquad \text{(A-5)}$$

En supposant que $\alpha(x)$ varie lentement avec x, on peut faire l'approximation

$$\frac{d^2\alpha(x)}{dx^2} << \left(\frac{d\alpha(x)}{dx}\right)^2 \qquad \text{(A-6)}$$

de sorte que l'équation s'écrit

$$\left(\frac{d\alpha(x)}{dx}\right)^2 = K^2(x) \qquad \text{(A-7)}$$

Ainsi lorsque l'électron atteint la barrière, c'est-à-dire pour $x>x_1$, $\alpha(x)$, qui doit être négatif, est donné par

$$\alpha(x) = -\int_{x1}^{x} K(x)\,dx \qquad \text{(A-8)}$$

La fonction d'onde s'écrit donc

$$\psi = \exp\left(-\int_{x1}^{x} K(x)\,dx\right) \qquad \text{(A-9)}$$

Le coefficient de transmission de la barrière, qui est donné par la probabilité de présence de l'électron en $x=x_2$, s'écrit donc

$$T = \psi^*\psi\Big|_{x=x2} = \exp\left(-2\int_{x1}^{x2} K(x)\,dx\right) \qquad \text{(A-10)}$$

Soit

$$T = \exp\left(-2\left(\frac{2m}{\hbar^2}\right)^{1/2}\int_{x1}^{x2}\left(U(x)-E\right)^{1/2}dx\right) \qquad \text{(A-11)}$$

A.2. Linéarisation de la fonction de distribution de Fermi

Dans un semiconducteur non dégénéré l'approximation de la fonction de distribution de Fermi par une fonction de Boltzmann permet d'obtenir des expressions analytiques des densités de porteurs. Mais lorsque le semiconducteur ne satisfait plus aux conditions justifiant l'approximation de Boltzmann, la distribution des porteurs doit nécessairement être représentée par la fonction de Fermi.

$$f(E) = \frac{1}{1 + e^{(E-E_F)/kT}} \tag{A-12}$$

Dans ce cas les densités de porteurs ne sont plus données par des expressions analytiques mais par les expressions intégrales (1-140 et 142)

$$n = N_c \, F_{1/2}(\eta) \tag{A-13-a}$$

$$p = N_v \, F_{1/2}(-\eta - \varepsilon_g) \tag{A-13-b}$$

avec $$F_{1/2}(x) = \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{\varepsilon^{1/2}}{1 + e^{\varepsilon - x}} d\varepsilon \tag{A-14}$$

où $\eta = (E_F - E_c)/kT$ $\quad \varepsilon = (E - E_c)/kT$ $\quad \varepsilon_g = E_g/kT$

Dans la pratique les fonctions intégrales $F_{1/2}(x)$ ne sont pas d'un usage aisé de sorte qu'il est souvent utile, dans l'étude de certains composants, de disposer d'expressions analytiques des densités de porteurs quel que soit le dopage du semiconducteur. Un exemple d'approximations est donné par les équations (1-145) représentées sur la figure (1-19).

Dans le cas d'un semiconducteur dégénéré on peut utiliser une autre approche de la simplification en linéarisant par intervalles la fonction de Fermi. Considérons par exemple un semiconducteur de type *n* dont le niveau de Fermi est situé dans la bande de conduction et étudions comment linéariser la fonction de distribution des électrons.

Considérons tout d'abord les états d'énergie situés au-dessous du niveau de Fermi à une distance de celui-ci supérieure à $2kT$, $(E_F\text{-}E)>2kT$. On peut pour ces états, négliger l'exponentielle devant 1 au dénominateur de la fonction de Fermi et écrire $f(E)\approx 1$.

Considérons maintenant les états d'énergie situés au-dessus du niveau de Fermi à une distance de celui-ci supérieure à $2kT$, $(E\text{-}E_F)>2kT$. On peut dans ce cas négliger 1 devant l'exponentielle au dénominateur de $f(E)$ qui se réduit alors à une distribution de Boltzmann. En fait, le matériau étant dégénéré la densité de porteurs d'énergie inférieure à $E_F + 2kT$ est considérablement plus importante que celle des

porteurs d'énergie supérieure à $E_F + 2kT$. On peut alors négliger ces derniers, ce qui revient à écrire $f(E) \approx 0$ pour $E > E_F + 2kT$.

Considérons enfin les états dont l'énergie est comprise entre $E_F - 2kT$ et $E_F + 2kT$. On peut, pour ces états, linéariser la fonction de distribution par un développement limité autour de E_F.

$$f(E) \approx f(E_F) + \left(\frac{d\,f(E)}{dE} \right)_{E_F} \left(E - E_F \right) \qquad \text{(A-15)}$$

avec $$\left(\frac{d\,f(E)}{dE} \right)_{E_F} = -\frac{1}{kT} \left(\frac{e^{(E-E_F)/kT}}{\left(1 + e^{(E-E_F)/kT}\right)^2} \right)_{E_F} = -\frac{1}{4\,kT}$$

Soit $$f(E) \approx \frac{1}{2} \left(1 - \frac{E - E_F}{2\,kT} \right) \qquad \text{(A-16)}$$

Dans un semiconducteur dégénéré on pourra donc linéariser la fonction de distribution des porteurs majoritaires par les expressions

$$E < E_F - 2kT \qquad f(E) \approx 1 \qquad \text{(A-17-a)}$$

$$E_F - 2kT < E < E_F + 2kT \qquad f(E) \approx \frac{1}{2} \left(1 - \frac{E - E_F}{2\,kT} \right) \qquad \text{(A-17-b)}$$

$$E_F + 2\,kT < E \qquad f(E) \approx 0 \qquad \text{(A-17-c)}$$

La figure (A-2) représente la fonction de Fermi linéarisée (Eq. A-17).

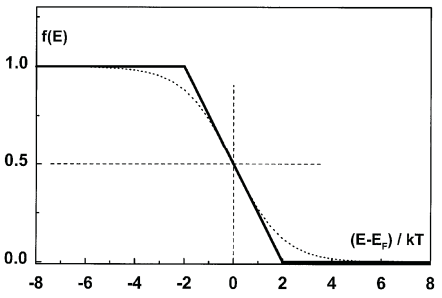


Figure A-2

La densité d'électrons dans la bande de conduction est donnée par l'intégrale

$$n = \int_{E_c}^{\infty} N_c(E)\, f(E)\, dE \qquad \text{(A-18)}$$

où $N_c(E)$ est la densité d'états dans la bande de conduction (Eq. 1-77).

En utilisant l'approximation linéarisée de la fonction de Fermi on peut calculer simplement l'intégrale (A-18) qui s'écrit sous différentes formes suivant la position du niveau de Fermi par rapport au bas de la bande de conduction, c'est-à-dire suivant le degré de dégénérescence :

i) $E_F < E_c - 2Kt$

L'expression (A-18) donne n=0 car le produit $N_c(E)\, f(E)$ est nul partout. En effet $N_c(E)$ est nul pour $E < E_c$ et $f(E)$ est nul pour $E > E_c$. En fait le semiconducteur est ici loin de la dégénérescence de sorte que la densité d'électrons est donnée par l'approximation de Boltzmann (Eq. 1-132-a).

ii) $E_c - 2kT < E_F < E_c + 2kT$

Le semiconducteur est voisin du seuil de dégénérescence, légèrement au-dessous ou légèrement au-dessus. Compte tenu du fait que $N_c(E)$ est nul pour $E<E_c$, l'utilisation des expressions (A-17) permet d'écrire l'expression (A-18) sous la forme

$$n = \frac{1}{2}\int_{E_c}^{E_F+2kT} N_c(E)\left(1 - \frac{E - E_F}{2\,kT}\right) dE \qquad \text{(A-19)}$$

iii) $E_c + 2kT < E_F$

Le semiconducteur est fortement dégénéré, l'utilisation des expressions linéarisées (A-17) donne

$$n = \int_{E_c}^{E_F-2kT} N_c(E)\, dE + \frac{1}{2}\int_{E_F-2kT}^{E_F+2kT} N_c(E)\left(1 - \frac{E - E_F}{2\,kT}\right) dE \qquad \text{(A-20)}$$

En explicitant $N_c(E)$ (Eq. 1-77) les expressions (A-19 et 20) s'écrivent respectivement

$$n = \frac{1}{2\pi^2}\left(\frac{2\,m_e}{\hbar^2}\right)^{3/2} \frac{1}{2}\int_{E_c}^{E_F+2kT} \left(E - E_c\right)^{1/2} \left(1 - \frac{E - E_F}{2\,kT}\right) dE \qquad \text{(A-21)}$$

$$n = \frac{1}{2\pi^2}\left(\frac{2\,m_e}{\hbar^2}\right)^{3/2} \left(\int_{E_c}^{E_F-2kT} N_c(E)\, dE + \frac{1}{2}\int_{E_F-2kT}^{E_F+2kT} N_c(E)\left(1 - \frac{E - E_F}{2\,kT}\right) dE\right) \qquad \text{(A-22)}$$

Ces expressions s'intègrent très facilement, on obtient ainsi les expressions analytiques suivantes où N_c est la densité équivalente d'états de la bande de conduction (Eq. 1-136-a) (un facteur 2/15 apparaît dans le calcul, il a été remplacé ici par le facteur $1/2^3$)

$E_c - 2kT < E_F < E_c + 2kT$ $\qquad n = \dfrac{N_c}{\sqrt{2\pi}}\left(\dfrac{E_F - E_C}{2kT} + 1\right)^{5/2}$ (A-23)

$E_c + 2kT < E_F$

$$n = \frac{N_c}{\sqrt{2\pi}}\left(\left(\frac{E_F - E_c}{2\,kT} + 1\right)^{5/2} - \left(\frac{E_F - E_c}{2\,kT} - 1\right)^{5/2}\right) \qquad \text{(A-24)}$$

La figure (A-3) représente l'évolution de la densité d'électrons n pour un semiconducteur non dégénéré dans l'approximation de Boltzmann (courbe a), et pour un semiconducteur dégénéré dans l'approximation analytique précédente (courbe b). Dans le régime de faible dégénérescence défini par $E_F < E_c+2kT$, l'approximation de Boltzmann donne une valeur approchée par excès alors que l'expression (A-23) donne une valeur approchée par défaut.

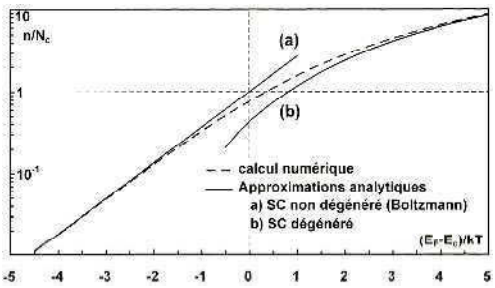


Figure A3 : Approximation

Quand le semiconducteur est très dégénéré, c'est-à-dire pour $E_F >> E_c+2kT$, un développement limité de l'expression (A-24) permet de retrouver le comportement métallique. Cette expression s'écrit

$$n = \frac{N_c}{\sqrt{2\pi}}\left(\frac{E_F - E_c}{2\,kT}\right)^{5/2}\left(\left(1+\frac{2\,kT}{E_F - E_c}\right)^{5/2} - \left(1-\frac{2\,kT}{E_F - E_c}\right)^{5/2}\right) \qquad \text{(A-25)}$$

Pour $2kT << E_F - E_c$, le développement au premier ordre des termes de la forme $(1\pm\varepsilon)^{5/2} \approx 1\pm\frac{5}{2}\varepsilon$, donne

$$n \approx 5\frac{N_c}{\sqrt{2\pi}}\left(\frac{E_F - E_c}{2\,kT}\right)^{3/2} \qquad \text{(A-26)}$$

En explicitant N_c (Eq. 1-136-a), qui varie comme $(kT)^{3/2}$, on retrouve le comportement métallique du semiconducteur avec une densité d'électrons indépendante de la température

$$n \approx \frac{8\pi}{3h^3}(2m_e)^{3/2}(E_F - E_c)^{3/2} \qquad \text{(A-27)}$$

(On a rétabli ici le facteur 2/15 qui avait été remplacé par 1/8 dans les expressions (A-23 et 24)).

On retrouve l'expression bien connue de l'énergie de Fermi des électrons de conduction d'un métal, ou d'un semiconducteur très dégénéré

$$E_F = E_c + \frac{h^2}{2\,m_e}\left(\frac{3n}{8\pi}\right)^{2/3} \qquad \text{(A-28)}$$

A.3. Effet Hall

Considérons un barreau semiconducteur de largeur l et d'épaisseur d soumis à un champ électrique longitudinal $\vec{E}(E_x,0,0)$ et un champ magnétique transverse $\vec{B}(0,0,B_z)$ (Fig. A-4).

Sous l'action du champ électrique les porteurs se déplacent longitudinalement avec une vitesse de dérive $\vec{v}(v_x,0,0)$ et transportent une densité de courant $\vec{j}(j_x,0,0)$.

Sous l'action du champ magnétique les porteurs sont soumis à une force de Lorentz $\vec{F} = q\vec{v} \wedge \vec{B}$ $(0,qv_xB_z,0)$ dans la direction y. Ils sont donc soumis à une force transversale qui a tendance à dévier leur trajectoire en leur donnant une composante de vitesse v_y non nulle. Une composante non nulle j_y de la densité de courant a donc tendance à s'établir transversalement. Mais le circuit étant ouvert dans la direction y aucun courant ne peut circuler dans cette direction, la composante j_y reste donc identiquement nulle. Des charges, négatives sur une face et positives sur l'autre, s'accumulent alors sur les faces latérales du barreau et créent un champ électrique transverse $\vec{E}(0,E_y,0)$. Ce champ exerce sur les porteurs une force qui, en équilibrant la force de Lorentz, maintient parallèle à l'axe x leur vecteur vitesse. C'est l'*Effet Hall*, découvert par Edwin Hall en 1879.

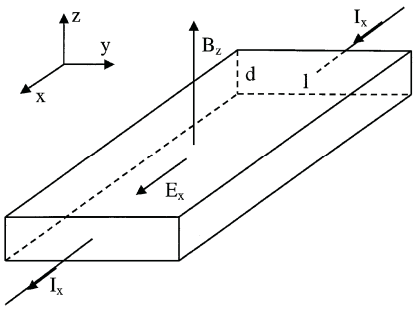


Figure A-4

A.3.1. *Modèle simple*

a. Semiconducteur de type n

Si le semiconducteur est de type n la densité de courant dans la direction longitudinale du barreau s'écrit

$$j_{nx} = -nev_{nx} = ne\mu_n E_x \tag{A-29}$$

où $v_{nx} = -\mu_n E_x$ est la vitesse de dérive des électrons.

Dans la direction transversale y, la force résultant de la composante de Lorentz et du champ de Hall s'écrit

$$F_y = -eE_y + ev_{nx}B_z = -e\left(E_y + \mu_n E_x B_z\right) \tag{A-30}$$

La densité de courant transversale s'écrit donc

$$j_{ny} = ne\mu_n\left(E_y + \mu_n E_x B_z\right) \tag{A-31}$$

Puisque $j_{ny} \equiv 0$ l'expression (A-31) donne

$$E_y = -\mu_n E_x B_z \tag{A-32}$$

E_y est appelé *champ de Hall*, la tension transversale V_y=-lE_y qui en résulte est appelée *tension de Hall*.

Il est plus commode d'exprimer ces grandeurs en fonction de la densité de courant j_{nx} car cette dernière est directement reliée à la grandeur mesurable qu'est le courant I_x traversant le barreau. Ainsi compte tenu des relations $j_{nx} = ne\mu_n E_x$ et $I_x = ld\, j_{nx}$, le paramètre local qu'est le champ de Hall et la grandeur macroscopique mesurable qu'est la tension de Hall s'écrivent respectivement

$$E_y = R_H\, j_{nx} B_z \tag{A-33-a}$$

$$V_y = -R_H \frac{I_x B_z}{d} \tag{A-33-b}$$

avec

$$R_H = -1/ne \tag{A-34}$$

Le paramètre R_H est appelé *coefficient de Hall*. Connaissant les quantités I_x, B_z et d, la mesure de la tension de Hall V_y permet d'atteindre le coefficient de Hall R_H. Le signe de ce dernier donne le type du semiconducteur, sa valeur donne la densité de porteurs majoritaires.

Les expressions précédentes montrent que le champ de Hall E_y est proportionnel à la densité de courant j_{nx} , le coefficient de proportionnalité a les dimensions d'une résistivité, on l'appelle ρ_{xy} , c'est la *résistivité de Hall*. De même la tension de Hall V_y est proportionnelle au courant I_x , le coefficient de proportionnalité a les dimensions d'une résistance, on l'appelle R_{xy} , c'est la *résistance de Hall*. Ces paramètres sont donnés par

$$\rho_{xy} = R_H\, B_z \tag{A-35-a}$$

$$R_{xy} = -\,R_H\, B_z/d \tag{A-35-b}$$

Si on complète la mesure de l'effet Hall par la mesure de la conductivité $\sigma = ne\mu_n$, on peut déduire la mobilité des porteurs

$$\mu_n = -R_H\sigma \tag{A-36}$$

b. Semiconducteur de type p

Dans un semiconducteur de type *p*, les expressions (A-29 et 30) deviennent respectivement

$$j_{px} = pev_{px} = pe\mu_p E_x \tag{A-37}$$

$$F_y = eE_y - ev_{px}B_z = e\left(E_y - \mu_p E_x B_z\right) \tag{A-38}$$

de sorte que la densité de courant suivant *y* s'écrit

$$j_{py} = pe\mu_p\left(E_y - \mu_p E_x B_z\right) \equiv 0 \tag{A-39}$$

Soit

$$E_y = \mu_p E_x B_z \tag{A-40}$$

En introduisant la densité de courant $j_{px} = pe\mu_p E_x$ et le courant total $I_x = ld\, j_{px}$, le champ et la tension de Hall s'écrivent respectivement

$$E_y = R_H j_{px} B_z \tag{A-41-a}$$

$$V_y = -R_H \frac{I_x B_z}{d} \tag{A-41-b}$$

avec

$$R_H = 1/pe \tag{A-42}$$

La résistivité et la résistance de Hall s'écrivent respectivement

$$\rho_{xy} = R_H\, B_z \tag{A-43-a}$$

$$R_{xy} = -\,R_H\, B_z/d \tag{A-43-b}$$

La mobilité des porteurs s'écrit

$$\mu_p = R_H \sigma \tag{A-44}$$

c. Semiconducteur compensé

Si les deux types de porteurs sont présents dans le matériau, les densités de courant dans les directions *x* et *y* s'écrivent respectivement

$$j_x = j_{nx} + j_{px} = e\left(n\mu_n + p\mu_p\right)E_x \tag{A-45}$$

$$j_y = j_{ny} + j_{py} = ne\mu_n\left(E_y + \mu_n E_x B_z\right) + pe\mu_p\left(E_y - \mu_p E_x B_z\right) \equiv 0 \tag{A-46}$$

Ainsi puisque $j_y \equiv 0$

$$E_y = \frac{p\mu_p^2 - n\mu_n^2}{p\mu_p + n\mu_n} E_x B_z \tag{A-47}$$

Soit, en faisant apparaître la densité de courant suivant *x*

$$E_y = R_H j_x B_z \tag{A-48}$$

$$V_y = -R_H \frac{I_x B_z}{d} \tag{A-49}$$

avec

$$R_H = \frac{1}{e} \frac{p\mu_p^2 - n\mu_n^2}{(p\mu_p + n\mu_n)^2} \tag{A-50}$$

Cette expression se ramène à $R_H=-1/ne$ pour un semiconducteur de type *n* et à $R_H=1/pe$ pour un semiconducteur de type *p*. Pour un matériau intrinsèque ou parfaitement compensé, $n=p=n_i$, le coefficient de Hall s'écrit

$$R_{Hi} = \frac{1}{n_i e} \frac{\mu_p - \mu_n}{\mu_p + \mu_n} \tag{A-51}$$

A.3.2. *Modèle plus élaboré*

Le modèle simple de calcul de l'effet Hall présenté ci-dessus est en bon accord avec l'expérience et permet de caractériser très correctement les semiconducteurs, par la mesure du type, du nombre de porteurs et de leur mobilité. Toutefois, nous devons préciser qu'une hypothèse simplificatrice importante est implicitement contenue dans ce modèle. Cette hypothèse est que tous les porteurs d'un même type ont la même vitesse de dérive, v_n pour les électrons, v_p pour les trous. En fait il n'en est pas ainsi car la vitesse de dérive est une valeur moyenne, certains porteurs sont plus rapides que la moyenne d'entre eux, d'autres plus lents. Le champ de Hall, qui est un effet moyen, n'équilibre donc pas la force de Lorentz pour tous les porteurs. Les plus rapides sont déviés dans un sens par la force de Lorentz qui pour eux n'est pas entièrement compensée par la force de Hall, les plus lents sont déviés dans l'autre sens par la force de Hall qui pour eux est supérieure à la force de Lorentz. Ainsi donc en présence du champ magnétique les trajectoires des porteurs ne sont pas toutes longitudinales, il existe des courants transversaux dont la somme algébrique est évidemment nulle puisque le courant total suivant *y* reste nul en raison de l'absence de connexion latérale. Il faut noter que les porteurs dont la trajectoire est déviée latéralement ont une composante longitudinale de vitesse qui de ce fait diminue, il en résulte une diminution du courant longitudinal c'est-à-dire de la conductivité du matériau, cet effet porte le nom de *magnétorésistance transversale*.

Le traitement rigoureux de ce cas général fait appel à l'équation de transport de Boltzmann, dont l'usage sort du cadre de cet ouvrage. Nous allons utiliser une autre approche en supposant que les porteurs d'un même type ont tous la même masse effective, c'est-à-dire que les bandes d'énergie sont paraboliques.

Pour tenir compte du fait que les porteurs sont soumis à des collisions, de la part des atomes du réseau cristallin ou de la part d'impuretés, on peut écrire l'équation fondamentale de la dynamique en faisant intervenir une force de frottement fonction de la vitesse de dérive du porteur

$$\vec{F} - \frac{m^*}{\tau}\vec{v_d} = m^* \frac{d\vec{v_d}}{dt} \tag{A-52}$$

Si la masse effective est la même pour tous les porteurs et leurs temps de collision indépendants de l'énergie, la force de frottement est proportionnelle à la vitesse.

Considérons un barreau semiconducteur de type *n* soumis à un champ électrique longitudinal $\vec{E}(E_x,0,0)$ et à un champ magnétique transverse $\vec{B}(0,0,B_z)$ constants. Le régime est stationnaire et l'équation (A-52) s'écrit

$$-e\vec{E} - e\vec{v_n} \wedge \vec{B} = \frac{m_e}{\tau}\vec{v_n} \tag{A-53}$$

Les projections de cette équation sur les axes de coordonnées s'écrivent

$$-eE_x - ev_{ny}B_z = \frac{m_e}{\tau}v_{nx} \tag{A-54}$$

$$-eE_y + ev_{nx}B_z = \frac{m_e}{\tau}v_{ny} \tag{A-55}$$

Dans les effets galvanomagnétiques un paramètre conditionne le fonctionnement du système, c'est la pulsation cyclotron définie par $\omega_c=eB/m$. En posant ici $\omega_c=eB_z/m_e$ les équations précédentes s'écrivent

$$-\frac{e\tau}{m_e}E_x - \omega_c\tau\ v_{ny} = v_{nx} \tag{A-56}$$

$$-\frac{e\tau}{m_e}E_y + \omega_c\tau\ v_{nx} = v_{ny} \tag{A-57}$$

On en déduit les expressions explicites des composantes de vitesse

$$v_{nx} = -\frac{e}{m_e}\left(\frac{\tau}{1+\omega_c^2\tau^2}E_x - \omega_c\frac{\tau^2}{1+\omega_c^2\tau^2}E_y\right) \tag{A-58}$$

$$v_{ny} = -\frac{e}{m_e}\left(\frac{\tau}{1+\omega_c^2\tau^2}E_y + \omega_c\frac{\tau^2}{1+\omega_c^2\tau^2}E_x\right) \qquad \text{(A-59)}$$

Si $dn(v_{nx})$ et $dn(v_{ny})$ sont les densités d'électrons dont les composantes de vitesses suivant x et y sont respectivement v_{nx} et v_{ny} , les densités de courant transportées par ces électrons sont

$$dj_{nx} = -ev_{nx}dn(v_{nx}) \qquad \text{(A-60)}$$

$$dj_{ny} = -ev_{ny}dn(v_{ny}) \qquad \text{(A-61)}$$

Les composantes j_{nx} et j_{ny} du vecteur densité de courant sont données par l'intégration de ces expressions sur toutes les valeurs de v_{nx} et v_{ny} respectivement. On obtient une expression simplifiée de ces composantes en supposant que tous les électrons ont la même masse effective, on peut alors définir les valeurs moyennes des composantes du vecteur vitesse par les relations

$$\langle v_{nx}\rangle = -\frac{e}{m_e}\left(\left\langle\frac{\tau}{1+\omega_c^2\tau^2}\right\rangle E_x - \omega_c\left\langle\frac{\tau^2}{1+\omega_c^2\tau^2}\right\rangle E_y\right) \qquad \text{(A-62)}$$

$$\langle v_{ny}\rangle = -\frac{e}{m_e}\left(\left\langle\frac{\tau}{1+\omega_c^2\tau^2}\right\rangle E_y + \omega_c\left\langle\frac{\tau^2}{1+\omega_c^2\tau^2}\right\rangle E_x\right) \qquad \text{(A-63)}$$

Si n est la densité d'électrons, les composantes du vecteur densité de courant $\vec{j} = -ne\langle\overrightarrow{v_n}\rangle$ s'écrivent alors

$$j_{nx} = \frac{ne^2}{m_e}\left(\left\langle\frac{\tau}{1+\omega_c^2\tau^2}\right\rangle E_x - \omega_c\left\langle\frac{\tau^2}{1+\omega_c^2\tau^2}\right\rangle E_y\right) \qquad \text{(A-64)}$$

$$j_{ny} = \frac{ne^2}{m_e}\left(\left\langle\frac{\tau}{1+\omega_c^2\tau^2}\right\rangle E_y + \omega_c\left\langle\frac{\tau^2}{1+\omega_c^2\tau^2}\right\rangle E_x\right) \qquad \text{(A-65)}$$

Le courant transversal étant nécessairement nul, on obtient E_y en écrivant que j_{ny}=0, soit

$$E_y = -\omega_c\frac{\left\langle\tau^2/(1+\omega_c^2\tau^2)\right\rangle}{\left\langle\tau/(1+\omega_c^2\tau^2)\right\rangle}E_x \qquad \text{(A-66)}$$

Si le champ magnétique est suffisamment faible pour que la condition $\omega_c\tau<<1$ soit vérifiée, l'expression précédente s'écrit

$$E_y = -\omega_c \frac{\langle \tau^2 \rangle}{\langle \tau \rangle} E_x \tag{A-67}$$

On obtient une expression analogue à l'expression (A-32) en explicitant la pulsation cyclotron $\omega_c=eB_z/m_e$ et en faisant apparaître explicitement la mobilité des électrons $\mu_n = e\langle \tau \rangle / m_e$, soit

$$E_y = -\mu_n \frac{\langle \tau^2 \rangle}{\langle \tau \rangle^2} E_x B_z \tag{A-68}$$

Comme dans le calcul simple développé précédemment, si on fait apparaître la densité de courant $j_{nx}=ne\mu_n E_x$ et le courant macroscopique $I_x=ld\, j_{nx}$, le champ de Hall E_y et la tension de Hall $V_y=-lE_y$ s'écrivent sous une forme analogue aux expressions (A-33-a et b)

$$E_y = R_H j_{nx} B_z \tag{A-69}$$

$$V_y = -R_H \frac{I_x B_z}{d} \tag{A-70}$$

avec

$$R_H = -r/ne \tag{A-71}$$

où

$$r = \langle \tau^2 \rangle / \langle \tau \rangle^2 \tag{A-72}$$

R_H est le *coefficient de Hall*, r est appelé *constante de Hall* .

En faisant apparaître explicitement la conductivité du matériau $\sigma=ne\mu_n$, dans l'expression (A-71), on obtient

$$\mu_H = r\mu_n = -R_H \sigma \tag{A-73}$$

μ_n est la mobilité de conduction, μ_H est appelé *mobilité de Hall.*

Si le temps de collision des porteurs est indépendant de l'énergie $\langle \tau^2 \rangle = \langle \tau \rangle^2$ de sorte que $r = 1$, on retrouve les expressions (A-33 à 36). Lorsque la relaxation des porteurs est assurée par les interactions avec le réseau cristallin on peut montrer que la constante de Hall vaut $r = 3\pi/8$. Lorsque cette relaxation est conditionnée par les collisions sur les impuretés cette constante vaut $r = 315\pi/512$. On constate que la valeur de r ne s'éloigne jamais beaucoup de la valeur 1 ce qui justifie le calcul simplifié développé précédemment.

Si le semiconducteur est de type p on obtient de la même manière les expressions

$$E_y = R_H j_{px} B_z \tag{A-74}$$

$$V_y = -R_H \frac{I_x B_z}{d} \tag{A-75}$$

$$R_H = r/pe \tag{A-76}$$

$$\mu_H = r\mu_p = R_H \sigma \tag{A-77}$$

A.3.3. *Magnétorésistance*

Revenons sur l'expression (A-64), en explicitant E_y donné par l'expression (A-66), la composante longitudinale de la densité de courant s'écrit

$$j_{nx} = \frac{ne^2}{m_e}\left(\left\langle \tau/(1+\omega_c^2\tau^2) \right\rangle + \omega_c^2 \frac{\left\langle \tau^2/(1+\omega_c^2\tau^2) \right\rangle^2}{\left\langle \tau/(1+\omega_c^2\tau^2) \right\rangle} \right) E_x \tag{A-78}$$

Soit, $j_{nx} = \sigma_{xx} E_x$ avec

$$\sigma_{xx} = \frac{ne^2}{m_e}\left(\left\langle \tau/(1+\omega_c^2\tau^2) \right\rangle + \omega_c^2 \frac{\left\langle \tau^2/(1+\omega_c^2\tau^2) \right\rangle^2}{\left\langle \tau/(1+\omega_c^2\tau^2) \right\rangle} \right) \tag{A-79}$$

- Domaine des champs faibles : Si la condition $\omega_c \tau << 1$ est vérifiée, on peut développer l'expression précédente en puissances de $\omega_c \tau$, $(1/(1+\varepsilon) \approx 1-\varepsilon)$, on obtient

$$\sigma_{xx} = \frac{ne^2}{m_e}\left(\left\langle (1-\omega_c^2\tau^2)\tau \right\rangle + \omega_c^2 \frac{\left\langle (1-\omega_c^2\tau^2)\tau^2 \right\rangle^2}{\left\langle (1-\omega_c^2\tau^2)\tau \right\rangle} \right) \tag{A-80}$$

Si en outre on néglige les termes d'ordre supérieur à ω_c^2 , on obtient

$$\sigma_{xx} = \frac{ne^2}{m_e}\left(\left\langle \tau \right\rangle - \omega_c^2 \left\langle \tau^3 \right\rangle + \omega_c^2 \frac{\left\langle \tau^2 \right\rangle^2}{\left\langle \tau \right\rangle} \right) \tag{A-81}$$

Dans la mesure où $\sigma_0 = ne^2 \left\langle \tau \right\rangle / m_e$, cette expression s'écrit

$$\sigma_{xx} = \sigma_0 \left(1 - \omega_c^2 \left(\frac{\langle \tau^3 \rangle}{\langle \tau \rangle} - \frac{\langle \tau^2 \rangle^2}{\langle \tau \rangle^2} \right) \right) \tag{A-82}$$

L'évaluation des moyennes permet de montrer que la quantité $\langle \tau^3 \rangle / \langle \tau \rangle$ est supérieure à la quantité $\langle \tau^2 \rangle^2 / \langle \tau \rangle^2$. Il en résulte que la conductivité σ_{xx} diminue quand le champ magnétique B_z augmente. Si le temps de collision des porteurs est indépendant de l'énergie, les quantités $\langle \tau^n \rangle$ et $\langle \tau \rangle^n$ sont égales de sorte que le coefficient de ω_c s'annule, la magnétorésistance disparaît.

- Domaine des champs forts : Si la condition $\omega_c\tau >> 1$ est vérifiée, on peut développer l'expression (A-79) en puissances de $1/\omega_c\tau$. En se limitant au premier terme du développement on obtient

$$\sigma_{xx} = \frac{ne^2}{m_e} \left(\frac{1}{\omega_c^2} \langle 1/\tau \rangle + \frac{1}{\langle 1/\tau \rangle} \right) \tag{A-83}$$

ou

$$\sigma_{xx} = \frac{ne^2}{m_e} \langle \tau \rangle \left(\frac{1}{\omega_c^2} \frac{\langle 1/\tau \rangle}{\langle \tau \rangle} + \frac{1}{\langle \tau \rangle \langle 1/\tau \rangle} \right) \tag{A-84}$$

Soit

$$\sigma_{xx} = \sigma_0 \left(\frac{1}{\omega_c^2} \frac{\langle 1/\tau \rangle}{\langle \tau \rangle} + \frac{1}{\langle \tau \rangle \langle 1/\tau \rangle} \right) \tag{A-85}$$

Lorsque le champ magnétique B_z tend vers l'infini le premier terme de l'expression précédente s'annule, et on obtient

$$\frac{\sigma_\infty}{\sigma_0} = \frac{1}{\langle \tau \rangle \langle 1/\tau \rangle} \tag{A-86}$$

Lorsque la relaxation est assurée par le réseau cristallin, l'évaluation des moyennes donne

$$\frac{\sigma_\infty}{\sigma_0} = \frac{9\pi}{32} \approx 0{,}88 \tag{A-87}$$

Il faut enfin noter que dans les semiconducteurs anisotropes la conductivité du matériau est aussi sensible à un champ magnétique parallèle au courant $(\vec{B} // \vec{j})$, il existe une *magnétorésistance longitudinale.*

A.3.4. *Effet Hall quantique*

Dans un système tridimensionnel tel que nous venons de l'étudier la résistance de Hall R_{xy} varie de façon continue proportionnellement au champ magnétique (Eq. A-35-b). Lorsque le gaz d'électrons est confiné dans un espace bidimensionnel, comme une structure MIS ou une hétérojonction (Fig. 10-1), il n'en est plus ainsi, la résistance de Hall est quantifiée et prend des valeurs discrètes dont la propriété remarquable est de ne faire intervenir que des constantes universelles, ces valeurs sont des sous-multiples de la quantité h/e^2. C'est l'*effet Hall quantique*, découvert par K. von Klitzing, G. Dorda et M. Pepper en 1980 (von Klitzing K.-1980).

Rappelons (Chap. 10) que dans une structure bidimensionnelle, comme une hétérojonction à dopage modulé par exemple (Fig. 12-1), les états électroniques sont quantifiés dans la direction perpendiculaire à l'interface et pseudo-continus dans le plan des couches. Il en résulte une redistribution de ces états en sous-bandes d'énergie (Fig. 10-3). Supposons comme dans la figure (12-1) que la population électronique est telle que seule la première sous-bande E_0 est peuplée.

Appliquons à cette structure un champ électrique longitudinal E_x dans le plan des couches et un champ magnétique transverse B_z perpendiculaire à l'interface. Les électrons, qui se déplacent suivant x sous l'action du champ électrique, voient leurs trajectoires déviées par le champ magnétique et décrivent dans le plan de l'interface des orbites circulaires dont le rayon est quantifié. La sous-bande de conduction est alors quantifiée en niveaux discrets appelés *niveaux de Landau*, d'énergie ε_n (n = 0,1,2,…) donnée par

$$\varepsilon_n = E_0 + \left(n + \frac{1}{2}\right)\hbar\omega_c \qquad \text{(A-88)}$$

où $\omega_c = eB_z/m_e$ est la pulsation cyclotron que nous avons déjà définie. La densité d'états, qui est constante dans chacune des sous-bandes en l'absence de champ magnétique (Fig. 10-4), prend la forme discrète d'une série de pics centrés sur chacun des niveaux de Landau et élargis par les processus de diffusion (Fig. A-5).

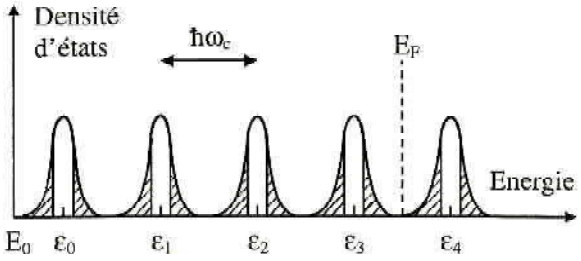


Figure A-5 : Densité d'états d'un gaz d'électrons bidimensionnel sous champ magnétique. Zones non hachurées : Etats étendus. Zones hachurées : Etats localisés.

La dégénérescence γ de chaque niveau et la distance entre deux niveaux successifs Δε sont respectivement données par

$$\gamma = \frac{e\,B_z}{h} \tag{A-89}$$

$$\Delta\varepsilon = \hbar\omega_c = \hbar\frac{e\,B_z}{m_e} \tag{A-90}$$

Notons que lorsque l'énergie cyclotron $\hbar\,\omega_c$ devient supérieure à la largeur des pics de densité d'états il existe des valeurs de l'énergie interdites aux électrons, c'est-à-dire pour lesquelles la densité d'états est nulle. Si, à très basse température, la population électronique est telle que le niveau de Fermi soit situé entre le $i^{ème}$ et le $(i+1)^{ème}$ niveau de Landau, il y a i niveaux pleins et la densité surfacique d'électrons s'écrit

$$n_s = i\,\gamma = i\frac{e\,B_z}{h} \tag{A-91}$$

Dans un système tridimensionnel la résistance de Hall R_{xy} donnée par les équations (A-34 et 35-b) s'écrit

$$R_{xy} = \frac{B_z}{e}\frac{1}{n\,d} \tag{A-92}$$

où le paramètre nd représente la densité surfacique d'électrons.

Dans le système bidimensionnel la densité surfacique n_s est donnée par l'équation (A-91) de sorte qu'en remplaçant la quantité nd par n_s dans l'expression précédente on obtient la résistance de Hall du système bidimensionnel, soit

$$R_{xy} = \frac{B_z}{e}\frac{1}{n_s} = \frac{1}{i}\frac{h}{e^2} \tag{A-93}$$

Quand le champ magnétique augmente, les niveaux de Landau se déplacent vers les hautes énergies (Eq. A-88) et en parallèle leur dégénérescence augmente (Eq. A-89). Ainsi les niveaux n, n-1, n-2, …traversent successivement le niveau de Fermi et se vident de leurs électrons dans les niveaux sous-jacents. Il en résulte (Eq. A-93) que la résistance de Hall R_{xy} reste constante tant que le niveau de Fermi reste entre deux niveaux de Landau et varie par saut de h/e^2 chaque fois qu'un niveau de Landau franchit le niveau de Fermi. R_{xy} est donc quantifié et sa courbe de variation avec le champ magnétique présente une allure de marches d'escalier avec des plateaux séparés par des marches de hauteur caractéristique h/e^2, ne faisant intervenir que des constantes universelles.

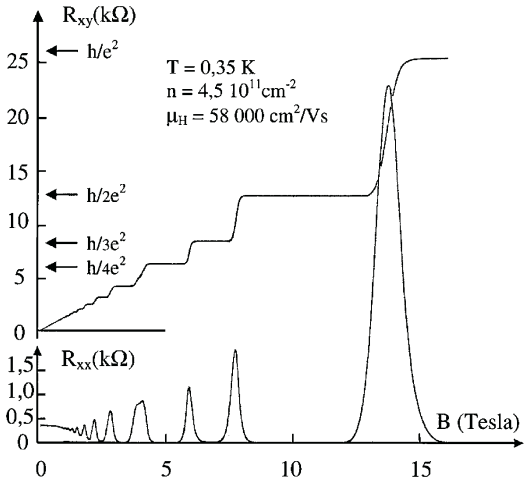


Figure A-6 : Variations de la résistance de Hall R_{xy} et de la résistance longitudinale R_{xx} en fonction du champ magnétique dans une hétérojonction GaAs/AlGaAs *(von Klitzing K. – 1984)*

Par ailleurs, à très basse température, lorsque le niveau de Fermi est situé entre deux niveaux de Landau le dernier niveau occupé est plein et le premier niveau vide est situé à l'énergie $\hbar\,\omega_c$ au-dessus. Il en résulte une absence totale de processus de diffusion car aucun état vide n'est accessible si la condition $\hbar\,\omega_c >> kT$ est vérifiée. La résistivité ρ_{xx} et la résistance longitudinale R_{xx} sont par conséquent nulles. La résistance R_{xx} n'est différente de zéro que chaque fois qu'un niveau de Landau franchit le niveau de Fermi car ce niveau est alors partiellement occupé ce qui autorise les processus de diffusion intra-niveau. La courbe de variation de la résistance longitudinale R_{xx} avec le champ magnétique présente alors une allure oscillatoire, c'est l'*effet Shubnikov-de Haas*, avec des pics correspondant chacun à un saut de la résistance de Hall R_{xy} d'un plateau à l'autre. D'autre part, à chaque plateau de R_{xy} correspond une valeur nulle de R_{xx} . La figure (A-6) représente les variations expérimentales de R_{xy} et R_{xx} mesurées dans une hétérojonction GaAs/AlGaAs. Les plateaux de R_{xy} ont des valeurs sous-multiples de h/e^2.

Cette explication simple de l'effet Hall quantique n'est cependant pas totalement satisfaisante car si elle explique correctement les valeurs des plateaux de R_{xy} et R_{xx} elle n'explique pas l'existence même de ces plateaux. En effet la densité surfacique n_s d'électrons étant constante le niveau de Fermi ne peut rester entre deux niveaux de Landau sur une plage importante de champ magnétique (parfois plusieurs teslas) alors que la densité d'états est nulle. Ceci signifie que les plateaux de R_{xy} et R_{xx} ne devraient pas exister.

Pour expliquer l'existence de ces plateaux nous devons considérer que certains états situés sur les flancs des niveaux de Landau sont des états localisés dans l'espace par un potentiel aléatoire créé par les impuretés et les défauts du système. Ces états localisés piègent des électrons qui de ce fait ne contribuent plus

au courant électrique. Cette hypothèse est tout à fait en accord avec les résultats expérimentaux qui montrent clairement que les échantillons de moins bonne qualité présentent les plateaux les plus larges. On doit ainsi considérer, centrés sur chaque niveau de Landau, deux types d'états (Fig. A-5): Au centre il existe une bande relativement étroite d'états étendus dont les électrons transportent le courant électrique, sur les flancs il existe des queues d'états localisés dont les électrons ne peuvent transporter le courant électrique. Les plateaux de R_{xy} et R_{xx} apparaissent lorsque le niveau de Fermi est situé dans les états localisés. R_{xx} est alors nul car la conductivité σ_{xx} est nulle du fait que les électrons sont piégés, or comme $\rho_{xx} = \sigma_{xx} \big/ (\sigma_{xx}^2 + \sigma_{xy}^2)$ le zéro de σ_{xx} assorti du non-zéro de σ_{xy} entraîne que ρ_{xx} et par suite R_{xx} sont nuls. Quant à la résistance de Hall R_{xy}, elle reste constante car les états localisés ne contribuent pas au courant. Les sauts de R_{xy} et les pics de R_{xx} apparaissent lorsque le niveau de Fermi franchit les états étendus. En effet ces niveaux ne sont alors que partiellement occupés ce qui autorise les processus de diffusion et par conséquent augmente ρ_{xx} qui passe par un maximum quand le niveau de Fermi est situé au centre du niveau, où la diffusion est maximum. D'autre part quand le niveau de Fermi atteint ces états étendus de nouveaux électrons libres participent à la conduction de sorte que R_{xy} saute d'un plateau au suivant.

Il faut toutefois noter que si ce dernier modèle explique l'existence des plateaux il semble en désaccord avec leurs valeurs. En effet dans chaque niveau de Landau les électrons situés sur les états localisés ne participent pas au courant, de sorte que la contribution d'un niveau de Landau plein devrait être inférieure à e^2/h. En fait il est montré théoriquement que la contribution des états étendus à la conductivité de Hall est amplifiée pour compenser exactement la non contribution des états localisés.

Notons enfin que la constante $R_K = h/e^2$ porte le nom de *constante de von Klitzing* (prix nobel de physique 1985). Mesurée avec une incertitude inférieure à 10^{-9}, cette constante est depuis 1990 utilisée comme étalon de résistance, sa valeur est

$$R_K = h/e^2 = 25\,812{,}807\ \Omega \qquad \text{(A-94)}$$

A.4. Ancrage du niveau de Fermi par les états de surface

Les semiconducteurs de types IV, III-V et II-VI sont caractérisés par une coordination tétraédrique, chaque atome de volume établit des liaisons purement covalentes ou partiellement ioniques avec ses quatre premiers voisins, chaque liaison étant saturée par deux électrons. La symétrie de translation du cristal regroupe ces états en bandes d'énergie, les combinaisons liantes sont saturées et donnent naissance à la bande de valence entièrement occupée, les combinaisons antiliantes sont totalement vides, elles donnent naissance à la bande de conduction. Le sommet de la première et le bas de la seconde sont séparés par le gap du matériau.

A la surface du semiconducteur le nombre de proches voisins de chaque atome est réduit à trois de sorte que chacun des atomes de surface présente vers l'extérieur une orbitale atomique non saturée que l'on appelle *liaison pendante*. Dans les éléments IV ces orbitales sont occupées par 1 électron, dans les composés III-V elles sont occupées par 3/4 ou 5/4 d'électron suivant que la surface se termine sur les cations (III) ou les anions (V). Leurs fonctions d'onde sont les hybridations sp^3 données par les expressions (1-38), leurs énergies sont les composantes hybrides $E_h=(E_s+3E_p)/4$ qui apparaissent dans les expressions (1-41). Un calcul en liaison forte montre que pour les éléments IV l'énergie E_h est résonnante avec la bande de valence alors que pour les composés III-V et II-VI cette énergie est résonnante pour les anions de surface mais située dans le gap pour les cations. La symétrie de translation du cristal dans le plan de la surface regroupe ces états en bandes d'énergie bi-dimensionnelles. Ces bandes de surface se positionnent généralement en partie dans les continuums d'états de volume où elles donnent des *états résonnants* et en partie dans le gap où elles donnent ce que l'on a coutume d'appeler des *états de surface*. Considérons par exemple les éléments IV, les liaisons pendantes sont occupées par un seul électron et donnent de ce fait naissance à des bandes à moitié occupées c'est-à-dire de type métallique. Rappelons que dans un système uni-dimensionnel ces bandes à moitié occupées éclatent en deux sous-bandes, l'une pleine à basse énergie l'autre vide à haute énergie en doublant la taille de la cellule élémentaire et minimisant l'énergie du système, c'est la *transition de Peierls*. Cette transition métal-isolant ne se manifeste en toute rigueur que dans un système uni-dimensionnel. Toutefois dans un système bi-dimensionnel, comme la surface, il se produit un phénomène analogue qui abaisse l'énergie par une *reconstruction de la surface*. Considérons par exemple la surface (111) du silicium ou du germanium, chaque atome de surface présente une liaison pendante qui doit être occupée par un seul électron. Mais, comme nous l'avons déjà signalé, les calculs en liaison forte montrent que l'énergie de cette liaison est résonnante avec la bande de valence, il en résulte qu'à l'équilibre thermodynamique cette liaison doit être saturée par deux électrons, puisque située au-dessous du niveau de Fermi. Ceci ne peut se produire que par un transfert d'électrons depuis une moitié des atomes de surface vers l'autre moitié. Une telle redistribution de charges nécessite un réarrangement des atomes de surface c'est-à-dire une reconstruction de la surface.

Outre ces états de surface que l'on peut qualifier d'intrinsèques, l'adsorption d'atomes étrangers établissant des liaisons avec les atomes de surface,

sature les liaisons pendantes et crée des états de surface spécifiques de l'adsorbat. Considérons par exemple un atome monovalent. En première approximation cet atome établit des liaisons avec un atome de surface pour former une molécule hétéropolaire. Les orbitales liantes de la molécule sont occupées par deux électrons provenant respectivement de l'atome et de la liaison pendante, ses orbitales antiliantes sont vides. Si les atomes sont répartis uniformément sur la surface la symétrie de translation dans le plan de la surface regroupe ces états extrinsèques en bandes d'énergie bi-dimensionnelles.

Les bandes de surface ont des largeurs typiques de l'ordre de 0,5 eV, c'est-à-dire nettement inférieures à celles des bandes de volume, en conséquence leur densité d'états sont très importantes. La densité d'atomes de volume étant de l'ordre de $5\ 10^{22}$ cm^{-3} , la densité surfacique est de l'ordre de $7\ 10^{14}$ cm^{-2}. Dans la mesure où chaque atome de surface donne une liaison pendante doublement dégénérée, la densité surfacique dans chacune des deux bandes de surface est de l'ordre de $7\ 10^{14}$ cm^{-2}. Chaque bande ayant une largeur de l'ordre de 0,5 eV, la densité d'états par unité d'énergie et unité de surface est par conséquent de l'ordre de $1{,}4\ 10^{15}$ eV^{-1} cm^{-2}.

Considérons à titre d'exemple un semiconducteur de type *n* dopé avec une densité N_d de donneurs. Dans le volume du semiconducteur le niveau de Fermi est voisin du bas de la bande de conduction. En surface, certains électrons sont piégés sur les états de surface de sorte que leur nombre dans la bande de conduction diminue et par suite celle-ci s'éloigne du niveau de Fermi, qui reste horizontal à l'équilibre thermodynamique. Une zone de déplétion s'établit alors sous la surface et une différence de potentiel, appelée *potentiel de surface*, existe entre la surface du semiconducteur et son volume. La charge électronique piégée en surface est égale et opposée à la charge positive développée dans la zone de charge d'espace que constitue la région de déplétion. Pour calculer l'ordre de grandeur de cette dernière il suffit d'intégrer l'équation de Poisson dans laquelle la densité de charge est donnée par $\rho = e(N_d^+ - N_a^- + p - n)$. Afin de simplifier les calculs nous appellerons N_d l'excès de donneurs par rapport aux accepteurs, nous supposerons tous ces donneurs ionisés et nous supposerons la zone de déplétion vide de porteurs sur une profondeur *w*. L'équation de Poisson s'écrit alors simplement

$$\frac{d^2V}{dz^2} = -\frac{e\,N_d}{\varepsilon} \tag{A-95}$$

En prenant l'origine des distances à la surface du semiconducteur, une première intégration avec la condition $E = -dV/dz = 0$ en $z=w$, donne

$$\frac{dV}{dz} = -\frac{e\,N_d}{\varepsilon}(z - w) \tag{A-96}$$

En intégrant une deuxième fois dans la zone de charge d'espace et en appelant V_s le potentiel de surface on obtient

$$V_s = V_{surface} - V_{volume} = -\frac{e N_d}{\varepsilon}\int_{z=w}^{z=0}(z-w)\,dz = -\frac{e N_d}{2\varepsilon}w^2 \qquad \text{(A-97)}$$

Le potentiel de surface, qui résulte ici du piégeage des électrons, est négatif. L'épaisseur w de la zone de déplétion est reliée à ce potentiel par la relation

$$w = \left(-\frac{2\varepsilon}{e N_d}V_s\right)^{1/2} \qquad \text{(A-98)}$$

La charge d'espace surfacique de déplétion s'écrit donc

$$Q_{dep} = e N_d\, w = \left(-2\varepsilon e N_d V_s\right)^{1/2} \qquad \text{(A-99)}$$

La densité surfacique d'électrons piégés en surface est par conséquent

$$n_s = \frac{Q_{dep}}{e} = \left(-\frac{2\varepsilon N_d V_s}{e}\right)^{1/2} \qquad \text{(A-100)}$$

Considérons le cas limite où la courbure des bandes est telle que le semiconducteur passe de type n en volume à type p en surface, la valeur limite correspondante du potentiel de surface est alors donnée par $V_s \approx -E_g/e$. Dans ces conditions la densité surfacique d'électrons piégés est donnée par

$$n_s \approx \frac{1}{e}\left(2\varepsilon N_d E_g\right)^{1/2} \qquad \text{(A-101)}$$

En prenant à titre d'exemple $E_g \approx 1$ eV et $\varepsilon \approx 12\varepsilon_0$, l'expression précédente montre que lorsque le dopage du semiconducteur varie de N_d=10^{16} cm^{-3} à N_d=10^{19} cm^{-3} , la densité surfacique d'électrons piégés varie de $n_s \approx 3{,}6\ 10^{11}$ cm^{-2} à $n_s \approx 1{,}1\ 10^{13}$ cm^{-2}. Ces n_s électrons viennent occuper la bande de surface vide dont la densité d'états est $N_s \approx 1{,}4\ 10^{15}$ eV^{-1}cm^{-2}. L'occupation de cette bande par ces n_s électrons entraîne donc une élévation du niveau de Fermi donnée par $\Delta E_F = n_s / N_s$. Avec N_d=10^{16} cm^{-3} et N_d=10^{19} cm^{-3} on obtient respectivement $\Delta E_F = 0{,}26$ meV et ΔE_F = 8,2 meV. La différence entre les deux cas de dopage est donc $\delta E_F \approx 8$ meV. Ainsi lorsque le dopage du semiconducteur varie de 3 ordres de grandeur le niveau de Fermi ne se déplace que de 8 meV.

Cela signifie que malgré une variation très significative du dopage du semiconducteur la grande densité d'états de surface impose une variation insignifiante du niveau de Fermi. Ceci est à l'origine de ce que l'on a coutume d'appeler l'*ancrage du niveau de Fermi* par les états de surface. Dans un semiconducteur de type n le niveau de Fermi reste pratiquement calé au bas de la bande vide d'états de surface quelle que soit l'importance du dopage (Fig. A-7-a). Dans un semiconducteur de type p le même phénomène entraîne un piégeage des trous dans la bande d'états de surface pleine. La grande densité d'états de cette

bande entraîne un ancrage du niveau de Fermi au voisinage du sommet de cette bande (Fig. A-7-b). Il faut noter que le gap entre les deux bandes de surface est faible (~0,2 eV) de sorte que le niveau de Fermi varie peu du type *n* au type *p* pour un même semiconducteur. Lorsque le dopage du semiconducteur varie les paramètres qui varient sont les positions, relativement au niveau de Fermi, des bandes de conduction et de valence en volume et leur courbure au voisinage de la surface, c'est-à-dire le potentiel de surface.

L'ancrage du niveau de Fermi par la surface du semiconducteur dépend de plusieurs facteurs dont les principaux sont la nature de la surface, son type de reconstruction, sa qualité cristallographique et enfin son éventuelle contamination par des atomes étrangers. Une surface (110) de GaAs fraîchement clivée par exemple ne présente aucun ancrage du niveau de Fermi, mais l'adsorption d'oxygène sur une faible fraction de cette surface suffit pour entraîner un ancrage du niveau de Fermi à 500 meV au-dessus de la bande de valence en surface.

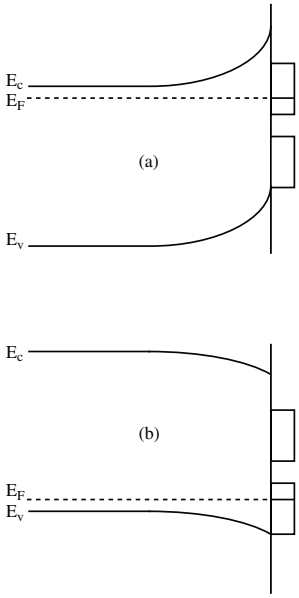


Figure A-7 : Ancrage du niveau de Fermi par les états de surface

A.5. Calcul du courant dans une structure MOS

Ce calcul important permet de relier le courant traversant un semiconducteur aux pseudo-niveaux de Fermi et au potentiel électrique présent dans le dispositif. Le calcul est fait pour un courant d'électrons mais le résultat s'applique également à un courant de trous.

La densité d'électrons peut s'écrire hors équilibre thermodynamique c'est-à-dire en présence d'un courant

$$n(x,y) = n_i e^{\frac{E_{Fn}(x) - E_i(x,y)}{kT}}$$

On considère pour simplifier une structure longitudinale, la position est repérée par la variable *x* et la profondeur par la variable *y*. On pourrait sans difficulté introduire la troisième dimension *z*. La grandeur $E_i(x,y)$ est l'énergie du silicium intrinsèque (environ au milieu du gap). Elle varie donc avec le potentiel électrique présent dans le semiconducteur comme il a été expliqué dans le chapitre 1.

$$E_i(x,y) = -eV(x,y) + C$$

La grandeur $E_{Fn}(x)$ est le pseudo-niveau de Fermi des électrons. A ce niveau de description cette grandeur pourrait apparaître dans un changement de variable comme une expression différente de la densité d'électrons puisqu'il y a correspondance entre *n* et $E_{Fn}(x)$. En fait il est nécessaire de faire une hypothèse physique supplémentaire en considérant que les électrons sont dans un état d'équilibre thermodynamique dans la bande de conduction. Il en est de même pour les trous dans la bande de valence. Cette hypothèse permet d'assimiler pseudo-niveau de Fermi et niveau de Fermi au niveau des contacts. Notons cependant que les pseudo-niveaux de Fermi sont définis localement (dépendance en fonction de la position) alors que les niveaux de Fermi sont définis pour l'ensemble du semiconducteur. On écrit pour les trous

$$p(x,y) = n_i e^{\frac{-E_{Fp}(x) + E_i(x,y)}{kT}}$$

Dans le cas d'un semiconducteur de type p traversé par un courant d'électrons (cas du NMOS), le courant de trous est nul ce qui implique que le pseudo-niveau de Fermi ne varie pas avec la position *x*. Ce calcul est fait en 1.5.1 c. On peut donc écrire

$$E_{Fn}(x) - E_i(x,y) = E_{Fn}(x) - E_{Fp} + E_{Fp} - E_i(x,y)$$

Exprimons maintenant le deuxième terme

$$E_{Fp} - E_i(x,y) = E_{Fp} - E_i(x,\infty) + E_i(x,\infty) - E_i(x,y)$$

Le choix *y* infini exprime simplement loin en profondeur dans le bulk.
Il suffit d'introduire le potentiel de Fermi du semiconducteur p et le potentiel électrique présent dans le semiconducteur.

$$E_{Fp} - E_i(x,\infty) = -e\phi_F$$

$$E_i(x,\infty) - E_i(x,y) = eV(x,y)$$

Le potentiel électrique est mesuré dans le modèle du MOS par rapport à celui en profondeur dans le bulk. Le niveau $E_i(x,\infty)$ dans le modèle du MOSFET ne dépend pas de *x*. La densité d'électrons s'écrit finalement

$$n(x,y) = n_i e^{\frac{E_{Fn}(x) - E_{Fp} - e\phi_F + eV(x,y)}{kT}}$$

Posons maintenant

$$eV_{CB}(x) = E_{Fp} - E_{Fn}(x) \tag{A-102}$$

Cette formule définit de manière rigoureuse la grandeur V_{CB} introduite dans le chapitre 8. En posant $\phi_t = \frac{kT}{e}$ on obtient finalement

$$n(x,y) = n_i e^{\frac{V(x,y) - \phi_F - V_{CB}(x)}{\phi_t}} \tag{A-103}$$

En profondeur dans le bulk la densité d'électrons s'écrit puisque le pseudo-niveau de Fermi est alors égal au niveau de Fermi commun aux électrons et aux trous.

$$n_b = n_i e^{\frac{-\phi_F}{\phi_t}}$$

En profondeur la densité de trous est la densité de dopants. Comme $n_b N_A = n_i^{\,2}$ on peut aussi écrire

$$n(x,y) = N_A e^{\frac{V(x,y) - 2\phi_F - V_{CB}(x)}{\phi_t}}$$

On écrit également

$$n(x,y) = n(x,\infty)\, e^{\frac{V(x,y) - V_{CB}(x)}{\phi_t}}$$

Dans le modèle du MOS, $n(x,\infty)$ est indépendant de *x*. Cette dernière relation permet de calculer facilement $\frac{\partial n}{\partial x}$

$$\frac{\partial n}{\partial x} = \frac{n(x,y)}{\phi_t}\left[\frac{\partial V(x,y)}{\partial x} - \frac{\partial V_{CB}(x)}{\partial x}\right]$$

On peut alors écrire le courant traversant une tranche *dy* dans le dispositif en fonction des composantes de diffusion et de dérive et de la largeur *W* du dispositif soit en intégrant la relation d'Einstein

$$dI(x,y) = Wdy\, e\, \mu_n\, n(x,y)\frac{\partial V(x,y)}{\partial x} - Wdy\, \mu_n\, \phi_t\, \frac{n(x,y)}{\phi_t}\left[\frac{\partial V(x,y)}{\partial x} - \frac{\partial V_{CB}(x)}{\partial x}\right]$$

$$dI(x,y) = Wdy\, e\, \mu_n\, n(x,y)\frac{\partial V_{CB}(x)}{\partial x} \qquad \text{(A-104)}$$

Remarquons que cette relation est valable en forte ou en faible inversion.
Si on suppose que la couche de conduction est très mince en surface, on obtient alors

$$I(x,y) = -W\mu_n\, Q'_I(x)\frac{\partial V_{CB}(x)}{\partial x}$$

Cette formule classique fait intervenir la charge d'inversion $Q'_I(x)$ par unité de surface.

Constantes physiques

Vitesse de la lumière dans le vide	$c =$ 2,997925 10^{8} ms^{-1}
Vitesse du son dans l'air	$v =$ 332 ms^{-1}
Permittivité du vide $(10^{7}/4\pi c^{2})$	$\varepsilon_o =$ 8,8542 10^{-12} Fm^{-1}
Perméabilité du vide $(4\pi 10^{-7})$	$\mu_o =$ 12,56637 Hm^{-1}
Constante de Boltzmann	$k =$ 1,38062 10^{-23} JK^{-1}
Constante de Planck	$h =$ 6,62620 10^{-34} Js
	$\hbar =$ h/2π = 1,0546 10^{-34} Js
Masse de l'électron	$m_o =$ 0,910956 10^{-30} kg
Masse du proton	$M_p =$ 1,67261 10^{-27} kg
Masse du neutron	$M_n =$ 1,674920 10^{-27} kg
Charge de l'électron	$e =$ - 1,60219 10^{-19} C
Rayon classique de l'électron $(e^{2}/m_o c^{2})$	$r_o =$ 2,81794 10^{-15} m
Rayon de Bohr $(\hbar^{2}/m_o e^{2})$	$a_o =$ 0,529177 10^{-10} m
Rydberg $(m_o e^{4}/2\hbar^{2})$	$R_y =$ 2,17991 10^{-18} J =13,606 eV
Magnéton de Bohr $(e\hbar/2m_o)$	$\mu_B =$ 9,2741 10^{-24} JT^{-1}
Constante de structure fine $(e^{2}/4\pi\varepsilon_o \hbar c)$	$\alpha =$ 7,29729 10^{-3}
	$\alpha^{-1} =$ 137,4
Constante de Stefan-Boltzmann	$\sigma =$ 5,6687 10^{-8} $Wm^{-2}K^{4}$
Unité de masse atomique	uma = 1,6598 10^{-27} kg
Nombre d'Avogadro	$N =$ 6,02217 10^{23} $mole^{-1}$
Volume molaire du gaz parfait	$V_o =$ 22,4207 10^{-3} $m^{3}mole^{-1}$
Constante des gaz parfait	$R =$ 8,317 JK^{-1}mole
Constante de Faraday	$F =$ 9,65219 10^{4} $Cmole^{-1}$
Constante gravitationnelle	$\gamma =$ 6,670 10^{-11} Nm^{2}/kg^{2}
Equivalent masse-énergie $(E = mc^{2})$	1 kg= 5,6100 10^{29} MeV
Longueur d'onde de Compton $(\hbar/mc)$	
de l'électron	$\lambda_e =$ 2,42626 10^{-12} m
du proton	$\lambda_p =$ 1,32141 10^{-15} m
du neutron	$\lambda_n =$ 1,31959 10^{-15} m
Energie thermique $(E_{th} = kT)$	
Température ambiante (T=300 K)	$E_{th} =$ 25,85 meV
Azote liquide (T=77,3 K)	= 6,67 meV
Hélium liquide (T=4,2 K)	= 0,36 meV
Cryostat à dilution (T≈30 mK)	= 2,59 μEv

Bibliographie

- ALOK D., BALIGA B.J., SiC and Rel. Mat. Conf. Proc., 749 (1996)
- ANDO T. - Phys. Rev. 13, 3468 (1975).
- ASHCROFT N.W., Physique des solides, EDP (1976)
- ASPNES D.E., STUDNA A.A., Phys. Rev. B27, 985 (1983).
- BAKER R.J., CMOS circuit design, (2005).
- BARNES J.J., SHIMOHIGASHI K., DUTTON R.W., IEEE Trans El Dev. ED-26, 446 (1979).
- BASSANI F., PARRAVICINI P., Electronics states and optical transitions in solids. Ed. R.A. BALLINGER - Pergamon Press, Oxford (1975).
- BINARI S.C., Electrochem. Soc. Proc. 95-21, 136 (1995).
- BASTARD G., Wave mechanics applied to semiconductor heterostructures, Ed. Les Editions de physique (1988).
- CHANG M.F., Sol. Stat. Electronics 38, 1972 (1995).
- COWLEY A.M., SZE S.M., J. Appl. Phys. 36, 3212 (1965).
- ESAKI L., TSU R., IBM Journal of Research and Development, 14, 61 (1970).
- FANET H., Micro et nano-électronique, Dunod (2006).
- FEYNMANN R., Lectures on computation, Perseus Books (2000).
- FRANCK, D.J., Device scaling limits or Si MOSFETs, proc. IEEE, vol. 89, march 2001.
- GROVE A.S., Physics and technology of semiconductor devices, Ed. John Wiley and Sons, New York (1967).
- GUNN J.B., Solid State Comm. 1, 88 (1963).
- HANDBOOK on semiconductors. Ed. North-Holland publishing Company - Amsterdam Series editor T.S. Moss (1981).
- HARRISON W.A., Electronic structure and the properties of solids, Ed. W.H. Freeman and company - San Francisco (1980).
- KITTEL C., Physique de l'état solide, Ed. Dunod (1983).
- KEYES R.W., Fundamental limits or silicon technology, Proc. IEEE, vol. 89, march 2001.
- LITTLEJOHN M.A., KWAPIEN W.M., GLISSON T.H., HAUSER J.R., HESS K., J. Vac. Sci. Technol. B1, 445 (1983).
- MATHIEU H., AUVERGNE D., MERLE P., Phys. Rev. B12, 5846 (1975).
- MATHIEU H., Thèse de Doctorat d'Etat, Montpellier (1969).
- MATHIEU H., LEFEBVRE P., CHRISTOL P., Phys. Rev. B46, 4092 (1992).
- MESSIAH A., Mécanique quantique, Dunod (1969).

- NAKAMURA S., SENOH M., NAGAHAMA S., IWASA N., YAMADA T., MATSUSHITA T., KIYOKU H., SUGIMOTO Y., KOZAKI T., UMEMOTO H., SANO M., CHOCHO K., Proc. 2th Int. Conf. Nitride Sem. - Tokushima, 444 (1997).
- NAKAMURA S., Physics and Applications of Group III nitride Semiconductor Compouds, Edit B.GIL, Oxford University Press, 391 (1998).
- NGO C.et H., Les semiconducteurs, Dunod (2003).
- NICOLLIAN E.H., BREWS J.R., MOS Physics and technology, Ed. John Wiley and Sons, New York (1982).
- NILES D.W., MARGARITONDO G., Phys. Rev. B34, 2923 (1986).
- NOZIERES P, Le problème à N corps, Dunod (1963)
- ONDA S., KUMAR R., HARA K., Phys. Stat. Sol., 162, 369 (1997).
- OUISSE T., Phys. Stat. Sol., 162, 339 (1997).
- PALMOUR J.W., ALLEN S.T., SINGH R., LIPKIN L.A., WALTZ D.G., SiC and Rel. Mat. Conf. Proc., 813 (1996).
- PAN J.N., COOPER J.A., MELLOCH M.R., Electronics Letters 31, 1200 (1995).
- PANKOVE J.I. 2th Euro. Conf. on Silicon Carbide and Related Materials. G1, Montpellier (1998).
- PIDGEON C.R., GROVES S.H., Phys. Rev. 186, 824 (1969).
- ROSS D.A., Optoelectronic devices and optical imaging techniques, Ed. The Macmillan Press, LTD, London (1979).
- SZE S.M., Physics of semiconductor devices, Ed. John Wiley and Sons, New York (1969).
- TSIVIDIS Y.P., The MOS transistor, Mc Graw-Hill (1988)
- VAN der ZIEL A., Solid State Physical electronics, Ed. Nick Holonyak Jr. Prentice-Hall Inc. New jersey (1976).
- VAPAILLE A., Physique des dispositifs à semiconducteurs - T1 Electronique du silicium homogène, Ed. Masson et Cie, Paris (1970).
- VON KLITZING K., DORDA G., PEPPER M., Phys. Rev. Lett. 45, 494 (1980).
- VON KLITZING K., EBERT G., Solid State Sciences 53, 242 (1984).
- WANNIER G.H., Elements of solid state theory-Cambridge University Press, Cambridge (1959).
- WEISBUCH C., Annales de Physique C2, 20, 353 (1995).
- WEISBUCH C., VINTER B., Quantum Semiconductor Structures, Academic Press (1991).

Index

absorbeur, 528
absorption, 499, 500
  fondamentale, 484
accepteur, 61
affinité électronique, 110
Air Masse, 498
amorphes, 527
amplificateur de tension, 462
ancrage du niveau de Fermi, 115, 239, 814
APCVD, 667
appauvrissement, 386
approximation
  adiabatique, 24
  à un électron, 24
  de la masse effective, 582
  des bandes paraboliques, 35
  du puits triangulaire, 599
  semi-classique, 34
attaque chimique, 669
atténuation, 499, 500
auto-aligné, 673
auto-alignement, 450
bande
  de conduction, 19, 29
  de valence, 19, 29
  interdite, 19, 29
  permise, 40
  plate, 229
barrière de Schottky, 123
base, 194, 198
  commune, 212
bâtonnets, 570
BESOI, 757
bipolaire, 344
blocage de Coulomb, 735, 737
boite quantique, 791
bonding, 673
bosons, 46
bottom-up, 752
brillance, 541
bruit, 124
  basse fréquence, 126, 455
  de génération-recombinaison, 131
  de grenaille, 125, 128, 455
    d'une diode, 217
  dit en *I/f*, 126
  induit par la grille, 460
  thermique, 125, 127, 455
    du canal, 455, 459
bulk, 386
canal, 343
  court, 416
capacité
  de couplage, 440
  de diffusion, 163, 166
  de transition, 161, 163
  différentielle, 243
  équivalentes, 444
  grille-source, 446
  parasites de couplage, 450
carborundum, 765
carbure de silicium, 761, 765
cavité Pérot-Fabry, 537
CD, 498, 566
CD-RW, 568
cellule
  de base, 3
  multicolores, 529
  photoconductrice, 512
  photovoltaïque, 516
centre de recombinaison, 90, 493
champ
  critique, 411
  de Hall, 800
  électrique effectif, 598
changement de phase, 568
charge
  d'inversion, 283
  d'accumulation, 231
  de déplétion, 231
  d'espace, 97, 138
  image, 112, 595

par unité de surface, 279
« virtuelle », 430
Chemical Solution Deposition, 668
chromaticité, 571
circuit CMOS, 673
claquage de la jonction, 177
coefficient
γ, 285
d'absorption, 494, 496
par porteurs libres, 546
d'amplification, 545
de capture et d'émission, 90
de corrélation, 126
de diffusion effectif, 182
de Hall, 800, 805
de réflexion, 496
de transfert des électrons, 332
de transmission, 713
collecteur, 194, 200
commun, 216
colorimétrie, 570
commutation, 466, 467
complexion, 43
composantes trichromatiques, 570, 572
condition de neutralité électrique, 180
conductance
de diffusion, 166
de drain, 351
de sortie, 436, 437, 438
du canal, 699
du fil quantique, 699
du fil, 699
conducteur, 19
conduction, 71
conductivité, 74
du semiconducteur, 104
cône, 570
d'extraction, 535
confinement quantique, 681
conservation du courant, 86
consommation, 467
statique, 474
constante
de diffusion ambipolaires, 506
de diffusion, 77
de Hall, 805
de Richardson, 119
de Rydberg, 12
de von Klitzing, 699, 811
diélectrique relative, 496
contact métal-semiconducteur, 227
coordonnées trichromatiques, 572
couche d'inversion, 115, 386
de l'hétérojonction, 644
couche tampon, 682
courant
de conduction, 74
de déplacement, 85
de diffusion, 77, 519, 520
de drain, 343, 344
de génération, 169, 519
de génération-recombinaison, 169
de pincement, 356
de recombinaison, 170
de saturation, 155, 346, 357, 366, 399
de seuil, 554, 555, 559
d'émetteur, 200
électromoteur, 524
transitoire, 426
transverse, 393
tunnel, 329
courants, 71
courbure des bandes, 230
cristaux
covalents, 2
ioniques, 2
moléculaires, 3
croissance électrolytique, 670
défaut, 672
dégénéré, 45, 55
délocalisées, 19
densité
de courant, 71
d'états, 8
bidimensionnelle, 680
par unité d'énergie, 40
équivalentes d'états, 56
spectrale, 126
d'intercorrélation, 219
dépôt
de type CSD, 668
en phase vapeur (CVD), 667
MOCVD, 668
par, 665
par laser pulsé (PLD), 666
diagramme de chromaticité CIE, 572
diamant, 761, 762
DIBL, 416, 421
diélectriques poreux, 675
diffusion, 71, 499

ambipolaire, 505
par les impuretés ionisées, 683
thermique, 672
diode
électroluminescente - LED, 531
laser, 542, 544
Schottky, 232, 237
directions de haute symétrie, 23
dislocation, 669
dispersion, 467
intermode, 500
distribution de Bose-Einstein, 47
divergence, 86
donneur, 61
donneurs et accepteurs, 61
dopages homogènes, 202
double hétérojonction, 563
drain, 343, 386
droite des pourpres purs, 573
durée de vie
des porteurs minoritaires, 90
non radiative, 493
radiative, 492
DVD, 498, 566
échelle de Wannier-Stark, 704
effet
body, 285, 399
d'avalanche, 178, 498
d'échange, 611
Destriau, 498
Early, 207
Gunn, 688
bidimensionnel, 686
Hall, 799
quantique, 808
RWH, 411, 687
Schottky, 112
Shubnikov-de Haas, 810
transistor, 194
tunnel, 122, 498
résonnant, 709
Wannier-Stark, 704
Zener, 178
efficacité de modulation, 539
électronégativité, 239
électrons
de valence, 11, 13
du cœur, 13
ELOG, 773, 787
émetteur, 194, 528
commun, 214
émission
de champ, 122
spontanée, 484
stimulée, 484
thermoélectronique, 116
énergie
de confinement, 581, 622
de désordre, 51
de Fermi, 52
interne, 43
libre, 51
enregistrement magnéto-optique, 568
enrichissement, 386
entropie, 51
épitaxie par jets moléculaires (MBE), 665
équation
de diffusion ambipolaire, 506
de Laplace, 97
de Poisson, 97
d'Ebers-Moll, 196, 201
espace
des phases, 43
réciproque, 7
étage inverseur, 470
étalon de résistance, 696, 699
état
bloqué, 469
conducteur, 469
cristallisé, 2
de surface, 114, 493, 812
d'interface, 115, 238
résonnant, 812
stationnaires, 11
évaporation, 664
facteur
de confinement, 719
de forme, 525
de réaction d'Early, 210
de remplissage, 525
d'émission spontanée, 730
faible
injection, 89
inversion, 286, 387
niveau d'injection, 179
fermions, 46
fil quantique, 696
flux chromatique, 572
fonction
d'Airy, 599, 605

de distribution, 43
de Fermi, 53
de Green, 28
de transfert, 126
enveloppe, 583
variationnelle, 605
force
électromotrice, 524
discontinuité, 308, 329, 330
injection, 179, 182
inversion, 387
fréquence de coupure, 369, 371, 559
gain
de la cellule, 514
modal, 719
périodique résonnant, 726
GaN, 771
gap, 19
direct, 31, 40, 485
indirect, 31, 40, 485
gaz d'électrons, 640
génération-recombinaison, 167, 197, 87
germanium, 29
gravure
anisotrope, 661
humide, 660
ionique réactive, 662
isotrope, 661
par plasma, 662
par *sputtering*, 662
sèche, 662
grille, 270, 344, 386
GRIN, 565
GRIN-SCH, 565
groupe
de symétrie, 3
ponctuel, 3
spatial, 3
Hamiltonien de Hartree-Fock, 25
harmonique sphérique, 12
haute température, 756
HD-DVD, 567
HEMT, 639
hétéroépitaxie basse température, 681
hétérojonction, 293, 297, 613, 764
hétérostructure, 613
de type I, 614
de type II, 614
homojonction, 297
HPCVD, 667
hybridation, 16
hypothèse
de faible injection, 197
de faible niveau d'injection, 148
ideality, 288, 402
IGFET, 344
implantation ionique, 670
impuretés, 90
indice
de comportement, 775
de l'interface, 239
de Miller, 5
de réfraction, 496, 547
indiscernabilité, 45
ingénierie de bandes, 681
injection électrique, 497
inscriptible, 566
intégrales de recouvrement, 23
interface, 108
inter-sous-bande, 682
inverseur, 464
inversion, 274
de population, 490, 543
ionicité, 239
isolant, 19
ITRS, 425
JFET, 343, 344
jonction
abrupte, 138
base-collecteur, 195, 196
émetteur-base, 195
quelconque, 161
largeur effective, 170, 202
laser
à boites quantiques, 734
à cascade quantique, 731
à cavité verticale-VCSEL, 725
à double hétérojonction, 719
à injection, 544, 545
à puits quantiques, 565, 718
bleu, 761, 787
sans seuil, 730
LED bleue, 784
LEO, 773
liaison pendante, 114, 812
libre parcours moyen, 72, 746
lieu spectral, 573
lift-off, 677
lithographie, 656, 657
à base de rayons X, 659

par contact, 658
par projection, 658
UV, 659
logique
CMOS, 470
MOS, 465
loi
de Coulomb, 85
de Joule, 75
de Moore, 423
d'Ohm, 75
longueur
de cohérence de la phase, 746, 680
de Debye, 99, 144
de diffusion, 96, 151
effective, 183
d'écran bidimensionnelle, 684
d'onde, 480
de de Broglie, 579
de Fermi, 746
LPCVD, 667
luminosité, 571
magnétorésistance
longitudinale, 807
transversale, 802
maille
élémentaire, 3
primitive, 3
masque, 657
masse effective, 32, 33, 34
de conductivité, 78
des électrons, 79
des trous, 82
de densité d'états, 42
de la bande de valence, 42
des électrons, 39
des trous, 39
longitudinale, 36, 601
transverse, 36, 601
matrice
admittance, 211
de diffusion, 682
de transfert, 711, 712
MBE, 773
MEE, 773
MESFET, 343, 344, 358
MESFET-GaAs à canal court, 371
mésoscopique, 680
Metal Organic Chimical Vapour
Deposition, 668
métal-vide-semiconducteur, 258
métaux, 2
méthode
APW, 27
BKW, 793
cellulaire, 27
des liaisons fortes, 26
LCAO, 26
MERIE, 662
OPW, 26
microcavité, 730
miroir de Bragg, 537, 725, 727
MISFET, 764
mobilité, 683
de Hall, 805
des porteurs, 73, 74
effective, 403
MOCVD, 668, 773
modèle
canal court, 405
canal long, 394
de diffusion, 323
de Kronig-Penney, 630
trichromatique, 570
petits signaux, 435
MODFET, 639, 640
monomode, 552
MOS
canal n, 404
canal p, 404
en commutation, 464
MOSFET, 343, 344
multimode, 552
multiplicateurs de Lagrange, 45
multipuit quantique, 624, 722
nanofil, 751
nano-impression, 677, 752
nanotube, 749
neutralité électrique, 97, 98
niveau
de Fermi, 52, 63, 66
du vide, 108, 228
de Landau, 808
nombre
d'états, 40
quantique
azimutal, 12
principal, 12
non dégénéré, 55
ohmique, 234, 236, 237

onde
de Bloch, 10
plane
augmentée, 27
orthogonalisée, 26
orbitale, 16
antiliante, 15
liante, 15
σ, 16
atomique, 12
moléculaire, 14
oscillateur de Bloch, 700
oscillations de Bloch, 701
oxyde
à haute permittivité, 422
de passivation, 673
paquet d'ondes, 11, 481
parallèle, 681
paramètre
α, 429
FF, 525
de bandes de valence, 38, 39
particules
identiques, 45
quasi-libres, 32
période, 480
perpendiculaire, 681
phénomène mésoscopique, 681
phonon
acoustique, 684
optique, 409, 684
photocourant, 517
de diffusion, 516
de génération, 516
photodétecteur, 504
photodiode
pin, 522, 783
à avalanche, 522
Schottky, 523
photoexcitation, 494
photon, 481
phototransistor, 523
HFET, 783
piège à électron, 90
pinch-off, 397
plan réticulaire, 4
point, 29
chromatique, 571
couleur, 571
polycristallines, 527
polytypisme, 766
pompé, 543
porteur
balistique, 408
chaud, 407, 408
majoritaire, 62
minoritaire, 62
potentiel
chimique, 52, 390
cristallin, 25
de contact, 267
de déplétion, 590
de Fermi, 267, 276
de Hartree, 590
de surface, 115, 284, 813
d'inversion, 590
électrique, 85
image, 112, 594
triangulaire, 598
première zone de Brillouin, 7
pré-structure, 228
problème à *n* corps, 25
procédé Czochralski, 669
processus de relaxation, 409
pseudo-potentiel chimique, 275, 390, 392
pseudo-continuité, 308, 329
pseudo-niveaux de Fermi, 53, 486
pseudo-potentiel, 27
puissance
d'un signal de bruit, 127
débitée, 524
puits, 673
quantique, 613, 696
couplés, 625
pulsation, 481
quantité de mouvement, 482
quantum de résistance, 748
quasi-particules libres, 32
rapport signal sur bruit, 127
rayon de Bohr, 12
effectif, 591
rayonnement électromagnétique, 480
réacteur à décharge inductive, 662
réacteur ECR, 662
recombinaisons directes, 88
recuit, 672
redresseur, 237
régime
d'accumulation, 259, 302
d'inversion, 261

de bandes plates, 259, 300
de déplétion, 261, 302
de faible inversion, 400
de forte inversion, 261, 394
de gel, 69
de pincement, 346
de saturation, 346, 353, 379, 414
de survitesse, 411
d'épuisement des donneurs, 69
dynamique, 210, 389, 431
intrinsèque, 68
linéaire, 346, 355, 373
quasi-statique, 389, 425
sous-linéaire, 377
statique, 389
de pincement, 397
régions neutres, 185
règle de l'affinité électronique, 298
réinscriptible, 566
relation
de dispersion, 10, 483
de Fick, 77
d'Einstein, 83
rendement, 526
global, 536
optique, 534
quantique
externe, 536
interne, 533
réponse, 514
réseau
cubique
centré, 6
faces centrées, 6
de Bravais, 7
résine, 657
résistance
d'accès à la base, 220
de base r_b, 201
de collecteur, 201
de Hall, 800
d'émetteur, 201
différentielle négative, 687, 688, 706
série, 155
tunnel, 738
de Hall, 800
retard, 467, 468
RIE, 771
roll-off, 421
RST, 687
rugosité des interfaces, 683
saturation de la vitesse, 407, 414
SCH, 565, 721
schéma électrique équivalent, 439
semi-classique, 680
semiconducteur, 19
à grand gap, 760
de type n, 62
de type p, 63
dégénéré, 57
extrinsèque, 61
intrinsèque, 60, 62
multivallée, 29, 600
univallée, 30
de types différents, 315, 322, 327
isotypes, 314, 321, 324
semi-isolant, 62
séries de sous-bandes d'énergie, 601
SET, 743
seuil
V_T, 399
effectif, 416
silicium, 29
polycristallin de grille, 673
sur isolant, 756
SIMOX, 757
Smart-cut, 757
source, 343, 386
sous-bande, 682
sous-gravure, 661
spacer, 640
sputtering, 662
RF, 665
stationnaires au sens large, 126
statistique
de Boltzmann, 43
de Bose-Einstein, 46
de Fermi-Dirac, 48
quantique, 45
structure
à multipuits quantiques, 624
de sous-bandes d'énergie, 584
GRIN-SCH, 722
MIS, 578
substrat diamant, 759
superradiante, 547
superréseau, 627, 703
nipi, 638
superstructure de bandes, 628
super-zone de Brillouin, 628

surface d'onde, 481
système
à *une* dimension, 697
cristallin, 4
cubique, 5
taux
de génération, 88
de génération
thermique, 87
de recombinaison, 88
net d'émission, 487, 491
technologie CMOS, 387
TEGFET, 639
température électronique, 408
temps
de collision, 72
de descente, 468
de montée, 467
de recouvrement, 171
de relaxation, 73
de l'énergie, 410
diélectrique, 104
du moment, 409
de réponse, 537
de la photodiode, 521
de stockage, 173
de transit, 431
tension
de bandes plates, 267, 394
de claquage, 178
de commutation de l'étage, 471
de cut-off, 347, 355
de diffusion, 140, 241, 295
de Hall, 800
de pincement, 349, 362
de saturation, 346
de seuil, 285
de l'étage, 469
Zener, 178
théorème
de Gauss, 86
de Wiener Khintchine, 126
théorie de Shockley-Read, 90
thermalisation, 486
tolérance au bruit, 467
top-down, 752
tranchée d'isolation, 673
transconductance, 351, 437, 438
transistor
à effet de champ, 343
à barrière de Schottky, 344, 358
à gaz d'électrons bidimensionnel-TEGFET, 639
à grille isolée, 344
à jonction, 344
à hétérojonction, 332
à un électron, 743
bipolaire, 193, 194
drift, 204
MOSFET, 385
unipolaire, 344
de Peierls, 812
radiative, 484, 485
perpendiculaire, 700
travail de sortie, 108
triangle des couleurs, 571
trou, 32
léger, 37, 78
lourd, 37, 78
TSR, 758
type à confinement séparé, 787
univallée, 601
varactors, 163
vecteur
d'onde, 10, 480
propagation, 10
primitifs, 3
vitesse
balistique, 408
de dérive, 73
de groupe, 33, 481
de phase, 481
de recombinaison, 251
à l'interface, 252
en surface, 252, 493, 494
de saturation, 408
drift, 73
thermique, 72
VLS, 752
wafer, 656
YAG, 790
zone
de charge d'espace, 143, 240
de déplétion, 230
de recouvrement, 450

SCIENCES SUP

Henry Mathieu
Hervé Fanet

6e édition

PHYSIQUE DES SEMICONDUCTEURS ET DES COMPOSANTS ÉLECTRONIQUES

Cet ouvrage, qui dresse un panorama complet du domaine des semiconducteurs et des composants électroniques, s'adresse aux étudiants de Master et aux élèves ingénieurs comme aux chercheurs et ingénieurs intéressés par une description générale des composants semiconducteurs.

Depuis leurs fondements jusqu'à leurs applications dans les composants, tous les phénomènes de la physique des semiconducteurs et des composants électroniques sont abordés et expliqués dans ce manuel, étape par étape, calcul par calcul, de façon détaillée et précise.

Cette sixième édition actualisée tient compte des progrès techniques et de quelques effets physiques qui interviendront dans les composants de demain. La description du transistor MOS ainsi que les procédés de fabrication associés ont été complétés pour intégrer les avancées technologiques les plus récentes. Des composants nouveaux à base de nanofils et nanotubes sont introduits. Elle propose de plus des exercices et des questions sur les points les plus importants, permettant à l'étudiant d'aborder des applications plus spécifiques (diode avalanche, thyristor, CCD…).

HENRY MATHIEU
est professeur à l'université Montpellier II.

HERVÉ FANET
est ingénieur de recherche au CEA LETI, chargé de coopérations scientifiques et d'actions de formation au sein du pôle MINIATEC (micro et nanotechnologies) de Grenoble.

MATHÉMATIQUES
PHYSIQUE
CHIMIE
SCIENCES DE L'INGÉNIEUR
INFORMATIQUE
SCIENCES DE LA VIE
SCIENCES DE LA TERRE

LICENCE MASTER DOCTORAT
1 2 3 4 5 6 7 8

ISBN 978-2-10-054134-8

www.dunod.com

DUNOD