Full-chip Monolithic 3D IC Design and Power Performance Analysis with ASAP7 Library

(Invited Paper)

Kyungwook Chang[1], Bon Woong Ku[1], Saurabh Sinha[2], and Sung Kyu Lim[1]
[1]School of ECE, Georgia Institute of Technology, Atlanta, GA
[2]ARM Inc., Austin, TX
k.chang@gatech.edu, limsk@ece.gatech.edu

ABSTRACT

In this paper, we present full-chip designs and their power, performance, and area (PPA) metrics using the ASAP7 process design kit (PDK) and library. Reliable cell library is a key element in evaluating new technological options such as monolithic 3D (M3D) ICs. Given an RTL, we conduct synthesis and place/route to obtain commercial-quality 2D and M3D IC designs and compare PPA. The ASAP7 library is highly useful to build high-quality designs that accurately reflect 7nm technology node. In addition, the full front-end and back-end access provided in ASAP7 allows us to see the impact of various device and interconnect parameters at the full-chip level for both 2D and monolithic 3D ICs. This work demonstrates the critical role of an academic PDK and library in enabling high-quality research in disruptive technologies such as M3D integration.

1. INTRODUCTION

3D integration has become one of potential technologies to continue Moore's law against the physical challenges on 2D device scaling such as lithography limitation and increasing parasitics on wire and contact. One of the well-known 3D IC techniques, through-silicon via (TSV)-based 3D ICs utilize 3D stacking by bonding pre-fabricated wafers with TSVs. Although it does not require considerable changes in the existing fabrication process, the large size of TSVs and their keep-out-zone (KOZ) hinder TSV-based 3D ICs from achieving dense vertical integration. Unlike TSV-based 3D stacking, in monolithic 3D (M3D) integration, transistors and metal wires are fabricated on multiple tiers sequentially, and connections between tiers are established by monolithic inter-tier vias (MIVs) as shown in Figure 1. Due to the sequential fabrication, the size of MIVs (50~100nm) are much smaller than that of TSVs (5~10μm), and hence, M3D ICs benefit from much higher vertical integration density as well as lower vertical connection parasitics.

Unfortunately, current commercial EDA tools do not support placing cells in 3D space. Therefore, several previous studies have shown alternative methodologies to accomplish cell placement on multiple tiers using 2D commercial EDA tools. In [1], the authors use dimensional shrinking technique to place cells in the half footprint of the 2D design, and partition the cells onto two tiers to obtain 3D cell placement. To tackle design rule check (DRC) violations accompanied by the dimensional shrinking, the authors in [2] utilize wire RC derating and cell projection technique to implement gate-level M3D designs. A block-level M3D implementation flow using design partition technique along with sets of dummy wires and anchor cells which mimic MIVs in 2D space is presented in [3].

While the performance, power, and area (PPA) benefit of M3D ICs in the existing technology nodes are explored widely [1, 4], those at the end of 2D device scaling is not identified and addressed
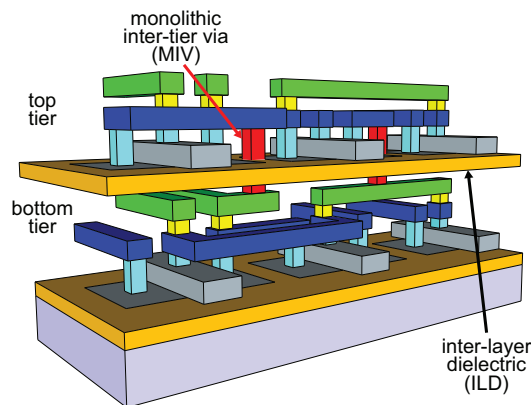

Figure 1: An example of monolithic 3D IC structure

actively. The insufficient research on future technology nodes are mainly because no open-source predictive library for future technology nodes are available to research community. Therefore, most of M3D integration studies on future technology nodes are performed either by directly scaling timing/power values in the library or by scaling some of key technology parameters from the existing technology nodes. The authors of [5, 6] present the power benefit of M3D ICs at 7nm technology node, but the approach to generate the 7nm cell library in [5] is based on a simple calculation with a scaling factor from intra-cell parasitics. In [6], the authors use the 7nm technology parameters based on previous publications, scaling layouts from NanGate 45nm library is prone to inaccuracy.

In sub-20nm technology nodes, devices have been transitioned from planar to FinFET to tackle degrading short channel effects, process variations and reliability degradation. Fortunately, as a result of recent active studies on device scaling and FinFET devices, an open-source library for 7nm technology node, which is known as ASAP7 process design kit (PDK) [7], is now available to research community. The library is known to be high quality based on historic trends and realistic parameters from previous publications, and it is very useful to build designs at 7nm technology node. In addition, since the back-end as well as front-end view is accessible, the library is a key enabler to perform in-depth analysis of PPA benefits of designs in 7nm technology node.

In this paper, we perform comprehensive study on the PPA benefits and trade-offs of M3D integration at the end of 2D device scaling utilizing ASAP7 library. The contribution of this work is as follows: (1) We use ASAP7 library to implement both 2D and gate-level M3D designs. (2) We perform in-depth studies on PPA benefits of M3D integration and examine the factors which impact the benefit. Finally, (3) we provide guidelines to maximize the ad-

Table 1: Key technology parameters used in ASAP7 library [7]

parameters		ASAP7 library
transistor model	V_{DD} (V)	0.7
	avail V_{th}	SLVT, LVT, RVT, SRAM
front end of line	L_G (nm)	21
	CPP (nm)	54
	fin pitch (nm)	27
	fin height (nm)	32
	fin thickness (nm)	6.5
	cell height (M2 track)	7.5TR
middle end of line	LISD width (nm)	24
	LISD pitch (nm)	54
back end of line	M1-M3 width (nm)	18
	M1-M3 pitch (nm)	36
	M4-M5 width (nm)	24
	M4-M5 pitch (nm)	48
	M6-M7 width (nm)	32
	M6-M7 pitch (nm)	64

vantage of M3D integration at 7nm technology node.

2. ASAP7 LIBRARY

2.1 Technology Details

In order to accurately estimate the advantage of M3D integration in 7nm FinFET technology node, a corresponding library is required. A 7nm library had not been available to research community, but recently, an open-source 7nm FinFET predictive PDK, ASAP7 PDK [7], is released. A 7.5 track standard cell library corresponding to the ASAP7 technology was also released. Table 1 summarizes the key parameters used in the library.

ASAP7 library utilizes BSIM-CMG SPICE models [9] with technology parameters from previous publications and the assumptions based on historic trends. In order to support multiple threshold voltages (V_{th}), they present four transistor models with different V_{th}s, which are SLVT, LVT, RVT, and SRAM in the order of decreasing drive-strength). SRAM devices are designed specifically for SRAM bit cell, which requires extremely low leakage power.

Assuming that fin pitch keeps being scaled at a rate of $0.8\times$ whereas fin height remains unchanged from previous technology nodes, 27nm fin-pitch and 32 nm fin-height has been used in the layout design. A contacted poly-pitch (CPP) of 54nm has been assumed, achieving a $0.9\times$ scaling factor from previous nodes. Middle of line (MOL) layers are processed with extreme ultraviolet (EUV) allowing two dimensional layout for local-interconnect source-drain (LISD).

In order to keep the design rules to be simple and maximize the cost effectiveness, they assume EUV lithography also for M1 to M3 of back end of line (BEOL), which results in 36nm pitch. On the other hand, self-aligned double patterned (SADP) is used for M4 to M7. The aspect ratio of metal to via is kept to 2. They use Copper (Cu) interconnects for metal layers in BEOL, and Cu vias with 1.5nm Tantalum Nitride (TaN) barrier layer (in order to prevent Cu atoms from contaminating the dielectric layer) with Ruthenium (Ru) liner (in order to reduce via resistance).

With the technology parameters and cell layouts, standard cell library containing 137 cells with four different VTs, and three corners for each VT (i.e. fast-fast (FF), typical-typical (TT), and slow-slow (SS) corner) are generated.

2.2 Tips for Design with ASAP7 library

In this section, based on our experience with ASAP7 library, we provide useful tips to implement designs with ASAP7 library.
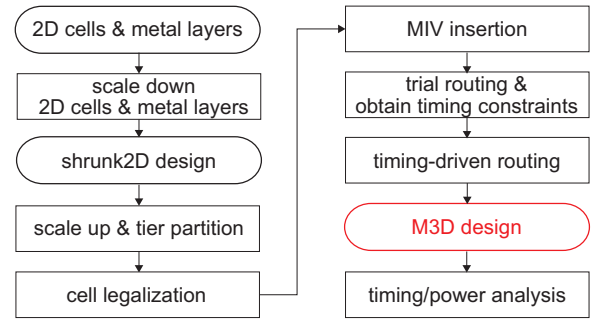
Figure 2: M3D design flow used in this work [1]

- In designs which have tight timing budget, during place-and-route (P&R), large number of buffers tend to be inserted as the maximum drive-strength of buffer cells in the library is $\times 13$. Larger drive-strength buffers can help in such designs.

- Due to the thin metal stack below M5 metal layer, in order to achieve static IR-drop less than 5% of nominal voltage (0.7V), setting top metal layer of power delivery network (PDN) as a metal layer above M6 is recommended assuming cell density of a design is more than 60%, the activity factor of primary input and sequential logic is 0.2 and 0.1, respectively.

- Target clock slew impacts the design quality and clock tree synthesis (CTS) time significantly. 20ps target clock is recommended as it prevents excessive clock buffer insertion (due to too small clock slew) as well as large power consumption due to high short-circuit current (due to too large clock slew).

- CTS time also heavily depends on the number of buffer and inverter candidates. Restricting the number of candidates around 10 helps reducing design time while maintaining the quality.

3. M3D DESIGN FLOW

Since current commercial EDA tools do not support 3D cell placement, [1] present a design flow which implements gate-level M3D designs using 2D EDA tools with dimensional shrinking technique, and Figure 2 shows the overview of the M3D design flow.

Assuming that the z dimension of a chip is negligibly small compared to its x-y dimensions, in order to place the cells of a M3D design with same standard cell area as the 2D counterpart but in two tiers (i.e., in the half footprint of the corresponding 2D design), the flow starts with scaling x-y dimension of cells and metal wires by $1/\sqrt{2}$. With the scaled cells and metal wires, a *shrunk2D design* is implemented with a regular P&R steps.

Once the *shrunk2D design* is implemented, only the cell information including the location of the cells is retained. Then, the size of all cells are scaled up to their original size, producing overlaps between the cells. To remove the overlaps and assign the cells onto the top and bottom tier (i.e., tier partitioning), we first divide the entire *shrunk2D design* into multiple square bins with same sizes, and perform min-cut area-balanced partitioning within each bin, so that half of the cells are placed on the top tier and the other half on the bottom tier while maintaining area skew as small as possible throughout the design. Then, cell legalization is performed for each tier, removing overlaps still remained even after tier partitioning.

In order to insert MIVs, the original metal stack is duplicated, so that the original metal stack represents the metal layers in the bottom tier, and the duplicated one accounts for the top-tier metal layers. The pin metal layer of every cell is annotated with the corresponding tiers, thereby pins of bottom-tier cells are located on the
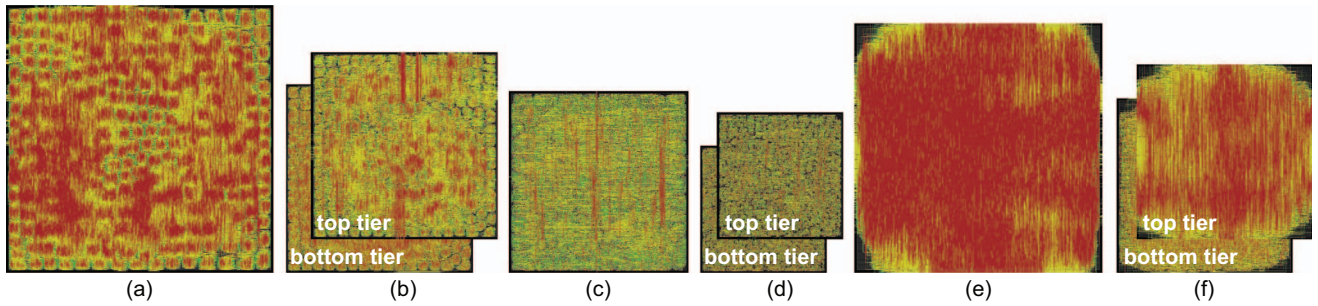
Figure 3: Die images of (a) AES-128 2D, (b) AES-128 M3D designs at 5GHz, (c) DCT 2D, (d) DCT M3D designs at 4GHz, (e) LDPC 2D, and (f) LDPC M3D designs at 2.5GHz using ASAP7 library

Table 2: Iso-performance comparison of the key design and power metrics of 2D designs with their M3D implementations. The percentage values in M3D designs are calculated with respect to their corresponding 2D designs.

metric	AES-128			DCT			LDPC		
	2D	M3D		2D	M3D		2D	M3D	
clock frequency (GHz)	5.0	5.0	(0.0 %)	4.0	4.0	(0.0%)	2.5	2.5	(0.0%)
footprint (um^2)	159×159	112×111	(-50.8%)	108×108	76×75	(-51.2 %)	148×147	104×104	(-50.3%)
cell count	205,459	199,617	(-2.8%)	92,518	88,087	(-4.8%)	72,765	70,096	(-3.7%)
std. cell area (um^2)	20,025	18,807	(-6.1%)	8,689	8,390	(-3.4%)	9,527	8,723	(-8.3%)
wire-length (um)	810,698	649,457	(-19.1%)	197,019	178,288	(-9.5%)	887,227	641,583	(-27.7%)
MIV count	-	61,709		-	18,918		-	28,299	
pin capacitance (pF)	239.9	227.3	(-5.3%)	98.3	95.5	(-3.8%)	104.1	97.3	(-6.6%)
wire capacitance (pF)	127.9	105.9	(-17.2%)	27.5	27.4	(-0.4%)	128.3	79.7	(-37.9%)
total capacitance (pF)	367.8	331.2	(-10.0%)	125.8	122.9	(2.3%)	232.4	177.0	(-23.9%)
switching power (mW)	161.5	128.2	(-20.6%)	44.8	41.1	(-8.3%)	45.7	36.8	(-19.5%)
internal power (mW)	75.1	71.7	(-4.5%)	31.5	30.2	(-4.2%)	28.6	24.9	(-12.9%)
leakage power (mW)	0.2	0.2	(0.0%)	0.1	0.1	(0.0%)	0.1	0.1	(0.0%)
total power (mW)	**236.9**	**200.2**	**(-15.5%)**	**76.3**	**71.4**	**(-6.4%)**	**74.4**	**61.8**	**(-16.9%)**

original metal stack, and those of top-tier cells utilizes the duplicated metal stack. The design with the duplicated metal stack and the annotated cells is routed, and the location of MIVs are determined as the location of vias connecting the top layer of the bottom tier and the bottom layer of the top tier.

Then, the netlists of each tier and the location of signal MIVs are used to perform trial routing, and we extract timing constraints of the top and bottom tier from the trial-routed design. With the timing constrains, timing-driven routing is performed on each tier producing the fully routed designs for each tier. With the top and bottom tier designs along with MIV parasitics, we perform static timing analysis (STA) to obtain timing/power metrics.

4. EXPERIMENTAL SETUP

In order to examine PPA benefits of M3D ICs over 2D designs at 7nm technology node, both M3D and 2D designs are implemented using AES-128, DCT, and LDPC designs from OpenCore as benchmarks. The LVT cell library of ASAP7 library is used to synthesize the logic and perform P&R. We use five metal layers of the ASAP7 library metal stack for 2D designs and both top and bottom tier of M3D designs. The footprint of each design is determined by setting the initial cell density with the synthesized netlist to 55% for AES-128 and DCT, 22% for LDPC designs since LDPC is a wire-dominated design, and the footprint of the design is determined by available routing resources rather than cell area. The target frequency of each benchmark is swept from 0.5GHz until their maximum frequency with 0.5GHz increments.

5. PPA ANALYSIS

We investigate the PPA benefits of M3D integration as well as analyze the impact of operating clock frequency, the characteristic of benchmarks, and the bin size selection in M3D design flow on the benefits. The GDS layouts and the key design/power metrics of both 2D and M3D designs of the benchmarks at their maximum frequencies are presented in Figure 3 and Table 2, respectively.

5.1 Impact of Operating Clock Frequency

The total power saving of the M3D ICs over the corresponding 2D implementations across the swept target clock frequencies are shown in Figure 4. We observe a clear trend of increasing power benefit as the target clock frequency increases.

To analyze the trend, Equation (1) is employed which describes the key components of dynamic power consumption of a design.

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

where P_{INT} is the internal power of a design, which represents power consumed by cells due to short-circuit current during cell input switching. The second term describes the switching power, which represents the power consumed by nets during wire-load dissipation. Wire-load of net is further decomposed into pin capacitance (C_{pin}) and wire capacitance (C_{wire}). α, V_{DD}, and f_{clk} are the switching activity factor of the net, the supply voltage, and the operating clock frequency, respectively.

The main advantage of M3D integration comes from wire-length reduction. Although M3D designs use same silicon area as their 2D counterparts, owing to multiple tiers, the wire-length between cells
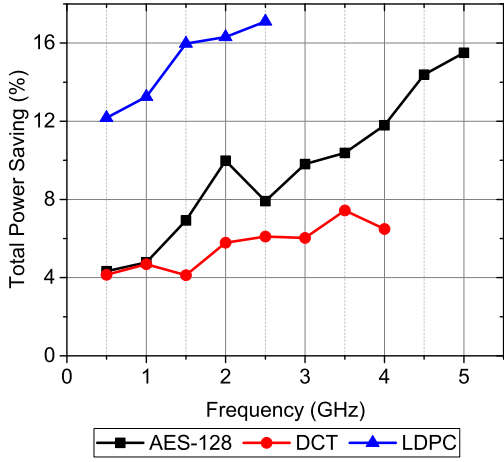
Figure 4: The total power saving of M3D implementations with respect to the 2D counterparts in 7nm technology node
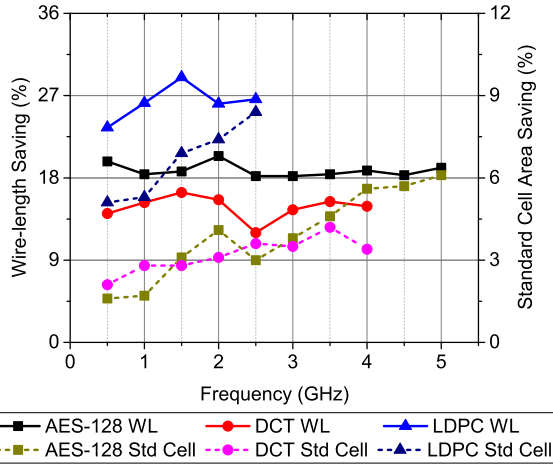

Figure 5: The wire-length and standard cell area saving of the implemented AES-128, DCT, and LDPC M3D designs across their swept target clock frequency points
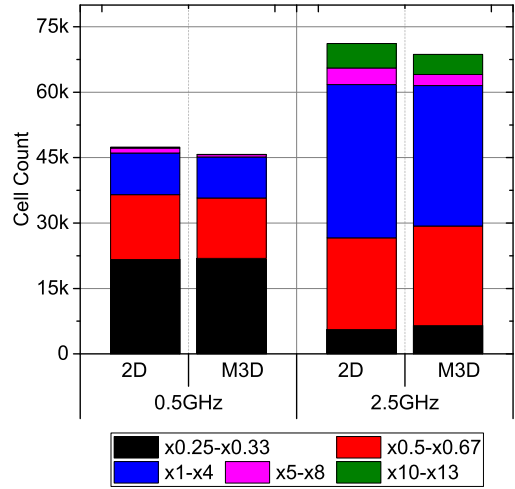

Figure 6: Comparison of the total cell count and cell drive-strength distribution of LDPC 2D and M3D implementations with the lowest and highest target clock frequency points

higher as target clock frequency increases. Whereas more number of buffers and higher drive-strength cells are utilized in 2D design in order to meet tight timing budget in high target clock frequency, M3D designs easily meet timing with fewer number of buffers and lower drive-strength cells due to the reduced wire parasitics. Figure 6 compares the standard cell count as well as cell drive-strength distribution of 2D and M3D LDPC designs at their lowest and highest target clock frequency points. The figure clearly shows M3D designs utilize fewer number of cells as well as smaller drive-strength cells, instead of using other larger variants, especially at higher target clock frequency.

It is important to note that at 7nm FinFET technology node, as FinFET based technologies have higher pin capacitance due to the 3D fin structure and the introduction of local interconnect MOL layers that contact the device terminals to M1, the ratio of C_{pin} to C_{wire} is higher than planar CMOS based technology nodes, and hence, the standard cell reduction becomes more important than wire-length reduction in order to achieve high total power reduction with M3D designs. Therefore, M3D designs with FinFET based designs at advanced technology nodes benefits more as target clock frequency increases compared to planar CMOS based designs.

5.2 Impact of Benchmarks

As shown in Figure 4, LDPC M3D implementations offer better power saving over the 2D counterparts compared to other benchmarks, showing up to 16.9% total power reduction.

The difference in the power saving is first attributed to two factors, the ratio of P_{INT} to the second term of Equation (1) which is the switching power, and the ratio of C_{pin} to C_{wire}. Since LDPC is a wire-dominated circuit, we observe that the switching power dominates the total power consumption of the LDPC 2D design, and furthermore, the ratio of C_{pin} to C_{wire} is much smaller in the LDPC design than other two benchmarks as shown in Table 2. Therefore, the switching power due to wire (i.e., C_{wire} power dissipation) occupies larger portion in total power consumption than P_{INT} and the switching power due to pin switching (i.e., C_{pin} power dissipation), and hence, the wire-length saving of the LDPC M3D designs has bigger impact on total power saving compared to other two benchmarks. Also, since the wire-length reduction

are reduced utilizing short vertical connections offered by MIVs and the reduced footprint. This effectively reduces C_{wire} in the second term of Equation (1), resulting in the reduced switching power. As the wire-length reduction is attributed to the reduced footprint and short vertical connection of M3D designs, it is not affected by the target clock frequency, and hence, the reduction is similar across the swept clock frequencies as shown in Figure 5.

In addition, the reduced wire-length affects the standard cell area saving, which is another major component of the power reduction of M3D ICs. The reduced wire parasitic, which comes from wire-length saving, eases meeting timing of M3D designs, allowing to use less number of cells and lower drive-strength cells, and hence, resulting in standard cell area reduction. The standard cell area saving reduces both P_{INT} and C_{pin} in Equation (1) as short-circuit current becomes smaller and the number of total transistors reduces. Figure 5 also shows an important trend of the standard cell area reduction depending on the target clock frequency. Unlike wire-length saving, the standard cell area saving tends to be

Table 3: Metal usage comparison of 2D and M3D designs for the benchmarks used in this paper. M3D △% is derived with respect to their corresponding metal layers in 2D counterparts. Metal usage of M3D is the averaged metal usage of the top and bottom tier.

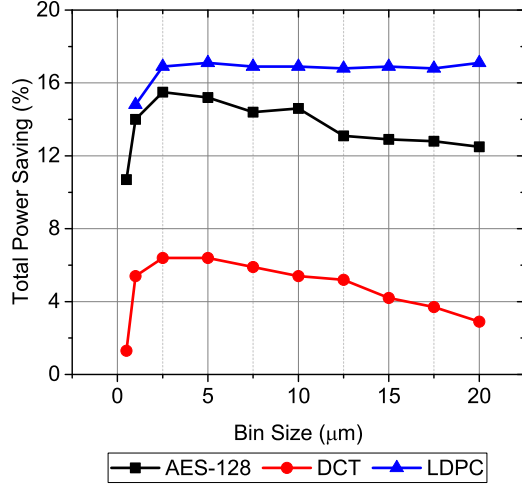| metal	AES-128			DCT			LDPC		
layer	2D	M3D	△%	2D	M3D	△%	2D	M3D	△%
M2	24.0%	22.9%	(-4.6%)	17.3%	16.9%	(-2.3%)	18.2%	12.6%	(-31.0%)
M3	35.1%	27.2%	(-22.5%)	21.5%	16.8%	(-22.1%)	36.9%	29.2%	(-21.0%)
M4	42.0%	41.8%	(-0.5%)	16.3%	13.2%	(-21.5%)	61.5%	45.5%	(-26.1%)
M5	21.0%	21.9%	(4.3%)	4.3%	6.0%	(15.1%)	38.0%	21.8%	(-42.6%)

Figure 7: The total power saving of the M3D implementations depending on bin size selection during tier partitioning

Figure 8: Layouts of DCT M3D designs (a) after tier partitioning, (b) after cell legalization with 20μm bin size, (c) after tier partitioning, (d) after cell legalization with 2.5μm bin size

is similar across the target clock frequencies as discussed in Section 5.1, the LDPC M3D designs constantly show better power saving regardless of the target clock frequency.

Another reason for the better power saving with LDPC designs is the routing congestion in LDPC 2D design. As shown in Figure 3 (e), unlike other 2D designs, LDPC 2D designs utilizes excessive amount of routing resources, showing 38.0% top metal layer resource usage as shown in Table 3. The congestion makes wires to be routed on detours rather than to use their optimal paths, increasing total wire-length of the designs. On the other hand, owing to the reduced wire-length of nets as well as the less number of nets due to the reduced number of cells, LDPC M3D designs effectively resolve the wire congestion, showing up to 42.6% routing resource reduction. As more wires are routed on their optimal paths, the wire-length saving of the LDPC M3D design in Table 2 reaches 27.7%, which is much higher than other two benchmarks.

5.3 Impact of Bin Size

As discussed in Section 3, the M3D design flow presented in [1] utilizes square bins in order to constrain area skew of a design. Figure 7 presents the impact of bin size selection on the total power saving of M3D designs. It clearly shows that the power benefits of the M3D designs are maximum with 2.5μm bins for AES-128 and DCT designs, whereas the LDPC M3D design is not highly affected by the size of bins.

In order to analyze the impact of bin size on M3D design quality, it is important to understand the trade-off between area skew and the number of cuts (i.e., the number of MIVs). Although the bin-based partitioning helps cells to be evenly distributed throughout the design, it also restricts the partitioning algorithm to obtain
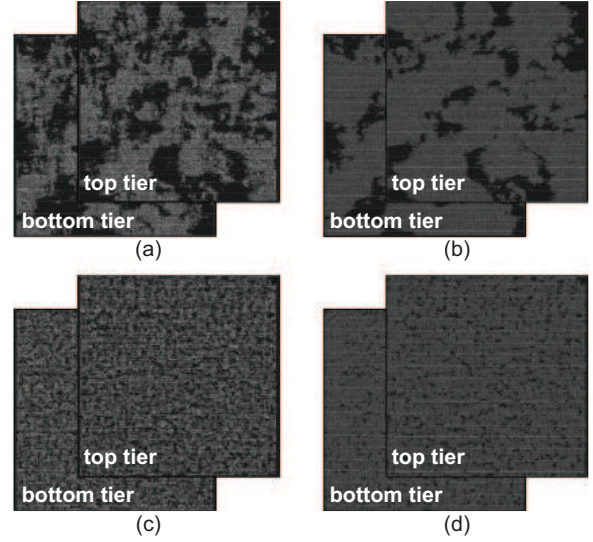
optimal solution (i.e., the minimum number of cuts) of the design producing unnecessarily large number of vertical connections.

Figure 8 presents intermediate M3D designs after tier partitioning step (i.e., before cell legalization) and after cell legalization step comparing 20μm and 2.5μm bin size selection for tier partitioning. As shown in Figure 8 (a), with large bins, neighboring cells are clustered and placed on a single tier, leaving excessive overlaps between cells even after tier partitioning. The overlaps are resolved during cell legalization, but during cell spread, as distance between cells are increased, the total wire-length of the design is increased. As shown in Figure 9, the wire-length overhead during cell legalization increases as the size of bin increases.

Although the small bin size ensures small area skew of M3D designs and minimizes the wire-length increase during cell legalization step, it makes the designs easily fall into local optimal solution, failing to minimize the number of MIVs. Figure 9 also presents the changing number of MIVs depending on bin size during tier partitioning step. With small bins, as the number of MIVs of a design is calculated by adding the minimum cuts of all bins, it is much higher than design with large bins, assigning too many neighboring cells onto different tiers and cutting short local interconnects. Although the M3D design flow assumes the z-dimension of a chip is negligibly small compared to x-y dimensions, assigning neighboring cells onto different tiers essentially increases wire-length of the interconnect since the wire needs to go through all the metal layers on the bottom tier to connect cells on different tiers. Note that this wire overhead happens after determining the location of MIVs (i.e., MIV insertion), hence, it is not reflected in Figure 9.
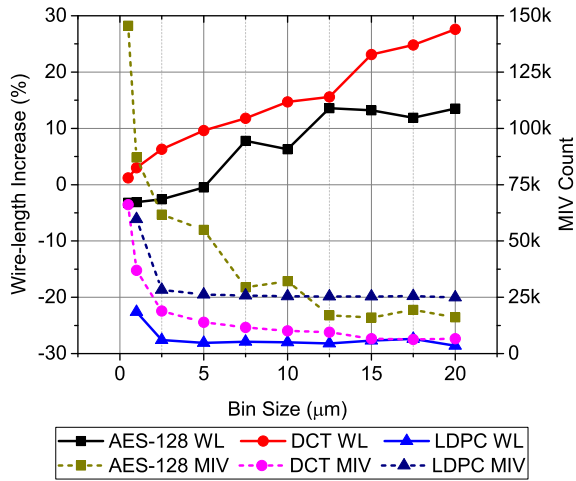
Figure 9: Wire-length increase due to cell legalization, and the number of MIVs across bin size cell selection

Compared to other two benchmarks, LDPC M3D designs tend to be not affected by the size of bins on their power saving. This is attributed to the low cell density of LDPC designs. As mentioned in Section 4, the initial target cell density of LDPC is set to much lower than other two benchmarks since the footprint of LDPC designs are determined by available routing resources, not by cell density. The low cell density prevents from producing overlap after tier partitioning because of ample empty space, even offering wire-length reduction during cell legalization as shown in Figure 9.

From the above discussion, we observe that the optimal bin size in the M3D design flow depends on the number of cells within a bin since large number of cells in a bin (i.e., large bin size) increases area skew whereas small number of cells (i.e., small bin size) increases the number of MIVs in a design degrading M3D design quality. We conclude that the optimal bin size are determined by the technology node and the cell density of a design since the number of cells depends on them. Therefore, designs at 7nm technology node with 60% cell density, $2.5\mu m$ bin size tends to offer the maximum power saving for the M3D design.

6. OBSERVATION

We summarize our observation from analysis on the PPA benefits of M3D integration with ASAP7 library.

- ASAP7 library is based on practical technology parameters, providing a way to build high quality designs at 7nm technology node. In addition, it is a key enabler to perform in-depth analysis of designs since it provides full front-end and back-end views.

- M3D designs offer power benefits by reducing wire-length and standard cell area. Standard cell area saving tends to increase as the target clock frequency of a design increases, whereas wire-length saving is similar across the target clock frequency. At 7nm FinFET technology node, due to 3D fin structure, reducing standard cell area impacts more on the power saving of M3D integration than wire-length saving.

- Wire-dominated design benefits more from M3D integration due to its high switching to internal power ratio, and wire to pin capacitance ratio. Furthermore, M3D designs achieve significant wire-length saving by resolving routing congestion.

- Bin size selection during tier partitioning impacts the quality of M3D designs significantly. It offers a trade-off between the area skew of a design and the number of MIVs, and depends on technology nodes as well as the cell density of a design. At 7nm technology node, $2.5\mu m$ bin size offers the best power benefit assuming cell density of a design is around 60%.

7. CONCLUSION

In this paper, we investigate PPA benefits of M3D ICs at 7nm technology node and provide guidelines to maximize the benfits. ASAP7 library, based on realistic technology assumptions for the 7nm node, is used for our study. We find that M3D integration provides PPA benefit over 2D designs showing up to 16.9% power saving at 7nm technology node. We also observe that more power reduction is achieved as the clock frequency increases and with wire-dominated circuits, and provide a guideline to set bin size in the M3D flow presented in [1]. Our work highlights the importance of an academic 7nm PDK and library to enable high quality VLSI and CAD research for advanced technology nodes.

8. REFERENCES

[1] S. A. Panth *et al.*, "Design and CAD Methodologies for Low Power Gate-level Monolithic 3D ICs," in *Proc. Int. Symp. on Low Power Electronics and Design*, 2014.

[2] B. W. Ku *et al.*, "Physical Design Solutions to Tackle FEOL/BEOL Degradation in Gate-level Monolithic 3D ICs," in *Proc. Int. Symp. on Low Power Electronics and Design*, 2016.

[3] K. Chang *et al.*, "Cascade2D: A design-aware partitioning approach to monolithic 3D IC with 2D commercial tools," in *Proc. IEEE Int. Conf. on Computer-Aided Design*, 2016.

[4] S. A. Panth *et al.*, "Power-performance study of block-level monolithic 3D-ICs considering inter-tier performance variations," in *Proc. ACM Design Automation Conf.*, 2014.

[5] 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.

[6] K. Chang *et al.*, "Power Benefit Study of Monolithic 3D IC at the 7nm Technology Node," in *Proc. Int. Symp. on Low Power Electronics and Design*, 2015.

[7] L. T. Clark *et al.*, "ASAP7," *Microelectronics Journal*, vol. 53, no. C, Jul. 2016.

[8] K. Chang *et al.*, "Frequency and Time Domain Analysis of Power Delivery Network for Monolithic 3D ICs," in *Proc. Int. Symp. on Low Power Electronics and Design*, 2017.

[9] N. Paydavosi *et al.*, "BSIM-SPICE Models Enable FinFET and UTB IC Designs," *IEEE Access*, vol. 1, pp. 201–215, 2013.

[10] M. Gerber *et al.*, "Next generation fine pitch Cu Pillar technology - Enabling next generation silicon nodes," in *IEEE Electronic Components and Technology Conf.*, 2011.