

# Power Benefit Study of Monolithic 3D IC at the 7nm Technology Node

Kyungwook Chang<sup>1</sup>, Kartik Acharya<sup>1</sup>, Saurabh Sinha<sup>2</sup>, Brian Cline<sup>2</sup>, Greg Yeric<sup>2</sup>, and Sung Kyu Lim<sup>1</sup>

<sup>1</sup>School of ECE, Georgia Institute of Technology, Atlanta, GA

<sup>2</sup>ARM Inc., Austin, TX

k.chang@gatech.edu, limsk@ece.gatech.edu

**Abstract**—Monolithic 3D IC (M3D) is one potential technology to help with the challenges of continued circuit power and performance scaling. In this paper, for the first time, the power benefits of monolithic 3D IC (M3D) using a 7nm FinFET technology are investigated. The predictive 7nm Process Design Kit (PDK) and standard cell library for both high performance (HP) and low standby power (LSTP) device technologies are built based on NanGate 45nm PDK using accurate dimensional, material, and electrical parameters from publications and a commercial-grade tool flow. In addition, we implement full-chip M3D GDS layouts using both 7nm HP and LSTP cells and industry-standard physical design tools, and evaluate the resulting full-chip power, performance, and area metrics. Our study first shows that 7nm HP M3D designs outperform 7nm HP 2D designs by 16.8% in terms of iso-performance total power reduction. Moreover, 7nm LSTP M3D designs reduce the total power consumption by 14.3% compared to their 2D counterparts. This convincingly demonstrates the power benefits of M3D technologies in both high performance as well as low power future generation devices.

## I. INTRODUCTION

As 2D scaling reaches its limit due to physical limits in channel length scaling, degrading process variations, lithography constraints and rising manufacturing costs, monolithic 3D IC (M3D) technology has come into the spotlight for continuing Moore's law. In M3D, the device layers are fabricated sequentially with nano-sized monolithic inter-tier vias (MIVs), which connect the top layer of the bottom tier and the bottom layer of the top tier. Because MIVs are extremely small, we can achieve much higher density and lower parasitics compared to through-silicon vias (TSVs), which is another method of 3D 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].

There are three different design styles in M3D: transistor-level, gate-level, and block-level M3D design. Transistor-level M3D design splits PMOS and NMOS into two tiers within a standard cell, and uses MIVs for intra-cell and inter-cell connections. It is the finest-grained design style, but takes significant effort because it requires completely new cell GDS layouts containing challenges in the power delivery network design. Gate-level M3D design, which is the focus of this paper, utilizes existing cells and places cells into tiers, using MIVs only for inter-cell connections. In block-level M3D design, functional blocks are floorplanned into multiple tiers. However, due to its coarse granularity, there is limit on fine-grained vertical integration.

One of the goals of this study is to understand the benefit and trade-offs involved in using M3D implementations at the end of silicon scaling, and hence we have targeted the 7nm technology node. By 7nm, devices will have transitioned from planar to FinFET in order to counteract the limits of degrading short channel effects, process variations and reliability degradation. Hence an important tool for

This research is supported by the Semiconductor Research Center under CADTS-2239.

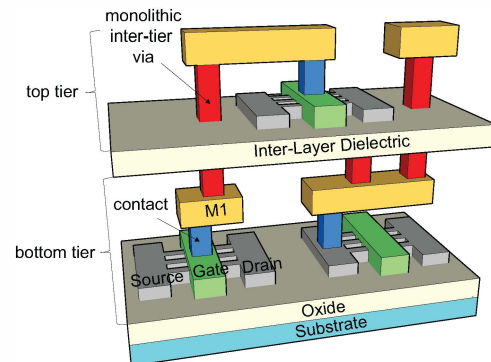


Fig. 1. A structure of monolithic 3D IC based on FinFET technology.

this study are predictive technology models for FinFETs (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 7nm technology has been investigated in [3]. However, the authors manually derived intra-cell RC parasitics with a simple calculation instead of utilizing commercial tools for extraction. Their library also contains only 6 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 benefits of gate-level M3D design at the 7nm technology node using FinFET transistors (see Fig. 1). The major contributions of our work are as follows: (1) We developed a predictive 7nm 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 EDA tools. (2) We used the developed 7nm libraries for gate-level M3D implementation. (3) We investigated the impact of our M3D technology on power consumption for both 7nm HP and LSTP cells using full-chip GDS layouts. To the best of our knowledge, this is the first work that studies full-chip designs at the 7nm node, both 2D and 3D implementations.

## II. 7NM PDK GENERATION

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

TABLE I  
KEY TECHNOLOGY PARAMETERS IN NANGate 45NM AND OUR 7NM PDK.

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

### A. Technology Modeling and PDK Generation

The 7nm PDK is defined based on 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 7nm PDK. Channel length scaling has been less aggressive in sub-45nm technology nodes and is no longer the primary parameter defining the technology node. However, contacted poly-pitch (CPP) and M1 pitch scale by about 0.7X every node and are better indicators of expected area scaling. Based on industry trends and [4], we settled on the values of 35 nm for M1 pitch and 48nm CPP for 7nm. To scale the 45nm layouts to 7nm dimensions, we used the geometric mean of the M1 pitch and CPP to get our scaling factor of 0.25<sup>1</sup>.

For interconnect dimensions, all X and Y wire dimensions are scaled from the 45nm 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 45nm PDK.

2) *Interconnect Modeling*: The 7nm 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 through M6 layers is determined to be 6.35 $\mu\Omega$ -cm, and 1.9 $\mu\Omega$ -cm for M7 through 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 [5]. Due to the increased resistivity and the diminished cross-sectional area, the sheet resistances of the 7nm technology are larger than that of NanGate 45nm 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

<sup>1</sup>Due to precision problems with the EDA tools, the scaling factor is rounded to 2 decimal places.

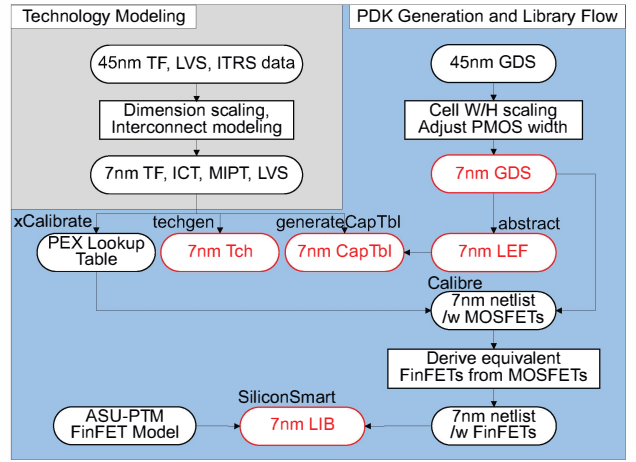


Fig. 2. Our 7nm PDK generation flow (based on NanGate 45nm PDK).

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 $\mu\Omega$ -cm [6]. Table I presents the resulting via resistance for each layers.

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 allows for low resistance contacts to the transistors. The resistivity of *W* contacts is determined to be 30 $\mu\Omega$ -cm as projected in [7], which yielded 27.3 $\Omega$  and 46.14 $\Omega$  for the resistance of Active-M1 contacts and Poly-M1 contacts, respectively.

### B. 7nm Standard Cell Library

1) *Layout Scaling*: Ever since the introduction of multiple patterning for min-pitch metals in sub-20nm 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 45nm layouts to 7nm dimensions, but MOL layers are not modeled in this study. This will result in some optimism when estimating cell level parasitics, but the larger scope of this study remains unaffected because important parameters such as transistor behavior and interconnect parasitics are accurately modeled. The goal of this study is to understand important trends and trade-offs when working with future technologies.

The cell widths and heights of the NanGate 45nm 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 to holes and hence PMOS transistors are sized wider. In sub-45nm 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 resulting in nearly equal current drive strengths as NMOS [8]. Hence, after scaling the 45nm planar layouts, the PMOS are sized equal to NMOS in order to balance cell rise and fall time.

An example 7nm cell GDS layout of NAND2\_X2 is compared with its NanGate 45nm 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.

LEF views are created from 7nm layout GDS files to be used for full-chip implementation. Interconnect dimensions and material

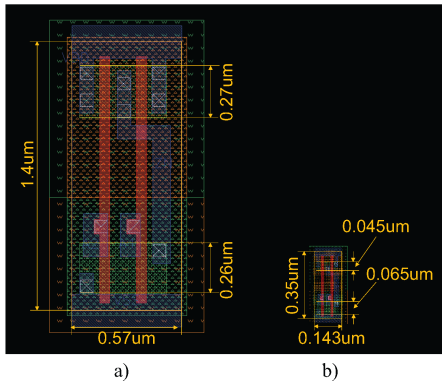


Fig. 3. Comparison of NAND2\_X2 cell GDS layouts between (a) NanGate 45nm PDK, (b) our 7nm PDK.

TABLE II

THE 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

properties discussed in previous subsections are coded in the MIPT file and is used to generate lookup tables for intra-cell parasitic data using *Mentor xCalibrate*. These lookup tables, along with the scaled cell GDS layouts and LVS file, are used to extract 7nm SPICE netlists with parasitics for every cell using *Mentor Calibre*.

2) *Planar Width to Quantized Fins*: Since our 7nm layouts are scaled from 45nm layouts assuming planar transistors, the device widths have to be appropriately quantized to fins. The number of fins in a standard cell is determined by the standard cell height and the ratio between metal pitch and fin pitch. We have 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 Table I shows, our scaled design has 10 M1 tracks, and we assume a fin-pitch of 25.5nm to fit 4 fins per FET, which is in line with industry trends [4]. With an assumption of 2 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 [9] to extract timing/power metrics (LIB) using *Synopsys SiliconSmart*.

### C. 7nm Library Characterization

We generated our 7nm HP and LSTP libraries with total 122 cells. Table III shows the comparison of cell delay, internal power-delay product (PDP) and leakage power of 10 selected cells between

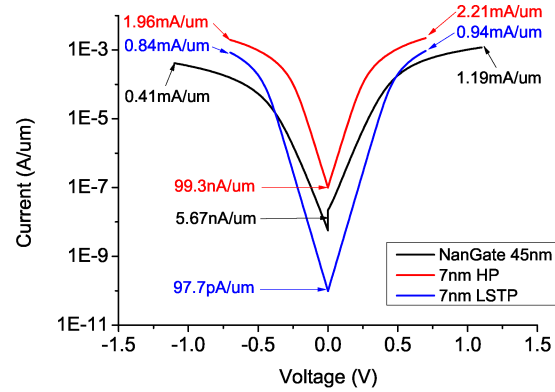


Fig. 4. Comparison of  $I_{on}$  and  $I_{off}$  in NanGate 45nm, 7nm HP, and 7nm LSTP transistor models.

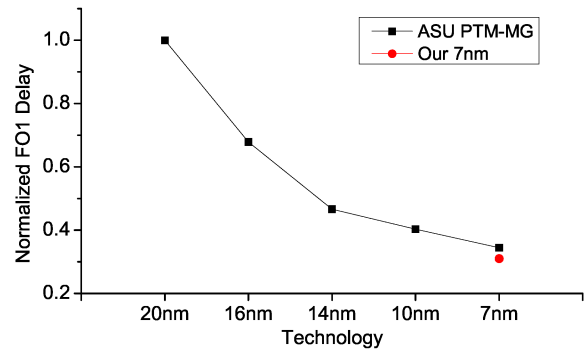


Fig. 5. Normalized FO1 cell delay of a 10-stage INV\_X4 chain.

NanGate 45nm, 7nm HP and 7nm LSTP libraries<sup>2</sup>. Fig. 4 also shows the I-V characteristics of the transistor models used in cell characterization.

Compared to NanGate 45nm library, 7nm HP library has 84.7% lower cell delay on average. Due to the decrease in cell delay,  $V_{DD}$  scaling and smaller input capacitance caused by the reduced dimensions, the internal PDP of our 7nm HP cells is reduced significantly (97.1% reduction on average). The 7nm LSTP library has longer cell delay compared to 7nm HP library because of lower leakage transistors but shows 69.1% cell delay reduction from NanGate 45nm library. Although the cell delay of 7nm LSTP cells is longer than 7nm HP cells, due to smaller  $I_{on}$  as shown in Fig. 4, the internal PDP of LSTP cells is lower than HP cells.

Fig. 5 shows the comparison of the 10-stage FO1 INV delay between the projected values in [2] and our 7nm 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.

## III. FULL-CHIP MONOLITHIC 3D IC DESIGN

### A. Full-chip M3D Design Flow

The methodology for M3D designs using each library is borrowed from [10]. Assuming that the z-dimension is negligibly small, the

<sup>2</sup>In order to obtain a fair comparison between different technology nodes, we set input slew to output slew of INV\_X4, and output capacitance to input capacitance of 4 INV\_X4 cells of corresponding technology.

TABLE III  
TIMING AND POWER COMPARISON BETWEEN NANGATE 45NM LIBRARY, OUR 7NM PDK HP, AND 7NM LSTP LIBRARIES FOR 10 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 %)

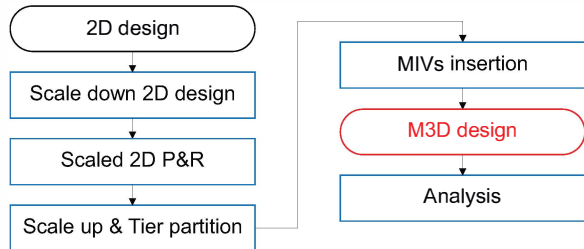


Fig. 6. The CAD methodology flow for generating M3D design from 2D design used in [10].

paper shows a CAD methodology that enables transforming a 2D design into a gate-level M3D design with 2 tiers. The overall flow for generating a M3D design with 2 tiers is shown in Fig. 6. Since the 2D design is divided into 2 tiers, the x-y dimensions of the initial 2D design (cell width, height, pin locations, metal width and pitch) are first scaled down by  $1/\sqrt{2}$ , so that all cells are placed into half the area achieved with 2D design.

The shrunk 2D design is fed to *Cadence Encounter*, and all the design stages including placement, post-placement optimization, CTS, routing, and post-route optimization are performed. The cells in the resulting shrunk 2D design are then scaled up to the original size causing overlaps in the 2D design. The overlapped 2D design is split into two tiers using a min-cut algorithm, so that half of the cells are located in bottom tier, and the other half in top tier.

To get separate designs for each tier, MIVs are inserted by utilizing a 2D router that can route pins on multiple metal layers. First, all metal layers within the cells are duplicated, thereby generating a new LEF. Then, we define two different types of cells, one for each tier. Pins for each cell type are mapped onto different layers depending on the tier (e.g. the bottom tier cells utilize the original metal layers, while the top tier cells use the duplicated metal layers). All the cells in the bottom and top tier are mapped onto their corresponding tier type, and forced into the same placement layer. This structure is fed into *Cadence Encounter*, and routed. Then, the locations of MIVs are determined, and the separate designs for each tier are generated. Table IV shows the major parameters of MIVs used for characterization in 45nm and 7nm technology.

Once the MIV locations are determined, the netlists for each tier are fed into *Synopsys PrimeTime* to derive timing constraints for each tier. Once the timing constraints are determined, we run timing-driven routing, and the result is again fed into *Synopsys PrimeTime* to get final timing/power metrics.

TABLE IV  
GEOMETRIC AND MATERIAL PROPERTIES OF MONOLITHIC INTER-TIER VIAS (MIVs) FOR 45NM AND 7NM TECHNOLOGY.

Parameters	45nm M3D	7nm M3D
MIV diameter (um)	0.1	0.025
MIV resistance ( $\Omega$ )	16	64
MIV capacitance (fF)	0.1	0.1

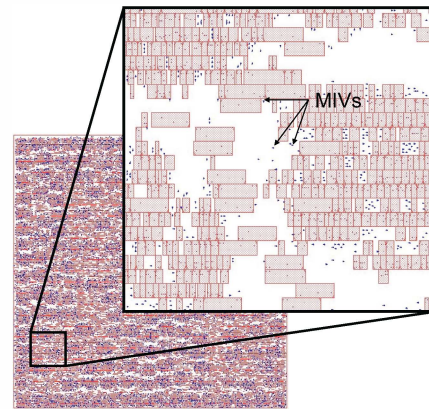


Fig. 8. Monolithic inter-tier via (MIV) placement in 7nm HP M3D implementation of AES.

### B. Full-Chip Design Analysis

To gauge the impact of M3D design on both 45nm and 7nm technology, we implemented full-chip designs using AES and JPEG obtained from [www.opencores.org](http://www.opencores.org). Fig. 7 shows the comparison of full-chip implementation of JPEG using 45nm, 7nm HP and 7nm LSTP library. Fig. 8 shows MIVs placement of 7nm HP M3D implementation of AES. Table V and Table VI show the iso-performance comparison of design metrics and power metrics of AES and JPEG between 45nm, 7nm HP and 7nm LSTP implementations, respectively.

1) *45nm M3D vs 45nm 2D*: As shown in Table V, our 45nm M3D implementations show smaller area compared to its 2D counterparts as the standard cell area is reduced by 23.6% on average. This reduction in standard cell area and MIV insertion result in wire-length reduction by 31.3% on average. The shorter wire-length also reduces the inter-cell parasitics, hence making it easier to meet timing. This affects the number of buffers inserted to meet timing closure, resulting in a significant drop in the total cell count (31.0% reduction).

These changes in design metrics also affect power consumption of the designs, as presented in Table VI. As the wire-length of designs is reduced, the net switching power of both design is reduced as well by

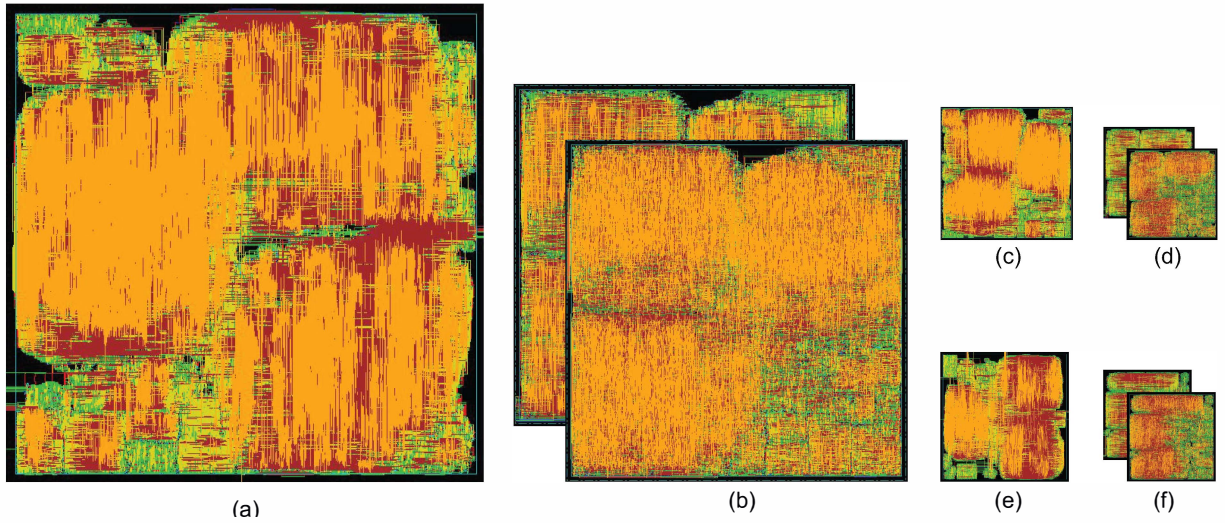


Fig. 7. GDSII die shots of a) 45nm 2D, b) 45nm M3D, c) 7nm HP 2D, d) 7nm HP M3D, e) 7nm LSTP 2D and f) 7nm LSTP M3D implementation of JPEG.

TABLE V  
COMPARISON OF VARIOUS DESIGN METRICS OF 2D AND M3D IMPLEMENTATION OF AES AND JPEG IN 45NM, 7NM HP, AND 7NM LSTP LIBRARY.  
THE PERCENTAGE VALUES IN M3D DESIGNS ARE COMPUTED WITH RESPECT TO THEIR 2D COUNTERPARTS.

Design	Parameter	45nm		7nm HP			7nm LSTP		
		2D	M3D	2D	M3D	%	2D	M3D	%
AES-128	footprint ( $\mu\text{m}^2$ )	700x700	500x500 (-49.0 %)	147x147	95x95 (-58.2 %)		134x134	95x95 (-49.7 %)	
	cell area ( $\mu\text{m}^2$ )	249,629	209,551 (-16.1 %)	13,195	11,194 (-15.2 %)		11,330	10,788 (-4.8 %)	
	cell count	195,683	155,703 (-20.4 %)	156,678	125,752 (-19.7 %)		127,056	126,809 (-0.2 %)	
	wire-length ( $\mu\text{m}$ )	3,645,882	2,753,181 (-24.5 %)	623,349	492,088 (-21.1 %)		473,124	366,947 (-22.4 %)	
	MIV count	-	86975	-	37539		-	41138	
JPEG	footprint ( $\mu\text{m}^2$ )	1050x1050	750x750 (-49.0 %)	291x291	200x200 (-52.8 %)		281x281	199x199 (-49.8 %)	
	cell area ( $\mu\text{m}^2$ )	995,379	686,893 (-31.0 %)	54,998	40,409 (-26.5 %)		51,685	40,095 (-22.4 %)	
	cell count	617,125	360,959 (-41.5 %)	489,385	304,704 (-37.7 %)		398,678	301,017 (-24.5 %)	
	wire-length ( $\mu\text{m}$ )	9,543,630	5,912,794 (-38.0 %)	2,504,079	1,348,328 (-46.2 %)		2,010,557	1,429,478 (-28.9 %)	
	MIV count	-	121082	-	105876		-	101846	

13.5% and 20.7% in AES and JPEG, respectively. The reduced total cell count results in an internal power reduction (9.7% on average), resulting in a total power reduction by 13.0% on average.

2) *7nm HP M3D vs 7nm HP 2D*: 7nm HP M3D designs pushes the bar even higher. The wire-length is reduced by 33.7% on average, leading to lower wire parasitics and better timing closure. The reduction in the wire-length also leads to a significant drop in the number of total cells in M3D designs (28.7% on average). This is because the placer and router need fewer buffers on signal and clock nets as the net distance among cells is lower in M3D implementation. We also see a 20.9% reduction on average in cell area because M3D designs use smaller drive-strength cells compared to its 2D counterparts.

Fig. 9 shows the drive-strength usage distribution in AES designs with X1 being the smallest cell variant and X32 the largest cell variant used during optimization. It is evident that M3D design uses smaller cell sizes which have lesser internal and leakage power. M3D designs have more cells with X1 drive-strength, and cell usage reduces significantly in M3D designs as we go from X4 to X32 variants. Both the observations are supported by the reduction in leakage and cell internal power in the benchmarks at 7nm HP.

We also notice a sharp drop in the net switching power by 40.7% in AES and 19.3% in JPEG which can be attributed to the reduced wire-length in M3D designs. This, combined with internal power

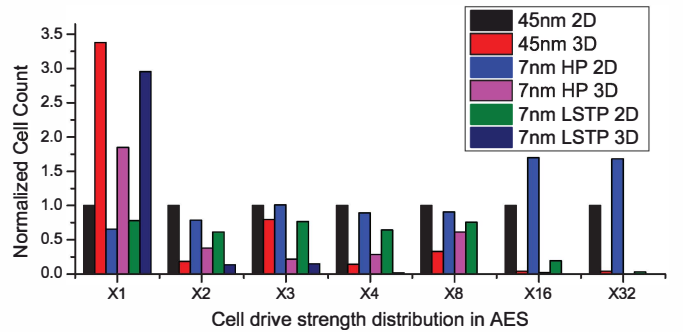


Fig. 9. Cell drive-strength distribution normalized to 45nm 2D in AES implementation using 45nm, 7nm HP, and 7nm LSTP Library.

reduction, leads to 18.0% total power reduction in AES and 15.6% in JPEG compared to their 2D implementations.

In AES, as we can easily meet timing with lower number of optimization buffers, the M3D design benefits more from net switching power than internal power. Considering the fact that in advanced technology nodes, net switching power is becoming more dominant in total power consumption due to the reduced dimensions, we can achieve more total power reduction in 7nm HP design than in 45nm designs as shown in Table VI.

TABLE VI  
COMPARISON OF PERFORMANCE AND POWER METRICS OF 2D AND M3D, AES AND JPEG IMPLEMENTATION OF 45NM, 7NM HP, AND 7NM LSTP LIBRARY. THE PERCENTAGE VALUES IN M3D DESIGNS ARE COMPUTED WITH RESPECT TO THEIR 2D COUNTERPARTS.

Design	Parameter	45nm		7nm HP		7nm LSTP	
		2D	M3D	2D	M3D	2D	M3D
AES-128	clock frequency (MHz)	870	870 (0.0 %)	5,000	5,000 (0.0 %)	2,500	2,500 (0.0 %)
	cell internal power (mW)	71.50	70.00 (-2.1 %)	45.60	44.00 (-3.5 %)	10.30	9.69 (-5.9 %)
	net switching power (mW)	24.40	21.10 (-13.5 %)	27.00	16.00 (-40.7 %)	8.35	6.83 (-18.2 %)
	clock switching power (mW)	21.00	18.30 (-12.9 %)	8.87	8.69 (-2.0 %)	3.68	3.72 (1.0 %)
	leakage power (mW)	3.436	1.832 (-46.7 %)	1.890	1.090 (-42.3 %)	0.003	0.002 (-25.9 %)
	total power (mW)	99.3	92.9 (-6.4 %)	74.5	61.1 (-18.0 %)	18.7	16.5 (-11.8 %)
JPEG	clock frequency (MHz)	467	467 (0.0 %)	870	870 (0.0 %)	196	196 (0.0 %)
	cell internal power (mW)	224.70	185.90 (-17.3 %)	50.50	44.90 (-11.1 %)	4.32	3.59 (-17.0 %)
	net switching power (mW)	69.00	54.70 (-20.7 %)	10.70	8.63 (-19.3 %)	2.16	1.80 (-16.4 %)
	clock switching power (mW)	64.60	51.90 (-19.7 %)	9.95	8.23 (-17.3 %)	1.97	1.66 (-15.9 %)
	leakage power (mW)	12.000	5.461 (-54.5 %)	6.093	3.232 (-47.0 %)	0.013	0.010 (-21.2 %)
	total power (mW)	305.70	246.10 (-19.5 %)	67.29	56.76 (-15.6 %)	6.49	5.40 (-16.8 %)

On the other hand, JPEG, which is a larger design with more number of nets and cells, the operating frequency is much lower than in AES, hence the M3D design mainly benefits from the total cell count reduction. However, as the technology advances, the wire parasitic is increased, hence reduction in buffer count is decreased. Therefore, we observed less benefit in cell internal power reduction in the 7nm HP design, resulting in relatively lower total power reduction compared to 45nm designs.

3) *7nm LSTP M3D vs 7nm LSTP 2D*: As discussed in the previous subsection, AES mainly benefits from net switching power reduction. In the 7nm LSTP design, the net switching power is already very low and dominated by the clock switch activity, hence the reduction in net switching power is lower than in 7nm HP design, although it is still higher than what we can achieve in the 45nm design. This leads to less total power saving (11.8%) compared to the 7nm HP design (18.0%).

In JPEG, because the net switching power is also highly dominated by the clock switch activity, the benefits from net switching power is relatively small compared to the HP counterpart. However, as the total cell reduction mainly affects the total power reduction, the relatively similar reduction in total cell count greatly affects total power consumption, showing greater total power reduction (16.8%) compared to its HP counterpart (15.6%).

#### IV. KEY FINDINGS

We summarize our findings when adopting M3D design in 7nm node for low power applications. First, we observed that as the cell internal power-delay product (PDP) of 7nm technology reduces significantly because of the increased via and sheet resistance, the net switching power becomes dominant. Therefore, reducing the wire-length is more important in scaled technologies in order to achieve the total power reduction. Second, M3D technology offers iso-performance power saving in both 45nm and 7nm nodes. In addition, we achieved significant power saving in both high performance and low power 7nm device models with M3D designs compared with their 2D counterparts. This convincingly shows that M3D offers consistent power saving across device generations and target applications.

Third, the saving in net switching power can be limited by the clock switching activity. We observed that with technology scaling, the power consumed by clock activity dominates the net switching power. Thus, the net switching power reduction from wire-length reduction can be limited in low power applications. But, M3D designs can still benefit from the reduction in cell internal power, especially in computation intensive designs. Fourth, M3D designs achieve power

saving mainly by buffer count and wire-length reduction. This leads to significant saving in cell internal power and net switching power, respectively. This saving is more prominent in larger-scale and/or wire-dominated designs. In addition, as the net switching power dominates the total power consumption in scaled technologies—especially for designs with high operating frequency—designers can achieve more total power savings with M3D technologies.

#### V. CONCLUSIONS

In this paper, we, for the first time, presented the impact of M3D technology on the power efficiency of 7nm FinFET based designs. We developed a predictive 7nm 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 7nm PDK for both HP and LSTP applications. The simulation studies show that our M3D design offer significant power and area benefits over 2D designs not only for older technology nodes with planar MOSFETs (i.e. 45nm technology) but also for future technologies using FinFETs.

#### REFERENCES

- [1] M. Okada *et al.*, “High-precision wafer-level Cu-Cu bonding for 3DICs,” in *Proc. IEEE Int. Electron Devices Meeting*, 2014.
- [2] S. Sinha *et al.*, “Design benchmarking to 7nm with FinFET predictive technology models,” in *Proc. Int. Symp. on Low Power Electronics and Design*, 2012.
- [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 Automation Conf.*, 2013.
- [4] M. Bardon *et al.*, “Group IV channels for 7nm FinFETs: Performance for SoCs power and speed metrics,” in *VLSI Technology (VLSI-Technology): Digest of Technical Papers, 2014 Symposium on*, 2014.
- [5] G. Lopez *et al.*, “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 Technology Conference*, 2007.
- [6] O. van der Straten *et al.*, “ALD and PVD Tantalum Nitride Barrier Resistivity and Their Significance in via Resistance Trends,” *ECS Transactions*, vol. 64, no. 9, pp. 117–122, 2014.
- [7] F. Liu *et al.*, “Subtractive W contact and local interconnect co-integration (CLIC),” in *Proc. IEEE Int. Interconnect Technology Conference*, 2013.
- [8] S.-Y. Wu *et al.*, “A 16nm FinFET CMOS Technology for Mobile SoC and Computing Applications,” in *Proc. IEEE Int. Electron Devices Meeting*, 2013.
- [9] S. Sinha *et al.*, “Exploring Sub-20nm FinFET Design with Predictive Technology Models,” in *Proc. ACM Design Automation Conf.*, 2012.
- [10] S. A. Panth, K. Samadi, Y. Du, and S. K. Lim, “Design and CAD Methodologies for Low Power Gate-level Monolithic 3D ICs,” in *Proc. Int. Symp. on Low Power Electronics and Design*, 2014.