

Received April 18, 2022, accepted June 9, 2022, date of publication June 17, 2022, date of current version June 24, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3184008

Monolithic 3D Semiconductor Footprint Scaling Exploration Based on VFET Standard Cell Layout Methodology, Design Flow, and EDA Platform

CHUNG-KUAN CHENG ^{1,2}, (Life Fellow, IEEE),

CHIA-TUNG HO ², (Graduate Student Member, IEEE),

DAEYEA LEE ², (Student Member, IEEE), AND BILL LIN², (Member, IEEE)

¹Department of Computer Science and Engineering, University of California at San Diego, La Jolla, CA 92093, USA

²Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093, USA

Corresponding author: Chung-Kuan Cheng (ckcheng@ucsd.edu)

This work was supported in part by NSF under Grant CCF-2110419.

ABSTRACT Continued scaling in accordance with Moore's law is becoming increasingly difficult. Pitch shrinkage and standard cell height reduction via design technology co-optimization with design rules have sustained this scaling until recently. However, we observe that standard cell device scaling is becoming saturated due to yield and cost. One way to continue device footprint reduction is by expanding in the third dimension via monolithic 3D integration, using for example stacked gate-all-around (GAA) devices, complementary FETs, vertical FETs, and 3D logic. However, using these footprint scaling approaches to increase device density creates new problems. Using vertical gate-all-around FET (VFET) technologies as a specific instance of 3D device scaling, we demonstrate that the key bottleneck to footprint scaling is the *pin density wall*. The footprint of a block is predominantly limited by the pin density as we increase the number of active device layers. While a full-blown paradigm shift on layout methodology, design flow, and electronic design automation (EDA) platform is not available now, we describe in this article three specific baby steps that can alleviate the pin density problem and demonstrate their potential benefits for footprint scaling: (1) allocating standard cell pin sideways and using block-level routing with the local interconnect layers; (2) using the backside of the substrate for the power distribution network; and (3) using the generation of more complex standard cells. We show via several core designs that a 42.6% reduction in the core area is achievable when a combination of these operations is employed.

INDEX TERMS 3D integration, DTCO, pin-density wall, routing congestion, STCO, VFET, VLSI scaling.

I. INTRODUCTION

Since the publication of Moore's law in 1965 [1], we have observed prominent efforts to push for and/or facilitate the scaling, including the Dennard's scaling prediction [2], the "More than Moore" roadmap [3], and a recent new metric proposal [4]. These self-fulfilling prophecies are essential for the continued growth of the market, expansion of the industry, and demand for research and development. Figure 1(a) illustrates the scaling roadmap of the technology nodes released by the IMEC team [5], [6]. There are three time windows:

- 1) The litho-centric era (-2013).

The associate editor coordinating the review of this manuscript and

- 2) The design technology co-optimization era (2012-2025).
- 3) The system technology co-optimization era (2022-).

(1) Up to the year 2013, the scaling has mainly relied upon pitch shrinkage. (2) However, starting from 2012, geometric reduction alone was no longer sufficient for the desired scaling. One approach has been to decrease the number of horizontal tracks of the standard cells to reduce the cell height (Fig. 2(a)). However, track reduction induces

routability problems. Therefore, the industry has been tuning the layout design rule parameters as a way to improve routability. This co-optimization between the design technology and track reduction has been able to sustain Moore's law. (3) However, after 2022, standard cell device scaling starts to

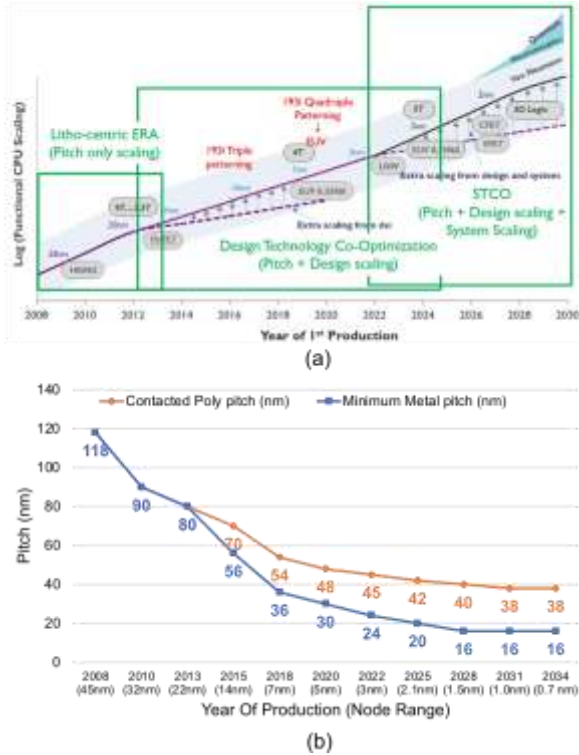


FIGURE 1. Scaling Roadmap. (a) Scaling roadmap [8]. (b) Device Architecture and Ground Rules Roadmap for Logic Devices. [Sources: International Technology Roadmap for Semiconductors, 2007/2009/ 2013 [9]. The International Roadmap for Devices and Systems, 2016/2018/2021 [10]].

reduction of the device footprint is by expanding in the third dimension, e.g. using stacked gate-all-around (GAA) devices, complementary FETs, vertical FETs, and 3D logic (Fig. 2(b)).

This approach to technology expansion changes the physical layout problem from a conventional planar device placement problem to a three-dimensional spatial arrangement problem.

In this article, we focus on vertical gate-all-around FET (VFET) technologies [14]–[17], which is a precursor (year 2027–2034) to 3D VLSI (year 2030–2034) (Figs. 1, 2(b)). The gate length and spacer thickness of VFET are less constrained than a conventional lateral FET as they are oriented vertically. Furthermore, the freedom of device ordering in VFET layouts leads to better layout optimization in terms of routing resources and area density. Recent studies [18], [19] describe a guideline with an interconnect structure to harvest the maximum advantages of 1-tier as well as many-tier VFETs which stack multiple transistors on the same transistor footprint [20], [21]. Also, Lee *et al.* [22] have proposed a Satisfiability Modulo Theories (SMT)-based

many-tier VFET standard cell synthesis automation framework and explored the impact of stacking multiple tiers on the footprint of standard cells and the building block area as shown in Fig. 3.

In the sequel, we will first describe the VFET architecture and current state-of-the-art layout approaches. We demonstrate that pin density becomes the bottleneck of footprint

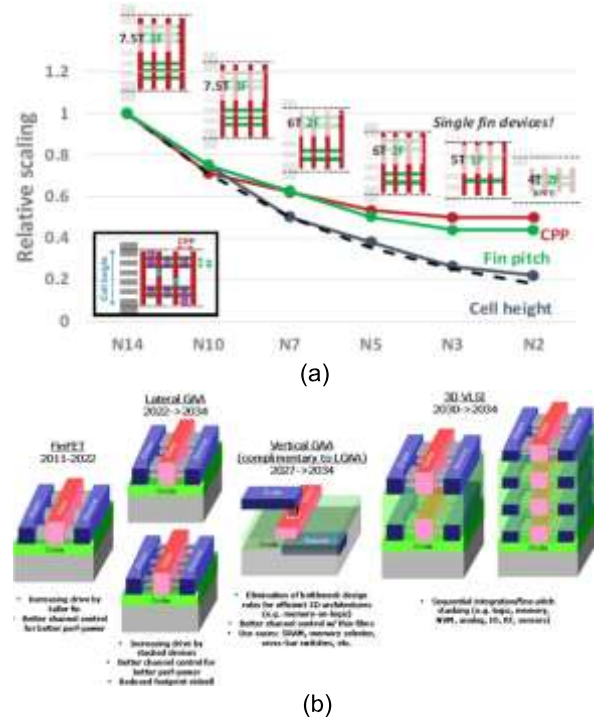


FIGURE 2. Scaling with design and system technology co-optimization. (a) Design technology co-optimization [11], [12]: Cell height reduction (7.5-4T) to enhance the scaling as contacted poly pitch (CPP) drawn in red and fin pitch (in green) scaling slows down. (b) System technology co-optimization [13]: After the period of FinFET (2011–2022), devices grow in the third dimension to reduce the footprint. GAA: Gate-All-Around.

scaling. The footprint of the block is dominated by the pin density, even if we increase the number of active device layers. Note that footprint shrinkage reduces the signal traveling distance and thus the footprint size is one fundamental metric of the power, performance, area, and cost of the technology. While a full-blown paradigm shift on layout methodology, design flow, and electronic design automation (EDA) platform is not available now, we use the following operations to demonstrate the potential benefits for footprint scaling.

- 1) Allocating standard cell pin sideways and using block-level routing with local interconnect layers: We incorporate back end of the line (BEOL) routing resources that were only used for building local interconnections inside standard cells in the block-level routing. Thus, routing with additional local

interconnect layers between the cells improves the routability compared with the conventional routing only utilizing the layers above the local interconnect layers in BEOL.

- 2) Backside power delivery network: We use the backside of the substrate for the power distribution network. This approach eliminates the BEOL routing for power pins and thus leaves more routing resources for signal pins.
- 3) Complex standard cell generation: We show that the pin density can be reduced with more complex standard cells, which is consistent with our derivation according to Rent's rule.

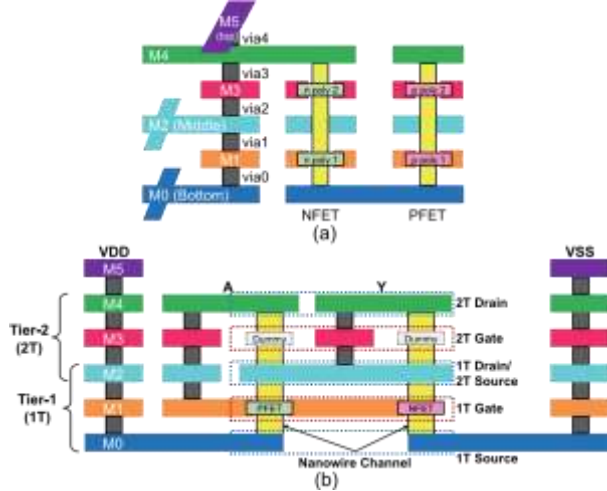


FIGURE 3. Block-level design utilization, standard cell, and building block area comparisons for 4.5T Lateral GAAFET and many-tier VFETs [22]. (a) design utilization ((total standard cell area)/(building block area)). (b) standard cell area presented as a normalized average area of M0 Core, M1 Core, and AES. (c) building block area presented as a normalized average area of M0 Core, M1 Core, and AES.

FIGURE 4. A sample of 2-tier VFET layout [22]. (a) layout architecture. (b) A profile view of a 2-tier VFET inverter.

II. BACKGROUNDS

A. MANY-TIER VERTICAL GATE-ALL-AROUND FET (VFET)

The VFET technology [14]–[19] is a successor of complementary FET and a precursor of monolithic 3D logic [8]. The standard cells are built in the local interconnect layers of the BEOL and cell pins are connected via the rest of the BEOL above the local interconnect layers [22]–[24]. Within the local interconnect layers, we can stack devices on multiple

layers. Figure 4 illustrates a VFET sample of two-device layers with (a) layout architecture, and (b) a layout sample of an inverter [22]. The source/gate/drain nodes of VFET device are vertically oriented (Fig. 4). The device layer (tier) is associated with three metal layers for source, gate, and drain interconnections. The gate poly layers are directly connected to the corresponding (odd numbered) metal interconnection layers (e.g., M1, M3). The routing on source/drain nodes (e.g., M0, M2, and M4) is bidirectional and on gates is unidirectional.

In the study, we use options of one to four device layers (tiers) for standard cell layout. The cell height can accommodate six horizontal metal tracks. The pins are allocated with a routability-driven threshold to keep a lower bound on the number of access points for each pin. The standard cells are designed with a rule-based satisfiability modulo theories (SMT) package so that the cell layout is optimized with the given VFET technology constraints [22]. For the block-level BEOL, we use five metal layers with unidirectional routing. Thus, the cell pins on top are connected unidirectionally or are extended to another layer to be routed in an orthogonal direction. The block-level logic design, placement, and routing are conducted using a commercial electronic design automation software suit.

B. PIN-DENSITY WALL

We use the term ‘‘pin-density wall’’ to express the limitation

that footprint scaling is stalled even if we increase the number of device layers in the VFET technology. Figure 5 illustrates the trend of block area vs. pin-density using 1-4 tiers on three different test cases. We calculate the pin-density extracted from an average number of external connections over one hundred windows over the block. For each size ranging from $0.1\mu m^2$ to $143\mu m^2$, the window is shaped as a square and its copies are evenly distributed over the block (Fig. 6). In general, pin-density is inversely proportional to the block area. The enlarged markers indicate the smallest feasible block area, labeled with the corresponding pin density. In general, the block area becomes smaller as we increase

the number of tiers. However, the amount of area decreased diminishes and all saturates around a pin-density in the range of 40.1-44.1 $\text{pins}/\mu\text{m}^2$. The pin-density is reaching its limit due to the following factors:

- 1) The reduction of the standard cell height with a decreasing number of horizontal routing tracks leads to decreasing routing resources and increasing pindensity. This trend is driven by the DTCO effort (Fig. 2(a)).
- 2) The addition of device layers that increases the device density, but the number of pins per cell remains the same. Therefore, the ratio of pins to footprint area increases.
- 3) The saturation of the metal pitch (Fig. 2(b)), which demands the same amount of area for pin access, even for more advanced technology nodes.

- 4) The conventional layout methodology, design flow, and EDA platform, which uses small-functional standard cells for flexibility, and two-staged two-dimensional block-level layout (i.e., cell placement and block-level routing). The smaller functional cells cause higher pin-density than larger cells. The two-staged twodimensional block-level layout limits the capability of fully utilizing the BEOL resources.

III. THREE STEPS TO ALLEVIATE THE PIN-DENSITY PROBLEM

A. STANDARD CELL PIN ON SIDEWAYS AND BLOCK-LEVEL ROUTING INCLUDING LOCAL INTERCONNECT LAYERS

We allocate standard cell pins to the top of the cell and allow sideways connections below the top layer accessible for blocklevel layout. The design rules for sideways connections

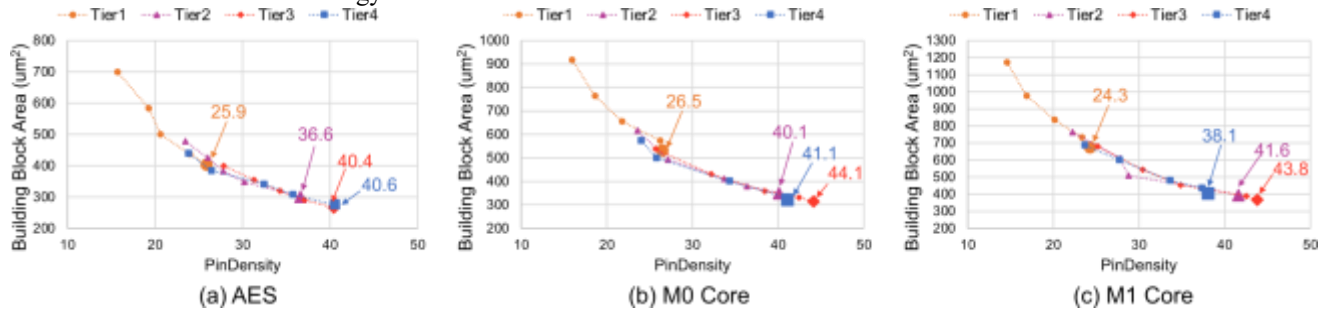


FIGURE 5. Plot of Building Block Area vs. Pin-density. (a) Plot of building block area of AES design as a function of pin-density per each #tier case from lower utilization to the maximum achievable utilization (i.e., minimum valid area). Utilization D (total standard cell area)/(building block area). The larger markers and data labels of each tier case represent the pin-density at the minimum valid area. (b) Same as in (a), but for M0 Core. (c) Same as in (a), but for M1 Core..

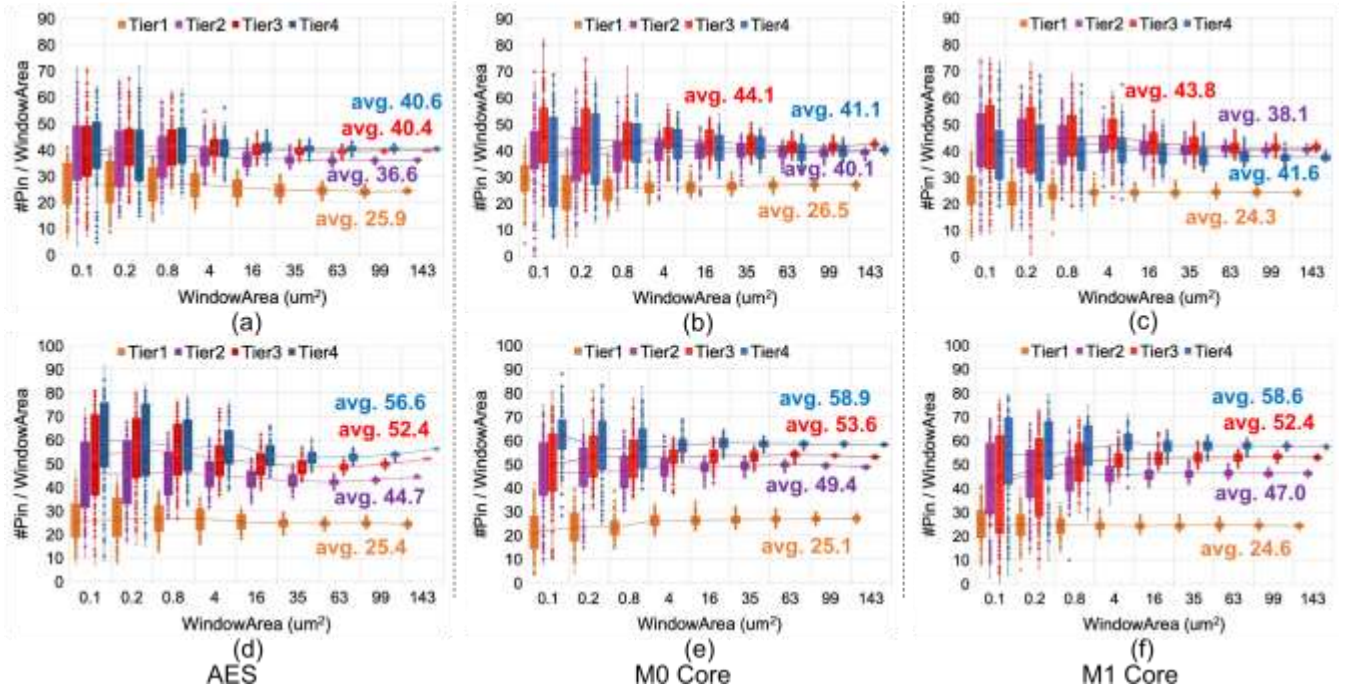


FIGURE 6. Comparison of the pin-density by routing methods. (a)-(c) Box and whisker plots of pin-density in different window areas of (a) AES, (b) M0 Core, (c) M1 Core design for 1/2/3/4-tier VFETs with the conventional routing method. Solid lines show mean line of each window area for 1-tier (orange), 2-tier (purple), 3-tier (red), and 4-tier (blue) VFETs. Data labels present the average mean values for each tier case. (d)-(f) Same data as in (a)-(c), but for the routing with local interconnect layers.

are set the same as those used in the standard cell generation (Section IV).

For conventional layout design flows, the standard cell pins are allocated on the top of the cell. We partition the routing layers into three groups: local layers for intra-standard cell layout, middle layers right above the local layers for intercell layout, and global layers above the other two groups for long-distance interconnect. The cell pins are located and routed from the top local interconnect layer to the middle and global layers in the block-level layout. The local layers within standard cells are not used for inter-cell routing because the utilization of the cell area is high, i.e., most cells are abutted together edge to edge.

In order to break the pin density wall with a low utilization rate, we propose to extend the cell pin sideways and incorporate local interconnect layers for block-level layout.

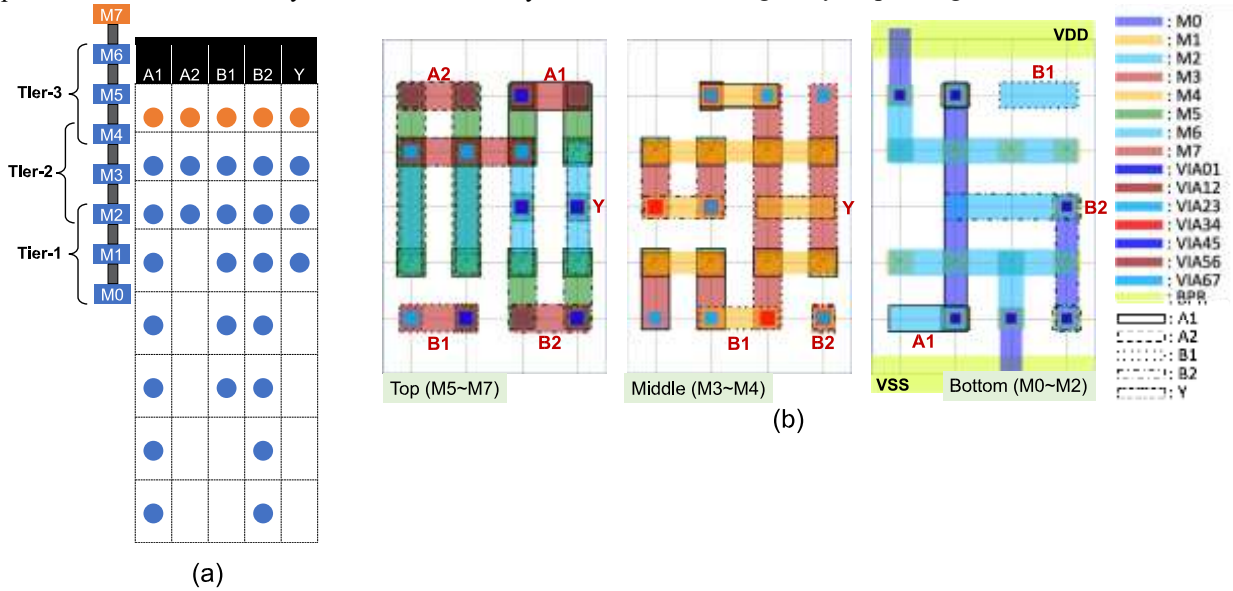


FIGURE 7. Examples of cells with pins on sideways. (a) Layer occupation of 3-tier AOI22-1 cell's I/O pins with top layer pin allocation constraint. A1/A2/B1/B2/Y represent the I/O pins and the solid circles in blue/orange colors indicate the occupied layers by each I/O pin. (b) 3-tier AOI22-1 cell layout which is generated with top layer pin allocation constraint. Each of left/middle/right figure shows top/middle/bottom-tier layers, respectively. The yellow boxes indicate the labels of I/O pins.

We explore the impact of cell pin allocation and block-level routing methodology on the building block area through the following approaches. (i) We synthesize standard cell libraries with pins on top and also accessible sideways. We perform the block-level layout using the middle and global layers, which is the case for Fig. 5. (ii) We use the same cell libraries but perform block-level layouts including the local layers.

Figure 7 shows a layout example of AOI22 $\times$ 1, where all I/O pins (i.e., A1, A2, B1, B2, and Y) occupy the top layer (solid circle in orange) with a threshold of accessible points and also extend sideways via local routing layers. Fig. 7(a) uses each column (pin) and row (local routing layer) to put a dot indicating the pin is accessible on the corresponding local routing layer. Fig. 7(b) illustrates the stick diagram of the layout with labels on the pins.

B. BACKSIDE POWER DELIVERY NETWORK (PDN)

We explore the impact of backside PDN as one technique to push against the pin density wall. Power pins take a portion of the pin resources and routing resources. For example, in our experimental settings (Section IV), we put columnwise power stripes for every 64 contact poly pitches, which take approximately 11.5% of total available vertical tracks in the lowest middle layer. These power stripes are then connected to the power pads on top of the chip via power grids in the middle and global layers. The backside PDN approach reduces the pin count and leaves more routing resources for signal pins. Chava *et al.* [25] showed that the backside PDN approach tackles the challenges such as the higher current demand, increased power density, and low power supply noise margin by separating the on-die PDN from the

conventional BEOL. In this study, we demonstrate that the approach can improve area utilization.

C. COMPLEX STANDARD CELL FOR MANY-TIER VFET

We explore complex standard cells to reduce the pin density for block-level layout. Merging more standard cells into larger complex cells can reduce the pin-density of each module according to Rent's rule [26]. Rent's rule has $T = tg^p$, where T is the number of pins, t is a constant, g is the number of gates, and p is a Rent's constant in a range of (0.5, 0.8) empirically. The large the module g , the ratio of pin T to gate g (i.e., $T/g = tg^{p-1}$) will become smaller since $p - 1 < 0$.

The goal to reduce the pin count using a complex standard cell approach for VFET is different from conventional FET designs. In conventional FET designs, we merge standard cells to reduce the total transistor count, increase circuit performance, decrease power, and potentially lower fabrication costs on clever combinations of cells. For example, a 2-2 AOI gate can be constructed with 8 transistors

in CMOS, compared to 16 transistors using two 2-input NAND gates (8 transistors), two inverters (4 transistors), and

reduced number of I/O pins (6 $\rightarrow$ 4). Thus, the reduced pin-density of complex cells can lead to the improvement of the

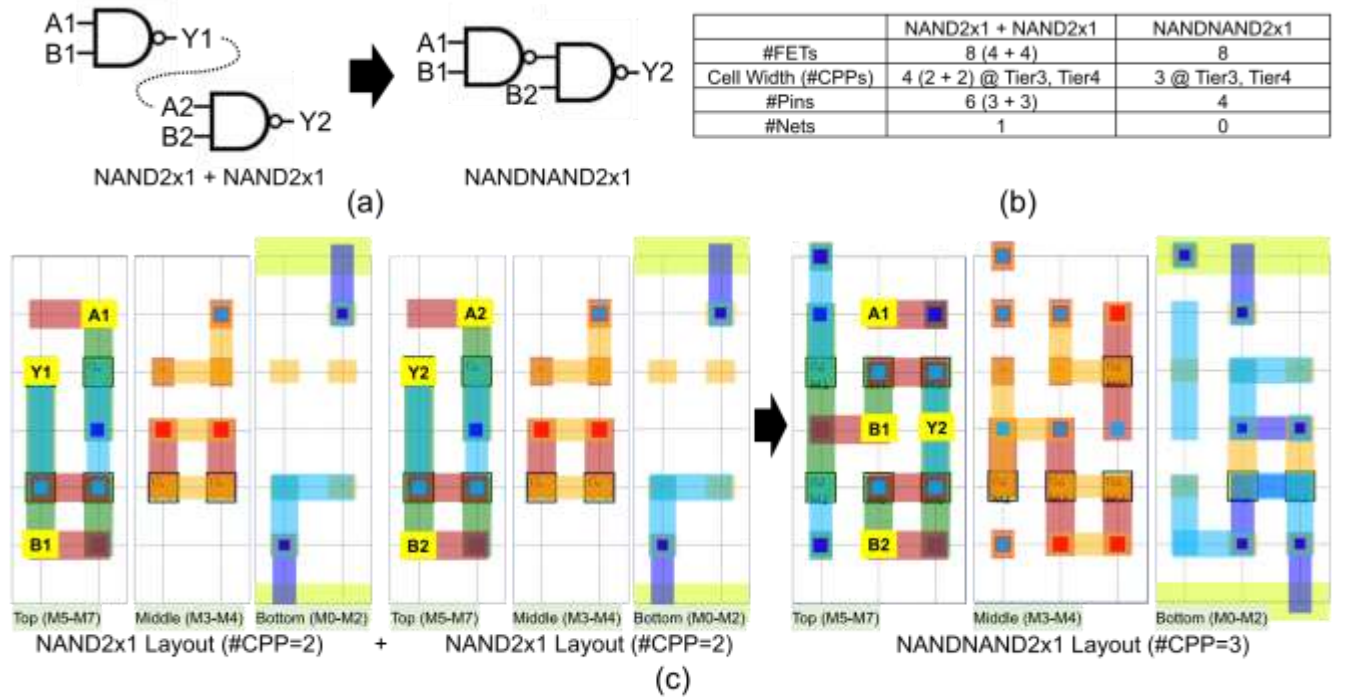


FIGURE 8. Examples of complex cells. (a) Schematic view of a NANDNAND2 1 cell merging two sequentially connected NAND2 1 cells. (b) Comparison of #FETs, cell width, #pins, and #nets between two separate NAND2 1 cells and one NANDNAND2 1 cell. (c) Layout view of merging two

a 2-input NOR gate (4 transistors). Therefore, in the conventional 2D process architecture, combinations of basic cells such as NAND, NOR, Inverter, Buffer, XOR, and XNOR can have such area benefit if the combined logic function can be implemented with a smaller number of transistors because the number of total transistors is a critical factor determining the area of cells.

For many-tier VFET architecture, we can merge two or more cells to reduce the pin count and also, decrease the footprint. Figure 8 illustrates an example of merging two sequentially connected NAND2x1 cells into a single merged cell in a 3/4-tier VFET structure. Since a NAND2 x 1 cell consists of 4 transistors, it respectively has 8 and 12 dummy transistors in each 3-tier and 4-tier configuration. Therefore, we can merge these two cells using these empty transistor placements. Theoretically, each 3/4-tier NAND2 x 1 cell has enough dummy transistors to accommodate one more set of transistors without increasing its footprint. However, since all the I/O pins must be located on the top layer for the pin-accessibility, the merged cell requires one more CPP to allocate all the I/O pins. Despite this overhead, the merged cell can be implemented with one fewer number of CPPs (4 $\rightarrow$ 3). In this example, the pin-density of the merged cell

NAND2 1 cells into one NANDNAND2 1 with one less #CPPs.

is also reduced because the connection between “Y1” and “A2” is locally routed inside the merged cell resulting in the

routability in the block-level P&R.

In this work, we generate four complex cells (i.e., NANDNAND2 x 1, NANDNOR2 x 1, NORNOR2 x 1, and NORNAND2 x 1) combining two sequentially connected NAND2 x 1/NOR2 x 1 cells that can be merged with the same reduction of footprint (4 #CPPs $\rightarrow$ 3 #CPPs) and the number of I/O pins (6 $\rightarrow$ 4). We only consider 3/4-tier VFETs because there is no area benefit in 1/2-tier architecture. Table 1 presents the comparison of the number of instances and the area occupation in the total design netlist related to the NAND2x1, NOR2x1, and additional complex cells between the original and the modified netlist accommodating complex cells. The total number of NAND2 x 1/NOR2 x 1 cells in AES / M0 Core / M1 Core is reduced from 5,020 $\rightarrow$ 3,414 / 4,946 $\rightarrow$ 3,826 / 4,474 $\rightarrow$ 3,799 by merging sequentially connected cells into additional complex cells, respectively. Each design has a different portion of NAND2x1/NOR2x1 cells ranging from 16.0% to 33.9% across the different number of tiers and netlists.

IV. RESULTS

A. EXPERIMENTAL ENVIRONMENT

1) STANDARD CELL GENERATION

To explore the scaling impact of many-tier VFETs on cell-level and block-level area, we select 30 representative

TABLE 1. Changes on the NAND2 1 and NOR2 1 cells' ratio by introducing complex cells (i.e., NANDNAND2 1, NANDNOR2 1, NORNOR2 1, and NORNAND2 1).

Design	Default Netlist		Modified Netlist w/ complex SDCs	
	InstanceName	#Instance	InstanceName	#Instance
AES	NAND2x1	2,573	NAND2x1	814
	NOR2x1	2,447	NOR2x1	994
			NANDNAND2x1	514
			NANDNOR2x1	375
			NORNOR2x1	361
			NORNAND2x1	356
	Total #Instance	5,020	Total #Instance	3,414
M0 Core	Area Occupation in Total Design	32.6% @Tier3 33.9% @Tier4	Area Occupation in Total Design	28.9% @Tier3 30.1% @Tier4
	NAND2x1	2,652	NAND2x1	1,256
	NOR2x1	2,294	NOR2x1	1,450
			NANDNAND2x1	424
			NANDNOR2x1	299
			NORNOR2x1	148
			NORNAND2x1	249
M1 Core	Total #Instance	4,946	Total #Instance	3,826
	Area Occupation in Total Design	23.8% @Tier3 25.6% @Tier4	Area Occupation in Total Design	21.7% @Tier3 23.4% @Tier4
	NAND2x1	3,287	NAND2x1	2,416
	NOR2x1	1,187	NOR2x1	708
			NANDNAND2x1	287
			NANDNOR2x1	178
			NORNOR2x1	91
M1 Core			NORNAND2x1	119
	Total #Instance	4,474	Total #Instance	3,799
	Area Occupation in Total Design	17.1% @Tier3 19.3% @Tier4	Area Occupation in Total Design	16.0% @Tier3 18.2% @Tier4

standard cells [27], [28]¹ from ASAP7 [29] process design

TABLE 2. Cell area comparison for 30 representative standard cells.

Cell	#Pin	#FET	#Net	Cell Area (#CPP)			
				Top-Allocation			
				Tier1	Tier2	Tier3	Tier4
AND2x2	3	6	7	4	3	2	2
AND3x1	4	8	9	4	3	2	2
AND3x2	4	8	9	5	3	3	2
AOI21x1	4	6	8	6	3	3	3
AOI22x1	5	8	10	8	4	4	3
BUFX2	2	4	5	3	2	2	2
BUFX3	2	4	5	4	2	2	2
BUFX4	2	4	5	5	3	2	2
BUFX8	2	4	5	10	5	4	3

DFFHQNx1	3	24	17	12	6	5	4
FAx1	5	24	17	12	6	4	4
INVx1	2	2	4	2	1	1	1
INVx2	2	2	4	2	2	2	2
INVx4	2	2	4	4	2	2	2
INVx8	2	2	4	8	4	3	2
NAND2x1	3	4	6	4	2	2	2
NAND2x2	3	4	6	8	4	3	2
NAND3x1	4	6	8	9	5	3	3
NAND3x2	4	6	8	18	9	6	5
NOR2x1	3	4	6	4	2	2	2
NOR2x2	3	4	6	8	4	3	2
NOR3x1	4	6	8	9	5	3	3
NOR3x2	4	6	8	18	9	6	5
OAI21x1	4	6	8	6	3	3	3
OAI22x1	5	8	10	8	4	4	3
OR2x2	3	6	8	4	3	2	2
OR3x1	4	8	9	4	3	2	2
OR3x2	4	8	9	5	3	3	2
XNOR2x1	3	10	9	8	4	3	3
XOR2x1	3	10	9	8	4	3	3
Average				7.0	3.8	3.0	2.6

kit library as specified in Table 2. Then, we generate one to four tiers of VFET cell libraries with six horizontal metal layers and buried power rails based on the SMT-based many-tier VFET standard cell synthesis framework [22]. The SMT-based framework formulates a conventional (sequential) cell layout process as a constraint satisfaction problem (CSP) with variables and constraints to integrate place-and-route steps into a multi-objective optimization problem. It adopts the state-of-the-art *lazy*-approach SMT solver Z3 [30] to solve the given optimization problem. Thus, given netlist information and cell architecture, the framework simultaneously obtains an optimal solution that strictly satisfies the constraints of transistor placement, in-cell routing, and conditional design rules. The clock and latch placements of the sequential cell (i.e., DFFHQNx1) are strictly ordered by the sequential datapath to optimize the cells' PPAC by adopting a cell partitioning feature [31]. Also, the I/O pins in each cell are allocated to keep the minimum two access points by the routability-driven threshold constraint. In this work, we limit the exploration of device stacking up to 4-tiers because it is observed that the cell footprint gain decreases as the number of tiers increases and the gain from 3-tier to 4-tier is less than 5% (i.e., 3.5%) as

¹ In this experiment, we select representative, typical types of standard cells carrying various structures of combinational and sequential logic cells by reflecting field engineers' opinions.

shown in Fig. 3(b) [22]. We adopt the same conditional design rule [31], [32] parameters (i.e., minimum area (MAR) = 1, end-of-line (EOL) = 0, via (VR) = 0, parallel run length (PRL) = 1, stepheight (SHR) = 1, and minimum pin opening (MPO) = 2) specified in the previous work [22].

2) BLOCK-LEVEL PLACEMENT AND ROUTING

We employ three open-source RTL designs [33], M0 Core, M1 Core, and AES, which respectively have 17K, 20K, and 14K instances. We perform the block-level analysis through the commercial P&R tool [34]. In this work, we removed timing constraints (i.e., setup and hold) to maintain the same netlist configuration (i.e., the same type/number of instances) regardless of the number of tiers, P&R options, and the routing methods so that we can focus on the impact of changes in the cell footprint, pin-accessibility, and P&R options on the routability of designs. We set the number of masks for each local layer of BEOL and use 36nm and 24nm for the contacted poly pitches (CPPs)/Vertical metal pitch and horizontal metal pitches, respectively, by applying the design parameters from previous works [8], [15]. We use five middle BEOL layers with unidirectional routing. The pitches and widths of middle BEOL layers are set by referring to the LEF/DEF language reference [35]. The front side power delivery network consists of the top metal-layer power meshes, intermediate power stripes, and standard cell power rails (BPR). Then, the power is delivered from the lowest middle BEOL layers, which is 4× wider than signal wires, to BPR using stacked vias and SuperVia models [36], respectively. The power stripes for the BPR standard cell rail are placed per every 64 CPPs [37]. We use the 300 #DRVs threshold to measure the valid block level area. As a common industrial practice, once the number of DRVs increases beyond 300, the block layout is considered too expensive to fix with laborious (sometimes, manual) engineering change orders.

B. RESULTS ANALYSIS

Figure 9 describes the results on three test cases (AES, M0 Core, and M1 Core) (Table 2) over three columns. The first row (a, b, c) shows the total standard cell area vs. the number of tiers. For standard cell library designs, the total cell area scales by $153.7\mu\text{m}^2/350.0\mu\text{m}^2 = 0.439$, $200.1/458.6 = 0.436$, $239.6/585.6 = 0.409$ from one tier to four tier technology for the three test cases. The area scaling from one tier to two tier technology is the most significant, $191.3/350.0 = 0.546$, $246.6/458.6 = 0.538$, $305.3/585.6 = 0.521$ and the drop slows down afterward. The benefit of more tiers is diminished by the VFET layout architecture (Fig. 4(a)) of limited routing resources in a narrow (small #CPPs) and tall (many tiers) space and the requirement that all pins extend to the top. For complex cell designs, the total area scales by a small percentage, $8.4\mu\text{m}^2/153.7\mu\text{m}^2 = 5.47\%$, $5.8/200.1 = 2.90\%$,

$3.5/239.8 = 1.46\%$ for four tier technology comparing with the standard cell library designs on three test cases. The percentage differences of the drop in the three cases are caused by the netlist component compositions. As described in table 1, the test case AES has more replacement of complex cells (number of instances drops from 5,020 to 3,414) and thus has more gain on the replacement. For this experiment,

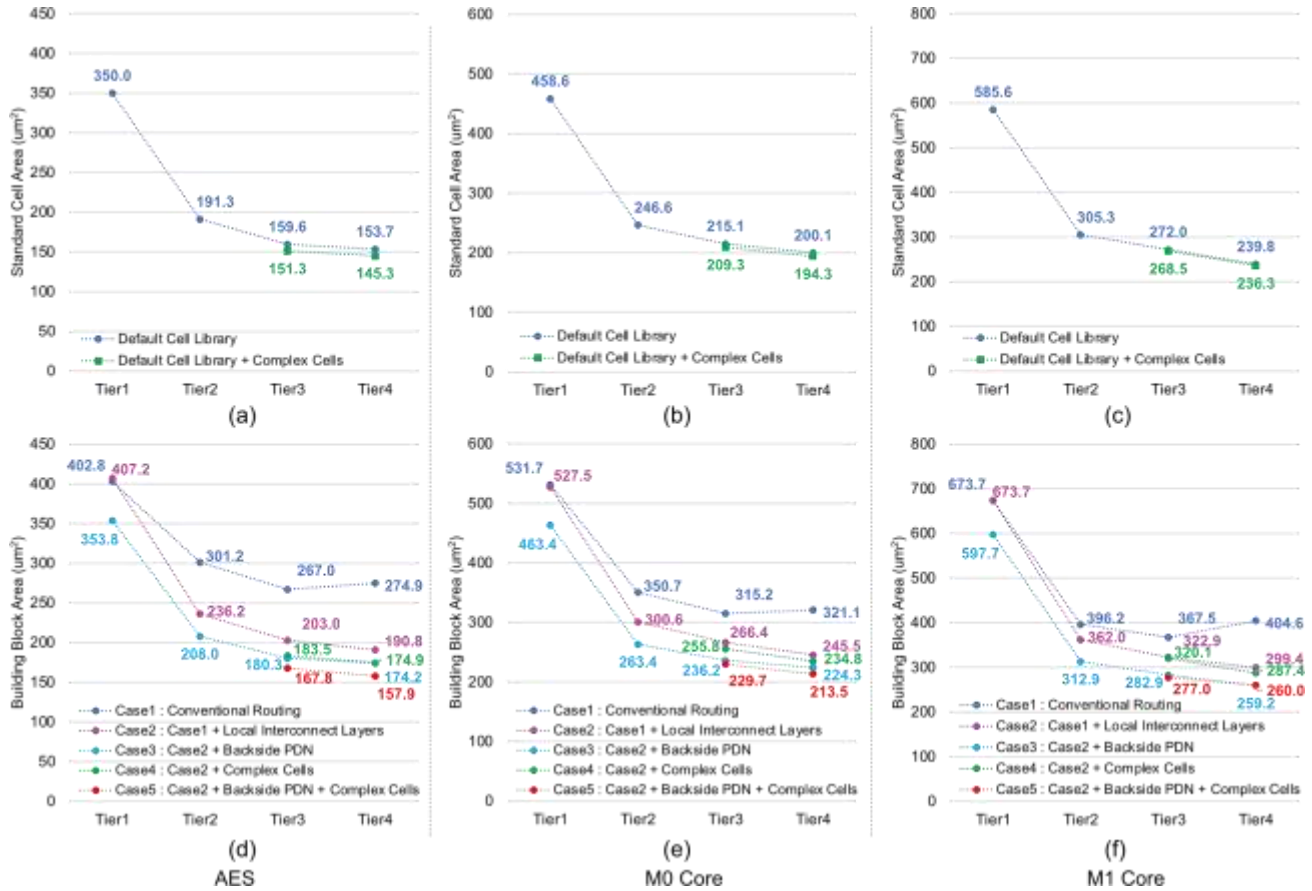


FIGURE 9. Block-level P&R results. (a) Standard cell areas of AES design with different cell libraries for 1/2/3/4-tier VFET cells. (b) Same as in (a), but for M0 Core design. (c) Same as in (a), but for M1 Core design. (d) Minimum valid building block areas of AES design with 1/2/3/4-tier VFET cells for various experimental configurations. (e) Same as in (d), but for M0 Core design. (f) Same as in (d), but for M1 Core design.

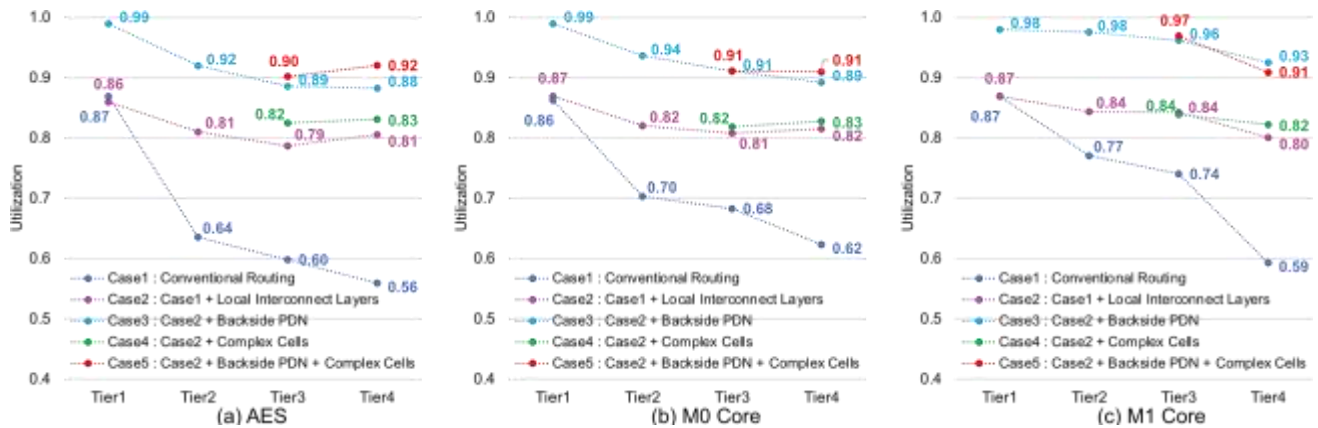


FIGURE 10. Block-level design utilizations. (a) Minimum valid utilization of AES design with 1/2/3/4-tier VFETs for various experimental configurations. Utilization = (total standard cell area)/(building block area). (b) Same as in (a), but for M0 Core design. (c) Same as in (a), but for M1 Core design.

we only adopt four complex cells, which are the most popular (in number) on the netlist. If we allow more and larger complex cells, we may see more benefits in cell area reduction.

Figures 9(d)-(f) show the block-level area vs. number of tiers. We have five cases: (1) conventional block-level routing excluding the local layers, (2) block-level routing with local layers, (3) block-level routing with local layers and

PDN on the backside, (4) block-level routing with local layers and complex cells, and (5) block-level routing with local layers, complex cells, and PDN on the backside. We observe the

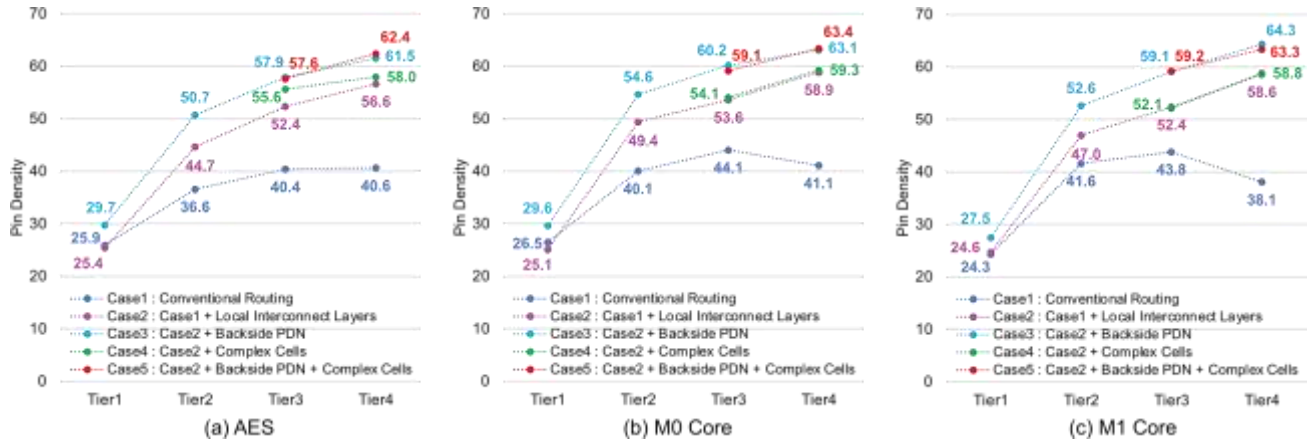


FIGURE 11. Comparison of pin-density. (a) Average pin-density of AES design with 1/2/3/4-tier VFETs for various experimental configurations. (b) Same data as in (a), but for M0 Core design. (c) Same data as in (a), but for M1 Core design.

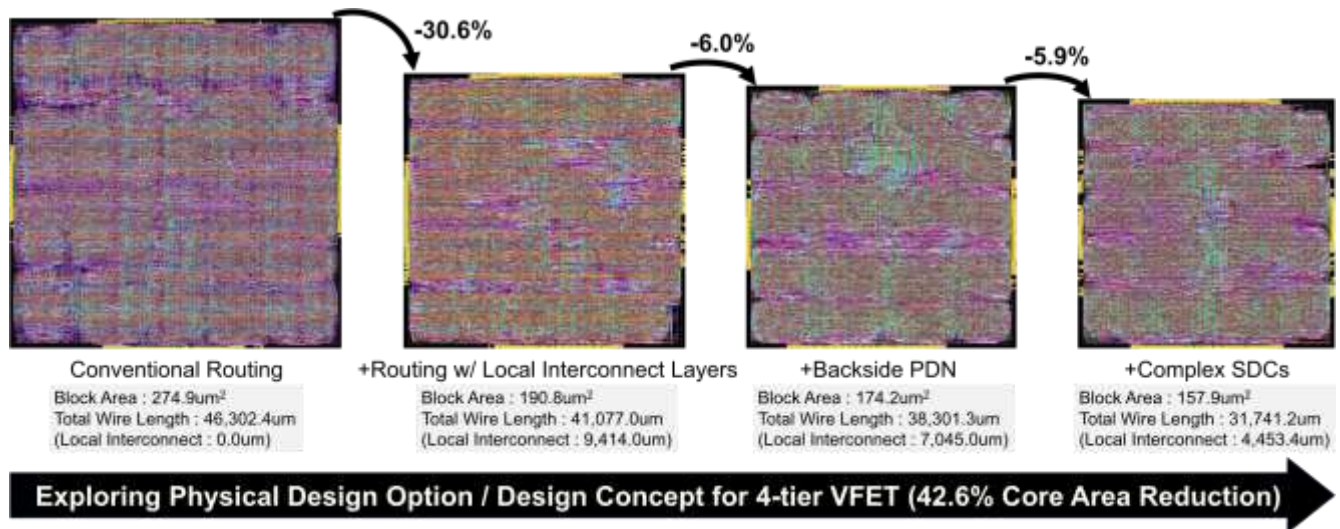


FIGURE 12. Block-level P&R results of AES for the physical design option and approaches. Block area reduces from $274.9 \mu m^2$ to $157.9 \mu m^2$ while total wire length from $46,302.4 \mu m$ to $31,741.2 \mu m$. The local interconnect (intercell connection below pin layer) reduces from $9,414.0 \mu m$ (the second map that allows local interconnect) to $4,453.4 \mu m$.

block-area scaling from (1) ($274.9, 321.1, 404.6 \mu m^2$) to (5) ($157.9, 213.5, 260.0 \mu m^2$) by (57.4, 66.5, 64.3) percents for (AES, MO Core, M1 Core) on 4 Tier technology with the corresponding improvement of utilization from (0.56, 0.62, 0.59) to (0.92, 0.91, 0.91) as shown in Fig. 10.

From (1) to (2), we observe most benefits: block areas scale to ($190.8, 245.5, 299.4 \mu m^2$) by (69.4, 76.5, 74.0) percents on 4-tier technology. The area reduction is caused by extra routing resources (local layers), which allow higher pin density. Fig. 11 shows that pin densities increase from ($40.6, 41.1, 38.1 \text{ pins}/\mu m^2$) to ($56.6, 58.9, 58.6 \text{ pins}/\mu m^2$) on 4-tier technology.

From (2) to (3), we use backside PDN to free more pin spaces and routing resources. Therefore, the block areas scale

to ($174.2, 224.3, 259.2 \mu m^2$) and the pin densities increase to ($61.5, 63.1, 64.3 \text{ pins}/\mu m^2$) on 4-tier technology (Fig. 11).

From (3) to (4), we use complex cells to merge some standard cells. We observe that block areas scale to ($174.2, 234.8, 287.4 \mu m^2$) on 4-tier technology. However, pin densities of option (4) ($58.0, 59.3, 58.8 \text{ pins}/\mu m^2$) are comparable to the pin density of option (2). Because the two options ((2) and (4)) use the same routing resources (local, middle, and global groups), they hit the same pin density wall. On the other hand, for option (4), the pin counts of the netlist are reduced by cell merging. Thus, option (4) renders smaller block areas than option (2).

From (4) to (5), we add backside PDN in addition to complex cells (option (4)). The block areas scale to ($157.9, 213.5, 260 \mu m^2$) and the pin densities increase to ($62.4, 63.4, 63.3 \text{ pins}/\mu m^2$) for 4-tier technology.

From (3) to (5), we add complex cells in addition to backside PDN (option (3)). The block area differences are

(174.2, 224.3, 259.2)-(157.9, 213.5, 260.0)=(16.3, 10.8, 0.8) μm^2 . For the first two test cases, the block areas are reduced. For the third test case, the benefit is buried by noise because the case has fewer cell merging (4,474 merged to 3,799)(Table 1) than the other two. However, the pin densities of these two options are comparable. Because options (3) and (5) use the same routing resources (local, middle, and global groups plus the PDN backside technology), options (3) and (5) hit the same pin density wall.

V. CONCLUSION

We have reported a comprehensive study of three possible approaches for alleviating the emerging wiring crisis by overcoming the *pin-density wall* in monolithic 3D semiconductor footprint scaling based on the VFET standard cell layout. We have observed that pin-density is the bottleneck for the conventional layout methodology, design flow, and EDA platform which use small-functional cells for flexibility, and two-staged two-dimensional block-level cell placement and routing. Throughout the exploration for many-tier VFET configurations up to four tiers, we show that the deterioration of area benefits from cell footprint scaling without proper metal pitch scaling can be significantly mitigated by increasing pin-densities by (i) utilizing the additional routing resources in the local interconnect layers of 3D cells, (ii) applying the backside PDN option, and (iii) increasing module size (i.e., complex cells) to reduce the pin-density in each module according to Rent's rule as shown in Fig. 12. Lastly, we find that there are still rooms to further explore, e.g., the higher parasitic resistance of many-tier VFETs and the thermal issue in a stacked logic transistors [21] call future research topics to obtain the maximum-achievable PPAC (power, performance, area, and cost) benefits through VFET.

REFERENCES

- [1] G. E. Moore, "Cramming more components onto integrated circuits," McGraw-Hill, New York, NY, USA, Tech. Rep., 1965.
- [2] R. H. Dennard, F. H. Gaensslen, V. L. Rideout, E. Bassous, and A. R. LeBlanc, "Design of ion-implanted MOSFET's with very small physical dimensions," *IEEE J. Solid-State Circuits*, vol. JSSC-9, no. 5, pp. 256–268, Oct. 1974.
- [3] *IRDS International Roadmap for Devices and Systems: More-Than-Moore White Paper*, IEEE, 2020.
- [4] S. K. Moore, "The node is nonsense," *IEEE Spectr.*, vol. 57, no. 8, pp. 24–30, Aug. 2020.
- [5] S. M. Y. Sherazi, M. Cupak, P. Weckx, O. Zografos, D. Jang, P. Debacker, D. Verkest, A. Mocuta, R. H. Kim, A. Spessot, and J. Ryckaert, "Standardcell design architecture options below 5 nm node: The ultimate scaling of FinFET and nanosheet," *Proc. SPIE*, vol. 10962, Mar. 2019, Art. no. 1096202.
- [6] R.-H. Kim, Y. Sherazi, P. Debacker, P. Raghavan, J. Ryckaert, A. Malik, D. Verkest, J. U. Lee, W. Gillijns, L. E. Tan, V. Blanco, K. Ronse, and G. McIntyre, "IMEC N7, N5 and beyond: DTCO, STCO and EUV insertion strategy to maintain affordable scaling trend," *Proc. SPIE*, vol. 10588, Mar. 2018, Art. no. 105880N.
- [7] *IRDS International Roadmap for Devices and Systems: Lithography*, IEEE, 2020.
- [8] S. M. Y. Sherazi, M. Cupak, P. Weckx, O. Zografos, D. Jang, P. Debacker, D. Verkest, A. Mocuta, R. H. Kim, A. Spessot, and J. Ryckaert, "Standardcell design architecture options below 5 nm node: The ultimate scaling of FinFET and nanosheet," *Proc. SPIE*, vol. 10962, Mar. 2019, Art. no. 1096202.
- [9] *ITRS Report*. [Online]. Available: <http://www.itrs2.net/itrs-reports.html>
- [10] *IRDS the International Roadmap for Devices and Systems: More Moore*, IEEE, 2021.
- [11] J. Ryckaert, A. Gupta, A. Jourdain, B. Chava, G. Van der Plas, D. Verkest, and E. Beyne, "Extending the roadmap beyond 3 nm through system scaling boosters: A case study on buried power rail and backside power delivery," in *Proc. Electron Devices Technol. Manuf. Conf. (EDTM)*, Mar. 2019, pp. 50–52.
- [12] Z. Tokei. [Online]. Available: <https://www.imec-int.com/en/imecmagazine/imec-magazine-november-2018/the-supervia-a-promising-scaling-booster-for-the-sub-3nm-technology-node>
- [13] *IRDS International Roadmap for Devices and Systems: Executive Summary*, IEEE, 2020.
- [14] S. Maheshwaram, S. K. Manhas, G. Kaushal, B. Anand, and N. Singh, "Vertical silicon nanowire gate-all-around field effect transistor based nanoscale CMOS," *IEEE Electron Device Lett.*, vol. 32, no. 8, pp. 1011–1013, Aug. 2011.
- [15] T. H. Bao, D. Yakimets, J. Ryckaert, I. Ciofi, R. Baert, A. Veloso, J. Boemmels, N. Collaert, P. Roussel, S. Demuyne, P. Raghavan, A. Mercha, Z. Tokei, D. Verkest, A. V.-Y. Thean, and P. Wambacq, "Circuit and process co-design with vertical gate-all-around nanowire FET technology to extend CMOS scaling for 5 nm and beyond technologies," in *Proc. 44th Eur. Solid State Device Res. Conf. (ESSDERC)*, Sep. 2014, pp. 102–105.
- [16] D. Yakimets, G. Eneman, P. Schuddinck, T. Huynh Bao, M. G. Bardon, P. Raghavan, A. Veloso, N. Collaert, A. Mercha, D. Verkest, A. Voon-Yew Thean, and K. De Meyer, "Vertical GAAFETs for the ultimate CMOS scaling," *IEEE Trans. Electron Devices*, vol. 62, no. 5, pp. 1433–1439, May 2015.
- [17] T. Huynh-Bao, J. Ryckaert, S. Sakhare, A. Mercha, D. Verkest, A. Thean, and P. Wambacq, "Toward the 5 nm technology: Layout optimization and performance benchmark for logic/SRAMs using lateral and vertical GAA FETs," *Proc. SPIE*, vol. 9781, Mar. 2016, Art. no. 978102.
- [18] T. Song, "Opportunities and challenges in designing and utilizing vertical nanowire FET (V-NWFET) standard cells for beyond 5 nm," *IEEE Trans. Nanotechnol.*, vol. 18, pp. 240–251, 2019.
- [19] T. Song, "Many-tier vertical GAAFET (V-FET) for ultra-miniaturized standard cell designs beyond 5 nm," *IEEE Access*, vol. 8, pp. 149984–149998, 2020.
- [20] H. Na and T. Endoh, "A new compact SRAM cell by vertical MOSFET for low-power and stable operation," in *Proc. 3rd IEEE Int. Memory Workshop (IMW)*, May 2011, pp. 1–4.
- [21] M. Li, J. Shi, M. Rahman, S. Khasanvis, S. Bhat, and C. A. Moritz, "Skybridge-3D-CMOS: A fine-grained 3D CMOS integrated circuit technology," *IEEE Trans. Nanotechnol.*, vol. 16, no. 4, pp. 639–652, Jul. 2017.
- [22] D. Lee, C.-T. Ho, I. Kang, S. Gao, B. Lin, and C.-K. Cheng, "Many-tier vertical gate-all-around nanowire FET standard cell synthesis for advanced technology nodes," *IEEE J. Explor. Solid-State Comput. Devices Circuits*, vol. 7, pp. 52–60, 2021.
- [23] W.-C. Wang and P. Gupta, "Efficient layout generation and design evaluation of vertical channel devices," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 35, no. 9, pp. 1449–1460, Sep. 2016.
- [24] W.-C. Wang, C. Zhao, and P. Gupta, "Assessing layout density benefits of vertical channel devices," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 37, no. 12, pp. 3211–3215, Dec. 2018.
- [25] B. Chava, K. A. Shaik, A. Jourdain, S. Guissi, P. Weckx, J. Ryckaert, G. V. D. Plaas, A. Spessot, E. Beyne, and A. Mocuta, "Backside power delivery as a scaling knob for future systems," *Proc. SPIE*, vol. 10962, pp. 19–24, Mar. 2019.
- [26] P. Christie and D. Stroobandt, "The interpretation and application of Rent's rule," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 8, no. 6, pp. 639–648, Dec. 2000.

- [27] L. Liebmann, D. Chanemougame, P. Churchill, J. Cobb, C.-T. Ho, V. Moroz, and J. Smith, "DTCO acceleration to fight scaling stagnation," *Proc. SPIE*, vol. 11328, 2020, Art. no. 113280C.
- [28] C.-K. Cheng, C.-T. Ho, D. Lee, B. Lin, and D. Park, "ComplementaryFET (CFET) standard cell synthesis framework for design and system technology co-optimization using SMT," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 29, no. 6, pp. 1178–1191, Jun. 2021.
- [29] L. T. Clark, V. Vashishtha, D. M. Harris, S. Dietrich, and Z. Wang, "Design flows and collateral for the ASAP7 7 nm FinFET predictive process design kit," in *Proc. IEEE Int. Conf. Microelectron. Syst. Educ. (MSE)*, May 2017, pp. 1–4.
- [30] N. Björner, A.-D. Phan, and L. Fleckenstein, "vZ—An optimizing SMT solver," in *Proc. TACAS*, 2015, pp. 194–199.
- [31] D. Lee, D. Park, C.-T. Ho, I. Kang, H. Kim, S. Gao, B. Lin, and C.-K. Cheng, "SP&R: SMT-based simultaneous place-and-route for standard cell synthesis of advanced nodes," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 40, no. 10, pp. 2142–2155, Oct. 2021.
- [32] D. Park, D. Lee, I. Kang, C. Holtz, S. Gao, B. Lin, and C.-K. Cheng, "Gridbased framework for routability analysis and diagnosis with conditional design rules," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 39, no. 12, pp. 5097–5110, Dec. 2020.
- [33] (2020). *OpenCores: Open-Source IP Cores*. [Online]. Available: <https://opencores.org/>
- [34] (2020). *Cadence Innovus User Guide*. [Online]. Available: <http://www.cadence.com>
- [35] (2020). *LEF/DEF Language Reference*. [Online]. Available: <http://www.ispd.cc/contests/18/lefdefref.pdf>
- [36] A. Gupta, S. Kundu, L. Teugels, J. Bommels, C. Adelman, N. Heylen, G. Jamieson, O. V. Pedreira, I. Ciofi, B. Chava, C. J. Wilson, and Z. Tokci, "High-aspect-ratio ruthenium lines for buried power rail," in *Proc. IEEE Int. Interconnect Technol. Conf. (IITC)*, Jun. 2018, pp. 4–6.
- [37] B. Chava, J. Ryckaert, L. Mattii, S. M. Y. Sherazi, P. Debacker, A. Spessot, and D. Verkest, "DTCO exploration for efficient standard cell power rails," *Proc. SPIE*, vol. 10588, pp. 89–94, Mar. 2018.



CHUNG-KUAN CHENG (Life Fellow, IEEE) received the B.S. and M.S. degrees in electrical engineering from the National Taiwan University and the Ph.D. degree in electrical engineering and computer science from the University of California at Berkeley, Berkeley, in 1984.

From 1984 to 1986, he was a Senior CAD Engineer at Advanced Micro Devices Inc. In 1986, he joined the University of California at San

Diego, San Diego, where he is currently a Distinguished Professor with the Computer Science and Engineering Department and an Adjunct Professor with the Electrical and Computer Engineering Department. He worked as a Principal Engineer at Mentor Graphics, in 1999. His research interests include medical modeling and analysis, network optimization, and design automation on microelectronic circuits. He was a recipient of the best paper awards, IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN, in 1997, and in 2002, the NCR Excellence in Teaching Award, School of Engineering, UCSD, in 1991, IBM Faculty Awards, in 2004, 2006, and 2007, the Distinguished Faculty Certificate of Achievement, UJIMA Network, UCSD, in 2013, and the Cadence Academic Collaboration Award, in 2016. He was appointed as a Honorary Guest Professor at Tsinghua University (2002–2008) and a Visiting Professor with the National Taiwan University, in 2011 and 2015. He was an Associate Editor of IEEE TRANSACTIONS ON COMPUTER AIDED DESIGN, from 1994 to 2003, and since 2020.



CHIA-TUNG HO (Graduate Student Member, IEEE) received the B.S. and M.S. degrees in electrical engineering and computer science from the National Chiao Tung University, Hsinchu, Taiwan, in 2011 and 2013, respectively. He is currently pursuing the Ph.D. degree with the University of California at San Diego, La Jolla. He was worked at Macronix as a Principal CAD Engineer for developing in-house design for manufacturing (DFM) flow. Then, he worked at Novatek as a Senior Research and Development Engineer to design digital circuits. After that, he joined Mentor Graphic for developing fastSPICE and Synopsys as a Senior Research and Development Engineer. He is also working as a Ph.D. Resident in X, the Moonshot Factory. His current research interests include DTCO pathfinding, power delivery network (PDN) optimization, power grid simulation, reinforcement learning for circuit design, and machine learning in VLSI.



DAEYEAL LEE (Student Member, IEEE) received the B.S. and M.S. degrees in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 2006 and 2008, respectively. He is currently pursuing the Ph.D. degree with the University of California at San Diego, San Diego. Since then, he has been with Samsung Electronics, Hwaseong, South Korea, where he is also a Principal Engineer of the FLASH Memory Product Engineering Department. His research interests include the automation of VLSI design and statistical learning for IC design, manufacturing, and VLSI testing.



BILL LIN (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering and computer science from the University of California at Berkeley, Berkeley, in 1985, 1988, and 1991, respectively. He is currently a Professor in electrical and computer engineering with the University of California at San Diego, San Diego, where he is actively involved with the Center for Wireless Communications (CWC), the Center for Networked Systems (CNS), and the Qualcomm Institute in industry-sponsored research efforts. His research has led to over 200 journals and conference publications, including a number of best paper awards and nominations. He also holds five awarded patents. He has served as the general chair and on the executive and technical program committee of many IEEE and ACM conferences, and he has served as an associate editor or a guest editor for several IEEE and ACM journals as well.

...