Seventh IEEE International Workshop on Testing Three-Dimensional, Chiplet-Based, and Stacked ICs

3DC-TEST

Electronic Workshop Digest

virtual interactive event, in conjunction with

International Test Conference
November 3-5, 2020
WORLD'S PREMIER ELECTRONICS TEST CONFERENCE

sponsored by
the IEEE Philadelphia Section
in concurrence with
the IEEE Computer Society’s Test Technology Technical Council
Welcome!
The 3DC-TEST Workshop focuses exclusively on test of and design-for-test for three-dimensional, chiplet-based, and stacked ICs (3D-SICs), including systems-in-package (SiP), package-on-package (PoP), 3D-SICs based on through-silicon vias (TSVs), micro-bumps, and/or interposers. While these stacked ICs offer many attractive advantages with respect to heterogeneous integration, small form-factor, high bandwidth and performance, and low power dissipation, there are many open issues with respect to testing such products. The 3DC-TEST Workshop offers a forum to present and discuss these challenges and (emerging) solutions among researchers and practitioners alike.

This edition revives a workshop last held in 2015 with the new interest in chiplets. The first six editions of this workshop were very successful: many attendees, strong technical programs with invited talks, reviewed papers, posters, demos, and corporate supporters. At the time of writing, all signs indicate that the seventh edition of the 3DC-TEST Workshop is going to be successful as its past six ones! Based on a mix of invited talks and presentations selected out of a large collection of submitted papers and extended abstracts, we have been able to put together an attractive technical program, containing presentations, papers, and panels. The budget of the Workshop has been helped by the generous financial contributions of six corporate supporters: Advantest, Galaxy Semiconductor, HiSilicon, proteanTecs, Synopsys, and Teradyne. Attention for the Workshop was boosted with the help of our three Media Partners: 3DinCites.com, ChipScaleReview.com, and MEPTEC.

We thank all who enabled this workshop: our moderators, presenters, and panelists, the workshop’s organizing and program committees, and the corporate supporters and media partners. As always, the real success of a workshop depends on the active participation of all workshop attendees. Here we count on contributions from you all! Enjoy the seventh edition of your 3DC-TEST Workshop!

General Co-Chair
Erik Jan Marinissen
IMEC (Belgium)

General Co-Chair
Yervant Zorian
Synopsys (USA)

Program Chair
Bapi Vinnakota
Broadcom (USA)

3DC-TEST takes place in conjunction with the IEEE International Test Conference (ITC) and is sponsored by the IEEE Philadelphia Section in concurrence with the IEEE Computer Society’s Test Technology Technical Council (TTTC).
Organizing Committee

**General Co-Chair:** Erik Jan Marinissen – IMEC (Belgium)

**General Co-Chair:** Yervant Zorian – Synopsys (USA)

**Program Chair:** Bapi Vinnakota – Broadcom (USA)

**Industrial Chair:** Marc Hutner – Teradyne (USA)

**Panel Chair:** E. Jan Vardaman – TechSearch International (USA)

**Finance Chair:** Chen-Huan Chiang – Intel (USA)

**Publicity Chair:** Françoise von Trapp – 3DInCites (USA)

**Web Chair:** Hardi Selg – Tallinn University of Technology (Estonia)

**Virtualization Chair:** Stefano di Carlo – Politecnico di Torino (Italy)
Program Committee

Saman Adham   – TSMC (Canada)
Michael Alfano – AMD (USA)
Dave Armstrong – Advantest (USA)
Sandeep Bhatia – Google (USA)
Krishnendu Chakrabarty – Duke University (USA)
Sreejit Chakravarty – Intel (USA)
Kun Young Chung – Qualcomm (USA)
Jon Colburn    – Nvidia (USA)
Eric Cormack  – DfT Solutions (UK)
Adam Cron      – Synopsys (USA)
Alfred Crouch  – AmidaTechnology (USA)
Marie–Lise Flottes – LIRMM (France)
Ferenc Fodor  – imec (Belgium)
Paul Franzon  – North Carolina State University (USA)
Phil Garrou    – Microelectronic Consultants of NC (USA)
Sandeep K. Goel – TSMC (USA)
Alan Hales     – Texas Instruments (USA)
Junlin Huang   – HiSilicon (China)
Hailong Jiao   – Peking University Shenzhen (China)
Gerard John    – Amkor Technology (USA)
Hongshin Jun   – Juniper Networks (USA)
Shuichi Kameyama – Ehime University (Japan)
Chien–Mo Li    – National Taiwan University (Taiwan)
Alan Liao      – FormFactor (USA)
Amit Majumdar  – Xilinx (USA)
Teresa McLaurin– ARM (USA)
Benoit Nadeau–Dostie – Mentor Graphics (CAN)
Brandon Noia   – AMD (USA)
Christos Papameletis – Cadence Design Systems (USA)
Mike Ricchetti  – Synopsys (USA)
Saghir Shaikh  – Broadcom (USA)
Raffaele Vallauri – Technoprobe (Italy)
Pascal Vivet    – CEA–LIST (France)
Michael Wahl   – University of Siegen (Germany)
Corporate Supporters

The 3DC-TEST Workshop gratefully acknowledges the financial support from the following companies.

Media partners of the 3DC-TEST Workshop 2020:

ChipScaleReview.com  The Future of Semiconductor Packaging

MEPTEC  THE NEXT GENERATION
Thursday, November 5, 2020

**Session 1: Opening**  
17:00 – 18:30 (EST)

**Moderator:** Saman Adham - TSMC (Canada)

- **17:00h:** Welcome Address  
  General Co-Chair: Erik Jan Marinissen – IMEC (Belgium)

- **17:10h:** Introduction to the Workshop Program  
  Program Chair: Bapi Vinnakota – Broadcom (USA)

- **17:15h:** Keynote Address Silicon Photonics Chiplets for Scaling AI and the Cloud – Technology, Design, and Test  
  Joris Van Campenhout – Fellow/Director – IMEC (Belgium)

- **18:00h:** The SEMI Heterogeneous Integrated Roadmap Test Working Group Update  
  Zoë Conroy – Cisco Systems (USA)

Friday, November 6, 2020

**Session 2: Probing Chiplets and 3D Dies**  
10:00 - 11:00 (EST)

**Moderators:** Gerard John – Amkor Technologies (USA)  
Saghir Shaikh – Broadcom (USA)

- **10:00h:** Probing Complexities of 3D-Stacked ICs – A Test Engineers’ Perspective  
  Ferenc Fodor, Bart De Wachter, Arnita Podpod, Michele Stucchi, Erik Jan Marinissen – IMEC (Belgium)

- **10:15h:** HBM2 Probing Challenges and Probe Card Architecture  
  Raffaele Vallauri, Alessandro Antonioli, Flavio Maggioni – Technoprobe (Italy)

- **10:30h:** A New Age of IC Packaging – Test Complexity and Coverage on Advanced Packages and HBM  
  Quay Nhin – FormFactor Inc. (USA)

- **10:45h:** Known Good Dies (KGD) Strategies Compatible with Direct Hybrid Bonding  
  Emilie Bourjot*, Paul Stewart, Clément Castan, Loic Sanchez, Gaelle Mauguen, Yorrick Exbrayat, Viorel Balan, Nicolas Bresson, Amandine Jouve, Frank Fournel, Florence Servant, Severine Cheramy – CEA-LETI (France); Nicolas Raynaud, Pascal Metzger – SET Corporation (France); Pascal Vivet – CEA-LIST (France)

**Session 3: Chiplet DfT and BIST**  
11:00 - 12:00 (EST)

**Moderators:** Adam Cron – Synopsys (USA)  
Dheepak Jayaraman – Facebook (USA)

- **11:00h:** Managing test & Repair of Die-to-Die High-Speed Interfaces in the Chiplet Era  
  Mike Ricchetti, Gurgen Harutyunyan, Yervant Zorian – Synopsys (USA)
### Session 4: The New 3D-DfT Standard: IEEE Std 1838™

Moderators: Vivek Chickermane – Cadence Design Systems (USA)
Pascal Vivet – CEA-LIST (France)

12:00h: IEEE Std 1838™ Introduction and the Move from a 1500-Centric TAM
Adam Cron – Synopsys (USA)

12:15h: Applying IEEE Std 1838™ to a 3DIC – A Case Study
Teresa McLaurin – ARM (USA)

12:30h: Leveraging Lessons-Learned on 2D-SOCs in Designing Parallel TAMs For 3C-SICs Based on IEEE Std 1838’s Flexible Parallel Port
Erik Jan Marinissen – IMEC (Belgium)

12:45h: Discussion

---

### Lunch Break

13:00 – 14:00 (EST)

---

### Session 5: Intel Foveros

Moderators: Sreejit Chakravarty – Intel (USA)
Teresa McLaurin – ARM (USA)

14:00h: Intel Foveros Technology: DFT And HVM Test Strategy
Wei Ming Lim, Terrence Huat Hin Tan, Sook Kwan Cheah, Kian Lek Koay, Sreejit Chakravarty – Intel (USA)

14:15h: Who’s at Fault? A Creative Way to Isolate and Debug Internal IO Failures
Devanraj Letchumanan, Ahmad Hisyamuddin Arshad – Intel (Malaysia)

14:30h: Pre-Silicon Validation Methodology Breakthrough for 3D-IC
Yip Wai Loon, Ng Hock Thien, Teo Bian Sim – Intel (Malaysia)

14:45h: Discussion

---

### Session 6: 3D Chiplets Novel Approaches

Moderators: Marc Hutner – Teradyne (Canada)
Rajamani Sethuram – NVidia (USA)

15:00h: Bunch of Wire (BoW) Inter-Chiplet Link Testing and Loopbacks
Shahab Ardalan – Ayar Labs; Marc Hutner – Teradyne (Canada);
Bapi Vinnakota – Broadcom (USA)
As the industry moves to new architectures to achieve the economic gains previously achieved with scaling, a new era of chiplets, including 3D ICs is emerging. Despite new names for packaging options, some issues remain — including test. The availability of known-good parts and a test strategy is still important. Co-design and design-for-test are essential. Developing a solution to allow the use of open market chiplets requires new strategies. This panel examines some of the issues related to test including alternatives to probing, developments in die sort, increased adoption of BIST, and the role of redundancy.

Moderator: Jan Vardaman – President – TechSearch International (USA)

Panelists: Dave Armstrong – Director of Business Development – Advantest (USA)  
Paul Franzon – Professor, Director of Graduate Programs – NCSU (USA)  
Gerard John – Sr. Director of Advanced Eng. Services – Amkor Technologies (USA)  
Bob Patti – President – NHanced Semiconductors (USA)  
John Yi – Principal Member of Technical Staff Product/Test Engineering – AMD (USA)  

16:55h: Closure of Workshop
General Co-Chair: Erik Jan Marinissen – IMEC (Belgium)
General Co-Chair: Yervant Zorian – Synopsys (USA)

17:00h: End of 3DC-TEST Workshop 2020
7th IEEE International Workshop on Testing Three-Dimensional, Chiplet-Based, and Stacked ICs

3DC-TEST
colocated with ITC

Erik Jan Marinissen
General Co-Chair
imec
Leuven, Belgium

Yervant Zorian
General Co-Chair
Synopsys
Mountain View, CA, USA

Bapi Vinnakota
Program Chair
Broadcom
Sunnyvale, CA, USA
Opening Session

Moderator: Saman Adham – TSMC, Canada

Welcome Address
- Erik Jan Marinissen – imec, Belgium – 3DC-TEST General Co-Chair
- Yervant Zorian – Synopsys, USA – 3DC-TEST General Co-Chair
- Bapi Vinnakota – Broadcom, USA – 3DC-TEST Program Chair

Keynote Address
- ‘Silicon Photonics Chiplets for Scaling AI and the Cloud – Technology, Design, and Test’
  Joris Van Campenhout – Fellow/Director – imec (Belgium)
A Short Walk Down Memory Lane...
(in the pre-Corona era)
3D-TEST 2010: Highlights

- **Keynote:**
  ‘Testing in a New Dimension’
  *Bob Patti – CTO/Founder, Tezzaron Semiconductor, Il, USA*

- **Invited Address:**
  ‘An Integrated Approach to Design and Test of 3D ICs’
  *Brion Keller, Senior Architect, Cadence Design Systems, USA*

- **Panel Session:** ‘Challenges and Solutions in 3D Wafer Probing’
  - **Moderator:** Erik Jan Marinissen – imec, Belgium
  - **Panelists:**
    - Marc Loranger – FormFactor, USA
    - Wayne Moorhead – Scanimetrics, Canada
    - Jay Orbon – Verigy, USA
    - Dan Rishavy – TEL, USA
    - Ken Smith – Cascade Microtech, USA
    - Andy Yin – Intel, USA

---

3D-TEST 2011

---
3D-TEST Workshop – Opening Session
Erik Jan Marinissen

3D-TEST 2011: Anaheim, CA

3D-TEST 2011: Highlights

- **Keynote:**
  ‘3-D SoC Packaging for Smart Mobile Devices: Current State and Challenges’
  *Hong Hao – Samsung Semiconductor, USA*

- **Invited Address:**
  ‘3D TSV Infrastructure: Challenges and Opportunities’
  *E. Jan Vardaman – TechSearch International, USA*
3D-TEST Workshop – Opening Session
Erik Jan Marinissen

3D-TEST 2012

3D-TEST 2012: Anaheim, CA

IEEE International Workshop on Testing Three-Dimensional Stacked Integrated Circuits
Past Editions: 2010–2015: Austin, TX; Anaheim, CA; Seattle, WA – http://3dtest.tttc-events.org
3D-TEST 2012: Highlights

- **Keynote:**
  ‘The Evolution of 3D-ICs: Three Road to Production of a 6.8B Transistor FPGA’
  *Ivo Bolsens – Xilinx, USA*

- **Invited Address:**
  ‘3D-Driven System Design – Present and Future’
  *Paul Franzon – North Carolina State University, USA*
3D-TEST 2013: Anaheim, CA

3D-TEST 2013: Highlights

- **Keynote:**
  ‘3D Solutions in the Coming Age of Terabit Communication’
  *Nick Ilyadis – Broadcom, USA*

- **Panel 1:**
  ‘Requirements for 3D Volume Production Testing’
  *AMD, Cisco Systems, Qualcomm, SanDisk, TSMC, Xilinx,*

- **Panel 2:**
  ‘How Will 3D-Testing Change the Test Supply Chain?’
  *Advantest, Amkor, GlobalFoundries, IMEC, Synopsys, TEL Test Systems,
  Third Millennium Test Solutions*
3D-TEST 2014: Highlights

- **Keynote:**
  ‘3D Rock from the Sun’
  Brion Keller – Cadence Design Systems, USA

- **Invited Address:**
  ‘What a Difference a Year Makes – Looking Back and Forward’
  Herb Reiter – eda2asic, USA

3D-TEST 2015

IEEE International Workshop on Testing Three-Dimensional Stacked Integrated Circuits
Past Editions: 2010–2015: Austin, TX; Anaheim, CA; Seattle, WA – http://3dtest.tttc-events.org

7th IEEE Intl. Workshop on Testing Three-Dimensional, Chiplet-Based, and Stacked ICs (3DC-TEST)
3D-TEST 2015: Highlights

**Keynote 1:**
‘New Paradigm Shift in 3-D Design and Testing’,
Jeff Rearick – Peter Li, Anson Li – AMD, China;
Bryan Black, Michael Alfano – AMD, USA

**Keynote 2:**
‘3D Integrated CMOS-Memristor Hybrid Circuits: Devices, Integration, Architecture, and Applications’
Kwang-Ting (‘Tim’) Cheng – UC Santa Barbara, USA

**Keynote 3:**
‘Known Good Die – Fantasy Land or Tomorrow Land?’
John Carulli and TM Mak – GLOBALFOUNDRIES, USA
Opening Session

Moderator: Saman Adham – TSMC, Canada

Welcome Address

- Erik Jan Marinissen – imec, Belgium – 3DC-TEST General Co-Chair
- Yervant Zorian – Synopsys, USA – 3DC-TEST General Co-Chair
- Bapi Vinnakota – Broadcom, USA – 3DC-TEST Program Chair

Keynote Address

- ‘Silicon Photonics Chiplets for Scaling AI and the Cloud – Technology, Design, and Test’
  Joris Van Campenhout – Fellow/Director – imec (Belgium)
In 2020, No Scenic Location for 3DC-TEST...
... But We Can Still See Many Familiar Faces!
General Chair: “Thank You”
Thank You to the Program Actors

- Speakers
- Keynoter, paper presenters
- Session Chairs
- Panelists and Panel Moderator

Without you, there would not be a workshop!
Thank You, Organizing Team

General Co-Chair
Erik Jan Marinissen
imec, Belgium

General Co-Chair
Yervant Zorian
Synopsys, USA

Program Chair
Bapi Vinnakota
Broadcom, USA

Industrial Chair
Marc Hutner
Teradyne, Canada

Panel Chair
Jan Vardaman
TechSearch Intl., USA

Finance Chair
Chen-Huang Chiang
Intel, USA

Publicity Chair
Françoise Von Trapp
3DInCites, USA

Virtualization Chair
Stefano di Carlo
Politecnico di Torino, Italy

Web Chair
Hardi Selg
TU Tallinn, Estonia
Thank You, Program Committee

Saman Adham – TSMC (CAN)
Michael Alfano – AMD (US)
Dave Armstrong – Advantest (US)
Sandeep Bhatia – Google (US)
Krish Chakrabarty – Duke University (US)
Sreejit Chakravarty – Intel (US)
Kun Young Chung – Qualcomm (US)
Jon Colburn – Nvidia (US)
Eric Cormack – DfT Solutions (UK)
Adam Cron – Synopsys (US)
Alfred Crouch – Amida (US)
Marie-Lise Flottes – LIRMM (FR)
Ferenc Fodor – IMEC (BE)
Paul Franzon – NC State University (US)
Phil Garrou – MCNC (US)
Sandeep K. Goel – TSMC (US)
Alan Hales – Texas Instruments (US)
Junlin Huang – HiSilicon (CN)

Hailong Jiao – Peking University (CN)
Gerard John – Amkor Technology (US)
Hongshin Jun – Juniper Networks (US)
Shuichi Kameyama – Ehime University (JP)
Chien-Mo Li – National Taiwan University (TW)
Alan Liao – FormFactor (US)
Amit Majumdar – Xilinx (US)
Teresa McLaurin – ARM (US)
Benoit Nadeau-Dostie – Mentor (CAN)
Brandon Noia – AMD (US)
Christos Papameletis – Cadence (US)
Mike Ricchetti – Synopsys (US)
Saghir Shaikh – Broadcom (US)
Raffaele Vallauri – Technoprobe (IT)
Pascal Vivet – CEA-LIST (FR)
Michael Wahl – University of Siegen (DE)
Thank You, Sponsors
7th IEEE International Workshop on Testing Three-Dimensional, Chiplet-Based, and Stacked ICs

colocated with ITC

Erik Jan Marinissen
General Co-Chair
imec
Leuven, Belgium

Yervant Zorian
General Co-Chair
Synopsys
Mountain View, CA, USA

Bapi Vinnakota
Program Chair
Broadcom
Sunnyvale, CA, USA
Introduction to the Workshop Program

- **Day 1: Friday November 6, 2020**
  - Session 1: Opening

- **Day 2: Friday November 6, 2020**
  - 10:00h: Session 2: Probing Chiplets and 3D Dies
  - 11:00h: Session 3: Chiplet DfT & BIST
  - 12:00h: Session 4: The New 3D-DfT Standard: IEEE Std 1838™
  - **13:00h: Lunch Break**
  - 14:00h: Session 5: Intel Foveros
  - 15:00h: Session 6: 3D Chiplets Novel Approaches
  - 16:00h: Session 7: Panel ‘Test Challenges in the New 3D and Chiplet World’
Meeting Logistics

- **Zoom**
  - 3DC-TEST is using a ‘regular meeting’ edition
  - You can do everything you are used to with ‘Zoom at home’
    - Speak to the entire meeting – workshops should be interactive
      - But please “mute” if you do not want to speak!
  - Chat for questions
  - Different views

- **Slack Workspace**  www.slack.com
  https://join.slack.com/t/3dct7-2020/shared_invite/zt-i3e9hwX-0OS5pjerRTo3B8Z9SXTdOw
Get The Most Out of the Workshop!

- Interact with your Peers on **Zoom Chat** and **Slack**
  - Workshops are all about open discussions!
  - Prize for interacting on Slack channel with Authors
  - Session channels: s[2 to 6]_<name of session>
- Check out on **Slack** material from corporate supporters
  - Both videos and PDFs
  - Demo from ProteanTecs
  - Prize for leaving message on each of the vendors
- Sponsor channels: Sponsor_******

Slack Workspace:  [www.slack.com](https://www.slack.com)
[https://join.slack.com/t/3dct7-2020/shared_invite/zt-i3e9hw0l-0OS5pjreRT03B8Z9SXTdOw](https://join.slack.com/t/3dct7-2020/shared_invite/zt-i3e9hw0l-0OS5pjreRT03B8Z9SXTdOw)
Session 7: Panel

Test Challenges in the New 3D and Chiplet World

- **Moderator:** Jan Vardaman – President – TechSearch International (USA)
- **Panelists:**
  - Dave Armstrong – Director of Business Development – Advantest (US)
  - Paul Franzon – Professor, Director of Graduate Programs – NCSU (US)
  - Bob Patti – President – NHanced Semiconductors (US)
  - John Yi – PMTS Product/Test Engineering – AMD (US)
Keynote Address

Silicon Photonics Chiplets for Scaling AI and the Cloud – Technology, Design, and Test

Joris Van Campenhout – Fellow/Director
imec (Leuven, Belgium)

Abstract: Artificial intelligence and cloud computing are driving an exponentially growing demand for optical interconnect bandwidth. From the datacenter network down to the chip level, silicon photonics is a prime technology to scale optical interconnects to the desired bandwidth density (>1Tbps/mm), power consumption (<1pJ/bit), and cost (<0.1$/bit). In this presentation, we give an overview of imec’s Silicon Photonics Platform (iSiPP), designed to realize optical I/O scaling by leveraging established CMOS manufacturing and advanced 3-D integration methods. We will discuss recently developed silicon photonics chiplet technology, featuring high-speed silicon optical devices, high-speed through-silicon vias (TSVs), and low-loss fiber coupling structures. We will describe existing electro-optical testing solutions and highlight some of the future testing needs.
Silicon Photonics Chiplets for Scaling AI and the Cloud – Technology, Design, and Test

by Joris Van Campenhout – imec (Leuven, Belgium)

Abstract
Artificial intelligence and cloud computing are driving an exponentially growing demand for optical interconnect bandwidth. From the datacenter network down to the chip level, silicon photonics is a prime technology to scale optical interconnects to the desired bandwidth density (>1Tbps/mm), power consumption (<1pJ/bit), and cost (<0.1$/bit). In this presentation, we give an overview of imec’s Silicon Photonics Platform (iSiPP), designed to realize optical I/O scaling by leveraging established CMOS manufacturing and advanced 3-D integration methods. We will discuss recently developed silicon photonics chiplet technology, featuring high-speed silicon optical devices, high-speed through-silicon vias (TSVs), and low-loss fiber coupling structures. We will describe existing electro-optical testing solutions and highlight some of the future testing needs.

Biography Speaker
Joris Van Campenhout is Fellow Silicon Photonics and Director of the Optical I/O industry-affiliation R&D program at imec, which covers the development of an industrially scalable short-reach optical interconnect technology based on silicon photonics. Prior to joining imec in 2010, he was a post-doctoral researcher at IBM’s Tj Watson Research Center (USA), where he developed silicon electro-optic switches for chip-level reconfigurable optical networks. He obtained a PhD degree in Electrical Engineering from Ghent University (Belgium) in 2007. Joris was granted nine patents and has (co-)authored over 100 papers in the field of silicon integrated photonics, which have received 9000+ citations.

See also:
- https://scholar.google.be/citations?user=h5GdrsYAAAAJ
Silicon Photonics Chiplets for Scaling AI and the Cloud – Technology, Design, and Test

Seventh IEEE International Workshop on Testing Three-Dimensional, Chiplet-Based, and Stacked ICs

November 5, 2020

Joris Van Campenhout – on behalf of imec’s Optical I/O and Silicon Photonics teams
imec, Kapeldreef 75, 3001 Leuven, Belgium
is a World-Leading Independent R&D and Innovation Hub for Nano- and Digital technologies, with a Global Presence

**Trusted partner** for companies, start-ups, and universities

- Customers: leading semiconductor (IDM and fabless) and systems companies
- Co-development partnerships with many leading suppliers of equipment, materials, and EDA
IMEC’s State-of-the-Art Fabrication Facilities
@Leuven Headquarters (25km east of Brussels, Belgium)

- Silicon Photonics Prototyping in 200mm FAB
- Advanced Silicon Photonics R&D (mostly) in 300mm FAB
Optical Interconnects: Landscape and Industry Trends
Optical Datacenter Network Connectivity

Hyper-Scale Cloud Datacenters

Massive # of Servers, Switches and Storage units (100,000+)

Massive Fiber Optic Interconnection Network

Large Installations (several 10,000m$^2$) requiring Long Interconnect Reach (500m+)
Datacenter Operators increasingly adopting 100G+ Single-Mode Optics (1310nm/1550nm)
Optical Datacenter Network Connectivity

Datacenter Switch Evolution

Sources:
- https://www.eetimes.com/broadcom-ships-25-6tbps-switch-on-single-7nm-chip/
- Mark Nowell, Cisco, “Pluggable Optics: pro’s and con’s”, OIF Webinar Oct 14, 2020

State-of-the-Art Datacenter Switch
- 25.6Tbps Switch Bandwidth
- 64x 400G ports (Pluggable Optics)
- >1kW chassis power
- 2x switch performance scaling every 2 years

Key Enablers
- Efficient, Scalable MMU & Pipe Architecture
- Leading-Edge Process Technology and IP
- Physical Design Expertise

Disruptive Improvements in Cost, Power, and Complexity
Six-Chip 25.6T System vs. a Single Die

Example System: 64 x QSFP-DD / OSFP 25.6Tbps in 2RU
High-Performance Computing & Artificial Intelligence

GPU/TPU/FPGAs in Advanced CMOS are moving to 20Tbps+ I/O bandwidths

**NVIDIA DGX-A100**

<table>
<thead>
<tr>
<th>Bandwidth Specifications NVIDIA A100</th>
</tr>
</thead>
<tbody>
<tr>
<td>On-package HBM</td>
</tr>
<tr>
<td>Off-package</td>
</tr>
<tr>
<td><strong>Total I/O</strong></td>
</tr>
<tr>
<td>Power</td>
</tr>
<tr>
<td>Network interface</td>
</tr>
</tbody>
</table>

Increasing compute demands in AI and high-performance computing are driving a need for faster and more scalable interconnects at the package, board, and rack/cluster level

Source: nvidia.com

Source: https://cloud.google.com/tpu
Co-Packaged Optics (CPO)
Moving beyond pluggable optics

- **25.6Tbps Switch - Pluggable Optics**
  - OSFP
  - QSFP-DD

- **51.2Tbps Switch - Co-Packaged Optics**

- **Scaling to 51.2Tbps switch capacity and beyond is challenging using pluggable optics**
- **Moving the optics closer to the switch IC can address the power and bandwidth challenges**
- **Co-Packaged Optics development is underway, see e.g. [http://www.copackagedoptics.com/](http://www.copackagedoptics.com/)**

Strawman concept. Source: Cisco & Microsoft/Facebook CPO consortium
Co-Packaged Optics (CPO)

Optical Modules Cross Section

Possible CPO footprints

Optical Module

Figure: Optical Module Cross Section

Cross section

Sources:
IMEC,
http://www.copackagedoptics.com/,
Rob Stone, Facebook, “Co-packaged Optics in the Datacenter”, OIF Webinar Oct 14, 2020,
Jef Hutchins, Ranovus, “Standardization of Co-Packaged Optics”, OIF Webinar Oct 14, 2020

CPO requires Optical Modules and Silicon Photonics Chiplets with bandwidth beyond 3.2Tbps
Optical Interconnect Landscape and Roadmap

Multi-Terabit/s Optical Interconnectivity needed by mid 2020’s, driven by Cloud and AI/HPC

Optical Interconnects will move into the rack (>1Tbps, >1m) and board (>1Tbps/mm, >10cm)

Total Optical Transceiver Volume expected to increase >10x from ~10M to 100M+ by 2025

Silicon Photonics is a key enabling technology for Scaling Optical Interconnects targeting link distances from 1cm to 100km+
Silicon Photonics (Chiplet) Technology
Silicon Photonic Integration

Silicon Photonic Integrated Circuits

- Si/SiN patterning with nanometer-scale accuracy (193i) [Low-loss, high-precision passive devices]
- Large Si/SiO₂ refractive index contrast of ~2 [scalable PIC density]
- Ge(Si) selective-area epitaxy [photodetectors/modulators]
- Low resistance contacts [high-speed optical devices]
- High-yield fabrication in existing CMOS fabs [200mm/300mm]
- Volume scalability [>1M units/year] & Efficiencies of scale [cost]
- Wafer-scale testing and 3-D packaging and assembly [TSVs, micro-bumps, ...]
- No monolithic integrated optical gain/lasing [need for hybrid solution]

Silicon Photonics = Leverage existing CMOS infrastructure for Photonic Integration
Silicon Photonic Integration
Integration of Passive and Active Optical Functions in Si and Ge on a Single Wafer

Fully Integrated Silicon Photonics Platform on SOI Wafers
- Passive devices patterned in SOI Substrates: BOX thickness = 1-2μm, Top Si thickness = 200-400nm
- Implemented on a standard CMOS toolset: 90/130nm (200mm) or <28nm (300mm)
- Selective-area epitaxial growth for realizing Ge(Si) active devices
Silicon Waveguides
Sub-Micron Si-on-Insulator Waveguides

- High index-contrast Si waveguide technology enables compact photonic circuits (~μm bends)
- But, more sensitive to sidewall roughness (propagation loss ~ 0.1-1dB/cm)
Fiber Coupling Interfaces
Surface-Normal Grating Couplers

Fiber Grating Couplers: <2-3dB insertion loss over ~25nm optical bandwidth to Standard Single Mode Fiber
Fiber Coupling Interfaces

Edge Couplers

**Top view schematic**

- **SiN waveguide**
- **Si waveguide**
- **Oxide cladding**
- **130 nm wide tip**
- **500 µm**

**Single-Mode Fiber**

- **Etched substrate region**
- **SiN waveguide**
- **SiN tapers**
- **SMF-28 fiber**
- **Etched facet**
- **Resealed openings**
  - for substrate removal etch

**Loss Characteristics**

- **<2.5dB loss in O-band**
- **<2.1dB loss in C/L-band**

Graphs showing coupling efficiency vs. wavelength for different etch and taper lengths.

Bosch etched substrate
Silicon Wavelength Multiplexing Devices (WDM)

Ring-based WDM Filters

Filter resonance condition

\[ m \cdot \lambda_{res} = L_{RT} \cdot n_{eff} \]
Silicon Ring Modulator
Design & Static Performance

- Highly doped p-n depletion diode phase shifter in a critically coupled ring resonator
- Extinction ratio of ER~3dB and insertion loss of IL~4.5dB from low drive swing (1Vpp)

### Parameter

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ring Radius</td>
<td>7.5</td>
<td>um</td>
</tr>
<tr>
<td>Quality factor</td>
<td>5000</td>
<td></td>
</tr>
<tr>
<td>Modulation efficiency</td>
<td>30</td>
<td>pm/V</td>
</tr>
<tr>
<td>Transmitter Penalty (1Vpp)</td>
<td>11.5</td>
<td>dB</td>
</tr>
<tr>
<td>Extinction Ratio (1Vpp)</td>
<td>3</td>
<td>dB</td>
</tr>
<tr>
<td>Insertion Loss (1Vpp)</td>
<td>4.5</td>
<td>dB</td>
</tr>
<tr>
<td>Heater tuning (5um RM)</td>
<td>230</td>
<td>pm/mW</td>
</tr>
</tbody>
</table>

- **Phase shifter cross section**
- **Electro-optic Response**
- **Thermo-optic Response**
- Tungsten Heater
- Light in
- Light out
- RF Signal
- Ground

- **Heater Control**
- **Ring resonator**
- Radius ~5-10µm

- **Rib Waveguide**
- **P+, P/N, N+**
- **P++**
Silicon Ring Modulator

High-Speed Performance

- Modulation bandwidth > 35GHz, low capacitance ~30fF
- Enables 60Gbps NRZ and 100Gbps PAM-4 Operation

**Small Signal E/O RF Response**

**Eye Diagrams**

**60Gbps NRZ**

**100Gbps PAM-4**

**Input Reflection Coefficient**

- S11-measurement
- S11-fitting

**Bias C_j (fF) R_S (Ohm)**

- 0V: 30.2, 39.3
O-band VPIN Ge Photodetector

Static Performance

“VPIN” GePD

Longitudinal TEM section

56Gbps NRZ Eye Diagrams

-1V bias

-2V bias

- High responsivity: $R = 0.89 \text{ A/W}$
- Low dark current: $I_{dc} = 15 \text{nA} (-1\text{V})$
- High Bandwidth: $f_{3\text{dB}} > 50\text{GHz} (-1\text{V})$
- Low capacitive load: <30fF
IMEC’s Mature Silicon Photonics Platform
Versatile 56Gb/s+ Silicon Photonics Technology

56-160Gb/s Silicon Ring Modulator
56-128Gb/s GeSi Electro-Absorption Modulator
56-106Gb/s Silicon Mach-Zehnder Modulator

Silicon WDM filters

High-density SOI waveguides (0.5-2dB/cm)
Integrated SiN waveguides (0.2-2.5dB/cm)
High-NA & SMF Edge Couplers (<2/3dB)
SMF Grating Coupler (2dB/5dB)

Fully Integrated Silicon Photonics Platform for 1310nm/1550nm Wavelengths
Silicon Photonics Chiplets for Co-Packaged Optics
Optical Module: Building Block Prototype
Silicon Photonics Test Chip

- 300mm Photonics SOI wafers
  - 220nm top Si
  - 2μm buried oxide layer
- FEOL:
  - 193i Si Waveguide Patterning
  - Selective Ge Epitaxial growth
- BEOL:
  - NiPtSi/W contacts
  - W heaters
  - 2-level Cu Interconnects
  - Cu microbumps, 50μm pitch
- Die size = 4mm x 5mm

- 2018 Symposium on VLSI Technology and Circuits: “Hybrid 14 nm FinFET - Silicon Photonics Technology for Low-Power Tb/s/mm² Optical I/O,” M. Rakowski et al., T20-5
Optical Module: Building Block Prototype
Co-designed FinFET CMOS chip

Transmitter Unit Cell (TX)

Receiver Unit Cell (RX)

Silicon Photonics Chip

FinFET CMOS Chip (mirror image)

Modulator Driver (270μm²)

TIA (<50μm²)  Output buffer

• 2018 Symposia on VLSI Technology and Circuits: “Hybrid 14 nm FinFET - Silicon Photonics Technology for Low-Power Tb/s/mm² Optical I/O,” M. Rakowski et al., T20-5
Optical Module: Building Block Prototype

TSV- and Microbump-assisted Hybrid Assembly

Hybrid FinFET CMOS – Silicon Photonics Transceiver

- powered and controlled using Through-Silicon Via’s (TSV) implemented in the Si Photonics Interposer
- bandwidth density beyond 1Tbps/mm²

• 2018 Symposium on VLSI Technology and Circuits: “Hybrid 14 nm FinFET - Silicon Photonics Technology for Low-Power Tb/s/mm² Optical I/O,” M. Rakowski et al., T20-5

Hybrid FinFET CMOS – Silicon Photonics Transceiver

- powered and controlled using Through-Silicon Via’s (TSV) implemented in the Si Photonics Interposer
- bandwidth density beyond 1 Tbps/mm²

Optical Module: Building Block Prototype
TSV- and Microbump-assisted Hybrid Assembly

SiPho Interposer Cross Section SEM

- 2018 Symposium on VLSI Technology and Circuits: “Hybrid 14 nm FinFET - Silicon Photonics Technology for Low-Power Tb/s/mm² Optical I/O,” M. Rakowski et al., T20-5
**Optical Module: Building Block Prototype**

**Test Results**

**Transmitter Static Transmission**

**Transmitter Eye diagrams (40Gb/s)**

**Receiver Eye diagram (40Gb/s)**

- **Basic functionality successfully demonstrated**
  - Transmitter with 4x WDM channels, each running at 40Gbps
  - Receiver running at 40Gbps

- **2018 Symposia on VLSI Technology and Circuits:**
  “Hybrid 14 nm FinFET - Silicon Photonics Technology for Low-Power Tb/s/mm² Optical I/O,” M. Rakowski et al., T20-5

- **2019 ECOC:**
  “TSV-enabled Hybrid FinFET CMOS – Silicon Photonics Technology for High Density Optical I/O,” D. Guermandi et al.
Optical Module: Scaling Up to 3.2Tbps

New Test Chip

Key Design parameters

- 50Gb/s NRZ modulation rate
- 8 parallel TX and RX lanes
- 8 WDM channels per lane
- 4 lanes for remote laser input
- 960 microbumps, 50µm pitch
- Additional bumps and/or TSVs would be needed for electrical wide I/O, power, ground and control lines
Silicon Photonics Wafer-Scale Testing
Wafer-Scale Electro-Optic Test
THOR test system in imec’s 200mm FAB

FormFactor CM300xi-SiPh probe station with Optical Instrumentation
Wafer-Scale Electro-Optic Test

Electro-Optic Process Control Monitor

Grating coupler-based testing

Automated fiber-to-grating alignment

Electro-Optic Process Control Monitor (gds layout)

Set of spiral waveguides. Horizontal distance input/output = 2750µm.

Set of gratings. Horizontal distance input/output = 500µm.

<table>
<thead>
<tr>
<th>Component</th>
<th>Test sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passive</td>
<td>Grating couplers</td>
</tr>
<tr>
<td>O-band 1310nm</td>
<td>Insertion loss, bandwidth, peak wavelength</td>
</tr>
<tr>
<td>C-band 1550nm</td>
<td>Waveguide spirals</td>
</tr>
<tr>
<td></td>
<td>Propagation loss, bend loss</td>
</tr>
<tr>
<td>Waveguide Crossings</td>
<td>Insertion loss, cross-talk</td>
</tr>
<tr>
<td>Waveguide Transitions</td>
<td>Insertion loss</td>
</tr>
<tr>
<td>Directional coupling</td>
<td>Power coupling, excess loss</td>
</tr>
<tr>
<td>Spltters</td>
<td>Insertion loss, excess loss</td>
</tr>
<tr>
<td>Active</td>
<td>Germanium photodetector</td>
</tr>
<tr>
<td>O-band 1310nm</td>
<td>Dark current, responsivity, resistance</td>
</tr>
<tr>
<td>C-band 1550nm</td>
<td>p-n diode phase shifter efficiency</td>
</tr>
<tr>
<td></td>
<td>VpiLpi</td>
</tr>
<tr>
<td></td>
<td>p-n diode phase shifter loss</td>
</tr>
<tr>
<td></td>
<td>Propagation loss</td>
</tr>
</tbody>
</table>
Process control monitor structures

Example #1: Fiber grating coupler (FGC) performance

- A straight waveguide with a grating coupler on both ends
- Measured quantity: wavelength dependent insertion loss, fiber to fiber
- Extracted device parameters
  - Fiber-to-waveguide insertion loss \(FtW\_IL\_AWL\) [dB]
  - Peak wavelength \(PWL\) [nm]
  - Peak wavelength IL \(FtW\_IL\) [dB]
  - 1dB bandwidth \(BW1\) [nm]
Process control monitor structures

Example #2: Waveguide propagation loss

- A set of (spiral) waveguides with increasing lengths
- Measured quantity: wavelength dependent loss vs. length
- Linear regression of IL vs L, #bends to obtain propagation and bend loss
Process control monitor structures

Gauging Measurement Reproducibility

- Repeated measurement of same wafer/dies/structures
  - bi-weekly frequency

- Evolution of 8 device parameters tracked
  - FtW IL, TE/TM, C/O band
  - Photodetector responsivity
  - Photodetector dark current

- Western Electric rules to detect measurement instability
In-Line Process Control and End-of-Line Electrical and Optical Test
Leverage CMOS metrology, and end-of-line wafer-scale E/O test
Towards Known-Good-Die Testing for SiPho Chiplets
Functional Devices
Example: Si Ring Modulator

Transmission

\[ Q = \frac{\lambda_{\text{resonance}}}{\text{FWHM}} \]

Ring Resonance Wavelength (nm)

Ring Quality Factor

Ring Modulation Efficiency (pm/V)

Wafer-Scale test of DC electro-optical device parameters
Towards KGD Testing of SiPho Chiplets

Basic requirements for a Multi-Tbps Transceiver

Need to enable simultaneous widely-parallel micro-bump Electrical and Optical Fiber Array Probing
Towards KGD Testing of SiPho Chiplets
High-throughput, parallel testing solutions

Solutions exist for parallel electro-optics testing with modest electrical pin count and pitch
Towards KGD Testing of SiPho Chiplets
Enabling electro-optic probing with 1000+ microbumps

Need to combine 32+ fiber array probe with 1000+ microbump probe card
Will need customized probe card form factor mechanically compatible with fiber array
Conclusion
Summary and Take-Away Messages

**Optical Interconnects** are an essential technology for scaling the Cloud and AI

**Silicon Photonics** is ideally positioned to develop dense, Terabit-scale optical interconnects
- Very high photonic integration density: 100-1000+ components per chip
- Multi-Tb/s optical I/O modules through spatial and wavelength-Division Multiplexing
- Low die cost, volume scalability, wafer-scale test
- Direct drive and tight integration with CMOS ICs using advanced 3-D assembly

An important remaining test challenge for **Silicon Photonics**
- Known-Good-Die testing of multi-Tbps SiPho chiplets
- Widely parallel multi-lane electro-optical testing
- Leverage testing solutions for HBM and wide I/O interfaces
Thank You!
embracing a better life
The SEMI Heterogeneous Integrated Roadmap Test Working Group Update

Zoë Conroy – Cisco Systems (USA)

The SEMI HIR (Heterogeneous Integrated Roadmap) TWG (Test Working Group) 2.5D and 3D Testing white paper addressed six key test challenges, based on the evolution of 2.5D/3D from complex die stacks through SiP. These test challenges included test flows, cost and resources, test access, testing heterogeneous die individually and in a single stack/package, debug and diagnosis of failing stacks/die and DfX (Design for Test, Yield, Cost and power). The paper was last updated in 2017. During 2020, the HIR TWGs have been revising and updating all the TWG white papers. This will continue into 2021.

The 2.5D and 3D testing white paper updates will include discussion on the test steps and tests that are needed to confirm that this 2.5D or 3D integration and assembly of sub-components is meeting its expectations. In particular, there will be focus on test methods and limitations for multiple device integrations and the interconnect between the various devices. Trade-offs in defining and implementing the test sequence for a multi-device integration, and the value and limitations of traditional test techniques such as concurrent and adaptive testing will also be addressed. Power and thermal management and cooling capabilities and limitations are also an important aspect on the integration and will be included. Finally, a new area to be added is the test methods and roadmap for devices to test, track, and adapt to the performance of their neighboring components.
Heterogeneous Integration Roadmap (HIR)

• Five organizations* CPMT, SEMI, EDS, EPS & ASME EPPD work in collaboration to deliver the Heterogeneous Integration Roadmap.

• It serves our profession, industry, academia, and research institutes, to meet the challenges of this new world of rapid market disruption and bold technology innovation.

• HIR has 18 Technical Working Groups (TWG) – Test is one.


*CMPT - Components Packaging Manufacturing Technology
SEMI - Global Industry Association representing electronic mfg and design supply chain.
EDS - Electronic Devices Society
EPS – Electronic Packaging Society
ASME EPPD - American Society of Mechanical Engineers Electronic and Photonic Packaging Division
HIR TECHNICAL WORKING GROUPS (TWGs)

Heterogeneous Integration Components
1. Single Chip and Multi Chip Packaging (including Substrates
2. Integrated Photonics
3. Integrated Power Devices
4. MEMS
5. RF and Analog Mixed Signal

Cross Cutting topics
7. Emerging Research Devices
8. Interconnect
9. Test

Integration Processes
10. SiP
11. 3D +2.5D
12. WLP (fan in and fan out)

Packaging for Specialized Applications
13. Mobile
14. IoT and Wearable
15. Medical and Health
16. Automotive
17. High Performance Computing

Design
18. Co-Design & Simulation – Tools & Practice
   • Device, package, subsystem & system levels
Test Technology Working Group Areas

Covering the breadth of topics involved in test requires a unique structure of white papers and tables.

Test Roadmap 2019

- Analog Mixed Signal Testing
- Logic Device Testing
- Memory Device Testing
- Photonic Device Testing
- RF Device Testing
- Specialty Device Testing
- Executive Summary
- List of Contributors

Device Handling & Contacting Roadmaps

- Cost of Test
- Test Methods 7
- 2.5D/3D Testing
- Adaptive Testing
- Burn-In & Reliability Testing
- Concurrent Testing
- DFT & SOC Test
- System Level Testing
- Test & Yield Learning

15 Whitepapers
This chapter will address six key test challenges, based on the evolution of 2.5D/3D from complex die stacks through SiP. These test challenges include:

1. test flows,
2. cost and resources;
3. test access;
4. testing heterogeneous die individually and in a single stack/package;
5. debug and diagnosis of failing stacks/die;
6. DfX (Design for Test, Yield, Cost and power.)
2019 2.5D & 3D Test Update

- 2.5D/3D is still, not yet, a mature and mainstream technology
  ➔ Test methods are still developing.

- So far, test of these devices has relied heavily on DFT and KGD to achieve a reasonable yield.

- Short-term challenges looking forward include:
  - Known-Good-Die:
    - How to achieve this and how to work around the lack of this (test/diagnosis/repair)
  - Interposer Testing:
    - The move to active circuitry in interposers will complicate test.
  - High-speed IO:
    - Contacting and testing those interfaces we can touch and testing those interfaces we can’t touch is becoming quite difficult.
2021 2.5D & 3D Test Update In Progress

• Additional areas of consideration with particular attention to **limitations and constraints** in the future:
  • Test methods / limitations for multiple device integration and interconnects between devices.
  • Value / limitations of traditional test techniques, standards (incl. concurrent and adaptive test).
  • Power and thermal management, cooling capabilities
  • Constraints / roadmap for multi-device probing /contacting.
  • Test methods / roadmap for devices to test, track, and adapt to the performance of their neighboring parts.
  • Effective ways to temporarily align fibers to silicon during test.
Test methods and limitations for a multiple device integration and the interconnect between various devices.

• Incoming die and die assemblies/stacks in the integration
  • Testing done on individual dies (KGD)
  • Managing multiple suppliers
  • DFT features that can be used at next integration level
  • Access to use features, run debug, fault isolate

• Interposer, TSV testing
  • Where is the industry at?
  • What can/needs be tested / what is possible to test /
  • What are limitations to achieve this?

• Repair / redundancy
  • Types of repair and redundancy
  • Future needs, limitations
Value / limitations of traditional test techniques, incl. concurrent and adaptive testing.

- Traditional vs. new test techniques needed
- Limitations/challenges to getting the new test techniques
- Application speed testing (e.g. PAM4, etc) needs and limitations
- Usefulness of IEEE Standards (1149.x, 1687, 1838, 1500)
- Cost vs test time discussion
Power and Thermal Management, Cooling Capabilities

• Thermal control of multiple die with different temp specs.
• Design and stacking to improve thermal management.
• How to confirm the performance over different thermal and power situations in co-packaged devices, including optical devices.
• Ways to confirm wavelength margins with regard to die temperature.
Constraints / roadmap for multi-device probing and contacting.

This area will include:

• Bump size limiting test
• Interposer probing/contacting/testing
• TSV probing
• Test techniques when hybrid bonding techniques are utilized
• Probing around stacks
• Inspections – AOI, Xray
• More discussion on areas to follow.
CALL to ACTION/HOW to COLLABORATE

- White Paper needs to be completed mid 2021.
- We would love to have new team members.
- Team meets weekly, average 2-3 times/month.
- Please reach out to me or Dave Armstrong for more details.
  - Email: zfconroy@cisco.com or
  - Email: dave.armstrong@advantest.com
- Please JOIN US!!
Thank You
Title
Probing Complexities of 3D-Stacked ICs – A Test Engineers’ Perspective

Authors
Ferenc Fodor, Bart De Wachter, Arnita Podpod, Michele Stucchi, Erik Jan Marinissen
imec, Kapeldreef 75, 3001 LEUVEN, Belgium

Abstract
Advancements in processing and scaling of through-silicon vias (TSVs), die- and wafer stacking, and general wafer processing techniques have opened the door towards ICs built by vertically stacking dies. Accompanied by standardized memory interfaces such as JEDEC’s Wide-IO and HBM, which consist of large arrays of fine-pitch micro-bumps, the semiconductor industry has created the building blocks for 3D-SICs. Benefits include heterogeneous integration of multiple dies with a reduced footprint and higher product yield with high-density, high-performance low-power interconnects.

Pre-bond test of these dies, i.e., testing prior to stacking, is vital to achieve an acceptable compound stack yield. To keep the associated costs acceptable, it is best to get pre-bond test access to the dies by probing directly on their large-array, high-density micro-bumps. As research institute with a mission to develop industry-relevant solutions, imec has challenged various suppliers in the wafer probe industry to address the challenges associated to probing on large arrays of fine-pitch micro-bumps. Several suppliers responded and imec has worked with them to co-develop and demonstrate high-accuracy wafer probe stations and advanced fine-pitch probe cards. This joint development work has contributed to the industrial uptake of 3D-SIC products, especially amongst the big hitters of the memory industry. Today, pre-bond die test through micro-bump probing is commonplace.

However, there are more challenges involved in testing complex 3D-SICs. Mid-bond tests of partially assembled stacks offer valuable information about the stacking process, while they also help to ensure a high product quality of the final stack products. Mid-bond tests come with their own set of probe challenges, e.g., in the form of thinned and flexible samples on tape frames.

In this presentation, we will use imec’s in-house wafer manufacturing and stack assembly flow of a seven-die flip-chip fan-out wafer-level package (FC-FOWLP) 3D-SIC with a complex stack architecture as a framework to illustrate typical 3D probing challenges that we encountered and how we were able to address them in collaboration with our suppliers.
PROBING COMPLEXITIES OF 3D-STACKED ICs
- a test engineers’ perspective -

Ferenc Fodor, and Bart De Wachter, Arnita Podpod, Michele Stucchi, Erik Jan Marinissen
Leuven, Belgium
in this presentation...

collaboration

development
the imec-FOWLP

-- problem definition --
INTRODUCING THE FOWLP TEST CHIP

FOWLP: FLIP-CHIP FAN-OUT WAFER-LEVEL PACKAGING

FOWLP is a 3D test vehicle intended to obtain practical leanings regarding:
- stacking and assembly technology development
INTRODUCING THE FOWLP TEST CHIP
FOWLP: FLIP-CHIP FAN-OUT WAFER-LEVEL PACKAGING
### INTRODUCING THE FOWLP TEST CHIP

<table>
<thead>
<tr>
<th>name</th>
<th>pseudo-logic</th>
<th>pseudo-memory</th>
<th>through-package via</th>
<th>Si-bridge</th>
</tr>
</thead>
<tbody>
<tr>
<td>function</td>
<td>mimics a processor, without active components</td>
<td>mimics a memory, without active components</td>
<td>vertical interconnect between two micro-bump arrays</td>
<td>horizontal interconnect between two micro-bump arrays</td>
</tr>
<tr>
<td>pre-bond test access</td>
<td>Cu pillars Wide-I/O2 @ 40µm Wide-I/O2 @ 20µm</td>
<td>Wide-I/O2 @ 40µm</td>
<td>Wide-I/O2 @ 40µm Wide-I/O2 @ 20µm</td>
<td></td>
</tr>
<tr>
<td>probing challenges</td>
<td>micro-bumps \textit{option}: diced samples</td>
<td>micro-bumps \textit{option}: diced samples</td>
<td>micro-bumps tape frames thin, diced samples</td>
<td>two probe cards required</td>
</tr>
</tbody>
</table>
### TEST MOMENTS DURING ASSEMBLY

<table>
<thead>
<tr>
<th>Name</th>
<th>Dies in Stack</th>
<th>Test Access</th>
<th>Probing Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mid-bond pre-mold</strong></td>
<td>logic Si-Bridges</td>
<td>Cu pillars on logic</td>
<td>thin samples pick and place inaccuracies</td>
</tr>
<tr>
<td><strong>Mid-bond post-mold</strong></td>
<td>logic Si-Bridges through-package vias</td>
<td>Cu pillars on logic</td>
<td>thin samples pick and place inaccuracies</td>
</tr>
<tr>
<td><strong>Mid-bond back-side</strong></td>
<td>logic Si-Bridges through-package vias</td>
<td>micro-bumps on TPV</td>
<td>micro-bumps thin samples pick and place inaccuracies</td>
</tr>
<tr>
<td><strong>Post-bond &amp; final</strong></td>
<td>logic Si-Bridges through-package vias</td>
<td>Cu pillars on logic</td>
<td>thin samples pick and place inaccuracies dicing tape on frame</td>
</tr>
</tbody>
</table>
PROBING CHALLENGES OF THE FC-FOWLP

1. Probing dense arrays of micro-bumps
2. Probing on ultra-thin wafers on tape
3. Probing singulated dies on tape
4. Probing on large tape frames
5. PTPA accuracy control*
6. Test execution and reporting*

micro-bumps
thinned (wafer)
diced (wafer)
on a frame
PTPA
software

*always present
VORTEX 2

-- our toolbox of solutions --
WHAT IS VORTEX 2?

measurement & analysis software

python, Excel, LabVIEW
VORTEX 2

probe station
**ADAPTED CASCADE CM300**
**PROBE STATION WITH THERMAL CONTROL**

- **parametric** & functional tests
- **pogo pad interface**
- **manual loading** through front-port of SEMI Std G74-0699 **tape frames**
- **misalignments auto-correction**
  - **AlignChip**: per die correction of **x**, **y**, **z**, and **θ**
  - **PreMapWafer**: optimized version of **AlignChip**

**micro-bump probing** of the TPV die thinned, diced, on a frame

**Translation x, y**
- [0, 25 μm]
- (25 μm, 50 μm]
- (50 μm, 75 μm]
- (75 μm, 100 μm]
- (100 μm, ∞)

**Misalignments**

<table>
<thead>
<tr>
<th></th>
<th>Avg.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>θ</strong></td>
<td>-0.497°</td>
<td>-0.637°</td>
</tr>
<tr>
<td><strong>x</strong></td>
<td>116μm</td>
<td>351μm</td>
</tr>
<tr>
<td><strong>y</strong></td>
<td>232μm</td>
<td>576μm</td>
</tr>
</tbody>
</table>

**[Marinissen et al. – ITC’18]**
VORTEX 2.2
advanced probe cards
THE PARADOX OF PROBING

**Objective 1**: good electrical contact

**Objective 2**: minimal impact probe mark
PROBING MICRO-BUMPS TO ENABLE PRE-BOND TEST ACCESS

1. INTRODUCTION

Micro-Bump Probe Targets
- imec’s PoR @40µm pitch
- Today’s advanced industry practice

Wide-I/O Micro-Bump Arrays
- WIO1: 1,200 micro-bumps @50/40µm pitch
- WIO2: 1,752 micro-bumps @40µm pitch
- HBM2: 4,258 micro-bumps @55µm pitch
MEMS probe card
- Probes: Ø25μm flat, 4100μm tall
- Contact with the whole probe surface
- Stable performance on Cu micro-bumps
- **No reflow** of Cu/Ni/Sn micro-bumps
- Operating over-travel: 50-60μm
- Membrane-based, MEMS probe card
- Probes: 6×6µm², 10µm tall
- Contact with the heel of probe
- Stable performance on Cu micro-bumps
- Can require reflow on Cu/Ni/Sn micro-bumps
- Operating over-travel: 110-130µm
VORTEX 2

measurement instruments
VORTEX 2.2’S HYBRID MEASUREMENT SETUP

STS T2 TEST HEAD AND PXI PLATFORM

for the Cu pillars

24 tester channels

1 × NI PXI/PXIe-2531
  - Switch Matrix (SMX)
1 × NI PXIe-4072
  - Digital Multimeter
1 × NI PXI/PXIe-4141
  - 4-channel Source Measurement Unit

upgraded to

1 × NI PXIe-4163
  - 24-channel Source Measurement Unit

for the micro-bumps

1200 tester channels

9 × NI PXIe-2535
  - FET-based SMX
1 × NI PXIe-4072
  - Digital Multimeter

[NI Whitepaper - Fodor et al.]
VORTEX 2
measurement and analysis software
TEST EXECUTIVE
ORIGINALLY IN LABVIEW, TRANSITIONING TO PYTHON

python libraries involved:
nidcpower, nidmm, niswitch, pyvisa, pandas, numpy

by changing the class arguments we can easily
- swap between projects, setups
- deploy the test executive on both probers
SNAPSHOTS
PROBE STATION AS A MICROSCOPE

wafer snapshotting is popular amongst our ‘customers’
- allows for a quick glimpse onto the wafer
- can explain measurement anomalies
- built-in function of Velox (probe station software)
- automatically called after measuring a wafer

**TPV during mid-bond back-side test**

**color change**
due to oxidation

**not a probe mark**

**rough surface** finish
from wafer thinning
PROBE MARK ANALYSIS
FOR PTPA ACCURACY ASSESSMENTS

prober PTPA – v – temperature

probe card accuracy

misalignment vector $\vec{m}$ (µm)
relative probe-mark area $a$ (%)

PTPA @T=32°C
PTPA @T=22°C

Fodor et al. – Probing Complexities of 3D-Stacked ICs – PUBLIC
UNDERSTANDING THE FC-FOWLP DATA

A DESIGN-SPECIFIC ANALYSIS TOOL

<p>| | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PM04-SHORTB</td>
<td>0.80562</td>
<td>0.410771</td>
<td>0.341333</td>
<td>0.442005</td>
<td>2.174273</td>
<td>1.618236</td>
<td>1.51276</td>
<td>2.962434</td>
<td>8.541298</td>
<td>9.938584</td>
<td>0.314802</td>
<td>0.284344</td>
<td>0.286562</td>
</tr>
<tr>
<td>PM06-SHORTB</td>
<td>0.384411</td>
<td>0.368521</td>
<td>0.348739</td>
<td>0.345312</td>
<td>0.533328</td>
<td>0.991715</td>
<td>0.87864</td>
<td>1.793683</td>
<td>7.341996</td>
<td>10.10949</td>
<td>0.304099</td>
<td>0.284583</td>
<td>0.286599</td>
</tr>
<tr>
<td>PM17-SHORTM</td>
<td>0.422974</td>
<td>0.451937</td>
<td>0.421245</td>
<td>0.391369</td>
<td>1.644582</td>
<td>2.444389</td>
<td>1.612898</td>
<td>3.279995</td>
<td>9.355931</td>
<td>11.15418</td>
<td>0.354817</td>
<td>0.329317</td>
<td>0.312614</td>
</tr>
<tr>
<td>PM19-SHORTM</td>
<td>0.37725</td>
<td>0.405135</td>
<td>0.431198</td>
<td>0.403057</td>
<td>1.954806</td>
<td>1.346746</td>
<td>1.305247</td>
<td>5.144133</td>
<td>9.355931</td>
<td>11.15418</td>
<td>0.354817</td>
<td>0.329317</td>
<td>0.312614</td>
</tr>
<tr>
<td>PM08-OPEN</td>
<td>8.44E-09</td>
<td>8.51E-09</td>
<td>8.45E-09</td>
<td>8.45E-09</td>
<td>8.44E-09</td>
<td>8.44E-09</td>
<td>8.46E-09</td>
<td>8.46E-09</td>
<td>8.42E-09</td>
<td>8.39E-09</td>
<td>8.29E-09</td>
<td>8.36E-09</td>
<td>8.35E-09</td>
</tr>
<tr>
<td>KPV1_S1-BCU_PM02</td>
<td>0.489512</td>
<td>0.490631</td>
<td>0.501166</td>
<td>0.507168</td>
<td>0.519753</td>
<td>0.498964</td>
<td>0.535672</td>
<td>0.630519</td>
<td>0.450585</td>
<td>0.447221</td>
<td>0.451709</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KPV1_S2-BCU_PM02</td>
<td>0.452701</td>
<td>0.455831</td>
<td>0.455083</td>
<td>0.462706</td>
<td>0.483823</td>
<td>0.507168</td>
<td>0.519753</td>
<td>0.498964</td>
<td>0.535672</td>
<td>0.630519</td>
<td>0.450585</td>
<td>0.447221</td>
<td></td>
</tr>
<tr>
<td>KPV1_S3-BCU_PM02</td>
<td>0.426228</td>
<td>0.429846</td>
<td>0.431717</td>
<td>0.435596</td>
<td>0.450473</td>
<td>0.452467</td>
<td>0.50524</td>
<td>0.536998</td>
<td>1.087874</td>
<td>17.0291</td>
<td>0.425467</td>
<td>0.426088</td>
<td></td>
</tr>
<tr>
<td>KPV1_S4-BCU_PM02</td>
<td>0.42574</td>
<td>0.428115</td>
<td>0.428125</td>
<td>0.432619</td>
<td>0.453875</td>
<td>0.456885</td>
<td>0.473023</td>
<td>0.604092</td>
<td>0.535672</td>
<td>0.922268</td>
<td>0.682253</td>
<td>0.421148</td>
<td>0.422354</td>
</tr>
<tr>
<td>KPV1_S5-BCU_PM02</td>
<td>0.52981</td>
<td>0.551481</td>
<td>0.550856</td>
<td>0.553989</td>
<td>0.573997</td>
<td>0.587114</td>
<td>0.583366</td>
<td>0.616886</td>
<td>0.922268</td>
<td>0.95904</td>
<td>0.545602</td>
<td>0.542725</td>
<td>0.543736</td>
</tr>
<tr>
<td>KPV1_S6-BCU_PM02</td>
<td>0.499316</td>
<td>0.497689</td>
<td>0.499065</td>
<td>0.511319</td>
<td>0.548969</td>
<td>0.499065</td>
<td>0.511319</td>
<td>0.548969</td>
<td>0.55521</td>
<td>0.558505</td>
<td>0.55323</td>
<td>0.625896</td>
<td>0.625896</td>
</tr>
<tr>
<td>KPV1_S1-MET1_PM02</td>
<td>1.046062</td>
<td>1.005075</td>
<td>1.061585</td>
<td>1.076421</td>
<td>1.074291</td>
<td>1.113458</td>
<td>1.086558</td>
<td>1.047076</td>
<td>18.50285</td>
<td>29.36514</td>
<td>1.053338</td>
<td>1.039662</td>
<td>1.045815</td>
</tr>
<tr>
<td>KPV1_S2-MET1_PM02</td>
<td>6.186409</td>
<td>1.135069</td>
<td>1.134185</td>
<td>1.22037</td>
<td>30.71179</td>
<td>21.51035</td>
<td>32.83534</td>
<td>29.01536</td>
<td>22.02703</td>
<td>11.9001</td>
<td>10.0928</td>
<td>1.125014</td>
<td>1.098749</td>
</tr>
<tr>
<td>KPV1_S3-MET1_PM02</td>
<td>1.167781</td>
<td>1.140137</td>
<td>1.138338</td>
<td>1.131716</td>
<td>1.110752</td>
<td>1.128977</td>
<td>1.109298</td>
<td>1.163591</td>
<td>8.85269</td>
<td>33.45475</td>
<td>1.125783</td>
<td>1.101728</td>
<td>1.094913</td>
</tr>
<tr>
<td>KPV1_S4-MET1_PM02</td>
<td>1.102741</td>
<td>1.143258</td>
<td>1.109435</td>
<td>1.085862</td>
<td>1.136531</td>
<td>2.732797</td>
<td>1.153217</td>
<td>1.190001</td>
<td>7.882666</td>
<td>12.96647</td>
<td>1.104228</td>
<td>1.085911</td>
<td>1.078523</td>
</tr>
<tr>
<td>KPV1_S5-MET1_PM02</td>
<td>1.102741</td>
<td>1.111393</td>
<td>1.107538</td>
<td>1.100409</td>
<td>1.095681</td>
<td>1.08287</td>
<td>1.078175</td>
<td>1.105993</td>
<td>25.06684</td>
<td>23.19842</td>
<td>1.093413</td>
<td>1.069482</td>
<td>1.068023</td>
</tr>
<tr>
<td>KPV1_S6-MET1_PM02</td>
<td>3.001099</td>
<td>2.886673</td>
<td>2.954925</td>
<td>2.811402</td>
<td>5.771777</td>
<td>5.897885</td>
<td>10.47391</td>
<td>8.44736</td>
<td>11.75188</td>
<td>800.756</td>
<td>2.893302</td>
<td>2.804444</td>
<td>2.723992</td>
</tr>
<tr>
<td>KPV1_S1_PM02</td>
<td>0.375053</td>
<td>0.337541</td>
<td>0.317541</td>
<td>0.318792</td>
<td>0.307548</td>
<td>0.350047</td>
<td>0.296298</td>
<td>0.331295</td>
<td>0.321298</td>
<td>0.318794</td>
<td>0.311294</td>
<td>0.312288</td>
<td>0.317544</td>
</tr>
<tr>
<td>KPV1_S2_PM02</td>
<td>-0.0762</td>
<td>-0.11993</td>
<td>-0.10744</td>
<td>0.031232</td>
<td>-0.08745</td>
<td>-0.00999</td>
<td>0.024985</td>
<td>-0.01249</td>
<td>-0.02623</td>
<td>0.064962</td>
<td>-0.12368</td>
<td>-0.03123</td>
<td>0.011243</td>
</tr>
</tbody>
</table>
# Understanding the FC-FOWLP Data

A Design-Specific Analysis Tool

![Diagram of a carrier wafer with interconnects and labels for Memory, Logic, Bridge, TPV, and Kelvin checks.]

<table>
<thead>
<tr>
<th>Memory</th>
<th>b5</th>
<th>TPV</th>
<th>b3</th>
<th>Bridge40</th>
<th>b1</th>
<th>Logic</th>
<th>Kelvin</th>
</tr>
</thead>
<tbody>
<tr>
<td>NW</td>
<td>0.0%</td>
<td>24.1%</td>
<td>8.18</td>
<td>72.5%</td>
<td>0.0%</td>
<td>Logic Dis</td>
<td>100%</td>
</tr>
<tr>
<td>SiBridge</td>
<td>72.5%</td>
<td>0.0180</td>
<td>0.0152</td>
<td>0.0152</td>
<td>0.0180</td>
<td>TPV</td>
<td>100%</td>
</tr>
<tr>
<td>TSV</td>
<td>34.6%</td>
<td>0.0466</td>
<td>17.7</td>
<td>128x 18.8</td>
<td>128x 18.8</td>
<td>Bridge20</td>
<td>100%</td>
</tr>
<tr>
<td>SiBridge</td>
<td>72.5%</td>
<td>57.0%</td>
<td>8.96</td>
<td>0.0148</td>
<td>0.0148</td>
<td>TPV</td>
<td>100%</td>
</tr>
<tr>
<td>SW</td>
<td>0.0%</td>
<td>0.0471</td>
<td>8.58</td>
<td>0.0189</td>
<td>0.0189</td>
<td>Bridge20</td>
<td>100%</td>
</tr>
<tr>
<td>Kelvin</td>
<td>72.5%</td>
<td>0.0232</td>
<td>18.2</td>
<td>0.0515</td>
<td>0.0515</td>
<td>TPV</td>
<td>100%</td>
</tr>
<tr>
<td>DaisyChain</td>
<td>0.0%</td>
<td>0.0180</td>
<td>0.0152</td>
<td>0.0152</td>
<td>0.0180</td>
<td>Bridge20</td>
<td>100%</td>
</tr>
</tbody>
</table>

**Probe Checks**

- 100% for Logic and TPV
- 57.0% for Bridge20
- 72.5% for Kelvin

---

Fodor et al. – Probing Complexities of 3D-Stacked ICs – PUBLIC
in conclusion
CONCLUSION

- In this presentation we have demonstrated a set of probing challenges that come with our seven-die FOWLP 3D architecture, namely:

  **probing challenges** during assembly of FOWLP
  1. Probing dense arrays of micro-bumps
  2. Probing on ultra-thin wafers on tape
  3. Probing singulated dies on tape
  4. Probing on large tape frames
  5. PTPA accuracy control
  6. Test execution and reporting

- Through active collaboration and development we have built a system that can handle the test challenges of emerging 3D architectures

  **addressed by Vortex 2.2 using**
  - advanced, low-force probe cards
  - probe station software
  - adapted probe station
  - various software solutions
Part of the work of Erik Jan Marinissen and Ferenc Fodor was performed in the project **SEA4KET**, Semiconductor Equipment Assessment for Key-Enabling Technologies (http://www.sea4ket.eu), sub-project 3DIMS, 3D Integrated Measurement System; this project received funding from the European Union’s Seventh Programme for research, technological development, and demonstration under grant agreement No. IST-611332

We also acknowledge the **EMPIR 14IND07-3D Stack project**
Title:
HBM2 probing challenges and probe card architecture

Authors:
Raffaele Vallauri, EVP Technology Operations - Technoprobe
Alessandro Antonioli, Business Development and Marketing – Technoprobe
Flavio Maggioni, VP – Product & Design Development - Technoprobe

Abstract:
The memory and data storage sectors have been subject to significant change recently, with the industry having changed more in the past 10 years than it did in the preceding 25. Even more drastic change is expected too as the memory sector continues to innovate in order to keep up with the introduction of game-changing innovations like AI.

HBM stands for ‘high-bandwidth memory’, a premium performance interface for 3D-stacked SDRAM (synchronous dynamic random-access memory). It maximizes data transfer rates in a small form factor that uses less power and has a substantially wider bus when compared to other DRAM solutions. For high-performance computing applications, industries planning to leverage AI, graphics card vendors and advanced networking applications, HBM provides data speed increases that are essential to helping drive industries forward.

The inception of HBM memory solutions has been followed by the introduction of HBM2 and HBM2E, which allow for more DRAM die to be utilized per stack, increasing capacities across the board.

Technoprobe developed a specific probing solution for HBM2 products based on TPEG™ MEMS T50 probe technology and on high density MLO solution.

In this paper a description of the device requirements and Technoprobe probing solution will be presented and also characterization data will be provided and discussed in details starting from joint presentation with SEC held in 2017 at SWTW.
HBM2 probing challenges and probe card architecture

Raffaele Vallauri, EVP Technology Operations - Technoprobe
Alessandro Antonioli, Director - Marketing BDM – Technoprobe
Flavio Maggioni, VP – Product & Design Development - Technoprobe
Agenda

• Technoprobe introduction
• 2D and 3D package integration
• Fine pitch probing solutions
• HBM2 at 55um pitch solution
• Characterization data
• Signal Integrity
• Next gen 2D/3D technologies
• Interconnection challenges
• Summary
• Follow-On Work
Technoprobe introduction

- Technoprobe is #2 WW Probe Card manufacturer
- Leader in Vertical Probe Cards addressed with patented TPEG™ MEMS probing technology
TPEG™ MEMS Vertical probe technology

- TPEG™ MEMS is a proprietary and patented process to manufacture probes currently used mainly in probe-cards:

- **TPEG™ MEMS Key Values:**
  - probe size tolerances are very tight
  - Force can be tailored to suit application
  - Current can be controlled even at lower pitches
  - Short probes for high frequency with controlled force
  - Long lifetime
  - probes are one-by-one replaceable and repairable even **onsite**
  - High Volume Manufacturing and easily scalable
2D and 3D package integration

- 2D/3D integration requires known good dies, i.e. devices fully tested at wafer sort, which means at highest performances
- No sockets can support fine pitch of HBM2 microcontacts

Source: Samsung
Fine pitch probing solutions
Technology

• Technoprobe has developed dedicated probing solution for fine pitch allowing full pad or bump contact at very low pitches

<table>
<thead>
<tr>
<th>PARAMETER</th>
<th>TPEG™ MEMS T60</th>
<th>TPEG™ MEMS T50</th>
<th>TPEG™ MEMS T40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Needle diameter</td>
<td>~ 1.3 mils equivalent</td>
<td>~ 1.2 mils equivalent</td>
<td>~ 1.0 mils equivalent</td>
</tr>
<tr>
<td>Tip shape</td>
<td>Pointed or Flat</td>
<td>Pointed or Flat</td>
<td>Pointed or Flat</td>
</tr>
<tr>
<td>Radial alignment accuracy and Z planarity (typical)</td>
<td>&lt;7 µm; Z electrical plan: Δ 20 µm</td>
<td>&lt;7 µm; Z electrical plan: Δ 20 µm</td>
<td>&lt;7 µm; Z electrical plan: Δ 20 µm</td>
</tr>
<tr>
<td>Min pitch and configuration</td>
<td>60 µm full array configuration</td>
<td>50 µm full array configuration</td>
<td>40 µm full array configuration</td>
</tr>
<tr>
<td>Pin Current (CCC ISMI 2009)</td>
<td>350 mA (std alloy) – 500 mA (special alloy)</td>
<td>330 mA (std alloy) – 470 mA (special alloy)</td>
<td>270 mA (std Alloy) - 330mA (HC alloy)</td>
</tr>
<tr>
<td>Force (at 3 mils OT)</td>
<td>2.2 g (±20%)</td>
<td>1.8 g (±20%) pointed</td>
<td>1.2 g (±20%) flat</td>
</tr>
</tbody>
</table>
Fine pitch probing solutions
Technology for HBM2

- TPEG™ T50 HC FLAT 1.2g has been developed to provide mechanical and electrical performances required in probe cards for HBM2
HBM2 at 55um pitch solution
Product example

• First paper work was presented in SWTW 2017 “HBM Micro Pillar Grid Array Probing Challenges” by Technoprobe and Samsung:
  • 3D-stacked DRAM grid array of 4942 micro-bumps at 55um pitch as its signal terminal
HBM2 at 55um pitch solution
Probe card

• MPGA - High density signal required Advanced MLO solution, which can be soldered onto main PCB or Interposer.
Full Characterization of probe card for HBM2

- **TPEG™ MEMS T50 1.2g flat** probe technology has been fully characterized on HBM2 to enable HVM. Main parameters
  - Contact Resistance and Cleaning
  - Probe Mark Area
  - Planarity
  - Alignment Error
  - Height loss
  - Current Carrying Capability (CCC)
Full Characterization of probe card for HBM2 Trial on Customer wafer

- Contact quality is key to deliver production oriented products:
  - Stable and low CRES is key to support HVM
Full Characterization of probe card for HBM2 CRES

• **CRES vs Overdrive**
  - 3 TDs are performed for every different OD on fresh bumps
  - 3 x 24 Cres measurements (pairs) = 72 Cres values are reported for each OD
  - Measuring Result: Spec in for OD 50~100µm
  - Max Cres = 0.3Ω - (Spec = under 2Ω)

• **CRES vs Cleaning**
  - 1000 TDs @ 75µm OD are performed
  - 5 runs of 200 TDs are done on available daisy chain areas:
  - Probe tip cleaning: 3 cleaning cycles TDs every 100 probing TDs
  - 3M pink paper; X-Y movement (30 µm L) – Cleaning OD: 30 µm

---

The value of Cres includes PC path Resistance (1.2Ω)

Without cleaning

< 0.3Ω (C_RES)

1.2Ω (Probe Card R path)
Full Characterization of probe card for HBM2 probe mark area

- Full probe pin populated Probe Head
  - Probe marks area inspection via confocal microscope at different OD = 50, 75, 100µm
  - Confocal microscope is used to obtain a 3D image of bump top surface
  - Measuring Result: Max 20% - Spec in (≤30%)
- Analyzed full wafer map

<table>
<thead>
<tr>
<th>OD</th>
<th>Average PM area</th>
<th>SD</th>
<th>MIN</th>
<th>MAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>50µm</td>
<td>14.4%</td>
<td>1.1%</td>
<td>12.6%</td>
<td>16.7%</td>
</tr>
<tr>
<td>75µm</td>
<td>16.6%</td>
<td>1.0%</td>
<td>14.7%</td>
<td>19.2%</td>
</tr>
<tr>
<td>100µm</td>
<td>17.9%</td>
<td>1.3%</td>
<td>15.8%</td>
<td>20.3%</td>
</tr>
</tbody>
</table>

### Probe Mark Area = (a/A)^2
- a : scrub diameter
- A : pillar diameter

![Confocal Microscope 3D image](image1.png)

![Bump slicing to calculate probe mark area](image2.png)

![Probe Mark Images @ OD 75µm](image3.png)
Full Characterization of probe card for HBM2 probe mark area

• To support HVM a bump damage (probe mark area) needs to be evaluated in different test conditions:
  • Probing overdrives from 50µm and 100µm
  • Testing temperatures of -5 °C, 30° C and 85° C
  • Contact times of 1s, 10s, 1min, 10min, 1h

• Custom prober setup was realized to achieve the level of stability required for this analysis
Full Characterization of probe card for HBM2 probe mark area

- Effect of overdrive, temperature and contact time
  - Low temperature (-5 °C) marks are smaller than RT marks by ~20%
  - Bump mark area increases with regular trends with respect to contact time

Prober-A
-5°C and +30°C

Prober-B
+30°C and +85°C

Bump mark area increases with increasing temperature, approximately by 50% moving from room temp to 85 °C. This is consistent with literature data reporting yield stress decay from 40%\(^1\) to 55%\(^2\) for SnAg solders in the same temperature range.

---

\(^1\) Shi et al., "Reliability assessment of PBGA solder joints using the new creep constitutive relationship and modified energy-based life prediction model", Proc of EPTC (IEEE) 2000

\(^2\) Pang et al., "Bulk Solder and Solder Joint Properties for Lead Free 95.5Sn-3.8Ag-0.7Cu Solder Alloy", Proc of ECTC 2003 (IEEE)
Full Characterization of probe card for HBM2 Alignment analysis

- Alignment Error
  - Measuring X-Y Alignment of needles using Vision method for all pins (Nominal pitch 55µm)
  - Measuring result: Max 7µm ➞ Spec in(< 8µm)

- Alignment Image @ OD 75µm
  - Prober camera is used to inspect the probe marks through the wafer
Full Characterization of probe card for HBM2 Planarity, Height Loss and CCC

- **Unloaded Planarity**
  - Measuring Result: Max $\Delta = 9 \mu m$ - Spec $\Delta 20 \mu m$

- **Height Loss**
  - Measure Bump height before and after Probing
  - Max under 1um $\Rightarrow$ Spec In ($\leq 3\mu m$)

- **Current Carrying Capability (CCC)**
  - Standard Method: ISMI ('09)
  - Sample of 4 needles were measured: CCC(mean) = 360 mA
# Full Characterization of probe card for HBM2

## Summary

<table>
<thead>
<tr>
<th>#</th>
<th>Items</th>
<th>Method</th>
<th>Spec</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alignment Error</td>
<td>Measuring the radial alignment(X-Y) error on PRVX4</td>
<td>&lt; 8μm</td>
<td>&lt; 7μm</td>
</tr>
<tr>
<td>2</td>
<td>Contact Resistance</td>
<td>Force V – Measure I method after remove internal path resistance</td>
<td>&lt; 1Ω</td>
<td>&lt; 0.3Ω</td>
</tr>
<tr>
<td>3</td>
<td>Planarity</td>
<td>Measuring full loading planarity using conductive check plate on PRVX4</td>
<td>Δ20μm</td>
<td>Δ9μm</td>
</tr>
<tr>
<td>4</td>
<td>Probe Mark Area</td>
<td>Measuring the PM area using Confocal Microscope at various OD(50,75,100um)</td>
<td>&lt; 30%</td>
<td>&lt; 20%</td>
</tr>
<tr>
<td>5</td>
<td>Height loss</td>
<td>Measuring the bump height using CAMTEK Eagle-I at various OD(50,75,100um)</td>
<td>&lt; 3μm</td>
<td>&lt; 1μm</td>
</tr>
<tr>
<td>6</td>
<td>Current Carrying Capability (CCC)</td>
<td>Measuring the CCC using ISMI '09</td>
<td>Max. 100mA</td>
<td>Max. 360 mA</td>
</tr>
<tr>
<td>7</td>
<td>Cleaning</td>
<td>Measuring the Cres every 1K TD Compare between No cleaning and cleaning at 75um OD</td>
<td>-</td>
<td>Need to clean each 100 TD</td>
</tr>
</tbody>
</table>
Signal Integrity

• High speed signals need to be tested at wafer sort.
• Full path simulation is key to build full functional probe card and predict behavior
• Full path can be measured by VNA: example below
Signal Integrity
HBM example

• Comparison simulation vs measure
  • Measure and simulation show similar insertion Loss: measure show just a bit higher loss due to 56cm coaxial cable not modelled in simulation
Signal Integrity
HBM example

• Eye Diagram measure
  • Measure helps to setup correctly tester and test program

INPUT

OUTPUT: eye diagram is open even if reduced
Sub-50µm product example
WIDE I/O2 wafer test chips at 40µm pitch (courtesy of IMEC)

- Scaling down pitch to 40µm with TPEG™ T40 / Arianna™ A40 depending on contact metallurgy
  - Φ 25µm Cu Micro-Bumps
    - 200nm SiN
    - Seed: 130nm Ta, 300nm Cu
    - Plating: 5µm Cu
    - Option: electroless 10nm NiB
  - Φ 15µm Cu/Ni/Sn Micro-Bumps
    - 200nm SiN
    - Seed: 130nm Ta, 300nm Cu
    - Plating: 5µm Cu + 1µm Ni + 3.5µm Sn
Sub-50µm product example
WIDE I/O2 wafer test chips at 40µm pitch (curtesy of IMEC)

• Wide I/O2 solutions adoptions in 2D/3D package applications
Interconnection challenges

Fine pitch solutions

• Traditional fine pitch interconnections from probe heads to PCB can be challenged due to the fine pitch and probe density

**Standard MLO**

- MLO Soldered to PCB through BGA
- Pitch Capability down to 80 μm

**Advanced MLO**

- MLO Soldered onto Interposer
- Pitch Capability down to 40/45 μm (depending on device configuration)

**Fine pitch MLC**

- HTCC/LTCC core and MLO build-up
- Pitch Capability down to 60 μm (depending on device configuration)
Interconnection challenges
40µm and smaller pitch interconnection

• When probing at 40 µm pitch and smaller pitches, the limitation becomes more on the space transformer than PH technology:

  • Micro-wired space transformer, which is limited in performance: low current, low frequency.

  • As a development path, Technoprobe is developing micro-interposer that will allow to overcome wired space transformer drawbacks.

  • Depending on layout it is possible to evaluate existing Fine Pitch MLO solution.

  • Si-Interposer is also an option.
Summary

• HBM2 solutions with 55µm pitch requires probe cards enabling full test at wafer sort
• Probe cards for HBM2 at 55µm pitch characterization has enabled mass production with TPEG™ MEMS T50 1.2g flat technology.
• In qualification next generation HBM solution requiring sub 50µm pitches: TPEG™ MEMS T40 has been proved as viable solution
Follow-On Work

• High volume manufacturing of sub 50µm solution down to 40µm pitch full array.
• Develop solution for sub 40µm pitches down to 30µm pitch
• HVM of Micro-Interposer space Transformer for sub 50µm pitches
3DC-TEST
Thank You
With the slowing of Moore’s Law, the industry is adopting a variety of advanced packaging methods to further advance heterogeneous integration of both logics and memory device. The breakthrough of 2.5D and 3D wafer level packaging technologies has opened up many new possibilities and test challenges. The vertically stack ICs and proximity of fanout methods on a small form factor package drive the need for a faster test speed with higher signal performance. A new level of advance probe card technology is required to address the required signal quality and the ability to successfully probe on TSV micro-bump structure in an ultra-low pitch grid array. This paper discusses test solutions for High-Bandwidth-Memory in various configurations and the implications on cost of test.
Known Good Dies (KGD) strategies compatible with Direct Hybrid bonding


Univ. Grenoble Alpes, CEA, LETI, 17 rue des Martyrs, 38000 Grenoble, France

SET Corporation, 131 impasse barteaud, 74490 Saint-Jeoire

Die-To-Wafer (D2W) direct hybrid bonding process, using copper/oxide mix interfaces, is identified by major microelectronic industrials as essential for the success of future logic & memory stacks, thanks to heterogeneous integration with low interconnection pitch and known good die (KGD) selection capabilities [1]. The current interconnection pitch demonstrated in LETI is 10µm [2]. 5µm is aimed by the end of 2021. Testing chips at wafer level before assembling adds significant value to Chip-to-Wafer approach by increasing yield of final product [3]. However, it rises integration challenges to make the test compatible with the direct bonding afterwards. At which level the test can be performed? Which solutions can be proposed to enable direct bonding after testing? This paper presents the complete solution developed at CEA-Leti of D2W direct bonding using KGD selection before stacking.

The first part of this presentation will review the D2W direct hybrid bonding as developed in CEA-Leti [4]. Direct hybrid bonding is based on the direct bonding of Cu/SiO2 surfaces. After an additional annealing, the bonding interface is sealed and Cu interconnections are rebuilt. The D2W hybrid bonding integration flow is depicted in Fig 1. First of all, two damascene bonding levels are created directly after the last interconnection level on both bottom wafer and top dies at wafer scale. The cross section in Fig.2 shows the damascene levels stack after CMP with 10µm bonding pitch. Main challenges for a good bonding quality lie in the realization of post CMP treatments without bonding surface degradation nor particle contamination. Thus, special care was carried out for dicing post treatment, die handling and stacking to limit defect on the bonding surface. Finally, dies were directly stacked with SET NEO HB stacking system [5] on surface prepared bottom wafer. Extremely good bonding quality was observed by scanning acoustic microscopy (SAM) as shown in Fig. 3 and lead to electrical yields higher than 75% (Fig.4). D2W hybrid bonding was successfully demonstrated.

In the second part, latest KGD development applied to D2W direct bonding will be presented. The great advantage of D2W in regards with W2W is the capability to select dies with good electrical performances before stacking (Fig. 5). KGD scheme strongly improves the final yield of the 3D system as demonstrated for conventional 3D stacking [1], [3]. However, KGD rises challenges of compatibility with high surface flatness and roughness requirement of direct bonding [6]. Indeed, electrical testing creates high topographies on testing pad which disable direct bonding. In this paper, several integration strategies are discussed to render KGD compatible with hybrid bonding. After the characterization of topography induced by test probes (Fig.6), post test planarization processes were developed to minimize defect. Finally, CEA-Leti successfully demonstrated a D2W hybrid bonding with KGD strategy, as shown by the SAM of bonded tested dies in Fig.7. These results demonstrate the feasibility of including a KGD strategy into the D2W direct bonding flow, which is a mandatory requirement to make D2W HB transferred into high volume manufacturing. Main conclusion of this presentation is the demonstration to KGD integration flow adapted to D2W direct hybrid bonding.

Acknowledgements: This work was partially funded thanks to the French national program “Programme d’investissements d’Avenir, IRT Nanoelec” ANR-10-AIRT-05”.

References:
Figure 1: Simplified process flow for Die-to-Wafer direct hybrid bonding.

Figure 2: X-section of an electrical test vehicle after damascene & CMP processing with 10µm interconnection pitch.

Figure 3: Bonding interface quality evaluation at wafer-level inspection: 90% of the dies present defect free interfaces.

Figure 4: Cumulative percentage of 1 daisy-chain element resistance (Rlink) for different daisy-chains length.

Figure 5: Integration of KGD module in the D2W hybrid bonding flow.

Figure 6: Optical interferometry of one probe mark on Cu, 3µm high topography measured.

Figure 7: Acoustic scan of D2W bonding with (a) no test and (b) 50µm overdrive test + CMP: no difference is noticed.
Known Good Dies (KGD) strategies compatible with D2W Direct Hybrid bonding


**SET Corporation**: N. Raynaud, P. Metzger

[emilie.bourjot@cea.fr](mailto:emilie.bourjot@cea.fr)
DIRECT HYBRID BONDING IN 3D TECHNOLOGIES

**µbumps**
- Pitch 20µm

**direct hybrid bonding**
- D2W
  - Pitch 10µm
- W2W
  - Pitch 1µm

**CoolCube™**
- Pitch 0.1µm

**Key enabler for 3D technologies**
KNOWN GOOD DIES FOR 3D

KGD to increase final product yield and decrease production costs

W2W N-layer stack

Wafer 1
Wafer 2
Wafer 3

w/o pre bonding test
with pre bonding test

Scrap
OK

Final product yield = f(wafer yield, bonding quality,...)

How does pre bond test impact direct hybrid bonding process?
SURFACE REQUIREMENTS FOR DIRECT BONDING

Direct bonding: adhesion of 2 surfaces by molecular bonds without adhesive

**Morphological criteria**
- Bow/wrap
- Topography
- Roughness

**Cleanliness criteria**
- Organic & particle contamination
- 1 µm particle ➔ 5 mm bonding defect

**Successful bonding = respect of these surface requirements**

[Image sources: Y. Beillard PhD work, Tong 99]
DIRECT HYBRID BONDING PROCESS & CHARACTERIZATION

Direct hybrid bonding levels realization
Damascene Cu/SiO$_2$

Introduction of electrical test

SAM for bonding quality
Bonding defect due to particule
IMPACT OF ELECTRICAL TEST ON CU SURFACE

Cantilever probes chosen for this work

Cantilever probe mark on Cu by optical interferometry

Surface requirements for direct bonding not respected

Incompatibility of testing and direct hybrid bonding
METHODOLOGY

Testing on Cu pads

- Mz BEOL levels
- HB levels
- Electrical test
- Hybrid bonding

Challenges:
- Minimization of overall topography
- Probe marks planarization

→ Bonding quality → product yield

How to minimize topography post testing to get a good bonding quality?

1. Testing: impact of test parameters
2. Design: impact of pad design
3. Process: impact of CMP process
1. TESTING : OVERDRIVE IMPACT ON TOPOGRAPHY

Topography measurement on tested Cu pad

Overdrive: ensure good electrical contact

- Topography range: 750nm to 3µm

Increased overdrive rises topography

- Overdrive: ensure good electrical contact

Final chuck position

Chuck position @ contact

- Topography range: 750nm to 3µm
1. TESTING: OVERDRIVE IMPACT ON BONDING QUALITY

Test + planarization + SAM post electrical test

Overdrive variation within 1 wafer

Overdrive 75µm

Bonding defect width 120µm

Overdrive 60µm

Bonding defect width 80µm

Overdrive 45µm

No bonding defect observed

Overdrive 30µm

No bonding defect observed

Reduced overdrive reduces bonding defect ➔ Trade-off with contact resistance
2. DESIGN : IMPACT OF PAD DESIGN ON BONDING QUALITY

Standard pads
Successfull integration of test pad with no bonding defect

Advanced pads
SAM post bonding w/o test
-75% topography

Bonding defect
Bonding OK

Successfull integration of test pad with no bonding defect
3. PROCESS : CMP PROCESS IMPROVEMENT

SAM after CMP post test

POR CMP Process

Improved CMP Process

Bonding defects = test pads

Bonding OK

Improved CMP process reduces overall topography
CONCLUSIONS ON KGD STRATEGIES COMPATIBLE WITH DIRECT HYBRID BONDING

Objectives: Define a test strategy compatible with direct hybrid bonding

High quality bonding achieved after testing on hybrid bonding level

KGD strategy compatible with direct hybrid bonding validated

- Defect topography reduction by testing parameters (overdrive)
  → Trade-off with electrical contact

- Successful integration of advanced test pads with no bonding defect

- CMP process improvement leading to minimum final defect topography
This work was partly funded thanks to the French national program “Programme d’Investissement d’Avenir”, IRT Nanoelec ANR-10-AIRT-05
Introduction

In the context of high performance computing, the integration of more computing capabilities with generic cores or dedicated accelerators for AI application is raising more and more challenges. Due to the increasing costs of advanced nodes and the difficulties of shrinking analog and circuit IOs, alternative architecture solutions to single die are becoming mainstream. Chiplet-based systems using 3D technologies enable modular and scalable architecture [1]. The current passive interposer solutions – silicon passive interposers [2] or organic substrates [3] – brings clear cost reduction by smart technology partitioning [4] and using the so-called Know Good Die (KGD) approach [5]. Nevertheless, they still lack flexible efficient long-distance communications, smooth integration of chiplets with incompatible interfaces, and easy integration of less-scalable analog functions, such as power management and system IOs.

In [6][7], we have presented the first CMOS active interposer, integrating i) power management without any external components, ii) distributed interconnects enabling any chiplet-to-chiplet communication, iii) system infrastructure, circuit IOs, and the associated Design-for-Test solution. The INTACT circuit prototype (fig 1) integrates 6 chiplets in FDSOI 28nm technology, which are 3D-stacked onto an active interposer in 65nm process, offering a total of 96 computing cores (Fig. 1).

![INTACT 3D Design-for-Test Solution and Chiplet-based Active Interposer Architecture](image)

In terms of complexity: 150,000 3D connections are performed using µ-bumps (20 µm pitch) between the chiplets and the active interposer, with 20,000 connections for system communication, using the various 3D-communication interfaces, called 3D-Plugs, and 120,000 connections for power supplies using the integrated voltage regulators (SCVRs); while 14,000 TSVs are implemented for power supplies and off-chip communication.

Testability Challenges

With such 3D active interposer, as for passive interposers, testability is raising various challenges. First, it is required to ensure Know-Good-Die (KGD) sorting to achieve high system yield [8]. This implies that the 3D test architecture must enable Electrical Wafer Sorting (EWS) test of the chiplet and the interposer (pre-bond test, before 3D assembly), and final test (post-bond, after 3D assembly in the circuit package). Moreover, due to fine pitch µ-bumps, reduced test access is observed, µ-bumps cannot be directly probed in test mode. This implies to include additional IO pads, which are only used for test purpose, and not in functional mode (see Fig. Fig. 2).
Finally, with 3D technologies, additional defects may be encountered, such as μ-bumps misalignments, TSV pinhole, shorts, etc. which lead to specific care for testing the 3D objects and interfaces. Another concern is also regarding the Automatic Test Pattern Generation (ATPG) engineering effort, where easy re-targeting of test patterns from pre-bond test to post-bond test should be proposed to reduce test development efforts.

Numerous researchers have addressed specific test solutions for 3D defaults [9][10], for testing generic 3D architectures using die wrappers and elevators [11], and for testing 2.5D passive interposers [12]. A standardization initiative on 3D testability has emerged with the P1838 proposal, with recent outcomes and results [13]. Nevertheless, no work addressed initially the testability of active interposers.

3D Design-for-Test Architecture

Within the INTACT architecture, the test of the 3D system must address the test of all the following elements: i) the regular standard-cell based logic, ii) all memories using BIST engines and Repair, iii) the distributed 3D interconnects and IOs: 3D connections of active links and passive links, which are implemented by micro-bumps, and finally iv) the regular package IO pads for off-chip communication through the TSVs.

In order to test the Active Interposer and its associated chiplets, the proposed 3D Design-for-Test architecture (Fig. 3) is based on the two following main Test Access Mechanisms (TAMs), as proposed earlier in [14]:
- A IJTAG IEEE1687 hierarchical and configurable chain, accessed by a primary JTAG TAP port, for testing all the interconnects and memories, based on the concept of “chiplet footprint”,
- A Full Scan logic network using compression logic, for reduction of test time and of number of test IOs.

By using IJTAG IEEE 1687, the JTAG chain is hierarchical and fully configurable: the JTAG chain provides dynamic access to any embedded test engines. The active interposer JTAG chain is designed similarly to a chain of TAPs on a PCB board. It is composed of “chiplet footprints”, which provide either access to the above 3D-stacked chiplet or to the next chiplet interface, and which are chained serially. The JTAG network is used to test and control the 3D active links, the 3D passive links, the off-chip interfaces, and the embedded test engines, such as the memory BISTs. This TAP chain presents a reduced area impact and reduced 3D pin count.

The Full Scan logic network offers efficient and parallel full scan test of the whole 3D system logic. In order to reduce the number of 3D parallel ports, compression logic is used in both the chiplets and the active interposer, with a classical tradeoff (shift time/pin count). Independent scan paths are used between the chiplets and the active interposer, to facilitate the test architecture integration.
Test CAD Flow and Test coverage

The proposed 3D Design-for-Test architecture has been designed and inserted using Tessent™ tools from Mentor, a Siemens Business [15]. By using IJTAG and IEEE1687, high level languages such as “Instrument Connectivity Language” (ICL) and “Procedural Description Language” (PDL) are provided and enable to handle the complexity of such a system. In particular, it is possible to fully-automate the test pattern generation of Memory BIST engines, from ATPG at chiplet level to ATPG of the same patterns within the full 3D system, enabling so-called test pattern retargeting. As presented in Table I, full testability is achieved for all logic, 3D interconnects and regular package IOs, and memory BIST engines, before 3D assembly and after 3D assembly.

** Table I: INTACT DESIGN-FOR-TEST RESULTS**

<table>
<thead>
<tr>
<th>DFT access</th>
<th>Active Interposer 65nm</th>
<th>Chiplet FDSOI 28nm</th>
<th>Full 3D System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Scan</td>
<td>32 scan chains,</td>
<td>182 scan chains,</td>
<td>#faults 134.8M</td>
</tr>
<tr>
<td></td>
<td>4 after compression,</td>
<td>16 after compression,</td>
<td>Test cov. 95.5%</td>
</tr>
<tr>
<td></td>
<td>#faults 5.7M,</td>
<td>#faults 21.5M,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Test cov. ~60%**,</td>
<td>Test cov. 97.06%,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1318 patterns</td>
<td>1790 patterns</td>
<td></td>
</tr>
<tr>
<td>IJTAG + Interec.</td>
<td>All IO pads pre-bond</td>
<td>All IO pads pre-bond</td>
<td>3D IO pads post-bond</td>
</tr>
<tr>
<td>Boundary Scan</td>
<td>2D pads (826)</td>
<td>2D pads (249)</td>
<td>(13 548)</td>
</tr>
<tr>
<td></td>
<td>3D IO (13 548)</td>
<td>3D IO (2.258)</td>
<td>27 patterns</td>
</tr>
<tr>
<td></td>
<td>81 patterns</td>
<td>68 patterns</td>
<td></td>
</tr>
<tr>
<td>BIST &amp; Repair</td>
<td>#BIST: 1</td>
<td>#BIST: 5</td>
<td>#BIST: 31</td>
</tr>
<tr>
<td></td>
<td>12 patterns / BIST</td>
<td>20 patterns / BIST</td>
<td>612 patterns</td>
</tr>
</tbody>
</table>

** Limited test coverage is reported by the tool within the interposer, this is due to the asynchronous NoC that can be tested using a dedicated test solution not reported here.

Using the proposed DFT architecture & generated test patterns, the full system was tested using an Automated Test Equipment (ATE):
- The 28nm chiplet has been tested at wafer level using a dedicated probe card, with a binning strategy.
- The active interposer has not been tested at wafer level, supposing the maturity of the 65nm technology and its high yield due to its low complexity. Nevertheless, its standalone DFT and dedicated IO test pads on the front face were initially planned and designed as mentioned above.
- The full INTACT circuit, after 3D assembly and packaging, has been tested within a dedicated package socket.
Conclusions and Perspectives

3D integration and Active Interposer open the way towards efficient integration of large-scale chiplet-based computing systems. Such scheme can be applied for integration of similar chiplets as presented with the INTACT circuit in this paper, but also for smooth integration of heterogeneous computing chiplets [16].

Regarding testability and DFT, the proposed solution allows to perform KGD sorting of both the chiplet and the active interposer, with final test of the full system. The DFT solution is based on existing DFT standard (IEEE1687 and compressed full scan) and tools. A chain of TAP, called chiplet footprint, offers a modular and scalable DFT for any number of chiplets within the active interposer. The solution has been successfully implemented and tested using the Mentor Graphics Tessent tool suite.

Regarding chiplet integration, it is currently complex to integrate chiplets from different sources, due to missing standards, even if strong standardization initiative are on-going [17, 18]. With passive interposers, wire-only interposers prevent the integration of chiplets using incompatible protocols, while active interposer enable to bridge them easily by ad-hoc logic within the active interposer. This has been proposed for instance as a generic connectivity, as adopted by zGLUE Inc. [16].

Regarding 3D technology, the technologies are still evolving to provide more advanced chiplet integration, with reduced pitches and improved thermo-mechanical behavior. Hybrid bonding technology initially devoted for Wafer to Wafer are also appearing for chip2wafer assembly, with reduced pitches (10µm pitch as of today) [20], while also proposing adequate solutions for KGD sorting on 3D copper pad physical interfaces [21].

Acknowledgements:

This work was partially funded thanks to the French national program "Programme d'investissements d'Avenir, IRT Nanoelec” ANR-10-AIRT-05”.

References:

[6] P. Vivet et al., “A 220GOPS 96-Core Processor with 6 Chiplets 3D-Stacked on an Active Interposer Offering 0.6ns/mm Latency, 3Tbs/mm2 Inter-Chiplet Interconnects and 150mW/mm2@ 82%-Peak-Efficiency DC-DC Converters”, 2020 IEEE International Solid-State Circuits Conference - (ISSCC).
[19] https://www.zglue.com
3D Design-for-Test Solution for Chiplet-based Active Interposer Architecture

P. Vivet\textsuperscript{a}, J. Durupt\textsuperscript{a}, S. Thuriès\textsuperscript{a}, D. Dutoit\textsuperscript{a}, E. Bourjot\textsuperscript{b}, S. Cheramy\textsuperscript{b}

\textsuperscript{a} Univ. Grenoble Alpes, CEA, LIST, 38000 Grenoble, France

\textsuperscript{b} Univ. Grenoble Alpes, CEA, LETI, 38000 Grenoble, France

{pascal.vivet@cea.fr}
High Performance Computing & Big Data

• More cores + more accelerators + more memory
  • Similar constraints are appearing for embedded HPC (Automotive, etc)
  • Need both highly optimized generic and specialized functions (i.e. ML/AI accelerator)
  • Need a « go-to-market » solution for sustainable system differentiation

• System designers must offer:
  • Modular and cost effective solutions
  • Energy efficiency of the system infrastructure
  • More on-chip memory bandwidth per core

➢ With advanced CMOS issues, « Single Die » solution is not viable anymore

P. Vivet, 3D CHIPLET TEST Workshop, 6 Nov. 2020
Chiplet Partitioning

• Chiplet motivations
  • Cost driven (*yield through KGD sorting*)
  • Modularity driven (*lego*)
  • Heterogeneous integration (*using the right technology*)

• Chiplet challenges?
  • Eco-system maturity,
  • Technology & Architecture partitioning,
  • Chiplet Interfaces, Thermal, Sign-off, 3D CAD flow,
  • And ... *testability and DFT architecture*

[D. Dutoit, Keynote, 3DIC’2014]
## Chiplet Partitioning: Solutions and Limitations

### Existing technologies

<table>
<thead>
<tr>
<th>3D Vertical Stacking</th>
<th>Organic Substrates</th>
<th>Passive interposer (2.5D)</th>
<th>Silicon bridges</th>
<th>Active Interposer - this work -</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="3D Vertical Stacking" /> INTEL, Foveros, Lakefiled CPU ISSCC'2020</td>
<td><img src="image2.png" alt="Organic Substrates" /> AMD, 4-chiplet / ISSCC'2018 / 8-chiplet + IODie / ISSCC'2020</td>
<td><img src="image3.png" alt="Passive interposer (2.5D)" /> TSMC, CoWos, VLSI'2019</td>
<td><img src="image4.png" alt="Silicon bridges" /> INTEL, EMIB bridge, ISSCC'2017</td>
<td><img src="image5.png" alt="Active Interposer" /> CEA, INTACT, ECTC'2019, ISSCC'2020</td>
</tr>
</tbody>
</table>

### But, some limitations

- Chiplet communication limited to side-by-side communication, not scalable
- How to integrate heterogeneous chiplets & differentiating functions?
- How to integrate less-scalable functions (IO’s, analogs, power management)?
Active Interposer : Principle

Scalable & Distributed NoCs
Any chiplet-to-chiplet traffic

Chiplets:
Clusters of Cores

Active Interposer

Power Management
Close to cores

SoC infrastructure
Analog, IOs, PHY, DFT

Additional features

➢ Mature CMOS technology (with low logic density to preserve system cost)

P. Vivet, 3D CHIPLET TEST Workshop, 6 Nov. 2020
Outline

• Chiplet-based design and Active Interposer
• INTACT Circuit Architecture
• 3D DFT architecture solution
• Test flow & results
• Conclusion & perspectives
Outline

• Chiplet-based design and Active Interposer
• INTACT Circuit Architecture
• 3D DFT architecture solution
• Test flow & results
• Conclusion & perspectives
6 Chiplets 3D-stacked on an Active Interposer

- **Chiplet Overview**
  - 4 cluster of 4 cores
  - Distributed L1$ + L2$ + L3$
  - Scalable Cache Coherency

- **Active Interposer**
  - Distributed flexible interconnects
  - Integrated SCVRs (1/chiplet)
  - Memory Controller & System IO’s
  - SOC Infrastructure, DFT

---

P. Vivet et al., ISSCC’2020

[100 100 100]
6 Chiplets 3D-stacked on an Active Interposer

• **Chiplet Overview**
  - 4 cluster of 4 cores
  - Distributed L1$ + L2$ + L3$
  - Scalable Cache Coherency

• **Active Interposer**
  - Distributed flexible interconnects
  - Integrated SCVRs (1/chiplet)
  - Memory Controller & System IO’s
  - SOC Infrastructure, DFT

> 2 technology nodes difference between chiplets & bottom die

[P. Vivet et al., ISSCC’2020]
INTACT Circuit Overview

• Die technologies
  • Chiplet: FDSOI 28nm, ULV + BodyBias, 22mm²
  • Active Interposer: CMOS 65nm, MIM option, 200mm²

• 3D technology integration
  • μ-bumps, 20µm pitch (150 k)
  • TSV middle, 40 µm pitch
  • Face2Face assembly on package substrate
  • 6 chiplets

3D integration and final package

[P. Coudrain et al., ECTC’2019]
Outline

• Chiplet-based design and Active Interposer
• INTACT Circuit Architecture
• 3D DFT architecture solution
• Test flow & results
• Conclusion & perspectives
3D Technology : Test Challenges & State-of-the-Art

• Test model?
  • New defects with 3D: misalignment, cracks, open, etc.

• Test flow? ➞ ensure Chiplet Known-Good-Die (KGD)
  • Pre-bond (Known-Good-Die), Mid-bond, Post-bond, Final Test
  • Yield versus test-cost trade-offs (various cost model existing)
    [D. Gitlin, S3S’2017]

• Test access & architecture?
  • Reduced test access due to small µ-bump pitch
  • Use TSVs as “elevators” for test signals
  • 3D Test Architectures based either on IEEE 1149.1 or IEEE 1500
  • New standard IEEE 1838 for 3D test

• Test of Interposers?
  • Passive interposers (2.5D): only wires, test after assembly
  • Active interposers: no proposed solution yet (... when this design was started in 2014 ...)
  • Need to improve 3D test flow for easy transition from 2D test to 3D test

P. Vivet, 3D CHIPLET TEST Workshop, 6 Nov. 2020
Proposed Solution for KGD on Active Interposer

• Flexible 3D-DFT architecture
  • Configurable and hierarchical IJTAG chain
  • IJTAG test interface, using IEEE1687 standard
  • Parallel scan chains, with test compression

Let’s chain the Chiplets on the active interposer as chips on a regular PCB board

• Good for
  • Modularity
  • Configuration
  • DFT timing closure

P. Vivet, 3D CHIPLET TEST Workshop, 6 Nov. 2020
**INTACT : 3D DFT IEEE1687 based Architecture**

- **Configurable hierarchical IJTAG chain – modularity driven –**
  - IJTAG test interface, using IEEE1687 standard
  - Boundary Scan Chains, for testing 3D interconnections
  - SIB/TDR, for executing Memory BISTs & Repair

- **Parallel scan chains, using test compression – test time driven –**
  - Full scan standard support
  - All dies have same // scan inputs
  - Each die has dedicated // scan outputs

---

**2 Test Access Mechanisms**

- **µ-buffer cell:**
  - Bi-directional 3D IO (*enables KGD test*)
  - Level Shifter
  - Pull down
  - ESD protection

- **µ-bumps 20µm pitch**

---

P. Vivet, 3D CHIPLET TEST Workshop, 6 Nov. 2020
INTACT : 3D DFT IEEE1687 based Architecture

- Configurable hierarchical IJTAG chain – *modularity driven* -
  - IJTAG test interface, using IEEE1687 standard
  - Boundary Scan Chains, for testing 3D interconnections
  - SIB/TDR, for executing Memory BISTs & Repair

- Parallel scan chains, using test compression – *test time driven* -
  - Full scan standard support
  - All dies have same // scan inputs
  - Each die has dedicated // scan outputs

2 Test Access Mechanisms

P. Vivet, 3D CHIPLET TEST Workshop, 6 Nov. 2020
Configurable 3D TAP Chain: Chiplet Footprint architecture

« Chiplet footprint » within Active Interposer
- Aligned with physical footprint (\textit{! helps timing !})
- « Present bit » for automatic die detection
- Configuration through TestDataRegister (TDR)
- Each TAP integrates a Die ID and a Chiplet ID
- A TAP chain with 2D / 3D / Bypass muxing

\[\text{Similar to a TAP serial chain on a PCB}\]

<table>
<thead>
<tr>
<th>Interposer</th>
<th>Mode</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre-bond test</td>
<td>2D mode</td>
<td>present=0. Bypass to next chiplet footprint</td>
</tr>
<tr>
<td>post-bond test</td>
<td>Reset or Bypass</td>
<td>At reset, chiplet is not connected; or select=0 to bypass a defective chiplet</td>
</tr>
<tr>
<td></td>
<td>3D mode</td>
<td>select=present=1, chiplet TAP is connected</td>
</tr>
</tbody>
</table>

\[J. \text{Durupt et al., ETS'2016}\]

- Passive links
- Active links

\[\text{Continuity test of 3D connections in multiple passes using boundary scan}\]
Outline

• Chiplet-based design and Active Interposer
• INTACT Circuit Architecture
• 3D DFT architecture solution
• Test flow & results
• Conclusion & perspectives
DFT Test Flow using IJTAG 1687

**Objective**
- Ease test pattern re-targeting from chiplet to package
  - Generate chiplet test patterns for KGD test
  - Retarget chiplet test patterns for final package test
  (as done for re-generation of IP-level test pattern to SOC-level)

**With IJTAG IEEE1687**, use high level coding languages
- ICL: “Instrument Connectivity Language”, test hierarchy
- PDL: “Procedure Description Language”, test protocol

Use Tessent™ tools for automatically retargeting
- test patterns
- test benches

Set `Current_design chiplet_top`
iCall BIST_controller_0.start_BIST

Set `Current_design 3D_interposer_top`
iCall chiplet0.BIST_controller_0.start_BIST

Preprocess inputs and create internal database

Convert 1687 solver results into Verilog testbench & WGL/STIL/SVF

Perform 1687 design rule checks

TCL API for 1687 introspections

Retarget PDLs from instrument level to wanted (chip) level (1687-Solver)

Retargetting

[Y. Fkih, 3D TESTWS @ ITC’2013] [ETS’2016]
# INTACT Design-for-Test Results

<table>
<thead>
<tr>
<th></th>
<th>Interposer 65nm</th>
<th>Chiplet FDSOI 28nm</th>
<th>Full 3D System</th>
<th>CAD Test Flow</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full Scan</strong></td>
<td>32 scan chains</td>
<td>182 scan chains</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4 after compression</td>
<td>16 after compression</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>#faults 5.7 M</td>
<td>#faults 21.5 M</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Test cov. ~ 60% **</td>
<td>Test cov. 97.1 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1318 patterns</td>
<td>1790 patterns</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Boundary Scan</strong></td>
<td>All IO pads in pre-bond</td>
<td>All IO pads in pre-bond</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- #2D pads (826)</td>
<td>- 2D pads (249)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- 3D IO cells (13 548)</td>
<td>- 3D IO cells (2 258)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>81 patterns</td>
<td>68 patterns</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BIST &amp; Repair</strong></td>
<td>#BIST = 1 12 patterns / collar</td>
<td>#BIST = 5 20 patterns / collar</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>81 patterns</td>
<td>68 patterns</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>#BIST = 31 612 patterns total</td>
<td>#BIST = 31 612 patterns total</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tessent TestKompress</td>
<td>Tessent ATPG</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Generation is available</td>
<td>Full ATPG automation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2D pads, 3D pads, 3D links)</td>
<td>(2D pads, 3D pads, 3D links)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tessent IJTAG</td>
<td>Full retargeting</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>is available with ICL/PDL</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Limited test coverage is reported by the tool within the interposer, due to the asynchronous NoC that can be tested using a dedicated test solution not reported here.**

- **Test of a very complex 3D system achieved (96 cores, 40MB Cache, > 20 k IOs)**
- **Full automation available for DFT insertion**
- **Full automation available for easy retargeting of BIST tests from 2D to 3D**
INTACT Test Strategy

• Know Good Die (KGD) strategy:
  • EWS chiplet test before BGA assembly
    • Probe card at wafer level
    • Test access on standard IO ring (not on 3D μ-bumps)
  • No Interposer test
    • Hypothesis of mature CMOS 65nm technology
    • DFT is available in case EWS test would be requested
  • Final test after packaging
    • Package level test in a socket

• How?
  • ATPG using Mentor Tessent © tools
    • Chiplet test patterns
    • Active interposer test patterns
    • Re-compute chiplet test patterns in context of interposer IO access
  • Binning strategy according to EWS results
Outline

• Chiplet-based design and Active Interposer
• INTACT Circuit Architecture
• 3D DFT architecture solution
• Test flow & results
• Conclusion & perspectives
Conclusion

• 3D DFT & Test of Active Interposer
  • Chaining chiplets with a configurable & hierarchical IJTAG chain
  • Supported by Mentor Graphics Tessent™ tools
  • Full DFT & test achieved on the INTACT circuit

• Chiplet’s based design is here and is pervasive
  • High Performance Computing, eHPC (automotive, CPS, …), AI accelerators,
  • With pitch reduction, new bonding techniques, upcoming die interfaces, etc.

• 3D TEST Standard
  • IEEE 1838 is now available (accepted March’2020)
  • JTAG + IEEE1500 compatible, with a Flexible Parallel Port (FPP)
Perspectives for 3D & Test

• 3D Test industrial solutions are now available:
  • For HBM memory, DFT is fully supported, using IEEE 1500 wrappers, available in HBM PHY IPs
  • But still many research issues to come ....

• « At speed » test of 3D interfaces: how? shared chiplet test responsibilities due to this context?

• Continue CAD tool developments
  • 3D Cost modelling tool (yield, test)
  • DFT tool alignment with IEEE 1838 standard
  • Interaction 3D test pattern ⇔ power grid effects ⇔ thermal effects (« Test HotSpots »)

• Continue pushing the limits of 3D pitch reduction for chiplet integration:
  • More parallelism, better energy efficiency, in line with chiplet aggressive CMOS nodes
  • From thermal compression & µbumps (20-40 µm target) to hybrid bonding (5 µm target)
  • 3D integration with better thermo-mechanical reliability
    ➤ Test challenges ????
    • Test on copper interfaces
    • Finer pitch => landing pads? do we need to test all 3D interconnects after 3D assy?

[A. Jouve, E. Bourjot, 3DIC’2019]
This work was partially funded thanks to the French national program "Programme d'investissements d'Avenir, IRT Nanoelec" ANR-I0-AIRT-05".

Many Thanks to Mentor Graphics for active collaboration on this research project
BIST and BISR-based 3DIC interconnect interface test and repair

Changming Cui 1, Zhe Liu 1, Junlin Huang 1

1 Hisilicon Technologies Co., LTD, Shenzhen, China

Abstract—In the current 3D stacked chips, the number of interconnect interfaces, including Through Silicon Via (TSV) and Hybrid Bond, is usually very large, ranging from hundreds to thousands or even tens of thousands in one inter-die surface. In terms of test quality, yield, and test cost, testing and repairing these interfaces is very important. This paper proposes a test and repair method for the 3 Dimension Integrated Circuits (3DIC) interconnect interface. The test of the inter-chip interconnection interface is implemented through the Build In Self-Test (BIST) circuit, and the interconnect interface repairing through the Build In Self-Repair (BISR) chain and the BISR controller automatically. This solution eliminates the quality risks of manual operations, has no additional requirements for the test procedure, and supports the subsequent test items to be tested based on the repaired state. In addition, this paper proposes a data repair compression algorithm based on the failure characteristics of interconnect interfaces, which can increase the repairable number of interconnect interfaces under the same EFUSE space size. This paper also introduces a solution for multiple repairs to adapt to the added faults caused by the multiple stacking process of 3DIC. And a BISR controller sharing method is discussed about, which enables multiple dies to share the same BISR controller, and other dies to only implement the BISR chain to reduce area overhead.

Keywords—3DIC Test, interconnect interface test

1. INTRODUCTION

With the ever-increasing scale of integrated circuits, the chip size of traditional 2D chips has become a bottleneck in chip development, and their large size also causes worse timing and lower yield. The 2.5D and 3D chips can effectively alleviate these problems and become an important way to achieve "more than Moore". Compared with 2D chips, the production and stacking process of 3DIC are more complex, and the stack types become more diverse. A variety of new interconnect interfaces, such as TSV and Hybrid Bond, bring about new failure models [1][2]. Therefore, 3DIC testing becomes particularly important, yet quite challenging [3].

Compared with 2D chips, the greatest difference of 3DIC lies in the interconnect interfaces between dies, such as TSV and Hybrid Bond. These interconnect interfaces are usually in very fine pitch (with a size of a few um), making it hard to probe on the test machine. Although the industry has introduced probe cards with very fine pitch [4], the probe pitch still cannot meet the requirements of Hybrid Bond (which is shorter than 10um), and the cost is relatively high. In addition, the number of interfaces is very large (one inter-die surface may comprise thousands or even tens of thousands of such interfaces), which imposes high requirements on the test coverage, and various methods, such as redundant repair, are in need to improve the yield. Generally, the repair of the interconnect interfaces is often achieved by software programming. The test machine analyzes the test report to locate the fault of the interconnect interface, employs the calculation method to analyze the repair data, and uses the program pattern to burn the repair data into the EFUSE. When the chip powers on next time, it is necessary to load the initial configuration, read and parse the repair data from the EFUSE, and load it to the repair circuit by register accessing to complete the repair flow of the interconnect interface. This method requires program analysis to calculate the repair value and edit the programming vector on the test machine, which increases the complexity of the test program, and introduces quality risks caused by manual operations. In addition, during the testing process, the repair data will be lost when the power is turned on again for subsequent test items, so it is difficult for the test machine to analyze the repair data and to control the accessing of the internal registers to complete the repair flow, which adversely affects the test quality of the subsequent test items. For TSV repairing, the industry has also proposed a variety of solutions. For example, some people propose an in-field TSV repair method, which adjusts the TSV redundancy selection through in-field configurable circuits [5]. And some people propose a repairable and reliable TSV set structure, using redundant resource sharing and reusing to improve the utilization of redundant TSVs [6]. These methods greatly reply on the participation of redundant control algorithm. If the algorithm is implemented on the chip by CPU hardware, the area overhead will be very large. If it is implemented outside the chip by using a program, the requirements for the test program are very high, which obviously influences the test process.

For the above problems, this paper proposes a BIST test method for the 3DIC interconnect interface and an automatic repair method based on the BISR chain. We can use BIST circuits to test the interconnect interface, automatically generate the repair data, and update the data to the fixed-sequence BISR chain, shifting the BISR chain data into BISR controller and programming it to EFUSE. At subsequent power-on, the chip reads the repair data from the EFUSE by the hardware circuit and shifts it to the original position in the BISR chain, which automatically completes the redundant selection of the interconnect interface. In this solution, the calculation, programming, and power-on loading of the repaired data are all automated by the hardware circuit, which eliminates the quality risk of manual intervention, has no additional requirements for
the test program, and supports the subsequent test items to be tested based on the repaired state, which improves the test quality.

The remainder of this paper is organized as follows. Section 2 introduces the IO BIST and repair methods, Section 3 the BISR data compression technology, Section 4 the IO multiple repair method, Section 5 the BISR controller sharing structure, Section 6 the test results, and Section 7 concludes this paper.

II. IO BIST TESTING AND REPAIRING

A. IO BIST testing

In many 3DIC chips, there are thousands of interconnect interfaces such as micro bumps or Hybrid Bonds between dies, and corresponding IO cells inside the dies. In order to test the connectivity and high-speed performance of the interconnect interface, the IOBIST circuit and repair circuit are designed. IO BIST and repair functions are implemented using Linear Feedback Shift Register (LFSR) circuits for testing and training the interconnect channels.

Each die can contain multiple sets of LFSR_UNIT, each LFSR_UNIT contains a LFSR_GEN module and a LFSR_CHK module to generate and compare LFSR data (Fig.1). Each LFSR_UNIT contains some data/control IOs, 1 clock micro IO, and 1 redundant micro IO. LFSR_UNIT can support internal die loopback test and inter-die interconnection test. According to different stacking manufacturing processes, we can configure the test mode to adapt to the test scenario.

Fig. 1. LFSR-based IO BIST diagram of 3DIC

B. IO repairing

Assuming each LFSR_UNIT contains 30 data/control IOs, 1 clock IO, and 1 redundant IO. The redundant IO can repair one bit in the 30 data/control IOs. During test, LFSR_GEN continuously generates LFSR data, and LFSR_CHK detects the received data. When LFSR_CHK detects an error data, it will report the “fail” flag and generate a 5-bit repair data. The data value can indicates the location of the error and guides the repairing: 5'b00001 indicates that IO bit 1 is failed, and the redundant IO is selected to replace IO bit 1; 5'b00010 indicates that IO bit 2 is failed, and the redundant IO is selected to replace IO bit 2; ... 5'b11110 indicates that IO bit 30 is failed, and redundant IO is selected to replace IO bit 30. If there is no failure in the LFSR_UNIT, the 5bit repair data is 5'b00000. In addition, for the case of 2 or more failing IOs, LFSR_CHK reports the “multi-fail” flag, and the 5bit repair data is 5'b11111, indicating that it cannot be repaired. Fig.2 shows the IO repair diagram.

At the same time, we can concatenate all LFSRUNIT corresponding BISR registers into a BISR chain, and connect the BISR chain with the BISR controller (Fig.3).

Fig. 2. IO repair diagram of 3DIC

Fig. 3. BISR chain and BISR controller diagram of one single die

When the IO BIST test is completed, all BISR data have been generated. Then BISR_TOP sends the efuse_busy signal to logic 1, making the BISR registers in the serial mode, and then shifts these data into the BISR controller through the BISR chain, compresses the data and programs it into EFUSE, and completes the IO repair progress. At the next power-on, the BISR controller automatically reads EFUSE data, decompresses the repair data, and shifts it into the corresponding position of the BISR chain. If the number of shift cycles is correct, the BISR
data value corresponding to each LFSR_UNIT is sure to be correct, and the redundancy IO selection is completed. The chip can automatically carry out the repair data calculation, programming, and power-on loading by circuits, without the need for an external test machine to control the calculation processes.

III. BISR DATA COMPRESSING

In the current 3DIC, the number of interconnection IOs is usually large, ranging from hundreds to thousands or even tens of thousands. However, as the yield of 3DIC interconnect interfaces is very high, there are fewer IOs that need to be repaired, and most of the BISR data is 0. From the perspective of the entire BISR chain, most of the values are continuous 0s, and some 5bit non-zero data will appear sporadically.

Aiming at the above characteristics, we propose the BISR value compressing technology. The main scheme is to compress consecutive 0 data segments and convert them into the count value in binary, while maintaining the original non-compression for non-zero data (Fig.4).

<table>
<thead>
<tr>
<th>EFUSE bit</th>
<th>15</th>
<th>14</th>
<th>13</th>
<th>12</th>
<th>11</th>
<th>10</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>addr 0</td>
<td>count number of continuous 0</td>
<td>uncompressed data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>addr 1</td>
<td>count number of continuous 0</td>
<td>uncompressed data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>addr n</td>
<td>count number of continuous 0</td>
<td>uncompressed data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 4. Example of BISR data compress scheme

Assuming that each address of EFUSE contains 16bit data, it can be divided into a piece of 11bit compressed data (continuous 0 count number) and a piece of 5bit uncompressed data (non-zero BISR data), so each EFUSE address can record at most 2047 continuous 0s. When the number of continuous 0s exceeds 2047, you can continue recording at the next address. For the failed IO group, the BISR value is 5bit non-zero data, which can be recorded in the uncompressed data segment. As mentioned above about the fault characteristics of the 3DIC interconnection signals, the value of the BISR chain is usually continuous 0, and a small amount of sporadic 5bit non-zero data is separated. Therefore, it can be considered that after compression, each address of EFUSE can record one fault. In this way, the repair ability of IO is basically not limited by the total length of the BISR chain, but is proportional to the size of the EFUSE storage.

Assuming that a chip has a total of \( n \) data/control interconnect interfaces, and every \( p \) interfaces form a group, each group has 1 redundant interface, and the group with the interconnect interface failure is not continuous, then the EFUSE storage requirement before compression \( M_{\text{pre-comp}} \) is:

\[
M_{\text{pre-comp}} = \lceil n/p \rceil \times \lceil \log_2 p \rceil
\]

And the storage requirement of 32bit width EFUSE after compression for the IO failure counts \( f \) is:

\[
M_{\text{post-comp}} = (f + 1) \times 16
\]

For example, if the chip has a total of 10,000 interconnect interfaces, every group has 30 interfaces, the EFUSE storage requirement before and after compression for different IO failure count are showed in Fig.5:

Fig. 5. The EFUSE storage requirement before and after compression for different IO failure count

It can be seen that when the supported IO failure count is less than 100, the compressed EFUSE space requirement is smaller than uncompressed. From the actual test results, the IO failure count is mostly of a single digit, and the EFUSE storage requirement is greatly reduced after compression. In addition, the greater significance of using compression technology is that the size of the EFUSE storage imposes no absolute constraint on the length of the BISR chain, and does only affect the repair capability.

IV. IO MULTIPLE REPAIR

For 3DICs, there are usually multiple stacking processes, and each stacking process may cause new failures. Therefore, in each stacking process, there are corresponding test flows (such as pre-bond test, mid-bond test, and post-bond test, which are shown in Fig.6) to ensure that failures will not be introduced into the subsequent processes [7] [8].

Fig. 6. Test flow planning of 3DIC

For the same interconnect interface, there may also be new faults in each stacking process, so the interface may be tested and repaired for multiple times. Based on this situation, a solution that supports multiple repairs is proposed. The scheme is to split the EFUSE space into multiple parts, and the repair data is only programmed in one of the parts at each test flow. In the next stacking process test flow, we perform a BIST test based on the latest repaired state, and at the same time identify the new fault IOs, update the BISR data to the corresponding location of the BISR chain, and move it into the BISR controller and program into EFUSE. At the subsequent power-on, the repair data of each part is read out simultaneously and OR into one
signal by bit, and then be shifted into the BISR chain. Two things need to be noticed here. First, the programmed BISR data only needs to include the newly generated BISR data at the current test flow, and there is no need to include the previous BISR data (because the previous BISR data will be read out and OR by bit at every power-on), otherwise it will affect the compression efficiency. Second, for EFUSE space division, more space should be reserved for the first repair, because usually the number of failures in the first test will be the largest, the subsequent stacking process has relatively little impact on the previous interconnect interfaces. For example, if the same interconnect interface in a 3DIC chip needs to go through three test and repair flows, then for an EFUSE with 128 addresses, we may assign the first 64 addresses to the first repair and 32 addresses to the subsequent second and third repairs respectively. This division plan here is for reference only and may vary according to different engineering data.

V. BISR CONTROLLER SHARING

In general, each die contains its own BISR controller and EFUSE for IO repairing, but we can leverage the BISR controller sharing to reduce area overhead. Especially in the case where one or more slave dies are interconnected with the master die, the BISR controller can be placed on the master die and be shared with the slave dies, and the slave dies only need to contain the BISR chains.

![Diagram of BISR controller sharing of 3DIC](image)

In fig. 7, die A is the master die, die B and die C are interconnected with die A respectively. Let’s have a look at the interconnect interfaces between die A and die B. After the BIST test, the LFSR_CHK on the receiver side generates the BISR data (marked in red), and then the BISR chains of the dies on both sides simultaneously shift to die A’s BISR controller, and pass through an OR gate before shifting into the BISR controller, so that the BISR data of the two dies is OR-operated by bit and merged into one signal, serially shifted into the controller. In this way, the data programmed in EFUSE is the data after the combination of the repair data of die A and die B. When the chips power on, the data will be shifted into the BISR chain of the two dies simultaneously, so that both the receiver and the sender of the interconnect interfaces have made the same redundant selection, ensuring the correctness of signal connectivity.

Similarly, for the test and repair between die A and die C, the same connection method is adopted, and the same BISR controller is shared by dividing the EFUSE storage space to reduce area overhead.

VI. TEST RESULT

In one of our logic on logic 3DICs, one master die is interconnected with three slave dies, and each die has thousands of interconnection IOs. We have designed BIST circuits and BISR chains in each die. Besides, the BISR controller and EFUSE are implemented in the master die, and the three slave dies share the BISR controller and EFUSE of the master die. It needs about 5000 bit EFUSE space with BISR data before compression, and with the BISR data compression technology, we use a 4096bit EFUSE to record the BISR data. Several flows are designed for this test, including low temperature and normal temperature final test.

The test result shows that the BIST circuit can accurately detect the connectivity and timing-related failures of the interconnect interfaces. Due to the low IO failure rate, the failed interconnect interface can be 100% repaired.

VII. CONCLUSION

This paper proposes a test and repair method for the 3DIC interconnect interfaces. The test data is sent and received in LFSR, and this method supports internal die loopback test and inter-die interconnection test. The IO repair is processed by the BISR chain and BISR controller circuits automatically, which reduces the programming complexity of the test machine and supports the test based on the repaired state. The BISR data compression technology is introduced to reduce the storage space requirement of repair data. This paper also proposes a solution for multiple repairs to adapt to the added faults caused by the multiple stacking process of 3DIC. Moreover, a solution for multiple dies to share one BISR controller is proposed, which reduces the area overhead.

REFERENCES


3DC-TEST
Nov 6–7, 2020
BIST and BISR-based 3DIC interconnect interface test and repair

Changming Cui, Zhe Liu, Junlin Huang
Hisilicon Technologies Co., LTD, Shenzhen, China
Content

• Background
• BIST and BISR-based interconnect interface test and repair
• BISR data compressing
• Multiple repair
• BISR controller sharing
• Conclusion
3DIC Technology landscape

### Package stacking
- **Multi-die Packaging**
  - Interposer “2.5D”
  - EMIIB bridge

### Die Stacking
- **µbump**

### Hybrid Bonding
- Wafer-to-Wafer
- Die-to-Wafer

### Wafer-to-Wafer Sequential Processing

### Transistor Stacking

<table>
<thead>
<tr>
<th>Pitch</th>
<th>400um</th>
<th>100um</th>
<th>40um</th>
<th>10um</th>
<th>1um</th>
<th>0.1um</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density</td>
<td>6.25</td>
<td>100</td>
<td>625</td>
<td>1E4</td>
<td>1E6</td>
<td>1E8</td>
</tr>
</tbody>
</table>

The count of interconnect interface grows rapidly, bringing challenges to testing and repairing.
3DIC interconnect interface test & repair

- Wafer probing [1]
  - >= 20um pitch
  - 40um pitch
  - Pitch gap between probe and Hybrid Bond
- Die Wrapper Register (IEEE1838-2019)

The flow of “Test – diagnosis – repair” is not so smooth

What we need is “test – repair” automatically

- TSV repairing
  - Router-based TSV repair [2]
  - Redundancy-sharing R^2-TSV [3]

TSV failing diagnosis and repair bring large area overhead or test flow complexity.

[2] Li Jiang, Fangming Ye, Qiang Xu, Krishnendu Chakrabarty, Bill Eklow. On effective and efficient in-field TSV repair for stacked 3D ICs. 2013 50th ACM/EDAC/IEEE Design Automation Conference (DAC), July 2013
BIST and BISR-based interconnect interface test and repair
BIST and BISR-Based interconnect interface test and repair

- LFSR-based IO BIST
  - Each IO group corresponds a LFSR_UNIT
  - Each LFSR_UNIT contains a LSFR_GEN and a LFSR_CHK
  - LFSR_UNIT can support internal die loopback test and inter-die interconnection test
BIST and BISR-Based interconnect interface test and repair

- BISR-based IO repairing
  - LFSR_CHK detects error data and generates repair value automatically.
  - For example, a group of 30 IOs corresponds 1 redundant IO, need 5bit repair data.

<table>
<thead>
<tr>
<th>Fail IO location</th>
<th>BISR data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bit[0]</td>
<td>5'b00001</td>
</tr>
<tr>
<td>Bit[1]</td>
<td>5'b00010</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Bit[29]</td>
<td>5'b11110</td>
</tr>
<tr>
<td>Multi-fail</td>
<td>5'b11111</td>
</tr>
<tr>
<td>No fail</td>
<td>5'b00000</td>
</tr>
</tbody>
</table>
BIST and BISR-Based interconnect interface test and repair

- BISR chain and BISR controller
  - Concatenate all BISR registers into a BISR chain.
  - BISR controller integrated: BISR data shift in -> program into EFUSE -> BISR data shift out(power-on) automatically.
  - Limited area overhead, no complex test flow.
  - No need for CHIP 2 ATE handshake, IO self-repair when power-on.
BISR data compressing
BISR data compressing

- The number of interconnection IOs is usually large -> BISR chain is much too long
- The yield of 3DIC interconnect interfaces is usually very high -> most of the BISR values are consecutive 0s, and some 5bit non-zero data will appear sporadically
- Compressing scheme: compress consecutive 0 data and convert them into the count value in binary, while maintaining the original non-compression for non-zero data

For the EFUSE which data width is 16bit, the compressing may presented below:

<table>
<thead>
<tr>
<th>EFUSE bit</th>
<th>15</th>
<th>14</th>
<th>13</th>
<th>12</th>
<th>11</th>
<th>10</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>addr 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>count number of continuous 0</td>
<td>uncompressed data</td>
</tr>
<tr>
<td>addr 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>count number of continuous 0</td>
<td>uncompressed data</td>
</tr>
<tr>
<td>addr n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>count number of continuous 0</td>
<td>uncompressed data</td>
</tr>
</tbody>
</table>
BISR data compressing

Assuming:

• a chip has a total of \( n \) interconnect interfaces
• every \( p \) interfaces form a group
• Each group has 1 redundant interface
• The group with the interconnect interface failure is not continuous
• EFUSE data width is 16bit
• IO failure count is \( f \)

EFUSE storage requirement before compression \( M_{\text{pre-comp}} \):

\[
M_{\text{pre-comp}} = \left\lceil \frac{n}{p} \right\rceil \times \left\lceil \log_2 p \right\rceil
\]

EFUSE storage requirement after compression \( M_{\text{post-comp}} \):

\[
M_{\text{post-comp}} = (f + 1) \times 16
\]

when the supported IO failure count is less than 100, the compressed EFUSE space requirement is smaller than uncompressed

After compressing, the size of the EFUSE storage does not absolute constraint the length of the BISR chain, but only affects the repair capability.
Multiple repair
Multiple repair

- Multiple stacking processes brings multiple test and repair flow.
- Multiple repair scheme:
  - Split the EFUSE space into multiple parts, each test flow program one of the parts.
  - The repair data of each part is read out simultaneously and OR into one signal by bit.
- Note:
  - The programmed BISR data only needs to include the newly generated BISR data at the current test flow, and there is no need to include the previous BISR data.
  - For EFUSE space division, more space should be reserved for the first repair, because usually the number of failures in the first test will be the largest.

<table>
<thead>
<tr>
<th>raw bit</th>
<th>count number of continuous 0</th>
<th>uncompressed data</th>
<th>count number of continuous 0</th>
<th>uncompressed data</th>
</tr>
</thead>
<tbody>
<tr>
<td>addr 0</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
</tr>
<tr>
<td>addr 00</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
</tr>
<tr>
<td>addr 01</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
</tr>
<tr>
<td>addr 02</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
</tr>
<tr>
<td>addr 03</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
</tr>
<tr>
<td>addr 04</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
</tr>
<tr>
<td>addr 05</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
</tr>
<tr>
<td>addr 06</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
</tr>
<tr>
<td>addr 07</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
</tr>
<tr>
<td>addr 08</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
</tr>
<tr>
<td>addr 09</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
</tr>
<tr>
<td>addr 10</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
</tr>
<tr>
<td>addr 11</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
</tr>
<tr>
<td>addr 12</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
</tr>
<tr>
<td>addr 13</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
</tr>
<tr>
<td>addr 14</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
</tr>
<tr>
<td>addr 15</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
</tr>
<tr>
<td>addr 16</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
<td>count number of continuous 0</td>
<td>compressed data</td>
</tr>
</tbody>
</table>

Decompressed data

OR

BISR chain
BISR controller sharing
BISR controller sharing

- For area overhead reducing, the BISR controller can be placed on the master die and be shared with the slave dies, and the slave dies only need to contain the BISR chains.
- The LFSR_CHK on the receiver side generates the BISR data (marked in red), and the BISR data should be OR operated by bit before shifting into the BISR controller.
- When the chips power on, the merged BISR data will be shifted into the BISR chain of the two dies simultaneously.
- The BISR controller is shared by dividing the EFUSE storage space when there are multiple interconnect interfaces and BISR chains.
Conclusion

• A BIST structure of interconnect interface is proposed, which supports internal die loopback test and inter-die interconnection test.

• The IO repair is processed by the BISR chain and BISR controller circuits automatically, which reduces the programming complexity of the test machine and supports the test based on the repaired state.

• The BISR data compression technology is introduced to reduce the storage space requirement of repair data.

• A solution for multiple repairs to adapt to the added faults caused by the multiple stacking process of 3DIC is proposed.

• A solution for multiple dies to share one BISR controller is proposed, which reduces the area overhead.
3DC-TEST

Thank You
Process-Resilient Fault and Error Tolerant DLL for Supporting Multi-Die Clock Synchronization

Jun-Yu Yang Shi-Yu Huang
Electrical Engineering Department, National Tsing Hua University, Taiwan

Abstract—A Delay-Locked Loop (DLL) circuit is often useful for the clock synchronization in a chip incorporating multiple functional dies. Previously we have shown in [14] that Triple-Module Redundancy (TMR) technique cannot work alone to provide fault and error tolerance for a DLL design unless it is enhanced by a timing correction scheme to nullify the adversary effect caused by the voter circuit’s delay. However, even such a timing correction scheme still has a serious drawback - it cannot track the process variation automatically. Therefore, in this work, we propose a process resilient fault and error tolerant DLL architecture featuring a new “dynamic timing correction scheme”. The process variation or even temperature effects can now be dealt with during runtime operations and thereby significantly improved performance can be achieved in terms of the overall phase error between the DLL’s input and output clock signals. Post-layout simulation under 5 process corners indicates that the worst-case phase error can be further slashed from 20ps down to only 10ps, or 50% improvement.

Index Terms — Delay-Locked Loop, Fault Tolerant, Soft Error Tolerant, Dynamic Timing Correction Scheme, Process Resilient

I. INTRODUCTION

A modern high-end IC could consist of a number of functional dies integrated on a silicon substrate. If the functional dies are designed and fabricated with different process technologies, then heterogeneous integration is needed. The clock subsystem on such an IC is more involved. In addition to the clock distribution network in each functional die, other complementary timing circuits such as Phase-Locked Loop (PLL) and Delay-Locked Loop (DLL) are often used to perform clock synchronization [1][2][3][4]. Functionally speaking, the PLL can be used to convert a low- to medium-frequency input clock signal (e.g., 10MHz) into a high-frequency on-chip clock signal (e.g., 1GHz). On the other hand, the DLL macros can be used to compensate for the clock latency differences in the different functional dies.

To achieve a robust IC with a very high reliability, various Fault and Soft Error Tolerance schemes (referred to as FET schemes in the sequel) are needed, to provide complete protection for the entire IC. Even though the FET schemes for the memory and the logic blocks are very well studied in the literature [5][6][7], the FET schemes for the clock subsystem are less developed and addressed. In some sense, the clock subsystem (including the PLLs, the DLLs, and the clock distribution networks) spans the entire silicon area in an IC. Without proper FET schemes, they could become the Achilles’ heel of an IC and cause some unexpected reliability problems.

Furthermore, the clock subsystem are delicate timing circuits and therefore particularly vulnerable to the environmental noises and interferences. Not only a timing glitch is detrimental, a radioactive bombardment of high-energy particles could be problematic as well. For example, a noise causing a timing drift of 100ps on the 1GHz high-speed clock signal will imply a loss of a 10% useful clock cycle time for a 1GHz clock signal, and thus, such a timing drift is likely to cause a functional failure already in the computational logic driven by the clock signal.

The rest of this paper is organized as follows. In Section II, we review the basics of a Delay-Locked Loop (DLL). In Section III, we first study a naïve FET-DLL design using the traditional Triple-Module Redundancy (TMR) [8]. We then point out a so-called “output-lagging” problem, which will render this naïve design useless. Next, we remedy the problem by introducing a timing correction scheme. In Section IV, we present the post-layout simulation results of our FET-DLL design using a 90nm CMOS process to demonstrate the significantly reduced Maximum Phase Error. In Section V, we conclude. The extended results will be report in the near future.

II. PRELIMINARIES

In this Section, we review the basic functionality of a DLL in the context of a multi-clock synchronization setting.

2.1 Basics of a DLL

Nowadays, a DLL can be easily made by all-digital cell-based circuits [9][10][11][12][13]. As shown in Fig. 2, a basic DLL consists of three major blocks: (1) a Tunable Delay Line (TDL), (2) a Phase Detector (PD), and (3) an overall controller. In addition, there be an extra component referred to as “Delay Under Compensation”. This component corresponds to clock latency across the clock delivery network.

The TDL is simply a delay line with the delay tunable by a control code issued by the controller. The PD is a simple circuit that compares the “phases” of two clock signals (i.e., the timing positions of the rising edges of two clock signals such as clk_out and clk_ref) and determine which of them is leading in time by producing a binary so-called lead/lag signal. In this figure, clk_ref is treated as the reference. The output lead/lag signal is ‘1’ if the other input clock signal, i.e., clk_out (denoting the output clock signal of the DLL), is leading in the phase.

* This work was sponsored in part by Ministry of Science and Technology (MOST) of Taiwan under research grants MOST-108-2218-E-007-062. We also acknowledge the help of Taiwan Semiconductor Research Institute (TSRI) for providing the access to the EDA tools.
Otherwise, the output *lead/lag* signal is ‘0’ (to indicate the *clk_out* is lagging *clk_ref*).

The entire operation of a DLL is often divided into two stages—i.e., phase-locking and phase-tracking. Initially, the output clock signal (namely *clk_out*) is not in-phase with the input clock signal (namely *clk_ref*). In the phase-locking stage, a phase-locked condition is achieved in a sense that the phase difference between *clk_out* and *clk_ref* has been reduced to a very small amount, as shown in the figure. After that, the DLL will proceed to the phase-tracking stage, in which the phase-locked condition is further maintained by constant tuning of the delay across the Tunable Delay Line (TDL) when necessary.

2.2 Basic FET-DLL Architecture

Fig. 2 shows a naive Fault and Error Tolerant DLL Architecture incorporating the TMR. It does not work very because of an output-lagging problem in a sense that the output clock signal may lag the input clock signal by more than 100ps when in the locked state due to the delay across the voter circuit.

To fix this problem, we have proposed a timing correction scheme as shown in Fig. 3 [14]. In a nutshell, we have inserted 3 so-called dummy voter circuits, namely, \{V1, V2, V3\}, each for one of the 3 primitive DLL instances, respectively. Each of them sits on the feedback path from a signal out of \{\phi 1, \phi 2, \phi 3\} back to the input of a Phase Detector. We denote the output signals of the 3 dummy voter circuits as \{fb1, fb2, fb3\}, respectively. Note that these 3 feedback signals are key signals in our design and we will mention them frequently in the rest of the paper.

Why the output-lagging problem can be solved nicely by this simple timing correction scheme is illustrated in Fig. 4. On the left-hand-side, we show the timing relationships among some key signals before the correction, including *clk_ref*, \{\phi 1, \phi 2, \phi 3\}, and *clk_out*. Suppose that the delay of the voter circuit and the dummy voter circuits are assumed to be the same for the time being and roughly denoted as \(\delta_{\text{voter}}\). We can see that before the correction, \{\phi 1, \phi 2, \phi 3\} are in-phase with *clk_ref* (due to the locking of the 3 primitive DLL instances), and *clk_out* is lagging ensemble \{\phi 1, \phi 2, \phi 3\} as well as *clk_ref* by \(\delta_{\text{voter}}\) (due to the voter circuit’s delay).

On the right-hand-side hand, we show the waveforms of key signals after the correction. Signal *clk_out* and \{fb1, fb2, fb3\} are all at the downstream of \{\phi 1, \phi 2, \phi 3\} by a voter circuit’s delay and so they are similarly lagging \{\phi 1, \phi 2, \phi 3\} by \(\delta_{\text{voter}}\) and thus *clk_out* is now in-phase with \{fb1, fb2, fb3\}. Since ensemble \{fb1, fb2, fb3\} are in-phase with *clk_ref* due to the locking operations of DLL-1, DLL-2, DLL-3, we therefore conclude by transitivity that *clk_out* is not only in-phase with \{fb1, fb2, fb3\}, but also in-phase with *clk_ref*. It can be seen
from the illustration that ensemble \{(\phi_1, \phi_2, \phi_3)\} is now ahead of \texttt{clk} \_\texttt{ref} by a proper amount of \(\delta_{\text{voter}}\), to make \texttt{clk} \_\texttt{out} properly in-phase with \texttt{clk} \_\texttt{ref}.

### 2.3 Performance of the Basic FET-DLL Design

Fig. 5 shows the layout of the basic FET-DLL design. The area of the whole design is \((370 \mu\text{m} \times 193 \mu\text{m}) = 0.072 \mu\text{m}^2\).

Table 1 shows the performances of 3 versions of DLL design, (1) Our primitive DLL without Fault and Error Tolerance, (2) A naïve TMR-DLL, and (3) A basic FET-DLL, in terms of the phase errors reported by post-layout simulation under the TT process corner. It shows that the phase error ranges are \([-10\text{ps}, -4\text{ps}]\), \([113\text{ps}, 117\text{ps}]\), \([12\text{ps}, 17\text{ps}]\), and the maximum phase errors are 10ps, 117ps, and 17ps, respectively. It indicates that the timing correction scheme does have help reduce the maximum phase error significantly from 117ps to 17ps.

This new architecture with the new dynamic timing correction scheme requires its own online calibration procedure to determine the control codes of the TDEs dynamically in the runtime. In this first report, the details are omitted for simplicity.

### IV. Experimental Results

The layout of the enhanced architecture is shown in Fig. 7. It area is \((430 \mu\text{m} \times 230 \mu\text{m}) = 0.999 \mu\text{m}^2\), or 37% increases than that of the basic architecture. We have performed post-layout simulation to extract the waveforms of certain key signals to verify its operation. The results of one such experiments are shown in Fig. 8. As expected, the DLL goes through the locking and the calibration processes before the normal operation. After calibration, all signals in \{\texttt{clk} \_\texttt{out}, \texttt{clk} \_\texttt{in}, v1, v2, v3\} are all aligned in their phases.

### III. Proposed Process Resilient FET-DLL

The above basic FET-DLL still has one limitation – it uses static technique to compensate for the extra delay across the voter circuit and thus cannot adapt to the process variation. In other words, the difference of the voter circuit’s delay and its three dummy circuits could translate into the extra phase error observed at the output clock signal of the DLL. To fix this problem, we propose an enhanced FET-DLL architecture incorporating a dynamic timing correction scheme, as shown in Fig. 6. Here, a number of Tunable Delay Elements (TDEs) have been inserted into the FET-DLL so that the above targeted delay mismatches can be dynamically tracked down and resolved.
Calibration. The

FS -13.44 y variati

FS -o. 5, pp.751 -

2.59. For a DLL, we are most concerned about the

16 Self

SF 10 rror Tolerance. In some

- 

[6, 11]-

[9, 18]-

5.97

3.82 9.47 r tolerant

20

TT dynamic compensation under different corners

SS

Max. Phase Error

Avg. Phase Error

Process Corners

Max. Phase Error (ps)

Avg. Phase Error (ps)

(b) clk_in, clk_out, v1, v2, v3 are all in-phase after calibration.

Fig. 8. Simulation waveforms of the new DLL.

The benefits of this enhanced architecture are summarized in Table 2. For a DLL, we are most concerned about the maximum phase error over a number of clock cycles (e.g., 1000) after it is locked. In the table, we also showed the phase range, and the average phase error during the observation window of time. It can be seen from Table 2 that the maximum phase error can be reduced from 18ps to 5ps during the TT process corner. If taking all 5 process corners into consideration, the basic DLL exhibit a worst-case maximum phase error at the FS process corner with an amount of 20ps. On the other hand, the enhanced DLL produces 10ps as the maximum phase error under 5 process corners. In the worst case, the overall maximum phase error at the FS process corner. If taking all 5 process corners into consideration, the maximum phase error during the observation window of time. It can be seen from Table 2 that the maximum phase error can be reduced from 18ps to 5ps during the TT process corner. If taking all 5 process corners into consideration, the basic DLL exhibit a worst-case maximum phase error at the FS process corner with an amount of 20ps. On the other hand, the enhanced DLL produces 10ps as the maximum phase error at the FF process corner. In the worst case, our enhanced version still enjoys a reduction of (20ps-10ps)/20ps = 50%.

Table 2. Phase Errors of enhanced DLLs based on post-layout simulation using a 90nm CMOS process.

<table>
<thead>
<tr>
<th>Process Corners</th>
<th>FF</th>
<th>SF</th>
<th>FS</th>
<th>SS</th>
<th>TT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase Error Interval (ps)</td>
<td>[11, 16]</td>
<td>[6, 11]</td>
<td>[5, 15]</td>
<td>[9, 14]</td>
<td></td>
</tr>
<tr>
<td>Max. Phase Error (ps)</td>
<td>16</td>
<td>11</td>
<td>15</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Avg. Phase Error (ps)</td>
<td>13.64</td>
<td>7.97</td>
<td>17.57</td>
<td>5.47</td>
<td>13.58</td>
</tr>
</tbody>
</table>

Table 2. Phase Errors of enhanced DLLs based on post-layout simulation using a 90nm CMOS process.

<table>
<thead>
<tr>
<th>Process Corners</th>
<th>FF</th>
<th>SF</th>
<th>FS</th>
<th>SS</th>
<th>TT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase Error Interval (ps)</td>
<td>[11, 16]</td>
<td>[6, 11]</td>
<td>[5, 15]</td>
<td>[9, 14]</td>
<td></td>
</tr>
<tr>
<td>Max. Phase Error (ps)</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg. Phase Error (ps)</td>
<td>13.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

V. CONCLUSION

In this paper, we have proposed an enhancement technique to make our previous Fault and Error Tolerant DLL (FET-DLL) even more robust under process variation and/or temperature effects. In this enhanced FET-DLL architecture, a dynamic timing correction scheme is incorporated, by which our new DLL can track and fix the delay variation of the three inserted dummy voters more easily during runtime operations. The benefits have been demonstrated by the post-layout simulations under 5 process corners. In the worst case, the overall maximum phase error of the DLL can be reduced from 20ps down to 10ps, or relatively a 50% improvement. From another aspect, this new result of 10ps is equivalent to the reference performance of a primitive DLL without Fault and Error Tolerance. In some sense, we have achieved an ideal condition that the performance degradation potentially induced by our fault and error tolerant architecture has been completely averted.

REFERENCES

Process Resilient Fault and Error Tolerant DLL for Supporting Multi-Die Clock Synchronization

Chun-Yu Yang  
Shi-Yu Huang (Speaker)  
EE Dept., National Tsing Hua University, Taiwan

Nov. 6, 2020
Outline

◆ Introduction
  - Problem, Background, Objective
◆ Fault and Error Tolerant (FET) DLL
◆ Experimental Results
◆ Concluding Remarks
Clock Distribution Problem
in a Heterogeneously Integrated Multi-Die IC

Objective: All Flip-Flops in every die receive the clock signal at the same time

- $S_0$: Chip-Level Clock Source Point
- $\{S_1, S_2, S_3, S_4\}$: Die-Level Clock Source Points
- $\cdots$: Clock Relays in the underlying interposer or RDL layer
- PLL: Phase-Locked Loop (for clock frequency multiplication)
Challenges

Different dies ➔ Different Within-Die Clock Latencies ➔ Flip-Flops (FFs) receive the clock signal with skews
DLL-Assisted Clock Synchronization in a Heterogeneously Integrated Multi-Die IC

Within-Die Latency

400ps
600ps
FF
FF
FF
FF
FF

300ps
700ps
FF
FF
FF
FF
FF

700ps
FF
FF
FF
FF
FF

300ps
300ps
500ps
FF
FF
FF
FF
FF

500ps
FF
FF
FF
FF
FF

Inter-Die Latency

S1 → DLL-1

S2 → DLL-2

S3 → DLL-3

S4 → DLL-4

Chip-Level Clock Source Point

{S1, S2, S3, S4} are Die-Level Clock Source Points

The die-to-die Clock Skew can be minimized by the inserted DLLs. (The width of a DLL box denotes its input-to-output delay)
Delay-Locked Loop (DLL) 
(can be made by only standard cells)

Operation of a DLL: Tune the delay line until a locking condition is reached

\[ \delta \text{ is multiple times the clock period} \]
**Performance We Care**

- **Maximum Phase Error of a DLL**

**Definition of maximum phase error:**
The worst-case phase error amount between clock_in and clock_out over a time frame (e.g., 1000 cycles) after the DLL is locked!

---

**Diagram:**

- **Bad:**
  - clock_in: 100ps
  - clock_out: 100ps
- **Good:**
  - clock_in: 10ps
  - clock_out: 10ps
Objective of This Work

Objective of this work:
Convert a given primitive Delay-Locked Loop Design, into a Fault and Soft Error Tolerant Architecture using Modified Triple-Module-Redundancy (TMR)

Outline

- Introduction
  - Problem, Background, Objective
- Fault and Error Tolerant (FET) DLL
- Experimental Results
- Concluding Remarks
Naïve TMR DLL

\[ \text{clk\_out} = \phi_1\phi_2 + \phi_2\phi_3 + \phi_3\phi_1 \]

\{\phi_1, \phi_2, \phi_3\} are all locked to \text{clk\_ref} 

\[ \Rightarrow \text{Yetclk\_out lags} \text{clk\_ref} \]
Output-Lagging Problem

Output Lagging Problem:
\( \Rightarrow \text{clk\_out lags clk\_ref} \) by a voter delay (~100ps)
Fault and Soft-Error Tolerant DLL
- using Static Timing Correction Scheme (ATS’20)

\{\phi_1, \phi_2, \phi_3\} are \textit{“one voter delay”} ahead of \textit{clk\_ref}
\implies \textit{clk\_out} is now \textit{in-phase} with \{\textit{fb1, fb2, fb3}\}, and also \textit{in-phase} with \textit{clk\_ref}
Outline

◆ Introduction
  - Problem, Background, Objective
◆ Fault and Error Tolerant (FET) DLL
  - (ATS’20) Architecture with Static Timing Correction
  - (This Work) Architecture with Dynamic Calibration
◆ Experimental Results
◆ Concluding Remarks
Process Variation Issue

Ideally, we have wished that
\[ \text{Delays of \{V1, V2, V3\} = Delay of the “Output Voter”} \]

But in reality,
\[ \text{there could be mismatches due to process variation.} \]
Architecture of a FET-DLL with Dynamic Calibration

- **clk_ref**
  - DLL-1
    - PD → Con
    - TDL
    - Delay-Locked Loop (DLL-1)
    - φ1
  - DLL-2
    - PD → Con
    - TDL
    - Delay-Locked Loop (DLL-2)
    - φ2
  - DLL-3
    - PD → Con
    - TDL
    - Delay-Locked Loop (DLL-3)
    - φ3

- Clk_out
  - VOTER
  - ω1
  - ω2
  - ω3

A TDE (Tunable Delay Element)
Calibration For One DLL At a Time
(The Primitive DLLs Take Turn to Calibrate)

**During Calibration**
- Tune $\omega_1$ until $\text{clk}_\text{out}$ locks to $\text{clk}_\text{ref}$
- $D(\phi_1 \rightarrow \text{clk}_\text{out}) = D(\phi_1 \rightarrow \text{fb}_1)$

**V1**
- $\text{clk}_\text{ref}$
- $\text{PD}$
- $\text{Con}$
- $\text{TDL}$
- $\phi_1$
- $\omega_1$

**V2**
- $\text{clk}_\text{ref}$
- $\text{PD}$
- $\text{Con}$
- $\text{TDL}$
- $\phi_2$
- $\omega_2$

**V3**
- $\text{clk}_\text{ref}$
- $\text{PD}$
- $\text{Con}$
- $\text{TDL}$
- $\phi_3$
- $\omega_3$

**Clk_out**
Outline

◆ Introduction
  - Problem, Background, Objective
◆ Fault and Error Tolerant (FET) DLL
◆ Experimental Results
◆ Concluding Remarks
Layout of FET-DLL (using a 90nm CMOS Process) - before and after Adding Dynamic Calibration

37% Area Overhead
Post-Layout Simulation Scenario: when there is a Short-Pulse Error at φ1

Output *clk_out* is not affected!
# Max. Phase Error Comparison

<table>
<thead>
<tr>
<th>DLL Version</th>
<th>Maximum Phase Error (ps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primitive DLL (not fault/error tolerant)</td>
<td>~10 ps (TT corner)</td>
</tr>
<tr>
<td>FET-DLL with Naïve TMR</td>
<td>~117 ps (TT corner)</td>
</tr>
<tr>
<td><strong>FET-DLL with Static Timing Correction</strong></td>
<td><strong>20 ps (5 Corners)</strong></td>
</tr>
<tr>
<td>FET-DLL with Dynamic Calibration</td>
<td>10 ps (5 Corners)</td>
</tr>
</tbody>
</table>
Concluding Remarks

The DLL is indispensable for a synchronous Chiplet Design. Yet, it was not clear previously how to make a DLL fault and soft-error tolerant.

Our research has shown that this is indeed achievable:
- **TMR** is applicable, but not directly
- We need **architecture** that incorporates not just “static timing correction”, but also “dynamic calibration” to make it robust & resilient
Thank You!
The IEEE standard for test integration of stacked ICs has been released and is rapidly being leveraged for test access in 3D stacked die and even 2.5D package applications. IEEE Std 1838™-2019, “IEEE Standard for Test Access Architecture for Three-Dimensional Stacked Integrated Circuits”, enables EDA tools to be automated to insert the standard so that designers can focus on functional innovation while having confidence in the fact that test access is available to help prove product quality. IEEE 1838 provides a mandatory serial data path and an optional parallel data path for test pattern application. The serial data path is implemented as a branch of IEEE 1149.1 using optional concepts further fleshed out in IEEE 1687.

This presentation will also update you on the architecture of IEEE Std 1838, provide some interesting development history, and explore some automation to help implement it in your next multi-die package project. The standard was not always heading in the direction that it ended up. This presentation will provide a background and some historical context to bring it to the point at which the published standard is now. For example, the initial direction that IEEE 1838 was heading involved using IEEE 1500 as the serial path interface between dies. It now uses IEEE 1149.1 and the SIB (Segment Insertion Bit) concept from IEEE 1687 to form that bit of the serial infrastructure. It will also lightly delve into some electronic design automation tools that will help implement the architecture for each die, and validate the implementation at the stack level.
IEEE Std 1838 Introduction and the Move from a 1500-Centric TAM
Serial and Parallel Port Introduction and Discussion

Adam Cron, Principal Engineer
3DC-TEST 2020
Package Possibilities

- 2.5D, 3D, 5.5D: many topologies
- Digital + memory is most prevalent
- Heterogeneous technology nodes
- Higher bandwidth
- Smaller footprint
- Lower power

Test Context

- Scan
- Compression
- Logic BIST
- Memory BIST
- Interconnect Test
- Interconnect Test
- Logic BIST
- Memory BIST
- Scan
- Compression

Functional (DFT) logic
Access Issues and Solutions

- TSVs not accessible
- Contacts too small and close together
- Contacts disappear

Die-Level Pre-Bond Test  Partial Stack Mid-Bond Test  Complete Stack Post-Bond Test
Access Issues and Solutions
Serial Port

Mandatory
Each Die Has a Primary TAP
Serial Access Up and Down

- Primary: PTAP1, WSP1
- Secondary: PTAP2, WSP2
- Primary: PTAP3, WSP3
Multi-Tower Support
Serial Access to Die Wrapper Register (DWR)
IEEE Std 1838 Serial Path Detail
3DCR Register
IEEE Std 1838 Serial Path – This Die
IEEE Std 1838 Serial Path – That Die

TDI  TMS  TCK  TRSTN  TDI

3DCR  FPPCONFIG  DR  BYPASS

DECODE  IR

PTAP
Die Wrapper Register
Pre-DFT Functional Path
Typical Dedicated Wrapper Cell at Boundary
Ignoring the “Analog” Component
Typical Shared Wrapper Cell
Inland Wrapping

Shore-Level Components
Constrained Inland Wrapping
Delay Test with No Inland Wrapping
Delay Test Penalty
Flexible Parallel Port (FPP)

Optional
Parallel Access
Flexible Parallel Port Lane Concept

- **Lane**
  - FPP_SEC
  - FPP_FROM_SIDE
  - FPP_FROM_CORE
  - FPP_TO_SIDE
  - FPP_TO_CORE
  - FPP_CLK_IN
  - FPP_CLK_OUT
  - FPP_PRI

- **Primary Port**
- **Secondary Port**

**die**

© 2020 Synopsys, Inc.
Design Specifications

- Flexible Parallel Port (FPP) control registration (driving signals labelled “Control”, below) can be distributed
Registered Lane, Up Only

Lane

die

FPP_TO_CORE

FPP_CLK_IN

FPP_PRI

FPP_SEC

FPP_SEC_EN

FPP_REGN_BYP

FPP_REGPU_BYP
Non-Registered Lane, Down Only

Lane

FPP_SEC

FPP_FROM_CORE

die

FPP_CORE_SEL

1 0

FPP_PRI_EN

FPP_PRI

FPP_PRI_EN
FPP Application

Scan Chains

Test Modes

Functional (DFT) logic
FPP Broadcast Application

Functional (DFT) logic
1500 vs. 1149.1
### 1149.1-Based vs. 1500-Based Solutions

<table>
<thead>
<tr>
<th></th>
<th>1149.1</th>
<th>1500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pins Required</td>
<td>5</td>
<td>5/bottom die, 8/other die</td>
</tr>
<tr>
<td>Timing closure</td>
<td>1 signals: TMS → TCK</td>
<td>4 signals: SelectWIR, CaptureWR, ShiftWR, UpdateWR → WRCK</td>
</tr>
<tr>
<td>Area overhead</td>
<td>TAP controller + 5 TSVs on each die</td>
<td>8 TSVs on each die</td>
</tr>
<tr>
<td>Test pattern reuse</td>
<td>Yes, reuse die level tests at stack and package level</td>
<td>More complicated pattern reuse for arbitrary state transitions</td>
</tr>
</tbody>
</table>
TAP Access Per Die

TAP Access Per Die

TAP1

TAP2

TAP3

TAP4

TDI TMS TCK TRST TDO

WSO SelectWIR WRCK WRSTN CaptureWR ShiftWR UpdateWR WSI
TMS Load to TCK
1500 Access Per Die
1500 Access Per Die
1500 Signal Paths

1500 Wrapped Core

- Optional User Defined Wrapper Parallel Port (WPP)
- Functional Inputs
- Wrapper Boundary Register
- Wrapper Bypass Register
- Wrapper Serial Input (WSI)
- Core FF
- Functional Outputs
- Wrapper Boundary Register
- Wrapper Instruction Register
- Wrapper Serial Output (WSO)
- Wrapper Serial Control (WSC)
- Mandatory Wrapper Serial Port (WSP)
Stack with 1500
Stack Under Discussion

TDI  TMS  TCK  TRST*  TDO

1  TAP1

2

3  TAP3

4  TAP4
TAP on Every Die
Memory with 1500 on TOP
Quick Evolution

= 1500 bus
Die-Level Cores Accessed with 1500
Stacked and Electrically Connected
1149.1 Controller Drives WSP
## 1149.1-Based vs. 1500-Based Solutions

<table>
<thead>
<tr>
<th></th>
<th>1149.1</th>
<th>1500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pins Required</td>
<td>5</td>
<td>5/bottom die, 8/other die</td>
</tr>
<tr>
<td>Timing closure</td>
<td>1 signals: TMS $\rightarrow$ TCK</td>
<td>4 signals: SelectWIR, CaptureWR, ShiftWR, UpdateWR $\rightarrow$ WRCK</td>
</tr>
<tr>
<td>Area overhead</td>
<td>TAP controller + 5 TSVs on each die</td>
<td>8 TSVs on each die</td>
</tr>
<tr>
<td>Test pattern reuse</td>
<td>Yes, reuse die level tests at stack and package level</td>
<td>More complicated pattern reuse for arbitrary state transitions</td>
</tr>
</tbody>
</table>
Implementation
Stack-Level Validation

Framework for any type of multi-die design
DFT Connections to Serial and Parallel 1838

1149.1/1838 TAP

- FPPCONFIG
- 3DCR

IP

PTAP

STAP

3DCR
Thank You

Please address questions to a.cron@ieee.org
Leveraging Lessons-Learned on 2D-SOCs in Designing Parallel TAMs for 3D-SICs Based on IEEE Std 1838’s Flexible Parallel Port

Erik Jan Marinissen
imec
Kapeldreef 75
3001 Leuven
Belgium
erik.jan.marinissen@imec.be

Abstract

3D-stacked ICs allow to keep the momentum of Moore’s Law going by exploiting the vertical dimension for further integration [1-3]. Just like all integrated circuits, 3D-SICs need to be tested for manufacturing defects [4]. This test should be a modular test, in which dies and interconnects between adjacent dies can be tested as stand-alone units. Test access from and to test equipment is via the stack’s external interface, which is typically located in the bottom die of the stack. Dies further up in the stack depend on appropriate 3D-DfT in the dies below them to pass their test stimuli up into the stack and their test responses vice versa. IEEE Std 1838™-2019 [5-7] standardizes 3D-DfT required in a single die, such that dies compliant to this standard constitute, once stacked, a consistent stack-level 3D test access architecture.

IEEE Std 1838™-2019 provides a single-bit (‘serial’) test access mechanism (TAM) to load instructions as well as test data into the stack through the well-known Test Access Port (TAP) of IEEE Std 1149.1. To accommodate transportation of large amounts of test data up and down the stack in a timely fashion, IEEE Std 1838™-2019 provisions a multi-bit (‘parallel’) TAM, named ‘Flexible Parallel Port’ (FPP) [8]. As test access requirements might vary widely between different 3D-SICs, the FPP is, as its name indicates, indeed flexible, viz. optional and scalable in both connectivity, functions, and bandwidth. IEEE Std 1838™-2019 only provides an FPP design template, from which die makers and stack integrator jointly need to decide what will be implemented in a specific 3D-SIC.

For modular, core-based 2D-SOCs, a large body of published work is available on optimizing parallel TAM architectures, such that their corresponding feasible test schedules result in minimal test data storage requirements on test vector memories of the test equipment and test application time [9–17]. In this presentation, we will leverage the ‘lessons-learned’ [18] from optimizing TAM architectures for 2D-SOCs in the context of 3D-SICs.
References


1. INTRODUCTION

Flexible Parallel Port Overview

- IEEE Std 1838’s Mandatory ‘Serial’ Test Access
  - Via TAP’s TDI and TDO
- IEEE Std 1838’s Motivation for FPP
  - Often, there is a need for more test bandwidth: “time is money”!
  - No ‘one-size-fits-all’ solution: FPP is optional and scalable in many parameters
- FPP Requirements
  - Transportation up/down the stack of test data, test control, and test clocks
  - Highly configurable: width, source/destination, direction, registration
  - Configuration possible (1) at design time and (2) at test time
  - FPP is a template that covers common, advanced, and exotic test scenarios
Leveraging Lessons-Learned on 2D-SOCs in Designing Parallel TAMs for 3D-SICs Based on IEEE Std 1838’s Flexible Parallel Port

Erik Jan Marinissen

Presentation Outline

1. Introduction
2. IEEE Std 1838™-2019 Compared to Other IEEE DfT Wrapper Standards
3. Flexible Parallel Port (FPP)
4. PTAM Design in 2D-SOCs
5. Lessons Learned on 2D-SOCs
6. Mapping IEEE Std 1838™ FPP on “My” 3D-Dft Architecture
7. Conclusion

2. IEEE Std 1838™-2019 Compared to Other IEEE DfT Wrapper Standards
Leveraging Lessons-Learned on 2D-SOCs in Designing Parallel TAMs for 3D-SICs Based on IEEE Std 1838’s Flexible Parallel Port

Erik Jan Marinissen

2. IEEE STD 1838™-2019 COMPARED TO OTHER IEEE DFT WRAPPER STANDARDS

IEEE P1838 Working Group

Several members of the IEEE P1838 Working Group assembled at the IEEE International Test Conference (ITC) in November 2019 in Washington, DC, USA

What is IEEE Std 1838™-2019?

- IEEE Std 1838 specifies mandatory and optional* 3D-DFT features per die:
  1. DWR : Die Wrapper Register
  2. SCM : Serial Control Mechanism
  3. FPP* : Flexible Parallel Port
- Such that
  - 3D-SIC can be modularly tested: dies and interconnect separate
  - Compliant dies in a 3D-SIC form a consistent test access architecture
  - Connect various modules-under-test to external test equipment

[Marinissen et al. – ETS 2016]
[Li et al. - ETS 2018]
Leveraging Lessons-Learned on 2D-SOCs in Designing Parallel TAMs for 3D-SICs Based on IEEE Std 1838’s Flexible Parallel Port

Erik Jan Marinissen

2. IEEE Std 1838™-2019 COMPARED TO OTHER IEEE DfT WRAPPER STANDARDS

IEEE Wrapper DfT Standards

IEEE Std 1149.1™: Test Board Assembly & Interc-ct

- Test focus: assembly of chips on printed circuit board: interconnect ⇒ EXTEST
- Test access: via serial (= one-bit) path: TDI-TDO
- No parallel access: not required for interconnect test and extra package pins very expensive
Leveraging Lessons-Learned on 2D-SOCs in Designing Parallel TAMs for 3D-SICs Based on IEEE Std 1838’s Flexible Parallel Port

Erik Jan Marinissen

2. IEEE Std 1838™-2019 COMPARED TO OTHER IEEE DFT WRAPPER STANDARDS

IEEE Std 1500™: Test of SOC with Embedded Cores

- Test focus: cores (⇒INTEST) and interconnects/logic (⇒EXTEST)
- Test access: via serial (= one-bit) path + via parallel (= multi-bit) path
- Parallel port WPI-WPO defined, but IEEE Std 1500 does not mandate much

IEEE Std 1838™: Test of 3D-Stacked Dies

- Test focus: interconnects (⇒EXTEST) and dies (⇒INTEST)
- Test access: via serial (= one-bit) path + via parallel (= multi-bit) path + loopback
- IEEE P1838 Working Group has spent much more effort to provide a template for the FPP lanes
2. IEEE Std 1838™-2019 COMPARED TO OTHER IEEE DfT WRAPPER STANDARDS

IEEE Std 1838™: Test of 3D-Stacked Dies

- Test focus: interconnects + 'shore' logic (⇒ EXTEST) and dies (⇒ INTEST)
- Test access: via serial (= one-bit) path + via parallel (= multi-bit) path + loopback

IEEE P1838 Working Group has spent much more effort to provide a template for the FPP lanes.

2. IEEE STD 1838™-2019 COMPARED TO OTHER IEEE DFT WRAPPER STANDARDS

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Modules</td>
<td>Packaged chips</td>
<td>Embedded cores</td>
<td>Stacked dies</td>
</tr>
<tr>
<td>System</td>
<td>Printed Circuit Board (PCB)</td>
<td>System-on-Chip (SOC)</td>
<td>3D-IC</td>
</tr>
<tr>
<td>Focus InTest</td>
<td>not applicable</td>
<td>Core-internal circuitry</td>
<td>Die-internal circuitry</td>
</tr>
<tr>
<td>ExTest</td>
<td>Board-level interconnect</td>
<td>Top-level circuitry</td>
<td>Die interconnects + 'shore' logic</td>
</tr>
<tr>
<td>I/O type</td>
<td>Package pins (= expensive!)</td>
<td>Core terminals (= cheap!)</td>
<td>Micro-bumps (= medium-priced)</td>
</tr>
<tr>
<td>Data Serial</td>
<td>1-bit: TDI-TDO</td>
<td>1-bit: WSI-WSO</td>
<td>1-bit: TDI-TDO</td>
</tr>
<tr>
<td>Parallel</td>
<td>not applicable</td>
<td>n-bit: WFI-WPO[n]</td>
<td>require loopback at every stack tier</td>
</tr>
<tr>
<td>Test Control</td>
<td>TCK, TMS, (TRSTn*)</td>
<td>WRCK, WSTn, SelectWIR, ShiftWR, CaptureWR, UpdateWR</td>
<td>TCK, TMS, TRSTn</td>
</tr>
<tr>
<td></td>
<td>+ 16-state Finite State Machine</td>
<td></td>
<td>16-state Finite State Machine</td>
</tr>
</tbody>
</table>

* = optional
Leveraging Lessons-Learned on 2D-SOCs in Designing Parallel TAMs for 3D-SICs Based on IEEE Std 1838’s Flexible Parallel Port

Erik Jan Marinissen

3. Flexible Parallel Port (FPP)

3. FLEXIBLE PARALLEL PORT (FPP)

IEEE Std 1838 Stack Model and Terminology

- **1838 avoids terms** based on relative physical position: ‘front’, ‘back’, ‘top’, ‘down’
  - Instead: use die order as seen by external test equipment
  - Assumption: all external I/Os are concentrated in one die
- **Die Identification**: numbering from die with external I/O → ‘first’, ‘last’ die
- **Primary Interface**: signals to the previous die (or external stack interface) = 1
- **Secondary Interface**: signals to a next die ≥ 0

[Marinissen et al. – ETS’16]
Leveraging Lessons-Learned on 2D-SOCs in Designing Parallel TAMs for 3D-SICs Based on IEEE Std 1838’s Flexible Parallel Port

Erik Jan Marinissen

3. FLEXIBLE PARALLEL PORT (FPP)

FPP Lane: One-Bit Test Data Transportation Hub

Lane Sources and Destinations

- Other die: Previous: FPP_PRI bi-directional
  Next: FPP_SEC bi-directional
- This die: Other lane: FPP_FROM_SIDE, FPP_TO_SIDE uni-directional
  Functional logic: FPP_FROM_CORE, FPP_TO_CORE uni-directional

FPP Lane Interconnectivity: Paths

- Every source can be connected to every destination
- Max. 4x4 – 2 = 14 paths
  - No path from bi-dir terminal to itself
  - Typical: << 14 paths

<table>
<thead>
<tr>
<th>from to</th>
<th>PRI</th>
<th>SEC</th>
<th>SIDE</th>
<th>CORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRI</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SEC</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SIDE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CORE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>


Leveraging Lessons-Learned on 2D-SOCs in Designing Parallel TAMs for 3D-SICs Based on IEEE Std 1838’s Flexible Parallel Port

Erik Jan Marinissen

3. FLEXIBLE PARALLEL PORT (FPP)

**FPP Path: Options**

- Destination with multiple Sources → Source Select – controlled from FPP Config
- Each path must have a Hold Element (neg-edge latch/FF) at its Destination
  - Optionally equipped with Hold Bypass – controlled from FPP Config
- Each path has zero or more pipeline registers (on pos-edge or neg-edge)
  - Optionally equipped with Pipeline Bypass – controlled from FPP Config
- If layout permits, pipeline registers can be shared between paths

**FPP Lane Control Signals**

- **Per Destination**
  - Source Select if \#Sources ∈ \{2, 3, 4\} 0 – 2
  - Output Enable only for FPP_PRI and FPP_SEC; only if output 0 – 2
- **Per Path**
  - Hold Bypass optional 0 – 1
  - Pipeline Bypass if pipeline bits are implemented; optional 0 – 1
Leveraging Lessons-Learned on 2D-SOCs in Designing Parallel TAMs for 3D-SICs Based on IEEE Std 1838’s Flexible Parallel Port

Erik Jan Marinissen

3. FLEXIBLE PARALLEL PORT (FPP)

FPP Channels

- **Channel** = Set of identical FPP Lanes
  - Identical Paths
  - Identical Clock signal
  - Identical FPP Config control signals

- **Sharing**: different Channels can share
  - Clock Signals
  - FPP Config control signals

- **Scalability**
  - Number of Lanes per Channel
  - Number of Channels per FPP

4. PTAM Design for 2D-SOCs
Leveraging Lessons-Learned on 2D-SOCs in Designing Parallel TAMs for 3D-SICs Based on IEEE Std 1838’s Flexible Parallel Port

Erik Jan Marinissen

4. PTAM DESIGN FOR 2D-SOCs

IEEE Std 1500 ‘ParallelTest’ Modes

- IEEE Std 1500 did recognize that a PTAM might be needed to get acceptable test times, especially for INTEST of large cores during high-volume production.

- But: IEEE Std 1500 only defines the parallel ports: WPI-WPO

4. PTAM DESIGN FOR 2D-SOCs

TestRail (=PTAM) Configurations

- Multiplexing Architecture
  - Single TestRail
  - Implementation with MUXes or as tri-state bus
  - Interconnect test cumbersome/difficult

- Daisychain Architecture
  - Single TestRail
  - Core: swap in/out (= bypass) of active daisychain
  - Allows sequential and parallel test of cores

- Distribution Architecture
  - Multiple TestRails
  - Need to distribute the overall TAM width over the various TestRails

[Goel & Marinissen – ITC 2002]
Leveraging Lessons-Learned on 2D-SOCs in Designing Parallel TAMs for 3D-SICs Based on IEEE Std 1838’s Flexible Parallel Port

Erik Jan Marinissen

4. PTAM DESIGN FOR 2D-SOCs

‘TR-Architect’ Software Tool

- **Given**
  - SOC cores details + their tests
  - SOC TestRail width \( W_{\text{max}} \)
  - Type of TestRail architecture

- **Determine**
  - \#TestRails + width per TestRail
  - Assignment of cores to TestRails
  - Wrapper design per core

- **Such that**
  - SOC test application time \( T \)
    (= ATE vector memory) is minimized

```
<table>
<thead>
<tr>
<th>Type-1 Idle Bits</th>
<th>Imbalanced TestRail test completion times</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type-2 Idle Bits</td>
<td>Module assigned to TestRail pf which width non Pareto-Optimal</td>
</tr>
<tr>
<td>Type-3 Idle Bits</td>
<td>Imbalanced scan chains in module</td>
</tr>
</tbody>
</table>
```

4. PTAM DESIGN FOR 2D-SOCs

Example Idle Bit Analysis to Optimize Architecture

Gain in ATE memory

Hybrid Architecture
All Hard Modules

Distribution Architecture
All Hard Modules

Gain in ATE memory

Hybrid Architecture
All Hard Modules

Gain in ATE memory

Hybrid Architecture
One Soft Module

Real Case Idle Bit Analysis to Optimize Architecture

Gain in ATE memory

Hybrid Architecture
All Hard Modules

Gain in ATE memory

Hybrid Architecture
Two Soft Modules
5. Lessons Learned on 2D-SOCs

What Did We Learn on 2D-SOCs That Helps for 3D-ICs

- No multiplexing PTAMs (includes ‘TestBus’)
- Hybrid Distribution + Daisychain Architecture gets you minimal test time / TDV
- Non-optimal architectures come from Distribution Architecture
  - Distribution of $W_{\text{max}}$ over various modules needs to be proportional to TDV of the module it serves
  - PTAM width needs to be integer
  - Accurate TDV data on modules (#SFFs, #test patterns) is needed at design start, but only becomes available when design is done
- Daisychain architecture is preferred for 2D-SOCs
  [Waayers, Morren, Grandi – ITC 2005]
Leveraging Lessons-Learned on 2D-SOCs in Designing Parallel TAMs for 3D-SICs Based on IEEE Std 1838’s Flexible Parallel Port

Erik Jan Marinissen

5. LESSONS LEARNED ON 2D-SOCs

Daisychain Makes Even More Sense in 3D-SICs

- In a 3D-SIC, Distribution Architecture implies that the dies below your die need to implement a PTAM while they are not allowed to use it – I think that will not fly!

6. Mapping IEEE Std 1838™ FPP on “My” 3D-DfT Architecture

[Li et al. – ETS 2018]
Leveraging Lessons-Learned on 2D-SOCs in Designing Parallel TAMs for 3D-SICs Based on IEEE Std 1838’s Flexible Parallel Port

Erik Jan Marinissen

6. MAPPING IEEE STD 1838™ FPP ON “MY” 3D-DfT ARCHITECTURE

Test@First 3D-DfT Architecture

- Test Access
  - **Serial**: mandatory 1-bit wide Serial Port
  - **Parallel**: optional n-bit wide Parallel Port

- In This Die
  - **ExTest**: WBR only
  - **InTest**: WBR + internal scan chains
  - **Bypass**: skip this die

- After This Die
  - **Elevate**: travel up to next die
  - **Turn**: turn downward to external I/Os

FPP Lane Example

- **Paths**
  - **Source** → **Destination**
    - 1. FPP_PRI → FPP_SEC
    - 2. FPP_PRI → FPP_TO_SIDE
    - 3. FPP_PRI → FPP_TO_CORE
    - 4. FPP_FROM_CORE → FPP_SEC
    - 5. FPP_FROM_CORE → FPP_TO_SIDE

Marinissen et al. [VTS’10, 3DIC’10, JETTA’12] US Patent 9,239,359

[Li et al. – ETS 2018]
Leveraging Lessons-Learned on 2D-SOCs in Designing Parallel TAMs for 3D-SICs Based on IEEE Std 1838’s Flexible Parallel Port

Erik Jan Marinissen

6. MAPPING IEEE STD 1838™ FPP ON “MY” 3D-DFT ARCHITECTURE

FPP Lane Example: Customization During Test

- ** Paths **
  
  **Source** | **Destination**
  --- | ---
  1. FPP_PRI | FPP_SEC
  2. FPP_PRI | FPP_TO_SIDE
  3. FPP_PRI | FPP_TO_CORE
  4. FPP_FROM_CORE | FPP_SEC
  5. FPP_FROM_CORE | FPP_TO_SIDE

- **Customization during Test**
  - Implement full FPP
    - Expensive in on-chip hardware
    - All options still available during test through configuration

Marinissen et al. [VTS’10, 3DIC’10, JETTA’12] US Patent 9,239,359

6. MAPPING IEEE STD 1838™ FPP ON “MY” 3D-DFT ARCHITECTURE

FPP Lane Example: Customization During Design

- ** Paths **
  
  **Source** | **Destination**
  --- | ---
  1. FPP_PRI | FPP_SEC
  2. FPP_PRI | FPP_TO_SIDE
  3. FPP_PRI | FPP_TO_CORE
  4. FPP_FROM_CORE | FPP_SEC
  5. FPP_FROM_CORE | FPP_TO_SIDE

- **Customization during Design**
  - Implement subset of FPP
    - Reduced on-chip hardware
    - Only implemented options are available

Marinissen et al. [VTS’10, 3DIC’10, JETTA’12] US Patent 9,239,359
Leveraging Lessons-Learned on 2D-SOCs in Designing Parallel TAMs for 3D-SICs Based on IEEE Std 1838’s Flexible Parallel Port

Erik Jan Marinissen

6. MAPPING IEEE STD 1838™ FPP ON “MY” 3D-DfT ARCHITECTURE

FPP Lane Example: Lock-Up Latch, Pipeline, Control

- **Paths**
  - Source       Destination
  - 1. FPP_PRI   → FPP_SEC
  - 2. FPP_PRI   → FPP_TO_SIDE
  - 3. FPP_PRI   → FPP_TO_CORE
  - 4. FPP_FROM_CORE → FPP_SEC
  - 5. FPP_FROM_CORE → FPP_TO_SIDE

- Mandatory **Hold Elements**
- Paths 1 and 2 have a shared optional single Pipeline Register
- FPP configuration signals: FPP_SEC_OE and !BYPASS

6. MAPPING IEEE STD 1838™ FPP ON “MY” 3D-DfT ARCHITECTURE

Test@First 3D-DfT Architecture

- Die
- ClkLane
- Func. Logic
- FPP_SEC Logic
- UPDATE
- TDI
- TCK
- TDO
- FPP_SEC
- FPP_FROM_SIDE
- FPP_TO_SIDE
- FPP_CLK_IN


Marinissen et al. [VTS’10, 3DIC’10, JETTA’12] US Patent 9,239,359


Marinissen et al. [VTS’10, 3DIC’10, JETTA’12] US Patent 9,239,359


Marinissen et al. [VTS’10, 3DIC’10, JETTA’12] US Patent 9,239,359

7. Conclusion

3D-SICs are best tested in a modular fashion and thus require DfT wrappers
- To minimize test time (= test cost) in HVM, these wrappers need parallel TAMs
- IEEE Std 1838™-2019 is the new IEEE standard for 3D-DfT wrappers
  - The standard includes a Flexible Parallel Port (FPP), on optional/scalable PTAM
- With 2D-SOCs, we have learned that the ‘single Daisychain’ is the best architecture
  - The total PTAM width depends on #package pins and is typically known at design start
  - All designers of modules/cores can then design their scan chains to that number
- In 3D-SICs, it makes even more sense to implement a Daisychain architecture
  - Avoids that lower dies need to implement PTAMs that pass through only...
Leveraging Lessons-Learned on 2D-SOCs in Designing Parallel TAMs for 3D-SICs Based on IEEE Std 1838’s Flexible Parallel Port

Erik Jan Marinissen
Abstract—3D stacked die technology is a major driving force that enables SOC products to incorporate multiple heterogeneous designs into a smaller package form factor. However, it introduces unique challenges to high volume manufacturing (HVM) that requires feasible & practical solutions to ensure mass market productization. This paper will discuss the various manufacturing challenges & its corresponding DFT & test strategies to make 3D stacking commercially viable.

Keywords—3D Stacked IC; High Volume Manufacturing; Test Strategy

I. INTRODUCTION

Figure 1 shows 2 top dies stacked together on top of a bottom die. The 2 top dies are connected to the bottom die thru an inter-die interconnect with special IOs, referred as internal I/O[1]. Interactions between the top dies and bottoms die are transacted thru this special I/Os. Thru Silicon Vias (TSV)s are used to deliver power & connect signals to the package C4 bumps. The die stacking significantly reduces the number of TSV connections needed while supporting more interconnects between active devices in the multiple dies.[2] The top dies signals can either be driven thru the internal I/Os or thru the TSV in the bottom die to the package C4 bumps. Meanwhile the bottom die's I/O will be driven thru the TSV directly to the package C4 bumps.

Figure 1: Diagram of a 3D stack die SOC showing I/O pads, TSVs and inter-die interconnects

II. WAFER LEVEL (SORT) TEST PROBING CHALLENGES

During wafer level testing or referred as “SORT” testing, each of the individual dies need to be fully tested before they are stacked together and tested at post assembly package test socket (hereinafter referred as “CLASS” testing). SORT is tested through the ubump which has a smaller pitch as compared to regular C4 bump. It is not feasible to make contact with every ubump. This poses a challenge for the inter-die interconnect signals that could number up to a few hundred and even thousands depending on the chip design. Some of the inter-die signals must be connected to the SORT probes to enable testing. Therefore to overcome the ubump pitch technology limitations, it is important to plan ahead to identify which signals are truly critical and then ensuring that their placements are adherent to the probing rules. Any accidental exclusion mistake could render the chip untestable at SORT. Selecting too many inter-die interconnect signals to be probed will increase the structural design complexity and the number of costly SORT probes. Hence, DFT survivability & override hooks are added to ensure that certain power-up sequences can be overridden thru TAP pins if there is an issue driving the inter-die connect I/Os at SORT. The number of test access ports thru the ubumps can be reduced by implementing internal die test compression/decompression scheme to reduce the need for test probes connectivity.

Figure 2: Layout constraint on inter-die interconnect signals due to pitch technology limitation that warrants carefully spaced out probing selections

Extensive X-validation during pre-silicon phase is a must to ensure that inter-die signals will not be accidentally propagating “X”s into the logic circuits and causing the test to fail. This is done by purposely injecting “X”s on all the unprobed inter-die signals on the simulation models. Additionally, die isolation DFT hooks were also implemented to ensure that each individual die can be tested in an independent manner at SORT & also in CLASS testing for test parallelism purposes.
The probing constraints also further amplifies the power delivery issues faced in 3D stacked technology to deliver sufficient power while ensuring sufficient number of ubumps are made probe-able thru pre-planned layout spacing. Additionally, current & power consumption modelling simulations must be comprehended thoroughly when intra-die test parallelism is being deployed to achieve efficient test time for HVM.

III. INTERNAL I/O & INTER-DIE INTERCONNECT TEST STRATEGY

Due to SORT probing limitation, not all ubumps can be probed and connected to the tester to be tested. This is a test coverage gap at the SORT that must be addressed to ensure that the defect does not flow downstream that would then incur costly post assembly package. There are hundreds of signals at the inter-die connect which is just too many to be connected to the probes. Spacing out all the ubumps appropriately to fulfil all the probing pitch requirements is not a feasible option.

The solution is to implement a DFT architecture with the emphasis on no-touch type of testing methodology for 3D stack technology as shown in Figure 3. We refer these inter-die signals IOs as Internal I/Os (IIO). These IIOs are grouped to a few family grouping according to their native functional speed and functions. It embodies the following features: (i) Reuse the existing functional flops instead of a standard boundary scan approach in order to optimize and reduce the DFT overheads (ii) Supports at-speed & slow (tap frequency) testing without the use of scan ATPG by utilizing the LFSR/MISR approach (iii) Analog loopback capabilities for each of the ubump (iv) Burn-in and parametric test support (v) CLASS test screening capabilities when the 2-dies are stacked together (vi) Analog monitoring capabilities & (vii) No Touch Leakage capabilities

The reset bring-up sequences to enter the DFT test mode to enable the internal IOs tests must be as simple as possible to avoid a chicken and egg scenario especially during inter-die testing due to dependency between the top and bottom die bring-up. All critical signals identified must have DFT override capabilities in place in their respective dies for survivability purposes.

Figure 3 shows the high-level internal IOs design which drives signals thru the interconnects between the 2 stacked dies.

<table>
<thead>
<tr>
<th>Test Type</th>
<th>Sort Coverage</th>
<th>Class Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>DC</td>
<td>Loopback at each internal IO driver at TAP clock frequency with static-0 and static-1 pattern</td>
<td>Sending static-0 and static-1 across the top and bottom die at TAP clock frequency</td>
</tr>
<tr>
<td>DC</td>
<td>No touch leakage (NTL) capabilities</td>
<td>No Touch Leakage tested on single die at one time only. Test preconditioning is needed on the other non-tested die</td>
</tr>
<tr>
<td>At Speed</td>
<td>Loopback at each internal IO at native clock frequency with pseudo random pattern (LFSR) and signature based response (MISR)</td>
<td>Sending LFSR from TX of one die to RX of another die at native frequency. MISR on the RX side will collect the incoming data to form a final signature response.</td>
</tr>
<tr>
<td>Stress</td>
<td>Loopback at each internal IO with a toggling 101010 pattern</td>
<td>Sending LFSR (101010 patterns) from TX of one die to RX of another die at native frequency. MISR on the RX side will collect the incoming data to form a final signature response.</td>
</tr>
<tr>
<td>Debug</td>
<td>Analog monitor for each family grouping</td>
<td>Analog monitor for each family grouping</td>
</tr>
<tr>
<td>Debug</td>
<td>Ability to pinpoint and mask out failing signal going into the MISR</td>
<td>Ability to pinpoint and mask out failing signal going into the MISR</td>
</tr>
</tbody>
</table>

Table 1: Testing feature summary for the Internal I/Os for SORT & CLASS test coverage

IV. THROUGH SILICON VIA (TSV) TESTING

The TSV on the bottom die is also a major component of the overall 3D stack technology which is used for (i) power delivery to the bottom die (ii) signal & power delivery to the top die itself. During CLASS package testing, IO loopback, boundary Scan (BSCAN) tests and rcomp code tests (for rcomp pins with external resistor on the test board) can be employed to screen out defects in the TSV. These tests supplement the mandatory parametric vcc-continuity/open/shorts tests at the package level. [3]
A test combination of high-speed (IO loopback) & slow-speed (BSCAN) can be complimentary to each other to gauge the overall health of the TSVs during HVM testing. It could be used as first level debug to isolate whether the defect is on the IO buffer itself or the TSV before further physical debug is performed. The Rcomp pins can also be muxed to an external resistor & tester channel card thru the use of a relay switch to provide the flexibility to provide parametric coverage on the Rcomp pin’s TSV.

Figure 4 shows the different IO topology related to the TSV and their corresponding coverage methods at CLASS.

V. 3D STACK IC TEST ACCESS MECHANISM

Another consideration point when executing 3D stack testing is the test port mechanism to access the top die when the dies are stacked together. Test access ports can be designed to be driven i) thru the TSV to the top die or ii) base die IO pins going thru the inter-die interconnect to the targeted top die as shown in Figure 5. Using the 1st option would mean additional dedicated CLASS package pins just to test out the top die versus reusing the same test access pins for both SORT & CLASS (2nd option). However, having a dedicated test access mechanism for the top die enables test content parallelism where both top & bottom dies can execute testing in parallel and independently of each other hence significantly reducing overall costly test time.

Test vector patterns & flows can be constructed in such a manner that a single test can be transposed seamless to work on the 2 different test access mechanisms by manipulating the different timing input/output result expectations. Vector patterns traversing thru the TSV to the top die will have lower latency versus going thru base die, inter-die to reach the top die's test access which will have multiple staging flops in between. These differences can be easily accounted for during the vector conversion flow to produce 2 set of test vectors accessing different test access using a single set of test collaterals.

Having this flexibility of choosing the different test access mechanism for the top die is essentially useful to serve multiple market segments with key product differential features with the same core design. For example, additional multi-chip packaging, CoPoP packages, EMIB bridges, etc on derivatives chips might indirectly eliminate the top die's direct access to the C4 packages thru the TSVs. The dual test access mechanism also provides a backup alternative if a particular test access path is found to be defective.

Figure 5: The multiple test access mechanism for the top die for testing flexibility

VI. RESULTS & LESSONS LEARNT

The solutions were imperative to resolve the challenges of bringing the 3D stacking technology to high volume manufacturing. One of the key lessons learnt here is that the importance of an early engagement and well planned technical readiness communication between all the stakeholders (architecture, DFT, front-end design, layout, test manufacturing) throughout the entire project phase cannot be understated. To further tighten the technical collaborations across multiple disciplines, the test manufacturing team have executed a shift-left initiative to be involved in DFT & pre-silicon content validation efforts to enable real-time feedback to design team from the very beginning. Comprehensive pre-silicon validation was vital to ensure that early assumptions were validated and verified before silicon arrives.

VII. SUMMARY & CONCLUSION

The purpose of this paper was to highlight the challenges and solutions taken to enable high volume manufacturing for silicon die that utilizes the 3D stacking technology. 3D stacking technology is truly ground-breaking as it enables more transistors to be packed into a very small form factor. The cost of this technology is that it comes with more complexity and challenges versus the traditional SOC or multi-chip design. However, the author is confident that in the future with more learnings & new innovations, these entry barriers can be reduced and overcome easily.
ACKNOWLEDGMENT

The authors would like to give the heartiest appreciation of thanks to all the wonderful folks that directly and indirectly involved in bringing the technology and solutions described in this paper into a reality for high volume manufacturing.

REFERENCES

Intel Foveros Technology: DFT And HVM Test Strategy

Wei Ming Lim, Terrence Huat Hin Tan, Sook Kwan Cheah, Kian Lek Koay, Sreejit Chakravarty

Presenter: Wei Ming Lim
Intel Corporation
Agenda

• Introduction to Intel Foveros
• Wafer Level Test Probing Challenges
• Inter-Die Interconnect Test Strategy
• Through Silicon Via (TSV) Testing
• 3D Stack IC Test Access Mechanism
• Results & Lessons Learnt
• Summary & Conclusion
“Foveros” is a new 3D technology invented at Intel that allows logic chips to be stacked for the 1st time, delivering high compute density and enabling a complete rethinking system architecture into a whole new dimension.
Wafer Level Test Probing Challenges

- Wafer level testing is done thru the uBump which have smaller pitch vs regular C4 bump
- Stringent selection in deciding which uBump signals needs to be probed
- Critical signals must have DFT survivability & override options (Validate assumption thru X-validation)
- Ensure sufficient ubumps are probed to deliver sufficient power during various testing scenarios
# Inter-Die Interconnect Test Strategy

<table>
<thead>
<tr>
<th>Test Type</th>
<th>Sort Coverage</th>
<th>Class Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DC Parametric</strong></td>
<td>Loopback at each internal IO driver at TAP clock frequency with static-0 and static-1 pattern</td>
<td>Sending static-0 and static-1 across the top and bottom die at TAP clock frequency</td>
</tr>
<tr>
<td><strong>DC Parametric</strong></td>
<td>No touch leakage (NTL) capabilities</td>
<td>No Touch Leakage tested on single die at one time only. Test preconditioning is needed on the other non-tested die</td>
</tr>
<tr>
<td><strong>At Speed</strong></td>
<td>Loopback at each Internal IO at native clock frequency with pseudo random pattern (LFSR) and signature based response (MISR)</td>
<td>Sending LFSR from TX of one die to RX of another die at native frequency. MISR on the RX side will collect the incoming data to form a final signature response.</td>
</tr>
<tr>
<td><strong>Stress</strong></td>
<td>Loopback at each Internal IO with a toggling 101010 pattern</td>
<td>Sending LFSR (101010 patterns) from TX of one die to RX of another die at native frequency. MISR on the RX side will collect the incoming data to form a final signature response.</td>
</tr>
<tr>
<td><strong>Debug</strong></td>
<td>Analog monitor for each family grouping</td>
<td>Analog monitor for each family grouping</td>
</tr>
<tr>
<td><strong>Debug</strong></td>
<td>Ability to pinpoint and mask out failing signal going into the MISR</td>
<td>Ability to pinpoint and mask out failing signal going into the MISR</td>
</tr>
</tbody>
</table>

- The Internal I/Os utilizes existing logic/flops, reducing the overhead cost
- Supports both TAP clock & at-speed/native speed frequency testing
- Inter-die interconnect testing is a critical screen at CLASS
Through Silicon Via (TSV) Testing

- **Mandatory parametric testing:**
  Vcc-continuity/Open/Shorts testing/Leakage
- A combination of HVM tests are needed to narrow down whether defect is due to TSV or otherwise.
- Further failure analysis can then be done on the unit
- **Key is to ensure that all TSVs are covered in testing!**
Test Access Ports for Top Die can be designed to be driven:

i) Directly thru the TSV to the top die

ii) Bottom die IO pins going thru the inter-die interconnect to the targeted top die

Considerations:

- Alternative test access path for Top Die
- Parallelism Testing for Top Die and Bottom Die
- Singular reusable ATE test vector across different test access ports
- Flexibility to respond to different market segment where primary test access mechanism might be no longer accessible (attaching DRAM, EMIB bridge, etc)
Results & Lessons Learnt

- Tight collaboration between multiple stakeholders and discipline
- Shift-Left initiative & influencing by High Volume Manufacturing (HVM) Team
- Validate all initial assumptions made during technical readiness in pre & post silicon activities
- Active & timely continuous feedback
- Post silicon data gathering fuels future improvements
Summary & Conclusion

• Despite various unique challenges in Foveros, Intel has successfully brought this new 3D advanced packaging technology into High Volume Manufacturing.

• Foveros will pave the road to further new advancement & innovation in packaging technology that will significantly improve interconnect density, power efficiency & scalability.

• This simply would not be possible without the hard work and dedication of fellow engineers who worked tirelessly to bring Foveros into reality.
3DC-TEST
Thank You
Who's at Fault? A Creative Way to Isolate and Debug Internal IO Failures

Devanraj Letchumanan, Intel Corporation
Ahmad Hisyamuddin Arshad, Intel Corporation

3DC-TEST
Nov 6-7, 2020
Problem Statement

- Increased die integration/stacking on a single package produces increased number of internal (non-touch) IOs for inter-die communication

- These IOs are susceptible to process defects, ESD events, circuit marginalities etc. during package assembly or otherwise

- With 3D/Foveros technology, fault isolation on internal IOs have become more complex due to probe and FA limitations

- An approach of utilising in-built IO DFT features to help engineers isolate the problematic die or circuitry is developed to address this challenge

*Foveros Example:*

*Monolithic die optical probe:*

With limited physical visibility on stacked dice & increased bump density, debugging internal IO failures is a lot more challenging

Successful defect finding and time to root-cause is key
Common loopback & inter-die data paths

- Near-end loopback test is key in diagnosing internal IO failures as well as determining subsequent fault isolation move and debug strategy.
- The significance of the loopback path within the analogue-front-end circuitry together with an inter-die test is essential in deriving a successful fault model for internal IOs.

While loopback test may not be required from test coverage perspective, it’s essential for debug and fault isolation.
# Fault diagnosis model

<table>
<thead>
<tr>
<th>DieA NELB</th>
<th>DieB NELB</th>
<th>Die-to-die</th>
<th>Failing mode(s) for NELB at PAD</th>
<th>Failing mode(s) for NELB isolated from PAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>fail</td>
<td>fail</td>
<td>fail</td>
<td>Short at A/B/Bump</td>
<td>Not Applicable for single defect</td>
</tr>
<tr>
<td>fail</td>
<td>fail</td>
<td>pass</td>
<td>Marginal/Resistive if fail at-speed only</td>
<td>Not Applicable for single defect</td>
</tr>
<tr>
<td>fail</td>
<td>pass</td>
<td>fail</td>
<td>Die A logic: TX if A-to-B OR RX if B-to-A failed</td>
<td>Die A logic: Pre-TX if A-to-B OR RX if B-to-A failed</td>
</tr>
<tr>
<td>fail</td>
<td>pass</td>
<td>pass</td>
<td>Die A logic</td>
<td>Die A logic: loopback path</td>
</tr>
<tr>
<td>pass</td>
<td>fail</td>
<td>fail</td>
<td>Die B logic: TX if B-to-A OR RX if A-to-B</td>
<td>Die B logic: Pre-TX if B-to-A OR RX if A-to-B</td>
</tr>
<tr>
<td>pass</td>
<td>fail</td>
<td>pass</td>
<td>Die B logic</td>
<td>Die B logic: loopback path</td>
</tr>
<tr>
<td>pass</td>
<td>pass</td>
<td>fail</td>
<td>Open at Bump</td>
<td>Open/Short at Diodes/Bump OR TX</td>
</tr>
</tbody>
</table>

*Table derived with the assumption of a single defect/failing mechanism*
Case study example

1. **Test** - A stacked die package had failed die-to-die interconnect test at-speed while passing at lowered frequency

2. **Diagnose** - By testing either die loopback tests for the failing IO, its results points to “DieB” Transmitter malfunction

3. **Isolate** - With this granularity, by performing silicon data dump with added stimulus, the path was identified which demonstrated a timing margin fault on further analysis

**Without sufficient DFT coverage, a more invasive debug/FA approach would be required to root-cause the failure, for example:**

<table>
<thead>
<tr>
<th>DieA NELB</th>
<th>Die-to-die (B to A only)</th>
<th>Failing mode(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>fail</td>
<td>fail</td>
<td>Short at A/B/Diodes/Bump</td>
</tr>
<tr>
<td>fail</td>
<td>pass</td>
<td>Die A logic</td>
</tr>
<tr>
<td>pass</td>
<td>fail</td>
<td>Open Bump OR Die B logic</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DieA NELB</th>
<th>DieB NELB</th>
<th>Die-to-die</th>
<th>Failing mode(s) for NELB at PAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>fail</td>
<td>fail</td>
<td>fail</td>
<td>Short at A/B/Bump</td>
</tr>
<tr>
<td>fail</td>
<td>fail</td>
<td>pass</td>
<td>Marginal/Resitive if fail at-speed NELB only</td>
</tr>
<tr>
<td>fail</td>
<td>pass</td>
<td>fail</td>
<td>Die A logic: TX if A-to-B OR RX if B-to-A failed</td>
</tr>
<tr>
<td>fail</td>
<td>pass</td>
<td>pass</td>
<td>Die A logic</td>
</tr>
<tr>
<td>pass</td>
<td>fail</td>
<td>fail</td>
<td>Die B logic: TX if B-to-A OR RX if A-to-B</td>
</tr>
<tr>
<td>pass</td>
<td>fail</td>
<td>pass</td>
<td>Die B logic</td>
</tr>
<tr>
<td>pass</td>
<td>pass</td>
<td>fail</td>
<td>Open at Bump</td>
</tr>
</tbody>
</table>

In the absence of bidirectional die-to-die interconnect and loopback tests on both dice, debug/FA would involve more components thus increasing the risk of omitting/removing the failing part, especially where invasive debug techniques are required
3DC-TEST
Thank You
Pre silicon validation methodology breakthrough for 3D IC

Wai Loon, Yip
Intel Technology Sdn Bhd,
Penang, Malaysia
wai.loon.yip@intel.com

Hock Thien, Ng
Intel Technology Sdn Bhd,
Penang, Malaysia
hock.thien.ng@intel.com

Bian Sim, Teo
Intel Technology Sdn Bhd
Penang, Malaysia
bian.sim.teo@intel.com

Abstract— The increase demand for functionality and miniaturization forces industry to enhance chip performance, without sacrificing valuable board space. This is achieved by stacking silicon wafers and interconnecting them vertically known as a three-dimensional integrated circuit (3D IC) or stacked die. Validation of the 3D IC during pre-silicon becomes a challenge as conventional monolithic model is no longer a viable method. The pre silicon challenge is always on the long simulation validation time, the sheer size of simulation model now that is it stacked and the validation coverage. This paper explains the pre silicon validation methodologies implemented in our facility today to achieve effective testability and debug capability of the stacked device.

Keywords— 3D IC; stack die; pre-silicon; validation

I. INTRODUCTION

A 3D IC also known as stack die is a combination of several dies vertically stacked and with a variation of multiple tiles too, illustrated in Fig 1. Besides providing a reduction in overall package footprint, stack die also substantially improve electrical performance through quicker transmissions requiring less energy to drive the signals. It is not feasible anymore for conventional monolithic full coverage testing methods in Resistor Transistor Logic (RTL) simulation model due to its high computational and memory consumption; and long simulation time that can span over several weeks for a test case.

With the stack die, new test methods need to be derived to overcome the challenges it presents in pre silicon simulation without compromising validation coverage.

II. PRE-SILICON VALIDATION METHODOLOGY

A. Multiple partial monolithic model testing

Testing a full stack die in a large monolithic model is close to impossible due to high processing memory and disk storage space required. Besides, it will also take several weeks per test for simulation to complete making it non pragmatic for debug activities. The solution is to breakdown the stacked die model to several smaller model with different test focus areas per model known as partial die testing. The model segregation could be main die, stack die with reduction of repeated chiplets and multiple partial stack die model with selected chiplets for test. This sometimes led to explosion of simulation models to manage and support by design but making test and validation manageable and pragmatic. Thus, the validation team needs to strike a balance in selecting the required models to achieve the full validation coverage with correct test strategy per model.

There several ways to break a full stack die model into smaller monolithic model, as below:

a) Standalone chiplets model testing
Separate it by chiplets itself, as show in Fig 2, which each of Die having one standalone model.

b) Partial IP on single die model
Further divide down the chiplets model into smaller model, normally use it for IP/partition testing

c) Shrinks down a full stack die model – Partial Stacked Die model.
To ensure all connection between dies is cover and basic functional between cross die is covered.

Fig 1. Simplified block diagram view portraying variations of 3D IC

Fig 2. Example of illustration of Partial stack model versus full stack model

Fig 3. Example of illustration of Partial IP on single die model.
B. X injection to validate standalone die isolation

One of the most effective test methods in pre-silicon validation to test die isolation connectivity is to perform ‘x’ injections to the die IO of the stack model. If there is a discontinuity in the signal path, the ‘x’ signal will cause the signal path in validation to fail due to ‘x’ propagation. Detail failures will then be investigated and traced in RTL simulation waveform to determine the source of the failure propagation thus detecting the broken isolation path that require fixes. This validation technique effectively detects that the die in test is properly isolated.

In this technique, it is assumed that one of the major DFT feature is to have a designed-in switch (that can be toggled through fuse) to ease die isolation during test. This particular switch primarily used for two reasons. The first is to initiate logical isolation of the stack dies Rx (input) pins. In wafer testing, these pins will be dangling into the main die so for stability reasons, it needs to be tied to a defined value. The second is to bypass the required handshake between chiplet and main die. This allows the functional boot to complete without chiplet hanging while waiting for non-existent chiplets to respond to message requests to main die. This designed-in feature needed to be well thought of in early design stage to link the chiplet respond network that is going to the main die to a switch that validation team can enable internal cross die signal with ease.

The X injection method can also successfully detect power availability of a die when tested on a power awareness model. This is because when a die is not properly powered up its IO and registers will not be at a deterministic state resulting in x propagation observation.

C. Controlling dies enabling for parallel and serial die testing

Another challenge that a stack device often encounter is power consumption. Depending on the functionality of the device, when there is occurrence of power constraint from the power supply from tester, certain test may be needed to be tested in serial instead of parallel. In parallel testing, all the dies in the device is powered up and validated at the same instance with no additional cost to test time. However parallel test can only be implemented when sufficient tester power supply could be supplied to the device else serial test is required. In serial testing, both dies are powered up where only the die is in test in fully out of reset and the other idle dies are halted at a minimum state to reduce device power consumption. Fig. 4 illustrates the test flow described.

Upon enabling the die isolation in parallel test flow described above, the external pins also needed to be managed accordingly for the device shared external pin to ensure the inputs and changes for the die in test is isolated from the idle die. Example in Fig5, external pin is gang together between Die 0 and Die 1, the criteria is to ensure that the “Idle Die” stays at a certain defined state with minimum voltage consumption.

This could easily be achieved by setting the voltage level of the pin to zero. To ensure invalid fail due to “Idle Die” TDO read and test port out read, need to be masked. This methodology on die selection could be fully validated TDO results and detail waveform analysis in pre silicon RTL simulation with the configurations described.

D. Die IO test

In pre-silicon stage, the luxury to access all the IO ports enable validation of the ports through loopback tests technique. Some of the IO ports are microbumps for internal die to die connectivity. Once a stack die is assembled, the direct accessibility to the microbumps IO ports no longer available, thus making test more complex. During pre-silicon
stage, a model file (turned on through a config switch during simulation run execution) can be setup to map the transmitter port, Tx back to receiving port, Rx. A defined value is then injected to the transmitter port and travel through the receiving port and being compared with the injected value for an expected match. Thus, partial model or individual die model could be used to identify the broken node in the IO network instead of using a full stack die model.

Fig 8. IO test without the need for full stack model.

E. Test pattern handling techniques

Capabilities to enable multiple chiplets testing and controlling of die enabling in parallel and serial test mentioned in previous sections requires thorough considerations and implementation thoughts in handling stack die test patterns. Following 3 aspects describes the required test pattern handling for stack die and its die.

This first is when one or more chiplets are combined, where 2-way communication occurs, the pattern pin groups also known as pattern domain need to be combined or stitched together to form a complete pattern to enable testing of both dies. This automation is required to merge the multiple die domains and to ensure the start and stop vector transactions are aligned.

Fig 9. Combining different pattern domain for 3 different dies.

The next aspect is pin convention planning required for multi-die stack die for die enabling control. As the pins for each die are identical, a naming convention for the differentiation is required. With this differentiation, controllability of the die enabling could be manage in test pattern algo rhythm to mask or unmask the pin response.

Lastly, the third aspect discussed here is the strategy to use same pattern source for stack multic和平 testing. As the cascaded dies are the same, it is logical to just duplicate the pattern for die1 from die0 pattern source instead of having multiple pattern source. This eases pattern source management and disk space usage optimization.

F. Test pattern validation

With today’s complexity of the large scaled design especially in the 3D IC design, simulation and emulation test pattern validation is no longer the optimum approach to be run. Simulation validation on a full stack die model easily consumes days or weeks, considering the debug iteration involved it is completely an out of scale approach. On the other hand, emulation validation could be relatively faster than simulation validation, but the premium cost of emulator hardware is hardly justified. 2 static checkers are introduced to replace the heavy time and resource consuming conventional validation method.

The first checker is named STIL Checker, where it is treated as a signoff checker to run various checks on a test source written in IEEE Standard Test Interface Language (STIL) format before test pattern conversion. The checker covers STIL block construction range from STIL format, signal, signal group, label, procedure, macro, timing and waveform table, user keyword and vectors. These comprehensive checks will make sure test pattern can be converted to ATE (Automated Test Equipment) format without error. With the complexity of pin convention for multi-die stack die, the offline checker will be the ideal solution to make sure all the pin convention are unique.

The second checker is named Pattern Checker, where it takes inputs from the golden test source (STIL) and HVM (High Volume Manufacturing) timing inherited test pattern (ATE format), converting both inputs file to FSDB waveform and performing waveform-to-waveform comparison. Comparing test pattern in ATE format to golden test source can eliminate all pattern conversion issue. Issue that we previously discussed about the multiple die domains alignment can be easily identified through the checker.

Pattern Checker flow diagram is shown in Fig. 10.
These offline testing capabilities are proven to be cost and time effective solution for stack die pre-silicon validation to cover the depth of the test content quality check in volume.

III. RESULT & CONCLUSIONS

Comprehensive pre-silicon validation increases the confidence of the silicon health of the final stacked device.

Before adopting these validation techniques, RTL validation in the past consume long simulation time particularly on huge simulation models which is not practical. This limitation poses a risk of incomplete validation test plans before silicon arrival. The challenge was to bring the validation time down with increased coverage.

This paper successfully discussed and present the pre-silicon validation methodologies of 3D IC. These methods have been proven to overcome the challenges described. They were successfully implemented on Intel 1st product built-in on Foveros technology, Lakefield which is stack of 2 dies and ongoing products. The high-test coverage that was achieved during pre-silicon was proven effective with the plug and play results during 1st silicon test. The isolation techniques enable us to test die independently were proven in actual silicon test flow today. The ability to test die in isolation for a stack device had tremendously expedites the root cause of test failures. Complimenting simulation validation, static pattern validation tools STIL and Pattern Checkers are proven to be cost and time effective solution to cover the depth of the test content quality checks in volume without the need for conventional computing resources in simulation.

Future work is to continuously prove this on multi stack die build on Foveros technology, hence bring the maximum effectiveness to the silicon product time to market with quality.

ACKNOWLEDGMENT

We would like to acknowledge the following reset validation and test vector team members who were not already listed as authors. Team members Chan, Xin Lei; Chong, Chee Mei; Lok, Wai Yee; Mohamed Nazir, Syed Nayeem B; Tan, Hsio Mei; Tan, Pei Jing; Teh, Pei Chin; Gupta, Pallav; Basheer, Raid; Yap, Rearon Guey Hwang; advisor Tan, Terrence Huat Hin.

REFERENCES


Title:
Bunch of Wire (BoW) Interchiplet Link Testing and Loopbacks
---
Authors:
Shahab Ardalan, Ayar Labs, CA, USA
Marc Hunter, Teradyne
Bapi Vinnakota, Broadcom, CA, USA
---
Abstract:
Multi-Chiplet Module (MCM or Multi-Chiplet in System-in-Package) designs have recently received a lot of attention as a mechanism to combat high SoC design costs and to economically manufacture large ASICs. The chiplet integration methodology is severely susceptible to the known good die issue where different chiplet can be manufactured by various vendors. Therefore, identifying a known-good-die before integration should be part of multi-chiplet module integration. Furthermore, the chiplet-to-chiplet interfaces (interchiplet communication) also should be an error-free link to guarantee the MCM operation. This is a challenging task where the transmitter and receiver of the link may have been developed by separate companies.

The open-domain specific architecture (ODSA) is a workgroup in the Open Compute Project. The ODSA workgroup created an open interface for interchiplet communication known as Bunch of Wire (BoW). This paper explains the existing different test modes for interchiplet communication in general and in details for BoW standard. The needs for simplex unidirectional as well as simplex bi-directional link in a clock forwarding link architecture will be discussed. In addition, this paper will cover different loopback tests architecture available for mesochronous systems in BoW interface.
3DC-TEST
Nov 6-7, 2020
Bunch of Wire (BoW)
Interchiplet Link Testing & Loopbacks

Shahab Ardalan, Ayar Labs, CA, USA
Marc Hutner, Teradyne, Canada
Bapi Vinnakota, Broadcom, CA, USA
Outline

• Background on ODSA
• BoW definition
• BoW test approach
• Test challenges going forward
ODSA: Accelerators and Chiplets

Domain-specific architectures (DSAs) to accelerate targeted compute-intensive workloads.

Chiplet: Die designed to be used with other die in a package, usually with proprietary interfaces.

DSAs built using chiplets with open standard D2D interfaces.

AI/ML/data workload explosion needs DSAs

IBM Power 9: potential modularity

Jeff Stuechli, Josh Friedrich, IBM – ODSA Workshop, IBM, San Jose, Sep. 2019
For more information

- **ODSA Wiki**: All workshops, weekly calls (open access)
- **Specification proposals**:
  - Bunch of Wires [GitHub repo](https://github.com) (open access)
  - [PIPE adapter](https://github.com), DiPort (open, but need to request access)
  - [ODSA PoC Demo](https://github.com) (open access), [ODSA PoC Implementation Specification](https://github.com) (open, but need to request access)
- **Papers**
  - [ODSA white paper](https://github.com), ODSA Wiki
  - S. Ardalan et al., “Bunch of Wire (BoW) An Open Die-to-Die Interface”, HOT Interconnect 2020
  - D. Jani, “Musings on Domain Specific Accelerators, Open Compute Project and Cambrian Explosion”, LinkedIn
ODSA D2D Interface

Use Case
Shrink board to package
Disaggregate a die

Transport Protocol
- Custom
- PCIe IP (off-the-shelf)
- CXL IP (off-the-shelf)
- CCIX IP (off-the-shelf)

Link Layer Protocol
- PIPE Interface
- Optional Mux
- LPIF Interface
- Optional Mux

PHY Interface
- PHY-specific PIPE Adapter
- PHY-specific LPIF adapter

PHY Adapter
- PCIe
- SerDes
- XSR, USR, ...

PHY Technology
- Optional Mux
- Protocol-integrated Link Layer
- PHY-specific Link Layer
- Optional Mux

Packaging Technology
- Organic substrate
- Interposers/bridges

Block/layer separator
- Std. interface, ODSA-supported
- New ODSA-defined

Switching Protocol
- New LPIF for transport protocol aggregation

Protocol-integrated Link Layer
- Optional Mux
- Cross chiplet switching
- D2D Link Layer
- Intra-Chiplet Buses

More information in the ODSA track at the OCP virtual summit in May 2020
**Bunch of Wires Interface**

- A set of backward compatible die-to-die parallel interfaces that provides the flexibility to trade off Throughput/wire, design complexity, cost, packaging technology.
- Be inexpensive to implement
  - Portable across process nodes ranging from 28nm to 5nm. Portable across multiple bump pitches. Have the Flexibility to support advancing packaging technology. Be unencumbered by technology license costs
- Very low power, < 1 pJ/bit, as defined by Tx IO Pad, wire and Rx IO Pad.
- Very low latency: <5ns without FEC, <15 ns with FEC.
  - Latency as defined from the PCS parallel interface at the source, through Tx interface, channel, Rx interface received at the PCS parallel interface at the receiver.
  - Based on experience, the 5 ns target meets the latency requirements of high-performance applications and has been demonstrated to be achievable.
- Throughput/Chip Edge target range (Rx+Tx):
  - 100Gbps/mm with all packaging options. As a reference example, be able to achieve this goal at a bump pitch of 150um and with a die edge stack depth no greater than 2 routing layer with organic laminate packaging
  - 1Tbps/mm with an advanced packaging option. As a reference example, be able to achieve this goal with a bump pitch of 50 um and with a die edge stack depth no greater than 4 routing layers with advanced packaging.
- Trace length ranges on laminate substrate:
  - Unterminated: <10mm. Terminated: <50mm. Enable higher throughput at very short reach < 1 mm.
- Single supply solution supporting a range of Vdd's compatible with new process technologies.
- Target BER of <1E-15 without FEC and with Optional FEC for ultra-low BER < 1E-25.
A Sample BoW Realization

- This is an example of realization of unidirectional link between Chiplet-A and B
BoW Link

- It is Clock Forwarding Architecture
- Tx is responsible to generate clock and send it to Rx
Based on SI simulation we know the cap will be settled with in 10UI
Suggested a stress pattern + PRBS9 or 31 based on application
Bi-Directional Test

- Short loopbacks mode for simplex bidirectional links
- Internal short loopback is not a full comprehensive test with the absence of real channel
- External short loopback is more comprehensive as the channel exist on ATE load board
Only Rx Chiplet

ATE Load Board

Logic

Golden Ref

Tx

Clock Source

DATA

CK

ATE Instrumentation

Rx

Logic

Chiplet-A

DATA

CK

Rx

Logic

Chiplet-A
• Tx will be tested in read channel with a golden ref Rx at the ATE side.
• Channel on ATE Load board should be match with BoW expected channel
Long Loop Back

- Full long loop back, where both Rx and both Tx are under test with real channel model in interposer
- This is verification setup and no ATE.
Long Loop Back in Shared Clock

- Full long loop back with shared clock between chiplet-A and B
- Generated Clock by chiplet-A is used by chiplet-B as transmitter clock
- Clock domain crossing could be an issue and needs to be considered
Applications for BoW for test?

Consider Mythical chip configuration

- Processor die only have D2D interfaces
- Coverage needed for traditional test cases with limited sacrificial pads
- Want both structural and functional coverage
- Could BoW be repurposed for IEEE1838 FPP?
[CALL to ACTION/HOW to COLLABORATE]

Various ways to be involved in ODSA BoW
• Help define BoW open interface
• Join member companies implementing ODSA PoC
• Join test discussion for BoW
3DC-TEST

Thank You
Designing Testable Systems With Chiplets

Active silicon interposer enables rapid development of SiPs and Chips using chiplets in a modular style but more importantly allow additional abilities for system testability. zGlue’s active silicon interposer, Smart Fabric, is used as a base chip to enable chiplet stacking, system power delivery, connectivity, and other management functions needed for successful productization of chiplet-based systems. One such essential function is the distributed built-in self-test (BIST) function. Distributed BIST enables manufacturing yield enhancement and the use of loose tolerance manufacturing lines resulting in cost optimization. A key aspect of the zGlue interposer is its ability to work with off-the-shelf chiplets in known good die (KGD) and chip scale package (CSP) format without dictating a footprint constraint on chiplets. This is achieved by making a fine pitch bump array on the zGlue interposer. A single ball on the chiplet die comes in contact with multiple bumps on the zGlue interposer, later referred to as a ‘one-to-many’ scheme. This way a spatial over-sampling of the die solder-balls ensures that zGlue interposer can attach to off-the-shelf IO solder ball geometries. A localized switch underneath each zGlue micro bump can then be programmed to connect it to power, ground, and different signal buses. Connection to the RF and sensitive analog signals are handled in RDL. The programmability of Smart Fabric bumps also opens up the possibility of repair after manufacturing. A scheme for electronic realignment makes it possible to compensate for the X, Y, and angular misalignment in the attachment of dies to a certain degree which alleviates one of the key manufacturing challenges. An important feature of the technology is the footprint agnostic assembly of components in a cost-conscious and high-volume compatible manner.

Another key aspect of this built-in self-test scheme augments the system level-test of the final assembly. BIST is implemented as a distributed feature to test open circuit, short circuit, and ohmic values before the system assembly or after the assembly and can take vectors piped in via a probe card or via package pins. Kelvin-probe formation underneath the chiplet IOs can help with failure analyses. Additionally, chiplet IOs can be probed selectively without the need for complicated test hardware. We have used such test mechanisms successfully for debug, bring-up, contact probe testing, as well as system-level structural testing. Additional functional features of Smart Fabric also help in testing of power scenarios, inter-chiplet connectivity, and functionality. With the successful delivery of the zGlue technology, a path to the development of a new area of 3D-IC has opened up. With this modular IC design style, we are able to support an ecosystem and effectively handle high-mix low volume devices that we expect with the growth of connected intelligent devices everywhere.

Reference:

Primary Author: Jawad Nasrullah
Company: zGlue Inc
Job Title: CEO
Address: 2627 Hanover St, Palo Alto, CA 94304
Phone: 6503878873
E-Mail: jawad@ieee.org

Invited By: Bapi Vinnakota
Keyword 1: Test
Keyword 2: SIP
Keyword 3: Chiplets
3DC-TEST
Nov 6-7, 2020
Designing Testable Systems with Chiplets

Jawad Nasrullah, CEO, zGlue Inc
11/6/2020
Outline

• zGlue Integration Platform
• Active Interposer BIST
• Yield and Testing Considerations
• Reliability Testing and Procedures
zGlue—Custom Chips on Demand
zGlue integration Platform

VDD1, VDD2, VDDn
GND
Analog Signal Fabric
Digital Signal Fabric

OTP
Control
Power Supplies
Peripherals

zGlue Smart Fabric
zGlue Smart Fabric Active Si Interposer

Programmable Routing Fabrics

a) Analog
b) Digital

Under Voltage/Over Temp Fault Detection

RGB LED Driver

PWM Controller

32 KB OTP

POReset

OnOff Controller w/ Debounce

Power Management Unit

Fabric BIST

Fabric Controller

SPI & PIC Interfaces

Integrated Voltage Regulators

Battery Regulation

Boost

SYS LDO

System Power

LD01

LD02

LD03

Programmable IO's

Programmable Passives

GPIO Expander

Level Translators
Active Interposer Chiplet BIST Scheme

Built in Testability Functions with a matrix of Cu pillars

1. Open/Short
2. IDDQ
3. Connectivity
4. Programmable Probe
5. Kelvin
6. Pull up/pull down
7. Cap insert
8. Level Shift
9. Programmable IO
10. Chiplet VID

Support for 400um balls to <55um u-bump pitch
zGlue OmniChip Reference Design

SIP for Wearables, Fitness, Bio

CortexM4 MCU + BLE
Temperature Sensing
Vibration, Steps (Accelerometer)
Roll and Pitch (Compass)
Battery Recharging
Heart Rate Sensor
zGlue Chiplet BIST Scheme

SLT: JTAG, SPI, I2C

CP: JTAG

CP: Probe Points
SLT: Wirebond/TSV

Built in Self Test

zGlue Surface Bumps

zGlue Si Interposer Chip

Attached Block Die

I
II
III
IV
V
VI
VII
VIII
IX
zGlue Chiplet BIST and Repair Scheme
Debug Signal Fabric

Observe any signal within the smart fabric at the output using software reconfiguration.
Designing and Manufacturing Steps—Chips / Chiplets

DESIGN STEPS

1- Architecture, IP block selection, Package Selection, Technology Selection
2- Design Entry
   IO Planning
   RTL Design and Verification
   Schematic Capture
   Component Placement
3- Chip Layout / Routing / Package Layout
4- Verification
   Functional
   Electrical
   Thermal
5- Send to Manufacturing

MANUFACTURING STEPS

1- Wafer Fab, Contact Probe (KGD?)
2- Package Fabrication (Optional), e-test
3- uBump/Assembly/CSP/BGA
4- Packaged Part Test
5- ESD and Latchup Tests (sampling)
6- Reliability Tests, HTOL (sampling)
7- Infant Mortality Burn-in
8- Final Test
Designing and Manufacturing Steps—Chiplet Integration

**DESIGN STEPS**

1. Architecture, Chiplet selection, Package Technology Selection

2. Design Entry
   - IO Planning
   - RTL Design and ESL Verification
   - Schematic Capture/Netlist Gen
   - Component Placement

3. Substrate/Interposer Routing and Layout

4. Verification
   - Functional
   - Electrical
   - Thermal

5. Send to Manufacturing

**MANUFACTURING STEPS**

1. Interposer Fab, Contact Probe

2. Substrate Fab, e-test

3. Chiplet Assembly/Packaging

4. Package Test

5. ESD and Latchup Tests (sampling)

6. Reliability Tests, HTOL (sampling)

7. Infant Mortality Burn-in
Testing for Heterogeneous Chips

DFT & Verification:
- Need ESL (transaction level) and IBIS Models for all Chiplets
- Need Mechanical Models for all Chiplets for Thermal/Mech Simulations
- Simple JTAG in each Chiplet

Incoming Material:
- Visual Inspection
- Quality of Chiplets to be guaranteed by the Vendor
- E-test for substrate
- CP test for Si Interposer

After Assembly:
- Inspection (e.g. x-rays)
- System level testing to verify assembly process (JTAG needed)
- Reliability testing
Yield Estimation Example

Incoming DPM in Chiplet C1-C4 = DPM_C < 100
Number of Chiplets = N = 4
Incoming DPM in Si Interposer = DRM_W < 100

Failed Assemblies due to incoming DPM
= N x DPM_C + DPM_W
= 4 x 100 + 100 = 500 (99.95%)  

Control of Incoming DPM_C is the key for Yield.

Assembly Yield numbers can be awesome.

System Level Test is the key to control outgoing DPM For Heterogeneous Chips.
Summary

• zGlue = Glue Chiplets in z Direction and Make Custom Chips with High Reliability.

• zGlue BIST is a key enabling technology for Chiplet Integration and production.

• Observability and Debug of Failures should be carefully planned. Built in Self Test schemes in active Silicon interposers are key enablers.

• IC Package Environmental Tests are critical to work out Chiplet Integration reliability concerns.

• System Level Tests can be the final shipping criterion.

• Beware of over-testing.

• Testing Technology available for licensing

For more info contact jawad@zglue.com
3DC-TEST

Thank You
Abstract submission for IEEE International Workshop on 3D & Chiplet Test

Title:
Universal Chip Telemetry™ (UCT) for quality and reliability monitoring of 2.5D packaging in 5nm and 7nm

Abstract:
High-speed, high-performance ICs rely today, more than ever, on advanced packaging. Driving high-performance computing applications, heterogeneous integration technologies, including 2.5D/3D stacking, tiling, intra-die routing, through-silicon vias, chiplets and fan-out, have and significantly increased system capabilities and performance. These in turn, have introduced new quality and reliability challenges due to an inherent high density and high frequency nature, with limited visibility after assembly. In this paper, a new deep data approach based on Universal Chip Telemetry™ (UCT) will be introduced. By applying advanced analytics to measurements extracted from on-chip monitoring IPs, high visibility is gained, at test and in mission. The silicon-proven technology provides signal integrity monitoring solutions for fault detection and repair, in mission-mode. Findings from Global Unichip Corp. (GUC) 5nm and 7nm High Bandwidth Memory (HBM2E, 3.2Gbps) testchips will be presented.

Contact:
Tamar Naishlos
Director of Marketing
tamarn@proteanTecs.com
Universal Chip Telemetry™ (UCT) for Quality and Reliability Monitoring of 2.5D Packaging in 5nm and 7nm

Nir Sever, Sr. Director of Product
November 6, 2020
Risks in D2D Connectivity

Challenges in CoWoS
- Single u-bump per signal
- Quality issues can be related to:
  - u-bump cracks
  - Assembly defects

QnR challenges
- Lack of redundancy
- High density
- High speed
- Lack of visibility for latent defects

Challenges in InFO
- Single trace per signal
- Quality issues can be related to:
  - Open & short
  - Bridge-short (signal to signal or to supply)

Risk of expensive
- Quality escapes
- In-field degradation

Interconnect quality issues may lead to full system failure

Image source: TSMC
Universal Chip Telemetry™ (UCT) for Visibility

- High resolution and widespread UCT Agents at no cost to area
- Mimic the design and monitor key parameters
- Operate in-situ and in mission-mode
- Readouts extracted using industry standard methods and formats
Proteus™ for Actionable Insights

Insights

- Performance measurement per pin
- Eye diagram per pin
- Rx signal integrity
- Lane integrity (organic substrate)
- u-bump integrity (silicon interposer)
Proletus™ for D2D Connectivity

• General Purpose
  – InFO™, CoWoS™, 3DFabric™, Glink™, HBM3 and counting…
• Based on ongoing tracking of the signal timing at the receiver (margin to failure)
  – Per lane, in mission mode, and no impact on signal
• Wide speed range: 2GHz (4Gb/s DDR) to 8GHz (16Gb/s DDR)
• Full eye visibility for DDR signals:
  – Setup and Hold to positive and negative edges of reference clock
  – Keep track of min, max and average (“Jitter”)
• Same Agent, all signal types:
  – Rx only, Tx only and Bidirectional
  – Single-ended and Differential
  – Single side monitoring (e.g. HBM3) or both
Complete System Level Solution

• IO Sensor:
  – Area efficient, operates on VDD core; per process node Hard IP
  – Designed for integration inside PHY hard macro
  – One IO Sensor per pin

• Agent controllers:
  – Synthesizable RTL for simple IC integration
  – APB standard bus interface to Host CPU
  – JTAG and I2C interfaces for external control
Near End and Far End Monitoring

Operates in Mission Mode (no special test mode required)

Δ = NE Pulse Delay

Δ = FE rising edge delay
Providing Visibility of Lane Integrity

- Signal integrity grading
- Performance grading
Silicon Proven: Lane Grading

Per pin signal quality map*:

Measured in GUC’s 5nm HBM2E 3.2 Gbps Test Chip

* CoWoS lines were intentionally routed beyond their spec limits
High Coverage NPI

- Go / No Go
- Worst case margin per group
- Low coverage, low volume
- Time consuming
- Labor intensive

Low visibility → Low confidence

- Parametric
- Full eye visibility per pin
- All pin coverage
- Fast, immediate, during standard test
- Analytics at the click of a button

High visibility → High confidence
Silicon Proven: Outlier Detection*

- Lane degradation monitoring and repair at test
- Based on Rx slew rate

Far-End Integrity Insight per Channel and Pins in Channel

Rx Signal Amplitude at Pin

Measured in GUC’s 7nm and 5nm HBM2E 3.2 Gbps Test Chips

* CoWoS lines were intentionally routed beyond their spec limits

© proteanTecs 2020. All rights reserved.
In-field **Degradation Monitoring & Predictive Maintenance**

- Predictive maintenance
- Alerts on faults before failures
- In mission mode
Silicon Proven: Degradation Monitoring

- Lane degradation monitoring and repair, in mission-mode*
- Based on near-end and far-end integrity insights

Measured in GUC’s 7nm and 5nm HBM2E 3.2 Gbps Test Chips

* ASIC buffer strength intentionally weakened to emulate u-bump resistance change
Thank you.
3DC-TEST
Nov 6–7, 2020
System Level Test for 3D- SIP

Sajjad Pagarkar & Sandeep Bhatia
DPPM Challenges
There are 3 major DFT challenges in 3D/5.5D SIPs to achieve low DPPM

➢ Inadequate test coverage of individual chiplets or monolithic dice, specially MSA & PHY circuits
➢ Lack of coverage of die-to-die interconnects
➢ Inability to test for subtle defects arising out intra die process skews among various chiplets, weak memory cells or walking wounded defects that may turn into hard defect with usage.

Potential Solution
While the DFT community continues to explore various ways of addressing above challenges, many companies are deploying SLT in production to supplement ATE screen testing to arrest residual DPPM
Marginal Defects are reaching a tipping point therefore for complex SOC the test strategy applying ATPG based and limited functional test on ATE is not sufficient to meet the quality expectations.

- 100% ATPG coverage does not translate to 100% defects coverage (due to limitations of fault models) e.g even advance fault models like Cell aware, SDD and PDF may NOT detect subtle FEOL or BEOL defects or marginal defects or defects sensitive to PVT combination
- 100% of the circuits are not covered by SCAN, Some of the circuits covered by SCAN are left out because of exceptions, hence collectively > 1% or ~300M transistors and associated BEOL are not tested at all.

- Developing Concurrent ATE test patterns for all combinations of scenarios is impractical
- 3D/5.5D Packaging complexity is driving new fault mechanisms.
  - Integrating multiple IPs with 100% yield does not give 100% yield
  - Traditional DFT tests do not cover (target) chip-to-chip or chip-to-chiplet interconnects. There are subtle AC interconnect (package) defects that get sensitised only when those nodes are run at speed and/or at mission mode PDN/Thermal condition
- Process Corner of each chip/chiplet on SIP is different (Details)
- Latent defects and unscreened walking wounded parts fail in the field (infant mortality or early life failures) within first 90 days.
What is SLT for Chip manufacturing screen

- **Enhance ATE functional Tests**
  - Several Native mode Functional test patterns can be generated using on-board chip manager CPU and internal SRAM memories & by building BIST engines to execute end to end sub system tests.
  - Also Classic DV test cases can be converted to ATE functional tests.

**Pros**
- No extra capital needed for additional insertion SLT system

**Cons**
- ATE test time is significantly more expensive than SLT, which limits functional test time to few seconds vs. minutes needed to expose critical paths by running mission mode workloads.
  - By virtue of its architecture it’s not feasible to build essential system peripherals on ATE, that are required to for running mission mode test cases.

- **Supplement ATE manufacturing screens with additional insertion of SLT screen using a dedicated system.**
  - HOT SLT of 1-2 hr/part also serves the purpose of Production burn-in.
  - SLT enables functional or BIST oriented regression tests for subsystems and individual cores.
  - Through IEEE1149.10 it enables execution of Advance Fault Models for logic and multiple-granular-memory-algorithms, at significantly lower COST compared to ATE.
System Level Test (SLT)

- Supervisory Processor (w PCIE, Ethernet, JTAG, ...)
- Thermal Controller
- Power Supplies

Diagram:
- SLT Front end
  - Management Engine
    - JTAG
  - FPGA
    - Socketed board
  - PCIe Test Interface
  - Test Generator
  - Other Comp.
What can we do improve DPPM

- Devise Better fault models with reasonable test time and vector depth to make it ATE worthy
- Consolidation of various IEEE standards & JEDEC standards evolved during last few years.
- EDA tools for IEEE1149.10/IEEE1500/IEEE1838 implementation to facilitate ATPG/MEMBIST test execution on SLT though HSIO (USB, PCIe etc.)
- EDA tools for quick implementation of Design for Systems test
  - Standardize Non-SCAN DFT in terms of Functional BIST engines for subsystems consisting of multiple chiplets & memory e.g. CPU subsystem, GPU subsystem, HBM subsystem, SERDES etc
- DFS (Design For SLT) - facilitate on-package system and/or sub-system level test without HLOS or SW stack.
3DC-TEST
Thank You
Individual IPs have unique process variation.
- It is impossible to pick and choose optimum process corners for a set of IP’s on a MCM.
- Thermal profile of each IP further complicates matching across PVT.
- We can strive to limit process spread of an individual IP with manufacturing screen Test Limits derived through detailed ATE characterization, Bench characterization and System correlation at SLT & BFT (board functional test).
- We need to build testability and controllability on individual IP to be able to screen them as desired.
- Work with foundry/s to set stringent test limits upstream towards Wafer Sort (WS) to reduce DPPM risk and scrap rates at FT.
Why do we need Chip level Burn-in as part of manufacturing screen
Traditional low speed BI System running ATPG/MBIST at 1-5MHz @125C, requires minimum 48Hrs to run 3-6 months lifetime v/s 1 to 2 hrs SLT at HOT traps majority of the ELF to minimize outgoing latent DPPM.
Session 7: Panel

Test Challenges in the New 3D and Chiplet World

- **Moderator:** Jan Vardaman – President – TechSearch International (USA)
- **Panelists:**
  - Dave Armstrong – Director of Business Development – Advantest (US)
  - Paul Franzon – Professor, Director of Graduate Programs – NCSU (US)
  - Bob Patti – President – NHanced Semiconductors (US)
  - John Yi – PMTS Product/Test Engineering – AMD (US)
Test Challenges in the New 3D and Chiplet World

E. Jan Vardaman
Moderator

RELEVANT, ACCURATE, TIMELY
Panel Members

- Dave Armstrong, Advantest
- John Yi, AMD
- Gerard John, Amkor Technology
- Bob Patti, NHanced Semiconductors
- Paul Franzon, North Carolina State University
Panel Questions

• What is different between 3D-stacked ICs and chiplet-based ICs?
  – Are there differences in technology, design, manufacturing flow, and especially test: Test flows, DfT, probing etc.?
  – What test/DfT technology do we need for chiplet-based ICs that we did not need for “conventional” 3D-ICs?

• What new test equipment is required for this new era of chiplets and 3D (or is what we have adequate)?

• Do we need Known Good Die or Chiplets or is Probably Good Die Sufficient? What level of testing is sufficient? Test interfaces? System test?

• Several companies indicate they are using AI-solutions to help with their test strategy, what solutions are you aware of?

• Are new probe methods needed?

• What is the role of inspection and what kind of inspection technology is required?
Chiplet Testing Issues

Paul Franzon,
Cirrus Logic Distinguished Professor,
Department of Electrical and Computer Engineering,
NC State University

paulf@ncsu.edu
Chiplets

“Chiplets”

- RISC V controller Chiplet
- Sparse CNN Chiplet
- Long Short Term Memory (LSTM) /MLP Chiplet
- Bitonic Sort Accelerator

Standalone

- HTM Accelerator

3DIC – 2 chip stack (GF 28)
Area Overhead

Chiplet IO take considerable area (RISC V core)

(3.2 x 1.6 mm)
# Multiple IO standards

<table>
<thead>
<tr>
<th>Standard</th>
<th>Source</th>
<th>Bandwidth Density</th>
<th>Throughput / lane</th>
<th>Latency</th>
<th>Energy/bit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced Interface Bus (AIB)</td>
<td>Intel (open)</td>
<td></td>
<td>Up to 2 Gbps</td>
<td>4 cycles</td>
<td>0.85 pJ</td>
</tr>
<tr>
<td>Multi-Die I/O (MDIO)</td>
<td>Intel</td>
<td></td>
<td>Up to 5.4 Gbps</td>
<td></td>
<td>0.5 pJ</td>
</tr>
<tr>
<td>High Bandwidth Memory (HBM2)</td>
<td>JEDEC</td>
<td></td>
<td>2.4 Gbps (2E)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XSR/USR</td>
<td>Rambus / OIF</td>
<td></td>
<td>112 Gbps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lipincon</td>
<td>TSMC</td>
<td>1.2 Tbps/mm²</td>
<td>8 Gbps</td>
<td></td>
<td>0.062 pJ</td>
</tr>
<tr>
<td>MoChi</td>
<td>Marvell</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bunch of Wires (BoW)</td>
<td>OCP / ODSA</td>
<td>1 Tbps/mm</td>
<td>56 Gbps (bidi)</td>
<td></td>
<td>0.7 pJ/bit</td>
</tr>
<tr>
<td>Bandwidth Engine</td>
<td>Mosys Inc.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMD</td>
<td>Zen2</td>
<td></td>
<td>10.6 Gbps</td>
<td></td>
<td>2 pJ/bit</td>
</tr>
</tbody>
</table>
Area Efficient DFT

• IO DFT
  – Many IO schemes implement redundancy so quality DFT leads to yield enhancement
  – IO lanes are tight: 40 um pitch today, 10 um tomorrow (?)
  – Need: Pitch-Efficient loopback, etc., testing

• Chiplet JTAG
  – We included a JTAG controller in each chiplet
    • Considerable additional area expense
  – Need a “lite” JTAG for chiplets?
How Much DFT?

• As my doctor tells me, only order a test if it leads to a decision
  – Applies to 3D and Chiplets too
  – E.g. Does the channel have redundancy?
Acknowledgements

Team members: Lee Baker, Sumon Dey, Weifu Li, Steve Lipa, Teddie Nigussie, Tse-Han Pan, Shep Pitts, Josh Schabel, Josh Stevens

Funding: DARPA, Google, USAF

paulf@ncsu.edu
Known Not Bad Die Success: Repair, Redundancy, and Pragmatism

September 17, 2020
How to Solve KGD for Advanced Packaging?

→Design For Repair!

What’s so Funny about Science? By Sidney Harris (1977)
HOW DID WE GET HERE??
End of Moore’s Law

Maintaining planar evolution so far... But, Scaling is getting difficult
- Sub-1nm hitting the limit of cell reliability → Enterprise?
- Tremendous investment cost required to continue → Consumer?

Shrinking chips
Number and length of transistors bought per $1
Transistor size, nanometers (nm)

Cost of Patterning

End of Growth of Single Program Speed?

40 years of Processor Performance

CISC 2X / 3.5 yrs (22%/yr)
RISC 2X / 1.5 yrs (52%/yr)
Multicore 2X / 3.5 yrs (23%/yr)
End of Dennard Scaling
End of the Line? 2X / 20 yrs (12%/yr)
Amdahl’s Law 2X / 6 yrs (12%/yr)

Apple A12 single thread performance (RISC ISA) = x86 Skylake single thread perf (SPEC), at much lower power, Anandtech 10/8/18

Based on SPECintCPU. Source: John Hennessy and David Patterson, Computer Architecture: A Quantitative Approach, 6th ed. 2018
Internet Of Things
More Than Moore

1. THIN 2D AND 3D ACTIVES
   - OPTO SOP
   - LASER
   - PHOTODETECTOR
   - WAVEGUIDE
   - NANOMAGNETICS

2. THIN FILM PASSIVES
   - CHIP-CAST EMBEDDED IC
   - DIGITAL SOP
   - THERMAL SOP
   - EBG & ISOLATION

3. SYSTEM INTERCONNECTIONS
   - ANALOG & RF SOP
   - ANTENNAS & MILLIMETERS
   - MEMS PACKAGING
   - MEMS

4. THERMAL INTERFACES AND STRUCTURES
   - GaAs RFIC
   - BIO-SENSOR
   - POWER & BATTERIES

GEORGIA TECH PRC

5. MULTI-FUNCTION MATERIALS

6. MIXED SIGNAL DESIGN AND TEST
   - 3D ICs

7. MECHANICAL DESIGN AND RELIABILITY

8. POWER SOURCES

---

MEMS
- Resonator
- Accelerometer
- MEMS/Cavity
- Mirror Arrays
- Cantilever
- Resonator
- Wafer on Silicon
- Chromatography
- Deep Trench Insulator
- Poly-Glass Trench

Microfluidics

High Voltage

3D IC
- Transistor
- Memory

NHANCED SEMICONDUCTORS
WHAT IS ADVANCED PACKAGING?
Span of Advanced Packaging

Packaging

3D-ICs
100-1,000,000/sqmm
1000-10M Interconnects/device

Peripheral I/O
- Flash, DRAM
- CMOS Sensors

1s/sqmm

Transistor to Transistor
- Ultimate goal

Wafer Fab

IBM/Samsung

100,000,000s/sqmm

IBM
Many Choices!

IME A-Star / Tezzaron Collaboration

Active Silicon Circuit Board

2 Layer Processor

3 Layer 3D Memory

Die to Wafer Cu Thermal Diffusion Bond

μBumps

C4 Bumps

Solder Bumps

Organic Substrate

level #0

level #1

level #2

level #3

level #4
Si Interposers

Bigger, Better, Faster
>50x50mm, Up to 6 layers, Lower R,C
8 Layer Logic Stack

- Wafer 8
- Wafer 7
- Wafer 6
- Wafer 5
- Wafer 4
- Wafer 3
- Wafer 2
- Wafer 1

Copper interconnect layers
SuperContacts
Supporting substrate

EM-4800 5.0kV 8.7mm x500 SE(M) 7/28/2015
HANCED SEMICONDUCTORS
5.5D Systems
System Densification

- Integrated Photonics
- 2.5D
- 3D
- Integrated power
- Integrated passives
SOUNDS GREAT!
WHAT’S THE CATCH?
Choices

• Wafer-to-wafer / Monolithic 3D
  – Best cost structure
  – Highest density interconnect
  – Fab processes
    • A messy fab issues
      – Particles
      – Materials
      – Non-standard sizes
      – Novel materials
      – Novel processes
• Interposers
  – Mixed fab and packaging flow
  – Add TSVs
• Chip Stacking - POP
  – Limited interconnect
  – Cost
Materials Opportunities

• Silicon Interposers
  – 2-3um L/S/D
  – Rs and Cs
  – Active is the future
  – Handling & handoff

• Organics Interposers
  – 5-6 um
    • Litho limits
    • Material planarity limits
  – Great cost structure
  – CTE Challenges
  – Large substrate

• Glass Interposers
  – Large substrate
Mixing Fab, Packaging and Assembly

Foundry \(\rightarrow\) ? \(\rightarrow\) Packaging \(\rightarrow\) ? \(\rightarrow\) Assembly

\(?\) Test what where when ?

Big hidden cost

Customer
Testing

• Significant planning required
• Careful analysis of yield cost
• New methodologies
  – High I/O count requires self-test
  – Deep embedding requires more effort for visibility
  – At speed test alternatives
• Embedding memory has numerous test issues
  – Standard test interface required.
• Self-repair / Self-redundancy
Data Points

• The future is chiplets… or at least really sophisticated multi-die packaging
  – Highly customized assembly flow
  – Provides:
    • Product Flexibility
    • Faster time to market
    • Reuse
    • Enables cost effective low volume production

• We don’t build enough of a given module type to get statistical reliability data
• We can’t inspect the latest generation of assembly technologies for defects
• KGD really is KNB – Known Not Bad
  – Probably as good as it gets
• 2.5/3D solutions have lots of I/O
  – Hard (many times impossible) to test
  – Costly to test
• Die probing is more difficult than wafer or package testing
• Probing causes damage
  – Can have worse effects than simply doing blind assembly
“Dis-Integrated” 3D Memory

- **DRAM layers**
  - 4xnm node

- **Controller layer**
  - Contains: sense amps, CAMs, row/column decodes and test engines. 40nm node

- **I/O layer**
  - Contains: I/O, interface logic and R&R control CPU. 65nm node

- 2 million vertical connections per layer per die

Better yielding than 2D equivalent!
Bi-STAR Repair Improves Yield
3D-Routing Node (NOC)

Leverage system level like redundancy schemes
3D NOC Interconnect
3D NOC Interconnect
Extensions of R&R

- **Spare Processors**
  - Virtually all advanced processors today

- **Smart Interposers**
  - Programmable routing
  - Intelligent power control

- **FPGA Repair Kits**
  - A logical extension of current chip repair kits

- **Redundant I/O**
  - Like HBM devices
RELIABILITY
How Do We Know If This Device Is Reliable?

• Hypothesis:
  – If we can measure the “Quality” of the assembly, we can infer the Reliability of the specific device tested.
A Plan

• Create universal test structures that are accessed via JTAG 1394 (IEEE1500).
• Measure R’s and C’s of alignment structures and interconnects using 1149.4 analog JTAG extension.
  – Electronic Verniers
  – Via Chains
  – Temperature sensors
  – PCM data
• Create ala carte test plans that bracket what tests are required based on the module content and assembly technologies employed
• Build a database of historical evidence to correlate actual reliability to measured “Quality” data
  – Starting with test devices that are built to validate the premise
One Slide:

Images from: A DFT Architecture for 3D-SICs Based on a Standardizable Die Wrapper; Erik Jan Marinissen et al

Physical

Augmented JTAG based on IEEE 1500: Add alignment sensing, 3D interconnect R/C measurement, power, temperature …

System level test, configuration, repair and validation

Objective is to “prove” specific device quality and improve reliability data.

Logical

Use Standardized DFT + 2.5/3D PCM methods to test Quality and derive Reliability

Use repair and redundancy to create KGD and obtain yield.

IEEE 1500 is well defined 2.5/3D DFT starting point building on 1394 standard. Plan is to add 1149.4 analog features targeting device manufacturing integrity.
embracing a better life