A CMOS Bandgap Reference Without Resistors

Arne E. Buck, Member, IEEE, Charles L. McDonald, Member, IEEE, Stephen H. Lewis, Fellow, IEEE, and T. R. Viswanathan, Fellow, IEEE

Abstract—This paper describes a bandgap reference fabricated in a 0.5-μm digital CMOS technology without resistors. The circuit uses ratioed transistors biased in strong inversion together with the inverse-function technique to produce a temperature-insensitive gain applied to the proportional to absolute temperature (PTAT) term in the reference. After trimming, the peak-to-peak output voltage change is 9.4 mV from 0 °C to 70 °C. It occupies 0.4 mm² and dissipates 1.4 mW from a 3.7-V supply.

Index Terms—Analog circuits, analog integrated circuits, CMOS analog integrated circuits, reference circuits, temperature.

I. INTRODUCTION

BANDGAP references add the forward bias voltage across a p–n diode with a voltage that is proportional to absolute temperature (PTAT) to produce an output that is insensitive to changes in temperature [1], [2]. The relative weighting of the voltages added is usually adjusted by trimming the ratio of two resistors. In standard digital CMOS technologies, models for the resistors may not be available or reliable. Also, in digital technologies, the area of such resistors is increased because silicide is often used to reduce the sheet resistance of the polysilicon and diffusion layers. As a result, the length and area of the required resistors is increased, increasing not only the cost, but also the susceptibility of the reference operation to substrate noise coupling. One way to overcome this problem is to use an extra mask to selectively block the silicide, but this mask also increases the cost. In some but not all technologies, the silicide block mask is required in electrostatic-discharge (ESD) protection circuits. This paper presents a circuit solution to the above problems: a bandgap reference without resistors [3]. This solution eliminates the need for resistor models and may eliminate the need for a silicide block mask when this mask is not required in ESD circuits. Also, the bandgap reference described here uses only MOS transistors biased in saturation or cutoff. The devices biased in saturation operate in strong inversion, for which accurate device models are usually available, simplifying the design process, especially in digital CMOS technologies.

This paper is organized as follows. Section II describes the concept, and Section III describes the circuits. The test results are presented in Section IV, and the conclusion is given in Section V.

II. CONCEPT

To produce a temperature-insensitive output, the bandgap reference applies a temperature-independent gain M of about 3–6 to the difference between the forward bias voltages across two diodes ΔV_D. Since resistors are not used, the required gain is obtained by using ratioed transistors together with the inverse function technique [4]. The main idea in the inverse function technique is to apply a pair of functions f and \( f^{-1} \) to ΔV_D so that \( f^{-1}(f(\Delta V_D)) = M(\Delta V_D) \). In circuit terms, f and \( f^{-1} \) might be a transconductance (possibly nonlinear) that maps ΔV_D to some current i. Then \( f^{-1} \), which is a scaled version of \( f^{-1} \), would be a transresistance that cancels the nonlinearity in i and provides the required gain M. Another approach used in this work is to choose \( f^{-1} \) so that \( f^{-1}(M(\Delta V_D)) = M(\Delta V_D) \). In this equation, the current i = f(ΔV_D) is multiplied by M in the current domain by a current mirror using transistors scaled 1 : M. The scaled current is converted to the output voltage by a properly selected transresistance \( f^{-1} \), which requires less gain but a wider dynamic range than \( f^{-1} \). The inverse-function approach works with any smooth nonlinearity, and knowledge of the exact functions is not required as long as it is possible to deduce the proper scaling.

III. CIRCUITS

Fig. 1 shows the core of the bandgap reference. The PTAT voltage \( \Delta V_D = V_{D2} - V_{D1} \) is applied across the differential pair \( M_3 - M_4 \), which acts as a transconductance f. The resulting current is multiplied by G using current mirror \( M_7 - M_5 \) and is delivered to differential pair \( M_1 - M_2 \), which operates as a...
transresistance $f_T^{-1}$ because of the negative feedback around $M_1$.

The voltage $V_{D2}$ falls to its minimum of about 0.6 V at the maximum temperature. If $V_{OUT} \approx 1.2$ V is maintained at the gate of $M_1$, the differential input to $M_1-M_2$ is up to about 0.6 V. To minimize the required $V_{GS}$, $V_t$ to operate both $M_1$ and $M_2$ in saturation, an intentional mismatch is introduced wherein the aspect ratios of $M_1$ and $M_2$ are larger than those of $M_3$ and $M_3$, respectively, by a factor $B$. To provide a qualitative understanding of the circuit behavior, the circuit can be analyzed using a simple square-law MOS model. With the transistor aspect ratios given in Fig. 1, the output voltage equation is derived from two gate–source loops around $M_1-M_2$ and $M_3-M_4$:

$$V_{OUT} = V_{D2} + \sqrt{AG} (V_{D2} - V_{D1}) = V_{D2} + \sqrt{AG} \Delta V_D. \quad (1)$$

In traditional bandgap references, a resistance value would be trimmed to adjust $\sqrt{AG}$ in this equation. Here, the parameter $A$ is the ratio of the size of the $M_3-M_4$ and $M_2-M_3$ differential pairs, and $G$ is the current-mirror gain from $M_7$ to $M_2$ as well as the ratio of the tail currents of the two differential pairs. On the prototype, $G$ is held constant while $A$ is trimmed by adjusting the aspect ratios of $M_3$ and $M_4$ under digital control. These transistors are each laid out as an array of devices with binary-weighted widths and equal lengths. These arrays can be adjusted digitally by means of transmission gates that connect the gates of individual array elements to the appropriate diode or to $V_{DD}$. Fig. 2 shows that $M_3$ is composed of five devices, labeled $M_{3a}$ to $M_{3e}$. Transistor $M_4$ uses a similar structure but with seven devices $M_{4a}$ to $M_{4g}$ in its array.

Fig. 3 shows the complete circuit. To improve the supply insensitivity, it uses a self-bias configuration that is a modification of a previously published circuit [5]. The current that flows is that which forces $\Delta V_D = V_{D2} - V_{D1} = V_{GS72} - V_{GS71}$. Most of the current in $D_2$ is provided by $M_{27}$, allowing the current in $M_{72}$ to be small enough that $V_{GS71} \approx V_t$. Therefore, $V_{GS72} - V_t \approx \Delta V_D$, and an equation for the drain current in $M_{72}$ is

$$I_{72} \approx \frac{k'}{2} \left( \frac{W}{L} \right) \gamma_2 (\Delta V_D)^2. \quad (2)$$

Transistors $M_{81}-M_{85}$ form a start-up circuit. It consists of a string of diode-connected transistors $M_{81}-M_{84}$. If the core of the reference is off when the power supply is applied to the circuit, $M_{85}$ turns on and pulls up on the gate of $M_{72}$, turning on the core. When the core turns on, the voltage from the gate of $M_{72}$ to ground rises high enough to turn off $M_{85}$, disconnecting the start-up circuit from the core.

Table I shows the device sizes. Transistor $M_3$ in Fig. 3 represents the components of the array of transistors $M_{3a}$ to $M_{3e}$ in Fig. 2 that are connected to diode $D_1$. The width of transistor $M_3$ can vary from 5 to 155 $\mu$m, and the nominal width is 80 $\mu$m. Similarly, $M_4$ represents the components of a similar array of transistors $M_{4a}$ to $M_{4g}$ that are connected to diode $D_2$. The width of transistor $M_4$ can vary from 5 to 635 $\mu$m, and the nominal width is 320 $\mu$m.

Diodes $D_1$ and $D_2$ are actually each formed with a PMOS transistor. For each of these transistors, the gate,
drain, and source are connected together. The combined
drain–gate–source node of $D_1$ is connected to the gate of $M_3$;
the combined drain–gate–source node of $D_2$ is connected
to the gate of $M_4$. The body of each of these PMOS transistors
is connected to ground. This approach was selected because
MOSIS provided models for these junctions but not for sub-
strate p-n-p transistors. However, these models may not be
accurate in part because BSIM3 models the source–drain diode
behavior adequately only when the diodes are reverse biased
[6]. To compensate for the possibility that the diode models
are inaccurate with only one fabrication cycle, the trim range
of the prototype was intentionally designed to be much larger
than simulations showed to be necessary. In practice, trimming
may not be required for low precision applications ($\pm 5\%$) once
the diodes are thoroughly characterized and the nominal gain is
adjusted accordingly.

IV. EXPERIMENTAL RESULTS

The circuit was fabricated in a MOSIS 0.5-$\mu$m n-well CMOS
process. Fig. 4 shows a die photograph. The area is 0.4 mm$^2$
without pads. Twenty-five parts were fabricated and tested at
25 $^\circ$C before trimming. The average output voltage is 1.1219 V,
and the standard deviation is 8.8 mV. Fig. 5 shows a plot of
the output voltage versus temperature for the trimmed refer-
ence with the lowest output variation from 0 $^\circ$C to 70 $^\circ$C. The
peak-to-peak variation is 9.4 mV. Fig. 6 shows a similar plot
in which four references are each trimmed to set the output at
45 $^\circ$C equal to the same target voltage (1.1195 V $\pm$ 0.5 mV).
The difference between the maximum voltage exhibited by any
of these references and the minimum exhibited by any of these
references between 0 $^\circ$C and 70 $^\circ$C is 11.4 mV peak to peak.

Fig. 7 shows a plot of supply rejection versus frequency. The
supply rejection is 45.1 dB at 10 Hz. The power dissipation is
1.4 mW from a 3.7-V supply. Table II summarizes the perfor-
mance.

V. CONCLUSION

This paper shows a bandgap reference circuit without resis-
tors that is compatible with a pure digital CMOS technology.

ACKNOWLEDGMENT

The authors would like to thank A. P. Brokaw and
G. Pietrobon at Analog Devices for their help with the noise
measurement.

REFERENCES

bandgap reference without resistors,” in IEEE Int. Solid-State Circuits
[6] W. Liu, MOSFET Models for SPICE Simulation, Including BSIM3v3 and