



Invited Paper: The Scope and Challenges of Scaling in Advanced Technologies

Chung-Kuan Cheng, Bill Lin, Byeonggon Kang, Yucheng Wang
University of California, San Diego
La Jolla, California, USA
ckcheng@ucsd.edu, billlin@eng.ucsd.edu, {b8kang, yuw132}@ucsd.edu

Abstract

Design Technology Co-optimization (DTCO) and System Technology Co-optimization (STCO) have become essential techniques to sustain Moore's law, while the geometric scaling has slowed down in the last decade. With new technology nodes are on the horizon, the anticipated scaling boost faces a potential hindrance known as the "pin density wall". This challenge arises from the shrinking cell area and the intricate 3D structure of advanced technology nodes, which limits the options for pin accessibility. Consequently, the advantages of cell area shrinkage in given existing advanced architectures may not translate well to block-level design. To address this issue, additional design methodologies regarding routability need to be explored.

In this work, we will describe the scope and potential benefits of different design knobs for standard cell design, device architectures, and block-level placement and route. In addition, we will cover the challenges and future research directions by investigating physical space constraints, cell design automation flow, and existing design tool limitations.

CCS Concepts: • Hardware → Best practices for EDA.

Keywords: Design Technology Co-Optimization, System Technology Co-Optimization, Standard-Cell Layout, Pin-Density Wall

ACM Reference Format:

Chung-Kuan Cheng, Bill Lin, Byeonggon Kang, Yucheng Wang. 2023. Invited Paper: The Scope and Challenges of Scaling in Advanced Technologies. In *2023 ACM International Workshop on System-Level Interconnect Pathfinding (SLIP '23)*, November 2, 2023, San Francisco, CA, USA. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3632409.3632841>



This work is licensed under a Creative Commons Attribution International 4.0 License.

SLIP '23, November 2, 2023, San Francisco, CA, USA

© 2023 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0474-1/23/11.

<https://doi.org/10.1145/3632409.3632841>

1 Introduction

Moore's Law [16], often regarded as a self-fulfilling prophecy, continues to exert a profound impact on the semiconductor industry by driving relentless advancements in computational capabilities and lithographic processes. Dennard's scaling prediction [9], the "More than Moore" roadmap [11], and a recent new metric proposal [17] serve as crucial drivers for the sustained expansion of the market and the escalating demands for research.

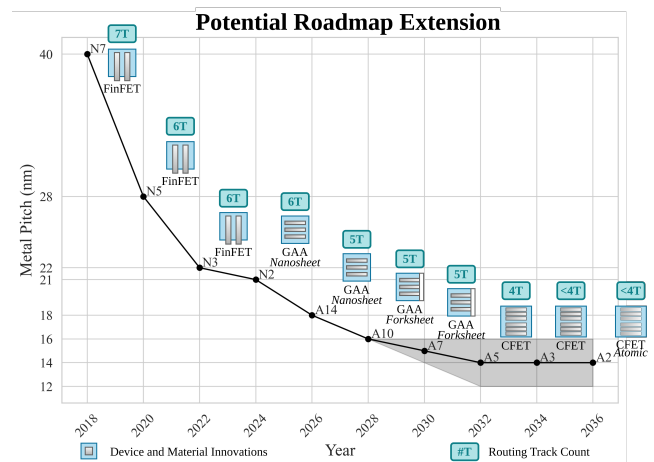


Figure 1. Latest scaling roadmap proposed by IMEC. Scaling continues with advanced device architecture and reduced horizontal tracks (cell height). However, the metal pitch will soon be saturated by physical limitations.

Figure 1 illustrates a potential roadmap extension [1]. We can recognize three key features regarding advanced technology nodes:

- Metal pitch scaling slows down from 22 nm at the N3 node in 2022 and will saturate at a range of 12-16 nm at the A5 node in 2037. In the meantime, the contacted poly pitch (CPP), which determines the cell width, scales at a much slower pace from 48 nm in 2022 to 38 nm in 2037 [2].
- We reduced the standard cell height by cutting down the number of horizontal tracks from 7T(Track) at the N7 node in 2018 to smaller numbers, which are expected to be smaller than 4T at A3 in 2034.

- We adopt advanced device architectures with a focus on scaling in 3D.

The above trend leaves some implications as follows:

- The reduction of cell height in terms of the number of horizontal routing tracks enforces the increase of extra numbers of CPPs to accommodate needed routing resources. Therefore, we have observed a huge amount of design technology co-optimization (DTCO) efforts to tune the conditional design rule to recover the cell width.
- We observe more routing usage in upper metal layers due to limited routing tracks (smaller cell area).

The overall effect leaves a denser pin density¹ with fewer routing resources for the block level layout². This kind of routing complexity can potentially lead to more routing congestion and a larger block area. Therefore, even though cell height and CPP are further reduced in recent technology, the overall cell area and block area do not reduce at the same scale.

Figure 2 demonstrates the flow chart of DTCO [18] that performs co-optimization between the process team and design team. In order to overcome the limitations of the one-way development in the past, both teams provide mutual feedback to each other to improve design rules and spice models.

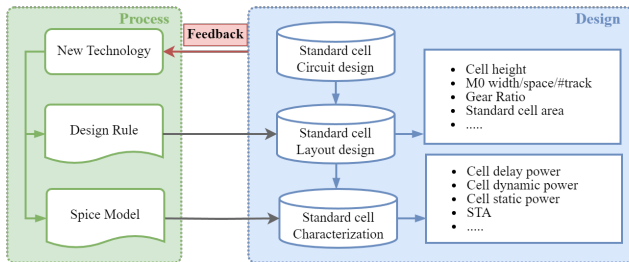


Figure 2. DTCO flow chart. Layout designers and process engineers are allowed to provide additional feedback to each other to work around design difficulties. Such co-optimization effort fosters a more future-proofing and efficient design workflow.

System technology co-optimization (STCO) [21] is introduced to further boost such a scaling challenge. STCO’s main goal is to discover new requirements for implementing future design knobs and to realize this technology through a system-level approach. One example is the Back-Side Power Delivery Network (BSPDN) [20]. To implement BSPDN, we use sophisticated process technologies like Through Silicon Via (TSV)

that penetrates the wafer with backside metals. STCO establishes a system environment connecting advanced process technology and logic scaling.

By combining efforts on DTCO/STCO, more progressive research in cell design is being explored. To systematically streamline the complicated design process, Cheng et al. introduce a Satisfiability-Modulo-Theory (SMT) based cell layout synthesis approach [6][7][14] to various transistor architectures. Recent development in PROBE3.0 [8] automates custom Process Design Kit (PDK) generation and incorporates Power, Performance, and Area (PPA) and IR drop prediction for DTCO. Combining these design methodologies with machine learning frameworks [5][8] enables a fast turnaround time for evaluating the effect of different cell architectures and technology parameters. With this established framework, we hope to observe and pinpoint the most influential parameters for the final cell designs.

However, contrary to the promised “scaling boost” brought by advanced technology node, we observe that these advancements do not always contribute to the overall PPA. This is because as the unit tile³ of a standard cell shrinks, the number of pins remains to be the same. Under the effect of design rules and limited routing resources, cells cannot be placed and routed without design rule violation. Hence, at the block level, the pin density (per square micron) cannot be further increased and we see limited performance gain for advanced nodes. We refer to such a challenge as *pin density wall*.

Our main contributions to this work are shown as follows:

- We present CFET and VFET for advanced transistor architectures and their potential area gain at the cell level.
- We showcase *pin density wall* through experiment result on VFET across three block-level designs, which emphasizes the area reduction in advanced devices exacerbate the pin density.
- We introduce design knobs that are not covered by previous exploratory frameworks from both design and technology perspectives. These additional knobs focus on pin optimization and could potentially mitigate such pin density problems to improve the overall block-level PPA.

In light of the well-established frameworks for DTCO and the maturation of pre-existing design parameters, we focus on the investigation of various design approaches regarding pin accessibility. We introduce design knobs that hold the potential for designers to navigate various standard cell designs for better block-level PPA. Integrating these design

¹We define pin density on the frontside (or Back-End-Of-Line (BEOL)) to be computed as, $\frac{\#Pins}{Area}$.

²The placement of and routing among standard cells to construct macro cell blocks

³In a standard cell layout, a unit tile is a space that spans between adjacent contacted poly along the x-axis and between power and ground along the y-axis. Unit tile area is computed by CPP $\times$ Cell Height, given by the design parameter.

knobs into the existing cell layout automation framework can also significantly expedite the exploration process.

Section 2 describes the fundamental challenges in standard cell designs and the various trends seen in future transistor architecture candidates. Section 3 demonstrates the pin density bottlenecks in both cell-level and block-level across different transistor architectures through experiments. Finally, Section 4 details pin-related design knobs and methodologies that could contribute to the improvement of block PPA.

2 Preliminaries

Standard cells, also known as digital logic gates, are the fundamental building blocks of modern integrated circuits. The standard cell library design begins with a tech-independent circuit generation. There are numerous combinational and sequential logic gates such as flip-flops (FF), latches, clock gates, etc. Designing an optimal layout for each circuit in a given technology node plays a pivotal role in improving block PPA. In recent technology, each metal layer is assumed to be unidirectional (i.e., each adjacent metal layer is orthogonal to their adjacent layers) for design efficiency and manufacturability[12]. To optimize cell layout design, we aim to explore the ideal transistor placement and routing path. This not only minimizes wire length and cell area but also enhances the accessibility of IO pins.

2.1 Design Rule Constraint for Routing

Designing standard cell layouts must adhere to stringent design rules to maximize yield. For instance, conditional rules like End-of-Line Rule (EOL) dictate minimum distances between adjacent metals, while the Overlap Rule (OVL) mandates specific metal-via overlap areas. As technology advances, these strict rules demand efficient metal placement to maintain layout quality.

Figure 3 demonstrates complicated metal interactions as we migrate a design from an old technology (Figure 3(a)) to a newer technology (Figure 3(b,c)) with scaled CPP and reduced horizontal track numbers. The dotted circle shows the active design rules, and the red bar the required distance. In Figure 3(b), the decreased CPP triggers design rule violations (DRV) (shown with the red cross) due to closer M0 metal spacing. Thus, in Figure 3(c), we shift the metal up by one track which causes an EOL violation. One solution is to move the metal to an upper layer which becomes a blockage in the block level routing and adversely affects pin accessibility. Thus, for efficient standard cell designs, a comprehensive approach is essential, considering both intra-cell and inter-cell routing.

2.2 Emerging necessity for 3D transistors

The cell height reduction demands the move to the third dimension for the devices [3][15][19]. In the following subsections, we describe two advanced device architectures.

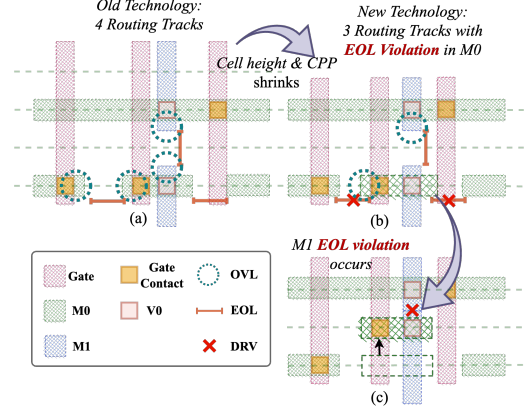


Figure 3. Migrating old Technology (a) to new technology (b,c) with less routing resources. (a) Layout in a previous technology node using 4 tracks; (b) Advanced node with 3 tracks and scaled CPP. The reduced spacing causes DRVs shown in red crosses; (c) Moving the middle M0 up by track causes a new DRV.

2.2.1 Complementary-FET. Complementary-FET (CFET) [7] structure stacks the PFET and NFET on top of each other (Figure 4). Thus, for the same net, stacked sources/drains or P/N gates can be shorted with a via (Figure 4(a,d)). Otherwise, the one at the bottom needs to extend to the upper layers with a long via and the one on top has to spare the space of a track for the long via to go through (Figure 4(b,c)). Hence, the layout has to observe the structure for inter- and intra-cell routability.

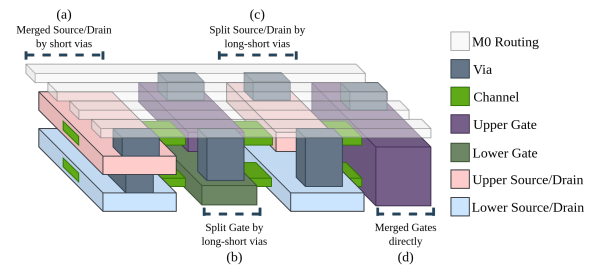


Figure 4. Complementary-FET structure where P/N FETs stack together. (a) Short via is used for connecting S/D with the same net. (b)&(c) If nets are different, S/G/D are split and a long-short via is used to connect M0 and contact. (d) stacked P/N gates of the same net are merged directly.

2.2.2 Vertical Gate-All-Around FET. Vertical Gate-All-Around FET (VFET) [6] structure allocates S/G/D into three metal layers (e.g., M0, M1, M2) and stacks the transistors on top of one another as shown in Figure 5. We refer to the three metal layers that contain a transistor as a *Tier*. The nanowires are used as vias to make channels. Here, we assume that routing is performed through the same metal

layers alongside S/G/D. Intra-cell routing for S/D is done bi-directionally while G remains to be uni-directional. BEOL remains to be uni-directional to match the preferred direction for global routing. Hence, VFET shrinks its placement footprint locally while adding more complicity to routing.

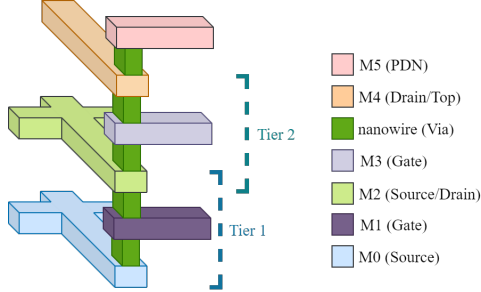


Figure 5. An example for a 2-tier Vertical Gate-All-Around FET (VFET) structure Source/Gate/Drain are separated across different interconnect layers. For M0 and M2, we allow bi-directional routing. In this example, IO pins are assigned at M4.

3 Area Scaling Bottlenecks in Advanced Technology Node

The goal of our experiment is to use pin density to demonstrate the limitation of scaling in advanced technology architectures. Pin density is defined at the cell level (with cell area) and at the block level (with block area). At the cell level, high pin density implies densely placed pins in a limited cell area. At the block level, high pin density per square micron represents a densely routed area which leads to better area utilization. However, using existing commercial layout tools and a given technology, high pin density at the cell level may not be able to improve the block-level pin density.

In the following, we first present a cell-level comparison between FinFET, CFET and VFET in terms of local pin density. We then present a block-level experiment using multiple-tiered VFET [6]. We list the following hypotheses for the test:

1. At the cell level, the area will reduce from FinFET to CFET and VFET architecture as the transistors are stacked in the third dimension.
2. We reduce the cell height with fewer horizontal track numbers in CFET, we need more upper metal layers for intra-cell routing.
3. As cell area shrinks in advanced device architecture, pin density will increase in block-level layout.
4. Cell area reduction will lead to a smaller block-level layout area.

Our cell layout design is synthesized using an SMT-based cell layout generation framework with the same netlist and metal pitch setting. The detailed settings are as follows:

- We keep CPP and metal pitches to be the same across all technologies.
- For cell level comparison, we keep the horizontal track count to be similar across different architectures, while the routing pitch remains to be the same: 4.5 tracks for FinFET, and for CFET[10], and 1 tier and 5 tracks for VFET[13].
- For block level comparison, we use from 1 to 4 tiers and 6 Tracks for VFET [6].

Our target block-level netlist is Advanced Encryption Standard (AES), M0-Core and M1-Core design.

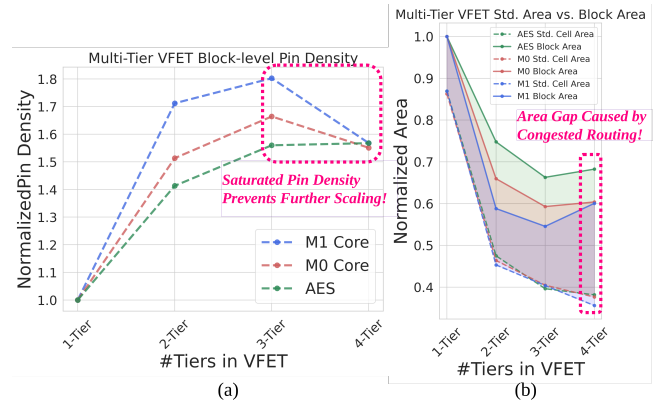


Figure 6. Pin density fails to increase consistently as cell-area reduces from 1-tier to 4-tier VFET architecture [6]. (a) Block-level pin density saturates at 3-tier and drops at 4-tier VFET architecture. (b) The area gap between the block area and the standard cell area increases as the tier count increases in VFET. More space is used for routing and filler cells, which degrades the block-level PPA.

Table 1 presents the comparison at the cell level for FinFET, CFET and VFET structures. The first column lists the cells. the second column shows that the cell area reduces from FinFET to CFET due to reduced cell height (track numbers) and even smaller to VFET due to reduced cell width. The third column shows that the total metal length reduces as the cell area decreases. The fourth column depicts that the number of M2 tracks increase as we can stack devices in the third dimension. The last column describes the cell pin density increases as an inverse function of the cell area.

Figure 6 shows a block-level comparison on VFET. The pin density increases from 1-tier to 3-tier but drops at 4-tier VFET structure (Figure 6(a)). In (Figure 6(a)), we show that the standard cell area decreases as the tier number increases. However, the total block area saturates at 3-Tier. The increased area gap between the block area and the standard cell area hints more space is taken to resolve the routing congestion. We refer to this phenomenon as *pin density wall*. Constrained by design rules and limited routing resources at

Table 1. Comparison table between cell metrics for FinFET/CFET/VFET given same design parameters. (CPP = Number of Contacted Poly, Cell Footprint = #Track $\times$ CPP, Metal Length = Number of edges occupied with each Via and M2 grid costs four times more, #M2 Track=Number of used tracks, Pin Density=#Pin / Cell Footprint)

Cell Name	#Pin	#FET	#Net	CPP & Cell Footprint						Metal Length (#Edge)			#M2 Track			Pin Density		
				FinFET	CFET	VFET	FinFET	CFET	VFET	FinFET	CFET	VFET	FinFET	CFET	VFET	FinFET	CFET	VFET
AND2x2	3	6	7	6	27	6	27	4	20	75	60	95	0	0	2	11.11%	11.11%	15.00%
AND3x1	4	8	9	6	27	6	27	4	20	91	68	116	0	0	4	14.81%	14.81%	20.00%
AND3x2	4	8	9	7	31.5	7	31.5	5	25	97	76	113	0	0	3	12.70%	12.70%	16.00%
AOI21x1	5	6	8	9	40.5	9	40.5	6	30	197	142	131	1	0	2	12.35%	12.35%	16.67%
AOI22x1	2	8	10	14	63	11	49.5	8	40	311	255	176	1	1	2	3.17%	4.04%	5.00%
BUFx2	2	4	5	5	22.5	5	22.5	3	15	61	40	67	0	0	2	8.89%	8.89%	13.33%
BUFx3	2	4	5	6	27	6	27	4	20	82	53	115	0	0	1	7.41%	7.41%	10.00%
BUFx4	2	4	5	7	31.5	7	31.5	5	25	88	59	136	0	0	1	6.35%	6.35%	8.00%
BUFx8	2	4	5	12	54	12	54	10	50	149	105	224	0	0	1	3.70%	3.70%	4.00%
DFFHQNx1	3	24	17	19	85.5	16	72	12	60	613	182	277	2	0	3	3.51%	4.17%	5.00%
FAx1	5	24	17	14	63	14	63	12	60	420	379	254	3	2	2	7.94%	7.94%	8.33%
INVx1	2	2	4	3	13.5	3	13.5	2	10	44	23	22	0	0	0	14.81%	14.81%	20.00%
INVx2	2	2	4	4	18	4	18	2	10	38	29	42	0	0	2	11.11%	11.11%	20.00%
INVx4	2	2	4	6	27	6	27	4	20	65	48	107	0	0	1	7.41%	7.41%	10.00%
INVx8	2	2	4	10	45	10	45	8	40	121	92	191	0	0	1	4.44%	4.44%	5.00%
NAND2x1	3	4	6	6	27	6	27	4	20	79	74	79	0	0	0	11.11%	11.11%	15.00%
NAND2x2	3	4	6	10	45	10	45	8	40	140	131	139	0	0	0	6.67%	6.67%	7.50%
NAND3x1	4	6	8	11	49.5	11	49.5	9	45	152	149	166	0	0	0	8.08%	8.08%	8.89%
NAND3x2	4	6	8	21	94.5	21	94.5	18	90	305	286	309	0	0	0	4.23%	4.23%	4.44%
NOR2x1	3	4	6	6	27	6	27	4	20	79	74	79	0	0	0	11.11%	11.11%	15.00%
NOR2x2	3	4	6	10	45	10	45	8	40	140	131	139	0	0	0	6.67%	6.67%	7.50%
NOR3x1	4	6	8	11	49.5	11	49.5	9	45	152	148	166	0	0	0	8.08%	8.08%	8.89%
NOR3x2	4	6	8	21	94.5	21	94.5	18	90	304	283	309	0	0	0	4.23%	4.23%	4.44%
OAI21x1	4	6	8	11	49.5	9	40.5	6	30	247	146	137	1	0	2	8.08%	9.88%	13.33%
OAI22x1	5	8	10	14	63	11	49.5	8	40	311	240	176	1	1	2	7.94%	10.10%	12.50%
OR2x2	3	6	8	6	27	6	27	4	20	75	60	95	0	0	2	11.11%	11.11%	15.00%
OR3x1	4	8	9	6	27	6	27	4	20	91	68	116	0	0	4	14.81%	14.81%	20.00%
OR3x2	4	8	9	7	31.5	7	31.5	5	25	97	76	113	0	0	3	12.70%	12.70%	16.00%
XNOR2x1	3	10	9	12	54	11	49.5	8	40	274	220	190	1	1	1	5.56%	6.06%	7.50%
XOR2x1	3	10	9	12	54	11	49.5	8	40	276	214	190	1	1	1	5.56%	6.06%	7.50%
Average	-	-	-	9.73	43.80	9.30	41.85	7.00	35.00	172.47	130.37	148.97	0.37	0.20	1.40	8.52%	8.74%	11.33%

frontside, cells cannot be densely placed and routed. Therefore, dense pin access induced by stacked 3D structures in Multi-Tier VFET causes routability issues, which hinder the benefit of cell area reduction.

4 Knobs for improving Routability

To overcome the constraints posed by design rules and pin density limitations, we discuss some "design knobs" to mitigate the routing challenges inherent in standard cell layout designs. Integrating these techniques into either a manual cell design process or a cell automation workflow can simplify routing complexities and enhance the block-level PPA.

4.1 Multi-Row Cells with Shared Gates for Routing Complexity Reduction

We use a multi-row layout to compensate for the routing congestion caused by cell height (horizontal track numbers) reduction. In Figure 7, we use an X2 (2 driver strength) cell and a FF to demonstrate the benefits of extending the cell layout from a single row to double rows. As the layout extends from a single row (Figure 7(a,c)) to a double row (Figure 7(b,d)), we have more routing tracks for intra-cell routing. We can even use poly to make the signal connection between two rows. Such a between-the-rows connection simplifies the

intra-cell routing and leaves space for inter-cell routing at the block level.

4.2 Circuit Topology Optimization

We transform the circuit topology such as net splitting and transmission gates to optimize the cell layout. Figure 8 shows that we can eliminate two M0, two M1, and one M2 by separating two nets when making a 2x driver cell, NAND2_X2. Figure 8(a) illustrates the original circuit diagram and layout that uses two copies of transistors to double the driving strength. Figure 8(b) shows that we split the net that connects the doubled devices into two. The transformation erases the need to make the connection and thus simplifies the routing. Discovering this kind of circuit topology is beneficial for routing blockage reduction.

4.3 Gate Pins for Improved Routability

We use gate pins to improve the flexibility of pin assignments. Fixed metal pin positions could be a burden on the global router at the P&R stage. Figure 9 illustrates the connection between two cells. In Figure 9(a), we fix pins on M0 layer. Thus, a metal segment on the third horizontal track is used to connect between the two cells. Figure 9(b) allows

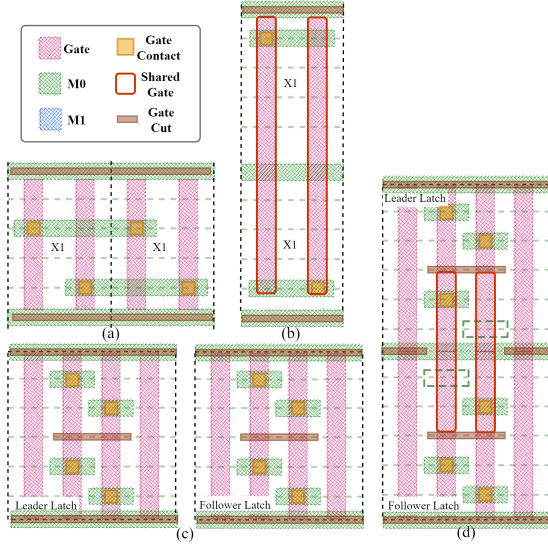


Figure 7. Single row vs. double row cells. (a) X2 cell in a single row; (b) Double row cell architecture that connects signals via poly between two rows; (c) Single row flip-flop (FF); (d) Double row FF with less metal usage.

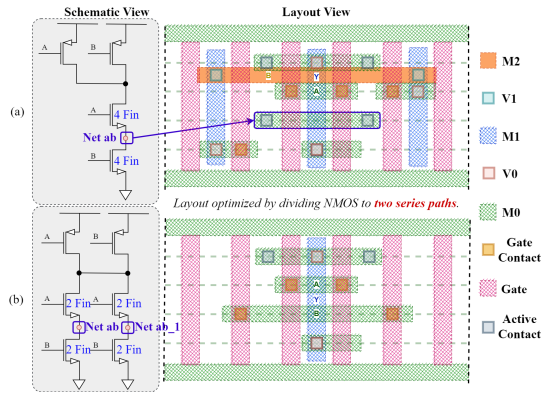


Figure 8. Two types of NAND2_X2 circuit topology and their corresponding layouts. (a) "Net ab" is connected via an M1 segment. (b) "Net ab" is split into "Net ab" and "Net ab_1" to simplify the layout.

pin openings on the gate. The extension of the metal segment from the right cell can make a direct connection. The flexibility of pin position reduces metal usage and overall routing congestion.

4.4 Pin Spacing in Advanced Node

While the IO pin count remains the same for logical cell designs, the continuous shrinkage in cell area in advanced device architecture stresses the need for pin assignment optimization. This trend is evident in Table 1, where deduction in track height significantly amplifies higher metal utilization. Any metal routing blockage, whether within the current cell

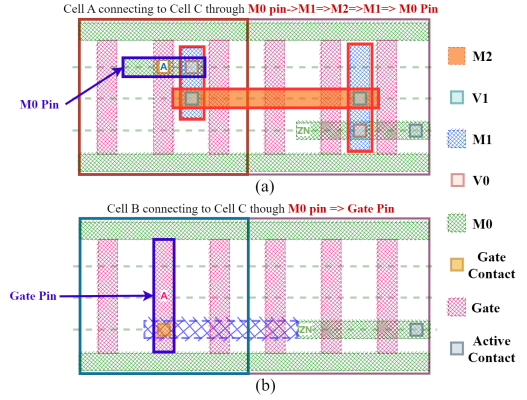


Figure 9. An example of fixed metal pins vs. gate pins. (a) Fixed M0 pin of INV_X1 design. Routing needs to be detoured from M0 all the way to M2, which is undesirable. (b) Gate pin for INV_X1. Routing can be directly connected with one M0 metal with a gate pin.

or adjacent cells, blocks the available track to route, rendering the pin hard to access. Figure 10 illustrates the necessity of separating pins not only to avoid existing metal patterns but also to consider the routing scenarios of adjacent pins, highlighting the critical role of this approach.

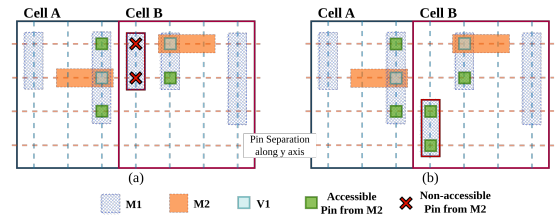


Figure 10. (a) Abutment between Cell A and Cell B causes inaccessible pins (from M2) in Cell B on M1. (b) Separating pins along the y-axis allows the pins to be accessible from M2.

4.5 Backside Routing for Routing Blockage Reduction

Previous Front Side Power Delivery Network (FSPDN) occupies routing tracks for power delivery. At the P&R stage, the tool has to place and route cells while avoiding PDN tracks. Recent developments in BSPDN and PowerVia [4] redistribute routing burden underneath the substrate.

We can expand this concept, and utilize Backside Metal (BM) (shown in Figure 11(c)) for intra-cell routing with a vertical structure of TSV and BM layer. Given the schematics of the circuit in Figure 11(a), the interconnecting metal highlighted in red in Figure 11(b) blocks the entire M0 track for routing. Figure 11(d) shows that the layout pattern can be improved by using empty tracks created through backside

routing. If all intra-cell routings can be removed using BM while only inter-cell routings remain at the frontside, routing blockage can be drastically reduced in P&R stage.

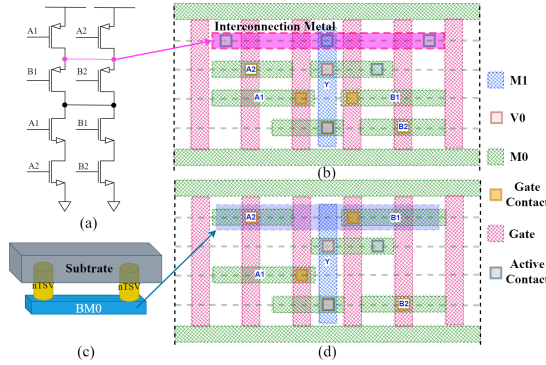


Figure 11. (a) AOI22_X1 circuit and interconnection net between series pmos. (b) AOI22_X1 layout and interconnection M0 layer. (c) Backside Metal0(BM0) that could substitute for the M0 layer. (d) Pin separation using newly acquired M0 track by using BM0.

4.6 Gear Ratio & Cell Legalization Optimization

Gear Ratio is defined as the ratio between CPP and M1 pitch. In the early stage of the DTCO process, the design team explores various gear ratios (GR) for each node in given design rule conditions. In recently advanced nodes, 2:3 GR with a tighter M1 pitch is preferred as it provides more routing resources than 1:1 GR. Nevertheless, utilizing a GR with a tighter M1 pitch generates offsets⁴ between the gate and M1 track, consequently restricting the locations where the cell can be placed.

In the global placement stage, standard cells are provisionally positioned with some overlapping. Subsequently, the tool relocates the instances to nearby locations to resolve this overlap, a step known as legalization. However, if a cell cannot be placed freely, its post-legalization location may diverge from its global placement location, potentially leading to routing congestion. Therefore, being able to legally place cells anywhere reduces the routing complexity and can lead to better PPA.

Offering multiple offset versions of cells simultaneously enhances placement possibility (PP) for GR in tighter M1 pitches such as 2:3 and 3:4. Commercial P&R tools can do legalization by having different options in all positions through an electrically equivalent (EEQ) swapping methodology by grouping multiple offset cells. Figure 12(a) and (b) shows an example of cells with multiple offset versions.

A methodology that avoids utilizing metal for potential PDN metal locations increases the likelihood of successful placement. Utilizing the characteristic of periodic PDN

⁴Offset refers to the distance between the very first gate of the cell and the first M1 grid position inside the cell.

metal placement and offset features, 2:3 GR architectures can achieve 100% PP as shown in Figure 12(d) and (e), unlike the 1:1 GR. In conclusion, by utilizing the two knobs, we can implement GR with additional vertical routing tracks without sacrificing placement possibility, which in turn can aid in enhancing the block's power, performance, and area (PPA).

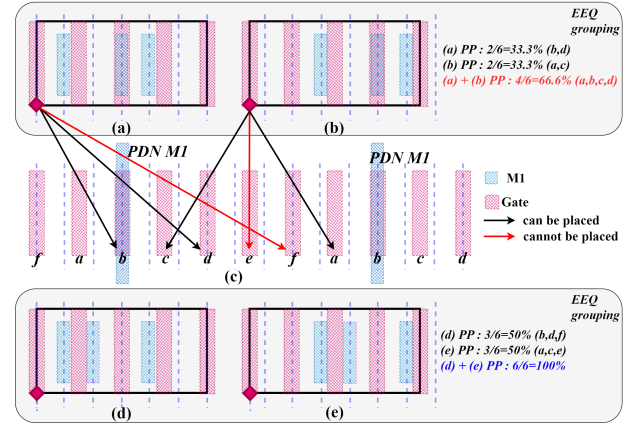


Figure 12. Under 2:3 GR, (a) has 33.3% PP, since it can be placed only b,d. (b) is the different offset version of (a). By using (a) and (b) together we can increase PP by 66.6%. (c) Possible placement location considering PDN. (d) A modified version of (a) with the M1 track left empty for tracks that PDN M1 can use. (e) A different offset version of (d). By using (d) and (e) together we can increase PP by 100%

5 Conclusion

Our scope for the scaling of advanced technology can be summarized as follows:

- Geometric scaling slows down in advanced nodes.
- The area scaling relies on track reduction, which complicates the routability.
- As demonstrated by *pin density wall*, cell utilization cannot be further increased in advanced nodes due to block-level design constraints.

In response to this, we recognize that the following challenges,

- Substrate back side routing: Novel architecture enables back side signal routing to alleviate the front side pin density wall.
- Standard cell designs with low track heights: Design flow and styles adopt low track heights such as gear ratio optimization and multi-row cell synthesis.
- PPA-driven block-level layout: We need to explore block-level layout flow and method optimization such as design rule parameter tuning and algorithm innovations.

For future work, we aim to integrate the discussed design knobs into existing cell layout automation frameworks for a

more effective DTCO/STCO process. Advanced node architecture necessitates a systematic "top-down" customization of cell libraries for each design and a "bottom-up" approach to understand the effect of design choices propagating through layers. This holistic approach holds the potential to unlock the scaling capabilities of advanced device architecture. Beyond layout optimization, we also recognize the potential of logical synthesis, as more efficient cell utilization can alleviate the global router's burden. With the novel 3D device architectures, evaluating pin accessibility for individual cells needs to target more sophisticated routing scenarios that encompass global routing vs. detailed routing, multiple layers, and multiple standard cell rows. Leveraging these design knobs and alternative scaling approaches broaden the scope of exploration and enables us to overcome the *pin density wall*.

Acknowledgement

We acknowledge the comments of the reviewers. This work was supported in part by NSF under Grant CCF-2110419, and TILOS (NSF CCF-2112665).

References

- [1] 2023. Smaller, better, faster: imec presents chip scaling roadmap. <https://www.imec-int.com/en/articles/smaller-better-faster-imec-presents-chip-scaling-roadmap>
- [2] H. Aoyama. [n. d.]. IrdS 2022 lithography - IEEE. https://irds.ieee.org/images/files/pdf/2022/2022IRDS_Litho.pdf
- [3] T. Huynh Bao et al. 2014. Circuit and process co-design with vertical gate-all-around nanowire FET technology to extend CMOS scaling for 5nm and beyond technologies. In *2014 44th European Solid State Device Research Conference (ESSDERC)*. 102–105. <https://doi.org/10.1109/ESSDERC.2014.6948768>
- [4] R. Chen et al. 2021. Design and Optimization of SRAM Macro and Logic Using Backside Interconnects at 2nm node. In *2021 IEEE International Electron Devices Meeting (IEDM)*. 22.4.1–22.4.4. <https://doi.org/10.1109/IEDM19574.2021.9720528>
- [5] C. Cheng, C. T. Ho, C. Holtz, D. Lee, and B. Lin. 2022. Machine Learning Prediction for Design and System Technology Co-Optimization Sensitivity Analysis. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 30, 8 (2022), 1059–1072. <https://doi.org/10.1109/TVLSI.2022.3172938>
- [6] C. Cheng, C. T. Ho, D. Lee, and B. Lin. 2022. Monolithic 3D Semiconductor Footprint Scaling Exploration Based on VFET Standard Cell Layout Methodology, Design Flow, and EDA Platform. *IEEE Access* 10 (2022), 65971–65981. <https://doi.org/10.1109/ACCESS.2022.3184008>
- [7] C. Cheng, C. T. Ho, D. Lee, B. Lin, and D. Park. 2021. Complementary-FET (CFET) Standard Cell Synthesis Framework for Design and System Technology Co-Optimization Using SMT. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 29, 6 (2021), 1178–1191. <https://doi.org/10.1109/TVLSI.2021.3065639>
- [8] S. Choi, J. Jung, A. B. Kahng, M. Kim, C-H Park, B. Pramanik, and D. Yoon. 2023. PROBE3.0: A Systematic Framework for Design-Technology Pathfinding with Improved Design Enablement. (2023). arXiv:2304.13215 [cs.AR]
- [9] R.H. Dennard, F. Gaensslen H, H. Yu, V L. Rideout, E. Bassous, and A. R LeBlanc. 1974. Design of ion-implanted MOSFET's with very small physical dimensions. *IEEE Journal of Solid-State Circuits* 9, 5, 256–268.
- [10] C. T. Ho. 2022. *Novel Computer Aided Design (CAD) Methodology for Emerging Technologies to Fight the Stagnation of Moore's Law*. University of California, San Diego.
- [11] IEEE. 2020. IRDS International Roadmap for Devices and Systems: More-than-Moore White Paper. (2020).
- [12] V. Kaushik et al. 2012. Design and manufacturability tradeoffs in unidirectional and bidirectional standard cell layouts in 14 nm node. In *Design for Manufacturability through Design-Process Integration VI*, Mark E. Mason (Ed.), Vol. 8327. International Society for Optics and Photonics, SPIE, 83270K. <https://doi.org/10.1117/12.916104>
- [13] D. Lee. 2022. *Logical Reasoning Techniques for VLSI Applications*. University of California, San Diego.
- [14] D. Lee, D. Park, C. T. Ho, I. Kang, H. Kim, S. Gao, B. Lin, and C. Cheng. 2021. SP&R: SMT-Based Simultaneous Place-and-Route for Standard Cell Synthesis of Advanced Nodes. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 40, 10 (2021), 2142–2155. <https://doi.org/10.1109/TCAD.2020.3037885>
- [15] S. Maheshwaram, S. K. Manhas, G. Kaushal, B. Anand, and N. Singh. 2011. Vertical Silicon Nanowire Gate-All-Around Field Effect Transistor Based Nanoscale CMOS. *IEEE Electron Device Letters* 32, 8 (2011), 1011–1013. <https://doi.org/10.1109/LED.2011.2157076>
- [16] G. E Moore. 1965. Cramming more components onto integrated circuits. McGraw-Hill New York, NY, USA:.
- [17] S. K Moore. 2020. The node is nonsense. *IEEE Spectrum* 57, 8 (2020), 24–30.
- [18] G. Northrop. 2011. Design technology co-optimization in technology definition for 22nm and beyond. In *2011 Symposium on VLSI Technology - Digest of Technical Papers*. 112–113.
- [19] E. Park and T. Song. 2023. Complementary FET (CFET) Standard Cell Design for Low Parasitics and Its Impact on VLSI Prediction at 3-nm Process. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 31, 2 (2023), 177–187. <https://doi.org/10.1109/TVLSI.2022.3220339>
- [20] D Prasad et al. 2019. Buried Power Rails and Back-side Power Grids: Arm® CPU Power Delivery Network Design Beyond 5nm. In *2019 IEEE International Electron Devices Meeting (IEDM)*. 19.1.1–19.1.4. <https://doi.org/10.1109/IEDM19573.2019.8993617>
- [21] J. Ryckaert. 2022. STCO: System-Technology Co-optimization. <https://www.imec-int.com/en/articles/unlocking-system-scaling-bottlenecks-system-technology-co-optimization>