CRANFIELD UNIVERSITY

IFTIKHAR AHMED SOOMRO

INTEGRATING SIMULTANEOUS BIDIRECTIONAL SIGNALLING IN THE TEST FABRIC OF 3D STACKED INTEGRATED CIRCUITS

SCHOOL OF AEROSPACE, TRANSPORT, AND MANUFACTURING

Doctor of Philosophy
Academic Year: 2018 - 2021

Supervisor: Dr. Mohammad Samie

Associate Supervisor: Prof. Ian K Jennions
July 2021

CRANFIELD UNIVERSITY

SCHOOL OF AEROSPACE, TRANSPORT, AND MANUFACTURING

Doctor of Philosophy

Academic Year 2018 - 2021

IFTIKHAR AHMED SOOMRO

INTEGRATING SIMULTANEOUS BIDIRECTION SIGNALLING IN THE TEST FABRIC OF 3D STACKED INTEGRATED CIRCUITS

Supervisor: Dr. Mohammad Samie
Associate Supervisor: Prof. Ian K Jennions

July 2021

This thesis is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

© Cranfield University 2021. All rights reserved. No part of this publication may be reproduced without the written permission of the copyright owner.

ABSTRACT

The world has seen significant advancements in electronic devices' capabilities, most notably the ability to embed ultra-large-scale functionalities in lightweight, area and power-efficient devices. There has been an enormous push towards quality and reliability in consumer electronics that have become an indispensable part of human life. Consequently, the tests conducted on these devices at the final stages before these are shipped out to the customers have a very high significance in the research community. However, researchers have always struggled to find a balance between the test time (hence the test cost) and the test overheads; unfortunately, these two are inversely proportional.

On the other hand, the ever-increasing demand for more powerful and compact devices is now facing a new challenge. Historically, with the advancements in manufacturing technology, electronic devices witnessed miniaturizing at an exponential pace, as predicted by Moore's law. However, further geometric or effective 2D scaling seems complicated due to performance and power concerns with smaller technology nodes. One promising way forward is by forming 3D Stacked Integrated Circuits (SICs), in which the individual dies are stacked vertically and interconnected using Through Silicon Vias (TSVs) before being packaged as a single chip. This allows more functionality to be embedded with a reduced footprint and addresses another critical problem being observed in 2D designs: increasingly long interconnects and latency issues. However, as more and more functionality is embedded into a small area, it becomes increasingly challenging to access the internal states (to observe or control) after the device is fabricated, which is essential for testing. This access is restricted by the limited number of Chip Terminals (IC pins and the vertical Through Silicon Vias) that a chip could be fitted with, the power consumption concerns, and the chip area overheads that could be allocated for testing.

This research investigates Simultaneous Bi-Directional Signaling (SBS) for use in Test Access Mechanism (TAM) designs in 3D SICs. SBS enables chip terminals to simultaneously send and receive test vectors on a single Chip Terminal (CT), effectively doubling the per-pin efficiency, which could be

translated into additional test channels for test time reduction or Chip Terminal reduction for resource efficiency. The research shows that SBS-based test access methods have significant potential in reducing test times and/or test resources compared to traditional approaches, thereby opening up new avenues towards cost-effectiveness and reliability of future electronics.

Keywords: 3D Stacked Integrated Circuits, System on Chip, Design for Testability, Simultaneous Bi-directional Signaling, Test Access Mechanism, Reduced Pin-Count Testing, Optimization

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor Dr. Mohammad Samie for his guidance and support. He allowed me the freedom to pursue my interests and provided directions when I needed a course correction. His friendly attitude and his willingness to help with any problem is something that I will always cherish. I am also grateful to Prof. Ian Jennions, my associate supervisor, for his valuable advice and support. Be it academic concerns, personal problems, or funding for training and software, Ian always rendered all the support I could have ever asked for.

I would also like to thank Prof. Nico Avdelidis, the head of IVHM Centre, for part-funding my registration which helped me sail through the disruptions caused by Covid-19. The insightful discussions and guidance from the IVHM industrial partners and DARTEC members during regular reviews and meetings had been very vital in steering this research in the right direction. I also thank George Yazigi, the Digital Systems Manager at DARTEC, for his help in the optimization aspects.

I would also like to thank Prof. Erik Larsson of Lund University Sweden, and Dr. Adil Ansari of Quaid-e-Awam University, Pakistan, for their valuable guidance during the initial stages of my research. Given their vast experience in testability, the insights I developed from discussions with them had been instrumental in shaping my research. I would also like to thank the IEEE P1838 standard working group for accepting me as a member. Through the regular group meetings, I was able to gain a breadth of knowledge regarding 3D SICs from the most brilliant industrial and academic experts in testing.

I am ever grateful to the Government of Pakistan for funding my Ph.D. and providing me with this lifetime opportunity.

Finally, a big THANK YOU to my wife Najma, for her patience and support at home allowing me to remain focused on my work, and Ami and Daddy for their prayers.

TABLE OF CONTENTS

ABSTRACT i
ACKNOWLEDGEMENTS iii
TABLE OF CONTENTS iv
LIST OF FIGURES vii
LIST OF TABLES x
LIST OF ABBREVIATIONS xi
1 INTRODUCTION 13
1.1 Abstract 13
1.2 Background 13
1.2.1 System on Chips and 3D Integrated Circuits 13
1.2.2 Integrated Circuit Testing 16
1.2.3 Built-in Self Test vs ATE Based External testing 19
1.3 TAM design Overview and Challenges in 3D SIC Testing 21
1.3.1 Research Gaps 24
1.3.2 Potential Impact 25
1.4 Research Objectives and Methodology 26
1.4.1 Aim 26
1.4.2 Objectives 26
1.4.3 Methodology 26
1.5 Organization of this thesis: 29
1.6 List of Published/Submitted Work 30
1.7 References 31
2 DESIGN AND INTEGRATION OF A TERNARY LOGIC BASED SIMULTANEOUS BIDIRECTIONAL SIGNALLING TRANSCEIVER IN PARALLEL TEST PORTS OF 3D STACKED INTEGRATED CIRCUITS 35
2.1 Abstract 35
2.2 Background and Motivation 36
2.3 Related Work 39
2.4 Proposed Approach 41
2.4.1 Ternary Level Coding 42
2.4.2 Test and Functional mode isolation 44
2.4.3 Integration with Boundary Scan 44
2.4.4 Vertical Access Considerations in mid- and post-bond testing 46
2.4.5 Pre-bond testing 47
2.4.6 Reference Sharing 48
2.5 SBS Transceiver Circuit Design 48
2.5.1 Transmitter 49
2.5.2 Ternary Decoder 51

2.6 Simulation results.......... 52
2.7 Conclusion.......... 55
2.8 References.......... 56
3 REDUCED PIN-COUNT TEST STRATEGY FOR 3D STACKED ICS USING SIMULTANEOUS BI-DIRECTIONAL SIGNALING BASED TIME DIVISION MULTIPLEXING.......... 60
3.1 Abstract.......... 60
3.2 INTRODUCTION.......... 60
3.3 MOTIVATION AND PRIOR WORK.......... 62
3.4 BACKGROUND.......... 66
3.4.1 Time Division Multiplexing.......... 66
3.4.2 Simultaneous Bi-directional Signaling.......... 69
3.5 METHODOLOGY.......... 72
3.5.1 Transceiver Design.......... 73
3.5.2 Pre- and Mid-Bond Testing.......... 76
3.5.3 Test Setup.......... 77
3.6 RESULTS AND DISCUSSION.......... 79
3.6.1 Power Consumption.......... 82
3.6.2 Signal Integrity under cross-coupling.......... 84
3.6.3 Comparison with relevant prior work.......... 85
3.7 CONCLUSION.......... 86
3.8 References.......... 87
4 OPTIMIZATION AND TEST TIME ANALYSIS OF SIMULTANEOUS BI-DIRECTIONAL SIGNALING BASED TEST ACCESS MECHANISM FOR 3D STACKED INTEGRATED CIRCUITS.......... 92
4.1 Abstract.......... 92
4.2 Introduction and Related Work.......... 92
4.3 Problem Formulation.......... 96
4.4 Proposed ILP Formulation.......... 100
4.4.1 Forming Test Channels between Tester and Dies.......... 101
4.4.2 Mapping Test Channel to Pins/ TSVs:.......... 102
4.4.3 Formulation of the Objective Function.......... 104
4.4.4 The Overall Linearized UDS/SBS TAM Optimization Model.......... 106
4.5 Results and Discussion.......... 108
4.6 Conclusion.......... 114
4.7 Appendix 4-I: Linearization of constraints in section 4.4.2.......... 115
4.8 References:.......... 117
5 AN INTEGRATED APPROACH FOR 3D SIC TESTING USING SIMULTANEOUS BIDIRECTIONAL AND CONVENTIONAL UNI-DIRECTIONAL SIGNALLING.......... 120
5.1 Abstract.......... 120

5.2 Introduction and Prior Art 120
5.3 ILP Formulation for the SBS-UDS Co-Design Methodology 124
5.3.1 Objective function formulation and linearization 128
5.3.2 The Overall ILP Formulation 130
5.4 Results and Discussion 131
5.4.1 SBS-UDS Co-design Analysis ε = 1 (TAT only) vs ε = 0.5 133
5.4.2 SIC Design and die position implications 139
5.4.3 Effects of ε variations on the optimal solution 141
5.5 Conclusion 143
5.6 Appendix 5-I: Calculation of Bounds on ε for multi-Objective Optimization 144
5.7 References 145
6 CONCLUSION AND FUTURE DIRECTIONS 148
6.1 Addressing the Aim and Objectives of the Research 148
6.2 Contributions to the existing body of knowledge 150
6.3 Limitations and Future Work 151
6.3.1 Transceiver Design 151
6.3.2 TSV Testing 152
6.3.3 Optimization 153
6.4 References 154
APPENDICES 156
Appendix A : Optimization Programming Language (OPL) based CPLEX codes 156

LIST OF FIGURES

Figure 1-1 An illustration of a System on Chip (SoC) 14

Figure 1-2: Stacked Integrated Circuits (a) 2.5D SIC (b) 3D SIC (c) 5.5D SIC [illustration adapted from [44]] 15

Figure 1-3 (a) A NOR gate (b) CMOS equivalent of a NOR gate [12] 18

Figure 1-4: Built in Self-Test (BIST) and External testing based test methodologies (a) An illustration of a Built in Self-Test (BIST) based test requirements (b) Scan-based external testing illustrating the basic el-ements required for SoC Testing: A Source, Sink, C 20

Figure 1-5: Problem Overview - TAM Design, Scheduling and Optimization for core-based ICs. 21

Figure 1-6: Research Methodology showing various phases of the research, the problem addressed during each phase and the relevant chapter of this thesis 27

Figure 2-1: A single test channel using: (a) Conventional signaling – uses two wires (b) SBS – uses one wire (c) A combination of Uni- and SBS – one wire between tester and Die 1, two between Die 1 and Die 2 37

Figure 2-2: Conventional vs SBS test ports (a) TAM design for uni-directional port and associated test schedule in (b), (c) TAM design for SBS port and associated test schedule in (d) 38

Figure 2-3: Test Channel for a 2 bit Scan-Chain using Ternary Encoding and Decoding at the Chip Terminal 42

Figure 2-4: Test and Functional mode isolation (in this illustration, the chip terminal is a functional output).. 43

Figure 2-5: Boundary scan compliant fully observable and controllable boundary scan cell [6] shown in the dashed box (a) without SBS (b) SBS integration outside functional path (c) SBS integration inside functional path 45

Figure 2-6: An illustration of SBS for accessing higher dies in a 3D SIC through TSVs.. 47

Figure 2-7: Reference Sharing 49

Figure 2-8: (a) Proposed SBS Transceiver Circuit (b) equivalent circuit for Vxm (c) equivalent circuit for Vxh (d) equivalent circuit for Vxl 50

Figure 2-9: Simulation results of the various signals in the proposed SBS Transceiver Circuit in Figure 2-8. (Verti-cal scale normalized to Vdd). 53

Figure 2-10: TSV Cross-coupling (a) Victim TSV (center) and 8 aggressor TSVs in a 3x3 Cluster (b) TSV-TSV Coupling model [41] (c) Histogram of Vx voltage levels at the receiver sampling time, under cross-coupling. 55

Figure 3-1: An example implementation of TDM for scan-test application based on [28].. 67

Figure 3-2: Control circuit for TDM (a) Generation of core and die clocks from global test clock using Ring Counters, the bottom RC generates die clocks, the top RC generates core clocks (2) Timing diagram of the die and core clocks.. 68

Figure 3-3: An illustration of Simultaneous Bi-directional Signaling (a) Conventional Uni-directional signaling using two wires (b) Simultaneous Bi-directional Signaling using one wire (c) Block diagram of SBS working principle – Test channel formation for a two-bit scan-chain....................... 70

Figure 3-4: Neural Network presentation of the ternary decoder...................... 71

Figure 3-5: Hyperplanes created by the neurons in Figure 3-4. 71

Figure 3-6: SBS-TDM based Test Access Mechanism for the example case of Figure 3-1. 72

Figure 3-7: SBS transceiver implementation (a) Transmitter (b) Equivalent electrical model of the transmitter when sending opposite signals (10,01) (c) Sense-Amplifier based Ternary Decoder (Receiver) 73

Figure 3-8: (a) Bypass method for Pre-bond testing. (b) Using SBS transceivers as buffers to access higher dies. 76

Figure 3-9: Pseudo-code for generation of 3D TDM based TAM output for a 1-bit global channel assuming identical dies in the stack 78

Figure 3-10: Waveforms of various signals using SBS transceiver design of Figure 3-7 in the test setup of Figure 3-6 simulated at 1.2 GHz Gclk frequency.. 80

Figure 3-11: Transceiver transient response with varying frequency. 81

Figure 3-12: Power consumption of a single channel for UDS (Figure 3-1) and SBS (Figure 3-6) transceiver based TDM schemes. 82

Figure 3-13: Eye diagram for the Vx signal at near end receiver side at 1.2GHz under TSV Cross coupling (all process corners). 82

Figure 4-1: The core wrapper design problem (a) An example core containing three scan chains of length, 10,9 and 5 and 5 I/Os (b) Serial concatenation of all scan elements for 1 bit TAM (c) Assignment of scan element to multi-bit TAM (3 in this case)... 96

Figure 4-2: TAM width assignment to cores [20] (a) Serial concatenation for sequential testing using full TAM width (b) Parallel testing using distributed TAM width.. 98

Figure 4-3: % improvement in TAT of SIC 1 with varying Pins and TSV limits111

Figure 4-4: Variation of the die test time with increasing number of test channels 111

Figure 4-5: % improvement in TAT of SIC 2 with varying Pins and TSV limits 112

Figure 4-6: % improvement in TAT of SIC 3 with varying Pins and TSV limits 112

Figure 4-7: Comparison of TAT and the percent TAT difference ∆TAT (%) using Conventional and SBS based TAMs with increasing number of dies in SIC 4. Global TSV limit has been fixed to 100 and Pins equal to (a) 50 (b)100 (c) 150 and (d) 200 113

Figure 5-1: TAT, Test Channels used and Chip Terminal design choice for SIC1 ($\boldsymbol{Pmax}$ =90, $\boldsymbol{TSVmax}$=190) 135

Figure 5-2: Reference CT impact on TAM Design (SIC1, $\boldsymbol{Pmax}$ =90, $\boldsymbol{TSVmax}$=190) (a) Chip Terminals designed as SBS with/ without reference CTs (red and green curves) and the resultant difference in TAT (dashed curve) (b) Effect of reference CTs on design choice 136

Figure 5-3: Comparison of SBS-UDS Co-design method against UDS only and SBS only TAMs (SIC1, $\boldsymbol{Pmax = 90}$, $\boldsymbol{TSVmax = 190}$) (a) Test Time (TAT) (b) Chip Terminal (CT) utilization 137

Figure 5-4: Effects of SIC construction on the TAM design using SBS-UDS codesign methodology for SIC 1 to 3 with easing cost constraints (a) TAT reduction (b) Test Channels formed (c) CTs designed as SBS or UDS (d) Pins and TSVs used 140

Figure 5-5: Multi Objective TAT and CT Minimization - Various pareto optimal solutions for different ε ($\boldsymbol{Pmax = 100}$, $\boldsymbol{TSVmax = 300}$) 142

LIST OF TABLES

Table 2-I: Mode Configuration States for SBS Integration with BSR in Figure 2-5(b) 43

Table 2-II: Average Power Consumption (in μW) 54

Table 3-I: SBS Transceiver States 71

Table 3-II: Binary-weighted transmitter performance 83

Table 3-III: Comparison with relevant works 86

Table 4-I: 3D SICs composition 108

Table 4-II: Test Time Comparison for SIC 1 using Uni- and Simultaneous Bi-directional Signaling 109

Table 4-III: Die level TAM for SIC 1 with 40 pins and 200 TSVs using SBS ... 110

Table 5-I: Chip Terminal utilization using ε = 1 and ε = 0.5 (SIC1, P_{max}= 10, TSV_{max}=30) 132

Table 5-II: Optimal solutions for SIC 1 using SBS-UDS Co-design (for SIC1, P_{max}= 10, TSV_{max}=30) 132

Table 5-III: ε bounds for SIC 1 (M = 5) 142

LIST OF ABBREVIATIONS

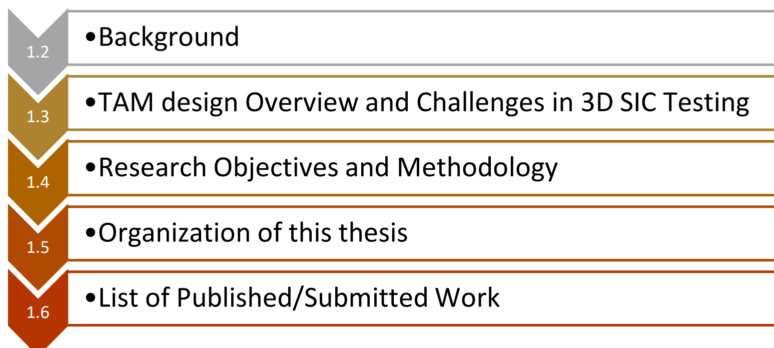
ASIC	Application Specific Integrated Circuit
BSR	Boundary Scan Register
CAD	Computer Aided Design
DfT	Design for Testability
DSI	Decoded Scan In
FPCT	Full Pin Count Testing
FPGA	Field Programmable Gate Array
FPP	Flexible Parallel Ports
HDL	Hardware Description Language
IC	Integrated Circuit
IP	Intellectual Property
IR	Instruction Register
JTAG	Joint Test Action Group
MVL	Multi Valued Logic
PCB	Printed Circuit Board
PRBS	Pseudo-Random Binary Sequence
PTAM	Parallel Test Access Mechanism
RPCT	Reduced Pin Count Testing
RTL	Register Transfer Level
Rx	Receiver
SBS	Simultaneous Bi-Directional Signalling
SBS TR	Simultaneous Bi-directional Signalling Transceiver
SC	Scan Chain
SerDes	Serializer De-Serializer
SI	Scan In
SIC	Stacked Integrated Circuit

SO	Scan Out
SoC	System on Chip
STAM	Serial Test Access Mechanism
STIL	Standard Test Interface Language
TAM	Test Access Mechanism
TAP	Test Access Port
TAT	Test Application Time
TB	Tri-Stated Buffer
TD	Ternary Decoder
TDI	Test Data In
TDM	Time Division Multiplexing
TDO	Test Data Out
TE	Test Enable
TG	Transmission Gate
TSV	Through Silicon Via
Tx	Transmitter
UDS	Uni-Directional Signalling

1 INTRODUCTION

1.1 Abstract

This Chapter is aimed to provide an overview of this research and the structure of the thesis. Section 1.2 begins with a background covering the basics of modular chip design philosophy, known as System on Chips (SoCs) and the 3D Stacked Integrated Circuits (SICs), followed by an overview of the fundamental principles of the Integrated Circuit (IC) test paradigm. In section 1.3, the focus is narrowed down to the particular area of chip testing being focused on in this research: the design of test accessibility in 3D SICs. The challenges in this area have been highlighted and the approaches being proposed to mitigate these challenges are discussed. Section 1.4 describes the research objectives and the methodological approach adopted to address these objectives. Section 1.5 provides a brief overview of the subsequent chapters, their scope and relevance to the aim and objectives. Section 1.6 lists the published and under-publication literary works relevant to this research.



1.2 Background

This section will cover the Design for Testability (DfT) basics for digital Integrated Circuits (ICs). It begins with describing the 2D System on Chips (SoCs) and the 3D Stacked Integrated Circuits (SICs), followed by the principles of electronic testing and its classification. The importance of testing and the associated problems in the manufacturing cycle are highlighted.

1.2.1 System on Chips and 3D Integrated Circuits

Traditionally, an Application Specific Integrated Circuit (ASIC) manufacturer would design the complete logic of the chip, which would then be forwarded to the foundry for manufacturing. With the increasing complexity of electronics, this approach is no longer

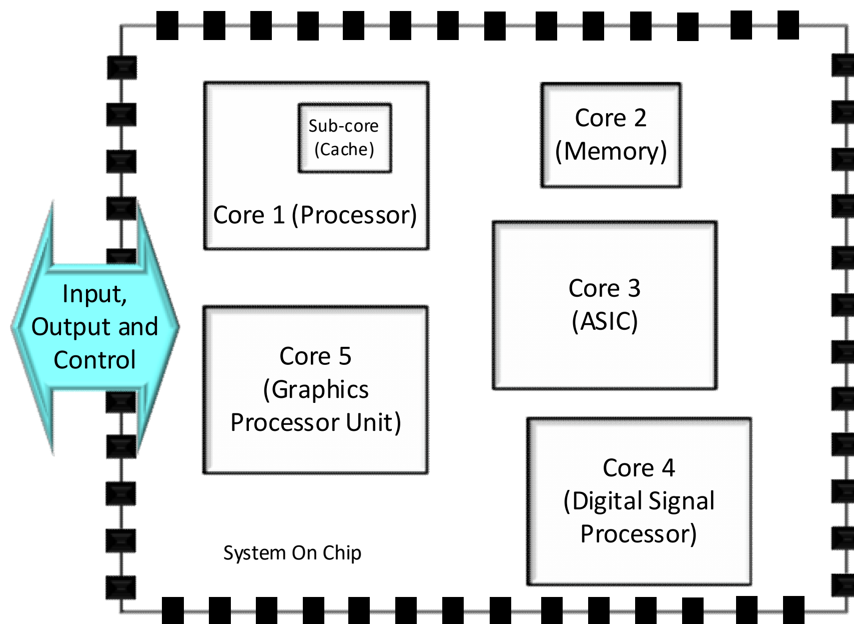


Figure 1-1 An illustration of a System on Chip (SoC)

feasible, as in many cases, the chip requirements are too complex to be designed by a single chip manufacturer. The semiconductor industry has thus moved towards a design-reuse philosophy, such that an ASIC manufacturer can simply re-use the pre-designed and optimised logic blocks (called Intellectual Property (IP) Cores or simply Cores), acquired from different vendors to produce an end product called a System on Chip (SoC). An illustration of an SoC consisting of various cores is shown in Fig. 1-1. The IP cores can be designed at various levels of abstraction such as Register Transfer Level (RTL) called Soft Cores, technology-dependent gate-level Netlist called Firm Cores or a physical layout description (Such as GDSII format) for a particular process technology which is called a Hard Core. The different levels of abstraction dictate the ability of an SoC manufacturer to modify the design of the IP Core, the highest flexibility being in Soft Core, followed by firm core, and the least flexible is the Hard Core.

Although this design-reuse philosophy facilitates the integration of cores to form complex designs, testing becomes a problem as it is now a shared responsibility between the core designer and the SoC manufacturer. The core designer needs to lay down the testability requirements of the individual core. In contrast, the SoC manufacturer, who is ultimately responsible for the end-product reliability, is required to design and insert test logic fulfilling test requirements of all the individual cores. In the case of the soft cores, the SoC manufacturer will have the liberty to design the test wrapper and insert a test mechanism that best fits the overall SoC design. However, the test wrapper may already be fixed for the firm or hard cores, resulting in a sub-optimal

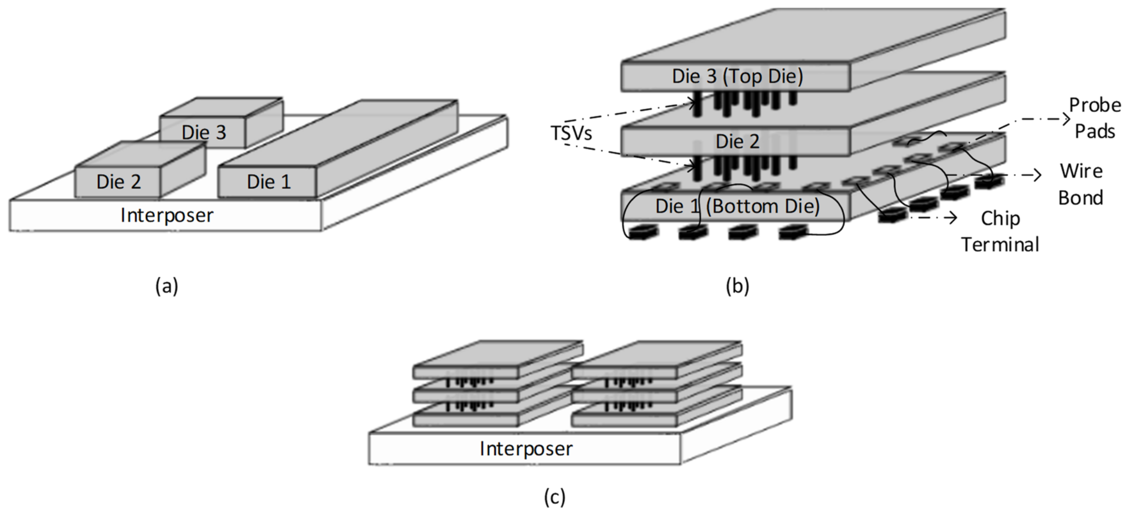


Figure 1-2: Stacked Integrated Circuits (a) 2.5D SIC (b) 3D SIC (c) 5.5D SIC [illustration adapted from [44]]

or inefficient testing solution. The choice of Core type is thus a trade-off between IP protection and facilitation of testing.

Technology miniaturization is becoming increasingly difficult owing to power, thermal dissipation and noise concerns below 6nm transistor technology nodes [1]. As a result, further scaling in the current 2D SoC designs may not readily be available. On the other hand, increased use of consumer electronics demand more functionality, performance, reduced power consumption and chip form-factor. Therefore, one viable option seems to be moving in the third dimension and producing 3-dimensional chip designs.

The most common method of producing 3D ICs is by stacking the individual 2D dies on top of each other, forming Stacked Integrated Circuits (SICs) [2]. This allows to embed more functionality with a reduced footprint and address other critical problems observed in 2D ICs, that of increasingly long interconnects and latency issues. This concept is already beginning to gain attention in processor design [3], FPGAs [4], Network on Chips (NoCs) [5], and chips have already been manufactured by leveraging the third dimension, such as DDR 3 memory ICs [6].

There are different ways in which SICs could be manufactured, as shown in Fig. 1-2 [7]. The Dies could be stacked horizontally on top of an Interposer, as shown in Fig. 2(a), in which case it is commonly referred to as 2.5D SIC. An interposer is simply a glass or silicon-based substrate containing electrical interconnect wiring. Another way is to stack dies on top of each other, forming a tower, with chip terminals wire-bonded to either top

or bottom die and the intermediate dies interconnected by Through Silicon Vias (TSVs). An SIC in such an arrangement is called 3D SIC and is shown in Fig. 1-2(b). Yet another possible arrangement is to combine 2.5D and 3D methods such that multiple 3D towers are interconnected horizontally on an Interposer. This arrangement is commonly referred to as 5.5D SIC as in Fig. 1-2(c). This thesis is primarily aimed at addressing 3D SICs as they appear to be the most prominent way in which SICs are to be manufactured.

3D SICs have the following benefits over 2D chips [2]:

- Increased transistor density per unit area, reduced form factor and chip area optimization.
- Reduction in the overall length of wire interconnections, resulting in reduced signal losses and improved latency.
- Reduction in PCB copper traces, as the dies are placed vertically, instead of chips being placed horizontally on a Printed Circuit Board (PCB).
- Allows integration of heterogeneous technology dies.

Despite the potential benefits of 3D integrations, the technology is yet in its infancy and requires innovative solutions to several challenges [8][9]. One of the most critical aspects being reliability and the design for testability which brings several new challenges. However, before discussing these test challenges (later in section 1.3), the basics of electronic testing are described in the following section, giving as much detail as necessary to provide the background for this thesis. A reader familiar with traditional DfT methodology in modular designs may skip to section 1.3.

1.2.2 Integrated Circuit Testing

Defects or bugs could be introduced inadvertently at various stages of the IC development process, including design, fabrication, and packaging. These could be due to human inaccuracies in the behavioural description of the circuit, errors introduced during compilation, synthesis, and optimization processes by the Computer-Aided Design (CAD) tools or imperfections of the manufacturing process. It is estimated that the cost of repairing a fault increases ten-fold as we transition from one stage to the following [10]. Therefore, testing at every stage is the key to reliability and cost reduction [11].

Based on the development stage, testing can be broadly classified as Pre- silicon or Post Silicon testing. As evident from the name, pre-silicon testing is done at the design stage of an IC in a computer environment (simulation or emulation) for design verification and identification of logic design errors. The design process starts with a behavioural model of the circuit in any language format such as C, System C or HDL, which describes the basic functionality of the IC. Once the functional design is finalized, it is converted to a lower-level design abstraction such as Register Transfer Level (RTL), which is then used to formulate the gate-level description (netlist) targeting the specific CMOS technology node. Finally, the physical layout of the silicon cores is synthesized. The transition from a higher to lower-level abstraction is usually done using CAD tools which may inadvertently introduce bugs during each synthesis step due to inherent limitations in CAD automation. These bugs are primarily identified, diagnosed, and debugged using test benches in a software simulation environment. While testing is the easiest at this stage as all the internal states of the circuit are both observable and controllable, pre-silicon testing suffers a severe drawback in that it is extremely slow due to the inherent limitations of the software environment.

Once silicon is fabricated, validation tests are conducted to ensure the silicon functions as per the desired high-level design. Pre-silicon tests can detect functional bugs; however, the manufacturing process induced faults such as opens, shorts, weak transistors, performance variation with temperature and EMI can only be tested after fabrication, i.e. Post-silicon. This makes post-silicon validation a crucial aspect of IC testing. It also enables verification using actual stimuli on a physical device at a clock speed which is orders of magnitude faster than the bandwidth-limited software environment in the pre-silicon phase.

Post-silicon testing can be undertaken using two methods depending upon the methodology used for test pattern generation, known as functional or structural testing. As the name suggests, the test patterns are generated using the knowledge of the chip’s functionality in functional testing. This has an advantage that the test bench generated at the pre-silicon stage can be re-used for post-silicon testing at the clock speed, which is orders of magnitude faster than simulation platforms. However, the manual generation of patterns becomes increasing difficult for complex systems, and the test patterns also increase exponentially. Consider an example of a NOR gate required to be tested as shown in Fig. 1-3 (example and the figure have been adapted from [12]). The output of

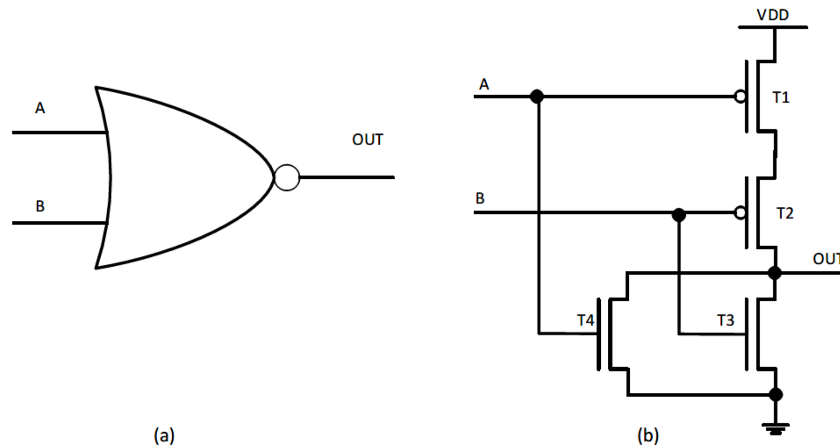


Figure 1-3 (a) A NOR gate (b) CMOS equivalent of a NOR gate [12]

a NOR gate is logic high when both A and B are zero and low for all other inputs. There are 4 possible input combinations of A and B, and one might think that if the logic works for one input, it should work for all other combinations as well? However, merely testing one or a subset of inputs combinations may not fully test the circuit. For example, consider the CMOS representation of the same NOR gate in Fig. 1-3(b), if the output node (OUT) is grounded due to any manufacturing flaw, the OUT node retains the capability to go to logic 0 (ground) but cannot be driven to logic 1 (VDD). Therefore, the only test pattern that could detect this fault is when both A and B are low. Additionally, some defects may manifest as transient faults, such as slow or stuck-open transistors. These faults are transition dependent and will therefore require a sequence of patterns, thus requiring 2^{2n} test patterns, where n is the number of inputs/states. In the example of Fig. 1-3, if n=2 , only 8 test patterns would be required, however for an 8 input gate, the pattern count increases to 65,536. Clearly, the exponential increase in test patterns for exhaustive functional testing is not scalable for larger circuits.

On the other hand, structural testing relies on verifying the operation of smaller structures within the whole design, such as gates and transistors and does not require the knowledge of the overall functionality. This is achieved by using fault models that describe the relationship between a fabrication defect and its effect on the circuit. Fault models can be generated at various levels of abstraction, such as the transistor level (stuck-open or close) or at the logic gate level (stuck-at, bridging model, Transition model, Gate Delay or Path delay models). The most commonly used model is the Single Stuck-at (SSA) fault model, in which it is assumed that a defect would manifest itself by

causing a node to be either stuck at zero (SA0) or 1 (SA1). If there are n nodes, there are 2n single stuck-at faults possible in the circuit, therefore unlike exponentially increasing test patterns in the functional testing, using the SSA model, the test pattern and hence the time increases linearly with the number of gates in the circuit.

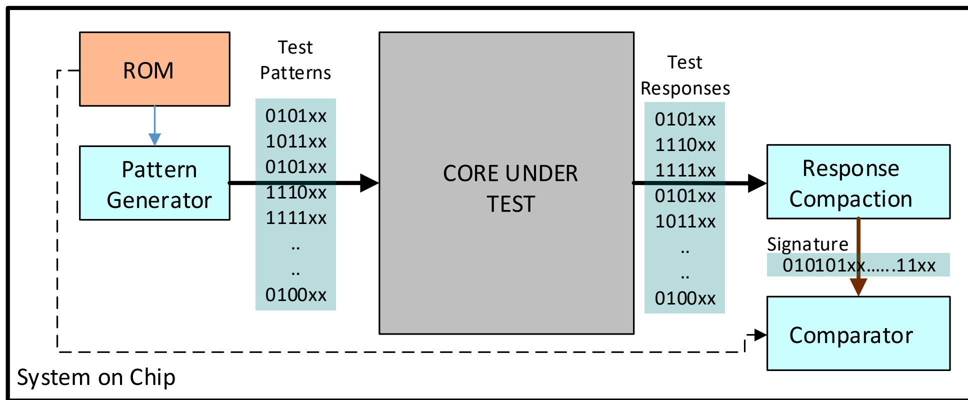
Structural testing immensely reduces the complexity of pattern generation as the fault models can be easily programmed to generate the test patterns, also known as Automatic Test Pattern Generation (ATPG), and the fault coverage can be quantified for a given fault model as:

$$Fault\ Coverage\ (FC) = \frac{Faults\ Detected}{Total\ Faults}.100\ (\%) \qquad \textbf{Eq 1-1}$$

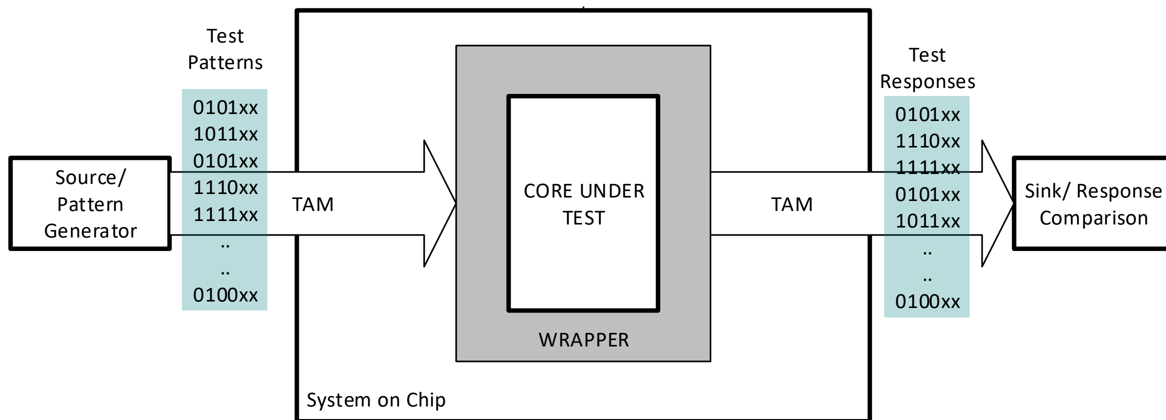
Increasing the fault coverage from 98% to 99.5% can double the required test patterns, significantly increasing test time and costs [11]. Moreover, testing at gate or transistor level requires controllability and observability at these levels needed to be built into the design specifically for test purposes, an approach commonly known as Design for Testability (DfT). The most widely used technique is to design access to the memory element within the design (usually the flip-flops), which are modified such that these are observable and controllable in test mode and are termed as scan-flops. The scan-flops are then concatenated into serial shift-registers, known as scan chains, such that the test data could be sequentially scanned in and out using the core's primary inputs and outputs. The scan-based DfT introduced hardware overheads, but the advantages of ATPG and pattern count reduction outweigh these limitations and has been widely used for chip testing [12].

1.2.3 Built-in Self Test vs ATE Based External testing

The test pattern generation is undertaken at the design stage, usually by the core manufacturer. The test pattern information is then translated into a human and machine-readable format using Standard Test Interface Language (STIL) [13]. The application of these test vectors requires a Test source and a sink, which could reside internally on the chip or external to the chip using Automatic Test Equipment (ATE), known as Built-in Self-Test (BIST) or External test respectively, as shown in Fig 1-4. BIST has an advantage in that the test patterns can be generated on-chip and applied at-speed,



(a)



(b)

Figure 1-4: Built in Self-Test (BIST) and External testing based test methodologies (a) An illustration of a Built in Self-Test (BIST) based test requirements (b) Scan-based external testing illustrating the basic el-ements required for SoC Testing: A Source, Sink, C

making it the fastest test application method. However, this comes at a price of significant overheads in the chip area occupied by the pattern generation and response comparison circuits. To reduce area overheads, the pattern storage memory block (ROM) can be replaced by random patterns generation circuits such as Linear Feedback Shift Registers (LFSRs); however, this does not guarantee the generation of all the required patterns, and therefore, the test quality may be significantly reduced. External testing using ATEs have the advantage of minimizing chip area overheads compared to BIST and the ability to undertake higher quality test using the most appropriate test patterns which can be stored on the external tester. ATE based external testing is, therefore, the most preferred method for scan-based testing. However, for external testing, a Test Access Mechanism (TAM) is required to transport the test to and from the tester, which resides external to the chip.

To summarise the discussion until this point, it is evident that scan-based (structural) testing is a highly scalable and reliable method for chip testing. It simplifies the pattern generation, allows defect modelling for modular testing, and significantly reduces pattern count compared to functional testing. However, the fine-grained observability and controllability require access to millions of internal flip flops through the chip terminals, which are very limited, necessitating serial structures (scan-chains) in which the data could be shifted one bit at a time for every test pattern. For external testing, the design of this accessibility is a highly complicated process due to the involvement of several design variables, while it is equally important as it significantly affects the test quality, time, and cost. The TAM design process and the challenges and implications on 3D SIC testing are discussed in the following section.

1.3 TAM design Overview and Challenges in 3D SIC Testing

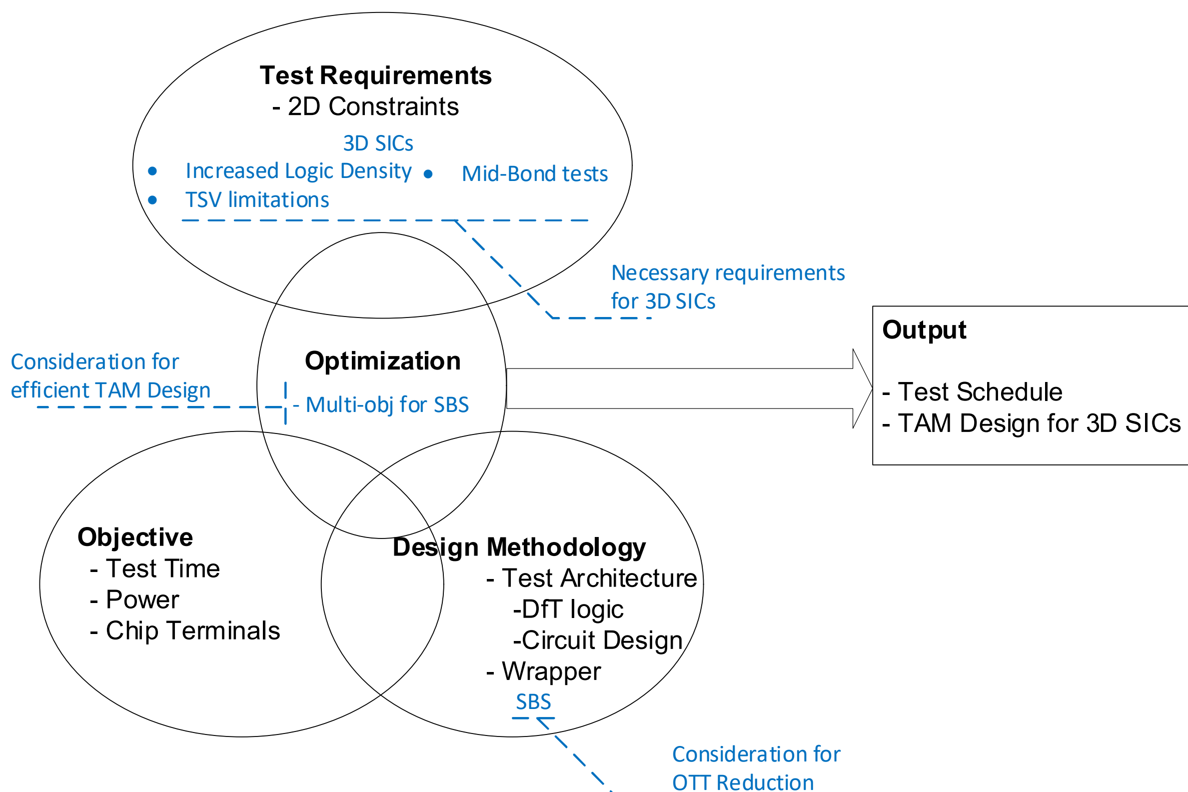


Figure 1-5: Problem Overview - TAM Design, Scheduling and Optimization for core-based ICs.

The fundamental purpose of the Test Access Mechanism (TAM) is to transfer Gigabits of test data to millions of scan flops to all the cores of an SoC. In an ideal situation, all

the cores in a die and all the scan chains in the cores would be tested simultaneously, in parallel, which would result in the minimum Test Application Time (TAT). Here, TAT is defined as the number of clock cycles required to apply test patterns to test the SoC or a 3D SIC. However, several constraints limit this approach, such as:

1) The Input/ Output chip terminal (Pins) used for communication with the outside world (including tester) are large structures, available only in limited quantities. Therefore, it may not be possible to access all cores and, in turn, all the scan chains at once.
2) There may be power and thermal dissipation limitations, which restrict the number of cores being tested simultaneously and the maximum test frequency.
3) Chip area and routing resources that could be dedicated for the test architecture may be limited.

Clearly, reduction of TAT under these restrictions become challenging. The task of the TAM design is to maximize the productivity of the allocated resources, which has been a focus of significant research in the past [14]. Several TAM design methodologies have been proposed in the literature which can broadly be classified as serial TAMs (based on JTAG,IJTAG) [15][16][17], packet data based TAMs (NoCs) [18][19][20][21], Parallel TAMs [22][23][24][25] or Reduced Pin Count Testing (RPCT) based TAMs [16][26][27][28][29]. Different TAM design methodology presents different trade-offs between TAT and resources. Briefly, Serial TAMs are resource-efficient but do not scale well with test complexity. Packet-data based TAMs require area and power-hungry routers and controllers, which cannot be inserted solely for test purposes and therefore only finds their application in NoC based devices. Parallel (or bus-based) TAMs and their variant RPCT based TAMs offer a balance between channels width and flexibility and are widely used methods. The detailed discussion into these TAM design methodologies follows in Chapters 2 and 3.

For a given chip design (test complexity and available resources) and a TAM design methodology, many possibilities would exist regarding the distribution of the test resources (primarily the test channel width) among all cores, and in turn, among all scan chains. The resources can be distributed temporally by testing a subset of cores in a given test session (known as *‘scheduling’*) and spatially within each session (commonly

referred to as *'design'*). Every design and schedule affects the test application time making this an optimization problem, as shown in Fig. 1-5. For a given chip design and TAM design methodology, the task of optimization is to provide the best (optimal) TAM design (i.e., channel width allocation) and schedule (the set of dies to be tested simultaneously in a test session) while keeping within the design and test constraints of the SoC. This problem is known to be NP-Hard, and a detailed description and review of the relevant literature are covered in Chapter 4. Apart from test time reduction, the optimization problem may also be formulated using other or (additional) objectives such as resource use minimization, in which case the solution optimal in test time may not be optimal for the revised objective. This aspect is further discussed in detail in Chapter 5.

3D SICs stacking brings about several additional and more complex challenges for the TAM design and Optimization problem [30][9][31][32][33] , such as:

a. The manufacturing process of 3D SICs introduce additional defects and necessitates multiple test instances [34]. In a 2D SoC, the chip is first tested at the wafer level by using microscopic probes to gain access to test pads, and later the final tests are conducted through standard chip terminals after it has been packaged. In 3D SICs, apart from the wafer and chip-level testing (known as **pre-bond** and **post-bond** testing, respectively), the stack has to be tested at every point during the stacking process. For example, an additional test has to be performed for a stack of 3 dies when Die 1 and 2 are stacked, followed by another test when Die 3 is added. These additional testing steps are known as **mid-bond testing** [35].

b. Higher transistor density in 3D SICs increases the probability of manufacturing defects, thus requiring a higher number of test pattern count but using a limited number of pins and inter-die connections (Through Silicon Vias (TSVs)) [36] [26].

c. TSVs are open-ended at this point (post-thinning), and testing un-bonded TSVs is a research area by itself [37].

The increased defect density leading to increased pattern count, the addition of several test instances, and the test access bottleneck caused by a limited quantity of the large TSV structures significantly exacerbate the TAM design problem [5], resulting in spiking

TAT and resources. Therefore, 3D testing has been a focus of significant research, with various approaches for design and optimization being reported.

While TSVs occupy a significant chip area, on the brighter side, they are capable of significantly high speeds due to lower channel resistance, providing opportunities for investigation of new avenues for efficient TAM designs. In particular, the existing TAM design methods have been based on an underlying assumption to use a Chip Terminal (CT) in a Uni-directional (simplex or half-duplex) fashion. This way, the existing functional CTs, which are mostly UDS based, can be re-used for test purposes without significant modifications. This greatly improves design effort and also does not incur additional area or power overheads. However, this method requires 2 x CTs to form a transceiver channel, one for the scan-in signal while the other for the scan-out signal. On the contrary, Simultaneous Bi-directional Signalling (SBS) based CT designs allow the utilization of a single CT to form a channel. The advantages are twofold; first it may be utilised in Parallel Test Ports based TAM (PTAM) designs to double the available TAM width for testing, or it can be employed in RPCT methods to further minimize the required CTs to form a TAM. However, the including of SBS has the following implications:

a. SBS requires ternary coding, therefore a transceiver circuit capable if SBS is to be included, which adds to the design effort.
b. SBS adds area and power overheads, which need to be accounted for.
c. In case SBS is being designed as an add-on to existing functional logic for use in test mode only, then the impact of SBS on the functional logic must be minimised.
d. The inclusion of SBS should not exclude the possibility of using standard UDS DfT logic such as JTAG BSRs, such that compatibility issues are avoided.

1.3.1 Research Gaps

Based on the above discussion, this research aims to explore the feasibility of using SBS in test mode for 3D SICs, with the primary aim of test time reduction and resource efficiency. Particularly, the following gaps in the existing research for 3D SICs testing have been identified:

- There is a need for research in SBS based communication methods to ensure full-duplex utilization of TSVs in testing.

- SBS Transceiver designs suitable for use in Parallel TAMs at low frequencies and in FPCT based TAMs at high frequency need to be studied.
- TAM design and optimization methodologies for SBS based methods need to be established.
- Re-usability of the SBS transceivers for employment in open-ended TSV testing may be studied. (This gap is not addressed in this thesis and has been earmarked for future work in Chapter 6)

The novelties of this work are highlighted in blue in Fig. 1-5. Firstly, the TAM design and scheduling framework for 3D SICs is investigated being an ongoing area of research. Secondly, a novel method of TAM design is proposed leveraging SBS for use in PTAMs, which to the best of the author's knowledge, have not been studied before. Finally, an optimization framework capable of TAM Design and Scheduling for SBS based TAMs or co-designs where TAM is a combination of conventional and SBS is proposed.

1.3.2 Potential Impact

- **Academic impacts:** This research will contribute to the existing literature by proposing design methods of TAM in 3D SICs using SBS and highlight different aspects and design considerations to achieve the same. This work is expected to trigger further research in different aspects of testing previously focused solely on conventional Uni-directional Signalling based methods, such as testing the TSVs and optimization algorithms. Furthermore, enhanced SBS transceiver circuit designs could also be explored.

- **Industrial impacts:** Reduction in test time will reduce product cost and will help to bring down the cost for consumer electronics. A resource-efficient solution to test 3D chips will help the semiconductor industry to realize 3D technology. Comparison with other TAM design techniques to quantify benefits vs costs will enable chip designers or researchers to have a fair idea of whether the proposed method is suitable for their particular chip design.

In the next section, the objectives of this research and the methodology adopted to achieve these are outlined.

1.4 Research Objectives and Methodology

1.4.1 Aim

Design and optimization of a resource-efficient Test Access Mechanism (TAM) in 3D digital electronics to reduce overall test-access time.

1.4.2 Objectives

Literature Review: Undertake a literature review to identify research gaps and propose a TAM design strategy to address the gaps identified.

Design: Study SBS design solutions for use in basic and advanced Test Access Mechanisms - Design SBS transceivers suitable for use in 3D IC test architectures.

Optimize: Formulate TAM optimization methodology for 3D Stacked Integrated circuits using SBS only and SBS+UDS co-design

Evaluation: Create TAM design and optimization evaluation frameworks – Validate proposed design; Analyze and quantify performance compared to conventional design methods.

1.4.3 Methodology

The research was divided into three distinct phases, as shown in Fig. 1-6. As highlighted earlier, scan-based testing using parallel test ports is the most widely used methodology because of high fault coverage. In Phase-1, the integration of SBS in the PTAM approach, based on the IEEE 1500 industry standard [38], was studied. First, an SBS transceiver suitable for low-frequency operation in PTAM was designed and analysed keeping in view the following considerations:

1. The added circuit to enable SBS communication may introduce overheads which must be kept to a minimum.
2. The design of the SBS circuit requires a mixed-signal circuit (analog+digital). Since the target implementation is in digital electronics, the transceiver circuit must be designed using standard CMOS transistor technologies.

The electrical design was validated using transient analysis using industrial chip design and simulation tools. In addition, performance under cross-coupling was validated, and

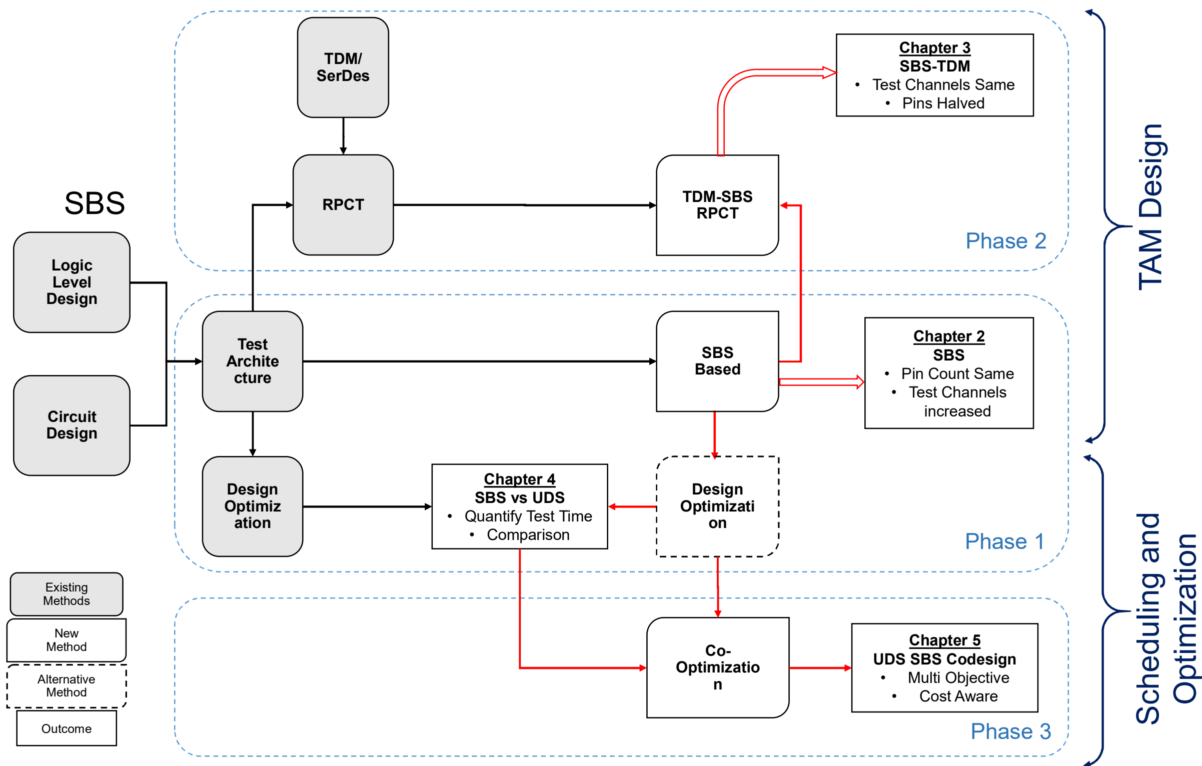


Figure 1-6: Research Methodology showing various phases of the research, the problem addressed during each phase and the relevant chapter of this thesis

a comparison in terms of power consumption was made with the baseline Uni-Directional Signaling (UDS) approach. As evident from Fig. 1-5, the electrical design methodology is one aspect of TAM design; an optimization formulation is required at the application-level (for test time comparisons).

In the second part of phase 1, an optimization framework for the SBS based TAM design was formulated for experiments using existing open-access ITC-02 2D SoC benchmarks circuits [39]. The optimization algorithm was based on Integer Linear Programming [24][23], and session-based scheduling [40]. For various SIC construction, comparisons were made for test time of TAMs based on SBS and UDS.

It may be highlighted that the test time of the IC is composed of two components: (1), the time required for application of the test (TAT) and (2), the time involved in probing of the Chip Terminals on the wafer, known as indexing time (T_{ind}) [41]. Hence, apart from TAT reduction, the test time may also be reduced by decreasing T_{ind}. The reduction in T_{ind} can be achieved by testing multiple dies in parallel, which effectively reduces the total number of touchdowns required by the prober to test the whole wafer. This method is commonly referred to as Multi-Site Testing (MST). However, the tester resources are

limited; therefore, the conventional testing using PTAM may not leave enough resources to be allocated to multiple dies. Under these scenarios, Reduced Pin Count Test (RPCT) methods can be used to test the die using a small subset of pins, thereby sparing tester resources for MST. However, the downside of this methodology is the reduced test coverage which results in lower yield at the wafer test. Therefore, there is a tradeoff to be made between lower wafer-level yield against the test time. The choice between conventional test methods and MST/RPCT is consequently dependent on specific chip design and test economics [42][43].

In phase 1, the transceiver design was restricted to low-frequency operation (suitable for PTAM based design); however, the design considerations for RPCT are considerably different. In phase 2, the feasibility of extending the SBS based method for use in Reduced Pin Count Test (RPCT) scenarios was studied.

1. Methods were proposed to integrate SBS with a Time Division Multiplexing (TDM) based TAM design method.
2. A test bench was implemented to validate functionality using 45nm Nangate Standard Cell Libraries and Verilog models. The TDM-SBS based PTAM output was validated against a high-level functional model implemented in MATLAB.
3. Experiments were conducted to analyse performance gains vs design limitations of SBS-TDM base Test Access Mechanisms under various design corners.

In phases 1 and 2, it was established that the SBS design incurs area and power overheads, and therefore under certain circumstances, the designer would opt to consider penalizing the use of SBS unless there is a considerable reduction in test time. This scenario poses a co-optimization problem in which test time is required to be minimized, considering the trade-offs of preferring SBS over UDS. In phase 3, a multi-objective optimization methodology was formulated to find the optimal trade-off between test time and the resources instead of focusing singularly on either.

Referring back to Fig. 1-5, the circuit level design for SBS based test architecture are covered in Phase 1 (focusing on basic TAM design based on PTAM) and Phase 2 (Advanced TAM design based on RPCT), whereas the application-level implications of SBS in terms of test time were studied in phase 1 (Single objective, SBS only) and phase

3 (Multi-objective, SBS and UDS co-design). In this way, using the research methodology in Fig. 1-6, a wholistic study covering the integration of SBS in the TAM design and optimization fabric of 3D SICs is presented.

1.5 Organization of this thesis:

The thesis is structured in the paper-based format, with every chapter as a self-contained document covering the background, prior art, methodology and results for a particular aspect of the overall problem, as shown in Fig 1-6. For flow considerations, Chapters 2 and 3 cover the electrical aspects of SBS based TAM designs, whereas Chapters 4 and 5 cover the application-level study by formulating the optimization of the TAM design framework. A summary of each chapter is as follows:

Chapter 2 presents the mainstay of the work undertaken in this thesis by studying the feasibility of using the SBS based test methodology in the primary form of chip testing using Parallel Test Ports. The standard approaches of design for testability using parallel Test ports are explained, and the logic level integration of SBS with the traditional methods is proposed. Finally, the chapter outlines the design requirements, presents a transceiver circuit suited to these conditions, analyses the transceiver design, and quantifies the performance constraints and costs regarding on-chip additional resource utilization.

In Chapter 3, the basic SBS based TAM design methodology of Chapter 2 is extended to advanced forms of TAM design presented in the literature such as Time Division Multiplexing (TAM) and Serializer De-Serializers (SerDes). These sophisticated TAM design methods are employed under various scenarios, such as when there are firm dies in the stack, test pins are severely limited, when significant test time reduction is required, or to ensure comprehensive tester resource utilization of high-end testers. The chapter underlines the core requirements for the SBS design employment in these scenarios, which differ from the primary test methodology of Chapter 3 and presents design modifications in the transceiver design. Experiments are conducted by employing SBS in the TDM based TAM design approach recently reported in the literature. The performance and design limitations of SBS in this scenario are examined under various design corners and compared with the relevant prior works.

In Chapter 4, a novel Integer Linear Programming based formulation is proposed, which is generalized to model both SBS and UDS type pin designs allowing incorporation of design and cost considerations of both approaches. This follows a detailed analysis of the SBS-based test method proposed in Chapter 2 at the application level, i.e., regarding the impact of using SBS on the test time compared to the traditional UDS based methods.

In chapter 5, the optimization framework of Chapter 4 is extended to a more generalized scenario which allows the co-existence of both UDS and SBS and co-optimises the cost to benefit ratio. The chapter highlights the trade-off between the test time reduction and the cost in terms of resources utilization (externally in the tester and on-chip resources). Keeping this in view, the ILP formulation inculcates multi-objective scenarios where the focus is finding the optimal trade-off between test time and the resources instead of focusing singularly on either. Detailed analysis on benchmark SOC circuits identifies different scenarios under which the SBS is more suited to UDS and vice versa and earmarks the instances where SBS and UDS may co-exist.

This thesis concludes in chapter 6, in which the prospects of the SBS based TAM design methodology presented in this thesis are discussed, and the directions for future research in this area are highlighted.

1.6 List of Published/Submitted Work

Journal Publications

I.A. Soomro, M. Samie, and I. K. Jennions, “Test Time Reduction of 3D Stacked ICs using Ternary Coded Simultaneous Bi-directional Signaling in Parallel Test Ports,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 0070, no. c, pp. 1–14, 2020.

Soomro, Iftikhar A., Mohammad Samie, and Ian K. Jennions. 2021. “Reduced Pin-Count Test Strategy for 3D Stacked ICs Using Simultaneous Bi-Directional Signaling Based Time Division Multiplexing.” IEEE Access 9: 75892–904. https://doi.org/10.1109/access.2021.3081359.

Under preparation for Journal Publication

I.A. Soomro, M. Samie, and I. K. Jennions, “An Integrated Approach for 3D SIC Test Access Mechanism Design and Optimization using Simultaneous Bidirectional and Conventional Uni-Directional Signalling”

Virtual Conference Presentations

J. Buu-Sao, M. Samie, M. Randa, et. al., "IoT Security – Hardware Perspective", December 2018, the IoT Day Slam 2018, VIRTUAL Internet of Things Conference: https://iotslam.com/session/iot-security-hardware-perspective/.

Test Standard Development

The author participated in the development of “IEEE Standard for Test Access Architecture for Three-Dimensional Stacked Integrated Circuits,” *IEEE Std 1838-2019*. pp. 1–73, 2020, as a member of the IEEE Standards Association Working group on the 1838 standard.

1.7 References

[1] S. Salahuddin, K. Ni, and S. Datta, “The era of hyper-scaling in electronics,” *Nat. Electron.*, vol. 1, no. 8, pp. 442–450, 2018.

[2] G. Philip, B. Christopher, and R. Peter, *Handbook of 3D Integration: Technology and Applications of 3D Integrated Circuits*. John Wiley & Sons, 2008.

[3] G. H. Loh, Y. Xie, and B. Black, “Processor Design in 3D Die-Stacking Technologies,” *IEEE Micro*, vol. 27, no. 3, pp. 31–48, May 2007.

[4] C. Ababei, P. Maidee, and K. Bazargan, “Exploring potential benefits of 3D FPGA integration,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3203, Springer, Berlin, Heidelberg, 2004, pp. 874–880.

[5] K. Kang, L. Benini, and G. De Micheli, “Cost-effective design of mesh-of-tree interconnect for multicore clusters with 3-D stacked L2 scratchpad memory,” *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 23, no. 9, pp. 1828–1841, 2015.

[6] 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, Jan. 2010.

[7] J. Knechtel, O. Sinanoglu, I. (Abe) M. Elfadel, J. Lienig, and C. C. N. Sze, “Large-Scale 3D Chips: Challenges and Solutions for Design Automation, Testing, and Trustworthy Integration,” *IPSJ Trans. Syst. LSI Des. Methodol.*, vol. 10, no. 0, pp. 45–62, 2017.

[8] J. P. Gambino, S. A. Adderly, and J. U. Knickerbocker, “An overview of through-silicon-via technology and manufacturing challenges,” *Microelectron. Eng.*, vol. 135, pp. 73–106, 2015.

[9] Y. Chen, D. Niu, Y. Xie, and K. Chakrabarty, "Cost-effective integration of three-dimensional (3D) ICs emphasizing testing cost analysis," *IEEE/ACM Int. Conf. Comput. Des. Dig. Tech. Pap. ICCAD*, pp. 471–475, 2010.

[10] L.-T. Wen, C.-W. Wang, and X. Wu, *VLSI Test Principles and Architectures*. Morgan Kaufmann, 2006.

[11] B. BAILEY, "Balancing The Cost Of Test." [Online]. Available: https://semiengineering.com/balancing-the-cost-of-test/. [Accessed: 03-Jun-2019].

[12] M. L. Bushnel and lVishwani D. Agrawal, *ESSENTIALS OF ELECTRONIC TESTING FOR DIGITAL, MEMORY AND MIXED-SIGNAL VLSI CIRCUITS*. KLUWER ACADEMIC PUBLISHERS, 2002.

[13] IEEE Computer Society, *1450-1999 - IEEE Standard Test Interface Language (STIL) for Digital Test Vector Data*, no. April. 1999.

[14] E. SPERLING, "Why Test Costs Will Increase," 2018. [Online]. Available: https://semiengineering.com/why-test-costs-will-increase/. [Accessed: 11-Jun-2021].

[15] R. Krenz-Baath, F. G. Zadegan, and E. Larsson, "Access time minimization in IEEE 1687 networks," in *Proceedings - International Test Conference*, 2015, vol. 2015-Novem.

[16] M. A. Ansari, J. Jung, D. Kim, and S. Park, "Time-Multiplexed 1687-Network for Test Cost Reduction," *IEEE Trans. Comput. -Aided Des. Integr. Circuits Syst.*, vol. 37, no. 8, pp. 1681–1691, Aug. 2018.

[17] Y. Fkih, P. Vivet, B. Rouzeyre, M. L. Flottes, and G. Di Natale, "A JTAG based 3D DfT architecture using automatic die detection," *Conf. Proc. - 9th Conf. Ph. D. Res. Microelectron. Electron. PRIME 2013*, vol. 1, no. 1, pp. 341–344, 2013.

[18] M. A. Ansari, D. Kim, J. Jung, and S. Park, "Hybrid test data transportation scheme for advanced NoC-based SoCs," *J. Semicond. Technol. Sci.*, vol. 15, no. 1, pp. 86–95, 2015.

[19] M. Nahvi and A. Ivanov, "Indirect test architecture for SoC testing," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 23, no. 7, pp. 1128–1142, 2004.

[20] E. Cota, M. Kreutz, C. A. Zeferino, L. Carro, M. Lubaszewski, and A. Susin, "The impact of NoC reuse on the testing of core-based systems," *Proc. IEEE VLSI Test Symp.*, vol. 2003-Janua, pp. 128–133, 2003.

[21] B. Aghaei and S. Babaei, "The new test wrapper design for core testing in Packet-Switched Micro-Network on Chip," *PEITS 2009 - 2009 2nd Conf. Power Electron. Intell. Transp. Syst.*, vol. 2, pp. 346–352, 2009.

[22] S. K. Goel and E. J. Marinissen, "Control-aware test architecture design for modular SOC testing," in *The Eighth IEEE European Test Workshop, 2003. Proceedings.*, 2003, vol. 2003-Janua, pp. 57–62.

[23] B. Noia, K. Chakrabarty, S. K. Goel, E. J. Marinissen, and J. Verbree, "Test-Architecture Optimization and Test Scheduling for TSV-Based 3-D Stacked ICs,"

IEEE Trans. Comput. Des. Integr. Circuits Syst., vol. 30, no. 11, pp. 1705–1718, Nov. 2011.

[24] V. Iyengar, K. Chakrabarty, and E. J. Marinissen, “Test Wrapper and Test Access Mechanism Co-Optimization for System-on-Chip,” *J. Electron. Test. Theory Appl.*, vol. 18, pp. 213–230, 2002.

[25] V. Iyengar, K. Chakrabarty, and E. J. Marinissen, “Test Access Mechanism Optimization Test Scheduling, and Tester Data Volume Reduction for System-on-Chip,” *IEEE Trans. Comput.*, vol. 52, no. 12, pp. 1619–1632, 2003.

[26] P. Georgiou, F. Vartziotis, X. Kavousianos, and K. Chakrabarty, “Testing 3D-SoCs Using 2-D Time-Division Multiplexing,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 37, no. 12, pp. 3177–3185, Dec. 2018.

[27] A. Sehgal, V. Iyengar, and K. Chakrabarty, “SOC test planning using virtual test access architectures,” *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 12, no. 12, pp. 1263–1276, Dec. 2004.

[28] A. Sanghani, B. Yang, K. Natarajan, and C. Liu, “Design and implementation of a time-division multiplexing scan architecture using serializer and deserializer in GPU chips,” in *Proceedings of the IEEE VLSI Test Symposium*, 2011, pp. 219–224.

[29] B. Li, B. Zhang, and V. D. Agrawal, “Adopting multi-valued logic for reduced pin-count testing,” *2015 16th Latin-American Test Symp. LATS 2015*, pp. 1–6, 2015.

[30] S. Deutsch and K. Chakrabarty, “Test and debug solutions for 3D-stacked integrated circuits,” *Proc. - Int. Test Conf.*, vol. 2015-Novem, pp. 1–10, 2015.

[31] K. Chakrabarty, M. Agrawal, S. Deutsch, B. Noia, R. Wang, and F. Ye, “Test and Design-for-Testability Solutions for 3D Integrated Circuits,” *IPSJ Trans. Syst. LSI Des. Methodol.*, vol. 7, no. 0, pp. 56–73, 2014.

[32] Y. Fkih, P. Vivet, M.-L. Flottes, B. Rouzeyre, G. Di Natale, and J. Schloeffel, “3D DFT challenges and solutions,” in *Proceedings of IEEE Computer Society Annual Symposium on VLSI, ISVLSI*, 2015, vol. 07-10-July, pp. 603–608.

[33] R. Wang, S. Deutsch, M. Agrawal, and K. Chakrabarty, “The hype, myths, and realities of testing 3D integrated circuits,” in *Proceedings of the 35th International Conference on Computer-Aided Design*, 2016, pp. 1–8.

[34] M. Taouil, S. Hamdioui, and E. J. Marinissen, “Quality versus cost analysis for 3D stacked ICs,” *Proc. IEEE VLSI Test Symp.*, vol. 1, 2014.

[35] M. Agrawal and K. Chakrabarty, “Test-cost modeling and optimal test-flow selection of 3-D-stacked ICs,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 34, no. 9, pp. 1523–1536, 2015.

[36] Y. Y. Wu and S. H. Huang, “TSV-aware 3D test wrapper chain optimization,” *2018 Int. Symp. VLSI Des. Autom. Test, VLSI-DAT 2018*, pp. 1–4, 2018.

[37] R. P. Reddy, A. Acharyya, and S. Khursheed, “A Cost-Effective Fault Tolerance Technique for Functional TSV in 3-D ICs,” *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 25, no. 7, pp. 2071–2080, 2017.

[38] "IEEE Standard Testability Method for Embedded Core-based Integrated Circuits," *IEEE Std 1500-2005*, pp. 1–136, 2005.

[39] E. J. Marinissen, V. Iyengar, and K. Chakrabarty, "A set of benchmarks for modular testing of SOCs," in *Proceedings. International Test Conference*, 2002, pp. 519–528.

[40] P. Varma and B. Bhatia, "A structured test re-use methodology for core-based system chips," in *Proceedings International Test Conference 1998 (IEEE Cat. No.98CH36270)*, 2002, pp. 294–302.

[41] A. C. Pfahnl, J. H. Lienhard V, and A. H. Slocum, "Thermal management and control in testing packaged integrated circuit (IC) devices," *SAE Tech. Pap.*, no. August 2016, 1999.

[42] A. Khoche, R. Kapur, D. Armstrong, T. W. Williams, M. Tegethoff, and J. Rivoir, "A new methodology for improved tester utilization," in *Proceedings International Test Conference 2001 (Cat. No.01CH37260)*, 2001, pp. 916–923.

[43] Q. Khasawneh, "Reducing the Production Cost of Semiconductor Chips Using (Parallel and Concurrent) Testing and Real-Time Monitoring,", Electrical Engineering Theses and Dissertations, 2019.

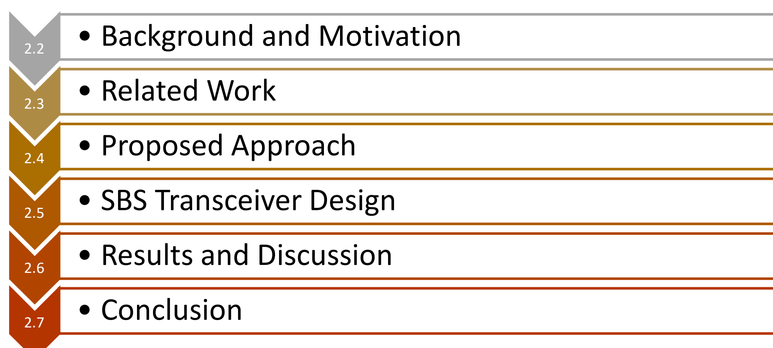
[44] E. J. Marinissen, T. McLaurin, and Hailong Jiao, "IEEE Std P1838: DfT standard-under-development for 2.5D-, 3D-, and 5.5D-SICs," in *2016 21th IEEE European Test Symposium (ETS)*, 2016, no. 1, pp. 1–10.

2 DESIGN AND INTEGRATION OF A TERNARY LOGIC BASED SIMULTANEOUS BIDIRECTIONAL SIGNALLING TRANSCEIVER IN PARALLEL TEST PORTS OF 3D STACKED INTEGRATED CIRCUITS

2.1 Abstract

This chapter discusses the electrical design aspects of Simultaneous Bi-Directional Signaling, focusing on its application in Parallel Test Access Mechanisms in the 3D Stacked ICs. Parts of this chapter have been peer-reviewed and published in IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems [1].

This chapter has been divided into 6 sections. In section 2.2, a brief background in scan-based testing is provided; the factors affecting the Test Application Time (TAT) are discussed, followed by a motivating example to explore the benefits possible using Simultaneous Bi-Directional Signalling. In section 2.3, the relevant prior art in TAT reduction and SBS is discussed. Section 2.4 discusses the working principles of SBS and the TAM design considerations for incorporating SBS in 3D SICs such that it does not interfere with the functional mode performance and standard DFT logic such as JTAG compliant boundary scan registers. In section 2.5, a transceiver circuit design suitable for low-frequency test vector transportation is presented for use in parallel test ports. Results are presented in section 2.6. The electrical design was validated using transient analysis using industrial chip design and simulation tools. Performance under cross-coupling is validated, and a comparison in terms of power consumption with the baseline UDS approach is presented. This chapter concludes in section 2.7.



2.2 Background and Motivation

As conventional 2D chips face the problem of diminishing scalability, 3D Stacked Integrated Circuits (SICs) are a promising way forward to address the demand for increased transistor density and power efficiency without increasing the package footprint. However, the realization of 3D SICs requires innovations to address new challenges unique to the vertical stacking [2][3][4]. One of the challenges is to meet the additional test requirements while keeping the test costs in check [5]. First, higher transistor density increases the probability of manufacturing defects and requires higher test vector volume for adequate coverage. Second, the partial 3D stacks also need to be tested during the manufacturing process, adding multiple test instances compared to 2D SoCs. Finally, the inter-die vertical connections (such as Through Silicon Vias (TSVs) are available in limited quantities, causing a bottleneck in transporting test vectors to dies higher up in the stack.

Testing of a core-based chip design involves three main components: a) The cores and the wrappers, b) a tester which generates required test vectors, controls the test operation and evaluates test response; and c) a Test Access Mechanism (TAM) which transports the test patterns/ responses between the cores and tester. Scan based testing is the most common way of Design for Testability (DFT) in core-based designs. The functional front end of the chip is designed as usual, and later the Flip Flops are made scan-test accessible by forming them into shift registers or scan chains using CAD tools. Testing is performed by sending in a set of pre-calculated test vectors to these scan chains and observing the response. Scan-based testing has been widely used in post-silicon testing focusing on improved fault coverage and minimal resource usage. The scalability of the ATPG and high fault coverage of scan-based testing, coupled with industry-accepted standards make scan-based testing a viable approach for 3D SIC testing. However, a typical scan-based testability requirement is that all the I/O pins and the internal flip-flops (storage) elements are made accessible (observable and controllable) for testing. Therefore, a Test Access Mechanism (TAM) must be designed and inserted into the design.

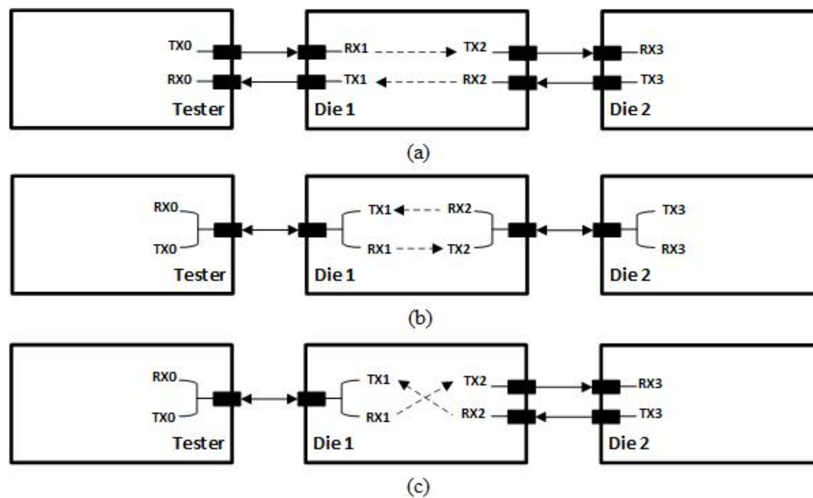


Figure 2-1: A single test channel using: (a) Conventional signaling – uses two wires (b) SBS – uses one wire (c) A combination of Uni- and SBS – one wire between tester and Die 1, two between Die 1 and Die 2

The TAM for test vector transport could be designed in several ways. The simplest being a Serial Test Access Mechanism (STAM) such as IEEE 1149.1 (aka JTAG) and its variant IEEE 1687 standard (iJTAG) [6][7]. However, STAM only has a single serial channel, which means that the data has to be shifted one bit at a time, severely limiting high data volume transfer for which a Parallel Test Access Mechanism (PTAM) is used. In a PTAM, such as that allowed by IEEE 1500 Standard [8], a chip's functional I/Os are temporarily used to enable data transfer on multiple test channels in parallel instead of one. It may be noted that the data is still shifted serially through the PTAM but using a higher number of test channels. Here, a 'channel' is defined as a single bit path capable of transporting a test vector to and from the Automatic Test Equipment (ATE) and the core under test. A similar architecture based on a combination of STAM and PTAM is recommended for 3D SICs in the recently established IEEE 1838 Standard [9].

In the conventional TAM design, a test channel is formed using a separate terminal for input and output; therefore, the number of test channels in a PTAM is half the available chip terminals, as shown in Fig. 2-1(a). This is because conventional chip terminals are designed to communicate in simplex (uni-directional) or half-duplex (bi-directional) configuration. In either case, only a single transmit-receive pair is active at a given time, and the data could only travel in one direction. This simplifies the hardware implementation and has been

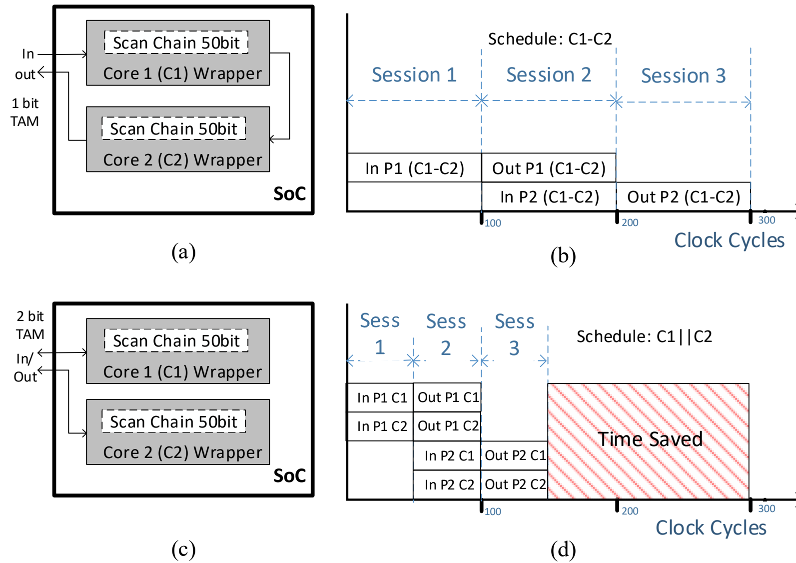


Figure 2-2: Conventional vs SBS test ports (a) TAM design for uni-directional port and associated test schedule in (b), (c) TAM design for SBS port and associated test schedule in (d)

sufficient in keeping the TAT of medium complexity chip designs down to an acceptable level; however, it does not scale well for more complex designs such as 3D SICs, which demand a higher number of test channels.

On the other hand, if a full-duplex configuration such as SBS is used in the TAM, the data could be shifted in and out at the same pin simultaneously, as shown in Fig. 2-1(b), resulting in channel width equal to the number of chip terminals. The parallelism introduced by SBS increases the number of test channels, significantly reducing the TAT. Consider the example of an SoC with two cores and two chip terminals. Each core has a single scan chain of 50 bits and requires two test patterns (say P1 and P2) for the scan-test. In a conventional TAM design using simplex/ half-duplex chip terminals, the two pins would form a 1 bit wide TAM. Consequently, both the scan chains could be concatenated to form a single channel of 50 + 50 =100 bits as shown in Fig 2-2(a). If however, an SBS based TAM is used, the resulting TAM would be two channels of 50 bits each as shown in Fig 2-2(c). The resulting schedule for both arrangements is shown in Fig 2-2(b) and (d). Cx-Cy denotes that the cores x and y will be connected in series, and Cx||Cy indicates the cores will be connected in parallel to form a test session. It is evident that the TAT in the case of uni-directional TAM with schedule C1-C2 is 300 clock cycles because of a single TAM channel, whereas for SBS TAM with schedule C1||C2, the TAT is only 150 cycles. It could, therefore, be concluded

that SBS ports increase the available TAM width, allowing more parallelism, which in the case of this example resulted in a reduction of 150 cycles in the TAT.

The added advantage of using SBS is that it can work in conjunction with the conventional TAM, as shown in Fig. 2-1(c). This means that instead of modifying all chip terminals to support SBS, only the most essential subset causing the bottleneck may be fitted with SBS while the remaining chip terminals operate as usual. This also implies that this method can be used to integrate hard-dies, in which the chip terminals are not modifiable.

2.3 Related Work

The rising test complexity and test times due to increasing transistor densities have attracted significant research from the scientific community into test cost reduction. Unlike 2D SoCs, which only required wafer and package level testing, 3D SICs involve addition mid-bond test instances (for the partially bonded stacks). Selection of appropriate test-flows (i.e. combinations of pre, mid and post-bond tests) in the 3D SIC manufacturing cycle offer different trade-offs among fault coverage, yield and test cost [10][11][12]. Nevertheless, testing at all instances promises better yield [13], as defective dies or stacks could be identified and removed during early production stages. However, given many test instances in 3D SICs, the effect of higher test time per die compounds multiple times, necessitating efficient and re-usable TAM designs offering lower test time.

The Test Application Time (TAT) of an SoC depends primarily on the test data volume (V), scan frequency (s_f) and the available number of test channels (T_{ch}) in a PTAM. A simplified estimate of the TAT can be given by $TAT = V/(s_f.T_{ch})$. Clearly, the TAT decreases with decreasing V and increasing s_f and T_{ch}. The test data volume (V) reduction can be achieved using test compression techniques [14]; Several test data compression techniques have been proposed using different compression coding methods such as Golomb coding [15], statistical coding [16] and Frequency-Directed Run-length (FDR) coding [17]. However, test compression methods necessitate on-chip decompressors adding significant area overheads, also beyond a certain point, test compression is likely to come at the cost of reduced fault coverage. The scan frequency (s_f) is limited by

thermal and design constraints; test scheduling has been frequently used to keep thermal properties within limits, allowing maximum possible scan frequencies [18][19]. However, the scan chain insertion is not optimized for performance, and the high switching activity during the shift and capture cycles consumes significant power. Therefore, the scan frequencies are usually limited to a few tens of MHz [20]. Most of the conventional TAM design methods, therefore, rely on either increasing the number of test channels (T_{ch}), which is limited by the chip pins or by efficient utilization of the available pins.

The works that involve improvement of pin-efficiency include Time Division Multiplexing (TDM) [21][22], Serializer and De-serializer [20][23] and using Multi-Valued Logic (MVL) [24]. The work in [21] has been concerned with reducing access time for serial Reconfigurable Scan Networks (RSNs). The authors in [22] proposed a TDM based method to reduce the overall test time of 3D SICs. The data is loaded through parallel buses at the scan frequency and is then serialized. The serial data can then be transferred from one die to another through TSVs operating at higher frequencies. Time de-multiplexers at the receiving end perform the serial to parallel conversion, and the data is shifted into scan chains at the scan frequency. Another approach to allow optimal utilization of tester resources was presented in [20] and [23]. In [20], the authors introduced the concept of virtual TAMs to utilize tester resources efficiently. Instead of operating the tester channels at a lower frequency to match scan frequency, Serializer and De-serializer (SerDes) were used to enable high-frequency data transfer between ATE to SoC I/O and low-frequency operation at the scan chains, thus maximizing ATE resource utilization. The approach adopted in this paper focuses on Simultaneous Bidirectional Signaling (SBS) at the chip terminal, which internally presents two virtual unidirectional I/Os to the SoC, just like UDS. The test time reduction techniques based on TDM and SerDes could be designed on top of these virtual I/Os providing further test time reduction, with an added advantage that SBS would require one pin, whereas UDS would need two.

The idea of SBS was initially reported in [25], following which improved designs capable of delivering up to 900Mbps and 8Gbps (450Mbps and 4Gbps in either direction, respectively) were proposed and tested on fabricated devices in [26] and [27]. While these works rely on voltage-mode signaling, further enhanced

design using current-mode differential transceivers were proposed in [28][29][30] for improved power efficiency in off-chip (chip to chip) [28][29] as well as on-chip (core to core) [30] signaling. In 3D SICs, SBS through TSVs has different design requirements, primarily because of negligible path resistance compared to off-chip communication. The authors in [31] presented an SBS transceiver design for vertical communication in 3D SICs through TSVs and achieved a data rate of 9.1Gbps. The performance of a single SBS channel has been shown to be better than two UDS channels, both in terms of power and the on-chip area [26][31]. The decrease in power consumption is attributed to lower switching activity as well as reduced voltage swing in SBS.

The research in SBS has been chiefly focused on functional communication at higher data rates. High-speed SBS is usually complicated by noise concerns such as common-mode noise rejection, echo cancellation, EMI considerations, and tight control of threshold voltage and comparator tolerances. However, as mentioned earlier, scan frequencies are typically a few tens of MHz, for which the design considerations become much relaxed, making testing a very viable application of SBS. The difference, however, is that unlike single-ended chip to chip channels where communication takes place between two transceivers, in this case, the chip terminal needs to be capable of functional communication in normal mode (assuming it is UDS) and SBS in test mode. To the best of our knowledge, TAM design methodology in 3D SICs using SBS has not been studied in the past.

2.4 Proposed Approach

Simultaneous Bi-directional signaling could be made possible by using ternary level encoding at the chip terminals (Pins and TSVs). The output from the sender and receiver is encoded into the ternary level at the chip terminal, and a decoding circuitry is used to convert back to binary levels. Here it may be noted that the chip-terminal could either be the chip pin (only in case of the die which connects to the PCB, mostly the lowermost die) or a TSV. It is assumed that the functional communication at the chip terminal is UDS, and the proposed design aims to add SBS capability specifically for use in the test mode.

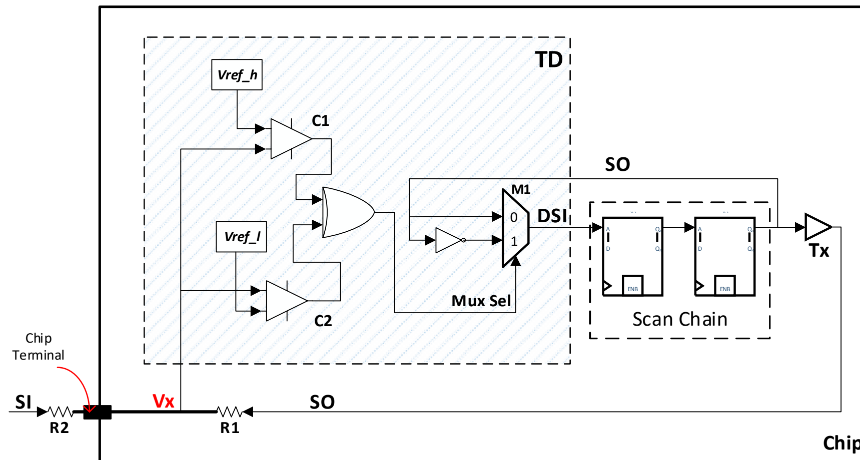


Figure 2-3: Test Channel for a 2 bit Scan-Chain using Ternary Encoding and Decoding at the Chip Terminal

2.4.1 Ternary Level Coding

The overall idea is illustrated in Fig. 2-3, using an example chip with a single scan chain comprising of two flip-flops (FFs). The output of the scan chain is denoted as Scan Out (SO) and is fed to the chip terminal using an inverter as a transmitter (Tx), which becomes active during SBS mode. The external end of the chip terminal is to be driven using a similar setup (not shown in the figure) in another die (in case of TSVs) or the tester (in case of the first die) and is denoted as Scan In (SI). Two series resistors of equal value R1 (located on-chip) and R2 (on tester/another die) form a voltage divider circuit. While the SI and SO signals are digital, the output of the voltage divider (Vx at the node X) will be ternary encoded and can take three distinct values i.e., 0 (Vxl), Vdd (Vxh), or ½ Vdd (Vxm) depending on if SI and SO are both low, high or in opposite states, respectively. It may be noted that the resistors R1 and R2 are explicitly shown for clarity; in actual implementations, the SI and SO line driver output impedances and the I/O cell resistance may be sufficient to form the voltage divider circuit.

The Ternary Decoder (TD), shown in the hatched block, receives a copy of the SO signal, which is tapped just before the Tx, where it remains in the binary state. The other input to the TD is the ternary encoded signal Vx, which is fed to one of the inputs to the two voltage comparators C1 and C2. The second input of the comparators is connected to high voltage reference Vref_h and low voltage reference Vref_l, which in this case are taken to be 2/3 Vdd and 1/3 Vdd, respectively. If Vx is 0 or Vdd, both comparators produce the same output (0 or

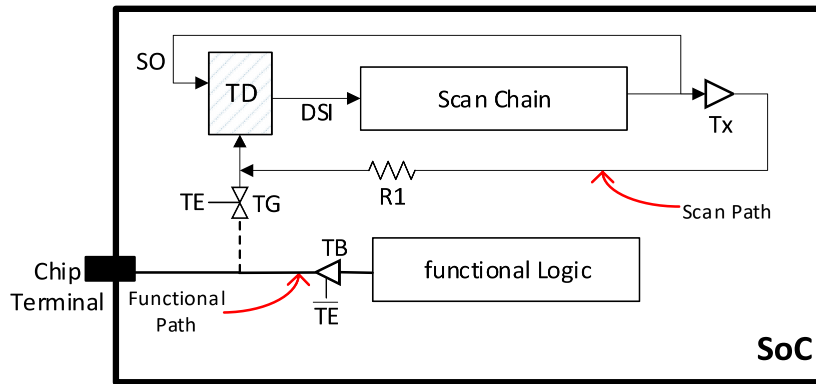


Figure 2-4: Test and Functional mode isolation (in this illustration, the chip terminal is a functional output)

Table 2-I: Mode Configuration States for SBS Integration with BSR in Figure 2-5(b)

Mode	TE	TB	TG	M1
Functional	0	Drive	Open	X
JTAG	0	Drive	Open	TDI
INTEST SBS	1	HiZ	Closed	DSI

1), and if Vx is equal to 0.5 Vdd, C2 produces a 1 (since 1/2Vdd > 1/3Vdd), and C1 produces a 0 (since 1/2Vdd < 2/3Vdd). The outputs of C1 and C2 could be fed to an XOR gate, which in turn controls a 2 to 1 Multiplexer M1. One input of the Multiplexer M1 receives the SO signal, while the other input received an inverted copy of the SO. The output of the Mux is the output of the TD and is denoted by Decoded Scan In (DSI) signal, which is the input to the Scan Chain. The TD forces the DSI to take on the same value as SI, simply by deciding whether it is the same as SO or the opposite, which is determined by the current state of the ternary encoded Vx signal. Of the 3 states of Vx, a 0 or a 1 indicate that the incoming signal must be the same as the outgoing signal (in which case Mux Sel will be low), whereas the third level (Vxm) would indicate the opposite (in which case Mux Sel will be high). In this way, the transmission and reception of the signal could be achieved simultaneously.

2.4.2 Test and Functional mode isolation

The use of SBS signaling in testing is complicated by the requirement of usability of the chip terminals in the functional operation of the chip as well. The design of the SBS must take into consideration that the functional path, which could be required to operate in the GHz range, may not be affected by the presence of SBS mode connections. Therefore, a Transmission Gate/ Analog Switch (TG) is inserted just before the TD on the ternary encoded signal Vx, and the drivers or the receiver at the functional side is designed as a Tristate Buffer (TB). These switches are controlled by the Test Enable (TE) signal of the Chip, which could either be provided externally through a dedicated chip terminal or could be sourced from the die-level JTAG IR decoder by loading an appropriate user-defined instruction via the TAP controller. The overall arrangement is illustrated in Fig 2-4. In this case, TE is de-asserted, and TG isolates the test side (dashed line) while TB is active, allowing the normal operation of the functional side (solid line). On the contrary, when TE is asserted, the functional side will be isolated.

2.4.3 Integration with Boundary Scan

JTAG is a widely used DFT feature that allows essential test accessibility features at all levels of system hierarchy, such as die, chip, circuit board, and system level. Boundary Scan Registers (BSR) is an essential component of JTAG, which allows the observability and controllability at the chip pins for test and debug purposes. Therefore, it would often be necessary to ensure that the incorporation of SBS does not affect the boundary scan capability of the chip. An illustration of an observable and controllable Boundary Scan Cell (BSC) (using an example implementation given in [6]) is shown in Fig 2-5(a). The proposed incorporation of SBS for a functional pin (in this case, an input) could be achieved, as shown in Fig. 2-5(b). As with the non-boundary scan chip terminal (Fig 2-4), the tri-state buffer is used as a receiver along with TG for isolation. A Multiplexer M1 is inserted between Test Data In (TDI) of the JTAG and R1 flip-flop of the BSR, such that depending on the state of control signal TE, either conventional boundary scan is selected (TE is low) or SBS is selected (TE is high) as shown in Table 2-I. The state of the multiplexers M2 and M3 will depend on the current instruction in the JTAG IR. This configuration supports all the functionality of a

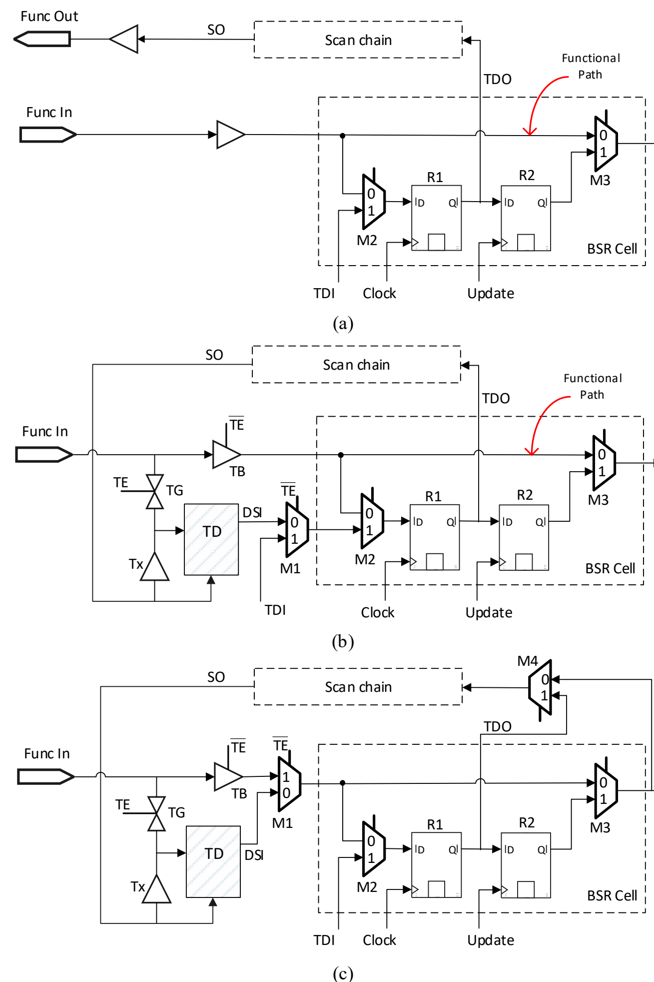


Figure 2-5: Boundary scan compliant fully observable and controllable boundary scan cell [6] shown in the dashed box (a) without SBS (b) SBS integration outside functional path (c) SBS integration inside functional path

conventional BSR, such as normal function, EXTEST, INTEST, SAMPLE, and PRELOAD, with an added option of SBS to support Parallel INTEST (for example, when using WPPs of the IEEE 1500 standard compliant core wrappers).

It may be noted that the implementation, shown in Fig. 2-5(b), does not add any logic in the functional path and does not incur any additional penalty on functional mode performance. However, this comes at the cost of increasing the scan-chain length by 1 bit, as the BSR Cell must be a part of the scan chain. An alternate arrangement is shown in Fig. 2-5(c), which allows the scan chain to be fed either from the BSR or from the chip terminal (using M4) while retaining all the functionality of Fig. 2-5(b). However, it necessitates the inclusion of the multiplexer M1 in the functional path, which may affect the functional mode performance. If the performance degradation is acceptable, the arrangement in

Fig. 2-5(c) is preferable because SBS can now be used for both functional and test mode communication along with UDS. Moreover, both these implementations also ensure that in case of a defect in the SBS circuitry, the testing could still be performed using conventional UDS based DFT resources, thus providing redundancy.

2.4.4 Vertical Access Considerations in mid- and post-bond testing

SBS implementation through TSVs is different in terms of transceiver design characteristics due to the low resistance and high capacitance of the TSV path. Therefore, further considerations are required for overall TAM design when accessing the higher dies through the first die. Consider, for example, the case in which the signal is required to be transmitted to the second die from the chip terminal of the first die, as illustrated in Fig. 2-6. In this case, the Ternary Encoded signal at the chip pin is required to traverse through more than one SBS Transceiver (SBS TR). The input signal SI1.0 at the 1st die pin must travel through the SBS TR1, followed by the SBS TR2 to reach the second die, while at the same time, the scan-out signal SO1.1 at the second die must travel to the chip pin in the opposite direction. It is clear that the signal traversal path may induce delays, which may become excessive when accessing dies that are further up in the stack. Therefore, bypass flip-flops must be inserted to ensure that the signal does not degrade when passing through multiple dies. The additional flip-flops ensure that the signal only propagates through one die at a time utilizing a full clock cycle. This problem is not specific to SBS and is relevant to any TAM design. The standard DFT practice for 3D SICs is expected to include mid-way flip-flops in the PTAM to avoid signal integrity issues altogether [9]. Fig. 2-6 shows the bypass flip-flops inserted in the scan in (By In) and scan out (By Out) paths, respectively. The addition of the flip-flops results in an increase in the scan path length for Die 2 and onwards, depending on the position of the die in the stack.

In order to calculate the test time of the cores in 3D stacked dies using SBS, the equation for core test time calculation of 2D SoCs given in [32] can be extended to stacked dies with bypass flip-flops. The test time of the core T_c is then given by:

$$T_c = (s + \max\{s_i, s_o\}).p + \min\{s_i, s_o\} + s - 1 \qquad (1)$$

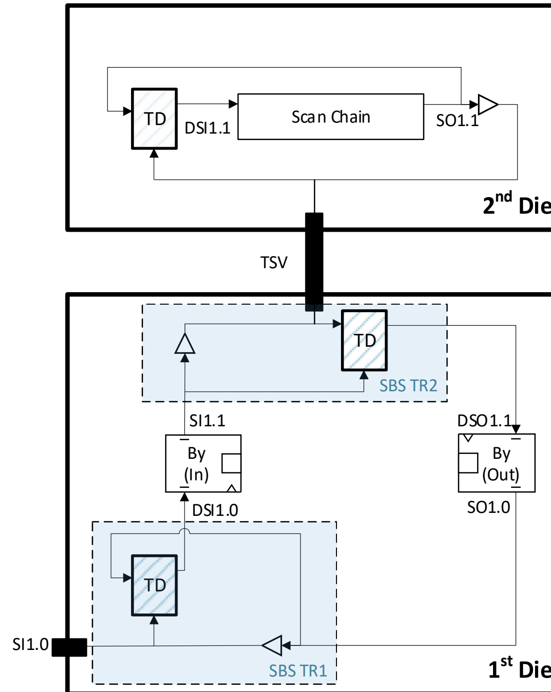


Figure 2-6: An illustration of SBS for accessing higher dies in a 3D SIC through TSVs

Where s is the position of the die in the stack, s_i and s_o represent the longest scan-in and scan-out chain lengths of the core wrapper, respectively, and p is the total number of test patterns required by the core. This affects the test time of every core in the SIC except the first die in which case s = 1, and the equation reduces to $T_c = (1 + \max\{s_i, s_o\}).p + \min\{s_i, s_o\}$ which is the same as given in [32] for 2D SoCs.

2.4.5 Pre-bond testing

The case of TSV accessible dies is complicated due to difficulty in probing the TSVs at the pre-bond stage. Current processes are capable of producing TSVs with pitch and diameter of less than 5μm [33][34][35], which is too small to be accessed through tester probes. Although a die may contain hundreds of TSVs, due to the stated probing issues, only a tiny subset of TSVs may be made accessible at the pre-bond stage using sacrificial probe pads. Therefore, the problem of pre-bond testing of TSV accessible dies can be considered similar to the pin accessible dies. The TAM design problem for the pre-bond testing may

only be required to ensure the maximal utilization of SBS resources in all test instances (pre- mid- and post- bond).

2.4.6 Reference Sharing

The comparators C1 and C2 of the Ternary Decoder shown in Fig. 2-3 require a high and low reference voltage (Vref_h and Vref_l). For low-frequency applications, the dies can have separate references generated locally on the die or sourced through the tester. However, for high-frequency applications, in order to couple the common mode noise to the receivers as well as to cancel out the effect of power supply variations, it may be necessary to have a shared reference between the dies, as shown in Fig. 2-7.

The requirement of two wires for reference generation between the first die and the tester and between the dies also reduces the number of test pins and TSVs available for the transportation of the test vectors. If the number of available test pins is denoted by P_{max} and the total number of TSVs in the entire stack is given by TSV_{max}, then the pins available for testing (W_{bi}) and TSVs available for testing (TSV_{bi}) using SBS is given by:

$$W_{bi} = P_{max} - 2 \quad (2)$$

$$TSV_{bi} = TSV_{max} - 2(M - 1) \quad (3)$$

Where M is the number of dies in the stack. Despite the reduction in test pins and TSVs in the SBS scheme, the number of test channels increases significantly compared to the conventional uni-directional approach since a single wire forms a channel in the former approach, whereas two wires form one channel in the latter approach.

2.5 SBS Transceiver Circuit Design

In this section, an example SBS Transceiver (SBS TR) design is presented for use in low-frequency test mode for 3D TSV communication. There are several ways in which SBS can be implemented. For high-frequency applications, differential mode communication is used [36] [37]. Although differential mode transceivers are power efficient and highly noise resistant, allowing very high

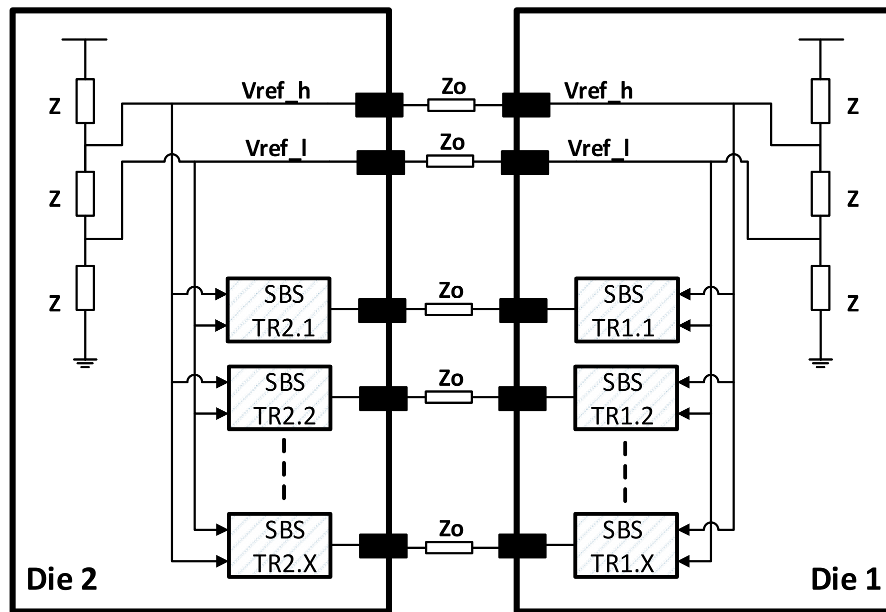


Figure 2-7: Reference Sharing

bandwidth, the requirement of two pins to form a channel limits its use in Parallel Test Ports, for which single-ended transceiver designs [31][38][39] is preferable.

The above works are designed for normal mode communication for high-speed data transfer. The implementation discussed below is intended to be used as an additional circuit for use in test mode such that a) the effect on functional performance is minimal b) Given the low-frequency requirements, the implementation is simplified and power efficient.

The main components of an SBS implementation shown in Fig. 2-8(a), where SBS communication takes between the 2nd Die and the 1st Die through a TSV. The Transmitter and the Ternary Decoder (TD) in both dies are similar; therefore, the detailed schematic is shown only for the 2nd die. In the following paragraphs, the circuit design is described in light of the different design options.

2.5.1 Transmitter

The transmitter can simply be designed as an inverter of appropriate size to allow the required bandwidth. However, this design may not be power efficient when the driving transmitters at either end of the channel are in opposite states, consuming static power, as reported by the authors in [31]. The proposed transmitter is built as an inverter with diode-connected MOSFETs to limit the static current, as shown in Fig. 2-8(a) [28]. The MOSFETs M_{NS} and M_{PS} perform

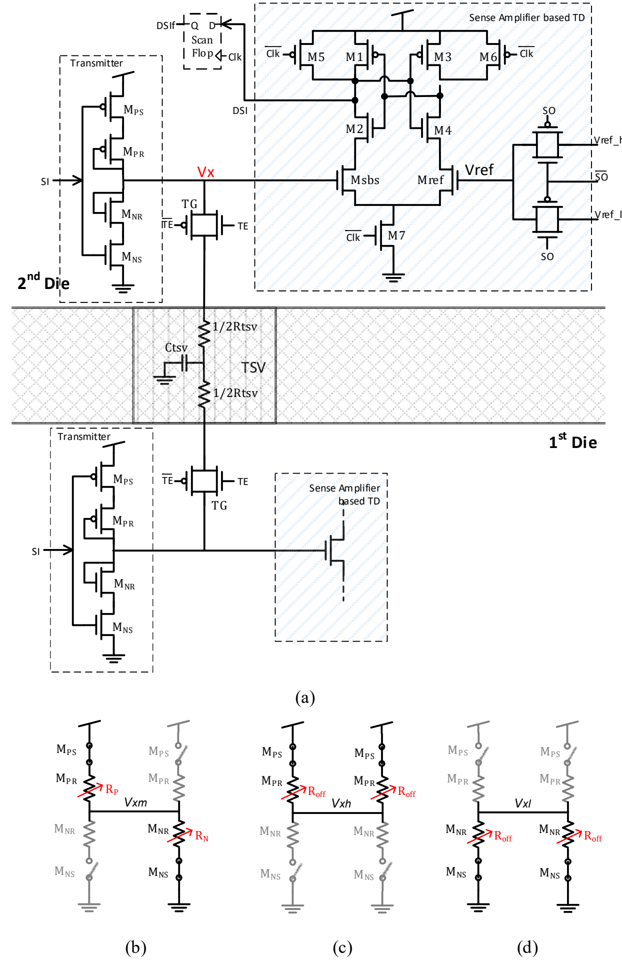


Figure 2-8: (a) Proposed SBS Transceiver Circuit (b) equivalent circuit for Vxm (c) equivalent circuit for Vxh (d) equivalent circuit for Vxl

the inverter switching and the diode-connected MOSFETs M_{NR} and M_{PR} serve as active series resistors, minimizing the static current and hence the static power consumption. Similar to a normal inverter, there is no static power consumption when the transmitters at both ends are either high or low.

The resistance of the diode-connected MOSFETs at a given instance is a function of V_x, and consequently the state of both the transmitters (11,10,01,00). Fig. 2-8(b) shows the equivalent circuit for the middle voltage level (V_{xm}) when one of the transmitters is high and the other is low (10, 01). Ignoring the TSV resistance and assuming the switching transistors as ideal, the middle voltage level $V_{xm} = V_{dd}(R_P/(R_P + R_N))$ where R_P and R_N are the resistance of M_{PR} and M_{NR} at $V_x = V_{xm}$. The resistance R_{on} of a diode-connected MOSFET can be approximated by:

$$R_{on} = \frac{V_{DS}}{I_d} \; where \; I_d = \mu C_{ox}\frac{W}{L}(V_{GS} - V_t)V_{DS} \quad (4)$$

therefore,

$$R_P = \frac{L_p}{\mu_p C_{ox} W_p (V_{xm} - V_{dd} - V_{t(p)})} \qquad (5)$$

$$R_N = \frac{L_n}{\mu_n C_{ox} W_n (V_{xm} - V_{t(n)})} \qquad (6)$$

Where L and W are the length and width of the diode-connected MOSFETs, μ is the mobility of the channel, C_{ox} is the oxide thickness, and V_t is the threshold voltage. It is clear that the W and L ratios of the diode-connected NMOS and PMOS can be adjusted to obtain the desired V_{xm} level. When both the transmitter inputs are low (00), as shown in Fig 2-8(c), V_x is pulled high through the M_{PS} and M_{PR}, however as V_x approaches V_{dd}, M_{PR} enters into the sub-threshold region and the resistance R_{P1} approaches the off resistance R_{off}, restricting the upper voltage swing to $V_{xh} = V_{dd} - V_{t(p)}$. Nevertheless, since there is a small conduction current in the cut-off region as well, V_x will gradually approach V_{dd} and hence $V_{dd} \geq V_{xh} \geq V_{dd} - V_{t(p)}$. Similarly, when both the transmitters are sending high (11), V_x is pulled low to $V_{xl} \leq V_{t(n)}$ as shown in Fig. 2-8(d). The upper and lower voltage swing can be further improved by using the body effect to reduce V_t.

In Fig 2-8(a) the transmission gate switch *TG* added after the transceiver ensures that the transceiver does not affect the normal mode operation (*TG* is open when *TE*=0). However, the diffusion parasitic capacitances of *TG* do appear in the functional path, but it is minimal compared to the TSV capacitance, and the effect on normal mode performance is expected to be negligible.

2.5.2 Ternary Decoder

The main component of the ternary decoder is the voltage comparator which could be designed either as a differential amplifier [38] or a voltage-sense amplifier [31]. The proposed receiver has been based on the latter, because of its simple design, robustness, and low power consumption. The circuit diagram of the sense-amplifier based TD is shown in Fig. 2-8(a). Note that the TD in Fig. 2-3 contained two comparators, an XOR gate and a multiplexer, however for the

area and power efficiency the proposed implementation is optimized such that the XOR gate and the multiplexer isn't required and only one comparator is used and the reference is switched between the high value (*Vref_h*) and low (*Vref_l*). When the scan-out signal is high (SO=1), the transmitter output is low and the signal *Vx* can only take the low and middle value and vice-versa. Therefore, the lower reference *Vref_l* is selected when SO=1 and *Vref_h* when SO=0. The reference switching is achieved using the transmission gate multiplexer with SO as the control signal.

The sense amplifier receives the ternary encoded input *Vx* and the reference voltage *Vref* at the gates of NMOS transistors Msbs and Mref, respectively. The transistor pairs M1, M2 and M3, M4 form two cross-coupled inverters, forming a regenerative latch. The sensing takes place during the positive half cycle. The transistors Msbs and Mref, depending on the voltages at the respective gates, will have different on-resistance and hence the voltage drop. In the negative half cycle, one of the nodes of the regenerative latch with the higher voltage is pulled high, and the lower voltage is pulled low, hence the comparator action. Since the input to TD was inverted by the transmitter, the inverting output of the sense-amplifier is taken as the output DSI, which can be fed directly to the scan chain. The transistors M5 to M7 are controlled by the clock and allow the latch action.

The proposed TD behaves like a neg-edge triggered flip-flop and introduces a delay of one clock cycle; however, it removes the requirement of a separate flip-flop required for Die-to-die communication, as shown in Fig. 2-6. As noted earlier in Sec. IV(D), the inclusion of the flip-flops is necessary from the DFT standpoint and is expected to be a part of the upcoming 3D IC DFT standard P1838 [9].

2.6 Simulation results

The SBS Transceiver was simulated with Cadence Spectre using 180nm technology. The design was limited to 50 MHz frequency, which was easily achieved using minimum size for all transistors in the transceiver. Assuming a TSV with 5µm diameter, 20 µm length, substrate doping concentration Na of $2x10^{15}/cm^3$, and Oxide Thickness of 200nm, the TSV was modelled based on [40] as a lumped RC circuit with a resistance *Rtsv* ≈ 100mΩ and capacitance *Ctsv* ≈ 30fF as shown in Fig. 2-8. The series inductance of the TSV was ignored as it

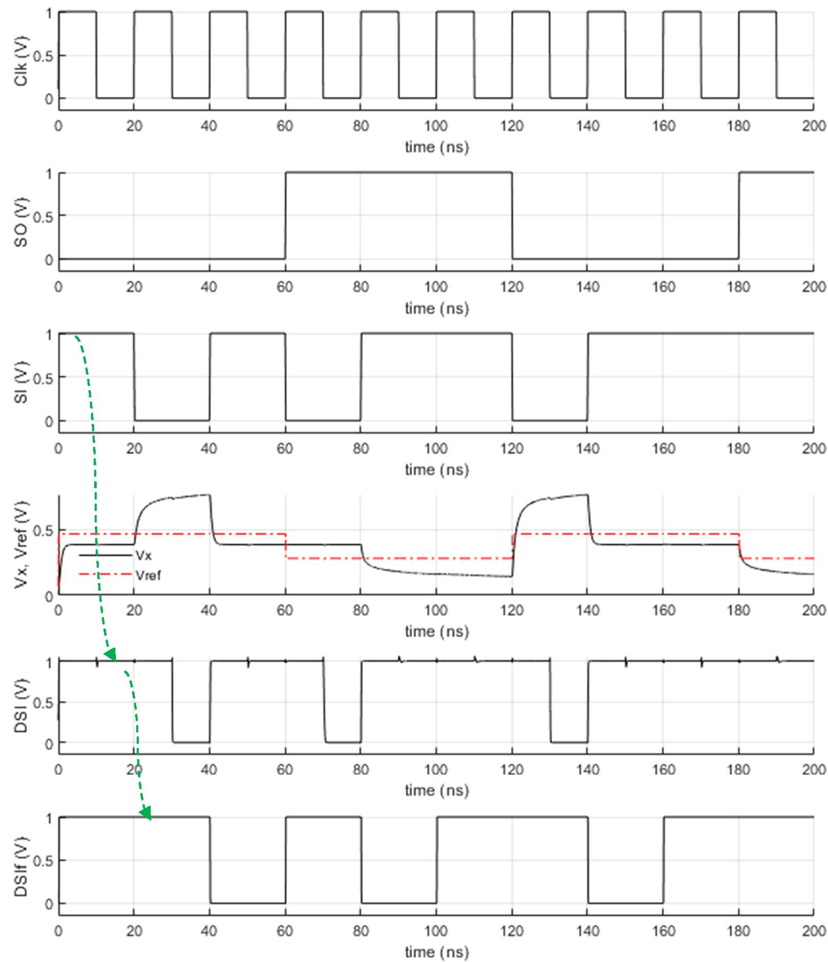


Figure 2-9: Simulation results of the various signals in the proposed SBS Transceiver Circuit in Figure 2-8. (Verti-cal scale normalized to Vdd).

is negligible at low frequencies. Minimum sized transistors provided a *Vx* swing with *Vxh*≥0.7Vdd, *Vxm*≈0.39Vdd, and *Vxl*≤0.2Vdd. *Vref_l* and *Vref_h* were chosen to be 0.28Vdd and 0.47Vdd, respectively. The voltage levels at various points for the four possible combinations of *SI* and *SO* are shown in Fig. 2-9. The output, *DSI*, of the Ternary Decoder, correctly reproduces the Scan-In (*SI*) signal and appears at the output of the first scan-flop (*DSIf*) with a delay of 1 clock cycle.

The power consumption of the proposed SBS transceiver was compared with the unidirectional transceiver designed as 4x-buffers. The average power consumption for a pair of transceivers (one channel), when transmitting and receiving the same Pseudo-Random Binary Sequence (PRBS), is given in Table 2-II. The SBS transceiver consumes approximately 27% more power compared to UDS and requires 19 minimum sized transistors per TSV. However, it may be noted that the TD samples the *Vx* at the negative edge of the clock cycle;

Table 2-II: Average Power Consumption (in μW)

	UDS	SBS	Diff (%)	SBS*	Diff* (%)
Transmitter	6.82	10.3	50.9	7.08	3.71
Receiver	2.61	1.76	-32.8	1.77	-32.37
Total	9.43	12.06	27.77	8.85	-6.28

*Transmitter turned off for the negative half cycle of the clock

therefore, one of the *TG* (hence the transmitter) can be turned off during the remaining (negative) half cycle of the clock, further reducing static power. This can be achieved with a NOR gate with $\overline{Clk}$ and $\overline{TE}$ as the inputs, whose output (and a complement, generated using an inverter) controls the *TG*. In this case, the SBS transceiver consumes approximately 6% lesser power compared to UDS (inclusive of the power consumed by the added circuit), at the expense of 6 additional transistors per channel.

TSVs are usually designed in clusters, and so, cross-coupling with the neighbouring TSVs is an essential concern in TSV communication. The performance of the SBS transceiver in the presence of noise coupling was studied, assuming a 3x3 TSV cluster with the centre TSV as the victim, as shown in Fig 2-10(a). Considering 10μm pitch between TSVs and silicon resistivity of 6.89 Ω.cm (for Na=2x10^{15}/cm^3), the values of the coupling capacitances $C_{si,ij}$ and resistances $R_{si,ij}$ of the silicon substrate between the victim TSV *i* and aggressor *j* were calculated using the coupling model described in [41], as shown in Fig 2-10(b). Assuming all the aggressors are being driven by similar SBS transceivers with different PRBS (for a 40μs simulation period), the histogram of the ternary encoded signal *Vx* at the receiver sampling time (negative clock edge) is shown in Fig 2-10(c). Given the low frequency of operation, almost all the couplings are eliminated during the positive half-cycle, before it is sampled by TD at the negative clock edge. In all cases, *Vx* remains stable at approximately *Vxm*≈0.39Vdd, *Vxh*≳0.7Vdd, and *Vxl*≲0.2Vdd. Also, as the drain current is not ideally zero in the sub-threshold regions, *Vxh* and *Vxl* will slowly tend to approach *Vdd* and *Gnd*, respectively. This explains the signal spread above *Vxh* and below *Vxl*, caused when the input of both transmitters does not change over consecutive cycles and augmented by the cross-coupling. Sufficient voltage margins exist

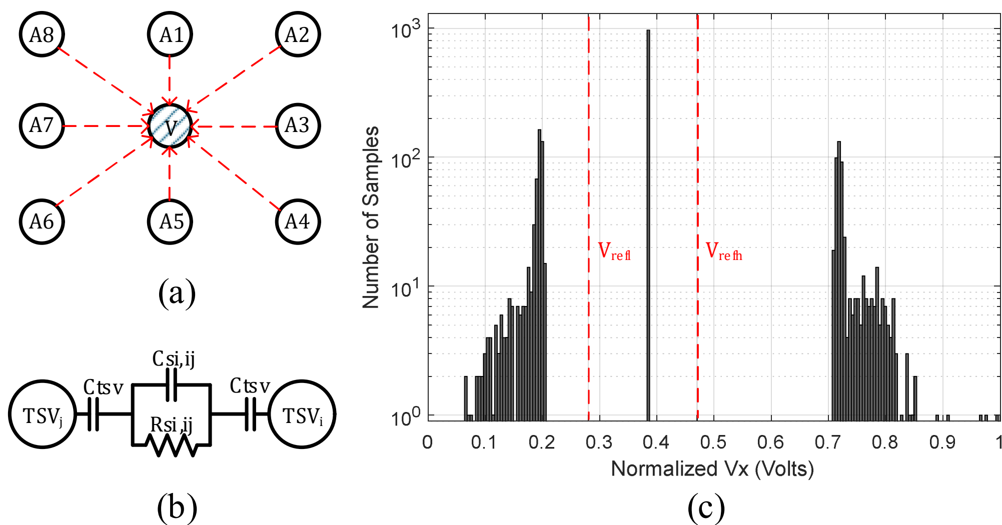


Figure 2-10: TSV Cross-coupling (a) Victim TSV (center) and 8 aggressor TSVs in a 3x3 Cluster (b) TSV-TSV Coupling model [41] **(c) Histogram of Vx voltage levels at the receiver sampling time, under cross-coupling.**

between *Vref* and *Vx* to account for the process variation (TD requires as little as 40mV difference between *Vx* and *Vref*). The transceiver operation was also verified across all process corners in the presence of noise coupling.

2.7 Conclusion

In this paper, a test accessibility architecture based on ternary encoded Simultaneous Bi-Directional Signaling (SBS), intended for use in parallel Test Access Mechanism (TAM) in System on Chip (SoC) based designs, is proposed. This method enables chip terminals to simultaneously send and receive test vectors, effectively doubling the per-pin efficiency during testing allowing additional parallelism for test time reduction. At the logic level, design considerations for incorporating SBS into PTAM while allowing functional mode utilization of chip terminal and co-existence with conventional uni-directional Signaling based DFT resources were presented. At the circuit level, a power-efficient SBS transceiver design suitable for the low-frequency operation was presented. The electrical design was validated using transient analysis using industrial chip design and simulation tools. Performance under cross-coupling was validated, and a comparison in terms of power consumption was made with the baseline UDS approach. Results suggest that the proposed transceiver consumes 27.7% more power than conventional Uni-Directional Signaling and requires 19 additional transistors, which is considered reasonable given the

potential halving of the test time. The transceiver's power consumption can be lowered by 33% using a transmitter control mechanism at the cost of additional control circuitry.

2.8 References

[1] I. A. Soomro, M. Samie, and I. K. Jennions, “Test Time Reduction of 3-D Stacked ICs Using Ternary Coded Simultaneous Bidirectional Signaling in Parallel Test Ports,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 39, no. 12, pp. 5225–5237, 2020.

[2] J. Knechtel, O. Sinanoglu, I. (Abe) M. Elfadel, J. Lienig, and C. C. N. Sze, “Large-Scale 3D Chips: Challenges and Solutions for Design Automation, Testing, and Trustworthy Integration,” *IPSJ Trans. Syst. LSI Des. Methodol.*, vol. 10, no. 0, pp. 45–62, 2017.

[3] R. S. Patti, “Three-dimensional integrated circuits and the future of system-on-chip designs,” *Proc. IEEE*, vol. 94, no. 6, pp. 1214–1224, 2006.

[4] J. P. Gambino, S. A. Adderly, and J. U. Knickerbocker, “An overview of through-silicon-via technology and manufacturing challenges,” *Microelectron. Eng.*, vol. 135, pp. 73–106, 2015.

[5] International Technology Roadmap for Semiconductors (ITRS), “ITRS 2.0 HETEROGENEOUS INTEGRATION CHAPTER: 2015,” 2015.

[6] “IEEE Standard for Test Access Port and Boundary-Scan Architecture,” *IEEE Std 1149.1-2013 (Revision IEEE Std 1149.1-2001)*, pp. 1–444, 2013.

[7] “IEEE standard for access and control of instrumentation embedded within a semiconductor device - 1687,” *IEEE Std 1687-2014*, pp. 1–283, 2014.

[8] “IEEE Standard Testability Method for Embedded Core-based Integrated Circuits,” *IEEE Std 1500-2005*, pp. 1–136, 2005.

[9] E. J. Marinissen, T. McLaurin, and Hailong Jiao, “IEEE Std P1838: DfT standard-under-development for 2.5D-, 3D-, and 5.5D-SICs,” in *2016 21th IEEE European Test Symposium (ETS)*, 2016, no. 1, pp. 1–10.

[10] B. SenGupta, D. Nikolov, A. Dash, and E. Larsson, “Test Flow Selection for Stacked Integrated Circuits,” *J. Electron. Test. Theory Appl.*, vol. 35, no. 4, pp. 425–440, 2019.

[11] M. Taouil, S. Hamdioui, and E. J. Marinissen, “Quality versus cost analysis for 3D stacked ICs,” *Proc. IEEE VLSI Test Symp.*, vol. 1, 2014.

[12] M. Agrawal and K. Chakrabarty, “Test-cost modeling and optimal test-flow selection of 3-D-stacked ICs,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 34, no. 9, pp. 1523–1536, 2015.

[13] M. Taouil, S. Hamdioui, K. Beenakker, and E. J. Marinissen, “Test cost

analysis for 3D Die-to-Wafer stacking," *Proc. Asian Test Symp.*, pp. 435–441, 2010.

[14] N. A. Touba, "Survey of Test Vector Compression Techniques," *IEEE Des. Test Comput.*, vol. 23, no. 4, pp. 294–303, Apr. 2006.

[15] A. Chandra and K. Chakrabarty, "Test data compression and decompression based on internal scan chains and Golomb coding," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 21, no. 6, pp. 715–722, 2002.

[16] A. Jas, J. Ghosh-Dastidar, and N. A. Touba, "Scan vector compression/decompression using statistical coding," in *Proceedings 17th IEEE VLSI Test Symposium (Cat. No.PR00146)*, pp. 114–120.

[17] A. Chandra and K. Chakrabarty, "Analysis of test application time for test data compression methods based on compression codes," *J. Electron. Test. Theory Appl.*, vol. 20, no. 2, pp. 199–212, 2004.

[18] V. Sheshadri, V. D. Agrawal, and P. Agrawal, "Power-Aware Optimization of SoC Test Schedules Using Voltage and Frequency Scaling," *J. Electron. Test. Theory Appl.*, vol. 33, no. 2, pp. 171–187, 2017.

[19] C. Richard M, K. K. Saluja, and V. D. Agrawal, "Scheduling Tests for VLSI Systems Under Power Constraints," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 5, no. 2, pp. 175–185, 1997.

[20] A. Sehgal, V. Iyengar, and K. Chakrabarty, "SOC test planning using virtual test access architectures," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 12, no. 12, pp. 1263–1276, Dec. 2004.

[21] M. A. Ansari, J. Jung, D. Kim, and S. Park, "Time-Multiplexed 1687-Network for Test Cost Reduction," *IEEE Trans. Comput. -Aided Des. Integr. Circuits Syst.*, vol. 37, no. 8, pp. 1681–1691, Aug. 2018.

[22] P. Georgiou, F. Vartziotis, X. Kavousianos, and K. Chakrabarty, "Testing 3D-SoCs Using 2-D Time-Division Multiplexing," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 37, no. 12, pp. 3177–3185, Dec. 2018.

[23] A. Sanghani, B. Yang, K. Natarajan, and C. Liu, "Design and implementation of a time-division multiplexing scan architecture using serializer and deserializer in GPU chips," in *Proceedings of the IEEE VLSI Test Symposium*, 2011, pp. 219–224.

[24] B. Li and V. D. Agrawal, "Applications of Mixed-Signal Technology in Digital Testing," *J. Electron. Test. Theory Appl.*, vol. 32, no. 2, pp. 209–225, 2016.

[25] K. Lam, L. R. Dennison, and W. J. Dally, "Simultaneous bidirectional signalling for IC systems," in *Proceedings., 1990 IEEE International Conference on Computer Design: VLSI in Computers and Processors*, 1990, pp. 430–433.

[26] R. Mooney, C. Dike, and S. Borkar, “A 900 Mb/s Bidirectional Signaling Scheme,” *IEEE J. Solid-State Circuits*, vol. 30, no. 12, pp. 1538–1543, 1995.

[27] R. J. Drost and B. A. Wooley, “An 8-Gb/s/pin simultaneously bidirectional transceiver in 0.35-um CMOS,” *IEEE J. Solid-State Circuits*, vol. 39, no. 11, pp. 1894–1908, Nov. 2004.

[28] H.-Y. Huang and R.-I. Pu, “Differential bidirectional transceiver for on-chip long wires,” *Microelectronics J.*, vol. 42, no. 11, pp. 1208–1215, Nov. 2011.

[29] P. Vijaya Sankara Rao and P. Mandal, “Current-mode full-duplex (CMFD) signaling for high-speed chip-to-chip interconnect,” *Microelectronics J.*, vol. 42, no. 7, pp. 957–965, Jul. 2011.

[30] N. Wary and P. Mandal, “Current-mode simultaneous bidirectional transceiver for on-chip global interconnects,” in *2015 6th Asia Symposium on Quality Electronic Design (ASQED)*, 2015, pp. 19–24.

[31] S. Park, A. Wang, U. Ko, L.-S. Peh, and A. P. Chandrakasan, “Enabling Simultaneously Bi-Directional TSV Signaling for Energy and Area Efficient 3D-ICs,” in *Proceedings of the 2016 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2016, pp. 163–168.

[32] E. J. Marinissen, S. K. Goel, and M. Lousberg, “Wrapper design for embedded core test,” in *Proceedings International Test Conference 2000 (IEEE Cat. No.00CH37159)*, 2000, pp. 911–920.

[33] B. Noia and K. Chakrabarty, “Pre-Bond Probing of Through-Silicon Vias in 3-D Stacked ICs,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 32, no. 4, pp. 547–558, Apr. 2013.

[34] B. Noia, S. Panth, K. Chakrabarty, and Sung Kyu Lim, “Scan Test of Die Logic in 3-D ICs Using TSV Probing,” *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 23, no. 2, pp. 317–330, Feb. 2015.

[35] D. L. Lewis and H. H. S. Lee, “A scanisland based design enabling prebond testability in die-stacked microprocessors,” in *2007 IEEE International Test Conference*, 2007, pp. 1–8.

[36] N. Wary and P. Mandal, “Current-Mode Full-Duplex Transceiver for Lossy On-Chip Global Interconnects,” *IEEE J. Solid-State Circuits*, vol. 52, no. 8, pp. 2026–2037, Aug. 2017.

[37] Y. Tomita, H. Tamura, M. Kibune, J. Ogawa, K. Gotoh, and T. Kuroda, “A 20-Gb/s Simultaneous Bidirectional Transceiver Using a Resistor-Transconductor Hybrid in 0.11-u CMOS,” *IEEE J. Solid-State Circuits*, vol. 42, no. 3, pp. 627–636, Mar. 2007.

[38] M. T. L. Aung, E. Lim, T. Yoshikawa, and T. T.-H. Kim, “Design of Simultaneous Bi-Directional Transceivers Utilizing Capacitive Coupling for 3DICs in Face-to-Face Configuration,” *IEEE J. Emerg. Sel. Top. Circuits*

Syst., vol. 2, no. 2, pp. 257–265, Jun. 2012.

[39] Jae-Yoon Sim, Young-Soo Sohn, Seung-Chan Heo, Hong-June Park, and Soo-In Cho, “A 1-Gb/s bidirectional I/O buffer using the current-mode scheme,” *IEEE J. Solid-State Circuits*, vol. 34, no. 4, pp. 529–535, Apr. 1999.

[40] G. Katti, M. Stucchi, K. De Meyer, and W. Dehaene, “Electrical Modeling and Characterization of Through Silicon via for Three-Dimensional ICs,” *IEEE Trans. Electron Devices*, vol. 57, no. 1, pp. 256–262, Jan. 2010.

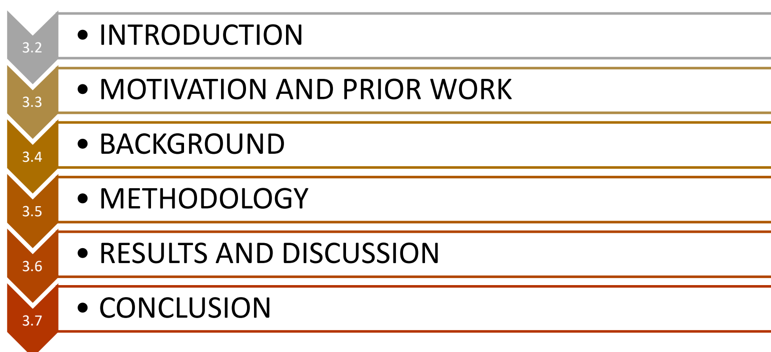
[41] T. Song, C. Liu, Y. Peng, and S. K. Lim, “Full-Chip Signal Integrity Analysis and Optimization of 3-D ICs,” *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 24, no. 5, pp. 1636–1648, May 2016.

3 REDUCED PIN-COUNT TEST STRATEGY FOR 3D STACKED ICS USING SIMULTANEOUS BI-DIRECTIONAL SIGNALING BASED TIME DIVISION MULTIPLEXING

3.1 Abstract

In Chapter 2, the electrical design considerations for use in Parallel Test Access Mechanism were discussed. As mentioned in Chapter 1 (section 1.4.3), Reduced Pin Count Testing (RPCT) is another frequently employed approach for test time reduction. In this chapter, the design considerations for integrating SBS in a 3D SIC test scenario based on RPCT are discussed, mainly focusing on a use-case based on Time Division Multiplexing (TDM). The contents of this chapter have been peer-reviewed and published in IEEE Access journal [1].

Section 3.2 introduces the problem of SBS integration with FPCT in 3D SICs, followed by a discussion into the prior art in section 3.3. In section 3.4, the working principles of TDM and SBS are illustrated with a particular focus on scan-based testing. In section 3.5, the SBS and TDM integration methodology for testing 3D SICs is presented. Section 3.6 discusses the results and section 3.7 concludes this chapter.



3.2 INTRODUCTION

The transistor density in 2D integrated circuits has been exponentially increasing following Moore's law over the past several decades; however, shrinking technology further is now proving difficult due to thermal and power constraints [2]. A promising way forward is to stack the individual dies vertically in the third dimension creating a single package known as 3D Stacked Integrated Circuit

(SIC). 3D SICs overcome the problems of increasing interconnect path delays and offer higher performance with a much smaller footprint. This concept has been applied successfully to manufacture processors [3][4] and memories [5].

One of the key enablers allowing 3D stacking of dies is the vertical interconnects between the dies, known as Through Silicon Vias (TSVs). However, TSVs bring about additional challenges for testing [6]. First, TSVs occupy a significant chip area, and therefore, there is a limit on the TSVs that could be included in the design, and even more so for testing purposes. The limited number of TSVs reduces the test channel width available for transportation of test vectors to and from the tester. Moreover, the increased number of dies means higher transistor count and an increased number of test patterns now need to be applied, and unlike 2D ICs, not just once but at several instances during stacking of the dies, such as pre- mid- and post-bond testing. It is evident that as the test vector volume and the number of test instances increase, the test access bottleneck caused by TSVs becomes significant.

Nevertheless, TSVs allow very high bandwidth owing to smaller channel resistance [7]. However, the same cannot be utilized in testing as the scan chains restrict the maximum shift frequency. The Flip-Flops, in the design, are converted to scan-enabled flops after the functional front-end design to enable scan-based tests. Consequently, the Flip-Flops, now concatenated as shift registers forming a scan-chain, are not optimized for timing. Because of these timing constraints and the thermal and power constraints associated with higher switching activity during testing, the maximum shift frequency of the scan-chain is restricted to a few tens of MHz and results in under-utilization of channel bandwidth and increased test times. One solution is to send in the test data at a high frequency and incorporate a mechanism to distribute the data among multiple scan chains at a lower frequency, such as by using Serializer-Deserializer (SerDes) or Time Division Multiplexing (TDM). By utilizing the full channel bandwidth, fewer pins or TSVs would be required and is termed as Reduced Pin Count Testing (RPCT) [8]. RPCT also results in a reduced number of test equipment channels needed

for testing a device, and the spare channels can be utilized to test multiple devices in parallel, also known as Multi-Site Testing [9][10].

This paper proposes a TDM based RPCT technique coupled with simultaneous Bi-Directional Signaling (SBS). In contrast with the conventional TDM based technique, which uses a communication channel in a uni-directional fashion by either sending or receiving data at a particular time, SBS allows simultaneous transmission and reception of the data in a channel. The advantage of such a technique could be a reduction in the number of Chip Terminal (CTs) needed for the required number of test channels for a given Test Access Mechanism (TAM) or a decrease in test time for a given number of test channels.

The rest of the paper is organized as follows. In section 3-3, the motivation for this work and the prior work in TAM design methods using RPCT techniques and SBS is presented. In section 3-4, we introduce the principle of operation of TDM and SBS. SBS-TDM based test strategy for 3D SIC is presented in section 3-5. Evaluation results of the proposed SBS-based method versus the UDS-based TDM test method are presented in section 3-6, followed by the conclusion in section 3-7.

3.3 MOTIVATION AND PRIOR WORK

Testing is of vital importance as it ensures defect-free and reliable devices. However, with the ever-advancing transistor density, the test times and hence the cost of chips increase significantly [11], opening the focus of significant research in such areas. The most commonly used Design for Test (DfT) strategy is scan-based testing, which involves shifting-in the test stimuli vectors serially, applying the test stimuli, and scanning out of the response vectors. The test application points are the memory elements or the flip-flops in the design, which are modified such that these are observable and controllable in test mode and are termed as scan-flops. The scan-flops are then concatenated into serial shift-registers, known as scan chains, such that the test data could be sequentially scanned in and out using the core's primary inputs and outputs. The stimuli could be propagated through the intermediate combinational logic, and the response could be read out for comparison with the expected response. With millions of flip-flops

expected in modern core-based design, a single serial scan mechanism such as an IEEE 1149.1 standard compliant JTAG port [12] may not be suitable in terms of test times. Parallel test ports, such as Wrapper Parallel Ports (WPP) of the IEEE 1500 Standard [13], allow the use of multiple serial scan channels by temporarily using the I/Os for test purposes. The test standard relying on similar serial/parallel test access mechanisms is also introduced for 3D-SICs [14].

Despite using the parallel test ports during testing, the exponentially increasing transistor density and the limited number of chip terminals do not allow all the scan chains to be accessed at once. Therefore, the test process is segmented into sessions, and during each session, only a limited number of cores are accessed and tested. A number of test patterns are scanned in one bit at a time, making the entire process significantly time-consuming and costly. In general, the test time increases with the test data volume (and hence the chip complexity) and decreases with the available channel width for parallel test ports and the scan shift frequency. Test compression methods have been frequently employed to reduce the test data volume [15]. However, this method requires additional on-chip resources such as decompressors and compactors. Also, beyond a certain point, test compression reduces the test coverage. Increasing the available channel width for parallel test ports significantly reduces the test time [16][17], but the bottleneck, in this case, is the limited number of chip terminals and TSVs in the design. Increasing the scan shift frequency offers a proportional reduction in test times; however, the most critical limiting factor, in this case, are the scan-chains that are not optimized for timing and operate on low frequency. This results in the loss of usable tester and the chip terminal/TSV bandwidth which are capable of much higher speeds.

RPCT based methods have two significant advantages. First, it allows optimal utilization of the tester and I/Os bandwidths, and secondly, it uses fewer chip terminals than Full Pin Count Testing. Additionally, in 3D-SICs, the wafers are thinned to expose the TSVs hidden in the substrate [6][18], and the thinned die may not be able to withstand the forces exerted by the tester probes; and therefore, only limited test channels may be available. Hence, RPCT naturally

lends itself to testing 3D-SICs. Techniques such as SerDes and TDM, which increase the scan frequency while addressing the scan chains' low frequency, have been frequently employed to achieve RPCT.

A TDM based access mechanism for serial Reconfigurable Scan Networks (RSNs), such as those based on IEEE 1687 standard (aka iJTAG) [19], was proposed in [20]. Unlike traditional scan design, an RSN allows dynamic scan path reduction as and when required. However, using a single serial access interface limits the practicability of such an approach for high volume test vector transportation. The future SICs are expected to contain millions of flip-flops necessitating high bandwidth scan-testing employing Parallel Test Ports. The authors in [21] and [22] highlighted the notion of using virtual TAMs and Serializer and De-serializer (SerDes). In [23], the authors discuss a combination of test vector compression and RPCT to reduce test times. The authors in [24] proposed using Multi-Valued Logic (MVL) for tester-to-chip communication. The use of MVL increases the data rate for a given clock frequency; however, they necessitate analog to digital converter with calibration schemes which may complicate the implementation and add significant chip area.

Much of the above work has been focused on 2D chips, and most these methods are applicable to 3D designs as well. Nonetheless, the implementation is not straightforward, and 3D technology-specific concerns must be considered. The Test Access Mechanism (TAM) design, which refers to the insertion of required logic between the chip's primary I/Os and the individual cores to enable test vector transport, is known to be an NP-Hard combinatorial problem. The design choice of a TAM such that the test requirements of all the dies in 3D SIC are met using limited resources significantly affects the test time. Several researchers focus on test time reduction by optimal TAM design for 3D SICs [25][26]. The authors in [27][28] proposed a TDM based RPCT test strategy in 3D-SICs. In [28], the authors showed significant improvement in test times using TDM compared to conventional TAM design methodology. The authors defined 'global channels' as communication channels traversing vertically through multiple TSVs/dies, as opposed to point-to-point communication channels (the same definition of 'global

channels' will be used in this article). The global channels were operated at a higher frequency, and the data were multiplexed to dies, and in turn to the cores at lower frequencies.

The focus of the previous research work in testing and test time reduction has been on conventional UDS signaling. This simplifies the design as the standard I/O cells can be utilized; however, Simultaneous Bidirectional Signaling (SBS) significantly improves the throughput of communication channels, significantly impacting test time. The author in [29] briefly discussed the dynamics of using SBS in chip testing and its future potential. The research in SBS has been mostly focused on the normal mode (as opposed to test mode), point to point communication links, focusing on throughput and power efficiency. Following the initial concept of SBS by the authors in [30], the researchers have proposed various methods to enable SBS. Broadly, the SBS design methodology can be classified into differential [31][32] or single-ended designs [30][33][34][35], current mode [31][32] or voltage mode transceivers [34][30][35], or based on channel characteristics such as on-chip [31][36] or off-chip communication [30][33]. 3D SICs offer a promising prospect for SBS due to the low channel impedance of the TSVs. The authors in [37][38] proposed single-ended SBS design methodologies for use in 3D SICs and reported significant improvement in chip area, throughput, and power consumption compared to 2 x UDS-based TSVs.

While both TDM/ SerDes and SBS methods improve the efficiency of a communication channel, the key difference is that the former is aimed at maximizing the use of available channel bandwidth, whereas the latter allows using a single pin to form the communication channel. The previous works have focused on using these methods separately in the chip testing scenario, with SBS used to decrease the test time [39], and TDM/ SerDes used to minimize test channels (pins/ TSVs) [28]. Nevertheless, a combination of both methods presents new prospects in the 3D SIC test, allowing minimization of test resources as well as test time reduction. This, however, presents several challenges. Firstly, the previous research in SBS has been focused on

conventional point-to-point communication, which simplifies the implementation. In this scenario, communication is required between a high-frequency source at the near-end (Scan-in vectors from the tester) and several far-end transmitters operating at much slower speeds (Scan-out vectors from the scan chains), for which the design is not straightforward. Secondly, unlike [39], where SBS was used in the Full Pin Count Test (FPCT) scenario with the low-frequency operation, the design considerations in this case are complicated by high-performance requirement.

This paper explores the feasibility of integrating SBS with TDM based RPCT method, with a particular focus on its application in testing 3D SICs. We present a potential SBS transceiver design capable of high-frequency operation and evaluate the design tradeoffs. The design challenges and possible solution of integrating SBS with 2D-TDM based Test Access Mechanism, which requires a global channel traversing through multiple TSVs, are studied. Moreover, the strategy to extend TDM-SBS based test methodology to pre- mid- and post-bond test instances of 3D SICs is presented.

3.4 BACKGROUND

This section describes the operation of TDM based test methodology, followed by an overview of Simultaneous Bidirectional Signaling. The working concept is illustrated using example cases that mainly focus on scan-based test architectures. The same examples will be subsequently used as the test cases for evaluation.

3.4.1 Time Division Multiplexing

Consider an example 3D-SIC with three dies, as shown in Fig. 3-1. It is assumed that every die is composed of 2 cores with a total of 3 scan chains and that all dies are identical (details are only shown for the first die for clarity). In general, the number of scan chains is far greater than the available number of test channels and depending on the test schedule, only a small subset of cores are

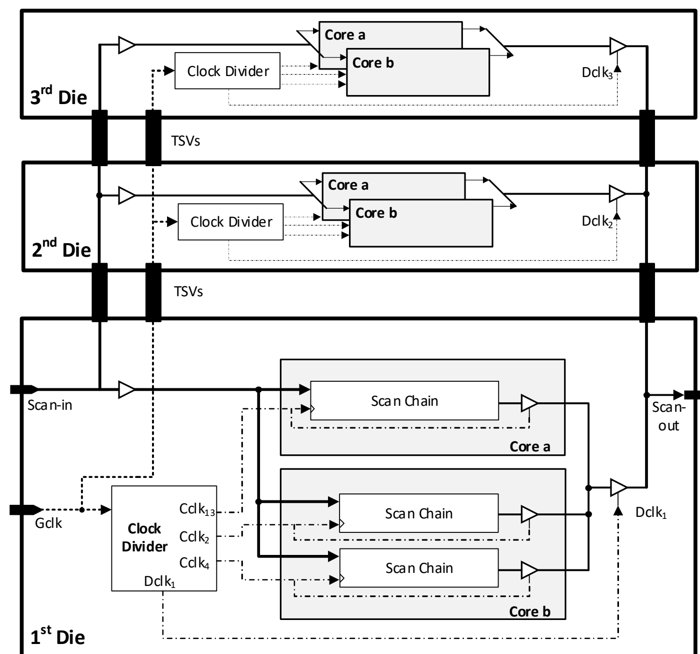


Figure 3-1: An example implementation of TDM for scan-test application based on [28]**.**

selected at a time. Therefore, this example represents the case for a particular test session in the overall test schedule. A single test channel is shown, which originates at the scan-in chip terminal and terminates at the bottom die's (1st die) scan-out chip terminal. It is assumed that the tester communicates with the 1st die through these chip terminals.

The TDM design for this example is based on the method proposed in [28] in which the authors proposed using separate 2D-TDM for the vertical (inter-die) and horizontal (intra-die) communication. At the input, the incoming data is available to all the dies, and in turn, the scan chains, using the global TSV channel. The data is demultiplexed from the Scan-in pin to the cores by controlling the scan chains' clock signal, such that data is scanned-in only to the scan chain, which receives the positive clock edge. At the scan-chain output, the data is multiplexed on the scan-out pin using two tri-state buffer stages. The first stage tri-state buffers (one for every scan-chain) are controlled such that only one buffer is active at a time, essentially serving as 3 to 1 multiplexer. In the second stage, tri-state buffer (one for every die) in-turn multiplexes the data onto the scan-out pin, one die at a time.

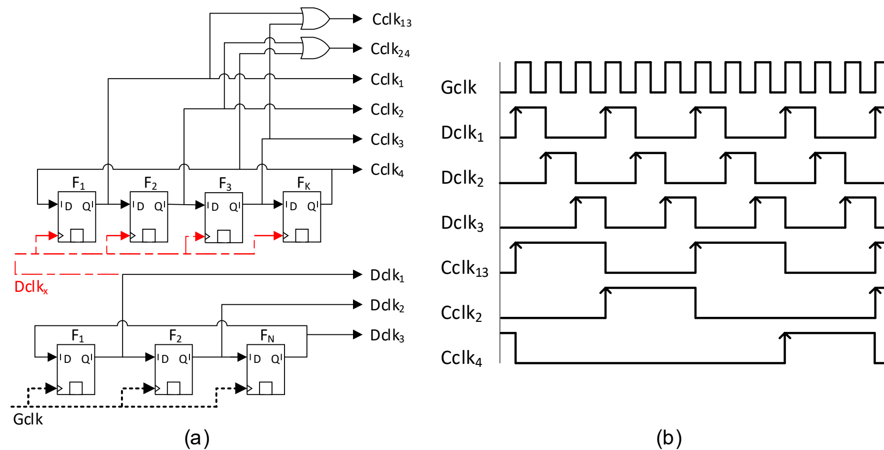


Figure 3-2: Control circuit for TDM (a) Generation of core and die clocks from global test clock using Ring Counters, the bottom RC generates die clocks, the top RC generates core clocks (2) Timing diagram of the die and core clocks

To appropriately demultiplex the data at the input and multiplex at the output, the above arrangement requires a control circuit to select the appropriate die, core, and hence the scan-chain at every clock cycle. This is achieved using a clock divider circuit that can be constructed using shift registers as a Ring Counter (RC). Fig. 3-2 (a) shows one implementation of the clock divider circuit. The incoming Global Clock signal (*Gclk*) is first divided to generate Die clock (*Dclk*). As there are three dies in this example, a 3-bit RC is used to generate three 120º out of phase die clocks ($Dclk_{1,2,3}$) running at 1/3rd of the *Gclk* frequency. To ensure this, the RC is initialized with a one-hot bit sequence (with only one bit set high - 100-bit sequence used in this case). The die clock serves two purposes; First, it allows multiplexing the data to the Scan-out pin by activating the second stage tri-state buffer of only one die at a time (Fig. 3-1). Secondly, it is used to derive the Core Clock (*Cclk*) signals for the individual cores/scan chains, as shown in Fig 3-2(a). The *Cclks* serve two other purposes, first, they allow demultiplexing the data from the scan-in terminal to the scan chains, and secondly, they activate the first stage tri-state buffer of the cores to be multiplexed at the output.

The choice of the number of flip-flops for *Cclk* generation would determine the lowest achievable frequency (*fmin*) and the number of scan chains serviceable by the *Cclk*. For *k* flip-flops, a minimum frequency of *Dclk/k* can be achieved, and at most, *k* scan-chain can be multiplexed. Multiples of *fmin* can also be produced by OR-gating alternate *Cclk* outputs, as shown in Fig. 3-2(a). In this way, different frequency clocks can be provided to scan-chains depending on the scan-

frequency supported by the core. For example, in Fig 3-1, the $Cclk_{13},$ which is twice the minimum core clock frequency (*Dclk/k),* is used to serve the scan-chain chain in the *core a* of the dies. Fig 3-2 (b) shows the timing diagram of the derived clock signals. The clock division using this arrangement produces a duty cycle with on-time equal to the clock period of the *Gclk*, ensuring the second stage tristate buffers are 'on' for the entire *Gclk* cycle.

3.4.2 Simultaneous Bi-directional Signaling

Conventional TAM design using TDM requires separate output and an input port to form a test channel, as shown in Fig. 3-3(a). The proposed TAM design is based on Simultaneous Bi-Directional Signaling (SBS) in which a channel could be formed using one pin only, as illustrated in Fig. 3-3(b). SBS is different from traditional bi-directional pins in that the latter can only be configured either as an input or an output at a given time (half-duplex) and is therefore considered to be UDS for test purposes.

The working principle of SBS is elaborated using an example die consisting of a single scan chain of two-bit length, as illustrated in Fig. 3-3(c). Using a conventional UDS scheme, two pins would be required, one to connect the input of the scan-chain to the Scan-In (*SI*) signal and another pin to connect the output of the scan-chain Scan-Out (*SO*) (Fig. 3-3a). However, using SBS, a single pin could transmit and receive *SO* and *SI* simultaneously. To achieve the same, the signal at the chip terminal is ternary encoded instead of binary. The scan-chain SO's output is fed to a transmitter *Tx1*, which could be designed as a buffer. The *Tx1* (with an output impedance of *R1*) drives the chip terminal from one end, whereas a similar transmitter *Tx2* (with output impedance *R2*) is assumed to be driving the same chip terminal with the *SI* signal from another die (In Fig. 3-3(c) the *R1* and *R2* depict transmitter internal impedance but are shown as external resistors for clarity). Depending on the state of the *SI* and *SO*, the voltage *Vx* at the chip terminal node will either be pulled low (0 V) or high (Vdd) in the case when both ends are being driven low or high, respectively; however, *Vx* will take on an intermediate value (*Vxm*) when both transmitters are in the opposite state

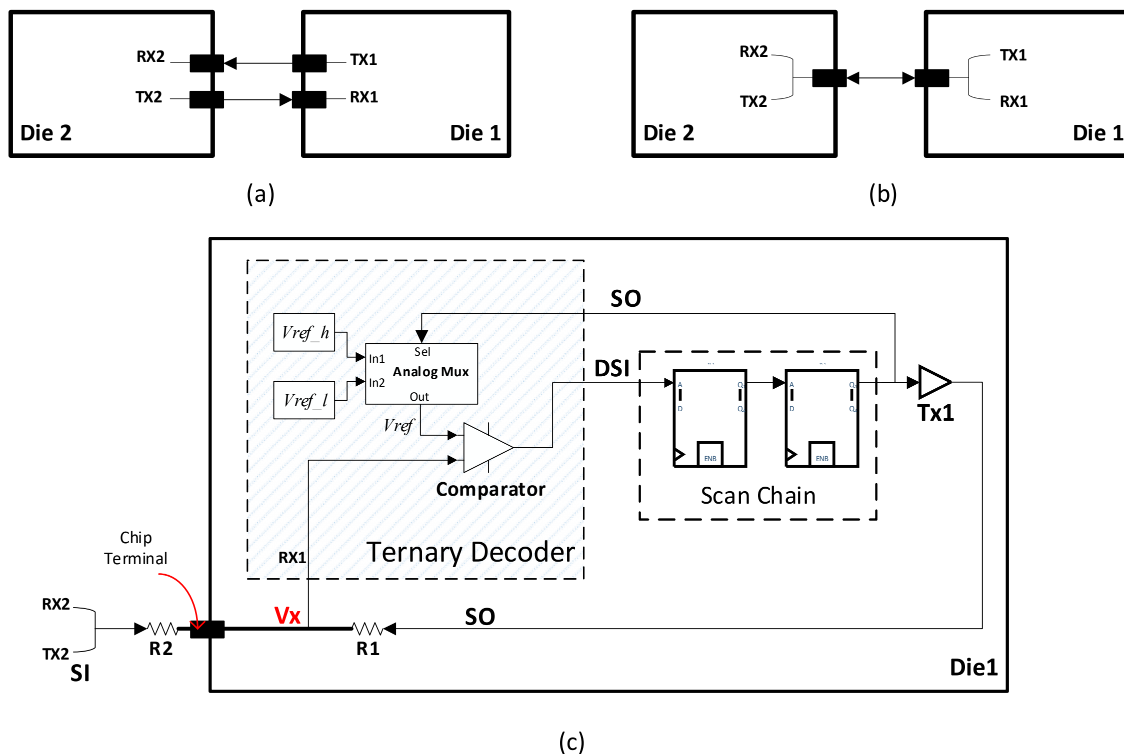


Figure 3-3: An illustration of Simultaneous Bi-directional Signaling (a) Conventional Uni-directional signaling using two wires (b) Simultaneous Bi-directional Signaling using one wire (c) Block diagram of SBS working principle – Test channel formation for a two-bit scan-chain

(10 or 01). The value of *Vxm* will depend on the impedances *R1* and *R2*, and assuming both to be the same, *Vxm* equates to ½Vdd.

The ternary encoded signal *Vx* at the chip terminal can now be used by each die to determine whether the incoming signal (*SI*) is the same or opposite of the signal (*SO*) being transmitted. The Ternary Decoder (TD) block shown in Fig. 3-3(c) receives the *Vx* signal and the *SO* signal (taken just before *Tx1*). To determine the *Vx* signal state, two reference voltages are required, a high reference voltage *Vrefh* that is midway between Vdd and *Vxm*, and a low reference voltage *Vrefl* halfway between *Vxm* and the ground. An analog multiplexer is employed, selecting *Vrefh* if *SO* is high and *Vrefl* when *SO* is low. A voltage comparator circuit is used to compare the *Vx* signal with the *Vref* (reference voltage from the Analog Multiplexer). Table 3-I lists all possible SI and SO values and the state of the transceiver in each case. When *SO* is 1, *Vx* can only take on the value of 1 (Vdd) or *Vxm* (½*Vdd). A high value at *Vx* implies that *SI* must also be high, whereas a ½*Vdd voltage level indicates *SI* must be zero.

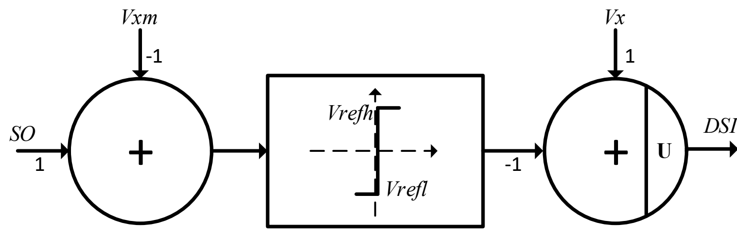


Figure 3-4: Neural Network presentation of the ternary decoder.

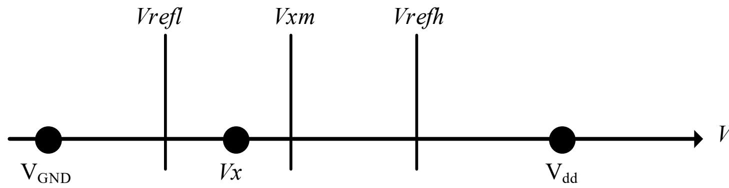


Figure 3-5: Hyperplanes created by the neurons in Figure 3-4.

Table 3-I: SBS Transceiver States

SO	SI	Vx	Vref	Vx>Vref	DSI
0	0	0	Vrefl	False	0
0	1	½*Vdd	Vrefl	True	1
1	0	½*Vdd	Vrefh	False	0
1	1	1	Vrefh	True	1

The comparator determines the same by comparing *Vx* with *Vref* from the Analog Multiplexer, which in this case (*SO*=1) would be *Vrefh*. Similarly, When *SO* is low, the comparator receives the lower reference *Vrefl* and compares it with *Vx,* which could either be low (meaning *SI*=0) or ½*Vdd (meaning *SI*=1). In all cases, the Decoded Scan-In (*DSI*) Signal produced by the comparator (and hence the TD) is the same as the original *SI*, as shown in Table 3-I.

A mathematical presentation of the ternary decoder can be constructed using a neural network shown in Fig. 3-4. The output DSI can be represented with (1), where *U* is the unit step function:

$$DSI = U\left(V_x - \left[V_{refl}U(V_{xm} - SO) + V_{refh}U(SO - V_{xm})\right]\right) \quad (1)$$

V_{GND} and V_{dd} in Fig. 3-5 are constant values, while Vx varies depending on the *SO* and *SI* voltages. The position of Vx is determined by the superposition of *SO*

and *SI* applied to the resistive network comprised from *R1* and *R2*, calculated using the following equation:

$$V_x = SI\,\frac{R_2}{R_1+R_2} + SO\,\frac{R_1}{R_1+R_2} \qquad (2)$$

The neurons in the network of Fig. 3-4 generate two hyperplanes, a fixed position hyperplane of *Vxm*, and a dynamic position hyperplane of *Vref*. This initially compares the input value *Vx* with the fixed hyperplane *Vxm*, then based on the result, it triggers *Vrefl* if *Vx*<*Vxm*, otherwise *Vrefh* if *Vx*>*Vxm*. Fig. 3-5 demonstrates a case when *Vrefl* is triggered as the valid hyperplane for the second neuron because of *Vx*<*Vxm*. Finally, in accordance with Table 3-I, the network generates logic value '1' as *Vx*>*Vrefl*.

3.5 METHODOLOGY

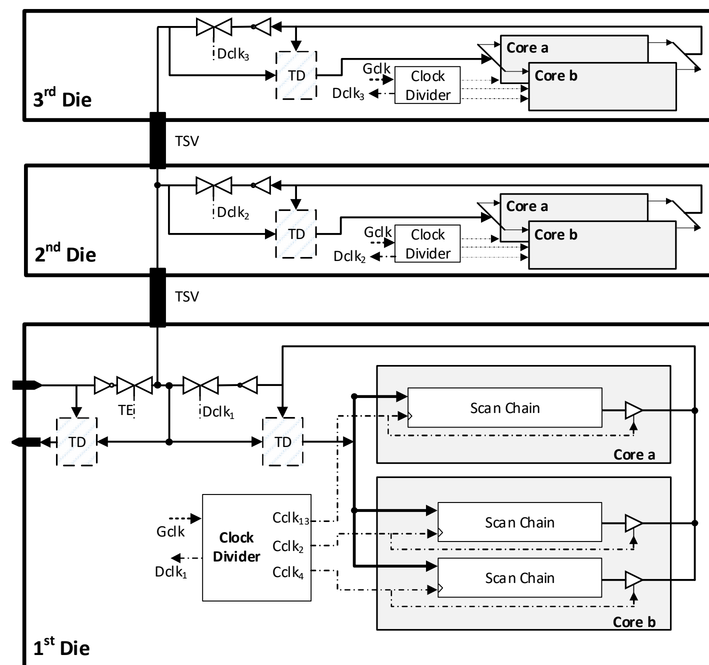


Figure 3-6: SBS-TDM based Test Access Mechanism for the example case of Figure 3-1.

This section demonstrates the feasibility of using SBS-based TDM using the test case of UDS-based TDM design presented in section III and modify it to include an SBS transceiver, as shown in Fig. 3-6. The design of an SBS transceiver is dependent on the characteristics of the channel; therefore, the design

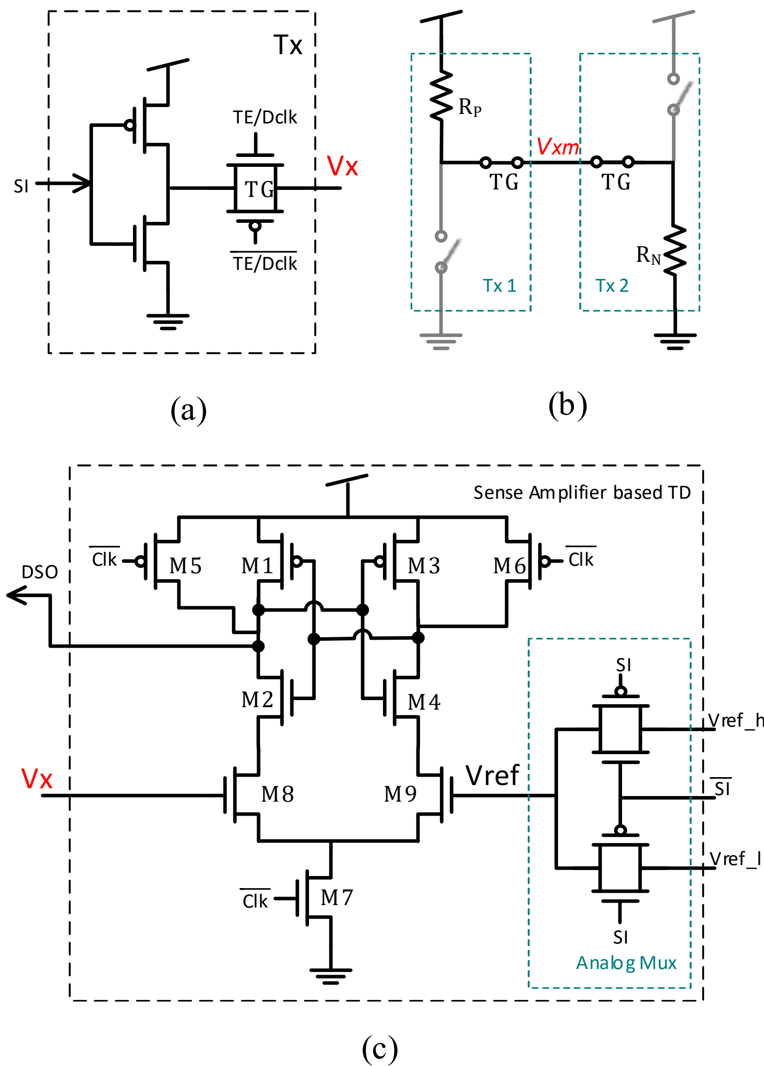


Figure 3-7: SBS transceiver implementation (a) Transmitter (b) Equivalent electrical model of the transmitter when sending opposite signals (10,01) (c) Sense-Amplifier based Ternary Decoder (Receiver)

considerations are different for the communication channel between the tester and the first Die, and for inter-die communication (using TSVs). As the TSV channel is much less resistive in nature [7], the design for an SBS transceiver is relatively simple and can be achieved using the core transistors. Therefore, to avoid complexity, we assume that tester-to-die communication is done using the existing UDS method, and we only propose SBS for inter-die communication through TSVs.

3.5.1 Transceiver Design

SBS transceivers can be implemented using several design methods. Differential mode transceivers [40] are used for high-frequency applications; however, the design is often complicated, and the requirement of two wires limits its use in

TAM design where single-ended one wire systems are preferable. Single-ended SBS transceivers for use in the TDM scheme require three main components, the transmitter, receiver, and the control circuit, including TDM switching circuitry.

The transmitter design mostly depends on two factors, the required performance and the power consumption. As the SBS transmitter involves ternary coding, the design deviates from the static CMOS logic to generate the intermediate voltage levels, resulting in high static currents. The transmitter was designed as an inverter, followed by a Transmission Gate (TG) acting as an analog switch, as shown in Fig. 3-7 (a). The TG allows turning off the transmitter when not in use; for instance, to limit static current or during functional mode operation. The TG also allows turning on and off the transmitters only during the specified intervals to allow time-division multiplexing. The transistor widths were carefully chosen to find the right balance between performance (maximum supported frequency) and acceptable power consumption. This trade-off is discussed further in section 3.6. The transmitter design in Fig. 3-7(a) is equivalent to a tri-stated inverter whose intermediate nodes between the Pull-up PMOS transistors and Pull-Down NMOS transistors are connected. While this functionality could also be achieved using a tri-state inverter, the transmission-gate based design has an advantage that during the on-state of the transmitter, the effective resistance of the transmission gate is a parallel combination of both the PMOS and NMOS; and therefore, has a lower resistance compared to either NMOS or PMOS. This arrangement reduces delay and improves performance.

The transmitter design must also account for the desired *Vxm* voltage level when the transmitters at either end are in the opposite states. Ignoring the TSV resistance and assuming the pass gate as an ideal switch, the two transmitters' equivalent electrical model when sending opposite signals (10,01) is shown in Fig. 3-7(b). If R_P and R_N denote the resistance of the PMOS and NMOS transistors of the inverter, respectively, the middle voltage level V_{xm} is given by:

$$V_{xm} = V_{dd}(R_P/(R_P + R_N)) \qquad (3)$$

Whereas in the triode region:

$$R_p = \frac{1}{\mu_p C_{ox} \frac{W_p}{L}(V_{GS} - V_t)} \qquad (4)$$

and,
$$R_n = \frac{1}{\mu_n C_{ox} \frac{W_n}{L}(V_{GS} - V_t)} \qquad (5)$$

For *Vxm*=1/2 Vdd, the factors $(V_{GS} - V_t)$ can be assumed to be similar for both PMOS and NMOS, also taking Cox and L to be the same, (3) reduces to:

$$W_p = \frac{\mu_n}{\mu_p} W_n \qquad (6)$$

Therefore, to achieve Vxm=1/2Vdd, the transistor widths ratio between PMOS and NMOS should be designed so that the on-state current of the low mobility PMOS is similar to higher mobility NMOS. Moreover, the TG transistor sizes are also chosen to be the same for the inverter, ensuring that the TG has a constant on-resistance with the same strength as the inverter.

The Ternary Decoder was designed using a sense-amplifier-based voltage comparator [41][42] and a pass-gate analog multiplexer, as shown in Fig. 3-7 (c). The sense amplifier is widely used as a voltage comparator due to its robust operation and power efficiency and is commonly used in high-performance Flash ADCs [43][44]. The pass-gate analog multiplexer selects the appropriate reference voltage depending on the outgoing signal SI, i.e. the high reference voltage *Vrefh* when SI=0 and lower reference voltage *Vrefl* when SI=1. The reference voltage is fed to one sensing input of the sense-amplifier (M9), whereas the other sensing input (M8) receives the ternary coded *Vx* signal. M8 and M9 act like variable resistors with values proportional to the respective gate voltage (*Vx* and *Vref*). The transistors M1 through M4 form two cross-coupled inverters. During the positive clock cycle, M5, M6, M2, and M4 turn on, charging the cross-coupled nodes of both inverters to Vdd. During the entire positive clock cycle, the cross-coupled transistors remain in the meta-stable state. During the negative clock half cycle, M7 turns on, providing a discharge path to the cross-coupled inverters; however, the inverters tend to discharge at different rates depending on the on-resistance, and hence the currents through M8 and M9, performing the

comparator action through regenerative feedback. The inverting output of the SA amplifier is chosen as the TD's output to reconstruct the original signal, which was inverted by the transmitter designed as an inverter.

3.5.2 Pre- and Mid-Bond Testing

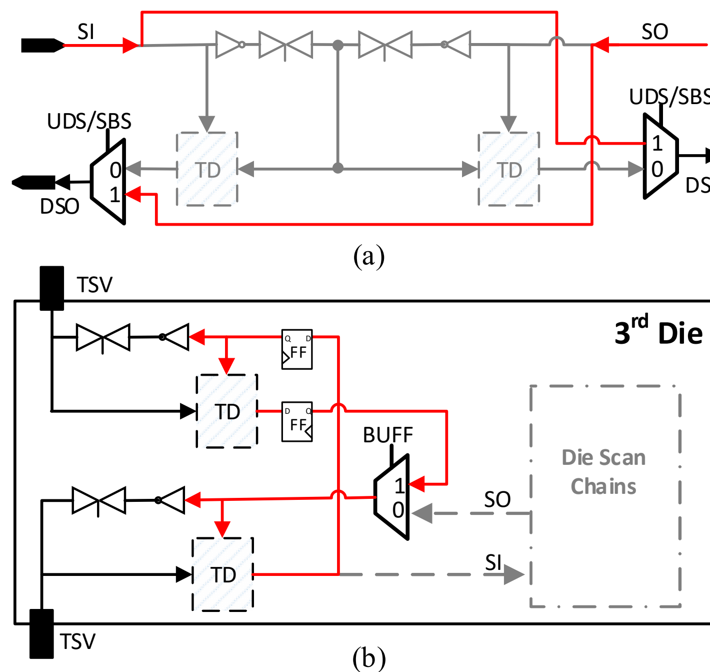


Figure 3-8: (a) Bypass method for Pre-bond testing. (b) Using SBS transceivers as buffers to access higher dies.

The implementation in Fig. 3-6 represents the case of post-bond testing in which the dies have been assumed to be already bonded. However, the dies may require testing before bonding, also known as pre-bond testing. As we have considered that the tester communicates using UDS, the pre-bond testing can be undertaken using UDS methods by bypassing the SBS transceivers. This can be achieved by multiplexing the output of the tester side and die side TDs, with the *SO* and *SI* signals, respectively, as shown in Fig. 3-8(a). The multiplexers may be configured to either select SBS or UDS using a single bit register accessible through JTAG. To minimize power consumption by the SBS transceiver when UDS is selected, the TGs can be turned off, and the TD clock could be disabled by using clock-gating; alternatively, the complete transceiver circuits could be

disabled by power-gating. Fig. 3-8(a) mainly depicts the case for the 1st die; similarly, the SBS transceivers may be bypassed for the other dies.

The mid-bond test instances are a subset of the post bond test problem. However, depending upon the number of dies stacked, the clock divider circuit may be multiplexed/ configured accordingly as the case would be in UDS based TDM. However, the addition or removal of the dies may affect the global channel's electrical characteristics, affecting the transmitters' performance and power consumption. The transmitter's performance may be adjusted by using a binary-weighted variable drive strength transmitter [30]. Depending on the requirement, the drive strength may be adjusted by enabling the desired inverters; for instance, a 3-stage transmitter would allow the adjustment of drive strength from 1x to 7x. Similar to the selection of UDS in pre-bond testing, the transmitter strength may be configured using JTAG.

3.5.3 Test Setup

To test the proposed transceiver, the example 3D stacked die and core structure shown in Fig. 3-1 was modified to include SBS transceivers, as shown in Fig. 3-6. The TSV was modelled as a lumped RC circuit [7], assuming a TSV structure with a length of 20 µm, a diameter of 5 µm, Tox (oxide thickness) of 200nm, and $2 \times 10^{15}/cm^3$ doping concentration for the substrate. The resultant RC model has a TSV capacitance of 30fF and a resistance of 100 mΩ.

To validate the test structure's output, the typical Capture, Shift, and Update cycle of the scan-chain was ignored, and continuous shifting was performed such that the output is the same as the input. However, it may be noted that unless all the scan-chains are of the same length and operate at the same frequency, the multiplexed output from the scan chains will be an interleaved form of the input. For instance, in the test structure used in this example, *core a* is being serviced using $Cclk_{1,3}$ which is twice the *fmin;* the scan sequence through this core will appear earlier at the output compared to other scan-chains. To validate the output, the correct/ expected multiplexed output, *SO(expected)*, for the TDM multiplexer can be modeled using the proposed pseudo-code as shown in Fig. 3-9.

```
    Pseudo-code
1   Input (for the current 1-bit global channel)
2     D              - The number of dies serviced by the
3   channel
4     S              - the number of scan chain per die
5     Lc             - length of cores, where c=1toS
6     x                     - vector with scan in bit pattern,
7     CycReg_die     - Die access sequence for die level clock
      CycReg_core    - Core access sequence for core clock

8
9   Generate a 3D array Scan_chains with D x S x Lc binary bins
10  Initialize clk=1, intermediate variables j=k=y=Input=sel_sc =[ ]
11  While clk = true
12    j= CycReg_core(1)                  // currently active core
13    For D dies,
14      Input=x(clk)                     // current scan-in bit
15      k= CycReg_die (end)              //currently active die
16      sel_sc =scan_chains(k, j , 1 to Lc )
17      y(clk) = sel_sc(end)
18      Right-shift sel_sc by 1 bit
19      sel_sc(1)=input
20      scan_chains(k, j , 1 to Lc )=sel_sc
21      clk++; if clk > length(x) then Break (while)
22      Right-Circular-shift CycReg_die by 1 bit
23    end For
24    Right-Circular-shift CycReg_core by 1 bit
25  end while
    Output y
```

Figure 3-9: Pseudo-code for generation of 3D TDM based TAM output for a 1-bit global channel assuming identical dies in the stack

The test setup in Fig. 3-6 has been limited to 3 dies; but, 3D SICs with any arbitrary number of dies may be tested using SBS based TDM. However, the signal integrity, when traversing multiple dies in a global TSV, must be ensured. In a UDS scheme, buffers can be inserted midway between two consecutive TSVs; in an SBS scheme, digital buffers cannot be used because of the ternary encoding. The authors in [34] discuss the use of accelerators, mid-way latches, and opposite-polarity transition encoding to improve SBS performance in highly lossy global links, and the authors in [45] propose using clamping circuits to address the same. The proposed design does not include any intermediate buffers; there will be an upper limit on the number of dies supported by the given transceiver design for a TDM global channel. The dies higher up the stack may be accessed by using SBS transceivers as buffers. For instance, for the test case in Fig. 3-6, the dies beyond the 3rd die may be accessed using SBS buffers as

shown in Fig 3-8(b). The use of this method has two implications, 1) every SBS buffer instance requires insertion of flip-flops resulting in an additional 1-bit delay in the scan shift cycle, and 2) The overall test schedule would require different sessions to test the buffered segments.

3.6 RESULTS AND DISCUSSION

Simulations were performed with Cadence Virtuoso using 45nm technology and the standard cells from 45nm Nangate Library [46], using 1v Vdd. The proposed transceiver was designed to achieve an operating frequency of 1.2GHz at the global TSV channel. The transmitter’s NMOS transistors were designed with 360nm width, and the PMOS transistors were designed as 1.6 x NMOS width giving the middle voltage level of approximately 0.5Vdd. The Sense-Amplifier was designed using 360nm NMOS width and 1.6x NMOS width for the PMOS transistors. The lower and upper reference voltages were chosen to be 300mV and 700mV, respectively, for a supply voltage of 1V. From here on, we define the scan-in side transceiver and the scan-out side transceivers as near-end and far-end transceivers, respectively. The transient simulation results for the near-end transceiver at 1.2GHz frequency are shown in Fig. 3-10. The near-end transmitter sends the scan-in (*SI*) signal, which is demultiplexed to 9 scan chains in 3 dies. Although there are 3 far-end transmitters, only one is active at a time, essentially behaving as a single transmitter, and hence the signal scan-out (*SO*) as seen by the near end transceiver is shown as *SO(expected)* and was computed using the pseudo-code in Fig. 3-9. The various intermediate signals in the TD are also shown, and the output of TD is denoted as Decoded Scan-Out (*DSO*). The *DSO* signal appears as a unipolar return to zero coded waveform due to the sense-amplifier's nature, which can be directly fed to the scan chains. The signal *SO* in Fig. 3-10 shows the first scan-flop’s reconstructed output and is similar to the *SO(Expected)* signal delayed by one cycle.

Fig. 3-11 compares the transceiver behaviour with varying frequency. The output of the transmitter (*Vx*) and the receiver (*DSO*) are shown for 0.6, 1.2, 1.8, and 2.2 GHz clocks. It may be noted that although there are 3 output states of the Vx signal (0, *Vxm,* and *Vdd*), there are six possible transitions, depending upon the

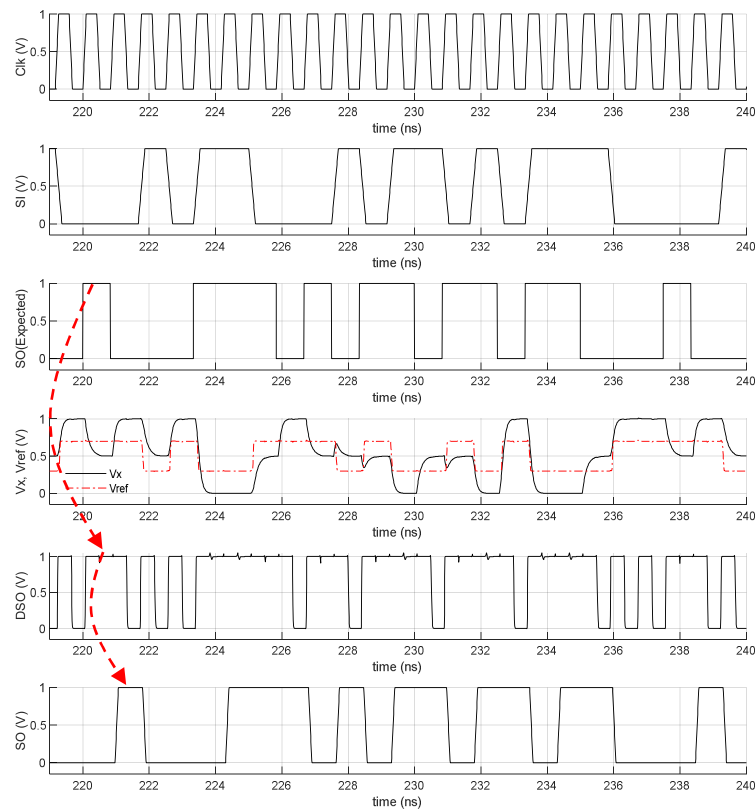


Figure 3-10: Waveforms of various signals using SBS transceiver design of Figure 3-7 in the test setup of Figure 3-6 simulated at 1.2 GHz Gclk frequency.

previous state, i.e. rail-trail (0-1,1-0), rail-mid (0-*Vxm*, 1-*Vxm),* and mid-rail (*Vxm*-1, *Vxm*-0). Fig. 3-11 shows the rail-rail and rail-mid transition, whereas the mid-rail transitions are omitted for clarity as the response is similar to rai-rail. For a given Vx transition, the dashed red markers show the reference voltage (low or high), which is dependent on SI signal being transmitted. Moreover, for ease of comparison, every interval on the horizontal axis depicts one clock cycle with waveforms for different frequencies accordingly scaled.

From Fig. 3-11, it is evident that the increasing frequency results in additional gate delay and transient time relative to the clock duration. This results in reduced timing margins and a lesser voltage difference between Vx and Vref at the TD input. While the former affects both the transmitter and receiver and is somewhat mitigated by accounting for timing pessimism, the effect of the latter is rather significant as lesser voltage margins decrease the robustness of the TD, especially in the presence of process variations and cross-coupling. It is interesting to note that although the rail-rail transitions involve a larger voltage

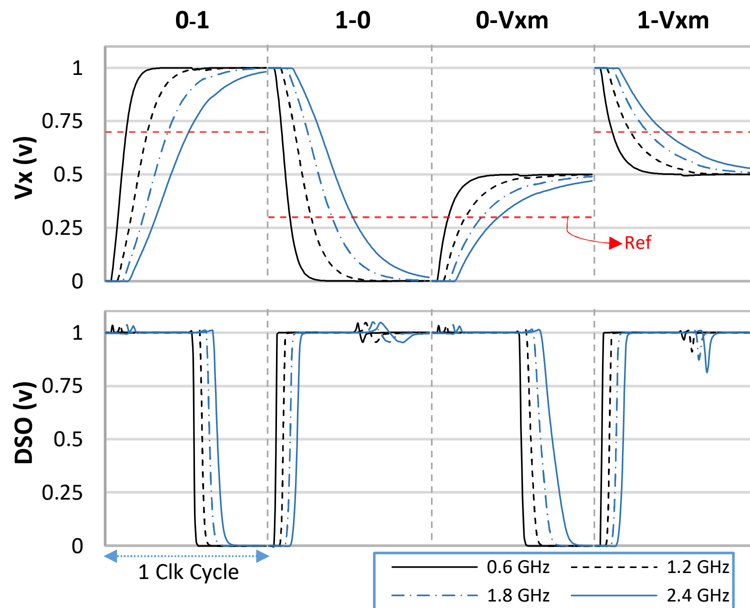


Figure 3-11: Transceiver transient response with varying frequency.

swing than the rail-mid swing, they are relatively faster (~60% compared to the rail-mid transitions). This is because the rail-rail transitions are only possible when both near- and far-end transmitters are being driven in the same direction, which effectively doubles the drive strength. For the same reason, the mid-rail transitions also exhibit similar behaviour, which may even be slightly better due to reduced voltage swing. On the other hand, there is effectively a single transmitter driving Vx for the case of rail-mid transitions, resulting in relatively slower transitions. Consequently, with increasing frequency, the transceiver performance is likely to degrade for the rail-mid transitions first; therefore, the transceiver may be designed considering the rail-mid response as the worst case.

The transceiver operation was verified across all process corners at 1.2GHz. A maximum variation of 40 mV was seen in the mid voltage level *Vxm* across various design corners. It may be noted that the *Vxm* seen by the near and far end transceivers may slightly differ further due to parasitic resistances. The minimum offset voltage (the difference between *Vref* and *Vx*) required to correctly resolve the middle voltage level *Vxm* from *Vrefl* or *Vrefh* (low or high reference voltages) levels was observed to be approximately 25mv. This gives a sufficient margin to account for voltage drops due to parasitics and additional statistical offset due to variability in the sense amplifier

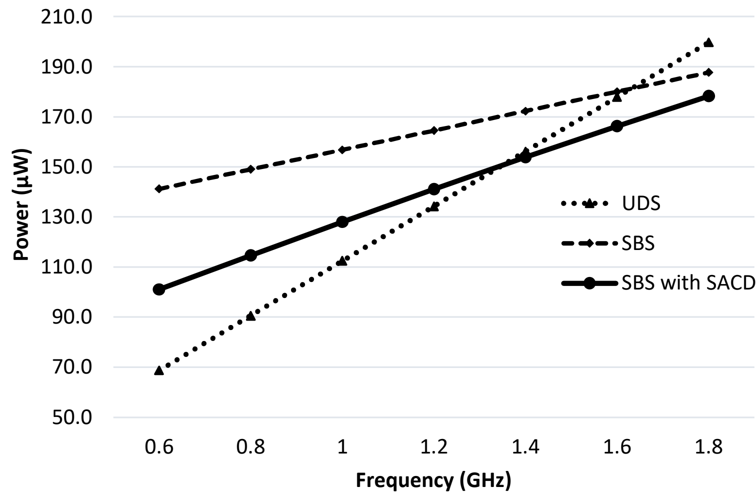


Figure 3-12: Power consumption of a single channel for UDS (Figure 3-1) and SBS (Figure 3-6) transceiver based TDM schemes.

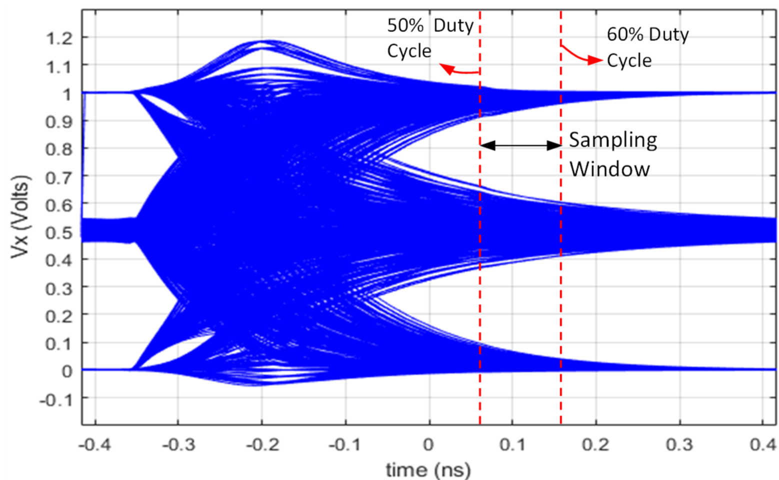


Figure 3-13: Eye diagram for the Vx signal at near end receiver side at 1.2GHz under TSV Cross coupling (all process corners).

3.6.1 Power Consumption

The power consumed by the test circuit designed using UDS and SBS based TDM is defined as the sum of average power consumed by the transmitters and receivers when a pseudo-random binary sequence is used as the input. At 1.2 GHz switching frequency, the total power consumption of the complete channel, including 1 x near-end transceiver and 3 x far-end transceivers of the global channel, was 164.5 µW for the SBS based design, which is 22.5% higher as compared to the UDS based scheme (134.2 µW), designed using 4x Buffers as transceivers. The power consumption trend of both designs with increasing

TABLE 3-II: BINARY-WEIGHTED TRANSMITTER PERFORMANCE

Sel bits	Tx Strength	Tx Power (µW)	Max Freq (GHz)
001	x1	52.28	0.67
010	x2	80.00	1.30
011	x3	107.00	1.77
100	x4	135.60	2.06
101	x5	162.80	2.21
110	x6	190.90	2.30
111	x7	218.30	2.46

frequency of operation is shown in Fig. 3-12. The power consumption of both methods increases with frequency. However, as the UDS-TDM can be designed using static CMOS, the static power component is minimal, and the overall power consumption is dominated by the dynamic power, which increases considerably with the frequency at the rate of 10.9 µW/100MHz. For the SBS based design, the dynamic power consumption is relatively lesser at 3.9 µW/100MHz, which can be attributed to the reduced voltage swing due to 3-level encoding and the relatively smaller transistor sizes. However, the major contributor to the overall power in SBS transceiver is the static power. As the static power consumption remains independent of the frequency, the power consumption of SBS at lower frequencies was observed to be higher than UDS. Nevertheless, due to lower dynamic power consumption, SBS consumes lesser power at higher frequencies as compared to UDS.

The static power consumption can be reduced by limiting the static current when both transmitters transmit the opposite signal, i.e., 10 or 01. To limit static currents and conserve power, designs such as capacitive coupling-based transmitters [38] or MOSFET Resistor (MOSR) coupled inverters [39] have been proposed. The capacitor-based design significantly improves static power consumption by blocking the steady-state current; however, the capacitor also blocks the noise discharge path making it prone to coupling noise, which may be significant in TDM based TAMs where global channels are required. The MOSR based transmitter limits the static power, but it also reduces the voltage swing and increases transient times/delays, and therefore only finds its use in low-frequency applications. The authors in [42] propose a Sense-Amplifier Completion Detector

(SACD) circuit that can turn off one of the transmitters after the TD has compared the inputs. The SACD was incorporated in the test circuit, and the SI side TG was turned off after completion of the sampling by the sense amplifier during every cycle. Fig. 3-12 compares the power consumption of an SACD-based design which reduces the static power by almost 18%; however, the additional circuit of the SACD adds to the dynamic power, which increases from 3.9 to 6.4 μW/100MHz.

Table 3-II presents the transmitter's average power consumption and the maximum frequency when a binary-weighted transmitter is used, as suggested in section IVB. The width for the x1 transmitter was chosen to be 180nm, which is twice the minimum technology width. The PMOS transistors were sized as 1.6 x NMOS. Similarly, the second and third stage transmitters were sized x2 and x4. The maximum frequency was estimated by measuring the 10-90% rise and fall times between rail to rail and rail to mid-level (Vxm). The reported maximum frequency (*fmax*) is 40% (0.8 x 0.5) of the frequency suggested by the transient response to account for the time required for sampling and sensing at the receiver, i.e. (*fmax*= 0.8 x (0.5/ max[rise-time, fall-time]). Clearly, decreasing the transmitter strength significantly reduces power consumption; however, the weaker transistors also limit the maximum achievable frequency. Therefore, there exists a trade-off between power consumption and maximum frequency, and an optimal transistor strength should be configured.

3.6.2 Signal Integrity under cross-coupling

As the TSVs are a relatively large structure traversing through the entire substrate, cross-coupling between TSVs becomes a significant concern. The authors in [47] presented a lumped RC model of the TSV cross-coupling, which was used to study the SBS transceiver's performance under cross-coupling noise. The RC values were calculated assuming silicon resistivity of 6.89 Ω.cm and a TSV pitch of 10μm. As the test circuit traverses 2 TSVs in the global channel, assuming every TSV is surrounded by 8 neighbouring aggressors (for a 3x3 cluster), and each neighbour driven by a PRBS sequence, the eye diagram of the Vx signal at the input of the near end (SI side) TD is shown in Fig 3-13. For

the given transmitter design for 1.2GHz frequency, the transceiver was verified working satisfactorily under cross-coupling in all process corners. Fig. 3-13 shows that most of the coupling noise diminishes at the negative clock edge (for 50% duty cycle); however, the sampling time may be delayed further by increasing the clock's duty cycle for more robust receiver performance. Moreover, the coupling noise may be reduced using TSV guard rings or using power/ ground TSVs among the neighbouring TSVs [48].

For a given sized transmitter, the eye height and width will decrease with increasing frequency. For correct transceiver operation, the transistor sizing of the transmitter, or in the case of the binary-weighted transmitter, the configuration of the transmitter is to be selected for the appropriate eye-opening at the desired frequency and power. The SA must be sized to account for the affordable offset voltage margin for the chosen reference voltages. The receiver sampling time may be adjusted using the clock duty cycle to provide sufficient timing margins.

3.6.3 Comparison with relevant prior work

Table 3-III compares this work with the other relevant work regarding power consumption improvement over UDS and maximum frequency. The results are compared with relevant previous works in TDM based 3D TAM design [28] and SBS designs for use in 3D SICs [37][38]. The authors in [28] reported 600μW average power consumption for a global channel of 3 dies, which is much higher than the SBS transceiver proposed in this paper (165μW). However, the authors used toggle input, 6x strength transmitters, and intermediate buffers resulting in higher power consumption. The power consumption of SBS and UDS is affected by various factors such as transistor technology, channel characteristics, and the bit patterns used. Therefore, we compare the percentage improvement in power consumption of the relevant SBS design over UDS designed in the same technology reported by the authors for a fair comparison. The SBS design in [38] consumes an estimated 37% more power compared to UDS based scheme,

Table 3-III: Comparison with relevant works

	Signaling Method	Bit pattern	Av. Power @ 1.2GHz	Power Improvement (%)	Max. Freq	Technology	Channel
This work	SBS/ TDM	PRBS	0.165mW	-22.5%	1.2GHz	45nm	-Multiple TSVs
Georgiou et al. [28]	UDS/ TDM	Toggle	0.6mW	NA	3.45GHz	45nm	-Multiple TSVs
Park et al. [37]	SBS	High Data Transition	0.154mW*	+31%	4.55GHz	28nm	-Single TSV
Aung et al. [37]	SBS	PRBS	0.106mW**	-37%**	>3.6GHz	65nm	-Capacitive Coupling

*estimated from reported energy efficiency at 4.55GHz clock
**estimated from reported static and dynamic power

which is slightly more than this work (22.5%). Park et al. in [37] reported an SBS transceiver design that was 33% more power-efficient than UDS transceivers for 2 x TSVs. However, the reported comparison was made at the maximum supported frequency, where it is expected to be more power-efficient, as suggested by the trend in Fig. 3-12. Another notable factor causing the difference in the power consumption and the maximum supported frequency is that the designs proposed in [37][38] were focused on a single TSV channel for point to point communication involving 2 x transceivers. On the contrary, a TDM based channel involves multiple dies/ TSVs in the channel involving multiple transceivers (1 near-end and 3 far-end transceivers used in our experiments). These factors increase power consumption and result in lesser frequency in this work than point-to-point communication channels.

Unlike full pin count test methods, which involve designing SBS transceiver at every chip terminal, RPCT based methods use only a subset of chip terminals, and therefore the area overheads are not a significant concern. In general, an SBS-based transceiver occupies an area similar to or slightly higher than the UDS counterpart designed for a similar frequency range [37].

3.7 CONCLUSION

A novel SBS-TDM based reduced pin count test strategy was proposed for testing TSV based 3D SICs. Design considerations for the transmitter, receiver, the control circuitry required for SBS based TDM, and associated trade-offs were presented. The transmitter was designed as a suitably sized inverter followed by a Transmission Gate, and the receiver designed as a Sense Amplifier. The power

consumption of the SBS based transceiver is dominated by the static power consumption of the transmitter, which can be minimized using appropriate transistor sizing and control. The limitations of the SBS-based test strategy in terms of the channel's electrical characteristics and possible solutions for the incorporation of SBS into pre-, mid-, and post-bond test instances were discussed. SBS transceiver can be bypassed for pre-bond testing allowing normal UDS communication, whereas adjustable strength SBS transmitters along with SBS buffers are proposed for mid-bond test insertions. Simulation results using an example test case in terms of power consumption and performance were presented. The proposed method consumed 22.5% more power while utilizing only half the number of TSVs than the UDS based design. The transceiver performance was verified across all process corners under cross-coupling from neighboring TSVs.

3.8 References

[1] I. A. Soomro, M. Samie, and I. K. Jennions, "Reduced Pin-Count Test Strategy for 3D Stacked ICs Using Simultaneous Bi-Directional Signaling Based Time Division Multiplexing," *IEEE Access*, vol. 9, pp. 75892–75904, 2021.

[2] R. Sharma and K. Choi, "Design of 3D Integrated Circuits and Systems," in *CRC Press*, CRC Press, 2014, pp. 157–174.

[3] M. B. Healy *et al.*, "Design and analysis of 3D-MAPS: A many-core 3D processor with stacked memory," in *IEEE Custom Integrated Circuits Conference 2010*, 2010, pp. 1–4.

[4] D. H. Kim *et al.*, "Design and analysis of 3D-MAPS (3D Massively parallel processor with stacked memory)," *IEEE Trans. Comput.*, vol. 64, no. 1, pp. 112–125, 2015.

[5] 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, Jan. 2010.

[6] H.-H. S. Lee and K. Chakrabarty, "Test Challenges for 3D Integrated Circuits," *IEEE Des. Test Comput.*, vol. 26, no. 5, pp. 26–35, Sep. 2009.

[7] G. Katti, M. Stucchi, K. De Meyer, and W. Dehaene, "Electrical Modeling and Characterization of Through Silicon via for Three-Dimensional ICs," *IEEE Trans. Electron Devices*, vol. 57, no. 1, pp. 256–262, Jan. 2010.

[8] H. Vranken, T. Waayers, H. Fleury, and D. Lelouvier, “Enhanced reduced pin-count test for full-scan design,” *J. Electron. Test. Theory Appl.*, vol. 18, no. 2, pp. 129–143, 2002.

[9] A. C. Evans, “Applications of semiconductor test economics, and multisite testing to lower cost of test,” in *International Test Conference*, 2003, pp. 113–123.

[10] A. H. Baba and K. S. Kim, “Framework for Massively Parallel Testing at Wafer and Package Test,” *Proc. - IEEE Int. Conf. Comput. Des. VLSI Comput. Process.*, pp. 328–334, 2009.

[11] M. L. Bushnel and IVishwani D. Agrawal, *ESSENTIALS OF ELECTRONIC TESTING FOR DIGITAL, MEMORY AND MIXED-SIGNAL VLSI CIRCUITS*. KLUWER ACADEMIC PUBLISHERS, 2002.

[12] “IEEE Standard for Test Access Port and Boundary-Scan Architecture,” *IEEE Std 1149.1-2013 (Revision IEEE Std 1149.1-2001)*, pp. 1–444, 2013.

[13] “IEEE Standard Testability Method for Embedded Core-based Integrated Circuits,” *IEEE Std 1500-2005*, pp. 1–136, 2005.

[14] “IEEE Standard for Test Access Architecture for Three-Dimensional Stacked Integrated Circuits,” *IEEE Std 1838-2019*. pp. 1–73, 2020.

[15] N. A. Touba, “Survey of Test Vector Compression Techniques,” *IEEE Des. Test Comput.*, vol. 23, no. 4, pp. 294–303, Apr. 2006.

[16] V. Iyengar, K. Chakrabarty, and E. J. Marinissen, “Test Wrapper and Test Access Mechanism Co-Optimization for System-on-Chip,” *J. Electron. Test. Theory Appl.*, vol. 18, pp. 213–230, 2002.

[17] X. Wu, Yibo Chen, K. Chakrabarty, and Yuan Xie, “Test-access mechanism optimization for core-based three-dimensional SOCs,” in *2008 IEEE International Conference on Computer Design*, 2008, pp. 212–218.

[18] E. J. Marinissen, “Challenges and emerging solutions in testing TSV-based 2 1 over 2D- and 3D-stacked ICs,” in *2012 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2012, pp. 1277–1282.

[19] “IEEE standard for access and control of instrumentation embedded within a semiconductor device - 1687,” *IEEE Std 1687-2014*, pp. 1–283, 2014.

[20] M. A. Ansari, J. Jung, D. Kim, and S. Park, “Time-Multiplexed 1687-Network for Test Cost Reduction,” *IEEE Trans. Comput. -Aided Des. Integr. Circuits Syst.*, vol. 37, no. 8, pp. 1681–1691, Aug. 2018.

[21] A. Sehgal, V. Iyengar, and K. Chakrabarty, “SOC test planning using virtual test access architectures,” *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 12, no. 12, pp. 1263–1276, Dec. 2004.

[22] A. Sanghani, B. Yang, K. Natarajan, and C. Liu, "Design and implementation of a time-division multiplexing scan architecture using serializer and deserializer in GPU chips," in *Proceedings of the IEEE VLSI Test Symposium*, 2011, pp. 219–224.

[23] M. S. Kawoosa, R. K. Mittal, M. Jalasuthram, and R. A. Parekhji, "Towards single pin scan for extremely low pin count test," *Proc. IEEE Int. Conf. VLSI Des.*, vol. 2018-Janua, pp. 97–102, 2018.

[24] B. Li and V. D. Agrawal, "Applications of Mixed-Signal Technology in Digital Testing," *J. Electron. Test. Theory Appl.*, vol. 32, no. 2, pp. 209–225, 2016.

[25] B. Noia, K. Chakrabarty, S. K. Goel, E. J. Marinissen, and J. Verbree, "Test-Architecture Optimization and Test Scheduling for TSV-Based 3-D Stacked ICs," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 30, no. 11, pp. 1705–1718, Nov. 2011.

[26] X. Wu, P. Falkenstern, K. Chakrabarty, and Y. Xie, "Scan-chain design and optimization for three-dimensional integrated circuits," *ACM J. Emerg. Technol. Comput. Syst.*, vol. 5, no. 2, pp. 1–26, Jul. 2009.

[27] M. A. Ansari, J. Jung, D. Kim, and S. Park, "Time-multiplexed test access architecture for stacked integrated circuits," *IEICE Electron. Express*, vol. 13, no. 14, pp. 20160314–20160314, 2016.

[28] P. Georgiou, F. Vartziotis, X. Kavousianos, and K. Chakrabarty, "Testing 3D-SoCs Using 2-D Time-Division Multiplexing," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 37, no. 12, pp. 3177–3185, Dec. 2018.

[29] B. G. West, "Simultaneous bidirectional test data flow for a low-cost wafer test strategy," in *International Test Conference, 2003. Proceedings. ITC 2003.*, 2004, vol. 1, no. January 2003, pp. 947–951.

[30] R. Mooney, C. Dike, and S. Borkar, "A 900 Mb/s Bidirectional Signaling Scheme," *IEEE J. Solid-State Circuits*, vol. 30, no. 12, pp. 1538–1543, 1995.

[31] H.-Y. Huang and R.-I. Pu, "Differential bidirectional transceiver for on-chip long wires," *Microelectronics J.*, vol. 42, no. 11, pp. 1208–1215, Nov. 2011.

[32] Y. Tomita, H. Tamura, M. Kibune, J. Ogawa, K. Gotoh, and T. Kuroda, "A 20-Gb/s Simultaneous Bidirectional Transceiver Using a Resistor-Transconductor Hybrid in 0.11-u CMOS," *IEEE J. Solid-State Circuits*, vol. 42, no. 3, pp. 627–636, Mar. 2007.

[33] Jae-Yoon Sim, Young-Soo Sohn, Seung-Chan Heo, Hong-June Park, and Soo-In Cho, "A 1-Gb/s bidirectional I/O buffer using the current-mode scheme," *IEEE J. Solid-State Circuits*, vol. 34, no. 4, pp. 529–535, Apr. 1999.

[34] C. J. Akl and M. A. Bayoumi, “Wiring-area efficient simultaneous bidirectional point-to-point link for inter-block on-chip signaling,” *Proc. IEEE Int. Freq. Control Symp. Expo.*, pp. 193–200, 2008.

[35] M. K. Jeon and C. Yoo, “A single-ended simultaneous bidirectional transceiver in 65-nm CMOS technology,” *J. Semicond. Technol. Sci.*, vol. 16, no. 6, pp. 817–824, 2016.

[36] P. Vijaya Sankara Rao and P. Mandal, “Current-mode full-duplex (CMFD) signaling for high-speed chip-to-chip interconnect,” *Microelectronics J.*, vol. 42, no. 7, pp. 957–965, Jul. 2011.

[37] S. Park, A. Wang, U. Ko, L.-S. Peh, and A. P. Chandrakasan, “Enabling Simultaneously Bi-Directional TSV Signaling for Energy and Area Efficient 3D-ICs,” in *Proceedings of the 2016 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2016, pp. 163–168.

[38] M. T. L. Aung, E. Lim, T. Yoshikawa, and T. T.-H. Kim, “Design of Simultaneous Bi-Directional Transceivers Utilizing Capacitive Coupling for 3DICs in Face-to-Face Configuration,” *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 2, no. 2, pp. 257–265, Jun. 2012.

[39] I. A. Soomro, M. Samie, and I. K. Jennions, “Test Time Reduction of 3-D Stacked ICs Using Ternary Coded Simultaneous Bidirectional Signaling in Parallel Test Ports,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 39, no. 12, pp. 5225–5237, 2020.

[40] N. Wary and P. Mandal, “Current-mode simultaneous bidirectional transceiver for on-chip global interconnects,” in *2015 6th Asia Symposium on Quality Electronic Design (ASQED)*, 2015, pp. 19–24.

[41] T. Na, S. H. Woo, J. Kim, H. Jeong, and S. O. Jung, “Comparative study of various latch-type sense amplifiers,” *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 22, no. 2, pp. 425–429, 2014.

[42] T. Kobayashi, K. Nogami, T. Shirotori, and Y. Fujimoto, “A Current-Controlled Latch Sense Amplifier and a Static Power-Saving Input Buffer for Low -Power Architecture,” *IEEE J. Solid-State Circuits*, vol. 28, no. 4, pp. 523–527, 1993.

[43] Y. Chen, P. I. Mak, J. Yang, R. Yue, and Y. Wang, “Comparator with built-in reference voltage generation and split-ROM encoder for a high-speed flash ADC,” *ISSCS 2015 - Int. Symp. Signals, Circuits Syst.*, pp. 2–5, 2015.

[44] J. Yang, Y. Chen, H. Qian, Y. Wang, and R. Yue, “A 3.65 mW 5 bit 2GS/s flash ADC with built-in reference voltage in 65nm CMOS process,” *ICSICT 2012 - 2012 IEEE 11th Int. Conf. Solid-State Integr. Circuit Technol. Proc.*, pp. 5–7, 2012.

[45] M. T. L. Aung, E. Lim, T. Yoshikawa, and T. T. H. Kim, “A 3-Gb/s/ch

simultaneous bidirectional capacitive coupling transceiver for 3DICs," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 61, no. 9, pp. 706–710, 2014.

[46] "Nangate 45nm FreePDK Library." [Online]. Available: https://si2.org/open-cell-library/.

[47] T. Song, C. Liu, Y. Peng, and S. K. Lim, "Full-Chip Signal Integrity Analysis and Optimization of 3-D ICs," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 24, no. 5, pp. 1636–1648, May 2016.

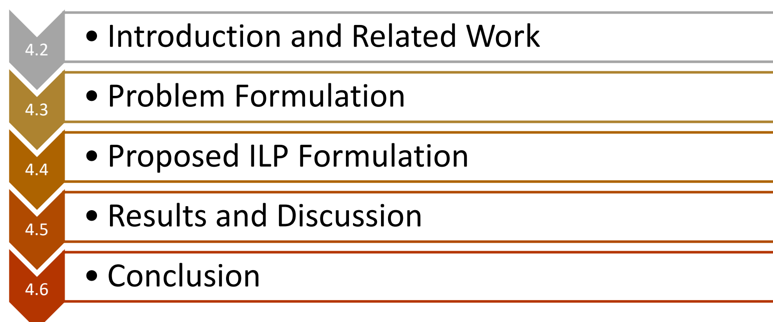
[48] Y. Peng, T. Song, D. Petranovic, and S. K. Lim, "Silicon effect-aware full-chip extraction and mitigation of TSV-to-TSV coupling," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 33, no. 12, pp. 1900–1913, 2014.

[49] M. T. L. Aung, E. Lim, T. Yoshikawa, and T. T. Kim, "Design of capacitive-coupling-based simultaneously bi-directional transceivers for 3DIC," *2011 IEEE Int. 3D Syst. Integr. Conf. 3DIC 2011*, pp. 1–4, 2011.

4 OPTIMIZATION AND TEST TIME ANALYSIS OF SIMULTANEOUS BI-DIRECTIONAL SIGNALING BASED TEST ACCESS MECHANISM FOR 3D STACKED INTEGRATED CIRCUITS

4.1 Abstract

The design methodology of an SBS based Test Access Mechanism in Parallel Test Ports of 3D SICs was presented in Chapter 2. In this chapter, the benefits of using SBS are quantified at the application level (Test Application Time), using 3D SICs based on ITC'02 benchmarks. The chapter begins with an introduction and an overview of the relevant prior work in the TAM optimization for test time reduction. In section 4.3, an overview of the combinatorial nature of the TAM design problem is discussed, which leads to the development of the TAM design as an optimization problem. In section 4.4, an Integer Linear Programming based optimization formulation is presented for SBS based TAM design, followed by the results and discussion in section 4.5, in which quantitative comparisons with the UDS based baseline methods are made. The chapter concludes in section 4.6. Parts of this chapter (the results and discussion section) have been published in [1].



4.2 Introduction and Related Work

The test time of ICs increases with design complexity and node density, and has been a subject of significant research over the years. A significant contributor to test time is the test vector transport phase, in which a large volume of data is required to be serially shifted into the internal scan chains [2]. 3D stacking brings about several additional and more complex challenges for test access [3][4][5][6].

First, higher transistor density increases the probability of manufacturing defects such as metal bridging, metal opens, via opens, and transistor defects. It, therefore, requires higher test vector volume for adequate coverage but without any significant increase in the chip terminals, which further tightens the test access bottleneck. Second, the manufacturing process of 3D SICs introduce additional defects and necessitates multiple test instances. Apart from wafer level and chip-level testing (known as pre-bond and post-bond testing respectively), the 3D SIC has to be tested at every point during the stacking process, known as mid-bond testing. Finally, the limited number of inter-die vertical connections (such as Through Silicon Vias (TSVs), micro-bumps, or wire-bonds) add a further test access restriction in addition to chip terminals when transporting test vectors to dies higher up in the stack. As a result of the stated challenges, the Test Application Time (TAT) in 3D SICs increase significantly compared to conventional 2D Chips, necessitating new test-access designs to bring down the test time and hence the test cost of the chips. Testing 3D SICs is therefore considered a major constraint and listed as one of the difficult challenges for the industry by the International Technology Roadmap for Semiconductors (ITRS) [7].

The test time of a die in a 3D-stack is a function of the test channels allocated to it. In general, as the number of test channels for a die are increased, more cores could be tested in parallel, and the test time for the die is reduced. In System on Chip (SoCs), every core has specific test requirements that are specified by the core designer and it is the job of SoC vendor to put in place an appropriate Test-Access Mechanism (TAM) that fulfills these requirements. The TAM could connect the cores in series, such that the scan-chains are tested sequentially, or the cores could be connected in parallel such that they could be tested simultaneously.

TAM design is shown to be an NP-Hard optimization problem [8] which has attracted significant research. Test time varies significantly for different TAM design choices, and one way of reducing it is to formulate optimization algorithms that points out the best possible TAM in the entire solution space with the

objective of minimizing TAT while adhering to the chip constraints. Most of the noticeable work has been focused on 2D SoC designs, and various approaches to solving the problem have been adopted such as Integer Linear Programming [8], Rectangle Bin Packing [9] and heuristics [10] [11]. While most of the work on 2D TAMs has been focused on the constraints imposed by available TAM width, some researchers [12][13] have also focused on thermal and power consideration, which is also an essential factor since significantly higher switching activity is observed during tests.

As mentioned previously, 3D SICs are considerably different and make the optimization process more challenging [3][4][5]. Wu et al. [14] presented a method to optimize the 3D TAM designed using wrapper design suggested in [8] using a heuristic combination of Integer Linear Programming (ILP), randomized rounding, and LP relaxation. In [15], the authors addressed the problem of scan-chain ordering and partitioning using the Genetic Algorithm and ILP combined with heuristics to reduce wire length and TAT. The authors in [16] proposed heuristics to design optimized TAM under a set of uncertainties. The authors in [17] proposed an ILP based model for designing a TAM for 3D SICs for UDS based TAM. It may be pointed out that the TAM optimization solution space is bounded by the physical layer of the TAM design. Most of the above works have relied upon conventional Uni-directional TAM design techniques and have not proposed any significant improvements in the physical design of the access infrastructure itself, specifically, the improvement in pin efficiency.

In [1], a novel Test Access Mechanism (TAM) design was proposed for 3D SICs that doubles the data transfer efficiency of the pins and TSVs. This was achieved by leveraging Simultaneous Bi-directional Signaling (SBS) for full-duplex test mode communication at chip terminals. SBS allows transmission and reception of test bits simultaneously compared to the conventional Uni-Directional Signaling (UDS) scheme in which the signal could travel in only one direction at a given time. Using SBS, a complete transmission and reception channel could be formulated using a single electrical path at the chip terminal instead of two,

effectively doubling the number of test-channels and increased parallelism in test scheduling.

Simultaneous Bi-Directional Signalling (SBS) doubles the channel bandwidth, while this statement may be sufficient to report the improvement expected in normal mode communication in which the exact usage of the link over a longer period may not be known, testing is a deterministic process in which the link utilization over the entire period of the test is known, and it is possible to calculate the exact improvements in test times. Furthermore, apart from the data rate at the communication channels (Pins/ TSVs), the test time of the 3D SIC also depends on the number of available test pins and TSVs, the construction of the SIC and the dies, their test requirements, and also the Test Access Mechanism (TAM), which may be different for both.

In order to calculate test times and draw comparisons between SBS and UDS, a design and scheduling methodology for SBS based TAMs is required. In [17], the authors proposed an ILP based TAM design and scheduling method for 3D SICs. However, as the method was intended for UDS based TAM, the authors have assumed a fixed relationship between the chip terminals and the test channels. For a generalised formulation usable for both SBS and UDS, the chip terminals and test channels and the relationship between them must be defined explicitly. In this article, we extend the 3D TAM design formulation of [17] such that it could be used for both UDS and SBS. We reformulate the problem such that constraints could be applied separately on test channels and chip terminals, which would also allow the extension of the formulation to a more complicated design problem requiring co-existence of SBS as well as UDS schemes in the 3D Stack (further discussed in Chapter 5).

To summarise, this paper addresses the following gaps in the previous research:

a. Propose a TAM design and Optimization methodology for 3D SICs that could be extended to SBS based TAM designs.
b. Study the improvements offered by SBS over conventional UDS TAMs in terms of test time.

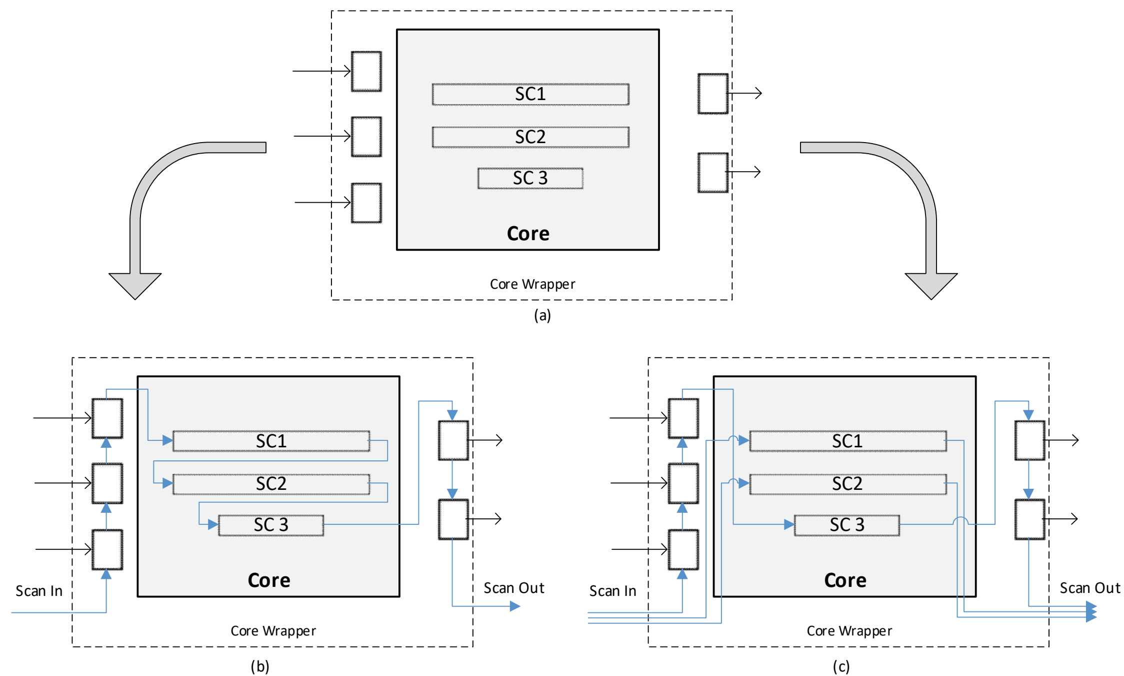


Figure 4-1: The core wrapper design problem (a) An example core containing three scan chains of length, 10,9 and 5 and 5 I/Os (b) Serial concatenation of all scan elements for 1 bit TAM (c) Assignment of scan element to multi-bit TAM (3 in this case)

Section 4.3 describes the TAM design problem, the requirement for optimization and briefly describes how the problem could be translated into Integer Linear Programs. Section 4.4 presents the ILP formulation to optimize UDS and SBS-based TAMs for 3D SICs. The results are presented in section 4.5 along with a discussion into the test time improvements offered by SBS with the conventional UDS. The paper concludes in section 4.7.

4.3 Problem Formulation

Modern chip design mostly follows a design re-use philosophy, in which the chip functionality is achieved using pre-designed IP blocks from different vendors. These functional blocks are also known as cores, and it is at this level of abstraction, the test requirements are specified in terms of the number of test patterns, the number of IOs of the core, and the number of scan chains and their lengths. The functional and test requirements must be fulfilled and are usually done by wrapping the core with additional logic. To ensure interoperability between the core and die designer, the wrapper is designed using standard methods [18]; however the channel width allocation choice is left to the designer.

For instance, the Core in Fig. 4-1 has five I/O chip terminals (CTs) and three scan chains of length, 10,9 and 5. The designer may choose to connect all the scan elements (Scan chains and CTs) into a single long chain, as illustrated in Fig. 4-1(b), or multiple smaller chains as in Fig 4-1(c), or any intermediate combination. Clearly, the choice would depend on the available test channels and will impact test times, therefore, an optimal choice is to be made.

The next level of abstraction is the die level, where the designer must now fulfil the test requirement of every core within the die. Similar to the core wrapper design, the die wrapper also provides several design choices into how the cores are connected and the allocation of test channels among cores [19][20]. The extreme cases are shown in Fig. 4-2, such as the cores are connected serially in Fig. 4-2(a) or in parallel Fig. 4-2(b) or some intermediate combination. Typically, a core is a standalone element and is tested in one go. The design choice in Fig 4-2(a), therefore, requires that all the cores are tested sequentially but using all the test channels and maximum bandwidth (this arrangement is also referred to as Daisychain architecture in [20]) . In sequential testing, once a core is tested, it is placed into the bypass mode and becomes transparent, allowing the other core to be tested. The sequential nature of the design choice in Fig. 4-2(a) adds up the die test time but reduces the individual core test time (offering an increased number of test channels). The arrangement in Fig. 4-2(b), also commonly known as Distributed Architecture [20], allows parallel execution of core tests but with reduced channel allocation (causing core test times to increase). Several other die level designs (Serial/ Parallel core combinations) will carry different implications on the required resources vs test times.

At the stack level, keeping in view the die level trade-offs (which in turn depend on core level design), the test channel resources (Pins and TSVs) must now be appropriately distributed among dies. In 3D SICs, the parallel ports are flexible and can be reconfigured to access higher dies as required. At the SIC level, a ‘3D session’ is defined as the duration in which specific dies in the SIC are tested in parallel while the remaining are idle. A sequence of sessions forms a schedule.

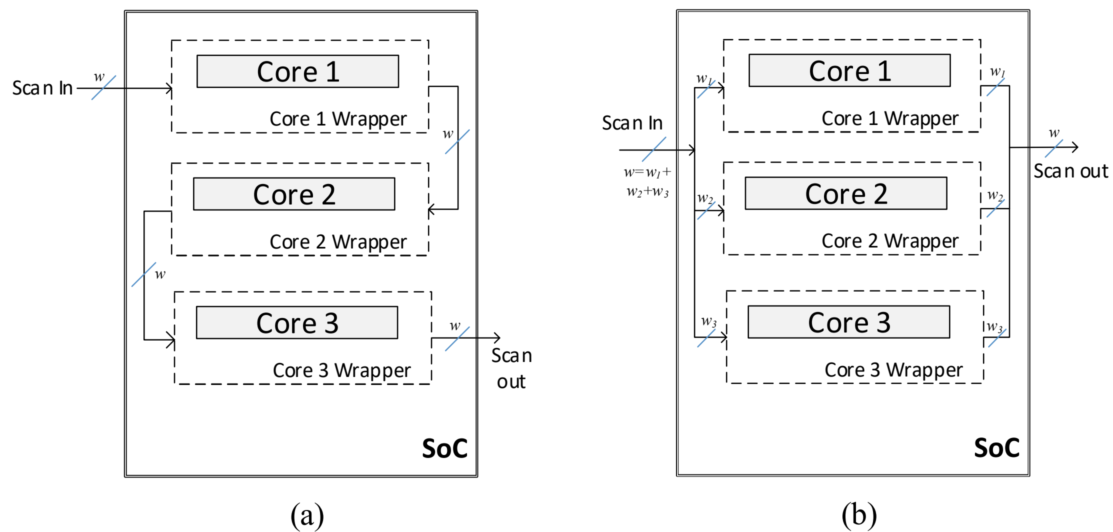


Figure 4-2: TAM width assignment to cores [20] **(a) Serial concatenation for sequential testing using full TAM width (b) Parallel testing using distributed TAM width**

Within a 3D Session, a full set of test channels is available to the dies being tested in the session, although as the dies will be tested in parallel, the channels must be divided appropriately among dies. The overall test time of the 3D stack is therefore governed by the choices made for resources allocated at the stack, die, and core level.

In an ideal situation, all the cores in a die and all the dies in a 3D SIC would be tested simultaneously, resulting in the minimum Test Application Time (TAT). Here, TAT is defined as the number of clock cycles required to apply test patterns to the entire 3D SIC. However, several constraints limit this approach, such as 1: It may not be possible to access all cores at the same time as there are limited chip terminals (Pins and TSVs), 2: there may be limitations posed by power and thermal dissipation which need to be adhered, and 3: the chip area that could be dedicated for the test architecture may be limited. Clearly, there is a trade-off between resources and the test time, and the task of the designer is to find the best possible solution for a TAM. However, the optimal resource allocation of the stack at all the above three levels is a combinatorial problem for which the computational complexity increases exponentially with the increasing design variables. The TAM design problem, combined with scheduling, is known to be NP-Hard [21].

The overall problem could be stated as: Given a SIC with M number of dies, each die has N cores such that N_b denotes the total number of cores in die b. Each core has certain I/Os, scan chains of specific lengths, and the number of test patterns to be applied. Let P_{max} denote the maximum number of pins available at the lowermost die and TSV_{max} denote the global TSV limit. *If the Test Access Mechanism could be designed either as UDS or SBS, determine the optimal TAM design that minimises the Test Application Time (TAT)* by:

a. Finding the 3D schedule (how dies should be tested) by optimal allocation of Test Channels to each die such that the upper limit on P_{max} and TSV_{max} is not exceeded.

b. Finding the optimal 2D schedule (how cores within the die should be tested) such that the allocated number of channels dictated by the SIC level schedule (in part a) is not exceeded.

In a Linear Programming (LP) problem, the aim is to minimize (or maximize) a linear objective function under a set of equality or non-equality linear constraints. For a vector variable x, a typical ILP problem is of the form

$$Minimize\colon\ \boldsymbol{A}x\ \ , subject\ to\colon\ \boldsymbol{B}x\ \leq \boldsymbol{C},$$

Where $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are constant value (or weights) matrices. x is the vector of non-negative variables, also known as decision variables, that are to be assigned optimal values. The assignments of x for which the constraints $\boldsymbol{B}x\ \leq \boldsymbol{C}$ are not violated are termed as feasible solutions and the task of the optimizer to identify the feasible solution associated with minimum $\boldsymbol{A}x$. In Integer Linear Programs (ILP), the optimization variables are restricted to integer values.

In the next section, the ILP formulation for the TAM design and optimization problem for SBS based 3D SICs is presented. For simplicity, only post-bond test scenarios of soft-dies and soft-cores have been considered. Moreover, only the test time involved in the shift cycle of the scan-test has been calculated as the rest can be ignored, being negligible.

4.4 Proposed ILP Formulation

To formulate the TAM design problem into ILP, the objective function and the constraints need to be defined as a set of linear equations of type $\boldsymbol{A}x$ and $\boldsymbol{B}x \leq \boldsymbol{C}$ respectively. In the next paragraph, we introduce the method of test time calculation, based on [17], which can be used as the objective function. The objective function requires the test time of individual dies, which is the function of test channels x. In section 4.4.1, the relationship and constraints are defined that translate x to the TAM width available to a die, such that communication between the tester and the die is possible. In section 4.4.2, the relationship between the test channels x and the chip terminals (Pins or TSVs) is established, such that constraints can be specified explicitly on chip terminals. Having defined the required relationships and constraints, the linearized objective function is presented in section 4.4.3 and the overall ILP formulation is summarised in section 4.4.4.

Assuming there are *M* dies in a stack, the position of each die in the stack is denoted by an index variable $m: m \in \{1,2,..,M\}$. The test time T of a given die, is a function of the number of test channels x allocated to that die, i.e. $T_m = f(x_m)$. The decision variable x could be assigned any non-negative integer value up to a maximum of x_{max} which denotes the upper bound on the maximum number of channels, such that $x_m: x \in \{0,1,\dots,x_{max}\}$. Further, assume there are t test sessions such that within a session, a subset of dies could be tested in parallel. In the extreme case where all the dies are tested individually in a separate session, there will be a maximum of *M* sessions, such that $t: t \in \{1,2,..,M\}$. If the channel allocation of each die is known, the Test Application Time for a given session (denoted by TAT_t) will be the maximum of the dies being tested in that session, i.e. $TAT_t = \max_{m \in t}(T_m)$. If the test time of all sessions is known, the overall TAT can be computed by the sum of the TAT_t of all the sessions. The overall test application time of the 3D stack is therefore given by:

$$TAT = \sum_{t=1}^{M} TAT_t = \sum_{t=1}^{M} \max_{m \in t}(T_m) \qquad \textbf{Eq 4-1}$$

In order to compute the TAT using Eq. 4-1, first, we need to compute every TAT_t by finding the optimal channel width allocation x, as well as the optimal arrangement of dies within each session t. In the following sub-sections, we describe the mathematical formulation of this problem by defining the linear programming constraints relating x and TAT_t.

4.4.1 Forming Test Channels between Tester and Dies

To allocate a channel width to die m, the Test Access Mechanism (TAM) can only be formed between the tester and the die if there is sufficient channel allocation at all the intermediate layers. Here we define a 'layer' the same way as in [17] as the communication channel between any two dies in the stack, or in case of the first die, between the tester and first die.

Let l be the set of all possible layers between the dies (1 to M), such that $l = 1$ indicates the interface between the tester and Die1, $l = 2$ indicate the layer between dies 2 and 1, and so on such that $l: l \epsilon \{0,1,\dots,M\}$. The non-negative continuous integer variable x_{lm} is now defined for every die m and layer l s.t $x_{lm}: 0 \le x_{lm} \le x_{max}, \forall m \epsilon \{1,2,..,M\}, \forall l \epsilon \{1,2,..,M\}$.

The variable x is the decision variable in the optimization problem, which allocates the channel width to each die. However, a die at position m only requires test channels (x) at layers at or below it. For ease of understanding, a variable TAM_{lm} is defined to denote the TAM width allocated to die *m*, defined only for required layers (i.e. $l \le m$):

$$TAM_{lm} = x_{lm} \; \forall m, \forall l \le m \qquad \textbf{Eq 4-2}$$

The constraints on test channels allocation can now be defined on TAM_{lm}. In order to create a test channel between the tester and a die m, the channels included in the layers at and below m should be at least of the same width as of the topmost layer for the die.

$$TAM_{(l-1)m} \ge TAM_{lm} \qquad \forall m \ge 2, \forall\, 2 \le l \le m \qquad \textbf{Eq 4-3}$$

Formulating this constraint in terms of TAM width, rather than the number of test channels allows any combination of SBS+UDS at each layer irrespective of the design at any other layer. As the layers for a given die are only defined for layers at or below m, TAM_{mm} can be considered as the width allocated to die m.

The following constraint ensures that at least a single width channel is allocated for die m, in all layers at or below m. This guarantees that all dies are tested.

$$TAM_{lm} \geq 1 \ \ \forall m, \forall l \leq m \qquad \textbf{Eq 4-4}$$

The authors in [8] reported that the test time of a die increase as the number of test channels are increased. However, beyond a certain channel width there is no further reduction in test time. In order to reduce the solution space of the optimization problem, an upper bound is introduced on the maximum channels (denoted by w_{max}) that could be allocated a die m, which could be computed using the method described in [8]. The following constraint ensures that the TAM width allocated to die (TAM_{mm}) must not exceed the maximum channel width supported by it.

$$TAM_{mm} \leq \ w_{max,m} \qquad \forall m \qquad \textbf{Eq 4-5}$$

It may be noted that the upper bound x_{max} is simply the maximum of $w_{max,m}$.

4.4.2 Mapping Test Channel to Pins/ TSVs:

The decision variable x will contain the optimal test requirements '*for each die'* in terms of test channels required '*at layers* $l \leq \ m$'. We now define a new decision variable for all layers $PMAP_l$ (for Pin Map), which indicates the chip terminal type (UDS or SBS) used to form the test channels in x.

First, we note that only a subset of dies in the stack maybe tested during every session t; therefore, the resources (Pins/ TSVs) need not be allocated to the dies not being tested in that session. A binary decision variable O_{mt}, that would optimally allocate dies to different sessions, is defined such that $O_{mt} = 1$ if a die m is tested in session t, and 0 otherwise. As previously highlighted, as there are M dies in the stack, at most M session would be required (in the worst case when

all dies are tested sequentially). Therefore, the upper bound on the range of t is M such that $t: t \in \{0,1,\ldots,M\}$. As every die must be included in one and only one session t, the following constraint is defined:

$$\sum_{t=1}^{M} O_{mt} = 1 \;\forall m \qquad \textbf{Eq 4-6}$$

Secondly, the relationship between the test channel type (SBS/UDS) and the number of chip terminals required to design the test channel type must be constructed. We define a variable s, which is 1 if UDS design is used and 2 for SBS design. As an SBS channel uses 1 pin and a UDS channel requires two pins, this is simply achieved by scaling x by a factor of $2/s$. This ensures that for UDS channels ($s = 1$), the variable x is doubled, whereas, for $s = 2$ (SBS), the variable x is unchanged.

Having now defined the test sessions and the relationship between a test channel and chip terminals, we can now construct the relationship between the pin requirement ($PMAP_l$) and test Channel requirement (x) for every session t as:

$$PMAP_l^t = \sum_{m=l}^{M} x_{lm} \,.\, O_{mt} \,.\frac{2}{s} \quad \forall t, \forall l \leq m \qquad \textbf{Eq 4-7}$$

The superscript t indicates that the $PMAP_l$ is for a given session t. The above equation simply states that the pins/ TSVs required for the session is the sum of chip terminal requirements for all the dies being tested in that session. The variable O_{mt} ensures that the test requirements for the dies which are not being tested in that session are zeroed out.

The variable $PMAP_l^t$ shows the requirement of chip terminals during every session in the schedule. The final construction of the TAM must fulfil the requirements of all sessions. This implies that PMAP for a session t, must also exist in all subsequent sessions. The above requirement can be satisfied if every element of $PMAP_l$ (the number of chip terminals of UDS or SBS type at every layer) takes on the maximum value among all sessions. Therefore, the following constraint is added:

$$PMAP_l = \max_t PMAP_l^t \qquad \forall l \qquad \textbf{Eq 4-8}$$

The constraints on the maximum Pins (P_{max}) and TSVs (TSV_{max}) can now be defined in terms of $PMAP_l$. As the SBS scheme requires 2 chip terminals at every layer to form the reference sharing network, the same must be accounted for in the final stack. The constraint on the number of available pins and TSVs can be imposed as:

$$PMAP_l + 2(s-1) \le P_{max} \quad for \ l = 1 \qquad \textbf{Eq 4-9}$$

$$\sum_{l=2}^{M} (PMAP_l + 2(s-1)) \le TSV_{max} \qquad \textbf{Eq 4-10}$$

For $s = 1$ (UDS) the factor $2(s-1)$ reduces to zero, and no chip terminals are reserved for reference network, whereas for SBS scheme ($s = 2$), the factor $2(s-1)$ reduces to 2, indicating that 2 Pins/ TSVs must be reserved at every layer for reference sharing. Eq. 4-9 ensures that the sum of all chip terminals at the first layer must be lesser or equal to the P_{max}. Eq. 4-10 ensures that the number of chip terminals included at or above the second layer must obey the global TSV limit TSV_{max}.

It may be noted that Eq. 4-7 contains the non-linear product of the decision variable x_{lm} and O_{mt}, whereas Eq. 4-8 contains a max function that is also non-linear. In order to formulate a linear program, the equations containing non-linear elements need to be linearized as discussed in Appendix 4-I.

4.4.3 Formulation of the Objective Function

The objective, in this case, is to minimize the TAT, which From Eq (4-1) is given by:

$$Minimize: \qquad TAT = \sum_{t=1}^{M} \max_{m \in t} (T_m) \qquad \textbf{Eq 4-11}$$

The test time of a die (T_m) is a non-linear function of the variable TAM_{mm}, which can be linearized by implementing T_m as a look up table T_{mw} for all possible values of channel widths $w: w \epsilon \{1,2,..,x_{max}\}$. A binary variable can then be used to select the test time for a given channel width. Consider a binary variable Y_{mw}, defined for all dies m and channel widths w. The following constraint ensures that for a die m, only the wth element of Y_{mw} is *1* if $w = TAM_{mm}$.

$$\sum_{w=1}^{w_{max}} Y_{mw}.w - TAM_{mm} = 0 \quad \forall m \qquad \textbf{Eq 4-12}$$

The die test time T_m can now be represented as the product of T_{mw} and Y_{mw}, i.e.:

$$T_m = \sum_{w=1}^{w_{max}} T_{mw}.Y_{mw} \qquad \textbf{Eq 4-13}$$

The domain of the max function can be equivalently written in terms of the variable O_{mt} as:

$$\max_{m \epsilon t} (T_m) = \max_{m} (T_m.O_{mt}) \qquad \textbf{Eq 4-14}$$

Using Eq. 4-20 and Eq. 4-21, the objective function of Eq. 4-18 can therefore be re-written as:

$$Minimize: \qquad TAT = \sum_{t=1}^{M} \max_{m} \sum_{w=1}^{w_{max}} T_{mw}.Y_{mw}.O_{mt} \qquad \textbf{Eq 4-15}$$

The above equation contains two non-linearities, the product of Y_{mw} and O_{mt}, and the max function of this product. The product of Y_{mw} and O_{mt} could be linearized by using standard linearization techniques using an additional binary variable YO_{mwt}. The following constraints ensure that the variable YO_{mwt} always takes on the value of the product $Y_{mw}.O_{mt}$:

$$Y_{mw} + O_{mt} - YO_{mwt} \leq 1 \quad \forall m \, \forall w \, \forall t \qquad \textbf{Eq 4-16}$$

$$Y_{mw} + O_{mt} - 2.YO_{mwt} \geq 0 \quad \forall m \, \forall w \, \forall t \qquad \textbf{Eq 4-17}$$

The above constraints ensure that YO_{mwt} is 1, only if both Y_{mw} and O_{mt} (and hence the product $Y_{mw}.O_{mt}$) are 1, in which case the first constraint reduces to $YO_{mwt} \geq 1$, and the second constraint reduces to $YO_{mwt} \leq 1$. Both of these constraints can only be satisfied if $YO_{mwt} = 1$. On the contrary, if any of the variables Y_{mw} and O_{mt} are zero, the term $Y_{mw} + O_{mt}$ is always less than or equal to 1, reducing the above constraints to $YO_{mwt} \geq -1$, and $YO_{mwt} \leq 0.5$ which forces YO_{mwt} to zero.

Similar to the linearization of the max function in Eq. 4-8, The max function in Eq. 4-22 can be linearized by treating the variable TAT_t as a decision variable, which takes the maximum value for every session t, using the following constraint:

$$TAT_t \geq \sum_{w=1}^{w_{max}} T_{mw}.YO_{mwt} \qquad \forall m \ \forall t \qquad \textbf{Eq 4-18}$$

It may be noted that unlike the linearization used for the max function in Eq. 4-8, only the constraint which describes the lower bound on TAT_t has been kept, and the constraints describing the upper bound as in Eq. 4-16 and Eq. 4-17 have been ignored. This is because TAT_t is directly a part of the objective function, which is to be minimized, and hence the constraints on lower bound become redundant.

4.4.4 The Overall Linearized UDS/SBS TAM Optimization Model

The overall optimization formulation can now be summarized as:

Given a design choice of a UDS ($s = 1$) or SBS ($s = 2$):

Objective $\qquad minimize \ TAT$

Subject to following constraints:

$$TAT = \sum_{t=1}^{M} TAT_t$$

$$TAT_t \geq \sum_{w=1}^{w_{max}} T_{mw}.YO_{mwt} \qquad \forall m \ \forall t$$

$$Y_{mw} + O_{mt} - YO_{mwt} \leq 1 \quad \forall m \ \forall w \ \forall t$$

$$Y_{mw} + O_{mt} - 2.YO_{mwt} \geq 0 \quad \forall m\, \forall w\, \forall t$$

$$\sum_{w=1}^{w_{max}} Y_{mw}.w - TAM_{mm} = 0 \quad \forall m$$

$$TAM_{lm} = x_{lm} \quad \forall m, \forall l \leq m$$

$$TAM_{(l-1)m} \geq TAM_{lm} \quad \forall m \geq 2\, , \forall\, 2 \leq l \leq m$$

$$TAM_{lm} \geq 1 \quad \forall m, \forall l \leq m$$

$$TAM_{mm} \leq w_{max,m} \quad \forall m$$

$$PMAP_l + 2(s-1) \leq P_{max} \quad for\ l = 1$$

$$\sum_{l=2}^{M} (PMAP_l + 2(s-1)) \leq TSV_{max}$$

$$\sum_{t=1}^{M} O_{mt} = 1 \quad \forall m$$

$$0 \leq XO_{lmt} \leq x_{lsm} \quad \forall l \leq m\, \forall m, t$$

$$XO_{lmt} - \alpha \quad O_{mt} \leq 0 \quad \forall l \leq m\, \forall m, t$$

$$XO_{lmt} - x_{lm} + (1 - O_{mt}).\alpha \geq 0 \quad \forall l \leq m\, \forall m, t$$

$$PMAP_l^t = \sum_{m=l}^{M} XO_{lmt} . \frac{2}{s} \quad \forall t, \forall l \leq m$$

$$PMAP_l \geq PMAP_l^t \quad \forall l\, \forall t$$

$$PMAP_l \leq PMAP_l^t + (1 - c_{lt}).\beta \quad \forall l\, \forall t$$

$$\sum_{t=1}^{M} c_{lt} = 1 \quad \forall l$$

4.5 Results and Discussion

Table 4-I: 3D SICs composition

SIC No.	No. of Dies	Composition of each SIC
1.	5	p93791,p34392,p22810,f2126,d695
2.	5	d695,f2126, p22810, p34392, p93791
3.	5	f2126, p22810, p93791, p34392, d695
4.	2 to 9	p93791,p34392,p22810,f2126,d695, q12710,h953,g1023,d281

In order to compare the test time of the conventional UDS TAM with SBS (assuming all pins and TSVs support SBS), experiments were conducted using 3 x handcrafted 3D SICs from the ITC 02 [22] Benchmarks reported in [17] and a fourth SIC in which the number of dies was gradually incremented from 2 to 9. The composition of each SIC is shown in Table 4-1. The computational complexity of the ILP increases exponentially with an increasing number of variables and constraints. The complexity can be approximated by the number of decision variables and constraints [21], which for this problem was $O(M^3.x_{max})$, where M is the number of dies in the stack and x_{max} is the maximum number of test channels possible. The ILP formulation was solved in IBM CPLEX and run-time using a desktop computer was under a second for most of the problems (as high as 4 seconds in some instances). However, for SIC4 in case of stacks with 8 and 9 dies, the run time was observed to be ranging from a few minutes to as high as 93 minutes.

The die level test time Tm were pre-computed for all possible TAM widths using the enumerative method $P_{NPAW_enumerate}$ described in [8] for Test-Bus architecture [23]. This was followed by a 3D TAM design using the ILP model described above. As both of these methods are based on ILP, the solution is always optimal.

Table 4-II: Test Time Comparison for SIC 1 using Uni- and Simultaneous Bi-directional Signaling

P_{max}	TSV_{max}	Uni-directional			Simultaneous bi-directional			% Δ TAT
		TAT (cycles)	Schedule	Channel Allocation	TAT (cycles)	Schedule	Channel Allocation	
20	200	5621203	'2|3-1-4|5'	'[10 7 3 9 1]'	3211830	'5-2-3|4-1'	'[18 18 10 8 18]'	42.86
30	200	3779806	'2|4-1|3-5'	'[12 11 3 4 15]'	2077911	'1|4-2|3|5'	'[23 19 8 5 1]'	45.03
40	**200**	2851522	'2|3-1-4|5'	'[20 14 6 18 2]'	**1541512**	**'1|2-3|4|5'**	**'[24 14 20 16 2]'**	45.94
50	200	2271852	'5-1-2|3|4'	'[25 14 6 5 25]'	1232343	'1|2|3|4-5'	'[23 14 6 5 48]'	45.76
60	200	1929521	'1|4-2|3|5'	'[25 20 9 5 1]'	1021037	'1|2|3|4|5'	'[28 16 7 6 1]'	47.08
70	200	1643683	'2|3|4|5-1'	'[35 19 8 7 1]'	863264	'1|2|3|4|5'	'[33 19 8 7 1]'	47.48
80	200	1454705	'1|2|3|4-5'	'[20 11 5 4 25]'	761220	'1|2|3|4|5'	'[37 23 9 8 1]'	47.67
90	200	1312582	'1-2|3|4|5'	'[45 25 10 9 1]'	688141	'1|2|3|4|5'	'[41 27 10 9 1]'	47.57
100	200	1150181	'1|2|3|4-5'	'[25 14 6 5 25]'	629794	'1|2|3|4|5'	'[45 30 11 10 2]'	45.24
110	200	1079197	'1|2|3|4|5'	'[26 15 7 6 1]'	579846	'1|2|3|4|5'	'[49 32 13 12 2]'	46.27
120	200	992118	'5-1|2|3|4'	'[29 17 7 7 23]'	545324	'1|2|3|4|5'	'[52 38 13 13 2]'	45.03
130	200	901726	'1|2|3|4|5'	'[31 18 8 7 1]'	545324	'1|2|3|4|5'	'[61 35 14 14 2]'	39.52
140	200	836701	'1|2|3|4|5'	'[34 20 8 7 1]'	545324	'1|2|3|4|5'	'[60 35 13 24 2]'	34.82
150	200	790304	'1|2|3|4|5'	'[36 21 9 8 1]'	545324	'1|2|3|4|5'	'[52 65 16 13 2]'	31.00
160	200	735361	'1|2|3|4|5'	'[38 23 10 8 1]'	545324	'1|2|3|4|5'	'[57 65 13 18 2]'	25.84

The optimization solution returns the TAM for the SIC including a 3D Schedule, 2D Schedule, and the resultant minimum possible test time (TAT). While the 3D scheduling used in this article has been described above, the 2D TAM design has been based on Test-Bus architecture [23]. The 2D and 3D scheduling policy and the test time calculation method used in this paper is further elaborated with the help of an example in the following paragraph.

Consider the third row of Table 4-II, which gives the optimal solution for SIC1 with 40 Pins and 200 TSVs. The optimal 3D Schedule for SBS is 1|2-3|4|5 with TAM widths of 24,14,20,16 and 2 channels, respectively, for dies 1 through 5. This implies that the 3D schedule has two sessions (separated by '-'); in session 1, Dies 1 and 2 will be tested in parallel using TAM widths of 24 and 14 channels, respectively, followed by session 2, in which the TAM will be reconfigured to access dies 3, 4 and 5 with 20, 16 and 2 channels respectively. The TAM width allocated to a die in a 3D schedule is further divided into sub-TAMs to form a 2D schedule. For instance, the 2D Schedules for all dies in SIC1 of the above example (row 3 of Table 4-II) are shown in Table 4-III. A TAM width of 16 channels

Table 4-III: Die level TAM for SIC 1 with 40 pins and 200 TSVs using SBS

	Die Level Schedule	Channels	T_{die}
Die1	[5-7-10-12-16-18-24-25-26-17-29-32] || [2-3-6-8-13-14-28-30] || [1-4-9-11-15-17-19-20-21-22-23-31]	7,8,9	1167503
Die2	[1-4-5-8-11-15-16] || [2-6-7-12-14-17-19] || [3-9-10-13-18]	2,5,7	1144129
Die3	[2-3-8-10-13-14-16-17-19-20-22-23-24-27] || [7-9-11-12-18-21-25-28] || [1-4-5-6-15-26]	4,5,11	368546
Die4	[1] || [2-3-4]	8,8	374009
Die 5	[2-3-4-5-7-9] || [1-6-8-10]	1,1	332743

was allocated to Die4, for which the optimal 2D Schedule is '[1] || [2-3-4]' with sub-TAMs of {8,8}. This Implies that the 16 channels allocated to the die are further divided into two sub-TAMs of 8 channels each. Sub-TAM 1 connects exclusively to core 1, and sub-TAM2 connects cores 2,3 and 4 in series. The cores within a sub-TAM are tested sequentially, and the test time of the sub-TAM is the sum of all core test time, calculated using the method in [24] for UDS and [1] for SBS. As all the sub-TAMs run tests simultaneously, the test time of the die Tm is the maximum among all sub-TAMs. In this way, for the given 2D schedule, the test time for Die 4 was 374009 cycles. T_{die} for the remaining dies is calculated similarly (Table 4-III), which gives the TAT of 1541512 cycles for the given 3D test Schedule.

Table 4-II shows the exact solutions for the TAT and the relevant schedule and channel allocation of SIC 1 using conventional and SBS based TAM when the number of pins of the bottom die is varied from 20 to 160 in steps of 10, and the global TSV limit is 200. It is evident that SBS offers significant TAT reduction; however, it may be noted that as the number of test-pins is increased, while the TAT in the case of both conventional and SBS TAMs decreases (as expected), the advantage offered by SBS decreases. The effect of variation of both P_{max} and TSV_{max} on the %TAT improvement for SIC 1 is shown in Fig. 4-3. A maximum improvement of 53.6% could be observed when P_{max}= TSV_{max} = 60 for this particular SIC. Below this number, the advantage of using SBS decrease to 19.6% (at P_{max} =20, TSV_{max}=20). This is because when the pin and TSV count is

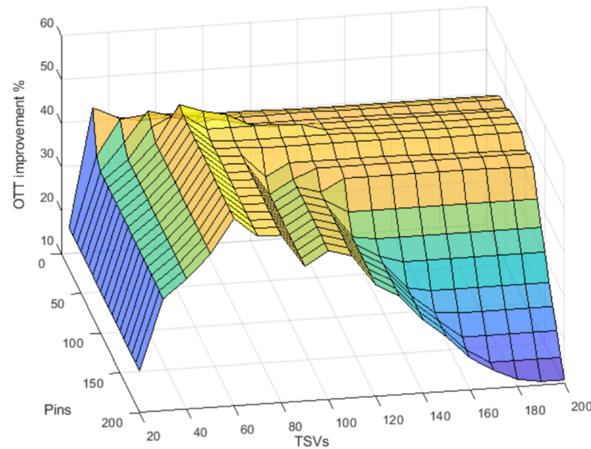


Figure 4-3: % improvement in TAT of SIC 1 with varying Pins and TSV limits

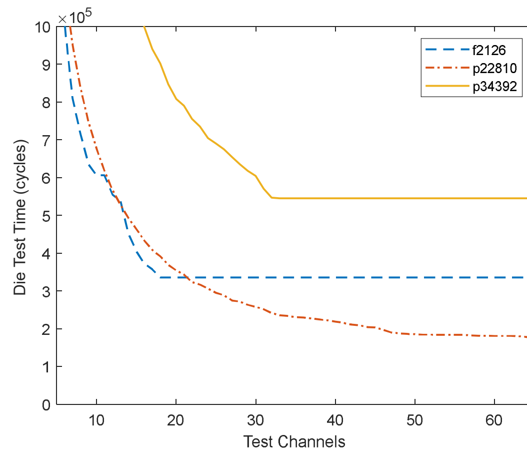


Figure 4-4: Variation of the die test time with increasing number of test channels

too low, the reference generation overheads become significant (in this case, 2 pins and 8 TSVs).

Fig. 4-3 also shows a decrease in the percent improvement using SBS as the P_{max} and TSV_{max} are increased (10.8% at x=y=200). This could be explained using Fig. 4-4 in which the test times of 3 SoCs (f2126, p22810 and p34392 of the ITC'02 benchmarks) are shown against the number of channels available for testing. It can be seen that as the number of test channels is increased, the test time of the die decreases until a certain point is reached, after which there is no further decrease. In the literature, this is commonly referred to as the Pareto-optimal point, at which the T_m does not decrease with the increase in the number of test channels and is constrained by the length of the longest scan chain in the die cores. When using SBS, the number of test channels increase quickly with

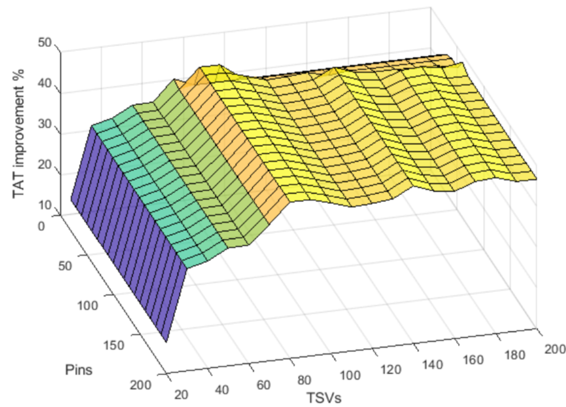


Figure 4-5: % improvement in TAT of SIC 2 with varying Pins and TSV limits

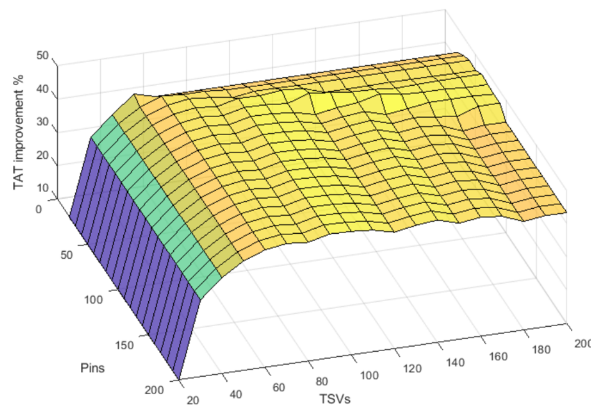


Figure 4-6: % improvement in TAT of SIC 3 with varying Pins and TSV limits

increasing pins/TSVs and reaches the Pareto-optimal point much earlier. However, as pins/TSVs are increased further, the conventional TAM scheme also reaches the Pareto-optimal point, at which both schemes will have the same testing time. It may also be mentioned that at this point, the TAT of SBS may even be slightly higher due to the inclusion of die-level bypass flip-flops.

Fig. 4-5 and Fig. 4-6 show the TAT improvement using SBS for SIC 2 and 3. In both cases, a maximum improvement of around 48% percent could be observed. Unlike SIC1, both SIC 2 and 3 do not exhibit the Pareto-optimality effect at TSV_{max}=P_{max}=200 (Pareto-optimal point does exist well beyond TSV_{max}=P_{max} =200 but not shown for clarity). This is because SIC1 has the most complex die (p93791) placed at the bottom and benefitted directly from the increase in pin count as well as the TSV count. However, SIC 2 and 3 have the most complex die placed on top and the middle, and require a relatively large increase in TSVs

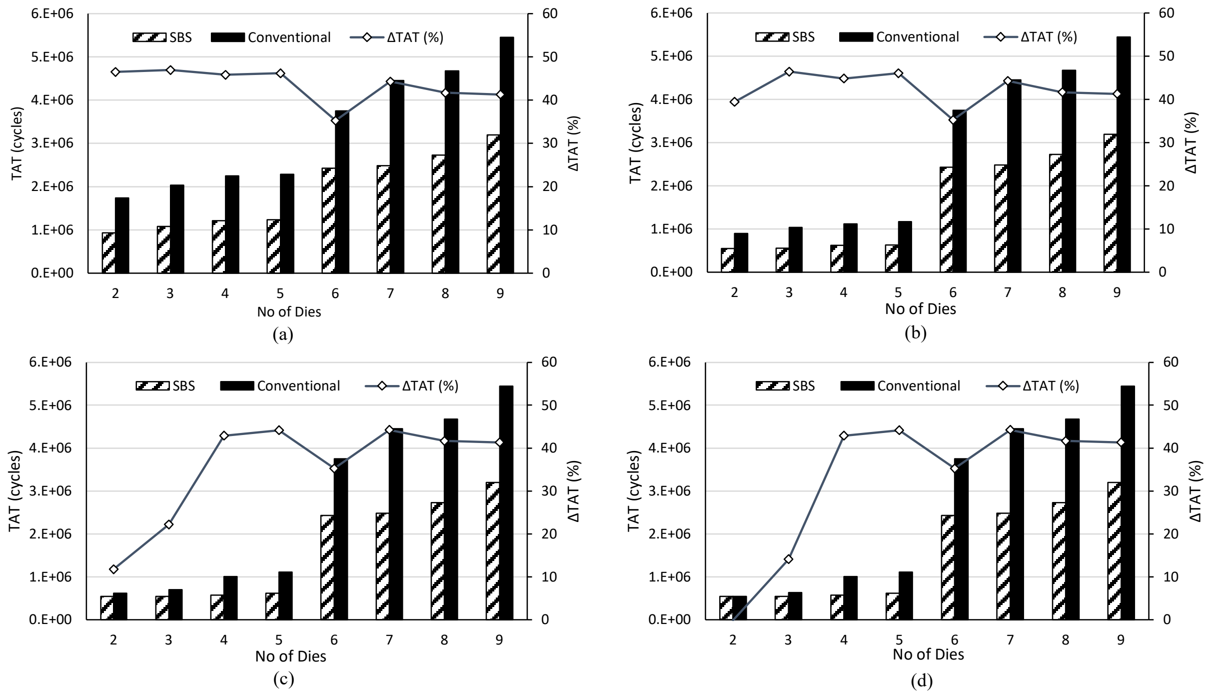


Figure 4-7: Comparison of TAT and the percent TAT difference ΔTAT (%) using Conventional and SBS based TAMs with increasing number of dies in SIC 4. Global TSV limit has been fixed to 100 and Pins equal to (a) 50 (b)100 (c) 150 and (d) 200

to allocate more channels to reduce Tm. For example, to increase 1 test-channel for die 5, which is the most complex die in SIC2, at least 4 TSVs are required to be added for SBS TAM (and 8 for UDS TAM). On the contrary, if the TAT limiting complex die is the third die as in SIC3, an increase in only 2 TSVs (4 for UDS TAM) would deliver the same result. Therefore, if complex dies are higher up in the stack, a higher number of TSVs would be required to bring down the test time to the Pareto-optimal point.

The average improvement in TAT over the range of Pins and TSVs considered is 40.5% for SIC1, 42.4% for SIC3, and 43% for SIC2.

In order to study the offered improvement in TAT using SBS with the increasing number of dies (and hence the complexity) of the SIC, the number of dies in SIC4 was incremented from 2 to 9 in a single die step. The global TSV limit was fixed to 100, and the percent improvement in TAT (ΔTAT %) when the number pins are 50, 100, 150, and 200 are shown in Fig. 4-7(a) through (d), respectively. It is evident that as the number of dies increase, ΔTAT(%) also increases with the SBS approach. The relatively lesser improvement observed for dies 2 and 3 when the number of pins equals 150 in Fig. 4-7(c) and 200 in Fig. 4-7(d) is because

both TAM design schemes are operating in the Pareto-optimal region. Moreover, the addition of the 6th die which is q12710 of the ITC'02 benchmark circuits, causes the relative improvement to dip down to 34% in all cases. This is due to the fact that q12710 SoC has only 13 scan-chains of lengths ranging from 413 to 1689 bits. Therefore, the test time of the SoC is now constrained by the length of the longest scan- chain of 1689 bits and remains the same as the TAM width is increased beyond 12 channels. Moreover, q12710 is placed high up in the stack and quickly becomes the source of TSV constraint and dominates the test time for the entire stack. In all other instances, SBS offers a significant improvement of up to 46%, and hence it is clear that this scheme scales well with the increasing SIC complexity. Also, the average ∆TAT(%) for the four cases of Fig. 4-7(a) through (d) are 43.5%, 42.4%, 35.4%, and 32.9%, respectively, i.e., the improvement offered by SBS is more pronounced when there is a lesser number of pins, and diminished as the pins are increased. In practical scenarios, the test channels (Pins and TSVs) are mostly very limited and that the advantage of SBS is likely to be more pronounced.

4.6 Conclusion

In this paper, an Integer Linear Programming based Test Access Mechanism optimization methodology was presented, which could be used for both conventional Uni-Directional Signalling and Simultaneous Bi-directional Signalling based designs. Experiments with four handcrafted 3D SICs showed that the test time reduction offered by SBS is dependent on the construction of the 3D Stack. Test time advantage of SBS is relatively higher for stacks that have complex dies higher up in the stack, and hence, are more complex from a testing perspective. Experiments also showed that test time improvements are also a function of the available pins and TSVs and vary between none and up to 53.6%. The advantage of SBS is more significant when the chip terminals are limited and scales well with the increasing number of dies in the SIC.

4.7 Appendix 4-I: Linearization of constraints in section 4.4.2

In Eq. 4-7, the product of the continuous variable x_{lm} and the binary variable O_{mt} can be linearized using standard linearization techniques by introducing an auxiliary variable XO_{lmt}. The elements of variable XO_{lmt} will equate to the product of x_{lm} and O_{mt} for all the sessions if the following constraints are imposed:

$$0 \leq XO_{lmt} \leq x_{lm} \qquad \forall l \leq m \; \forall m, t \qquad \textbf{Eq 4-19}$$

$$XO_{lmt} - \alpha \;\; O_{mt} \leq 0 \qquad \forall l \leq m \; \forall m, t \qquad \textbf{Eq 4-20}$$

$$XO_{lmt} - \; x_{lm} + (1 - O_{mt}).\, \alpha \geq 0 \qquad \forall l \leq m \; \forall m, t \qquad \textbf{Eq 4-21}$$

Here α is a constant (also called 'Big M' in the linear programming context) which takes the value of the upper bound on x_{lm}, i.e. $\alpha = x_{max}$. The first constraint simply delineates the upper and lower bounds for the individual elements of XO_{lmt}. The second and third constraints ensure that for a given l, m and t, XO_{lmt} is zero if O_{mt} is zero, else XO_{lmt} will be equal to x_{lm}. This can be understood by considering the cases when O_{mt} is either zero or one:

$\boldsymbol{O_{mt}}$ **is zero:** In this case, the second constraint reduces to $XO_{lmt} \leq 0$. As the first constraint defines the lower bound on XO_{lmt} to be zero, XO_{lmt} cannot be a negative number, hence $XO_{lmt} \leq 0$ can only be satisfied if XO_{lmt} is assigned the value of zero. The third constraint reduces to $XO_{lmt} \geq x_{lm} - \alpha$, since $\alpha = x_{max}$ (the maximum possible value for x_{lm}), the right-hand side either equates to a negative number or zero and becomes a redundant constraint that remains satisfied regardless of the value assigned to XO_{lmt}. Therefore, to satisfy all constraints XO_{lmt} must be zero.

$\boldsymbol{O_{mt}}$ **is one:** In the second case when O_{mt} equals 1, the constraints are required to ensure that XO_{lmt} must equate to x_{lm}. In this case, the third constraint reduces to $XO_{lmt} \geq x_{lm}$ and when read in conjunction with the first constraint, $XO_{lmt} \leq x_{lm}$, can only be satisfied if XO_{lmt} takes the value of x_{lm}. The second constraint reduces to $XO_{lmt} \leq \alpha$, which is always satisfied and becomes a redundant constraint.

Equation 4-7 can now be re-written in its linear form as:

$$PMAP_l^t = \sum_{m=l}^{M} XO_{lmt} \cdot \frac{2}{s} \quad \forall t, \forall l \leq m \qquad \textbf{Eq 4-22}$$

To linearize the max function in Eq 4-8, an auxiliary binary variable c_{lt}, is introduced which is defined for every l and t and introduce a Big-M constant β such that $\beta = \max(P_{max}, TSV_{max})$. Following linearization constraints are now added:

$$PMAP_l \geq PMAP_l^t \qquad \forall l \ \forall t \qquad \textbf{Eq 4-23}$$

$$PMAP_l \leq PMAP_l^t + (1 - c_{lt}).\beta \qquad \forall l \ \forall t \qquad \textbf{Eq 4-24}$$

$$\text{Where} \quad \sum_{t=1}^{M} c_{lt} = 1 \quad \forall l \qquad \textbf{Eq 4-25}$$

The first constraint simply states that the value assigned to $PMAP_l$ must be at least as high as the highest among all the sessions. For instance, if there are 2 x sessions, then $PMAP_l \geq PMAP_l^1$ and $PMAP_l \geq PMAP_l^2$. This constraint defines the lower bound on the value $PMAP_l$ can take; however, it remains satisfied for all values greater than that. The second constraint defines the upper bound on the $PMAP_l$, such that it doesn't exceed the maximum value among all sessions. This is achieved by ensuring that the auxiliary variable $c_{lt} = 1$ only for session t which has the highest value, and zero for the others.

The second constraint can only be satisfied if c_{lt} is 1 only for the highest value of $PMAP_l^t$. Continuing from the previous example where there are only 2 x sessions, assuming the true case is $PMAP_l^1 \geq PMAP_l^2$ such that $PMAP_l^1 = \max_t PMAP_l^t$, we have the following two possible cases:

Case 1: $c_{l1} = 1$, $c_{l2} = 0$: In this case, Eq. 4-16 reduces to $PMAP_l \leq PMAP_l^1$ and $PMAP_l \leq PMAP_l^2 + \beta$. Both constraints can only be satisfied for $PMAP_l = PMAP_l^1$ (which is indeed the max).

Case 2: $c_{l1} = 0, c_{l2} = 1$: In this case, Eq. 4-16 reduces to $PMAP_l \leq PMAP_l^1 + \beta$ and $PMAP_l \leq PMAP_l^2$. The constraint $PMAP_l \leq PMAP_l^1 + \beta$ is redundant and remains satisfied for both $PMAP_l = PMAP_l^1$ and $PMAP_l = PMAP_l^2$. However, the constraint $PMAP_l \leq PMAP_l^2$ cannot be satisfied as $PMAP_l^2 \leq PMAP_l^1$ and the first constraint (Eq. 4-15) requires that $PMAP_l \geq PMAP_l^1$.

4.8 References:

[1] I. A. Soomro, M. Samie, and I. K. Jennions, “Test Time Reduction of 3-D Stacked ICs Using Ternary Coded Simultaneous Bidirectional Signaling in Parallel Test Ports,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 39, no. 12, pp. 5225–5237, 2020.

[2] N. A. Touba, “Survey of Test Vector Compression Techniques,” *IEEE Des. Test Comput.*, vol. 23, no. 4, pp. 294–303, Apr. 2006.

[3] J. Knechtel, O. Sinanoglu, I. (Abe) M. Elfadel, J. Lienig, and C. C. N. Sze, “Large-Scale 3D Chips: Challenges and Solutions for Design Automation, Testing, and Trustworthy Integration,” *IPSJ Trans. Syst. LSI Des. Methodol.*, vol. 10, no. 0, pp. 45–62, 2017.

[4] H.-H. S. Lee and K. Chakrabarty, “Test Challenges for 3D Integrated Circuits,” *IEEE Des. Test Comput.*, vol. 26, no. 5, pp. 26–35, Sep. 2009.

[5] E. J. Marinissen, “Challenges and emerging solutions in testing TSV-based 2 1 over 2D- and 3D-stacked ICs,” in *2012 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2012, pp. 1277–1282.

[6] S. Panth and S. K. Lim, “Probe-Pad Placement for Prebond Test of 3-D ICs,” *IEEE Trans. Components, Packag. Manuf. Technol.*, vol. 6, no. 4, pp. 637–644, Apr. 2016.

[7] International Technology Roadmap for Semiconductors (ITRS), “ITRS 2.0 HETEROGENEOUS INTEGRATION CHAPTER: 2015,” 2015.

[8] V. Iyengar, K. Chakrabarty, and E. J. Marinissen, “Test Wrapper and Test Access Mechanism Co-Optimization for System-on-Chip,” *J. Electron. Test. Theory Appl.*, vol. 18, pp. 213–230, 2002.

[9] Yu Huang *et al.*, “Optimal core wrapper width selection and SOC test scheduling based on 3-D bin packing algorithm,” in *Proceedings. International Test Conference*, 2002, pp. 74–82.

[10] S. K. Goel and E. J. Marinissen, “Control-aware test architecture design for modular SOC testing,” in *The Eighth IEEE European Test Workshop, 2003. Proceedings.*, 2003, vol. 2003-Janua, pp. 57–62.

[11] E. Larsson, K. Arvidsson, H. Fujiwara, and Z. Peng, “Efficient Test Solutions for Core-Based Designs,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 23, no. 5, pp. 758–775, May 2004.

[12] K. Chakrabarty, “Design of system-on-a-chip test access architectures under place-and-route and power constraints,” in *Proceedings of the 37th conference on Design automation - DAC ’00*, 2000, pp. 432–437.

[13] C. Yao, K. K. Saluja, and P. Ramanathan, “Power and Thermal Constrained Test Scheduling Under Deep Submicron Technologies,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 30, no. 2, pp. 317–322, Feb. 2011.

[14] X. Wu, Yibo Chen, K. Chakrabarty, and Yuan Xie, “Test-access mechanism optimization for core-based three-dimensional SOCs,” in *2008 IEEE International Conference on Computer Design*, 2008, pp. 212–218.

[15] X. Wu, P. Falkenstern, K. Chakrabarty, and Y. Xie, “Scan-chain design and optimization for three-dimensional integrated circuits,” *ACM J. Emerg. Technol. Comput. Syst.*, vol. 5, no. 2, pp. 1–26, Jul. 2009.

[16] S. Deutsch, K. Chakrabarty, and E. J. Marinissen, “Robust Optimization of Test-Access Architectures Under Realistic Scenarios,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 34, no. 11, pp. 1873–1884, Nov. 2015.

[17] B. Noia, K. Chakrabarty, S. K. Goel, E. J. Marinissen, and J. Verbree, “Test-Architecture Optimization and Test Scheduling for TSV-Based 3-D Stacked ICs,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 30, no. 11, pp. 1705–1718, Nov. 2011.

[18] “IEEE Standard Testability Method for Embedded Core-based Integrated Circuits,” *IEEE Std 1500-2005*, pp. 1–136, 2005.

[19] V. Iyengar, K. Chakrabarty, and E. J. Marinissen, “Recent advances in test planning for modular testing of core-based SOCs,” in *Proceedings of the 11th Asian Test Symposium, 2002. (ATS ’02).*, 2002, pp. 320–325.

[20] J. Aerts and E. J. Marinissen, “Scan chain design for test time reduction in core-based ICs,” *IEEE Int. Test Conf.*, pp. 448–457, 1998.

[21] E. G. Coffman, Jr., M. R. Garey, D. S. Johnson, and R. E. Tarjan, “Performance Bounds for Level-Oriented Two-Dimensional Packing Algorithms,” *SIAM J. Comput.*, vol. 9, no. 4, pp. 808–826, Nov. 1980.

[22] E. J. Marinissen, V. Iyengar, and K. Chakrabarty, “A set of benchmarks for modular testing of SOCs,” in *Proceedings. International Test Conference*, 2002, pp. 519–528.

[23] P. Varma and B. Bhatia, “A structured test re-use methodology for core-based system chips,” in *Proceedings International Test Conference 1998 (IEEE Cat. No.98CH36270)*, 2002, pp. 294–302.

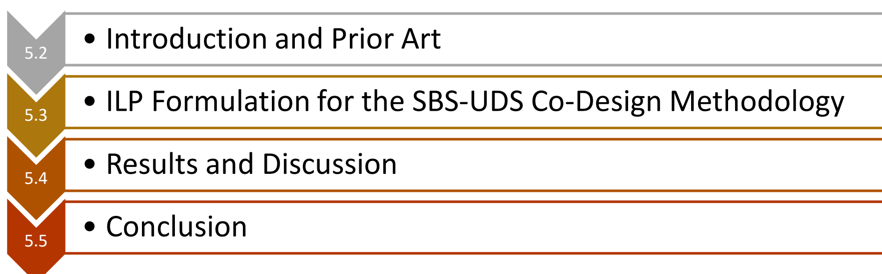
[24] E. J. Marinissen, S. K. Goel, and M. Lousberg, “Wrapper design for embedded core test,” in *Proceedings International Test Conference 2000 (IEEE Cat. No.00CH37159)*, 2000, pp. 911–920.

5 AN INTEGRATED APPROACH FOR 3D SIC TESTING USING SIMULTANEOUS BIDIRECTIONAL AND CONVENTIONAL UNI-DIRECTIONAL SIGNALLING

5.1 Abstract

In Chapter 4, a TAM design and optimization formulation was presented, which could be used for TAM designs based entirely on either UDS or SBS. However, both methods have their strengths and weaknesses. On one hand, UDS involves lesser design effort and is power efficient; on the other hand, SBS could be used to either double the number of test channels or reduce the chip terminal count used in testing. Clearly, combining both SBS and UDS opens up greater avenues for efficient TAM designs taking advantage of the strengths of either method where required. In this chapter, these limitations are addressed by proposing an optimization framework for ***co-design* of** TAM using a combination of SBS and UDS. Additionally, the focus of chapter 4 was solely on test time reduction, whereas the resources required for the TAM (in terms of Chip Terminals) were ignored. In this Chapter, a ***multi-objective formulation*** is proposed that allows a balance between the test time reduction and chip terminal utilization.

In section 5.2, an introduction to the test time and resource-aware TAM design philosophy is provided along with the prior art in this area. In section 5.3, an Integer Linear Programming based formulation is proposed for the SBS-UDS co-design problem while also accounting for chip terminal minimization. Section 5.4 discusses the results of the co-design methodology using 3 handcrafted SICs based on ITC'02 benchmarks. The results suggest that the proposed method out-performs the TAMs based entirely on SBS or UDS in terms of TAT reduction while providing an optimal balance between CTs used. Section 5.5 concludes the chapter.



5.2 Introduction and Prior Art

In order to perform scan-based testing, a Test Access Mechanism (TAM) is required such that the test patterns can be transported to and from the tester. A

TAM is composed of test channels, with every channel capable of serially shifting a single bit of test data. The number of test channels dictates the width of the TAM and a larger TAM allows more test data to be shifted per clock cycle. The added parallelism reduces the test vector shifting time, which is the dominant factor in the overall Test Application Time (TAT). Nonetheless, every test channel requires added Chip Terminals (CTs), which includes device pins (for the chip to tester communication), and in the case of 3D Stacked Integrated Circuits (SICs), the Through Silicon Vias (TSVs) for intercommunication between dies. However, CTs consume a significant chip area and are available in limited quantities. Therefore, the test channels must be appropriately allocated to every die, and in turn to every core. A large number of design choices exist having different implications on the test time. Therefore, TAM design is an optimization problem which is known to NP-Hard and has been a focus of significant research over the years.

While testing is vital to ensure reliable operation of the IC, it inevitably takes up on-chip resources that could otherwise be used for functional purposes and given that testing is mostly a one-off requirement, the test resources are considered undesirable overheads. These overheads may include the area consumed by the added circuitry for test purpose (such as core wrapper and boundary scan registers), the routing overheads, as well as the power consumption of the chip. Therefore, A TAM design and optimization methodology focusing singularly on optimal channel allocation for test time reduction is not a holistic scenario and the resources allocated for the test purpose must also be accounted for. However, the complexity of the TAM design increases significantly as more dependencies are factored in. Most of the TAM design algorithms simplify the problem by considering only a sub-set of cost factors. In practice, test hardware area and power are the biggest concerns as the former reduces available on-chip real estate for ICs main functionality, whereas the latter causes the chip temperature to increase which must be kept within limits. Several factors affect the area overheads, however for 3D SICs the number of TSVs designated for test purpose have the most significant impact [1][2][3]. Unlike conventional landing pads

created only on the top metal layer for probing or bonding, TSVs pass through the entire substrate, and therefore, the active transistor layers are not available at the TSV site. Moreover, the keep out zones add further chip area loss, and consequently, the number of test TSVs must be kept minimal. The chip's power budget, on the other hand, is dictated by the die thermal dissipation capability. During testing, the increased switching activity due to the shift and capture cycles occurring simultaneously throughout the core cause significantly greater power consumption than the functional mode operation [4].

Clearly, if there are unlimited TSVs and no power restrictions, the test time could be minimal; otherwise, some form of the trade-off must be made among these factors, which is the underlying concept of multi-objective optimization. Various authors have proposed methods to efficiently design TAMs while accounting for essential facets in addition to the main objective of test time reduction. The authors in [5][4][6] proposed power-aware designs, methods to reduce routing overheads were proposed in [4][6], whereas the authors in [4][7][8] proposed test mechanisms while accounting for thermal constraints. The authors in [9] presented temperature aware scheduling methods and studied the effect of ordering test vectors on the chip's temperature. The authors in [10] proposed a pin-aware optimization by re-using the existing NoC for test access. While the above works were related to on-chip resources, the authors in [11] presented a TAM optimization method that efficiently utilizes the tester resources as well. The authors showed that increasing the TAM widths of cores may not result in test time reduction but might increase the tester memory depth requirements, which could be otherwise utilized to test other dies in parallel.

The insights developed for resource-aware optimization literature produced for 2D Chips may be extended to the 3D ICs, with some modifications according for the third dimension. However, some challenges are unique to TAM design for 3D ICs. The 3D SICs require multiple test instances in the form of pre-, mid- and post bond tests. The authors in [12] have discussed that optimization must account for all test insertions/ instances to ensure efficient utilization of test resources. The authors proposed bandwidth adapters to exploit available bandwidth at different

test instances and used dynamic programming to design TAM. Being computation friendly, meta-heuristic optimization algorithms have been frequently employed to solve TAM design problem. The authors in [13] proposed Cukoo Search for test time and TSV minimization, whereas for power-aware TAM designs, Genetic Algorithms (GA) [14] and Strength Pareto Evolutionary (SPE) optimization algorithm [15] have been employed. In [16][17] , the authors proposed heuristics for test time and TSV minimization by considering the I/O Cell bindings associated with scan chains. In [18], the authors approach the problem of test TSV minimisation, based on the argument that the addition of wrapper cells at every TSV adds area overheads, but increase fault coverage. The authors also proposed methods to optimize the use of test TSVs against benefits in terms of fault coverage and test time. In [19], the authors proposed an ILP formulation for 3D SICs under thermal and wire length minimization constraints. Apart from TAT reduction, the selection of optimal test sets and test insertion in the overall 3D chip manufacturing has also been shown to be an important optimization concern [20] [21].

The authors in [22] proposed a TAM design method based on Simultaneous Bidirectional Signaling (SBS), using which a test channel could be designed using a single chip terminal. Using this method, significant test time improvement if up to 53% were reported. However, the test time improvement was reported considering the extreme cases in which the whole TAM was designed either as SBS or UDS. It may be noticed that a TAM does not have to be designed solely using UDS or SBS; instead, a combination of these may be used. This may be particularly important given that the SBS design incurs additional circuitry and power overheads, and therefore the designer may consider penalizing the use of SBS unless there is a considerable reduction in test time. Moreover, under certain conditions the designer may opt to force UDS design at certain locations in the stack, such as if some of the dies in the stack are already fabricated or wrapped (hard or firm dies). This scenario poses a co-design problem in which test time is required to be minimized but using a combination of SBS and UDS considering the trade-offs of preferring either scheme. Another limitation of the work in [22] is

that the optimization method was focused solely on TAT reduction; however, SBS could also be used to reduce the chip terminals used for testing.

This article addresses the above limitations by proposing an optimization model for the SBS-UDS co-design problem while also accounting for chip terminal minimization. In Chapter 4, the TAM design model presented in [23] was adapted for SBS-based TAMs. In this work, we extend the model to the co-optimization problem such that the design choice for the test channel type (SBS and UDS) becomes the decision of the optimization problem. This requires modelling several considerations within the optimization problem, such as the cost of using SBS in terms of power consumption and the benefits of using SBS in terms of additional test channels. Further, in addition to the original objective of test time reduction, we extend the problem by including chip terminal reduction as a secondary objective in this paper. This ensures that the gains in test time reduction are weighted against the chip terminal utilization such that an optimal balance is achieved between the test cost and test resources. In the next section, the SBS-UDS co-design problem is formulated as an Integer Linear Program.

5.3 ILP Formulation for the SBS-UDS Co-Design Methodology

For the SBS-UDS co-design, the overall problem of Chapter 4 could be re-stated as: Given a SIC with M number of dies, each die has N cores such that N_b denotes the total number of cores in die b. Each core has certain I/Os, scan chains of specific lengths, and the number of test patterns to be applied. Let P_{max} denote the maximum number of pins available at the lowermost die, TSV_{max} denote the global TSV limit and C_{max} denote the maximum cost budget. *If the Test Access Mechanism can be designed using a combination of UDS and SBS chip terminals, determine the optimal TAM design that minimises the Test Application Time (TAT) as well as Chip Terminals (CTs) by*:

> a) Finding the 3D schedule (how dies should be tested) by optimal allocation of Test Channels to each die such that the upper limit on P_{max}, TSV_{max} and cost budget C_{max} is note exceeded.

b) Finding the optimal 2D schedule (how cores within the die should be tested) such that the allocated number of channels dictated by the SIC level schedule (in part a) is not exceeded.

In Chapter 4, it was shown that the Test Application Time (TAT) of a 3D SIC is the sum of the test times of individual sessions, whereas the test time of the session is the maximum test time among the dies (T_m) being tested in the session:

$$TAT = \sum_{t=1}^{M} \max_{m \epsilon t} (T_m) \qquad \textbf{Eq 5-1}$$

Here, m refers to the index of the die in the stack and t indicates a particular test session, that is if there are M dies in the stack, the range of m and t is given by $m: m \epsilon \{1,2,..,M\}$ and $t: t \epsilon \{1,2,..,M\}$. T_m is a function of the channel width allocated to the die, which is denoted by the integer variable x. In Chapter 4, x was only defined for every layer l and die position m, however, for the SBS-UDS co-design problem an additional index to indicate the channel type, i.e. UDS or SBS is to be defined. Let a variable $s \epsilon \{1,2\}$ indicate whether a chip terminal (Pin or TSV) is designed as UDS (s = 1) or SBS (s = 2); x_{lsm} is re-defined s.t $x_{lsm}: 0 \leq x_{lsm} \leq w_{max}, \forall m \epsilon \{1,2,..,M\}, \forall l \epsilon \{1,2,..,M\}, \forall s \epsilon \{1,2\}$. It may be noted that s was similarly used to denote UDS and SBS channel types in Chapter 4 and as a scripting variable whose value remained constant, whereas, in this formulation, it is a decision variable that is to be assigned by the optimization algorithm.

Similar to Chapter 4, a variable TAM_{lm} is used to indicate the TAM width dedicated for each die, defined only for layers below the die. As the TAM width is the sum of UDS and SBS test channels, TAM_{lm} is the sum of variable x along s. The constraints on TAM are defined such that: a. In order for the tester to communicate with a die m, TAM must be designed at all layers between the die and tester (i.e. for all $l \leq m$), b. There must be at least a single test channel for each die, and c. TAM_{lm} for each die is upper bounded by the maximum supported TAM width, denoted by $w_{max,m}$. The above requirements are respectively modelled as:

$$TAM_{lm} = \sum_{s=1}^{2} x_{lsm} \ \forall m, \forall l \leq m \qquad \textbf{Eq 5-2}$$

$$TAM_{(l-1)m} \geq \ TAM_{lm} \quad \forall m \geq 2 \,, \forall\, 2 \leq l \leq m \qquad \textbf{Eq 5-3}$$

$$TAM_{lm} \geq 1 \ \forall m, \forall l \leq m \qquad \textbf{Eq 5-4}$$

$$TAM_{mm} \leq \ w_{max,m} \quad \forall m \qquad \textbf{Eq 5-5}$$

The now redefined 3-dimensional decision variable x will contain the optimal test requirements '*for each die'* in terms of test channels required and the composition of the test channels, i.e. the number of UDS and SBS channels at every layer. Accordingly, to define the relationship between the test channels and the chip terminals, the decision variable $PMAP$ is redefined to include the channel type at each layer by adding an index s, i.e $PMAP_l$ to $PMAP_{ls}$. Furthermore, as $PMAP_{ls}$ is the final TAM design which is required to satisfy the test requirements of all individual sessions, the relationship between the final pinmap $PMAP_{ls}$ and the pinmap $PMAP_{ls}^t$ for session t, is as follows:

$$PMAP_{ls} = \max_{t} PMAP_{ls}^{t} \quad \forall l \ \forall s \qquad \textbf{Eq 5-6}$$

The chip terminals needed for a test session require the information regarding: a. the die(s) being tested within the session t, and b. the channels required for those dies. The latter is already modelled using variable x, to model the former, we use a binary variable O_{mt}, defined for every die m and session t, to indicate which dies are being tested within a session. The following constraint ensures that O_{mt} equals to 1 if and only if a die m is tested in session t, and 0 otherwise:

$$\sum_{t=1}^{M} O_{mt} = 1 \ \forall m \qquad \textbf{Eq 5-7}$$

Using the scheduling information contained in O, and the test channels requirement of a given die in x, the CT requirement of individual sessions could be expressed as follows:

$$PMAP_{ls}^{t} = \sum_{m=l}^{M} x_{lsm} . O_{mt} . \frac{2}{s} \quad \forall s, \forall t, \forall l \leq m \qquad \textbf{Eq 5-8}$$

In the SBS-UDS co-design problem, the decision of whether to allocate additional CTs at a given layer for reference sharing would depend on if there are any SBS test channels at or above that layer. Hence the reference CT overheads are not constant and must be defined as an optimization variable within the problem. This is achieved by introducing a binary decision variable R_l, defined for every layer such that R_l= 0 if the sum of all SBS channels at or above the layer is zero, and R_l= 1 otherwise. This formulation follows the requirement that if reference voltages are to be sourced from the tester, then it must traverse through all the dies at and below the layer at which SBS is required to be designed. Using $PMAP$ and R_l, the constraint on the number of available pins and TSVs can now be imposed as:

$$\sum_{s=1}^{2} \left(PMAP_{ls} + 2. R_l . (s-1) \right) \leq P_{max} \quad l = 1 \qquad \textbf{Eq 5-9}$$

$$\sum_{l=2}^{M} \sum_{s=1}^{2} \left(PMAP_{ls} + 2. R_l . (s-1) \right) \leq TSV_{max} \qquad \textbf{Eq 5-10}$$

To formulate the behaviour of R_l, following constraints are imposed using the ‘Big M’ method (γ is a large constant s.t. $\gamma = TSV_{max} + P_{max}$, i.e. the maximum possible chip terminals at any layer):

$$\sum_{l'=l}^{M} PMAP_{l's} \geq 1 - \gamma . (1 - R_l) \quad s = 2, \forall l \qquad \textbf{Eq 5-11}$$

$$\sum_{l'=l}^{M} PMAP_{l's} \leq \gamma . R_l \quad s = 2, \forall l \qquad \textbf{Eq 5-12}$$

The above constraints (Eq. 5-11 and Eq. 5-12) ensure that if the sum of SBS channels at or above a layer l ($\sum_{l'=l}^{M} PMAP_{l's}$) is greater than or equal to 1, R_l

must equate to 1 for both the constraints to be satisfied, while the opposite is true if ($\sum_{l'=l}^{M} PMAP_{l's}$) is equal to 0. Consequently, the variable R_l forces the factor $2.R_l.(s-1)$ to zero (in Eq. 5-9 and Eq. 5-10) if there are no SBS channels used at or above that layer.

The costs of designing a chip terminal using SBS and UDS can now be imposed on $PMAP_{ls}$. We express the cost using a variable C_s, where C_1 and C_2 denote the cost of a designed chip terminal using SBS and UDS, respectively, and C_{max} denotes the maximum allowable cost. The problem can now be constrained in terms of cost as:

$$\sum_{s=1}^{2} \sum_{l=1}^{M} PMAP_{ls}.\ C_s \leq C_{max} \qquad \textbf{Eq 5-13}$$

To summarise the formulation until this point, the variable x provides the information of the number and type of test channels to be included at every die interface for a particular die, and the variable O provides the optimal test schedule, i.e. the number of sessions and the subset of dies being tested within each. These variables were later used to extract the channel width dedicated to a die (TAM) and the number of chip terminals required through the entire stack ($PMAP_{ls}$). In the following subsections, the objective function using this information is constructed.

5.3.1 Objective function formulation and linearization

Multi-objective optimization involves a combination of more than one objective, which in this case is: $minimize\ (f_1(x,O),\ f_2(x,O))$, where f_1 is the test time TAT and f_2 is the Chip Terminal (CT) count. There are several methods to formulate the overall (scalarised) objective function using a combination of individual objectives, such as exponential weighting, weighted products and weighted sum [24]. In this case, the objective is formulated as a weighted sum given by:

$$minimize\ \ (1-\varepsilon).CT + \varepsilon.TAT \qquad \textbf{Eq 5-14}$$

Here ε is the scaling factor or weightage that would determine which objective is preferred. It may be noted that the TAT and CT are contradicting objectives; therefore, different solutions will exist depending upon the choice of ε. To decide the appropriate value for ε, the bounds on the individual objectives must be approximated, which are discussed in Appendix 1.

The CT or the chip terminal count is the sum of pins and TSVs utilized by the solution, which is given by the LHS of Eq. 5-9 and Eq. 5-10. Combining these, CT can be written as:

$$CT = \sum_{l=1}^{M} \sum_{s=1}^{2} . \left(PMAP_{ls} + 2. R_l . (s-1) \right) \qquad \textbf{Eq 5-15}$$

It was shown in Chapter 4 that the expression for TAT in Eq 5-1 could be written as below by rewriting the domain of the max function as $\max_{m \epsilon t}(T_m) = \max_{m}(T_m . O_{mt})$, and expressing T_m as the product of T_{mw} and Y_{mw}, i.e.:

$$TAT = \sum_{t=1}^{M} \max_{m} \sum_{w=1}^{w_{max}} T_{mw} . Y_{mw} . O_{mt} \qquad \textbf{Eq 5-16}$$

In this equation, T_{mw} is a look up table indicating all possible values of channel widths $w: w \epsilon \{1,2,..,x_{max}\}$, and Y_{mw}, is a binary variable constrained such that only the wth element is *1* for a given die m, if a TAM width of size w is allocated to that die (by using the constraint $\sum_{w=1}^{w_{max}}(Y_{mw} . w - TAM_{mm}) = 0 \quad \forall m$). The non-linearities in the objective function (the product of Y_{mw} and O_{mt}, and the max function of this product) are linearized similar to Chapter 4 using standard linearization methods as follows:

$$Y_{mw} + O_{mt} - YO_{mwt} \leq 1 \quad \forall m \, \forall w \, \forall t \qquad \textbf{Eq 5-17}$$

$$Y_{mw} + O_{mt} - 2. YO_{mwt} \geq 0 \quad \forall m \, \forall w \, \forall t \qquad \textbf{Eq 5-18}$$

Here YO_{mwt} is an auxiliary binary variable, which takes the value of the product $Y_{mw} . O_{mt}$. Similarly, we use a non-negative continuous integer variable TAT_t to linearize the max function using the following constraint:

$$TAT_t \geq \sum_{w=1}^{w_{max}} T_{mw}.YO_{mwt} \qquad \forall m\, \forall t \qquad \textbf{Eq 5-19}$$

It may be noted that Eq. 5-8 contains the non-linear product $x_{lsm}.O_{mt}$, and Eq. 5-6 contains the max function which is non-linear. Eq. 5-8 can be linearized using the Big-M method, using an auxiliary variable XO and a constant α ($\alpha = x_{max}$) as follows:

$$0 \leq XO_{lsmt} \leq x_{lsm} \qquad \forall l \leq m\, \forall s, m, t \qquad \textbf{Eq 5-20}$$

$$XO_{lsmt} - \alpha \;\; O_{mt} \leq 0 \qquad \forall l \leq m\, \forall s, m, t \qquad \textbf{Eq 5-21}$$

$$XO_{lsmt} - x_{lsm} + (1 - O_{mt}).\alpha \geq 0 \qquad \forall l \leq m\, \forall s, m, t \qquad \textbf{Eq 5-22}$$

Eq. 5-6 is linearized using an auxiliary binary variable c_{lst}, and a Big-M constant β such that $\beta = \max(P_{max}, TSV_{max})$

$$PMAP_{ls} \geq PMAP_{ls}^{t} \qquad \forall l\, \forall s\, \forall t \qquad \textbf{Eq 5-23}$$

$$PMAP_{ls} \leq PMAP_{ls}^{t} + (1 - c_{lst}).\beta \qquad \forall l\, \forall s\, \forall t \qquad \textbf{Eq 5-24}$$

$$\text{Where} \quad \sum_{t=1}^{M} c_{lst} = 1 \quad \forall l\, \forall s \qquad \textbf{Eq 5-25}$$

5.3.2 The Overall ILP Formulation

The overall optimization formulation can now be summarized as:

$$minimize \;\; (1 - \varepsilon).CT + \varepsilon.TAT$$

Subject to constraints:

$$CT = \sum_{l=1}^{M} . \sum_{s=1}^{2} . (PMAP_{ls} + 2.R_l.(s-1))$$

$$TAT = \sum_{t=1}^{M} TAT_t.$$

$$TAM_{lm} = \sum_{s=1}^{2} x_{lsm} \;\; \forall m, \forall l \leq m$$

$$TAM_{(l-1)m} \geq \ TAM_{lm} \qquad \forall m \geq 2\,, \forall\ 2 \leq l \leq m$$

$$TAM_{lm} \geq 1 \ \ \forall m, \forall l \leq m$$

$$TAM_{mm} \leq \ w_{max,m} \qquad \forall m$$

$$\sum_{s=1}^{2} \quad (PMAP_{ls} + 2. R_l. (s-1)) \quad \leq P_{max} \quad l = 1$$

$$\sum_{l=2}^{M} \ \sum_{s=1}^{2} \quad (PMAP_{ls} + 2. R_l. (s-1)) \ \leq TSV_{max}$$

$$\sum_{l'=l}^{M} PMAP_{l's} \geq 1 - \gamma. (1 - R_l) \qquad s = 2, \forall l$$

$$\sum_{l'=l}^{M} PMAP_{l's} \leq \gamma. R_l \qquad s = 2, \forall l$$

$$\sum_{s=1}^{2} \ \sum_{l=1}^{M} PMAP_{ls}.\ C_s \leq C_{max}$$

$$\sum_{t=1}^{M} O_{mt} = 1 \ \forall m$$

$$\sum_{w=1}^{w_{max}} Y_{mw}. w - TAM_{mm} = 0 \quad \forall m$$

$$Y_{mw} + O_{mt} - YO_{mwt} \leq 1 \quad \forall m\ \forall w\ \forall t$$

$$Y_{mw} + O_{mt} - 2. YO_{mwt} \geq 0 \quad \forall m\ \forall w\ \forall t$$

$$TAT_t \geq \ \sum_{w=1}^{w_{max}} T_{mw}. YO_{mwt}. \qquad \forall m\ \forall t$$

$$0 \ \leq \ XO_{lsmt} \ \leq \ x_{lsm} \qquad \forall l \leq m\ \forall s, m, t$$

$$XO_{lsmt} - \alpha \ \ O_{mt} \leq 0 \qquad \forall l \leq m\ \forall s, m, t$$

$$XO_{lsmt} - \ x_{lsm} + (1 - O_{mt}). \alpha \geq 0 \ \ \forall l \leq m\ \forall s, m, t$$

$$PMAP_{ls}^{t} = \ \sum_{m=l}^{M} XO_{lsmt} \ . \frac{2}{s} \quad \forall s, \forall t, \forall l \leq m$$

$$PMAP_{ls} \geq PMAP_{ls}^{t} \qquad \forall l\ \forall s\ \forall t$$

$$PMAP_{ls} \leq \ PMAP_{ls}^{t} + (1 - c_{lst}). M2 \qquad \forall l\ \forall s\ \forall t$$

$$\textstyle\sum_{t=1}^{M} c_{lst} \ = 1 \quad \forall l \forall s$$

5.4 Results and Discussion

Table 5-II: Optimal solutions for SIC 1 using SBS-UDS Co-design (for SIC1, *Pmax*= 10, *TSVmax*=30)

'Cmax'	TAT	schedule	TAM width dies 1 to 5	CT Type Layer 1 to 5	Pins Used	TSVs Used	Power used
10	17259266	'1|4-2|3-5'	'[3 2 1 1 1]'	'[8 6 2 2 2; 0 0 0 0 0]'	8/10	12/30	10/10
20	8154757	'1|2|3|4-5'	'[4 2 1 1 1]'	'[0 8 4 2 2; 10 0 0 0 0]'	10/10	16/30	18.4/20
30	7548985	'1-5-2|3|4'	'[8 4 2 2 2]'	'[0 4 8 4 4; 10 8 0 0 0]'	10/10	28/30	28.2/30
40	7240424	'1-3|5-2|4'	'[8 6 7 2 1]'	'[0 0 4 4 2; 10 10 9 0 0]'	10/10	29/30	34.9/40

Table 5-I: Chip Terminal utilization using ε = 1 and ε = 0.5 (SIC1, *Pmax*= 10, *TSVmax*=30)

$C_{(\max)}$	Single Objective (TAT) ε = 1			Multi Objective (CT+TAT) ε = 0.5			CT reduction
	CT Type	Total UDS CTs	Total SBS CTs	CT Type	Total UDS CTs	Total SBS CTs	
10	'[8 6 2 2 2; 0 0 0 0 0]'	20	0	'[8 6 2 2 2; 0 0 0 0 0]'	20	0	0
20	'[0 8 4 2 2; 10 0 0 0 0]'	16	10	'[0 0 4 2 2; 10 6 0 0 0]'	8	16	2
30	'[0 4 8 4 4; 10 8 0 0 0]'	20	18	'[0 0 8 4 4; 10 10 0 0 0]'	16	20	2
40	'[0 0 4 4 2; 10 10 9 0 0]'	10	29	'[0 0 0 0 2; 10 10 10 4 0]'	2	34	3

Experiments were conducted using three handcrafted 3D SICs using ITC 02 benchmarks SOCs as described in [22] and [23]. The models were solved for optimal solutions using CPLEX, and the run-time was between a few milliseconds to 8 seconds using a desktop machine with a 3.2 GHz Corei5 CPU and 16 GB memory. For the experiments in this article, only power costs have been considered. For the UDS based test channel, a normalized value of 1 is used, whereas for the SBS channel, the power consumption has been approximated to be 30% higher than UDS [22]. The experiments conducted are aimed at identifying the trade-offs of using full TAM design using either as SBS or UDS as compared to a combination of both. Although only power relative costs have been used in the experiments, the results can be used to interpret the generalized situation in which SBS is to be penalized for other factors such as area consumption, routing overheads and design effort.

In section 5.4.1, the results for co-optimization are presented when the objective focuses singularly on TAT reduction. The effects of SBS costs in the TAM design choice in terms of CT type is discussed. Moreover, the effects of the inclusion of the CT reduction objective are studied when the objectives are given equal weightage (ε=0.5), representing the case in which CTs are reduced only if there is no further reduction in TAT. It is demonstrated that the proposed optimization is better than single-objective optimization by removing redundant CTs in the pareto optimal cases. In section 5.4.2, the implications of SIC construction on the SBS-UDS co-design are studied. Finally, in section 5.4.3, the trade-offs associated with multi-objective optimization are presented when different weights are assigned.

5.4.1 SBS-UDS Co-design Analysis ε = 1 (TAT only) vs ε = 0.5

Table 5-I shows the optimal solution obtained for SIC1 with P_{max} and TSV_{max} fixed to 10 and 30 respectively. The 'schedule' column shows the optimal combination of dies to be tested in parallel within a session, and the 'TAM width' column shows the optimal channel widths for dies 1 through 5. For example, for $C_{max} = 20$, the optimal schedule is composed of two sessions (demarcated by '-'), with dies 1 through 4 tested in parallel in session 1 with channels widths of 4,2,1 and 1 respectively, followed by die 5 in the next session with a single test channel. To allow the dies to be tested as per the schedule while ensuring that the required test channels can be formed between the tester and the die, the optimal chip terminal design type is shown in the column under the heading CT Type. The first row indicates the number of UDS chip terminals, whereas the second row shows the SBS chip terminals required at layers 1 through 5, respectively. For the example of $C_{max} = 20$, layer 1 is designed using 10 SBS pins forming 8 test channels (2 CTs reserved for Reference sharing), out of which 4 channels will be used to test die 1 and remaining channels for dies 2, 3 and 4. Layers 2 through 4 have been designed using 8,4,2 and 2 UDS chip terminal forming 4,2,1 and 1 channels, respectively. It is evident from the last three columns that using this Pinmap, the test requirements of all dies are fulfilled without violating any constraints on P_{max}, TSV_{max} and C_{max}.

Table 5-II compares the solutions obtained for the same problem in Table 5-I but using the ε = 0.5 in the objective function, representing the case when both objective have equal weights. Since TAT is orders of magnitude greater than CTs, for the ε = 0.5, the solver will only reduce CTs when there is no further decrease in TAT. The optimal TAT, schedule and TAM allocation remains the same using both objective functions (as in Table 5-I); however, the multi-objective formulation returns the solutions having different PMAP. which minimizes CT count. The solver achieves this in two ways. First, it only creates the minimum number of test channels required for the given TAT. This follows the discussion in [22][25][26] in which the authors indicated that TAT might not necessarily decrease further until a sufficient number of additional test channels (and hence the CTs) are added, therefore for a given upper limit on CTs, there will exist a region in which TAT does not decrease with additional test channels, also known as pareto-optimal region. In this region, several redundant design choices are possible using a different number of test channels giving the same TAT. If the objective is singularly to reduce TAT, the solver may return any solution as long as it is feasible. However, in this case, as CT reduction is added to the objective, it forces the solver to return the solution which uses the least number of CTs whenever it encounters pareto-optimal front in terms of TAT. Secondly, using the SBS-UDS co-design methodology, the solver will have the choice of several SBS/UDS combinations to create the same number of test channels. Further CT reduction may be possible by increasing the SBS CTs as long as the cost budget is not exceeded.

The pareto-optimal region and the SBS/UDS design choices are better illustrated in Fig. 5-1, which compares the TAT for SIC1 with the number of test channels used and the composition of test channels (UDS or SBS). The solid line indicates the sum of UDS and SBS test channels (at all layers including Pins and TSVs) used for the corresponding test time shown in dashed line. Figure 5-1 shows that the test time reduction is a monotonically non-increasing function of the number of test channels. For the given P_{max} and TSV_{max}, the TAT is in the pareto-optimal

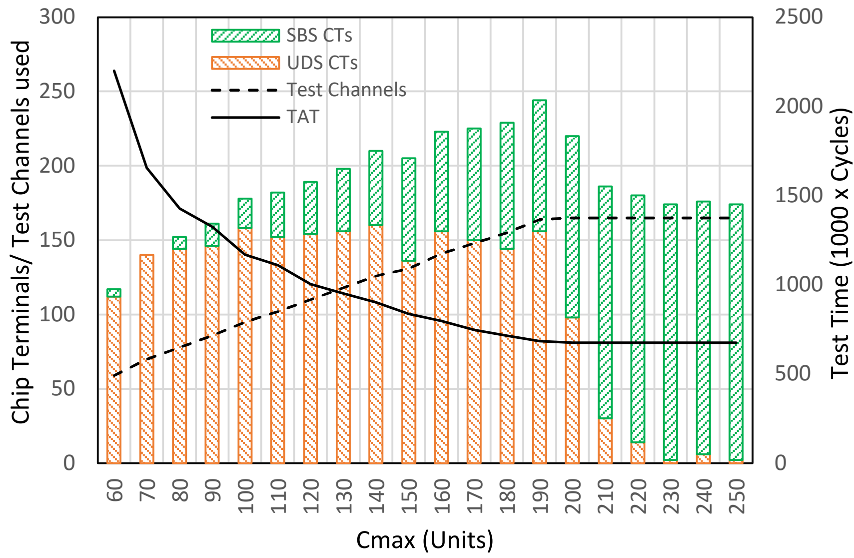
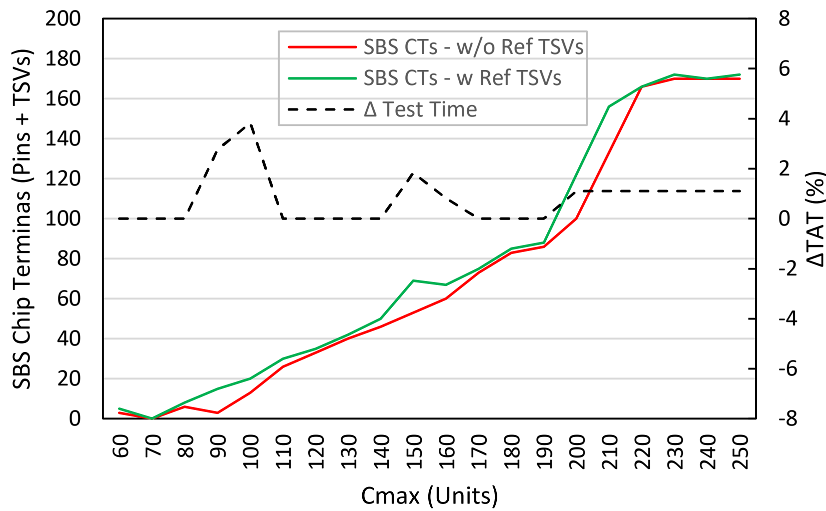


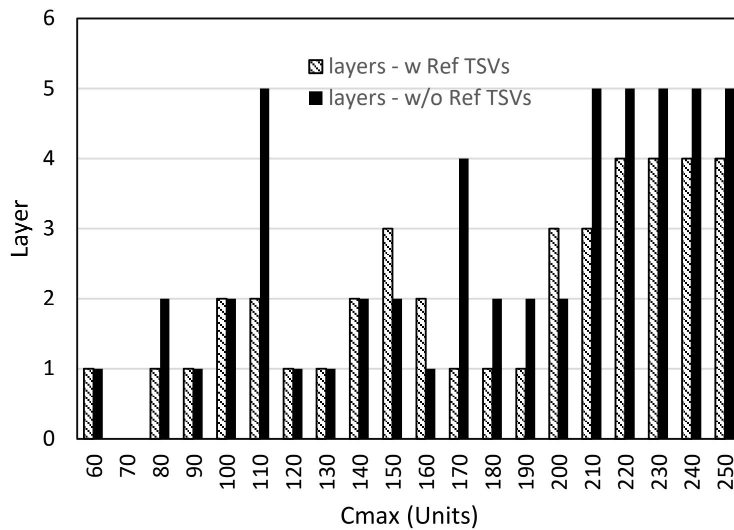
Figure 5-1: TAT, Test Channels used and Chip Terminal design choice for SIC1 (P_{max} =90, TSV_{max}=190)

region when $C_{max} > 190$. Beyond this point, the solver does not utilize more than 244 CTs (out of 280), although additional test channels are possible but will not result in TAT reduction. It may be noted that discussion so far has been in terms of CTs; as CTs include both pins and TSVs, either of them may cause pareto-optimal region depending on the design of the SIC, which is discussed later in section 5.4.2. In the region below 190, the solver increases the number of test channels using different SBS and UDS combinations. As the constraint on maximum power is relaxed, the solver tends to favour SBS over UDS. When the power budget is low, the optimal TAM solution prefers UDS which is more power efficient. If the power constraints are further relaxed, the solver increases the number of test channels by finding the optimal UDS and SBS trade-off that minimizes CT consumption.

The proportion of the SBS test channels increases as the power constraints are relaxed; however, in some cases the solver prefers UDS even though additional SBS test channels may be created without violating the power constraints. For instance, in Fig. 5-1 for C_{max} 150, 131 test channels are formed using 136 UDS and 69 SBS chip terminals. As the power constraints are further eased to 160, the solver increases the test channels to 141, but it decreases SBS terminals to



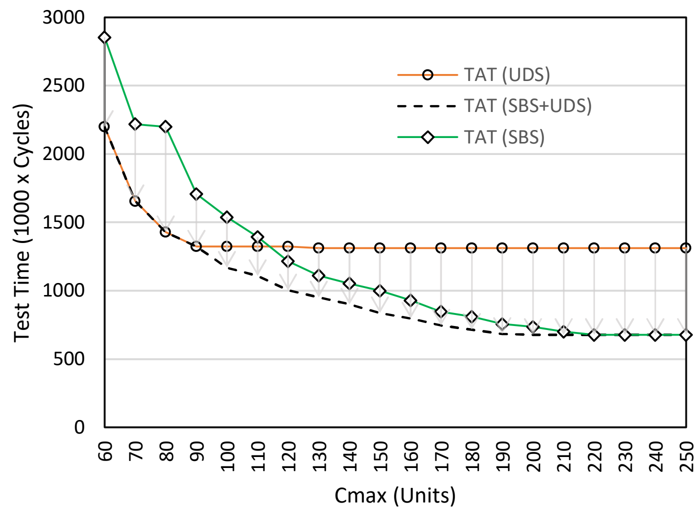
(a)



(b)

Figure 5-2: Reference CT impact on TAM Design (SIC1, P_{max} =90, TSV_{max}=190) (a) Chip Terminals designed as SBS with/ without reference CTs (red and green curves) and the resultant difference in TAT (dashed curve) (b) Effect of reference CTs on design choice

67 and increases UDS terminals to 156. This is because the solver also tries to avoid designing SBS as it requires additional dedicated CTs for reference sharing. Fig. 5-2 further elaborates the impact of reference CTs inclusion on the SBS-UDS co-design methodology. The solid curves in Fig. 5-2(a) show the total number of CTs designed as SBS when reference CT overheads are included in the problem (green curve) and when these are ignored/ excluded (red curve). As only two chip terminals are required at every layer to share reference voltage, there is no significant effect on the design preference. The red and green curves mostly stay close suggesting that in both cases the solver utilizes a similar



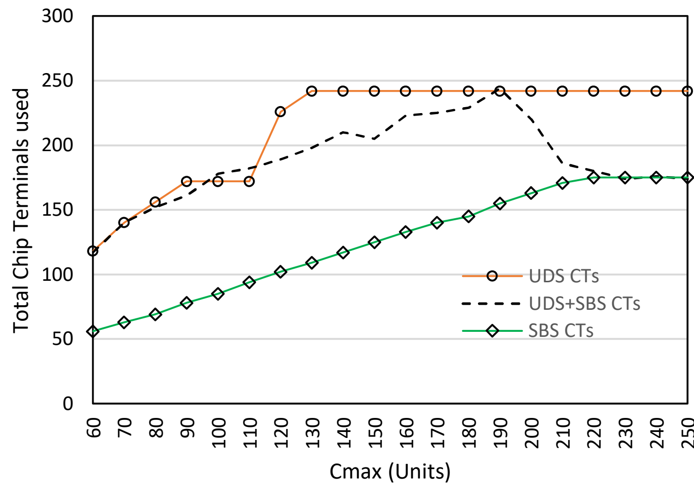


Figure 5-3: Comparison of SBS-UDS Co-design method against UDS only and SBS only TAMs (SIC1, P_{max} = 90, TSV_{max} = 190) (a) Test Time (TAT) (b) Chip Terminal (CT) utilization

proportion of CTs as SBS. However, as the green curve is mostly above the red, this indicates that the solver slightly reduces the preference for SBS if reference TSVs are to be included. To quantify it's effects on the TAT, the % difference for the TAT is shown in the dashed line with corresponding secondary vertical axis on the right. The effect on TAT is mostly negligible (0-1%), however in some cases it may be as high 3.8% (for C_{max} = 100). Figure 5-2(b) shows the layers in the stack at which SBS is included (with or without reference CT consideration). As noted earlier, inclusion of SBS at a given layer requires CTs at all layers below it. The proposed co-optimization minimizes the reference CT overheads by designing SBS starting from bottom up and avoids designing SBS on higher dies.

The discussion so far has focused on the analysis of co-design methodology. We now compare the co-design performance against the TAMs designed purely using SBS or UDS. It may be noted that the co-optimization formulation presented in section 5.3 can be easily reduced to SBS only and UDS only problem by including a constraint $\sum_{l=1}^{M} PMAP_{ls} = 0,$ (for s=1 or s=2 respectively). This constraint simply forces the solution not to include any UDS or SBS design choice at any of the layers. Figure 5-3(a) compares the optimal test time for SIC 1 when the TAM is designed entirely using UDS, SBS and using the proposed co-design method using both schemes. In the low-power budget region (<90), the UDS based TAM design offers lower test time than SBS. This is because, in this region, more test channels could be formed using UDS due to lower power consumption, and also because the reference TSV overheads of SBS are significant. On the contrary, in the high-power budget region (>220), The SBS based TAM offers significant test time reduction compared to UDS, since more channels can now be formed within the power allowance, and also because the reference TSV overheads become insignificant compared to the total test channels formed. It is evident that the SBS-UDS co-design model prefers the full UDS and SBS based schemes in these regions. However, in the intermediate region, the co-design offers more test time reduction as compared to either UDS or SBS alone. The solver identifies the layers which cause the test access bottleneck and designs SBS channels at those layers only, while designing UDS channels where possible with the objective of minimizing the overall test time.

Fig. 5-3(b) compares the chip terminals (Pins and TSVs) utilized by the TAM design methods out of the maximum allowance of 90 Pins and 190 TSVs. As expected, the UDS based TAM would require the highest number of chip terminals, whereas SBS significantly reduces the chip terminal utilization. However, keeping in view the test time reduction in Fig. 5-3(a) and CT utilization in Fig. 5-3(b), the co-design method clearly balances both, which also validates the correctness of the proposed co-design ILP formulation.

5.4.2 SIC Design and die position implications

Fig. 5-4 shows the effects of SIC construction on the TAM design using SBS-UDS co-design methodology for the 3 x handcrafted SICs reported in [23]. It may be noted that all the SICs in this experiment are composed of the identical dies but stacked in different order. In SIC1, the die complexity (and hence T_{mw}) decreases bottom up, for SIC2 the opposite is true whereas SIC3 has the most complex die in the middle. From Fig. 5-4(a) it is apparent that the TAT for SIC 1 reduces steeply with increasing C_{max} (which in turn increases the number of CTs that can be used) and enters the pareto optimal region much earlier as compared to SIC 2 and 3. This behaviour is in line with observation made in [23], in which the authors highlighted that for SICs composed of same dies, lower test times can be achieved if the complex dies are placed at the bottom (such as SIC 1). The reason is that dies higher up in the stack require test channels to be formed at all the intermediate layers, and hence require more resources. For SIC 1 the complex die dominating the test time is at the bottom of the stack, therefore the test time decreases with every pin while requiring only a few TSVs. On the contrary, SIC 2 and 3 require larger increment in TSVs for every pin increment to be able to reduce TAT.

From Fig. 5-4(b), It is apparent that in order to allow TAT to decrease, the Pins and TSVs need to increase with a certain ratio between them, which we denote by ‘P/T’. For SIC 1, P/T is around 1, meaning that test time should decrease for 1:1 increase in pin and TSV channels. For SIC 1, at around $C_{max} = 240$, there is enough power to create the maximum channels possible (100+100), and there is no further decrease in test time despite additional TSVs being available. On the other hand, SIC 2 requires a P/T ratio of 0.3 (i.e., for every 1 pin, there should be 3 TSVs to make a difference in TAT), and for SIC3, the P/T is observed to be 0.5. Consequently, for SIC2 and 3 the test time keeps decreasing beyond $C_{max} = 240$ (Unlike SIC1). It is worth noting that that the test time reduction is directly proportional to the P/T; however, for SICs requiring lower P/T, any TSV count

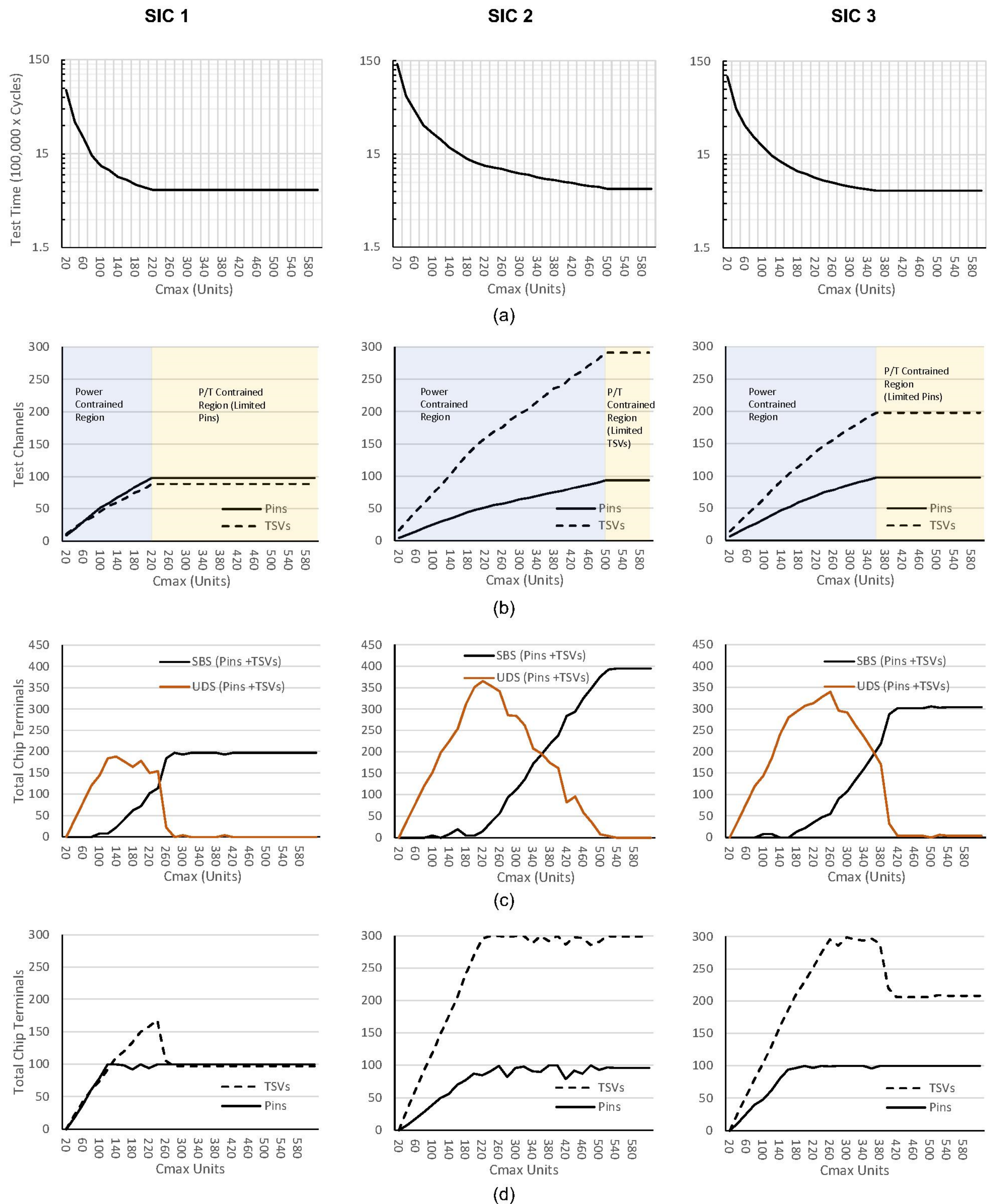


Figure 5-4: Effects of SIC construction on the TAM design using SBS-UDS codesign methodology for SIC 1 to 3 with easing cost constraints (a) TAT reduction (b) Test Channels formed (c) CTs designed as SBS or UDS (d) Pins and TSVs used

above $P_{max}/(P/T)$ does not yield any further TAT reduction. Fig. 5-4(b) can therefore be demarcated into two regions:

a. Power Constrained region: Additional test channels can be formed for the required P/T if the power constraints are relaxed.

b. P/T Constrained: There is enough power budget, but additional test channels cannot be formed for the required P/T either due to insufficient Pins or TSVs.

Fig. 5-4(c) shows the channel type preference to form the test channels in the Power Constrained Region in Fig. 5-4(b). The response for all the SICs is similar in these regions; the solver initially prefers UDS and gradually adds SBS test channels where TAT reduction is possible. Fig. 5-4(d) shows the TSVs and Pins used (UDS and SBS) to form the test channels in Fig. 5-4(b) for a TAT in Fig. 5-4(a). For SIC 1 and 3, when the P/T constrained region is reached, as further TAT reduction is not possible, the solver now focuses on reducing the Pins or TSVs used as much as the power budget allows. For SIC 2, however, no such decrease is observed. This is because coincidently, P/T for SIC 2 is 0.3 and ratio between P_{max} and TSV_{max} is also 1:3 (P_{max} 100 TSV_{max} 300). Therefore, all the Pins and TSVs could be utilized to reduce TAT and hence pin minimization is not possible. It may be pointed out that the pareto-optimal TAT for SIC1 and SIC3 is the same (618176 cycles), whereas, for SIC 2, it is higher (635845 cycles) which indicates that SIC 1 and 3 are pin-constrained whereas SIC 2 is TSV-constrained. A slight increase in the TSV_{max} would in fact, bring the pareto-optimal TAT for SIC2 to 618176 cycles as well.

5.4.3 Effects of ε variations on the optimal solution

The approximation of lower and upper bound on ε was discussed in Appendix 1. For SIC1 (for $M = 5$ and assuming $w_{max} = 65$), the calculated bounds are shown in Table 5-III (columns 2 and 3). From Table 5-III, it is evident that CT_{lb} of 10 and CT^{ub} of 1950 are rather extreme values. Therefore, the approximation in Appendix 1 represent an ideal scenario. The minimum and maximum number of CTs would be known in practice, and the bounds can be further tightened. For instance, if Pmax is taken to be between 10 and 100, and TSV_{max} between 30 and 300, the corresponding TAT can be computed by solving the problem for

Table 5-III: ε bounds for SIC 1 (M = 5)

	Ideal case		10 ≤ Pins≤ 100 and 30≤TSVs≤300	
	lb	*ub*	*lb*	*ub*
TAT	$\max_{m}(T_{m,w_{max}})$ 544579	$\sum_{m=1}^{M}(T_{m1})$ 54735722	(solve for $P_{max} = 100, TSV_{max} = 300$ and ε =1) 618176	(solve for $P_{max} = 10, TSV_{max} = 30$ and ε =1) 7240424
CT	$2M$ 10	$w_{max}.\ M\ (1+M)$ 1950	$P_{min} + TSV_{min}$ 40	$P_{max} + TSV_{max}$ 400
ε (x10^-5)	$CT_{lb}\ /\ TAT^{ub}$ 0.0183	$CT^{ub}\ /\ TAT_{lb}$ 358.0748	$CT_{lb}\ /\ TAT^{ub}$ 0.5525	$CT^{ub}\ /\ TAT_{lb}$ 64.7065

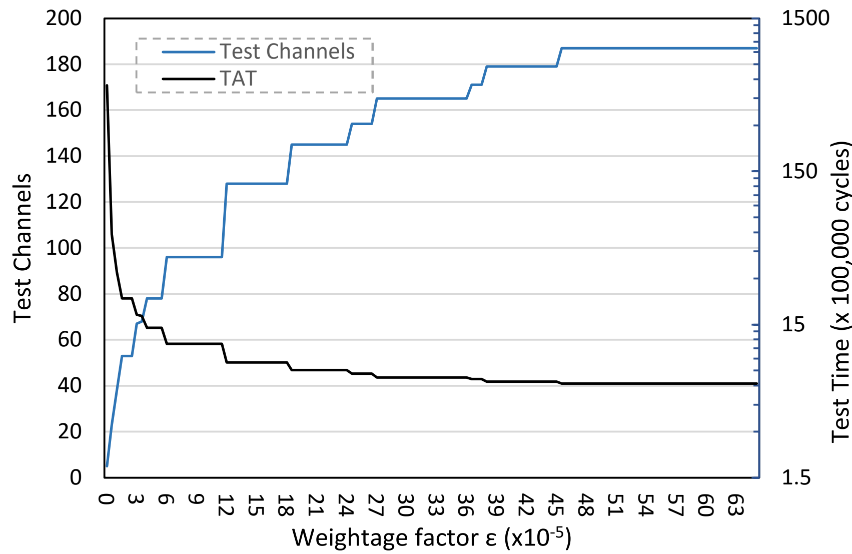


Figure 5-5: Multi Objective TAT and CT Minimization - Various pareto optimal solutions for different ε ($P_{max} = 100, TSV_{max} = 300$)

these Pin and TSV limits using ε=1 (TAT reduction only). Columns 4 and 5 in Table 5-III show the revised bounds on ε for the known Pmax and TSV_{max} case, which are significantly tighter as compared to the ideal case. The tighter bounds on ε allow the designer to explore only the regions in which multi-objective optimization is feasible.

Fig. 5-5 compares the trade-off between reduction in TAT as compared to the increase in test channels when the weightage ε is varied. For the case of Fig. 5-5, the power limit has been set to a high value (300), and the ε was varied from 0 to 65 (x10^{-5}) (from Table 5-III), the two extremities represent the case when the

objective reduces to pin minimization (For ε equals to zero) or TAT reduction only ($\varepsilon > 64.7 \times 10^{-5}$). In the former case, every die has been allocated a single channel (using 2 UDS pins and 8 UDS TSVs i.e. 2 for every layer), resulting in a very high TAT (54.7 million cycles), whereas in the latter case, the test time and the channel formation is 618176 cycles, which is the same as the pareto-optimal region in Fig. 5-4(a) for SIC1. In between the two extremities, several options exist, the optimal balance being at the cross-over point ($\varepsilon = 4 \times 10^{-5}$) at which the TAT is 1.4 million cycles (97.4% lesser to the worst-case TAT at ε =0) and test channels utilization is 78 (55.8% lesser compared to the worst-case channel utilization at $\varepsilon = 65 \times 10^{-5}$). However, it may be noted that the TAT is plotted on the logarithmic scale (base 10), whereas the CTs using a linear scale. Clearly, for increasing CT count, there is an exponential decrease in TAT. Nevertheless, as both objectives are essential but contradicting, there is no perfect solution and the choice of ε remains a designer preference.

5.5 Conclusion

This paper proposes an SBS and UDS co-design strategy for test time and chip terminal minimization of 3D Stacked Integrated Circuits. The problem was formulated as an Integer Linear Programming based optimization. The performance of the proposed method in terms of TAT reduction and CT utilization was demonstrated. Experiments conducted using 3D SICs based ITC'02 benchmark SoCs suggest that the proposed methodology out-performs the TAMs based entirely on SBS by providing optimal balance between TAT and CT count. The proposed method prefers UDS when the power budget is low and SBS when sufficient a power budget is available. In the intermediate region, the proposed method suggests a combination of both. Important factors governing the design choice of SBS and UDS were identified. The tendency of the optimal TAM design to include SBS depends on the power consumption of the SBS channels, the overall power budget and the construction (regarding test complexity) of the 3D Stack. A multi-objective optimization focusing on TAT and CT reduction was presented. An important factor when using multi-objective optimization is the choice of weightage factor ε. Methods to approximate the bounds on ε were

presented, and the design trade-offs between TAT reduction and required test channels were discussed.

5.6 Appendix 5-I: Calculation of Bounds on ε for multi-Objective Optimization

In section 5.3.1, the objective function comprising of both the TAT and CT was formulated as follows:

$$minimize \quad (1-\varepsilon).CT + \varepsilon.TAT.$$

It is apparent that the problem could be easily switched between TAT reduction only, by setting $\varepsilon = 1$, or as a CT minimization problem by using $\varepsilon = 0$. For ε values between 0 and 1, the solver will return a solution minimizing both CT and TAT depending upon the value of ε. In general, TAT is several orders of magnitude higher than the number of chip terminals, therefore for appropriate values of ε, the upper and lower bounds on both TAT and CT must be approximated.

The minimum number of CTs would be in the case when all dies are tested sequentially using a single test channel, therefore for a UDS based design, the lower bound on the number of chip terminals $CT_{lb} = 2M$, where M is the number of dies in the stack. The test time associated with CT_{lb} will be the highest, representing the upper bound on TAT, i.e. $TAT^{ub} = \sum_{m=1}^{M}(T_{m1})$.

In order to approximate the upper bound on CT and lower bound on TAT (i.e. CT^{ub} and TAT_{lb}), the bounds on the individual dies are required to be approximated first. The authors in [25] had demonstrated that for a given core in a die, there is no further improvement in test time for a TAM width greater than k_{max}. This upper bound on the cores can then be used to approximate the upper bound on the TAM width (w) for a die (m), which we denote by $w_{max,m}$. While $w_{max,m}$ will be different for each die, for simplicity, we assume a single value w_{max} such that $w_{max} = \max_{m} (w_{max,m})$. The highest number of CTs will be required if all dies are tested in parallel using w_{max} TAM width and all TAMs are designed using UDS ($2.w_{max}$ CTs required for each die). As previously noted in section 5.3, in

order to communicate with the tester, a die would require w_{max} channels (2x w_{max} CTs) at all layers below it, therefore, the total chip terminals required by M dies is given by:

$$CT^{ub} = 2.w_{max} \sum_{m=1}^{M} m$$

It is clear that the summation component of the above equation is a sum of an arithmetic series from 1 to M and with a common difference of 1, therefore:

$$CT^{ub} = 2.w_{max}. \frac{M(M+1)}{2}$$

OR

$$CT^{ub} = w_{max}. M (M+1)$$

Similarly, the lower bound on TAT, when all dies are tested in parallel using wmax TAM width for each die, is given by

$$TAT_{lb} = \max_{m} (T_{m,w_{max}})$$

The bounds on ε can now be approximated as:

$$\frac{CT_{lb}}{TAT^{ub}} \leq \varepsilon \leq \frac{CT^{ub}}{TAT_{lb}}$$

5.7 References

[1] J. P. Gambino, S. A. Adderly, and J. U. Knickerbocker, “An overview of through-silicon-via technology and manufacturing challenges,” *Microelectron. Eng.*, vol. 135, pp. 73–106, 2015.

[2] S. S. Muthyala and N. A. Touba, “Reducing test time for 3D-ICs by improved utilization of test elevators,” *IEEE/IFIP Int. Conf. VLSI Syst. VLSI-SoC*, vol. 2015-Janua, no. January, 2015.

[3] P. Georgiou, F. Vartziotis, X. Kavousianos, and K. Chakrabarty, “Testing 3D-SoCs Using 2-D Time-Division Multiplexing,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 37, no. 12, pp. 3177–3185, Dec. 2018.

[4] Li Jiang, Qiang Xu, K. Chakrabarty, and T. M. Mak, “Integrated Test-Architecture Optimization and Thermal-Aware Test Scheduling for 3-D

SoCs Under Pre-Bond Test-Pin-Count Constraint," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 20, no. 9, pp. 1621–1633, Sep. 2012.

[5] C. Y. Ooi, J. P. Sua, and S. C. Lee, "Power-aware system-on-chip test scheduling using enhanced rectangle packing algorithm," *Comput. Electr. Eng.*, vol. 38, no. 6, pp. 1444–1455, 2012.

[6] S. K. Roy and C. Giri, "Test Architecture Optimization for Post-bond Test and Pre-bond Tests of 3D SoCs Using TAM Reuse," *IETE J. Res.*, 2021.

[7] T. E. Yu, T. Yoneda, K. Chakrabarty, and H. Fujiwara, "Thermal-Safe Test Access Mechanism and Wrapper Co-optimization for System-on-Chip," in *16th Asian Test Symposium (ATS 2007)*, 2007, pp. 187–192.

[8] D. Xiang, K. Chakrabarty, and H. Fujiwara, "Multicast-Based Testing and Thermal-Aware Test Scheduling for 3D ICs with a Stacked Network-on-Chip," *IEEE Trans. Comput.*, vol. 65, no. 9, pp. 2767–2779, 2016.

[9] N. Aghaee, Z. Peng, and P. Eles, "A Test-Ordering Based Temperature-Cycling Acceleration Technique for 3D Stacked ICs," *J. Electron. Test. Theory Appl.*, vol. 31, no. 5–6, pp. 503–523, 2015.

[10] M. Richter and K. Chakrabarty, "Optimization of test pin-count, test scheduling, and test access for NoC-based multicore SoCs," *IEEE Trans. Comput.*, vol. 63, no. 3, pp. 691–702, 2014.

[11] V. Iyengar, K. Chakrabarty, and E. J. Marinissen, "Test Access Mechanism Optimization Test Scheduling, and Tester Data Volume Reduction for System-on-Chip," *IEEE Trans. Comput.*, vol. 52, no. 12, pp. 1619–1632, 2003.

[12] Y. W. Lee and N. A. Touba, "Unified 3D test architecture for variable test data bandwidth across pre-bond, partial stack, and post-bond test," *Proc. - IEEE Int. Symp. Defect Fault Toler. VLSI Syst.*, pp. 184–189, 2013.

[13] A. Zhu, C. Xu, J. Wu, Z. Li, and J. Niu, "Three-dimension test wrapper design based on multi-objective cuckoo search," *Open Cybern. Syst. J.*, vol. 8, no. 1, pp. 104–110, 2014.

[14] T. Kaibartta, C. Giri, H. Rahaman, and D. K. Das, "Approach of genetic algorithm for power-aware testing of 3D IC," *IET Comput. Digit. Tech.*, vol. 13, no. 5, 2019.

[15] W. Marrouche, R. Farah, and H. M. Harmanani, "A Strength Pareto Evolutionary Algorithm for Optimizing System-On-Chip Test Schedules," *Int. J. Comput. Intell. Appl.*, vol. 17, no. 02, p. 1850010, 2018.

[16] F. H. Tang, H. Y. Kao, S. H. Huang, and J. F. Li, "3D Test Wrapper Chain Optimization with I/O Cells Binding Considered," *IEEE 2019 Int. 3D Syst. Integr. Conf. 3DIC 2019*, pp. 3–6, 2019.

[17] Y. Y. Wu and S. H. Huang, "TSV-aware 3D test wrapper chain optimization," *2018 Int. Symp. VLSI Des. Autom. Test, VLSI-DAT 2018*, pp.

1–4, 2018.

[18] M. Agrawal, K. Chakrabarty, and R. Widialaksono, “Reuse-based optimization for prebond and post-bond testing of 3-D-stacked ICs,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 34, no. 1, pp. 122–135, 2015.

[19] X. Wu, Y. Chen, K. Chakrabarty, and Yuan Xie, “Test-access mechanism optimization for core-based three-dimensional SOCs,” *Microelectronics J.*, vol. 41, no. 10, pp. 601–615, 2010.

[20] M. Agrawal and K. Chakrabarty, “Test-cost modeling and optimal test-flow selection of 3-D-stacked ICs,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 34, no. 9, pp. 1523–1536, 2015.

[21] B. SenGupta, D. Nikolov, A. Dash, and E. Larsson, “Test Flow Selection for Stacked Integrated Circuits,” *J. Electron. Test. Theory Appl.*, vol. 35, no. 4, pp. 425–440, 2019.

[22] I. A. Soomro, M. Samie, and I. K. Jennions, “Test Time Reduction of 3-D Stacked ICs Using Ternary Coded Simultaneous Bidirectional Signaling in Parallel Test Ports,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 39, no. 12, pp. 5225–5237, 2020.

[23] B. Noia, K. Chakrabarty, S. K. Goel, E. J. Marinissen, and J. Verbree, “Test-Architecture Optimization and Test Scheduling for TSV-Based 3-D Stacked ICs,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 30, no. 11, pp. 1705–1718, Nov. 2011.

[24] R. T. Marler and J. S. Arora, “Survey of multi-objective optimization methods for engineering,” *Struct. Multidiscip. Optim.*, vol. 26, no. 6, pp. 369–395, 2004.

[25] V. Iyengar, K. Chakrabarty, and E. J. Marinissen, “Test Wrapper and Test Access Mechanism Co-Optimization for System-on-Chip,” *J. Electron. Test. Theory Appl.*, vol. 18, pp. 213–230, 2002.

[26] Brandon Noia · Krishnendu Chakrabarty, *Design-for-Test and Test Optimization Techniques for TSV-based 3D Stacked ICs*. Springer International Publishing Switzerland, 2014.

6 CONCLUSION AND FUTURE DIRECTIONS

6.1 Addressing the Aim and Objectives of the Research

The aim of the research was 'Design and optimization of a resource-efficient Test Access Mechanism (TAM) in 3D digital electronics to reduce overall test-access time.' In this research, a novel Test Access Mechanism based on Simultaneous Bi-directional Signaling has been investigated for use in 3D Stacked Integrated Circuits. This work demonstrates how SBS can be used in conjunction with existing test methodologies covering various abstractions levels of the TAM design methodology, ranging from finer details governing circuit design and logic integration to the overall impact (in terms of test time) at the application level. Experiments suggest that SBS offers a significant advantage over conventional signalling, reducing test time by as much as 53.6%. Alternatively, the benefits could be manifested in terms of Chip Terminal reduction, or a balance between both could be achieved (as per design requirements) using the proposed co-design methodology. The individual objectives are discussed below.

Literature Review: Undertake a literature review to identify research gaps and propose a TAM design strategy to address the gaps identified.

The factors affecting the TAM design methodology and their effects in terms of test time were evaluated in light of the past research as discussed in Chapters 2 and 3. In terms of TAM design, the need for chip terminal efficiency was identified, and Simultaneous Bi-directional Signaling was proposed being highly suitable for use in scan-based TAMs. Consequently, the need for optimization methods for SBS based TAM was identified as discussed in Chapters 4 and 5.

Design: Study SBS design solutions for use in basic and advanced Test Access Mechanisms - Design SBS transceivers suitable for use in 3D IC test architectures.

TAM design considerations of integrating SBS for use in Parallel Test Ports and advanced test methods based on Reduced Pin Count Testing (RPCT) were presented in Chapters 2 and 3, respectively, covering both logic and circuit level design considerations:

- **Logic level**: TAM design considerations for the incorporation of SBS in 3D SICs such that it does not interfere with the functional mode performance and standard DFT logic such as JTAG compliant boundary scan registers were presented.
- **Circuit level**: The transceiver designs suitable for use in either test method (PTAM and RPCT) have been proposed.

Optimize: Formulate TAM optimization methodology for 3D Stacked Integrated circuits using SBS only and SBS+UDS co-design.

TAM optimization methodology for SBS only and SBS+UDS co-design was developed and has been presented in Chapters 4 and 5.

Evaluation: Create TAM design and optimization evaluation frameworks – Validate proposed design; Analyze and quantify performance compared to conventional design methods.

Separate evaluation methods were required for the design and optimization aspects.

- **Design:** The validation and evaluation methods have been discussed in Chapter 2 for PTAM and Chapter 3 for TDM based TAM. The transceiver operation for PTAM was validated using a simple case of a 2-bit scan chain, designed using commercial 180nm technology and comparing the scan-in signal with the scan-out signal. The capture cycle was ignored, and continuous shifting was performed. For TDM based TAMs, a use-case stack of 3 dies was implemented in Cadence Virtuoso based on existing TDM based methods using academic 45nm Nangate technology. In the TDM case, the output is an interleaved form of the scan-in signal. Therefore, a functional model of the stack was implemented in MATLAB to generate the expected scan-out signal followed by a comparison with the scan-out from the transient simulations. In both cases, the signal integrity was verified across all process corners and under cross-coupling from 8 adjacent TSVs assuming a 3x3 cluster. Power comparison was made with UDS based designs implemented in the same technology as 4x

buffers. For TDM based design, the transceiver performance with varying test clock frequencies were discussed and compared with the prior works in TDM and SBS.

- **Optimization:** Experiments were performed on 3D Stacks based on ITC02 benchmarks SoCs, using session-based scheduling for post-bond test scenarios assuming soft-cores. The model was initially implemented in MATLAB; however, it was migrated to Optimization Programming Language and solved using a commercial optimization tool (IBM CPLEX) due to high computation times. Comparisons were made by designing TAMs based on UDS and SBS under similar TSV and Pin constraints. For the multi-objective co-design problem, three sets of experiments were conducted focusing on: 1) The effect of SBS related design overheads on the design choice of the TAM in terms of UDS or SBS. Comparisons were also made for the co-design method with SBS only and UDS only TAMs. 2) The implications of SIC construction on the SBS-UDS co-design methodology. 3) Effects of variation in objective weightage on the TAM design and schedule.

6.2 Contributions to the existing body of knowledge

The original contributions of this thesis are:

- A novel method of improving chip terminal efficiency for application in test mode using simultaneous bi-directional signalling.
- SBS Transceiver circuit design suitable for Parallel Test Port based TAMs and quantitative analysis of the of the cost in terms of power and area, and benefits in terms of test time reduction.
- An Integer Linear Programming based optimization formulation for 3D SICs, capable of designing a TAM using conventional UDS design, SBS or both.
- Extension of the ILP model to multi-objective optimization scenarios for Test Application Time (TAT) and Chip Terminal (CT) reduction.

- SBS Transceiver circuit for use in Reduced Pin Count Test (RPCT) methods and the methodology to integrate SBS with existing Time Division Multiplexing (TDM) based approaches.

6.3 Limitations and Future Work

6.3.1 Transceiver Design

An important consideration in the SBS transceiver design was observed to be the static power consumption of the transmitter, which limits the maximum achievable performance. Therefore, the goal in this research had been to keep the power consumption as close to UDS as possible. In the Parallel TAM based design, the static power consumption was easily controlled owing to the following factors:

1. Low-frequency requirement.
2. Lesser channel impedance due to point-point communication involving a single TSV.

Because of the above favourable factors, diode-connected MOSFETs could be inserted in the transmitter resulting in significantly reduced static power consumption. However, in the case of TDM based RPCT based methods, the above factors are no more favourable. Therefore, the following approach was adopted:

1. To allow high frequencies, the diode-connected MOSFETs in the transmitter were removed.
2. Transistor sizes were chosen carefully to allow a balance between the required frequency but also limit power consumption. To allow reconfigurability, as an extension, a binary-weighted transmitter was proposed providing adjustability after device fabrication.
3. A transmitter control circuitry using Sense Amplifier Completion Detector (SACD) was incorporated in the design to remove static power after the receiver has completed the sensing operation.

Nevertheless, future research may focus on power-efficient SBS transceiver designs suitable for single-ended implementations in test mode. With increased power efficiency, the tighter design margins may be eased in favour of increased

safety margins and robustness. The potential areas of further investigation include:

1. Investigation of capacitor coupled transmitter designs [1]. The capacitors block the static power, which significantly improves the overall power consumption.
2. SACD is already shown to be capable of removing static power after completion of the sensing operation. A similar detection and control circuit may be investigated that senses the completion of transmitter operation (i.e. when the output transition has reached a steady-state) [2]. Focus may be restricted only on the rail-mid transitions as the other transition do not involve static power consumption, as highlighted in Chapter 3. Such an arrangement will improve power consumption and alleviate the requirement of having to carefully select transmitter sizes, thereby reducing the overall design effort and increased robustness.
3. Utilization of Carbon-Nanotube Field Effect Transistors (CNFETs) for ternary logic generation [3] and CMOS-memristors for the Ternary Decoder designs [4].

6.3.2 TSV Testing

At the pre-bond stage, the TSV has an internal end that is accessible using on-chip test architectures and an external end connected after the stacking and remain open at the pre-bond stage. The open end of the TSVs cannot be directly probed because of their small size (*as small as 5μm),* high density and fragility of the thinned wafer [5]. However, defects may be introduced in TSV manufacturing due to various reasons, necessitating pre-bond TSV tests such that the defective dies do not propagate to the stacking stage. Therefore, pre-bond testing of TSVs is an active area of research (as previously highlighted in Chapter 1).

The equivalent model of the TSV comprises capacitance and resistance (inductance is negligible and ignored), as previously discussed in chapter 2. Any faults in the TSV manifest as a change in the electrical properties, hence the variation in TSV impedance. The previous work in TSV testing [6][7][8] have

proposed methods to quantify these changes to characterise a TSV as defective or fault-free. The process generally involves applying a test stimulus and observing the response using on-chip hardware. The response of the faulty and correct TSV differs due to different impedance and exhibits as a change in the rise/fall time constant, which is then sensed/ measured using voltage comparators. Using these methods, the TSV faults may be detected. It is interesting to note that SBS based test methodology inherently provides an ability to apply stimuli (using the transmitter and the TG) at the internal end of the TSV and provides a sense amplifier to monitor the response compared to reference levels. Therefore, methods employing re-usability of SBS hardware may be explored for use in open-ended TSV testing, so as to avoid the overheads involved in designing of separate TSV test hardware.

6.3.3 Optimization

The ILP based optimization used in this thesis assumes that the additional test channels created using SBS can all be used to create a Test Access Mechanism which could assign any (additional) number of cores in parallel. However, the TAM design is restricted by the allowable power budget during testing and therefore, only a limited number of cores can be tested simultaneously. Therefore, under certain circumstances, the additional test channels may not be used to test additional cores in parallel, but only the existing TAM width of the cores can be increased. As evident from the results in chapter 4, the test time of core does not necessarily decrease with an increasing number of test channels (in the Pareto-optimal regions). Under these situations, if the power budget does not allow the addition of another core in the session, the additional channel width may not be useful, and the difference between the TAT using SBS and UDS may not be significant.

Secondly, the optimization formulation has been restricted to the post-bond (complete stack) test scenarios in this thesis. Although the SBS based proposed TAM can be re-used in all the test instances, it has been previously reported by the authors in [9] that the TAM design optimal for one test instance may not necessarily be optimal for other instances.

As future work, the optimization formulation may be extended to consider the above-mentioned limitations. However, an important consideration to be made is the significant increase in the problem complexity and consequently the computational time. In such cases, the role of meta-heuristic optimization methods may be investigated [10]–[12].

6.4 References

[1] M. T. L. Aung, E. Lim, T. Yoshikawa, and T. T.-H. Kim, "Design of Simultaneous Bi-Directional Transceivers Utilizing Capacitive Coupling for 3DICs in Face-to-Face Configuration," *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 2, no. 2, pp. 257–265, Jun. 2012.

[2] R. Kumar, H. Deshpande, G. Choi, A. Sprintson, and P. Gratz, "Bidirectional interconnect design for low latency high bandwidth NoC," *ICICDT 2013 - Int. Conf. IC Des. Technol. Proc.*, pp. 215–218, 2013.

[3] C. Vudadha, A. Surya, S. Agrawal, and M. B. Srinivas, "Synthesis of Ternary Logic Circuits Using 2:1 Multiplexers," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 65, no. 12, pp. 4313–4325, 2018.

[4] A. Tankimanova and A. P. James, "Neural Network-Based Analog-to-Digital Converters," in *Memristor and Memristive Neural Networks*, InTech, 2018.

[5] E. J. Marinissen, "Challenges and emerging solutions in testing TSV-based 2 1 over 2D- and 3D-stacked ICs," in *2012 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2012, pp. 1277–1282.

[6] P. Y. Chen, C. W. Wu, and D. M. Kwai, "On-chip testing of blind and open-sleeve TSVs for 3D IC before bonding," *Proc. IEEE VLSI Test Symp.*, pp. 263–268, 2010.

[7] Y. Lou, Z. Yan, F. Zhang, and P. D. Franzon, "Comparing through-silicon-via (TSV) void/pinhole defect self-test methods," *J. Electron. Test. Theory Appl.*, vol. 28, no. 1, pp. 27–38, 2012.

[8] Y. W. Lee, H. Lim, and S. Kang, "Grouping-Based TSV Test Architecture for Resistive Open and Bridge Defects in 3-D-ICs," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 36, no. 10, pp. 1759–1763, 2017.

[9] B. Noia, K. Chakrabarty, and E. J. Marinissen, "Optimization methods for post-bond testing of 3D stacked ICs," *J. Electron. Test. Theory Appl.*, vol. 28, no. 1, pp. 103–120, 2012.

[10] X. Yang, *Nature-Inspired Metaheuristic Algorithms*. Luniver, 2010.

[11] K. Hussain, M. N. Mohd Salleh, S. Cheng, and Y. Shi, "Metaheuristic research: a comprehensive survey," *Artif. Intell. Rev.*, vol. 52, no. 4, pp.

2191–2233, 2019.

[12] K. A. Smith, “Neural networks for combinatorial optimization: A review of more than a decade of research,” *INFORMS J. Comput.*, vol. 11, no. 1, pp. 15–34, 1999.

APPENDICES

Appendix A : Optimization Programming Language (OPL) based CPLEX codes

SBS or UDS based TAM Design (Chapter 4)

```
/*********************************************
 * OPL 20.1.0.0 Model
 * Author: s286972
*********************************************/

  //Data
 int M = 5;
 int w_max = 65;
 int Pmax = ...;
 int TSVmax = ...;


  //Index
range l = 1..M;
range m = 1..M;
int s = 2; //*********************************** UDS or SBS selection 1/2
range t = 1..M;
range w = 1..w_max;  // {int} w = {1,2,3}; //I to w_max


 //data
int T[m][w]= ...; int T2[m][w]= ...; int T3[m][w]= ...;
int BIGM= w_max;  int BIGM2=maxl (Pmax,TSVmax);

 //Decision Variables

dvar int+ x[l][m];
dvar int+ o[m][t] in 0..1;
dvar int+ PMAPmax[l];
dvar int+ y[m][w] in 0..1;
dvar int+ PMAP[l][t];
dvar int+ TAT[t];
dvar int+ yo[m][w][t] in 0..1;
dvar int+ xo[l][m][t];
dvar int+ TAM[l][m];
dvar int+ c[l][t] in 0..1;
dvar int+ R[l] in 0..1;

// Obj Function
dexpr float cost= sum(tt in t) TAT[tt];


// Model
minimize cost;
subject to {

   forall (tt in t, mm in m)
     c9: TAT [tt] >= (sum (ww in w) (T[mm,ww]) * yo[mm,ww,tt])  ;

   forall (mm in m, ww in w, tt in t){
     L1:  y[mm,ww] + o[mm,tt] - yo[mm,ww,tt]  <= 1;
     L2:  y[mm,ww] + o[mm,tt] - 2* yo[mm,ww,tt]  >= 0;
     }

    forall (mm in m)
    c7: sum (ww in w) (y[mm][ww]*ww) - TAM[mm][mm] == 0;

    forall (mm in m, ll in l:ll<=mm)
    cty: TAM[ll][mm] == x[ll][mm];
```

```
  forall (mm in m: mm>=2, ll in l:2<=ll<=mm)
    c5: TAM[ll-1,mm] == TAM [ll,mm];

    forall (mm in m, ll in l:ll<=mm)
  c1: TAM[ll][mm] >= 1;

  forall (mm in m)
    c6: TAM [mm][mm] <= w_max;

     c12r: PMAPmax[1] + 2*(s-1)  <= Pmax;   //
   c13r: sum (ll in l: ll>=2) (PMAPmax[ll] +2*(s-1)) <= TSVmax;


  forall (mm in m)
  c10v2: sum (tt in t) o[mm][tt] == 1;

    forall (mm in m, ll in l:ll<=mm, tt in t){
    l31: xo[ll][mm][tt] - BIGM* o[mm][tt] <= 0;
    l32: xo[ll][mm][tt] - x[ll][mm] <= 0;
    l41:  xo[ll,mm,tt] - x[ll,mm] + (1 - o[mm,tt])*BIGM   >= 0;}



     forall (mm in m, ll in l:ll<=mm, tt in t)
       IE5: PMAP[ll][tt] == sum (mm in m) (xo[ll,mm,tt]*(2/s));


   forall (mm in m, ll in l:ll<=mm, tt in t){
     L6: PMAPmax[ll] >= PMAP[ll,tt];
         L62: PMAPmax[ll] <= PMAP[ll,tt] + (1 - c[ll,tt])*BIGM2;
         L63:  sum (ttt in t) c[ll,ttt] == 1;
        }

  /*********************************************
 * OPL 20.1.0.0 Data
 * Author: s286972
*********************************************/


SheetConnection sheet ("Chap4_SIC_T_SBS.xlsx");
T from SheetRead(sheet,"SIC1!$A$1:$BM$5");
T2 from SheetRead(sheet,"SIC2!$A$1:$BM$5");
T3 from SheetRead(sheet,"SIC3!$A$1:$BM$5");

Pmax= 100;
TSVmax=300;
```

SBS/UDS co-design (Chapter 5)

```
/*********************************************
 * OPL 20.1.0.0 Model
 * Author: s286972
 *********************************************/


//Data
 int M = 5;//...;
 int w_max = 65;
 int Pmax = ...;
 int TSVmax = ...;
 int Cmax = ...;
 float Cs[1..2] = ...;
 float alpha = ...;

  //Index
range l = 1..M;
range m = 1..M;
```

```
range s = 1..2;
range t = 1..M;
range w = 1..w_max;  // {int} w = {1,2,3}; //I to w_max


 //data
int T[m][w]= ...; int T2[m][w]= ...; int T3[m][w]= ...;
int BIGM= w_max;  int BIGM2=maxl (Pmax,TSVmax); int BIGM3= Pmax+TSVmax; //maxl(maxl(T));

 //Decision Variables

dvar int+ x[l][s][m]; //decision variable, pos int
dvar int+ o[m][t] in 0..1;
dvar int+ PMAPmax[l][s];
dvar int+ y[m][w] in 0..1;
dvar int+ PMAP[l][s][t];
dvar int+ TAT[t];
dvar int+ yo[m][w][t] in 0..1;
dvar int+ xo[l][s][m][t];
dvar int+ TAM[l][m];
dvar int+ c[l][s][t] in 0..1;
dvar int+ R[l] in 0..1;
dvar int+ cost2;

// Obj Function
dexpr float cost= alpha * (sum(tt in t) TAT[tt] ) + (1-alpha) * cost2;


// Model
minimize cost;
subject to {

  cost2 == sum(ll in l, ss in s) (PMAPmax[ll,ss] *2/ss + 2*R[ll]*(ss-1));

  forall (mm in m, ll in l:ll<=mm)
  c1: sum (ss in s) x[ll][ss][mm] >= 1;

  forall (mm in m, ll in l:ll<=mm)
    cty: TAM[ll][mm] == sum (ss in s) x[ll][ss][mm];

  forall (mm in m: mm>=2, ll in l:2<=ll<=mm)
    c5: TAM[ll-1,mm] == TAM [ll,mm];

  forall (mm in m)
    c6: TAM [mm][mm] <= w_max;

  forall (mm in m)
    c7: sum (ww in w) (y[mm][ww]*ww) - TAM[mm][mm] == 0;

  forall (mm in m)
  c10v2: sum (tt in t) o[mm][tt] == 1;

  forall (mm in m, ll in l:ll<=mm, ss in s, tt in t){
    l31: xo[ll][ss][mm][tt] - BIGM* o[mm][tt] <= 0;
    l32: xo[ll][ss][mm][tt] - x[ll][ss][mm] <= 0;
    l41:  xo[ll,ss,mm,tt] - x[ll,ss,mm] + (1 - o[mm,tt])*BIGM   >= 0;}

  forall (mm in m, ll in l:ll<=mm, ss in s, tt in t)
    IE5: PMAP[ll][ss][tt] == sum (mm in m) xo[ll,ss,mm,tt];

  forall (mm in m, ll in l:ll<=mm, ss in s, tt in t){
     L6: PMAPmax[ll,ss] >= PMAP[ll,ss,tt];
     L62: PMAPmax[ll,ss] <= PMAP[ll,ss,tt] + (1 - c[ll,ss,tt])*BIGM2;
     L63:  sum (ttt in t) c[ll,ss,ttt] == 1;
        }

forall(layer in l){
    Cr1: sum (ll in l: ll>=layer) PMAPmax[ll,2] >= 1 - BIGM3*(1 - R[layer]);
    Cr2: sum (ll in l: ll>=layer) PMAPmax[ll,2] <=  BIGM3* R[layer];
    }
```

```
    c12r: sum (ss in s) (PMAPmax[1,ss] *2/ss + 2*R[1]*(ss-1))  <= Pmax;   //
    c13r: sum (ss in s, ll in l: ll>=2) (PMAPmax[ll,ss] *2/ss + 2*R[ll]*(ss-1)) <= TSVmax;

forall (mm in m, ww in w, tt in t){
     L1:  y[mm,ww] + o[mm,tt] - yo[mm,ww,tt]  <= 1;
     L2:  y[mm,ww] + o[mm,tt] - 2* yo[mm,ww,tt]  >= 0;
     }

forall (tt in t, mm in m)
    c9: TAT [tt] >= (sum (ww in w) (T[mm,ww]) * yo[mm,ww,tt])  ;

cost1: sum (ss in s) (Cs[ss] * sum(ll in l)  PMAPmax[ll,ss]   ) <= Cmax;

}



/*********************************************
 * OPL 20.1.0.0 Data
 * Author: s286972
 *********************************************/


SheetConnection sheet ("SIC_T.xlsx");
T from SheetRead(sheet,"SIC1!$A$1:$BM$5");
T2 from SheetRead(sheet,"SIC2!$A$1:$BM$5");
T3 from SheetRead(sheet,"SIC3!$A$1:$BM$5");

//M= 5;
Pmax= 100;
TSVmax=300;
Cmax = 560;
alpha = 0.0002;
Cs = [1, 1.3];
```