TUP-Springer **Project**

Yangdong Deng
Wojciech P. Maly

3维超大规模集成电路

——2.5维集成方案

3-Dimensional VLSI

A 2.5-Dimensional Integration Scheme

清华大学出版社

Springer

Yangdong Deng
Wojciech P. Maly

3维超大规模集成电路

—— 2.5维集成方案

3-Dimensional VLSI

A 2.5-Dimensional Integration Scheme

清华大学出版社　　Springer

内 容 简 介

本书提出一种新的 3 维超大规模电路集成方案，即 2.5 维集成。根据这一集成方案实现的电子系统将由多层单片集成芯片叠加而成，芯片间将由极细小尺度的“垂直联线”实现电气连接。这一新集成方案能够在很大程度上克服积累成品率损失的问题。

本书从制造成本和设计系统性能两方面探讨 2.5 维集成的可行性。首先，作者建立了一个成本分析模型来比较各种典型集成方案，分析数据表明 2.5 维集成具备制造成本上的优越性。从设计性能角度，作者完成了全定制和专用集成电路两种设计风格的一系列设计实例研究，从而证明了 2.5 维集成能够实现传统单片集成不能达到的系统性能。同时，为了实现 2.5/3 维集成电路版图，作者也开发了第一代 2.5 维/3 维物理设计 EDA 工具。

本书适合集成电路工艺开发人员和决策人士、集成电路设计人员、电子设计自动化研发人员和决策人士参考。

版权所有，侵权必究。侵权举报电话：010-62782989　13701121933

图书在版编目(CIP)数据

3 维超大规模集成电路：2.5 维集成方案＝3-Dimensional VLSI: A 2.5-Dimensional Integration Scheme：英文/邓仰东，(美)马利 (Maly, W. P.)著. —北京：清华大学出版社，2010.1
ISBN 978-7-302-21165-5

Ⅰ.3…　Ⅱ.①邓…　②马…　Ⅲ.超大规模集成电路－英文　Ⅳ.TN47

中国版本图书馆 CIP 数据核字(2009)第 180065 号

责任编辑：陈志辉
责任校对：王淑云
责任印制：孟凡玉
出版发行：清华大学出版社　　**地　　址**：北京清华大学学研大厦 A 座
http://www.tup.com.cn　　**邮　　编**：100084
社 总 机：010-62770175　　**邮　　购**：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质 量 反 馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：北京雅昌彩色印刷有限公司
经　　销：全国新华书店
开　　本：153×235　**印　张**：13.25　**字　数**：175 千字
版　　次：2010 年 1 月第 1 版　　**印　　次**：2010 年 1 月第1 次印刷
印　　数：1～1200
定　　价：68.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题，请与清华大学出版社出版部联系调换。联系电话：(010)62770177 转 3103　　产品编号：029613-01

Preface

Today we are seeing strong demand for integrating more functionality onto silicon. Nonetheless, we are soon approaching the limit of Moore's law. In fact, the fundamental physics laws preclude the scaling of CMOS devices below a certain dimension. On the other hand, so far no alternative technologies are likely to mature and replace CMOS in the coming 15 years. Then how could the semiconductor industry continue to provide integration capacity for constantly increasing functionality?

3-D integration is a natural solution to address the above problems. Orthogonal to shrinking feature size, a 3-D integrated VLSI system would deploy multiple device layers to improve integration density. Moreover, since the vertical inter-chip interconnects could provide a shortcut to break long signal paths, a 3-D IC would have opportunity for improved circuit performance. Inspired by the great potential, many 3-D integration schemes and fabrication technologies have been proposed in the last a few years.

As pioneers in this new 3-D arena, the authors of this book designed a new 3-D integration scheme, so-called 2.5-D integration. According to this concept, a VLSI system is built as a 3-dimentional assembling of monolithic chips with small-scaled inter-chip interconnections. With a carefully designed, incremental and hierarchical testing methodology, this approach would largely overcome the accumulative yield loss problem hindering other 3-D integration schemes.

In this book, the authors evaluated the feasibility of the 2.5-D integration from both cost and performance perspectives. They established an analytical cost model to compare the manufacturing cost of different VLSI integration styles. The cost analysis shows that the 2.5-D scheme could offer significant cost saving over other schemes. Secondly, the authors performed design case studies on real-world designs. These studies demonstrate the strong potential of 2.5-D integrated designs for higher performance. To study the characteristics of 3-D layouts, the authors constructed a prototype EDA tool-chain consisting of 2.5/3-D floorplanning, placement, and routing tools. With these tools, a synthesized netlist could be automatically implemented as manufacturable layout.

To the best of my knowledge, this book is the first one to give a complete overview of the 3-D integration problem. It would provide valuable information for readers from various communities, such as semiconductor fabrication process developers, IC designers, and EDA R&D practitioners. The book could also serve as an excellent reference for graduates majoring in microelectronics.

Prof. Zhihua Wang
Professor
Institute of Microelectronics
Tsinghua University
Beijing, China

November, 2008

此为试读，需要完整PDF请访问：www.ertongbook.com

Acknowledgements

This book is based on the first author's Ph.D. work. So we would like to thank his Ph.D. committee members, Dr. Wilfred Haensch, Prof. Larry Pileggi, Prof. Radu Marculescu, and Dr. Herman Schmit, for reviewing this research work and providing valuable feedback. We are extremely grateful to Dr. Herman Schmit and Dr. David Whelihan, who kindly provide us with the complete design data of PipeRench. During the whole period of this research, we have been discussing with many other researchers and their opinions have greatly improved the quality of this work. Here we would like to show our gratitude to Prof. Rob A. Rutenbar, Prof. Qiao Lin, Dr. P. K. Nag, Prof. Sung-Kyu Lim, Prof. Yuan Xie, Prof Peng Li, Dr. Jingcao Hu, Prof. Chunsheng Liu, Julia Fei, Tao Lin, Dr. Thomas Zanon, Dr. Yi Wang, and many others. The authors are also thankful for Prof. Zhihua Wang for writing the preface.

Contents

List of Figures and Tables ix

1 **Introduction** 1
  1.1 2.5-D Integration 5
  1.2 Enabling Technologies 8
    1.2.1 Fabrication Technology 8
    1.2.2 Testing Methodology and Fault Tolerance Technique 9
    1.2.3 Design Technology 10
  1.3 Objectives and Book Organization 13
  References 16

2 **A Cost Comparison of VLSI Integration Schemes** 21
  2.1 Non-Monolithic Integration Schemes 22
    2.1.1 Multiple-Reticle Wafer 23
    2.1.2 Multiple Chip Module (MCM) 23
    2.1.3 Three-Dimensional (3-D) integration 24
  2.2 Yield Analysis of Different VLSI Integration Approaches 26
    2.2.1 Monolithic Soc 28
    2.2.2 Multiple-Reticle Wafer (MRW) 28
    2.2.3 Three-Dimensional (3-D) Integration 30
    2.2.4 2.5-D System Integration 31
    2.2.5 Multi-Chip Module 34

2.2.6 Summing Up 35
2.3 Observations 37
References 38

3 Design Case Studies

3 Design Case Studies 42
3.1 Crossbar 43
3.2 A 2.5-D Rambus DRAM Architecture 46
3.2.1 Tackle the Long Bus Wire 46
3.2.2 Serialized Channel in the 3rd Dimension 48
3.3 A 2.5-D Redesign of PipeRench 50
3.3.1 The 2.5-D Implementation 52
3.3.2 Simulation Results 54
3.4 A 2.5-D Integrated Microprocessor System 56
3.4.1 A 2.5-D Integrated Microprocessor System 57
3.4.2 An Analytical Performance Model 62
3.4.3 Detailed Performance Simulation for Reduced Memory Latency 66
3.5 Observations 69
References 71
4 An Automatic 2.5-D Layout Design Flow 74
4.1 A 2.5-D Layout Design Framework 75
4.1.1 2.5-D Floorplanning 77
4.1.2 2.5-D Placement 78
4.1.3 2.5-D Global Routing 78
4.2 Observations 81
References 81

5 Floorplanning for 2.5-D Integration 83
5.1 Floorplan Level Evaluation—Category 2 Circuits 87
5.1.1 Technique 87
5.1.2 Results 89
5.2 Floorplan Level Evaluation—Category 3 Circuits 91
5.2.1 Technique 91
5.2.2 Results 92
5.3 Thermal driven floorplanning 93
5.3.1 Chip Level Thermal Modeling and Analysis for 2.5-D Floorplanning 95
5.3.2 Coupled Temperature and Leakage Estimation 99
5.3.3 2.5-D Thermal Driven Floorplanning Techniques 105
5.3.4 Experimental results 107
5.4 Observations 111
References 113

6 Placement for 2.5-D Integration 117
6.1 Pure Standard Cell Designs 119
6.1.1 Placement Techniques 120
6.1.2 Benchmarks and Layout Model 123
6.1.3 Evaluation of Vertical Partitioning Strategies 125
6.1.4 Wire length scaling 126
6.1.5 Wire length reduction 129
6.1.6 Wire Length vs. Inter-Chip Contact Pitch 133
6.2 Mixed Macro and Standard Cell Designs 134
6.2.1 Placement Techniques 136
6.2.2 Results and Analysis 138

6.3 Observations 140
References 142

7 A Road map of 2.5-D Integration 144
7.1 Stacked Memory 145
7.2 DRAM Integration for Bandwidth-Demanding Applications 147
7.3 Hybrid System Integration 151
7.4 Extremely High Performance Systems 155
7.4.1 Highly Integrated Image Sensor System 155
7.4.2 Radar-in-Cube 158
References 160

8 Conclusion and Future Work 164
8.1 Main Contributions and Conclusions 165
8.2 Future Work 168
8.2.1 Fabrication Technology for 2.5-D Systems 169
8.2.2 Testing Techniques for 2.5-D Integration 171
8.2.3 Design Technology for 2.5-D Integration 173
References 186

Index 188

List of Figures and Tables

Figure 1.1 Actual chip complexity increases faster than Moore's law 2
Figure 1.2 An imaginary 2.5-D system (*see colour plate*)................................ 5
Figure 2.1 Total consumed silicon area of multiple-reticle wafer.................... 30
Figure 2.2 Silicon area of the 2.5-D implementation with 4 slices of chips 33
Figure 2.3 Silicon area of the 2.5-D implementation.. 34
Figure 2.4 Silicon area of the MCM implementation 35
Figure 2.5 Silicon area comparison of different integration schemes 36
Figure 2.6 System planning for future VLSI systems.. 38
Figure 3.1 Stick diagram of a monolithic crossbar (*see colour plate*) 44
Figure 3.2 Stick diagram of a 2.5-D crossbar (*see colour plate*) 45
Figure 3.3 Rambus DRAM 46
Figure 3.4 2.5-D Rambus DRAM 48
Figure 3.5 RDRAM memory system 49
Figure 3.6 3-D Rambus DRAM: 4-channel configuration................................. 50
Figure 3.7 Original monolithic implementation of PipeRench 51
Figure 3.8 Critical path of PipeRench system.. 52
Figure 3.9 The 2.5-D re-design of PipeRench (*see colour plate*).................... 53
Figure 3.10 Alpha 21364 floorplan and memory bus placement 58
Figure 3.11 A 2.5-D stacked microprocessor and DRAM 60
Figure 3.12 A diagram of computer system... 60
Figure 3.13 CPI calculation 63

此为试读，需要完整PDF请访问：www.ertongbook.com

Figure 3.14 CPI with regard to main memory latency and L2 cache miss rate (*see colour plate*) 65
Figure 3.15 IPC Speedup by reduced memory latency 68
Figure 4.1 A 2.5-D layout synthesis framework 76
Figure 4.2 2.5-D routing graph 79
Figure 5.1 2.5-D floorplanning 87
Figure 5.2 A floorplan example 89
Figure 5.3 Insert a 0-weight cell 91
Figure 5.4 2.5-D thermal-driven floorplanning flow 95
Figure 5.5 A 3-D IC with two stacked chip layers in a package 96
Figure 5.6 Thermal interactions between a region of the top transistor layer to all other regions on both transistor layers (not all interactions are drawn) 98
Figure 5.7 Thermal simulation of a set of floorplans with varying total area and aspect ratio (only one stacked layer is shown for each case) 99
Figure 5.8 Modeling the temperature dependency of the leakage power using a linear model 101
Figure 5.9 Leakage power distribution is confined within the placed circuit blocks 103
Figure 5.10 The distribution of wire length and temperature gradient 109
Figure 5.11 Temperature snapshots of the thermal driven floorplanning with Benchmark AMI49. Both the maximum temperature and the temperature gradient are reduced during the optimization (*see colour plate*) 111
Figure 6.1 2.5-D placement problem (*see colour plate*) 119
Figure 6.2 2.5-D placement process 121
Figure 6.3 Wire length reductions vs. vertical partitioning 126

Figure 6.4 Monolithic and 2.5-D placements for the same design................ 127
Figure 6.5 A profile of wire length reduction 128
Figure 6.6 Wire length reductions of standard cell placement....................... 130
Figure 6.7 Wire length distribution of one design... 132
Figure 6.8 Interconnect power comparison—2-D and 2.5-D solutions.......... 133
Figure 6.9 Wire length vs. pitch of inter-chip contact pitch........................... 134
Figure 6.10 Block splitting during mixed placement...................................... 138
Figure 6.11 Wire length reductions of mixed placement 140
Figure 7.1 Road map for the development of 2.5-D ICs................................ 145
Figure 7.2 Flash memory capacity in cellular phones (adapted from)........... 146
Figure 7.3 Peak memory bandwidths of major NVidia GPUs 148
Figure 7.4 Intel's wire-bonded stacked Chip Scale Packaged flash memory (courtesy of Intel Corporation) 148
Figure 7.5 Normalized clock rate vs. peak memory bandwidth of NVidia..... 149
Figure 7.6 Tile-based multiprocessor architecture... 151
Figure 7.7 A multi-chip wireless handset solution (courtesy of Texas Instruments)... 152
Figure 7.8 Passive components in package.. 155
Figure 7.9 An image sensor system digram 156
Figure 7.10 A 2.5-D camera/IR sensor system 158
Figure 7.11 Computational demands for military radar systems (adapted from) 159
Figure 7.12 Block diagram of a radar system 159
Figure 7.13 2.5-D implementation of a radar system.................................... 160
Figure 8.1 Area power I/O for 2.5-D integration (*see colour plate*).............. 168
Figure 8.2 MEMS based inter-chip contact (*see colour plate*) 170
Figure 8.3 Design flow for 2.5-D ICs 184

Table 1.1 Design variables involved in designing a 2.5-D system 11
Table 2.1 Wafer bonding based 3-D integration technologies 25
Table 2.2 Values for the major parameters of our cost model 28
Table 3.1 SPICE simulation on the critical path 55
Table 3.2 Configuration of target microprocessor 58
Table 3.3 SPEC2000 benchmark programs under study 67
Table 3.4 IPC improvement by Reduced Memory Latency 68
Table 5.1 2-D and 2.5-D floorplans for Category 2 designs 90
Table 5.2 2-D and 2.5-D floorplans for Category 3 designs 93
Table 5.3 2.5-D thermal-driven floorplans with different weighting factors for thermal cost 108
Table 5.4 3-D floorplans with and without thermal concern 110
Table 6.1 Placement benchmarks 123
Table 6.2 Worst-case wire length reduction for nets with large fan-out 129
Table 6.3 Wire length comparison of standard cell placements 131
Table 6.4 Mixed Layout Benchmarks 135
Table 6.5 Wire length characteristics of mixed placement 139

1 Introduction

Yangdong (Steve) Deng

Institute of Microelectronics, Tsinghua University
Beijing 100084, P. R. China, dengyd@tsinghua.edu.cn

Wojciech P. Maly

Electrical and Computer Engineering Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213-3891 USA, maly@ece.cmu.edu

Abstract In this chapter we elaborate on the need for new 3-dimensional VLSI paradigms by extrapolating the trend of technology development. On such a basis, we will propose our target 2.5-D integration scheme, and then explain its advantages. The fabrication, testing, and design technologies to enable the 2.5-D scheme are explained. Finally we are going to introduce the objectives and organization of this book.

Keywords 3-dimensional VLSI, 2.5-D integration, inter-chip contact, interconnection, fabrication, test, design technology.

The semiconductor industry has been and will continue to be driven by the consumer demands for superior performance and functionality. To keep pace with such demands, it is essential to maintain the momentum of shrinking process feature size so as to pack more devices on a single silicon die. As a matter of fact, the complexity of the integrated circuit (IC) system has always been growing at the speed delineated by the Moore's Law since the invention of the first integrated circuit. From the beginning of the 1990s, the speed of increasing complexity has even been accelerated with the introduction of broadband and multimedia applications. One such exemplar application is illustrated in Fig. 1.1, where each dot representing the number of gates on a given generation of NVidia's flagship graphic processing unit (GPU)[1]. The dotted line indicates the number of gates predicted by the Moore's Law. Clearly, the GPU chips would integrate a greater number of transistors than that predicted by the Moore's Law. Similar trends could be observed in other applications domains like wireless chipsets[2].

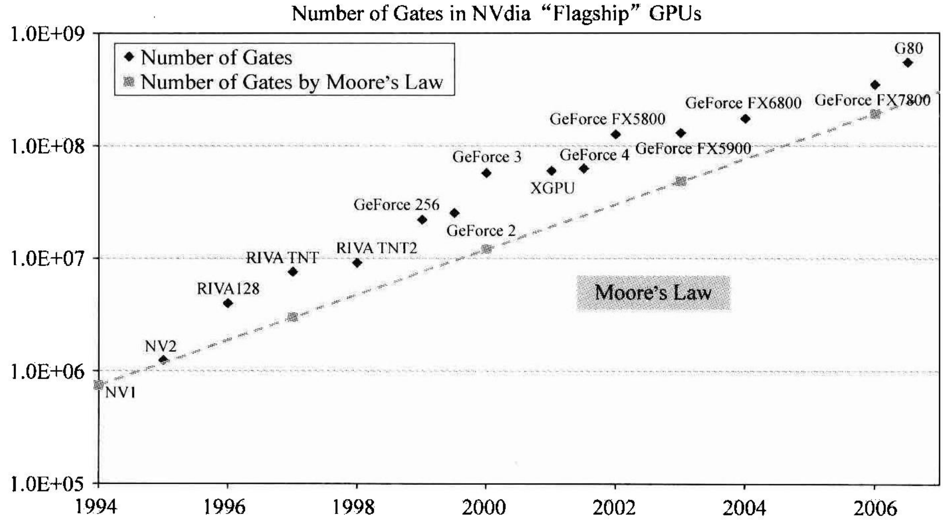

Figure 1.1 Actual chip complexity increases faster than Moore's law

Despite the strong need for more silicon real estate, the basic physics laws would not allow an unlimited scaling of device dimension. The limit would have to be reached in the next 10 – 20 years, if no replacement technologies come up during this time frame.

Meanwhile, emergent very large scale integration (VLSI) systems are incurring overwhelming complexity as the main-stream process technology is now moving to the 45 nm node. Among many difficulties, the following three problems are inherent to the very nature of monolithic integration:

Interconnection Performance Historically, the functionality to be integrated in a single chip at every technology generation has always exceeded the capacity provided by pure scaling. To accommodate the extra transistors, the chip size has always been increasing since the invention of the first IC[3]. The problem is that, the interconnection length, especially worst-case interconnection length, has to increase accordingly. Starting from the 0.25 μm technology node, the interconnection delay of long on-chip wires has become the dominant part determining system performance[3]. Unfortunately, interconnection delay is very hard to predict before the circuit is actually laid out. As a result, IC architects usually take considerable efforts to manage those long wires with the help of advanced electronic design automation (EDA) software.

Mixed Technology Integration Modern System-on-Chips (SoCs) typically have to integrate heterogeneous, mixed-technology components. The technology heterogeneity certainly complicates the underlying fabrication processes. The fabrication cost of today's semiconductor processes is already skyrocketing with the shrinking of the feature size[4]. A single mask set as well as the corresponding probe for digital ASICs is reported to soon reach $5 million at the 45 nm technology node[5,6], while the price of a finished wafer in a RF-CMOS process is higher than that in a pure CMOS process by at least 15%[7]. Meanwhile, it is worth mentioning that certain RF circuits would not benefit from a finer process

此为试读，需要完整PDF请访问：www.ertongbook.com