Mismatch Analysis and Statistical Design at 65 nm and Below

Larry Pileggi, Gökçe Keskin, Xin Li, Ken Mai and Jon Proesel
Carnegie Mellon University
5000 Forbes Ave. Dept. of ECE, Pittsburgh, PA 15213 USA
{pileggi,gkeskin,xinli,kenmai,jproesel}@andrew.cmu.edu

Abstract- **Transistor sizing to control random mismatch is investigated. Input offset voltage of 65nm bulk CMOS SRAM sense amplifiers are measured to analyze NMOS and PMOS threshold voltage (Vtn, Vtp) variation effects and compare them with statistical models and Pelgrom model predictions. A linear statistical response surface model (RSM) relating input offset to Vtn and Vtp is shown to agree well with measured results. Designs optimized using the RSMs produce circuits with 25% lower input offset voltage spread at a cost of 10% more active device area. Statistical models for post-manufacturing configuration are postulated and shown for sub-65nm technologies.**

Keywords: Mismatch model, sense amplifier, input offset voltage

I. Introduction

Pelgrom's models [1] are universally applied for determining sufficient sizing of CMOS transistors to control mismatch due to random process variations. It is commonly applied to a wide range of circuit building blocks, such as differential amplifiers, current sources, and sense amplifiers. Importantly, since Pelgrom's analysis describes the mismatch between two elements, its application to circuits with more than one dominant source of mismatch is not straightforward [2] and can result in considerable overdesign.

To investigate Pelgrom's model for capturing mismatch for 65nm bulk CMOS, we consider its application for input offset voltage of the latch type sense amplifier in Fig. 1. This circuit was chosen because: a) the input offset is not easily described by an analytical formula; and b) the input offset is generally dominated by the random mismatch of the input NFETs. Both of which make this a classic circuit for application of the Pelgrom model.

Along with the Pelgrom model, a simple linear response surface model (RSM) for the offset is constructed based on Monte Carlo simulations. One version of the sense amp is designed using this RSM and following a statistical optimization methodology similar to that in [3]. The sense amplifiers are fabricated and tested to compare the statistical measurement results. While the required sizing is reasonable for handling mismatch for 65nm, we will show that below the 65nm node, post manufacturing configuration will eventually be the most effective way to manage the increase random variations that result in mismatch.

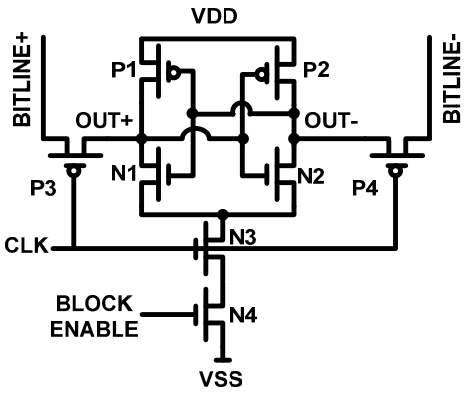

Fig. 1. Latch type sense amplifier (LTSA).

II. Pelgrom's Model and The Latch Type Sense Amplifier

When used in SRAM circuits, latch type sense amplifiers (LTSAs) are enabled only after memory bitlines reach a differential voltage swing level that is detectable [4]. For a robust read operation, this swing voltage must be larger than the input offset voltage of the LTSA. Bitlines are discharged by an individual memory cell, therefore, input offset voltage of the LTSA is an important factor in determining the overall parametric yield of the memory at a chosen operating frequency.

LTSAs operate at high speed by using the positive feedback formed by the (N1, P1) and (N2, P2) inverter pairs [5]. Unfortunately, this positive feedback makes it difficult to derive analytical expressions for the offset voltage and yield estimates. Previous work for modeling the offset ranges from simple analyses [4] following Pelgrom's model [1] to complex equations based on imprecise square law models for transistors [2].

Pelgrom's model on the matching of a process parameter P (e.g., threshold voltage) between two circuit elements is given by the equation [1]:

$$\sigma^2(\Delta P) = \frac{A_P^2}{WL} + S_P^2 D_x^2 \qquad (1)$$

where $\sigma^2(\Delta P)$ is the variance of the difference of parameter P between the elements; A_P and S_P are the area and spacing

proportionality constants for parameter P; W and L are the dimensions of each element; and D_x is the spacing between them. If the elements are laid out in close proximity with a layout style that eliminates most sources of systematic offset (e.g., common centroid), mismatch is mostly dominated by the $A_P^2/(WL)$ term in (1). Pelgrom's model is particular useful for initial design of circuits where a certain performance specification is dominated by a *single* mismatch source, P, such as the LTSA in Fig. 1. With proper selection of transistor lengths, and careful layout to avoid systematic variations, the input offset voltage is dominated by the threshold voltage mismatch of NMOS transistors N1 and N2 when the input voltages are close to V_{DD}.

In an SRAM, before the read operation, both bitlines are precharged to a voltage V_{pc}. During the first phase of the read operation; CLK is pulled low, one of the bitlines is discharged by the accessed memory cell, and the other is held at V_{pc}. The bitline discharge is relatively slow since SRAM cells are formed by small transistors to maximize memory density. Bitline voltages pass through P3 and P4, then are stored at the output nodes OUT+ and OUT–. Transistor N4 is included to disable a block of LTSAs when used in an SRAM design, hence it will be enabled for our analysis. When CLK is pulled high, the cross coupled inverters turn on and strong positive feedback pulls the output nodes to complementary logic levels quickly. It is desirable to perform CLK low to high transition as early as possible for reduced power consumption and increased memory speed. However, for a correct read operation, voltage differential at OUT+ and OUT– must be larger than the input offset voltage of the LTSA when CLK turns high.

If V_{pc} is kept near V_{DD}, after one of the bitlines is discharged and CLK goes high, (N1, N2) pair is on while (P1, P2) pair is off. A decision is well under way when one of the outputs reaches $V_{DD}–V_{tp}$ and one of the PMOS transistors turns on. Therefore, mismatch of (P1, P2) is less important and mismatch of (N1, N2) is the dominant factor on offset when V_{pc} is near V_{DD}. Such conditions meet requirements for application of Pelgrom's model.

Fig. 2 shows the scatter plot of the standard deviation of input offset voltage (σ_{Offset}) with respect to Diff(Vtn) and Diff(Vtp) from Monte Carlo simulations when $V_{pc}=V_{DD}=1.0V$. The axes are normalized to their respective standard deviations and:

$$Diff(Vtn) = VtN1 - VtN2 \quad Diff(Vtp) = VtP2 - VtP1 \qquad (2)$$

$$\sigma_{Diff(Vtn)} = \sigma_{VtN1}\sqrt{2} \qquad \sigma_{Diff(Vtn)} = \sigma_{VtP1}\sqrt{2} \qquad (3)$$

For modeling purposes we assume that the input offset is dominated by random mismatch and that the (VtN1, VtN2) and (VtP1, VtP2) pairs are independent, identically distributed random variables representing the threshold voltages.

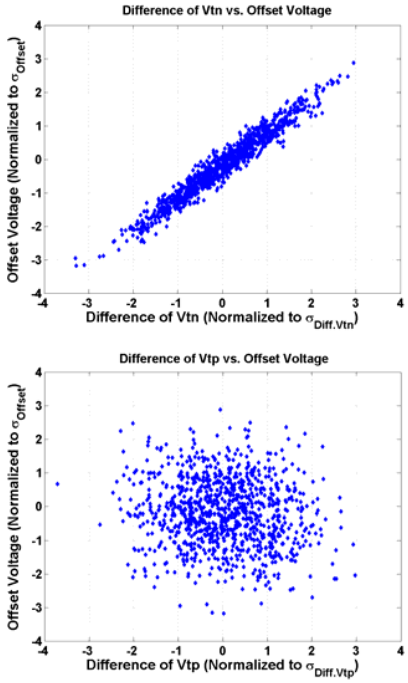

Fig. 2. Offset voltage vs. Diff(Vtn) and Diff(Vtp), $V_{pc}=V_{DD}=1.0V$.

As expected, Fig. 2 shows a linear relationship between offset and Diff(Vtn), while there is little or no correlation with Diff(Vtp). Therefore, we apply the following first order model for offset:

$$Offset = \left(a \times Diff(Vtn)\right) + \left(b \times Diff(Vtp)\right) + c \qquad (4)$$

We characterize this model using data for four different precharge voltages (0.7, 0.8, 0.9 and 1.0V) while keeping V_{DD} constant at 1.0V. Compared to simulations, the modeling error of the linear approximation for each case is shown to be within 3%, thus confirming that input offset is indeed dominated by random threshold voltage mismatch in terms of the simulation models.

Fig. 3 shows the correlation coefficients (CC) of Diff(Vtn) and Diff(Vtp) with offset for different V_{pc}'s. As expected, as V_{pc} is lowered, P1 and P2 turn on sooner during the latching phase and their mismatch has an increasingly stronger impact on the offset. The correlation coefficient between offset and Diff(Vtp) increases correspondingly, and Pelgrom's model is no longer directly applicable when the second source of mismatch becomes significant. Measurement results in the next section will support this expected behavior.

Using the RSM model in (4), the transistors are sized to reduce the mismatch by 25% while minimizing the area and power impact. Simulation results for this statistically optimized design are also shown in Fig. 3. The optimization results in larger NFETs (N1, N2) but smaller PFETs (P1, P2), since the offset impact of the former are greater. We will compare this RSM model and Pelgrom's formula with measurement results.

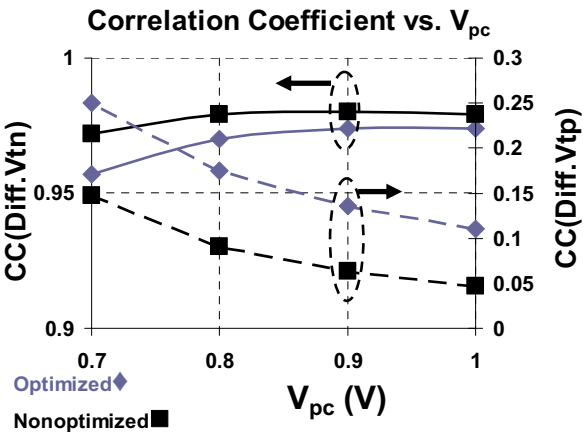

Fig. 3. Correlation Coefficient (CC) of Offset with Diff(Vtn), Diff(Vtp).

III. Measurement Results

We designed and fabricated arrays of optimized and non-optimized LTSAs in a commercial 65nm bulk CMOS technology (Fig. 4). Each chip includes 2048 LTSAs, 50% based on the original design and 50% based on the statistically optimized design. Each differential output is connected to a D-flipflop, and the flops are connected as a scan chain. The chip is wire-bonded in a PGA package, and mounted on a printed circuit board (PCB) using a socket. Inputs (bitlines) are heavily decoupled with both on and off-chip capacitance, and the input differential voltages are externally applied. After inputs are set, LTSAs and flops are clocked a few times to clear any potential metastability. After clocking, flops are put to scan mode and digital outputs are pushed through the scan chain at a low frequency. The scan chain output is read by a logic analyzer probe on the PCB. An automated test setup sweeps the inputs in steps of a few mV over a wide input range, and the switching point of each LTSA is determined by post processing the outputs [6]. The tests are repeated at four different precharge voltages.

Histograms for the measured input offset distribution for the circuits are shown in Fig. 5. More than 12k samples for each design are collected from 12 different die. Bins are normalized to the standard deviation of the offset voltage of the non-optimized circuit ($\sigma_{Offset,NO}$), and results are shown for V_{pc}=1.0V. The improved offset spread for the statistically optimized circuit is apparent in the measurement results. Simulated and measured values of $\sigma_{Offset,NO}$ are within 5%, and this difference is even less for the optimized circuit ($\sigma_{Offset,O}$).

Fig. 6 shows a comparison of the $\sigma_{Offset,NO}/\sigma_{Offset,O}$ ratio for a simple area ratio based on Pelgrom model:

$$SimpleAreaRatio = \sqrt{\frac{Area(N1+N2), OptimizedCircuit}{Area(N1+N2), NonOptimizedCircuit}} \quad (5)$$

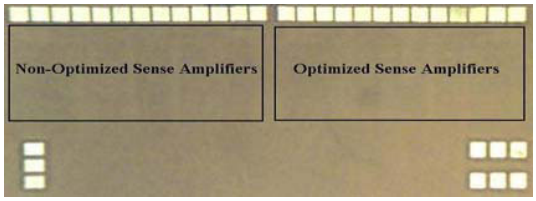

Figure 4. Die photo of the manufactured chip.

The efficacy of Pelgrom's model is apparent in Fig.6; at the region where σ_{Offset} is dominated by the mismatch of N1 and N2 (i.e., V_{pc}=1.0V), the Pelgrom model predicts the performance of the circuit very well. As V_{pc} is decreased and mismatch of P1 and P2 further impacts σ_{Offset}, Pelgrom's ratio is no longer applicable. The RSM model in [3] is nearly overlapping with the simulation curve in Figure 6. Such an RSM model can be an effective model of performance (σ_{Offset}) during initial design in general, and can further suggest design trade-offs in the circuit and for statistical optimization.

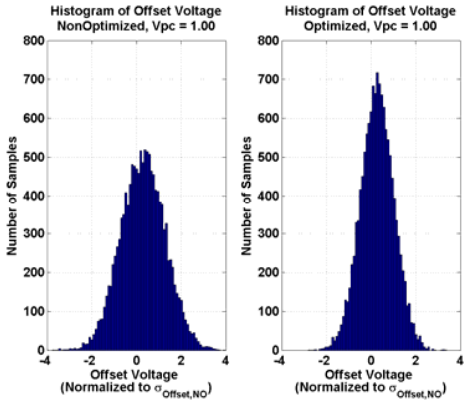

Figure 5. Histogram of input offset voltage (measured).

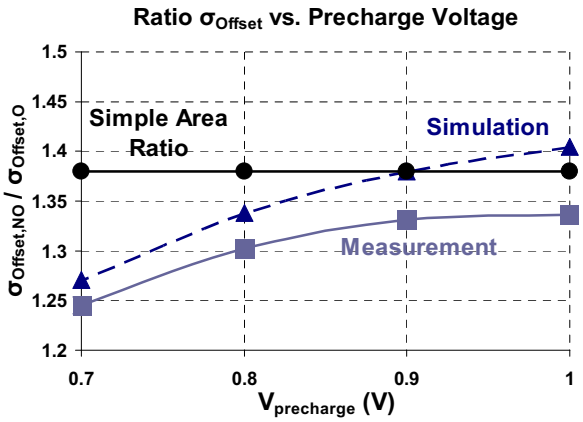

Fig. 6. Comparison of σ_{Offset} ratio with simple area ratio based on Pelgrom model.

IV. Statistical Element Selection

Even with more accurate RSM models, the scaling of analog designs will be limited due to the large device sizes required to accommodate mismatch specifications. For 45nm and below, the random variations will become more dominant for general bulk CMOS. Instead of oversizing devices to *average out* random mismatch, the randomness can be used to improve the ability to provide post-manufacturing configuration to match devices and sub-components with *statistical element selection (SES)*.

Consider the LTSA in Fig. 1 that is designed with multiple sub-components connected in parallel. The number of sub-components would be determined by Pelgrom's model or the RSM in (4). Next consider a post-manufacturing capability of enabling only a subset of the sub-components for matching, but being able to scan through the subset combinations to pick a good, or even the best sets. The random variations help to create a very large population of choices for matching. It can be shown that there is an *exponential* increase in the number of subsets with a linear increase in the number of elements in a set. Correspondingly, there is an exponential decrease in offset voltage spread among a set of selectable elements with a linear increase in area (Fig. 7), as compared to a 1/sqrt(area) relationship following the Pelgrom model [7].

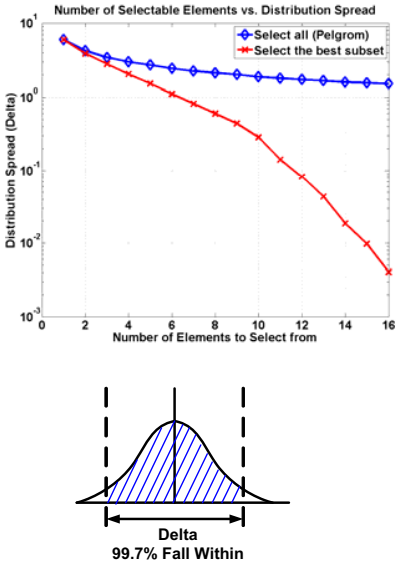

Fig. 7. Comparison of SES with Pelgrom style matching.

At extreme scaling of CMOS, a wide range of circuits can potentially benefit from an SES approach. Sense amplifiers can be made reconfigurable to reduce input offset voltage, resulting in a higher parametric yield for the memory. The selection can be performed with digital logic and stored in the memory. In general, the exponential scaling with area and relatively low analog complexity makes SES extremely competitive with other potential post-manufacturing calibration techniques.

Conclusions

A linear response surface model for relating input offset voltage of latch type sense amplifiers to threshold voltages variations has been described. The model is compared the design choices specified by the Pelgrom model. The response surface model was used to demonstrate the efficacy of statistical sizing for the LTSA. Measurements from a 65nm bulk CMOS testchip were used to compare the original and optimized designs and their corresponding models. The statistically optimized design resulted in a 25% decrease in the standard deviation of the input offset voltage at a cost of 10% increase in active area for the LTSA. A statistical element selection (SES) algorithm was proposed as the next step to control random variations in scaled CMOS processes.

Acknowledgements

The authors would like to thank Umut Arslan and Mark McCartney of Carnegie Mellon for the helpful discussions during the course of this work.We are very grateful for the access to 65nm fabrication from IBM, and the implementation support from John Cohn, Jack Pekarik, Randy Wolf and Ida Pucino. The authors also acknowledge the support of the Focus Center for Circuit & System Solutions (C2S2), one of five research centers funded under the Focus Center Research Program, a Semiconductor Research Corporation Program, and the National Science Foundation under contract CCF-0702278.

References

[1] M.J.M. Pelgrom, A.C.J. Duinmaijer, and A.P.G. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, pp. 1433-1439, October 1989.

[2] R. Singh and N. Bhat, "An offset compensation technique for latch type sense amplifiers in high-speed low-power SRAMs," *VLSI Systems, IEEE Transactions on*, vol. 12, pp. 652-657, June 2004.

[3] X. Li, P. Gopalakrishnan, Y. Xu, and L.T. Pileggi, "Robust analog/RF circuit design with projection-based performance modeling," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 26, pp. 2-15, January 2007.

[4] S.J. Lovett, G.A. Gibbs, and A. Pancholy, "Yield and matching implications for static RAM memory array sense-amplifier design," *IEEE J. of Solid-State Circuits*, vol. 35, pp. 1200-1204, August 2000.

[5] B. Wicht, T. Nirschl, and D. Schmitt-Lansiedel, "Yield and optimization of a latch type voltage sense amplifier," *IEEE J. of Solid-State Circuits*, vol. 39, pp. 1148-1158, July 2004.

[6] R. Ho, "On-chip wires: Scaling and efficiency," *Ph.D. Thesis*, Stanford University, 2003.

[7] X. Li, B. Taylor, Y-T. Chien, L. Pileggi, "Adaptive post-silicon tuning for analog circuits: concept, analysis and optimization," *ICCAD*, 2007.