Using Repair & Redundancy with KGD to Produce Cost Effective 2.5 and 3D Devices

Bob Patti, CTO

rpatti@tezzaron.com
How to Solve KGD for 3D?

Design For Repair!

What’s so Funny about Science? By Sidney Harris (1977)
Span of 3D Integration

Tezzaron 3D-ICs
100-1,000,000/sqmm
1000-10M Interconnects/device

1s/sqmm
Peripheral I/O
- Flash, DRAM
- CMOS Sensors

100,000,000s/sqmm
Transistor to Transistor
- Ultimate goal

Packaging

Wafer Fab

Tezzaron Semiconductor
Tezzaron Semiconductor

TSV Pitch ≠ Area ÷ Number of TSVs

- TSV pitch issue example
  - 1024 bit busses require a lot of space with larger TSVs
  - They connect to the heart and most dense area of processing elements
  - The 45nm bus pitch is ~100nm; TSV pitch is >100x greater
  - The big TSV pitch means TOF errors and at least 3 repeater stages

![Diagram of TSV pitch and area relationship](image)

10um TSV
20um Pitch

1024 bit bus
Single layer interconnect

Tezzaron Semiconductor
TSV Pitch ≠ Area ÷ Number of TSVs

- TSV pitch issue example
  - 1024 bit busses require a lot of space with larger TSVs
  - They connect to the heart and most dense area of processing elements
  - The 45nm bus pitch is ~100nm; TSV pitch is >100x greater
  - The big TSV pitch means TOF errors and at least 3 repeater stages
### 3D Interconnect Characteristics

<table>
<thead>
<tr>
<th>Material</th>
<th>SuperContact™ I 200mm Via First, FEOL</th>
<th>SuperContact™ III 200mm Via First, FEOL</th>
<th>SuperContact™ IV 200mm Via First, FEOL</th>
<th>Interposer TSV</th>
<th>Bond Points</th>
<th>Die to Wafer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size L X W X D W in Bulk</td>
<td>1.2 μ X 1.2 μ X 6.0μ</td>
<td>0.85 μ X 0.85 μ X 10μ</td>
<td>0.60 μ X 0.60 μ X 2μ</td>
<td>10 μ X 10 μ X 100 μ Cu</td>
<td>1.7 μ X 1.7 μ Cu</td>
<td>3 μ X 3 μ Cu</td>
</tr>
<tr>
<td>Minimum Pitch</td>
<td>&lt;2.5 μ</td>
<td>1.75 μ</td>
<td>1.2 μ</td>
<td>30/120 μ</td>
<td>2.4 μ</td>
<td>5 μ</td>
</tr>
<tr>
<td>Feedthrough Capacitance</td>
<td>2-3fF</td>
<td>3fF</td>
<td>0.2fF</td>
<td>250fF</td>
<td>&lt;&lt;</td>
<td>&lt;25fF</td>
</tr>
<tr>
<td>Series Resistance</td>
<td>&lt;1.5 Ω</td>
<td>&lt;3 Ω</td>
<td>&lt;1.75 Ω</td>
<td>&lt;0.5 Ω</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
</tbody>
</table>

**Small fine grain TSVs are fundamental to 3D enablement**
A Closer Look at Wafer-Level Stacking

Dielectric(SiO2/SiN)
Gate Poly
STI (Shallow Trench Isolation)
W (Tungsten contact & via)
Al (M1 – M5)
Cu (M6, Top Metal)

“Super-Contact”
Next, Stack a Second Wafer & Thin:
Stacking Process Sequential Picture

Two wafer Align & Bond → Course Grinded → Fine Grinded

→ After CMP → Si Recessed

High Precision Alignment

Misalign=0.3um

Top wafer

Bottom wafer
Then, Stack a Third Wafer:

1st wafer: controller

2nd wafer

3rd wafer

Tezzaron Semiconductor
Finally, Flip, Thin & Pad Out:

This is the completed stack!
3rd Si thinned to 5.5um
2nd Si thinned to 5.5um
SiO₂
1st Si bottom supporting wafer
“Dis-Integrated” 3D Memory

Memory Layers from DRAM fab

Controller Layer from high speed logic fab

2 million vertical connections per lay per die
Gen2 Octopus Stack

2 Layer Stacked Device (SEM)

DRAM Control/Logic

DRAM Memory Cells
Gen4 “Dis-Integrated” 3D Memory

2 million vertical connections per layer per die

I/O layer contains: I/O, interface logic and R&R control CPU. 65nm node

Better yielding than 2D equivalent!

Controller layer contains: senseamps, CAMs, row/column decodes and test engines. 40nm node

DRAM layers 42nm node
3D DRAMs

Octopus I

- 1-4Gb
- 16 Ports x 128bits (each way)
- @1GHz
  - CWL=0 CRL=2 SDR format
  - 5ns closed page access to first data (aligned)
  - 12ns full cycle memory time
  - 288GB/s data transfer rate
- Max clk=1.6GHz
- Internally ECC protected, Dynamic self-repair, Post attach repair
- 115C die full function operating temperature
- JTAG/Mailbox test&configuration

Octopus II

- 4-64Gb
- 64-256 Ports x 64bits (each way)
- @1GHz
  - 5-7ns closed page access to first data (aligned)
  - 12ns full cycle memory time
  - 2TB/s data transfer rate
The 3D Problem

Stacking reduces yields:
Industry average = ±50% yield on memory wafers
Stacking compounds the yield problem:
Yield of an n-layer stack = (0.5)^n
Impact:
For a 4-layer (stacked) chip:
\[
\frac{1}{(4)^2} = 6\% \text{ yield}
\]

6% yield is not economically viable
BiSTAR Architecture

• Intelligent self repair and super robust factory testing
• 256 BIST sequencers run independently in parallel
• Each sequencer reads/writes to 8 arrays at once
• Any single error trips a global error for that BIST sequencer
• BIST sequencer uses high-speed isolation algorithm to detect where error occurred
• Repeated errors can be masked out to speed up testing
• Soft redundancy mapping
• Post assembly and field repair
What Can Bi-STAR™ Test & Repair?

- Bad memory cells
- Bad line drivers
- Bad sense amps
- Shorted word lines
- Shorted bitlines
- Leaky bits
- Bad secondary bus drivers
- Bad CAMS
The “Killer” App: Split-Die

Tezzaron 3D DRAM

- Embedded Performance with far superior cost/density.
- 110nm DRAM node has better density than 45nm embedded DRAM.
- 1000x reduction in I/O power.

Proven Technology!
Logic on Memory

Memory

92 pads
(528 total pads at edge, stagger 250um pad, 125um pitch ~1500 available pads)

8 DRAM ports
16x21 pad array

>10μf bypass caps
SS ~4,000pf
Logic on Memory

Memory

172 pads

92 pads
(528 total pads at edge, stagger 250um pad, 125um pitch ~1500 available pads)

199 I/O Bondpoints/side

8 DRAM ports
16x21 pad array

Memory also acts as interposer

>10μf bypass caps
SS ~4,000pf
OC768 Packet Engine

<table>
<thead>
<tr>
<th>Dual PPC 64x ARM SOC</th>
<th>FPGA (CPU Augmentation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRAM</td>
<td></td>
</tr>
<tr>
<td>Flash</td>
<td></td>
</tr>
<tr>
<td>CAM</td>
<td></td>
</tr>
<tr>
<td>CAM</td>
<td></td>
</tr>
<tr>
<td>FPGA (Packet Cracker)</td>
<td></td>
</tr>
<tr>
<td>Stack Controller</td>
<td></td>
</tr>
</tbody>
</table>
3D-Routing Node (NOC)

Leverage system level like redundancy schemes
3D NOC Interconnect
3D NOC Interconnect
Data Paths: On-ramp/Off-ramp
130nm Implemented Node
Fault Tolerant
Self-configuring/Re-configuring

Stack Manager
Processor0

TBS

TBU

TBC

Stack Manager
Processor1

TBS

TBU

TBC

routing nodes

routing nodes
2.5/3D Circuits

IME A-Star / Tezzaron Collaboration

3 Layer 3D Memory

FPGA (4Xnm)

2 Layer Processor

Active Silicon Circuit Board

Organic Substrate

Solder Bumps

μBumps

C4 Bumps

Die to Wafer Cu Thermal Diffusion Bond

level#0

level#1

level#2

level#3

level#4

IME A-Star / Tezzaron Collaboration

Tezzaron Semiconductor 10/23/2012
Memory Module

64Gb DRAM
64Gb DRAM
64Gb DRAM
64Gb DRAM

HUB

1024 bits each way @ 2GT/s

SERDES 28Gb/s

64Gb DRAM spare
Physical Module

Tezzaron Semiconductor
2020 “Processing Cube” Stack Design

Complete system
20TFlops
200G packets/s
16 x 10Tb links
~1200W
• “Volume” 2.5D and 3D Manufacturing in 2013
• Interposers
• Future interposers with
  – High K Caps
  – Photonics
  – Passives
  – Power transistors
• Cu-Cu, DBI, Oxide, Au/In 3D assembly
3D Inflection Point

- Tezzaron
- With Novati

The Effect of 3D on Devices

- Transistors
- Transistors with 3D
- Frequency
- Frequency with 3D
- HPC Frequency
Summary

• 3D has numerous and vast opportunities!!
  – Best of class integration of
    • Memory
    • Logic
    • RF
    • MEMS
  – New design approaches
    • Design For Repair (DFR)
  – DFT is not “just a good idea” anymore
    • Targeted DFT: Concentrate on single point failures

Sensors
Computing
MEMS
Communications