OPERATIONAL AMPLIFIER SPEED AND ACCURACY IMPROVEMENT

THE KLUWER INTERNATIONAL SERIES IN ENGINEERING AND COMPUTER SCIENCE

ANALOG CIRCUITS AND SIGNAL PROCESSING

Consulting Editor: **Mohammed Ismail.** *Ohio State University*

Related Titles:

LOW POWER ANALOG CMOS FOR CARDIAC PACEMAKERS
Silveira and Flandre
ISBN: 1-4020-7719-X

MIXED-SIGNAL LAYOUT GENERATION CONCEPTS
Lin, van Roermund, Leenaerts
ISBN: 1-4020-7598-7

HIGH-FREQUENCY OSCILLATOR DESIGN FOR INTEGRATED TRANSCEIVERS
Van der Tang, Kasperkovitz and van Roermund
ISBN: 1-4020-7564-2

CMOS INTEGRATION OF ANALOG CIRCUITS FOR HIGH DATA RATE TRANSMITTERS
DeRanter and Steyaert
ISBN: 1-4020-7545-6

SYSTEMATIC DESIGN OF ANALOG IP BLOCKS
Vandenbussche and Gielen
ISBN: 1-4020-7471-9

SYSTEMATIC DESIGN OF ANALOG IP BLOCKS
Cheung & Luong
ISBN: 1-4020-7466-2

LOW-VOLTAGE CMOS LOG COMPANDING ANALOG DESIGN
Serra-Graells, Rueda & Huertas
ISBN: 1-4020-7445-X

CIRCUIT DESIGN FOR WIRELESS COMMUNICATIONS
Pun, Franca & Leme
ISBN: 1-4020-7415-8

DESIGN OF LOW-PHASE CMOS FRACTIONAL-N SYNTHESIZERS
DeMuer & Steyaert
ISBN: 1-4020-7387-9

MODULAR LOW-POWER, HIGH SPEED CMOS ANALOG-TO-DIGITAL CONVERTER FOR EMBEDDED SYSTEMS
Lin, Kemna & Hosticka
ISBN: 1-4020-7380-1

DESIGN CRITERIA FOR LOW DISTORTION IN FEEDBACK OPAMP CIRCUITE
Hernes & Saether
ISBN: 1-4020-7356-9

CIRCUIT TECHNIQUES FOR LOW-VOLTAGE AND HIGH-SPEED A/D CONVERTERS
Walteri
ISBN: 1-4020-7244-9

DESIGN OF HIGH-PERFORMANCE CMOS VOLTAGE CONTROLLED OSCILLATORS
Dai and Harjani
ISBN: 1-4020-7238-4

CMOS CIRCUIT DESIGN FOR RF SENSORS
Gudnason and Bruun
ISBN: 1-4020-7127-2

ARCHITECTURES FOR RF FREQUENCY SYNTHESIZERS
Vaucher
ISBN: 1-4020-7120-5

THE PIEZOJUNCTION EFFECT IN SILICON INTEGRATED CIRCUITS AND SENSORS
Fruett and Meijer
ISBN: 1-4020-7053-5

CMOS CURRENT AMPLIFIERS; SPEED VERSUS NONLINEARITY
Koli and Halonen
ISBN: 1-4020-7045-4

OPERATIONAL AMPLIFIER SPEED AND ACCURACY IMPROVEMENT

Analog Circuit Design with Structural Methodology

by

Vadim V. Ivanov

Texas Instruments, Inc.

and

Igor M. Filanovsky

University of Alberta

KLUWER ACADEMIC PUBLISHERS

NEW YORK, BOSTON, DORDRECHT, LONDON, MOSCOW

eBook ISBN: 1-4020-2517-3
Print ISBN: 1-4020-7772-6

Created in the United States of America

Visit Kluwer Online at: http://kluweronline.com
and Kluwer's eBookstore at: http://ebooks.kluweronline.com

Dedication

To my father Valery Nikolayevich Ivanov who led and inspired me to become an engineer

Vadim Ivanov

Contents

Preface

"Operational Amplifier Speed and Accuracy Improvement" focuses on the analog integrated circuit design methodology that is pushing the state of art limits. OpAmp development is used as an example, but the methodology is applicable in any area of the analog IC design. This work is useful for analog IC designers who would like to create new and superior circuits, as well as for graduate students who want to leapfrog the lengthy process of detailed studying of the huge legacy of analog circuits and accelerate their way to professional excellence.

The basics of this methodology, which we call structural design, were developed in 1960s and 1970s in the USSR, and were used in development of control systems for hydrofoil ships and cruise missiles. Except for a few recent papers, there are no adequate references to this methodology in English. In its analytical part, the structural design approach is close to the area of modern philosophy called *systems thinking*.

We have tailored the structural design methodology for analog IC development. This approach has influenced the designs of OpAmps, references, instrumentation and power amplifiers developed by the Tucson division of Texas Instruments, Inc. (former Burr-Brown). Effectiveness of this methodology has been confirmed by more than 30 patents and patent applications received and filed in last few years.

The circuits shown in this book have been used in micropower (< 1uA of I_q), and high power (3A load current) OpAmps, in the fastest CMOS amplifiers developed by industry, in the most accurate CMOS and bipolar OpAmps, and in many general purpose OpAmps as well.

In chapter 1, we describe the basic steps of analog design, outline the situation with modern analog processes, discuss the requirements of modern

OpAmps, and review the basic parameters and characteristics of bipolar and CMOS transistors that are important for successful analog design.

Chapter 2 outlines the application of signal graphs in the structural design methodology, discusses the content of analog cell libraries and proposes additional cells for these libraries that are proven to be useful in the analog design.

Chapter 3 is dedicated to the OpAmp biasing: supply-insensitive, proportional to the absolute temperature, and other biasing cores; current sources with high output impedance and low saturation voltage; low-noise charge pumps for bootstrapping the tail current source.

Chapter 4 examines structures which improve the power to speed ratio of the OpAmp, while maintaining high gain. The gain stage and amplification stage are differentiated, and the voltage and current gain boost circuits are discussed.

Chapter 5 discusses the input stages (including rail-to-rail stages with stable transconductance), the offset and temperature drift trimming techniques, and input protection circuits.

Chapter 6 describes the intermediate OpAmp stages - primarily the folded cascode which is an essential part of any amplifier with a rail-to-rail or single-supply input. The improvement of this stage's parameters, voltage gain boost and voltage clamping are discussed.

Chapter 7 is dedicated to the output class AB stage, its control structure, regulation and stability of the quiescent current, with emphasis on a low supply and rail-to-rail output capability.

Chapter 8 describes the implementation of special function circuits which protect or extend the boundaries of the OpAmp functionality (slew rate boost, current limiting, fast shutdown and start-up, fast overload recovery).

Chapter 9 gives a top-down OpAmp design example. Some practical tricks and honing of the common sense in distributing the current budget, in choice of the component dimensions is the subject of this chapter. The reader is moving from the general idea to final implementation and test results.

The Appendix contains the article "Structural properties and linear transformations in the multidimensional systems with symmetric links", which reveals part of the theory behind the structural design methodology. This is an adapted translation from Russian of the article written by Valery Ivanov, one of the inventors of the structural design methodology, to whom both authors are conveying their respect and gratitude.

Authors

Acknowledgments

We both express our gratitude to David Spady (Texas Instruments, Tucson) for comments on manuscript, to Misha Ivanov (Texas Instruments, Germany) who read and commented on chapters 1-5, to Rod Burt (Texas Instruments, Tucson), who took a burden to review the book.

We are thankful to David Jones, Wally Meinel, Sergey Alenin, Greg Johnson (Texas Instruments, Tucson), to Prof. Rob Fox (University of Florida) who assisted with critique and encouragement.

Prof. Mohammed Ismail (Spirea AB, Sweden) persuaded us to write this book.

One of the authors would like to express his gratitude to the staff of Electronics Research Laboratory, Delft University of Technology, The Netherlands, and especially to Dr. C. J. M. Verhoeven and Dr. J. R. Long. The unique congenial atmosphere of this laboratory encouraged this author to writing this book and making the first steps to its completion.

Notations

OpAmp	operational amplifier
A_v	voltage gain
β	current gain of the bipolar transistor
C	capacitor value
C_L	load capacitance
C_{ox}	sheet capacitance of the gate oxide
C_M	Miller capacitor value
C_{eqv}	total equivalent capacitance at the node
ESD	electro-static discharge
f	frequency
g_m	small signal transconductance
GBW	gain-bandwidth product
f_T	transit frequency of the transistor
i	small-signal current
I	current
I_C	collector current
I_B	base current
I_{BD}	body current
I_D	drain current
I_E	emitter current
I_q	quiescent current
I_S	source current
I_s	reverse current of the p-n junction
k	Boltzman's constant
K	$= \mu C_{ox}(W / L)$
L	MOS transistor channel length

M	MOS transistor
normal temperature	27^{o}C = 300K default simulation temperature for SPICE
PTAT	proportional to the absolute temperature
q	charge of the electron
Q	bipolar transistor
r_{bp}	base body resistance
r_{dp}	drain body resistance
r_{ep}	emitter body resistance
r_{sp}	source body resistance
R	resistance
R_L	load resistance
S_E	bipolar transistor emitter area
T	absolute temperature
TC	temperature coefficient
um	micrometer
v	small-signal voltage
V	voltage
V_A	Early voltage
V_{BE}	base-emitter voltage
V_{CE}	collector-emitter voltage
V_{DD}	positive supply voltage
V_{DS}	drain-source voltage
V_{SS}	negative supply voltage
V_{GS}	gate-source voltage
V_T	$=kT/q$, thermal voltage
V_{TH}	threshold voltage
W	MOS transistor channel width
μ	charge carrier mobility

Chapter 1

INTRODUCTION

The operational amplifier (OpAmp) concept was introduced by Tellegen [1] in 1954 under the name of "ideal amplifier". The first OpAmps with discrete transistors appeared in production in 1956. One of the first analog ICs was an OpAmp developed by R. Widlar in 1964. The operational amplifier is still the integrated circuit with highest production volume. Thousands of OpAmps have been designed, and new ones are being introduced by the semiconductor industry every month to fulfill the demands of innumerable applications and to take advantage of new technologies.

In an ideal world, where every analog application is unique, there would be a unique amplifier tailored for every particular application. In practice, this situation is impossible for obvious reasons, so every industrial OpAmp, being targeted on the specific application area, is overdesigned in an attempt to broaden the customer base and outpace the competition. Usually these additional features do not visibly change the production costs. This is true for stand-alone OpAmps as well for those used for systems on a chip, where overdesign is done in order to save on the development efforts and to reuse the same layout in different parts of the system.

This means that, in addition to the accomplishment of the main parameters like speed, power consumption and accuracy, the OpAmp designer should also implement a number of specific functions. Examples of such specific functions are numerous. For the DAC buffer that has to settle during a clock period, it may be a specific slewing behavior. For the audio or DSL amplifier, it may be the requirement to minimize the signal distortions. For a video amplifier that is used as a multiplexer, it may be a fast start up and shutdown with high output impedance. For the buffer of a switched-capacitor ADC with the rail-to-rail input range it may be the zero saturation voltage to the negative rail and high load capacitor tolerance; in addition, it may be the requirement of minimal delay when coming out of saturation (small overload recovery time). While the realization of the main OpAmp

parameters and the involved trade-offs are widely considered in the literature, the design of these specific features is left to the ingenuity and experience of the designer.

Analog design may be approached in two different ways. The first way, well represented in the literature and which is more common in academia, is a detailed consideration and overanalysis of some relatively simple circuits obtained in the "divine enlightenment". The second, more common to industry designers and less covered by the literature, as these people rarely publish their results, is a cookbook methodology where the past solutions are being reused with the smallest possible modifications.

This book is about analog design methodology, which we call the structural design. This methodology may be applied to design of any analog circuit or system, yet the OpAmp design process including the development of special requirements may be a good example of the methodology application. The top down design process of the OpAmp is considered, starting from selection of the gain structure (the explanation of this term is given below). The implementation of this gain structure as well as gradual addition of various specific functions in CMOS/BiCMOS/bipolar processes is considered. Most of the circuits shown in this book are new and have been used in recent industrial ICs. Yet it is not another cookbook with analog circuit recipes. The goal of this work is to arm the reader with a tool helping to invent the solution for any analog design problem and, at the same time, be reasonably sure that this solution is one of the best possible for any given process and set of constraints.

The theoretical basis for structural design was created in 1960-s and 70-s in the former USSR and applied to design various weapon control systems. Nearly all publications on the topic were classified and normally were not available in the literature, even in Russian. The adapted translation of the article *Structural properties and linear transformations in multidimensional systems with symmetric links* written by Valery Ivanov is given in Appendix to this book.

The authors have adapted this theory to analog IC design. The efficiency of this approach has been verified developing numerous standard and application-specific ICs: OpAmps, instrumentation and power amplifiers, voltage references, translinear circuits, low-dropout voltage regulators, switching DC-DC converters, microwave and ultrasound sensor interfaces, electric motor controls, and signal processing circuits of inductive angle sensors, accelerometers and other devices. The structural design methodology, in the author's opinion, matches intuitive approach used by the best analog designers [11], and corresponds to the modern philosophy called *systems thinking* [12].

1.1 Organization of the book

The book structure is shown in fig. 1-1.

Chapter 2 describes the main steps and tools of the structural design methodology:

- using the signal graphs for the system synthesis and analysis at the level of abstraction preceding the circuit diagram;
- feedback as a universal tool to control any parameter which needs to be regulated or stabilized;
- feedback loop interaction, frequency stability and compensation in the multiloop systems;
- voltage and current sensors and amplifiers for the local feedback loops;
- circuit design and optimization steps;
- quick preliminary selection of the circuits that are worth of detailed consideration.

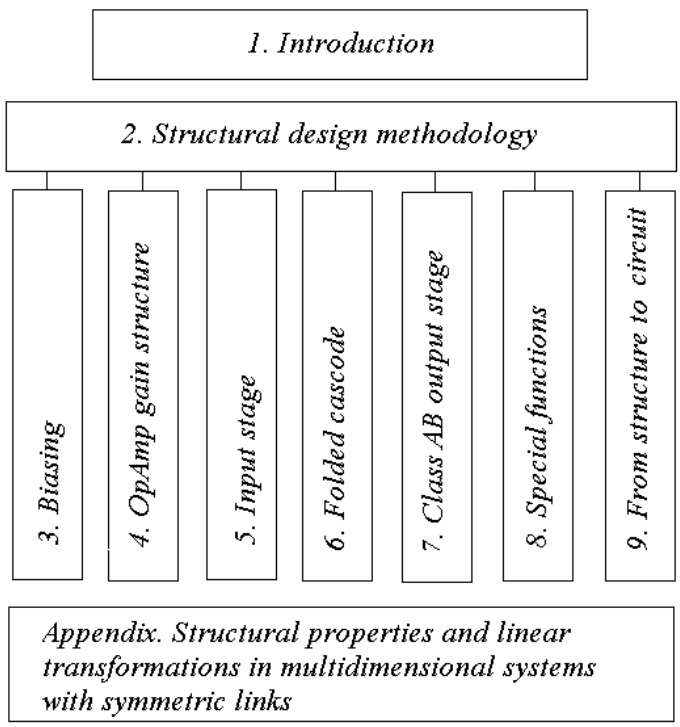


Figure 1-1. Book organization

The OpAmp designers often leave the development of biasing voltage and current sources for the finishing steps of the work. As a result, some surprises caused by the real components stay unnoticed until the silicon is

back and it is too late to make changes and there is no time to investigate and fix the parameter degradation of the OpAmp with real biasing. The design of reliable circuit foundation: PTAT current sources and other biasing cores, high-impedance current sources with low saturation voltage and low-noise charge pumps are discussed in chapter 3. These circuits are relatively simple and provide good examples of the structural design approach.

The general view of the OpAmp structure, namely the number of gain stages and meaning of the gain stage are presented in chapter 4. How many gain stages are really needed? What are the complications of multistage structure and when it is reasonable to use it? Which compensation scheme is preferable? How improvement of the voltage gain in a single stage can be done? What is the difference between bipolar and CMOS OpAmp structure? How the amplifier is made tolerant to the load capacitance? All these questions are discussed in this chapter.

Chapter 5 considers some options and trade-offs in the design of the input stage: realization of the rail-to-rail operation, the improvement of common-mode voltage rejection, and the influence of the flicker and high-frequency noise. New techniques for offset and temperature drift trimming during the final test (after packaging) are included in this chapter.

Chapter 6 is devoted to the OpAmp intermediate stages: folded cascode stage with floating current source, gain boost techniques, and discusses the role of voltage followers.

Class AB operation and output stage are considered in chapter 7. We are showing the general structure of class AB stage, its modifications and the resulting trade-offs. Possible stability problems, as well as the current gain erosion in CMOS amplifiers due to the impact ionization are also discussed in this chapter. In addition, we are discussing here how to achieve stability of the stage quiescent current with the supply voltage variations.

Realization of various special functions and the corresponding circuits for different applications are described in chapter 8. The following problems are covered (and the solutions are proposed): slew boost circuits, fast start-up and shut down, overload recovery time improvement, load current limiting circuits.

Chapter 9 considers top-down practical OpAmp design example. Some common sense points on choosing the transistor sizes and the operating points are clarified as well.

One theoretical article has been added to the book as Appendix. The article, “Structural properties and linear transformations in multidimensional systems with symmetric links” considers the equivalent structure transformations in the special class of multidimensional linear system important for electronic circuit design. These systems may be reduced to the universal differential structure with common-mode feedback, and the properties of this reduced structure are analyzed. Knowledge of these

properties allows the designer to predict the main features of a complex circuit from the single glance after referencing it to the basic underlying structure. The situation here becomes similar to the case of a single feedback loop system: many properties of such a system and its behavior can be established after brief analysis of the circuit.

1.2 Analog design steps and tools

The analog design process can be divided in the following iterative steps (fig. 1-2): marketing input, conception/structure, circuit development, design verification, layout and silicon verification. Often these steps are interacting with the automated test and design of trimming elements, and are followed by silicon debugging if necessary.

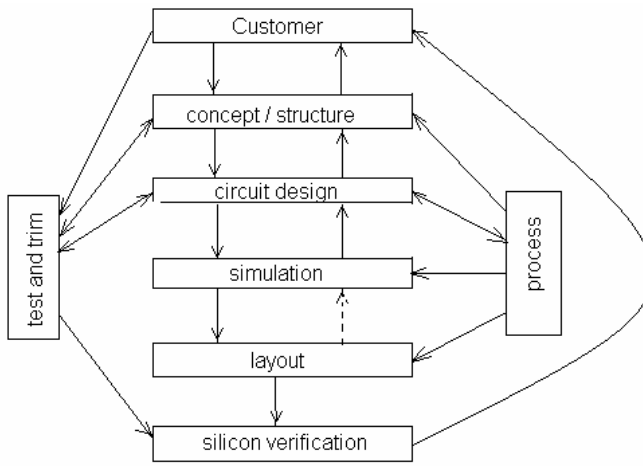


Figure 1-2. OpAmp development steps

The designer gets little help from CAD tools during the first two steps. Yet he/she relies heavily on the CAD simulation software and models during the design verification. Then the designer participates during layout step and silicon verification. Analog designers still need a soldering iron, but the romantic days in Bob's Pease style are practically over, and the solder is heated only if some mistakes have occurred.

At the circuit structure and stage topology development it is most important to develop the "feeling", or intuitive understanding of the circuit operation and its main features. Such feeling allows the designer to choose the direction for improvement of the structure or the circuit, and to evaluate

multiple choices in an acceptable time. There can be no innovation or a significant advance without such a feeling.

At the structural level this feeling comes from the preceding analysis of the operation of the general structures, such as one-dimensional system with a single feedback loop, and from the analysis of the general differential structure with common-mode feedback (Appendix) for the multi-dimensional systems.

At the circuit topology design level this feeling is based on understanding of the operation of components and elementary cells like differential stage or current mirror. How deep should we study and keep in memory the operation of these components and cells?

One of the quantifiable definitions of human intelligence is the number of factors that are taken into account while making a decision or choice. Regardless of how intelligent the designer is, this number is always limited. The CMOS and bipolar models described in the literature [9, 13, 64] are grossly excessive for understanding of the circuit function and for establishing an intuitive feeling. A list of most essential dependencies for CMOS and bipolar transistor parameters, compiled on the basis of the author's experience, is discussed in the subchapter 1.5.

Simulation can only verify that the circuit is operating in accordance with the feeling and expose the hidden problems. This verification must be as accurate as possible, so the most precise and detailed models have to be used on this step. The authors have not seen any dramatic difference between various simulators, but the accuracy of the models used is very important (try to get something better than free models supplied by foundries!). The least accurate are modeling of the saturation behavior of bipolar transistors and modeling of CMOS devices in the weak inversion. Sometimes it is even better to design the circuit in such a way that the components do not operate in the badly modeled region, for example, by adding clamps to prevent the saturation of bipolar transistors.

More detailed models do take more computation power. Analog circuits (or the analog units of mixed-signal systems) usually do not contain more than a few hundred transistors. The modern workstation has sufficient memory and speed for any kind of analysis in simulating the circuits of this size, and the computing time is not a problem. The only exception may be the simulation of transients in the switching circuits such as chopping or auto-zero OpAmps. The special software for fast simulation of such circuits has been developed. Simplified macromodels of the components can also be used, but using them does sacrifice accuracy, and the final simulations need to be done in the traditional way.

The inputs for layout design include the approximate floor plan, matching requirements for the critical components and the minimum width for the high-current buses. The hand-made layout of the signal path wiring

can make a difference for very high-speed amplifiers (>500 MHz bandwidth). The schematic for layout should be prepared in such a way so the automated wiring tools and the layout/schematic comparison software will find the ground or supply bus loops and avoid them. It is also necessary to avoid the supply modulation by the currents of load and digital units, etc. This can be done by drafting separate wires and subcircuit nodes for the large and small currents, for power, signal and digital grounds and substrate connections while joining these nodes only on the top-level schematic.

Use of the parasitic extraction tools and simulation with the parasitics introduced by layout can sometimes reveal problems with a circuit. However, only a circuit with design flaws can be sensitive to parasitics to the extent that they noticeably change the circuit behavior. If, after the parasitics extraction, the circuit behaves differently then a detailed investigation should be done to find design mistakes. That may cause very significant changes to the circuit, even on conceptual level.

The on-chip heat distribution should be taken into account when placing output stage relatively to the input and biasing core components. Bonding of the chip to the package is usually not even, and this causes a non-uniform heat transfer, thus, the thermal distribution models are not accurate. As a result, simulations of the temperature distribution do not provide much of an improvement in comparison to the placement in accordance with a common sense.

The mechanical stress of packaging significantly shifts the offset and changes the component matching [3], especially in processes without surface planarization. These effects can be reduced several times by coating the die with soft polyamide or another similar layer. The components that define accuracy should be laid out in a way to compensate for the mechanical stress: in quads, be interdigitated, be placed close to the chip center or the axis of minimal stress.

The common problem of substrate noise has gotten more attention lately [2], especially in the analysis and optimization of the systems on chip. Very little can be done during the OpAmp layout design stage to reduce the substrate noise leakage. To be efficient, the space separation between noise-emitting and noise sensitive units should exceed the die thickness (150 um or more), so the required isolation areas may exceed the total area occupied by the OpAmp in CMOS submicron process. If the problem of substrate noise is anticipated, then the appropriate process has to be chosen (dielectric isolation processes, processes with high-ohmic epitaxial layer and low-ohmic substrate, processes with both isolated N and P wells). Some special system solutions (time separation between noise-emitting and noise-sensitive unit operation, synchronization) should be used in combination with the proper circuit techniques like differential signal processing, utilization of the components insensitive to the substrate noise, etc.

The final touches to layout include shielding of the input wires by the surrounding grounded metal. For the high-speed and micropower amplifiers, shielding of the output bus and other fast-switching nodes helps to avoid the problem of noise injection into the substrate or underlying components.

The least number of mistakes occur when those are anticipated. The important nodes for easy debugging may be fitted with small pads (around 5x5 um^2) on the top-layer metal so they nodes can be easily probed.

Design for testability became the common methodology in digital circuits. The OpAmps may also include a circuitry for switching between different test modes (for example, connection of the internal current to an external pin during some operations of the final test), and a circuitry for recognition of the control signals during package-level trim through the existing external pins, etc.

Participation of the designer in the silicon verification and test varies from company to company. Usually, the test engineer and technicians design the test procedure, and the designer's involvement is necessary if the silicon is not operating as expected.

1.3 Modern analog processes

Just a few years ago the situation in the analog process development was grim: the dedicated analog companies (ADI, LTC, Maxim, Burr-Brown, etc.) were too small to afford the new sophisticated fine line equipment. The foundries like TSMC or UMC were providing for the analog designers only slightly modified digital processes from their older depreciated production lines, and the giants like Intel or Texas Instruments did not consider analog circuits as a prime business area. One of the very few instances carrying an active development of the analog processes was design of the BCD family for the power management in Castelletto division of ST Microelectronics [4].

The recent economy slowdown has been done wonders for the analog process development in big companies, where it got much more attention and investments. It is the area that was least affected by the economy slump, and the importance of analog design is now well recognized by the high management.

New analog processes that emerged recently from the evaluation stage to the industry mainstream could be divided in 3 groups:

1. Processes for high-speed wireless applications and internet hubs comprising heterojunction and SiGe components and exotic chemistry [5];

2. Processes for power management providing high-current, high-voltage MOS components, thick copper wiring layer on the top, fine-feature CMOS with double wells both for analog signal processing and for massive amounts

of logic, including the processes with dielectric isolation to avoid the substrate noise problems [6, 7];

3. Processes for precision analog design, comprising accurate, low-TC thin-film resistors, capacitors with no measurable voltage coefficient, wide spectrum of CMOS transistors rated from 3 to 40V, with low V_{TH} and low noise properties, and bipolar transistors including vertical PNPs [8]. The minimum CMOS channel length in this group is larger than 0.25-0.35 um because the short channel does not bring significant advantages to the precision analog circuits. The equipment used for manufacturing is capable for much finer geometries allowing one to improve the component matching.

New OpAmps are being developed using all these groups. Some new OpAmps were also being developed as a part of the systems on chip, in the digital processes. The newest digital processes usually implement, in addition to the fast thin-oxide 90 nm (or even 60 nm) channel transistors, also devices with longer channel and thicker oxide. These devices are actually used for analog units to avoid leakage, charge tunneling and other problems of the short-channel devices.

Practically all new processes for analog design are modular, with the capabilities to add new features and components, like additional metal layers or flash memory. This flexibility involves a few additional process steps and masks. The base mask count can be as high as 20, but the die cost is still relatively low thanks to using the depreciated equipment and smaller die size. Usually, for new OpAmps the cost of test and packaging exceeds the silicon cost.

New processes are creating the base for arrival of the new analog IC generation. This generation will be faster, more accurate and smarter, but it will also require a new kind of analog designer capable of using the new process capabilities not only for the incremental parameter improvement, but also for the quantum leaps ahead in all areas of analog design, including OpAmp circuits.

1.4 Trends and requirements of the OpAmp design

To be successful on the market, the new OpAmp should not only be the best in industry by at least one parameter important for the targeted applications, but also conform to the industry standards and customer expectations by the rest of parameters.

Industry standards include:

- reliability performance, proved by the various tests (life test consisting of at least 168 hours at 150^{o}C, set of the mechanical, humidity, etc. tests);
- electrostatic discharge protection (ESD protection components can consume the lion's share of the die area in CMOS OpAmps);
- package and pin assignment;

- operating temperature range (commercial 0 to 70°C, industrial –40 to 125°C, military -55 to 125°C).

Customer expectations include:

- absence of oscillations while operating with any gain (the OpAmps which are not stable with unity gain are almost extinct),
- tolerance to at least 100 pF of the load capacitance and to a small or nonlinear load resistance,
- at least 90 dB of the open-loop gain;
- rail-to-rail output stage,
- absence of strange effects like output inversion when the input signal moves out of the specified range, I_q spike during output saturation, etc.
- wide range of the supply voltages.

The rail-to-rail output stage initially became a feature of the low-voltage CMOS OpAmps in order to improve the output operating range but developed into the customer expectation even for the high-voltage amplifiers as it simplifies application designs. The rail-to-rail input stage capability is also developing in a customer expectation to any OpAmp by the same reason.

The supply voltage range should exceed the supply limits of the targeted applications. This is a quantum scale, and there is almost no customer base difference, say, between 6 and 10 V of the maximum voltage supply for OpAmps (or between 1.9 and 2.2 V on the low side). On the high side, these voltage steps are: 3.7 V (3.3 V + 10% digital supply); 5.5 V (5V +10%); 12 V; 22V; 27V; 33 V. On the low side it is 1.8 V (two half dead alkaline cells), 2.2 V (two NiCd or NiMH cells), 2.5 V (Li ion battery), 2.7 V (3 V – 10%), 4.5 V (5 V – 10%).

The expectations to the high-frequency amplifiers are much lower due to the luck of competition, until recent times, and to the much better analog proficiency of the high-speed OpAmp users (it is easier to explain to these people why one can not implement certain features, and how to use the OpAmp as is in the application, than to develop the all-good OpAmp).

The compliance with some standards and expectations can significantly increase die area and take a large share of the development time, especially for such features as ESD protection of CMOS OpAmps or prevention of the output phase inversion when the input is out of specified range for the bipolar OpAmps.

In developing new OpAmps the main targets are the improvement of:

- speed to power ratio;
- accuracy to cost ratio;
- speed to cost ratio.

Constant upgrading of the processes and new circuit techniques complement each other in this development. The circuit enhancements include:

- reducing the number of gain stages while preserving the high gain [14];
- new package-level offset and temperature drift trimming techniques [15];
- auto-zero and chopping offset elimination circuits [16].

The OpAmps breaking speed limits in SiGe processes are being designed for the signal processing in wireless and video applications. In other areas creation of the champions in speed or accuracy regardless of cost happens rarely, as the market for the record-breaking part usually does not yet exist, and its financial success is questionable.

The same may be said about the very low supply voltage (<1.8V) designs. In spite of much attention in the recent literature [11] there is still almost no market demand for these OpAmps.

1.5 Essential parameters of bipolar and MOS transistors.

The Gummel-Poon model of the bipolar transistor lists 45 parameters [112] and still is not accurate enough to simulate the saturation behavior or junction breakdown region. The BSIM 3.3 model of the MOS transistor has more than 50 coefficients [113] not counting noise and gate leakage parameters. All these variables are useful for the circuit simulations. However, only very few numbers and equations have to be remembered for creative work, and the shapes of few dependencies and some qualitative relationships (not how much, but more/less, grow/decrease) are much more important.

The list of essential curves and parameters varies with application and personal opinion. This subchapter mentions only the most important, from the author's point of view, component parameters that are necessary for the circuit topology design. Not surprisingly, it also matches the list of questions usually asked during interview with a prospect designer by the experienced engineers in different companies. A more detailed description of the component operation can be found in textbooks on analog design (for example, chapter 1 of [9] or [13]) or in dedicated to modeling books [64, 114].

1.5.1 Bipolar transistor

Important DC values

Fig. 1-3 shows the typical dependence I_C (V_{CE}), V_{BE} = *const*. It can be divided in three regions: saturation, active region and breakdown. The breakdown region potentially can be used to design a voltage reference; otherwise, the transistor operation both in the breakdown and in saturation regions should be avoided.

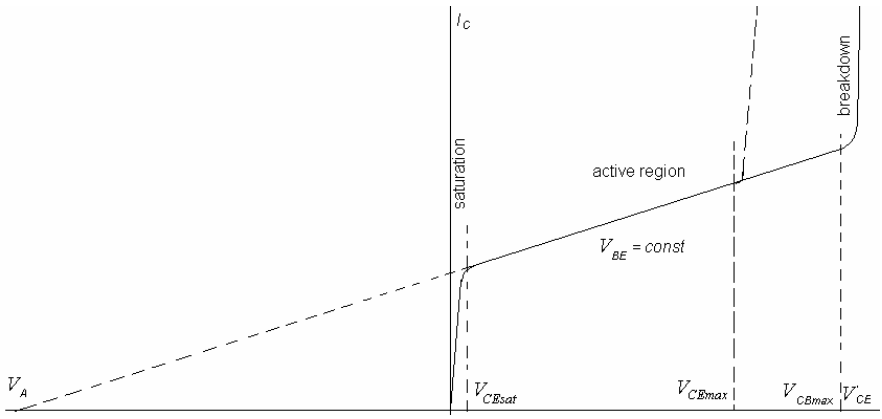


Figure 1-3. Transistor output characteristic $I_C(V_{CE})$

The breakdown region is not included usually in the SPICE models. If the operating voltage can reach this region then a Zener diode has to be added to the model for more realistic simulations.

The dependence shown in fig.1-3 assumes a zero base voltage source resistance. The bipolar transistor base current depends on V_{CE}. With increase of V_{CE} the base current may even become negative, effectively moving the breakdown region left if the signal source has non-zero resistance.

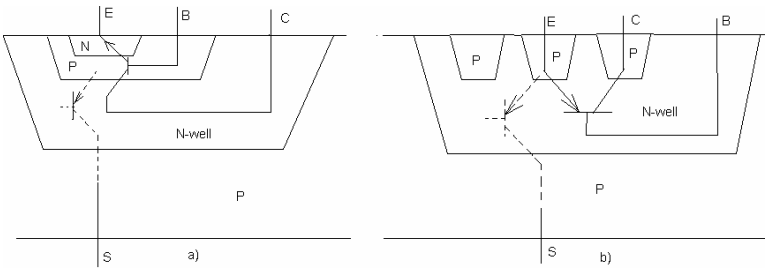


Figure 1-4. Parasitic structures of integrated bipolar transistors

The vertical NPN transistor includes a substrate PNP transistor (fig. 1-4a) that can become active in saturation. This parasitic transistor limits the minimum V_{CE} of NPN device to about 100 mV, regardless of how much current is being pumped in the base.

The lateral PNP transistor (fig. 1-4b) includes a vertical parasitic transistor causing the substrate leakage of the emitter current. The value of this leakage current can be less than 1% of I_E if the process has a highly doped buried layer, and up to 50% or more if there is no buried layer and the

well thickness is comparable with the base width. The saturation voltage of the lateral PNP device may be much lower (down to 10 mV).

The breakdown of the emitter-base junction in reverse biasing can occur at a much lower voltage than that of the collector-base junction, and an adequate input protection circuit may be necessary.

Operating point parameters

The current gain, $\beta = I_C/I_B$ is usually in the range of 50-200 for normal temperature but it can be as low as 20 or as high as 5000 (for transistors with very thin base). The gain, β, may increase with temperature. It also gradually increases with the collector current, reaching the peak value at the collector current equal to the parameter *IKF* of the Gummel-Poon model and sharply drops at larger currents. The gain β is the least matching and the least stable of transistor parameters (10% mismatch is possible even for the closely located differential pair).

The gain, β, increases with V_{CE} as the leakage current of the collector-base junction subtracts from the base current I_B.

The Early voltage, V_A, (fig. 1-3) is the parameter defining the small-signal collector resistance $r_{ce} = V_A / I_C$.

Important temperature and stress dependencies

The base – emitter voltage, V_{BE}, (fig. 1-5) is in the range of 0.5 – 0.6 V for the silicon diffused transistors with realistic currents and normal temperature, and 0.7-0.8 V for the high frequency transistors with poly emitters.

For variations of V_{BE} with the current I_C and the emitter area, S_e, it is important to remember (fig. 1-5) that $V_{BE} = V_T \ln(I_C / I_S)$. Here $I_S \approx \alpha T^4 S_e$, called the reverse current of the p-n junction, is proportional to the emitter area and is heavily depending on temperature. The logarithmic dependence between collector current and voltage is accurate with current variation up to 100 dB (five decades of current change).

The thermal voltage V_T is proportional to the kT/q. It can vary from approximately 22 to 30 mV at the 27°C temperature for the different processes. V_T changes very linearly with the temperature.

V_{BE} has a slightly nonlinear negative temperature coefficient of approximately -2 mV/K. The bow-like nonlinear part of $V_{BE}(T)$ dependence from temperature is called curvature, and is the main source of error in the bandgap voltage references. V_{BE} also varies with mechanical stress to which the PNP transistors are about 3 times less sensitive [3].

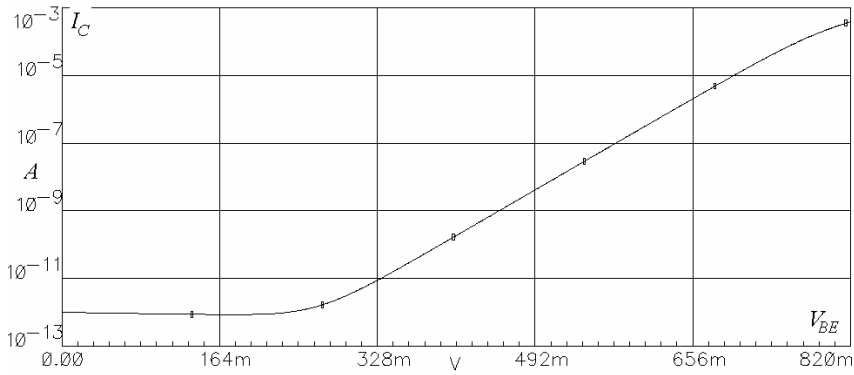


Figure 1-5. I_C (V_{BE}) on logarithmic scale

Small-signal parameters

The emitter and collector silicon body resistances (r_{ep} and r_{cp}) should be taken into account (for example, for the output device sizing) for transistors operating at high currents where the voltage drop across these resistors is noticeable.

The small-signal transconductance, $g_m = \delta I_C / \delta V_{be} = I_C/V_T$, defines the small-signal emitter resistance $r_e = (1/g_m + r_{ep})$, and the base resistance $r_b = (\beta * r_e + r_{bp})$, r_{bp} – silicon body resistance of the base.

Dynamic parameters

The small-signal model of bipolar transistor is shown in fig. 1-6. It includes:

capacitances C_{be}, C_{cb},

parasitic capacitor to the substrate C_{cs} between collector and substrate for the vertical devices or C_{bs} between base and substrate for the lateral ones, both are shown but only one will appear in the model used,

small-signal base-emitter resistance r_{be},

the silicon body resistance of the base r_{bp}

and small-signal collector – emitter resistance r_{ce} defined by V_A and I_C .

C_{cb}, C_{cs} and C_{bs} are decreasing when the voltage rise across the corresponding junction increases.

The capacitor $C_{be} \approx C_{jbe} + \tau_F g_m$ represents the capacitance of the p-n junction plus the delay caused by the transition time of the charge carriers through the base region. This current-dependent (via g_m) capacitor makes the frequency compensation of bipolar circuits much more challenging than the compensation of circuits in CMOS.

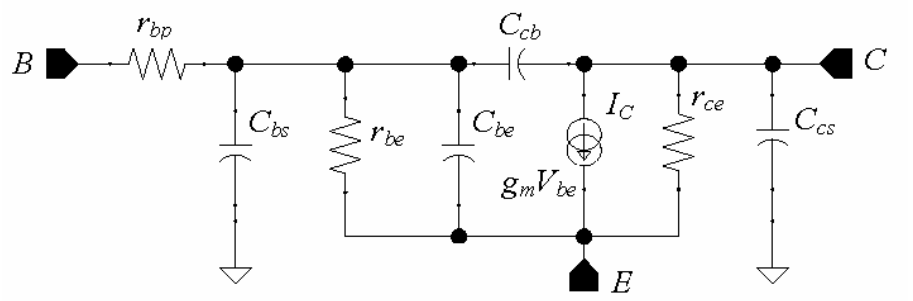


Figure 1-6. Small-signal model of the bipolar transistor

Noise and matching

The equivalent series noise resistor for the bipolar transistor is $r_{eqs} \approx r_e/2 + r_{bp}$ and the equivalent parallel noise resistor $r_{eqp} \approx 2\beta r_e$ [10, p. 85]. This means that the value of high-frequency noise is inversely proportional to the square root of emitter current with an additional term defined by the silicon body resistance of the base.

The low frequency flicker noise is process-dependent and is represented by the parameter *KF* of the Gummel-Poon model. This noise, to some extent, is inversely proportional to the square root of the transistor overall emitter area. The 1/*f* flicker noise corner for the bipolar transistor is usually located at 100-200 Hz.

The base-emitter voltage mismatch, V_{off}, between adjacent transistors of the same layout and orientation normally does not exceed 0.5 mV (3σ). The temperature drift of V_{off} is proportional to the absolute value of V_{off} itself at the normal temperature.

In modern fine-line processes matching of the lateral devices may be better than vertical ones because lateral dimensions are controlled in production better than vertical diffusions.

High-voltage and reliability effects

Large reverse voltages across p-n junction are causing the impact ionization and, with further increase, the avalanche breakdown. The avalanche itself does not cause the reliability problems. It is the heat generated during the breakdown that destroys the bipolar transistor when the current is not limited.

The impact ionization current of the collector-base junction flows through the base, subtracts from the base current and effectively increases β. At the voltage V_{CEmax} this impact ionization current gets larger than the base current at zero V_{CB} voltage, and the current through the base inverses its sign.

1.5.2 MOS transistor

The MOS transistor has a very complicated theoretical model. However, for the designer it is more manageable than the bipolar transistor model, especially if one uses only very essential parameters.

Important differences from bipolar transistor

The dependence $I_D(V_{DS})$ is similar to that of the bipolar transistor (fig. 1-7). The differences are the following:

- the saturation region ($V_{DS} > V_{GS} - V_{TH}$) for MOS transistor corresponds to the active region for bipolar one, and the saturation region for bipolar transistor corresponds to the triode region ($V_{DS} < V_{GS} - V_{TH}$ or $V_{DS} < 4V_T$ if $V_{GS} \leq V_{TH}$) for the MOS one;
- there is no definite Early voltage value for the MOS transistor as the equivalent parameter increases with the channel length;
- operation in the breakdown region is absolutely forbidden as the MOS transistor will be irreversibly altered or destroyed;
- there is no parasitic structure active in the triode region, and it is a very useful region of operation where transistor is used as a passive resistor.

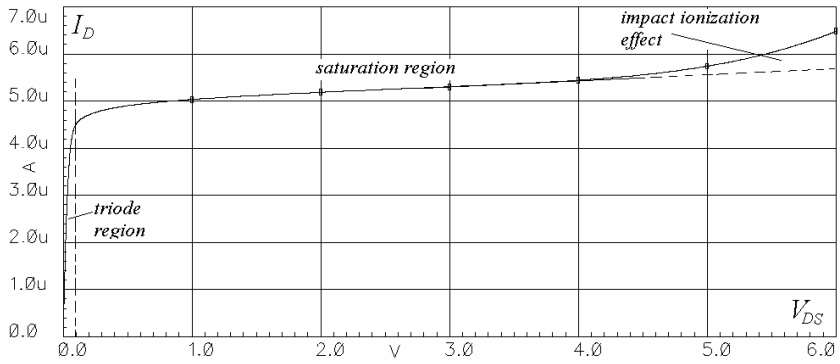


Figure 1-7. $I_D(V_{DS})$ dependence

DC characteristics and parameters

The dependence I_D (V_{GS}) in the saturation region is shown in fig. 1-8. It is divided into the triode ($V_{GS} > V_{TH}$), and weak inversion ($V_{GS} < V_{TH}$) or subthreshold, regions.

In strong inversion $I_D = \frac{K}{2}(\frac{W}{L})(V_{GS} - V_{TH})^2$ where $K = \mu C_{ox}$.

In weak inversion $I_D \approx K_w \exp(\frac{V_{TH} - V_{GS}}{nV_T})$ where K_W is a coefficient that is easier to estimate via simulations and experiments than to present analytically.

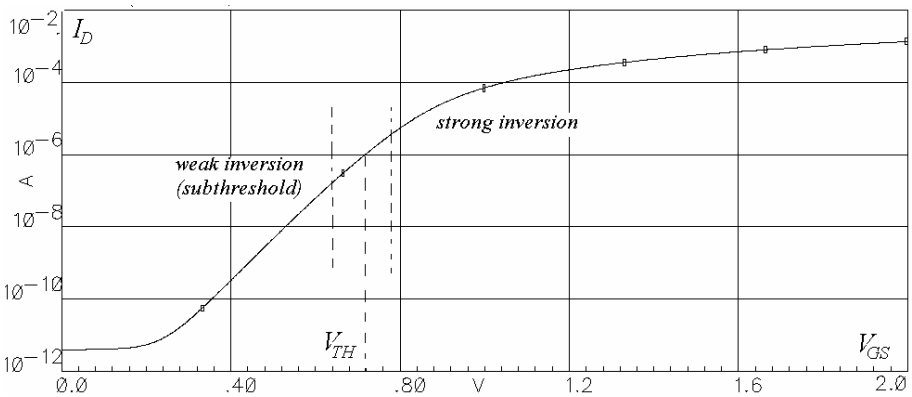


Figure 1-8. $I_D(V_{GS})$ on logarithmic scale

The transition between the triode and saturation regions is soft and is located at $V_{DS} \approx 4V_T + (V_{GS} - V_{TH})$ or approximately at $4V_T$ if $(V_{GS} - V_{TH}) < 0$.

The threshold voltage, V_{TH}, with zero body to source bias is a process-controlled parameter. Its value is set depending on the minimal channel length in the process. For example, it is usually set to 0.6-0.8 V for NMOS and 0.8-1.0 V for PMOS in the 0.6 um process. These values are smaller for fine-line processes. The transistors without V_{TH} adjustment (so-called natural transistors) sometimes are also available. V_{TH} of the natural NMOS transistor can be in the 0.1-0.3 V range, and V_{TH} of the natural PMOS can be in the 1.0-1.2 V range.

The body to source bias V_{BS} changes the value of V_{TH}. This change is approximately half of the body-source bias. The forward bias of the body to source should not exceed $V_{BS} < (V_{BE} - 4V_T)$ to prevent the current flow through the source to body p-n junction.

When accessible in the process, the body can be used as a second gate to control the transistor.

Temperature dependencies

For a constant drain current I_D the temperature dependence of V_{GS} is linear with a slight curvature. The temperature coefficient of V_{GS} is negative

when operating in weak or mild inversion, and gets positive in very strong inversion ($V_{GS} \geq 2\ V_{TH}$).

Small signal parameters

The MOS transistor in weak inversion behaves similar to the bipolar transistor, and the MOS transconductance, g_m, is approximately half of that for the bipolar transistor carrying the same current, i.e. $g_m \approx I_D / 2V_T$

In strong inversion $g_m = K(\frac{W}{L})(V_{GS} - V_{TH})$. For any given operating current g_m can be increased by increasing *W/L* ratio – but only while the transistor is operating in strong inversion. In weak inversion, g_m reaches its maximum value and no longer depends on size.

When the transistor is controlled by the body potential: $g_{mb} = K'(\frac{W}{L})(V_{GS} - V_{TH})$ where $K' \approx (0.1\text{-}0.3)K$.

The drain-source resistance in the triode region while in strong inversion is $1/R_{DS} \approx K(\frac{W}{L})(V_{GS} - V_{TH})$

The small-signal drain-source impedance can be represented as $r_{ds} = V_A/I_D$ where V_A is a function of the channel length.

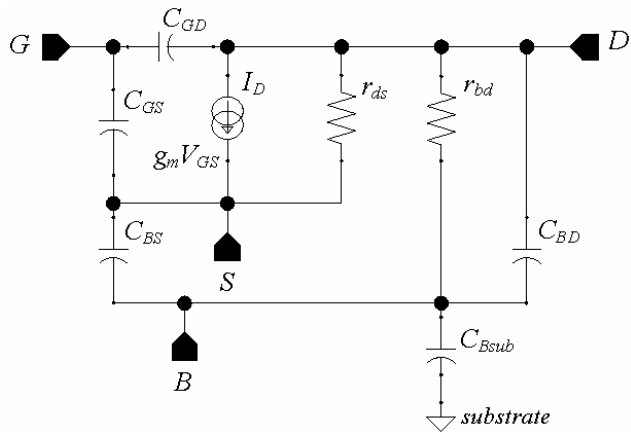


Figure 1-9. The dynamic model of MOS transistor

Dynamic parameters

The small-signal model of a MOS transistor is shown in fig. 1-9. It includes:

- oxide capacitances C_{GS}, C_{GD};

- parasitic junction capacitors to the body, C_{BD} between the drain and body and C_{BS} between the source and body, proportional to the source and drain areas plus sidewalls;
- parasitic junction capacitor between body and substrate (well capacitance) C_{Bsub}, which is larger than any other capacitor in the equivalent circuit;
- small-signal drain – source resistance r_{ds};
- resistor r_{bd} presenting impact ionization current between drain and body.

All junction capacitors decrease with increase of the voltage across the corresponding junction.

Noise and matching.

The equivalent series noise resistor for the MOS transistor is $r_{eqs} \approx 2/3g_{m.}$ The equivalent parallel noise resistor is very large [10, p. 85] and can be neglected. This means that the high-frequency noise of the MOS transistor in weak inversion is approximately twice as large as that for the bipolar transistor with the same current. This ratio increases with drain current when transistor is in strong inversion.

The low-frequency flicker noise is process-dependent and is usually 10-50 times larger than that for the bipolar transistor of the comparable size. This noise is inversely proportional to $\sqrt{WL}$. The 1*/f* flicker noise corner for the MOS transistors can also be located at the frequency of 100-200 Hz (i.e. comparable with that of bipolar transistors), but only for relatively large size devices ($WL > 1000\text{um}^2$).

The mismatch of V_{GS} voltages between two interdigitated transistors depends on their size. It can be expressed as $1\sigma(V_{ofst}) = V_{mtch} / \sqrt{WL}$ where W and L are in um and V_{mtch} is the empirical process-dependent constant. V_{mtch} can be as high as 50 mV for old processes, 20-25 mV for some foundry processes, and as small as 5 mV for the newer processes for precision analog design.

The spread of V_{GS} voltages does not depend on the current, so current matching is better for transistors operating in strong inversion because the same offset is related to larger values of V_{GS}.

The temperature drift and absolute value of V_{GS} mismatch do not correlate as to the corresponding values in bipolar transistors.

High-voltage and reliability effects.

Large voltages between the drain and source cause the impact ionization currents from the drain to body. This impact ionization current is noticeable even with 1-2 V of V_{DS}. Its effect can be seen in the right side of fig. 1-7 where it changes the slope of the $I_D(V_{DS})$ curve.

The same impact ionization with larger V_{DS} can cause the time drift of V_{TH} (hot electron effect). This V_{TH} change can be as large as few hundred millivolts. It depends on the exposure time and is irreversible.

Further increase in V_{DG} and/or V_{GS} voltage will cause the destruction of the gate oxide. The avalanche breakdown voltages of the drain-body, source-body, and body-substrate junctions are normally much larger than the oxide breakdown voltage, and the gate oxide gets ruined first.

Chapter 2

STRUCTURAL DESIGN METHODOLOGY

The common attitude towards analog design is that is more of a black magic than engineering. The existing textbooks on analog IC design consist of very detailed description of the component operation [9, pp. 1-202, 10, pp. 1-204] followed by examples. Lack of the circuit synthesis theory is supplanted with the detailed circuit analysis of relatively simple circuits. Such an approach is not able to produce innovative results.

The majority of industry designers reuse the working solutions from their past experience, with the necessary adjustments to a new process or to a different trade-off of the parameters. Those chosen few who do create new circuits consider them more of a result of "divine enlightenment" [11].

In the digital world the designer's active role ends at the functional description of the circuit in VHDL or another formal language while the software takes the rest of the work: (circuit synthesis, layout and test generation). In the analog world there are no formal rules for circuit synthesis. While nobody even thinks about patenting a digital circuit implementation, every new and useful analog circuit is considered valuable intellectual property. Transistor count in digital circuits has climbed to a billion, yet, not every analog circuit has more than a hundred active components. Such place of analog electronics is caused, among other reasons, by the difficulty in understanding the analog functionality. This includes the dynamic stability, noise, distortions, operation limits, multiple physics effects (thermal, optical, mechanical, etc.), and is aggravated by the individuality of the problems to be solved by analog circuits.

From the general point of view, the structural design is close to the philosophy of systems thinking [12, 17], heuristic mathematics [19] and methods of innovation theory [18]. It is a set of formal rules of how to approach the analog synthesis problem, and how to look for an acceptable solution. The application of this methodology generates many different circuits unless there is a fundamental contradiction in the problem to be

solved. Some of these circuits are known, and almost always some are new. Nearly all are useful, with the final selection based on the parameters of secondary importance.

The structural design methodology of analog integrated circuits consists of the following iterative steps:

- description of the system functionality at the high level of abstraction using the signal flow graph;
- equivalent transformations and modifications of the system graph to the forms where all important parameters are controlled by dedicated feedback loops;
- implementations of these graphs using the library of elementary cells;
- simplified, single-glance estimation of possible implementations and selection of the most attractive ones;
- consideration of the non-linear effects (saturation, short circuit, input overload) and adding to the circuit the necessary clamps and protection units;
- simulations of a small number of chosen implementations and selection of the final circuit.

The main tool to create the desired system property or characteristic (and this is the essence of the structural design) is the systematic usage of the dedicated and efficient feedback loops controlling each and every important parameter. Optimization of the component parameters can be used only if all other reserves for the structural and circuit topology improvement are exhausted, which, in practice, means never.

The structural design methodology can be described as a view point at the place where one can find a solution for the problem at hand, and where the more you look, the more you find, plus the tools for fast estimation of the solution shortcomings. As noted by Paul Brokaw, “good solution will appear obvious from the proper viewpoint” [11, p. 127].

2.1 Consider good circuits only

Thousands of different amplifiers can be designed using just 2 transistors: the combinations of NMOS/PMOS types, low/high input impedance, low/high output impedance, different types of feedback, etc. An average OpAmp has more than 100 transistors and the number of their combinations is larger than number of atoms in the galaxy. From one hand, one may be reasonably sure that, if it is possible at all, there exists the combination of components that is performing as desired. On the other hand, one need rules to select a set of possibly good circuits without their detailed consideration to avoid the waste of time analyzing bad circuit variants.

One may state that all good circuits look similar, and every bad has its specific problems, but we should not care about these problems as we can just move to a better circuit topology.

The basic features of a good circuit from the structural design point of view are the following:

1. All important parameters are controlled by the dedicated efficient feedback loops.
2. Each of these feedback loops is stable dynamically and behaves like a first- or a second-order system.
3. The circuit is robust to potential variations of the component parameters and operation modes.
4. All possible non-linear effects like saturations, shut-offs, etc. do not cause dramatic consequences such as conditional stability, shut down, phase inversion, etc.

The feedback loops controlling the main parameters of a unit (or a block) have to be efficient. The open-loop gains should be large enough that the variations of component values, external noise, etc. do not cause unacceptable errors, and the block transfer function is defined by the sensor and feedback network parameters. This efficiency can be improved using control theory methods: addition of the astatic units with transfer functions like $(1+T_0 s)^{-1}$ or units $(1+T_0^2 s^2)^{-1}$ having infinite gain at the resonance frequency $1/T_0$, positive feedbacks, parallel links, etc. Various gain boost methods that are specific for microelectronics (see chapters 4 and 6) can also be used.

Each feedback loop may have a stability problem and may need compensation. There are a multitude of compensation techniques available [22]. But with the structural approach, the system may be designed in a way that no capacitive compensation is required, by using the single-stage amplifiers as building cells for the local loops and feedforward links. Sometimes this design is not possible and regular compensation methods must be used. The stability problems and compensation of multiloop systems are considered in subchapter 2.3.

The robustness of the circuit versus component variations is the general requirement of IC design. This robustness is verified using the corner model libraries, Monte-Carlo methods, corner simulation tools, etc. For the designer it also means that any parametric optimization is worthless: if the optimum is sharp then the circuit is not usable, if it is dull then the choice of component value based on common sense is good enough.

The circuits discussed in the literature are usually simplified to illustrate the main idea. They rarely comprise the clamps or other units that are used to prevent saturation, shut down or other non-linear effects. But in any industrial OpAmp at least every third component is added to prevent these effects. For example, the die area of a CMOS OpAmp is often dominated by

electrostatic discharge protection components. Analysis of the circuit operation during start-up, with input overload or with shorted output consumes at least half of the design time. Some of these protection circuits are discussed in chapter 8.

With practice, the selection of the good/bad circuits starts to operate on the subconscious level. There are thousands of circuits that are used in industry which are not good from the structural design point of view. All of these "bad" circuits do have specific problems, which may or may not be important in their application. Most of these circuits are the result of very hard work, as it is very difficult to find a new and working solution, and also have a sentimental value to their authors. However, they can be easily improved problem is looked upon using the structural design approach.

2.2 System description and analysis with signal flow graphs

Any design starts with the clear conception of the system goals and the knowledge of the component base (process) at hand. The system description includes the list of input and output variables, the functional connections between them, input and output limitations, disturbances (noise, mismatch, etc.), and acceptable static and dynamic errors. It can be presented in the forms of algebraic and differential equations, text descriptions, graphs, tables, and macromodels. In the OpAmp design this is often formulated by a similarity criterion (make an amplifier similar to another but with certain improvements). Some requirements may not be present in the initial description, and the familiarity with the applications is essential for the designer. There are always a number of trade-offs and the possibility of additional functional features to add to product. Familiarity with the application helps to acquire an understanding of what is important for customers.

The next step is theoretical or, better yet, experimental confirmation that the design goals are achievable and do not contradict each other.

Examples of the trade-offs are the power to noise ratio or the output power to the power consumption ratio (always < 1). The discovery of the theoretical limits can be a very creative process. A challenging and feasible functional description may determine an innovative solution.

The graphical presentation of the functional description is easiest for the mind to grasp. There are two equivalent methods of the graphical presentation: block diagrams and signal flow graphs. Block diagrams are used very often. However, the signal flow graph may be preferable, as it is faster to draw, and the formal rules for equivalent graph transformations have been developed [25]. Long time ago the signal flow graphs used to be a common tool for electric circuits analysis and, sometimes, for synthesis as

well [93, 94]. Fortunately, these applications of graphs get more attention again [23, 24, 28].

The difference between a structure (or a graph) and a circuit reflects the difference in the level of abstraction. The components of a structure are macromodels, transfer functions, links and nodes of the signal flow graph or the block diagram. The components of a circuit are real transistors, resistors, capacitors, etc. or their symbols in the circuit diagram. The transition from the structure to the circuit is not unique, and there is always a set of circuits implementing the structure.

Fig. 2-1 gives a simple OpAmp circuit diagram (fig. 2-1a) and its signal graphs with different degrees of detailing.

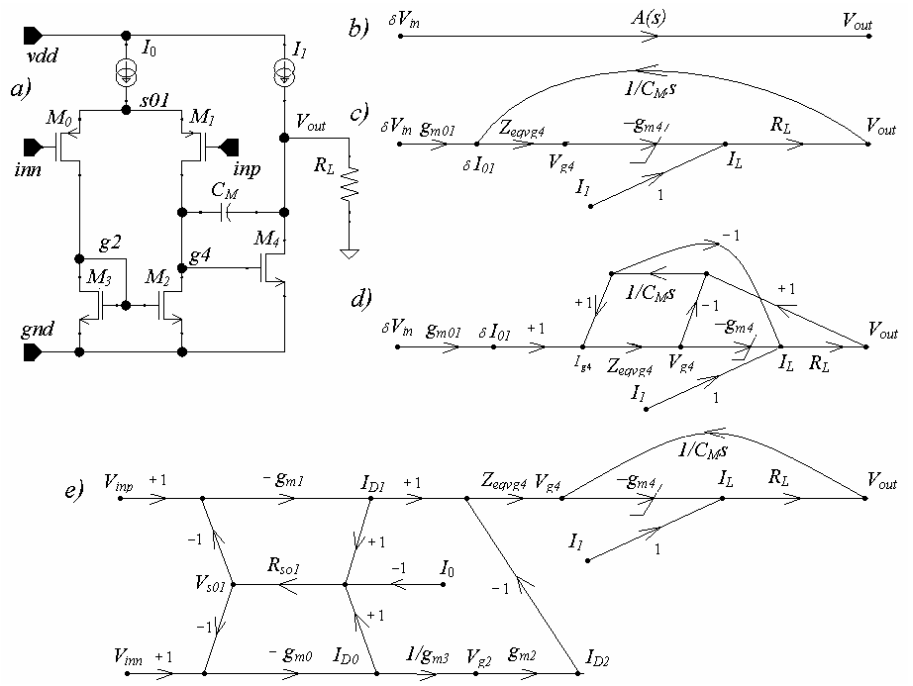


Figure 2-1. OpAmp functional presentation with different signal flow graphs

Fig. 2-1b represents the functional definition of the OpAmp as a unit with the transfer function $A(s)$. Any OpAmp can be described this way.

Fig. 2-1c describes this OpAmp as a two-stage circuit with feedback path via Miller compensation capacitor C_M. This description neglects the signal feed forward via this compensation capacitor.

The effects of the signal feedforward through C_M are included in the graph of Fig. 2-1d. This feedforward causes the faster phase degradation at the high frequencies in comparison with other compensation methods.

Canceling of this effect can be done, as it is follows from the graph, by breaking the feed forward link, for example, by adding a follower in series with Miller capacitor. Such break of an undesirable link is one of the steps in the structural design.

This last graph (fig. 2-1e) represents in more detail the operation of input differential stage and the current mirror. The left part of this graph is what is called in the structural design "the general structure with common-mode feedback". This graph may represent a rather general multidimensional system as discussed in the Appendix.

Using the graph of fig. 2-1c, one can see that the amplifier open-loop gain is equal to $A = g_{m01} R_{eqvg4} g_{m4} Z_L$. The transconductances g_{m01} of the input differential pair and g_{m4} of the output device M_4 are limited by the corresponding values of current. There is no control over the load, and the only viable way left for designer to increase the open-loop gain is to increase the equivalent resistance R_{eqvg4} at the node *g4* (fig. 2-1a). This resistance is defined by the internal feedbacks in the transistors M_1 and M_2. The graph detailing these feedbacks and the ways to break these links and to boost the gain are discussed in chapter 4.

It is important to realize that the application of traditional graph theory rules to the transformations of system graphs is limited by the linear systems. The real systems with limitations can be presented by the signal flow graphs with limitation links. The inverted L-shape line is chosen here to denote these links. An example representing this limiting link for g_{m4} is shown in fig. 2-1b. A word of caution should be said about these systems. Smooth nonlinearities may be represented by the links with transfer functions that can vary in some range. Using the piece-wise approximation and choosing different combinations of the link transfer values one can represent a nonlinear system by a set of graphs. One can then apply the traditional graph theory to each graph of this set. Yet, the graph transformations of these partial graphs will not necessary lead to valid realization of the system or its blocks (units). Hence, the system that includes nonlinear links needs more attention, as the rules of traditional theory being applicable to each graph of the set may be not applicable to the whole system.

Every connection of two components or units may create a new feedback loop. There are two practical rules for electronic circuits that should be obeyed to eliminate unnecessary interactions:

- do not load a current source (high output impedance) on the high-impedance load;
- do not load a voltage source (low output impedance) on the low-impedance load.

The system design starts using ideal current and voltage sources. Every time when the ideal component is replaced with a real one the effects of the

new "natural" feedback loop should be anticipated. If the exclusion of nonideality is essential for the system operation then these components should be "idealized" in the structural design way, namely, by adding the efficient feedback loop consisting of the sensor, amplifier and actuator.

The description of connections of passive R, C and active components as parts of feedback loops in a signal flow graph allows the designer abstracting from the nature of energy in these components. This means that one can use the design solutions found on the structural level, regardless of the physical nature of the component base. This similarity of the models describing different kinds of energy has been noticed long ago [33]. It can be shown that this similarity is a consequence of the energy preservation law. On the structural level there is no difference between design of the bipolar or CMOS OpAmp, of the cruise missile control or of the corporation management structure.

2.3 Frequency stability in the multiloop system

Every feedback loop may become unstable. Good systems are not only stable: they are robust to the parameter and condition variations. This means that they should have the excess stability.

The dynamic properties of a system include:

- excess stability of the system in all possible conditions (including stability of the periodic oscillations if this is the system operating mode);
- low sensitivity (or unresponsiveness) of the dynamic parameters to the variation of component parameters; it can be used as a merit of the excess stability;
- dynamic errors, settling time, settling behavior.

Excess stability and low sensitivity in the system should not depend on which nodes are the inputs and which are the outputs.

Every real system is non-linear. Again, it is convenient to represent such a system as a set of linear structures with a different structure for each operating point. For the overall system stability it is sufficient (but not necessary) that each of the structures in the set is stable [92].

The practice shows that the transient processes in the systems with excess stability always look like the transient processes of a second-order system. Most often estimation of the linear system dynamics is done using Bode plots as graphic presentation with phase margin as a merit of excess stability. But the conditions of the excess stability and unresponsiveness of the dynamic parameters can also be expressed by the overshoot function M_m of the closed loop step transient response [20].

All systems with the same overshoot factor M_m can be called dynamically equivalent. The advantages of using M_m to estimate the excess stability are

hidden in the fact that it is easy to observe M_m by simulation of the transient process, and that overshoot factor M_m has a narrow range for good systems (1.1 to 1.5). Also, for non-minimal phase systems the phase margin as a merit of excess stability is overly pessimistic (for example, an low dropout regulator with a capacitive load can have a normally unacceptable 5^o phase margin but still demonstrate a good step response with M_m approximately equal to 1.5).

The sufficient condition of the frequency stability in the multiloop system requires that each and every feedback loop within this system be stable [92, chapter 8]. Using this condition is aggravated by the fact that the connection of one loop to another loop (or even a component) can change the characteristics of both loops. The best example, probably, is the connection of the capacitive load to the OpAmp that can make it unstable.

Connection of two different loops is the most frequent case in design of electronic systems. This connection can take place in two variants:

- loops have a common node (fig. 2-2a)
- loops have a common link (fig. 2-2b).

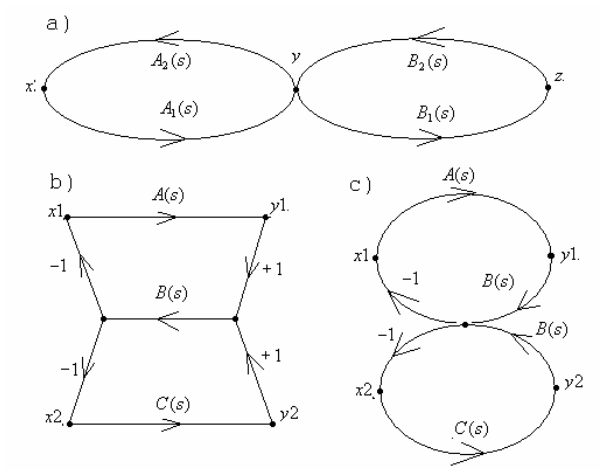


Figure 2-2. Connections of feedback loops in two-dimensional system

The loop connection of fig. 2-2b type can be transformed to the form of fig. 2-2c which is equivalent to fig. 2-2a.

When two loops have a common node then one of them is the load or signal/noise source to another and vice versa. If the equivalent impedance of one loop at the connection node is small (so that this loop is a voltage source), and is loaded by a relatively large or comparable equivalent impedance of the second loop then, providing that both loops are "good" (have excess stability and robust to the small parameter variation), such

connection will not create any stability problems. In the opposite situation such connection can significantly change parameters in at least one of the loops. If this can cause the stability problem, then such connection should be avoided by a structural change (for example, by excluding the node loading with the follower), or by the corrected ratio of the equivalent impedances (for instance, by an additional or diminished gain in one of the loops).

The compensation problems can be avoided altogether if one decides to use, for the local feedback loops, single-stage amplifiers stable with any load The structural design answer to the question "How do you compensate the circuit?" may be "we usually do not", because the circuit structure and topology may be chosen in the way that no dedicated compensation capacitors are needed.

In a few cases when multistage amplifiers need to be used, the feed forward links may provide for the system stability [26, 27].

2.4 Elementary building cells

The next step after finalizing the system structure is its implementation using the library of elementary cells. This library is process specific, and may be hierarchical. The basic library cells (first level of hierarchy) should have the following common features:

- cell is used in almost every IC;
- cell does not need frequency compensation;
- cell should have close to the "ideal" input/output impedance, which allows one to connect this cell to another without possible problems requiring detailed analysis;
- cell is robust to the component and process variations.

Each elementary cell is usually modified for the application. The simplest modifications are the component size changes to reflect the different current or voltage operating point. The structural changes should be applied when a better functional performance is needed, and component parameter adjustment is not sufficient.

Some time ago there was (and still is) the discussion about usefulness of the standard analog cell library. Few companies even developed the analog arrays with circuit and layout libraries [29]. Such libraries are almost never used without modifications in practical analog design. If the cell operation is not clearly understood by designer, then it is better not to use this cell. If there are no questions about the cell operation, then the designer is always willing to modify it in order to better fit the application.

The traditional set of elementary cells (fig. 2-3) includes the single-transistor amplifiers (common-gate (fig.2-3a), common-drain (fig. 2-3b), common-source (fig. 2-3c); simple (fig. 2-3d) and cascoded (fig. 2-3e) current mirrors; the differential source-coupled pair with differential (fig. 2-

3f) and single-ended (fig. 2-3g) output [13, pp. 125-181; 9, pp 193 – 287]. The similar cells can be drawn using PMOS, NPN and PNP transistors.

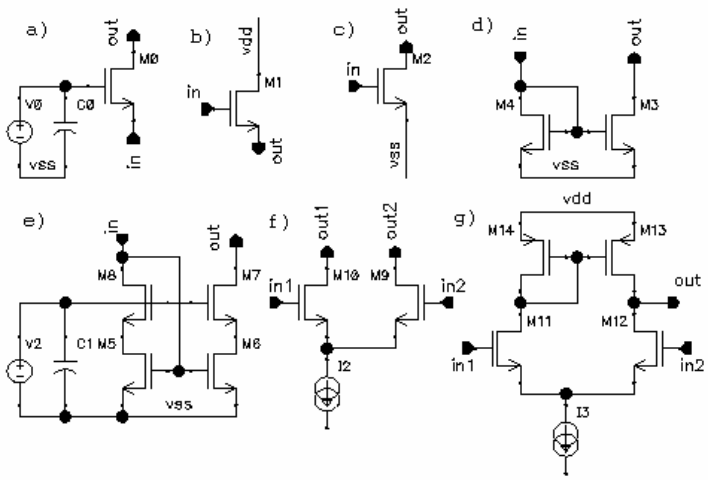


Figure 2-3. Traditional elementary cells of analog ICs

The real voltage sources always have large high-frequency impedance and this effect can be very detrimental for the cell operation at high frequencies. The capacitors at the gates of the common-gate transistors in fig. 2-3a and 2-3e are needed when using any real voltage source.

The structural design solutions to increase the output impedance and to decrease the saturation voltage of the current mirror are considered in chapter 3.

We consider that, in order to improve the design flexibility, this traditional set of cells in fig. 2-3 has to be augmented at least by addition of the current-input differential amplifiers with differential and single-ended outputs (fig. 2-4).

There are 16 types of differential amplifiers classified by the ratio of the differential and common-mode input and output impedance to the signal source and load. Out of this variety, only a differential pair (fig. 2-3f, large input and output, both differential and common-mode impedance) is in general use. Yet, the stages with low differential and high common-mode input impedance and with current output (fig. 2-4) add, as it will be demonstrated in the following chapters, a significant flexibility to design, and deserve to be a part of designer's armory.

The cell in fig. 2-4a has a relatively low input impedance which is equal to $r_{in} = \delta V_{in} / \delta I_{in} = 2 / g_m$, where g_m is defined by I_0 and the ratio of aspect ratios $(W_{M3}L_{M1})/(W_{M1}L_{M3})=(W_{M2}L_{M0})/(W_{M0}L_{M2})$.

The differential output current is equal to $\delta I_{out} = \delta I_{in}(\frac{W_1}{L_1})/(\frac{W_1}{L_1}+\frac{W_3}{L_3})$ and is limited by the product $I_0(W_{M1}L_{M3})/(W_{M3}L_{M1})$.

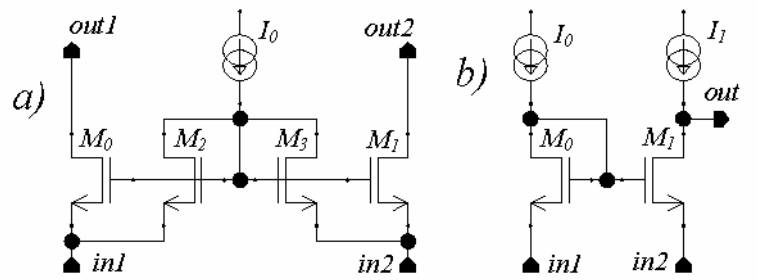


Figure 2-4. Additional cells useful in the structural design: current input amplifiers

The transconductance characteristic (output current versus input voltage) of the fig. 2-4a cell is shown in fig. 2-5a. It is similar to the characteristic of the differential pair of fig. 2-3f.

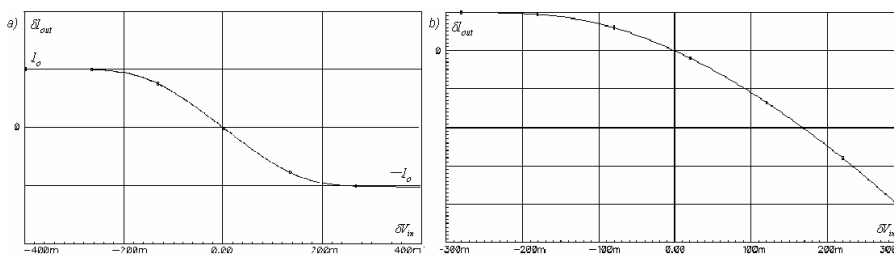


Figure 2-5. Differential output current versus input voltage for the fig.2-4 cells

The output current of the cell shown in fig. 2-4b is limited by I_1 (I_1=I_2) only in one direction. For this cell the transconductance characteristic is not symmetric (fig. 2-5b), and is virtually unlimited in the other direction.

Each CMOS analog IC includes the voltage sources for the gates of cascode transistors, so these cells also should belong to the basic library. A diode-connected transistor operating in strong inversion is used very often (fig. 2-6a). A word of warning may be useful. The temperature coefficient (TC) of V_{gs} of such a transistor may not match the temperature coefficients of the cascading transistors of the current sources that this cell is serving. For example, in the typical current distribution circuit of fig. 2-6b the cascoding transistors M_1/M_2 are usually short-channel and operate in (or close to) weak

inversion, so the TC of their V_{gs} is different from TC of the V_{gs} of transistor M_0, and the drain voltages of M_3/M_4 will vary with temperature.

For better process and temperature robustness, the cell of fig. 2-6c may be preferable [31].

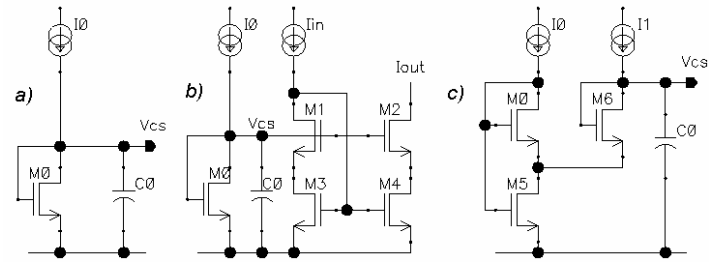


Figure 2-6. Voltage sources for the cascode transistor gate bias

For the bipolar current sources simple cascoding (fig. 2-7a) is less common due to the dramatic changes in operation if any transistor becomes saturated. The current distribution in bipolar circuits can be done using emitter degeneration resistors (fig. 2-7b). The FET transistors if available on the process may be used for cascoding (fig. 2-7c). Yet, the feedback current control circuits (subchapter 3.4) are preferable. The circuits of fig. 2-7b and 2-7c may also be included in the library.

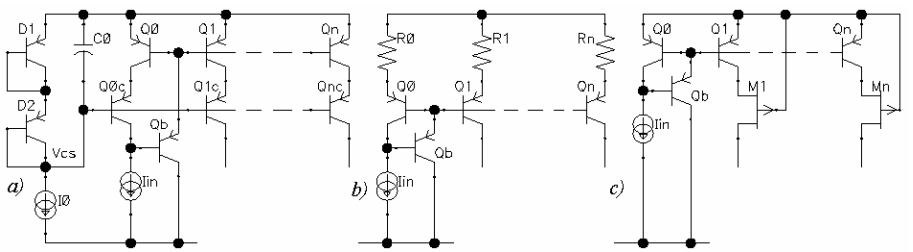


Figure 2-7. Current sources and current distributors in bipolar circuits

Any analog circuit can function only if its node voltages and transistor operating points are within specified boundaries. A common design task is limiting these node voltages and transistor currents to the operating area. The simplest nonlinear cells often used for "soft" limitation are the diode-connected transistors and transistors in common-gate configuration (fig. 2-8a and 2-8b).

The diode-connected transistor can be used for current rectification (fig. 2-8d). The common-gate circuit can be used for this purpose as well (fig. 2-8e) where $V_{out} \le V_{gs}$.

The same circuits can be used as the voltage limiters (fig. 2-8f) where $V_{out} \le V_0 + V_{gs}$ and (fig. 2-8g) where $V_{out} \le V_{gs}$. The limiting characteristics of both cells are "soft" and defined by the dependence $I_D(V_{gs})$ (fig. 2-8c). The slope of this curve and "softness" of limitation can, to some extent, be adjusted with the ratio *W/L* of the transistor. A sharper limiting characteristic, if needed, can be achieved using a dedicated feedback circuit comprising a reference for the limiting voltage, an amplifier and the current shunting transistor (chapter 8).

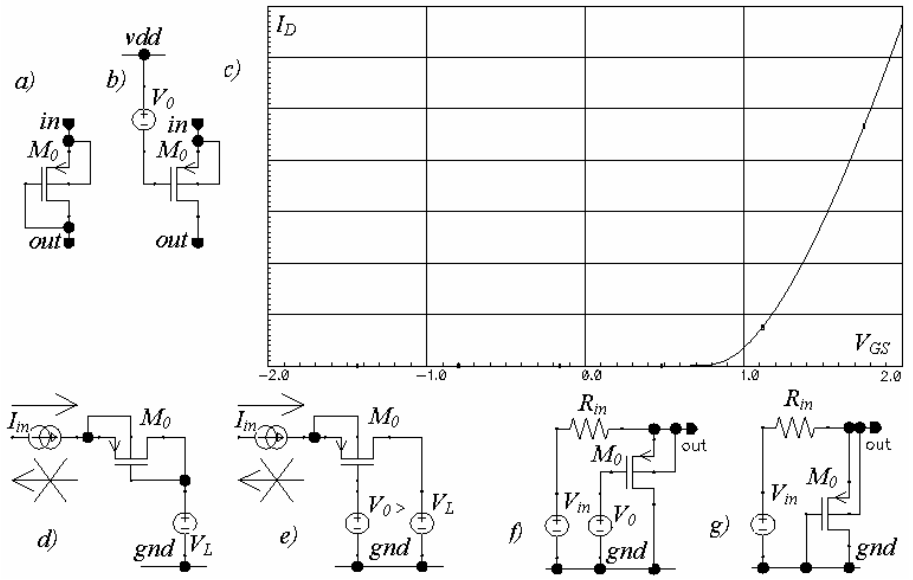


Figure 2-8. Elementary cells for voltage and current limiting

The bipolar limiters are similar to the MOS ones with the corresponding substitution of one type of transistor with another.

Another nonlinear cell frequently used in the design of class AB stages is a selector of smaller or larger value from two currents or voltages. Fig. 2-9a shows the maximum voltage selector cell suitable for both bipolar and CMOS realizations. In this circuit, the output voltage $V_{out} \approx (\max(V_{in1}, V_{in2}) - V_{be})$.

The CMOS-only minimum current selector is shown in fig. 2-9b [32]. All transistors in this circuit should be matched. If $I_{in1} >> I_{in2}$ then $I_{out} = I_{dM3} \approx I_{in2}$ and if $I_{in2} >> I_{in1}$ then $I_{out} = I_{dM2} \approx I_{in1}$, with a smooth transition between these two states. The circuit accuracy can be improved

(fig. 2-9c) by a dedicated feedback loop controlling $V_{gsM0}=V_{gsM1}$ when $I_{in2}>>I_{in1}$ and $V_{dsM3}=V_{gsM5}$ when $I_{in1}>>I_{in2}$.

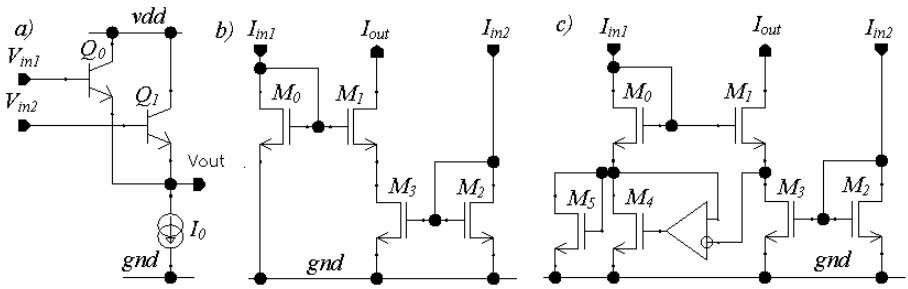


Figure 2-9. Extremum selector cells

The designer's basic library by no means is limited by the above considered cells. Additional cells may be created for specific applications and for non-generic processes. For example, JFET transistors, if available on the process, can be used in input differential pairs; the low- and high-V_{TH} "natural" MOS transistors can be used for clamps, for voltage sources and cascoding; the high-voltage depletion-mode MOS transistors for cascoding, etc.

2.5 Summary

The structural design methodology has solid theoretical and philosophical foundation and can be applied in any control systems development, from amplifiers to social systems. In the area of analog IC design, its main features are the following:

1. Any system parameter is perfected and achieved by structural changes, i.e. by addition of new components and links, and feedback, which is the main consideration for any component connection and the main tool of the designer. The circuit optimization by the variations of component parameters is acceptable only when the structural means of improvement are exhausted.

2. The first run of the system synthesis and analysis is performed on the level abstracted from the energy forms, i.e. with graphic representation of the system under design using signal flow graphs.

3. The circuit synthesis is executed by iterative process of the structure implementations using a library of cells, and a detailed simulation verifies implementation.

4. The circuits and systems that do not comply with definition of "good" systems are rejected (even if they appear acceptable).

Hence, the structural design methodology is a set of rules describing where and how to look for solutions to the problem. The results are usually not isolated and specific, and different circuit solutions for the same problem can be created by designers with different background and experience or by the same designer if more time is allowed to tackle the same problem.

Chapter 3

BIASING

Current and voltage biasing is the foundation that allows any analog circuit to function without errors. Bias circuits are relatively simple, and may serve as a good practical introduction to the structural design.

Verification of the biasing circuit has to be done by transient simulation, with slow ramping of the supply voltage first, followed by switching of the load current and supply voltage. AC small-signal simulations may be deceiving. For example, it is interesting to notice that AC simulations of the phase margin in the voltage regulators provide results that would be unacceptable in the OpAmps (5^o phase margin, for example). Yet, the simulations of step-response transient or experimental verification of this response on the silicon chip show quite sufficient excess stability. Also, the start-up problems in biasing circuits are likely to cause conditional instability, and AC simulations do not detect this circuit problem.

Start-up circuits are the subject of many discussions, and a frequent silicon fault cause. Paul Brokaw formulated the following practical rule: if on the circuit diagram one can draw a closed line around the supply bus that crosses only drains of MOS devices or collectors of bipolar transistors, then such circuit has the second stable condition when all components are off [106].

One can find four types of current biasing in OpAmps: proportional to absolute temperature (PTAT) currents; g_m-matching; temperature-independent currents; and currents with a negative temperature coefficient (TC).

The bandwidth of an OpAmp with Miller compensation is equal to $\omega_t = g_m / C_M$, where g_m is the transconductance of the input stage and C_M is the Miller compensation capacitance. g_m of the bipolar differential stage is inversely proportional to absolute temperature, i.e., $g_m = \frac{I_C q}{kT}$. A similar

relationship is valid for a MOS differential pair operating in a weak inversion, and g_m of such a pair has similar TC.

As a result, biasing with PTAT current for bipolar or MOS differential pair operating in a weak inversion will provide the input stage with a stable transconductance versus temperature, and, hence, a stable OpAmp bandwidth.

In the absence of slew boost circuits (chapter 5) the OpAmp slew rate is equal to $dV_{out}/dt = I_0/C_M$, where I_0 is the tail current of the input stage, and C_M is the Miller capacitor of the second stage. With PTAT biasing the slew rate will increase with temperature.

The Class AB output stage is another block where PTAT biasing may be necessary. The load capacitance, C_L, adds an additional pole with the time constant $T_{out} = C_L/g_{mout}$ to the OpAmp transfer function. Here, g_{mout} is the sum of the transconductances of the output devices. In the absence of a load current the value of g_{mout} is minimal. The output devices are biased by the quiescent current, I_q, and this current is usually small. The size of the output devices is usually chosen to supply the maximum load current. Without a load these devices operate in weak inversion. To keep g_{mout} stable versus temperature, the current I_q must be a PTAT current.

A MOS differential pair operating in strong inversion has different from the PTAT temperature coefficient of g_m. For this case, to keep the bandwidth stable, the g_m-matching biasing circuit may give better results [34]. As will be shown (subchapter 3.2), g_m-matching biasing is also advantageous for MOS differential pair operating in weak inversion.

The temperature independent and negative TC biasing currents are used in offset and offset temperature drift trimming circuits (chapter 5). In some applications, it is more important to keep I_q and the slew rate of the OpAmp stable rather than its bandwidth, and this is achieved with temperature-independent biasing.

3.1 PTAT biasing circuits

It is known [9] that the difference of V_{be} voltages of two bipolar transistors, with the emitter area ratio equal to N, and biased by equal currents, is equal to

$$V_{PTAT} = V_T \ln N \tag{3.1.1}$$

If this voltage is applied to a constant resistor, then the current in this resistor will be the required PTAT current (a similar result is obtained when two matched transistors are biased with two different currents that set the current density ratio equal to N).

Now, we have to find a circuit that will be able, using a dedicated feedback, to support the condition where V_{PTAT} voltage is applied to a

constant resistor. The equation (3.1.1) is the basis of dozens of different PTAT generators. Fig. 3-1a shows a popular circuit that satisfies this structural design approach and can be realized in any CMOS process.

Fig. 3-1b shows the signal graph of this circuit. This signal flow graph is very similar to the graph of differential structure with common-mode feedback (see Appendix) and to the graph of differential input stage (fig. 2-1d). The difference is in the sign of the loop transfer: the feedback in loop 1 is positive and in loop 2 is negative. The steady-state operation of the circuit is achieved when both feedback loops have the same gain, providing that the gain $A(s) = g_{m0,1}Z_{g4}g_{m4}Z_{g5}$ is large (here $g_{m0,1}$ is the transconductance of the pair M_0/M_1, Z_{g4} is the equivalent impedance at the node g_4, g_{m4} is the transconductance of M_4, and Z_{g5} equivalent impedance at node g_5). Any difference in the gains through the links g_{m5}/g_{m0} and $g_{m6}/(R_0+1/g_{m1})$ would cause a difference of V_{s0} and V_{s1} and change in V_{g5} through operation of the amplifier M_0-M_4 loaded by the diode-connected M_7.

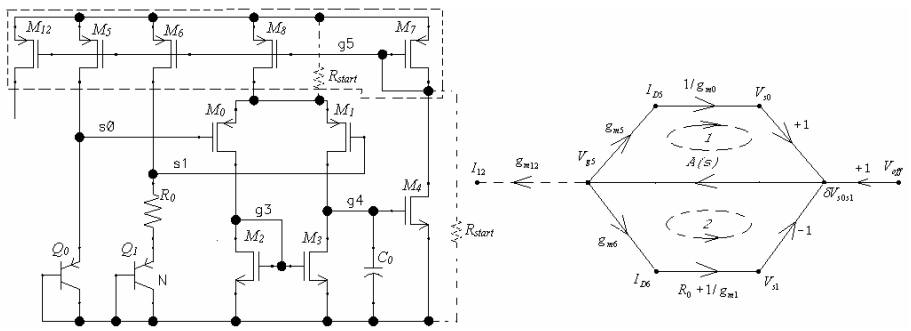


Figure 3-1. PTAT current generator

The circuit has the following sources of errors:

1. The temperature coefficient of resistor R_0; if zero-TC resistors are not available, then a limited compensation for this error can be achieved by assembling R_0 from resistors with different TC. For example, a poly-Si resistor having positive TC and N-well diffused resistor in series with negative TC would result in temperature-independent resistor[109].

2. The offset voltage, V_{off}, of the M_0-M_4 amplifier adds to the PTAT voltage (between Q_0 and Q_1), so that the output current becomes

$$I_{PTAT} = \frac{(V_{off} + V_T \ln N)}{R_0} \qquad (3.1.3)$$

This error can be decreased by replacing of the M_0/M_1 pair with bipolar transistors, if available, and better matching of M_2/M_3 in the current mirror. The offset voltage of the bipolar differential pair is also PTAT, so this stage

may be intentionally mismatched, and the PTAT offset voltage of the amplifier will sum with the PTAT voltage generated by Q_0/Q_1. In this way one can reduce the influence of the amplifier random offset and noise, as well as increase the closed-loop gain (transconductance of the bipolar pair is larger than MOS with the same tail current).

3. Current mismatch of M_5/M_6 and emitter size mismatch of Q_0/Q_1 are equivalent to the variation of N. The current mismatch of M_5/M_6 also varies with V_{ds} when V_{DD} is changing, and the error introduced by this mismatch is a cause of the finite PSRR.

4. Power supply variation is also causing the difference of V_{ds} voltages for the pair M_5/M_6 and M_7, so the current I_{D7} in M_7 is different from I_{D5} or I_{D6}. This changes the transconductance of M_4 and results in the change of voltage at *g4* node. This last change divided by the voltage gain of the M_0-M_1 pair may be represented as an equivalent change of V_{off} voltage. To improve PSRR, this circuit may be powered from a subregulated voltage source (subchapter 3.5).

Other properties of this circuit are following.

The minimum supply voltage of the circuit is equal to $V_{be} + V_{gsPMOS} + V_{sat} \approx 1.5$ V. The circuit has a second stable operating point when no current flows (following the Brokaw rule: dashed line in fig. 3-1 encircles the positive supply bus and, in absence of the R_{start}, on each intersection with the wire there is the drain only in one of the sides). A start-up can be provided by a small current introduced between the node *g5* and ground or between the sources of M_0/M_1 and V_{DD} as shown by dashed line resistors R_{start} (see subchapter 3.7).

The frequency stability of the circuit is defined by the negative feedback loop *s1-g4-g5-s1*. This is a 3-stage amplifier loop (the classification is given in chapter 4), and it does require compensation. The open-loop gain of this loop is: $A_L(s) = g_{m01} Z_{g4} g_{m4} Z_{g5} g_{m6} (R_0 + 1/g_{m0})$. Compensation by a Miller capacitor between the nodes *g4* and *g5* is not advisable, since the high-frequency ripple on the positive supply rail will be fed directly into the high-impedance node *g4*, reducing the PSRR at the frequency where gain of the amplifier $A(s)$ starts to roll off. The compensation is better achieved with the parallel capacitor C_0 and by the diode-connected M_7 which serves to decrease the open-loop gain by decreasing $Z_{g5} = 1/g_{m7}$.

Modifications of the PTAT biasing core.

Alternative PTAT generator is shown in fig. 3-2. It addresses some of the shortcomings of previous circuit. The error caused by the amplifier offset is eliminated, and the minimum supply voltage is decreased to about 1 V using the PTAT current generator shown in fig. 3-2.

The signal graph of this circuit is similar to the graph of previous circuit. Two variables are compared (this time it is the collector current of Q_0 and the drain current of M_5), and, with the high gain of Q_2, the transmissions via the side branches of the graph are equal.

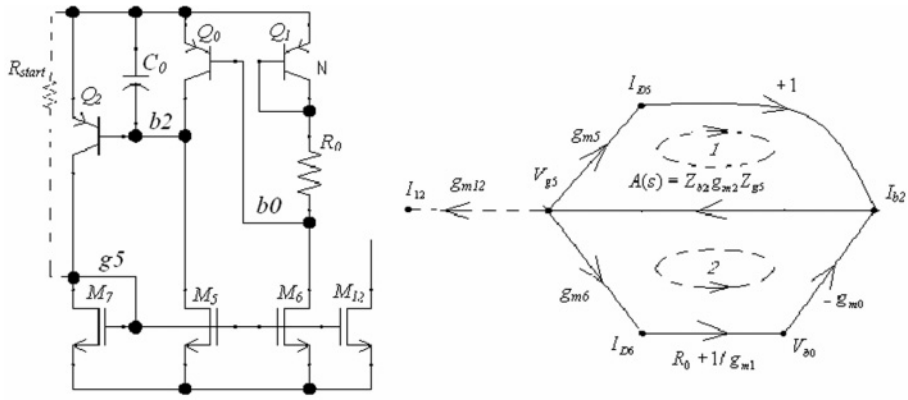


Figure 3-2. PTAT current generator using lateral bipolar transistors

The two equal currents, I_{D5} and I_{D6}, are required to establish the PTAT voltage and are obtained via matching by the same mechanism as in the previous circuit. Yet, the amplifier structure is no longer symmetrical. The negative feedback loop has 3 gain stages, and needs frequency compensation (C_0 and M_7). This circuit also needs a start-up that can be accomplished with a small current introduced between V_{DD} and *g5* (as shown by resistor R_{start}).

This circuit is designed using lateral PNP bipolar transistors. This transistor is available on any CMOS process (it may be a lateral NPN for N-substrate processes). Like any lateral bipolar transistor, it includes the vertical PNP parasitic structure to the substrate (see fig. 1-4). The emitter current leakage through this parasitic structure can be quite high. However, this leakage does not affect the PTAT voltage as R_0 is moved to the collector of Q_1.

The minimum supply voltage for the circuit of fig. 3-2 is (V_{sat}+V_{gsNMOS}) $\approx$ 0.8 V.

The PTAT circuits shown in fig. 3-3 also use lateral PNP transistors.

It is easy to see that the circuits of fig. 3-3 are also described by the same type of graph as that shown in fig. 3-1. The reader can verify that both circuits are using the same idea realized with similar transformation of the signals. The circuit shown in fig. 3-3b comprises the current-input amplifier $M_0/M_1/M_2/M_3$ instead of the voltage input amplifier $Q_2/Q_3/M_2/M_3/M_4$ of fig. 3-3a. The negative feedback loop (M_0-M_5-Q_0) of the circuit with current-

input amplifier has only one gain stage, and the circuit is stable without a dedicated frequency compensation capacitor.

Contrary to the previous circuit, the circuits of fig. 3-3 are sensitive to the leakage via parasitic transistor. Yet, the buried N+ layer, if available, reduces this leakage to less than 1% of collector current. Also, this leakage may be less than 5% for modern processes allowing a very narrow base ($\leq$0.35 um) of lateral transistors. Hence, the circuits of fig. 3-3 should be used in the process providing the narrow-base transistors and/or the buried N+ layer.

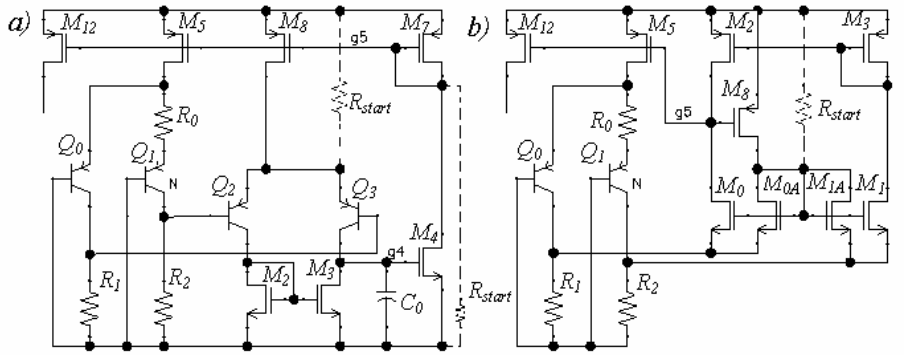


Figure 3-3. PTAT current generator with low-voltage supply

Both circuits need a start-up using a small current source as shown by dashed line resistors. It can be a resistor between *g5* and negative supply, between V_{DD} and the emitters of Q_2, Q_3 (fig. 3-3, a) or between V_{DD} and the gates of M_0, M_1 (fig. 3-3, b).

The minimum supply voltage for the circuits of fig. 3-3 is $V_{sat}+V_{gsPMOS}+V_{R4/5} \approx 1.2$ V, and it can be less than 1 V for fine-line processes with smaller V_{TH}. Their applications with minimum supply voltage of 1.1 V, and implementation in any process with bipolar devices having insignificant substrate leakage do not represent any difficulties.

3.2 MOS gm-matching biasing

For MOS differential pair operating in strong inversion g_m decrease with temperature is stronger, and biasing with a PTAT current source is not sufficient to stabilize the pair transconductance g_m (fig. 3-4).

The circuit for g_m-matching is biasing a transistor to obtain the temperature independent transconductance. Then, by matching the current density of this transistor with input differential pair, one can obtain the required temperature stabilized transconductance of this pair.

An example of this circuit is shown in fig. 3-5. We remind that for MOS transistors in strong inversion the drain current can be described by following equation: $I_D = K\frac{W}{L}(V_{GS} - V_{TH})^2$. If two transistors M_1 and M_2 have equal drain currents (this is supported by matching M_6 and M_5), and $W_1L_2/W_2L_1 = N$, then [34]

$$g_{m2} = \frac{\sqrt{2}(1-\sqrt{1/N})}{R_0} \qquad (3.2.1)$$

If the transistors of the OpAmp input stage match with M_2 then g_m of the input stage is defined, as a first approximation, by R_0 only, and does not depend on temperature.

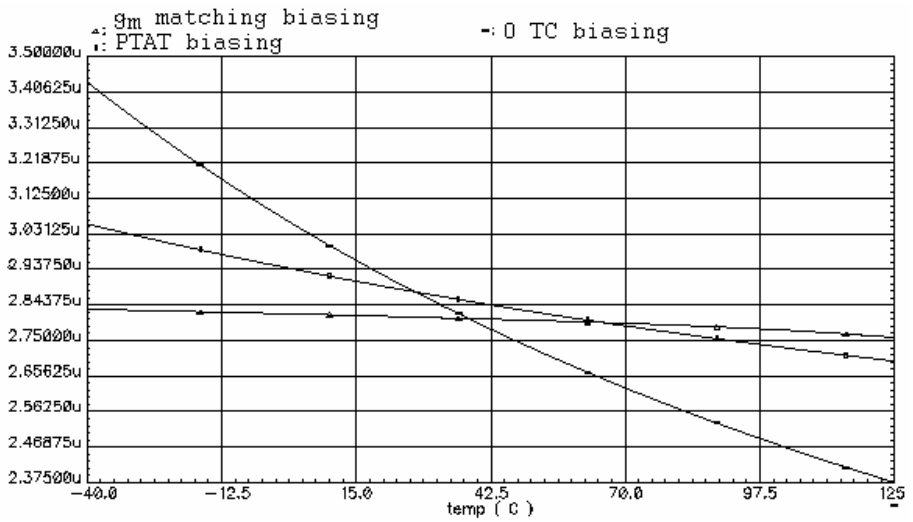


Figure 3-4. g_m versus temperature of MOS differential pair (strong inversion)

To provide the result 3.2.1 one needs two matched currents. They can be developed, by mirroring, from the same diode-connected transistor. If one organizes a high gain amplifier that amplifies the difference of two required currents and controls the voltage at node *g5*, then the problem will be solved. Hence, one needs a system described by the now familiar two-loop graph, and from this graph one arrives to the circuit shown in fig. 3-5. The current of M_2 is compared with the current of M_6. The difference is amplified and is used to establish the voltage at the drain of M_7 from which both required currents will be obtained.

The circuit has both positive and negative feedback loops. The negative feedback loop includes three gain stages and should be compensated. The

compensation is provided by the capacitor C_0 and the loop gain degeneration using the diode-connected M_7.

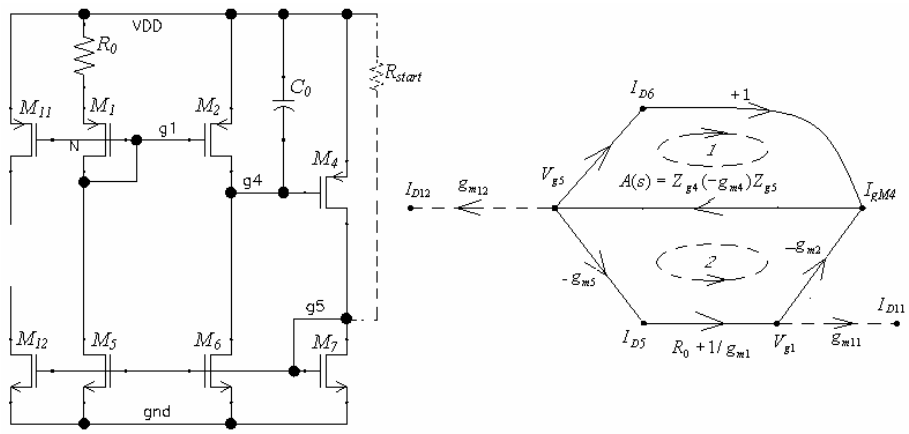


Figure 3-5. g_m-matching biasing circuit

The circuit also needs a start-up. It can be done by a small current source connected between the node *g5* and V_{DD} line, as shown.

The PSRR of this circuit is relatively good because $V_{g4} \approx V_{g1}$ and the V_{ds} of the transistors that should be matched (M_5/M_0 and M_1/M_2) are approximately equal. The circuit is operational with the supply voltage down to $(V_{gsPMOS}+V_{sat}) \approx 1.2$ V.

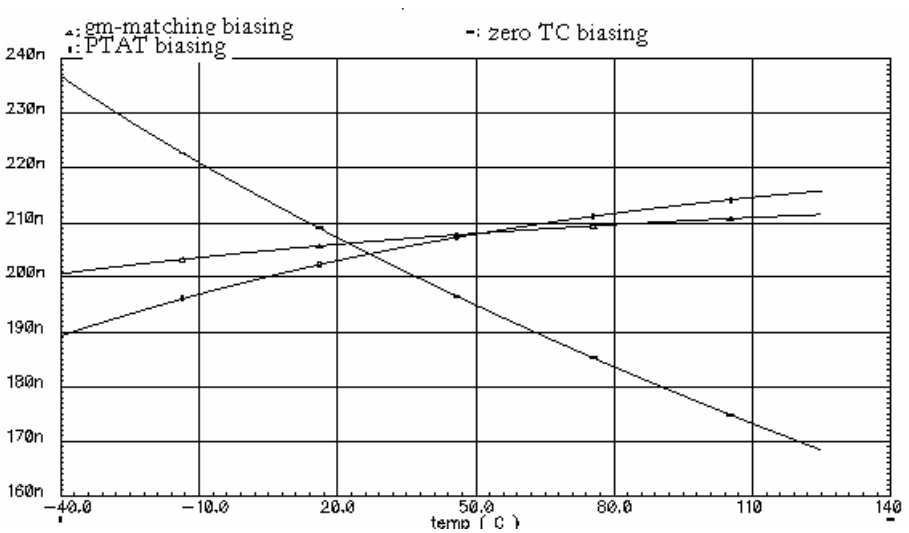


Figure 3-6. g_m versus temperature of the MOS differential pair (weak inversion)

Equation (3.2.1) is valid when transistors M_1, M_2 are operating in strong inversion. The equations for these transistors operating in weak inversion are too cumbersome for any practical use, however, the results obtained in silicon and simulations show that g_m-matching biasing circuits still provide better results then PTAT biasing ones (fig. 3-6).

3.3 Negative-TC and zero-TC current generators

Negative-TC current generators can be used for temperature measurements in the protection units of power OpAmps (chapter 8), and for trimming of the offset voltage temperature drift (chapter 5).

The most accurate variable with negative TC that one can find in IC components is the bipolar transistor emitter-base voltage, V_{be}. It has TC that is approximately -2 mV/°C. The temperature dependence of V_{be} is not perfectly linear, and has a bow-like curvature that can cause about 4 mV (0.7%) of error in the -40 to 125°C range. The process variation of V_{be} between wafer lots can be as large as 20 mV (3%).

This voltage is transformed into a current with negative TC. Fig. 3-7a shows the simplest generator of a current that is equal to V_{be}/R_0. Nonlinearity of the resistor TC is another source of error.

The minimum supply voltage for this generator is $(V_{gs} + V_{be} + V_{sat}) \approx 2$ V. The current I_0 should have a linear dependence on temperature (PTAT, zero- or negative TC).

For smaller supply voltages the circuits of fig. 3-7b or 3-7c may be recommended.

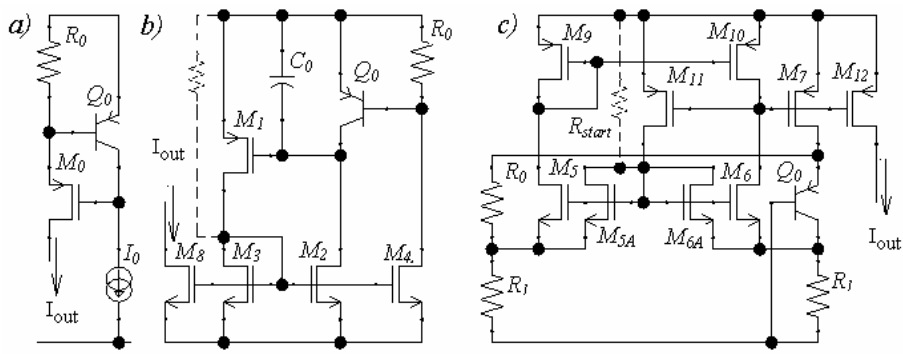


Figure 3-7. V_{be}/R current generators

The circuit of fig. 3-7 b has three gain stages in the negative feedback loop (M_1-M_4-Q_1), and needs compensation (C_0 and gain-degeneration diode M_3). The minimum supply voltage is $(V_{gs} + V_{sat} + V_{R1}) \approx 1.3$ V.

The circuit of fig. 3-7c is utilizing the current-input amplifier ($M_5/M_6/M_{10}/M_{11}$) in the negative feedback loop and does not need compensation. The minimum supply voltage is $(V_{gs} + V_{sat} + V_{R23}) \approx 1.3$ V.

The reader can verify that the circuits of fig. 3-7b and c are built using the idea of two matched currents produced from the same diode-connected transistor. These two currents appear in two corresponding loops, and the difference of the currents is amplified with a high gain amplifier.

The circuits of fig. 3-7b and 3-7c also need a start-up that can be provided by a small current via R_{start}.

Zero-TC current sources are used for biasing in the applications where stability of the OpAmp quiescent current, I_q, versus temperature is more important than stability of the bandwidth. They can also be used for trimming of the voltage offset and offset temperature drift.

An obvious way to get a zero-TC current is to combine PTAT and V_{be}/R currents. These currents can be provided by two separate current generators, as well as combined within one circuit as shown in fig. 3-8.

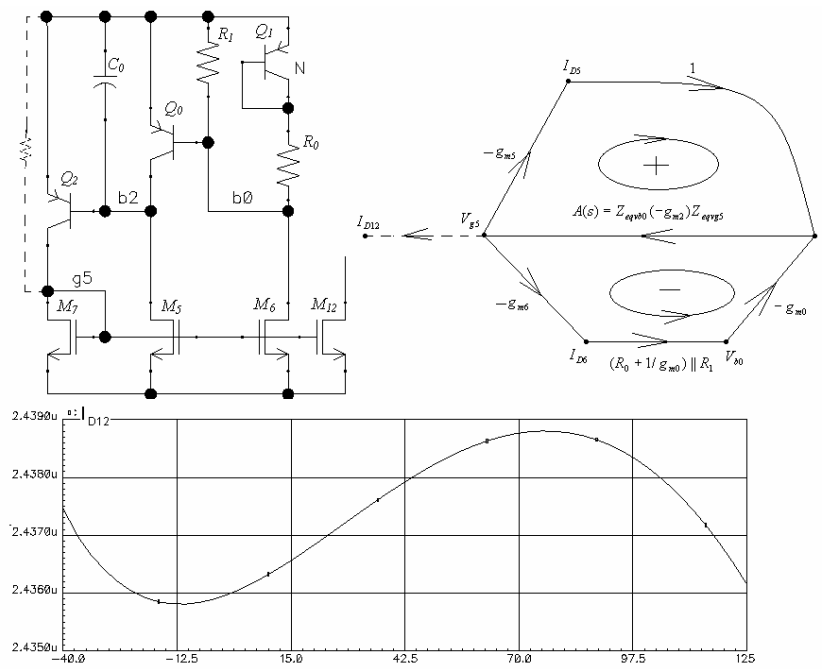


Figure 3-8. Zero-TC current generator

The drain current of M_6 should be a combination of PTAT current with V_{be}/R currents. Due to the transistor matching, M_5 and M_6 should have equal

drain currents. To keep this current as required the designer has to add an amplifier that amplifies the difference between the current in M_5 and an arbitrary current (the collector current of Q_0). If the amplifier gain is sufficiently high, and the amplifier output provides a finite voltage at the diode-connected transistor so that this voltage can be used to develop the matched currents in M_5 and M_6 then the required current in M_6 will be obtained. It is not surprising that the circuit of fig. 3-8 will be described by the same graph as circuit in fig. 3-2, only with a different transfer function in the negative feedback loop. An exact equation describing the temperature behavior of this circuit may be available, yet a modern designer usually relies on physical principle of circuit operation, and the optimal operation point is found using simulations. Using this approach (and that is the one that we are trying to develop), the designer finds that the circuit can provide very good results (about 0.1% current variation with temperature) as shown on the simulation plot with N=8, R_0=52 k, R_1=710 k.

The circuit of fig. 3-8 is operational with supply voltage as small as 1V. It has three gain stages in the negative feedback loop and requires compensation (C_0/M_7).

A zero-TC current generator having a similar structure but using a current-input amplifier (and, as a result, does not require compensation) is shown in fig. 3-9. In the circuit of fig. 3-9 the current of M_{12} is zero-TC. It is combined from the current though R_1, which is V_{be}/R_1 (i.e., negative-TC) and the currents through R_2 and R_3 that are both PTAT.

Both circuits of fig. 3-8 and fig. 3-9 need a start-up, for example, by a small current between V_{DD} and $g5$ for the circuit of fig. 3-8 or between V_{DD} and the gates of M_0, M_3 for the circuit of fig. 3-9 as shown by the dashed line resistors.

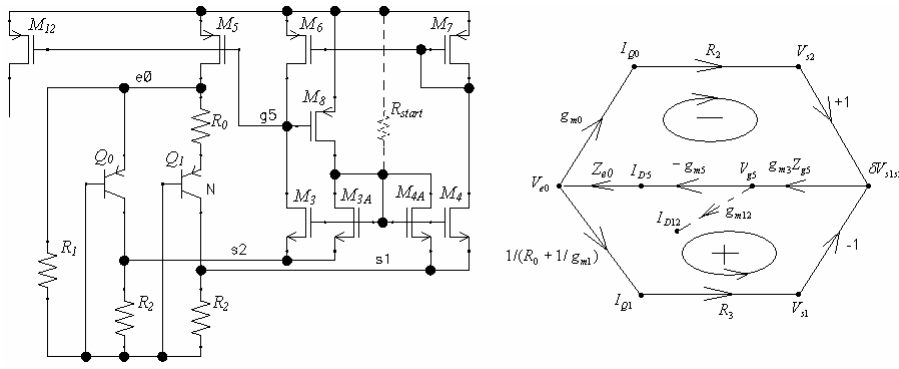


Figure 3-9. Zero-TC current generator stable without compensation capacitors

Fig. 3-10 shows a single cell that generates all three types of current biasing generators by adding two current input amplifiers. Two new feedback loops do not require compensation.

The first additional feedback loop uses PTAT voltage across R_2 as the input signal. Transistors M_3 and M_5 with the current mirror M_6, M_9 and M_{10} form the current-input amplifier. The current through R_5, if the gain of the amplifier M_6/M_{10} is high enough, is a PTAT current as well. But the drain current of M_6, and so of M_9, are PTAT currents. Hence, the current of M_{10} is PTAT, and so is the current of M_{11} that can be sourced to any load. If the current of M_5 is small then the current I_{D10} is proportional to the current through R_2 with R_5/R_2 ratio and can be mirrored to the any load by M_{11}.

The second feedback loop uses a fraction $(R_1/(R_1+R_{1A})$ of the V_{be} voltage as an input signal for the current-input amplifier $M_{15}/M_{16}/M_{17}$. If the amplifier biasing currents $I_{D14} = I_{D13} << V_{be}/(R_1+R_{1A})$ then the current through M_{15} is proportional to V_{be}: $I_{D15} = V_{be}\dfrac{R_{1A}}{R_6(R_1+R_{1A})}$.

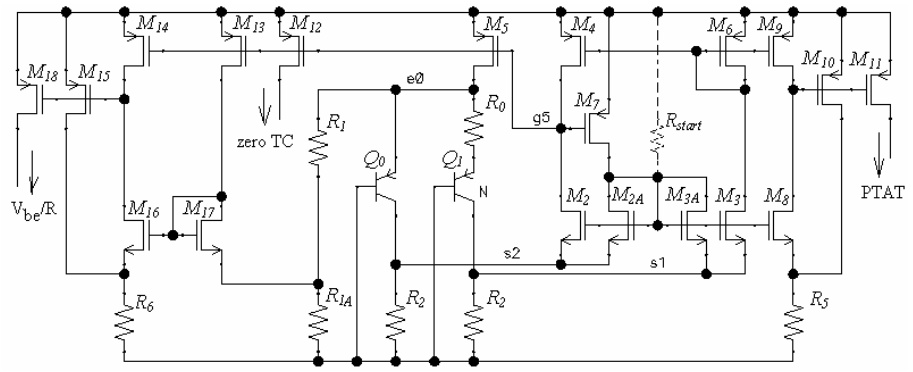


Figure 3-10. PTAT, V_{be}/R and zero-TC current generator

The minimum supply voltage for the circuit of fig. 3-10 is $(V_{gsPMOS}+V_{sat}+V_{s1s2}) \approx 1.2\text{V}$.

Using the current-input amplifiers allows one to avoid the compensation capacitors while preserving low supply voltage capability. The biasing currents of these amplifiers cause errors in the circuit transfer coefficients. If these errors are not acceptable, than the voltage-input variants should be used with the drawbacks of requiring compensation capacitors and/or larger supply voltage.

3.4 Current mirrors and sources

Nearly every block in an OpAmp contains at least one current source. The important parameters of a current source are the following:

- output resistance (or variation of the current in the output voltage range);
- output capacitance;
- saturation voltage (or how close the output voltage can swing to the supply rail);
- power supply rejection (or variation of the current in the power supply range)
- noise.

Usually most of the current sources are obtained by mirroring from the primary biasing core units (considered above). Using the cascoded current mirrors provides increased output resistance.

In bipolar circuits no current sources should be mirrored from the same core if one of them can enter in saturation, even briefly (for example, on start-up). Existing models are not sufficient for adequate simulation of saturated bipolar transistor, and the simulation results may be different from the silicon operation if saturation is not prevented.

The practical gate bias voltage sources always have some source impedance especially noticeable at high frequencies. Every transistor has a parasitic capacitance between its gate and drain. These two factors may cause high-frequency signal leakage between the loads of different current sources mirrored from the same origin. To avoid this, a good practice is to add a bypass capacitor across the cascode gate bias voltage source (C_0 in fig. 3-11).

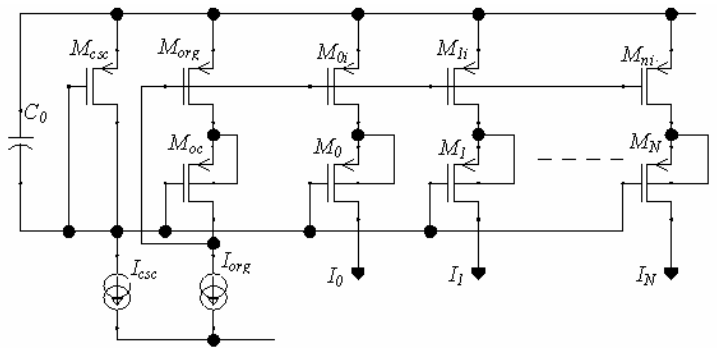


Figure 3-11. Set of cascoded current sources

The gate bias voltage source for cascode transistors can be created using the current I_{csc} and a diode-connected M_{csc} (fig. 3-11) For low voltage

designs, and for better control over temperature of the drain-source voltages of mirror transistors M_{org}-M_{xi}, a more complicated cell [31], as shown in fig. 2-6b, may be recommended.

To minimize noise, the transconductances of mirror transistors (M_{org}, M_{xi} in fig. 3-11) must be small [108]. This results in longer channel devices operating in strong inversion, and having larger parasitic capacitances. At the same time, the current source should introduce as little of parasitic capacitance as possible in the unit where it is used. The cascoding transistors (M_0-M_n in fig. 3-11) isolate large parasitic capacitances of the mirror transistors from the load. Transistors M_0-M_n should be as small as possible to decrease their parasitics (i.e., they should be of minimal allowed length and minimum possible, for a given current and chosen V_{gs} width).

Impact ionization leakage current between the drain and body of CMOS transistors becomes noticeable even with a relatively small (1 V) drain-source voltage (see fig. 1-7). Such drain-body leakage decreases the output impedance and accuracy of the current source. This effect can be especially damaging in 2-stage amplifiers with an open-loop gain defined as the product of a transconductance and equivalent impedance at the current summing node. To eliminate this effect, it is desirable to connect the body of cascode transistors to their sources; in this case the total output current of the source will still be equal to the drain current of the mirror transistor.

The body to source connection is not possible for single-well processes where the bodies of all NMOS transistors are inherently connected to the substrate (or of all PMOS for the process with N-type substrate). Two structural design solutions (applying dedicated feedback) for this problem are shown in fig. 3-12.

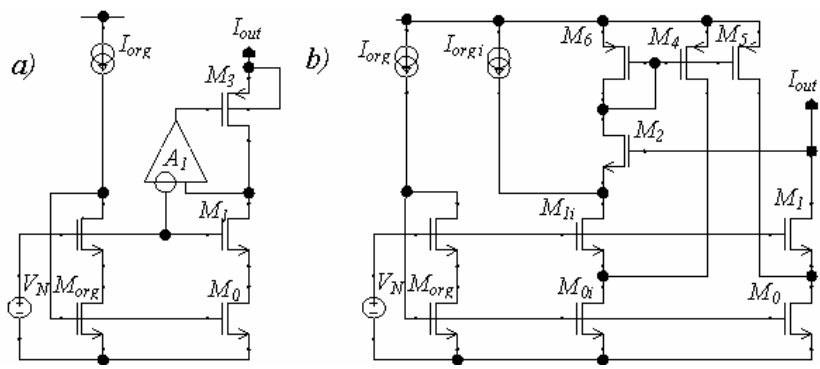


Figure 3-12. Decrease of the impact ionization leakage effects in the current source.

In fig. 3-12a [35], in order to reduce the drain voltage of M_1 and simultaneously to preserve the high output impedance we organize a local

feedback loop including M_3 and amplifier A_1. The current of M_3 is compared with the drain current of M_1, and the high loop gain provides equality of these two currents. At the same time, due to the fact that the second input point of A_1 is at the gate of M_1, the drain-source voltage of M_1 will be equal to its gate-source voltage. This minimizes impact ionization in M_1. The PMOS transistor M_3 has its body connected to the source and impact ionization effects in this transistor do not change I_{out}.

The circuit of fig. 3-12b utilizes compensation of the leakage by a current derived from the leakage measurements in the reference transistor. It is the feedback loop that detects the amount of the leakage current and then subtracts it from the initial current source. The leakage detection feedback loop consists of the current source M_{1i}/M_{0i} which matches M_1/M_0, transistor M_2 which keeps the drain voltage of M_{1i} approximately equal to the drain voltage of M_1 (with V_{gs} difference), the reference current I_{orgi} which matches the original current I_{org} (and the desirable current I_{out} of the source) and the current mirror $M_4/M_5/M_6$. M_4 and M_5 are matched, but M_4 and M_6 have a large ratio of their aspect ratios so there is a considerable gain in the M_2-M_6-M_4 feedback loop. As a result, the current of M_4 is close to the drain-body leakage of M_{1i} and the matching current of M_5 is providing necessary compensation for the leakage in the current I_{out} of transistor M_1.

Another factor decreasing the output impedance of the cascoded current source is the finite output impedance of both transistors. The internal feedback loop existing in these sources may have not sufficient gain to provide the high output impedance. This effect can be somewhat diminished by increasing the channel length of both mirroring and cascoding transistors, but, as has been mentioned, the cascoding transistor needs to be as short and as small as possible to decrease the parasitic drain capacitance. In this case an additional local feedback loop can be a solution.

In this respect the circuit of fig. 3-12a is also one of the ways to apply feedback in order to improve the output impedance of the current source.

Two more ways to apply the local feedback and improve the current source output impedance are shown in fig. 3-13. The circuit of fig. 3-13a was developed for the OpAmp gain boosting [38] but it can be used to increase the current source output impedance as well. The feedback loop A_1-M_1 keeps the drain voltage (and, consequently, I_D) of M_0 constant and equal to V_{ds} (this voltage is set by V_0 and M_2) regardless of V_{Iout} variation, as long as V_{Iout} is below $V_{DD} - V_{ds}$.

The circuit of fig. 3-13b comprises the feedback loop A_2-M_5. It keeps the V_{ds} voltages of M_4 and M_5 equal. It ensures that the drain currents of M_4 and M_5 are equal as long as there is gain in the feedback loop and V_{ds} of M_4/M_5 is larger than the offset voltage of A_2.

A regular cascoded mirror source also has a significant trioding voltage (250-400 mV). When the drain voltage decreases below this level the output

impedance sharply drops. Trying to use a long channel device to increase the output impedance increases this voltage and reduces the available swing headroom. The minimal output voltage of the current source shown in fig. 3-12a should not be less than V_{gsM3}. In this respect the circuits shown in fig. 3-13 are improving the output impedance without using long-cannel devices and at the same time decrease the minimum output voltage.

Fig. 3-14 shows how the circuit of fig. 3-13b can be implemented in 0.6u CMOS process. The implemented circuit uses the current-input amplifier. The output current versus voltage characteristic of this circuit may be compared with the characteristic of a regular cascoded current source, both designed for the same I_{out} = 10 uA, V_1=1.3 V. Using the feedback loop improves the current source output resistance in about 50 times and decreases the trioding voltage from 350 mV to 60 mV. Both parameters can be improved further by increasing the gain of the amplifier M_7/M_8.

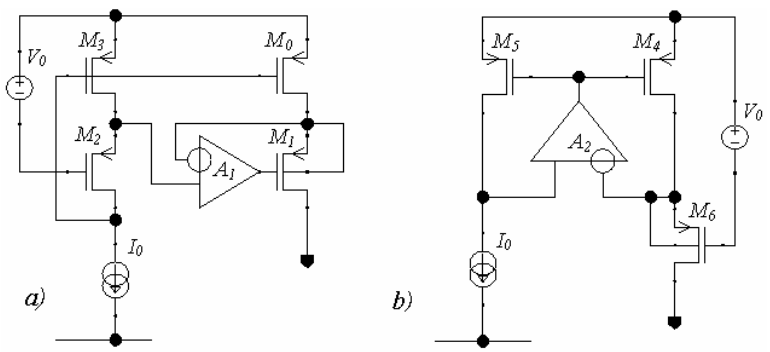


Figure 3-13. Improvement of the current source output impedance

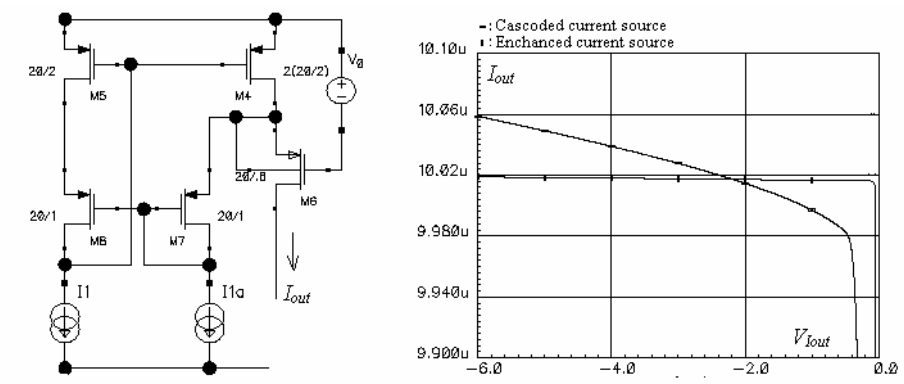


Figure 3-14. Output currents of the regular and improved current sources versus voltage

The current source of fig. 3-14 can be used as a tail current source of the OpAmp input stage. It provides a wider common-mode voltage range and improves CMRR (chapter 5) in comparison with using an ordinary cascode current source.

Another structural design solution for the tail current source is shown in fig. 3-15. Here the feedback loop comprises the sensors (transistors M_{0a} and M_{1a}) that are matching M_0/M_1 of the OpAmp input stage. The gain booster A_1 and M_2, M_3 are providing the required loop gain.

The feedback loop sets $I_{D0a} + I_{D1a} = I_0(R_0 - R_1)/R_1$ (it is assumed that $I_0 = I_{0a}$). The loop does not need compensation in CMOS implementation but with bipolar components a compensating capacitor in parallel with I_{0a} may be required.

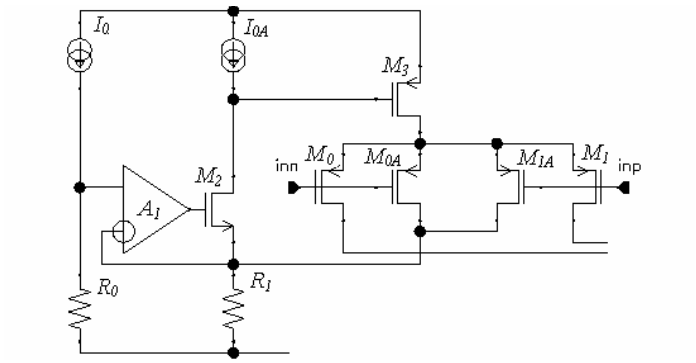


Figure 3-15. Tail current source for the OpAmp input stage

The circuits of fig. 3-12 to 3-15 illustrate the main principle of the structural design: when parameter is important for the circuit operation it should be controlled by a dedicated feedback loop. Then this parameter will be defined by the quality of the parameter sensor and by the gain in the loop.

3.5 Subregulated biasing

PSRR of the OpAmp is defined by design of the input stage and by stability of the core biasing currents versus supply voltage variations. The ultimate solution is to use the core with high PSRR, and, in addition, to subregulate the voltage supply for the biasing core. Such subregulated biasing also has a significantly better high-frequency PSRR [43].

The subregulated biasing supply can be clamped to the positive (fig. 3-16a) or negative (fig. 3-16b) rail of the main power supply. The biasing core output current has to be mirrored and distributed to different subcircuits in the OpAmp circuit. For the N-well CMOS process the NMOS current

sources with significant output voltage may have an extra error due to the impact ionization effects, and for accurate mirroring it may be better to clamp the biasing supply to the positive rail and to distribute the currents using PMOS mirrors. If the impact ionization effect is diminished (see fig. 3-12) then one can tie the biasing to the negative rail and to use NMOS current sources for current distribution.

The subregulators shown in fig. 3-16 should supply the voltage $V_{sbr}=V_{gsPMOS}+V_{gsNMOS}$ to the biasing core. The circuit operation, for example, for the circuit fig. 3-16a can be described the following way. The circuit includes the controlling transistor M_0 and the series connection of M_3 and M_4 diodes operating as a reference. When the load current diminishes, the current in the M_3/M_4 branch increases. Then the current in M_2 increases, and in the negative feedback loop M_0-M_2-M_1 a signal is provided to the gate of controlling transistor M_0. This signal reduces the total current required for the load and reference diode, and returns the voltage at the reference diode and, hence, the voltage $V_{gsNMOS}+V_{gsPMOS}$ to the required value.

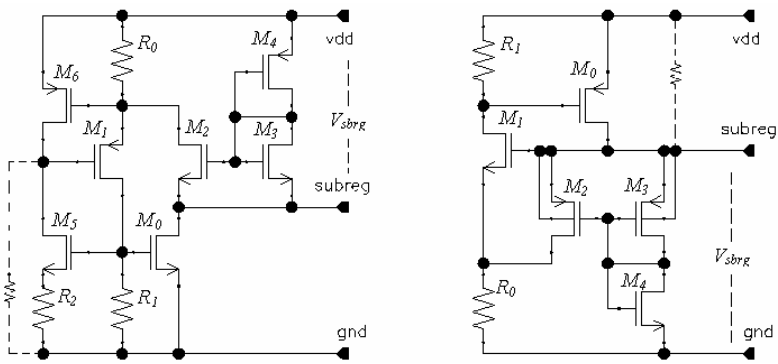


Figure 3-16. Subregulated ($V_{gsNMOS}+V_{gsPMOS}$) supply for the biasing core

Subregulator of fig. 3-16a includes also the positive feedback loop M_1-M_5-M_6 in order to exclude a necessity in the additional biasing current for M_5. This loop is requiring a very small start-up current source (it may be implemented as shown, by the dashed line resistor).

The operation of the subregulator shown in fig. 3-16b is similar. It includes negative (M_0-M_2-M_1) and positive (M_1-M_0) feedback loops. It also needs a start-up current (also shown by the dashed line resistor).

Both circuits do not need compensation. Negative feedback loops in these subregulators should be carefully separated from the main biasing core.

MOS transistors M_4 can be replaced with diodes to decrease the subregulated voltage to ($V_{gs}+V_{be}$). The minimum supply voltage should exceed V_{sbrg} by 50-100 mV.

Any of the considered above mechanisms to create a system with two feedback loops and common gain can be used to control the subregulator output voltage. Fig. 3-17 shows the self-subregulated g_m-matching biasing core and its signal graph.

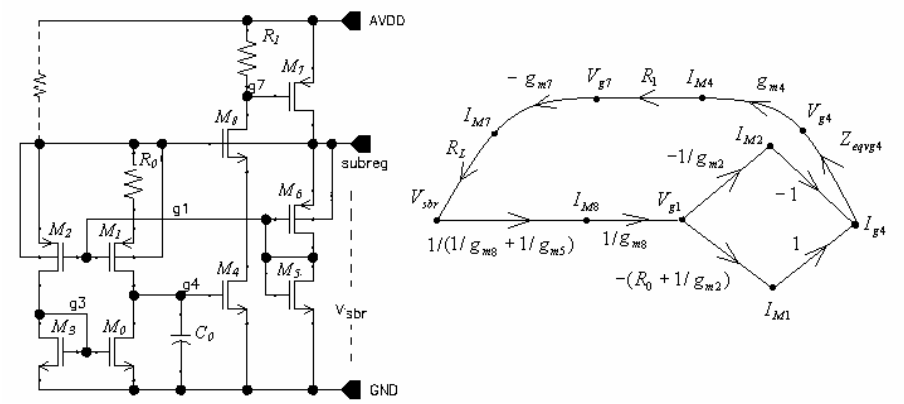


Figure 3-17. Self-subregulated g_m-matching biasing core

This circuit includes the positive (M_1-M_4-M_7) and negative (M_2-M_3-M_0-M_4-M_7) feedback loops as well, it requires a small current for start-up (provided by a resistor), and a frequency compensation (C_0).

The subregulated voltage is V_{sbrg}=(V_{gsPMOS}+V_{gsNMOS}). It can be decreased by replacement of M_5 by a diode, by resistor or even by shortening it (this circuit still will be operational because $V_{gsPMOS} > V_{gsNMOS}$).

3.6 Low-noise bootstrap charge pump

The charge pump can be used inside an OpAmp to create a bootstrapped voltage source for the input stage tail current source. Such approach allows achieving the input rail-to-rail range without second input stage [39]. This charge pump should have very low-noise as all voltage ripples are leaking to the tail current and consequently to the OpAmp output voltage. Even if the charge pump frequency is above OpAmp bandwidth, the inherent nonlinearities in the OpAmp units will cause demodulation and in-band errors.

The charge pump noise has two different components: the digital noise from the switching logic operation, and noise resulting from the output voltage ripples.

Digital switching produces sub-nanosecond spikes of the current in the supply bus and substrate noise due to the capacitive currents caused by fast-changing voltages on the logic nodes. The supply bus noise can be filtered

out by a subregulated logic supply: any subregulator is too slow to process digital switching spikes of the supply current, and serves as a low-pass filter. The circuits in fig. 3-16 are suitable for this purpose. These subregulators have large high-frequency output impedance as the output devices (M_0 in fig. 3-16) work in the common-source mode, so it is does not pass the current spikes from load to the supply rails.

The substrate noise can be decreased by the layout means [2] and by the slowing down of logic switching (charge pump control circuits usually do not need the top speed realized in the process). Such slowdown can be achieved by lowering the logic supply voltage and by reduction of the transistor *W/L* ratio inside logic gates.

A good solution for reduction of the logic-caused, both supply and substrate, noise would by using the current-mode logic similar to bipolar I^2L, but such cell libraries are not common for modern CMOS processes.

Output voltage ripples are the common problem of traditional charge pumps [115]. These voltage ripples can be reduced by some orders of magnitude in the charge pump using current mode of operation (fig. 3-18).

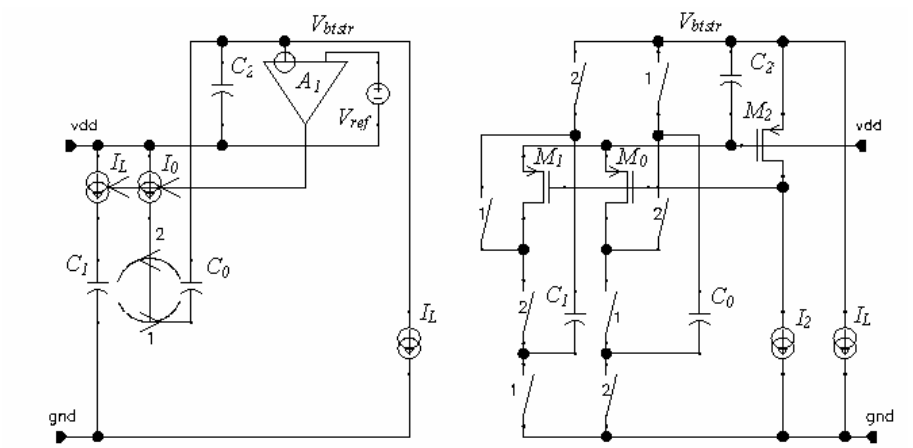


Figure 3-18. Charge pump with suppressed ripples in the output voltage

This charge pump connects the capacitors C_0, C_1 sequentially in series with the regulated output voltage V_{btsr}. The amplifier A_1 controls $I_1=I_0=I_L$. During Phase 1 the first of switched capacitors (say, C_0), is being discharged into the output small storage capacitor C_2 by the constant current I_0. As this current that charges C_2 is equal to the discharging current I_L, there is no voltage variation or ripple at the output during this phase period. At the same time the second capacitor C_1 is being charged by the current $I_1=I_0$. During Phase 2 C_1 is discharging to the capacitor C_2 and C_0 is charging. Thus, the storage capacitor C_2 serves as a deglitching bypass as the output voltage

ripple occurs during interchange of the capacitors only, and it is defined by the ratio of the parasitic capacitances of the bottom plates of C_1, C_0 and the capacitance of the switches to the value of C_2. This ratio can be made quite small. The duration of this ripple is also small, about few nanoseconds only. A simple implementation of this principle is shown in fig. 3-18b where A_0 and $V_{ref} = V_{gsM2}$ are realized by M_2, and I_1 / I_0 are implemented by M_1 and M_0.

3.7 Start-up circuits

Any biasing core needs a small current source to provide the start-up [40, 119, 120]. Usually the value of this current is not important as long as it does not exceed certain range. Few choices for implementation of such current sources are shown in fig. 3-19: very long channel, diode-connected MOS transistor; JFET (if available) with resistor between gate and source; large resistor (for example, using a pinched off diffused layer); leakage element made of the PMOS/PNP structure.

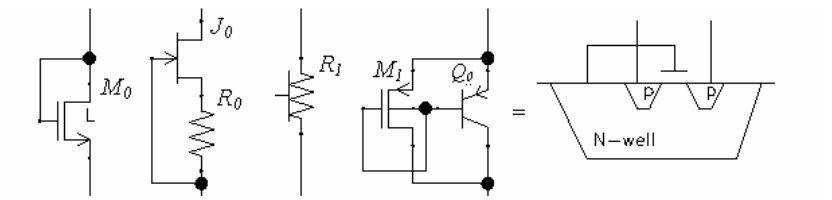


Figure 3-19. Start-up current sources

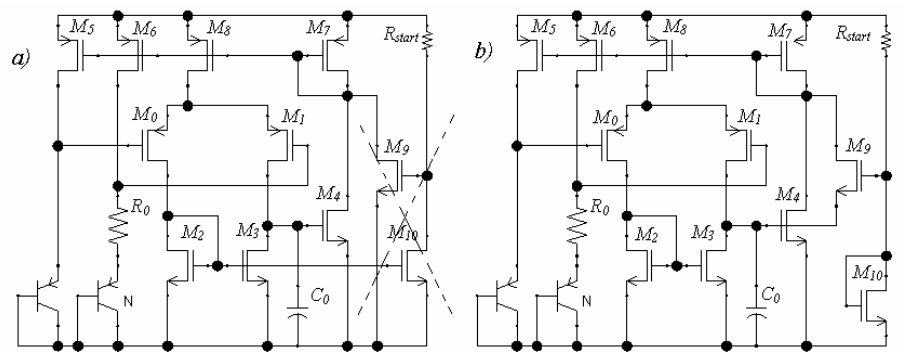


Figure 3-20. Start-up with disable

The start-up current can flow through the core constantly or it may be disconnected when the core is active. In the last case it may create a new feedback loop that can be conditionally unstable: the start-up current can

launch the biasing core, then, when the transient process is over and the start-up current is switched off, the core will shut down.

Hence, such start-up circuit may create a multiloop system (the core loop plus start-up loop). From the practical point of view, the sufficient condition of stability in the multiloop system is stability of each and every loop. It is usually sufficient if the start-up feedback loop is stable by itself without regarding that it is OFF during normal operation. This stability is easier to provide if the interaction between the start-up loop and the rest of the system is minimal. Fig. 3-20a shows a popular yet problematic start-up circuit (for explanation see subchapter 4.5). The interaction of the start-up loop with existing loops is reduced in the circuit of fig. 3-20b, and this circuit is stable.

Chapter 4

OPAMP GAIN STRUCTURE, FREQUENCY COMPENSATION AND STABILITY

OpAmp circuit development starts from the definition of a structure which provides the required gain and speed. Some decisions on the circuit topology may be predetermined. For example, the industry standard of the rail-to-rail output stage means that the output transistors should operate in the common-source (common-emitter) configuration. The industry standard of the differential input with large input impedance invariably means the differential input pair. The size of the output devices and the quiescent current of the output stage are defined by the maximum load current and by the required tolerance to a capacitive load (subchapter 4.5). The size of the input devices and the input stage tail current are defined by the noise requirements.

In structural design, frequency compensation is an integral part of the circuit design. Almost always it should be attained via topology means, by use of dedicated feedback loops and feedforward links. For any given process, the usual price for larger speed and high-frequency gain is larger operating current and dissipated power. Compensation capacitors decrease these, already paid for, speed and high frequency gain and should only be considered as a last resort, when all other compensation methods are exhausted. For example, the OpAmp described in [44] comprises 18 feedback loops and has only one compensation capacitor.

Quiescent current to speed ratio is one of the most important figures of merit of design quality, provided that the other goals have been met. For this figure, an OpAmp gain structure is very important because it defines the boundary of achievable.

Generally, each transistor in the signal path not only amplifies the voltage or current or both, but also introduces a signal delay. It means that generally number of stages should be as small as possible. But the system

can have a large number of poles and still be dynamically equivalent to a second-order system if the frequencies of these poles are significantly larger (at least twice) than the system bandwidth. Delays from different stages make different contributions in the total delay. Delays of some stages may be disregarded.

For the gain structure design it is useful to create a new definition, and to distinguish amplification stages and gain stages. We define *gain stage as the current and/or voltage amplification stage that introduces a pole within the OpAmp bandwidth*. Addition of a gain stage in the signal path complicates the overall OpAmp frequency compensation; addition of an amplification, but not gain, stage does not affect the OpAmp dynamics. This differentiation allows one to synthesize a set of dynamically equivalent circuits without detailed analysis of each of them. The proposed definition also underlines using of amplification stages in the OpAmp design.

Gain and amplification stages

Let us consider how the elementary stages may be assessed in accordance with this new classification. We assume that the requirements for the OpAmp bandwidth are known.

A transistor can be used in one of three configurations: common-source, common-gate and common-drain for MOS or, equivalently, common-emitter, common-base and common-collector for bipolar transistors.

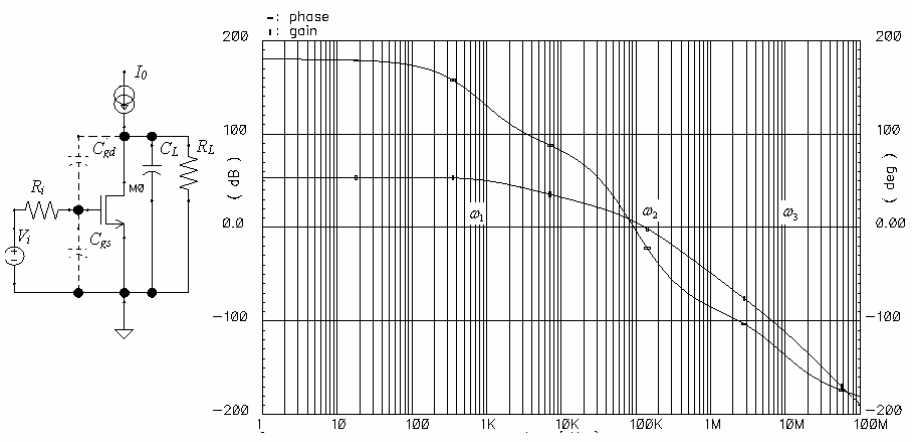


Figure 4-1. Transistor in the common-source configuration

A common-source transistor in the signal path provides both voltage and current amplification, and is always a gain stage (fig. 4-1) with the main pole frequency of $\omega_1 = 1/R_L C_L$, where R_L and C_L are the equivalent resistance and capacitance at the drain node. It also has a signal feedforward through

C_{gd} causing 180^0 additional drop in the phase at the frequency $\omega_2 = g_m / C_L$ where the voltage gain drops to unit. The resistance of the signal source creates one more pole ($\omega_3 = 1/R_i C_{gs}$). If both the source impedance and load capacitance are large then the common-source transistor creates more than one pole in the OpAmp transfer function, and should be counted as more then one gain stage for a fast assessment of stability.

The faster drop of the phase at higher frequencies, while still preserving gain, due to the feed forward through C_{gd} makes this stage a non-minimal phase stage. This fact makes the phase margin on the Bode plots a questionable stability merit, at least for the OpAmps comprising common-source stages.

Transistor in the common-gate configuration (fig. 4-2) provides only voltage amplification, and whether it may or may not be considered as a gain stage depends on the signal source impedance. In the first approximation, if $R_i >> 1/g_m$ then it is not a gain stage, if $R_i << 1/g_m$ then it is a gain stage.

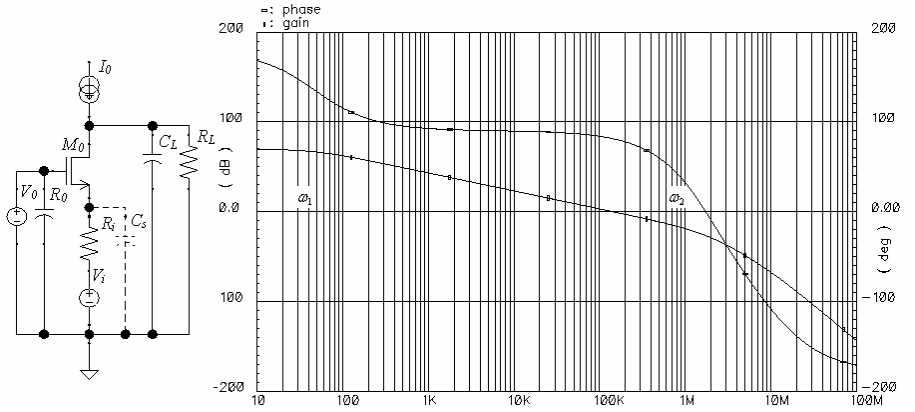


Figure 4-2. Transistor in the common-gate configuration

The stage with transistor in the common-gate configuration has a pole at $\omega_1 = 1/R_L C_L$. With large resistance of the signal source this pole would be at the same frequency even without transistor. This pole should be disregarded when assessing the number of gain stages. But if the common-gate transistor stage has another pole at $\omega_2 = g_m / C_s$, where C_s is the equivalent capacitance at the source, and this pole is located within the OpAmp bandwidth, then this pole returns this transistor back to the count of the gain stages. To avoid this, both C_s and g_m may be controlled by transistor sizing, by biasing current or even by an additional local feedback loop.

The capacitor C_0 in fig. 4-2 emphasizes the fact that all real voltage sources have internal resistance that can significantly degrade dynamics of

the common-gate transistor. Placing of such capacitor before the start of simulations should be a routine practice for every common-gate transistor in the signal path. This capacitor can be removed later if the simulations show that it has no influence on the amplifier performance.

The transistor in the common-drain configuration (fig. 4-3) has a pole at $\omega_1 \approx g_m / C_L$ and is not considered as a gain stage when C_L is small. However, if C_L is large then such transistor should also be considered as a gain stage. Again, if one controls g_m by transistor sizing, biasing current or by a local feedback loop (subchapter 4.3), these measures may eliminate this stage from the "black list" of the gain stages.

If the signal source resistance is large then the common-drain stage creates an additional pole with the frequency of $\omega_2 = 1/R_i C_{gd}$ and is considered as a gain stage.

The current mirror stage (fig. 4-4a) is similar dynamically to the common-source configuration with $R_i = 1/g_m$. Hence, it should be counted as a gain stage. The differential pair in the non-inverting configuration (fig. 4-4b) comprises a common-drain transistor (not a gain stage) and a common-gate transistor with signal source resistance $R_i = 1/g_m$. This makes it dynamically a gain stage.

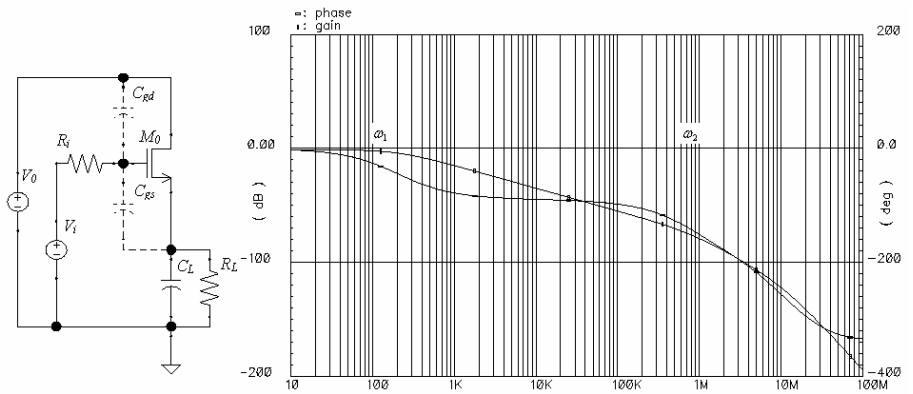


Figure 4-3. Transistor in the common-drain configuration

In summary, transistors in common-source configuration, current mirrors and differential stages in the signal path create poles in the OpAmp transfer function, and should be considered gain stages. The common-drain stage provides large current amplification without meaningful delay if the load impedance is larger than the transistor source impedance. The common-gate

stage provides voltage amplification with no delay if the signal source impedance is larger than the transistor source impedance.

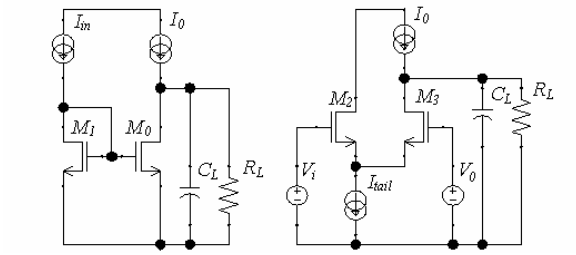


Figure 4-4. Current mirror and differential pair

These considerations are very useful for a single-glance estimation of the loop stability, and anticipation of the possible problems while creating set of circuit solutions. The exact calculations of the time constants and transfer functions do not make sense in practice as simulation results are always more accurate and faster. Also, the circuits without robustness, at the edge of oscillations, belong to the realm of bad circuits, devaluating the effort of exact stability calculations.

How many gain and amplification stages does an OpAmp need?

By industry standards, an OpAmp should provide at least 90 dB of voltage gain (120 dB for precision OpAmps), as well as 120 dB of current gain.

A single MOS or bipolar transistor in a common-source or common-gate configuration can provide about 40 dB of voltage gain. This means that an OpAmp should have at least 3 voltage amplification stages. One of these amplification stages has to be a transistor in the common-source configuration to provide the signal inversion (one of the OpAmp inputs is an inverting input), and must, therefore, to be a gain stage. Voltage input OpAmps with the rail-to-rail output comprises at least two common-source transistors in the signal path: the input differential stage, and the output devices. This means that there are at least two gain stages in these OpAmps.

The current gain of a MOS transistor in the common-source or common-drain configuration is infinity and the problem with current gain in CMOS OpAmps never arises. The current gain of a bipolar transistor (β) can be as low as 20 at low temperatures, so a bipolar OpAmp should comprise at least 5 current amplification stages or employ MOS or FET components, if available, in the signal path. Fortunately, none of these stages need to be a gain stage.

Fig. 4.5 shows three amplifiers consisting of one, two and three identical gain stages g_m with load (*R, C*) and compensation circuits.

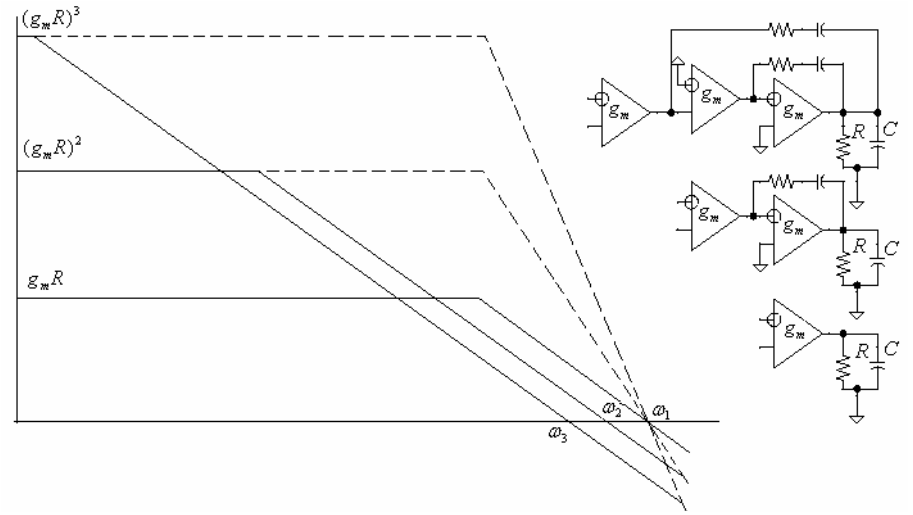


Figure 4-5. Bandwidth versus number of gain stages

The single gain stage amplifier does not need compensation and has unity-gain bandwidth ω_1. The two-stage amplifier needs a Miller or other compensation circuit, and, in order to achieve 70 degrees phase margin, its bandwidth should be limited to $\omega_2 = \omega_1 / 3$ [45, subchapter 4.4, 22, chapter 5]. In the three-stage amplifier the achievable bandwidth is limited to $\omega_3 = \omega_1 / 9$ in order to retain a reasonable phase margin.

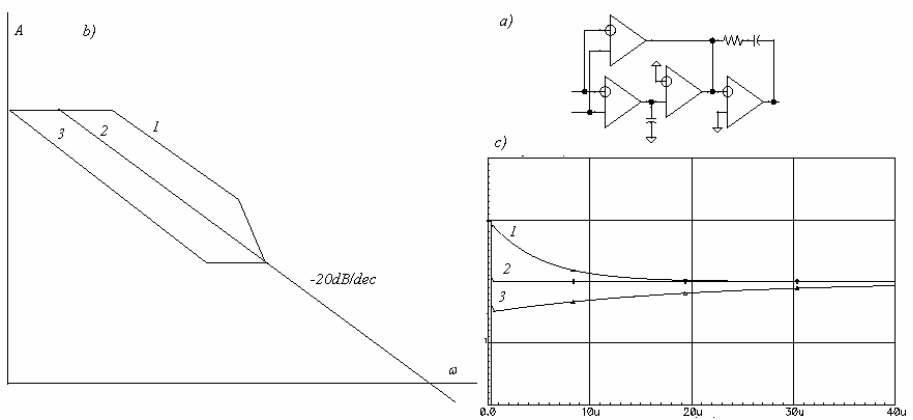


Figure 4-6. Step response of the OpAmp with pole-zero doublets

These relationships are very approximate, but in practice a bandwidth decrease by 3 times for each additional gain stage in the OpAmp signal path is optimistic.

One method used to combine a large number of gain stages and to obtain a wide bandwidth is to use a parallel feedforward links [26, 27, 46]. The main problem when using parallel links is matching of the poles and zeros of the branches to create a smooth -20 dB/dec gain versus frequency dependence without pole-zero doublets (fig. 4-6b, curve 2). If this is not achieved, and the dependence looks like that of fig. 4-6b curve 1 or 3, then the step response of the OpAmp may have a long settling (fig. 4-6c). The evaluation of settling time and dynamic errors due to these doublets can be found in [47].

The output stage of an OpAmp usually takes the lion's share of the current budget and has the largest transconductance in order to reduce the distortions and increase the OpAmp tolerance to capacitive loads. It rules out all parallel structures with the feedforward link to the output. To be efficient, such link should have the same or larger transconductance as the bypassed link, and this means the larger consumed current.

High-speed OpAmps and OpAmps optimized for the power to speed ratio should have only a single gain stage in the signal path while achieving the required voltage and current gain by increasing the number of the amplification, but not gain, stages. Examples of such an approach are the current-input OpAmps for video applications [48] or Norton-type OpAmps [107].

Voltage-input OpAmps with rail-to-rail output do not have less than two gain stages. In most applications this should be the limit while achieving the desired voltage and current gain with amplification stages and gain boost circuits (subchapter 4.3).

An exception from this rule are the OpAmps for professional audio: it can be shown that, providing that the non-linearity of the output stage is fixed, an OpAmp with three gain stages may have slightly smaller distortions in the range of medium frequencies (300Hz-20 kHz) than a two-stage OpAmp with the same current budget and with three times wider bandwidth.

4.1 Voltage and current gain boost

Voltage gain boost is the generalization of the cascode approach to increase the output impedance of transistor. In terms of the structural design, it is the use of active feedback loop to reduce the influence of parasitic feedbacks which are affecting voltage or current gain in the amplification stage. The block diagram of a popular voltage gain boost circuit [38], and its signal flow graph are shown in fig. 4-7.

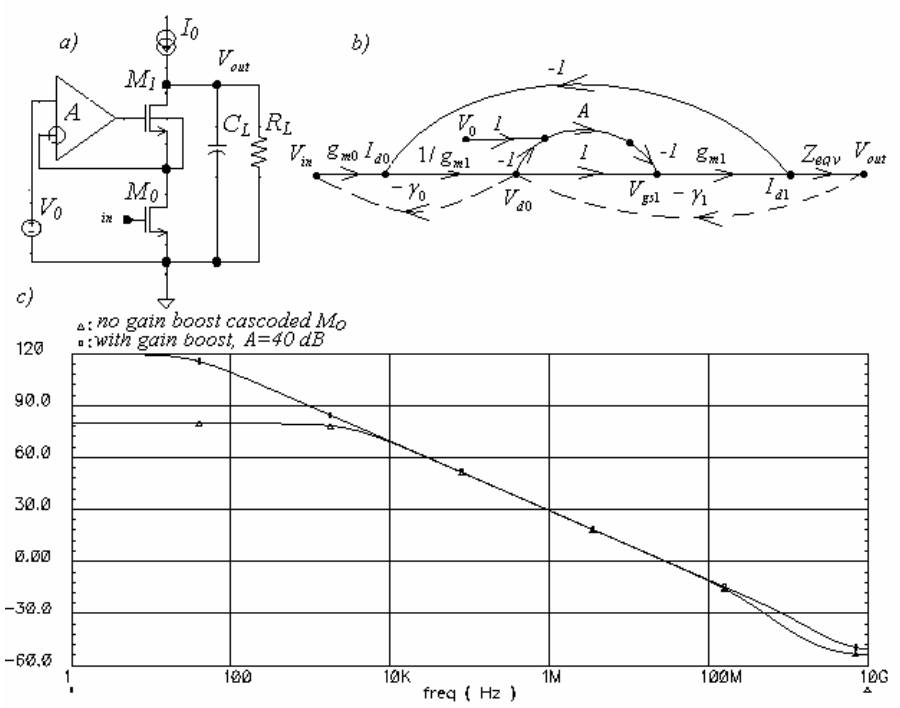


Figure 4-7. Voltage gain boost

The sequence of horizontal branches in this graph represents the transmission of the input signal to the output in the cascode connection of M_0 and M_1. To boost the output impedance of M1, it is placed in the negative feedback loop in conjunction with amplifier *A*. If *A* is very high, then any AC finite value signal appearing *after* the amplifier requires that the AC signal at the amplifier input is nearly zero (we will write, for simplicity, simply zero). This requires that the AC component of V_{d0} be zero, and that the AC component of I_{d0} also be zero. Hence, the AC signal I_{d1} is $I_{d1} = g_{m0}V_{in}$, and the AC component of V_{out} is $V_{out} = g_{m0}Z_{eqv}V_{in}$. The voltage gain is $g_{m0}Z_{eqv}$. Here Z_{eqv} is the equivalent impedance at the node *out*, and it consists of the parallel connection of load, R_L, inner impedance of the current source I_0, and drain impedance of M_1. If the load is the gate of the next amplification stage, and the current source is designed as suggested in chapter 3, and has very high inner impedance, then Z_{eqv} is dominated by the drain impedance of M_1.

Transistor M_0 has a parasitic feedback link γ_0. Yet, because the AC component of V_{d0} is zero, this link is not operable. Transistor M_1 also has an

internal feedback γ_1 (finite output impedance, short-channel effects, etc.) that affects its V_{gs1} as the drain voltage changes. Again, as V_{d0} is zero, this parasitic link is suppressed, and $V_{gs1} = I_{d1} / g_{m1}$ as long as $A(s)$ is high.

Fig. 4-7c shows the gain/frequency response of the stage without and with the gain boost. The last dependence has a smooth -20dB/dec slope. No doublets are present as long as the load capacitance is larger than the sum of the parasitic capacitances of M_1 and I_0. The circuit not only improves DC and low-frequency gain (by the value of A), but also the high-frequency gain of the stage: the gain in the A-M_1 loop decreases the voltage V_{d0}, and, hence, the impedance seen at the source of M_1. This impedance is the load for M_0, and a smaller load means a smaller time constant of M_0 associated with C_{gd} (suppressed Miller effect).

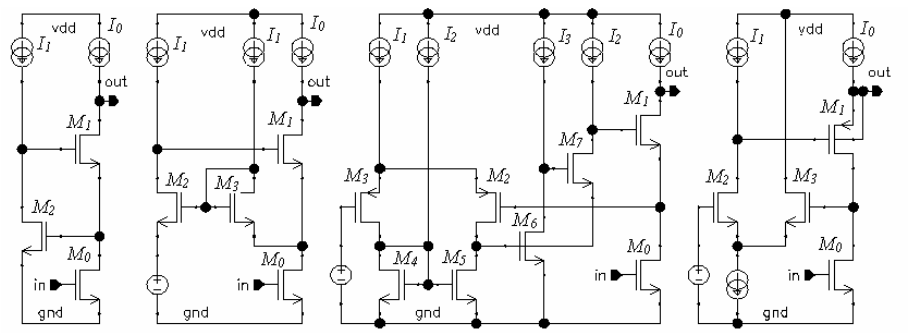


Figure 4-8. Implementations of the voltage gain boost

Some implementations of the circuit of 4-7a are shown in fig. 4-8. The circuit of fig 4-8a is the simplest one, with the amplifier A implemented by a single transistor M_2. The minimum output voltage for this circuit is equal to $(V_{gs}+V_{sat})$.

The circuit of fig. 4-8b employs the current-input amplifier M_2/M_3. This circuit can have an output voltage as small as $2V_{sat}$.

Fig. 4-8c shows the "gain boosted" gain boost amplifier comprising a differential pair M_2/M_3 with current mirror M_4/M_5 and gain stage M_6 boosted by the cascode transistor M_7. Such circuit can achieve 120 dB or more of voltage amplification.

Unless the body of M_1 can be connected to the source, the feedback loop through A does not affect the M_1 drain to body leakage (impact ionization current). This leakage can significantly reduce Z_{eqv}, especially at high temperatures and large drain voltages. One of the solutions to eliminate this effect is shown in fig. 4-8d with the cascoding transistor M_1 of opposite type [35].

Current gain boost is, in structural design terms, the use of an active feedback loop to suppress parasitic feedbacks in the amplification stage in order to decrease the stage output conductance. It is less known but no less useful, especially in the case of bipolar amplifiers because it also boosts the stage's input impedance.

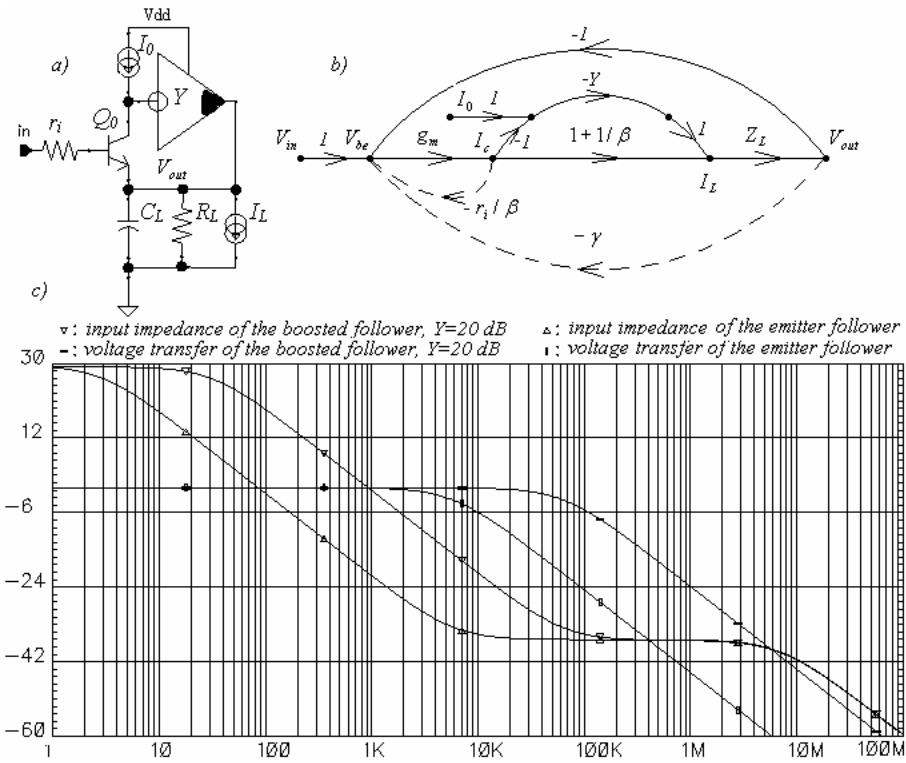


Figure 4-9. Current gain boost

The circuit block diagram and the signal graph for current boost in a common-collector amplification stage are shown in fig. 4.9. The series of the horizontal branches represent the transmission of the input signal in an ordinary emitter follower. Introducing the negative feedback loop with the high-gain amplifier $-Y$, one achieves that AC component of I_C of Q_0 is zero.

As one can see from the characteristics of fig. 4-9c, the amplifier improves the output bandwidth in the case with a capacitive load and the input impedance of the emitter follower by *Y*.

Several implementations of the structure given in fig. 4-9a are shown in fig. 4-10.

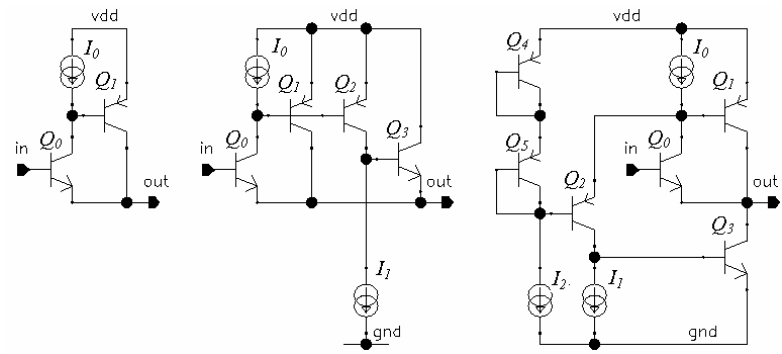


Figure 4-10. Current gain boost implementations

The circuit of fig. 4-10a is the simplest implementation with the amplifier *Y* implemented as a single transistor Q_1. In the circuit of fig. 4-10b the amplifier *Y* comprises an additional current boost amplifier Q_2/Q_3 for larger gain. The circuit of fig. 4-10c is a class AB boosted follower that can both sink and source current to the load.

All of these circuits can be redesigned with CMOS transistors.

The circuit of fig. 4-9a does not suppress the parasitic feedback, γ ,caused by variation of the collector-base voltage (base width modulation). This limits the input impedance regardless how large the gain *Y* is. As shown in fig. 4-9c, the input impedance levels at low frequencies have the same values for both boosted and nonboosted followers.

This deficiency can be fixed if the voltage bias input of the current amplifier *Y* is connected to the follower input as shown in fig. 4-11. In this case the collector voltage of Q_0 will follow its base voltage, so no internal feedback occurs. As show the normalized simulation results (fig. 4-11b) the input resistance characteristic does not level off at low frequencies as in fig. 4-9.

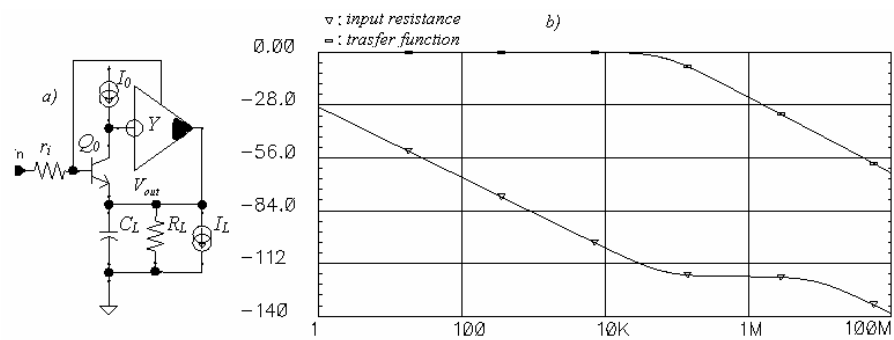


Figure 4-11. Current boost without limitation of the input resistance

4.2 Frequency compensation

Frequency compensation in structural design is integrated in the design process starting from the development of the gain structure. Stability is provided for by use of the single-stage amplifiers for local feedback loops, as well as by extensive use of the gain boost techniques. It allows to avoid or to minimize the need in dedicated compensation capacitors while achieving any required gain.

Good circuits should be robust and have excessive stability. Historically, the main merit of the stability was the phase margin [22, p. 38]. AC parameters, including phase margin, are the fastest to obtain in simulations. But the phase margin can be too pessimistic when the circuit is not a minimal phase system, like, for example, an OpAmp with Miller compensation and a capacitive load. AC simulations also do not take into account non-linear effects, so the conditional instability may be overlooked. For modern workstations, the simulation time of transient process is not a restrictive factor, and the step overshoot ratio M_m may be a more reliable criterion of stability. For good systems M_m should be in the range from 1.1 to 1.5.

The compensation tasks can be separated into three groups:

- overall OpAmp compensation, with the goal of achieving a smooth -20 dB/dec slope of the open-loop gain up to twice the unity-gain frequency;
- compensation of internal feedback loops;
- prevention of conditional instability.

An OpAmp needs the constant -20 dB/dec slope of the gain/frequency dependence to avoid the settling time degradation (see fig. 4-6), and for stable operation with any closed-loop feedback resistive ratio that can be set by a customer. The overall OpAmp compensation is extensively considered in literature [9, 10, 13, 22, 45]. For a two-stage OpAmp there is only one practical approach, namely, the Miller compensation [22, pp. 76-94]. It has three versions:

a) compensation capacitor C_m in the second stage feedback (fig. 4-12a),
b) C_m driven by the output voltage follower (fig. 4-12b), and
c) the so-called cascoded Miller compensation [49, 50] (fig. 4-12c) comprising C_m and a current follower (it can be implemented as a common-gate transistor).

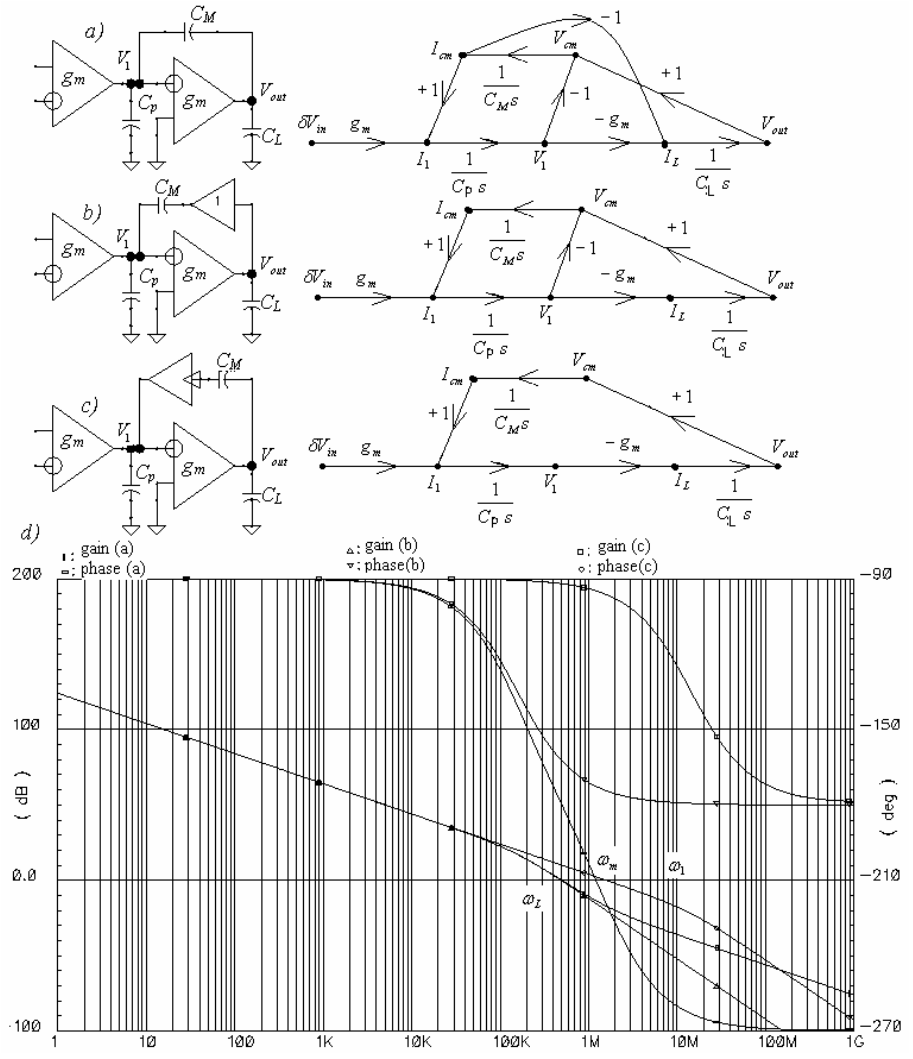


Figure 4-12. Two-stage amplifiers with Miller compensation

Simulation results using macromodels for each version of the Miller compensation are shown in fig. 4-12d. The simulated circuits include the OpAmps consisting of two identical transconductance stages with g_m=10^{-4} A/V, Miller capacitors C_M=10pF, parasitic input capacitances of the second stage C_p=0.1pF, and load capacitances C_L=100pF.

One can see that at low frequencies the open-loop gain/phase frequency dependencies are identical for all circuits. Circuits 4-12a and 4-12b have their first pole at $\omega_L = g_m / C_L$. The signal feedforward through C_M for the

circuit 4-12a starts at $\omega_M = \omega_L C_M / C_L$ causing the additional 90^o phase shift within the gain decrease slope of -20 dB/dec. Amplifier with cascoded Miller compensation of the circuit 4-12c has only one pole at $\omega_1 = g_m / C_p$.

With ideal macromodels, the Miller compensation versions 4-12b and 4-12c have clear advantages over the version 4-12a, providing less phase shift at high frequencies. One may note also that the versions 4-12b and 4-12c have simpler signal graphs as well. The version 4-12c has the simplest graph including only one loop, and the familiar method of stability evaluation by the phase margin is adequate only for this case.

These advantages may disappear in implementation. It is difficult to design a rail-to-rail voltage follower with low high-frequency output impedance and acceptable current budget for version 4-12b. It is equally difficult to design the current follower with the low high-frequency input impedance for the version 4-12c. Fig. 4-13 shows the results of the same simulations with more realistic 10 kOhm resistances for the followers.

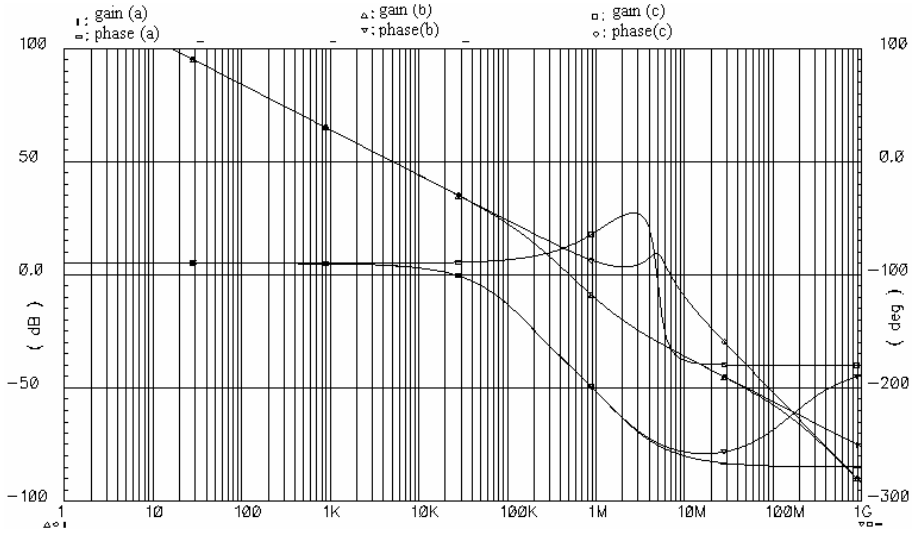


Figure 4-13. Miller compensation with non-ideal followers

As shown, the circuits of fig. 4-12a and 4-12b are now identical at the frequencies of interest, and the circuit of fig. 4-12c has sharp changes of the gain and phase characteristics. This indicates a possibility of oscillations in the circuit in the loop around the second stage.

The phase characteristic of the version of fig. 4-12a can be somewhat improved by adding a resistor in series with the capacitor C_M. This small modification makes this version the most popular solution for two-stage OpAmp compensation.

A compensated two-stage OpAmp has only one pole in the band of interest, and can be considered as a single-stage amplifier. This is the principle of nested Miller compensation for three-stage amplifiers [22]: last two stages are compensated with the Miller capacitor C_{M2}, and a larger Miller capacitor C_{M1} is connected around these two stages (fig. 4-14).

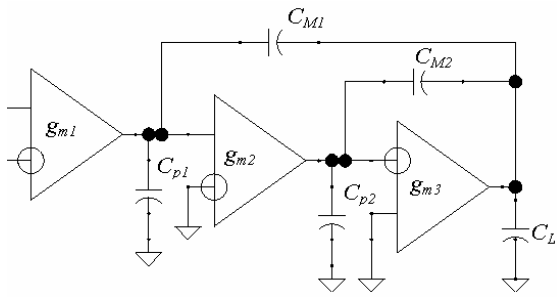


Figure 4-14. Nested Miller compensation of three-stage OpAmp

Both feedback loops of this compensation can also be realized by any of the versions shown in fig. 4-12.

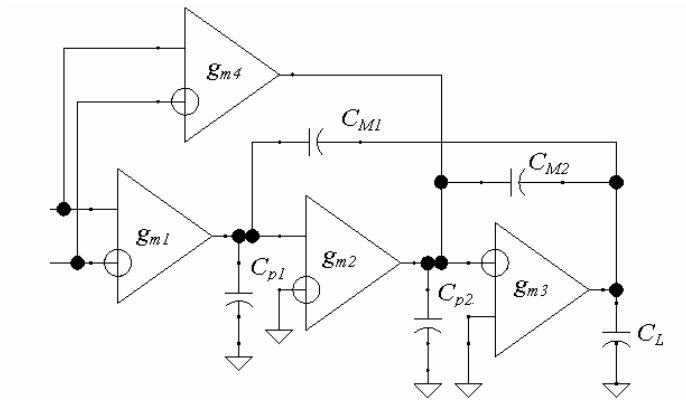


Figure 4-15. Multipath Miller compensation of the three-stage OpAmp

A three-stage amplifier can achieve a bandwidth close to that of a two-stage amplifier when a feedforward link bypassing the first two stages (fig. 4-15) is used (the so-called multipath Miller compensation [10, pp. 229-232]).

A smooth -20 dB/dec slope of the open-loop gain can be achieved if $g_{m1} / g_{m4} = C_{M2} / C_{M1}$.

Compensation of the internal amplifiers is different from the overall OpAmp compensation. Feedback configuration of internal amplifiers is known (usually they operate with unity feedback). A single-stage structure, not requiring dedicated compensation capacitors, is often acceptable when there is no need for rail-to-rail output or high input impedance. Examples of such circuits have been presented in figs. 3-3 b, 3-7 c, and 3-9.

Even with a two gain stage structure there is no necessity for -20 dB/dec slope of the open-loop gain at all frequencies, so the compensation of a two-stage amplifier with a feedforward link is acceptable. The principle of this feedforward compensation is shown in fig. 4-16.

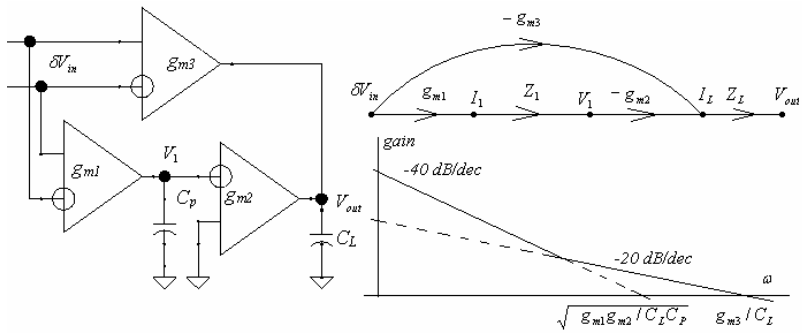


Figure 4-16. Feedforward compensation

Such an amplifier may be designed with sufficient excess stability if the transition point from the -40dB/dec to -20dB/dec slope is at least five times below the unity-gain frequency, i.e. if $g_{m3} \geq 2.5\sqrt{g_{m1}g_{m2}C_L/C_p}$. A large value of g_{m3} means more current in this stage, and if the high speed is not required then a simple Miller or parallel compensation with capacitor may become more attractive, as, for example, in the three-stage amplifier of the core biasing circuit shown in fig. 3-1.

Conditional instability

Conditional instability is probably the second, after ESD protection failures, cause of “bad silicon”, and is not predicted by AC simulations. Externally it appears as low frequency (few orders below unity-gain frequency) oscillations with large amplitude. These oscillations frequently occur in the biasing core and then propagate through the system. The tracing of their exact origin may be difficult.

Such oscillations can occur in the feedback loop comprising two or more gain stages if, for example, the output signal range of the previous stage is wider than the linear input range of the following stage. This situation is illustrated in fig. 4-17. In this three-stage feedback loop the second stage has

the input linear range smaller than the output voltage swing of the first stage. The small-signal frequency compensation is provided by sufficient length of the -20 dB/dec slope of the amplitude frequency response in the vicinity of the unity-gain frequency (fig. 4-17b). Assume that with these three gain stages the response has -60 dB/dec slope for low frequencies.

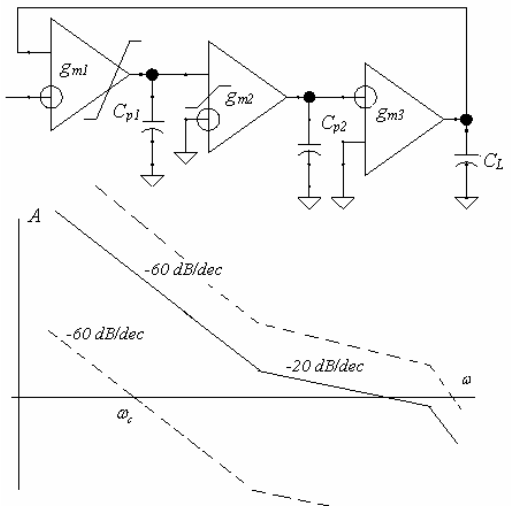


Figure 4-17. Conditional instability due to the limit mismatch

The output signal of the first stage may exceed the linear input range of the second stage during start-up or due to input overload. On AC plots it is equivalent to a shift of the amplitude frequency response up or down as shown by the dashed lines. With this new response, the amplifier is not stable any more as it has the gain slope of -60 dB/dec or more around a new unity-gain frequency. If, during the ensuing transient process, the output of the previous stage overshoots the linear range of the following stage in the opposite direction then the situation repeats, and the amplifier oscillates.

Such behavior may, for example, occur in the biasing circuit of fig. 3-1. During start-up, when a supply voltage step is applied, the voltage at the node *g4* of the stage M_0/M_1 can be close to zero (the compensation capacitor is not charged), and M_4 is off. Then, due to the delay in the feedback loop, the voltage V_{g4} can rise too high and increase the gain in the loop so that the loop becomes unstable (usually this is indicated by a very high current in the biasing core). Then the negative feedback with increased loop gain turns the transistor M_4 off, reducing the voltage V_{g4} to zero, and the oscillation starts. This oscillation is shown in fig. 4-18. The same circuit can be perfectly stable if started softly, by a slower increase of the power supply voltage.

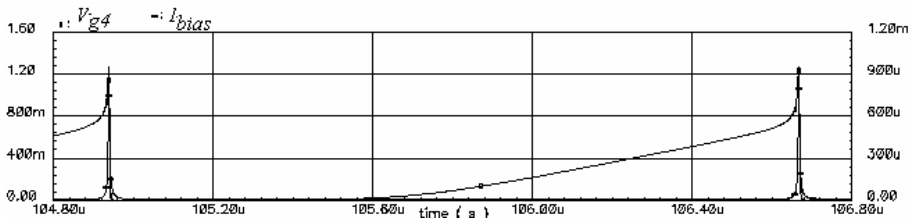


Figure 4-18. Conditional instability in the biasing cell of fig. 3.1

Suppression of this kind of oscillation can be done by the following methods:

- clamping the first stage output voltage so its signal limitation will better match the input linear range of the following stage (M_8 in fig. 4-19a);
- expanding the second stage linear range (for example, with source degeneration resistor R_1 in fig. 4-19b);
- expanding the frequency range of -20 dB/dec slope of open-loop frequency response (with a larger capacitor C_0, for example) so the circuit is stable with wider gain variations.

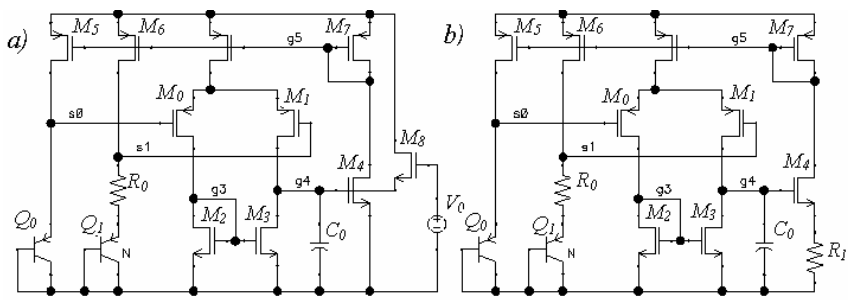


Figure 4-19. Compensation of the conditional instability

Another kind of conditional oscillations may occur due to interaction of the feedback loops if one of these loops is unstable. An example is interaction of the start-up and PTAT cell in fig. 3-20a. The shutdown transistor M_{14} of the start-up circuit creates the unstable feedback loop M_9-M_6-M_1-M_0-M_{10}-M_9. During normal operation this loop should be disabled. Yet, during the transient process, after the supply voltage step is applied, it may become active and cause low-frequency oscillations.

The sufficient (but not necessary) practical cure to avoid this kind of oscillations is stability of each feedback loop in the system [92, chapter 8].

One more example of the conditional instability problem is shown in fig. 4-20. In this circuit, the all of the PMOS devices of the biasing core are cascoded in order to improve PSRR. During the transient process, when the

currents of M_4, M_9 are low, the negative feedback loop M_{11}-M_7 is not operational and the equivalent impedance at node *g5* is high, causing an increase of the gain in the main feedback loop M_4-M_6-M_1-M_4 and, consequently, the oscillations in this loop.

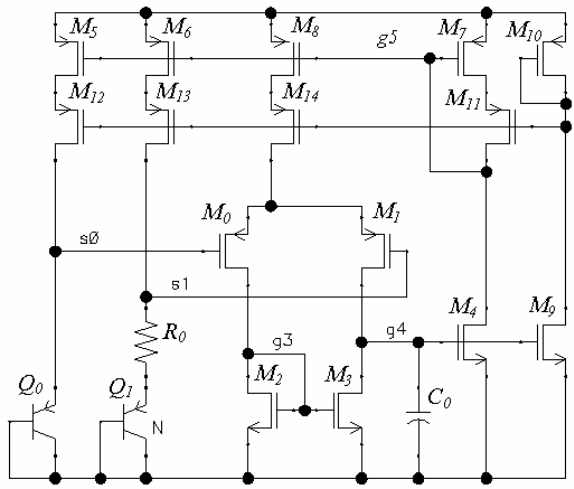


Figure 4-20. Conditional instability due to the interaction of different feedback loops

The cure for problem in this circuit is the elimination of the unwanted feedback loop, i.e. the deletion of M_{11}.

It is easy to fix the conditional instability problem when it has been detected and traced to its origin. But there is no universal recipe for detection of all unforeseen nonlinear effects. The meticulous simulations of circuit transients, in all possible modes of operation, may help to solve these problems. These simulations take a lion's share of the overall design time.

4.3 Rail-to-rail IO OpAmp structure

The OpAmps with rail-to-rail output started as low-voltage CMOS OpAmps in order to improve their signal/noise ratio. Now, this requirement has become an industry standard for almost all OpAmp types due to convenience in the applications. Rail-to-rail output stage requires the complementary output devices in the common-source (common-emitter) configuration.

The rail-to-rail input stages are not used in a majority of applications. They are, however, still desirable in order to avoid conditional instability effects in the customer's system, for example, in the case when common-mode signal unpredictably changes during start-up or input overload.

The most popular rail-to-rail input stage consists of NMOS/PMOS differential pairs, and of the tail current control circuit to keep transconductance of the stage relatively constant versus common-mode voltage [10, chapter 4]. Other practical circuits comprise only one differential pair with the bootstrapped tail current source [39], employ low-V_{th} transistors [51], or the depletion-mode transistors [52].

To achieve a good power to speed ratio it is preferable that the output transistors and input differential pair(s) are the only gain stages in OpAmp.

The general block diagram for this type of OpAmp is shown in fig. 4-21.

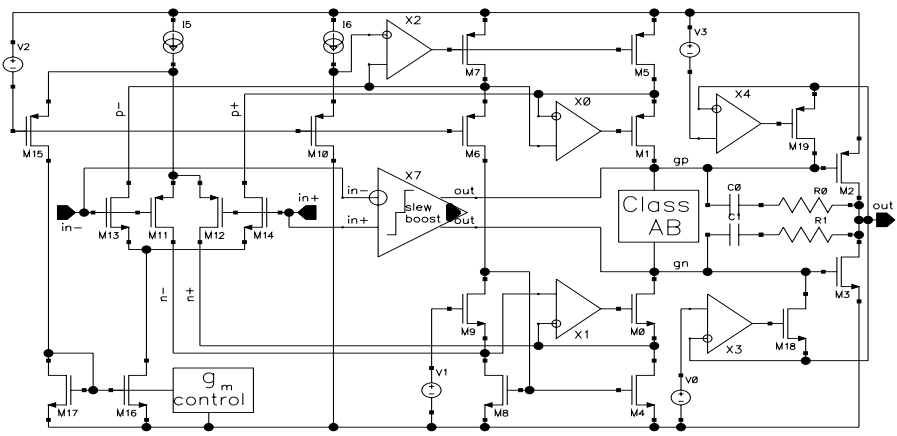


Figure 4-21. General block diagram of two-stage OpAmp with rail-to-rail operation

In addition to the input and output stages it comprises biasing, the link between input and output stages (usually done as shown, using folded cascode stages [53]), and the circuits for functional improvement: gain boost, slew rate boost, overload recovery time improvement, shutdown, trimming, etc.

As any two gain stage structure this OpAmp requires compensation. It can be done as shown with *RC* links between the gates and drains of output devices. To employ cascoded Miller compensation no dedicated circuitry is needed and the capacitors can also be connected between the output and the inputs of the folded cascode (the sources of M_0 and M_1). Due to a noticeable source resistance (equal to $1/g_m$) of the folded cascode transistors the pure cascoded Miller compensation is not feasible and may cause oscillations (see the previous part). Combining the cascoded and regular Miller capacitors may provide compensation with the best capacitive load capability for any given quiescent current I_q of the output stage [54].

The overall area of the input devices of an OpAmp is defined by the offset and flicker noise requirements (subchapter 1.5.2). The noise of the following stage reflects to the input-referred noise by being divided by the gain of the input stage. It means that the transconductance of the input stage should be maximized. It results in wide and short channel transistors of the input stage, operating in weak inversion. The only exclusions to this rule are the high-frequency OpAmps for video applications where the low-frequency noise and offset are less important than the speed so the input devices can be smaller and operate in strong inversion.

The contribution of the mirror transistors of the folded cascode (M_4, M_5, M_7, M_8) to the OpAmp noise and offset should be comparable to or less than that of input devices. To achieve this, the overall area of these transistors may be comparable with the area of input devices but their transconductance should be minimized. This g_m decrease can be achieved by using resistors between the sources and the power bus (to be efficient, at least a 100 mV voltage drop across these resistors is necessary) and by choice of the larger channel length. These transistors almost always operate in strong inversion.

The noise contribution of M_4, M_5, M_7, M_8 also increases with current. If a slew rate boost circuit is present, the current of the folded cascode (I_D of M_6) should be 2-3 times smaller than the input stage tail current, to make this noise contribution negligible. Without the slew rate boost this current may be equal to the tail current in order to have a symmetric slew rate behavior.

The source resistance and equivalent capacitances at the inputs of the folded cascode (sources of M_1, M_0) create an additional pole (g_{m0}/C_{eqv0} or g_{m1}/C_{eqv1}) in the signal path. To keep these poles from affecting the OpAmp stability, the transconductance of M_1, M_0 has to be maximized. The current of M_1, M_0 is fixed to the folded cascode current, and is chosen on the basis of the slew rate or noise considerations. The maximization of the transconductance of M_1 and M_0 is achieved using minimum channel length and larger width, which results in the weak inversion operation mode of M_1 and M_0.

The size of output transistors is chosen considering the worst-case available gate drive voltage and the OpAmp load drive requirement. This consideration usually results in very wide devices. When there is no load, then these transistors conduct only the output stage quiescent current; that almost always forces them to operate in weak inversion. The capacitances C_{gd} of the output devices may be so large that they can replace the dedicated Miller capacitors and no other compensation is necessary [44].

The capacitance C_{gs} of these devices creates another pole in the OpAmp transfer function. This pole is usually the main limiting factor of the maximum achievable OpAmp bandwidth. The output devices should have the minimum channel length allowed in the process in order to be smaller and decrease C_{gs}.

The block diagram of a bipolar OpAmp with rail-to-rail output is shown in fig. 4.22. It differs from the CMOS OpAmp by:

- addition of the current gain boost A_3,
- addition of the circuit to prevent the output phase inversion,
- addition of the input overvoltage protection circuit (subchapter 5.4).

Complementary rail-to-rail input stages are not used in industrial bipolar OpAmps. The base current flow changes direction at the switching point between PNP/NPN input pairs, which creates problems in applications.

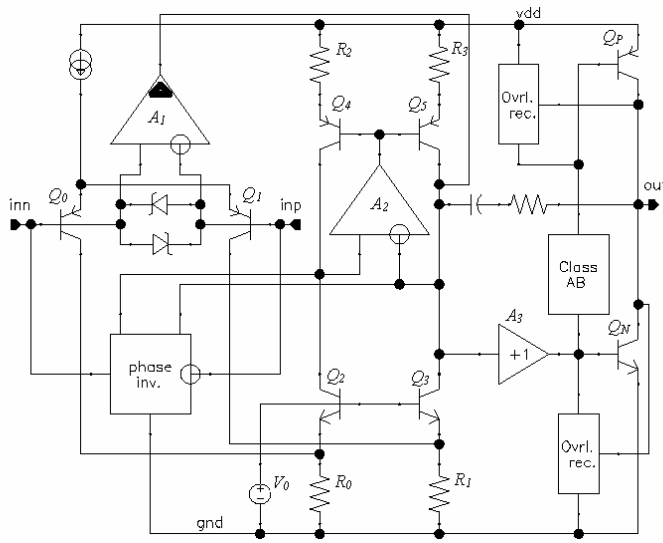


Figure 4-22. General block diagram of the bipolar OpAmp

The current gain of bipolar transistors is limited and can be as low as 20, especially at negative temperatures. In order to sink or source 10 mA of load current, while having only a few nanoamperes of the input current, the bipolar OpAmp should comprise four or five current amplification stages. With two-stage voltage gain structure this can be achieved using the follower (current gain boost circuit A_3). This follower may be implemented as shown in fig. 4-10 or by using FET or MOS transistors when available.

Voltage gain boost can be implemented as before in CMOS OpAmp structure of fig. 4-21 or as shown in fig. 4-22, by the current mirror Q_4/Q_5 and amplifier A_2. This principle will be discussed in more details and variants in chapter 5.

The input overload protection is usually required because an inverse voltage across the base-emitter junction of integrated bipolar transistor is often limited between 2 to 5 V.

When the input common-mode voltage drops below the negative rail, the collector-base junction of input transistors starts to conduct. This results in the inversion of the amplifier gain. To avoid the undesirable changes at the output, the phase inversion protection clamps have to be added.

The overload recovery units in the bipolar amplifiers are not aimed to limit the base voltage of the output devices. It does not exceed 0.6-0.7 V by the nature of bipolar transistor. The main goal is to prevent deep saturation of the output devices resulting in significant recovery time.

The OpAmps of fig. 4-21 and 4-22 have the structure with only two gain stages. There are multiple examples in literature and in industry of OpAmps with three or more gain stage. However, one can show that such OpAmps are inferior in power to speed ratio, and practically in any other parameter, to the OpAmp with two gain stages designed for the same process and the same quiescent current. These multistage OpAmps were created on the assumption that additional gain stages are necessary to increase the open-loop gain. In fact, with proper gain boost as discussed above and will be discussed further in chapter 5, any amount of DC voltage and current amplification can be achieved with just one (in dynamic sense) gain stage.

Chapter 5

INPUT STAGE

The input stage determines most of the noise and accuracy properties of an OpAmp and, to some degree, its bandwidth and slew rate.

An input stages with differential input and output can be classified into 16 types depending on the stage's input and output differential and common mode impedances in relation to the signal source and the load. The output of an input stage is usually the drain or collector of a transistor. Since the transistor drain/collector has high output impedance, the practical number is limited to 4 types of the current-output (with high output impedances) stages: 1) high differential and common-mode input impedance; 2) high differential and low common-mode; 3) low differential and high common-mode; 4) low both differential and common-mode impedance.

Voltage input OpAmps, with high differential and common-mode input impedances, became the industry mainstream after a few unsuccessful attempts to introduce the so-called Norton amplifiers (with low input impedances) in 1970-s and 80-s [107]. OpAmps with voltage input usually use differential transistor pairs (fig. 2-3f) in their input stages.

Due to the speed advantages of current signal processing, the stages with high common-mode and low differential input impedances are often used in very fast OpAmps for video applications. Design of these OpAmps usually involves common-gate configurations (fig. 2-4) in the input stages.

The input stages of the OpAmps in systems on chip are not defined by customer preference, and one could expect to meet a larger variety of circuit solutions. Nevertheless, tradition and absence of the application guides usually limits them to the same two types.

The main parameters for small-signal operation of the voltage input stage are the transconductance, output resistance and parasitic output capacitance. For large-signal operation the important characteristics are dependence of transconductance on the differential input voltage, the output current limits,

the differential and common-mode input voltage ranges, and some special properties, if any, such as signal transfer inversion in the bipolar stage or the MOS gate oxide breakdown when the input signals exceed specified ranges.

The OpAmp error parameters (defined mostly by the input stage) include offset voltage, temperature offset voltage drift, common-mode and power supply rejection ratios, flicker and high-frequency noise, and input bias currents.

In CMOS OpAmps both the offset voltage and flicker noise are inversely proportional to the square root of the overall area of the input devices [108, p. 236, 465]. As a result, the input devices are usually the largest one. These large input transistors have significant parasitic drain capacitance that, in combination with the input resistance of the next, folded cascode, stage, create an additional pole in the OpAmp transfer characteristic. This pole may convert the folded cascode stage into a gain stage and limit the overall OpAmp bandwidth.

Flicker noise, also called 1/f noise, defined mostly by the surface effects in transistors. Since MOS transistors operate along the surface of the silicon, flicker noise can be a significant degrading factor in OpAmp DC performance. The process steps that heal the lattice damage close to the surface may reduce this noise. With large input devices, the tail current of the stage defines, almost exclusively, the high-frequency (thermal) noise. This makes the tail current the second largest, after the output stage, contributor in the overall current budget, being between 10 to 40% of the OpAmp quiescent current, I_q.

Noise optimization causes the operation of the MOS input differential pair in weak inversion. This is valid for the most industrial OpAmps with only one exception known to the authors [105].

The offset voltage temperature drift of bipolar differential pair is proportional to the offset value [9, p. 234], so trimming of the offset reduces the temperature drift as well. Offset and offset temperature drift for MOS differential pairs are not correlated, and the MOS offset drift can be as high as 20 uV/^{o}C regardless how small is the offset. In a correctly designed OpAmp (if the transistor operation mode is preserved with temperature) the temperature drift is constant up to 110-120 ^{o}C. A noticeable curvature caused by the junction leakage currents may be observed above this limit. These currents are proportional to T^4 (subchapter 1.5), and their effect is negligible at low temperatures.

The offset and offset drift both vary when mechanical stress is applied to the chip. Special die coating steps may decrease this stress. The stress sensitivity of V_{be} or V_{gs} voltages is proportional to their charge carrier mobility [3], making PMOS or PNP transistors with symmetric topology approximately 3 times less sensitive to the mechanical stress then NMOS or NPN ones.

Time variation (ageing) of the offset voltage is caused mainly by the time-dependent mechanical pressure on the chip, and, to a smaller degree, by the never-ending process of dopant diffusion, continuing at room temperature as well.

Differential pair offset is caused by the geometrical or diffusion mismatches between transistors. It varies with the change of each of the electrical parameters defining V_{gs} (or V_{be}): tail current, V_{ds} (or V_{ce}) voltage, and V_{bs} voltage in MOS transistors. In order to improve the OpAmp CMRR and PSRR, the input pairs should operate in "greenhouse" conditions. They should have a constant tail current that is independent on the common-mode input or supply voltage, a constant V_{bs}, and a stable V_{ds} or V_{ce} voltage (subchapter 5.2).

The existing auto-zero and chopping techniques and their combinations allow one to eliminate or strongly reduce the offset and flicker noise [89]. These techniques improve the OpAmp DC parameters with the drawback of complicated time-domain behavior. This behavior can be masked, and it is difficult to detect its peculiarities using standard frequency domain measurements [60]. In the absence of clear understanding or application guides, it is limiting the number of applications.

The input bias current of a bipolar differential pair is defined by the value of the tail current and the transistor current gain. The lowest value of the tail current is dictated by the noise requirements. Other means to decrease the bias current are base current cancellation and use of the super-β transistors.

The current gain, β, is the poorest matching parameter of a transistor, and base current cancellation allows one to reduce the bias current no more than 10-20 times. Base current cancellation circuitry increases the input current noise and current offset, which can be detrimental for applications with high driving impedance, i.e. exactly where one would want to use this technique. Another difficulty is that the common-mode input voltage of an OpAmp with base current cancellation is not able to reach the power supply rail.

Super-β lateral transistors are readily available in any fine-line CMOS process. For example, current gain of lateral PNP transistors with 0.35 um base width exceeds 500. For bipolar transistors the Early voltage is inversely proportional to the current gain so cascoding of the super-β input pair may be necessary [90].

The input bias current of CMOS OpAmps is defined mostly by the leakage of the ESD protection devices. The simplest devices of this type are large diodes, connecting the input pin to the substrate and/or to V_{DD} rail. Their leakage is strongly dependent on temperature and common-mode input voltage and can be as high as 5-10 nA at 125^{o}C.

Input bias currents in the femto-amp range can be achieved using JFET input transistors in a silicon-on-insulator process [55].

5.1 Rail-to-rail input stages with stable gm

The small-signal bandwidth of an OpAmp with Miller compensation, C_M, and input stage transconductance g_m is $GBW = g_m / C_M$, and is limited by the high frequency poles originating from the parasitic capacitors of the output devices, folded cascode stage input resistance, etc. The process variations of C_M can be as high as 10%. The variation of g_m should be in the same range or less to keep the bandwidth spread within approximately same limits. With g_m-matching biasing (subchapter 3.2) the variation of input stage transconductance is defined mostly by the spread of the biasing core (fig. 3.5) resistor values. The process variation of these resistors is usually also within 10%. Sometimes this resistor is being trimmed at the probe or final test to improve the repeatability of I_q and the OpAmp dynamic properties.

The bandwidth stability versus temperature is defined by the stability of g_m. The problem of g_m stability in the rail-to-rail complementary input stages has had a lot of attention in recent literature [10, pp. 102-119]. To optimize for noise, the input differential pairs in most of the industrial CMOS OpAmps operate in weak inversion. The transconductance g_m of MOS transistor in weak inversion is proportional to their drain current, so the overall transconductance of the NMOS/PMOS input stage is constant if the sum of the tail currents is kept constant in the full range of the common-mode input voltage.

A circuit for switching of the current from the PMOS to the NMOS input pair when the common-mode input voltage approaches V_{DD} is shown in fig. 5-1.

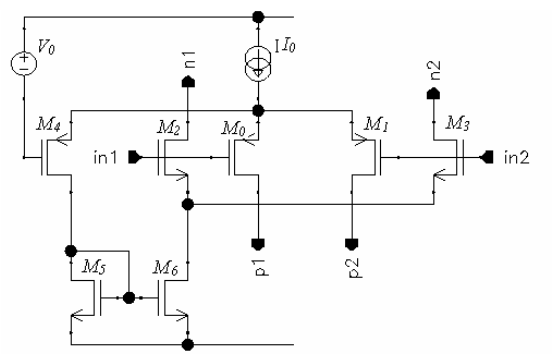


Figure 5-1. Tail current switching between PMOS and NMOS input pairs

Equal g_m of the PMOS and NMOS differential pairs when running the full tail current can be achieved by scaling of the current mirror M_5/M_6. The voltage $V_0 \geq V_{gs} + V_{sat}$ defines the switching point. Often this point is placed as close as possible to the V_{DD} potential because PMOS transistors usually have better matching and less flicker noise than NMOS ones.

The minimum supply voltage for this rail-to-rail input stage is equal to $(V_{gsPMOS} + V_{gs\ NMOS} + 2V_{sat}) \approx 1.8$ to 2.2 V depending on the process. At voltages below this level, the stage exhibits a forbidden voltage range [10, p. 100] in the middle of the common-mode input voltage range due to the fact that the transistors of both of the differential pairs are shut-off. This minimum supply can be decreased to 1.5 – 1.8 V by biasing the bulk of PMOS transistors as shown in fig. 5-2. This bias arrangement makes this input stage suitable for practically all application needs existing at the present time.

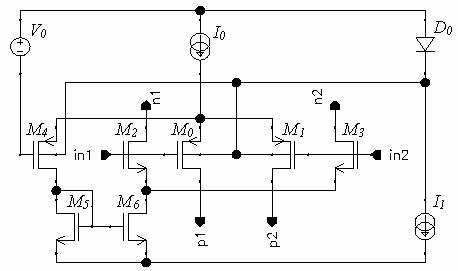


Figure 5-2. Complimentary rail-to-rail input stage with 1.8V minimum supply

When the common-mode input voltage is approaching V_{DD}, the negative bulk bias decreases the threshold voltage V_{TH} of PMOS input devices, allowing additional 200-300 mV of operating range for this differential pair. Using a diode for the bulk biasing ensures that the parasitic bipolar structures of the PMOS transistor will not be activated.

The direct current transfer between PMOS/NMOS differential pairs operating in strong inversion may cause up to a 40% spike of the input stage g_m. Many circuits addressing this problem have been reviewed in [10, pp. 102-119]. A simple low-power equalization circuit to reduce the spike is shown in fig. 5-3 [14, 56].

In this circuit, a replica of the PMOS pair tail current is generated by M_0'/M_1', and a replica of the transfer current – by M_4'. The translinear circuit M_7-M_9 with minimum-selection function [32] decreases the NMOS tail current at the switching region. Simulation results of g_m versus the common-mode input voltage (with and without the equalization circuit) are given in fig. 5-3b. They show that this circuit can provide less than 5% variation of g_m during the current transfer between the input pairs.

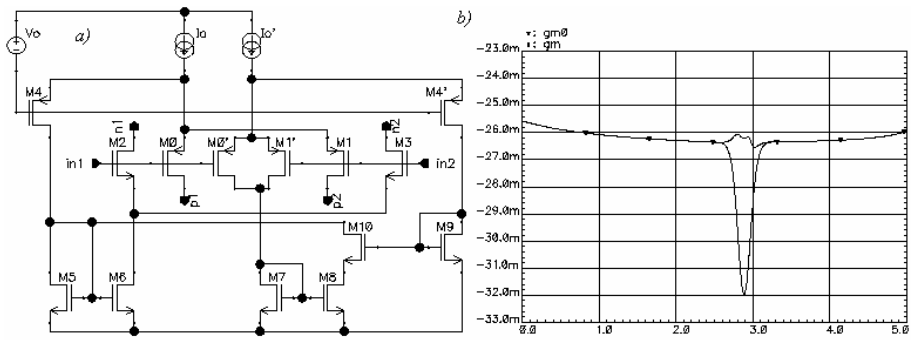


Figure 5-3. Control of the input stage g_m in PMOS/NMOS switching point

Switching between PMOS and NMOS input differential pairs also changes the OpAmp noise properties and complicates slewing behavior. In addition, it causes a few millivolts change of the input offset. This is well known, and usually is taken into account in correctly designed applications, so only one (usually PMOS) input differential pair is used during normal operation. The NMOS pair is still necessary, even if the top part of the common-mode region is not used, in order to avoid conditional instability during start-up or during large transients in an application.

A rail-to-rail input stage with a single PMOS differential pair and with a bootstrapped voltage source for the tail current can be used for applications prohibiting operation with the switching point [39]. This bootstrap voltage source can be created using an on-chip charge pump. The amplitude of the output ripples in this pump should be minimized; otherwise these ripples will leak to the OpAmp output, increasing the high-frequency noise. They may also be the cause of a DC error due to rectification on various component nonlinearities. The amplitude of the ripples is reduced using the low-noise charge pump described in subchapter 3.6. Also, it is desirable to set the charge pump operating frequency higher than the OpAmp *GBW* so that the switching noise will be filtered by the OpAmp itself and will not appear at the frequencies of interest.

5.2 CMRR/PSRR improvement

The design of an OpAmp insensitive to the power supply and common-mode voltage variations starts from having a supply-insensitive, preferably subregulated, biasing core (subchapter 3.5). A subregulated core not only has superior DC stability, but is also less sensitive to high frequency ripples of the main supply. Such ripples are unavoidable in the mixed-mode chips where digital or switching units operate simultaneously with analog blocks.

The biasing core current must be distributed to the critical for accuracy units with good, of high output impedance, current mirrors. To minimize the output capacitance of the current sources, the cascoding transistors must be as short and narrow as feasible. Use of the short transistors somewhat decreases the output resistance of these current mirrors. Improvement of the high output impedance can be done with structural design methods as discussed in subchapter 3.4. Bypass capacitors between the supply rails and the gates of the cascoding transistors should be routinely used to improve high frequency PSRR/CMRR.

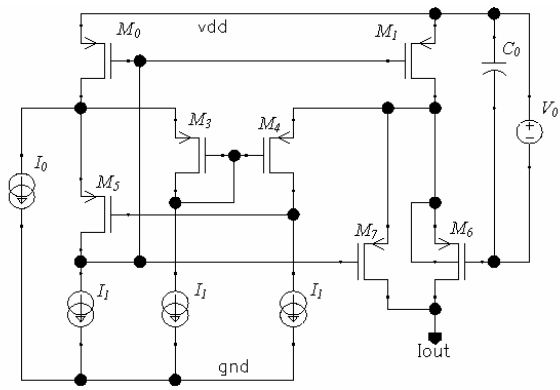


Figure 5-4. Tail current source

Stability and output impedance of the input differential pair tail current source is very important for good CMRR. The structures of such current sources have been considered in subchapter 3.4. The example of fig. 5-4 illustrates further current source modification in accordance with the structural design approach. The circuit of fig. 3-14 is taken as the initial design. If it is found that the output resistance and the voltage range of this circuit are still not satisfactory, then one has to use another feedback loop to boost the gain and improve the quality of the first loop. This is what is done in the current source of fig. 5-4. In addition to the initial current mirror M_0/M_1 with the reference current I_0 and cascoding transistor M_6, it now includes now the current-input amplifier M_5 where the gain boost (M_3/M_4 biased by small currents I_1) is added. When the voltage at the I_{out} node is approaching V_{DD}, M_5 starts to enter triode operation, and V_{dsM1} decreases. The amplifier $M_5/M_3/M_4$ raises the gate-source voltage of M_0/M_1 to keep the drain voltages of M_0/M_1 equal. In addition, it turns on M_7 in order to decrease the voltage drop across M_6. As a result, the output current of this current source starts to vary only when voltage between I_{out} node and V_{DD} becomes comparable with the offset voltage of M_3/M_4 pair.

Cascoding of the input differential pair is the technique to place it in the "greenhouse" environment and to improve its CMRR and PSRR [90]. With a good tail current source, the CMRR of an OpAmp with a cascoded input stage can be as high as 100-120 dB [57, 59]. Until recently, it was used mostly in bipolar amplifiers designed for ±15 V supply because the input cascoding circuits decrease the input common-mode voltage range and usually preclude its extension down to the supply rail.

The principle of the input pair cascoding is shown in fig. 5-5a.

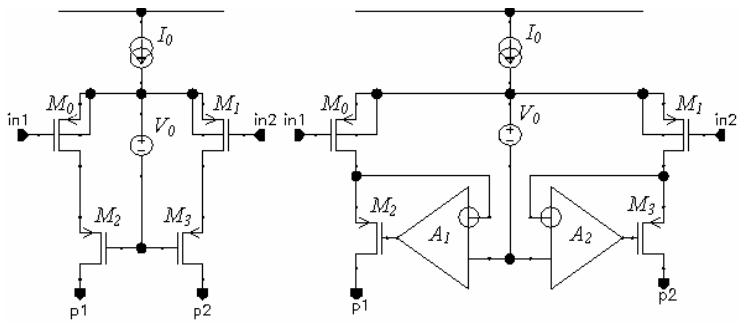


Figure 5-5. Cascoding of the input differential pair

The voltage source, V_0, is equal to ($V_{dsM0,M1}$ + $V_{gsM2,M3}$). $V_{dsM0,M1}$ should be kept stable in the operating range of input common-mode voltages. But $V_{gsM2,M3}$, however, depends on V_{ds} and varies with the common-mode input voltage varies. A better stability of $V_{dsM0,M1}$ voltage can be achieved using feedback loops (fig. 5-5b). The gates of the cascoding transistors M_2/M_3 do not need to be connected together, and each input device has a dedicated feedback loop as shown. In this structure, the amplifiers A_1, A_2 provide a stable V_{ds} for M_0 and M_1, and, simultaneously, a voltage gain boost (subchapter 4.3, fig. 4-8) for the cascoding transistors M_2 and M_3. The separate control of V_{dsM1} and V_{dsM0} also improves the slewing behavior and simplifies the design of the clamp circuits preventing the input differential overload.

An implementation of the structure shown in fig. 5-5b is given in fig. 5-6.

For the input differential pair operating in weak inversion, the voltage V_{ds} can be as small as 4-5 V_T. This should be taken into account when defining the voltage V_0 that is implemented here with resistor R_0 and a PTAT (as V_{ds} should be proportional to V_T) current source I_1. The amplifiers A_1, A_2 are realized using the differential pairs M_4/M_5 and M_6/M_7.

The cascoding transistors M_2/M_3 are in the signal flow path and they introduce a pole defined by the parasitic drain capacitance of M_0/M_1 and the source resistance of M_2/M_3. Large input transistors M_0, M_1 have significant drain-to-body capacitances, and this pole can make M_2/M_3 a gain stage and

complicate the OpAmp compensation. The amplifiers A_1/A_2 reduce the input impedance at the sources of M_2/M_3 and improve the high frequency behavior as illustrated in the bottom right part of fig. 4-7c. Also, transistors M_2/M_3 do not affect the noise parameters, and they can be much smaller now than the input pair. Then the pole created by the drain capacitance of M_2/M_3 and the input resistance of the following folded cascode (in case of its presence) will be located at a higher frequency and will not influence the overall OpAmp bandwidth.

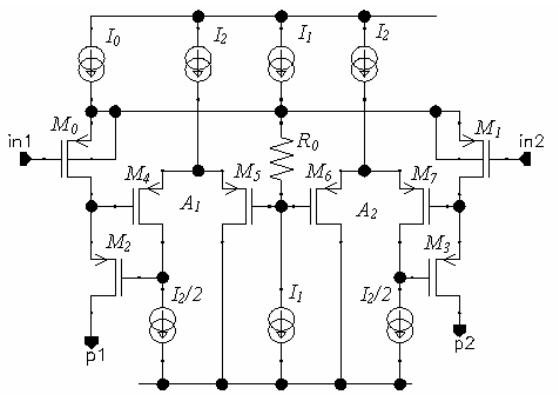


Figure 5-6. Cascoded input differential pair with voltage gain boost

Using the current sources $I_2/2$ instead of the current mirrors pursues the same goal. In this case one eliminates the current mirror delay and improves the stability in the A_1-M_2 and A_2-M_3 feedback loops.

The bulk to source voltage of the input differential pair should also be constant to improve the CMRR. In the circuits of fig. 5-5 and 5-6, the bulk of the input devices is connected to the source. The common bulk of these large input devices has a significant parasitic capacitance to the substrate. During OpAmp slewing in the follower mode, this capacitance has to be charged or discharged, and the current in this capacitance is added to or subtracted from the tail current [110, part 5]. In the absence of slew rate enhancement circuit (subchapter 8.1), this additional charging/discharging results in the large-signal step response discontinuity and asymmetric slewing. In the spirit of structural design, the problem can be solved with the addition of a feedback loop controlling the bulk voltage, so that the tail current source, during the transient process with slewing, is not loaded, and the required current charging the body-substrate capacitance is supplied by the feedback amplifier (see below, the amplifier A_3 in fig. 5-7).

The voltage at the gates of M_2/M_3 does not reach the values that are below the negative supply rail voltage (apart from a possibility of a very complicated negative charge pump with the circuits of substrate voltage

control). Yet, the extension of the common-mode input range to the negative rail (single-supply operation capability) is very important, especially in low-voltage applications, as the negative rail is often used as the signal ground.

Input common-mode voltage range still can include negative rail with operational cascoding if $V_{gsM0,M1} > (V_{dsM0,M1} + V_{gsM2,M3})$. It can be achieved by the bulk voltage control of both the input pair M_0/M_1 and the cascoding transistors M_2/M_3 as shown in fig. 5-7 [58].

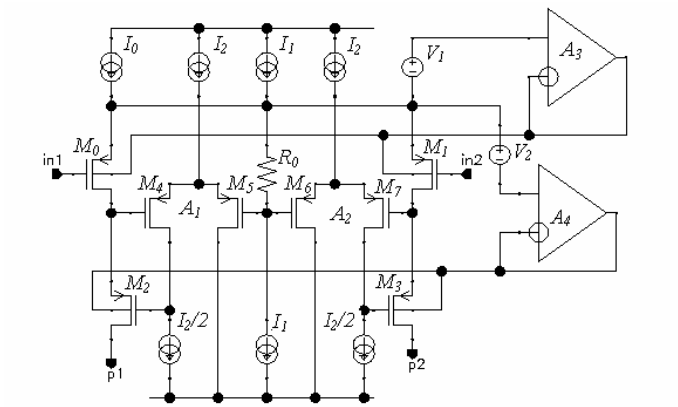


Figure 5-7. Cascoded input stage with body voltage control and single-supply capability

Positive biasing of the M_0/M_1 bulk by V_1 increases $V_{gsM0,M1}$, negative biasing of the M_2/M_3 bulk by V_2 is decreasing $V_{gsM2,M3}$. One can achieve the required $V_{gsM0,M1} > (V_{dsM0,M1} + V_{gsM2,M3})$ and allow some headroom for the $I_2/2$ current sources when inputs are at the negative rail. The source V_2 should be smaller than V_{be} to avoid activation of the parasitic lateral and vertical bipolar PNP transistors with the bulk of M_2/M_3 as the base.

Implementation of the structure of fig. 5-7 results in an OpAmp with better than 100 dB CMRR and with single-supply input capability [59].

The "natural" p-channel MOS transistors (without V_{TH} adjustment process step), if available, have a larger V_{TH}, and can be used in the input pairs in the circuits shown in fig. 5-5 and 5-6. These transistors may have even better matching than regular PMOS transistors. With "natural" p-channel MOS and regular PMOS transistors used for cascoding, such input stage can have a single-supply capability even without bulk voltage control.

If the process allows design of PNP transistors with a negligible substrate leakage then these transistors can be used for cascoding of the input differential pair (fig. 5-8). For processes with a minimum channel length of more than 0.35um, the voltage V_{TH} of PMOS transistors is larger than V_{be}. For smaller geometries, the "natural" PMOS transistors with larger V_{TH} may be used in the input stage. In this case, even with the inputs at the negative

rail, the equation $V_{gsM0,M1} > V_0 = (V_{dsM0,M1} + V_{be})$ may be satisfied, and the single-supply input capability is preserved without any additional circuitry.

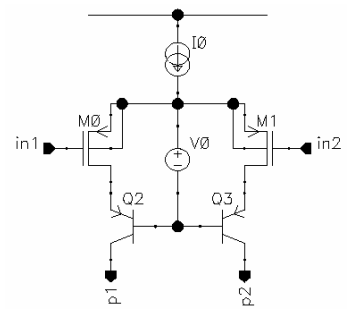


Figure 5-8. Input stage cascoding with bipolar transistors

The high-frequency PSRR is defined by the compensation scheme of the OpAmp as a whole, providing that the biasing is subregulated and the critical current sources have gate bypass capacitors. With simple Miller compensation schemes (fig. 4-12a) and a rail-to-rail output stage, the compensation capacitors are connected between the output and the gates of the output transistors (C_0/C_1 in fig. 4-21). Any high-frequency ripple at the supply lines is transmitted to the output and is suppressed by the OpAmp gain at the ripple frequency. The only way to reduce this output ripple more is to use different compensation schemes. Using enhanced Miller compensation (fig. 4-12b or 4-12c), using single-stage OpAmp structure not requiring any compensation, or employing the parallel compensation scheme [22, chapter 4.2] may drastically improve the ripple suppression.

5.3 Trimming techniques

The offset voltage and its temperature drift are usually the most objectionable error parameters of an OpAmp. Even the very first monolithic OpAmp had external pins to connect variable resistors for offset trimming in the application. Various on-chip trimming techniques were developed for this purpose. They include zener-zapping, laser resistor trim, laser link cutting, and, lately, using on-chip DACs controlled by EPROM.

Zener-zapping was one of the first IC trimming techniques [73, 74]. It consists of melting a reverse-biased p-n diode by a large current (usually in single ampere range). An unmelted diode has only a few picoampere leakage current. The resistance of this diode after overheating and melting is only a

few Ohms. This technique is convenient to change the value of the resistor-diode network (fig. 5-9).

Zener diodes are available, without additional masks and cost, in almost any bipolar/BiCMOS process. For this reason, the zener-zapping technique has been widely used for many years, and still remains the only trimming method available even in some modern processes [4]. It has, however, many drawbacks: a large die area is required for the contact pad to each diode; short life of the test probe needles carrying amp-range currents during zapping; offset variation caused by the on-chip heat waves.

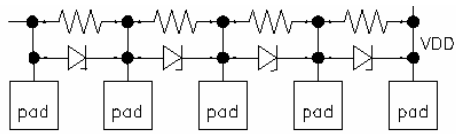


Figure 5-9. Trimming with zener-zapping

Fusing on-chip metal links (fig. 5-10) has the same drawbacks and advantages as zener zapping. This technique also requires current in the ampere-range, and contact pads for each link. The heat waves have the same effect. And it is also available on any process.

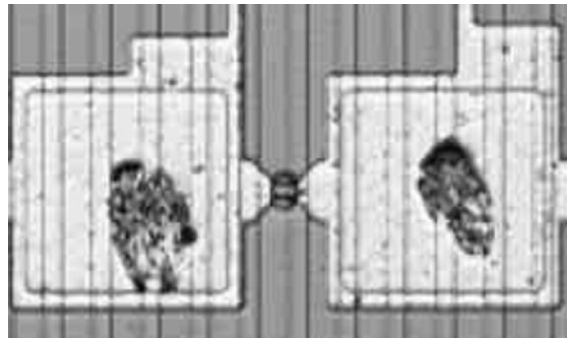


Figure 5-10. Fused metal link above diffused resistor

Laser cutting of thin-film resistors and links was first developed for hybrid integrated circuits [75] and later transferred to silicon by analog - oriented companies (Burr-Brown, ADI, LTC) [76]. The technique employs the evaporation of a narrow path of a thin metal layer by a laser beam. A typical laser-trimmed resistor is shown in fig. 5-11.

The coarse trim part is on the left. The laser beam trace significantly increased the length and decreased the width of this resistor and causes the large change in its value. Similar traces in the fine trim pad on the right just

slightly decrease the width of the resistor, making the value change much smaller.

Laser cutting of aluminum links has also been in use for a long time [77]. Another variation of this technique, laser cutting of poly-Si links in standard CMOS processes was developed in the 90s. This procedure is used for probe trimming of CMOS OpAmps.

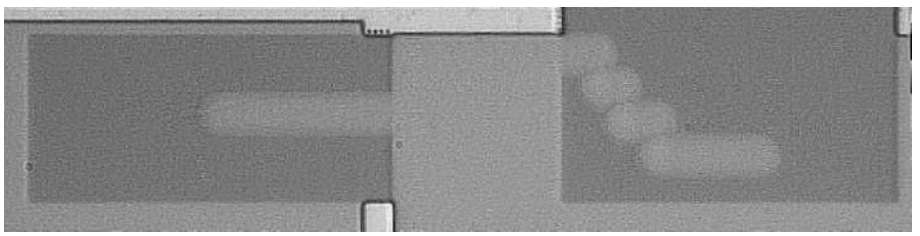


Figure 5-11. Laser-trimmed thin-film resistors (coarse and fine trim)

The drawbacks of laser trimming technique are the following. It requires additional process steps to protect silicon surface below the trimmed elements and, in some cases, opening the die passivation above the cut area. Such opening can reduce the climate robustness of the chip, so special materials between the resistor and glass that absorb the evaporated metal may be used. This trimming can only be done at the wafer probe test before circuit packaging. The links after laser cutting still retain some conductivity.

The fusing of poly links in a CMOS process is a recent advancement (the fuse layout is shown in fig. 5-12a). For the fine-line processes, the poly fusing current can be as small as 20-40 mA. This allows using relatively small on-chip transistors (M_0 in fig. 5-12b,c) for fuse programming.

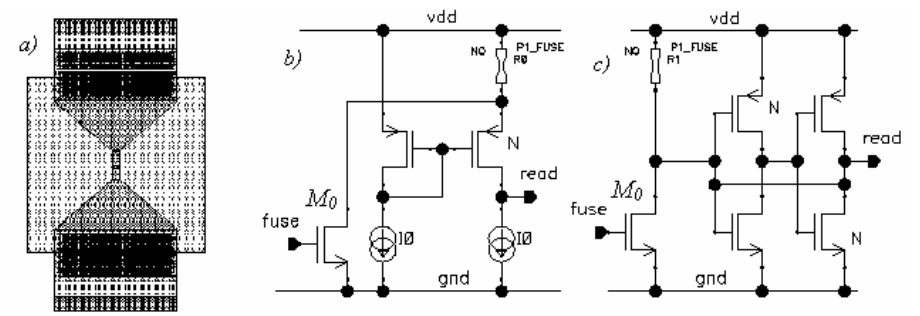


Figure 5-12. Poly link and poly fuse ROM

Fuse state reading can be done with a current-input differential amplifier (fig. 5-12b) or by a latch with mismatched input devices (fig. 5-12c). If the fuse in fig. 5-12c is intact then it defines the state of the latch. With a burned fuse, the latch state is set during start-up by the mismatch of input devices.

Poly fuses do not require any additional process steps. They can be programmed during the final test, allowing package-level trimming of the OpAmps. Only 20 to 50 bits of memory are necessary to store the offset and drift trimming information. Such amount of memory in poly fusing form results in a relatively minor increase of the chip area and does not significantly inflate the production cost.

EPROM or flash memory is available on some of the modern processes. It costs few additional steps and masks. Programming of the flash memory requires a high voltage supply that may be provided internally, from an on-chip charge pump, or externally, via an additional pin. Using flash memory to store the trimming information is justified if the OpAmp is a part of a system that already includes a large quantity of memory and the means for memory programming.

Industrial OpAmps are usually manufactured with standard pin-outs, and no additional pins are available to control the ROM programming process during the trim and test of packaged units. Methods to control in-package trimming while using existing IC pins are yet to be developed.

5.4 Offset and temperature drift trimming

OpAmp offset trimming can be done using the following methods (fig. 5-14):

1. Size adjustment of the input transistors by connecting/disconnecting the sources of fractional transistors (fig. 5-14a) to the sources of main devices. Laser cutting or fusing of poly links performs the source disconnection. As the input voltage has the rail-to-rail span, the commutation in packaged units should be done by NMOS/PMOS switching pairs.

2. Size adjustment of the input devices by switching the drains of fractional transistors (fig. 5-14b). Due to the residual link resistance, laser cutting is not used. The drains of disconnected fractions should connect to the supply rail to avoid parasitic effects. The commutation switches using only one type of transistors are usually sufficient as the drain potentials are usually near one of the supply rails.

3. Adjustment of the current mirror degeneration resistors (fig. 5-14c). This method used to be the most popular technique for laser offset trimming. Resistor network trimming can also be applied using metal or poly fuses and zener zapping. CMOS switches here are not optimal due to their relatively large, supply- and temperature-dependent turn-on resistor R_{on}.

4. Adjustment of the load current sources of the input pair's drains by switching the sources of fractional current sink transistors (fig. 5-14d). This can be done by fusing metal or poly links or by zener zapping.

5. Same as 4., only by switching the drains of fractional current sink transistors (fig. 5-14e). It can be done using MOS switches with one type of transistors.

6. Addition of a trimmable current source to one of the current sinks (fig. 5-14f).

The offset voltage temperature drift may be as important to the customer as the offset voltage itself. It was mentioned that the temperature drift of the bipolar amplifier is proportional to the offset. Trimming offset reduces the temperature drift as well. Drift values of 0.1-0.2 uV/°C are realistic for high precision trimmed bipolar amplifiers.

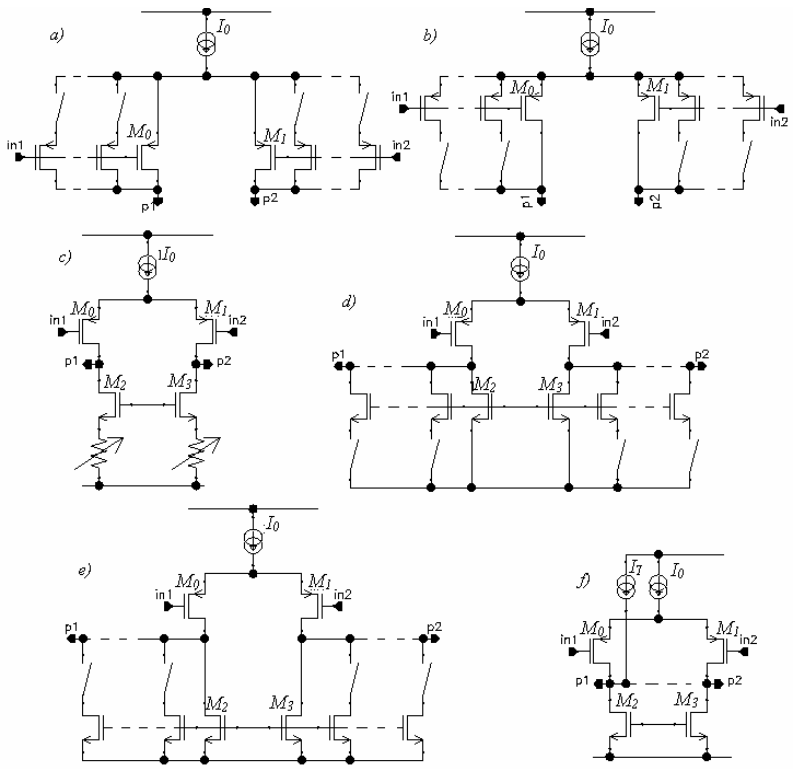


Figure 5-13. Offset trimming techniques

Trimming of both the offset and offset temperature drift has been realized for JFET input amplifiers as well. The procedure includes two temperature offset measurements, calculations and trimming of the offset-controlling

variables (resistors or currents) [55]. This procedure can only be used in the wafer probe test (i.e., with unpackaged units) because it requires the knowledge of the OpAmp ID. On a wafer, this ID is defined by the OpAmp location. The offset measurements are done for all amplifiers on the wafer at the first temperature, then the temperature of the test environment changes, (heated up or cooled down) and the measurements are repeated. The trimming is performed using the data of the first measurement as a reference.

An automated test of packaged parts at an arbitrary temperature is technically realistic. However, tracking the identity of a packaged unit up to the final test is unfeasible unless the ID number is programmed into each unit. Storing this ID number would at least double the amount of required on-chip ROM. It would also require the addition of a communication port to the OpAmp in order to read this ID back into test system. As a result, keeping the packaged amplifier ID is feasible only if the OpAmp is a part of the system already having flash memory and a communication port.

The temperature drift of a MOS differential pair is not correlated with the offset while operating in either weak or strong inversion. Temperature drifts on the order of 10 uV/^{o}C are not uncommon. In a correctly designed OpAmp, without drastic changes of input stage or biasing operation mode with temperature, this drift is relatively constant. A significant curvature in the offset versus temperature dependency may appear only at 110-120 ^{o}C. This curvature is attributed to the junction leakage currents rapidly growing at high temperatures.

Measurement results of the offset versus temperature (normalized to zero at –55^{o}C) dependencies for a set of OpAmps are shown in fig. 5-14.

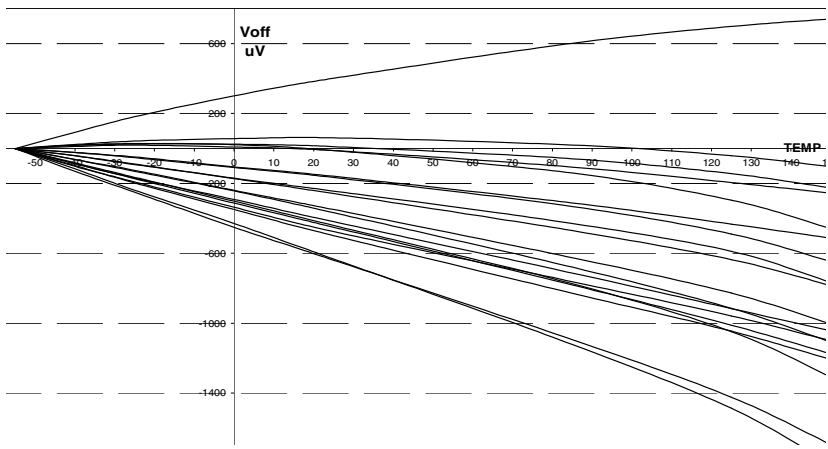


Figure 5-14. Offset versus temperature for the typical CMOS OpAmp

A recently introduced method [15, 62] allows one to implement independent trimming of the offset and first-order temperature drift without amplifier identification. As one can see from fig. 5-14, the dependence of offset versus temperature is approximately linear. This method relies on this fact, and uses two temperature test results to eliminate this main linear component. This drift trimming can be done in large production batches. After testing and trimming at the first temperature, rejection of bad units, etc., the whole batch proceeds to the second step in any unit order. No amplifier ID, except the data on what trims have been done to which batch, is required.

Using this method, drifts in the range of 0.5-1 uV/^{o}C (box method) for CMOS OpAmps can be achieved. Further reduction of the drift is limited by the curvature in the offset versus temperature dependence.

5.5 Input protection

The inputs of a CMOS OpAmp should be protected against electro-static discharges (ESD), latch-up and the gate oxide breakdown in case of an excessive input differential voltage. It is also advantageous if the OpAmp retains some functionality when the input common-mode voltage exceeds the specified range.

The inputs of a bipolar OpAmp should also be ESD and latch-up protected as well as protected against overheating that may be caused by an excessive differential input voltage. In addition, one has to prevent output phase inversion when the common mode input voltage is outside the specified range. This last requirement has become an industry standard. The OpAmp functionality when the common-mode input voltage is outside the range is not yet required, but desirable.

ESD sensitivity is tested by discharging a capacitor through a special RC circuit at each pin. There are three standard RC circuits for ESD testing called: machine model, body model and a combined model, with the survival voltage requirements for each variant [63]. ESD failures are one of the leading causes of layout redesigns.

In many fine-geometry low-voltage processes, ESD protection consists of edge-triggered SCRs or high-current switches connected between the V_{DD} rail and the substrate (fig. 5-15). Diode and SCR combinations, connected between the input/output pins and the supply rails, may also be used.

ESD protection circuits invariably include large p-n junctions. They are a source of large leakage currents, especially at high temperatures, which degrade the precision parameters of an OpAmp and are a source of parasitic capacitances that deteriorate the parameters of high-speed OpAmps.

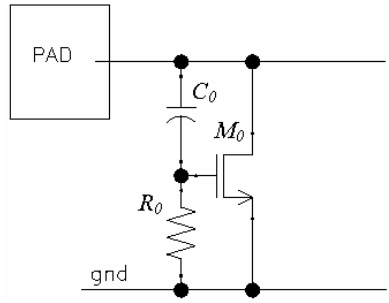


Figure 5-15. Edge-triggered ESD protection

ESD protection of the inputs can be improved using *R-NMOS* network R_1/M_1 as shown in fig. 5-16. R_1 can be as large as it does not noticeably affect the OpAmp noise and bandwidth ($GBW << 1/2\pi R_1 C_{gM0}$).

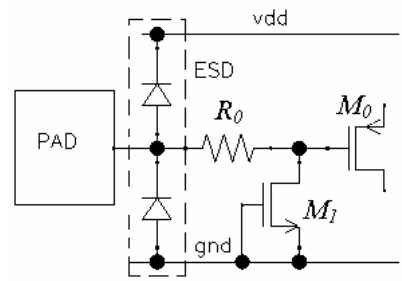


Figure 5-16. Input ESD protection

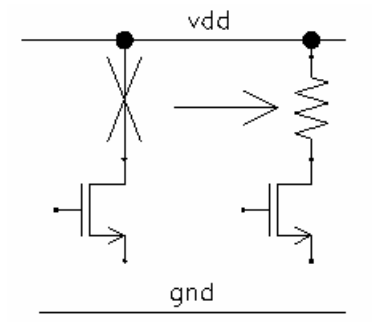


Figure 5-17. Input ESD protection

A direct connection of NMOS drains to the V_{DD} line in single-well processes should be avoided in all parts of the circuit (fig. 5-17). During an

ESD event, every N-well connected to the V_{DD} acts as a Zener diode conducting current between supply rails. If unlimited, this current can fuse the narrow metal line from transistor to the V_{DD} bus.

The voltage breakdown of bipolar transistors does not cause any irreversible effects as long as there is no overheating. This property of devices simplifies the design of ESD protection for bipolar circuits.

N-P-N-P SCR structures do present on every CMOS chip, and the latch-up is possible [84]. The robustness to latch-up is tested by running a 100 mA current pulse between all pins of the powered part and comparing the quiescent current of the part before and after test. Latch-up is prevented by layout means: using guard rings, larger spacing between components and component placement in a way that parasitic SCR structures activation is prevented.

Input overvoltage protection should prevent amplifier damage in case of a long-term (ideally, infinitely long) connection of the inputs to supply rails of opposite polarity. The protection circuit (fig. 5-18) includes resistors and voltage clamps of any available type. ESD protection resistors are usually a part of this circuit as well. In CMOS amplifiers, this protection is a must when the oxide breakdown voltage of the input transistors is less than the supply voltage.

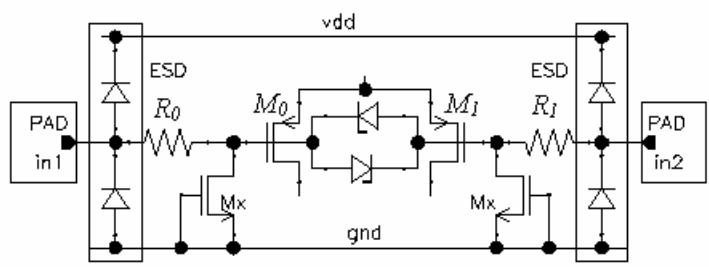


Figure 5-18. Input overvoltage protection

Overvoltage protection of bipolar amplifiers can be designed in the same way, but the resistor values should be as small as possible because these resistors increase the amplifier noise due to the presence of input transistor base currents. Using small resistors requires that the voltage clamp should be designed for larger currents. For example, an OpAmp in an SO-8 package can dissipate up to 0.5 W, therefore the current through the input protection network for a 20-V supply OpAmp can be as large as 50 mA and the resistors can be as small as 200 Ohm.

The main input stage of the CMOS OpAmp created to achieve the design goals in accuracy and speed may not have rail-to-rail capability. At the same time, rail-to-rail outputs have became an industry standard. Hence, the input

voltages for an OpAmp in an application where it works in cascade connection with similar OpAmps, can exceed the specified input range during start-up or a sudden transient. To avoid unpredictable behavior of the system, a low-quality input stage covering the rest of the input common-mode range and, thus, providing the rail-to-rail span may be added to the OpAmp.

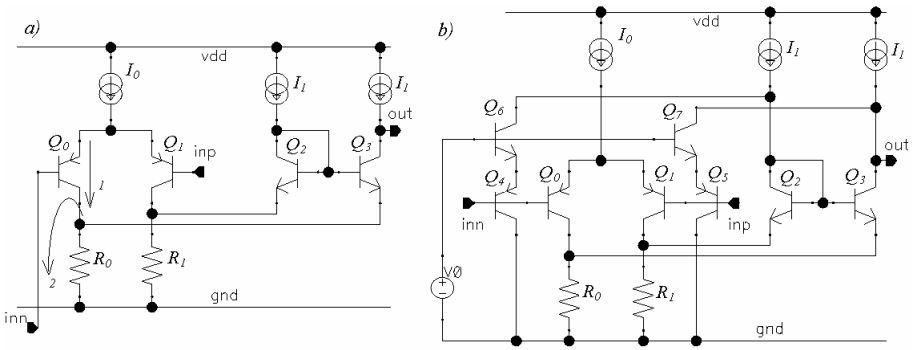


Figure 5-19. Output phase inversion in single-supply bipolar OpAmps

Phase inversion of the output voltage in bipolar OpAmps is explained in fig. 5-19a. In the bipolar stage Q_0/Q_1 with folded cascode transistors, Q_3/Q_4, the collector current of Q_0 during normal operation flows in the direction 1. Let's assume it increased the potential at the emitter of Q_3, and the output voltage is close to the V_{DD}. When the voltage of the input *inn* drops below *gnd*, the collector-base junction of Q_0 becomes forward-biased and its current starts to flow in the direction 2. This decreases the potential at the emitter of Q_3 and the output moves closer to the zero. Hence, the output voltage has the phase inversion.

One of the solutions to this problem is shown in fig. 5-19b. When the input common-mode voltage drops below operating range, the parallel input stage consisting of Q_5-Q_8 becomes active and prevents the output signal phase inversion.

Chapter 6

INTERMEDIATE AMPLIFICATION STAGES

In the general case, the connection of the input stage to the output stage requires a voltage shift. For bipolar amplifiers, the current gain obtained by a direct connection of the input and output stages is not sufficient. The tasks of level shifting and adding gain (voltage and/or current) are performed by intermediate stages. Some other requirements may also necessitate the inclusion of an intermediate stage. For example, the amplifier in fig. 6-1a does not need an intermediate stage in most of the common mode input range. However, it becomes not operational with the input signal close to the negative rail. The single-supply input capability requires the addition of an intermediate stage, in this case the folded cascode (fig. 6-1b).

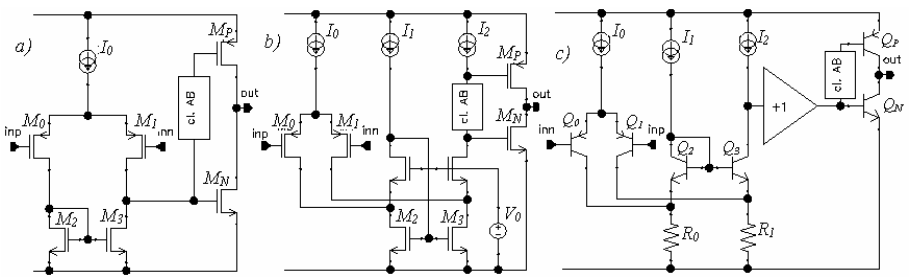


Figure 6-1. Connection of the input and output stages

In bipolar OpAmps, the current gain should be at least 90-100 dB, which, assuming a β between 50 and 200, means that at least four current amplification stages are necessary. To achieve the sufficient current gain, a voltage follower is introduced usually between the input and output stages (fig. 6-1c).

The intermediate stages should not deteriorate the OpAmp speed, and, therefore, they must be an amplification, and not a gain, stages. This requirement limits the available options to the common-gate and common-drain transistor configurations. Some local feedback loops may be necessary to improve the input and output impedances or voltage and current gain of these configurations (subchapter 4.3) and to avoid their conversion from amplification into gain stages.

The differential signal from the input stage should be converted somewhere into a non-differential form. This usually means the presence of a current mirror in at least one signal path from the inputs to output, like in fig. 6-1a. This mirror can be in a folded cascode form as in fig. 6-1b.

The voltage noise V_{nmr} (and the offset) of the current mirror transistors increase the input-referred noise by $V_n = V_{nmr} g_{m2} / \sqrt{2} g_{m1}$ (a similar result is valid for offset) where g_{m1} is the transconductance of input pair and g_{m2} is the transconductance of mirror transistors. In order to decrease V_n, the area of the mirror transistors should be comparable with that of input pair and g_{m2} should be as low as feasible. The mirror should therefore consist of transistors operating in strong inversion with longer channels and include source (emitter) degeneration resistors if possible.

The mismatch and noise of the current sources I_1, I_2 in the folded cascode also contribute to the input offset and noise, with the transfer coefficient $1/g_{m1}$. Often these currents are chosen equal or larger than the tail current I_0 in order to get the symmetric slewing behavior of the OpAmp.

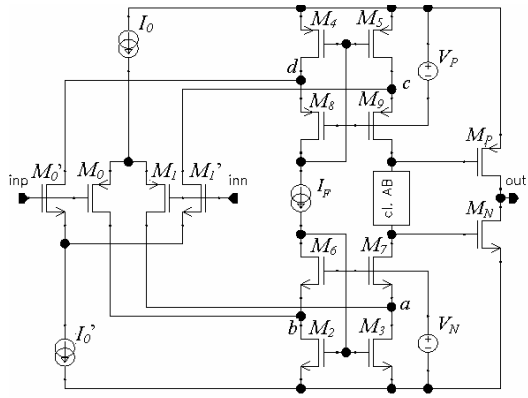


Figure 6-2. Input stage with complementary folding cascode

The error caused by the mismatch of currents I_1, I_2 can be lesser if these currents are smaller than the tail current I_0. Current sources I_1, I_2 can be replaced by the floating current source I_F and the current mirror M_4, M_5 as shown in fig. 6-2. In this circuit, the noise of I_F does not influence the input-

referred noise, but the effects of the current mirror M_4, M_5 offset and noise are equivalent to those of the mirror M_2, M_3. They should be added to the offset and noise introduced by M_2, M_3.

The complementary folded cascode with floating I_F also provides for easy switching between the outputs of the PMOS and NMOS pairs in the rail-to-rail input stage.

The folded cascode is not the only option for connecting the input and output stages while preserving the single-supply operation capability. One can use also the voltage follower/shifter (fig. 6-3).

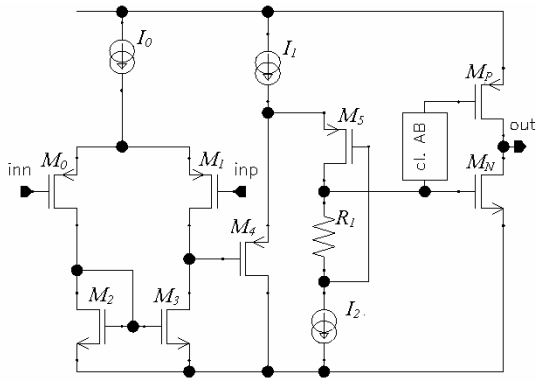


Figure 6-3. Use of the voltage shifter for connecting the input and output stages

This voltage shifter comprises M_4, M_5, and R_1. To have the level shift that does not depend on transistor dimensions one has to choose $I_2 \approx I_1/2$. The voltage drop across R_1 is $V_{sfht} = I_2 R_1 \le (V_{gs} - V_{sat})$ and prevents M_5 from entering the triode operation. When the inputs are close to the negative rail, the gate voltage of M_N is limited by the relation $V_{gsMN} < V_{gsM1} + V_{shft}$ as the gate potential of M_4 is not higher than the source potential of M_1.

For many applications, the simple circuits considered above do not have sufficient gain. This deficiency may be corrected by using the same circuits with addition of gain boost (see subchapter 6.3).

6.1 Floating current source

The floating current source is useful not only in the symmetric folded cascode, but in other applications as well. It is convenient to include it as a basic cell of a designer's library.

A circuit for implementation of the floating current source is shown in fig. 6-4 [44].

The circuit comprises the current source I_{ref}, and the transistor M_{ref}

creates the reference gate-source voltage V_{gs} for the transistor M_{10}. To obtain the best matching of the drain currents for these two transistors, the drains of M_{10} and M_{ref} should be at the same voltage. This is provided by the diode M_{11}. The drain (and source) current of M_{11} matches I_{ref} as long as A_1 is keeping $V_{gsM10} = V_{gsMref}$.

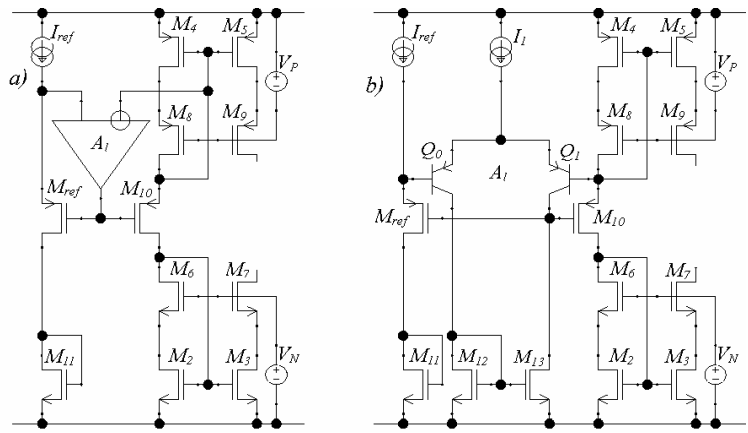


Figure 6-4. Floating current source implementation

The implementation of A_1 shown in fig. 6-4b uses PNPs Q_0, Q_1 and the current mirror M_{12}/M_{13}. The bipolar transistors provide better current accuracy and have a lower offset voltage. The base potentials of Q_0/Q_1 are less than V_{DD} by the voltage V_{gsPMOS}. The voltage $V_{be} < V_{gsPMOS}$ so there is sufficient headroom for the current source I_1.

The minimum supply voltage of the complementary folded cascode (fig. 6-2) with the floating current source of fig. 6-4 is $(V_{gsPMOS} + V_{gsNMOS} + V_{sat}) \geq$ 2.2 V, if the transistors of the current mirrors operate in strong inversion and have $V_{gs} > V_{TH}$. This minimal voltage can be reduced to $(V_{gs} + 2V_{sat}) \leq 1.5$ V, if the transistors in one of the current mirrors are employed in the current control feedback loop (fig. 6-5).

In this circuit, I_{ref} and M_{ref} create the reference voltage V_{gs} for M_8 (as it was in the previous circuit for M_{11}), but the amplifier A_1 controls the current of the mirror transistors M_4/M_5 instead of the dedicated transistor M_{11}. As in the previous circuit, the drain (and source) current of M_8 matches I_{ref} as long as A_1 has enough gain to keep $V_{gsM8} = V_{gsMref}$. The currents of M_4 and M_5 can be equal to the current of M_8, or different from it when the NMOS input pair currents (see fig. 6-2) flow in M_4 and M_5 in addition to I_F. In this case, the M_8-A_1 path is the functional replacement of the wire connecting the drain of M_8 with the gate of M_4 in the circuit of fig. 6-4.

The amplifier A_1 can be implemented (fig. 6-5 b) as a differential pair M_{12}/M_{13} loaded by the current mirror M_{14}/M_{15}.

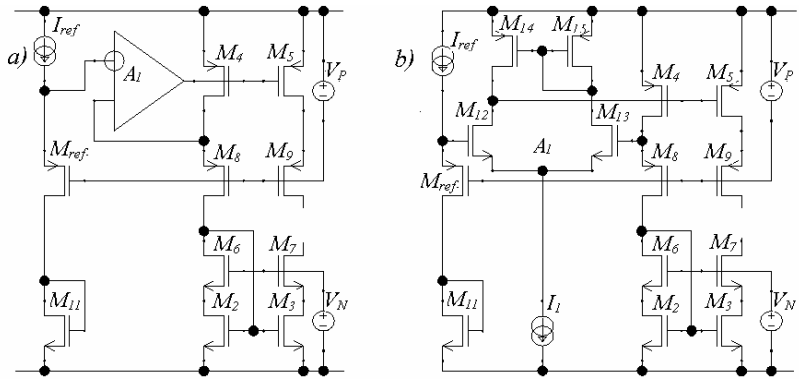


Figure 6-5. Folded cascode with floating current source and low supply voltage

The variation of the floating current source current with changes of the supply voltage or temperature may be detrimental for OpAmp PSRR and offset voltage temperature drift. This variation can be decreased by a higher gain in the current control feedback loop. Improving M_{ref}/M_4 matching will help as well. The last measure is achieved, for example, by cascoding these transistors (fig. 6-6) so that their drain voltages are equal and do not depend on the supply voltage variation.

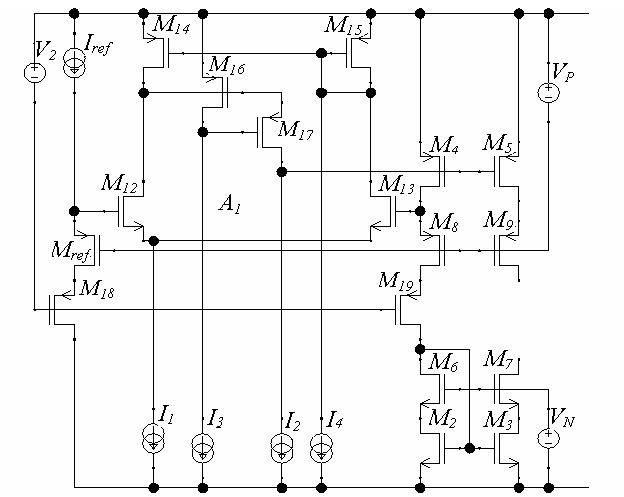


Figure 6-6. Improvement of the floating current source accuracy

Here A_1 is realized using the differential pair M_{12}/M_{13}, the current mirror M_{14}/M_{15} and the gain boost M_{16}/M_{17} (it is similar to the gain boost shown in fig. 4-8c). M_8 and M_{ref} are cascoded with M_{18} and M_{19} and the reference voltage for the gates of cascoding transistors should be $V_2 \geq V_P + V_{sat}$. The voltage V_{dsM17} is equal to $V_{gsM4} - V_{gsM16}$, and to keep the gain boost operational, M_{16} should operate in the weak inversion.

6.2 Current mirrors of the folded cascode

The folded cascode stage comprises current sources and current mirrors (fig. 6-2). This stage is necessary to achieve single-supply input capability. It also provides the differential to single-ended signal conversion. The stage should be designed with the goal to minimizing the detrimental effects to the accuracy and dynamic parameters of the OpAmp. The folded cascode can also be used for the OpAmp voltage gain boost, and it is more convenient to transfer the boost circuits into this stage than to use them in the input stage.

The noise and mismatch in the current mirrors M_2/M_3 and M_4/M_5 of the folded cascode can significantly increase the overall OpAmp noise and offset.

The sum of the voltage V_{nmr} noise/offset of the mirror transistors causes the equivalent noise/mismatch in the folded cascode output current $I_{nmr} = V_{nmr}g_{m2}$, and equivalent input-referred noise/offset equal to $2I_{nmr}/g_{m1}$.

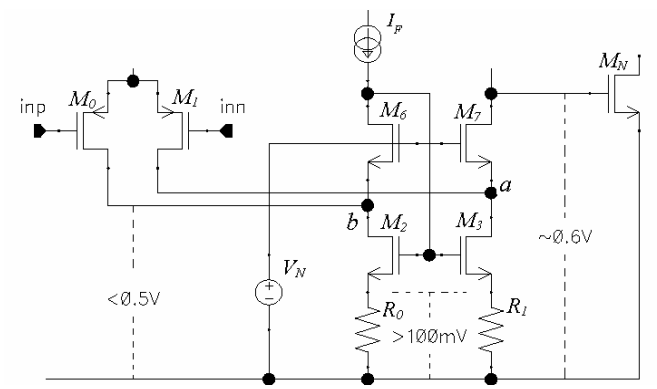


Figure 6-7. Voltage limitations in the folded cascode

The voltage offset and flicker noise of the MOS transistors is inversely proportional to the square root of their area. To reduce the contribution of the mirror transistors M_2/M_3 and M_4/M_5 in the overall noise/offset their overall area should be comparable with that of the input differential pair(s).

In order to decrease g_{m2}, these mirror transistors should have longer channels and operate in strong inversion.

The current through the mirror transistors also affects g_{m2}, and the choice of a smaller current decreases the noise contribution I_{nmr}. Another method to decrease g_{m2} is to use the source (emitter) degeneration resistors (fig. 6-7). To be efficient, the voltage drop across these resistors should be equal to at least to $4V_T$ (i.e. >100 mV at 20°C).

To achieve the single-supply capability, the voltages at the folded cascode inputs *a, b* should be $V_{a,b} < (V_{gsMO,M1} - V_{sat}) \approx 0.5$ V.

A voltage drop of 100-150 mV across the resistors leaves about 300 mV for the drain source voltage V_{ds} of M_2/M_3. At the same time, the triode operation of M_2/M_3 should be avoided because it will cause an increase of the input-referred noise and a drop of the OpAmp open-loop gain. This means that M_2/M_3 should be in strong inversion but the overdrive $(V_{gs} - V_{TH})$ can be only 200-300 mV. There is little room for V_{dsM3} variations and $V_{a,b}$ must be tightly controlled. The design of high voltage side of the folded cascode (transistors M_4/M_5-M_8/M_9 in fig. 6-2) should be based on similar considerations.

The drains of M_7 and M_9 are often connected directly to the gates of the output devices M_N and M_P (fig. 6-7). The sizes of M_N and M_P are usually chosen to provide the maximum load current with relatively small V_{ds} and V_{gs} voltages. In the absence of a load, these transistors operate in weak inversion with $V_{gs} < V_{TH}$. Whether the triode operation of M_7 and M_9 is acceptable or not depends on the gain boost scheme. For example, it should be avoided with the gain boost as in fig. 4-21, as transistor M_7 (M_0 in fig. 4-21) is part of the gain boost feedback loop and its triode operation will effectively break this loop. The triode operation can lower an already tight value of $V_{a,b}$, especially on the low side as $V_{TH\,NMOS} < V_{THPMOS}$.

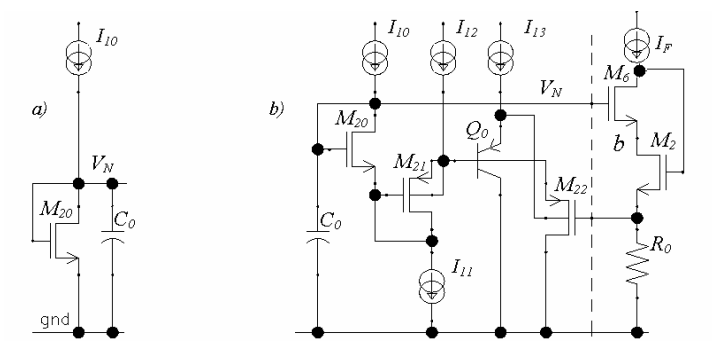


Figure 6-8. Control of the input voltage of folded cascode

The drain-source voltages on cascoding transistors are defined by the reference current I_{ref} and voltage sources V_P and V_N (see fig. 6-7). These voltages can be obtained using a current source and diode-connected MOS transistor (fig. 6-8a).

If this simple circuit does not provide good control of the drain-source voltages over temperature and process variations, then the circuit shown in fig. 6-8b, where the voltage $V_{dsM\,2}$ does not depend on V_N, may help to solve the problem.

In this circuit, M_6, M_2 and R_0 belong to the folded cascode circuit. The current ratio I_{10}/I_F should be equal to the aspect ratio of M_{20} and M_6. The current I_{11} should be equal to $I_{11} = I_{10} + I_{12}/2$ so that the drain current of M_{22} is also equal to $I_{12}/2$. If transistors M_{21}/M_{22} are matched then:

$V_{dsM\,2} = |V_{gsM\,22}| - |V_{gsM\,21}| = |V_{THM\,22}| - |V_{THM\,21}| = |V_{THM\,22}| - |V_{THP}|$.

But the threshold voltage of M_{22} is larger than $|V_{THP}|$. Therefore the voltage drop across M_2 is equal to $V_{dsM\,2} = \gamma(\sqrt{V_{be} + |2\phi_f|} - \sqrt{|2\phi_f|}) \approx 150 - 300$ mV.

Small capacitors similar to C_0 (about 0.5 pF) should be routinely added to every voltage source in the circuit to avoid deterioration of high-frequency behavior and unnecessary delays in transistors in the common-gate configuration. The decision to remove them in the final design steps depends on simulation results, especially on transient simulations.

Both input and current mirror transistors have a large area. This results in large equivalent parasitic capacitances at the folded cascode input nodes a and b (see fig. 6-2), with $C_a \approx C_b = C_{gd1} + C_{db1} + C_{gd3} + C_{db3} + C_{gs7}$. This capacitance, in combination with the source resistance $1/g_{m7}$ of M_7 (or $1/g_{m9}$ of M_9), creates an additional pole in the signal transfer path and can convert the folded cascode into a gain stage if $g_{m7}/C_a < 2GBW$.

Hence, to preserve good dynamic properties of the OpAmp, the transconductances g_{m7} and g_{m9} of M_7 and M_9 should be high, but the area of each transistor M_7 or M_9 should be small. The transistors M_7 and M_9 should have minimal gate length and be wide enough to operate with $V_{gs} \approx V_{TH}$. Any further increase of transistor width does not noticeably improve $g_{m7,9}$ and only increases the parasitic capacitances. The gain boost using the scheme of fig. 4-7 will also decrease the source resistance and improve the high-frequency behavior of M_7 or M_9.

During the OpAmp slewing or during sharp changes of the input voltages, the tail current of the input stage and the capacitive currents through C_{gd} of the input devices flow through the folded cascode inputs (fig.6-9). The values of these currents can exceed I_F.

Any realistic value of current can be sourced through (flow into) the input *b*. Such current acts to turn off M_6 by increasing the potential at *b*, so M_2 is turned on by I_F and runs (sinks) the current I_F and all required extra current.

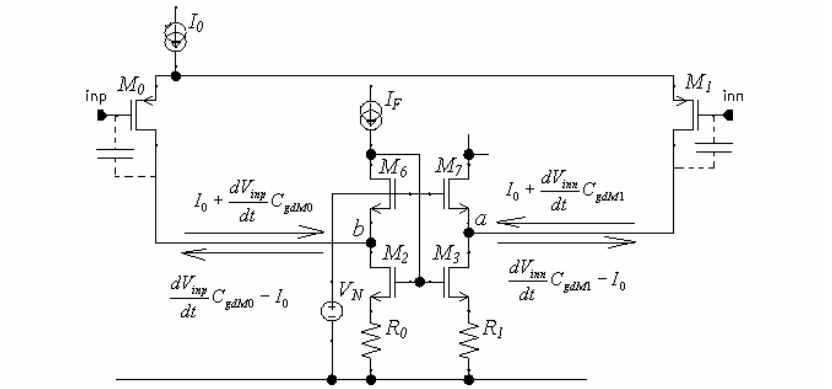


Figure 6-9. Effects of the large current through folded cascode inputs

Any value of the sinking current can flow from the input *a*. This happens when it is necessary to charge the capacitance C_{gd} of M_1 . In this case, the source of M_7 acts as the voltage clamp in common-gate configuration, providing the required extra charge current for the capacitance.

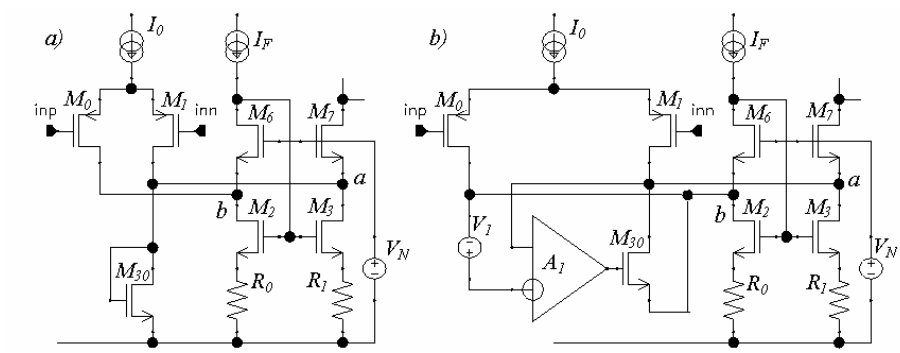


Figure 6-10. Voltage clamp at the folded cascode input

Sinking a large current from the input *b* can cause a drop of potential at this input below the negative rail. This results in activation of the parasitic body diode of M_2. Layout precautions (guard ring, spacing) must be taken to prevent a latch-up or other detrimental effect. In most layouts, a guard ring is

already in place in order to improve the matching of M_2, M_3, and the drop of potential of input *b* may even stay unnoticed.

Sourcing a current that exceeds I_F through the input *a* results in the shutdown of M_7 and an uncontrollable rise of the voltage at this input. To provide a faster recovery time and prevent a delay after the end of the transient (and possible conditional instability), the voltage at this input should be clamped. An example of such clamping is shown in fig. 6-10a.

The diode-connected transistor M_{30} in the circuit of fig. 6-10a limits the voltage rise at the input *b* by the value that is somewhat above V_{TH}. It may leak some current during normal operation, if V_a is close to V_{TH}, increasing the OpAmp offset. In this case, a dedicated feedback (always feedback!) circuit (fig. 6-10b) will help [44]. This feedback also sharpens the clamp operation. The voltage V_1 sets the threshold voltage of this loop. The amplifier A_1 can be implemented as a current-input amplifier (fig. 2-4), with the threshold voltage source realized as the offset voltage of two mismatched input transistors. When the voltage at the node *a* exceeds that at *b* more than the value of this offset, M_{30} is turned on and sources all the extra current to the input *b*.

6.3 Direct voltage gain boost in folded cascode

A folded cascode may be used to boost the DC gain of the OpAmp shown in fig. 6-2. The gain boost principle shown in fig. 4-7 is used in many of existing circuits. Some implementations of this approach are shown in fig. 6-11.

The circuit of fig. 6-11a [65] uses the amplifier consisting of the differential pair M_{10}/M_{11} and the current mirror M_{12}/M_{13}. The bulk of M_{10}/M_{11} should be connected to V_{DD} bus, so $V_{gsM10,M11} > V_{gsM7}$ and there is enough room for V_{dsM11}. The drain capacitance of large input transistor and current mirror transistors act as a capacitive load in the gain boost feedback loop. This loading does not create any stability problems with the single-stage amplifier of fig. 6-11a but it rules out use of any 2-stage amplifier implementation.

The gain and offset parameters of the gain boost amplifier itself can be improved if one decides to use the folded-cascode implementation of this boost amplifier (fig. 6-11b) [66]. The folded cascode transistors in this boost amplifier introduce some additional delay, and this amplifier may require frequency compensation.

Using a current-input amplifier (M_{10}/M_{11} as shown in fig. 6-11c) [67] is, probably, the fastest and simplest solution. It does, however, add some noise and offset to the OpAmp introduced by the noise and mismatch of the current sources I_1 and I_2.

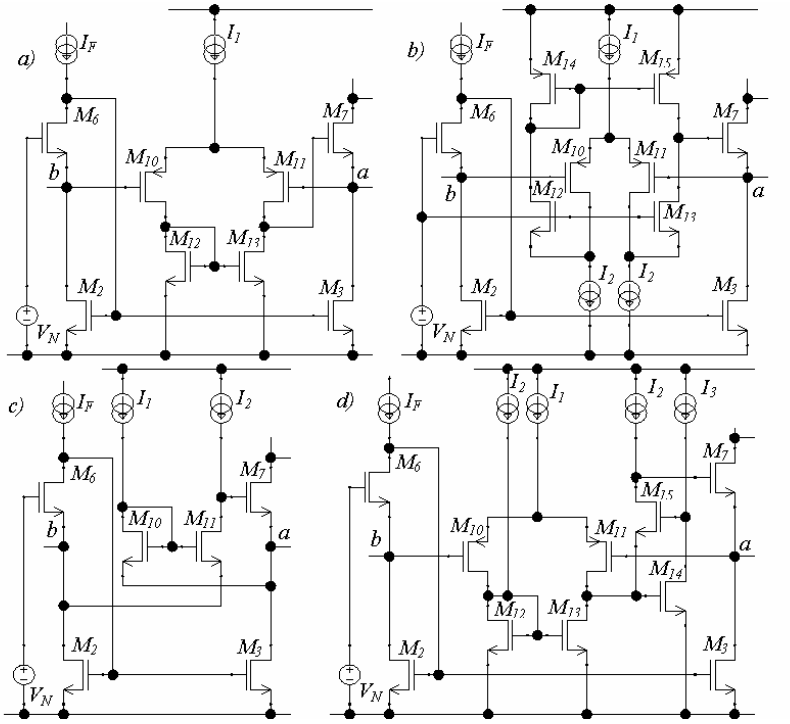


Figure 6-11. Implementations of the voltage gain boost in folded cascode

The gain boost amplifier of fig. 4-8c can also be used within the gain boost of the main OpAmp to achieve very high DC gains. This implementation is shown in fig. 6-11d.

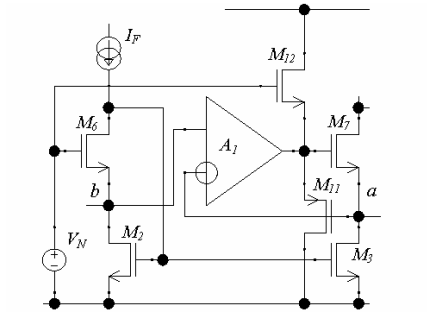


Figure 6-12. Limiting of the cascode transistor gate voltage during slewing

During OpAmp slewing, the boost amplifier is in nonlinear operation, and the gate voltage of M_7 may be of any value between the supply rails.

Clamps (transistors M_{11} and M_{12}) may be used (fig. 6-12) to limit this voltage and to eliminate delays in the settling of the OpAmp.

6.4 Voltage gain boost utilizing current mirrors

The simple current mirror circuit is frequently used as an active load of the differential pair. In subchapter 3.4 we considered improvement of the current source output impedance. Current sources with improved output impedance may also be used as active loads for differential pairs. An example of such an application is shown in fig. 6-13a. The signal flow graph of this circuit is shown in fig. 6-13b. With a simple current mirror this stage would have the gain of $g_{m1}Z_a$, where Z_a is the equivalent impedance at the node *a*. Adding the amplifier A_1 increases this gain approximately by a factor of $A_1 g_{m2} Z_a$, where g_{m2} is the transconductance of mirror transistors M_2, M_3.

The negative feedback loop M_2-*b*-A_1 ensures that the circuit does not have pole-zero doublets bringing a slow settling.

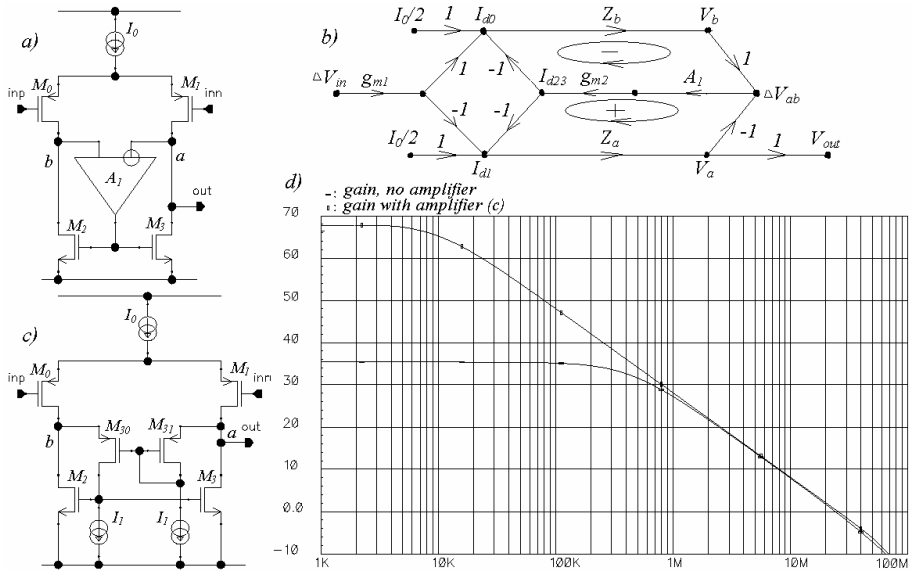


Figure 6-13. Gain boost employing the input stage current mirror

Fig. 6-13c shows the implementation of this circuit using the current-input amplifier M_{30}/M_{31}. The simulation results in fig. 6-13d compare the open loop voltage gain of the circuit of fig. 6-13c and of the input stage without gain boost. This simulation uses the models of transistors in CMOS generic 0.6um process ($M_{0,1}$ = 100/0.6, $M_{2,3}$ = 100/2, $M_{30,31}$ = 20/0.6, I_0 = 10 uA, I_1 = 1 uA, C_L=1 pF).

Voltage-input implementations of A_1 add an additional gain stage in the negative feedback loop and would require compensation.

The gain boost structure used in the input stage may also be used in the cascoded current mirror. The gain boost circuits in this case may also reuse the current mirror components.

Four possible implementations of this approach in application to the folded cascode current mirror are shown in fig. 6-14 [68].

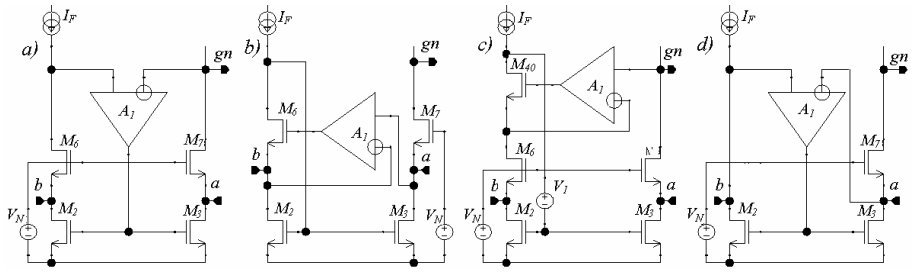


Figure 6-14. Structure 6-13 adapted for using in the folded cascode circuit

Fig. 6-14a shows the straightforward modification of the folded cascode. In the circuit of fig. 6-14b the transistor M_6 is reused as part of the feedback amplifier A_1. In the circuit of fig. 6-14c a dedicated transistor M_{40} is added to A_1. This circuit may require a voltage shift source V_1 to provide sufficient headroom voltage V_{dsM40}. The circuit of fig. 6-14d is equivalent to that of fig. 6-14a, but transistor M_6 now becomes optional and has been removed.

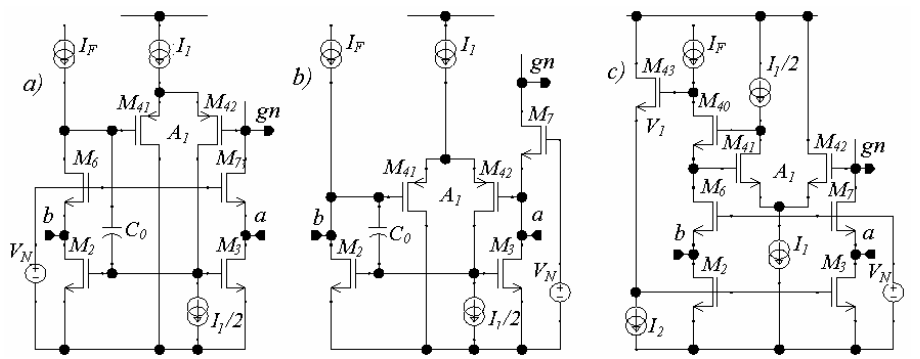


Figure 6-15. Current mirror voltage gain boost in folded cascode

The current mirror modification with gain boost circuit may also be used as a load in the non-complementary folded cascode. An example of such an application in the bipolar OpAmp was shown in fig. 4-22.

Two implementations of the current mirror gain boost of fig. 6-14a and 6-14c in application to the folded cascode are shown in fig. 6-15. Any one of the amplifiers shown in fig. 6-11 can be used for realization of the gain boost in the circuit given in fig. 6-14b.

The circuits of fig 6-15a and b (corresponding to the structure shown in fig. 6-14a) employ a differential pair M_{41}/M_{42} as the amplifier A_1. In case of a low supply voltage, the cascoding transistor M_6 may be omitted (fig. 6-15b). The feedback loop M_{41}-M_{42}-M_2-M_{41} may have a two gain stage amplification structure and need the capacitor C_0 for compensation. Using the current source $I_1/2$ instead of the current mirror transistor also improves the circuit stability.

The circuit of fig 6-15c (corresponding to the structure shown in fig. 6-14c) uses M_{41}/M_{42} as A_1 and the voltage shift V_1 provided by V_{gs} of M_{43}. The gain boost amplifier is operational when the voltage at the node *gn* is $V_{gs} < V_{gn} < 2V_{gs}$.

6.5 Voltage follower

Voltage follower is the current amplification stage. It is very useful in bipolar OpAmps where it is necessary to have 4 to 5 current amplification stages while preserving the structure of one or two voltage gain stage. It can also be used in CMOS power OpAmps to drive the output devices having large gate capacitance while avoiding any speed degradation of the overall amplifier.

A voltage follower/shifter can also be used as an intermediate stage between input and output. It may be considered as an alternative to the folded cascode in two gain stage OpAmp structure.

The most useful application of the follower is, probably, as the output stage in the OpAmps with single gain stage structure, especially when the rail to rail output is not required.

The simplest follower is realized as the transistor in the common-drain (common-collector) configuration. Its output impedance is $R_{out} = 1/g_m$. This impedance can be reduced using current gain boost circuits. The basic structure of this approach was shown in fig. 4-9 and some implementations were given in fig. 4-10. Fig. 6-16 shows additional implementations of the current gain boost.

The voltage follower of fig. 6-16a has received a lot of attention recently [70]. It has, however, very limited input voltage range, and this deficiency is eliminated in the circuit of fig. 6-16b. Further modification of this circuit (fig. 6-16c) allows it to sink and source large load currents. The shoot-through, or quiescent, current of M_3/M_1 is defined by the voltage source V_0.

The follower characteristics can be further improved using an amplifier which provides the voltage gain boost in the current gain boost loop (fig. 6-

16d). The implementation of this gain boost in the circuit with high sink/source output capability may be reduced to one transistor (fig. 6-16e).

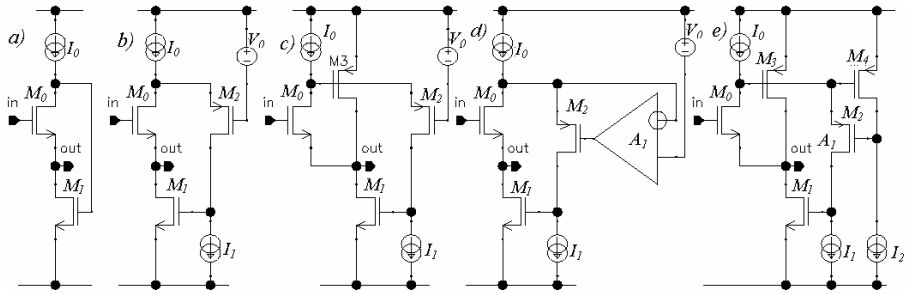


Figure 6-16. Voltage followers

Using the voltage follower as the bipolar OpAmp intermediate stage is shown in fig. 6-17.

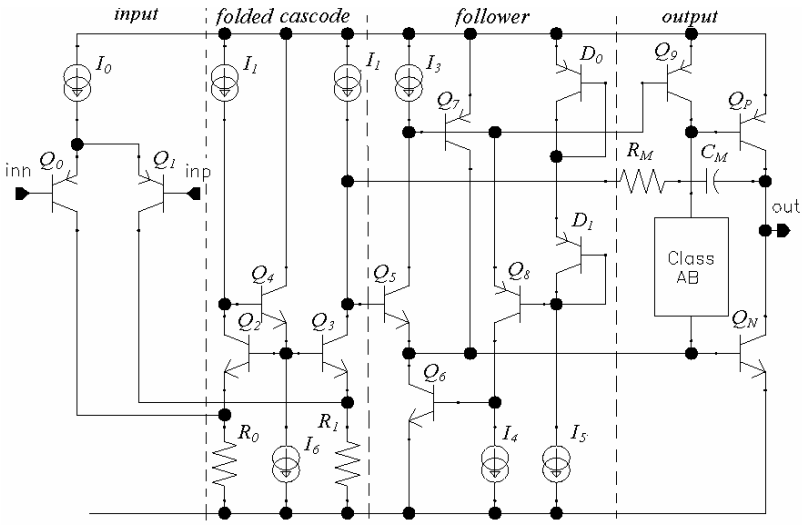


Figure 6-17. Voltage follower as bipolar OpAmp intermediate stage

This OpAmp has two voltage gain stage structure (the input differential pair Q_0/Q_1 and the output transistors Q_N/Q_P), yet includes four equivalent current amplification stages (Q_0/Q_1, Q_P/Q_N, and Q_5 with its own gain boost Q_7 and Q_8, Q_6). The voltage follower Q_5 has the same structure as the circuit of fig. 6-16c (current gain boost with high sink/source capability). The voltage V_0 is formed by the diodes D_0/D_1 and I_4. If the OpAmp open-loop voltage gain is not sufficient, then the folded cascode should be augmented

with the voltage gain boost as discussed in this and the previous chapter. If more current gain is necessary, then an additional gain boost in the voltage follower Q_5-Q_7 can be used as shown in fig. 6-16e. Other details of this amplifier will be clarified in chapter 7.

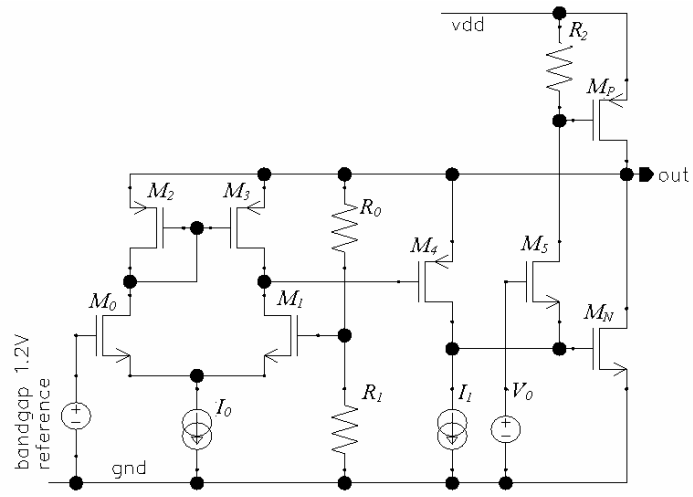


Figure 6-18. Follower as a low drop-out voltage regulator output stage

Fig. 6-18 shows the output stage of the low drop-out voltage regulator or reference. The single gain stage structure (M_0/M_1 with the follower M_4) ensures the regulator stability with any load capacitance [91]. The follower M_4 has the current gain boost circuit M_5/M_N/M_P with high sink/source capability.

Chapter 7

CLASS AB OUTPUT STAGE

The OpAmp output stage should provide minimal non-linear distortions, low current consumption when the amplifier is idle, and high output current capability when loaded. In addition, this stage should provide the required capacitive load tolerance and not limit the OpAmp slew rate.

The load capacitance introduces the pole with the frequency of $\omega_L = g_{mout} / C_L$ [71] in the OpAmp open loop transfer function. For a chosen OpAmp bandwidth, the output stage transconductance, g_{mout}, should be sufficient so that this pole does not cause the OpAmp to oscillate in the required load capacitor range. Except for very fast amplifiers for video applications, the industry standard requires that any OpAmp should tolerate at least 100 pF of load capacitance. This figure defines the minimum value of g_{mout}.

In a correctly designed OpAmp the distortions should be defined by the output stage as long as input stage is in linear operation (i.e., the OpAmp is not in the slewing mode). Then, in the feedback loop, the distortions will lessen with an increase of the open-loop gain.

The distortions are measured in the test systems as deflections of the input-referred error from the straight line. A low-gain OpAmp may have a large input error and low distortions at the same time if it's output stage amplification is constant. Observation of the amplifier's input-referred error may give indication on the presence and value of distortions. A typical dependency of the OpAmp input-referred error as a function of the output voltage is shown in fig. 7-1. Any sharp change in the slope of this error dependence indicates a change of the amplifier open-loop gain at this point and the occurrence of nonlinear distortions.

The error dependence also indicates the character of the distortions. A non-symmetric error curve indicates the second-order distortions in the

output signal, while sharp changes of the error around zero indicates the presence of the high-order harmonics, etc. [116].

This dependence of error from the output voltage may look different for different velocities of the signal change. Both the value and spectrum distribution of distortions depend on the signal frequency and amplitude. In applications, some distortions may be more acceptable than others. For example, a second harmonic of the vacuum tube audio amplifier is more easily tolerated than the higher-order harmonic distortions typical for class AB transistor stages.

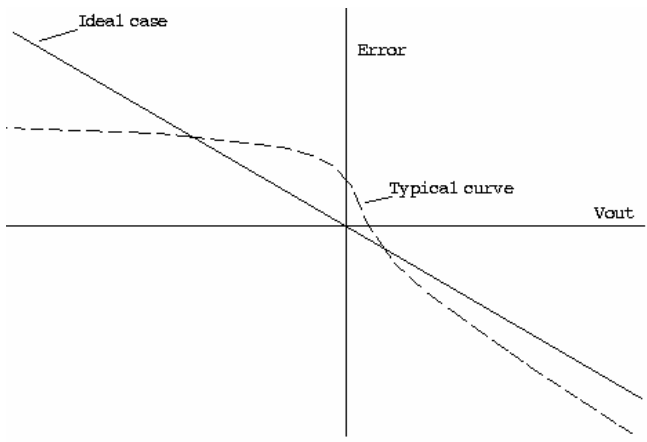


Figure 7-1. Ideal and typical dependencies of input error on output voltage

The transconductance of the output stage rises with the load current. It is, perhaps, feasible to make it constant using local feedback loops, and, thus, to design a very low distortion amplifier.

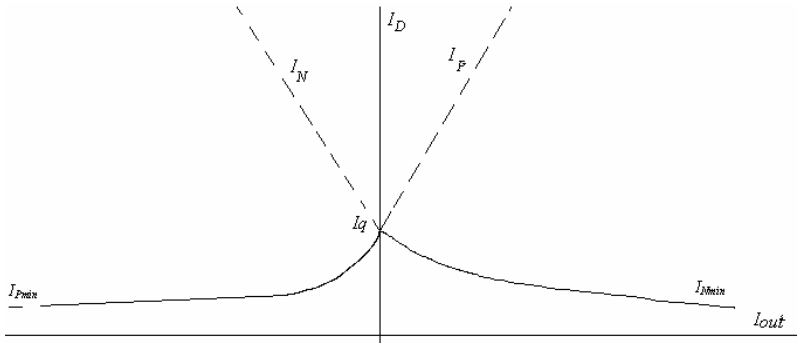


Figure 7-2. Currents of the output transistors versus output current in class AB stage

The standard approach, nevertheless, is different. The output stage devices are used in the mode that is called class AB operation. The typical dependencies of the output transistor currents in the class AB stage are shown in fig. 7-2. The designer tries to guarantee the minimum g_{mout} by controlling the quiescent current, I_q, of the output devices in no load condition.

The value of I_q and the dimensions of the output devices define the minimum value of g_{mout}. The exact values of I_{Nmin} and I_{Pmin} or the exact shape of the dependencies of the device currents versus I_{out}, especially after the cross-over point, are not important. Usually I_{Nmin} and I_{Pmin} should have non-zero values only in order to avoid the complete shutdown of the stage transistors and ensuing delays while turning them on.

7.1 Class AB stage structure

In the general case, a class AB output stage comprises of high-side and low-side power transistors and is characterized by two controlled variables, namely, the output voltage $V_{out}=(I_P - I_N)Z_L$, defined by the OpAmp input signal, and the quiescent current, $I_q = min(I_P, I_N)$.

It is rational to represent a class AB output stage by the general graph of differential structure with common-mode feedback (fig.7-3a). Indeed, we have two output transistors that supply two currents I_P, I_N that are subtracted to become the load current I_L. From the other side, these output transistors have their common quiescent current I_q and this should be controlled. This variable I_q is shown in this graph as the input variable, due to the fact that this value is frequently produced via other independent variables and one has to create the control system that takes care of both transistors simultaneously. In this case, one has to measure the currents I_P and I_N using the current sensors F_P and F_N, and to compare the average of these two currents with I_q and then use this in the feedback control loop.

This loop should have an amplifier with a negative gain $-G$ (we assume here a current amplifier). The output of this amplifier should be split using the distributor or splitter, providing two signals to the gate node variables I_{gP} and I_{gN} via the links E_1 and E_2. The input signal V_{in} should arrive to the same nodes via the transconductance amplifiers A_1 and A_2. If the gate node variables are currents, then the summation is done on the gate node resistances R_1 and R_2 and the gate driving voltage variables V_{gsP} and V_{gsN} could be obtained.

It is interesting to note that if the system satisfies the conditions $E_1=E_2$, $R_P g_{mP} = R_N g_{mN}$, then the operation of the common-mode feedback and I_q control does not vary the amplifier gain from the input to the load.

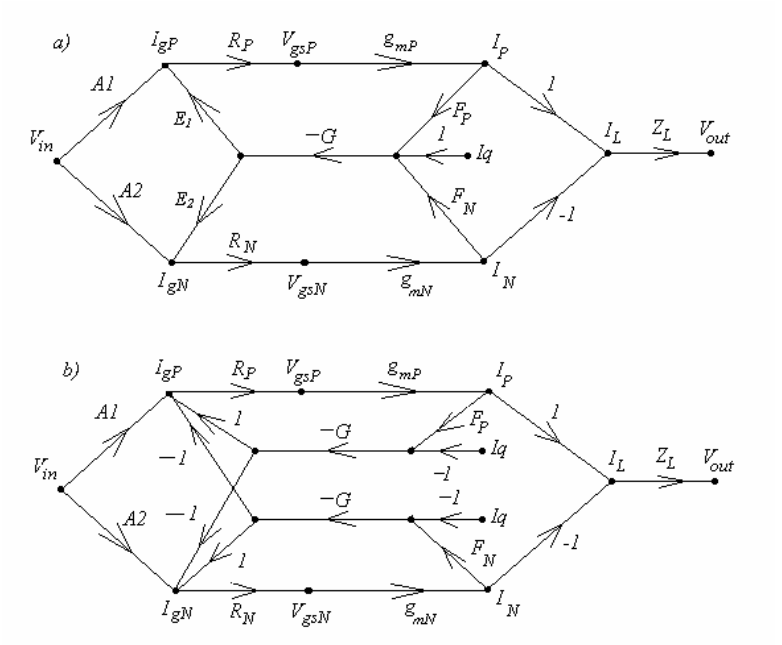


Figure 7-3. Class AB stage structure

For an ideal class AB linear stage, in addition to the previous conditions, one should require that A_1=A_2 and g_{mP} = g_{mN} = const. These conditions should be met for all output currents. In real circuits, g_{mN} and g_{mP} are changing with currents I_N and I_P which causes changes of the stage gain.

The equivalent graph transformations of the general structure (see Appendix) can be performed, and one of the useful variants is shown in fig. 7-3b (with E_1 = E_2 = *1*). Here the common link -*G* is divided in two separate links, and the splitters divide the output signal of each link into parts with opposite signals. An example of such a voltage splitter is a transistor with equal value resistors in the drain and source, where the output signals from these resistors will be of equal values and opposite sign. Another example of a splitter is the differential transistor pair, where the drain currents have opposite sign and their sum is constant.

Each output transistor, represented by its drain current in fig. 7-3b, has its own feedback loop controlling the quiescent current. These feedback loops do not affect the overall gain if $R_P g_{mP} = R_N g_{mN}$: all current generated in the quiescent current control loop and added to the gate of one transistor is subtracted from the gate of opposite transistor.

In addition, every particular implementation, while being represented generally by one of the graphs of fig. 7-3, may have additional local feedbacks or substructures, and/or some links may be absent.

The output devices may be realized using Darlington or composite transistors [10, pp. 144-148]. The links F_N and F_P may employ a logarithmic or other functional cells such as minimum selectors, etc. Each realization will result in a unique class AB stage behavior. However, the proposed class AB general structure is a useful guide for the design of high quality linear amplifiers and for a single-glance estimation of possible stage problems.

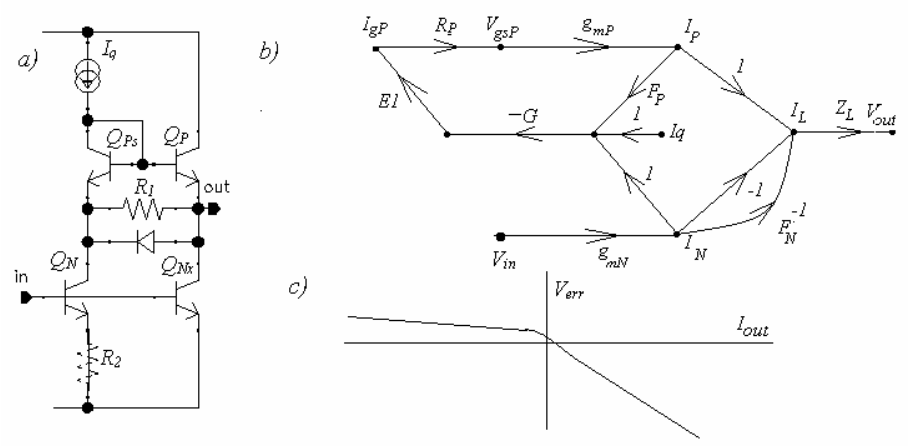


Figure 7-4. Example of class AB stage, its signal graph and error curve

In the amplifiers found in literature, some links may be missing. For example, the all n-p-n stage of fig. 7-4a was used in many industrial OpAmp output stages starting from the very first transistor amplifiers [78] and up to modern days [52, 79, 9 p. 395]. The stage is described by the signal graph of fig. 7-4b. In this circuit, the input voltage, V_{in}, sets the current I_N of Q_N (hence, here $A2$=1). When the current I_N is large then the voltage drop across R_1 (limited by the diode) is large as well and there is practically no current I_P through Q_P (the feedback loop is not in operation). When I_N decreases, the feedback loop, including R_1, Q_{PS} and I_q, is activated and controls the I_P current. Operation of this loop can be described by the links F_P (this non-linear function is defined by the dependence of the ratio of the currents I_{PS}, I_P and the voltage drop across R_1), by $-G$ and E_1 (both are unity gains here), and the equivalent input resistance, R_P, which depends on I_P.

In this structure, the links A_1 and E_2 are missing. This makes the structure non-symmetric, and the stage has different gains for sourcing and sinking currents, resulting in non-symmetric input error dependence as shown in fig. 7-4c.

Transistor Q_N may be supplemented by Q_{NX} and R_2. This helps to forward-compensate the inherent stage nonlinearity. From the structural design point of view, Q_{NX} and R_2 add another link F_N^{-1} between I_N and I_L.

This link is nonlinear and symmetrical to the link F_P if $R_1=R_2$ and emitter area ratios $Q_{Ns}/Q_N = Q_P/Q_{Ps}$. As a result, the stage gain when sinking the load current becomes somewhat matching the stage gain when sourcing it (via the loop G-R_1-F_P).

One of the most known bipolar output stages is shown in fig. 7-5a [9, subchapter 5.5.3]. The signal graph of this stage (fig. 7-5b) is nearly identical to the graph of fig. 7-3a, with all links present except of A_1.

Here, the nonlinear links F_N and F_P are the logarithmic dependencies of the voltages V_{be} on their corresponding collector currents. The reference quiescent current, I_q, is defined here indirectly via comparison of ($V_{beQN} + |V_{beQP}|$) with the voltage drop across the diodes Q_{Ps} and Q_{Ns}. The gain, G, is defined by the equivalent impedance at the base of Q_P. One can see that in this stage

$$|V_{beQP}| + V_{beQN} = |V_{beQPs}| + V_{beQNs}$$

Assuming that Q_P matches Q_{Ps} and Q_N matches Q_{Ns} (taking in account their emitter area ratio N), then

$$I_N \cdot |I_P| = (NI_q)^2$$

The error curve of the amplifier with this stage is symmetric and the slope at each point is defined by the dependence of the stage transconductance g_m on current (fig. 7-5c).

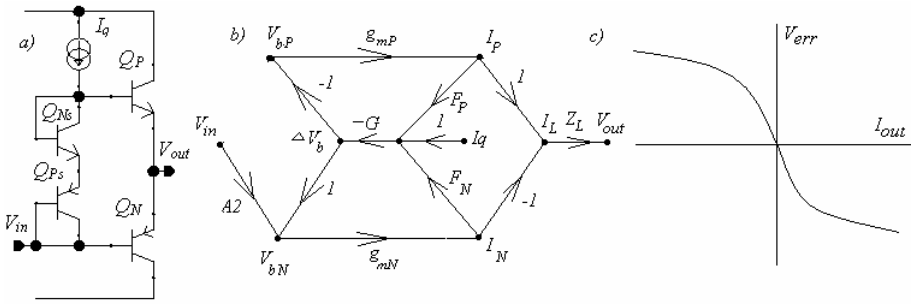


Figure 7-5. Output stage of 741 OpAmp

Numerous implementations of the structure given by fig. 7-3a for bipolar and CMOS output stages, under the name of "feedback class-AB biasing" can be found in [10, 22, 45] and other works of Prof. J. Huijsing and his group.

The stage shown in fig. 7-6a is a frequently used rail-to-rail CMOS output stage [72; 10, p. 159]. The signal graph of this stage (fig. 7-6c) has the structure of fig. 7-3b. The error curve of the OpAmp with this stage is similar to the curve of fig. 7-5c.

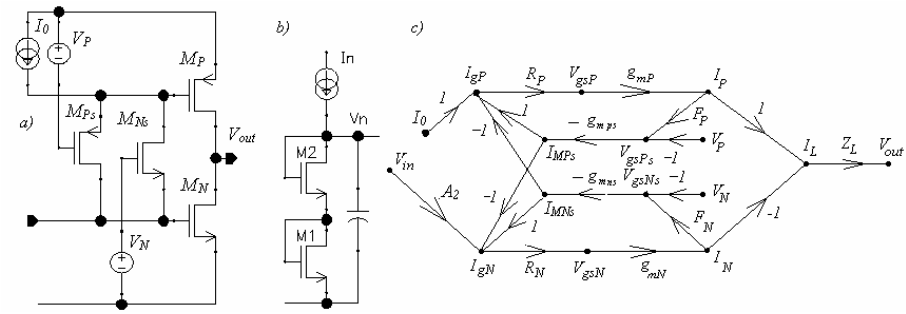


Figure 7-6. Output stage with the separate common-mode feedbacks

Class AB operation in this circuit is controlled by two feedback loops. Transistor M_{Ns} provides the local feedback loop for M_N, adding the current I_{MNs} when the current of M_N is decreases (links F_N and g_{mns} of fig. 7-3c) and acting to preserve the voltage at the gate of M_N and, simultaneously, to turn the transistor M_P on (the source and drain of M_{Ns} realize the distributor function operation). The operation of M_{Ps} is similar in relation to M_P. The non-linear function $F_N = V_{gsNs}(I_N)$ for M_N may be obtained from the relationship $V_N = V_{gsNs} + V_{gsN}$ ($F_P = V_{gsPs}(I_P)$); for M_P it is obtained in a similar way.

The voltage sources V_N and V_P are usually implemented using two diode-connected transistors (fig. 7-6b). Then the currents I_0 and I_n define I_q of the stage, providing that M_1 is matched with M_N, and M_2 with M_{NS} (similar implementation and matching requirement are done for the high-side source V_P).

These last two examples illustrate "good" circuits, in the structural design definition: as the stage I_q is important, it is controlled by an efficient feedback loop. Circuits without such feedback should generally be considered "bad" and be discarded. Such "bad" circuits appearing in the literature (for example, [61, 117, 119]) usually have problems that do not allow them to mature into industrial products: speed penalty due to the additional gain stages in the signal path, excessive non-linearity, etc.

The speed required from the Class AB stage depends on how the signal is delivered from the input stage to the gates of output devices.

If both links, A_1 and A_2, of fig. 7-3 are present and identical, then the class AB control can be slow. In this case, the increase of current at the gate of one output transistor goes simultaneously with the decrease of current at another gate. The delay in the two-loop control system of this stage should still be small enough to prevent the complete shutdown of one of the output transistors.

If the links $A_1$and A_2 are not identical or one of them is missing altogether, like in fig. 7-4, then the class AB circuit is in the main signal gain path and needs to be at least 3-5 times faster than the whole OpAmp. Delay in this stage would worsen the overall stability and/or cause the complicated settling behavior when settling of the class AB circuit ends later than overall OpAmp settling.

7.2 Generation and improvement of class AB circuits

A set of class AB circuits for any process can be generated implementing the general structures of fig 7-3 with the toolbox of elementary cells. The cells specific for Class AB circuits should include current sensors, amplifiers and the cells with specific nonlinear functions like minimum selector functions, signal splitters, etc.

Some current sensors implementations are shown in fig. 7-7.

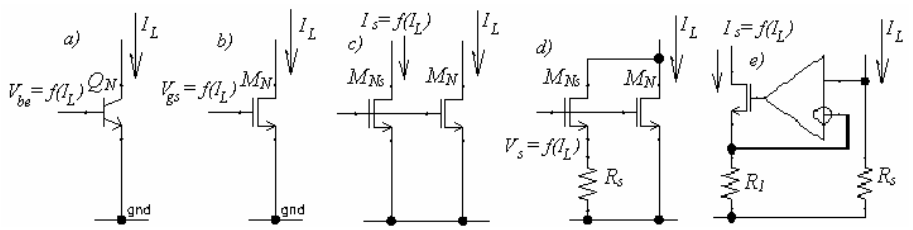


Figure 7-7. Current sensors

The voltages V_{be} and V_{gs} are monotonic functions of the collector or drain current, and they can be used for measurement of current (7-7a, b) [45]. A matching transistor of a small area with the gate-source connected in parallel to the output device is an accurate current sensor (fig. 7-7c). To improve power efficiency, a resistor, R_s, can be inserted in the sensor source wire to decrease the sensor current, I_s, even further (fig. 7-7d). The voltage drop on this resistor can also be used for current sensing. The direct measurement using a small resistor in the current bus is shown in fig. 7-7e where $I_s=I_L R_s/R_1$.

Single-stage and current input amplifiers are preferable in the class AB circuit feedback loops as well as in other local feedback loops. This helps to avoid the compensation problems (subchapter 2-4).

Nonlinear cells should provide a function similar to the minimum selector circuits [45] with their output signal $F_{1,2} \approx \min(in_1, in_2)$. The requirements for these cells are the following:

- output is monotonic versus either of inputs,

- good accuracy versus temperature and process variations at the point where both inputs are equal,
- output does not go to zero when both inputs are non-zero.

These nonlinear cells in bipolar amplifiers often employ the logarithmic dependence of V_{be} on the collector current as in the circuit of fig. 7-5. The voltage difference between the bases of Q_N and Q_P is defined as the product of their collector currents. In the rail-to-rail bipolar stage, one of the V_{be} voltages can be voltage-shifted as shown in fig. 7-8a [10, p. 169].

Other realizations of nonlinear cells are shown in fig. 7-8b [80], 7-8c [45, p. 79] and 7-8d [32].

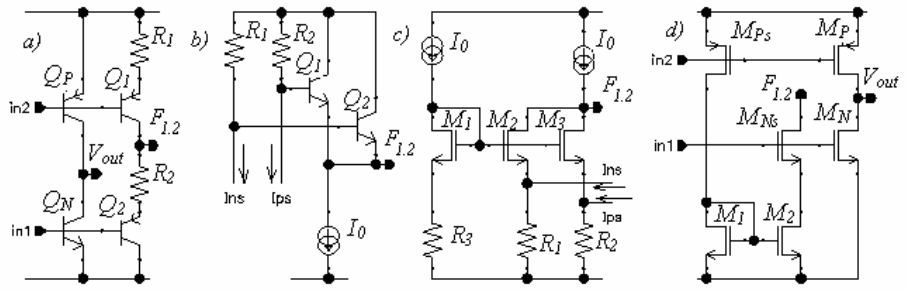


Figure 7-8. Nonlinear cells for class AB circuits

The cells of fig. 7-8b, 7-8c can employ either bipolar or CMOS transistors. The cells shown in this figure do not exhaust all possible solutions. More circuits can be found in [32, 82, 83].

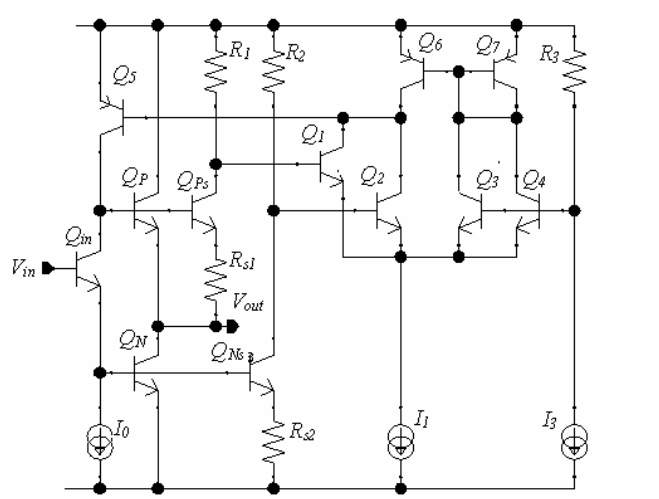


Figure 7-9. NPN class AB output stage

An example of a class AB output stage for a micropower bipolar amplifier with NPN output devices is shown in fig. 7-9 [81, 88]. This stage was designed by direct implementation of the structure shown in fig. 7-3a. The stage uses the current sensors of fig. 7-7c (Q_{Ps}, Q_{Ns}), the nonlinear cell of fig. 7-8b (R_1, R_2, Q_1, Q_2) and the differential stage Q_1-Q_4 as a feedback loop amplifier with the current mirror Q_6/Q_7 and transistor Q_5 providing additional gain. Transistors Q_1 and Q_2 of the non-linear cell are reused as part of the feedback amplifier. The quiescent current reference is formed by I_3 and R_3, where R_3 should match $R_1 = R_2$.

Another example of a class AB output stage is the rail-to-rail CMOS output stage shown in fig. 7-10a. This Class AB stage is also designed by direct implementation of the structure shown in fig 7-3a. It consists of the current sensors of fig 7-7c (M_{Ns}, M_{Ps}), the nonlinear unit of fig. 7-8d and the transistor M_A as the feedback amplifier G.

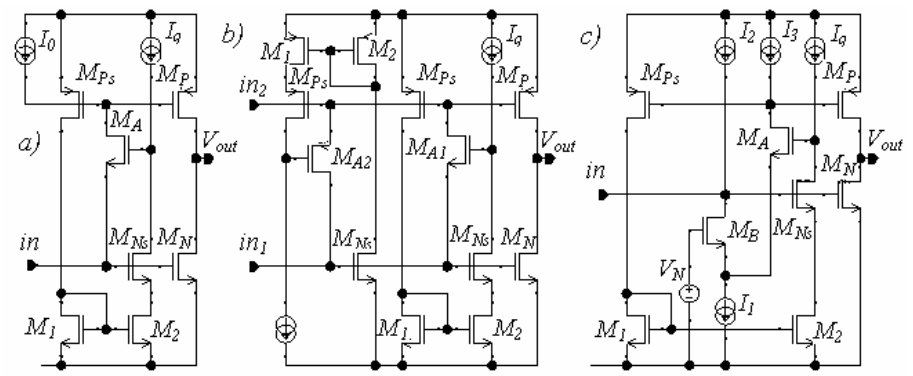


Figure 7-10. CMOS rail-to-rail class AB stages

If the input signal is applied to both gates of M_N and M_P, then the class AB circuit may be symmetrical as shown in fig. 7-10b.

For lower supply voltage, the amplifier M_A of fig. 7-10a can be replaced by the differential stage M_A, M_B (fig. 7-10c), or a more complicated low-voltage amplifier.

Numerous realizations of the structure 7-3a for rail-to-rail output stages can be found in [10, pp. 169-176, 45, subchapter 3.5].

The implementations of the structure shown in fig. 7-3b do not require a minimum-like function units as shown in fig. 7-8, but they need a cells with a transfer function similar to limiting rectifiers. The transfer characteristics F_N and F_P of these cells allow one to switch off one of the local feedback loops when the current of one of the output devices increases above the I_q level. At this time the opposite side transistor current drops significantly

below I_q and this transistor is not providing gain any more. Both feedback amplifiers G should have a differential output.

The simplest current limiting cell is a transistor in the common-gate configuration (fig. 7-11a), where the current can flow in only one direction (it can be called a current rectifier). The simplest voltage rectifier is a diode-connected transistor (fig. 7-11b). Accuracy of the limitation can be improved using an amplifier (fig. 7-11c and 7-11d), with the penalty of amplifier delay. The current-input amplifier of fig. 7-11e has a smaller delay but its biasing current runs through the load.

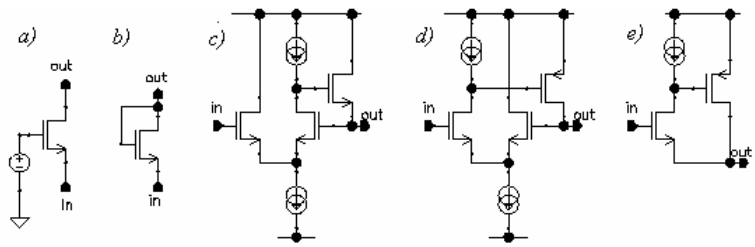


Figure 7-11. Limiting cells

In the stage of fig. 7-6a, the output devices themselves provide the current sensing (as it is suggested by fig. 7-7b). Transistors M_{Ns} and M_{Ps} also provide a limiting function as in fig. 7-11a as well as switching (when an input voltage turns on M_{Ns} and it starts to drive M_P, the transistor M_{Ps} will be off). Both M_{Ns} and M_{Ps} are reused as feedback amplifiers with their source and drain acting as differential outputs (splitters). This is probably one of the most elegant implementations of the class AB output stage. It does, however, have the following drawbacks:

- large minimum supply voltage ($V_{dd} \geq V_{gsP} + V_{gsN} + V_{sat} \approx 2.5$V);
- strong dependence of I_q on the supply voltage, especially using short-channel transistors;
- low input impedance of the stage caused by the drain-body leakage of M_{Ns} or M_{Ps} when implemented in a single-well process and with large supply voltage.

This low input impedance can limit the open-loop gain of a two-stage amplifier (for example, in the OpAmp of fig. 4-21). Transistors M_{Ns} and M_{Ps} of fig. 7-6a may have large drain voltages causing impact ionization in the drain region and high drain to body currents. If the body can be connected to the source then this leakage does not create any problems. But in a single-well process, the body of at least one transistor is connected to the substrate. For instance, in most of the 0.6 um processes this leakage would limit the open-loop gain to 80-90 dB in a two-stage OpAmp with a 5V supply voltage.

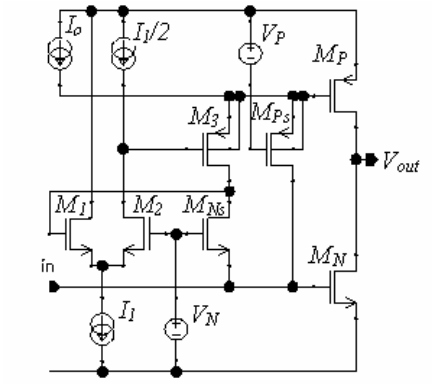


Figure 7-12. Improvement of the input impedance of Class AB output stage

A solution to this problem is shown in fig. 7-12 [35]. The stage has the same structure as that of fig. 7-6, yet includes an additional dedicated local feedback loop. This loop comprises the amplifier M_1, M_2, M_3 and controls $V_{dMns} \approx V_N$ eliminating the impact ionization effects in M_{Ns}. The bulks of M_{Ps} and M_3 are connected to their sources, so the drain leakage in these transistors does not affect the equivalent input resistance that is limited now by the output resistance of the signal source and the current source I_0.

The poles in the OpAmp transfer function, which are located far in the left half-plane and limit the OpAmp maximum achievable bandwidth, are defined largely by the size of output devices. For example, increase of the M_N and M_P channel length from 0.6 um to 1.8 um, while increasing the width to keep the same current capability, decreases the achievable *GBW* by 5 times. This means that in order to improve the OpAmp speed, the output devices should be as short as possible.

The V_{gs} voltage in short-channel transistors is strongly dependent on V_{ds}. This correlation causes a significant error in I_q measurements. This is valid not only for the stage of fig. 7-6, but for any other stage using the same current sensing approach, i.e. where the output devices also act as the current measuring devices. As a result, the variation of I_q versus supply voltage in the stage of fig. 7-6 can be up to 2 times different when the transistors M_N and M_P have a length of L=0.6 um. This error is much larger for finer geometries.

This problem can be solved by controlling the reference voltage sources V_N and V_P. The circuit of fig. 7-13 [118] shows V_N generation depending on the V_{ds} variation in the sensor devices.

The voltage $V_N = (V_{gsM1} + V_{gsM2})$ is generated here in the feedback loop M_1-M_3-M_2. In this loop, M_1 matches M_N, and M_2 matches M_{Ns}. The voltage V_{dsM1} now tracks V_{dsMN} as the gate of M_3 is connected to V_{out}. The transistor

M_4 prevents the current source $2I_1$ from saturation when V_{out} approaches the positive rail. The currents through M_1 and M_2 are equal to I_1.

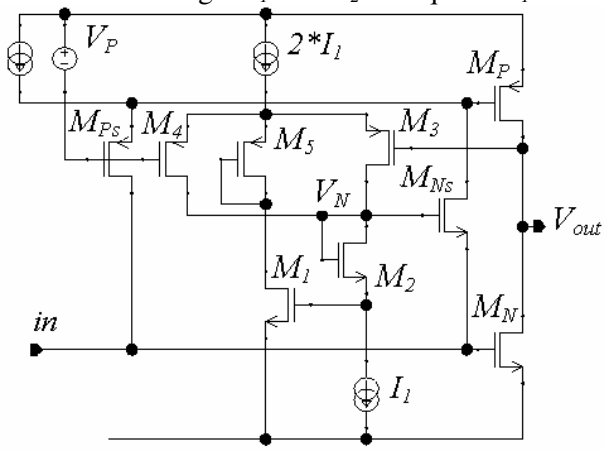


Figure 7-13. Improvement of the I_q versus supply voltage stability

The mismatch of the V_{ds} voltages for M_2 and M_{Ns} does not produce any significant error as these transistors may have a longer channel and therefore less variation of V_{gs} versus V_{ds}.

Using the circuit 7-13 for the generation of both V_N and V_P voltages ensures a stability of I_q versus supply variations of better than 5%, even with a 0.5 um channel length of the output devices [87].

Fig. 7-14 shows another example of the class AB output stage implemented in accordance with the structure of fig 7-3b. The stage is capable of operating with the supply voltage of $V_{dd} \geq V_{gs} + 2V_{sat} \approx 1.3$ V [85]. The I_q control feedback loop for the output transistor M_N consists of the current sensor M_{Ns}, reference current I_{Nq}, diode-connected limiting transistor D_1 and amplifier with differential output at the drains of M_3 and M_4. The resistor R_0 defines the gate potential of M_2 when this feedback loop is not in operation, and decreases the excessive loop gain. Transistor M_4 is wider than M_3, and this mismatch defines the control range.

When the drain current of M_{Ns} drops below I_{Nq}, the difference of these currents starts to flow through D_2 and creates a voltage drop across R_0. The amplifier M_3/M_4 steers more current to the gates of M_N/M_{Ns}. This increases the gate potential of the M_N/M_{Ns} parallel connection. Hence, the value of the M_{Ns} current is kept stable, and defined by I_{Nq}, R_0 and the size ratio between M_3 and M_4. One may notice that any current that is added or subtracted from the gate of M_N is subtracted or added to the gate of M_P. Hence, the operation of this loop does not affect the overall stage gain.

The operation of the feedback loop for M_P (this loop includes $M_{PS} - D_2 - M_1/M_2$) is identical as for M_N.

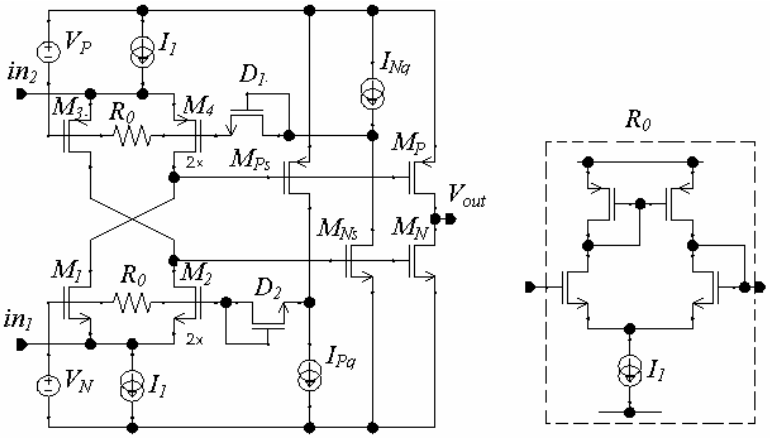


Figure 7-14. Low voltage supply class AB output stage with two separate feedback loops

This stage was used in the nanopower ($I_q \approx$ 1uA) OpAmp [86]. To avoid large resistor values and the loading of V_1 and V_2, R_0 was implemented as an active circuit shown in fig. 7-14b, which is providing $R_0 = 1/g_m$. Transistors M_2 and M_3 were reused as part of the folded cascode stage preceding this class AB stage. Currents I_1 are the outputs of the current mirrors of this folded cascode.

The examples considered above demonstrate the main points of the structural design methodology procedure: start from the general structure and its equivalents; generate a set of circuit solutions based on the elementary cell library; improve the important parameters using additional local feedback loops.

Chapter 8

SPECIAL FUNCTIONS

OpAmps comply with their performance specifications when their input signal, supply voltage and load environment fall within specified boundaries. Operation far outside of these boundaries may cause irreversible damage, such as, for example, OpAmp destruction by large supply voltage. If the violation of the operating boundaries is less severe, then it usually results in various recovery delays: start-up and input overload delays, the delay caused by limited slew rate, etc.

Every OpAmp should include a number of circuits that guard, limit or extend the boundaries of linear operation. Circuits for protecting of the input stage were discussed in chapter 5. In addition, an amplifier capable of overheating itself with a short circuit load should have a temperature shutdown and/or current limiting circuit. Additionally, OpAmp may be equipped with slew rate boost, with circuits reducing overload recovery delay, and with fast start-up. This chapter discusses these additional protection and performance enhancement measures and circuits.

8.1 Startup and shutdown

Despite of proliferation of the OpAmps with 1uA of quiescent current, not every OpAmp can be powered by the moonlight. Shutdown became a very common feature for the OpAmps used in battery-powered applications, allowing to extend the battery life. During shutdown the OpAmp should consume a very small supply current, typically less than 10 uA, and have output in the high-impedance state.

The shutdown control signal usually comes from the digital part of the system. The digital and analog ground lines are almost always at different potentials. A differential shutdown input is often required, with *OFF* and

OFF_GND pins, and with a signal threshold of 1 to 2 V, to be compatible with CMOS logic. A circuit that can accept a differential input signal with different DC levels can be implemented as shown in fig. 8-1.

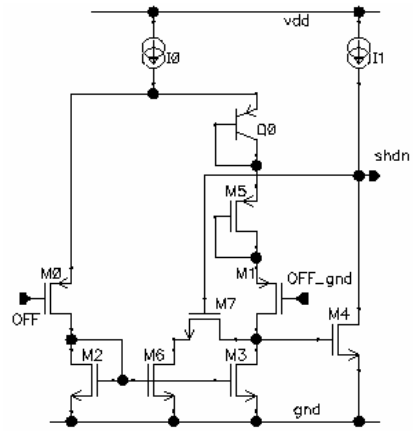


Figure 8-1. Differential shutdown input with hysteresis

This circuit comprises the comparator including transistors M_0 to M_4. The diodes M_5 and Q_0 set the offset $V_{thrs}= V_{gsM5}+V_{be} \approx 1.5$ V. The comparator is switched from one state to another by the difference of logic level signals which are present at the comparator inputs. The threshold hysteresis is created by the positive feedback transistor M_6 and M_7 which prevent multiple switching in case of noisy input. The bias currents, I_0 and I_1, can be very small and may be derived even from the start-up source.

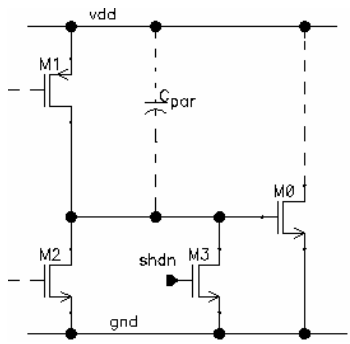


Figure 8-2. Prevention of the shoot-through currents caused by capacitive gate charging

The OpAmp can be in shutdown, yet the application system may be running. This means that the supply, input and output voltages may still

changing. The currents caused by these variations can be integrated by the capacitances on internal nodes and may cause uncontrollable shoot-through currents. To prevent them, every floating transistor gate in the OpAmp should be, during shutdown, connected to the supply rail or to the source of the subsequent transistor (fig. 8-2).

During shutdown, the gate voltage of M_0 in the circuit of fig. 8-2 may be undefined if both M_1 and M_2 are off. This voltage can be of any value, and if a current flow path is present between the V_{DD} supply rail and the drain of M_0 then a large shoot-through current is possible. The switch M_3, controlled by the shutdown logic signal, prevents this effect.

An OpAmp in shutdown still requires protection from destruction by the input or output voltages, so the clamps for the input voltage limiting (subchapter 5.4) should operate even in this mode. The OpAmp could be loaded with an inductive load, for example, a long cable when load is located in a remote part of the system. The currents generated by such a load require a path to flow to the supply rails. Such a path can be provided by the diodes in the ESD protection, or by the body diodes of the output devices. If these elements are absent, then a dedicated protection circuit should be added.

On power up, the OpAmp requires a start-up time that approximately equals its step settling time plus the start-up delay. The latter delay usually dominates. The start-up delay is defined by the settling time of the biasing core and the settling time of the bias current sources in the amplifier. A shorter start-up time can save battery energy, provide quicker signal acquisition time and lessen the time period when the OpAmp must be kept on. A short start-up time is especially important in applications where OpAmps, with separate input signals, operate on a common load, and the OpAmp shutdown function is used for signal multiplexing [103].

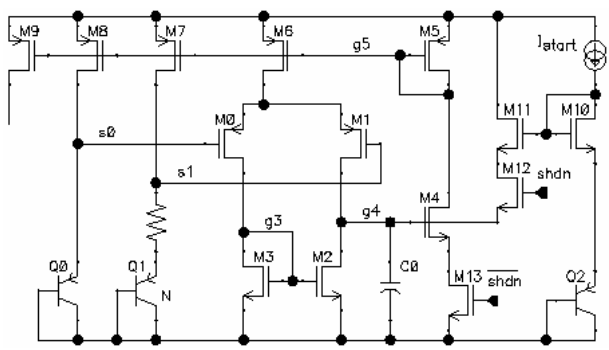


Figure 8-3. Speeding up of the bias core start-up

Most of the PTAT or g_m-matching biasing cores, described in subchapters 3.2 and 3.3, have a long start-up time because one has to charge the compensation capacitor (or the parasitic gate capacitor if there is no dedicated compensation), and the only available charging current is a very small current from the start-up source. This start-up time can be reduced if the compensation capacitor remains charged in shutdown mode. This precharged voltage should approximately match the capacitor voltage required for normal OpAmp operation. An example of this circuit is shown in fig. 8-3 [95].

Here, in the PTAT biasing core Q_0-Q_1 and M_0-M_8, the compensation capacitor C_0 is precharged during the OFF time to the voltage $(V_{gs} + V_{be} - V_{TN}) > V_{C0} > V_{be}$, by reusing the start-up current source I_{start}. The switch M_{13} disables the core, and I_{start} is the only current consumed during shutdown. To turn the OpAmp on, the switch M_{12} is turned off and M_{13} is turned on. The capacitor C_0 is disconnected from the charging circuit by the switch M_{12}. When M_{13} turns on, the capacitor is already charged to nearly required voltage, and M_4 immediately starts conducting the current close to its steady state value. All biasing current sources will be turned on immediately as well. The overall OpAmp start-up time can be reduced to as little as 100 ns [105].

8.2 Temperature shutdown

Conventional bipolar and CMOS cells such as current mirrors can only operate at the temperatures below 160-170°C. At higher temperatures the junction leakage currents become significant, and the voltages V_{be} or V_{gs} drop, are making circuit design much more complicated [96].

If an OpAmp sinks or sources a large load current (shorted load due to a system fault, incorrect application design, etc.), it can overheat and self-destruct. For example, SOIC-8 package can dissipate about 0.5 Wt with 50°C overheating. With a 5V supply, an OpAmp using this package would need temperature protection if its short-circuit current could exceed 100 mA.

Temperature protection is a must for power OpAmps and line drivers where the load and heat dissipation are not predictable.

Usually, the temperature protection circuit shuts down the OpAmp or its output stage at a temperature of about 150-160°C (with about 20°C of temperature hysteresis before turning the OpAmp back on).

Temperature protection circuits require temperature sensors. The most frequently used are shown in fig. 8-4.

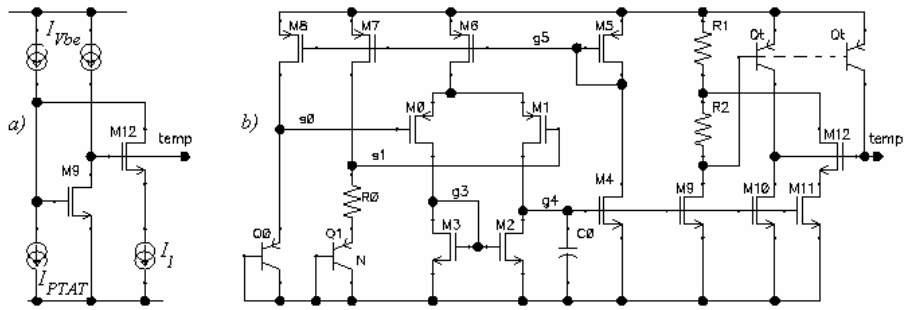


Figure 8-4. Temperature protection circuits

The circuit of fig. 8-4a is a one-transistor comparator. Here, I_{Vbe} is the proportional to V_{be} current (i.e. has negative TC) and I_{PTAT} is a PTAT current. When the temperature is low, then $I_{PTAT} < I_{Vbe}$, and transistor M_9 is on and the output signal is LOW. The protection trip temperature is set by the difference (I_{PTAT} - I_{Vbe}). When the temperature increases, and $I_{PTAT} = I_{Vbe}$, the comparator output signal starts to increase. The positive feedback transistor will be on and accelerate the comparator transition to HIGH at the output. The portion I_1 of the current I_{Vbe} is intercepted by M_{12} and the circuit returns to LOW at the output when the decrease in temperature results in $I_{PTAT} = (I_{Vbe} - I_1)$. Hence, the current I_1 sets the hysteresis width.

Another circuit (fig. 8-4b) that can be conveniently used with a PTAT core (for example, the one shown in fig. 3-1) uses the positive temperature coefficient of resistors and the negative temperature coefficient of V_{be} voltage of lateral PNP transistors. These transistors can be placed around heat-generating areas and connected in parallel. At low temperatures transistors Q_t are off, transistor M_{10} is on, and the output voltage of the comparator Q_t, M_9, M_{10} is LOW. At the trip temperature, transistors Q_t will be on, and the positive feedback transistors M_{11}, M_{12} will accelerate the comparator transition to the HIGH state. The current in M_{11} and the value of R_1 will set the hysteresis width and the reverse trip temperature.

The location of the temperature sensors is important for protection of high-current OpAmps, with large output devices and large temperature gradients across the chip. Bipolar transistors are prone to secondary breakdown caused by hot spots. Temperature sensors distributed over the power device surface were critical in order to catch these hot spots before the silicon is destroyed [97]. However, bipolar transistors are not likely to be used in power OpAmps any more, because MOS transistors occupy less die area, do not create hot spots and are more reliable. The location of temperature sensors along the periphery of power devices, between them and the rest of the circuitry are usually sufficient for OpAmp protection.

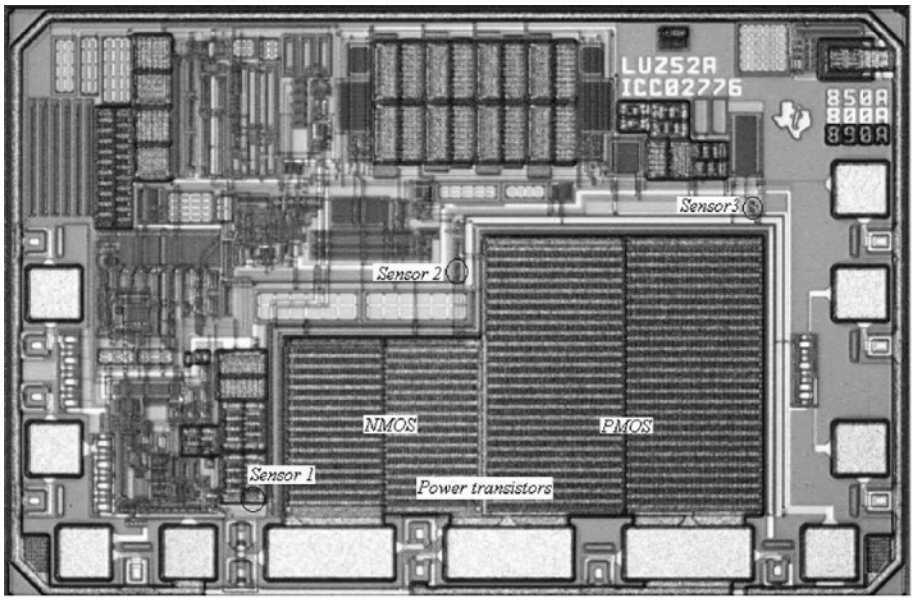


Figure 8-5. Location of the temperature sensors in power OpAmp

Fig. 8-5 shows the locations of bipolar sensor transistors for temperature protection of 3A, 16 V OpAmp with PMOS and NDMOS output devices [44].

8.3 Output current limiting

An OpAmp can be equipped with a current limiting circuits either to protect itself or the load. The latter case occurs when the load safe operating area is limited or when the load current or power is restricted by regulations. An example could be the government regulations on power line communications.

OpAmp overload can result in a local or general overheating and eventual die destruction in the absence of protection. Thermal shutdown is a measure that protects an OpAmp against general overheating. Output current limiting prevents local thermal effects such as metal or bondwire fusing, via or contact meltdown. It may also extend the OpAmp operating time after the load fault. During this time, rest of the chip warms up and the general thermal shutdown can be triggered. OpAmp protection can be achieved even without current limiting at all if a local thermal protection is implemented [97].

The metal width and via quantity for a given current requirement are usually based on long-term reliability concerns. Fusing current far exceeds

the load current that the OpAmp is designed for, so as long as the limit for load current is above the specified operating range, the accuracy of the current limiting circuit is not important.

Delay of the current limiting circuit is also not important because the time stretch of the destructing thermal effects far exceeds any electrical transient processes in the protection circuit.

One of the most popular current limiting circuits (fig. 8-6a) is based on a dedicated feedback loop (again a good example of the structural approach) [98]. In this loop the actuator transistor, M_0, is connected between the gate and source of the output transistor M_N. In normal operation this transistor is off, and does not interfere with the OpAmp. The sensor transistor M_{Ns}, connected in parallel with the output transistor, measures the load current. When this load current exceeds the limit, the voltage V_{R0} exceeds the reference voltage V_0, and the amplifier A_1 turns on the actuator transistor M_0. This transistor starts to conduct and removes the signal from the gate *gn* of output transistor M_N.

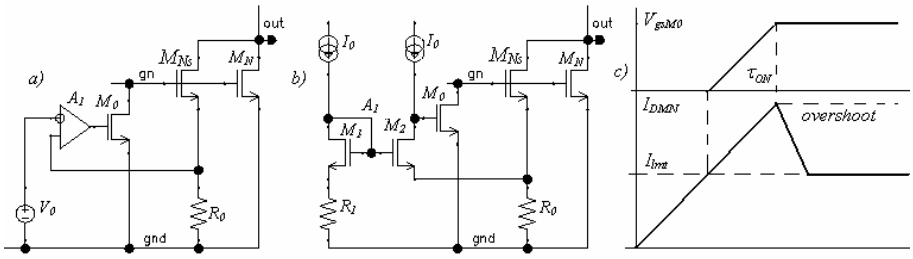


Figure 8-6. Current limiting for OpAmp protection

Fig. 8-6b shows a full implementation of the limiting circuit. Reference voltage here is $V_0=I_0R_1$ (assuming $R_1>>R_0$), and the amplifier is realized as the current amplifier M_1, M_2. During normal operation the gate potential of M_0 is close to the negative rail, and the circuit has a delay τ_{ON} defined by the time required to charge the parasitic capacitances at the gate node to the potential of V_{gs} (fig. 8-6c). This delay causes the output current to overshoot. Dependence of the voltage across R_0 on I_{DMN} is nonlinear, and the variation of the limiting current, I_{lmt}, exceeds the production or temperature spread of the resistor values and the spread of reference voltage V_0. Errors of 50% in I_{lmt} are not uncommon, even with relatively small 10% resistor spread.

Current limiting circuits for load protection should usually have tighter limit tolerances (less than 10%), and delay may not be tolerated as well. In essence, these two requirements distinguish them from the OpAmp protection circuits; they also respond to the output current, yet, this response should be faster and more accurate. In addition, the current limit in this case

should often be application-adjustable. Circuits providing these improvements are shown in fig. 8-7 [99].

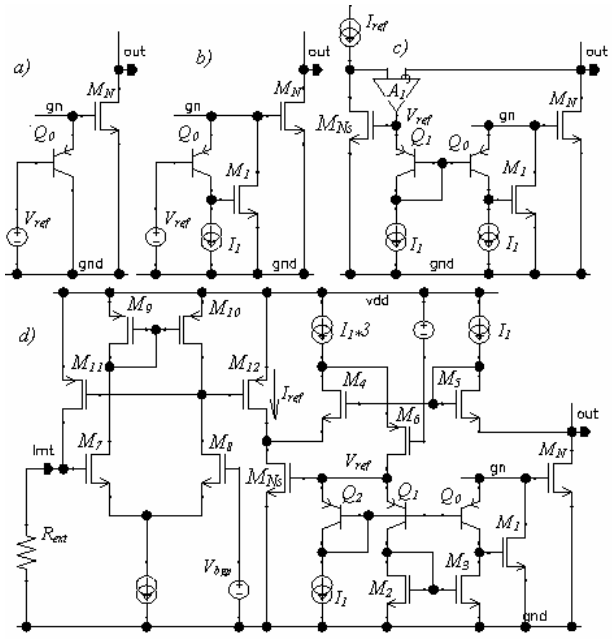


Figure 8-7. Current limit for load protection

The operation principle is illustrated by fig. 8-7a. When the gate voltage of the power transistor, M_N, increases above (V_{ref} + V_{be}), then transistor Q_0 starts to conduct and prevents V_{gsMN} from increasing. This limitation mechanism works faster than the signal transfer through an OpAmp, so there is no load current overshoot.

However, there can be up to 0.2 V difference between V_{gsMN} corresponding to the condition when Q_0 starts to conduct, and value of V_{gsMN} when this gate voltage can be considered as finally limited. This "softness" of the limiting characteristics can be lessened by introduction of the simple feedback loop including the gain boost amplifier I_1/M_1 as shown in fig. 8-7b.

For load protection, the output device current should be limited even if this device starts to operate in the triode region. Circuit providing this feature is shown in fig. 8-7c. It comprises the sensor transistor M_{Ns}, which matches the output transistor M_N with the aspect ratio N, the current source $I_{ref}=I_{lmt}/N$ and the feedback loop through A_1. This amplifier holds the drain voltage of M_{Ns} equal to the output voltage, even when M_N is in the triode mode. When M_N is in saturation, M_{Ns} is in saturation as well, and $V_{ref} = V_{TN} + \sqrt{I_{ref} / K_{NS}}$. If the gate voltage of M_N is less than V_{ref}, transistor

M_1 is OFF. Transistor M_1 will start to conduct when the drain current of M_N achieves the limit value, I_{lmt}, and the voltage at the gate of M_N becomes higher than V_{ref}. When M_N enters the triode region then the current sensor transistor M_{Ns} enters triode region as well, and the voltage V_{ref} increases because M_{Ns} must take the same current I_{ref}. This increase of V_{ref} allows a higher maximal voltage at the gate of M_N to obtain $I_{lmt}=I_{ref}N$.

The detailed implementation of this load protection circuit is shown in fig. 8-7d. Source I_{ref} uses the bandgap reference voltage V_{bgp} and the amplifier M_7-M_{12}, so that an external resistor, R_{ext}, sets $I_{ref} = V_{bgp}/R_{ext}$ (one could also use an external current source). Amplifier A_1 is realized as the current-input amplifier M_4/M_5 with cascode M_6. The output transistor gate voltage limiter comprises Q_0 with current gain boost M_1 and additional gain realized with Q_1/Q_2 and the mirror M_2/M_3.

Error of the current limiting circuit of fig. 8-7c is defined almost exclusively by the mismatch of M_N/M_{Ns} and can be as low as 5%. This circuit has practically no delay or current overshoot [44].

8.4 Slew rate enhancement

The tail current of an input differential pair usually limits its output current (fig. 8-8, curve 1). This, in turn, results in a limited OpAmp slew rate. In two-stage OpAmp, the slew rate is $dV_{out}/dt \le I_{tail}/C_M$, where C_M is the total Miller compensation capacitance.

Slew rate enhancement circuits can be divided in two groups depending on the application requirements:

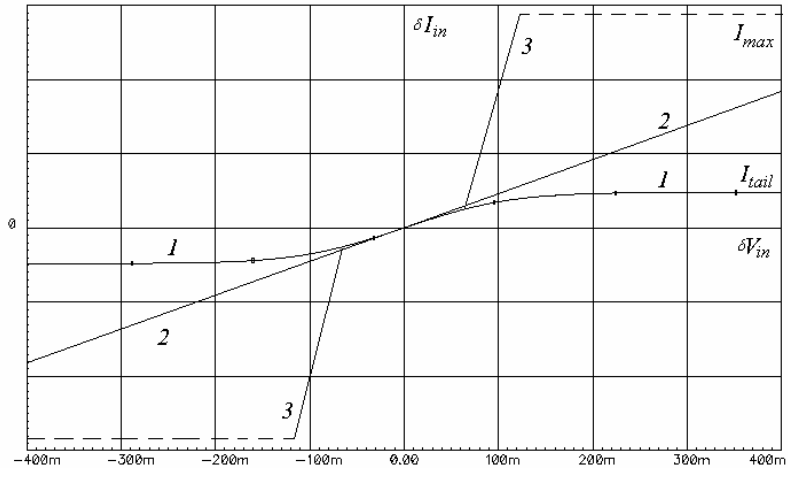


Figure 8-8. Input stage current versus differential input voltage

1) OpAmps processing "natural" signals like audio or DSL should have minimal distortions of the input signal. The OpAmp must keep linearity in the differential input voltage range as wide as possible (curve 2 in fig. 8-8).

2) In a significant number of applications the signals are "artificial", and it is more important to settle the OpAmp output voltage within a fixed time with the same absolute error (not proportional to the input step). Examples of this are video amplifiers (video signal consists of relatively small AC voltage with large synchronization pulses), and DAC and ADC buffers which have to settle within 1 LSB in less than one clock period. To achieve this fast settling, one must provide a nonlinear output from the input stage (curve 3 of fig. 8-8).

The approximation to the curve 2 can be realized using circuits similar to the ones controlling the minimum device current in the class AB output stage [45]. Structure of such circuits are also described by the signal graphs close to that ones of the class AB output stages (fig. 7-3).

Fig. 8-9a shows an example of circuit modifying the tail current of input stage when the large differential input voltage is applied [100]. Signal graph of this circuit (fig. 8-9b) is similar to that for the general class AB structure (fig. 7-3a) except of removing the node for the load current.

The input transistors, M_0 and M_1, have the matching current sensors, M_{0s} and M_{1s}, connected in parallel with them. Output current of the nonlinear cell, $M_2/M_3/M_4$, is defined by the smaller of the two sensor currents. The smallest of the sensor transistor currents through M_2 or M_3 defines the voltage V_{g5}. This voltage controls the input stage tail current through M_5.

The simulation results of fig. 8-9c show the currents for each of the input devices, M_0 and M_1, and the difference of these currents for the values of R_0=10k, I_0=10 uA, and all W/L=100/0.6 um. Both transistors, M_0 and M_1, are always on (each of them preserves a certain minimal current), and the difference of their current increases with the input differential voltage. There is no limit to this difference, and the nonlinearity of this characteristic is usually acceptable.

One can also derive that for $V_{in1}=V_{in2}$ and $M_2/M_3/M_4$ operating in strong inversion the tail current is

$$I_{tail} = 2I_0 + (4/R_0)\sqrt{I_0/[2K(W/L)]}(\sqrt{2}-1) \quad \text{where } K = \mu_N C_{ox}.$$

The graph of fig. 8-9b can be equivalently transformed by the splitting the common-mode feedback link (fig. 8-10a). Fig. 8-10b shows an implementation of this graph [101].

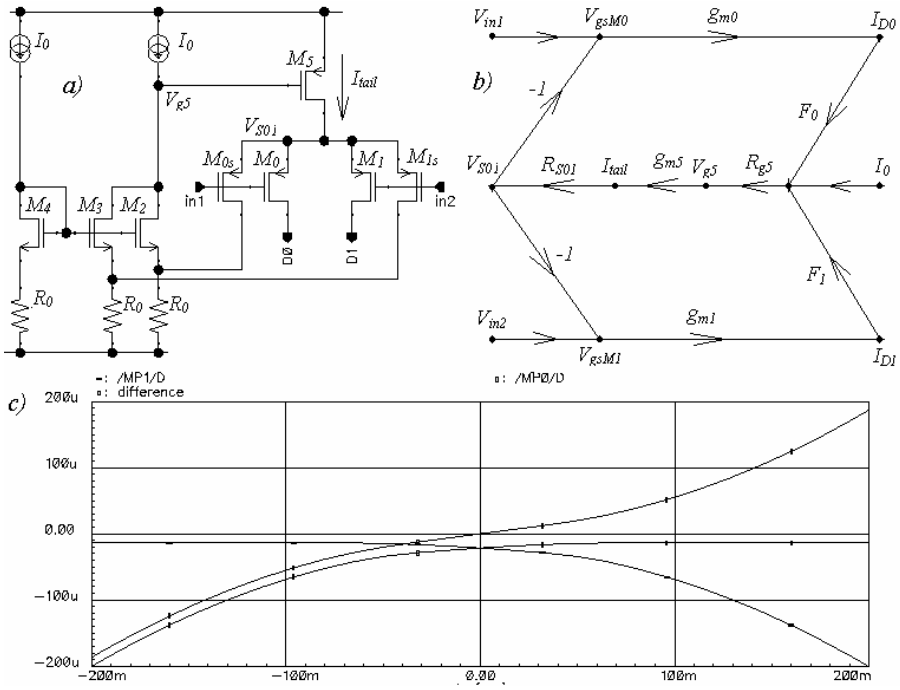


Figure 8-9. Tail current enhancement circuit derived from general differential structure

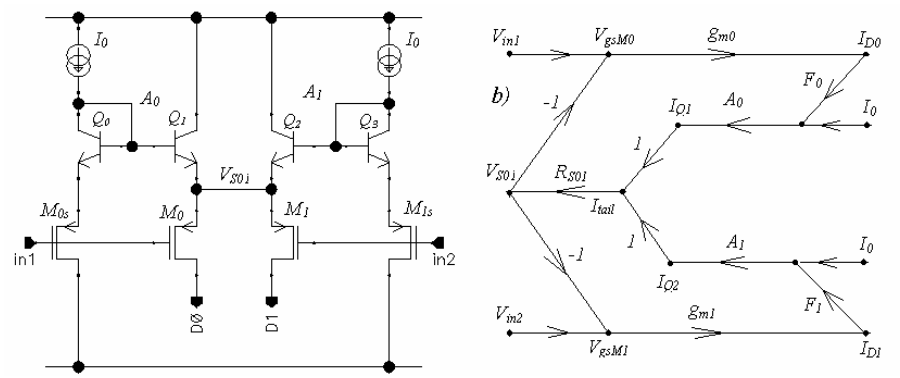


Figure 8-10. Tail current enhancement circuit with split feedback structure

Each of the input devices, M_0 and M_1, has a feedback loop (M_{0s}-Q_0-Q_2 and M_{1s}-Q_1-Q_3). These loops increase the total tail current when the current of any of input transistors decreases due to an increase of the differential input signal. Operation of this circuit is even closer to the operation of the circuits controlling the minimum device current in the class AB output stage

[45]. The output current versus differential input voltage dependence is similar to the one shown in fig. 8-9c.

The currents of input devices usually pass through the folded cascode on the way to the gates of output transistors. It is there that these currents charge the Miller (and gate) capacitances. A folded cascode stage can pass almost unlimited current through the inverting input. Yet, the current through the non-inverting input is limited by I_F. To pass the extra current due to the tail current increase, one must clamp the inverting input. An example of this circuit is shown in fig. 8-11.

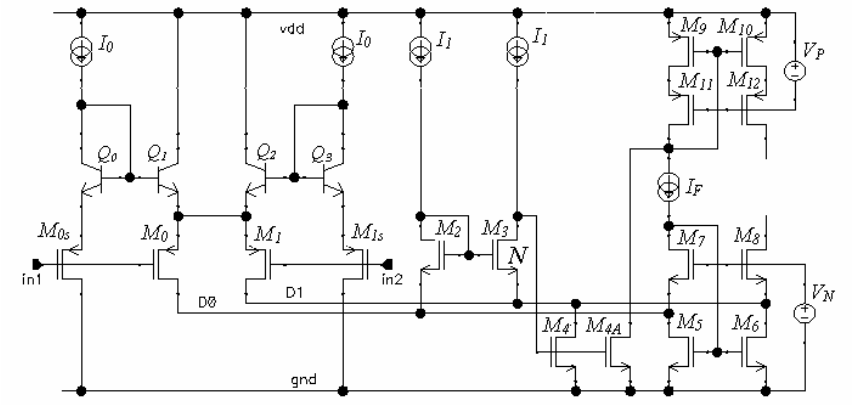


Figure 8-11. Slew rate enhanced stage with folded cascode and clamping circuit

The clamp comprises the current-input amplifier with a mismatched input, M_2, M_3 and M_4. This mismatch results in the triode operation of M_3, and the transistors M_4, M_{4A} being off. When the voltage at the folded cascode input, *D1*, starts to raise, the transistor M_4 starts to conduct and passes all extra current from M_1 to the supply rail. Current of M_4 is mirrored by M_{4A} and flows around I_F through the inverting input of the opposite side M_9-M_{12} of the folded cascode.

The transconductance of the input stage with an increased tail current does not change drastically with the increase of the input differential voltage. Large-signal dynamics of the OpAmp with such an input stage is close to the small-signal dynamics. No additional compensation problems occur usually. Only if the input stage has a very nonlinear *I/V* dependence (fig. 8-8, curve 3) then the large-signal transconductance becomes much larger the small-signal one. This sharp increase of transconductance can cause conditional instability effects in the OpAmp step response as shown in fig. 8-12.

If the non-linear circuit (which provides the sharp increase of input stage current up to I_{max}) does not have any delays, the step response looks like in fig. 8-12a. It includes the initial fast slewing with $dV_{out}/dt = I_{max}/C_M$, then,

when the output voltage is close to the final value, the slew boost circuit is disabled and the OpAmp settles at the target output voltage (the dashed line shows the large signal step response without slew rate enhancement).

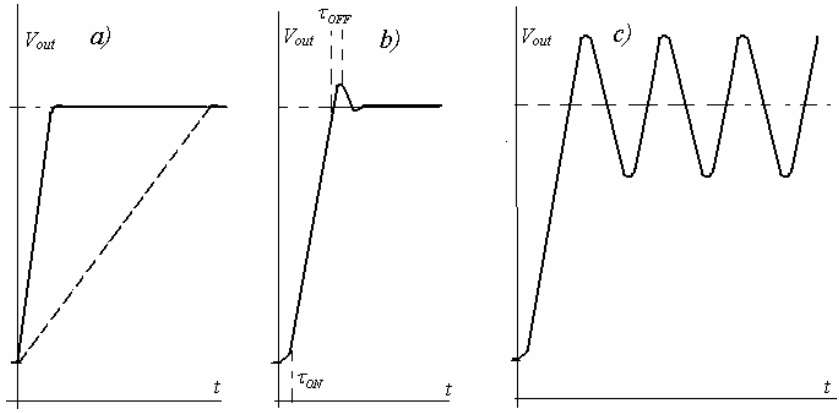


Figure 8-12. Conditional instability with nonlinear input stage

In the practical case, when the non-linear unit has delays, τ_{ON} and τ_{OFF}, the step response looks as shown in fig. 8-12b. One has the delay, τ_{ON}, before the start of the fast slewing, then fast slewing, then an overshoot due to the τ_{OFF}, and only then the final settling.

If τ_{OFF} is too large, the overshoot can trigger the operation of the non-linear unit in the opposite direction. This may cause conditional instability, where the OpAmp output voltage starts to oscillate (fig. 8-12c). The τ_{OFF} delay is critical for normal operation of the slew rate enhancement unit.

The implementation of the input stage with a non-linear slew rate boost as shown in fig. 8-13 does not have this deficiency [102]. Circuit comprises the main input differential pair, M_0/M_1, and two non-linear units for enhancement of positive and negative slewing. The input differential pairs (M_2/M_3 and M_{2A}/M_{3A}) of each circuit are mismatched by K aspect ratio, and this mismatch determines their turn-on threshold voltage. When the input differential voltage exceeds the threshold, one of the current-input amplifiers (M_4/M_5 or M_{4A}/M_{5A}) turns on, and one of the transistors, M_6 or M_{6A}, conducts. Current of M_6 or M_{6A} is mirrored with M_7/M_8 or M_{7A}/M_{8A} and flows through the gates, *gn* and *gp*, of the output devices, M_N and M_P (not shown here). This provides the slewing current increase similar to that of curve 3, fig. 8-8, with $I_{max} = 2I_1N$.

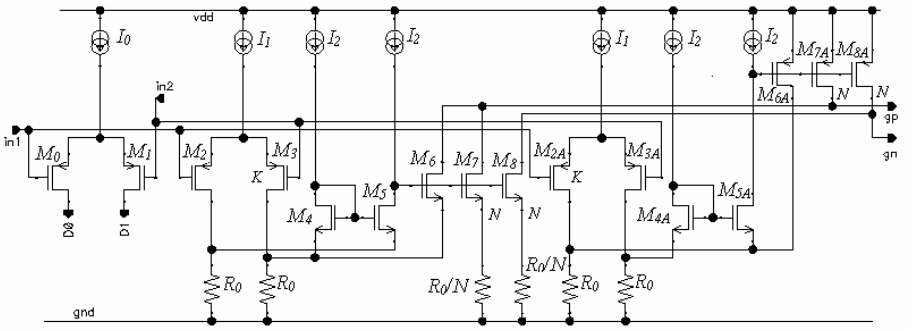


Figure 8-13. Input stage with non-linear slew rate boost

The τ_{ON} delay is defined by the time necessary to charge the parasitic capacitance at the gates of M_6 or M_{6A} by the bias currents I_2 from zero to the required V_{gs} voltage. Due to the feedback operation, during ON time the delay τ_{OFF} is much smaller: the gate voltages of M_6 or M_{6A} should change by 100-200 mV only to turn these transistors off.

8.5 Overload recovery

When the OpAmp output is close to the supply rail, the one of the output transistors may enter the triode region. Assuming that this is the case for n-channel transistor, M_N (fig. 8-14a), the currents, I_{D1} and I_{D0}, from the input stage still can flow through the folded cascode and charge the gate of M_N to an unpredictable and high potential. When the output voltage returns to normal then this gate potential will eventually recover, but with the overload recovery delay, τ_{ovl} (fig. 8-14 b).

Limitation of the gate voltage of M_N (or M_P, for the high side) decreases this delay. The gate clamp (fig. 8-15a) serves the purpose. However, this simple approach does not eliminate τ_{ovl} completely. Also, in the power OpAmp, the V_{gs} of M_N (and M_P) can achieve quite large values even during normal operation when running large output current. Such a clamp would interfere with this operation. A better solution using the structural approach, i.e. a dedicated feedback loop controlling the overload recovery, is shown in fig. 8-15b, with a full implementation given in fig. 8-15c [104].

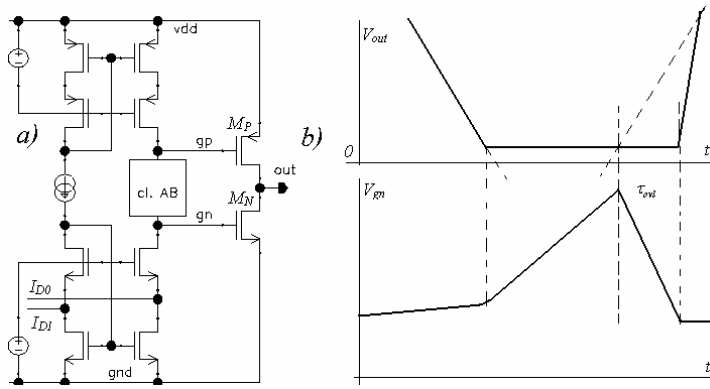


Figure 8-14. Overload recovery delay

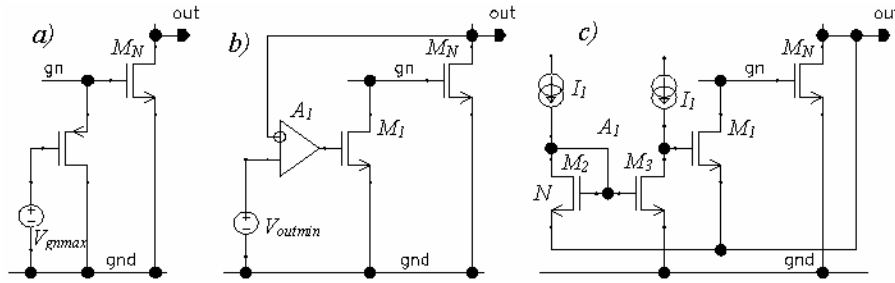


Figure 8-15. Gate voltage clamping to prevent overload recovery delay

In the circuit of fig. 8-15b, the amplifier A_1 turns the transistor M_1 on when the output voltage drops below V_{outmin}. This amplifier is implemented using a current-input amplifier (M_2 and M_3). Voltage V_{outmin} is set by the size mismatch ($V_{out\,\min} \approx \sqrt{2I_1 / [K(W/L)_3]}$, where $K = \mu_N C_{ox}$). Source of M_1 can be connected to the output to improve the speed (in this case, when the output voltage drops below V_{outmin}, the source voltage of M_1 decreases and the gate voltage increases simultaneously). During normal operation, when the output voltage is above V_{outmin}, M_1 is off and does not interfere with the OpAmp operation. When the output voltage drops below V_{outmin}, M_1 starts to conduct and limits the gate voltage of the output transistor.

Chapter 9

FROM STRUCTURE TO CIRCUIT

Currents, voltages, sizes and honing of the common sense

Analog design starts with an outline of the circuit structure and the generation of a few, possibly acceptable, circuit topologies, using piece of paper. The next steps of the design process are the drawing of exact circuit diagrams using CAD schematic capture software, choosing the voltages, currents and component dimensions, and running of the simulations which should verify what the designer already knows about the circuit.

Some unpredictable problems do arise in each and every design, and this move from structure to the final circuit diagram must be divided into several iterative steps in order to quickly locate problem sources as well as to find the performance boundaries. The following sequence of the design steps may be suggested:

1) Creation of ideal macromodels for the circuit units and simulation of the OpAmp using a mix of these macromodels and real components. For example, all biasing circuits initially are replaced with ideal current or voltage sources, and amplifiers in the local feedback loops are replaced with their macromodels. The real components are the input transistor pair and the output devices only. The OpAmp should include a bare minimum of units or blocks that is necessary for functionality, and should be stripped of all unnecessary units such as gain boost, folded cascode, etc.

Simulations at this step should clarify the maximum achievable bandwidth, choice of compensation scheme, amplifier noise properties, capacitive drive capability, and the distribution of the OpAmp current budget between the units. Requirements to the amplifier units also should be determined at this step. Questions like the following should be answered: what should the output impedance and voltage range for each of the current sources be? What is their acceptable value spread resulting from the process parameter variations and temperature coefficients? What are the limits for the offset, noise, speed and gain of internal amplifiers?

2) Addition of the functional enhancement units represented by their macromodels (with ideal current and voltage sources for their bias, ideal amplifiers, etc). Simulations should prove that the amplifier works as expected with these ideal units and define their main requirements.

3) Generation of the circuit implementations for the functional units and simulations to verify how close these units are to their macromodels. Special attention needs to be paid to the operation at the edge and outside of the normal boundary conditions (i.e. saturation of the current sources; heavy loading of the voltage sources (for example, by capacitive currents during signal transients); start-up operation; tests with supply voltage up/down steps.

4) Sequential replacement of the macromodels of the units with the real circuit implementations and ensuing simulations to determine whether any problems occur. Any real cell, clearly, does not operate better than the ideal one. The damage done by the cell parasitic effects to the overall OpAmp parameters should be estimated. Remedies should be applied if this damage is not acceptable.

5) Simulations of the full OpAmp circuit and verification of the compliance with all requirements at all operating temperatures and process corners.

It is interesting to notice that the next step after conception of the circuit topology, namely choosing the transistor dimensions and biasing current values, can almost always be done using common sense only. An amplifier with a sharp dependence of the performance parameters of these values is not robust and hence is not good, and the search of a dull extremum may be not worth the design time. "Good enough" is a very frequent comment on the choice of sizes and currents. If unit properties are not satisfactory with "good enough" components, then the designer should look for a new structural solution, i.e. for a different circuit topology.

As an example, in this chapter the design, step by step, of the industry-quality OpAmp using the generic CMOS process with single N-well and 0.5 um minimal channel length is considered. The main requirements to this OpAmp are the following:

- total I_q is 50 uA (typical),
- supply voltage range from 2.2 to 5.5 V,
- offset less than 5 mV (3σ),
- rail to rail output within 100 mV from supply voltages when running 5 mA load current,
- slew rate 2 V/us
- rail-to-rail input.

9.1 General considerations of transistor sizing and biasing

Generally, in order to reduce parasitic capacitances, the size of any transistor should be as small as possible. The minimum transistor area (*WL*) is defined by the matching requirements, and the *W/L* aspect ratio is defined by maximum current and operating mode. Biasing currents should also be as small as possible while meeting the noise and speed requirements.

CMOS transistor sizing.

Development of common sense in the evaluation of transistor sizes requires, without any doubts, some background data. For any given process, a few simulations, numbers and reference curves help to create the basis for this common sense. For CMOS transistors these numbers and curves include:

1) The values for NMOS and PMOS threshold voltages, V_{TH}, and their boundaries for the given process. In our example the nominal V_{TH} is 0.75V for NMOS and -0.89V for PMOS transistors.

2) Value of the voltage V_{mtch} (subchapter 1.5.2) for NMOS and PMOS transistors. This empiric value usually can be found in the process documentation. It helps to choose the overall area (*WL*) of transistors that have to match each other for any desired voltage offset or to estimate a possible mismatch of two transistors that are sized from other considerations. Usually $1\sigma(V_{off}) = V_{mtch} / \sqrt{WL}$, where W and L are in microns (the numerical ratios given below will always be in um/um). In our example V_{mtch} is equal to 25 mV for both transistor types.

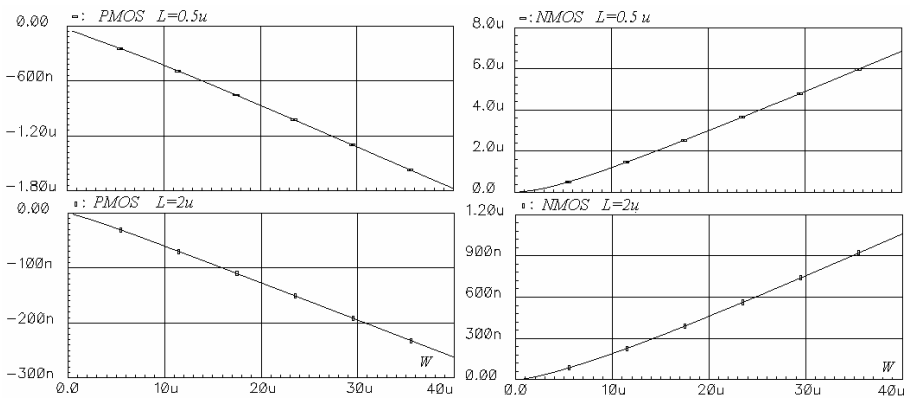


Figure 9-1. I_D versus transistor width for $V_{gs}=V_{TH}$

3) Dependencies of I_D versus W for L=L_{min} and, say, L=2u, with V_{gs} = V_{TH} (fig. 9-1). These dependencies show the border between weak and strong inversion modes of operation. The drawn width and length are different from these values in real transistors, yet, for longer transistors (with L=2 um) this difference is negligible and the results of fig. 9-1 can be scaled (current for the same V_{gs} will be inversely proportional to L). For the minimum-length transistors, the difference between the drawn and real dimensions is significant, and the curves for L=L_{min} are necessary because such transistors are chosen most often in order to decrease parasitic capacitances.

The curvature at the small transistor width reflects the narrow channel effect. As a result, narrow transistors occupy larger area to run the same current, and they introduce more parasitics. If a wide transistor is required, it is preferable to use transistors with lesser number of gates and larger W of each module.

4) The dependencies of I_D versus V_{gs} for transistors with L=L_{min} and L= 2 um, with W large enough so that narrow channel effects are not noticeable, say W=20 um (fig. 9-2), and small V_{ds} = 0.2 V.

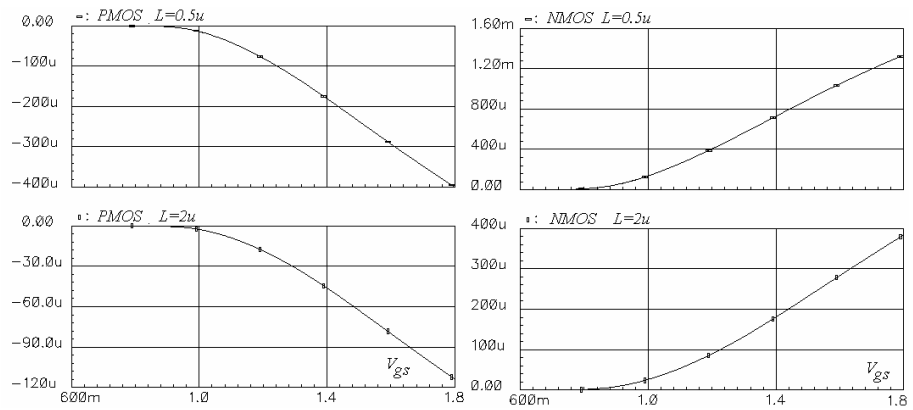


Figure 9-2. I_D versus V_{gs} for W=20um, L= 0.5 um and L=2 um

These dependencies help the designer to size the output devices and to choose the transistor dimensions for any given current if it is important to avoid the triode mode.

5) The simulations of the g_m versus W for a given current, for example, I_D=1 uA, as in fig. 9-3, illustrate the limits of g_m control with MOS transistor dimensions.

Curve of the g_m versus W becomes flat as soon as the increase of W puts the transistor in subthreshold mode. Any further increase of W can only increase parasitic capacitances. In weak inversion, g_m of NMOS and PMOS transistors for the same I_D have close values.

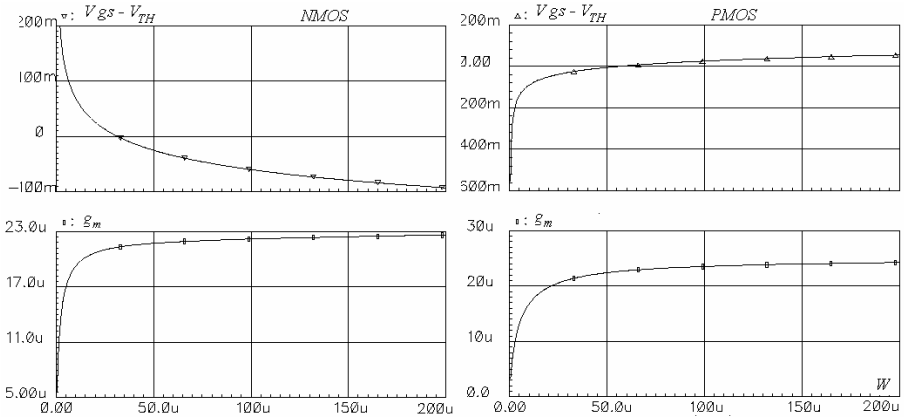


Figure 9-3. g_m versus W for L=2 um and I_D=1uA

Bipolar transistor sizing.

The times when each bipolar device was tailored to its place in the circuit are long gone. In most design kits it is possible to choose only the multiple factor of a few standard modules. Parameters important for sizing include:

1) Value of parameter *IKF* for the smallest geometry device or the dependence of its β versus collector current, I_C. The *IKF* value reflects the collector current when β starts to decrease sharply. It can be used to determine the minimum size of the device for a given current value. Also, the speed of the bipolar transistor increases with current and is highest when the current is around *IKF*.

2) Collector and emitter body resistances. After sizing the transistor using *IKF,* the voltage drop on the body resistances should be checked. Due to these voltage drops and with required V_{CE}, transistor can operate in highly undesirable saturation mode.

3) The base body resistance of the smallest geometry device. This figure is used to size noise-critical transistors. The base body resistance of the noise critical devices should be smaller than the emitter resistance defined by the collector current: $r_{body} < 1/g_m = V_T / I_C$.

Biasing currents.

The total value of the biasing current can be picked from the available current budget. 70-80% of this budget should be used to improve the key current-dependent parameters of the OpAmp: noise, bandwidth, and capacitive drive capability.

In a correctly designed OpAmp, the input-referred noise is dominated by the noise of the input differential pair. Its equivalent noise resistance for high frequency noise is inversely proportional to g_m [10, p. 85]. The transconductance, g_m, for any given current in a MOS transistor does not increase any more once enlargement of the *W/L* puts the transistor in weak inversion mode (fig. 9-3). Hence, the input stage tail current defines the high-frequency noise, and the value of this current should be set on the basis of noise requirements (therefore, the realistic noise requirements should reflect the available current budget).

The output devices are sized on the basis of maximum load current and saturation voltage requirements. These devices are large and have large C_{gd} capacitance. Frequency compensation of the two-stage OpAmp can be done using the C_{gd} only [44]. Unfortunately, C_{gd} is nonlinear versus voltage, so Miller compensation capacitors, at least of equal or larger than C_{gd}, are usually added to increase the bandwidth stability. In a two-stage OpAmp, the bandwidth is equal to $g_m/(C_M+C_{gd})$ where g_m is the transconductance of input stage. This transconductance depends also on the input stage tail current. The bandwidth requirement, hence, sets another limitation on the minimum value of the tail current.

The OpAmp capacitive load introduces a time constant which is equal to C_L/g_{m2}, limiting the OpAmp bandwidth. Here, a g_{m2} is the sum of the transconductances of output devices. In the absence of a load current, the output devices, designed to carry large currents, operate in weak inversion. Their g_{m2} is defined exclusively by the I_q of the output stage. The requirement on the OpAmp tolerance to capacitive load sets the limit for minimum current in the output stage for any given OpAmp bandwidth.

Current consumption in the rest of OpAmp can be reduced by design, and should not dominate the total current budget. Biasing currents in other parts of the OpAmp can be as small as it is comfortable for designer, say, 1% or even 0.1% of the total current budget. The bottom value of the biasing currents is limited by junction leakages. These leakage currents can be as large as 10 nA for very big (about 10,000 um^2) transistors in the ESD protection circuits. But for the relatively small devices in the current mirrors, the leakage currents would hardly exceed 10 pA, at least at the temperatures below 160°C.

9.2 Design step one: input and output devices and currents

Implementation of the chosen structure (two-stage OpAmp) starts from the sizing of the input differential pair and output devices and simulations of the simplest possible OpAmp (fig. 9-4) to find the minimal achievable noise

and the maximum bandwidth within a given current budget (or, inversely, one can find the minimal current to realize these two required parameters).

In our example, 1σ of the offset voltage, V_{off}, should be 5/3 = 1.6 mV = $V_{mtch}/\sqrt{WL}$. The input devices area should be $(WL) > (V_{mtch}/1.6)^2 > 250$ um^2.

Tail current can initially be set at 10% of the total budget, i.e. to 10 uA or 5 uA in each of the input transistors. Using the curves from fig. 9-1 for L=2 um, and assuming operation in weak inversion, one can scale and find that, with 5 uA of the I_D, the required W/L ratio of PMOS input transistors should exceed 300 and of NMOS exceed 100. One may choose PMOS device with W/L = 300/1 (15 devices with W/L = 20/1) and NMOS device with W/L = 200/2 um (10 devices with W/L = 20/2), to satisfy both g_m and overall area requirements.

Now let's look at the output devices. The maximum V_{gs} for the output devices with a 2.2 V supply voltage can be around 1.5-1.6 V while running 5 mA of I_D. For the smallest overall size, these devices should be of minimum length, L = 0.5 um. To achieve 5 mA current with 1.5 V of V_{gs}, their W, in accordance with the curves of fig. 9-2, should exceed 600 um (30 of 20/0.5 devices) for PMOS and 200 um (10 of 20/0.5 devices) for NMOS (the calculations in this and previous paragraphs are done with some margins). These values may be doubled to ensure the sufficient output current capability versus temperature and process variations.

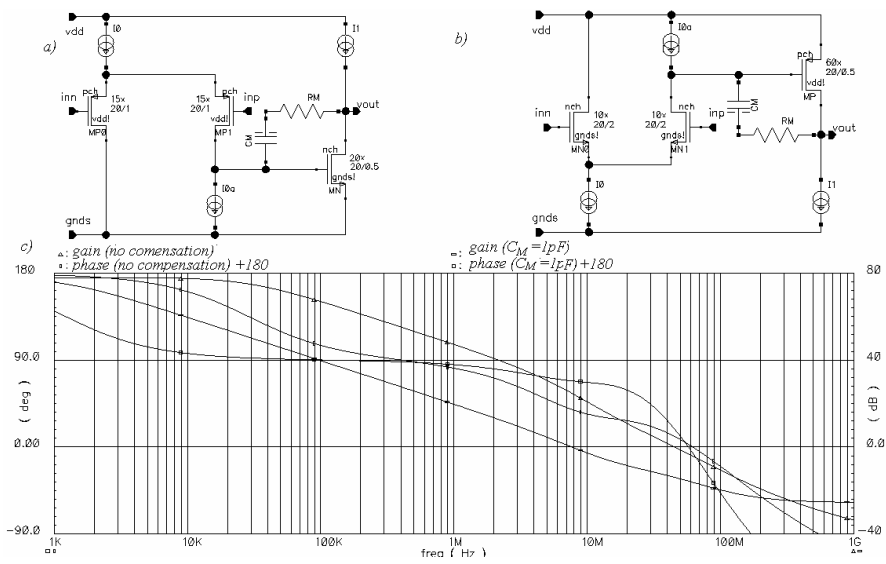


Figure 9-4. Circuits for estimation of maximum achievable bandwidth

Simplified circuits shown in fig. 9-4a and 9-4b include only the input pair and only one output device (M_N or M_P). Final circuit will include a class AB output and other circuits, but this will be added later. Quiescent current, I_q, of the output stage can be set initially to 50% of the total budget, i.e. to 25 uA.

Simulations at this step include the open loop gain and phase frequency responses for these simplified circuits, and it is rational to obtain results for the circuits with and without compensation (fig. 9-4c). Point where the phase response of uncompensated amplifier is crossing 0^o (approximately 30-40 MHz in fig. 9-4c)) is defined by the parasitic capacitors, delays in transistors and by the feedforward through C_{gd} of the output device. This figure indicates 3 to 5 times the maximum OpAmp bandwidth achievable for the chosen output device size and bias currents.

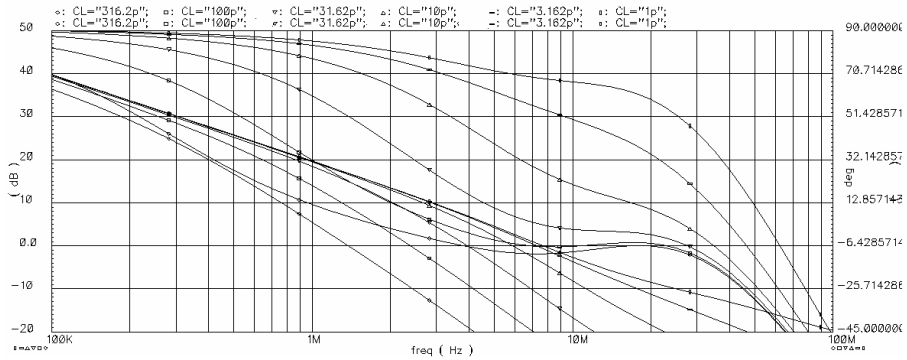


Figure 9-5. Stability deterioration with capacitive load

With a 1pF Miller compensation capacitor, the amplifier has a 6 MHz bandwidth with 65^o phase margin (without any capacitive load). Results of AC simulations for the load capacitance in the range of 1pF to 300 pF (with the step of two values per decade) are shown in fig. 9-5. OpAmp bandwidth starts to diminish noticeably even for 30 pF load capacitance. The amplifier remains stable for any load capacitance in the chosen range. However, for the load of 300 pF, the phase margin is only 6^o.

Simulations of small-step (say, 10 mV) transient responses (fig. 9-6) should verify these results.

OpAmp tolerance to the load capacitance improves with a larger current in the output device. An improvement (limited in practice) can also be achieved using the cascoded Miller compensation [50]. Another way to improve the speed and capacitive load tolerance is to change the OpAmp gain structure from two-stage to single-stage. Regretfully, there is no known

way to combine the rail-to-rail output capability and voltage (high input impedance) input in the single-stage OpAmp.

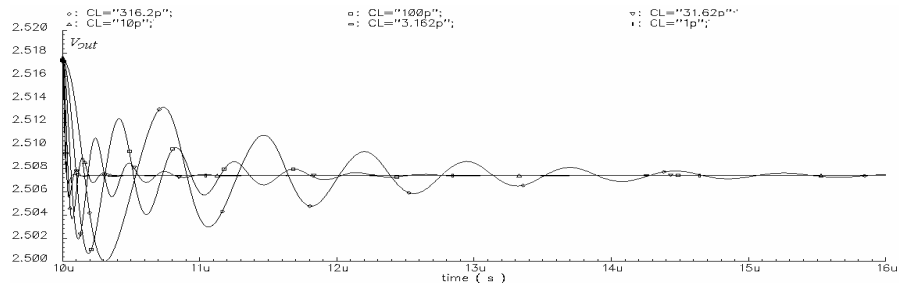


Figure 9-6. Small-signal step transient responses for the range of capacitive loads

At this design step, the designer must realize that if a better tolerance to a load capacitance is required, and there is no available current budget, then the only means left is to slow down the amplifier using a larger C_M or a smaller transconductance of the input stage.

If the design involves a high-speed OpAmp (with more than 200 MHz bandwidth) then the ESD cells and wirebond macromodels should be added at this first step, and their impact on high-frequency gain and phase response should be evaluated in the same manner.

OpAmp noise is another parameter whose achievable boundaries are determined at this design step. Simulated input-referred noise for the amplifier of our example is shown in fig. 9-6.

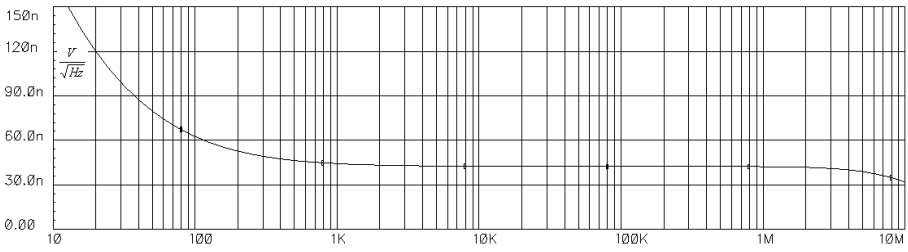


Figure 9-7. Spectrum density of the input-referred noise

This noise characteristic is flat at high frequencies and becomes proportional to 1/f at low frequencies. High-frequency noise is defined by the g_m of the input devices [10, p. 85]. In our example, with 10 uA tail current, it is 40 nV / $\sqrt{Hz}$. This value is only about two times larger than the corresponding figure for a bipolar transistor pair carrying the same tail current. An advantage of a MOS transistor is the absence of input current, so

the tail current can be increased as necessary to reduce the high-frequency noise.

The flicker (1/f) noise corner in this example is around 70-100 Hz. Contrary to the common opinion [10], CMOS transistors in modern processes can be competitive with bipolar ones in terms of the noise performance. Flicker noise is inversely proportional to $\sqrt{WL}$ of the input devices and can also be reduced. The trade-off for smaller flicker noise is increased parasitic capacitances at the drains of input devices. An excessive increase of their area may limit the OpAmp speed.

The OpAmp slew rate, in absence of a slew enhancing circuit, is $dV_{out}/dt \approx g_{m1}/C_M$ and can also be determined at this design step. If this value is not satisfactory, then the OpAmp topology should be changed and design of slew enhancement circuits (chapter 8) should be planned.

9.3 Folded cascode

The next step in the OpAmp circuit design is the addition of the circuit connecting input and output stages. This connection is most often done using a folded cascode stage (chapter 6) as shown in fig. 9-8.

If slew rate symmetry is not important, then the bias current of the folded cascode, I_F, can be smaller than the input stage tail current. If the slew rate is important, then slew boost cells must be added, and the folded cascode current can still be small. In our example, slew rate of only 2 V/us is required, so the current I_F has been arbitrarily chosen to equal half of the tail current, i.e. to 5 uA.

Current mirrors of the folded cascode (M_2/M_3 and M_4/M_5 of fig. 9-8) can significantly increase the noise and offset of the OpAmp (chapter 6.2). Transconductance of the mirror transistors should be as low as possible. This can be achieved by operating them in a strong inversion and/or by adding of the source degeneration resistors. For better matching and lower flicker noise, the overall area (WL) of these transistor pairs should be similar to the area of input devices, i.e. should be at least 300 um^2 for in example.

To be efficient, the voltage drops across the source degeneration resistors should be at least 3-4 V_T, i.e. about 100 mV, better yet, 200 mV. This leaves us approximately 250 mV for the V_{ds} of the mirror transistors. These devices, as it has been mentioned, should not operate in the triode region, as this will worsen the noise performance and significantly reduce the open-loop gain. Hence, the effective V_{gs} voltage becomes limited by $(V_{gs} - V_{TH}) < 250$ mV.

The quiescent current through each of the mirror transistors and degeneration resistors is equal to $(I_F + I_{tail}/2) = 10$ uA.

Hence, the resistor values should be equal to 200mV/10uA = 20 kOhm. The aspect ratio W/L of PMOS mirror transistors, based on simulations of

fig. 9-2 for 10uA of I_D and (V_{gs}-V_{TH}) = 250mV, should be approximately 10, and for NMOS transistors $W/L \approx 3$. The size 60/6 (3 devices of 20/6) for PMOS and 30/10 for NMOS should satisfy both the overall area requirement and strong inversion condition.

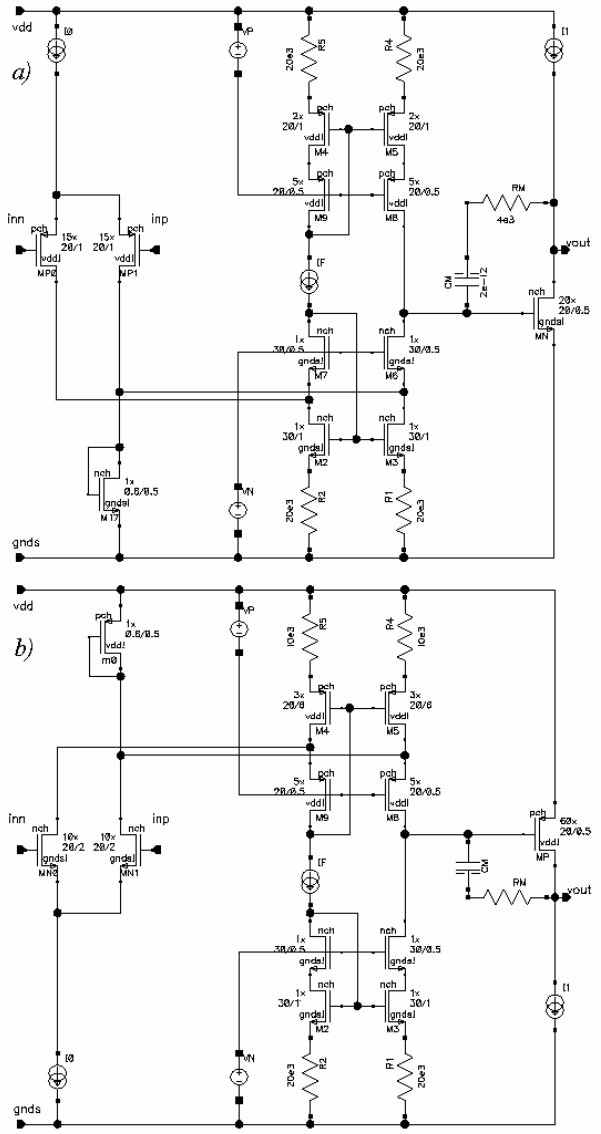


Figure 9-8. Addition of the folded cascode

Cascoding transistors (M_6 and M_8 of fig 9-8) should have a high transconductance and introduce small parasitic capacitances. They should be

of minimum length and may operate in the weak inversion region. Current through these transistors is equal to I_F (5uA) and, again using fig. 9-1 (but this time the top graphs), we set the size of M_8 equal to 100/0.5 um (5 devices of 20/0.5) and that of M_6 to 30/0.5. In the absence of a gain boost circuit, M_7 should match M_6 and M_9 should match M_8.

V_{gs} of the NMOS output device can be as low as 550 mV over temperature and process variations. The V_{ds} of M_6 should be at least 100 mV to avoid triode operation. Source potential of M_6 should be equal to 450 mV. It defines the voltage source V_N, $V_N = (V_{gsM7} + 450\text{ mV})$.

On the high side of the folded cascode $|V_{gsMP}| > 0.7$ V. The voltage drop across M_8 should be at least 200 mV, so the voltage V_P is larger: $V_P = (V_{gsM9} + 500\text{ mV})$.

The minimum-size transistors (M_{17} in fig 9-8a and M_0 in fig. 9-8b) clamp the folded cascode input voltages during slewing.

The circuit should be simulated again to find AC and noise characteristics. One must check if the deterioration of OpAmp speed and noise parameters produced by the addition of the folded cascode is acceptable. Corrections should be done if necessary. Two corrections were done in this example.

1. The delay in the current mirrors created a dip in the gain and phase frequency responses around unity-gain frequency, decreasing the phase margin to 40°.

To increase the phase margin, one can add a 5pF capacitor between the negative supply rail and the gates of M_2 and M_3. The gain/phase frequency responses without and with this capacitor are shown in fig. 9-9. However, the slower mirror with capacitor introduces a pole/zero doublet at low frequencies that results in a longer OpAmp settling.

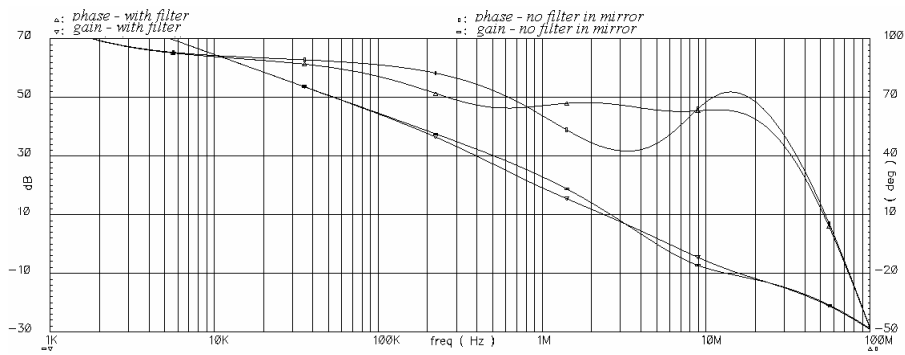


Figure 9-9. Phase margin increase due to the delay in the current mirror

Another way to reduce the mirror delay is to decrease its transistor sizes while increasing the voltage drop on the source degeneration resistors in order to keep the noise low. This second possibility was chosen.

2. The high-frequency noise changed by only a small amount (less than 10%) but the flicker noise tripled. To limit this change to a mere 100%, the area of the NMOS devices, M_2 and M_3, should be increased to 1200 um^2 each. Instead of this one can increase the source degeneration resistors, which was chosen.

The AC and noise simulation results with R_1-R_5 = 20 kOhm, $(W/L)_2=(W/L)_3$ = 30/1, $(W/L)_3=(W/L)_4$ =(40/1)=2*(20/1) (these values are shown in fig. 9.8a, the initially chosen are in fig. 9-8b) are shown in fig. 9-10. The pole/zero doublet is absent, the bandwidth increases, and, to keep it at the previous value, the compensation capacitor C_M was increased to 2 pF.

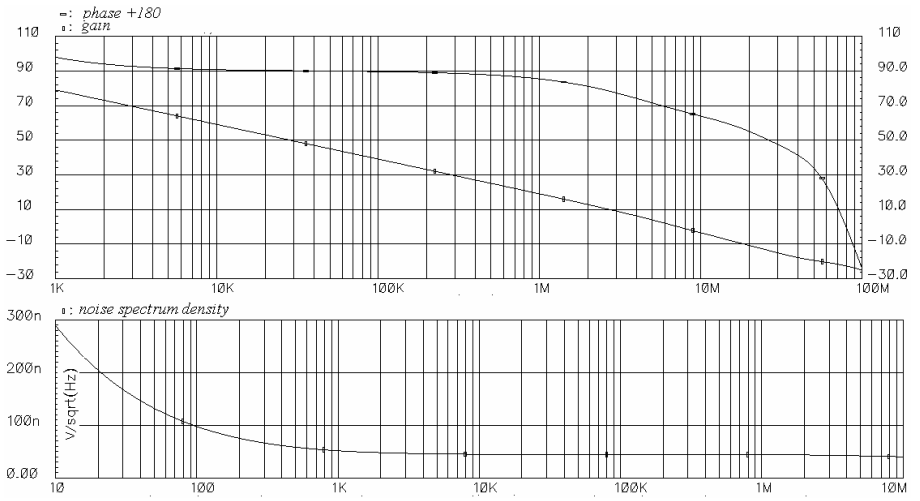


Figure 9-10. Gain/phase and noise versus frequency for final size of the folded cascode components

Dependence of the transistor operating points in the folded cascode over temperature and process corners also should be checked at this point. Operation of anyone of the $M_2/M_3/M_4/M_6/M_8$ in the triode region strongly decreases the DC gain (20 dB or more).

9.4 Class AB output stage

The next step in this design will be to create a class AB output stage. We will do this as shown in fig. 9-11 by replacing the transistor/current source combinations of fig. 9-8 (M_N/I_1 and M_P/I_1) by the simplest variant of class

AB stage shown in fig. 7-6. The minimum supply voltage for resulting circuit is equal to $V_{DDmin} = (V_{gsMN} + V_{gsMP} + V_{sat})$.

In the output stage, one of the power devices, at any instant, is operating in subthreshold. It means that either $V_{gsMN} < V_{THN}$ or $|V_{gsMP}| < |V_{THP}|$. Hence, even with a 2.2 V of the supply, there are 1.5-1.6 V still available for the gate drive of the opposite side transistor.

Transconductance of the transistors M_{10} and M_{11} defines the gain of quiescent current control feedback loops (see the signal-flow graph in fig. 7-6c). There is no need to struggle for precise control and have high gain in these loops. In fact, a lesser gain improves the OpAmp linearity. Transconductance of M_{10} and M_{11} can be low, they may operate in strong inversion and have a small area, say, 10/1 um (5 of 2/1 devices) for M_{10} and 30/2 (5 of 6/2 devices) for M_{11}.

Current I_F, supplied by the folded cascode, can be split approximately equally between M_{10}/M_{11} (2.5 uA in each). The body of the PMOS transistor, M_{11}, should be connected to its source to prevent OpAmp gain erosion caused by the drain-body leakage at a large supply voltage. In a single-N well process, it is impossible to do the similar body connection in M_{10}.

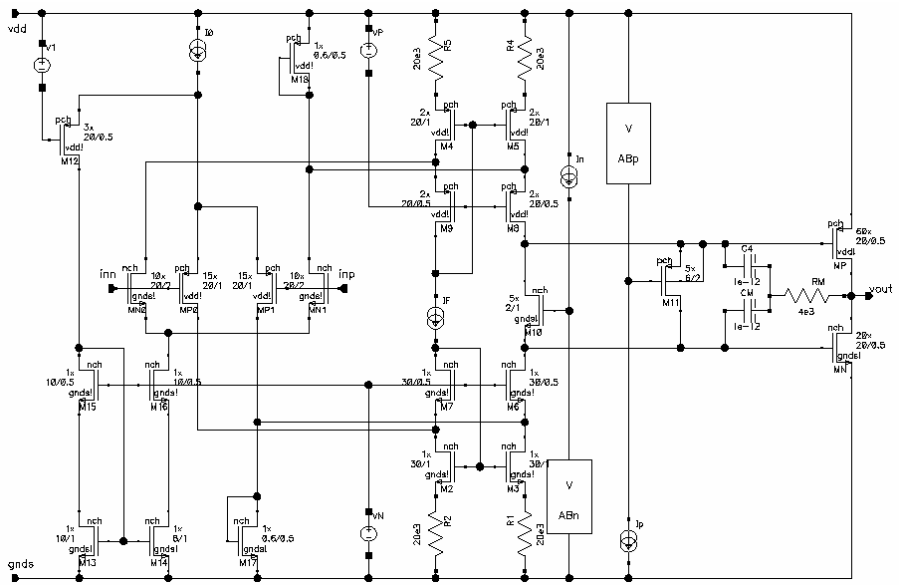


Figure 9-11. OpAmp with class AB circuit and tail current switch

The simplest gate voltage generator for the gate bias in class AB circuit is shown in fig. 7-6b, where M_1 of this figure must match M_N of fig. 9-10 and M_2 must match M_{10} with corresponding aspect ratios. Value of I_n (or I_p) is

not important. This can be as small as it is comfortable for the designer, for example, one can take $I_p = I_n \approx 0.5$uA which is about 1% of the total I_q.

In operation, the drain voltages of M_N and M_{10} devices do not match the drain voltages of the diodes in the gate voltage generator. This results in an I_q error. This error changes with the supply variation, and it can be simulated separately from the OpAmp using the circuit shown in fig. 9-12a. Here the I_{13} and I_{14} represent the currents of the folded cascode. Voltage sources V_1 and V_2 substitute the opposite transistor of the output stage. In simulations, the $V_1=V_2=V_0/2$ and $I_3 = I_4 = I_F/2 = 2.5$ uA were taken. The I_q of the whole output stage would be close to the average of I_{DMN} and I_{DMP} currents. Simulation results for supply range of 2.2-5 V with the gate voltage generators of fig. 7-6b are shown in fig. 9-12b.

Variation of I_{DMP} current is from 25 uA to 85 uA, and clearly is not acceptable as it will significantly increase the OpAmp I_q. Variation of I_{DMN} current is lesser: it is from 25 to 37 uA only. Yet, it would be better to decrease it as well. Simulations with a fixed drain voltage of M_{10} and M_{11} show that these variations almost exclusively depend on the mismatch between the drain voltage of M_N (or M_P) and the voltage of the corresponding diode in the gate voltage generators V_{ABn} (or V_{Abp}).

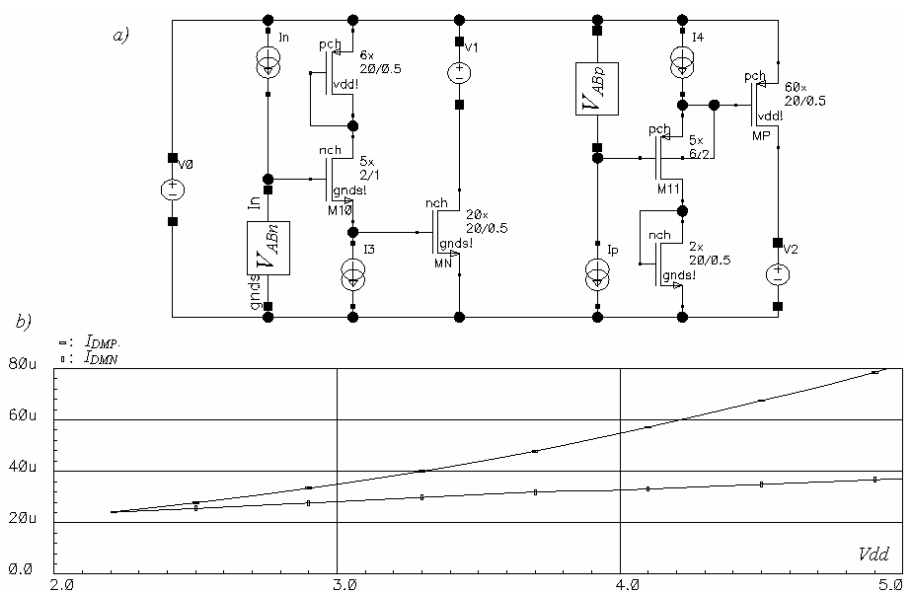


Figure 9-12. Simulation of the quiescent current I_q versus supply voltage

An improvement for V_{AB} generator, shown in fig. 7-13 [118], is applied now to V_{ABp} of our example. This improved generator is shown in fig. 9-13a, and its connection with the output transistor is shown in fig. 9-13b.

In this generator the drain voltage of transistor M_{Ps} varies together with the drain voltage of M_P due to the operation of the feedback loop M_2-M_1. Current through M_1 is equal to the I_p, the current through M_{Ps} is equal to the $(2I_p - I_p) = I_p$. To have the ratio $I_{DMP}/I_p = 50$, transistors M_{Ps} and M_P should have the same aspect ratio.

This feedback loop comprises a one-stage amplifier (M_{Ps}) and, hence, is stable without any dedicated compensation.

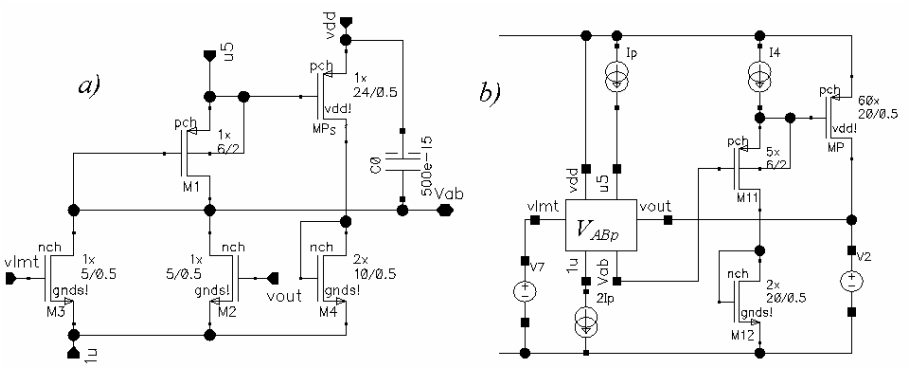


Figure 9-13. Biasing voltage generator for PMOS transistor in the class AB circuit

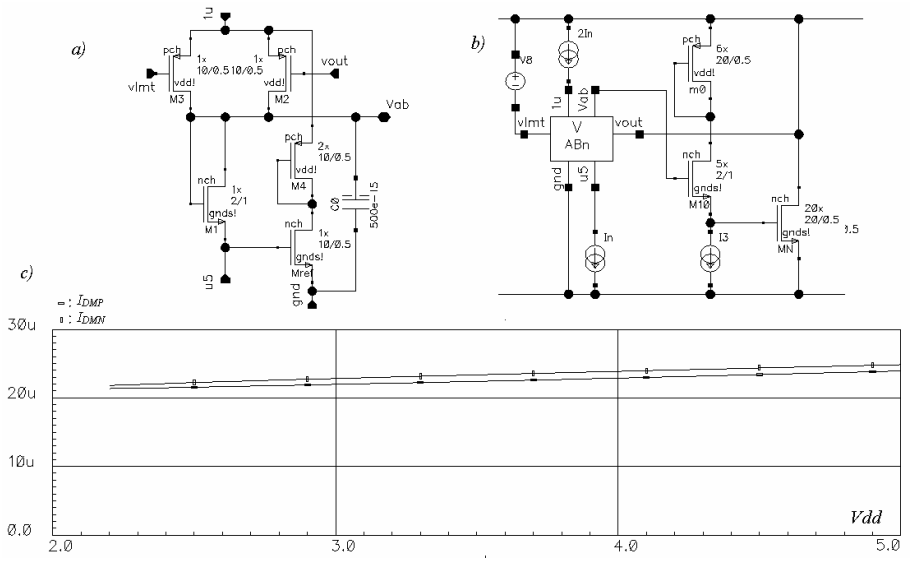


Figure 9-14. Biasing voltage generator for NMOS transistor in the class AB circuit and simulation of the quiescent current I_q versus supply voltage

The purpose of the transistor M_3 is to disable the loop operation when the output voltage is approaching the negative rail and to prevent the saturation of the current source $2I_p$. Gate of the M_3 can be connected to any voltage source within 1-2 V of the negative rail. Gate lengths of the M_2 and M_3 can be minimal, and their width is chosen using fig. 9-1 for operating current I_p = 0.5 uA and targeting the weak inversion mode.

Transistor M_{11} is in the OpAmp signal gain path and operates in the common-gate connection. To avoid extra delay, the high-frequency impedance of the gate voltage source should be low, and this is achieved using the capacitor C_0.

Complementary circuit is added as the gate voltage generator for transistor M_{10}. Circuit is shown in fig. 9-14a, and its connection with output transistor, M_N, is given in fig 9-14b. Simulation results for the both gate voltage generators are given in fig. 9-14c.

As one can see, with an improved gate voltage generators the variation of the output stage I_q versus supply voltage does not exceed 20%, which is usually sufficient.

Body of M_{10} in a single N-well process is the substrate. With a 5V supply, the V_{ds} of this transistor may be close to 4 V and the drain-body leakage current may significantly decrease the OpAmp DC gain by reducing the equivalent resistance seen at the gate of M_P (see chapter 4). Circuit of fig. 9-15 shows a feedback loop which keeps the drain voltage of M_{10} small and, hence, the leakage negligible [35].

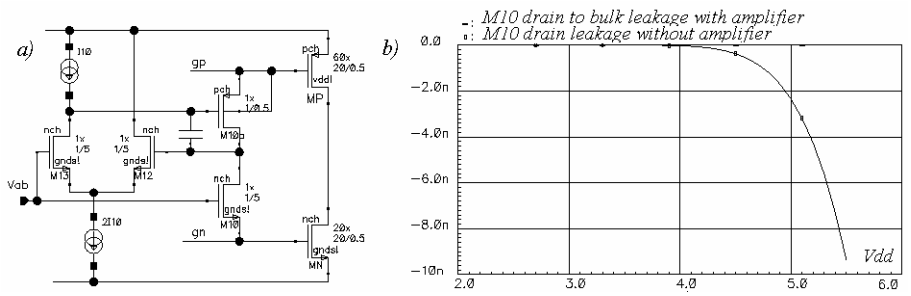


Figure 9-15. M_{10} drain-body leakage reduction

It comprises a PMOS transistor M_{10a} and the differential pair M_{12}/M_{13}, biased by the current sources $2I_{10}$ and I_{10}. This loop includes two gain stages (M_{12}/M_{13} and M_{10A}) and may be unstable. Frequency compensation can be done using a capacitor between the gate and drain of M_{10a} or by decreasing the loop gain using a smaller *W/L* of M_{12} and M_{13}. Here, the current I_{10} is chosen equal to 0.5uA, M_{12} and M_{13} have *W/L*=1/5 both, and M_{10a} has *W/L*=

1/0.5. This loop is compensated with a 0.05pF capacitor between the gate and drain of M_{10a}.

The OpAmp of fig. 9-11 has been updated with the performance enhancement units shown in fig. 9-13, 9-14 and 9-15 (see the final circuit in fig. 9-21). It is time now to simulate class AB operation and verify its performance. Showing the output transistor currents versus output voltage results are in fig. 9-16a. This simulation was performed with the OpAmp in inverting configuration (closed loop gain G=-1, R_L=500 Ohm), by swinging the V_2.

The input error of the OpAmp in this test configuration is shown in fig. 9-16b. This error does not have any sharp variations near 2.5V, where the load current switches between the output devices. This means that the class AB circuit does not affect the OpAmp gain. The curve is relatively linear (any sharp change in this curve slope would mean significant distortions).

This characteristic changes rapidly when the output voltage approaches the supply rails, and M_N (or M_P) starts to operate in the triode region. These points are within 100 mV from the rail voltages, with 5 mA load current, as required.

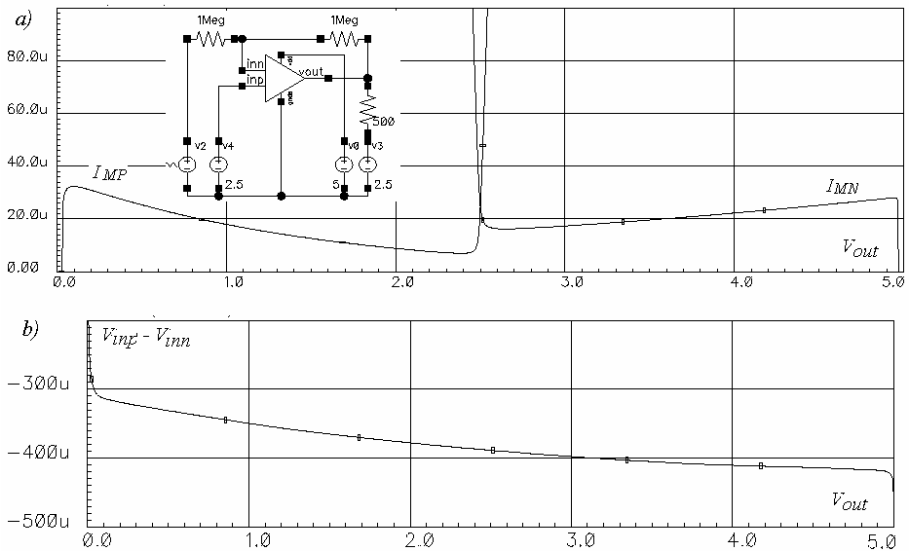


Figure 9-16. Currents of the output transistors and input offset with R_L=500 Ohm

The DC gain of the OpAmp can be estimated by the ratio of the output voltage swing to the input voltage change. Using this estimation, one finds that, with a 500 Ohm load, the open-loop gain is about 94 dB. This may be low for some applications, but this parameter can be improved as shown in the next subchapter.

In the circuit of fig. 9-11, the input differential pairs were merged together. Both NMOS and PMOS input differential pairs in this OpAmp operate in weak inversion so that their g_m depends only on the tail current. In this case, a simple tail current switch (V_1, M_{12} and the current mirror M_{13}-M_{16}) is sufficient to provide stability of the input stage g_m during common-mode input voltage changes through the switching point between the two pairs (subchapter 5.1). To narrow the switching range of the common-mode input voltage, transistor M_{12} may operate close to weak inversion while running the full tail current. Using fig. 9-1 again and assuming a 10uA tail current for M_{12} one can choose its size as 60/0.5 (3 devices of 20/0.5). The sizes of the mirror transistors can be smaller, and their dimensions are chosen as 10/1 for M_{13}/M_{14}, 10/0.5 for M_{15} and 8/0.5 for M_{16} (the NMOS tail current should be slightly smaller for the same transconductance).

AC gain/phase simulations are now repeated for the OpAmp with its full output stage and for both PMOS and NMOS input stages to ensure that the delays of the class AB circuit do not affect the OpAmp bandwidth and phase margin (fig. 9-17).

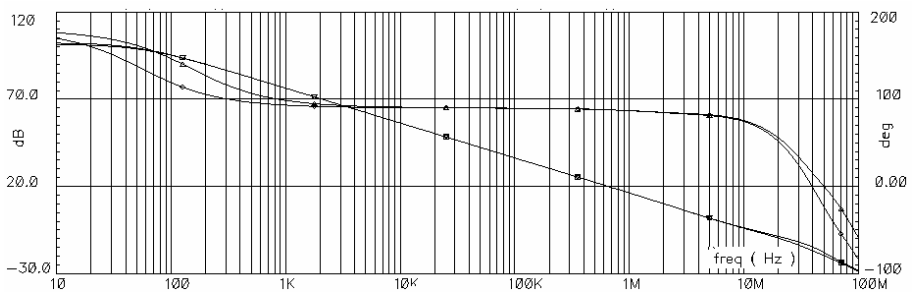


Figure 9-17. Gain and phase (+180°) when operating with PMOS and NMOS inputs

Simulations of the step-response transients for large and small steps in the follower configuration and with different common-mode input voltages are also necessary at this stage (fig. 9-18a). These simulations demonstrate that the OpAmp has a positive slew rate of 2.6 V/us, and negative slew rate of 2.2 V/us. Settling behavior does not show any abnormalities.

The small-signal step transients in the required range of load capacitances verify stability with the capacitive load (fig. 9-18b). As can be seen, the settling frequency and settling time change with load capacitance. This means that the bandwidth of the OpAmp changes. But the OpAmp is stable with a capacitive load of up to 1 nF (only part of the responses is shown to avoid cluttering).

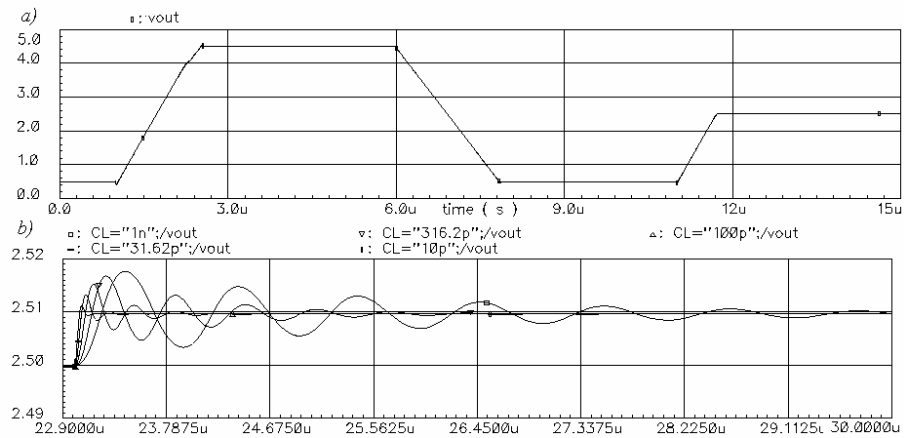


Figure 9-18. Large and small signal step responses

9.5 Gain boost and folded cascode current source

DC gain of the two-stage OpAmp is proportional to the equivalent resistance at the gates of output devices (chapter 4). DC gain in this example is around 90 dB, with 500 Ohm load. This is not sufficient for precision applications. In addition, it will drop below 90 dB over temperature and process variations, which would not satisfy industry standards and customer expectations.

Increase of the DC gain can be achieved by boosting the folded cascode output impedance as discussed in subchapters 6.3 and 6.4.

The gain boost structure of fig. 4-7 has been chosen for this design example. A gain boost amplifier is also gain boosted as shown in fig. 6-11d. Detailed schematics of the gain boost amplifiers, with transistor sizes, are shown in fig. 9-19a and 9-19b. All of the biasing currents are 0.5 uA.

Circuit of fig. 6-5 has been chosen for implementation of the floating current source. The detailed schematic of the floating current source feedback amplifier is in fig. 9-19c. This amplifier requires a capacitor C_0 to speed up operation of the folded cascode current mirror (it provides the high-frequency bypass of the amplifier M_0/M_1).

In the presence of gain boost amplifiers, the offset of these amplifiers defines the matching of the drain voltages in the folded cascode current mirror, and matching of the cascode transistors (M_6/M_7 and M_8/M_9 in fig. 9-11) is no longer necessary. Transistor M_9 is a part of the floating current source. It should have a longer channel and operate in strong inversion to improve the accuracy of the bias current I_F. Size of M_9 has been chosen to be

10 of (5/10) devices and the size of M_{Ps} is 5/10. The reference current (in M_{Ps}) is 0.5 uA.

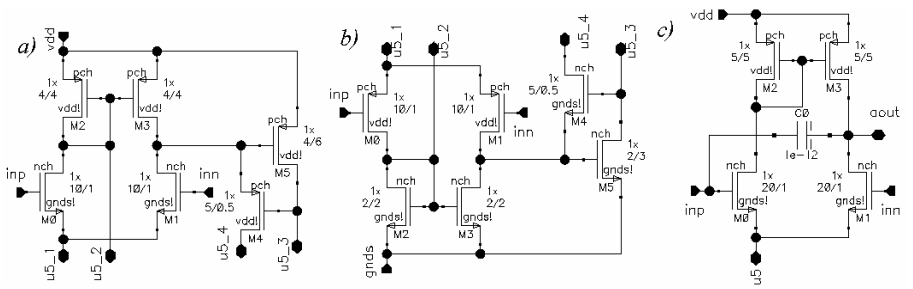


Figure 9-19. Amplifiers of the gain boost and floating current source

The final circuit of the OpAmp, with all cells including class AB, gain boost and floating current source is shown in fig. 9-21. Biasing cells, discussed later, are also shown.

To avoid cluttering of the schematic, all current source connections are labeled, with the value of current and number. Gate connections of the cascoding transistors are also labeled using *pcsc* for PMOS voltage source and *ncsc* for NMOS source notations.

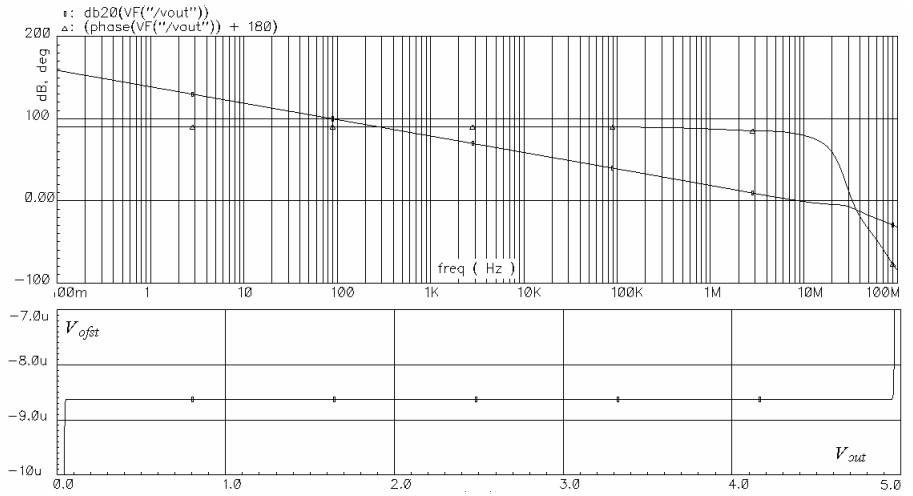


Figure 9-20. Gain/phase and input error of the OpAmp with gain boost

At this step, the AC open loop gain/phase versus frequency and input offset, V_{off}, versus V_{out} should be simulated again to verify the expected improvements in the gain (fig. 9-20).

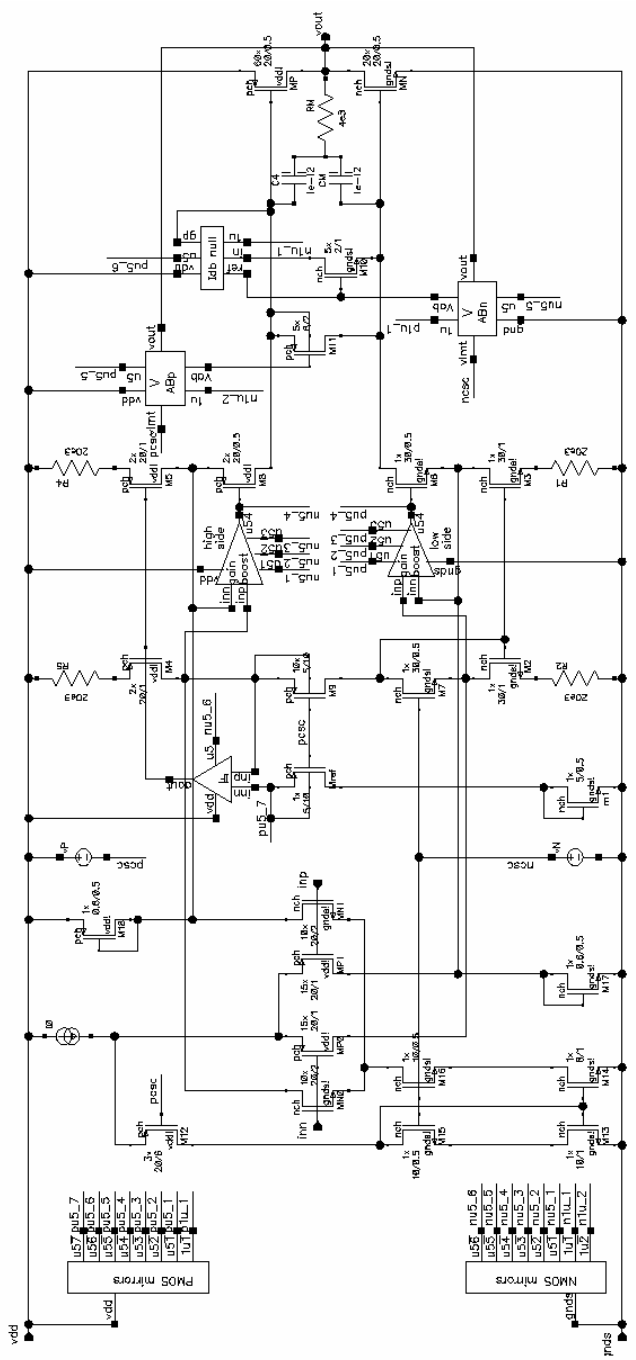


Figure 9-21. OpAmp full schematic

9.6 Biasing

The final step in the OpAmp circuit development is the design of the biasing core, design of the voltage sources for folded cascode (they can be reused for other cascoding transistors in the circuit) and replacement of the ideal current sources with the real ones.

The OpAmp in the example requires the following types of currents:

- g_m-matching TC for the input stage tail current (subchapter 3.2);
- negative TC (subchapter 3.3) for the folded cascode to avoid the transistors triode operation that may be caused by the voltage drop on resistors;
- with zero temperature coefficient (subchapter 3.3) for the rest of the circuit in order to keep the overall I_q stable with temperature.

The complete circuit diagram of the biasing cell is shown in fig. 9-22. First of all, the cell comprises the subregulated g_m-matching core (transistors M_0/M_2, resistors $R_1/R_2/R_3$) similar to that of fig. 3-17, where the current-input amplifier replaces the voltage amplifier. This amplifier (input stage M_5-M_9, loaded with the active mirror M_{11}/M_{12}, and output stage M_{13}) does not require frequency compensation, and the capacitor C_0 of 0.05 pF is included to shorten the core transient when the supply voltage changes in steps. Transistor M_0 is sized to operate approximately in the same inversion region as the OpAmp input devices. Currents of M_0 and M_2 are 0.25 uA each, set by adjustment of the R_3 and of their aspect ratio $(W/L)_0/(W/L)_2$ (the device M_0 should be wider). The voltage drop across the resistors R_1 and R_2 is chosen to be approximately 300 mV. The core start-up circuit (subchapter 3-7) includes transistor M_{17} and the leakage resistor R_{start}. Gate of the M_{17} is biased to $2V_{be}$ below V_{DD} by Q_0 and Q_1, and when M_{13} starts to conduct and sets the voltage V_{sbrg}, the transistor M_{17} will be turned off. The diode-connected transistor, M_{16}, (normally turned off) eliminates the possible, but unwanted, condition of the circuit when M_4 is in triode region. In this case, there would be no current in $M_5/M_6/M_7/M_9$, and, therefore, no regulation in the core feedback loop.

Results of the transient simulations of the g_m-matching core, while slowly ramping the supply voltage and while switching it between 5 V and 2.2 V, are shown in fig. 9-23. The core starts to operate at supply of approximately 1.6 V. There are no current spikes while starting. Variation of the core current versus supply voltage does not exceed 1% statically and 5% dynamically in the full supply range.

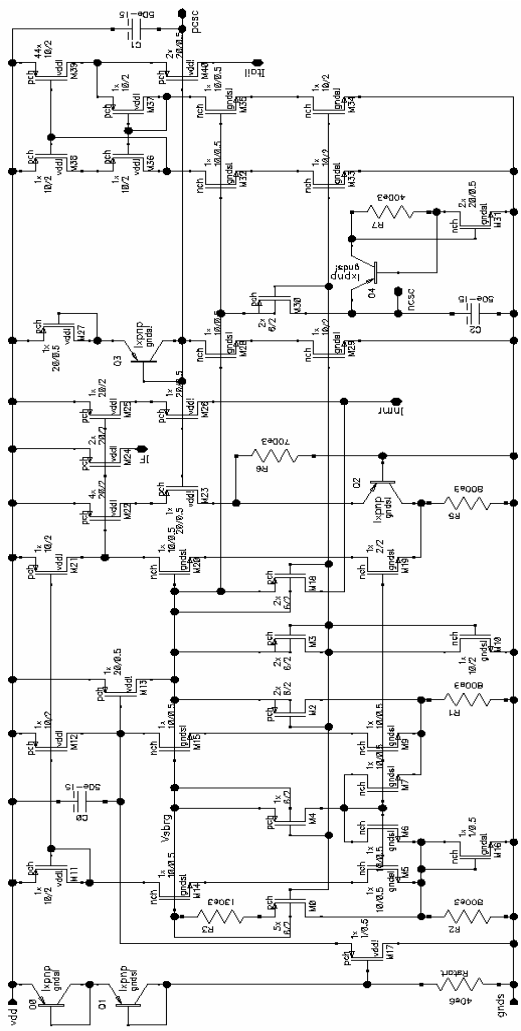


Figure 9-22. Biasing circuit

The negative TC V_{be}/R current core employs transistor Q_2 and the stabilizing feedback loop Q_2-M_{19}-M_{20}-M_{22}-M_{23}-Q_2 (fig. 9-22). This feedback loop controls $I_{D22} \approx I_{eQ2} + V_{be}/R_6$. Current I_{eQ2} is set by resistor R_5, and, if this resistor is large, then $I_{eQ2} << V_{be}/R_6$ and $I_{D22} \approx V_{be}/R_6$. Current I_{D22} is then mirrored by M_{24}, and the current I_{D24} (labeled as I_F) is used in the OpAmp folded cascode circuit.

Finally, I_{D25} of the M_{25} with a negative TC and the current I_{D18} with a positive TC are summed together to create a zero TC current, I_{nmr}, that is used as the reference current for the rest of the biasing current sources.

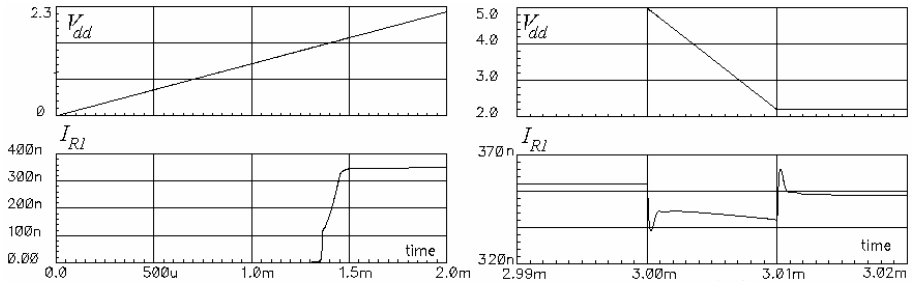


Figure 9-23. gm-matching core start up and stability versus supply voltage variations

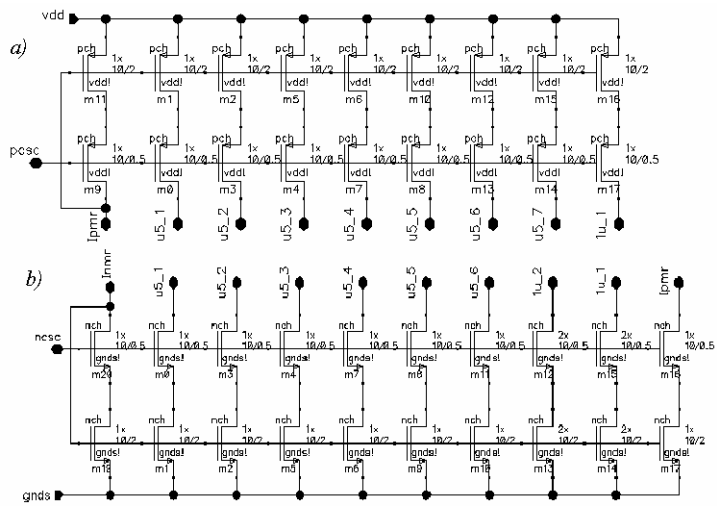


Figure 9-24. Current mirror cells of the biasing circuit

This biasing unit also includes the voltage sources for the gates of the cascoding transistors in the OpAmp. These voltages are created by the voltage drop on Q_3 and M_{27} using the current source M_{28} and M_{29} (for *pcsc* voltage line), and by the voltage drop on the subcircuit Q_4, M_{31} and R_7, using the current source M_{30} (for *ncsc* line). These voltages are $V_{pcsc} = (V_{dd} - V_{gs27} - V_{be3})$ and $V_{ncsc} = (V_{gs31} + V_{be4} - I_{D30}R_7)$.

The tail current source, I_{tail}, consists of the cascoded mirror $M_{38}/M_{39}/M_{40}$ with the boosted output resistance in accordance with the circuit shown in fig. 3-14. This tail current source provides a current with a positive TC in order to keep g_{m2} of the OpAmp input stage constant (see subchapter 3.2).

The zero-TC current (through the pin I_{nmr}) is used in the NMOS and PMOS current mirrors cells (fig. 9-24a and 9-24b) to create the rest of biasing currents for the OpAmp.

9.7 Finale of the amplifier design

Once the biasing cell is complete, one has to replace the ideal voltage and current sources in fig. 9-21 by ones from the cell. The ESD cells and wirebond macromodels must then be added, and the final simulations performed. They include:

1. Simulations of AC open loop gain/phase frequency responses (similar to those of fig. 9-21a) while sweeping the load resistance, capacitance and current, as well as common-mode input voltage in their required ranges;

2. Simulations of large- and small-step transients (similar to that of fig. 9-18) in the follower and negative unity gain connections while sweeping the load capacitance and output current;

3. Simulation of transients when entering and leaving output saturation (overload recovery).

All of these simulations should be repeated using all process corner models and at all the extreme temperatures of the required range. If any problems are discovered, correction must be found and applied and all simulations repeated.

Simulations of the CMRR and PSRR usually do not produce credible results. Even with intentionally mismatched input components, the silicon measurement results are always worse than the simulated ones, often by an order of magnitude. Still, such simulations help to define the variables that affect the values of CMRR and PSRR, and they should be done as well.

The parameters of the resulting OpAmp based on the simulations are given in table 9-1.

Finally, the designer should prepare schematics for layout and write layout instructions.

Schematics for the layout preparations include designation of matching components, high-current buses (these buses require wider than minimum metal width) and virtual separation of the signal and power supply rails with metal resistors. In fig. 9-21 the high-current buses are shown with thick lines and the output stage has separate supply nodes. OpAmp root diagram of fig. 9-25 comprises ESD cells and metal resistors connecting the signal and power supply rails together.

Layout instructions include the maximum allowable die dimensions (defined by the package) and an approximate floor plan with the locations of key components and bonding pads.

Package and part pin out (this is standard for all OpAmps, except for ones in new packages) define bonding pad locations. There is very little

freedom in the moving of the bonding pads, especially in high-speed or power OpAmps, where the bondwire parasitics can significantly affect the parameters.

Placement of the output devices is also important. They generate heat that can affect other components. They are connected to the supply and output bonding pads via wide metal, and these buses metal, especially in power OpAmps, can occupy more die area than the transistors themselves.

The components defining the voltage offset (the input pair and folded cascode mirrors) should be located in the least mechanically stressed areas (closer to the die center and along die axis) [3].

Table 9-1. Example OpAmp parameters

Parameter	Min.	Typ.	Max.
Input offset, mV		1	5
temperature drift, uV/°C			10
PSRR uV/V, V_S = 1.1V to 3V, V_{CM} = (V-) +0.3V			200
Input voltage range	(V-)-0.3		(V+)+0.3
CMRR V_{supply} = ±2.5V -2.5V <V_{CM} < 1V or 1.5V < V_{CM} <2.7 V, uV/V			200
Noise			
Input noise voltage density 1kHz nV/sqrt(Hz)			50
Input noise peak-to-peak, 0.1 to 10 Hz, uV			3
Open-loop gain, R_L = 500 Ohm, dB	100		
Output			
Voltage swing to rail, I_L = 5mA, mV	100		
Output current, mA			20
Frequency response			
GBW, R_L = 10 kOhm, CL = 10 pF, MHz	5		7
Slew rate, V/us	2		
Settling time, 1V step, 0.1%, us			3
Stability with capacitive load, G=+1, pF	1000		
Power supply, total, V	2.2		5.5
Quiescent current, uA		50	60
Temperature range, °C	-55		125

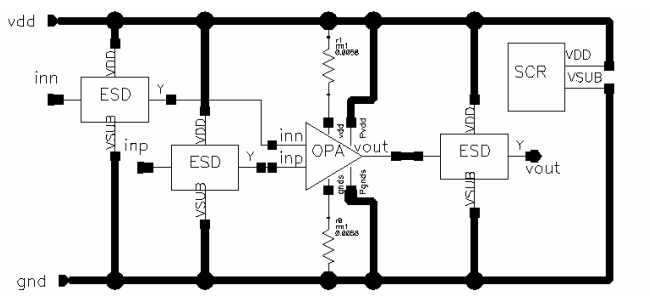


Figure 9-25. OpAmp root circuit diagram for layout

Example of the OpAmp layout (similar to the design example) is shown in fig. 9-26 [14]. The layout of the signal path is straight along the die axis, starting from the input stage and going via the folded cascode to the output devices. Biasing, gain boost and other service units are placed on the sides. The empty areas on the die are filled with supply bypass capacitor (empty areas on the die, if any, may always be filled with something useful to enhance the performance).

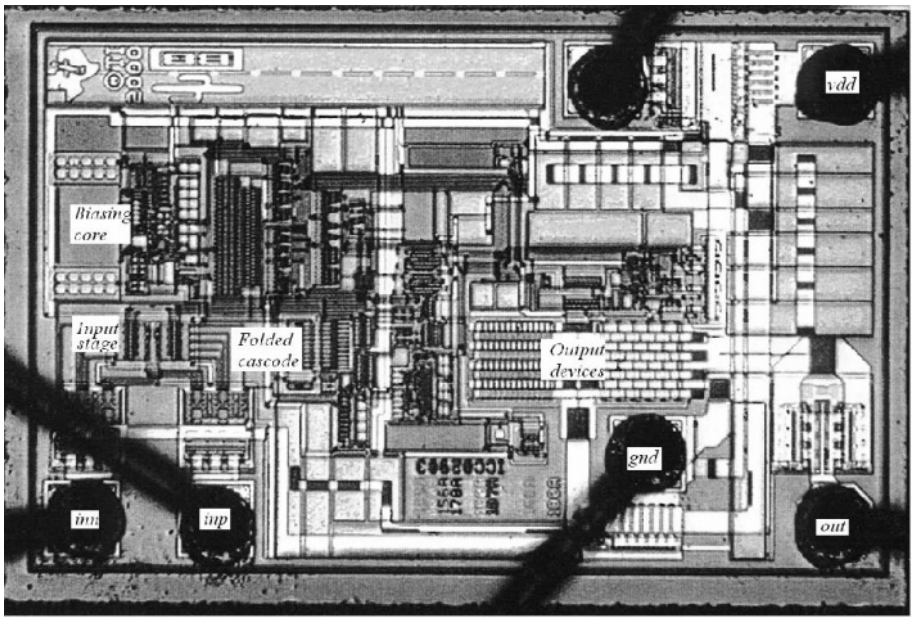


Figure 9-26. OpAmp layout

Appendix. Structural properties and linear transformations in the multidimensional systems with symmetric links

In this Appendix we consider the basic properties of the one class of multidimensional linear systems. This particular class important for design of electronic systems, including operational amplifiers, can be reduced to the systems with two input and two output variables and one common mode feedback. The equivalent structural graphs are presented, and the transfer coefficients for common mode and differential mode components of the input and output signals are derived. These transfer coefficients are also used to evaluate the influence of variations of the structure parameters. Some simple transistor amplifiers with n inputs and n outputs are shown as examples.

1. We consider a special class of multidimensional (multichannel) linear systems with symmetric links that can be represented by a graph of fig. A1a or by equations:

$$z_i = -q_i^{'}(u_i + r_i x_c) \qquad (i = \overline{1,n}) \qquad \text{(A.1)}$$

$$x_c = \rho_c y_c\,; \qquad y_c = \sum_{i=1}^{n} c_i z_i$$

where u_i are input signals, z_i are output signals, c_i, $q_i^{'}$, r_i $(i = \overline{1,n})$ and ρ_c are transfer coefficients. The same transformation of the input vector **u** into output vector **z** may also be represented by the equivalent signal graph in fig. A1b.

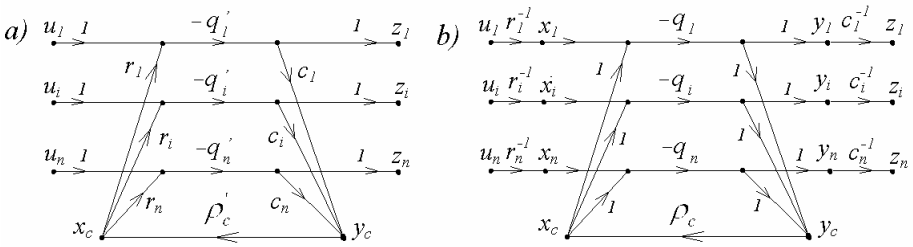


Figure A1. Graph of the multidimensional system with common-mode feedback

In this figure $q_i = r_i q_i^{'} c_i$, and the signals $x_i = r_i^{-1} u_i$ and $y_i = c_i z_i$ $(i = \overline{1,n})$ are inner variables. The systems described by the graphs of fig. A1 may be called the *systems with common mode feedback,* which can be positive or negative.

2. Any channel a out of n can be chosen and the rest of $(n-1)$ channels could be combined in a single channel b. It allows one to limit consideration by the two-dimensional system of fig. A2, where

$$q_b = \sum_{i\neq a}^{n} q_i; \qquad x_b = \frac{1}{q_b}\sum_{i\neq a}^{n} q_i x_i = \frac{1}{q_b}\sum_{i\neq a}^{n} q_i r_i^{-1} u_i; \qquad y_b = \sum_{i\neq a}^{n} c_i^{-1} y_i \tag{A.2}$$

This substitution simplifies the analysis, and, at the same time, helps to visualize, without complicated matrix formalism, the general properties of the system and effects of the variations of the system parameters.

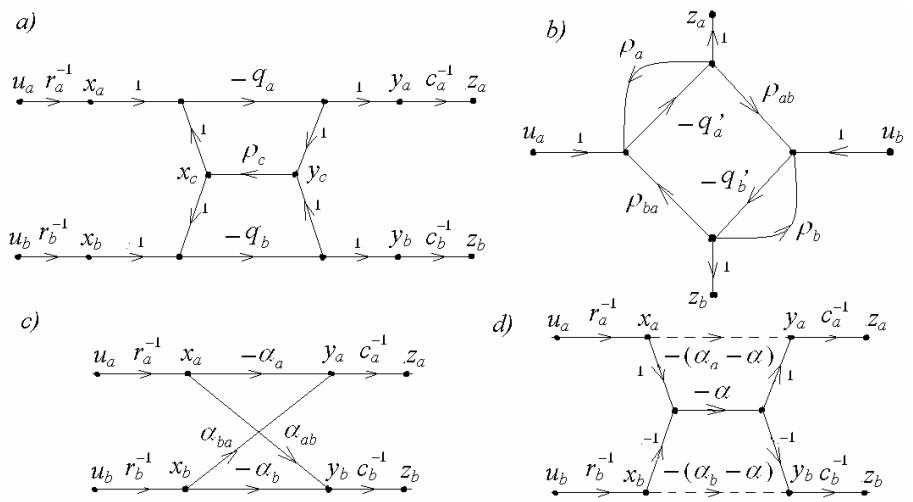


Figure A2. Graph of the two-dimensional system with common-mode feedback

3. The graph of fig. A2a may be transformed in other equivalent forms that may be more convenient for analysis and synthesis of particular electronic systems. The considered two-channel systems can be easily reduced to the equivalent bridge representation (fig. A2b). Signal graphs shown in fig. A2a and A2b are equivalent if

$$\begin{aligned} &\rho_a = r_a\rho_c c_a; \qquad \rho_{ba} = r_a\rho_c c_b; \qquad q_a = r_a q_a' c_a \\ &\rho_{ab} = r_b\rho_c c_a; \qquad \rho_b = r_b\rho_c c_b; \qquad q_b = r_b q_b' c_b \end{aligned} \tag{A.3}$$

From these relationships one can find that

$$\frac{\rho_a}{\rho_{ab}} = \frac{\rho_{ba}}{\rho_b} = \frac{r_a}{r_b}; \qquad \frac{\rho_a}{\rho_{ba}} = \frac{\rho_{ab}}{\rho_b} = \frac{c_a}{c_b} \tag{A.4}$$

The bridge structures have a high sensitivity to variations of feedback parameters that violate the relationships satisfying (A.4), especially when $r_a = r_b = 1$ and/or $c_a = c_b = 1$. This property is routinely used in instrumentation (bridge sensors).

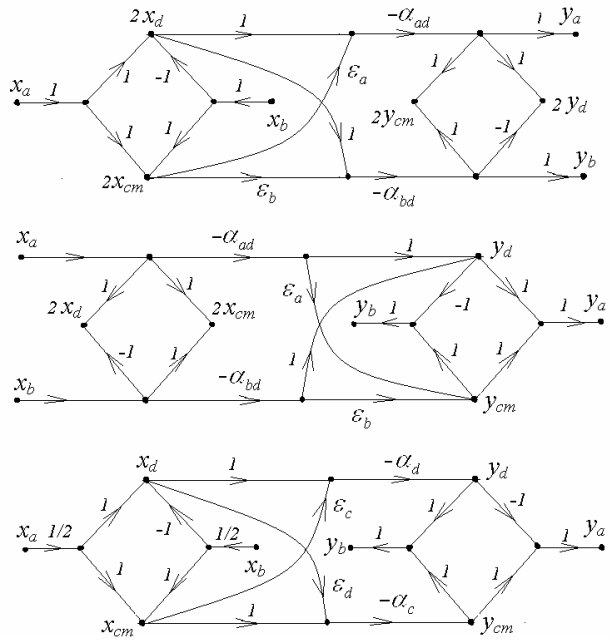


Figure A3. Two-dimensional systems with eliminated feedback

4. Using the rules of signal graph transformations [25] the graph of fig. A1a can be modified into forms without feedback (fig. A2c and d) where:

$$z_a = -c_a^{-1}(\alpha_a r_a^{-1} u_a - \alpha\, r_b^{-1} u_b)$$
$$z_b = -c_b^{-1}(\alpha\, r_a^{-1} u_a - \alpha_b\, r_b^{-1} u_b) \tag{A.5}$$

$$\alpha_a = \frac{q_a(1+\rho_c q_b)}{1+\rho_c(q_a+q_b)}; \qquad \alpha_b = \frac{q_b(1+\rho_c q_a)}{1+\rho_c(q_a+q_b)};$$
$$\alpha_{ab} = \alpha_{ba} = \alpha = \frac{q_a \rho_c q_b}{1+\rho_c(q_a+q_b)}. \tag{A.6}$$

If

$$\rho_c \cdot \min(q_a, q_b) >> 1 \tag{A.7}$$

then the transfer coefficients become

$$\alpha_a = \alpha_b = \alpha = (q_a^{-1} + q_b^{-1})^{-1} = \tilde{q} \tag{A.8}$$

Introducing differential and common mode components of the input signal

$$x_d = (x_a - x_b)/2; \qquad x_{cm} = (x_a + x_b)/2; \tag{A.9}$$

one transforms the graph of fig. A2b into the graph of fig. A-3a. For this graph, one has in the channel a,

$$y_a = -2\alpha_{ad}(x_d + \varepsilon_a x_{cm}) \tag{A.10}$$

where

$$\alpha_{ad} = \frac{1}{2}(\alpha_a + \alpha) = \frac{q_a/2 + \rho_c q_a q_b}{1+\rho_c(q_a+q_b)} \approx \tilde{q}/2 \tag{A.11}$$

is the gain of the differential signal in the channel a, and

$$\varepsilon_a = \frac{\alpha_a - \alpha}{\alpha_a + \alpha} = \frac{1}{1 + 2\rho_c q_b} \tag{A.12}$$

is the attenuation of the common mode signal in the channel a. For channel b

$$y_b = -2\alpha_{bd}(x_d + \varepsilon_b x_{cm}) \tag{A.13}$$

and, substituting b instead of a and changing signs, one can find similar to (A.11) and (A.12) expressions for the differential gain, α_{bd}, and common mode attenuation, ε_b, in the channel b.

The same parameters ε_a, ε_b, α_{ad}, α_{bd} define the dependence of the differential output signal $y_d = (y_a - y_b)/2$ and common mode signal $y_{cm} = (y_a + y_b)/2$ as a function of the variable x_a (graph of fig. A-3b):

$$(y_d)_a = -\alpha_{ad} x_a; \qquad (y_{cm})_a = -\varepsilon_a (y_d)_a \tag{A.14}$$

or as a function of x_b:

$$(y_d)_b = -\alpha_{bd} x_b; \qquad (y_{cm})_b = -\varepsilon_b (y_d)_b \tag{A.15}$$

One can also find the dependencies of the differential, y_d, and common mode, y_{cm}, components from x_d and x_{cm} (graph of fig. A3c). After simple substitutions one obtains that

$$\begin{aligned} y_d &= -\alpha_d (x_d + \varepsilon_c x_{cm}); \\ y_{cm} &= -\alpha_c (\varepsilon_d x_d + x_{cm}), \end{aligned} \tag{A.16}$$

where

$$\begin{aligned} \alpha_d &= \alpha_{ad} - \alpha_{bd} = \frac{2\rho_c q_a q_b + (q_a + q_b)/2}{1 + \rho_c (q_a + q_b)}; \\ \varepsilon_c &= \frac{\alpha_a - \alpha_b}{\alpha_{ad} - \alpha_{bd}} = \frac{(q_a - q_b)}{(q_a + q_b) + 4\rho_c q_a q_b}; \end{aligned} \tag{A.17}$$

and

$$\begin{aligned} \alpha_c &= \frac{\alpha_a + \alpha_b}{2} - \alpha = \frac{(1/2)(q_a + q_b)}{1 + \rho_c (q_a + q_b)} \approx \frac{1}{2\rho_c}; \\ \varepsilon_d &= \frac{\alpha_a - \alpha_b}{\alpha_a + \alpha_b - 2\alpha} = \frac{q_a - q_b}{q_a + q_b}. \end{aligned} \tag{A.18}$$

The equations (A.10)-(A.18), and the corresponding graphs of fig. A3 represent different cases useful in design of instrumentation and amplification systems.

5. When the left side in (A.7) increases the attenuation of the common mode components x_{cm} and y_{cm} also increases. This allows one to amplify even very small differential signal x_d and to resolve very small difference y_d or to obtain the signal proportional to a very small difference of two parameters:

$$2x_d = u_0 (r_a^{-1} - r_b^{-1}); \qquad y_d = y_0 (c_a^{-1} - c_b^{-1}), \tag{A.19}$$

and $u_a = u_b = u_0$ or $y_a = y_b = y_0$ even if the channel gains are different, i.e. $q_a \neq q_b$.

If, in addition to (A.7), the gain of one channel is much larger than the gain of another one, for example,

$$q_b >> q_a \text{ then } \tilde{q} \to q_a. \tag{A.20}$$

Also, as follows from (A.6) and (A.10):

$$\alpha \approx \alpha_a \approx \alpha_b \approx \tilde{q} \approx q_a \tag{A.21}$$

The condition (A.20) may occur if the channel a has a local feedback, or the gain of this channel has changed due to nonlinear distortions or variation of bias with time. It may also occur if the channel b represents the equivalent of $(n-1)$ channels in the n -channel system and

$$q_b = \sum_{i \neq a}^{n} q_i \approx (n-1)q >> q_a \qquad \text{(A.22)}$$

The n two-channel systems that obtained from a single n -channel system by choosing one channel as a -channel and grouping all other channels into b -channel may have different transmissions α_a ($a = \overline{1,n}$) for the differential mode signal $(x_a - x_b)/2$. Yet, the attenuation parameters of the common mode signal $\varepsilon_a = (2\rho_c q_b)^{-1}$ and $\varepsilon_b = (2\rho_c q_a)^{-1}$ at the input and output of each pair of channels may be very small if the condition (A.7) is satisfied.

6. The two-channel structure of fig. A-2 may be interpreted as a general structure of a control system where q_b , r_b , c_b are transfer coefficients of the controlled object, its input sensors and the output transducers, and q_a , r_a , c_a are transfer coefficients of the model representing desirable properties. The parameter ρ_c represents the feedback amplifier amplifying the difference between the outputs of the object and the model.

For example, one can consider as a model the compensation unit included in the feedback link (fig. A-4a) or in the feedforward link (fig. A-4b).

If the inequalities A.7 and A.20 are valid, and $r_a^{-1} >> r_b^{-1}$ (see fig. A-2a) then the outputs of the object and the model, as it follows from A.4, are approximately equal, i.e.

$$y_a \approx c_a z_a \approx y_b \approx c_b z_b \approx q_a r_a^{-1} u_a \qquad \text{(A.23)}$$

and defined by the parameters q_a , r_a and the control signal u_a of the model.

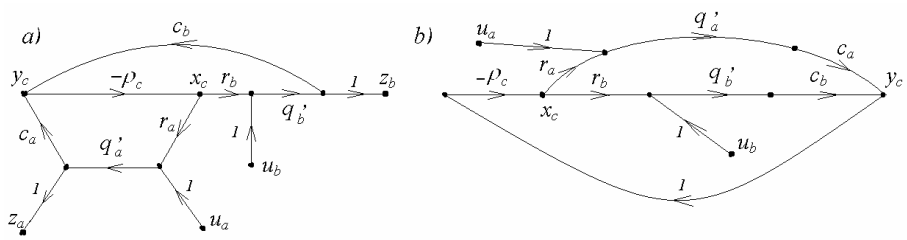


Figure A4. Systems with models

These properties are preserved for any dimension n_b of the object ($n_a + n_b = n$, $n_a >> n_b$). The relationships A.9 - A.15 allow one to evaluate the degree of independence of the control system parameters on the object properties.

7. If one changes the sign of the link transfer in channel a or in feedback as shown in fig. A-5a ($r_a, c_a, \rho_{ab}, \rho_{ba} < 0$) then the common-mode feedback suppresses the differential (and not the common-mode as before) component of the input and output signals. In other words,

the parametric differences between channels as well as the difference between input signals due to different biasing, noise, etc. are reduced in this system.

All previous relationships are valid for this structure as well, but $\varepsilon, \varepsilon_a, \varepsilon_b$ now mean attenuation of the differences while $\alpha_{ad}, \alpha_{bd}, \alpha_d$ are now the transfer coefficients of the common-mode signals.

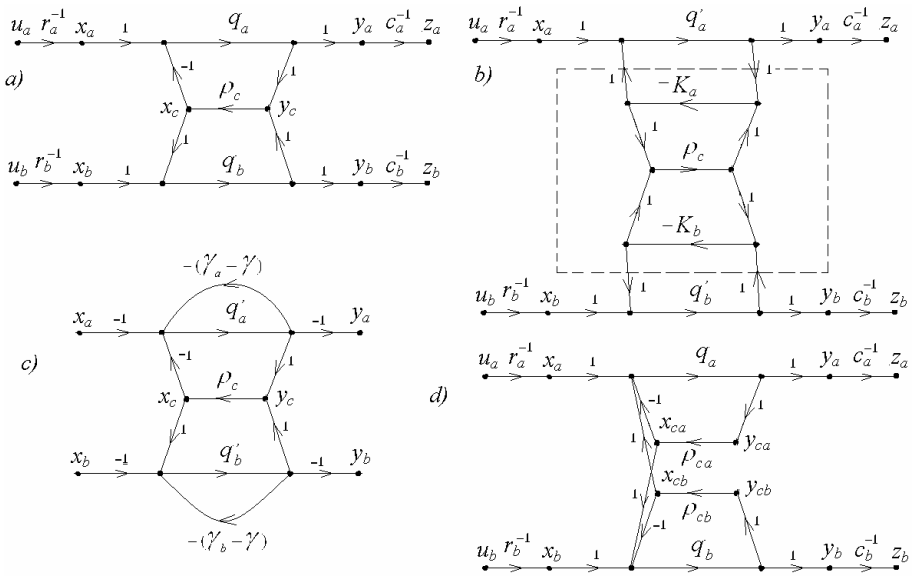


Figure A5. Useful modifications of two-dimensional system graph

8. Some other useful modifications of the general structure are shown in fig. A5b and A5c. The figure A-5b highlights the fact that the common-mode feedback link in the general structure of fig. A-2a can also have, in turn, the same structure. It can also be seen as if each of the links q_a', q_b' has its own local feedback but the parameters of these feedbacks are defined by the main common-mode link.

The signal inversions can be implemented at the input links and the units q_a', q_b' can have their own local feedback links $-(\gamma_a - \gamma)$ and $-(\gamma_b - \gamma)$ as shown in fig. A-5c. The resulting link transfers are:

$$q_a = \frac{q_a'}{1+(\gamma_a - \gamma)q_a'} \text{ and } q_b = \frac{q_b'}{1+(\gamma_b - \gamma)q_b'} \tag{A.24}$$

If the condition (A.7) is satisfied for this structure, and the structure is changed to the form without feedbacks as in fig. A-2d, then

$$\alpha = \frac{q_a q_b \gamma}{1+\gamma(q_a + q_b)} \approx (\frac{1}{q_a} + \frac{1}{q_b})^{-1} \approx (\frac{1}{q_a'} + \frac{1}{q_b'} + \frac{1}{\rho_c}) \tag{A.25}$$

Other relationships for this structure can be derived in the same way as A.8 – A.18.

The fig. A-2a common mode feedback link can also be split in two links. This creates the equivalent structure shown in fig. A-5d. Here each of $q_a^{'}, q_b^{'}$ links has its own feedback, and these feedbacks affect also the opposite signal paths.

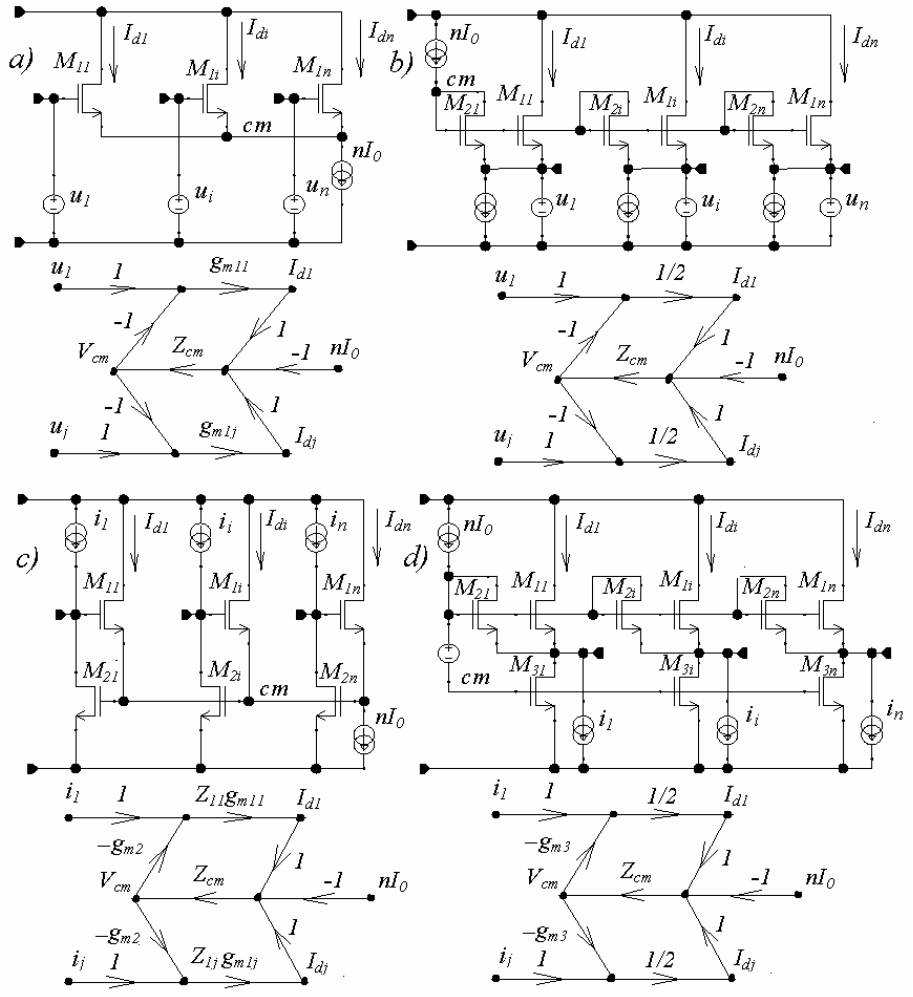


Figure A-6. Four types of input stages

9. With addition of the cross links at input and output, the general structure with common-mode feedback can be considered as functionally complete unit for implementation of linear multidimensional systems (in the same sense as NOR and NAND cells are functionally complete for logic operations). For example, it can be used to represent voltage and current relationships in arbitrary electrical networks.

Table A1 shows three types of summing circuits for current or voltage signal sources having a common load. The complementary distributing circuits of the signal from the

common source between loads are shown in Table A2. All these circuits can be described by the same graph with common-mode feedback as shown in the bottom row of the tables. The choice of the input and output variables in the graph defines the type of the circuit it represents.

10. The differential input stages of the voltage and current amplifiers can also be reduced to the general structure as shown in fig. A6. Fig. A6a represents the voltage controlled current sources (large output impedances, high input impedances). The output current of each source depends on the difference between the input voltage of this source and the average input voltage, namely

$$I_i - I_0 = g_{mi}(u_i - \overline{u}) \; ; \; \overline{u} = \frac{1}{n}\sum_{i=1}^{n} u_i \quad i = \overline{1,n} \tag{A.26}$$

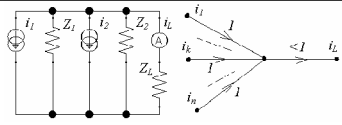
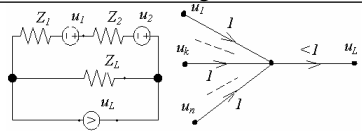
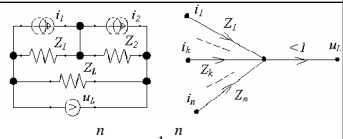
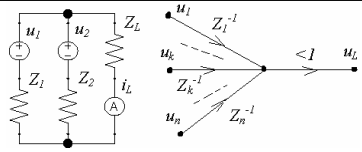
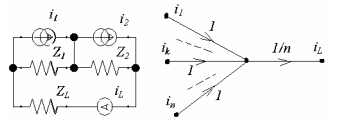
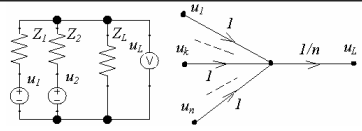
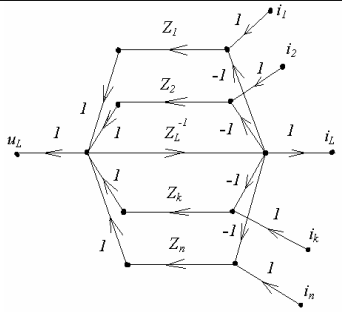
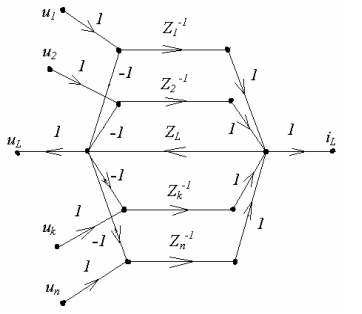
The same structure may represent a completely different stage of fig. A6b. This stage has a large common-mode and small differential input resistance. The input resistance of each channel is equal to $1/g_{mi}$, yet, the output currents are expressed by equation (A.26).

The input stage with large differential and small common-mode input impedance is shown in fig. A-6c. The input stage where both low differential and common mode impedances are low is shown in fig. A-6d.

Numerous implementations of these stages are possible depending on the component cell library at hand.

11. The differential structure with common-mode feedback can model very wide variety of multidimensional systems. This structure helps to understand the common properties and differences of these systems, and, which is the most important, to synthesize the circuits with desired properties.

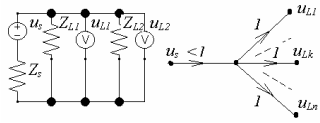
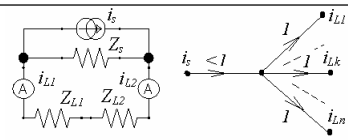
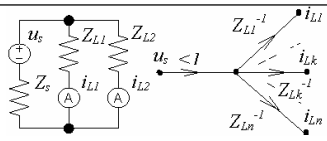
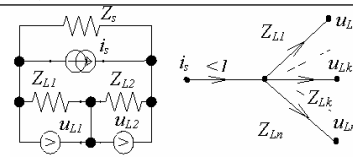
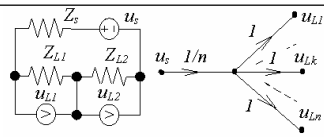
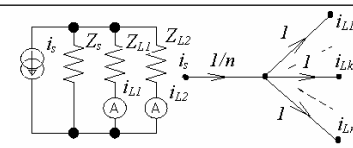
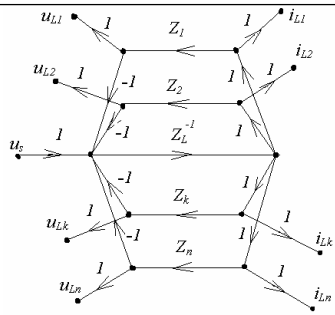
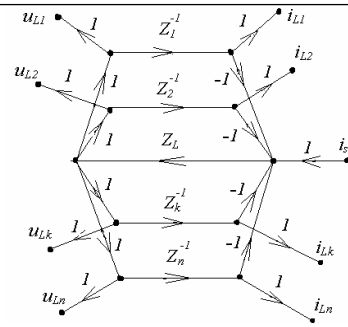
Table AA1. Summing circuits using passive components

	Current	Voltage	Application
Simple summing	 $i_L = (1+\sum_{k=1}^{n} G_k^{-1})^{-1} \sum_{k=1}^{n} i_k =$ $= \sum_{k=1}^{n} i_k$ if $\sum_{k=1}^{n} \left\| G_k^{-1} \right\| << 1$ and $z_L \to 0$	 $u_L = (1+\sum_{k=1}^{n} G_k)^{-1} \sum_{k=1}^{n} u_k =$ $= \sum_{k=1}^{n} u_k$ if $\sum_{k=1}^{n} \left\| G_k \right\| << 1$ and $z_L \to \infty$	Power circuits with ideal sources Feedback circuits with ideal instrument
Weighted summing	 $u_L = (1+\sum_{k=1}^{n} G_k)^{-1} \sum_{k=1}^{n} z_k i_k \approx$ $\approx \sum_{k=1}^{n} z_k i_k$ if $\sum_{k=1}^{n} \left\| G_k \right\| << 1$ and $z_L \to \infty$	 $i_L = (1+\sum_{k=1}^{n} G_k^{-1})^{-1} \sum_{k=1}^{n} i_k =$ $= \sum_{k=1}^{n} i_k$ if $\sum_{k=1}^{n} \left\| G_k^{-1} \right\| << 1$ and $z_L \to 0$	Low power circuits, only with ideal instrument
Averaging summing	 $i_L = \frac{\sum_{k=1}^{n} G_k^{-1} i_k}{\sum_{k=1}^{n} G_k^{-1} + 1} \approx \frac{\sum_{k=1}^{n} z_k^{-1} i_k}{\sum_{k=1}^{n} z_k^{-1}}$ $i_L = \frac{1}{n} \sum_{k=1}^{n} i_k$ if $z_k = z_0$	 $u_L = \frac{\sum_{k=1}^{n} G_k u_k}{\sum_{k=1}^{n} G_k + 1} \approx \frac{\sum_{k=1}^{n} z_k^{-1} u_k}{\sum_{k=1}^{n} z_k^{-1}}$ $u_L = \frac{1}{n} \sum_{k=1}^{n} u_k$ if all $z_k = const = z_0$	Low power circuits with ideal instrument Power circuits if $G_k = G_0$
Summing circuit graph			

Notes. 1. Ideal current instrument $z_i \to 0$; ideal voltage instrument $z_i \to \infty$; ideal voltage source $z_i \to 0$; ideal current source $z_i \to \infty$.

2. $G_k = z_L^{-1} z_k$, $k = \overline{(1,n)}$; $G_k^{-1} = z_L z_k^{-1}$.

Table AA2. Distributing circuits using passive components

	Voltage	Current	Application
Simple distributing	 $u_{Lk} = (1+\sum_{k=1}^{n} G_k)^{-1} u_s \approx u_s$ if $\sum_{k=1}^{n} \lvert G_k \rvert >> 1$ and $z_s \to 0$	 $i_{Lk} = (1+\sum_{k=1}^{n} G_k^{-1})^{-1} i_s \approx i_s$ if $\sum_{k=1}^{n} \lvert G_k^{-1} \rvert >> 1$ and $z_s \to \infty$	Power circuits with ideal source Measurements with ideal instruments
Weighted distributing	 $i_{Lk} = z_{Lk}^{-1}(1+\sum_{k=1}^{n} G_k^{-1})^{-1} u_s \approx z_{Lk} i_s$ if $\sum_{k=1}^{n} \lvert G_k^{-1} \rvert >> 1$ and $z_s \to 0$	 $u_{Lk} = z_{Lk}(1+\sum_{k=1}^{n} G_k)^{-1} i_s \approx z_{Lk} i_s$ if $\sum_{k=1}^{n} \lvert G_k \rvert >> 1$ and $z_s \to \infty$	Input circuits with ideal instruments or sources
Averaging distributing	 $u_{Lk} = \frac{G_k^{-1} u_s}{1+\sum_{k=1}^{n} G_k^{-1}} \approx \frac{G_k^{-1} u_s}{\sum_{k=1}^{n} G_k^{-1}} \approx \frac{u_s}{n}$ if $\sum_{k=1}^{n} \lvert G_k^{-1} \rvert >> 1$; $z_{Lk} = z_0$	 $i_{Lk} = \frac{G_k i_s}{1+\sum_{k=1}^{n} G_k} \approx \frac{G_k i_s}{\sum_{k=1}^{n} G_k} \approx \frac{i_s}{n}$ if $\sum_{k=1}^{n} \lvert G_k \rvert >> 1$; $z_{Lk} = z_0$	Input circuits with ideal sources
Distributor graph			

Notes. 1. Ideal current instrument $z_i \to 0$; ideal voltage instrument $z_i \to \infty$; ideal voltage source $z_s \to 0$; ideal current source $z_s \to \infty$.

2. $G_k = z_{Lk}^{-1} z_s$, $k = (\overline{1,n})$; $G_k^{-1} = z_{Lk} z_s^{-1}$.

References

1. B.D.H. Tellegen ”La recherche pour une serie complete d’elements de circuit ideaux non-lineares”, *Rendiconti Del Seminario Matematico e Fisico di Milano*, vol. 25, pp. 134-144, April 1954.
2. E. Charbon, R. Gharpurey, P. Miliozzi, R. Meyer, and A. Sangiovanni-Vincentelli, *Substrate noise: analysis and optimization for IC design*, Dordreht, The Netherlands: Kluwer, 2001.
3. F. Fruett, and G.C.M. Meijer *The piezojunction effect in silicon integrated circuits and sensors*, Dordrecht, The Netherlands: Kluwer, 2002
4. A. Andreini, C. Contiero, and P. Galbiati, “BCD technology for smart power ICs”, in *Smart Power ICs : technologies and applications*, B. Murari, F. Bertotti, and G.A. Vignola, eds., New York: Springer, 1996.
5. W. Bakalski et al., “A Monolithic 2.45GHz Power Amplifier in SiGe-Bipolar with 0.4W Output Power and 53% PAE at 2V”, *Proc. ESSCIRC-2002*, pp. 223-226, Florence, Italy, Sept. 24-26, 2002.
6. C. Contiero, A. Andreini, and P. Galbiati, “Roadmap Differentiation and Emerging Trends in BCD Technology”, *Proc. ESSDERC-2002*, pp. 275-282, Florence, Italy, Sept. 24-26, 2002.
7. M. Berkhout, “A class D output stage with zero dead time”, *Proc. ISSCC-2003 Digest Tech. Papers*, pp. 1-8, San Francisco, Feb. 9-13, 2003.
8. M. C. Wilson et al., “A 12Volt, 12GHz Complementary Bipolar Technology for High Frequency Analogue Applications”, *Proc. ESSDERC-2002*, pp. 375-378, Florence, Italy, Sept. 24-26, 2002.
9. P. Gray, and R. Meyer, *Analysis and design of analog integrated circuits*, New York: Wiley, 1993.
10. J. H. Huijsing, *Operational amplifiers*, Dordrecht, The Netherlands: Kluwer, 2001.
11. J. Williams, ed., *Analog circuit design*, Oxford, England: Butterworth-Heinemann, 1991.
12. J. O’Connor, and I. McDermott, *The art of systems thinking,* Thorsons, 1997.
13. D. Jones, and K. Martin, *Analog integrated circuit design*, New York: Wiley, 1997.
14. V. Ivanov, and S. Zhang “250 MHz CMOS Rail-to-Rail IO OpAmp”, *Proc. ESSCIRC-2002*, pp. 183-186, Florence, Italy, Sept. 24-26, 2002.

15. V. Ivanov, W. Meinel, and J. Zhou, "Method and circuit for trimming offset and temperature drift for operational amplifiers and voltage references", US pat. 6,614,305, 2003
16. A. Tang, "A 3uV-Offset Operational Amplifier with 20 nV/$\sqrt{Hz}$ Input Noise PSD at DC Employing both Chopping and Autozeroing", *ISSCC-2002 Digest Tech. Papers*, vol. 1, pp. 386-387, San Francisco, Feb. 3-7, 2002.
17. G. M. Weinberg, *An Introduction to General Systems Thinking*, New York: Dorset House, 2001
18. G. Altshuller, *And Suddenly the Inventor Appeared*, Worcester, MA: Technical Innovation Center, Inc., 1996.
19. G. Polya, *How to solve it*, Princeton University Press, 1971.
20. В. Н. Иванов "Динамически-эквивалентные системы", *Чувствительность систем управления*, т. 3, стр. 296-304, Владивосток, 1975, (V. N. Ivanov, "Dynamically equivalent systems", in *Control systems sensitivity,* vol. 3, pp. 296-304, Vladivostok, 1975, in Russian).
21. В.Н. Иванов "Принципы построения мажоритарных систем управления", *Технические средства автоматики*, Москва: Наука, 1971 (V.N. Ivanov "Design principles of the control systems with voting", in *Technical means of automatics*, Moscow: Nauka Pub. House, 1971, in Russian).
22. R. Eschausier, and J. Huijsing, *Frequency compensation techniques for low-power operational amplifiers*, Dordrecht, The Netherlands: Kluwer, 1997.
23. A. Ochoa, "A systematic approach to the analysis of general and feedback circuits and systems using signal flow graphs and driving point impedance" ", *IEEE Trans. Circuits and Systems*, Part II, vol. 45, no. 2, pp. 187–195, Feb. 1998.
24. H. Schmid, "Circuit transposition using signal-flow graphs", *Proc. ISCAS-2002*, vol. 2, pp. 25-28, Phoenix, AZ, May 26-29, 2002.
25. S. Mason, "Feedback theory—Further properties of signal flow graphs", *Proc. IRE*, vol. 44, no. 7, pp. 920–926, July 1956.
26. B. Thandri, and J. Silva-Martínez, "A Robust Feedforward Compensation Scheme for Multistage Operational Transconductance Amplifiers With No Miller Capacitors", *IEEE J. Solid-State Circuits*, vol. SC-38, no. 2, pp. 237-243, Feb. 2003.
27. V. Ivanov, "High-gain amplifier with multipath forward feeding for frequency compensation" US pat. 5,917,376, 1998.
28. W.-H. Ki "Signal Flow Graph Analysis of Feedback Amplifiers", *IEEE Trans. Circuits and Systems*, Part I, vol. 47, no. 6, pp. 926-933, June 2000.
29. *Exar Inc. product catalog*, 1985.
30. E. Seevinck, *Analysis and synthesis of translinear integrated circuits*, New York: Elsevier, 1988.
31. B. Minch, "A low-voltage MOS cascode bias circuit for all current levels", *Proc. ISCAS-2002*, vol. 3, pp. 619-622, Phoenix, AZ, May 26-29, 2002.
32. E. Seevinck, "CMOS translinear circuits" in *Analog Circuit Design*, Dordrecht, The Netherlands: Kluwer, pp. 323-336, 1996.
33. G. F. Olson, *Dynamic analogies*, Princeton, NJ: Van Nostrand, 1958.
34. J. Steininger, "Understanding of the wide-band MOS transistors", *IEEE Circuits and Devices Magazine*, vol. 6, no. 3, pp. 26 –31, May 1990.
35. S. Sanchez, V. Ivanov, and G. Johnson, "Rail-to-rail class AB output stage of the operational amplifier with wide supply range", US pat. 6,545,538, 2003.
36. B. Hosticka et al., "Design methodology for analog monolithic circuits", *IEEE Trans. Circuits and Systems*, Part I, vol. 41, no. 5, pp. 387-394, May 1994.

37. B. Hosticka et al., “CMOS operational amplifier with nearly constant settling time”, *Proc. IEE*, Part G, vol. 137, no. 4, pp. 309-314, Aug. 1990.
38. K. Bult, and G. Geelen “A fast-settling CMOS OpAmp for SC circuits with 90 dB DC gain”, *IEEE J. Solid-State Circuits*, vol. SC-25, no. 6, pp. 1379-1384, Dec. 1990.
39. *OPA363 data sheet and application notes*, *www.ti.com*.
40. A. Boni, "Op-Amps and Startup Circuits for CMOS Bandgap References With Near 1-V Supply", *IEEE J. Solid-State Circuits*, vol. SC-37, no. 10, pp. 1339-1343, Oct. 2002.
41. W. M. Leach, “On the application of the Thevenin and Norton equivalent circuits and signal flow graphs to the small-signal analysis of active circuits”, *IEEE Trans. Circuits and Systems*, Part I, vol. 43, no. 11, pp. 885-893, Nov. 1996.
42. R. D. Kelly, “Electronic circuit analysis and design by driving-point impedance techniques”, *IEEE Trans. Education*, vol. E-13, pp.154-167, Sept. 1970.
43. Y. Hu, and M. Sawan, “A 900 mV high PSRR CMOS voltage reference dedicated to implantable micro-devices”, *Proc. ISCAS-2003*, vol. 1, pp. 373-376, 25-28 May, Bangkok, Thailand, 2003.
44. V. Ivanov, and D. Baum “A 3A 20 MHz BiCMOS/DMOS power operational amplifier: structural design approach”, *ISSCC-2003 Digest Tech. Papers*, pp. 1-10, San-Francisco, Feb. 9-13, 2003.
45. K. de Langen, and J. Huijsing, *Compact low-voltage and high-speed CMOS, BiCMOS and bipolar operational amplifiers*, Dordrecht, The Netherlands: Kluwer, 1999.
46. J. Harrison, and N. Weste, “A 500 MHz anti-alias filter using feed-forward OpAmps with local common-mode feedback”, *ISSCC-2003 Digest Tech. Papers*, pp. 1-9, San-Francisco, Feb. 9-13, 2003.
47. B. Kamath, R. Meyer, and P.Gray “Relationship between frequency response and settling time of operational amplifiers”, *IEEE J. Solid-State Circuits*, vol. SC-9, no. 6, pp. 347-352, Dec. 1974.
48. *OPA620 data sheet and application notes*, *www.ti.com*
49. R. Reay, and G. Kovacs “An unconditionally stable two-stage CMOS amplifier”, *IEEE J. of Solid-State Circuits*, vol. 30, no. 5, pp. 591-594, May 1995.
50. B. Ahuja, “An improved frequency compensation technique for CMOS operational amplifiers”, *IEEE J. Solid-State Circuits*, vol. SC-18, no. 12, pp. 629-633, Dec. 1983.
51. V. Ivanov, “CMOS input stage with wide common-mode range”, US pat. 6,509,795, 2003.
52. R. Griffith, R. Vyne, R. Dotson, and T. Petty, “A 1V BiCMOS rail-to-rail amplifier with n-channel depletion-mode input stage”, *ISSCC-1997Digest Tech. Papers*, pp. 352-353, 484, Feb. 6-8, San Francisco, 1997.
53. R. Blauschild, “Differential amplifier with rail-to-rail capability”, US pat. 4,532,479, 1985.
54. R. Burt, private communication.
55. *OPA627 data sheet and application notes*, *www.ti.com*.
56. S. Zhang, and V. Ivanov, “Rail-to-rail CMOS input stage of the operational amplifier with constant transconductance” US pat. 6,462,619.
57. *OPA277 data sheet and application notes*, *www.ti.com*.
58. V. Ivanov, J. Zhou, and W. Meinel, “Operational amplifier input stage and method”, US pat. 6,642,789, 2003.
59. *OPA725 data sheet and application notes*, *www.ti.com*.
60. A. Tang, “Bandpass spread spectrum clocking for reduced clock spurs in autozeroed amplifiers”, *Proc. ISCAS-2001*, vol. 1, pp. 663-666, May 6-9, Sidney, 2001.

61. T. Stockstad, and H. Yoshizawa, “A 0.9 V 0.5 uA rail-to-rail CMOS OpAmp”, *IEEE J. Solid-State Circuits*, vol. 37, no. 3, pp. 286-292, March 2002.
62. V. Ivanov, J. Zhou, and W. Meinel, “Method and circuit for trimming offset and temperature drift for operational amplifiers and voltage references”, US pat. 6,628,169, 2003.
63. A. Wang et al., “A Mixed-mode ESD protection circuit simulation-design methodology”, *IEEE J. Solid-State Circuits*, vol. 38, no. 6, pp. 995-1006, 2003.
64. Y. Tsividis, *Operation and modeling of MOS transistor*, New York: McGraw-Hill, 1987.
65. V. Ivanov, “Rail-to-rail input/output operational amplifier and method”, US pat. 6,150,883, 2000.
66. R. Wassenaar et al., “Design aspects of rail to rail CMOS OpAmp”, Proceedings of the 1st VLSI Workshop, pp. 23-28, Columbus OH, May 1997.
67. G. Johnson, and V. Ivanov, “Rail-to-rail input/output operational amplifier and method”, US pat. 6,356,153, 2002.
68. S. Sanchez, V. Ivanov, and W. Meinel, “Amplifier gain boost circuitry and method”, US pat. Application, 2003.
70. J. Ramirez-Anguilo et al., “The flipped voltage follower: a useful cell for low-voltage low-power circuit design”, *Proc. ISCAS-2002*, vol. 3, pp. 26-29, Phoenix AZ, May 26-29, 2002.
71. E. Vittoz, “Low-power low-voltage limitations and prospects in analog design” in *Low-power Low-voltage Integrated filters and Smart Power*, Dordrecht, The Netherlands: Kluwer, 1994
72. D. Monticelli, “Class AB output circuit with large swing”, US pat. 4,570,128, 1986.
73. G. Erdi, “A Precision Trim Technique for Monolithic Analog Circuits”, *IEEE J. Solid-State Circuits*, vol. SC-1, no. 6, pp. 412-416, Dec. 1975.
74. J. Price, “Programmable circuit”, US pat. 3,191,151, 1962.
75. M. Cohen, B. Unger, and J. Milkosky, "Laser machining of thin films and integrated circuits", Bell Syst. Tech. Journal., vol. 47, pp. 385-405, 1968.
76. E. Swenson, “The evolution of laser processing in semiconductor manufacturing”, *Proc. 26th Midwest Symp. Circuits and Systems*, p. 212, Puebla, Mexico, Aug. 15-16, 1983.
77. A. Sypherd, and N. Solman, “Automatic laser encoding of semiconductor ROM”, *Proc. National Electronic Conference*, pp. 206-209, Dec. 9-11, Chicago, IL, 1968.
78. G. Chaplin, and A. Owens, "Some transistor input stages for high-gain d.c. amplifiers", Proc. IEE, Part B, No. 21, pp. 249-257, May 1958.
79. F. Goodenough, “Rail-to-Rail Op Amps Use Depletion-Mode PMOSFETs”, *Electronic Design*, p. 127, Apr. 16, 1992.
80. J. Fonderie, and J. Huijsing “Operational amplifier with 1V rail-to-rail multipath-driven output stage”, *IEEE J. Solid-State Circuits*, vol. 26, no. 12, pp. 1817-1824, Dec. 1991.
81. V. Ivanov, “Push-pull stage of the operational amplifier”, US pat. 5,973,564, 1998.
82. P. Bruschi, D. Navarrini, and M. Piotto, “A high current drive CMOS output stage with tunable quiescent current limiting circuit”, *IEEE J. Solid-State Circuits*, vol. 38, no. 8, pp. 1416-1420, Aug. 2003.
83. E. Seevinck, *Analysis and synthesis of translinear integrated circuits*, PhD Thesis, University of Pretoria, 1981.
84. R. R. Troutman, *Latchup in CMOS technology : the problem and its cure*, Dordrecht, The Netherlands: Kluwer, 1986.
85. V. Ivanov, “Fast rail-to-rail class AB output stage having stable output bias current and linear performance”, US pat. 6,366,169, 2002.
86. *OPA349 data sheet and application notes*, *www.ti.com*.

87. *OPA379 data sheet and application notes*, *www.ti.com*.
88. *OPA244 data sheet and application notes*, *www.ti.com*.
89. C. Enz, and G. Temes, "Circuit techniques for reducing the OpAmp imperfections: autozeroing, correlated double sampling and chopper stabilization", *Proc. IEEE*, vol. 84, no. 11, pp. 1584-1614, 1996.
90. G. Erdi, "Common-mode rejection in monolithic operational amplifiers", *IEEE J. Solid-State Circuits*, vol. SC-5, no. 6, pp. 365-367, 1970.
91. *REF31xx data sheet and application notes*, *www.ti.com*.
92. Е. Попов, *Теория линейных систем автоматического управления и регулирования*, Москва: Наука, 1989, (E. Popov, *Linear control system theory,* Moscow, Nauka Pub. House, 1989, in Russian).
93. I. Horowitz, *Synthesis of feedback systems*, London, England: Academic Press Inc.: 1963.
94. S. Mason, and H. Zimmermann, *Electronic circuits, signals and systems*, New York: Wiley, 1960.
95. S. Zhang, and V. Ivanov, "Quick turn-on enable/disable bias control circuit", US pat. 6,400,207, 2002.
96. R. Widlar, and M. Yamatake, "Dynamic safe-area protection for power transistors employs peak-temperature limiting", *IEEE J. Solid-State Circuits*, vol. 22, no. 1, pp. 77-84, Feb. 1987.
97. R. Widlar, and M. Yamatake, "A monolithic power op amp", *IEEE J. Solid-State Circuits*, vol. 23, no. 2, pp. 527 –535, April 1988.
98. B. Murari, "Power Integrated Circuits: Problems, Tradeoffs, Solutions", *IEEE J. Solid-State Circuits*, vol. 13, no. 3, 1978.
99. D. Baum, and V. Ivanov, "Overcurrent protection circuit and method", US pat. Application, 2002.
100. M. Ivanov, *Low-Voltage Low-Power CMOS Operational Amplifier Stages with Class AB common-mode feedback,* MSEE thesis, Ohio State University, Columbus, 1997.
101. D. Baum, and V. Ivanov, "Slew rate boost circuit and method", US pat. 6,437,645, 2002.
102. S. Zhang, G. Johnson and V. Ivanov, "Slew rate boost circuit and method", US pat. 6,359,512, 2002.
103. B. Carter, "Video multiplexer uses high-speed op amps", *EDN*, 8/21, p. 87, 2003.
104. S. Zhang, and V. Ivanov, "Overload recovery circuit and method", US pat. 6,317,000, 2001.
105. *OPA355 data sheet and application notes*, *www.ti.com*.
106. P. Brokaw, "Bandgap reference design", MEAD Microelectronics Class on Power Management, Lausanne, 2003.
107. J. Haslett, "Noise performance of new Norton OpAmps", *IEEE Trans. Electronic Devices*, vol. 21, pp. 571-577, Sept. 1974.
108. B. Razavi, *Design of Analog CMOS Integrated Circuits*, New York: McGraw Hill, 2001.
109. I. Filanovsky, and L. Najafizadeh "Zeroing in On a Zero-Temperature Coefficient Point", *Proc. IEEE 45th Int. Midwest Symp. Circuits and Systems*, vol. 1, pp. 271-274, Tulsa, OK, Aug. 2002.
110. J. Solomon, "The monolithic Op Amp: a tutorial study", *IEEE J. Solid-State Circuits*, vol. 9, no. 12, pp. 314-332, Dec. 1974.
111. *Low-voltage design for portable systems*, Special Topic Evening Session, *ISSCC-2002*, Feb. 3-7, San Francisco, 2002.
112. H. Gummel, and H. Poon, "An integral charge control models of bipolar transistors", *Bell Systems Technical Journal*, vol. 49, pp 827-852, 1970.
113. *HSPICE user's manual*, Meta-Software, Inc., 2000.

114. D. Foty, *MOSFET Modeling with SPICE: Principles and Practice*, Upper Saddle River, NJ: Prentice-Hall, 1997.
115. J. F. Dickson, “On-chip high-voltage generation in NMOS integrated circuits using an improved voltage multiplier technique”, *IEEE J. Solid-State Circuits*, vol. 11, pp. 374–378, June 1976.
116. G. Palmisano, G. Palumbo, and G. Pennisi, “Harmonic distortions in class AB output stages”, *IEEE Trans. Circuits and Systems*, Part II, vol. 45, no. 2, pp. 243-250, Feb. 1998.
117. F. You et al., “Low-voltage class AB buffers with quiescent current control”, *IEEE J. Solid-State Circuits*, vol. 33, no. 6, pp. 915-920, Jun. 1998.
118. V. Ivanov, “Class AB output stage with stable quiescent current”, US pat. Application, 2003.
119. R.M. Fox, "Design-oriented analysis of DC operating-point instability", *Proc. ISCAS-1995*, vol. 1, pp. 109-112, May 1995.
120. R.M. Fox, "A general operating-point instability test based on feedback analysis", *Proc. ISCAS-1998,* v.3, pp. 550-553, Orlando FL, May-June 1998

Index