

RTL-to-GDS Design Tools for Monolithic 3D ICs

"Invited Talk"

Jinwoo Kim¹, Gauthaman Murali¹, Pruek Vanna-iampikul¹, Edward Lee¹, Daehyun Kim¹,
Arjun Chaudhuri², Sanmitra Banerjee², Krishnendu Chakrabarty², Saibal Mukhopadhyay¹,
and Sung Kyu Lim¹

¹School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta GA, USA

²Department of Electrical and Computer Engineering, Duke University, Durham NC, USA

jinwookim@gatech.edu;limsk@ece.gatech.edu

ABSTRACT

In this paper, we propose RTL-to-GDS design flow for monolithic 3D ICs (M3D) built with carbon nanotube field-effect transistors and resistive memory. Our tool flow is based on commercial 2D tools and smart ways to extend them to conduct M3D design and simulation. We provide a post-route optimization flow, which exploits the full potential of the underlying M3D process design kit (PDK) for power, performance and area (PPA) optimization. We also conduct IR-drop and thermal analysis on M3D designs to improve the reliability. To enhance the testability of our M3D designs, we develop design-for-test (DFT) methodologies and integrate a low-overhead built-in self-test module into our design for testing inter-layer vias (ILVs) as well as logic circuitries in the individual tiers. Our benchmark design is RISC-V Rocketcore, which is an open source processor. Our experiments show 8.1% of power, 19.6% of wirelength and 55.7% of area savings with M3D designs at iso-performance compared to its 2D counterpart. In addition, our IR-drop and thermal analyses indicate acceptable power and thermal integrity in our M3D design.

KEYWORDS

Monolithic 3D IC, Physical design (EDA), CNFET, Design-for-test, ILV dual-BIST

1 INTRODUCTION

Monolithic 3D IC (M3D) is introduced to surmount the difficulties of a traditional 2D IC due to the process node scaling and the high design complexity, and outperform 2D IC in terms of power, performance and area (PPA). Moreover, M3D maximizes the benefit of 3D IC achieving higher device density when compared to TSV-based 3D stacking [1]. In this technology, transistors are processed tier-by-tier on the same wafer. M3D offers massive inter-die connections using nanoscale inter-layer vias (ILVs) which are not possible with TSV [3]. Therefore, M3D integration results in significantly reduced area and higher performance when compared to TSV-based 3D die stacking. However, ILV testing is fundamental to effective defect screening and quality assurance due to the high integration density

in M3D ICs [10]. ILVs can be tested together with logic and memory, and we need design-for-test (DFT) method for ILVs for defect isolation and yield learning.

Carbon nanotubes (CNTs) has been proposed as a new alternative over existing silicon while silicon-based semiconductor technology improves its performance and energy efficiency through scaling down. Carbon nanotube field-effect transistor (CNFET), which is fabricated with aligned CNTs in parallel, becomes a promising device beyond silicon field-effect transistor (SiFET) due to its performance [16]. Moreover, the recent research indicates that CNFETs are applicable to a complex digital VLSI design as beyond-silicon technology [8].

In this paper, we claim the following contributions: (1) We present RTL-to-GDS design flow for monolithic 3D ICs (M3Ds) using 3D CNFET PDK; (2) We generate a CNFET-based M3D benchmark design (RISC-V Rocketcore) using our M3D design flow and provide IR-drop and thermal analysis framework for design validation; (3) We perform PPA analysis to observe the benefits of CNFET-based M3D designs; (4) We discuss a design-for-test (DFT) methodology that detects defects of inter-layer vias (ILVs) which connect two adjacent tiers and present Rocketcore design with DFT circuitry and ReRAM to show the feasibility of integration.

2 MONOLITHIC 3D IC DESIGN TOOL

2.1 Overall M3D Design Flow

Our proposed monolithic 3D IC (M3D) design flow based on Shrunk-2D flow [12] is shown in Figure 1(a). Shrunk-2D flow is the first commercial-quality RTL-to-GDS flow for M3D designs, which has inspired the following works [4, 11]. This approach introduces an initial pseudo-3D design named *shrunk-2D* which has elements scaled down to $1/\sqrt{n}$ of the original dimensions for n -tier design. We prepare the scaled library exchange format (LEF) files in which standard cells and metal dimensions are scaled down to $1/\sqrt{2}$ for 2-tier pseudo-3D design. Since our flow handles 3D PDK, we generate these files only with top-tier elements.

Our M3D design flow starts from a synthesized 2D netlist and information of macro blocks if the design contains pre-defined blocks such as memory modules. In the floorplanning stage, we pre-place macro blocks to corresponding tier. We set a placement blockage on the area where each macro block is placed to enable the proper standard cell placement: a partial blockage for 2D macro block and a full blockage for 3D macro block. With the generated floorplan and shrunk technology files, we generate *shrunk-2D* place-and-route

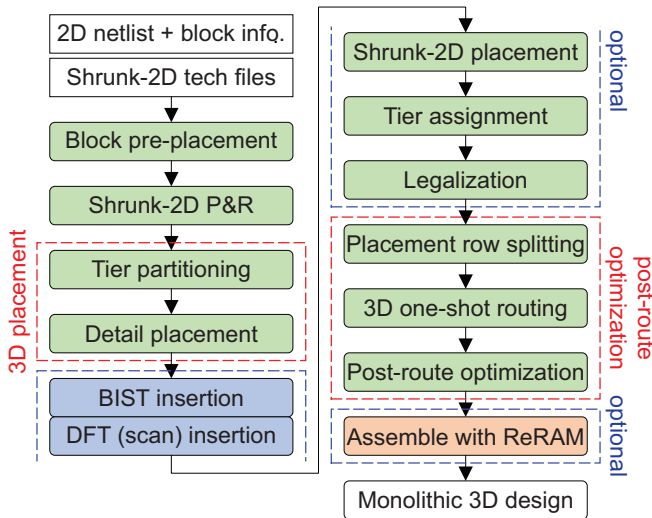
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

ICCAD '20, November 2–5, 2020, Virtual Event, USA

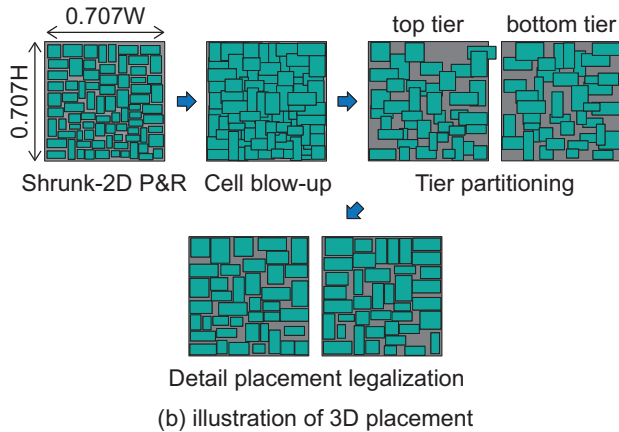
© 2020 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8026-3/20/11.

<https://doi.org/10.1145/3400302.3415780>



(a) overall M3D design flow



(b) illustration of 3D placement

Figure 1: Monolithic 3D design flow using 3D PDK. The optional flows are marked as blue.

(P&R) result using a commercial 2D P&R tool. In generated *shrunk-2D* result, all cells are shrunk into half size and placed in 50% of 2D footprint.

With *shrunk-2D* design, we perform 3D placement with in-house tools as illustrated in Figure 1(b). Shrunk standard cells in the design are first blown up to their original size and these cells are partitioned into two different tiers with bin-based Fiduccia-Mattheyses (FM) min-cut partitioning [7]. In tier partitioning stage, we divide the footprint into square bins and perform FM min-cut algorithm on each bin. Remaining cell overlaps due to blown-up cells are legalized in the tier-by-tier detail placement stage, which uses refinement command of commercial 2D P&R tool. Finally, we import both tier designs into 2D tool, and generate placement information including both tiers in a single design exchange format (DEF) file.

We also design built-in self-test (BIST) modules to validate inter-layer vias (ILVs) in our M3D design. The detailed discussion of design-for-test (DFT) methodology is given in Section 5. During

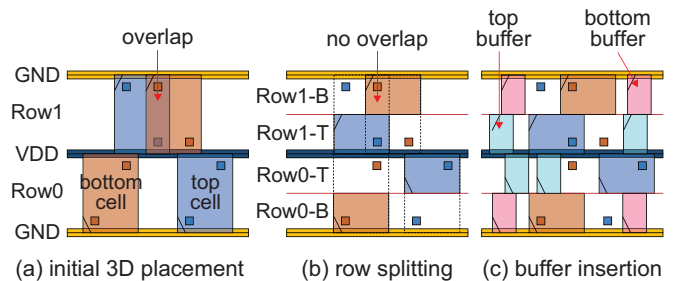


Figure 2: Row splitting scheme in post-route optimization flow.

the post-route optimization stage which details are in Section 2.2, we perform the routing of M3D design and design optimization to meet the timing constraints.

2.2 3D Post-route Optimization

In 3D post-route optimization stage, we utilize the row splitting concept in Compact-2D flow [11] because none of 3D-routing-aware optimization is available in Shrunk-2D flow. To solve the cell overlap issue during buffer insertion, we split a placement row into the top and bottom rows. Moreover, we halve the heights of standard cells in 3D macro LEF files to fit those cells into split placement rows.

Figure 2 illustrates the post-route optimization stage in our flow. First, we split each row into two horizontally split rows. In Row0, the bottom half is reserved for the bottom tier, and the top half for the top tier. The order in Row0 is reversed in Row1 to align standard cells to the power/ground rails. As a result, the placement overlap is fully legalized while accommodating every cell in the design on the final footprint. With split placement rows, we perform timing optimization using commercial 2D P&R tool, restore split placement rows back to the original and generate the final M3D design.

3 CARBON NANOTUBE M3D DESIGN

3.1 Carbon Nanotube Transistor PDK

In this paper, we use a custom carbon nanotube field-effect transistor (CNFET) process design kit (PDK) for our implementations as shown in Figure 3. This PDK has been derived from [14] and includes 2 layers of CNFETs, 2 layers of ReRAM, and 11 metal routing layers. More recently, commercial scale fabrication technology and associated PDK have been demonstrated considering multiple layers of transistors and ReRAM devices [15].

Unlike traditional silicon FET, CNFET has the bottom-gate structure which gate is placed at the bottom, and design rules are declared to bypass CNFETs for effective routing. Moreover, *n*-type and *p*-type CNFETs are distinguished by different doping layers for source and drain of CNFET. Four metal layers in ReRAM tiers are used for the routing purpose through ReRAM bypass connections for the design without ReRAM modules.

3.2 3D Random Access Memory

We implement 3D SRAM using the stackable subarrays generated from our memory compiler tool. The conceptual view of our

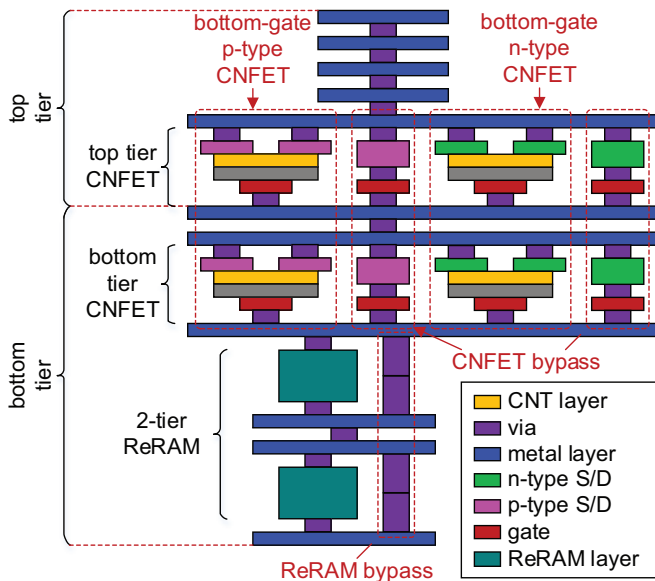
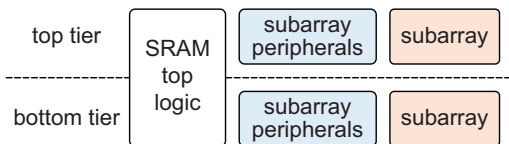
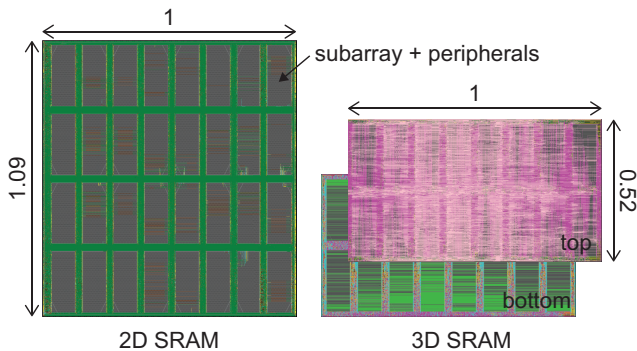


Figure 3: Vertical stack-up of CNFET 3D PDK.



(a) conceptual vertical view of CNFET 3D SRAM



(b) GDS layouts of CNFET SRAMs (16kB): 2D vs. 3D

Figure 4: Conceptual view and GDS layout of CNFET 3D SRAM.

3D SRAM design is shown in Figure 4(a). Subarrays in each tier are designed to use only the metal layers right below and above each CNFET layer and not to share metal layers with each other. We generate 3D SRAM designs with different capacities using our monolithic 3D design flow that treats subarray modules as macro blocks. The size of subarray in each design is 64×64 and the design results of 3D SRAMs are shown in Figure 4(b) and Table 1. In case of 512×256 SRAM design, 3D design shows 48.2% of area, 25.0% of total power and 40.9% of WNS savings compared to 2D design.

Table 1: Normalized PPA comparison of 2D and 3D SRAM designs.

	64×128 (1kB)		512×256 (16kB)	
	2D	3D	2D	3D
Area	1×1	0.57×1.04	1×1	1.01×0.52
Std. cell count	1	1.14	1	0.81
Total wirelength	1	0.65	1	0.79
Cell internal power	0.10	0.13	0.34	0.35
Net switching power	0.08	0.04	0.39	0.13
Leakage power	0.002	0.004	0.01	0.01
Subarray power	0.81	0.81	0.26	0.26
Total power	1	0.98	1	0.75
Worst negative slack	1	0.08	1	0.59

3.3 Carbon Nanotube Rocketcore M3D Design

We choose Rocketcore [2] with a single core as our benchmark architecture, and compare M3D and 2D IC designs in terms of power, performance and area (PPA) metrics to demonstrate the superiority of our M3D tool. As our Rocketcore design only has SRAM modules, we use four metal layers in ReRAM layers as routing resources. In the design stage, we synthesize RTL netlist by using Synopsys Design Compiler and perform P&R with Cadence Innovus. The final analysis is performed by Cadence Tempus (timing and power), Cadence Voltus (IR-drop) and Ansys Fluent solver (thermal).

Figure 5 shows GDS layouts of Rocketcore designs in both 2D and M3D. Both design include 1kB SRAM as tag array and 16kB SRAM as data array, both of which are 3D SRAMs. In M3D design, all standard cells are not overlapped in macro cells because 3D SRAMs fully occupy both top and bottom CNFET tiers.

Table 2 summarizes the normalized comparison of 2D and M3D designs. M3D design has 55.7% smaller footprint and 19.6% of wirelength reduction when compared to 2D design. The shorter wirelength also results in the smaller net switching power by 13.1%, which in turn reduces the total power. Moreover, due to the use of 3D SRAM which consumes lower power than 2D SRAM, the overall power has reduced by 8.2% in M3D design. However, the worst negative slack in M3D design has increased by 107.0%, which comes from the characteristics of standard cells in the bottom CNFET tier. CNFETs in top and bottom tier show different timing characteristics because each top and bottom tier has different layer-stacking: BEOL-CNFET-BEOL in top tier and ReRAM-CNFET-BEOL in bottom tier. This characteristic difference should be further minimized to maximize the PPA savings in M3D design.

3.4 The Effect of Post-route Optimization

As we discussed in Section 2.2, we perform post-route optimization to improve the timing performance of M3D design. However, the overheads exist in terms of switching power, wirelength and cell count. We implement the M3D designs with and without post-route optimization stage to observe the effect of it as shown in Table 3.

In the optimized design, the cell count has increased by 1.23% compared to non-optimized design due to the insertion of additional buffers. The total wirelength has also increased by 2.55% which leads

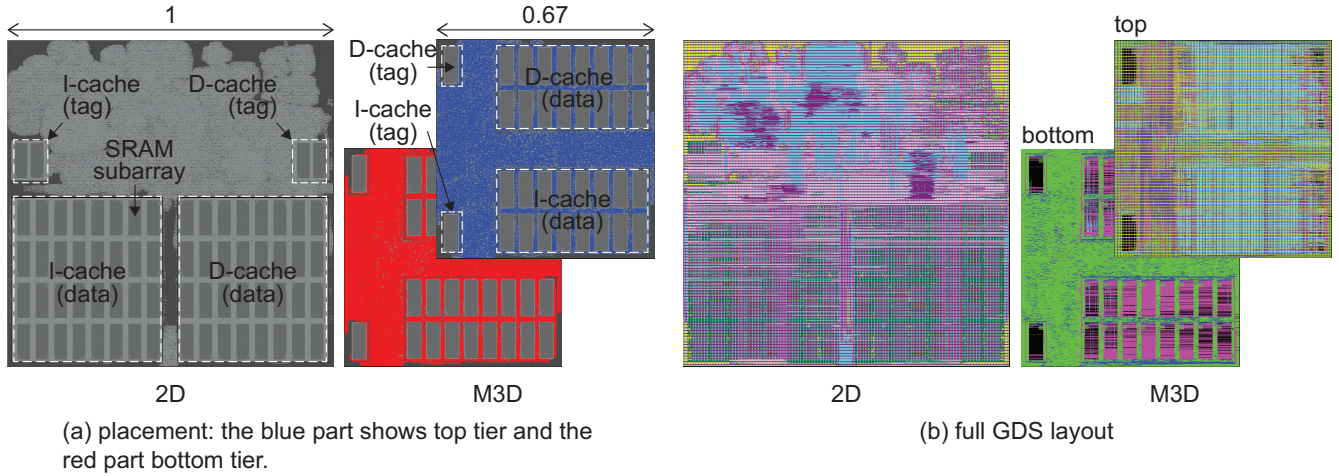


Figure 5: GDS layouts of carbon nanotube RocketCore: 2D vs. M3D.

Table 2: Normalized iso-performance comparison of 2D and Monolithic 3D (M3D) designs.

	normalized 2D	M3D percentage gain over 2D (%)
Area	1×1	-55.7
Std. cell count	1	-11.0
Total wirelength	1	-19.6
Cell internal power	0.39	-1.6
Net switching power	0.15	-13.1
Leakage power	0.06	-2.2
SRAM power	0.40	-13.7
Total power	1	-8.2
Worst negative slack	1	+107.0

Table 3: Normalized iso-performance comparison of with and without post-route optimization.

	non-optimized	percentage gain of optimized (%)
Std. cell count	1	+1.23
Total wirelength	1	+2.55
Cell internal power	0.39	-2.56
Net switching power	0.15	+4.09
Leakage power	0.06	+3.39
SRAM power	0.40	-
Total power	1	-0.36
Worst negative slack	1	-78.39

to 4.09% increase in the net switching power. However, the worst negative slack (WNS) has significantly improved by 78.39% without the total power overhead. This comparison shows the promising advantages of post-route optimization flow to resolve timing issue without extra overhead.

4 IR-DROP AND THERMAL ANALYSIS OF CNFET-BASED M3D DESIGN

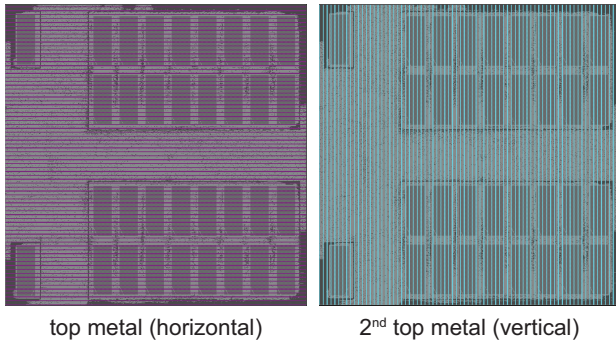
4.1 IR-drop Analysis

We conduct IR-drop analysis to validate the robustness of the power delivery network (PDN) in M3D design. Our PDN design flow takes the design constraints of PDN as initial inputs, such as the size of power/ground (P/G) bump array and the physical dimensions of PDN grids. We first uniformly place P/G bump array and calculate the number of P/G stripes which lie within each P/G bump pair. Then, our tool re-calculates the pitch of P/G grid and places PDN grids in designated layers. Finally, we connect P/G bumps to PDN grids. The top two metal layers are used for our PDN design as shown in Figure 6(a). P/G stripes are overlapped with the existing memory blocks because our memory blocks do not use the topmost metal layers for their own P&R.

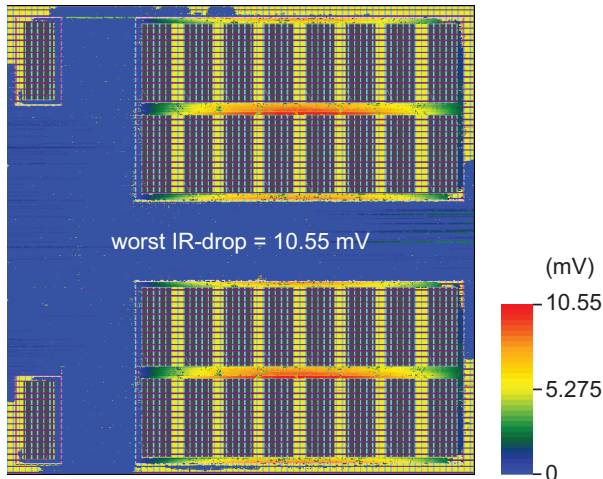
For IR-drop analysis, we generate the power grid library using 3D parasitic technology file and perform the analysis using Cadence Voltus. Figure 6(b) shows IR-drop map of M3D Rocketcore design. PDN is placed on the top two metal layers and the standard cells in both top and bottom tier are fed by the PDN. Therefore, M3D design has longer worst IR-drop path than 2D design which only uses standard cells in one (top) tier. Nevertheless, the worst IR-drop is $10.55mV$, which is acceptable at 0.16% of supply voltage.

4.2 Thermal Analysis

We perform thermal analysis of RocketCore design by modeling the entire design as a 3D thermal cube. Figure 7(a) shows an overview of our thermal analysis. We associate each layer (metal, semiconductor, dielectric) in the design with its thermal conductivity value. The power analysis of RocketCore generates a power density map, which is used as an input to the thermal analysis flow. Power density multiplied by thermal conductivity gives the temperature gradient of each layer across the die. Therefore, we divide the entire RocketCore design into multiple 3D cubes, generate a common power density and thermal conductivity for each cube and solve for temperature distribution of each cube using ANSYS Fluent solver.



(a) power delivery network (PDN) design



(b) IR-drop map of M3D design

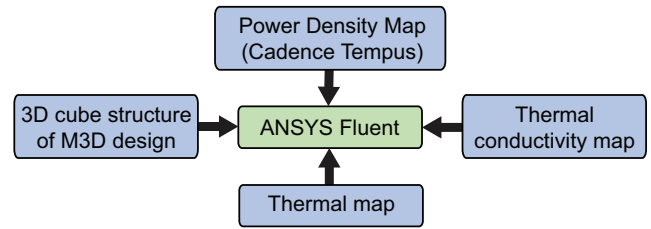
Figure 6: IR-drop map of M3D carbon nanotube RocketCore design.

On consolidating the temperature distribution across all the 3D cubes, we obtain the thermal map of the overall RocketCore design.

Figure 7(b) shows the power density map and the thermal map of our benchmark design. The power density across the die is similar as seen from the predominant yellow regions in the power density map, and is almost equal to $2,900W/m^2$. Assuming a uniform metal/dielectric layer distribution across all the thermal cubes, the thermal properties of thermal cubes do not vary much. Therefore, on an average the temperature variation across the entire die of RocketCore is very small, which is $0.02K$ approximately. Assuming the room temperature to be $300.00K$, the temperature of die shoots to a maximum of $300.02K$ with skywater CNFET PDK as shown in Figure 7(b).

5 BUILT-IN SELF-TEST SOLUTION FOR INTER-LAYER VIAS

Well-known fault models for an ILV are shorts, opens, and stuck-at faults (SAFs). Particle contamination and metal diffusion lead



(a) overview of the 3D thermal analysis

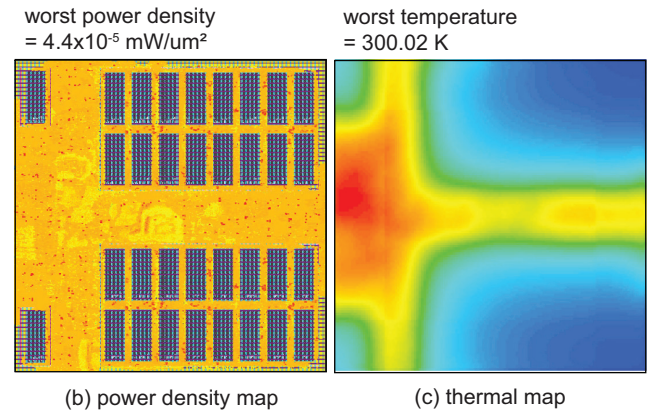


Figure 7: Overview and results of the 3D thermal analysis.

to shorts. When an ILV fails to land on a contact pad, an open is created causing the ILV resistance to increase significantly.

To test ILVs, methods such as [13] deploy one scan flop per ILV, resulting in large area overhead and test time [9]. Interconnect test methods based on ATPG [6] are less effective for testing ILVs as I/O pins are present only on one tier in an M3D IC. As a result, the activated ILV faults have to be propagated through multiple ILVs and tiers, thereby increasing the risk of ILV-fault masking due to faults in the logic gates and hindering fault detection. In [5], a BIST framework has been proposed that can effectively detect single and multiple SAFs, shorts, and opens in ILVs. The proposed BIST methodology achieves nearly 100% fault coverage (both single and multiple faults) of the ILVs with only two test patterns. In this work, for the first time, the proposed ILV-BIST framework is fully automated and integrated with the design flow for CNFET-based M3D ICs.

5.1 XOR-BIST Architecture for Fault Detection

The BIST architecture to test for faults in ILVs is illustrated in Figure 8. On the output side of the ILVs, 2-input XOR gates are inserted between neighboring ILVs. For a set of N ILVs placed in an 1D array-like manner where every ILV has at most two nearest neighbors, $(N - 1)$ XOR gates are inserted. The XOR outputs are fed as inputs to a space compactor which is an optimally balanced AND tree with $(N - 1)$ inputs and a 1-bit output signature Y_1 . By observing Y_1 , it can be determined whether a fault is present in the ILVs under consideration. Test patterns are fed to the inputs of the ILVs from an input source V_{in} ; V_{in} feeds an inverter chain that generates complementary signals to adjacent ILVs in the test mode.

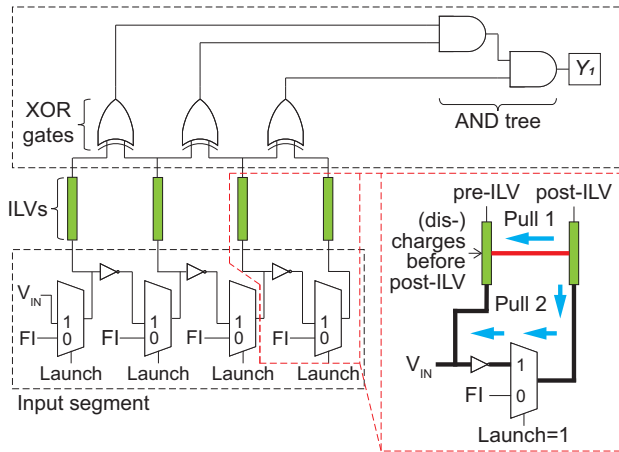


Figure 8: Illustration of the XOR-BIST architecture.

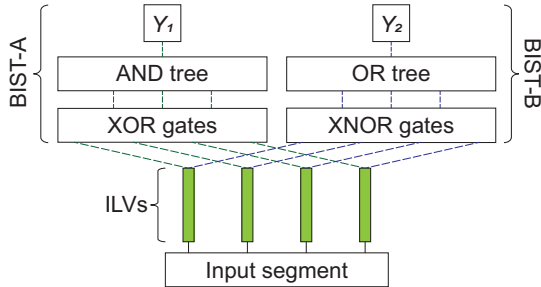


Figure 9: Illustration of the dual-BIST architecture.

A 2:1 multiplexer is present at the input of every ILV to switch between test mode and mission mode (functional input— FI) based on the $Launch$ signal.

The ILVs are tested in two clock cycles by switching V_{in} . The test patterns to the ILVs are “010...” ($V_{in} = 0$) and “101...” ($V_{in} = 1$) in the first and second cycles, respectively. It can be proven that a group of ILVs does not contain a hard fault if and only if Y_1 is 1 in both clock cycles. Aliasing occurs only when all ILVs are alternately stuck at 0 and 1, leading to masking of the ILV faults. However, the likelihood of the occurrence of such a scenario is $\frac{2}{3^n}$, where n is the number of ILVs under test.

The inverter chain-based method of driving the ILVs in the test mode leads to a deterministic hard-short behavior. If a short is present between two ILVs, the ILV appearing first (pre-ILV) in the path of the incoming test signal from V_{in} , via the inverter chain, will drive the other ILV (post-ILV); this is illustrated in the inset of Figure 8. It is because the short provides a path of lower resistance (“pull 1”) compared to the path through the multiplexer and inverter (“pull 2”).

5.2 Dual-BIST Architecture

The BIST design described in Section 5.1 may be affected by SAFs, which in turn can potentially mask ILV fault(s). To reduce the likelihood of masking, a second propagation path is added from the ILV outputs to a 1-bit signature Y_2 . The topology of this path to Y_2

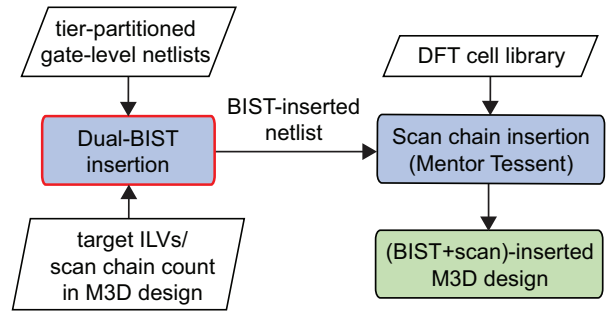


Figure 10: Tool flow of ILV dual-BIST insertion. The stage using our custom Python script are marked as red.

(BIST-B) is identical to that of the path from the ILV outputs to Y_1 (BIST-A). The XOR and AND gates in BIST-A are substituted with the corresponding logical dual gates (XNOR and OR, respectively) in BIST-B, as shown in Fig. 9. The ILVs under test, along with the “dual-BIST” engine, are considered to be fault-free if and only if $Y_1 = 1$ and $Y_2 = 0$ for both test patterns. With the “dual-BIST” architecture, it can be proven that a single fault in the dual-BIST engine cannot mask ILV fault(s). Furthermore, the probability of masking due to multiple faults in the dual-BIST engine is negligible.

5.3 Automation of Dual-BIST Insertion

We generate BIST-inserted M3D design using our fully automated BIST-insertion flow shown in Figure 10. Our custom in-house tool, implemented using Python, takes as inputs: (1) tier-partitioned gate-level netlists of the M3D tiers, (2) target ILVs to be BIST-inserted, and (3) target number of scan chains to insert in the full M3D design. Note that one dual-BIST engine tests a group of eight ILVs and accordingly, using (2), the tool determines the number of dual-BIST engines to be inserted in the M3D design. Using (1) and (2), dual-BIST insertion is carried out in a non-intrusive manner; Figure 11 illustrates the dual-BIST insertion method. Using (3), the tool generates a dofile for scan-chain insertion that follows the BIST-insertion stage; see Figure 10. Using the generated dofile, scan chains are inserted in the BIST-inserted M3D design using Mentor Tessent.

5.4 Overhead for Dual-BIST

We evaluate the impact of dual-BIST on the PPA metrics of Rocketcore M3D benchmark. For clear comparison of the impact of BIST, we apply the same design flow of Section 2 with block-level tier-partitioning. Following the flow described in Section 5.3, the dual-BIST is then inserted tier-wise to generate a BIST-inserted M3D design as shown in Figure 12.

Table 4 compares the power consumption and area overheads of the BIST-inserted Rocketcore design (BI) and the non BIST-inserted design (N-BI). The target number of ILVs to be tested is 64. Therefore, eight dual-BIST engines are inserted in N-BI to generate BI, with each engine dedicated to testing eight ILVs. Compared to the N-BI, the impact of dual-BIST on the circuit’s PPA is minimal and within acceptable limits.

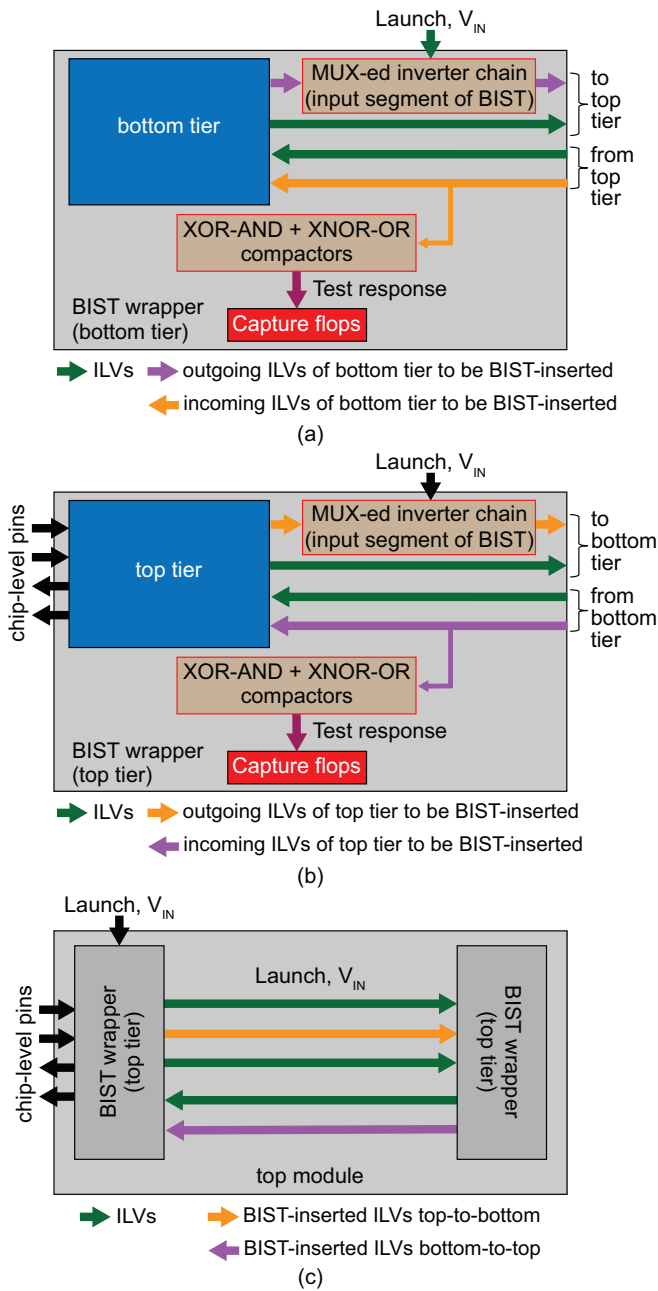


Figure 11: Non-intrusive dual-BIST insertion in a partitioned two-tier M3D design with I/O pins present on the top tier: (a) Generation of BIST-inserted bottom tier: insertion of BIST as a wrapper around the bottom tier’s netlist, (b) Generation of BIST-inserted top tier: insertion of BIST as a wrapper around the top tier’s netlist, (c) Generation of the top module: top-level instantiation of the BIST-inserted top and bottom tiers.

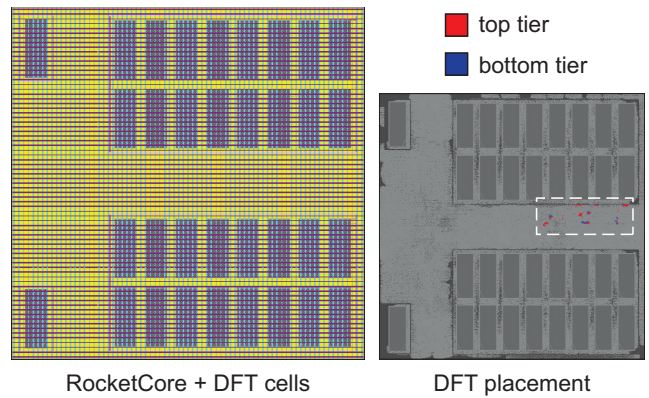


Figure 12: GDS layout of M3D Rocketcore design integrated with ILV-BIST.

Table 4: Impact of dual-BIST on PPA metrics of M3D Rocketcore design.

	non-BISTed	BISTed percentage gain over non-BISTed(%)
Std. cell count	1	+0.27
Total wirelength	1	+9.69
Total power	1	+0.62

6 CONCLUSION

In this paper, we presented RTL-to-GDS tool flow using 3D CNFET PDK which includes 2 layers of CNFETs, 2 layers of ReRAM and 11 metal routing layers. Our tool flow also includes a low-cost BIST solution for ILVs, which are all necessary for high-quality monolithic 3D IC (M3D) designs. Moreover, we provided the post-route optimization method to improve the timing performance of M3D designs, and conducted IR-drop and thermal analysis flow for the design validation. Using our M3D design flow, we generated CNFET-based Rocketcore design to demonstrate the possibility of its M3D integration. Moreover, we integrated the dual-BIST module into M3D Rocketcore design to observe the impact of BIST on M3D design. Through our experiments, our M3D design shows 8.2%, 19.6% and 55.7% savings in terms of power, wirelength and area respectively when compared to 2D design.

ACKNOWLEDGEMENTS

This research is funded by the DARPA ERI 3DSoc Program under Award HR001118C0096.

REFERENCES

- [1] K. Arabi, K. Samadi, and Y. Du. 3D VLSI: A Scalable Integration Beyond 2D. In *Proceedings of the 2015 Symposium on International Symposium on Physical Design*, ISPD '15, page 1–7, New York, NY, USA, 2015. Association for Computing Machinery.
- [2] K. Asanović et al. The Rocket Chip Generator. Technical Report UCB/EECS-2016-17, EECS Department, University of California, Berkeley, Apr 2016.
- [3] P. Batude et al. 3-D Sequential Integration: A Key Enabling Technology for Heterogeneous Co-Integration of New Function With CMOS. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 2(4):714–722, Dec 2012.
- [4] K. Chang et al. Cascade2D: A Design-aware Partitioning Approach to Monolithic 3D IC with 2D Commercial Tools. In *2016 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 1–8, Nov 2016.
- [5] A. Chaudhuri et al. Built-in Self-Test for Inter-Layer Vias in Monolithic 3D ICs. In *2019 IEEE European Test Symposium (ETS)*, pages 1–6, 2019.
- [6] D. Erb et al. Multi-cycle Circuit Parameter Independent ATPG for interconnect open defects. In *2015 IEEE 33rd VLSI Test Symposium (VTS)*, pages 1–6, 2015.
- [7] C. M. Fiduccia and R. M. Mattheyses. A Linear-Time Heuristic for Improving Network Partitions. In *19th Design Automation Conference*, pages 175–181, June 1982.
- [8] G. Hills et al. Modern Microprocessor Built from Complementary Carbon Nanotube Transistors. *Nature*, 572:595–602, August 2019.
- [9] A. Jutman. Shift Register based TPG for At-speed Interconnect BIST. In *2004 24th International Conference on Microelectronics (IEEE Cat. No.04TH8716)*, volume 2, pages 751–754 vol.2, 2004.
- [10] A. Koneru, S. Kannan, and K. Chakrabarty. A Design-for-Test Solution Based on Dedicated Test Layers and Test Scheduling for Monolithic 3-D Integrated Circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 38(10):1942–1955, 2019.
- [11] B. W. Ku, K. Chang, and S. K. Lim. Compact-2D: A Physical Design Methodology to Build Commercial-Quality Face-to-Face-Bonded 3D ICs. In *Proceedings of the 2018 International Symposium on Physical Design*, ISPD '18, page 90–97, New York, NY, USA, 2018. Association for Computing Machinery.
- [12] S. Panth et al. Design and CAD Methodologies for Low Power Gate-level Monolithic 3D ICs. In *2014 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED)*, pages 171–176, Aug 2014.
- [13] R. Pendurkar, A. Chatterjee, and Y. Zorian. Switching Activity Generation with Automated BIST Synthesis for Performance Testing of Interconnects. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 20(9):1143–1158, 2001.
- [14] M. M. Shulaker et al. Monolithic 3D integration: A path from concept to reality. In *2015 Design, Automation Test in Europe Conference Exhibition (DATE)*, pages 1197–1202, 2015.
- [15] T. Srimani et al. Heterogeneous Integration of BEOL Logic and Memory in a Commercial Foundry: Multi-Tier Complementary Carbon Nanotube Logic and Resistive RAM at a 130 nm node. In *IEEE VLSI*, 2020.
- [16] H. . P. Wong et al. Carbon Nanotube Field Effect Transistors - Fabrication, Device physics, and Circuit Implications. In *2003 IEEE International Solid-State Circuits Conference, 2003. Digest of Technical Papers. ISSCC.*, pages 370–500 vol.1, Feb 2003.