A Precision CMOS Voltage Reference Exploiting Silicon Bandgap Narrowing Effect

Bo Wang, *Student Member, IEEE*, Man Kay Law, *Member, IEEE*, and Amine Bermak, *Fellow, IEEE*

***Abstract*— A compact voltage-mode bandgap voltage reference (BGR) is presented. Instead of using overhead circuits, the silicon bandgap narrowing effect is exploited for bipolar junction transistor's (BJT) curvature reduction and residual curvature correction. Prototype measurements in a 0.18-μm standard CMOS process show that the curvature of the BJT is effectively reduced from its inherent 3.6 mV to 1.4 mV. The proposed BGR measures a minimum temperature coefficient of 8.7 ppm/°C from −55 °C to 125 °C by batch trimming one resistor. After a curvature trimming, it further improves to 4.1 ppm/°C. The BGR has a minimum supply voltage of 1.3 V, 4.3 μA nominal current consumption, 0.03%/V line sensitivity, and 2 mV/mA load sensitivity at 25 °C. The output rms noise in the 0.1∼10-Hz band measures 10.23 μV.**

***Index Terms*— Bandgap narrowing (BGN), bipolar junction transistor (BJT) curvature reduction, BJT noise, CMOS bandgap voltage reference (BGR), curvature correction, process spread, temperature coefficient (TC).**

I. Introduction

MANY efforts have been made to design voltage references with ultralow supply for process scaling or nanowatt power for portable electronic devices [1]. However, the achieved accuracy has remained stagnant. For emerging applications, such as distributed sensor networks, accurate on-chip voltage references are essential for signal recognition and data processing, but costly overhead circuits are unaffordable [2].

Among the bandgap voltage reference (BGR) topologies, the bipolar junction transistor (BJT)-based designs can achieve accurate and scalable outputs [3] and are less sensitive to process variations compared with the pure-MOSFET ones [4]. As summarized in [5], many BJT-based designs have been reported, since it was first implemented by Widlar, such as the Kuijk topology or the Brokaw cell, which are all first-order temperature compensated BGRs designed by adding a complementary-to-absolute-temperature voltage $V_{\rm be}(T)$ to a scaled proportional-to-absolute-temperature (PTAT) voltage V_T. The achieved accuracy was, however, limited to a few tens of ppm/°C to hundreds of ppm/°C mainly because of the inherent curvature in $V_{\rm be}(T)$.

More accurate voltage references can be obtained with multiple temperature trimming [6], but are not favored because of their cost. Alternative methods to improve the BGR precision are curvature minimization or curvature correction using circuit techniques. Song and Gray [7] mentioned $V_{\rm be}(T)$'s curvature can be minimized by biasing a BJT's collector with a strongly temperature-dependent current. Nevertheless, producing such a current with a large temperature coefficient (TC), for example, using translinear circuits [8] not only increases power and chip area but is not always feasible in CMOS process [9, p. 81]. Curvature correction using the temperature dependence of the BJT's base current was utilized in [10] to achieve high precision. However, the small curvature signal in the BJT base current mandates a large bias current to generate the required correction signal, making it power hungry. In [2], a second-order TC compensation using the opposite TC characteristics of a MOS and a BJT is proposed. However, the relatively high power consumption and the large MOS process spread make this approach unattractive in many ultralow power high-accuracy applications. Introducing nonlinearity into the PTAT voltage $V_{\rm pt}(T)$ can also correct $V_{\rm be}(T)$'s curvature after voltage summation. However, such a method relies on accurate auxiliary circuits or device trimming to guarantee that the introduced nonlinear term matches well with that of $V_{\rm be}(T)$ [11]. The piecewise-linear technique has a greater flexibility in terms of curvature correction [12]. However, it still requires overhead circuits which consume extra power as in [2] and [11] to produce the nonlinear signals.

To avoid dissipating power driving the curvature correction circuits, this paper presents a BGR that only exploits the temperature characteristics of the BJT itself to reduce the $V_{\rm be}(T)$ curvature and to perform residual curvature correction. As a result, a precise and low-power voltage reference can be designed. The remainder of this paper is structured as follows. Section II presents the principle of curvature reduction and residual curvature correction. The circuit implementation is elaborated in Section III. Error sources are identified and discussed in Section IV, while Section V shows the experimental results before drawing a conclusion in Section VI.

II. Nonlinearity Minimization

The temperature characteristic of the base–emitter voltage $V_{\rm be}(T)$ of BJTs has been well modeled in [13] and

Manuscript received March 6, 2015; revised May 4, 2015; accepted May 14, 2015. Date of publication June 4, 2015; date of current version June 17, 2015. This work was supported in part by the Hong Kong Innovation and Technology Fund under Grant ITS/195/14FP and in part by the Research Committee of University of Macau under Grant MYRG2015-00140-AMSV. The review of this paper was arranged by Editor G. Niu.

B. Wang and A. Bermak are with the Smart Sensory Integrated Systems Laboratory, Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong (e-mail: timepicker@gmail.com; eebermak@ust.hk).

M. K. Law is with the State Key Laboratory of Analog and Mixed-Signal VLSI, Macau University, Macau 999078, China (e-mail: mklaw@umac.mo).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TED.2015.2434495

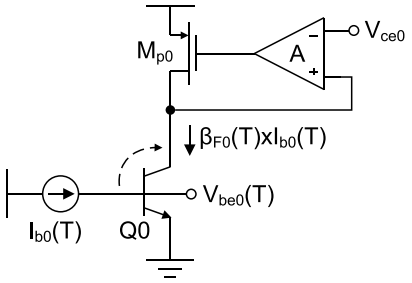

Fig. 1. Base-biased n-p-n bipolar transistor utilizing the temperature dependence of $\beta_{F0}(T)$ (silicon BGN effect) for $V_{\rm be0}(T)$ curvature reduction.

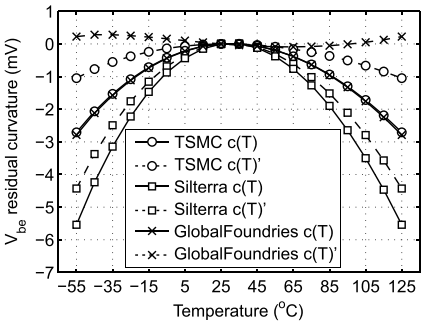

Fig. 2. Simulated $c(T)$ and $c(T)'$ from TSMC, Silterra, and GlobalFoundries 0.18-μm process, respectively.

can be expressed as the sum of a constant term, a proportional to T and a nonlinear term

$$V_{\rm be}(T) = \left[V_{g0} + (\eta - \zeta)\frac{kT_r}{q}\right] - \lambda T + c(T) \tag{1}$$

in which the curvature $c(T)$ is

$$c(T) = (\eta - \zeta)\frac{k}{q}\left(T - T_r - T\ln\frac{T}{T_r}\right) \tag{2}$$

where V_{g0} is the extrapolated bandgap voltage at 0 K, k is the Boltzmann constant, q is the electron charge, η is a process parameter [9, p. 21], T is the absolute temperature in Kelvin, T_r is a reference temperature, λ is the extrapolated $V_{\rm be}(T)$ slope at T_r, and ζ is the temperature exponent of the BJT collector current [i.e., $I_c \propto (T/T_r)^{\zeta}$].

It is implied by (2) that a minimal $|\eta - \zeta|$ is beneficial to yield a smaller $|c(T)|$. However, reducing $|\eta - \zeta|$ with overhead circuits like [8] burns extra power, which motivates this design to investigate the temperature characteristics of BJT itself for compensation.

A. Bandgap Narrowing in BJT

In modern CMOS processes, as the silicon impurity doping concentration becomes particularly high ($>10^{18}$ cm^{-3}), the silicon bandgap narrowing (BGN) occurs [14]. For BJT, due to the BGN induced by the silicon lattice deformation in the heavily doped emitter, its forward current gain becomes strongly temperature dependent [10], which is described by (3) in the Gummel–Poon model

$$\beta_{F0}(T) = \beta_{f0}\left(\frac{T}{T_r}\right)^{X_{\rm TB}} \tag{3}$$

where β_{f0} is its nominal value at T_r and $X_{\rm TB}$ is its temperature exponent which is positive and increases with technology scaling.

B. $c(T)$ Reduction

Instead of directly compensating $c(T)$ using $\beta_{F0}(T)$ as in [10], $c(T)$ is first reduced in this design. By biasing an n-p-n transistor via its base with $I_{b0}(T)$, its collector current $I_{c0}(T)$ would be $\beta_{F0}(T) \times I_{b0}(T)$, as shown in Fig. 1. $X_{\rm TB}$ from $\beta_{F0}(T)$ is then embodied in $I_{c0}(T)$. If $I_{b0}(T)$ is PTAT, ζ in (2) is $1 + X_{\rm TB}$ and (2) is revised to be

$$c(T)' = (\eta - 1 - X_{\rm TB})\frac{k}{q}\left(T - T_r - T\ln\frac{T}{T_r}\right). \tag{4}$$

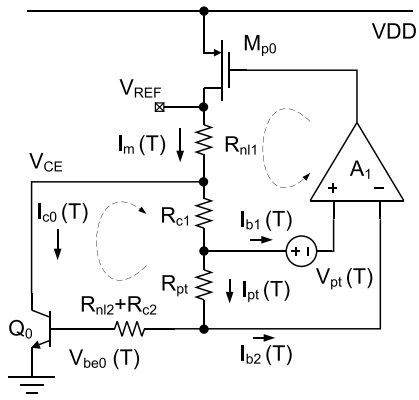

Fig. 3. Proposed BGR topology with curvature reduction and residual curvature correction.

If $0 < X_{\rm TB} < 2(\eta - 1)$, $|c(T)'|$ can be reduced by a fraction of $1 - |1 - X_{\rm TB}/(\eta - 1)|$ when compared with $|c(T)|$. Typically, η is $3 \sim 4$ and the above inequality holds for the most CMOS processes.

Fig. 2 shows the simulated $c(T)$ and $c(T)'$ for BJTs in three different 0.18 μm processes from Taiwan Semiconductor Manufacturing Company (TSMC), Silterra, and GlobalFoundries, in which curvature reduction percentages of 63%, 15%, and 82% are observed, respectively. Notice that the simulated curvature reduction amount does not precisely follow (4) because $c(T)'$ is modeled based on the underlying physics of silicon instead of device measurement. For a process whose $X_{\rm TB}$ is small compared with its η, such as Silterra process, beta multiplication like Darlington pair can be used to produce a larger effective $X'_{\rm TB}$ and the curvature reduction method utilizing BGN effect still applies.

C. BGR Topology

The proposed BGR topology with curvature reduction is shown in Fig. 3. A shunt-feedback loop, consisting of an amplifier A_1 and a pass transistor M_{p0}, regulates the voltage across $R_{\rm pt}$. The input pair of A_1 consists of two matched vertical transistor $Q_{1,2}$, which have an emitter area ratio of $1{:}p$ and have the same collector bias. Therefore, A_1 has a PTAT offset voltage $V_{\rm pt}(T)$. The current generated via $R_{\rm pt}$

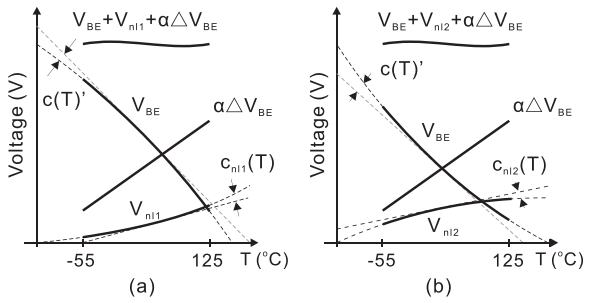

Fig. 4. Relationship between voltage and temperature, showing the curvature and curvature summation. (a) $0 < X_{\rm TB} < \eta - 1$. (b) $\eta - 1 < X_{\rm TB} < 2(\eta - 1)$ (not drawn to scale, and exaggerated).

is [11]

$$I_{\rm pt}(T) = \frac{V_{\rm pt}}{R_{\rm pt}} = \frac{1}{R_{\rm pt}} \cdot \frac{kT}{q} \ln(p). \tag{5}$$

$I_{\rm pt}(T)$ mainly serves as Q_0's base bias and the curvature in $V_{\rm be0}(T)$ becomes $c(T)'$. In Fig. 3, $R_{c1,2}$ are used to amplify $V_{\rm pt}(T)$, thereby compensating the first-order TC of $V_{\rm be0}(T)$.

After $c(T)$ reduction, the residual $c(T)'$ can still be large in a process like Silterra (Fig. 2) and further correction is preferred. In this topology, three $X_{\rm TB}$-related nonlinear signals can be used to linearize $V_{\rm be0}(T)$, including Q_0's collector current I_{c0} and the base currents $I_{b1,2}$ of $Q_{1,2}$, if their collector currents are PTAT. However, for different $X_{\rm TB}$ values, the characteristics of $c(T)'$ can vary and a different correction method is needed, as explained in Section II-D.

D. $c'(T)$ Correction

If $0 < X_{\rm TB} < \eta - 1$, based on (4), the shape of $c(T)'$ is concave and can be corrected by adding (subtracting) a convex (concave) signal that has the same amplitude, as shown in Fig. 4(a). Because $X_{\rm TB}$ is small, $I_{b1,2}$ are not the optimal compensation signals due to their small nonlinear content. In this case, a convex voltage V_{nl1} developed by R_{nl1} is exploited, as shown in Fig. 3, which is

$$V_{nl1}(T) = R_{nl1}[\beta_{F0}(I_{\rm pt} - I_{b2}) + I_{\rm pt} + I_{b1}]. \tag{6}$$

The curvature content in $V_{nl1}(T)$ is obtained by linearizing (6) at T_r

$$c_{nl1}(T) \approx \beta_{f0} R_{nl1} I_{\rm pt}(T_r) \times \left[\left(\frac{T}{T_r} \right)^{X_{\rm TB}+1} - (1 + X_{\rm TB}) \frac{T}{T_r} + X_{\rm TB} \right]. \tag{7}$$

If $\eta - 1 < X_{\rm TB} < 2(\eta - 1)$, $c(T)'$ is convex and can be corrected with a concave signal, as shown in Fig. 4(b). In this case, a voltage V_{nl2} is developed by R_{nl2} with a current $I_{\rm pt} - I_{b2}$ flowing through it, which is highly nonlinear and is expressed as

$$V_{nl2}(T) = R_{nl2}(I_{\rm pt} - I_{b2}) \tag{8}$$

$V_{nl2}(T)$ has a curvature content

$$c_{nl2}(T) = -R_{nl2} \frac{I_c(T_r)}{\beta_{f2}} \times \left[\left(\frac{T}{T_r} \right)^{1-X_{\rm TB}} - (1 - X_{\rm TB}) \frac{T}{T_r} - X_{\rm TB} \right]. \tag{9}$$

As a result, the overall curvature left in $V_{\rm REF}$ is $c_{\rm ref}(T) = c(T)' + c_{nl}(T)$; where $c_{nl}(T)$ is either $c_{nl1}(T)$ or $c_{nl2}(T)$ based on the $X_{\rm TB}$ value of the adopted process. R_{nl1} or R_{nl2} values are then sized to minimize $c_{\rm ref}(T)$ in order to produce a precision $V_{\rm REF}$.

E. Voltage Summation

After $c(T)$ reduction and $c(T)'$ correction, as shown in Fig. 3, $V_{\rm REF}$ is generated by adding the PTAT voltage developed by $R_{c1,2}$ and the nonlinear voltage $V_{nl}(T)$ into $V_{\rm be0}(T)$. The expression for $V_{\rm REF}$ is

$$V_{\rm REF} = V_{\rm be0}(T) + R_{c1} I_{b1} - R_{c2} I_{b2} + (R_{\rm pt} + R_{c1} + R_{c2}) I_{\rm pt}(T) + V_{nl}(T) \tag{10}$$

in which $V_{nl}(T)$ is either $V_{nl1}(T)$ or $V_{nl2}(T)$.

In the topology, the base current of Q_0 is not purely PTAT due to I_{b2} (much smaller than $I_{\rm pt}$), and $V_{\rm be0}$ is revised to be

$$V_{\rm be0}(T) = \frac{kT}{q} \ln \frac{\beta_{F0}(I_{\rm pt} - I_{b2})}{I_{s0}} \approx V_{\rm be0}|_{I_{b0}=I_{\rm pt}} - \frac{R_{\rm pt}}{\ln p} I_{b2} \tag{11}$$

where I_{s0} is the saturation current of Q_0. Combining (10) and (11), $I_{b1,2}$ induced nonlinearities can be minimized if the resistors are sized such that

$$\frac{R_{c2} + \dfrac{R_{\rm pt}}{\ln(p)}}{R_{c1}} = \frac{\beta_{F2}}{\beta_{F1}}. \tag{12}$$

In (12), though $\beta_{F1,2}$ are different due to the different collector current densities of $Q_{1,2}$, the ratio β_{F2}/β_{F1} remains almost constant at different temperatures, collector biasing conditions, and process corners. The first-order TC in (10) can be cancelled by appropriately sizing $R_{c1,2}$.

III. Implementation

A. Circuit Description

The complete BGR circuit is shown in Fig. 5 with the nominal device sizing shown in Table I. $V_{\rm REF}$ is developed at the drain of M_{p0} by resistors $R_{c1,2}$, $R_{\rm pt}$, and R_{nl2}, where R_{nl2} is used to correct the simulated convex shape residual curvature. $Q_{1,2}$ and $M_{p1,2}$ form the error amplifier, with $R_{\rm ss}$ determining its tail current as well as performing emitter degeneration to avoid thermal runaway. A Miller capacitor C_c is used to stabilize the BGR. Because $Q_{1,2}$ have different emitter areas, to maintain $V_{\rm pt}(T)$'s linearity, a p-plus resistor R_b is added to match their base resistances (to a first order). Moreover, symmetrical dummy BJTs $Q_{m1,2}$ are included to match the collector-to-base and collector-to-substrate leakages of $Q_{1,2}$ [15].

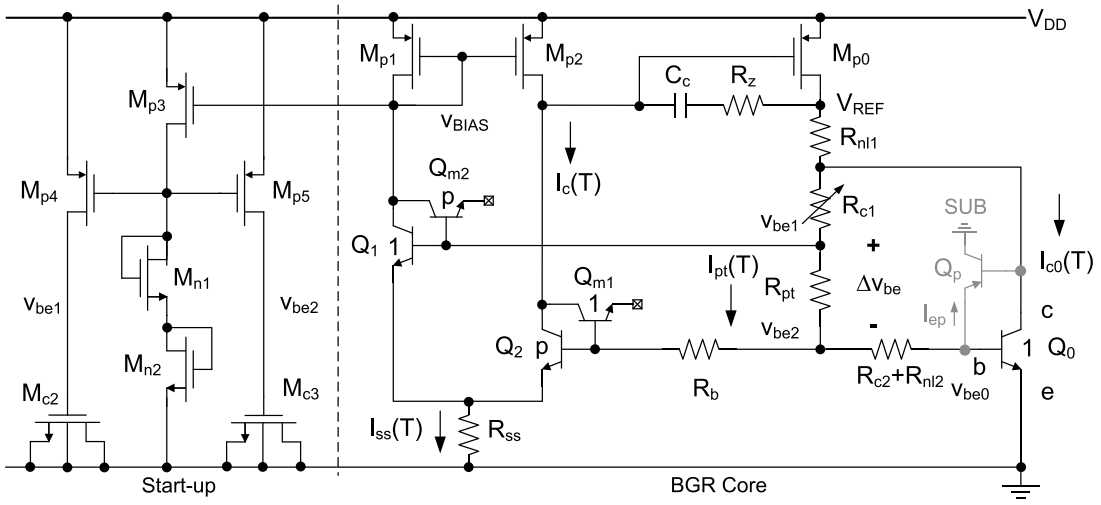

Fig. 5. Proposed BGR utilizing silicon BGN for bipolar transistor's curvature reduction and residual curvature correction.

As this circuit is a self-biasing topology, a start-up circuit consisting of $M_{p3,4,5}$, $M_{n1,2}$, and $M_{c2,3}$ is included. $M_{p4,5}$ are long and high-$V_{\rm th}$ devices for minimal OFF-state leakages. In addition, due to current amplification by Q_1, small $M_{p4,5}$ are preferred to avoid large initial current excitation in the BGR, which requires a long time to recover or even results in start-up failure.

B. Biasing Current

To maintain the validity of (5), the bias current $I_c(T)$ for $Q_{1,2}$ should be much larger than the BJT saturation current I_s. It should, however, be smaller than the bipolar transistor's knee current in case $Q_{1,2}$ enter their high-injection regions, which alters β_{F2}/β_{F1}, thus invalidating (12). In this design, $I_c(T_r)$ is 0.35 μA for stable $\beta_{f1,2}$ ($\pm$0.4% with $\pm$30% $I_c(T_r)$ change) and a minimal current consumption.

Other than $I_c(T)$, the amplitude of $I_{\rm pt}(T)$ determines the overall system power as it will be amplified by β_{f0} (26 at 25 °C). In this design, $I_{\rm pt}(T_r)$ is 0.1 μA considering power consumption and chip area, leaving a simulated 0.47-mV curvature in $V_{\rm be0}(T)$, which can be further corrected by R_{nl2}. The optimized compensation resistor R_{nl2} is 590 kΩ. With a 1:7 emitter area ratio of $Q_{1,2}$, $\beta_{F1}/\beta_{F2} \approx 1.18$, $R_{c1} = 1.8$ MΩ and $R_{c2} = 1.29$ MΩ can cancel the first-order TC in $V_{\rm REF}$. The overall chip area in this design is dominated by the resistor size. In area-limited systems, larger $I_{\rm pt}(T_r)$ can be used for a smaller resistor, at the expense of an increased current consumption.

C. Line Sensitivity

Fig. 6 is the small-signal model of the proposed BGR, where the input impedances $r_{\pi_{1,2}} \propto 1/I_{b1,2}$ of $Q_{1,2}$ are neglected because their steady-state currents $I_{b1,2}$ are in the nanoampere range. The small base resistors of $Q_{0,1,2}$ are also ignored. As this model only focuses on low-frequency analysis, the circuit capacitances are also not shown. At low frequency, output change $v_{\rm ref}$ due to supply noise $v_{\rm dd}$ is

$$\frac{v_{\rm ref}}{v_{\rm dd}} \approx \frac{R_T[1 + g_m r_{\rm op}(1 - A_{\rm dd})]}{g_m r_{\rm op} A_v R_{\rm pt} + \beta_{F0} r_{\rm op} + (1 + r_{\rm op}/r_{Q_0}) R_T} \quad (13)$$

in which $R_T = R_{\rm pt} + R_{c1,2} + R_{nl2} + r_{\pi 0}$; $A_{\rm dd}$ is the supply sensitivity of the amplifier and $A_v \propto V_{\rm AR}/V_T = 500$ is its open-loop gain, where $V_{\rm AR} = 13$ V is the reverse early voltage of an n-p-n bipolar transistor; $r_{\pi 0} = 260$ kΩ is Q_0's small-signal input impedance and $r_{Q_0} \propto V_{\rm AR}/I_{c0}$ is its output impedance; and $g_m = 92.5$ μS and $r_{\rm op} \propto 1/(\lambda I_{M_{p0}}) = 3$ MΩ are the transconductance and output impedance of M_{p0}, respectively. Since $M_{p1,2}$ are diode connected, supply noise directly couples into the feedback loop and modulates the gate voltage of M_{p0} with respect to its source, $A_{\rm dd} \approx 1$ and a 60-dB power supply rejection (PSR) can be achieved at low frequency. Increasing the effective r_{p0} while keeping $A_{\rm dd} \approx 1$ like using a self-cascode current mirror can improve the PSR, but a higher supply is required.

TABLE I
NOMINAL DEVICE PARAMETERS OF THE BGR

Transistor	W/L (μm/μm)	Res	kΩ	Others	
M_{p1}, M_{p2}	4*12.6/5.4	R_{ss}	400	C_c	5.4 pF
M_{p4}, M_{p5}	0.36/1.8	R_b	0.082	R_Z	120 kΩ
M_{c2}, M_{c3}	20/5	R_{pt}	451	p	7
M_{n1}, M_{n2}	0.22/36	R_{c1}	1796	R_{nl1}	0
M_{p3}	2*12.6/5.4	R_{c2}	1290		
M_{p0}	60*25/0.5	R_{nl2}	589.6		

D. Output Noise

The noise model of the BGR is also shown in Fig. 6, in which E_{A_v} and $I_{n1,2}$ are the amplifier's input referred voltage and current noise [3], respectively. E_{p0} is the noise of M_{p0}. I_{n0} is the base-referred noise of Q_0. Since noise from different devices are uncorrelated, the BGR output noise density (Volts^2/Hz) is a sum of the contribution from each noise source, as expressed in (14), as shown

$$\overline{v_{\mathrm{REF},n}^2} = (A_c-1)^2 E_{\mathrm{pt}}^2 + A_c^2\left(E_{A_v}^2 + \frac{E_{p0}^2}{A_v^2}\right) + E_{c1}^2 + A_n^2\left[E_{c2}^2 + E_{nl2}^2 + (R_{c2}+R_{nl2}+r_\pi)^2 I_{n2}^2 + r_\pi^2 I_{n0}^2\right] + R_{c1}^2 I_{n1}^2 \quad (14)$$

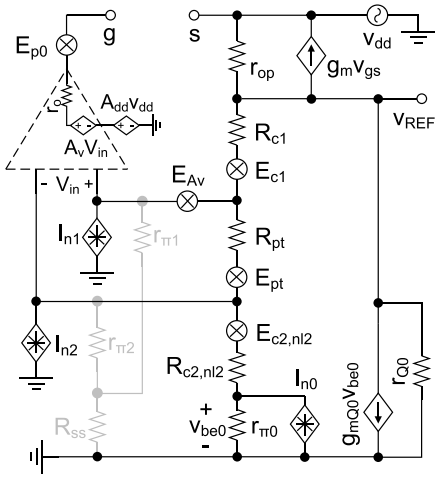

Fig. 6. Small-signal model of the BGR with device noise.

at the top of the page, in which $A_c = R_T/R_{\mathrm{pt}} \approx 10$ is the closed-loop gain of the BGR and $A_n = r_{Q_0}/(r_{Q_0} + R_{c1}) \approx 1/2$ is a noise scaling factor. In (14), $I_{n1,2}$ are mostly shot noise due to the ultrasmall base currents of $Q_{1,2}$ [16, p. 757]. $I_{n0} \propto I_{\mathrm{pt}}^{\alpha}/f$ has a larger flicker content but it is attenuated by A_n. As a result, the low-frequency noise of $\overline{v_{\mathrm{REF},n}^2}$ is dominated by the flicker contents in E_{A_v} and E_{pt} as they are amplified by A_c. The noise spectral density E_{A_v} and E_{pt} can be expressed as [16, pp. 769 and 775], [17]

$$E_{A_v}^2 = \frac{2}{g_{m_{Q1}}^2}\left(2qI_c + \frac{8kT}{3}g_{m_{p1}} + \frac{g_{m_{p1}}^2 K_{f_1}}{WL|_{m_{p1}}C_{ox}}\cdot\frac{1}{f}\right) \quad (15)$$

$$E_{\mathrm{pt}}^2 = 4kTR_{\mathrm{pt}} + V_{\mathrm{pt}}^2\cdot\frac{K_{f_2}}{WL|_{R_{\mathrm{pt}}}}\cdot\frac{1}{f}. \quad (16)$$

For E_{A_v}, its flicker term can be minimized using large $M_{p1,2}$ given the designed $I_c(T_r)$. For the flicker term in E_{pt}, because $V_{\mathrm{pt}} = T\cdot k/q\cdot \ln p$ and $A_c \approx \lambda'/(k/q\cdot\ln p)$, changing V_{pt} cannot reduce $\overline{v_{\mathrm{REF},n}^2}$ (ignoring the change of λ'). E_{pt} can only be reduced using larger resistor dimensions, which is a tradeoff with the chip area.

IV. Device- and Process-Induced Errors

Unlike the PTAT drift, the interdie curvature spread in BGRs cannot be corrected with only one-point trimming. To control the reference precision, device and process nonidealities-induced curvature and curvature spreads need to be identified.

A. BJT Nonideality

1) Parasitic Bipolar Transistor: As shown in Fig. 5, a parasitic p-n-p transistor Q_p is formed by the p^+ base and n^+ collector of Q_0 to the p-type substrate after fabrication. A leakage current I_{ep} exists and the induced error is

$$\Delta V_{I_{\mathrm{eq}}} \approx -\frac{kT}{q}\frac{I_{\mathrm{eq}}(T)}{I_{\mathrm{pt}}(T)}. \quad (17)$$

Based on Ebers–Moll model, the magnitude of I_{eq} is comparable with the saturation current of a p-n-p bipolar transistor. As a result, $\Delta V_{I_{\mathrm{eq}}}$ is negligible given a picoampere-level I_{eq} and a microampere-level I_{pt}.

2) Nonideal V_{be}–I_c Relationship: Considering base-width modulation and carrier recombination of Q_0, the expression for $V_{\mathrm{be0}}(T)$ is modified to be [18]

$$V_{\mathrm{be0}}(T) = n_F\frac{kT}{q}\ln\frac{I_{c0}(T)}{I_{s0}(T)(1+V_{\mathrm{BC}}/V_{\mathrm{AF}} - V_{\mathrm{BE}}/V_{\mathrm{AR}})}$$
$$\approx n_F\left(V_{\mathrm{be0|Ideal}} + \frac{kT}{q}\frac{V_{\mathrm{REF}}}{V_{AF}}\right) \quad (18)$$

where $n_F = 1.0048$ is the forward current emission coefficient and $V_{\mathrm{AF}} = 14.3$ is Q_0's forward early voltage. Since n_F and V_{AF} are temperature invariant, the error term in (18) is PTAT and the linearity of V_{be0} is not impaired. Though n_F has a lot-to-lot spread, it only causes a PTAT drift. Similarly, for $Q_{1,2}$, the nonideal $V_{\mathrm{be}} - I_c$ introduces a multiplicative term n_F into $V_{\mathrm{pt}}(T)$, while the drift caused in V_{REF} is also PTAT [ignoring the small change in $I_c(T)$].

B. Resistor Nonideality

Using R_{pt} with a negative first-order TC α_1 and a positive second-order TC α_2 can increase the temperature exponent ζ of $I_{c0}(T)$. However, the amplitudes of α_1 and α_2 are difficult to control and their variation induced intradie curvature spread cannot be PTAT trimmed. For a resistor with α_1 (ignoring the effects of α_2) in the order of 1×10^{-3} (available in most if not all CMOS processes), a $\pm30\%$ α_1 variation-induced interdie spread is about ±0.2 mV. In this design, among the provided resistor types, the unsalicided poly resistor is used for its smallest TCs with $\alpha_1 = -86.7\times10^{-6}$ and $\alpha_2 = 0.60\times10^{-6}$. The resultant intradie curvature spread with $\pm30\%$ $\alpha_{1,2}$ variations is only ±0.03 mV.

C. Process Spreads

1) Resistor: In this design, if the resistors are matched, (9) and (12) remain unaltered. However, R_{pt} spread (ΔR_{pt}) affects $I_{\mathrm{pt}}(T)$ and thereby alters $V_{\mathrm{be0}}(T)$. ΔR_{pt} induced change in V_{REF} is

$$\Delta V_{\Delta R_{\mathrm{pt}}} = -\frac{kT}{q}\ln\left(1+\frac{\Delta R_{\mathrm{pt}}}{R_{\mathrm{pt}}}\right) \approx -\frac{kT}{q}\frac{\Delta R_{\mathrm{pt}}}{R_{\mathrm{pt}}}. \quad (19)$$

As ΔR_{pt} is temperature invariant [11], $\Delta V_{\Delta R_{\mathrm{pt}}}$ is PTAT and will not introduce an additional curvature in V_{REF}.

2) Bipolar Transistor: The main spreads in BJT are its saturation current (ΔI_{s0}), its forward current gain ($\Delta\beta_f$) and the temperature exponent ($\Delta X_{\rm TB}$). For Q_0, $V_{\rm be0}(T)$ can be affected by its saturation current spread and current gain spread, but the introduced errors are PTAT [9, pp. 29–30]. For a matched $Q_{1,2}$ pair, their spreads do not affect $V_{\rm pt}(T)$ and (12) still holds. However, (7) and (9) are sensitive to $\beta_{f0,2}$ and their variation could cause incomplete or over curvature correction in $c_{\rm ref}(T)$. In this design, as $c_{nl2}(T)$ is used, curvature spread due to $\Delta\beta_{f2}$ is

$$\Delta V_{\beta_{f2}} \approx -\frac{\Delta\beta_{f2}}{\beta_{f2}} c_{nl2}(T) \approx \frac{\Delta\beta_{f2}}{\beta_{f2}} c(T)'. \tag{20}$$

For a ±30% current gain variation, the induced $\Delta V_{\beta_{f2}}$ is about ±0.14 mV, which is nonlinear and is left in $V_{\rm REF}$.

In most CMOS processes, the $X_{\rm TB}$ variation is not modeled. As simulated, a ±30% $X_{\rm TB}$ change causes ∓0.63 mV $c(T)'$ variation. Notice that $X_{\rm TB}$ spread-induced error cannot be corrected with on-chip signals except by trimming $R_{nl1,2}$.

D. Device Mismatch

In Fig. 5, the collector current ratio of $Q_{1,2}$ is subject to the drain current mismatch $\delta I_{\rm ds}$ of $M_{p1,2}$. When referred to the output, $\delta I_{\rm ds}$ induced voltage variation is

$$\Delta V_{I\rm ds} = \delta I_{\rm ds} \frac{kT}{q} \left(\frac{R_{nl2} + R_{c1} + R_{c1}}{R_{\rm pt}} + \frac{1}{\ln p} \right). \tag{21}$$

For a small $I_c(T)$, $M_{p1,2}$ are biased in their weak inversion regions and $\delta I_{\rm ds}$ is temperature dependant [19]. As the temperature dependence model for $\delta I_{\rm ds}$ is unavailable, $\Delta V_{I\rm ds}$ as a whole should be minimized for higher design confidence. Referring to the modeled $1.7\% \cdot \mu{\rm m}^{-1}$ drain current mismatch of interleaved pMOS pair, $M_{p1,2}$ with 12.6 μm × 4/5.4 μm (width/length) are used and $\delta I_{\rm ds} < 0.1\%$ is expected. As a result, $\Delta V_{I\rm ds}$ only varies by ±0.14 mV.

Saturation current mismatch between $Q_{1,2}$ has a similar effect as that of $\delta I_{\rm ds}$, but it is temperature invariant if on-chip stress is minimized [9, p. 34] and the induced error can be PTAT trimmed. Resistor mismatch between $R_{\rm pt}$ and $R_{c1,2}$ induced error can also be PTAT trimmed.

However, current gain mismatch δ_{β_f} between $Q_{1,2}$ and resistor mismatch δ_R between $R_{c1,2}$ weaken (12). The induced non-PTAT voltage change at the output is

$$\Delta V_{\delta_R,\delta_{\beta_f}} \approx (\delta_R + \delta_{\beta_f}) \frac{R_{c1}}{\beta_{F1}} I_c(T) \tag{22}$$

where $\delta_{\beta_f} = 1\%$ and $\delta_R = 1\%$ under pessimistic estimations, curvature in $\Delta V_{\delta_R,\delta_{\beta_f}}$ is only ±0.04 mV.

To sum up, based on the error analysis, the bipolar transistor's saturation current, forward current gain spreads, and resistor spread dominate the final PTAT error in $V_{\rm REF}$, which can be trimmed with the designed trimming network. Instead, $X_{\rm TB}$ variation dominates the interdie curvature spread and limits the BGR precision after only a PTAT trimming.

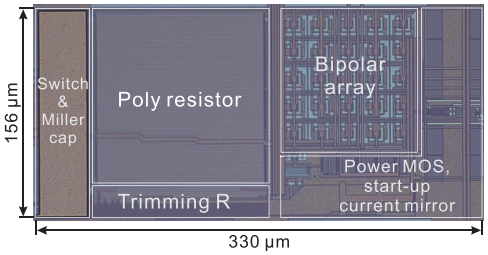

Fig. 7. Chip microphotograph of the BGR.

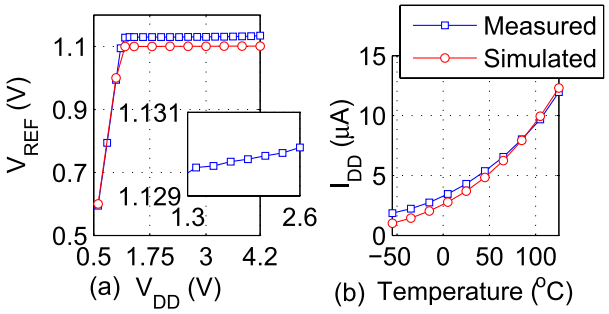

Fig. 8. Measured (a) $V_{\rm REF}$ as a function of $V_{\rm DD}$ at 25 °C and (b) BGR current consumption as a function of temperature.

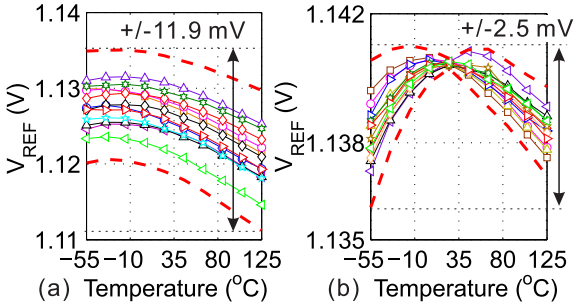

Fig. 9. Inaccuracy of $V_{\rm REF}$ from 12 samples (a) without trimming and (b) after batch trimming resistor R_{c1} at 25 °C. Boundary dotted lines: ±3σ values.

V. Verification and Discussion

A. Prototype Measurement

Fig. 7 shows the micrograph of the prototype BGR, which occupies an area of 0.05 mm^2. The voltage reference $V_{\rm REF}$ is measured under different supplies, as shown in Fig. 8(a). For device reliability, the allowable supply range of the BGR is from 1.3 to 2.6 V with an average line regulation of 0.03%/V (0.34 mV/V). The BGR's nominal current consumption is 4.3 μA and increases to 12 μA at 125 °C [Fig. 8(b)].

As shown in Fig. 9(a), the spread (3σ, box method) of the untrimmed $V_{\rm REF}$ is ±1.06% from −55 °C to 125 °C with an averaged TC of 36.5 ppm/°C. After batch trimming [11] of R_{c1} at 25 °C (by applying the average trimming code from four measured samples to all the 12 samples), the resultant spread is ±0.22% [Fig. 9(b)] and its TCs range from 8.7 ppm/°C to 18.1 ppm/°C [Fig. 10(a)].

As opposed to other current-mode BGR topologies, this BGR can drive a large resistive load. Fig. 10(b) shows the relationship between $V_{\rm REF}$ and the load current I_o under different supplies. At $V_{\rm DD}$ = 1.4 V, for I_o > 10 mA,

TABLE II
PERFORMANCE SUMMARY OF THE RECENTLY REPORTED PRECISION BGRS

Reference		[2]	[3]	[4]	[6]	[11]	[20]	[21]	This work
Year		2014	2007	2007	2012	2011	2012	2007	2014
Technology (μm)		CMOS 0.18	BiCMOS 0.5	CMOS 0.35	BiCMOS 0.5	CMOS 0.16	CMOS 0.35	CMOS 0.35	CMOS 0.18
Active area (mm^2)		0.036	0.4	0.045	0.04	0.12	0.1	1.2	0.05
Supply (V)		1.2	1	0.9 - 4.0	1.6 - 5	1.8±10%	2.5	1.4	1.3 - 2.6
Nominal current (μA)		36	20	0.055	25[3]	55	38	162	**4.3**
Output (V)		0.767	0.191	0.67	1.285	1.088	0.618	0.858	1.1402
Line sensitivity (%/V)		0.1	0.025	0.27	0.027	-	0.039	-	0.03
Temp. range (°C)		-40 to 120	-40 to 125	0 to 80	-40 to 100	-40 to 125	-15 to 150	-20 to 100	**-55 to 125**
Samples		8	32	20	5	61	5	11	12
Inaccuracy (%, 3σ)	Untrim	-	0.6[1]	9.3[1]	-	0.75	-	-	1.06
	Trim	-	-	-	0.21	0.15	-	-	0.22
TC1$^\sharp$ (ppm/°C)		-	11.04[2]	10[2]	-	5-12	-	12.4[4]	**8.7-18.1**
TC2* (ppm/°C)		3.4-6.9	-	-	5-7.2	-	3.9	-	**4.1, 5.2**
PSR (dB)		84@10Hz	-	47@100Hz	70@10Hz	74@DC	-	68@100Hz	54@100Hz
Noise (μV)		$5.42\sqrt{Hz}$ (@320Hz)	$0.04\sqrt{Hz}$ (@100Hz)	-	-	$6.1_{\rm rms}$ (0.1-10Hz)	$0.02_{\rm rms}$ (0.1-50Hz)	$9.1_{\rm rms}$ (0.1-10Hz)	$10.23_{\rm rms}$ (0.1-10Hz)

$^\sharp$without curvature trimming; *with curvature trimming.
1. Extracted from the corresponding literature. 2. Average of the measured samples.
3. Worst case current consumption. 4. Two-point trimming.

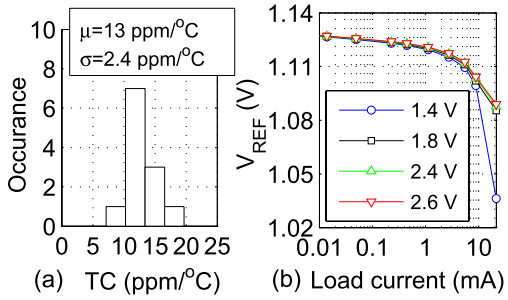

Fig. 10. (a) TC distribution of the PTAT trimmed 12 samples. (b) $V_{\rm REF}$ for different loading current I_o and supply $V_{\rm DD}$.

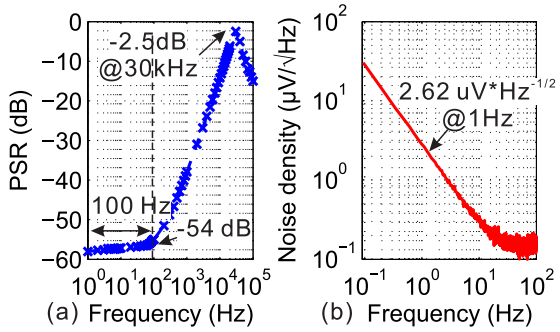

Fig. 11. (a) PSR of the proposed BGR from 1 Hz to 100 kHz without loading. (b) Output noise spectrum from 0.09 to 100 Hz at 25 °C without output filtering capacitor.

M_{p0} (Fig. 5) cannot sustain the output anymore. With higher supply, this BGR has an averaged load regulation of $\Delta V_{\rm REF}/\Delta I_o = 2$ mV/mA up to 20 mA loading current.

As shown in Fig. 11(a), the BGR PSR is 54 dB at 100 Hz and the flat rejection band is relatively narrow because of the small current consumption. The output noise spectrum from 0.09 to 100 Hz without an external filtering capacitor is shown in Fig. 11(b) and the calculated RMS noise within the 0.1 to 10-Hz band is 10.23 μV, which is mostly contributed by the poly resistor $R_{\rm pt}$.

B. Discussion

The measured minimum curvature in $V_{\rm be0}(T)$ from the prototype is 1.4 mV, which is reduced by 61.1% compared with its original 3.6 mV inherent curvature, validating the proposed curvature reduction effectiveness, as shown in Fig. 12(a). However, as shown in Fig. 8(b), the measured BGR current has a weaker temperature dependence compared with that from the simulation. Because such temperature dependence mainly comes from $\beta_{F0}(T)$, a smaller fabricated $X_{\rm TB}$ is expected. In this design, the curvature in $V_{\rm be0}(T)$ was measured to be concave and another concave-shaped voltage developed by

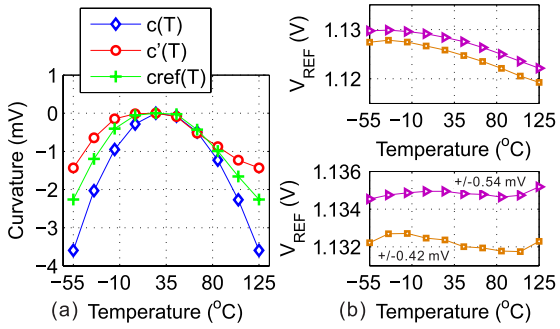

Fig. 12. (a) Curvature performance: BJT intrinsic curvature $c(T)$ before curvature reduction; measured $V_{\rm be0}(T)$ curvature $c(T)'$ after curvature reduction; and measured curvature $c_{\rm ref}(T)$ in $V_{\rm REF}$. (b) Measured $V_{\rm REF}$ before (top) and after (bottom) the compensation by R_{nl1} for two samples.

R_{nl2} results over compensation, which is shown in Fig. 12(a), where the curvature in $V_{\rm REF}$ is even larger than that in $V_{\rm be0}(T)$. Fortunately, such over compensation is systematic and can be recovered by batch tuning R_{nl1}. To demonstrate the effectiveness of $c'(T)$ correction using on-chip nonlinear signals, two samples are measured by adding an external

resistor R_{nl1} (1.5 kΩ). A final curvature as small as 0.84 mV (4.1 ppm/°C) can be achieved [Fig. 12(b)], which validates the curvature correction of $c(T)'$ discussed in Section II-D. It is worth mentioning here that for a process whose X_{TB} is comparable with $\eta - 1$, the shape of its $c'(T)$ deserves careful examination before designing the residual correction circuit.

Table II compares the performance of our design against other recently reported precision BGRs. Notice that the comparison among BGRs is not straightforward because for different designs, the adopted processes, circuit complexities, and optimization targets vary significantly. Meanwhile, the adopted trimming method and the number of trimmed devices also affect the reference accuracy. In this paper, the focus is to design a precision reference while without sacrificing the BGR current consumption and trimming effort. It can be observed that, without adding overhead circuits, the achieved accuracy of this BGR is in line with [2]–[4], [6], [11], [20], and [21] but consumes 4–40 times less current (except [4]). Meanwhile, to achieve the presented TC1, only batch trimming one resistor at room temperature is needed because of the BJT base–emitter voltage curvature reduction and interdie curvature spread control.

VI. Conclusion

Instead of using overhead circuits to passively compensate the existing curvature in V_{be}, an active curvature reduction technique exploiting the silicon BGN effect to design a precision BGR is proposed. Such scheme overweighs most curvature-corrected BGRs in terms of current consumption and design complexity. The design considerations and detailed device analysis in this BGR are presented and its performance is verified experimentally.

Acknowledgment

The authors would like to thank S. F. Luk for his technical support and S. Mohamad for the fruitful discussion.

References

[1] Y. Osaki, T. Hirose, N. Kuroki, and M. Numa, "1.2-V supply, 100-nW, 1.09-V bandgap and 0.7-V supply, 52.5-nW, 0.55-V subbandgap reference circuits for nanowatt CMOS LSIs," *IEEE J. Solid-State Circuits*, vol. 48, no. 6, pp. 1530–1538, Jun. 2013.

[2] B. Ma and F. Q. Yu, "A novel 1.2-V 4.5-ppm/°C curvature-compensated CMOS bandgap reference," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 4, pp. 1026–1035, Apr. 2014.

[3] K. Sanborn, D. Ma, and V. Ivanov, "A sub-1-V low-noise bandgap voltage reference," *IEEE J. Solid-State Circuits*, vol. 42, no. 11, pp. 2466–2481, Nov. 2007.

[4] G. De Vita and G. Iannaccone, "A sub-1-V, 10 ppm/°C, nanopower voltage reference generator," *IEEE J. Solid-State Circuits*, vol. 42, no. 7, pp. 1536–1542, Jul. 2007.

[5] C.-W. Kok and W.-S. Tam, *CMOS Voltage References: An Analytical and Practical Perspective*. New York, NY, USA: Wiley, 2013.

[6] Z.-K. Zhou *et al.*, "A 1.6-V 25-μA 5-ppm/°C curvature-compensated bandgap reference," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 59, no. 4, pp. 677–684, Apr. 2012.

[7] B. Song and P. R. Gray, "A precision curvature-compensated CMOS bandgap reference," *IEEE J. Solid-State Circuits*, vol. 18, no. 6, pp. 634–643, Dec. 1983.

[8] I. M. Filanovsky and Y. F. Chan, "BiCMOS cascaded bandgap voltage reference," in *Proc. Midwest Symp. Circuits Syst.*, vol. 2. Aug. 1996, pp. 943–946.

[9] M. A. P. Pertijs and H. Huijsing, *Precision Temperature Sensors in CMOS Technology*. Dordrecht, The Netherlands: Springer-Verlag, 2006.

[10] I. Lee, G. Kim, and W. Kim, "Exponential curvature-compensated BiCMOS bandgap references," *IEEE J. Solid-State Circuits*, vol. 29, no. 11, pp. 1396–1403, Nov. 1994.

[11] G. Ge, C. Zhang, G. Hoogzaad, and K. A. A. Makinwa, "A single-trim CMOS bandgap reference with a 3σ inaccuracy of ± 0.15% from -40 °C to 125 °C," *IEEE J. Solid-State Circuits*, vol. 46, no. 11, pp. 2693–2701, Nov. 2011.

[12] G. Rincon-Mora and P. E. Allen, "A 1.1-V current-mode and piecewise-linear curvature-corrected bandgap reference," *IEEE J. Solid-State Circuits*, vol. 33, no. 10, pp. 1551–1554, Oct. 1998.

[13] Y. Tsividis, "Accurate analysis of temperature effects in I/SUB c/V/SUB BE/ characteristics with application to bandgap reference sources," *IEEE J. Solid-State Circuits*, vol. 15, no. 6, pp. 1076–1084, Dec. 1980.

[14] W. C. Dillard and R. C. Jaeger, "The temperature dependence of the amplification factor of bipolar-junction transistors," *IEEE Trans. Electron Devices*, vol. 34, no. 1, pp. 139–142, Jan. 1987.

[15] R. A. Pease, "A new Fahrenheit temperature sensor," *IEEE J. Solid-State Circuits*, vol. 19, no. 6, pp. 971–977, Dec. 1984.

[16] P. R. Gray, P. J. Hurst, S. H. Lewis, and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 4th ed. New York, NY, USA: Wiley, 2001.

[17] R. Brederlow, W. Weber, C. Dahl, D. Schmitt-Landsiedel, and R. Thewes, "Low-frequency noise of integrated polysilicon resistors," *IEEE Trans. Electron Devices*, vol. 48, no. 6, pp. 1180–1187, Jun. 2001.

[18] J.-S. Yuan and J. J. Liou, "An improved early voltage model for advanced bipolar transistors," *IEEE Trans. Electron Devices*, vol. 38, no. 1, pp. 179–182, Jan. 1991.

[19] P. Andricciola and H. P. Tuinhout, "The temperature dependence of mismatch in deep-submicrometer bulk MOSFETs," *IEEE Electron Device Lett.*, vol. 30, no. 6, pp. 690–692, Jun. 2009.

[20] C. M. Andreou, S. Koudounas, and J. Georgiou, "A novel wide-temperature-range, 3.9 ppm/°C CMOS bandgap reference circuit," *IEEE J. Solid-State Circuits*, vol. 47, no. 2, pp. 574–581, Feb. 2012.

[21] R. T. Perry, S. H. Lewis, A. P. Brokaw, and T. R. Viswanathan, "A 1.4 V supply CMOS fractional bandgap reference," *IEEE J. Solid-State Circuits*, vol. 42, no. 10, pp. 2180–2186, Oct. 2007.

Bo Wang (S'12) received the B.Eng. degree in electrical engineering from Zhejiang University, Hangzhou, China, in 2010, and the M.Phil. and Ph.D. degrees in electronic and computer engineering from The Hong Kong University of Science and Technology (HKUST), Hong Kong, in 2012 and 2015, respectively.

He is currently a Research Associate with the S2IS Laboratory, HKUST.

Man Kay Law (M'11) received the B.Sc. degree in computer engineering and the Ph.D. degree in electronic and computer engineering from The Hong Kong University of Science and Technology, in 2006 and 2011, respectively.

He is currently an Assistant Professor with the State-Key Laboratory of Analog and Mixed-Signal VLSI, Macau University, Macau, China.

Amine Bermak (M'99–SM'04–F'13) received the M.Eng. and Ph.D. degrees in electronic engineering from Paul Sabatier University, Toulouse, France, in 1994 and 1998, respectively.

He is currently a Professor with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology. He also serves as the Director of Computer Engineering and the Director of the M.Sc. degree program in integrated circuit design.