Behzad Razavi

The Bandgap Reference

Since its inception in the late 1960s, the bandgap circuit has served as an essential component in most integrated circuits. This simple, robust idea provides a temperature-independent (TI) voltage and a proportional-to-absolute-temperature (PTAT) current. In this article, we study the principles of bandgap circuit design.

A Brief History

Semiconductor technology does not directly offer any electric quantity that is nominally independent of the ambient temperature. Thus, temperature independence has generally been envisioned in the form of combining two phenomena that have opposite temperature coefficients (TCs). For example, a resistor with a low TC can be constructed by placing in series, with proper weighting factors, a positive-TC resistor and a negative-TC resistor.

Applying this idea to voltage quantities proved more difficult. It had been long recognized that the voltage across a forward-biased diode has a negative TC (if its bias current does not change much with temperature). However, a positive-TC voltage was missing. In 1964, Hilbiber of Fairchild Semiconductor observed that two diode stacks biased at different current densities can provide a TI voltage [1]. In 1965, Widlar, from the same company, more explicitly showed that the base-emitter voltages of two transistors biased at different current densities had a PTAT difference [2] and in 1971 introduced the first bandgap circuit (Figure 1) [3]. This was followed by another topology presented by Brokaw in 1974 (Figure 2) [4] and many others later.

The rise of CMOS technology in the 1970s posed the question of whether a stable voltage reference could be created without the use of bipolar devices [5]. However, it was observed that the high-threshold mismatch of MOS transistors leads to significant error and drift in such references. Subsequent work therefore focused on the "native" bipolar transistor available in standard CMOS processes [6], [7]. Figure 3 shows an example similar to Brokaw's cell [8], except that the current-measuring resistors are moved from the collectors to the emitters because the former must be tied to V_{DD}. (Early CMOS technologies used an *n*-substrate and hence accommodated a vertical npn transistor.) A useful feature of this topology is that the op amp does not drive resistors and can thus maintain a high loop gain.

A useful feature of this topology is that the op amp does not drive resistors and can thus maintain a high loop gain.

The Basic Idea

As mentioned in the previous section, the voltage across a *pn* junction and hence the base-emitter voltage, V_{BE}, of a bipolar transistor exhibit a negative TC. It can be shown that, for a constant collector current,

$$\frac{\partial V_{BE}}{\partial T} = \frac{V_{BE} - (4+m)V_T - E_g/q}{T}, \quad (1)$$

where T is the absolute temperature, $m \approx -3/2$, $V_T = kT/q$, and E_g the bandgap

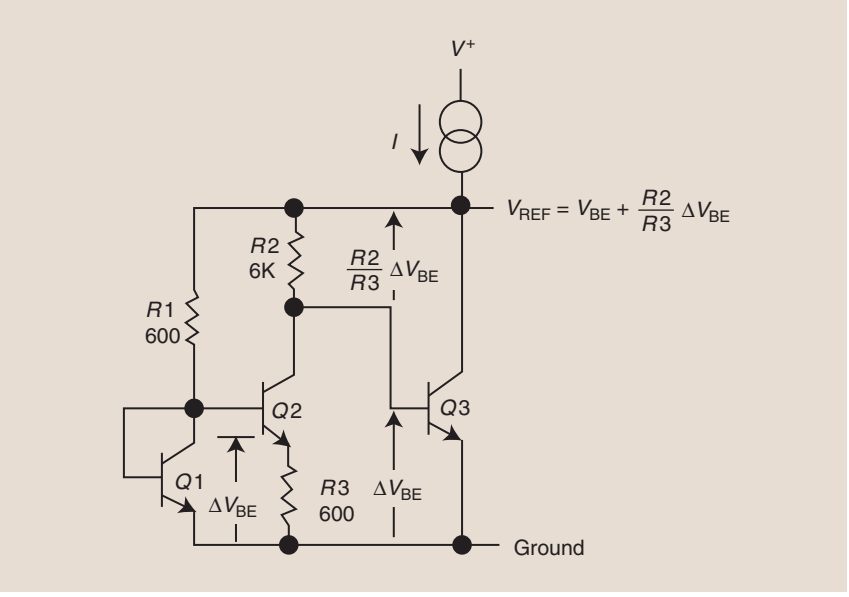

FIGURE 1: Widlar's bandgap circuit.

Digital Object Identifier 10.1109/MSSC.2016.2577978
Date of publication: 2 September 2016

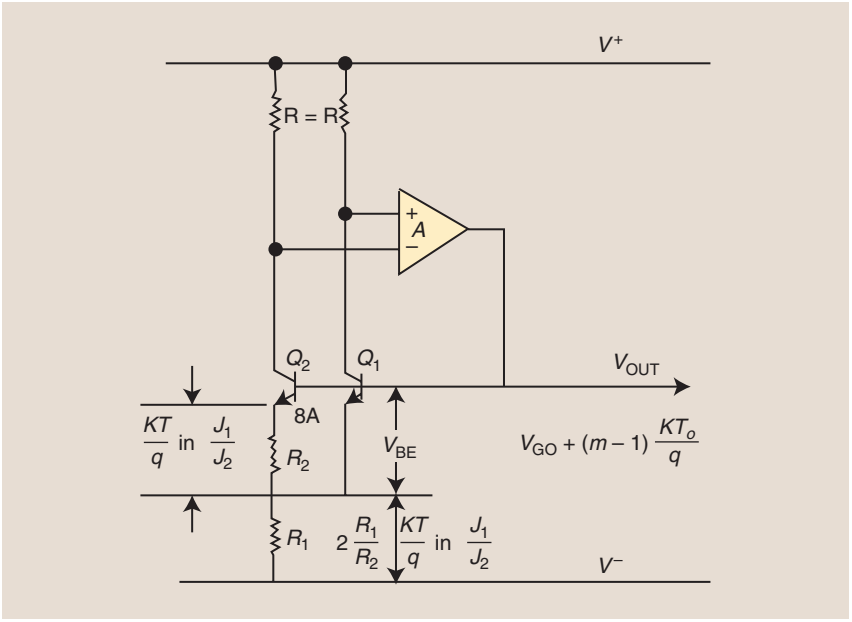

FIGURE 2: Brokaw's bandgap circuit.

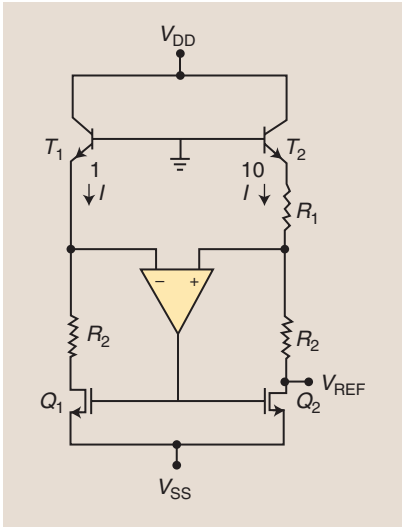

FIGURE 3: The CMOS bandgap circuit proposed by Gregorian et al.

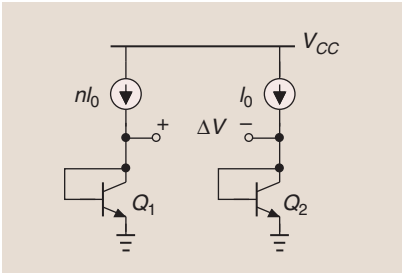

FIGURE 4: The generation of a PTAT voltage.

energy. With typical current densities, $V_{BE} = V_T \ln(I_C/I_S) \approx 750\,\text{mV}$, yielding a TC of about –1.5 mV/K at room temperature.

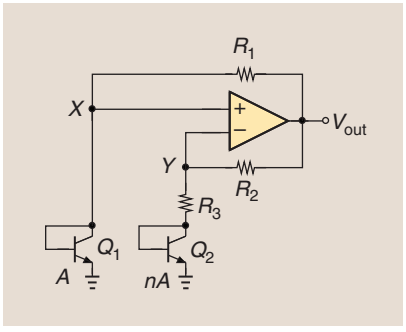

FIGURE 5: The bipolar bandgap circuit proposed by Kujik.

We wish to create a voltage with a TC equal to +1.5 mV/K. As shown in Figure 4, we bias two identical diode-connected bipolar transistors at different current levels, obtaining different current densities and

$$\Delta V = V_{BE1} - V_{BE2} \tag{2}$$

$$= V_T \ln \frac{nI_0}{I_S} - V_T \ln \frac{I_0}{I_S} \tag{3}$$

$$= V_T \ln n. \tag{4}$$

The same result holds if the bias currents are equal but $I_{S2} = nI_{S1}$, i.e., if Q_2 consists of n parallel units, each identical to Q_1. We postulate that

$$V_{REF} = V_{BE} + V_T \ln n \tag{5}$$

can have a zero TC if n is chosen properly. That is, we must calculate the necessary value of n and find a circuit that adds V_{BE} and $V_T \ln n$. The temperature coefficient of ΔV is equal to $+(k/q)\ln n = +0.087 \ln n$ mV/K. To reach +1.5 mV/K, we require $\ln n \approx 17.2$ and, therefore, $n \approx 2.95 \times 10^7$, an impractically large scaling factor for the bias currents or the transistors.

The term "bandgap reference" can be appreciated by calculating $\partial V_{REF}/\partial T$ from (5), setting it to zero, finding $V_T \ln n$ from the result, and substituting in (5)

$$V_{REF} = \frac{E_g}{q} + (4 + m)\, V_T. \tag{6}$$

This expression suggests that the value of V_{REF} extrapolated to $T = 0\,\text{K}$ is equal to the bandgap voltage, E_g/q.

The problem of $n = 2.95 \times 10^7$ can be mitigated if $V_T \ln n$ is somehow "amplified" before it is added to V_{BE}. A compact implementation of this idea is shown in Figure 5 [9], where V_{BE1} travels through a noninverting amplifier with a gain of $1 + R_2/R_3$ (if $g_{m2}^{-1} \ll R_3$) and V_{BE2} sees a gain of $-R_2/R_3$ (if $g_{m1}^{-1} \ll R_1$). That is,

$$V_{out} = V_{BE1}\left(1 + \frac{R_2}{R_3}\right) - V_{BE2}\frac{R_2}{R_3} \tag{7}$$

$$= V_{BE1} + (V_{BE1} - V_{BE2})\frac{R_2}{R_3}. \tag{8}$$

One can prove that this relation holds even if we do not assume $g_{m2}^{-1} \ll R_3$ and $g_{m1}^{-1} \ll R_1$. We select $R_1 = R_2$ so that Q_1 and Q_2 carry equal currents, and an emitter area ratio of n to one, thereby obtaining $V_{BE1} - V_{BE2} = V_T \ln n$ for the voltage sustained by R_3. It follows that

$$V_{out} = V_{BE1} + \frac{R_2}{R_3} V_T \ln n \tag{9}$$

$$= V_{BE2} + \left(1 + \frac{R_2}{R_3}\right) V_T \ln n. \tag{10}$$

We can now choose $(1 + R_2/R_3)\, V_T \ln n \approx 17.2$ with a moderate value for n, e.g., in the range of 10–20.

A flaw in our results is that the collector current is assumed constant in (1) but it is PTAT in this circuit. Fortunately, this issue is easily remedied by reducing the factor of 17.2 to about 16.

In standard CMOS technologies, the circuit of Figure 5 faces three issues.

First, a CMOS op amp directly driving the feedback resistors must deal with gain-power-stability trade-offs. Second, it is difficult to realize bipolar transistors whose collectors are not grounded; the vertical structure available in typical CMOS processes utilizes the *p*-substrate as the collector. Third, the op amp offset is amplified by a factor of $1 + R_2/R_3$ as it appears in V_{out}, causing error and temperature drift. For example, $n = 10$ translates to $1 + R_2/R_3 \approx 7.5$.

In this article, we study the principles of bandgap circuit design.

To address these issues, we first recognize that resistors R_1 and R_2 in Figure 5 provide equal bias currents, a function that can be realized by controlled current sources. In the spirit of Figure 3, we arrive at the topology depicted in Figure 6(a). If M_1 and M_2 are identical, then $V_{BE1} - V_{BE2} = V_T \ln n$ and $V_{out} = V_{BE2} + (1 + R_2/R_3) V_T \ln n$. The bipolar transistors are implemented as vertical *pnp* structures. As with Figure 3, resistor $R_2' = R_2$ can be inserted to ensure $V_{DS1} = V_{DS2}$, suppressing current mismatch due to channel-length modulation.

The op amp offset is still amplified by a factor of $1 + R_2/R_3$. This issue is ameliorated by choosing a large n, say, 20–30, and/or establishing a nonunity ratio between I_{E1} and I_{E2}. Specifically, if $(W/L)_1 = m(W/L)_2$, the bipolar transistor current densities differ by another factor of m, leading to

$$V_{out} = V_{BE2} + (1 + \frac{R_2}{R_3}) V_T \ln(n \cdot m). \quad (11)$$

For example, with $n = 20$ and $m = 5$, we require $1 + R_2/R_3 \approx 3.7$. This means that the designer must still pay close attention to the op amp's offset.

While simple and robust, the bandgap reference in Figure 6(a) suffers from several undesirable effects: 1) the output tends to contain substantial flicker (and thermal) noise contribution from the op amp and M_1 and M_2, 2) the circuit potentially exhibits poor supply rejection, especially due to that of the op amp, 3) if driving a heavy load capacitance at the output, the circuit can become unstable, especially in cases where A_1 itself contains more than one stage, and 4) if the op amp begins with a high output level at power-up, the two branches may remain off indefinitely. The remedies include the use of large MOS transistors to address the first issue, the choice of an op amp topology for high supply rejection for the second issue, and adding a start-up circuit to deal with the third. As shown in Figure 6(b), a weak additional branch creates a large initial imbalance at the input of the op amp, forcing its output to go low. After the circuit turns on and reaches the desired operating point, M_3 and M_4 turn off.

Low-Voltage Bandgap References

The basic bandgap expression, $V_{BE} + 17.2 V_T$, takes on a value of about 1.25 V at $T = 300$ K, defying direct implementation in today's low-voltage technologies. We describe one low-voltage example and refer the reader to [10] and [11] for others.

For a low-voltage bandgap reference to have a zero TC, it must produce an output of the form $\alpha(V_{BE} + 17.2 V_T)$, where $\alpha < 1$ is a scaling factor. For example, beginning with the branch shown in Figure 7(a) and writing $V_{out} = V_{BE3} + R_4 I_1$, where I_1 is a PTAT current, we observe that $\partial V_{out}/\partial T$ can be zero but V_{out} is still around 1.25 V. How do we create a fraction of this value without first generating such a high voltage? Let us tie a resistor from the output node to ground [Figure 7(b)], surmising that the resulting voltage division creates a zero-TC output with a smaller magnitude. Since $(V_{out} - V_{BE3})/R_4 + V_{out}/R_5 = I_1$, we have

$$V_{out} = \frac{R_5}{R_5 + R_4}(R_4 I_1 + V_{BE3}), \quad (12)$$

concluding that the output can be arbitrarily small and have a zero TC. We now derive the PTAT current, I_1, from the circuit of Figure 6(a) and arrive at the topology shown in Figure 7(c) [12]. With a voltage drop of $V_T \ln n$ across R_3, we have

$$V_{out} = \frac{R_5}{R_5 + R_4}(V_{BE3} + \frac{R_4}{R_3} V_T \ln n), \quad (13)$$

where the MOSFETs are assumed identical.

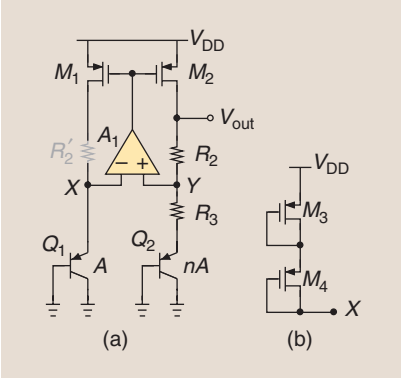

FIGURE 6: (a) A CMOS bandgap circuit and (b) a start-up branch.

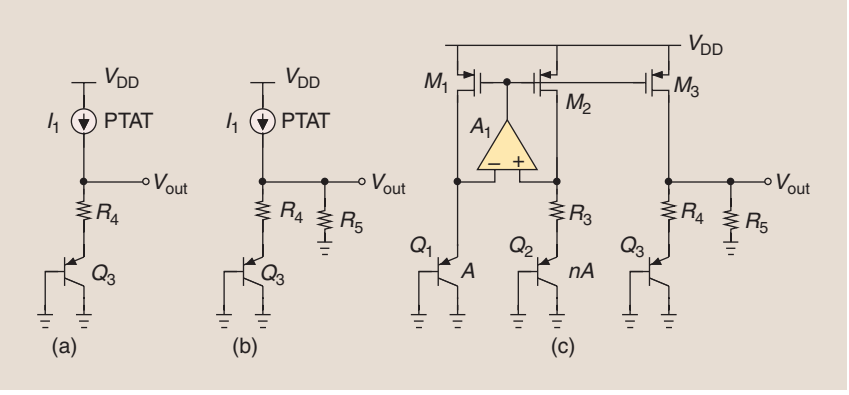

FIGURE 7: (a) A single branch generating a TI voltage, (b) the use of voltage division to lower the output, and (c) a complete low-voltage reference.

This circuit requires a minimum supply voltage equal to $V_{BE1} + |V_{DS1}|$, which amounts to 700–800 mV, provided that the op amp does not pose a higher limit. If technology scaling continues to reduce V_{DD}, we may, in the future, resort to voltage doublers so as to accommodate the unscalable V_{BE}, or exploit the exponential characteristics of MOSFETs in the subthreshold region [6].

Answers to Last Issue's Questions

1) Explain in the time domain why $1 - z^{-1}$ represents a high-pass function.

Suppose the input signal changes slowly during one clock cycle. If subjected to a one-clock-cycle delay, such a signal still resembles itself, yielding a small output for a $1 - z^{-1}$ function. On the other hand, if the input contains faster dynamics, then the result of $1 - z^{-1}$ exhibits greater excursions.

2) Explain why the comparator clock jitter in a discrete-time Σ-Δ modulator is not critical.

Since the digital-to-analog converter signal is given enough time to settle, the comparator clock jitter has no effect on the feedback signal. It only increases or decreases, by a slight amount, the time available for settling.

References

[1] D. Hilbiber, "A new semiconductor voltage standard," in *Proc. Int. Solid-State Circiuts Conf. Dig. Tech.*, Philadelphia, PA, 1964, pp. 32–33.

[2] R. J. Widlar, "Some circuit design techniques for linear integrated circuits," *IEEE Trans. Circuit Theory*, vol. 12, no. 4, pp. 586–590, Dec. 1965.

[3] R. J. Widlar, "New developments in IC voltage regulators," *IEEE J. Solid-State Circuits*, vol. 6, no. 1, pp. 2–7, Feb. 1971.

[4] A. P. Brokaw, "A simple three-terminal IC bandgap reference," *IEEE J. Solid-State Circuits*, vol. 9, no. 6, pp. 388–393, Dec. 1974.

[5] R. A. Blauschild, P. A. Tucci, R. S. Muller, and R. G. Meyer, "A new NMOS temperature-stable voltage reference," *IEEE J. Solid-State Circuits*, vol. 13, no. 9, pp. 767–774, Dec. 1978.

[6] Y. P. Tsividis and R. W. Ulmer, "A CMOS voltage reference," *IEEE J. Solid-State Circuits*, vol. 13, no. 6, pp. 774–778, Dec. 1978.

[7] E. A. Vittoz and O. Neyroud, "A low-voltage CMOS bandgap reference," *IEEE J. Solid-State Circuits*, vol. 14, no. 3, pp. 573–577, June 1979.

[8] R. Gregorian, G. Wegner, and W. E. Nicholson, "An integrated single-chip PCM voice codec with filters," *IEEE J. Solid-State Circuits*, vol. 16, no. 4, pp. 322–333, Aug. 1981.

[9] K. E. Kujik, "A precision reference voltage source," *IEEE J. Solid-State Circuits*, vol. 8, pp. 222–226, June 1973.

[10] H. Banba, H. Shiga, A. Umezawa, and T. Miyaba, "A CMOS bandgap reference circuit with sub-1-V operation," *IEEE J. Solid-State Circuits*, vol. 34, no. 5, pp. 670–674, May 1999.

[11] C. J. B. Fayomi, G. I. Wirt, H. Achigui, and A. Matsuzawa, "Sub-1-V CMOS bandgap reference design techniques: A survey," *Analog Integr. Circuits Signal Process.*, vol. 62, pp. 141–157, Feb. 2010.

[12] H. Neuteboom, B. M. Kup, and M. Janssens, "A DSP-based hearing instrument IC," *IEEE J. Solid-State Circuits*, vol. 32, pp. 1790–1806, Nov. 1997.

SSC

CORRECTION

In the article "Memory Interfaces: Past, Present, and Future" [1], Figure 16 was incorrect due to a production error. The corrected figure is printed below.

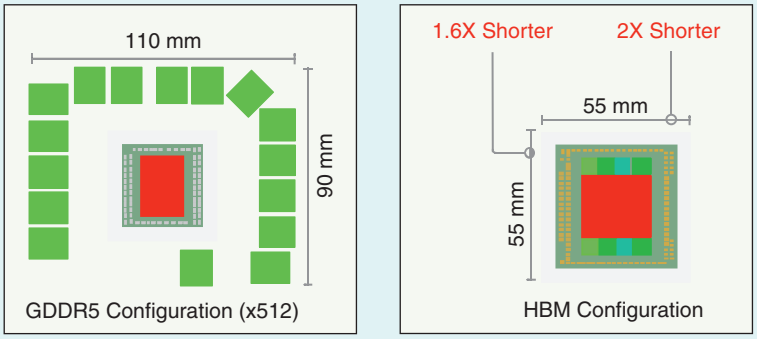

FIGURE 16: Configuration examples for (a) GDDR5x and (b) HBM.

Digital Object Identifier 10.1109/MSSC.2016.2599700
Date of publication: 2 September 2016