Common-Centroid Layout for Active and Passive Devices: A Review and the Road Ahead

Nibedita Karmokar†, Meghna Madhusudan†, Arvind K. Sharma†, Ramesh Harjani†, Mark Po-Hung Lin‡, Sachin S. Sapatnekar†
†University of Minnesota, USA ‡National Yang Ming Chiao Tung University, Taiwan

***Abstract*—This paper presents an overview of common-centroid (CC) layout styles, used in analog designs to overcome the impact of systematic variations. CC layouts must be carefully engineered to minimize the impact of mismatch. Algorithms for CC layout must be aware of routing parasitics, layout-dependent effects (for active devices), and the performance impact of layout choices. The optimal CC layout further depends on factors such as the choice of the unit device and the relative impact of uncorrelated and systematic variations. The paper also examines scenarios where non-CC layouts may be preferable to CC layouts.**

I. Introduction

In analog/mixed-signal (AMS) circuits, process variations cause unpredictability in circuit performance parameters. AMS circuits are built so that they are less sensitive to the absolute value of process-induced variability of a device or passive (which is hard to control), but are still sensitive to the differential variability between devices (which are more controlled). For example, in differential structures such as differential pairs (DPs) in an operational transconductance amplifier (OTA), the use of matching is effective in reducing variations in OTA performance. Several other analog structures, e.g., active devices in planar and FinFET technologies (e.g., current mirrors) and passives (e.g., resistor/capacitor arrays), require matching.

This paper overviews common-centroid layout [1], one of the most widely used techniques for reducing process-induced differential mismatch in analog layouts. The common-centroid (CC) technique creates a layout for a set of k elements, with each device i consisting of s_i units. CC layout ensures that the centroids of the units of each device coincide.

The layout of the DP in Fig. 1 could be organized into an array of two elements, devices A and B (i.e., $k = 2$), each consisting of $s_A = s_B = 2$ unit cells [1]. The CC technique lays out devices in a 1D or 2D array such that in each dimension of the array, the centroids match. Given that the location of unit i of device j is (x_i^j, y_i^j), for a 1D layout such as Fig. 1, the CC criterion is:

$$\frac{1}{s_A}\sum_{i=1}^{s_A} x_1^A = \frac{1}{s_B}\sum_{i=1}^{s_B} x_2^B \tag{1}$$

In the figure, this is met using the "ABBA" sequence.

A 2D CC layout pattern is symmetric around both the X- and Y-axis. In a general 2D array,

$$\frac{1}{s_1}\sum_{i=1}^{s_1} x_i^1 = \frac{1}{s_2}\sum_{i=1}^{s_2} x_i^2 = \cdots = \frac{1}{s_k}\sum_{i=1}^{s_k} x_i^k \tag{2}$$

$$\frac{1}{s_1}\sum_{i=1}^{s_1} y_i^1 = \frac{1}{s_2}\sum_{i=1}^{s_2} y_i^2 = \cdots = \frac{1}{s_k}\sum_{i=1}^{s_k} y_i^k \tag{3}$$

In contrast, *interdigitated* layouts alternate the placement of the fingers (or FinFET unit cells) of each of the k devices, e.g., the 1D layout in the sequence "ABAB" shown in Fig. 1. Interdigitated schemes do not have a common centroid for the devices: in the figure, the centroid for the A cells lies to the left

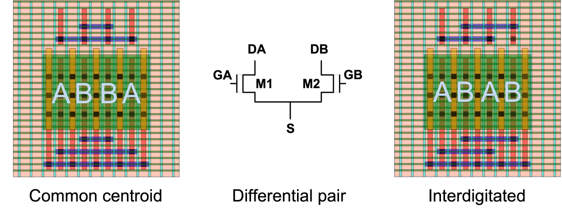


Fig. 1: Common-centroid and interdigitated layout of a 2× differential pair in a FinFET technology.

of that for the B cells. Generally speaking, CC layouts have been considered to be better for matching process-induced variations than other alternatives such as interdigitated patterns, and are widely used to match circuit elements. However, it should be noted that CC layouts may involve more complex routing and larger routing parasitics than other alternatives.

The rationale for using CC layouts is that they cancel out linear systematic variations due to first-order process gradients. A variation Δp in process parameter p induces a small perturbation, ΔP, in the circuit performance parameter P. This can be modeled using a linear Taylor series expansion,

$$\Delta P = \mathcal{S}_p \Delta p \tag{4}$$

where $\mathcal{S}_p = \partial P/\partial p$ is the sensitivity at the nominal point. Using the centroid as the origin, the variations are modeled by a plane $\Delta p = \alpha \cdot x$ where α is the (possibly unknown) gradient of the variation. In the horizontal dimension x,

$$\Delta P = \alpha \mathcal{S}_p \cdot x \tag{5}$$

i.e., for linear variations, i.e., constant α, the performance P is a linear function of x, the location of each device.

Under this linear assumption, the CC criterion ensures that the sum of variations over all devices cancel each other out. In Fig. 1, let us say that p represents the threshold voltage and P the drain current. Since $\Delta p = \alpha \cdot x$, the parameter p of device A is shifted by -2α for the leftmost unit cell and $+2\alpha$ for the rightmost unit cell with respect to the value at the centroid. From Eq. (5), the drain current shifts by $-2\alpha\mathcal{S}_{p,A}$ and $+2\alpha\mathcal{S}_{p,A}$, adding up to a net shift of zero. Using a similar notation, currents in the devices of B shift by currents shift by $-\alpha\mathcal{S}_{p,B}$ and $\alpha\mathcal{S}_{p,B}$, also creating a net shift of zero. A similar argument justifies CC in 2D layouts.

Additionally, the aspect ratio of a CC layout is typically close to a square [2], for which the maximum distance from the origin is smaller than any other rectangle, thus limiting the magnitude of systematic variation over the layout.

II. Modeling On-Chip Variations

Process-induced on-chip variations can be categorized as either *systematic* variations, which can be modeled predictably, or

random variations, which can only be represented statistically. Variations can also be classified as:
Global variations: These affect all like elements on a chip similarly, and do not cause mismatch between elements on a die and are well modeled using process corners.
Local variations: These include systematic gradient-based variations and local random variations that can be modeled using spatially correlated models [3]–[5], whereby elements that are closer to each other on a chip have lower mismatch. These variations do not significantly affect small arrays [6]–[8].

Local systematic variations are often represented using linear or nonlinear models [8], while random variations are modeled using distributions. Local random variations can be characterized by their spatial correlation [5], [9]: *uncorrelated variations* affect even adjacent elements independently, while *spatially correlated variations* show a correlation trend that decays with the distance between the elements. This is captured by a metric called the *correlation distance* [10], [11] (uncorrelated variations have a correlation distance of zero).

The total variation in a process parameter is given by:

$$\Delta P = g + u + s \tag{6}$$

where g, u, and s are, respectively, the global, local uncorrelated, and local spatially correlated variations, with variances σ_g^2, σ_u^2, and σ_s^2. The mean of ΔP is zero, and its variance is

$$\sigma_P^2 = \sigma_g^2 + \sigma_u^2 + \sigma_s^2 \tag{7}$$

A. Modeling Systematic Variations

1) Process Gradients: We illustrate a gradient-based systematic variation model [12], [13] for a capacitive array. The nominal value of the oxide thickness at the origin (the center of the array) is t_0, resulting in a unit capacitor value of C_u. The capacitance C_k, $1 \leq k \leq n$, at location (x_k, y_k) is shifted due to systematic variations in the oxide thickness, t_k, to

$$C_k^* = \sum_k C_u \frac{t_0}{t_k} \tag{8}$$

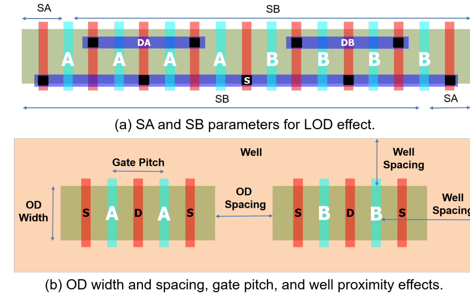
$$\text{where } t_k = t_0 + \gamma(x_k \cos\theta + y_k \sin\theta) \tag{9}$$

Here, γ and $0 \leq \theta \leq 180°$ are the linear oxide gradient magnitude and angle at the origin.

A metric for systematic variation in CC capacitor arrays is the maximum ratio mismatch between any capacitor pair. In an array of n capacitors, if the ideal capacitance ratio is $C_1 : C_2 : \cdots : C_n$ and $C_1^* : C_2^* : \cdots : C_n^*$ is the capacitance ratio due to systematic variations, the maximum ratio mismatch is:

$$M_{sys} = \max_{p,q \in \{1,\cdots,n\}, p \neq q} \left| \frac{(C_p^*/C_q^*) - (C_p/C_q)}{(C_p/C_q)} \right| \tag{10}$$

2) Layout-Dependent Effects: In advanced technology nodes, layout-dependent effects (LDEs) [14]–[16] induce shifts in the performance parameters of transistors. This shift depends on relative positions of features in the layout. The most common LDEs, shown in Fig. 2, include:
Well proximity effect (WPE) is seen in nanoscale CMOS nodes, where high-energy ions are used to create a deep retrograde well profile [16]. However, high-energy ions scatter at the edge of the photoresist and change the doping profile, modifying the



(a) SA and SB parameters for LOD effect.
(b) OD width and spacing, gate pitch, and well proximity effects.

Fig. 2: Layout-dependent effects.

threshold voltage V_{th} of a device based on its distance from the well edge (shown for device B in Fig. 2(b)). This effect is commonly knows as WPE [16], and WPE-induced mismatch can be minimized by keeping well edges far from devices or by maintaining equal well spacing for matched devices.
Length of diffusion (LOD) [17], results in variations in stress on transistors, and hence its V_{th}, due to changes in the length of the diffusion region. The impact of LOD is described by two parameters, SA and SB, the distances from poly-gate to the diffusion/active edge on either side of the device. For a device of gate length L_g and n unit cells [18]:

$$\Delta V_{th} \propto \frac{1}{\text{LOD}} = \sum_{i=1}^{n} \left(\frac{1}{\text{SA}_i + 0.5L_g} + \frac{1}{\text{SB}_i + 0.5L_g} \right) \tag{11}$$

Fig. 2(a) illustrates SA and SB for unit cells of devices A and B. Matched devices must have same values of SA and SB, in order to match their threshold voltage shift, ΔV_{th}.
Oxide definition (OD) spacing and width [14], also known as oxide spacing effect (OSE), is illustrated in Fig. 2(b). This effect changes the stress in a transistor due to variations in spacing between the OD regions (active areas), therefore altering V_{th}. Moreover, the stress induced in a transistor varies with the OD width (active area width). Mismatch can be avoided by maintaining same OD width and spacing.
Gate pitch variations causes the stress induced in a transistor to shift [14], as shown in Fig. 2(b) for device A. As gate pitch increases, the volume of the stressor material around the poly increases, resulting in increased stress in the transistor channel that perturbs V_{th}. In analog cells, the mismatch is minimized by using the same poly pitch.

The use of identical unit cells for matched devices can be used to cancel out all LDEs except LOD and WPE [19]. Specifically, the gate/poly pitches are uniform for CC analog blocks; by construction, the unit cell approach ensures that the OD width is uniform; the y-direction OD spacing (OSE) is uniform for each transistor due to the use of a row-based unit cell placement approach, and the x-direction spacing is uniform due to diffusion sharing. Therefore, the focus must be on optimizing LOD and WPE mismatch through the use of dummies and using placement techniques.

B. Modeling Correlated Variations

Spatially-correlated variations in capacitor arrays can be modeled by the following correlation function for two capacitors separated by a Euclidean distance of r [20]:

$$\rho_s(r) = (\rho_0)^r \tag{12}$$

where $0 < \rho_0 < 1$ and l depend on the process: l plays a similar role as correlation distance (a typical value is 1mm).

For two capacitors, $C_p = pC_u$ and $C_q = qC_u$, with p and q unit capacitors, respectively, their correlation coefficient is:

$$\rho_{pq} = \frac{Cov(p,q)}{\sigma_p \sigma_q} \tag{13}$$

The corresponding variances and covariances are:

$$\sigma_p^2 = \sigma_u^2 (p + 2S_p)\,;\ \sigma_q^2 = \sigma_u^2 (q + 2S_q)\,;\ Cov(p,q) = \sigma_u^2 S_{pq}$$

$$S_p = \sum_{a=1}^{p-1} \sum_{b=a+1}^{p} \rho_{ab}\ ;\ S_q = \sum_{a=1}^{q-1} \sum_{b=a+1}^{q} \rho_{ab}\ ;\ S_{pq} = \sum_{a=1}^{p} \sum_{b=1}^{q} \rho_{ab}$$

For t capacitors, the average correlation coefficient ρ_{avg} over all $C(t,2) = t(t-1)/2$ pairwise combinations of capacitors

$$\rho_{avg} = \frac{1}{C(t,2)} \sum_{p=1}^{t-1} \sum_{q=p+1}^{t} \rho_{pq} \tag{14}$$

A widely used metric for CC capacitor arrays is the capacitance ratio mismatch due to random variations, given by:

$$M_{rand} = \max_{p,q \in \{1,\cdots,n\}, p \neq q} var\left(\frac{C_p}{C_q}\right) \tag{15}$$

It can be shown that $var(C_p/C_q)$ is given by

$$\frac{A_f^2}{C_u^2 q^4 WL} \left[q^2 (p + 2S_p) + p^2 (q + 2S_q) - 2pqS_{pq}\right] \tag{16}$$

where W and L are the device width and length, respectively.

For active devices, the spatial correlation model of [10], [11] models device variations using the following correlation functions between two devices that are a distance r apart:

$$\rho_g(r) = 1, \rho_u(r) = 0, \rho_s(r) = e^{-(r/R_L)^2} \tag{17}$$

where R_L is the *correlation distance*. For devices i and j,

$$\mathrm{cov}(P_i, P_j) = \sigma_g^2 + \rho_s(r)\sigma_s^2 \tag{18}$$

The correlation coefficient $\rho(r)$ between P_i and P_j is:

$$\rho(r) = \frac{\mathrm{cov}(P_i, P_j)}{\sigma_{P_i}\sigma_{P_j}} = \frac{\sigma_r^2 + \rho_s(r)\sigma_s^2}{\sigma_P^2} \tag{19}$$

A plot of $\rho(r)$ is shown in Fig. 3.

C. Modeling Random Variations

Uncorrelated random variations due to random dopant fluctuations (RDF) [21] or line edge roughness (LER) [22] can be reduced by using larger devices. One of the most widely-used transistor variational models in analog design was proposed by Pelgrom [23]. The model quantifies the mismatch in a parameter P (e.g., V_{th}) of two devices as the sum of two random variables corresponding to the uncorrelated component, u, and a spatially correlated component, s. The variance of the mismatch is given by

$$\sigma^2(\Delta P) = \sigma_u^2 + \sigma_s^2 \tag{20}$$

$$\text{where } \sigma_u^2 = \frac{A_P^2}{WL}\ ;\ \sigma_s^2 = S_P^2 r^2$$

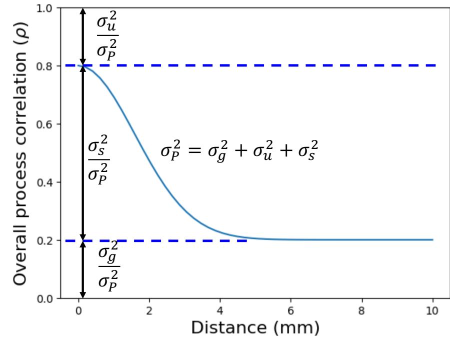


Fig. 3: Overall process correlation as a function of distance between two devices (adapted from [10], [11]).

Here, A_P and S_P are technology-dependent proportionality constants, r is the distance between the devices, and σ^2 represents the variance of the corresponding random variable. The first component depends on the area of the transistor and can be diluted by using large-sized transistors, while the second depends on the distance between components, and can be mitigated by layouts that reduce the distance between devices. Similar models for capacitors [24], [25] use

$$\sigma_u^2 = \frac{A_f^2}{WL} \tag{21}$$

where W is the width and L the length of C_u, and A_f is a mismatch coefficient similar to Pelgrom's coefficient.

III. Common-Centroid Capacitor Layout

CC capacitor layouts are essential for capacitor networks in many AMS integrated circuits, such as charge-scaling digital-to-analog converters (DAC), successive-approximation-register analog-to-digital converters (SAR ADC), switched capacitor filters, and other circuits requiring charge storage elements. For example, the CC layout applies to the binary-weighted capacitor network in the charge-scaling DAC in Fig. 4, in order to achieve highly matched capacitance ratios while reducing unwanted parasitics. The quality of CC capacitor layouts depend on unit capacitor structures, placement styles, and routing among the unit capacitors. The impact of variation must be translated into circuit performance metrics, such as nonlinearity and power consumption.

A. Performance Modeling

Different circuits may adopt different performance metrics. For the N-bit DAC in Fig. 4, the most important performance criteria include nonlinearity metrics, the differential nonlinearity

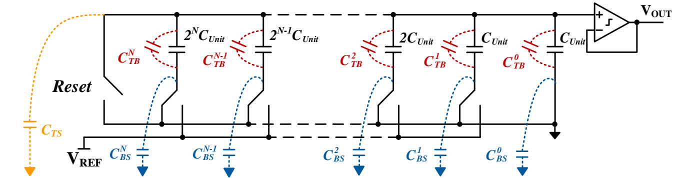


Fig. 4: Parasitic capacitors, C_{TB}, C_{TS}, C_{BS}, in the capacitor network of a charge-scaling digital-to-analog converter [26], [27], which may have great impact on overall circuit performance and power consumption.

(DNL) and integral nonlinearity (INL). DNL quantifies the degree to which each output step varies from the ideal step, which can be calculated by Eq. (22), while INL describes the maximum deviation between the ideal output of a DAC and the actual output level, which can be calculated by Eq. (23), where V_{LSB} is the ideal output voltage difference corresponding to any two adjacent digital codes, known as the least significant bit (LSB). If either DNL or INL of a DAC is worse than ± 1 LSB, it may result in a non-monotonic transfer function, or a missing code. To design an even more robust DAC, both DNL and INL are limited within ± 0.5 LSB.

$$DNL_i = \frac{V_{OUT}(i+1) - V_{OUT}(i) - V_{LSB}}{V_{LSB}}, \forall i = 0, \ldots, 2^{N-1}. \quad (22)$$

$$INL_i = \frac{V_{OUT}(i) - V_{OUT}^{ideal}(i)}{V_{LSB}}, \forall i = 0, \ldots, 2^{N-1}. \quad (23)$$

In both equations, the output voltage, V_{OUT}, may be far from the ideal value, $V_{OUT}^{ideal}(i)$, without a well-matched CC layout.

Power consumption is another critical metric, as many AMS circuits are used in mobile battery-powered devices. Larger unit capacitors can reduce the impacts from the mismatch due to process variation and routing parasitic, but may result in significantly higher power consumption and chip area. Therefore, a better tradeoff between capacitor mismatch and power consumption must be considered for CC capacitor layouts. Lin *et al.* [26], [27] have shown that minimizing routing parasitic mismatch can result in smaller required unit capacitance, and hence lower power consumption and area.

B. Unit Capacitor Structures

With the advancement of process technologies, more metal layers are available for chip design and manufacturing [28]. Instead of applying MIM capacitors, the metal-oxide-metal (MOM) capacitor structures, such as fingers [24], [29], [30], sandwiches [31], pillars [32]–[34], and mortise-tenon [35], and vertical bars [36] are preferable. The perspective view from top and the side views at three different cross sections of these MOM capacitors are shown in Figs. 5(a), (b), (c), and (d), respectively. Compared to MIM capacitors, MOM capacitors offer advantages of lower manufacturing cost as well as higher capacitance density due to multiple metal layers and shrinkage of metal width/spacing in advanced process nodes.

Among MOM capacitors, the fingered structure is the easiest to implement and has the highest capacitance density. Fingered structures are FinFET-technology-friendly as lower metal layers must be gridded with constant widths and pitches. However, in non-FinFET nodes, they may produce various unwanted parasitics after routing, as seen Fig. 4, leading to unexpected gain loss or higher switching energy. The large parasitic capacitance between the top plate and substrate, C_{TS}, may lead to significant gain loss [30]. In addition to C_{TS}, the routing for the finger structure may also induce large parasitic capacitance, C_{TB}, between the top plate and the bottom plate of different capacitors in SAR ADC due to coupling between the fingers and the metal wires connecting different capacitors.

The sandwich structure, pillar structure, and mortise-tenon structures can effectively reduce some unwanted parasitics and make routing easier. For example, the C_{TS} in these

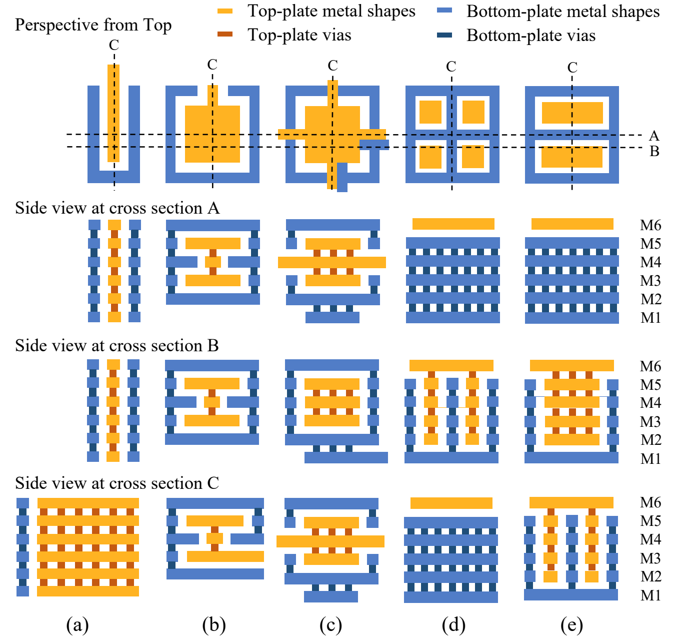


Fig. 5: Perspective view from top and different cross section views of popular MOM capacitor structures. (a) Finger [29]; (b) Sandwich [31]; (c) Pillar [32]; (d) Mortise-tenon I [35]; (e) Mortise-tenon II [28].

structures is eliminated because the top-plate metal shapes are enclosed by the bottom-plate metal shapes. However, these three structures are more complex, the corresponding capacitance densities are not as high as the finger structure. A parameterized mortise-tenon structure considering various dimensions and layers was introduced in [28] for fast unit capacitor generation while achieving high capacitance density for various unit capacitance values.

C. CC Capacitor Array Construction

A common step in CC layout is to first compute the array size, attempting to create a structure that is as close to a square as possible. If the total number of unit capacitors is less than the array size, then dummy cells are added to complete the array. An outer ring of dummy cells is often added to an array to avoid fringe effects for cells at the periphery of the array.

Early approaches to CC capacitor placement were *heuristic*. In [12] the concept of rectangles and circles was used to develop a placement and routing algorithm. The work of [20] showed that a higher dispersion degree between two capacitors can ensure a higher correlation coefficient and lower variation. In [37], a heuristic non-CC placement algorithm was proposed to increase correlation among capacitors, improving correlated random variations at the expense of systematic variations.

A second class of methods formulates the problem as an *integer linear program* [38]. The constraints include exclusivity, which slots exactly one unit capacitor into each location; ratio requirements that ensure that the number of the unit capacitors should be exactly equal to the required number; and a routing constraint, where each unit capacitor uniquely selects one track in one of its four adjacent channels for routing, where each track spans an entire channel between capacitors.

A third class of methods uses *structured methods* for creating CC layouts with high dispersion is based on the *chessboard* layout style, proposed in [39] and developed in [40], [41]. The method, focused on binary-weighted capacitor ratios, is illustrated for a six-bit DAC in Fig. 6(a). First, all unit capacitors of C_6 are placed on the same color in a chessboard style. Next, C_5 is placed on the other color as shown in Fig. 6(b). The process alternates between the colors of the chessboard until all unit capacitors are placed. To perform the placement of the capacitors from C_2 to C_{N-1}, [42] proposed a partition-centering based symmetry placement algorithm considering the impact of parasitics. In [43], the chessboard placement method is generalized to nonbinary capacitor ratios based on a hybrid chessboard placement method that aims to obtain the lowest DNL and while precisely match the routing wirelength with the capacitor ratio values. Chessboard routing involves numerous vias, which can cause a degradation in 3dB frequency due to high via resistances, particularly in FinFET nodes. In [44], the 3dB frequency is improved using a block chessboard method that places capacitors in a chessboard pattern at various granularities is presented.

A fourth class of methods uses *iterative* techniques based on stochastic optimization algorithms such as simulated annealing (SA). The work of [13] proposes a common centroid placement to maximize dispersion while respecting the adjacency constraint of non-integer capacitor ratios to reduce systematic and random mismatches simultaneously. The unit capacitors of a pair from the pair sequence are placed symmetrically with respect to the CC point of the placement matrix starting from the innermost circle and going outward direction. The work of [13] presents three operations during perturbation of the pair sequence to increase the degree of dispersion, and devises a procedure to maintain feasible placement that fulfills the adjacency constraint after each perturbation. However, the placement does not consider routing complexity.

A placement method based on the center-based corner block list (C-CBL) was proposed in [45] for CC placement, using a grid-based approach to place the devices uniformly so that they can average out the parasitic effects. After generating several feasible placements by varying the column numbers and eliminating redundant solutions, SA is used to perturb the global sequence pair and they re-defined the moves to perturb the position of the sub-devices. However, routing considerations are not accounted for.

The SA-based CC layout generation method in [46] performs simultaneous placement and global routing, performing a search over a *pair sequence* representation of a CC placement. In each pair (u_i, u_j), u_i and u_j are symmetrically placed about the CC point, and pair sequence lists pairs in nonincreasing order of distance from the CC point. Perturbation of a pair sequence can result in a different CC layout with the same dimensions. A core step in [46] is routability analysis, which finds the overlapped channel spans among different connected capacitor groups. Next, the largest number of required routing tracks in a channel is minimized, attempting to achieve one track per channel. Finally, detailed routing (Fig. 7) first routes trunk wires in channels, then branch wires within the array by using breadth-first search, and lastly, bridge wires that symmetrically connect all trunk wires.

(a) $k=N$ (b) $k=N-1$ (c) placement P_1

Fig. 6: A 6-bit example of the chessboard placement method. The first two steps ($k = N = 6$ and $k = N-1 = 5$) and the final result, denoted by P_1 [40].

trunk wire
bridge wire
branch wire

Fig. 7: Routing topology of the net, n_3, consisting of two vertical trunk wires, one horizontal bridge wire, and many horizontal or vertical branch wires in [46].

In [27] an approach that constructs a minimum spanning tree (MST) to connect the top plates of all the disjoint connected components is presented. The approach defines a CP-sequence that encodes the unit capacitor size, routing topology, and routing patterns. A genetic algorithm is employed to find the best configuration of the CP sequence for both power minimization and parasitic matching, using the fitness function

$$\Phi = \left(\alpha \frac{C_{unit}}{C_{unit,max}} + \beta \frac{\text{DNL}_{\text{max}}}{0.5} + \gamma \frac{\text{INL}_{\text{max}}}{0.5} \right)^{-1} \quad (24)$$

where a penalty function is used so that if DNL_{max} or INL_{max} exceeds 0.5 LSB, it is set to ∞. Finally, a shielded routing problem is formulated as a small ILP that adds shields in a way that keeps capacitor ratios close to their ideal values.

IV. Common-Centroid Transistor Layout

Algorithms for CC layout of capacitor arrays are not directly applicable to transistor arrays, where considerations such as diffusion sharing and LDEs must be taken into account. CC layouts to minimize systematic variations in transistor arrays have been studied in [2], [47]–[51].

The works in [48], [49] present constructive algorithms to generate CC patterns for transistor arrays. Thermal effects are also considered for placement generation in [49]. In [47], the notion of dispersion, the degree to which the unit cells of a transistor are distributed throughout a layout, is used to compare layouts and methods for generating maximally dispersive layouts are presented. However, the proposed techniques can only be applied to arrays with two transistors. None of these methods addresses the routing problem, or the issue of diffusion sharing between transistors.

A framework for constructing a 2D CC array to maximize diffusion sharing is based on [2]. Representing the nodes as vertices and source-drain connections between transistor finger

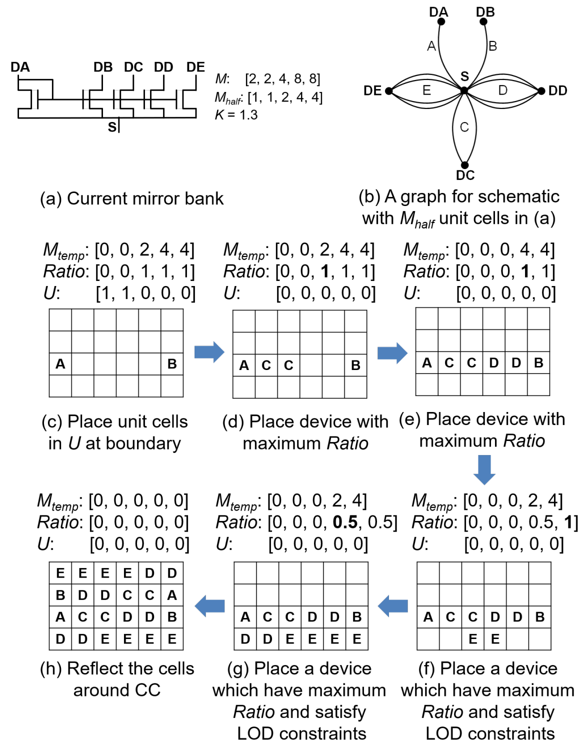


Fig. 8: (a) A current mirror bank. (b) Its M_{half} graph. (c)–(f) Steps in the proposed common-centroid algorithm. [19]

as edges, a "half diffusion graph," M_{half}, is first created by halving the number of edges between vertices. Fig. 8(a) shows a schematic of the example circuit consists of five devices, A, B, C, D, and E, whose multiplicity matrix $M = [2, 2, 4, 8, 8]$ represents, in the same order, the number of unit cells of these five devices. The graph for the circuit is shown in Fig. 8(b). An extension in [19] considers the number of unit cells is odd and full diffusion sharing is not possible. In this case one cell is placed at the edge of the layout. An Euler path is then found on this graph (in [2], this is done through expensive enumeration) to create half of the layout: this is then reflected about the CC point to create the full layout.

The work in [19], incorporated into ALIGN [52]–[54], introduces several improvements over prior methods. First, it improves upon the expensive enumeration of Euler paths in [2]. Second, it accounts for LDEs and parasitic mismatches in constructing the CC layout. Third, the routing method is made electromigration-aware and IR-drop-aware by creating a fishbone structure whose wire widths are optimized. The approach is illustrated in Fig. 8(c)–(f). At each step, cells of the device with the largest fraction (*Ratio*) of unplaced cells are added to the layout matrix. To improve dispersion, cells are placed alternately to the left and right of the CC point. To minimize LOD mismatch if a device has already been placed in the column (in a different row), another device is prioritized.

For example, in the circuit of Fig. 8(a), first, device C is selected: at this point no device can share the diffusion region, and C is a device with the highest *Ratio* value. Its placement in X is shown in Fig. 8(d). Thereafter, *Ratio* is updated, and device D, which now has the largest value in *Ratio*, is placed as shown in the figure. At this point, the row is filled and we move to the next row. The procedure is repeated until all cells are placed, as shown in Fig. 8(f)–(g).

In [50], a CC placement for FinFET technologies considering the impact of gate misalignment is studied. Due to gate misalignment, the position of the printed gate of a FinFET may deviate from the expected position, increasing the threshold voltage and decreasing the FinFET drain current. By carefully arranging the orientations of all FinFETs within a current mirror or a differential pair, the ratio of the drain current among different transistors in a current mirror or a differential pair can be perfectly matched [55]. A new quality metric is proposed for evaluating current ratio matching among transistors in a current mirror under gate misalignment and parasitic resistance in a CC array. The placement algorithm, focused on current mirror structures, is diffusion-sharing-aware and maximizes unit cell dispersion to optimize current ratio while maximizing the dispersion degree. Routing is performed using a parasitic-aware technique based on a minimum spanning tree method.

V. Is Common-Centroid Layout Always Necessary?

A. Impact of Layout on Performance

In [56], two issues that affect the performance of analog transistor array during layout are analyzed:

Layout-dependent effects (LDEs): LDEs induce shifts in transistor performance parameters stemming from their relative position in the layout, as described in Section II-A2. Fig. 9 shows three layouts (clustered, ID, and CC). From Eq. 11, each layout experiences different LOD variations. In the clustered layout, SA [SB] for the leftmost unit cell A is the same as SB [SA] for the rightmost cell B, resulting in the same LOD. A similar observation is made for the ID layout, but in the CC layout, from Eq. (11), LOD for the inner B cells differs from that for the outer A cells, causing mismatch.

Routing parasitic mismatch: From Fig. 9, the CC layout inherently shows a mismatch between the length of the drain/source connections (and hence the wire parasitics) for devices A and B. No such mismatch is seen for the ID or clustered layout. In FinFET technologies, where the wires have significant resistance, this can be a significant performance issue. As design rules specify unidirectional wire routing each layer, detours for parasitic matching are not possible, and moving to another layer involves vias that cause large resistances jumps, making resistance matching even harder.

The impact of parasitic mismatch and LOD is more critical for smaller devices. For larger devices, these effects can be avoided by changing device placement, e.g., in Fig. 9, mismatch can be reduced by using two rows of transistors, with A and B swapped in the second row, to ensure that both LOD and routing parasitics for A and B match even for CC.

The effective transconductance [57] captures the impact of interconnect parasitics in DPs on performance:

$$G_m = \frac{g_m(v_{in} - v_s)}{v_{in}} = \frac{g_m(v_{in} - i_{ac}R_s)}{v_{in}} \tag{25}$$

where v_{in} and v_s are the small-signal input and source voltages, respectively; g_m is the transistor transconductance; R_s is the parasitic resistance from the transistor source to its AC ground node (the point where small-signal currents cancel); and i_{ac} is the small-signal current through R_s.

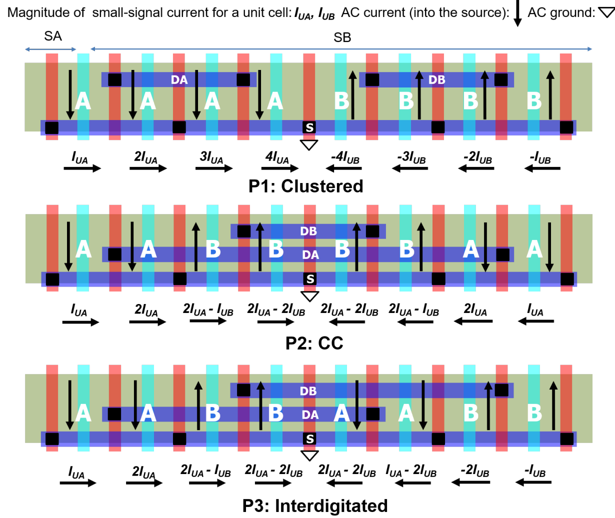


Fig. 9: Clustered, CC, and ID layouts with routing connections.

For a DP, each unit cell of device A carries a positive small-signal current of magnitude I_{UA} and device B carries a negative small-signal current of magnitude I_{UB}. The locations of AC ground and the AC currents in a DP are annotated in Fig. 9. For the clustered pattern, the current through R_s increases from the leftmost/rightmost unit cell (A/B) to the AC ground, and is larger than for CC or ID. Consequently, v_s is higher and G_m is lower (from (25)) than for CC or ID.

The effective small-signal currents through R_s are very similar for CC and ID, but due to R_s mismatch between device A and B, the CC pattern is inferior to the ID pattern [57]. Thus, the ID layout provides the best G_m, the CC layout is next best, and the clustered layout is the worst of the three.

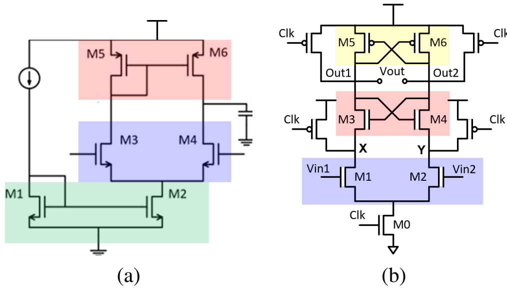


Fig. 10: Schematics: (a) 5T-OTA (b) StrongARM comparator

B. Evaluation of Circuit-Level Performance

We consider clustered, CC, and optimized layouts of a 5T-OTA and a StrongARM comparator, for two values of the correlation length, R_L: 10μm and 1000μm.

5T-OTA: The 5T-OTA in Fig. 10(a) uses a DP (M3, M4) [W/L = 46μm/14nm], an NMOS CM (M1, M2) [18.4μm/14nm], and a much smaller structure for the PMOS CM (M5, M6) [2.3μm/14nm]. Its input-referred offset is sensitive to mismatch [58]. The optimized OTA layout uses a CC pattern for the larger structures, the DP and the NMOS CM. These cells have 80 and 40 unit cells for each device, respectively, and the CC patterns for these have four rows that can match both LOD and parasitics. A clustered pattern is used for the small PMOS CM. The mean of the offset is affected by layout parasitics and LDEs, and the CC layout is worse than the optimized layout. The culprit in CC is the PMOS CM with four unit cells per device, arranged in a single row, which creates high parasitics and LOD mismatch.

The offset stdev is affected by both uncorrelated (u) and spatially correlated (s) variations. For correlation distance R_L = 1000μm, the total variations are dominated by u for all layout pattern. For R_L = 10μm, the clustered pattern is clearly worse. The optimized case has the best σ and good μ.

StrongARM comparator: The StrongARM comparator in Fig. 10(b) uses a DP (M1, M2) [6.1μm/14nm], an NMOS cross-coupled pair (CCP) (M3, M4) [3.1μm/14nm], a PMOS CCP (M5, M6) [1.6μm/14nm], and switches. Its dynamic input offset is sensitive to mismatch between X and Y [59]. All blocks in the comparator are small, and the optimized layout uses the clustered pattern. CC shows capacitance mismatch between X and Y, and ID has higher parasitics at these nodes.

The dynamic offset is a nonlinear function of V_{th} mismatch and parasitics [59]. Its mean is higher under CC due to parasitic mismatch and inherent LOD mismatch in the DP and CCP. Like the 5T-OTA, at R_L = 1000μm, μ and σ are similar to the u-only case. At R_L = 10μm, for the optimal clustered layout, spatial variations impact mismatch, and the nonlinear relationship with dynamic offset causes both μ and σ for the clustered case to degrade. For CC, spatial variations at both R_L values have modest effects: μ and σ are similar to u-only, but worse than the optimized case.

VI. Future Directions and Conclusion

This article has provided a survey of techniques used for CC layout for transistors and passives. Although a great deal of work has been performed in building CC layouts in the past, much work is still to be done. The ability to build low-power solutions depends greatly on developing capabilities of building well-matched low-capacitance structures with low parasitics, and this remains an open problem. There is early work on understanding when CC layouts are preferable to non-CC layouts, and vice versa, but further studies are necessary.

The advent of new technologies – FinFETs and gate-all-around FETs (GAAFETs/nanoribbons) – brings in a number of new challenges for CC layout. For example, in lower metal layers, all wires may be required to be on a grid, with constant pitch and width; all wires in lower metal layers must be unidirectional; the cost of "turning" from the horizontal to the vertical direction, or vice versa, can incur high via resistances. Moreover, MOM capacitors are greatly preferred over MIM structures as the cost of moving from lower metal layers to higher metal layers can similarly incur high resistances over via stacks. Device structures are susceptible to stress and require dummy placements to maintain stress [15], [60].

References

[1] A. Hastings, *The Art of Analog Layout*. Upper Saddle River, NJ: Prentice-Hall, 2001.
[2] D. Long, *et al.*, "Optimal two-dimension common centroid layout generation for MOS transistors unit-circuit," in *Proc. ISCAS*, pp. 2999–3002, 2005.
[3] Y. Abulafia and A. Kornfeld, "Estimation of FMAX and ISB in microprocessors," *IEEE TVLSI*, vol. 13, no. 10, pp. 1205–1209, 2005.

[4] L. T. Pang, *et al.*, "Measurement and analysis of variability in 45 nm strained-Si CMOS technology," *IEEE JSSC*, vol. 44, no. 8, pp. 2233–2243, 2009.

[5] P. Friedberg, *et al.*, "Modeling within-die spatial correlation effects for process-design co-optimization," in *Proc. ISQED*, pp. 516–521, 2005.

[6] K. J. Kuhn, *et al.*, "Process technology variation," *IEEE Trans. Electron Devices*, vol. 58, no. 8, pp. 2197–2208, 2011.

[7] K. Kuhn, *et al.*, "Managing process variation in Intel's 45nm CMOS technology," *Intel Technol. J.*, vol. 12, no. 2, pp. 93–109, 2008.

[8] M. Orshansky, *et al.*, "Impact of spatial intrachip gate length variability on the performance of high-speed digital circuits," *IEEE TCAD*, vol. 21, no. 5, pp. 544–553, 2002.

[9] B. E. Stine, *et al.*, "Analysis and decomposition of spatial variation in integrated circuit processes and devices," *IEEE T. Semicond. Manuf.*, vol. 10, no. 1, pp. 24–41, 1997.

[10] J. Xiong, *et al.*, "Robust extraction of spatial correlation," in *Proc. ISPD*, pp. 2–9, 2006.

[11] J. Xiong, *et al.*, "Robust extraction of spatial correlation," *IEEE TCAD*, vol. 26, no. 4, pp. 619–631, 2007.

[12] D. Sayed and M. Dessouky, "Automatic generation of common-centroid capacitor arrays with arbitrary capacitor ratio," in *Proc. DATE*, pp. 576–580, 2002.

[13] C.-W. Lin, *et al.*, "Mismatch-aware common-centroid placement for arbitrary-ratio capacitor arrays considering dummy capacitors," *IEEE TCAD*, vol. 31, no. 12, pp. 1789–1802, 2012.

[14] J. V. Faricelli, "Layout-dependent proximity effects in deep nanoscale CMOS," in *Proc. CICC*, pp. 1–8, 2010.

[15] M. G. Bardon, *et al.*, "Layout-induced stress effects in 14nm & 10nm FinFETs and their impact on performance," in *Proc. IEEE Symp. on VLSI Technology*, pp. T114–T115, 2013.

[16] T. Hook, *et al.*, "Lateral ion implant straggle and mask proximity effect," *IEEE Trans. Electron Devices*, vol. 50, no. 9, pp. 1946–1951, 2003.

[17] K. W. Su, *et al.*, "A scaleable model for STI mechanical stress effect on layout dependence of MOS electrical characteristics," in *Proc. CICC*, pp. 245–248, 2003.

[18] P. G. Drennan, *et al.*, "Implications of proximity effects for analog design," in *Proc. CICC*, pp. 169–176, 2006.

[19] A. K. Sharma, *et al.*, "Performance-aware common-centroid placement and routing of transistor arrays in analog circuits," in *Proc. ICCAD*, 2021.

[20] P.-W. Luo, *et al.*, "Impact of capacitance correlation on yield enhancement of mixed-signal/analog integrated circuits," *IEEE TCAD*, vol. 27, pp. 2097–2101, 2008.

[21] D. J. Frank, *et al.*, "Monte Carlo modeling of threshold variation due to dopant fluctuations," in *Proc. VLSI Symp.*, pp. 171–172, 1999.

[22] P. Oldiges, *et al.*, "Modeling line edge roughness effects in sub 100 nanometer gate length devices," in *Proc. SISPAD*, pp. 131–134, 2000.

[23] M. J. Pelgrom, *et al.*, "Matching properties of MOS transistors," *IEEE JSSC*, vol. 24, no. 5, pp. 1433–1439, 1989.

[24] V. Tripathi and B. Murmann, "Mismatch characterization of small metal fringe capacitors," *IEEE TCAS-I*, vol. 61, no. 8, pp. 2236–2242, 2014.

[25] H. Omran, *et al.*, "Matching properties of femtofarad and sub-femtofarad MOM capacitors," *IEEE TCAS-I*, vol. 63, no. 6, pp. 763–772, 2016.

[26] M. P.-H. Lin, *et al.*, "Parasitic-aware sizing and detailed routing for binary-weighted capacitors in charge-scaling DAC," in *Proc. DAC*, pp. 1–6, 2014.

[27] M. P.-H. Lin, *et al.*, "Parasitic-aware common-centroid binary-weighted capacitor layout generation integrating placement, routing, and unit capacitor sizing," *IEEE TCAD*, vol. 36, no. 8, pp. 1274–1286, 2017.

[28] T.-W. Wang, *et al.*, "Late breaking results: Automatic adaptive MOM capacitor cell generation for analog and mixed-signal layout design," in *Proc. DAC*, 2020.

[29] P. J. A. Harpe, *et al.*, "A 26uW 8 bit 10 MS/s asynchronous SAR ADC for low energy radios," *IEEE JSSC*, vol. 46, no. 7, pp. 1585–1595, 2011.

[30] J.-Y. Lin and C.-C. Hsieh, "A 0.3 V 10-bit 1.17 f SAR ADC with merge and split switching in 90 nm CMOS," *IEEE TCAS-I*, vol. 62, no. 1, pp. 70–79, 2014.

[31] C.-C. Liu, *et al.*, "A 10-bit 50-MS/s SAR ADC with a monotonic capacitor switching procedure," *IEEE JSSC*, vol. 45, no. 4, pp. 731–740, 2010.

[32] G.-Y. Huang, *et al.*, "A 10b 200MS/s 0.82 mW SAR ADC in 40nm CMOS," in *Proc. A-SSCC*, pp. 289–292, 2013.

[33] S.-H. Wan, *et al.*, "A 10-bit 50-MS/s SAR ADC with techniques for relaxing the requirement on driving capability of reference voltage buffers," in *Proc. A-SSCC*, pp. 293–296, 2013.

[34] W. Kim, *et al.*, "A 0.6 V 12 b 10 MS/s low-noise asynchronous SAR-assisted time-interleaved SAR (SATI-SAR) ADC," *IEEE JSSC*, vol. 51, no. 8, pp. 1826–1839, 2016.

[35] N.-C. Chen, *et al.*, "High-density MOM capacitor array with novel mortise-tenon structure for low-power SAR ADC," in *Proc. DATE*, pp. 1757–1762, 2017.

[36] P.-Y. Chou, *et al.*, "Matched-routing common-centroid 3-D MOM capacitors for low-power data converters," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 25, no. 8, pp. 2234–2247, 2017.

[37] J.-E. Chen, *et al.*, "Placement optimization for yield improvement of switched-capacitor analog integrated circuits," *IEEE TCAD*, vol. 29, no. 2, pp. 313–318, 2010.

[38] P.-Y. Chou, *et al.*, "An Integrated Placement and Routing for Ratioed Capacitor Array based on ILP Formulation," in *Proc. VLSI-DAT*, pp. 1–4, 2016.

[39] R. S. Soin, *et al.*, *Analogue-Digital ASICs: Circuit Techniques, Design Tools and Applications*. Stevenage, U.K.: Peter Peregrinus Ltd., 1991.

[40] F. Burcea, *et al.*, "A new chessboard placement and sizing method for capacitors in a charge-scaling DAC by worst-case analysis of nonlinearity," *IEEE TCAD*, vol. 35, no. 9, pp. 1397–1410, 2015.

[41] C.-C. Huang, *et al.*, "Performance-driven unit-capacitor placement of successive-approximation-register ADCs," *ACM TODAES*, vol. 21, no. 1, pp. 1–17, 2015.

[42] C.-C. Huang, *et al.*, "PACES: A partition-centering-based symmetry placement for binary-weighted unit capacitor arrays," *IEEE TCAD*, vol. 36, no. 1, pp. 134–145, 2016.

[43] Y. X. Ding, *et al.*, "PASTEL: Parasitic matching-driven placement and routing of capacitor arrays with generalized ratios in charge-redistribution SAR-ADCs," *IEEE TCAD*, vol. 39, no. 7, pp. 1372–1385, 2019.

[44] N. Karmokar, *et al.*, "Constructive common-centroid placement and routing for binary-weighted capacitor arrays," in *Proc. DATE*, 2022. (to appear).

[45] Q. Ma, *et al.*, "Analog placement with common centroid constraints," in *Proc. ICCAD*, pp. 579–585, 2007.

[46] M. P.-H. Lin, *et al.*, "Common-centroid capacitor layout generation considering device matching and parasitic minimization," *IEEE TCAD*, vol. 32, no. 7, pp. 991–1002, 2013.

[47] C. C. McAndrew, "Layout symmetries: Quantification and application to cancel nonlinear process gradients," *IEEE TCAD*, vol. 36, no. 1, pp. 1–14, 2016.

[48] V. Borisov, *et al.*, "A novel approach for automatic common-centroid pattern generation," in *Proc. SMACD*, pp. 1–4, 2017.

[49] M. P.-H. Lin, *et al.*, "Thermal-driven analog placement considering device matching," *IEEE TCAD*, vol. 30, no. 3, pp. 325–336, 2011.

[50] P. H. Wu, *et al.*, "Parasitic-aware common-centroid FinFET placement and routing for current-ratio matching," *ACM TODAES*, vol. 21, no. 3, p. 39, 2016.

[51] M. F. Lan and R. Geiger, "Gradient sensitivity reduction in current mirrors with non-rectangular layout structures," in *Proc. ISCAS*, pp. 687–690, 2000.

[52] K. Kunal, *et al.*, "ALIGN: Open-source analog layout automation from the ground up," in *Proc. DAC*, pp. 77–80, 2019.

[53] T. Dhar, *et al.*, "ALIGN: A system for automating analog layout," *IEEE Des. Test*, 2021.

[54] "ALIGN: Analog layout, intelligently generated from netlists," Software repository, accessed November 1, 2021. https://github.com/ALIGN-analoglayout/ALIGN-public.

[55] M. Fulde, *et al.*, "Analog design challenges and trade-offs using emerging materials and devices," in *Proc. ESSCIRC*, pp. 123–126, 2007.

[56] A. K. Sharma, *et al.*, "Common-centroid layouts for analog circuits: Advantages and limitations," in *Proc. DATE*, 2021.

[57] B. Razavi, *Design of Analog CMOS Integrated Circuits*. New York, NY: McGraw-Hill, 2nd ed., 2016.

[58] P. R. Kinget, "Device mismatch and tradeoffs in the design of analog circuits," *IEEE JSSC*, vol. 40, no. 6, pp. 1212–1224, 2005.

[59] B. Razavi, "The StrongARM latch [a circuit for all seasons]," *IEEE Solid-St. Circ. Mag.*, vol. 7, no. 2, pp. 12–17, 2015.

[60] S. K. Marella, *et al.*, "Optimization of FinFET-based circuits using a dual gate pitch technique," in *Proc. ICCAD*, pp. 758–763, 2015.