Threshold voltage based CMOS voltage reference

Y. Dai, D.T. Comer, D.J. Comer and C.S. Petrie

Abstract: The paper describes a CMOS voltage reference design that uses the temperature dependence of NMOS and PMOS threshold voltages to form a temperature-insensitive reference. No diodes or parasitic bipolar transistors are used. The circuit architecture accommodates a wide range of output voltages. A test chip is fabricated using a 0.5 μm CMOS process. The prototype achieves a temperature coefficient of 32 ppm/°C for a temperature range of −10°C to 80°C and a supply voltage sensitivity of 10 mV/V.

1 Introduction

A precision voltage reference circuit is very important in the design of many mixed-signal and analogue integrated circuits such as oscillators, PLLs and data converters. Voltage references are required to be stable over process, power supply voltage, and temperature variations.

Precision CMOS voltage reference circuits have typically been implemented using different variations of the Widlar bandgap circuit [1, 2]. Over the years, the performance of CMOS bandgap references has improved significantly [3–6]. In these designs, either diodes or bipolar junction transistors (BJTs) are used to generate a 'proportional to absolute temperature' (PTAT) voltage, which is then used to compensate the negative temperature coefficient (TC) of the bipolar junction drop (V_{BE}). Bipolar junction transistors present several problems in the implementation of bandgap references: (i) the parasitic bipolar transistor in a CMOS process is usually not very well characterised; (ii) the output voltage is fixed since the base of the parasitic transistor is grounded; and (iii) curvature compensation is needed for precision applications because V_{BE} is not a linear function of temperature over the entire temperature range. Improvements have been developed to address these problems. In [7], techniques to compensate the base current and base resistance of the transistors are discussed. The work reported in [4] and [8] developed a sub-1 V voltage reference, rather than the 1.25 V of earlier designs. Curvature compensation techniques are used in [7, 9, 10] to improve temperature stability. [3] and [7] use chopper stabilising and correlated-double sampling (CDS) techniques, respectively, to address offset voltage of the op amp. In [11], a current mode voltage reference is developed in which currents proportional to absolute temperature and inversely proportional to absolute temperature are summed to create a voltage reference. In [12], the mutual compensation of mobility and threshold voltage temperature effects is investigated.

This paper describes a new architecture for a precision CMOS voltage reference that overcomes the problems listed above. The voltage reference does not use any diodes or BJTs. Two linear voltages, one associated with PMOS threshold voltage and the other associated with NMOS threshold voltage, are combined to generate a reference voltage that is stable with temperature. The reference voltage dependency on CMOS device mobility, which is nonlinear with temperature, is eliminated. The performance of the proposed voltage reference is shown to be comparable to bandgap circuits.

2 Temperature characteristics of MOS devices

The basic operation of long-channel MOS devices is described by the following equations [14]. In the active region:

$$I_D = \frac{1}{2}\mu_n C_{ox}\left(\frac{W}{L}\right)(V_{GS} - V_{tn})^2 \tag{1}$$

and in the triode region:

$$I_D = \mu_n C_{ox}\left(\frac{W}{L}\right)\left((V_{GS} - V_{tn})V_{DS} - \frac{V_{DS}^2}{2}\right) \tag{2}$$

where C_{ox} is the gate capacitance per unit area, μ_n is the mobility of electrons near the silicon surface (μ_p is used for holes), W and L are gate width and length and V_{tn} is the threshold voltage of an NMOS device (V_{tp} is used for a PMOS device).

In these equations, there are two important temperature dependent parameters: mobility, μ_n and threshold voltage, V_{tn}. The threshold voltage is a linear function of temperature as modelled in [13]:

$$V_T(T_{device}) = V_T(TNOM) + \left(KT1 + \frac{KT1L}{L_{eff}} + KT2V_{BS}\right) \times \left(\frac{T_{device}}{TNOM} - 1\right) \tag{3}$$

where $V_T(TNOM)$ is the threshold voltage at a nominal temperature, $KT1$ is the TC of the threshold voltage, $KT1L$ is the channel-length coefficient of the threshold voltage's temperature dependence and $KT2$ is the bulk-bias coefficient of the threshold voltage's temperature dependence. This linear model has been found to provide an excellent fit for the I–V characteristics at various temperatures [13].

The mobility factor, on the other hand, is a nonlinear function of temperature [13] and its temperature

IEE Proceedings online no. 20040217

doi:10.1049/ip-cds:20040217

Paper first received 16th December 2002 and in revised form 29th April 2003. Online publishing date: 26 January 2004

The authors are with the Department of Electrical and Computer Engineering, Brigham Young University, Provo UT 84602, USA

dependence is more complicated [12]:

$$U0(T_{device}) = U0(TNOM) \times \left(\frac{T_{device} + 273.15}{TNOM + 273.15}\right)^{UTE} \quad (4)$$

where $U0(TNOM)$ is the mobility at a nominal temperature and UTE is the temperature coefficient for $U0$, which is typically in the range -2.0 to -1.5.

Since mobility is a nonlinear function of temperature, it is difficult to build voltage references that rely on MOSFET characteristics. A common compensation method involves using a PTAT voltage generated by bipolar transistors [1, 2]. However, the PTAT voltage has a constant TC; therefore this method results in a zero-TC only at one certain temperature point. Curvature compensation is then needed for better performance [9, 7, 10]. In [15] and [19], the mobility dependence is completely cancelled by multiplying a current component which is proportional to mobility and a current component which is inversely proportional to mobility using transistors operating in the subthreshold region.

In this design, the mobility factor is cancelled completely without using subthreshold characteristics. Voltages proportional to the threshold voltage of both a PMOS device and an NMOS device, thus linear with temperature, are generated and then subtracted to form the voltage reference.

3 Proposed voltage reference circuit

The proposed voltage reference is based upon two voltages, V_P and V_N, that are proportional to PMOS device threshold voltage and NMOS device threshold voltage, respectively. By subtracting V_P and V_N, the TCs are cancelled and the resulting voltage can be adjusted to the desired output reference voltage. The concept is illustrated in Fig. 1.

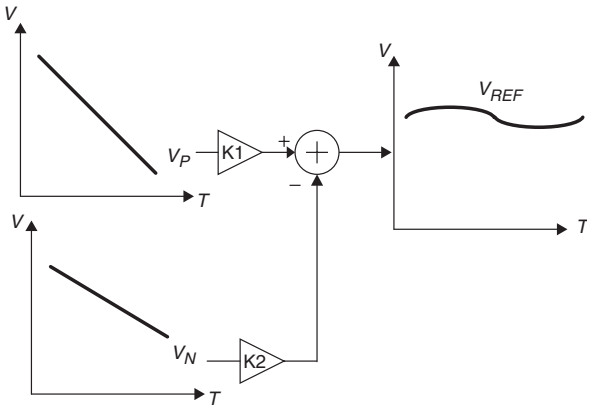

Fig. 1 *Voltage reference concept*

Figure 2 shows the proposed voltage reference schematic. In block I, V_P is generated using PMOS transistors MP_1 and MP_2. Block II generates V_N using NMOS transistors MN_1 and MN_2. The two voltages are then combined by a subtractor to generate a zero-TC reference voltage in block III. Owing to the op amps, V_P and V_N are both supply-independent. MP_3 and MP_4 are start-up devices to avoid an operating point at 0 V; they are off during normal operation.

3.1 Linear voltage from PMOS device

In block I of Fig. 2, a linear voltage is generated from PMOS devices MP_1 and MP_2. The op amp A_1 keeps the

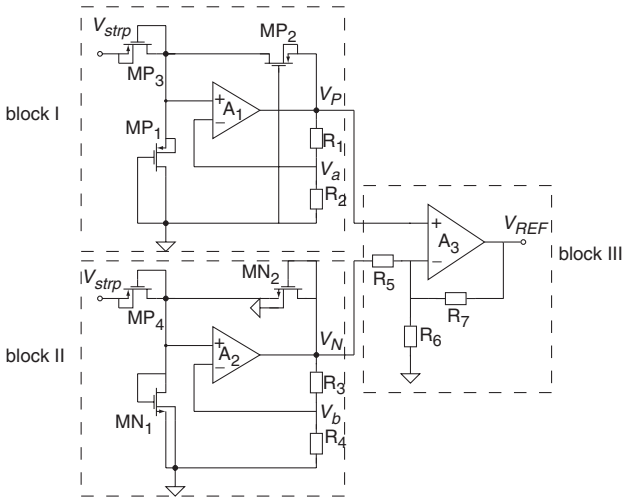

Fig. 2 *Voltage reference schematic*

drain voltage of transistor MP_2 equal to the voltage of node Va. By scaling R_1 and R_2 properly, MP_1 operates in the active region while MP_2 operates in the triode region. Transistors MP_1 and MP_2 form a positive feedback loop around the op amp, but this feedback is weak and the resistive negative feedback dominates.

There is a zero operating point for this circuit when V_P is zero and no current flows through MP_2 and MP_1. With node Vstrp at about 1.25 V, MP_3 serves as a start-up device. When the circuit operates at zero current, MP_3 turns on and starts to inject current into MP_1. This current will raise the drain voltage of MP_1 to about 1.4 V and op amp A_1 starts to operate. At this operating point, MP_3 is cut off completely and does not affect normal operation. Since the accuracy of the startup voltage is not important, V_{strp} is derived from the power supply directly using a simple voltage divider.

In (3), $KT1L$, the channel-length coefficient for threshold voltage variation with temperature, is small for this process and the threshold voltages of MP_1 and MP_2 can be considered equal. Even if different device lengths produces slightly different threshold voltages for MP_1 and MP_2, these threshold voltages still have a linear dependence on temperature according to (3).

The currents through MP_1 and MP_2 are given by the following two equations, respectively:

$$I_1 = \frac{1}{2}\mu_p C_{ox}\left(\frac{W}{L}\right)_1 (V_a - V_{tp})^2 \quad (5)$$

$$I_2 = \mu_p C_{ox}\left(\frac{W}{L}\right)_2 \left[(V_P - V_{tp})(V_P - V_a) - \frac{(V_P - V_a)^2}{2}\right] \quad (6)$$

Assuming no offset for op amp A_1

$$V_a = \frac{R_2}{R_1 + R_2} V_P$$

The high input impedance of the op amp forces the currents through MP_1 and MP_2 to be the same,

$$I_1 = I_2 \quad (7)$$

Combining (5), (6) and (7) leads to

$$[1 - \alpha^2(1+\beta)]\left(\frac{V_P}{V_{tp}}\right)^2 - 2[1 - \alpha(1+\beta)]\left(\frac{V_P}{V_{tp}}\right) - \beta = 0 \quad (8)$$

where

$$\alpha = \frac{R_2}{R_1 + R_2} \text{ and } \beta = \frac{(W/L)_1}{(W/L)_2}$$

The mobility dependence in (5) and (6) has been completely cancelled.

Solving for V_P,

$$V_P = \frac{1 - \alpha(1+\beta) + (1-\alpha)\sqrt{1+\beta}}{1 - \alpha^2(1+\beta)} V_{tp} \quad (9)$$

Therefore V_P/V_{tp} is a constant decided by the W/L ratio of MP_1 and MP_2 and the resistance ratio of R_1 and R_2. From (3) the threshold voltage V_{tp} is observed to be a linear function of temperature, so V_P will be strongly linear with temperature. In (7), mobility is cancelled even when the offset voltage of the amplifier A_2 is considered.

3.2 Linear voltage from NMOS device

V_N is generated using NMOS devices MN_1 and MN_2, as shown in block II of Fig. 2. Here, both MN_1 and MN_2 operate in the active region; otherwise, a temperature stable supply must be generated and applied to the gate of MN_2 to force triode region operation. Op amp A_2 forces the source voltage of transistor MN_2 to be the same as the voltage of node Vb. Both MN_1 and MN_2 can operate in the active region by scaling R_3 and R_4. Again, the start-up device MP_4 keeps the circuit away from the zero operating point then is cut off during normal operation. The positive feedback factor is small since the W/L ratio of MN_1 is smaller than that of MN_2; therefore, the resistive negative feedback dominates.

The currents through MN_1 and MN_2, I_1 and I_2, are given by the following equations

$$I_1 = \frac{1}{2}\mu_n C_{ox}\left(\frac{W}{L}\right)_1 (V_b - V_{tn0})^2 \quad (10)$$

$$I_2 = \frac{1}{2}\mu_n C_{ox}\left(\frac{W}{L}\right)_2 (V_N - V_b - V_{tn})^2 \quad (11)$$

where V_{tn} is threshold voltage of NMOS device and V_{tn0} is the threshold voltage with zero V_{SB}.

Assuming no offset for the op amp

$$V_b = \frac{R_4}{R_3 + R_4} V_N$$

Devices MN_1 and MN_2 have the same drain currents, that is,

$$I_1 = I_2 \quad (12)$$

Solving (10), (11) and (12) for V_N gives

$$V_N = \frac{V_{tn} - \sqrt{\beta} V_{tn0}}{1 - \alpha(1 + \sqrt{\beta})} \quad (13)$$

where

$$\alpha = \frac{R_4}{R_3 + R_4} \text{ and } \beta = \frac{(W/L)_1}{(W/L)_2}$$

Again, the mobility dependence in (10) and (11) has been cancelled and V_N is proportional to threshold voltage. Device MN_2 has a non-zero source to bulk voltage V_{SB} that changes with temperature. However, the resulting nonlinear effect in its threshold voltage V_{tn} due to body effect is typically small. So V_N is expected to be linear with temperature. Again, in (12), even when the offset voltage of the amplifier A_2 is considered, mobility and hence its nonlinear dependence on temperature, is still cancelled.

3.3 The subtractor

Both V_P and V_N have negative TCs but they are different in value. Hence, they can be subtracted with different weighting values to form a near zero-TC output voltage. A third op amp A_3 is used to form a subtractor circuit as shown in block III of Fig. 2.

The transfer function of this subtractor is

$$V_{REF} = \left(1 + \frac{R_7}{R_5} + \frac{R_7}{R_6}\right) V_P - \frac{R_7}{R_5} V_N \quad (14)$$

Equation (14) can be re-written

$$V_{REF} = \frac{R_7}{R_5}\left[\left(1 + \frac{R_5}{R_6} + \frac{R_5}{R_7}\right) V_P - V_N\right] \quad (15)$$

While other op amp based subtractors are more common [16], this circuit provides more freedom in choosing the relative gain values of the two inputs.

In (15), the coefficient

$$\left(1 + \frac{R_5}{R_6} + \frac{R_5}{R_7}\right)$$

is used to cancel the TC difference of voltages V_P and V_N while the coefficient R_7/R_5 is used to set the reference voltage to a desired level.

The two variables R_7/R_6 and R_5/R_6 offer independent control of temperature compensation and reference voltage value. This subtractor overcomes the problem of a fixed reference voltage of 1.25 V in the conventional CMOS bandgap circuit and provides a convenient method of calibration by trimming the resistors R_5 and R_7.

3.4 Operational amplifier offset effects

The effect of the offset voltages of the operational amplifiers are not considered in the previous sections. Considering the offset voltage of operational amplifier A_1,

$$V_a = \frac{R_2}{R_1 + R_2} V_P + V_{os1}$$

Then (8) changes to

$$[1 - \alpha^2(1+\beta)]\left(\frac{V_P}{V_{tp}}\right)^2 - 2\left[1 - \alpha(1+\beta)\left(1 - \frac{V_{os1}}{V_{tp}}\right)\right]\left(\frac{V_P}{V_{tp}}\right) - \beta + (1+\beta)\left[\frac{2V_{os1}}{V_{tp}} - \left(\frac{V_{os1}}{V_{tp}}\right)^2\right] = 0 \quad (16)$$

Since $V_{os1} \ll V_{tp}$, the terms including V_{os1} are very small. The effect of V_{os1} on V_P is therefore very small. Considering the offset voltage of operational amplifier A_2,

$$V_b = \frac{R_4}{R_3 + R_4} V_N + V_{os2}$$

(13) changes to

$$V_N = \frac{V_{tn} - \sqrt{\beta} V_{tn0}}{1 - \alpha(1 + \sqrt{\beta})} + \frac{(1 + \sqrt{\beta}) V_{os2}}{1 - \alpha(1 + \sqrt{\beta})} \quad (17)$$

Again, the offset voltage of amplifier A_2 produces only a small offset in V_N since $V_{os2} \ll V_{tn}$. For the subtractor, (15) changes to

$$V_{REF} = \frac{R_7}{R_5}\left[\left(1 + \frac{R_5}{R_6} + \frac{R_5}{R_7}\right)(V_P + V_{os3}) - V_N\right] \quad (18)$$

Since $V_{os3} \ll V_P$, the offset voltage of amplifier A_3 results in a very small offset for V_P. This offset can be compensated by R_7/R_5.

The overall noise performance of the voltage reference depends on the transistors (MN_1, MN_2, MP_1 and MP_2) and

the operational amplifiers (A_1, A_2 and A_3). Assuming ideal operational amplifiers, the dominant noise source is flicker noise, which is proportional to $I_{ds}^2/W_{eff}L_{eff}$ [17] and can be minimised using big devices for transistors MN_1 and MP_1. The design of a low noise operational amplifier is beyond the scope of this paper.

4 Circuit implementation and trimming

The circuit is implemented in a standard CMOS process. The sizes of the MOSFETs and resistors chosen are tabulated in Table 1. Fairly large resistors are used since high sheet resistance polyresistors are available in this process. The width of transistors MP_1 and MP_2, along with MN_1 and MN_2, are set the same for better matching. Also, relatively large devices are used and they are placed close in the layout to minimise the mismatch of threshold voltages. Good threshold voltage matching is required for the previous analysis of (9) and (13) to hold; however, it is not required to preserve the linearity of V_P and V_N with temperature. The lengths of MP_1 and MN_1 are large to make sure the positive feedback factors are smaller than the negative feedback factors.

Table 1: Table of device sizes

Device	Size	Unit	Device	Size	Unit
MP_1	2.4/3.6	μm/μm	R_3	120	kΩ
MP_2	2.4/24	μm/μm	R_4	60	kΩ
R_1	60	kΩ	R_5	50	kΩ
R_2	120	kΩ	R_6	124.5	kΩ
MN_1	6/3.6	μm/μm	R_7	144.275	kΩ
MN_2	6/54.6	μm/μm			

4.1 Op amps

The op amps A_1 and A_2 of Fig. 2 are implemented using the folded cascode amplifier in Fig. 3. A PMOS source follower is used to give a good output swing range, which is needed at low temperatures. The op amp is designed to have a gain of 95 dB.

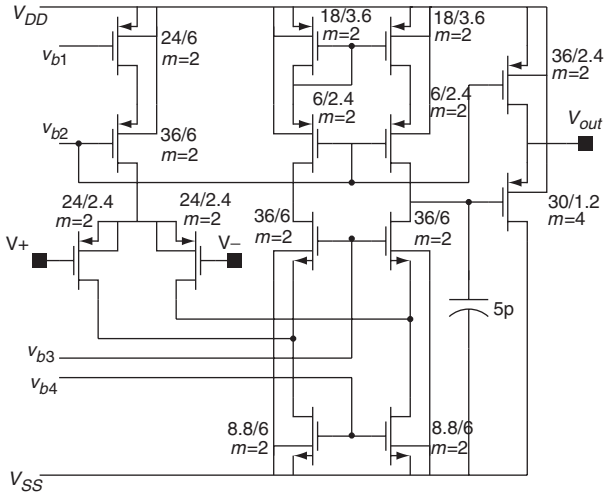

Fig. 3 *Op amp for linear voltages*

4.2 Calibration

The output voltage may deviate from its designed value due to process parameter variation. This voltage variation can be corrected with a simple two-step calibration technique. The first step is to calibrate for zero-TC, that is, decide upon coefficient

$$\left(1+\frac{R_5}{R_6}+\frac{R_5}{R_7}\right)$$

in (15). After V_P and V_N are measured, this coefficient is set to their TC ratio so that the resulting TC is minimised. The second step is to calibrate for voltage level. With values of V_P and V_N, the coefficient R_7/R_5 is set to amplify the resultant voltage to the desired level. These two coefficients can be adjusted independently by trimming resistors R_5 and R_7 in Fig. 2.

5 Experimental results

The proposed design was fabricated in a 0.5 μm n-well double-poly triple-metal CMOS process. The threshold voltages of NMOS and PMOS devices are 0.70 V and 0.98 V at room temperature, respectively. The reference circuit occupies 360 μm × 260 μm. With a power supply of 5.0 V, the circuitry generating V_P and V_N dissipates 0.97 mW at room temperature (power consumption of the subtractor circuit is not included). Figure 4 shows a micrograph of the test chip.

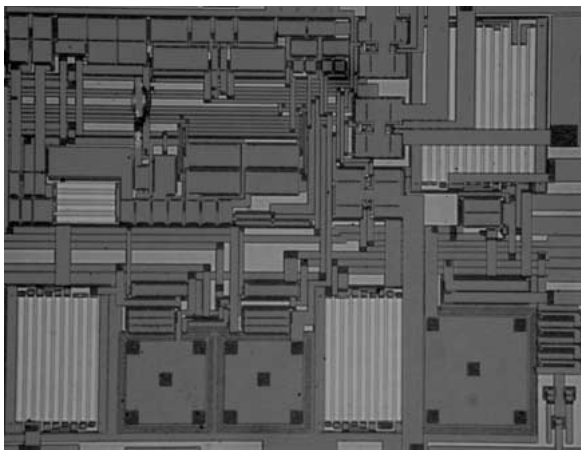

Fig. 4 *Micrograph of voltage reference*

Experimental results demonstrated that the on-chip compensation was insufficient; therefore, an external RC network (a 470 Ω resistor and 0.1 μF capacitor in series) is connected to the output pins of both op amps A_1 and A_2 to stabilise them. These networks eliminate oscillation but do not affect the operation of the voltage reference. The offset voltages of op amps A_1 and A_2 cause V_P and V_N to deviate from their designed values. The offset voltage is compensated by trimming the size of either the inverting input or noninverting input device.

The two linear voltages are measured at different temperatures with power supply voltage at 4.5 V, 5.0 V and 5.5 V. V_P is shown in Fig. 5 along with simulated results, while Fig. 6 shows measured and simulated results for V_N.

The variation due to power supply is small (not visible in Figs. 5 and 6), only about 2.6 mV for V_P and 1.8 mV for V_N. Both voltages are linear functions of temperature. There is a relatively large deviation from the simulation results; this is largely due to the threshold voltage variation of this process. With better process control and characterisation of MOSFET threshold voltage, simulation can be better correlated to measured results.

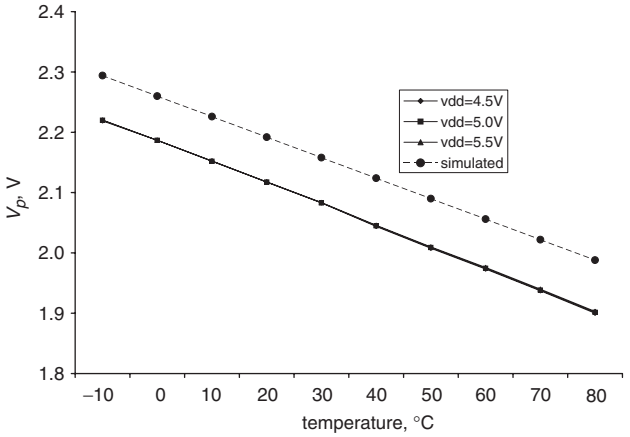

Fig. 5 *Simulated and measured results for linear voltage V_P*

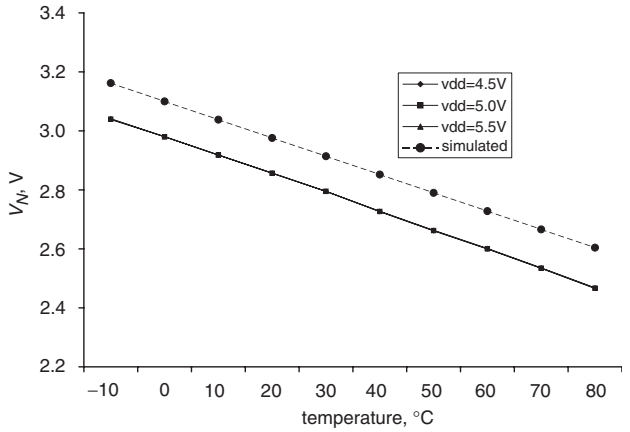

Fig. 6 *Simulated and measured results for linear voltage V_N*

The measured data for voltages V_P and V_N are used to select the coefficients of the subtractor as described in the previous subsection. An external op amp is used as the subtractor, but on-chip resistors are used. Temperature coefficient calibration is performed by adding a series resistor of 380 Ω to R_5. The reference voltage level is not calibrated in this case, for simplicity.

The measured reference voltage as a function of temperature and supply voltage is shown in Fig. 7, demonstrating a TC of 32 ppm °C. The variation due to power supply is less than 10 mV (or just 1%) when the supply voltage changes from 4.5 V to 5.5 V. The equivalent power supply rejection is 10 mV/V.

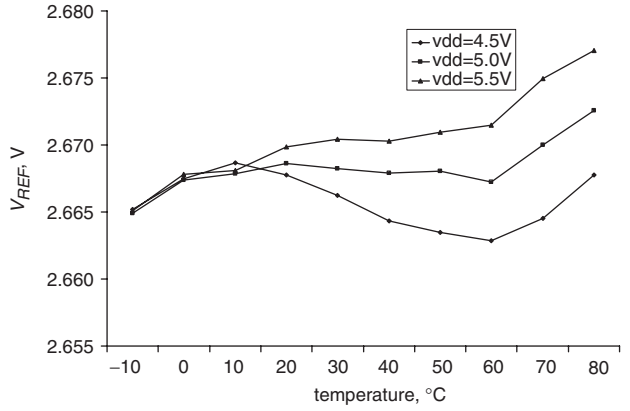

Fig. 7 *Measured reference voltages*

6 Conclusions

A CMOS voltage reference that uses the temperature dependence of PMOS and NMOS threshold voltages is described. Unlike conventional voltage reference design, no diodes or bipolar transistors are used. Instead, two negative TC voltages, both independent of power supply voltage and MOSFET mobility, are generated from MOSFET threshold voltages and then subtracted to generate a low TC voltage reference. The MOSFET threshold voltage has demonstrated good linearity with temperature. This voltage reference can be designed for a wide range of voltage levels.

A test chip is fabricated using a standard 0.5 μm CMOS process and experimental data is presented to verify the theoretical analysis. Performance is comparable to present day bandgap references. The proposed circuit is suitable for on-chip voltage reference generation in precision analogue/digital mixed systems built in a standard CMOS process.

7 Acknowledgments

The authors would like to thank AMI Semiconductor, Utah Research and Design Center for the fabrication service and the Signaling Research Group, Intel Labs, Hillsboro, OR, USA for research support of this project.

8 References

1 Wildar, R.J.: 'A simple three-terminal IC bandgap reference', *IEEE J. Solid-State Circuits*, 1971, **9**, (6), pp. 2–7
2 Brokaw, A.P.: 'New developments in IC voltage regulators', *IEEE J. Solid-State Circuits*, 1974, **6**, (6), pp. 388–392
3 Sanduleanu, M.A.T., Tuijl, A.J.M., and Wassenaar, R.F.: 'Accurate low power bandgap voltage reference in 0.5um CMOS technology', *Electron. Lett.*, 1998, **34**, (10), pp. 1025–1026
4 Banba, H., Shiga, H., Umezawa, A., Miyaba, T., Tanzawa, T., Atsumi, S., and Sakui, K.: 'A CMOS bandgap reference circuit with sub-1-V operation', *IEEE J. Solid-State Circuits*, 1999, **34**, (5), pp. 670–673
5 Tsividis, Y.P., and Ulmer, R.W.: 'A CMOS voltage reference', *IEEE J. Solid-State Circuits*, 1978, **13**, (6), pp. 774–778
6 Vittoz, E.A., and Neyroud, O.: 'A low-voltage CMOS bandgap reference', *IEEE J. Solid-State Circuits*, 1979, **14**, (3), pp. 573–577
7 Song, B.S., and Gray, P.R.: 'A precision curvature-compensated CMOS bandgap reference', *IEEE J. Solid-State Circuits*, 1983, **18**, (6), pp. 634–643
8 Leung, K.N., and Mok, P.K.T.: 'A sub-1-V 15-ppm/CMOS bandgap voltage reference without requiring low threshold voltage device', *IEEE J. Solid-State Circuits*, 2002, **37**, (4), pp. 526–530
9 Meijer, G.C., Schmale, P.C., and Zalinge, K.V.: 'A new curvature-corrected bandgap reference', *IEEE J. Solid-State Circuits*, 1982, **17**, pp. 1139–1143
10 Lee, I., Kim, G., and Kim, W.: 'Exponential curvature-compensated BiCMOS bandgap references', *IEEE J. Solid-State Circuits*, 1994, **29**, (11), pp. 1396–1403
11 Harrison, W.T., Connelly, J.A., and Stair, R.: 'An improved current-mode CMOS voltage reference'. Proc. 2001 Southwest Symposium, SSMSD, 2001, pp. 23–27
12 Filanovsky, I.M., and Allam, A.: 'Mutual compensation of mobility and threshold voltage temperature effects with applications in CMOS circuits', *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, 2001, **48**, (7), pp. 876–884
13 Liu, W.: 'MOSFET models for SPICE simulation' (John Wiley & Sons, Inc., New York, USA, 2001)
14 Johns, D.A., and Martin, K.: 'Analog integrated circuit design' (John Wiley & Sons, Inc., New York, USA, 1996)
15 Lee, C.H., and Park, H.J.: 'All-CMOS temperature independent current reference', *Electron. Lett.*, 1996, **32**, (14), pp. 1280–1281
16 Sedra, A.S., and Smith, K.C.: 'Microelectronic circuits' (Oxford University Press, New York, USA, 1997, 4th edn.)
17 Cheng, Y., and Hu, C.: 'MOSFET modeling & BSIM3 user's guide' (Kluwer Academic Publishers, Boston, USA, 1999)
18 Enz, C.C., and Temes, G.C.: 'Circuit techniques for reducing the effects of op amp imperfections', *Proc. IEEE*, 1996, **84**, (11), pp. 1584–1614
19 Giustolisi, G., Palumbo, G., Criscione, M., and Cutri, F.: 'A low-voltage low-power voltage reference based on subthreshold MOSFET', *IEEE J. Solid-State Circuits*, 2003, **38**, (1), pp. 151–154