Compact Calibration Circuit for Large Neuromorphic Arrays

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

Instituto de Microelectrónica de Sevilla, IMSE-CNM, CSIC and Universidad de Sevilla.

E-mail: bernabe@imse.cnm.es

Abstract

Low current applications, like neuromorphic circuits, where operating currents can be as low as few *nano* amps or less, suffer from huge transistor mismatches, resulting in around or less than 1-bit precision. Here we present a new calibration approach based on individually calibratable current sources made with MOS transistors of digitally adjustable length, which require only *N* unit transistors. The scheme includes a translinear circuit based tuning scheme, which allows to expand the operating range of the calibrated circuits with graceful precision degradation, over 4 decades of operating currents. Experimental results are provided for 5-bit resolution DACs operating at 20*nA*.

I. Introduction

Over the last 20 years a vast amount of neuromorphic VLSI systems have been reported ([1]-[2], just to mention a few) 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 *nano* amps 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 that plagues most of the so far reported neuromorphic chips [1]-[2].

To keep mismatch low without increasing transistor sizes nor operating currents, the only known solution is calibration. Some researchers have reported calibration techniques based on floating-gate MOS transistors [3]-[4] in standard CMOS processes. However, these techniques require a special know-how, which makes difficult the path towards marketable products. Recently, some neuromorphic systems with in-pixel RAM based calibration techniques have been reported [5]-[6], which exploit the use of compact current DACs made with calibratable MOS ladder structures [7]. The drawback of this approach is that it uses a one-point calibration principle, which limits the final precision to 3-bit for practical transistor currents and sizes. In this paper we present another way of implementing a digitally adjustable MOS, much more compact than a MOS ladder structure, which makes viable a multi-point calibration. Also, we introduce some extra translinear circuitry which enables to sweep operating currents of the calibrated current sources without requiring recalibration. This way, circuit precision degrades smoothly when changing operating currents.

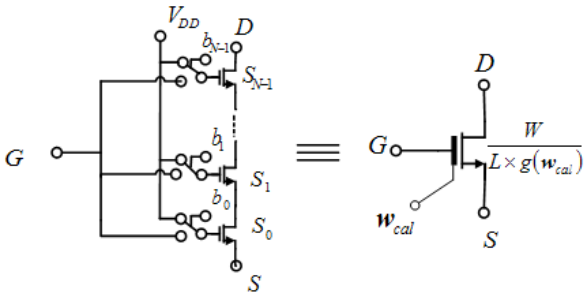

Fig. 1: New calibration circuit. (a) Circuits schematics and (b) symbol.

II. MOS with digitally adjustable length

Previously reported in-pixel RAM based calibration circuits [5]-[6] were based on the use of MOS ladder structures [7]. In this paper we present a new approach to digitally adjust the size of a MOS transistor using a compact circuitry. Fig. 1.a shows the schematics and Fig. 1.b shows the symbol of the proposed circuit. As it can be seen in Fig. 1.a, the circuit is composed of 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}. Transistors can be designed for example with sizes: $S_{N-1} = W/L$,$S_{N-2} = 2W/L$,....,$S_0 = (2^{N-1})W/L$. This can be implemented physically by using transistors of size W/L (one for S_{N-1}, two in parallel for S_{N-2}, ... 2^{N-1} in parallel for S_0). This way, each segment would be equivalent to a transistor of size $S_i = W/(L/2^{N-i-1})$. Consequently, the digitally adjustable transistor in Fig. 1(a) would be equivalent to one of width W whose length is digitally adjustable through the N-bit digital word $\mathbf{w}_{cal} = \{b_{N-1}b_{N-2}\ldots b_2 b_1 b_0\}$. The equivalent length is given by $L_{eq} = L \times g(\mathbf{w}_{cal})$, where

$$g(\mathbf{w}_{cal}) = \sum_{i=0}^{N-1} \frac{b_i}{2^{N-1-i}} \qquad (1)$$

Digital word $\mathbf{w}_{cal}$ is stored on in-pixel static RAM cells at startup. The optimum calibration words are obtained from a calibration procedure followed the first time the chip is used [5]-[6].

This transistor can be used as part of a current mirror[1], as shown in Fig. 2, to provide a calibration current

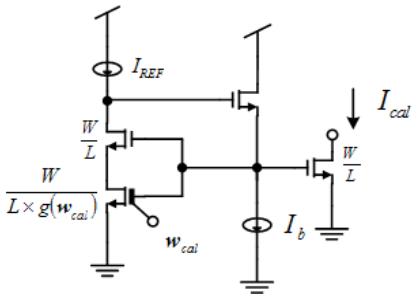

Fig. 2: Application to a calibration current source.

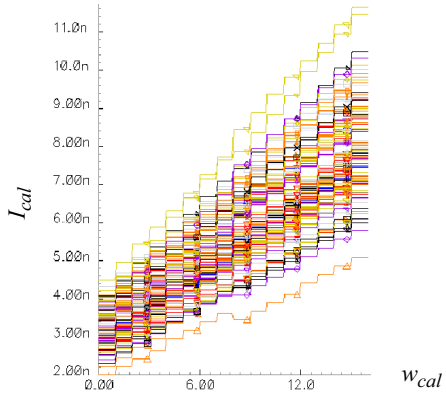

Fig. 3: Montecarlo simulation (with 100 iterations) of the circuit in Fig. 2, using a 4-bit digitally controlled length MOS.

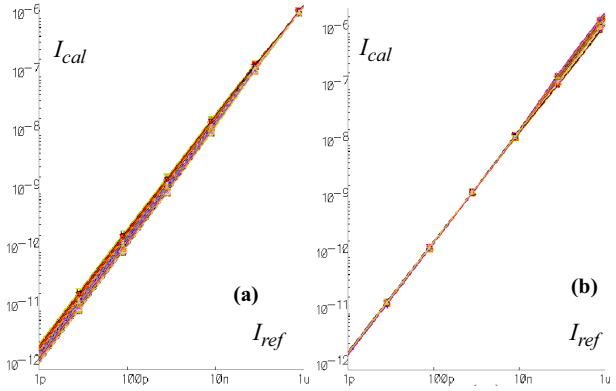

Fig. 4: Montecarlo simulation results for the circuit in Fig. 2(b) when sweeping I_{REF}. (a) Before calibration with w_{cal}=15 for all Montecarlo iterations. (b) After calibration with optimum w_{cal} for each iteration.

$I_{cal} = I_{REF} \times (g(\mathbf{w}_{cal}) + 1)$. Fig. 3 shows the simulated stairs of I_{cal} as function of $\mathbf{w}_{cal}$ (using a 4-bit digitally-controlled-length MOS) with $I_{REF} = 3nA$, using unit MOS sizes of $1\mu m/4\mu m$, and models for a $0.35\mu m$ standard CMOS process. Fig. 4(a) shows I_{cal} as function of I_{REF} before calibration, with $\mathbf{w}_{cal} = 15$ for each of the 100 simulated Montecarlo iterations. The mismatch at $I_{REF} = 3nA$ is $\Delta I_{cal}/I_{REF} = 110\%$ and at $I_{REF} = 1pA$ is 130%. Using the results in Fig. 3 (for $I_{REF} = 3nA$), one can compute for each Montecarlo iteration the optimum value of $\mathbf{w}_{cal}$ for minimum spread at I_{cal}. After setting this optimum set of values for $\mathbf{w}_{cal}$, the resulting I_{cal} as function of I_{REF} is shown in Fig. 4(b). Now the mismatch at $I_{REF} = 3nA$ has been reduced to 4% (4.6 bits).

From a practical point of view, it is not efficient to follow the previous unit transistor based sizing strategy. Note that transistor of size S_{N-1} is the most critical for mismatch (since this segment contributes the largest L). However, it uses only one unit transistor, while transistor of size S_0 is the least critical for mismatch and uses 2^{N-1} units. In practice it is more efficient to use one single transistor for each MOS segment and adjust its size to have a similar effect. Furthermore, from a statistical point of view, since mismatch plays an important role, the steps of the final stair cases will not be all equal. The maximum step heights will limit the final calibration capability. Therefore, it is important to minimize this maximum possible step height. To do this, the nominal stair case should be designed with some intentional 'down-steps', so that when mismatch introduces random variations the extra redundancy compensates for eventual large up-steps. The intentional 'down-steps' are introduced by adjusting the transistor geometries so that periodically a current division factor lower than 2 is implemented. Fig. 5, for example, shows Montecarlo simulation results of a 5-bit structure that uses one single transistor per segment and has intentional down-steps. Simulated transistor sizes were {2/3, 2/1.8, 2/1.4, 2/1, 2/0.7}.

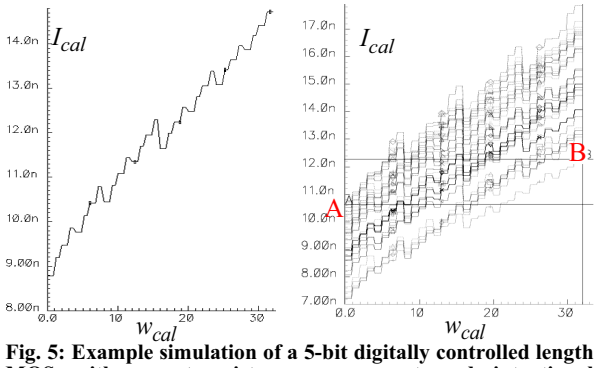

Fig. 5: Example simulation of a 5-bit digitally controlled length MOS with one transistor per segment and intentional down-steps. (a) Nominal mismatch-less simulation. (b) Monte Carlo simulation with 100 iterations.

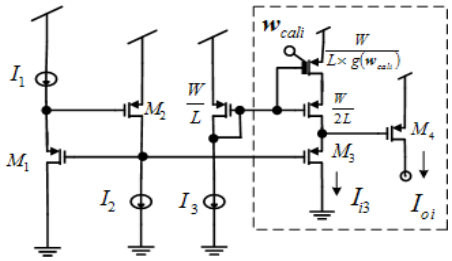

Fig. 6: Translinear circuit for tuning operating range of calibration circuit.

III. Translinear circuits for tuning

The calibration technique shown in Fig. 2 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

1. Here we use a sub-pico-ampere current mirror topology [8], since we want to use eventually I_{REF} values down to the *pico ampere* range [6].

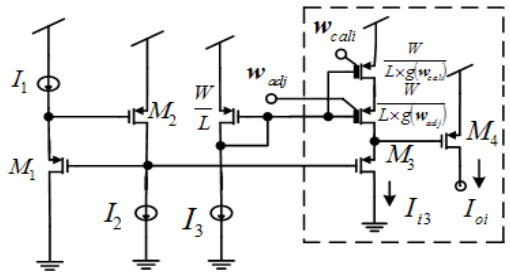

Fig. 7: Strategy for optimizing calibration range.

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 introduce 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. 6. 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, $I_{i3} = I_3/(2+g(\mathbf{w}_{cal_i}))$ and

$$I_{oi} = \frac{I_1 I_2}{I_3}(2+g(\mathbf{w}_{cal_i})) \qquad (2)$$

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}

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. 5(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 their bits for calibration, resulting in reduced calibration capability. Consequently, in practice, it will be desirable to be capable to adjust the relative

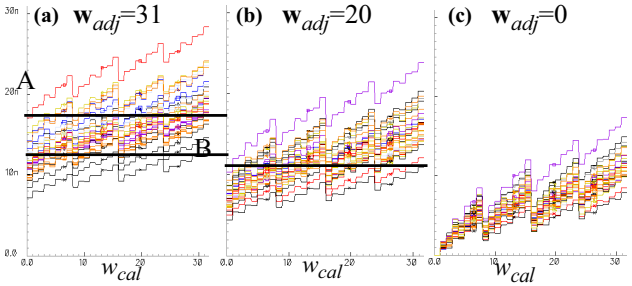

Fig. 8: Simulation results. (a) Simulated stairs for w_{adj}=31, (b) for w_{adj}=16, (c) and w_{adj}=0. Vertical scale is the same for the three graphs.

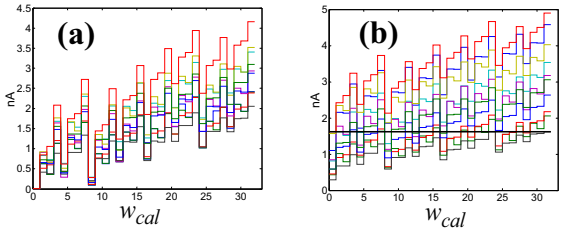

Fig. 9: Experimentally measured output currents for the circuit in Fig. 7. (a) for w_{adj} = 0, (b) for optimum w_{adj}. The horizontal line in (b) is the target value, which is cut/touched by all 10 traces.

positions of points 'A' and 'B' electronically. For this, we have implemented a global optimization strategy.

In the circuit, shown in Fig. 7, two digitally-controlled-length transistors are used. One of them is adjusted locally, as in Fig. 6, but the other is adjusted globally. Thus all gates of its transistor segments (see Fig. 2(a)) are shared by all pixels and controlled from the periphery. As a result,

$$I_{oi} = \frac{I_1 I_2}{I_3}(g(\mathbf{w}_{adj}) + g(\mathbf{w}_{cal_i})) \qquad (3)$$

Fig. 8 shows the resulting simulated staircases for 3 different values of global control word $\mathbf{w}_{adj}$. For one extreme ($\mathbf{w}_{adj}$ = 31, as in Fig. 8(a)) 'A' is above 'B', and the array has very poor calibration capability. For the other extreme ($\mathbf{w}_{adj}$ = 0, as in Fig. 8(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 $\mathbf{w}_{adj}$ = 20 as in Fig. 8(b), which sets points 'A' and 'B' to be close. The optimum value of $\mathbf{w}_{adj}$ is sample dependent.

V. Experimental results

A test prototype microchip was fabricated in a standard 0.35μm CMOS process. Ten 5-bit current DACs with area $18\times 14\mu m^2$ were fabricated.

Each of the ten DACs uses five replicas of the circuit in Fig. 7, 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.

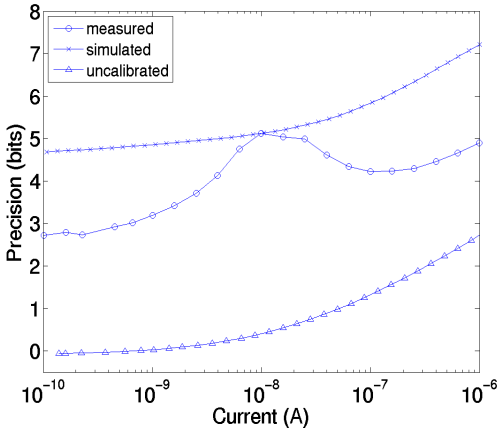

Fig. 10: Measured precision of calibratable and tunable current source with the approach of Fig. 7. Trace with circles: measured precision after calibration (with optimum $\mathbf{w}_{cal_i}$ for each of the ten current sources). Current sources were calibrated at $10nA$. Trace with triangles: measured precision before calibration ($\mathbf{w}_{cal_i} = 0$ for all current sources). Trace with crosses: precision after calibration, obtained through simulations.

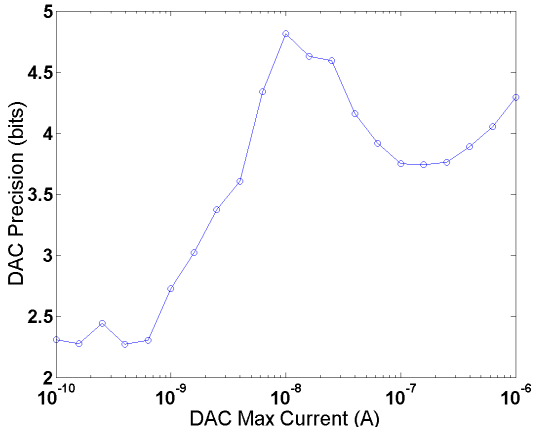

Fig. 11: Measured Precision for the ten 5-bit DAC samples that use the tuning strategy of Fig. 7. DACs were calibrated with MSB at $10nA$ and at 16ºC.

Fig. 9(a) shows the experimentally measured output currents for ten replicas of the circuit in Fig. 7, when setting $\mathbf{w}_{adj} = 0$. Peripheral bias currents were made equal to $I_1 = I_2 = I_3 = 10nA$, and all calibration words $\mathbf{w}_{cal_i}$ ($i = 1, \ldots 10$) 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. 9(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 these circumstances is $|\Delta I_{oi}|_{max} = 0.57nA$, which corresponds to 5.7%, at a nominal current of $I_b = 10nA$. If this were the current source controlled by the most significant bit of a current DAC (with $20nA$ maximum range), it would limit the DAC precision to $-\ln(|\Delta I_{oi}|_{max}/2I_b)/\ln 2 = 5.13$bits. To verify how calibration degrades when changing bias conditions, we swept I_2 in Fig. 7 between $100pA$ and $1\mu A$. The maximum current spread among all 10 calibrated current sources is shown in the trace with circles in Fig. 10. We can see the 10 samples maintain a precision of 4 bits for currents above $3nA$ after calibration. The trace with triangles are measurements obtained before calibration ($\mathbf{w}_{cal_i} = 0$, for all i). . Horizontal axis is average of I_{oi} among all ten samples. We also show in Fig. 10 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 ($10nA$). The reason is that in this particular circuit (Fig. 7) calibration degrades because of mismatch in MOS transistor slope factor. Mismatch in this parameter is not modelled in our simulator.

Fig. 11 shows the matching precision among 10 current sources calibrated at $10nA$ at 16ºC. Now we use 5 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 10 fabricated DACs is shown in Fig. 11 .

VI. Conclusions

A new compact calibration scheme for current sources is presented. The approach is illustrated for current sources operating in the nano ampere range. A tuning scheme has been proposed for sweeping the operating range over four decades. The scheme achieves high precision at the calibration point and degrades more gracefully as operating current is increased.

VII. Acknowledgements

This work has been supported by the Spanish research grant TEC2006-11730-C03-01 (Samanta2) and the local Andalusian administration grant P06-TIC-01417 (Brain Systems). JALB was supported by a national scholarship from CSIC (Spanish Ministry of Education and Science).

VIII. References

[1] M. Mahowald, *An Analog VLSI Stereoscopic Vision System*, Kluwer Academic Publishers, 1994.

[2] K. A. Zaghloul and K. Boahen, "Optic nerve signals in a neuromorphic chip: Parts 1 and 2," *IEEE Trans.Biomed Eng.*, vol. 51, pp. 667-675, 2004.

[3] 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 Trans. on Circuits and Systems, Part II*, vol. 48, No. 1, pp. 4-11, Jan. 2001.

[4] Y. L. Wong, M. H. Cohen, and P. A. Abshire, "128x128 floating gate imager with self-adapting fixed pattern noise reduction," *Proc. of the IEEE 2005 Int. Symp. on Circuits and Systems* (ISCAS'05), vol. 5, pp. 5314-5317, 2005.

[5] J. Costas-Santos, T. Serrano-Gotarredona, R. Serrano-Gotarredona and B. Linares-Barranco, "A Contrast Retina with On-chip Calibration for Neuromorphic Spike-Based AER Vision Systems," *IEEE Trans.on Circuits and Systems I:Regular Papers.* July 2007.

[6] 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. on Circuits and Systems I: Regular Papers*, vol. 53, No. 12, pp. 2548-2566, Dec. 2006.

[7] 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. on Neural Networks*, vol. 14, No. 5, pp. 1207-1216, September 2003.

[8] 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, No. 8, pp. 1353-1363, August 2003.