Self-Aligned 3D Chip Integration Technology and Through-Silicon Serial Data Transmission

THÈSE N^{O} 5182 (2011)

PRÉSENTÉE LE 29 SEPTEMBRE 2011
À LA FACULTÉ SCIENCES ET TECHNIQUES DE L'INGÉNIEUR
GROUPE DE SCIENTIFIQUES STI
PROGRAMME DOCTORAL EN MICROSYSTÈMES ET MICROÉLECTRONIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Fengda SUN

acceptée sur proposition du jury:

Prof. M. A. Ionescu, président du jury
Dr J.-M. Sallese, directeur de thèse
Prof. J. Brugger, rapporteur
Dr W. Escher, rapporteur
Prof. C. Lallement, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2011

To my parents Gangbo and Jingbo

Acknowledgements

With thanks to the great mind of science, Swiss beautiful landscape and École Polytechnique Fédérale de Lausanne, my lovely mother school, I finished my dissertation. My thanks are boundless. At first, I would like to express my great appreciation to Prof. Jacques Giovanola, the dean of the doctoral school, for his strong supports to help me complete my PhD study. Through the communications with Prof. Giovanola, I felt not only the warm and kind understanding, but also the high requirement of a prestigious university has on his graduates. And I would like to thank Dr. Jean-Michel Sallese for accepting me as his student and guiding me through the final tough months to complete this thesis. Without his confidence, encouragement and patience, I can image how much more difficult it will be. This is a most precious experience.

I thank Prof. Mihai Adrian Ionescu for helping me arrange the administrative issues and would like to be the president of my PhD defense. I would like to extend my gratitude and thanks to Dr. Werner Escher, Prof. Christophe Lallement and Prof. Jürgen Brugger for accepting to be the jury members, reading and evaluating this thesis.

The cooperation with Dr. Thomas Brunschwiler is an unforgettable beautiful memory in my life. I admire his passion for science, his devotion to work, his fairness and sincerity. He encouraged me to make progress and firmed my confidence in studies. The colleagues in IBM, Zurich offered me happiness and support as well during my visits. I especially thank Ute Drechsler and Martin Witzig for their precious help on the process.

I greatly thank Prof. Vijaykrishnan Narayanan, Prof. Yuan Xie for the guidance and the cooperation of Asit K. Mishra, Guangyu Sun, Jin Ouyang, and Cong Xu during my visit in Pennsylvania State University. And I thank all the team mates and friends, Hsiang-Yun Cheng, Jishen Zhao, Qiaosha Zou, Jing Xie, Dimin Niu, Wenhui Hu, Tao Zhang, Yibo Chen, Matt Poremba in Penn State University for sharing an interesting short semester with me.

I thank Dr. Deniz Tezcan for giving me the chance to visit IMEC, Belgium. I thank Philip Normans for sharing his knowledge on microfabrication which was really in need for a beginner like me at that time. I thank Dr. Fabrice Duval for sharing his precious experience on polymers. And I enjoyed a lot the exciting travels with Dr. Iaci Pereira during the three months in Belgium, the ice skating and museum visits with Dr. Koen De Munck, Dr. Minoglou Kiki, Arno Aarts, and the happy time with Dr. Pham Nga, Dr. Majeed Bivragh, Goedele Potoms and Rita Verbeeck.

I thank the clean room staff in CMI for the support during my work and I would like to thank Yvan Deillon for the help on dicing and Boris Lunardi for configuring the coating system which ensured the success of the work in the extremely tight schedule. I thank Dr. Harald Van Lintel for sharing the grinding machine and doing the grinding experiment with me. I thank Dr. Sandro Carrara for letting me use their fume hood and offering some chemicals for trial. I also would like to thank Dr. Marc Pastre, Dr. Alex Schmid, Dr. Alain Vachoux and Joseph Guzzardi for helping me with the PC and Linux configuration.

Alessandro Cevrero, Panagiotis Athanasopoulos and I had a busy, productive and happy night together in the office until 3am to catch a deadline. And I enjoyed the days with Andrea Cavallini and Cristina Boero when we shared some experimental tools. I thank Massimo Spina for kindly offering his expertise on AFM, although in the end, the two fine images are not used in this thesis.

I thank Yin Yuan for accompanying me and providing valuable suggestions since we became friends. And I thank all my Chinese friends in EPFL and University of Lausanne, especially Dr. Meng Shen, Yan Yan, Yu Lu, Dr. Hongyan Bi, Hui Yang, Dr. Li Jin, Dr. Hua Zhang, Dr. Peng Xu, Liang Qiao, Hui Chen, Yu Chen, Cheng Yu, Shenqi Xie, Ji Cao, Xiaoqing Zhou, Dr. Lina Huang, Dr. Tiankai Zhu, Dr. Yu Bai, Dr. Zhanbin He, Qunzi Zhang, Chenxiu Lin and many more, for sharing their colorful and exciting lives.

I would like to thank Marie Halm, the secretary of my doctoral program. Moved by her kindness and consideration when I was looking for an advisor four years ago, I decided to come to this beautiful city for my PhD study. Throughout my study here, she offered excellent guidance on the doctoral school formalities and real warm help on my living. I would like to thank Ms. Marie Ligier and Ms. Marianne Wannier for their support as well.

Along the way, many of my teachers in all levels of training and shaping my mind with great hands that I greatly appreciate and wish to acknowledge, especially Ms. Huiyun Ni, Ms. Rong Cao, Ms. Hong Shen, Mr. Fengyi Sun, Mr. Zhenzhen Song, Prof. Jinming Dong, Prof. Bingru Yang, Prof. Dezheng Zhang and etc.

I owe my thanks to my father Dr. Gangbo Sun and mother Jingbo Cao, for giving me a family full of love, trust and devotion to each other. I thank my dad for opening a beautiful world well-connected in all aspects for me through his history study and teaching me the depth and length of a life. My lovely mom, being so serene, tender, organized, diligent and wise, is a life-long example for me, and I thank her. You are also my mentors I cherish forever.

Abstract

The emerging three-dimensional (3D) integration technology is expected to lead to an industry paradigm shift due to its tremendous benefits. Intense research activities are going on about technology, simulation, design, and product prototypes. This thesis work aims at fabricating through-silicon vias (TSVs) on diced processor chips, and later bonding them into a 3D-stacked chip. How to handle and process delicate processor chips with high alignment precision is a key issue. The TSV process to be developed also needs to adapt to this constraint.

Four TSV processes have been studied. Among them, the ring-trench TSV process demonstrates the feasibility of fabricating TSVs with the prevailing dimensions, and the whole-through TSV process achieves the first dummy chip post-processed with TSVs in EPFL although the dimension is rather large to keep a reasonable aspect ratio (AR).

Four self-alignment (SA) techniques have been investigated, among which the gravitational SA and the hydrophobic SA are found to be quite promising. Using gravitational SA, we come to the conclusion that cavities in silicon carrier wafer with a profile angle of 60 ° can align the chips with less than 20 μm inaccuracies. The alignment precision can be improved after adopting more advanced dicing tools instead of using the traditional dicing saws and larger cavity profile angle. Such inaccuracy will be sufficient to align the relatively large TSVs for general products such as 3D image sensors. By fabricating bottom TSVs in the carrier wafer, a 3D silicon interposer idea has been proposed to stack another chip, e.g. a processor chip, on the other side of the carrier wafer.

But stacking microprocessor chips fabricated with TSVs will require higher alignment precision. A hydrophobic SA technique using the surface tension force generated by the water-to-air interfaces around the pads can greatly reduce the alignment inaccuracy to less than 1 μm. This low-cost and high throughput SA procedure is processed in air, fully-compatible with current fabrication technologies, and highly stable and repeatable. We present a theoretical meniscus model to predict SA results and to provide the design rules. This technique is quite promising for advanced 3D applications involving logic and heterogeneous stacking.

As TSVs' dimensions in the chip-level 3D integration are constrained by the chip-level processes, such as bonding, the smallest TSVs might still be about 5 μm. Thus, the area occupied by the TSVs cannot be neglected. Fortunately, TSVs can withstand very high bandwidths, meaning that data can be serialized and transmitted using less numbers of TSVs. With 20 μm TSVs, the 2-Gb/s 8:1 serial link implemented saves 75% of the area of its 8-bit parallel counterpart. The quasi-serial link proposed can effectively balance the inter-layer bandwidth and the serial links' area consumption. The area model of the serial or quasi-serial links working under higher frequencies provides some guidelines to choose the proper serial link design, and it also predicts that when TSV diameter shrinks to 5 μm, it will be difficult to keep this area benefit if without some novel circuit design techniques.

As the serial links can be implemented with less area, the bandwidth per unit area is increased. Two scenarios are studied, single-port memory access and multi-port memory access. The expanded inter-layer bandwidth by serialization does not improve the system performance because of the bus-bottleneck problem. In the latter scenario, the inter-layer ultra-wide bandwidth can be exploited as each memory bank can be accessed randomly through the NoC. Thus further widening the inter-layer bandwidth through serialization, the system performance will be improved.

Keywords: 3D integration, 3D chip integration, memory-on-logic stacking, through silicon via (TSV), post-processing, self-alignment (SA), gravitational SA, hydrophobic SA, surface tension force, self-assembled monolayer (SAM), serial transmission, inter-layer bandwidth, serial link, Network on chip (NoC), chip multi-processor (CMP), serializer and deserializer (SerDes), multiplexer (MUX), demultiplexer (DEMUX)

Résumé

L'émergence de la technologie d'intégration 3D devrait repousser les limites technologiques suite aux avantages que promet celle-ci. Des recherches poussées sont en cours portant sur divers aspects, tels que la simulation, le design et les prototypes en tant que produits. Dans ce contexte, le sujet de cette thèse réside dans la fabrication de connections (TSV) réalisées à travers le silicium afin de fusionner en un seul empilement 3D des puces électroniques déjà découpées. Optimiser la manipulation et l'intégration de ces puces électroniques avec une grande précision demeure un défi majeur qui doit être pris en compte par le procédé de fabrication de ces "TSV".

Quatre procédés ont été étudiés. Parmi ceux-ci, le "ring-trench" est celui qui a permis de démontrer qu'il était possible de fabriquer ces "TSV" avec les dimensions requises. Ainsi a pu être réalisée à l'EPFL la première connexion à travers le silicium sur un échantillon témoin, et ceci bien que les dimensions étaient relativement grandes pour obtenir un rapport géométrique raisonnable.

D'autres parts, quatre techniques d'auto alignement de ces puces ont été évaluées, parmi lesquelles l'alignement par gravité et l'alignement obtenu sous l'effet de la tension superficielle. Tous deux très prometteurs. L'alignement par gravité démontre qu'il est possible d'atteindre une précision inférieure à 20 micromètres lorsque l'angle de la cavité, qui sert d'alignement dans le "wafer" de silicium, est de 60 °. Une meilleure précision pourrait être obtenue si des outils de découpes plus évolués étaient utilisés, ainsi qu'avec des angles plus verticaux. Cependant, de telles imprécisions restent compatibles pour aligner des "TSV" relativement larges qui sont destinés à des applications telles que les capteurs d'images. Par ailleurs, en fabricant des "TSV" sur la face arrière du "wafer" en silicium qui sert d'intermédiaire pour ces placements, nous avons proposé l'idée d'interposer un autre film de silicium afin d'empiler une puce sur l'autre face, telle que par exemple un processeur.

Toutefois, empiler des microprocesseurs nécessite une plus grande précision dans l'alignement. Celui ci peut être réalisé avec la technique d'auto alignement basé sur la tension superficielle qui apparait en présence de surfaces hydrophiles et hydrophobes. Dans ce cas, la précision devient meilleure que le micromètre. Cette technique, à bas coût et à fort rendement, est compatible avec les technologies standards. Elle est en outre stable et reproductible. Nous présentons également une étude théorique qui permet de prédire l'alignement des puces et qui pourrait être utilisée pour définir des règles de "design". Il est à noter que cette approche est prometteuse dans le sens où elle permettrait de combiner des circuits assez hétérogènes.

Etant donné que les dimensions de ces TSV dépendent également des procédés de fabrication de la puce elle-même, tels que la connexion, leurs dimensions minimales devraient être de 5 micromètres, pas moins. Ainsi, la surface qui leur est associée pourrait ne plus être négligeable. Mais un des avantages de ces "TSV" est qu'ils offrent une bande passante extrêmement large, ce qui implique que des données peuvent être transmises en série à travers un nombre réduit de ces éléments. Si l'on considère des "TSV" de 20 micromètres de diamètre, une transmission série de type 2-Gb/s 8:1 réduirait de 75% la surface occupée par une transmission équivalente de 8-bit sous un format parallèle. Il s'avère que la transmission partielle en série peut compenser de manière efficace la surface occupée par cette technologie, tout en restant compatible avec des bandes passantes réalistes. Ainsi, le calcul de la surface utilisée pour la transmission série ou pseudo-série à plus haute fréquence permet de choisir la meilleure topologie. Il permet aussi de souligner que lorsque les "TSV" atteindront 5 micromètres, cet avantage sera difficile à conserver en termes de gain en surface sans devoir introduire de nouvelles techniques de "design" de ce genre de circuit.

Par ailleurs, étant donné que les connections de type série demandent moins de surface, la largeur de bande par unité de surface en sera augmentée. Deux scénarios sont étudiés, à savoir un accès mémoire "single port" ou bien "multiple-port".

Mais nous avons mis en évidence que l'accroissement de la largeur de bande entre les différentes couches obtenue par la transmission série n'améliore pas forcément les performances du système, ceci à cause de l'engorgement provenant du bus de transmission lui-même.

Finalement, dans une dernière analyse, nous mentionnons que la transmission entre les différentes couches à ultra haute fréquence pourrait être exploitée dans le cas de la mémoire qui serait accessible de manière aléatoire à travers le "Network on Chip". Ainsi, les performances du système seraient améliorées en augmentant la bande passante au moyen de la transmission série entre les couches

Mots clefs: 3D integration, 3D chip integration, memory-on-logic stacking, through silicon via (TSV), post-processing, self-alignment (SA), gravitational SA, hydrophobic SA, surface tension force, self-assembled monolayer (SAM), serial transmission, inter-layer bandwidth, serial link, Network on chip (NoC), chip multi-processor (CMP), serializer and deserializer (SerDes), multiplexer (MUX), demultiplexer (DEMUX)

Glossary

Acronym	Signification
2EH	2-ethyl-1-hexanol
ALD	Atomic Layer Deposition
AR	Aspect Ratio
BARC	Bottom-Layer Anti-Reflective Coating
BCB	Benzocylcobutene
BEOL	Back End of Line
BHF	Buffered Hydrofluoric Acid
CMI	Center of Micronanotechnology
CML	Current Mode Logic
CMP	Chemical Mechanical Polishing
CMP	Chip Multiprocessor (in Chapter 6)
CSCM	Chip Scale Camera Module
C-to-W	Chip-to-Wafer
CTE	Coefficient of Thermal Expansion
CVD	Chemical Vapor Deposition
DEMUX	Demultiplexer
DMA	Distributed Memory Architecture
DRAM	Dynamic Random Access Memory
DRIE	Deep Reactive Ion Etching
DSC	Differential Scanning Calorimetry
FEOL	Front End of Line
FIB	Focused Ion Beam
F-W SA	Fully-Wetted Self-Alignment
HAR	High Aspect Ratio
HPIMS	High Power Impulse Magnetron Sputtering
HVM	High Volume Manufacturing
ICP	Inductive Coupled Plasma
ICV	Inter Chip Via
IMP	Ionized Metal Plasma
ITRS	International Technology Roadmap for Semiconductors
LPCVD	Low Pressure Chemical Vapor Deposition
KGD	Known Good Die
KOD	Keep Out Distance
KOZ	Keep Out Zone
MEK	Methylethyl
MEOL	Middle End of Line
MOCVD	Metal Organic Chemical Vapor Deposition
MUX	Multiplexer
NoC	Network on Chip
NUCA	Non-Uniform Cache Access
P-A SA	Pad-Assisted SA
PECVD	Plasma Enhanced Chemical Vapor Deposition
PGMEA	Propylene Glycol Monometryl Ether Acetate
PVD	Physical Vapor Deposition
P-W SA	Partially-Wetted SA

Acronym	**Signification**
RC	Routing Computation
RDL	ReDistribution Layer
RIE	Reactive Ion Etching
RMS	Root Mean Square
SA	Self-Alignment
SACVD	Sub Atmospheric Chemical Vapor Deposition
SAM	Self-Assembled Monolayer
SEM	Scanning Electron Microscope
SerDes	Serializer and Deserializer
SMA	Shared Memory Architecture
S-NUCA	Static Non-Uniform Cache Access
SOI	Silicon on Insulator
SOD	Spin-On Dielectrics
TCV	Through Chip Via
TSV	Through Silicon Via
UCA	Uniform Cache Access
VC	Virtual Channel
W-to-W	Wafer-to-Wafer

Content

Chapter 1 Introduction........ 1
- 1.1 History........ 1
- 1.2 Key Technologies of 3D Integration........ 3
  - 1.2.1 Through Silicon Via........ 3
  - 1.2.2 Thinning........ 11
  - 1.2.3 Bonding........ 13
  - 1.2.4 3D Integration Schemes........ 13
- 1.3 3D Applications........ 14
- 1.4 3D Circuit Design and System Architecture........ 15
- 1.5 Outline of This Thesis........ 15
- Reference........ 16

Chapter 2 TSV Fabrication Technologies........ 20
- 2.1 Target........ 21
- 2.2 Core-Hole TSV Process........ 21
  - 2.2.1 Process Flow........ 21
  - 2.2.2 Silicon Dry Etch........ 21
  - 2.2.3 Spray Coating in High Aspect Ratio (HAR) Topography........ 31
  - 2.2.4 Core-Hole TSV Detailed Process........ 33
- 2.3 Ring-Trench TSV Process........ 36
  - 2.3.1 Process Flow........ 36
  - 2.3.2 Filling Polymer in HAR Microstructures and Pattern Transfer........ 37
  - 2.3.3 Electroplating Mold for Cu Electroplating........ 42
  - 2.3.4 Ring-Trench TSV Detailed Process........ 45
  - 2.3.5 A Modified Ring-Trench TSV Process........ 49
- 2.4 Step TSV Process: An Alternative Solution........ 53
  - 2.4.1 Process Flow........ 53
  - 2.4.2 Step TSV Detailed Process........ 53
- 2.5 Whole-Through TSV Process: A Feasible Solution to TSV Post-Processing on Dies........ 54
  - 2.5.1 Introduction........ 54
  - 2.5.2 Parylene C on Through-Wafer Holes and Double-Side Sputtering of Al on Parylene........ 55
  - 2.5.3 Die Thinning........ 55
  - 2.5.4 BEOL Metal (Al) as Dry Etch Hard Mask........ 57
  - 2.5.5 Whole-Through TSV Detailed Process........ 59

2.6 Summary 60
Reference 62
Chapter 3 Alignment and Stacking Techniques for Die-Level 3D Integration 65
3.1 Feasibility of Die-Level Processing and Stacking 65
3.1.1 Materials for Temporary Die Attachment 65
3.1.2 Precise Die Attachment/Stacking 66
3.2 Picturing SA Techniques for 3D 67
3.2.1 Gravitational SA 67
3.2.2 Electrostatic SA 68
3.2.3 Magnetic SA 68
3.2.4 Hydrophobic SA 69
3.3 Silicon Interposer and Gravitational SA Technique 69
3.3.1 The Idea 69
3.3.2 Proof of Concept 73
3.3.3 Discussion of the Process, Further Improvements and Conclusion 74
3.4 Magnetic SA Technique 76
3.4.1 The Idea 76
3.4.2 Theoretical Analysis 76
3.4.3 Initial Experiment Results and Summary 77
Reference 77
Chapter 4 Hydrophobic SA for Heterogeneous 3D Integration 79
4.1 Background 79
4.1.1 The Advantage of Hydrophobic SA over the Other SA Techniques 79
4.1.2 Hydrophobic SA 80
4.2 Our Method 81
4.3 Experiment Evaluation 82
4.3.1 Experimented Self-Alignment Processes 82
4.3.2 Sample Preparation 83
4.3.3 Experiment Results 85
4.4 Meniscus Model 89
4.4.1 Analytical Model of SA 89
4.4.2 Modeling Results 90
4.5 A Revisit of Magnetic SA in Comparison of Hydrophobic SA 93
4.6 Summary 94
Reference 95
Chapter 5 Inter-Layer Serial Links for 3D Chip Integration 97

5.1 Introduction 97
5.2 TSV Characteristics and Performance 97
  5.2.1 TSV Area 97
  5.2.2 TSV Electrical Performance 99
  5.2.3 TSV I/O Power Estimation 104
  5.2.4 Summary 105
5.3 A 2-Gb/s Inter-Layer Serial Link 105
  5.3.1 Circuit Design for Each Tier 105
  5.3.2 TSV Insertion and Layout Simulation 107
  5.3.3 Results 108
5.4 A 8-Gb/s Inter-Layer Quasi-Serial Link 109
5.5 Inter-Layer Serial Link Design Methodology 110
5.6 Summary 113
Reference 114
Chapter 6 System-Level Exploration of Inter-Layer Serial Transmission 117
6.1 Background 117
  6.1.1 Memory-on-Logic 3D Stacking and Our Research Region 117
  6.1.2 Networks-on-Chip (NoCs) 118
  6.1.3 UCA cache and NUCA Cache 118
  6.1.4 3D Memory-on-Core System Architectures 119
6.2 Proofs for Inter-Layer Serialization 120
  6.2.1 Area Considerations 120
  6.2.2 Power Consumption Considerations 121
6.3 Serialized Inter-Layer Transmission in 3D SMA 122
  6.3.1 Simulation Platform 122
  6.3.2 Simulation Results 123
6.4 Serialized Inter-Layer Transmission in 3D DMA 123
  6.4.1 Simulation Platform 123
  6.4.2 Simulation Results 124
6.5 Summary 126
Reference 127
Chapter 7 Conclusion and Outlook 129
7.1 General Conclusion 129
  7.1.1 Technology Aspects 129
  7.1.2 Through-Silicon Serial Data Transmission Issues 130
7.2 Outlook 131

Chapter 1

Introduction

1.1 History

The invention of transistor by John Bardeen, Walter Brattain and William Shockley in 1947 and the integrated circuit by Jack Kilby in 1959 were great innovations. Since the Moore's Law was postulated in 1965, the semiconductor industry continuously followed these scaling rules. In 1980's, people saw the possible end of the Moore's Law. Although this limit has always been postponed, the end of the Moore's law is becoming a reality: we are facing atomic scale issues in ultra deep miniaturization. Researchers have been working hard to find replacement solutions. Intensive research is going not only on "exotic" semiconductor materials, like diamond and graphene, but also on computing technologies such as quantum and bio-inspired computation. However, among all the solutions, exploring the 3D vertical space using existing semiconductor infrastructures is surely the most natural and prospective choice. This is known as "3D integration". It is truly not a new idea. William Shockley made a technology patent entitled "Semiconductive Wafer and Method of Making the Same" in 1965 [1]. It is actually what we call "Through Silicon Vias (TSV)" or "3D vias" now, which is the key of the recent 3D technology. But no one really paid attention to this idea at that time.

In 1985, a Noble Physics laureate, Richard Feynman, said in his address "the Computing machines in the Future" and I quote, "*Another direction of improvement (of computing power) is to make physical machines three dimensional instead of all on a surface of a chip. That can be done in stages instead of all at once - you can have several layers and then add many more layers as time goes on.*" In the following year, Yoichi Akasaka published a paper in Proceedings of IEEE with a title "Three-Dimensional IC Trends" [2]. Since then, 3D research has been world-wide popular in different levels for the past 20 years. Today its era has come.

The Advanced Semiconductor Engineering Group (ASE Group, the world's largest provider of independent semiconductor manufacturing services in assembly and test) has proposed a 3D roadmap in Fig. 1.1, showing the aimed 3D applications with the predicted time to market.

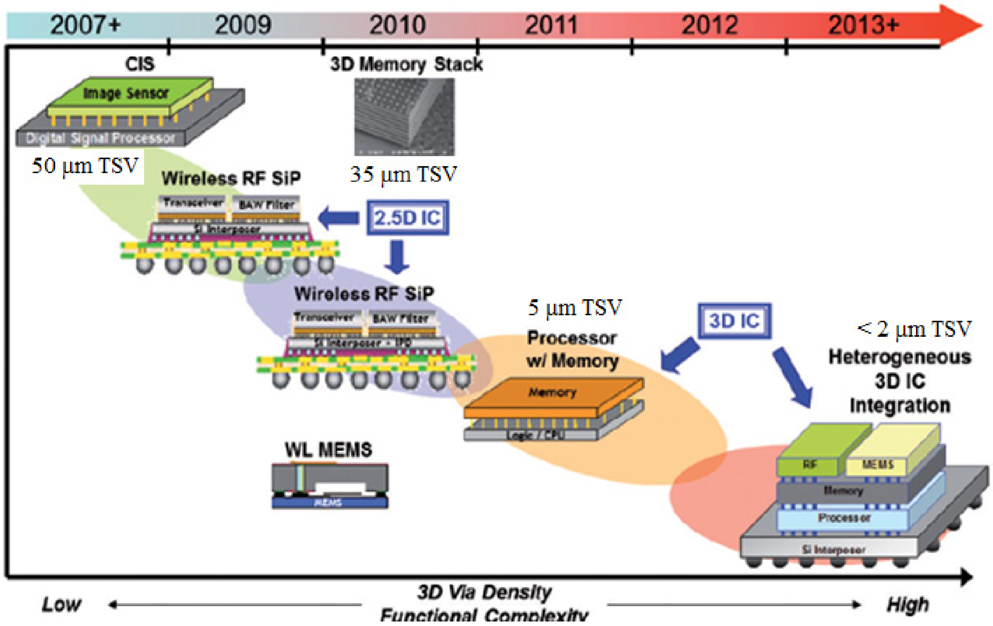
Consumer electronics, for example, cell phones, digital cameras, PDAs audio players and mobile gaming, demand innovative form factors and styling. By removing the on-board connections by vertical interconnections in the chip stack, the form factor is greatly reduced. Toshiba's commercialized Chip Scale Camera Module (CSCM) equipped with a Dynastron® CMOS image sensor manufactured with Through-Chip Via (TCV) technology. A downsizing of approximately 64% compared to conventional module has been achieved (Fig. 1.2) [3].

Now, considering computing performance, for a uni-core processor, 3 GHz clocking is almost the upper limit and represents the end of the line in clock speed improvements. In one third of a nanosecond, light moves about 4 inches or 10 cm, not much distance when the chips have miles of internal wiring.

Instead of pursuing higher clock frequency, more cores are stuffed into one chip. But the number of cores we can put in a single chip is an issue. From the application point of view, except for some critical ones like scientific calculation and videos that multi-core architectures will gain efficiency, other applications like web surfing certainly do not require so many cores. Making multi-core chips workable is not only very challenging with respect to the technology itself, but also making the memory wall problem even more difficult, since larger on-chip caches are then required to improve the performances. But inevitably, cache hit ratio declines with size, and so does the performance. What is required is a system-level solution, higher memory capacity with new memory architectures. Directly stacking memory on core or distributing cores and memories into several layers creates more design possibilities with higher system performance.

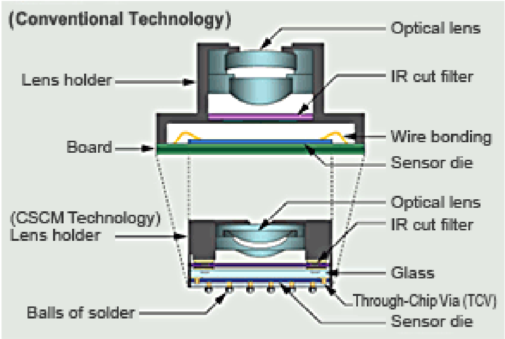
People also put hope on 3D integration to enable the super stacking of heterogeneous chips with different functions such as MEMS, RF, CPU, memories and etc. Without 3D, the one-plane implementation is form factor prohibitive or even not functional.

Great potential and significance of 3D integration have been clearly demonstrated. However, to get the enabling technologies and to achieve successful 3D systems, circuit designers and system architects should have a good understanding of the technology. In the next sections, we will introduce the key technologies in 3D integration, the upper levels of the 3D ecological systems including the EDA tool support and 3D circuit and system study.



2007+
2009
2010
2011
2012
2013+
CIS
Image Sensor
Digital Signal Processor
50 μm TSV
3D Memory Stack
35 μm TSV
Wireless RF SiP
Transceiver
BAW Filter
Si Interposer
2.5D IC
Wireless RF SiP
Transceiver
BAW Filter
Si Interposer + IPD
5 μm TSV
Processor
w/ Memory
Memory
Logic / CPU
3D IC
< 2 μm TSV
Heterogeneous
3D IC
Integration
RF
MEMS
Memory
Processor
Si Interposer
WL MEMS
MEMS
3D Via Density
Functional Complexity
Low
High

Figure 1.1 3D Roadmap. Courtesy of ASE Group.



(Conventional Technology)
Optical lens
Lens holder
IR cut filter
Board
Wire bonding
Sensor die
(CSCM Technology)
Lens holder
Optical lens
IR cut filter
Glass
Balls of solder
Through-Chip Via (TCV)
Sensor die

Figure 1.2 Toshiba commercialized CSCM with TCV (= TSV) technology [3].

1.2 Key Technologies of 3D Integration

1.2.1 Through Silicon Via

TSV is the soul of 3D technology. It is the interconnection penetrating the wafer or chip that transmits signals from one stratum to another. Basically, to get a TSV, we need to first open some via holes, then form an isolation layer around the sidewall and finally deposit conducting materials inside the holes. Two main stream TSV processes exist, front end of line (FEOL) TSV process and back end of line (BEOL) TSV process. Later, a middle end of line (MEOL) TSV process has been proposed. TSVs can also be processed after bonding. As it is not as popular as the other three TSV processes, it will be introduced as the last solution. The four processes are illustrated in Fig. 1.3.

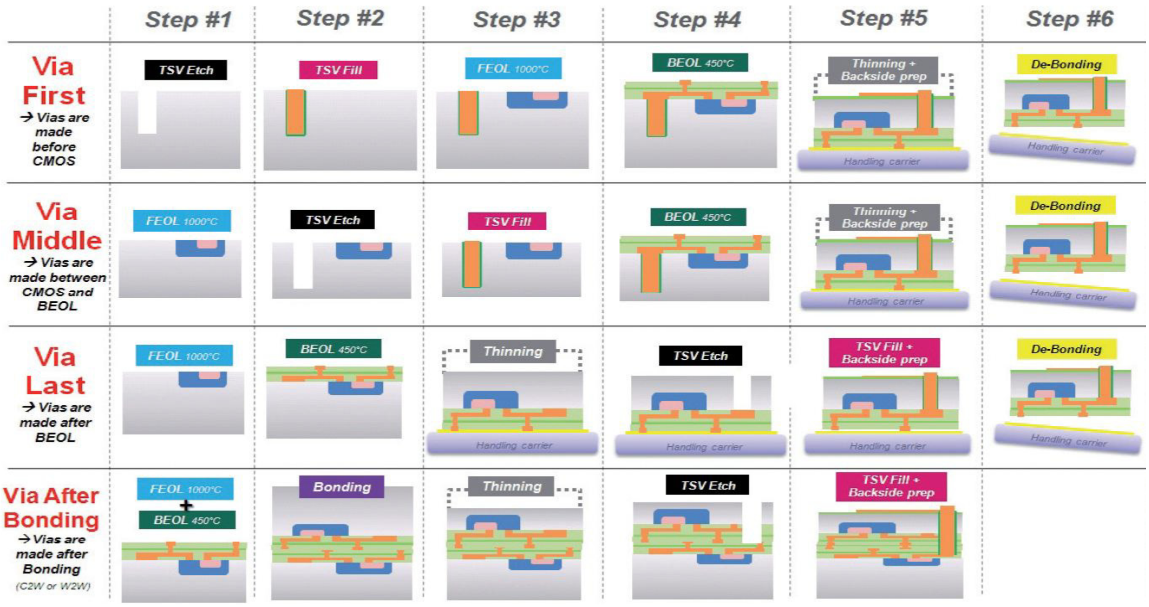

Step #1
Step #2
Step #3
Step #4
Step #5
Step #6
Via First
→ Vias are made before CMOS
TSV Etch
TSV Fill
FEOL 1000°C
BEOL 450°C
Thinning + Backside prep
Handling carrier
De-Bonding
Via Middle
→ Vias are made between CMOS and BEOL
Via Last
→ Vias are made after BEOL
Thinning
TSV Fill + Backside prep
Via After Bonding
→ Vias are made after Bonding
(C2W or W2W)
Bonding

Figure 1.3 Main TSV integration scenarios [3]. Courtesy of Yole Development.

FEOL TSV or Via-First Approach

For FEOL TSV process, TSV holes are etched prior to CMOS transistor or MEMS microstructure fabrications. For TSV used in CMOS chips, the materials need to be CMOS-compatible. Thermal oxidation of silicon for TSV isolation can be used, which is a big advantage since it achieves conformal lining without the limitation of topography, while other methods such as plasma enhanced chemical vapor deposition (PECVD) and sputtering processes cannot.

An early trial was presented by Chow *et al.* in 2000 [4]. In this process, 20 μm circular holes were etched by double-side DRIE process through a 400 μm thick wafer. Through-hole insulation and conduction layer were achieved by thermal oxidation of silicon and Low Pressure Chemical Vapor Deposition (LPCVD) of polysilicon. Next, phosphorus diffusion at 1000 °C rendered the polysilicon conducting.

In 2008, Kawano *et al.* presented polysilicon TSVs, with 18 μm diameter and 140 μm depth, working in a 8-layer dynamic random access memory (DRAM) stack [5] (Fig. 1.4). In the future, it is expected that tiny and dense TSVs will have less than 1 μm diameter and only a few micron height. Only polysilicon can be used as conducting material for FEOL TSV since metal cannot withstand high temperatures. Thus, via resistivity will be higher compared to metals. Many challenging problems exist, e.g. handling of thinned and fragile wafers, wafer-level bonding to achieve interlayer connections and so on (will be discussed in Section 1.2.2 and Section 1.2.3). This is a foundry-only solution.

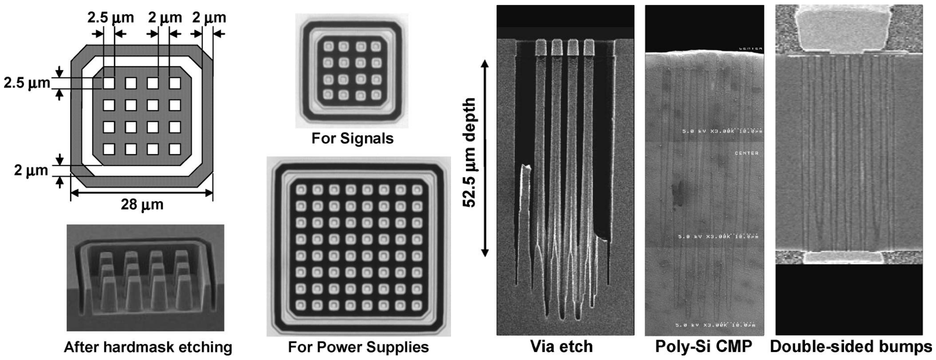

2.5 µm
2 µm
2 µm
2.5 µm
2 µm
28 µm
After hardmask etching
For Signals
For Power Supplies
52.5 µm depth
Via etch
Poly-Si CMP
Double-sided bumps

Figure 1.4 Poly-Si TSV process proposed by Kawano *et al*. [5]. (a) Design rules for poly-Si TSVs and plan views before poly-Si CVD. (b) Poly-Si TSV cross sections at several process steps.

From a MEMS processing perspective, the "via-first" approach is by all means preferred in most cases since completed delicate MEMS structures will not benefit from additional wafer level post processing and handling. In 2007, Bauer presented a unique via-first process for low-frequency 3D MEMS applications (Fig. 1.5) [6]. The process began with the formation of 10 - 15 µm wide ring trenches using DRIE process, achieving the necessary high aspect ratio features up to 600 µm thick low-resistivity silicon substrate. Then the wafer was subject to a high temperature filling of the trenches by a dielectric material. Finally, a chemical mechanical polishing (CMP) process was applied to the backside of the wafer until the via plugs were isolated from the bulk of the wafer. Typical resistance in a standard via connection of 100 µm diameter in a 400 µm thick substrate is on the order of 1 Ω and such a resistance will constitute an unacceptable loss in most RF designs.

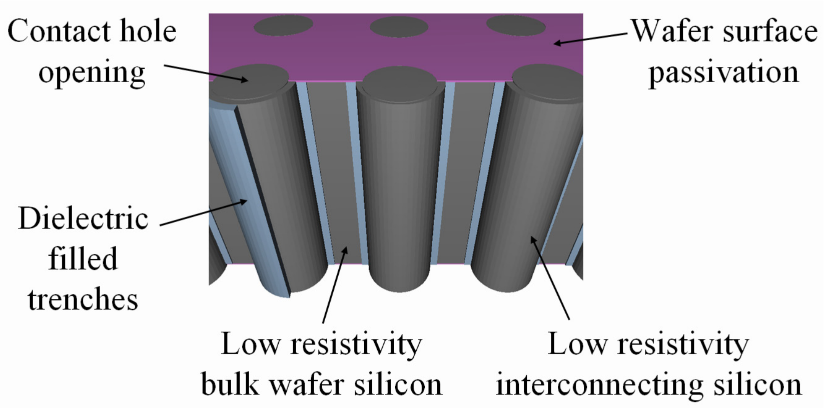

Contact hole opening
Wafer surface passivation
Dielectric filled trenches
Low resistivity bulk wafer silicon
Low resistivity interconnecting silicon

Figure 1.5 Schematic cross section of the TSV structure proposed by Bauer [6].

BEOL TSV or Via-Last Approach

Most research institutes and companies went for BEOL solutions. In this via-last approach, TSVs are fabricated after the device integration and the BEOL metal processes. The thermal budget is thus constrained by BEOL metal process that must be kept below 450 °C. Even temperatures lower than 300 °C are preferred. If the chip is exposed to temperatures greater than 450 °C, CMOS metallization will start diffusing significantly into silicon, resulting in non-functional circuits [7, 8].

Post-processing might also induce device damages. Technologies such as e-beam evaporation generate X-ray radiation that might cause threshold voltage shifts in transistors and deteriorates the circuits [8], so that should be avoided. Reactive Ion Etch (RIE) and Deep Reactive Ion Etch (DRIE) processes used to fabricate TSV holes may also cause some radiation-induced damages in CMOS circuits due to charging from high energy UV photons, ions or electrons [9]. Moreover, plasma processing might cause dielectric breakdown of gate dielectrics [9]. To minimize the potential damages to CMOS devices, BEOL TSVs are better processed from the back-side of the wafer or chip. This is called "backside TSV". Conversely, "frontside TSV" exists as well.

Early developed BEOL TSVs were etched through the full thickness of the wafer. The diameters ranged from about 100 μm to more than 200 μm, mainly depending on the etching technology used, either KOH or DRIE.

KOH-etched TSVs have to be quite large because of the etching mechanism (Fig. 1.6). In 1994, Linder *et al*. presented the first results of vias fabricated by partial metallization of KOH-etched wafer through-holes [10]. The process was based on photolithography in the KOH etch pit using photosensitive polyimide and subsequent electroplating of Au into the polyimide mould. The resistance of the vias was around 20 mΩ and their capacitance in the range of 2 pF to 5 pF. In 1996 Linder presented an improved via process in which electrodeposited photoresist was employed for patterning the via metallization [11]. A via fabricated according to this process is shown in Fig. 1.7-a. This via has a resistance of 10 mΩ only and a parasitic capacitance in the range of 2 pF to 5 pF. In 1996, Burger *et al*. fabricated multiple via Al metal connections using KOH-etched through-holes with a shadow mask. The via resistance was reported to be below 5 mΩ and the parasitic capacitance less than 1 pF. Finally, in 1998/1999, Heschel *et al*. proposed a process based on electrodeposition of photoresist used as a mould for electroplating of copper via metal from evaporated Ti/Au seed layer [12, 13, 14]. Parasitic capacitance and resistance of these vias were further reduced to 0.7 pF and in the range of 700 mΩ. This is shown in Fig. 1.7-b.

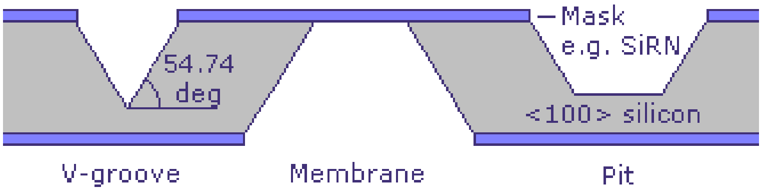
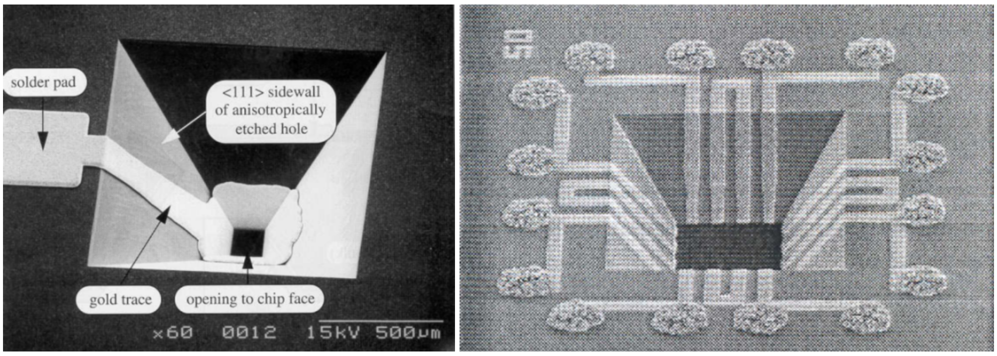


Figure 1.6 KOH etching mechanism.



a) b)

Figure 1.7 TSV technologies presented by (a) Linder *et al*. [11] and (b) Heschel *et al*. [13].

In the continuous effort to increase the via density, DRIE producing smaller and straight TSVs was adopted. With the same wafer thickness, the via aspect ratio was increased which rendered even more difficult the problem of sidewall isolation and metallization. Conventional physical vapor deposition (PVD) methods were not able to achieve highly conformal deposition on the via sidewall.

In 2003, Rasmussen *et al*. presented a hallow TSV structure using Parylene as dielectric isolation. Highly conformal deposition of TiN/Cu by MetalOrganic Chemical Vapor Deposition (MOCVD) was employed to metallize parylene. A 100 μm wide and 380 μm deep through-wafer TSVs with 5 μm thick parylene for isolation and 1 μm thick Cu layer for conduction was fabricated. The via resistance and capacitance were 161 mΩ - 177 mΩ and 1.7 pF [16].

The benefit coming with TSV shrinkage is reduced TSV parasitics and thus improved electrical performance. An example of coaxial shield TSV structure is illustrated in Fig. 1.8. This was proposed by Ho *et al.* [16] in 2006 for high frequency application. The process consists in: 1) 300 μm diameter through wafer (250 μm thick) via-hole etching; 2) 1 μm SiO_2 and 1000 Å silicon nitride insulation liner deposition; 3) 1 μm thick Cu ground plane deposition along the via and the substrate; 4) screen-printing of 112 μm thick SU8 into the via holes for the dielectrics; 5) removing core centers in the SU8-filled via holes by photolithography; 6) bottom-up electroplating to fill the center cores with Cu. With this proposed coaxial shield via structure, the TSVs fabricated in low-resistivity silicon substrate achieved a minimum RF signal transmission loss under 40 GHz. It also provides an elegant method to achieve thick dielectrics along the via sidewall.

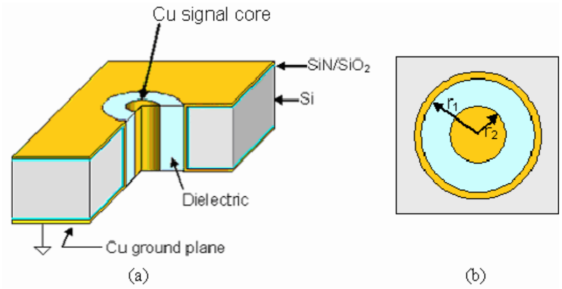
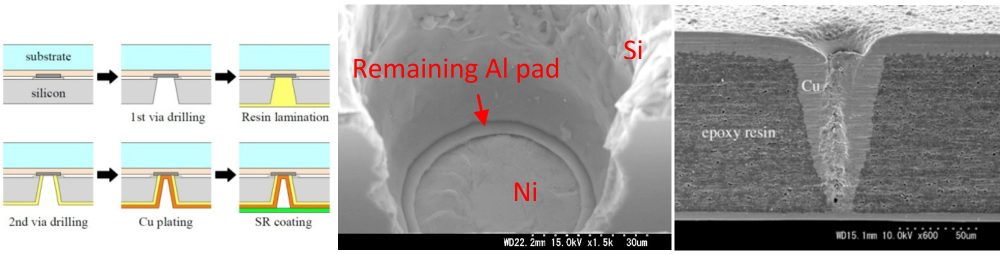


Figure 1.8 Coaxial shield TSV structure [16].

Although these TSV densities and their number were quite limited because of their large size, 3D demonstrators had already shown a great reduction of form factor, thus making this technology commercially competitive. An early attempt of commercial application of vias was in a MEMS microphone consisting of stacked chips electrically interconnected by vias [12, 17]. In 2008, Toshiba produced the first commercial 3D CMOS image sensor with a cheap and streamlined TSV (they call "Through Chip Via (TCV)") process [3]. For instance, the CMOS image sensor wafer was attached to a handling glass wafer by patterned adhesives and thinned by backside grinding and polishing. A first via was drilled by a pad-stop (stop at Ni plated above the Al pads) laser ablation technique from the wafer backside with backside alignment to the pads. A resin was laminated using vacuum lamination equipment to completely fill the via holes. A second via drilling was performed at the center of the first via. Subtractive or semi-additive PCB electroplating processes was employed to pattern the 10 μm Cu plate on the back of the wafer and inside the vias. Typical via diameter was 60 μm at the top and 30 μm at the bottom. The process flow and the cross-section of Cu plated via are shown in Fig. 1.9 [19, 20].

To optimize the area and get better electrical characteristics of the TSVs in order to use them for more advanced 3D applications, an effort was done to develop smaller TSVs with higher density. Wafers were thinned down to less than 100 μm, mostly 50 μm, and sometimes even 15 μm. Trade-offs exist. Thicker wafers can be easily handled and thinned, so a tall TSV is preferred. Then, In order to keep the via density, higher aspect ratios are necessary. But the aspect ratio is constrained by many processing steps, such as etching, deposition and electroplating.

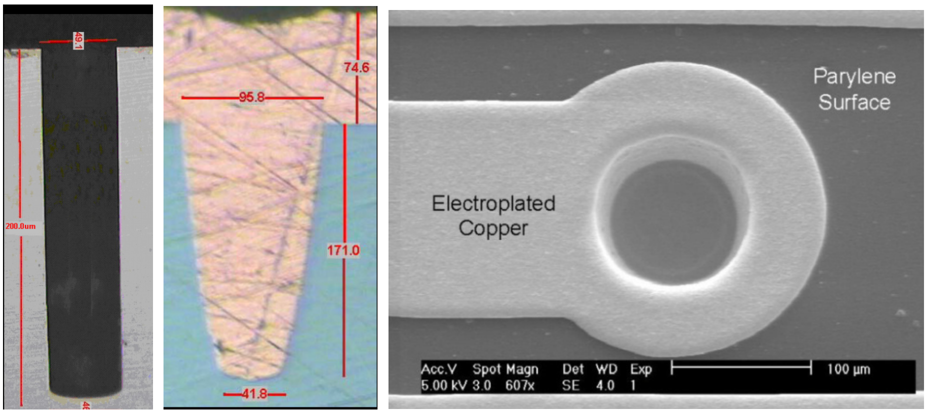


(a) (b) (c)

Figure 1.9 Toshiba TSV technology. (a) Via process flow used in Toshiba 3D image sensor. (b) First via drilling. (c) Cross-section of Cu-plated via.

Nagarajan *et al*. proposed a tapered TSV process in 2006 [21]. 200 μm-deep 100 μm-wide 80°-tapered via holes were etched by a dual-etch process (DRIE followed by continuous silicon etch) and later completely filled with copper by a DC plating process with equal ON/OFF time of 20 ms. The optical images of the via cross-sections in different process steps are shown in Fig. 1. 10-a and Fig. 1. 10-b. No electrical characteristic has been reported.

Another sloped TSV process for 3D wafer-level packaging was presented by Tezcan *et al.* in 2007 [22]. A silicon wafer was glued on a carrier pyrex wafer and thinned down to 100 μm. Then 50 - 100 μm wide 80 °-tapered vias were etched until the landing Al pads, and covered with parylene for isolation. Subsequently, the insulator material at the bottom of the via was removed in order to reach the metal contact. The related photolithography step was done through spray resist coating technique. Finally a seed layer was sputtered, followed by Cu plating. With 140 μm top diameter, 100 μm bottom diameter, 100 μm height, 1 - 2 μm thick parylene as dielectric isolation and 5 - 20 μm thick copper electroplated on the via sidewall for current conduction. A via resistance in the range of 20 - 30 mΩ was reported. A cross-section view of such a via is shown in Fig. 1.10-c.



(a) (b) (c)

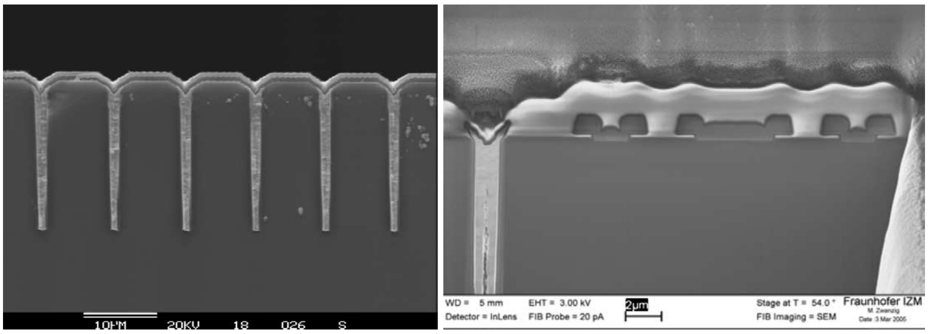
Figure 1.10 TSV technology proposed by Nagarajan *et al*. [21] and Tezcan *et al*. [22]. (a) Via hole after 1st DRIE in [21]. (b) Via hole after 2nd etch in [21] and filled with copper. (c) Via cross section view in [22].

Next, in 2007, Benkart *et al*. presented a 5 µm-long and 10 µm-tall square TSV process with a thermal budget below 300 °C [23]. The wafer was prepared in a unique way for thinning which includes markers definition for backside alignment and epitaxial overgrowth with a high boron doped etch stop layer in the silicon wafers before CMOS fabrication. Mechanical grinding was performed on the backside of the wafer which was glued to a carrier wafer, followed by chemical etching down to the etch stop layer. By adding a step of anodic oxidation at < 50 °C to the low temperature CVD of SiO_2, leakage current between the signal lines or into the substrate was decreased, so the isolation was enhanced. The vias were filled with Cu electroplating. As it was a through-hole TSV technology, metal redistribution was required to get the connection between the TSV and the circuitry.

Very recently in 2011, Ohara *et al.* reported their development of 5 µm-diameter backside TSV technology [24]. The chip was supported on a glass or silicon carrier wafer and thinned down to 15 µm by grinding and CMP. Then, about 1 µm SiO_2 was deposited as an isolation layer / hard mask . The TSVs were created with Bosch process and then lined with SiO_2 (500 nm). The bottoms of the insulated TSVs were opened by anisotropic SiO_2 etching using the thicker backside oxide layer as partially sacrificial mask for the etching. This is a much cheaper and simpler method to achieve TSV contact comparing to the afore-mentioned way Tezcan *et al*. applied in 2007 [ ].

The above TSVs are all based on plated copper as the conducting material. To further decrease the TSV dimension, plating is not sufficient to get a uniform coverage on very high aspect ratio structures. Instead, CVD has to be used.

In 2005, a 1.2 µm-diameter, 25 µm-tall TSV technology was demonstrated by Wieland *et al*. [25] (Fig. 1.11). The dielectric layer deposited by sub atmospheric chemical vapor deposition (SACVD) was highly conformal O_3/TEOS oxide. The vias were then filled with W using Metal Organic Chemical Vapor Deposition (MOCVD). See [26] for details.



(a) (b)

Figure 1.11 TSV technology proposed by Wieland *et al*. [25]. (a) Test wafer with W-filled Inter Chip Vias (ICVs) and interconnect metallization. (b) FIB image of a metal gate NMOS transistor (w/l = 20/2) with W-filled ICV near active area.

As shown above, the fabrication of many TSV sizes have been demonstrated so far. Keeping in mind that the semiconductor industry is constrained by the cost, the main focus now is improving each process step in order to achieve higher throughput. Hereafter are some technological trends to achieve this goal.

DRIE processes providing higher etch rate, smaller scallop, smoother via sidewall, higher selectivity of silicon to mask, and higher aspect ratio have been demonstrated [27, 28].

MOCVD and Atomic Layer Deposition (ALD) are relatively high-cost processes. Less expensive physical vapor deposition (PVD) methods have been demonstrated with good results on high AR structures. Weichart *et al*. presented a High Power Impulse Magnetron Sputtering (HPIMS) technique which achieved continuous layers of Ti and Cu inside TSVs with an AR of 10 [29]. Ionized Metal Plasma (IMP) Sputtering has been demonstrated to achieve conformal seed layer on the via sidewalls [30]. IMP sputtering is a magnetron sputtering system coupled with an inductively coupled plasma (ICP; RF 13.56 MHz) which ionizes gas atoms and a substrate bias (RF 27 MHz) that accelerates the ionized atoms towards the substrate. RF copper deposition with re-sputter technology is also able to improve copper step coverage, as demonstrated by Wu *et al.* [31]. TaN is deposited prior to sputter etch. So, after sputter etch, Ta sputtering will achieve a better coverage on the sidewall. This helps to improve the copper RF sputtering coverage.

Furthermore, completely filling the vias without void is timely. Less timing-consuming electroplating is being worked at to lower the cost. Some solutions have been proposed [32, 33, 34].

Electrografting has been demonstrated by Alchimer® to provide extremely uniform film coating stacks (isolation, barrier and fill layers) on TSVs with aspect ratios of 10:1 or 20:1 [35]. This fluid-based process eliminates voids, opens, and shorts, enables higher yields while also saves cost substantially.

MEOL TSV or Via-Middle Approach

BEOL TSV is a matter of supply chain issue. Processing TSVs on a chip fabricated somewhere else is not welcome since every house and foundry wants to make responsibility clear. So an intermediate approach known as "middle end of line (MEOL) TSV process" is the actual preferred solution. This approach leverages existing IC foundry infrastructure to fabricate TSVs after the FEOL processing and prior to BEOL processing. So after the chips are achieved, TSV structures are already fabricated inside. Later the wafers are to be thinned to expose TSVs from the wafer backside. After dicing, known good dies (KGDs) are to be stacked to the 3D stacks.

One advantage of MEOL TSV process over the via-last approach is the larger process temperature window. Here, the temperature can be as high as the BEOL metal process (450 °C). This is a foundry-level solution as well.

In 2011, Redolfi *et al*. presented a 300 mm industry-compliant via-middle TSV module, integrated with an advanced high k/metal gate CMOS process platform [36]. TSVs were fabricated following the Bosch process after contact definition and before the first metal layer. The target for the copper diameter was 5µm and the via depth in the silicon substrate was 50 µm. Dense structures had a pitch of 10µm. The vias were filled with TEOS/O3 oxide to reduce via-to-substrate capacitance and leakage and a Ta layer was used as the Cu-diffusion barrier. The via was finally filled with electroplated copper. Copper was thermally treated before CMP to minimize copper pumping effects (Fig. 2.21). The processing (Fig. 1.12 and Fig. 1.13) was integrated as part of a 65 nm node CMOS fabrication module and validated with regular monitoring of physical parameters. The module was tested and integrated to a thinning and backside passivation flow. However, the electrical characteristics of these TSVs were not presented.

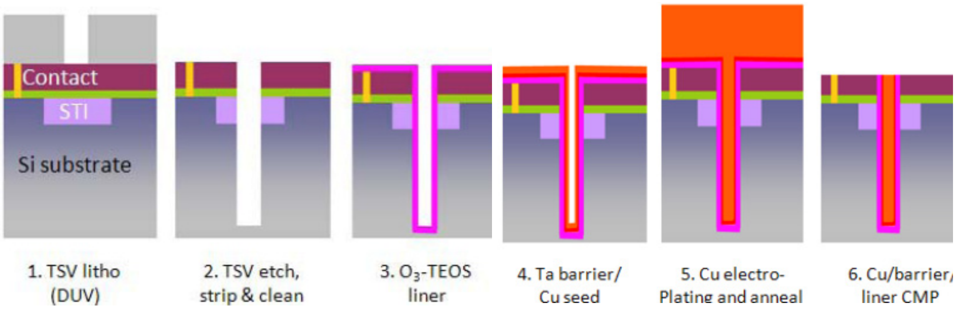

Contact
STI
Si substrate
1. TSV litho (DUV)
2. TSV etch, strip & clean
3. O_3-TEOS liner
4. Ta barrier/ Cu seed
5. Cu electro-Plating and anneal
6. Cu/barrier/ liner CMP

Figure 1.12 Major process steps of the MEOL TSV technology proposed by Redolfi *et al*.[36].

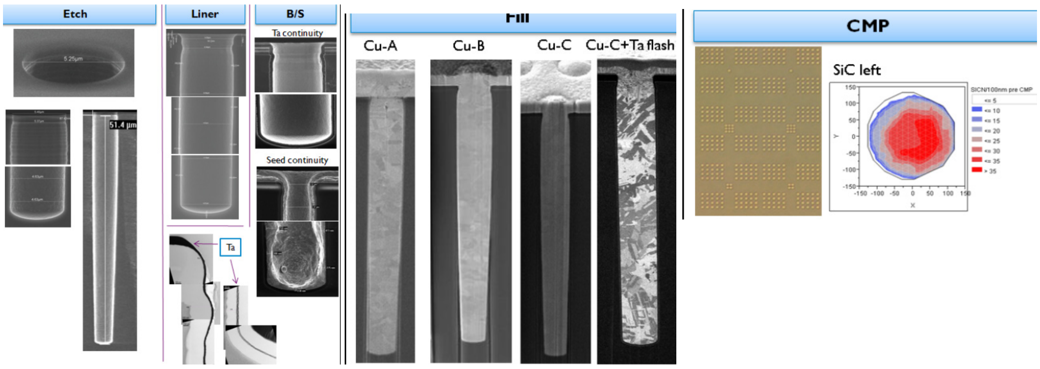

Etch
Liner
B/S
Ta continuity
Seed continuity
Ta
51.4 µm
Fill
Cu-A
Cu-B
Cu-C
Cu-C+Ta flash
CMP
SiC left

Figure 1.13 Results of the major process steps in the MEOL TSV technology proposed by Redolfi *et al*. [36].

Bringing 3-D devices based on through-silicon via (TSV) is taking longer than expected. It is only in this year (2011) that a breakthrough has been achieved by Samsung. A 1-Gbit DRAM with a 512-pin wide I/O interface intended for mobile applications such as smartphones and tablet computers was announced [37]. The chip was implemented in a manufacturing process technology somewhere between 50 nm and 59 nm. It was based on a MEOL TSV technology and housed in a 3-D package. Shipments are targeted for 2013. The fabricated TSVs (Fig. 1.14) were 7.5 µm in diameter, and exhibited a resistance of 0.22 - 0.24 Ω and a capacitance of 47.4 fF.

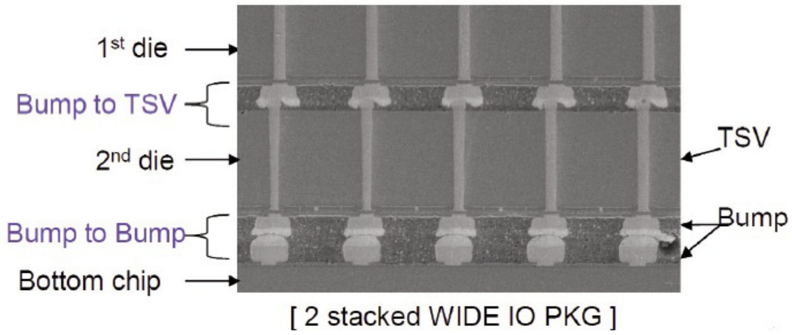


Figure 1.14 TSV technology proposed by Samsung [37].

BEOL TSV or Via-Last TSV Processed after Bonding

Another option is to fabricate TSVs after chip-to-wafer (C-to-W) or wafer-to-wafer (W-to-W) bonding. The detailed introduction to bonding techniques can be found in the next section.

In this approach, TSVs are etched through the silicon substrate and the bonding interface landing at the expected metal pad made during BEOL metal process. The thermal budget depends on the bonding and BEOL metal process, so it can be the lowest among the four solutions discussed so far. Lincoln lab SOI TSV process can be viewed as such a process (Fig. 1.15). Basically, a SOI wafer with integrated CMOS devices is face-to-face bonded to the bottom carrier wafer. The buried SiO_2 layer in the upper SOI wafer is then etched away, leaving the thin upper part of the SOI wafer. TSVs are then etched through the thin upper wafer and the dielectric layer of the bottom wafer, finally landing on targeted metal pads [38].

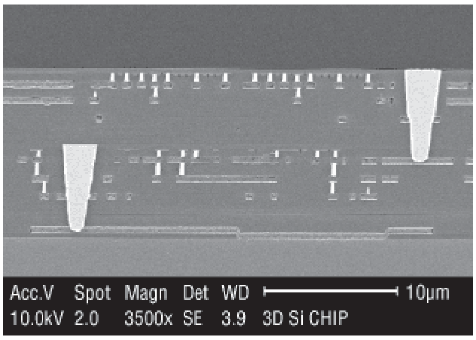

Acc.V Spot Magn Det WD 10μm
10.0kV 2.0 3500x SE 3.9 3D Si CHIP

Figure 1.15 TSV technology proposed by Lincoln Lab [38].

Summary of TSV Technologies

As explained in this section, different technologies have been developed and tested to integrate TSVs into the standard CMOS process flow. The preferred via size is now less than 30 μm; eventually, it will go down to 5 μm for memory-on-logic stacks and ≤ 2 μm for heterogeneous 3D integration (Fig. 1.1). The drives to get via size as small as possible from a technology point of view are: 1) less thermal expansion mismatch between Si (2.5e-6/°C) and Cu (17.5e-6/°C); 2) lower cost using less Cu plating time to fill the vias; 3) higher throughput; 4) more space for routing. From an electrical performance point of view, smaller vias are with less parasitic capacitance meaning potential shorter delay and less power consumption.

1.2.2 Thinning

Thinning or wafer thinning is one of the enabling technologies for 3D integration at the point of enabling TSV shrinking. Generally, the silicon wafer to be thinned is temporarily glued to a carrier wafer and then thinned from the backside to less than 50 μm. A summary of thinned silicon wafer thickness for 200 mm or 300 mm wafers is shown in Table 1.1. For general products, the minimum thickness of thinned wafer will maintain at 30 µm. For ultra-thin package, the minimum thickness of thinned wafer achieved was 20 µm, and will eventually go down to 10 µm. Whether going further down to 5 µm is still doubtful. Anyhow, transistors require a certain bulk thickness to ensure the expected performance. And the technological difficulties in the process itself need to be solved. The loading could damage the devices.

The thinning process usually starts from lapping or grinding the wafer down to a certain thickness which is thicker than the target. Then, polishing is used to refine the rough surface (Fig. 1.16) and to remove the stress concentrated at the cracks caused by the rough grinding. Other techniques to further relieve this stress can be processed after the polishing. For example, CMP and DRIE. As reported by Chen *et al.* [39], the chip strength can be enhanced after DRIE thinning.

A critical issue for successful thinning is temporary gluing/bonding of the silicon wafer with the carrier or handling wafer. Without a good solution, delamination (Fig. 1.17-a), bubble (Fig.1.17-b) and wafer chipping (Fig.1.17-a) could happen. Thickness uniformity is important as well (Fig. 1.17-c). To settle these problems, many techniques have been proposed [42, 43] and many materials have been tested [42, 43, 45, 46].

The time to perform thinning is related to the TSV fabrication approach adopted (Section 1.2.1). Finally, a wafer thinned to less than 50 µm is of good flexibility (Fig. 1.18).

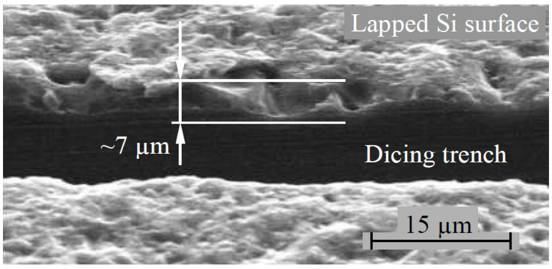


Figure 1.16 Surface after lapping and defect depth of a Si sample lapped with 3 µm Al_2O_3 abrasive particles. Pits as deep as 7 µm have been found in SEM micrographs [44].

Table 1.1 Thinned Silicon Wafer Thickness 200 mm / 300 mm [40, 41].

Year of Production	Min. Thickness (microns) For General Products, Such As Image Sensors	Min. Thickness (microns) For Ultra-Thin Packages, Such As Smart Card
2007	50	20
2008	50	20
2009	50	15
2010	50	15
2011	40	10
2012	40	10
2013	30	10
2014	30	10
2015	30	5
2016	20	5

🟧 Manufacturable solutions are known.
🟥 Interim solutions are known.

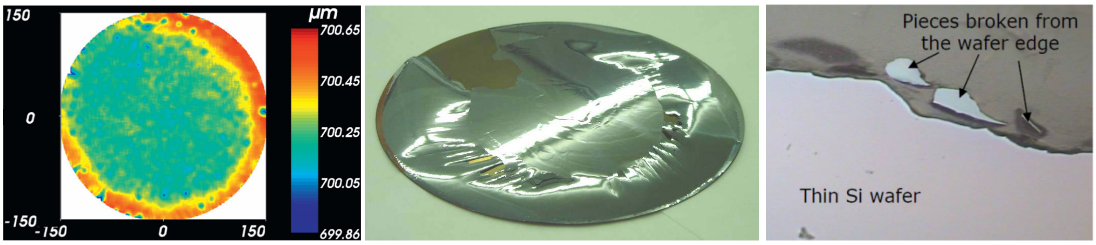


Figure 1.17 Main problems caused by thinning. (a) Thickness nonuniformity and warpage. (b) Thin wafer delamination and breakage. (c) Chipping at the wafer edge.

Figure 1.18 50 μm-thick 200 mm silicon wafer showing flexibility [42].

1.2.3 Bonding

As another key enabling technology for 3D integration, bonding plays an important role to achieve the final stack.

Many bonding techniques have been studied for 3D integration, among which we will highlight direct oxide bonding [47], polymer bonding [48], Cu-Cu thermal compression bonding [49], and Cu/Sn/Cu eutectic bonding [50, 51]. The first three techniques are wafer-level processes, while the last one mainly works at chip level and depends on flip chip technology.

For wafer-level bonding processes, the electrical connections between the stack are usually achieved through metal or oxide fusion under very high pressure and usually high temperatures (> 350 °C). The surface cleanliness and roughness are the key factors deciding the bonding quality. Surface pretreatment to remove oxidation is very complicated and the surface roughness requires < 1.0 nm Root-Mean-Square (RMS). The great advantage is that they can achieve very small TSV dimensions (< 5 μm) and very high TSV densities (pitch < 10 μm). From the research point of view, direct oxide bonding and polymer bonding should work with the BEOL TSV after bonding (Section 1.2.1, Part 4), while Cu-Cu thermal compression bonding should be proceeded after TSVs are fabricated (Section 1.2.1, Part 2 and 3). The cost of wafer-bonding processes is still very high, so it is not widely adopted as the chip-level bonding process nowadays.

In the chip-level eutectic bonding processes, the connections are achieved by thermal compression on the eutectic solder bumps or balls. Thus, the bonding temperatures required are usually lower than the wafer-level processes, and the requirements on surface pretreatment are moderate. The cost is also low. But the bump size and density are limited. The dimension varies from minimum of about 8 μm to several tens of micrometers.

1.2.4 3D Integration Schemes

Bonding technologies lead to a separation of 3D integration schemes into two main streams, which are named by Lau as "3D silicon integration" and "3D chip integration" [52]. Along this thesis, our work will cover many aspects of the 3D chip integration technology.

For 3D silicon integration, stacking is processed at the wafer level. Because the wafer-level bonding processes, e.g. direct oxide bonding, polymer bonding and Cu-Cu thermal compression bonding, are directly connecting two pads, no bump is required and so the footprint of inter-layer connections (TSV plus bonding/landing pad) can be decreased. Thus, tiny TSV dimensions as small as a few microns are achievable in this scheme. As mentioned in Section 1.2.3, wafer-level bonding technologies are costly and technologically challenging. Not only so, it also lacks of EDA design tools and is waiting for the whole product chain support.

For 3D chip integration, TSVs are fabricated on the wafers, mostly using the via-middle approach. Then, the chips with TSVs and bonding bumps are diced and bonded to the bottom chip stack or wafer using flip chip technology. As the dimension of the bumps is limited by the technology, the footprint of the inter-layer connection cannot be as small as that in 3D silicon integration. Today, 30 μm-diameter TSV is used for 3D chip integration; eventually, it is expected to go down to 5 - 10 μm.

The concept of interposer, no matter it is a 2.5 interposer or 3D interposer, is popular in 3D chip integration. The difference between 2.5D and 3D is just whether dies are stacked on one side or both sides to the interposer. ("2.5D" means that TSVs might not be needed if one flip chip bonding is performed for the stacking). Interposers are also categorized in passive and active ones, depending on whether or not active devices are fabricated. If not, the TSV process will benefit from a wider temperature process window.

Some successful interposers are already on the market, such as Xilinx [53, 54, 55] and Toshiba [3], and more are on the way. Actually many vendors are picturing 3D products with interposers in the near future. Section 1.3 will talk about it more.

The wide adaptation of interposers is motivated by the inadequacy of EDA tools, testing schemes and industrial standardization. First, the commercial EDA tools do not support 3D design and simulation with TSVs. Then, splitting the inter-layer connection components from the circuit is the safest choice: signal integrity (SI) on the chip is ensured by circuit designers, while the TSVs in the interposer are analyzed separately to ensure the SI. Moreover, chips fabricated with standard technologies ask for a signal redistribution layer (RDL) to match the I/Os designed according to different industrial standards. Moreover, fabricating the RDL and TSVs in the interposers instead of in the chips can minimize the risk of chip damages and increase the yield. The benefit from the increased yield outweighs the additional costs brought by the additional wafer as interposer and the interposer fabrication.

To summarize this section, we have completed an overview of the existing 3D fabrication technologies. In this perspective, our work will cover the key technologies, i.e. TSV, bonding and thinning. But it should be noted that there are still many aspects not mentioned, e.g. metrology and testing. In the next section, we will mainly look at the existing 3D applications.

1.3 3D Applications

Actually all the commercial 3D products are based on interposers, either glass or silicon. Until 2010, in the 3D interposer market [56], logic stacking (logic on logic and logic on memory) solutions are still lacking because small TSV dimensions (5 μm in diameter) and large TSVs density are required.

Xilinx presented their 3D FPGA work in 2010 as the first 3D stacking using an advanced logic technology (28 nm Virtex-7 LX2000T multi die FPGA) (Fig. 1.19) [53, 54, 55]. In this approach, four dies were communicating through the metal stacks fabricated in the silicon interposer and they were at the same time connected to the package substrate through TSVs. However, it is still not stacking logic dies vertically on each other.

1.4 3D Circuit Design and System Architecture

In the past 10 years, many aspects concerning 3D circuit design techniques have been studied, e.g. placement and partitioning, EDA, testing schemes, thermal and power analysis and so on [57, 58, 59, 60, 61, 62]. System architects mainly optimized computing systems by focusing on memory and/or logic stacking. Many interesting works have been presented but only a few are provided here [63, 64, 65, 66, 67, 68].

In these studies, great benefits have been demonstrated after using TSVs. One of the popular topics these days is the wide I/O (bandwidth) interface gained with TSVs. Less power consumption with greatly increased bandwidth has been demonstrated in the recent Samsung memory stacking [37].

A close and plain examination of TSV as a connection itself could be helpful to achieve further benefits. Firstly, TSVs can work at very high frequencies over 60 GHz [69]. In 3D digital systems, this inborn high bandwidth has not been exploited and is wasted on the slow data transmission rate. Secondly, many works were dedicated to improve the performance through novel inter-layer bus or routers (for Network on Chip (NoC)) designs. But for system architects, the physical dimension is usually merged in the study, so TSVs' real physical dimension is ignored in many cases and thus its influence on circuits and systems has not been explored.

In all, to address these two issues requires a well-understanding of both technology feasibility and system requirements.

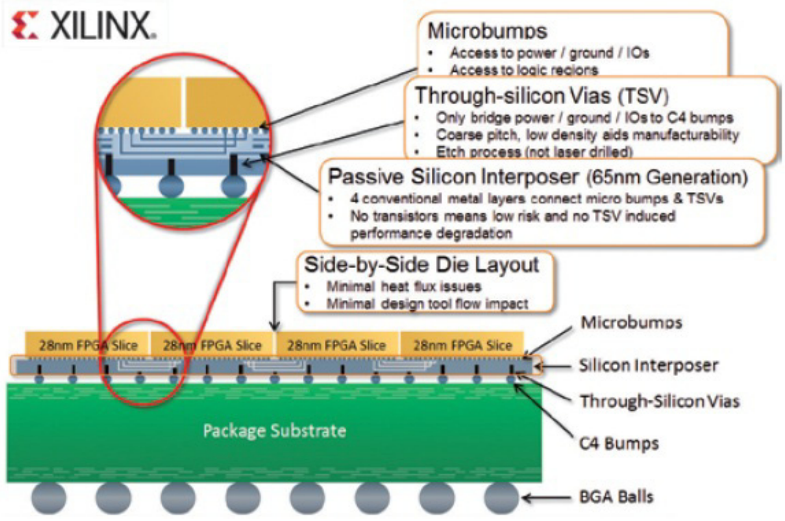

XILINX
Microbumps
• Access to power / ground / IOs
• Access to logic regions
Through-silicon Vias (TSV)
• Only bridge power / ground / IOs to C4 bumps
• Coarse pitch, low density aids manufacturability
• Etch process (not laser drilled)
Passive Silicon Interposer (65nm Generation)
• 4 conventional metal layers connect micro bumps & TSVs
• No transistors means low risk and no TSV induced performance degradation
Side-by-Side Die Layout
• Minimal heat flux issues
• Minimal design tool flow impact
28nm FPGA Slice
28nm FPGA Slice
28nm FPGA Slice
28nm FPGA Slice
Microbumps
Silicon Interposer
Through-Silicon Vias
C4 Bumps
Package Substrate
BGA Balls

Figure 1.19 Xilinx introduced 3D silicon interposers with TSVs for wide I/O interface in FPGA products [53]. Courtesy of Xilinx.

1.5 Outline of This Thesis

This thesis work started in September 2007 when controversies of 3D technologies widely existed both in academy and industry.

The general aim was to develop a die-level 3D stacking scheme at EPFL. So, the work had to cover all the aspects of the key 3D technologies in order to provide a well-balanced and workable solution. The existing works mainly adopted BEOL TSV approach at the wafer level. However, in the context of a university project, our resource was limited at diced standard chips fabricated in the foundry. To post-processing TSVs on these singulated dies was of a greater challenge. Firstly, the way to handle and align the dies was a big problem. Secondly, we were facing many technology constraints, such as lacking of necessary instruments for 3D integration, e.g. a wafer grinding and polishing machine for wafer thinning and the die bonding machine.

Our work to tackle these technology problems will be presented in **Chapters 2, 3 and 4**.

In **Chapter 2**, a thorough study of the TSV fabrication feasibility based on the clean room facilities in Center of MicroNanotechnology, EPFL (hereafter referred to as CMI) will be performed. All the four TSV processes replying on wafer-level processes will be detailed in that chapter.

In **Chapter 3**, the feasibility of precise chip attachment to a carrier wafer with the available materials and technologies will be studied first. This is the prerequisite of the die-level TSV post-processing. Seeing the limitations, self-alignment (SA) techniques were studied for a better alignment and a workable process. Four SA techniques, namely "gravitational SA", "hydrophobic SA", "electrostatic SA" and "magnetic SA", will be well studied, while the most successfully one, the hydrophobic SA will receive special attention in **Chapter 4**. SA techniques will be demonstrated as good solutions to achieve the final chip stacking as well as temporary die attachment and alignment techniques to post-process TSVs on dies.

Our work in the first four chapters will well-establish a 3D chip integration platform based on die-level post-processing. These studies will lead to a more realistic and practical study in the following chapters.

In **Chapter 5**, we will continue with a discussion of TSV characteristics, from silicon area cost to electrical performance and power consumption. These will clarify the key reasons of going for 3D. As TSVs' area is actually not negligible, and TSVs support very high bandwidths, exploiting this high bandwidth through inter-layer serialization could be a good method to save the precious silicon area consumed by the TSVs. We will discuss about the circuit implementation of the inter-layer serial links and provide a theoretical analysis of the future of this technique with smaller TSVs.

In **Chapter 6**, the serial links will be applied into 3D memory-on-core system simulators to evaluate the system performance. The TSV dimensions assumed in the simulation are practical numbers from the 3D chip integration technology.

In **Chapter 7**, we will conclude the thesis with an outlook on future developments.

Reference

1. W. Shockley, "Semiconductive Wafer and Method of Making the Same, " United States Patent 3,044,909, July 17, 1962.
2. Y. Akasaka, "Three-Dimensional IC trends," *Proceedings of IEEE*, Vol. 74, No. 12, 1986, pp 1703.
3. Iwate Toshiba Electronics, Product name: Chip Scale Camera Module: CSCM, http://www.toshiba-iwashiba.co.jp/eng/product/cscm.htm.
4. 3D-IC & TSV Interconnects 2010 Reports, Yole Development.
5. E. M. Chow, A. Partridge, C. G. Quate, and T. W. Kenny. "Through-Wafer Electrical Interconnects Compatible with Standard Semiconductor Processing." *Processings of Solid-State Sensor and Actuator Workshop*, pages 343-346, Hilton Head Island, SC, USA, June 4-8, 2000.

6. M. Kawano, N. Takahashi, Y. Kurita, K. Soejima, M. Komuro, and S. Matsui, "Three-Dimensional Packaging Technology for Stacked DRAM With 3-Gb/s Data Transfer", *IEEE Transactions on Electron Devices*, Vol. 55, No. 7, July 2008.
7. T. Bauer, "High Density Through Wafer Via Technology," *NSTI-Nanotech 2007*, www.nsti.org, ISBN 1420061844, Vol. 3, 2007
8. H. Baltes, O. Paul, and O. Brand, "Micromachined Thermally Based CMOS Microsensors," *Proceedings of the IEEE*, Vol. 86, No. 8, pp. 1660-1678, 1998.
9. A. D. Yalçinkaya, "Micromechanical Resonators for Low-Power, Low-Voltage Systems," PhD thesis, Mikroelektronik Centret (MIC), Technical University of Denmark, June, 2003.
10. L. L. Lauer, J. L. Shohet, and R. W. Hansen, "Measuring Vacuum Ultraviolet Radiation-Induced Damage," *Journal of Vacuum Science Technology A*, **21 (4):** 1253-1259, July/August 2003.
11. S. Linder, H. Baltes, F. Gnaedinger, and E. Doering, Fabrication Technology for Wafer Through-Hole Interconnections and Three-Dimensional Stacks of Chips and Wafers, pp. 349-354, 1994.
12. S. Linder, "Chip Stacks for Memory Applications," PhD Thesis, Physical Electronics Laboratory, ETH Zürich, 1996.
13. M. Heschel, J. F. Kuhmann, S. Bouwstra, and M. Amskov, "Stacking Technology for a Space Constrained Microsystem," *Proceedings of the 11th IEEE International Conference on Micro Electro Mechanical Systems (MEMS ’ 98)*, pages 312–317, Heidelberg, Germany, January 26–30, 1998.
14. M. Heschel, "Multiple Through-Wafer Interconnects for Stacking of Microelectromechanical Devices," PhD thesis, Mikroelektronik Centret (MIC), Technical University of Denmark, June, 1999.
15. M. Heschel and S. Bouwstra," Conformal Coating by Photoresist of Sharp Corners of Anisotropically Etched Through-Holes in Silicon," *Sensors and Actuators A: Physical*, Vol. 70, 1998.
16. F. E. Rasmussen, "Electrical Interconnections Through CMOS Wafers," PhD Thesis, the Technical University of Denmark, 2003.
17. S. W. Ho, V. S. Rao, O. Kalandar, N. Khan, S. U. Yoon, V. Kripesh, " Development of Coaxial Shield Via in Silicon Carrier for High Frequency Application," *EPTC*, 2007.
18. Sonion MEMS A/S, Product Description: 3 mm^3 Smart Silicon Microphone. http://www.soionmems.dk, 2002.
19. K. Takahashi and M. Sekiguchi, " Through Silicon Via and 3-D Wafer/Chip Stacking Technology," *Symposium on VLSI Circuits Digest of Technical Papers*, 2006.
20. M. Sekiguchi, H. Numata, N. Sato, T. Shirakawa, M. Matsuo, H. Yoshikawa, M. Yanagida, H. Nakayoshi and K. Takahashi, "Novel Low Cost Integration of Through Chip Interconnection and Application to CMOS Image Sensor," *56th Electronic Components and Technology Conference*, 2006
21. R. Nagarajan, L. Ebin, L. Dayong, S. C. Seng, K. Prasad and N. Balasubramanian, "Development of a Novel Deep Silicon Tapered Via Etch Process for Through-Silicon Interconnection in 3-D Integrated Systems," *56th Electronic Components and Technology Conference*, pp. 383 - 387, 2006.
22. D. S. Tezcan, N. Pham, B. Majeed, P. De Moor, W. Ruythooren and K. Baert, "Sloped Through Wafer Vias for 3D Wafer-Level Packaging," *57th Electronic Components and Technology Conference*, pp. 643 - 647, 2007.

23. P. Benkart, A. Munding, A. Kaiser, E. Kohn, A. Heittmann, H. Huebner, U. Ramacher, "Three-Dimensional Integration Scheme with a Thermal Budget Below 300 °C," *Sensors and Actuators A*, Vol. 139, pp. 350-355, 2007.
24. Y. Ohara, Y. Watanabe, K. Lee, T. Fukushima, M. Koyanagi, T. Tanaka, "Development of 5 μm Diameter Backside Cu TSV Technology for 3D LSI," *The International Conference on Electronics Packaging (ICEP)*, 2011.
25. R. Wieland , D. Bonfert, A. Klumpp, R. Merkel, L. Nebrich, J. Weber, P. Ramm, "3D Integration of CMOS Transistors with ICV-SLID Technology," *Microelectronic Engineering*, Vol. 82, pp. 529–533, 2005.
26. A. Klumpp, P. Ramm and R. Wieland, "3D-Integration of Silicon Devices: A Key Technology for Sophisticated Products," *Proceeding of Design, Automation and Test in Europe*, pp.1678 - 1683, 2010.
27. W. H. Teh, R. Caramto, T. Chidambaram, W. Wang, S. R. Arkalgud, T. Saito, K. Maruyama, and K. Maekawa,"300-mm Production-Worthy Magnetically Enhanced Non-Bosch Through-Si-Via Etch for 3-D Logic Integration," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 23, No. 2, May 2010.
28. D. Haynes, "Advances in the Etching of TSVs for IC Packaging Applications," *SEMICON*, 2008.
29. J. Weichart, M. Elghazzali, S. Kadlec, and A. P. Ehiasarian, "PVD Processes in High Aspect Ratio Features by HIPIMS," *Proceedings of the 52nd Annual Technical Conference of Society of Vacuum Coaters*, Albuquerque, NM 87122-1914, United States, May 9-14, 2009 Santa Clara, (CA), 201 (2009).
30. B.-H. Cho, J.-J. Yun, and W.-J. Lee, "Filling of Very Fine Via Holes for Three-Dimensional Packaging by Using Ionized Metal Plasma Sputtering and Electroplating," *Japanese Journal of Applied Physics*, Vol.46, pp. 1135-1137, 2007.
31. H.-J. Wu, R. Shaviv, M. Sriram, W. Wu, A. Pradhan, K. J. Park, J. O'loughlin, K. Chattopadhyay, T. Mountsier and G. Dixit, "Improving Interconnect Reliability via Optimized Barrier/Seed," *Semiconductor International*, Vol. 31, No. 11, Oct. 1, 2008.
32. H. Y. Li, E. Liao, X. F. Pang, H. Yu, X. X. Yu, J. Y. Sun, "Fast Electroplating TSV Process Development for the Via-Last Approach," *60th Electronic Components and Technology Conference,* 2010.
33. A. Keigler, Z. Liu, J. Chiu, J. Drexler, "Sematech 3D Equipment Challenges: 300 mm Copper Plating," Sematech Workshop on Equipment challenges for 3D Interconnect, San Francisco, CA, July 2008.
34. I. Teelinck, "Electrochemical Copper Deposition in IC Manufacturing," PhD thesis, University of Gent, 2002.
35. Electrografting, http://www.alchimer.com/electrografting#.
36. A. Redolfi, D. Velenis, S. Thangaraju, P. Nolmans, P. Jaenen, M. Kostermans, U. Baier, E. Van Besien, H. Dekkers, T. Witters, N. Jourdan, A. Van Ammel, K. Vandersmissen, S. Rodet, H.G.G. Philipsen, A. Radisic, N. Heylen, Y. Travaly, B. Swinnen, E. Beyne, "Implementation of an Industry Compliant, 5×50 μm, Via-Middle TSV Technology on 300 mm Wafers," *61th Electronic Components and Technology Conference*, pp. 1384 - 1388, 2011.
37. J.-S. Kim, C. S. Oh, H. Lee, D. Lee, H.-R. Hwang, S. Hwang, B. Na, J. Moon, J.-G. Kim, H. Park, J.-W. Ryu, K. Park, S.-K. Kang, S.-Y. Kim, H. Kim, J.-M. Bang, H. Cho, M. Jang, C. Han, J.-B. Lee, K. Kyung, J.-S. Choi, Y.-H. Jun, "A 1.2V 12.8GB/s 2Gb Mobile Wide-I/O DRAM with 4×128 I/Os Using TSV-Based Stacking," *ISSCC*, 2011.

38. J. A. Burns, B. F. Aull, C. K. Chen, C.-L. Chen, C. L. Keast, J. M. Knecht, V. Suntharalingam, K. Warner, P. W. Wyatt, and D.-R. W. Yost, "A Wafer-Scale 3-D Circuit Integration Technology," *IEEE Transactions on Electron Devices*, Vol. 53, No. 10, October 2006.
39. S. Chen, T.-Y. Kuo, H.-T. Hu, J.-R. Lin, and S.-P. Yu, "The Evaluation of Wafer Thinning and Singulating Processes to Enhance Chip Strength," *55th Electronic Components and Technology Conference*, 2005.
40. International Technology Roadmap for Semiconductors, 2009.
41. International Technology Roadmap for Semiconductors, 20011.
42. K. De Munck, "Generic Building Blocks for 3D Integration and Their Application on Hybrid CMOS Image Sensors," PhD thesis, Katholieke Universiteit Leuven, 2008.
43. L. W. Schaper, S. L. Burkett, S. Spiesshoefer, G. V. Vangara, Z. Rahman, S. Polamreddy, "Architectural Implications and Process Development of 3-D VLSI Z-Axis Interconnects Using Through Silicon Vias," *IEEE Transactions on Advanced Packaging*, Vol. 28, Issue 3, August 2005.
44. A. Munding, "Interconnect Technology for Three-Dimensional Chip Integration," phd thesis, University of Ulm, 2007.
45. Quickstick 135, http://www.emsdiasum.com/microscopy/products/materials/adhesives.aspx.
46. Thermal Release Tape REVALPHA, http://www.nitto.com/product/datasheet/e_parts/010/.
47. Ziptronix ZiBond™ Direct Wafer Bonding Technology, http://www.ziptronix.com/techno/zibond.html.
48. R. J. Gutmann, J.-Q. Lu, Y. Kwon, J. F. McDonald, and T. S. Cale, "Three-Dimensional (3D) ICs: A Technology Platform for Integrated Systems and Opportunities for New Polymeric Adhesives," *First International IEEE Conference on Polymers and Adhesives in Microelectronics and Photonics*, pp. 173 - 180, Potsdam, Germany, October 2001.
49. B. Swinnen, W. Ruythooren, P. De Moor, L. Bogaerts, L. Carbonell, K. De Munck, B. Eyckens, S. Stoukatch, D. S. Tezcan, Z. Tőkei, J. Vaes, J. Van Aelst, E. Beyne, "3D Integration by Cu-Cu Thermo-Compression Bonding of Extremely Thinned Bulk-Si Die Containing 10 μm Pitch Through-Si Vias," *International Electron Devices Meeting*, 2006.
50. M. Baum, C. Jia, M. Haubold, M. Wiemer, "Eutectic Wafer Bonding for 3-D Integration," *3rd Electronic System-Integration Technology Conference (ESTC),* 2010.
51. C.-J. Zhan, J.-Y. Juang, Y.-M. Lin, Y.-W. Huang, K.-S. Kao, T.-F. Yang, S.-T. Lu, J. H. Lau, T.-H. Chen, R. Lo, and M. J. Kao, "Development of Fluxless Chip-on-Wafer Bonding Process for 3DIC Chip Stacking with 30 μm Pitch Lead-Free Solder Micro Bumps and Reliability Characterization," *61st Electronic Components and Technology Conference*, 2011.
52. J. H. Lau, "Evolution and Outlook of TSV and 3D IC/Si Integration," *12th Electronics Packaging Technology Conference*, pp. 560-570, 2010.
53. B. Banijamali, S. Ramalingam, N. Kim, and C. Wyland, "Ceramics vs. Low-CTE Organic Packaging of TSV Silicon Interposers," *61st Electronic Components and Technology Conference*, 2011.
54. B. Banijamali, S. Ramalingam, K. Nagarajan, and R. Chaware, "Advanced Reliability Study of TSV Interposers and Interconnects for the 28nm Technology FPGA," *61st Electronic Components and Technology Conference*, 2011.

55. N. Kim, D. Wu, D.Kim, A. Rahman, and P. Wu, "Interposer Design Optimization for High Frequency Signal Transmission in Passive and Active Interposer using Through Silicon Via (TSV)," *61st Electronic Components and Technology Conference*, 2011.
56. Next Generation Package Substrates: 3D Silicon and Glass Interposers - Technologies, applications and market, Report, Yole Development, August 2010.
57. M. Takamiya, K. Onizuka, and T. Sakurai, "3D-Structured On-Chip Buck Converter for Distributed Power Supply System in SiPs," *IEEE International Conference on Integrated Circuit Design and Technology and Tutorial*, pp. 33-36, 2008.
58. R. Lipton, "3D-Vertical Integration of Sensors and Electronics," *Nuclear Instruments and Methods in Physics Research A*, Vol. 579, pp. 690-694, 2007.
59. J. Deguchi, T. Watanabe, T. Nakamura, Y. Nakagawa, T. Fukushima, J.-C. Shim, H. Kurino, T. Abe, M. Tamai, M. Koyanagi, "Three-Dimensionally Stacked Analog Retinal Prosthesis Chip," *Japanese Journal of Applied Physics*, Volume: 43, Issue: No. 4B, Publisher: Japan Soc. Appl. Phys, Pages: 1685-1689, 2004.
60. C. Ababei, Y. Feng, B. Goplen, H. Mogal, T. Zhang, K. Bazargan, and S. Sapatnekar, "Placement and Routing in 3D Integrated Circuits," *IEEE Design & Test of Computers*, pp. 520-531, 2005.
61. S. M. Alam, D. E. Troxel, C. V. Thompson, "Layout-Specific Circuit Evaluation in 3-D Integrated Circuits," *Analog Integrated Circuit and Signal Processing*, Vol. 35, Issue 2-3, 2003.
62. S. Sridharan, M. DeBole, G. Sun, Y. Xie, V. Narayanan, "A Criticality-Driven Microarchitectural Three Dimensional (3D) Floorplanner," *Asia and South Pacific Design Automation Conference*, 2009.
63. X. Wu, Y. Chen, K. Chakrabarty, Y. Xie, "Test-Access Mechanism Optimization for Core-Based Three-Dimensional SoCs," *IEEE International Conference on Computer Design*, 2008.
64. P. Garrou, C. Bower, P. Ramm*, Handbook of 3D Integration: Technology and Applications of 3D Integrated Circuits*, 2008.
65. J. Kim, C. Nicopoulos, D. Park, R. Das, Y. Xie, V. Narayanan, M. S. Yousif, C. R. Das, "A novel dimensionally-decomposed router for on-chip communication in 3D architectures," *Proceedings of the 34th annual international symposium on Computer architecture*, 2007.
66. A. K. Misha, X. Dong, G. Sun, Y. Xie, N. Vijaykrishnan, C. R. Das, "Architecturing On-Chip Interconnects for Stacked 3D STT-RAM Caches in CMPs," *Proceeding of the 38th annual international symposium on Computer architecture*, June 4-8, 2011.
67. H. Sun, J. Liu, R. Anigundi, N. Zheng, J.-Q. Lu, K. Rose and T. Zhang, "Design of 3D DRAM and Its Application in 3D Integrated Multi-Core Computing Systems," *Design & Test of Computers*, Issue 99, 2009.
68. D. H. Woo, N. H. Seong, D. L. Lewis, H.-H. S. Lee," An Optimized 3D-Stacked Memory Architecture by Exploiting Excessive, High-Density TSV Bandwidth," Proceedings of *the 16th International Symposium on High-Performance Computer Architecture*, pp.429-440, Bangalore, India, January, 2010.
69. S. Kannan, S. S. Evana, A. Gupta, B. Kim and L. Li, "3-D Copper Based TSV for 60-GHz Applications," *Electronic Components and Technology Conference*, pp. 1168-1175, 2011.

Chapter 2

TSV Fabrication Technologies

2.1 Target

As mentioned in Chapter 1 already, we aim at a die-level post-processing platform for 3D integration in which chips with post-processed (or BEOL) TSVs are face-to-back stacked into a 3D chip. Hereafter, during the process, the maximum temperature can be tolerated is 450 °C. In this chapter, we endeavor to develop the smallest TSV structure compatible with our existing clean room facilities [1].

TSVs are to be fabricated from the backside of the substrate. The main advantage is reduced risk of damaging the devices during processing. Via hole etching will land on bottom metal layers such as metal 1 (MET1) or metal 2 (MET2) layer. The expected benefit is utilizing silicon area more efficiently as the upper metal layers are still free for routing. If backside TSV is the choice, the wafer or chip needs to be thinned already before the via hole etching.

Normally, metal patterns for landing and signal distribution are to be fabricated on the front side of the wafer and covered with dielectric layer. Then, this side of the wafer is glued to a carrier wafer exposing the blank backside. Afterwards, wafer-level thinning is performed to get the targeted thickness. Finally, TSVs are processed on the back of the thinned wafer with the mechanical support from the carrier wafer.

2.2 Core-Hole TSV Process

2.2.1 Process Flow

Fig. 2.1-1 illustrates the wafers ready for TSV fabrication. As mentioned, the landing metals are already fabricated, and the wafer is already thinned to 50 μm and processed with the carrier wafer. Tapered or vertical via holes (here, tapered) are etched through the silicon substrate until the landing metal (Fig. 2.1-2). Then dielectrics are deposited on the via sidewall (Fig. 2.1-3), an d later opened at the bottom to get metal contact to the chip (Fig. 2.1-4). Then, barrier and seed layers are deposited (Fig. 2.1-5). Before electroplating, a resist mold is to be fabricated to define the areas for Cu electroplating (Fig. 2.1-6). Afterwards, chemical mechanical polishing (CMP) is used to planarize Cu top surface (Fig. 2.1-7). Finally, resist is to be stripped (Fig. 2.1-8).

2.2.2 Silicon Dry Etch

Plasma etching in the form of reactive ion etching (RIE) is widely used within IC manufacturing. RIE is suitable for shallow etching, however for features requiring larger etch depths with high aspect ratios, the etch rate, mask selectivity and anisotropy of the standard RIE process is no longer acceptable. In the mid 1990s, specialized equipment for etching of deep features in silicon was introduced. This equipment uses inductively coupled plasma (ICP) etching and is therefore often used as a synonym for deep reactive ion etching (DRIE). The principle of the ICP etch process is based on the technique invented by Lärmer and Schilp [2], known as the Bosch process. Today, DRIE equipment working on the basis of the Bosch process is made by a number of manufactures, e.g. Alcatel, Oxford, STS, and Unaxis. The tool we used for via etching experiments is an Adixen AMS200 etcher [3] that is optimized for DRIE in Silicon (Si), Silicon on Insulator (SOI) and dielectrics.

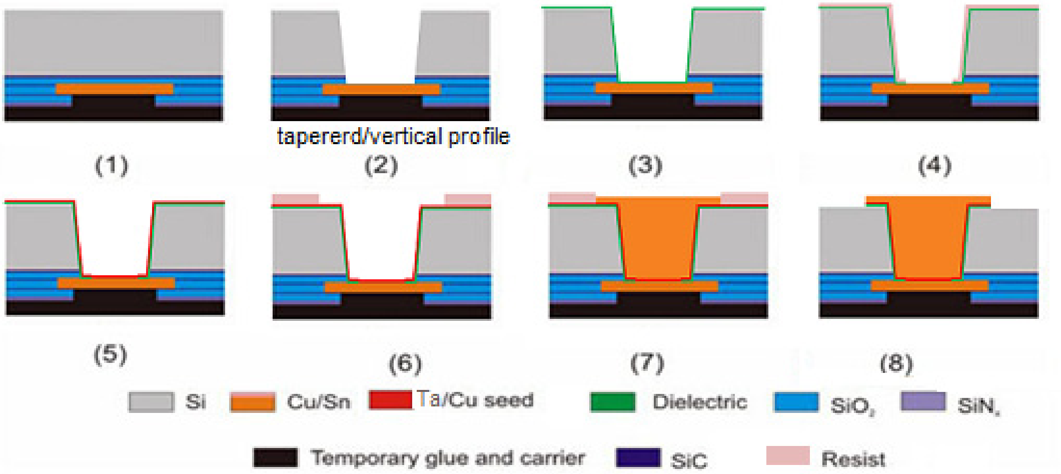


Figure 2.1 Core-hole TSV process flow.

The system is equipped with a single ceramic process chamber source for 4 inch wafers and an ICP plasma source. The maximum RF coil power is 3 KW and the frequency is 13.56 MHz. During processing, the wafer is electrostatically clamped to the chuck, and helium cooled. Another 13.56 MHz RF or low frequency pulse power signal (called platen power) is fed to the chuck to vary the bias potential of the wafer with respect to the plasma, thus controlling the energy of the incident ions. A throttle valve controls the pressure in the chamber in combination with a turbomolecular pump. In our experiments, standard 4 inch <100> p/n-doped (0.1 - 100 Ω·cm) silicon wafers were used. The patterns of vias were transferred to a 5-μm thick AZ9260 resist layer by photolithography.

In the following sections, we will describe three RIE processes, including an existing Bosch process and two tapered etching processes.

Bosch Process

In a sulfur hexaflouride (SF_6)-based plasma, SF_6 etches silicon isotropically by initially dissociating the relatively inert molecule into atomic fluorine radicals by the following reactions.

First, the radical is formed through the following formula:

$SF_6 + e^- \rightarrow S_xF_y + + S_xF_y^* + F^* + e^-$ (1)

The plasma contains some ions, which attack the wafer from a nearly vertical direction and also assist fluoride radicals to etch silicon, which results in volatile fluorides such as SiF_4 (see Eq. (2)).

$Si + F^* \rightarrow SiF_x$ (volatile gas) (2)

Next, a short polymer deposition cycle is performed to build up a Teflon-like material just after silicon etching. This passivation layer (C_xF_y) is deposited on the sidewalls and at the base of the etched feature. It occurs by both ionization and dissociation of the octafluorocyclobutane (C_4F_8) gas (Eq. 3), and this polymerization comes from C and CF_x radicals binding under ion bombardment (Eq. (4)).

$$C_4F_8 + e^- \rightarrow CF_x + CF_x^* + F^* + e^- \tag{3}$$

$$CF_x^* \rightarrow nCF_2 \text{ (fluorocarbon polymer)} \tag{4}$$

Optimal processing is obtained by a careful control of passivation and etching times, wafer temperature, ion energy and the process variables that control species fluxes.

Since the polymer dissolves very slowly only in the chemical portion of the etch (Eq. (5)), and since ions hit and attack the passivation layer at the bottom of the trench (but not along the sides), passivation tends to take place only on the sidewalls which protects them from etching. This enables an anisotropic etching feature despite the isotropic etching nature of the fluorine radical (in addition, this creates periodic nanoscale ripples called "scallops" along the sidewall).

$$nCF_2 + F^* \rightarrow CF_x^* \rightarrow CF_2 \text{ (volatile gas)} \tag{5}$$

A SEM image of microtrenches with different widths etched by the Bosch process is shown in Fig. 2.2. The well-known aspect ratio dependent etch rate (ARDE) effect can be observed. Etch rates for different TSV diameters are plotted in Fig. 2.12.

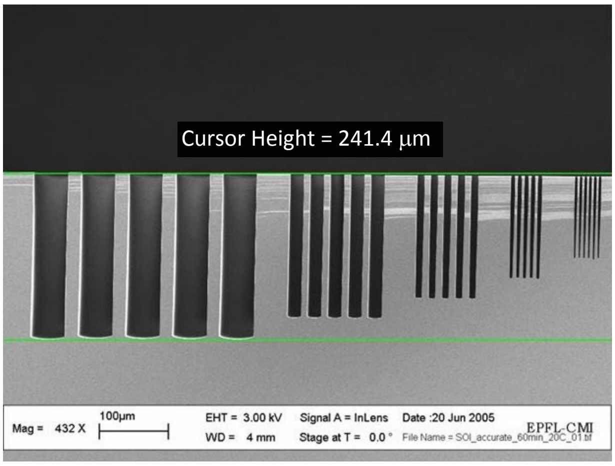


Figure 2.2 Bosch process etched microtrenches in CMI [3].

Tapered Silicon Etch Based on Modified Bosch Process

To achieve tapered profile, a possible approach is to take advantage of the parameter ramping capability of the etching system. Burkett *et al.* reported a slope angle of 80.4 degree by using a seven-module method with a continuously increasing passivation time [4]. Another method is to insert an isotropic etching step subsequent to the Bosch process [5]. The sidewall angle depends mainly on the extra isotropic etching time. Both of them require long etch duration. Actually, for Bosch process, increasing the C_4F_8 flow rate, the passivation layer formed will be thicker and a final tapered profile can be achieved.

In Fig. 2.3, the main features of a tapered TSV profile are illustrated. For a good tapered TSV etching, small undercut, constant sidewall slope, smooth surface and no local bowing are the main requirements.

In our experiment, two factors that affect the etching process were studied: C_4F_8 flow rate and substrate temperature. The setting for each parameter is shown in Table 2.1.

The SEM images in Fig. 2.4 show cross sections of vias with an additional C4F8 gas flow varying from 40 sccm to 100 sccm. Other parameters adopt the baseline values. With high C4F8 flow rate, the sidewall roughness is greatly improved, the bowing is reduced, the top surface is planar, and the undercut is minimized as well.

As shown in Fig. 2.5, when temperature decreases from 20 °C to 10 °C, the slope angle increases from 60 ° to 80 ° because of decreased etch rates under lower temperatures. One problem is that mouse bites appear on top of the vias at lower temperature (i.e. 10 °C) while this doesn't happen at a temperature of 20 °C. In silicon etching, mouse bite profiles are found when the isotropic etch component is large and when, under high temperature conditions, the sidewall polymer layer becomes locally too thin to withstand the isotropic etching. Therefore, our observation seems to contradict this theory.

To clear this confusion, the experiment was reproduced at 20 °C. Even worse results were observed: now, the mouse bites also occur on the sidewall. The reason for this inconsistent experimental result could be the chamber cleanliness and condition. Once the plasma is ignited, the temperature of the chamber walls tends to increase over time, eventually settling at a steady-state condition. Volatile by-products become released from the chamber walls as the temperature increases. These by-products react with the fluorine radicals from the SF_6 etch step, resulting in a net reduction of etchant species. The sputtering of different quantities of photoresist also contributes to the larger volume of the contaminant species within the chamber walls. Again, these species are released and act as a fluorine "sink" as the wall temperature increases. Other mask materials, such as silicon oxide, can be sputtered and could potentially compound this effect. Cleaning the chamber removes carbon byproducts deposited on the chamber walls. After cleaning the chamber, mouse bites appear maybe because the etching radicals concentration increases although the same SF_6 flow rate is used, leading to higher etch rate. In practice, for AMS 200, Bosch process (recipe: SOI_accurate) etch rate can even double after chamber cleaning.

Vias of different nominal diameters are etched using the baseline recipe for 10 minutes. SEM images of via cross sections with different nominal diameters are shown in Fig. 2.6. When the via nominal diameter decreases from 100 μm to 20 μm, the etch rate drops down from 5 μm/min to 2.71 μm/min (Fig. 2.12). The local bowing remains around 3 μm for all via sizes. The problem is this negative angle (local bowing) on top of the via that somehow will shadow via sidewall from particle deposition during sputtering. An unselective anisotropic silicon etching could be a solution to remove this negative angle.

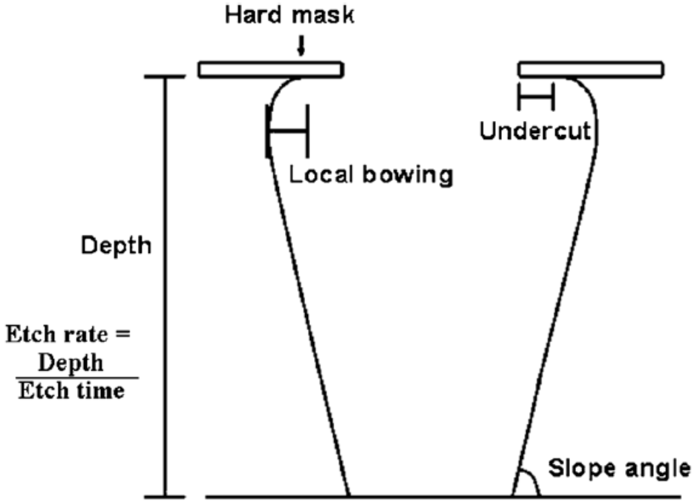
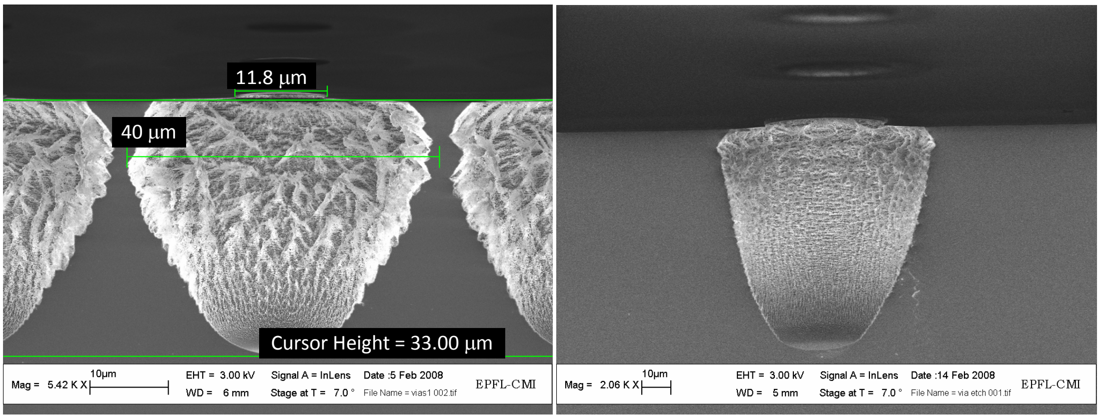


Figure 2.3 Schematic of a tapered via hole, with via parameter definition (local bowing, undercut, slope angle and depth) [6].

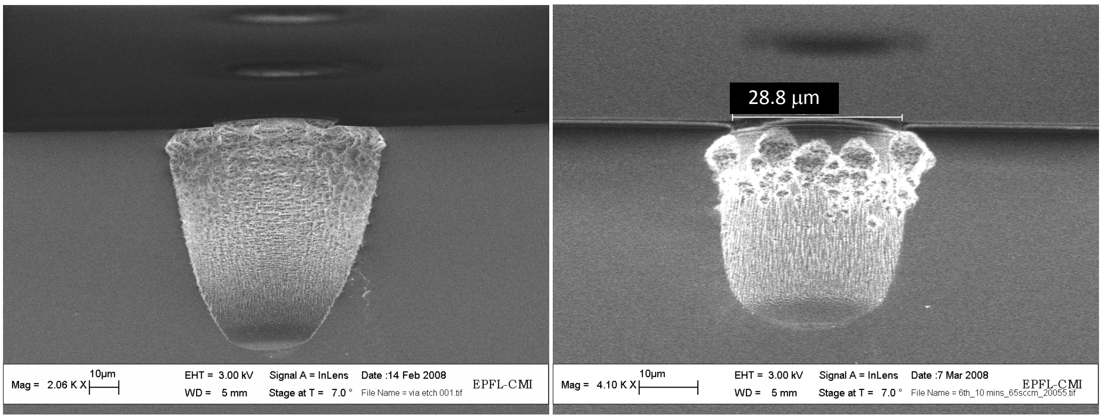
Table 2.1 Settings for Each Parameter Used for the Modified Bosch Process.

Parameters	Investigation values	Baseline	Optimized
Coil power (W)	1200	1200	1200
SF6 flow (sccm)	500	500	500
C4F8 flow (sccm)	40, 60, 80, 100	60	60
O2 flow (sccm)	200	200	200
Platen bias (W)	40	40	40
Pressure (mbar)	5e-2	5e-2	5e-2
Platen temperature (°C)	20, 10	20	0
Etching duration (min)	10	10	10


11.8 μm
40 μm
Cursor Height = 33.00 μm
Mag = 5.42 K X
10μm
EHT = 3.00 kV
WD = 6 mm
Signal A = InLens
Stage at T = 7.0 °
Date :5 Feb 2008
File Name = vias1 002.tif
EPFL-CMI
Mag = 2.06 K X
10μm
EHT = 3.00 kV
WD = 5 mm
Signal A = InLens
Stage at T = 7.0 °
Date :14 Feb 2008
File Name = via etch 001.tif
EPFL-CMI

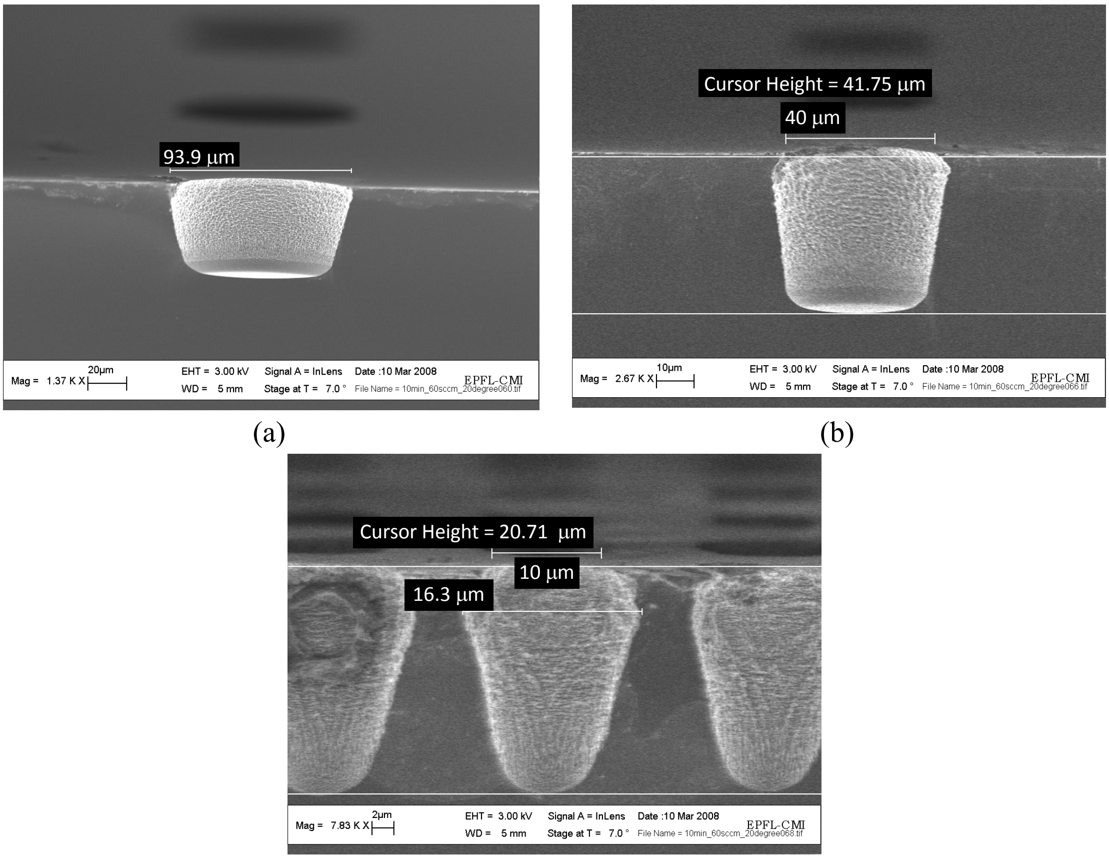
(a) (b)

Figure 2.4 SEM cross section images of tapered via holes etched with modified Bosch process when C4F8 flow rate increases from 40 sccm (a) to 100 sccm (b).


28.8 μm
Mag = 2.06 K X
10μm
EHT = 3.00 kV
WD = 5 mm
Signal A = InLens
Stage at T = 7.0 °
Date :14 Feb 2008
File Name = via etch 001.tif
EPFL-CMI
Mag = 4.10 K X
10μm
EHT = 3.00 kV
WD = 5 mm
Signal A = InLens
Stage at T = 7.0 °
Date :7 Mar 2008
File Name = 6th_10 mins_65sccm_20055.tif
EPFL-CMI

(a) (b)

Figure 2.5 SEM cross section images of tapered via holes etched with modified Bosch process when the temperature decreases from 20 °C (a) to 10 °C (b).



(a) (b)

(c)

Figure 2.6 SEM cross section images of the via holes with different nominal diameters of (a) 80 μm, (b) 30 μm, (c) 10 μm, etched with modified Bosch process.

Continuous Silicon Tapered Etch Process

The parameters influencing the etching are C_4F_8 flow rate, O_2 flow rate, platen bias, pressure and substrate temperature. We mainly studied the C_4F_8 flow rate, pressure and substrate temperature, other parameters having the values listed in Table 2.2. Note that whenever a given parameter was investigated, the others were set to the recipe baseline.

The sidewall angle can be modified by controlling the ratio of C_4F_8 during the etching. Adding C_4F_8 to the SF_6 gas flow in a plasma is known to drastically increase F-radical concentration. The silicon etching rate increases due to a decrease in F-radical density, but then decreases because of the deposition of carbon fluorine film and the dilution of F-radicals. SEM images in Fig. 2.7 show via cross sections with the C_4F_8 gas flow varying from 40% (280 sccm) to 57% (400 sccm), keeping the plate power at 70 W. The etch rate, undercut and slope angle as a function of the C_4F_8 flow rate are shown in Figure 2.8.

From Fig. 2.9, when decreasing the platen temperature from 20 °C to 10 °C, the undercut in 100 μm vias decreases drastically from over 20 μm to about 10 μm; obviously, significant improvement of sidewall smoothness can be seen for pattern opening as small as 20 μm; conversely, the etching rate is lower at 10 °C than at 20 °C, but the slope angle remains almost independent of the temperature. Further decreasing the temperature down to 0 °C confirms this trend.

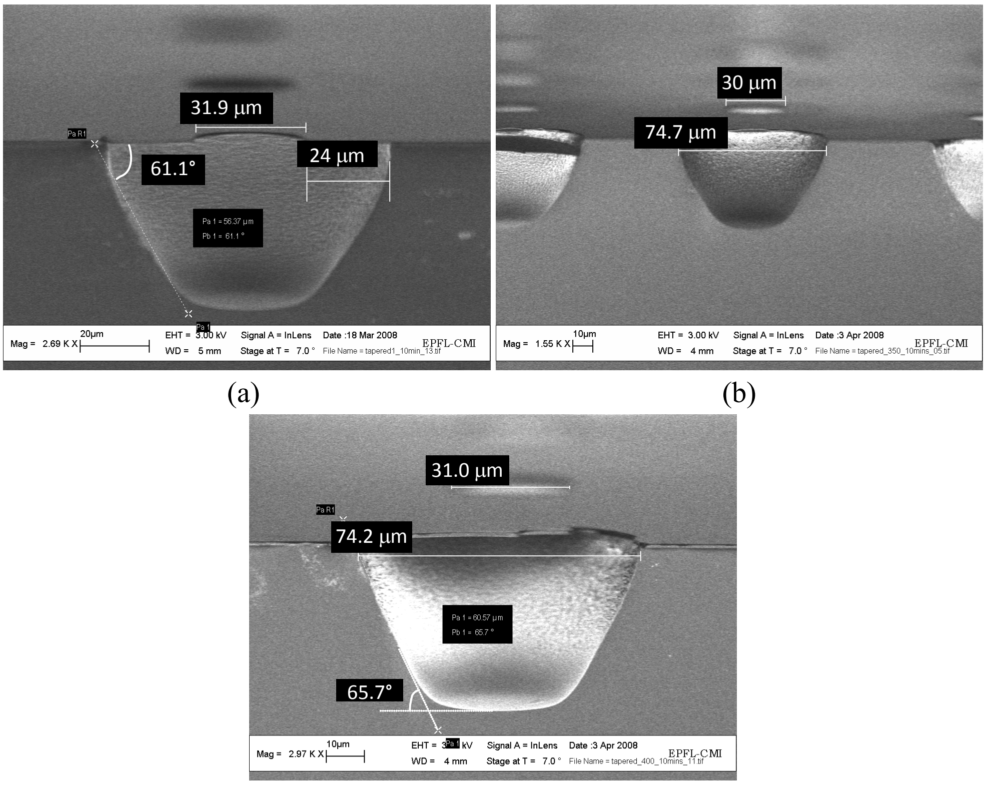
Setting the flow rates of C_4F_8 and SF_6 to 400 sccm and maintaining the temperature at 0 Celsius degree, the slope angle increases from 77.7 ° to 84.3 ° when increasing the platen power from 70 W to 100 W (Fig. 2.10). The selectivity of resist to silicon is lower under 100

Walts. The over-consumed resist pattern left during etching might be the reason of the "sliced" rough sidewall and chipped via top opening.

An optimized recipe has been developed from a detailed analysis of the previous experiments, as shown in Table 2.2. Fig. 2.11 shows SEM cross sections for six relevant widths corresponding to 10 min etching using this recipe. The most important observation is that the slope angle remains the same regardless of the via width. Compared to Bosch process, ARDE is moderate in this case (also plotted with numbers in Fig. 2.12). It can be concluded that the via shape results from a balance between etching and passivation, and that the etching mechanism remains the same regardless of the via size.

Table 2.2 Settings for Parameters Used for Continuous Etching Process.

Parameters	Investigation values	Baseline	Optimized
Coil power (W)	1200	1200	1200
SF6 flow (sccm)	400, 600, 1000	400	400
C4F8 flow (sccm)	280, 350, 400	400	400
O2 flow (sccm)	200	200	200
Platen bias (W)	70, 100	70	70
Pressure (mbar)	5e-2	5e-2	5e-2
Platen temperature (°C)	20, 10, 0	20	0
Etching duration (min)	10	10	10



(a)

(b)

(c)

Figure 2.7 SEM cross section images of the via holes etched with continuous Si etching process using different amount of additional C_4F_8 gas flow: (a) 280 sccm, (b) 350 sccm, (c) 400 sccm.

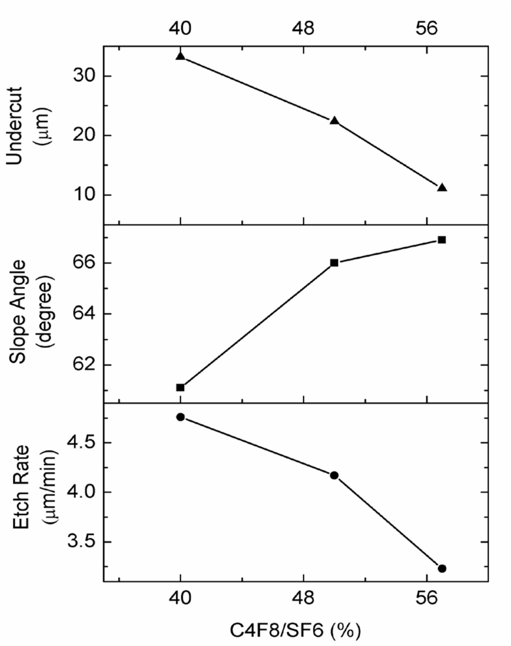
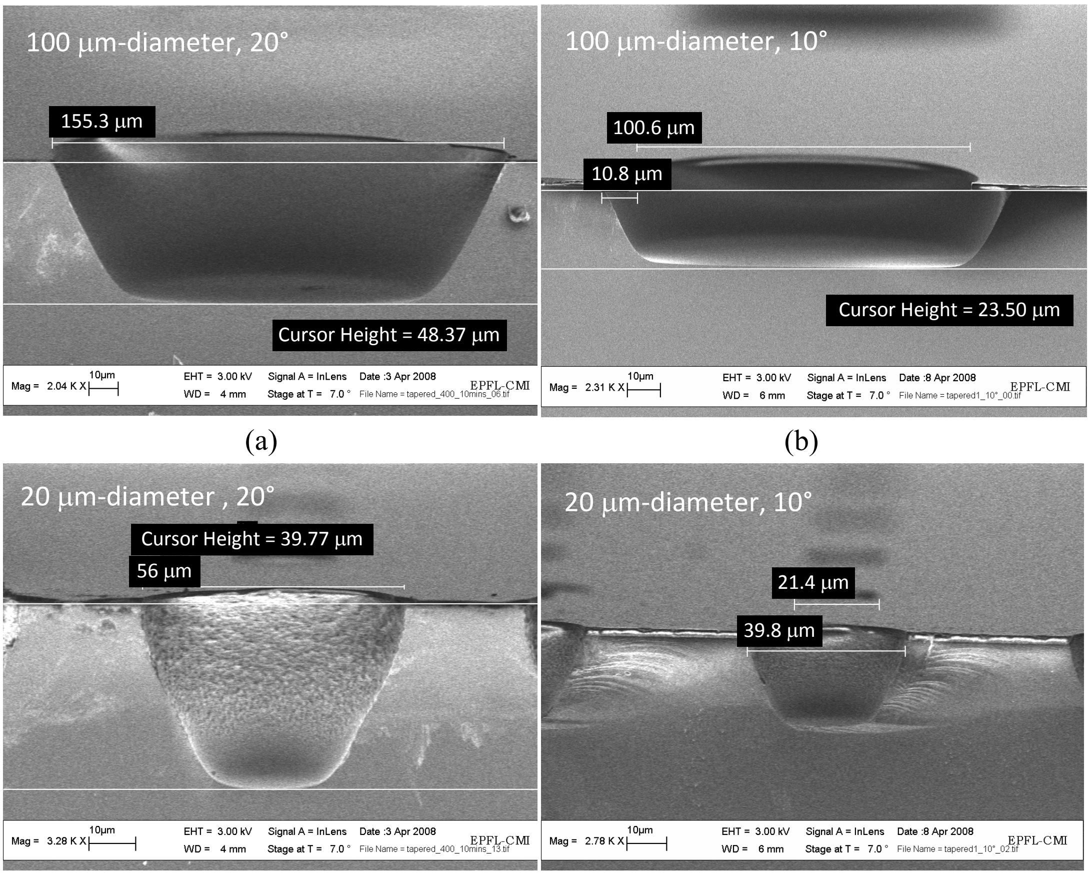


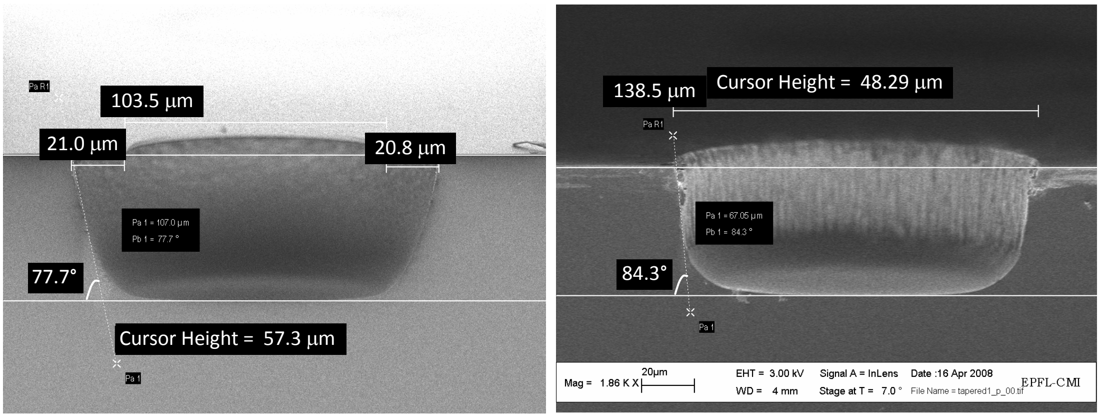
Figure 2.8 Etch rate, slope angle, and undercut using different amount of C_4F_8 for continuous Si etching. SF_6 flow rate uses baseline number, 700 sccm. Nominal TSV diameter is 30 µm.


100 µm-diameter, 20°
155.3 µm
Cursor Height = 48.37 µm
100 µm-diameter, 10°
100.6 µm
10.8 µm
Cursor Height = 23.50 µm
20 µm-diameter , 20°
Cursor Height = 39.77 µm
56 µm
20 µm-diameter, 10°
21.4 µm
39.8 µm
EPFL-CMI

(a) (b)

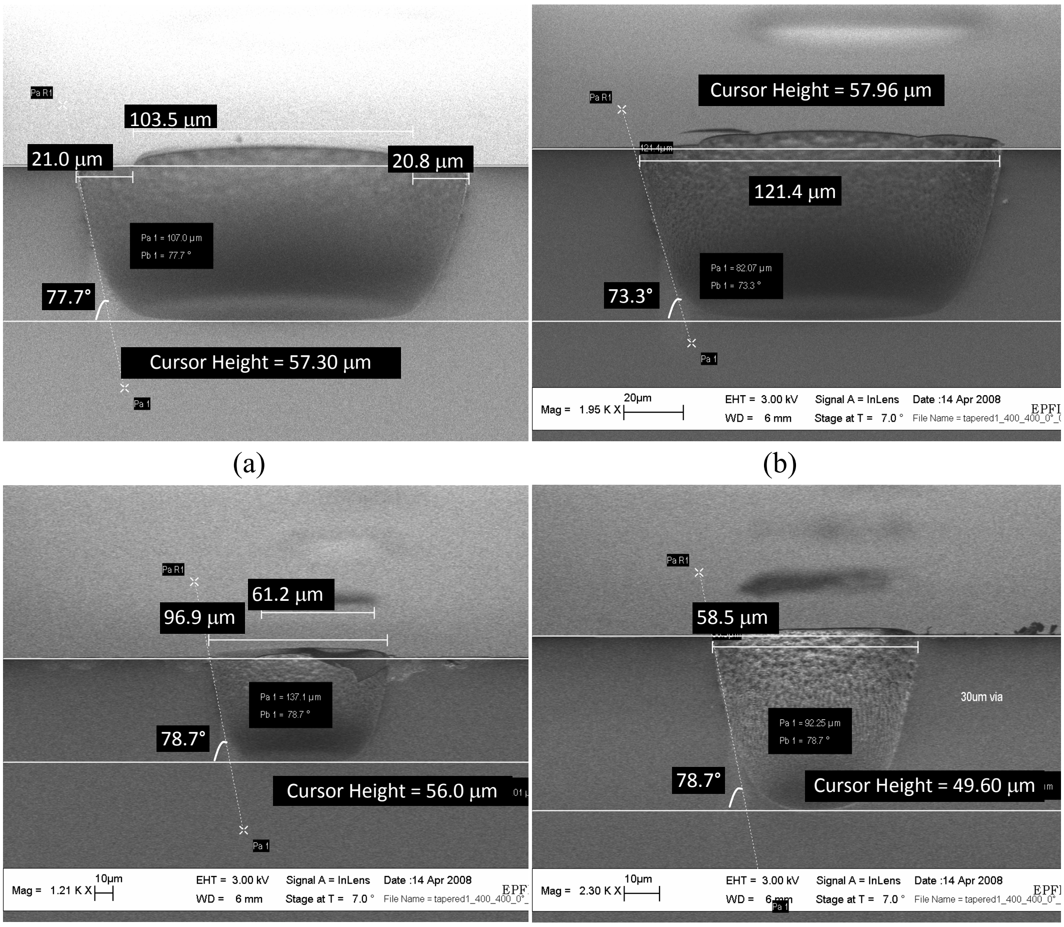
(c) (d)

Figure 2.9 SEM cross-section views of the via holes after the continuous silicon etching. The nominal TSV diameters and the platen temperatures are marked on the images.


103.5 µm
21.0 µm
20.8 µm
Pa 1 = 107.0 µm
Pb 1 = 77.7 °
77.7°
Cursor Height = 57.3 µm
138.5 µm
Cursor Height = 48.29 µm
Pa 1 = 67.05 µm
Pb 1 = 84.3 °
84.3°
20µm
Mag = 1.86 K X
EHT = 3.00 kV
WD = 4 mm
Signal A = InLens
Stage at T = 7.0 °
Date :16 Apr 2008
File Name = tapered1_p_00.tif
EPFL-CMI

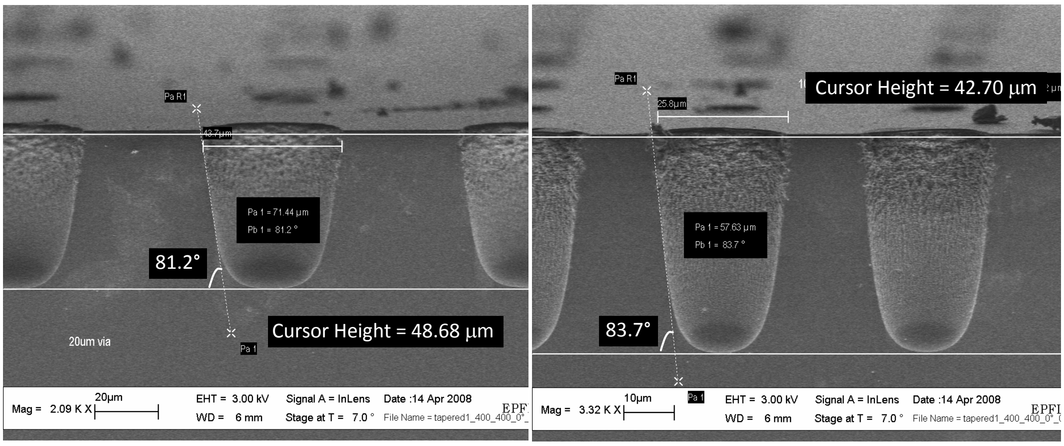
(a) (b)

Figure 2.10 Etching results of continuous etching under different platen powers. SF_6: 400 sccm, C_4F_8: 400 sccm, substrate temperature: 0 °C. (a) Platen power: 70 W, (b) Platen power: 100 W.


103.5 µm
21.0 µm
20.8 µm
Pa 1 = 107.0 µm
Pb 1 = 77.7 °
77.7°
Cursor Height = 57.30 µm
Cursor Height = 57.96 µm
121.4 µm
Pa 1 = 82.07 µm
Pb 1 = 73.3 °
73.3°
20µm
Mag = 1.95 K X
EHT = 3.00 kV
WD = 6 mm
Signal A = InLens
Stage at T = 7.0 °
Date :14 Apr 2008
61.2 µm
96.9 µm
Pa 1 = 137.1 µm
Pb 1 = 78.7 °
78.7°
Cursor Height = 56.0 µm
10µm
Mag = 1.21 K X
58.5 µm
30um via
Pa 1 = 92.25 µm
Pb 1 = 78.7 °
78.7°
Cursor Height = 49.60 µm
Mag = 2.30 K X

(a) (b)

(c) (d)


Pa R1
43.7µm
Pa 1 = 71.44 µm
Pb 1 = 81.2°
81.2°
Cursor Height = 48.68 µm
Pa 1
20um via
Mag = 2.09 K X
20µm
EHT = 3.00 kV
WD = 6 mm
Signal A = InLens
Stage at T = 7.0°
Date :14 Apr 2008
Pa R1
25.8µm
Cursor Height = 42.70 µm
Pa 1 = 57.63 µm
Pb 1 = 83.7°
83.7°
Pa 1
Mag = 3.32 K X
10µm
EHT = 3.00 kV
WD = 6 mm
Signal A = InLens
Stage at T = 7.0°
Date :14 Apr 2008

(e) (f)

Figure 2.11 SEM cross-section images of the via holes etched using the optimized continuous silicon etching recipe. TSV nominal diameters: (a) 100 µm; (b) 80 µm; (c) 60 µm; (d) 30 µm; (e) 20 µm; (f) 10 µm.

Summary

Continuous etching has demonstrated faster etching rates than the modified Bosch process since the cycle iteration is eliminated. For large TSV dimensions, both tapered processes have higher etching rates than the Bosch process. Both tapered processes show a more obvious ARDE effect than the Bosch process. The etch rates for all the three processes are shown in Fig. 2.12.

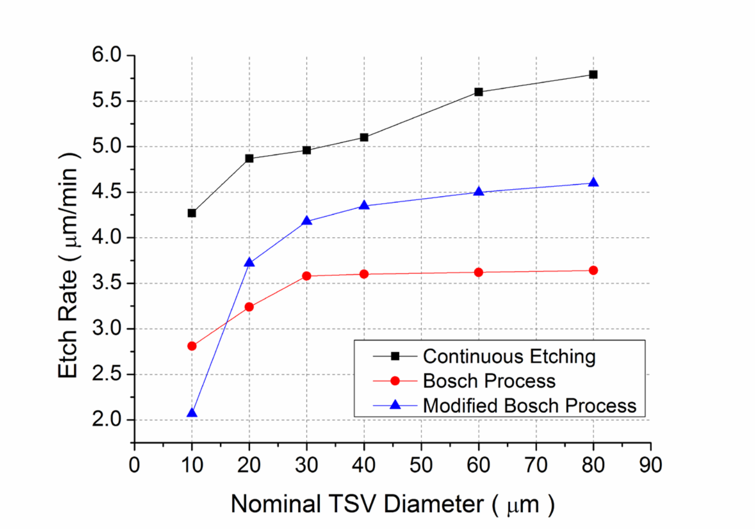


Figure 2.12 Etch rates for the via holes with different nominal diameters in Bosch process, modified Bosch process and continuous etching process.

The undercuts for vias with different nominal widths are shown in Fig. 2.13. Modified Bosch process leads to almost the same undercut for all via sizes, while the undercuts in the continuous etching process are aspect ratio dependent. This demonstrates different etching principles for the two processes. The modified Bosch process keeps the well-known advantage on small undercuts. On the other hand, continuous etching process has a higher etching rate, lesser local bowing and smoother surface.

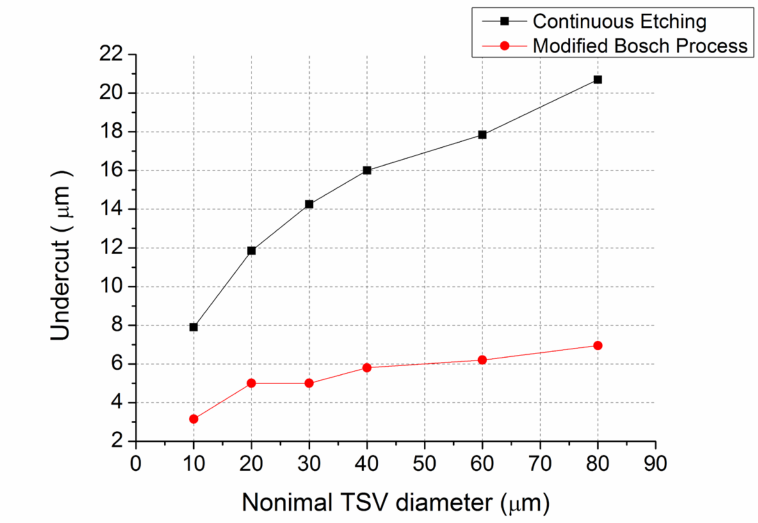


Figure 2.13 Undercuts for via holes with different nominal diameters in continuous etching and modified Bosch process.

Both processes have successfully demonstrated their ability to achieve a slope angle of about 80 °. But continuous etching process can avoid local bowing on top of the TSV, while the modified Bosch process still cannot solve this problem (Fig. 2.14).

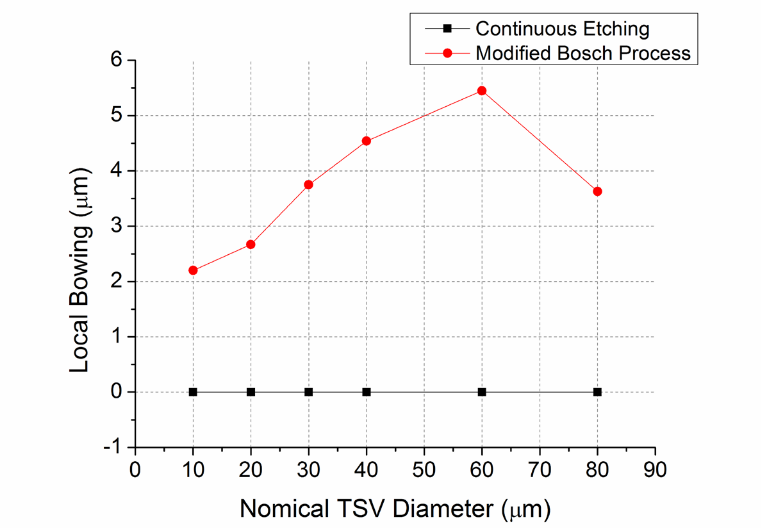


Figure 2.14 Local bowings for via holes with different nominal diameters in continuous etching and modified Bosch process.

Using the continuous etching process, the minimum TSV size that can be achieved is about 26 μm in diameter and 50 μm in depth. Since we would like to avoid the negative angle (local bowing) on top of the TSV to improve the sputtering coverage, we chose continuous etching for tapered silicon etching. For vertical TSVs, we aim at a similar TSV size which is 20 μm in diameter and 50 μm in depth.

2.2.3 Spray Coating in High Aspect Ratio (HAR) Topography

Spin coating is the most common method of coating planar surfaces. It can be used for high topography surfaces if certain modifications are made, e.g. varying the spin speed to allow sufficient time for the solution to flow and spread into the deep features. The problem is the uniformity of the coated layer that is not as good as for planar wafers and not always sufficient for the required resolution. In our case, the structures are TSV holes with a 1-2 μm sputtered silicon dioxide as isolation. Our aim is to remove the silicon dioxide at the bottom of via holes to get a metal connection between TSV and metal structures on the chip (e.g. MET2).

Direct spray coating of photoresists appears to be a promising technique for coating irregular surfaces. A schematic of the spraying system is shown in Fig. 2.15. Diluted resist is pumped through the nozzle of a spray head. Micron-sized resist droplet result and approach the substrate. During the flight phase, a certain amount of solvent evaporated. At an insufficient evaporation rate, it will prevent the droplets from sticking to the substrate; if evaporated too much, the increased droplet viscosity will prevent the resist from spreading on the substrate. The evaporation rate for each droplet as a function of the droplet surface is determined by temperature, droplet velocity, solvent saturation, solvent composition and concentration. The spray coating performances concerning the wetting, edge coverage and homogeneity are influenced by the initial coverage of the droplets from the atomized spray, the resist adhesion to the substrate, the resist surface tension, and its viscosity. Besides the chemical and physical resist properties, the atomized spray formation mechanism (droplet size distribution), the flight of the droplet to the substrate (evaporation, the spray angle), and the solvent evaporation out of the growing/grown resist film (time dependant surface tension and viscosity) determine the spray coating result with respect to homogeneous, smooth, and closed films.

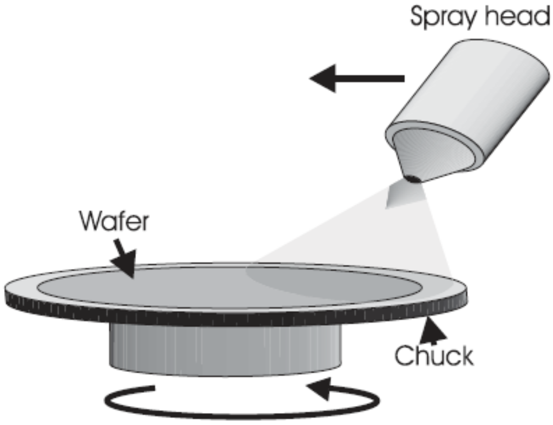


Figure 2.15 Schematic of the spraying system [7].

We use the spray coating system EVG150 from EVGroup. In our experiment, the only factor studied is the resist composition. The other parameters are all set according to the recommended values. The resist used was AZ9260 (Clariant Corporation) which is formulated with propylene glycol monometryl ether acetate (PGMEA).

The spray coating system is typically operated using resist solutions with a viscosity lower than 20 cSt in order to get the proper droplet size distribution. The photoresist can be diluted with solvents which 1) are same or similar chemistry as main photoresist solvent, e.g. the carrier solvent of the resist; 2) have no chemical reaction with the photoresist in mixture or during deposition and exposure; 3) are of high vapor pressure.

The first choice is PGMEA which is a carrier solvent used for photoresists. But using only PGMEA, the diluted resist will have a flowing effect resulting in an accumulation of photoresist at the bottom and depletion at the top corner of the cavities. Thus, the evaporation rate of the diluted resist is to be increased, but without increasing the viscosity. Methylethyl (MEK) is then introduced because of its higher evaporation rate than PGMEA. But with only MEK and no PGMEA in the mixture, it will tend to produce a rougher coating.

With a mix ratio of original AZ9260, PGMEA and MEK = 4 : 8 : 90, a 400 nm-thick AZ9260 is uniformly coated along the vertical via sidewall (Fig. 2.16).

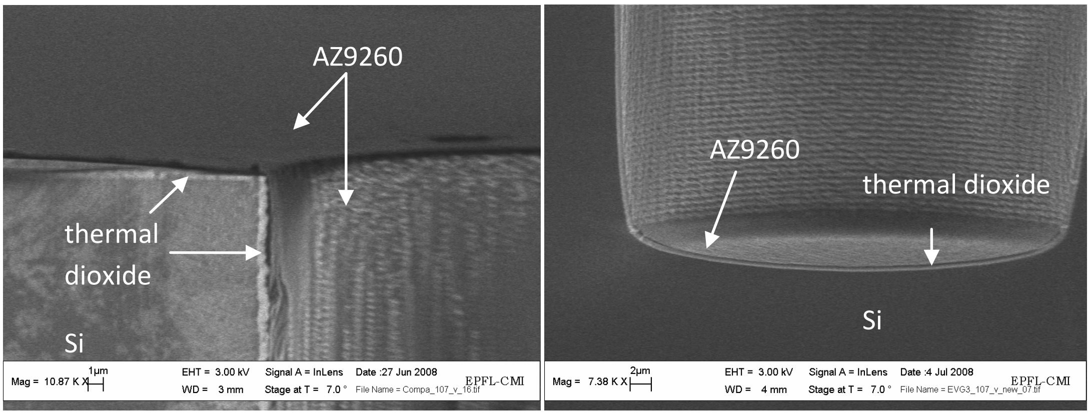


Figure 2.16 SEM cross section images of TSVs after spray coating and a soft bake.

2.2.4 Core-Hole TSV Detailed Process

The core-hole TSV process follows the process flow illustrated in Fig. 2.1.

Without wafer thinning and thin-wafer handling methods, our experiments can only be processed on blank full-thickness silicon wafers. However, our process flow is well-adapted to our initial aim. In our experiment, our process starts from etching blind via holes in cleaned 4 inch full-thickness (525 μm) <100> silicon test wafers without metal landing.

Before etching, positive-tone photoresist AZ9260 (Clariant Cooperation) is spin-coated on the wafer surface and patterned through photolithography to define the areas opened for via hole etching. The diameter of the via openings varies from 5 μm to 100 μm. Via holes are etched 50 μm deep into the silicon wafer using dry etch.

Deep reactive ion etching (DRIE) by Bosch process was adapted to provide via holes with straight sidewalls. The TSVs are with 20 μm diameter and 50 μm height. Due to the alternating nature of etching and passivation cycles used in the Bosch process, sidewall scalloping naturally occurs. The size of the scallops is well-controlled to avoid blocking of materials during dielectric isolation layer deposition, and the undercut is well-controlled as well to get a precise pattern transfer of the TSV size. Tapered TSVs with 80 ° profile are etched using continuous silicon etching process which has with the merit of almost no local bowing (Fig. 2.11).

After etching, the photoresist is stripped by a positive-tone remover heated at 65 °C for 10 minutes (Remover 1165 [10]). Then the structures were cleaned using buffered hydrofluoric acid (NH_4F 40% + HF 50%, also called "BHF") to remove any fluoropolymer residues formed during the etching process. Afterwards, 1 micron SiO_2 is sputtered using Spider 600 [9] as the dielectric isolation layer. In order to achieve metal contact from TSV to the landing pads, we should first use photolithography to define these areas, and then etch away the exposed SiO_2 (Fig. 2.1-4). Note that so far, the following experimental steps have been only tested on vertically etched via holes.

To be able to perform lithography at the bottom of the via, the resist first needs to cover the via sidewall uniformly. AZ9260 (the most used resist in CMI) was tested first. After spin coating, the resist was trapped on the upper part of the vias (Fig. 2.17). The same occurred for tapered-profile holes. We also tried diluted AZ9260 (called "AZ92XX" in CMI), but no improvement was observed. Then we proceeded with the spray coating method as it is known for its ability to transfer pattern on high topography surfaces [7, 8]. The optimal recipe consists in blends of AZ9260 : PGMEA : MEK with a mix ratio of 4 : 8 : 90, achieving a uniform ~400 nm-thick AZ9260 coverage on both tapered and vertical profiles.

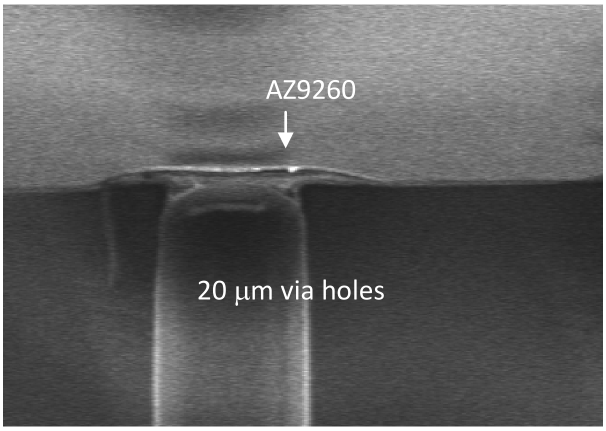


Figure 2.17 AZ9260 trapped on top of the via holes after spin coating.

The next step is resist exposure at the bottom of the via holes. In most cases, photolithography is performed on low topography microstructures on the wafer or directly on the planar wafer surface. During exposure, the mask comes into contact with the wafer surface or is held in proximity in order to get more precise pattern transfer (additional information to be found in [11]). The exposure time is calculated by the required energy divided by the light intensity.

In our case, light intensity is greatly reduced at the via bottom that is located 50 μm below the wafer surface; thus, the exposure dose, or equivalently, the exposure duration needs to be increased. Normally, for 400 nm AZ 9260 on the wafer surface, 1.4 second of exposure under light density of 10 mW/cm^2 is sufficient, but we tried an exposure of 3 seconds.

After resist development, the wafer is dipped in BHF to remove the exposed SiO_2. The etch rate of silicon dioxide in BHF is about 80 nm/min. So, about 12.5 minutes is required to completely remove 1 μm-thick SiO_2. In our experiment, 7 min BHF etching was performed. Then, the wafers were cleaved for SEM cross-section inspection. Successful dioxide etching indicates that the exposure time to open the area for BHF etching is fine (Fig. 2.18) since after 7 minutes, the dioxide is etched almost half down to 500 nm.

Worth to be mentioned, for the sample used in Fig. 2.18 and Fig. 2.19, a 1 μm-thick uniform thermal oxide is grown along the structured wafer surface. Because of a process temperature as high as 1000 °C, thermal oxidation is not adapted for BEOL TSV process. However, the silicon dioxide deposited by sputtering on the bottom of the vias is thinner than that deposited on the upper part of the same vias. Then, if the aspect ratio is too high, this difference can be deleterious.

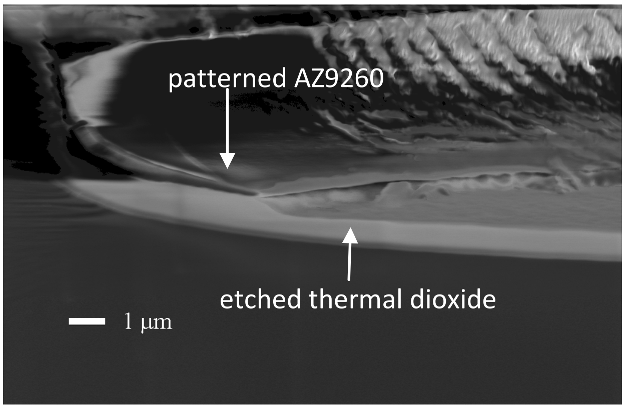


Figure 2.18 SEM image of the bottom of a via hole after 7 minutes BHF etching. Thermal dioxide thickness: 1 μm.

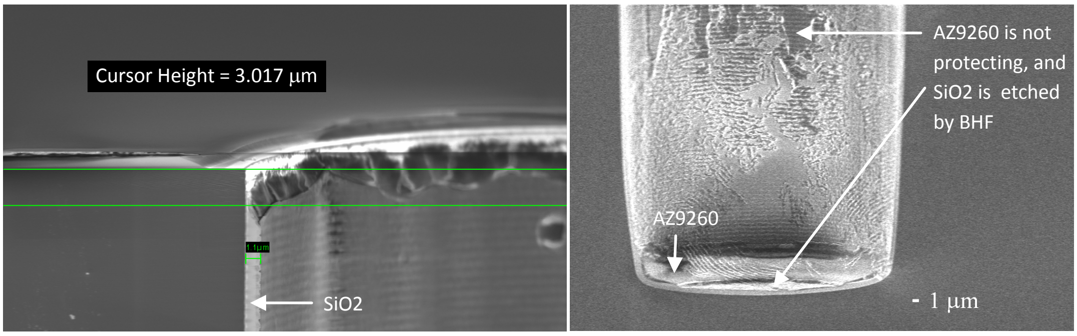
Although the silicon dioxide on the bottom of the via is fairly etched, the dioxide on the sidewall and the via top are etched or damaged partially at the same time. SEM images of the damaged TSV sidewall and top corner after BHF SiO_2 etching are shown in Fig. 2.19. One possible reason is that when the exposure dose is too high, light scattering in the resist film also exposes the "dark" parts of the resist (the part protected by the mask from exposure) as well as the thin resist layer on the sidewalls. In the top corners, photons are scattered the most which explains why the top corner endures the most severe damage. Another possible reason is that HF diffuses into the resist film and leads to resist peeling during the etching or after the subsequent rinsing because of resist swelling and the etching of the underlying SiO_2 not well-protected by the resist.

Following these arguments, three solutions might help: 1) increasing the resist thickness; 2) decreasing the via opening size in the Cr mask or using tapered vias so that the sidewall and top corner are further away from the light source; 4) using anisotropic SiO_2 dry etch instead of isotropic BHF wet etch. As the selectivity of photoresist and SiO_2 in the adapted process is $\geq$ 4:1, 400 nm photoresist is sufficient for etching 1 μm SiO_2.

Here comes another interesting idea that consists in directly performing SiO_2 anisotropic etching after it is deposited. As long as the SiO_2 on the bottom is removed, the SiO_2 on the wafer surface is also etched away. However, the dielectric on the sidewall is barely etched, so leaving a good dielectric layer around the TSV. The imperfect top TSV part can later be removed by CMP. Parylene can work as an alternative to SiO_2 providing a room temperature deposited uniform thickness isolation layer. The limitation is that it only works for vertical TSVs.

We can already observe some disadvantages of this process based on the experimented steps. First, it does not have a good scalability since process parameters have to be adjusted for each specific TSV size. Second, the aspect ratio of TSV is rather limited because of the difficulty to open the dielectrics at the bottom of the vias, which in other words means that the diameter of the TSV has to be rather large, at least several tens of microns. The process is very challenging, not as straightforward as it was expected at first. But for some TSV processes, for example, the MEOL TSV processes, without bottom dielectric opening, this process is the best sequence for on-going TSV fabrication.

In our experiment, we did not continue with this process and moved towards the "ring-trench TSV" process. We will see how this new approach offers good scalability of TSV diameter and avoids bottom dielectric opening process in the next section.



(a) (b)

Figure 2.19 SEM image of the damaged vertical TSVs after 7 minute BHF SiO_2 etching. (a) Top corner of the vertical vias after resist stripped. (b) Sidewall and bottom of the vertical vias with resist left on the sidewall.

2.3 Ring-Trench TSV Process

2.3.1 Process Flow

Ring-trench process can be briefly described as following: etching the ring trenches (Fig. 2.20-2) until the landing metal, filling the ring trenches with a polymer (Fig. 2.20-3), removing Si and SiO_2 core (Fig. 2.20-4 and Fig. 2.20-5), deposition of Ta as an adhesion and barrier layer, further deposition of Cu or Au as seed layer (Fig. 2.20-6), patterning the area dedicated to electroplating (Fig. 2.20-7), filling the via holes with copper by electroplating, and finally surface CMP and resist stripping (Fig. 2.20-8). This process was first proposed by IMEC, Belgium [12]. The third step in the process flow was developed in collaboration with IMEC.

This process has the following advantages. First, no lithography is required to expose the metal contact area at the bottom of the vias. Second, good scalability is achieved. Indeed, for TSVs with different diameters, the ring trench width is the same, so the silicon etching rate is almost the same all over the wafer. With the same via depth, once the polymer filling is successful, TSVs with different diameters can be processed at the same time. Third, the polymer working as the TSV dielectric layer is much thicker than that in core-hole TSV process, providing better isolation. Fourth, polymers can perform as stress relief or buffer from the reliability point of view. Polymers have been shown to increase device reliability by reducing "copper pumping", where copper pops out from the TSV holes (Fig. 2.21) during thermal cycling due to the combination of CTE mismatch, aspect ratio of the holes, and the hole diameter [13].

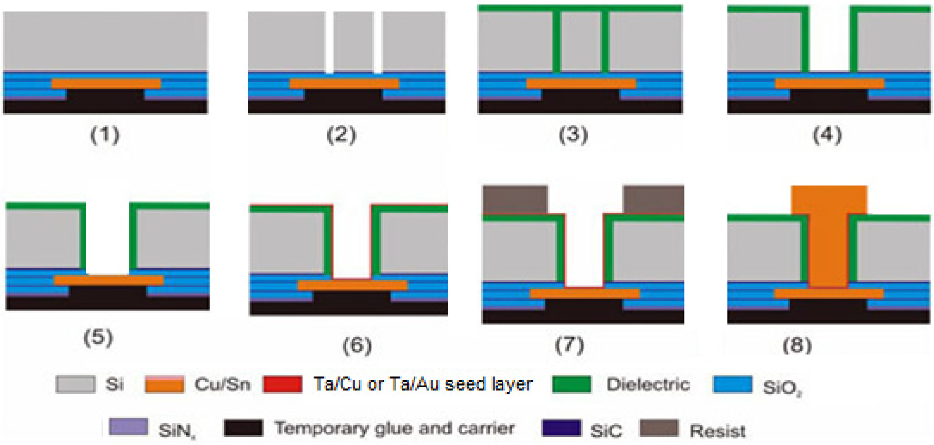


Figure 2.20 Ring-trench TSV process flow.

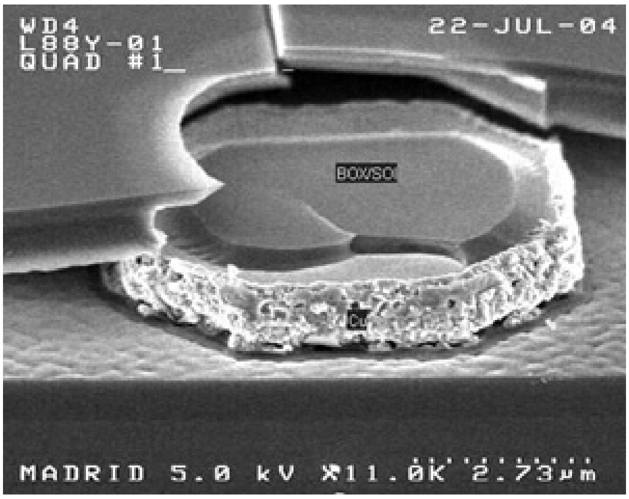


Figure 2.21 Copper pumping during thermal cycling. Copper pops out and breaks the dielectric layer around the TSV. Courtesy Tezzaron.

2.3.2 Filling Polymer in HAR Microstructures and Pattern Transfer

Filling deep trenches (AR = 10:1) with polymers for TSV dielectric isolation is the key step for the ring-trench TSV process. The ring-trenches (Width: 5 μm; Inner diameter: 5 μm, 20 μm or 50 μm; Depth: 50 μm) have been etched by DRIE (Fig. 2.22). After etching, the photoresist was stripped by soaping in a 65 °C remover for 10 minutes. Afterwards, the wafers were cleaned using BHF to remove any fluoropolymer residues formed during etching. Three polymers have been tested to fill in the ring trenches. Some of their properties are listed in Table 2.3.

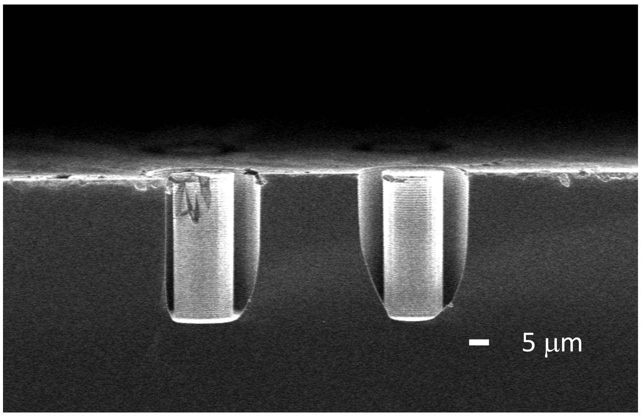


Figure 2.22 50 μm deep ring trenches etched in a silicon wafer.

Table 2.3 Materials Used for Polymer Filling.

Material	Tone	Viscosity	Shrinkage in Z-axis	Supplier
RH8023	negative	1000 cP	12%	Rohm & Hass
AZ9260	positive	500 cSt	Not known	Clariant
CCV-S300	positive	Not known, but lower than the other two	Not known	JSR

Polymer Filling Issues: 1. Viscosity

To fill in polymers in higher aspect ratio microtrenches, Duc *et al.* [14] demonstrated epoxy-based SU8 photoresist deep filling up to 30 μm. Mahfoz–Kotb *et al.*[15] investigated the use of a low- and high-viscosity spin-on dielectrics (SODs) and concluded that a viscous SODs, such as benzocylcobutene (BCB), is a better filler than SU8 and can fill 100 μm-deep cavities. Dang *et al.* concluded that low-viscosity SODs were not efficient in filling 100 μm-deep trenches compared to viscous polymers. In all these experiments, at first sight, viscosity seems to be a key parameter.

Normal spin coating of highly viscous AZ9260 without any additional surface treatment led to some trapping on the upper part of the via (Fig. 2.17). No improvement was obtained with lower viscous resists such as polyimide and AZ92xx. Even adding a one-minute pause during the spin coating procedure (Table 2.4, step 3) to allow resist flowing deeper into the trenches, not much improvement could be observed. For RH8023 and CCV-S300, similar behavior was observed.

Therefore, we came to the conclusion that parameters other than viscosity (or solid content) must play a role in the TSV filling properties, for example, the polymer wetting properties.

Polymer Filling Issues: 2. Wettability

Photoresists are usually hydrophobic, and thus have a good wettability or adhesion on hydrophobic surfaces. Actually, our samples were cleaned with BHF before resist coating; thus the resulting trench sidewall reveals an H-passivated and hydrophobic surface with a very good adhesion for resists. The above results seem to show that good resist affinity does not necessarily lead to good filling because in those trials, the resists did not creep into the via holes.

Our following experiment seemed to provide a contradictory demonstration. Adding a 5-minute PGMEA prewetting at the beginning of the process, a better filling was observed. For RH8023, resist fills half of the via holes, as shown in Fig. 2.23. With PGMEA left on the wafer surface, the wafer surface tension is reduced providing better wettability to the resists.

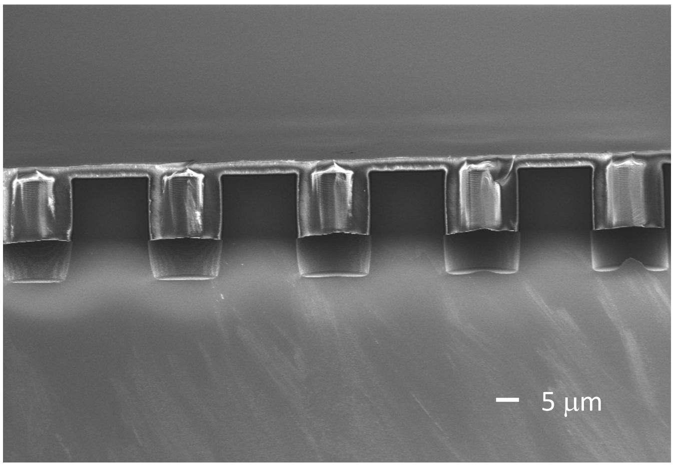


Figure 2.23 Ring trenches filled with RH8023 with PGMEA prewetting.

Another experiment was done to clear this confusion. After wafer cleaning with piranha, we observed complete trench filling with all the three resists tested, which is exactly the goal we wanted to achieve. One or two minutes' O_2 plasma can work as alternative of piranha as well. The spin coating procedures follow the parameters shown in Table 2.4. The resists went down to the bottom of the trenches and left no void in the filled parts, as shown in Fig. 2.24 and Fig. 2.25.

As known, in piranha cleaning and O_2 plasma, oxidation happens on silicon, so the surface for resist filling is hydrophilic, which in other words means that the surface is of no affinity or poor wettability to resist). Poor wetting of the resists boosts the filling performance. The improvement brought by prewetting would be rather explained from a mechanical action of the solvent. Duval *et al.* also came to that conclusion.

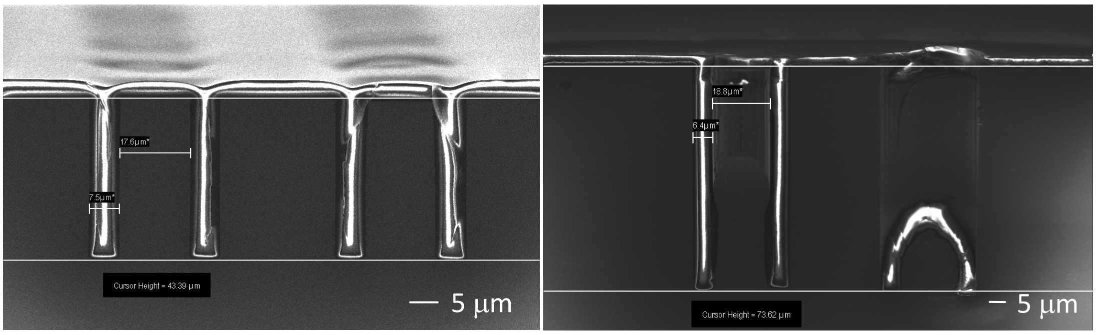
Polymer Filling Issues: 3. Faster and Simpler filling

Once the resist can flow down into the trench, the viscosity of the polymer starts to play a role on the filling performance. Polymers with higher viscosity are filling the trenches more efficiently. For RH8023 which is of the highest viscosity, a single spin coating is sufficient to completely fill in both 50 μm and 100 μm deep trenches (Fig. 2.24) whereas for AZ9260, which is of a lower viscosity, a second spin coating is required (Fig. 2.25-a). Finally, for CCV-S300 which has the lowest viscosity, up to three spin coatings are necessary to fill up only three fifth of the total depth (Fig. 2.25-b). Note that PGMEA prewetting is only applied for the first spin coating.

To sum up, the most critical factor for filling is surface wettability. Oxidized surface after piranha or O_2 plasma having no affinity with resists helps the resist to flow into the bottom of the deep trenches. This is the key to achieve complete filling without void. Once the resists are able to reach the bottom of the trenches, high viscosity polymers are preferred since more resists can stay in the trenches after maximum-speed 2nd spinning. Finally, solvent prewetting helps filling as it eases the polymer to flow down along the sidewall.

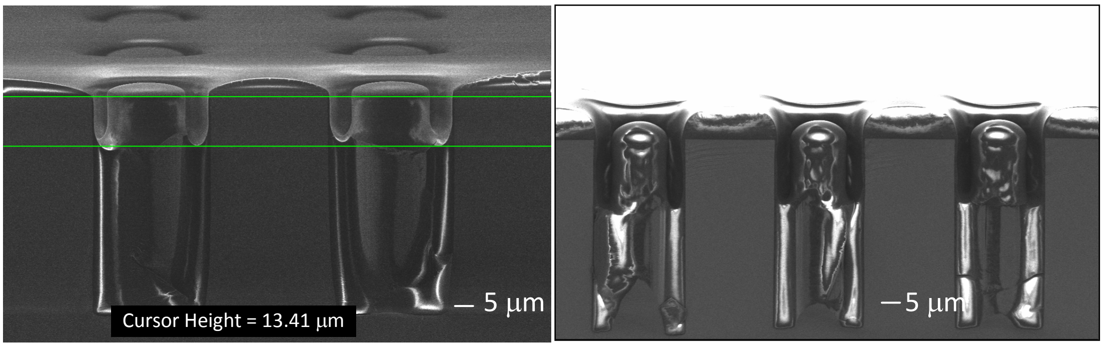
Table 2.4 Polymer Coating Procedure.

Step	Processing	RH8023	AZ9260	CCV-S3000
1	PGMEA prewetting	PGMEA left for 5 minutes, and then spun off	PGMEA left for 5 minutes, and then spun off	PGMEA left for 5 minutes, and then spun off
2	1st spin	500 rpm for 8 s, acceleration: 500 rpm/s	400 rpm for 8 s, acceleration: 400 rpm/s	300 rpm for 8 s, acceleration: 300 rpm/s
3	Stay for 1 min	0 rpm	0 rpm	0 rpm
4	2nd spin	5000 rpm for 1 min, acceleration: 2500 rpm/min	5000 rpm for 1 min, acceleration: 2500 rpm/min	5000 rpm for 1 min, acceleration: 2500 rpm/min
Repeating NO. 2-4	Repeat coating	0 time	1 time	2 times
5	Prebaking	140 °C for 3 min	110 °C for 3 min	110 °C for 5 min
6	Thermal cure	200 °C for 60 min, natural cool	140 °C for 60 min, natural cool	200 °C for 60 min, natural cool



(a) (b)

Figure 2.24 SEM cross section images of the ring trenches after RH8023 spin coating and prebaking. (a) AR = 10:1. (b) AR = 20:1.



(a) (b)

Figure 2.25 SEM cross section images of the ring trenches after spin coating and prebaking. (a) AZ9260 after one spin coating. (b) CCV-S300 fills 3/5 of the trenches after 3 times of spin coating. The trench depth is 50 µm.

Baking Issues

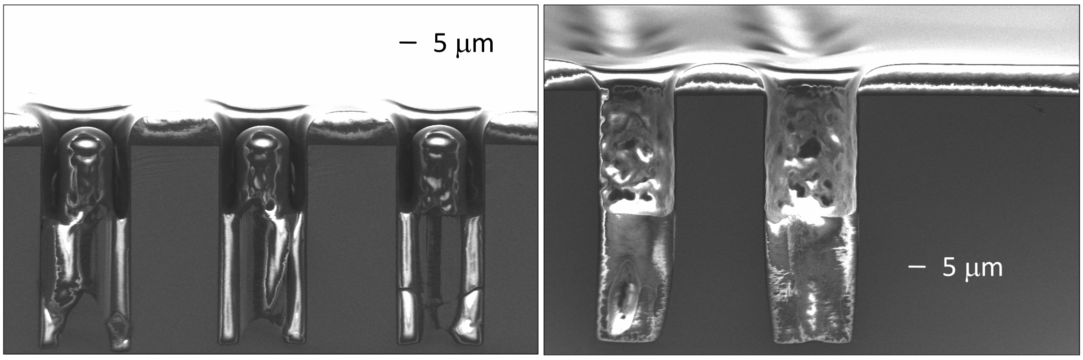
After filling (prebake done already), polymers have to be cured to achieve higher mechanical and chemical stability. High shrinkage of the resist during prebake and curing might damage the microstructures. Various curing methods such as UV fast cure are available, however, we only experimented the most used one that is thermal curing.

The RH8023 resist remains almost unaffected after curing while the CCV-S300, it shrinks from three fifth to about a half of the depth (Fig. 2.26). Therefore, CCV-S300 is not a good choice. AZ9260 film endures severe brownishing and cracking after a thermal curing at 140 °C for 1 hour (Fig. 2.27). AZ9260 reacts with atmospheric oxygen and embrittles above approx. 120-130 °C. The different CTE of resist/substrate leads to cracking of the resist which could be suppressed by a moderate cooling ramp. However, for a long time at high temperature (one hour at 140 °C and a few hours for slow cooling), even a slow cooling cannot avoid cracks. Neither is AZ9260 a good choice.

On the other hand, the results for RH8023 are very encouraging, and thus RH8023 will be the one used for polymer filling in ring-trench TSV process. The detailed experimental procedure for thermal curing of RH8023 is plotted in Fig. 2.28.

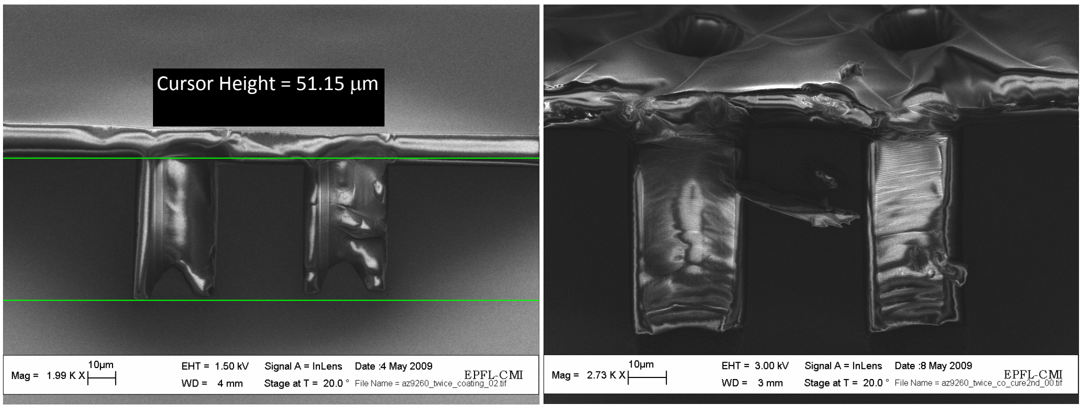
For further work, characterizing polymerization is necessary. Differential scanning calorimetry (DSC) could be an interesting method. Melting points T_m and glass transition temperatures T_g for most polymers are available from standard compilations (Fig. 2.29) which can be measured by DSC. The degree of polymerization is shown by the melting point, T_m, for example. T_m depends on the molecular weight of the polymer, so lower polymerization will have lower melting points than expected in the data sheet of the polymer.

Other faster curing methods are to be tried. For example, UV curing. For UV curing, typically, seconds-to-minutes' exposure times are sufficient while for thermal curing several hours are required to wait for the slow temperature ramp and natural cooling.



(a) (b)

Figure 2.26 Comparison of CCV-S300-filled ring trenches after prebaking (a) and after thermal curing (b).


Cursor Height = 51.15 μm
Mag = 1.99 K X
10μm
EHT = 1.50 kV
WD = 4 mm
Signal A = InLens
Stage at T = 20.0 °
Date :4 May 2009
File Name = az9260_twice_coating_02.tif
EPFL-CMI
Mag = 2.73 K X
10μm
EHT = 3.00 kV
WD = 3 mm
Signal A = InLens
Stage at T = 20.0 °
Date :8 May 2009
File Name = az9260_twice_co_cure2nd_00.tif
EPFL-CMI

(a) (b)

Figure 2.27 Comparison of AZ9260-filled ring trenches after prebaking (a) and after thermal curing (b).

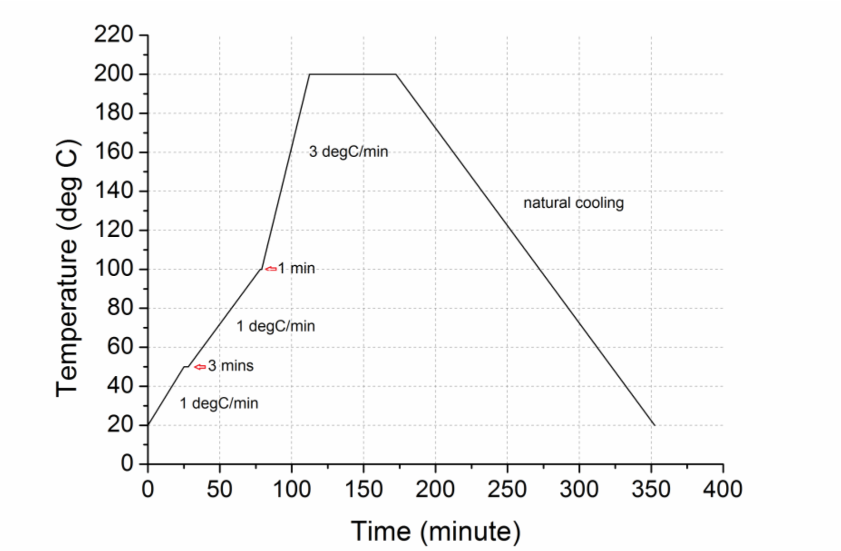


Figure 2.28 Experimental procedure for thermal curing of RH8023.

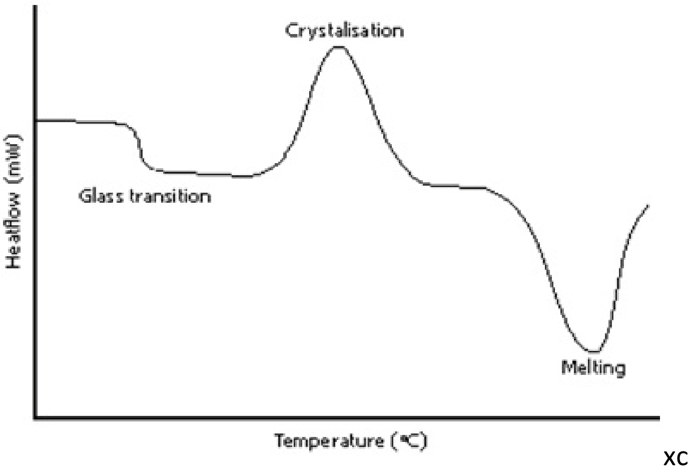


Figure 2.29 A typical DSC curve [17].

2.3.3 Electroplating Mold for Cu Electroplating

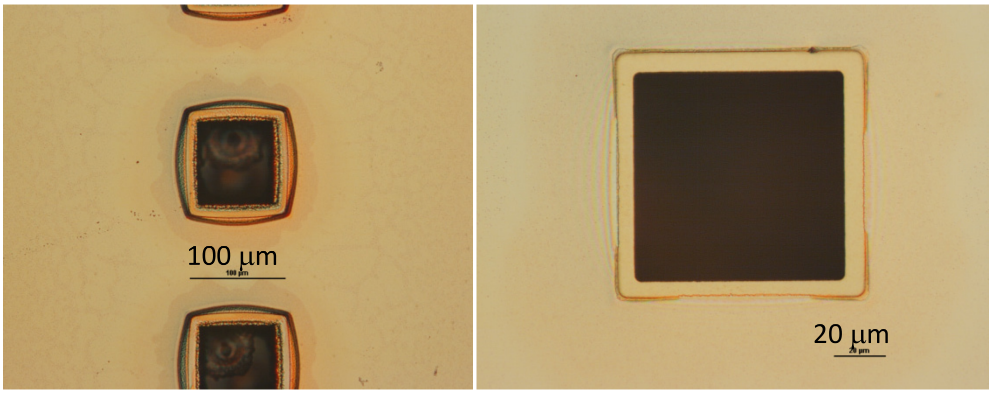
To pattern across high aspect ratio microstructures such as via holes, negative photoresists should be used since they allow precise photolithography on the planar wafer surface. For electroplating applications, the resist also needs to be thick enough to provide thick patterns over 10 μm. RH8023 is a potential candidate. Before RH8023 spin coating, 60nm Ta and 200 nm Au are sputtered. Right before coating, a few seconds' O_2 plasma is performed on the Au surface to remove residual organics. The detailed lithography procedure follows Step 1 - 6 listed in Table 2.5. Hard bake or even curing steps (Step 7) are sometimes performed to increase the thermal and chemical stability of developed resist structures for subsequent processes such as electroplating, wet or dry chemical etching. Because the experiment so far focused on lithography, Step 7 was not performed.

Table 2.5 Lithography Procedure for 7 μm-Thick RH8023.

Step	Process	Parameters
(0)	Prewetting[1]	5 min PGMEA
1	Spin coat	3500 rpm, 30 secs
2	Pre-bake	145 deg C, 300 μm proximity, 3 mins
3	Exposure[2]	400 mJ/cm^2
4	Post Exposure Bake	100 deg C, 150 μm proximity, 5 mins
5	Development	60 secs × 2 puddle, 2.38% TMAH at 23 deg C
6	Post Development Bake	140 deg C, 150 μm proximity, 30 secs
7[3]	Cure (Oven)	100 deg C, 30 mins

1: Not optimized.
2: Exposured under "UV, h, i line" spectrum with light intensity of "8.0 mW/cm^2".
3: Not performed so far in our experiment.

The result shows that many residues sticking on the sidewalls and via openings (Fig. 2.30-a). An additional prewetting (Step 0) was added into the lithography procedure to solve this problem. Clean and precise pattern openings can then be observed (Fig. 2.30-b).



(a) (b)

Figure 2.30 RH8023 electroplating mold on 100 µm-wide TSV holes with Ta/Au seed layer. (a) Without PGMEA prewetting. (b) With PGMEA prewetting.

The SEM cross-section views of circular TSVs (Fig. 2.31) with different diameters show that: the resist above 10 µm diameter via holes cannot be fully developed (Fig. 2.31-a); many resist residues and damages exist on the seed layer, which are evidenced on the sidewall of 20 µm-diameter via (Fig. 2.31-b); when via holes are larger than 30 µm, the resist molds show good shapes with clean and complete sidewalls (Fig. 2.31-c, d, e).

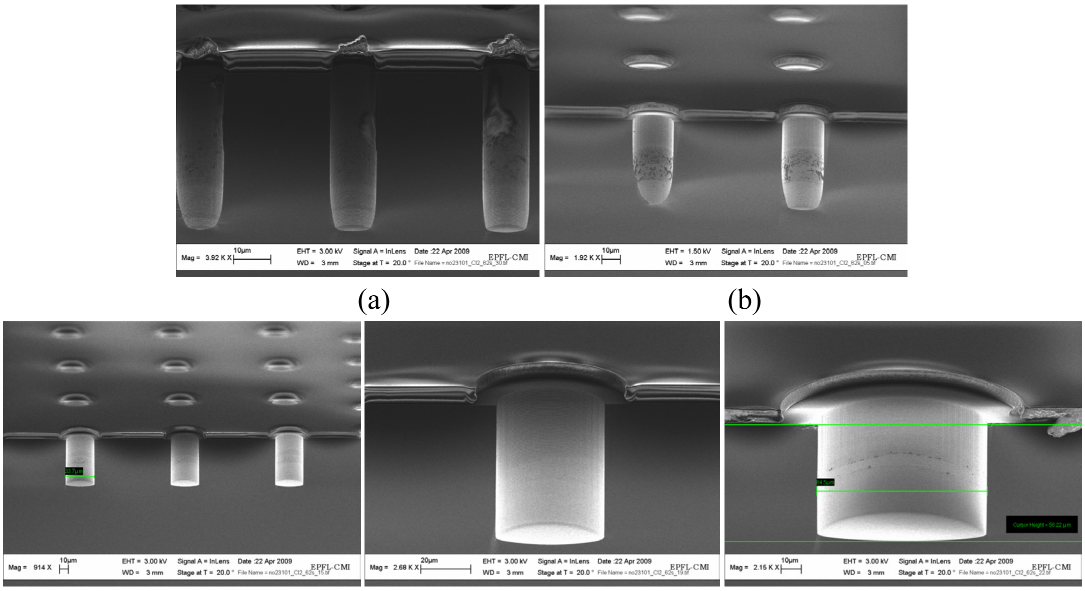
The residues and damages on the metal seed layer (Fig. 2.32) will cause current inhomogeneity which could be detrimental for electroplating, especially when the vias are to filled completely by copper. Electrolessplating of nickel or gold can be performed to get a better seed layer, but this technique is not available in our clean room. To remove completely the residues, development time was increased, but with no improvement, which suggests that the residues are rather exposed already. Note also that overexposure usually leads to clogged openings for small microstructures. A possible explanation for presence of residues and for the closed via openings could be photon reflectivity on the metal-covered substrate.

To test this hypothesis, a bottom-layer anti-reflective coating (BARC) benchmarked as AZ Barli II [18] was spin-coated before RH8023 coating. As shown in Fig. 2.33, the openings above the 10 µm-diameter via holes are much better defined after lithography (comparing to Fig. 2.31-1); the sidewall is well protected from damages as well. The problem is that this BARC is of positive tone and remains on the via sidewalls after development of RH8023, which is a negative photoresist. Another dry etch is then necessary to remove this layer. However, both copper and gold are not allowed to be exposed directly to plasma in our etchers. So, this BARC was not used for our final process.

Another method that can be adopted to reduce residues is to decrease the prebake temperature, e.g. from 140 °C to 130 °C. This is a known method to reduce the scum on Al pads. However, we should note that this still cannot solve the clogged opening issue on 10 µm-diameter openings.

Worth to be mentioned, repeating the procedure in Table 2.5 using Ti/Pt seed layer, no matter with or without prewetting, much resist residues remain both on top of the via and on the via sidewall (Fig. 2.34). It was reported that more scattering happened with Pt than with Au [19], so with the same exposure dose, the resist coated on Pt is now overexposed.

To sum up, it comes out that Ta/Au works better than Ti/Pt as the seed layer, our electroplating mold technique only works for vias with diameter ≥ 30 µm; and fortunately, the aspect ratio of the via hole has no influence on pattern transfer.


(a) (b)

(c) (d) (e)

Figure 2.31 Electroplating molds on vias with various diameters of (a) 10 μm; (b) 20 μm; (c) 30 μm; (d) 40 μm; (e) 80 μm.

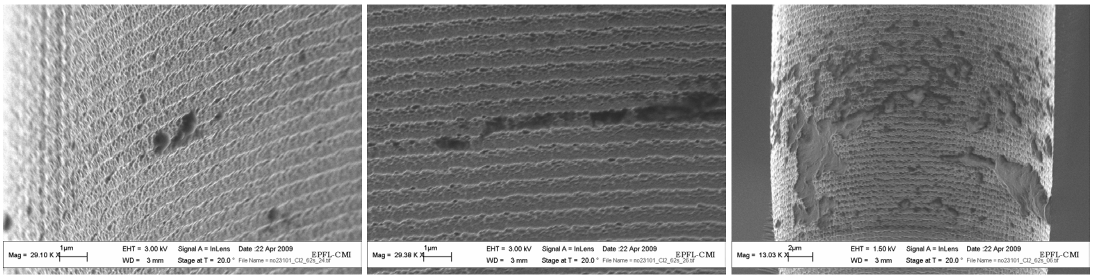

(a) (b) (c)

Figure 2.32 Residues and damages on the via sidewall after resist development. (a) Residues on the sidewall of 100 μm-diameter via; (b) Damages on the sidewall of 100 μm-diameter via; (c) Damages on the sidewall of 20 μm-diameter via.

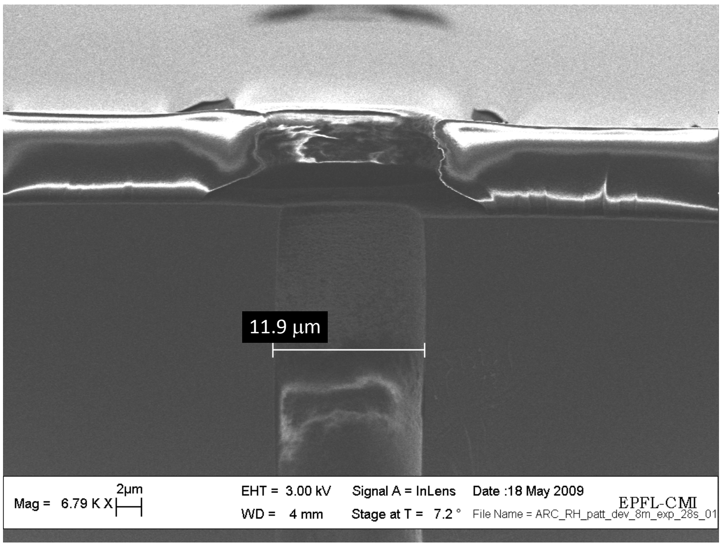
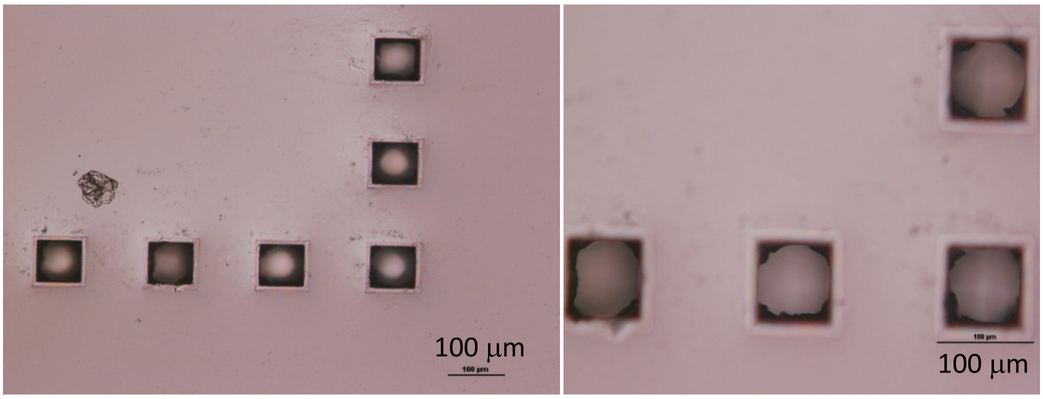

11.9 µm
Mag = 6.79 K X
2µm
EHT = 3.00 kV
WD = 4 mm
Signal A = InLens
Stage at T = 7.2 °
Date :18 May 2009
File Name = ARC_RH_patt_dev_8m_exp_28s_01
EPFL-CMI

Figure 2.33 Lithography cross a 10 μm-diameter via hole with a BARC coated below RH8023.



(a) (b)

Figure 2.34 RH8023 Electroplating mold on 100 µm-wide via holes with Ti/Pt seed layer. (a) Via top after development. (b) Much resist residue on the via sidewall after development.

2.3.4 Ring-Trench TSV Detailed Process

As for core-hole TSV, ring-trench TSV process started with via hole etching in a 4 inch silicon wafer without landing on real metal layer. A 5-µm thick AZ9260 was spin-coated and patterned on the wafer surface to define the ring patterns. Bosch process was then applied to etch the ring trenches 50 µm-deep into the silicon substrate. After resist stripping, the wafer was cleaned using BHF to remove the fluropolymers from etching, and then cleaned with piranha or 1 min oxygen plasma.

Filling Ring Trenches

To completely fill the ring trenches, RH8023 was our baseline material. The baseline process was: 1) Piranha-etch; 2) 5-minute PGMEA prewetting; 3) slow-rate spinning, 300 rpm/s for 8 seconds; 4) pause during 1 min to allow resist trench filling; 5) maximum-speed spinning at 5000 rpm/s for 60 seconds; 6) and at last thermal cure at 200 °C for 90 minutes. The SEM cross-section image of a completely filled trench after curing is shown in Fig. 2.35.

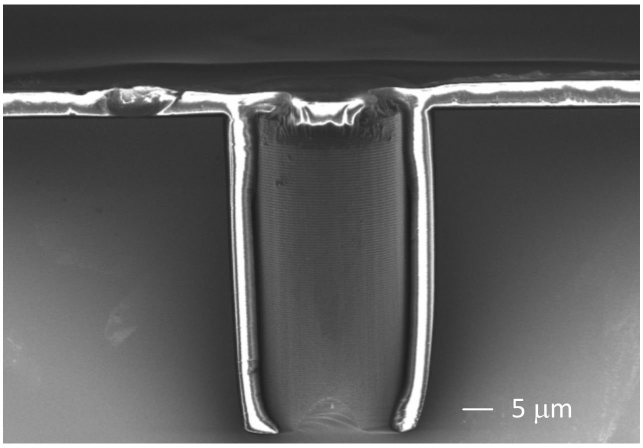
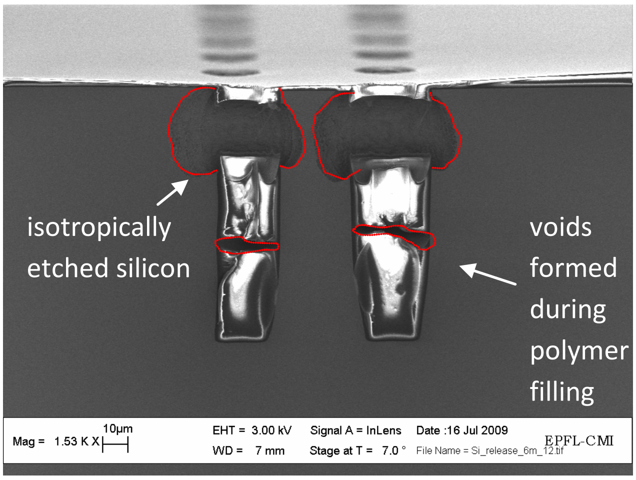


Figure 2.35 SEM cross section image of a ring-trench TSV after RH8023 filling and curing.

After thermal curing, 7 - 8 µm thick RH8023 still covered the wafer planar surface. Another photolithography was then performed on top of the RH8023 layer using 5 microns of AZ9260. AZ9260 above TSV central cores was exposed and solved in the developer. Anisotropic polymer etching using O_2 with an etching rate of ~ 1 µm /min was performed in STS multiplex ICP etcher [35] for 8 minutes. All the RH8023 above TSV central cores was completely removed, thus exposing the silicon cores for removal. Aside these cores remained a 4 µm-thick RH8023 that would work as the etching mask and the final dielectric isolation.

Etching the Silicon Core

After the central silicon core is exposed, dry etch was performed to remove the silicon inside the polymer rings. Both anisotropic (Bosch) and isotropic silicon etching have been experimented.

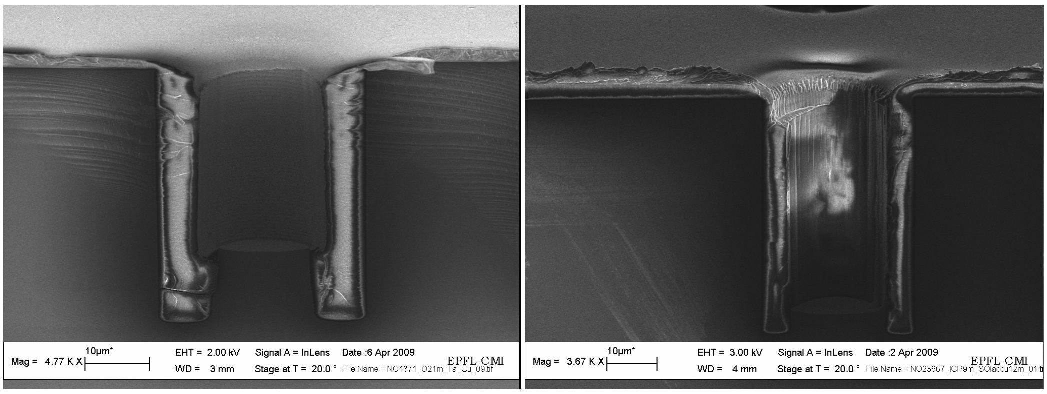


isotropically etched silicon
voids formed during polymer filling
Mag = 1.53 K X
10µm
EHT = 3.00 kV
WD = 7 mm
Signal A = InLens
Stage at T = 7.0 °
Date :16 Jul 2009
File Name = Si_release_6m_12.tif
EPFL-CMI

Figure 2.36 SEM cross section image of damage on Si substrate after Si core removal using isotropic silicon etching because of leakage defect in the polymer.

Using isotropic etching, any silicon area not well-protected by dielectrics was to be etched, leaving a big hole on the TSV sidewall (Fig. 2.36). Conversely, anisotropic silicon etching usually did not show this problem. But after Bosch process, the sidewall was much rougher especially around the top opening, possibly because of the passivation layer grown. BHF etching and O_2 plasma were used to decrease sidewall roughness, but did not bring any improvement. Fig. 2.37-b shows a SEM cross section image of the TSV after the silicon core has been removed by Bosch process, then a 30 second O_2 plasma and Ta/Cu sputtering.

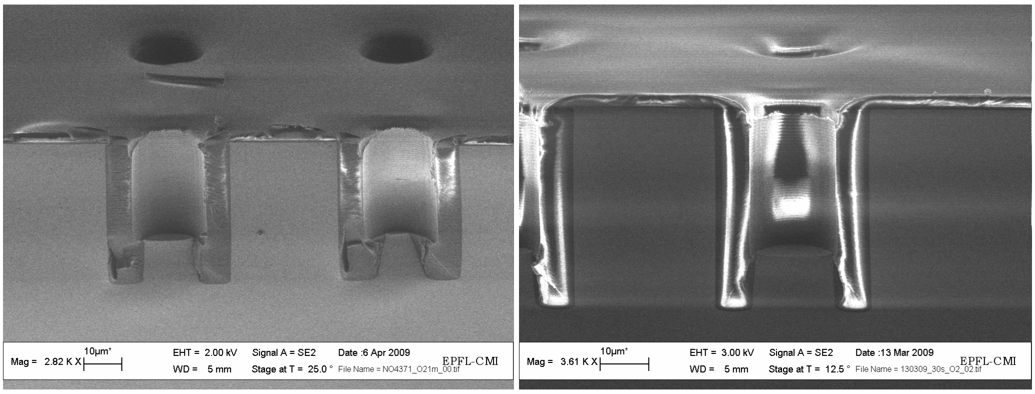
Isotropic silicon etching achieved smoother and cleaner TSV sidewall (Fig. 2.37-a and Fig. 2.38) and thus was chosen as the baseline process. As shown in Fig. 2.38, the TSV top corner was well-covered by the polymer. After 30 second O_2 plasma, the top corner was rounded and the sidewall got smoother (Fig. 2.38-b), which helped to get a better metal coverage in the following steps.



Mag = 4.77 K X
10µm*
EHT = 2.00 kV
WD = 3 mm
Signal A = InLens
Stage at T = 20.0 °
Date :6 Apr 2009
File Name = NO4371_O21m_Ta_Cu_09.tif
EPFL-CMI
Mag = 3.67 K X
10µm*
EHT = 3.00 kV
WD = 4 mm
Signal A = InLens
Stage at T = 20.0 °
Date :2 Apr 2009
File Name = NO23667_ICP9m_SOIaccu12m_01.t
EPFL-CMI

(a) (b)

Figure 2.37 SEM cross-section images of TSVs after silicon core removal ((a) after isotropic Si etch or (b) after anisotropic Si etch), 30 second O_2 plasma and Ta/Cu seed layer sputtering.


Mag = 2.82 K X
10μm*
EHT = 2.00 kV
WD = 5 mm
Signal A = SE2
Stage at T = 25.0 °
Date :6 Apr 2009
File Name = NO4371_O21m_00.tif
EPFL-CMI
Mag = 3.61 K X
10μm*
EHT = 3.00 kV
WD = 5 mm
Signal A = SE2
Stage at T = 12.5 °
Date :13 Mar 2009
File Name = 130309_30s_O2_02.tif
EPFL-CMI

(a) (b)

Figure 2.38 SEM cross section images of ring-trench TSVs after central silicon core removal. (a) After isotropic silicon etching. (b) After isotropic silicon etching and O_2 plasma.

Alignment Accuracy to Define Silicon Holes

Alignment precision is a key issue to get a good dielectric isolation around the via top corners. Irregular corners will obstruct metal particles during sputtering and cause failure in the following processing steps. Silicon can be etched through the leakage path caused by some misaligned exposure during isotropic silicon etching, (Fig. 2.39). High aspect ratio copper electroplating is sensitive to the uneven current density distribution due to the irregular corners.

In addition, TSV top corner suffers from the largest thermal mechanical stress. The copper pumping problem makes via top corner the most delicate part of TSV structure.

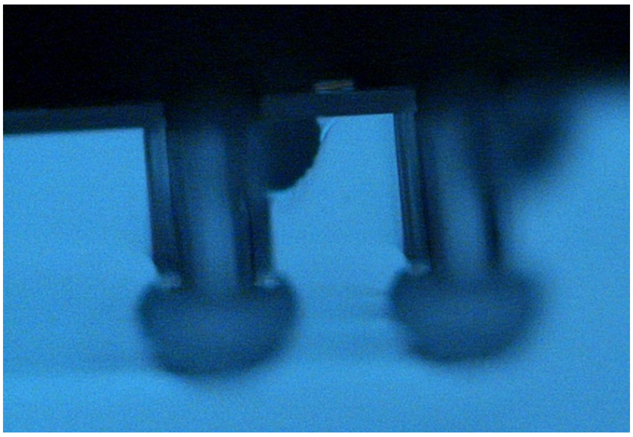
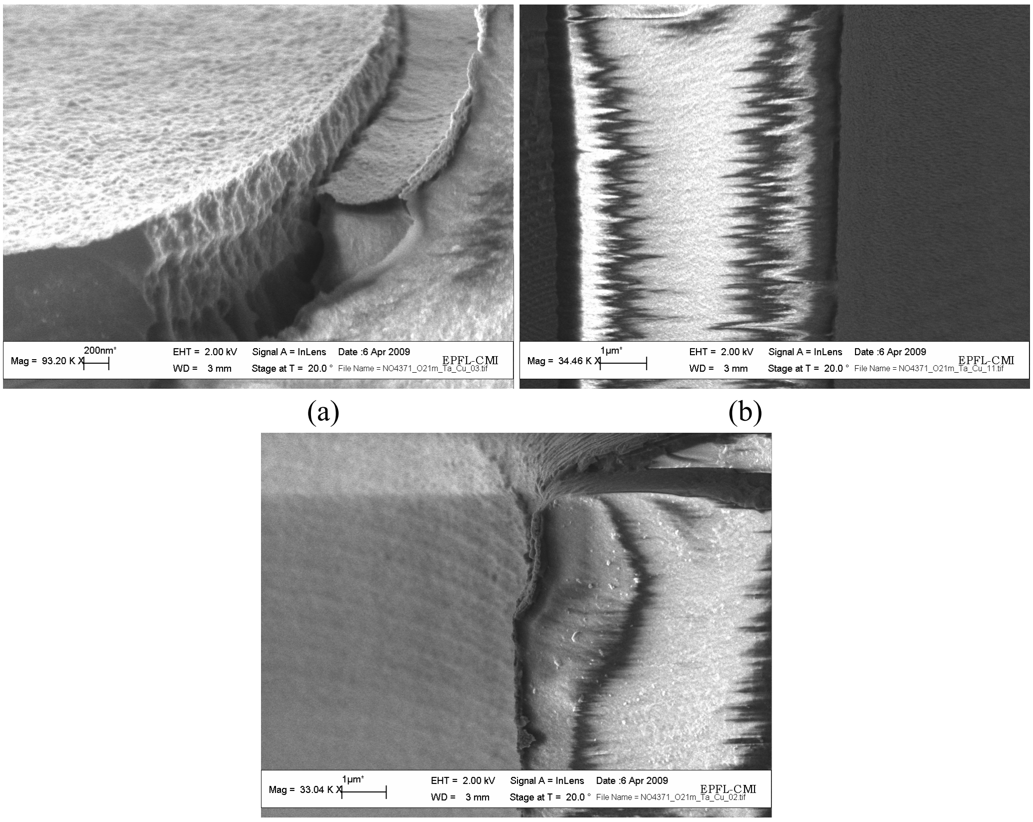


Figure 2.39 Microscopic picture of a failed TSV because of leakage on the top corner during plasma etching.

Shown in Fig. 2.20-4, for real chips, silicon dioxide and pre-metal dielectrics (PMD) are to be etched to expose the metal. In our process, this step is not necessary so far and is ignored.

Metal Layer Sputtering

Layers of 70 nm Ta and 500 nm Cu were sputtered on the wafer surface following the process flow in Fig. 2.20-5. Deposition of the Ta layer improves the adhesion and serves as a diffusion barrier for Cu which is the seed layer for electroplating. The metal sputtering system used is BAS [36]. The SEM image of the TSVs after metal sputtering is already shown in Fig. 2.37-a, and is magnified in Fig. 2.40. The thickness of copper is reduced with the increasing depth. We also used 70 nm Ta and 200 nm Au as the seed layer in some samples for electroplating.


Mag = 93.20 K X
200nm
EHT = 2.00 kV
WD = 3 mm
Signal A = InLens
Stage at T = 20.0 °
Date :6 Apr 2009
EPFL-CMI
Mag = 34.46 K X
1µm
EHT = 2.00 kV
WD = 3 mm
Signal A = InLens
Stage at T = 20.0 °
Date :6 Apr 2009
EPFL-CMI
Mag = 33.04 K X
1µm
EHT = 2.00 kV
WD = 3 mm
Signal A = InLens
Stage at T = 20.0 °
Date :6 Apr 2009
EPFL-CMI

(a) (b)

(c)

Figure 2.40 SEM images of a ring-trench TSV after 60nm Ta and 200 nm Cu sputtering. (a) Bottom part of TSV. (b) Middle part of the TSV. (c) Upper part of TSV. TSV depth 35 μm, diameter 20 μm, AR 1.75 (the maximum AR for copper sputtering > 3 (Figure 2.42)).

Electroplating

An electroplating mold was then fabricated above the Ta/Au seed layer around the TSVs through photolithography using the negative resist RH8023 (Section 2.3.3). Resist residues were found on the sidewall, as well as seed layer damages. Electroless plating of nickel can be processed to enhance metal layer coverage, but was not available. So, for electroplating, we continued without the electroplating mold.

After Ta/Cu seed deposition, a ~10 μm thick layer of copper is electroplated [20] on the whole wafer surface without the electroplating mold. Because of copper plasticity, it is very difficult to get a clear cross section of the TSVs through cleavage. Grinding was performed to get a better view of the cross section, as shown in Fig. 2.41. Fig. 2.42 shows another example without grinding. The electroplated Cu layer is rather grainy and brittle after we tried some replenisher. In both images, copper grew faster on the via top, so with time, voids would be formed inside of the TSV holes. This is a notorious problem for Cu-filled TSVs.

To summarize what has been achieved in this TSV process, most technological steps have been implemented with good results. However, copper electroplating remains an issue. Having a good electroplating will open the way to metal patterning on top of the copper layer. In addition, chemical mechanical polishing after electroplating will help planarize the surface.

Since bottom up electroplating seemed to be a viable approach, we modified this process to the "Step TSV process" (Section 2.4) so as to achieve TSV structures with completely filled metal cores. In the mean time, we further modify this process by using parylene as the filling metal and a layer of sputtered Al as the conducting metal (Section 2.3.5).

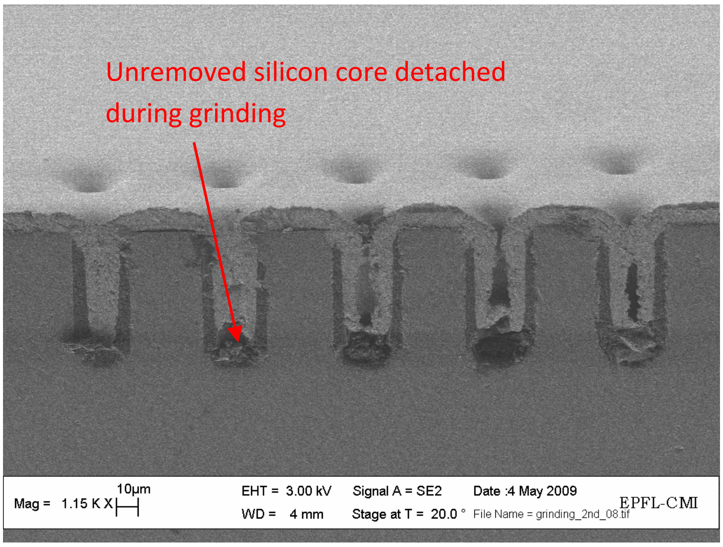

Unremoved silicon core detached
during grinding
Mag = 1.15 K X
10µm
EHT = 3.00 kV
WD = 4 mm
Signal A = SE2
Stage at T = 20.0 °
Date :4 May 2009
File Name = grinding_2nd_08.tif
EPFL-CMI

Figure 2.41 SEM cross section image of ring-trench TSVs after uniform copper electroplating.

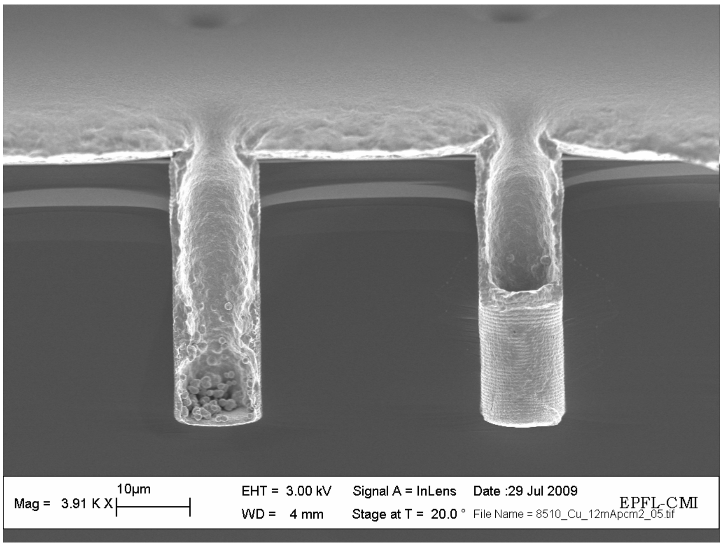

Mag = 3.91 K X
10µm
EHT = 3.00 kV
WD = 4 mm
Signal A = InLens
Stage at T = 20.0 °
Date :29 Jul 2009
File Name = 8510_Cu_12mApcm2_05.tif
EPFL-CMI

Figure 2.42 SEM cross section image of via holes after uniform copper electroplating. Seed layer: 60 nm/500 nm Ta/Cu. AR: ~3.

2.3.5 A Modified Ring-Trench TSV Process

Parylene

Parylene (sometimes known as paralene) is a vacuum deposited plastic film used to coat many types of substrates. The parylene coating process is unique in coating technology and is best described as a vapor deposition polymerization. Without an intermediate liquid phase, the parylene coatings are formed from a gaseous monomer and are created directly on the surface at room temperature. These coatings are completely conformal, of uniform thickness and pinhole free. Parylene completely penetrates narrow spaces, thus it is able to cover sharp edges, points, flat surface, crevices or exposed internal surfaces without voids. So, this makes parylene very promising for ring trenches filling.

Parylene is impermeable to most solvents, and in fact not soluble in all organic solvents up to a temperature of 150 °C [21]. Due to the room temperature deposition process - without the need for subsequent curing - no thermal stresses are induced in the coating and the overall residual stress level is therefore very low. Good comparisons of the Parylenes can be found in [22] which indicate that Parylene C is first with a fairly high dielectric strength, second with decreased CTE (comparing to SiO_2) close to the ones of metals which could minimize thermal mismatch, and third with very low permeability to moisture and corrosive gases which makes it the best barrier among the Parylenes and even superior to almost any other polymeric material.

The Idea of Using Parylene as Isolation and An Al Layer for Conduction in TSV

The main idea in this modified ring-trench TSV process is to use parylene to fill ring trenches. As shown in Fig. 2.36, voids may exist in the filled resist sidewalls. So far, only spin-coated polymers have been studied. Parylene is well-known for its perfect covering uniformity on irregular topographies. So, it makes sense to consider it as an alternative for trench filling.

So far, only Cu was experimented as the conducting metal. As a common metal used for on-chip metallization, Al can also work as an alternative. The process is then simplified as no barrier layer is needed for Al.

Hereafter, the ring-trench TSV process was repeated with the replacers: parylene and aluminum.

Result

SEM images of a ring-trench TSV after ring trench filled with Parylene C and center silicon core removed are shown in Fig. 2.43. This result looks quite promising. Using parylene, TSVs with higher aspect ratios comparing to those filled with spin-coated polymer can be aimed at.

Previously sputtered Cr and Ti have been reported to have good adhesion with Parylene C when a plasma pretreatment of the Parylene surface is included [23, 24, 25]. So, copper is still a viable metal if using Ti as the adhesion and barrier layer. But this is not our aim in this section and we propose to continue with Al sputtering.

In our experiment, a 1 μm Al was sputtered on 2 μm parylene coated on via holes. The resulting SEM images in Fig. 2.44 show how Al thickness varies along the increasing depth. For a 20 μm-diameter via opening, 1 μm Al is at the top (Fig. 2.44-1.1); the Al film thickness reduces to 500 nm reaching a via aspect ratio of 1.5 (Fig. 2.44-2); it further reduces to 230 nm (Fig. 2.44-3), and finally only achieves 170 nm at the bottom reaching a via aspect ratio of 3 (Fig. 2.44-4). The maximum aspect ratio for our Al sputtering process is only about 2 (Fig. 2.44-5). For copper, the aspect ratio is at least 3 (Fig. 2.42).

For Cu and Al deposition in our experiment, dc planar magnetron sputtering were performed. Our results conform to what has been reported by Park *et al.* [26] in 1991, "(comparing to Al) the thickness of metal at the bottom of a hole was greater for Cu and that the sidewall coverage of Cu was continuous".

They attributed this to the near normal cosine distribution of the sputtered Cu (Fig. 2.45-a) which enhances the forward throw of sputtered particles. Sputtered Al showed a preferential emission distribution directed at about 40 ° from the target normal (Fig. 2.45-b) which would promote lateral growth at the top, resulting in void formation within the hole as well as thinner deposits on the bottom of the hole.

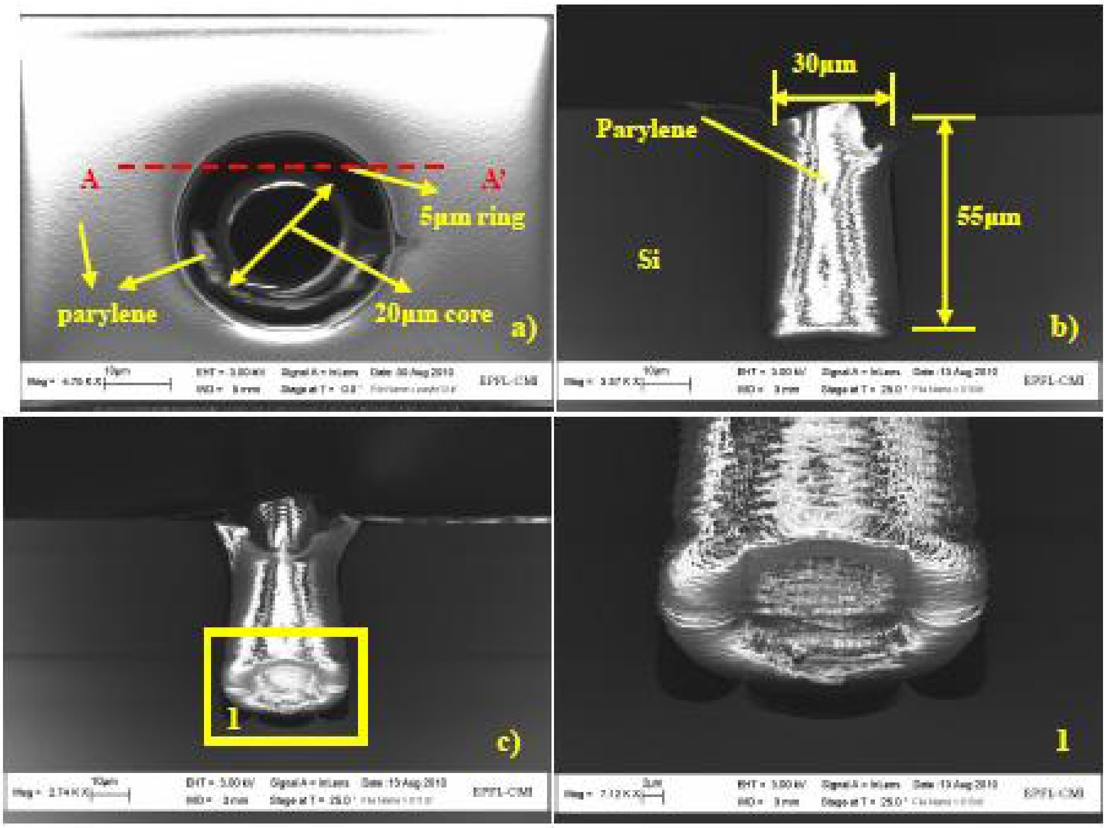


Figure 2.43 SEM images of a ring-trench TSV after ring trench filled with parylene and center silicon core removed.

To summarize the new trials we have done in this section for ring-trench TSV process:

- Parylene is a promising dielectric material for very high AR trench filling;
- Sputtered Al can replace Cu as long as the required TSV AR is less than 2. For TSVs with an AR higher than 2 and lower than 3, sputtered Cu is a proper solution. For TSV with an AR larger than 3, dc magnetron sputtering becomes unsuitable. CVD should be the solution. Improved physical vapor deposition (PVD) methods, e.g. sputter-sputter etch [27] (ion milling) iteration [28] and High Power Impulse Magnetron Sputtering (HPIMS) [29] are options as well. Sputter-sputter etch iteration method asks for the least effort for us.

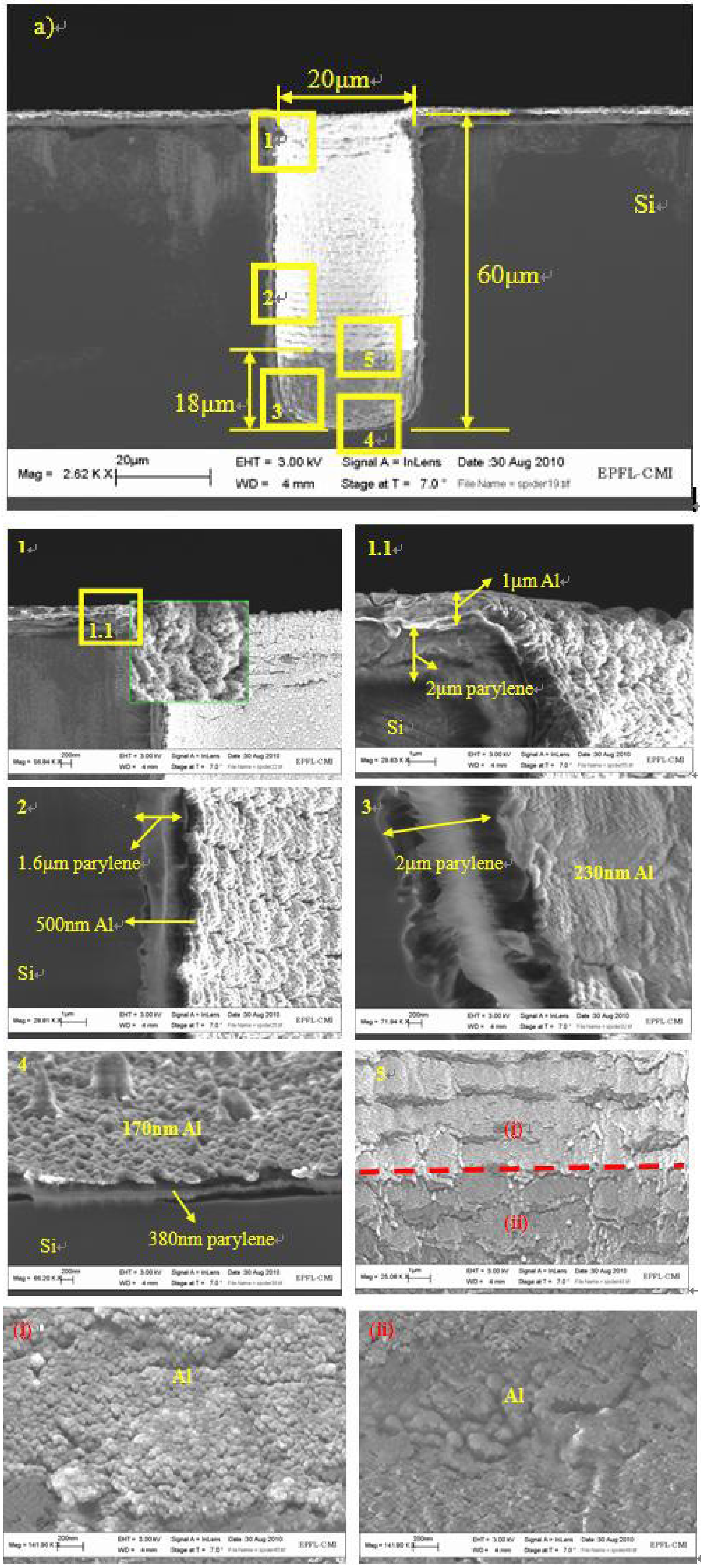

a)
20µm
60µm
18µm
Si
1
2
3
4
5
Mag = 2.62 K X
20µm
EHT = 3.00 kV
WD = 4 mm
Signal A = InLens
Stage at T = 7.0 °
Date :30 Aug 2010
File Name = spider19.tif
EPFL-CMI
1
1.1
1.1
1µm Al
2µm parylene
Si
2
1.6µm parylene
500nm Al
Si
3
2µm parylene
230nm Al
4
170nm Al
380nm parylene
Si
5
(i)
(ii)
(i)
Al
(ii)
Al

Figure 2.44 SEM images showing Al sputtering aspect ratio limitation of about 2.1.

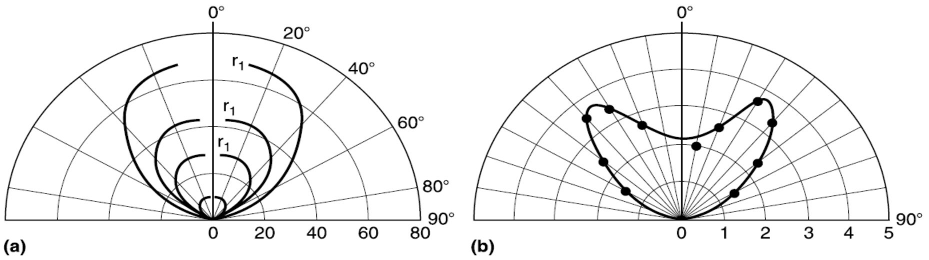


Figure 2.45 Angular distribution of (a) sputtered Cu and (b) sputtered Al [26].

2.4 Step TSV Process: An Alternative Solution

2.4.1 Process Flow

As discussed in previous two TSV processes, landing on metal layer is the first crucial step for backside TSV fabrication. The original idea of this process is to provide Al landing pads for TSVs, to test the bottom-up Cu electroplating and to perform electrical characterizations on TSVs.

The process flow is shown in Fig. 2.46. The process starts from 330 μm-deep KOH silicon etching from the backside of the 380 μm thick <100> wafer using 200nm Si_3N_4 as etching mask (Fig. 2.46-1). After a 2 μm SiO_2 deposition (sputtered), 20 nm/50 nm/2 μm/100nm Ti/Pt/Al/Pt are sputtered on the backside of the wafer (Fig. 2.46-2). Afterwards, the wafer is flipped to allow ring-trench TSVs to be processed in the wafer front side (Fig. 2.46-3 to Fig. 2.46-9).

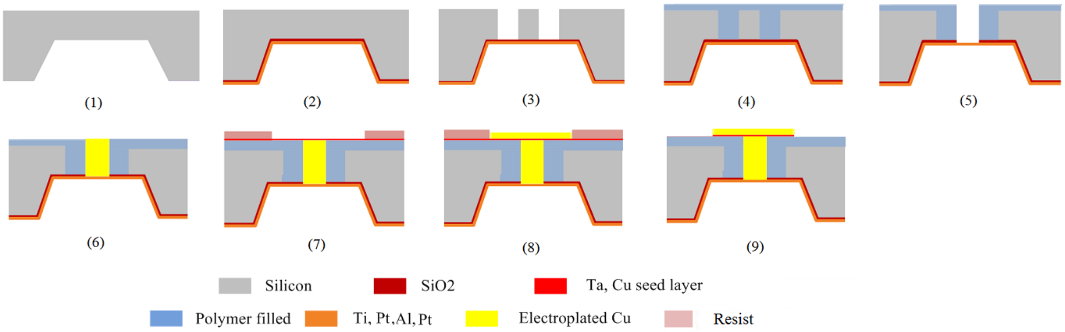


Figure 2.46 Step TSV process.

2.4.2 Step TSV Detailed Process

The first run could not be completed with the expected result. Several problems happened during the process.

First, the wafer became too fragile with the big cavities on the backside (Fig. 2.47).

Second, patterning on high aspect ratio topography requires more complex methods like spray coating. These techniques are not going to be used in this process for simplicity reasons. To allow electrical measurements on TSVs, only two TSVs will be connected using the sputtered metal on the backside of the wafer. The footprint for these TSVs is about 100 μm × 100 μm, meaning that the aspect ratio of the cavities will be around 4. To reach that depth, DRIE with high etching rate of about 4 μm/min only requires 1.5 hours; however, the vertical sidewall poses difficulties on metal coverage using sputtering. KOH in <100> silicon wafers provides a slope of 54.76° and is thus a better choice.

Third, if over-etching occurs during TSV ring trench etch, the metal layer could be damaged and the wafer would not be clamped any more because of He leakage. The etch rate would become intractable because of irregular wafer cooling as well, since He gas flow cools at the wafer backside. A solution is then to glue a carrier wafer below the processed wafer in case of film fracture. A good option is Quickstick 135. Only a few small dips of Quickstick 135 are able to provide strong adhesion between the two wafers during processing. Quickstick is soluble in acetone and can be cleanly thoroughly by iterative rinsing of acetone and isopropanol.

Fourth, to continue TSV process on the front side, backside alignment is required. The alignment markers now are covered by the carrier wafer. Thus, the wafer needs to be released before photolithography. Coaters usually have their own vacuum channel designs in order to hold the wafer in position during wafer transfer or coating. These big cavities cause vacuum leakage and annoying wafer handling difficulty afterwards. The process complexity is greatly increased. In all, step-TSV is not a promising process for TSV fabrication.

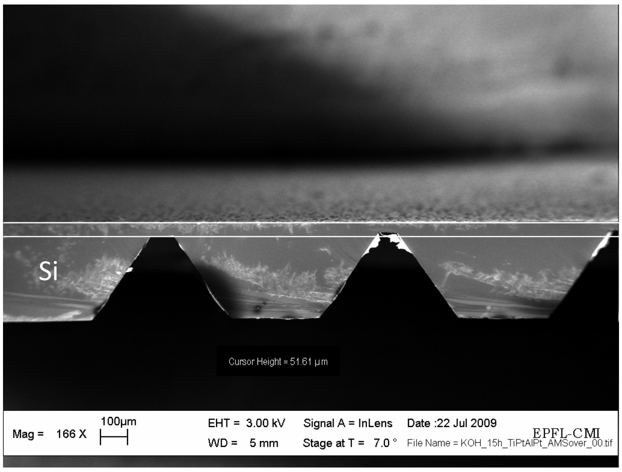


Figure 2.47 350 µm-deep KOH etched cavities after 40 nm/100 nm/2 µm Ti/Pt/Al sputtering.

2.5 Whole-Through TSV Process: A Feasible Solution to TSV Post-Processing on Dies

2.5.1 Introduction

Direct copper electroplating to fill TSV holes completely is a major problem in ring-trench TSV process (Section 2.3). Large voids appear in the copper pillar because copper grows faster around the top corners of TSVs. Bottom-up electroplating from the seed layer deposited on large and deep KOH-etched cavities was not successful as well in step TSV process (Section 2.4). So, small TSV dimensions seems not supported by our current technologies. It happened that big TSVs (at least 200 µm high, over 30 µm wide) were required in another project. So, we went for other approaches. For such large aspect ratio (~7), SiO_2 deposition in via holes, neither sputtering nor PECVD, is not going to get uniform coverage on the sidewalls [30]. Double-side sputtering and parylene deposition are both workable solutions. Actually the latter is more preferred for simple operation (single run required for parylene depostion), good chemical stability and good dielectric isolation (Section 2.3.5). Metal layer sputtering suffers from the same coverage limitations (also discussion in Section 2.3.5).

The newly-setup atomic layer deposition (ALD) process in CMI that can now be used to deposit a thin barrier and adhesion layer was not available at the time being. We assumed that a barrier layer can be deposited. Since through-wafer via holes were used, bottom-up electroplating could be a viable approach. Bottom-up electroplating also does not require wafer-level processing in experimental stages. Then, considering that TSV process needs to be processed on singulated dies, a die thinning process can be developed using the available grinder if the target is 200 µm die thickness. The detailed process is described in Section 2.5.3. A critical issue is to define TSV areas on the dies. To do so, two methods were proposed. The first one was to use the metal (usually Al) on the chip as hard mask to define TSV areas during the chip design phase. This study will be performed in Section 2.5.4. The second one was to apply a stencil mask to define the TSV areas for etching similar to what will be shown in Fig. 2.57. Summing up these studies, the final process is described in Section 2.5.5.

2.5.2 Parylene C on Through-Wafer Holes and Double-Side Sputtering of Al on Parylene

Parylene coating is by nature a double-side process. 8-µm parylene was deposited on the wafer surface and via holes. Afterwards, 2 µm Al layers were sputtered. According to Fig. 2.44 in Section 2.3.5, the maximum aspect ratio that can achieve is about 2, implying that for via holes with 80 µm diameter and aspect ratio of 3, Al will not cover the whole sidewall until the bottom. So, the wafer has to be flipped to allow metal sputtering from the other side of the wafer.

The SEM cross-section view of a whole-through TSV is shown in Fig. 2.48-a. Since Al in the middle of the TSVs (Fig. 2.48-b) sees double sputtering, it presents a metal thickness similar to what is found on the via top parts (Fig. 2.48-c) that only received a single layer.

Good uniformity of Al in the whole via hole sidewall is successfully achieved. Some cracks in the Al film shown in Fig. 2.49 could come from internal stress of Al film, CTE mismatch between Al and parylene, or cleavage due to a greater elasticity of parylene. In further trials, we suggest to deposit a thin layer of Ti below Al as adhesion layer. We also consider modifying the parylene surface using plasma to enhance metal adhesion on parylene [31, 32].

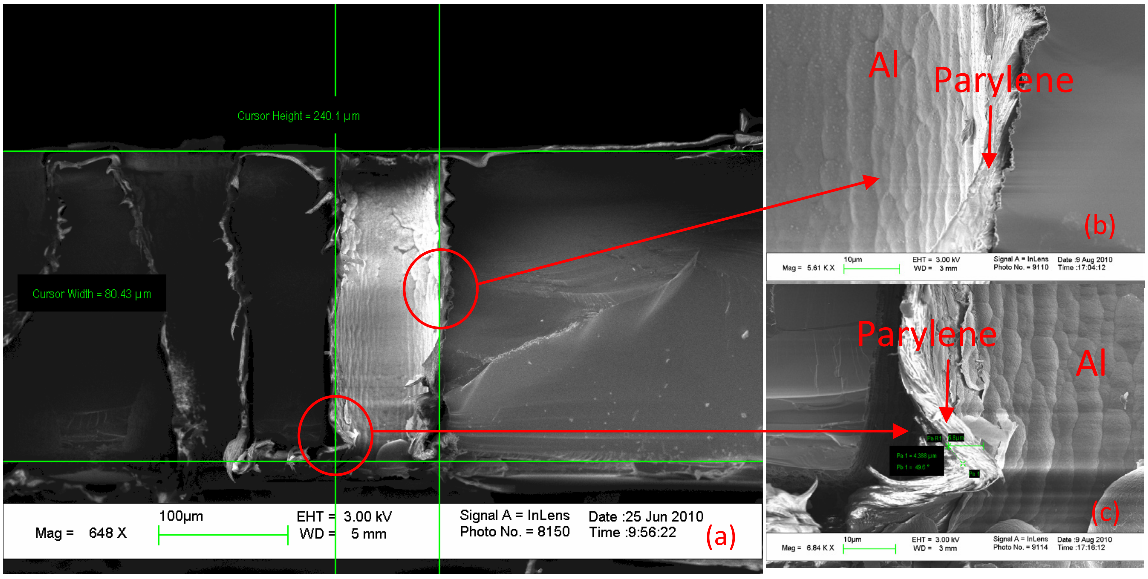

Cursor Height = 240.1 µm
Cursor Width = 80.43 µm
Mag = 648 X
100µm
EHT = 3.00 kV
WD = 5 mm
Signal A = InLens
Photo No. = 8150
Date :25 Jun 2010
Time :9:56:22
(a)
Al
Parylene
(b)
Mag = 5.61 K X
10µm
EHT = 3.00 kV
WD = 3 mm
Signal A = InLens
Photo No. = 9110
Date :9 Aug 2010
Time :17:04:12
Parylene
Al
Pa R1
6.8µm
Pa 1 = 4.388 µm
Pb 1 = 49.6 °
Pa 1
(c)
Mag = 6.84 K X
10µm
EHT = 3.00 kV
WD = 3 mm
Signal A = InLens
Photo No. = 9114
Date :9 Aug 2010
Time :17:16:12

Figure 2.48 SEM cross-section views of a through-wafer TSV hole after 10 µm (8 µm achieved in reality) parylene deposition and double-side 200 nm/2 µm Ti/Al sputtering. (a) The whole via hole. (b) The middle part of the via hole. (c) The bottom of the via hole.

2.5.3 Die Thinning

Thinning experiments have been done at chip level. Small silicon parts cleaved from a wafer were thinned through grinding and polishing. TSV holes are already etched in the silicon. The grinding and polishing machine is composed of a rotating plate (Fig. 2.50-a), a moving arm (Fig. 2.50-b) and a cylindrical mechanical system (Fig. 2.50-c).

The sample was held in the center part with a polishing resist and pressed on the rotating plate with an adjustable force. The sample was thinned and polished using different grain sizes of silicon carbide paper grains. Water was used as lubricant.

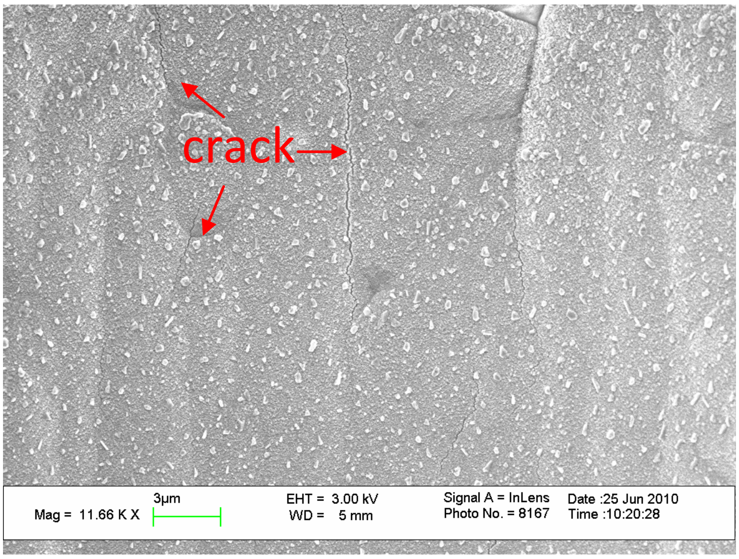
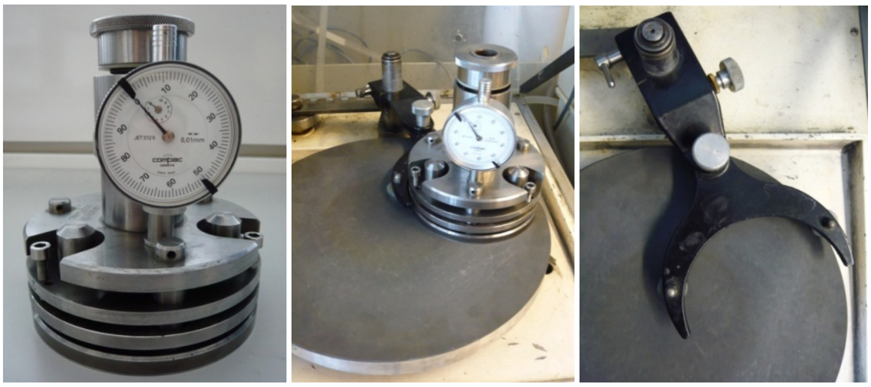

crack
3µm
Mag = 11.66 K X
EHT = 3.00 kV
WD = 5 mm
Signal A = InLens
Photo No. = 8167
Date :25 Jun 2010
Time :10:20:28

Figure 2.49 SEM image showing cracks in Al film sputtered on parylene.



(a) (b) (c)

Figure 2.50 Thinning equipments. (a) Rotating plate with a moving arm. (b) Mechanical holder. (c) Complete system.

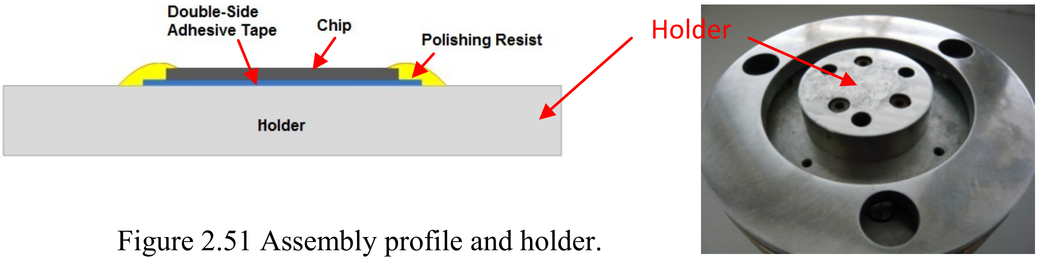

Double-Side Adhesive Tape
Chip
Polishing Resist
Holder
Holder

Figure 2.51 Assembly profile and holder.

To hold the sample, a solid polishing resist was used after being softened at higher temperature. The metallic holder was heated on a hotplate (set at 150 °C). A first test was done by gluing the sample on the holder with the polishing resist only, but the resist suffered from significant erosion due to water lubrication and stress during grinding. The sample was laid on a double-side adhesive tape to protect the front-side TSV holes; then, a resist frame was set to protect the sample edges. The profile view of the assembly with the holder is shown in Fig. 2.51.

Thinning speed depends on grain size, pressing force and table rotation speed. The maximum rate used was 4.5 µm/min. Thicknesses was measured throughout the grinding and polishing process on five different points of the sample (center and corners) with a digital dial gauge. Samples with original thickness of 525 µm were successfully thinned down to 100 µm to expose the TSV holes. The resulting thin, shiny and reflecting silicon pieces are shown in Fig. 2.52.

The 100 µm-diameter opening (Fig. 2.53-a) shows less fractures than the 50 µm-diameter ones (Fig. 2.53-b). Smaller via holes require more delicate grinding sequence. In the real cases, if it is for via-middle processes, thinning will be done when different materials presenting. A typical sequence is: grinding to obtain a depth close to the target thickness, polishing and smoothening the rough surface and further smoothening of the surface and exposure of Cu pillar TSVs with plasma dry etch. The chip strength can be improved by 55% using plasma etching [33].

In our experiment, the chips were thinned even down to 50 µm. However they became very fragile. A more delicate gluing and releasing method would be required. Since in this thesis the manipulation of ultra-thin chips is not the aim, basic performance of the grinding machine will not the only thinning solution. The potential application of such a chip thinning process for our TSV process and 3D stacking schemes will be exposed later on in Chapters 3 and 4.

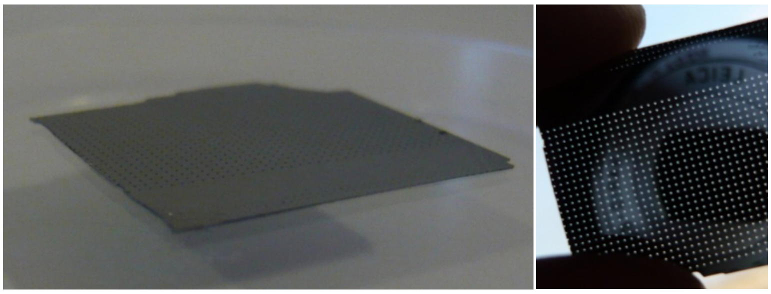
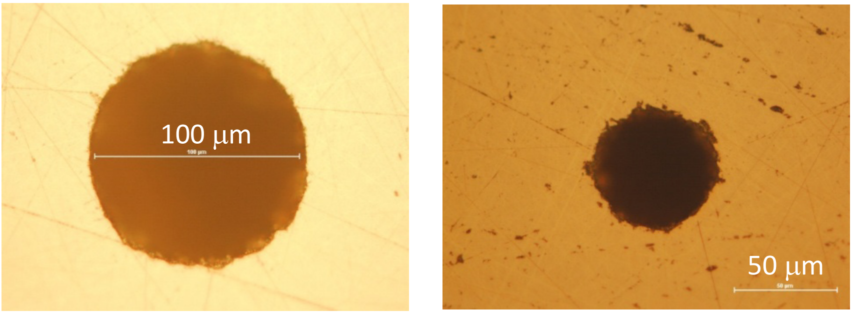


Figure 2.52 100 µm thick thinned Si component with TSV holes.



(a) (b)

Figure 2.53 100 µm deep TSV holes dry etched in silicon and opened on the other side by thinning. TSV diameter: (a) 100 µm; (b) 50 µm.

2.5.4 BEOL Metal (Al) as Dry Etch Hard Mask

This section discusses an original "self-aligned" mask technique proposed for TSVs.

After sticking the chip to a carrier wafer using a tiny dip of Quickstick 135, dielectrics on the chip (including the passivation layer which is usually silicon nitride or silicon dioxide and on-chip dielectric which is usually silicon dioxide) were etched down to MET2 using Adixen 200 etcher in RF mode. Recipe "SiO2_PR_2:1" gives an SiO_2 etch rate of 0.27 µm/min with a selectivity SiO2:PR = 2:1. It works on silicon nitride as well.

The CMOS chip used in this experiment was fabricated in 0.35 µm technology. Each layer's height is approximately 1.5 µm. Thus for simple estimation, to expose one metal layer should take about 6 minutes.

The sample shown in Fig. 2.54 has undergone SiO_2 etching for 13 minutes, with three metal layers clearly exposed. For confidentiality reason, the dimension of the structure is not to be disclosed in Fig. 2.54. The profile of some exposed Al interconnection structures is measured by VEECO optical profile and shown in Fig. 2.55. The height difference between two metal layers is about 1.7 µm which matches our expectation which is 1.5 µm.

Since Al pads could detach from the chip and expose the covered via multi-pillar structure to connect to another metal layer, these Al pads could be the most delicate parts in this process (Fig. 2.56).

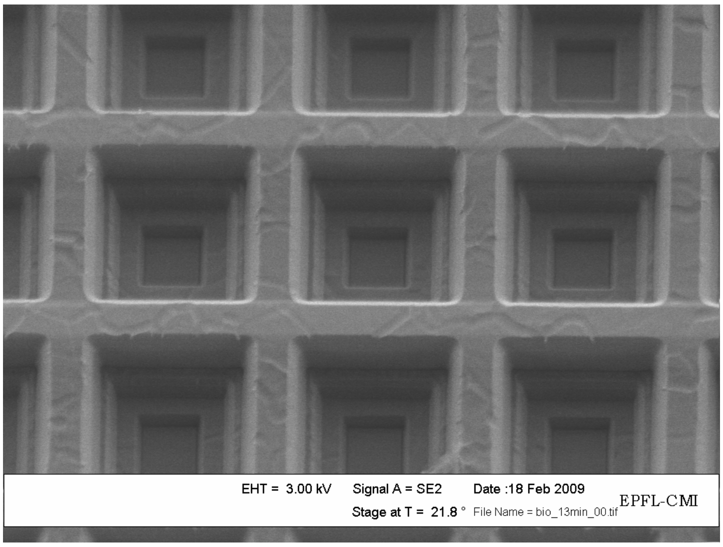


Figure 2.54 SEM image of exposed Al structures after passivation layer etching.

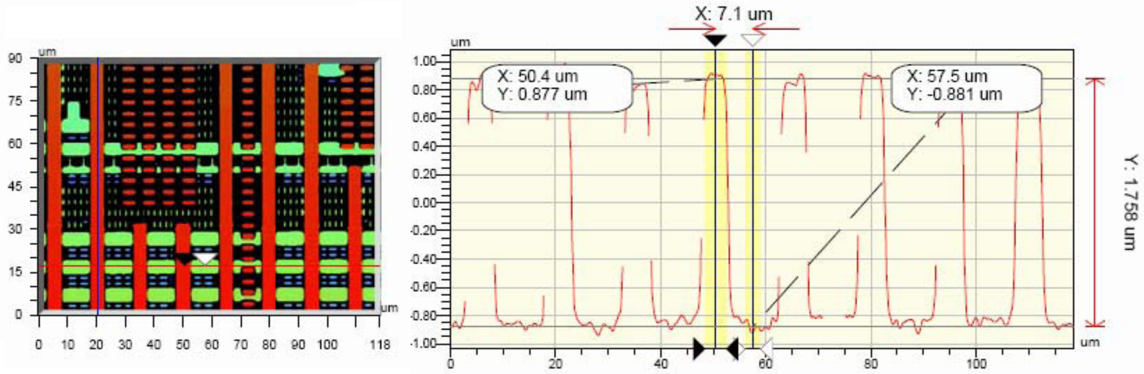


Figure 2.55 Profile images of some metal line structures in different layers measured by VEECO optical profiler.

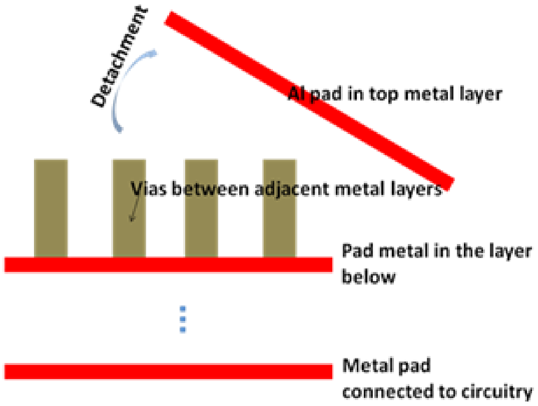

Detachment
Al pad in top metal layer
Vias between adjacent metal layers
Pad metal in the layer below
Metal pad connected to circuitry

Figure 2.56 Illustration of Al pad metal detachment during SiO_2 etching.

From our results, it can be concluded that using a metal layer (Al) already available on the chip in order to define self aligned TSVs and protect the circuit during etching could be quite interesting and new for TSV post-processing on diced chips.

Whether it will violate chip design rule needs to be examined. Usually, slotting (also known as "cheesing") and metal filling are to be introduced on the chip. Slotting involves punching holes in all large metal areas, and metal filling ensures that a certain range of the chip area is covered with metal. So, a whole piece of metal seems not allowed. But if the metal is non-functional, it does not have to satisfy design rules, which complicates the design phase.

As mentioned in Chapter 1, front side chip post-processing affects yield. In some biosensor applications, the idea of post-processing on the chip front side to expose the buried metal structures to achieve sensor arrays with higher density has been proposed and demonstrated [34]. Because this post-processing is at wafer level, the pad area can still be protected from plasma etching in advance using photolithography, so this difficulty about the pad damage is not confronted. A stencil mask technique could work well to protect the pad and further to define the TSV patterns (Fig. 2.57).

Anyway, this process can only work for chips with relatively simple circuit designs and functions, such as sensors. We do not expect it to work for delicate CPU chips.

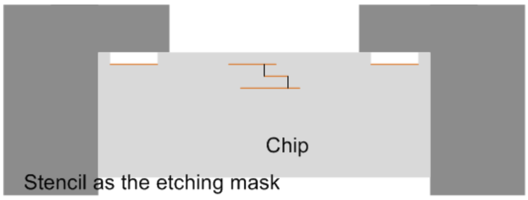


Figure 2.57 Stencil as etching mask to protect areas unwanted to be exposed to plasma etch.

2.5.5 Whole-Through TSV Detailed Process

A simplified TSV process flow is shown in Fig. 2.58. A die can be post-processed with through-wafer TSV holes using the technique we illustrated in Figure 2.57. Then parylene is to be deposited on the whole die surface. The die is then placed on a wafer with sputtered seed layer of Ta/Cu (80 nm/200 nm), followed by bottom-up electroplating. After some trials of bottom-up electroplating, the development of this process was carried on by other colleagues. Details will be disclosed here.

Since this approach was initiated by myself, I write some notes about the initial bottom-up electroplating trials. The adhesion between the dielectrics and copper is very important.

Without the adhesion layer deposited, some copper pillars detached from the holes after die separation.

A final result of this process is presented here to complete our TSV development with a completely-filled TSV structure (Fig. 2.60). A bottom-up filled whole-through Cu TSV has been fabricated in a singulated dummy silicon die. The top surface after electroplating is rather rough. CMP can be processed afterwards to achieve planar surface in the expectation of die-level stacking processes.

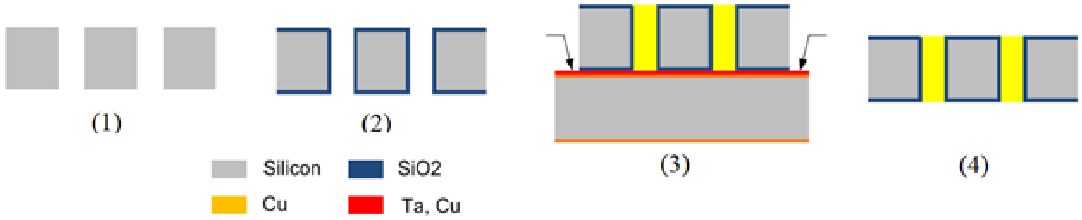


Figure 2.58 Whole-through TSV process.

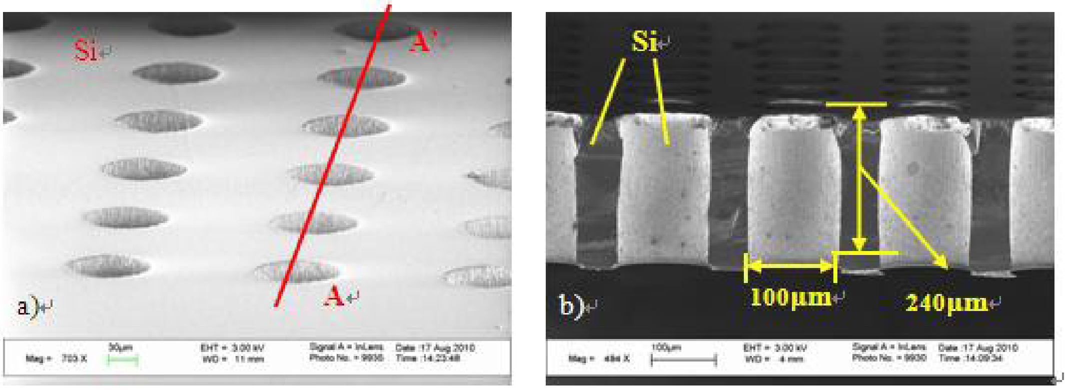


Figure 2.59 SEM images of through wafer Si etching. (a) Top view. b) Cross-section A-A').

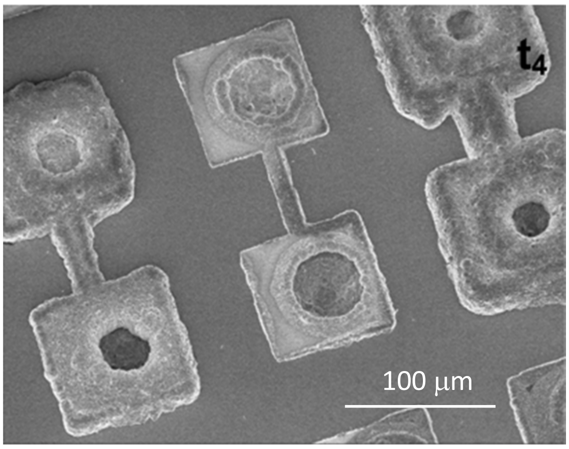


Figure 2.60 SEM image of whole-through TSVs after bottom-up electroplating.

2.6 Summary

Four TSV processes have been developed in this chapter. Ring-trench TSV process has demonstrated a state-of-the-art TSV technology where TSV dimensions are relatively "small" (20 µm in diameter, 50 µm in height). The whole-through TSV process, although could only be relatively big (100 µm in diameter, 380 µm in height), is our first die with post-processed TSVs. Some comments on these processes are summarized in Table 2.6.

Table 2.6 Benchmarking the TSV Processes Developed.

TSV processes	Achievements and Advantages	Problems and Disadvantages
Core-Hole TSV	1. It is a straightforward and widely adopted TSV process. 2. SiO_2 is the best dielectric material for the isolation. 3. A spray coating technique can be developed with further experiments. 4. Two tapered silicon etching processes are developed.	1. Patterning the SiO_2 on the via bottom to open metal connection is challenging. 2. Going for higher AR, sputtering will not satisfy the need. 3. Electroplating to completely fill the vias will be a problem. 4. It is with a poor scalability. The process needs to be adjusted to each TSV dimension. 5. The "copper pumping" problem might crack the SiO_2 around the via top.
Ring-Trench TSV	1. It is with a good scalability. 2. Polymer as dielectric work as stress reliever which increases reliability. 3. High aspect ratio (>10:1) and small diameter (10 μm) TSVs can be expected. 4. High AR (upto 20:1) trenches are filled with polymers by spin coating. 5. An electroplating mold fabrication process is developed.	1. Polymer is of lower dielectric permittivity than SiO_2. 2. Completely filling the vias with copper electroplating is the main difficulty, and still has to be solved. 3. The aspect ratio of TSVs is limited by the dc planar magnetron sputtering technique we use.
Step TSV	1. Experience on double-side wafer processing was gained. 2. It provides some initial information about DRIE landing on metal which is an important step in BEOL TSV fabrication.	1. The wafer becomes too fragile for processing. 2. The process became rather complicated when a carrier wafer is required to hold the sample wafer. 3. It is not a useful TSV process.

Continuing the table in the previous page.		
Whole-Through TSV	1. It demonstrates the first dummy die post-processed with TSVs in EPFL. 2. Bottom-up copper electroplating was done to achieve complete filling. 3. With whole-through TSVs, no metal landing is required. This can avoid less device failure.	1. The alignment accuracy is still to be improved. 2. A planarization solution is to be developed. 3. The TSV dimension is still limited. 4. The adhesion of the copper core to the sidewall still needs to be improved. Without a proper solution, "copper pumping" problem will easily happen.

Reference

1. CMI, http://cmi.epfl.ch.
2. F. Lärmer and A. Schilp, "Method of Anisotropically Etching Silicon," *US Patent: US5501893*, March 26 1996. Applicant: Robert Bosch GmbH, Stuttgart, Germany.
3. AMS 200 DSE, http://cmi.epfl.ch/etch/AMS200.php.
4. S. Burkett, L. Schaper, T. Rowbotham, J. Patel, T. Lam, Abhulimen U., Boyt D. D., Gordon M., Cai T., "Material Aspects to Consider in the Fabrication of Through-Silicon Vias," *Mater. Res. Soc. Symp. Proc.* 970 2007
5. R. Nagarajan, L. Ebin, L. Dayong, S. C. Seng, K. Prasad and N. Balasubramanian, "Development of a Novel Deep Silicon Tapered Via Etch Process for Through-Silicon Interconnection in 3-D Integrated Systems," *56th Electronic Components and Technology Conference*, pp. 383 - 387, 2006.
6. R. Li, Y. Lamy, W. F. A. Besling, F. Roozeboom and P. M. Sarro, "Continuous Deep Reactive Ion Etching of Tapered Via Holes for Three-Dimensional Integration," *Journal of Micromechanics Microengineering*, Vol. 18, No. 12, 2008.
7. N. P. Pham, J. N. Burghartz, and P. M. Sarro, "Spray Coating of Photoresist for Pattern Transfer On High Topology Surfaces," *Journal of Micromechanics Microengineering*, Vol. 15, pp. 691-697, 2005.
8. L. Yu, Y. Y. Lee, F. E. H. Tay, C. Iliescu, "Spray Coating of Photoresist for 3D Microstructures with Different Geometries," *Journal of Physics: Conference Series 34*, pp. 937-942, 2006.
9. Spider 600, http://cmi.epfl.ch/thinfilms/Spider600.php.
10. Remover 1165, http://cmi.epfl.ch/organisation/materials.php#Produits%20chimiques.
11. Suss MA6 mask and wafer aligner, http://cmi.epfl.ch/photo/MA6_en.php.
12. Y. Civale, D. S. Tezcan, H. G. G. Philipsen, F. F. C. Duval, P. Jaenen, Y. Travaly, P. Soussan, B. Swinnen, and E. Beyne, "3-D Wafer-Level Packaging Die Stacking Using Spin-on-Dielectric Polymer Liner Through-Silicon Vias," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, Vol. 1, No. 6, June 2011.
13. http://www.imaps.org/abstracts/system/new/abstract_preview.asp?abstract=10imaps092
14. T. Duc, G. Lau, and P. M. Sarro, "Polymeric Thermal Microactuator with Embedded Silicon Skeleton: Part II–Fabrication, Characterization, and Application for 2-DOF Microgripper," *J. Microelectromech. Syst.*, vol. 17, no. 4, pp. 823–831, Aug. 2008.

15. H. E. M. Kotb, K. Isoird, F. Morancho, L. Théolier, and T. D. Conto, "Filling of Very Deep, Wide Trenches by Benzocyclobutene Polymer," *Microsyst. Technol.*, vol. 15, no. 9, pp. 1395–1400, 2009.
16. F. F. C. Duval, C. Okoro, Y. Civale, P. Soussan, and E. Beyne, "Polymer Filling of Silicon Trenches for 3-D Through Silicon vias Applications," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 2011.
17. Differential Scanning Calorimetry, http://en.wikipedia.org/wiki/Differential_scanning_calorimetry#DSC_curves.
18. AZ Barli II, http://www.microchemicals.eu/technical_information/anti_reflective_coating_photoresist.pdf
19. G. Teowee, K. C. McCarthy, T. J. Bukowski, T. P. Alexander, S. Motakefr and D. R. Uhlmann, "Optical Properties of Au and Pt-Doped PLT 28 Films," *Proceedings of The 10th IEEE International Symposium On Applications Of Ferroelectrics Volume*, 1996.
20. Electroplating of Metal with Desaules Wet Bench, http://cmi.epfl.ch/thinfilms/Galvano_intro.php.
21. M. Heschel, "Multiple Through-Wafer Interconnects for Stacking of Microelectromechanical Devices," PhD thesis, Mikroelektronik Centret (MIC), Technical University of Denmark, June, 1999.
22. A. D. Yalçinkaya, "Micromechanical Resonators for Low-Power, Low-Voltage Systems," PhD thesis, Mikroelektronik Centret (MIC), Technical University of Denmark, June, 2003.
23. J. Gobet, J.-P. Thiebaud, F. Crevoisier, and J.-M. Moret, "IC Compatible Fabrication of Through-Wafer Conductive Vias," *Proceedings of the SPIE - The International Society for Optical Engineering*, Vol. 3223, 1997.
24. T. N. Pornsin-Sirirak, Y.-C. Tai, H. Nassef, and C. M. Ho, "Flexible Parylene Actuator for Micro Adaptive Flow Control," *The 14th IEEE International Conference on Micro Electro Mechanical Systems, MEMS 2001*, pages 511–514, 2001.
25. M. Liger, N. Pornsin-Sirirak, and Y.-C. Tai, "Large-Area Electrostatic-Valved Skins for Adaptive Flow Control on Ornithopter Wings," *Solid-State Sensor and Actuator Workshop*, pages 247–250, Hilton Head Island, SC, USA, June 2–6, 2002.
26. Y. H. Park, A. H. Chung and M. A. Ward, "Step Coverage Evaluation Of Copper Films Prepared By Magnetron Sputtering," *Proceedings of Eighth International IEEE VLSI Multilevel Interconnection Conference*, pp.295 - 297, 1991.
27. G. C. Schwartz, K. V. Srikrishnan, and A. Bross, editors. Handbook of Semiconductor Interconnection Technology, Chapter V: Metalization (by G.C. Schwartz), page 313. Marcel Dekker, Inc., 1998.
28. J. Weichart, M. Elghazzali, S. Kadlec, and A. P. Ehiasarian, "PVD Processes in High Aspect Ratio Features by HIPIMS," *Proceedings of the 52nd Annual Technical Conference of Society of Vacuum Coaters*, Albuquerque, NM 87122-1914, United States, May 9-14, 2009 Santa Clara, (CA), 201 (2009).
29. H.-J. Wu, R. Shaviv, M. Sriram, W. Wu, A. Pradhan, K. J. Park, J. O'loughlin, K. Chattopadhyay, T. Mountsier and G. Dixit, "Improving Interconnect Reliability via Optimized Barrier/Seed," *Semiconductor International*, Vol. 31, No. 11, Oct. 1, 2008.
30. N. Ranganathan, L. Ebin, L. Linn, L. W. Sheng Vincent, O. K. Navas, V. Kripesh and N. Balasubramanian, "Integration of High Aspect Ratio Tapered Silicon Via for Through-

Silicon Interconnection," *58th Electronic Components and Technology Conference*, pp. 859-865, 2008.

31. Q. Yu, J. Deffeyes, and H. Yasuda, "Engineering the Surface and Interface of Parylene C Coatings by Low-Temperature Plasmas," *Progress in Organic coatings*, Vol. 41, Issue: 4, pp. 247-253, 2001.
32. T. N. Pornsin-Sirirak, Y.-C. Tai, H. Nassef, and C. M. Ho, "Flexible Parylene Actuator for Micro Adaptive Flow Control," *The 14th IEEE International Conference on Micro Electro Mechanical Systems, MEMS 2001*, pp. 511-514, 2001.
33. S. Chen, T.-Y. Kuo, H.-T. Hu, J.-R. Lin, and S.-P. Yu, "The Evaluation of Wafer Thinning and Singulating Processes to Enhance Chip Strength," *55th Electronic Components and Technology Conference*, 2005.
34. L. Berdondini, P. D. van der Wal, N. F. de Rooij, and M. Koudelka-Hep, "Development of an Electroless Post-Processing Technique for Depositing Gold as Electrode Material on CMOS Devices," *Sensors and Actuators B*, Vol. 99, pp. 505-510, 2004.
35. STS Multiplex ICP dry etcher, http://cmi.epfl.ch/etch/STS.php.
36. Balzers BAS 450, sputter, single chamber, multi-target system, http://cmi.epfl.ch/thinfilms/Bas450.php.

Chapter 3

Alignment and Stacking Techniques for Die-Level 3D Integration

After discussing what are the main limitations of current solutions for die-level stacking, we will introduce four self-alignment (SA) techniques that have been presented in previous publications and explore two of them that have been part of our 3D stacking research activity, namely, gravitational SA and magnetic SA. Another one called "hydrophobic SA" will be addressed in Chapter 4.

3.1 Feasibility of Die-Level Processing and Stacking

To stack individual dies into a 3D chip, the foremost and most important considerations are:

- Thermal budget
- Die dimension in each layer
- Number of layers to be stacked and the stacking scheme (face-to-back or face-to-face)
- TSV dimension
- Handling method of single dies and placement precision

The post-processing 3D scheme requires a thermal budget below 450 °C to avoid damages to the devices and structures already integrated on the chips. The 2nd to 4th considerations should be decided by the targeted application.

Basically, our aim is to implement a two-layer multi-processor stack. Two by two LEON [1] cores are assumed to be placed on each layer of the stack, communicating through inter-layer serial or quasi-serial links which will be detailed in Chapter 5. Circuits with the link designs are to be fabricated in 90 nm technology and then diced into dies of about 5 mm × 5 mm. The ring-trench TSVs were to be processed from the chip backside. The last consideration will be the focus of this chapter. The experimental aspects of die attachment to the carrier wafer will be discussed in Section 3.1.1 while an in-depth discussion on high precision alignment methods will be addressed in the rest part of this chapter.

3.1.1 Materials for Temporary Die Attachment

To attach single dies to a carrier wafer, three materials have been experimented along this work:

- Temporary mounting wax Quickstick 135 [2]
- Double-side adhesive tapes such as Revalpha (from Nitto Denko) [3].
- Parylene

Quickstick 135

Quickstick 135 is a transparent thermoplastic material available in solid sticks. This material offers the advantages of becoming fluid above 71 °C and being chemically stable during processing up to 300 to 400 °C. Quickstick can be dissolved into acetone and then be spun over a wafer surface [4] for temporary bonding. In our die-level process, the carrier wafer (silicon or pyrex) is clamped on a spin coater with a temperature increasing up to 135 °C where some Quickstick is spawn on the bonding sites. Next, a certain force is applied on top of the chip to get a thinner and uniform adhesive layer between the chip and the carrier. Finally, the carrier is released from the vacuum holder and cooled down. After processing, Quickstick can be completely removed in acetone.

One problem inherent to this method could be the non-uniformity in wax thickness which in turn can generate some pattern distortion due to inhomogeneous exposure. In addition, during dry etching, a thick layer of Quickstick will hurdle the heat accumulated on the die to dissipate through the backside carrier wafer, and thus might influence the etching results as well. However, a very thin layer of Quickstick is usually sufficient to hold the chip and avoid such problems.

Revalpha Double-Side Thermal Release Adhesive Tape

Revalpha double-side thermal release adhesive tape is also a good candidate and easy to use. The major limitation of a tape is its incompatibility with further processing steps. Originally, the tape is used for dicing. Therefore, the tape is generally incompatible with high temperature and aggressive chemistry. The releasing temperature is 200 °C and it could deteriorate and leave residues if the temperature reaches 300 °C. In addition, the 100 μm-thick tape might bring unpredicted etching results due to poor heat dissipation on the chip backside.

Parylene

Parylene can be used for temporary mounting as well due to its high uniform and conformal coating ability and good chemical stability. But an additional step of O_2 plasma etching is then required to remove it. The adhesion strength is relatively weak compared to Quickstick and Revalpha tapes, but we found it good enough for dies manipulation.

3.1.2 Precise Die Attachment/Stacking

The placement precision achieved by hand is indeed unsatisfying. Modern pick-and-place tools operate at very high speeds to achieve reasonable process throughput (> 10,000 placements per hour), but this comes at the expense of placement accuracy. Even while operating at slow speeds, many pick-and-place operations will not achieve chip placement accuracies better than +/- 20 μm [5]. To fabricate 20 μm-diameter TSVs, die misalignment cannot exceed a few microns [25]. Flip chip bonders achieve higher accuracies. For example, SUSS FC250 claims bonding accuracies to 1 micron [6]. However, these machines were not available at CMI.

Using transparent pyrex wafer as the carrier, backside alignment could be an interesting approach to align the flipped dies to some desired positions. Pyrex wafer is the most adopted carrier in 3D integration: it has a CTE close to that of silicon and its transparency allows backside inspection. In practice, cameras to take backside images for alignment references need to be adjusted to focus on die surface that has been raised by the carrier and the glue. Since this modification cannot be done at CMI, backside alignment through pyrex carrier wafer is not a viable approach.

Despite these constraints, other technical problems exist concerning the limited opened regions for alignment inspection in the wafer chuck of the MA6 aligner [7]. Only the patterns in two blank regions of the wafer can be observed during backside alignment (Fig. 3.1), meaning that dies need to be located in these two regions in order to be aligned.

With such imbalanced topography, many problems will happen during etching, CMP and lithography. One solution could be placing multiple dummy dies to reach a balanced pattern distribution on the wafer surface. Aside from these technical problems, it can be foreseen that such a process will be of high cost and low yield.

Since none of the above solutions seems to be interesting, developing die (temporary) attachment techniques with self-alignment ability becomes mandatory. Also, to successfully achieve the 3D-stacked chip, we must develop novel methods that lead to high quality permanent bonding (with satisfying alignment precision). Our "home-made" SA techniques should work both for temporary bonding and for permanent bonding, still with a minimum modification of the basic set up.

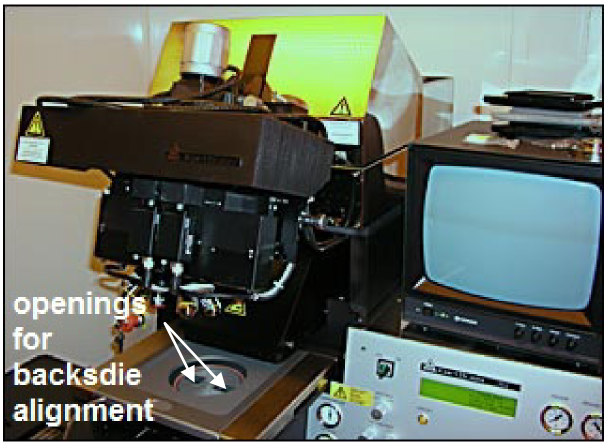


Figure 3.1 Karl Suss MA6 aligner.

3.2 Picturing SA Techniques for 3D

Four major SA techniques named gravitational SA [8, 9, 10], electrostatic force SA [11], magnetic SA [12, 13, 14, 15, 23] and hydrophobic SA [Chapter 4-2, -3, -4, -5, -6, -7, -8], have been studied before. In the rest of this section, we will analyze the feasibility of using these techniques for 3D stacking.

3.2.1 Gravitational SA

Gravitational SA has been originally proposed as an easy, direct and cheap method for heterogeneous system integrations [8, 9, 10]. As reported in [10], with a dicing inaccuracy of 25 μm, the pitch size and pad size have to be beyond 50 μm. Therefore, the SA precision is the main limitation of gravitational SA techniques, as illustrated in Fig. 3.2. Higher accuracy can only be gained by adopting more sophisticated and more expensive dicing methods, e.g. DRIE and laser ablation. For 3D systems with limited numbers of large TSVs, this could be a viable approach. In Section 3.3, a more detailed study will be performed.

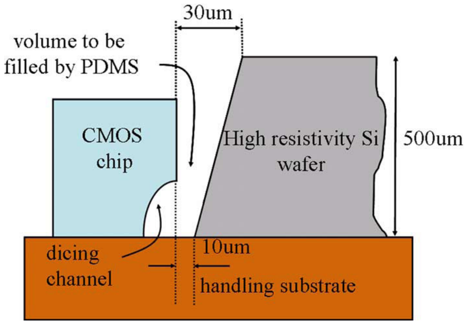


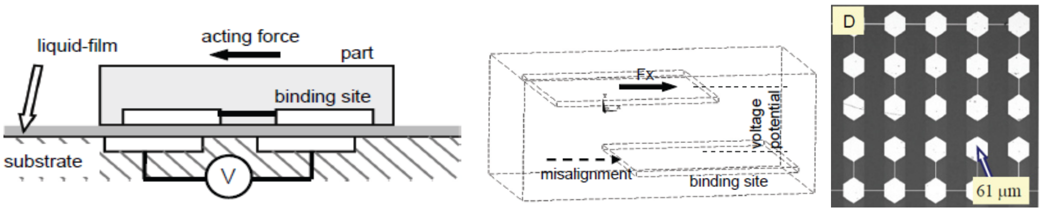
Figure 3.2 Gravitational SA precision limitation [10].

3.2.2 Electrostatic SA

Self-alignment of micro-structures can also be accomplished with electrostatic attraction force. As proposed by Dalin *et al.* [11] in 2009, parts with microstructures shown in Fig. 3.3-c can be successfully aligned with a constantly applied voltage of 50 V. The concept of electrostatic SA is shown in Fig. 3.3-a while the finite-element (FE) model for computations is shown in Fig. 3.3-b. Only a small area occupied by the microstructures (about 2% of the total die area) can generate a large alignment force to align the chip.

Using this technique in 3D integration has some constraints that limit its practical application.

- Circuit designers have to insert these microstructures as floating metal structures into the layout before sending the final design to the foundry. First, these metal structures violate design rules although they still might be tolerated. Second, circuit designers usually don't work with concepts of microstructures.
- Connecting these microstructures in a 3D stack with relatively high voltages is not compatible with low voltage digital design which can suffer from electrostatic discharges.
- These nonfunctional microstructures will permanently stay in the 3D stack, which creates process difficulties such as bonding and metal redistribution.



(a) (b) (c)

Figure 3.3 Electrostatic SA proposed by Dalin *et al.* [11]. (a) Concept. (b) Detail of a finite element model for computations of the electrostatic force. (c) An example of a fabricated test structure on substrate for self-assembly experiments with electrostatic attraction.

3.2.3 Magnetic SA

Magnetic SA has been proposed as an alternative for monolithic heterogeneous integration of compound semiconductor devices with silicon integrated circuits as well [12, 13, 14].

An interesting magnetic SA concept for 3D was proposed by Shet *et al.* (Fig. 3.4) [13, 14, 23]. Soft magnetic layer 12 deposited on the microparts become magnetic when there is a permanent magnet 26 below the substrate. Then, the magnetic attraction guides the microparts to move into the matching cavities. When 26 moves away, the parts can still get closely attached to the substrate because of the hard magnetic layer 15 deposited in the cavities. The alignment has to take place sequentially from bigger parts to smaller ones to avoid mismatch.

For 3D integration with a large variation of die sizes, whether magnetic SA is a viable approach or not still requires further discussion. Some initial analysis will be performed in Section 3.4.

Another interesting magnetic SA technique is illustrated in Fig. 3.5 [15]. It presents some potential application in some unique 3D systems where 3D stacking is achieved by folding. The whole system can be fabricated on flexible substrate. Ultra-thin dies are embeded in polymer, then folded, aligned and bonded.

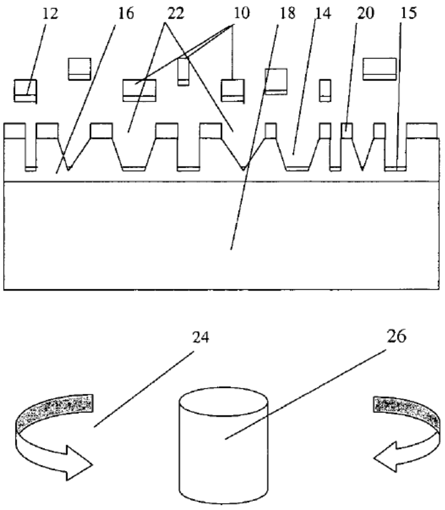

12
16
22
10
18
14
20
15
24
26

Figure 3.4 Magnetic-assisted SA technique proposed by Shet *et al.* [14]. Note: 12 - soft magnetic layer; 15 - hard magnetic layer; 26 - permanent magnet.

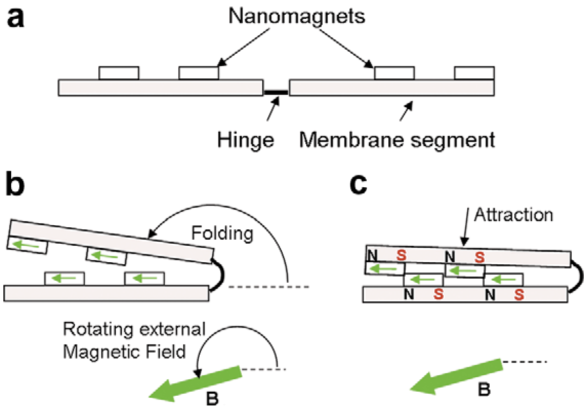

a
Nanomagnets
Hinge
Membrane segment
b
Folding
Rotating external
Magnetic Field
B
c
Attraction
N
S
N
S
N
S
N
S
B

Figure 3.5 Magnetic SA technique proposed by Nichol *et al.* [15].

3.2.4 Hydrophobic SA

A complete discussion will be performed in Chapter 4.

3.3 Silicon Interposer and Gravitational SA Technique

3.3.1 The Idea

Temporary Die Attachment and TSV Post-Processing

The dies are temporarily fixed in the etched recesses in the carrier, the precisely-defined positions of the recesses provide the alignment references. Then with some alignment markers on the carrier, precise wafer-level alignment can be achieved. A detailed process flow is proposed and illustrated in Fig. 3.6 to elaborate this technique.

The depth of these recesses is almost the full wafer thickness. With such recesses in an array in the wafer, the wafer becomes very fragile for handling. So, we propose to bond first a pyrex wafer on the backside of the silicon wafer to enhance the wafer strength.

A 380 µm-thick double-side polished silicon wafer is first anodically bonded to a pyrex wafer using Carl Suss SB6 [16] (Fig. 3.6-1). Silicon-pyrex anodic bond has been demonstrated to be mechanically stable and strong enough to withstand a large shear force of at least 1 MPa, or further reactive ion and wet chemical etching processes [17]. So, this stack should be sufficient to stand the following process steps. Before anodic wafer bonding, a 200 nm silicon nitride layer will have been already deposited on the exposed silicon surface. This will be used as the etching mask during cavity etching.

A key issue in this process is how we provide the alignment markers. Note that the TSVs are to be processed after the silicon side is thinned, so any structure for alignment will disappear if they are located on the exposed surface of the silicon wafer. We propose to make the alignment markers on the exposed surface of pyrex wafer since backside alignment components are equipped in the MA6 mask aligner. So now, a photolithography step can be performed on the pyrex exposed surface to define the markers for backside alignment (Fig. 3.6-2). After resist development, 2 µm pyrex etching will be done to transfer the marker patterns to the wafer followed with resist stripping (Fig. 3.6-3). Afterwards, the wafer stack will be flipped.

The shape and location of the cavities are well defined using backside alignment to the etched markers on the pyrex carrier (Fig. 3.6-4). A subsequent nitride etching will be performed. After resist stripping, the chip cavities will be etched through the whole silicon wafer (Fig. 3.6-5) with the remaining nitride as the etch mask.

A sloped profile is preferred for better tolerance of chip dimension variations after dicing. Our tapered silicon etching process developed in Chapter 2 provides 80° slope while KOH silicon etching provides a wider angle around 60°. Both etching processes can be adapted here because nitride hard mask is of very high etch selectivity. After KOH etching, the nitride can be stripped by HF [18], while for dry etch, it will be etched right after the silicon etching.

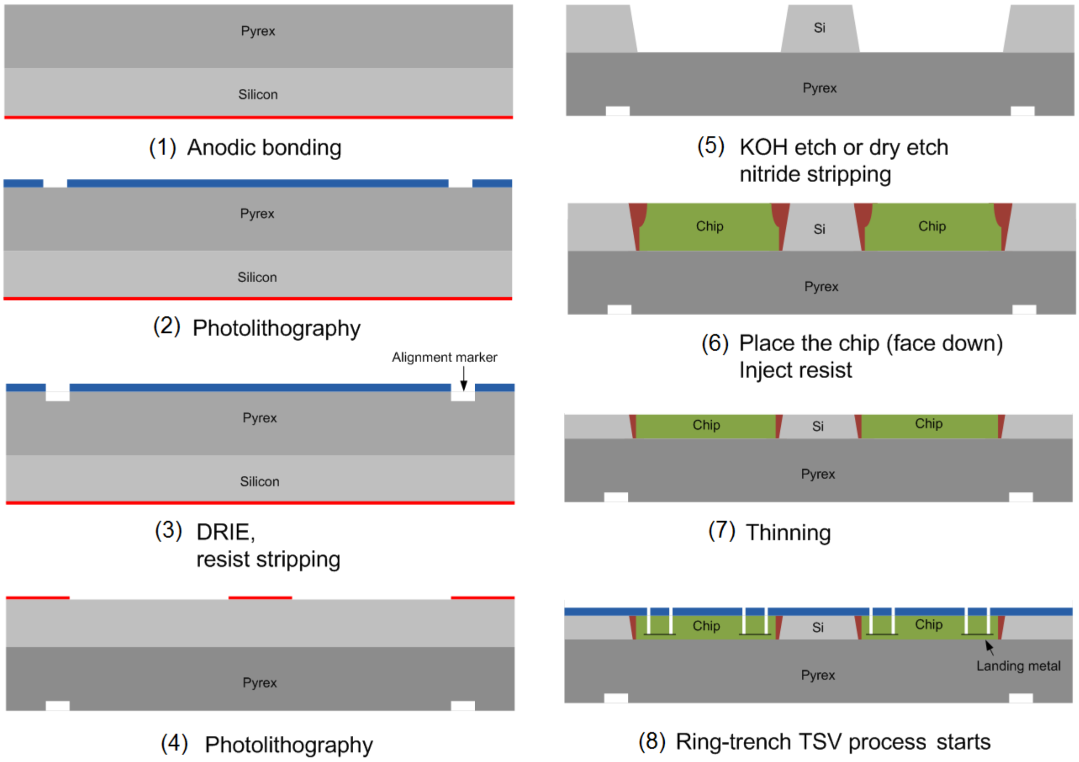
Afterwards, the chip will be placed face-down into the recesses and then fixed by injecting resists to seal the holes (Fig. 3.6-6). It has been reported by Fukushima *et al.* [19] that serious wafer warpage after thinning might be caused by improper resist underfill, and a proper resist should be of high viscosity and with a low CTE to avoid edge chipping after thinning. The resist is then cured according to its specification. Afterwards, the silicon side is thinned down to 50 µm through a first grinding and then polishing process (Fig. 3.6-7). CMP could be a good choice for the final planarization.

The stack is now ready for TSV fabrication. The next photolithography step is to pattern the ring trenches on the embedded chips. Backside alignment according to the alignment marks etched on the pyrex wafer (Fig. 3.6-8). Then, the process can be continued with the ring-trench TSV process described in Section 2.3.

When TSVs are filled by electroplating, the redistribution layer will be electroplated as the same time. Metal bumps can be fabricated on the bonding pads defined in the redistribution layer, but since bump fabrication is not the main focus of this thesis work, we just reference the related process flow [20].

After bump reflow is over, chips will be released from the carrier substrate by etching away the resist. The final thin die with post-processed TSVs and bumps is illustrated in Fig. 3.7.

Following the above process flow, our initial aim of post-processing TSVs on diced chips can be fulfilled. The chips released as shown in Fig. 3.7 will be ready for further stacking and permanent bonding.



(1) Anodic bonding

(2) Photolithography

(3) DRIE, resist stripping

(4) Photolithography

(5) KOH etch or dry etch nitride stripping

(6) Place the chip (face down) Inject resist

(7) Thinning

(8) Ring-trench TSV process starts

Figure 3.6 Post-processing TSVs on dies using gravitational SA technique.

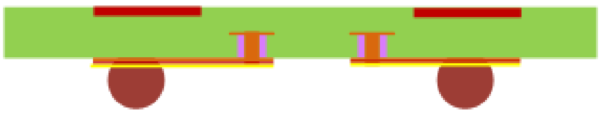


Figure 3.7 The final thin die with post-processed TSVs and bumps.

Permanent Stacking using Silicon Interposer

Using the dies post-processed with TSVs and bumps shown in Fig. 3.7, we are going to stack them permanently. Usually, flip chip bonders are the right tools for this task, but are not available. Then we had to propose again the use of etched cavities for alignment. Luckily, this method has another advantage on yield. The serially-processed flip chip process will require the unfinished stack going through multiple heating, pressing and cooling bonding cycles, which would damage the samples. Using the cavities, the multi-layer bonding could be done within one cycle and thus would reduce the risk of failure. The temperature gradient in the stack could be a problem, but as long as we have no flip chip bonder, this could be the only workable solution.

In Fig. 3.7, the TSV dimension is about 30 µm, as ring-trench TSV in Chapter 2 will be used (20 µm diameter copper core plus 5 µm wide ring isolation layer). The bump is normally larger than the TSV, could be, for example, 65 μm. With these relatively large dimensions, the mentioned SA precision limitation of gravitational SA can be tolerated. The thickness of the wafer needs to be well-calibrated to the final stack thickness after bonding. Finally, resist underfill is to be injected and cured to fix the chip stack. An illustration is shown in Fig. 3.8.

We further developed this idea by introducing TSVs into the carrier wafer itself in order to allow another layer of stacking (Fig. 3.8 and Fig. 3.9). This 3D silicon interposer solution well matches the memory-on-core stacking applications. 30 μm TSVs can provide the via density required in a 3D DRAM stacking. This idea is of the inborn merit of silicon interposers which allow signal redistribution. Nowadays, the inputs and outputs of memories and CPUs are all designed following different standards, so they cannot be stacked directly as they are. Processing TSVs on the delicate and complicated CPU cores have not been demonstrated yet. With the proposed stacking scheme, CPUs could be directly integrated into the 3D stacked systems.

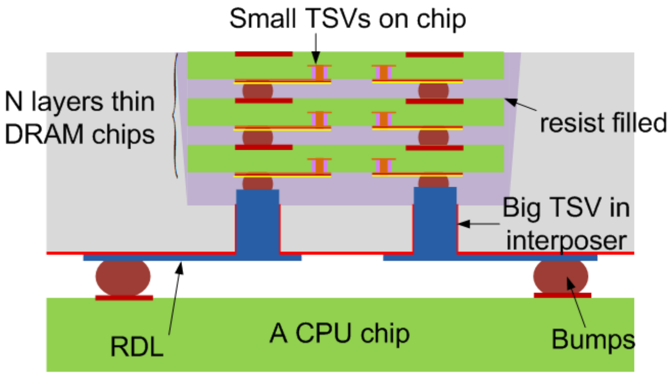


Figure 3.8 Suitable Si interposer application, taking memory-on-logic stacking for example. Figure not drawn to scale.

To have a clear view, the silicon interposer in Figure 3.8 is extracted and detailed in Fig. 3.9. As a demonstrator, TSV diameters do not need to be as small as a few tens of micrometers. As mentioned in Chapter 2, complete via filling is the greatest problem for us. Lower aspect ratio TSVs could be more realistic.

The optimal TSV dimension was 100 μm in diameter, and 100 μm in depth. Even though copper should be used for better conductance, this metal can contaminate our dry etchers and should be avoided. As the samples had to be opened on the wafer backside to expose TSVs in the cavities, nickel electroplating was used instead.

In our SA experiment, dummy dies for assembly were fabricated from glass wafers to allow direct inspection of the alignment precision.

To summarize the discussion on gravitational self-assembly concepts, the structure shown in Fig. 3.9 was found to be a good one that we aimed at. For this purpose, four different aspects of the process were to be investigated:

- Should use KOH or dry etching to etch the recesses used to hold the chip, since chip dimension variation after dicing is not clear yet?
- Assess the alignment accuracy supported by gravitational SA using normal dicing process with dicing saw.
- Make clear of the fabrication problems.
- Demonstrate low aspect ratio TSV fabrication as a by-product.

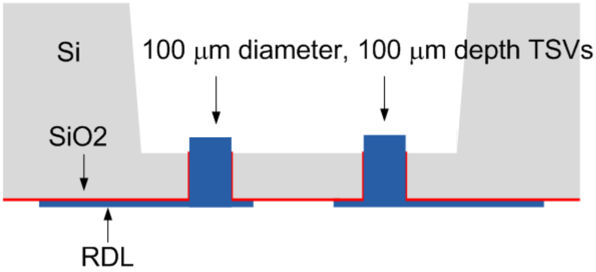


Figure 3.9 The silicon interposer to be fabricated.

3.3.2 Proof of Concept

A simplified process flow was then proposed to investigate the above problems, see Fig. 3.10. A 4 inch double-side polished 380 μm-thick <100> silicon wafer with 100 nm double-side thermal oxide was used. The cavity openings had been pattened through lithography using AZ9260, and the silicon dioxide exposed was removed by dry etch.

Square TSV holes of 100 μm × 100 μm were first patterned on the silicon dioxide layer through lithography using AZ9260 and silicon dioxide dry etch, and then etched 100 μm deep into silicon. 500 nm SiO_2 was then sputtered on TSVs as a dielectric isolation. Afterwards, a 60 nm/200 nm Ta/Au seed layer was sputtered.

Nickel electroplating was then processed trying to fill the via holes completely. A SEM image of the nickel TSVs after electroplating is shown in Fig. 3.11. Nickel grows uniformly on the whole surface. After the 100 μm via hole is completely filled, overgrowth of nickel appears on the wafer surface as the nickel bump dimension is enlarged to about 150 μm.

Then the wafer was immersed in the 40 °C KOH bath for 14 hours to etch cavities down by 280 μm. Additional 15 minute overetch was performed to expose the 5 μm-high TSV pillar extrusions shown in Fig. 3.12. Most TSVs seemed to be correctly filled (Fig. 3.12-a), while some showed big voids (Fig. 3.12-b).

The matching patterns of the TSVs on the chips were evaporated Cr/Au square pads defined by the lift-off technique. The glass wafer was diced with a standard dicer [26].

The glass dies were then released above the cavities and landed on the bottom. Successful alignment demonstrates a SA accuracy less than 20 μm (Fig. 3.13-a), while failed trials evidence a quite large pattern mismatch (Fig. 3.13-b). These failures can be attributed to a smaller chip dimension after dicing because the dicing blade was wider than the defined dicing channel and to a large profile angle achieved by KOH etching that provides over tolerance on the chip landing positions.

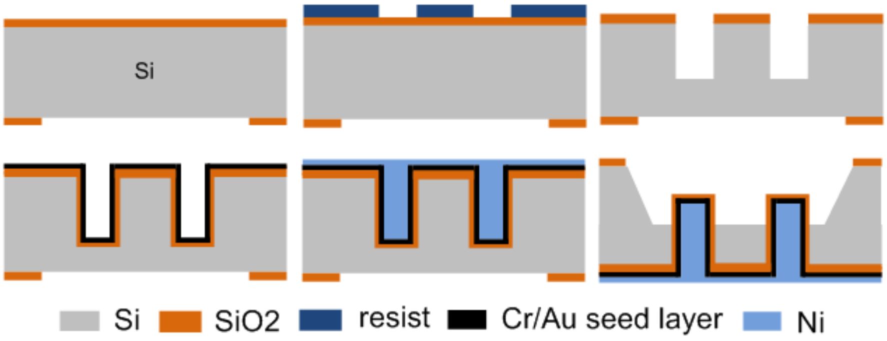


Figure 3.10 Simplied process sequence for the first trail of the silicon inteposer with TSVs.

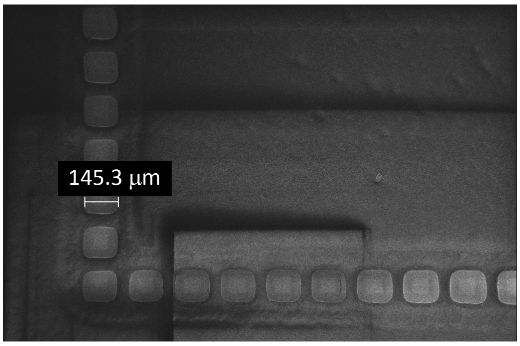
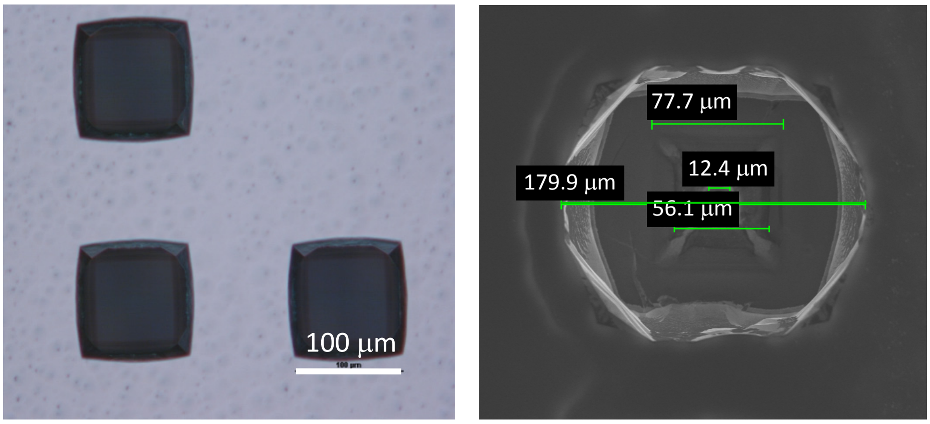
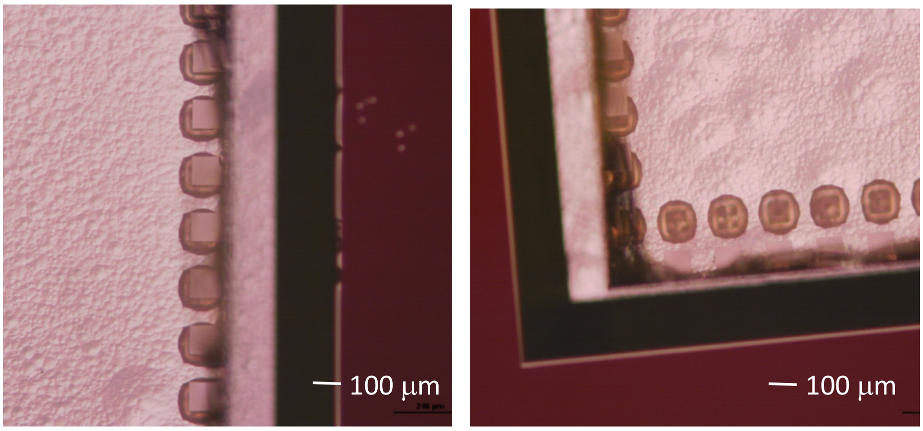


Figure 3.11 Ni TSV after CMP on the front side.



(a) (b)

Figure 3.12 Ni TSVs exposed after backside KOH silicon etching.



(a) (b)

Figure 3.13 Microscopic view of assembly. (a) Successfully aligned. (b) Not aligned.

3.3.3 Discussion of the Process, Further Improvements and Conclusion

From the above experience, it is considered that etching silicon with KOH is not the good choice. First, KOH etching leads to over sloped profiles which degrades the SA accuracy. Some other reasons are related to the process.

As shown in Fig. 3.12, the silicon around the TSVs cannot be totally removed because KOH etches silicon along the crystal plane. In addition, KOH etching generates some problems regarding the process. Alkali ions are detrimental to IC circuitry and potassium ions generated from etching represent a hazard since they can diffuse across the silicon substrate when heated during additional process steps (diffusivities of alkali metals can be found in [21, 22]). These ions are also a possible source of contamination in clean room.

Based on the above arguments, another wet cleaning process after KOH etching is desirable. This cleaning solution call "neutralization" in CMI consists of 6 parts of deionized water, 1 part of HCl, and 1 part of H_2O_2. As seen in Fig. 3.10, the front-side Ni is exposed in the sample. Although nickel is not attacked by KOH bath, it reacts with the neutralization bath. So, this last cleaning step becomes detrimental to this process flow.

To protect nickel from etching, a 200 nm SiO_2 layer was sputtered above the nickel layer. However, the wafer suffered greatly from bowing because of large internal stress in the electroplated nickel layer (Fig. 3.14).

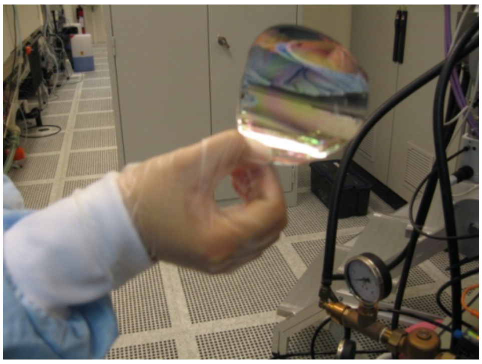


Figure 3.14 Wafer bowing after SiO_2 sputtered on Ni.

Instead of KOH, dry etch should be used. Several modification of the process can be proposed. First, profile angles can be adjusted using dry etch (detailed information to be found in Section 2.2.2). Second, copper can be used for TSV filling, and electroplating mold should be used to define the area for electroplating and reduce the difficulty of doing CMP on very thick copper layer after uniform copper electroplating.

To avoid the risk of copper contamination in the dry etchers, a thick thermal oxide layer (2 μm) is to be grown on the whole wafer surface. Etching process is also to be monitored carefully in order to avoid over etch. Later the exposed SiO_2 around Cu pillars can be etched using 10:1 buffer HF (10 H_2O : 1 49% HF) [24]. Thereafter, an improved process flow illustrated in Fig. 3.15 can be proposed.

To summarize, gravitational SA technique is demonstrated to be a promising approach for 3D systems with limited TSV dimension and density. The alignment accuracy can be improved with our improved solutions. However, it will always be constrained by the dicing accuracy.

The SA technique can be integrated with 3D silicon interposer concept to enable memory and CPU core stacking. High volume memories can be achieved by stacking multiple memories stacked in a single bonding process. CPU cores can be stacked as such with the help of redistribution layers fabricated on the silicon interposer. Higher yield and lower cost of the 3D assembly can be expected from this technology.

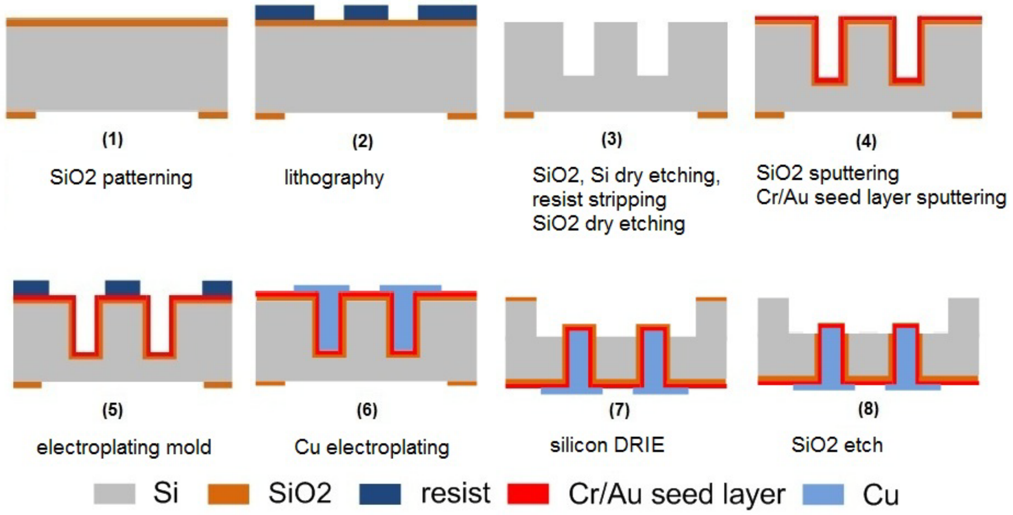
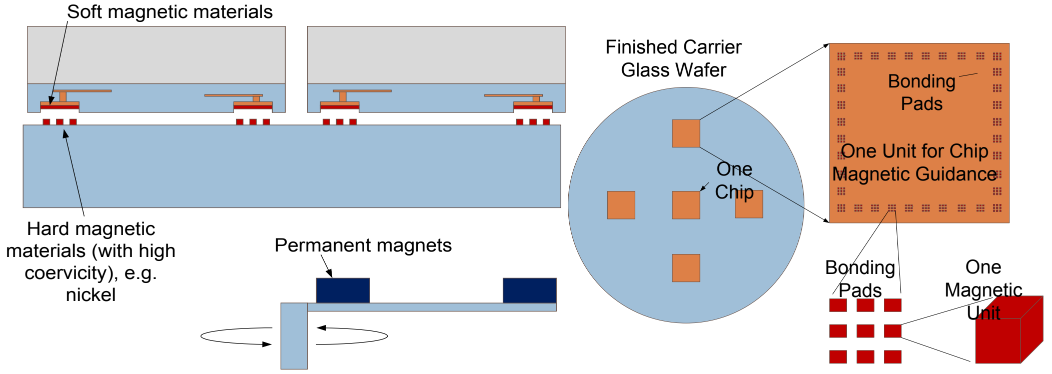

(1)
SiO2 patterning
(2)
lithography
(3)
SiO2, Si dry etching, resist stripping SiO2 dry etching
(4)
SiO2 sputtering Cr/Au seed layer sputtering
(5)
electroplating mold
(6)
Cu electroplating
(7)
silicon DRIE
(8)
SiO2 etch
Si
SiO2
resist
Cr/Au seed layer
Cu

Figure 3.15 Improved process flow to fabricate the 3D silicon interposer proposed in Fig. 3.12.

3.4 Magnetic SA Technique

3.4.1 The Idea

The magnetic SA technique is illustrated in Fig. 3.16. The rotating permanent magnets magnetizes the hard magnetic materials and the soft magnetic materials on the pads and the carrier. The generated magnetic force will then guide the chip to self-align.


Soft magnetic materials
Hard magnetic materials (with high coervicity), e.g. nickel
Permanent magnets
Finished Carrier Glass Wafer
Bonding Pads
One Unit for Chip Magnetic Guidance
One Chip
Bonding Pads
One Magnetic Unit

(a) (b)

Figure 3.16 Proposed magnetic SA technique. (a) A simplified illustration. (b) Magnets distribution topography on the carrier wafer.

3.4.2 Theoretical Analysis

The magnetic force on one pad with the permanent magnet sitting 1 mm below is simulated using the COMSOL model illustrated in Figure 3.17. Only one hard magnet is used in this simulation. We also simulated the case with an array of such cubic hard magnets. More effort is still required to fix the details, but some findings could be interesting to be shared.

- The magnetic force in Z direction is 1000 times larger than the forces in X and Y directions. The chip would rather stick to the carrier.

- Thinner soft magnetic layer on the pad is of larger magnetic force. The surface area of the magnetic layer is more influential than the height.
- With a pad dimension of 100 μm, the alignment force on one pad is on the order of 10^{-7} N.

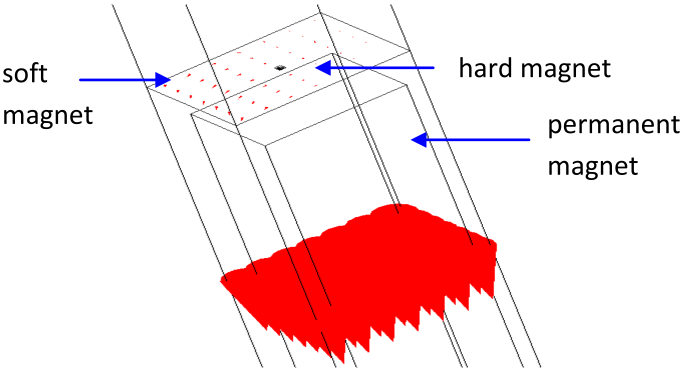


Figure 3.17 COMSOL simulation of magnetic forces on the pad with the permanent magnet 1mm away.

3.4.3 Initial Experiment Results and Summary

Since nickel is the most common magnetic material available in CMI, it was used to assemble the structures. With a hard magnet on the backside of the moving carrier wafer, no chip movement was observed. We tried to use water as lubricant, but the chip started to float away. Our first impression was that the surface tension of water could generate a larger alignment force than that originating from the magnets. The work then proceeded to hydrophobic self-alignment technique that will be described in the next chapter.

Some drawbacks of magnetic SA techniques have already been seen, such as the use of magnetic materials hardly compatible with standard fabrication processes. Monitoring and understanding the magnetic field and magnetic forces generated in thin film structures is quite complex. We could expect applying this technique in other applications, for example, nano device self-assembly. For ultra-thin die stacking, as we mentioned in Section 3.2.3, 3D stacking by folding ultra-thin dies connected by polymer could be an interesting study.

Reference

1. LEON4 Processor, http://www.gaisler.com/cms/index.php?option=com_content&task=view&id=338&Itemid=231.
2. Quickstick 135, http://www.emsdiasum.com/microscopy/products/materials/adhesives.aspx.
3. Thermal Release Tape REVALPHA, http://www.nitto.com/product/datasheet/e_parts/010/.
4. Koen De Munck, "Generic Building Blocks for 3D Integration and Their Application on Hybrid CMOS Image Sensors," PhD thesis, Katholieke Universiteit Leuven, 2008.
5. C. A. Bower, E. Menard, P. E. Garrou, "Transfer Printing: An Approach for Massively Parallel Assembly of Microscale Devices," *58th Electronic Components and Technology Conference*, 2008.
6. SUSS FC250, http://www.photonics.com/Article.aspx?AID=11792.
7. SUSS MA6 mask and wafer aligner, http://cmi.epfl.ch/photo/MA6_en.php.

8. H.-J. J. Yen and J. S. Smith, "Fluidic Self-Assembly for the Integration of GaAs Light-Emitting Diodes on Si Substrates," *IEEE Photonics Technology Letters*, Vol. 6, No. 6, June 1994.
9. J. S. Smith, H.-J. J. Yen, "Method for Fabricating Self-Assembled Microstructures," *United States Patent 5,783,856*, Jul. 21, 1998.
10. H. Sharifi, T.-Y. Choi, and S. Mohammadi, "Self-Aligned Wafer-Level Integration Technology with High-Density Interconnects and Embedded Passives," *IEEE Transactions on Advanced Packaging*, Vol. 30, No. 1, pp. 11-18, February 2007.
11. J. Dalin and J. Wilde, "Self-Assembly of Micro-Parts using Electrostatic Forces and Surface Tension," *59th Electronic Components and Technology Conference*, 2009.
12. C. G. Fonstad, Jr., "Magnetically-Assisted Statistical Assembly - A New Heterogeneous Integration Technique," https://dspace.mit.edu/retrieve/3964/AMMNS013.pdf, February 22, 2004.
13. S. Shet, Vishal R. Mehta, A. T. Fiory, M. P. Lepselter, and N. M. Reavindra, "The Magnetic Field-Assisted Assembly of Nanoscale Semiconductor Devices: A New Technique ," *Journal of Microelectronics*, October 2004.
14. R. M. Nuggehalli, A. T. Fiory, S. Sudhakar, "Method of Magnetic Field Assisted Self-Assembly," *United States Patent 7,217,592 B2*, May 15, 2007.
15. A. J. Nichol, P. S. Stellman, W. J. Arora, G. Barbastathis, " Two-Step Magnetic Self-Alignment of Folded Membranes for 3D Nanomanufacturing," *Microelectronic Engineering*, Vol. 84, pp. 1168–1171, 2007.
16. SUSS Substrate Bonder SB6, http://cmi.epfl.ch/packaging/SB6.php.
17. M. Despont, H. Gross, F. Arrouy, C. Stebler and U. Staufer, "Fabrication of a Silicon-Pyrex-Silicon Stack by A.C. Anodic Bonding," *Sensors and Actuators A: Physical*, Vol. 55, Issues 2-3, pp. 219-224, 31 July 1996.
18. Chapelle chimique pour le recyclage des plaques (Reclaim), http://cmi.epfl.ch/thinfilms/Reclaim.php.
19. T. Fukushima, Y. Yamada, H. Kikuchi, and M. Koyanagi, "New Three-Dimensional Integration Technology Using Self-Assembly Technique," *IEEE International Electron Devices Meeting*, pp. 348 - 351, 2005.
20. Bumping Design Guide, Revised May 2009, © FlipChip International LLC, 2009.
21. S. Linder, H. Baltes, F. Gnaedinger, and E. Doering, In *Fabrication Technology for Wafer Through-Hole Interconnections and Three-Dimensional Stacks of Chips and Wafers*, pp. 349-354, 1994.
22. U. Münch, N. Schneeberger, O. Paul, H. Baltes, and E. Doering, "Thin Film Front Protection of CMOSWafers Against KOH," *Proceedings of SPIE - The International Society for Optical Engineering* 3514, pp. 124–133, 1998.
23. R. D. Rivero, S. Shet, M. R. Booty, A. T. Fiory, and N. M. Ravindra, "Modeling of Magnetic-Field-Assisted Assembly of Semiconductor Devices," *Journal of Electronic Materials*, Vol. 37, No. 4, 2008.
24. K. R. Williams, K. Gupta and M. Wasilik, "Etch Rates for Micromachining Processing—Part II," *Journal of Microelectromechanical Systems*, Vol. 12, Issue 6, pp. 761 - 778, 2003.
25. S. Farrens, "Wafer and Die Bonding Technologies for 3D Integration," *MRS Proceedings, Symposium E.*, Fall 2008.
26. Scie DISCO DAD321, http://cmi.epfl.ch/packaging/Disco.php.

Chapter 4

Hydrophobic SA for Heterogeneous 3D Integration

4.1 Background

Heterogeneous 3D integration allows the combination of disparate technologies such as silicon-on-insulator (SOI), bulk complementary metal–oxide–semiconductor (CMOS), analog and even microelectromechanical systems (MEMS) dies into a single, tightly-coupled chip stack. Circuit quality is improved because each layer is manufactured in its own highly-optimized manufacturing process.

Currently, people are going for silicon interposer solutions as the intermediate approach. However, the limitations regarding the relatively large TSV dimensions and low density are still matters of intensitve study since they directly affect inter-layer bandwidth in 3D stacked circuits.

In future, removing the interposer and directly stacking chips with dense TSV interconnections is necessary. According to the International Technology Roadmap for Semiconductors (ITRS) [1], TSVs that will be implemented in future technologies will be less than 2 μm in diameter with a pitch of 4 μm.

To be able to achieve inter-layer connections through such tiny TSVs, sub-micron alignment accuracies are required.

Sub-micron accuracy can be achieved with dedicated wafer-level bonding tools. It is the method people have used the most in semiconductor industry. However, this scheme is of lower yield and higher cost for heterogeneous 3D stacking mainly for two reasons: first, a single defect on one strata results in a nonfunctional chip stack; second, all dies in the stack are constrained to have the same size, which means that valuable silicon area could be wasted. Die-level stacking allows the stacking of known good dies (KGD) with disparate dimensions. However, in high-volume manufacturing, serial processing, such as die-to-die or die-to-wafer bonding using flip-chip bonding tools, is costly due to the low throughput. SA techniques allow the manipulation of multiple dies in parallel within one single processing step and thus become viable technologies for 3D integration.

4.1.1 The Advantage of Hydrophobic SA over the Other SA Techniques

For 3D die-level stacking, four criteria for a successful SA technique are to be considered:

- precise pad-to-pad alignment from die to carrier;
- compatibility with prevailing bonding techniques, such as eutectic soldering or thermal compression bonding;
- process stability and reliability;
- minimum or no additional patterns introduced on the chip.

These four criteria are critical especially for advanced 3D applications such as heterogeneous 3D systems.

The gravitational SA technique has been experimented in Chapter 3 and shows a limited alignment accuracy which depends on the dicing inaccuracy and recess precision. A silicon interposer is also to be fabricated to handle the dies at wafer level, which introduces process complexity and cost.

On the other hand, SA based on electrostatic force needs to implement passive components and conducting paths to generate electrical field. These redundant structures are not desired on the chip.

Finally, SA based on magnetic force has been briefly studied in Chapter 3 as well. Some drawbacks have been found, such as the use of magnetic materials which are hardly compatible with standard fabrication process. Monitoring magnetic forces in thin film structures is also difficult. We could expect to apply this technique in other applications, e.g. nano-device self-assembly.

According to these evaluations, it comes out that hydrophobic SA could be the simplest and the least invasive one, making this technology the best candidate for heterogeneous 3D integration.

4.1.2 Hydrophobic SA

Keeping in mind the four concerns mentioned in Section 4.1.2 when examining the hydrophobic SA techniques proposed in previous publications [2-7], we are further convinced about the potential of hydrophobic SA usage in 3D even though each technique has limitations.

In [6], two examples of multi-chip SA were discussed, namely SA on silicon or on polymeric substrate. In both cases, a layer of SiO_2 was deposited on the binding site (the entire chip surface area: 5 mm × 5 mm) to provide locally a more hydrophilic area than the rest exposed silicon substrate surface. The chip surface was completely covered with SiO_2.

Several hundreds of nanoliters of aqueous solution or resins were applied onto the binding sites. Then, the chip was placed on the liquid drop. The fluid wetted the complete surface up to the die periphery. And the SA was induced by the surface tension force resulting from surface energy minimization of the water-air meniscus interface around the die periphery. So, this method resulted in periphery-to-periphery SA (Fig. 4.1), which will be called as such throughout this thesis. Further study about the effects of die-size precision, possible SiO_2-SiO_2 thermal compression bonding after SA, and the automation of this multi-chip SA technique were presented in [7].

In [7], the authors used an expensive and precise dry etching with subsequent dicing which is not compatible with high-volume manufacturing (HVM) standard dicing processes. In HVM standard dicing processes, chip dimension variations are typically of ±15 to 25 µm mainly because of temperature instabilities, waved dicing effects, blade wear of the blade, and optical misalignment. In case of periphery-to-periphery alignment, the misalignment of the patterns relatively to the chip periphery caused by dicing inaccuracies must also be considered.

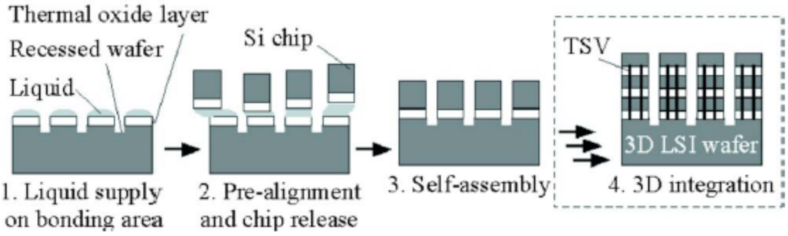
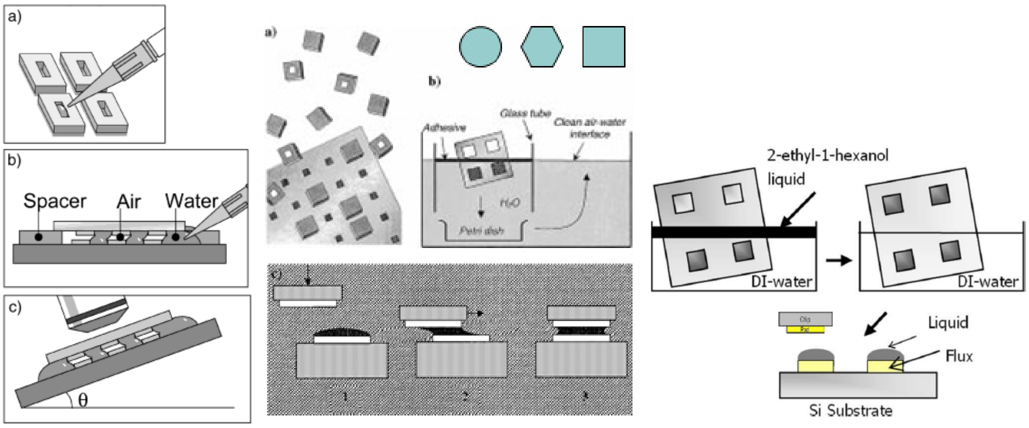


Figure 4.1 Periphery-to-periphery SA technique proposed by Fukushima *et al.* [7].

Some pattern-to-pattern hydrophobic SA techniques have been demonstrated [2, 4, 5].

For instance, in [4], a two-inch glass wafer and a three-inch silicon wafer with hydrophobic matching grid or frame pattern have been aligned by means of minimization of the meniscus surface (Fig. 4.2-a). This method will be referred to as "frame-to-frame SA" in this thesis. The alignment accuracy was better than 1 μm, but some of the water was trapped between the substrates could not be easily released.



(1) (2) (3)

Figure 4.2 Hydrophobic SA techniques. (1) Wafer-level frame-to-frame hydrophobic SA technique proposed by Martin *et al.* [4]. (2) A SA procedure using surface tension force generated by lubricants, proposed by Böhringer *et al.* [2]. (3) Pad-to-pad SA technique proposed by Chang *et al.* [5].

In [2], SA was achieved through capillary forces created by an hydrophobic organic lubricant covering hydrophobic areas in the presence of water (Fig. 4.2-b). Small dies of 150 to 400 μm in size were successfully assembled. The influence of various hydrophobic patterns with respect to the alignment accuracy was studied as well.

The use of UV-curable adhesives or liquid solders to form a reliable bond interface was also proposed in [8]. Hereby, the bonding temperature must be below 100 °C because of the water environment. This is a constraint that imposes using some low melting-point solder, which is incompatible with further reflow process and reliability needs.

A SA technique compatible with metal bonding was discussed in [5] (Fig. 4.2-c). Here, the binding sites on the carrier consisted in eutectic solder bumps. The flux covering the screen-printed solder of very high viscosity, so 2-ethyl-1-hexanol (2EH) was introduced as lubricant. 2EH exhibits a good wettability on the hydrophobic Au area and poor wettability on silicon. Since pads were used as SA structures, this process can be specified as "pad-to-pad SA". However, the pad-to-pad SA error was about 13.2 μm, which is too high for 3D integration which requires sub-micron accuracy. Such a high misalignment could come from imprecise pattern matching and large frictions due to the rough surface of the solder bumps.

4.2 Our Method

In this section, we will explain the basic principle of a modified technique based on surface tension driven SA. First, we will recall some properties of hydrophilic and hydrophobic surfaces in the relation with our specific clean room processes.

The formation of alkanethiol self-assembled monolayers (SAMs) on patterned gold surfaces results in local hydrophobicity, see [2, 3, 4, 5]. The gold is first cleaned with an oxygen plasma or in piranha. The contact angle of water on the SAM-covered Au surface is close to 110 °, whereas the rest of the chip – whether it is silicon, silicon dioxide or silicon nitride – is hydrophilic with a contact angle ranging from nearly 0 ° (right after cleaning, e.g. piranha) to about 30 ° (after being exposed for hours to air). Therefore, water dispensed on the chip surface wets the hydrophilic areas quickly and stops at the periphery of the SAM-covered Au pads. This water-air interface will be used later as the alignment reference.

This implementation of patterned hydrophobic layers is compatible with standard IC-fabrication technology. The uppermost layer of the solder pads is immersion gold, preventing oxidation of the underlying nickel wetting layer. It is suitable as a template for the SAM growth. The area between the pads is covered with a hydrophilic passivation layer of silicon oxide or silicon nitride. Accordingly, a high wettability contrast can be achieved with batch-process-compatible alkanethiol SAM processing, without the need for any extra lithography step. The pad-to-pad SA technique of multiple dies in parallel is illustrated in Fig. 4.3.

Prior to the SA procedure, a defined amount of water is dispensed in the center of the binding site in the carrier wafer. The glass dies are then pre-aligned in proximity to the binding sites. The dies are brought into contact with the water on the carrier and then released. The water trapped between the floating die and the carrier now wets the entire hydrophilic surface. The relaxation of the water-air interface achieves the final position of the die relative to the carrier. During water evaporation, the distance between the dies is reduced until contact between the die and the carrier is achieved. The subsequent Au-Au thermo-compression bonding results in a reliable electrical interconnect.

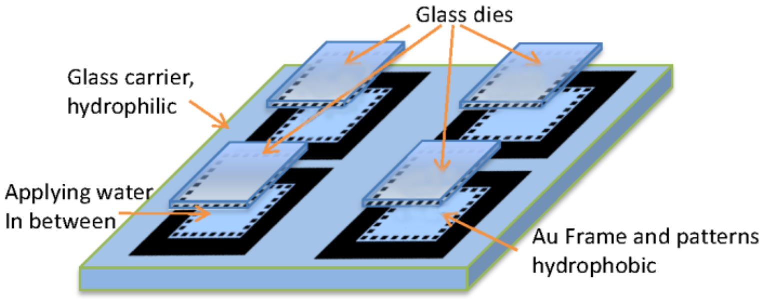

Glass dies
Glass carrier, hydrophilic
Applying water In between
Au Frame and patterns hydrophobic

Figure 4.3 Illustration of the proposed pad-assisted SA technique with multiple dies aligned in parallel. The black surfaces are hydrophobic. Therefore, water is contained between the die and the Au frame on the carrier only. The pads on the two parts are congruent, further supporting the alignment procedure, to achieve high pad-to-pad accuracy. After SA, thermo-compression bonding can be performed.

4.3 Experiment Evaluation

4.3.1 Experimented Self-Alignment Processes

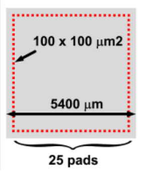
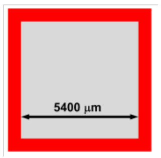
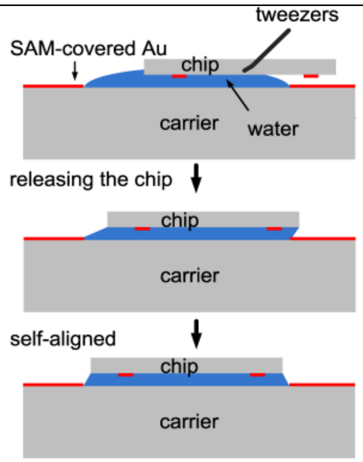
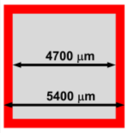
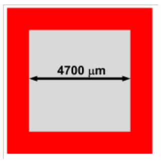
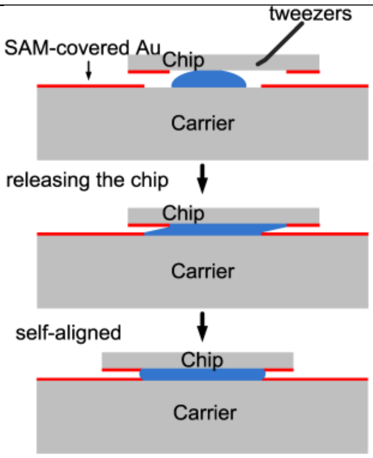
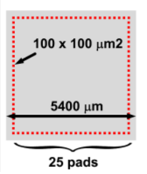
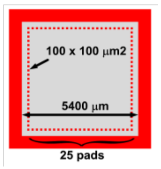
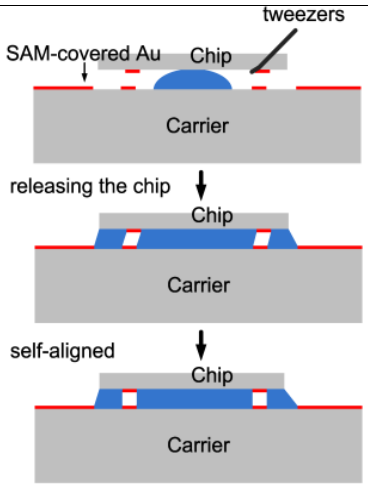
Three different SA procedures are benchmarked in terms of their pad-to-pad alignment quality. Therefore, alignment markers were added in the center of the die and of the carrier. The pattern design and the resulting wetting are presented schematically in Table 4.1. The procedures are referred according to the water-wetting topology, namely: "fully-wetted (F-W) SA", "partially-wetted (P-W) SA" and "pad-assisted (P-A) SA". F-W SA is actually peripheral-to-peripheral SA. Water covers all the binding sites. And the whole chip surface is wetted by water. The target is to study peripheral-to-peripheral SA's tolerance of large dicing inaccuracies obtained from normal dicing processes using the dicing blade. P-W SA and P-A SA are both structure-to-structure SA.

The purpose of P-W SA is to study whether or not the SA force generated from the well-defined water menisci along the frame is sufficient to align the chip with a better precision. Similarly, we will study if in the P-A SA technique the water-to-air interfaces can be successfully generated around the pads and if the corresponding SA forces are sufficient to align the chip with a good precision.

4.3.2 Sample Preparation

Glass dies and glass carriers fabricated from a 550-μm-thick float glass wafer were used as specimens to inspect the pad-to-pad alignment with the microscope during the experiments. 50 nm Cr/100 nm Ti/150 nm Au layers were evaporated onto the glass substrates and patterned using lift-off technology. The dies were designed to be of the same size as the binding site in the carrier. The die dimension is 5400 μm as shown in Table 4.1. After die singulation with a dicing saw, the die dimension was reduced to 5250 μm, with a precision of ± 15 to 25 μm. Glass chipping during dicing resulted in a wavy die periphery (Fig. 4.5). Subsequently, the chips and carriers were rinsed in deionized (DI) water to remove particles. The protective resist layer was striped in a positive-tone remover and the gold surface was cleaned in piranha solution for 5 minutes. After rinsing in water, the chips and the carriers were immersed into 1 mMol 1-octadecanethiol (1-ODT, $CH_3(CH_2)_{17}SH$, 98%, Aldrich) for 24 hours. The samples were rinsed with ethanol to remove SAM residues and to prevent oxidation of the surface. Prior to the SA experiment, the specimens were dried using N_2.

Table 4.1 SA Procedures with Pattern Design and Water Wetting Topology.

Name	Design	SA Procedure
F-W SA	Chip  100 x 100 μm2 5400 μm 25 pads Carrier 	 tweezers SAM-covered Au chip carrier water releasing the chip chip carrier self-aligned chip carrier
P-W SA	Chip  4700 μm 5400 μm Carrier 	 tweezers SAM-covered Au Chip Carrier releasing the chip Chip Carrier self-aligned Chip Carrier
P-A SA	Chip  100 x 100 μm2 5400 μm 25 pads Carrier  100 x 100 μm2 5400 μm 25 pads	 tweezers SAM-covered Au Chip Carrier releasing the chip Chip Carrier self-aligned Chip Carrier

4.3.3 Experiment Results

Fully-Wetted SA

In this procedure, the entire bottom surface of the die was wetted, and the die pads were thus not functionalized. Accordingly, the water-air interface responsible for the alignment forces spans from the die periphery to the carrier frame pattern, so that die-periphery to carrier-frame alignment is expected. Ten experiments with varying amounts water ranging from 0.5 μl to 10 μl were performed. The water amount was controlled by a micropipette.

At lower water amounts, the pad-to-pad alignment accuracy improved from ~50 μm to ~10 μm (Fig. 4.4). At such a low water content, the die is in contact with the carrier from the very beginning of the assembly process, and the force from the meniscus is too small to align the specimen. As optimal, a water amount of 0.8 μl to 2.0 μl with a resulting 10 μm pad-to-pad alignment accuracy could be identified. This result is primarily defined by the pad-to-periphery dicing inaccuracy.

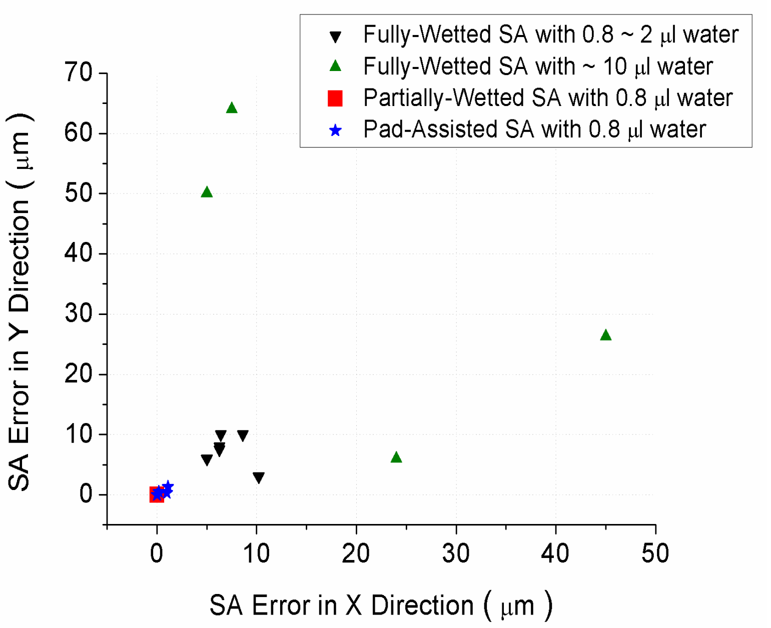


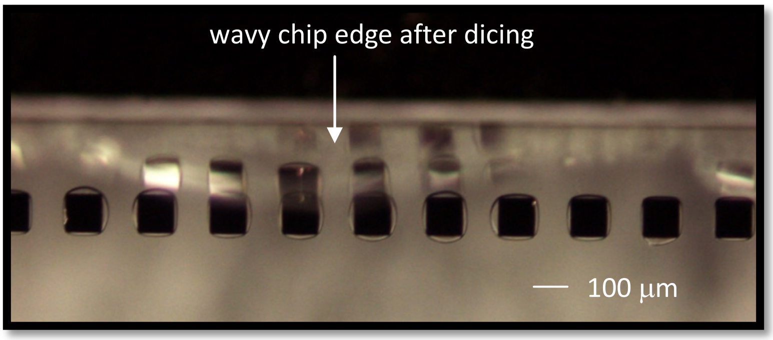
Figure 4.4 Experimental pad-to-pad alignment inaccuracy in the x- and the y-direction of ten samples for fully-wetted, partially-wetted, and pad-assisted SA procedures.

Partially-Wetted SA

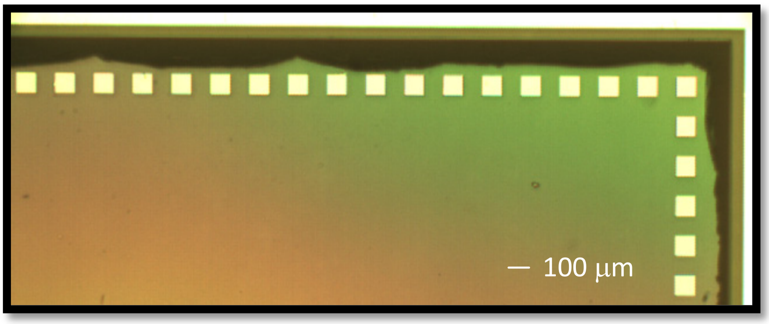
The use of a hydrophobic frame on the die prevents water from reaching the die periphery and precisely define the water between the frames on both sides of the assembly. With a water amount of 0.8 μl, a submicron pad-to-pad alignment accuracy below the resolution of the optical microscope could be achieved in all ten trials (Fig. 4.4, red squares). This proves the robustness of the process against dicing inaccuracies. Unfortunately, according to standard chip-fabrication design rules, large metal frames around the chip periphery are not accepted because of reliability and processing concerns.

Pad-Assisted SA

As shown in partially-wetted SA experiments above, well-defined water menisci from the die pattern to the carrier pattern are the key to improve pad-to-pad alignment accuracy. Therefore, also the pads of the die were functionalized to form pad-to-pad menisci (Table 4.1). The alignment is now performed via the forces exerted on the die from the peripheral and the pad-to-pad menisci. Again, 0.8 µl of water was dispensed in the center of the binding site. An initial chip alignment precision of less than the size of a pad is now needed to form the water-air interface between the corresponding pads. The die was gently brought into contact with the fluid surface to prevent water from penetrating the hydrophobic pad surface. Then the die was released and got accelerated by the meniscus forces, resulting in oscillations with an exponentially decaying amplitude for 1.7 s (the case in Fig. 4.6) until the system relaxed completely. Usually, the pad-assisted SA process finishes within 0.5 s to 2 s. The pad-to-pad meniscus forming an air bubble underneath each pad is now visible (Fig. 4.5-a). After water evaporation (Fig. 4.5-b), pad-to-pad alignment accuracies of less than 1 µm (on average, 0.33 µm) were observed for ten trials (Fig. 4.4, blue stars).



(a)



(b)

Figure 4.5 Optical microscope photographs of the pad-assisted SA process. (a) The water-air interfaces around the pads are visible during water evaporation. (b) Pad-to-pad alignment after water evaporation.

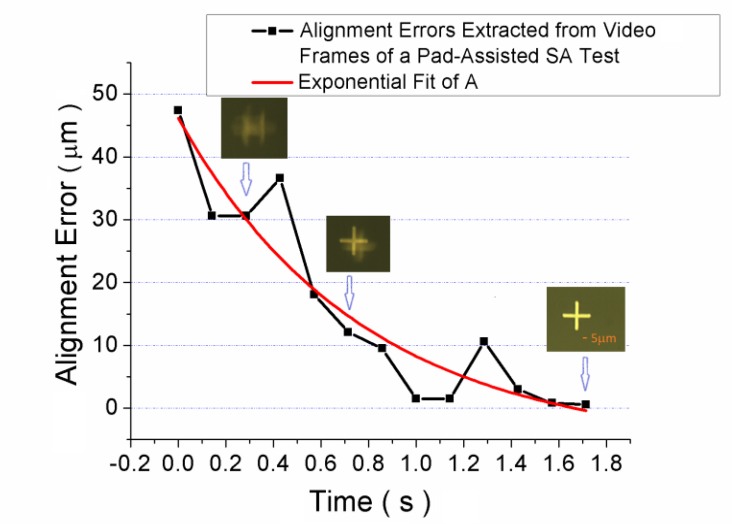


Figure 4.6 SA dynamics: Alignment error amplitude measured by optical-microscope video-imaging during the pad-assisted SA process until complete relaxation is achieved. (Pad dimension: 100 µm).

The robustness of the pad-assisted SA procedure with respect to the pre-alignment accuracy was studied at an initial offset of one pad (100 µm), one pad plus one pitch (300 µm in this case), and half the die dimension (about 2 mm). All cases resulted in improved alignment compared to the fully-wetted case (Fig. 4.7). Video imaging revealed the existence of trapped air on the carrier pads even without a mating die pad during the alignment phase. Finally, pad-to-pad menisci could be observed with the optical microscope after relaxation for all tests. It seems that the pad-to-pad meniscus formation is reversible and occurs if a die pad is close by. However, the resulting alignment accuracy was affected by the initial displacement and resulted in average values of 0.33 µm, 3.35 µm and 7.63 µm for the three cases. This effect could be explained by the lower air bubble formation yield for large displacements. The fewer pad-to-pad menisci available, the higher the predominance of the peripheral (for chip, it is periphery; for carrier, it is frame) meniscus force. Accordingly, the misalignment asymptotically approaches the value of the fully-wetted procedure.

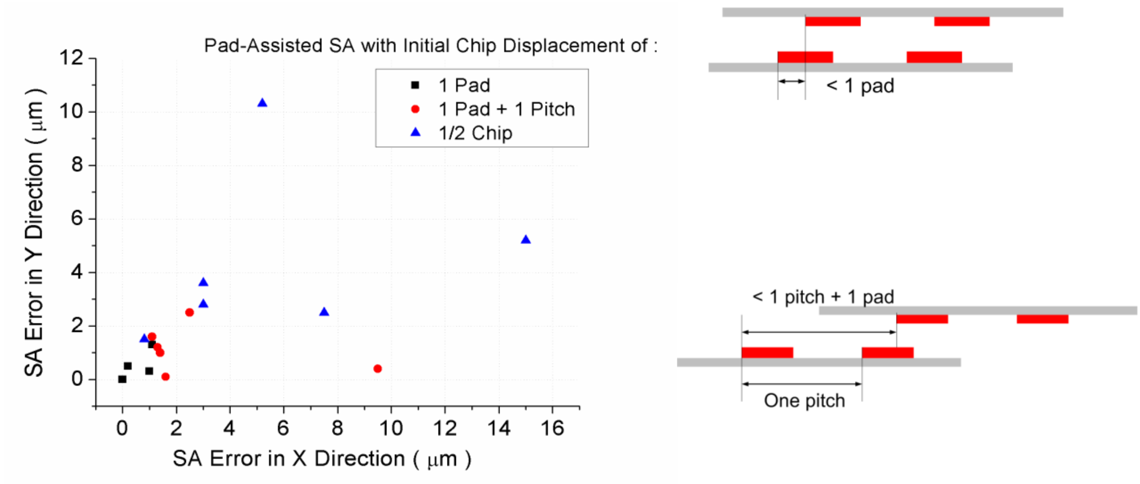


Figure 4.7 Experimental pad-to-pad alignment inaccuracy in the *x*- and *y*-direction of pad-assisted SA process with different initial chip displacements.

To increase the pad-to-pad menisci force, four columns of pads were designed on the die and carrier side. The experiments were carried out at a pre-misalignment of about six-pad-pitch. Four out of the six trials resulted in a one-pad-line misalignment. When the chip finally stays with one-pad-line misalignment, not all the pads are aligned to their matching pads, but the alignment force induced by the increased number of pads seems to be larger than the force induced by the peripheral meniscus. This is an important finding, highlighting the need for design methods that enable a prediction of the optimal number of pads.

Pad-Assisted SA with Gold Studs and Subsequent Thermo-Compression Bonding

15 µm-high and 40 µm-wide gold studs were deposited by electroplating in the center of the carrier pads for later Au-Au thermo-compression boding (Fig. 4.9-a). SA experiments resulted in poor quality as when using the fully-wetted procedure.

Note that after standard alkanethiol treatment, the contact angle of water on the rough top surface of the bump does not exceed 90 ° instead of being about 110 ° on a flat SAM-covered Au surface. This could lead to an absence of air trapped on the carrier pads. However, air bubbles are still observable on many of the carrier pads. Unfortunately, with these air bubbles, the alignment forces generated by the water-air menisci around the pads cannot move the chip to a well-aligned position as in the case with no stud. The most likely reason could be the large friction force between the rough Au stud top surfaces and the Au pads on the chip. To fully determine the reasons, more experiments are needed, such as measurements of the friction factor [9]. If the friction force is the reason leading to SA failure, further planarization of the stub surface will be helpful to improve SA.

In our experiment an additional ring (diameter: 2 mm, width: 0.5 mm) or pad patterns in the chip and binding site center without Au studs were implemented to overcome this issue and generate hydrophobic areas (as described in [2, 3] and shown in Fig. 4. 10). The ring pattern (Fig. 4.10-b) led to a sub-micron accuracy even in the presence of Au studs. The pad array also requires a pre-alignment better than one pad dimension (the same as Fig. 4.7)

Here, we propose to explain why ring structures lead to much better results: although water menisci are formed along the rings, the chip weight is too large to be held by these menisci and so water will finally spread over all the hydrophilic areas. But a portion of the water already gets into the area outside of the ring through the hole on the right. And water wets the whole hydrophilic area almost immediately. During this period, the chip is held above the final position where the Au studs touch the upper Au pads, so the friction forces have not started to perform or have not fully performed against SA. Also the ring structure forces the water to spread over the binding site symmetrically.

Finally, the die and carrier with Au studs were exposed to 380 °C and 20 N force after SA. The shear test resulted in ductile cohesive fracture, indicated a high bond quality (Fig. 4.9-b).

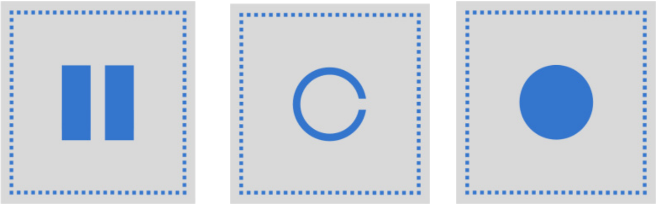
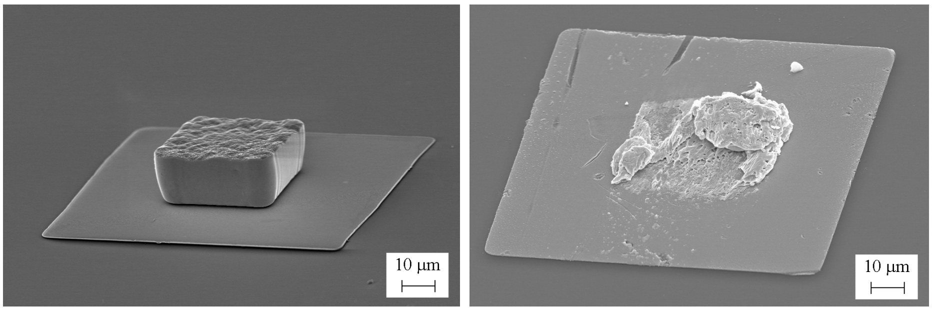


Figure 4.8 Center patterns introduced to the center of the chip and the binding site.



(a) (b)

Figure 4.9 SEM image of (a) a 15-µm-high Au-stud prior to thermo-compression bonding. (b) Au-Au interface after shear test, indicating a ductile cohesive fracture.

4.4 Meniscus Model

4.4.1 Analytical Model of SA

To understand the experimental results in more details and to be able to design reliable SA patterns, a modeling framework is needed. This can be addressed through a meniscus model that we intend to develop in this section.

The absolute value of the force F acting on the die is equal to relative increment of the meniscus surface area ∂A with respect to a chip displacement ∂s times the water surface tension γ:

$$F = -\gamma \frac{\partial A}{\partial s} = -\frac{\partial E}{\partial s}, \tag{1}$$

where E is the total surface energy originating from the water-air free interfaces.

According to figure 4.7, any misalignment in the x-direction will generate a net force that can be described as a superposition of a force induced by the peripherals and a force originating from the pad menisci. This reads

$$F_x = F_{peri} + F_{pad} = -\frac{\partial E_{peri}}{\partial \Delta X_{chip}} - \frac{\partial E_{pad}}{\partial \Delta X_{pattern}} = \gamma_{SAM} \times \left(\frac{\partial A_{peri}}{\partial \Delta X_{chip}} + \frac{\partial A_{pad}}{\partial \Delta X_{pattern}}\right) . \tag{2}$$

Accordingly, the periphery-to-periphery and the pad-to-pad (pattern-to-pattern) displacement are called ΔX_{chip} and $\Delta X_{pattern}$, respectively. The area of the water-air interface along the periphery is calculated as

$$A_{peri} = \frac{1}{2} \times (D_{chip} + D_{frame}) \times \left(\sqrt{\left(\frac{D_{frame}}{2} - \frac{D_{chip}}{2} + \Delta X_{chip}\right)^2 + H^2} + \sqrt{\left(\frac{D_{frame}}{2} - \frac{D_{chip}}{2} - \Delta X_{chip}\right)^2 + H^2} \right) \tag{3}$$

where H represents the gap between the die and carrier, and D_{chip} and D_{frame} denote the dimension of the die and carrier frame, respectively (Fig. 4.10, left).

The pad-to-pad meniscus area is described as follows:

$$A_{pad} = NO_{bubble} \times D_{pad} \times \sqrt{\left(\Delta X_{pattern} - N \times P_{pad}\right)^2 + H^2} , \tag{4}$$

with the total number of air bubbles NO_{bubble}, the pad size D_{pad}, and the number of misaligned pads N at a pad pitch P_{pad}. The die moves as long as the surface tension force is not zero. To mimic the dicing inaccuracy, an offset d of 25 µm is introduced ($\Delta X_{chip} = \Delta X_{pattern} + d$) (Fig. 4.10, right).

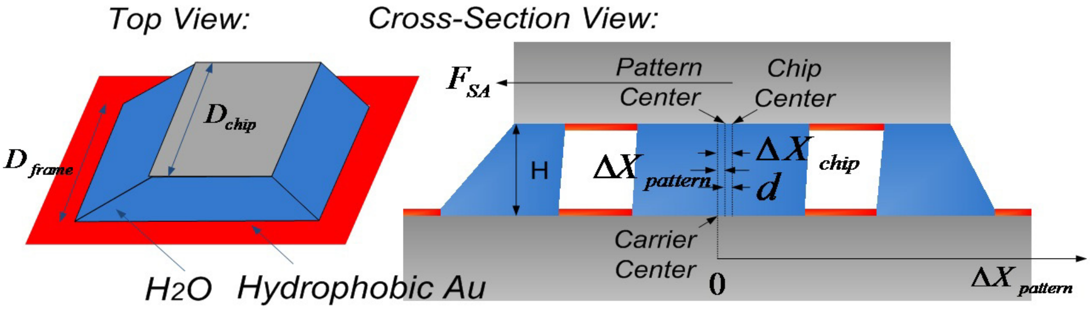


Figure 4.10 Meniscus model of pad-assisted SA.

4.4.2 Modeling Results

Fully-Wet SA Force and Optimal Water Amount

Figure 4.11 shows the resulting alignment force of the fully-wetted SA obtained from the model at different die-to-carrier heights *H*. In all cases, zero force is observed at a pattern misalignment $\Delta X_{pattern}$ of 25 μm because of the dicing inaccuracy that has been introduced. The influence of a larger carrier frame D_{frame} compared with the die size D_{chip} of 150 μm results in a reduced slope at chip displacements $\Delta X_{chip} < 75$ μm and small gaps. The increase of the meniscus surface on one side of the die periphery to the carrier frame is compensated by the reduction of meniscus area on the other side as long as the die stays within the carrier frame. A strong increase in the alignment force is observed if the die is displaced across the carrier frame. Small amounts of water would accordingly result in a higher alignment accuracy if we assume a yield force to move the floating die (induced, e.g., by particles in the fluid, tilting errors, frictions, etc.). The optimal water height roughly ranges from 35 μm (0.88 μl) to 85 μm (2.13 μl) as shown in Fig. 4.5. With over amount of water (H = 300 μm), the SA force is not big enough to align the chip to the precise position. With too little water (H = 15 μm), the alignment forces sharply reduces after the chip completely enters the frame. Expressed in the curve in Fig. 4.11, a plateau is shown (the curve with black dots). Before the chip completely enters the frame, the SA forces generated on each side of the chip work together to pull the chip into the binding site, while after the chip completely sits above the frame, they start to work against and canceling each other. This prediction corresponds with our experimental observation, in which the highest alignment accuracy is at 0.8 to 2 μl of dispensed water.

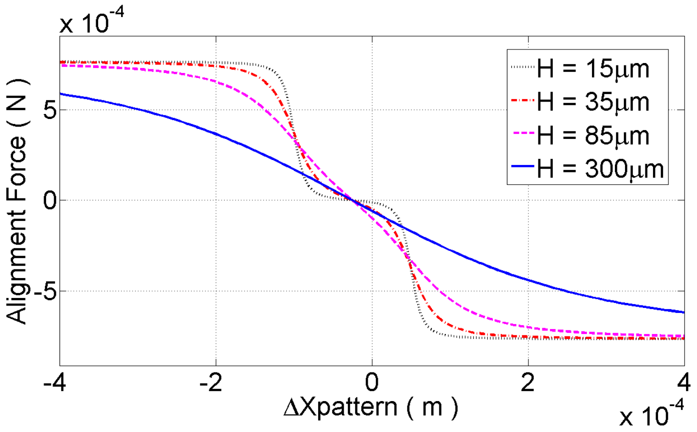


Figure 4.10 Modeled alignment force at different die-to-carrier heights H for fully-wetted SA.

Pad-Assisted SA Force

The superposition of alignment forces for the pad-assisted SA procedure is presented in Fig. 4.12 for one and four columns of 25 pads on each side of the chip (red dashed and pink dash-dotted lines) and a gap height of 35 μm. The periphery and pad forces (blue solid and gray dotted line) are shown individually for the one column case with the assumption of a constant number of bubbles. Again, the equilibrium position of the periphery case (the case if only consider the SA force from the periphery) shifts 25 μm from the pad case (the case if only consider the SA force from the pads). The superposition of the two alignment forces, actually the x-intercept of this combined force decides the final location where the chip will stay after SA.

The zero force locations of the superposition for one or four pad columns result in alignment accuracies of 1.76 μm and 0.53 μm, respectively. This is in good agreement with the observations reported in Fig. 4.4. And in experiment, better alignment accuracy was achieved. The main reason is that the frame is larger than the die so that the water menisci around the chip periphery in reality is not so precisely defined as it is now in the model; thus, the alignment force from the periphery in reality is smaller than the model calculation.

The model also predicts the need for a higher pre-alignment precision for multiple columns than for a single column because there are multiple equilibrium positions.

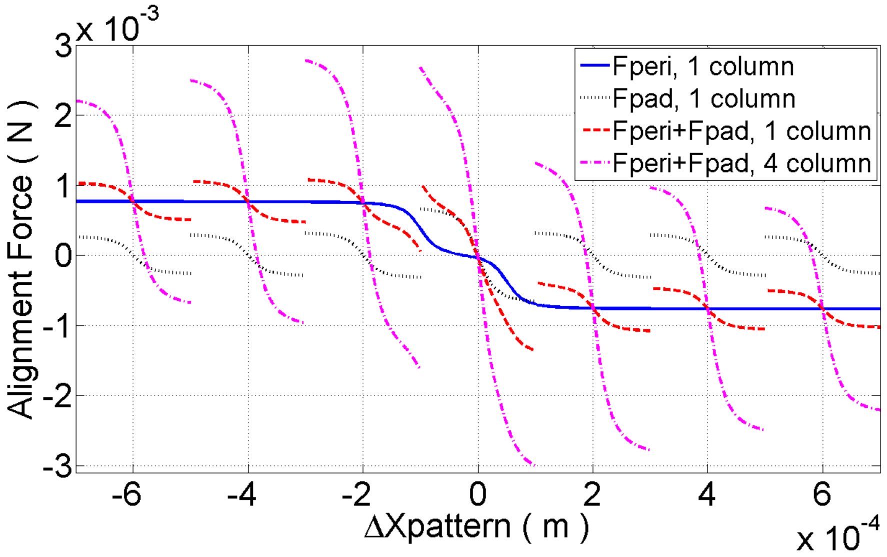


Figure 4.11 Modeled alignment forces of pad-assisted SA at a die-to-carrier height H of 35 μm when assuming one and four columns of 25 pads on each side of the chip.

Sensitivity Analysis on Pad-Assisted SA

A sensitivity analysis with dicing misalignment d, the number of bubbles formed on pads NO_{bubble}, and the die tilting error as parameters is presented in Fig. 4.13.

As expected, the offset $\Delta X_{pattern}$ of the equilibrium position (equivalently, SA error) increases with increasing d (Fig. 4.13, up). The model validates our experiment results on the great tolerance of the pad-assisted SA technique on dicing inaccuracy. Decreasing the dicing inaccuracy could improve the SA accuracy.

The total length of the functioning menisci plays a crucial role in SA (Fig. 4.13, middle). With additional pads and available bubbles, the accuracy is improved, which is consistent with our observation in Fig. 4.7 that a lower alignment accuracy was observed for large initial displacements, resulting in a low yield of bubble formation. It also suggests that with only 15 pads, a comparable sub-micron SA accuracy can be achieved already. A design rule concerning the minimum pad number to achieve a targeted SA accuracy can be proposed based on this calculation result.

The SA procedure is also quite robust against gravitational forces (Fig. 4.13, down). A tilting error of 5° only results in an additional misalignment of 1 μm.

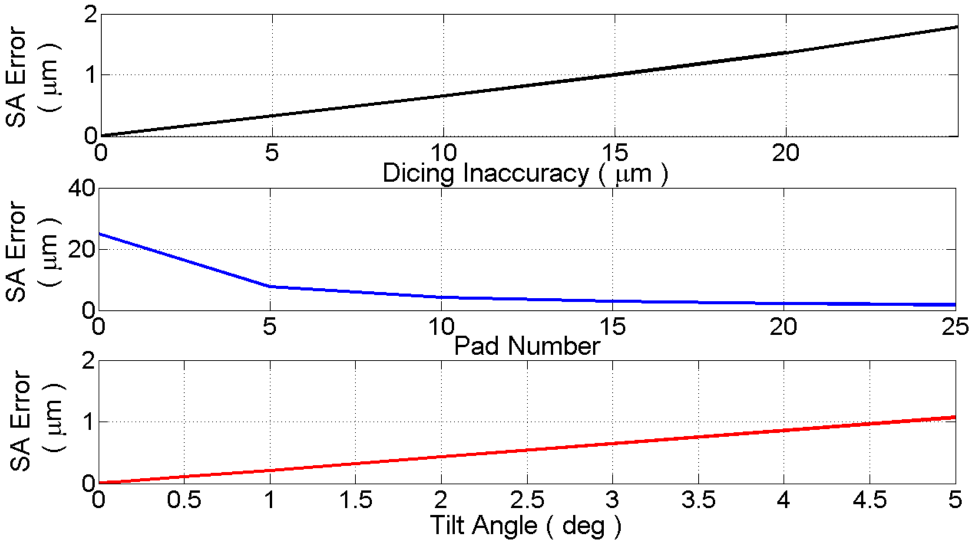


Figure 4.12 Results of the sensitivity analysis of the pad-assisted SA procedure with varying dicing accuracy, pad number, and tilting angle.

Comparison of Fully-Wet SA, Partially-Wet SA and Pad-Assisted SA

The alignment forces for the three SA processes are compared in Fig. 4.14. The design parameters for these cases are listed in Table 4.1. The equilibrium position of the fully-wetted case is shifted by the dicing misalignment of -25 μm (the case shown in Fig. 4.11 and the blue curve in Fig. 4.14-b) and exhibits a reduced slope because the carrier frame is 150 μm larger than the die. Pad-assisted and fully-wetted SA result in a steep slope close to the zero offset position which means better alignment accuracies.

The highest alignment forces close to the zero offset position are achieved by pad-assisted SA thanks to the increased effective meniscus perimeter. The chip is then guided to the equilibrium position with the largest SA force (red curve in Fig. 4.14-a). However, its equilibrium is 1.78 μm offset to the zero offset position ($\Delta X_{pattern} = 0$) (red curve in Fig. 4.14-b).

Perfect alignment can only be achieved by partially-wetted SA (see the x-intercept of the black curve in Fig. 4.14-b). The pity is that the frame around the chip periphery is not practical for standard chip fabrication.

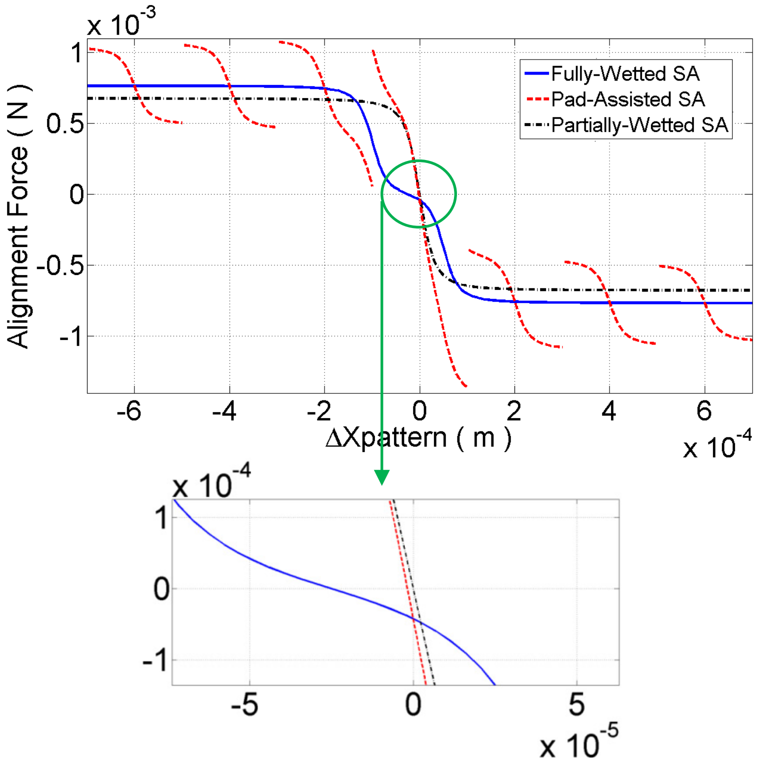


Figure 4.13 Alignment force benchmarking of individual SA procedures at a gap of 35 μm.

4.5 A Revisit of Magnetic SA in Comparison of Hydrophobic SA

From our calculations in Section 4.4, the alignment forces provided by hydrophobic SA are on the scale of 10^{-3} ~ 10^{-4} N. If we revisit Chapter 3, we find that the alignment forces, or in other words, the horizontal magnetic force, generated in each pad (a soft magnet magnetized) are on the scale of 10^{-7} N. To get a magnetic SA force comparable to the surface tension force would require thousands of pads. This is not realistic from the design aspect of view. Furthermore, these pads are already quite large, i.e. 100 μm by 100 μm.

We will also see in Chapter 3 that the magnetic force along the normal axis is on the scale of 10^{-4} N, a thousand times larger than the lateral magnetic SA force component. Then, the induced friction force will be very large as we assume a dry contact. Introducing a liquid medium could then be helpful. This was indeed tested with water, but it failed because chip flowed away from the binding site. Other lubricants than water such as optimized air pressure and vibration are known to improve magnetic assembly processes, but not tried.

To make magnetic SA usable for us, a complicated design and validation needs to be performed as many factors are to be studied, e.g. the liquid medium, the vibration amplitude and frequency, the topology of the magnets, the movement of the permanent magnets, the magnetic materials and so on. Another question for us which could be an important one is that whether the cavities are necessary. If so, additional process complexity and cost will be introduced.

To sum up the above comparison, hydrophobic SA outperforms all the other SA methods, thus really becomes the most suitable SA method for 3D integration because of its high alignment precision, low cost, high process efficiency and reliability.

4.6 Summary

The proposed pad-assisted SA procedure is compatible with standard manufacturing technology and, as we have shown, results in sub-micron pad-to-pad alignment accuracy even at large dicing misalignments. A design methodology calculating the alignment forces from meniscus growth has been proposed and validated. A compromise between small pads and a large number of pads is needed to achieve an optimal balance between the resulting accuracy and the pre-alignment precision. We have also demonstrated a successful SA with 15 μm tall Au studs for subsequent thermo-compression bonding using additional alignment patterns in the die center. To sum up, the batch-process-compatible pad-assisted SA procedure is a promising candidate to reduce the assembly cost of future 3D chip stacks.

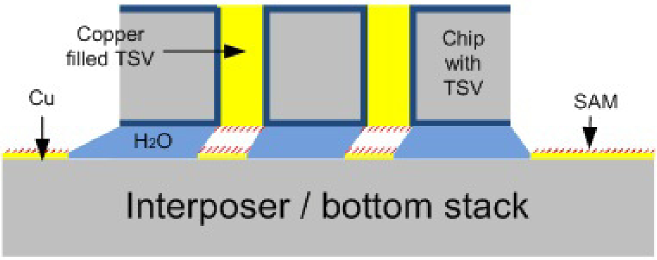
4.6.1 Improvement of Pad-Assisted Self-Alignment Process

Experiments with alternative number, dimension and placement of the pads are worthwhile to be done to finalize the design rules. Hereafter, we give some directions that should be pursued for future studies.

1) As shown in the bonding experiment, the shape of the bumps is not critical. It can be bumps or balls. The same for the material used for the bumps. Au bumps, Cu bumps, or even eutectic solder balls are all good candidates. Bumpless bonding could also be an interesting option.
2) Other metals such as silver, platinum and copper are known to support alkanethiol treatment for self-assembled monolayer growth [13]. Copper is a good replacer for gold since it is widely used in TSV filling. In addition, in our initial experiment, alkanethiol SAM modified copper surface reaches a contact angle of about 120 degrees, which is even larger than gold. But then we must prevent the surface from oxidation during processes, which is a key to ensure successful SA with copper pads.
3) Other alkanethiols with shorter chains are to be tested as well. This will lead to a study on low temperature bonding after SA is done.
4) An automatic multi-die pick and place tool is to be designed. With a rough initial placement (e.g. less than one pad), more stable and better precision can be expected.
5) This technique should work for wafer-level alignment as well.

4.6.2 Co-Integration of Pad-Assisted Self-Alignment with TSV

As mentioned before, using copper is a natural choice if we need to integrate SA structures with TSV. SAM can be grown on the backside exposed and CMPed TSVs. Moreover, no redundant process is required to deposit gold on copper. Bumpless bonding between the contacting copper surfaces can be performed to achieve interlayer metal connection. It is a very good way to reduce the footprint of TSV by removing the bigger metal bumps. Hereby, a die-level 3D integration scheme for heterogeneous 3D systems is proposed and illustrated in Fig. 4.15.



Proposed 3D integration scheme:
A silicon interposer w/o TSVs or an unfinished 3D chip stack
+ a chip with TSVs to be stacked
+ surface-tension-driven SA
+ bumpless/bump bonding

Figure 4.15 A viable die-level 3D integration scheme to achieve through silicon interconnections.

Reference

1. International Technology Roadmap for Semiconductors, 2007.
2. U. Srinivasan, D. Liepmann, and R. T. Howe, “Microstructure to Substrate Self-Assembly Using Capillary Forces,” *J. Microelectromech. Syst.,* Vol. 10, No. 1, pp. 17-24, 2001.
3. K. F. Böhringer, U. Srinivasan, and R. T. Howe, “Modeling of Capillary Forces and Binding Sites for Fluidic Self-Assembly,” *Proc. 14th Int. Conf. Micro Electro Mechanical Systems, Interlaken, Switzerland, Jan.,* 2001.
4. B. R. Martin, D. C. Furnange, T. N. Jackson, T. E. Mallouk, and T. S. Mayer, “Self-Alignment of Patterned Wafers using Capillary Forces at a Water-Air Interface,” *Adv. Func. Mater.*, Vol. 11, No. 5, pp. 381-386, 2001.
5. C.-S. Chang, R.-H. Uang, and E. Wu, “Microchip Self-Assembly on a Substrate Using Plasma Treatment,” *IEEE Trans. Adv. Packaging,* Vol. 31, No. 2, pp. 404-409, 2008.
6. T. Fukushima, T. Konno, T. Tanaka, and M. Koyanagi, “Multichip Self-Assembly Technique on Flexible Polymeric Substrate,” *Proc 58th Electronic Components and Technology Conf.*, pp. 1532-1537, 2008.
7. T. Fukushima, E. Iwata, J. Bea, M. Murugesan, K.-W. Lee, T. Tanaka, and M. Koyanagi, “Evaluation of Alignment Accuracy on Chip-to-Wafer Self-Assembly and Mechanism on the Direct Chip Bonding at Room Temperature,” *IEEE International 3D System Integration Conf.*, Munich, Germany, November 2011.
8. W. Zheng, J. Chung, and H. O. Jacobs, “Fluidic Heterogeneous Microsystems Assembly and Packaging,” *J. Microelectromech. Syst.,* Vol. 15, No. 4, pp. 864-870, 2006.
9. K. Hiratsuka, A Bohno and H Endo, “Water Droplet Lubrication between Hydrophilic and Hydrophobic Surfaces,” *J. Phys.: Conf. Series*, Vol. 89, 2007.
10. H. Sharifi, T.-Y. Choi, and S. Mohammadi, "Self-Aligned Wafer-Level Integration Technology With High-Density Interconnects and Embedded Passives," *IEEE Trans Adv. Packaging*, Vol. 30, No. 1, pp. 11-18, 2007.
11. J. Dalin and J. Wilde, " Self-Assembly of Micro-Parts Using Electrostatic Forces and Surface Tension," *Proc 59th Electronic Components and Technology Conf.*, pp. 1517-1524, 2009.

12. A. J. Nichol, P. S. Stellman, W. J. Arora, G. Barbastathis, "Two-Step Magnetic Self-Alignment of Folded Membranes for 3D Nanomanufacturing," *Microelectronic Engineering*, Vol. 84, pp.1168–1171, 2007.
13. J. Christopher Love, Lara A. Estroff, Jennah K. Kriebel, Ralph G. Nuzzo, and George M. Whitesides, "Self-Assembled Monolayers of Thiolates on Metals as a Form of Nanotechnology," *Chem. Rev.*, Vol. 105, pp. 1103-1169, 2005.

Chapter 5

Inter-Layer Serial Links for 3D Chip Integration

5.1 Introduction

Over years, multiple types of links have been proposed for on-chip communication [1, 2]. One of them, known as "synchronous parallel link", transmits data-words of fixed size along with the clock required for synchronization. Unfortunately, apart from very high demand on area for large data-words, this type of link increasingly suffers from synchronization issues as the size of data-word increases. In order to battle the latter problem, "asynchronous parallel link" has been proposed [2]. However, in the asynchronous case, even though there is no need for any clock transmission, additional circuitry required for synchronization may infringe the overall system performance.

Replacing these long 2D connections with short TSVs, synchronization problem is expected to be relieved. However, as we will see in Section 5.2.1, the silicon area occupied by wide parallel TSVs are prohibitive, as the TSV dimension in use nowadays is still large (20-30 μm in diameter). Another point is that, parallel links usually work at low clock frequencies, much lower than what TSV supports. Therefore, using TSVs in parallel is quite unproductive.

Serializing the parallel data using serializer and deserializer (SerDes) and transmitting the serialized data through a single TSV at a much higher data rate becomes a possible approach to save the large area consumed by the TSVs.

TSV characteristics will be first clarified and then some post-processing design rules for TSV insertion will be proposed in Section 5.2. In Section 5.3, through the implementation of a 2-Gb/s inter-layer serial link which is composed of an 8:1 multiplexer (MUX), a 1:8 demultiplexer (DEMUX) and TSVs, area benefit gained through serialization will be demonstrated. Placing such multiple links in parallel, a new concept of link referred as "inter-layer quasi-serial link" will be proposed in Section 5.4. For both cases, the large transistor density offered by 90 nm CMOS technologies is exploited. We would expect more benefits if these could follow the trend in miniaturization driven by more advanced CMOS processes, since most, if not all, TSV fabrication technologies are not linked with the actual transistor scaling. Within the limited time frame of this thesis, serial links working under higher frequencies will not be implemented physically because more complicated and advanced circuit design techniques are required [9, 10, 12, 13, 14, 15]. Instead, a theoretical model will be proposed in Section 5.5 based on the design principles of these links to fully study the area benefit that could be achieved by serialization.

5.2 TSV Characteristics and Performance

5.2.1 TSV Area

A simplified TSV structure shown in Fig. 5.1 illustrates the footprint of a TSV. The preferred via size now is less than 30 μm, eventually, it will go down to 5 μm for memory-on-logic applications and even 2 μm for heterogeneous integration. The via size should be as small as possible. The pitch between two TSVs is mainly decided by the bonding technology because the bumps are usually larger than the interconnects.

Aside from the physical dimensions, the Keep-Out Zone (KOZ) or Keep-Out Distance (KOD) should also be considered. Because TSVs generate some thermal-mechanical stress to the substrate, and the longitudinal and transverse stresses influence the carrier mobility near the vicinity of the TSV [16], active devices are not to be placed in the KOZ in order to ensure device reliability. A CMOS transistor should be kept about 6.7 μm away from the 20 μm-diameter TSV [19]. And the KOZ is mainly decided by the diameter of the TSV, rather than the height [17].

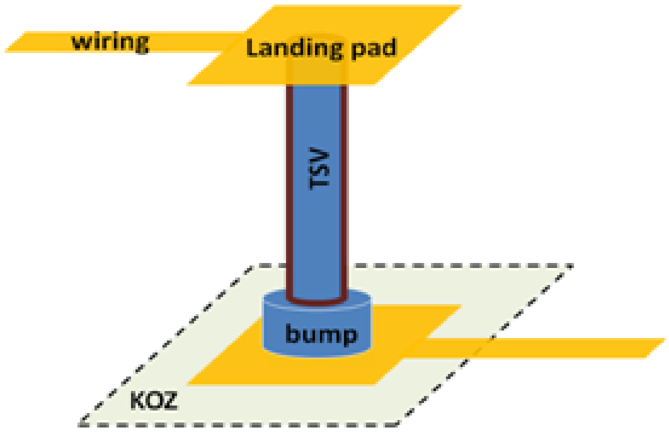


Figure 5.1 The comings of TSV area consumption.

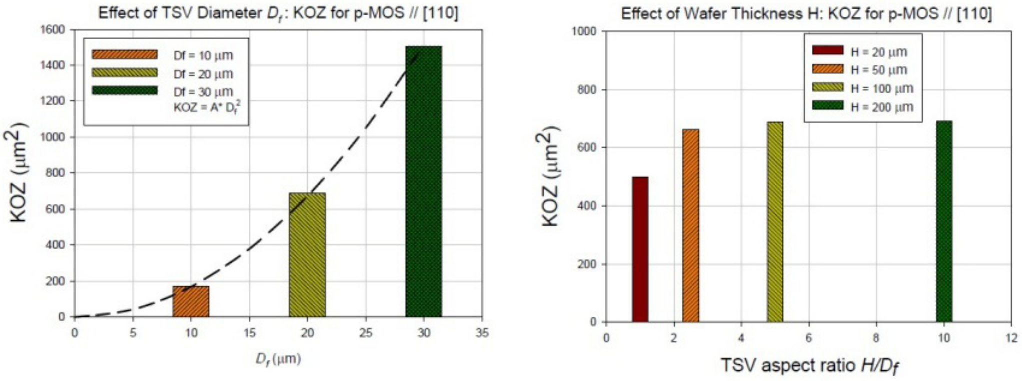


Figure 5.2 KOZ for TSVs of different sizes. Courtesy of University of Texas.

As will be mentioned in Chapter 6, a scaled version of Sun's Rock core in 32 nm technology would consume 3.5 mm^2 area. If one TSV consumes 40 μm × 40 μm area (diameter 20 μm, pitch 40 μm or KOZ 6.7 μm), then the same area for one core can only hold about 2200 TSVs. So, for the current TSV technology (20 - 30 μm diameter), finding a technique to decrease the required TSV number is of great significance.

5.2.2 TSV Electrical Performance

A cross-section view of 3D interconnection structures is shown in Fig. 5.3. For a ground-signal (G-S) TSV pair, a lumped model shown in Fig. 5.4 can be proposed. Because the bumps' size dominates the pitch between TSVs, they are included in this model.

Given the physical dimensions and material properties, the parameters in the lumped model can be calculated using Equation (1) - (7). Two TSV diameters were chosen for the discussion: 20 μm (represents the state of the art TSV technology) and 5 μm (represents the future TSV technology for 3D chip integration). The parameters for the bumps were set following the numbers proposed in [18]. All the dimensions used in the model are listed in Table 5.1.

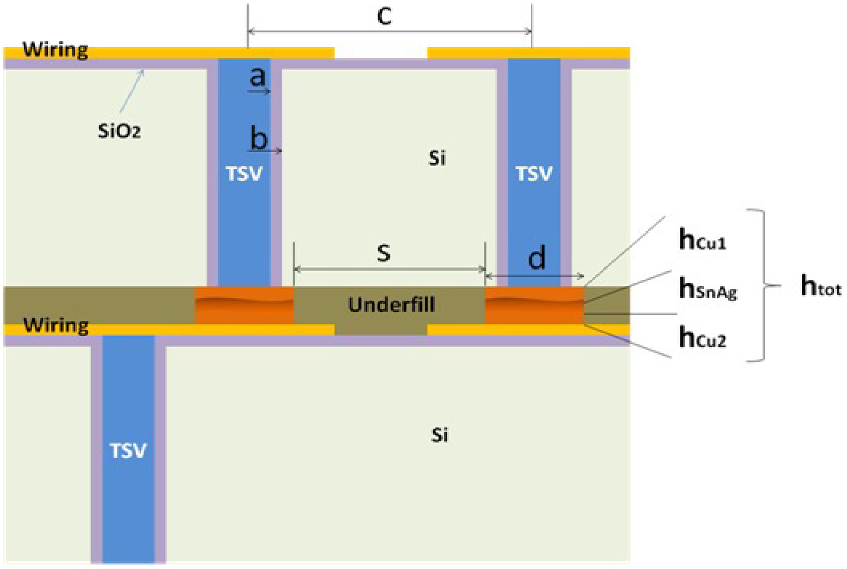

c
Wiring
a
SiO2
b
TSV
Si
TSV
s
d
hCu1
hSnAg
htot
hCu2
Underfill
Wiring
TSV
Si

Figure 5.3 Cross-section illustration of 3D interconnections including TSVs and bumps.

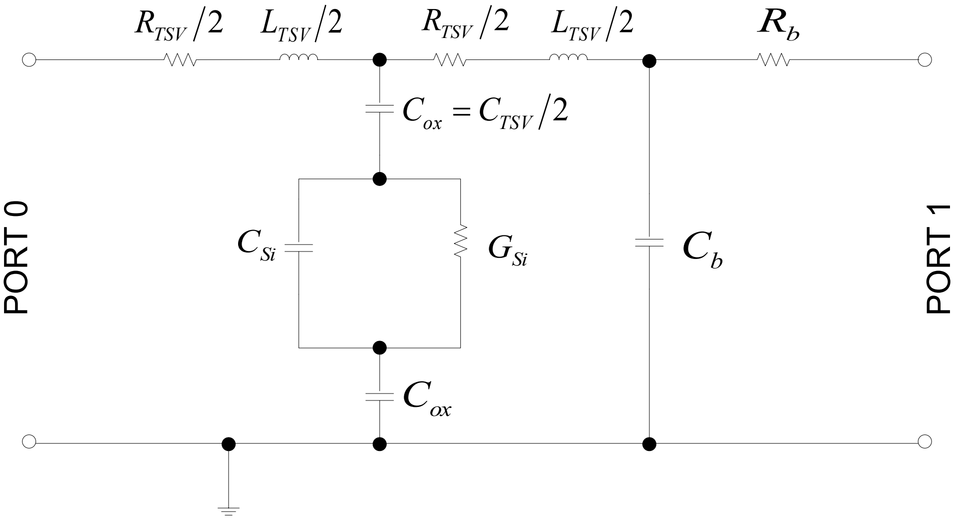

$R_{TSV}/2$
$L_{TSV}/2$
$R_{TSV}/2$
$L_{TSV}/2$
R_b
$C_{ox} = C_{TSV}/2$
PORT 0
C_{Si}
G_{Si}
C_b
PORT 1
C_{ox}

Figure 5.4 Lumped model of Ground-Signal (G-S) 3D interconnection pair.

Table 5.1 Studied TSV Geometries.

Dimensions (unit: μm)	Definition	State of the art	Future
a	Radius of copper core	10	2.5
b	Radius of TSV	11	2.6
c	TSV pitch	100	10
h	TSV height	50	15
d	Bump diameter	35	5
s	Spacing between bumps	100	0
h_{Cu1}	thickness of first copper layer	17	0
h_{SnAg}	thickness of SnAg layer	15	0
h_{Cu2}	thickness of second copper layer	0	0
h_{tot}	$h_{Cu1} + h_{SnAg} + h_{Cu2}$	32	0

1) TSV part:

TSV resistance R_{TSV}: $R_{TSV} = \rho_{Cu}\frac{h}{\pi a^2}$ (1)

TSV oxide capacitance C_{TSV}: $C_{TSV} = \frac{2\pi\varepsilon_{ox}h}{\ln(b/a)}$ (2)

TSV inductance L_{TSV} [19]:

$$L_{TSV} = \frac{\mu_0 h_{TSV}}{2\pi}\left\{\begin{array}{l}\left[\ln\left(\frac{h}{a}+\sqrt{\left(\frac{h}{a}\right)^2+1}\right)+\frac{a}{h}-\sqrt{\left(\frac{a}{h}\right)^2+1}\right]\\ -\left[\ln\left(\frac{h}{c}+\sqrt{(\frac{h}{c})^2+1}\right)+\frac{c}{h}-\sqrt{(\frac{c}{h})^2+1}\right]\end{array}\right\} \tag{3}$$

2) Substrate coupling parameters are developed based on the model proposed in [20] :

Silicon substrate capacitance C_{Si}: $C_{si} = \frac{\pi\varepsilon_{si}h}{\ln[(c/2b)+\sqrt{(c/2b)^2-1}]}$ (4)

Silicon substrate conductance G_{Si}: $G_{si} = \frac{\sigma_{si}C_{si}}{\varepsilon_{si}} = \frac{\pi\sigma_{si}h}{\ln[(c/2b)+\sqrt{(c/2b)^2-1}]}$ (5)

3) Bump part:

Bump resistance R_b:

$$R_b = R_{Cu1} + R_{SnAg} + R_{Cu2} = \rho_{Cu1}\frac{h_{Cu1}}{\pi(d/2)^2} + \rho_{SnAg}\frac{h_{SnAg}}{\pi(d/2)^2} + \rho_{Cu2}\frac{h_{Cu2}}{\pi(d/2)^2} \tag{6}$$

Bump capacitance C_b: $C_b = \frac{2\pi\varepsilon_r\varepsilon_0 h_{tot}}{\ln(\frac{S+d/2}{d/2})}$ (7)

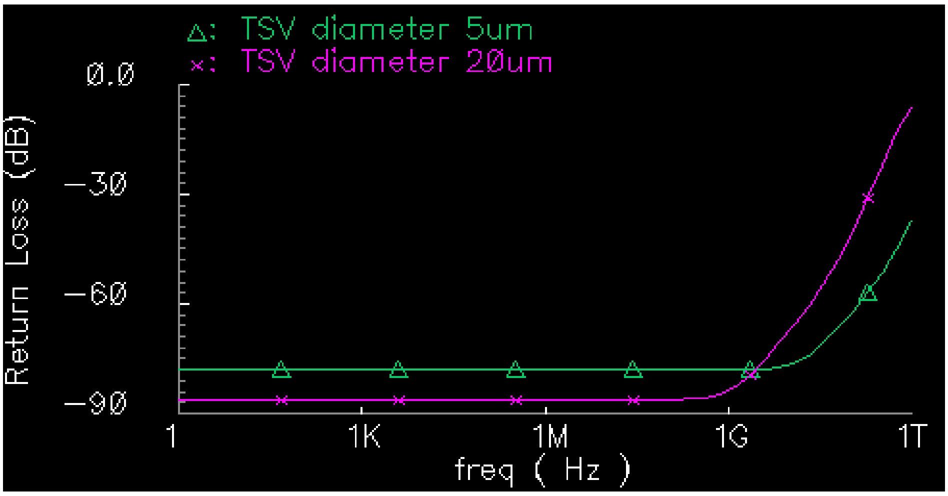
where ε_{ox} is the relative permittivity of silicon dioxide, 4.2; ε_r is the relative permittivity of underfills, 3.3; σ_{si} is the conductivity of silicon; ε_0 is vacuum permittivity; $\rho_{Cu}$is the resistivity of copper; ρ_{SnAg} is the resistivity of the bump.

Based on the above parameters, a two-port simulation was done using Spectre. 50 Ohms impedance matching was assumed in the simulation, so we were actually looking at an ideal case in which the resulting return loss and insertion loss of the G-S TSV pair only counted for the TSV structure itself. As shown in Fig. 5.5, TSVs of both dimensions can provide ultra-high bandwidth in this ideal case. Obviously, since we use a quite low frequency equivalent representation of the TSVs (we ignore magnetic induction since we do not solve Maxwell equations), the model is not accurate above some dozen of GHz. Nevertheless, the trend figured out from Fig. 5.5 demonstrates that the TSV of 5-μm diameter is of less return and insertion loss than its 20-μm diameter counterpart, and thus can support higher bandwidths. So, in reality, the distribution layer design will decide the final TSV bandwidth. In literature, TSV connections have been reported to work under 20 GHz [25], 60 GHz [24], and even 170 GHz [23].

As shown in Table 5.3, TSVs' capacitance is about 25 times smaller than the I/O capacitance of a standard I/O buffer in 90 nm technology. Using TSVs instead of off-chip connections, CMOS drivers could be sufficient and the use of power-hungry I/O drivers, such as low-voltage differential signaling (LVDS) or Stud Series Terminated Logic (SSTL) drivers, could be avoided. To demonstrate this speculation, we further simulated two parallel TSVs (20 μm diameter, 40 μm pitch) in time domain with only CMOS drivers. Here we considered a 50 μm MET2 wiring between TSVs and drivers on both sides to get a more realistic analysis. All the parasitics for MET2 wiring, i.e. Rw (resistance of the wire), Cs_w (self-capacitance of the wire), Cc_w (coupling capacitance between two wires), are in 90 nm technology. Wirings are assumed to be of minimum width. Cc_w is neglected, since the two wires can be far away from each other when routed to TSVs. The timing model can thus be illustrated as shown in Fig. 5.6. To get the minimum delay from the 50 μm wiring metals, a 20× inverter INV20 replace the first minimum-sized inverter INV1. Both INV1 and INV20 are from Faraday 90 nm standard cell library.

The simulation result is shown in Fig. 5.7. The delays with and without TSVs are summarized in Table 5.2. The addition of TSVs leads to both a degradation of rise and fall time and a larger crosstalk between the signals. Using stronger driver INV20 instead of INV1 enables to restore timing responses, although it brings larger crosstalk during switching.

When working on higher frequencies, careful interconnect design and optimization is required. For video processing and other high-bandwidth requirements, a coaxial interconnect design, with each signal TSV surrounded by four ground/buffer TSVs, may be required. Xu *et al.* [21] performed a good study about TSV couplings with various position arrangements. The TSV topology providing the least coupling is shown in Fig. 5.8 [21].



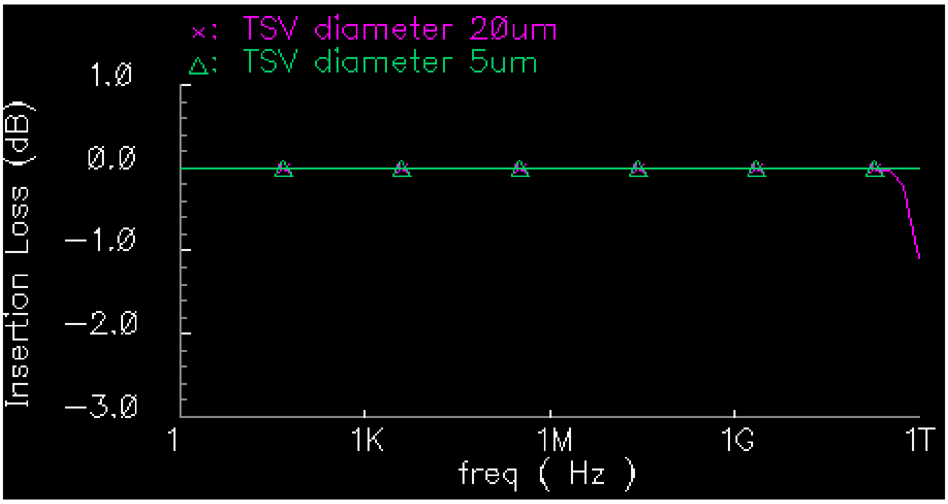


Figure 5.5 Return loss and insertion loss of a Ground-Signal TSV pair with ideal impedance matching under upto 1 THz.

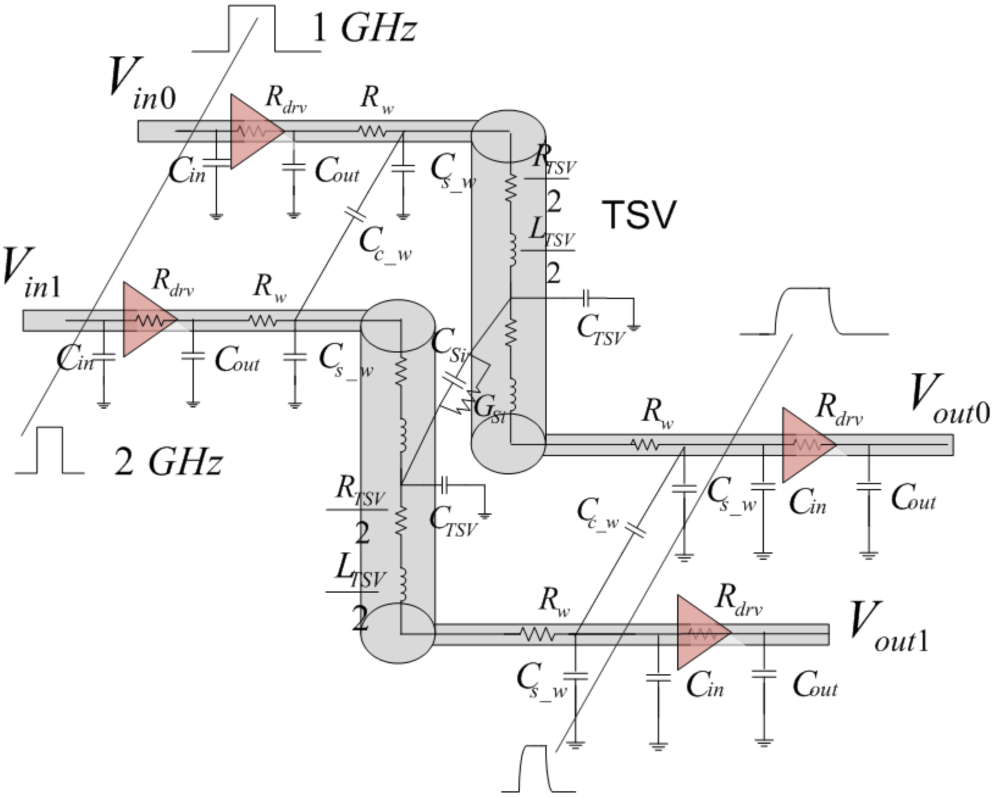

1 GHz
V_{in0}
R_{drv}
R_w
C_{in}
C_{out}
C_{s_w}
R_{TSV}
2
L_{TSV}
2
TSV
C_{c_w}
V_{in1}
R_{drv}
R_w
C_{TSV}
C_{in}
C_{out}
C_{s_w}
C_{Si}
G_{Si}
2 GHz
R_w
R_{drv}
V_{out0}
R_{TSV}
2
C_{c_w}
C_{s_w}
C_{in}
C_{out}
C_{TSV}
L_{TSV}
2
R_w
R_{drv}
V_{out1}
C_{s_w}
C_{in}
C_{out}

Figure 5.6 TSV timing model.

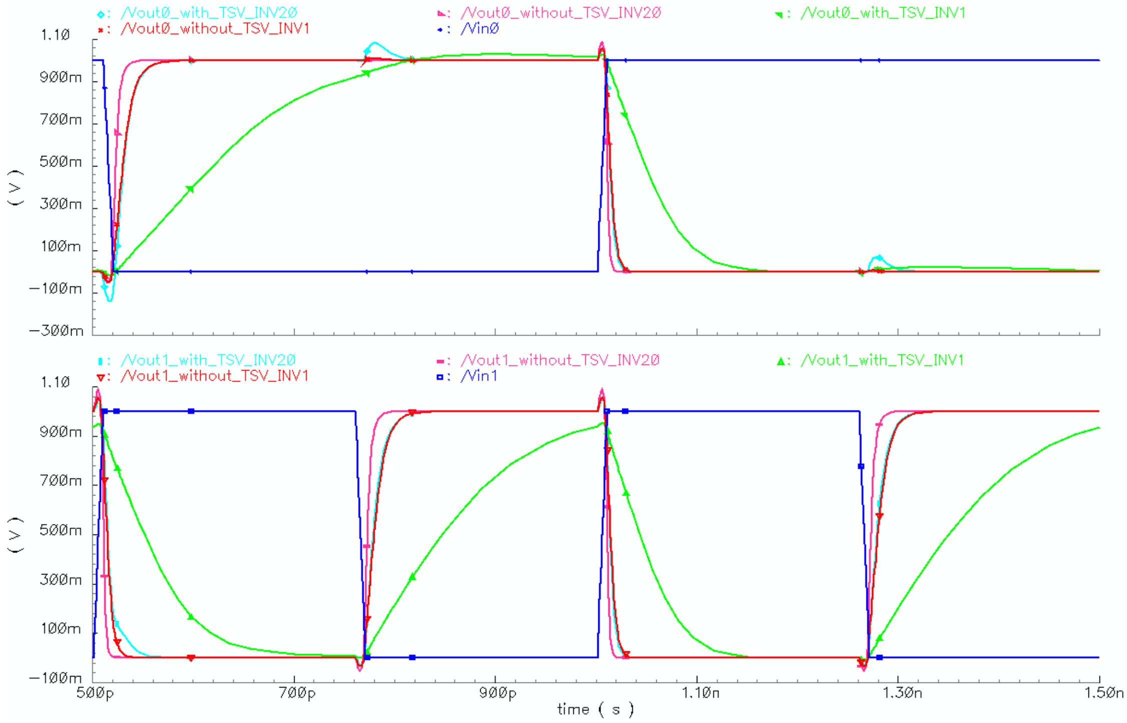


Figure 5.7 Timing performance of two identical 3D connections separately working under 1 GHz and 2 GHz.

Table 5.2 Delay with or w/o TSV.

Driver	Delay (ps)	
	Without TSV	*With TSV*
INV1	11.12	63.07
INV20	6.29	10.45

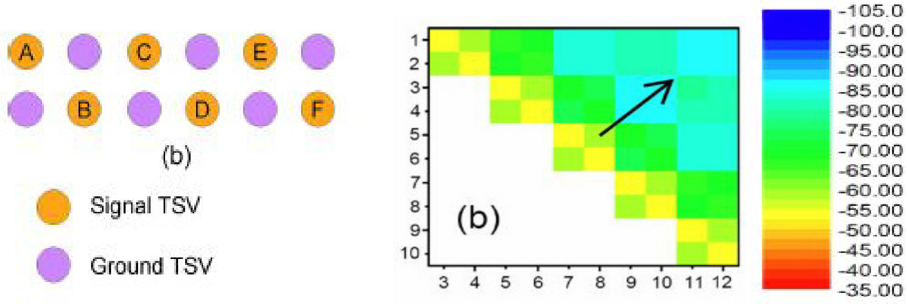

A
B
C
D
E
F
(b)
Signal TSV
Ground TSV
1
2
3
4
5
6
7
8
9
10
3
4
5
6
7
8
9
10
11
12
(b)
-105.0
-100.0
-95.00
-90.00
-85.00
-80.00
-75.00
-70.00
-65.00
-60.00
-55.00
-50.00
-45.00
-40.00
-35.00

Figure 5.8 TSV arrangement providing the least coupling.

5.2.3 TSV I/O Power Estimation

In Section 5.2.2, through the timing analysis of TSVs, it has been demonstrated that CMOS drivers are sufficient to drive TSV connections. The scenarios for off-chip connections and on-chip TSV connections are illustrated in Fig. 5.9. The corresponding values are listed in Table 5.3.

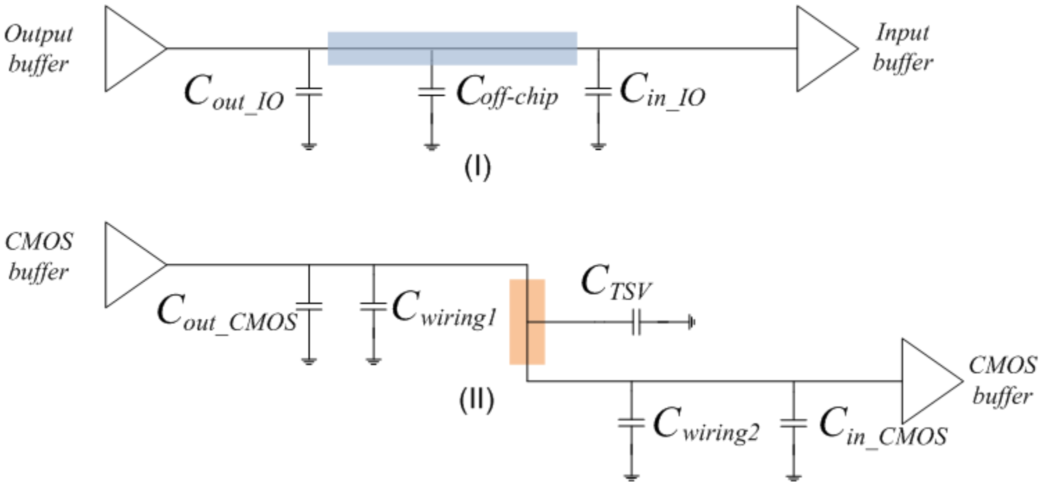

Output buffer
C_{out_IO}
$C_{off\text{-}chip}$
C_{in_IO}
Input buffer
(I)
CMOS buffer
C_{out_CMOS}
$C_{wiring1}$
C_{TSV}
(II)
$C_{wiring2}$
C_{in_CMOS}
CMOS buffer

Figure 5.9 Scenarios for dynamic power consumptions of (I) Off-Chip I/O connection; (II) On-chip TSV connection.

Table 5.3 TSV Capacitances and I/O Capacitances.

Type	Value
$C_{out\_I/O}$	2.5 pF
$C_{in\_I/O}$	2.5 pF
$C_{off-chip}$	5 pF
C_{TSV} i	80 fF
C_{out_CMOS} ii	1-10 fF
C_{in_CMOS} ii	45 fF
$C_{wiring1}$, $C_{wiring2}$ iii	100 fF

i. Corresponding to TSV dimension: 5 µm diameter, 15 µm height, 0.1 µm SiO_2 isolation, 10 µm pitch.
ii. An inverter in 90 nm technology.
iii. 250 µm global wiring in 90 nm technology.

The dynamic power consumption of the off-chip connection in Scenario I, P_{IO}, can be calculated as follows:

$$P_{IO} = \alpha\left(C_{out\_IO} + C_{off-chip} + C_{in\_IO}\right){V_{IO}}^2 f$$

where α is the activity factor, C_{out_IO} is the intrinsic capacitance of the output I/O buffer, $C_{off-chip}$ is the capacitance of on-board connection between two chips, C_{in_IO} is the input capacitance of the input I/O buffer, V_{IO} is the core power supply voltage, which is 2.5 V in Faraday 90 nm technology, f is the working frequency of the off-chip connection.

The dynamic power consumption of the on-chip TSV connection in Scenario II, P_{TSV}, can be calculated as follows:

$$P_{TSV} = \alpha\left(C_{out\_CMOS} + C_{wiring1} + C_{TSV} + C_{wiring2} + C_{in\_CMOS}\right){V_{core}}^2 f$$

where α is the activity factor, C_{out_CMOS} is the intrinsic capacitance of the CMOS driver, C_{TSV} is the capacitance of a TSV, C_{in_CMOS} is the input capacitance of the CMOS driver, V_{core} is the core power supply voltage, which is 1.0 V in Faraday 90 nm technology, f is the TSV connection's working frequency. The capacitances of the 250 μm global interconnect wiring to a TSV on both sides, $C_{wiring1}$and $C_{wiring2}$, are about 100 fF [22]. Relatively long on-chip wirings (250 μm) to the TSV are assumed here in order to get a more pessimistic result for the power estimation of the TSV on-chip connection.

Assuming the activity factors and the working frequencies for both scenarios are the same, an on-chip TSV connection will only consume about 1% of the dynamic power consumed by its off-chip counterpart. This estimation result demonstrates the potential power saving and high bandwidth gained through ultra-wide I/O interface using TSVs.

5.2.4 Summary

To summarize this section, we can first conclude that based on an "electrical" model, TSVs can provide very high bandwidths even over dozen of gigahertz, and smaller TSVs should provide better electrical performance. Secondly, we notice that TSVs consume large area and this constraint should be seriously considered in circuit and system design.

5.3 A 2-Gb/s Inter-Layer Serial Link

The proposed architecture for the 2-Gb/s inter-layer serial link is shown in Fig. 5.10. The serial link operates at 2 Gb/s and is designed based on existing solutions detailed in [4] (3-Gb/s examples can also be found in [5]).

5.3.1 Circuit Design for Each Tier

The 8:1 MUX schematic is depicted in Fig. 5.11, as proposed in [4]. The circuit generates a high-level pulse of the duration $1/f_{out}$ (f_{out}: MUX's output frequency). The pulse is transmitted through eight shift-registers (labeled as S1 to S8) synchronized with the bit clock of 2 GHz. Data latched in the load registers (labeled as L1 to L8) are extracted and sampled by a retiming flop. The circuit has been implemented in 90 nm CMOS using basic components from a Faraday commercial standard cell library. The back-end phase has been carried out with SoC encounter 7.1 by Cadence®. The entire back-end design including placement, clock tree insertion, and routing has been performed. Finally, a sign-off parasitic extractor has been used to verify the circuit timing after layout. The propagation delay along the critical path in worst-case Process Voltage Temperature (PVT) simulation is 588 ps, which corresponds to a maximum operating frequency of 1.7 GHz. This requires slowing down the NoC clock to 212 MHz (212 MHz = $1.7/2^3$ GHz, 3 represents $8(2^3)$:1 serialization ratio). In typical corner, the circuit has a critical path of 500 ps, which complies with the 2 GHz specification. The total area is 22 μm by 20 μm (440 μm^2).

The clock divider is implemented by two latches, arranged in a master-slave configuration. This building block performs a division by two, while further division may be obtained by connecting by-two-dividers in series. 1 GHz, 500 MHz and 250 MHz clock signals are all generated in this way. The 300 ps skew existing between the 2 GHz and 250 MHz clocks does not hamper the MUX's functionality. The area of a by-two-divider is 21.2 μm^2, and the circuit can operate at its working frequency in the slow corner.

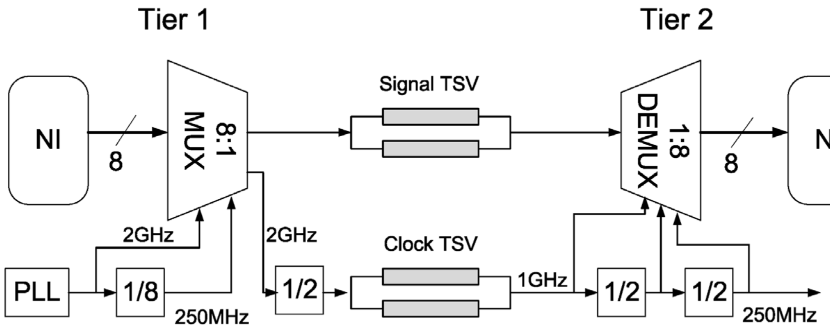

Tier 1
Tier 2
NI
8
8:1
MUX
Signal TSV
1:8
DEMUX
8
NI
2GHz
2GHz
Clock TSV
PLL
1/8
250MHz
1/2
1GHz
1/2
1/2
250MHz

Figure 5.10 Architecture of the 2-Gb/s inter-layer serial link. "NI" stands for "Network Interface".

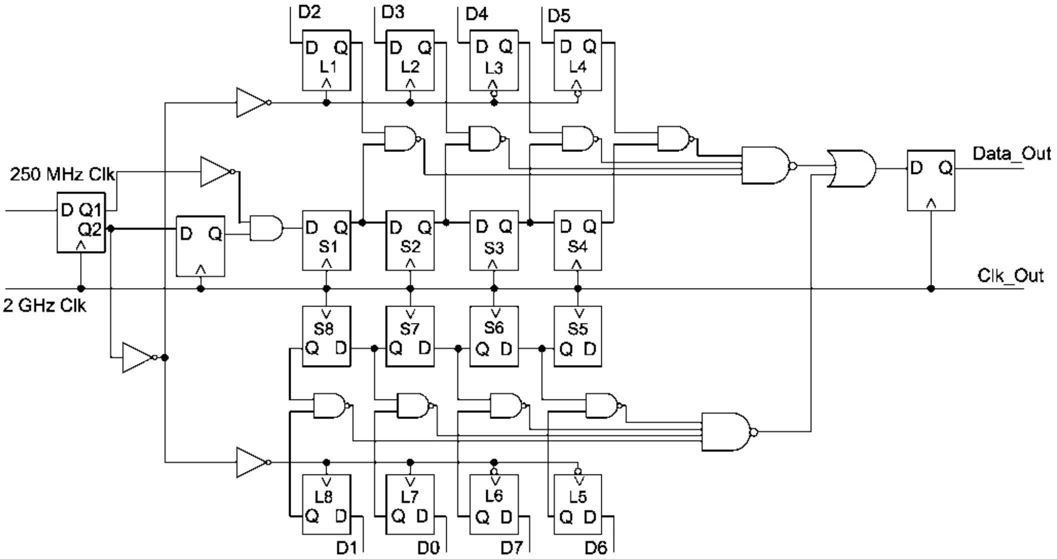

D2
D3
D4
D5
D Q
L1
L2
L3
L4
250 MHz Clk
D Q1
Q2
D Q
S1
S2
S3
S4
Data_Out
Clk_Out
2 GHz Clk
S8
S7
S6
S5
Q D
L8
L7
L6
L5
D1
D0
D7
D6

Figure 5.11 Schematic of the 8:1 MUX.

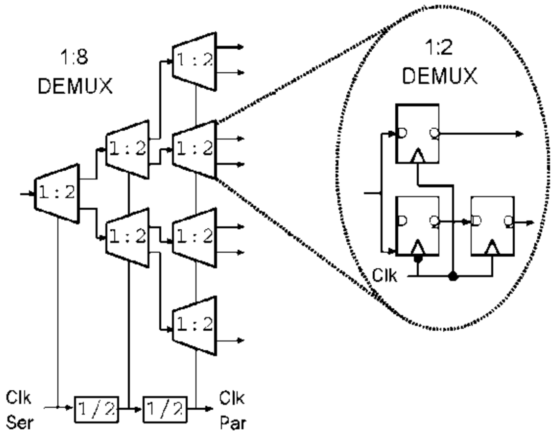

1:8
DEMUX
1 : 2
1:2
DEMUX
Clk
Clk
Ser
1/2
1/2
Clk
Par

Figure 5.12 Schematic of the 1:8 DEMUX.

The 1:8 DEMUX is designed based on the circuit proposed in [5]. The circuit schematic is depicted in Fig. 5.12. It consists of a tree of 1:2 DEMUXes. Each tree level operates at different frequency. Fig. 5.12 also shows the circuit diagram of the basic building block, a 1:2 DEMUX, which is formed by a rising-edge triggered flop and a falling-edge triggered flop followed by a rising-edge triggered flop, which in turn samples on the falling edge and outputs data on the rising edge. The DEMUX circuit has been implemented in 90 nm technology. The design could demultiplex 4-bit data at 8 Gb/s, synchronized with 1 GHz clock in the slow corner. The layout area is 10 μm by 44.32 μm (443.2 μm^2).

Considering how TSVs can be integrated into the design, we made some simplifications.

- TSV's height was limited to approximately 50 μm, much shorter than the wavelength associated with the 1 GHz signal, so a matching network was not required.
- Just used CMOS drivers for TSVs, no power-hungry standard I/O drivers was used.

5.3.2 TSV Insertion and Layout Simulation

Fig. 5.13 illustrates the design rules of TSV insertion. The diameter and spacing of TSVs used in this design are both 20 μm (the diameter here can be viewed as the footprint of the TSV including the bumps). TSV connections from Tier2 land on the top of the metal pads in Tier 1 while TSVs to connect Tie 1 are fabricated from the chip backside and landed on MET2 layer of Tier 2.

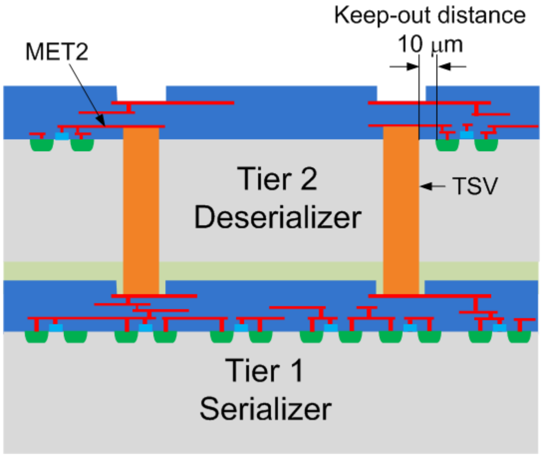


Figure 5.13 Design rules concerning TSV insertion.

MUX is assigned on Tier 1 and DEMUX is Tier 2. For the MUX, because TSV landing pads are placed on the top metal layer, all the layers under them are free for routing and placing active devices (e.g. transistors). In this implementation, we have placed the MUX core and power rings under the TSV landing pads. For the DEMUX in Tier 2, because metal2 has been used for TSV backside landing, active devices are placed away from the TSV connection region. A keep-out distance of 10 μm is kept between TSV and the adjacent transistors in order to be consistent with Keep-Out Zone (KOZ) analysis in Fig. 5.2 and to be immune to electrical coupling as well as to misalignments inherent to TSV post-processing.

Investigation of the solution scheme discussed so far will be assessed through extensive simulations that will be carried out with commercial EDA tools.

First, a LEF file with specifications to define TSV landing pads as input/output is generated. A redundant TSV is arranged for each signal considering the currently not so ensured yield of TSVs. Power and ground connections are also placed in order to provide a more realistic case of study. Finally placement and routing of the design with the TSV pins specified as inputs or outputs of the circuit are done.

5.3.3 Results

The final layouts for the 8:1 MUX and 1:8 DEMUX are shown in Fig. 5.14. The areas for each components and the total silicon area for the link are summarized in Table 5.4. TSV array's area includes the 20 μm spacing between TSVs and the 10 μm blocking distance from TSV to active devices. So, a single TSV's area is considered as given by 40 μm by 40 μm (1600 μm^2).

Since the entire system has been designed using the same commercial standard cell library, we experience some performance loss due to the 1.7 GHz maximum operation frequency for the 8:1 MUX which in turn limits the NoC clock to 212.5 MHz. A careful full-custom design of flops and combinational gate forming the system would allow further improvement.

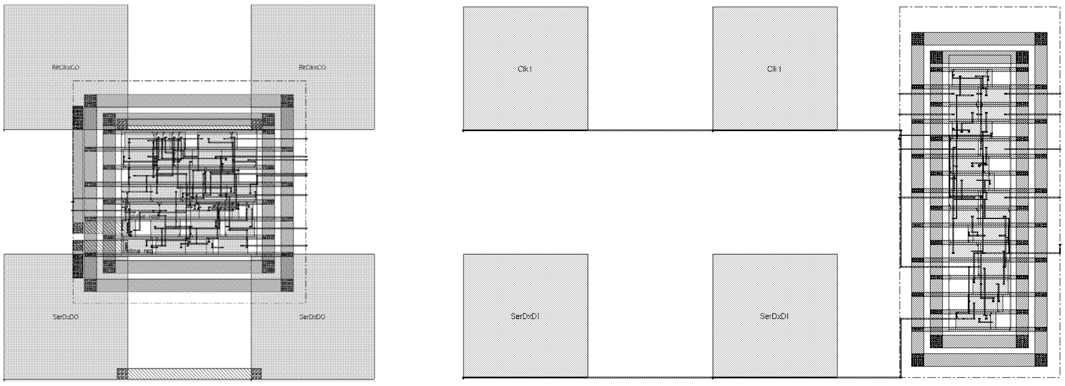


Figure 5.14 Layouts for 2-Gb/s serial link. Layout of a) the 8:1 MUX; b) the 1:8 DEMUX.

Table 5.4 Area Summary for the 2-Gb/s Serial link and its Parallel Counterpart.

Component	Area (μm^2)	
	Serial link	*8-bit Parallel Link*
8:1 MUX	440.0	0
1:8 DEMUX	443.2	0
Clock Divider	21.2	0
Single TSV	1600	1600
TSV Number	2×1×TSV_signal+2×1×TSV_clk = 4×TSV	2×8×TSV_signal+2×TSV_clk = 18×TSV
TSV Array	6400	28800
Total Link (silicon)	7304.4	28800
Area %	25.36%	100%

5.4 A 8-Gb/s Inter-Layer Quasi-Serial Link

The complete system architecture for the 8-Gb/s quasi-serial link is depicted in Fig. 5.15. The main architecture parameters are set according to the Scalable Communication Core (SCC) architecture from Intel [3]. SCC implements a 32-bit physical transfer digit and the NoC operates at 250 MHz. These parameters are applied to our design.

Four 2-Gb/s serial links that have been designed in Section 5.3 are now placed in parallel and therefore will act as a 8-Gb/s quasi-serial link. Such a network bandwidth (or router link width, 64 bits) satisfies the need of NoCs such as SCC. To accommodate with this architecture, clock signal must be synchronized and so clock TSVs are shared among the four 2-Gb/s serial links. In addition, the redundancy of the clock TSV was increased to four. The design rules are the same as in Section 5.3.

The area summary of the link is shown in Table 5.5. The large TSV (20 μm-diameter) is not suitable for parallel link architecture because a 32-bit interlayer bus with double-redundancy would occupy approximately 0.1 mm^2. This would correspond to 40K gates in 90 nm Faraday CMOS standard cell library [7]. This gate density corresponds roughly to the area of a modern embedded processor if it is optimized for area. For instance, the ARM9 architecture synthesized in 90 nm TSMC library [8], occupies 0.2 mm^2, which is only two times larger than an array of sixty-four TSVs. The situation is exacerbated due to the presence of multiple parallel buses in a single chip.

On the other hand, a quasi-serial link greatly reduces the area consumption by almost 80%. The final area of the link is equivalent to approximately 8K gates making the design economically viable. Nevertheless, a shift towards more advanced processes (density of 1 Mgate/mm^2 or more) may support more aggressive serialization such as 16:1 MUX and 1:16 DEMUX.

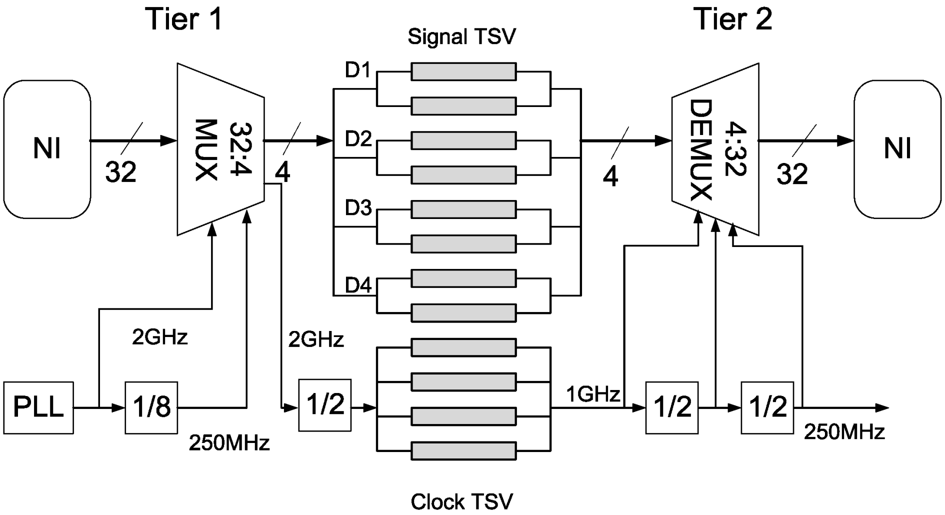


Figure 5.15 Architecture of 8Gb/s quasi-serial link.

Table 5.5 Area Summary for the 8-Gb/s Serial link and its Parallel Counterpart.

Component	Area (μm^2)	
	Quasi-Serial Link	*32-bit Parallel Link*
32:4 MUX	1282.6	0
4:32 DEMUX	1185.4	0
Clock Divider	21.2	0
Single TSV	1600	1600
TSV Number	2×4×TSV_signal+4×1×TSV_clk =12×TSV	2×32×TSV_signal+2×TSV_clk = 66×TSV
TSV Array	19200	105600
Total Link (silicon)	21689.2	105600
Area %	20.54%	100%

5.5 Inter-Layer Serial Link Design Methodology

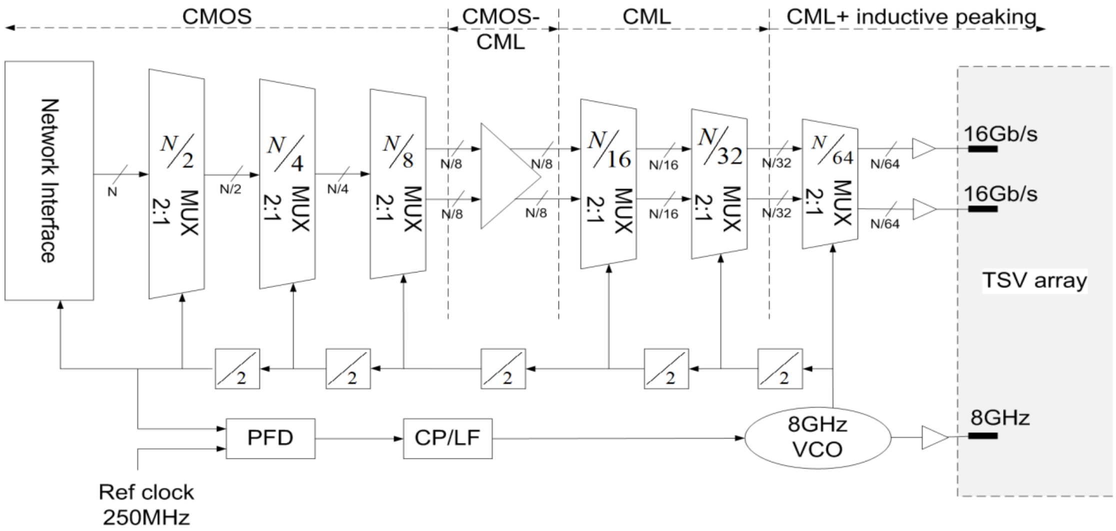
In the previous sections, it has been demonstrated that serial and quasi-serial links bring great benefits on area and design complexity. In this section, the discussion will further continued on the serial links working under higher frequencies to fully exploit the high bandwidth supported by TSVs. As mentioned in the beginning of this chapter, the final physical implementation of the links working under higher frequencies will not be our aim due to the limited time frame of the thesis. Instead, we developed a theoretical model to study the links' area.

The whole link architecture is illustrated in Fig. 5.16. The maximum serialization ratio is 64:1. Higher than 64:1, or in other words, the link width is larger than 64, the link will be implemented as quasi-serial links. As the basic clock frequency is set at 250 MHz, the highest data rate to be achieved with 64:1 MUX will be 16 Gb/s. The reasons for this design and the way we build the link area model will be explained as following:

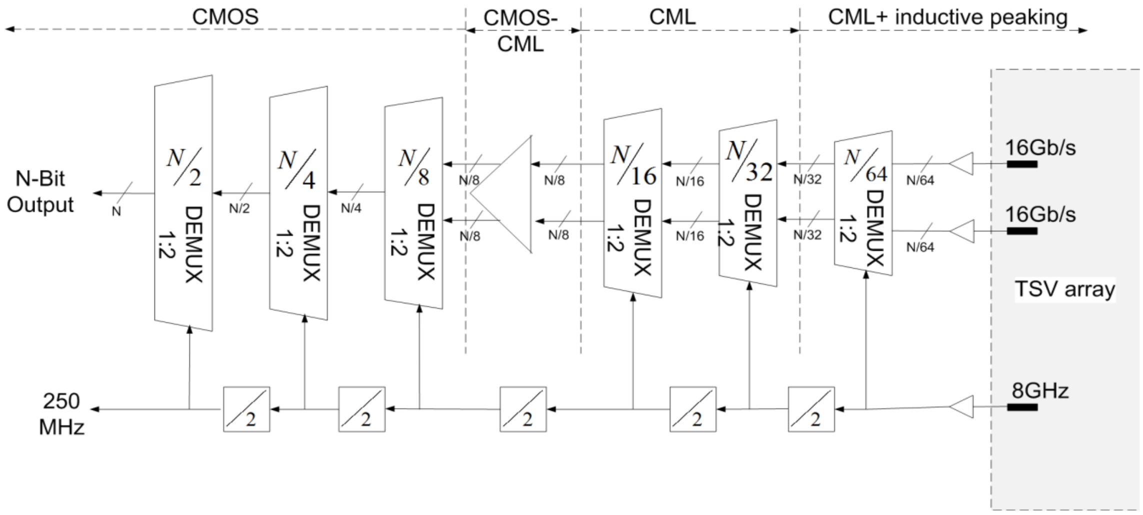
1) The maximum data rates of the other chips, using CMOS flip-flop circuits, are limited by the toggle frequency of CMOS logic [5]. For example, the maximum clock frequency of a 0.18-μm CMOS chip is 3.1 GHz [12], and that of a 0.15-μm CMOS chip is 3.0 GHz [4]. Cell areas in a 90 nm standard CMOS cell library are used for our calculation in this frequency range.
2) Having the tolerance to the common mode noise, current-mode logic (CML) is capable of faster operation than the conventional CMOS logic and has been adopted in 10 Gb/s systems [5] (Fig. 5.17-a). The CML library developed internally in EPFL is used to predict the link area.
3) For serial links with data rate higher than 10 Gb/s, inductors are to be introduced into the design; if even higher than 20 Gb/s, multiple inductors are required (Fig. 5.17-b). Although two inductors in one stage can be implemented interleaved to save area [10],

its large area still could be detrimental to the serial link. So, in our link design in Fig. 5.16, going further to data rate higher than 20 Gb/s is not meaningful. For links with larger link width, quasi-serial link should be a better solution. Please note that we try to minimize the negative effect of the inductors' large area by assuming that the two inductors in each stage of the MUX/DEMUX are implemented interleaved.

4) The TSV arrays' area is calculated by summing individual TSV areas, but in reality, when TSVs are working under higher frequencies, a careful signal integrity design is necessary. This will influence the final area consumed by TSVs. Being case-specific, such issue is not considered in our calculation. Considering that the current TSV yield is not ensured, double redundancy is used for signal TSVs while four-time redundancy is used for clock TSVs.
5) Clock divider designs can achieve very small sizes, such as 9.2 µm × 5.2 µm in [11] and 20 µm × 80 µm in [13]. Because clock division might be shared among other parts of the chip, we decided to minimize its effect on the total link area by choosing the smaller one [11].
6) The total area of the 3D inter-tier link is obtained by summing all the areas for the MUX, the DEMUX, the clock divider and the TSV array. It is assumed that all active devices are placed outside of the KOZ of the TSV array. Thus, MUX's area is part of the total area. This is different from our circuit implementation in Section 5.3 and 5.4.
7) For the circuits designed in CMOS logic, the CMOS logic area is multiplied by a correction factor 1/0.7 in order to account for the exact area needed for routing the signals and the power/clock rails; another 1/0.6 correction factor is applied to the CML area for the same reason. Routing signals from the active area to the TSV array also consumes some area, but this is case-specific and thus will not be taken into account in this theoretical study.



(a)


CMOS
CMOS-CML
CML
CML+ inductive peaking
N-Bit Output
N
N/2
N/4
N/8
N/16
N/32
N/64
N/2 DEMUX 1:2
N/4 DEMUX 1:2
N/8 DEMUX 1:2
N/16 DEMUX 1:2
N/32 DEMUX 1:2
N/64 DEMUX 1:2
16Gb/s
16Gb/s
TSV array
8GHz
250 MHz
2

(b)

Figure 5.16 3D serial link architecture. (a) N:1 MUX. (b) 1:N DEMUX. The maximum serialization ratio is 64:1. Higher than 64:1, the links will be implemented as quasi-serial links. As the clock frequency is set at 250 MHz, the highest data rate to be achieved with 64:1 MUX will be 16 Gb/s.

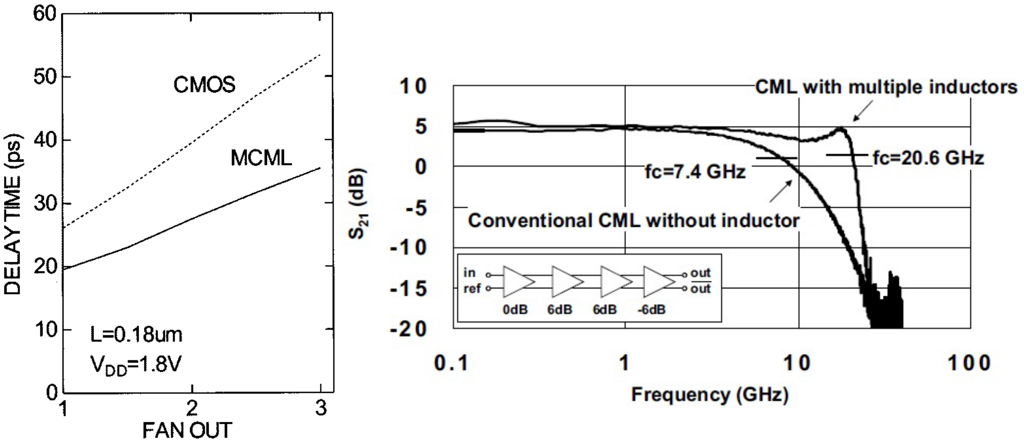
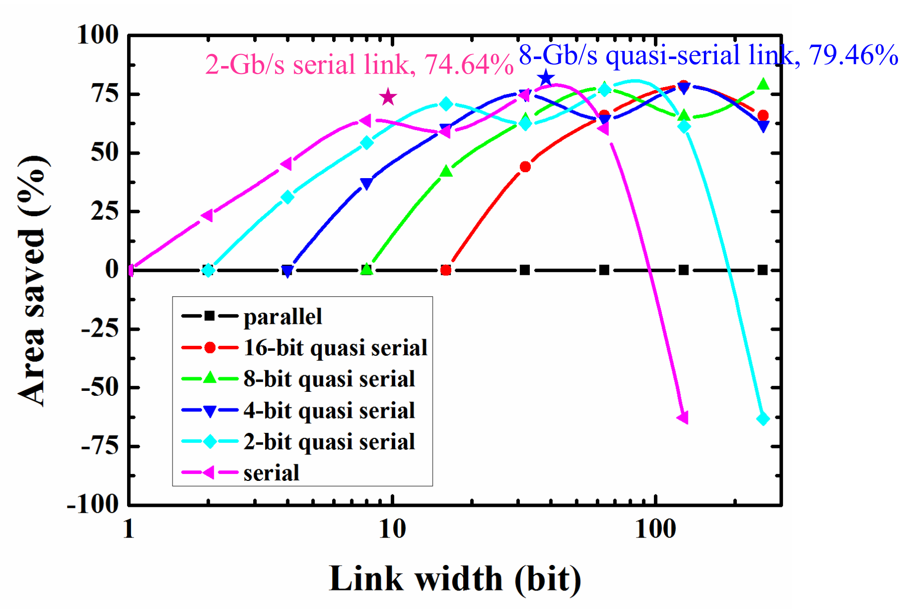


Figure 5.17 Different circuit design techniques used for the serial link design. (a) Comparison of CMOS logic and CML. The delay time as a function of fan-out [5]. (b) Performance enhanced by introducing inductors [15].

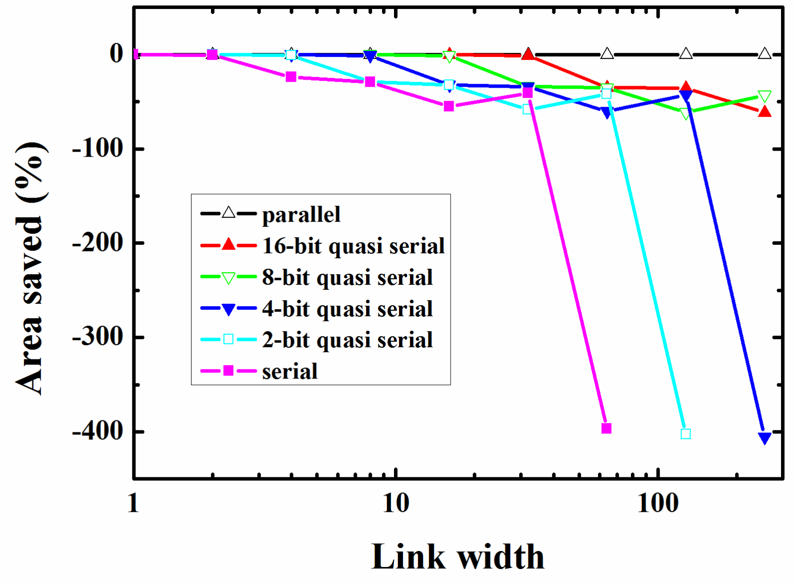
Based on the above principles, the architecture for any N-bit serial or quasi-serial link can be derived. For the calculation, two TSV diameters are considered, 20 μm and 5 μm. The final results are shown in Fig. 5.18. Here, the main observations will be listed as below:

1) As can be seen in Fig. 5.18-a, through serialization, up to 75% area saving can be achieved. Our modeling results are in good accordance with the physical implementation results of the two serial links designed in Section 5.3 and 5.4.
2) When the link width increases more than 64 bits, serial link's area can be twice the area of its parallel counterpart because inductors must now be introduced (Fig. 5.18-a). So, serial link shows clear limitations when the serialization ratio or the data rate is too high. By using quasi-serial strategy, links with wider width can be implemented with much smaller area and operate at lower frequency.
3) As the yield of TSV improves, the redundancy of TSV can be reduced and this might bring less area benefit.

4) Using the conventional circuit design techniques, for TSV diameter as small as 5 µm (Fig. 5.18-b), area gain through serialization cannot be maintained any more. Some novel design technique has been reported which avoids the use of inductors but maintains the high data rate, e.g. [14]. Such designs have great potential to be used in more advanced 3D stacked systems with tiny TSVs and ultra-wide interlayer bandwidths. This will be our main discussion in Chapter 6.



a) TSV dimension = 20 µm



b) TSV dimension = 5 µm

Figure 5.18 Area saved in various inter-tier communication strategies.

5.6 Summary

In this chapter, we exploited the large bandwidth offered by the state of the art TSV technology to investigate its use in inter-layer link design. The proposed inter-layer quasi-serial link achieves five times less area than the traditional synchronous parallel link. This approach can be considered as a low-cost and efficient inter-tier communication solution for

3D systems like 3D image sensors and 3D NoCs. In these systems, the circuit working frequency is relatively low, but at the same time achieving the minimum packaging form factor is also the biggest concern. For memory-on-logic stacking and heterogeneous 3D systems, the TSV dimension required is only a few micrometers. Serialization does not help to save area any more using conventional circuit techniques [4, 5, 6, 9, 10]. Does it mean that serial link has no future in high-performance 3D applications? This will be further discussed in Chapter 6.

Reference

1. R. Dobkin, A. Morgenshtein, A. Kolodny, R. Ginosar, "Parallel vs. Serial On-Chip Communication," *SLIP'08 Proceedings of the 2008 International Workshop on System Level Interconnect Prediction*, 2008.
2. S. Ogg, E. Valli, B. Al-Hashimi, A. Yakovlev, C. D'Alessandro, L. Benini, "Serialized Asynchronous Links for NoC," *DATE'08 Design, Automation and Test in Europe Conference and Exhibition*, 2008.
3. D. Arditti Ilitzky, J. D. Hoffman, A. Chun, and B. Perez Esparza, "Architecture of the Scalable Communications Core's Network on Chip," *IEEE Micro*, Vol. 27, No. 5, pp. 62-74, September 2007.
4. M. Kurisu, M. Kaneko, T. Suzaki, A. Tanabe, M. Togo, A. Furukawa, et al, "2.8-Gb/s 176-mW Byte-Interleaved and 3.0-Gb/s 118-mW Bit-Interleaved 8:1 Multiplexers with a 0.15-µm CMOS Technology," *IEEE Journal of Solid-State Circuits*, Vol.31, No.12, pp. 2024-2029, December 1996.
5. A. Tanabe, M. Umetani, I. Fujiwara, T. Ogura, K. Kataoka, M. Okihara, H. Sakuraba, "0.18-µm CMOS 10Gb/s Multiplexer/Demultiplexer ICs using Current Mode Logic with Tolerance to Threshold Voltage Fluctuation," *IEEE Journal of Solid-State Circuits*, Vol. 36, No. 6, pp. 988-996, 2001.
6. S. Yasuda, Y. Ohtomo, M. Ino, Y. Kado, "3-Gb/s CMOS 1:4 MUX and DEMUC ICs," *IEICE Trans. Electron*., Vol. E78-C, No. 12, pp. 1746-1753, 1995.
7. http://www.faraday-tech.com
8. http://www.arm.com/products/CPUs/ARM968E-S.html
9. K. Kanda, D. Yamazaki, T. Yamamoto, M. Horinaka, J. Ogawa, H. Tamura, H. Onodera, "40 Gb/s 4:1 MUX/1:4 DEMUX in 90 nm Standard CMOS," *ISSCC'05 IEEE International Solid-State Circuits Conference*, pp. 152-153, 2005.
10. D. Kehrer, H.-D. Wohlmuth, H. Knapp, M. Wurzer, A.L. Scholtz, "40 Gb/s 2:1 Multiplexer and 1:2 Demultiplexer in 120 nm CMOS," *ISSCC'03 IEEE International Solid-State Circuits Conference*, Vol. 1, pp. 344-498, 2003.
11. M. Motoyoshi, M. Fujishima, "58.8/39.2 GHz Dual-Modulus CMOS Frequency Divider with 9.2 x 5.2 µm Core Size," *Electronics Letters*, Vol.43, No.2, pp. 98-100, January 18, 2007.
12. K. Nakamura, M. Fukaishi, H. Abiko, A. Matsimoto, and M. Yotsuyanagi, "A 6-Gb/s CMOS Phase-Detecting DEMUX Module Using Half-Frequency Clock," *Symp. VLSI Circuits Dig. Tech. Papers*, 1998, pp. 196-197.
13. C. Cao, and K.O. Kenneth, "A Power Efficient 26-GHz 32:1 Static Frequency Divider in 130-nm Bulk CMOS," *IEEE Microwave and Wireless Components Letters*, Vol. 15, No. 11, November, 2005.
14. A. Mineyamal, T. Suzuki, H. Itol, S. Amakawal, N. Ishiharal, and K. Masul, "A 20 Gb/s 1:4 DEMUX with Near-Rail-to-Rail Logic Swing in 90 nm CMOS process," *IEEE MTT-*

S International Microwave Workshop Series on Signal Integrity and High-Speed Interconnects (IMWS2009-R9), pp. 119-122, 2009.

15. T. Yamamoto, M. Horinaka, D. Yamazaki, H. Nomura, K. Hashimoto, H. Onodera, "A 43Gb/s 2:1 Selector IC in 90 nm CMOS Technology," *IEEE International Solid-State Circuits Conference (ISSCC) Digest of Technical Papers*, Vol.1, pp. 238 - 239, 2004.
16. J. Yang, K. Athikulwongse, Y. Lee, S. Lim, and D.Z. Pan, "TSV Stress Aware Timing Analysis with Applications to 3D-IC Layout Optimization," *47th Design Automation Conference*, 2010.
17. http://www.electroiq.com/index/packaging/packaging-blogs/ap-blog-display/blogs/ap-blog/post987_4325661074814860314.html
18. J. Roullard1, S. Capraro1, A. Farcy, T. Lacrevaz1, C. Bermond1, P. Leduc, J. Charbonnier, C. Ferrandon, C. Fuchs, B. Flechet, "Electrical Characterization and Impact on Signal Integrity of New Basic Interconnection Elements Inside 3D Integrated Circuits," *61st Electronic Components and Technology Conference*, 2011.
19. J. Cho, J. Shim, E. Song, J. S. Pak, J. Lee, H. Lee, K. Park and J. Kim, "Active Circuit to Through Silicon Via (TSV) Noise Coupling," *IEEE 18th Conference on Electrical Performance of Electronic Packaging and Systems*, 2009.
20. E.-X. Liu, E.-P. Li, W.-B. Ewe, H. M. Lee, "Multi-physics Modeling of Through-Silicon Vias with Equivalent-Circuit Approach," *IEEE 19th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*, pp. 33-36, 2010.
21. Z. Xu, A. Beece, D. Zhang, Q. Chen, K.-N. Chen, K. Rose and J.-Q. Lu, "Crosstalk Evaluation, Suppression and Modeling in 3D Through-Strata-Via (TSV) Network," *IEEE 3DIC Conference*, Munich, November 16-18, 2010.
22. T. Kgil, S. D'Souza, A. Saidi, N. Binkert, R. Dreslinski, S. Reinhardt, K. Flautner, T. Mudge, "PicoServer: Using 3D Stacking Technology To Enable A Compact Energy Efficient Chip Multiprocessor," *15th International Conference on Architectural Support for Programming Language and Operating Systems (ASPLOS)*, 2001.
23. S. Hu, Y.-Z. Xiong, J. Shi, L. Wang, B. Zhang, D. Zhao, T. G. Lim, X. Yuan, "THz-Wave Propagation Characteristics of TSV-Based Transmission Lines and Interconnects," *60th Electronic Components and Technology Conference*, pp. 46-50, 2010.
24. S. Kannan, S. S. Evana, A. Gupta, B. Kim and L. Li, "3-D Copper Based TSV for 60-GHz Applications," 61st *Electronic Components and Technology Conference*, pp. 1168-1175, 2011.
25. C. Ryu, J. Lee, H. Lee, K. Lee, T. Oh, J. Kim, "High Frequency Electrical Model of Through Wafer Via for 3-D Stacked Chip Packaging," *Electronics System Integration Technology Conference*, pp. 215-220, 2006.

Chapter 6

System-Level Exploration of Inter-Layer Serial Transmission

In this chapter, we will discuss about serial links' application in more advanced 3D computational systems, that is, the memory-on-logic 3D systems. After clarifying where we stand and some background architectures, we will provide some initial proofs on potential area and power benefits of inter-layer serialization for memory-on-logic 3D systems in Section 6.2. In Section 6.3 and Section 6.4, we will evaluate the system performance after serial inter-layers are inserted to our memory-on-core Chip-Multiprocessor (CMP) simulators. We will summarize our results in Section 6.5.

6.1 Background

6.1.1 Memory-on-Logic 3D Stacking and Our Research Region

The great difficulty of bridging the large and growing gap between CPU speed and memory speed had been well-discussed since 1990's. Computational ability was envisioned to be combined with the off-chip memory system into the same chip [1]. 3D stacking outperforms the 2D solutions by removing the common substrate noise problem when putting analog and digital components on the same chip, and by allowing heterogeneous technologies.

Technically, the first-coming scenario is to directly stack memories on CPUs. Each layer is designed following the corresponding standards and then the I/O mismatch is solved by a redistribution layer. By replacing the power-consuming chip I/Os and long 2D inter-chip connections with short TSVs, 80% of the interface power consumption or 60% of the total memory system power consumption can be saved in a DRAM on logic stacking running mobile applications [2].

A more advanced stacking is to stack custom memories on CPUs. Not only off-chip main memory can be stacked on the logic, but also large on-chip last-level cache are to be split and moved to an upper layer since these days, they are getting larger and almost comparable to the logic areas. Large memories are split into multiple banks, thus multi-port memory access is supported in this scenario, ultra-wide inter-layer bandwidth solves the bus-bottleneck problem between the CPU and the memory. Some components can also be decomposed into 3D to get better performances.

For memory-on-logic stacking, TSVs are to be as small as possible. With 3D chip integration, the eventual TSV diameter could be 5 to 10 μm. It will still take some years to get to 3D silicon integration in which TSV diameter will decrease to less than 1 μm. Most of the previous architecture studies are actually picturing at 3D silicon integration as TSVs are assumed to be as small as 1 μm or 2 μm, and of negligible area. So, more realistic solutions dealing with memory-on-logic 3D systems stacked using 3D chip integration technology (TSV diameter ≥ 5 μm) will be useful.

Our work in this chapter will help clarify the following problems:

- How much area a TSV array with 5 μm-diameter TSVs consumes? We can foresee that it is not negligible.
- Is inter-layer serialization going to help improve the performance of the memory-on-logic 3D systems stacked with 3D chip integration technology (TSV diameter ≥ 5 μm) ?

6.1.2 Networks-on-Chip (NoCs)

A packet-based NoC provides a scalable interconnection fabric for connecting the processor nodes, the on-chip shared cache banks and the on-chip memory controllers [3]. On-chip routers and links constitute this scalable communication backbone. A neat description of the generic NoC router can be found in [10], and is quoted here,

"*A generic NoC router has P input and P output channels/ports; typically P = 5 for a 2D mesh, one from each cardinal direction and one from the local node. The main components of a router are the routing computation unit, virtual channel arbitration unit, switch arbitration unit, and a crossbar to connect the input and output ports. The routing computation unit is responsible for determining the next router based on the packet address and the virtual channel (VC) in the next router for each packet. The virtual channel arbitration unit arbitrates amongst all packets requesting access to the same VCs and decides the winners. The SA unit arbitrates amongst all VCs requesting access to the crossbar and grants permission to the winning packets/flits. The winners are then able to traverse the crossbar and be placed on the output links.*

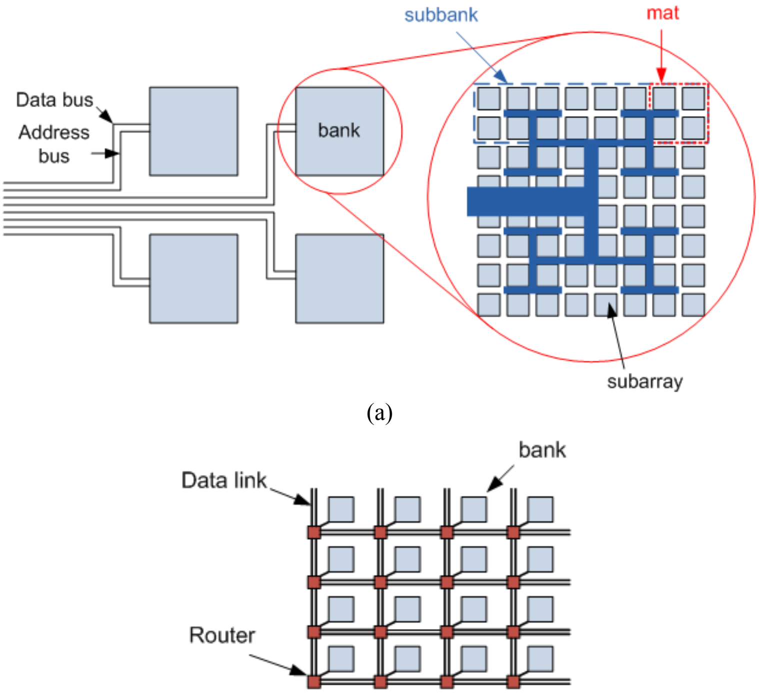
State-of-the-art wormhole switched NoCs devote two to four pipeline stages to these components [4] and typically employ dimension-ordered routing (e.g. X-Y routing) to route packets in the network. The two arbitration stages (virtual channel arbitration unit and switch arbitration unit), where a router must choose one packet/flit among several packets/flits competing for either a common output VC or a crossbar output port, play a major role in selecting packets for transmission. Current router implementations use simple, local arbitration policies such as round robin to decide which packet should be scheduled next."

6.1.3 UCA cache and NUCA Cache

Fig. 6.1-a illustrates the layout of a CACTI-style Uniform Cache Access (UCA) cache (CACTI: An integrated cache and memory access time, cycle time, area, leakage, and dynamic power model which tutorials can be found in [5]). At the highest level, a data array is composed of multiple identical banks. Each bank can be concurrently accessed and has its own address and data bus. Each bank is composed of multiple identical subbanks with one subbank being activated per access. In this example, each bank has four subbanks and each subbank has four mats. Each mat is also partitioned into multiple subarrays (four in this case), where all the subarrays in a single mat share the predecoding logic. Each mat is connected to the cache interface using an H-tree distribution network, also shown in Figure 6.1-a.

For cache read, write, fill, and write-back operations, addresses and data are routed through the H-tree between the cache controller interface and the target subbank. The number and physical organization of banks and sub-banks were chosen to maximize overall IPC, after an exhaustive exploration of the design space. Much performance is lost by requiring worst-case uniform access in a wire-delay dominated cache. Multiple banks can mitigate those losses, if each bank can be accessed at different speeds, proportional to the distance of the bank from the cache controller. In our banked cache models, each bank is independently addressable, and is sized and partitioned into a locally optimal physical sub-bank organization.

Several Non-Uniform Cache Access (NUCA) cache architectures have been proposed (Fig. 6.1-b) [6]. Taking the most basic static NUCA (S-NUCA) cache as an example, they have two advantages over the UCA organization previously described. First, accesses to banks closer to the cache controller incur lower latency. Second, accesses to different banks may proceed in parallel, reducing contention. We call these caches S-NUCA caches, since the mappings of data to banks are static, and the banks have non-uniform access times.



(a)

(b)

Figure 6.1 Illustrations of cache architectures. (a) UCA cache. (b) NUCA cache.

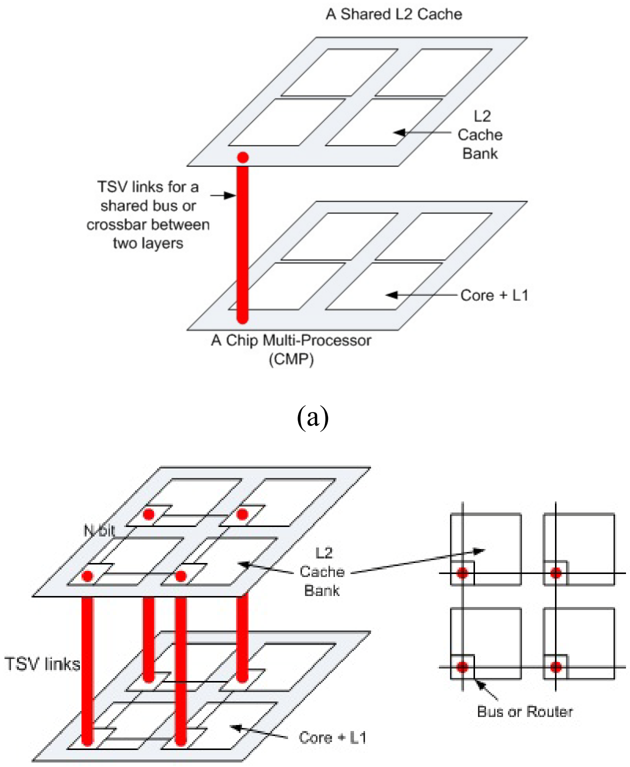
6.1.4 3D Memory-on-Core System Architectures

3D Shared Memory Architecture (SMA)

A 3D shared memory model is proposed for direct memory-on-logic stacking as shown in Fig. 6.2-a. In this model, each core has its own private L1 cache. Each L1 cache is connected to the last level cache (here suppose L2 cache) which locates on the upper layer of chips and is shared among the cores through a crossbar or a bus. An arbiter is needed to decide the connection priority between the core and the cache. The model however does not exploit the full advantage of 3D cache memory in the form of ultra-wide bandwidth. All cores communicate to the cache through a given port and therefore it is bandwidth limited when multiple requests are made simultaneously.

3D Distributed Memory Architecture (DMA)

A 3D distributed memory model is proposed and illustrated in Fig. 6.2-b. Every node, either core or bank, is connected through a network on chip. TSV links are implemented as vertical buses with an interface at each router. Coherence can be maintained between the different set of banks through a directory based system. This option allows the implementation of multiple ports' parallel access to different bank layers, thus exploiting the ultra-wide bandwidth feasible within 3D core-memory stack.



(a)

(b)

Figure 6.2 Inter-layer TSVs are used to implement vertical connections between cores and cache banks. (a) 3D shared memory model. (b) 3D distributed memory model.

6.2 Proofs for Inter-Layer Serialization

6.2.1 Area Considerations

For a vertical parallel link, its link width is actually that of the NoC router (in DMA) or of the bus (in SMA). Considering that routers or buses are working at baseline frequencies around 3 GHz, differential signal transmission is assumed. Using the topology illustrated in Fig. 6.3 where four ground TSVs around one signal TSV, each pair of differential signals (1 bit signal) will require four TSVs. Here, we assume a TSV dimension of 5 μm in diameter with a pitch of 10 μm. The vertical links are one-directional, but we also need to consider the requirement for bi-directional data transmission for the cache. In Fig. 6.4, we compare the areas of the TSVs required for a 1MB L2 cache and the cache itself.

The area of the 1MB L2 cache was estimated using CACTI under 90 nm and 32 nm technologies. And the TSV area overhead is calculated as $S_{TSV\_overhead} = S_{TSVs} / S_{cache}$, where $S_{TSV\_overhead}$ is the TSV area overhead, S_{TSVs} is the total area of the TSV connections to the 1MB L2 cache, and S_{cache} is the area of the 1MB L2 cache.

In 90 nm technology, the TSV array has reasonable area overhead (< 8%) and thus can be tolerated. However, under 32 nm technology, the scaling in device area greatly increases the TSVs area overhead which can be as large as 60%. Thus, from figure 6.4, if 10% area overhead is considered tolerable, under 32 nm technology, the link width of the router or the bus width is constrained to less than 32B.

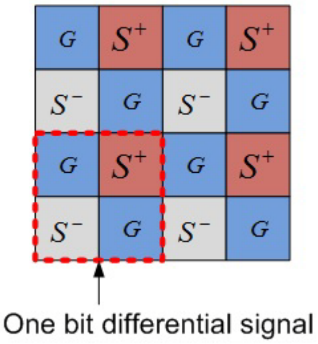


Figure 6.3 Illustration of TSV topology for four bits' signals.

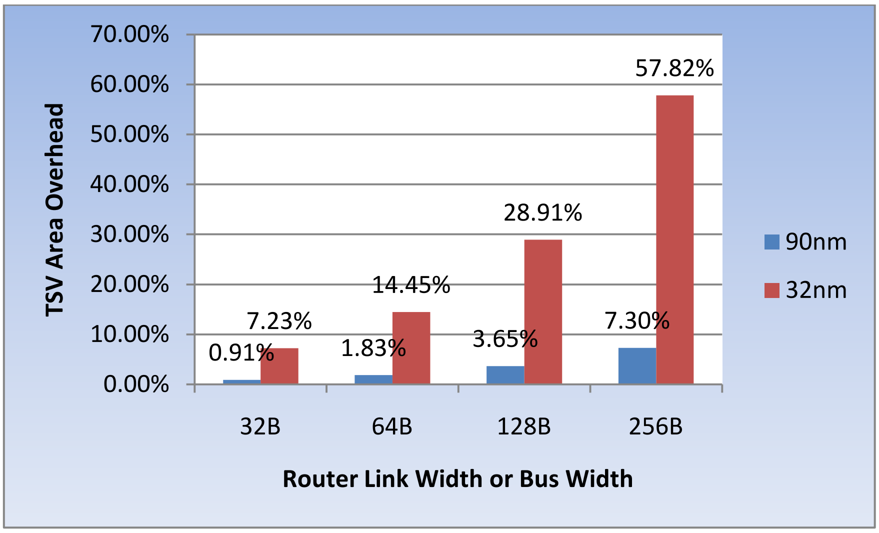


Figure 6.4 TSV array area overhead when TSVs are of 5 µm diameter and with a pitch of 10 µm versus the area of 1MB L2 cache implemented in 90 nm and 32 nm technology.

Now, we consider the areas of the serial links. The designs of 25 Gb/s 1:4 DEMUX and 13 Gb/s 4:1 MUX in 90 nm CMOS presented in [7] can be borrowed for our analysis of inter-layer serialization. The areas for the MUX and DEMUX are respectively 30 × 18 μm^2 and 29 × 40 μm^2 [7]. The area of the 4-bit parallel link is calculated by assuming 16 TSVs (4 TSV × 4), while the area of the corresponding serial link is calculated by adding the areas of the 4:1 MUX, the 1:4 DEMUX and four TSVs. Clock division is not considered in this case.

The result of the calculation is the following: without area-consuming inductors in the design, the 4:1 serial link only takes 36.8% of the area of its 4-bit parallel counterpart. In other words, the original silicon area used to implement a 4-bit parallel link can contain two or three of these 4:1 serial links. With the same duration and silicon area, 2 or 3 times of data bits can be transmitted vertically. Therefore, inter-layer bandwidth is more than doubled through serialization.

6.2.2 Power Consumption Considerations

In Chapter 5, we have already discussed about the great power reduction obtained by replacing the standard chip I/Os with TSVs. Now, we should check whether the power overhead of the serial link will hamper this benefit.

With the MUX and DEMUX designs in [7], the MUX and DEMUX's power consumptions without output buffer, P_{MUX} and P_{DEMUX} are respectively 2.3 mW and 4.8 mW. The self capacitance of a TSV with 5 μm diameter, 10 μm pitch, 100 nm thick silicon dioxide isolation and 15 μm height is only 82.9 fF. Thus, the dynamic power of a TSV P_{TSV} calculated from $\alpha C_{TSV} V^2 f$ (α: activity factor, [0,1]; C_{TSV}: self-capacitance of TSV, 82.9 fF; V: voltage, 0.9 V; f: frequency, 3 GHz) is only 10 μW, and so it can be ignored. Therefore, total power consumption of a vertical 4:1 serial link is dominated by the MUX and DEMUX's elements and is about 7 mW.

The static power of a 1 MB L2 cache (8-associativity, 1 bank, 32B cache line size (also known as cache block size), under 32 nm technology node) P_{cache} is estimated to be 239.3 mW. So, concerning the caches, the total power of the 16 serial links is almost half of the power consumed by the cache itself. On the other hand, the maximum power dissipation of a Ultrasparc IIIi processor working at 1 GHz is 64 W, meaning that it will dominate anyway. We could come to a suggestion that serial link is power-efficient for logic stacking, but maybe is not so suitable for pure 3D memory stacking.

6.3 Serialized Inter-Layer Transmission in 3D SMA

6.3.1 Simulation Platform

Our baseline configuration is a 8-core in-order processor using Ultra SPARC-III ISA. Each core has its own 64KB L1 cache. One shared 8 MB L2 cache is placed on the top layer. In 45 nm technology, the area of the core is 6.8 mm^2. By using CACTI, we further obtain that one cache layer fits to approximately 8 MB SRAM L2 cache, assuming that the cache layer has similar area as the core layer. The configurations are detailed in Table 6.1.

We used the Simics toolset [8] for performance simulation. The L2 cache line size is kept the same as the L1 cache line size for simpler analysis. As Simics does not model bus architectures connecting cores and memories, it is not possible to modify the bus width as we wanted. As an alternative, we modified the cache line sizes, from 64B to 512B, and assumed that L1 and L2 cache line sizes are the same as the bus width. Finally, a Solaris v10 operating system has been installed on the emulated machine.

We used a set of workloads from PARSEC [9]. For each benchmark we fast forwarded the benchmark to the program phase of interest and warm up the caches, then 500 million of cycles were simulated.

Our primary performance evaluation metrics is the average execution time of the eight cores and the L2 cache miss rate. A cache miss refers to a failed attempt to read or write a piece of data in the cache, which results in a main memory access with much longer latency. As the last level cache in this configuration, less L2 cache miss rate indicates less latency brought by main memory access.

Table 6.1 System Configuration of 3D SMA

Processor	8-core, in order , 3 GHz
L1	64KB DL1/IL1 per core, 2-way associate, 2 cycles latency, cache line size studied: 64B, 128B, 256B and 512B
L2	8MB shared cache, 1 MB per bank, 8-way associate, 10 cycles latency, cache line size is the same as L1
Memory	60 cycles latency, 16 MB large page

6.3.2 Simulation Results

Enlarging the cache line size is found to be helpful to reduce L2 cache miss rate (Fig. 6.5). However, it has no improvement on the overall performance. Indeed, Fig. 6.6 demonstrates that although serialization is able to enlarge inter-layer data bandwidth, the system performance is still greatly constrained by the bus-bottleneck problem.

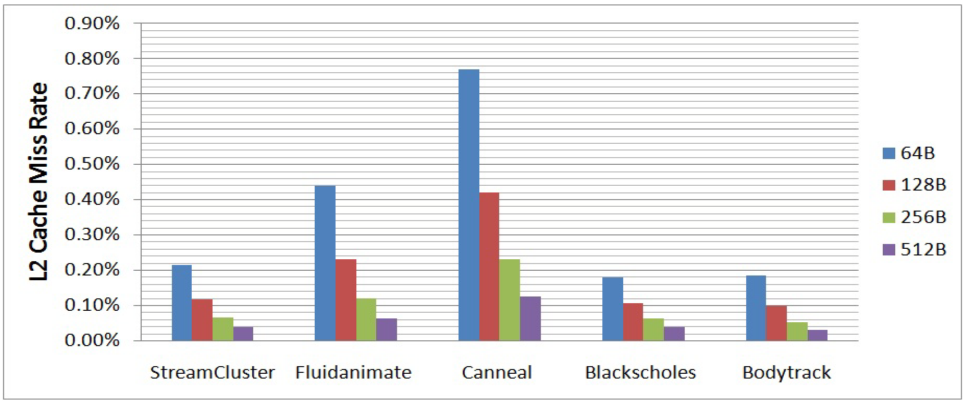


Figure 6.5 L2 cache miss rate with various benchmarks from PARSEC suite.

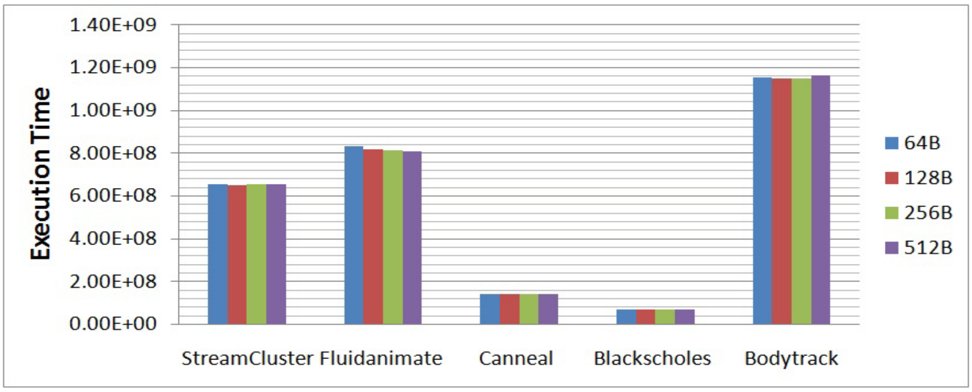


Figure 6.6 Execution time with various benchmarks from PARSEC suite.

6.4 Serialized Inter-Layer Transmission in 3D DMA

6.4.1 Simulation Platform

We implemented a trace-driven cycle-accurate hybrid NoC/cache simulator for CMP architectures. Table 6.2 provides the baseline configuration which contains 64 cores on the top layer and 64 L2 cache banks on the bottom layer (each layer laid out as a 8×8 mesh NoC).

Our CPU model consists of a fetch/issue/commit out-of-order pipeline with a reorder buffer. The memory instructions are modeled through the detailed memory hierarchy and network model mentioned in Table 6.2. The memory hierarchy uses a two-level directory-based MESI (Modified, Exclusive, Shared, Invalid) cache coherence protocol. A 3D network connects the cores, L2 cache banks, and memory controllers.

We assume 3.5 mm^2 area for each core at 32 nm technology based on a scaled version of Sun's Rock core. The SRAM die, placed directly below the core die, is partitioned into 64 banks. Based on estimates from CACTI, a 1MB SRAM cache bank and associated router/controller have an area of 3.03 mm^2, roughly equal to the area of one core as shown in Table 6.2.

Each router uses a state-of-the-art two-stage micro-architecture based on [4]. We used the deterministic X-Y-Z routing algorithm, finite input buffering, wormhole switching, and virtual-channel flow control. A message (or packet) consists of 16 64-bit flits and a header flit. So, each parallel link is of 64 bits. With serialization, if N serial links can be contained within the same area, the link width can be increased to N times of the parallel link width.

The components of the routers were implemented in structural Register-Transfer Level (RTL) Verilog and then synthesized in Synopsys Design Compiler using a 32 nm technology node. The vertical interconnects were modeled as 2D wires with equivalent resistance and capacitance. The resulting designs operate at a supply voltage of 1 V and a clock speed of 3 GHz [10].

We chose 7 bandwidth-sensitive applications in multi-programmed SPEC 2006 benchmarks according to Table 3 in [10]. For each application, we simulated at least 50 million instructions (3200 million instructions across the 64 processors). Four scenarios were studied. In Section 6.2.1, we have demonstrated that using 4:1 serial link achieves only 36.8% area of 4 bit parallel link. So, the corresponding link width can be 2.7 times of the link width of the parallel link. In our case, it will be 172 bits. For our simulation, we studied four link widths, 64-bit, 128-bit, 256-bit and 512-bit concerning different serialization ratios.

With larger link widths, the area of arbitration unit and thus the area of the router will be increased because of the increased circuit complexity of the crossbar. In order to focus on the vertical link areas and keep routers' area out of discussion, we modified the VC number and the buffer depth accordingly. The details can be found in Table 6.3.

Table 6.2 Baseline Processor, Cache, Memory and Network Configuration for 3D DMA.

Processor	3GHz processor
Fetch/Exec/Commit width	2 instruction per cycle in each core; only 1 can be memory operation
L1 caches	32 KB per-core (private), 4-way set associate, 128B block size
L2 caches	1MB banks, shared, 16-way set associative, 128B block size
Main memory	4GB DRAM
Network Router	2-stage wormhole switched, virtual channel flow control, number of VCs per port depends on the serialization ratio
Network Topology	3D network, each layer is an 8x8 mesh, each node in layer 1 has a router, private L1 cache, each node in layer 2 has an L2 bank, 4 memory controllers (1 at each corner node in layer2), number of bi-directional links in each layer depends on serialization ratio.

Table 6.3 Router Scaling for Different Flit Widths to Keep the Same Silicon Area.

	Link width (flit width)	VC	Buffer depth
I	64-bit	8	10
II	128-bit	8	5
III	256-bit	5	4
IV	512-bit	4	3

6.4.2 Simulation Results

The simulation results on network latencies and normalized IPCs are shown in Fig. 6.7 and Fig. 6.8.

1) Great benefit of serialization for the bandwidth sensitive applications is evidenced, since all of the seven benchmarks used are for such applications. For computing intensive application, the performance should also improve, but maybe will not be so obvious. To demonstrate this expectation, running more computing intensive applications is necessary.
2) The memory bottleneck problem has been solved by multi-port memory access scheme enabled by NoC architecture. This is a great advantage of 3D DMA over 3D SMA.
3) As the link width of the router is increased, the equivalent flit width is increased. So with the same message size, the flit number in one message is decreased. Basically, a message occupies one VC and only releases it after the transmission is over. So, the VC channel can be released in a shorter time for sending one message. The first benefit is that the network congestion can be decreased, and so as the network latency. This is demonstrates in Fig. 6.7. The second benefit is that decreasing the VC number and the buffer depth to keep the same router area is not expected to degrade the router's performance. This is in accordance with our actual deeds when we built the simulation platform.

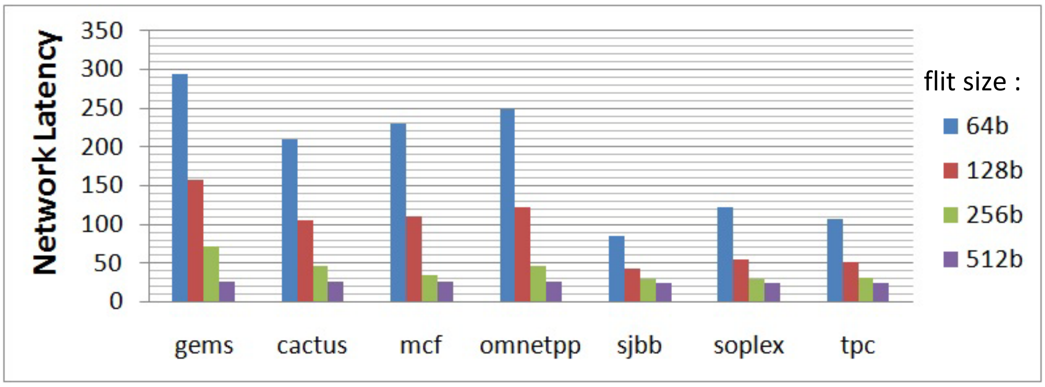


Figure 6.7 Network latencies with seven bandwidth-sensitive benchmarks.

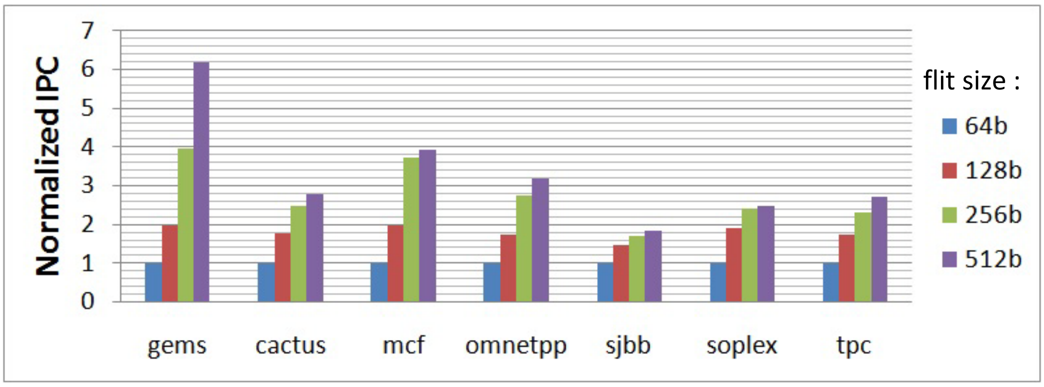


Figure 6.8 Normalized IPC with seven bandwidth-sensitive benchmarks.

4) With a much wider link widths (> 256 bits), the performances, both IPC and network latency, are dampened. The reason is that to transfer a longer flit, longer time is required for the router to process the data. For a router, a longer flit is not always preferred.

5) In our simulation, we set aside the feasibility of circuit implementation temporarily. However, to achieve 8 or 16 times more router link widths, the corresponding link designs have not been demonstrated. So far, the 4:1 serial link design we used for the area benefit proof is, to our knowledge, already with the smallest area and power. It only achieves 2.7 times of the link width. As the baseline clock frequency is high as 3 GHz, higher serialization based on conventional design techniques will require inductors, implying that the area gain will become negligible in that case. Thus the performance improvement will be compromised by the greatly enlarged serialization area overhead. With more advanced technologies, 8:1 serialization could be attained, but higher serialization could also be unrealistic.
6) For a 4:1 serial link, the highest clock frequency required is twice the baseline frequency (in our case, 6 GHz). Additional clock design is necessary, although in our area calculation, we did not consider this part. In the real design, it could introduce a lot of design complexity. For a 2:1 serial link, the baseline clock is sufficient. No additional cost for the clocking generation or division is introduced. The 2:1 MUX and 1:2 DEMUX designs are easier and smaller. So, we can foresee that 2:1 serial link could enable a comparable link width enlargement with a much lower design complexity and could be a good choice as well.

6.5 Summary

Through the work of this chapter, we have studied about the 3D memory-on-core system architectures for the 3D chip integration in which TSVs are not area-negligible. The L2 last-level cache is placed on top of the cores.

First, the potential ultra-wide bandwidth of 3D memory-on-logic stacking is achieved through multi-port memory access scheme. Directly stacking memory on cores, no matter how large the bus width is, will always be constrained by the bus-bottleneck problem and therefore can only be an intermediate solution.

Second, for 3D chip integration in which 5 μm-diameter TSVs are used, the link width of the router or the bus needs to be kept low, e.g. 32B to keep the TSV area overhead reasonable for the corresponding cache or cache bank (<10%). So, the inter-layer bandwidth is constrained by the limited link width of the router.

Through serialization, more serialized vertical links can be stuffed in the same area used for one parallel vertical link, so the inter-layer bandwidth (the link width of the router in NoC) can be expanded, and therefore, the system performance can be improved.

Moreover, considering the serial link implementation, a high serialization ratio (>8:1) is not preferred. A 4:1 serial link has been demonstrated in 90 nm technology, while 8:1 could be achievable in more advanced technologies. 2:1 will be an easy and efficient serialization ratio, it does not need for higher clock frequency than the baseline frequency and still can lead to comparable system performance improvement.

Finally, we still see space here for further exploitation of TSVs' high bandwidth (≥ 60 GHz) in future using novel serial link designs.

All in all, serialization is demonstrated to be an efficient and economic way to expand inter-layer bandwidth and improve the system performance for the memory-on-core 3D systems stacked using 3D chip integration technology and in which the TSV diameters are at least 5 μm.

Reference

1. D. Burger, J. R. Goodman, and A. Kagi, "Memory Bandwidth Limitations of Future Microprocessors," *23rd International Symposium on Computer Architecture*, May, 1996.
2. M. Facchini, T. Carlson1, A. Vignon, M. Palkovic, F. Catthoor, W. Dehaene, L. Benini, P. Marchal, "System-Level Power/Performance Evaluation of 3D Stacked DRAMs for Mobile Applications," *Design, Automation & Test in Europe Conference & Exhibition*, pp. 923 - 928, 2009.
3. W. J. Dally and B. Towles, "Route Packets, Not Wires: On-Chip Interconnection Networks," *Proceedings of Design Automation Conference*, pp. 684 - 689, 2001.
4. L.-S. Peh and W. J. Dally, "A Delay Model and Speculative Architecture for Pipelined Routers," *7th International Symposium on High-Performance Computer Architecture*, pp. 255-266, 2001.
5. CACTI, http://www.hpl.hp.com/research/cacti/.
6. C. Kim, D. Burger, S. W. Keckler, "An Adaptive, NonUniform Cache Structure for WireDelay Dominated On-Chip Caches," *10th International Conference on Architectural Support for Programming Language and Operating Systems (ASPLOS)*, 2001.
7. T. Sekiguchi, S. Amakawa, N. Ishihara, and K. Masu, "Inductorless 8.9 mW 25 Gb/s 1:4 DEMUX and 4 mW 13 GB/s 4:1 MUX in 90 nm CMOS," *Journal of Semiconductor Technology and Science*, Vol. 10, No. 3, September 2010.
8. Simics, http://www.virtutech.com/.
9. PARSEC benchmark suite, http://parsec.cs.princeton.edu/.
10. A. K. Misha, X. Dong, G. Sun, Y. Xie, N. Vijaykrishnan, C. R. Das, "Architecturing On-Chip Interconnects for Stacked 3D STT-RAM Caches in CMPs," *Proceeding of the 38th annual international symposium on Computer architecture*, June, 2011.

Chapter 7

Conclusion and Outlook

In this thesis, I have performed a bottom-up study of through-silicon interconnections achieved by the dominating 3D chip integration technology nowadays. My study tried to cover all levels of the 3D-stacked systems, from technology fabrication to circuit implementation, and finally to system performance evaluation. This concluding chapter summarizes the main results of this work, and presents some perspectives.

7.1 General Conclusion

7.1.1 Technology Aspects

A 3D die-level stacking platform based on post-processing has been successfully developed.

The initial targeted application was logic stacking, to be specific, two identical multiprocessor chips had to be stacked and connected through TSVs. As these chips used for stacking could only be affordable as diced, TSVs had to be post-processed on singulated dies, meaning that the solution had to be different from the other prevailing TSV processes as they were all carried out at the wafer level. Although the cost of the chips seemed to be lowered and could be more compatible with university constrains, actually the technology itself is more challenging. Aside from the many problems inherent to TSV fabrication, a key problem arose, that was how to attach the chips on a carrier wafer with a good precision compatible with further processing steps. Another problem was how we can stack the chips without a die-level bonder. The lack in CMI of some key instruments for 3D fabrication further increased these difficulties.

Through my work on the four TSV processes reported in Chapter 2, the feasibility of TSV fabrication in CMI has been fully explored. TSVs with various dimensions and aspect ratios have been experimented. For each problem, either I have provided or proposed some interesting solutions. I consider that these concepts should enable fast and successful future development. This can be highlighted by two major results:

- The ring-trench TSV process that we developed has successfully demonstrated the feasibility of fabricating TSVs with the prevailing dimensions (20-30 μm in diameter and 50 μm of height). At the time being, this was an advanced technology. The failure of electroplating process was the main unsettled problem to get the final TSV structure.
- Based on bottom-up electroplating in through-wafer large TSV holes, we demonstrated the first dummy chip post-processed with TSVs.

In Chapters 3 and 4, I have conducted a complete study of self-alignment techniques for 3D integration. These techniques included gravitational SA, hydrophobic SA, magnetic SA and electrostatic force SA, among which the first two were found quite promising.

- Using gravitational SA, dies were aligned according to the edges of the cavities on the carrier wafer with about 15 μm inaccuracy. This value was mainly generated from dicing inaccuracy itself and imprecise shape with low profile angle (60 °) of the KOH-etched cavities. This can be reduced to the few micron accuracy required for die-level TSV (20-30 μm in diameter) post-processing if a more accuracy dicing is applied and a larger profile angle (80 °) to better confine the chip is used. Following the process flow proposed in Fig. 3.15, TSVs and bonding bumps are able to be post-processed on dies.

- A silicon interposer with TSVs fabricated on the bottom of the cavities allows multiple thin dies to be self-aligned in the cavities and later be bonded with one thermal compression bonding. This can be a practical solution when flip chip bonder is not available. This interposer also allows another more delicate chip to be stacked without post-processing, which makes it an interesting approach for memory-on-logic stacking.
- The proposed hydrophobic SA technique using surface tension force generated by water menisci around the pads greatly improved structure-to-structure alignment, leading to an accuracy of less than 1 μm. The problem of dicing inaccuracy (±15 - 25 μm) caused by high-volume manufacturing dicing process using dicing saws was gracefully settled. The SA is processed in air, supports chip assembly of various chip dimensions and weights, and has great tolerance on disturbance and instability. After self-alignment, the well-aligned chips can be processed with the subsequent bonding steps to achieve solid metal connection. This technique is fully compatible with current fabrication technologies as no additional structure or "exotic" material or special processing step needs to be introduced. The theoretical meniscus model proposed is also an efficient design tool to predict SA results. It is a low-cost, highly precise, and high throughput multi-chip assembly technique that is well suited for high-end 3D applications with tiny TSV dimensions such as logic stacking and heterogeneous 3D integration.

7.1.2 Through-Silicon Serial Data Transmission Issues

For 3D chip integration, TSV dimensions are limited by the chip assembly process. Currently, the TSV dimension is still as large as 20 - 30 μm; eventually, it will decrease to several microns, e.g. 5 - 10 μm, but not much less. The area consumed by TSVs is therefore far from negligible.

In Chapter 5, the feasibility of transmitting serialized data through TSVs in order to save the area consumed by the currently-large TSVs has been demonstrated. The physically implemented 2-Gb/s 8:1 serial link only consumes 20% of the area occupied by its parallel counterpart. A "quasi-serial link" concept proposed is able to balance the inter-layer bandwidth and the serial link's area consumption. According to the serial link design methodologies, great area benefit can be gained using through-silicon serial data transmission for the current TSV dimensions, but to keep this benefit along with TSV shrinkage, novel circuit design techniques are required.

In Chapter 6, through-silicon serial data transmission technique is introduced for the multi-core memory-on-logic 3D systems in which TSVs are only 5 μm in diameter. We demonstrated that as serial links can be implemented using less area, the bandwidth per unit area is increased. Two scenarios were studied, single-port memory access and multi-port memory access. The expanded inter-layer bandwidth by serialization could not improve the system performance in the former scenario because the inter-layer bandwidth was still constrained by the bus bottleneck. However, in the latter scenario, the ultra-wide inter-layer bandwidth has been exploited as each memory bank is accessible randomly by the cores through the NoC. Further expanding this bandwidth through serialization, the system performance showed great improvement.

Through-silicon serial data transmission not only is an efficient technique to save area, but also can increase inter-layer data bandwidth and improve the system performance.

7.2 Outlook

Although quite some work has been done in this emerging technology, there is still a lot to do. A few ideas are briefly discussed below:

First, the final implementation of the 3D stacked chip using the developed techniques is to be done. Integrating the thinning process with the TSV technology should be one of the first tasks. A good copper electroplating process for via filling is another one. Although die-level TSV processing will not be an industrial solution, it still has some significance in the current stage for research as 3D technology has not been standardized and the infrastructure has not been well-established yet.

Second, the hydrophobic SA technique should be further developed. As having been discussed in the end of Chapter 4, cheaper materials such as copper can be experimented to replace Au. Self-alignment directly according to the exposed copper TSVs (can be viewed as pads) is an interesting and meaningful work. As bumpless (pad-to-pad) bonding can be performed in this case, higher TSV density will be achieved and TSV dimensions could be decreased further. Experiments to understand the SA failures after the introduction of Au studs are still necessary, as well as a study of low-temperature bonding techniques. An automatic pick and place process will be meaningful to get better alignment accuracy and higher yield. Design rules can be further developed based on a more precise and controlled experimental procedure.

Third, for memory-on-logic 3D systems, only low serialization ratio of 2:1 or 4:1 are supported with real circuit implementations now because the parallel inter-layer link already works at 3 GHz. To fully exploit the high bandwidth of TSVs, developing novel serial link design techniques is the main critical issue.

A seemingly widely-agreed saying is also adopted in this thesis, i.e. the eventual TSV dimensions in 3D chip integration may not be able to shrink to less than 1 μm as they will be in 3D silicon integration, thus 3D silicon integration will be the ultimate solution for the advanced 3D systems such as heterogeneous system and pure logic stacking. Through our work, we do see more potential of 3D chip integration competing with 3D IC integration in these tiny TSV domains.

First, the TSV dimension could shrink to less than 5 μm, as the alignment accuracy now can be decreased to less than 1 μm using our hydrophobic SA technique and the novel bumpless or copper pillar bonding technology supports higher TSV density.

Second, even with relatively larger TSVs, the circuit and system performance can still be improved, for instance, by adopting the serialization technique that we have proposed. We believe there will be more options.

List of Figures

Figure 1.1 3D Roadmap. Courtesy of ASE Group.. 2
Figure 1.2 Toshiba commercialized CSCM with TCV (= TSV) technology. 2
Figure 1.3 Main TSV integration scenarios. Courtesy of Yole Development........ 3
Figure 1.4 Poly-Si TSV process proposed by Kawano *et al*.. (a) Design rules for poly-Si TSVs and plan views before poly-Si CVD. (b) Poly-Si TSV cross sections at several process steps. 4
Figure 1.5 Schematic cross section of the TSV structure proposed by Bauer. 4
Figure 1.6 KOH etching mechanism. 5
Figure 1.7 TSV technologies presented by (a) Linder *et al*. and (b) Heschel *et al*. 6
Figure 1.8 Coaxial shield TSV structure. 6
Figure 1.9 Toshiba TSV technology. (a) Via process flow used in Toshiba 3D image sensor. (b) First via drilling. (c) Cross-section of Cu-plated via. 7
Figure 1.10 TSV technology proposed by Nagarajan *et al*. and Tezcan *et al*.. (a) Via hole after 1st DRIE. (b) Via hole after 2nd etch and filled with copper. (c) Via cross section view. 8
Figure 1.11 TSV technology proposed by Wieland *et al*.. (a) Test wafer with W-filled Inter Chip Vias (ICVs) and interconnect metallization. (b) FIB image of a metal gate NMOS transistor (w/l = 20/2) with W-filled ICV near active area........ 9
Figure 1.12 Major process steps of the MEOL TSV technology proposed by Redolfi *et al*.. 10
Figure 1.13 Results of the major process steps in the MEOL TSV technology proposed by Redolfi *et al*. 10
Figure 1.14 TSV technology proposed by Samsung. 11
Figure 1.15 TSV technology proposed by Lincoln Lab. 11
Figure 1.16 Surface after lapping and defect depth of a Si sample lapped with 3 μm Al_2O_3 abrasive particles. Pits as deep as 7 μm have been found in SEM micrographs 12
Figure 1.17 Main problems caused by thinning. (a) Thickness nonuniformity and warpage. (b) Thin wafer delamination and breakage. (c) Chipping at the wafer edge. 13
Figure 1.18 50 μm-thick 200 mm silicon wafer showing flexibility 13
Figure 1.19 Xilinx introduced 3D silicon interposers with TSVs for wide I/O interface in FPGA products. Courtesy of Xilinx. 15
Figure 2.1 Core-hole TSV process flow. 22
Figure 2.2 Bosch process etched microtrenches in CMI. 23
Figure 2.3 Schematic of a tapered via hole, with via parameter definition (local bowing, undercut, slope angle and depth) 24
Figure 2.4 SEM cross section images of tapered via holes etched with modified Bosch process when C4F8 flow rate increases from 40 sccm (a) to 100 sccm (b). 25
Figure 2.5 SEM cross section images of tapered via holes etched with modified Bosch process when the temperature decreases from 20 °C (a) to 10 °C (b). 25
Figure 2.6 SEM cross section images of the via holes with different nominal diameters of (a) 80 μm, (b) 30 μm, (c) 10 μm, etched with modified Bosch process. 26
Figure 2.7 SEM cross section images of the via holes etched with continuous Si etching process using different amount of additional C_4F_8 gas flow: (a) 280 sccm, (b) 350 sccm, (c) 400 sccm. 27

Figure 2.8 Etch rate, slope angle, and undercut using different amount of C_4F_8 for continuous Si etching. SF_6 flow rate uses baseline number, 700 sccm. Nominal TSV diameter is 30 μm. 28
Figure 2.9 SEM cross-section views of the via holes after the continuous silicon etching. The nominal TSV diameters and the platen temperatures are marked on the images. 28
Figure 2.10 Etching results of continuous etching under different platen powers. SF_6: 400 sccm, C_4F_8: 400 sccm, substrate temperature: 0 °C. (a) Platen power: 70 W, (b) Platen power: 100 W. 29
Figure 2.11 SEM cross-section images of the via holes etched using the optimized continuous silicon etching recipe. TSV nominal diameters: (a) 100 μm; (b) 80 μm; (c) 60 μm; (d) 30 μm; (e) 20 μm; (f) 10 μm. 30
Figure 2.12 Etch rates for the via holes with different nominal diameters in Bosch process, modified Bosch process and continuous etching process. 30
Figure 2.13 Undercuts for via holes with different nominal diameters in continuous etching and modified Bosch process. 31
Figure 2.14 Local bowings for via holes with different nominal diameters in continuous etching and modified Bosch process. 31
Figure 2.15 Schematic of the spraying system 32
Figure 2.16 SEM cross section images of TSVs after spray coating and a soft bake 33
Figure 2.17 AZ9260 trapped on top of the via holes after spin coating. 34
Figure 2.18 SEM image of the bottom of a via hole after 7 minutes BHF etching. Thermal dioxide thickness: 1 μm. 35
Figure 2.19 SEM image of the damaged vertical TSVs after 7 minute BHF SiO_2 etching. (a) Top corner of the vertical vias after resist stripped. (b) Sidewall and bottom of the vertical vias with resist left on the sidewall 36
Figure 2.20 Ring-trench TSV process flow. 36
Figure 2.21 Copper pumping during thermal cycling. Copper pops out and breaks the dielectric layer around the TSV. Courtesy Tezzaron. 37
Figure 2.22 50 μm deep ring trenches etched in a silicon wafer. 37
Figure 2.23 Ring trenches filled with RH8023 with PGMEA prewetting. 38
Figure 2.24 SEM cross section images of the ring trenches after RH8023 spin coating and prebaking. (a) AR = 10:1. (b) AR = 20:1. 40
Figure 2.25 SEM cross section images of the ring trenches after spin coating and prebaking. (a) AZ9260 after one spin coating. (b) CCV-S300 fills 3/5 of the trenches after 3 times of spin coating. The trench depth is 50 μm. 40
Figure 2.26 Comparison of CCV-S300-filled ring trenches after prebaking (a) and after thermal curing (b). 41
Figure 2.27 Comparison of AZ9260-filled ring trenches after prebaking (a) and after thermal curing (b). 41
Figure 2.28 Experimental procedure for thermal curing of RH8023. 41
Figure 2.29 A typical DSC curve 42
Figure 2.30 RH8023 electroplating mold on 100 μm-wide TSV holes with Ta/Au seed layer. (a) Without PGMEA prewetting. (b) With PGMEA prewetting. 43
Figure 2.31 Electroplating molds on vias with various diameters of (a) 10 μm; (b) 20 μm; (c) 30 μm; (d) 40 μm; (e) 80 μm. 44
Figure 2.32 Residues and damages on the via sidewall after resist development. (a) Residues on the sidewall of 100 μm-diameter via; (b) Damages on the sidewall of 100 μm-diameter via; (c) Damages on the sidewall of 20 μm-diameter via. 44

Figure 2.33 Lithography cross a 10 μm-diameter via hole with a BARC coated below RH8023. 44
Figure 2.34 RH8023 Electroplating mold on 100 μm-wide via holes with Ti/Pt seed layer. (a) Via top after development. (b) Much resist residue on the via sidewall after development. .. 45
Figure 2.35 SEM cross section image of a ring-trench TSV after RH8023 filling and curing. 45
Figure 2.36 SEM cross section image of damage on Si substrate after Si core removal using isotropic silicon etching because of leakage defect in the polymer..... 46
Figure 2.37 SEM cross-section images of TSVs after silicon core removal ((a) after isotropic Si etch or (b) after anisotropic Si etch), 30 second O_2 plasma and Ta/Cu seed layer sputtering. 46
Figure 2.38 SEM cross section images of ring-trench TSVs after central silicon core removal. (a) After isotropic silicon etching. (b) After isotropic silicon etching and O_2 plasma..... 47
Figure 2.39 Microscopic picture of a failed TSV because of leakage on the top corner during plasma etching..... 47
Figure 2.40 SEM images of a ring-trench TSV after 60 nm Ta and 200 nm Cu sputtering. (a) Bottom part of TSV. (b) Middle part of the TSV. (c) Upper part of TSV. TSV depth 35 μm, diameter 20 μm, AR 1.75 (the maximum AR for copper sputtering > 3 (Figure 2.42))...... 48
Figure 2.41 SEM cross section image of ring-trench TSVs after uniform copper electroplating..... 49
Figure 2.42 SEM cross section image of via holes after uniform copper electroplating. Seed layer: 60 nm/500 nm Ta/Cu. AR: ~3..... 49
Figure 2.43 SEM images of a ring-trench TSV after ring trench filled with parylene and center silicon core removed..... 51
Figure 2.44 SEM images showing Al sputtering aspect ratio limitation of about 2.1..... 52
Figure 2.45 Angular distribution of (a) sputtered Cu and (b) sputtered Al 53
Figure 2.46 Step TSV process..... 53
Figure 2.47 350 μm -deep KOH etched cavities after 40 nm/100 nm/2 μm Ti/Pt/Al sputtering. 54
Figure 2.48 SEM cross-section views of a through-wafer TSV hole after 10 μm (8 μm achieved in reality) parylene deposition and double-side 200 nm/2 μm Ti/Al sputtering. (a) The whole via hole. (b) The middle part of the via hole. (c) The bottom of the via hole. 55
Figure 2.49 SEM image showing cracks in Al film sputtered on parylene..... 56
Figure 2.50 Thinning equipments. (a) Rotating plate with a moving arm. 56
Figure 2.51 Assembly profile and holder..... 56
Figure 2.52 100 μm thick thinned Si component with TSV holes..... 57
Figure 2.53 100 μm deep TSV holes dry etched in silicon and opened on the other side by thinning. TSV diameter: (a) 100 μm; (b) 50 μm..... 57
Figure 2.54 SEM image of exposed Al structures after passivation layer etching..... 58
Figure 2.55 Profile images of some metal line structures in different layers measured by VEECO optical profiler..... 58
Figure 2.56 Illustration of Al pad metal detachment during SiO_2 etching..... 59
Figure 2.57 Stencil as etching mask to protect areas unwanted to be exposed to plasma etch. 59
Figure 2.58 Whole-through TSV process..... 60
Figure 2.59 SEM images of through wafer Si etching. (a) Top view. b) Cross-section A-A').60
Figure 2.60 SEM image of whole-through TSVs after bottom-up electroplating..... 60
Figure 3.1 Karl Suss MA6 aligner..... 67

Figure 3.2 Gravitational SA precision limitation. 68
Figure 3.3 Electrostatic SA proposed by Dalin *et al.*. (a) Concept. (b) Detail of a finite element model for computations of the electrostatic force. (c) An example of a fabricated test structure on substrate for self-assembly experiments with electrostatic attraction. 68
Figure 3.4 Magnetic-assisted SA technique proposed by Shet *et al.*..................................... 69
Figure 3.5 Magnetic SA technique proposed by Nichol *et al.* 69
Figure 3.6 Post-processing TSVs on dies using gravitational SA technique. 71
Figure 3.7 The final thin die with post-processed TSVs and bumps...................................... 71
Figure 3.8 Suitable Si interposer application, taking memory-on-logic................................. 72
Figure 3.9 The silicon interposer to be fabricated.. 73
Figure 3.10 Simplied process sequence for the first trail of the silicon inteposer with TSVs. 73
Figure 3.11 Ni TSV after CMP on the front side. 74
Figure 3.12 Ni TSVs exposed after backside KOH silicon etching. 74
Figure 3.13 Microscopic view of assembly. (a) Successfully aligned. (b) Not aligned. 74
Figure 3.14 Wafer bowing after SiO_2 sputtered on Ni... 75
Figure 3.15 Improved process flow to fabricate the 3D silicon interposer proposed in Fig. 3.12. 76
Figure 3.16 Proposed magnetic SA technique. (a) A simplified illustration. (b) Magnets distribution topography on the carrier wafer. 76
Figure 3.17 COMSOL simulation of magnetic forces on the pad with the permanent magnet 1mm away. 77
Figure 4.1 Periphery-to-periphery SA technique proposed by Fukushima *et al.* 80
Figure 4.2 Hydrophobic SA techniques. (1) Wafer-level frame-to-frame hydrophobic SA technique proposed by Martin *et al.*. (2) A SA procedure using surface tension force generated by lubricants, proposed by Böhringer *et al.*. (3) Pad-to-pad SA technique proposed by Chang *et al.* 81
Figure 4.3 Illustration of the proposed pad-assisted SA technique with multiple dies aligned in parallel. The black surfaces are hydrophobic. Therefore, water is contained between the die and the Au frame on the carrier only. The pads on the two parts are congruent, further supporting the alignment procedure, to achieve high pad-to-pad accuracy. After SA, thermo-compression bonding can be performed. 82
Figure 4.4 Experimental pad-to-pad alignment inaccuracy in the *x*- and the *y*-direction of ten samples for fully-wetted, partially-wetted, and pad-assisted SA procedures. 85
Figure 4.5 Optical microscope photographs of the pad-assisted SA process. (a) The water-air interfaces around the pads are visible during water evaporation. (b) Pad-to-pad alignment after water evaporation. 86
Figure 4.6 SA dynamics: Alignment error amplitude measured by optical-microscope video-imaging during the pad-assisted SA process until complete relaxation is achieved. (Pad dimension: 100 μm) 87
Figure 4.7 Experimental pad-to-pad alignment inaccuracy in the *x*- and *y*-direction of 87
Figure 4.8 Center patterns introduced to the center of the chip and the binding site. 88
Figure 4.9 SEM image of (a) a 15-μm-high Au-stud prior to thermo-compression bonding. (b) Au-Au interface after shear test, indicating a ductile cohesive fracture. 89
Figure 4.10 Modeled alignment force at different die-to-carrier heights H for fully-wetted SA. 90
Figure 4.11 Modeled alignment forces of pad-assisted SA at a die-to-carrier height *H* of 35 μm when assuming one and four columns of 25 pads on each side of the chip. 91

Figure 4.12 Results of the sensitivity analysis of the pad-assisted SA procedure with varying dicing accuracy, pad number, and tilting angle. 92
Figure 4.13 Alignment force benchmarking of individual SA procedures at a gap of 35 μm. 93
Figure 5.1The comings of TSV area consumption. 98
Figure 5.2 KOZ for TSVs of different sizes. Courtesy of University of Texas. 98
Figure 5.3 Cross-section illustration of 3D interconnections including TSVs and bumps...... 99
Figure 5.4 Lumped model of Ground-Signal (G-S) 3D interconnection pair. 99
Figure 5.5 Return loss and insertion loss of a Ground-Signal TSV pair with ideal impedance matching under upto 1 THz. 102
Figure 5.6 TSV timing model. 102
Figure 5.7 Timing performance of two identical 3D connections separately working under 1 GHz and 2 GHz. 103
Figure 5.8 TSV arrangement providing the least coupling. 103
Figure 5.9 Scenarios for dynamic power consumptions of (I) Off-Chip I/O connection; (II) On-chip TSV connection. 104
Figure 5.10 Architecture of the 2-Gb/s inter-layer serial link. 106
Figure 5.11 Schematic of the 8:1 MUX. 106
Figure 5.12 Schematic of the 1:8 DEMUX. 106
Figure 5.13 Design rules concerning TSV insertion. 107
Figure 5.14 Layouts for 2-Gb/s serial link. Layout of a) the 8:1 MUX; b) the 1:8 DEMUX.108
Figure 5.15 Architecture of 8Gb/s quasi-serial link. 109
Figure 5.16 3D serial link architecture. (a) N:1 MUX. (b) 1:N DEMUX. The maximum serialization ratio is 64:1. Higher than 64:1, the links will be implemented as quasi-serial links. As the clock frequency is set at 250 MHz, the highest data rate to be achieved with 64:1 MUX will be 16 Gb/s. 112
Figure 5.17 Different circuit design techniques used for the serial link design. (a) Comparison of CMOS logic and CML. The delay time as a function of fanout. (b) Performance enhanced by introducing inductors 112
Figure 5.18 Area saved in various inter-tier communication strategies. 113
Figure 6.1 Illustrations of cache architectures. (a) UCA cache. (b) NUCA cache. 119
Figure 6.2 Inter-layer TSVs are used to implement vertical connections between cores and cache banks. (a) 3D shared memory model. (b) 3D distributed memory model. 120
Figure 6.3 Illustration of TSV topology for four bits' signals. 121
Figure 6.4 TSV array area overhead when TSVs are of 5 μm diameter and with a pitch of 10μm versus the area of 1MB L2 cache implemented in 90 nm and 32 nm technology. 121
Figure 6.5 L2 cache miss rate with various benchmarks from PARSEC suite. 123
Figure 6.6 Execution time with various benchmarks from PARSEC suite. 123
Figure 6.7 Network latencies with seven bandwidth-sensitive benchmarks. 125
Figure 6.8 Normalized IPC with seven bandwidth-sensitive benchmarks. 125

List of Tables

Table 1.1 Thinned Silicon Wafer Thickness 200 mm / 300 mm. 12
Table 2.1 Settings for Each Parameter Used for the Modified Bosch Process 25
Table 2.2 Settings for Parameters Used for Continuous Etching Process. 27
Table 2.3 Materials Used for Polymer Filling. 37
Table 2.4 Polymer Coating Procedure 39
Table 2.5 Lithography Procedure for 7 µm-Thick RH8023. 42
Table 2.6 Benchmarking the TSV Processes Developed. 61
Table 4.1 SA Procedures with Pattern Design and Water Wetting Topology 84
Table 5.1 Studied TSV Geometries. 100
Table 5.2 Delay with or w/o TSV. 103
Table 5.3 TSV Capacitances and I/O Capacitances. 104
Table 5.4 Area Summary for the 2-Gb/s Serial link and its Parallel Counterpart. 108
Table 5.5 Area Summary for the 8-Gb/s Serial link and its Parallel Counterpart. 110
Table 6.1 System Configuration of 3D SMA 122
Table 6.2 Baseline Processor, Cache, Memory and Network Configuration for 3D DMA. 124
Table 6.3 Router Scaling for Different Flit Widths to Keep the Same Silicon Area 124

Fengda SUN

Address:	Ch. de Veilloud 16, CH-1024, Ecublens, Switzerland
Phone:	+41 78 762 8826
Email:	sunfengda@gmail.com
Nationality:	China
Date of Birth:	March 23, 1983
Marital Status:	Single
Language:	Mandarin, English

Education

09/2007–09/2011 *Ph.D., Microelectronics*
École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

09/2005–09/2007 *M.S., Electrical Engineering*
Florida International University (FIU), USA

09/2001–06/2005 *B.S., Computer Science*
The University of Science and Technology, Beijing (USTB), China

Research Experience

05/2011–09/2011 *Research Assistant*
EPFL, Switzerland

- PhD thesis: “Self-Aligned 3D Chip Integration Technology and Through-Silicon Serial Data Transmission”.

10/2010–02/2011 *Visiting Scholar*
Microsystems Design Laboratory (MDL), Pennsylvania State University (PSU), USA

- Studied system-level application of inter-tier serial links in 3D memory-on-logic systems.

10/2008–12/2008 *Intern*
IMEC, Belgium

- Conducted experiments on polymer filling of high aspect ratio micro-structures in silicon substrates. Studied polymer curing and characterization methods such as differential scanning calorimetry (DSC).

09/2005–06/2007 *Research Assistant*
Beihang University (BUAA)–FIU Dual Master Program on Integrated Circuit Design, Beijing, China

- Implemented I^2C, 32-point FFT, 32-order FIR filter, Viterbi Encoder/Decoder and DMAC based on AMBA rules using Verilog HDL.
- Gained experience on designing op-amp circuits and bandgap reference circuits.
- Studied blind signal processing algorithms.

02/2005–06/2005 *Research Assistant*
The University of Science and Technology, Beijing, China

  - Studied Bayesian theory of forecasting and decision-making for macro-economics application and built the web interfaces and the database using Java, JSP, and SQL.

Publications

- Fengda Sun, Alessandro Cervero, Panagiotis Athanasopoulos, Yusuf Leblebici, “Design and Feasibility of Multi-Gb/s Quasi-Serial Vertical Interconnects Based on TSVs for 3D ICs,” *19th IFIP/IEEE International Conference on Very Large Scale Integration (VLSI-SoC’10)*, Madrid, September 2010 (poster).
- Fengda Sun, Thomas Brunschwiler, Yusuf Leblebici, “Surface-Tension-Driven Multi-Chip Self-Alignment Techniques for Heterogeneous 3D Integration,” *61th Electronic Components and Technology Conference (ECTC’61)*, Orlando, USA, May 2011 (oral presentation).
- Fengda Sun, Asit K. Mishra, Vijaykrishnan Narayanan, “Using Cost-Effective Serial TSVs for Optimizing Bandwidth Sensitive Applications in 3D-Stacked CMPs,” Conference paper to be submitted.

Honors

- Recipient of the 1^{st} Entrance Scholarship at Beihang University in 2005
- Recipient of the 2^{nd} National Scholarship at the University of Science and Technology, Beijing in 2004
- Recipient of SONY Scholarship from SONY, China at the University of Science and Technology, Beijing in 2003

September 14, 2011