

Impact and Design Guideline of Monolithic 3-D IC at the 7-nm Technology Node

Kyungwook Chang, Kartik Acharya, Saurabh Sinha, Brian Cline, Greg Yeric,
and Sung Kyu Lim, *Senior Member, IEEE*

Abstract—Monolithic 3-D (M3D) IC is one of the potential technologies to break through the challenges of continued circuit power and performance scaling. In this paper, for the first time, we demonstrate the power benefits of M3D and present design guideline in a 7-nm FinFET technology node. The predictive 7-nm process design kit (PDK) and the standard cell library using both high-performance (HP) and low-standby-power (LSTP) device technologies are developed based on NanGate 45-nm PDK using accurate dimensional, material, and electrical parameters from publications and a commercial-grade tool flow. We implement full-chip M3D designs utilizing industry-standard physical design tools, and gauge the impact of M3D technology on performance, power, and area metrics. We also provide the design guidelines as well as a new partitioning methodology to improve M3D design quality. This paper shows that M3D designs outperform 2-D counterparts by 16% and 16.5% on average in terms of isoperformance total power reduction with 7-nm HP and LSTP cell library, respectively. This demonstrates the power benefits of M3D technology in both HP and low-power future generation devices.

Index Terms—7-nm technology, design quality, FinFET, monolithic 3-D (M3D) IC.

I. INTRODUCTION

AS TECHNOLOGY scaling faces its physical limits in channel length scaling, degrading process variations, lithography constraints, increased parasitics, and rising manufacturing costs, monolithic 3-D (M3D) takes center stage in continuing Moore's law. In the M3D technology, the devices are fabricated onto multiple tiers sequentially with nanosized monolithic intertier vias (MIVs), which connect the top metal layer of the bottom tier and the bottom metal layer of the top tier. Because MIVs are extremely small, we can achieve much higher density and lower parasitics compared with through-silicon vias, which is another method of 3-D design. Thanks to the enhancement of fabrication technology such as higher alignment precision and thinner die, we can harness the true benefit of M3D with fine-grained vertical integration [1].

Manuscript received October 11, 2016; revised January 18, 2017; accepted February 20, 2017. Date of publication April 5, 2017; date of current version June 23, 2017. This work was supported by ARM Inc.

K. Chang, K. Acharya, and S. K. Lim are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: k.chang@gatech.edu; kartik.acharya@gatech.edu; limsk@ece.gatech.edu).

S. Sinha, B. Cline, and G. Yeric are with ARM Inc., Austin, TX 78735 USA (e-mail: saurabh.sinha@arm.com; brian.cline@arm.com; greg.yeric@arm.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TVLSI.2017.2686426

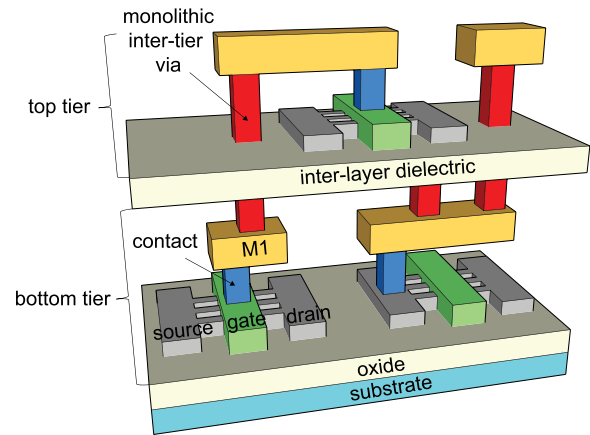


Fig. 1. Structure of M3D IC based on FinFET transistors.

This paper is to understand the benefit and tradeoffs involved in using gate-level M3D, which utilizes places standard cells into multiple tiers and uses MIVs only for intercell connections, at the end of silicon scaling, and hence, we have targeted the 7-nm technology node. By 7 nm, devices will have transitioned from planar to FinFET in order to counteract the limits of degrading short-channel effects, process variations, and reliability degradation. Therefore, important tools, and, to the best of our knowledge, the only open-source 7-nm FinFET transistor models for this study are Predictive Technology Models for Multi-Gate (PTM-MG) [2].

While M3D technology based on planar MOSFETs has been studied actively, FinFET implementations have not been widely explored. A study on the benefits of M3D on a 7-nm technology has been investigated in [3]. However, the authors manually derived intracell RC parasitics with a simple calculation instead of utilizing commercial tools for extraction. Their library also contains only six cells and did not consider the structure and effects of FinFET technology during cell design, which is prone to inaccuracies.

In this paper, we present the power benefit of gate-level M3D technology at the advanced technology nodes (i.e., 7-nm FinFET technology node) (see Fig. 1). The major contributions of this paper are as follows: 1) we developed a predictive 7-nm process design kit (PDK) based on FinFET transistors and corresponding high performance (HP) and low-standby-power (LSTP) standard cell libraries with 122 cells using commercial-grade electronic design automation (EDA) tools; 2) we used the developed

TABLE I
KEY PARAMETERS IN NANGate 45-nm AND OUR 7-nm PDK

Parameters		NanGate 45nm	Our 7nm
V_{DD} (V)		1.1	0.7 (-36.4 %)
L_G (μm)		0.0500	0.0125 (-75.0 %)
M1 Pitch (μm)		0.1400	0.0350 (-75.0 %)
Contacted Poly Pitch (CPP) (μm)		0.1900	0.0480 (-74.7 %)
Cell height (M1 track)		10TR	10TR (-75.0 %)
M1	width (μm)	0.0700	0.0174 (-75.1 %)
	thickness (μm)	0.1300	0.0348 (-73.2 %)
	diel. thickness (μm)	0.2500	0.0673 (-73.1 %)
	sheet resistance (Ω/\square)	0.3800	1.8200 (378.9 %)
VIA1	via resistance (Ω)	5.0000	36.4000 (628.0 %)
M4	width (μm)	0.1400	0.0350 (-75.0 %)
	thickness (μm)	0.2800	0.0700 (-75.0 %)
	diel. thickness (μm)	0.5700	0.1425 (-75.0 %)
	sheet resistance (Ω/\square)	0.2100	0.9070 (331.9 %)
VIA4	via resistance (Ω)	3.0000	8.7200 (190.7 %)
M7	width (μm)	0.4000	0.1000 (-75.0 %)
	thickness (μm)	0.8000	0.2000 (-75.0 %)
	diel. thickness (μm)	1.6200	0.4050 (-75.0 %)
	sheet resistance (Ω/\square)	0.0750	0.0950 (26.7 %)
VIA7	via resistance (Ω)	1.0000	0.8330 (-16.7 %)
M9	width (μm)	0.8000	0.2000 (-75.0 %)
	thickness (μm)	2.0000	0.4000 (-80.0 %)
	diel. thickness (μm)	4.0000	0.8000 (-80.0 %)
	sheet resistance (Ω/\square)	0.0300	0.0475 (58.3 %)
VIA9	via resistance (Ω)	0.5000	0.2960 (-40.8 %)

7-nm libraries to implement full-chip gate-level M3D designs using the state-of-art gate-level M3D design flow [4]; 3) we investigated the impact of gate-level M3D technology on power consumption at the 7-nm FinFET technology node for both HP and LSTP cells using the full-chip GDS layouts; and 4) we presented guidelines and a new clock tree partitioning methodology to maximize the benefit of M3D technology in the advanced technology nodes. To the best of our knowledge, this is the first work that studies both full-chip 2-D and M3D designs at the 7-nm FinFET technology node.

II. 7-nm PDK GENERATION

In order to properly evaluate the benefits of M3D technology on a 7-nm FinFET node, the corresponding PDK is needed for standard cell design and M3D synthesis, place and route (P&R). Since an open-source 7-nm FinFET PDK is not readily available to the research community, we created our own and validated it. We started with NanGate 45-nm PDK and scaled all technology parameters to values corresponding to the 7-nm node. Fig. 2 presents the procedure used to develop our predictive 7-nm PDK.

A. Technology Modeling and PDK Generation

The 7-nm PDK is defined based on the minimum dimensions of each layer in the process and accurate modeling of the transistor and interconnect behavior.

1) *Dimensional Scaling*: Table I shows the minimum dimensions and material properties assumed in the 7-nm PDK. Channel length scaling has been less aggressive in sub-45-nm technology nodes and is no longer the primary parameter defining the technology node. However, contacted poly-pitch (CPP) and M1 pitch scale by about 0.7 times every node and are better indicators of expected area scaling. Based

on industry trends and [5], we settled on the values of 35 nm for M1 pitch and 48-nm CPP for 7 nm. To scale the 45-nm layouts to 7-nm dimensions, we used the geometric mean of the M1 pitch and CPP to get our scaling factor of 0.25.¹

For interconnect dimensions, all $x - y$ dimensions of wires are scaled from the 45-nm PDK by the same scaling factor of 0.25, but the aspect ratios (thickness/width) are set to 2 based on ITRS projections. The dielectric thicknesses are scaled proportionately from the 45-nm PDK.

2) *Interconnect Modeling*: The 7-nm PDK requires accurate modeling of interconnect parameters, such as conductor sheet resistance, via, and contact resistance. We assumed copper (Cu) is used for metal layers, and the resistivity of the M1–M6 layers is determined to be 6.35 and 1.9 $\mu\Omega\text{-cm}$ for M7–M10, based on ITRS projections. One of the main reasons for the increased resistivity is the increased scattering experienced at grain boundaries within the Cu wires [6]. Due to the increased resistivity and the diminished cross-sectional area, the sheet resistances of the 7-nm technology are larger than that of NanGate 45-nm PDK, as shown in Table I.

For vias, Cu is assumed for via material with tantalum nitride (TaN) barrier. A barrier is necessary between a Cu via and the corresponding dielectric layer in order to prevent Cu atoms from diffusing into and contaminating the dielectric layer. The resistivity of Cu is based on ITRS projections, while the resistivity of TaN is determined to be 2000 $\mu\Omega\text{-cm}$ [7]. Table I presents the resulting via resistance for each layer.

Contacts from M1 to active and poly utilize tungsten (W) instead of Cu because of their excellent step coverage and gap fill abilities, especially for high-aspect ratio fills. Additionally, tungsten silicide (WSi₂) allows for low resistance contacts to the transistors. The resistivity of W contacts is determined to be 30 $\mu\Omega\text{-cm}$ as projected in [8], which yielded 27.3 and 46.14 Ω for the resistance of Active-M1 contacts and Poly-M1 contacts, respectively.

3) *Layout Scaling*: Ever since the introduction of multiple patterning for min-pitch metals in sub-20-nm nodes, tungsten local interconnects (also called middle of line, MOL layers) are used for cell level routing. Since standard cell layouts with these features are not available publicly, we scaled 45-nm layouts to 7-nm dimensions, but MOL layers are not modeled in this paper. This will result in some optimism when estimating cell level parasitics, but the larger scope of this paper remains unaffected, because important parameters such as transistor behavior and interconnect parasitics are accurately modeled. The goal of this paper is to understand important trends and tradeoffs when working with future technologies.

The cell widths and heights of the NanGate 45-nm library were shrunk along the $x - y$ dimension with the scaling factor derived in Section II-A1. For planar transistors, electron mobility is higher compared with holes, and hence, pMOS transistors are sized wider. In sub-45-nm technologies, strain engineering improves carrier mobility and has been an important knob to improve performance every technology node. Additionally, pMOS transistors benefit more from strain

¹Due to precision problems with the EDA tools, the scaling factor is rounded to two decimal places.

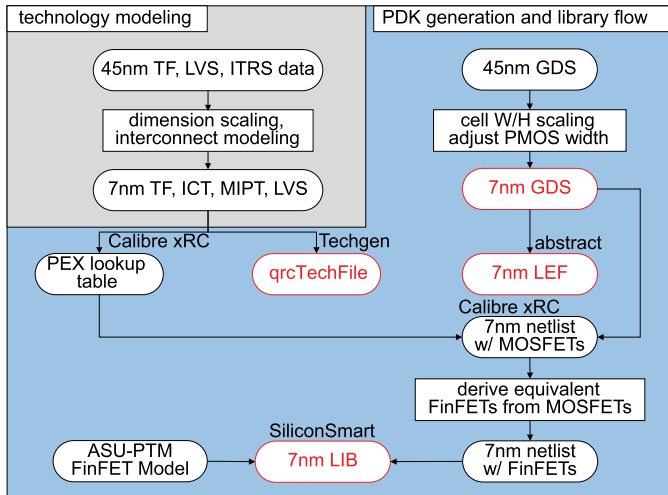


Fig. 2. Our 7-nm PDK generation flow (based on NanGate 45-nm PDK).

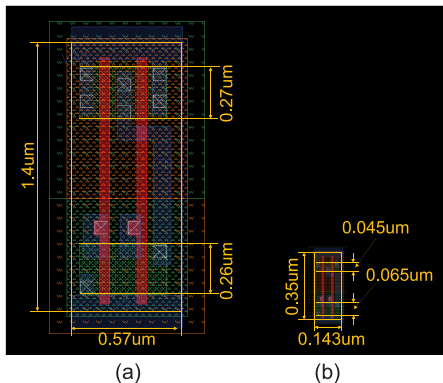


Fig. 3. Comparison of NAND2_X2 cell GDS layouts between (a) NanGate 45-nm PDK and (b) our 7-nm PDK.

resulting in nearly equal current drive strengths as nMOS [9]. Hence, after scaling the 45-nm planar layouts, the pMOS is sized equal to nMOS in order to balance cell rise and fall time.

An example 7-nm cell GDS layout of NAND2_X2 is compared with its NanGate 45-nm PDK layout in Fig. 3. As shown in the figure, though cell height and width are scaled down according to the geometric scaling factor, pMOS width is shrunk further to balance drive strength.

B. 7-nm Standard Cell Library

LEF views are created from 7-nm layout GDS files to be used for full-chip implementation by Cadence abstract generator. Interconnect dimensions and material properties discussed in Section II-A are coded in an MIPT file and is used to generate lookup tables for intracell parasitic data using Mentor Graphics Calibre xRC. These lookup tables, along with the scaled cell GDS layouts and LVS file, are used to extract 7-nm SPICE netlists with parasitics for every cell using Mentor Graphics Calibre xRC.

1) *Planar Width to Quantized Fins*: Since our 7-nm layouts are scaled from 45-nm layouts assuming planar transistors, the device widths have to be appropriately quantized to fins. The maximum number of fins in a standard cell is determined by the standard cell height and the ratio between metal pitch

TABLE II
MAXIMUM NUMBER OF FINS AND THE FINGER COUNT
IN VARIOUS DRIVE-STRENGTH INVERTERS

	max. # of fins	# of fingers
INV_X1	1	1
INV_X2	2	1
INV_X4	4	1
INV_X8	4	2
INV_X16	4	4
INV_X32	4	7

and fin pitch. We have the assumptions for dummy fins in our layouts, which are also required to make room for gate contacts between the FETs and to allow isolation between FETs in adjacent cell rows. Therefore, the number of fins in a pMOS and nMOS pair is limited by the number of M1 tracks subtracted by the number of dummy fins. As shown in Table I, our scaled design has 10 M1 tracks, and we assume a fin pitch of 25.5 nm to fit four fins per FET, which is in line with industry trends [5]. With an assumption of two dummy fins per pMOS and nMOS pair, dividing the transistor width by fin pitch gives us the number of fins for that device.

Table II shows the maximum number of fins as well as the number of fingers derived using our method for various drive-strength cells of inverter. The low drive-strength inverters (i.e., INV_X1 to INV_X4) gain strength by increasing their number of fins, while the high drive-strength inverters (i.e., INV_X8 to INV_X32) do so by increasing their number of fingers.

Using the method, we generated the new SPICE netlists with FinFETs. We, then, used the netlists and ASU PTM-MG FinFET transistor models for both HP and LSTP applications [10] to extract timing and power metrics (LIB) using Synopsys SiliconSmart.

C. 7-nm Library Characterization

We generated our 7-nm HP and LSTP libraries with total 122 cells. Table III shows cell delay, internal power-delay product (PDP), and leakage power of ten selected cells comparing NanGate 45 nm, 7-nm HP, and 7-nm LSTP libraries.² Fig. 4 also shows the $I-V$ characteristics of the transistor models used in cell characterization.

Compared with NanGate 45 nm, 7-nm HP library has 84.7% lower cell delay on average. Due to the decrease in cell delay, voltage scaling, and smaller input capacitance, the internal PDP of our 7-nm HP cells is reduced significantly. A leakage power consumption of 7-nm HP cells is also 69.5% smaller on average, mainly due to the reduced dimension and supply voltage even though 7-nm HP transistor model shows higher I_{OFF} . The 7-nm LSTP library has longer cell delay compared with 7-nm HP library because of smaller I_{ON} of lower leakage transistors as shown in Fig. 4, but the internal PDP of LSTP cells is lower than HP cells. Since 7-nm LSTP transistor model is designed to specifically reduce its leakage power consumption, it exhibits much smaller I_{OFF} than the

²In order to obtain a fair comparison between different technology nodes, we set input slew to be output slew of INV_X4, and output capacitance to be input capacitance of four INV_X4 cells of corresponding technology.

TABLE III
TIMING AND POWER COMPARISON BETWEEN NANGate 45 nm, OUR 7-nm HP, AND 7-nm LSTP LIBRARIES FOR TEN SELECTED CELLS

cell name	cell delay (ps)			internal PDP (fJ)			leakage power (nW)		
	45nm	7nm HP	7nm LSTP	45nm	7nm HP	7nm LSTP	45nm	7nm HP	7nm LSTP
AND2_X2	56.5	9.2 (-83.7 %)	18.7 (-67.0 %)	3.84	0.122 (-96.8 %)	0.104 (-97.3 %)	26.4	10.9 (-58.7 %)	0.028 (-99.9 %)
BUF_X4	44.4	7.3 (-83.5 %)	15.4 (-65.4 %)	16.25	0.186 (-98.9 %)	0.156 (-99.0 %)	41.5	13.8 (-66.7 %)	0.049 (-99.9 %)
DFF_X2	114.9	15.9 (-86.2 %)	33.4 (-70.9 %)	7.25	0.430 (-94.1 %)	0.396 (-94.5 %)	200.6	41.2 (-79.5 %)	0.139 (-99.9 %)
INV_X4	21.4	4.1 (-80.8 %)	8.3 (-61.4 %)	5.97	0.103 (-98.3 %)	0.084 (-98.6 %)	40.3	11.1 (-72.4 %)	0.012 (-100.0 %)
MUX2_X2	75.1	10.2 (-86.4 %)	20.8 (-72.2 %)	8.37	0.172 (-97.9 %)	0.145 (-98.3 %)	78.1	22.2 (-71.6 %)	0.061 (-99.9 %)
NAND2_X2	38.8	6.5 (-83.3 %)	12.6 (-67.5 %)	1.78	0.080 (-95.5 %)	0.066 (-96.3 %)	29.6	10.7 (-63.8 %)	0.012 (-100.0 %)
NOR2_X2	46.1	6.6 (-85.6 %)	12.9 (-72.0 %)	4.48	0.090 (-98.0 %)	0.078 (-98.3 %)	42.4	11.1 (-73.8 %)	0.014 (-100.0 %)
OR2_X2	59.7	9.2 (-84.6 %)	18.6 (-68.8 %)	9.06	0.114 (-98.7 %)	0.099 (-98.9 %)	32.7	11.0 (-66.3 %)	0.032 (-99.9 %)
XNOR2_X2	60.2	8.6 (-85.8 %)	16.9 (-71.9 %)	4.92	0.163 (-96.7 %)	0.138 (-97.2 %)	69.4	17.2 (-75.2 %)	0.046 (-99.9 %)
XOR2_X2	67.7	8.8 (-87.0 %)	17.4 (-74.2 %)	4.31	0.162 (-96.2 %)	0.141 (-96.7 %)	48.9	16.4 (-66.5 %)	0.038 (-99.9 %)

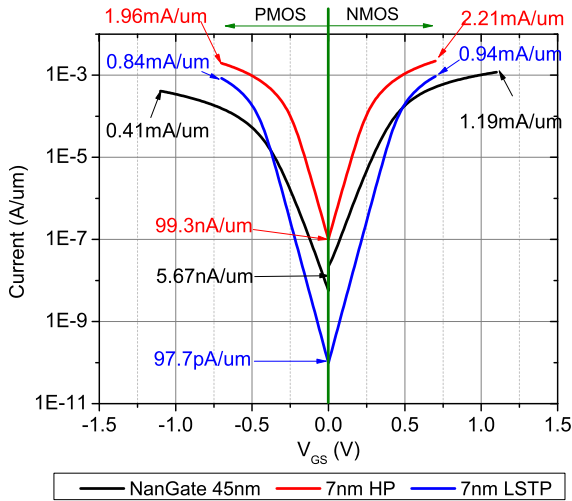


Fig. 4. Comparison of I_{ON} and I_{OFF} of unit width NanGate 45 nm, 7-nm HP, and 7-nm LSTP transistor models. Values for 7-nm transistor models are derived by measuring current flowing the transistor with single fin and normalizing the current by effective width ($2 * H_{FIN} * T_{FIN}$).

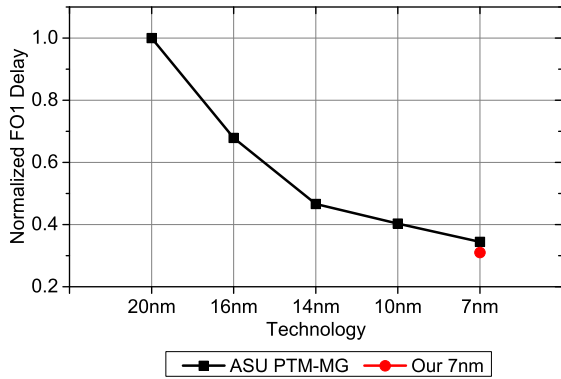


Fig. 5. Normalized FO1 cell delay of a ten-stage INV_X4 chain.

45-nm transistor model, which does not take leakage power reduction into account in its design.

Fig. 5 shows the comparison of the ten-stage FO1 INV delay between the projected values in [2] and our 7-nm extracted cell. Our INV cell delay is within 10% of the projections made in [2]. Considering that both approaches utilize the same transistor models, the plot shows the accuracy of our cell level parasitics and, hence, the efficacy of our PDK.

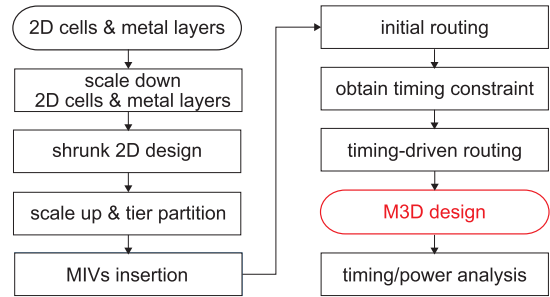


Fig. 6. CAD methodology flow for implementing a gate-level M3D design from 2-D cell library and metal stack used in [4].

III. FULL-CHIP MONOLITHIC 3-D IC DESIGN

A. Full-Chip M3D Design Flow

The methodology for implementing M3D designs is borrowed from [4]. Assuming that the z -dimension of an M3D design is negligibly small, this paper shows a CAD methodology, which implements a gate-level M3D design with two tiers. The overall flow is shown in Fig. 6. Since a 2-D design is divided into two tiers in the corresponding M3D design, the $x - y$ dimensions of all cells and metal layers (e.g., cell width, height, pin locations, metal width, and pitch) in the PDK are first scaled down by $1/\sqrt{2}$, so that all cells are placed into half of the area of the 2-D counterpart.

The shrunk cells and the metal layers are fed to Cadence Innovus, and all the design stages, including placement, pre-CTS optimization, CTS, post-CTS optimization, routing, and postroute optimization, are performed, generating a shrunk 2-D design. From the resulting shrunk 2-D design, only cell placement information, such as number of cells, their drive-strength, and $x - y$ location, is retained, and every other information including routing is discarded.

The cells in the shrunk 2-D design are, then, scaled up to the original size creating overlaps in the design. In order to remove the overlap, the design is divided into multiple square bins on xy plane, and for each bin, cells are partitioned into two tiers using area-balanced min-cut partitioning algorithm, so that half of the cells in the bin is assigned on the bottom tier and the other half on the top tier, determining z -location of each cell. After partitioning cells for all bins, cells in each tiers are legalized in order to remove overlap remained even after partitioning, deriving final $x - y$ location of cells.

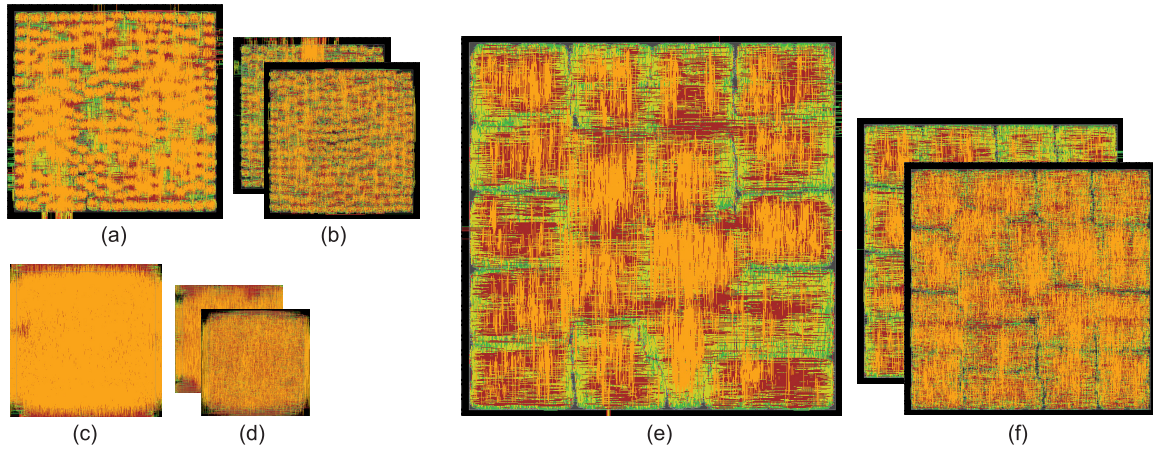


Fig. 7. AES-128 (a) 2-D, (b) M3D designs at 12 GHz, LDPC, (c) 2-D, (d) M3D designs at 2.5 GHz, FFT, (e) 2-D, and (f) M3D designs at 5 GHz implemented with 7-nm HP library.

With the legalized cell placement of both tiers, the location of MIVs is determined by utilizing a 2-D router that can route pins on multiple metal layers. First, all metal layers used in the design are duplicated, thereby generating a new metal stack. Then, we define two different types of every standard cell and macro, one for each tier. Pins for each cell type of cells and macros are mapped onto different metal layers depending on the tier. Pins in the bottom tier cells are located on the original metal layers, while pins in the top tier cells are located on the duplicated metal layers. After mapping all cells and macros onto their corresponding tier type, they are forced into the same placement layer. This structure is fed into *Cadence Innovus*, and routed. Then, the locations of MIVs are determined to be the location of vias connecting the original metal layers and the duplicated metal layers. The diameter of MIVs in the 7-nm technology node is set to be two times of its M1 width (i.e., $0.0348 \mu\text{m}$), and we set the resistance and capacitance of MIV as 64Ω and $0.1 fF$, respectively.

Once the MIV locations are determined, two separated designs, top tier and bottom tier designs, are created, and trial routing is performed for each tier using the location of legalized cells and MIVs of each tier. The trial-routed designs are fed into Synopsys PrimeTime to derive timing constraints for each tier. Once the timing constraints are determined, we run timing-driven routing, which results in the final M3D design. The M3D design is again fed into Synopsys PrimeTime for final timing and power analysis.

B. Full-Chip Design Analysis

To investigate the impact of M3D technology on power saving at the 7-nm technology, we synthesized and performed P&R 2-D and M3D designs with several different target clock frequencies for each benchmark using the developed 7-nm PDK. AES-128, FFT and LDPC from [11] are used as benchmarks. The chip area of each design is determined by targeting cell utilization to be 60% for AES-128 and FFT, and to 40% for LDPC, since LDPC is wire-dominated circuit, so that the chip area is determined not by cell utilization, but the available routing resources. Statistical power simulation

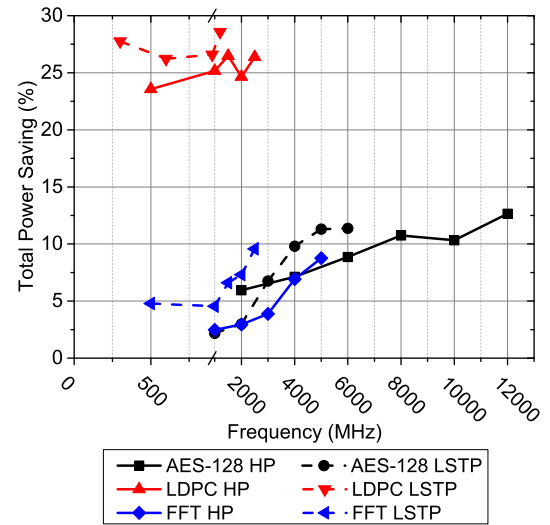


Fig. 8. Total power saving of M3D designs over the corresponding 2-D designs across target clock frequencies for three benchmarks implemented with 7-nm HP and 7-nm LSTP library.

is used to derive power metrics of the implemented designs, and considering dark silicon in the advanced technology nodes [12], only 30% of sequential logics (i.e., flip-flops) are assumed to be powered ON at an instance. The toggle rate for sequential logics and primary inputs is set to 40% and 20%, respectively. Fig. 7 shows 2-D and M3D implementations of the benchmarks at their highest clock frequency using 7-nm HP library. Tables IV and V show the design and the power metrics of the three benchmarks comparing M3D designs to their 2-D counterparts, respectively.

1) *Impact of Target Clock Frequency*: Fig. 8 shows the total power saving of M3D designs compared with their 2-D counterparts across target clock frequencies. Fig. 8 clearly shows a trend that the power benefit of M3D technology is increasing as the target clock frequency increases.

To interpret the trend, we use the following equation, which describes the components of dynamic power consumption of a design:

$$P_{\text{dyn}} = P_{\text{INT}} + \alpha \cdot (C_{\text{pin}} + C_{\text{wire}}) \cdot V_{\text{DD}}^2 \cdot f_{\text{clk}}. \quad (1)$$

TABLE IV

COMPARISON OF DESIGN METRICS OF 2-D AND M3D IMPLEMENTATIONS OF THE AES-128, LDPC, AND FFT WITH 7-nm HP, AND 7-nm LSTP LIBRARY. THE PERCENTAGE VALUES IN M3D DESIGNS ARE COMPUTED WITH RESPECT TO THEIR 2-D COUNTERPARTS

design	parameter	7nm HP		7nm LSTP	
		2D	M3D	2D	M3D
AES-128	clock frequency (MHz)	12,000	12,000 (0.0 %)	6,000	6,000 (0.0 %)
	footprint (μm^2)	140x139	101x101 (47.6 %)	140x140	99x99 (49.9 %)
	cell area (μm^2)	13,353	12,564 (5.9 %)	14,358	12,930 (9.9 %)
	wire-length (μm)	539,859	464,552 (13.9 %)	479,148	442,215 (7.7 %)
	avg. net size	3.008	2.790 (7.2 %)	3.022	2.854 (5.6 %)
	MIV count	-	47,370	-	46,676
LDPC	clock frequency (MHz)	2,500	2,500 (0.0 %)	1,200	1,200 (0.0 %)
	footprint (μm^2)	96x95	68x68 (49.8 %)	96x95	68x68 (49.5 %)
	cell area (μm^2)	4,130	3,970 (3.9 %)	4,258	3,787 (11.1 %)
	wire-length (μm)	630,826	493,876 (21.7 %)	617,071	464,610 (24.7 %)
	avg. net size	3.402	3.028 (11.0 %)	3.356	3.089 (8.0 %)
	MIV count	-	22,067	-	22,913
FFT	clock frequency (MHz)	5,000	5,000 (0.0 %)	2,500	2,500 (0.0 %)
	footprint (μm^2)	245x245	173x174 (49.9 %)	245x245	173x173 (50.0 %)
	cell area (μm^2)	39,452	37,156 (5.8 %)	37,591	36,463 (3.0 %)
	wire-length (μm)	1,424,382	1,159,516 (18.6 %)	1,361,842	1,036,165 (23.9 %)
	avg. net size	3.164	3.129 (1.1 %)	3.247	3.190 (1.8 %)
	MIV count	-	85,379	-	78,476

TABLE V

COMPARISON OF POWER METRICS OF 2-D AND M3D, AES-128, LDPC, AND FFT IMPLEMENTATIONS WITH 7-nm HP AND LSTP LIBRARY. THE PERCENTAGE VALUES IN M3D DESIGNS ARE COMPUTED WITH RESPECT TO THEIR 2-D COUNTERPARTS

design	parameter	7nm HP		7nm LSTP	
		2D	M3D	2D	M3D
AES-128	clock frequency (MHz)	12,000	12,000 (0.0 %)	6,000	6,000 (0.0 %)
	cell internal power (mW)	87.5	76.1 (13.0 %)	29.4	27.0 (8.2 %)
	net switching power (mW)	73.4	64.9 (11.6 %)	47.1	40.8 (13.4 %)
	leakage power (mW)	2.148	1.178 (45.2 %)	0.004	0.004 (0.0 %)
	total power (mW)	162.8	142.2 (12.7 %)	76.5	67.8 (11.4 %)
	clk power (mW)	0.783	0.775 (1.0 %)	0.533	0.529 (0.8 %)
LDPC	clock frequency (MHz)	2,500	2,500 (0.0 %)	1,200	1,200 (0.0 %)
	cell internal power (mW)	10.8	9.2 (14.8 %)	2.8	2.4 (14.3 %)
	net switching power (mW)	46.2	32.7 (29.2 %)	20.4	14.2 (30.4 %)
	leakage power (mW)	0.382	0.336 (12.0 %)	0.001	0.001 (0.0 %)
	total power (mW)	57.4	42.2 (26.5 %)	23.2	16.6 (28.4 %)
	clk power (mW)	0.154	0.140 (9.1 %)	0.100	0.096 (4.0 %)
FFT	clock frequency (MHz)	5,000	5,000 (0.0 %)	2,500	2,500 (0.0 %)
	cell internal power (mW)	139.7	130.4 (6.7 %)	37.4	34.8 (7.0 %)
	net switching power (mW)	75.3	65.7 (12.7 %)	31.6	27.6 (12.7 %)
	leakage power (mW)	4.122	3.835 (7.0 %)	0.013	0.012 (7.7 %)
	total power (mW)	219.1	199.9 (8.8 %)	69.0	62.4 (9.6 %)
	clk power (mW)	5.096	4.812 (5.6 %)	3.024	2.726 (9.9 %)

The first term, P_{INT} , is cell internal power of the cells, and the second term describes net switching power, where C_{pin} is the pin capacitance of the cells, C_{wire} is the wire capacitance in the design, and f_{clk} is the operating clock frequency.

The M3D technology offers power benefit in two ways, by reducing wire length and by reducing standard cell area. The first component of the power saving of M3D technology comes from the reduced C_{wire} value in the second term of 1 due to wire-length reduction. Since an M3D design utilizes short vertical connections (MIVs) instead of long horizontal wires in xy plane, it reduces the wire length of design in the half footprint of the 2-D counterpart. Since wire length is closely related to wire capacitance C_{wire} , as shown in Fig. 9, the reduced footprint of the M3D design helps to reduce the net switching power of the design. Fig. 9 also shows that the wire-length reduction of M3D designs tends to be

similar regardless of their operating clock frequencies. This is because the wire length saving comes from the structural characteristic of M3D design (i.e., the short vertical connection and the reduced footprint), which is not affected by the operating clock frequency of designs.

The second component of the power saving of M3D technology is attributed to its standard cell area saving. As wire length is reduced with the M3D technology, parasitics of wires are also decreased. The reduced parasitics help an M3D design to meet timing more easily. Hence, the M3D design utilizes less number of buffers and lower drive-strength cells, resulting in standard cell area reduction. Fig. 10 shows the cell drive-strength distribution of FFT 7-nm HP 2-D and M3D designs at 5000-MHz operating clock frequency with X1 being the smallest cell variant and X32 the largest cell variant. It is evident that the M3D design uses smaller cell sizes, utilizing

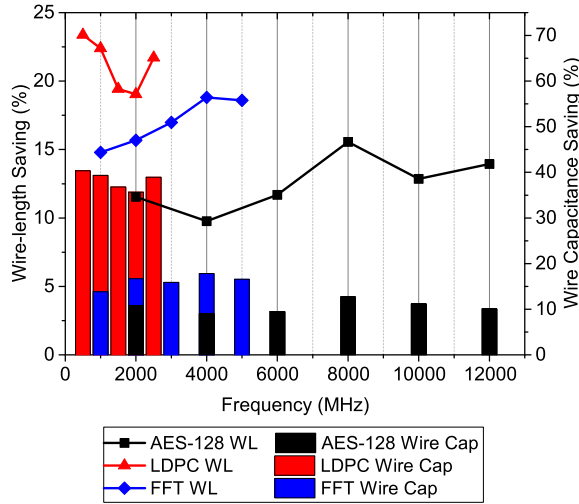


Fig. 9. Wire length and the corresponding wire capacitance saving of AES-128, LDPC, and FFT M3D designs utilizing 7-nm HP library.

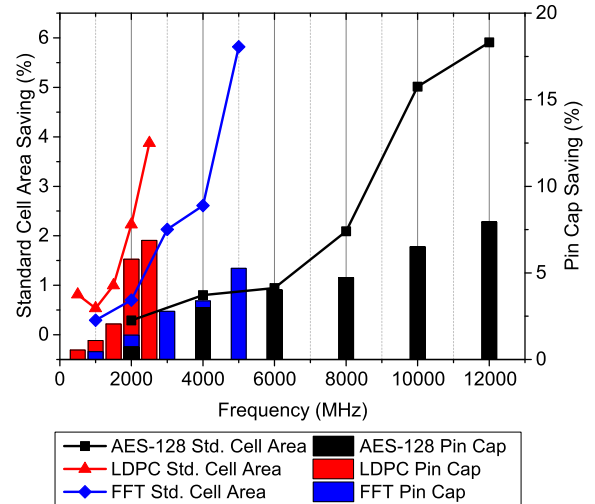


Fig. 11. Standard cell area and pin capacitance saving of M3D in AES-128, LDPC, and FFT 7-nm HP designs.

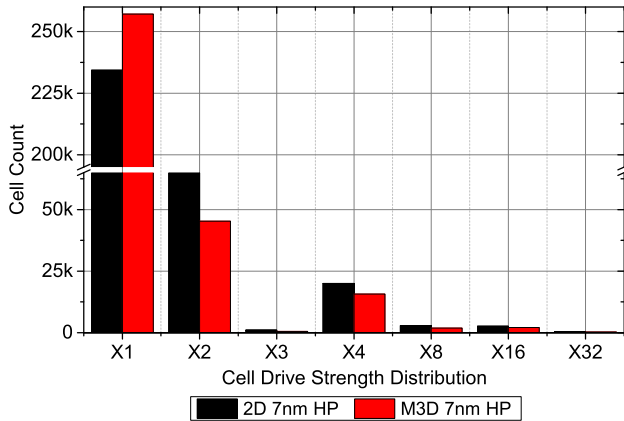


Fig. 10. Cell drive-strength distribution of FFT 7-nm HP 2-D and M3D implementations. Values are normalized to each cell variant of 2-D design.

more cells with X1 drive-strength cells, instead of using other larger variants. The standard cell area saving helps to reduce both the first term, P_{INT} , and the second term by reducing C_{pin} in 1. Fig. 11 shows the high correlation between standard cell area and pin capacitance of a design. Another trend to note in the figure is that unlike wire-length saving, the standard cell area saving is increasing as operating clock frequency increases. This is because 2-D designs utilize more buffers and higher drive-strength cells to meet timing especially in high operating clock frequency, whereas M3D designs meet timing with lesser number of buffers and lower drive-strength cells due to wire-length saving.

2) *Impact of Characteristic of Benchmarks*: We observe the significant difference in total power saving between LDPC and other benchmarks, as shown in Fig. 8. While the total power saving of AES-128 and FFT M3D designs ranges from 2.18% to 12.65%, LDPC M3D designs show 23.57%–28.59% power saving depending on their operating clock frequency. To explain the difference, we rewrite 1 as follows:

$$P_{dyn} = P_{INT} + a \cdot C_{pin} \cdot V_{DD}^2 \cdot f_{clk} + a \cdot C_{wire} \cdot V_{DD}^2 \cdot f_{clk}. \quad (2)$$

First, we observe that the ratio of the net switching power to the cell internal power of LDPC 7-nm HP 2-D design (4.27) is much higher than the other two designs (AES-128: 0.84 and FFT: 0.54) in Table V, which means that the first term is much smaller than the other two terms in 2 in the LDPC design.

In addition, from Table IV, we observe that the ratio of wire length to standard cell area of LDPC designs is $150.74 \mu\text{m}^{-1}$, which is also much higher than AES-128 ($40.43 \mu\text{m}^{-1}$) and FFT ($36.10 \mu\text{m}^{-1}$) designs. Hence, the ratio of C_{wire} to C_{pin} is much larger in LDPC designs than that in the other two benchmarks, since C_{pin} and C_{wire} are highly correlated with standard cell area and wire length of a design, respectively.

Considering that the standard cell area saving reduces the first two terms in 2, and the wire-length saving reduces the last term, we conclude that the total power saving of LDPC designs heavily depends on the wire-length saving of the design rather than the standard cell area saving.

Moreover, the M3D technology achieves more wire-length saving in LDPC than other two benchmarks. We observe that the ratio of wirelength to the footprint in an LDPC 7-nm HP 2-D design ($69.17 \mu\text{m}^{-1}$) is also larger than the other two designs (AES-128: $27.74 \mu\text{m}^{-1}$ and FFT: $23.73 \mu\text{m}^{-1}$). This metric along with the average net size in Table IV indicate that wires are more congested in LDPC. Since an M3D design helps reduce the wire congestion by utilizing both top and bottom metal layers, it effectively reduces wire length in LDPC designs. By comparing Fig. 7(c) and (d), we see the efficacy of M3D technology in reducing wire congestion of the design.

3) *Impact of Cell Library*: The trend difference between 7-nm HP and 7-nm LSTP designs can also be explained by the ratio of C_{wire} to C_{pin} in 2. Note that the wire parasitics of both libraries are remained the same, because they share the same metal layer configuration. Considering the fact that pin capacitance of 7-nm HP library cells tend to be higher than that of 7-nm LSTP library cells due to higher parasitics in HP devices, the ratio of the third term in 2 to the total dynamic power is larger in 7-nm LSTP designs than in 7-nm HP designs. This indicates that the designs with LSTP library

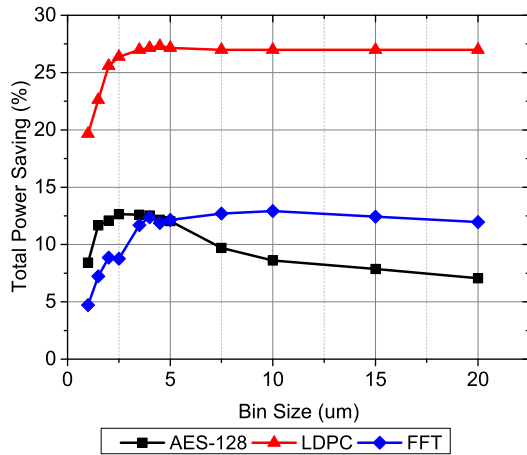


Fig. 12. Impact of bin size selection on the total power saving of the M3D design over the 2-D design for AES-128 at 12 GHz, LDPC at 2.5 GHz, and FFT at 5 GHz implemented with the 7-nm HP library.

cells take more advantage from wire-length saving of M3D technology than those using HP library cells. Since M3D designs reduce the wire length more effectively in LDPC compared with other two benchmarks as discussed in Section III-B2, we observe the largest total power saving in LDPC 7-nm LSTP M3D designs than others.

IV. M3D DESIGN SPACE EXPLORATION

Since the design quality of an M3D design heavily depends on the tier partitioning methodology of the M3D design flow described in Section III-A, we perform design space exploration on the tier partitioning scheme, specifically on bin size selection and clock tree partitioning.

A. Bin Size Selection

As described in Section III-A, in order to partition cells into two tiers, a shrunk 2-D design is, first, divided into multiple square bins on xy plane, and area-balanced min-cut partitioning algorithm is performed for each bin.

The impact of the bin size on the total power saving of M3D designs over 2-D designs is shown in Fig. 12. For each benchmark, twelve M3D designs are implemented from a shrunk 2-D design using different bin sizes, and the power consumption of the resulting M3D designs is compared with the 2-D design. Fig. 12 shows that the total power saving of all three benchmarks is maximum at the bin size of 3–4 μm , and the total power saving is ranged from 8.4% to 12.7% for AES-128 design depending on the bin size. Another important trend to note is that unlike LDPC and FFT M3D designs, the total power saving of AES-128 M3D designs sharply decreases after a 5- μm bin size.

Considering that cell count and drive strength of cells are not changed during tier partitioning step (they are determined and fixed while implementing the shrunk 2-D design), this difference on the total power saving comes mainly from the wire-length saving due to the bin size selection, which shows a similar trend, as shown in Fig. 13.

Fig. 14 explains the reason why the bin size selection affects the wire-length saving of an M3D design. If the bin

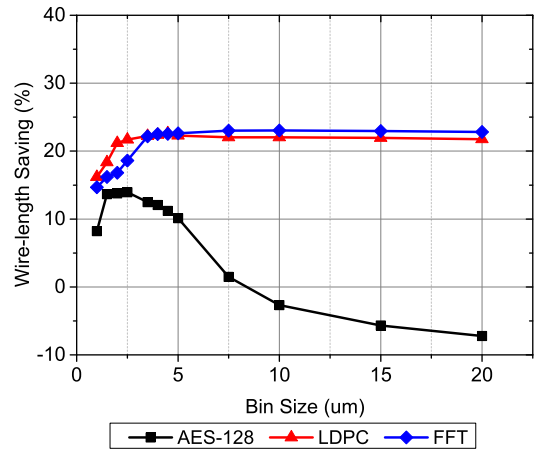


Fig. 13. Impact of bin size during tier partitioning on wire-length saving of 7-nm HP designs.

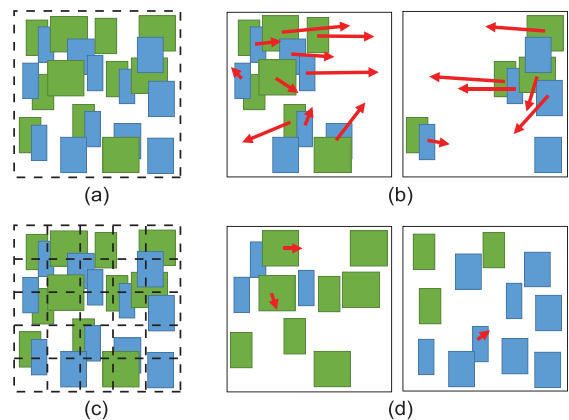


Fig. 14. Two extreme cases showing the impact of the bin size on the M3D design quality. (a) Shrunk 2-D design with only one huge bin after scaling cells up to their original size and (b) corresponding partitioning result. (c) Shrunk 2-D design with very small sized bins and (d) corresponding partition result. Dashed lines and red arrows indicate bins and cell movement during legalization, respectively.

size during tier partitioning is more than 5 μm (i.e., large bin size), area-balanced min-cut partitioning algorithm finds global optimal solution more easily, minimizing the number of connections between two tiers. However, neighboring cells in the local area tend to be clustered and placed on a single tier altogether, since cells with a global connection are more likely to be partitioned in this case, remaining overlap between cells even after tier partitioning. This overlap increases cell movement while legalizing cells [red arrows in Fig. 14(b)], resulting in wire-length overhead. The local cell clustering due to large bin size becomes more severe if a design has clustered design structure as AES-128 in Fig. 15(a), which indicates that the design has the large number of local wires, but very few global wires. This is because while the partitioning algorithm partitions very few global connections, the clustered neighboring cells in the local area tend to be placed on a single tier altogether, increasing wire-length overhead during cell legalization.

On the other hand, in the case of bin size less than 3 μm (i.e., small bin size), it does not suffer from overlap after tier partitioning because of its fine-grained partitioning scheme as shown in Fig. 14(c) and (d), but it is more likely to fall

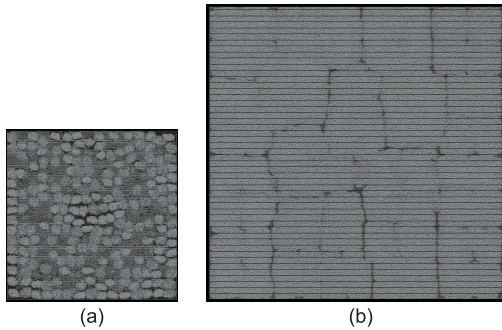


Fig. 15. Comparison of cell placement of (a) AES-128 and (b) FFT shrunk 2-D designs. AES-128 has clustered design structure, while FFT has evenly distributed cell placement.

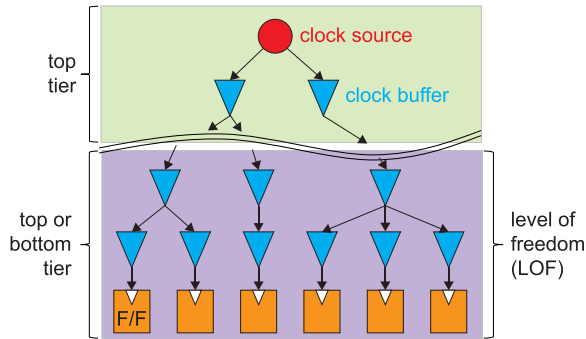


Fig. 16. Fixed cells and cells which are free to partition when $LOF = 2$. Note that the root of a clock tree, the clock source, is always placed on the top tier.

into local optimal solution during tier partitioning, splitting unnecessarily large number of local interconnects into two tiers, showing the sharp decrease in wire-length saving of M3D designs at the small bin sizes in Fig. 13.

B. Clock Tree Partitioning Methodology

In this section, two clock tree partitioning methodologies, level-of-freedom (LOF)-based partitioning [13] and prioritized clock tree partitioning, to improve the clock tree of an M3D design are presented. As discussed in Section III-A, $x - y$ location of clock cells in an M3D design, including clock buffers and flip-flops, is first determined during implementing its shrunk 2-D design. Then, during the tier partitioning step, z -location of each clock cells is determined depending on the clock tree partitioning methodology.

1) *Level-of-Freedom-Based Partitioning*: We define the LOF of clock cells as the distance from leaf nodes of a clock tree, and clock cells within the distance are free to be partitioned either bottom or top tier, as shown in Fig. 16. All other cells whose distance is larger than the LOF are fixed on one of the tiers (e.g., top tier) before partitioning. For example, if $LOF = 1$, only leaf cells of a clock tree (i.e., flip-flops), and the clock buffers that drive the leaf cells are free to be partitioned on either tiers, and every other clock cells are fixed on the top tier before partitioning.

Table VI shows the number of clock cells fixed to the top tier before tier partitioning depending on LOF and the resulting number of clock MIVs. Since clock MIVs are used when a

TABLE VI
NUMBER OF THE FIXED CLOCK CELLS FOR EACH LOF AND THE RESULTING CLOCK MIVs OF THE BENCHMARK DESIGNS IMPLEMENTED WITH 7-nm HP LIBRARY

LOF	AES-128		LDPC		FFT	
	# fixed clk cell	# clk MIVs	# fixed clk cell	# clk MIVs	# fixed clk cell	# clk MIVs
0	1,057	379	239	64	1,644	1,498
1	732	390	176	67	121	1,522
2	497	385	128	65	61	1,510
3	305	402	83	78	42	1,507
4	162	411	65	71	30	1,503
5	77	403	49	79	21	1,506
max	0	413	0	246	0	5,485

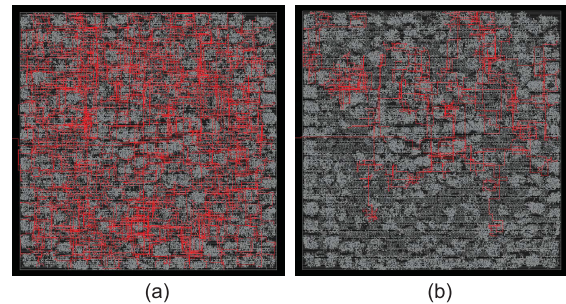


Fig. 17. Clock trees on the top tier span from the root of the clock tree to the first MIV encountered along the branches of AES-128 7-nm HP design when (a) $LOF = 1$ and (b) $LOF = \max$.

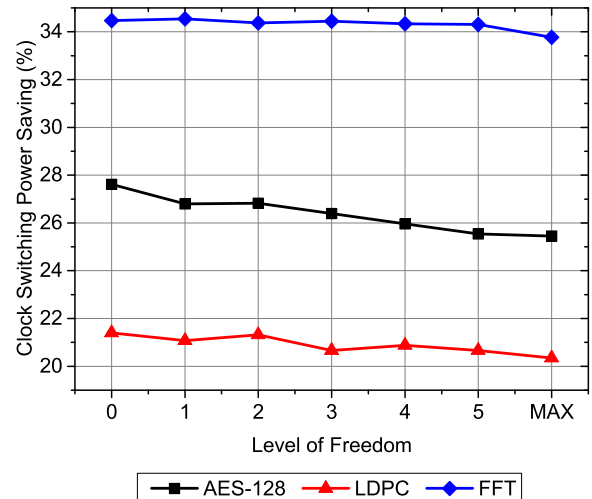


Fig. 18. Impact of LOF on the clock switching power saving of 7-nm HP M3D designs.

parent cell in a clock tree is assigned to the different tier from its child cell, the number of clock MIVs is increasing as LOF increases. Fig. 17 compares the clock trees on the top tier, which span from their root of the clock tree to the first MIV encountered along the branches when $LOF = 1$ and $LOF = \max$. Fig. 17 clearly shows that the clock tree spanning from the clock source is larger as LOF decreases, since more number of clock cells are fixed on to the top tier with low LOF.

Fig. 18 shows the impact of LOF on clock switching power saving of M3D designs over the corresponding 2-D designs, showing that clock switching power saving decreases as LOF

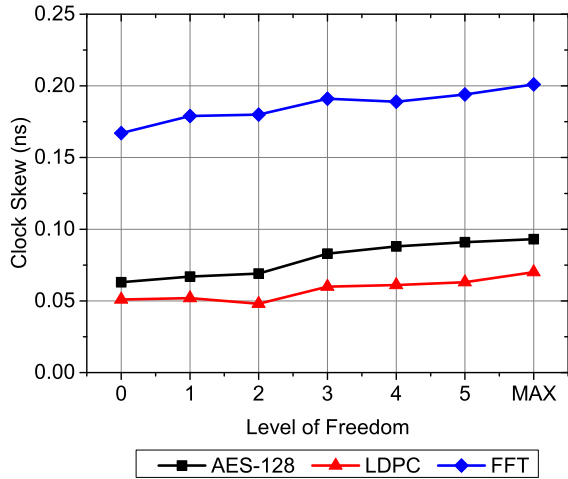


Fig. 19. Clock skew of 7-nm HP M3D designs depending on LOF in LOF-based partitioning.

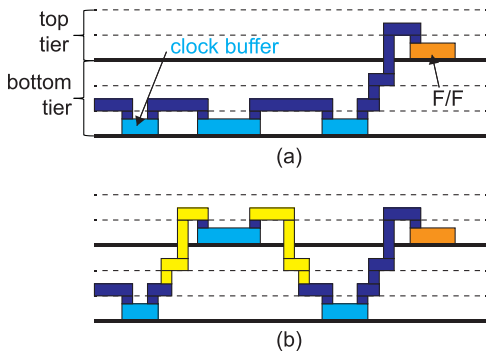


Fig. 20. Example comparing clock trees partitioned with (a) low LOF and (b) high LOF, showing that the extra wire (yellow wires) is required to connect clock cells on different tiers.

increases. However, total power consumption of the designs does not vary by a large magnitude, since the clock power consumption is less than 3% of the total power consumption for all benchmarks. On the other hand, the clock skew of the designs is significantly affected by LOF, as shown in Fig. 19.

The decreased clock switching power saving and the increased clock skew in an M3D design with high LOF are attributed to the increased clock wire length of the design. As LOF affects only on the tier assignment of clock cells (i.e., z -location), not on $x - y$ location of the cells, if LOF increases, the number of connection crossing the tiers is increasing, which, in turn, requires extra wires to establish the connection, as shown Fig. 20 (yellow wires). These extra wires differentiate clock delay at the clock sink of each branch, hence, increasing the clock skew of the M3D design. In addition, they also increase the parasitics of the clock tree, which increases the clock switching power consumption of the design.

2) *Prioritized Clock Tree Partitioning*: In this section, we propose prioritized clock tree partitioning methodology in order to further improve the quality of the clock tree of an M3D design. In the proposed methodology, a tier partitioning stage is divided into two phases, clock cell pretier partitioning

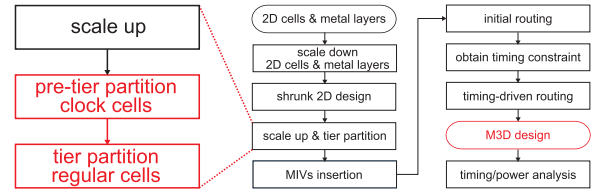


Fig. 21. Proposed gate-level M3D design flow with prioritized clock tree partitioning to improve the clock tree of a design.

ing phase and regular cell tier partitioning phase, as shown in Fig. 21. The entire clock tree of a design, which includes clock buffers as well as flip-flops, is first partitioned into two tiers using the area-balanced min-cut partitioning algorithm. The partitioned clock cells are fixed on the assigned tier and are not changed thereafter. Then, regular cells are partitioned, so that the cell area of the top and bottom tiers is balanced.

Different from LOF-based partitioning with $LOF = \max$, which treats clock cells as regular cells and partitions all cells at the same time, the proposed methodology gives clock cells higher priority over regular cells while balancing the area of clock cells on the top and the bottom tiers.

As discussed in Section IV-B1, in order to minimize the clock skew and the clock switching power of an M3D design, it is important to reduce the number of connection between clock cells crossing tiers. Therefore, in clock cell pretier partitioning phase of the proposed methodology, larger bin size is utilized for clock cells, compared with the bin size which would be used for regular cells, to minimize the cut size of the clock tree during area-balanced min-cut partitioning algorithm. The larger bin size may cluster neighboring clock cells in local area and place all of them onto one tier, as shown in Fig. 14(b). However, considering that the number of clock cells is small compared with total cell count of a design, the issue is automatically resolved during partitioning regular cells in the next phase by placing more regular cells on the other tiers. In this paper, the bin size for clock cell pretier partitioning is determined to be maximum in the range that regular cell tier partition is able to balance standard cell area between two tiers.

Table VII shows the bin size used for clock cell pretier partitioning, and compares the design and power metric of the proposed methodology with those with LOF-based partitioning with $LOF = 0$, which shows the best clock skew and clock switching power saving in Section IV-B1. The bin size used for clock cell pretier partitioning phase for AES-128 and LDPC 7-nm HP designs is set to be the same as entire design size (i.e., single bin for entire design). On the other hand, the bin size for FFT 7-nm HP design, which has higher clock cell density compared with other two benchmarks, is set to 1/8 of entire design size (i.e., 64 bins for entire design), since bin size larger than that clusters too many clock cells and place them into single tier, leading highly imbalanced clock cell area between tiers in local area, which cannot be resolved in regular cell tier partitioning phase.

As shown in Table VII, the proposed methodology successfully reduces the number of clock MIVs by finding global optima with a large bin size. The reduced clock MIVs mini-

TABLE VII
CLOCK METRIC COMPARISON OF THE PROPOSED METHODOLOGY
WITH LOF-BASED PARTITIONING WITH LOF = 1
FOR THE 7-nm HP DESIGNS

design	parameter	LOF=0	proposed
AES-128	bin size (<i>um</i>)	-	101
	# clock MIVs	379	118 (-68.9 %)
	clk skew (<i>ns</i>)	0.063	0.058 (-7.9 %)
	clk sw power (<i>mW</i>)	0.321	0.307 (-4.4 %)
LDPC	bin size (<i>um</i>)	-	68
	# clock MIVs	64	26 (-59.4 %)
	clk skew (<i>ns</i>)	0.051	0.043 (-15.7 %)
	clk sw power (<i>mW</i>)	0.072	0.070 (-2.4 %)
FFT	bin size (<i>um</i>)	-	21.75
	# clock MIVs	1,498	1,220 (-18.6 %)
	clk skew (<i>ns</i>)	0.167	0.138 (-17.4 %)
	clk sw power (<i>mW</i>)	1.954	1.942 (-0.6 %)

mize the extra wire for connecting clock cells on different tiers, achieving 2.5% clock switching power reduction and 13.7% clock skew reduction on average compared with LOF-based partitioning with LOF = 0. An interesting point to note is that while the clock switching power saving is proportional to clock MIV reduction because of the reduced wire-length overhead, the benefit on the clock skew does not because the reduced number of clock MIVs also reduces the clock latency of the shortest branch of the clock tree as well as the longest branch although it benefits more on the longest branch.

V. KEY FINDINGS

We summarize our findings when adopting M3D technology in the 7-nm node for low-power applications.

First, an M3D technology offers isoperformance power saving at the 7-nm technology node. We achieved significant power saving in both HP and LSTP 7-nm device models with M3D designs compared with their 2-D counterparts. This convincingly shows that M3D offers consistent power saving across device generations and target applications.

Second, an M3D design shows steady wire-length saving over operating clock frequency, which leads net switching power reduction. Wire-dominated designs using LSTP devices show more wire-length saving resolving congestion issue in a 2-D design.

Third, standard cell area saving of M3D technology is increasing as operating clock frequency increases, and the saving affects both cell internal power and net switching power of a design.

Fourth, designs with LSTP devices benefits more from the M3D technology than HP devices, because the ratio of wire capacitance to total capacitance of design is larger in LSTP designs.

Fifth, bin size during tier partitioning significantly affects the total power saving. While small bin size is prone to fall into local optimal solution during area-balanced min-cut partitioning, large bin size increases wire length during legalization of cells after partitioning especially for designs with clustered design structure.

Sixth, although clock tree partitioning methodology does not have large impact on total power saving of M3D designs at the advanced technology nodes, it affects clock skew of the

designs. With the LOF-based clock tree partitioning, in order to reduce clock skew of the M3D designs, fixing all clock buffers on a tier is recommended.

Seventh, the proposed clock tree partitioning methodology, prioritized clock tree partitioning, improves the quality of the clock tree of a design by giving clock cells higher priority over regular cells, which, in turn, reduces wire-length overhead during clock tree design of M3D designs.

VI. CONCLUSION

In this paper, we, for the first time, presented the impact of M3D technology on the power efficiency of 7-nm FinFET-based designs. We developed a predictive 7-nm PDK and a corresponding library using commercial-grade tools that accurately model dimensional and material properties accounting for device behavior, cell-level, and interconnect parasitics. We built full-chip GDS layouts of M3D design using the generated 7-nm PDK for both HP and LSTP applications. To improve the benefit of M3D technology further, we performed design space exploration of tier partitioning scheme in M3D design flow, and offered a guideline for bin size selection and proposed prioritized clock tree partitioning methodology. The simulation studies show that our M3D design offer significant power and area benefits over 2-D designs for future technologies using FinFETs.

REFERENCES

- [1] M. Okada *et al.*, "High-precision wafer-level Cu-Cu bonding for 3DICs," in *IEDM Tech. Dig.*, 2014, pp. 27.2.1–27.2.4.
- [2] S. Sinha *et al.*, "Design benchmarking to 7 nm with FinFET predictive technology models," in *Proc. Int. Symp. Low Power Electron. Design*, 2012, pp. 15–20.
- [3] Y.-J. Lee, D. Limbrick, and S. K. Lim, "Power benefit study for ultra-high density transistor-level monolithic 3D ICs," in *Proc. ACM Design Autom. Conf.*, 2013, p. 104.
- [4] S. A. Panth *et al.*, "Design and CAD methodologies for low power gate-level monolithic 3D ICs," in *Proc. Int. Symp. Low Power Electron. Design*, 2014, pp. 171–176.
- [5] M. Bardon *et al.*, "Group IV channels for 7 nm FinFETs: Performance for SoCs power and speed metrics," in *Symp. VLSI Technol. (VLSI-Technology), Dig. Tech. Papers*, Jun. 2014, pp. 1–2.
- [6] G. Lopez, R. Murali, R. Sarvari, K. Bowman, J. Davis, and J. Meindl, "The impact of size effects and copper interconnect process variations on the maximum critical path delay of single and multi-core microprocessors," in *Proc. IEEE Int. Interconnect Technol. Conf.*, Jun. 2007, pp. 40–42.
- [7] O. van der Straten, X. Zhang, K. Motoyama, C. Penny, J. Maniscalco, and S. Knupp, "ALD and PVD tantalum nitride barrier resistivity and their significance in via resistance trends," *ECS Trans.*, vol. 64, no. 9, pp. 117–122, 2014.
- [8] F. Liu *et al.*, "Subtractive W contact and local interconnect co-integration (CLIC)," in *Proc. IEEE Int. Interconnect Technol. Conf. (IITC)*, Jun. 2013, pp. 1–3.
- [9] S.-Y. Wu *et al.*, "A 16 nm FinFET CMOS technology for mobile SoC and computing applications," in *IEDM Tech. Dig.*, Dec. 2013, pp. 9.1.1–9.1.4.
- [10] S. Sinha *et al.*, "Exploring sub-20 Nm FinFET design with predictive technology models," in *Proc. ACM Design Autom. Conf.*, 2012, pp. 283–288.
- [11] *OpenCores*, accessed on Oct. 11, 2016. [Online]. Available: <http://www.opencores.org>
- [12] H. Esmaeilzadeh, E. Blem, R. S. Amant, K. Sankaralingam, and D. Burger, "Dark silicon and the end of multicore scaling," in *Proc. IEEE Int. Symp. Comput. Archit.*, Jun. 2011, pp. 365–376.
- [13] K. Acharya *et al.*, "Monolithic 3D IC design: Power, performance, and area impact at 7 nm," in *Proc. 17th Int. Symp. Quality Electron. Design (ISQED)*, Mar. 2016, pp. 41–48.



Kyungwook Chang received the B.S. degree in electrical and computer engineering and the M.S. degree in electrical engineering and computer science from Seoul National University, Seoul, South Korea, in 2007 and 2010, respectively. He is currently pursuing the Ph.D. degree with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA.

His current research interests include monolithic 3-D IC technology, advanced technology nodes, low power designs, and 3-D IC design methodology.



Kartik Acharya received the B.E. degree in electrical engineering from the Maharaja Sayajirao University of Baroda, Vadodara, India, in 1999, and the M.S. degree in electrical engineering from Syracuse University, Syracuse, NY, USA, in 2000. He is currently pursuing the Ph.D. degree with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA.

He is currently a Power Grid Analysis Methodology Engineer with IBM, Austin, TX, USA. His current research interests include electrical analysis

for power and noise integrity, and design methodologies for advanced technology nodes.



Saurabh Sinha received the B.Tech. degree in electronics and instrumentation engineering from the National Institute of Technology at Rourkela, Rourkela, India, in 2006, and the M.S. and Ph.D. degrees in electrical engineering from Arizona State University, Tempe, AZ, USA, in 2008 and 2011, respectively.

He is currently a Staff Research Engineer with ARM Research, Austin, TX, USA. His current research interests include predictive technology, design-technology co-optimization at advanced technology nodes, and disruptive technologies, such as 3D-ICs and their impact on design.



Brian Cline received the B.S. degree in electrical engineering from The University of Texas at Austin, Austin, TX, USA, in 2004, and the M.S. and Ph.D. degrees in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 2006 and 2010, respectively.

From 2006 to 2010, he was a Graduate Fellow with Semiconductor Research Corporation, Durham, NC, USA. He is currently a Principal Research Engineer with the ARM Research Group, Austin. His current research interests include design technology

co-optimization, low-power circuit design, variation-aware computer aided design tool development, and very large-scale integration design optimization for high-performance and low-power designs.



Greg Yeric received the Ph.D. degree in microelectronics from The University of Texas at Austin, Austin, TX, USA, in 1993.

He joined Advanced Products Research and Development Laboratories in Embedded Nonvolatile Memory Process Integration, Motorola, Austin, TX, USA, subsequently researching on technology development roles at TestChip Technologies, Plano, TX, USA, HPL Technologies, San Jose, CA, USA, and Synopsys, Mountain View, CA, USA. He is currently an ARM Fellow and the Director of Future Silicon

Research in ARM, Austin, where he is involved in design-technology co-optimization and predictive technology.



Sung Kyu Lim (S'94–M'00–SM'05) received the B.S., M.S., and Ph.D. degrees from the University of California at Los Angeles, Los Angeles, CA, USA, in 1994, 1997, and 2000, respectively.

He joined the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA, in 2001, where he is currently a Dan Fielder Endowed Chair Professor. He has authored the book *Practical Problems in VLSI Physical Design Automation* (Springer, 2008). His current research interests include modeling, architecture, and

electronic design automation (EDA) for 3-D ICs. His research on 3-D IC reliability is featured as Research Highlight in the Communication of the ACM in 2014. His 3-D IC test chip published in the IEEE International Solid-State Circuits Conference in 2012, is generally considered the first multi-core 3-D processor ever developed in academia.

Dr. Lim was a recipient of the National Science Foundation Faculty Early Career Development (CAREER) Award in 2006. He received the Distinguished Service Award in 2008. He received the Best Paper Awards from the IEEE Asian Test Symposium in 2012 and the IEEE International Interconnect Technology Conference in 2014. He received the Class of 1940 Course Survey Teaching Effectiveness Award from the Georgia Institute of Technology, in 2016. He was on the Advisory Board of the ACM Special Interest Group on Design Automation from 2003 to 2008. He was an Associate Editor of the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION SYSTEMS from 2007 to 2009. He has been an Associate Editor of the IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS since 2013. He has served on the Technical Program Committee of several premier conferences in EDA.