Leveraging Monolithic 3D Integrated Circuit Technology for Emerging Applications

A Dissertation Presented

by

Chen Yan

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Electrical Engineering

Stony Brook University

December 2018

Stony Brook University

The Graduate School

Chen Yan

We the dissertation committee for the above candidate for the
Doctor of Philosophy degree,
hereby recommend acceptance of this dissertation.

Dr. Emre Salman - Advisor of Dissertation
Associate Professor, Department of Electrical and Computer Engineering

Dr. Milutin Stanacevic - Chairperson of Defense
Associate Professor, Department of Electrical and Computer Engineering

Dr. Peter Milder - Defense Committee Member
Associate Professor, Department of Electrical and Computer Engineering

Dr. Qiaoyan Yu - Defense Committee Member
Associate Professor, Department of Electrical and Computer Engineering,
University of New Hampshire

This dissertation is accepted by the Graduate School

Richard Gerrig
Dean of the Graduate School

Abstract of the Dissertation

Leveraging Monolithic 3D Integrated Circuit Technology for Emerging Applications

by

Chen Yan

Doctor of Philosophy

in

Electrical Engineering

Stony Brook University

2018

Three-dimensional (3D) integrated circuits (ICs) have emerged as a practical solution to some of the critical issues encountered in planar technologies such as longer global interconnects and difficulty in scaling the transistors. Through-silicon via (TSV) based 3D ICs have attracted significant attention during the past decade with emphasis on both fabrication and design methodologies. However, a typical TSV diameter is in the range of several micrometers, which is multiple orders of magnitude larger than nanoscale transistor dimensions. Thus, bulky TSVs not only restrict the integration density, but also limit the power and performance advantages of vertical integration due to significant TSV capacitance. More recently, interest on monolithic 3D integration has grown due to encouraging developments on sequentially fabricating multiple transistor layers on a single substrate.

In monolithic vertical integration, stacked transistors are sequentially fabricated after the bottom layers have been manufactured. Communication among the tiers is achieved by monolithic inter-tier vias (MIVs). MIVs have comparable size to conventional on-chip metal vias since multiple tiers can be aligned with lithographic alignment precision. Thus, MIV based 3D integration enables significantly higher interconnect density as compared to TSV based vertical integration.

In this dissertation, an open source cell library based on a full custom design of each cell as well as a fully functional process design kit (PDK) are developed for transistor-level monolithic 3D integration technology. The power and timing characteristics of each cell are fully characterized with both SPICE-level simulations and a commercial library characterization tool to ensure accuracy. The proposed cell library is used to evaluate the power and timing characteristics of multiple benchmark circuits and a large scale 128-point FFT core with approximately 330K cells. Experimental results demonstrate that at a clock frequency of 1.5GHz, the FFT core implemented with the proposed monolithic 3D library consumes 51% less footprint as compared to conventional 2D technology. The 20% reduction in wirelength enables approximately 22% reduction in net switching power. The entire proposed library and related files for tool integration are publicly available to facilitate future research.

The effect of routing congestion on timing characteristics is stronger in monolithic 3D technology due to the significant reduction in footprint. Routing congestion increases the communication latency, thus degrades the quality of the IC. The cell-level number of routing tracks plays an essential role in routing congestion. Three versions of the cell library with different heights are developed to investigate the effect of the number of routing tracks on area, power, and delay characteristics. Simulation results of a large FFT core demonstrate that an optimum number of routing tracks exists which achieves 14% and 12.5% reduction in, respectively, overall power consumption and worst negative slack as compared to the conventional 2D technology, while consuming 38% less physical area. Clock tree characteristics of the large FFT core are also investigated.

The security of integrated circuits has emerged as a fundamental issue due to the threats from the globalized semiconductor supply chain. Monolithic 3D integrated circuits enable not only ultra-high density device integration, but also introduce novel opportunities and challenges on managing hardware security. The proposed PDK and cell library are used to develop an efficient logic camouflaging method for monolithic 3D ICs. Logic camouflaging is a layout-level technique to thwart image analysis based reverse engineering attacks. Full custom cell libraries are developed and characterized to camouflage large-scale 2D and 3D circuits. The methodology

is implemented and evaluated using a SIMON block cipher and several ISCAS'89 benchmark circuits.

Thermal management of monolithic 3D ICs is more challenging than 2D circuits due to the reduced circuit area (hence, higher power density) and the low thermal conductivity of the inter-layer dielectric material, situated between adjacent device layers. Therefore, accurate and efficient analysis of the thermal behavior is crucial for the reliability of monolithic 3D ICs. To determine the heat propagation and track the formation of hot spots, the thermal integrity of monolithic 3D circuits is explored at the physical level. Results demonstrate that the steady-state temperature increases by approximately 2 to 4 °C in monolithic 3D circuits as compared to the conventional 2D technology. The increase in temperature is not as significant due to two primary factors: (1) additional process layers in monolithic 3D technology are sufficiently thin, (2) power consumption is reduced by approximately 10%.

The proposed open source monolithic 3D process design kit and related design guidelines described in this dissertation provide significant insight into important characteristics of monolithic 3D ICs, such as 1) footprint, timing and power consumption, 2) routing congestion, 3) the effect of the number of routing tracks in each cell, 4) circuit camouflaging for security, and 5) thermal management.

Table of Contents

List of Figures

List of Tables

ACKNOWLEDGEMENTS

First and foremost I would like to thank my advisor, Prof. Emre Salman. It has been such an honor to be his Ph.D. student. I appreciate all his tremendous contributions of time, ideas, and funding to make my Ph.D. career productive and memorable. Also, I want to express my sincere gratitude for his valuable feedback, advice, and encouragement, for his patience, motivation, and immense knowledge. I could not have imagined having a better advisor for my Ph.D. study. Thankful for him providing an excellent example as a successful researcher and professor.

Besides my advisor, I would like to thank my managers, Dr. Pengbo (Shelley) Sun and Alexander Martin, from GLOBALFOUNDRIES, who provided me with excellent internship opportunities. Special thanks to my mentors, Dr. Qi Qi, and Dr. Guoqing Deng, for their tremendous support. Also, profound gratitude goes to Prof. Qiaoyan Yu from University of New Hampshire, for her insightful discussions and ideas.

I am very thankful to all my defense committee: Prof. Sangjin Hong, Prof. Milutin Stanacevic, Prof. Peter Milder and Prof. Qiaoyan Yu, for their valuable comments and questions.

I owe a high debt of gratitude to all my fellow labmates in NanoCAS Labo-

ratory: Hailang, Zhihua, Weicheng, Tutu, Mallika, Manav, Ivan and Krithika for stimulating discussions, for the late nights we were working together before the deadlines and for the great times we shared in the last five and a half years. Thank you for making my Ph.D. life not only about simulations but also friendship. I wish all of you the best of luck and a bright future.

Last but not least, I would like to thank my family: my parents Xinmin Yan, Yingjie Liu and my partner Chris Perkins Yan for their endless emotional support and love throughout my Ph.D. career and my life in general. I dedicate this thesis to them.

Chapter 1

Introduction

Technology scaling has been a powerful driver for the continuous development of the microelectronics industry. However, conventional CMOS technologies face critical challenges as technology nodes keep shrinking, particularly below 10 and 7nm. Recently, several CMOS foundries decided to stop 7nm FinFET technology development [7–9], such as GLOBALFOUNDRIES (GF) [10] and United Microelectronics Corporation (UMC) [11]. 3D integrated circuits with through-silicon via (TSV) or wire bonding technology have been explored to enable the continuous growth in transistor density [12]. In TSV based 3D integration, multiple dies are thinned, aligned, and vertically bonded, thereby enabling shorter global interconnects (and therefore reduced power consumption) and heterogeneous integration [13].

Monolithic 3D integration has received growing attention due to ultra-high density device integration with the smallest vertical interconnects [14]. In monolithic vertical integration, stacked transistors are sequentially fabricated after the bottom layers have been manufactured. Communication among the tiers is achieved by

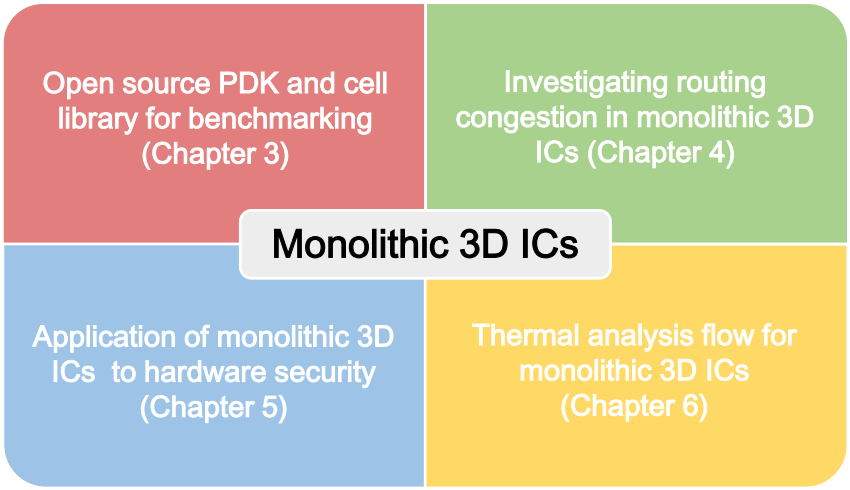


Figure 1.1: Summary of the contributions.

monolithic inter-tier vias (MIVs). MIVs have comparable size to conventional on-chip metal vias since multiple tiers can be aligned with lithographic alignment precision [15]. Thus, MIV based 3D integration enables significantly higher interconnect density as compared to TSV based vertical integration.

A monolithic 3D standard cell library integrated into a conventional design flow is developed to facilitate the design process of complex ICs in monolithic 3D technology. Furthermore, the routing complexity increases in monolithic 3D ICs due to the significant reduction in footprint. Thus, routing congestion aware libraries are developed to mitigate this issue. Increasing power density leads to thermal issues in vertically integrated circuits. A thermal analysis flow is proposed to explore monolithic 3D heat dissipation. In addition to these challenges, 3D ICs introduce new opportunities on managing hardware security, such as split manufacturing techniques and logic camouflaging, as shown in this thesis. These challenges and opportunities are summarized in Chapter 2.

In this thesis, four contributions are proposed on monolithic 3D integration technology, as summarized in Figure 1.1. An open source cell library based on a full custom design of each cell is developed for transistor-level monolithic 3D integration [16] based on existing 2D 45 nm process design kit. The proposed cell library is used to evaluate the power and timing characteristics of multiple benchmark circuits and a large scale 256-point FFT core [17]. Multiple versions of the library with different cell heights are also developed to investigate the tradeoffs among routability, timing, power and area characteristics. This analysis is essential since routing congestion is one of the primary physical design issues in monolithic 3D ICs. Clock tree characteristics of the large FFT core are also investigated.

For 3D hardware security, an efficient logic camouflaging method for monolithic 3D ICs is presented. Full custom cell libraries are developed and fully characterized to camouflage large-scale 2D and 3D circuits. The area, power, and timing overhead of circuit camouflaging is evaluated both at the cell- and chip-levels.

Higher temperatures and thermal gradients are expected for 3D ICs due to increased power densities. An accurate exploration of the thermal behavior of monolithic 3D circuits is provided by combining the proposed cell library and a publicly available thermal simulator which supports cell-level analysis.

The rest of this thesis is organized as follows. Background on 3D ICs and existing works on monolithic 3D ICs are provided in Chapter 2. The proposed open source monolithic 3D standard cell library and process design kit are presented in Chapter 3. The proposed routing congestion aware monolithic 3D standard cell library is introduced in Chapter 4. The proposed hardware-efficient camouflaging methodology for monolithic 3D ICs is elaborated in Chapter 5. Explorations of the thermal behavior in monolithic 3D ICs and related analysis flow are described in

Chapter 6. Finally, the thesis is concluded in Chapter 7 with a brief discussion on future research directions.

Chapter 2

Background

As indicated by International Roadmap for Devices and Systems (IRDS) [18], after 2024, there will not be enough headroom for 2D geometry scaling. 3D ICs have emerged as a compelling solution for a portion of the issues experienced in planar technologies such as longer global interconnects and difficulty in scaling the transistors [19–21]. Vertical integration technologies are expected to be critical performance boosters, as predicted by IRDS [18]. Several different 3D manufacturing technologies are introduced in Section 2.1. Existing works on monolithic 3D ICs are summarized in Section 2.2.

2.1 Three-Dimensional Integration

3D ICs stack 2D dies along the vertical dimension. As compared to traditional 2D planar technologies, 3D ICs achieve significant benefits, such as smaller footprint, shorter interconnect, lower power consumption and higher communication bandwidth. Also, 3D ICs provide opportunities for highly heterogeneous and mul-

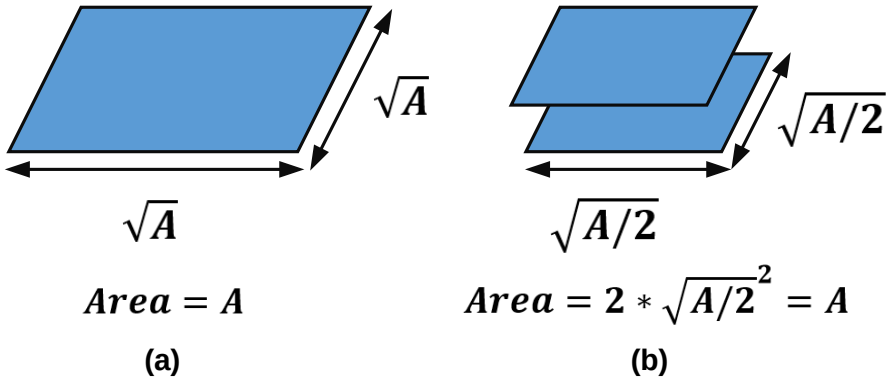


Figure 2.1: Reduction in wirelength where the conventional 2D circuit is implemented in two planes: (a) Planar integrated circuit with an area *A*, (b) Two-plane 3D integrated circuit with an area *A/2* for each plane.

tifunctional systems and new hardware implementations for circuit security.

The core competitiveness of 3D integration is the decrease in the length of the longest interconnects within an IC [20]. As shown in Figure 2.1 (a), assuming a conventional 2D circuit with an area *A*, the longest interconnect in this planar IC has a length

$$L_{max,2D} = 2 \times \sqrt{A}. \tag{2.1}$$

If the same circuit is implemented with two stacked dies with an area *A/2* for each plane, the overall area of the system will not change, but the length of the longest interconnect for this 3D IC will be reduced to

$$L_{max,3D} = 2 \times \sqrt{A/2}, \tag{2.2}$$

which is $\sqrt{2}/2$ of $L_{max,2D}$. Therefore, by increasing the number of dies in a 3D IC,

the length of the longest interconnect can be further reduced.

2.1.1 Manufacturing Technologies for 3D ICs

There are two primary types of 3D IC manufacturing process: a sequential or a parallel process [20]. In the case of a parallel process, a 3D IC is manufactured by bonding multiple wafers or bare dies. Alternatively, in a sequential process, the devices and metal layers of the higher planes of the stack are fabricated layer by layer on top of the first plane.

3D ICs can be classified in terms of the level of vertical interconnect hierarchy. In general, 3D integration contains technologies such as system-in-package (SiP), through-silicon via (TSV)-based 3D integration, monolithic 3D integration, and contactless coupled 3D ICs.

2.1.1.1 System-in-Package

A system-in-package (SiP) is a variety of ICs contained in a single package [22], where the interconnections of these circuits are implemented by the third dimension by using wire bonding, peripheral vertical interconnects, area array vertical interconnects, or metalization between the faces of a 3D stack, as shown in Figure 2.2 [20].

An SiP will considerably increase the packaging efficiency by improving the die-to-package area ratio and reducing the package footprint at the same time. Also, SiP provides a full commercialized variant of vertical integration. Thus, many firms have leveraged SiP products [3, 23].

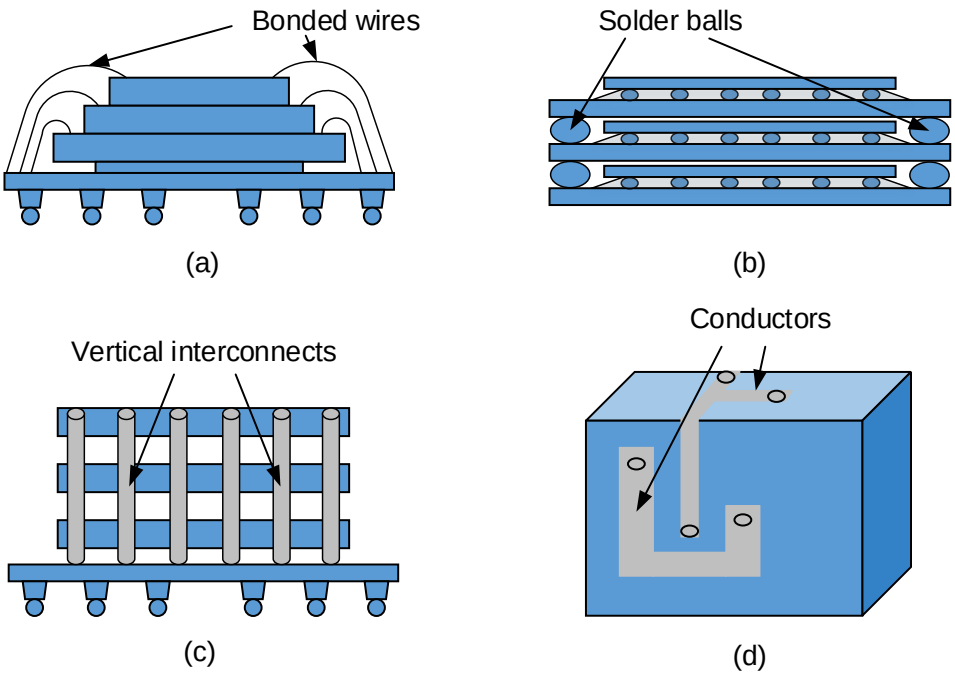


Figure 2.2: Examples of SiP: (a) wire-bonded SiP [1], (b) solder balls at the perimeter of the planes [2], (c) area array vertical interconnects, and (d) interconnects on the faces of the SiP [3].

2.1.1.2 3D ICs with Through-Silicon Vias (TSV)

Through-silicon via (TSV) based 3D ICs have attracted vital attention throughout the past decade with emphasis on both fabrication and design methodologies [24–28]. In TSV based 3D integration, multiple dies are thinned, aligned, and vertically bonded, thereby, enabling shorter global interconnects (and therefore reduced power consumption) and heterogeneous integration [24, 26, 29–36], as shown in Figure 2.3.

However, a typical TSV diameter is in the range of several micrometers, which is multiple orders of magnitude greater than nanoscale transistor dimensions. Furthermore, TSVs induce thermo-mechanical stress in the front-end-of-line (FEOL) layer, thereby impacting the transistor behavior [37]. Thus, bulky TSVs not only restrict the integration density but also limit the power and performance advantages of vertical integration due to significant TSV capacitance and stress [38–41].

2.1.1.3 Monolithic 3D ICs

More recently, interest on monolithic 3D integration has grown due to encouraging developments on sequentially fabricating multiple transistor layers (particularly the thermal characteristics) [14, 42, 43]. As depicted in Figure 2.4, in monolithic vertical integration, stacked transistors are sequentially fabricated after the bottom layers have been manufactured. Communication among the tiers is achieved by monolithic inter-tier vias (MIVs). A critical challenge in the fabrication of monolithic 3D ICs is to minimize the detrimental effect of the manufacturing process of the top tier on bottom tier [44]. Thus, significant research on the fabrication side has focused on developing low thermal budget processes, typically less than 500-600°C for the upper tiers [45, 46].

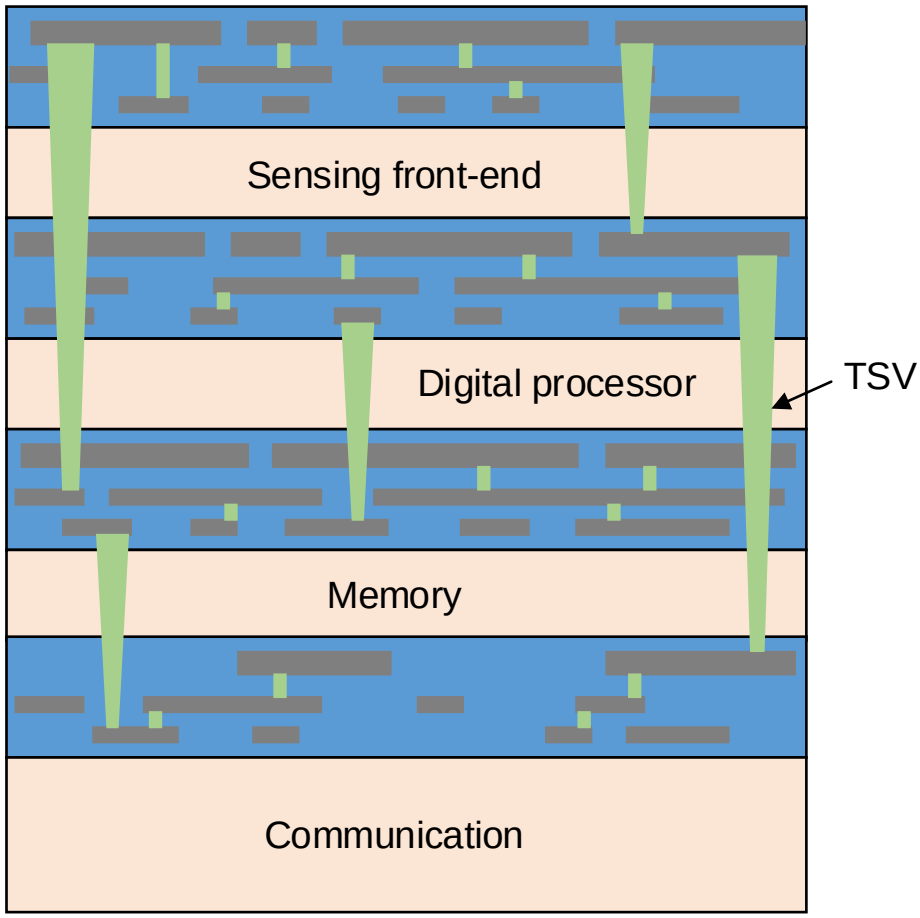


Figure 2.3: Three-dimensional integration of diverse planes using TSV technology [4].

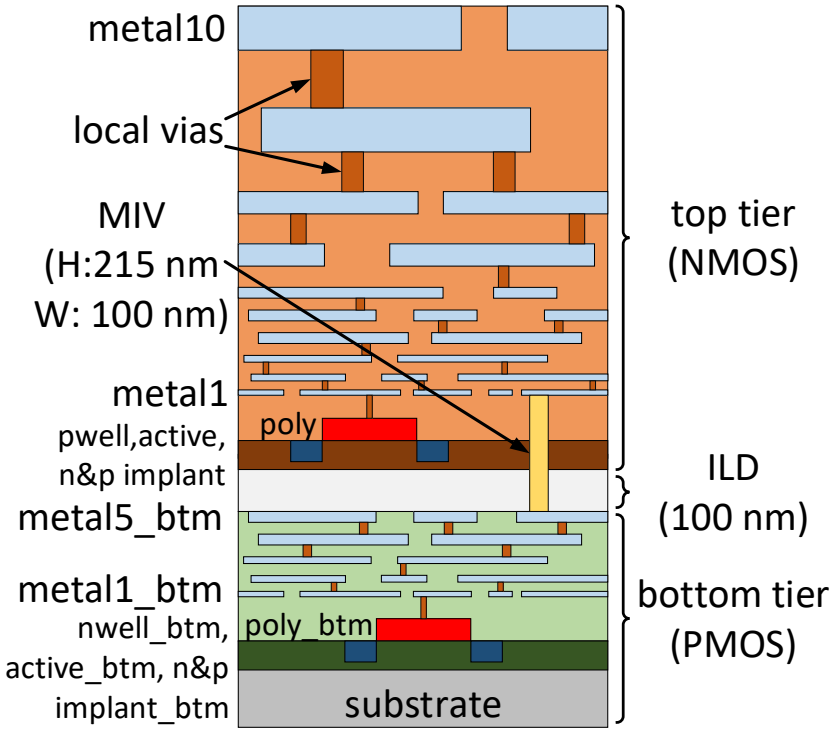


Figure 2.4: Cross-section of the transistor-level monolithic 3D technology with two tiers. The top tier hosts the nMOS transistors whereas the pMOS transistors are placed within the bottom tier.

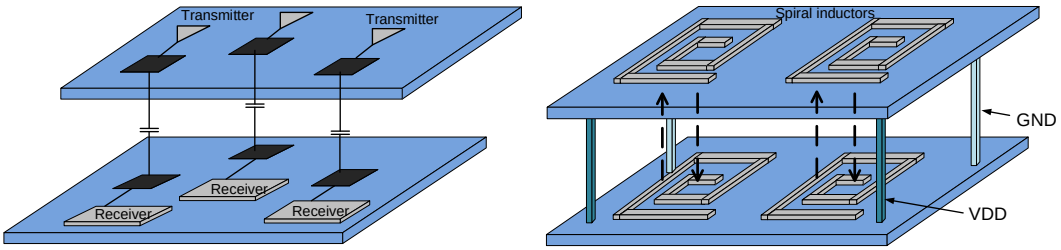


Figure 2.5: Contactless 3D integrated circuits: (a) capacitively coupled 3D integration, (b) inductively coupled 3D integration.

MIVs have comparable size to conventional on-chip metal vias since multiple tiers can be aligned with lithographic alignment precision [15]. Thus, MIV based 3D integration enables significantly higher interconnect density as compared to TSV based vertical integration.

2.1.1.4 Contactless 3D ICs

Contactless 3D ICs represent another technique for communication among circuits situated on different planes. This technique relies on coupling of electrical or magnetic fields.

As shown in Figure 2.5 (a), in capacitively coupled 3D integration, on each plane there are small on-chip parallel capacitors. Transmitter and receiver modules are needed to drive the capacitors to modulate the received signals. Since transmitter and receiver electronic equipments consume significant power, capacitively coupled 3D integration is comparatively power inefficient.

In inductively coupled 3D ICs, as illustrated in Figure 2.5 (b), there is a spiral inductor located on every plane of a two-plane 3D IC. Similar to the capacitively coupled 3D integration, some specialized circuits are necessary for transmission

and receiving signals, thus, inductively coupled 3D ICs exhibit high power consumption as well. Also, the size of the inductors is typically very large [47], particularly for conventional CMOS technology where the fabrication of on-die inductor is relatively area inefficient.

2.1.2 Design Methodologies for Monolithic 3D ICs

Monolithic 3D integration offers multiple benefits over conventional 2D circuits. To utilize vertical integration efficiently, the technology should support a very high density of vertical interconnects with dimensions comparable to nanoscale devices, as achieved by monolithic 3D integration. There are three design styles for monolithic 3D technology: transistor-level, gate-level, and block-level, as described below.

2.1.2.1 Transistor-level Monolithic 3D ICs

In transistor-level monolithic 3D integration, as shown in Figure 2.6, nMOS and pMOS transistors within a circuit are separated into two different tiers. For example, the pull-down network of each gate is placed within one tier whereas the pull-up networks are placed in another tier. This approach not only achieves fine-grained 3D integration with intra-cell MIVs but also enables the individual optimization of the bottom and top tier devices [48]. Existing design automation methodologies (with modifications) can be used for this approach.

2.1.2.2 Gate-level Monolithic 3D ICs

In gate-level monolithic 3D integration, as shown in Figure 2.7, multiple cells

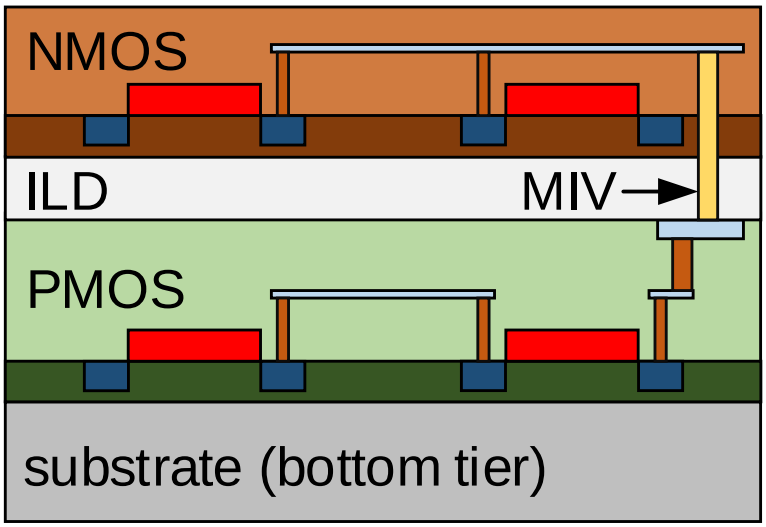


Figure 2.6: Design methodology for monolithic 3D technology: transistor-level monolithic 3D.

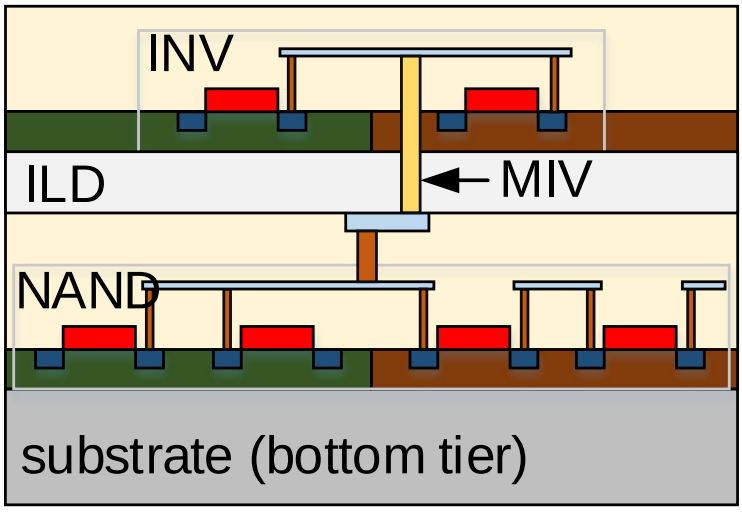


Figure 2.7: Design methodology for monolithic 3D technology: gate-level monolithic 3D.

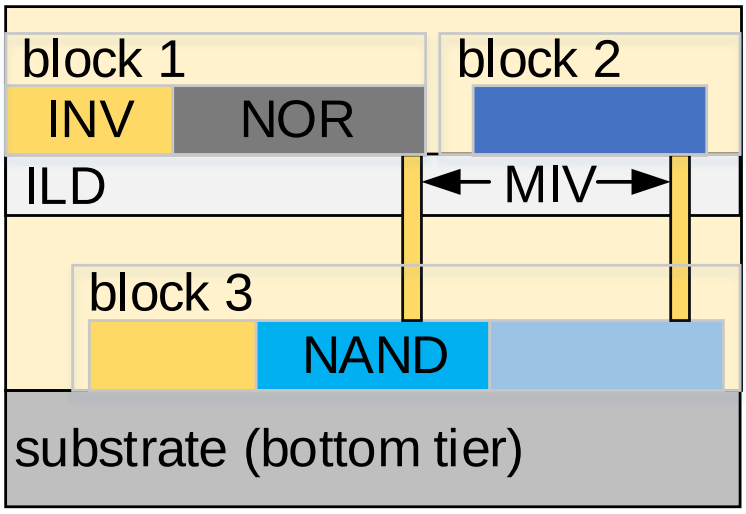


Figure 2.8: Design methodology for monolithic 3D technology: block-level monolithic 3D.

within a functional block are partitioned into multiple tiers. MIVs are utilized for inter-cell communication. There is no area overhead for each cell, and a sufficiently high integration density can be achieved due to placement flexibility of the MIVs. However, novel partitioning algorithms are needed. An unbalanced partition can degrade the power and performance characteristics of gate-level monolithic 3D ICs.

2.1.2.3 Block-level Monolithic 3D ICs

Finally, block-level monolithic 3D integration, as shown in Figure 2.8, represents a more coarse-grain integration where the partitioning of the integrated circuit is achieved based on individual functional blocks. In this method, the benefits of monolithic 3D technology cannot be fully utilized since the high density MIVs is

not sufficiently exploited. Furthermore, proper partitioning algorithms need to be carefully applied.

The primary focus in this thesis is on transistor-level monolithic 3D integration since the benefits provided by the small dimensions of the vertical interconnects can be utilized the most. Furthermore, individual device optimizations per plane are possible in transistor-level monolithic 3D technology.

2.2 Existing Research on Monolithic 3D ICs

2.2.1 Fabrication-Level Research

As shown in Figure 2.9, a 50nm 3D sequential structure on 10nm silicon was successfully fabricated by CEA-LETI [5]. CEA-LETI has also been collaborating with STMicroelectronics [49] and IBM [50] to develop a monolithic 3D technology with InGaAs nFET devices fabricated sequentially over SiGe pFET devices on a silicon-on-insulator (SOI) substrate.

Monolithic 3D Inc. has announced the manufacturing process of monolithic 3D wafers [51]. Transistors are built with c-Si films above the copper/low k materials to avoid alignment issues of bonding fabricated wafers. In order to not damage copper/low k layers, transistor fabrication temperature should be less than 500 - 600°C [46, 52].

2.2.2 Design-Level Research

Hsueh *et al.* have demonstrated a monolithic 3D vertical cross-tier compute-in-memory (CIM) SRAM cell fabricated using low-cost TSV-free FinFET-based

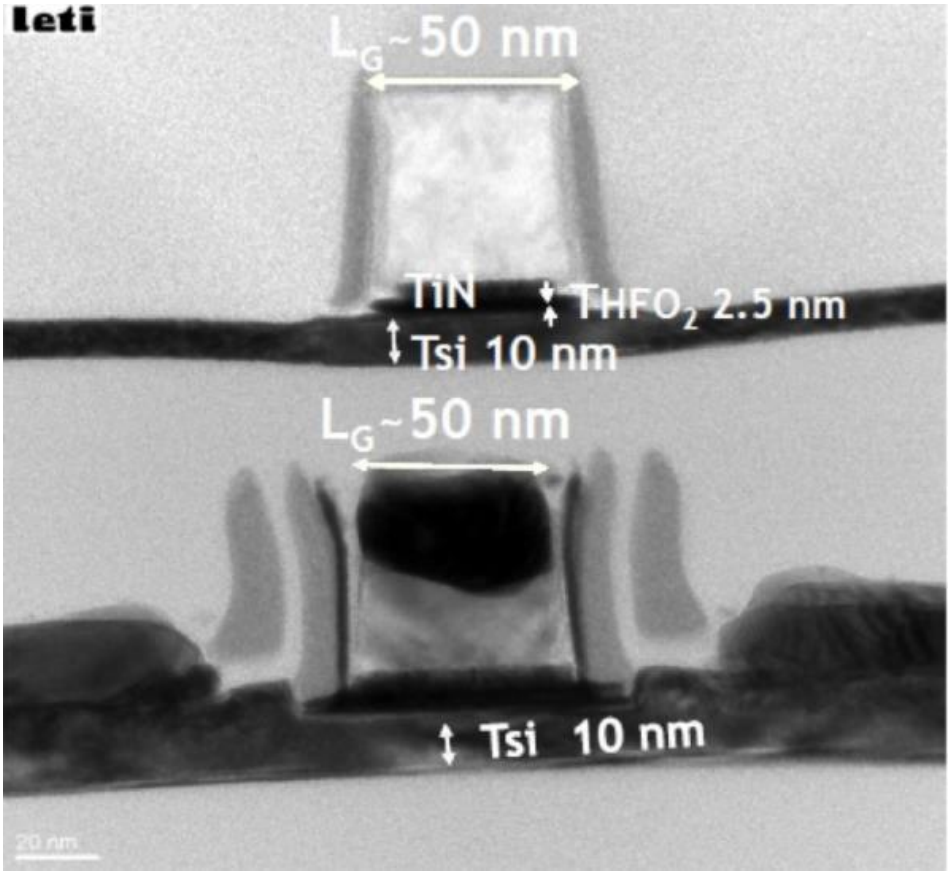


Figure 2.9: Cross sectional image of two stacked devices fabricated through monolithic 3D integration technology [5].

3D IC technology [53]. Proposed 9T 3D CIM SRAM cell is able to compute NAND/AND, OR/NOR and XOR/XNOR operations within a single memory cycle. Proposed scheme enables the fabrication of energy and area efficient circuits for the cost-aware intelligent Internet of things (IoT) devices. As compared to the 2D version, the monolithic 3D SRAM device reduces area by 51%.

Liu and Lim have investigated the design tradeoffs in monolithic 3D ICs considering both transistor-level and gate-level monolithic integration [54]. Useful physical design guidelines and insight into the routability issue have been provided. The effects of inter-tier process variation have also been investigated. In this work, however, authors have assumed that the monolithic 3D gates and traditional 2D gates have the same power and timing characteristics. This assumption is not accurate due to different parasitic impedances within a 3D monolithic cell and the existence of MIVs.

Lee *et al.* have fixed this limitation by individually characterizing the timing and power consumption of transistor-level monolithic 3D cells [55]. The power characteristics of several 3D monolithic benchmark circuits have been investigated and compared with 2D versions at similar timing constraints. The authors, however, have adopted the cell-folding method and used the same pull-up and pull-down networks as in 2D cells. MIVs have been inserted in between these two networks. As a result, the proposed 3D cells are not optimized for the footprint. Also, in [55], the timing constraints are relatively relaxed, which may prevent to investigate the behavior of the monolithic 3D technology under tighter clock frequency constraints.

Shi *et al.* have demonstrated the power benefits of transistor-level monolithic 3D ICs through the custom design of a cell library in 14nm FinFET technology, also utilizing the cell-folding method [12]. A dedicated track is assumed for the

MIVs. A detailed cell-level *RC* extraction methodology is described. The authors, however, did not investigate the timing characteristics of the benchmark circuits. The effect of routing congestion on timing constraints and power consumption is not discussed.

Chapter 3

Mono3D: Open Source Cell Library for Monolithic 3D ICs

In this chapter, an open source standard cell library for design automation of large-scale transistor-level monolithic 3D ICs is proposed. A 256-point, highly parallelized FFT core with 1.6M cells is implemented with the proposed library. Power and timing characteristics of monolithic 3D ICs are quantified. The effect of the signal integrity and routing congestion on timing characteristics is investigated. The proposed open source cell library facilitates future research on multiple aspects of monolithic 3D technology.

The rest of this chapter is organized as follows. The details of the proposed open source cell library, characterization, and design flow are provided in Section 3.1. Power/timing and several important physical design characteristics of cells and an FFT core with monolithic 3D implementation are investigated in Section 3.2. Finally, this chapter is summarized in Section 3.3.

3.1 Open Source Cell Library for Monolithic 3D ICs

The characteristics of the proposed cell library are described in Section 3.1.1. The design flow to integrate the proposed library into the design process is discussed in Section 3.1.2.

3.1.1 Library Development

In this work, the *Mono3D*, an open source standard cell library for transistor-level monolithic 3D technology is developed in 45 nm technology [16]. *Mono3D* consists of two tiers where each tier is based on the 2D 45 nm process design kit *FreePDK45* from North Carolina State University (NCSU) [56]. Thus, the process and physical characteristics (transistor models and characteristics of the on-chip metal layers) are obtained from the *FreePDK45*. Similar to [12, 55], the pull-down network of a CMOS gate (nMOS transistors) is built within the top tier whereas the pull-up network (pMOS transistors) is fabricated within the bottom tier. Note that the processing temperature of the top tier is constrained to be less than 500-600°C [45] to not damage the transistors within the bottom tier. This relatively low processing temperature, however, degrades the quality of the top-tier devices. Thus, pMOS devices (that already have lower mobility) are placed within the bottom tier. As such, the proposed cell library can only be used for transistor-level monolithic 3D approach since MIVs exist within each standard cell to connect nMOS and pMOS devices. The transistor device characteristics are the same as in 2D *FreePDK45*. Thus, any processing temperature related degradations are not considered. However, the impact of novel devices/models and manufacturing steps for 3D monolithic integration can be captured by replacing/modifying the device

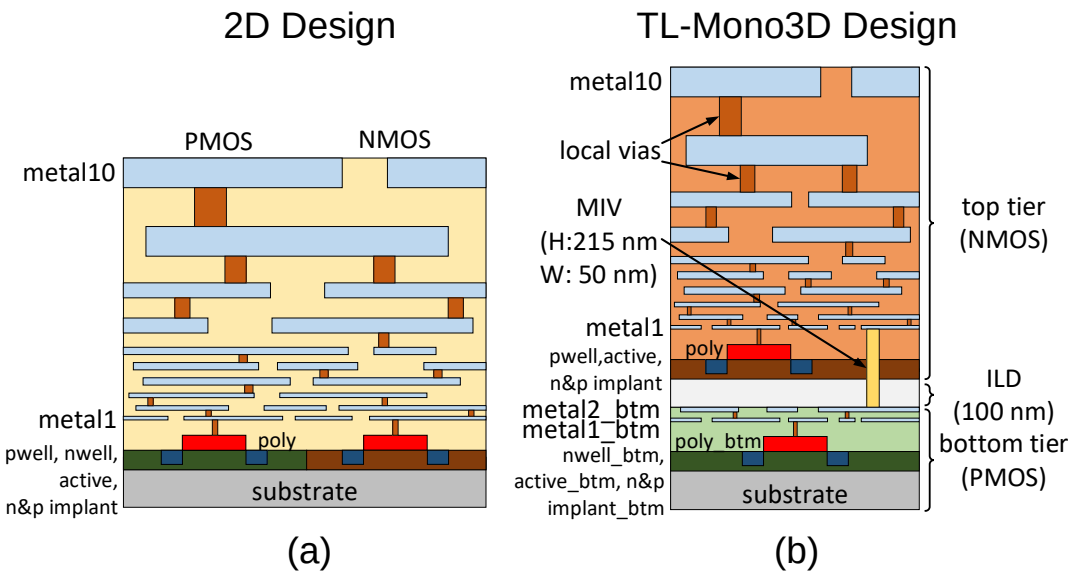


Figure 3.1: Cross-sections of the conventional 2D and transistor-level monolithic (TL-Mono) 3D technology with two tiers. The top tier hosts the nMOS transistors whereas the pMOS transistors are placed within the bottom tier.

models within the provided process design kit. System-level effects of varying device characteristics (due to, for example, the manufacturing steps of the top tiers) can, therefore, be investigated.

In the proposed *Mono3D*, two metal layers are allocated to the bottom tier (metal1_btm and metal2_btm), as illustrated in Figure 3.1. These metal layers are primarily for routing the intra-cell signals.

The top tier is separated from the bottom tier with an inter-layer dielectric (ILD) with a thickness of 100 nm. Inter-tier coupling is minimized at this thickness, as experimentally validated [15]. The 10 metal layers that exist in 2D *FreePDK45* are maintained the same for the top tier in *Mono3D*. The intra-cell connections that span the two tiers are achieved by MIVs. Each MIV has a width of 50 nm and a height of 215 nm [57].

AND2X1	INVX2
AOI21X1	INVX4
BUFX2	LATCHNEG
BUFX4	MUX2X1
CLKBUF1	NAND2X1
CLKBUF2	NOR2X1
CLKBUF3	OAI21X1
DFFPOSX1	OR2X1
FILL	XNOR2X1
INVX1	XOR2X1

Table 3.1: List of standard cells in the monolithic 3D library.

Currently, 20 standard cells exist in *Mono3D*, as listed in Table 3.1. In addition to the fundamental cells, multiple clock buffers and a latch are included. Each cell is developed with a full-custom design methodology using a cell stacking technique. As opposed to [12,55] where the power (within the bottom tier) and ground (within the top tier) rails overlap, in the proposed *Mono3D*, the power rail is located at the top of the bottom tier, and ground rail is located at the bottom of the top tier. These power and ground rails at each cell row are connected to the system-level power network through power and ground rings placed during the placement and routing process.

A specific track is allocated for intra-cell MIVs, which are distributed within the cell to minimize the interconnect length and reduce the cell height. Each cell within the 2D *NanGate* library has 14 routing tracks. In this thesis, monolithic 3D cell libraries are developed with 9 tracks. The number of tracks plays an essential role in chip-level routing congestion, a primary issue in monolithic 3D ICs. The cell heights in *Mono3D* is 1.52 μm. This cell height is 38% shorter than the standard cell height (2.47 μm) in *NanGate* cell library [58].

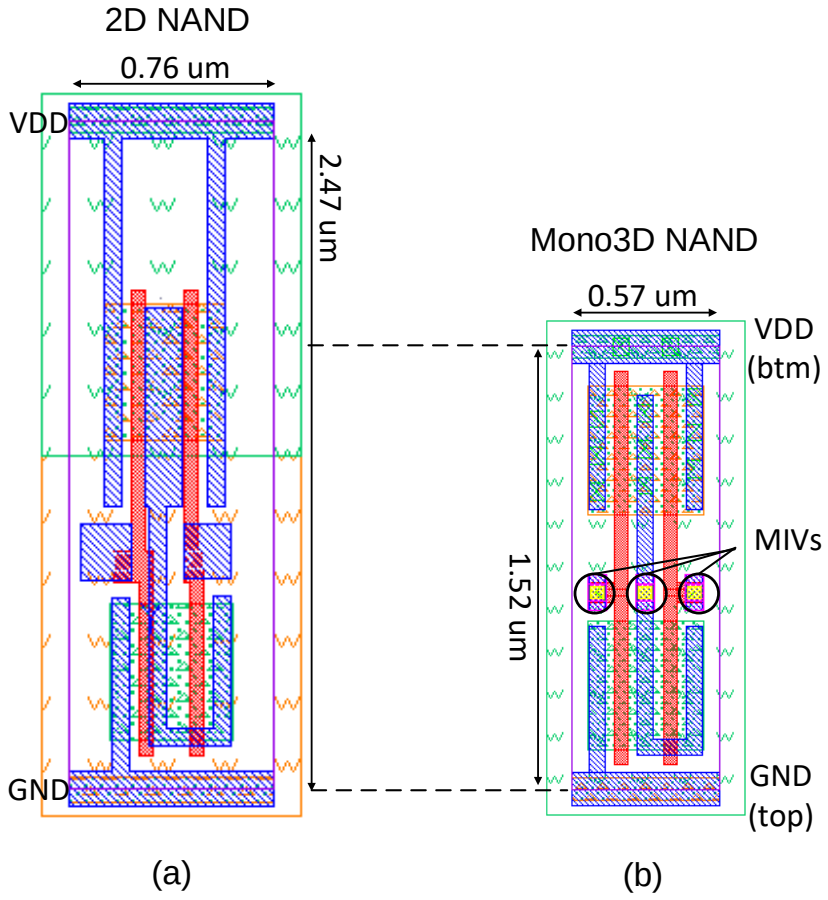


Figure 3.2: Comparison of the layout views of a NAND gate in (a) traditional 2D technology with 14 routing tracks in each cell, (b) monolithic 3D technology with 9 routing tracks, illustrating the three MIVs used to connect the top and bottom tiers.

The layout of a NAND cell is illustrated in Figure 3.2 in both 2D and 3D monolithic technologies. Cell dimensions and the three MIVs are highlighted. Similarly, a 2D D-flip-flop cell and 3D monolithic D-flip-flop cell within *Mono3D* are compared in Figure 3.3.

In this case, the top and bottom tiers are separately depicted. Note that the width of the 3D flip-flop cell increases by approximately 7% due to MIVs and intra-cell

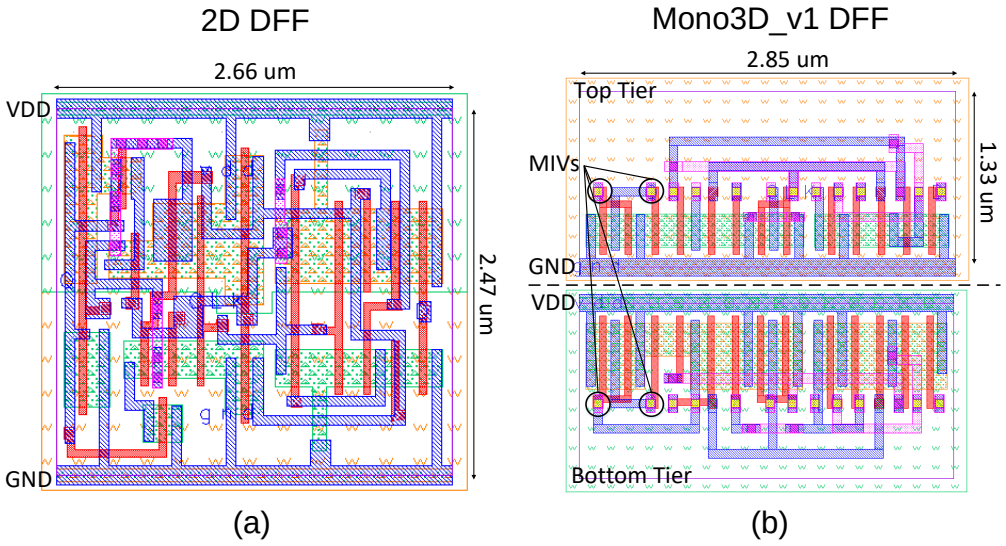


Figure 3.3: Comparison of the layout views of a D-flip-flop in traditional 2D and transistor-level monolithic 3D technology. The top and bottom tiers are separately depicted for the 3D technology.

routing. Also, note that particular emphasis is given to provide white space at the top tier (depending upon the number of routing tracks) to avoid pin block issue induced routing congestion.

3.1.2 Design Flow

The design flow adopted in this work and the modifications required for 3D monolithic technology are depicted in Figure 3.4. A new technology file (*.tf*) is generated for *Mono3D* to include all of the new layers (interconnects, via, ILD, and MIV). Based on these modifications, a new display resource file (*.drf*) is generated to develop full-custom layouts of the 3D cells. The design rule check (DRC), layout versus schematic (LVS) and parasitics extraction (PEX) are performed using *Calibre* [59]. The DRC rule file is modified to include new features for the additional

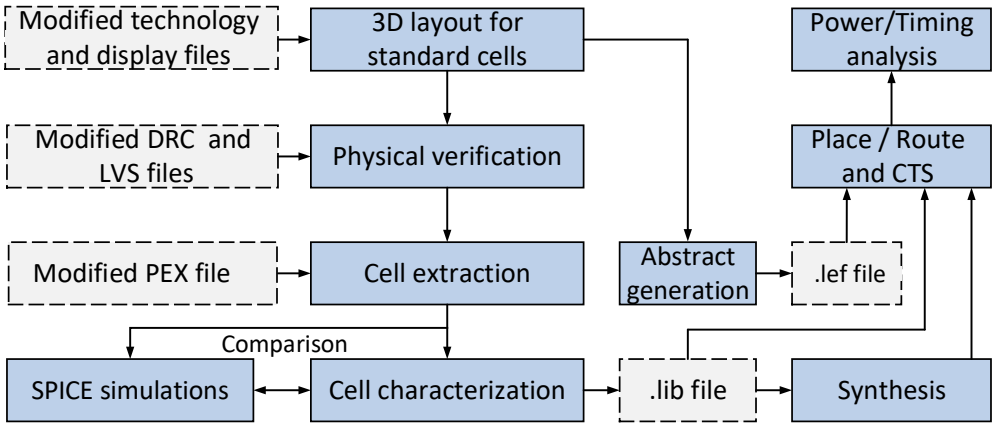


Figure 3.4: Integration of the proposed open source cell library into design flow, illustrating the required modifications.

metal layers, vias, transistors, ILD and MIV. For example, the minimum spacing between two MIVs is equal to 120 nm, producing an MIV pitch of 170 nm.

The LVS rule file is also modified for the tool to be able to identify transistors located in separate tiers independently. The extracted netlist with MIVs is analyzed to accurately obtain the interconnections between nMOS (within the top tier) and pMOS (within the bottom tier) transistors. The *RC* extraction rule file is modified to be able to recognize the new device tier, new metal layers, and MIVs. For metal interconnects, intrinsic plate capacitance, intrinsic fringe capacitance, and near-body (coupling) capacitance are considered between silicon and metal, and metal and metal. A single MIV is characterized by a resistance of 5.5 Ωs and a capacitance of 0.04 fF, based on [12] where device-level extraction is performed. The only parasitic component that is not considered during the extraction process is the tier-to-tier coupling capacitance. As experimentally demonstrated in [15], this component is negligible when the inter-layer dielectric is 100 nm thick.

After *RC* extraction, 3D cells are characterized by *Encounter Library Charac-*

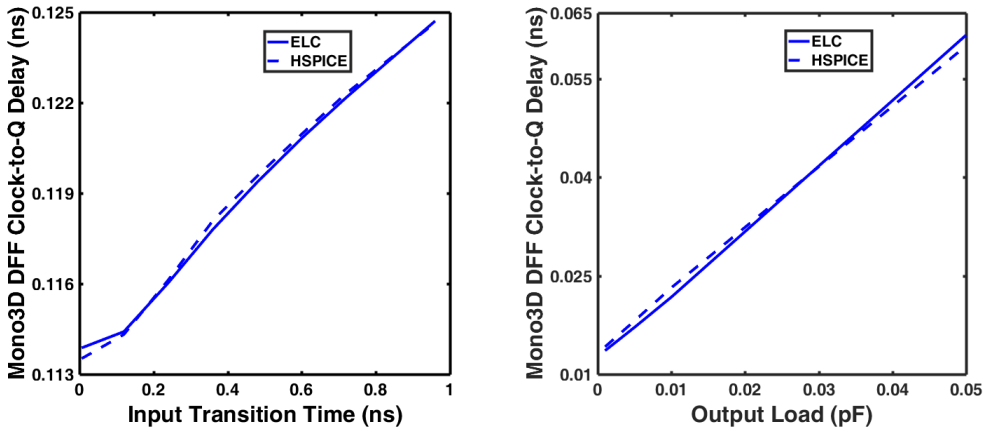


Figure 3.5: Comparison of the automated cell library characterization process with *HSPICE* simulations: (a) clock-to-Q delay of a 3D D-flip-flop as a function of input transition time at constant load capacitance of 0.01 pF, (b) clock-to-Q delay of a 3D D-flip-flop as a function of load capacitance at constant input transition time of 120 ps.

terizer (ELC) [60] to obtain the timing and power characteristics (lookup tables) of each cell. The extracted 3D cell netlists are also simulated with *HSPICE* [61] to ensure the accuracy of the characterization process. For example, the clock-to-Q delay of the 3D D-flip-flop cell (as a function of input transition time and output load capacitance) obtained by *HSPICE* simulations is compared with *ELC* results (look-up tables) in Figure 3.5. As shown in this figure, *ELC* results sufficiently match with *HSPICE* simulations (with an average error of less than 4%), demonstrating the accuracy of the characterization process. More details on the area, timing, and power characteristics of the 3D cells and comparison with 2D cells are provided in Section 3.2.

The *.lib* file for the *Mono3D* generated by *ELC* is converted into the *.db* format, which is used for circuit synthesis, placement, clock tree synthesis, and routing.

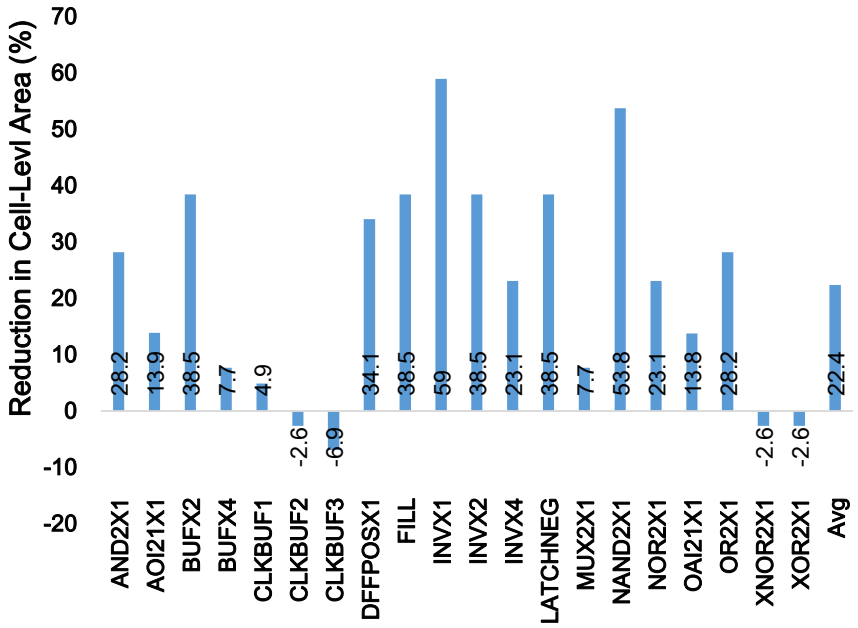


Figure 3.6: Percent reduction in area achieved by each monolithic 3D cell as compared to the 2D cells. Results are provided for the 3D library, *Mono3D*.

Since all of the I/O pins of the 3D cells are located within the top tier, existing physical design tools can be used for these steps.

3.2 Simulation Results

3.2.1 Cell-Level Evaluation

3.2.1.1 Area

Cell-level area improvement obtained by monolithic 3D technology is shown in Figure 3.6. According to this figure, the reduction in cell area varies from -6.9%

to 59.0% in *Mono3D*, depending upon the specific cell. An average improvement of 22% is achieved for *Mono3D*. Note that a negative percent implies that the cell area increases as compared to the 2D cell. This behavior occurs for cells where the reduction in cell height causes a considerable increase in cell width. Similarly, the average area reduction is not as significant as the reduction in cell height since, on average, the cell width slightly increases due to MIVs and intra-cell routing within the reduced cell footprint.

3.2.1.2 Delay and Power Consumption

HSPICE simulations are performed on the extracted 3D netlists to compare monolithic 3D technology with the conventional 2D technology at the cell level. At 1.1 V power supply, 50 ps transition time, and 27°C temperature, average delay and power consumption are analyzed, as listed in Table 3.2 for 2D and 3D technology. According to this table, *Mono3D* cells have, on average, 3.22% higher propagation delay and 0.46% lower power consumption as compared to the 2D standard cells. This slight increase in delay is due to denser cell layout, producing additional coupling capacitances and MIV impedances. Note that in a DFF cell, both delay (clock-to-Q delay) and power are improved as compared to 2D cells since the DFF cell has relatively longer average interconnect length where the monolithic 3D technology is helpful. Also note that the cell-level change in delay and power highly depends upon the individual cell layout, interconnects, and MIVs.

Cells	Delay (ps)		Power (μW)	
	2D	3D	2D	3D
AND2X1	17.63	19.5 (10.6%)	2.82	3.01 (6.7%)
AOI21X1	13.68	13.69 (0.1%)	3.32	3.34 (0.6%)
BUFX2	17.89	17.86 (-0.2%)	14.04	13.92 (-0.9%)
BUFX4	15.82	15.29 (-3.4%)	29.00	28.98 (-0.1%)
CLKBUF1	27.01	27.32 (1.2%)	64.07	62.52 (-2.4%)
CLKBUF2	39.57	40.17 (1.5%)	93.25	90.88 (-2.5%)
CLKBUF3	51.73	53.04 (2.5%)	121.4	119.0 (-2.0%)
DFFPOSX1	41.69	34.72 (-16.7%)	26.75	27.18 (1.6%)
INVX1	6.73	6.42 (-4.6%)	4.69	4.68 (-0.2%)
INVX2	6.54	6.32 (-3.4%)	9.31	9.24 (-0.8%)
INVX4	6.44	6.29 (-2.3%)	18.01	18.16 (0.8%)
MUX2X1	16.25	17.23 (6.0%)	5.81	6.15 (5.9%)
NAND2X1	10.22	9.78 (-4.3%)	1.63	1.58 (-3.1%)
NOR2X1	11.41	12.16 (6.6%)	1.63	1.69 (3.7%)
OAI21X1	12.89	12.88 (-0.1%)	3.27	3.25 (-0.6%)
OR2X1	18.33	21.12 (15.2%)	2.54	2.85 (12.2%)
XNOR2X1	36.05	41.76 (15.8%)	12.66	14.13 (11.6%)
XOR2X1	35.59	42.41 (19.2%)	12.53	14.24 (13.6%)
Average	21.42	22.11 (3.22%)	23.71	23.60 (-0.46%)

Table 3.2: Average delay and power characteristics of 2D and monolithic 3D cells with 9 (*Mono3D*) routing tracks. The percent changes with respect to 2D cells are listed.

3.2.2 System-Level Evaluation

The proposed open source *Mono3D* cell library is used to investigate the footprint, power, and timing characteristics of several benchmarks with a various number of gates, ranging from 2.7K to 1.6M. For the conventional 2D technology and synthesis, the 45 nm *NanGate* cell library and the *FreePDK45* process kit are used, whereas for the monolithic 3D technology, the proposed *Mono3D* library is used (all libraries have the same type of cells for the fair comparison). Circuits are

Circuit	SIMON	s38584	FFT64	FFT128	FFT256
No. Gates	2,697	22,273	89,991	691,839	1,467,815
DFF	392	1,426	18,963	96,746	190,025
INV	206	6,798	14,069	74,945	152,803
NAND	9	916	16,266	104,613	193,864
NOR	332	1,040	5,477	81,293	210,231
AND	138	8,559	20,468	106,012	197,293
OR	1,228	3,534	1	102,242	246,543
MUX	258	0	6,881	17,374	34,924
OAI	134	0	7,866	108,614	242,132
XOR/XNOR	0	0	0	0	0

Table 3.3: Number and type of cells for each benchmark circuit operating at 500 MHz.

synthesized using *Synopsys Design Compiler* [62] at 500 MHz (relaxed timing constraint with no timing violations) and 1.5/2 GHz (tighter constraint with negative slack) clock frequencies. Note that for the relatively small benchmarks SIMON (lightweight encryption core) and s38584 (academic benchmark), the high-frequency constraint is 2 GHz whereas, for larger FFT cores (64-, 128-, and 256-point [17]), the high-frequency constraint is 1.5 GHz. The synthesized netlists are placed (at 70% placement density) and routed using *Cadence Encounter* [63]. The overall number of gates and the number of each cell are listed in Tables 3.3 and 3.4, for, respectively, 500 MHz and 1.5/2 GHz. According to these tables, those cells that achieve an above-average reduction in the area are typically used more than the other cells by the synthesis process, thereby maximizing the reduction in system-level footprint.

Circuit	SIMON	s38584	FFT64	FFT128	FFT256
No. Gates	3,456	35,804	96,123	770,370	1,617,784
DFF	392	1,426	18,963	96,746	190,025
INV	294	7,653	23,992	79,375	254,358
NAND	648	2,531	15,056	220,357	213,765
NOR	509	2,101	0	62,881	187,063
AND	650	15,057	19,258	322,196	218,183
OR	547	7,036	3,502	83,353	243,049
MUX	33	0	6,882	46,588	66,489
OAI	383	0	8,470	108,847	244,852
XOR/XNOR	0	0	0	0	0

Table 3.4: Number and type of cells for each benchmark circuit operating at 1.5/2 GHz.

3.2.2.1 Footprint and Wirelength

The comparison of footprint and overall wirelength in 2D and 3D designs are listed in Table 3.5 and Table 3.6, respectively.

As an example, the layout views of the 2D and 3D versions of the 128-point FFT core are depicted in Figure 3.7 [64], illustrating the effect of the number of tracks on the chip-level footprint.

According to Table 3.5 and Table 3.6, benchmark circuits developed with transistor-level monolithic 3D libraries consume, on average, 32.4% less area as compared to conventional 2D designs.

3.2.2.2 Power Characteristics

The power consumption of 2D and monolithic 3D designs is compared in Table 3.7 and Table 3.8.

All of the three components of power consumption (internal, switching, and

Operating Frequency		500 MHz			
Circuit	Design Style	Footprint (mm^2)	Change (%)	Wirelength (μm)	Change (%)
SIMON	2D	0.0110	-	30,260	-
	3D	0.0073	**-33.6**	24,027	**-20.6**
s38584	2D	0.077	-	174,114	-
	3D	0.051	**-34.2**	142,442	**-18.2**
FFT64	2D	0.45	-	965,796	-
	3D	0.30	**-33.3**	771,457	**-20.1**
FFT128	2D	2.54	-	12,205,011	-
	3D	1.59	**-37.2**	9,240,148	**-24.3**
FFT256	2D	5.72	-	40,787,944	-
	3D	4.36	**-23.7**	33,404,062	**-18.1**

Table 3.5: Comparison of footprint, wirelength, and the number of DRC violations (vios) in 2D and monolithic 3D technologies with 9 (*Mono3D*) routing tracks in each 3D cell, operating at 500 MHz. The percent changes with respect to 2D cells are listed.

Operating Frequency		1.5/2 GHz				
Circuit	Design Style	Footprint (mm^2)	Change (%)	Wirelength (μm)	Change (%)	DRC vio
SIMON	2D	0.0122	-	44,242	-	0
	3D	0.0084	**-31.1**	26,449	**-40.2**	2
s38584	2D	0.079	-	203,703	-	0
	3D	0.051	**-35.4**	164,039	**-19.5**	53
FFT64	2D	0.59	-	1,202,699	-	0
	3D	0.42	**-28.8**	975,095	**-18.9**	11
FFT128	2D	2.94	-	15,201,864	-	0
	3D	1.84	**-37.5**	11,407,021	**-24.9**	7
FFT256	2D	5.88	-	39,094,466	-	0
	3D	4.58	**-22.1**	33,983,460	**-13.1**	29

Table 3.6: Comparison of footprint, wirelength, and the number of DRC violations (vios) in 2D and monolithic 3D technologies with 9 (*Mono3D*) routing tracks in each 3D cell, operating at 1.5/2 GHz. The percent changes with respect to 2D cells are listed.

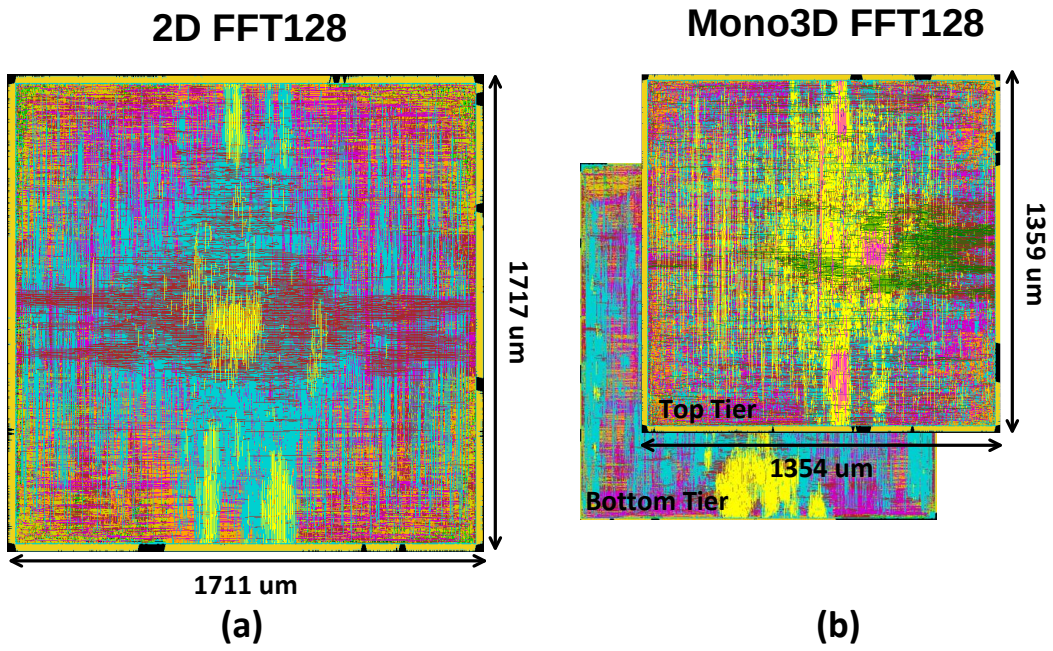


Figure 3.7: The layout views of a highly parallelized 128-point FFT core in (a) conventional 2D technology with 14 routing tracks in each cell, (b) transistor-level monolithic 3D technology.

Operating Frequency		500 MHz			
Circuit	Design Style	Power component (mW)			
		INT	SWI (Change)	LK	Total (Change)
SIMON	2D	8.67	4.48 (-)	0.571	13.73 (-)
	3D	9.88	2.82 **(-37%)**	0.548	13.25 **(-3.49%)**
s38584	2D	55.04	13.80 (-)	2.472	71.31 (-)
	3D	51.43	11.24 **(-19%)**	2.558	65.23 **(-8.52%)**
FFT64	2D	352	160.8 (-)	17.4	530 (-)
	3D	356	125.5 **(-21.95%)**	17.0	498 **(-6.04%)**
FFT128	2D	2,365	924.8 (-)	119.5	3,510 (-)
	3D	2,309	726.0 **(-21.50%)**	118.8	3,154 **(-10.14%)**
FFT256	2D	4,852	1,956 (-)	252.2	7,060 (-)
	3D	5,176	1,571 **(-19.7%)**	254.3	7,001 **(-0.83%)**

Table 3.7: Comparison of power consumption in 2D and monolithic 3D technologies 9 routing tracks in each cell, operating at 500 MHz. *INT, SWI*, and *LK* refer, respectively, to internal, switching (net), and leakage power.

Operating Frequency		2 GHz / 1.5 GHz			
Circuit	Design Style	Power component (mW)			
		INT	SWI (Change)	LK	Total (Change)
SIMON	2D	34.15	14.61 (-)	0.546	49.3 (-)
	3D	34.72	11.54 **(-21%)**	0.537	46.8 **(-5.07%)**
s38584	2D	223.4	58.16 (-)	2.822	284.4 (-)
	3D	214.0	47.01 **(-19%)**	2.870	263.9 **(-7.21%)**
FFT64	2D	713	324.7 (-)	17.7	1,056 (-)
	3D	720	252.8 **(-22.14%)**	17.8	990 **(-6.25%)**
FFT128	2D	7,891	2,859 (-)	144.9	10,895 (-)
	3D	6,863	2,309 **(-19.24%)**	145.9	9,318 **(-14.47%)**
FFT256	2D	9,674	3,919 (-)	288.1	13,881 (-)
	3D	10,286	3,210 **(-18.09%)**	287.9	13,784 **(-0.69%)**

Table 3.8: Comparison of power consumption in 2D and monolithic 3D technologies with 9 routing tracks in each cell, operating at 1.5/2 GHz. *INT, SWI*, and *LK* refer, respectively, to internal, switching (net), and leakage power.

leakage) are provided. Internal power is consumed due to the intra-cell device and interconnect capacitances and short-circuit current during the switching activity of a cell. Switching power is absorbed by the inter-cell interconnect (net) capacitances. Due to the considerable reduction in overall wirelength in monolithic 3D designs, the switching power is reduced, on average, by 24% at 500 MHz.

Note that the change in internal power in 3D designs depends upon the specific circuit. For example, for some of the benchmarks, the internal power consumed by the 3D versions is slightly less (such as s38584 and FFT128) whereas, for some others (such as SIMON, FFT64, and FFT256), 3D versions consume somewhat more internal power than the 2D counterpart. This variation depends upon the number of times each cell is used in the circuit since the 3D cell power may increase or decrease depending upon the specific cell (see Table 3.2). For example, comparing the cell type and number of FFT128 and FFT256 (listed in Tables 3.3 and 3.4), FFT256

contains a significantly higher number of OR, NOR, and MUX gates. According to Table 3.2, the 3D versions of these gates consume more power as compared to the traditional 2D gates. Since the internal power is still the dominant power component in these benchmarks, this fluctuation significantly affects the overall power savings, despite a consistent and reasonable reduction in switching power in all of the benchmarks. For example, up to 10% (at 500 MHz) and 14% (at 1.5 GHz) reduction in overall power consumption is achieved for FFT128. For FFT256, however, the power reduction is only 0.8% (at 500 MHz) and 0.7% (at 1.5 GHz) due to an increase in the internal power component of the 3D versions.

3.2.2.3 Timing Characteristics

The timing characteristics of the 2D and monolithic 3D circuits are compared in Table 3.9 where the worst slack (WS), worst negative slack (WNS), total negative slack (TNS), and number of timing violations are listed at both 500 MHz (with no timing violations) and 1.5/2 GHz (with timing violations).

An important observation from Table 3.9 is that the timing characteristics of FFT256 are degraded when monolithic 3D circuits are considered. This degradation is due to 1) higher average cell delay for monolithic 3D technology and 2) routing congestion. However, some of the 3D benchmarks (SIMON, s38584, FFT64, and FFT128) outperform the 2D counterparts at both 500 MHz and 1.5/2 GHz operating frequencies. At 500 MHz, the positive slack increases. At 1.5/2 GHz, the WNS, TNS, and number of violations are reduced. Alternatively, for FFT256 (where the number of OR and NOR gates is significantly higher), the 3D versions cannot outperform the 2D version due to higher cell-level delays of 3D OR and 3D NOR gates.

Operating Frequency		500 MHz	2 GHz / 1.5 GHz		
Circuit	Design Style	WS (ns)	WNS (ns)	TNS (ns)	No. Violations
SIMON	2D	0.326	0.051	0	0
	3D	0.429	0.048	0	0
s38584	2D	0.760	-0.216	-10.98	252
	3D	0.767	-0.205	-5.459	169
FFT64	2D	0.561	-0.070	-151.799	1,107
	3D	0.606	-0.055	-125.390	1,792
FFT128	2D	0.21	-0.104	-516.100	8,097
	3D	0.23	-0.091	-302.582	6,221
FFT256	2D	0.145	-0.152	-1029.1	18,404
	3D	0.024	-0.202	-1322.4	19,431

Table 3.9: Comparison of timing characteristics in 2D and monolithic 3D technologies with 9 routing tracks in each cell. *WS*, *WNS*, and *TNS* refer, respectively, to worst slack, worst negative slack, and total negative slack.

According to the simulation results, for relatively low-performance applications with relaxed timing constraints, monolithic 3D technology can be leveraged to achieve the highest reduction in footprint (therefore cost) by developing highly dense 3D cell layouts. For high-performance applications with tighter timing constraints, however, interconnects and the routing process plays a significant role in system timing and power consumption. In this case, metal resources for routing (such as the number of tracks) should be carefully considered to alleviate routing congestion and prevent timing degradation at the expense of slightly reduced savings in footprint.

3.3 Summary

In this chapter, an open source transistor-level monolithic 3D cell library is developed and integrated into a digital design flow. The proposed library is used to investigate several essential characteristics of monolithic 3D ICs such as footprint, timing and power consumption at both relaxed and tight timing constraints. The results of a 128-point FFT core operating at 1.5 GHz demonstrate that the monolithic 3D technology can reduce the footprint and overall power consumption by, respectively, 38% and 14%.

The entire proposed library and related files for tool integration are publicly available to facilitate future research in some of the critical design aspects of monolithic 3D technology such as thermal integrity and design-for-test methodologies as well as manufacturing aspects such as the effects of tier-specific device characteristics on system-level performance.

Chapter 4

Routing Congestion in Monolithic 3D ICs

In the previous chapter, it was demonstrated that routing congestion is an important issue for monolithic 3D ICs due to the reduced footprint. In this chapter, multiple versions of a cell library for transistor-level monolithic 3D integration are developed. The effect of the number of routing tracks on area, power, and delay characteristics is investigated by developing three versions of the cell library with different cell heights. This analysis is important since routing congestion is one of the primary physical design issues in monolithic 3D ICs. The primary clock tree characteristics of monolithic 3D ICs are also discussed.

The rest of the chapter is organized as follows: Contributions of this work are summarized in Section 4.1. The details of the proposed open source cell libraries, characterization, and comparison with 2D cells are provided in Section 4.2 for each version. The effect of the number of routing tracks on congestion, power, and

timing characteristics is investigated in Section 4.3 by developing a large-scale 3D FFT core. Finally, the chapter is concluded in Section 4.4.

4.1 Contributions of This Work

An open source cell library for monolithic 3D ICs has recently been developed [65] and is publicly available [16]. In this work:

- Additional versions of the 3D monolithic cell library with different cell heights are developed and fully characterized.
- These additional libraries are utilized to investigate the effects of the number of tracks on chip-level routing congestion (e.g., number of interconnect related design rule check violations), power, and timing (worst negative and total negative slack) characteristics.
- It is demonstrated that an optimum number of routing tracks exists to minimize chip-level power and delay characteristics in monolithic 3D ICs.

4.2 Routing Congestion Aware Cell Library for Monolithic 3D ICs

In this work, the previously developed standard cell library for *Mono3D* is extended to consider the different number of routing tracks.

Currently, 20 standard cells exist in *Mono3D*, as listed in Table 3.1. Each cell has been developed with a full-custom design methodology using a cell stacking

technique. A specific track is allocated for intra-cell MIVs, which are distributed within the cell to minimize the interconnect length and reduce the cell height. Each cell within the 2D *NanGate* library has 14 routing tracks. Alternatively, in this study, three monolithic 3D cell libraries are developed with the different number of tracks: 8-track (*Mono3D_v1*), 9-track (*Mono3D_v2*), and 10-track (*Mono3D_v3*). The number of tracks plays an essiential role in chip-level routing congestion, a primary issue in monolithic 3D ICs. The cell heights in *Mono3D_v1*, *Mono3D_v2*, and *Mono3D_v3* are, respectively, 1.33 μm, 1.52 μm, and 1.71 μm. These cell heights are, respectively, 46%, 38%, and 31% shorter than the standard cell height (2.47 μm) in *NanGate* cell library [58].

The layout of a NAND cell is illustrated in Figure 4.1 in both 2D and 3D monolithic technologies with three different cell heights. Cell dimensions and the three MIVs are highlighted. The design flow and the modifications required for 3D monolithic technology are described in [65].

4.3 Experimental Results

Cell-level simulation results and comparison of 3D cells (in each version of the library) with 2D cells are provided in Section 4.3.1. A large circuit is analyzed in Section 4.3.2 to quantify the benefits of transistor-level 3D technology for more complex circuits where interconnects play a more dominant role in determining system performance and power consumption.

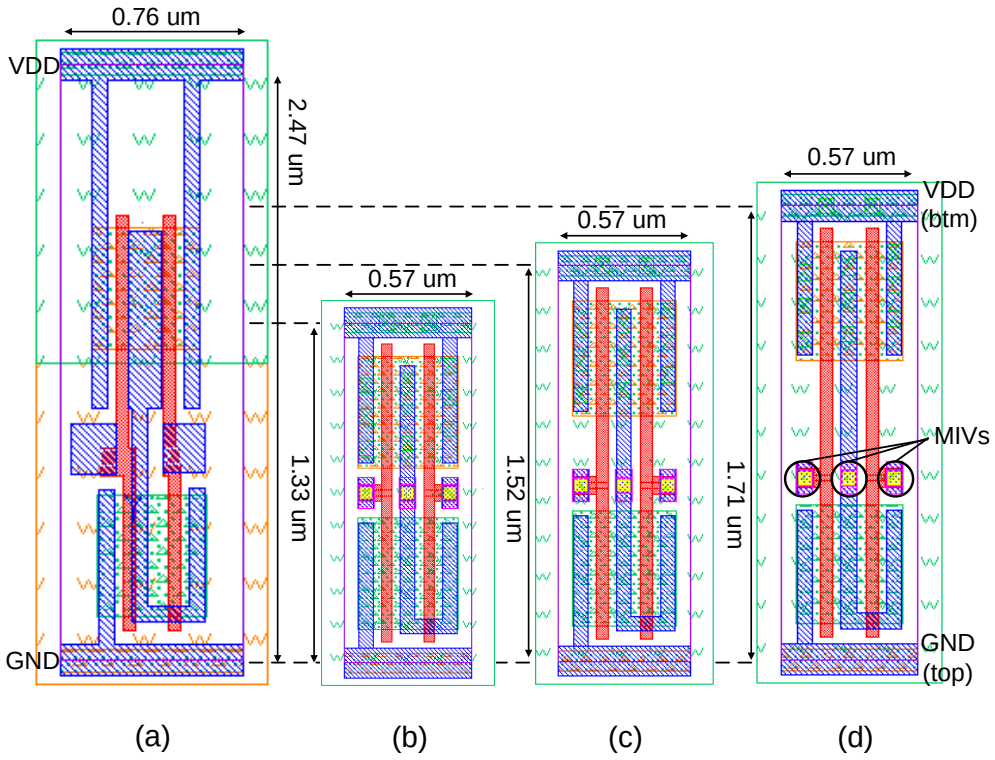


Figure 4.1: Comparison of the layout views of a NAND gate in (a) traditional 2D technology with 14 routing tracks in each cell, (b) monolithic 3D technology with 8 routing tracks, (c) monolithic 3D technology with 9 routing tracks, and (d) monolithic 3D technology with 10 routing tracks, illustrating the three MIVs used to connect the top and bottom tiers.

4.3.1 Cell-Level Evaluation

4.3.1.1 Area

As shown in Figure 4.2, cell-level area reduction varies from 6.5% to 64.1% in *Mono3D_v1*, -6.9% to 59.0% in *Mono3D_v2*, and -13.5% to 53.8% in *Mono3D_v3*, depending upon the specific cell. An average improvement of 32%, 22%, and 14% is achieved for, respectively, *Mono3D_v1*, *Mono3D_v2*, and *Mono3D_v3*. Note that a negative percent implies that the cell area increases as compared to the 2D cell. This behavior occurs for cells where the reduction in cell height causes a considerable increase in cell width. Similarly, the average area reduction is not as large as the reduction in cell height since, on average, the cell width slightly increases due to MIVs and intra-cell routing within the reduced cell footprint.

4.3.1.2 Delay and Power Consumption

As mentioned in Chapter 3, *HSPICE* simulations are performed on the extracted 3D netlists to compare monolithic 3D technology with the conventional 2D technology at the cell level. The power supply voltage is set to be 1.1 V, the transition time is 50 ps, and the temperature is set to be 27°C. We analyzed the average delay and power consumption, as listed in Table 4.1 and Table 4.2 for 2D and each version of the 3D technology. According to this table, *Mono3D_v1* cells have, on average, 2.15% (3.22% in *Mono3D_v2* and 3.78% in *Mono3D_v3*) higher propagation delay and 0.93% (0.46% in *Mono3D_v2* and 0.08% in *Mono3D_v3*) lower power consumption as compared to the 2D standard cells. This slight increase in delay is due to denser cell layout, producing additional coupling capacitances and MIV impedances.

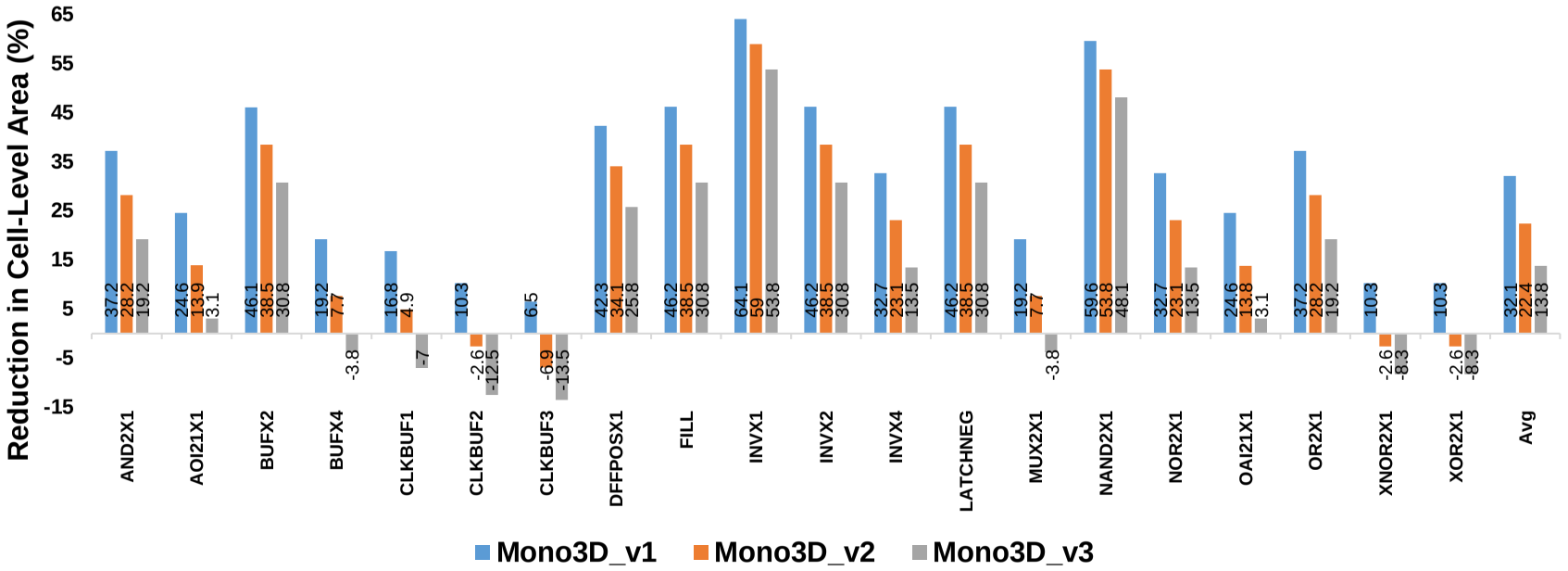


Figure 4.2: Percent reduction in area achieved by each individual monolithic 3D cell as compared to the 2D cells. Results are provided for each 3D library, *Mono3D_v1*, *Mono3D_v2*, and *Mono3D_v3*.

Cells	**Delay (ps)**			
	2D	3D_v1	3D_v2	3D_v3
AND2X1	17.63	19.27 (9.3%)	19.5 (10.6%)	19.52 (10.7%)
AOI21X1	13.68	13.58 (-0.7%)	13.69 (0.1%)	13.80 (0.9%)
BUFX2	17.89	17.56 (-1.8%)	17.86 (-0.2%)	17.94 (0.3%)
BUFX4	15.82	14.97 (-5.4%)	15.29 (-3.4%)	15.76 (-0.4%)
CLKBUF1	27.01	27.28 (1.0%)	27.32 (1.2%)	27.34 (1.2%)
CLKBUF2	39.57	40.05 (1.2%)	40.17 (1.5%)	40.24 (1.7%)
CLKBUF3	51.73	52.74 (2.0%)	53.04 (2.5%)	53.27 (3.0%)
DFFPOSX1	41.69	34.54 (-17.2%)	34.72 (-16.7%)	34.93 (-16.2%)
INVX1	6.73	6.10 (-9.4%)	6.42 (-4.6%)	6.50 (-3.4%)
INVX2	6.54	6.08 (-7.0%)	6.32 (-3.4%)	6.40 (-2.1%)
INVX4	6.44	6.08 (-5.6%)	6.29 (-2.3%)	6.38 (-0.9%)
MUX2X1	16.25	17.21 (5.9%)	17.23 (6.0%)	17.27 (6.2%)
NAND2X1	10.22	9.76 (-4.5%)	9.78 (-4.3%)	9.89 (-3.2%)
NOR2X1	11.41	11.78 (3.2%)	12.16 (6.6%)	12.18 (6.8%)
OAI21X1	12.89	12.72 (-1.3%)	12.88 (-0.1%)	12.93 (0.3%)
OR2X1	18.33	20.89 (14.0%)	21.12 (15.2%)	21.14 (15.3%)
XNOR2X1	36.05	41.32 (14.6%)	41.76 (15.8%)	41.92 (16.3%)
XOR2X1	35.59	41.95 (17.9%)	42.41 (19.2%)	42.69 (19.9%)
Average	21.42	21.88 (2.15%)	22.11 (3.22%)	22.23 (3.78%)

Table 4.1: Average delay characteristics of 2D and monolithic 3D cells with 8 (*Mono3D_v1*), 9 (*Mono3D_v2*), and 10 (*Mono3D_v3*) routing tracks. The percent changes with respect to 2D cells are listed.

Cells	**Power (μW)**			
	2D	3D_v1	3D_v2	3D_v3
AND2X1	2.82	2.98 (5.7%)	3.01 (6.7%)	3.03 (7.4%)
AOI21X1	3.32	3.33 (0.3%)	3.34 (0.6%)	3.35 (0.9%)
BUFX2	14.04	13.71 (-2.4%)	13.92 (-0.9%)	13.97 (-0.5%)
BUFX4	29.00	28.97 (-0.1%)	28.98 (-0.1%)	29.14 (0.5%)
CLKBUF1	64.07	62.17 (-3.0%)	62.52 (-2.4%)	62.84 (-1.9%)
CLKBUF2	93.25	90.43 (-3.0%)	90.88 (-2.5%)	91.05 (-2.4%)
CLKBUF3	121.4	118.6 (-2.3%)	119.0 (-2.0%)	119.1 (-1.9%)
DFFPOSX1	26.75	27.13 (1.4%)	27.18 (1.6%)	27.39 (2.4%)
INVX1	4.69	4.64 (-1.1%)	4.68 (-0.2%)	4.72 (0.6%)
INVX2	9.31	9.15 (-1.7%)	9.24 (-0.8%)	9.35 (0.4%)
INVX4	18.01	17.99 (-0.1%)	18.16 (0.8%)	18.36 (1.9%)
MUX2X1	5.81	6.14 (5.7%)	6.15 (5.9%)	6.17 (6.2%)
NAND2X1	1.63	1.57 (-3.7%)	1.58 (-3.1%)	1.61 (-1.2%)
NOR2X1	1.63	1.66 (1.8%)	1.69 (3.7%)	1.73 (6.1%)
OAI21X1	3.27	3.24 (-0.9%)	3.25 (-0.6%)	3.26 (-0.3%)
OR2X1	2.54	2.84 (11.8%)	2.85 (12.2%)	2.87 (12.9%)
XNOR2X1	12.66	14.12 (11.5%)	14.13 (11.6%)	14.18 (12.0%)
XOR2X1	12.53	14.16 (13.0%)	14.24 (13.6%)	14.28 (13.9%)
Average	23.71	23.49 (-0.93%)	23.60 (-0.46%)	23.69 (-0.08%)

Table 4.2: Average power characteristics of 2D and monolithic 3D cells with 8 (*Mono3D_v1*), 9 (*Mono3D_v2*), and 10 (*Mono3D_v3*) routing tracks. The percent changes with respect to 2D cells are listed.

Circuit	Freq (GHz)	Design Style	Area (mm^2)	Wirelength (m)	DRC Vios
FFT128	0.5	2D	2.54	12.2	-
		3D_v1	1.55 **(-39.1)**	9.4 **(-23)**	-
		3D_v2	1.59 **(-37.2)**	9.2 **(-24)**	-
		3D_v3	1.77 **(-30.3)**	10.2 **(-16)**	-
	1.5	2D	2.94	15.2	0
		3D_v1	1.76 **(-40.2)**	11.6 **(-24)**	568
		3D_v2	1.84 **(-37.5)**	11.4 **(-25)**	7
		3D_v3	2.04 **(-30.6)**	13.0 **(-15)**	0

Table 4.3: Comparison of the area, wirelength, and number of DRC violations in 2D FFT and monolithic 3D FFT with 8 (*Mono3D_v1*), 9 (*Mono3D_v2*), and 10 (*Mono3D_v3*) routing tracks in each cell. *DRC vios* refers to the number of design rule check (DRC) violations.

4.3.2 System-Level Evaluation

The proposed open cell libraries are used to develop a parallel 128-point FFT core with approximately 330K cells, operating at 500 MHz (relaxed timing constraint) and 1.5 GHz (tighter timing constraint). The layout views of the 2D and 3D versions of the FFT core are depicted in Figure 4.3, illustrating the effect of the number of tracks on the chip-level footprint. Specifically, as compared to the 2D version, the footprint and overall wirelength are reduced, respectively, by 40% and 24% in *Mono3D_v1*, 38% and 25% in *Mono3D_v2*, and 31% and 15% in *Mono3D_v3*, when the operating frequency is 1.5 GHz. These results are listed in Table 4.3.

At 500 MHz, no DRC violations are reported. At 1.5 GHz, however, *Mono3D_v1* exhibits approximately 600 violations due to routing congestion. These violations are reduced to 9 for *Mono3D_v2* and are completely eliminated for *Mono3D_v3*.

At 1.5 GHz, the FFT core implemented with *Mono3D_v2* library consumes ap-

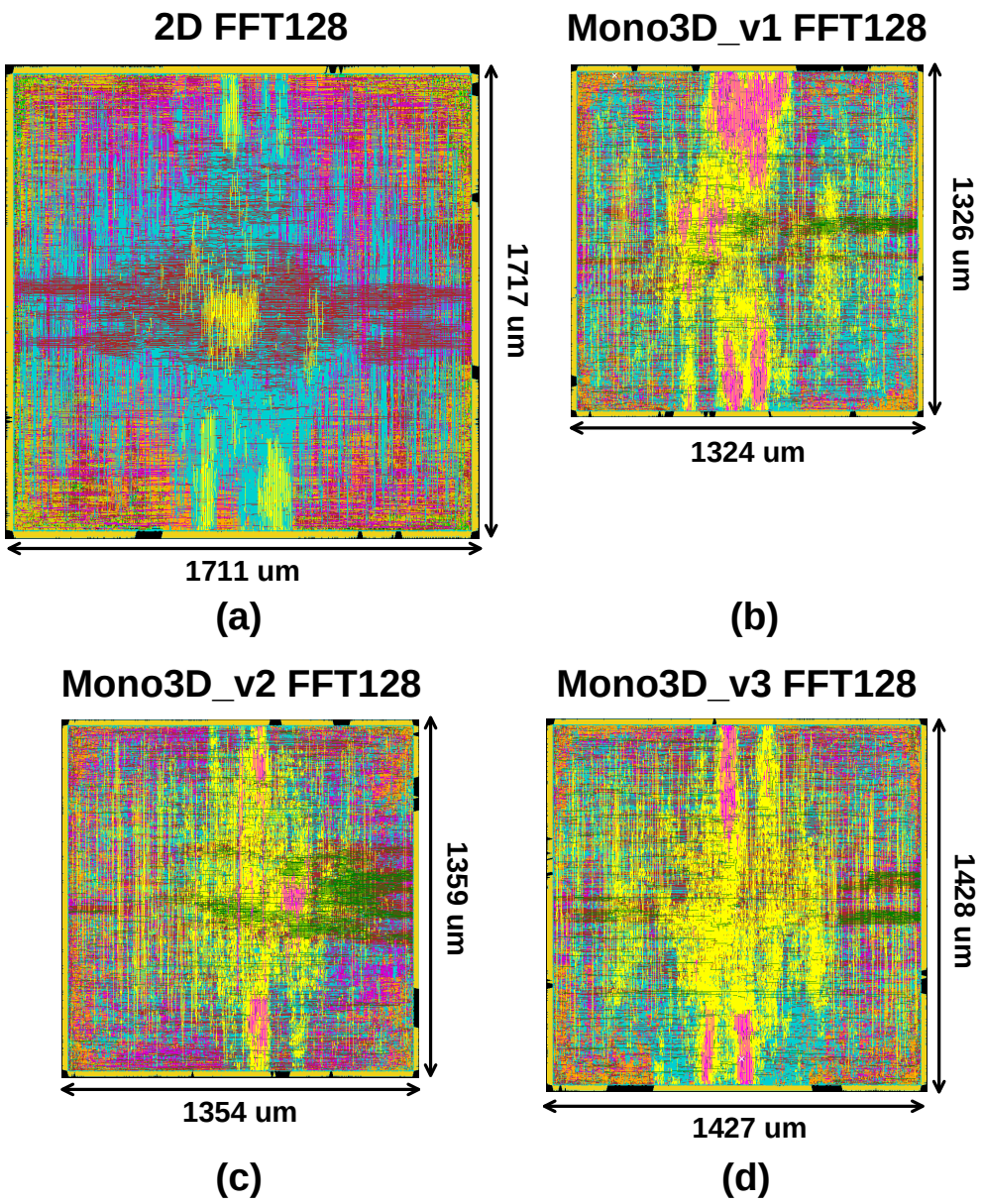


Figure 4.3: The layout views of a highly parallelized 128-point FFT core in (a) conventional 2D technology with 14 routing tracks in each cell, (b) transistor-level monolithic 3D technology with 8 routing tracks, (c) monolithic 3D technology with 9 routing tracks, and (d) monolithic 3D technology with 10 routing tracks.

Operating Frequency		500 MHz			
Circuit	Design Style	Power component (mW)			
		INT	SWI (Change)	LK	Total (Change)
FFT128	2D	2,365	924.8 (-)	119.5	3,510 (-)
	3D_v1	2,340	750.3 **(-18.87%)**	119.4	3,210 **(-8.55%)**
	3D_v2	2,309	726.0 **(-21.50%)**	118.8	3,154 **(-10.14%)**
	3D_v3	2,302	749.7 **(-18.93%)**	119.9	3,172 **(-9.63%)**

Table 4.4: Comparison of power consumption in 2D FFT and monolithic 3D FFT with 8 (*Mono3D_v1*), 9 (*Mono3D_v2*), and 10 (*Mono3D_v3*) routing tracks in each cell, operating at 500 MHz. *INT, SWI*, and *LK* refer, respectively, to internal, switching (net), and leakage power.

Operating Frequency		1.5 GHz			
Circuit	Design Style	Power component (mW)			
		INT	SWI (Change)	LK	Total (Change)
FFT128	2D	7,891	2,859 (-)	144.9	10,895 (-)
	3D_v1	6,936	2,351 **(-17.77%)**	146.5	9,437 **(-13.38%)**
	3D_v2	6,863	2,309 **(-19.24%)**	145.9	9,318 **(-14.47%)**
	3D_v3	6,884	2,333 **(-18.40%)**	145.3	9,363 **(-14.06%)**

Table 4.5: Comparison of power consumption in 2D FFT and monolithic 3D FFT with 8 (*Mono3D_v1*), 9 (*Mono3D_v2*), and 10 (*Mono3D_v3*) routing tracks in each cell, operating at 1.5 GHz. *INT, SWI*, and *LK* refer, respectively, to internal, switching (net), and leakage power.

Operating Frequency		500 MHz	1.5 GHz		
Circuit	Design Style	WS (ns)	WNS (ns)	TNS (ns)	NO. Violations
FFT128	2D	0.021	-0.104	-516.100	8,097
	3D_v1	0.016	-0.116	-725.474	9,118
	3D_v2	0.023	-0.091	-302.582	6,221
	3D_v3	0.022	-0.097	-319.003	6,574

Table 4.6: Comparison of timing characteristics in 2D FFT and monolithic 3D FFT with 8 (*Mono3D_v1*), 9 (*Mono3D_v2*), and 10 (*Mono3D_v3*) routing tracks in each cell. *WS, WNS* and *TNS* refer, respectively, to worst slack, worst negative slack and total negative slack.

proximately 20% less interconnect power as compared to 2D technology. The internal power is also reduced by approximately 13%, partly due to the type of cells used in the design and partly due to the reduction in short-circuit power (since the interconnect lengths are shorter and signal transitions are faster). Overall, the monolithic 3D technology achieves approximately a 15% reduction in power, as listed in Table 4.4 and Table 4.5.

The timing characteristics of the 2D and monolithic 3D circuits are compared in Table 4.6 where the worst negative slack (WNS), total negative slack (TNS), and a number of timing violations are listed. The target clock frequencies are 500 MHz and 1.5 GHz. An important observation from Table 4.6 is that the timing characteristics are degraded when 8 routing tracks (*Mono3D_v1*) are considered. This degradation is due to 1) higher average cell delay for monolithic 3D technology and 2) routing congestion. For each of the three benchmarks, all WS, WNS and TNS increase. However, if the number of routing tracks in each cell is increased to 9 (*Mono3D_v2*), the timing characteristics of all of the 3D benchmarks outperform 2D designs at both 500 MHz and 1.5 GHz operating frequencies. At 500 MHz,

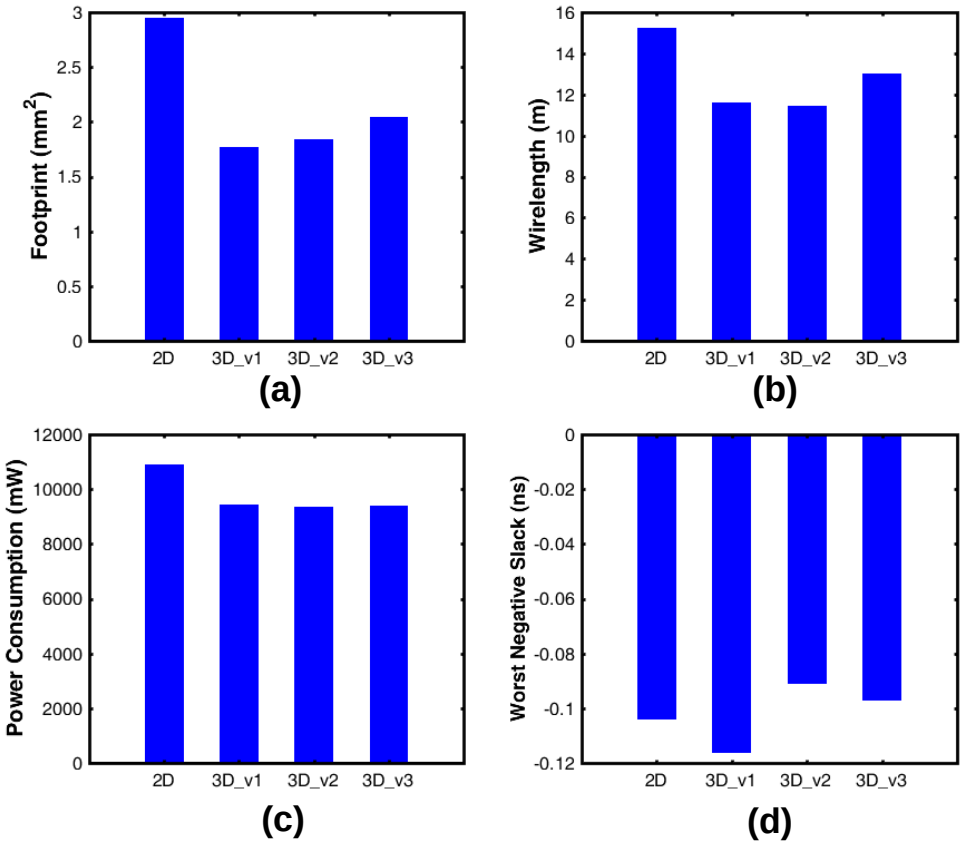


Figure 4.4: Summary of results obtained from the FFT128 operating at 1.5 GHz: (a) footprint, (b) wirelength, (c) power consumption, and (d) worst negative slack.

the positive slack increases. At 1.5 GHz, the WNS, TNS, and number of violations are reduced. Thus, similar to power characteristics, *Mono3D_v2* achieves the best timing characteristics.

The simulation results obtained from FFT128 are summarized in Figure 4.4 when the clock frequency is 1.5 GHz.

Since clock networks play a significant role in both performance and power in large-scale circuits, the clock tree synthesis (CTS) results of the FFT core are also

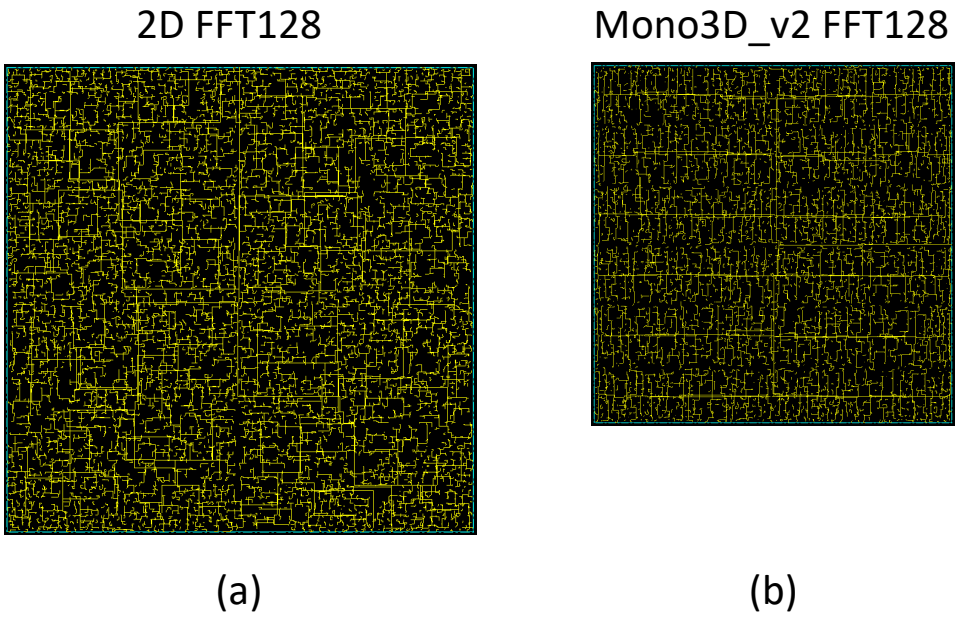


Figure 4.5: Clock tree floorplan of a 128-point FFT core with approximately 97K flip-flops: (a) traditional 2D technology and (b) monolithic 3D technology with 9 routing tracks in each cell (*Mono3D_v2*).

reported to quantify the benefits of monolithic 3D technology in clocking. The clock trees obtained by *Encounter* for both 2D and 3D technologies (*Mono3D_v2*) are shown in Figure 4.5. The number of sinks for both designs is 96,796. Both the skew and slew constraints are set to 100 ps. The smaller 3D footprint is helpful for enhancing primary clocking characteristics, as listed in Table 4.7. Due to the reduced footprint, the number of clock buffers is reduced from 8,231 to 6,427, which reduces the clock internal power by approximately 28%. The clock wirelength is also minimized by 28% and the clock net power is reduced by about 27%. The overall clock power is reduced by 28%.

The clock tree of the 2D design exhibits slew violations, which are fixed in the 3D clock network (due to shorter and therefore less resistive clock nets). The

Circuit	128-point FFT core	
Design style	2D	3D_v2
Number of sinks	96,796	96,796
Number of buffers	8,231	6,427
Clock wirelength (mm)	732	527
Clock cap (pF)	1,052	719
Max. buffer slew (ps)	110.1	98.7
Max. sink slew (ps)	107.8	95.1
Skew (ps)	51.7	45.5
Clock internal power (mW)	2,419	1,738
Clock switching power (mW)	2,430	1,767
Clock leakage power (mW)	9.2	7.3
Overall clock power (mW)	4,858	3,512

Table 4.7: Comparison of primary clock tree characteristics of 2D FFT and monolithic 3D FFT.

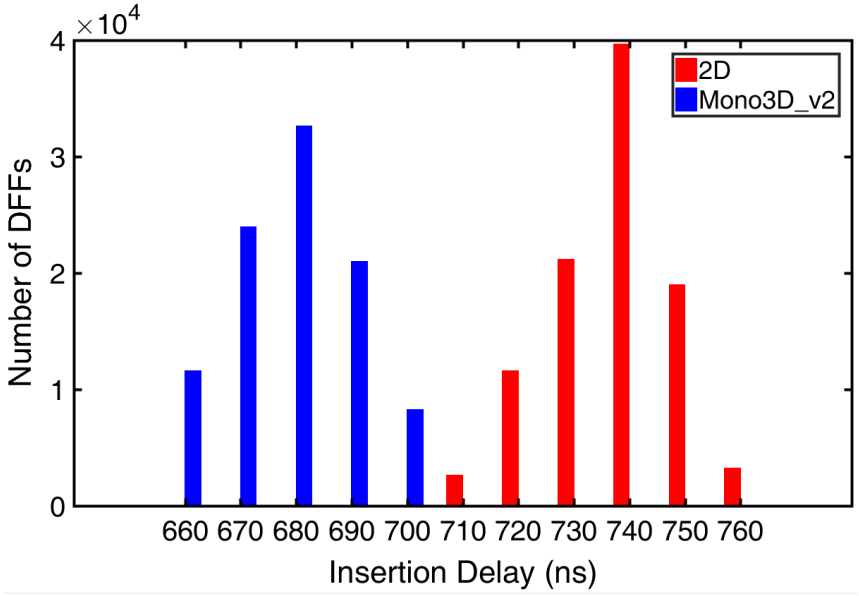


Figure 4.6: Comparison of clock insertion delay histograms for 2D FFT and monolithic 3D FFT.

global skew decreases from 51.7 ps in 2D FFT core to 45.5 ps in 3D FFT core, implemented by *Mono3D_v2*. The 3D FFT core also exhibits lower clock insertion delays, as shown in Figure 4.6 where the insertion delay histograms are compared for 2D and 3D designs. Lower insertion delays are helpful in reducing the variation-induced skew or corner-to-corner skew variation.

4.4 Summary

In this chapter, several open source transistor-level monolithic 3D cell libraries are developed and integrated into a digital design flow. The proposed libraries are used to investigate several essential characteristics of monolithic 3D ICs such as 1) footprint, timing and power consumption at both relaxed and tight timing constraints, 2) routing congestion, 3) the effect of the number of routing tracks in each cell, 4) clock tree. The results of a large-scale FFT core operating at 1.5 GHz demonstrate that the monolithic 3D technology can reduce the footprint and overall power consumption by, respectively, 38% and 14%. The effect of routing congestion on timing characteristics is stronger in monolithic 3D technology, where the cell-level number of routing tracks plays an important role. An optimum number of routing tracks exists that achieves the most significant improvements in both power and timing characteristics.

Chapter 5

Hardware-Efficient Logic Camouflaging for Monolithic 3D ICs

Circuit camouflaging is a layout-level technique to thwart image analysis based reverse engineering attacks. An efficient dummy contact based camouflaging method for monolithic 3D ICs is proposed. 3D ICs achieve ultra-high density device integration enabled by fine-grained monolithic inter-tier vias. Standard cell libraries are developed to evaluate the effects of circuit camouflaging on large-scale 2D and monolithic 3D ICs. These libraries are used to design a camouflaged SIMON (lightweight block cipher) and several academic benchmarks. Simulation results demonstrate that the monolithic 3D technology is highly useful to facilitate the utilization of camouflaging technique against reverse engineering attacks. At the expense of a slight degradation in timing characteristics, monolithic 3D technology eliminates not only the area but also the power overhead related to camouflage.

The rest of the chapter is organized as follows: A brief background and con-

tributions of this work are summarized in Section 5.1. Proposed camouflaged 2D and monolithic 3D integrated circuit cell libraries are provided in Section 5.2. Simulation results for SIMON block cipher and several academic benchmarks are presented in Section 5.3. The chapter is summarized in Section 5.4.

5.1 Background and Contributions

The security of ICs has emerged as a fundamental issue due to the threats from the globalized semiconductor supply chain [66–70]. Circuit camouflaging is an effective technique to thwart reverse engineering attacks that try to recover the original netlist through scanning electron microscopy (SEM) images [71–74]. The circuit obfuscation level achieved by the camouflaging technique, however, depends upon the number and location of the camouflaged gates [72]. Thus, these parameters play an important role in achieving the desired attack resilience. A larger number of camouflaged gates strengthens the countermeasure at the expense of significant overhead in the area, power, and delay characteristics [72].

Recently, monolithic 3D technology [13,14,43] has been conceptually proposed as a countermeasure against reverse engineering since it has the potential to reduce the overhead of traditional circuit camouflaging [75]. Transistor-level camouflaged logic locking method for monolithic 3D integrated circuit security has also been proposed [76]. These studies follow the highly encouraging recent developments on monolithic 3D technology that relies on sequentially fabricating multiple transistor layers [52]. Monolithic 3D ICs enable not only ultra-high density device integration but also introduce novel opportunities and challenges on managing hardware security [75–79]. For example, existing split manufacturing techniques developed

primarily for through silicon via and interposed based vertical integration are not applicable to monolithic 3D ICs. Unlike TSV based 3D ICs, in monolithic 3D ICs, all of the tiers are manufactured sequentially by the same foundry. Thus, splitting the system functionality into multiple tiers is not useful to protect monolithic 3D ICs from reverse engineering and hardware intellectual property (IP) piracy attacks from untrusted foundries.

5.1.1 Circuit Camouflaging

Circuit camouflaging is a method to obfuscate logic function by making subtle changes to the physical layout of standard cells [71–73]. The primary goal of camouflage is to disguise the circuit against a reverse engineer who utilizes SEM pictures to recover the original chip design. For example, from the SEM image analysis, a camouflaged logic cell appears to be a 2-input NAND gate. In practice, however, that cell can be a 2-input NOR gate. For example, NAND and NOR cells are designed to look identical where the actual function depends upon the real and dummy contacts. The camouflaged 2D NAND and NOR gates with both dummy and real contacts/vias are shown, respectively, in Figures 5.1 (c) and (d) [80]. As a reference, the non-camouflaged NAND and NOR gates are also illustrated in Figures 5.1 (a) and (b), respectively. This wrong perception can be achieved by small changes in metal contacts and vias [81]. As shown in Figure 5.2, the contacts between metal 1 (M1) and the polysilicon layers are physically connected on the left cell, but disconnected after a thin layer of contact material on the right cell. The top views of M1 (and polysilicon) for the gates with or without camouflaging are identical. The same technique can be applied to vias between metal layers. As demonstrated in [71], the conventional and camouflaged 2-input AND gates have

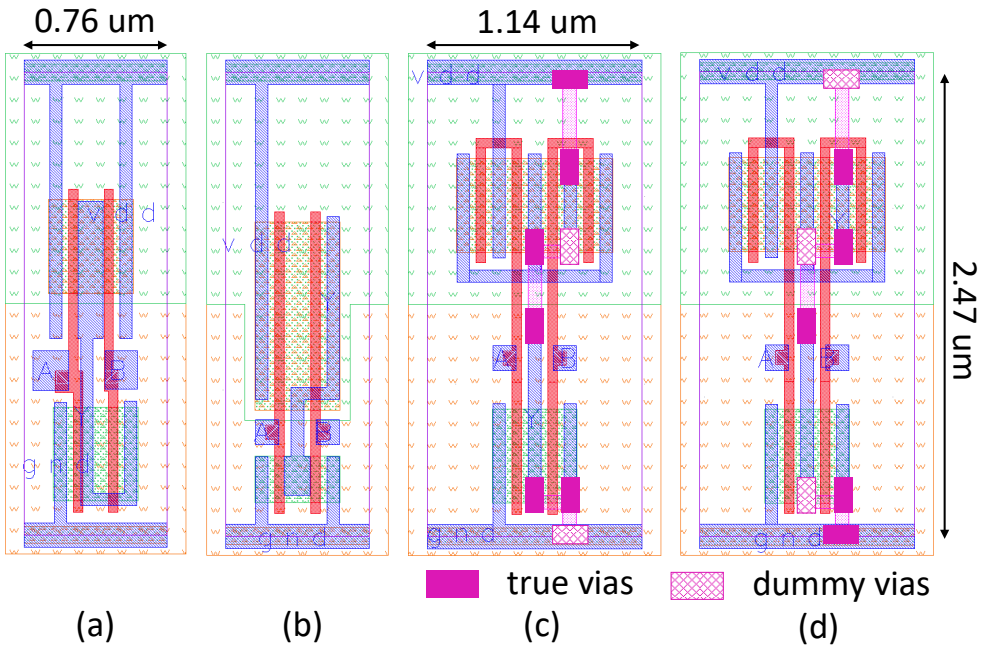


Figure 5.1: Standard cell layouts in 2D technology: (a) conventional NAND, (b) conventional NOR, (c) camouflaged NAND, and (d) camouflaged NOR.

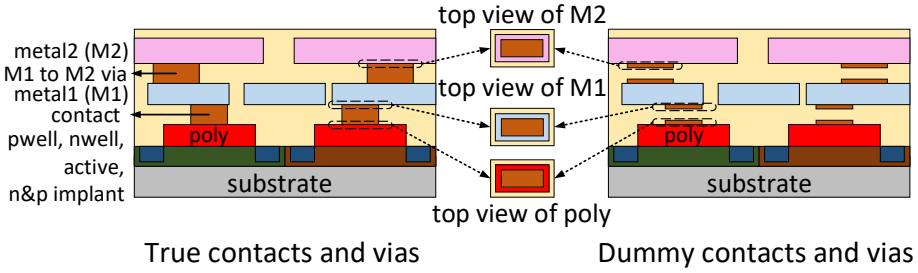


Figure 5.2: Illustration of true and dummy contacts/vias.

the same SEM image. Thus, an attacker cannot entirely rely on SEM image analysis to successfully extract the correct circuit netlist. Since reverse engineers cannot partially etch a layer [74], circuit camouflaging with dummy contacts/vias has become an effective method to obscure the original circuit.

5.1.2 Contributions of This Work

The primary contributions of this work are as follows:

- Camouflaged monolithic 3D cells are developed and fully characterized by using cell stacking method.
- A chip-level analysis is performed on fully placed and routed 2D and monolithic 3D circuits to quantify and compare the area, power, and delay overhead of camouflaging technique. An important observation is that the cell-level power overhead (that is typically reported in existing work) is compensated by the reduction in chip-level interconnect power for monolithic 3D technology.
- Advantages and limitations of circuit camouflaging for 3D monolithic technology are discussed. To the best of the authors' knowledge, this study is the first work that quantitatively investigates the overhead of circuit camouflaging on monolithic 3D ICs.

5.2 Logic Camouflaging for Monolithic 3D ICs

The characteristics of the standard cell libraries for proposed camouflaged methods of both the conventional 2D and monolithic 3D ICs are described in Chap-

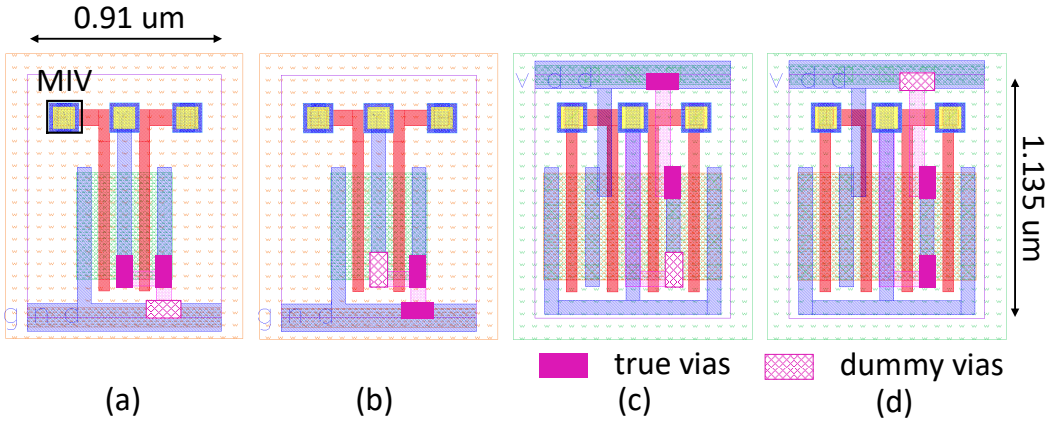


Figure 5.3: Camouflaged cell layouts in monolithic 3D technology: (a) top tier of NAND gate, (b) top tier of NOR gate, (c) bottom tier of NAND gate, (d) bottom tier of NOR gate.

ter 5.2.1. Cell-level simulation results and comparisons among regular 2D cells, camouflaged 2D cells, and camouflaged monolithic 3D cells are provided in Chapter 5.2.2.

5.2.1 Camouflaged Cells in 2D and Monolithic 3D Technologies

Two camouflaged standard cell libraries are developed. The first one is for conventional 2D technology whereas the second one is for monolithic 3D technology with inter-tier vias. Both of these libraries are generated based on the 2D 45 nm process design kit *FreePDK45* [56]. Thus, the process and physical characteristics such as transistor models and on-chip metal layers are obtained from *FreePDK45*.

As shown in Figure 3.1, in the transistor-level 3D monolithic technology, there are two tiers where the top tier is used for nMOS transistors whereas the pMOS transistors are placed within the bottom tier, similar to [55].

Regular Standard Cells		Camouflaged Standard Cells
INVX1	INVX2	NAND2X1 & NOR2X1
CLKBUF1	CLKBUF2	AND2X1 & OR2X1
DFFPOSX1	FILL	XNOR2X1 & XOR2X1

Table 5.1: List of standard cells in the camouflaged 2D and monolithic 3D libraries.

Twelve standard cells are developed in both 2D and 3D camouflaged cell libraries, as listed in Table 5.1. Of these cells, NAND, NOR, AND, OR, XOR, and XNOR are camouflaged. For example, NAND and NOR cells are designed to look identical where the actual function depends upon the real and dummy contacts. This behavior also holds for AND/OR and XOR/XNOR cell pairs.

In camouflaged monolithic 3D cells, the power rail is located at the top of the bottom tier and the ground rail is located at the bottom of the top tier. MIVs are distributed within the cell to minimize the interconnect distance and reduce the cell height, as shown in Figure 5.3, where the camouflaged 3D NAND and NOR gates are illustrated. Both the top [see Figures 5.3(a) and b)] and bottom tiers [see Figures 5.3(c) and d)] in each cell look identical from the top view.

In camouflaged 3D cells, the cell height is 1.135 μm, which is 54% smaller than the standard cell height (2.47 μm in *Nangate* 45 nm cell library [58]. The top tier metal layers and true/dummy contacts of these camouflaged cells are illustrated in Figure 5.4 for both NAND and NOR cells. Note that contrary to non-camouflaged cells that utilize only metal 1 for intra-cell routing, camouflaged cells require both metal 1 and metal 2 for routing, which affects both the cell-level (see Section 5.2.2) and chip-level (see Section 5.3) area, power and timing characteristics.

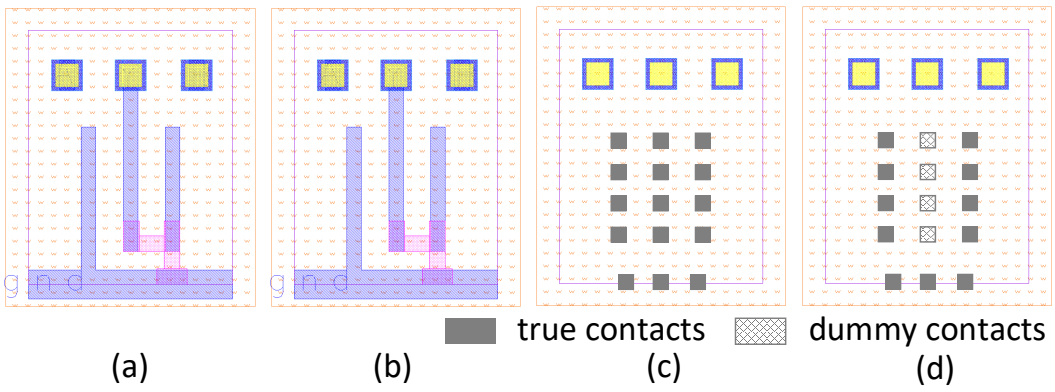


Figure 5.4: Metal layers and true/dummy contacts within the top tier of camouflaged monolithic 3D cells: (a) NAND metal layers, (b) NOR metal layers, (c) NAND contacts, (d) NOR contacts.

5.2.2 Cell-Level Evaluation

The effect of camouflaging on the cell-level area, delay, and power consumption is investigated for both 2D and 3D technologies. The results are listed in Table 5.2.

5.2.2.1 Footprint

For the 2D camouflaged cells, the increase in cell area varies from 0 to 50%, depending upon the cell type. For example, for XNOR and XOR gates, there is no overhead in the area since the transistors in both cells have the same sizes. Thus, it is not necessary to upsize the cells to make them look identical. Furthermore, the inherent cell area is sufficiently large, leaving sufficient space for intra-cell routing needed for camouflaging the cells.

For camouflaged monolithic 3D standard cells, the cell area is reduced as compared to non-camouflaged 2D cells due to the inherent advantage of monolithic 3D technology. This reduction in cell area varies from 34% to 51%. Despite more than

Std Cell	Design	Area (μm^2)	Delay (ps)	Power (μW)
NAND_2D	2D	1.88	7.61	1.28
	2D_C	2.82 (**50%**)	8.93 (**17%**)	1.82 (**42%**)
NAND_3D	3D	0.6 (**-68%**)	8.95 (**18%**)	1.45 (**13%**)
	3D_C	1.04 (**-45%**)	10.3 (**35%**)	1.97 (**54%**)
NOR_2D	2D	1.88	8.73	1.41
	2D_C	2.82 (**50%**)	8.05 (**-8%**)	1.80 (**28%**)
NOR_3D	3D	1.04 (**-45%**)	9.05 (**4%**)	1.42 (**1%**)
	3D_C	1.04 (**-45%**)	8.22 (**-6%**)	1.82 (**29%**)
AND_2D	2D	2.87	14.2	2.28
	2D_C	3.66 (**28%**)	17.0 (**20%**)	2.98 (**31%**)
AND_3D	3D	1.27 (**-58%**)	15.3 (**8%**)	2.32 (**2%**)
	3D_C	1.42 (**-51%**)	18.1 (**27%**)	2.99 (**31%**)
OR_2D	2D	2.87	15.4	2.26
	2D_C	3.66 (**28%**)	15.6 (**1%**)	2.75 (**22%**)
OR_3D	3D	1.42 (**-51%**)	16.9 (**10%**)	2.35 (**4%**)
	3D_C	1.42 (**-51%**)	17.0 (**10%**)	2.76 (**22%**)
XNOR_2D	2D	4.67	29.6	9.82
	2D_C	4.67 (**0%**)	31.2 (**5%**)	10.3 (**5%**)
XNOR_3D	3D	3.10 (**-34%**)	31.4 (**6%**)	10.2 (**4%**)
	3D_C	3.10 (**-34%**)	32.9 (**11%**)	10.5 (**7%**)
XOR_2D	2D	4.67	29.3	10.1
	2D_C	4.67 (**0%**)	30.6 (**4%**)	10.5 (**4%**)
XOR_3D	3D	3.10 (**-34%**)	31.7 (**8%**)	10.4 (**3%**)
	3D_C	3.10 (**-34%**)	32.5 (**11%**)	10.7 (**6%**)

Table 5.2: Area, average delay and power characteristics of conventional 2D, camouflaged 2D (2D_C), monolithic 3D, and camouflaged monolithic 3D (3D_C) standard cells. All of the percentages are with respect to conventional 2D results.

50% reduction in cell height, the average area reduction is less than 50% due to camouflaging overhead and MIVs. The area of the non-camouflaged 3D cells is also listed in the table as a reference.

5.2.2.2 Delay and Power Consumption

HSPICE simulations are performed on the extracted non-camouflaged 2D, 3D, and camouflaged 2D and 3D netlists to compare the cell-level power and delay characteristics at nominal operating conditions. Non-camouflaged 2D results are considered as the baseline for all of the percentages reported here. In general, the 2D camouflaged cells of this work have significantly less delay and power overhead as compared to [72]. As listed in Table 5.2, for the 2D camouflaged cells, the percent change in average propagation delay varies from -8% (for the NOR gate) to 20% (for the AND gate), while for the monolithic 3D technology, it differs from -6% (for the NOR gate) to 38% (for the NAND gate). Thus, except the NOR gate, camouflaging increases the delay in both 2D and 3D technologies due to additional interconnects and vias. For the camouflaged NOR gate, the size of the nMOS has been increased from 0.25 μm to 0.5 μm (since the size of each nMOS in the NAND gate is 0.5 μm due to series connection), thereby lowering the average propagation delay.

For the camouflaged 2D cells, the increase in power consumption varies from 4% (for the XOR gate) to 42% (for the NAND gate), while for the camouflaged monolithic 3D cells the power overhead is between 6% (for the XOR gate) and 54% (for the NAND gate). Note that the camouflaged monolithic 3D cells have higher propagation delay and higher power consumption as compared to camouflaged 2D standard cells.

According to Table 5.2, camouflaged 3D cells have, on average, 7.82% higher propagation delay and 2.33% higher power consumption as compared to the camouflaged 2D cells. This slight increase in the delay and power is due to the MIV impedances and the denser cell layout that produces additional coupling capacitances. Also, note that the camouflaged standard cells can be further optimized to reduce the delay and power at the expense of reduced improvement in the footprint. Thus, for monolithic 3D technology, the significant reduction in cell area is achieved at the cost of a slight increase in cell-level power and delay characteristics. The chip-level implications of these effects are investigated in the following chapter.

5.3 Chip-Level Simulation Results

Proposed camouflaged 2D and camouflaged monolithic 3D standard cell libraries are used to investigate the system-level power and timing characteristics with existing physical design tools.

Specifically, standard cells listed in Table 5.1 are used to generate four camouflaged circuits: a SIMON block cipher (balanced Feistel cipher to fulfill the security concerns of sensitive and hardware constrained applications [82]) and three ISCAS'89 academic benchmarks.

5.3.1 Experimental Setup

The proposed camouflaged cells are characterized (after *RC* extraction) with *Encounter Library Characterizer (ELC)* to obtain timing and power characteristics. SIMON and ISCAS'89 benchmark circuits are synthesized using *Synopsys Design*

Circuit		SIMON	s35932	s38417	s38584
NO. Gates		903	21,281	22,978	20,423
NCC	DFF	168 (18.6%)	1,728 (8.1%)	1,636 (7.1%)	1,426 (7.0%)
	INV	23 (2.5%)	7,349 (34.5%)	11,253 (49%)	6,784 (33.2%)
CC	NAND	529 (58.6%)	0	554 (2.4%)	974 (4.8%)
	NOR	20 (2.2%)	0	848 (3.7%)	1,096 (5.4%)
	AND	1 (0.01%)	11,052 (51.9%)	8,121 (35.3%)	7,642 (37.4%)
	OR	0	1,152 (5.4%)	566 (2.5%)	2,498 (12.2%)
	XOR	1 (0.01%)	0	0	1 (0.005%)
	XNOR	161 (17.8%)	0	0	2 (0.01%)

Table 5.3: The overall number of gates and the distribution of camouflaged and non-camouflaged cells. *NCC*, and *CC* refer, respectively, to Non-Camouflaged Cells, and Camouflaged Cells.

Compiler. Synthesized netlists are placed (at 70% placement density) and routed using *Cadence Encounter*. The clock frequency is 0.5 GHz for all of the circuits.

5.3.2 System-Level Evaluation

The distribution of cells in camouflaged SIMON cipher and ISCAS'89 benchmarks is listed in Table 5.3. The percentage of camouflaged cells varies significantly depending upon the circuit.

The SIMON cipher used in this work has 168 D flip-flops and 735 gates. (including the inverter, NAND, NOR, AND, OR, XNOR and XOR). Three larger ISCAS'89 benchmark circuits (s35932, s38417, s38584) are also used to evaluate camouflaged 3D monolithic circuits. The number of cells for these benchmarks varies from 17,892 to 22,179.

In 3D monolithic technology, 15 metal layers (both top and bottom tiers) are used for signal routing since 10 metal layers of the top tier are not sufficient to

successfully route the circuit due to the significant reduction in footprint.

Note that for these benchmarks that are relatively larger than the SIMON cipher, 10 metal layers of the top tier are not sufficient to successfully route the circuit due to the significant reduction in footprint. Thus, 15 metal layers (both top and bottom tiers) are used for signal routing in these benchmarks.

Contrary to the SIMON benchmark where experimental results are obtained for both 10 and 15 metal layers of monolithic 3D designs, for ISCAS benchmarks, 10 metal layers are not sufficient to route the camouflaged placed design due to the significant reduction in footprint. Thus, 15 metal layers (both top and bottom tiers) are used for signal routing.

5.3.2.1 Footprint and Wirelength

Physical layouts of the conventional 2D, camouflaged 2D, and camouflaged monolithic 3D implementations of the SIMON block cipher are depicted in Figure 5.5. Approximately 80% of the gates is camouflaged. The area and overall wirelength characteristics in conventional 2D, camouflaged 2D, and camouflaged monolithic 3D are listed in Table 5.4. According to this table, in camouflaged 2D circuits, the area, and wirelength increase, respectively, by 21.1% and 11.3%. For the camouflaged monolithic 3D circuit, however, the area and overall wirelength are reduced, respectively, by 37.7% and 15.7% as compared to the conventional 2D implementation.

For the three larger ISCAS'89 benchmarks, the average area increase for camouflaged 2D circuits is 17.5%. Alternatively, for camouflaged monolithic 3D technology, the area is reduced, on average, by 47.5%. The increase in the overall wirelength for camouflaged 2D circuits is highly design dependent and varies from

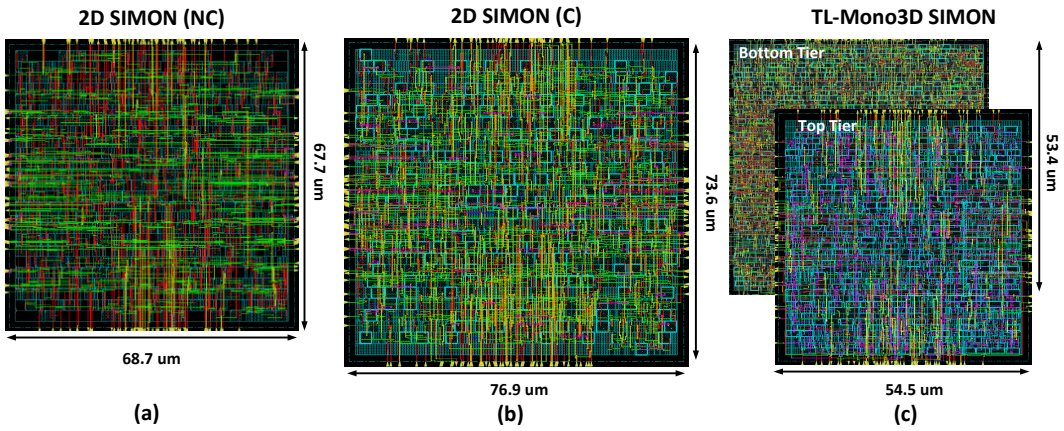


Figure 5.5: The layout views of SIMON cipher in (a) conventional 2D technology without camouflaging, (b) conventional 2D technology with camouflaging, (c) transistor-level monolithic 3D technology with camouflaging.

Circuit	Design style	Area (mm^2)	Change (%)	Wirelength (μm)	Change (%)
SIMON_2D	2D	0.0047	-	10694	-
	2D_C	0.0057	**21**	11905	**11**
SIMON_3D	3D	0.0025	**-47**	8530	**-20**
	3D_C	0.0029	**-38**	9008	**-16**
s35932_2D	2D	0.086	-	163247	-
	2D_C	0.102	**18.6**	173471	**6.3**
s35932_3D	3D	0.042	**-51**	125921	**-23**
	3D_C	0.045	**-48**	127852	**-22**
s38417_2D	2D	0.080	-	133106	-
	2D_C	0.092	**15**	145519	**9.3**
s38417_3D	3D	0.041	**-49**	108975	**-18**
	3D_C	0.042	**-48**	110415	**-17**
s38584_2D	2D	0.080	-	170008	-
	2D_C	0.095	**18.8**	199865	**17.6**
s38584_3D	3D	0.040	**-50**	129650	**-24**
	3D_C	0.042	**-48**	135329	**-20**

Table 5.4: Area and wirelength characteristics in conventional 2D, camouflaged 2D (2D_C), monolithic 3D, and camouflaged monolithic 3D (3D_C) circuits. All of the percentages are with respect to conventional 2D results.

Circuit	Design style	Power Component (mW)			
		Gate	Int.	Leakage	Total
SIMON_2D	2D	5.57	1.43	0.20	7.20
	2D_C	5.89	1.59	0.25	7.73 **(7.4%)**
SIMON_3D	3D	4	1.12	0.20	5.32 **(-26%)**
	3D_C	4.18	1.17	0.25	5.60 **(-22%)**
s35932_2D	2D	66.42	16.45	3.48	86.35
	2D_C	73.31	17.58	4.42	95.13 **(10.2%)**
s35932_3D	3D	64.77	12.32	3.00	80.09 **(-7%)**
	3D_C	68.56	13.53	3.99	86.09 **(-0.3%)**
s38417_2D	2D	59.42	14.31	2.77	76.5
	2D_C	63.17	15.09	3.45	81.7 **(7%)**
s38417_3D	3D	59.64	11.03	2.45	73.12 **(-4%)**
	3D_C	60.93	11.89	3.14	75.96 **(-1%)**
s38584_2D	2D	58.74	14.03	2.77	75.54
	2D_C	62.56	16.50	3.55	82.61 **(9%)**
s38584_3D	3D	57.42	10.7	2.57	70.69 **(-6%)**
	3D_C	58.57	11.73	3.38	73.69 **(-2%)**

Table 5.5: Comparison of power consumption in conventional 2D, camouflaged 2D (2D_C), monolithic 3D, and camouflaged monolithic 3D (3D_C) circuits. *INT* refers to interconnect power. All of the percentages are with respect to conventional 2D results.

approximately 6.3% (for s35932) to 17.6% (for s38584). For camouflaged monolithic 3D technology, the overall wirelength can be reduced, on average, by 19.7%.

5.3.2.2 Power Characteristics

The power consumption of the conventional 2D, camouflaged 2D, and camouflaged monolithic 3D circuits are compared in Table 5.5. All of the three components of power consumption (gate, interconnect, and leakage) are provided. The camouflaged 2D circuits consume, on average, 8.3% more power than the conven-

tional 2D version. This increase is primarily due to the rise in camouflaged gate power (see Table 5.2) and longer interconnects. In camouflaged 2D cells, an additional metal layer is needed for intra-cell routing to make the two cells with different logic functions look identical. Since this metal layer occupies a routing track for inter-cell routing, the overall interconnect length increases (see Table 5.4), thereby increasing the net power. Alternatively, camouflaged monolithic 3D circuits consume, on average, 6.3% less power than the conventional 2D version. This reduction is primarily due to reduced area and therefore shorter interconnects. Thus, an important observation for monolithic 3D technology is that the cell-level power increase due to camouflaging (see Table 5.2) is compensated by the reduction in interconnect power. Also note that in SIMON cipher and s38584, gate power is slightly reduced in 3D technology despite the increase at the cell-level power consumption. This behavior is due the reduced interconnect length in 3D technology which improves the average signal slew (due to lower interconnect resistance), which in turn reduces the short circuit power (one of the components of gate power). Thus, depending upon the type of the cell used by the chip and overall interconnect characteristics, the camouflaged 3D technology may increase or decrease the overall gate power.

5.3.2.3 Timing Characteristics

The worst slack (from the slowest timing path) of the conventional 2D, camouflaged 2D, and camouflaged monolithic 3D circuits are compared in Table 5.6. Note that the timing constraints are satisfied in all of the circuits at 0.5 GHz frequency. According to this table, camouflaging degrades the timing characteristics for both 2D and 3D technologies since the slack is reduced. The average reduction

Circuit	Design Style	Worst Slack (ns)
SIMON_2D	2D	0.921
	2D_C	0.770 **(-16%)**
SIMON_3D	3D	0.885 **(-4%)**
	3D_C	0.745 **(-19%)**
s35932_2D	2D	1.210
	2D_C	1.127 **(-7%)**
s35932_3D	3D	1.064 **(-12%)**
	3D_C	1.035 **(-14%)**
s38417_2D	2D	0.617
	2D_C	0.396 **(-36%)**
s38417_3D	3D	0.406 **(-34%)**
	3D_C	0.279 **(-55%)**
s38584_2D	2D	0.734
	2D_C	0.721 **(-2%)**
s38584_3D	3D	0.583 **(-21%)**
	3D_C	0.578 **(-22%)**

Table 5.6: Timing characteristics in conventional 2D, camouflaged 2D (2D_C), monolithic 3D, and camouflaged monolithic 3D circuits. All of the percentages are with respect to conventional 2D results.

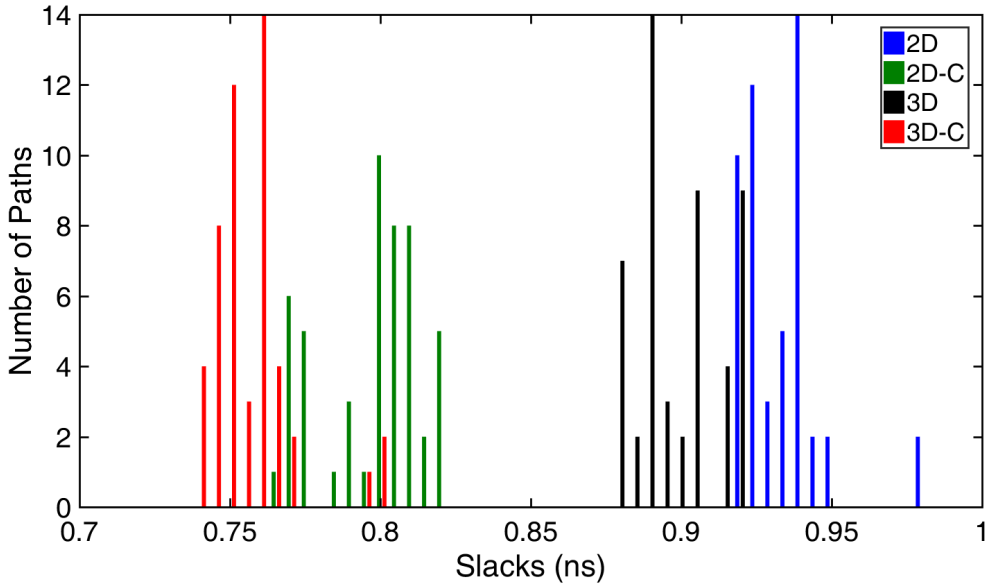


Figure 5.6: Slack distribution of the 50 slowest paths in SIMON cipher for 2D technology without (2D) and with (2D-C) camouflaging, and monolithic 3D technology without (3D) and with camouflaging (3D-C).

in slack is approximately 15.2% (equivalent to 5.85% percent of the clock period) for camouflaged 2D and 27.4% (equal to 10.6% percent of the clock period) for camouflaged 3D circuits. Note that in s38417, a relatively larger reduction in worst slack is observed. This characteristic is due to the presence of a large number of NAND and AND gates along the critical path (specifically 54 out of 110 gates). According to Table 5.2, the largest increase in cell delay is observed for these two gates for both 2D and 3D camouflaging.

To better observe the change in timing characteristics, the slack histogram of the 50 slowest paths is provided in Figure 5.6, where the effect of 2D and 3D camouflaging on slack is illustrated. 2D camouflaging degrades the slack by approximately 120 ps (6% of the clock period). 3D camouflaging causes an additional

degradation of about 50 ps (concerning non-camouflaged 2D) due to larger cell-level delays.

5.4 Summary

Circuit camouflaging has recently received attention to thwarting image analysis based reverse engineering attacks. The number of camouflaged gates, however, should be sufficiently high to ensure its efficacy, which incurs significant overhead. The benefits provided by monolithic 3D technology in circuit camouflaging has been investigated at the cell- and chip-levels. Both 2D and 3D camouflaged cell libraries have been developed and fully characterized. The results obtained from fully placed and routed SIMON cipher and several academic benchmark circuits demonstrate that monolithic 3D technology is highly effective in eliminating not only the area but also the power overhead of circuit camouflaging at the expense of a slight degradation in timing characteristics.

Chapter 6

Thermal Management and Analysis of Monolithic 3D ICs

Monolithic 3D ICs support ultra-high density device integration through fine-grained connectivity enabled by monolithic inter-tier vias. Monolithic 3D technology can potentially save up to 50% in area while exhibiting comparable to or better performance than 2D circuits. However, thermal management of monolithic 3D ICs is more challenging due to the reduced area (hence, higher power density) and the low thermal conductivity of the inter-layer dielectric material, situated between adjacent device layers. Therefore, carefully and accurately analyzing the thermal behavior is crucial for the reliability of monolithic 3D integration. In this chapter, to analyze heat propagation and track the formation of hot spots in monolithic 3D ICs, the thermal integrity of the monolithic 3D circuits is explored at the physical level.

The rest of the chapter is organized as follows. Background of thermal issues in

modern ICs and contributions of this work are summarized in Section 6.1. Proposed thermal analysis flow for monolithic 3D ICs is provided in Section 6.2. Simulation results for a secure cipher and a 128-point FFT core are presented in Section 6.3. The chapter is summarized in Section 6.4.

6.1 Background and Contributions

6.1.1 Thermal Management

Thermal management has emerged as one of the most challenging issues in high-performance ICs due to the increasing transistor density and power consumption, and clock speed [83]. Furthermore, thermal issues caused by higher power densities in sub-90 nm advanced CMOS technologies result in performance degradation and significant leakage currents [84]. With the help of modern device technologies such as FinFET and SOI, further scaling of silicon technology by suppressing leakage currents has been possible [6]; however, due to self-heating and the low thermal conductivity of the materials involved, these technologies can be more sensitive to thermal issues [85].

6.1.1.1 Thermal Issues in Monolithic 3D ICs

3D integration has attracted significant attention in recent years due to its ability to outperform traditional planar CMOS technologies [44, 86–90]. Long interconnect delay and heterogeneous technology integration are a few examples of issues that could be addressed by 3D integration [91]. A number of innovative process technologies including wafer thinning, etching and filling of high aspect ratio holes

in silicon, inter-strata bonding are used to develop a functional 3D circuit [92, 93]. Many different 3D integration technologies have been demonstrated, such as face-to-face integration [92] and back-to-face integration [94].

Nanoscale monolithic inter-tier via (MIV) connects the vertical device layers and has recently become a key component of vertical integration [13]. Monolithic 3D ICs exhibit significant benefits regarding electrical performance (as mentioned in Chapter 3), but they are also expected to suffer from thermal issues due to higher power densities [95–98].

6.1.1.2 Full-Chip Thermal Analysis

Heat is a form of energy that can be transferred from one system to another as a result of temperature difference. This process is known as heat transfer [99, 100]. The full-chip thermal analysis requires understanding and solving the heat transfer theory [97].

The flow of on-chip heat is represented by a parabolic partial differential equation (PDE) [6, 101],

$$\rho C(x,\mu)\frac{\partial \mu(x,t)}{\partial t} - \nabla(\kappa(x,\mu)\nabla\mu(x,t)) = f(\mu,x,t), \quad in \quad \Omega \times [0,T], \tag{6.1}$$

$$(\kappa(x,\mu)\nabla\mu)n = \eta(\mu_a - \mu) \quad on \quad \partial\Omega \times [0,T], \tag{6.2}$$

$$\mu(x,0) = \mu_0 \quad in \quad \bar{\Omega} = \Omega \cup \partial\Omega, \tag{6.3}$$

where ρ is the material density (in kg/m^3), the function $C(x,\mu)$ denotes the heat ca-

pacity of the chip material (in J/kgK), $\kappa(x,\mu)$ is the thermal conductivity (in W/mK), $\mu(x,t)$ is the temperature (in [K]) at time t, the function $f(\mu,x,t)$ (in W/m^3) is the power density dissipated by the active layers in the system. The power density is comprised of two parts: $f_1(x,t)$ and $f_2(\mu,x,t)$. $f_1(x,t)$ is the dynamic power and $f_2(\mu,x,t)$ is the leakage power. The leakage power has exponential dependence on temperature [102–105]. $[0,T]$ is the time interval, η (in W/m^2K) is the thermal transmittance at the boundaries $\partial\Omega$.

Finite element method (FEM) discretizes the entire chip's temperature field and forms a system of linear equations showing the relation between the temperature distribution within the chip and the power density distribution [97]. The initial boundary condition in (6.1), (6.2), and (6.3) can be solved numerically by using the FEM and implicit time integration methods [106].

6.1.2 Contributions of This Work

The primary contributions of this work are as follows:

- An analysis flow is developed that utilizes the proposed monolithic 3D cell libary, PDK, and an existing thermal analysis tool [6] for computing steady-state and transient thermal profiles of monolithic 3D ICs.
- Thermal behavior of several monolithic 3D circuits is explored.

6.2 Proposed Thermal Analysis Flow

Proposed flow consists of two primary parts: monolithic 3D cell library characterization to obtain power data (see Section 6.2.1) and use of an existing GDS level

thermal analysis tool [6] (see Section 6.2.2).

6.2.1 Standard Cell Library Review

Standard cell timing and power modeling are described in Section 6.2.1.1 and Section 6.2.1.2, respectively.

6.2.1.1 Standard Cell Timing Modeling

Standard cell timing models provide relatively accurate timing information for different switching activities of the cell instances [107]. In liberty format (.lib), there are look-up table based models called Non-Linear Delay Model (NLDM) which specify the delay and timing checks for different cell timing arcs [108]. The delay for these timing arcs depends on two variables:

- The capacitance load at the cell output pin
- The transition time (slew) of the signal at the cell input pin.

These two variables are independent and build up a two-dimensional look-up table in the NLDM delay model. When calculating the delay values, two methods can be applied: interpolation and extrapolation. If the input slew and output load are in the range, then the interpolation is applied. Extrapolation is utilized when these two variables are out of range, however, extrapolation has lower accuracy as compared to the interpolation.

In one cell, for each pin, there are two tables for transition time: $rise\_transition$ and $fall\_transition$, and two tables for delay time: $rise\_delay$ and $fall\_delay$. Following is an example of $cell\_rise$ delay representation:

```
// monolithic_3d_v2.lib:
timing() {
      related_pin : "A";
      timing_sense : negative_unate;
      cell_rise(delay_template_10x7) {
        index_1 ("0.006, 0.12, 0.24, 0.36, 0.48, 0.6, 0.72, 0.96,
            1.08, 1.2"); /* input slew */
        index_2 ("0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1"); /* output
            load*/
        values ( ... );
      }
      ...
    }
```

6.2.1.2 Standard Cell Power Modeling

Power is dissipated due to two components: active power and standby or leakage power [109, 110]. The standard cell library includes both of these components for each pin within the cells.

The switching activity induces the active power at the cell input pins and output pin. Charging of the output load and the internal cell switching are two sources of the active power.

Since the internal switching power depends upon the type of the cell, it is included in the cell library [108]. Following is an example of the description of a cell

internal power in the library:

```
// monolithic_3d_v2.lib:
internal_power() {
      related_pin : "A";
      rise_power(energy_template_10x7) {
        index_1 ("0.006, 0.12, 0.24, 0.36, 0.48, 0.6, 0.72, 0.96,
            1.08, 1.2"); /* input slew */
        index_2 ("0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1"); /* output
            load */
        values ( ... );
      }
      ...
     }
```

Leakage power contribution is from two phenomena: subthreshold current and gate oxide tunneling [111, 112]:

$$P_{leakage} = P_{subthreshold} + P_{gate-leakage}. \tag{6.4}$$

The subthreshold MOSFET leakage power strongly depends upon temperature [113]; however, the leakage contributed through gate oxide tunneling is relatively insensitive to temperature [114]. Standard cell library specifies leakage power for each cell. An example of the leakage power specification in the library is provided below:

```
// monolithic_3d_v2.lib:
cell (INVX1) {
  ...
  cell_leakage_power : 53.0162;
  ...
}
```

6.2.2 Thermal Analysis Flow

The proposed thermal analysis flow is introduced in this section. A brief introduction of the thermal analysis tool used in this thesis, Manchester Thermal Analyzer (MTA) [6], is described in Section 6.2.2.1. The data flow and required files are presented in Section 6.2.2.2.

6.2.2.1 Thermal Analyzer Introduction

There are several existing thermal analysis tools for 2D and 3D ICs [6], such as HotSpot [115], 3D ICE [116], and 3D Thermal-ADI [117, 118]. However, some of these tools focus only on linear representations and some of them rely on time-consuming methods.

Manchester Thermal Analyzer (MTA) is a versatile, fast, and accurate tool for thermal analysis [119]:

- It can analyze both the steady-state and transient thermal profiles.
- It supports both linear and non-linear models.

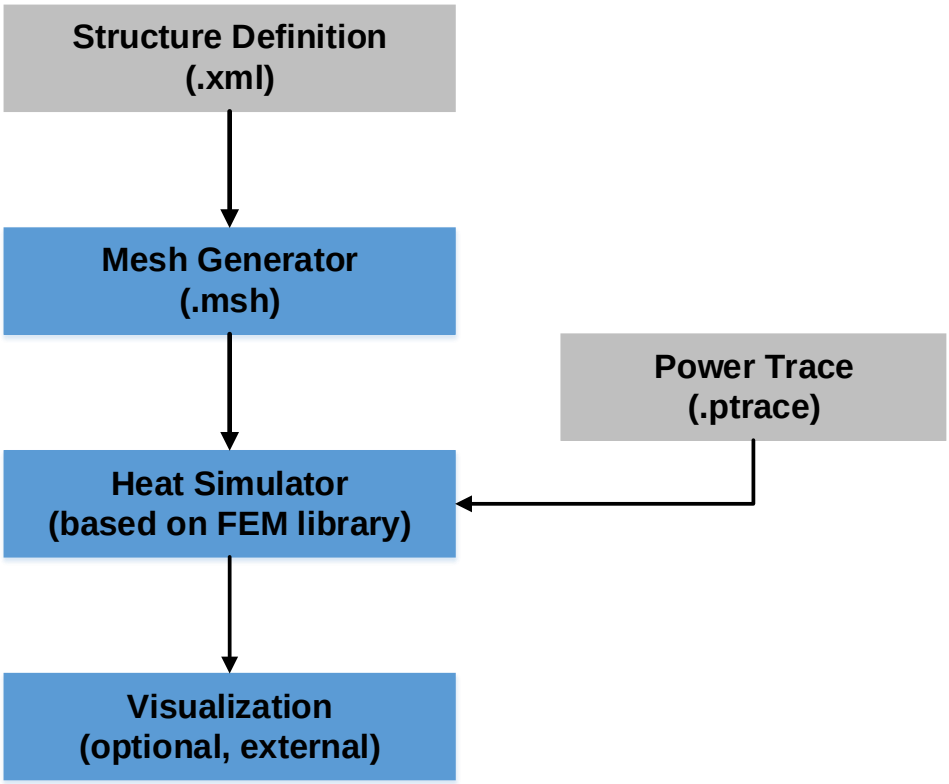


Figure 6.1: The MTA flow with user-supplied input files highlighted in grey [6].

- It can compute highly accurate temperature profiles at near-optimal computational cost.

A flow chart of the simulation process of the MTA is shown in Figure 6.1. The input documents for MTA are layer thickness file with related process information, design exchange format (.def), library exchange format (.lef), liberty (.lib), standard parasitic extraction format (.spef), static power trace and transient power trace. Detailed introduction of thermal analysis data flow through MTA is discussed in Section 6.2.2.2.

Layer Name	Thickness (μm)
metal10	2.0
metal9	2.0
metal8	0.82
metal7	0.82
metal6	0.29
metal5	0.29
metal4	0.29
metal3	0.12
metal2	0.12
metal1	0.12
poly	0.085
inter tier dielectric	0.1
metal2_btm	0.12
metal1_btm	0.12
poly_btm	0.085
substrate	1.0

Table 6.1: Layer thicknesses in proposed monolithic 3D technology.

6.2.2.2 Thermal Analysis Data Flow

The MTA provides an entirely programmed 3D mesh generation tool. This generator can create the mesh by parsing the bounding box of the integrated circuit from the XML file [6, 106]. Thickness information of each layer is required in the XML file. Thickness information for each layer in proposed monolithic 3D technology is provided in Table 6.1.

The mesh generator creates the computational mesh of the related design. The initial mesh does not have any hanging nodes, and its development depends on the circuit topology [6]. Figure 6.2 is an example of ISCAS'89 benchmark circuit s27 [120] in the computational mesh of the XY plane generated by MTA mesh generator.

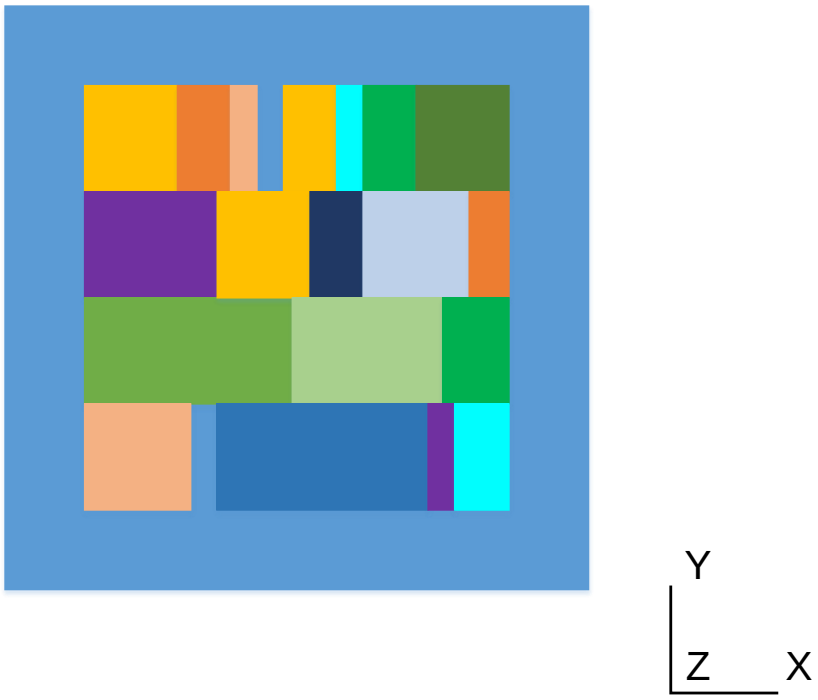


Figure 6.2: Computation mesh of ISCAS'89 s27 generated by MTA.

Design Exchange Format (.def) is usually generated by Place and Route tools for post-layout analysis [121]. It is used to present the physical layout of the IC, and it can be used along with Library Exchange Format (.lef) file to represent the complete circuit layout design.

Library Exchange Format (.lef) file provides information of the technology, such as the metal layer, via layer information and via generation rules. It is the abstract view of standard cells, which includes PR boundary, pin position, and metal layer parasitic impedance data. Liberty (.lib) file contains cell-specific power and timing data, as described earlier.

Standard Parasitic Extraction Format (.spef) has the description of parasitic information of a design in an ASCII exchange format. It is primarily used to pass parasitic information from one tool to another tool [108].

Static power trace and transient power trace are generated through different flows:

- Static power trace generation design flow has been described in Chapter 3.1.2: RLT-level benchmark circuits are synthesized through synthesis tools such as Cadence Genus [122] or Synopsys Design Compiler [62] after linking the design to the standard cell library. A gate-level netlist of the design can be generated. This netlist will be the input to the next design step: Place and Route. After place and route, the detailed routed netlist and the .spef file of the design which contains extracted parasitic impedance data is generated for static timing analysis, as shown in Figure 6.3.

- Transient power analysis flow is shown in Figure 6.4. Procedures before Place and Route are the same as the static power design flow. The routed

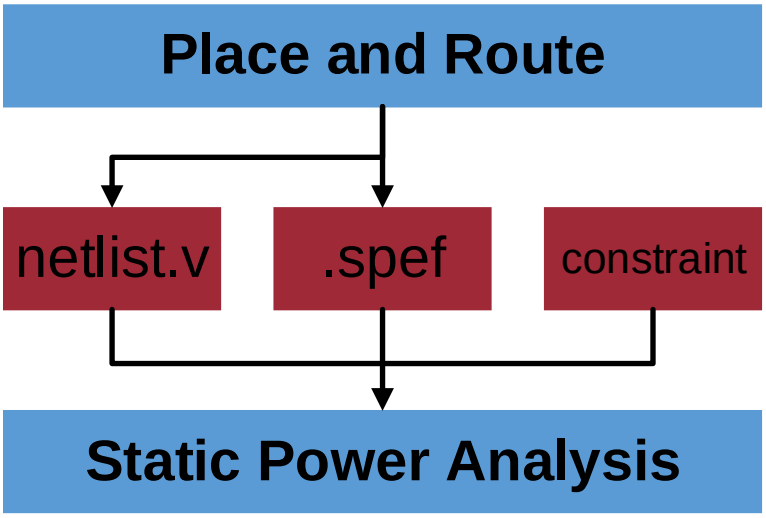


Figure 6.3: Static power analysis data flow.

circuit netlist and the extracted parasitic data can be collected after place and route process. A timing constraint for post-layout level design is required at this stage for timing closure. The output of static timing analysis will be a Standard Delay Format (.sdf) file which describes the timing information and constraints. A Value Change Dump (.vcd) file is generated after gate-level simulations. A .vcd file contains a series of time-ordered value changes for the signals in a given simulation model [123, 124]. There are four sections in a .vcd file: header section, variable definition section, $dumpvar section and value change section. After running PrimeTime PX [125], a .fsdb waveform file can be generated. The timing and power information for each cell are included in this file. The data can be exported to column based (txt, excel) format by waveform viewer such as Synopsys Waveview [126]. An example

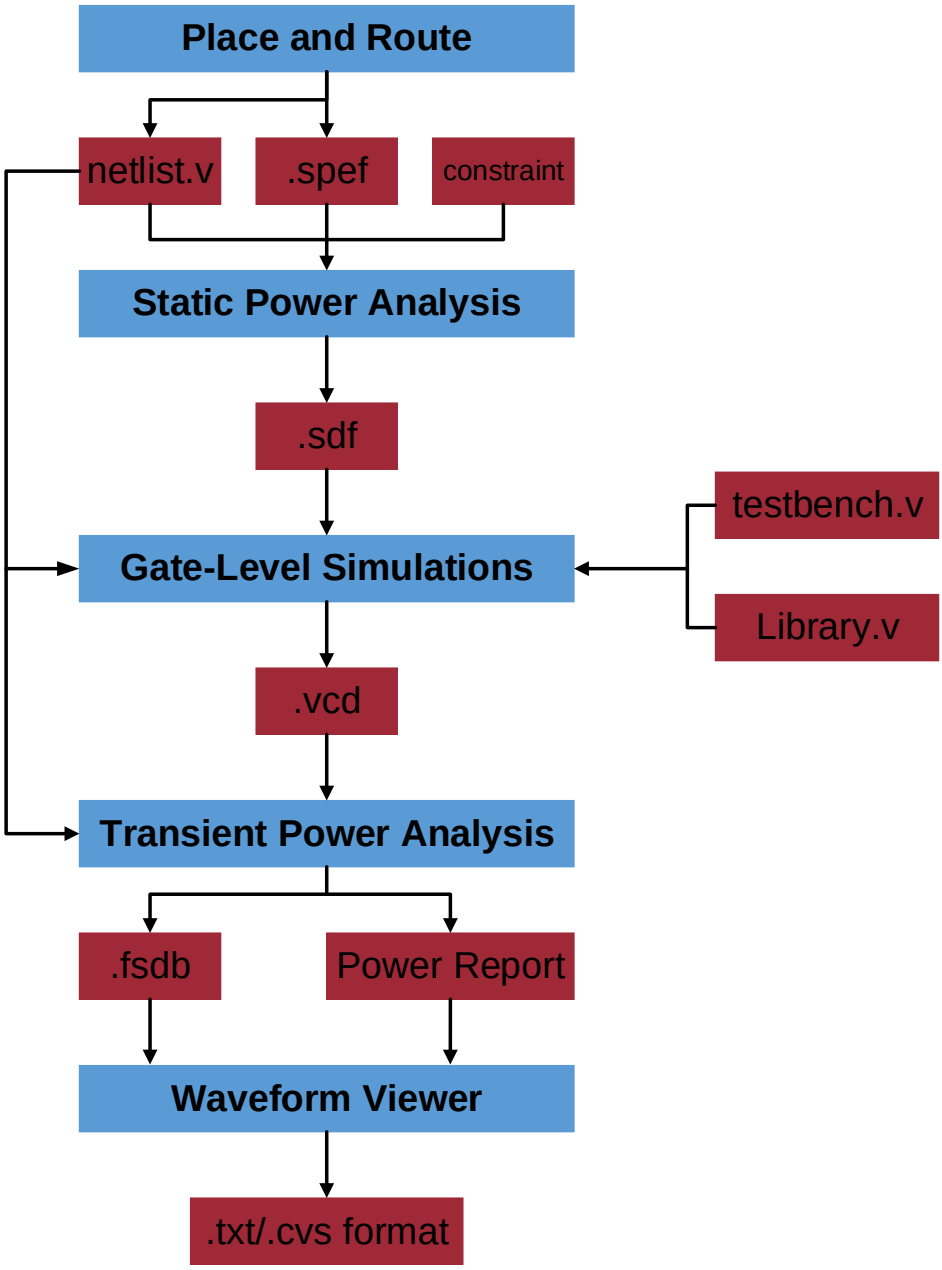


Figure 6.4: Transient power analysis data flow.

Figure 6.5: Transient power waveform of AES128.

of transient power data from .fsdb file is shown in Figure 6.5. The X-axis represents the timing information, and the Y-axis represents power data of one datapath.

6.3 Simulation Results

The test cases introduced in this work are 128-bit encryption core AES128 and a 128-point FFT core FFT128 with testbenches for static and transient analysis. FFT128 has been introduced in Chapter 3 in detail. AES (Advanced Encryption Standard) is a specification for electronic data encryption published by American National Institute of Standards and Technology (NIST) in 2001 [127–129]. It describes a symmetric-key algorithm where the block size of AES is restricted to 128 bits and the key size can be 128, 196 or 256 bits.

Operating Frequency		250 MHz			
Circuit	Design Style	Footprint (mm^2)	Change (%)	Wirelength (μm)	Change (%)
AES128	2D	0.914	-	4,870,687	-
	3D	0.568	**-37.9**	4,027,709	**-17.3**
FFT128	2D	2.54	-	11,161,750	-
	3D	1.71	**-32.7**	8,546,622	**-23.3**

Table 6.2: Comparison of footprint and wirelength in 2D with 14 routing tracks and monolithic 3D technologies with 9 (*Mono3D*) routing tracks in each 3D cell, operating at 250 MHz. The percent changes with respect to 2D cells are listed.

Simulations results of footprint and power for AES128 and FFT128 are described in Section 6.3.1. Thermal analysis results are discussed in Section 6.3.2.

6.3.1 Footprint and Power

6.3.1.1 Footprint

The layouts of conventional 2D and monolithic 3D AES128 and FFT128 with operating frequency of 250 MHz are shown in Figure 6.6. Footprint and overall wirelength data of monolithic 3D and conventional 2D designs are listed in Table 6.2. These results are also compared in Figure 6.7.

According to these tables and figures, monolithic 3D AES128 achieves a %37.9 reduction in chip footprint and %17.3 reduction in overall wirelength, and monolithic 3D FFT128 achieves a %32.7 reduction in chip footprint and %23.3 reduction in total wirelength as compared to their conventional 2D counterparts at operating frequency of 250 MHz.

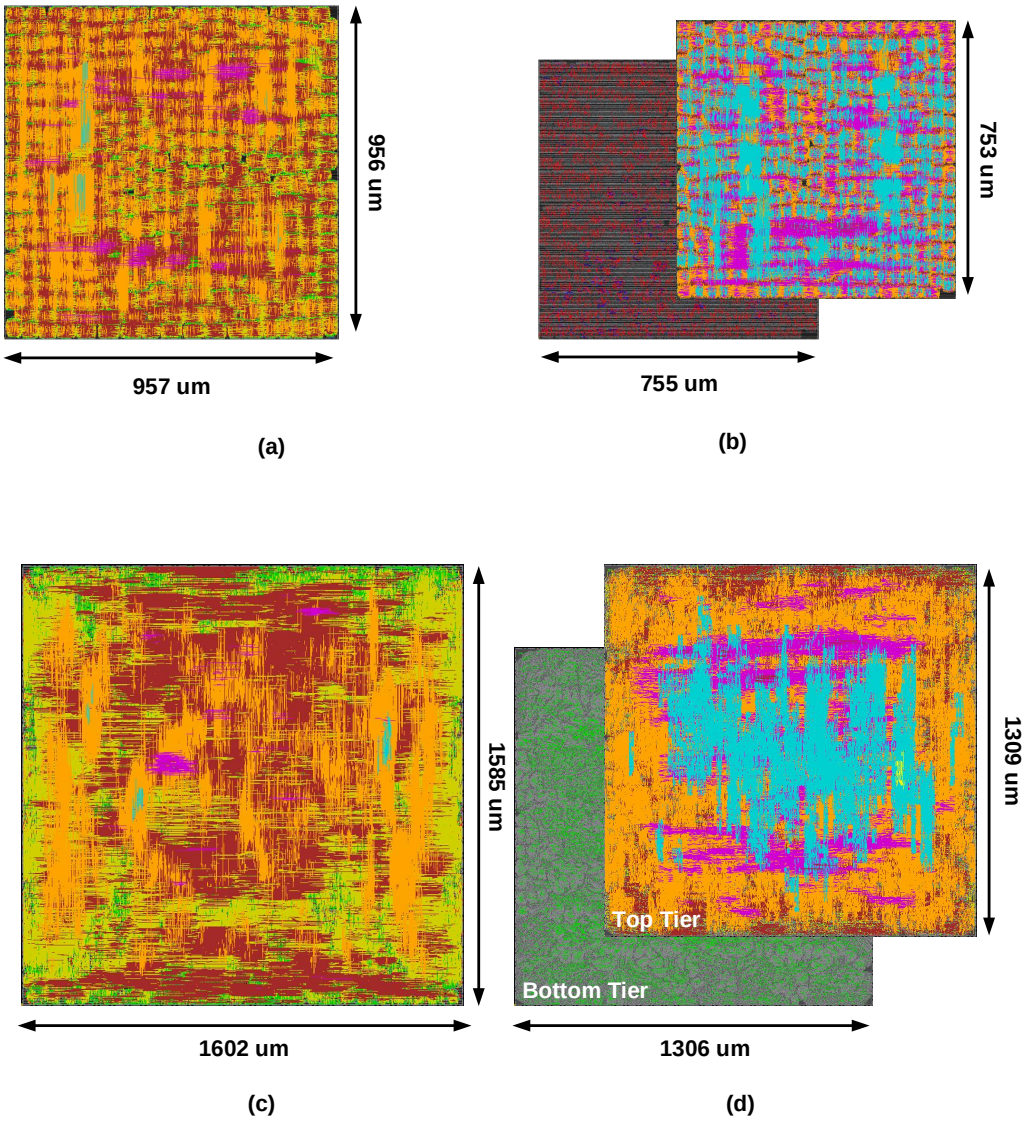
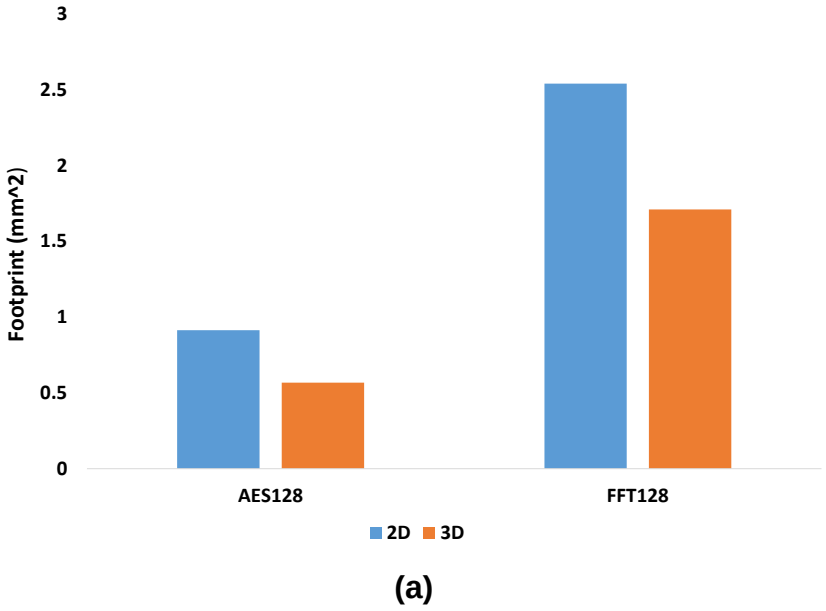
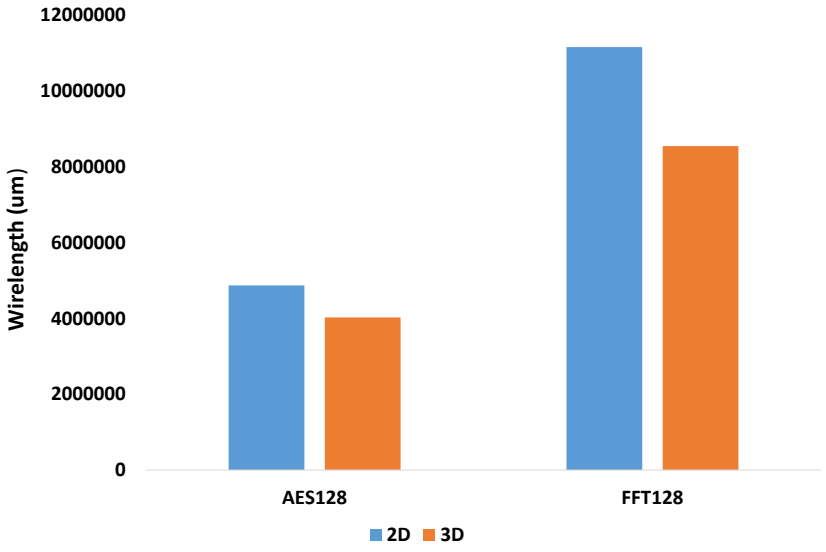


Figure 6.6: The layout views of (a) AES128 core in conventional 2D technology, (b) AES128 core in transistor-level monolithic 3D technology, (c) FFT128 core in conventional 2D technology, (d) FFT128 core in transistor-level monolithic 3D technology.



(a)



(b)

Figure 6.7: The footprint and wirelength comparison of AES128 and FFT128 cores in conventional 2D technology and transistor-level monolithic 3D technology.

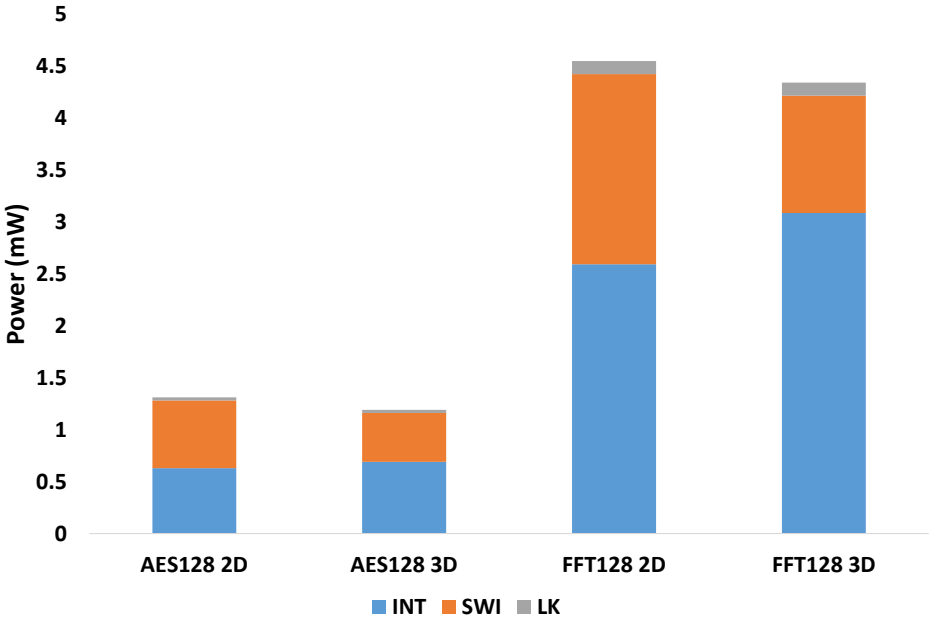


Figure 6.8: The power comparison of AES128 and FFT128 cores in conventional 2D technology and transistor-level monolithic 3D technology. *INT, SWI*, and *LK* refer, respectively, to internal, switching (net), and leakage power.

6.3.1.2 Power

The power consumption of conventional 2D and monolithic 3D AES128 and FFT128 is listed in Table 6.3. Three components of power consumption (internal power, switching power, and leakage power) are provided. Because of the considerable reduction in overall wirelength in monolithic 3D designs, the switching power is reduced by %27.7 for AES128 and %27.3 for FFT128 respectively at 250 MHz. The overall power consumption is compared in Figure 6.8.

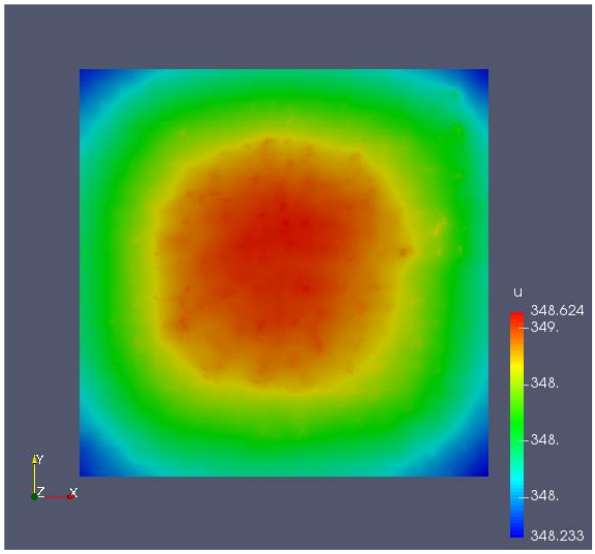
Operating Frequency		250 MHz			
Circuit	Design Style	Power component (mW)			
		INT	SWI (Change)	LK	Total (Change)
AES128	2D	0.63	0.65 (-)	0.03	1.31 (-)
	3D	0.69	0.47 **(-27.7%)**	0.03	1.19 **(-9.16%)**
FFT128	2D	2.591	1.829 (-)	0.125	4.545 (-)
	3D	2.884	1.329 **(-27.3%)**	0.126	4.339 **(-4.53%)**

Table 6.3: Comparison of power consumption in 2D and monolithic 3D technologies 9 routing tracks in each cell, operating at 250 MHz. *INT, SWI*, and *LK* refer, respectively, to internal, switching (net), and leakage power.

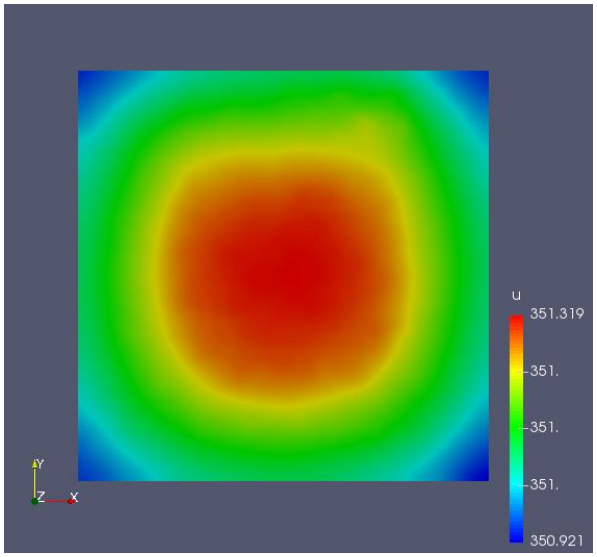
6.3.2 Temperature

The steady-state temperature characteristics of conventional 2D and monolithic 3D technologies are shown, respectively, in Figure 6.9 and Figure 6.10. Results show that the AES128 circuit based on the monolithic 3D technology saves 37.9% of silicon area while the temperature is 2.3 °C higher than the conventional 2D version. The FFT128 circuit shows 33% improvement in the area while the temperature is 3.8 °C higher in steady-state and 2.1 °C higher in transient thermal analysis with practical switching activities for 15 ms.

According to the above results, the temperature characteristics in monolithic 3D technology exhibit only slight increase as compared to 2D technology. This relatively small increase is due to the higher power densities. Large temperature difference is not observed between 2D and 3D implementations due to (1) the cross-sectional thicknesses do not change significantly since the additional 3D process layers are sufficiently thin, and (2) power is reduced by approximately 10% in 3D implementations.

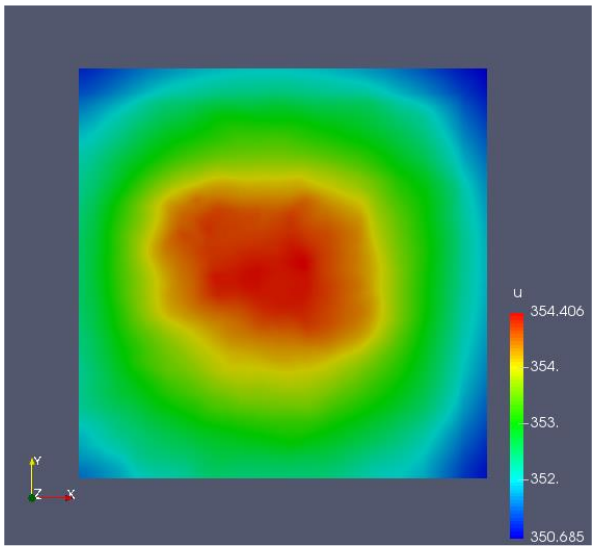


(a)

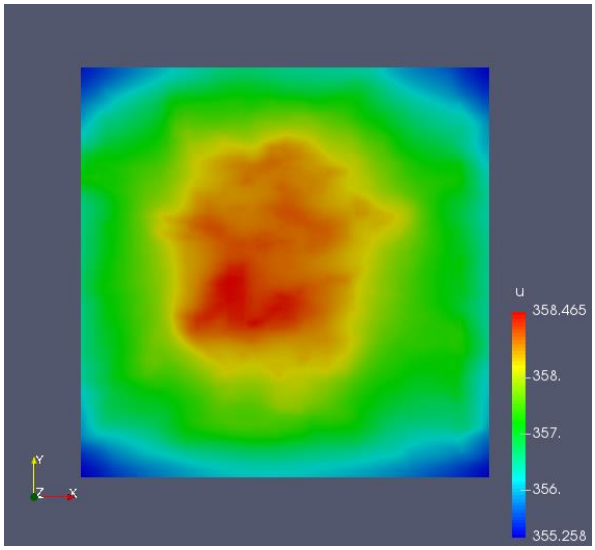


(b)

Figure 6.9: Spatial temperature distribution of AES128 core in (a) conventional 2D technology, (b) in transistor-level monolithic 3D technology.



(a)



(b)

Figure 6.10: Spatial temperature distribution of FFT128 core in (a) conventional 2D technology, (b) in transistor-level monolithic 3D technology.

6.4 Summary

In this chapter, an open-source monolithic 3D circuit process design kit is used to develop a thermal analysis flow for monolithic 3D ICs. An existing thermal analyzer based on the finite element method is used for GDS-level static and transient heat simulations. The test cases presented in this chapter are the AES128 and FFT128 circuits with appropriate testbenches for steady-state and transient thermal analysis.

Chapter 7

Conclusions and Future Directions

The monolithic 3D technology is effective to reduce circuit footprint and power dissipation. In this thesis, an open source transistor-level standard cell library for monolithic 3D ICs, a routing congestion aware monolithic 3D transistor-level standard cell library, and a hardware-efficient logic camouflaging for monolithic 3D ICs are proposed and all of the related automation files are made publicly available to facilitate future research on various important aspects of 3D monolithic integration such as thermal integrity, design for test (DFT), and interaction between the manufacturing/device development and the design process. These contributions are summarized in Section 7.1. Several possible future works are discussed in Section 7.2.

7.1 Thesis Summary

3D ICs have emerged as a practical solution to some of the critical issues encountered in planar technologies. An open source transistor-level standard cell

library for monolithic 3D ICs is proposed in Chapter 3. Simulation results of a large-scale FFT core demonstrate that the monolithic 3D technology can reduce the footprint and overall power consumption by, respectively, 38% and 14%. The entire proposed library and related files for tool integration are publicly available to facilitate future research.

An open source routing congestion aware transistor-level monolithic 3D standard cell library is developed and integrated into a digital design flow, as described in Chapter 4. Simulation results with the proposed cell library demonstrate that the effect of routing congestion on timing characteristics is stronger in monolithic 3D technology, where the cell-level number of routing tracks plays an essential role. An optimum number of routing tracks exists that achieves the most significant improvements in both power and timing characteristics.

The security of 3D ICs has emerged as a fundamental issue due to the threats from the globalized semiconductor supply chain. To thwart reverse engineering attacks, circuit camouflaging is a useful technique. Therefore, a hardware-efficient logic camouflaging method for monolithic 3D ICs is proposed in Chapter 5 to investigate the benefits provided by monolithic 3D technology in circuit camouflaging. Simulation results demonstrate that the monolithic 3D technology is highly effective in eliminating not only the area but also the power overhead of circuit camouflaging at the expense of a slight degradation in timing characteristics.

Monolithic 3D integration is a promising technique to improve integration density. 3D ICs, however, suffer from more stringent thermal issues when compared to equivalent implementations in traditional single-die technologies. Therefore, efficient analysis of thermal behavior for monolithic 3D integration is essential for mitigating thermal issues. The results of Chapter 6 provide an overview of the heat

dissipation to analyze thermal characteristics in monolithic 3D integration technology. A thermal analysis flow for monolithic 3D ICs is introduced.

7.2 Future Directions

Future work includes a more in depth study of thermal issues in monolithic 3D ICs and comparison with conventional 2D technologies. The impact of several physical design characteristics on temperature can be investigated such as power distribution networks, cell placement density, MIV density, and number of routing tracks. In terms of hardware security, the side channel resistance of monolithic 3D circuits can be studied. Another interesting direction is the investigation of hardware Trojans that can span multiple tiers within a monolithic 3D circuit. All of these future works should closely follow the latest developments in monolithic 3D process/manufacturing to ensure that the PDK is up to date and represents a practical scenario.

Bibliography

[1] M. Karnezos, "3D Packaging: Where All Technologies Come Together," in *Proceedings of the IEEE/CPMT/SEMI International Electronics Manufacturing Technology Symposium*, 2004, pp. 64–67.

[2] J. Miettinen, M. Mantysalo, K. Kaija, and E. O. Ristolainen, "System Design Issues for 3D System-in-Package (SiP)," in *Proceedings of the IEEE International Electronic Components and Technology Conference*, 2004, pp. 610–615.

[3] S. F. Al-Sarawi, D. Abbott, and P. D. Franzon, "A Review of 3-D Packaging Technology," *IEEE Transactions on Components, Packaging, and Manufacturing Technology: Part B*, vol. 21, no. 1, pp. 2–14, 1998.

[4] "International Technology Roadmap for Semiconductors (ITRS)," 2011.

[5] "CEA-LETI." [Online]. Available: http://electroiq.com/blog/2014/07/cea-leti-monolithic-3d-is-the-solution-for-further-scaling/

[6] S. Ladenheim, Y.-C. Chen, M. Mihajlović, and V. F. Pavlidis, "The MTA: An Advanced and Versatile Thermal Simulator for Integrated Systems," *IEEE*

Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2018.

[7] "7 nm process offerings." [Online]. Available: https://en.wikipedia.org/wiki/7_nanometer

[8] "GlobalFoundries Halts 7nm Work." [Online]. Available: https://www.eetimes.com/document.asp?doc_id=1333637

[9] "UMC not to rejoin race to develop 7nm technology." [Online]. Available: http://www.taipeitimes.com/News/biz/archives/2018/09/04/2003699736

[10] "GLOBALFOUNDRIES." [Online]. Available: https://www.globalfoundries.com/

[11] "United Microelectronics Corporation." [Online]. Available: http://www.umc.com/English/

[12] J. Shi, D. Nayak, M. Ichihashi, S. Banna, and C. A. Moritz, "On the Design of Ultra-High Density 14nm Finfet Based Transistor-Level Monolithic 3D ICs," in *Proceedings of the IEEE Computer Society Annual Symposium on VLSI*, Jul. 2016, pp. 449–454.

[13] C. Yan and E. Salman, "Mono3D: Open Source Cell Library for Monolithic 3-D Integrated Circuits," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 3, pp. 1075–1085, 2018.

[14] S. Wong, A. El, P. Griffin, Y. Nishi, F. Pease, and J. Plummer, "Monolithic 3D Integrated Circuits," in *Proceedings of the IEEE International Symposium on VLSI Technology, Systems and Applications*, Apr. 2007, pp. 1–4.

[15] P. Batude, M. Vinet, A. Pouydebasque, C. Le Royer, B. Previtali, C. Tabone, L. Clavelier, S. Michaud, A. Valentian, O. Thomas *et al.*, "GeOI and SOI 3D Monolithic Cell Integrations for High Density Applications," in *Proceedings of the IEEE International Symposium on VLSI Technology*, Jun. 2009, pp. 166–167.

[16] "Mono3D, Open Source Cell Library for Transistor-Level Monolithic 3D Integration." [Online]. Available: http://www.ece.stonybrook.edu/~emre/Mono3D.zip

[17] P. Milder, F. Franchetti, J. C. Hoe, and M. Püschel, "Computer Generation of Hardware for Linear Digital Signal Processing Transforms," *ACM Transactions on Design Automation of Electronic Systems*, vol. 17, no. 2, pp. 15:1–15:33, Apr. 2012.

[18] "International Roadmap for Devices and Systems (IRDS)." [Online]. Available: https://irds.ieee.org/

[19] E. Salman and E. G. Friedman, *High Performance Integrated Circuit Design*. McGraw-Hill Professional, Aug. 2012.

[20] V. F. Pavlidis and E. G. Friedman, *Three-Dimensional Integrated Circuit Design*. Morgan Kaufmann, July 2010.

[21] E. Salman, M. H. Asgari, and M. Stanaćević, "Signal Integrity Analysis of a 2-D and 3-D Integrated Potentiostat for Neurotransmitter Sensing," in *Proceedings of the IEEE Biomedical Circuits and Systems Conference*, 2011, pp. 17–20.

[22] C. Yan, Z. Gan, and E. Salman, "In-Package Spiral Inductor Characterization for High Efficiency Buck Converters," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, 2017, pp. 1–4.

[23] S. Stoukatch, C. Winters, E. Beyne, W. De Raedt, and C. Van Hoof, "3D-SiP Integration for Autonomous Sensor Nodes," in *Proceedings of the IEEE Electronic Components and Technology Conference*, 2006, pp. 5–pp.

[24] D. H. Kim, "3D-MAPS: 3D Massively Parallel Processor with Stacked Memory," in *Proceedings of the IEEE International Solid-State Circuits Conference*, Feb. 2012, pp. 188–190.

[25] V. F. Pavlidis and E. G. Friedman, "Interconnect-Based Design Methodologies for Three-Dimensional Integrated Circuits," *Proceedings of the IEEE*, vol. 97, no. 1, pp. 123–140, Jan. 2009.

[26] S. M. Satheesh and E. Salman, "Power Distribution in TSV-Based 3D Processor-Memory Stacks," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2, no. 4, pp. 692–703, Dec. 2012.

[27] H. Wang, M. H. Asgari, and E. Salman, "Efficient Characterization of TSV-to-Transistor Noise Coupling in 3D ICs," in *Proceedings of the ACM international conference on Great lakes symposium on VLSI*, 2013, pp. 71–76.

[28] S. Fang and E. Salman, "Low Swing TSV Signaling Using Novel Level Shifters with Single Supply Voltage," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, 2015, pp. 1965–1968.

[29] E. Salman, "Noise Coupling Due to Through Silicon Vias (TSVs) in 3-D Integrated Circuits," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, 2011, pp. 1411–1414.

[30] S. M. Satheesh and E. Salman, "Effect of TSV Fabrication Technology on Power Distribution in 3D ICs," in *Proceedings of the ACM international conference on Great lakes symposium on VLSI*, 2013, pp. 287–292.

[31] H. Wang and E. Salman, "Power Gating Methodologies in TSV Based 3D Integrated Circuits," in *Proceedings of the ACM international conference on Great lakes symposium on VLSI*, 2013, pp. 327–328.

[32] L. Brunet, P. Batude, C. Fenouillet-Beranger, P. Besombes, L. Hortemel, F. Ponthenier, B. Previtali, C. Tabone, A. Royer, C. Agraffeil *et al.*, "First Demonstration of a CMOS over CMOS 3D VLSI CoolCub Integration on 300mm Wafers," in *Proceedings of the IEEE Symposium on VLSI Technology*, 2016, pp. 1–2.

[33] H. Wang and E. Salman, "Enhancing System-Wide Power Integrity in 3D ICs with Power Gating," in *Proceedings of the IEEE International Symposium on Quality Electronic Design*, 2015, pp. 322–326.

[34] C. Erdmann, D. Lowney, A. Lynam, A. Keady, J. McGrath, E. Cullen, D. Breathnach, D. Keane, P. Lynch, M. De La Torre *et al.*, "A Heterogeneous 3D-IC Consisting of Two 28 nm FPGA Die and 32 Reconfigurable High-Performance Data Converters," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 1, pp. 258–269, 2015.

[35] H. Wang and E. Salman, "Resource Allocation Methodology for Through Silicon Vias and Sleep Transistors in 3D ICs," in *Proceedings of the IEEE International Symposium on Quality Electronic Design*, 2015, pp. 528–532.

[36] S. M. Satheesh and E. Salman, "Design Space Exploration for Robust Power Delivery in TSV Based 3-D Systems-On-Chip," in *Proceedings of IEEE International SOC Conference*, 2012, pp. 307–311.

[37] S. K. Lim, *Design for High Performance, Low Power, and Reliable 3D Integrated Circuits*. Springer Science & Business Media, 2012.

[38] D. H. Kim, K. Athikulwongse, and S. K. Lim, "A Study of Through-Silicon-Via Impact on the 3D Stacked IC Layout," in *Proceedings of the ACM International Conference on Computer-Aided Design*, Nov. 2009, pp. 674–680.

[39] H. Wang and E. Salman, "Decoupling Capacitor Topologies for TSV-Based 3-D ICs with Power Gating," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 23, no. 12, pp. 2983–2991, 2015.

[40] I. Savidis and E. G. Friedman, "Closed-Form Expressions of 3D Via Resistance, Inductance, and Capacitance," *IEEE Transactions on Electron Devices*, vol. 56, no. 9, pp. 1873–1881, Sep. 2009.

[41] H. Wang, M. H. Asgari, and E. Salman, "Compact Model to Efficiently Characterize TSV-to-Transistor Noise Coupling in 3D ICs," *The VLSI journal on Integration*, vol. 47, no. 3, pp. 296–306, 2014.

[42] O. Billoint, H. Sarhan, I. Rayane, M. Vinet, P. Batude, C. Fenouillet-Beranger, O. Rozeau, G. Cibrario, F. Deprat, O. Turkyilmaz *et al.*, "From

2D to Monolithic 3D: Design Possibilities, Expectations and Challenges," in *Proceedings of the ACM International Symposium on Physical Design*, Mar. 2015, pp. 127–127.

[43] M. Vinet, P. Batude, C. Tabone, B. Previtali, C. LeRoyer, A. Pouydebasque, L. Clavelier, A. Valentian, O. Thomas, S. Michaud *et al.*, "3D Monolithic Integration: Technological Challenges and Electrical Results," *Microelectronic Engineering*, vol. 88, no. 4, pp. 331–335, Apr. 2011.

[44] P. Batude, M. Vinet, A. Pouydebasque, C. Le Royer, B. Previtali, C. Tabone, J.-M. Hartmann, L. Sanchez, L. Baud, V. Carron *et al.*, "Advances in 3D CMOS Sequential Integration," in *Proceedings of the IEEE International Electron Devices Meeting*, Dec. 2009, pp. 1–4.

[45] F. B. Claire, B. Previtali, P. Batude, F. Nemouchi, M. Casse, X. Garros, L. Tosti, N. Rambal, D. Lafond, H. Dansas *et al.*, "FDSOI Bottom MOSFETs Stability versus Top Transistor Thermal Budget Featuring 3D Monolithic Integration," *Solid-State Electronics*, vol. 113, pp. 2–8, Nov. 2015.

[46] B. Rajendran, R. S. Shenoy, D. J. Witte, N. S. Chokshi, R. L. DeLeon, G. S. Tompa, and R. F. W. Pease, "Low Thermal Budget Processing for Sequential 3D IC Fabrication," *IEEE Transactions on Electron Devices*, vol. 54, no. 4, pp. 707–714, Apr. 2007.

[47] C. Yan, Z. Gan, and E. Salman, "Package-Embedded Spiral Inductor Characterization with Application to Switching buck Converters," *Microelectronics Journal*, vol. 66, pp. 41–47, 2017.

[48] P. Batude, B. Sklenard, C. Fenouillet-Beranger, B. Previtali, C. Tabone, O. Rozeau, O. Billoint, O. Turkyilmaz, H. Sarhan, S. Thuries *et al.*, "3D Sequential Integration Opportunities and Technology Optimization," in *Proceedings of the IEEE International Interconnect Technology Conference*, May. 2014, pp. 373–376.

[49] "STMicroelectronics." [Online]. Available: https://www.st.com/content/st_com/en.html

[50] "IBM." [Online]. Available: https://www.ibm.com/us-en/

[51] "Monolithic 3D Inc." [Online]. Available: http://www.monolithic3d.com/

[52] C. Fenouillet-Beranger, B. Previtali, P. Batude, F. Nemouchi, M. Casse, X. Garros, L. Tosti, N. Rambal, D. Lafond, H. Dansas *et al.*, "FDSOI Bottom MOSFETs Stability Versus Top Transistor Thermal Budget Featuring 3D Monolithic Integration," in *Proceedings of European Solid State Device Research Conference*, 2014.

[53] F.-K. Hsueh, H.-Y. Chiu, C.-H. Shen, J.-M. Shieh, Y.-T. Tang, C.-C. Yang, H.-C. Chen, W.-H. Huang, B.-Y. Chen, K.-M. Chen *et al.*, "TSV-Free FinFET-Based Monolithic 3D+-IC with Computing-in-Memory SRAM Cell for Intelligent IoT Devices," in *Proceedings of the IEEE International Electron Devices Meeting*, Dec. 2017.

[54] C. Liu and S. K. Lim, "A Design Tradeoff Study with Monolithic 3D Integration," in *Proceedings of the IEEE International Symposium on Quality Electronic Design*, Mar. 2012, pp. 529–536.

[55] Y.-J. Lee, D. Limbrick, and S. K. Lim, "Power Benefit Study for Ultra-High Density Transistor-Level Monolithic 3D ICs," in *Proceedings of the ACM Design Automation Conference*, May. 2013, pp. 104:1–104:10.

[56] "FreePDK45." [Online]. Available: http://www.eda.ncsu.edu/wiki/FreePDK45:Contents

[57] S. A. Panth, K. Samadi, Y. Du, and S. K. Lim, "Design and CAD Methodologies for Low Power Gate-Level Monolithic 3D ICs," in *Proceedings of the ACM International Symposium on Low Power Electronics and Design*, Aug. 2014, pp. 171–176.

[58] "Nangate Cell Library." [Online]. Available: http://www.nangate.com

[59] "Mentor Graphics Calibre." [Online]. Available: https://www.mentor.com/products/icnanometerdesign/verificationsignoff/

[60] "Cadence Encounter Library Characterizer (ELC)." [Online]. Available: http://www.cadence.com/products/di/library-characterizer/pages/default.aspx

[61] "Synopsys HSPICE." [Online]. Available: https://www.synopsys.com/tools/Verification/AMSVerification/Circuit-Simulation/HSPICE/Pages/default.aspx

[62] "Synopsys Design Compiler." [Online]. Available: http://www.synopsys.com/Tools/Implementation/RTLSynthesis/Design-Compiler/Pages/default.aspx

[63] "Cadence Encounter." [Online]. Available: https://www.cadence.com/content/cadencewww/global/enUS/home/tools-/digitaldesignandsignoff/hierarchicaldesignandfloorplanning/innovus-implementationsystem.html

[64] C. Yan and E. Salman, "Routing Congestion Aware Cell Library Development for Monolithic 3D ICs," in *Proceedings of the IEEE International Conference on Rebooting Computing*, Nov. 2017.

[65] C. Yan, S. Kontak, H. Wang, and E. Salman, "Open Source Cell Library Mono3D to Develop Large-Scale Monolithic 3D Integrated Circuit," in *Proceedings of IEEE International Symposium on Circuits and Systems*, May. 2017, pp. 2581–2584.

[66] M. Tehranipoor and C. Wang, *Introduction to hardware security and trust*. Springer Science & Business Media, 2011.

[67] R. S. Chakraborty, S. Narasimhan, and S. Bhunia, "Hardware Trojan: Threats and Emerging Solutions," in *Proceedings of the IEEE International High Level Design Validation and Test Workshop*. IEE, 2009, pp. 166–171.

[68] M. Rostami, F. Koushanfar, and R. Karri, "A Primer on Hardware Security: Models, Methods, and Metrics," *Proceedings of the IEEE*, vol. 102, no. 8, pp. 1283–1295, 2014.

[69] M. Potkonjak, A. Nahapetian, M. Nelson, and T. Massey, "Hardware Trojan Horse Detection Using Gate-Level Characterization," in *Proceedings of the ACM Annual Design Automation Conference*, 2009, pp. 688–693.

[70] M. Majzoobi, F. Koushanfar, and M. Potkonjak, "Testing Techniques for Hardware Security," in *Proceedings of the IEEE International Test Conference*, 2008, pp. 1–10.

[71] R. P. Cocchi, J. P. Baukus, L. W. Chow, and B. J. Wang, "Circuit Camouflage Integration for Hardware IP Protection," in *Proceedings of the ACM/IEEE Design Automation Conference*, 2014, pp. 1–5.

[72] J. Rajendran, M. Sam, O. Sinanoglu, and R. Karri, "Security Analysis of Integrated Circuit Camouflaging," in *Proceedings of the ACM Computer and Communications Security*, Nov. 2013, pp. 709–720.

[73] R. P. Cocchi, L. W. Chow, J. P. Baukus, and B. J. Wang, "Method and Apparatus for Camouflaging a Standard Cell Based Integrated Circuit with Micro Circuits and Post Processing," Aug. 13 2013, US Patent 8,510,700.

[74] L. W. Chow, J. P. Baukus, B. J. Wang, and R. P. Cocchi, "Camouflaging a Standard Cell Based Integrated Circuit," Apr. 3 2012, US Patent 8,151,235.

[75] P. Gu, S. Li, D. Stow, R. Barnes, L. Liu, Y. Xie, and E. Kursun, "Leveraging 3D Technologies for Hardware Security: Opportunities and Challenges," in *Proceedings of the IEEE/ACM International Great Lakes Symposium on VLSI*, 2016.

[76] J. Dofe, C. Yan, S. Kontak, E. Salman, and Q. Yu, "Transistor-Level Camouflaged Logic Locking Method for Monolithic 3D IC Security," in *Proceedings of the IEEE Asian Hardware Oriented Security and Trust Symposium*, Dec. 2016.

[77] J. Dofe, Q. Yu, H. Wang, and E. Salman, "Hardware Security Threats and Potential Countermeasures in Emerging 3D ICs," in *Proceedings of the IEEE/ACM International Great Lakes Symposium on VLSI*, 2016.

[78] J. Dofe, Z. Zhang, Q. Yu, C. Yan, and E. Salman, "Impact of Power Distribution Network on Power Analysis Attacks in Three-Dimensional Integrated Circuits," in *Proceedings of the IEEE/ACM Internation Great Lakes Symposium on VLSI*, 2017.

[79] C. Yan and E. Salman, "Physical Design of Monolithic 3D ICs with Applications to Hardware Security," in *Proceedings of the Government Microcircuit Applications & Critical Technology Conference*, 2018.

[80] C. Yan, J. Dofe, S. Kontak, Q. Yu, and E. Salman, "Hardware-Efficient Logic Camouflaging for Monolithic 3-D ICs," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 6, pp. 799–803, 2018.

[81] A. Vijayakumar, V. C. Patil, D. E. Holcomb, C. Paar, and S. Kundu, "Physical Design Obfuscation of Hardware: A Comprehensive Investigation of Device- and Logic-Level Techniques," *IEEE Transaction. on Information Forensics and Security*, vol. 12, no. 1, Jan 2017.

[82] R. Beaulieu, S. Treatman-Clark, D. Shors, B. Weeks, J. Smith, and L. Wingers, "The SIMON and SPECK Lightweight Block Ciphers," in *Proceedings of the ACM/IEEE Design Automation Conference*, Jun. 2015.

[83] S. Im and K. Banerjee, "Full Chip Thermal Analysis of Planar (2-D) and Vertically Integrated (3-D) High Performance ICs," in *Proceedings of IEEE In ternatioinal Electron Devices Meeting*, 2000, pp. 727–730.

[84] M. M. Waldrop, "The Chips Are Down for Moores Law," *Nature News*, vol. 530, no. 7589, p. 144, 2016.

[85] C. Xu, S. K. Kolluri, K. Endo, and K. Banerjee, "Analytical Thermal Model for Self-Heating in Advanced FinFET Devices with Implications for Design and Reliability," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 32, no. 7, pp. 1045–1058, 2013.

[86] J. Van Olmen, A. Mercha, G. Katti, C. Huyghebaert, J. Van Aelst, E. Seppala, Z. Chao, S. Armini, J. Vaes, R. C. Teixeira *et al.*, "3D Stacked IC Demonstration Using a Through Silicon Via First Approach," in *Proceedings of the IEEE International Electron Devices Meeting*, 2008, pp. 1–4.

[87] S. Gu, P. Marchal, M. Facchini, F. Wang, M. Suh, D. Lisk, and M. Nowak, "Stackable memory of 3d chip integration for mobile applications," in *Proceedings of the IEEE International Electron Devices Meeting*, 2008, pp. 1–4.

[88] M. M. Shulaker, T. F. Wu, A. Pal, L. Zhao, Y. Nishi, K. Saraswat, H.-S. P. Wong, and S. Mitra, "Monolithic 3D Integration of Logic and Memory: Carbon Nanotube FETs, Resistive RAM, and Silicon FETs," in *Proceedings of the IEEE International Electron Devices Meeting*, 2014, pp. 27–4.

[89] Y. Ghidini, M. Moreira, L. Brahm, T. Webber, N. Calazans, and C. Marcon, "Lasio 3D NoC Vertical Links Serialization: Evaluation of Latency and Buffer Occupancy," in *Proceedings of the IEEE International Symposium on Integrated Circuits and Systems Design*. IEEE, 2013, pp. 1–6.

[90] M. Li, J. Shi, M. Rahman, S. Khasanvis, S. Bhat, and C. A. Moritz, "Skybridge-3D-CMOS: A Fine-Grained 3D CMOS Integrated Circuit Tech-

nology," *IEEE Transactions on Nanotechnology*, vol. 16, no. 4, pp. 639–652, 2017.

[91] K. Banerjee, S. J. Souri, P. Kapur, and K. C. Saraswat, "3-D ICs: A Novel Chip Design for Improving Deep-Submicrometer Interconnect Performance and Systems-on-Chip Integration," *Proceedings of the IEEE*, vol. 89, no. 5, pp. 602–633, 2001.

[92] P. Morrow, C.-M. Park, S. Ramanathan, M. Kobrinsky, and M. Harmes, "Three-Dimensional Wafer Stacking Via Cu-Cu Bonding Integrated With 65-nm Strained-Si/low-k CMOS Technology," *IEEE Electron Device Letters*, vol. 27, no. 5, pp. 335–337, 2006.

[93] S. Pozder, R. Chatterjee, A. Jain, Z. Huang, R. E. Jones, and E. Acosta, "Progress of 3D Integration Technologies and 3D Interconnects," in *Proceedings of the IEEE International Interconnect Technology Conference*, 2007, pp. 213–215.

[94] R. P. Cocchi, L. W. Chow, J. P. Baukus, and B. J. Wang, "Method and Apparatus for Camouflaging a Standard Cell Based Integrated Circuit with Micro Circuits and Post Processing," Aug. 13 2013, US Patent 8,510,700.

[95] M. Li, S. Khasanvis, J. Shi, S. Bhat, M. Rahman, and C. A. Moritz, "Towards Automatic Thermal Network Extraction in 3D ICs," in *Proceedings of the IEEE/ACM International Symposium on Nanoscale Architectures*, 2016, pp. 25–30.

[96] K. Athikulwongse, M. Ekpanyapong, and S. K. Lim, "Exploiting Die-to-Die Thermal Coupling in 3-D IC Placement," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22, no. 10, pp. 2145–2155, 2014.

[97] S. S. Sapatnekar, "Addressing Thermal and Power Delivery Bottlenecks in 3D Circuits," in *Proceedings of the IEEE Asia and South Pacific Design Automation Conference*, 2009, pp. 423–428.

[98] P. Emma, A. Buyuktosunoglu, M. Healy, K. Kailas, V. Puente, R. Yu, A. Hartstein, P. Bose, and J. Moreno, "3D Stacking of High-Performance Processors," in *Proceedings of the IEEE International Symposium on High Performance Computer Architecture*, 2014, pp. 500–511.

[99] C. A. Yunus and J. G. Afshin, "Heat and Mass Transfer: Fundamentals and Applications," *Tata McGraw-Hill, New Delhi, India*, 2011.

[100] "What is Heat." [Online]. Available: https://www.physicsclassroom.com/class/thermalP/Lesson-1/What-is-Heat

[101] M. N. Ozisik, "Heat Transfer: A Basic Approach," 1985.

[102] H. Sultan and S. R. Sarangi, "A Fast Leakage Aware Thermal Simulator for 3D Chips," in *Proceedings of the European Design and Automation Association Conference on Design, Automation and Test*, 2017, pp. 1737–1742.

[103] C. Yan, H. Zhu, D. Zhou, and X. Zeng, "An Efficient Leakage-Aware Thermal Simulation Approach for 3D-ICs Using Corrected Linearized Model and Algebraic Multigrid," in *Proceedings of the European Design and Automation Association Conference on Design, Automation and Test*, 2017, pp. 1207–1212.

[104] Y. Ma, K. Wang, S. Dong, Y. Wang, and X. Hong, "Thermal Effects of Leakage Power in 3D ICs," in *Proceedings of the IEEE international Conference on Green Circuits and Systems*, 2010, pp. 578–583.

[105] W.-L. Hung, G. M. Link, Y. Xie, N. Vijaykrishnan, and M. J. Irwin, "Interconnect and Thermal-Aware Floorplanning for 3D Microprocessors," in *Proceedings of the IEEE International Symposium on Quality Electronic Design*, 2006, pp. 6–pp.

[106] S. Ladenheim, Y.-C. Chen, M. Mihajlovic, and V. F. Pavlidis, "IC Thermal Analyzer for Versatile 3-D Structures Using Multigrid Preconditioned Krylov Methods," in *Proceedings of the IEEE international Conference on Computer-Aided Design*, 2016, pp. 1–8.

[107] R. Trihy, "Addressing Library Creation Challenges from Recent Liberty Extensions," in *Proceedings of the ACM/IEEE Design Automation Conference*, 2008, pp. 474–479.

[108] J. Bhasker and R. Chadha, *Static Timing Analysis for Nanometer Designs: A Practical Approach*. Springer Science & Business Media, 2009.

[109] N. H. Weste, D. Harris, and A. Banerjee, *CMOS VLSI Design*. Pearson India, 2010.

[110] S.-M. Kang and Y. Leblebici, *CMOS Digital Integrated Circuits*. Tata McGraw-Hill Education, 2003.

[111] S. Mukhopadhyay, A. Raychowdhury, and K. Roy, "Accurate Estimation of Total Leakage Current in Scaled CMOS Logic Circuits Based on Compact

Current Modeling," in *Proceedings of the ACM Design Automation Conference*, 2003, pp. 169–174.

[112] S. Mukhopadhyay, C. Neau, R. T. Cakici, A. Agarwal, C. H. Kim, and K. Roy, "Gate Leakage Reduction for Scaled Devices Using Transistor Stacking," *IEEE Transactions on Very Large Scale Integration Systems*, vol. 11, no. 4, pp. 716–730, 2003.

[113] G. Giustolisi, G. Palumbo, M. Criscione, and F. Cutri, "A Low-Voltage Low-Power Voltage Reference Based on Subthreshold MOSFETs," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 1, pp. 151–154, 2003.

[114] N. Kimizuka, T. Yamamoto, T. Mogami, K. Yamaguchi, K. Imai, and T. Horiuchi, "The Impact of Bias Temperature Instability for Direct-Tunneling Ultra-Thin Gate Oxide on MOSFET Scaling," in *Proceedings of IEEE International Symposium on VLSI Technology*, 1999, pp. 73–74.

[115] W. Huang, S. Ghosh, S. Velusamy, K. Sankaranarayanan, K. Skadron, and M. R. Stan, "HotSpot: A Compact Thermal Modeling Methodology for Early-Stage VLSI Design," *IEEE Transactions on Very Large Scale Integration Systems*, vol. 14, no. 5, pp. 501–513, 2006.

[116] A. Sridhar, A. Vincenzi, M. Ruggiero, T. Brunschwiler, and D. Atienza, "3D-ICE: Fast Compact Transient Thermal Modeling for 3D ICs with Inter-Tier Liquid Cooling," in *Proceedings of the IEEE International Conference on Computer-Aided Design*, 2010, pp. 463–470.

[117] T.-Y. Wang and C. C.-P. Chen, "3-D Thermal-ADI: A Linear-Time Chip Level Transient Thermal Simulator," *IEEE Transactions on Computer-Aided*

Design of Integrated Circuits and Systems, vol. 21, no. 12, pp. 1434–1445, 2002.

[118] A. Fourmigue, G. Beltrame, and G. Nicolescu, "Efficient Transient Thermal Simulation of 3D ICs with Liquid-Cooling and Through Silicon Vias," in *Proceedings of the IEEE Europe conference on Design, Automation & Test*. European Design and Automation Association, 2014, p. 74.

[119] Y.-C. Chen, S. Ladenheim, H. Kalargaris, M. Mihajlovic, and V. F. Pavlidis, "Computationally Efficient Standard-Cell FEM-Based Thermal Analysis," in *Proceedings of the IEEE international Conference on Computer-Aided Design*, 2017, pp. 490–495.

[120] F. Brglez, D. Bryan, and K. Kozminski, "Combinational Profiles of Sequential Benchmark Circuits," in *Proceedings of the IEEE International Circuits and Systems*, 1989, pp. 1929–1934.

[121] "Design Exchange Format." [Online]. Available: https://www.vlsi-exoert.com/2015/09/DEF.html

[122] "Cadence Genus." [Online]. Available: https://www.cadence.com/content/cadence-www/global/en_US/home/tools/digital-design-and-signoff/synthesis/genus-synthesis-solution.html

[123] "Value Change Dump." [Online]. Available: https://en.wikipedia.org/wiki/Value_change_dump

[124] J. Becker, M. Huebner, and M. Ullmann, "Power Estimation and Power Measurement of Xilinx Virtex FPGAs: Trade-Offs and Limitations," in *Proceed-*

ings of the IEEE International Integrated Circuits and Systems, 2003, pp. 283–288.

[125] "Synopsys Primetime PX." [Online]. Available: https://www.synopsys.com/support/training/signoff/primetimepx-fcd.html

[126] "Synopsys Waveview." [Online]. Available: https://www.synopsys.com/verification/ams-verification/waveform-analysis-debug/custom-waveview.html

[127] J. Daemen and V. Rijmen, *The Design of Rijndael: AES-the Advanced Encryption Standard*. Springer Science & Business Media, 2013.

[128] "Advanced Encryption Standard." [Online]. Available: https://en.wikipedia.org/wiki/Advanced_Encryption_Standard

[129] "AES Core Modules." [Online]. Available: https://opencores.org/project/aes_128_192_256