Spring 5-18-2019

Performance-Aware and Power-Efficient Three Dimensional (3D) Integrated Circuit (IC) Design Utilizing Evolutionary Algorithms

Nahid Mirzaie
Southern Methodist University, nmirzaie@smu.edu

PERFORMANCE-AWARE AND POWER-EFFICIENT THREE DIMENSIONAL (3D) INTEGRATED CIRCUIT (IC) DESIGN UTILIZING EVOLUTIONARY ALGORITHMS

Approved by:

_____

Prof. Duncan L. MacFarlane
Professor of Electrical and Computer Engineering, Advisor

_____

Prof. Ronald A. Rohrer
Professor of Electrical and Computer Engineering

_____

Prof. Jennifer Dworak
Associate Professor of Electrical and Computer Engineering

_____

Prof. Joseph Camp
Associate Professor of Electrical and Computer Engineering

_____

Prof. Mohammad Khodayar
Assistant Professor of Electrical and Computer Engineering

PERFORMANCE-AWARE AND POWER-EFFICIENT THREE DIMENSIONAL (3D) INTEGRATED CIRCUIT (IC) DESIGN UTILIZING EVOLUTIONARY ALGORITHMS

A Dissertation Presented to the Graduate Faculty of

Lyle School of Engineering

Southern Methodist University

in

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

with a

Major in Electrical and Computer Engineering Department

by

Nahid Mirzaie

May 18, 2019

ACKNOWLEDGMENTS

I would like to begin by thanking my former advisor, Dr. Gyung-Su Byun, for the opportunity to perform this research. I am deeply grateful to my current advisor Dr. Duncan L. MacFarlane, for his continious guidance and support. I was also fortunate to work with Dr. Ronald A. Rohrer who was an amazing advisor; his patience and valuable comments helped me as I concluded my Ph.D. research. Profound thanks to the members of my Ph.D. dissertation committee: Dr. Jennifer Dworak, Dr. Joseph Camp, and Dr. Mohammed Khodayar, for providing insightful comments and feedback during the final stages of this dissertation.

I owe my deep gratitude to my amazing friends: Nathan Palkki, Arefeh Safaei Moghadam, Lael Badakhshanian, Shadi and Donya Zaheri, Nadere Mansouri, Hedie Ashrafi, Mahdi Heidarizad, Amin Mansoorifar, and Bardia Heidari who have always been supportive like a family, when I was thousands of miles away from home.

I am grateful to my parents and sisters, Elham, Mahtab, Nasim, and Ghazal for their devotion all throughout my life. They always wanted me to dream big, work hard, and achieve my goals. Profound thanks to my brother-in-law, Ali, for his sincere support and help when I was experiencing a difficult time in my life.

PUBLICATIONS

[1] **N. Mirzaie** and G. S. Byun, "3D Pipeline ADC Utilizing TSV/Design Optimization and Memristor Ratioed Logic," *IEEE Transactions on Very Large-Scale Integration Systems (TVLSI)*, 2018.

[2] **N. Mirzaie** and G. Byun, "An Optimal Design Methodology for Yield-Improved and Low-Power Pipelined ADC," *IEEE Transactions on Semiconductor Manufacturing (TSM)*, vol. 31, no. 1, 2018.

[3] A. Alzahmi, **N. Mirzaie**, and G. S. Byun, "3D Power Delivery Network's Sub-Blocks and Regulator Placement Optimized by Evolutionary Algorithm," *IEEE Transactions on Components, Packaging and Manufacturing Technology (TCPMT)*, 2017.

[4] **N. Mirzaie**, H. Shamsi, and G. S. Byun, "Resilient design of current steering DACs using a transistor level approach," *Analog Integrated Circuits and Signal Processing* 9, vol. 90, no. 1, pp. 29-41, 2017.

[5] **N. Mirzaie**, H. Shamsi, and G. S. Byun, "Yield-aware sizing of pipeline ADC using a multiple-objective evolutionary algorithm," *International Journal of Circuit Theory and Applications (IJCTA)*, vol. 45, no. 6, 2017.

[6] **N. Mirzaie**, H. Shamsi, and G. S. Byun, "Automatic design and yield enhancement of data converters," *Journal of Circuits, Systems and Computers,* vol.26, no. 01, 2017.

[7] **N. Mirzaie**, A. Alzahmi, C. Lin, and G. Byun, "A Performance-Aware I/O Interface for 3D Stacked Memory Systems," *IEEE Ubiquitous Computing, Electronics & Mobile Communication Conf.*, New York City, 2017.

[8] **N. Mirzaie**, A. Alzahmi, C. Lin, and G. S. Byun, "A Low-Power and Performance-Efficient SAR ADC Design," *IEEE International SoC Design Conf. (ISOCC)*, Korea, Sept. 2017.

[9] **N. Mirzaie**, A. Alzahmi, C. Lin, and G. S. Byun, “A 3D Flash ADC Structure for High-Speed Communication Applications,” *The 8th IEEE Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, NV, 2018.

[10] A. Alzahmi, **N. Mirzaie**, C. Lin, and G. S. Byun, "Low-Power and High-Performance 2.4 GHz RF Transmitter for Biomedical Application," *IEEE Ubiquitous Computing, Electronics & Mobile Communication Conf. (UEMCON)*, New York City, NY, 2017.

[11] A. Alzahmi, **N. Mirzaie**, C. C. Lin, and G. S. Byun, "High-Performance RF-Interconnect for 3D Stacked Memory," *IEEE International SoC Design Conf. (ISOCC)*, Jeju, Korea, Sept. 2017.

[12] C. Lin, **N. Mirzaie**, A. Alzahmi, and G. S. Byun, "Design of a Pre-distortion Power Amplifier for Ku-Band/5G. applications, " *IEEE International Electro/Information Technology Conference (EIT) 2018*, Rochester, 2018.

[13] **N. Mirzaie**, A. Alzahmi, and G. S. Byun, "Reliability-Aware 3D Clock Distribution Network Using Memristor Ratioed Logic," *IEEE Transactions on Very Large-Scale Integration Systems (TVLSI)*, 2018 (under review).

Nahid, Mirzaie

B.S., University of Kashan, Kashan, 2007
M.S., Iran University of Science and Technology, 2009

Performance-Aware and Power-Efficient Three Dimensional (3D) IC Design Utilizing Evolutionary Algorithms

Advisor: Professor Duncan L. MacFarlane

Doctor of Philosophy conferred May 18, 2019

Dissertation completed January 28, 2019

High-performance, low-power, and high-accuracy circuit design with a compact silicon area is necessary for a wide range of current applications. Three-dimensional integrated circuits (3D ICs) can be a promising solution that overcomes the technical barriers in high-performance communication systems by using their capabilities of chip area reduction, improved signal integrity, and reduced routing complexity. In addition to the 3D integration, to achieve the optimum power efficiency, signal quality, and minimum supply noises on a heterogeneous stacking, utilizing computational intelligence methods by considering all design parameters (e. g., TSV and transistor sizing and placement) can be significantly favorable.

This dissertation presents multiple novel 3D IC designs in digital and mixed-signal systems with their fabricated chip prototypes. An evolutionary approach has also been utilized on the design to further improve the circuit performance. Exploiting the multi-objective evolutionary algorithms on the circuit sub-blocks and TSVs has provided the overall system specification enhancement regarding performance, power efficiency, delay, noise, and silicon area for the proposed 3D architectures.

The implemented novel 3D prototypes include analog to digital converter (ADC), clock distribution network (CDN), power distribution network (PDN), and high speed I/O memory

interface. The proposed on-chip 3D architectures are analyzed, designed, and fabricated in 65nm complementary metal–oxide–semiconductor (CMOS) technologies at 1.0 V. The measured results show utilizing both 3D IC integration and evolutionary algorithms have significantly improved several important circuit factors regarding performance, power, jitter, latency, accuracy, chip area, and speed in all the implemented prototypes.

TABLE OF CONTENTS

List of Figures

List of Tables

Acronyms

Abbreviation	Description
3D	Three-dimensional
2D	Two-dimensional
ADC	Analog to digital converter
CDN	clock distribution network
CMOS	complementary metal–oxide–semiconductor
CPU	central processing unit
DEC	digital error correction
DRAM	Dynamic Random-Access Memory
ENOB	Effective number of bits
FFT	Fast Fourier transform
FoM	Figure-of-merit
GA	Genetic algorithm
GS	Gain stage
HBM	High bandwidth memory
HFSS	High frequency structure simulator
HPC	High-performance computing
HSTL	High-speed transceiver logic
I/O	Input/output
IC	Integrated Circuit
IMD	inter-metal dielectric
IRC	International Roadmap Committee
ISI	Inter-symbol interference
JEDEC	Joint electron device engineering council

Abbreviation	Description
LDO	Low-dropout voltage regulator
LPDDR	Low power double data rate
MD	Multi-drop
MDAC	Multiplying digital-to-analog converter
MEMS	Micro-Electro-Mechanical Systems
MRL	memristor ratioed logic
NMOS	N-type metal-oxide-semiconductor
ODT	On-die termination
OTA	Operational transconductance amplifier
p2p	point to point
PCB	Printed circuit board
PDN	power distribution network
PLL	phase-locked loop
PMOS	P-type metal-oxide-semiconductor
PVT	Process, voltage, and temperature
RF	Radio frequency
RMS	Root mean square
RX	Receiver
S/H	Sample and hold
SFDR	Spurious-free dynamic range
SNDR	Signal to noise and distortion ratio
SOC	system on chip
SSTL	Stub Series Terminated Logic
TL	Transmission line
TSI	Through-silicon interposers

Abbreviation	Description
TSV	Through-silicon via
TX	Transmitter
UGBW	Unity gain bandwidth
VLSI	very large system integration

To my beloved parents, Maryam and Hassan,

who have made great sacrifices and provided their love and support throughout the journey of my life.

Chapter 1

INTRODUCTION

1. 1 Introduction

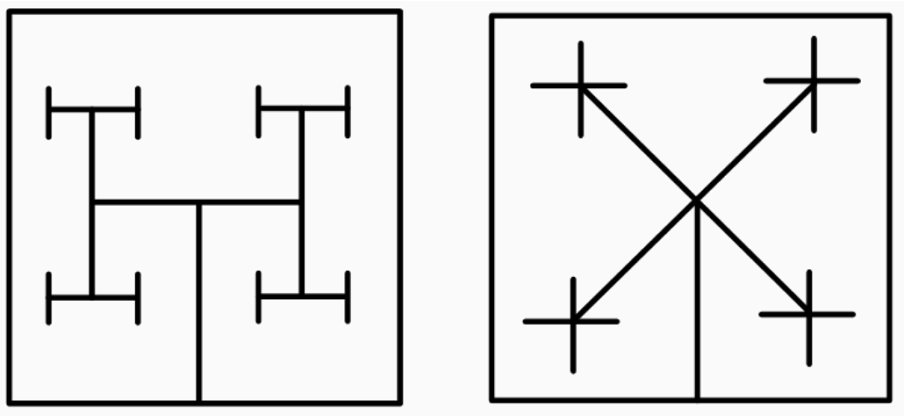
The continuous scaling of advanced CMOS technologies makes interconnection a key limiting factor in latency and power/area efficiency. Three-dimensional integrated circuit (3D IC) can be a promising solution that overcomes the technical barriers in high-performance communication systems by using their capabilities of addressing fundamental limitations in on-chip interconnects [1], [2].

Recently, some analog/mixed-signal IC prototypes have explored 3D design advantages in areas such as stacked memory chips, memories stacked on central processing unit (CPU) [3], [4], co-integration of antenna and radio frequency micro-electro-mechanical systems (RF-MEMS) [5], [6], and image sensor layer stacking on the memory and CPU layers [7], [8]. However, to date, 3D analog to digital converter (ADC) design has not been investigated extensively because of the ADC complicated structure compared to other mixed-signal IC prototypes [9], [10]. As ADCs (e.g. pipeline ADC) include different types of analog/digital sub-blocks such as sub-ADC, sub-DAC, digital error correction, sample and hold, amplifier, comparator, etc.

In a conventional two dimensional (2D) ADC design, one of the major challenges is

routing congestion on the die leading to increased latency and degraded signal integrity due to the long metal interconnections. Also, the conventional ADC prototype has the lack of analysis on inherent key design challenges such as clock distribution network (CDN) latency, power overhead, and on-chip routing complexity. Thus, a considerable amount of research concerning CDN has already been carried out to reduce the skew and jitter in 2D IC [11], [12]. For example, some proposed techniques such as H-tree and X-tree can provide clock distribution symmetry [13], [14]. In H-tree approach, the primary clock driver is connected to the center of the main "H" structure. The clock signal is transmitted to the four corners of the main "H." These four close to identical clock signals provide the inputs to the next level of the H-tree hierarchy, represented by the four smaller "H" structures. The distribution process continues through several levels of progressively smaller "H" structures. The final destination points of the H-tree are used to drive the local registers or are amplified by local buffers which drive the local registers. Thus, each clock path from the clock source to a clocked register has practically the same delay. X-tree clock distribution network also consists of a global clock interconnect similar to the H-tree network which branches out in symmetrical X type topology. This branching continues until the registers are reached as shown in Figure 1.1.

However, H/X-tree clock networks cause more design complexity, power consumption, and die area overhead in ADC design and can not address the mentioned current technology scaling issues, properly.



(a) (b)

Figure 1.1 Two conventional structures for clock distribution networks (a) H-tree (b) X-tree

In addition to the ADC design challenges, advances in mobile devices such as smartphones demand to elaborate and improve computation capabilities and video processing to achieve greater aggregate bandwidth and higher energy efficiency [15]. For instance, the Samsung Galaxy Note 7 and OnePlus 3 are equipped with 4 GB and 6 GB DRAM, respectively [16].

Multiple chip package (MCP) technology can stack several chips into a single package, offering increased spatial density and performance benefits, while reducing overall power consumption. This enables designers to pack more functionality into a smaller form factor, facilitating the development of smaller electronic devices. By utilizing MCP technology in the future of very large system integration (VLSI), processors shift to multi-core architectures to increase the CPU performance [17]. In spite of the processor improvements, the bottleneck between a CPU core and a DRAM is still existing and may be alleviated by increasing the data rate and number of I/O pins between the CPU core and the DRAM. In Figure 1.1 is shown a schematic of two-dimensional (2D) CPU core-DRAM

link (i. e. memory interface) including N I/O into/from a channel arranged on a printed circuit board (PCB).

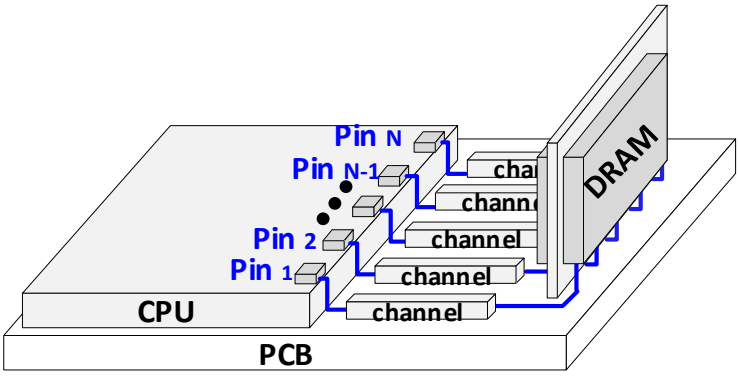


Figure 1.2 Memory interface for smart mobile phone devices

The key issue of a 2D mobile memory interface is bandwidth limitation due to multi-drop channel impedance discontinuities and degraded signal integrity. The increasing demand for low power DRAM has driven joint electron device engineering council (JEDEC) to standardize low power double data rate (LPDDR) for mobile platforms. Mobile DDR is a type of 2D double data rate synchronous DRAM for mobile computers. The complete name is low power double data rate synchronous dynamic random access memory (LPDDR-SDRAM) which emerged as the mainstream choice for most of the mobile devices. There are three generations of LPDDRs standardized by JEDEC calling LPDDR2, LPDDR3, and LPDDR4.

- LPDDR2 can have up to 0.8 Gb/s/pin data rate. It also improved the design of as smartphones, tablets, personal digital assistants, global positioning system (GPS) devices, and other mobile devices by increasing memory density, enhancing performance, shrinking size, reducing power consumption as well as increasing

battery life.

- LPDDR3 was designed to fulfill the performance and address the demands for more memory density in the latest mobile device generation such as notebooks, tablets, smartphones, and other connected devices on the 4G networks. LPDDR3 can offer higher bandwidth, boosted power efficiency, and denser memory over its innovative predecessor, LPDDR2. LPDDR3 can have the same quality in power-efficiency and signaling interface as LPDDR2, which allows a fast clock stop/start, smart array management, and low-power self-refresh.
- LPDDR4 is defined as the fourth generation of LPDDR which helps device developers have no concern in terms of power consumption in handset applications. It also offers improved performance and cost. LPDDR4 is designed to considerably enhance memory data rate and make mobile devices more efficient. The eventual goal for LPDDR4 id operating at an I/O rate twice that of LPDDR3. To achieve this goal, its topology is totally redesigned. It was changed from a one-channel die with 16 bits per channel to a two-channel die with 16 bits per channel and totally of 32 bits.

In addition to LPDDR, wide I/O is another current memory interface. Wide I/O provides a considerable speed improvement over LPDDR2 while preserving vertically through silicon via (TSV) interconnects and optimized packaging. Combined, these characteristics creat wide I/O to provide the ever-increasing speed, and power efficiency required by any type of mobile devices. Wide I/O delivers memory bandwidth four-time as much as the previous version of the standard, with lower power consumption.

Wide I/O Mobile DRAM is a revolution in memory technology that can address industry

demands for more integratin as well as enhanced performance, bandwidth, delay, power efficiency, and packaging form factor. Wide I/O Mobile DRAM uses the 3D stacking with TSV interconnections and memory chips directly stacked upon a system on chip (SoC). According to the International Roadmap Committee (IRC), the number of CPU cores and each core's frequency are increasing by 1.4x and 1.05x per year, respectively, as shown in Figure 1.3. The CPU performance is also projected to increase 1000x in 2024 compared to 2009 [18]. Nevertheless, the I/O pin data rate is expected to be only improved by 10x in 2024 compared to 2009 due to the channel loss and crosstalk, and the number of I/O pins increase 2x or 4x at the same period [18], [19]. Thus, the I/O pin bandwidth, i.e., data rate per pin between cores and memories, and the CPU performance need to both scale at the same rate to achieve the best possible data transfer performance. Moreover, the energy reduction per bit of I/O transceiver has a much slower rate compared with the projected I/O pin bandwidth for future mobile memory advances [17]. Thus, significant improvement in energy efficiency and data rate for future advances in mobile memory is required.

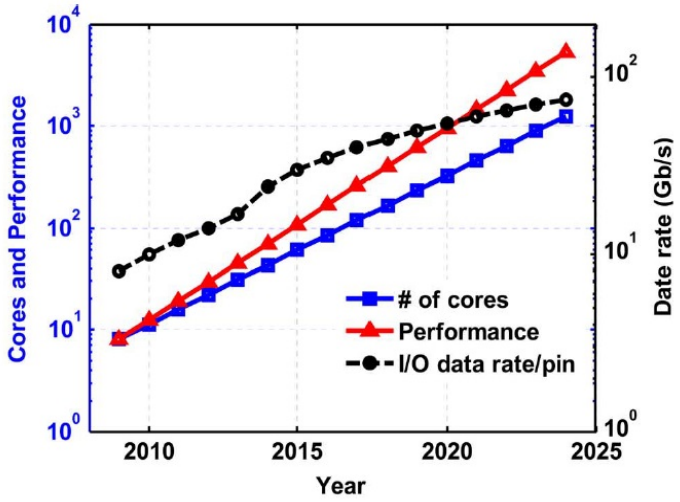


Figure 1.3 Projected number of cores, performance, and I/O data rate for future application [20]

1. 2 Research contribution

This research has concentrated on addressing some of the most challenging issues in digital and mixed-signal IC design, such as analog to digital converter (ADC) and memory interface, by investigation and considering 3D integration for each circuit. The research includes design and implementation of 3D CDN in a pipeline ADC, design and development of a reliable 3D PDN in a baseband memory interface, and finally a high-speed and energy efficient 3D I/O interface design.

In addition to the 3D integration, a novel optimization algorithm based on a multi-objective evolution has been developed and applied to provide substantial design improvement. The 3D designed and fabricated circuits are as follows.

1.2.1 3D pipeline ADC

High-performance, low-power, and high-accuracy analog-to-digital converters (ADCs) with compact die areas are necessary for a wide range of current applications. Among various ADC architectures, the pipeline converter is widely used in Nyquist sampling applications that require a combination of high resolution and high throughput. The second chapter of this dissertation describes a prototype design of pipeline ADC architecture with a novel three-dimensional (3D) clock distribution network (CDN) utilizing through-silicon via (TSV)-induced benefits. It also implements memristor ratioed logic (MRL) as the basic elements of digital error correction (DEC) sub-block to further decrease the chip area, delay, and power consumption. Additionally, an optimization technique using computational intelligence is applied to maximize the overall system performance. The

proposed prototype of a 10-bit 3D pipeline ADC is designed in a 65nm digital CMOS technology and shows significant improvement regarding dynamic performance, energy efficiency, silicon area, and clock accuracy as compared with conventional two-dimensional (2D) ADC designs.

1.2.2 3D power delivery network in a memory interface architecture

Chapter 3 of this dissertation presents a 3D power delivery network (PDN) using a novel common-centroid regulator placement to build a reliable I/O interface. To analyze the main benefits of the proposed on-chip 3D PDN, the regulator, most crucial block, is integrated into the PDN. Through-silicon via (TSV) and I/O links are also utilized to analyze the power noise accurately. Moreover, by utilizing an evolutionary algorithm, a set of interesting design points namely Pareto fronts are generated for both TSV and LDO regulator that allows a designer to select a suitable solution that meets power, performance, and other PDN goals based on the circuit application. The proposed 3D PDN can significantly improve energy efficiency, supply noises, and power/signal quality of the 3D I/O.

1.2.3 3D and energy efficient baseband transceiver for I/O interface

An energy and performance-aware 3D baseband transceiver utilizing CMOS driver for memory interface applications is described in chapter 4. The 3D integration is exploited utilizing through-silicon vias (TSVs) to attain the high-performance and low-power. Exploiting the 3D integration technology allows for the direct placement of memory on top of a microprocessor, significantly reducing the wire-delay between the two, and thereby

alleviating memory latency and bandwidth constraints.

The existing and future trends of memory interface architecture are shown in Figure 1.4. To increase the bandwidth of memory especially in high-performance computing and video/graphics applications, the I/O pin data rate should increase [21]. However, it is very challenging to increase the data rate beyond 10 Gb/s/pin. For instance, LPDDR3 and LPDDR4 accomplish 1.6 Gb/s and 4.2 Gb/s, respectively. The LPDDR3 and LPDDR4 utilize conventional two-dimensional (2D) point to point memory interface as shown in Figure 1.4 (a) [22]. 2D multi-drop memory interface which implements multiple DRAMs could be installed for higher bandwidth in the emerging applications such as cloud computing and data mining. However, the multi-drop memory interface increases power consumption and leads to routing congestion on PCBs as shown in Figure 1.4(b). 2D multi-drop channels also suffer from a high capacitive load. For example, the transmitter (TX) in a multi-drop structure is large regarding transistor sizing compared to a point-to-point to support driving a considerable load. Thus, the multi-drop I/O interface consumes significant power compared to the point-to-point because of the huge distance that the data travels between CPU and memory which deteriorate the bandwidth. Recent point-to-point and multi-drop I/O interfaces achieve 5.5 Gb/s/pin at 7.8 mW and 5.8 Gb/s/pin at 14.198 mW, respectively [23], [24].

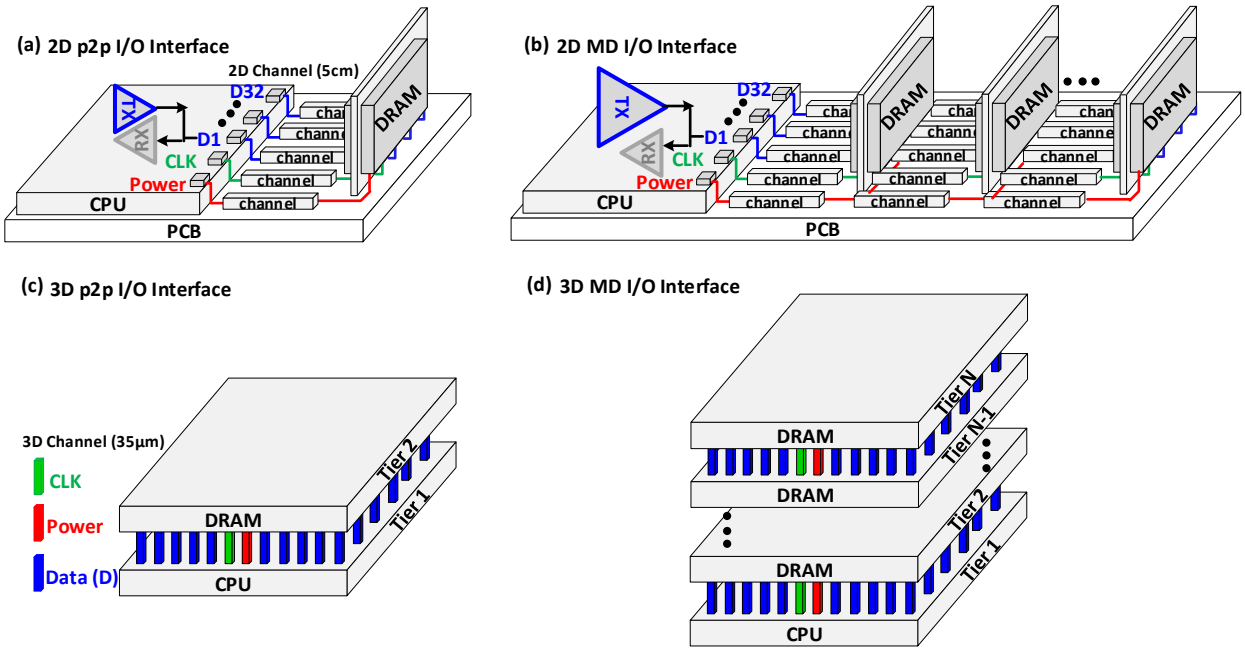


Figure 1.4 Memory interface structure trends

To obtain high energy efficiency, the I/O transceivers need to operate at high frequency and low power [25]. However, inter-symbol interference (ISI) and high-frequency distortion decrease the maximum operating frequency of 2D long transmission line (T-Line) (e.g., 5-30cm) that used in point-to-point and multi-drop I/O[26].

Increasing the number of I/O pins could be a solution to increase the performance by utilizing 3D integration such as wide I/O or high bandwidth memory (HBM) as shown in Figure 1.4(c) and (d), respectively. For example, the number of I/O pins are 32, 512, and 1024 for conventional DRAM, wide I/O, and HBM, respectively [27], [28]. 3D integration enables multiple stacking of chips vertically using very short channel (e.g., 50 – 200 µm), through-silicon-via (TSV) [29]–[31]. 3D IC design also significantly improves bandwidth/power efficiency, and decreases latency, noise, and chip area by minimizing signal path lengths,.

1. 3 Conclusion

The goal of this research is developing an array of energy efficient and high-performance 3D circuit designs in memory interface and mixed-signal systems. This dissertation presents novel approaches for three-dimensional pipeline ADC, power delivery network, and I/O interface to address their most important design challenges such as high power consumption, delay, noise, jitter, and improve their performance specifications.

Chapter 2

3D PIPELINE ADC UTILIZING TSV/DESIGN OPTIMIZATION AND MRL

2. 1 Introduction

Pipeline ADCs implement a large number of critical clock phases (i.e., non-overlapping sampling clocks, MDAC clock, and sub-ADC clock) to synchronize all gain stages. Hence, when the clock signal paths to the gain stages which should be synchronous, the clock skew must be minimized in the ADC design. Some techniques such as H-tree and X-tree can provide clock distribution symmetry as explained in section 1.1 [13], [14], [32]. However, H/X-tree networks increase the complexity of synchronous SOC systems and consequently the complexity of the clock network and hence it increases the clock power even if the clock frequency may not scale anymore. Therefore, the main fraction of the total power consumption of highly synchronous systems, e.g. high-resolution pipeline ADCs, would be due to the clock network.

In a pipeline ADC with unsynchronized sampling clocks, the jitter and noise reduce the signal to noise and distortion ratio (SNDR), introduce spurious emissions, and thereby further limit the spurious-free dynamic range (SFDR) [33]–[35]. Thus, the key

performance parameters such as SNDR and SFDR can become critical limitations of sampling accuracy when the routing length and complexity increase. While depends on various factors, the effect of broadband jitter on the clock signal is given as (2.1).

$$SNDR_{JITTER} = -20log_{10}(2\pi f_{IN} * T_{JITTER}) \quad (2.1)$$

where f_{IN} is the input frequency and T_{JITTER} is the jitter in Root mean square root mean square (RMS). Figure 2.1 (a) shows a long and complex on-chip clock routing in a conventional pipeline ADC. Since the clock signal is a sampling time reference in all gain stages, it has to be distributed throughout the die using metal interconnects [36]. As a consequent, for a high-resolution pipeline ADC, the clock delay and skew/jitter to the gain stages, that must be accurately synchronized, increase relatively, as shown by the red solid line in Figure 2.1 (b) [37]. The created clock skew/jitter can directly affect the maximum converter speed. However, the proposed 3D ADC can significantly decrease the clock jitter which is often dominated by the white noise floor of the clock signal.

In addition to the 3D implementation, memristive devices can further extend the capabilities of ADC performance. The memristor can be described as a 2-port passive element with variable resistance (memristance) [38], [39]. Figure 2.1 (c) shows the memristor symbol, model, and key characteristics of the current-voltage curve. The thick black line on the symbol shows the negative polarity of the memristor. When the current flows from the negative to the positive polarity, the memristance increases and follows the R_{OFF} line on the curve; while in the positive to the negative polarity of the current flow, the memristance decreases and follows the R_{ON} line on the curve. In this work, a hybrid CMOS-memristor logic family, namely MRL, has been utilized in the digital blocks of the

pipeline ADC to decrease the clock delay further and increase the logic density [39], [40]. By incorporating the MRL into the proposed 3D ADC, we can also decrease the power consumption, interconnection complexity, and parasitic capacitance.

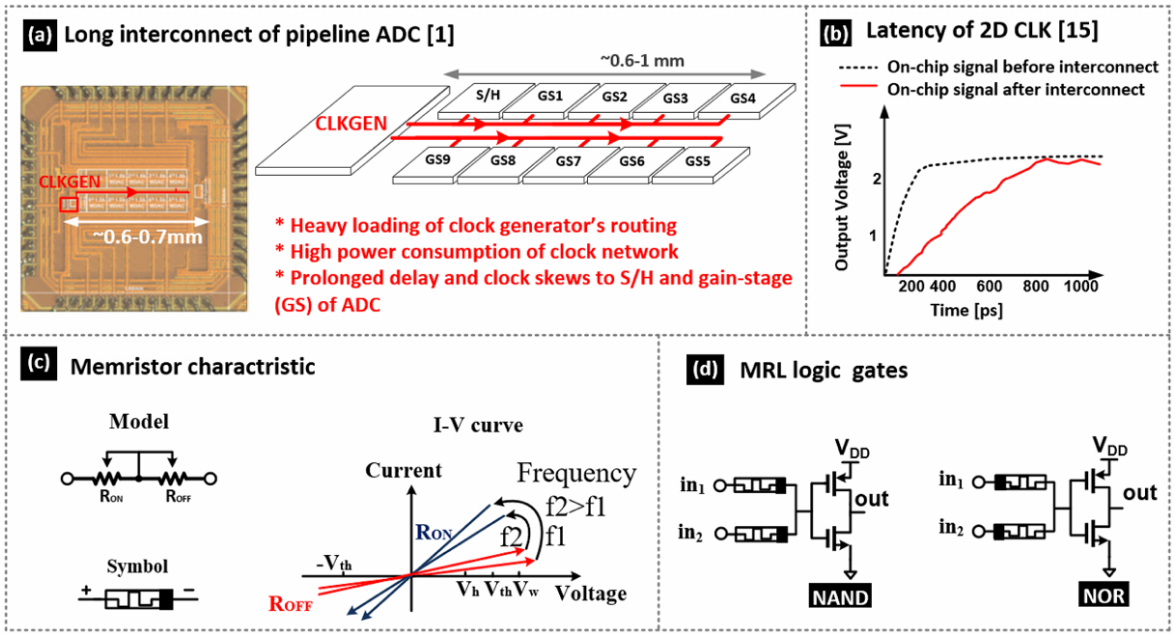


Figure 2.1 2D pipeline ADC interconnection issues and MRL model
a) Conventional 2D pipeline ADC with considerable power consumption, clock delay and heavy clock loading b) A sample 2D clock after passing through the on-chip track c) Memristor symbol, model and characteristic d) MRL logic basic elements

Figure 2.1 (d) shows the MRL basic logic gates. In MRL, OR and AND logic gates are based on memristors, and CMOS inverters are added to improve the logic function and create the output signal restoration [41]. Both OR and AND logic gates are composed of two series connected memristors that have opposite polarity. MRL occupies 67.6% and 75% reduced silicon area in NAND and NOR logic configurations, respectively. As a result, the propagation delay for a full adder, utilizing the MRL-based logic gates, can be decreased by about 30% which can improve the speed of the ADC digital sub-blocks. Therefore, MRL can lower the number of transistors on the chip for the same functionality

to improve the speed, decrease power consumption, and save the die area.

This chapter analyzes and proposes a novel ADC design exploiting 3D CDN, MRL logic, and computational intelligence. The proposed scheme is designed and analyzed on a 10-bit pipeline ADC as a prototype because the pipelined architecture has most successfully fulfilled the features of advanced CMOS technologies [36], [42]. Utilizing a hierarchical block-based evolutionary algorithm has maximized the overall ADC performance such as dynamic specifications, power, accuracy, and silicon area [43]. To the best of author's knowledge, the proposed 3D pipeline ADC design in which MRL logic and evolutionary algorithm are incorporated to obtain the optimal design point and inter-tier routing has never been investigated.

Although the proposed MRL/3D integration approach looks to arouse the extensive interest of industry and academia and have a promising contribution to future high-speed ADC generation, the manufacturing technology is not yet common. Thus, commercially available memristor circuitry is not expected to exist in the near future.

The rest of this chapter is organized as follows. Section 2.2 discusses the proposed 3D ADC design architecture. Section 2.3 presents the implemented TSV/wire model and MRL logic. The evolutionary approach is discussed in chapter 2.4 to find the optimal transistor sizing and TSV geometry. Section 2.5 shows the simulation and evaluated results. The conclusion and the suggested future work are explained in section 2.6.

2. 2 Introducing the proposed 3D ADC structure

Figure 2.2 (a) shows a 10-bit pipeline architecture considered for the proposed 3D

implementation. This architecture is composed of a sample and hold (S/H) block, four 2.5-bit gain stages followed by a 2-bit back-end flash ADC and DEC block.

In a pipeline ADC, quantization is broken down into multiple steps implemented by a cascade of gain stages. By inserting a sample-and-hold block at the beginning of the conversion and also each gain stage, all the gain stages can perform the conversion concurrently. This allows a conversion throughput same as a flash ADC.

The analog input signal is captured by the sample and hold block and transferred to the gain stage to be quantized by a coarse sub-ADC. This digital output is then converted back to an analog signal by a sub-DAC inside each gain stage and subtracted from the original input signal. The generated residue signal is basically the quantization error for each gain stage which is then passed onto the following stages for more accurate digitization. In order to have the same dynamic range for all gain stages, the residue at each stage is amplified by a precise operational transconductance amplifier (OTA) to scale up the residue signal to the full-range. The transfer characteristic of a 2-bit coarse gain stage inside the pipeline ADC is shown in fig 2.2 (a).

In pipeline ADC design, back-end gain stages can be scaled down to reduce their input-referred noise. However, preceding work has illustrated that for a 10-bit pipeline ADC, the current structure has the minimum power consumption and input-referred noise [43]. Thus, the pipeline structure, shown in Figure 2.2 (a), has been employed. Figure 2.2 (b) shows the proposed 3D pipeline ADC architecture by placing analog and MRL-based digital circuits into the top and bottom tiers, respectively. This architecture can significantly improve the critical clock synchronization and de-skewing in high-resolution pipeline ADCs by fully utilizing the advantage of vertically short TSV channels.

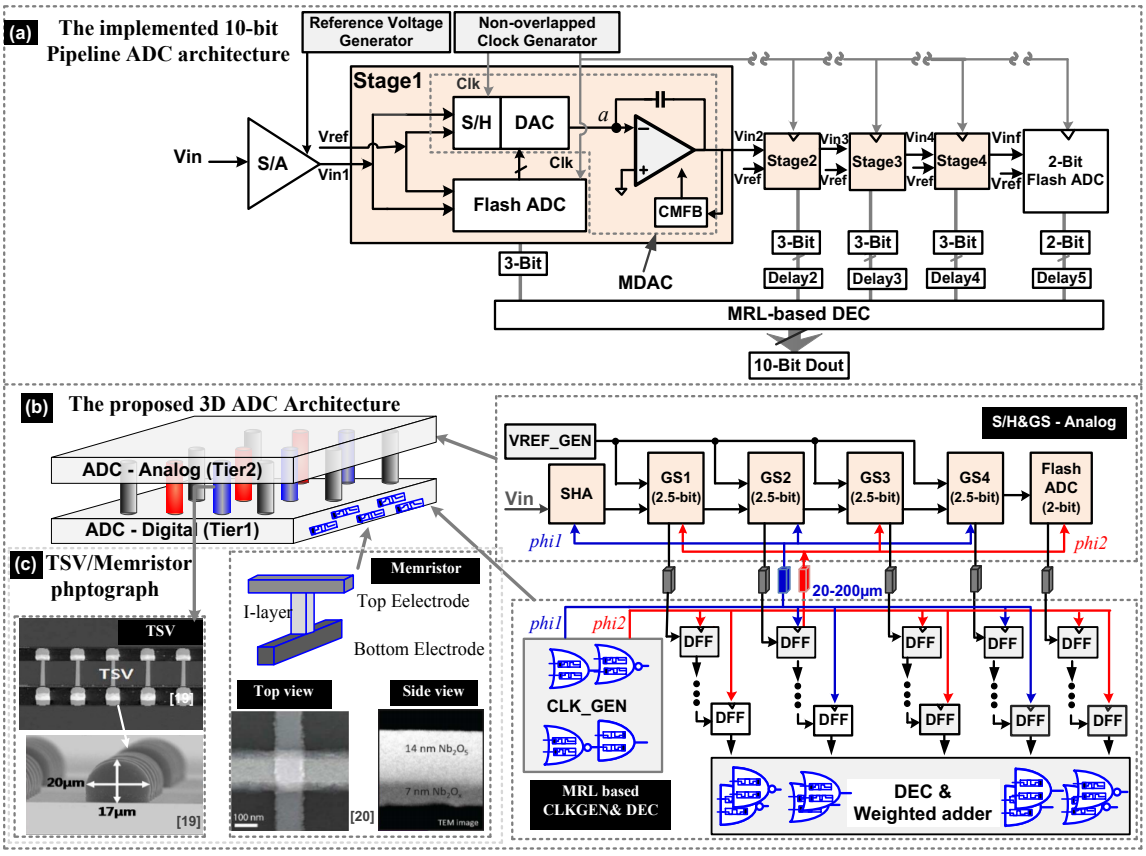


Figure 2.2 The implemented 10-bit pipeline ADC
a) conceptual diagram and b) Proposed 3D pipeline ADC architecture c) TSV channel and Memristor microphotographs

The implemented 3D architecture also considerably reduces the design complexity and increases the accuracy due to the suppression of the inter-tier clock skew/ latency and improved clock synchronization. Figure 2.2 (c) shows the physical structure and microphotographs of TSV channels, micro-bumps, and memristors (i.e., side view and top view) [44], [45].

Figure 2.3 illustrates in detail how the clock jitter and timing skew can significantly affect the conversion accuracy in a high-speed pipeline ADC. Figure 2.3 (a) and (b) show different clocks of each sub-block inside a 1.5-bit gain stage and their switching timing priority, respectively. Since in a pipeline ADC regeneration nodes are inherently coupled

to the input of the comparator through the transistor parasitic capacitances, this coupled voltage, known as kickback noise, can disturb the input signal and result in the error at the outputs [46]. Figure 2.3 (c) depicts a conventional approach to reduce the kickback noise effect on each gain stages of the pipeline ADC and also generate two non-overlapping conversion phases. The tracking phase of the comparator (phi_{COMP}) is delayed with respect to the phase (phi_{1a}) to prevent their coincidence at the regeneration node. However, this time interval and also the non-overlapping timing period of two basic conversion phases (i. e., phi_1 and phi_2) can be degraded when the clock distribution network length becomes longer. Therefore, by considering the proposed 3D CDN-Based ADC, utilizing the TSV channels, delay skew between phi_{1a} and phi_{COMP} can be significantly reduced, and thereby the kick-back noise further decreases. It also can maintain more precise two non-overlapping pipelined conversions. The following describes the implemented models of TSV and wire for the 3D CDN-based pipeline ADC.

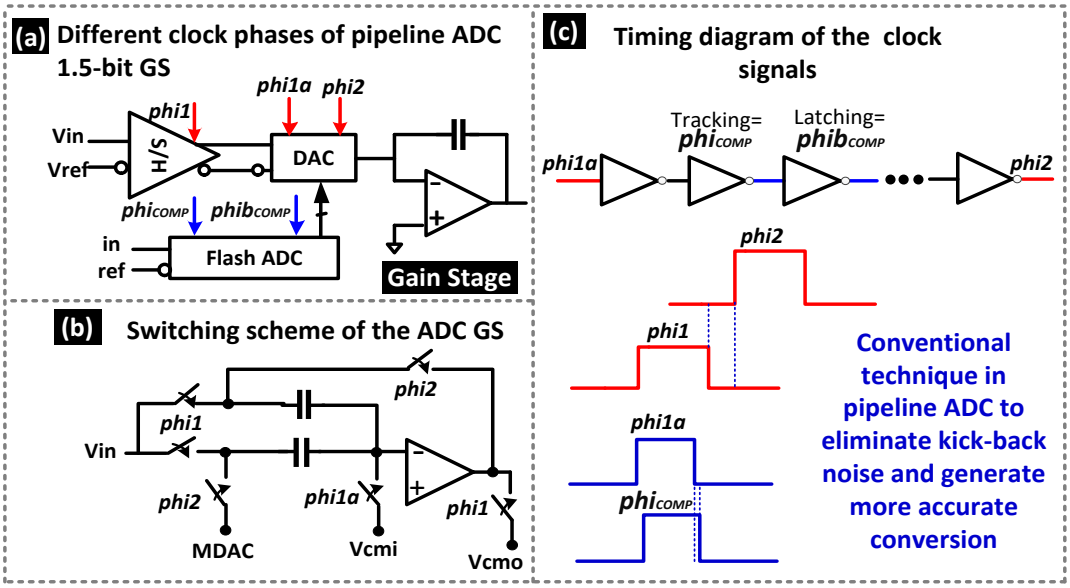


Figure 2.3 Delayed tracking time scheme of the pipeline ADC clocks for preventing kickback noise and clock skew

2. 3 Proposed TSV and wire models

2.3.1 TSV model

To optimize the 3D ADC structure, it is necessary to implement the TSV channel model. The electrical model and design parameters of the signal and power/ground TSV channels are illustrated in Figure 2.4 (a) where R_{TSV} , L_{TSV} , C_{TSV}, R_{si}, and C_{si} represent TSV resistance, inductance, capacitance, and silicon substrate resistance and capacitance, respectively [47][48][49][50][29]. The TSV inductance is modeled with a calculation of the two-wire transmission line [51]. The scalable electrical lumped model is simulated by a 3D EM simulator such as HFSS, as shown in figure 2.4 (b). The TSV model includes the coupling between two adjacent TSVs that can degrade the 3D channel signal integrity. Each TSV is surrounded by a layer for isolating the TSV from the conductive silicon substrate. This model also considers other parasitic effects between the micro-bump and the silicon substrate in the inter-metal dielectric (IMD), micro-bump capacitors and bottom oxide layer capacitor. The analytic equations are derived from the physical configuration including the design parameters. A detailed evaluation is discussed and analyzed in [48], [29]. In the high-frequency structure simulator (HFSS) model, five metal layers (M2-M6) are implemented to emulate a TSV channel inside a silicon substrate. For analog/mixed-signal generally, the noise-sensitive devices or interconnects need to be shielded by ground. Thus, one signal TSV, e.g., clock and data, is surrounded by two ground TSVs inside the silicon substrate.

The evolutionary parameters are shown in figure 2.4 (c) where h_{TSV} , d_{TSV} , t_{ox}, and P_{TSV} are TSV height, TSV diameter, insulator thickness, and the pitch between two TSVs,

respectively. Moreover, ρ_{TSV}, μ_0, ε_0, μ_r, and $\varepsilon_{r,IMD}$ represent TSV electrical resistivity, magnetic constant, electrical constant, relative magnetic and electrical permeability of inter-metal dielectric (IMD) layer, respectively. The TSV design parameters such as capacitance and resistance can be affected by the d_{TSV}, h_{TSV}, and P_{TSV}. Therefore, a computational intelligence-based optimization by using the TSV design parameters is performed, and the optimal trade-offs between power, latency and signal integrity are achieved.

The model includes coupling between two adjacent TSVs that can degrade the 3D channel signal integrity. Each TSV is surrounded by an insulation layer formed for isolating the TSV from the conductive silicon substrate and the corresponding capacitance (C_{TSV}), shown in (2.2), where ε_0 , $\varepsilon_{r,ox}$, l_{TSV}, r_{TSV}, and t_{ox} represent permittivity of free space, relative permittivity of silicon oxide dielectric, hight of TSV, radius of TSV, and insulator thickness, respectively. The TSV resistance is also analytically formulated from the geometric parameters according to (9). The first part reflects DC TSV resistance (R_{dc}), and the second term, R_{ac}, corresponds to the surface resistance (R_{ac}). As the frequency increases, current flows more on the surface of TSVs and causes R_{ac} to be a more dominant term of (2.3). Thus, R_{ac} should be considered to measure the frequency effect. This parameter is also called skin resistance. The ρ_{TSV}, μ_{TSV}, σ_{TSV}, and f represent TSV electrical resistivity, magnetic permeability, metal conductivity, and operating frequency. The TSV inductance is modeled, as in equation (2.4) where μ_o permeability of free space in addition to previously mentioned parameters, with a calculation of the two-wire transmission line [51]. The analytic equations are derived from the physical configuration including the design parameters. A detailed evaluation is discussed and analyzed in

[29][47][48][49].

$$C_{TSV} = \varepsilon_0 \varepsilon_{r,ox} \frac{2\pi . l_{TSV}}{4 \ln[\frac{r_{TSV}+t_{ox}}{r_{TSV}}]} \quad (2.2)$$

$$R_{TSV} = R_{dc} + R_{ac} = \left(\frac{\rho_{TSV} l_{TSV}}{\pi.(r_{TSV})^2}\right) + \left(\frac{l_{TSV}\sqrt{\pi.f.\mu_{,TSV}.\sigma_{TSV}}}{r_{TSV}.\sigma_{TSV}}\right) \quad (2.3)$$

$$L_{TSV} = \frac{\mu_0 l_{TSV}}{2\pi} \times \left(\ln(\frac{2.l_{TSV}}{\pi.(r_{TSV})^2} - \frac{3}{4})\right) \quad (2.4)$$

The silicon resistance, R_{si}, and capacitance, C_{si}, can be calculated from (2.5) and (2.6), respectively. The ε_{si}, d, and σ represent silicon permitivity, TSV pitch, and silicon conductivity, respectively. The scalable equivalent electrical lumped model is simulated and verified in HFSS as shown in Figure 2.4 (c).

$$R_{si} = \frac{\varepsilon_{si}}{C_{si}\sigma} \quad (2.5)$$

$$C_{si} = \frac{\pi \varepsilon_{si} l_{TSV}}{\ln\left\{\frac{d}{2\,r_{TSV}} + \sqrt{\left(\frac{d}{2\,r_{TSV}}\right)^2 - 1}\right\}} \quad (2.6)$$

As the frequency increases, the current flows on the surface of the TSVs. Dependency to the frequency is measured by the skin depth. The TSV inductance is modeled with a calculation to the "two-wire interconnection line model" [48]. The silicon resistance, R_{si}, and capacitance, C_{si}, have been also considered as two optimization parameters.

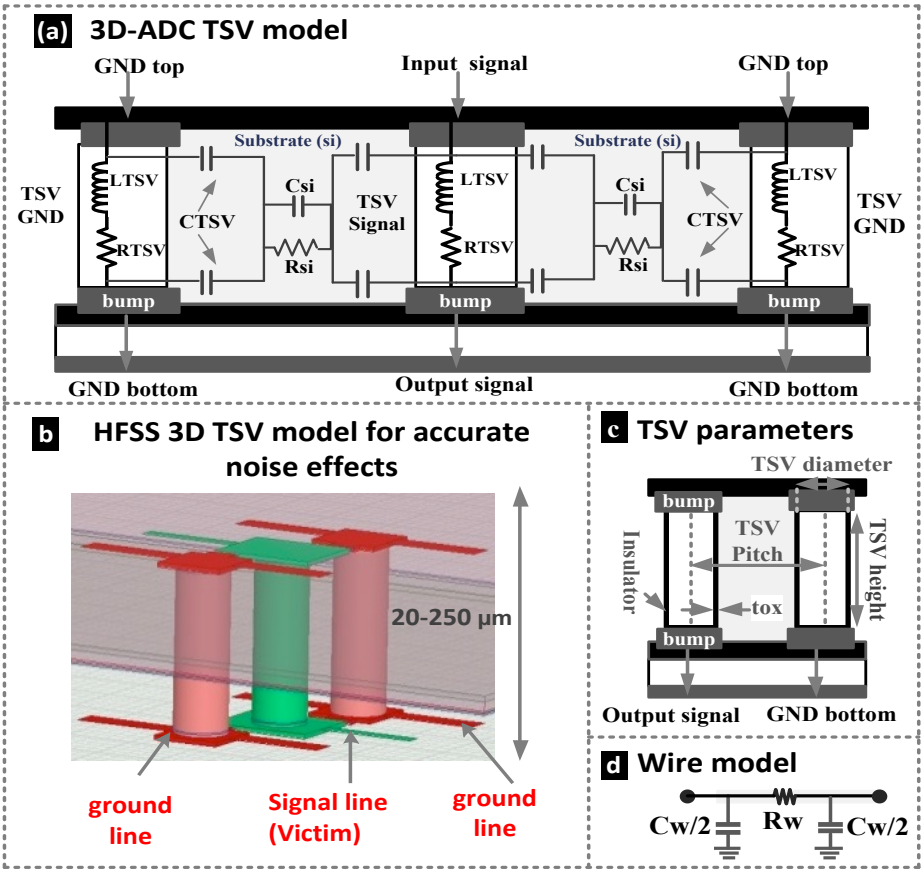


Figure 2.4 (a) Modeling of coupling between TSVs (b) HFSS model (c) TSV design parameters (d) wire model

2.3.2 Wire model

Latency, noise, and power consumption optimization in 3D design require TSV, wire model, RC estimation, and routing improvement. To estimate wire parameters shown in Figure 2.4 (d), the metal layers assigned to inter-block signals must be defined. In conventional 2D designs, top metals are typically wide and thick. Thus, they are rarely implemented for signal routing. Due to this fact, using a wide portion of these metal layers for power delivery does not impact the overall signal interconnect length. As M4-M6 layers are the commonly used top metal type, we have considered the same metal stacking. The

appropriate RC values for power/ground wire are evaluated in R_w and C_w in equations (2.7), (2.8), (2.9), (2.10) considering copper for wire material, respectively. w, t , h, and d are wire width, wire thickness, wire-ground space, and wire-wire space, respectively [52].

$$\mathrm{R_w} = \rho \frac{\mathrm{l_{wire}}}{\mathrm{A_{wire}}} \tag{2.7}$$

$$\mathrm{C_w} = \mathrm{C_{1w}} + \mathrm{C_{2w}} \tag{2.8}$$

$$\mathrm{C_{1w}} = \varepsilon\left(\left(\frac{\mathrm{w}}{\mathrm{h}}\right) + 2.97\left(\frac{\mathrm{t}}{\mathrm{h}}\right)^{0.232}\right) \tag{2.9}$$

$$\mathrm{C_{2w}} = \varepsilon\left(0.23\left(\frac{\mathrm{w}}{\mathrm{d}}\right) + 1.22\left(\frac{\mathrm{t}}{\mathrm{d}}\right)^{1.384}\right)\left(\frac{\mathrm{d}}{\mathrm{h}}\right)^{-0.0398} \tag{2.10}$$

2. 4 The evolutionary process of TSV and sub-blocks

To achieve the optimum power efficiency, supply noises, and power/signal quality of a heterogeneous pipeline ADC, the proposed architecture utilizes an evolutionary method by considering all design parameters, i.e., analog sub-blocks sizing, wire geometery, TSV geometry and their arrangement, simultaneously.

When considering co-optimization of TSV structure and circuit sub-blocks, different performance parameters, i.e., die area overhead and coupling in TSVs or speed and power consumption of the ADC, conflict with one another. Thus, the optimization method requests trade-offs, namely Pareto-fronts, to find optimum design points among performance functions that compromise a computational [53], [54].

To achieve the best performance and power/area efficiency, the proposed 3D CDN-based design utilizes an evolutionary method by considering all design parameters (i.e., ADC

transistor sizing, CDN design parameters, and TSV model variables such as width, height, and pitch) concurrently. In this novel co-optimization, the key TSV geometry model and circuit design parameters of the 3D pipeline ADC are individually defined as the input variable vectors in MATLAB. A kind of MATLAB-HSPICE toolbox interface is created to simulate the proposed CDN-based ADC utilizing the evolutionary solver [43], [55], [56]. After each HSPICE simulation, MATLAB links to HSPICE to evaluate desired performance parameters and feed the new design points to the genetic algorithm [54], [57]. The algorithm considers both effects of inter-tier TSV geometry and design parameters such as transistor sizing. Thus, the optimization problem explores design space to find trade-off points which are called Pareto-fronts through the MATLAB-HSPICE link shown in figure 2.5 [58]. The evolutionary engine is a GA since it is one of the most effective solvers for multi-objective problems [59].

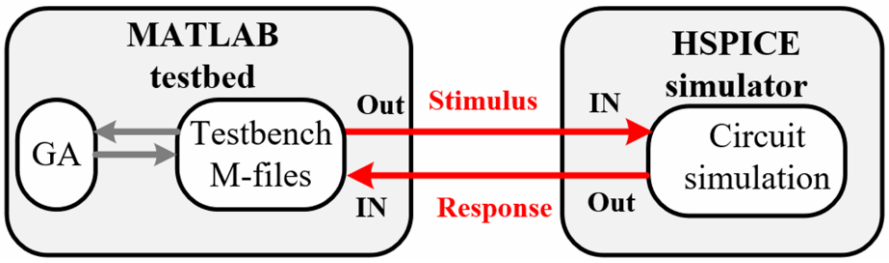


Figure 2.5 Matlab-HSPICE link to perform the evolutionary process

To the best of our knowledge and based on the references, the simultaneous optimization of 3D ADC/CDN design and TSV geometry is first reported in this work; although there are some works on ADC design and TSV model optimization. In this co-optimization of 3D pipeline ADC design and TSV geometry, different performance parameters conflict with one another (i.e., signal integrity and coupling parameters of TSVs, and latency and

power parameters for ADC). Thus, the optimization problem requests trade-offs that would compromise a general synthesis computational efficiency.

2.4.1 Evolutionary process of TSV

Figure 2.6 (a) shows the performance parameters of the TSV optimization. Given an n-dimensional decision variable vector $\mathrm{x} = \{\mathrm{x}_1, \ldots, \mathrm{x}_\mathrm{n}\}$ in the solution space X (i.e., X represents the whole design parameter space), the TSV optimization problem can be formulated as to find a vector x^* that minimizes a set of K objective functions $z(x^*) = \{z_1(x^*), \ldots, z_k(x^*)\}$. The solution space X is generally restricted by a set of equality constraints denoted by $\mathrm{g_j}(x^*) = \mathrm{b_j}$ for j = 1, …, m, and bounds on the decision variables. The GA tries to obtain all TSV sizing parameters through a reduced set of independent variables, namely gene. These variables shape the chromosome string. The corresponding string is shown in Figure 2.6 (b) where d_{bump}, h_{bump}, and h_{imd} represent micro-bump diameter, micro-bump height, and inter-metal dielectric (IMD) height, respectively. The GA's fitness function, shown in (2.11), including all TSV output signal specification, finds the optimum design parameter of TSV. For example, the better fitness function has a lower settling time, error band, overshoot percentage, peak time, and higher peak absolute value.

$$fitness\ function = \frac{k_0 \times Settling\ time + k_1 \times errorband + k_2 \times Overshoot + k_3 \times peak_{time}}{k_4 \times Peak_{absolute_{value}}} \tag{2.11}$$

2.4.2 Evolutionary process of ADC sub-blocks

To improve the overall performance of the 3D ADC, it is necessary to attain the optimal

sub-block design points and the scaling factor of gain stages by utilizing the optimal TSV geometry obtained from the first evolutionary process. To improve the efficiency and minimize search space, a hierarchical multi-objective optimization is implemented.

Figure 2.6 (c) shows an example of each sub-block chromosome in the evolutionary process; where, WLp_{11}, WLp_{12}, show transistors width (W) to length (L) ratio inside each analog sub-block and p represents the transistor pair number. The best (minimum) fitness function value can be obtained using (2.12).

$$fitness\_function = -\frac{k_0 \times SNDR + k_1 \times (FOM)^{-1}}{k_2 \times Power + k_3 \times Area} \tag{2.12}$$

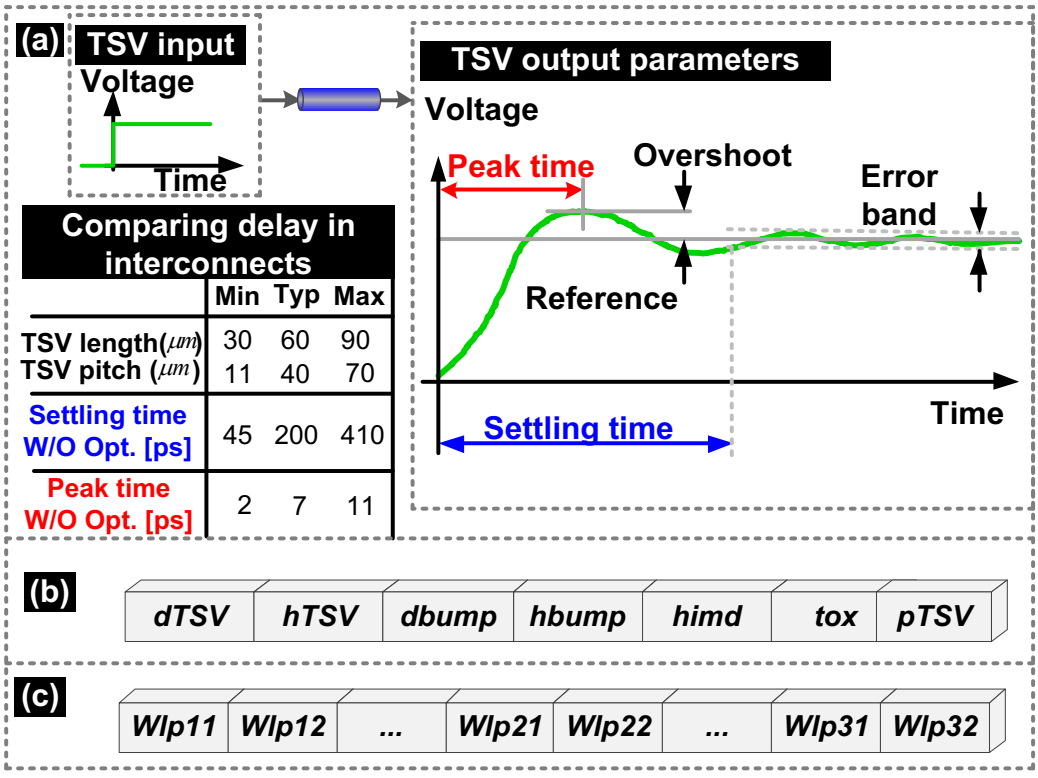


Figure 2.6 (a) Time corresponding performances of the TSV fitness function (b) TSV and (c) ADC sub-block evolutionary design parameters in a GA chromosome string

Since dynamic performances improve the ADC applicability and give a better

understanding of its most critical parameters such as linearity and spectral purity, SNDR is the most effective term of the fitness function. Higher SNDR can also improve the bit rate. Thus, the relevant coefficient, k_0, is chosen so that the SNDR term in the numerator (2.12) is several times more effective than others. The second key factor is figure-of-merit (FoM), with coefficient k_1, and consists of total circuit power consumption, Pw, sampling frequency, fs, and effective number of bits (ENOB).

$$FOM = \frac{Pw}{2^{ENOB} \times f_s} \tag{2.13}$$

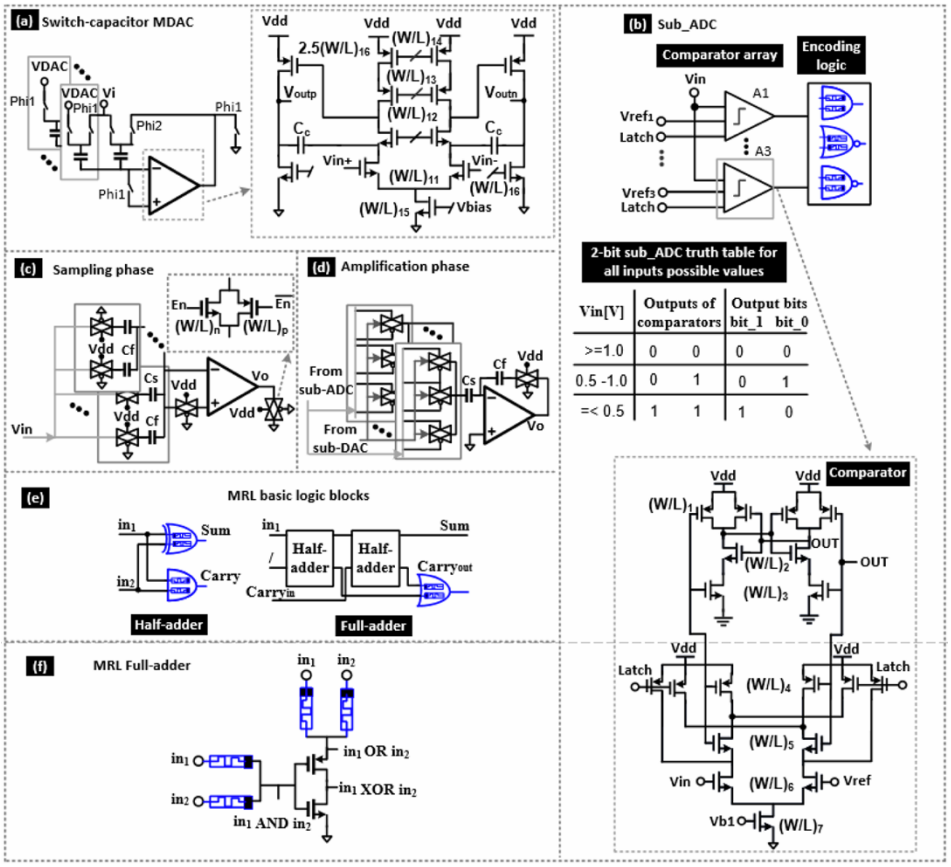
The third significant parameter that GA tries to minimize is power consumption. In (2.12), k_2 changes the effect of power consumption on the fitness value. Since a high-speed pipelined ADC power consumption is mostly dominated by the analog blocks, the evolutionary process tries to minimize this portion of the power consumption. The silicon area minimization is the last effective factor, extracted from the applied elements sizing. Thus, k_3 is assigned in a way to have the least impact on fitness function compared to the other terms.

2. 5 Simulation process

The proposed architecture described in chapter 2 was implemented on a 10-bit pipeline ADC at 120 MS/s in 65nm technology with a 1.0V power supply. The circuit simulations are performed by using Spectre simulator and HSPICE-MATLAB interface in MATLAB R2016a environment. To evaluate the performance accurately, the post-layout extracted simulation was performed by considering the parasitic resistance and capacitance of the 3D ADC design.

2.5.1 The pipeline ADC topology

Figure 2.7 shows the schemes of the optimized ADC sub-blocks as the case study. The switch-capacitor multiplier DAC (MDAC) is shown in Figure 2.7 (a). To have a fast output settling time, low-voltage operation, and high-gain amplifier, a two-stage telescopic cascade topology is implemented.



Vin[V]	Outputs of comparators		Output bits bit_1	bit_0
>=1.0	0	0	0	0
0.5 -1.0	0	1	0	1
=< 0.5	1	1	1	0

Figure 2.7 The designed ADC schematic
(a) Switch-capacitor scheme, inside the applied 2.5-bit stages and MDAC circuits (b) sub-ADC and all possible input signal Vin values and their corresponding output digital codes (c) sampling and (d) amplification phases using transmission gate (TG) switches, (e) schematic of the MRL-based half and full-adder, using 2-input NAND/NOR gates (f) Full adder MRL-based implemented in the DEC block

For improved stability and phase margin, a Miller compensation scheme is utilized in the

gain stage amplifier. Figure 2.7 (b) shows the sub-ADC block generates three coarse output bits from the input signal (V_{in}). A comparator-latch is also utilized to synchronize further any potential clock skews from the 3D CDN. Figure 2.7 (c) and (d) show the operation of a multiplier MDAC (MDAC) utilizing the transmission gate (TG) switches in signal sampling and amplification phases, respectively. Figure 2.7 (e) and (f) show the schematic of the MRL-based half and full-adder, using 2-input NAND/NOR gates, implemented in the DEC block. As it was pointed out, both logic NAND and NOR gates consist of two memristive devices where their polarity is the only structural difference. Since the memristor voltage divider degrades the output voltage (i.e., $R_{OFF}>>R_{ON}$), CMOS inverter is integrated into an MRL OR/AND logic gates to improve the output voltage. However, the delay time of the logic gates changes with the capacitance of the CMOS gate and consequently needs optimization.

As illustrated in Figure 2.7 (a), the gain stage introduces seven design variables which consist of six-transistor sizes and one Miller capacitor value. Using equal size transistors for the implemented switches (transmission gates) adds a new evolutionary variable. The comparator transistor sizing also introduces seven new variables as shown in Figure 2.7 (b). A total of 14 variables can be used as a chromosome vector by the evolutionary algorithm leading to fast and accurate performance optimization.

The evolutiony process of the whole 3D pipeline ADC topologey as a multi-objactive optimization is illustrated in Figure 2.8.

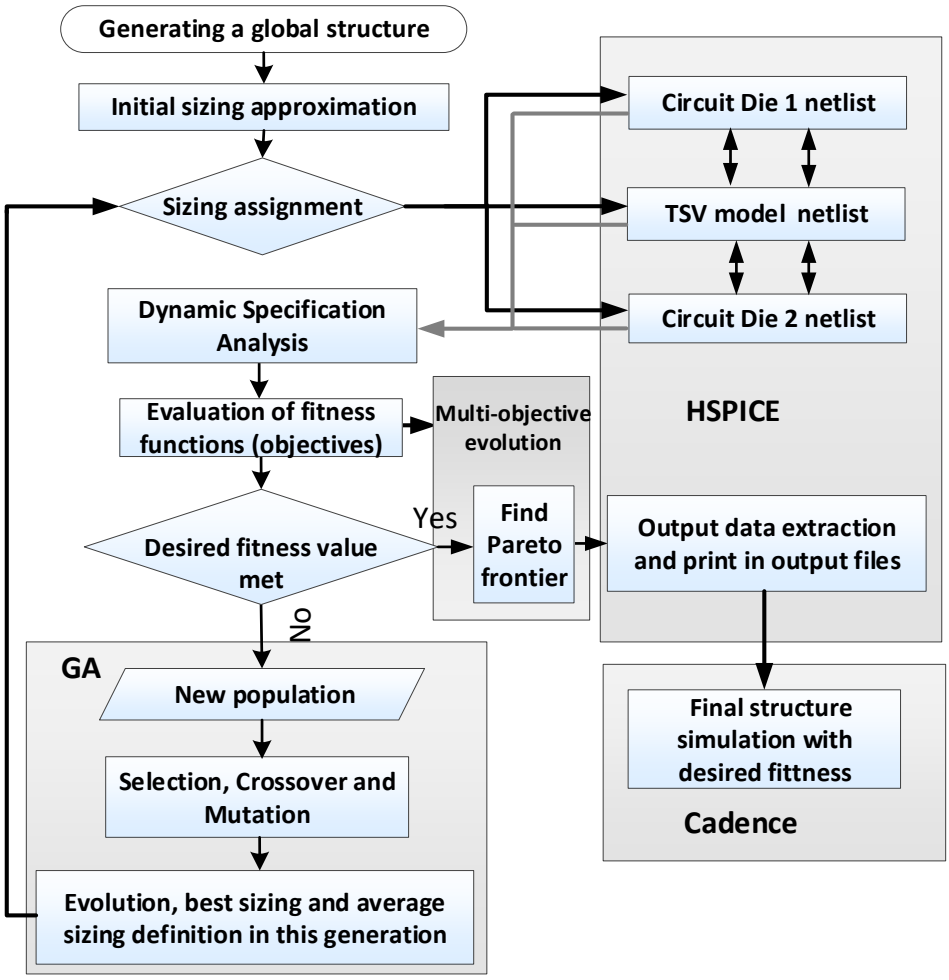


Figure 2.8 Multi-objective evolution of the designed 3D pipeline ADC

2.5.2 The simulation results

The fitness function of the pipeline ADC is constrained to give at least 8-effective bit resolution at the Nyquist rate. When the proposed co-optimization algorithm is used to a 10-bit pipeline ADC, the acceptable solutions in each iteration should have SNDR more than 49.9 dB using a 1024-point Fast Fourier Transform (FFT). Figure 2.9 shows the improvement of the TSV channel performance parameters such as latency, power consumption, and output voltage overshooting utilizing the evolutionary process on TSV

over 100 iterations. Power simulations are performed based on the power consumption per TSV in a 3×4 TSV array. Before the evolutionary process, the 3× 4 TSV array generates 0.35 µW average power. As expected, the TSV latency (i.e., from 25ps to sub 1ps), I/O power per channel (i.e., from 45µW to 21µW) and output voltage overshooting (i.e., from 0.8V to 0.05V) are minimized after the evolutionary process.

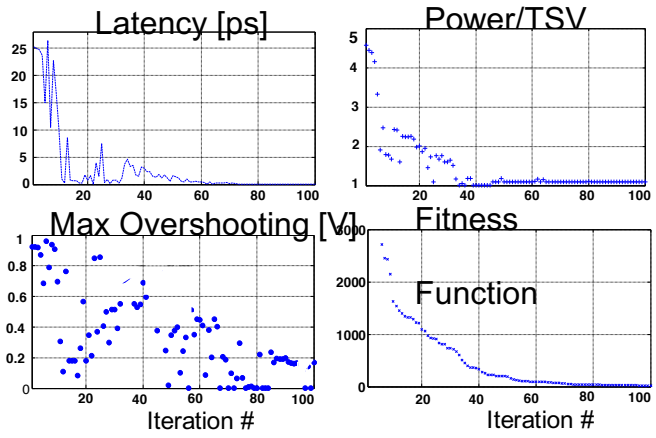


Figure 2.9 Performance optimization of TSV channel in 100 iterations

Diagrams of the multi-objective optimization process to explore the trade-offs between power consumption and latency of the proposed 3D ADC, considering the optimized TSVs are shown in Figure 2.10 Based on the trend of the *Pareto-front* plot, we can evaluate the most optimal design point for a high-performance and low-power 3D ADC architecture. The illustrated diagrams show how the implemented evolutionary algorithm can decrease the power consumption and delay. It also shows when the genetic algorithm is operating after each generation the optimal solutions get closer and the convergence rate increases.

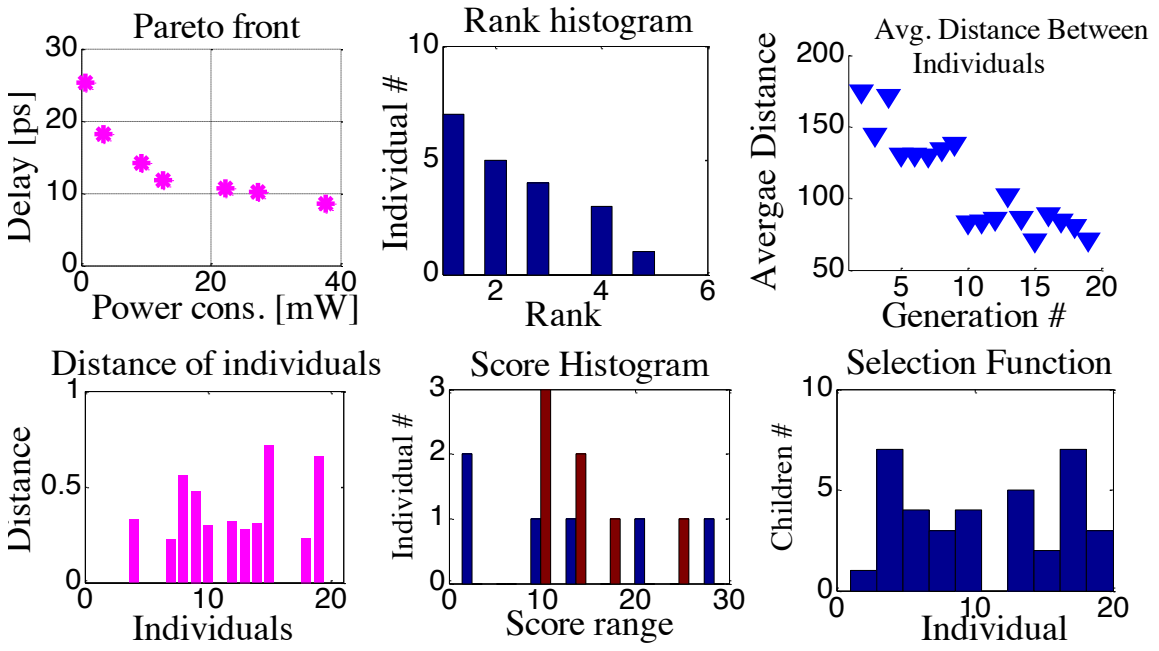


Figure 2.10 Multi-objective evolutionary process of the proposed ADC

Table 2.1 shows a summary of the optimization parameters [54]. By utilizing the multi-objective hierarchical optimization, the computational overhead is reduced because of limiting the search space. The table shows the entire design cycle took about 3 hours on high-performance computing (HPC) server with 128GB RAM 16-core Intel(R) Xeon(R) CPU X5560 @ 2.80GHz 107 processor.

Table 2.1 The evolutionary specifications of the proposed algorithm

NO. of variables	**Population size**	**Generation No.**	**Crossover**	**Mutation probability**	**Stopping condition**	**Total performance improvement(%)**	**Total convergence time (hr.)**
15	40	20	Scattered	0.05	Stall generation	23	3

Figure 2.11 shows the final TSV channel output waveforms by considering the TSV/ADC sub-block design parameters and MRL logic. As expected, the optimized output voltage is significantly improved regarding the delay, power, and overshooting with the obtained design parameters. The dotted lines show the final ADC output voltage waveform after each iteration. Green and red lines show the ADC output voltage waveform with and without optimization, respectively. The black line shows the CDN output voltage signal in a 2D pipeline ADC [43].

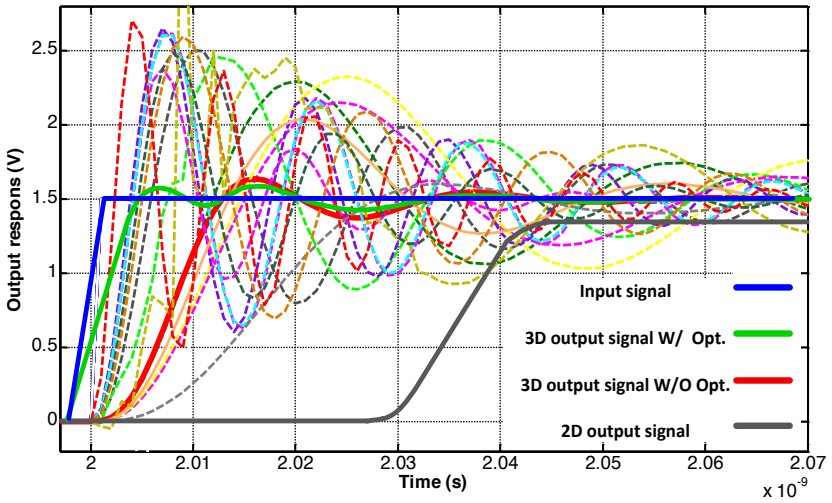
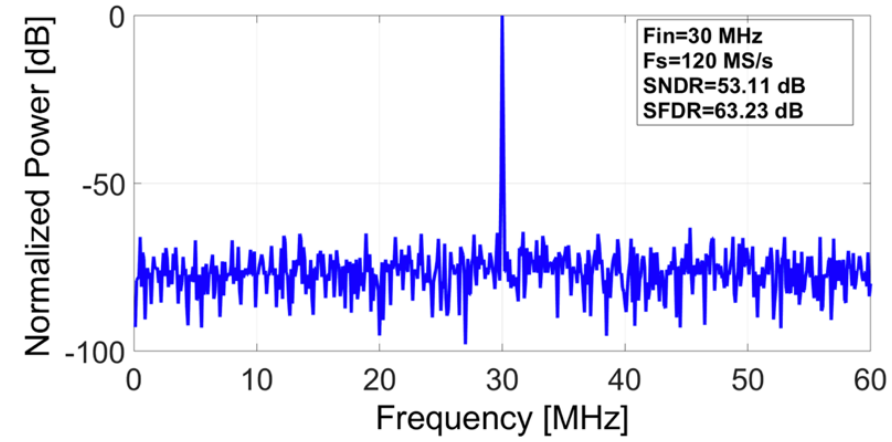


Figure 2.11 Evolution of 3D TSV output signal compared to the corresponding 2D signal

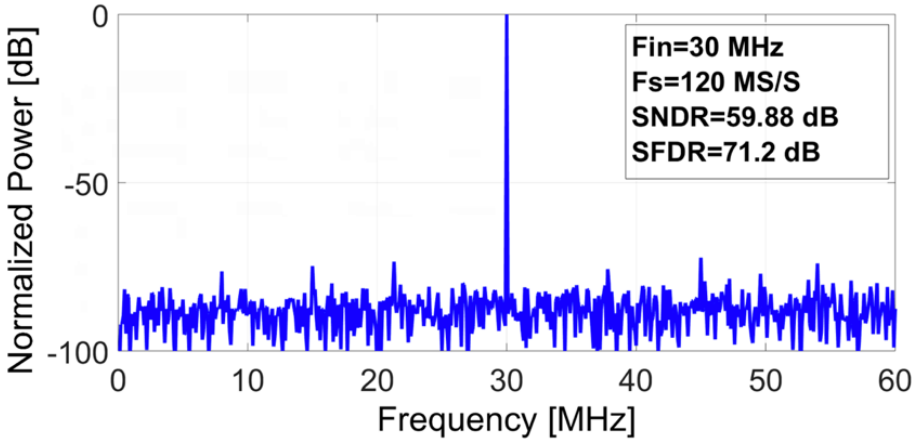
To verify performance improvement of the proposed 3D ADC design vs. 2D conventional design, we also implemented Figure 2.12 (a) and (b) show the ADC output power spectrum for 2D and 3D designs, respectively. As predicted, the proposed CDN-based 3D ADC design utilizing the co-optimization methodology results in significant improvement of the performance (i.e., SNDR increased from around 53 dB (2D) to around 60 dB (3D /w MRL and optimization)).

Figure 2.13 shows the power consumption and the signal delay of each investigated case (i.e., the conventional 2D ADC, 3D ADC using MRL logic with and without the optimization).

This figure shows the signal delay has significantly decreased by considering the TSV vertical connections instead of 2D long interconnections. It also shows that the proposed 3D ADC utilizing the co-optimization algorithm as well as MRL logic can improve power efficiency compared with a 2D design.



(a)

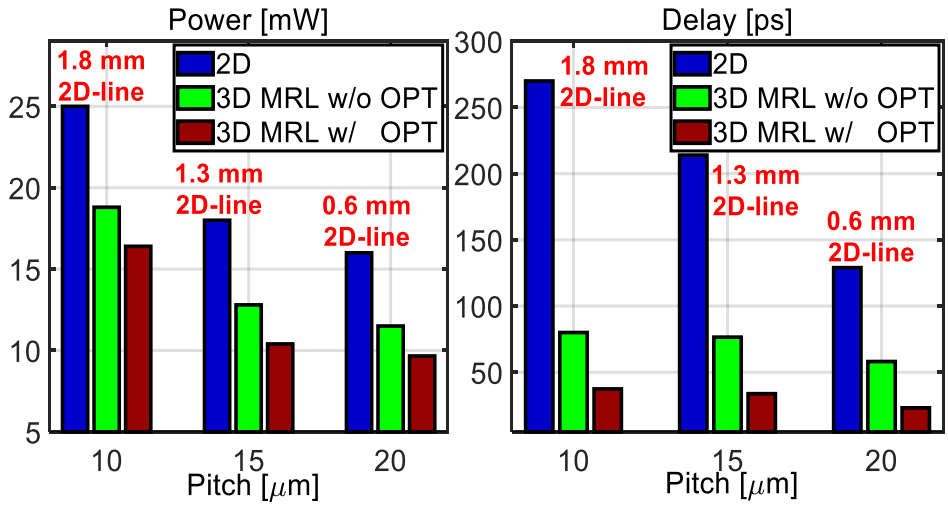


(b)

Figure 2.12 ADC power spectrum of (a) a conventional 2D ADC and (b) the proposed 3D ADC.

The proposed designs offers more than 7 dB improvement in both SNDR and SFDR

Moreover, as the TSV pitch decreases, the power consumption can be increased dramatically (i.e., from 19mw to 12mW).



(a) (b)

Figure 2.13 2D vs. non-optimized/optimized 3D ADC using MRL logic
(a) power consumption and (b) clock delay for different TSV placement and interconnection line length with 60μm TSV height

Table 2.2 compares the performance of the state-of-arts pipeline ADCs and the proposed 3D ADC.

Table 2.2 Performance summary and comparison

	Previous works			**This work**	
	TVLSI 15 [60]*	TCASI 16 [61]*	JSSCC 15 [62]*	3D	3D MRL /w Opt.
Bit #	10	14	12	10	10
Supply [V]	1.2	1.8	1	1	1
Tech. [nm]	130	180	40	65	65
Sample rate [MHz]	200	250	195	120	120
SFDR [dB]	63	87	82	68.7	71.2
SNDR[dB]	53	68	64.8	57.3	59.9
Power[mW]	38	300	53	11.4	9.6
FOM [fJ / Conv.]	552	570	157	158	98.8

(* Measurement results)

As summarized in Table 2.2, in an identical technology, this design can achieve considerably reduced power consumption than other techniques which implement particular power saving policies such as amplifier sharing, sample and hold less design or switched amplifier methods [31]-[33]. Moreover, a higher performance specification has been achieved and determined after considering both evolutionary process and MRL architecture (i. e., SNDR is increased from 57.3 dB to 59.9dB). In addition, implementing TSV-aware optimization and MRL logic has reduced FoM up to 37%.

The results confirm that the proposed 3D design algorithm searches for a circuit with higher conversion accuracy and lower power dissipation, while other parameters such as SNDR, and FOM are improved at the same time. Figure 2.14 shows the 3D pipeline ADC layout and block-diagrams of analog and digital tiers, implemented in the 3D simulation. It also shows the power and delay bar diagrams that compare 2D with 3D ADC simulation results.

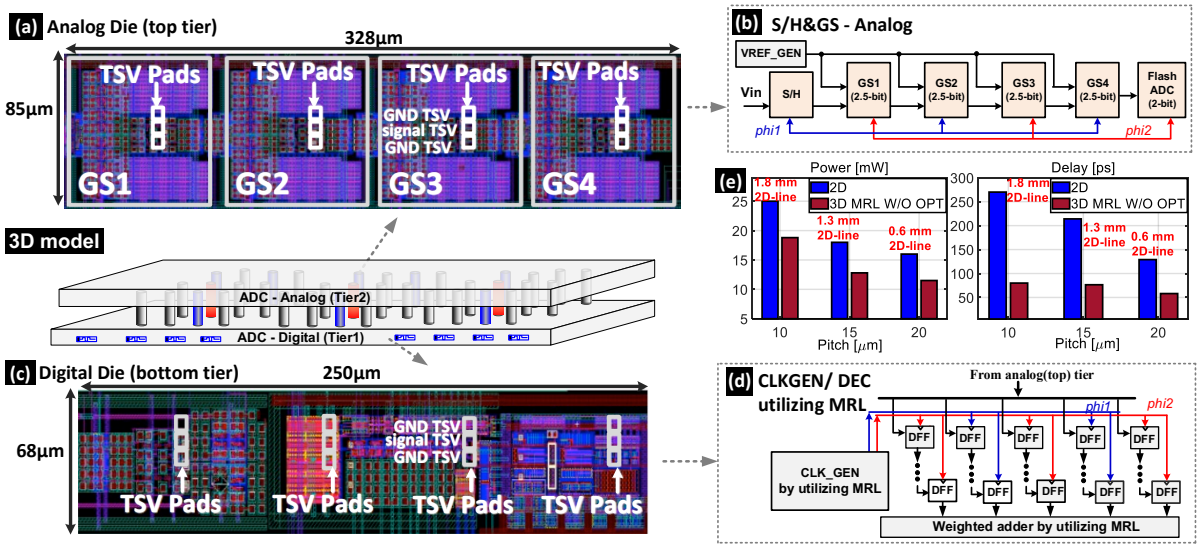


Figure 2.14 The designed ADC layout
(a) Analog/top die layout and (b) scheme, (c) digital/bottom tier layout and (d) scheme of the 3D pipeline ADC, (e) comparison bar diagrams for the ADC power and signal delay

2. 6 Conclusion

This chapter presented a 3D pipeline ADC design with a novel CDN architecture by utilizing 3D channel and MRL to improve the conversion accuracy, dynamic performance, and power/area efficiency. In addition to implementing MRL digital blocks, implementing a multi-objective evolutionary algorithm on the analog blocks of the proposed 3D ADC has further provided the overall performance improvement regarding power, delay, and silicon area efficiency. The prototype pipelined ADC achieves 59.9 dB SNDR at 120 MS/s with 9.6 mW power consumption [10]. It also has a 98.8 fJ/conversion-step, which shows a significant improvement among the designs compare.

Chapter 3

3D POWER DELIVERY NETWORK

3. 1 Introduction

In a mixed-signal system on chip (SoC) design, providing a robust power distribution network (PDN) is a critical factor to enhance the overall performance [63]. A well-designed PDN is essential to reduce supply noises, tolerate large variations of load current, having better both power and signal integrity of 3D integrated heterogeneous devices, and support failure-resistant chip operations.

The power delivery design of a conventional 2D system on chip (SoC), as shown in Figure 3.1 (a), is a critical challenge in advanced semiconductor technologies because of the high operating frequency, more power density, and low supply voltage. As the frequency of heterogeneous devices increases, while the supply voltage scales down, the design of optimized PDN becomes more critical. In this chapter, a novel 3D PDN design for a memory I/O interface is proposed. A computational intelligence, i.e., evolutionary process, is also utilized to further improve the topology and find the optimum design points of the 3D PDN structure.

In a 3D PDN packaging, latency and integrity can inherently be improved due to vertically short interconnections (3D TSVs or 2.5D micro-bumps). However, as the number of 3D I/O channels increases, i.e., 1024 TSVs in a wide-IO interface [64], the TSV I/O transceiver should be co-integrated into the PDN to evaluate the PDN performance accurately.

Figure 3.1 (b) shows an on-chip 3D-PDN which is composed of a phase-locked loop (PLL), TSV I/O array, and clock/signal distribution networks. In this complex configuration of 3D PDN, the critical parameters such as TSV resistance, capacitance, and inductance sizing can be the most important design variables for the optimization. For example, TSV connection on 3D PDN has small inductance value compared to the traditional bond wire connection. Thus, when the frequency of the I/O interface increases reaching above gigahertz, i.e., 2 Gb/s in high bandwidth memory (HBM) [65], the fraction of inductance value, i.e., 200 pH, causes disruptive frequency-dependent impedance and entire system failure. According to the international technology roadmap for semiconductor (ITRS) number of 3D I/O array will be increased significantly in the near future [20]. Therefore, a prudent PDN topology planning can have key effects on the heterogeneous 3D IC device operation. For instance, severe crosstalk and supply voltage fluctuations can occur due to a poor PDN topology and dense 3D TSV array.

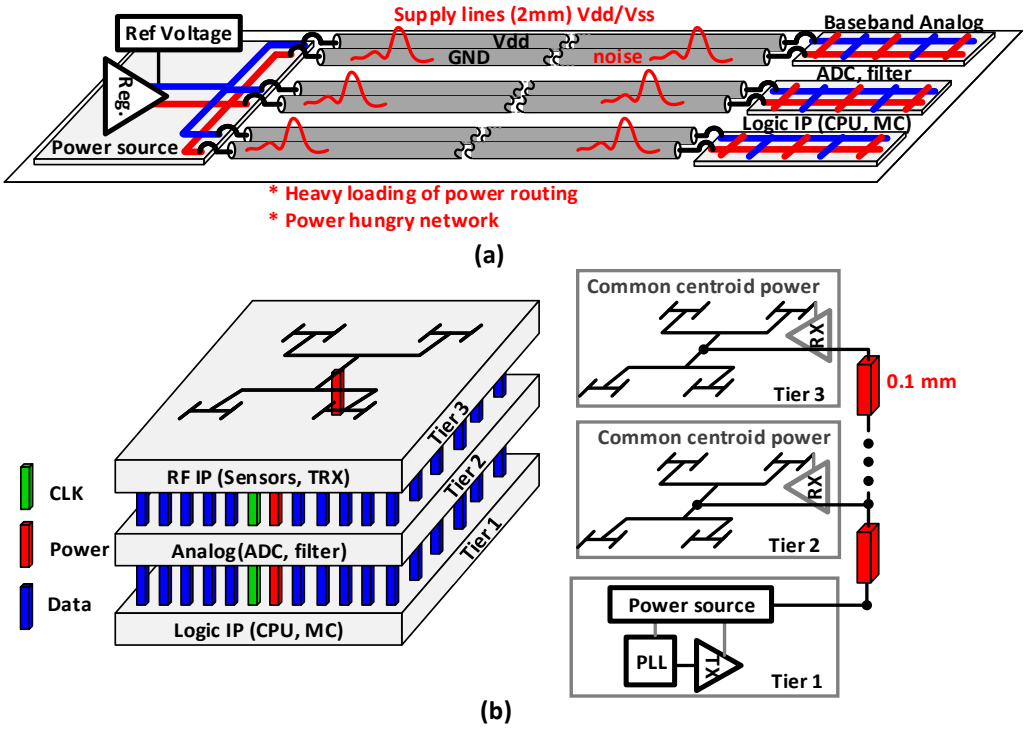


Figure 3.1 Diagram of a conventional (a) 2D and (b) 3D PDN

Integrating low-dropout voltage regulator (LDO) into the PDN provides a significant improvement in eliminating power supply noises and forms a complete on-chip 3D PDN. Also, implementing multiple LDOs close to on-die noise sources in a distributive manner can further reduce the power supply noise [66][67][68]. On the other hand, the complex interaction between the LDOs, TSVs, and I/O interface would cause unstable power delivery network and result in severe degradation of circuit performance or design failure.

In the proposed work, the PDN stability and chip functionality are accomplished by a proper LDO placement among multiple stacked tiers. Moreover, rigorous analysis for different LDO's placement and active/passive device sizing is implemented to find the optimum 3D PDN performance.

Recently, many works have been undertaken to explore 3D PDN design. Work in [69]–[77]. shows the impact of TSV sizing and allocation on 3D PDN, and [70] analyzes different TSV densities and aspect ratios. A novel topology for 3D PDN is taken into consideration for TSV placement and package parasitics in [71]. Floor-planning and optimization of PDN are discussed in [72]. Electrothermal modeling and 3D PDN optimization under electrothermal constraints are presented in [73], [74]. Works in [75]–[77] discuss power supply noises and decoupling insertion, while [78] investigates placement and planning of power and ground TSVs in 3D PDN. Work in [79] designs and analyzes PDN impedance including TSV effects on HBM interposer. An HBM and micro-bump (μbump) test methods are explained in [28] to improve the chip testability. To the best of our knowledge, this is the first work that ensures 3D PDN stability by taking into consideration the complex interaction between LDOs, TSVs, and I/O interface. It also finds the optimum active/passive devices sizing and placement for LDO and TSV to achieve the best performance for the 3D PDN and I/O interface.

Since the academic fabrication of 3D IC utilizing TSVs or 2.5D through-silicon interposers (TSIs) is challenging, a feasible packaging technique such as 3D micro-bumps [50] with similar dimensions, i.e., width/pitch, has been implemented to optimize the TSV geometry for the proposed 3D PDN. However, the proposed architecture could be applied to any 2.5D silicon interposer and future 3D TSV-based ICs.

3. 2 Proposed performance-aware 3D PDN using common centroid regulator placement

To reduce severe crosstalks and supply voltage fluctuations from dense TSV I/O array and 2D complex power routing, a performance-aware 3D PDN utilizing common

centroid regulator placement has been proposed. It comprises a low dropout (LDO) voltage regulator placed on the middle tier (Tier 2), a TSV array as well as straightforward transmitter and receiver to form the I/O interface. Figure 3.2 (a) shows the 3D integration of the proposed architecture. A multi-drop link is formed by CPU connected to the multi-DRAM chips through TSV channels exploiting 3D integration technology. Since the distance between CPU and each DRAM is very short, i.e., 35µm, the signal integrity is improved significantly compared to the traditional 2D DRAM I/O interface [80]. TSV channel enables the output data to be transmitted vertically from CPU to DRAM where the data is going to be recovered. Every signal TSV is bounded by ground TSVs to shield the noise-sensitive interconnect and improve the signal integrity. 3D TSV pitch of adjacent TSVs can be increased two or three times the TSV diameter to minimize crosstalk and timing distortion. Consequently, the proposed architecture uses five TSV channels as shown in Figure 3.2 (b). The labeled TSVs represent ground, data, and power channels.

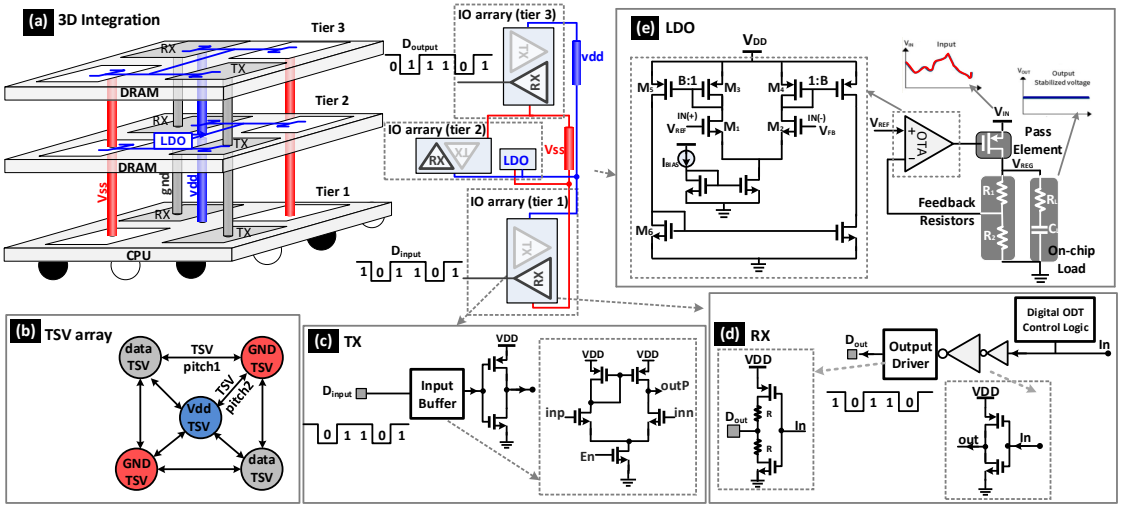


Figure 3.2 Proposed performance-aware 3D PDN using a common centroid LDO placement (a) 3D integration (b) TSV Array (c) TX (d) RX (e) LDO

3. 3 3D I/O interface

As shown in Figure 3.2 (c), the proposed 3D transmitter (TX) consists of an input buffer and output driver. On the transmitter side, the generated data (D_{input}) is fed into the input buffer followed by an output driver. The output driver utilizes CMOS logic to drive the TSV channel and enable data transmission. To support high-frequency operation and minimize delay, voltage offset, mismatch, and noise effect, an input buffer using a self-biased differential amplifier is exploited [81].

Regarding the receiver (RX) side as shown in Figure 3.2 (d), it consists of a digital on-die termination (ODT) [82], pre-driver, and output driver. Because TX and RX communicate by the common mode voltage, ODT is utilized to amplify incoming data from TSV channel. It also controls the common mode voltage and improves signal integrity by eliminating impedance mismatch. The pre-driver utilizes inverters to enhance the recovered data, and the output driver uses push-pull configuration with two resistors in series with transistors to avoid impedance mismatch and reduce sensitivity to process, voltage, and temperature (PVT) corners [83].

3.3.1 Low dropout voltage regulator for 3D PDN

LDO regulators are an essential part of the power management system that provides constant, stable, and reliable voltage supply rails [84], [85]. Figure 3.3 shows an example of a simple NMOS low dropout (LDO) voltage regulator including a pass element, control circuit, and the output load. NMOS devices are not widely used in LDO designs, but they simplify the explanation of LDO performance.

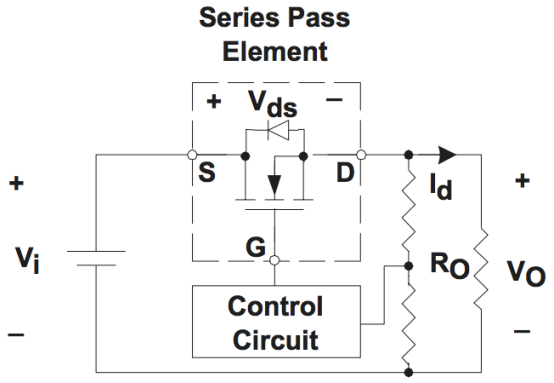


Figure 3.3 A simple NMOS LDO

The critical parameters to evaluate LDO performance can be categorized into three groups namely quiescent currents, operating voltages, and regulating performance [86]. The evolutionary algorithm tries to find the LDO sizing with better regulating performance, higher efficiency and lower operating voltages. Quiescent current, or ground current, is the difference between input and output currents. Minimum quiescent current is necessary for maximum current efficiency. Quiescent current consists of bias current (such as band-gap reference, sampling resistor, and error amplifier) and drives current of the series pass element, which does not contribute to output power. The value of quiescent current is mostly determined by the series pass element, topologies, ambient temperature, etc.

The mentioned parameters can define the LDO metrics such as dropout voltage, load regulation, and transient output voltage variation. Figure 3.4 shows the two regions of pass element operation, linear and saturation. In the linear region, the series pass element acts as a series resistor. In the saturation region, the device becomes a voltage-controlled current source. Voltage regulators usually operate in the saturation region.

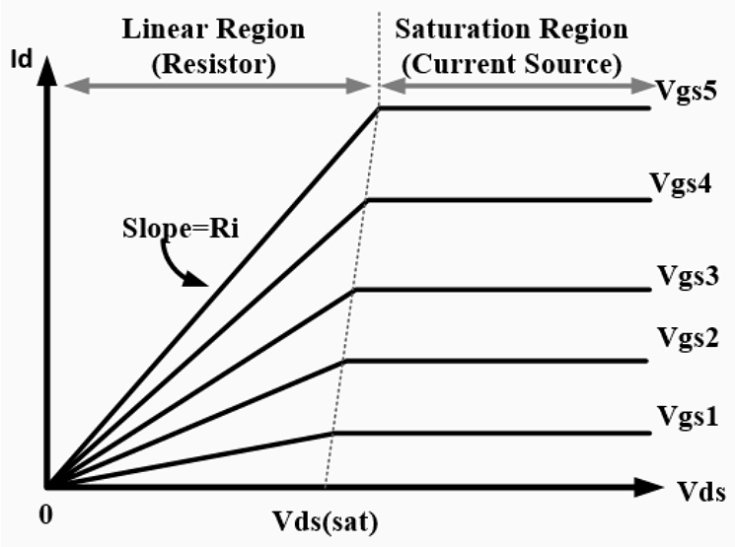


Figure 3.4 I-V Characteristic of n-channel MOSFET

As shown in Figure 3.2 (e), the utilized LDO in the proposed architecture consists of an error amplifier, a voltage reference, a pass transistor, and a feedback network which forms the regulator loop. The error amplifier uses a symmetrical operational transconductance amplifier (OTA). A common source PMOS device is chosen as a pass element. It exhibits the best overall design and yields a good compromise of drop-out voltage, quiescent current flow, output current, and speed [86]. The pass device must be physically large to obtain low drop-out voltage and high output current characteristics. Depending on the reference voltage (V_{REF}) and feedback resistors ($R_1$and R_2), the output voltage (V_{out}) can be calculated as in (3.1). When the LDO supplied by the suitable input voltage (V_{IN}), the V_{OUT} altered based on node voltages of the pass transistor. When the sum of V_{OUT} and threshold voltage of the pass transistor (V_{TH}) is less than the gate voltage of the pass transistor, the on-chip decoupling capacitor (C_L) is charged because the pass transistor is ON. In contrast, C_L is discharged and V_{OUT} reduces if the values of V_{REF} , R_1 and R_2 turn the pass transistor OFF.

$$V_{REG} = V_{REF} * (1 + \frac{R_1}{R_2}) \quad (3.1)$$

Dropout voltage ($V_{DROP-OUT}$) determines least input/output differential voltage where circuit fails to regulate against more reductions in input voltage which changes switch "on" resistance (R_{ON}) and load current (I_{LOAD}). The load regulation is the output impedance (R_O) of the circuit as seen in (3.3). R_{O-PASS} is the output impedance of the pass transistor while F and A_O are feedback factor and open loop gain of the system respectively. Last, transient voltage variation (V_{TVV}) occurs as the load current changes suddenly. The transient voltage variation is controlled by closed-loop bandwidth, a decoupling capacitor (C_L) and load current. The worst scenario happens when the load current steps from zero to its maximum ($I_{LOAD-max}$) as shown in (3.4) where Δt is the time needed for LDO to respond, and ΔV_{TVV} is output voltage variation. The transient voltage variation must be minimized to fulfill accuracy requirement of the system. In this design, the regulated output voltage V_{REG} and power supply noises are critical measurements for power/signal integrity [87].

$$V_{DROP-OUT} = R_{ON} * I_{LOAD} \quad (3.2)$$

$$R_O = \frac{\Delta V}{\Delta I} = \frac{R_{O-PASS}}{1+A_O*F} \quad (3.3)$$

$$\Delta V_{TVV} = \frac{\Delta t * I_{LOAD-max}}{C_L} \quad (3.4)$$

In this LDO architecture, a classical two-stage OTA is utilized [54]. It has an external bias current I_{BIAS} and single ended output. The first stage of OTA comprises a differential NMOS input pair M_1, M_2, and two PMOS transistors M_3, M_4 as an active load.

The drain current of $M_1$and M_2, I_{BIAS}/2, is mirrored to M_5 and M_6, respectively. As shown in (3.5), the transconductance of OTA (G_m) is dominated by a gain factor (B) and a transconductance of either input pair [54]. Thus, B can be changed according to the ratio of $(W/L)_{(5,6)}$: $(W/L)_{(3,4)}$ while $g_{m(1,2)}$ depends on DC current ($\sqrt{I_{BIAS}/2}$) which can be modified by $(W/L)_{(1,2)}$. Moreover, the voltage gain of OTA (A_v) is displayed on (3.6) where Z_L is the output impedance of the OTA which can be calculated as given in (3.7). OTA has high output impedance due to $r_{0(M_8)}||r_{0(M_6)}$ results in easier frequency compensation as C_L forms only dominate pole. Therefore, modifying C_L is the way of ensuring the stability of the system since it controls the gain bandwidth and hence the phase margin.

$$G_m = 2 * g_{m(1,2)} * B \tag{3.5}$$

$$A_v = G_m * Z_L \tag{3.6}$$

$$Z_L = \frac{R_0||R_L}{1+ R_0||R_L*C_L} \tag{3.7}$$

3. 4 On-chip 3D PDN models

In the literature, the basic electrical model of TSV is explored in [88]–[91], and [40]. TSV to TSV and TSV to wire couplings have a significant effect on signal integrity and power consumption due to their immense value (tens of femtofarads) [92], [93]. For an accurate simulation, the coupling effects including the TSV and wire dimensions have to be considered in the calculation [94]. For the layout, a conventional power delivery design method is implemented with middle metal layers to connect M1 layer in VDD/VSS rails to the top metal in global wires [95], [96]. On each tier, power delivery network is

usually structured as a regular mesh with metal tracks running in perpendicular to each other. Figure 3.5 (a) and (b) show the PDN mesh corresponding electrical model and the wire model, respectively [74].

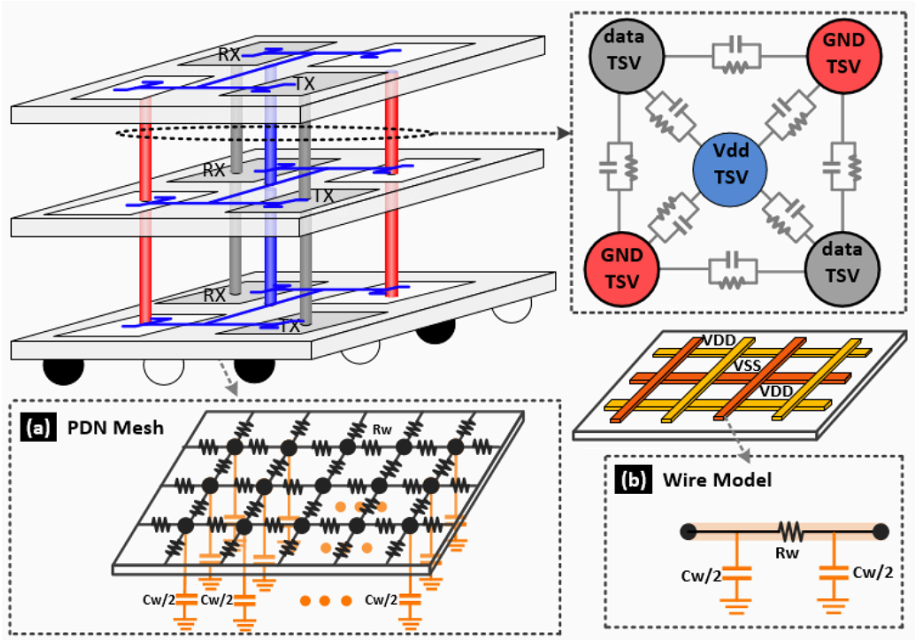


Figure 3.5 (a) PDN Mesh layout (b) Wire Model

3. 5 Power/ground wire model for 3D PDN

There are previous works on 3D PDN and TSV optimization such as in [76], [78] respectively. However, this work finds the best LDO placement and sizing for on-chip 3D PDN taking into consideration TSV coupling and I/O interface by the implementation of an intelligence computation, i.e., Evolutionary process. The implemented model for the PDN wiring and TSVs is similar to section 2.5 and 2.7, respectively.

3. 6 Evolutionary process of the PDN sub-blocks

To achieve optimum power efficiency, supply noises, and power/signal quality of a heterogeneous 3D PDN, the proposed architecture utilizes an evolutionary method by considering all design parameters, i.e., LDO placement, TSV geometry, and LDO circuit sizing, simultaneously.

When considering co-optimization of TSV structure and circuit sub-blocks, different performance parameters, i.e., area overhead and TSV coupling or speed and power consumption, conflict with one another. Thus, the optimization method requests trade-offs, namely Pareto-fronts, to find optimum design points among performance functions that compromise a computational efficiency [53], [54]. To attain the optimum performance of the proposed architecture regarding power consumption, voltage stability, and supply fluctuations, it is essential to find the best design points of the sub-blocks. It can be achieved by sizing their devices in the circuits such as LDO-error amplifier and TSV geometry. To improve the efficiency and reduce search space, a hierarchical multi-objective optimization is implemented, as explained in section 2.7. The chromosome size would be different depending on each sub-block structure. The best (minimum) fitness value of LDO optimization can be obtained using (3.8).

$$\text{Fitness Function} = \frac{k_0 \times pwr + k_1 \times V_{fluctuation}}{k_2 \times V_{REG}} \tag{3.8}$$

where pwr and $V_{fluctuation}$ represent the circuit power consumption and output voltage fluctuation, respectively.

In the fitness function, $k_0, k_1, \text{ and } k_2$ are relevant coefficients of the power

consumption, voltage fluctuation, and regulated output voltage, respectively. Then, the fitness function in (3.8) tries to extract all relevant PDN sub-blocks design parameters utilizing a genetic algorithm evolutionary process to achieve the best performance.

3. 7 Experimental Results

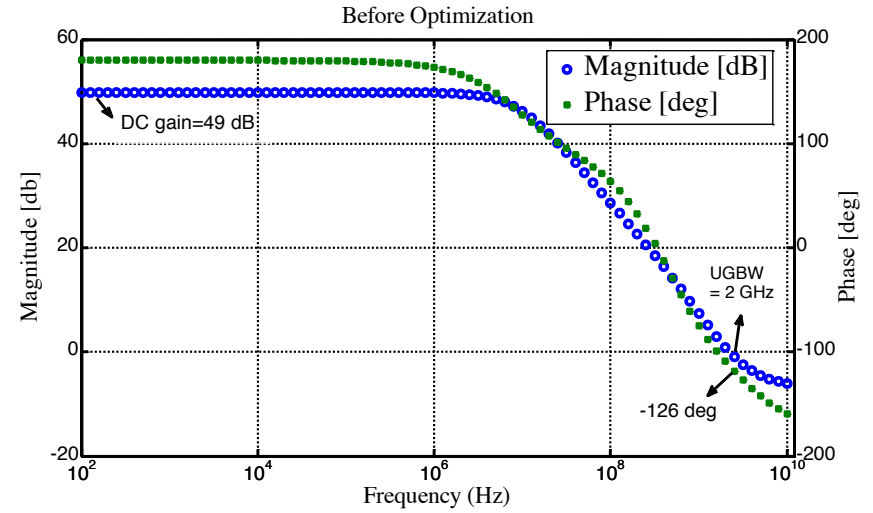
TSV geometry, LDO placement/sizing, and TSV/LDO interaction are the factors that significantly improve power efficiency, supply noises and obtain power/signal integrity. The evolutionary process is implemented on the above-mentioned factors. The circuit simulation is conducted by using Cadence Spectre and HSPICE-MATLAB interface in MATLAB R2016a environment. The post-layout extracted simulation was performed considering parasitic capacitance and resistance. Table 3.1 displays a summary of the optimization parameters [97]. we first designed each architecture and found initial design value and range for each variable. Since the initial values are located near the optimal point for the first generation, it can increase the capability of finding the optimal global design points and minimize the computational burden and convergence time. The creation of the first generation is also a trivial process to prevent the optimization of being trapped in local minima instead of the global minima. The convergence time increases considerably without evaluation of the initial operating points. The entire design took around 4 hours using high-performance computing (HPC) server with 128GB RAM 16-core Intel(R) Xeon(R) CPU X5560 @ 2.8GHz 107 processor.

Table 3.1 The Evolutionary Specifications of The Proposed Algorithm

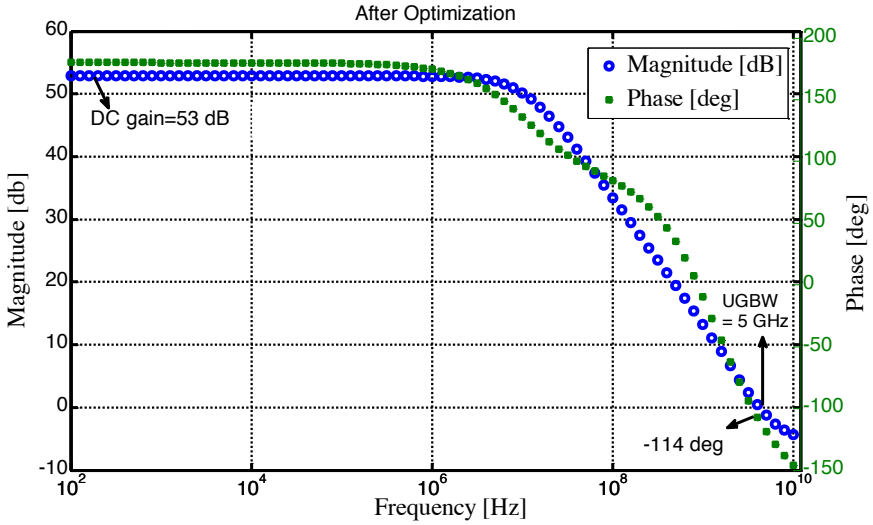
Number of Variables (sub-blocks/TSV)	14
Population Size	40
Number of Generation	20
Crossover	scattered
Mutation probability	0.05
Stopping Condition	Stall Generation
Total Performance Enhancement (%)	35
Total Convergence Time (hr.)	4

3.7.1 Evolutionary process of LDO regulator

Before the co-optimization of the entire 3D PDN sub-block sizing and placement, the evolutionary process is implemented on LDO to improve its performance parameters such as regulated output voltage and stability. First, the evolutionary process is implemented on the error amplifier inside the LDO. Figure 3.4 shows the improvement of its phase margin, bandwidth, and DC gain. After the evolutionary process, DC gain, bandwidth, and phase margin are improved from 49dB, 2GHz, and 54° as displayed in Figure 3.6 (a) to 53dB, 5GHz, and 66° respectively as seen in Figure 3.6 (b).



(a)



(b)

Figure 3.6 Simulated OTA in LDO (a) before and (b) after the evolution

Second, the evolutionary process is executed on the overall LDO performance. Figure 3.7 shows the regulated output voltage waveform after each optimization iteration. It also displays the output voltage waveform with (/w) and without (w/o) optimization (opt). As expected, the optimized output is significantly improved regarding the regulated voltage and stability.

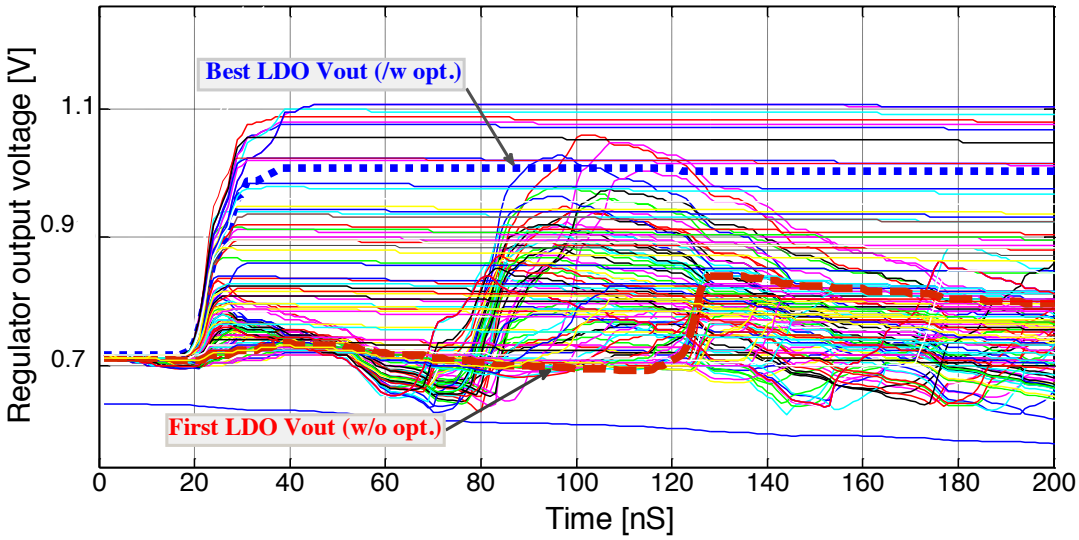


Figure 3.7 Evolution of LDO regulated output signal

The implemented evolutionary process on TSV can relatively improve the key performance parameters of the heterogeneous 3D stacked devices. Figure 3.8 shows the final output voltage waveform after each optimization iteration. It also displays the output voltage waveform with and without the optimization. As predicted, the optimized output voltage is significantly improved regarding latency, power, and overshooting. The black line shows the output voltage signal in the state-of-the-art 2D interconnect.

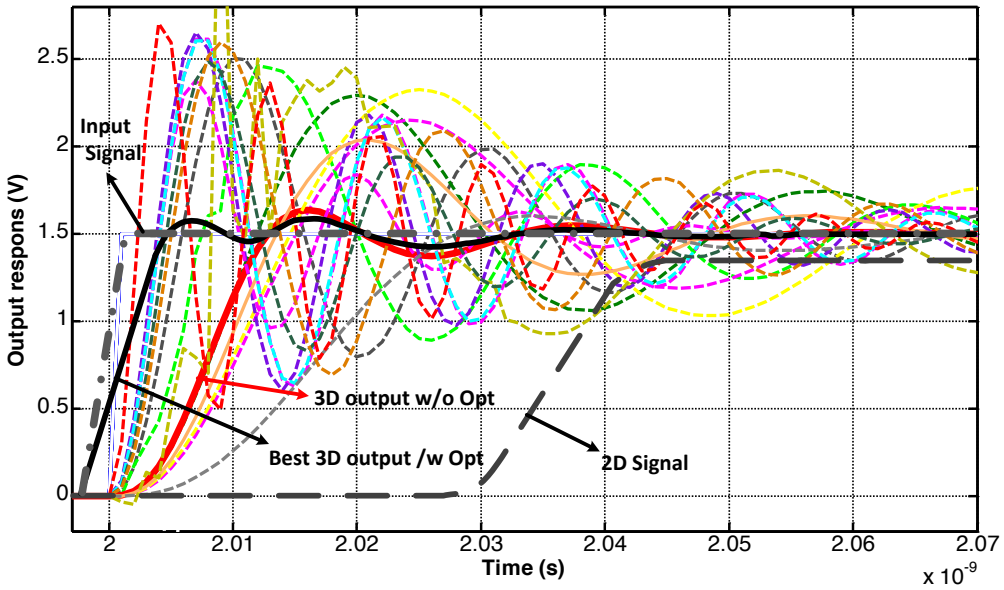


Figure 3.8 Evolution of 3D TSV output signal compared to the corresponding 2D signal

The recommended TSV channel design parameters, summarized in Table 3.2, are utilized to achieve the best 3D PDN performance.

Table 3.2 Recommended TSV Design Parameters for 3D Pdn

TSV Length (Height)	35µm
TSV Pitch	45µm
TSV Diameter	25µm

3.7.2 Impact of LDO placement on 3D PDN performance

Integrating LDO in the on-chip 3D PDN provides significant improvements in power supply noises and signal/power integrity. Thus, six performance aware cases of LDO placement in 3D PDN are simulated as shown in Figure 3.9.

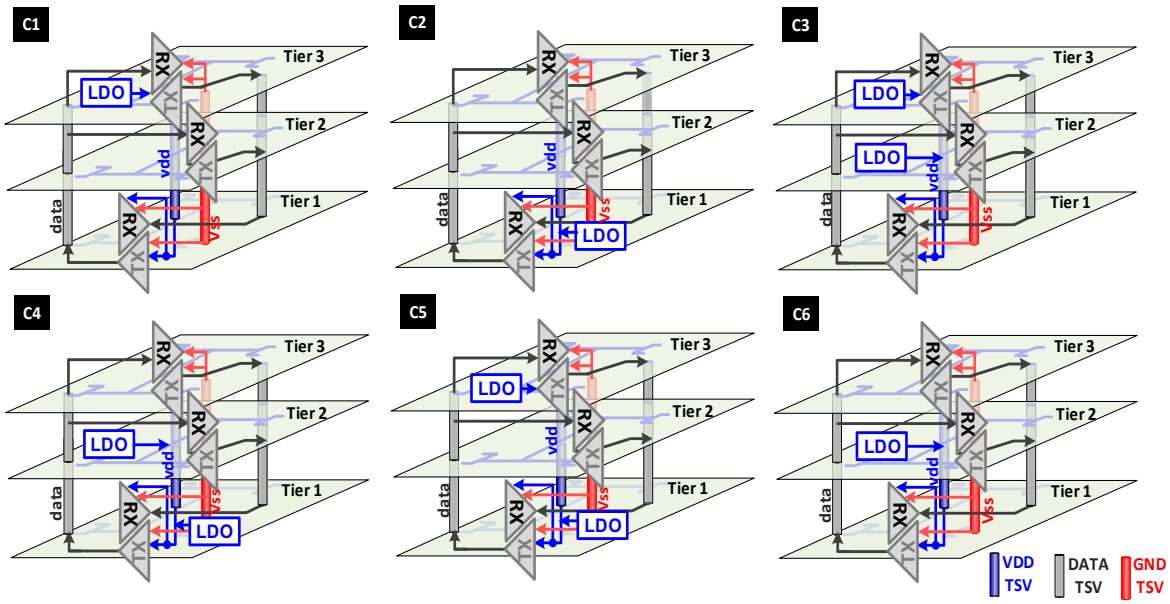


Figure 3.9 The Experimental setup for C1, C2, C3, C4, C5, C6 for on-chip 3D PDN

In case of C1, (0-0-1), LDO is placed on tier 3. LDO is placed on tier 1 in case of C2, (1-0-0). In case of C3, (1-1-0), LDOs are on tier 1 and 2. In case of C4, (0-1-1), LDOs are on tier 2 and tier 3. LDOs are placed on tier 1 and 3 in case of C5, (1-0-1). LDO is located on tier 2 in case of C6, (0-1-0). After obtaining the optimized sub-blocks and TSVs, the co-optimization is implemented to find the optimum architecture considering the complex interaction between TSVs, LDO and I/O interface for all cases. Since the power supply noise, i.e., maximum dc voltage droop is worst at the topmost tier, the experimental results are performed on tier 3. Transient analysis has been carried out to measure the PDN output voltage. Figure 3.10 demonstrates stable waveforms of the output voltage in cases of C3-to-C6. Moreover, to find the best LDO placement of 3D PDN, a performance comparison in power consumption, PDN output voltage, delay, and output voltage fluctuation is summarized in Table 3.3. The peak-to-peak noise introduced to the input supply voltage ($V_{in}\ tx$) is ~170 mV. The results significantly improved after implementing

the optimization algorithm. As expected, when the number of LDO decreases, the PDN power consumption reduces. In standby and power-nap mode of the heterogeneous 3D device application, LDO can be always on, thus the number of utilized LDOs should be minimum. For instance, C5 (1-0-1) has the best performance in terms of voltage fluctuation, but it consumes the power of 1.46X higher than C6 (0-1-0) after optimization. Consequently, the 3D PDN utilizing the common centroid placement of LDO (C6) has stable output voltage and minimum power consumption among C3-to-C6. It also has 10.2 mV output voltage fluctuations after optimization. Thus, C6 has the optimum overall performance.

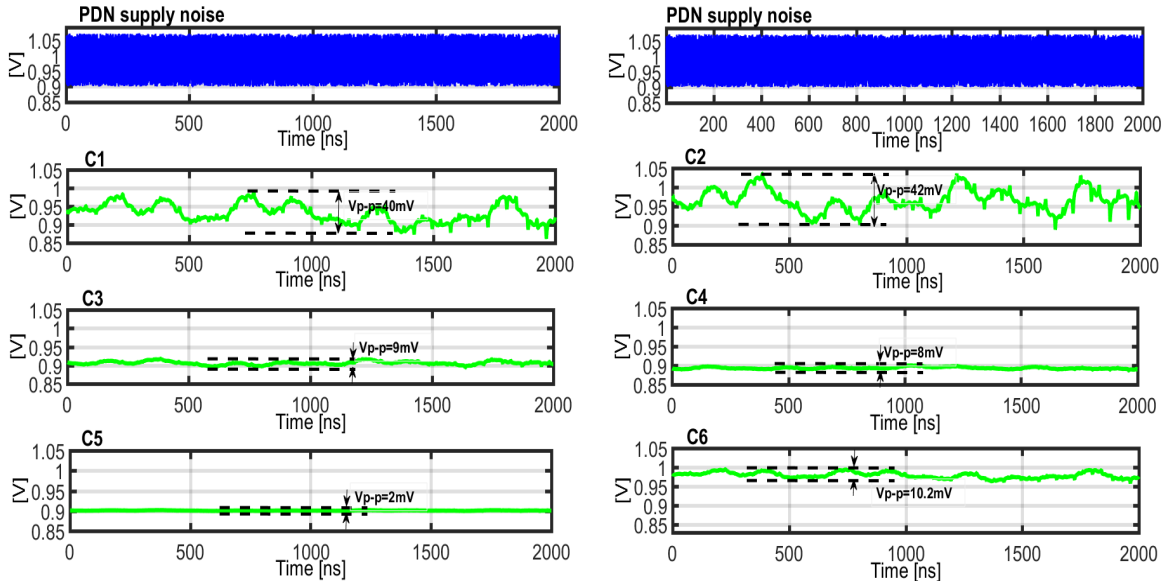


Figure 3.10 Transient analysis demonstrating PDN output voltage Vs. input supply noise

For verification, the performance-aware 3D PDN with common centroid placement of LDO has been fabricated in a 65nm CMOS technology. Because the academic 3D TSVs and TSIs prototyping are challenging, we implemented the proposed 3D PDN by using the feasible 3D micro-bumps which could provide similar dimensions, i.e., 25 μm diameter

and 45µm pitch, of the latest TSVs. The optimized micro-bumps geometry can be easily adapted to any 2.5D silicon interposer-and future 3D TSV-based ICs.

Table 3.3 Simulated Results of Different Cases of 3D PDN Architecture

PDN Case #	LDOs Placement T (1-2-3)	Total Power (mW)		PDN DC output (V)		PDN Output Voltage Fluctuation (mV)		Delay(ps)	
		w/o opt	**/w opt**	w/o opt	**/w opt**	w/o opt	**/w opt**	w/o opt	**/w opt**
C1	0-0-1	6.8	**4.25**	0.9	**0.92**	112	**40**	82	**49**
C2	1-0-0	6.3	**4.18**	0.9	**0.95**	118	**42**	85	**51**
C3	1-1-0	8.4	**5.95**	0.9	**0.91**	70	**9**	97	**88**
C4	0-1-1	9.2	**6.26**	0.9	**0.9**	65	**8**	93	**86**
C5	1-0-1	8.8	**6.16**	0.9	**0.9**	30	**2**	95	**87**
C6	0-1-0	6.6	**4.22**	0.9	**0.96**	75	**10.2**	78	**46**

The 3D structure block diagram is shown in Figure 3.11 (a). Figure 3.11 (b) shows die photographs of the fabricated chips including transceivers, LDO, and 3D pads.

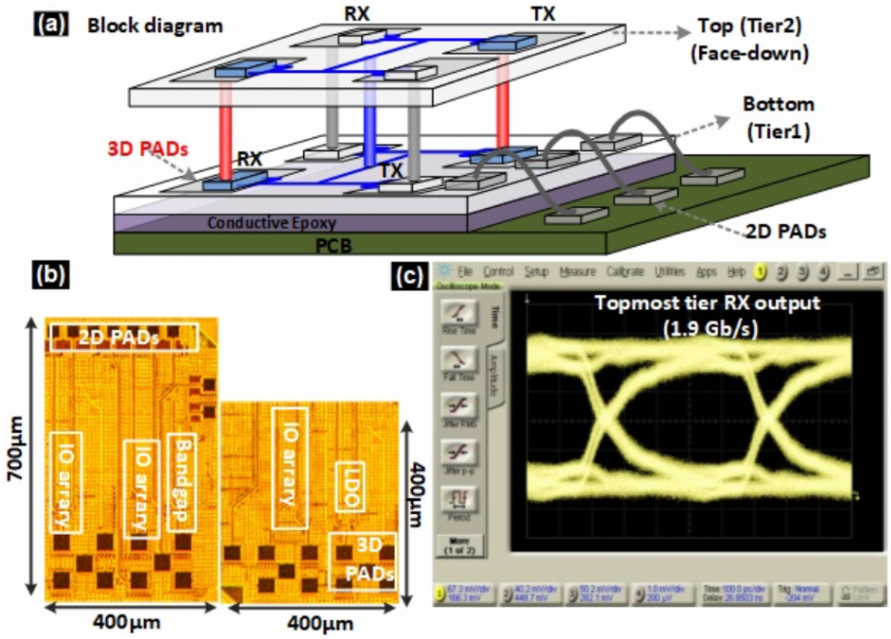


Figure 3.11 (a) IC prototyping for the proposed 3D-PDN (C6) (b) Microphotograph of each tier die, and (c) measured eye diagram

The accomplished data rate is up to 1.9 Gb/s and the power consumption is around 4.22 mW. Figure 3.11 (c) shows the measured eye diagram of the topmost RX output of the proposed architecture to verify that the I/O interface is functioning properly as the power/signal integrity and stability of supply are achieved.

3. 8 Conclusion

This chapter presented a performance-aware 3D PDN utilizing a common-centroid regulator placement to increase power/signal quality, aggregate I/O data bandwidth, and improve energy efficiency [98]. To complete system-on-chip (SoC) design and include the power management circuit, LDO regulator is integrated into the I/O interface and data/power TSVs. An evolutionary algorithm was conducted to achieve the minimum energy consumption and supply noises considering the complex interaction among TSVs, LDO, and I/O interface. Consequently, the fabricated chip achieved a total power of 4.22mW and 1.9 Gb/s data rate for the integrated I/O interface.

Chapter 4

3D Baseband Memory I/O Interface

4. 1 Introduction

The critical demand for higher-power efficiency, wider bandwidth operation, and more compact packaging in high-performance computing (HPC) and memory interface design has significantly increased as the mobile devices continue to accelerate media-intensive data communication and graphics processing capabilities. However, the current 2D memory interface technologies for multi-drop memory bus have critical limitations, especially limited bandwidth, high-power consumption, low-density packaging, and large form-factors due to 2D's boundary pad/packaging limitations [15], [99], [100]. High-bandwidth memory (HBM) solutions are promising to induce the latest innovations in IC packaging technologies by utilizing vertically short interconnects and die stacking. The basic HBM structure includes a main bottom logic die and the top stacked DRAM dies, with the vertical interface as shown in Figure 4.1 (b) [15], [28], [101]. The power and ground channels are common to support all the signal channels. The memory (i.e. DRAM) and the host chip (i.e. CPU) interconnection can be implemented using two solutions, chips-on-inter-poser (2.5D) and chip-on-chip (3D). In 2.5D HBM, the DRAM is flipped and stacked on the interposers (i.e., micro-bumps). This type of stacking (i.e., face-to-face) has wide I/O interconnects which decrease the heat dissipation [102]. HBM can provide high-performance

systems because of providing scalability for memory capacity, smaller silicon area, and reduced I/O power dissipation [101]. In a situation that DDR3 memory interfaces operate at maximum 2Gb/s bandwidth with power efficiency of 15.8pJ/b/pin, the state-of-the-art 4-stacked HBM maximum bandwidth is up to 256 GB/s [21], [25], [28], [103], [104]. The first generation of HBM (i.e. HBM1) works at a maximum 1 Gb/s/pin data rate with 4 stacked memories. While HBM2 generation can provide around 2Gb/s/pin maximum speed in 2/4/8 number of stacked memories in CMOS technologies. Thus, the HBM generations can fulfill the requirements for a wide range of high-bandwidth applications. It also has been shown that HBM can significantly increase signal integrity, throughput, and memory capacity. However, fabrication of 8 stacked dies in HBM2 looks challenging and rarely implemented properly [21], [25], [101]–[104]. Current mobile memory interfaces are based on two I/O topology buses (i.e., point-to-point and multi-drop) which can use voltage or current mode signaling [105], [106].

Point-to-point topologies support one device at each end of the signal line while in multi-drop interface several devices, in this case memory modules, are connected to the memory controller with shared interconnects. Point-to-point topologies have better signal integrity properties and permit higher bus speeds, but cannot be upgraded with multiple modules. While point-to-point channels are advantageous in achieving higher data rates because of the absence of undesired reflections that occur at each stub of multi-drop channels, they are not suitable for high-capacity, high-throughput memory systems such as transaction servers or cloud computing nodes due to their prohibitively large PCB routing area connecting the memory chips.

Figure 4.1 (a) and (b) show the conventional 2D and 3D DRAM topologies in point-to-point and multi-drop interface, respectively. Figure 4.1 (c) shows a simplified circuit schematic of the conventional multi-drop memory interface using a simple push-pull CMOS logic with key

advantages of simplicity and low silicon area. However, the data bandwidth is limited and very sensitive to on-die-termination (ODT) schemes in the multi-drop cases. Thus, 2D I/O utilizing stub series terminated logic (SSTL) or high-speed transceiver logic (HSTL) are used in DRAM-based DDR memories to extend the bandwidth (i.e. 2 Gb/s in 3D point-to-point I/O) [107], as shown in Figure 4.1 (d) and (e), respectively. In SSTL architecture, each busline has its own termination resistor at one end that connects the busline to the termination voltage. The SSTL bus has also a serial resistor, R_S, between the busline and the I/O to reduce the reflection. HSTL input buffer has strong noise resistance ability and is usually used as a clock input buffer. HSTL output buffer is the I/O interface of choice for high-speed memory applications and is ideal for driving address busses to multiple memory banks. There are four types of HSTL output buffers and the commonly used type (i.e., HSTL_I) is shown in Figure 4.1 (e).

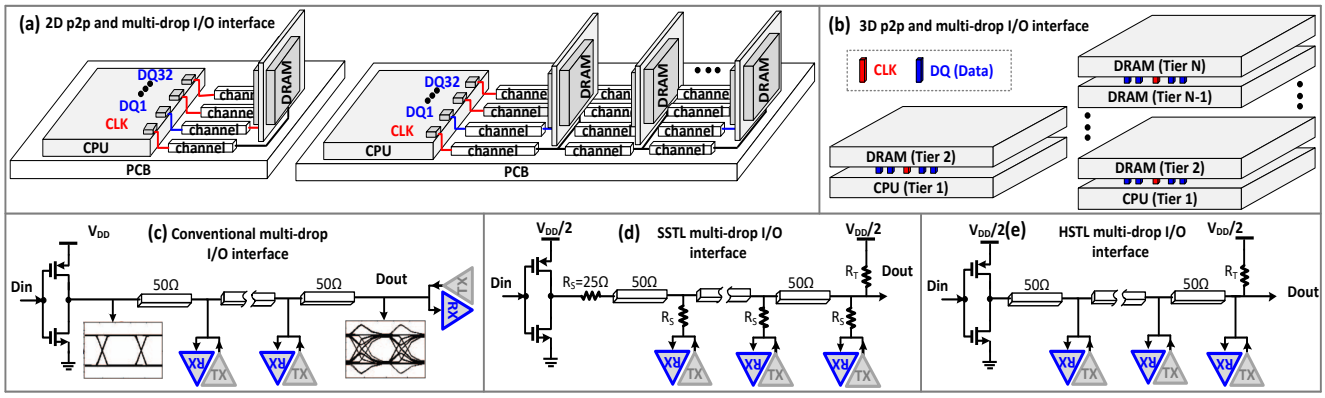


Figure 4.1 Typical (a) 2D and (b) 3D point-to-point and multi-drop memory I/O interface; simplified scheme 2D (c) conventional (d) SSTL (e) HSTL interfaces

As mentioned earlier, in a 2D multi-drop bus, memory devices are mounted on multiple dual in-line memory modules (DIMMs) which causes significant signal integrity degradations, severe crosstalk, and prolonged latency. The problems are due to channel impedance discontinuity, impedance mismatch, undesired added disturbances, and heavy/long loading. The attenuation increases with channel length and signal frequency. 2D PCB metal line can cause serious

impedance mismatch problem if the circuit is not designed carefully. Especially, when adding modules in order to expand memory, the channel termination situation changes and consequently impedance mismatch is hard to avoid. A desired impedance matching network increases the circuit design complexity. Moreover, SSTL and HSTL I/O interface have inherent limitations such as increased power consumption and sensitivity to ODT termination configuration. Also, their packaging size increases due to using termination and series resistors [108], [82].

Within the same bandwidth, 3D TSV channels have less signal attenuation compared to 2D PCB metal lines due to their shorter length. Therefore, they can be simply integrated into the system chip package and can be less affected by the memory package capacitance. In 3D HBM with heterogeneous structure, the stacked DRAM tiers on top of the logic die are interconnected through TSV channels which enable data communication between CPU and multiple DRAMs. Utilizing stacked-memories, TSV channels or micro-bumps, HBM can offer higher signal integrity, power/area efficiency, and bandwidth compared to the conventional 2D structure. This improvement can be in both point-to-point and multi-drop memory interfaces. Moreover, the yield challenges related to a massive heterogeneous memory system packaging and volume production can be resolved and it leads to achieving high yield in packaging.

Despite the aforementioned merits, the current 3D IO prototype using push-pull I/O signaling still has some limitations on bandwidth, signal integrity, and future power-scalability. Also, a comprehensive analysis and design of 3D I/O considering both bandwidth/energy efficiency and termination impacts on the overall 3D system performance does not exist yet. To overcome these technical challenges, we have proposed an energy-efficient point-to-point and multi-drop memory interface using novel resistive feedback for utilizing in HBM. Furthermore, for the highest bandwidth/energy efficiency and more compact packaging, the 3D transceiver, and

termination topologies including TSV channels are fully optimized through a multi-objective evolutionary algorithm.

4. 2 The proposed transceiver circuit and I/O interface

Figure 4.2 shows a conceptual view of the proposed 3D architecture for a four-drop source-synchronous interface where the TX, on the CPU tier (i.e., first tier), is connected to the three stacked DRAM tiers through the TSV channels and sends data to them. The proposed interface also utilizes a switched-diode termination (SDT) to have a low-power on-chip single-ended signaling through a 3D vertical channel. Since the CMOS inverters can provide enough gain while only consuming the small amount of power, the proposed TX/RX employ an inverter-based cascade amplifier to minimize the power consumption. Due to the simplicity of the proposed 3D memory I/O interface, it can be fully compatible to mainstream low power double data rate (LPDDR) HBM based on joint electron device engineering council (JEDEC) specification [109], [110].

In order to design high-speed signal TSV channels and perform signal integrity analysis for the advanced 3D IC design, much research has been done to model and analyze the TSVs. In this work, TSVs are electrically modeled using the design parameters such as geometric and material information up to gigahertz range [48], [111]. Implementation details of the key building blocks of the proposed 3D memory I/O interface architecture are discussed below.

4. 3 Transmitter

Figure 4.2 (a) shows the TX side of the proposed architecture consisting of an inverter with a feedback resistor parallel connection as an equalizer driver. It performs a pre-emphasis operation

by adding current over the transition time of D_{IN}.

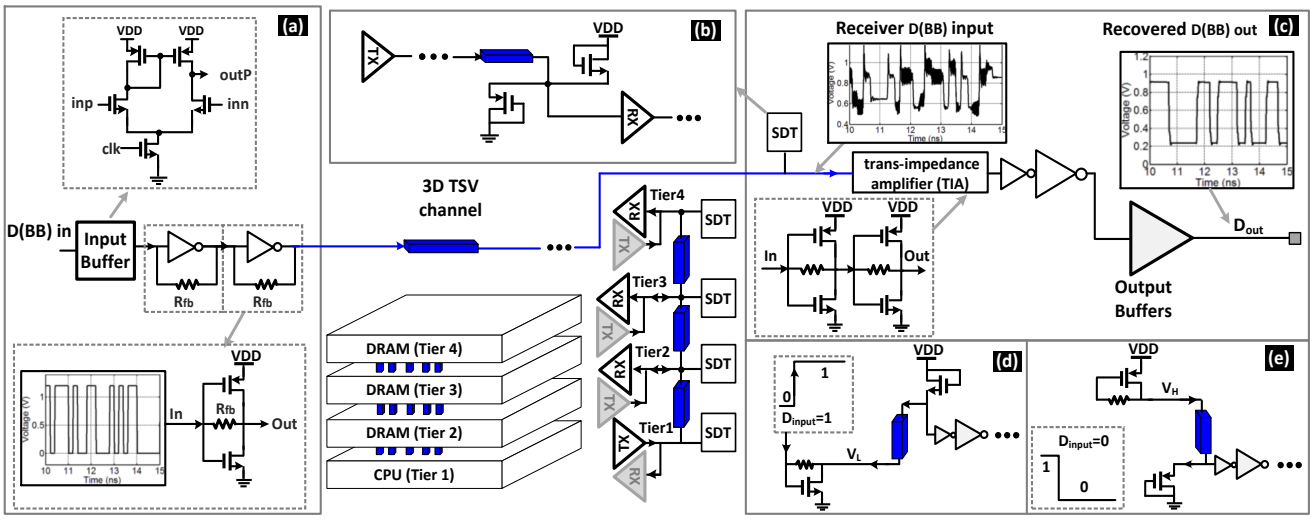


Figure 4.2 schematic of the proposed 3D transceiver blocks
(a) transmitter (b) termination (c) receiver (d) high and (e) low input signaling scheme

This architecture can also provide an increased bandwidth/power efficiency since the resistive feedback gives smaller output voltage swing and less sensitivity to the termination in case of multi-drop I/O. The CMOS logic works as a TX output driver which drives the TSV channel and enables data transmission. Also, to minimize the delay, voltage offset, mismatch, and noise effect, an input buffer using a self-biased differential amplifier is implemented. The baseband TX output is matched with the baseband RX termination and controls the threshold voltage levels (high /low) of the input CMOS inverter. The TX and RX communicate via common-mode signaling.

4. 4 Termination

Using diode nonlinearity as low-impedance termination at the receiver leads to a reduced voltage swing, improved bandwidth performance, and causes impedance matching. As shown in Figure 4.2 (b), an nMOS diode in series with a pMOS diode from V_{DD} to V_{SS} shape the RX diode termination in this work. At each moment, only one diode (pMOS or nMOS) shapes the termination to avoid

the short-circuit current through V_{DD}-V_{SS} path [111]. This inverter cascading configuration as well as diode termination nulls the minor changes of the current input by matching.

4. 5 Receiver

Figure 4.2 (c) shows the proposed RX circuitry. To recover data coming from the channel, an inverter-cascade voltage amplifier has been implemented. This topology offers a large low-frequency voltage gain and unity gain bandwidth (UGBW) with a quite limited 3dB bandwidth. Since the maximum output data rate is limited by 3-dB bandwidth, an equalizer circuit should be implemented to increase the bandwidth. To accomplish this aim, a feedback branch has been shaped using a resistor across two stages of the inverter amplifiers. As illustrated in Figure 4.2 (d) and (e) the TX output controls the RX input voltage by means of the termination circuit and depending on this output value, the created signal path causes the termination to be switched into a diode connected MOS either to V_{DD} or V_{SS}.

4. 6 TSV channel

In the proposed architecture, the TSV channel design parameters have considered same as section 2.4. Each the signal channel is surrounded by two ground channels inside the silicon substrate, simulated in both HSPICE and HFSS. The ground TSVs provide shielding for the signal TSV and make the 3D routing signal integrity less noise-sensitive. The analytic equations are derived from the physical configuration including the design parameters. A detailed evaluation is discussed and analyzed in [49], [29]. As mentioned in chapter 2, the proposed design utilizes a heuristic multi-objective evolutionary method to compromise the general computational efficiency of the optimization process. A multi-objective algorithm is utilized for the evolution because some performance parameters such as power, bandwidth, and area efficiency are non-differentiable and

non-commensurable. In case of n-dimensional decision variable (i.e., $x = \{x_1, \dots, x_n\}$) in the solution space X which shows the entire design space, the system optimization should be formulated in a way to find a vector x* which gives a minimum set of *K* objective functions $z(x^*) = \{z_1(x^*), \dots, z_k(x^*)\}$ considering a general restriction as $g_j(x^*) = b_j$ for j = 1, …, m, and bounds on the decision variables. As an example of fitness function (i.e. ff_1, ff_2) (4.1) and (4.2) are proposed for TSV and circuit co-optimization, respectively.

To fulfill this objective, a solution with lower overall fitness would require lower settling time, error band, overshoot percentage, and peak time and higher peak absolute value.

$$ff_1 = \frac{k_0 \times \text{Settling time} + k_1 \times \text{errorband} + k_2 \times \text{Overshoot} + k_3 \times \text{peak\_time}}{k_4 \times \text{Peak\_absolute\_value}} \tag{4.1}$$

$$ff_2 = \frac{k_0 \times \text{Power consumption} + k_1 \times \text{Sillicon area}}{k_2 \times \text{Signal integrity}} \tag{4.2}$$

The genetic algorithm tries to extract all relevant TSV and circuit sub-block parameters by sizing through a reduced set of independent variables, namely gene. These variables shape the chromosome strings as explained in section 2.7.

4. 7 Measured results

The proposed point-to-point, 4-drop, and 8-drop transceivers were fabricated in a 130 nm CMOS process to demonstrate multi-drop bus in a memory stacking system and all of them were taken to be measured. The memory interfaces were assembled in a 2.5-D micro-bump-based structure [21]. To verify the results, three cases were studied based on Figure 4.3 schematics (i.e. point-to-point, 4-drop, and 8-drop).

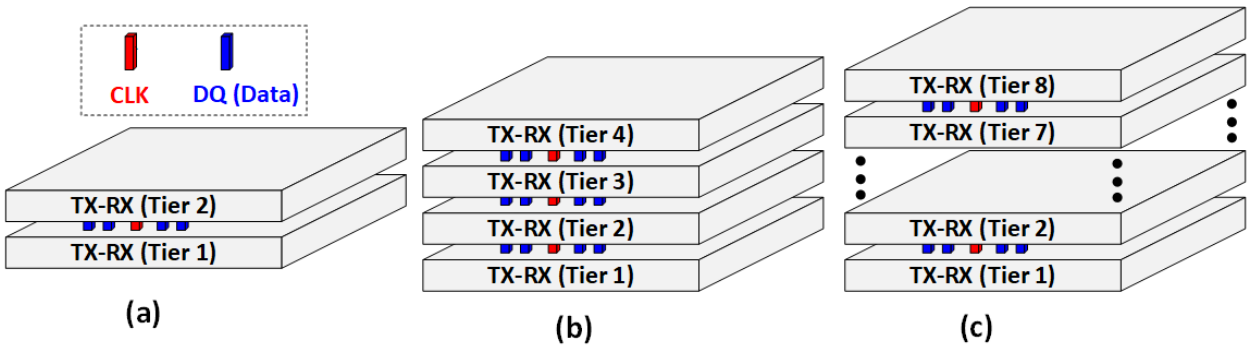


Figure 4.3 Schematic of the implemented 3D memory interface architecture

(a) point-to-point (b) 4-drop (c) 8-drop

Die photos of the test layouts for the top and bottom tiers are shown in Figure 4.4. Figure 4.4 (a) shows the bottom tier of the 8-drop transceiver, Figure 4.4 (b) shows the top tier of the 8-drop transceiver, the top tier of 2- drop and 4-drop transceivers are shown in Figure 4.4 (c), and finally the bottom tier of 2-drop and 4-drop transceivers is shown in Figure 4.4 (d).

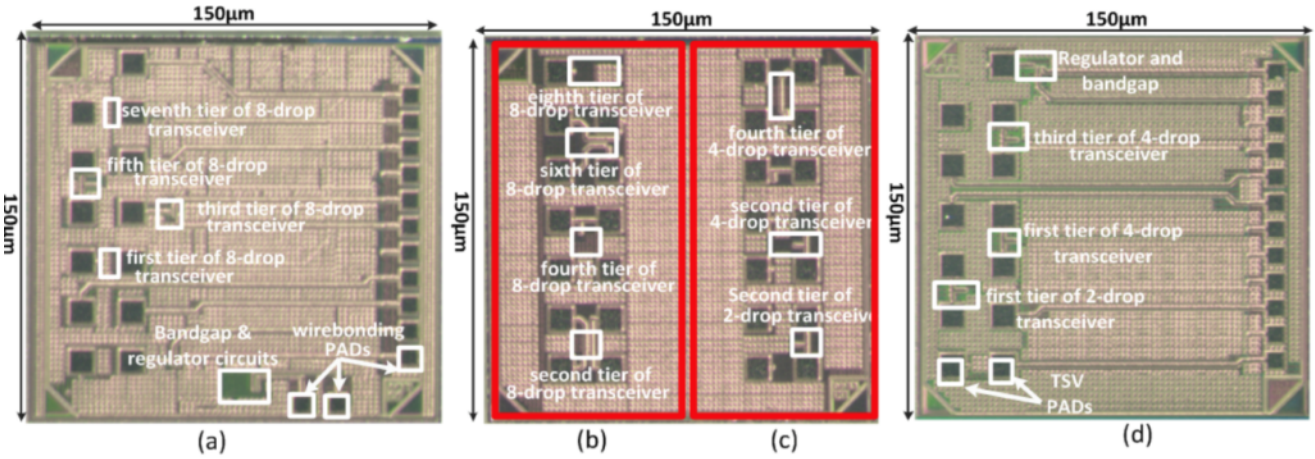
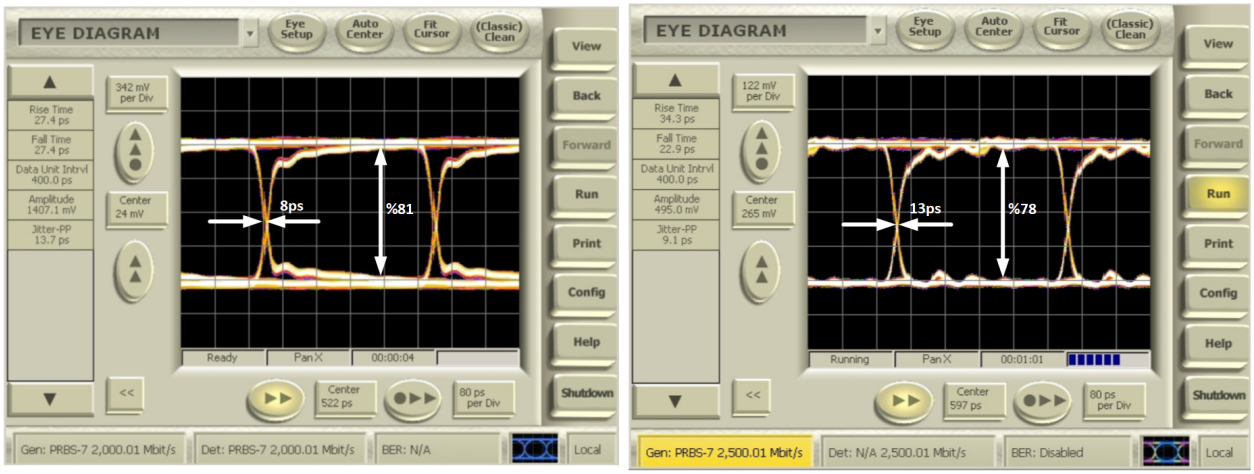


Figure 4.4 Die photos of the top and bottom tiers are shown
(a) bottom tier of 8-drop transceiver (b) top tier of 8-drop transceiver (c) top tier of 2- drop and 4-drop transceiver (d) bottom tier of 2-drop and 4-drop transceivers

Figure 4.5 (a) and (b), show the measured eye diagrams of the implemented 3D point-to-point memory interface at 2 GHz and 2.5 GHz, respectively. As we can realize from this figure, the designed point-to-point memory interface can achieve data rate even beyond 2.5 GHz which is a significant improvement in 3D memory interface bandwidth compare to the published works

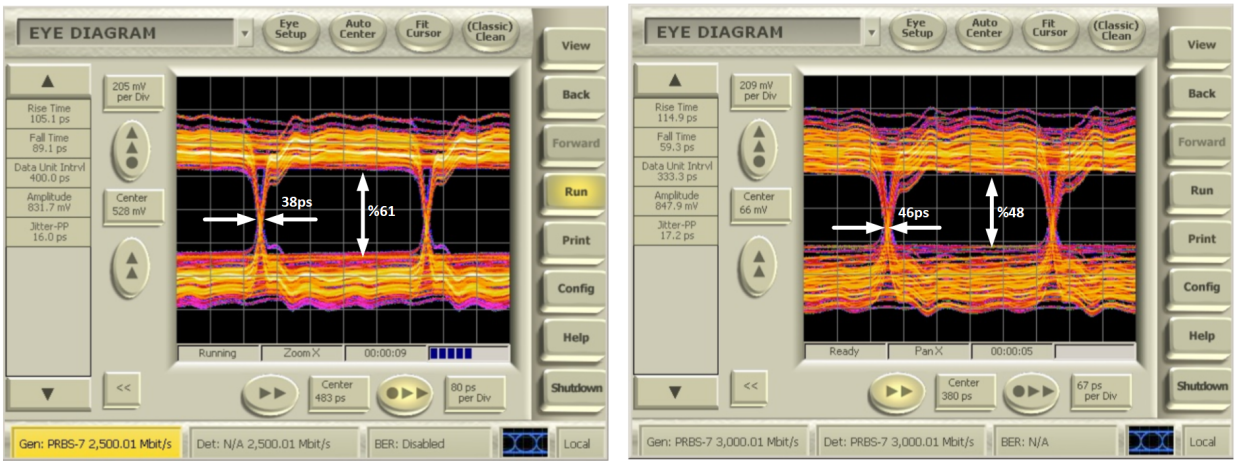
so far. As TSVs have many improved electrical characteristics, the parallel termination circuits can be removed; consequently, the capacitive load becomes considerably low. Also, the BBTX buffers and the logic die interconnections are able to drive outputs, properly.



(c) (d)

Figure 4.5 measured eye-diagram for 3D 2-drop I/O interface at (a) 2 Gb/s (b) 2.5 Gb/s

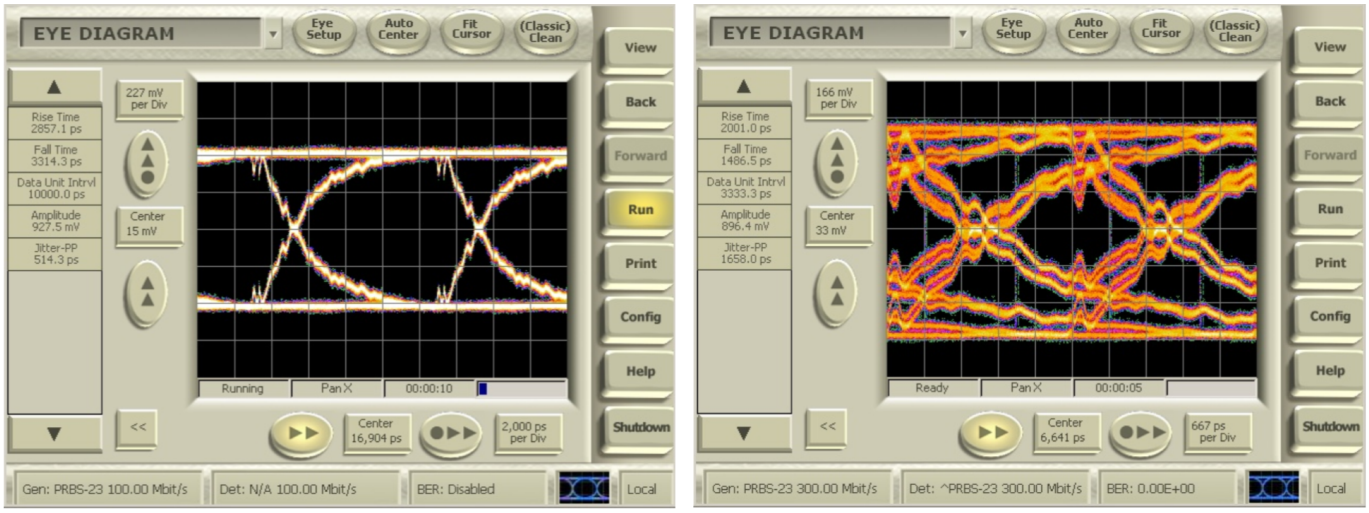
Figure 4.6 (a), shows the measured eye diagrams of the 3D 4-drop memory interface at 2.5 GHz. This figure shows for a 4-drop memory interface the eye-diagram has %61 heigh and 38ps deterministic jitter. While Figure 4.6 (b) shows the degradation of the measured eye diagram parameters for the 4-drop memory interface at 3 GHz. As this figure shows, for a 4-drop memory interface at 3 GHz, the eye-diagram has %48 height and 46 ps deterministic jitter.



(a) (b)

Figure 4.6 measured eye-diagram for 3D 4-drop I/O interface at (a) 2.5 Gb/s (b) 3 Gb/s

Figure 4.7 (a) and (b), show the measured eye diagrams of the fabricated 3D 8-drop memory interface at 100 MHz and 300 MHz, respectively. This figure shows the maximum bandwidth for an 8-drop memory interface based on this work design is around 300 MHz and further frequency increment would not have an accurate data transfer.



(a) (b)

Figure 4.7 measured eye-diagram for 3D 8-drop I/O interface at (a) 100 Mb/s (b) 300 Mb/s

Figure 4.8 shows an analysis of power consumption versus the supply voltage in a point-to-point 3D memory interface. The proposed architecture power consumption decreases quadratically as supply voltage scales down due to P=CV^2f. The proposed baseband I/O interface accomplishes around 3 Gb/s/pin with power consumption of 7.841 mW. Thus, this architecture provides a promising solution to increase power efficiency and data rate for the future of mobile memory interface.

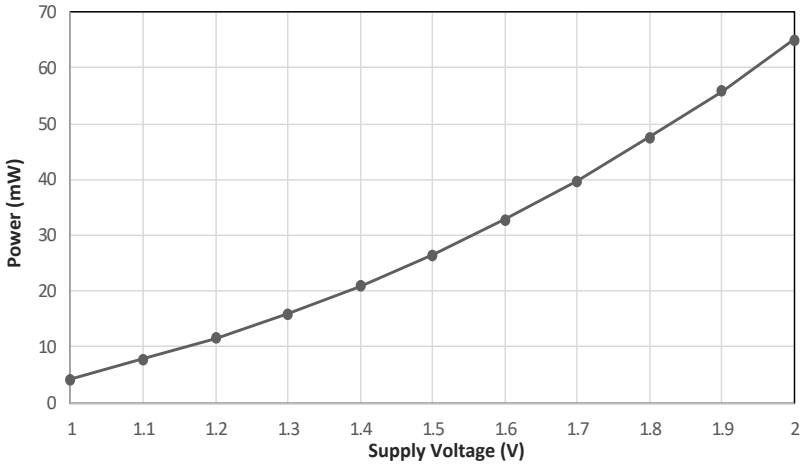


Figure 4.8 Simulated power consumption for 3 Gb/s in 130nm CMOS technology.

4. 8 Conclusion

This chapter presented a 3D low-power voltage-mode single-ended transceiver and the interface in both point-to-point and multi-drop bus configuration for on-chip signaling [112]. The transmitter architecture utilizes novel resistive feedback amplifying method to drop signaling power and the TSV channel swing is limited to 50 mV. Moreover, the amplifier uses cascaded inverters to improve the bandwidth. Combining the proposed architecture with an evolutionary algorithm provides the optimum energy and performance efficiency as well as more compact packaging. To verify the feasibility, three transceiver topology cases (point-to-point, 4-drop, and

8-drop) are implemented in 130 nm CMOS technology at 1.2 V supply. Achieving an energy efficiency of 9.5 fJ/b/mm at 3 Gb/s, the proposed 3D memory interface schemes can be considered for future mixed-signal 3D systems and their global signal routing.

Chapter 5

Conclusion and Future Work

5. 1 Conclusion

This dissertation lays the groundwork for a new approach, namely 3D integration, in analog, digital and mixed-signal systems packaging. The idea of 3D IC integration has been implemented on different circuits (i. e. pipeline ADC, power distribution network, and clock distribution network) to minimize the interconnection length and consequently the related noise, crosstalk, die area, and power consumption.

The advantages of 3D stacking of mixed-signal ICs can be summarized as follows. 1) ideally eliminating the switching noise disturbance due to the shared substrate in the 2D case, and therefore avoiding the unnecessary guard rings or deep well process; 2) elevating the performance and reducing the power consumption of digital circuit by applying different process technologies to digital and analog designs, respectively (e.g., a 32nm digital design can be stacked with a 90nm analog design); 3) also shrinking the concerned footprint and shortening the length of global interconnects.

In addition to the 3D structure, some complementary approaches such as evolutionary algorithms and MRL are utilized to further improve the accuracy, bandwidth, performance, and power/area efficiency. The illustrated results show significant improvements in all types of the

implemented topologies in term of performance, bandwidth, power, and silicon area efficiency.

5. 2 Future Work

The proposed novel design for 3D integration can be a promising solution to alleviate many issues exposed in the current technology scaling of IC design. This approach can increase the capability of heterogeneous integration in mixed-signal systems (e.g. ADCs and DACs design) and address the design issues in conventional analog/digital mixed-signal.

3D integration of different types of ADCs and DACs such as Flash ADC, SAR ADC, current steering DAC can be a revolution in the mixed-signal systems which addresses the most key issues of the current process nodes in IC design. Thus, the future work should engage in the fabrication of the mentioned circuits to prove the benefits of the 3D application.

The 3D integration also requires the address of increased reliability challenges, caused by technology scaling which leads to strong electric fields and degrades the overall system robustness. Thus, as a part of future work, we can investigate the robustness of the 3D systems to consider different forms of uncertainty in the 3D integration and try to optimize the systems in a way to have the maximum reliability. Also, since technology scaling creates more significant process variations, the yield-aware design plays a crucial role in the future of 3D integration and packaging. The work can be focused on extracting the models for the different fabrication process, design, and environmental parameter optimization.

REFERENCES

[1] Y. Lim, S. Member, and M. P. Flynn, "Pipeline ADC Using Self-Biased Ring Amplifiers," *IEEE J. Solid-State Circuits*, vol. 50, no. 10, pp. 2331–2341, 2015.

[2] J. Knechtel, I. L. Markov, and J. Lienig, "Assembling 2-D blocks into 3-D chips," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 31, no. 2, pp. 228–241, 2012.

[3] D. Malta *et al.*, "TSV-Last, Heterogeneous 3D Integration of a SiGe BiCMOS Beamformer and Patch Antenna for a W-Band Phased array Radar," *Proc. - Electron. Components Technol. Conf.*, vol. 2016–Augus, pp. 1457–1464, 2016.

[4] W. Liu, G. Chen, X. Han, Y. Wang, Y. Xie, and H. Yang, "Design Methodologies for 3D Mixed Signal Integrated Circuits : a Practical 12-bit SAR ADC Design Case."

[5] K. Kim, "Silicon technologies and solutions for the data-driven world," *Dig. Tech. Pap. - IEEE Int. Solid-State Circuits Conf.*, vol. 58, pp. 8–14, 2015.

[6] M. M. Shulaker, T. F. Wu, M. M. Sabry, H. Wei, H. P. Wong, and S. Mitra, "Monolithic 3D Integration : A Path From Concept To Reality," *2015 Date*, pp. 1197–1202, 2015.

[7] T. Kondo *et al.*, "A 3D stacked CMOS image sensor with 16Mpixel global-shutter mode and 2Mpixel 10000fps mode using 4 million interconnections," *IEEE Symp. VLSI Circuits, Dig. Tech. Pap.*, vol. 2015–Augus, no. V, pp. C90–C91, 2015.

[8] R. Li, C. Jin, S. C. Ong, T. G. Lim, K. F. Chang, and S. W. Ho, "Embedded wafer level

packaging for 77-GHz automotive radar front-end with through silicon via and its 3-D integration," *IEEE Trans. Components, Packag. Manuf. Technol.*, vol. 3, no. 9, pp. 1481–1488, 2013.

[9] N. Mirzaie, A. Alzahmi, C.-C. Lin, and G.-S. Byun, "A 3D flash ADC structure for high-speed communication applications," in *2018 IEEE 8th Annual Computing and Communication Workshop and Conference, CCWC 2018*, 2018, vol. 2018–Janua.

[10] N. Mirzaie, A. Alzahmi, H. Shamsi, and G. Byun, "Three-Dimensional Pipeline ADC Utilizing TSV/Design Optimization and Memristor Ratioed Logic," *IEEE Trans. Very Large Scale Integr. Syst.*, 2018.

[11] S. Chan and P. Restle, "A resonant global clock distribution for the cell broadband engine processor," *IEEE J. Solid-State Circuits*, vol. 44, no. 1, pp. 64–72, 2009.

[12] L. Zhang, A. Carpenter, B. Ciftcioglu, A. Garg, M. Huang, and H. Wu, "Injection-locked clocking: A low-power clock distribution scheme for high-performance microprocessors," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 16, no. 9, pp. 1251–1256, 2008.

[13] X. Chen, T. Zhu, W. R. Davis, S. Member, and P. D. Franzon, "Adaptive and Reliable Clock Distribution Design for 3-D Integrated Circuits," vol. 4, no. 11, pp. 1862–1870, 2014.

[14] M. Seok, D. Blaauw, and D. Sylvester, "Robust clock network design methodology for ultra-low voltage operations," *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 1, no. 2, pp. 120–130, 2011.

[15] B. Leibowitz *et al.*, "A 4.3 GB/s Mobile Memory Interface With Power-Efficient Bandwidth Scaling," vol. 45, no. 4, pp. 889–898, 2010.

[16] K. Zhong, D. Liu, L. Long, J. Ren, Y. Li, and E. H. M. Sha, "Building NVRAM-Aware Swapping Through Code Migration in Mobile Devices," *IEEE Trans. Parallel Distrib.*

Syst., vol. 28, no. 11, pp. 3089–3099, 2017.

[17] "Maximizing Performance," 1988. .

[18] "International Technology Roadmap for Semiconductor."

[19] F. O'Mahony *et al.*, "A 27Gb/s forwarded-clock I/O receiver using an injection-locked LC-DCO in 45nm CMOS," *Dig. Tech. Pap. - IEEE Int. Solid-State Circuits Conf.*, vol. 51, pp. 452–454, 2008.

[20] International Roadmap Committee (IRC), "International Technology Roadmap for Semiconductortle," 2009. .

[21] K. Sohn *et al.*, "A 1.2 V 20 nm 307 GB/s HBM DRAM With At-Speed Wafer-Level IO Test Scheme and Adaptive Refresh Considering Temperature Distribution," vol. 52, no. 1, pp. 250–260, 2017.

[22] J. Song, S. Hwang, H. W. Lee, and C. Kim, "A 1-V 10-Gb/s/pin Single-Ended Transceiver With Controllable Active-Inductor-Based Driver and Adaptively Calibrated Cascaded-Equalizer for Post-LPDDR4 Interfaces," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 65, no. 1, pp. 331–342, 2018.

[23] W. Choi, T. Kim, J. Shim, H. Kim, G. Han, and Y. Chae, "A 1V 7.8mW 15.6Gb/s C-PHY transceiver using tri-level signaling for post-LPDDR4," *Dig. Tech. Pap. - IEEE Int. Solid-State Circuits Conf.*, vol. 60, pp. 402–403, 2017.

[24] H. W. Lim *et al.*, "A 5.8-Gb/s Adaptive Integrating Duobinary DFE Receiver for Multi-Drop Memory Interface," *IEEE J. Solid-State Circuits*, vol. 52, no. 6, pp. 1563–1575, 2017.

[25] T. Y. Oh *et al.*, "A 7Gb/s/pin 1 Gbit GDDR5 SDRAM with 2.5 ns bank to bank active time and no bank group restriction," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, pp. 107–118, 2011.

[26] M. Jalalifar and G. S. Byun, “An energy-efficient mobile PAM memory interface for future 3D stacked mobile DRAMs,” *Proc. - Int. Symp. Qual. Electron. Des. ISQED*, pp. 675–680, 2014.

[27] W. I. O. Dram *et al.*, “Design Challenges in 3-D SoC Stacked With a 12.8 GB/s TSV Wide I/O DRAM,” vol. 6, no. 3, pp. 364–372, 2016.

[28] D. U. Lee *et al.*, “A 1.2 v 8 Gb 8-channel 128 GB/s high-bandwidth memory (HBM) stacked DRAM with effective I/O test circuits,” *IEEE J. Solid-State Circuits*, vol. 50, no. 1, pp. 191–203, 2015.

[29] J. Kim *et al.*, “High-frequency scalable electrical model and analysis of a through silicon via (TSV),” *IEEE Trans. Components, Packag. Manuf. Technol.*, vol. 1, no. 2, pp. 181–195, 2011.

[30] K. Takahashi and M. Sekiguchi, “Through Silicon Via and 3-D Wafer/Chip Stacking Technology,” *2006 Symp. VLSI Circuits, 2006. Dig. Tech. Pap.*, vol. 00, no. c, pp. 89–92, 2006.

[31] U. Kang *et al.*, “8 Gb 3-D DDR3 DRAM using through-silicon-via technology,” *IEEE J. Solid-State Circuits*, vol. 45, no. 1, pp. 111–119, 2010.

[32] J. S. Wang, C. Y. Cheng, P. Y. Chou, and T. Y. Yang, “A Wide-Range, Low-Power, All-Digital Delay-Locked Loop with Cyclic Half-Delay-Line Architecture,” *IEEE J. Solid-State Circuits*, vol. 50, no. 11, pp. 2635–2644, 2015.

[33] H. Mafi, M. Yavari, and H. Shamsi, “Digital Calibration of DAC Unit Elements Mismatch in Pipelined ADCs,” *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 63, no. 1, pp. 34–45, 2016.

[34] B. Brannon, “Aperture uncertainty and ADC system performance,” *Appl. Note AN-501.*

Analog Devices, Inc., pp. 1–4, 2000.

[35] S. Chaganti, L. Xu, and D. Chen, "A low-cost method for separation and accurate estimation of ADC noise, aperture jitter, and clock jitter," *Proc. IEEE VLSI Test Symp.*, 2017.

[36] L. Sumanen, M. Waltari, and K. A. I. Halonen, "A 10-bit 200-MS/s CMOS parallel pipeline A/D converter," *IEEE J. Solid-State Circuits*, vol. 36, no. 7, pp. 1048–1055, 2001.

[37] P. J. Restle, K. A. Jenkins, A. Deutsch, and P. W. Cook, "Measurement and modeling of on-chip transmission line effects in a 400 MHz microprocessor," *IEEE J. Solid-State Circuits*, vol. 33, no. 4, pp. 662–665, 1998.

[38] H. E. Memristor and T. Thermistors, "Memristive devices and systems," *Proc. IEEE*, vol. 64, no. 2, pp. 209–223, 1976.

[39] S. Kvatinsky, G. Satat, N. Wald, E. G. Friedman, A. Kolodny, and U. C. Weiser, "Memristor-based material implication (IMPLY) logic: Design principles and methodologies," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 22, no. 10, pp. 2054–2066, 2014.

[40] X. Li, J. Qin, and J. B. Bernstein, "Compact modeling of MOSFET wearout mechanisms for circuit-reliability simulation," *IEEE Trans. Device Mater. Reliab.*, vol. 8, no. 1, pp. 98–121, 2008.

[41] S. Kvatinsky *et al.*, "MAGIC — Memristor-Aided Logic," vol. 61, no. 11, pp. 895–899, 2014.

[42] L. Yu, M. Miyahara, and A. Matsuzawa, "A 9-bit 1.8 GS/s 44 mW Pipelined ADC Using Linearized Open-Loop Amplifiers," *IEEE J. Solid-State Circuits*, vol. 51, no. 10, pp. 2210–2221, 2016.

[43] N. Mirzaie and G. Byun, "An Optimal Design Methodology for Yield-Improved and Low-

Power Pipelined ADC," *IEEE Trans. Semicond. Manuf.*, 2017.

[44] J. S. Kim *et al.*, "A 1.2 v 12.8 GB/s 2 Gb mobile wide-I/O DRAM with 4 × 128 I/Os using TSV Based Stacking," *IEEE J. Solid-State Circuits*, vol. 47, no. 1, pp. 107–116, 2012.

[45] A. Ascoli, S. Slesazeck, H. Mähne, R. Tetzlaff, and T. Mikolajick, "Nonlinear Dynamics of a Locally-Active Memristor," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 62, no. 4, pp. 1165–1174, 2015.

[46] P. M. Figueiredo and J. C. Vital, "Kickback Noise Reduction Techniques for CMOS Latched Comparators," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 53, no. 7, pp. 541–545, 2006.

[47] J. S. Pak *et al.*, "PDN impedance modeling and analysis of 3D TSV IC by using proposed P/G TSV array model based on separated P/G TSV and chip-PDN models," *IEEE Trans. Components, Packag. Manuf. Technol.*, vol. 1, no. 2, pp. 208–219, 2011.

[48] S. Piersanti, F. De Paulis, A. Orlandi, M. Swaminathan, and V. Ricchiuti, "Transient Analysis of TSV Equivalent Circuit Considering Nonlinear MOS Capacitance Effects," *IEEE Trans. Electromagn. Compat.*, vol. 57, no. 5, pp. 1216–1225, 2015.

[49] T. S. V Modeling and N. Coupling, "TSV Modeling and Noise Coupling in 3D IC."

[50] J. B. Lee, J. L. Aw, and M. W. Rhee, "3-D TSV Six-Die Stacking and Reliability Assessment of 20-μm-Pitch Bumps on Large-Scale Dies," *IEEE Trans. Components, Packag. Manuf. Technol.*, vol. 7, no. 1, pp. 33–38, 2017.

[51] F.C. Wu and et al., "Empricial Models for wiring Capacitances in VLSI," in *IEEE int. Sym. On Circuits and Systems*.

[52] S. C. Wong, T. G. Y. Lee, D. J. Ma, and C. J. Chao, "An empirical three-dimensional crossover capacitance model for multilevel interconnect VLSI circuits," *IEEE Trans.*

Semicond. Manuf., vol. 13, no. 2, pp. 219–227, 2000.

[53] N. Mirzaie, H. Shamsi, and G.-S. Byun, "Resilient design of current steering DACs using a transistor level approach," *Analog Integr. Circuits Signal Process.*, vol. 90, no. 1, 2017.

[54] S. Jung, J. Lee, and J. Kim, "Yield-aware pareto front extraction for discrete hierarchical optimization of analog circuits," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 33, no. 10, pp. 1437–1449, 2014.

[55] N. Mirzaie, H. Shamsi, and G.-S. Byun, "Automatic Design and Yield Enhancement of Data Converters," *J. Circuits, Syst. Comput.*, vol. 26, no. 01, p. 1750018, 2017.

[56] N. Mirzaie, H. Shamsi, and G. S. Byun, "Resilient design of current steering DACs using a transistor level approach," *Analog Integr. Circuits Signal Process.*, vol. 90, no. 1, pp. 29–41, 2017.

[57] J. Ruiz-Amaya, M. Delgado-Restituto, and Á. Rodríguez-Vázquez, "Transistor-level synthesis of pipeline analog-to-digital converters using a design-space reduction algorithm," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 58, no. 12, pp. 2816–2828, 2011.

[58] N. Mirzaie and G. S. Byun, "An Optimal Design Methodology for Yield-Improved and Low-Power Pipelined ADC," *IEEE Trans. Semicond. Manuf.*, vol. 6507, no. c, pp. 1–6, 2017.

[59] F. A. Fortin, D. Rainville, M. A. G. Gardner, M. Parizeau, and C. Gagné, "DEAP: Evolutionary Algorithms Made Easy," *J. Mach. Learn. Res.*, vol. 13, pp. 2171–2175, 2012.

[60] M. Chu, B. Kim, and B.-G. Lee, "A 10-bit 200-MS/s Zero-Crossing-Based Pipeline ADC in 0.13um CMOS Technology," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 23, no. 11, pp. 2671–2675, 2015.

[61] X. Zheng *et al.*, "A 14-bit 250 MS / s IF Sampling Pipelined ADC in 180 nm CMOS

Process," pp. 1–12, 2016.

[62] R. Sehgal, F. Van Der Goes, and K. Bult, "A 12 b 53 mW 195 MS/s Pipeline ADC with 82 dB SFDR Using Split-ADC Calibration," *IEEE J. Solid-State Circuits*, vol. 50, no. 7, pp. 1592–1603, 2015.

[63] J. Kim, "Chip-Package Hierarchical Power Distribution Network Modeling and Analysis Based on a Segmentation Method," vol. 33, no. 3, pp. 647–659, 2010.

[64] S. Choi, "Eye-diagram estimation and analysis of High-Bandwidth Memory (HBM) interposer channel with crosstalk reduction schemes on 2.5D and 3D IC," in *2016 IEEE Int. Symposium on Electromagnetic Compatibility (EMC)*, 2016, pp. 425–429.

[65] H. Lee *et al.*, "Design of Non-Contact 2Gb / s I / O Test Methods For High Bandwidth Memory (HBM)," pp. 169–172.

[66] J. F. Bulzacchelli *et al.*, "Dual-Loop System of Distributed Microregulators With High DC Accuracy , Load Response Time Below 500 ps , and 85-mV Dropout Voltage," vol. 47, no. 4, pp. 863–874, 2012.

[67] Z. Zeng, X. Ye, Z. Feng, and P. Li, "Tradeoff Analysis and Optimization of Power Delivery Networks with On-Chip Voltage Regulation," pp. 831–836, 2010.

[68] P. Zhou, D. Jiao, C. H. Kim, and S. S. Sapatnekar, "Exploration of On-Chip Switched-Capacitor DC-DC Converter for Multicore Processors Using a Distributed Power Delivery Network," vol. 2, pp. 0–3, 2011.

[69] H. He, J. J. Q. Lu, and X. Gu, "Analysis of TSV geometric parameter impact on switching noise in 3D power distribution network," *ASMC (Advanced Semicond. Manuf. Conf. Proc.*, pp. 67–72, 2014.

[70] H. He, J. J. Q. Lu, Z. Xu, and X. Gu, "TSV density impact on 3D power delivery with high

aspect ratio TSVs," *ASMC (Advanced Semicond. Manuf. Conf. Proc.*, pp. 70–74, 2013.

[71] M. B. Healy and S. K. Lim, "A novel TSV topology for many-tier 3D power-delivery networks," pp. 1–4, 2011.

[72] P. Falkenstern, Y. Xie, Y. W. Chang, and Y. Wang, "Three-dimensional integrated circuits (3D IC) floorplan and power/ground network co-synthesis," *Proc. Asia South Pacific Des. Autom. Conf. ASP-DAC*, no. 3d Ic, pp. 169–174, 2010.

[73] A. Todri, S. Kundu, P. Girard, A. Bosio, L. Dilillo, and A. Virazel, "A study of tapered 3-D TSVs for power and thermal integrity," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 21, no. 2, pp. 306–319, 2013.

[74] A. Todri-Sanial, S. Kundu, P. Girard, A. Bosio, L. Dilillo, and A. Virazel, "Globally constrained locally optimized 3-D power delivery networks," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 22, no. 10, pp. 2131–2144, 2014.

[75] H. He and J. J. Q. Lu, "Modeling and Analysis of PDN Impedance and Switching Noise in TSV-Based 3-D Integration," *IEEE Trans. Electron Devices*, vol. 62, no. 4, pp. 1241–1247, 2015.

[76] G. Huang, M. S. Bakir, A. Naeemi, and J. D. Meindl, "Power delivery for 3-D chip stacks: Physical modeling and design implication," *IEEE Trans. Components, Packag. Manuf. Technol.*, vol. 2, no. 5, pp. 852–859, 2012.

[77] H. He, X. Gu, and J. Q. Lu, "Modeling of switching noise and coupling in multiple chips of 3D TSV-based systems," *Proc. - Electron. Components Technol. Conf.*, pp. 548–553, 2014.

[78] S. Yao, X. Chen, Y. Wang, Y. Ma, Y. Xie, and H. Yang, "Efficient region-aware P/G TSV planning for 3D ICs," *Proc. - Int. Symp. Qual. Electron. Des. ISQED*, no. 612111074, pp. 171–178, 2014.

[79] S. C. K Cho, Y Kim, S Kim, H Lee, "Power distribution network (PDN) design and analysis of a single and double-sided high bandwidth memory (HBM) interposer for 2.5D Terabtye/s bandwidth system," in *2016 IEEE International Symposium on Electromagnetic Compatibility (EMC)*, 2016, pp. 96–99.

[80] X. Zhao, Y. Wan, M. Scheuermann, and S. K. Lim, "Transient modeling of TSV-wire electromigration and lifetime analysis of power distribution network for 3D ICs," *IEEE/ACM Int. Conf. Comput. Des. Dig. Tech. Pap. ICCAD*, vol. 2, no. 2, pp. 363–370, 2013.

[81] Mel Bazes, "Two Novel Fully Complementary Self-Biased CMOS Differential Amplifiers," *IEEE J. Solid State Circuits*, vol. 26, no. 2, pp. 16–21, 1991.

[82] G. S. Byun, Y. Kim, J. Kim, S. W. Tam, and M. C. F. Chang, "An energy-efficient and high-speed mobile memory I/O interface using simultaneous Bi-directional dual (base+RF)-band signaling," *IEEE J. Solid-State Circuits*, vol. 47, no. 1, pp. 117–130, 2012.

[83] C. Yoo *et al.*, "A 1.8-V 700-Mb/s/pin 512-Mb DDR-II SDRAM with on-die termination and off-chip driver calibration," *IEEE J. Solid-State Circuits*, vol. 39, no. 6, pp. 941–951, 2004.

[84] B. M. King, "Advantages of using PMOS-type low-dropout linear regulators in battery applications," *Analog Appl.*, no. February, 2000.

[85] R. J. Milliken, J. Silva-martínez, S. Member, and E. Sánchez-sinencio, "Full On-Chip CMOS Low-Dropout Voltage Regulator," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 54, no. 9, pp. 1879–1890, 2007.

[86] G. A. Rincon-Mora and P. E. Allen, "A low-voltage, low quiescent current, low drop-out regulator," *IEEE J. Solid-State Circuits*, vol. 33, no. 1, pp. 36–44, 1998.

[87] B. S. Lee, “Understanding the Terms and Definition of LDO Voltage Regulators,” 1999.

[88] D. E. Khalil, Y. Ismail, M. Khellah, T. Karnik, and V. De, “Analytical model for the propagation delay of through silicon vias,” *Proc. 9th Int. Symp. Qual. Electron. Des. ISQED 2008*, pp. 553–556, 2008.

[89] I. Savidis and E. G. Friedman, “Closed-form expressions of 3-D via resistance, inductance, and capacitance,” *IEEE Trans. Electron Devices*, vol. 56, no. 9, pp. 1873–1881, 2009.

[90] T. Bandyopadhyay, R. Chatterjee, D. Chung, M. Swaminathan, and R. Tummala, “Electrical modeling of through silicon and package vias,” *2009 IEEE Int. Conf. 3D Syst. Integr. 3DIC 2009*, 2009.

[91] C. Xu, H. Li, R. Suaya, and K. Banerjee, “Compact AC modeling and performance analysis of through-silicon vias in 3-D ICs,” *IEEE Trans. Electron Devices*, vol. 57, no. 12, pp. 3405–3417, 2010.

[92] D. H. Kim, S. Mukhopadhyay, and S. K. Lim, “TSV-aware interconnect length and power prediction for 3D stacked ICs,” *Proc. 2009 IEEE Int. Interconnect Technol. Conf. IITC 2009*, pp. 26–28, 2009.

[93] D. H. Kim and S. K. Lim, “Through-silicon-via-aware delay and power prediction model for buffered interconnects in 3D ICs,” in *IEEE Int. Interconnect Technol. Conf.*, 2010.

[94] D. H. Kim, S. Mukhopadhyay, and S. K. Lim, “Fast and accurate analytical modeling of through-silicon-via capacitive coupling,” *IEEE Trans. Components, Packag. Manuf. Technol.*, vol. 1, no. 2, pp. 168–180, 2011.

[95] A. Alzahmi, N. Mirzaie, and G. Byun, “3-D Power Delivery Network’s Subblocks and Regulator Placement Optimized by Evolutionary Algorithm,” *IEEE Trans. Components, Packag. Manuf. Technol.*, 2017.

[96] S. H. Nauman H. Khan, Syed M. Alam, “Power Delivery Design for 3-D ICs Using Different Through-Silicon Via (TSV) Technologies,” *IEEE Trans. VERY LARGE SCALE Integr. Syst.*, vol. 19, no. 4, pp. 647–658, 2011.

[97] X. Yan, Q. Wu, C. Hu, and Q. Liang, “Design Electronic Circuits Using Evolutionary Algorithms,” *J. Next Gener. Inf. Technol.*, vol. 1, no. 1, pp. 127–139, 2010.

[98] A. Alzahmi, N. Mirzaie, and G. Byun, “3-D Power Delivery Network ’ s Subblocks and Regulator Placement Optimized by,” vol. 7, no. 12, pp. 1–9, 2017.

[99] G. H. Loh, B. Black, and K. Bernstein, “Design Space Exploration for 3D Architectures,” vol. 2, no. 2, pp. 65–103, 2006.

[100] D. Stow, I. Akgun, R. Barnes, P. Gu, and Y. Xie, “Cost Analysis and Cost-Driven IP Reuse Methodology for SoC design Based on 2 . 5D / 3D Integration Invited Paper,” 2016.

[101] H. Jun, S. Nam, H. Jin, J. C. Lee, Y. J. Park, and J. J. Lee, “High-Bandwidth Memory (HBM) Test Challenges and Solutions,” *IEEE Des. Test*, vol. 34, no. 1, pp. 16–25, 2017.

[102] Y. Peng, D. Petranovic, K. Samadi, P. Kamal, Y. Du, and S. K. Lim, “Inter-die Coupling Extraction and Physical Design Optimization for Face-to-Face 3-D ICs,” *IEEE Trans. Nanotechnol.*, vol. 17, no. 4, pp. 634–644, 2018.

[103] K. Kaviani *et al.*, “A Tri-Modal 20Gbps / link Differential / DDR3 / GDDR5 Memory Interface,” pp. 234–235, 2011.

[104] S. S. T. Association, “DDR3 SDRAM Specification,” 2006.

[105] R. Bashirullah, W. Liu, and R. K. Cavin, “Current-mode signaling in deep submicrometer global interconnects,” *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 11, no. 3, pp. 406–417, 2003.

[106] W. Bae, B. J. Yoo, and D. K. Jeong, “Design of CMOS 5 Gb/s 4-PAM transceiver frontend

for low-power memory interface," *ISOCC 2012 - 2012 Int. SoC Des. Conf.*, pp. 49–52, 2012.

[107] Y. F. Hstlo, S. Gao, J. Chen, J. Ma, and T. Liu, "Design and Verification of CMOS HSTL Buffers -YF_HSTLO18," *2005 6th Int. Conf. ASIC*, pp. 620–623, 2005.

[108] M. Kumanoya, T. Ogawa, and Y. Konishi, "Trends in high-speed DRAM architectures," *IEICE Trans. Electron.*, vol. 79, no. 4, 1996.

[109] R. Cataldo, G. Korol, R. Fernandes, G. Sanchez, D. Matos, and C. Marcon, "Evaluation of emerging TSV-enabled main memories on the PARSEC benchmark," *2016 IEEE Int. Conf. Electron. Circuits Syst. ICECS 2016*, pp. 408–411, 2017.

[110] D. W. Chang *et al.*, "Reevaluating the latency claims of 3D stacked memories," *Proc. Asia South Pacific Des. Autom. Conf. ASP-DAC*, pp. 657–662, 2013.

[111] I. M. Yi *et al.*, "A 40-mV-swing single-ended transceiver for TSV with a switched-diode RX termination," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 61, no. 12, pp. 987–991, 2014.

[112] N. Mirzaie, A. Alzahmi, C.-C. Lin, and G.-S. Byun, "A performance-aware I/O interface for 3D stacked memory systems," *2017 IEEE 8th Annu. Ubiquitous Comput. Electron. Mob. Commun. Conf.*, no. October, pp. 369–371, 2017.