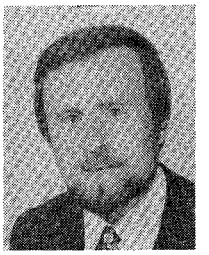
[5] E. Bächle *et al.*, "Monolithic crosspoint arrays for private automatic branch exchanges (PABX)," *IEEE J. Solid-State Circuits*, vol. SC-12, pp. 407-415, Aug. 1977.
[6] A. Weissberger *et al.*, "Processor family specializes in dedicated control," *Electron.*, vol. 8, pp. 84-89, July 1976.
[7] J. M. Herman *et al.*, "Second generation I^2L/MTL: A 20 ns process/structure," *IEEE J. Solid-State Circuits*, vol. SC-12, pp. 93-101, Apr. 1977.
[8] O. Jakits, "Integrierte Injektionslogik, ein neuartiges Prinzip für Digitalschaltungen," *Valvo Berichte*, vol. 18, pp. 215-226, 1974.
[9] J. Saltich *et al.*, "Processing technology and AC/DC characteristics of linear compatible I^2L," *IEEE J. Solid-State Circuits*, vol. SC-11, pp. 478-485, Aug. 1976.



G. Bergmann was born in Wiesenthal/Neisse, Germany, on November 18, 1943. He received the Dipl.-Ing. and the Dr.-Ing. degrees in electronic engineering from the University of Stuttgart, Stuttgart, Germany, in 1970 and 1975, respectively.

From 1970 to 1975 he was employed as Scientific Assistant to Professor Dosse, Head of the Department of Semiconductor Devices, working on the dynamic behavior of junction field-effect transistors. In 1975 he joined the AEG-Telefunken Research Laboratories, Ulm, Germany, where he is working in the field of large-scale integrated circuits, especially in circuit design.

A Low-Voltage CMOS Bandgap Reference

ERIC A. VITTOZ, MEMBER, IEEE, AND OLIVIER NEYROUD

***Abstract*—The CMOS bandgap voltage reference described here uses the bipolar substrate-transistor, and the bipolar-like source-to-drain transfer characteristics of MOS transistors in weak inversion to implement a voltage source that is proportional to absolute temperature (PTAT). A first version of PTAT source is derived from a circuit described previously. A second version is based on a novel cell that can be stacked to obtain the desired voltage. Both versions operate down to 1.3 V with a current drain below 1 μA. A stability of 3 mV over 100°C has been obtained with a few nonadjusted samples.**

Experimental results suggest some possible improvements to extend this stability to every circuit.

I. Introduction

THE FAST PROGRESS in MOS LSI and VLSI technologies allows the implementation of increasingly sophisticated systems on a single chip. Although these systems are mainly based on digital circuits, they usually need some analog circuitry to perform functions difficult or impossible to realize by digital means. For economic reasons, these analog circuits must be realizable without tighter control or additional steps in the process.

During the last few years, a series of new solutions were proposed to meet these requirements, mainly in the areas of A/D conversion and filtering [1]. Recently, a voltage reference that uses the difference between gate-source voltages in an enhancement/depletion process was described [2].

Bandgap voltage references [3] are widely used in bipolar technology where they reach very high standards of accuracy and stability [4], [5]. Their principle relies on the fact that the voltage V_J across a p-n junction that is forward biased by a constant current increases fairly linearly with decreasing temperature, and tends toward the bandgap value V_{G0} when the absolute temperature is extrapolated to 0. If a voltage that is proportional to absolute temperature (PTAT) is added to V_J to exactly compensate the difference between V_J and V_{G0}, one obtains a total voltage that is independent of temperature.

The realization of a bandgap reference in MOS technology has long been hindered by the difficulty of realizing a PTAT voltage. Solutions to this problem may be found by considering the behavior of a MOS transistor in weak inversion.

For a negligible density N_{SS} of fast surface states, the drain current in weak inversion may be expressed as [6]

$$I_D = SI_{D0}\, e^{V_G/nU_T} \left(e^{-V_S/U_T} - e^{-V_D/U_T}\right) \tag{1}$$

where S is the effective geometrical shape factor of the transistor; V_G, V_S, and V_D are the gate, source, and drain potentials with respect to the substrate, and $U_T = kT/q$. The slope factor n is fairly controllable, whereas the characteristic current I_{D0} is very sensitive to process and temperature.

This relation is valid below the weak inversion limit, which may be expressed as [6]

$$I_D \ll \beta U_T^2 \tag{2}$$

where β is the well-known strong inversion transfer parameter.

A PTAT voltage source based on the exponential $I_D - V_G$ characteristics included in (1) has been used in a CMOS bandgap reference described recently [7]: a voltage proportional to nU_T was extracted as the voltage difference across two stacks of three diode-connected transistors ($V_S = 0$, $V_D = V_G$) having

Manuscript received October 16, 1978; revised December 20, 1978.
The authors are with the Centre Electronique Horloger S.A., 2000 Neuchâtel, Switzerland.

0018-9200/79/0600-0573$00.75

different values of I_D/S. A slightly parabolic temperature dependence was observed, mainly due to the nonnegligible variation of the slope factor n with temperature. Furthermore, the need to sum up three gate-to-source voltages is not compatible with low-voltage operation.

It is interesting now to compare the MOS transistor in weak inversion with a bipolar transistor having high-value direct and inverse gains. By using the simple fundamental Ebers-Moll model assuming ideal junctions, the collector current may be expressed as

$$I_C = I_S\,(e^{V_{BE}/U_T} - e^{V_{BC}/U_T}). \tag{3}$$

Comparison of (3) and (1) shows that a MOS transistor in weak inversion is equivalent to a bipolar transistor with the additional possibility of adjusting the characteristic current I_S by means of the gate voltage V_G.

This paper will describe two experimental versions of a CMOS bandgap reference based on this bipolar-like I_D (V_S, V_D) characteristic of the MOS transistor in weak inversion. Both versions have the advantage of being independent of the slope factor n and may operate at a supply voltage as low as 1.3 V.

II. Principle of the Reference

Fig. 1 shows the block diagram of the CMOS voltage reference supplying an output voltage V_R. The diode voltage V_J is obtained across the base-emitter junction of a bipolar substrate-transistor T_S which is available in any CMOS technology by using the n-type substrate as the collector, a p-type well as the base, and an n^+ diffusion as the emitter. This transistor is biased by a constant current source I_E which does not have to be more accurate than about ±10 percent. The current reference may thus be an n-channel transistor with gate-source voltage equal to V_R.

The PTAT source supplies a voltage V_2, which is compared to

$$V_1 = \frac{R_2}{R_1 + R_2} \cdot V_R - V_J \tag{4}$$

by a differential amplifier A. The output S of this differential amplifier drives the gate of the p-channel regulating transistor T_R. The loop is at equilibrium at

$$V_R = \left(1 + \frac{R_1}{R_2}\right)(V_J + V_2) \tag{5}$$

which is independent of temperature, according to the bandgap reference scheme, if

$$V_J + V_2 = V_{G0} \tag{6}$$

at any temperature.

The reference voltage V_R may be adjusted to any value above V_{G0} by properly choosing the ratio R_1/R_2. The resistors can be implemented by means of p-well strips, and binary-weighted pads may be provided for fine adjustment.

If transistor T_R is omitted so that V_R equals V_{CC}, the circuit operates as a voltage level detector: the state of S changes when V_{CC} crosses the value V_R given by (5).

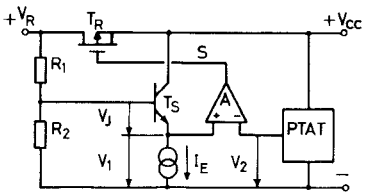


Fig. 1. Block digaram of the bandgap voltage reference.

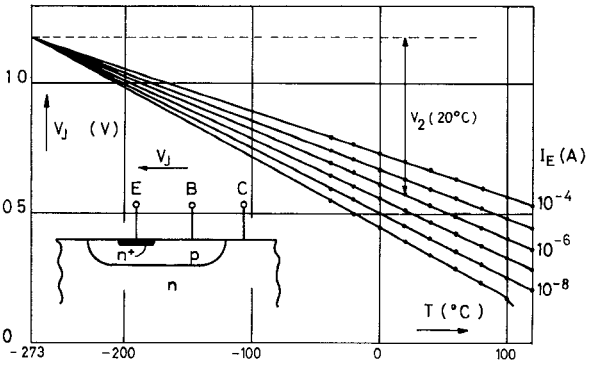


Fig. 2. Base-emitter voltage of the substrate-transistor. Emitter area is 12 × 12 μm^2 on mask. Experimental points are represented by dots.

Measurements of the base-emitter voltage variation with temperature on a typical bipolar substrate-transistor obtained in a low-voltage CMOS process are shown in Fig. 2. As for any good transistor, the curves for various emitter currents converge linearly towards the extrapolated gap value of about 1.2 V. Statistical measurements at 20°C and 1 μA emitter current yield a standard deviation of less than 2 mV. Values of the gain are of the order of 100; the base current is thus negligible.

From these measurments, a value of the required V_2 for any current and temperature can be obtained. An example at 1 μA and 20°C is V_2 = 670 mV.

III. Realization of the PTAT Source, Version 1

Fig. 3 shows the circuit diagram of a first version of the PTAT source, which is derived from a circuit used with bipolar transistors [8]. It is a closed loop made up of two current-mirrors T_1-T_3 and T_4-T_2, with a loop gain larger than 1. If transistors T_1 and T_3 are in the same p-well and operate in weak inversion, application of (1) yields [6]

$$V_0 = U_T \ln \frac{S_3}{S_1} \cdot \frac{S_2}{S_4} + \Delta V_0 \tag{7}$$

where an offset voltage ΔV_0 has been introduced to account for the mismatch of both pairs T_1-T_3 and T_4-T_2. Measurements on pairs of similar transistors show that, at low current, the main part of this statistical mismatch appears as a small differential gate voltage ΔV_G, which is independent of the common values of V_S, V_D, and I_D applied to both transistors. Furthermore, ΔV_G is found to be fairly independent of temperature.

ΔV_0 can therefore be calculated by adding ΔV_{Gn} to the gate voltage of the n-channel transistor T_3 and ΔV_{Gp} to that of the p-channel transistor T_4. If all transistors operate in weak

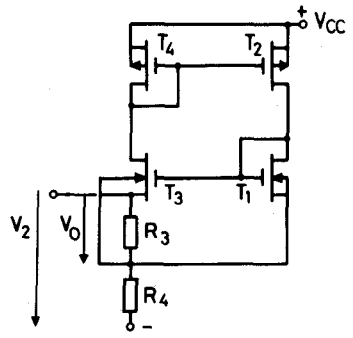


Fig. 3. Circuit diagram of the PTAT source, version 1. Transistors T_1 and T_3 must operate in weak inversion.

inversion, application of (1) yields

$$\Delta V_0 = \frac{\Delta V_{Gn}}{n_n} - \frac{\Delta V_{Gp}}{n_p} \tag{8}$$

where n_n and n_p are the slope factors of the n-channel and p-channel transistors, respectively. Since ΔV_{Gn} and ΔV_{Gp} are not statistically correlated, the variance of ΔV_0 is the sum of those of the right-hand terms of (8).

The minimum value of V_{CC} is limited by the necessity to maintain transistors T_2 and T_3 in saturation. For equal values of p- and n-channel threshold voltage, application of (6) yields the following sufficient condition:

$$V_{CC} > V_{G0} - V_J + 3U_T + |V_{G4}| \tag{9}$$

where $3U_T$ is the minimum drain-source voltage ensuring saturation in weak inversion [6]. If T_2 and T_4 operate in weak inversion, $|V_{G4}|$ is slightly smaller than the threshold voltage $|V_{Tp}|$ of p-channel transistors. Correct operation for V_{CC} as low as 1.2 V is thus ensured if $|V_{Tp}| \leqslant 0.4$ V.

On the other hand, if T_2 and T_4 operate in strong inversion, the expression for ΔV_0 becomes

$$\Delta V_0 = \frac{\Delta V_{Gn}}{n_n} - \Delta V_{Gp} \frac{2U_T}{|V_{G4} - V_{Tp}|}. \tag{10}$$

The effect of the mismatch in the pair T_2-T_4 may thus be reduced at the expense of an increase in the minimum value of V_{CC}.

The maximum value V_0 practically obtainable is of the order of 100 mV. Therefore, it must be multiplied up to the necessary value V_2 by means of resistors R_3 and R_4. The ratio V_2/V_0 is fairly insensitive to any mismatch in the pair T_4-T_2 if the current in the right-hand branch of the circuit is made very small by choosing $S_2 \ll S_4$.

An experimental circuit has been implemented in low-voltage CMOS technology with the following values of W/L (in micrometers on the mask):

T_1	18/8	T_2	40/8
T_3	7200/8	T_4	800/8.

The loop gain is thus 20 and the current flowing through R_4 differs by only 5 percent from that flowing through R_3. Channel length is the same for both transistors of each pair in order to eliminate the uncertainty caused by lateral diffusion. This leads to a very large T_3, which is 400 times wider than T_1.

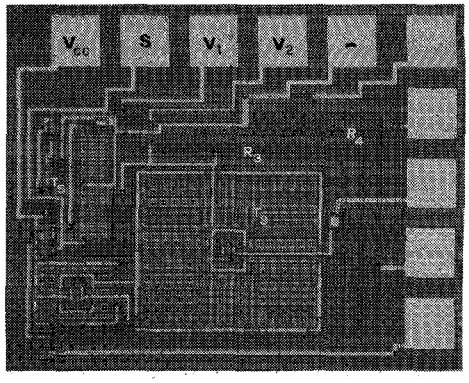


Fig. 4. Photomicrograph of the first version of the reference. The active area of 0.43 mm^2 is mainly occupied by transistor T_3 and resistor R_4.

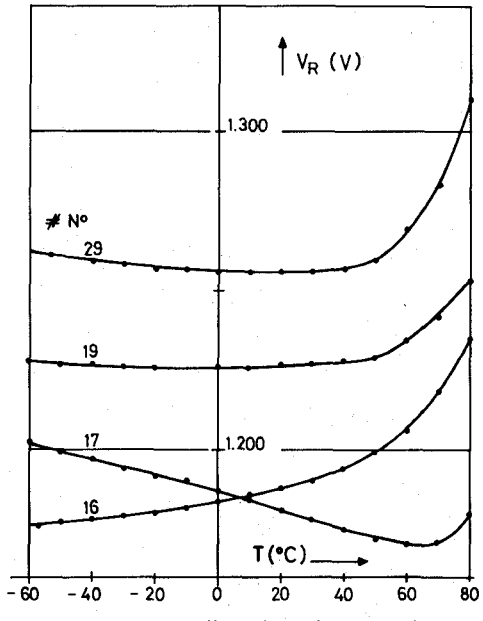


Fig. 5. Measurements on unadjusted random samples. Number 19 exhibits a variation of 3 mV from -60° to +50°C.

Resistances R_3 and R_4 are implemented as p-well strips covered by n^+ diffusions; nominal values are 230 kΩ and 1.75 MΩ. All transistors operate in weak inversion in order to allow the smallest possible value of V_{CC}.

Fig. 4 is a photograph of the circuit that includes all elements of Fig. 1, except the regulating transistor T_R and the resistive divider R_1/R_2. It operates, therefore, as a voltage-level detector with output S switching when V_{CC} reaches the value V_R centered on V_{G0}. Chip area is 0.43 mm^2, mainly due to transistor T_3 and resistor R_4.

IV. Experimental Results with Version 1

The critical value V_R of V_{CC} that corresponds to a change of state of output S has been measured as a function of temperature, without making any adjustments. The results for a random sample of four circuits among 40 are shown in Fig. 5.

It can be noticed that V_R varies fairly linearly up to 40°C, although the slope changes from sample to sample. The fast increase above 40°C is due to an increasing importance of leakage currents. There is also a spread of V_R at a given temperature, which results from the offset voltage ΔV_0

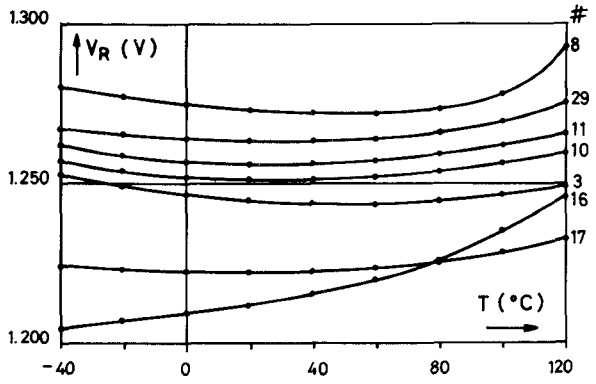


Fig. 6. Measurements on random samples with V_2/V_0 adjusted to 7.8 by low-value external resistors (R_4 = 220 kΩ).

multiplied by V_2/V_0; the offset of the differential amplifier is comparatively negligible. This spread has a standard deviation of the order of 50 mV (measured over 50 circuits); it could be cancelled by introducing an adjustable divider R_1-R_2. The best sample, number 19, exhibits a variation of 3 mV in the range -60° to +50°C, which corresponds to 22 ppm/°C. The total current drain is about 0.7 μA at 20°C.

The variation of slope from sample to sample is due to the spread of the resistance ratio R_4/R_3. This can be demonstrated by fixing the voltage ratio V_2/V_0 to the same value for all samples, by means of low-value external resistors. Results are shown in Fig. 6 for seven random samples, including some of Fig. 5. It can be seen that the spread in temperature coefficients is widely reduced. Another interesting point is that although the currents are now increased slightly beyond the low-inversion limit ($I_{D1} \cong 2\,\beta_1\,U_T^2$), the PTAT source still operates correctly, giving just a slight increase in the central value of V_R. The importance of leakage currents, however, is reduced and the circuit operates at higher temperatures. Most circuits now exhibit a variation of less than 8 mV in the range -50° to +100°C, which corresponds to 45 ppm/°C.

V. Realization of the PTAT Source, Version 2

The basic cell of a novel PTAT voltage source is shown in Fig. 7. The two transistors T_a and T_b are in the same p-well, which may be connected to the source of T_a or to any more negative potential. If both transistors operate in weak inversion, (1) may be applied, which yields

$$V_0 = U_T \ln\left(1 + \frac{S_b}{S_a}\right) + \Delta V_0. \tag{11}$$

This result is independent of the current I flowing through the transistors, as long as this current is smaller than the weak inversion limit and much larger than the leakage currents. Furthermore, the common gate potential must exceed that of the intermediate point by 3 or 4 U_T in order to maintain T_b in saturation. Again, an offset voltage ΔV_0 is added because of the mismatch of the two transistors. By adding ΔV_{Gn} to the gate voltage of transistor T_b, ΔV_0 is found to be equal to $\Delta V_{Gn}/n_n$ for $S_b \gg S_a$.

Experimental results on this cell with $S_b/S_a = 10$ are reported in Fig. 8 for various values of current I. The dimensions of transistor T_a correspond to $\beta_a U_T^2$ = 20 nA.

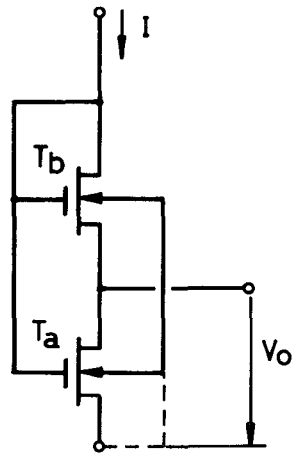


Fig. 7. Elementary PTAT voltage source, version 2. Both transistors must operate in weak inversion.

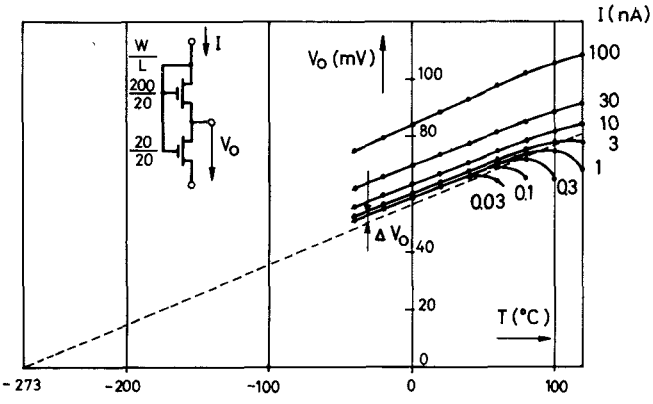


Fig. 8. Typical measurements on the PTAT cell, version 2. Theoretical value of V_0 for perfectly matched transistors is represented by a dotted line.

Up to about 40°C, measurements for currents ranging from 30 pA to 1 nA all coincide on the same straight line; this line is fairly parallel to the theoretical behavior with an offset ΔV_0 of about 3 mV. Due to the effect of leakage currents, the curves deviate from this straight line at a temperature that increases with I.

If the current is increased above 1 nA, condition (2) for weak inversion ceases to be fulfilled. Measurements still line up on parallel straight lines, but these straight lines are progressively shifted upwards. The slope only starts to change significantly when T_a fully operates in strong inversion, at about 100 nA.

Values of V_0 practically obtainable with this cell are limited to about 100 mV. Higher values may be obtained by stacking a certain number of cells as shown in Fig. 9. The mode of operation is still the same, except that the current I supplied to each cell goes through the bottom transistor T_a of all subsequent cells. This must be taken into account in the calculation of V_2.

The advantage of stacking a number of cells, instead of multiplying the voltage V_0 of a single cell up to V_2, relies on the fact that the offset voltages ΔV_0 of the various cells are not statistically correlated. The standard deviation of ΔV_0 is therefore multiplied by $(V_2/V_0)^{1/2}$ and the spread of V_R is reduced. As for version 1, the effect of any mismatch in p-channel transistors may be made negligible by operating these

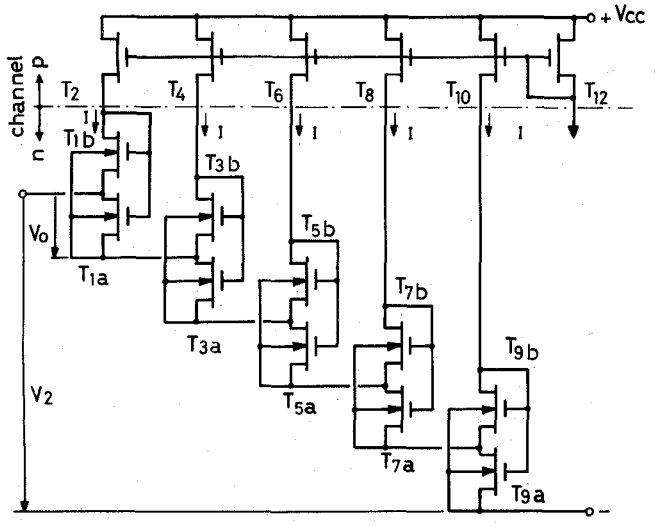


Fig. 9. Stack of elementary PTAT cells.

transistors in strong inversion. The minimum value of V_{CC} is limited by the necessity to maintain the voltage across T_2 above the saturation value V_{sat}, which yields

$$V_{CC} > V_{G0} - V_J + V_{sat} + V_{Tn} \tag{12}$$

where V_{Tn} is the threshold voltage of n-channel transistors.

It would be possible to have all cells in the same p-well connected to the most negative potential, but substrate modulation would increase V_{Tn} and, therefore, the minimum value of V_{CC}.

Preliminary results have been obtained on a first voltage reference based on this circuit. The spread of V_R has a standard deviation reduced to 35 mV, and the peak-to-peak variation of the temperature coefficient is 110 ppm/°C, without any adjustment. However, the average value of the temperature coefficient is not centered on 0, and the influence of leakage currents start to appear at room temperature.

VI. Conclusion

Two versions of a bandgap reference based on the well-controlled I_D (V_S, V_D) characteristics of MOS transistors in weak inversion have been designed and tested. Both can operate with a supply voltage as low as 1.3 V with a current drain below 1 μA. Variations of 3 mV over more than 100°C have been obtained on a few samples of the first version and are probably within reach for each circuit with one adjustment at a fixed temperature. The rather poor results obtained with the second version may be attributed to the nonoptimum design. There seems to be no basic reason why this scheme should be inferior to the first one, and it has the advantage of containing no resistive dividers in the PTAT and has a reduced spread at fixed temperature. The temperature range of both versions may be extended by increasing the current levels somewhat above the weak inversion limit (without really entering strong inversion, where control on the temperature coefficient would be lost), and by optimizing the shape and size of those of the transistors that are sensitive to leakage currents. Although the spread in reference voltages at a fixed temperature might be somewhat reduced by optimizing the design, it does not seem possible to avoid any adjustment to reach an accuracy of a few millivolts. Furthermore, long-term stability is still to be proved. But it must be recalled that the purpose is not to compete with the best bipolar references currently available, but to build the best possible reference on a CMOS chip with no additional steps in the process.

References

[1] D. A. Hodges, P. R. Gray, and R. W. Brodersen, "Potential of MOS technologies for analog integrated circuits," *IEEE J. Solid-State Circuits*, vol. SC-13, pp. 285–294, June 1978.

[2] R. A. Blauschild, P. Tucci, R. S. Muller, and R. G. Meyer, "An NMOS voltage reference," in *ISSCC Dig. Tech. Papers*, 1978, pp. 50–51.

[3] R. J. Widlar, "New developments in IC voltage regulators," *IEEE J. Solid-State Circuits*, vol. SC-6, pp. 2–7, Feb. 1971.

[4] A. P. Brokaw, "A simple three-terminal IC bandgap reference," *IEEE J. Solid-State Circuits*, vol. SC-9, pp. 388–393, Dec. 1974.

[5] G. C. Meijer and J. B. Verhoeff, "An integrated bandgap reference," *IEEE J. Solid-State Circuits*, vol. SC-11, pp. 403–406, June 1976.

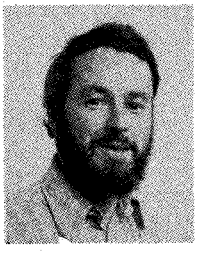
[6] E. Vittoz and J. Fellrath, "CMOS analog integrated circuits based on weak inversion operation," *IEEE J. Solid-State Circuits*, vol. SC-12, pp. 224–231, June 1977.

[7] Y. P. Tsividis and R. W. Ulmer, "A CMOS reference voltage source," in *ISSCC Dig. Tech. Papers*, 1978, pp. 48–49.

[8] M. P. Timko, "A two-terminal IC temperature transducer," *IEEE J. Solid-State Circuits*, vol. SC-11, pp. 784–788, Dec. 1976.

Eric A. Vittoz (A'63–M'72) was born in Lausanne, Switzerland, on May 9, 1938. He received the M.S. and Ph.D. degrees in electrical engineering from the Federal Institute of Technology, Lausanne, in 1961 and 1969, respectively.

After spending one year as a Research Assistant, he joined the Centre Electronique Horloger S.A., Neuchâtel, Switzerland, in 1962, where he became involved in micropower integrated-circuit developments for the watch, while preparing a thesis in the same field. Since 1971, he has been Associate Director of this Laboratory, supervising advanced electronic watch developments. His actual field of research is in very low-power CMOS integrated analog circuits. He also lectures and supervises student work in integrated-circuit design at the Federal Institute of Technology (EPF-L), Lausanne.



Olivier Neyroud was born in Chardonne, Switzerland, on March 24, 1951. He received the M.S. degree in electrical engineering from the Federal Institute of Technology, Lausanne, Switzerland, in 1976.

He then joined the Centre Electronique Horloger S.A., Neuchâtel, Switzerland, and is involved with micropower linear IC and watch systems developments.