Bias Current Generators with Wide Dynamic Range

Author(s):
Delbrück, Tobias; van Schaik, André

Publication date:
2005-06

Permanent link:
https://doi.org/10.3929/ethz-b-000032498

Originally published in:
Analog Integrated Circuits and Signal Processing 43(3), https://doi.org/10.1007/s10470-005-1606-1

Analog Integrated Circuits and Signal Processing, 43, 247–268, 2005

Bias Current Generators with Wide Dynamic Range

TOBI DELBRÜCK AND ANDRÉ VAN SCHAIK

[1] *Institute of Neuroinformatics, ETH Zürich, and the University of Zürich, Winterthurerstr. 190, CH-8057, Zürich, Switzerland*
[2] *School of Electrical and Information Engineering, The University of Sydney, NSW 2006 Australia*
E-mail: tobi@ini.phys.ethz.ch, andre@ee.usyd.edu.au

Received September 24, 2004; Revised November 1, 2004; Accepted December 15, 2004

Abstract. Mixed-signal or analog chips often require a wide range of biasing currents that are independent of process and supply voltage and that are proportional to absolute temperature. This paper describes CMOS circuits that we use to generate a set of fixed bias currents typically spanning six decades at room temperature down to a few times the transistor off-current. A bootstrapped current reference with a new startup and power-control mechanism generates a master current, which is successively divided by a current splitter to generate the desired reference currents. These references are nondestructively copied to form the chip's biases. Measurements of behavior, including temperature effects from 1.6 and 0.35 μ implementations, are presented and nonidealities are investigated. Temperature dependence of the transistor off-current is investigated because it determines the lower limit for generated currents. Readers are directed to a design kit that allows easy generation of the complete layout for a bias generator with a set of desired currents for scalable MOSIS CMOS processes.

Key Words: bias current generator, current reference, current splitter, current divider, PTAT

1. Introduction

Analog or mixed-signal CMOS chips usually require a number of fixed reference currents for biasing amplifiers, determining time constants and pulse widths, powering loads for static logic, and so on. Chips will often have large pluralities of identical circuits (e.g. pixels, column amplifiers, or cells) that require nominally identical biases. The required currents can extend over many decades. For instance, consider a chip with circuits that span timescales from ns to ms and that uses subthreshold g_m-C filters with 1 pF capacitors. The rise time T—which we take as the timescale—of simple g_m-C circuits scales as C/g_m. The transconductance g_m of a transistor in subthreshold operation scales as I/U_T, where I is the bias current and U_T is the thermal voltage. Thus, such a chip would require bias currents $I = CU_T/T$ from 10 uA to 10 pA—a range of six decades.

Bias current references are often left out in experimental chips because designers assume that these "standard" circuits can easily be added in later revisions when the chip's design is productized. As a result, chips are designed that must later be individually tuned for correct operation. These biases are often specified by directly setting bias transistor gate voltages using off-chip components. However, the required voltages depend on chip-to-chip variation in threshold voltage. If these voltages are generated by potentiometers or supply-referenced digital-to-analog converters (DACs), they are sensitive to supply ripple. The supply currents depend in an exponential way on temperature. Potentiometers and DACs consume macroscopic amounts of power and are expensive items for consumer goods. More importantly, each chip requires individual tuning, which can be difficult and time-consuming, especially if the space of tuning parameters has many dimensions. Furthermore, any drift in off-chip components can be difficult to correct.

Some designers generate scaled versions of the desired currents using off-chip resistors that supply current to on-chip scaling current mirrors [1]. Although this biasing scheme is straightforward, it requires a pin for each independent bias, is sensitive to threshold voltage, requires a possibly wasteful regulated power supply, and can necessitate bulky on-chip current mirrors when a small on-chip current is required. For example, if we take a maximum feasible off-chip resistance of

1 MΩ, then a 1 pA on-chip current requires a bulky scaling current mirror—or series of mirrors—with the ratio 10^6:1.

Neither of the foregoing design choices guarantees the feasibility of manufacturing the chip in quantity. Here we show the architecture of the bias generator circuits that we have been using regularly (e.g. [2–5]) to derive a wide-ranging set of fixed bias currents from a single generated master current. This paper is an expanded version of the work originally presented in [6], with new analysis and measurements. Section 2 describes the circuits, Section 3 the measurements, Section 4 the design kit, and Section 5 concludes this paper.

2. Biasing Circuits

The proposed circuit shown in Fig. 1 generates the master current I_m and scaled copies of it. The master current is subdivided to form a set of smaller references, which are copied by the circuits described in Section 2.3 to form the individual biases. The total supply current in the core bias generator circuit is $3I_m$, consisting of $2I_m$ in the master bias and I_m copied to the splitter. In the following discussion, transistor "off-current" means the saturation drain current of a transistor with gate and source both tied to the bulk. This current is also known as the subthreshold leakage current and is not the diode or junction leakage current from active region to bulk.

2.1. The Master Bias

The master current I_m is generated by the familiar bootstrapped current reference attributed to Widler [8, 9] and first reported in CMOS by Vittoz and Fellrath [10] (see also textbooks such as [1, 11–13]). Transistors M_{n1} and M_{n2} have a gain ratio $(W_{n1}/L_{n1})/(W_{n2}/L_{n2}) = M$. Since the currents in the two branches are forced to be the same by the mirror M_{p1}–M_{p2}, the ratio in current density in the M_n's sets up a difference in their gate-source voltage, which is expressed across the load R. Resistance R and ratio M determine the current. The master current I_m that flows in the loop is computed by equating the currents in the two branches. In subthreshold, this equality is expressed by $I_m = e^{\kappa V_n/U_T} = Me^{(\kappa V_n - I_m R)/U_T}$, where κ is the back-gate or body-effect coefficient (also known as $\kappa = 1/n$), resulting in the remarkably simple yet accurate formula

$$I_m = \log(M)\frac{U_T}{R}, \quad U_T = \frac{kT}{q} \tag{2.1}$$

U_T is the thermal voltage. The voltage V_R across the load resistor R does not depend on the resistance R in subthreshold and provides a direct measurement of temperature.

$$V_R = \log(M)U_T \tag{2.2}$$

Above threshold, an analogous computation yields another formula that is not very accurate but still useful

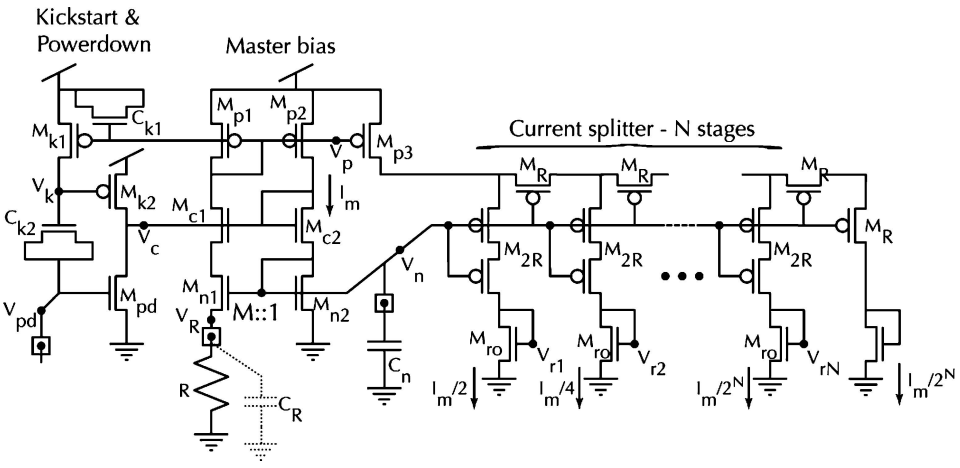

Fig. 1. Bias generator core circuits. I_m is the master current, and R is the external resistance. Transistor sizes are in units of λ (the scalable length parameter) and are listed in Table 1. C_{k1} and C_{k2} are MOS capacitors. M_R and M_{2R} are identically sized unit transistors. The squares represent bonding pads and recommended external connections.

Table 1. Transistor and capacitor sizing for the circuit in Fig. 1. Transistor width to length ratios (W/L's) are given in λ, the MOSIS [7] scalable parameter, and are 24/6 unless listed differently. A minimum length transistor is 2 λ long, so 2 λ is usually the process technology dimension (e.g. $\lambda = 0.4\ \mu$m for a 0.8 μm technology); for submicron processes, λ is sometimes slightly larger than this (e.g. a MOSIS 0.35 μ process has $\lambda = 0.2\ \mu$).

Transistor W/L	
M_{n2}	24/6
M_{n1}	M^* 24/6
M_{p1}, M_{p2}, M_{p3}	76/65
M_{c1}, M_{c2}	24/6
C_{k1},C_{k2}	132/20
M	40
M_R, M_{2R}	24/12
M_{pd}, M_{k1}, M_{k2}	6/6
Capacitance	
C_n	∼10 pF
C_{k1}, C_{k2}	∼1 pF

for estimating the required resistance.

$$I_m = \frac{2}{\beta_n R^2}\left(1 - \frac{1}{\sqrt{M}}\right)^2, \quad \beta = \mu_n C_{ox}\frac{W_{n2}}{L_{n2}} \quad (2.3)$$

Here μ_n is the electron-effective mobility, and C_{ox} is the unit-gate oxide capacitance. In strong inversion the current decreases with R^2, while in weak inversion it decreases as R. Hence—and as shown later by the data in Fig. 8—the estimated I_m is approximately the sum of Eqs. (2.1) and (2.3). With ideal transistors I_m does not depend on supply voltage or threshold voltage, but is closely proportional to absolute temperature (PTAT) in subthreshold. In reality it is slightly affected by the supply voltage through drain conductance and also by mismatch of the threshold voltage and β between the transistors in the current mirrors.

This master bias circuit is often called the constant-g_m circuit because the g_m of a transistor biased with current I_m is independent of temperature for both weak and strong inversion. The transconductance of a transistor with W/L the same as M_{n2} biased with current I_m is given by

$$g_m = \overbrace{\frac{\kappa \log M}{R}}^{\text{weak}}, \qquad \overbrace{\frac{2(1 - 1/\sqrt{M})}{R}}^{\text{strong}} \quad (2.4)$$

These g_m depend only on R, M, and κ, and the β-dependence of the strong inversion master current has also disappeared [14]. Thus, g_m does not depend on temperature in either weak or strong inversion if R and κ are independent of temperature. As discussed in [15] in this issue, this temperature-independence holds only if the transistor is of the same type as M_{n1} and running in the same operating regime. Therefore, we expect that circuits that are biased from the splitter outputs will have some degree of temperature-dependent g_m. Temperature dependence of the bias generator is discussed in more detail in Section 3.6.

2.1.1. Power Supply Sensitivity To decrease the DC power supply sensitivity of the master bias current, the drain resistances of the transistors are increased by using long M_p's and cascoding M_{n1} with M_{c1} and M_{c2}. This choice minimizes the size of the entire generator. We chose not to cascode the p-FETs to preserve headroom. Razavi computes the power supply sensitivity of the master bias current as an exercise ([12], example 11.1). The result of this small-signal analysis is interesting and a bit surprising in that the sensitivity vanishes if the M_{p2} mirror output transistor in Fig. 1 has infinite drain resistance. In other words, if the p-mirror copies the current perfectly, the output resistance of the M_{n1} (or M_{c1}) transistor is irrelevant. Why is this plausible? If the p-mirror copies perfectly, it is impossible for the n-mirror to have unequal output current. Therefore, the original premise of the circuit that is expressed in Eqs. (2.1) and (2.3) is satisfied and power supply variation has no effect. One might still think that finite drain conductance in M_{n1} increases the gain of the M_{n1}–M_{n2} mirror, but this is not the case. Drain conductance does not increase incremental mirror gain; it only increases output current, and incremental current gain is what determines the master bias current. Simulations of the master bias circuit that increase the length of the M_{n1}–M_{n2} mirror (which decreases the mirror's output conductance) slightly *increase* the master bias current.

An additional interpretation of the result of Razavi's analysis is that increasing M reduces supply variation. This interpretation is also reasonable because increasing gain in a feedback loop decreases the effects of component imperfection. All of these effects are shown in the simulation results of the supply sensitivity of the low-voltage version of the master bias circuit (see Section 3.7.1). The results suggest that making transistors M_{p1} and M_{p2} long, excluding the M_{c1}–M_{c2} cascode, and using a large M are likely good alternative choices for low-voltage operation.

2.1.2. Stability The ratio M is not critical as long as it is substantially larger than 1. We have used values from 20 to 120, and the measurements shown here come from the design kit described in Section 4, which uses $M = 40$. A very large ratio can destabilize the circuit through the parasitic capacitance C_R on V_R. A common error in this circuit (and one not mentioned in any of the standard texts or original references) is to have too much capacitance C_R, which can cause large-signal limit-cycle oscillations. The circuit can be stabilized by making the compensation capacitor C_n several times C_R. In practice, we usually bring out V_n to a bonding pad, where we can use an external capacitance to ensure that the master bias can be stabilized. Nicolson and Phang [15] show a new topology for the master bias circuit that requires much less compensation capacitance. Lichtsteiner proposed another compensation scheme [16] that places C_n between the two legs of the master bias circuit so that destabilizing swings in one branch are compensated by swings in the other branch. For example, if the current increases in the left branch, the downward movement of the left-branch voltage causes a downward movement in the right-branch voltage, which counteracts the increase in current. He demonstrated in simulation that this arrangement is stable even when $C_R/C_n = 1,000$, but we have not yet implemented this arrangement.

As an aside, this large-signal instability is not easy to analyze, because a small-signal analysis shows that all poles are almost always in the left-half plane regardless of what values are chosen for capacitance and M [16]. The circuit is therefore nearly always small-signal stable. Even when the circuit is small-signal stable, it can still easily be large-signal unstable. If one simulates the circuit behaviorally with subthreshold dynamics and infinite power supply rails, it is also large-signal stable, and any oscillation eventually damps out. The large-signal instability arises from the extremely nonlinear (exponential) large-signal characteristics of the current mirror transistor M_{n1} and the low impedance of C_R at high frequencies, allowing transient positive feedback that approaches or even exceeds unity gain. This can be understood by considering that in steady state, M_{n1} and M_{n2} carry identical currents, but M_{n1} is source-degenerated in DC by R, so that in DC the current gain from M_{n1} to M_{n2} is less than 1. At high frequencies, C_R provides a virtual short to ground for the source of M_{n1}. In this condition the gain from M_{n1} to M_{n2} approaches 1, at least when I_m is subthreshold where $g_m = \kappa I / U_T$ irrespective of the transistor geometry. When I_m is above threshold, the high frequency g_m of M_{n1} will be greater than that of M_{n2} when they carry the same current because the M_{n1} overdrive will be lower. It is therefore possible that when I_m is above threshold, positive feedback can exceed unity gain; and the larger the M ratio, the larger this effect. In any case, this nonlinearity causes a response to a perturbation that can easily cause the voltages to hit the power rails, shutting off the current in the mirrors so that the oscillation cannot catch up with itself and limit-cycle oscillations continue forever. Coupled with the extremely expansive nonlinearity of M_{n1}, this positive feedback makes limit-cycle oscillations simple to generate, e.g. on startup. Slowing the other branch with C_n reduces the positive feedback to the gate of M_{n1} and prevents the instability. Section 3.1 shows measurements of this instability.

2.1.3. Startup and Power Control Very small current—traditionally but incorrectly called "zero current"—in both branches of the bootstrapped current mirror circuit can also be a stable or metastable operating point [11]. Although this state can definitely occur in implementations, why it does is not so easy to see. A straightforward analysis shows that when a current mirror's input transistor goes out of saturation, the output of the mirror reduces to the off-current, but the mirror's incremental current gain is reduced only by a factor of the back-gate coefficient κ. This situation is illustrated in Fig. 2, which shows the degenerated mirror over a wide range of currents. The output of the mirror is the output transistor's off-current when the input current is zero, and the current gain is $M\kappa$, where $\kappa = 1$ in this example. Thus, the total current gain around the loop when both mirrors are in their "off" state, with both gate-source voltages zero, is $M\kappa_n\kappa_p$, where p and n refer to the n- and p-type back-gate coefficients, and this factor is certainly larger than one in most implementations. Considering substrate leakage from the drain junctions does not change this situation, but a conductance from V_n to ground can produce a stable "off" state. Ordinarily there is no such intentional conductance, but substrate leakage from the ESD protection structures in the V_n pad or across the drain of M_{pd} can act as such a conductance. In addition, off-current or substrate leakage from M_{k2} can supply some of the current sunk by M_{n1}, also reducing the gain. In extensive behavioral as well as transient SPICE simulations, we have not been able to produce a true "off" state. However, we have observed in practice that an

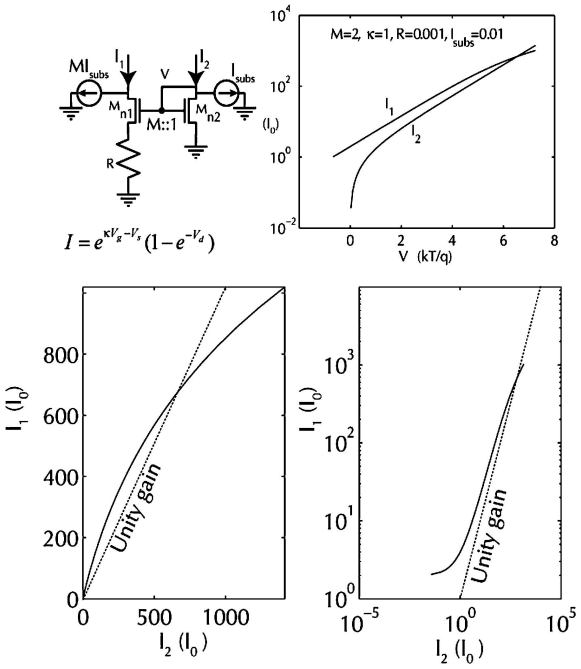

Fig. 2. Detailed action of degenerated mirror showing entire range of operating currents, from "off" to intended currents. Current is normalized to the M_{n2} off-current, and voltages are in units of U_T. R is in units U_T/I_0. Bottom left shows the currents on a linear scale whereas bottom right shows them on a log scale.

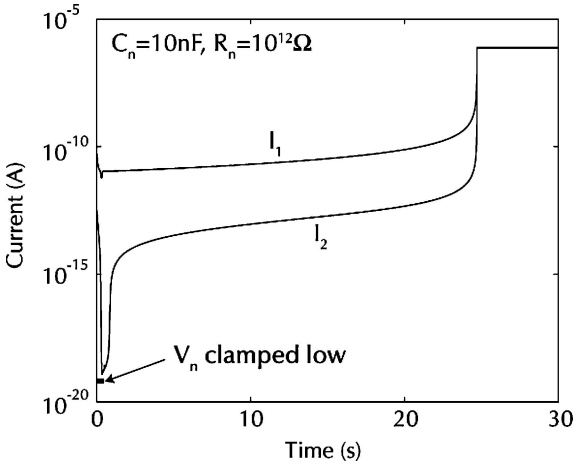

Fig. 3. Simulation of slow restart when off state is intentionally produced by clamping V_n low. A resistor R_n was connected between V_n and ground, and a large capacitance $C_n = 10$ nF was used to demonstrate slow self-restart.

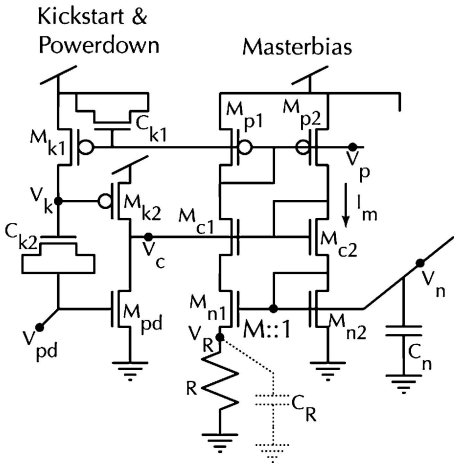

Fig. 4. Close-up of startup and power control circuits.

intentionally produced "off" state (produced by using an extra transistor to tie V_n momentarily to ground) can be stable at room temperature for many seconds. Increasing the temperature, which increases the transistor off-current and the junction leakage current, decreases the duration of this metastable operating point. Whatever the cause, the simulation in Fig. 3 shows that escape from an "off" state can be very slow, even when it is unstable, so a startup circuit is necessary to escape this parasitic operating point quickly when power is applied.

A large number of startup mechanisms are currently in use [12, 17, 18]. In the present circuit we use a new 4-transistor startup circuit that transiently injects current into the current mirror loop on power-up and then shuts itself off completely. Unlike many other startup circuits, this mechanism is process-independent because it does not depend on threshold or supply voltage and does not require any special devices. The inventors of this startup circuit request anonymity, although they have agreed to its description here. It is used on a commercial product that has shipped over 100 million units.

To make the explanation of this startup circuit clearer, part of Fig. 1 is reproduced as Fig. 4.

Transistors M_{k1}, M_{k2}, and M_{pd}, and MOS capacitors C_{k1} and C_{k2} enable the startup and power-control functionality. The loop is kick-started on power-up by the current flowing from M_{k2}, which is "on" until V_k is charged to Vdd by M_{k1}, which then shuts off. C_{k2} holds V_k low on power-up (V_{pd} is at ground), while C_{k1} ensures that V_p is initially held near Vdd, holding M_{k1} "off" so that the kick-start can occur. C_{k1} and C_{k2} must be large enough so that sufficient charge flows into the loop to get it going; we usually use about 1 pF. C_{k1}

and C_{k2} are MOS capacitors to avoid the necessity of a special capacitor layer, such as a second polysilicon layer. The polarity of the MOS capacitors is arranged so that they operate in inversion (C_{k2}) or accumulation (C_{k1}) when they need to. (C_{k1} has another important role that is discussed later.) While the bias generator is operating, essentially zero current flows in this startup circuit. The charge injected by M_{k2} is a complex function of circuit parameters, but the essential point is that M_{k2} is not shut off until the master bias has current flowing in it.

If the master bias circuit ever falls into a metastable low-current state, there is no rapid automatic recovery. We have not experienced this problem or been able to produce it in simulations or experiments by manipulation of the power supply, but it is possible that such a circumstance could arise under, for instance, very deep brown-out conditions with slow recovery of the supply voltage. Capacitor C_{k1} is important here because it tends to hold the gate-source voltage of M_{p1}—and hence its current—constant when Vdd changes. If C_{k1} is not included, a sudden drop in Vdd can transiently turn off M_{p1}, possibly leading to an unintended and extended shutdown of the master bias.

In some systems the ability to completely shut off all bias currents and then restart them is desirable, such as for a sensor chip that needs only periodic activation by an external periodic wakeup signal. We have included a method to enable this "soft" power control by input V_{pd}, which is grounded for normal operation. Raising V_{pd} to Vdd turns off the master bias and the derived biases by pulling V_c to ground through M_{pd} and shutting off the current in the loop. Yanking V_{pd} back to ground yanks V_k low, through C_{k2} (M_{k1} is "off"), and the start-up circuit restarts the current as before. While V_{pd} is high, no current flows in $M_{pd,}$ because V_k is at Vdd and M_{k2} is off. A conductive path to ground from V_n (say, through a leaky C_n) could require pumping V_{pd} for a few cycles at a sufficient rate to move the current mirror loop to a regime of positive feedback. But if, as usual, there is no DC path other than through M_{n2}, a single downward transition on V_{pd} is sufficient for restart, as demonstrated in the measurement in Fig. 10.

2.2. The Current Splitter

The master current is copied to a Bult and Geelen [19] current splitter, which successively divides it to form a geometrically spaced series of smaller currents. At each branch, half of the current is split off, and the rest continues to later stages. The last stage is sized to terminate the line as though it were infinitely long. In Fig. 1, M_R and the two M_{2R} transistors (which each have the same W/L as the M_R transistor) form the R-2R network; the octave splitter is terminated with a single M_R transistor. The splitter has N stages, and the current flowing transversely out from the splitter at the kth stage is $I_m/2^k$. The final current is the same as the penultimate current. Our transistor sizing for the octave splitter is given in Table 1. The reference voltage for the p-FET gates in the splitter should be a low voltage to minimize splitter supply- voltage requirements, but it needs to be high enough to saturate the diode-connected n-type output transistors. We use the master bias voltage V_n, which conveniently scales correctly with master bias current. We chose p-FET devices for the splitter because they are built in an n-well implanted in a p-substrate and can be protected from the effects of parasitic photocurrents simply by covering them with metal.

The current splitter principle accurately splits currents over all operating ranges, from weak to strong inversion, independent of everything but the effective device geometry. In this R-2R splitter, behavioral independence from the operating regime is most easily understood by following each transistor's operation back from the termination stage and observing that the transverse M_{2R} and lateral M_R transistors share the same source and gate voltage and that the transverse transistor is in saturation. It can be easily observed that combining series and parallel paths causes half the current to flow into each branch at each stage without any assumption about channel operating conditions. It is also easy to see that, looking from the input terminal, the entire splitter forms a "compound transistor" that has an effective W/L equal to one of its M_R or M_{2R} unit transistors.

That the transverse transistor is in saturation can be observed as follows. Assuming that n- and p-type transistors have comparable threshold voltages and ignoring back-gate effect, the source of this compound splitter transistor will be at approximately 2 V_n. The drain will be at V_n because this is the gate voltage of the compound M_{ro} diode-connected readout transistor. Therefore, the splitter will have about V_n across its "drain-source", ensuring that it is in saturation. The same will hold for the individual transverse transistors in the splitter because they and the corresponding M_{ro} will carry only scaled copies of I_m.

Figure 1 shows an R-2R splitter—built from unit transistors—that splits by octaves, but we have also built decade splitters by using M_R and M_{2R} with different aspect ratios. However, we strongly recommend the use of unit transistors in an R-2R configuration. We have discovered subtle effects that act differentially on transistors with differing aspect ratios. These effects cause non-ideal splitter behavior, especially in deep subthreshold, and are discussed in Section 3.3.1.

The diode-connected M_{ro} transistors read out the currents to make copies for individual biases. This arrangement allows for non-destructive readout at the cost of mismatch in the M_{ro} transistors.

2.3. *Generating Individual Biases*

An individual bias (V_{bn}, V_{bp}) is generated by copying a splitter current using one of the cells shown in Fig. 5. A *p*-type bias is generated by (a). First, the splitter current is copied using a cascoded transistor for better accuracy. This current is drawn from a diode-connected *p*-MOS transistor with the desired W/L ratio. The W/L ratio of the transistor is the same as the W/L ratio used in the user's circuit. The resulting gate voltage is then used as the bias voltage and is wired to other parts of the chip. We have used this "voltage routing" distribution method exclusively, although "current routing"—where the splitter current is copied and routed to the place it is needed—is, of course, also feasible if the bias is required in only a small number of places [1]. An *n*-type bias is generated by (b). The *p*-type mirror used for the *n*-type bias is cascoded for better accuracy, and the bulk of the cascode transistors is tied to the gate voltage of the mirror. This arrangement provides a bit more headroom because the back-gate bias is reduced, which reduces the required gate-source voltage. A differential pair is used by (c) to enable fine-tuning of the programmed bias in a controlled manner by external input V_{tune}. Tying V_{tune} to V_c programs half of the splitter current, and the actual value can be varied from zero to the full splitter current. An additional diode-connected copy of the W/L transistor could be used at the drain of the V_c transistor to improve circuit symmetry, but because the bias is tunable this extra transistor might be superfluous.

2.3.1. Bypass Decoupling of Individual Biases A diode-connected transistor sinking current I_b and operating in subthreshold has a gate or drain conductance $g \approx I_b/U_T$. This means that the bias voltage for a small bias current will have a high impedance and can easily be disturbed by other signals on the chip that are capacitively coupled to it, e.g. by crossing wires or by drain-gate parasitic capacitance. The simplest remedy is to bypass the bias with a large capacitance to the appropriate power rail (Vdd for *p*-type and ground for *n*-type; see Fig. 5), which is easy to do if the bias is brought off-chip. Bypassing the bias has the additional benefit of greatly reducing the effect of power-supply ripple on the bias current. It is important to bypass to the appropriate power rail so that the bias voltage is better stabilized relative to the appropriate transistor source voltage. The parasitic capacitance to the other rail will then have much less effect on the gate-source voltage. In a production chip, a pad may not be economically justifiable, but in a prototype chip we strongly advise bringing all biases out to pads anyway.

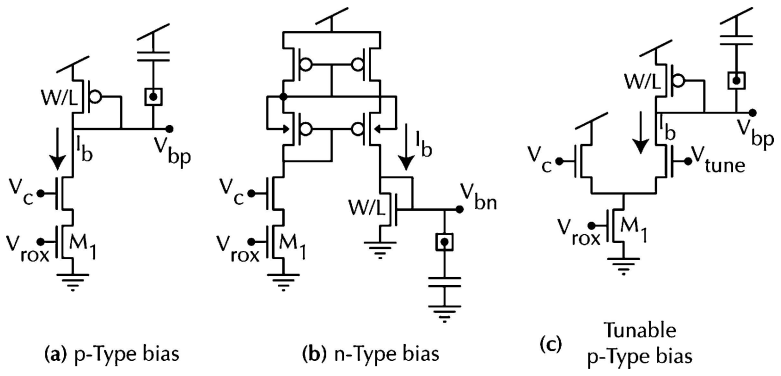

Fig. 5. Generating individual biases from the current splitter outputs. M_1 has same W/L as M_{ro} in Fig. 1. The square attached to each capacitor represents a bonding pad.

If the chip will be exposed to light, care must be taken when bringing these generated bias voltages off-chip because ESD protection structures in the bonding pads can produce significant currents under illumination (e.g. several nA under 1 klux). When the programmed bias current is very small, these parasitic photocurrents in the bonding pads can significantly perturb the bias currents. In addition, parasitic conductance between package pins can significantly affect the generated biases when bias currents are in the sub-nA range, especially under humid conditions.

We have also investigated active buffering of the generated bias voltages to reduce the effects of coupling. The total capacitance on the bias voltage is often large when a large number of identical circuits (e.g. pixels) are biased. On one chip, we tried using a source-follower arranged in a current conveyer structure, as shown in Fig. 6(a), to buffer the low- current biases, but we observed that even with a huge current ratio I_{buf}/I_b of 10^4 (1 μA/100 pA), transient capacitive coupling to the bias produced a systematic, activity-dependent shift of the bias current. The source-follower has very unsymmetrical large-signal characteristics and acts as a peak detector for transients coupling to the bias line, so the net result is a systematic shift in the bias current that depends on the frequency and amplitude of transient coupling. If the individual capacitive coupling is small, then this effect is probably insignificant unless it is synchronous. We have not experimentally investigated the use of a linear amplifier as the buffering element, as shown in Fig. 6(b), but simulations suggest that it would not exhibit a systematic bias current shift. However, the resonance frequencies of this circuit must be considered because the amplifier may be driving a large capacitance, so the time constants at the input and output nodes of the amplifier can be comparable. If the resonance frequencies are comparable to the disturbance frequencies, the buffer can amplify the disturbance rather than suppress it. The circuit in Fig. 6(c) is more stable (assuming that the amplifier can stably drive its load capacitance) because the amplifier's internal time constants will generally be much smaller, but it introduces random and systematic offset to the bias voltage.

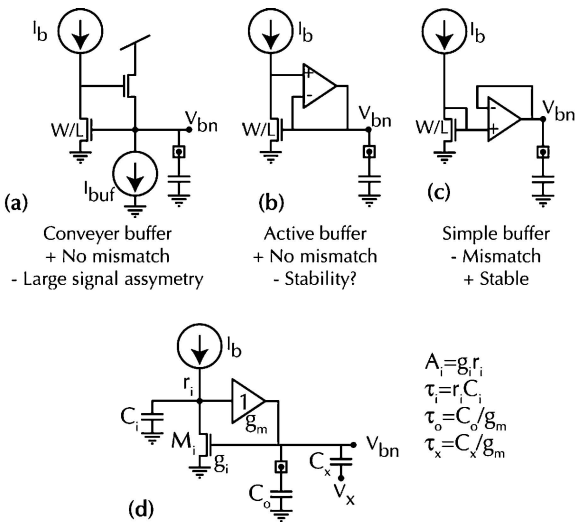

Fig. 6. Active bias-voltage buffering circuits. The desired bias voltage V_{bn} for a bias current I_b is actively buffered to other parts of the chip. (a) is simple but has large signal asymmetry that can systematically shift bias value in response to coupled disturbances. (b) is linear and does not introduce significant mismatch but must be biased correctly, as must (a), to avoid resonance amplification. (c) is straightforward but can introduce additional random and systematic mismatch. (d) shows the circuit analyzed for biasing conditions.

We can calculate the condition that avoids resonance in Fig. 6(b) using the equivalent circuit in Fig. 6(d). The response to a sinusoidal disturbance V_x coupled through capacitance C_x is given by (2.5), where the parameters are shown in the figure.

$$\frac{V_{bn}}{V_x} \approx \frac{\tau_x s(\tau_i s + 1)}{(\tau_i s + 1)(\tau_o s + 1) + A_i} \tag{2.5}$$

To achieve critical damping, condition (2.6) must apply to the buffer amplifier time constant.

$$\tau_o < \frac{\tau_i}{4A_i} \tag{2.6}$$

Here τ_o is the "time constant" of the output of the unity-gain buffer amplifier, which drives (usually) a large capacitance. τ_i is the time constant of the amplifier input node, which consists of the small input capacitance C_i and the large input resistance $r_i \approx V_E/I_b$, where V_E is the Early voltage at the input node. A_i is the gain of the input node looking from the gate of M_i.

If $\tau_o = \tau_i$, we have the condition of maximum resonance, and $Q = \sqrt{A_i}/2$, which will typically be about 10. When the circuit is properly biased according to (2.6), however, the equivalent time constant of the high pass filter shrinks to (2.7):

$$\tau = \frac{\tau_i}{4A_i} = \frac{r_i C_i}{4 r_i g_i} = \frac{C_i}{4g_i} \approx \frac{C_i U_T}{I_b} \tag{2.7}$$

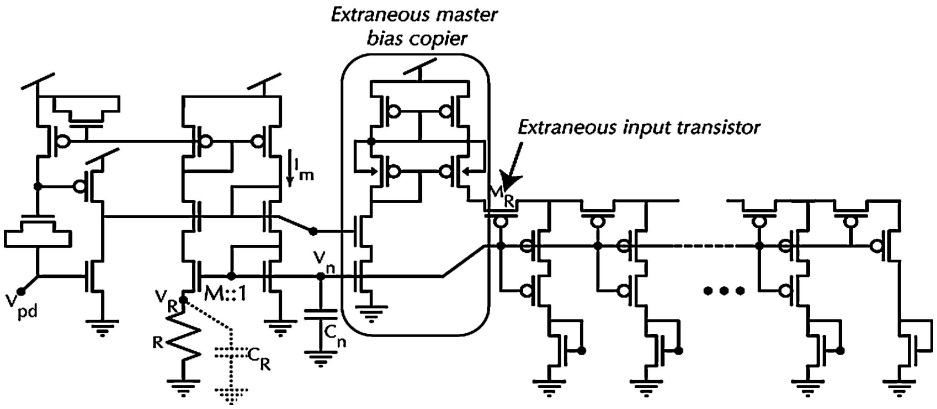

Fig. 7. Measured bias generator circuit.

Compared with the diode-connected schemes of 2.3.1, which have a high-pass time constant of $\tau = C_o U_T / I_b$, the actively buffered bias high-pass time constant is reduced by a factor C_o/C_i, which can be many decades.

3. Measured Characteristics

We have used the bias generator circuits described in this paper in several generations of CMOS process technology (1.6, 0.8, and 0.35 μ) with no striking differences in performance. Here we show measurements and discuss the limits of operation. Most of the results shown here come from bias generators with 20-stage octave splitters built in two different 3.3 V, 0.35 μ processes. These layouts were generated by the design kit described in Section 4.

While writing this paper, we realized that the fabricated circuits require more voltage headroom than they should. Figure 7 shows the measured circuit. Compared with the proposed circuit of Fig. 1, it copies the current from the master bias to the splitter using a doubly cascoded mirror and includes an extra p-FET on the front of the splitter. These additional transistors reduce possible headroom and provide no advantage; their effect is visible, e.g. in the measurements shown later in Fig. 17. (These shortcomings have not been corrected in the design kit.)

3.1. Master Bias

The variation with resistance R of the master current I_m and the voltage V_R across the load resistor R are shown in Fig. 8, along with the theoretical values given by Eqs. (2.1) and (2.3) and SPICE BSIM3v3 simulation results. The theory gives a reasonable estimate for the measured values in the subthreshold range of operation, and the SPICE simulation does even better. The above-threshold model is not very accurate. The exact behavior is usually not important for practical purposes

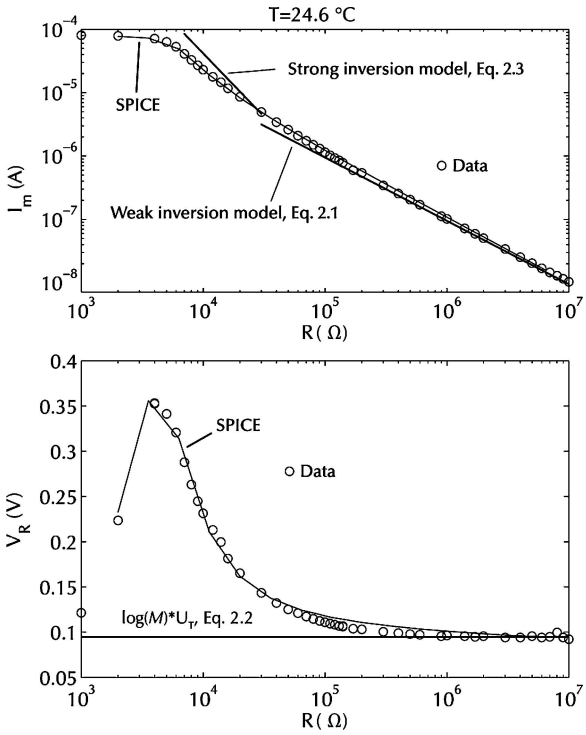

Fig. 8. Master current I_m and voltage V_R versus resistance R.

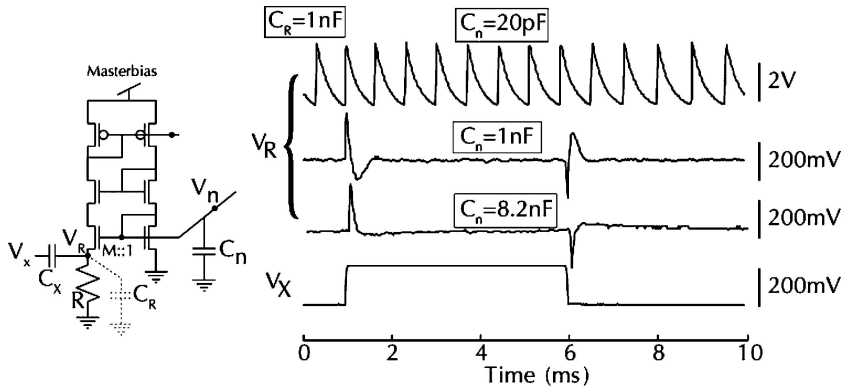

Fig. 9. Measurements of master bias stability. V_X is an external signal capacitively coupled into the V_R node.

because R is generally external and is selected for the desired operating point. The maximum possible current is determined by the power supply voltage and the headroom required by the current mirrors.

Measurements that demonstrate possible master bias instability are shown in Fig. 9. We used large external C_R and C_n to intentionally produce the various regimes of stability and coupled an external perturbation V_X to the V_R node using a 220 pF C_X capacitor with a 200 mV square wave input. C_R was fixed at 1 nF, and we varied C_n between 20 pF (the oscilloscope probe), 1 nF, and 8.2 nF. When $C_n/C_R = 1$, the master bias is marginally stable; C_n/C_R = 8.2 makes the master bias circuit unconditionally stable. The master bias oscillated when C_n consisted of only the 20 pF oscilloscope probe, corresponding to an undesirable situation where excess capacitance is allowed on the V_R node.

3.2. Power Control Circuits

The power control behavior is shown in Fig. 10. V_{pd} is initially high (the powered "off" state) and is then brought low. V_n first jumps upward when M_{k2} injects a packet of charge and then increases exponentially until it reaches its stable value. An unusually large $C_n = 10$ nF was used to demonstrate that a large C_n does not affect restart except to delay it.

A conductive path to ground from V_n (e.g. from an oscilloscope probe) can make it harder to restart the master bias current because the transient current injected by the transient drop in V_{pd} is leaked away by the resistive load faster than the positive feedback can restore it. In this case, the master current can still be

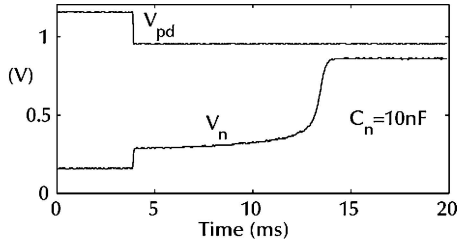

Fig. 10. Power control with V_{pd} (V_{pd} is rescaled). There was a 30-second delay since the last cycle. A high-impedance analog buffer was used to prevent a conductive path from V_n to ground.

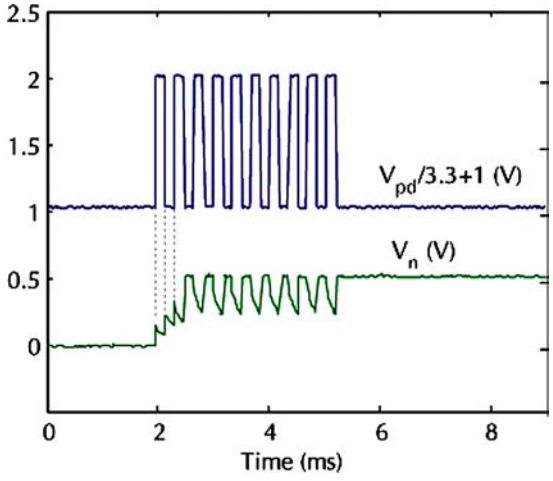

Fig. 11. Restarting the master bias by using multiple pulses on V_{pd} when there is a conductive path from V_n to ground.

restarted by a series of low-going pulses on V_{pd}, as illustrated in the measured data in Fig. 11. In this setup, the resistive load was a 10 MΩ oscilloscope probe, and three to four pulses were necessary to restart the

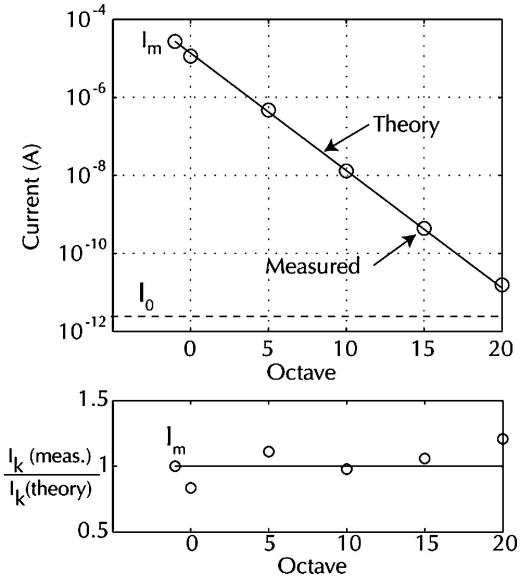

Fig. 12. Octave splitter behavior.

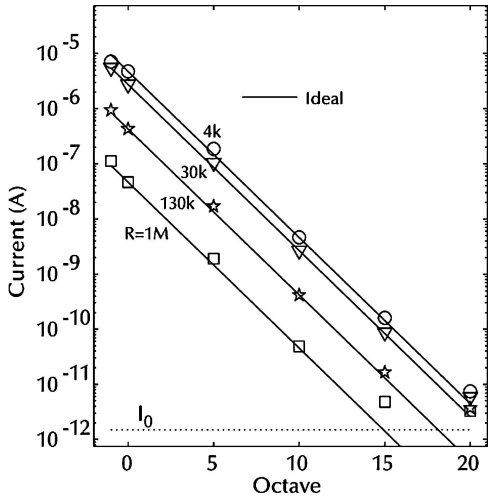

Fig. 13. Measurements of the splitter currents using a variety of master bias resistances R. The measured nFET off-current is shown as I_0.

master bias. The capacitance C_n was the oscilloscope probe capacitance of about 20 pF.

3.3. Current Splitter

The behavior of the octave splitter is shown in Fig. 12. A separate n-type transistor with W/L = 24/6 was used to measure the splitter currents. We connected the splitter output voltages successively to the transistor's gate while holding its drain in saturation and measuring its drain current using a Keithley 6430 source measure unit (www.keithley.com). To ensure that this test transistor, which is located far distant from the bias generator on the chip, had the same threshold voltage and body effect as the M_{ro} transistors reading out the splitter current, we measured them each in a diode-connected arrangement.

The splitter behavior is amazingly ideal over 20 octaves (6 decades) spanning strong to weak inversion. A current of 10 pA is reliably generated from a master current of 10 uA. Each splitter current is within 10% of the ideal predicted value. The measured n-type transistor off-current ($V_g = V_s = 0$, V_d in saturation) of the test transistor used to measure the splitter currents is 3 pA at room temperature. Figure 13 shows more measurements of the current splitter, with a range of master currents determined by varying the external resistance R. Each set of measurements is ideal to within about 10% down to a few times the transistor off-current I_0.

Imagers and focal plane arrays generally require biases that are not affected by illumination. However, illumination creates parasitic photocurrents in all uncovered transistor source and drain regions as well as in covered, native-type transistors that can collect diffusing minority carriers. These parasitic currents have particularly significant effects on transistors with low currents or with large areas of affected junctions. The effect of parasitic photocurrent can be greatly reduced by covering transistors with opaque metal and ensuring that native transistors (ones that are built in the wafer substrate) are surrounded by guard rings. These guard rings are generally built using well implants and help to absorb diffusing minority carriers that could otherwise create parasitic photocurrents in the native transistors.

In the design kit layout, the splitter and individual bias devices are covered with metal and surrounded by n-well guard bars to protect against parasitic photocurrents. Immunity to illumination is illustrated by the data in Fig. 14, which compares the master and splitter currents with and without chip illumination. The measured drain parasitic substrate photocurrent induced in the test transistor is also shown in this plot. We used chip illumination of 460 lux by uncovering the chip in our office, which is lit by fluorescent illumination. (In conjunction with $f/3$ optics, this *chip* illumination level corresponds to *scene* illumination of about 15 klux—about the ambient light on a cloudy day [20].) Direct

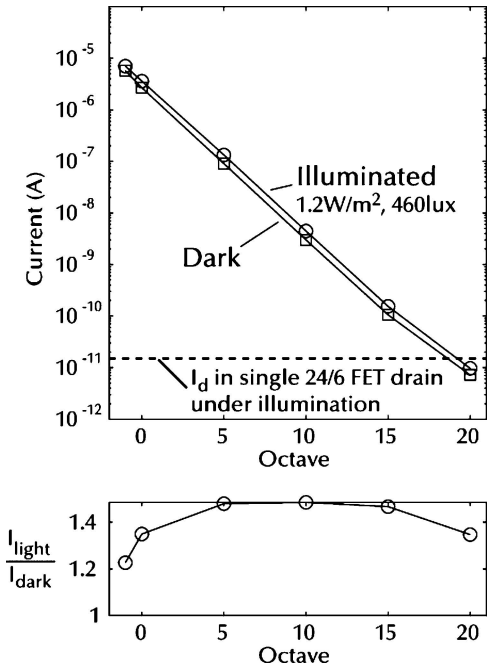

Fig. 14. Effect of illumination on generated currents.

illumination with 460 lux increases I_m by about 20%, and it has a maximal effect of a 50% increase in the splitter output currents that peaks for the middle splitter outputs. Therefore, this layout is suitable for use in imagers or focal plane arrays as long as slight changes in all currents can be tolerated. The n-well guard bars are at least 20 λ wide. We did not cover the master bias, thinking that the effect of parasitic photocurrent would be insignificant because the master current is much larger than the parasitic photocurrent from any one junction. We also neglected to place a guard bar at the start of the splitter where it abuts the master bias circuit. We believe that the 20% effect on the master bias arises from a direct effect of parasitic photocurrent, mostly on its large M_{n1} transistor, and that some of the remaining effect in the splitter comes from the lack of a guard bar at its starting side. Covering the master bias with metal would most likely greatly reduce these remaining effects.

The smallest current that can be generated is limited by the off-current of transistors. In the case of the measured 0.35 μ chip, this I_0 is about 3 pA. The minimum possible generated current is a few times I_0. The currents in the last stages of the splitter can approach the junction leakage currents, which are typically much smaller, but a bias generated from a copy of these currents is limited to the off-current because the bias transistors are presumably in saturation. Generation of still smaller currents would require techniques such as those outlined in [21, 22], where the splitter output current is copied by a source-biased current mirror whose gate input is shifted downward relative to the drain by a source-follower voltage shifter.

3.3.1. Splitter Nonidealities When not Using Unit Transistors In Section 2.2 we advised the use of unit transistors in the current splitter because we have observed nonidealities in non-R-2R splitters built without using unit transistors. The data in Fig. 15 illustrate these nonidealities, which were measured from a bias generator predating the design kit and built in a 1.6 μm process. This splitter used transistor W/L ratios of 24/81 for the lateral M_R and 24/10 for the transverse $M_{2R,}$ and was terminated with a 24/9 M_R. By following back from the termination of the splitter, it is easy to see that this should generate decade steps in splitter current (the ratio of the last two stages in the decade splitter is 9:1). Some measurements have two values because there were two separate biases (n-type and p-type) with the same current level. The measurements show that as the currents enter weak inversion, they are larger than predicted by the theory but that the nonideality is well modeled by the SPICE simulation. The following discussion is presented with the caveat that when we use unit transistors in an R-2R configuration, these nonidealities disappear both in simulation and in reality, and we are still not certain of their underlying device physics cause.

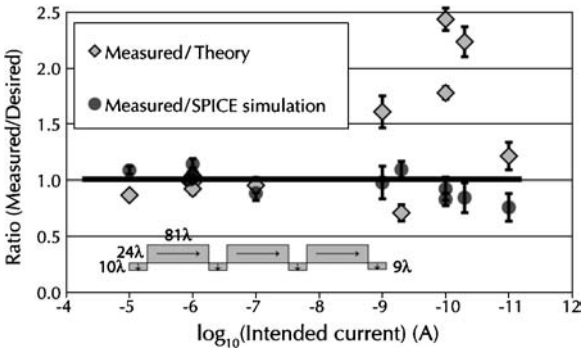

Fig. 15. Nonidealities of decade splitter built not using unit transistors. Points show ratio of measured current to ideal current and measured current to SPICE-simulation current for all instrumented splitter taps and individual biases. Error bars show variation over five chips. Inset shows current splitter transistor sizing.

The nonideality is predicted by SPICE BSIM3v3 simulations of the circuit, but it was hard to trace down its cause. The effect appears only in the parts of the current splitter operating in weak inversion and is not a channel-length modulation effect that only appears in subthreshold. Channel-length modulation would tend to have a greater effect on the shorter transverse M_{2R} transistors, which would have the opposite effect than what is observed. Any nonideality that increases current through a shorter transistor more than through a longer one has the wrong sign—it would make too much current split off in the early stages, leaving too little for later ones. In other words, such a nonideality would increase the "slope" of the splitter as viewed, like the curve in the top half of Fig. 12. What we actually observe is that this "slope" is decreased.

The nonideality turns out to be a complex mixture of threshold shift and short-channel transistor effects. It is illustrated in the SPICE simulation results shown in Fig. 16 of the final two stages of the decade splitter. Ideally the currents should be in the ratio 9:1. The plot shows the ratio as a function of injected current I_{in}. With long transistors the size of the ones we built, the ratio drops to about 6:1 in subthreshold, indicating that too much of the current is going into the branch with the longer transistors. This is a surprising result and a huge effect, representing a deviation of 50% from ideality. When we reduce the length of both transistors by a factor of 4, the nonideality flips over and we see the more familiar short-channel effect. In this case, the ratio is much larger than expected, about 20:1. In summary, these measurements and simulations suggest that it is dangerous to rely on length scaling even for very long transistors that are substantially wider than minimum width, and especially in subthreshold operation. The octave splitter does not have this problem, because it is built from unit transistors.

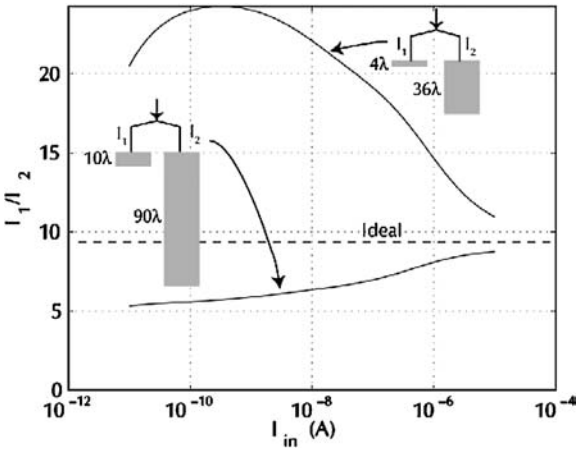

Fig. 16. SPICE simulations of the terminating decade splitter stage with two transistor length scales, as illustrated in the insets. Transistor width W = 24 λ.

3.4. Matching

Generated bias currents will be mismatched owing to inherent transistor mismatch. Although we have not extensively characterized mismatch, the data in Fig. 15 offers guidance. It was measured from a set of five chips, and the error bars indicate the chip-to-chip variation in measured currents. In the strong inversion region the variations are under 5%, while in the weak inversion region they grow to about 20%. These variations are probably acceptable in many applications. From another design fabricated in a 0.35 μ process, we have anecdotally observed that final bias currents match specified values with a variability of about 10% in strong inversion to 50% in weak inversion using transistor sizing as given in Table 1. Thus, it can be expected that matching is possible to within the resolution of the current splitter over the entire range.

3.5. Power Supply Sensitivity

Figure 17 shows measured sensitivity of the 0.35 μ bias generator to power supply voltage along with SPICE simulations. The master current was 5 μA, and the splitter outputs were measured by using the on-chip test transistor with a fixed drain-source voltage of 0.3 V. Each curve is normalized by its ideal value. There is a rather poor qualitative match between measured results and simulation except with regards to the power supply requirements. In this 3.3 V process, where the threshold voltages are $V_{Tn} = 0.49$ V and $V_{Tp} = -0.71$ V, the master bias requires a power supply voltage of about 1.75 V, and the splitter requires about 2.25 V to operate so that all transistors are correctly saturated. An additional 0.5 V is required to operate the splitter because of the unfortunate choice of splitter input current shown in the measured circuit (Fig. 7). Simulations of the proposed circuit, such as those discussed in Section 3.7, show that removing this mirror and directly supplying the

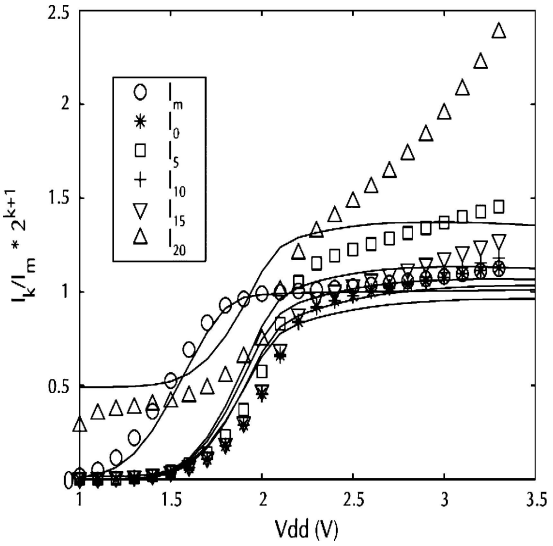

Fig. 17. Measured power supply sensitivity of the master bias current and splitter output currents. $I_m = 1\ \mu$A, and each curve is normalized by its ideal value. The test transistor used to monitor the current was held at 0.3 V drain voltage, and $R = 30\ \text{k}\Omega$ was used to make a master current of about 5 μA.

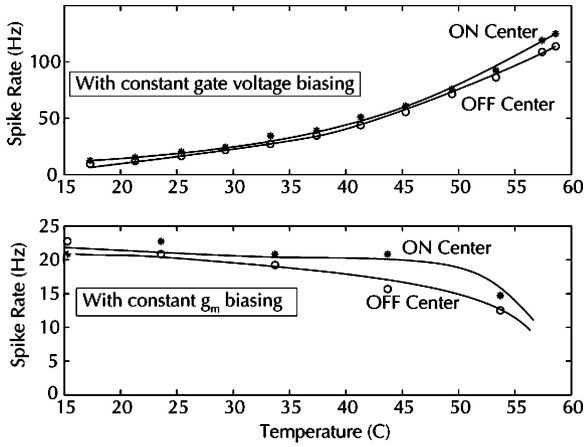

Fig. 18. Using static gate-voltage biases leads to exponential temperature dependence of current, as illustrated in these measurements from the system reported in [3]. Each plot shows the spike rate (rate of current-to-frequency converter) for different cells on the chip as a function of temperature. The top measurements were collected with static gate-voltage bias, whereas the bottom measurements used the bias generator circuits reported here. Using the bias generator leads to a much more stable operation. In fact, the cells slow down and finally stop firing as temperature increases; this arises from the increase in substrate leakage current on an oversized critical transistor.

splitter from the master bias significantly reduces the bias generator supply voltage requirements. We think that the poor match between simulation and measurement in Fig. 17 with regards to the supply sensitivity arises from leakage pathways in the bonding pads.

3.6. Temperature Dependence

Temperature dependence of biasing circuits is clearly important for real-world applications. Sometimes it is desirable to have a bias current that results in a constant g_m; at other times it may be desirable to have a constant current—for instance, when that current determines a slew rate or pulse width. The current generators presented here will act as PTATs when the master bias is operated in subthreshold, where $I_m = \log(M)U_T/R$, so that they fit well with circuits requiring constant g_m, but applications requiring temperature-independent constant current will have to employ different techniques. If a bias current determines the level of a current pulse and the pulse width is determined by the reciprocal of another bias current (e.g. as in a silicon model of a synapse), then using a PTAT generator will make the product of pulse height and width constant and result in a fixed-size charge packet. It should be kept in mind that PTAT sources vary their output by a factor of only 1.5 over a temperature range of -20 to 100°C, which would be acceptable for some applications.

To clearly show that constant gate-voltage biasing has very poor temperature sensitivity, the example data in Fig. 18 collected from the system described in [3] compares constant gate-voltage and constant-g_m biasing using the circuits described here. Over a range of 15°C to 50°C, the constant gate-voltage behavior changes by a factor of more than 5, whereas using the bias generator circuits results in a variation of only about 20%, most of which is due to parasitic substrate leakage.

We measured temperature dependence using a thermal wand (Temptronic Thermostream TP04100A; www.temptronic.com) to control the chip temperature. This thermal wand (or "elephant") is a benchtop device that blows heated or chilled air from a small tip that can be directed at a packaged chip. A thermocouple under the chip package measures the package temperature, and the thermal wand uses this measured temperature in a feedback loop to accurately set the package temperature. We found it difficult to explore temperatures near 0°C because water condensation from our standard (nondried) compressed air source created conductive paths that corrupted the low-current measurements;

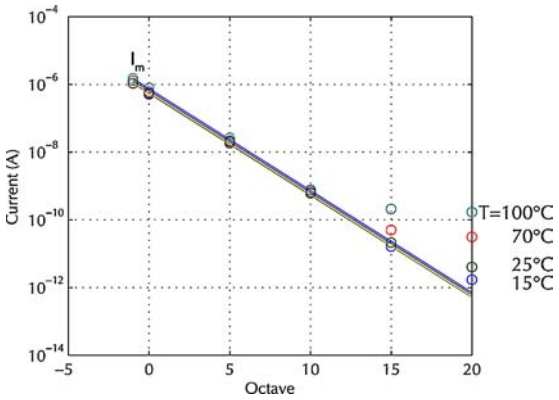

Fig. 19. Temperature measurements of bias generator core circuit. Plot shows master bias and splitter behavior at four temperatures. The straight lines show theoretical $I_m/2^{(\text{Octave}+1)}$ predictions of the current.

therefore, we varied the temperature from 100 to 15°C, which is a factor of 1.3 in absolute temperature. We controlled only the temperature of the chip, leaving the external resistor R at room temperature to make interpreting the data more straightforward.

The results of the measurements of the bias generator are shown in Fig. 19. The main observation is that increasing temperature slightly increases the master bias current but does not affect the splitter except to increase the minimum possible current. This increase in minimum current is consistent with the expected increase of the transistor off-current with temperature.

3.6.1. Influence of Temperature on Minimum Current Increasing temperature will increase transistor off-current through the increase of carrier density in the channel, thus increasing the minimum possible current that can be generated from the current splitter and decreasing the range of currents that can be generated. It is of interest to understand this phenomenon. We first discuss how this effect arises and how it is related to measured parameters such as the threshold voltage. We then show measurements of temperature effects and compare them with the theory.

The value of the off-current in (3.1) comes from a commonly accepted expression (e.g. [23–25]) for subthreshold current that includes the threshold voltage V_T:

$$I_0 = N_s U_T^2 \beta(T) \exp\left(\frac{-\kappa V_T}{U_T}\right) \tag{3.1}$$

Here N_s is a dimensionless preexponential that accounts in part for the concentration of carriers in the source. It is dimensionless because the rest of the expression has the familiar units βV^2. One factor of U_T accounts for the effective density of states in the channel at the source end. It depends on temperature because it arises from integration of a Fermi distribution over the (unknown) energy density of states in the channel, and higher temperature spreads the electrons over more energy states, increasing the effective number of states in the channel [26]. The other factor of U_T is part of the diffusion coefficient, which is given by the Einstein relation to the mobility $(kT/q)\mu(T)$. The mobility is a weak function of temperature $\mu(T) = \mu_r(T_r/T)^k$ where T_r is a reference temperature and k ranges from 1 to 2 [27]. Increasing temperature increases density of channel states and the diffusion coefficient, but these increases are nearly compensated by reduction of mobility.

The use of constant κ in (3.1) is inaccurate. When the channel is near flat-band, κ changes significantly with gate voltage and is also different than its value at threshold because the depletion capacitor is just starting to form and changes rapidly with surface potential. Nonetheless, for this analysis we will simply use the value of κ at threshold, which is very close to the value over most of the subthreshold range. In (3.1) we have ignored additional parameters, such as V_{off}, that appear in SPICE BSIM3v3 [24, 25] and that connect weak and strong inversion operation in a sensible way, but that are rendered meaningless for physical interpretation because they are subverted for curve fitting in automatic parameter extraction.

Temperature effects in the exponential last term in (3.1) are dominant; the term expresses the concentration of carriers at the source end of the channel as a function of the barrier potential, or "activation potential," $V_a = \kappa V_T$ between source and channel. This form can be misleading because if one assumes that V_a is a temperature-independent constant equal to κV_T, then the fit to reality is very poor. The threshold voltage decreases with temperature increase because the carriers are hotter and lower gate voltages are required for the same channel concentration. We will approach the problem of understanding temperature variation of the off-current by including the variation of V_T with temperature. We can compute the temperature sensitivity of (3.1) as follows, where all temperature dependence

that is not exponential is ignored:

$$\begin{aligned}\log I_o &= \text{const} - \frac{\kappa V_T}{U_T} \\ \frac{d \log I_0}{dT} &= \frac{dI_0/dT}{I_0} \\ &= \frac{\kappa}{U_T}\left(\frac{V_T}{T} - \frac{dV_T}{dT}\right) \end{aligned} \tag{3.2}$$

The first term in (3.2) represents the effect of average carrier energy for the barrier, while the second term represents the change in barrier height. Temperature sensitivity of the threshold voltage is well known [1, 23, 28] and is given by

$$\begin{aligned}\frac{dV_T}{dT} &= \left(\frac{1}{\kappa} - \frac{1}{2}\right)\frac{(2\phi_F - V_{BG})}{T} \\ &= \frac{(2\phi_F - V_{BG})}{T} \text{ when } \kappa = 0.66 \end{aligned} \tag{3.3}$$

where $2\phi_F = 2U_T \log(N_A/n_i(T))$ is the surface potential at threshold, and $V_{\text{BG}} = 1.206V$ is the band gap of silicon at 300°K [29]. The value of dV_T/dT ranges from -3 mV/°K to -1 mV/°K as depending on channel doping and oxide thickness [27]. $2\phi_F$ is the surface potential that brings the channel to a state of inversion that equals the channel doping N_A. The intrinsic concentration at $T = 300$°K is $n_i = 1.0 \cdot 10^{10}/\text{cm}^3$ (and not the commonly accepted value of $1.45 \cdot 10^{10}/\text{cm}^3$ that has clearly been shown to be incorrect by 45% [30]). Using (3.3) in (3.2), we can numerically evaluate (3.2) for representative values of threshold voltage and channel doping to obtain

$$\frac{\text{d log I}_0}{\text{dT}} = \frac{0.66}{25\text{ mV}}\left(\frac{500\text{ mV}}{300^\circ\text{ K}} - \frac{-2\text{mV}}{^\circ\text{K}}\right) \approx \frac{9\%}{^\circ\text{K}} \tag{3.4}$$

The two terms are comparable in (3.4), so both change in carrier energy and barrier height are significant. Measurements of current versus gate-source voltage at various temperatures for single transistors are shown in Fig. 20. Separate measurements of just the off-current are shown in the insets. The off-current very closely follows the classical form where $\log I_0$ is linear in $1/T$. The extracted activation potential is $V_a = 0.66$ V

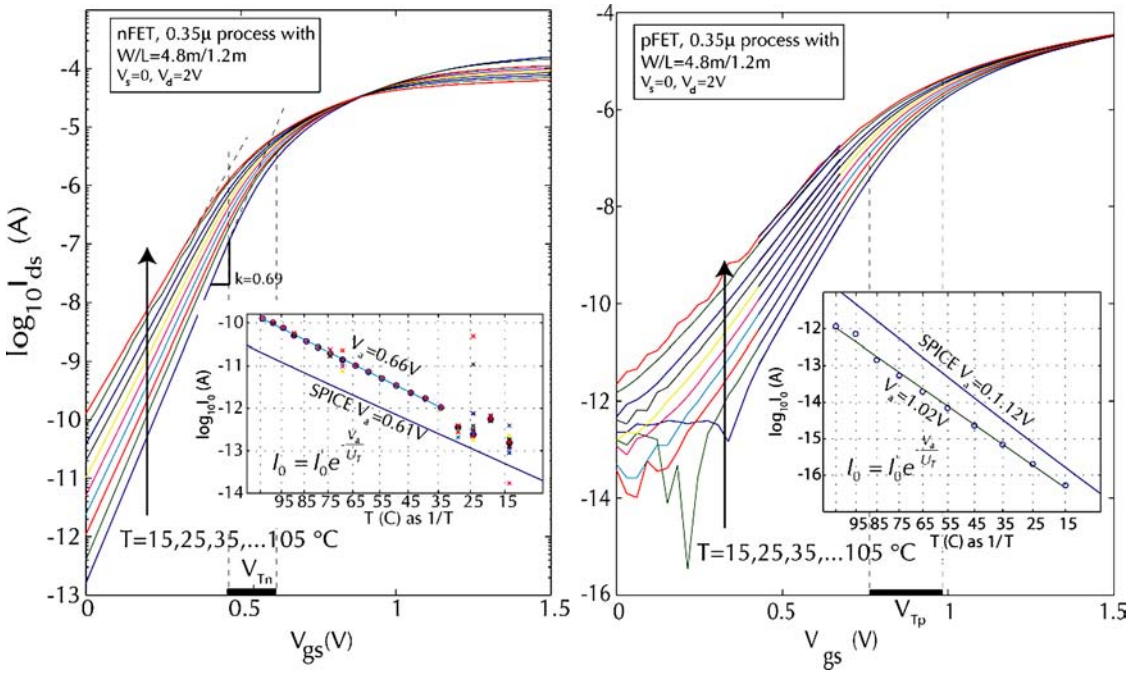

Fig. 20. Measured transistor temperature effects. The main plots shows I_{ds} vs. V_{gs} with temperature as a parameter. The inset plots show the log off-current log(I_0) as a function of 1/T along with the fitted activation potential V_a in (3.1). For the nFET the off-current was directly measured, whereas for the pFET it was inferred from the intercept of the fits to the I_{ds} vs. V_{gs} curves. SPICE BSIM3v3 simulation results for the off-current using vendor process parameters are also shown in the insets.

for the n-FET and $V_a = 1.01$ V for the p-FET. The value for V_a for the n-FET means that, at room temperature, the off-current increases about 9%/°K, or a doubling every 8°K. The n-FET off-current grows as large as 100 pA at T = 100°C, but at room temperature is about 1 pA in this process. A higher threshold voltage implies a smaller minimum current, but (3.2) further implies that a higher threshold voltage (larger V_a, smaller I_0) will result in a larger fractional variation of I_0 with temperature. We can see that, generally, the deeper in subthreshold the transistor operates, the larger the temperature sensitivity.

As the insets in Fig. 20 show, BSIM3v3 qualitatively models the temperature sensitivity of the off-current, but the quantitative correspondence is not very good: the magnitude of off-current differs by about a decade from the measured values, the activation potential differs by about 10%, and the discrepancies are in opposite directions for the two types of transistors.

3.6.2. PTAT Behavior of Master Bias Current It is also of interest to measure whether the master current I_m is truly a PTAT current. A measurement of a master bias circuit built in a 0.35 μ process is shown in Fig. 21 as the master bias current I_m and the voltage V_R versus temperature for weak and strong inversion operation. R was left at room temperature to simplify interpretation. The theory claims that in weak inversion operation, the master current is a PTAT current. Therefore, the line fitted to the measured data should intersect the origin. It almost intersects the origin, and V_R comes even closer. Simulations of temperature dependence, however, say that even for an above-threshold master bias current, the current should still be approximately PTAT. This is demonstrated in the lower part of Fig. 21, which shows the same measurements for strong inversion operation of the master bias circuit. In this case, the current rises more steeply with temperature, which is expected, since according to (2.3), $I_m \propto 1/\beta_n \propto T^k$ with k ranging from 1 to 2.

3.6.3. Influence of Temperature Dependence on Choice of Resistance Of course, temperature also affects passive components in a system. A fully integrated chip would include the resistor R on-chip. We would advise against the use of a diffused resistor when building an imager chip because it would collect diffusing minority carriers generated by light in the local substrate unless it was well protected. It is sometimes beneficial for several reasons to put part of the resistance on-chip and the rest off-chip. Putting part of R on-chip increases the stability of the master bias circuit because it degenerates the gain of M_{n1} as seen from the bonding pad. On-chip and off-chip resistance behavior with temperature can also sometimes be made to cancel each other. Figure 22 shows measurements of two resistors: a standard carbon axial through-hole resistor of 120 kΩ and an on-chip unsilicided polysilicon resistor of 40 kΩ with aspect ratio 2,500 λ/2λ built in a 1.6 μ process. They were measured to see how their temperature coefficients compare with each other and with the temperature coefficient of the PTAT master bias current. First, the magnitudes of the temperature coefficients of both types of resistors are much lower than PTAT. Second, the two types of resistors behave

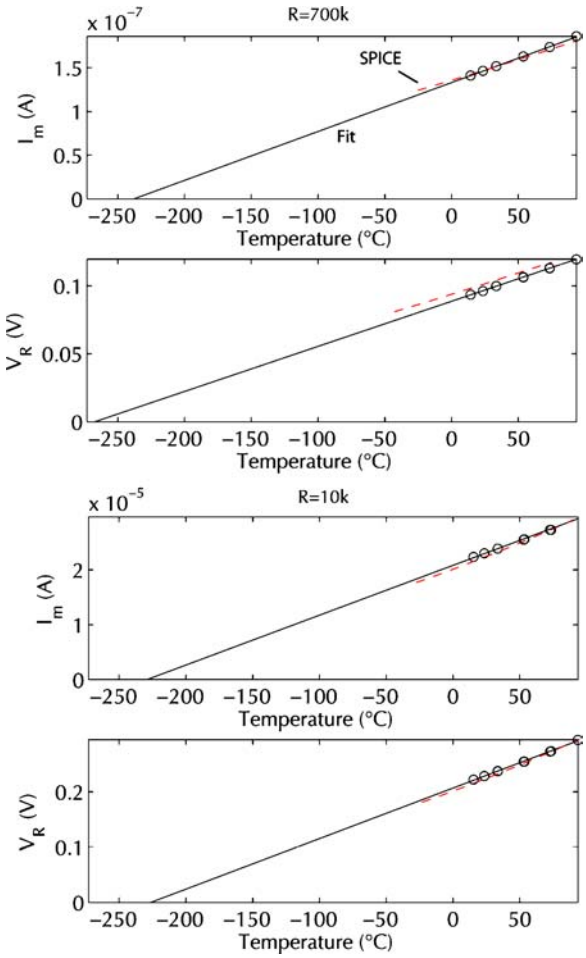

Fig. 21. Measurements of I_m and V_R plotted versus temperature to show absolute temperature dependence.

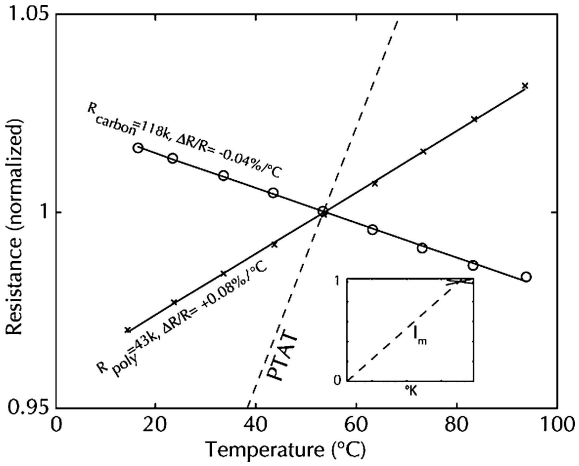

Fig. 22. Comparing temperature effects on a carbon resistor, polysilicon resistor, and master bias PTAT current. The inset shows on an absolute temperature scale the relative influence of temperature coefficient on the measured resistors and the PTAT master current.

oppositely: The carbon resistance goes down with temperature (as though the carbon grains come closer), whereas the polysilicon resistance goes up with temperature (as though the mobility decreases). The silicon and carbon resistors have roughly equal but opposite temperature coefficients, so they could in principle be balanced, but the exact values of temperature coefficient matter and may not be known ahead of time.

In a contemporary submicron process, the temperature coefficients of polysilicon and all diffused resistors are usually positive, probably because mobility decreases with temperature, although dedicated high-resistance polysilicon can have a negative temperature coefficient [31, 32]. For off-chip components, the temperature coefficients of surface-mount thin-film chip resistors are typically less than ± 200 ppm/°C [33]. Not worrying about it and just using a standard off-chip carbon resistor with a temperature coefficient of $-1{,}000$ ppm/°C will decrease the bias current away from PTAT only by about 10% over 100°C, and using a metal film off-chip resistor will decrease this effect even more.

3.7. Low-Voltage Bias Current Generator

Many designers are presently concerned with low-voltage operation because, as process technologies scale down, maximum supply voltages are becoming lower, and battery-powered low-power applications are of increasing importance. Since we have built these circuits only in 3.3 V or 5 V processes where low voltage is not a great concern, we cannot provide experimental results from a low-voltage process. The simplest modification to reduce the required supply voltage requirements is to remove all cascodes from the circuit. Removing cascodes could increase supply sensitivity, but as discussed in Section 3.4, if the master bias p-mirror can be built with high output resistance, this sensitivity can be minimized.

To study low-voltage operation limits of the present circuit, we removed all cascodes, leaving all remaining transistor scaling identical to the values in Table 1. We ran SPICE simulations of the bias generator using BSIM3v3 model parameters publicly available from MOSIS for a contemporary 0.18 μm non-epitaxial substrate mixed-signal RF process with a maximum power supply of 1.8 V (MOSIS-run T44E). In the master bias circuit we removed transistors M_{c1} and M_{c2} and tied the drain of M_{pd} to V_n. In the individual bias circuits we removed the n- and p-type cascodes. As mentioned in Section 2.2, the entire current splitter has an effective equivalent transistor W/L equal to a single one of its unit transistors. Hence, we expect that the bias generator requires slightly more supply voltage than $|V_{Tn}| + |V_{Tp}|$ to operate so that transistors that should be in saturation are saturated.

We modeled a bias generator that generates a master current of 1 μA and that has an octave current splitter of 20 stages. In Fig. 23 the top set of traces show the results of a DC sweep of the power supply Vdd. The master bias current and selected splitter currents are normalized by their ideal values and are plotted on a log scale. The data show that in this process, with approximately equal threshold voltages of 0.5 V for n- and p-type transistors, the bias generator is usable down to about 1.25 V. The last splitter output is about 10 times higher than it should be because the off-current in this process is substantially higher than the desired 1 pA splitter current. The two lower sets of traces in Fig. 23 show the results of a transient simulation with the supply voltage Vdd and the power-down input V_{pd} varied as shown.

3.7.1. Supply Sensitivity We also studied in simulation the power supply sensitivity of the master bias current, as discussed in Section 2.1.1, to understand the constraints on transistor sizing. Using the low-voltage version of the master bias circuit, we programmed

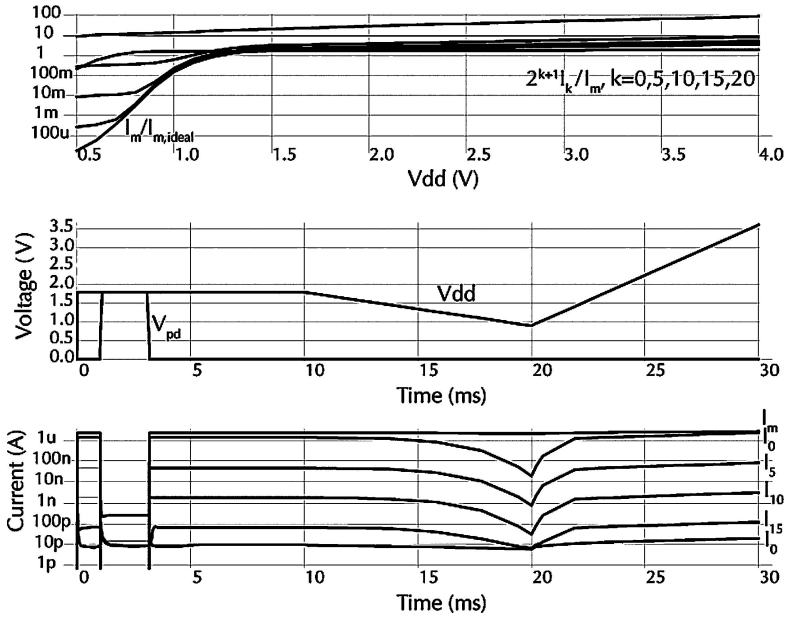

Fig. 23. Simulation results for a master bias and 20-stage current splitter from a 0.18 μ 1.8 V process. The top traces are from a DC sweep of the power supply voltage and the lower traces show a transient simulation where V_{dd} and power control V_{pd} were both varied.

a master current of about 400 nA and measured the supply sensitivity while varying the transistor lengths *LP* for the M_{p1}–M_{p2} mirror, *LN* for the M_{n1}–M_{n2} mirror, and the gain multiplier M. These simulation results are shown in Fig. 24. We assume that transistor drain resistance scales with transistor length, which is valid for a range of transistor lengths that are not too short (where short-channel effects dominate) or too long (where impact ionization dominates). These plots are useful for estimating the supply sensitivity. Using the transistor sizing in Table 1, supply sensitivity of the low-voltage master bias circuit in the 0.18 μ process is about 8%/volt. The benefit of using a large M is clearly visible in the lower plot. One might think that doubling the n-mirror output resistance would halve the supply sensitivity, but here it has only a small effect. Only when the n-mirror transistor lengths are made comparable to the p-mirror transistor lengths does the n-mirror transistor length begin to be significant. These observations are consistent with the discussion in Section 2.1.1.

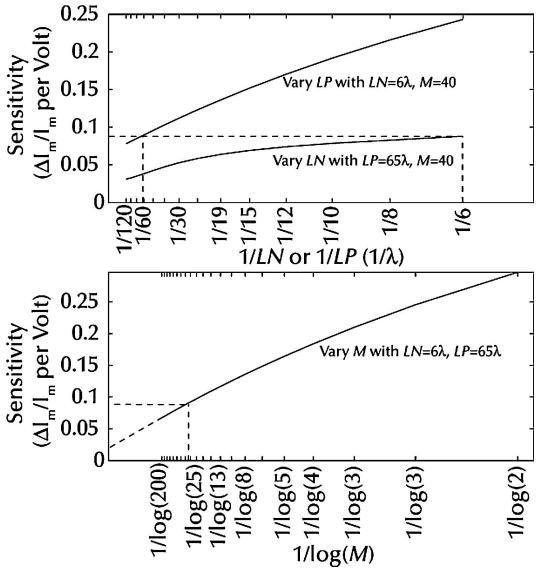

Fig. 24. Scaling of power supply sensitivity of the low-voltage version of the master bias current with circuit parameters. The dashed lines indicate default values, with LN = 6 λ, LP = 65 λ, and $M = 40$.

4. Design Kit

One of the authors (T.D.) has developed a design kit that makes it simple to construct a complete bias generator when using Tanner design tools (www.tanner.com/eda). A prerequisite for successful use of this kit is knowledge of the necessary bias currents. The process parameters and desired bias currents are specified in the schematic using parameter cells

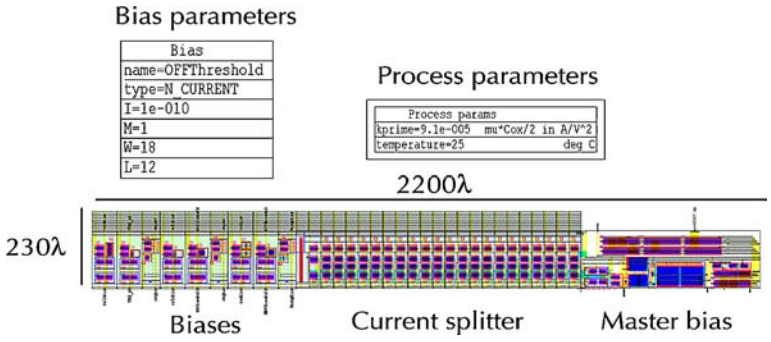

Fig. 25. Design kit use and generated layout. The user uses parameter cells to specify the bias currents and their types, as shown by the example cell, defines the process parameters, and then runs the compiler to generate the layout. The final layout has an area of 0.02 mm^2 in a 0.35 μ process.

such as the one shown in Fig. 25. A layout compiler parses the netlist from the schematic and computes the range of biases, the number of required splitter cells, and the master bias current. The required resistance R is estimated and reported. Bias current values can be chosen arbitrarily but can be programmed to only 2^{-k} of the master current; therefore, an arbitrarily chosen current will always be programmed to within 33% of the desired value. The compiler then builds the layout of the complete generator using a set of predefined layout and routing cells. The layout is clearly labeled for connection to the user's circuits, and a log file is written to show what has been generated. A variety of SPICE test bench files are provided to assist circuit simulation. The final generated layout has dimensions similar to the example shown in Fig. 25, which generates nine biases and occupies an area of about 0.02 mm^2 in a 0.35 μ process. Including 10 pF MOS bypass capacitors on each individual bias would approximately double the area.

The design kit layout cells are based on MOSIS [7] (www.mosis.org) scalable λ-based design rules, with double-metal, single-poly processes; thus the layout is compatible with any MOSIS CMOS process, including deep submicron processes. The cells are shielded by metal and have guard rings, so they are suitable for use in imager or focal plane arrays. No special techniques are used to minimize device mismatch except for the use of rather large geometry and the regularity of the current splitter. Two compiled bias generators built in a 0.35 μ process were the source of most of the data presented here. Readers are referred to www.ini.unizh.ch/~tobi/biasgen for the free kit.

Many industrial productions hold back wafers at various process steps during fabrication. With modifications, this design kit could generate layout so that a current for each bias could be determined by a metal mask. This capability would be useful if values for necessary bias currents are not known at the time the chip is first produced. Modifications that would allow desired currents to be programmed from a serial shift register are also possible but require more chip area and a digital interface to the chip. We have not developed this technology because we have been interested mostly in completely integrated chips that do not require external components.

5. Conclusion

The biasing circuits described in this paper enable designers with requirements for a wide range of bias currents to generate them systematically. Currents can be generated ranging from strong inversion to a few times the transistor off-current. Chips with known requirements for a wide range of bias currents can benefit significantly from the use of these circuits. Use of experimental chips by naïve users can be much easier to support if the parameters are truly fixed and the chips are not dependent on fine tuning of external parameters. The chip designers themselves will have the satisfaction of understanding the operation of the chip and knowing that it could probably be manufactured in quantity. The design kit described in Section 4 makes it simple to add these biasing circuits to any chip (especially a chip designed with MOSIS scalable rules) and provides reasonable assurance that they will function correctly the first time.

Acknowledgments

Oliver Landolt told us about current splitters and compound transistors. Discussions with Michael Godfrey about device physics were helpful, as were reviewers' detailed suggestions. The authors collaborated on this work at the Telluride Neuromorphic Engineering Workshop from 1999 to the present. This work was supported by the University of Zürich, ETH Zürich, the Institute of Neuromorphic Engineering and EC 5th framework project CAVIAR (IST-2001-34124).

References

1. P.R. Gray, P.J. Hurst, S.H. Lewis, and R.G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 4th edition. John Wiley, 2001.
2. A. van Schaik and S. Shamma, "A neuromorphic sound localizer," *Analog Integrated Circuits and Signal Processing*, vol. 39, pp. 267–273, 2004.
3. T. Delbruck and S.C. Liu, "A silicon early visual system as a model animal," *Vision Research*, vol. 44, pp. 2083–2089, 2004.
4. P. Lichtsteiner, T. Delbruck, and J. Kramer, "Improved ON/OFF temporally differentiating address-event imager," in *11th IEEE International Conference on Electronics, Circuits and Systems (ICECS2004)*, Tel Aviv, Israel, 2004, pp.
5. X. Arreguit, A. van Schaik, F. Baudin, M. Bidiville, and E. Raeber, "A CMOS motion detector system for pointing devices," *IEEE Journal of Solid State Circuits*, vol. 31, pp. 1916–1921, 1996.
6. T. Delbruck and A. van Schaik, "Bias current generators with wide dynamic range," in *2004 International Symposium on Circuits and Systems*, Vancouver, Canada, 2004, pp. I-337–tI340.
7. D. Cohen and G. Lewicki, "MOSIS-The ARPA silicon broker," in *Second Caltech Conference on VLSI*, Pasadena, 1981, pp. 29–44.
8. R.J. Widlar, "Design techniques for monolithic operational amplifiers," *IEEE Journal of Solid State Circuits*, vol. SC-4, pp. 184–191, 1969.
9. R.J. Widlar, "Some circuit design techniques for linear integrated circuits," *IEEE Transactions on Circuit Theory*, vol. CT-12, pp. 586–590, 1965.
10. E. Vittoz and J. Fellrath, "CMOS analog integrated circuits based on weak inversion operation," *IEEE Journal of Solid-State Circuits*, vol. 12, pp. 224–231, 1977.
11. T.H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, Cambridge Univesity Press, 1998.
12. B. Razavi, *Design of Analog CMOS Integrated Circuits*, McGraw-Hill Higher Education, 2000.
13. R.J. Baker, H.W. Li, and D.E. Boyce, *CMOS Circuit Design, Layout, and Simulation*, IEEE Press, 1998.
14. J.M. Steininger, "Understanding wide-band MOS transistors," *IEEE Circuits and Devices*, vol. 6, pp. 26–31, 1990.
15. S. Nicolson and K. Phang, "Improvements in biasing and compensation of CMOS opamps," in *2004 Intl. Symposium on Circuits and Systems (ISCAS 2004)*, Vancouver, Canada, 2004, pp. 665–668.
16. P. Lichtsteiner, "Bias current generator. Semester Project Report," Inst. of Neuroinformatics, Zurich 2001.
17. G.A. Rincon-Mora, *Voltage References: From Diodes to Precision High-Order Bandgap Circuits*, John Wiley, 2001.
18. V.V. Ivanov and I.M. Filanovsky, *Operational Amplifier Speed and Accuracy Improvement: Analog Circuit Design With Structural Methodology*, Kluwer Academic Publishers, 2004.
19. G. Bult and G. Geelen, "An inherently linear and compact MOST-only current division technique," *IEEE Journal of Solid-State Circuits*, vol. 27, pp. 1730–1735, 1992.
20. A. Rose, *Vision: Human and Electronic*, Plenum Press, New York, 1973.
21. B. Linares-Barranco and T. Serrano-Gotarredona, "On the design and characterization of femtoampere current mode circuits," *IEEE Journal of Solid-State Circuits*, vol. 38 pp. 1353–1363, 2003.
22. B. Linares-Barranco, T. Serrano-Gotarredona, R. Serrano-Gotarredona, and C. Serrano-Gotarredona, "Current mode techniques for sub-pico-ampere circuit design," *Analog Integrated Circuits and Signal Processing*, vol. 38, pp. 103–119, 2004.
23. E. Vittoz, "Electronics Laboratories Advanced Engineering Course on CMOS IC Design '02, Advanced Analog IC Design, Lausanne, Switzerland, August 26–30, 2002." EPFL-Lausanne, Switzerland, 2002.
24. Y. Cheng, M. Chan, K. Hui, M. Jeng, Z. Liu, J. Huang, K. Chen, J. Chen, R. Tu, P.K. Ko, and C. Hu, "BSIM3v3.2.2. MOSFET Model Users' Manual," Department of Electrical Engineering and Computer Sciences, University of California, Berkeley 1999.
25. M.-C. Jeng, "Design and modeling of deep-submicrometer MOSFETs," University of California Berkeley, Berkeley, PhD EDL M90/90, 1990.
26. C.A. Mead, *Analog VLSI and Neural Systems*, Addison-Wesley, Reading, MA, 1989.
27. S.M. Sze, *Physics of Semiconductor Devices*, 2nd edition, John Wiley & Sons, 1981.
28. P.E. Allen and D.R. Holberg, *CMOS Analog Circuit Design*, Harcourt Brace Jvanovich College Publishers, 1987.
29. M.A. Green, "Instrinsic concentration, effective densities of states, and effective mass in silicon," *Journal of Applied Physics*, vol. 67, pp. 2944–2954, 1990.
30. A.B. Sproul and M.A. Green, "Instrinsic carrier concentration and minority-carrier mobility of silicon from 77 to 300 K," *Journal of Applied Physics*, vol. 73, pp. 1214–1225, 1993.
31. W.A. Lane and G.T. Wrixon, "The design of thin-film polysilicon resistors for analog IC applications," *IEEE Transactions on Electron Devices*, vol. 36, pp. 738–744, 1989.
32. A. Hastings, *The Art of Analog Layout*. Prentice Hall, Upper Saddle River, NJ, 2001.
33. T. Williams, *The Circuit Designer's Companion*. Butterworth-Heinemann Ltd., Oxford, 1991.

Tobi Delbrück is a group leader at the Inst. of Neuroinformatics (INI), part of ETH Zurich and the University of Zurich, Switzerland. His main interest is in developing application-specific low-power vision sensor chips. In 1993 he graduated with a PhD in Computation and Neural Systems from Caltech, where he worked in Carver Mead's laboratory. He co-invented with Mead a widely-used adaptive photoreceptor circuit and invented the bump circuit. Subsequently he worked for several years for Arithmos, Synaptics, National Semiconductor, and Foveon, where he was one of the founding employees. In 1998 he moved to Switzerland to join INI. In 2002 he was lead developer of a tactile luminous floor used in INI's exhibit "Ada: Playful Intelligent Space" experienced by more than a half million visitors to the Swiss National Exhibition. He has been awarded 7 patents, and has written or coauthored 8 journal papers, 14 peer-reviewed conference papers, 4 book chapters, and 1 book.

André van Schaik obtained his M.Sc. in electronics from the University of Twente in 1990. In 1991–1994, he worked at CSEM, Neuchâtel, Switzerland, in the Advanced Research group of prof. Eric Vittoz. In this period he designed several analogue VLSI chips for perceptive tasks, some of which have been industrialised. A good example of such a chip is the artificial, motion detecting, retina in Logitech's Trackman Marble TM.

From 1994 until 1998, he was a research assistant and Ph.D. student with Prof. Vittoz at the Swiss Federal Institute of Technology in Lausanne (EPFL). The subject of his Ph.D. research was the development of analogue VLSI models of the auditory pathway. In 1998 he was a post-doctorate research fellow at the Auditory Neuroscience Laboratory at the University of Sydney.

In April 1999 he became a Senior Lecturer in Computer Engineering for the School of Electrical & Information Engineering at the University of Sydney. He is now a Reader in the same School and Head of the Computing and Audio Research Laboratory. His research interests include analogue VLSI, neuromorphic systems, wireless sensor networks, human sound localisation, and virtual reality audio systems.