A Calibration Technique for Very Low Current and Compact Tunable Neuromorphic Cells: Application to 5-bit 20-nA DACs

Juan A. Leñero-Bardallo, Teresa Serrano-Gotarredona, and Bernabé Linares-Barranco

Abstract—**Low current applications, like neuromorphic circuits, where operating currents can be as low as a few nanoamperes or less, suffer from huge transistor mismatches, resulting in around or less than 1-bit precisions. Recently, a neuromorphic programmable-kernel 2-D convolution chip has been reported where each pixel included two compact calibrated digital-to-analog converters (DACs) of 5-bit resolution, for currents down to picoamperes. Those DACs were based on MOS ladder structures, which although compact require $3N + 1$ unit transistors (N is the number of calibration bits). Here, we present a new calibration approach not based on ladders, but on individually calibratable current sources made with MOS transistors of digitally adjustable length, which require only N-sized transistors. The scheme includes a translinear circuit-based tuning scheme, which allows us to expand the operating range of the calibrated circuits with graceful precision degradation, over four decades of operating currents. Experimental results are provided for 5-bit resolution DACs operating at 20 nA using two different translinear tuning schemes. Maximum measured precision is 5.05 and 7.15 b, respectively, for the two DAC schemes.**

Index Terms—**Analog, calibration, mismatch, subthreshold.**

I. INTRODUCTION

OVER THE LAST 20 years, a vast amount of neuromorphic VLSI systems have been reported which usually consist of large arrays of special processing pixels. Since pixel size has to be of reduced size and power consumption, analog design techniques are used with transistors of small size operating with nanoamperes or less. This yields necessarily high mismatch. Although reported neuromorphic VLSI systems have revealed interesting, powerful, and fast information sensing and processing capabilities, they still have not evolved clearly to specific marketable products. One of the main reasons for this is the unavoidable excessive mismatch. For example, a 2.5 μm$\times$ 1.5 μm nMOS at 20 nA has a mismatch of $\sigma \approx 8\%$ (see [4, Fig. 3]). Defining LSB as 6σ, this yields a precision of $\log_2 6\sigma = 1$ b or, to design a 5-bit current source at 20 nA, one needs a 160 μm$\times$ 10 μm nMOS [4].

To keep mismatch low without increasing transistor sizes nor operating currents, the only known solution is calibration.
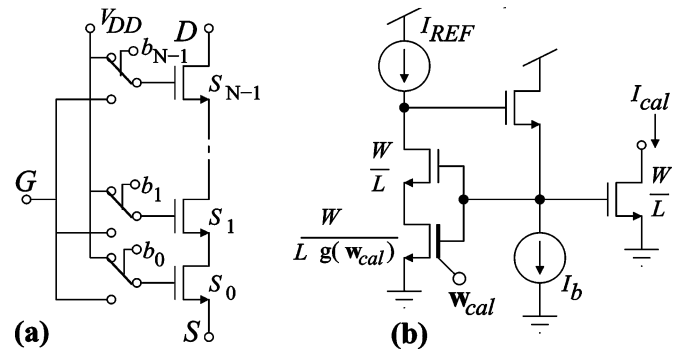
Fig. 1. (a) Schematics of proposed digitally controlled length MOS transistor. (b) Application to a calibration current source.

Some researchers have reported calibration techniques based on floating-gate MOS transistors [5], [6] in standard CMOS processes. However, these techniques require large area and special know-how. Recently, some neuromorphic systems with in-pixel RAM-based calibration techniques have been reported [1]–[3], [7], which exploit the use of compact current digital-to-analog converters (DACs) made with calibratable MOS ladder structures [8]. The drawback of this approach is that it uses a one-point calibration principle, which limits the final precision to 3 bits for nano amp currents and practical transistor sizes. In this paper, we present another principle with which we have achieved up to 7.15 b.

II. MOS WITH DIGITALLY ADJUSTABLE LENGTH

Previously reported in-pixel RAM-based calibration circuits [1], [2] were based on the use of MOS ladder structures [8].

With these structures, we obtained in the past [8] 4.4 bits at 2 μA with 16 5 μm$\times$ 5 μm nMOS transistors (total active area $= 400$ μm^2), for the same 0.35-μm CMOS technology we are using in the present work.

In this paper, we present a new approach to digitally adjust the equivalent size of a MOS transistor using a more compact circuitry. Fig. 1(a) shows the schematics of the new digi-MOS (digitally controlled-length MOS). There are N transistor segments between terminals D and S. Each segment is either enabled by connecting its gate to terminal G or disabled by connecting its gate to V_{DD} (for noise-sensitive applications, this node should be a low-noise V_{DD}). Transistor sizes can be, for example, $S_{N-1} = W/L$, $S_{N-2} = 2W/L$ ..., and $S_0 = (2^{N-1}) W/L$. This can be implemented physically by using unit transistors of size W/L
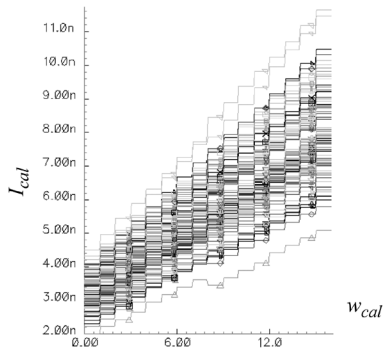
Fig. 2. Monte Carlo simulation (with 100 iterations) of the circuit in Fig. 1(b), using a 4-bit digitally controlled length MOS.
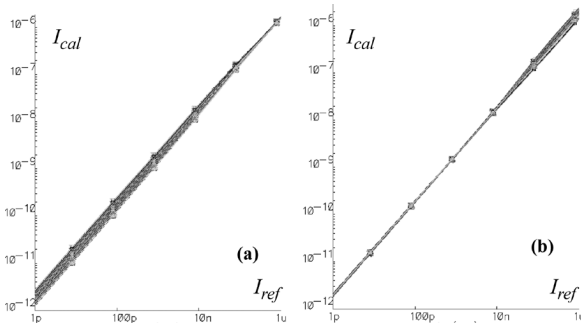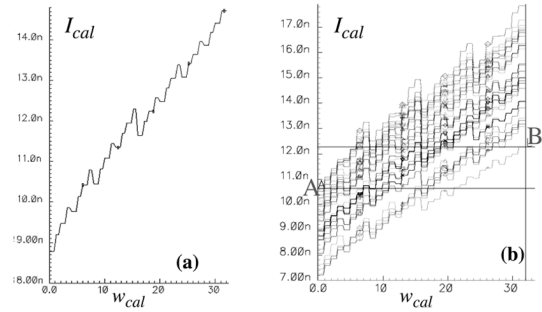

Fig. 4. Example simulation of a 5-bit digitally controlled length MOS with one transistor per segment and intentional down-steps. (a) Nominal mismatch-less simulation. (b) Monte Carlo simulation with 100 iterations.

Fig. 3. Monte Carlo simulation results for the circuit in Fig. 1(b) when sweeping I_{REF}. (a) Before calibration with $\mathbf{w} = 15$ for all Monte Carlo iterations. (b) After calibration with optimum $\mathbf{w}$ for each iteration.

(one for S_{N-1}, two in parallel for $S_{N-2}, \ldots 2^{N-1}$ in parallel for S_0). This way, each segment would be equivalent to a transistor of size $S_i = W/\left(L/2^{N-i-1}\right)$. The operation of this circuit is region-independent and can be analyzed by simple transistor series/parallel association [10]. Consequently, the digitally adjustable transistor in Fig. 1(a) would be equivalent to one of width W and digitally adjustable length $L_{\mathrm{eq}} = L \times g\left(\mathbf{w}_{\mathrm{cal}}\right)$ where $\mathbf{w}_{\mathrm{cal}} = b_{N-1}\ldots b_1 b_0$ and $g(\mathbf{w}_{\mathrm{cal}}) = \sum_{i=0}^{N-1} b_i 2^{1+i-N}$. This transistor can be used as part of a current mirror,[1] as shown in Fig. 1(b), to provide a calibration current $I_{\mathrm{cal}} = I_{REF} \times (g\left(\mathbf{w}_{\mathrm{cal}}\right) + 1)$. Fig. 2 shows the simulated stairs of I_{cal} as a function of $\mathbf{w}_{\mathrm{cal}}$ (using a 4-bit digitally controlled-length MOS) with $I_{\mathrm{REF}} = 3$ nA, using unit MOS sizes of 1 μm/4 μm, and models for a 0.35-μm standard CMOS process. Fig. 3(a) shows I_{cal} as function of I_{REF} before calibration, with $\mathbf{w}_{\mathrm{cal}} = 15$ for each of the 100 simulated Monte Carlo iterations. The mismatch at $I_{\mathrm{REF}} = 3$ nA is $\Delta I_{\mathrm{cal}}/I_{\mathrm{REF}} = 110\%$ and at $I_{\mathrm{REF}} = 1$ pA is 130%. Using the results in Fig. 2 (for $I_{\mathrm{REF}} = 3$ nA), one can compute for each Monte Carlo iteration the optimum value of $\mathbf{w}_{\mathrm{cal}}$ for minimum spread at I_{cal}. Once setting this optimum set of values for $\mathbf{w}_{\mathrm{cal}}$, the resulting I_{cal} as function of I_{REF} is shown in Fig. 3(b). Now, the mismatch at $I_{\mathrm{REF}} = 3$ nA has been reduced to 4% (4.6 b).

From a practical point of view, it is not efficient to follow the previous unit transistor-based sizing strategy, because the

number of unit transistors doubles with number of bits. In practice, it is more efficient to use one single transistor for each segment (bit) and adjust its size to have a similar effect. Furthermore, from a statistical point of view, we are not looking for nice uniform staircases, but for a (random) coverage. The maximum step heights will limit the final calibration capability. Therefore, it is important to minimize this maximum possible step height. To do this, we design the nominal stair case with some intentional "down-steps," so that when mismatch introduces random variations the extra redundancy (coverage) compensates for eventual large up-steps. Fig. 4, for example, shows Monte Carlo simulation results of a 5-bit structure that uses one single transistor per segment and has intentional down-steps. Simulated and fabricated transistor sizes are $\{2/3, 2/1.8, 2/1.8, 2/1, 2/0.7\}$. Consequently, total active area is now 16.6 μm^2.

III. TRANSLINEAR CIRCUITS FOR TUNING

The calibration technique shown in Fig. 1 requires to recalibrate all circuits when there is a global change in the operating current I_{REF}. In practice, it is desirable to allow a change in the operating current I_{REF} without requiring recalibration. Note that all transistors introduce mismatching and calibration compensates for the combination of all mismatches of all transistors. The mismatch introduced by each transistor is dependent on its operation current and bias conditions. To have calibration less sensitive to bias conditions one should use topologies that change bias conditions for as few transistors as possible. To achieve this, we use tunable translinear circuits, which will allow us to keep fixed bias currents for some transistors, including the digitally controlled-length ones. This is shown in Fig. 5. The circuitry comprised by broken lines is replicated once per pixel, but the rest is implemented only once at the periphery. Transistors M_1 to M_4 form a translinear loop, thus $I_{oi} = I_1 I_2/I_{i3}$. Local current I_{i3} is mirrored from the peripheral global current I_3, through a current mirror with a local digitally-controlled-length MOS. To achieve a factor-2 calibration range, we include two transistors in series for this current mirror output. One of fixed size $W/2L$ and the other calibratable. Consequently, $I_{i3} = I_3/\left(2 + g\left(\mathbf{w}_{\mathrm{cal}_i}\right)\right)$ and $I_{oi} = (I_1 I_2/I_3) \times (2 + g(\mathbf{w}_{\mathrm{cal}_i}))$. With this circuit, one can maintain (after calibration) constant currents I_3 (and I_{i3}) and I_1, while tuning I_2 globally to scale up or down all local currents I_{oi}.

---

[1]Here we use a subpico-ampere current mirror topology [11], since we want to use eventually I_{REF} values down to the *pico ampere* range [2].
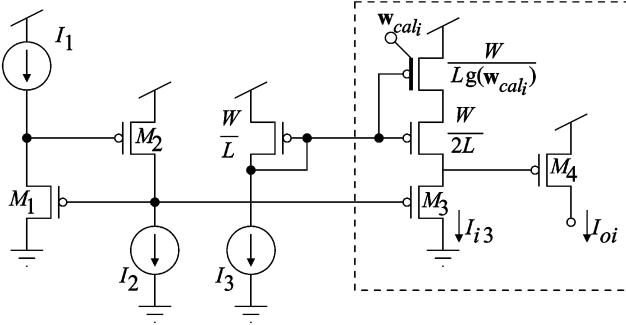
Fig. 5. Translinear circuit for tuning operating range of calibration circuit.
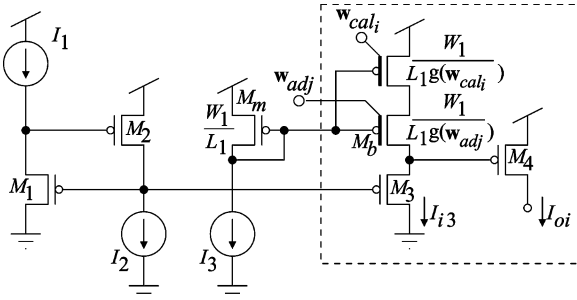
Fig. 6. First strategy for optimizing calibration range.

Fig. 7. Simulation results for first strategy. Simulated stairs for: (a) $\mathbf{w}_{\mathrm{adj}} = 31$, (b) for $\mathbf{w}_{\mathrm{adj}} = 16$, and (c) $\mathbf{w}_{\mathrm{adj}} = 0$. The vertical scale is the same for the three graphs.

IV. OPTIMIZING CALIBRATION RANGES

For calibration, the goal is to find the optimal horizontal line that cuts through all stairs and produces the minimum dispersion among all stairs. Note in Fig. 4(b) points "A" (top value of left side) and "B" (bottom of right side). If B is below A, the maximum dispersion after calibration will be high, because there will be no horizontal line cutting all stairs. If A is below B, it is possible to find for each stair a value close enough to the desired horizontal line cutting all stairs. For optimum calibration it is desired that A be close to B, so that final calibration words may spread over the whole range. The resulting relative position of points A and B depends on the resulting mismatch distribution of the array and the resulting process corner of the sample. One can design the nominal case to have A as close as possible to B, but then many fabricated samples will result with A higher than B, yielding poor calibration capability. On the other hand, if one designs the nominal case for A conservatively lower than B, then many samples will not take advantage of all of their bits for calibration, resulting in reduced calibration capability. Consequently, in practice, it will be desirable to be capable to adjust the relative positions of points A and B electronically. For this, we have implemented two different global optimization strategies.

In the first strategy, shown in Fig. 6, two digitally controlled-length transistors are used. One of them is adjusted locally, as in Fig. 5, but the other is adjusted globally. Thus, all gates of its transistor segments [see Fig. 1(a)] are shared by all pixels and controlled from the periphery. As a result, $I_{oi} = (I_1 I_2 / I_3) \times (g(\mathbf{w}_{\mathrm{adj}}) + g(\mathbf{w}_{\mathrm{cal}_i}))$. Fig. 7 shows the resulting simulated staircases for three different values of global control word $\mathbf{w}_{\mathrm{adj}}$. For one extreme [$\mathbf{w}_{\mathrm{adj}} = 31$, as in Fig. 7(a)], A is above B, and the array has very poor calibration capability. For the other extreme [$\mathbf{w}_{\mathrm{adj}} = 0$, as in Fig. 7(c)], A is at the bottom and the horizontal
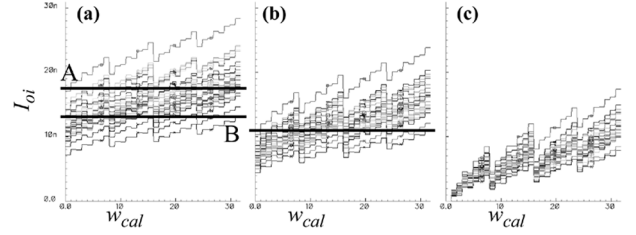
lines cut only a reduced range of the stairs, thus reducing significantly the available number of bits for calibration. The optimum solution is an intermediate one, in this case $\mathbf{w}_{\mathrm{adj}} = 20$ as in Fig. 7(b), which sets points A and B to be close. The optimum value of $\mathbf{w}_{\mathrm{adj}}$ is sample-dependent. Sizing of the extra transistor is not critical but should guarantee proper adjustment of A versus B for all process corners.

The second global optimization strategy is shown in Fig. 8. Here, the translinear circuit has been replicated twice, so that there are two of such translinear circuits in parallel. One of them uses local calibration through local digital control word $\mathbf{w}_{\mathrm{cal}_i}$. The other is adjusted globally and only the output transistor M_x of its translinear set is replicated once per pixel. This allows for a larger size of this transistor and, consequently, less mismatch. The purpose of the locally calibrated translinear circuit is to compensate for the mismatch at M_x. Fig. 9 shows simulation results for this circuit. In Fig. 9(a) all peripheral bias currents I_k and I_k' ($k = 1, 2, 3$) were set to 10 nA. The result is A being lower than B and a reduced range for the calibration words. Fig. 9(b) shows the contribution of only the bottom locally adjustable subcircuit (I_{i4} in Fig. 8). Note that, for $\mathbf{w}_{\mathrm{cal}_i} = 0$, the bottom circuit does not add current to I_{oi}. Consequently, in Fig. 9(a), for $\mathbf{w}_{\mathrm{cal}_i} = 0$, the mismatch is produced only by the upper M_x transistors. Note that this left part of the stairs will be fixed if peripheral currents I_k' are maintained fixed. The tuning strategy consists now in scaling peripheral currents I_k until obtaining the optimum situation shown in Fig. 9(c). In this case, we have set all $I_k = 4.5$ nA. After finding the optimum calibration words, the resulting operating point can be scaled by adjusting simultaneously only peripheral currents I_2 and I_2'.

V. EXPERIMENTAL RESULTS

A test prototype microchip was fabricated in a standard 0.35-μm CMOS process. Twenty 5-bit current DACs were fabricated. Ten of them used the first calibration range optimization strategy (Fig. 6), and the other ten used the second one (Fig. 8). We use the digi-MOS structures of Fig. 1(a) with five transistors of sizes $\{2/3, 2/1.8, 2/1.8, 2/1, 2/0.7\}$. Power supply was set to $V_{\mathrm{DD}} = 3.3$ V.

Each of the first ten DACs uses five replicas of the circuit in Fig. 6, one for each bit. The nominal output currents of each (I_{oi}) were adjusted to be binarily scaled. Consequently, at the periphery, we need five groups of current sources $\{I_1, I_2, I_3\}$ and five groups of transistors $\{M_1, M_2, M_m\}$, one for each bit. However, these five groups of peripheral current sources and transistors are shared by all ten DACs.
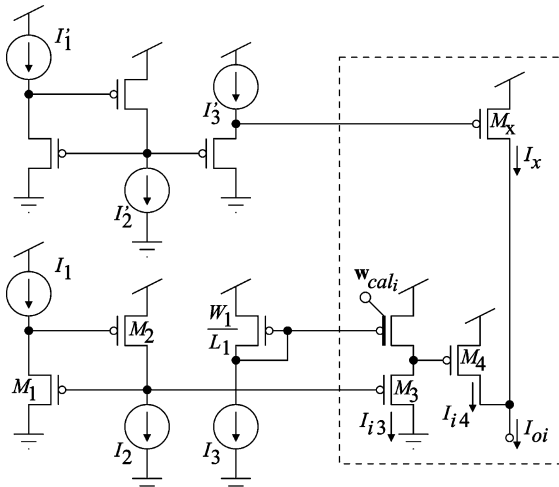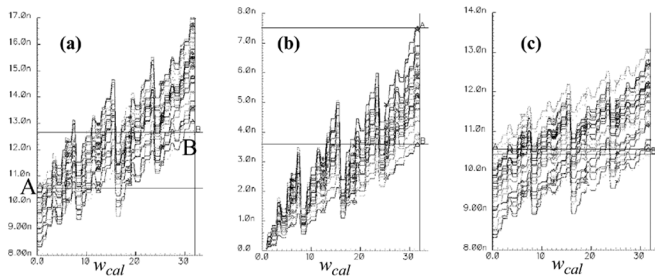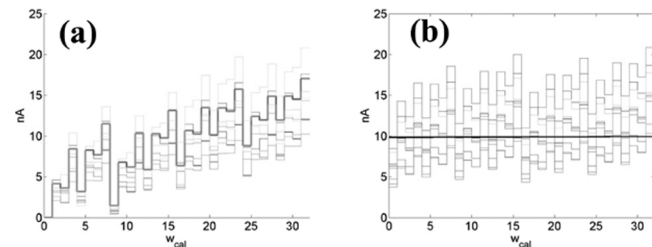
Fig. 8. Second strategy for optimizing calibration range.

Fig. 9. Simulation results for second strategy. (a) For all bias current equal to 10 nA. (b) Details of the bottom calibratable subcircuit I_{i4}. (c) Results for turning bias currents I_i down to 4.5 nA.

Fig. 10. Experimentally measured output currents for the circuit in Fig. 7: (a) for $w_{adj} = 0$ and (b) for optimum w_{adj}. The horizontal line in (b) is the target value, which is cut/touched by all ten traces.

Each of the second ten DACs uses five replicas of the circuit in Fig. 8. Again, for each of the ten DACs, the circuitry is replicated five times (one per bit), and the peripheral circuitry (outside broken lines in Fig. 8) is shared, per bit, by all ten DACs.

The area of the circuit layout inside broken lines is $18 \times 14 \mu m^2$ for Fig. 6 and $17 \times 14 \mu m^2$ for Fig. 8, excluding latches.

Fig. 10(a) shows the experimentally measured output currents for ten replicas of the circuit in Fig. 6, when setting $\mathbf{w}_{adj} = 0$. Peripheral bias currents were made equal to $I_1 = I_2 = I_3 = 10$ nA, and all calibration words $\mathbf{w}_{cal_i}$ $(i = 1, \dots 10)$ were swept simultaneously from 0 to 31. After repeating this measurement for all possible $\mathbf{w}_{adj}$ values, the optimum value for $\mathbf{w}_{adj}$ corresponds to the situation where the top left value is closest to the bottom right one. This case is shown in Fig. 10(b). At this point, we can obtain the ten optimum calibration words $\mathbf{w}_{cal_i}$ that render the minimum variation. The maximum output current spread obtained under
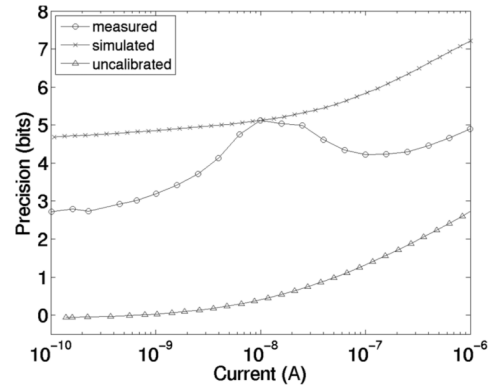
Fig. 11. Measured precision of calibratable and tunable current source with the approach of Fig. 6. Trace with circles: measured precision after calibration (with optimum $\mathbf{w}_{cal_i}$ for each of the ten current sources). Current sources were calibrated at 10 nA. Trace with triangles: measured precision before calibration ($\mathbf{w}_{cal_i} = 0$ for all current sources). Trace with crosses: precision after calibration, obtained through simulations.
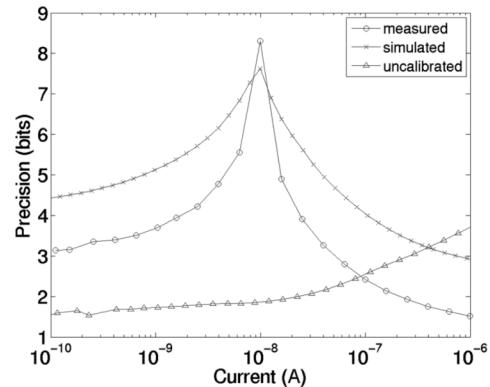
Fig. 12. Measured precision of calibratable and tunable current source with the approach of Fig. 8. Trace with circles: measured precision after calibration. Current sources were calibrated at 10 nA. Trace with triangles: measured precision before calibration ($\mathbf{w}_{cal_i} = 0$ for all current sources). Trace with crosses: precision after calibration, obtained through simulations.

these circumstances is $|\Delta I_{oi}|_{max} = 0.57$ nA, which corresponds to 5.7%, at a nominal current of $I_b = 10$ nA. If this were the current source controlled by the most significant bit of a current DAC (with 20-nA maximum range), it would limit the DAC precision to $-\ln\left(|\Delta I_{oi}|_{max}/2I_b\right)/\ln 2 = 5.13$ b. To verify how calibration degrades when changing bias conditions, we swept I_2 in Fig. 6 between 100 pA and 1 μA. The maximum current spread among all ten calibrated current sources is shown in the trace with circles in Fig. 11. The trace with triangles are measurements obtained before calibration ($\mathbf{w}_{cal_i} = 0$, for all i). We can see that the ten samples maintain a precision of 4 bits for currents above 3 nA. The horizontal axis is the average of I_{oi} among all ten samples. We also show in Fig. 11 the resulting precision after calibration obtained through simulations, shown with crosses. Note that it is over optimistic, except for the point at which calibration was done (10 nA). The reason is that usually circuit simulators do not model mismatch of slope factor (or gamma). Since the new length-controlled digi-MOS [Fig. 1(a)] is sensitive to body-effect, such mismatch affects performance, although it is not detected by most simulators.

In a similar way, Fig. 12 shows the measured precision before and after calibration of ten calibratable and tunable current
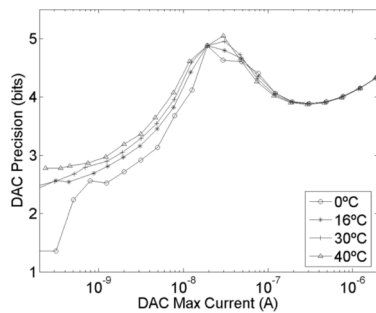
Fig. 13. Measured precision for the ten 5-bit DAC samples that use the first tuning strategy of Fig. 7. DACs were calibrated with MSB at 10 nA and at 16 °C. After calibration, precision is characterized sweeping operating current for different temperatures.
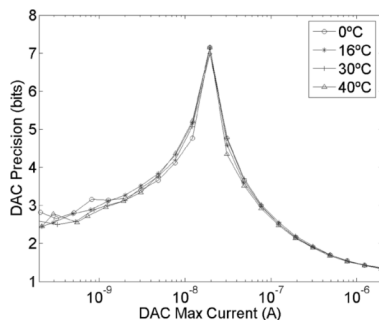

Fig. 14. Measured precision for the ten 5-bit DAC samples that use the second tuning strategy of Fig. 8. DACs were calibrated with MSB at 10 nA and at 16 °C. After calibration, precision is characterized sweeping operating current for different temperatures.

sources that follow the approach depicted in Fig. 8. Note that now the mismatch before calibration is less than that in Fig. 11. This is because now the area used by digitally controlled-length transistor M_b in Fig. 6 is available for transistor M_x in Fig. 8, which can be made larger. With the structure of Fig. 8, we obtain a much better precision at the calibration point (8.30 bits at 10 nA), but degrades rapidly, specially for high currents. The precision after calibration obtained by simulation is slightly pessimistic at the calibration point (7.63 bits at 10 nA), but it degrades optimistically as operating current departs from the calibration point (again because mismatch in slope factor (gamma) is not modeled).

Figs. 11 and 12 show the matching precision among ten current sources calibrated at 10 nA. Now we use five of these sources, calibrated at $\{10, 5, 2.5, 1.25, 0.625\}$ nA, to build a 5-bit current DAC. The matching precision obtained among the ten fabricated DACs is shown in Fig. 13 (for the tuning scheme of Fig. 6) and in Fig. 14 (for the tuning scheme of Fig. 8). The DACs were calibrated at 16 °C, and the figures also illustrate the DACs behavior when temperature is changed between 0 °C and

40 °C. We can see that the effect of temperature is not severe for the lower current range, while for higher currents the DACs are almost insensitive to temperature variations.

VI. CONCLUSION

A new compact calibration scheme for current sources is presented. The approach is illustrated for current sources operating in the nano ampere range. Two tuning schemes are proposed for sweeping the operating range over four decades. The first one achieves less precision at the calibration point but it degrades more gracefully as the operating current is increased (it shows over 4-bit precision for currents larger than 10 nA). The second one achieves higher precision at the calibration point but precision degrades more as current increases (4 bits is achieved only for currents between 8–40 nA). Test prototypes have been fabricated and extensively tested and characterized. As an example application, current DACs of 5-bit resolution and 20-nA range have been fabricated and characterized.

REFERENCES

[1] J. Costas, T. Serrano-Gotarredona, R. Serrano-Gotarredona, and B. Linares-Barranco, "A spatial contrast retina with on-chip calibration for neuromorphic spike-based AER vision systems," *IEEE Trans. Circuits Syst. 1, Reg. Papers*, vol. 54, no. 7, pp. 1444–1458, Jul. 2007.

[2] R. Serrano-Gotarredona, T. Serrano-Gotarredona, A. Acosta-Jimenez, and B. Linares-Barranco, "A neuromorphic cortical-layer microchip for spike-based event processing vision systems," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 53, no. 12, pp. 2548–2566, Dec. 2006.

[3] R. J. Kier, J. C. Ames, R. D. Beer, and R. R. Harrison, "Design and implementation of multipattern generators in analog VLSI," *IEEE Trans. Neural Netw.*, vol. 17, no. 4, pp. 1025–1038, Jul. 2006.

[4] R. Serrano-Gotarredona, L. Camuñas-Mesa, T. Serrano-Gotarredona, J. A. Leñero-Bardallo, and B. Linares-Barranco, "The stochastic I-pot: A circuit block for programming bias currents," *IEEE Circuits Syst. I, Reg. Papers*, vol. 54, no. 8, pp. 1760–1764, Aug. 2007.

[5] R. R. Harrison, J. A. Bragg, P. Hasler, B. A. Minch, and S. P. Deweerth, "A CMOS programmable analog memory-cell array using floating-gate circuits," *IEEE Circuits Syst. II, Analog Digit. Signal Process.*, vol. 48, no. 1, pp. 4–11, Jan. 2001.

[6] Y. L. Wong, M. H. Cohen, and P. A. Abshire, "128 × 128 floating gate imager with self-adapting fixed pattern noise reduction," in *Proc. ISCAS*, 2005, vol. 5, pp. 5314–5317.

[7] S. Shah and S. Collins, "A temperature independent trimmable current source," in *Proc. ISCAS*, May 2002, vol. 1, pp. 713–716.

[8] B. Linares-Barranco, T. Serrano-Gotarredona, and R. Serrano-Gotarredona, "Compact low-power calibration mini-DACs for neural massive arrays with programmable weights," *IEEE Trans. Neural Netw.*, vol. 14, no. 5, pp. 1207–1216, Sep. 2003.

[9] K. Bult and J. G. M. Geelen, "An inherently linear and compact MOST-only current division technique," *IEEE J. Solid-State Circuits*, vol. 27, no. 6, pp. 1730–1735, Dec. 1992.

[10] C. Galup-Montoro, M. C. Schneider, and I. J. B. Loss, "Series-parallel association of FETs for high gain and high frequency applications," *IEEE J. Solid-State Circuits*, vol. 29, no. 9, pp. 1094–1101, Sep, 1994.

[11] B. Linares-Barranco and T. Serrano-Gotarredona, "On the design and characterization of femtoampere current-mode circuits," *IEEE J. Solid-State Circuits*, vol. 38, no. 10, pp. 1353–1363, Oct. 2003.