A Physical Interpretation of the Distance Term in Pelgrom's Mismatch Model results in very Efficient CAD

B. Linares-Barranco and T. Serrano-Gotarredona

Abstract

In 1989 Pelgrom et al. published a mismatch model for MOS transistors, where the standard quadratic deviation of the mismatch in a parameter between two identical transistors, is given by two independent terms: (1) a transistor size-dependent term and (2) an inter-transistor distance-dependent term. To include the distance term, some researchers have developed CAD tools based on the so called σ-Space Methodology, which result in very computationally expensive algorithms. Such algorithms become non-viable even for circuits with a reduced number of transistors. On the other hand, by understanding and interpreting correctly the physical origin of Pelgrom's model distance term, one can implement in a straight forward manner this mismatch contribution in a CAD tool. Furthermore, the computational cost results negligible and viable for any number of transistors.

Introduction

The mismatch model proposed in 1989 by Pelgrom et al. [1] models the standard deviation in the mismatch of property P between two identical MOS transistors of width W and length L separated by a distance D (from center to center) in the layout, as

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

This model was experimentally verified by Pelgrom by measuring mismatch on many dies, fabricated on many runs, and including many identical transistors per die, as well as many transistor sizes per die (see Fig.2 in [1]). The model was verified for different foundries and technologies. Theoretical derivations of this model (see Appendix) reveal that the two terms are originated by two different means. The size dependent term is caused by local random fluctuations of material and technological properties of transistors. This random-induced mismatch term decreases with transistor area. On the other hand, the distance term in eq. (1), which is size independent (common for all sizes), is originated by gradients along the dies and wafers. For a particular die, the gradient surface can be approximated by a plane. Depending on the die position within the wafer, a different gradient plane will result. Consequently, in practice this gradient plane has a random nature from die to die. Since Pelgrom measured many dies, his models includes the random characterization of these random gradient planes. Layout techniques, like common-centroids, can eliminate the impact of gradient-induced mismatch, and consequently eliminate the distance term from eq. (1). However, layout techniques cannot eliminate the size-dependent term, because of its random nature.

1. Physical Interpretation and Implementation of Pelgrom's Mismatch Model in a CAD tool

The random-induced size-dependent term is quite straightforward to add in a CAD circuit simulation tool. First, one needs to know the critical transistor mismatch parameters, whose random fluctuations impact transistor currents. They are usually a small number of parameters. Pelgrom suggested 2 main ones (V_{T0} and β) and a secondary one (γ) [1]. Bastos et al. added mobility degradation [2]-[3], Serrano et al. suggested a total of 5 relevant parameters [4]-[5], and recently models extending from weak to strong inversion have been proposed with no more than 5 parameters [6]-[8]. For more sophisticated transistor models like BSIM, about 16 parameters would be required [9]. Consider a generic mismatch relevant parameter P. Let us assume P_{mean} is the mean value predicted by the manufacturer. Usually the manufacturer also characterizes global inter-die variations ΔP_{global} common for all transistors in the same die, whose standard deviation $\sigma(\Delta P_{global})$ would characterize variations from die to die. Finally, a local term ΔP_i, has to be added for each transistor in the circuit, such that its standard deviation is characterized by eq. (1). This local mismatch term includes two components, a random size-dependent component ΔP_{rand_i} and gradient induced size-independent component ΔP_{grad_i}.The standard deviation of the random component is given by

$$\sigma(\Delta P_{rand\_i}) = \frac{A_P}{\sqrt{2 W_i L_i}} \qquad (2)$$

where W_i and L_i are width and length of transistor i. The factor 2 in the denominator accounts for the fact that each transistor is deviated from a nominal mismatch-less transistor. Parameters A_P are provided by the manufacturer for the mismatch relevant parameters. Sometimes, correlations between these parameters are also characterized. In these cases, it is convenient to reflect them when generating the random numbers ΔP_{rand_i} for each parameter [4].

Now let us add the distance term of eq. (1). Let us assume we have the layout of our circuit and we know the central coordinates of each transistor i in the layout (x_i, y_i). Let us assume also, that the manufacturer provides parameter S_P of the distance term for parameter P in eq. (1). Let us assume also, that for each

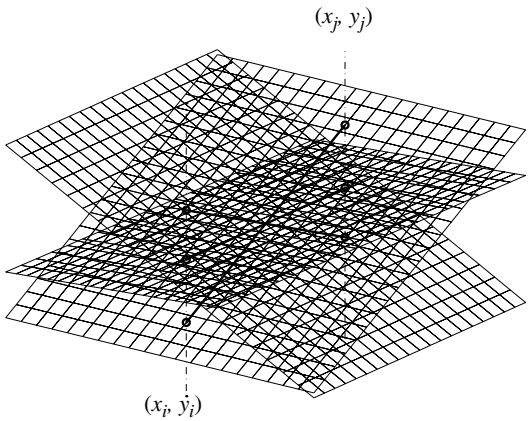

Fig. 1: Illustration of random gradient planes for transistor property *P* that gives rise to a distance dependent mismatch term ΔP between the two transistors at coordinates (x_i, y_i) and (x_j, y_j).

fabricated die, we can approximate the gradient of P along the die by a plane

$$P(x, y) = Ax + By + C \tag{3}$$

where $C = P_{mean} + \Delta P_{global}$, and A and B are random numbers. Consider now two transistors i and j located at coordinates (x_i, y_i) and (x_j, y_j). The mismatch in property P caused by the gradient plane of the die is

$$\Delta P_{grad\_ij} = A(x_i - x_j) + B(y_i - y_j) \tag{4}$$

Repeating this for many dies by generating random numbers A and B for each die, we can compute

$$\sigma^2(\Delta P_{grad\_ij}) = \sigma^2(A)(x_i - x_j)^2 + \sigma^2(B)(y_i - y_j)^2 \tag{5}$$

Assuming symmetry of the random planes $\sigma(A) = \sigma(B)$ (no preferred directions), results in

$$\sigma^2(\Delta P_{grad\_ij}) = \sigma^2(A)[(x_i - x_j)^2 + (y_i - y_j)^2] = \sigma^2(A)D_{ij}^2 \tag{6}$$

where D_{ij} is the distance between transistors i and j. Comparing eq. (6) with the distance term in eq. (1), reveals that

$$S_P = \sigma(A) = \sigma(B) \tag{7}$$

Consequently, if the manufacturer provides parameter S_P we can generate the random gradient planes for each simulated die.

Summarizing, according to Pelgrom's 1989 findings [1], a CAD tool implementing all these mismatch components of a MOS transistor should compute for each transistor i and for each of its mismatch relevant parameters P_i a deviation including the following terms

$$P_i = P_{mean} + \Delta P_{global} + \Delta P_{rand\_i} + \Delta P_{grad\_i} \tag{8}$$

To compute ΔP_{rand_i} a random number needs to be generated for each transistor i. However, to compute ΔP_{grad_i} only two random numbers need to be computed for all NMOS transistors (and another two for all PMOS) in the same circuit, which are parameters A and B, characterized by eq. (7). With these two random numbers, and the central coordinate of each transistor i, the term ΔP_{grad_i} would be given by

$$\Delta P_{grad\_i} = Ax_i + By_i \tag{9}$$

Note that this way of interpreting Pelgrom's distance term cancels out gradient effects when using common centroid layout techniques. Assume a differential pair where each transistor is split into two. This is consistent with Pelgrom's distance term prediction for common centroid layouts [1] (see (28) in Appendix).

2. σ-Space Methodoly

Shortly after Pelgrom's seminal paper was published, some authors developed a means to implement the distance term in eq. (1) in CAD environments [9]-[11], called sigma-space analysis or design. This methodology has been further developed by people in the CAD community [12]. However, here the distance term is not considered to model the statistics of the possible gradient planes from die to die. It is considered that within the same die, there is another random component per transistor ΔP_{d_i}, such that when computing the difference between two transistors separated by a distance D_{ij}, the standard deviation of this difference obeys

$$\sigma_{Dij}^2 = \sigma^2(\Delta P_{d\_i} - \Delta P_{d\_j}) = S_P^2 D_{ij}^2 \tag{10}$$

In principle, there could be many ways to generate random numbers for an arbitrary number of transistors in a circuit satisfying eq. (10). The way proposed in [9] would be as follows. Consider there are N transistors. Then for each of them we compute the following N correlated random numbers

$$\begin{aligned}
\Delta P_{d_1} &= 0 \\
\Delta P_{d_2} &= A_{22} R_2 \\
\Delta P_{d_3} &= A_{32} R_2 + A_{33} R_3 \\
&\dots \\
\Delta P_{d_N} &= A_{N2} R_2 + A_{N3} R_3 + A_{N4} R_4 + \dots + A_{NN} R_N
\end{aligned} \tag{11}$$

where R_i are random numbers normally distributed with zero mean and standard deviation equal to 1. Coefficients A_{ji} are computed by using eqs. (10) for each transistor pair. Since there are a total of $N(N-1)/2$ transistor pairs there is a total of $N(N-1)/2$ nonlinear quadratic equations to solve with a total of $N(N-1)/2$ parameters A_{ji} in eqs. (11). If eqs. (10) were linear, there would be one unique solution. However, since they are nonlinear we should expect many possible solutions. The computational cost of solving these equations grows exponentially with the number of transistors.

Obviously, the random gradient plane solution discussed previously (see eq.(9)) should be one of the solutions of the formulation of eqs. (11), since both interpretations satisfy the mathematical distance statistics of eq. (10). However, the computational cost for eqs. (7) and (9) is much smaller than for eqs. (10) and (11), and specially when there is a large number of transistors in the circuit. Surprisingly, a more detailed analysis of the solutions of eqs. (11) reveals that the gradient plane is the only possible solution.

3. Conclusions

We have seen that the distance dependent term $(S_P^2 D^2)$ in Pelgrom's mismatch model is equivalent to consider for each die a random gradient plane $Ax + By$, such that $\sigma(A) = \sigma(B) = S_P$. The solution of the formulation known as sigma-space analysis for predicting this mismatch term in a CAD tool, is also this random plane. However, the mathematical formulation used is extremely complex, is computationally expensive, and the computational cost grows exponentially with the number of transistors, resulting non-viable for moderate and large circuits. However, the random plane physical interpretation of this mismatch component results in a very simple mathematical formulation, very easy to implement in a CAD tool, and without almost any computing penalty.

Appendix. Theoretical Derivation of Pelgrom's Model

In the 1989 Pelgrom paper [1], the authors did not include the theoretical derivation of the mismatch model of eq. (1). Here, we provide this derivation, which to our knowledge is not available anywhere else.

Fig. 2 shows two transistors of size $W \times L$ located at coordinates (x_1, y_1) and (x_2, y_2), respectively. Let us define the position of the pair as its middle point

$$x_{12} = \frac{x_1 + x_2}{2} \quad , \quad y_{12} = \frac{y_1 + y_2}{2} \tag{12}$$

Let us assume that property P of a transistor can be obtained by averaging a certain density function P over its area

$$P_1(x_1, y_1) = \frac{1}{WL} \int \int_{area(x_1, y_1)} P(x', y') dx' dy' \tag{13}$$

The density function $P(x', y')$ is assumed to reflect wafer gradients as well as random noise components.

Under these assumptions, the mismatch in property P for the transistor pair located at (x_{12}, y_{12}) is given by

$$\begin{aligned} \Delta P(x_{12}, y_{12}) &= P_1(x_1, y_1) - P_2(x_2, y_2) = \\ &= \frac{1}{WL} \int \int_{area(x_1, y_1)} P(x', y') dx' dy' - \frac{1}{WL} \int \int_{area(x_2, y_2)} P(x', y') dx' dy' = \\ &= \frac{1}{WL} \int \int_{\Re^2} G(x_{12} - x', y_{12} - y') P(x', y') dx' dy' \end{aligned} \tag{14}$$

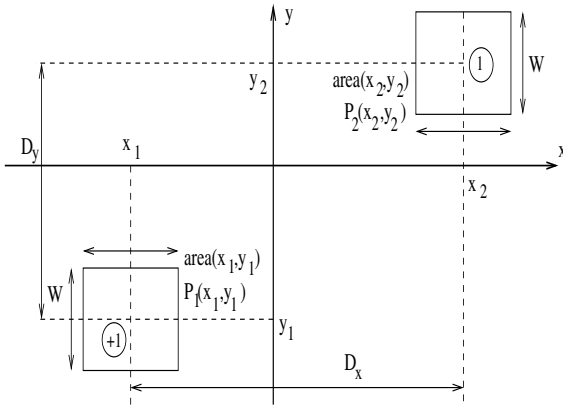

Fig. 2: Position and Coordinates of two Transistors

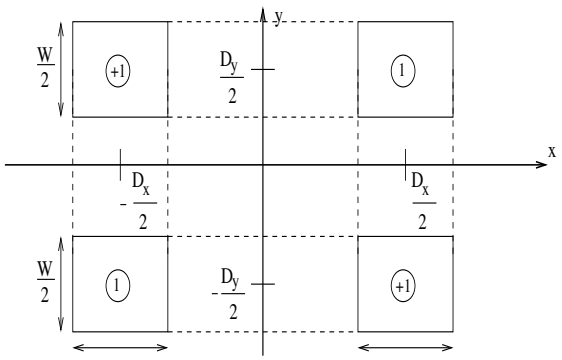

Fig. 3: Layout Configuration for a Transistor Pair using Common Centroid

where $G(x, y)$ is a geometry function which is +1 inside the region of the transistor at (x_1, y_1), -1 inside the one at (x_2, y_2), and zero elsewhere. Taking the Fourier Transform in eq. (14) yields,

$$\Delta \mathcal{P}(\omega_x, \omega_y) = \frac{1}{WL} \mathcal{G}(\omega_x, \omega_y) \mathcal{P}(\omega_x, \omega_y) \tag{15}$$

where $\Delta \mathcal{P}(\omega_x, \omega_y)$ is the Fourier Transform of $\Delta P(x_{12}, y_{12})$, $\mathcal{G}(\omega_x, \omega_y)$ is the one of $G(x, y)$, and $\mathcal{P}(\omega_x, \omega_y)$ is the one of $P(x, y)$.

For the layout of Fig. 2 it can be shown that when choosing our coordinate system such that $D_y = 0$,

$$\left| \frac{1}{WL} \mathcal{G}(\omega_x, \omega_y) \right| = \frac{\sin\left(\frac{\omega_x L}{2}\right)}{\frac{\omega_x L}{2}} \frac{\sin\left(\frac{\omega_y W}{2}\right)}{\frac{\omega_y W}{2}} \left\{ 2 \sin\left(\frac{\omega_x D_x}{2}\right) \right\} \tag{16}$$

For a pair of transistors in a common centroid configuration, as shown in Fig. 3, it would be

$$\left| \frac{1}{WL} \mathcal{G}(\omega_x, \omega_y) \right| = \frac{\sin\left(\frac{\omega_x L}{2}\right)}{\frac{\omega_x L}{2}} \frac{\sin\left(\frac{\omega_y W}{4}\right)}{\frac{\omega_y W}{4}} \left\{ 2 \sin\left(\frac{\omega_x D_x}{2}\right) \sin\left(\frac{\omega_y D_y}{2}\right) \right\} \tag{17}$$

Fig. 4 shows a wafer in which contour lines of constant property P have been drawn. In the wafer, at coordinate (x_{12}, y_{12}) a pair of transistors is drawn.

Assuming that when averaging $\Delta P(x_{12}, y_{12})$ all over the wafer we have $\overline{\Delta P}\big|_{Wafer} \approx 0$ we can write that

$$\sigma^2(\Delta P) = \frac{1}{\Omega} \int \int_{\Omega} \Delta P^2(x_{12}, y_{12}) dx_{12} dy_{12} \tag{18}$$

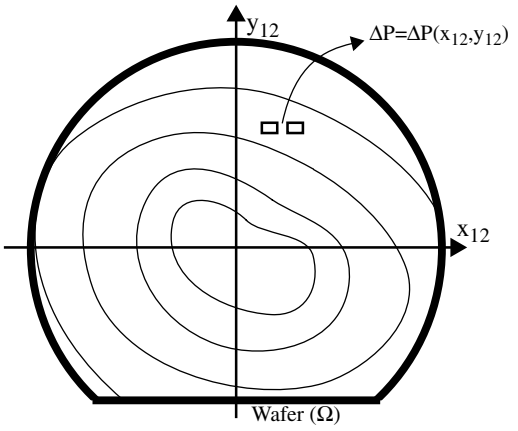

Fig. 4: Wafer Gradients

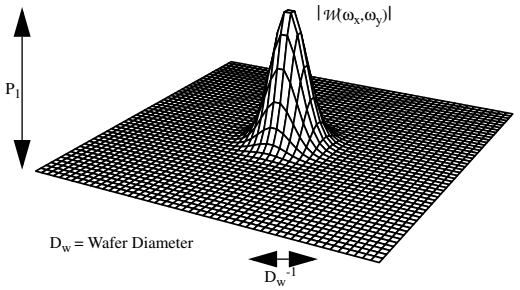

Fig. 5: Approximate Shape of Frequency Domain Function $\mathcal{W}()$

where Ω is the area of the wafer. Applying Poisson's Theorem to eq. (18) results in

$$\sigma^2(\Delta P) = \frac{1}{4\pi^2\Omega}\int_{-\infty}^{\infty} d\omega_x \int_{-\infty}^{\infty} d\omega_y \left| \frac{1}{WL} G(\omega_x, \omega_y) \mathcal{P}(\omega_x, \omega_y) \right|^2 \qquad (19)$$

Pelgrom models function $\mathcal{P}(\omega_x, \omega_y)$ as having two components [1]:

$$\mathcal{P}(\omega_x, \omega_y) = P_o + \mathcal{W}(\omega_x, \omega_y) \qquad (20)$$

where P_o is a constant (frequency independent) representative of random white noise and $\mathcal{W}(\omega_x, \omega_y)$ is a wafer map component responsible for long distance gradients along the wafer. The spatial frequency content of function $\mathcal{W}(\omega_x, \omega_y)$ is for frequencies of the order of D_w^{-1}, where D_w is the wafer diameter. Therefore, function $\mathcal{W}(\omega_x, \omega_y)$ can be assumed to have a shape of the type depicted in Fig. 5, and consequently we can assume that

$$\mathcal{W}(\omega_x, \omega_y) = \begin{cases} \sim P_1 & \text{if } \left(\frac{-1}{D_w} \le \omega_x \le \frac{1}{D_w}\right), \left(\frac{-1}{D_w} \le \omega_y \le \frac{1}{D_w}\right) \\ 0 & \text{otherwise} \end{cases} \qquad (21)$$

Therefore, eq. (19) can be written as

$$\sigma^2(\Delta P) = \frac{1}{4\pi^2\Omega W^2 L^2}\int_{-\infty}^{\infty} d\omega_x \int_{-\infty}^{\infty} d\omega_y |G|^2 |P_o + \mathcal{W}|^2 = \\ = \frac{1}{4\pi^2\Omega W^2 L^2}\{|P_o|^2 Y_1 + Y_2\} \qquad (22)$$

Assuming a transistor pair as in Fig. 2, it would be

$$Y_1 = \int_{-\infty}^{\infty} d\omega_x d\omega_y |G|^2 = 8\pi^2 WL \qquad (23)$$

and

$$Y_2 = \int_{-\infty}^{\infty} d\omega_x \int_{-\infty}^{\infty} d\omega_y |G|^2 [P_o^* \mathcal{W} + P_o \mathcal{W}^* + |\mathcal{W}|^2] \qquad (24)$$

Since $D_w \gg D_x, W, L$ we only need to evaluate G in eq. (24) near the origin. From eq. (16) $G(\omega_x, \omega_y) \approx \omega_x D_x L W$. Thus,

$$Y_2 \approx D_x^2 L^2 W^2 k_o' \\ k_o' = \int_{-\frac{1}{D_w}}^{\frac{1}{D_w}} d\omega_x \int_{-\frac{1}{D_w}}^{\frac{1}{D_w}} d\omega_y \omega_x^2 [P_o^* \mathcal{W} + P_o \mathcal{W}^* + |\mathcal{W}|^2] \qquad (25)$$

where $P_o^* \mathcal{W} + P_o \mathcal{W}^* + |\mathcal{W}|^2 \approx P_o^* P_1 + P_o P_1^* + |P_1|^2 = k_o$ is constant and

$$Y_2 = \frac{4 k_o D_x^2 L^2 W^2}{3 D_w^4} \qquad (26)$$

This results in

$$\sigma^2(\Delta P) = \frac{2|P_o|^2}{\Omega WL} + \frac{k_o D_x^2}{3\pi^2 \Omega D_w^4} = \frac{A_P^2}{WL} + S_P^2 D_x^2 \\ A_P^2 = \frac{2|P_o|^2}{\Omega} \quad , \quad S_P^2 = \frac{k_o}{3\pi^2 \Omega D_w^4} \qquad (27)$$

For a common centroid configuration, by changing the geometry function, it can be shown that the result is

$$\sigma^2(\Delta P) = \frac{2|P_o|^2}{\Omega WL} + \frac{k_o D_x^2 D_y^2}{36\pi^2 \Omega D_w^6} = \frac{A_P^2}{WL} + S_P^2 \frac{D_x^2 D_y^2}{12 D_w^2} \qquad (28)$$

Note that the distance term has been reduced an amount of the order $(D_y / D_w)^2$, which is very small. Consequently, from a practical point of view, the distance term can be considered to have disappeared. Perfect gradient planes would cancel out exactly this distance term. However, in practice the gradients not always are perfect planes and might have a higher order curvature component. This is why in eq. (28) there results a residual distance term.

References

[1] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching Properties of MOS Transistors," *IEEE Journal of Solid State Circuits*, vol. 24, No. 5, pp. 1433-1440, 1989.

[2] J. Bastos, *Characterization of MOS Transistor Mismatch for Analog Design*, Ph. D. Dissertation, Katholieke Universiteit Leuven, 1998.

[3] M. Steyaert, J. Bastos, R. Roovers, P. Kinget, W. Sansen, B. Graindourze, A. Pergoot, E. Janssens, "Threshold voltage mismatch in short-channel MOS transistors," *Electronics Letters*, vol. 30, No. 18, pp.1546-1548, 1 Sept. 1994.

[4] T. Serrano-Gotarredona and B. Linares-Barranco, "Systematic Width-and Length Dependent CMOS Transistor Mismatch Characterization and Simulation," *Int. J. of Analog Integrated Circuits and Signal Processing*, vol. 21, pp. 271-296, December 1999.

[5] T. Serrano-Gotarredona and B. Linares-Barranco, "A new five-parameter MOS transistor mismatch model," *IEEE Electron Device Letters*, vol. 21, No. 1, pp. 37-39, Jan. 2000.

[6] J. Croon, M. Rosmeulen, S. Decoutere, and W. Sansen, "An Easy-to-Use Mismatch Model for the MOS Transistor," *IEEE J. of Solid State Circuits*, vol. 37, No. 8, pp. 1056-1064, August, 2002.

[7] P.G. Drennan and C.C. McAndrew, "Understanding MOSFET mismatch for analog design," *IEEE J. of Solid-State Circuits*, vol. 38, No. 3, pp. 450-456, March 2003.

[8] T. Serrano-Gotarredona and B. Linares-Barranco, "CMOS Transistor Mismatch Model valid from Weak to Strong Inversion," *Proc. 2003 Europ. Solid-State Circ. Conf.* (ESSCIRC'03), pp. 627-630, September 2003.

[9] C. Michael and M. Ismail, "Statistical modeling of device mismatch for analog MOS integrated circuits," *IEEE J. of Solid-State Circuits*, vol. 27, No. 2, pp. 154-166, Feb. 1992.

[10] C. Michael and M. Ismail, *Statistical Modeling for Computer-Aided Design of MOS VLSI Circuits*, Boston, MA: Kluwer Academic, 1993.

[11] C. Michael, C. Abel, and M. Ismail, "SMOS: A CAD-compatible statistical model for analog MOS integrated circuit simulation," *Int. J. Circuit Theory Applicat.*, vol. 20, pp. 327-347, Mar. 1992.

[12] H. Zhang, Y. Zhao, and A. Doboli, "ALAMO: an improved s-space based methodology for modeling process parameter variations in analog circuits," *Proc. of the conference on Design, automation and test in Europe,* pp. 156-161, 2006.