

3D IC Power Benefit Study Under Practical Design Considerations

Taigong Song¹, Moongon Jung², Yang Wan³, Yarui Peng¹, and Sung Kyu Lim¹

¹School of ECE, Georgia Institute of Technology, Atlanta, GA, USA

²Intel Corp., Santa Clara, CA, USA

³Google Inc., Mountainview, CA, USA

taigong.song@gatech.edu

Abstract—Despite many predictions that 3D IC is the solution for future low-power electronics, few studies describe how this can happen in real designs. In this paper, we investigate the practical design factors that affect the power consumption of 3D IC using a commercial-grade large-scale benchmark (OpenSPARC T2). In particular, we investigate the impact of power distribution network (PDN) in designer’s perspective. Our study shows that PDN significantly affects several important design metrics in addition to the total power.

I. INTRODUCTION

One of the technologies that has been highly anticipated to reduce power significantly is 3D IC. Despite the importance, however, few studies describe the impact of various design metrics in 3D ICs. [1] performs a thorough study on high-performance processors and reports the impact in various metrics such as power & thermal. In [2], authors claim that an order of magnitude higher power efficiency can be achieved using an architecture that fully utilizes the 3D interconnects. In [3], authors reported the importance of different bonding styles (face-to-face or face-to-back) and block-folding for power reduction in 3D ICs. In this paper, we present a study on a practical design factor that can significantly impact power consumption in 3D IC: power distribution network (PDN). Our study is based on OpenSPARC T2 [4], which is derived from a commercial 64-bit microprocessor. We use Synopsys 28nm PDK [5] with nine metal layers to build GDSII layouts. Using our commercial-grade RTL-to-GDSII flow [6], we design block-level 2D and 2-tier 3D layouts under various 3D technology configurations. These designs are timing-closed, power-optimized, and analyzed using sign-off quality timing, power, and noise calculation tools [6]. With these designs and CAD tools, we study how PDN impacts 3D layouts in various metrics such as area, wirelength, and power.

II. IMPACT OF POWER DISTRIBUTION NETWORK DESIGN

In this section, we describe how we implement both 2D and 3D block-level designs with PDN in detail. Then, based on our layout simulations, we show how PDN impact in core and full-chip level.

A. PDN Design Specifications

Table I describes the details of our PDN. Based on the specifications, we place our PDN at the initial design stage before placement and routing. We choose PDN width/pitch considering the alignment with routing tracks, and we do not place a fixed PDN for M1 and M2. This is because for M1, standard cells already contain VDD/VSS lines, and a fixed PDN for M2 acts as placement blockages. Figure 1 shows some metal layers with PDN in actual layouts. In our 3D

This work is supported by Intel Corporation through Semiconductor Research Corporation (ICSS Task 2293) and the Center for Integrated Smart Sensors funded by the Ministry of Science, ICT & Future Planning of the Korean Government under the Global Frontier Project (CISS-2012366054194).

TABLE I
PDN SPECIFICATIONS USED IN OUR 2D AND 3D DESIGNS. # TRACKS SHOW THE MAX NUMBER OF SIGNAL WIRES THAT CAN FIT IN BETWEEN TWO ADJACENT P/G WIRES.

	Local	Intermediate	Global		
	M3	M4 - M6	M7	M8	M9
Metal width/pitch	56/152nm	112/228nm	224/456nm		
PDN density (%)	10.5	14.9	18.0	21.4	24.9
PDN width (nm)	208	340	2048		
PDN pitch (nm)	1,976	2,280	11,400	9,576	8,208
# tracks	11	8	20	16	13

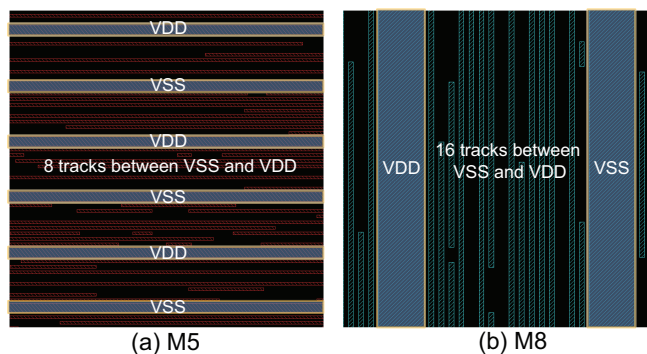


Fig. 1. PDN snapshots in (a) M5 and (b) M8

designs, we use two 3D bonding technologies shown in Table II. For TSV technology, which the dies are stacked in face-to-back fashion as in Figure 2 (a), we assume a 3um TSV in diameter. For face-to-face (F2F) technology, we assume that F2F vias connect the top-metals as in Figure 2 (b) and has diameter of 0.448um.

B. How Does PDN Affect 3D IC Placement and Routing?

In 3D ICs, the location of their 3D connections such as TSVs and F2F vias are affected by the PDN. Unless back-side RDL is used, a TSV must satisfy two constraints to be safely placed in a 3D IC: First, TSV location should not overlap with standard cells/memory macros and M1 wires in die 0. Second, the landing pad of a TSV in die 1 should not overlap with the top metal PDN. In case of face-to-face bonding, a F2F via must be placed in an empty space of the top-metal on both die 0 and die 1 where there is no PDN. Figure 3 shows how PDN affects the F2F via locations.

In terms of routing, both 2D and 3D designs suffer from the reduced routing resources due to the PDN. For example, in a 2D load/store (LSU) unit in OpenSPARC T2 core, we see heavy routing congestions and design rule violations (DRVs) increase, because PDN occupies significant routing space [see Figure 4 (a)]. Modules that require more routing resources will suffer more from PDN, resulting

TABLE II
3D IC TECHNOLOGY PARAMETERS USED IN THIS PAPER

Name	diameter (um)	height (um)	pitch (um)	R (Ω)	C (fF)
TSV	3.0	20	5.016	0.5	10.0
F2F	0.448	0.380	0.912	0.1	0.2

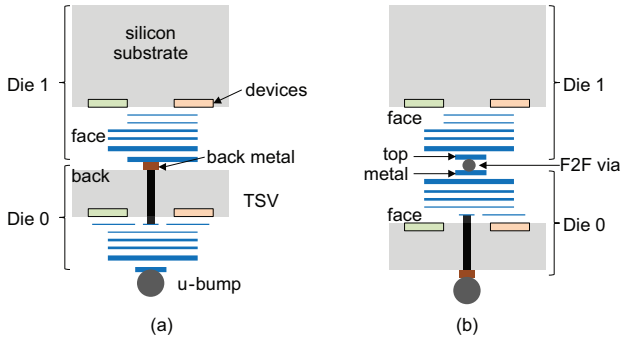


Fig. 2. Illustration of our 3D stackup: (a) face-to-back using TSVs, and (b) face-to-face (F2F) using F2F vias.

in more congestion and DRVs. The same happens in 3D as well. However, 3D IC benefits from the shorter wirelength that is made possible by using TSVs and F2F vias. Therefore, 3D IC suffers less from the lack of routing resources due to PDN. In Figure 4 (b), we see that the routing congestion problem is reduced in 3D. Note that this LSU is “folded”, where the macros and gates are partitioned into 2-tiers. This is shown to improve design quality [7]. We call this “folded LSU”.

C. T2 Single Core Results

The OpenSPARC T2 single core consists of 13 functional unit blocks (FUBs). Each FUB is synthesized, placed, routed, and optimized with Synopsys 28nm cell library using our commercial-grade RTL-to-GDSII CAD tools [6]. Then, we assemble these FUBs into the top-level module. As described earlier, some of these FUBs can be “folded”, where the FUB occupies two tiers instead of one. These are then floorplanned together with other folded and/or non-folded FUBs.

We use two V_{th} cells in our design. These are regular- V_{th} (RVT) and high- V_{th} (HVT). HVT cells consume lower cell power and leakage power than RVT, but is 30% slower. Given an initial timing constraint, we first design the individual FUBs and the top-level module (= entire core), and then perform static timing analysis (STA) using Synopsys PrimeTime to obtain a new set of timing constraint for the pins between the FUBs. With the new timing constraints, we redesign and optimize each FUB for timing closure. By going through these design optimizations for several iterations, we improve the design quality and meet the target timing of 1.5ns clock period (= 677MHz).

Table III describes the impact of PDN on our 3D design that does not contain any folded FUBs. Under the same area, we design two cores with and without PDN for comparison. We see that the T2 core with PDN uses more wirelength (+6.2%) and buffers (+10.0%) with more DRVs (+950). However, this increase is mostly from intra-block designs. In addition, we see +7.5% of total power increase, most of which comes from the net power increase (+12.4%).

Regarding the increase in net power, Table IV shows a detailed analysis. We see that the wirelength increase (+6.2%) is not that significant compared with the wire power increase (+21.2%). This

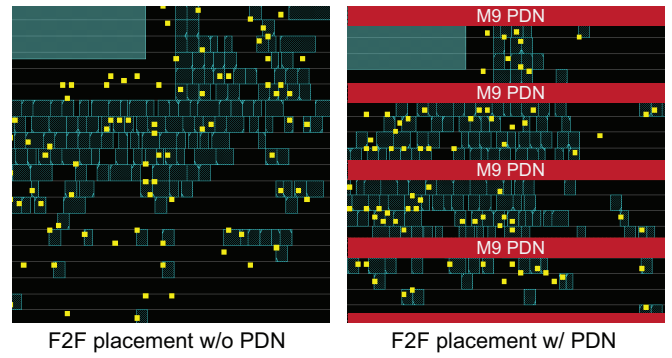
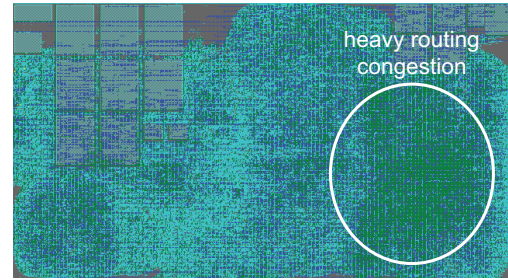
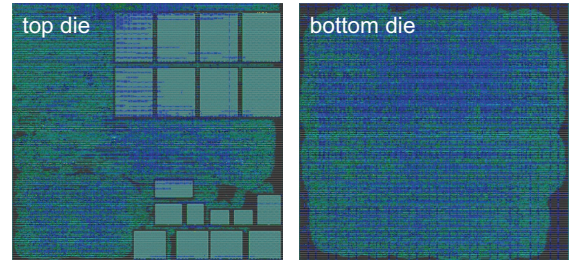


Fig. 3. F2F via (= yellow squares) location affected by the PDN. F2F vias cannot overlap with M9 PDN wires and thus not placed at their optimum locations.



(a) 2D LSU, w/ PDN, #DRV = 99



(b) 3D folded LSU, w/ PDN, #DRV = 0 for both dies

Fig. 4. Congestion map showing the impact of PDN on routing in a load/store (LSU) module. Green illustrates the area where the routing need exceeds the routing resource capacity (high congestion).

means that the routing detour in 3D core due to the PDN is not a significant factor for power increase. However, note that the wire capacitance increases by 23.1%. Pin capacitance is the capacitance of input pins of a cell, and wire capacitance is the pure capacitance that a wire sees. This indicates that high coupling capacitance forms between the PDN and signal wires, and in between signal wires as well. In short, the total power increases in designs with PDN due to the wire capacitance and wire power increase caused by the PDN, which causes the net power to increase.

D. T2 Full-chip Results

We investigate the impact of PDN on full-chip level. OpenSPARC T2, a commercialized chip with 500M transistors, consists of 53 top-level modules including eight cores (SPC), one cache crossbar (CCX), eight L2 cache data banks, (L2D) eight L2 cache tags (L2T), and eight L2 cache miss buffers (L2B). Our target clock frequency for CPU is 500MHz and 250MHz for I/O. Figure 5 shows the GDSII layouts of our 3D designs of full-chip T2. We design three 3D IC layouts: face-to-back (= TSV) 3D with no folded FUBs [see Figure 5 (a)],

TABLE III
PDN IMPACT ON 3D T2 CORE W/O FOLDING (TSV)

T2 Core	No PDN	PDN	Δ
Footprint (mm ²)	1.50	1.50	-
Wirelength (m)	19.04	20.23	+6.2%
# Cells	421.5k	434.4k	+3.1%
# HVT cells	409.1k (97.1%)	413.6k (95.2%)	-
# Buffers	128.2k	141.0k	+10.0%
# DRV	587	950	-
Total power (mW)	264.6	284.4	+7.5%
Cell (mW)	62.9 (23.8%)	64.3 (22.6%)	+2.2%
Net (mW)	129.9 (49.1%)	146.0 (51.3%)	+12.4%
Leakage (mW)	71.8 (27.1%)	74.1 (26.1%)	+3.2%

TABLE IV

PDN IMPACT ON NET POWER. PDN INCREASES WIRE CAPACITANCE AND WIRE POWER SIGNIFICANTLY.

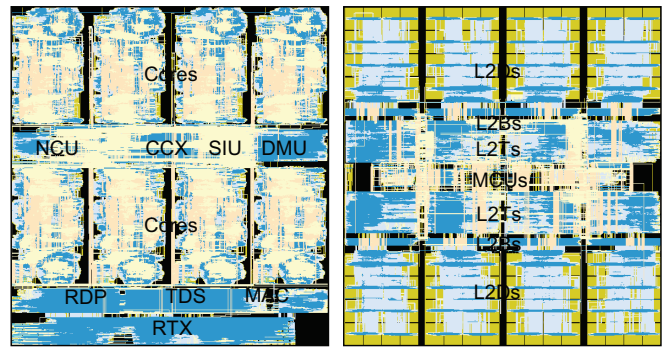
		No PDN	PDN	Δ
Wirelength (m)		19.04	20.23	+6.2%
Cap (pF)	Pin cap	1342.1	1359.3	+1.3%
	Wire cap	1723.2	2120.4	+23.1%
	Total net cap	3065.3	3479.7	+13.5%
Net power (mW)	Pin power	56.2	56.7	+0.9%
	Wire power	73.7	89.3	+21.2%
	Total net power	129.9	146.0	+12.4%

face-to-back (= TSV) 3D with folded FUBs, and face-to-face (= F2F) 3D with folded FUBs [see Figure 5 (b)]. In the non-folded design, all cores are located in one die, and all L2 caches are placed in another die. Five blocks are folded for maximum power benefit (SPC, L2B, L2T, CCX, and RTX) in the folded designs.

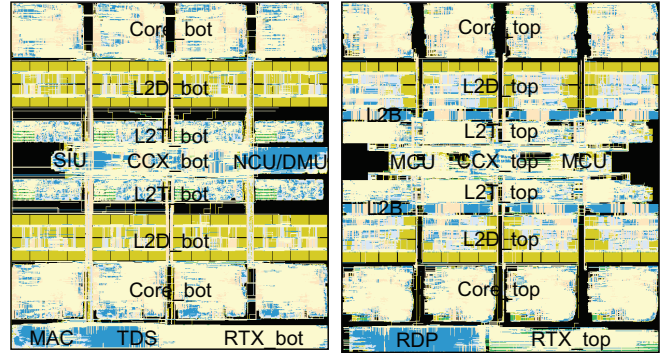
Figure 6 shows our full-chip results. We emphasize that 3D designs with folding lead to benefits over 2D from PDN impact. For example, block-folded 3D with PDN showed improved power reduction than without PDN (-1.5% more in folded F2F and -0.9% more in folded TSV). However, 3D design without folding loses its power reduction when we insert PDN (no PDN: -11.1% > PDN: -10.2%). This is because the power increase ratio due to the PDN between designs are different. For example, 2D shows +7.1% total power increase when PDN is inserted, but 3D without folding shows +8.4%, which is more than 2D. 3D designs without folding do not obtain any intra-block-level benefit from folding. Therefore, 3D may lose its power benefit depending on the intra FUBs that are designed in 2D. However, block-folded TSV shows +6.6% and F2F shows +5.4% total power increase by PDN. Thus, in order to benefit from PDN impact, we must optimize both inter and intra-level designs to minimize wirelength and congestion. In full-chip designs, large FUBs that require significant routing resources can be optimized using the folding technique resulting in shorter total wirelength and fewer buffers, which reduces the impact from PDN. In summary, block-folding for shorter wirelength leads to smaller power increase after PDN insertion. Managing FUBs to have sufficient routing resources is critical to reduce the impact from PDN for low power design.

III. CONCLUSIONS

In this paper, we demonstrated the impact of an important design factor: PDN. We found that PDN impacts the 3D IC design in various metrics such as TSV/F2F via placement and metal routing. Designing (= folding) modules into 3D not only helps to reduce wirelength for better design, but also relieves the impact from having less routing resources due to PDN. Our full-chip benchmark design built in 3D with folding showed more power reduction with PDN comparing to the 2D counterpart due to the less wirelength used in the total design, which led to less influence caused by the PDN.



(a) 3D w/o folded blocks (#TSV: 3,263)



(b) 3D w/ folded blocks (#F2F: 101,555)

Fig. 5. GDSII layouts of 2-tier full-chip T2 (left: bottom die, right: top die)

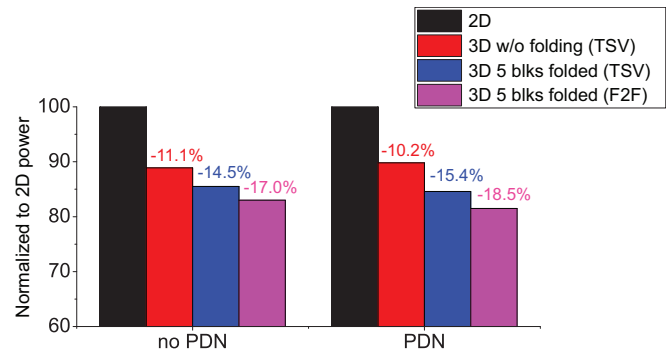


Fig. 6. PDN impact on power: full-chip T2

REFERENCES

- [1] P. Emma *et al.*, "3D Stacking of High-Performance Processors," in *High-Performance Computer Architecture, The Twentieth International Symposium on*, Feb 2014.
- [2] T. Sekiguchi *et al.*, "1-Tbyte/s 1-Gbit DRAM Architecture Using 3-D Interconnect for High-Throughput Computing," *Solid-State Circuits, IEEE Journal of*, vol. 46, no. 4, pp. 828–837, April 2011.
- [3] M. Jung *et al.*, "On enhancing power benefits in 3D ICs: Block folding and bonding styles perspective," in *Design Automation Conference (DAC), 2014 51st ACM/EDAC/IEEE*, June 2014, pp. 1–6.
- [4] Oracle, OpenSPARC T2. [Online]. Available: <http://www.oracle.com>.
- [5] Synopsys, 32/28nm Generic Library. [Online]. Available: <http://www.synopsys.com>.
- [6] D. H. Kim *et al.*, "3D-MAPS: 3D Massively Parallel Processor with Stacked Memory," in *Proc. Int. Solid-State Circuits Conf.*, 2012.
- [7] M. Jung *et al.*, "How to reduce power in 3D IC designs: A case study with OpenSPARC T2 core," in *Proc. IEEE Custom Integrated Circuits Conference*, 2013, pp. 1–4.

