DATE’09 Friday Workshop on

3D Integration

— Technology, Architecture, Design, Automation, and Test —

Electronic Workshop Digest

Palais des Congrès Acropolis – Nice, France
Friday April 24, 2009
Preface

3D Integration is a promising technology for extending Moore’s momentum in the next decennium, offering higher transistor density, faster interconnects, heterogeneous technology integration, and potentially lower cost and time-to-market. But before 3D chips can be produced, new capabilities are needed: process technology, architectures, design methods and tools, and manufacturing test solutions. The goal of this Workshop is to bring together researchers, practitioners, and others interested in this exciting and rapidly evolving field, in order to update each other on the latest state-of-the-art, exchange ideas, and discuss future challenges.

General Chair
Erik Jan Marinissen
IMEC
Leuven, Belgium

Program Co-Chair
Yann Guillou
ST-Ericsson
Grenoble, France

Program Co-Chair
Geert Van der Plas
IMEC
Leuven, Belgium

DATE’09 Friday Workshop on
‘3D Integration’
Acropolis, Nice, France – April 24, 2009
# Table of Contents

**Keynote Address:** The Promise of Through-Silicon Vias  
Sitaram Arkalgud – Sematech, US ................................................................. 6

**Invited Talk:** Requirements for Design-for-3D Environment  
Riko Radojcic – Qualcomm, US ................................................................. 40

**Paper 2.1:** 3D Integration Perspective for Multimedia Products  
Dominique Henoff, Laurent Bonnot – ST Microelectronics, FR .......................... 69

**Paper 2.2:** Z-Axis Interconnections: Fabrication and Electrical Performance  
Voya R. Markovich, Rabindra N. Das, Michael Rowlands, John Lauffer –  
Endicott Interconnect Technologies, US ........................................................ 86

**Paper 2.3:** Impact of Design Choices on 3D SiC Manufacturing Cost  
Dimitrios Velenis, Michele Stucchi, Erik Jan Marinissen, Eric Beyne – IMEC, BE .... 88

**Paper 3.1:** Clock and Power Distribution Networks for 3D Integrated Circuits  
Ioannis Savidis, Eby G. Friedman – University of Rochester, US;  
Vasilis F. Pavlidis, Giovanni De Micheli – LSI-EPFL, CH ................................. 114

**Paper 3.2:** Hierarchical Cache System for 3D-Multi-Core Processors in Sub 90nm CMOS  
Kumiko Nomura, Keiko Abe, Shinobu Fujita, Yasuhiko Kurosawa, Atsushi Kageshima –  
Toshiba Corp., JP ............................................................................................... 117

**Paper 3.3:** Test Strategies for 3D Die-Stacked Integrated Circuits  
Dean L. Lewis, Hsien-Hsin S. Lee – Georgia Institute of Technology, US ........... 149

**Panel Session:** “The Future of 3D Integration From All Angles”  
Moderator: Pol Marchal – IMEC, BE ............................................................... 193
Panelists: Krishnendu Chakrabarty – Duke University, US ................................. 194  
Lisa McIlrath – R3Logic .................................................................................. 208  
Paul Siblerud – Semitool, US ........................................................................... 212  
Nicolas Sillon – CEA-LETI, FR .......................................................................... 216  
Pascal Urard – ST Microelectronics, FR .............................................................. 218  
Geert Van der Plas – IMEC, BE ........................................................................ 223
### Table of Contents (continued)

#### Posters

1. **3D Technologies and Data Structures – An Overview**  
   Robert Fischbach, Jens Lienig – Dresden University of Technology, DE  
   …………………………………… 233

2. **System-Level Exploration of 3D Interconnection Schemes**  
   Kostas Siozios, A. Papanikolaou, Alexandros Bartzas, Dimitrios Soudris –  
   National Technical University of Athens, GR  
   …………………………………… 236

3. **3D Integrated Smart Antenna Systems**  
   Nakul Haridas, Tughrul Arslan – University of Edinburgh, UK  
   …………………………………… 239

4. **3D Integration Program Overview**  
   Laurent Bonnot, Pascal Ancéy, Damien Riquet, Pascal Urard –  
   ST Microelectronics, FR; David Henry, Astrid Astier – CEA Léti, FR  
   …………………………………… 241

5. **Topology Exploration and Buffer Sizing for 3D Networks-on-Chip**  
   Alexandros Bartzas, Kostas Siozios, Dimitrios Soudris –  
   National Technical University of Athens, GR  
   …………………………………… 244

6. **Closed-Form Equations for Through-Silicon Via Parasitics in 3D ICs**  
   Roshan Weerasekera, Dinesh Pamunuwa, Matt Grange –  
   Lancaster University, UK; Hannu Tenhunen, Li-Rong Zheng – KTH, SE  
   …………………………………… 247

7. **3D-NOCs, TSVs, Asynchronous Circuits, and Serial Vertical Links**  
   A. Sheibanyrad, F. Pétrot – TIMA/SLS, FR  
   …………………………………… 251

8. **Modular Modeling of RF Behavior of Interconnect Structures in 3D Integration**  
   Jörn Stolle, Sven Reitz, Peter Schneider, Andreas Wilde –  
   Fraunhofer Institute for Integrated Circuits, DE  
   …………………………………… 254

9. **Evaluating Noise Coupling Issues in Mixed-Signal 3D ICs**  
   Liuchun Cai, Ramesh Harjani – University of Minnesota, US  
   …………………………………… 257

10. **Design, Verification and Simulation of 3D Circuit**  
    Guruprasad Katti, Bart De Wachter, Marc Nellis, Morin Dehan, Miroslav Cupak,  
    Kris Croes, Gerd Beeckman, Pol Marchal, Michele Stucchi – IMEC, BE;  
    Wim Dehaene – Katholieke Universiteit Leuven, BE  
    …………………………………… 259

11. **Examination of Delay and Signal Integrity Metrics in TSVs**  
    Matt Grange, Roshan Weerasekera, Dinesh Pamunuwa –  
    Lancaster University, UK; Hannu Tenhunen – KTH, SE  
    …………………………………… 260
<table>
<thead>
<tr>
<th>Posters</th>
</tr>
</thead>
</table>
| **12. On-Chip Waveform Capturing Functionality Partitioned for 3D Realization**  
Yuuki Araga, Yoji Bando, Takushi Hashida, Makoto Nagata – Kobe University, JP …… 265 |
| **13. Thermal Aware Test Scheduling for Stacked Multi-Chip Modules**  
Vinay N.S., Virendra Singh – Indian Institute of Science, IN; Erik Larsson – Linköping University, SE ……………………………………………………………………………… 267 |
| **14. A Prospective Analysis of High-Frequency Cross-Coupling Mechanisms for the Next 3D Packaging Generation**  
Roberto Antonicelli – ST-Ericsson, BE ……………………………………………………………… 273 |
| **15. Power Integrity Issues in 3D ICs using TSVs**  
Waqar Ahamd, Qiang Chen, Roshan Weerasekera, Hannu Tenhunen, Lirong Zheng – KTH, SE ……………………………………………………………………………………… 274 |
| **16. Core Test Wrapper Optimization for 3D ICs with TSVs**  
Brandon Noia, Krishnendu Chakrabarty – Duke University, US …………………………… 276 |
| **17. Application of Substrate Noise Simulation Methodology to 3D Stacking**  
S. Bronckers, Geert Van der Plas, Pol Marchal – IMEC, BE; Y. Rolain – Vrije Universiteit Brussel, BE ……………………………………………………………………………… 278 |
| **18. Impact of Thinning on 65nm Device Performance**  
| **19. Bandwidth Optimization for Through Silicon Via Bundles in 3D ICs**  
Awet Yemane Weldezion, Li-Rong Zheng, Hannu Tenhunen – KTH, SE; Roshan Weerasekera, Dinesh Pamanuw – Lancaster University, UK ………………………………………………… 283 |
| **20. Wafer-Level Based Manufacturing Technologies for Realization of TSV and 3D-Based Applications**  
Stefan Pargfrieder, Daniel Burgstaller, Otto Bobenstetter, Bioh Kim – EV Group, AU … 288 |
| **21. SOC Test Architecture and Method for 3D-IC**  
Chih-Yen Lo, Yu-Tsao Hsing, Li-Ming Denq, Cheng-Wen Wu – National Tsing Hua University, TW ……………………………………………………………………………… 315 |
| **22. Vertical Links for 3D Network-on-Chip**  
Igor Loi, Federico Angiolini, Luca Benini – Università di Bologna, IT …………………… 317 |
The Promise of Through-Silicon Vias

Sitaram Arkalgud – Sematech, US

Abstract:

The twin drivers of all advances in the semiconductor industry have been ever increasing performance and productivity. With tremendous strides in lithography and device development over the last several decades, achieving both has been possible. However, with fundamental issues and cost concerns surrounding new technology elements at 32nm and below, the viability of traditional lithography and device scaling to stay on the productivity curve becomes questionable. One of the technologies gaining popularity has been Through-Silicon Vias (TSVs) for stacking chips in the third dimension. This presentation will discuss the merits of 3D TSVs, state of the art on 3D, the risks and challenges involved, the timeline, the need for understanding the cost implications and manufacturability, and the necessity for standardization and classification.
DATE 2009

The Promise Of Through Silicon Vias

Sitaram Arkalgud PhD

Director – Interconnect

April 24, 2009
Outline

• The Promise of TSVs
• Process & Integration Considerations
• Challenges
• The Need for Roadmaps
• Conclusions
IC Scaling – the Future

- Shrinks have driven density and performance
  - Lithography requires immersion, double exposure, double patterning and EUV to maintain shrink roadmap
  - Transistors require high k dielectrics, metal gates and 3D structures (FinFETs)
  - Metallization requires complex low k dielectric integration

  ➔ Steadily rising Process Complexity

- 3D stacking can provide high density
  - With delayed introduction of above elements of conventional scaling
  - Without area driven yield penalty (stacking smaller die instead of a large 2D die)*

- Short Through Silicon/Substrate Vias (TSVs) can reduce interconnect delay compared to long lines in 2D**

  ➔ TSVs can contain rising complexity by extending technology lifecycles

* L Smith (SEMI) - 3D Architectures for Semiconductor Integration and Packaging 2007
** S. Vitkavage (SEMI) – 3D Architectures for Semiconductor Integration and Packaging 2005
Through Silicon/Substrate Vias

SiP:
- Low cost
- High Functionality

SoC:
- High Performance
- High Functionality

TSV:
- High Performance
- High Functionality
- Lower Power
- Lower Cost?
Benefits of 3D ICs With TSVs

• Improved form factor
  – Can provide smaller footprint and/or increased density
  – Can result in higher yield

• High Functionality
  – Improved integration of heterogeneous technologies, materials and signals over SoC

• High Performance
  – High bandwidth and shorter wires (lower RC)

• Lower Power Consumption
  – Shorter wires and lower overall I/O count

• Cost Containment Potential
  – **Device level**: Addresses slowdown in productivity gain (from scaling)
  – **Die level**: Functionality, performance, power reduction (compared to SoC)
  – **Wafer level**: Parallelism of fab processing
  – **Factory level**: Optimizes cost structure for each “level” in 3D stack
  – **Market level**: Short Time To Market for products
Process & Integration Considerations
Key Process Elements of 3D ICs

- TSV Reactive Ion Etch
- Dielectric liner, barrier and Cu seed
- Void free TSV fill – process and materials
- Bonding of wafers or dies – process and materials
- Thinning and handling of wafers or dies

- Depending on the integration, certain process steps may be optional
- The order of processing can also change
• Considerable experience available from DRAM deep trench capacitor formation
• Conventional (SF6/O2) and Bosch (SF6/O2/C4F8) RIE options available
Dielectric Liner Layer Deposition

- Dielectric liner layers must have high aspect ratio coverage capability
  - Organics conformally deposited by CVD
  - PECVD dielectrics
  - Thermal CVD dielectrics
Barrier and Seed Metallization

- The silicon will be contaminated by copper if the barrier layer is penetrated.
- Copper seed layer must be continuous or the plating will not proceed.

- High aspect ratio TSVs will require CVD or ALD seed layers (Eg: Ru)
Copper Filled TSV

Enclosure of a void can be bad in PVD or CVD. Can be catastrophic in plating, because it generally encloses plating solution. Upon subsequent heating of the wafer:
Via Filling Material

- **Doped CVD Polysilicon**
  - Very high conformality
  - Generally limited to ~ 1 µm in thickness
  - Resistivity is ~ 500 x copper
  - Temperature tolerance consistent with FE processes

- **CVD Tungsten**
  - Limited to ~ 1 µm in thickness by stress
  - Vias limited to about 1.5 µm (must be closed from the sides by conformal deposition)
  - More limited in temperature than polysilicon, oxidation prone, better in via-last or via-mid integration

- **Plated Copper**
  - Lowest resistivity
  - Requires continuous barrier and seed metal (aspect ratio limitation)
  - Can fill very large vias, either conformally or bottom up
  - Inconsistent with via-first; can be used in via mid or via last applications

- **Solder**
  - Similar to Cu
  - Easily available in Assembly and lends itself to via last
Polysilicon’s via resistance limits its use in TSVs
Aspect ratio limitation causes thinner substrates/handling issues for Cu filled TSVs
### Permanent Bond Technologies

<table>
<thead>
<tr>
<th>Type</th>
<th>Technology</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dielectric</td>
<td>Adhesive</td>
<td>Via-Last&lt;br&gt;Medium temperature/pressure&lt;br&gt;Defect tolerant</td>
</tr>
<tr>
<td></td>
<td>Oxide Fusion</td>
<td>Via-Last&lt;br&gt;Low temperature/pressure&lt;br&gt;High sensitivity to defects&lt;br&gt;High flatness/roughness requirements</td>
</tr>
<tr>
<td>Metallic</td>
<td>Cu – Cu&lt;br&gt;Au – Au Alloys</td>
<td>Via-First&lt;br&gt;Readily available in fab or assembly&lt;br&gt;Sensitive to defects, metal coverage&lt;br&gt;High temperature/pressure</td>
</tr>
<tr>
<td>Hybrid</td>
<td>Dielectric + Metal</td>
<td>Via-First&lt;br&gt;Simultaneous mechanical and electrical bonds&lt;br&gt;Avoids underfill requirement&lt;br&gt;Higher requirements on process control and material properties</td>
</tr>
</tbody>
</table>
Cu-Cu Bond with Dielectric Recess

- Cu-Cu bonded interface; no Cu- SiO₂ contact observed
- ~10 nm diameter voids observed
# 3D-Specific Unit Process Challenges

<table>
<thead>
<tr>
<th><strong>Wafer Fab Centric</strong></th>
<th><strong>Challenges</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>– TSV RIE</td>
<td>Productivity</td>
</tr>
<tr>
<td>– Metallization</td>
<td><strong>Barrier/seed step coverage (AR-driven)</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Metal fill throughput</strong></td>
</tr>
<tr>
<td></td>
<td><strong>CMP of thick Cu</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Assembly Centric:</strong></th>
<th><strong>Challenges</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>– Bonding</td>
<td><strong>Choice of bond type – temperature, force, alignment accuracy, defectivity, reliability, productivity</strong></td>
</tr>
<tr>
<td>– Thinning/dicing</td>
<td><strong>Ultrathin wafer/die handling, stress, TTV, and die strength (driven by TSV diameter and aspect ratio)</strong></td>
</tr>
<tr>
<td>– Handle wafer</td>
<td><strong>Several options – need to be narrowed based on integration, technical merit, and productivity</strong></td>
</tr>
</tbody>
</table>
Metrology Needs

- Via depth - SEM, TEM, Confocal
- Via defectivity - SEM, TEM, Confocal
- Bond interface defectivity - Scanning Acoustic, IR
- Alignment accuracy - IR, optical/electrical verniers
- Wafer thickness - Capacitance, IR
- Residual stress - Micro-Raman
Wafer/Wafer vs Die/Wafer Integration

• Processing hundreds of die in parallel of can reduce costs considerably in WtW integration

• But …
  – Wafer sizes have to be the same
    • No mixing and matching of wafer sizes
  – Die sizes have to match up
    • Heterogenous integration will require design rethink
  – Yield has to be (very) high
    • Restricts choice of technologies
    • Biggest cost factor

➔ Die to Wafer integration has better cost structure in the near term
  • Serial processing, but …
  • Favorable stack yields though use of KGD
  • Mix and match capability offers very short Time To Market potential
Yield Impact – WtW and DtW

Compare DTW @ 60 dph align/bond throughput with WTW

DTW starts out more expensive than WTW because of align/bond ($2) and test ($0.40) costs, but becomes relatively less expensive as defect density increases.
Integration Options

- Via-first vs. via-last
- Face-to-face vs. back-to-face
- Bonding: Dielectric (BCB, oxide) vs. metallic (Cu, Au, etc) vs. hybrid
- Fab process (front end) vs. assembly (back end)
- ....
Integration Option Decision Tree

• What are the needs?
  – Density
  – Performance
  – Functionality

• What products address these needs?
• What integrations support these products?
• What reliability specs are driven by these integrations?
• Business: What is the cost differential compared to current technology? Will the current technology actually work in the future?
Product Requirements

• Via count
• Via density
  – Drives via dimensions and pitch
  – Aspect ratio is fixed by fill material choice
  – In turn, drives die/wafer thickness
• Redistribution layer
  – Relaxes pitch requirements
  – Permits die shrinks, without redesign of primary levels
3D Drivers

**Product Needs/Applications**

**Density/form factor**
- Memory stacking

**Functionality/Performance**
- Processor + SRAM cache

- Logic + SRAM + DRAM + Flash Memory
- 3D equivalent of SoC
- Beyond CMOS  
  - Stacking CMOS “levels” with optoelectronics, MEMS, biosensors, etc.

**Technologies**

- Via-last
- Wafer-to-wafer or die-to-wafer

- Two-level, face-to-face wafer-to-wafer or die-to-wafer

Die-to-wafer heterogeneous integration
Challenges
SEMATECH Survey at SEMICON West: Effort Needed for a Manufacturable 3D

Top 5 Categories:

2006
- Design Tools’ Availability
- Product Design in 3D
- Thermal Modeling
- Test Methodology, Prototyping Capabilities
- Standards

2007
- Product Design in 3D
- Design Tools’ Availability
- Test Methodology, Thermal Modeling, Equipment/Wafer Standards
- Yield Modeling (new category)
- Tool development, Basic Integration,

<table>
<thead>
<tr>
<th>Capabilities</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tool Development</td>
<td>60</td>
<td>71</td>
</tr>
<tr>
<td>Materials Screening</td>
<td>46</td>
<td>44</td>
</tr>
<tr>
<td>Unit Process Development</td>
<td>63</td>
<td>60</td>
</tr>
<tr>
<td>Metrology</td>
<td>63</td>
<td>58</td>
</tr>
<tr>
<td>Basic Integration</td>
<td>69</td>
<td>71</td>
</tr>
<tr>
<td>Early Reliability</td>
<td>68</td>
<td>65</td>
</tr>
<tr>
<td>Prototyping Capabilities</td>
<td>73</td>
<td>59</td>
</tr>
<tr>
<td>Design Tools' Availability</td>
<td>86</td>
<td>78</td>
</tr>
<tr>
<td>Product Design in 3D</td>
<td>84</td>
<td>80</td>
</tr>
<tr>
<td>Test Methodology</td>
<td>73</td>
<td>76</td>
</tr>
<tr>
<td>Roadmapping</td>
<td></td>
<td>67</td>
</tr>
<tr>
<td>Chip to Chip Wiring Standards</td>
<td>72</td>
<td>50</td>
</tr>
<tr>
<td>Equipment and Wafer Handling Standards</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>Thermal Modeling</td>
<td>83</td>
<td>76</td>
</tr>
<tr>
<td>Yield Modeling</td>
<td></td>
<td>72</td>
</tr>
</tbody>
</table>

Surveys had 30–50 participants including IDMs, fabless companies, A&P and Universities.
SEMATECH Survey at SEMICON West:
Key Business Challenges

### Top 3 Business Challenges:

**2006:** Cost, Timing and Co-Design  
**2007:** Cost, Timing and Standards/Co-Design/Prototyping

<table>
<thead>
<tr>
<th>Challenge</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unclear Cost Benefits</td>
<td>72</td>
<td>67</td>
</tr>
<tr>
<td>Unclear Performance Benefits</td>
<td>32</td>
<td>41</td>
</tr>
<tr>
<td>Timing of 3D Technology Introduction</td>
<td>56</td>
<td>52</td>
</tr>
<tr>
<td>Design Non Recurring Expense (NRE)</td>
<td>24</td>
<td>7</td>
</tr>
<tr>
<td>Does Not Align to Product Portfolio</td>
<td>24</td>
<td>2</td>
</tr>
<tr>
<td>Lack of Standards</td>
<td>12</td>
<td>35</td>
</tr>
<tr>
<td>Fab and Packaging Co-Design Issues</td>
<td>40</td>
<td>35</td>
</tr>
<tr>
<td>Unavailability of (or lack of access to) Prototyping Capabilities</td>
<td>12</td>
<td>35</td>
</tr>
<tr>
<td>Other</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Surveys had 30–50 participants from IDMs, fabless companies, A&P and Universities

DATE 2009. Nic D. Fraser
3D Program at SEMATECH

Option Assessment

- **Material:**
  - Silicon
  - SOI
  - Other

- **Via Formation:**
  - Via First
  - Via Last

- **Bonding:**
  - Metal
  - Dielectric
  - Polymer
  - Hybrid

- **Integration:**
  - Face to Face
  - Die to Face
  - Wafer to Wafer
  - Die to Wafer

Option Narrowing

- **Cost Modeling**
- **Performance Evaluations**
- **Risk Assessments**
- **Product Requirements**
- **Tool Benchmarking**
- **Process Benchmarking**

Technology Development

- **Materials**
- **Unit Process Integration**
- **Device Performance**
- **Circuit Performance**
- **Early Reliability**
- **Design Rules**
- **Design & Test**
- **Equipment Hardening**
- **Cost Effectiveness**

Manufacturing Solution

- **300mm test bed for member evaluations**

**Industry Infrastructure: Roadmaps, Standards, Methodologies**

2005 - 2007

2008 - 2010
The Need For Roadmaps ....
TSV Roadmaps (I)
(in Emerging Technologies section of 2007 ITRS Interconnect chapter)


<table>
<thead>
<tr>
<th>Table INTO6</th>
<th>High Density Through Silicon via Draft Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Year of Production</strong></td>
<td>2007</td>
</tr>
<tr>
<td>DRAM % Pitch (nm) (contacted)</td>
<td>63</td>
</tr>
<tr>
<td>MPU/ASIC Metal 1 % Pitch (nm) (contacted)</td>
<td>68</td>
</tr>
<tr>
<td>MPU Physical Gate Length (nm)</td>
<td>23</td>
</tr>
<tr>
<td><strong>Minimum Interlayer HDTV Contact Pitch (μm)</strong></td>
<td></td>
</tr>
<tr>
<td>High Density</td>
<td>3.2–5.0</td>
</tr>
<tr>
<td>HDTV Diameter (μm)</td>
<td></td>
</tr>
<tr>
<td>High Density</td>
<td>1.6–2.5</td>
</tr>
<tr>
<td><strong>Maximum Via Density (cm²)</strong></td>
<td></td>
</tr>
<tr>
<td>High Density</td>
<td>9.77E+0.6</td>
</tr>
<tr>
<td><strong>Minimum Face-to-Face Pitch (μm)</strong></td>
<td></td>
</tr>
<tr>
<td>High Density</td>
<td>5.00</td>
</tr>
<tr>
<td><strong>Maximum Layer Thickness (μm)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Total Thickness Variation (μm)</strong></td>
<td>&lt;1</td>
</tr>
</tbody>
</table>
# TSV Roadmaps (II)
(2007 ITRS Assembly & Packaging Section p 34)


<table>
<thead>
<tr>
<th>Year of Production</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numbers of stacked die using TSV</td>
<td>3 - (8)</td>
<td>6</td>
<td>9</td>
<td>&gt;9</td>
<td>&gt;9</td>
<td>&gt;9</td>
<td>&gt;9</td>
</tr>
<tr>
<td>Minimum TSV pitch</td>
<td>10.0</td>
<td>8.0</td>
<td>6.0</td>
<td>5.0</td>
<td>4.0</td>
<td>3.8</td>
<td>3.6</td>
</tr>
<tr>
<td>TSV maximum aspect ratio**</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
</tr>
<tr>
<td>TSV exit diameter(um)</td>
<td>4.0</td>
<td>4.0</td>
<td>3.0</td>
<td>2.5</td>
<td>2.0</td>
<td>1.9</td>
<td>1.8</td>
</tr>
<tr>
<td>TSV layer thickness for minimum pitch</td>
<td>50</td>
<td>20</td>
<td>15</td>
<td>15</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

**This applies for small diameter vias. The larger diameter vias will have a smaller aspect ratio.
ITRS 2008 HDTSV Update

- 2008 Update has been released on public website [http://www.itrs.net/](http://www.itrs.net/)

---

<table>
<thead>
<tr>
<th>Table INTC6 High Density Through Silicon Via Specification [1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>----------------------------------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Minimum Interlayer HDTSV Pitch (µm)</strong> [4]</td>
</tr>
<tr>
<td><strong>Minimum Layer Thickness (µm)</strong> [5]</td>
</tr>
<tr>
<td><strong>Bonding overlay accuracy, δ (3 sigma) (µm)</strong> [6]</td>
</tr>
<tr>
<td><strong>Minimum size bonding pad Øpad =ØTSV +2.1 (µm)</strong></td>
</tr>
<tr>
<td><strong>Minimum pad spacing, STSV (µm)</strong> [7]</td>
</tr>
<tr>
<td><strong>Minimum contact pitch, PTSV= Øpad + STSV (µm)</strong> [8]</td>
</tr>
</tbody>
</table>

- Addresses only wafer to wafer integration
- Assembly & Packaging Table has not changed
- Lack of electrical specs and corresponding links in other sections
  - Needs inputs from Design/System Drivers

Product Driven Roadmaps

Table SYSD1  Major Product Market Segments and Impact on System Drivers

<table>
<thead>
<tr>
<th>Market Drivers</th>
<th>SOC</th>
<th>Analog/MS</th>
<th>MPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Portable/consumer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Size/weight ratio: peak in 2004</td>
<td>Low power paramount</td>
<td>Migrating on-chip for voice processing, A/D sampling, and even for some RF transceiver</td>
<td>Specialized cores to optimize processing per microwatt</td>
</tr>
<tr>
<td>2. Battery life: peak in 2004</td>
<td>Need SOC integration (DSP, MPU)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Function: 2×2/2 years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Time-to-market</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II. Medical</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Cost: slight</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Time-to-market (-1/2 every 2 years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Function:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Form factor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Durability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Conservation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III. Networking</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Bandwidth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Reliability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Time-to-market</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Power:</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Necessary to**
  - Drive consensus on broad applications areas
  - Develop coherence and critical mass around design, test and technology R&D
  - Provide a timeline for development
- **Current TSV roadmaps in the ITRS are technology driven “bottoms up” roadmaps**
  - Need a top down assessment from the System Drivers chapter
- **Risk**
  - Development of technology looking for solutions
  - Development of custom solutions for niche applications

Summary

• 3D can be the new engine to keep the industry on the productivity curve
  – High performance
  – High functionality

• Product requirements and roadmaps are key to driving coherence and critical mass for 3D R&D

• 3D is not a single technology element
  – Productivity benefits can be realized at several levels
    • Requires industry wide coordination across technology, design and test
Abstract:

Qualcomm’s roadmap for 3D technology is outlined, and the corresponding requirements for a holistic design environment necessary to define and implement optimized 3D products are described. The focus is on the design environment and EDA tools necessary for ‘Stage 1’ class of products, consisting of a functionally partitioned two-die stack. The design environment requirements are segregated into three classes of methodologies and the associated EDA technologies.

(a) “TechTuning” technologies required to co-optimize process technology and chip design requirements, and to define and validate the design rules and models required for 3D Design Authoring,

(b) “PathFinding” technologies required to co-optimize system and technology specifications, and to define the optimum architecture for the 3D process and design, and to generate the constraints required for Design Authoring, and

(c) “Design Authoring” flow and the EDA technology upgrades required to implement chip design for 3D stacks.

The status of the collaborative efforts, supported by Qualcomm and a set of partners, to develop and evaluate each of these technologies is summarized. Key results and challenges will be presented.
Design Eco-System for 3D

DATE 09 – W5 Workshop
3D Integration – Technology, Architecture, Design, Automation & Test

Riko Radojcic, Qualcomm
5775 Moorehouse Dr.
San Diego, CA, 92121, USA
rikor@qualcomm.com
## Qualcomm TSS Roadmap

<table>
<thead>
<tr>
<th></th>
<th>TSV</th>
<th># TIERS</th>
<th># VIAS</th>
<th>TSV Diameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAGE1</td>
<td>Cu via after FEOL</td>
<td>2</td>
<td>1000’s</td>
<td>6um/ 50um</td>
</tr>
<tr>
<td>STAGE2</td>
<td>Cu via after FEOL</td>
<td>2-3</td>
<td>10,000s</td>
<td>2-3um/ 25um</td>
</tr>
<tr>
<td>STAGE3</td>
<td>Cu via after FEOL</td>
<td>3 or more</td>
<td>&gt;</td>
<td>&lt;2-3um/ &lt;25um</td>
</tr>
</tbody>
</table>

- Projections based on current realities and outlook
- It is still early days in the technology and sourcing life cycle

**Develop Design Infrastructure for Stage 1**

*but with Focus on Road Map*

TSVs => smaller / Si => thinner / options & challenges => proliferating
Product Design Eco System...

- Generic View of the Chip Design Eco-System
  - a structured flow with many procedures and steps
- Manufacturing Process Information flows UP the flow - *in form of constraints*
  - Rules, tech files, models....
- Product Design Information flows DOWN the flow - *in form of specifications*
  - RTL, NL, GDS ..
- Historically the Hand-Offs and Interfaces were Reasonably Well Defined
  - Open and ‘de-facto’ Standards
  - Between ‘ecosystems’ and Across the Levels of Hierarchy

System Design Eco-System
- architecture, software, etc...

Chip Design Eco-System

Process Tech. Eco-System
- materials, equipment, etc..
Design Eco-System Challenge: Costs

- Chip Design Expensive = 10’s of M$
- Part of the Cause: Iterative Re-Dos and Re-Spins
  - Instability in System Specifications
  - Instability in Process Technology

- This is an Especially Acute Concern with 3D Technologies
  - Allows new and untried degrees of freedom in architecture
  - Allows new and untried sensitivities in the manufacturing process

Riko Radojcic: Qualcomm Inc
Eco-System for 3D Design

- Segment Design Eco-System into 3 Groups
  - “Design Authoring” – actual chip design
    - (Expensive) Actual Chip Design – Output GDS
  - “PathFinding” – system space exploration
    - (Cheap) Quick & Dirty System Design – Output Clean Specs
  - “TechTuning” – physical space exploration
    - (Cheap) Multiphysical Chip Simulation – Output Clean Constraints
PathFinding

a Design Co-Optimization Methodology for TSS
System Level – to – Chip Level
What is this “PathFinding” anyways…

**Ideal**: Black Box Design System that Seamlessly Optimizes Process & Design
- Leverage Existing Design Flow Paradigm
- Optimization = Looping Through the Flow
  - otherwise it is not ‘optimization’
- Current Practices
  - Do Nothing
  - Informal, ad-hoc, guru wing it… or
  - Complete Trial Design

**PATHFINDING IS** a Design Flow that allows:
- **Fast & Iterative** Looping
  - Adjustable trade-off of speed vs. accuracy
- **Structured & Holistic** Analyses
  - A systematic ‘practice’ for optimization
- **Predictive** Analyses
  - It has to be based on predictive models

I/P Variables:
- Process option A, B, C
- Design option X, Y, Z … etc..

O/P Estimates:
- Die Area
- Power
- Yield

Magical “Black Box”
3D System Design Exploration

- Generic Wireless Product Requirements
  - Optimized **Power** and **Cost** for given Performance
- Intrinsic 3D TSS Technology Value Propositions
  - Wire length distributions
  - Heterogeneous technologies w/o off-chip loads
  - Die Sizes and Form Factors
- But … Complex Design Space Trade Off Exploration

  - **Chip Specs**
    - Si Technology Choice
      - 90nm, 65nm, 45nm …
      - Flavors & LM’s..
    - Embedded features
      - On board memories ..
      - On board analog …
    - TSS Flavor Choices
      - Via 1st, Middle, Last…
      - Via size & KOA …
      - I/O configurations …

  - **System Specs**
    - Architecture
      - NoC vs Bus …
      - Hardware vs Software…
    - Performance Requirement
      - Clock, bandwidth, latency…
      - Power and Cost
    - Partitioning
      - Memory/Analog on Logic…
      - Existing vs Custom die …
    - Size and Cost
      - Package selection …

Need Spatially Aware System Simulation Environment

Riko Radojcic : Qualcomm Inc
PathFinding-for-3D Road Map

- **Phased Approach**
- **Past**: Spatially Un-Aware Design Space Exploration
- **Phase 1**: Physical PathFinding
- **Phase 2**: Integrated Architectural PathFinding

**Phase 1: Physical PathFinding**
- Partition & Floorplan & Virtual Route
- Existing Design Specification
  - but at Multiple Levels of Hierarchy
- 3D Technology 'knobs'
- PathFinding-for-3D Road Map
- Phased Approach
- Past: Spatially Un-Aware Design Space Exploration
- Phase 1: Physical PathFinding
- Phase 2: Integrated Architectural PathFinding

**Phase 2: Architectural PathFinding**
- Different Architectures
- Different Target Technologies
- Different Packages
- Integrated Flow with all Knobs

**Past**
- spreadsheets
  - no spatial awareness
  - guru
  - arch level only

**PathFinding**: Spatially Aware Design Space Exploration
- Output Specs rather than Working products
- Focus on Relative Sensitivity Analyses rather than Absolute Accuracy
- Focus on Flexibility rather than Correctness
PathFinding: from System to GDSII in ‘no time’

Product requirements

1. System-level design exploration for 3D
   - Functional partitioning
   - Block-diagram (spreadsheet)
   - System-level design (incl. 3D partitioning)

2. RTL-elaboration: bridging gap between system & physical design
   - Component-level synthesis (Optional)
   - RTL synthesis (for various Technologies)
   - Gate-level architecture (soft, hard macro, gate-level netlist)

3. Physical design prototyping for 3D
   - 3D stack design
   - 3D Thermal/Cost analysis
   - Constraints for design authoring

IP-models (virtual tech.)
- Functional Model: Blackbox, behavioral, RTL
- Parametric Model: Performance, power, cost

Architecture in SystemC (RTL/behavioral/black boxes)

Power/performance/cost

Component lib (virtual 2D tech.)
- Gate lib (virtual 2D tech.)

Gate-level architecture (soft, hard macro, gate-level netlist)

Component-level synthesis (Optional)
- RTL synthesis (for various Technologies)

Performance/Power/Cost

Physical design lib (virtual 2D tech.)
- DRM (virtual 3D tech.)

3D stack design

Riko Radojcic: Qualcomm Inc
Example Physical PathFinding Demonstration

- **Scenario A:** two side by side die on the PCB/ SIP

- **Scenario C:** existing Memory on Logic connected using RDL and TSVs

- **Scenario D:** custom Memory on Logic connected using TSVs

---

**Power:**
- Scenario C → 3x less than Scenario A
- Scenario D → 10x less than Scenario A

**Delay:**
- Scenario C → 5x less than Scenario A
- Scenario D → 25x less than Scenario A
TechTuning
a Technology Co-Optimization Methodology for TSS Chip Level – to – Process Technology
New 3D TSS Stack Design Considerations

- **Traditional Thermal & Mechanical Domains of Concerns**
  - Thermal & Mechanical analyses typically ignored at chip design level
  - Mostly a Package or System level concern
  - At chip level managed through max die size & max power rules

- **3D Technology is a Chip Level System Integration Technology**
  - 3D Technology = Direct & Intimate Interaction Across Several Die
  - Cannot Ignore Thermal and Mechanical Interactions during Stack Design
    - Cannot manage thermal & mechanical issues at package level only
    - Cannot treat a die as a monolithic slab of Si only
    - Cannot manage empirically \(\Leftrightarrow\) Need Ability to Simulate
Thermal Concerns & Requirements

Example Concerns:
• **Hot Spot (T > X°C)** on any Tier due to Stack Power Dissipation
  • vs. physical and layout factors
    – thickness, uBump distribution, underfill & potting properties, etc…
  • vs. chip design factors
    – floorplans, routing, power management schemes, PNG, etc…
  • vs. use and application factors
    – power distribution and densities on each tier vs P and V corners, etc..
• **Thermal Effect on Device Performance & Variability**
  • Use conditions, layout configurations, package specs…

Stack Design Requirements
• Design rules for bump and TSV placement and density
• Thermal application rules and guidelines inc Chip Design
• Thermal hot spot analysis and TSS stack sign off

Infrastructure Needs
• **Model Format + Metrology System & Test Structures**
• **DATA:** thermal material properties + Validation Data
Mechanical Concerns & Requirements

Example Concerns:

- **Effect of Warpage on Stack Manufacturability**
  - vs temperatures, die sizes, process flow, underfill properties, thickness...

- **Die size constraints for tier1/tier2 combinations**
  - relative alignments, temperatures, process flow, underfill properties, thickness

- **Strain Effect on device Performance and Variability**
  - strain booster relaxation vs thinning, TSV & uBump proximity..

Product Requirements

- **Design rules**
- **Mechanical application rules and guidelines inc Chip Design**
- **Mechanical hot spot analysis and TSS stack sign off**

Infrastructure Needs

- **Model Format + Metrology System & Test Structures**
- **DATA : mechanical material properties + Validation data**
3D Stack Design & Process Exploration

- TSS is a System Integration Technology
  - Must Include Thermal & Mechanical Concerns in Stack Design
- Process & Design Must be Co-Optimized for Thermo-Mechanicals
  - Especially due to the novelty in this domain
  - c.f. the challenges with Pb-Free packages and Low K technologies
- But … Complex Process - Design Space Trade Off Exploration

- Process Knobs
  - Materials
    - Epoxies & dielectrics
    - Conductors …
  - Process Structure
    - Plating, depositions
    - Grinding, CMP…
  - Process Integration Flow
    - Form – Fit – Function
    - Thermal Budget
    - Stress Budget …

- Chip Design Knobs
  - Die Sizes & Alignment
    - Absolute & Relative….
  - Chip Design
    - Partitioning & Floorplans
    - Power distribution
  - Bump Design
    - counts…
    - densities
  - Layout Design Rules
    - linewidths, pitches, KOA …

Need Physically Aware Chip Simulation Environment

Riko Radojcic : Qualcomm Inc
Thermo-Mechanical Implementation Road Map

- **Phased Approach**
- **Current**: gross global constraints
- **Phase 1**: Thermal & Mechanical Rules based on Analyses Tools
- **Phase 2**: In Situ Design-for-Thermal & Mechanical

**Past:**
- simple constraints
  - die size / max power
  - some placement limits
- no specific DRC

**Phase 1: TechTuning**
- interface to design: complex rules
- compliance: hot spot checker
- tools focus: ANALYSES
- Input models: Quasi PHYSICAL
- Performance req: slow is OK

**Phase 2: Design for X ....**
- interface to design: in situ tools & flow
- compliance: design simulation
- tools focus: CORRECTION
  - Input models: COMPACT models
  - Performance req: must be fast

**TechTuning**: SPICE-like Thermal and Mechanical Simulator
- Analyses vs. in-situ use during Physical (digital) Design
- An In-Between simulation layer between T-CAD and E-CAD
- Accurate and based on a blend of Physical and Numerical model
- Supported by the Entire Supply Chain

Riko Radojcic: Qualcomm Inc
TechTuning Hierarchy

- **Challenge: Multi-Scale & Multi-Physics**
  - No single integrated solution (today)
  - Use sub-modeling approach w/ Hierarchical Model & Infrastructure
  - Leverage existing infrastructure as much as possible

---

<table>
<thead>
<tr>
<th>mm's</th>
<th>μm's</th>
<th>nm's</th>
</tr>
</thead>
<tbody>
<tr>
<td>Package</td>
<td>Stack</td>
<td>Chip</td>
</tr>
<tr>
<td>Gate</td>
<td>Feature</td>
<td></td>
</tr>
</tbody>
</table>

- **Multi-Physics**
  - Thermal
  - Stress
  - Comments
  - Elect.

- **Challenge: Multi-Scale & Multi-Physics**
  - No single integrated solution (today)
  - Use sub-modeling approach w/ Hierarchical Model & Infrastructure
  - Leverage existing infrastructure as much as possible
Modeling and Simulation Granularity

- Simulation (&modeling) Occurs at Different Abstraction Levels
- Dependent on Granularity of Process and Design Input
  - Material parameters
  - Layout parameters
- Only Some Combinations Make Sense
  - e.g. if process input is monolithic Si slab - there is no point of any design input beyond die size
  - e.g. if the design input is a LEF/DEF there is no point of detailed material properties
- TechTuning Goal is to Get Thermal & Mechanical Simulations at GDS level
- Ergo require “Smear” models and/or “Compact” model
  - Should be easier to manage at both, the source and the use end

Riko Radojcic : Qualcomm Inc
Evaluation Results: Cadence Thermal Solution

- **The Solution**
  - Integrated in the FE Environment
  - Targeted for Design-for-Thermal
    - Power & Performance optimization
  - Extended to Encompass 3D TSS
- **The Procedure**
  - Import package Boundary Conditions
    - Elegant interface to FEA simulator
    - Extract package Thermal Resistance
  - Import Design Data for all Tiers
    - DEF / LEF layout
    - Power models
  - Import Process Data for all Tiers
    - Material properties & geometries
  - Simulate Thermal Map (s)
    - First for each Die in the Stack
    - Then iterate to adjust for the interactions in the Stack
- **Characteristics**
  - Very fast simulator
    - Compatible with Phase 2
  - Integrated Flexibility for ‘what-if’s’
    - Explore design / process options
Evaluation Results: Synopsys FAMMOS

- **Trial Evaluation: SNPS**
  - Fammos TCAD Solution tuned for TSS
  - Seismos DFM solution (TB tuned for TSS)

- **Effect of TSV on Device Mobility**
  - fn. of KOA

- **Effect of Thinning on Surface Stress**
  - fn. of Thickness and Layout

- **Effect of μBump on Stress**
  - fn. of material properties & thermal history

- **Need Si Data for Calibration**
TechTuning RoadMap?

- This is a loooong term project

Need to define a practical and synchronized methodology for Model Formats, Material Characterization and Simulation Solutions

- But a Vital one to Enable Proliferation of 3D Technology
Design Authoring

a Physical Design Flow for TSS Stack Design Spec – to – GDS
**Design Authoring Implementation Road Map**

- **Phased Approach for Design Authoring**
  - **Current**: “2D” Design = multiple layers on single one-sided die
  - **Phase 1**: “2.5D” Design = multiple layers on single two-sided die
  - **Phase 2**: “3D” Design = multiple layers on multiple two-sided die

- **Current 2.5D Implementation**: enabling design with two-sided die
  - Analyses: extraction and physical verification for single 2 sided die
  - Physical Design: automated place and route on TSV and RDL layers
  - Assumes PathFinding has partitioned the system into ‘2D’ tiers
2.5D Design Authoring: Stage -1 Requirements

**Required EDA Tool Updates**

Mostly require recognition of Double Sided Die:
- **Physical Design (P&R)** – may need ability to do timing driven TSV placement & connect
- **Extraction** – need ability to understand TSV and backside RDL all in a single netlist

**Limited need for Two Tier Awareness:**
- **Verification** - need to be able to verify whole stack – including TSV, u-Bump, RDL, up through T2 pad locations

**Ability to deal with TSS specific factors:**
- **Thermo-Mechanical signoff & “TechTuning”**
- **Voltage Dependent RC for STA**
- **Design of Chip Utilities (PNG, Clocks..)**
### Status of 2.5D Production Flow Tools

<table>
<thead>
<tr>
<th>Function</th>
<th>Stage-1 Requirement</th>
<th>2.5D Status &amp; Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synthesis</td>
<td>NO</td>
<td>• Assume Synthesis is all in 2D, with clean Specs for Tier to Tier communication defined in PathFinding</td>
</tr>
<tr>
<td>Floorplan</td>
<td>YES</td>
<td>• Assume that complex TSS floor planning restrictions dictated by Tier-2 are well Defined in PathFinding ✓Need TSV and Backside awareness</td>
</tr>
<tr>
<td>Layout</td>
<td>No</td>
<td>• Layout is all in 2D. No need for any UpDates</td>
</tr>
<tr>
<td>STA</td>
<td>NO</td>
<td>• Assume that the voltage dependence of TSV capacitance can be handled as a corner case</td>
</tr>
<tr>
<td>PDN</td>
<td>No</td>
<td>• Assuming basically 2 different power networks that have been well spec’ed out n PathFinding</td>
</tr>
<tr>
<td>P&amp;R</td>
<td>Maybe</td>
<td>✓Including the Backside routing between TSVs and uBumps, if any,</td>
</tr>
<tr>
<td>Extraction</td>
<td>YES</td>
<td>✓Recognize TSV and BM1 in a single NL</td>
</tr>
<tr>
<td>Verification</td>
<td>YES</td>
<td>✓Physical verification across both tiers</td>
</tr>
</tbody>
</table>

**Good Enough for Functionally Partitioned Stacks**
How Should it all Play Together

Trade Off Sensitivities
Predictive Models
Specifications
Chip Utilities
Design Enablement
Hot Spot Sign-Off

PathFinding
TechTuning

Products

✔ Optimize Specs by Looping in PathFinding
✔ Optimize Technology by Looping in TechTuning
✔ Design Products without Looping
Design Eco-System for 3D

THANK YOU

Riko Radojcic, Qualcomm
5775 Moorehouse Dr.
San Diego, CA, 92121, USA
rikor@qualcomm.com
3D integration perspective for multimedia products

DATE09  3D Integration workshop
Apr.24th 2009  Nice - FRANCE

Home Entertainment & Displays Group
Home Video Division - R&D –
Advanced Technology & Innovation  Dominique Hénoff

Technology R&D
Silicon Technology Development - Advanced CMOS
Technology Platform Sustaining - Tech & FE/BE Qual  Laurent Bonnot
Key factors for 3D integration in consumer

- Multimedia scope
- Solution cost
- Packaging constraints
- Process/design constraints
- IO types vs signal integrity
- Design integration & CAD flow

- How this can push/limit us in next process generation (2010)?
Multimedia product focus

New services
New usages

Environment
Friendly

User Experience
Audio
Graphics
Video
Remote Control

Social Networks
Target Ads
Map, shopping, news
Over-The-Top streaming
Multimedia product details

• Rules
  – Hardware:
    • Development on design platforms set for generic node
    • Production on half node with direct shrink
      – Ultimately no sign-off step for shrunk version
    – Software: cumulative maturity based on two OS (OS21, Linux)

• R&D cycles vs process nodes and SOC sizes:
  – 2003-2008: 90/80nm ~ 0.33Mgates/mm²  SOC: 30-90Mtr
  – 2005-2009: 65/55nm  ~ 0.63Mgates/mm²  SOC: 90-180Mtr
  – 2007-2011: 45/40nm ~ 1.15Mgates/mm²  SOC: 200-250Mtr
  – 2009-2015: 32/28nm ~ 2.50Mgates/mm²  SOC: >500Mtr

• Consumer is production enabler
  – Fast process ramp-up from devt D0 to std production D0
  – « system » : HW and SW maturity at the same time
Solution cost

• Based on H264 decoders with intensive integration
  – New digital features to take the benefit of Mgate/mm² steps (e.g. networking)
  – New known parts to reduce « BOM » (e.g. Satellite tuner) and new functions (e.g. homeplug)
  – Analog/MS PHy’s (e.g. USB) and high speed interfaces (e.g. DDR) with standard dependencies.

• Limitations
  – Number of interconnections
  – Process options/cost/fab cycle vs integration
  – Die vs substrate technologies
  – Power max (whatever nbr of decoders): 5W
  – Robust schedule  f(CAD flow, integration complexity)
Packaging constraints

• Cost of substrate (2010)
  – BE manufacturing costs strongly leveraged
  – FCBGA 35*35 or 27*27 range
  – FCBGA 6L HDI = 2* FCBGA 4L HDI
  – FCBGA 150µm pitch = 1.5 * FCBGA 180µm pitch
  – Large part of consumer product cost

• Interconnections nbr impact
  – FCBGA 180µm pitch
  – 4L HDI substrate
  – 1028 interconnects/620 signals
  – Interposer area ~ 45mm²

• RLC values

• Silicon substrate co-design

• Supply chain: locations of Si fab, EWS and assembly
Process/design constraints

- Design robustness and reliable libs
- Low power design solutions
- No size reduction of main analog and IO areas below 65nm
- Mono GO2: 30 Ang (1V5-1V8) or 50Ang (2V5-3V3)
- Process type: CMOS bulk LP vs GP

<table>
<thead>
<tr>
<th></th>
<th>65LP</th>
<th>45LP</th>
<th>40LP</th>
<th>40G</th>
<th>32LP</th>
<th>32G</th>
</tr>
</thead>
<tbody>
<tr>
<td>VDD Volts</td>
<td>1.2</td>
<td>1.10</td>
<td>1.10</td>
<td>0.90</td>
<td>1.00</td>
<td>0.85</td>
</tr>
<tr>
<td>Gate type</td>
<td>Poly SiON</td>
<td>Poly SiON</td>
<td>Poly SiON</td>
<td>Poly SiON</td>
<td>HKMG</td>
<td>HKMG</td>
</tr>
<tr>
<td>Speed Tpd ps</td>
<td>Ref.</td>
<td>~ / 1.2</td>
<td>≈ Ref</td>
<td>&gt; / 2</td>
<td>&gt; / 1.5</td>
<td>~ / 3</td>
</tr>
<tr>
<td>Pdyn nW/MHz</td>
<td>Ref.</td>
<td>~ / 1.5</td>
<td>~ / 1.5</td>
<td>~ / 2</td>
<td>~ / 3</td>
<td>~ / 3.5</td>
</tr>
<tr>
<td>Pleak nW/gate @ 25°C</td>
<td>Ref. x10</td>
<td>~ x1.15</td>
<td>~ x1.3</td>
<td>&gt; x35</td>
<td>≈ Ref x9</td>
<td>&gt; x20 x10</td>
</tr>
<tr>
<td>Pleak nW/gate @ 125°C</td>
<td>x10</td>
<td>x10</td>
<td>x10</td>
<td>x10</td>
<td>x10</td>
<td>x10</td>
</tr>
</tbody>
</table>
IO types vs signal integrity

- **Standard IO**
  - Compatible with any GO1 and GO2, Fmax 200MHz

- **High speed IOs**
  - High speed memory interfaces (e.g. DDR)
    - Increasing MHz/Mbps while decreasing voltage range
    - Break point on DDR3/800MHz/1.6Gbps at 1V5
  - AC coupled serial links (e.g. SATA)
    - Differential links, impedance adaptation, GO2/voltage adapted
  - DC coupled serial links (e.g. HDMI)
    - Standard voltages blocking voltage decrease
    - If possible redesign (cascode) this means long design cycle and test chips validation.

- **Specific analog parts (e.g. DAC’s)**
  - Requiring redesign on die or on board.

- **Signal integrity**
  - **Wire bonding**: R and C OK, L impacts DDR integrity
  - **Flip chip**: R and L OK, possible C limitation
  - **Cu pillar**: RLC OK
  - **TSV**: R and L OK, possible C limitation
  - **POP**: 2 or 3 times flip-chip constraints, but better than board routing. Power limitation from bottom part.
Design integration & CAD flow

- **3D co-design**
  - Handling 2 or 3 different DB’s
  - Handling at least 2 different tool types (Silicon and substrate/board routers)
  - Allow iteration on IO/power grid change interactively with substrate (cost, power, thermal, signal integrity).

- **Power and Thermal analysis**
  - Link between tool results and silicon results
  - Accurate budgeting

- **Signal integrity**
  - Accurate models

- **Embedding sensors (performance, thermal, Vreg)**
32/28 nm case study (1/6): choice vs constraints

- **Pitch vs substrate cost**
  - Flip chip required for DDR3, 180µm pitch by 2010
  - Micro Cu pillar for stacked die
- **Stacking / multi chip module**
  - Stacking otherwise issue with power, performance and signal integrity
  - Face 2 Face approach allowing slug or heat-sink on top die
- **Process and schedule vs die split**
  - « Top »: Digital chip sized for DDR3 high speed bandwidth priority. Minimum of area penalty by moving out analog IP’s and minimizing interconnections.
  - Limited nbr of additional interconnections for Top/Bottom die I/F.
  - « Bottom »: Analog chip required to embed 50Ang features (HDMI, DAC, USB2, …) => « Active Interposer »
  - IP reuse enabling schedule optimization (split devt, validated analog IP’s).
- **3D integration solution**
  - Top/Bottom: Die2Wafer, Bump reference pitch ~ 1000 interconnects
  - Bottom/Substrate: TSV x2 bump pitch with RDL at x3 bump pitch
  - Substrate: FCBGA 4L HDI, pin pitch >x10 bump pitch
32nm case study (2/6): possible solutions

- Stacked+active interposer with TSV
  - RLC OK for DDR3/800MHz
  - Top/bottom die 32nm/65nm split with DDR3
  - Bottom size depends on nbr of IO (align with 4L substrate)
- Test chip to validate assumptions.

- Flip-chip+interposer in wire bonding
  - A) Stacked
    - RLC vs features
    - Active interposer
  - B) « companion »
    - « MCM » routability
    - RLC « 2xFC »
    - Passive interposer?

- DDR3/800MHz performance?
- DDR3 bare die vs shrinkability

- FC+ interposer in WB + package
  - Feasibility of package report on Si interposer?
    - Routing density on interposer?
    - DDR3 size to embed?
    - No roadmap

- POP structure
  - Top package with DDR
  - At least 3 « bumping » stack, C?
32nm case study (3/6): product driver

Advanced CMOS process
1V & 1.8V

Digital functions; DDR3 800MHz
PHY for DDR3

>1000 micro-interconnections,
bump reference pitch

Landing pads

1000 TSV
(TSV pitch = 2x bump pitch)

Wafer thickness
= TSV pitch

~1000 bumps. RDL pitch x3 bump pitch.
Target 4L HDI FCBGA

CATRENE project 3DIM3
ST contribution

2.5V 3.3V (50Ang) and 1.xV
PHYs (for HDMI, usb, sata, PCIe),
DACs, …
32nm case study (4/6): technology details

- Top chip
- Bottom chip
- Backside RDL & UBM
- TSV
- Bumps

PBGA 35*35
4L HDI
1-2-1
32nm case study (5/6): stack and EDA

- Stacked pitch complexity
  - Pitch ratios to handle on same flow: x1, x2, x3, >x10
- Rerouting assignment feasible but in 2D way
- Signal integrity in 2D way
- Power plates and power pins adjustment
- Signal integrity, power dissipation, thermal impacts
- DFM not defined for stack (F2F hot spots, ESD, …)
- Abstraction/models versus implementation thru a real 3D flow

Bottom die border
Top die border

Yellow: pillar bumps
Red: TSV’s
Blue: substrate balls
Green: signal routing thru stack
32nm case study (6/6): product ASP vs mono and 3D stack die

- **Analysis**
  - Monochip « P » area 32nm = 2.0 x area 32nm « 3D stack »
  - Interposer « P-2 » area 65nm = 1.5 x area 32nm « 3D stack »
  - ASP erosion 10%/yr

- Result: ~45% of cost on substrate, ~15% on assembly, ~40% on Silicon
- **Short term 3D integration interest**: cost similar to single die, but better time to market (IP reuse).

---

**65/32nm ASP hypothesis**

![Graph showing ASP ratios over time]

- Monochip requires longer IP validation
- 15% decrease in ASP ratio
Conclusion

• Enablers:
  – Cost effective technologies and supply chain, EDA solutions.

• Existing
  – Moving towards heterogeneous 3D integration.
  – First interconnect technology bricks
  – Substrate design/cost challenged

• To be successful:
  – Supply chain biz model
  – TSV roadmap and RDL rerouting
  – Real 3D EDA flow
    • Data Base management with modeling
    • Accuracy thru flow, top-down and bottom-up
    • Associated solution cost model
  – EDA tools and models fitting with supply chain
Thank you for your attention

Great thanks to all ST colleagues from TRD, ATM and HED for their support
Z-axis Interconnections: Fabrication and Electrical Performance

Voya R. Markovich, Rabindra N. Das, Michael Rowlands and John Lauffer
Endicott Interconnect Technologies, Inc., 1093 Clark Street, Endicott, New York, 13760.
Telephone No: (607) 755-1978, E-mail: voya.markovich@eitny.com

The demand for high-performance, lightweight, portable computing power is driving the industry toward miniaturization at a rate not seen before. Electronic packaging is evolving to meet the demands of higher functionality in ever smaller packages. To accomplish this, new packaging needs to be able to integrate more dies with greater function, higher I/O counts, smaller pitches, and greater heat densities, while being pushed into smaller and smaller footprints. One strategy allows for metal-to-metal z-axis electrical interconnection (Figure 1) of subcomposites during lamination to form a composite structure. Conductive joints are formed during lamination using an electrically conductive adhesive (ECA). As a result, one is able to fabricate structures with vertically-terminated vias of arbitrary depth. Replacement of conventional plated through holes (PTHs) with vertically-terminated vias opens up additional wiring channels on layers above and below these vertical interconnections, and eliminates via stubs which cause reflective signal loss. Vertically terminated vias facilitate a more space-efficient package redesign. Conductive adhesives have been used for Z-interconnect applications.

Figure 1: Photograph of z-interconnect laminate chip carrier having four signal wiring planes with a stripline transmission line structure shown in cross section.

Typically, adhesives formulated using controlled-sized micro particles have been used to fill small diameter holes for Z-interconnect applications. In the present study, micro-filled adhesives were modified to improve overall electrical, mechanical, and reliability performance. Nanoparticles were mixed with micro-particles to improve sintering behavior of the adhesives. Addition of a conducting polymer will increase overall mechanical strength without compromising electrical conductivity. A variety of metals including Ag and Cu have been used to make the conductive adhesives. The adhesive was applied onto epoxy, or LCP, or PTFE-based substrates by printing or coating. The content of metal in the adhesives ranged from 60% to 90% by weight. The effects of the polymer binder, particle size and loading parameters on electrical performance are presented. Adhesives were characterized by SEM and optical microscopy to ascertain particle dispersion and interconnection mechanisms. A Keithley micro-ohmmeter was used for electrical characterization. Adhesives exhibited volume resistivity ranging from $10^4$ ohm-cm to $10^6$ ohm-cm depending on composition, particle size, and loading of the adhesives. It was found that with increasing curing temperature, the volume resistivity decreased due to sintering of metal particles (Figure 2). Adhesives modified with nanoparticles or conducting polymer showed 30-50% resistance drop when cured at 275°C instead of 200°C. The electrical properties of adhesives fabricated from metal-epoxy composites showed a stable resistance for pot life ranging from 72 to 168 hrs.

Figure 2: SEM micrographs for the nano-micro filled silver based conducting adhesives; (A) un-sintered at 200 °C, and (B) sintered at (275 ±10) °C.

This work also deals with adhesion issues between adhesives and the substrates to which they are mated. Adhesives formulated with a conducting polymer exhibited tensile strength with Gould’s JTC-treated Cu ≥ 3800 PSI, and as low as 1800 PSI for a conducting polymer- LMP based system. The paper presents a reliability assessment of adhesive joints conducted by testing samples exposed to aging, pressure cooker tests (PCT), IR-reflow, and solder shock. The effect of IR reflow (simulated assembly conditions) on the bond strength of adhesive joints was measured using 90 degree peel test, as well as by microstructural examination. It was found that aging generally caused a decrease in peel strength.

A few optimized metal-epoxy adhesives were used for hole fill applications to fabricate Z-axis interconnections in laminates. Conductive joints were formed during composite lamination using an electrically conductive adhesive. Z-axis interconnection was achieved using joining cores. Through holes in the joining cores, formed by laser or mechanical drilling and having diameters ranging from 50 µm to 250 µm, were filled with an optimized electrically conductive adhesive. The adhesive-filled joining cores (Figure 3) were laminated with circuitized subcomposites to produce a composite structure. High temperature/pressure lamination was used to cure the adhesive in the composite and provide Z-interconnection among the circuitized subcomposites.
As a case study, we have designed and built a circuit board test vehicle (TV) to make new RF structures, using Z-axis interconnection (Z-interconnect) building blocks. A typical 50-ohm stripline was designed with a ground-signal-ground structure. The stack-up had 23 metal layers, including 5 0S1P joining cores and 6 2S1P signals cores. Teflon-based low loss materials were used for the dielectric layers. Laminated conducting joints show low resistance in the range of 5-12 milliohm per joint. Conductive paste showed good signal transmission to 25GHz. Z-interconnect construction allows wide (7.4 mils) and narrow (3.6mils) lines in the same stack-up. Wide lines (7.4 mils) in organic full Z-interconnect board have signal loss less than 0.6 dB/inch at 10GHz, and less than 1.0 dB/inch at 16GHz. A full-Z taconic board can carry 25Gbps at least 12". Very low-loss multi-GHz, controlled-impedance transmission lines can be built using Z-interconnect with organic dielectric materials. Electrically, S-parameter measurements showed (Figure 4) very low loss at multi-gigahertz frequencies. The measured insertion loss for narrow, short lines and wide, long lines are similar. Figure 5 shows eye diagrams for Z and non-Z interconnect joints. We considered eye diagram simulation using measured S-parameters from 6 inch length net for four Z-interconnect joints and compared with no Z-interconnect joints. It is clear from eye diagram simulation that Z-interconnect does not degrade 10Gbps data.

Z-interconnect can be used in single and multi-chip applications. By designing an organic package without electrical stubs and without through holes, high wiring density and excellent electrical performance can be achieved. Novel means of providing vertical electrical interconnection in organic substrates can help semiconductor packaging keep pace with the needs of the semiconductor marketplace. This effort is an integrated approach on three fronts: materials development and characterization, fabrication, and design and electrical characterization at the board level.

Figure 3: SEM micrographs of adhesive-filled joining cores

Figure 4: Insertion loss as a function of frequencies (A) narrow line (80um), (B) wide Stripline (180um), and (C) medium Stripline (130um).

Figure 5: Eye Diagram Simulation using measured S-parameters, 6" net length (A) non Z-interconnect joints and (B) four Z-interconnect joints.
Impact of Design Choices on 3D SiC Manufacturing Cost

Dimitrios Velenis, Michele Stucchi, Erik Jan Marinissen and Eric Beyne
Introduction

- Choice of Going Vertical
- Choice of 3D Technology
- Choice of TSV Repair and Redundancy
- Choice of 3D Testing
- Conclusions
3D – SIC (Stacked IC) Process

- TSV fabrication steps (via first)
  - Deep SI etching
  - Via Oxide deposition
  - Cu seed deposition
  - Cu plating
  - CMP

- Backside thinning and bonding
  - Carrier wafer
    - Temporary carrier bonding
  - Carrier wafer
    - Backside thinning
  - Carrier wafer
    - Expose Cu nails
  - Carrier wafer
    - Permanent bonding
  - Carrier wafer
    - Carrier de-bonding
IMEC 3D Cost Model

- Real 3D integration process
  - Extrapolated from IMEC data/tools to production environment

- Cost of 3D integration processing
  - Cost of Equipment ( depreciated)
  - Cost of Facilities
  - Personnel Cost
  - Material Cost

- Qualitative answers to cost questions
  - Explore cost and benefits of 3D integration

- General assumptions
  - Two tier homogeneous 3D systems
    - CMOS logic on both top and bottom die
    - Even higher advantages for 3D heterogeneous systems
Outline

• Introduction

Choice of Going Vertical

• Choice of 3D Technology

• Choice of TSV Repair and Redundancy

• Choice of 3D Testing

• Conclusions
To 3D or Not 3D?

- System on a single die / two stacked chips
  - At which system size is it cost effective to go vertical?
  - Demonstration of cost analysis
  - Can be applied prior to integration decision

- System integration options
  - Single Die (2D)
  - Two stacked chips – 3D (Die-to-Wafer Stacking)
  - Two stacked chips – 3D (Wafer-to-Wafer Stacking)

- Costs and values considered
  - Cost of CMOS wafer processing - $2K per wafer
  - Cost of 3D processing – tools, clean room, personnel, materials
  - CMOS process yield: 90% per cm²
  - Stacking process yield: 95%
Effect of System Size on Die Area

- Wafer size: 200mm (8 inch), 65nm process technology
- Considering same top and bottom die size for D2W and W2W stacking
- Die size increase not linear with gate count
3D integration: cost effective for large systems
D2W: stacking on Known Good Die (KGD) only
D2W yield is better than W2W yield
Outline

• Introduction

• Choice of Going Vertical

▷ Choice of 3D Technology

• Choice of TSV Repair and Redundancy

• Choice of 3D Testing

• Conclusions
SIC – WLP TSV comparison

- Stacked IC processing (via first approach)
  - Via diameter: 5µm
  - Via pitch: 10µm – determines TSV area on-chip
  - Via depth: variable (25-50µm)
  - Higher TSV density

- Wafer Level Processing (via-last approach)
  - Via diameter: 35µm
  - Via pitch: 60µm – determines TSV area on-chip
  - Via depth: variable (40-150µm)
  - Lower process cost

- Considering number of TSVs per stack
Effect of Number of TSVs on Die Area

- Increase of die size
  - Due to increase in number of TSVs only
  - Circuit size remains the same
  - Number of drawn chips/wafer shown – No yield loss considered
Cost per 3D stack (Considering yield loss)

- CMOS Process Yield: 90% per cm²
- Stacking Yield: 95%
- 50M gates per 3D Stack
Outline

- Introduction
- Choice of Going Vertical
- Choice of 3D Technology

Choice of TSV Repair and Redundancy

- Choice of 3D Testing
- Conclusions
TSV redundancy and repair

- Faulty TSV
  - Non-conducting TSV
  - Leaking TSV
  - High Capacitance TSV (high signal delay)

- TSV repairing
  - Detect a faulty TSV
  - Re-direct signal routing to non-faulty TSVs

- TSV repair schemes
  - 2:1 repair – 1 repair TSV for every 2 functional*
  - 4:2 repair – 2 repair TSVs for every 4 functional*
  - TSV Doubling: 1 redundant TSV for every 1 functional
  - TSV Tripling: 2 redundant TSVs for every 1 functional

*U. Kang et al. 8GB 3D DRAM Using Through-Silicon-Via Technology, ISSCC 2009
Yield Comparison Repair vs. Redundancy

Assumptions:
- 1000 functional TSVs
- Non-correlated TSV defects
Introduction

Choice of Going Vertical

Choice of 3D Technology

Choice of TSV Repair and Redundancy

Choice of 3D Testing

Conclusions
3D Testing Cost

- 3D Testing
  - Testing at individual die level
    - Determine **Known Good Die** (KGD) for stacking
  - Testing at 3D stack level
    - Final stack test

- Individual die test options
  - No Test
  - Built-In-Self-Test (BIST)
  - Reduced-Pad-Count-Test (RPCT)

- 3D Testing assumptions
  - Final, Post-stacking Test performed in ALL cases
  - Test Bottom die
    - Avoid stacking cost if die is faulty
  - Scan chains in place on all stacked dice
    - To be utilized in the Final, Post-stacking Test
3D Testing Cost Components

- Test time and equipment overhead
  - Individual test bottom die
  - Individual test top die
  - Final stack test

- Die area overhead
  - Flip-flop Scan chains for bottom die
  - Flip-flop Scan chains for top die
  - I/O pads on top die
    - Assuming No TSV probing for test
    - System I/O pads on bottom die accessible for testing
      - No additional I/O pads are required on bottom die

Determine KGD
Test Options for Top Die

- No Test
  - Area Overhead: Scan Chains for Final Test
- BIST
  - Internally generated, large number of test patterns
  - Additional logic circuit area on-chip
  - Limited number of I/O pads (minimum five)
  - Moderate Fault Coverage (~90% for logic circuits)
- RPCT
  - Externally applied, small number of test patterns
  - Additional I/O Pads required to reduce testing time
  - No additional logic on-chip
  - Good Fault Coverage (~99% for logic circuits)
**Total 3D Stack Cost – yield based**

- **Bottom die**: RPCT testing using system I/O pads
- **Top die test assumptions**:
  - Scan chain overhead: 5% of circuit area
  - BIST area overhead: Additional 5% of circuit area, 5 I/O pads
  - RPCT area overhead: 30 I/O pads
  - Test time: No. of test patterns / No. of scan chains
Total 3D Stack Cost – yield based

- Break even points:
  - 94% die yield per cm² for RPCT
  - 90.5% die yield per cm² for LBIST
Outline

• Introduction

• Choice of Going Vertical

• Choice of 3D Technology

• Choice of TSV Repair and Redundancy

• Choice of 3D Testing

Conclusions
Conclusions

- IMEC 3D Cost Model
  - Estimate 3D integration manufacturing cost
  - Evaluate impact of design choices
- 3D integration choice
  - Effective for high gate counts
- 3D Technology choice
  - Reduced die size pays off for higher processing cost
- TSV repair and redundancy choice
  - TSV fault and detection scheme required
- 3D Testing choice
  - Testing helps to keep system cost low
  - Cost model used to decide based on processing yield

- Keep die size low
  - Higher number of stacks/wafer significantly reduces system cost
Acknowledgments

Clock and Power Distribution Networks for 3-D Integrated Circuits

Ioannis Savidis and Eby G. Friedman
Department of Electrical and Computer Engineering
University of Rochester,
Rochester, New York 14627, USA
{iosavid, friedman}@ece.rochester.edu

Vasilis F. Pavlidis and Giovanni De Micheli
LSI-EPFL
1015-Lausanne, Switzerland
{vasileios.pavlidis, giovanni.demicheli}@epfl.ch

Abstract – Global interconnect design for three-dimensional integrated circuits is a crucial task. Despite the importance of this task, limited results related to global issues have been presented. Challenges in reliably distributing power, ground, and the clock signal within a multi-plane integrated system are discussed in this paper. The design of two 3-D test circuits addressing these issues is described. Candidate 3-D topologies for both power and clock distribution networks are also presented. Design implications due to the different design approaches are discussed. Experimental and simulation results of the 3-D clock and power distribution architectures, respectively, are provided. Both of the test circuits are fabricated by the 3-D fabrication process developed at MIT Lincoln Laboratories (MITLL). The design of the clock and power distribution networks is discussed.

I. CLOCK DISTRIBUTION NETWORKS FOR 3-D ICS

Distributing the clock signal in 3-D ICs is a complex and challenging task as sequential elements synchronized by the same clock signal can be located on multiple planes [1]. In 2-D circuits, symmetric interconnect structures, such as H- and X-trees, are widely utilized to distribute the clock signal across a circuit [2]. An extension of an H-tree to three dimensions does not guarantee equidistant interconnect paths from the root to the leaves of the tree. The impedance of the TSVs can increase the time for the clock signal to arrive at the leaves of the tree on specific planes as compared to the time for the clock signal to arrive at the leaves of the tree located on the same plane as the clock driver [3]. Consequently, asymmetric topologies can be potential candidates for distributing the clock signal in a 3-D integrated system.

Another fundamental issue in the design of clock distribution networks is low power consumption, since the clock network dissipates a significant portion of the total power consumed by a synchronous circuit [2]. This demand is stricter for 3-D ICs due to the increased power density and related thermal concerns.

A test circuit investigating the distribution of the clock signal has been recently designed. The test circuit is fabricated by the MITLL process which is a fully depleted silicon-on-insulator process (FDSOI). The minimum feature size of the devices in this test circuit is 180 nm, with one polysilicon layer and three metal layers interconnecting the devices on each wafer. The dimensions of the TSVs in this test circuit are 1.75 μm × 1.75 μm [5]. Each block contains about 30,000 transistors and includes the same logic circuit but implements a different clock distribution architecture.

The function of the logic circuit common to the three blocks is to emulate different switching patterns and load conditions for the clock distribution networks under investigation. Random switching patterns are generated in each block by combining pseudorandom number generators and several groups of four-bit counters that switch current loads over different time intervals.

Each of the three blocks includes a different clock distribution structure, which are schematically illustrated in Fig. 1. The dashed lines depict vertical interconnects implemented by TSVs. These architectures combine different topologies, such as H-trees, rings, and meshes [2]. The fabricated 3-D test circuit is illustrated in Fig. 2. The highest operational frequency is 1.4 GHz. Clock skew measurements indicate that the topology in Fig. 1a produces, on average, the lowest skew as compared to the other two topologies. The clock skew among the planes is greater for the local mesh topology (see Fig. 1b) as compared to the H-tree topology, primarily due to the unbalanced clock load for certain local meshes. Alternatively, the clock distribution network that includes the global rings exhibits the highest clock skew among all topologies due to the greater difference in distance that the clock signal traverses on each plane [7].

II. POWER DISTRIBUTION NETWORKS FOR 3-D ICS

Another important global issue is the design of robust power distribution architectures for 3-D ICs [8]. In planar ICs where flip-chip packaging is adopted as the
packaging technique, an array of power and ground pads are allocated throughout the surface of the die to provide the necessary current for the circuit to operate properly. Increasing current densities and faster current transients, however, further complicate the design of power distribution networks. Three-dimensional integration alleviates the restraints on using interconnect metal layers for power distribution networks through topologies that are not available in two-dimensional circuits. With 3-D technologies, individual planes can potentially be dedicated to power delivery. Three different 3-D power distribution networks are proposed for the first time. A comparison of the noise on both the power and ground networks for all three topologies is provided. Several tradeoffs among different TSV densities are also explored.

The power networks are included on a test vehicle, which will also be fabricated by MITLL. The target technology is a fully depleted silicon-on-insulator (FDSOI) 150 nm process, where the dimensions of the TSVs have been scaled to 1.5 μm x 1.5 μm [6]. The topologies explored on this test circuit are schematically illustrated in Fig. 3.

Interdigitated power/ground lines are used in all of these topologies where a grid is desirable. There are two primary objectives for the test circuit: (i) to explore the benefits of a dedicated power/ground plane, and (ii) to investigate the effect that a higher TSV density has on the noise characteristics of the power network. The upper left and lower topologies shown in Fig. 3 are used to address objective (i), while the upper two topologies address objective (ii). The difference between the upper left and upper right topologies is the number of TSVs, where the latter topology includes 50% more TSVs. The TSVs are only located on the periphery of the upper left power distribution network, whereas the upper right topology includes additional TSVs on the power lines crossing the middle of the circuit block.

Each test block includes identical circuitry. The basic logic blocks are shown in Fig. 4. Power supply noise generators draw varying current from the power lines. These noise generators are placed on each plane of each power network topology. Voltage sense circuitry is included on all of the planes within each test block to measure the noise on both the power and ground lines. The only plane that does not include any circuitry is the plane for power and ground (see Fig. 3c). An average of the output voltage from the sense circuitry on each plane within each test block is used to compare the topologies depicted in Fig. 3.

**Fig. 2** Fabricated 3-D circuit investigating different clock distribution schemes.

The power networks are included on a test vehicle, which will also be fabricated by MITLL. The target technology is a fully depleted silicon-on-insulator (FDSOI) 150 nm process, where the dimensions of the TSVs have been scaled to 1.5 μm x 1.5 μm [6]. The topologies explored on this test circuit are schematically illustrated in Fig. 3.

Interdigitated power/ground lines are used in all of these topologies where a grid is desirable. There are two primary objectives for the test circuit: (i) to explore the benefits of a dedicated power/ground plane, and (ii) to investigate the effect that a higher TSV density has on the noise characteristics of the power network. The upper left and lower topologies shown in Fig. 3 are used to address objective (i), while the upper two topologies address objective (ii). The difference between the upper left and upper right topologies is the number of TSVs, where the latter topology includes 50% more TSVs. The TSVs are only located on the periphery of the upper left power distribution network, whereas the upper right topology includes additional TSVs on the power lines crossing the middle of the circuit block.

Each test block includes identical circuitry. The basic logic blocks are shown in Fig. 4. Power supply noise generators draw varying current from the power lines. These noise generators are placed on each plane of each power network topology. Voltage sense circuitry is included on all of the planes within each test block to measure the noise on both the power and ground lines. The only plane that does not include any circuitry is the plane for power and ground (see Fig. 3c). An average of the output voltage from the sense circuitry on each plane within each test block is used to compare the topologies depicted in Fig. 3.

**Fig. 3** Three 3-D power distribution networks within the test circuit, (a) TSVs only on the periphery, (b) 50% higher TSV density as compared to (a), and (c) dedicated plane for power and ground distribution.

**Fig. 4** Power supply noise generators and voltage sense circuitry included in the circuit blocks depicted in Fig. 3.

**REFERENCES**


Hierarchical Cache System for 3D-Multi-Core Processors in sub 90nm CMOS

Kumiko Nomura*, Keiko Abe*, Shinobu Fujita*, Yasuhiro Kurosawa**, Atsushi Kageshima***
Advanced LSI Technology Laboratory, Corporate R&D Center, Toshiba Corporation*,
Center of Semiconductor Research and Development, Semiconductor Company, Toshiba Corporation**
System LSI Division, Semiconductor Company, Toshiba Corporation***
E-mail: {kumiko.nomura*, keiko2.abe*, shinobu.fujita*, yasuhiro.kurosawa**, atsushi.kageshima**}@toshiba.co.jp

Abstract

Three-dimensional (3D) integrated circuits have the potential to reduce interconnect length and improve digital system performance for multi-core processor systems. We introduce a novel hierarchical cache system containing 3D bus with crossbar and common bus for multi-core processor having 16 cores in a sub-90nm CMOS generation. Performance of the present 3D system is typically better than that of 2D system. Furthermore, performance of the 3D system in "65 nm" and "45 nm" CMOS technology is superior to that of the 2D system in "45 nm" and "32 nm" CMOS technology, respectively. It indicates that the present 3D circuit design for the hierarchical cache can replace CMOS scaling in a sub-65nm CMOS generation.

1. Introduction

Instead of CMOS scaling, use of multi-core processor is another trend to increase performance for various on-chip applications. As the dimension of CMOS is decreased, the number of PEs (processor elements) implemented on the chip is increased, as a result, the processor performance is improved [1]. Additionally, increase in cache capacity is effective to improve the performance. An efficient architecture for this is a shared cache having hierarchical topology [1]. Because of these trends of the PE and cache, the interconnect length used between PEs and cache is increased. Furthermore, the interconnect delay increases with CMOS scaling for the sub-90nm CMOS technology. Thus, the interconnect delay is becoming larger than logic gate delay. One of the most effective way to reduce interconnect delay is to use three-dimensional (3D) circuits. Considering these situations, 3D multi-core processor with hierarchical cache is one of the most suitable applications of 3D circuit using through-silicon-via (TSV).

Recently TSV-based 3D circuits have been more likely [2] and 3D circuits and their design have been studied intensively [3]. Previous works covered 3D multi-core processor design focusing on 3D circuits themselves, such as stacked DRAM with 3D multi-layer design [4] and cache (SRAM) with 3D multi-layer design [5]. However, there was little study to study the 3D hierarchical cache topology for multi-core processors from the viewpoint of 3D interconnect design.

This paper describes design and performance analysis of the 3D hierarchical cache based on the analysis of interconnect delay in sub-90nm CMOS.

2. Interconnect delay and components

Based on an interconnect delay theory, π model [5], the interconnect delays as a function of interconnect length were calculated using data in ITRS 2006[6]. Considering data on R&D of TSV technology, the diameter of TSV fabricated in near future industries is 5 to 10um with a length of 20 to 40um. Previous work has shown that the delay of TSV-based vertical interconnects is small enough to be neglected compared to that of conventional interconnect in CMOS according to the results of precise calculation[7], since the diameter of TSV is much larger and its aspect ratio is much smaller than that of the CMOS interconnects.

To evaluate the performance of 3D shared cache, we used an in-house cache/bus simulator. In this simulation, we evaluated the processing time for bus access from the time when PE requests data in memory to the time when the data reach the PE. To calculate interconnect delay, dimensions of PE, cache and bus were determined. For the PE dimension, we referred the design of a quad processor fabricated in a 65nm CMOS technology [8] considering CMOS technology migration from 65nm to 45nm or 32nm CMOS. For the cache, we used CACTI 5.2 [9] to determine the area and the access time for each CMOS technology. The CACTI uses “perfectly scaled” CMOS technology even in 32nm process, which is overestimated and unrealistic. Thus we assume that the area and the access time latency of the cache in 32nm process are the same as in 45nm process. This is also because, in the case of large cache memory, CMOS dimension of SRAM in 32nm process is generally almost the same as that in a 45nm process owing to the issues of process variation and power consumption. Further adjustment of the CACTI for the simulation in sub-65nm CMOS will be discussed later.

3. 16-core processor with crossbar circuits

Simulation for 16-core processor is performed to quantify the impact of 3D circuits in 90nm, 65nm, 45nm, and 32nm CMOS technology. Fig. 1 (1) shows the hierarchical topology of 16-core microprocessor with an original bus topology effective for 3D-hieracal cache. This processor has L1-caches in PEs and L2 L2-caches, two L3-caches and one L4-cache outside PE. L2- and L3-cache, and L3- and L4-caches are connected by a common bus. It has also the four 4x4 crossbar bus circuits connecting four PEs to four L2-caches since the bus connection for communication between PEs and L2 cache is much more frequently used compared to that between L2-chaches and other layer caches. Our preliminary simulation has indicated suitability of use of the cross-bar bus for PEs and L2-caches. Based on the original topology in Fig. 1(1), we designed 2D and 3D circuits (See Fig. (2) and (3)). In the 3D circuits, circuit blocks such as PEs and cache themselves are 2D circuits with one tier of silicon substrate, and these circuit blocks are connected three dimensionally using TSV-based bus interconnects, as shown in Fig.1 (3). Since increase in the number of 3D-circuit layers worsens heat conductance in the vertical direction, and temperature is drastically increased, we use 6 tiers for 3D-circuits. Based on the analysis of layout of 3D-circuits, the area overhead of TSV is less than 3%, where the diameter of TSV is less than 10um, and is small enough to be acceptable.

3.1. Simulation Parameter

Let C L2, C L3, and C L4 be the cache capacity of L2, L3, and L4 caches, respectively. In each CMOS technology, three cases for cache capacity (C L2, C L3, C L4) are selected as follows: (I) (C L2, C L3, C L4)=(256KB, 8MB, 24MB), (II) (C L2, C L3, C L4)=(512KB, 20MB, 60MB), (III) (C L2, C L3, C L4)=(768KB, 48MB, 144MB).

As the number of PEs of this microprocessor is 4 times that for the quad-core [8], total capacity of L2 and L3 caches in case (III) is selected as simply 4 times of that used in[8], respectively. The cache capacities in cases of (I) and (II) are selected by linearly decreasing those in case (III).

For each case, cache miss rates P L2, P L3, and P L4 are selected as follows: (A) (P L2, P L3, P L4)=(1%,2%,0%), (B) (P L2, P L3, P L4)=(10%,20%,0%), (C) (P L2, P L3, P L4)=(20%,50%,0%).

To avoid long-time communication between L4 cache and a main memory, we assume that all read requests can be completed in the L4 cache (cache miss rate is 0%). This assumption does not affect performance estimation largely.
Placement of PE and cache is fixed so that each interconnect path can be the shortest. Each interconnect delay is calculated using the results in section 2. Note that the interconnect delay of TSV can be ignored for the reason stated in section 2.

CMOS miniaturization cost is increasing drastically.

4. Summary
We introduce a 16-core processor using a novel 3D hierarchical cache topology with hybrid crossbar and common bus. This 3D system is designed based on thorough analysis of interconnect delay in sub-90nm CMOS generation and 3D-bus topology. We evaluate dependence of the performance on cache capacity, cache miss rate and CMOS technology, considering the actual CMOS situation. From these results, 3D system can improve performance of the 2D system by over 20%. This improvement ratio is close to that obtained by CMOS scaling for one CMOS generation. Furthermore, by adjustment of actual CMOS scaling, performance of the 3D system in “65 nm” and “45 nm” CMOS technology is superior to that of the 2D system in “45 nm” and “32 nm” CMOS technology, respectively. It indicates that in the present 3D system, the design change from 2D- to 3D-circuit design is superior to the miniaturization based on CMOS scaling in a sub-65nm CMOS generation.

Fig. 1 16PE crossbar circuit (1) Topology of 16PE circuit, (2)2D circuit layout of (1), (3) 3D circuit layout of (1)

3.2. Simulation Results
Figures 2(1)-4(4) show the simulation results. The average of the processing time for 200000 cache access operations is shown. By comparing the processing time of 2D and 3D systems, an improvement ratio of 3D system to 2D system for each CMOS technology is also plotted in these figures.

In each CMOS technology, the systems having the higher cache miss rate have longer processing times. Because this circuit uses a common bus to connect between upper level caches such as L2 and L3 cache, the bus is occupied by some bus masters very frequently. In the case of high cache miss rate like case (C), the processing time becomes longer with increasing cache capacity. This is due to the increase in both cache access time and bus waiting time. Thus, cache capacity and miss rate greatly affect average processing time for both 2D system and 3D system.

For a comparison between 2D and 3D, the improvement ratio of 3D to 2D becomes larger with increasing the cache capacity, since cache with large capacity has large dimension and interconnect path from the cache increases. In these cases, the 3D circuit has a big advantage in terms of improvement of processor performance. The improvement ratio of 3D to 2D also increases with increasing CMOS scale. This trend clearly indicates that interconnect delay affects the performance more severely with CMOS scaling. In particular, there are many cases of an improvement ratio exceeding 20% for 32nm CMOS technology, which is attributed to large interconnect delay. This improvement ratio is close to that obtained by CMOS scaling for one CMOS generation.

As stated in the above section, in sub-65nm CMOS generation, the cache access calculated from CACTI is different from that of actual data. If the cache access time for 90nm CMOS is 1, that for 65nm is about 1.17, and that for 45nm is about 1.30, estimated from various industries’ data. Adjusted processing time using these correction factors for sub-65nm CMOS, is shown in Fig.3. These results indicate that performance of the 3D system in “65 nm” CMOS technology is superior to that of the 2D system in “45 nm” CMOS technology, and that performance of the 3D system in “45 nm” CMOS technology is superior to that of the 2D system in “32 nm” CMOS technology.

It indicates that in the present case, the design change from 2D- to 3D-circuit is superior to the miniaturization based on CMOS scaling in a sub-65nm CMOS generation. This fact is also advantageous from a cost saving point of view, since

**Reference**
Hierarchical Cache System for 3D-Multi-Core Processors in sub-90nm CMOS

K. Nomura, K. Abe, S. Fujita, Y. Kurosawa, A. Kageshima

Toshiba corporation
What is the best application of 3D-IC?

Main Effect of use 3D-IC
- Decrease in Interconnect delay, especially for long interconnects between blocks

Multi-core (many-core) processor
Issues

3D design and performance analysis have been studied only for *blocks* in multi-core processors.

- **3D DRAM** [G.L. Loi et al, DAC 2006],
- **3D Cache design** [G. Loh et al, IEEE Micro 2007],
- **3D Core for Processor** [K. Puttaswamy et al, 2007]

“Black box” simulator for processors where we cannot observe effects of decreasing interconnect directly.

Interconnect delay is severely large only for sub-90 nm CMOS. But, most of previous studies use 130 or 90 nm-CMOS
Objective

To clarify improvement of multi-core processor performance using 3D-IC in sub-90nm CMOS technology.
Outline of main part

• Design for Case Study:
  Hierarchal Cache for Multi-core
• Evaluation conditions
• Results and discussion to compare 3D-IC with 2D-IC
• Area Overhead Evaluation
In the circuit based on this topology, processing time of one request created by PE sometimes becomes long.
Initial situation: PE1 is now accessing to L4 cache and PE4s request is waiting at L2 cache.
Design for Case Study: Hierarchal Cache for Multi-core[3/9]

Conventional hierarchal cache topology.

[To avoid such long bus waiting] Increase the parallel processing

PE3 creates new request.
The processing time of PE3 request is long, since PE3 must wait until PE1 and PE4 requests finish.

Use of Crossbar bus is better solution to increase the parallel processing.
Design for Case Study: Hierarchal Cache for Multi-core[2/5]

What kind of Crossbar bus is better.

<table>
<thead>
<tr>
<th>Crossbar Size</th>
<th>Parallel Processing</th>
<th>Circuit Size</th>
<th>Arbitration</th>
</tr>
</thead>
<tbody>
<tr>
<td>2x2 XBAR (MUX)</td>
<td>Low</td>
<td>Small</td>
<td>Fast</td>
</tr>
<tr>
<td>4x4 XBAR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8x8 XBAR</td>
<td>High</td>
<td>Large</td>
<td>too slow</td>
</tr>
</tbody>
</table>

8x8 XBAR bus has high parallel processing ability.
8x8 XBAR bus size is very large.
Arbitration of 8x8 XBAR bus is much slower than that of 2x2 XBAR bus.

we decided to use 4x4 XBAR bus in our topology design.
Hierarchal topology design.

Our Design hierarchal topology of 16-core microprocessor

Each PE includes one L1 cache.
Four 4x4 Crossbar Bus(XBAR bus) are used to connect between lower level caches.
The common bus is used to connect between upper level caches.
Design for Case Study: Hierarchal Cache for Multi-core[4/5]

2D circuit layout

- 4x4XBAR is assigned to each corner position.
- L3 and L4 caches are assigned to the center position.
Design for Case Study: Hierarchical Cache for Multi-core[4/5]

2D circuit layout

- 4x4 XBAR is assigned at each corner.
- L3 and L4 caches are assigned at the center position.

4PE can connect to L2 caches at the same time through 4x4 2D-XBAR bus.
Design for Case Study: Hierarchal Cache for Multi-core[5/5]

3D circuit layout

4PE and 4 L2 are assigned in 4 tiers. TSV-based 3D-XBAR bus are used for connecting between 4PE and 4 L2.
Design for Case Study: Hierarchal Cache for Multi-core[5/5]

3D circuit layout

4 PEs can connect to L2 cache at the same time through TSV-based 3D crossbar bus.
Outline of main part

• Design for Case Study: Hierarchal Cache for Multi-core
• Evaluation conditions
• Results and discussion to compare 3D-IC with 2D-IC
• Area Overhead Evaluation
Interconnect delay

Evaluate conditions[1/4]

Calculate interconnect delay using interconnect theory (pi-model) and the data in ITRS 2006

- The long interconnect length is from 1mm to 10mm in our design.

- The interconnect delay with repeater is from 0.1ns to 1ns.
Thought Silicon Via (TSV) delay

Considering data on R&D of TSV technology, the diameter of TSV is 5-10μm in the near future.

-RC delay of TSV is less than 0.1 ps with 5 μm diameter
Thought Silicon Via (TSV) delay

Compare TSV delay with interconnect delay

TSV delay < 0.1 ps !! (negligibly small)
- TSV delay is much small enough to neglect.
PE and Cache

- **PE:** $4.38mm \times 8.43mm (65nm CMOS)$
  refer to the size of Intel quad-core processor (Tukwila)
  [B. Stackhous et al, ISSCC 2008]

- **Cache performance**
  use CACTI 5.2 to determine cache area and cache access time. [S. Thoziyoor et al, HP Lab, 2008]

- **Cache capacity and Cache miss rate**
  Processor performance strongly depends on the cache capacity and cache miss rate.

<table>
<thead>
<tr>
<th></th>
<th>Cache Capacity</th>
<th>Cache miss rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$(C_{L2}, C_{L3}, C_{L4})$</td>
<td>$(P_{L2}, P_{L3}, P_{L4})$</td>
</tr>
<tr>
<td>(I)</td>
<td>(256KB, 8MB, 24MB)</td>
<td>(A) (1%, 2%, 0%)</td>
</tr>
<tr>
<td>(II)</td>
<td>(512KB, 20MB, 60MB)</td>
<td>(B) (10%, 20%, 0%)</td>
</tr>
<tr>
<td>(III)</td>
<td>(768KB, 48MB, 144MB)</td>
<td>(C) (20%, 50%, 0%)</td>
</tr>
</tbody>
</table>
Outline of main part

• Design for Case Study:
  Hierarchical Cache for Multi-core
• Evaluation conditions
• Results and discussion to compare 3D-IC with 2D-IC
• Area Overhead Evaluation
Results and discussion to compare 3D-IC with 2D-IC[1/4]

The average processing time (PE-> cache->PE)

- The bus waiting time becomes large with increasing cache miss rate.
- Difference between 2D and 3D bus waiting time becomes large.

Cache capacity and cache miss rate in 90nm

Cache Capacity: (768KB, 48MB, 144MB)
The average processing time (PE->cache->PE)

Results and discussion to compare 3D-IC with 2D-IC[2/4]

The reduction time by changing from 2D to 3D circuit is almost same in each CMOS technology.
The design change from 2D to 3D circuit is superior to the miniaturization based on CMOS scaling in sub-65nm CMOS.

- Advantageous from a cost point of view, since CMOS miniaturization cost is increasing drastically.
Results and discussion to compare 3D-IC with 2D-IC [4/4]

Improvement Ratio (3D vs 2D)

- Improvement ratio becomes large with increasing CMOS technology
- 3D system can improve performance of 2D system by over 20%.
Outline of main part

- Design for Case Study: Hierarchal Cache for Multi-core
- Evaluation conditions
- Results and discussion to compare 3D-IC with 2D-IC
- Area Overhead Evaluation
Area Overhead Evaluation [1/2]

Detail of crossbar bus area in one tier

- via pitch is 2 times of TSV diameter.
- calculate the crossbar bus area for each TSV diameter
Crossbar bus area Overhead for PE in each CMOS technology

Area Overhead Evaluation[2/2]

- TSV diameter is less than 10um in the near future.
- Area overhead for PE is less than 3%.
- Area overhead is much small enough to neglect.
Summary


- Evaluate interconnect delay in sub-90nm CMOS technology based on real technology conditions.

- Improvement factor over 20% in multi-core processor performance using 3D-IC in sub-90nm CMOS technology

- Consider crossbar area overhead of 3D circuit and this area overhead is less than 3% for TSV 10um in diameter.
Power Consumptions

Calculate the number of repeater buffers inserted in 2D and 3D Crossbar bus area.

<table>
<thead>
<tr>
<th></th>
<th>2D</th>
<th>3D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Num of repeater</td>
<td>295808</td>
<td>14592</td>
</tr>
</tbody>
</table>

The wire area: decreasing total 38% power consumption.

Decreasing voltage of 3D circuit to be same performance of 2D circuit.

Total 30% power consumption can be decreased
Where is Crossbar bus used?

- Parallel Processing:
  - Low
  - High

- #N of TSV (TSV Area):
  - Small
  - Large

- Arbitor Complexity:
  - Short
  - Long

XBAR use Level
- Common Only
  - Non
- Hybrid
  - First level
- Multi-stage XBAR
  - First & Second level

Second Level
First Level

Hybrid Case selected
ABSTRACT

3D integration technology is a radical new chip assembly technology that promises greater numbers of devices on chip, increased performance, and reduced power consumption. However, in order for this technology to be economically viable, we must be able to test each die before it is bonded to the rest of the die to form a stacked chip. Pre-bond test presents interesting new challenges because chip functionality and connectivity is partitioned across different pieces of silicon. We will explore these challenges and present test strategies to address them. Our solutions are simple extensions of current scan-based test technology, enabling simple integration of 3D into current test systems. Our results show that full pre-bond test can be achieved at equal or lower cost than testing an equivalent planar design.

1. INTRODUCTION

3D die stacking is a promising new technology that enables the tight integration of multiple silicon die in a vertical stack [3][6]. In its simplest form, multiple planar designs are stacked vertically, significantly increasing on-chip device count. In addition, these planar layers do not need to be homogeneous as they could be designed separately using different design flows. As such, 3D stacking makes the integration of heterogeneous devices onto a single package possible. At the opposite end of the spectrum, individual circuits may be split across multiple die in the stack. Such designs promise simultaneous reductions in delay, power, and area. Unfortunately, 3D die stacking also presents tough new challenges to industry including power delivery, routing, 3D aware design flows, and thermal dissipation, along with many others. One chief challenge is testability, specifically testing each individual die pre-bond (i.e. before the die are bonded together to form a complete chip stack). Before bonding, design functionality is partitioned across multiple die. Depending on the partitioning styles, only partial circuits may exist on any single layer. A simple solution is “bond and pray,” where the stack is fully assembled and then tested as a complete design. This is not a practical solution because the manufacturing yield will fall off exponentially as the number of stacked die increases. This is turn imposes an economic limit on the number of die that can be stacked.

In this paper, we will explore the range of challenges presented by pre-bond test and discuss several techniques that can enable this crucial technology. We will consider the different granularities of 3D partitioning as well as the special needs of the so-called “hardcore” hardware. We will also present case studies representative of these varying challenges and demonstrate the application of such test strategies.

2. TECHNOLOGY PARTITIONING

We begin at the granularity of technology partitioning. At this granularity, die manufactured in different technologies, for example high-speed CMOS and high-density DRAM, are bonded together in a 3D stack. Each individual die is effectively a planar design where TSVs replace what would normally be off-chip connections. Of course, TSVs are orders of magnitude smaller than that board or package wiring, so 3D significantly reduces latency and increases bandwidth between the components.

The test challenges here are minimal. Each design is completely functional, and each die can be individually tested just as they would be in standard system with off-chip DRAM. Traditional boundary scan can be employed to test the quality of the actual TSV bonds. One small concern: once the die are bonded into a stack, all but the topmost die are physically inaccessible to probing. Thus, the test hardware must be designed in such a way as to allow logical access to these lower layers. Fortunately, current test hardware designs are hierarchical in nature [2], so 3D test can be seen as simply another level in the hierarchy.

3. ARCHITECTURAL PARTITIONING

The next smaller granularity is the block level. At this granularity, the die are manufactured in the same technology, but different functional units are partitioned across the stack. For example, we might stack the arithmetic units on top of the register file to reduce the length of the operand and result buses. Such a partitioning scheme increases the test difficulty; a planar test strategy that calls for one functional unit to supply test data to a neighboring functional unit is no longer a viable strategy if these units exist on different die. Once again, extensions to current scan-based test techniques can be applied to achieve the required test coverage. In this case, it comes down to establishing controllability of input signals from TSVs and observability of output signals to TSVs. Such functionality can be provided with well-studied test techniques like PRPG, MISR, and BILBO hardware [8].

3.1 Architecture Test Strategy

To enable test for this granularity, we adopt the scan island test architecture as implemented in the Alpha 21364
processor [2]. In this design, the processor is subdivided into logically independent islands with the use of scan chains. The scan chains all connect to an island scan port (ISP). The ISPs are connected in series to the central scan controller (CSC), which then interfaces to the ATE for in-house test or the IEEE 1149.1 TAP for board-level test.

Applying the concept of scan islands to 3D, we find that each individual silicon layer is already a perfectly isolated island pre-bond. Thus we adopt the scan island test architecture for 3D test (Figure 1). ISPs are replaced with layer test controllers (LTC). Pre-bond each LTC interfaces directly with the ATE to run the layer test. Post-bond the LTCs are connected serially to the CSC, just as in the 21364. This hierarchal approach gives us full test access at each stage of manufacturing.

Of course, there’s no reason for each layer to be limited to a single island. As appropriate for the test requirements of a specific design, a layer may be subdivided into multiple scan islands. Each ISP would then connect in series to the LTC.

With this general test architecture in place, we must consider the individual cross-layer signals. Just as the Alpha test team inserted scan registers to isolate islands, we insert them to establish controllability and observability of TSV signals. In the worst case, two registers are required per TSV, one on the sending layer to observe the test output and one on the receiving layer to provide the test input. Figure 2 shows an example of this design applied to a high-speed pipelined adder. Note the insertion of pass FETs to disable the test signals post-bond. One distinct advantage of 3D is that it is possible to avoid adding any extraneous gates to the operational path. The cost of this is design is that the test registers become useless post-bond. Traditional mux insertion is another option, of course. If we choose this route, the pair of control and observation latches on each TSV can be used to enable boundary-scan-like test of the TSVs themselves post-bond. This may be desirable as it enables very quick verification of the TSV connections.

These varying test strategies highlight the flexibility of our test architecture. It is up to each 3D test team to decide which option is best for a particular product.

Figure 1: A generic 3D implementation of the scan island architecture. (a) shows individual scan registers connected in series to form scan chains and these scan chains connected to the CSC. (b) shows the LTCs connected in series to the CSC for post-bond test.

Figure 2: Shown is a three-stage pipelined adder which first adds the low-order bits, then adds the high-order bits, and finally computes the associated flags. Attached are injection and observation scan-flops which are integrated into one of the layer’s scan chains. Thick lines indicate multi-bit structures (e.g. thick lines represent buses and thick nFETs represent one nFET per bit in the associated buses).

Figure 3: A floorplan for a two-layer die stack split by architectural block. The gray areas between and around blocks represents whitespace within the floorplan.

3.2 Architecture Experiments

To evaluate our test architecture, we took the widely-studied Alpha 21264 as a case study. To evaluate the overhead of each scan cell, we produced a layout for one in 0.25μm TSMC technology, which closely matched the technology used in Alpha’s 21264A product. The design is a pair of 8T latches, consuming a total silicon area of 75.8μm². To determine the total number of scan cells required, we employed a published 3D floorplanner[15] (Figure 3). From this floorplan, we counted the number of cross-layer signals and calculated the total overhead by assuming a worst-case two scan cells per TSV—this is a worst-case scenario because in a real 3D design many functional latches could also serve in the pre-bond test capacity. In total, 4794 cells were required, for a total silicon overhead of 0.165%, a negligible cost. Detailed results of this case study can be found in [10].

4. CIRCUIT PARTITIONING

The finest partitioning granularity is the circuit level. At this level we see a variety of partitioning schemes from simple sub-block partitions to very ambitious transistor-level partitions, where even individual circuits are split across
several layers. Here pre-bond test becomes trickiest, if not completely impossible. Unlike technology- and block-level partitioning, the sheer number of TSV connections can overwhelm the latch-insertion techniques in the previous section; the number of latches required is simply too great. Fortunately, we see an interesting trend in testing these circuit-partitioned designs. In general, test cost increases superlinearly with circuit complexity. As a result, the cost of separately testing each sub-block or sub-circuit individually plus the cost of testing the TSV connections post bond is usually similar to, and can be significantly less than, the cost of testing the equivalent planar design.

4.1 Circuit Design

For circuit partitioning, we examine a bit-sliced Kogge-Stone adder and a port-split register file[14]. In the bit-sliced adder design, the odd bits are computed on one layer and the even bits in the other (Figure 4). Data is shared between the layers only in the first stage of the computation when the neighboring bits are summed together. Such a design is fairly described as sub-block partitioning, a coarse-grained example of circuit partitioning.

On the opposite end of the circuit-partitioning spectrum is the port-split register file (Figure 5). This 3D design targets the many-ported register files required for today’s very wide out-of-order microprocessors, some of which require as many as twenty access ports. Port-splitting effectively targets the quadratic growth in SRAM cell size, significantly reducing the size compared to a planar design. A smaller cell size means shorter word- and bitlines, which means smaller drivers, so port-splitting is a win in every part of the register file design. Port-splitting is a wonderful showcase of the power of 3D design.

4.2 Circuit Test

With two very different 3D designs, we require two very different test strategies. Testing the Kogge-Stone adder is straight-forward. With only two TSVs per adder bit, we can effectively apply scan test to this partitioning. Conveniently, the two signals that cross the layer boundary are already observable on the source layer. Thus, unlike above, we only need one scan register per signal to act as a test control on the receiving layer, reducing the scan test overhead by half.

Testing the port-split register file, on the other hand, requires a completely new approach. The bottom layer, which contains the actual storage cells, can be tested with any standard RAM test, e.g. walking ones. The other layers, obviously, cannot. To test these layers, we propose transmit test. The idea is simple; we place a test vector on one write port, pass it through the pass FETs, and read it from the read ports, all in a single test cycle. This fully exercises the address decoders, write drivers, pass FETs, and sense amplifiers, enabling full coverage of stuck-at faults. In order to run this test, at least one write port and one read port is required on each layer, which becomes a DFT requirement for the design team to meet. Fortunately, the number of read ports is usually well-balanced with the number of write ports (no worse than two-to-one), so meeting this constraint is not overly difficult. Of course, one port could always be made read/write just for the sake of test if required.

Running a transmit test is easy enough. We must be able to source test addresses to the decoders and test vectors to the write ports. We must also be able to read test vectors from the read ports. All of these requirements are already met by the test hardware that is in place for standard RAM test. The only change would be in the control logic, which would have to alter the timing of read and write enable signals to allow for test vector transmission.

4.3 Circuit Experiments

For our experiments, we produced two-layer implementations of the adder and register file. To evaluate our test strategies, we used two different tool flows. For delay and energy measurements, we used the 3DMagic [4] design tool, which can produce 3D VLSI layouts. These layouts were then extracted to HSPICE for simulation.

To evaluate fault coverage, we used the FlexTest tool from Mentor Graphics[5]. This tool takes VHDL models as input and both produces a set of test vectors and calculates the fault coverage. To determine the total cost of 3D test, we sum the cost of testing the bottom layer pre-bond, the top layer pre-bond, and the TSVs post-bond.

We will briefly summarize the results here; a full analysis is available in [11]. Confirming the prior work, both 3D designs out-performed their planar counterparts in area, delay, and energy simultaneously. Most interesting, however, is that both 3D designs were cheaper to test as well. The planar Kogge-Stone adder required 313 test patterns, while the 3D equivalent required slightly less, 301 patterns—146 for the top layer, 145 for the bottom, and ten for the TSVs. The results are even better for the register file. Using Suk
...and Reddy’s Test B[1], the planar register file requires 8192 accesses. Our 3D register file, on the other hand, requires only 4864 test accesses—4096 to apply Test B to the bottom layer, 256 to execute a transmit test on the top layer, and 512 accesses to test the TSVs—a 40% reduction in test cost. This result showcases how 3D can not only significantly improve circuit performance but in some cases significantly reduce test cost as well.

Of course, a transmit test will not work if it takes too long or consumes too much power. And there is reason to worry because the transmit test requires a very long bit line to be charged through two small pass FETs in series. To test the practicality of transmit test, we simulated the test environment in HSPICE as well. Transmission takes approximately 1.5 ns, compared with 1.4 ns for a planar read and 1.0 ns for a 3D read. The energy results are similar; 0.16 pJ for transmission versus 0.15 pJ for a planar read and 0.13 pJ for a 3D read. These results demonstrate the feasibility of transmission test in a real test environment.

5. IMPLICATION TO CLOCK AND POWER DISTRIBUTION

Functional logic is not the only part of the design affected by 3D integration. We must also consider the chip’s hard-core: its power and clock networks. Thankfully, to minimize IR drop, modern processors employ dense grids of power and ground wiring. Such grid designs ensure that these critical nets are fully connected within each and every die in the stack. Unfortunately, the clock distribution network is not so simple.

Previous design work has shown that the optimal 3D clock distribution is one in which the clock is distributed on a single layer and provided to other layers through TSVs are the leaves of the clock tree[13] (Figure 6(a)). Such a design minimizes the wirelength of the clock tree, thus also minimizing power consumption. However, it also creates thousands of disconnected clock domains on each layer (excepting the distribution layer) pre-bond. Without a functional clock, all of the test methods described above are completely useless. The simplest solution is to distribute the clock through a fully connected tree on each layer; these layers can then be connected with a single central TSV (Figure 6(b)). Unfortunately, this design is very wasteful due to the redundant nature of the trees.

In a pre-bond testable 3D design, we have to have an operational clock pre-bond, but we don’t want the massive power cost of the redundant trees. We propose two potential solutions, which are currently works-in-progress. The first and simpler design starts with an optimized test-unaware 3D clock tree as described above. Then, to enable pre-bond test, each layer—excepting the layer that is already fully connected—is augmented with a pre-bond test tree that connects all the leaf nodes together. This tree includes a gating signal; pre-bond the tree is on to enable test, post-bond it is off so power is not wasted switching the redundant distribution wiring.

The second design is more complicated but also more useful. Basically, in this design, even the distribution portion of the original 3D tree can be gated. The effect of this slight alternation is that the distribution wiring on any layer can be used to clock all of the leaf nodes throughout the stack. Thus, if one distribution network fails, e.g. due to electromigration, another layer can be switched on to allow the stack to continue operating. Thus we use the redundancy in the previous design to increase product lifetime. The downside is the complexity of the design. It is of the utmost importance to minimize skew in the clock network. In the first design, this is fairly simple because there is only one post-bond clock to design. In the second design, there are n trees to design, much more difficult. Additionally, each distribution net will intersect the leaves of the clock tree at slightly different points (plus or minus a couple TSV delays), exacerbating the skew problem. Whether the benefits of the more complicated option justify the costs or if redundant low-skew trees are even possible remains to be seen. However, either design promises a combination of pre-bond testability and optimal post-bond operation.

6. RELATED WORK

3D test is still in the very early stages. Below are contributions others have made to this emerging field. Mak identified a list of challenges facing 3D test going forward[12]. Wu et al. studied scan chain ordering in a 3D stack to minimize wirelength[16]; this work does not consider pre-bond test. Jiang et al. studied the total test time of 3D systems, factoring in pre-bond test[7]. More recently, Lee and Chakrabarty identified the research challenges still yet to be addressed in 3D-ICs[9].

7. SUMMARY

In this paper, we have identified a variety of challenges facing 3D test engineers and presented a number of potential solutions. Taken together, these test strategies provide a framework for enabling pre-bond test in 3D integrated designs. More importantly, these test techniques can be easily integrated into existing test plans and executed on existing test systems. This greatly reduces the barrier to 3D adoption in industry. With these test solutions, we can finally realize the amazing potential 3D integration promises the semiconductor industry.

8. ACKNOWLEDGMENT

This work was supported in part by C2S2, a center of the Focus Center Research Program, a Semiconductor Research Corporation Program and an NSF grant CCF-0811738.
9. REFERENCES


Test Strategies for 3D Die-Stacked Integrated Circuits

Dean L. Lewis, Xin Shao
Professor Hsien-Hsin S. Lee, Professor Sung Kyu Lim
Workshop on 3D Integration, DATE 2009
Nice, France, April 2009
http://arch.ece.gatech.edu/mars.html
3D Integration

- **Technology**
  - Stack multiple active layers vertically
  - Tightly integrate with die to die (d2d) vias

- **Benefits**
  - Routing freedom
  - Higher performance
  - Lower power

Motivation

![Graph showing the relationship between Single Layer Yield and Stack Yield for different numbers of layers. The graph demonstrates a decrease in Stack Yield as Single Layer Yield increases, with a more pronounced effect for higher numbers of layers.](image)

- **Motivation**:
  - Single Layer Yield
  - Stack Yield

- **Legend**:
  - Blue: 1 Layer
  - Orange: 2 Layers
  - Yellow: 4 Layers
  - Green: 8 Layers
  - Red: 16 Layers

- **Key Points**:
  - As Single Layer Yield increases, Stack Yield decreases.
  - The decrease is more significant with a higher number of layers.

- **Example**:
  - At 100% Single Layer Yield, Stack Yield for 1 Layer is 100%, reducing to 80% for 16 Layers.
3D Assembly

- **Wafer to Wafer**
- **Wafer to Die**
- **Die to Die**
3D Integration and Testability

Technology Level

Architecture Level

Circuits Level

BitLine0
BitLine1
BitLine2

ALU 1
ALU 2
ALU 3
ALU 4

CMOS
DRAM
Analog
Architectural Partitioning and Test
Pre-bond Test Challenges
Incomplete Circuits

Architectural Level

Fetch
Decode
Reorder Buffer
Issue
Out of Order Execution
Commit

I Cache
D Cache

Complete Architectural Design

Circuit Level

Complete Register File Design
Pre-bond Test Challenges
Incomplete Circuits

Architectural Level

Circuit Level

Fetch

Reorder Buffer

Issue

Commit

Possible Pre-bond Partition

Pre-bond Circuit

Lewis and Lee. ITC 2007
Pre-bond Test Challenges
Incomplete Circuits

Architectural Level

- Fetch
- Reorder Buffer
- Issue
- Commit

Circuit Level

10000+

Possible Pre-bond Partition
Pre-bond Circuit
Test Strategy

Alpha 21364

3D Pre-bond Test

IEEE 1149.1 TAP
CSC
ISP
ISP

LTC
CSC
IEEE 1149.1 TAP
LTC
Layer Border Scan Flops

Post-bond Operation

Low Add

Pipeline Stage Register

High Add

Flags
Layer Border Scan Flops

Pre-bond

Low Add
Pipeline Stage Register
High Add
Flags
Layer Border Scan Flops

Add Scan Flops

Si

Low Add

Pipeline Stage Register

High Add

Flags

So
Layer Border Scan Flops

Add Control Signal

Test_Enable

Si

Low Add

High Add

Pipeline Stage Register

Flags

So
Layer Border Scan Flops

Pre-bond Test

Si

Low Add

Pipeline Stage Register

High Add

Flags

So
Layer Border Scan Flops

Post-bond Operation

Si

Low Add

Pipeline Stage Register

High Add

Flags

So
Scan Flops Not Required

Low Add

High Add

Flags

Pipeline Stage Register
Scan Flops Not Required

Si

Low Add

Pipeline Stage Register

High Add

Flags

So
Scan Flops Not Required

Pre-bond

Si

Pipeline Stage Register

Low Add

High Add

Flags

So
Scan Flops Not Required

Post-bond

Si

Low Add

High Add

Pipeline Stage Register

Flags

So
Experiment

- Based on 21264 Architecture
- Floorplanned microarchitecture blocks
  - Two die layers
- Determined widths of inter-die buses
- Laid out scan cell

Results

Layer 1

Layer 2

<table>
<thead>
<tr>
<th>Component</th>
<th>Layer 1</th>
<th>Layer 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scan Cell Size</td>
<td>75.8 μm²</td>
<td></td>
</tr>
<tr>
<td>Inter-die Vias</td>
<td>2397</td>
<td></td>
</tr>
<tr>
<td>Scan Cell Count</td>
<td>4794</td>
<td></td>
</tr>
<tr>
<td>Two cells per via</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area</td>
<td>0.363 mm²</td>
<td></td>
</tr>
<tr>
<td>Overhead</td>
<td>0.165%</td>
<td></td>
</tr>
</tbody>
</table>

Lewis and Lee. *ITC 2007*
Circuit Partitioning and Test
Bit-Split Kogge-Stone Adder

Planar

3D
64-Bit Adder Layout - Planar
64-Bit Adder Layout – 3D

- **Reductions**
  - Area – 34%
  - Latency – 18%

- **Power** – 13%
- **Test Patterns** – 0%
Port-Split Register File

Planar

3D
Register File Layout - Planar
Register File Layout - 3D

- 39% Area Reduction
- Delay Reduction
  - Read – 25%
  - Write – 45%
- Energy Reduction
  - Read – 15%
  - Write – 27%
Port-Split Test

- 40% Test Pattern Reduction
- Performance
  - Delay – 110%, 148% of read delay
  - Energy – 110%, 130% of read power
3D Clock Tree Design
Clock Routing

Power/Routing Optimized  Pre-bond Test Optimized

Layer 1

Layer 2

Layer 1

Layer 2
Real Clock Tree
Real 3D Clock Tree
Testable 3D Clock Tree
More Trees Are Better

Pre-bond Test

Reliability
Conclusion

- We can test it!
- Pre-bond test can be mostly implemented with low-risk extensions to existing test methodology
- Some 3D designs will require new approaches to test, but these are not show-stoppers
- 3D may even significantly reduce test cost
Thank You
Pre-bond Test Challenges
Wafer Probing

Probing

Probe Pads
Why should we not tape next year?
Testing of 3D Integrated Circuits: More Challenges than Solutions

Krishnendu Chakrabarty
Department of Electrical and Computer Engineering
Duke University
Durham, North Carolina
USA
A barrier to widespread 3D adoption: Insufficient understanding of 3D testing issues and the lack of design-for-testability (DFT) techniques
Test Challenges

• Today’s 3D chip designs do not consider test cost and the implications of design decisions on testability
  • Gap between anticipated (perceptual) benefits and practical value

• Among all EDA challenges for 3D IC design, tools and methodologies for 3D IC testing are regarded as the “No.1 challenge” (keynote speech at the 2007 3D Architecture Conference by Ted Vucurevich, former CTO of Cadence Design Systems)
Test Challenges

- Traditional IC manufacturing: Wafers are probed and individual dies tested (a process referred to as wafer sort) before they are packaged.
- 3D integration: How to sort individual wafers before they are bonded together?

Yield enhancement: Bond pre-tested wafers, or sort wafers first and stack matched dies (based on speed or power)
Test Challenges

• Serious obstacles to pre-bond testing of wafers in 3D integration
  – Wafer probing is a problem for F2F bonding
    • Bottom die has up to hundreds of thousands of Copper pads
    • Their small size and large numbers make probing of signals impossible
    • Top wafer cannot be probed from the Copper side. TSVs are buried and C4 pads are not fabricated prior to bonding
  
• In F2B stacking, top die is more testable since the C4 pads can be fabricated on the top layer
  – However, top wafer has to be thinned ⇒ problems of ultra-thin wafer processing
  – Probe card applies a force of 3-10 g per probe ⇒ probe force per wafer as high as 60-120 kg!
  – TSVs have low fracture strength
Test Challenges

- **Known Good Die**
  - Concerns: quality of the incoming bare (unpackaged) dies and wafers prior to stacking
  - Testing needed to ensure structural integrity and performance of stacked modules

- **Significant differences between the KGD problem for MCM testing and for 3D IC testing (“Pretty Good Die”?)**
  - Wafer probing is a major limitation for thinned wafers and pre-bond wafers in 3D integration: few probe contacts and wafer touchdowns
  - Due to design partitioning in 3D ICs, any given layer in a stacked 3D IC does not always include complete functionality
    - Type of tests that can be applied at this stage is limited
  - Smaller form factors and higher performance
    - Test solutions must target more speed- and layout-related defects that arise at nanoscale dimensions
  - More structural constraints: Limits on the number of TSVs and the availability of I/Os only on the layer with the C4 pads
New Defect Types

• Alignment, stacking, and thinning of wafers add extra steps to the manufacturing process
• Foreign particles caught between wafers during bonding can cause voids, peeling, and delamination
• Wafers must be precisely aligned for bonding
  – Alignment problems are likely to lead to imperfect via connections
• Edge effects
  – Inter-wafer gap is typically greater at the edges and the bond is weaker
  – Bonded edges vulnerable to chipping, peeling, and delamination
• Cracking can occur during the stacking process, due to loading forces, backside grinding, and die thinning
• Random open defects can result from dislocations, oxygen trapped on the surface, voids formation, and mechanical failures
Testing of the TSV “Electronic Pillars”

• Fabricated TSV dimensions have been reported in the range of 0.4 microns (small) and 200 microns (large)
  – Process improvements can allow manufacturers to integrate over 60,000 TSVs in a package
• TSVs must be tested for defects such as underfill and problems due to thermal effects
• Defective TSVs will prevent test access to logic blocks on different layers
  – Consider redundant TSVs?
    • Considerations: “Keep-out area”, how many extra TSVs? Where?
Thermal and Power-Delivery Concerns

• Oxide encapsulating each layer make upper tiers increasingly vulnerable to thermal effects
• Thermal fatigue also a potential problem for C4 bumps
  – Test patterns must therefore be generated to target “hot” regions on different layers as well the “hot” TSVs
• Thermal problems are accompanied by power-delivery concerns in 3D ICs (“two sides of the same coin”)
• Need for test patterns that can accurately evaluate IR-drop and di/dt effects
  – Similar testing problems are being addressed for today’s planar ICs
  – New methods for high-density 3D ICs need to have better scalability.
Test Economics

- Close coupling between test-cost analysis and 3D physical design optimization
- Wafer cost, bonding cost model, die area, test cost related to various stacking options

(Yuan Xie, Penn State University)
Test Requirements

• Access for TSV pads at all level of the 3D package assembly
  – ESD requirements for all TSV pads

• Standard DFT access to test the TSV stacks
  – Something like a Boundary scan access to TSV ports
  – Provide test access to all the test features on every die
  – Similar to providing test access to chips on a board

• Hierarchical access to test controllers
  – As 3D stacks get taller, test controllers get buried.
Some Possible Solutions

- Test-access mechanisms and test scheduling
Test Access
(Prasad Mantri, Sun Microsystems)
Partial-Functionality Testing
(H. H. Sean Lee, Georgia Tech)

3D Memory Port-Splitting (four ports, two for each die layer)

3D Scan-island Design
Applications Timeline → Tools Timeline

Drivers:
- Imagers
- RF / Mixed Signal
- Heterogeneous Integration

The technology pull is real – But, 3D design requires tools designed for 3D

© R³LOGIC 2009 – ALL RIGHTS RESERVED
SoC Economics

Price-point reached
3D Package – Chip Co-Design

- Match chip I/O’s to interposer TSV’s
- Match TSV’s to RDL routes
- Typically large difference in via pitches from chip to TSV to package
- Verify thermal / mechanical / electrical effects
Architecture Evaluation

- True Front-to-Back 3D Design Flow
  - Multiple process handling
  - Full interoperability with 2D infrastructure
  - 3D partitioning at System-C level
  - Thermal interface / backannotation
  - Chip-Package co-design

Add Rapid Thermal / Power Evaluator Here

© R³LOGIC 2009 – ALL RIGHTS RESERVED
The development of IC technology is driven by the need to increase performance and functionality while reducing size, power and cost. The continuous pressure to meet those requirements has created innovative, small, cost-effective 3-D packaging technologies. 3-D packaging can offer significant advantages in performance, functionality and form factor for future technologies.

Key process technologies enabling 3-D chip:

- Via formation
- Insulator, barrier and seed deposition
- Copper filling (plating),
- CMP
- Wafer thinning
- Die to Wafer /chip alignment, bonding and dicing

This presentation will investigate these techniques that require interdisciplinary coordination and integration that previously have not been practiced. We will review the current state of 3-D interconnects and the of a cost effective Via-first TSV integrated process.

The speaker will incorporate information from the EMC3D consortium members:

- Applied Materials
- Semitool
- EV Group
- Fraunhofer
- LETI
- SAIT
- KAIST
- Texas A&M
- Rohm and Haas
- Enthone
- AZ

Through collaboration with research partners, the consortium will develop processes for creating micro vias between 5 and 30 µm on thinned 50 µm 200 and 300 mm wafers using both via-first and via-last techniques.

Speaker Information

Paul Siblerud is the VP of ElectroChemical Deposition for SEMITOOL and the program manager for the EMC3D international consortium.

Phone: 406-752-2107
E-Mail: psiblerud@semitool.com
3D TSV integration; consortium approach

Paul Siblerud
Chairman EMC³D
V.P. Semitool

www.emc3d.org
3D in Leti

Around 50 researchers directly involved
Mixed backgrounds: MEMS, µelectronics and Packaging

200 mm facilities
Current investments on 300 mm line for to 3D and WLP
Our philosophy for 3D technologies development

A generic Toolbox

- Bonding / Alignement
- Interstrata connexions
- TSV
- Thin wafer handling
- Chip to wafer planarisation
- Thermal management

- Advanced research
- Design
- Electrical tests
- Reliability

- TSV for imagers (transferred)
- Multimedia smart cards
- Stacked Mixed technologies
- TSV-free Stack (Via Belt Techno.)
- High density 3D
TSV Technology applied to miniaturized Image Sensors

DATE  Conference 2009
Nice , April 24th, 2009

Pascal Urard
Technology R&D / Central Cad & Design solutions
« COB »
2Mp RawBayer sensor, 3um pixel size

WLP « TSV » 2Mp RawBayer sensor, 1.75um pixel size
Mobile Imaging Driving Factors

3D INTEGRATION ARENA

Process complexity
200mm 300mm
90nm ; 65nm
FSI vs BSI
Supply chain model

Physical Size
Resolution
Pixel size
Module height
Optics

Cost

Performances
Low Light
SNR ratio
Crosstalk
Speed

Cost

Scene Illumination
Luminance SNR post CM

SNR10(Y): measured and Simulation

851 Gr3
851 Simulation (Measured Grade123 QE)
Through Silicon Via Pros and Cons

• Pros
  – Allow even smaller package outline
  – No pad extension needed
  – Lower sensitivity to foreign material at Camera assembly
  – Wire bonding compatible layout
  – Reflow process compatible
  – Better interconnect routing capability

• Cons
  – More complex technology
    • Glass
    • Silicon
    • Back-end processes
  – Cost
Thank you for your attention

Thanks to Jean Luc Jaffard & Imaging development team at ST Microelectronics
DATE 3D Integration workshop
“The Future of 3D Integration from all Angles”

The analog/RF angle

G. Van der Plas, IMEC
What’s this with analog and 3D?

my Boss wants to fit the VCO of our new 32n design in a 100x100u area!?

But the inductor takes 400x400u??

Can this 3D stuff help me?

Should we tell him he can put his VCO on a ‘cheap’ 130n layer in a 3D chip stack?

Or maybe integrate the high Q inductor into the package??
What’s this with analog and 3D?

We got this nice SiPs where I can get my GaAs amp tightly connected to CMOS controller to linearize and get better efficiency.

But the parasitics are still too large for my bandwidth.

Can this 3D stuff help me?

Should we tell him that with 3D he can put his GaAs and CMOS so tightly together that there are virtually no parasitics in between?
What’s this with analog and 3D?

All this stuff about full system integration up to antennas, great but I don’t see it happening. Try to put a mmWave antenna on silicon and you’ll see why.

Can this 3D stuff help me?

Should we tell him that with 3D he can put mmWave antennas in package improving system integration and performance?
What’s this with analog and 3D?

>75% of SoCs have 'analog' content
So what’s this with 3D

For analog:
- Improved performance
- More functionality
- Lower power
- Cost effective?
The 5 challenges for 3D integration seen from analog/RF perspective

- Pathfinding for analog/RF in 3D
  - the opportunity is also posing the biggest challenge.
  - What to partition where: RF in separate die, High Q passives in package, ...
  - What technology for tiers: go for BiCMOS, GaAs, ... ?
  
  - 1M $ question: in all this complexity who can get 'the 3D-SoC architecture' right?

  ⇒ a need for bottom-up models, a structured flows & costs modeling

- Analog/RF Test
  - testing analog/mixed signal is difficult, mostly functional and application specific
  - how to test when analog/RF is spread all over and not functional before 3D integration, what with KGD test?

  ⇒ a need for testing of incomplete analog functions

- Electro-Magnetics (EM)
  - analog/RF is very sensitive to EM coupling.
  - Increased integration brings close EM couplings.
  - Every 6dB of interference means 1b less in dataconverter performance

  ⇒ a need for understanding EMC issues in 3D

- Thermal
  - analog/RF is notorious for generating high heat (hot spots): PMU, PA, ...
  - and also for being very susceptible to thermal gradients

  ⇒ a need for thermal modeling capabilities to verify and optimize performance of analog/RF

- Reliability & Mechanical Stress
  - analog/RF is heterogeneous: GaAs, GaN, ...
  - analog/RF pushes even more materials in the 3D stack with different mechanical properties

  ⇒ a need for mechanical modeling capabilities to verify and optimize performance of analog/RF in stacks

Let's work on these
Mnemonic

- Pf
- E
- Th
- PfATEThR
- R
Disclaimer

- any similarity of analog designers to actual persons is purely coincidental. Analog designers are basically nice persons.

@ imec we have programs to convince girls to choose for technical professions.
aspire invent achieve
Introduction

As predicted by Moore’s Law, modern technologies allow high density integrated circuits with billions of transistors. This remarkable progress has been mostly achieved by reducing structure sizes. However, further reduction of the lithographic structures is becoming increasingly expensive. Another enabler of Moore’s Law has been the introduction of new technologies (e.g. copper interconnects, SOI, strained silicon). Currently it is becoming increasingly obvious that new technologies and methodologies rather than structure reduction are the key to further performance enhancements. The employment of 3D integration technologies in which the active devices are placed in multiple layers is one promising possibility to achieve a performance boost. Although discussed for some decades, 3D integration has only recently gained practical importance.

Power consumption and performance of a chip are mainly influenced by interconnects. Global wires which do not scale well need a growing number of repeaters. 3D integration allows to shorten interconnects which results in an improvement of performance and lower power consumption. Higher data bandwidths are possible and enable the design of high efficient cache structures for microprocessor architectures. Smaller footprint sizes increase the yield of fabrication and improve the usage for mobility devices with tightened weight requirements. New system designs are possible due to the option to combine heterogeneous technologies. A complete new set of devices, such as vertical transistors, could dramatically change the way integrated circuits are designed.

If there are so many advantages, why isn’t 3D integration widely used yet? One reason is the fact that until now the scaling of the semiconductor structures has been easier to achieve than investing into a new methodology. However, the obvious advantages of the 3D integration also cause challenges to deal with in both the technology and the physical design. Typical problems in 3D integration technologies are reliability, alignment accuracy and testability issues. The physical design process for 3D structures is even more complex than its 2D counterpart. The third dimension dramatically increases the solution space. Besides the higher complexity, additional constraints such as stringent thermal issues must be addressed.

In order to handle such complex problems in physical design, a feasible modeling of the layout problem is necessary. Specifically, all design steps need an efficient data structure to represent the real geometries of modules or blocks inside the algorithmic space (i.e., computer memory). Currently developed 3D data structures realize such a mapping and allow the use of efficient optimization approaches.

In the first part of this talk we will give an overview of the different 3D integration technologies that are in use today. In the second part, modern data structures used for 3D integration are presented with a special emphasis on 3D floorplanning problems.

3D Integration Technologies

Traditionally, an integrated circuit has been comprised of one active device layer covered with several metal layers. The transistors are integrated on the active layer and the metal layers are used to realize the interconnections between them. This two dimensional approach is currently reaching its limits because even with shrinking structure sizes the wire length increases into dimensions that pose a severe problem to electrical signal properties. Furthermore, the maximum circuit area is limited by the given signal propagation. Hence, it has been becoming obvious that with conventional 2D technology the recent trend towards multifunctional mixed-signal devices is hardly reachable.

One solution to this problem has been the increased implementation of so-called 3D circuits based on 3D integration technologies. Here, the active devices are not limited to one layer but are also placed on top of each other. One famous example for the usage of this technology is the stacking of memory dies to enable high capacity memory cards. Enabling technologies for this kind of integration
are, among others, through silicon vias and inductive or capacitive coupling.

A typical distinguishing feature of 3D integration technologies is the level on which the vertical integration takes place. 3D integration on chip level such as System on Package (SoP) and System in Package (SiP) are often classified as 3D packaging. 3D integration on wafer level is another promising 3D integration technology. Examples for the latter include wafer stacking or donor wafer bonding. These 3D integration technologies are presented in this talk in detail. We also compare and characterize them regarding their future potential.

3D integration gives plenty of opportunities to improve performance and functionality of future circuits. Along with this new technology, new challenges arise that will be addressed in this talk. Such challenges include technological problems, like high accuracy alignment or reliability issues, and physical design issues such as considering new thermal constraints.

### 3D Data Structures

In electronic design automation, specifically during physical design, data structures are used to store information about various layout elements and their properties. Such data structures are an abstract model of the corresponding design problem.

Efficient data structures integrate additional helpful properties of the layout elements, such as direct access to neighbors. They build up a solution space which then can be investigated with an optimization method like simulated annealing. This solution space should be non-redundant, as small as possible and include the best solutions. Furthermore, an efficient implementation of a data structure must allow fast operations like rotations, exchanges and the transformation from the abstract representation into the real geometries.

Early 3D physical design flows have utilized classical proven data structures, such as one slicing tree for one active device layer. However, these so called 2.5D approaches have the disadvantage that tight linking between the layers has been neglected, thus, preventing for example a successful thermal-driven design.

It is obvious that modern 3D data structures should support full 3D layouts. The majority of the currently developed 3D data structures are extensions of their well known 2D predecessors. One example is the T-Tree where the efficient properties of its 2D version (B*-Tree) could be maintained by implementing a ternary tree (instead of a binary tree) that enables the representation of the third dimension. However, some of these 3D extensions of data structures lose their properties which make them so efficient in the 2D case. This requires finding new data structures specifically designed for the 3D integration which are truly able to fulfill the new demand determined by 3D integration technologies.

In this talk, an overview of 2D data structures is given that enable efficient 3D representations by transforming them into the third dimension. Furthermore, currently published new 3D data structures independent of a particular 2D predecessor are discussed. Data structures available for 3D floorplanning are then analyzed and compared in order to draw conclusions about their future potential.

Selected material about 3D integration technologies and 3D data structures can be found in the references. Motivation and challenges of 3D integration are pointed out in [1] and [3]. In [4] and [5], two widely used 3D integration technologies are presented. Recent 3D data structures are described in [2], [6] and [7].

### References


3D Technologies and Data Structures

An Overview

Robert Fischbach, Jens Lienig
Institute of Electromechanical and Electronic Design
Dresden University of Technology
fischbach@ifte.de, lienig@ifte.de

As predicted by Moore’s law, modern technologies allow high density integrated circuits with hundreds of millions of transistors. This remarkable progress has been mostly achieved by reducing structure sizes. However, further reduction of the lithographic structures is becoming increasingly expensive. Another enabler of Moore’s law has been the introduction of new technologies (e.g., copper interconnects, SOI, strained silicon). Currently, it is becoming more and more obvious that new technologies and methodologies rather than structure and construction allow for further performance enhancements. The employment of 3D integration technologies, in which the active devices are placed in multiple layers, is one promising possibility to achieve a performance boost. Although discussed for some decades, 3D integration has only recently gained practical importance.

3D Integration Technologies

3D INTEGRATION TECHNOLOGIES can be divided into 3D packages and 3D integrated circuits. One of the most enhanced 3D packaging method is the System on Package. It combines the advantages of several packaging technologies into an enhanced functionality printed circuit board. Many diverse components can be combined on such a system, e.g., digital and analog components, embedded passives, optical modules, filters, and antennas. 3D integrated circuits consist of multiple device layers which are integrated directly on wafer level. The most common technologies are layer growth, donor wafer bonding and wafer level stacking.

3D Data Structures

In ELECTRONIC DESIGN AUTOMATION, specifically during physical design, data structures are used to store information about various layout elements and their properties. Such data structures are an abstract model of the corresponding design problem. Efficient data structures integrate additional helpful properties of the layout elements, such as direct access to neighbors. They build up a solution space which can then be investigated with an optimization method like simulated annealing. This solution space should be non-redundant, as small as possible and include the best solutions. Furthermore, an efficient implementation of a data structure must allow fast operations like rotations, exchanges and the transformation from the abstract representation into the real geometries.

Summary

3D INTEGRATION is becoming a new force keeping Moore’s law still holding in today’s nano era. Numerous 3D data structures have been recently developed in order to make 3D technologies accessible to an efficient and automatic design. To further exploit the advantages of the extra dimension in 3D integrated circuits, a layout designer and tool developer needs to be aware of this rapid development.

This table presents most common 3D data structures applied to 3D floorplanning problems with regard to their publication date, runtime complexity of the operations, size of the solution space and main characteristics. It contains an abstract representation and includes full implementations, grid structures, and graph-based representations.

<table>
<thead>
<tr>
<th>Data Structure</th>
<th>Year</th>
<th>Complexity</th>
<th>Solution Space</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D sub-Transitive Cluster Graph</td>
<td>2018</td>
<td>3</td>
<td>Three or two sequences (bid)</td>
<td>Three transitive graphs</td>
</tr>
<tr>
<td>T-Tree</td>
<td>2004</td>
<td>3</td>
<td>T-Tree诱人</td>
<td>Ternary tree, nodes: modules, branches: neighbor information</td>
</tr>
<tr>
<td>3D Rounding-Square Graph</td>
<td>2005</td>
<td>3</td>
<td>3D Rounding-Square Graph 4</td>
<td>3D grid structure</td>
</tr>
<tr>
<td>3D Skewed Tree</td>
<td>2005</td>
<td>3</td>
<td>3D Skewed Tree 4</td>
<td>3D grid structure</td>
</tr>
<tr>
<td>Labeled Tree and Dual/Sequence</td>
<td>2006</td>
<td>3</td>
<td>Labeled Tree and Dual/Sequence 4</td>
<td>3D tree, inner nodes: slices, leaves: sequences</td>
</tr>
<tr>
<td>Sequence of modules, number sequence and tree</td>
<td>2007</td>
<td>3</td>
<td>Sequence of modules, number sequence and tree</td>
<td>Efficient data structures represent a physical representation as an abstract model in which several operations can preserve through the solution space. Cost functions are implemented utilizing knowledge from the data structure. A 3D data structure is an abstraction layer between a problem-specific optimization strategy (metaheuristic) and the physical 3D floorplanning problem.</td>
</tr>
</tbody>
</table>
System-Level Exploration of 3-D Interconnection Schemes

K. Siozios, A. Papanikolaou, A. Bartzas and D. Soudris
ECE Dept. National Technical University of Athens

Abstract — 3-D chip stacking is the big next step in system integration. Even though the process technology is maturing, many issues related to system implementation are not fully understood yet. Furthermore, this 3-D integration comes in many different flavors. Adoption in mainstream industrial design flows requires the availability of tools that allow designers to evaluate the cost and benefits of 3-D integration, compared to conventional system implementations. In this paper we propose a software-supported framework for architecture-level exploration and system prototyping on 3-D ICs. The results show that 3-D integration provides significant improvements among others in wire-length, delay, power/energy consumption, as well as die area, compared to conventional 2-D implementations.

I. INTRODUCTION

Three dimensional (3-D) integration technology offers to designers new opportunities for the implementation of complex electronic systems. The System-on-Chip (SoC) paradigm of cramming as much functionality as possible on a single die is giving way to the 3-D IC where the same functionality can be implemented in a number of vertically stacked die. The increased number of neighbors results to significant reduction on total wire-length, which results to improvement on system’s performance and power/energy consumption. Additionally, the 3-D ICs offer three unique advantages compared to conventional approaches, namely the higher system integration, the enable of heterogeneous integration, and the capability for wire-length reduction.

The advantages of 3-D chip stacking seem obvious from a theoretical point of view. However, up to now, there are no EDA tools that can handle this new technology available yet. Therefore 3-D SIC has not percolated up the design abstraction levels. Additionally, even before physical design tools for 3-D become commercially available, path-finding tools will be needed. Such tools will enable the design community among others to explore the design/technology search space and identify for which systems 3D-SIC technology is beneficial.

Related work [1, 2, 3, 4] has been focusing on three major parallel tracks, namely physical design or die/package co-design tools, modeling of 3-D technologies and case-studies. We have to mention that up to now; only few design tools that support these tasks are available from academia. On the other hand, EDA vendors are not yet releasing CAD tools, due to the limited market potential and the fact that 3-D has not yet become a mainstream technology.

In this paper we propose a software-supported methodology that enables exploration of system implementation using conventional and 3-D integration solutions. Such a methodology is also called path-finding, as it aids in establishing a path towards a design/technology sweet-spot in the domain of 3-D integration, while it also can be used to co-optimize design and technology development. Our goal is to make early search-space exploration easier and faster and to enable designers to find the short-list of interesting design and technology choices that meet their specific design objectives, including power consumption and performance as well as fabrication cost.

II. THE PROPOSED PATH-FINDING FRAMEWORK FOR 3D ICs

Figure 1 depicts the proposed path-finding framework, which consists of two steps: i) 3-D stack generation, where the system architecture is realized by appropriately assigning IP components to die, and ii) system prototyping, where a physical prototyping of the stack is performed and performance metrics are estimated. The main attributes of the proposed path-finding framework can be summarized, as follows: i) it has to be relatively accurate to allow designers to make decisions at a very early stage of the design flow, ii) fast, iii) flexible, and iv) to allow feedback to incrementally improve design quality. The input to the framework is a system description in a Hardware Description Language (HDL). Initially, this input is synthesized, in order to retrieve the system’s IP blocks, as well as their appropriate communication.

A. 3-D Stack Generation

During this step all the 3-D related decisions are made.

The first step of 3-D stack generation task partitions the system. At this level the connections between partitions are minimized while respecting constraints, like keeping IP blocks that are fabricated onto non-compatible technology (i.e. DRAM and logic) on different partitions and balancing the area occupied by IP blocks assigned to each partition.

Next, the partitions are assigned to physical die and technology selection per die is performed. This also implies the selection of the appropriate instantiation for each IP component based on the technology of the corresponding die. The optimization objectives during this sub-step include the total system’s power consumption, performance and fabrication cost.

The following sub-step builds a prototype of the stack by deciding on the order of the die in the stack (which goes to the bottom, which to the top, etc.) and the choice of 3-D interconnection option (TSV, wire-bond, etc.). This sub-step is equally important to the previous, as it can alleviate numerous design issues related to the operation frequency and thermal stress of the derived 3-D IC.

Once the 3-D stack generation procedure is complete, the efficiency of the derived 3-D partitioning is estimated based on high-level models, so as to provide fast estimations of the main design metrics regarding the delay, power consumption, area and yield. Whenever the resulting 3-D chip does not meet the system specifications, there is a feedback loop back to the partitioning step to allow designers to modify some of the
decisions already made, like using different IP blocks or different die assignment options.

In contrast to existing approaches for system partitioning, the proposed one provides significant more accurate results. More specifically, rather than focusing solely on minimizing the edge-cut, we are also pay effort to handle a number of design parameters (such as the area imbalance among die, the distribution/variation of power sources over the die, etc).

B. System Prototyping

The result of the previous step is a high level prototype of the 3-D stack that already specifies its main properties, number of die, interconnection options, IP block to die assignment, etc. During the second step of the proposed methodology, we aim to provide a physical prototype which will incorporate physical design information. The main purpose of this step is to determine the spatial location of all blocks, including those that provide inter-die communication (TSVs, pads for wire-bonds, etc), while minimizing the main cost parameters.

Since the decisions related to the 3-D stack have already been taken in the previous step, we can make system prototyping with an extended version of existing 2-D physical design tools. In our flow we perform physical design using the floor-planning and global routing functionalities of Cadence SoC Encounter.

III. RESULTS

For demonstration purposes we show results using a MPEG4 encoder [5] implementation onto 3-D ICs implemented in 90nm technology. More specifically, we provide results about stack implementations that use different 3D interconnection technologies, namely wire-bonds between pads on the different die and Through Silicon Vias. Table I summarizes the performance of the MPEG4 encoder implemented in 2-D/3-D integration technologies. Based on the results, we can conclude that 3-D integration using wire-bonding and TSVs lead to a communication frequency improvement (as compared to 2-D) of about 13% and 21%, respectively. Similarly the gains in power consumption are 11% and 19%, respectively. Wirelength is reduced by up to 16%.

<table>
<thead>
<tr>
<th>Table I: Evaluation for MPEG4 implementation into 2-D/3-D ICs</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-D</td>
</tr>
<tr>
<td>Frequency (Hz)</td>
</tr>
<tr>
<td>3.73E+05</td>
</tr>
<tr>
<td>Power (W)</td>
</tr>
<tr>
<td>Area (um²)</td>
</tr>
<tr>
<td>Wirelength (um)</td>
</tr>
</tbody>
</table>

IV. CONCLUSIONS

A novel software-supported framework for system prototyping on 3-D ICs was proposed. It can evaluate the cost and benefits for alternative 3-D process technologies, as compared to conventional system implementation. Also, a case study scenario affecting the evaluation of MPEG4 encoding on two 3-D ICs was provided. The experimental results shown average gains on performance, power and wirelength, as compared to 2-D implementation, about 21%, 19% and 16%, respectively.

REFERENCES


Figure 1: The proposed framework for evaluating 3D ICs.
System-Level Exploration of 3-D Interconnection Schemes

K. Siozios, A. Papanikolaou, A. Bartzas and D. Soudris
Microprocessors and Digital Systems Lab
National Technical University of Athens, Greece
Email: {ksiop, antonis, alexis, dsoudris}@microlab.ntua.gr
WEB: http://www.microlab.ntua.gr

3D integration is the big next step in system integration, but methodologies and tools are required to evaluate the cost and quantify the benefits of this emerging technology

Proposed Methodology

MPEG4 System Implementation @ 3-D IC

Physical implementation of MPEG4 system @ 3-D IC (Cadence Layout)
3D Integrated Smart Antenna Systems

Nakul Haridas and Tughrul Arslan

Institute for Integrated Micro and Nano Systems, School of Engineering and Electronics,
The University of Edinburgh, Edinburgh, Scotland, United Kingdom.

Motivation

The wireless communications market has lately emerged as the most dominating and competitive market, with annual 3G revenues for operators expected to reach US$ 118 billion by 2010. With the developing countries, primarily the Asia-Pacific region now accounting for more than half the mobile subscribes worldwide [2]. Studies by key market analysts have clearly demonstrated that there is a significant need for smart antennas in the wireless communication market. This includes WLAN, WiMAX, UMTS, UWB, CDMA, Bluetooth applications. Smart antenna technology has the capability to extend the range, increase the quality of service, and make more effective use of channel and bandwidth capacity for nearly every wireless communications technology. The WTRS market study published in May 2008 indicates that the smart antenna chipset sales reached US$ 985 million 2008 and is expected to top US$ 2 billion by 2010 [3].

The next generation of such devices will be smaller, low power and will become more intelligent suitable for autonomous environments in which the whole system can adapt to various environmental conditions. Integrating silicon based micro antennas, with MEMS, with reconfigurable system on chip communication architectures along with a variety of signal processing algorithms will be the key to achieving the smart antenna systems that can cater to future demands of the wireless market.

Proposed 3D Integrated Smart Antenna systems

The demands of fabricating cheaper, smaller, and lighter electronic products offering better performance and increased functionalities are continuously growing. The number of electronic devices on a single chip is rapidly increasing, and thus initiating the need of multilevel interconnections. According to International Technology Roadmap for Semiconductors (ITRS) 2005, integrated circuit (IC) chip size will be in the order of 30 nm by 2010. Such a nanosized IC will carry more than 100 million transistors, which will further require more than 100,000 I/Os for next level packaging.

The paper presents a new proposed 3D integration process, illustrated in figure 1, specifically targeting smart antenna systems. A holistic approach is followed towards the design of a fully integrated reconfigurable System-on-Chip (SoC), with Micro Electromechanical Systems (MEMS), in order to provide a multistandard architecture for the antenna or sensor based system. Current work, at the University of Edinburgh, on the development of smart antenna systems has resulted in a number of antenna structures that are designed to work for a wide spectrum of communication bands, as illustrated in figure 2. The application of MEMS to control these antennas, has resulted in tailoring the radiation pattern through reconfiguration. This is illustrated in figures 4 and 5.

Figure 1 3D Integrated Smart Antenna
Herein the application of through silicon vias (TSV) comes into important play as it would form the vital feed for the antenna system. This would eventually connect to the MEMS substrate, as shown in figure 1. The MEMS form the reconfigurable low power switching network, and would connect to a mixed signal substrate. This would be finally connected to a CMOS substrate that will form the reconfigurable System on Chip system. Figure 3, illustrates the antenna with a probe fed via, which would connect to the MEMS substrate through 3D Wafer stacking. Getting all these individual technologies into a single package is task undertaken by this work.

3D wafer stacking represents a wafer level packaging technique wherein specific components, such as logic, memory, sensors, A/D converters, etc., are fabricated on separate wafer platforms and then integrated onto a single wafer-scaled package using through-wafer interconnect (TWI) [1]. 3D multilayered packaging offers a promising solution for building sophisticated chips with benefits of small size and lower cost. The development of 3D integration technologies is mainly motivated by shorter chip-to-chip interconnects, which leads to reduced parasitics, reduced power consumption and higher baud rate compared to conventional lateral chip placement.

References

2. 3G Press release, Asia will Lead Global Growth in Wireless, http://www.3g.co.uk/PR/Jan2003/4709.htm
Abstract: This presentation will give a global view of ST approach regarding 3D integration solutions for VLSI products:

- Motivation and targets
- Contents and results from an active demonstrator
- Need for CAD solution breakthrough, to enable efficient exploration of opened new design space, and secure implementation
- Conclusion and perspectives

A real case study is presented in a separate paper (“3DI Perspective for Multimedia Products”, D Hénoff).

A) Motivations and target:
ST started to consider 3D integration solutions to handle in a comprehensive way the following factors:

- Technical performance: high pin count, electrical features (frequency, power consumption …), thermal issues …
- Miniaturization (X Y Z)
- Cost and Time-to-Market (IP re-use)

ST is having a progressive approach for this disruptive technology:

- capitalization on imaging industrial line (TSV experience and assets)
- driven by actual applications (example: multimedia consumer products)
- starting with a full integration of suitable process bricks, including final packaging
- preparing future generations with step-by-step implementation of more challenging features developed by advanced R&D

Target is cost effective solutions, enabling heterogeneous integration and fitting performances required by ST products.

B) Functional demonstrator and perspectives
A Test Vehicle has been designed, manufactured, packaged and tested. It is a complex 45nm die stacked Face-to-Face on an active interposer designed with a 0.13µm technology. External connections are achieved with TSVs. It includes 1000 inter-die connections, 600 TSVs and 500 bumps connected to a laminate.

Full functionality and compliance with electrical specification is demonstrated.
This demonstrator of the 1st 3DI technology solution has been co-developed by ST and CEA-LETI laboratory in Grenoble, enabling a rapid industrialization phase. The main focus for this first technology generation is to succeed the integration of the overall process, enabling fast product introduction. Further generations that are planned in our roadmap will first reduce the critical dimensions (TSV diameter, inter-die connection pitch), and then enable complex stacks (3 layers or more, several dice on a given layer).

C) **Needed CAD breakthrough**

3D integration technologies are bringing opportunities and challenges for product designers. Advanced tools are required to enable an efficient exploration of the design space: exercise new architectures, split design in different layers, select the most appropriate technologies (silicon dice, 3DI schemes, packaging solution), deal with thermal dissipation issues, take maximum benefits of reduced interconnection lengths, balance technical features and manufacturing cost. CAD solutions are already available to support 3D design up to certain extent. Real 3D design solutions are still to be developed.

D) **Conclusion**

3D integration enables heterogeneous combinations, taking full benefits of advanced nodes (high performance digital blocks, memories …) and mature technologies (analog, RF …). It will be more and more a key enabler for time-to-market, high performance and cost effective products. This disruptive approach is an opportunity but also a challenge for all semiconductor industry areas:

- Technology developments (TSV, interconnection bonding, handling, packaging, test, reliability, failure analysis …)
- Product development (architectures, partitioning …)
- CAD solutions (exploration tools, libraries, Design For Test, Implementation tools)
- Industrial solution (costs, logistics, lead-times …)

Excellence in each area is important, but coherent overall integration is absolutely required to succeed.
**3D Integration Program Overview**

**Why 3D Integration?**
- Device bandwidth: clock frequency on flat line with 2DI
- Heterogeneous integration: RF, analogue, logic, memory, sensors
- Power reduction
- Form factor
- Modularity, IP re-use, Time-to-market
- Cost: 2D scaling reaching its limits (costs, physics, performances)

**ST approach**
- Cost-effective heterogeneous integration, bringing appropriate performances for targeted products

**Active Demonstrator contents**
- 45nm device
- 1056 Face-to-Face intercos: 588 TSV; 482 bumps; in BGA 864 lead free balls, 1.0 mm pitch
- TSV for top and bottom chips access
- Daisy chains & Delay chains with Face-to-Face paths, TSV and RDL paths
- Voltage regulator driving a 45nm IP
- Mechanical stress sensors
- Thermal sensors (top and bottom dice)
- IO boundary-scan

**Product development overall process**
- Product features
  - Silicon techno node data
  - Silicon techno 3D data
  - Packaging techno data
  - Modelling
  - Libraries
  - Implementation Tools
  - DFT Techniques

**Conclusion: a major breakthrough**
- Technology developments
- Product developments (architectures, thermal dissipation ...)
- CAD solutions (IPs, exploration and implementation tools, DFT)
- Supply chain (costs, logistics, lead-times ...)

**Active demonstrator Gallery**
- Whole chain from design up to packaging exercised
- Functionality demonstrated with first samples, end December 2008

**3D system CAD flow overview**

**Summary**
- Whole chain from design up to packaging, including intermediate and final tests, has been exercised in 2008
- First targeted products identified
- CAD flow: currently based on 2D toolbox, partially addressing inter-die interactions
1. Introduction

Future integrated systems will contain billions of transistors [9], composing tens to hundreds of IP cores. These IP cores, implementing emerging complex multimedia and network applications, should be able to deliver rich multimedia and networking services. An efficient communication among these IP cores (e.g., efficient data transfers) can be achieved through utilization of the available resources. An architecture that is able to accommodate such a high number of cores, satisfying the need for communication and data transfers, is the Network-on-Chip (NoC) architecture [2, 5].

Furthermore, the emerging three-dimensional (3D) integration and process technologies allow the design of multi-level Integrated Circuits (ICs). As illustrated in [8], this creates new design opportunities in NoC design. In order to satisfy the demands of emerging systems for scaling, performance and functionality 3D integration is a way to accommodate these demands [3]. For example, a considerable reduction can be achieved in the number and length of global interconnect using three-dimensional integration. On deciding whether to choose a two-dimensional (2D) or 3D NoC as an architecture it is shown in [1, 4] that 3D NoCs are advantageous, providing better performance.

In this work we present an exploration methodology designing alternative 3D NoC architectures. We define as 3D NoCs these architectures composed of many layers, where each layer is a two-dimensional NoC grid, where the grids are the same for all the layers, composed of elements of the same type(s). The main objective of the methodology is to derive heterogeneous 3D NoC topologies with a mix of 2D and 3D routers and vertical link interconnection patterns that perform best to the incoming traffic. Furthermore, the combination of priority-based QoS and buffer sizing techniques is proposed for the first time. The starting point of the proposed methodology is an already optimized mapping [7] which is 32% better than other solutions. In this way additional improvements in latency and energy consumption can be achieved. Moreover, the proposed methodology is applied in computationally intensive applications, e.g. DSP, mapped into 2D and 3D NoC mesh architectures. The cost factors we consider are: a) energy consumption; b) average packet latency and i) total switch block area.

2. Methodology

An overview of the proposed methodology is shown in Figure 1. In order to perform the exploration for alternative topologies of 3D NoC architectures, we have used as a basis the WormSim NoC Simulator [6] that utilizes wormhole switching. As it is shown in Figure 1 now the simulator supports 3D NoC architectures (3D Mesh and 3D Torus) and vertical link interconnection patterns [1]. The 3D architectures to be explored may have a mix of two- and three-dimensional routers, ranging from very few 3D routers to only 3D routers (100% vertical interconnection link presence). The output of the simulation is a log file containing the relevant cost factors we evaluate, such as overall latency, average latency per packet and the energy breakdown of the NoC, providing numbers for link energy consumption, crossbar and router energy consumption etc. From these energy figures we calculate the total energy consumption of the 3D NoCs. In order to steer the exploration we are based on different patterns (as they were presented in [1]). The proposed 3D NoCs can be constructed by placing a number of identical two-dimensional NoCs on individual layers, providing communication by inter-layer vias among vertically adjacent routers. This means that the position of silicon vias is exactly the same for each layer. Hence, the router configuration is extended to the third dimension, while the structure of the individual logic blocks (IP cores) remains unchanged. Moreover, when real applications are used, instead of synthetic traffic, the priority assignment and buffer sizing algorithms are employed (after the application mapping phase) in order to achieve opti-
mizations regarding the latency and the energy consumption of the system.

3. Experimental Results

On top of the synthetic traffic schemes (uniform, transpose and hotspot) three different traffic loads were used (heavy, medium and low). In this way, by altering the packet generation rate it is possible to test the performance of the NoC. The heavy load has 50% increased traffic, whereas the low one has 90% decreased traffic compared to the medium one respectively. The behavior of the NoCs in terms of the average packet latency is shown in Figure 2. In this Figure the latency is normalized using as basis the average packet latency of the full connectivity 3D NoC under medium load and for each traffic scheme. It can be seen the impact of the traffic load (latency increases as the load increases) and that the NoCs can cope with the increased traffic as well as the differences between the different traffic schemes.

![Figure 2. Impact of traffic load on 2D and 3D NoCs (for all different types of traffic used).](image)

Furthermore, we have performed extensive simulations of the behavior of three DSP applications (a) MMS b) MWD and c) VOPD) in two-dimensional and three-dimensional NoCs. The simulation framework employed is based in the one used in [1] and has been extended in order to support priority assignment and buffer sizing (in order to minimize the buffer space in the routers of the NoC). The simulation results of the proposed integrated approach are presented in Figures 3 and 4. The NoC topologies that were tested were 2D and 3D mesh ones of 64 (Figure 3) and 144 nodes (Figure 4) respectively. In these Figures are presented the normalized values of the average packet latency, total energy consumption and the energy x delay product (EDP). The baseline for the normalization is the values of the simulations considering best effort (BE) traffic type and same number of slots (6) for every buffer in NoC for each case.

As can be seen in Figure 3 the EDP is approximately 9% lower in the best case. The integrated approach achieved a reduction in both the average latency as well as the energy consumption and that achieved the desirable behavior. The results for the 144-node NoCs are shown in Figure 4. The maximum reduction of the EDP cost function is 12%. The gains, although

![Figure 3. Experimental results on 64-node 2D and 3D architectures](image)

![Figure 4. Experimental results on 144-node 2D and 3D architectures](image)

it might be considered as not large, they are achieved in addition of an already optimized mapping, thus highlighting the relevance of the proposed methodology.

4. Conclusions

We have presented a framework able to perform topology exploration, priority assignment and buffer sizing for 3D NoC architectures. It can employ different vertical interconnection patterns and evaluate their performance using synthetic or real application traffic.

References

Incorporation of different 3D integration techniques
• Incorporation of buffer sizing and QoS techniques into the exploration phase
• Adaptive routing (power-aware)
• Semi-automated topology generation in RTL

Supported by E.C. funded MOSART IST project, http://www.mosart-project.org
Closed-Form Equations for Through-Silicon Via (TSV) Parasitics in 3-D Integrated Circuits

(Extended Abstract)

Roshan Weerasekera∗†, Dinesh Pamunuwa†, Matt Grange †
†Centre for Microsystems Engineering,
Faculty of Science & Technology,
Lancaster University, Lancaster LA1 4YR, UK.
Email: {r.weerasekera,d.pamunuwa,m.grange}@lancaster.ac.uk

Hannu Tenhunen*, Li-Rong Zheng*
* Department of Electronics, Computer, and Software Systems,
KTH School of Information and Communication Technologies,
ELECTRUM 229, 164 40 Kista, Sweden.
Email: {lirong,hannu}@kth.se

3-D integration using TSVs to interconnect multiple silicon dies in a single chip can offer significant improvements over 2-D Integrated Circuits (ICs) in performance, heterogeneous integration, footprint and integration density [1]. Accurate electrical models of TSV structures are essential in estimating delay, signal integrity (SI) and power integrity (PI) of circuits and interconnects in the design and verification of 3-D ICs. To the authors’ best knowledge, no cohesive DC parasitic parameter models for TSVs in a bundle have been reported in the literature [2], [3], [4]. This paper proposes a set of self-consistent equations for resistance (R), capacitance (C) and inductance (L) of TSVs in a bundle, and presents a reduced-order equivalent circuit including capacitive and inductive coupling. The analytic forms for R, C and L eliminate the need for a computationally expensive field solver and enable efficient delay, SI and PI related analyses early in the design flow.

A 3D-IC is shown in Fig. 1 in which two bulk silicon dies are bonded on top of each other and electrically interconnected using TSVs. These TSVs are assumed to have a uniform circular cross-section. The material commonly used for TSVs is Cu, with an annular dielectric barrier (SiO$_2$ or Si$_3$N$_4$) surrounding the Cu cylinder. The insulating dielectric acts as a barrier to the Cu TSV, preventing the migration of Cu ions into the Si substrate which can degrade device performance by inducing leakage currents. It also electrically isolates the Cu cylinder from the substrate, providing improved isolation to power and ground planes. Further, a thin annular TiN layer is usually deposited between the Cu and SiO$_2$ layers, which acts as an adhesion layer and also concentrates the current in the Cu bar due to its high resistivity [5]. This TiN barrier layer has been neglected for the sake of simplicity and to reduce computational time in the field solver, since its inclusion has an apparently negligible effect on the parasitic parameters.

The general methodology adopted in the modelling is to fit equations to empirical data obtained from a field solver for a range of physical dimensions using analytical forms suggested by physical laws. In a 3-D chip stack, the likely configuration for TSVs is in a regular matrix, for which a representative unit is a 3 × 3 bundle (see Fig. 2). Such a structure has been simulated in a 3-D/2-D quasi-static electromagnetic-field solver specifically used for parasitic extraction of electronic components [6]. It is assumed that the TSV structure and silicon substrate is floating as layout level information describing adjacent ground layers as defined by nearby metal lines distributing power and ground is not present early in the design flow. Also, it is assumed that the substrate is highly resistive [3] as is typical for low noise applications. Further,

This research is supported by European Union research funding under grant FP7-ICT-215030 (ELITE) of the 7th framework programme.
the thickness of the SiO$_2$ barrier $d_b$ is set to be constant equal to 0.2µm as it is fixed for a given technology. The simulated ranges for length ($l_v$), radius ($r_v$), and inter-via spacing ($s_v$) (see Fig. 2) are $20\mu$m $\leq l_v \leq 140\mu$m, $10\mu$m $\leq r_v \leq 45\mu$m and $40\mu$m $\leq s_v \leq 140\mu$m respectively, in corresponding steps of $40\mu$m, $5\mu$m and $20\mu$m. These values are representative of most TSV technologies commonly reported.

Resistance of a TSV can be described using the traditional function of conductivity ($\sigma$) and cross-sectional area:

$$R_{via} = \frac{l_v}{\sigma \pi r_v^2},$$

(1)

The model of (1) is accurate to within 98% of the simulated values.

In a TSV bundle that comprises an $m \times m$ matrix, all self and coupling terms are given by:

$$C_{bundle} = \begin{bmatrix}
C_{i,1} & -C_{1,2} & \cdots & -C_{1,n} \\
-C_{2,1} & C_{2,2} & \cdots & -C_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
-C_{n,1} & -C_{n,2} & \cdots & C_{n,n}
\end{bmatrix},$$

(2)

where $n = m^2$. In (2), the diagonal element $C_{i,i}$ represents the sum of the self ($C_{i,i}$) and inter-via coupling capacitances ($C_{i,j}$) as given in (3):

$$C_{i,i} = C_{i,0} + \sum_{j=1}^{n} C_{i,j}.$$  

(3)

The capacitance matrix is sparse; the main diagonal and adjacent diagonals representing coupling terms to nearest neighbours are populated while the other entries are vanishingly small in comparison (Refer Fig. 3). With reference to the naming convention given in Fig. 2 the distances from M TSV to N, E, S and W TSVs are the same, while the distances to NE, NW, SE and SW TSVs are also equal. Therefore, the capacitance formulae for the total capacitance of M, N, and NE TSVs ($C_{i,i}$), and the coupling terms to their nearest neighbours ($C_{i,j}$) as defined in Fig. 2 are a representative unit for a TSV bundle of any size.

With reference to Fig. 2, the self capacitance $C_s = C_{i,0}$ of a TSV is of the form:

$$C_s = C_{tsv} - k_1 C_{tsv} e^{(k_2 \frac{L_v}{R_v} + k_3 \frac{B}{L_v})} \left[ k_4 \left( \frac{l_v}{r_v} \right)^{k_5} + k_6 \left( \frac{r_v}{l_v} \right)^{k_7} + k_8 \right],$$

(4)

where $C_{tsv}$ is the capacitance of an isolated TSV given by:

$$C_{tsv} = \frac{63.34 \varepsilon_0 d_v}{l_0 \left( 1 + 5.26 \frac{l_v}{r_v} \right)}$$

(5)

and the constants in (4) are defined in the first three rows of Table I. In (4) the constants $k_2$ and $k_3$ are negative, and therefore as $r_v$ approaches infinity, $C_s$ approaches $C_{tsv}$, the capacitance of an isolated TSV given in (5), with a maximum error contained to 6% for the simulated range.

$$\text{Fig. 3: Capacitive coupling between the centre TSV of a 7 } \times \text{ 7 bundle with its surrounding TSVs. Values are normalized to the total capacitance of the center TSV.}$$

The formula for the coupling capacitance ($C_c = C_{i,j}$) terms of (2) for $i \neq j$ (defined in Fig 2) is of the form:

$$C_c = \frac{k_1 \varepsilon_0 d_v}{\ln \left( k_2 \frac{L_v}{r_v} \right)} \left[ 1 + k_3 \left( \frac{p_v}{r_v} \right)^{k_4} + k_5 \left( \frac{l_v}{r_v} \right)^{k_6} + k_7 \left( \frac{r_v}{l_v} \right)^{k_8} \right],$$

(6)

with the constants $k_1, \ldots, k_8$ corresponding to $C_{c,1}, C_{c,2}, \ldots, C_{c,d}$ defined in the last three rows of Table I respectively.

As can be seen in Table I, all models have a minimum accuracy over the full simulated range of approximately 90%. It should be noted that the $C_{s,MT}$ value is valid only when $C_{s,MT} = 0.09$, where $C_{s,MT}$ ($C_{i,i}$) is the total capacitance of the M TSV. Below this range $C_{s,MT}$ values are so small that they are negligible for any meaningful delay, SI or PI analysis.

For those geometries the self capacitance values are in fact negligible for any meaningful delay, SI or PI analysis. Comparisons between the calculated and extracted $C_s$ values for M, N, and NE TSVs for the whole range have maximum absolute errors of 2.3%, 3.6% and 2.9% respectively.

The self ($L_s$) and mutual ($L_{lm}$) inductance terms for a TSV bundle are defined by:

$$L_{bundle} = \begin{bmatrix}
L_{1,1} & L_{1,2} & \cdots & L_{1,n} \\
L_{2,1} & L_{2,2} & \cdots & L_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
L_{n,1} & L_{n,2} & \cdots & L_{n,n}
\end{bmatrix},$$

(7)

where diagonal elements represent the self inductance terms, and off diagonal elements the mutual inductance terms. Inductive coupling is long range and therefore the inductance matrix is well populated, with all elements being non-negligible (See Fig. 4). The self inductance ($L_s$) can be estimated from:
As the intended use of the compact models is to calculate resistive, inductive, and capacitive parasitic parameters of TSV bundles
in 3-D ICs were proposed in this paper. These parasitic models were shown to exhibit fidelity; when the model extracted parasitics were used in circuit simulations, the final error in the metrics of delay and noise amplitude when compared to the same simulation using field solver extracted parasitics was less than the error in the parasitics themselves, showing the usefulness of the proposed models.

REFERENCES


Closed-Form Equations for Through-Silicon Via Parasitics in 3-D Integrated Circuits

Roshan Weerasekera*, Dinesh Pamunuwa*, Matt Grange*, Hannu Tenhunen*, and Li-Rong Zheng*

*Department of Electronics, Computer and Software Systems, KTH School of Information and Communication Technologies, Forum 120, 164 40 Kista, Sweden

Introduction:

- This work proposes a set of self-consistent equations for resistance (R), capacitance (C) and inductance (L) of TSVs in a bundle, and presents a reduced-order equivalent circuit including capacitive and inductive coupling.
- These equations eliminate the need for a computationally expensive field solver and enable efficient delay and SI and PI related analyses early in the design flow.

Physical Modelling:

- Various TSV structures have been simulated in a field solver and their electrical parameters extracted.
- Using curve fitting techniques and analytical insight, semi empirical formulae for capacitance and inductance with varying geometrical and physical parameters are being finalised.

![Physical Dimensions & material properties](image)

- TSVs have Cu filled uniform circular cross-section, and are surrounded by an annular TiN adhesion layer and SiO2 dielectric barrier layer.
- TiN layer has been neglected for the sake of simplicity and to reduce computational time.
- Assumed that TSV structure and silicon substrate is floating, and the substrate is highly resistive.

Resistancenotation Description Simulated Range

r0 TSV radius 10 μm – 40 μm

r1 TSV length or height 20 μm – 140 μm

dth 0.1μm SiO2 dielectric thickness 0.2 μm

dsh separation of two TSVs 20 μm – 200 μm

C copper Conductivity of Copper 106×10^6 S/m

epoi Relative permittivity of SiO2 3.9

epoi Relative permittivity of air 1.0

permeability of air 4.84×10^-7 Tm

permeability of air 1.2566×10^-9 Tm

![Representative unit for a TSV bundle](image)

- Capacitance:
  - The capacitive coupling terms to nearest neighbors dominate, and the coupling terms to nonadjacent lines are mostly insignificant.
  - Within nearest neighbors the lateral terms are more significant than the diagonal terms.

  Self-Capacitance model is of the form:
  \[ C_{a} = \frac{C_{TSV}}{1 - k_{a} r_{1}^{2}} + \frac{1}{1 - k_{a} r_{1}^{2}} \left( \frac{1}{r_{1}} + \frac{1}{r_{2}} + \frac{1}{r_{3}} + \frac{1}{r_{4}} \right) \]

  where TSV is the capacitance of an isolated TSV

  Coupling-Capacitance model is of the form:
  \[ C_{c} = \frac{k_{c} r_{1}^{2}}{\log \left( 1 + \frac{r_{1}}{r_{2}} \right)} \]

  Percentage Error in Capacitive parasitics estimated the order (a)...(f) given in Table

![Percentage Error in Capacitive parasitics estimated the order (a)...(f) given in Table](image)

- Inductance:
  - Inductive coupling is significant within the entire bundle

  Self-Inductance model is of the form:
  \[ L_{a} = 0.16 \mu H \ln \left( 1 + 0.9 \frac{r_{2}}{r_{1}} \right) \]

  Mutual-Inductance model is of the form:
  \[ L_{m} = 0.199 \mu H \ln \left( 1 + 0.438 \frac{r_{2}}{r_{1}} \right) \]

  Percentage error in the predicted self (a) and mutual (b) inductance terms in comparison with the field solver.

![Percentage error in the predicted self (a) and mutual (b) inductance terms in comparison with the field solver](image)

Model Fidelity and Conclusions:

- Absolute errors in Delay and Coupled Noise from recorded maximum errors in each parasitic component are given in the following Table.
- Inductance is not considered as its effect is negligible.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>2%</th>
<th>5%</th>
<th>0%</th>
<th>0%</th>
</tr>
</thead>
<tbody>
<tr>
<td>R_m</td>
<td>2%</td>
<td>5%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>C_m</td>
<td>2%</td>
<td>5%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>L_m</td>
<td>2%</td>
<td>5%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>R_m</td>
<td>2%</td>
<td>5%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>C_m</td>
<td>2%</td>
<td>5%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>L_m</td>
<td>2%</td>
<td>5%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>R_m</td>
<td>2%</td>
<td>5%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>C_m</td>
<td>2%</td>
<td>5%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>L_m</td>
<td>2%</td>
<td>5%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>R_m</td>
<td>2%</td>
<td>5%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>C_m</td>
<td>2%</td>
<td>5%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>L_m</td>
<td>2%</td>
<td>5%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

- In all cases, the error is less than the error in the parasitic values predicted by the models.
- The errors in delay and noise from the worst-case combination of errors in parasitic values are contained within 10% and 14% respectively.
Modular Modeling of RF Behavior of Interconnect Structures in 3D Integration

Jörn Stolle, Sven Reitz, Peter Schneider, Andreas Wilde
Fraunhofer Institute for Integrated Circuits, Division Design Automation
Dresden, Saxony, Germany
Email: joern.stolle@eas.iis.fraunhofer.de, Tel: +49 (351) 4640 789

Keywords - 3D integration, RF behavior, interconnect structures, interchip vias, modular modeling, embedding

1 Abstract

Currently, 3D integration is still characterized by stacking of similar dies (e.g. memories or cpus) which are mostly connected to each other at their outer boundaries using bond wires. Nevertheless, first steps are done towards the combination of different functional units (sensor, micro controller, radio ...) which are usually not designed for 3D integration. The dies are adapted to current packaging technologies, like e.g. TQFP, concerning mounting, location and size of pads … [1]. Interposers containing rewiring and interchip vias (ICVs) are often used to be able to connect these units together [2].

The future potential will lay in real 3D systems. There we will encounter short signal paths enabling among others performing very efficient signal processing for sensor array structures, parallel computing … But, due to dense integration of the 3D system a lot of physical interactions within the 3D systems exist.

2 Introduction

3D integration offers a variety of capabilities for new system concepts and form factors. Especially the influence of integration technology on system function has to be considered preferably early in the design process. Different materials and manufacturing technologies cause different properties of layers, which are located closely on top of each other, and there are a lot of degrees of freedom for the system architecture [3].

Due to the dense interconnections between the stacked layers parasitic effects play a more and more important role. The electrical behavior of the interconnect structures is e.g. characterized by skin and proximity effect by reason of their decreasing geometrical dimensions and increasing operating frequencies which may influence the electrical function of the entire 3D system [4]. Signal integrity, cross talk and interconnect delays have also to be considered.

To enable the electrical function of the system place and route in 3D have to be performed [5]. The RF behavior of the interconnection structures has to be taken into account during these steps to ensure the overall system behavior. Additionally, the placing of the functional blocks has to respect their thermal power loss to reduce or avoid hot spots [6] but this topic is not within the scope of this paper. A real 3D design from scratch is a multi-criteria optimization task with a huge amount of design parameters.

To derive the RF behavior of the variety of structures efficient modeling approaches and simulation algorithms are necessary as key methods [7]. Therefore, an appropriate methodology is required for multilevel and multiphysics analysis of interconnect structures as well as entire stacked 3D systems.

3 Modular modeling approach

The main idea is to create a generic hierarchical modeling approach for multiphysics problems in 3D systems to covering support of technology development, analysis of particular interconnect structures and investigation of the entire stacked system [8].

The first step is to generate parametric models of basic structures, which can be modified concerning geometrical and material properties as well as the represented physical domain (thermal, electrical, mechanical, …). In this paper we deal with the electrical domain. These parametric models, so-called basic modules, can be stored in libraries and can be used to build up models of more complex structures.

The modular approach allows on one hand detailed analysis of the behavior of single basic modules as function of the parameters, and on the other hand, investigation of the built-up more complex 3D structures.

Complementary, network and behavioral models for electrical simulations are provided for these basic modules as well as for the composed structures as SPICE models or using modeling languages like VHDL-AMS and Verilog-AMS. These parametric electromagnetic behavioral models can be combined with the electrical model of the entire system to enable the complete system simulation considering parasitic effects.

4 Application

The investigation of the electromagnetic behavior is illustrated using the example of an interconnect structure in the RF domain considering wave propagation as well as skin and proximity effects which may influence the electrical function of the entire 3D system.

Fig. 1 shows the RF model for an interconnect structure
formed by three ICVs, a transmission line and again three ICVs.

Fig. 1 CST Microwave Studio model of the interconnect structure

Two different approaches will illustrate the possibilities to calculate the electrical behavior of interconnect structures.

The first opportunity is to solve the Maxwell equations of the entire structure in a PDE solver, e.g. the full wave simulation tool Microwave Studio (MWS) from CST. This needs a lot of simulation time and system resources.

Second, a modular modeling approach can be exerted. The single structures are extracted from the overall interconnect structure and afterwards simulated in MWS. A reduction of computing time and resources as well as the generation of basic modules are the result. These basic modules are stored in a library and composed to the more complex structure via embedding techniques. The comparison of the computed transmission behavior of the entire and the embedded structure is shown in Fig. 2.

The modular modeling approach including embedding techniques as well as the parameterization of equivalent network models for SPICE will be discussed in the full paper.

5 Conclusions

The main goal is design support for 3D systems. Therefore, the adaptation of design flows and the integration of results within these design flows are important tasks. Depending on the design task, models on different levels of abstraction can be used. Our modular modeling approach covers detailed simulations of electromagnetic effects using PDE solvers. For analog and mixed signal systems these are usually SPICE net lists or behavioral models programmed in VHDL-AMS or Verilog-AMS. Digital system design is supported by models for crosstalk and signal-dependent delays which are derived from more detailed models and analysis.

Furthermore, electromagnetic behavioral models can be generated from the entire system to combine these models with models of semiconductor devices to investigate the RF behavior of the whole system.

Acknowledgment

This paper is partly based on the project e-CUBES which is supported by the European Commission under support-no. IST-026461. The authors of this paper are solely responsible for its content.

References


Brief Biography

Jörn Stolle studied Electrical Engineering at Dresden University of Technology from 2000-2007. Since 2007 he is with the Fraunhofer Institute for Integrated Circuits, Design Automation Division, Dresden. His main research interests include RF simulation, physical modeling and optimization.
Modular Modeling of RF Behavior of Interconnect Structures in 3D Integration

Jörn Stolle, Sven Reitz, Peter Schneider, Andreas Wilde
Fraunhofer Institute for Integrated Circuits IIS, Design Automation Division EAS, Dresden, Saxony, Germany

Goals of Modeling and Simulation
– Flexible modeling, simple modification of material and geometry parameters
– Hierarchical modeling approach for multi-physics problems and design support for 3D systems
– Support and verification of design decisions
– Input for design methods, e.g. Floorplanning, Place&Route

Important for electrical / RF behavior
– Investigation of frequency dependence
– Ampacity per via or area

Modular modeling approach
– Extract basic geometrical models of the real structure
– Detailed simulations of electromagnetic effects using PDE solver (e.g. CST Microwave Studio)
– The modular approach allows detailed analyses of the behavior of single basic modules and also the investigation of the built-up more complex 3D structures
– Generate network and behavioral models using languages like SPICE or VHDL-AMS for system level design

Embedding of interconnect structures

Basic steps
– Generate parametric models of basic structures
– Transform the s-parameters (electrical behavior) of the basic modules to t-parameters
– Multiply the t-parameters and re-transform to s-parameters to get the electrical behavior of the complete complex structure

Derivation of network models

The presented work is partly based on
– the integrated project e-Cubes which is supported by the European Commission under support-no. IST-026461 and
– the national project VSI which was supported by the German Bundesministerium für Bildung und Forschung, support-no. 01M 2999 A.

Contact
Fraunhofer Institute for Integrated Circuits IIS, Design Automation Division EAS, Dresden
Zeunerstraße 38
01069 Dresden
Germany
Joern Stolle
Phone +49 351 4640789
Email Joern.Stolle@eas.iis.fraunhofer.de
Technology for 3D ICs is slowly becoming mainstream such that integrating complete mixed digital-analog-RF systems is not only possible but almost unavoidable. Because of the close proximity of the different substrates switching noise from digital circuits are easily coupled through to severely degrade the performance of sensitive analog/RF circuits, as illustrated in Fig 1. In this paper, we evaluate the noise coupling problem for mixed digital-RF circuits and propose techniques to reduce this problem. In particular, we evaluate the efficacy of using separate tiers for digital and analog/RF circuits and the use of Faraday cages to shield sensitive circuit blocks. Verification is done using a combination of EM simulation and measurements from fabricated test structures on the 0.18µm fully-depleted silicon-on-insulator (FDSOI) technology[1] shown in Fig 2.

Fig 3 shows the microphotograph for the 5mm \times 5mm 3D IC test chip. Spiral inductor test structures without Faraday cages in tier 3 are used as the reference to evaluate the designs with Faraday cages (Fig 3b) on the same tier (tier 3). For the first experiment noise is injected (transmitter) and measured (receiver) with the use of identical inductors. The transmitter consisted of four inductors at different distances from the receiver. The distances between transmitter inductor centers and receiver inductor centers are 710µm, 510µm, 435µm and 430µm. For this design, the Faraday cage consists of the back metal (BM1) on tier 3, the back metal (BM1) on tier 2 and intertier via between tier 3 and tier 2. The simple diagram for the solid plate Faraday cage and the inductor is shown in Fig. 4 (a). One port of the inductor is connected to the test pad through the external port connection. The other port of the inductor is connected to the ground via. De-embedding structures shown in Fig 3(c) were used to eliminate the influence of the transition region between the probe, probe contact and the device under test. The measurement results shown here are after de-embedding. Ansoft HFSS was used for our electromagnetic (EM) simulations [2].

For the second set of experiments the noise coupling problem for mixed digital-RF circuits was directly evaluated by including multiple FPGA tiles in combination with a LNA on the test chip. The LNA is usually the first active stage in RF receivers and amplifies the input signal from the antenna and suppresses noise contributions from subsequent stages. Therefore, the LNA usually sets the noise performance for the entire RF system. The LNA design, shown in Fig. 4b, includes several gate (g_n) in the common gate circuit [3]. For an input impedance equal to 1/g_n the noise figure is given by \( F_{CG-LNA} = 1 + \gamma/(2\alpha) \), where the additional factor of 1/2 is a result of the g_n-boosted technique. The LNA design uses an external balun to transform the single ended antenna signal to differential signals. We utilize its presence to provide a DC ground and a RF signal path for the LNA. An on-chip balun-transformer is possible but was omitted for design flexibility. The Faraday cage designs for the LNA are similar to those designed for the inductors. The test structures for LNAs with/without Faraday cages in tier 3 are shown in Fig. 3(d) and (e). The noise generators for this test are digital circuits (FPGA tiles & ring oscillators) are included on different tiers (tier 2, tier 3). A simple one stage design was selected for this set of experiments due to the limited a priori RF model knowledge prior to chip fabrication.

Fig 5 shows measured and EM simulation results for the reference structures in Fig 3 (a). Measurements were done from 50MHz to 1GHz with the help of an Agilent 8719Es two-port network analyzer. The graph plots the measured S_{21} with the probes in the air and measured and simulated values of S_{21} at different separations between the two structures. The separation between the structures is varied from 430µm to 710µm. We note a good matching between measurement and simulations. It should be noted that increasing the distance between the transmitter and receiver by about 200µm reduces the crosstalk by only 5dB at 1GHz.

Measurement and EM simulation results for Faraday cage enclosed receiver and transmitter shown in Fig. 3 (b) indicated that the metal external port connection outside the Faraday cage was the primary contributor to the increase in crosstalk with frequency. We were unable to eliminate the connection to the outside for our prototype, however, we were able to validate close agreement between measurement and simulation results, as shown in Fig. 5. We therefore have used our previously calibrated EM simulations to show the impact of these cages on noise coupling. In our experiments we evaluated the impact of using i) one vs. two Faraday cages (for noise generator and sensitive circuit) and ii) separating sensitive and noisy circuits using different tiers. Our simulation results, shown in Fig. 6 show that using different tiers increases the isolation by 10dB while using even one Faraday cage improves isolation by over 75dB. Therefore, to limit area overhead only sensitive analog/RF circuits need be enclosed in Faraday cages.

We are currently in the process of lab testing the LNA performance with and without Faraday cages while varying the noise source to different tiers. Measured LNA results will be presented at the workshop. Simulated S-parameter results for our LNA design are shown in Fig. 7. The input reflection coefficient (S_{11}) is less than -14dB and the maximum power gain is approximately 10 dB. The rest of the performance values are listed in Table I. Our early results from inductors are promising and show that Faraday cages have the potential to provide very good noise isolation in fully integrated 3D ICs. We conjecture that LNA tests will further enforce this conclusion showing that 3D ICs provide the potential for compact high performance mixed-signal circuits, where sensitive analog/RF circuits can be in close proximity to DSP circuits.

REFERENCES

### Table of Contents

1. **Noise Generation Circuits**
   - Digital Circuits
   - Switching Regulator
   - Analog Circuits
   - RF IC Blocks

2. **Sensitive Circuits**
   - Inductor DR
   - Capacitor DR
   - Synthesizer

---

**Fig. 1:** Typical noise generators and sensitive circuits in mixed-signal ICs

**Fig. 2:** The 3-tier 3D IC cross-section used for prototype designs [1]

**Fig. 3:** Micrograph of test die (a) inductors without Faraday cage on tier 3 (b) inductors with Faraday cage on tier 3 (c) de-embedding structures (d) LNA without Faraday cage on tier 3 (e) LNA with Faraday cage on tier 3 (f) digital circuits

**Fig. 4:** (a) Solid plate Faraday cage and (b) LNA circuit schematic

**Fig. 5:** Measurement and simulation results without Faraday cage [Fig. 3(a)]

**Fig. 6:** Simulation results for different structures (d= 430µm)

**Fig. 7:** Simulated S-parameters for wideband inductorless LNA

---

### Table I: Summary of simulation results for single-stage wideband LNA

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>0.18µm</td>
</tr>
<tr>
<td>$S_{11}$</td>
<td>$-14dB$</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>1 – 6GHz</td>
</tr>
<tr>
<td>NF</td>
<td>&lt;3dB</td>
</tr>
<tr>
<td>IIP3</td>
<td>+5dBm</td>
</tr>
<tr>
<td>Power gain ($S_{21}$)</td>
<td>10dB</td>
</tr>
<tr>
<td>Power dissipation</td>
<td>5.4mW</td>
</tr>
</tbody>
</table>
Evaluating Noise Coupling Issues In Mixed Signal 3D

Liuchun Cai and Ramesh Harjani, University of Minnesota, Minneapolis, MN

Outline

- Introduction: Motivations for 3D mixed-signal circuits
- Motivations for 3D
- Mixed-Signal Systems
  - Simulation and measurement results
  - Conclusions

Motivations for 3D

- Mixed signal circuits have become pervasive in ultra-performance computing, storage, networking, and imaging.
- The increased complexity of mixed-signal systems and increasing interconnects between chips and dies has created noise issues.
- 3D integration to mitigate noise coupling issues.
- Advantages of 3D:
  - Reduced wire lengths between dies to improve performance and reduce crosstalk.
  - Improved power delivery and thermal management.
- Critical problems with 3D mixed-signal:
  - Noise coupling from digital circuits can degrade analog performance.
- Noise coupling from 3D mixed-signal devices:
  - Can be reduced through wafer bonding, alignment, and thinning.
- Resulting in increased power and area due to alignment steps.
- Large-scale wafer bonding can cause warpage and alignment issues.
- Great path-length matching due to shorter critical wire lengths.
- Good path-length matching due to shorter critical wire lengths.
- Low noise amplifier (LNA)

Experimental Setup: Test bed

- Test bed to verify isolation.
- Isolation provided by separate tiers.
- Isolation provided by Faraday cage.

Experimental Setup: Faraday Cages and LNA

- Faraday cage with solid top and bottom layers.
- LNA circuits with multiple FPGA dies.
- De-embedding structures and FPGA: (c) and (f)

3D Fabrication Steps (I)

- Wafer bonding
- Die bonding
- Die bonding using wafer bonding technology.
- Bond metal layer in tier 3 is formatted and etched bond pads.
- Results and discussions: With Faraday cage.

Results and discussions: No Faraday cage

- Separation between structures varies from 430 µm to 710 µm.
- Bonding and alignment between measurements and simulations.
- Increasing bond pad from 200 µm to 260 µm results in almost 90% reduction in overall crosstalk.

Results and discussions: With Faraday cage

- Changes in frequency do not change crosstalk explicitly.
- Changes in frequency do not change crosstalk explicitly.
- Changes in frequency do not change crosstalk explicitly.

2D RF Coupling Problem

- Coupling problem at high frequencies.
  - digitally less than 1dB.
  - coupling below 100 µm at frequencies above 1GHz.

3D Fabrication Steps (II)

- Ansoft HFSS was used for EM simulations.
- Simulation structures correspond to the test structures.
- Simulation structures correspond to the test structures.

Mitigating noise coupling issues

- 3D layer description
- Simulation and measurement results
- Conclusions

Advantages of 3D for mixed-signal/RF circuits

- Mixed-signal circuits
  - Multiple functions on a single die.
  - Increased performance and reduced area.

Experimental Setup: Die Photo

- Solid plate Faraday Cage.

Electromagnetic Simulation Setup

- Ansoft HFSS was used for EM simulation.
- Simulation structures correspond to the test structures.
- Simulation structures correspond to the test structures.

Results and discussions: Impact of tiers

- LNA performance within Faraday cages while varying the noise source to different layers.
- The simulated results for LNA are listed as following.

Results and discussions: LNA

- LNA sees a broadband g×1inverted common gate-circuit.
- LNA performance within Faraday cages while varying the noise source to different layers.
- The simulated results for LNA are listed as following.

Conclusions

- Developed techniques for crosstalk reduction in 3D ICs.
  - Using Faraday cages.
- Measurement and simulation results show:
  - One Faraday cage increases isolation by 70 dB.
  - Two Faraday cages, increases isolation by 120 dB.
- 3D ICs with Faraday cages may be the preferred technology for high performance mixed signal systems.
- 3D ICs with Faraday cages may be the preferred technology for high performance mixed signal systems.
- Future systems have small area - large digital.
- Isolation of digital noise critical for high performance systems.
- Faraday cages are possible in 2D technologies.
- Faraday cages not possible in 2D technologies.
- 3D good for mixed-signal systems.
Design, Verification and Simulation of Three Dimensional (3D) Circuits

Guruprasad Katti1, Bart De Wachter, Marc Nelis, Morin Dehan, Miroslav Cupak, Kris Croes, Gerd Beeckman, Pol Marchal, Michele Stucchi, Kristin De Meyer1, Wim Dehaene1

IMEC, Kapeldreef 75, Leuven, B-3001, Belgium, email: guruprasad.katti@imec.be,

1EE Department, K.U. Leuven, Kasteelpark Arenberg 10, Leuven, B-3001, Belgium

Abstract—3D ICs provide a promising option to keep pace with Moore’s Law. Along with the technological efforts to realize reliable 3D ICs, EDA/CAD efforts leading to a robust 3D Design environment are essential to design, test and verification of high performance 3D circuits. This paper discusses the implementation of 3D design environment leveraging the existing 2D design platforms. Verification methodology constituting 3D implementation of 3D design environment leveraging the existing performance 3D circuits. This paper discusses the environment are essential to design, test and verification of high reliable 3D ICs, EDA/CAD efforts leading to a robust 3D Design

While the technology efforts are key to realizing reliable 3D-SIC, the design efforts require a robust 3D Design Environments [1] (3DDE). The initial steps building 3DDE are described in this paper. Although industry currently lacks 3D design tools, existing 2D EDA tools can be leveraged to enable the design, test and verification of 3D-SIC. In this paper, existing popular Full Custom Design Framework II [2] (DFII) is enabled to tape-out 3D ICs. Design verification in the form of Design Rule Checks (DRC), Layout Versus Schematic (LVS) are accomplished with the help of upgraded Calibre [3] rule decks. Spectre [4] simulations on the extracted circuit netlist generated using the Analog Design Environment (ADE) are used to test the functionality of the designed 3D circuits.

Section 2 illustrates the methodology to translate existing 2D platforms to design and layout 3D circuits in DFII. A case study of a 2D/3D RO circuit encompassing the entire design flow is illustrated in section 3. Section 4 concludes the paper.

I. INTRODUCTION

Performance of deep-sub micrometer Very Large Scale Integrated (VLSI) circuits is being increasingly dominated by interconnects due to decreasing wire pitch and increasing die size [1]. Additionally, heterogeneous integration of different technologies in one single chip is becoming increasingly desirable [2-3], for which planar (2D) ICs may not be suitable. 3D Stacked Integrated Circuit (3D-SIC) technology stack various 2D active device layers in the vertical direction achieving better performance in addition to minimal form factor. These vertical layers are then interconnected by using the Through Silicon Via (TSV) as shown in Fig. 1.

While the technology efforts are key to realizing reliable 3D-SIC, the design efforts require a robust 3D Design Environments [1] (3DDE). The initial steps building 3DDE are described in this paper. Although industry currently lacks 3D design tools, existing 2D EDA tools can be leveraged to enable the design, test and verification of 3D-SIC. In this paper, existing popular Full Custom Design Framework II [2] (DFII) is enabled to tape-out 3D ICs. Design verification in the form of Design Rule Checks (DRC), Layout Versus Schematic (LVS) are accomplished with the help of upgraded Calibre [3] rule decks. Spectre [4] simulations on the extracted circuit netlist generated using the Analog Design Environment (ADE) are used to test the functionality of the designed 3D circuits.

Section 2 illustrates the methodology to translate existing 2D platforms to design and layout 3D circuits in DFII. A case study of a 2D/3D RO circuit encompassing the entire design flow is illustrated in section 3. Section 4 concludes the paper.

II. 3D DESIGN ENVIRONMENT

Although industry lacks 3D chip design tools, popular 2D design tools can be updated to enable 3D circuit design. Existing 2D design flow using industry standard 2D tools along with key support files is shown in Fig. 2. The DFII platform supports the creation of schematic/netlist and layout entry. The DRC errors are identified with the help of Calibre rule decks while LVS rule decks confirm the sanity of the layout w.r.t. the schematic/netlist. Spectre or other SPICE simulators integrated in the ADE environment are used to test the functionality of the designs. Technology rule file (i.e. 2D.tf) consists of the technology layer information necessary to fabricate an IC. Display file (display.drf) supports viewing and layout of these layers. Calibre DRC and LVS rules are instrumental in identifying and correcting DRC and LVS errors. The schematic/netlist can be simulated with the help of SPICE using relevant device model files.

The first step towards upgrading the existing 2D design environment to 3D is to change the technology file 2D.tf to 3D.tf such that all the existing 2D FEOL and BEOL layers are tagged with the purpose "T" and data type 0 for the top die and “B” and data type 1 for the bottom die in case the 3D stacking is limited to 2 dies. Also, a new TSV layer is added as via that connects Nth metal layer on the bottom die and 1st metal layer on the top die [6]. Stream out numbers for all top and bottom
layers remain the same while a suitable number is assigned to the TSV layer. To aid 3D layout, the display file *display.drf* is updated such that the borders of all the top blocks are viewed as a solid line while the bottom layout layers appear as dotted lines. Resulting layer map in LSW and the top and bottom 2D inverter layouts is shown in Fig. 3. With this infrastructure, designers can layout 3D circuits as seamlessly as they are used to layout 2D circuits.

Next step is to upgrade the 2D DRC rules to 3D DRC rules. With the 3D DRC rules, the design rules are checked for the top and bottom dies one after the other. Design rules pertaining to the TSV width, spacing and FEOL keep out areas are also added to the rule decks. For the IMEC process [6] with TSV width = 5µm and pitch = 10µm, FEOL keep out spacing is conservatively chosen to be 7.5µm away from the center of the TSV. A standard pCell of the TSV is used in the layout to avoid errors concerning the size of the TSV.

For the LVS purpose, TSV is treated as via connecting the N\textsuperscript{th} metal on the bottom die to the 1\textsuperscript{st} metal on top die. With this scheme the devices on the top and bottom dies are extracted separately and the devices on the top and bottom dies are connected through TSV. 2D and 3D CMOS circuits are then extracted using Calibre and simulated using ADE integrated in DFI with the help of existing platform MOS device models.

### III. RING OSCILLATOR – A CASE STUDY

RO is a basic digital circuit to analyze a particular technology. 41 stage 2D/3D RO circuits with varying configurations as shown in Fig. 4 were designed with the help of the 3D design environment. All ROs are followed by an 8-stage frequency divider (divide by 256) circuit such that the resulting waveform frequency is in the 1-10MHz range enabling measurements. DRC errors generated during the layout creation are shown in Fig. 5. It can be seen that the DRC rules can be modified such that the errors on the top die are concatenated with “_0” while errors on the bottom die are concatenated with “_1” to ease the process of DRC cleaning. Moreover, an automatic script is created to translate existing 2D DRC/LVS and LPE platform rules to respective 3D rules.

For the LVS purpose, TSV is treated as via connecting the N\textsuperscript{th} metal on the bottom die to the 1\textsuperscript{st} metal on top die. With this scheme the devices on the top and bottom dies are extracted separately and the devices on the top and bottom dies are connected though TSV. 2D and 3D CMOS circuits are then extracted using Calibre and simulated using ADE integrated in DFI with the help of existing platform MOS device models.

Once the 3D DRC is clean and LVS is found to be in sync with the schematic, spectre simulations on the extracted layouts are performed for varying RO configurations. Output waveforms for V\textsubscript{dd} = 1.2V for various RO configurations are shown in Fig. 6 while the table 1 shows RO delay and power for various RO configurations.
It can be seen that the RO delay and power for all RO configuration are same. This is because TSV is treated as a short wire connecting the Nth metal on the bottom die and METAL1 on top die during the 3D circuit extraction. Parasitic RC elements of the TSV are ignored. Next step is to perform parasitic TSV extraction using caliper and accomplish 2D/3D circuit simulations using parasitic TSV model. TSV parasitic extraction will be essential not only to analyze the effect of varying TSV architectures on 3D circuit performance but also to design optimal 3D circuits.

### Table 1. RO Delay and Power simulation results

<table>
<thead>
<tr>
<th>RO Configuration</th>
<th>RO Delay [ns]</th>
<th>RO Frequency [MHz]</th>
<th>RO Power [mW]</th>
</tr>
</thead>
<tbody>
<tr>
<td>RO_T</td>
<td>242.034</td>
<td>4.131</td>
<td>1.604</td>
</tr>
<tr>
<td>RO_B</td>
<td>242.034</td>
<td>4.131</td>
<td>1.604</td>
</tr>
<tr>
<td>RO_HBHT</td>
<td>242.034</td>
<td>4.131</td>
<td>1.604</td>
</tr>
<tr>
<td>RO_ATOB</td>
<td>242.034</td>
<td>4.131</td>
<td>1.604</td>
</tr>
<tr>
<td>RO_ABOT</td>
<td>242.034</td>
<td>4.131</td>
<td>1.604</td>
</tr>
<tr>
<td>RO_Cascaded</td>
<td>242.034</td>
<td>4.131</td>
<td>1.604</td>
</tr>
</tbody>
</table>

### IV. SUMMARY & CONCLUSIONS

Existing 2D design tools DFII and Calibre are successfully leveraged to develop 3D circuits. An entire design flow from schematic/netlist entry to test and verification of small circuit blocks is demonstrated with the help of a 2D/3D Ring Oscillator case study. As the 3D design environment created uses a set of popular 2D design tools, the 3D design environment would seamlessly aid design engineers sign-off and tape-out trustworthy 3D circuits. The methodology can as well be extended to stack more than two dies to build multilayer 3D ICs.

### REFERENCES


Examination of Delay and Signal Integrity Metrics in Through Silicon Vias

Matt Grange, Roshan Weerasekera, Dinesh Pamunuwa
Centre for Microsystems Engineering, Faculty of Science and Technology, Lancaster University, Lancaster, LA1 4YR, United Kingdom
{m.grange, r.weersekera, d.pamunuwa}@lancaster.ac.uk

Hannu Tenhunen
Department of Electronics, Computer, and Software Systems, KTH School of Information and Communication Technologies, ELECTRUM 299, 164 40 Kista, Sweden
Email: hannu@kth.se

Abstract—This article discusses results from simulations of signaling in Through Silicon Vias (TSVs) with an emphasis on latency and signal integrity effects. Data from field solver simulations is used for TSV parasitics and employed in SPICE simulations. A reduced order electrical circuit is proposed for lone TSVs as well as bundled structures and switch-factor based delay models are derived to calculate rise times in a 3x3 bundle. Furthermore Signal Integrity (SI) issues in coupled TSVs are briefly discussed.

I. INTRODUCTION

3-D integration is a promising new technology that is currently being developed by a number of organizations worldwide. Circuits are stacked at the wafer, die, chip or package level to decrease the footprint, increase speeds, and allow for integration of various technologies on one chip, among other potential improvements [1]. One viable solution to provide connectivity between the layers is the Through Silicon Via (TSV). Holes are etched in the silicon and filled with copper surrounded by an adhesive layer and a dielectric barrier. TSVs can be implemented down to a diameter of less than 5µm and densities are currently increasing up to 1000 I/Os per device [2]. As limitations on wire bonding interconnects are becoming apparent in recent chip designs with increasing numbers of layers, the TSV provides an enticing alternative to 3-D stacking interconnects.

As the use of TSVs is a fairly recent concept, their effects on signaling within a 3-D circuit are not well documented. It is the goal of this paper to provide a preliminary assessment of the effect a TSV has on signal integrity within a realistic context. Using circuit models derived from field solver simulations [3], the parasitic effects of TSVs in various configurations and sizes can be simulated within a SPICE engine. As 3-D integration is coming closer to the implementation phase, it is necessary for chip designers to better understand signal characteristics in TSVs to expedite the design process.

Explorations into signal effects caused by TSV parasitics are not widely covered in literature. There are several examples of TSV explorations such as an investigation into propagation delay in [4] and a further examination of parasitic modeling [5], [6], but to the authors’ knowledge, comprehensive simulations for TSVs in a bundle have not been published and useful models for calculations of delay, SI and Power Integrity (PI) have not been widely released.

The organization of this paper is as follows: In section II, the trends associated with lone and bundled TSV parasitics are described and an appropriate equivalent electrical circuit is discussed. Next, the parasitic properties of the TSV are tested for their significance and a reduced order model is proposed. Section III examines the effects of parasitic crosstalk between TSVs in a 3x3 bundle and switch-factor based delay models that capture the effect of crosstalk on signal latency are presented. Section IV discusses Signal Integrity issues in a bundle, such as the effect of using grounded TSVs as shields, and finally Section V ends with our conclusion.

II. TSV PARASITICS

In order to investigate electrical properties for TSVs within a bundle, field solver simulations were carried out to determine parasitic and coupling values for realistic geometrical configurations. The equivalent electrical circuit diagram for a TSV is as shown in Figure 1. This is a conventional T-model wire segment including parasitic resistance, inductance and capacitance to ground as well as capacitive and inductive coupling to a neighboring TSV.

Figure 1. TSV Electrical Equivalent Circuit

The parasitics were extracted for TSV structures with radii of 5-40µm and lengths of 20-140µm. Mutual capacitance and inductance values were extracted for the same geometrical structures, with TSV pitches ranging from 50µm to 260µm in a 3x3 bundle. As with on-chip interconnects, the trends for wire capacitance, inductance and resistance are largely the same for TSVs. As the length of the via increases, resistance, capacitance and inductance also increases. As the width increases, the capacitance goes up but the inductance and resistance decrease. The capacitive parasitic terms in different sized bundles of TSVs are visible in the plots shown in Figures 2, 3 and 4.
In all simulations displayed in this paper employing the circuit diagram in Figure 1, the TSV model was driven by an inverter size of 10 and loaded by a minimum sized inverter in 0.35µm technology. The driver was found to be the optimum size, given the TSV parasitics, by a series of sweeps and the minimum-sized inverter represents the pin load for each vertical interconnect. A 50ps rise time was employed throughout all of the simulations.

A. Significance of RLC Parasitics

Simulations were performed for the entire range of resistance, inductance and capacitance values as determined by the field solver. The resistance was swept from 0-500mΩ, and the output waveforms were plotted to observe variations in delay. As can be seen in Figure 5, the TSV resistance within the considered range of is so small that it has no observable effect on the output waveform. The inductance was then swept from 0-500pH, the extracted range, for rise times down to 1ps, revealing no significant contribution as seen in Figure 6. Finally, the capacitance was swept from 0-500fF showing a significant effect on latency of the output waveform, as seen in Figure 7. These results appear to show that the electrical model for a cylindrical TSV can be reduced to a purely capacitive model. The resistive and inductive parasitics are small enough to be neglected in any delay simulations, which reduces the complexity of the electrical model significantly.

B. Distributed vs lumped model

In addition to parametric sweeps, simulations were conducted to determine if distributed models were necessary to attain accurate results. The model was segmented into 2, 5, and 10 sections and output waveforms examined to show no significant effect on the signal from increasing the number of segments within the parasitic range determined by the field solver. The relatively low resistive and inductive terms reduce the necessity for a distributed model for simulation of signals within the considered range.

III. Crosstalk in a 3x3 TSV bundle

This section investigates the effects of crosstalk between TSVs organized in a 3x3 bundle. By employing electrical models derived from the field solver simulations, various switching patterns are simulated to analyze crosstalk effects between these structures. The coupling capacitance between two TSVs is a function of radius, length and inter-via spacing, as well as dielectric barrier thickness and permittivity, and increases monotonically with increasing radius and decreasing spacing. As capacitive and inductive coupling can have detrimental effects on bandwidth and signal integrity, the crosstalk between adjacent structures must be examined to determine the most efficient use of area and TSV sizing to maximize signal throughput and reliability.

As in the on-chip case, the capacitive coupling terms to nearest neighbors dominate over the coupling terms to non-adjacent lines, which are mostly insignificant. Within the set of nearest neighbors the lateral terms are more significant than the diagonal terms. This is observable in Figure 8 and is due to the fact that the diagonal neighbors are partly shielded by the lateral conductors and the non-adjacent lines are almost completely shielded by the ring of adjacent lines.
determined switch factors for the various switching patterns are defined in Table 1. Here $C_s$ is the self capacitance, $C_l$ the lateral coupling capacitance, $C_d$ the diagonal coupling capacitance and $R_0$ the driver resistance.

$$t_d = 0.69R_D(C_s + K_1C_l + K_2C_d) \quad (1)$$

<table>
<thead>
<tr>
<th>Switching Pattern</th>
<th>$K_1$</th>
<th>$K_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Victim</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lateral</td>
<td>3.4</td>
<td>5.2</td>
</tr>
<tr>
<td>Diagonal</td>
<td>9.0</td>
<td>10.6</td>
</tr>
</tbody>
</table>

Table 1. Switch Factors for Delay Estimation

The minimum accuracy of this equation over the entire range was greater than 92%, principally due to the negligible parasitic resistance inherent in the TSV. These switch-factor-based delay equations facilitate the integration of TSV interconnects into established on-chip static timing methodology.

IV. SIGNAL INTEGRITY SIMULATIONS

In order to fully capture the effect of crosstalk on delay and coupled noise amplitude under real-world conditions, simulations were carried out with pseudo-random bit streams (PRBS) at the victim and aggressor inputs in a 3x3 bundle to generate the eye diagrams at the output. All drivers were size 10 inverters while every TSV was loaded with a minimum-sized inverter. The example geometry chosen was a bundle with radii, length and pitch of 15µm, 20µm, 50µm respectively.

The eye diagram in Figure 12 shows the response of the victim line when the PRBS speed is 10 GBPS with signal rise and fall times of 10ps. It is clear that the eye is very narrow and the variation in delay has widened to an unacceptable

A. Predicting delay in a TSV bundle

Given that the investigations reveal a lumped capacitive equivalent circuit as being sufficiently accurate, the switching pattern dependant delay within a bundle can be accurately captured by a first-order Elmore delay model. For the entire range of geometrical configurations considered, the delay can be accurately estimated by (1) where the empirically
level. At this speed the crosstalk completely overpowers the signal on the victim.

For the considered range, simulations show that resistance and inductance are mostly negligible for latency and SI considerations and therefore signal propagation through an isolated TSV as well as a TSV in a bundle can be analyzed by considering the capacitance alone. Tests were also carried out to determine if the via should be treated as a lumped or distributed model. The results show that no benefit is conferred by considering a distributed model due to the relatively low resistance and a single lumped section is sufficiently accurate.

Furthermore, crosstalk effects between TSV structures in a 3x3 bundle were examined. Capacitive crosstalk is far greater than inductive crosstalk, such that inductance can be ignored in most cases. Due to the reduced complexity of the TSV electrical model as proposed in this paper, a simplified delay formula based on the Elmore delay and empirical switch factors were proposed to estimate delay in a TSV bundle with a maximum error contained to within 8% over the entire simulated range. These equations allow for preliminary assessment of delay for worst, nominal and best case switching scenarios in accordance with well-established timing analysis practice. Finally, simulations were carried out using eye diagrams to further investigate SI issues, demonstrating the effect of capacitive coupling in a TSV bundle with random switching patterns. Shielding the lateral TSVs in a bundle was shown to increase signal reliability and allow for faster speeds through the structures. It is expected that this study will provide the basis for further explorations through the recommendation of the equivalent circuits as well as the investigations on the relative importance of the various parasitic terms, providing insight into signaling schemes over TSV interconnects.

V. CONCLUSION

TSVs represent an important interconnection option for 3D ICs but have not received significant attention in the literature with regard to their signaling characteristics. In this paper, parasitic extraction has been carried out using a field solver for typical geometrical configurations achievable in current processes in order to examine their variation with physical dimensions and carry out delay and signal integrity explorations. As suggested by field theory the significant capacitive coupling terms in a bundle are restricted to nearest neighbors, while the mutual inductance terms are significant throughout the bundle.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support of European Union research funding under grant FP7-ICT-215030 (ELITE) of the 7th framework programme, and its collaborative partners Qimonda GMbh in Germany, LETI in France, and HyperStone in Germany.

REFERENCES


Examination of Delay and Signal Integrity Metrics in Through Silicon Vias
Matt Grange*, Roshan Weerasekera*, Dinesh Pamunuwa* and Hannu Tenhunen*  
*Centre for Microsystems Engineering, Faculty of Science and Technology, Lancaster University, Lancaster LA1 4YR, United Kingdom  
*Department of Electronics, Computer and Software Systems, KTH School of Information and Communication Technologies, Forum 120, 164 40 Kista, Sweden

Introduction

• Through Silicon Vias (TSV) are a promising interconnect technology to meet the demands of fast and reliable vertical signaling in 3-D Integrated circuits. As the use of TSV structures for system interconnects is a fairly recent concept, the effects on signaling across them within a 3-D circuit are not well documented. It is the aim of this work to provide a preliminary assessment of the effect a TSV has on signal integrity within a realistic context.

• Spectre simulations were performed on extracted parasitics for a range of TSV dimensions and pitches to determine the characteristic behavior of these structures for signal integrity (SI) and delay analysis. A reduced electrical model is proposed and a first-order delay model is derived based on extensive simulations.

TSV Electrical Equivalent Circuit Model

• Field solver simulations were carried out to determine parasitic and coupling values for realistic geometrical configurations of Through Silicon Vias as lone structures, and in bundles. The equivalent electrical circuit diagram for a TSV as shown in Figure 1 was used throughout all circuit level simulations.

• The parasitics were extracted for TSV structures with radii of 5-40um and lengths of 20-140um. Mutual capacitance and inductance values were extracted for the same geometrical structures, with TSV pitches ranging from 50um with on-chip interconnects, the trends for wire capacitance, resistance, and inductance also increases. As the width increases, the capacitance goes up but the inductance and resistance decrease.

Significance of RLC Parasitics

• Simulations were performed for the entire range of resistance, inductance and capacitance values as determined by the field solver.

• The resistance was swept from 0-500mΩ, and the output waveforms were plotted to observe variations in delay. As can be seen in Figure 2, the TSV resistance within the considered range of is so small that it has no observable effect on the output waveform.

• The inductance was then swept from 0-500µH, the extracted range, for rise times down to 1ps, revealing no significant contribution as demonstrated in Figure 3.

• Finally, the capacitance was swept from 0-500F showing a significant effect on latency of the output waveform, as seen in Figure 4. These results appear to show that the electrical model for a cylindrical TSV can be reduced to a purely capacitive model. The resistive and inductive parasitics are small enough to be neglected in any delay simulations, which reduces the complexity of the electrical model significantly.

Crosstalk in a 3x3 TSV Bundle

• Figure 5 shows a sweep of the maximum range of couple capacitances in all bundle sizes of 8 surrounding aggressors switching on a central victim. Figure 6 shows the worst case inductive coupling under the same scenario.

• As in the on-chip case, the capacitive coupling terms to nearest neighbors dominate over the coupling terms to non-adjacent lines, which are mostly insignificant. This is observable in Figure 7.

• In the case of inductance, the coupling is significant within the entire bundle because magnetic field lines tend to permeate the length and breadth of the global structure, again analogous to the on-chip case. This relationship can be observed in Figure 8.

Delay Prediction in a TSV Bundle

• Given that the investigations reveal a lumped capacitive equivalent circuit as being sufficiently accurate, the switching pattern dependant delay within a bundle can be accurately captured by a first-order Elmore delay model. For the entire range of geometrical configurations considered, the delay can be accurately estimated by (1) where the empirically determined switch factors for the various switching patterns are defined in Table 1. Here C is the self capacitance, C the lateral coupling capacitance, C the diagonal coupling capacitance and di the driver resistance.

\[ I_d = 0.69 R_d (C_s + K_1 C_k + K_2 C_d) \]  

Table 1: Switch pattern-based delay

<table>
<thead>
<tr>
<th>Switching Pattern</th>
<th>K_1</th>
<th>K_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Victim Lateral Diagonal</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>3.4</td>
<td>5.2</td>
</tr>
<tr>
<td>9.0</td>
<td>10.6</td>
<td></td>
</tr>
</tbody>
</table>

Signal Integrity Simulations

• In order to fully capture the effect of crosstalk on delay and coupled noise amplitude under real-world conditions, simulations were carried out with pseudo-random bit streams (PRBS) at the victim and aggressor inputs in a 3x3 bundle to generate the eye diagrams at the output. All drivers were size 10 inverters while every TSV was loaded with a minimum-sized inverter. The example geometry chosen was a bundle with radii, length and pitch of 15µm, 20µm, 50µm respectively.

• The eye diagram in Figure 9 shows the response of the victim line when the PRBS speed is 10Gbps with signal rise and fall times of 10ps. It is clear that the eye is very narrow and the variation in delay has widened to an unacceptable level. At this speed the crosstalk completely overpowers the signal on the victim.

• Since the lateral (Northern, Southern, Eastern, and Western) neighbors in the bundle contribute the majority of the capacitive coupling, an obvious strategy to counteract capacitive crosstalk at high signaling speeds is to use these lines as shields. As seen in Figure 10, this effectively eliminates the majority of the coupling and allows for higher bit rates through the interconnect.

The authors would like to acknowledge the financial support of European Union research funding under grant FP7 ICT-215030 (ELITE) of the 7th framework programme, and its collaborative partners Daimonda GmbH in Germany, LETI in France, and HyperStone in Germany.

©2009 Matt Grange, m.grange@lancaster.ac.uk Presented at DATE 2009 10 Integration Workshop

Figure 1: Capacitive coupling effect in a bundle  
Figure 2: Inductive coupling effect in a bundle  
Figure 3: Capacitive sweep  
Figure 4: Inductive sweep  
Figure 5: Maximum Inductive coupling  
Figure 6: Capacitive coupling in a 3x3 bundle  
Figure 7: Capacitive coupling effect in a bundle  
Figure 8: Inductive coupling effect in a bundle  
Figure 9: 10Gbps unshielded victim in a bundle  
Figure 10: 10Gbps shielded victim in a bundle

* Corresponding author.
On-chip waveform capturing techniques have evolved in the field of signal and power integrity studies [1]. Their benefits can be widely extended in a 3D integrated system, to probe hidden environments of circuit operation deeply inside stacked chips and to support diagnosis and testability of analog/RF and mixed-signal complex functionality.

On-chip waveform capturing functionality is partitioned in 3D realization, where probing front end (PFE) circuits are attached to signals of interest in a target chip and decoupled from the rest of waveform acquisition functions on a different chip, such as voltage and timing generation, as well as data processing and storing.

The compactness is pursued in the design of PFE, consisting of source followers (SFs) for sensing and a latched comparator (LC) for discretization. A prototype setup includes PFEs on a target chip (Chip A), a back-end processor with memory macros embedded on Chip B for executing program codes of an efficient waveform acquisition algorithm, as well as off-chip measurement equipments and PC.

The number of transactions between PC and measurement equipments exhibits the highest impact on the system throughput of waveform acquisition. An efficient algorithm determines the most approximate value of PFE output voltage at a particular sampling timing. The algorithm is based on the essential fact that the voltage difference between adjacent digitized points becomes small for the smaller sampling interval (oversampling). The initial value is set at the latest value determined in the previous timing. A voltage step is sized in an order from coarse search to fine one. In addition, four-times parallel data channels in a digital data processing unit (DPU) realizes four times (4x) acceleration of the algorithm, as proven by measurements.

Test chips were fabricated in a 90-nm CMOS technology. High-voltage transistors prepared for I/O cells are used in the front-end part including PFE and DPU, allowing single supply voltage operation for rail-to-rail input voltage range of signal capturing. Analog components such as capacitors are not needed, due to the elimination of analog S/H and ADC stages, and minimize the area consumption of PFE. The back-end data processors use 1.0-V core transistors.

Dynamic performance is evaluated through FFT analysis applied to captured waveforms. A 1-MHz sinusoid given to PFE is digitized with the sampling interval of 5 ns, equivalently to the sampling frequency of 200 MHz. Frequency components for the signal bandwidth of interest of 100 MHz, exhibits SFDR of 57.1dB and SNDR (signal to noise distortion ratio) of 48.9dB. In addition to the 2nd and 3rd harmonics of the input signal, spurious components are due mainly to board-level coupling of the reference frequency at 10 MHz that is distributed among measurement equipments and that of the clock frequency at 25 MHz for the back-end data processing, respectively.

These experiments prove that the inclusion of small-area PFE circuitry realizes on-chip monitoring of various analog and digital internal signals, which potentially enhances diagnosis and testability of 3D integration.

On-chip Waveform Capturing Functionality Partitioned for 3D Realization

Yuuki Araga, Yoji Bando, Takushi Hashida, Makoto Nagata, Kobe University

- Inclusion of small area probing circuitry realizes on-chip monitoring.
  - On-chip waveform capturing technique have evolved in the field of signal and power integrity studies.
  - The technique can be widely extended in a 3D integrated system, to probe hidden environments of circuit operation deeply inside stacked chips.
  - Small-area probing front end circuitry realizes on-chip monitoring of various analog and digital internal signals.

* This work is partly supported by Semiconductor Technology Academic Research Center (STARC).

**DPU: Data Processing Unit
*PFE: Probing Front End

On-chip Waveform Capturing Setup

- The Probing Front End (PFE) array can probe multi signals in hidden environment of bottom chip (Chip A).
- PFE consists of Source Followers (pSF, nSF) and Latched Comparator (LC) for discretization.
- A back-end processor with memory macros embedded on Chip B.

Design of PFE

- Schematic and physical layout example of PFE, contains p- and n-channel SFs.
- Simulated DC transfer of PFE, showing that p- and n-channel SFs cover low and high side of rail-to-rail input voltage.

Throughput of waveform acquisition

- A discretization algorithm for capturing on-chip waveform with PFE, minimizing the number of transactions per a data point.
- Four-channel parallel discretization with 4x DPU realizes four times acceleration of the algorithm.

Experiments

(A) Minimum setup of on-chip waveform capturing system.
(B) Demo of waveform capturing (amplitude modulation signal).
(C) 1-MHz sinusoids are digitized with sampling interval of 5ns.
(D) FFT analysis of 1-MHz captured sinusoid.
Thermal Aware Test Scheduling for Stacked Multi-Chip-Modules

Vinay NS∗
Indian Institute of Science
Bangalore, India
nsvinay@cedt.iisc.ernet.in

Erik Larsson†
Linkoping University
Sweden
erila@ida.liu.se

Virendra Singh‡
Indian Institute of Science
Bangalore, India
viren@serc.iisc.ernet.in

ABSTRACT
In an attempt to increase the area utilization of multi-chip packages, manufacturers have started looking at 3D packaging. These stacked structures have low thermal capabilities, and hence the thermal issues get aggravated during testing. In this paper, thermal aware test scheduling of stacked structures is proposed. Care is taken to schedule the stacked cores appropriately to prevent local heating and achieve uniform vertical temperature spread. Concurrent testing of dies is considered for reduced test time and suitable test architecture is assumed. A new partition scheme is proposed for the power profile based on true power, as the global peak power model gives a pessimistic estimate of the power profile. The partitioning proposed is a trade-off between a cycle accurate model and a global peak power model. The algorithms are based on heuristics; use thermal models for stacked structure and power constraints for scheduling purposes. System modeling is done considering the various factors that affect the heat flow and die temperatures during testing. Experimental results show that the proposed technique achieves uniform temperature spread across the stack.

Keywords
Thermal aware, Test scheduling, 3D

1. INTRODUCTION
The benefits of incorporating more than one die on a substrate are numerous and compelling. Multiple chip Module(MCM)solutions provide better performance through shorter interconnects as compared to single chip packages connected through the printed circuit board fabric. They have lower inductance, capacitance, crosstalk, and power consumption. The latest development in MCM is the 3D orientation of chips called as stacked-MCM. Stacked-MCMs have silicon wafers stacked one above the other and wire-bonded to form a package. This form of packaging is becoming quite popular in recent times with many major IC manufacturers getting in the race.

The stacked structure has certain drawbacks which have to be addressed. The most important among these are the thermal issues. Due to the stacked nature, inner chips encounter higher temperature resistance and have higher cooling down times. This results in overheating and may damage the chips. Thus care has to be taken during all the stages involving high temperature dissipation and thermal issues have to be addressed during all the phases of the packaging process. Several approaches have been proposed to achieve good thermal spread during the design and placement for planar chips, but the thermal issues during chip placement in a stack have not been addressed. The most recent ones [1] tries to achieve a smooth temperature profile and addresses the issue during the placement of modules in the stack for floorplanning. The same can be extended for multiple dies on a stack, except that the freedom is restricted in the horizontal direction. Apart from this fabrication and packaging issues also have to be addressed. [2], [3] discuss the related issues during fabrication and packaging. Testing is one of the most important phases of IC fabrication and deployment. Hence it is necessary to address thermal issues during testing phase. Not many attempts have been made in this regard. [4], [5] discuss thermal aware testing for planar SoCs. The techniques proposed here cannot be directly used for 3D structures.

In this paper we present a thermal aware test scheduling scheme for stacked MCMs. The rest of the paper is organized as follows; section 2 discusses the test architecture for stacked MCM. Section 3 describes the thermal model used. Section 4 describes the system modeling aspects. Section 5 discusses the algorithms for thermal aware test scheduling. Sections 6 and 7 have the experimental results and conclusions.

2. TEST ARCHITECTURE
The standard boundary scan test-architecture can be extended to obtain test architecture for an MCM. Various such methods have been mentioned in [6]. The different methods offer tradeoff between Test Access Mechanism complexity and test flexibility. The test mechanism for a stacked MCM should provide the following vital features.
• Provision for testing all the dies on the stack.
• Built in self test for certain dies.
• Concurrent testing of dies on the stack.

Test architectures such as parallel-MCM Test Controller(MTC) and reconfigurable MTC[?] provide the essential test features stated above. We assume one of the architectures and proceed.

3. THERMAL CHARACTERIZATION
Thermal modeling of the stacked structure is essential for the proposed test scheduling. The accuracy of the thermal model used is critical and hence the different models must be studied in detail. We use the thermal model proposed in [7] here. This model which is resistive in nature is simple as well as accurate. It models the interface with resistances and the heat sources(dies being tested) as capacitors. The thermal model used is as shown in Figure 1. The side wall thermal resistance is high and hence is neglected. Once the thermal model is obtained we can model various physical parameters mathematically to reflect the effect it has on the system.

4. SYSTEM MODEL

Figure 1: Thermal model of the stack structure

This section deals with modeling of the system. We define variables describing the attributes of the system such as die temperature, heat flow in the stack etc. These variables are later used by the algorithm to generate the cost function.

The naming of dies is done from the bottom of the stack. The structure shown below has N dies stacked. The bottommost is numbered 1 and the topmost is N. The resistance values can be obtained from the thermal impedance matrix. We have simplified the structure by assuming equal resistances(practically almost equal) between die interfaces. The justification for this assumption is given in the later part of the paper. The list below gives the other definitions.

• Die-number (i): Die number indicates the location of the die in the stack. The bottom most die is numbered as 1 and the topmost as N.

Let us understand some of the non-intuitive parameters stated above.

4.1 Die Exposure

As seen from Figure 2, the thermal impedance seen by a single wafer with a planar orientation is lesser than that of a wafer in a stack. Therefore while considering the power profile of these wafers in the stack scaling has to be done. If not, the power constraint considered for the structure will be inaccurate. We scale the power profile with the die-exposure constant to account for the above stated reason. Let us derive an expression for die-exposure.

Assuming equal die-interface thermal resistance $R$ and assuming that a capacitor acts as open for steady flow of heat we proceed to derive an expression.

Looking at Figure 2, we have, for die $i$ in the stack, the thermal impedance $Z_i$ is,

$$Z_i = i \times R(1 + 1 - i) \times R$$

Simplifying, we get,

$$Z_i = i \times (N + 1 - i) \times R/(N + 1)$$

Now for a bare die the thermal impedance $Z_b$ is

$$Z_b = R/2$$
Thus, we have
\[ DE_i = 2 \frac{i \cdot (N + 1 - i)}{(N+1)} \] (1)

This is independent of the thermal resistance and hence the initial assumption of equal interface resistance does not introduce significant errors.

### 4.2 Thermal Influence

![Figure 3: Heat flow in stack](image)

When a die is being tested, i.e. the die under test (DUT) the heat dissipated during testing flows through either of the arms as shown in Figure 3. The ratio of heat through the arms is proportional to the ratio of the thermal impedance of the two arms. The thermal influence of the DUT on the other dies in the stack is quantified by the heat flow in the arm. The expression for thermal influence is given below.

\[ T_{i,j} = R_j/R'_j \] (2)

When \( d_{i,j} > 0 \),
\[ T_{i,j} = \frac{(N + 1 - i)}{i} \] (3)

When \( d_{i,j} < 0 \),
\[ T_{i,j} = \frac{i}{(N + 1 - i)} \] (4)

where \( R_j \) and \( R'_j \) are the thermal resistances of arms containing die \( j \) and not containing die \( j \) respectively.

### 4.3 Adjacency factor

![Figure 4: Effect of heat flow on Temperature of die](image)

While the thermal influence due to the DUT is same for all the dies on the same arm, the actual effect of this heat flow on the temperature of the wafer is different. The adjacency factor models this. The temperature of the die can be derived with the help of thermal-influence and interface resistances and is explained below. From Figure 4 we can say that adjacency factor is a function of die-number and interface resistance. Also the temperature of die \( j \) when die \( i \) is being tested is equal to the product of the heat flowing in the arm and ratio of the resistance to die \( j \) from \( i \) to the total resistance in the arm. Now the heat flowing in this arm is quantitatively given by the thermal influence. The ratio of resistances is taken care of by the adjacency factor.

Looking at Figure 4 the ratio of the resistances is given by,
\[ A_{i,j} = f(i, j, R) \]

When \( d_{i,j} > 0 \),
\[ A_{i,j} = \frac{j \cdot R \cdot R'}{i \cdot R} = \frac{j}{i} \] (5)

When \( d_{i,j} < 0 \),
\[ A_{i,j} = \frac{(N + 1 - j) \cdot R \cdot R'}{(N + 1 - i) \cdot R} = \frac{(N + 1 - j)}{(N + 1 - i)} \] (6)

The box representation for the power profile of a die is not efficient. This is because the difference between the maximum and minimum power in the profile is huge. Global peaks give a pessimistic estimate for the box representation as explained in Figure 5. Here we subdivide the power profile based on the local peaks to obtain more efficient partitions. Also partitioning the power profiles gives flexibility as to when the die is actually tested and this time can be chosen based on the time for cooling. More partitions may result in higher complexity; hence we put a limit on the maximum number of partitions and use the above method. The above partitioning results in better approximation than the simple box model.

![Figure 5: Conflict graph for a five-stack structure](image)

The above definitions model the system parameters quite accurately. With these parameters, the scheduling can be performed. The word core is used to indicate one of the die
power profile partitions. In the next section we discuss the problem in detail and also describe the scheduling algorithms proposed.

5. TEST SCHEDULING

We propose a thermal aware test scheduling algorithm based on the parameters defined previously. In these algorithms we try to schedule the partitioned cores within the power constraints. The problem is stated as follows.

Problem formulation: Given a stack consisting of \( N \) dies with power profiles \( P_1, P_2, ..., P_N \) and a well defined thermal model with all the interface resistances approximated to \( R_t \), with \( P_{\text{max}} \) being the power constraint, find an efficient thermal aware scheme involving modeling and test-scheduling of dies such that the power constraint is not exceeded. In this regard we look at two algorithms which use different heuristics and try to achieve the above goals.

The partitioning further sub-divides the power profile \( P_i \) into, what we call, cores. \( P_1 \) thus consists of cores \( C_1, C_2, ..., C_m \). The number of partitions is taken to be 5.

5.1 A Simplistic Approach to Thermal Aware Scheduling

In this approach importance is given to the thermal constraint rather than the time for testing. Here we assume that concurrent scheduling of conflicting cores cannot be done. For example the top-most and the bottom most dies see little impedance on one of the arms and hence are not affected by any activity in the rest of the structure i.e. these dies do not conflict with any of the other dies in the stack. The dies at the centre see almost equal impedances in the two arms and hence thermal activity anywhere affects the die i.e. they conflict with all the dies in the stack. Thus from the above concept we can prepare a conflict graph and use it for scheduling. A clear threshold must be chosen to define dies with equal resistances in the two arms. We do this by plotting the resistance in the two arms for all the dies and then define a variance around the point of intersection and then consider all the dies lying within this variance to have equal resistance. The dies lying within the circle, conflict with all the dies in the stack. For any die outside the circle of equal resistance the conflict is only with the dies lying in the path of lower resistance. The scheduling begins by selecting one of the cores. The graphs for different dies are then run through to find out if the core is under conflict. The set of cores with no conflict are selected and scheduled after verifying for the power constraint condition. In this approach the most important task would be to obtain the conflict graph after clearly identifying the dies with resistances nearly equal in both the arms.

For the conflict graph shown in Figure 6 the schedule order can be 1 and 5, 2, 3 and finally 4 assuming \( m=1 \).

The flow of the algorithm is as explained below.

1. for all \( i \) such that \( 0 \leq i \leq N \) do
2. Prepare a conflict graph
3. end for
4. for all \( i \) such that \( 0 \leq i \leq N \) do
5. Scale the power profiles according to the die exposures
6. end for

The conflict graph after clearly identifying the dies with resistance nearly equal in both the arms.

Figure 6: Conflict graph for a five-stack structure.

7. for all \( i \) such that \( 0 \leq i \leq N \) do
8. Power profile is partitioned using the partitioning algorithm
9. end for
10. repeat
11. Schedule list = NULL.
12. Select one of the unscheduled cores and add to the schedule list
13. Total power = 0
14. repeat
15. Check for a non-conflicting core (cores).
16. Find the sum of the powers of the cores selected
17. Add the selected core to the schedule list
18. until Power constraint is exceeded
19. Remove the last added core from the schedule list.
20. Schedule the cores in the schedule and remove from the unscheduled cores list.
21. until All the cores are scheduled

This method considers the weights of the conflict graph as infinity. A better method would be to make the weights finite. By doing so the complexity of the algorithm is increased but more concurrency and controllability can be brought in. In the next algorithm proposed we calculate the cost of each die with respect to the other dies and use this cost to schedule the dies.

5.2 Least Intrusive Algorithm

In this approach we calculate the cost function for all the cores with respect to activity in the other cores and schedule those cores which satisfy the power constraint and also have the least total cost. Here for scheduling we consider both the power constraint and total cost for concurrently scheduled cores. The maximum total cost provides controllability and can be adjusted depending on the requirements. We schedule a core for testing and then select the core which has the least cost with respect to the scheduled core and check if it satisfies the power and the total cost constraint. If it does we schedule it and then look at cores with the same strategy (In the \( k \) th sub-iteration i.e. when \( (k-1) \) cores have been selected for scheduling, the core with the lowest cost considering all the selected cores has to be chosen). If it does not we go to the core with the next least cost and check to see if it satisfies the constraints. Once this iteration is done we remove the tested cores and then apply the same algorithm for the remaining cores. The cost function is calculated with the help of the system parameters.
The cost function should consider various factors such as thermal flow and junction temperature. With this in mind, a cost function, with adjacency factor and thermal influence, can be arrived upon. The junction temperature at each of the die interface is given quantitatively by the product of the thermal-influence and adjacency factor. The cost function can be defined suitably and more parameters can be added without much change to the algorithm proposed.

The flow of the schedule algorithm is explained below.

1. for all $i$ such that $0 \leq i \leq N$ do
2. Power profile is partitioned using the partitioning algorithm
3. end for
4. for all $i$ such that $0 \leq i \leq N$ do
5. Scale the power profiles according to the die exposures
6. end for
7. for all $i$ such that $0 \leq i \leq N$ do
8. The cost function is calculated using the system variables.
9. end for
10. repeat
11. Schedule list = NULL.
12. Select one of the unscheduled cores and add to the schedule list
13. Total power = 0
14. Total cost = 0
15. repeat
16. Select a core which has the lowest cost with respect to the (cores) core selected.
17. Find the sum of the powers of the cores selected
18. Find the sum of the cost of the cores selected
19. Add the selected core to the schedule list
20. until Both power constraint and total cost are exceeded
21. Remove the last added core from the schedule list.
22. Schedule the cores selected and remove the scheduled cores from the list.
23. until All the cores are scheduled

6. CONCLUSION
In this paper we propose two algorithms for efficient testing of stacked structures with importance given to thermal activity. We demonstrate two methods employing entirely unique methods to achieve thermal aware testing. The two heuristics look at reducing the local heating, an other approach could be to look at the actual flow of heat in the stack during testing. This approach is being looked at to further reduce the test time.

7. REFERENCES
[2] Bruce Kim, Member, IEEE, Madhavan Swaminathan, Member, IEEE, Abhijit Chatterjee, Senior Member, IEEE, and David Schimmel, Member, IEEE, A Novel Test Technique for MCM Substrates, , February 1997, IEEE transactions on components, packaging, and manufacturing technology-U-part b, vol. 20, no. 1.
Thermal Aware Test Scheduling for Stacked Multi-Chip-Modules

Vinay N S, CEDT, Indian Institute of Science, v_nv@serc.iisc.ernet.in
Erik Larsson, Linkoping University, Sweden, erlia@ida.liu.se
Virendra Singh, SERC, Indian Institute of Science, viren@serc.iisc.ernet.in

Introduction
The benefits of incorporating more than one die on a sub-strate are numerous and compelling. Multiple chip Module(MCM)solutions provide better performance through shorter interconnects as compared to single chip packages connected through the printed circuit board fabric. They have lower inductance, capacitance, crosstalk, and power consumption. The latest development in MCM is the 3D orientation of chips called as stacked-MCM. The stacked structure has certain drawbacks which have to be addressed. The most important among these are the thermal issues. Due to the nature of stacked, inner chips encounter higher temperature resistance and have higher cooling down times. This results in overheating and may damage the chips. Test scheduling algorithms for a stacked structure is addressed here.

Test Architecture for an MCM
The standard boundary scan test-architecture can be extended to obtain test architecture for an MCM. Various such architectures have been proposed. The different methods offer tradeoff between Test Access Mechanism complexity and test feasibility. The test mechanism for a stacked MCM should provide the following vital features:
- Provision for testing all the chips on the stack.
- Built in self test for certain chips.
- Concurrent testing of chips on the stack.

Test architectures such as parallel-MCMTest Controller(MTC) and reconfigurable MTC provide the essential test features stated above.

Thermal Characterization and System Modelling
Thermal modeling of the stacked structure is essential for the proposed test scheduling. A Capacitive-resistive model is used as it is simple and intuitive. It models the interface with resistances and the heat sources(chips being tested) as capacitors. Side wall thermal resistance is high and hence is neglected.

System Parameter Definitions
We define variables describing the attributes of the system such as die temperature and heat flow in the stack. These variables are later used by the algorithm to generate the cost function. The naming of chips is done from the bottom of the stack. The structure is assumed to have N chips stacked. The bottommost is numbered 1 and the topmost N. The resistance values can be obtained from the thermal impedance matrix. The list below gives the other definitions:

- Die number (i): Die number indicates the location of the die in the stack. The bottom most die is numbered as 1 and the topmost as N.
- d_{ij}: This is the distance between two chips in the stack and is defined as the difference between the die numbers d_{ij} = i - j.
- Die-exposure (DIE): Indicating the thermal exposure of die i.
- Thermal influence (T_{ij}): Thermal-influence value indicates the effect of the thermal state of die i on j.
- Power profile (P_i): It is the rectangular box profile model of the single die i.
- Modified power profile (P_{mi}): The power profile which accounts for Die-exposure.
- Partitioned power profile (P_{mij}): The partitioned modified power profile.
- Adjacency factor (A_{ij}): This indicates how close die j is to a die i thermally.

Problem formulation: Given a stack consisting of N chips with power profiles P_1, P_2, . . ., P_N and a well defined thermal model with all the interface resistances approximated to R, with P_{max} being the power constraint, find an efficient thermal aware scheme involving modeling and test-scheduling of chips such that the power constraint is not exceeded. In this regard we look at two algorithms which use different heuristics and try to achieve the above goals.

The partitioning further sub-divides the power profile P_i into, what we call, cores. P_i thus consists of cores C_{1}, C_{2}, . . ., C_N. The number of partitions is taken to be S. The cost function is calculated using the system variables.

Least Intrusive Algorithm
1. for all i such that 0 ≤ i ≤ N do
2. Prepare a conflict graph
3. end for
4. for all i such that 0 ≤ i ≤ N do
5. Scale the power profiles according to the die exposures
6. end for
7. for all i such that 0 ≤ i ≤ N do
8. Power profile is partitioned using the partitioning algorithm
9. end for
10. repeat
11. Schedule list = NULL.
12. Select one of the unscheduled cores and add to the schedule list
13. Total power = 0
14. repeat
15. Check for a non-conflicting core (cores).
16. Find the sum of the powers of the cores selected
17. Add the selected core to the schedule list
18. until Power constraint is exceeded
19. Remove the last added core from the schedule list.
20. Schedule the cores in the schedule and remove from the unscheduled cores list.
21. until All the cores are scheduled

Test Scheduling algorithms
1. for all i such that 0 ≤ i ≤ N do
2. Power profile is partitioned using the partitioning algorithm
3. end for
4. for all i such that 0 ≤ i ≤ N do
5. Scale the power profiles according to the die exposures
6. end for
7. for all i such that 0 ≤ i ≤ N do
8. The cost function is calculated using the system variables
9. end for
10. repeat
11. Schedule list = NULL.
12. Select one of the unscheduled cores and add to the schedule list
13. Total power = 0
14. Total cost = 0
15. repeat
16. Select a core which has the lowest cost with respect to the (cores) core selected.
17. Find the sum of the powers of the cores selected
18. Add the selected core to the schedule list
19. until Both power constraint and total cost are exceeded
20. Remove the last added core from the schedule list.
21. Schedule the cores selected and remove the scheduled cores from the list.
22. until All the cores are scheduled

Figure 1: A stacked structure with the thermal resistances and capacitances shown. The heat emitting chips are modelled as capacitors and the interface resistances between adjacent chips are modelled as resistances.

Figure 2: The thermal impedance seen by a single wafer with a planar orientation is lesser than that of a wafer in a stack. Therefore while considering the power profile of these wafers in the stack scaling has to be done. This scaling factor (die-exposure) is defined to be the ratio of the thermal-impedance in the two cases.

Figure 3: When a die is being tested, the heat dissipated during testing flows through either of the arms. The ratio of heat through the arms is proportional to the ratio of the thermal impedance of the two arms. The thermal influence of the DUT on the other chips in the stack is quantified by the heat flow in the arm.

Figure 4: While the thermal influence due to the DUT is same for all the chips on the same arm, the actual effect of this heat flow on the temperature of the wafer is different. The adjacency factor models this.

Figure 5: The box representation for the power profile of a die is not efficient. This is because the difference between the maximum and minimum power in the profile is huge. Global peaks give a pessimistic estimate. Here we subdivide the power profile based on the local peaks to obtain more efficient partitions.

Figure 6: When d_{ij} < 0, A_{ij} = \frac{1}{(N+1-j) + R} \times \frac{1}{N+1-i + R}

Figure 7: The flow of the algorithm is as explained below.
1. for all i such that 0 ≤ i ≤ N do
2. Prepare a conflict graph
3. end for
4. for all i such that 0 ≤ i ≤ N do
5. Scale the power profiles according to the die exposures
6. end for
7. for all i such that 0 ≤ i ≤ N do
8. Power profile is partitioned using the partitioning algorithm
9. end for
10. repeat
11. Schedule list = NULL.
12. Select one of the unscheduled cores and add to the schedule list
13. Total power = 0
14. repeat
15. Check for a non-conflicting core (cores).
16. Find the sum of the powers of the cores selected
17. Add the selected core to the schedule list
18. until Power constraint is exceeded
19. Remove the last added core from the schedule list.
20. Schedule the cores in the schedule and remove from the unscheduled cores list.
21. until All the cores are scheduled

Figure 3: When a die is being tested, the heat dissipated during testing flows through either of the arms. The ratio of heat through the arms is proportional to the ratio of the thermal impedance of the two arms. The thermal influence of the DUT on the other chips in the stack is quantified by the heat flow in the arm.

Figure 4: While the thermal influence due to the DUT is same for all the chips on the same arm, the actual effect of this heat flow on the temperature of the wafer is different. The adjacency factor models this.

Figure 5: The box representation for the power profile of a die is not efficient. This is because the difference between the maximum and minimum power in the profile is huge. Global peaks give a pessimistic estimate. Here we subdivide the power profile based on the local peaks to obtain more efficient partitions.
A Prospective Analysis of High-Frequency Cross-Coupling Mechanisms for the Next 3D Packaging Generation

Roberto Antonicelli, PhD, IEEE member
ST–Ericsson, Connectivity and Broadcast, R&D
Excelsiorlaan, 44–46 – B1930 Zaventem – Belgium
roberto.antonicelli@stericsson.com – www.stericsson.com

3D Integrated circuits consist of multiple device layers which are integrated directly on wafer level. In comparison to 2D packages, 3D integrated circuits achieve a higher level of integration, thus keeping Moore’s law still on-track.

One of the common limitations encountered in densely integrated high-frequency or high-speed systems is represented by the excessive level of mutual coupling which affects different cells within the same IC or through adjacent IC’s, usually creating higher levels of interference and cross-coupling and disturbing the regular functioning of the system.

Design paradigms are also under discussion. Design and layout competence borders are not well defined, and sequential 2D design is being replaced by a concurrent 3D approach.

Principal coupling mechanisms
- Basic coupling (capacitive and inductive), often not avoidable, creating
- Clock harmonics, mixing with RF transceivers in critical frequency bands, with adverse effect on sensitivity
- Power and ground intersections, giving rise to
- Pure magnetic, from trace to coil, from coil to coil, creating unwanted in-band spurious, intermodulation products, etc, with adverse effects on modulation performance

Coupling mechanisms shall be foreseen during common design and layout phase, and proper EDA tools need be available, with consolidated flow and proven sign-off.

Basic coupling mechanisms
1. Inductive coupling (TSV to TSV, coil to coil, trace to trace, …)
   - Depends on via technology, metallization, sizes
   - Via are often isolated with clearance > Tradeoff area – performance 3D simulations mandatory to establish minimal confidence
2. Capacitive coupling (pad to pad, polygon to polygon, …)
   - Stacked IC can be as close as few to tens of microns
   - Active areas and BEOL’s of separate IC’s are subject to heavy coupling one to each other, usually 20 times the coupling within stacked SIP’s

Extensive use of back-annotations is mandatory

10um 200um

TSV’s

TSV

Capacitive coupling

Inductive coupling

1. Harmonics of digital clocks or fast-switching signals are often under-estimated: with reasonably short rise-time, strong harmonics can easily fall into GSM, GPS, Bluetooth or WLAN operating bands
2. Desensitization, blocking and spurious interferers are to be expected with adverse impact on design quality and timing
3. Such heavy consequences can be avoided by early prediction of critical coupling in function of aggressor/victim pair 3D localization
4. On-field requirements and customer breakout needs to be addressed before the kick-off of multiple layout sessions

Power and ground intersections

Clock harmonics

Conclusions
- Challenging packaging techniques are made available to keep Moore’s law alive
- Densely integrated packages result in closer vicinity of several spurious generators and sensitive receivers and PLL’s
- Need an early, rational, simulation–based intervention structured along several axes
- What’s next:
  - Holes in the CAD flow (management of libraries, tools and decks)
  - Lack of ownership in the field of electrical and thermal modeling (redistribution runners, passivation layers, TSV’s, under ball material, etc.)
Power Integrity Issues in 3D Integrated Chips Using TSVs (Through Silicon Vias)

(Extended Abstract)

Waqar Ahmad*, Qiang Chen*, Roshan Weerasekera*, Hannu Tenhunen*, Lirong Zheng*

* Department of Electronics, Computer, and Software Systems,
KTH School of Information and Communication Technologies,
ELECTRUM 229, 164 40 Kista, Sweden.
Email: {ahmadw,qiangch,roshan,hannu,lirong}@kth.se

Power supply instability and losses caused by long interconnects in 2D attracted the designers towards 3D integration technology in order to achieve low form factor and jump beyond Moor’s law. Designers are now exploiting the integration into the third dimension. At very high frequencies (of the order of GHz), the inductance is prime player. The effects of on-chip inductance on supply integrity have been neglected in the past by arguing that package inductance is so high that on-chip inductance is negligible. However, today’s packages based on BGAs have very small inductance as compared to on-chip inductance. High frequency switching currents tend to follow low inductive paths. Also, the author in [5] claim about 30% and 60% improvement in metal area usage for power grids in the 90nm and 45nm processes, respectively, by comprehensive modeling of on-chip power grid inductance. Therefore, inclusion of inductance in power integrity analysis is imperative now since it causes ground bounce or ringing effect at high frequency switching thereby producing supply noise. The ringing effect adds further challenges to power supply design for 3D ICs as practically supply can be injected only from bottom of 3D stacks and inclusion of vertical TSVs further adds to inductance. The length of inductive loop is increased gradually by adding more chips vertically.

Assume that TSVs connect the global supply and ground grids of chips. The load on TSVs is reduced gradually as we move from bottom to top. The load of whole integrated chips is distributed accordingly as we add TSVs on each floor uniformly. Fig.1 shows an example power supply scheme where 3 chips are stacked vertically through 2x4 pairs of TSVs. In this paper we represent each chip by a capacitor which behaves like the switching of on-chip load through a clock edge during transient analysis. Inductance and resistance of TSVs have been taken from[2] for 5 micrometer diameter. The results in Fig2.show that ground bounce increases as we move from bottom chip to top chip (as loop inductance is increased). In Fig3 we analyzed 8-chips stack using 8-pairs of TSVs which shows variation of ground bounce with respect to number of chips per stack and number of TSVs per stack. The ground bounce increases by increasing number of chips in a stack whereas it reduces by increasing number of TSVs in a given stack as a fraction of chip supply voltage. This ground bounce can be minimized by using on-chip distributed decoupling capacitor scheme proposed in [1].

References:


Fig.1 Power distribution scheme for 3 Chips stacked through 4 pairs of TSVs between adjacent chips. Load of each chip is divided into 4 equal parts. Similarly, N chips can be stacked using M TSV pairs, thereby distributing the load of N chips in M parts.

Fig.2 Maximum ground bounce (as percentage of supply voltage) on each chip in 3D vertical stacks (using TSV) vs the number of chips used in each stack.

This research is supported by European Union research funding under grant FP7-ICT-215030 (ELITE) of the 7th framework program.
Core Test-Wrapper Optimization for 3D ICs with Through Silicon Vias

Brandon Noia and Krishnendu Chakrabarty
Department of Electrical and Computer Engineering
Duke University
Durham, NC 27708
USA

November 23, 2008

ABSTRACT

As technology scales, the International Technology Roadmap for Semiconductors projects that on-chip communication will require new design approaches to achieve system-level performance targets. Since global interconnects do not scale with newer technologies, intermediate and global interconnects of current microprocessors and ASICs contribute to a significant power consumption and cause major signal integrity problems.

Three-dimensional integrated circuits (3D ICs) can potentially overcome the barriers in interconnect scaling, thereby offering an opportunity to continue performance improvements using CMOS technology. In a 3D chip, multiple device layers are stacked together with direct vertical interconnects such as through silicon vias (TSVs) tunneling through them. Consequently, one of the most important benefits of a 3D chip over a traditional two-dimensional (2D) design is the reduction in global interconnect. Other benefits of 3D ICs include: (i) higher packing density and smaller footprint due to the addition of a third dimension to conventional two-dimensional layouts; (ii) higher performance due to reduced average interconnect length; (iii) lower interconnect power consumption due to the reduction in total wiring length; (iv) support for the realization of mixed-technology chips.

Even though manufacturing techniques for 3D integration are nearly mature and 3D ICs offer tremendous benefits, there are several challenges that hinder the successful adoption of 3D architectures. A major concern today is the lack of systematic/structured manufacturing test and design-for-testability (DFT) solutions for 3D ICs. Among all EDA challenges for 3D IC design, tools and methods for 3D IC testing are regarded as the “No.1 challenge”, according to a keynote speech at the 2007 3D Architecture Conference by Ted Vucurevich, CTO of Cadence Design System.

Test access is one of the key DFT challenges for 3D ICs. In a core-based design implemented using 3D technology, large cores are likely to be partitioned in a “fine-grained manner” across multiple layers in a 3D stack. The internal scan chains for these cores are also likely to span multiple layers. To test a system chip designed using several such cores, a modular test solution can be adopted to reduce test complexity. An IEEE Std. 1500 wrapper and a test-access mechanism (TAM) must be optimized to minimize the test-application time for the core. However, the core terminals and scan I/Os need to be chained in the wrapper using TSVs if they span multiple 3D layers.
In this workshop presentation, we will describe early results on the optimization of a core test wrapper for the modular testing of 3D ICs. Given the placement of a core across multiple layers and the TAM width provided to it for testing, the goal of the optimization method is to minimize the test-application time under constraints on the number of TSVs available for wrapper design. It is important to limit the number of TSVs because designers need them for the routing of power/ground, clock, and functional signals. Even though TSVs can be fabricated with small via-hole diameters and pitch sizes, they require the use of a “keep out” area on each layer, which reduce density. Moreover, TSVs can be major yield limiters and reliability bottlenecks for 3D ICs. Therefore, the test-access architecture must be optimized under given constraints on the number of available TSVs.

Researchers have shown that the wrapper optimization problem for traditional (two-dimensional) ICs is NP-hard. Our optimization approach is therefore based on heuristic algorithms. These scalable algorithms adopt the bin design methods used in earlier work, but they include two new components: (i) enumeration of admissible partitions of the TSV limit across wrapper chains; (ii) bin design under constraints on TSV limits for the bins. We will also present integer linear programming models, which can provide useful lower bounds based on LP-relaxation and Lagrangian relaxation.
Application of substrate noise simulation methodology to 3D-stacking

S. Bronckers(1,2), G. Van der Plas(1), P. Marchal(1), Y. Rolain(2)
(1) IMEC, Belgium — (2) Vrije Universiteit Brussel, Belgium

Abstract—Substrate noise problems continue to harass the smooth co-habitation between the analog and digital circuitry on a System-On-a-Chip. Stacking the analog circuits on top of the digital circuitry can improve the immunity against substrate noise. This paper explores with simulations the benefits of 3D-stacking related to substrate noise. Simulations show that a low frequencies 3D-stacking can suppress the substrate noise coupling with more than 100 dB. However this reduction in substrate noise coupling can be jeopardized by the impedance of the PCB ground interconnect.

I. INTRODUCTION

Today’s mobile phone features a host of ICs ranging from analog/RF circuits, memory, digital signal processors, etc... New devices include image displays, MP3 devices, image sensors. Given the large number of ICs, placing them on printed-circuit boards (PCBs) and routing them to be properly interconnected in a tight layout is a daunting task. Integrating all the functionality on a single die is at this moment not possible. To further reduce the cost factor of mobile phones, the semiconductor industry tends to move to 3D-stacked IC’s. Such an IC is also known as a system-in-a-cube. A 3D-IC is a chip with more layers of active electronic components, integrated both vertically as horizontally into a single system. The semiconductor industry is hotly pursuing this promising technology in many ways. One way to stack 3D IC’s is the Die-on-die method [1]. The electronic circuitries are built on two or more dies, which are then stacked on top of each other. The different layers are connected with each other through vertical connections, called through-silicon vias (TSVs). This technology offers many significant benefits. First of all, 3D-ICs save on IC footprint area. Furthermore, thanks to the TSVs the signals do not have to go off-chip anymore which greatly reduces the power consumption, the propagation delay and extends the usable signal bandwidth. Also the cost can be greatly reduced because in comparison with the SoC, the electronics circuitries do not need to be integrated in the same technology. For example the power management unit can be fabricated in a cheap 1µm technology and the fast digital processor can be designed in an expensive 45nm technology.

Stacking the analog circuits on top of the digital circuitries can also increase the immunity against substrate noise [2]. This paper explores with simulations the benefits of 3D-stacking for reducing the impact of substrate noise on an analog/RF circuit. Therefore we consider the stacking of two dies. The bottom die contains a substrate contact that serves to emulate the switching activity of the digital circuitry. The top die contains an 48-53 GHz mm-wave LC-VCO that acts as the victim. The 3D-IC experiment is described in section II. The methodology to predict the impact of substrate noise in 3D-ICs is described in section III. The corresponding simulations are discussed in section IV and compared against the traditional SoC solution.

II. DESCRIPTION OF THE 3D-IC EXPERIMENT

The 3D-IC experiment is set up based on the Die-on-die method of stacking. This method consists in building the different circuitries that are possibly realized in different technologies, on two or more dies. In our case we consider the stacking of two dies. The bottom die consists of a substrate contact (see Fig 2). Such a substrate contact forms a resistive connection to the substrate and replaces the switching activity of the digital circuitry in this experiment. Hence, the substrate contact is considered as the aggressor. The substrate of the bottom die is unaltered.

The top die consists of the 48-53 GHz LC-VCO that is designed in a UMC 0.13µm CMOS technology (see Fig 2). The gain block of the oscillator consists of a cross-coupled NMOS transistor pair and a NMOS current mirror (see Fig. 3). NMOS transistors are preferred over PMOS transistors because they exhibit a lower parasitic capacitance for the same transconductance. In this way, a higher tuning range can be achieved. The frequency tuning of the LC-VCO is obtained by changing the capacitance value of the resonant LC-tank. The tunable capacitors are implemented as NMOS transistors whose drain and source node are shorted. The inductor of the...
A similar simulation methodology is used as described in Ref. [4]. However this methodology is extended to 3D-ICs. The simulation approach consists of two simulations. First an EM simulation [5] simulation is performed. This simulation includes the on-chip interconnects, the substrate and the passive components. Since the substrate noise signals are sufficiently small, the interconnects and the substrate can be considered to behave linear. Next, the results of the EM simulation are used together with the RF models of the devices to perform a SpectreRF [6] simulation. The resulting waveforms on the different terminals of the simulation model will give the designer insight in the different substrate noise coupling mechanisms (see Fig. 4). This section discusses the different simulations that are needed to characterize the 3D-IC experiment.

III. IMPACT SIMULATION METHODOLOGY

A. EM simulation

Fig. 2. 3D-view of the 3D-IC experiment.

Fig. 3. Schematic of the mm-wave LC-VCO

Fig. 4. Impact simulation approach

Fig. 5. Cross-section of the 3D-stacked experiment.
The simulation setup is similar to Ref. [4]. The layouts of the bottom and top circuitries are streamed into the EM environment. In this environment the substrate, the silicon dioxide and the Pwell are added for each die (see Fig. 5). An air box is added on top of the top die and a BCB layer connects both dies with each other. Furthermore, one port is placed at the substrate contact of the bottom die. On the top die, ports are placed at the connections of the different bias lines, at the terminals of the transistors and the differential output of the VCO. A complete cross-section of the experiment is shown in Fig. 5.

This EM model is simulated from DC up to 60 GHz with a minimum solved frequency of 50 MHz and a maximum $\Delta S$ of 0.01. It takes 7.5 hours on a HP DL145 Server to simulate the S-parameters at the different ports.

### B. Circuit simulation

A simulation model is constructed that fully characterizes the 3D-IC. The interconnects, the substrate and the passive components of both dies are represented by a S-parameters box resulting from the EM simulation. The transistors are represented by their RF model and are properly connected to the S-parameter box. On this complete simulation model the designer can apply any circuit analysis. The goal of this experiment is to predict the power of the substrate noise induced sideband spurs. To that end, a Periodic AC analysis is performed on this simulation model with SpectreRF. The analysis takes 2 minutes on a HPUX9000 platform.

### IV. REDUCING THE IMPACT OF SUBSTRATE NOISE

An LC-VCO is very sensitive to substrate noise coupling [7], [8]. A substrate noise signal which couples into a VCO will modulate the oscillator signal both in frequency as in amplitude. These modulation effects cause sideband spurs to appear around the local oscillator. In this section the power of the sideband spurs in a 3D-IC design is compared to the traditional SoC solution.

3D-stacking improves the immunity against substrate noise. The amount of gained isolation however strongly depends on the grounding scheme. In Ref.[4] where the impact of substrate noise was predicted on the same VCO in an SoC context, all the ground bond pads were shorted to a zero potential ground. This makes sense since all the ground bond pads do share the same ground plane. This is not the case when the 3D-IC is considered. The ground connection of the substrate contact is connected on PCB to the ground connection of the VCO with for example bond wires. The impedance between the node where both ground connections are merged to each other and the ideal ground will determine the amount of gained isolation.

We consider two cases:

- In the first case both dies are connected separately to the ideal ground. This means that the impedance between the node where the ground connection of the different dies are merged in the ideal ground is zero. This is the ideal case. In that case substrate noise couples capacitively through the silicon dioxide of the bottom die toward the substrate of the top die (see Fig. 6). The resulting spurs are also modulated in frequency up to 100 MHz but are not decreasing anymore with a rate of -20dB/decade as one would expect from resistive FM coupling. The height of the spurs is constant due to the capacitively coupling. Therefore the gained immunity against substrate noise is frequency dependent when compared to the SoC solution. 3D-stacking improves the immunity with 100 dB for a substrate noise signal of 1 MHz. For a noise signal of 800 MHz the immunity improves with 40 dB (see Fig. 6).

- In the second case a small resistor of only 0.5 $\Omega$ is placed between the node where both ground connections meet each other and the ideal ground (see Fig. 7). In that case, the injected substrate noise couples resistively toward the PCB and causes a voltage fluctuation over the external resistor of 0.5 $\Omega$. This voltage fluctuation is also sensed by the VCO and hence the VCO suffers from ground bounce. Ground bounce also causes FM modulated spurs in the spectrum of the VCO. Compared to the traditional SoC solution the improvement in substrate noise immunity is only 20 dB for a noise signal of 100 MHz. For a noise signal of 800 MHz the improvement is only 10 dB (see Fig. 7).

### V. CONCLUSION

This paper explores with simulations the opportunity of 3D-stacked circuits related to the immunity against substrate noise. Therefore a dedicated simulation experiment is set up where a very sensitive 48-53 GHz LC-VCO is chosen as a victim. The die of the LC-VCO is stacked on top of a die.
that mainly consists of a substrate contact. This substrate contact acts as an aggressor and is used to inject substrate noise. The simulations revealed that 3D-stacking reduces the substrate noise induced spurs with 100 dB for a substrate noise signal with a frequency of 1 MHz. However special attention should be paid to the PCB grounding scheme. The ground interconnects of the different dies should be connected on the PCB as close as possible to the ideal ground.

REFERENCES

Impact of thinning and packaging on a deep sub-micron CMOS product

1. Objectives
   - Understand impact of thinning on product and device performance
   - Determine risk to product yield from thinning & packaging
   - Evaluate metrology solutions for thin wafer characterization
   - Evaluate wafer thinning techniques

2. Experimental flow
   - 300 mm product wafers: 65nm low power CMOS
   - Mount to background tape and thin @ Amkor (#7)
   - Optical analysis of @ IMEC (#4, #5)
   - Physical analysis of @ IMEC (#4, #5)
   - Package assembly test @ Amkor (#9)
   - Final characterization test @ Qualcomm (#9)

3. Wafer thinning at IMEC
   - Wafer grinding: Coarse grind + soft clean (SPM/APM based)
   - CMP (10um)
   - Chipset for voice and data solutions for WCDMA (UMTS), EDGE and GSM/GPRS networks.
   - Analog die
   - Package Substrate

4. Thinned wafer characterization by Raman spectroscopy at IMEC
   - Raman spectroscopy measures the frequency of lattice vibrations (phonons).
   - The Raman peak is sensitive to stress, it shifts to higher frequencies for compressive stress and to lower frequencies for tensile stress.
   - Peak asymmetry and height (intensity) can be correlated to damage in Si.
   - Nine measurements were made on each thinned wafer. Si reference was measured before and after each sample measurement.
   - Summary and Conclusions: Within the measurement error, there is no damage visible (no increase of L1-L1' or L1-L1'' and no decrease of the intensity). In addition, no additional stress was detected (no shift of the peak position within the measurement error).

5. Thin wafer Moiré measurements at IMEC
   - 65 and 100µm wafers were diced and subjected to Topography and Deformation measurement
   - Plastic deformation/hysteresis observed at temperatures above 175°C
   - Based on extrapolation of measurements on larger die, the warpage of a single die of 8x8mm can be predicted to be 0.8x0.5 µm when heated up to 250°C

6. Wafer debonding at EVG
   - Schematic image of EVG debonding flow
   - Successful debonding, cleaning and handling of 25 µm thin product wafers

7. Wafer thinning at Amkor
   - Edge View – No Die Crack
   - Sliced thinning process
   - Die handling

8. Sample parametric data for IMEC-thinned wafers
   - Long channel, thin gate PMOS
   - No parameter shift observed in wafers thicker than 50 µm

9. Functional test results from packaged product
   - DAC characterization
   - Failures unrelated to thinning

10. Conclusions
    - Observed shift in long channel device performance in wafers 50 µm and thinner
    - No change in product performance due to thinning
    - Considering maturity of process, yield loss in packaging of thinned, stacked die is negligible
    - Production-level metrology needs to be developed
    - Two wafer thinning techniques show acceptable results based on electrical data
    - Industry needs solutions for handling and shipping thinned wafers
Bandwidth Optimization for Through Silicon Via (TSV) bundles in 3D Integrated Circuits

Awet Yemane Weldezion*, Roshan Weerasekera+, Dinesh Pamunuwa+, Li-rong Zheng*, and Hannu Tenhunen*

*School of Information and Communication Technologies, Department of Electronics, Computer, and Software Systems, KTH Royal Institute of Technology, Forum 120, Kista SE 16440, Sweden
{aywe, lirong, hannu}@kth.se

+Centre for Microsystems Engineering, Department of Engineering, Lancaster University, Lancaster LA1 4YW, UK
{r.weerasekera, d.pamunuwa}@lancaster.ac.uk

Abstract

Through silicon vias (TSVs) are the backbone of 3D integration technology connecting vertically stacked ICs. Parallel TSVs in the form of bundles are used for vertical signaling. In this paper, we present the ways of maximizing the total bandwidth of a TSV bundle placed in a fixed area by varying the density and the geometries. The ways of optimizing the total bandwidth using analytical methods for a bundle of TSVs placed in a structure with a fixed area and length are examined. The result shows that for uniformly distributed TSVs, maximum bandwidth by proportional placement of fewer number of TSV in the bundle can be achieved.

1. Introduction

TSVs are primarily used as wires for vertical signaling while at the same time as a mechanical support strengthening the 3D structure. In interconnect analysis; a local wire is characterized by short wire length as compared to a global wire that possibly has its length increased with scaling [1]. TSVs can be treated as local wires reasonably because like local wires, TSV length gets shorter with scaling. Moreover, a TSV driven by a device, the TSV RC effect is insignificant and the overall device resistance is the dominant factor. According to the ITRS, the diameter of high density - HDTSV decreases by a factor of half per year [2] making TSVs to be densely placed into a small area and in consequence the coupling effect increases sharply.

These changes in density and TSV geometries define the overall signaling characteristics such as cross-talk, propagation delay and ultimately the total bandwidth. Thus the challenge is how to effectively distribute few TSVs in a given area and attain maximum total bandwidth with less coupling. There are other relevant works such as [6] with focus on bandwidth optimization method for planar wires in 2D integrated circuits. In this paper, we extend the bandwidth optimization method for 3D integrated circuits, for the first time. We examine the ways of optimizing the total bandwidth based on the compact models for a bundle of TSVs placed in a structure with a fixed area and length. Given the assumptions that the TSV behaves like local wire, a lumped model is accurate enough to estimate the delay and hence the bandwidth.

Figure 1. $3 \times 3$ TSV structure (a) top view for $N_{TSV} = 9$ (b) single TSV
2. TSV delay and bandwidth modelling

The bundle of TSVs shown in Figure 1, with number, \(N_{tsv} = 3 \times 3\), is made out of copper vias embedded in silicon substrate. Each via is surrounded by SiO\(_2\) dielectric barrier \(d_b\). In the compact models described in [3], it has been shown that for such TSV structure the RLC extraction equivalent values can be deduced through analytical method for a specific range of radius, \(10 \mu m < r_v < 45 \mu m\) and spacing, \(40 \mu m < s_v < 180 \mu m\). The total capacitance \(C_t\) and inductance \(L_t\) is the sum of the self capacitance \(C_s\) and inductance \(L_s\) and mutual capacitance \(C_m\) and inductance \(L_m\), respectively. The effective TSV resistance \(R_t\) remains the same in all conditions.

\[
R_t = \frac{\rho_{cu}}{\pi r_v} \Omega
\]  
(1)

The via pitch, \(p_v\), between two TSV for a given \(N_{tsv}\) can be varied by changing the area of the TSV structure \(A_{str}\) of Figure 1a. The \(p_v\), is uniformly distributed and measured as the distance between two lateral TSVs from center to center.

\[
p_v = \frac{\sqrt{A_{str}} - 2(r_v + d_b)}{\sqrt{N_{tsv}} - 1} \text{ for } N_{tsv} > 2
\]  
(2)

In the compact model [3], it has been shown that the total capacitance \(C_t\) and total inductance \(L_t\) of any TSV in a uniformly distributed bundle can be expressed in a closed form equation as a function of the radius, \(r_v\), the distance between two TSVs, i.e., pitch, \(p_v\), and TSV length, \(l_v\).

\[
C_t = \frac{63.34\epsilon_0 l_v}{ln(1 + 5.26\frac{l_v}{r_v})} + \frac{\epsilon_0 l_v}{ln(\frac{\rho_{cu}}{2p_v})} [k_1(p_v r_v)^{k_2} + k_3(\frac{p_v}{l_v})^{k_4}]
\]  
(3)

\[
L_t = \frac{\mu l_v}{2\pi} \ln(1 + \frac{2.84\epsilon_0 l_v}{\pi r_v}) + \mu l_v \ln(1 + 0.34\frac{l_v}{p_v}) + \mu l_v \ln(1 + 0.24\frac{l_v}{p_v})
\]  
(4)

where \(k_i\) is a constant extracted through simulation based on the geometric position of each TSV in the structure. The compact model equations are used to derive the Elmore delay, \(T_{d,RC}\), for the lumped model in Figure 2 [4].

Figure 2. RC model of TSV with driver and load

This model represents a single TSV in the bundle. For simple series RC dominant wire the Elmore delay can accurately estimate the delay. The Elmore delay for rise time between 0%-50% is the sum of individual node delays of the circuit can be described as follows.

\[
T_{d,RC} = \ln(2)\Sigma RC
\]  
(5)

Now based on these equations, we extract the time delay for the equivalent RC circuit shown in Figure 2 as follows.

\[
T_{d,RC} = (R_{drv}(C_{drv} + C_t + C_L) + R_t(C_t + C_L)) \ln(2)
\]  
(6)

Equation (6) contains two terms, delay from the driver throughout the wire and delay due to the total wire impedance. With further rearrangement

\[
T_{d,RC} = \ln(2)(R_{drv}C_{drv} + (R_{drv} + R_t)C_L + (R_t + R_{drv})C_t)
\]  
(7)

Now if we substitute \(R_t\) by equation (1) and \(C_t\) by equation (3), we get the time delay for each TSV in the worst case.

\[
T_{d,RC} = \ln(2)((R_{drv} + (R_{drv} + (\rho_{cu} \frac{l_v}{\pi r_v})^{k_1} C_t)) + \ln(2)((R_{drv} + (\rho_{cu} \frac{l_v}{\pi r_v})^{k_2} (\frac{63.34\epsilon_0 l_v}{ln(1 + 5.26\frac{l_v}{r_v})}) + \frac{\epsilon_0 l_v}{ln(\frac{\rho_{cu}}{2p_v})} [k_1(p_v r_v)^{k_2} + k_3(\frac{p_v}{l_v})^{k_4}]))
\]  
(8)

Given the assumptions that the TSV behaves like local wire, the lumped model is accurate enough to estimate the delay and hence the bandwidth. The bandwidth (BW) of N numbers of TSV (\(N_{tsv}\)) in any given structure can be estimated by measuring the rate of bit streaming of each TSV multiplied by the total number of TSVs in the bundle.

\[
BW_{max} = \frac{N_{tsv}}{T_{d,RC}}
\]  
(9)

This is the maximum BW that can be achieved considering the worst case delay found in the mid-TSV of Figure 1a. To be in the safe region, it is sufficient approximation to assume that the average BW is only one-third less than that of the maximum.

\[
BW_{avg} = \frac{2}{3} BW_{max}
\]  
(10)

Substituting \(N_{tsv}\) by re-arranging equation (2) and \(T_{d,RC}\) by equation (8) and with some simplification gives...
us the BW expressed in terms of $BW = F(r_v, p_v)$. The model is the RC equivalent of a TSV found in the middle of the bundle which is worst affected by the coupling effect. Thus we derive the worst case expression for the total bandwidth from the Elmore delay, $T_d, RC$, and the number of TSVs in the bundle, $N$ as a function of $r_v$ and $p_v$ which is shown in equation 11. To get the maximum bandwidth, we use differential equation to derive the local maxima of this expression or use computational tools such as Matlab as shown in Figure 3(a) and b as described in [6]. From these Figures the BW within range can be evaluated and the maximum BW can be easily extracted.

3. Discussion and Conclusions

In practice, compared to the long horizontal metal wires, TSVs behave more like local wires. They are short and fat wires with radius in micro meter range. It is conclusive from the analysis that the method of BW optimization is effective for ideal drivers to attain maximum bandwidth with number of TSVs fewer than the maximum possible $N_{tsv} < N_{max}$. For example, the plot in Figure 3(a) and b. shows that with ideal drivers in a fixed area of $A = 1 \text{ mm}^2$, though it is possible to distribute maximum of 296 TSVs, the maximum bandwidth is found only when the number of TSVs = 64. The coupling effect increases when more than 64 TSVs are stacked affecting the BW negatively, where as less than 64 TSVs do not give the maximum possible BW. When both ends of the TSVs are wired to a driver, the maximum bandwidth is achieved only by stacking as many TSVs as possible. This is because for real drivers, the delay is dependent on the driver impedance and the TSV impedance change has small significance on the total delay. The maximum number of TSV is achieved only when the smallest radius TSV with pitch distance is used. With scaling, when the TSV impedance is significant enough to drop the delay ($R_t >> R_{drv}$), this optimization method can be used to design and fabricate a circuit with fewer TSVs and higher bandwidth.

4. Acknowledgments

Funding support from the European Commission under grant FP7 – ICT – 215030 (ELITE) of the 7th framework program is gratefully acknowledged.

References


Through silicon vias (TSVs) are the backbone of 3D integration technology connecting vertically stacked ICs. Parallel TSVs in the form of bundles are used for vertical signaling. We propose ways of maximizing the total bandwidth of a TSV bundle placed in a fixed area by varying the density and the geometries. The ways of optimizing the total bandwidth using analytical methods for a bundle of TSVs placed in a structure with a fixed area and length are examined.

The result shows that for uniformly distributed TSVs, maximum bandwidth by proportional placement of fewer numbers of TSVs in the bundle can be achieved.

### TSV compact parasitic models

\[
C_T = \frac{\frac{4.344}{\ln (1+4.24\frac{r_c}{d})}}{2}\left[k_1 \left(\frac{d_s}{r_c}\right) + k_3 \left(\frac{d_s}{r_c}\right)^3\right]
\]

\[
R_T = \frac{\rho_0}{d_s}\sqrt{T}
\]

### Equivalent Circuit

![Equivalent Circuit Diagram](image)

### TSV delay & bandwidth modelling

TSVs in micrometer range are short and fat wires can be characterized as lumped local interconnects.

\[
T_{del,LC} = \ln (2) \sum RC
\]

(Elmore delay model for lumped wires)

\[
T_{del,LC} = \frac{1}{2}\left(\frac{r_c}{d} + \frac{N_{TSV} \times (r_c + d_s)}{d_s}\right)
\]

(Elmore delay equivalent for Figure 2)

\[
BW_{max} = \frac{1}{2}C \sqrt{r_c d_s}
\]

(Average bandwidth estimation)

\[
BW_{avg}(r_c, d_s) = \frac{1}{2}C \sqrt{r_c d_s}
\]

(Bandwidth expression in terms of geometric parameters)

### Methods of finding optimized bandwidth

The maximum bandwidth of \(BW(r_c, d_s)\) expression and the corresponding geometric values, can be found in two ways.

1. Use differential equation to derive the local maxima within a range.
2. Use computational tools such as Matlab, find the maximum point in range (eg. Figure 3) and the corresponding geometric values.

### Conclusions

- Increasing TSV density introduces more coupling effect, worst delay and degrades bandwidth
- With scaling, when a radius \(r_c\) in nanometer range is possible to process and when the TSV impedance is significant enough to drop the delay \((R_l >> R_{imp})\), this BW optimization method can be used to design and fabricate 3-D chips with fewer TSVs and higher bandwidth.
- Shielding with dummy TSVs or thicker dielectric barrier which can potentially increases bandwidth is also worth to investigate.

### Contact

**Awet Yemane Weldezion**
Royal Institute of Technology, KTH, IC/TC/ECOS/ESD, Isafjordsgatan 39, 8th Floor, Forum 120, 16440 Kista, Sweden
Phone: +46 8 790 41 97, Fax: +46 8 751 17 93,
Email: awwe@kth.se, http://www.ipack.kth.se/awwe/
ABSTRACT

After several years of research, 3D Interconnect technology utilizing Through Silicon Via (TSV) connections is moving from universities to the industry and is making inroads into first commercial devices such as image sensors. In this move, cost effective manufacturing methods are one of the key enablers that allow for adoption of this revolutionary technology across a broad range of applications. One specific challenge is the cost effective manufacture of the actual TSV structures.

Today’s deposition techniques for barrier and seed layers are not always standing up to the challenges presented by substrates with TSV type structures. They same is true for conventional spray coating techniques that have been used to lay down layers of organic materials such as photo resists or dielectrics over topographically challenged substrates.

This poster presents a new method for coating substrates with these deep structures. Driven by customer requirements, specialized and unique processing equipment has been developed with the goal to meet requirements of coating substrates with DRIE etched via structures with vertical sidewalls. The newly developed NanoSpray technology covers a wide range of via geometries with aspect ratios up to 1:5 and via diameters ranging from tens of microns to two hundred micron and allows for conformal coverage of the surface topography. The layer thickness on the bottom and the sidewalls of the vias, the top edges and the top surface layer of the substrate can be altered by adjusting the process parameters.

Another challenge for 3D integration is seen into vertical stacking and integration of individual wafers on top of each other as well as most recently also the chip-to-wafer integration scheme, which will be explained and discussed as well.

For aligned permanent wafer-to-wafer (W2W) bonding, EVG developed multiple direct and indirect alignment methods to stack various types of substrates as well as improve alignment accuracy (typically 0.5-3µm, 3σ) with minimal z-axis movement. A recently developed alignment method (SmartView®) enables the use of any unique features on the front-side of the device wafers for W2W alignment. Our unique bonding process and equipment, based on the separation of the alignment process from the bonding process, are being widely adopted for wafer-level 3D integration; Cu-to-Cu bonding, solder-based bonding, fusion bonding, polymer bonding, and hybrid bonding.

Chip-to-wafer (C2W) bonding, typically through metal-metal thermo-compression mechanisms such as Cu-Cu bonding or Cu-Sn intermetallic compound (IMC) bonding, is generally used for heterogeneous chip stacking. However, as the diffusion rate is proportional to temperature, pressure and time, it is not economically feasible to perform bonding process at a single-die level. We developed a new advanced chip-to-wafer (AC2W) bonding concept which splits the bonding process into two sub-processes. The temporary pre-bonding with alignment is performed on a pick-and-place machine, followed by the permanent bonding of dies as a batch process in a specially designed bond chamber. In AC2W bonding, the center position and the absolute value of total force is determined by the location and population of chips, which enables a true known-good-die (KGD) stacking.

As the thin (<100µm) silicon wafers which are commonly used for TSV formation also exhibit increased instability and fragility. The lack of mechanical stability and the increased fragility present a major challenge to maintaining high yield levels in volume manufacturing environments.

A reliable support and handling solution is needed to overcome the above-mentioned challenges while maintaining yield levels compatible with low-cost, high-yield manufacturing processes. The solution of choice must enable safe, reliable handling of the substrates through back-thinning and backside processing while being compatible with existing (already installed) equipment lines and manufacturing processes.

The most accepted handling solution for UltraThin® wafers is the use of temporary bonding and debonding techniques utilizing a rigid carrier wafer to provide sufficient mechanical support. Once the product wafer is temporarily bonded to the carrier wafer, it is ready for backside processing including back-thinning, through-silicon via (TSV) formation, etc. The product wafers can either be pre-processed wafers with CMOS devices as well as e.g. substrates with high-topographies like solder-balls.
After completion of the backside processing steps, the product wafer can be released from the carrier wafer and proceed to final packaging processes which will be covered as well.

The EMC-3D consortium has proven the process capability and reliability, where details will be provided as well.
EV Group

Wafer-Level Based Manufacturing Technologies for Realization of TSV and 3D-Based Applications

Stefan Pargfrieder
Daniel Burgstaller
Otto Bobenstetter
Bioh Kim

Design, Automation & Test in Europe
20-24 April, 2009 Nice, France
I. W2W Bonding

II. Advanced C2W Bonding

III. TSV NanoSpray Coating
IC Market Drivers

Density
Achieving the highest capacity per volume

Form Factor
driven

Performance
driven
Mid-term driver > 2010

Cost driven
Long-term driver > 2012

3D vs. “More Moore”
Can 3D be cheaper than going to the next lithography node?

“More than Moore”
Heterogeneous Integration
Co-integration of RF + logic + memory + sensors in a reduced space

Electrical Performances
Higher interconnect speed and reduced parasitics

3D IC
Optimum Market Access Conditions

Form Factor driven
Short-term driver > 2008

Source: Yole Development

3D IC

CPU
GPU

MEMS
RF

CIS
Flash

DRAM

Source: Yole Development

Short-term driver
> 2008

Long-term driver
> 2012

Mid-term driver
> 2010

Density
Achieving the highest capacity per volume

More than Moore
Heterogeneous Integration
Co-integration of RF + logic + memory + sensors in a reduced space

Electrical Performances
Higher interconnect speed and reduced parasitics

3D vs. “More Moore”
Can 3D be cheaper than going to the next lithography node?

“More than Moore”
Heterogeneous Integration
Co-integration of RF + logic + memory + sensors in a reduced space

Electrical Performances
Higher interconnect speed and reduced parasitics

3D IC
Optimum Market Access Conditions

Form Factor driven
Short-term driver > 2008

Source: Yole Development
3D Integration

Multi-chip Integration Scheme: SoC vs. 3D SiP

2D Interconnect
Long connection, Shared bus

SoC Integration
Large die, No outside connections

3D SiP Interconnect
Short, direct lines

System on Chip

Cost / Function
Time to Market

System Complexity

Source: ITRS
TSV Interconnects

TSV electrodes can provide vertical connections that are the shortest and most plentiful.

Wafer Level 3D Integration

- Long Connection: High Power consumption
- Short Connection: Low Power consumption
- Low Density: High Density
- Poor Heat Dissipation: Good Heat Dissipation
- RC Delays: Reduced RC Delays
- High Impedance: Low Impedance
- Large Area: Smallest Area
- Challenging Interposers: Simple Interposers
- I/O Pitch limitations: Less I/O Pitch limitations

TSV interconnects provide solutions to many limitations of current SiP and chip stacking methods.
Requirements for TSV Integration

Robust and precise thinning process flow
Handling concepts for a thin wafer
Electrical interconnects through a thinned wafer
  - Via etch: shape, angle, and scallop control
  - Insulator/Barrier/Seed: conformity and adhesion control
  - Metal fill: fill robustness and speed control
  - Metal removal: surface smoothness and overpolish control
Suitable bonding process: alignment, bonding, and dicing

Need cooperation with industrial leaders in equipment, materials, and technology !!!

Achieved CoO < $200 per wafer
TSV Adoption by Industry

3D WLP Encapsulation

Source: SAIT

3D Interposer Module

Source: NXP

Only reliable bonding enables reliable 3D architectures!!!
Bonding Methods for TSV Integration

Direct Oxide Fusion (with metal)

Solder-based (IMC)

Polymer Adhesive (BCB)

Cu-Cu

Hybrid (Cu/BCB)

Courtesy of Ziptronix DBI™

Courtesy of Fraunhofer IZM

Courtesy of RPI

Courtesy of Tezzaron

Courtesy of RPI
I. Wafer-to-Wafer (W2W) Bonding

- High Throughput Platform
- Sub-µ Accuracy Alignment
- Plasma Activation Cleaning
- Modular Design
- Multiple Processes
- Process Controllability
Process Controllability

40, 60, 100kN

\[
< 1\% \text{ Force Accuracy} \\
< 2\% \text{ WEC Accuracy}
\]

Up to 550°C

+/-0.5°C*

with 60°C/min

& -23°C/min

* Based on 300mm Cu-Cu bond configuration and process

< 1 E-5 mbar

Up to 300mm
Proprietary alignment technique for F2F bonding with non-IR transparent wafers

No Z-travel
- Locate bottom wafer alignment marks with top objectives
- Digitize image
- Store position

No re-focussing
- Align top wafer to digitized image

Perfect result
- Restore bottom wafer position
- Bring wafers in contact
SmartView® Alignment: Accuracy

Misalignment

~400 measurements

< 1µm post-bond alignment

| dx, max | < 250 nm |
| dy, max | < 250 nm |

\[ dx, 3\sigma \leq 120 \text{ nm} \]

\[ dy, 3\sigma \leq 120 \text{ nm} \]
Cu-Cu Bonding

Bond strength vs. Bonding process & temp.

- Acetic Acid Clean, H₂/Ar Purge
- HCl Clean, H₂/Ar Purge
- HCl Clean, N₂ Purge

Minimum Limit for Processing

Our data (9.46J/m² @413°C)

R. Tadepalli and C. V. Thompson
Proc. IITC, 2003

4 point bending method

Original bonding interface

R. Tadepalli and C. V. Thompson
Proc. IITC, 2003
Plasma-assisted Direct Oxide Bonding

Low Temperature Plasma Activation

Si to SiO₂ wafer bonding:
surface energy vs. annealing time (at 300°C)

IR transmission images of a blank and a patterned PECVD wafers, 200mm diameter, bonded with plasma activation (surface energy: 2.0 – 2.7 J/m²)
Polymer Bonding

Scanning acoustic microscopy (Sonix AW Vision 3000) of 200mm Si wafers bonded with BCB

Courtesy of RPI

CMOS SOI

BCB

Si

SiO2

BCB

Cu

SiO2

Si

1 μm

 Courtesy of Freescale

 Courtesy of RPI
II. Advanced Chip-to-Wafer (AC2W) Bonding

**Thermo-compression Bonding**
with controlled force application
and chip thickness tolerance

**High Throughput**
with two subsequent processes

**Cost-effective Technology Mix**

**High Reliability**
Gang Bonding
AC2W Process Flow

1. Temporary Bonding (Flip-Chip Bonder)
   - Velocity: 3000 chips/hour (avg)
   - Alignment: 10µm (3σ)

2. Permanent Bonding (Chip-to-Wafer Bonder)
   - Throughput: 1-2 wafers/hour
   - Up to 300°C, up to 300mm, 20kN

Force

Temperature

Ambient
Features vs. Benefits

**Features**
- Target bonding methods: **Thermo-compression** (Cu-Cu, Cu-Sn IMC, and Hybrid) with a long process time
- Controlled force application
- Chip thickness tolerance by compliant layer

**Benefits**
- Technology mix
- Higher functional density
- Reduced cost

*Courtesy of Infineon*

*Courtesy of RPI*
Post-Bond Alignment Accuracy

Chip Alignment @ EVG 540 C2W Bonder

Alignment accuracy ≤ ±3µm even with solder-based bonding !!!
Recent data : ≤ ±2µm achievable
III. TSV NanoSpray Coating

Photosensitive masking layer

Passivation layer

Organic insulation layer
NanoSpray Coating

- Unique spray process that is based on a spray mist created by ultrasonic nozzles
- Significant improvement in refined dispense and targeted positioning of the spray stream
- Homogeneous coatings of features up to 300µm deep and down to 20µm diameter with various aspect ratio
- Positive and negative resists have been qualified.
Masking Layer for Via Bottom Etch

100µm×200µm

a) Dry etching of through-wafer vias
b) Deposition of dielectric layer
c) Spray coating of photoresist
d) Patterning of dielectric layer
e) Metallization using Cu plating
Summary

W2W Bonding  AC2W Bonding

TSV NanoSpray Coating

Memory Stack

CIS

Memory on Logic

Source: Intel

Source: Samsung
Thank You!

www.EVGroup.com

Triple i - The key to your success
SOC Test Architecture and Method for 3D-IC

Chih-Yen Lo, Yu-Tsao Hsing, Li-Ming Denq, and Cheng-Wen Wu
Department of Electrical Engineering
National Tsing Hua University
Hsinchu, Taiwan

Abstract

Three-dimensional (3D) integration has been considered the most promising solution to extend the life of Moore’s law in semiconductor manufacturing technology. However, these stacked dies will face the severe problem of exponential decay in their quality if the currently employed post-bond testing is not changed. In Figure 1(a), We propose a practical system-on-chip (SOC) test strategy for 3D-IC based on reconfigured wafer-on-wafer 3D integration technology [1]. To consider the yield issues of 3D-IC manufacturing, we perform known-good-die (KGD) test before die stacking. The details of finding KGD is shown in Figure 1(b). By our special arrangement in SOC test integration, we can flexibly execute logic or memory testing with simple test configuration and small area overhead. After KGDs are obtained, we are ready to mount them layer by layer. Every time a new KGD is mounted on the original stacked chip, we perform through-silicon-via (TSV) test for 3D interconnect verification between the two top-most layers. If necessary, our proposed test scheme also supports extra KGD test in every layer of the stack with neither extra test circuit nor modified test application. Therefore, the yield of the stacked chips can be further guaranteed.

In order to fulfill the requirement mentioned above, we propose a test architecture named Test Access Control System for 3D-IC (TACS-3D). Figure 2(a) shows the block diagram of TACS-3D which consists of an extended JTAG/IEEE 1149.1 Test Access Port (TAP) Controller and multiplexer-based test access mechanism (TAM) buses. For applying logic testing, TACS-3D features IEEE 1500 Wrapper Control, hierarchical test control, at-speed test (for
transition faults), functional and scan test, heterogeneous test protocols, etc. [2]. In order to save the control signal pins/TSVs, TACS-3D is further extended to support memory BIST (MBIST) in the stacked chips by adding MBIST Start Register (MSR) in the Test Controller and defining a special TAM session. Furthermore, the 3D interconnect verification can be easily applied through the operations of Single/Cascade Register (SCR) and Bypass Flag Register (BFR). Figure 2(b) shows the detailed setting of pre-stack KGD test, parallel TSV test and optional KGD test for the bottom layer of the stacked chips. Based on our proposed test scheme and TACS-3D, the yield issues of the 3D-IC can be easily done by flexibly applying SOC test before and after dies are mounted. In addition, shorter overall test time is expected due to uniform test interface and small test-control requirements.

Figure 2: (a) Block diagram of TACS-3D and (b) examples of stacked-chip setting.

References


Abstract:

3D technologies hold the promise to further enable system performance increase in a time where device scaling has become increasingly challenging. Networks on chip (NoCs) have been proposed to address the complexity of interconnecting an ever-growing number of cores, memories and peripherals. NoCs are a promising choice for implementing scalable 3D interconnect architectures. However, the development of 3D NoCs is still at an early development stage.

Design trade-offs and design technology support for 3D NoCs are yet to be explored in depth. To be useful for a NoC infrastructure, a vertical wire should not be used in isolation; instead, to simplify routing, it is better to create buses of such wires. Given the physical proximity of the Through-Silicon Vias (TSVs) in link, concerns related to capacitive coupling within such buses may arise, therefore increasing the load of the structure.

The aim of this work, is to present a semi-automated design flow for 3D NoCs. The first contribution of our work is the integration in a physical design environment for digital circuits of a circuit-level model for vertical interconnects (TSVs), based on accurate three-dimensional parasitic extraction. This is a requirement for performing accurate analysis, post P&R, of a digital design macro with vertical connectivity. As a second main contribution, we extend a two-dimensional NoC switch architecture with support for vertical links. Our third contribution is the development of a prototype design flow for automatic instantiation of three dimensional NoCs. Finally, we present a case study where a planar NoC topology is folded and implemented across two chip layers.

In order to evaluate our method, first, we conducted TSV analysis for different link configuration, varying both the number of wires for each link, and the technology parameters, like TSV pitch, height, diameter, and oxide thickness between TSV and bulk silicon. Normalized RC extracted data has been compared with the capacitance and resistance of horizontal wires (Metal_1 to Metal_9), therefore showing a gain of two orders of magnitude in RC delay improvement.

In the second step, we evaluate the impact of 3D interconnect on a typical communication channel composed of 2 switches and a planar link and a vertical link. To do so, we implemented both 2-D and 3-D versions of this communication channel. Then we extracted parasitic (R and C) information from these two implementations to compare the performance of two different designs. Based on our experiments (Figure 1) the 3-D implementation had around 10 percent improvement over the 2-D one, thanks to the advantage of TSVs and the shorter links.

Figure 1: Maximum frequency achievable by switches in 2D and 3D flows, for varying switch cardinalities and two different flow controls.

Moreover, to facilitate the 3-D NoCs design flow we also propose a semi-automatic flow to map a fully 2 dimensional NoCs into a 3-D one by splitting the 2-D design into multiple stacked designs. As a validation of our flow, we present a NoC implementation based on a 2D 3x2 quasi-mesh and migrate it to a 3D arrangement (Figure 2). The 3D mapping is achieved by splitting in two halves the mesh and overlapping them in separate chip layers, with communication achieved through TSVs. The stacked topology has exactly the same functionality of the two-dimensional implementation.

Figure 2: Layouts for (a) the 2D 3x2 mesh, (b) one of the halves of its 3D re-implementation, and (c) shows a detail of the switch and the corresponding vertical links.
**EXPERIMENTAL RESULTS**

- **Semi-Automatic Design Flow**
- **Modeled Component Library → Accurate Synthesis and P&R**

**3D Flow**

- Topology Partitioning
  - Cut along the links to minimize TSV's count and to simplify the routing
  - Layers are connected by vertical vias

**Timing Models**

- **T-Network for each TSV**
- **Coupling and Crosstalk**
- **Contact resistance between the bonding pads (ThermoCompression)**

**STALL/GO**

- Delay is given by a sum of parasitics
- Horizontal wire to via base
- Via delay (includes R of bases)
- Horizontal wire from via top
- Load
- For a whole via of 50µm, delay is 16/16.5ps (SOI/bulk)
- For a 1.5mm horizontal link, delay is around 200ps

**ACK/NACK**

- User
- Compiler
- Platform
- System