3D Integration:

New Opportunities for Speed, Power and Performance

Robert Patti, CTO

rpatti@tezzaron.com
Why We Scale?

Speed | Power | Cost | Size

>180nm 130nm 90nm 65nm 45nm 28nm 22nm 16nm

Advantages

COST PER GATE REDUCTION TRENDS

What can 3D do for us?

SPEED / PERFORMANCE ISSUE  The Technical Problem
# 3D Stacking Approaches

<table>
<thead>
<tr>
<th>Chip Level</th>
<th>Device Level</th>
<th>Wafer Level</th>
</tr>
</thead>
</table>
| • Ziptronix  
  • Xan3D  
  • Vertical Circuits  
  Amkor : 4S CSP (MCP) | • Stanford  
  • Besang | • Infineon/IBM  
  • RPI  
  • ZyCube |

<table>
<thead>
<tr>
<th>Irvine Sensors : Stacked Flash</th>
<th>Matrix: Vertical TFT</th>
<th>Tezzaron</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Irvine Sensors: Stacked Flash" /></td>
<td><img src="image2" alt="Matrix: Vertical TFT" /></td>
<td><img src="image3" alt="Tezzaron" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Samsung : Stacked Flash</th>
<th><img src="image4" alt="Samsung: Stacked Flash" /></th>
<th><img src="image5" alt="Tezzaron: Stacked Flash" /></th>
</tr>
</thead>
</table>

**Tezzaron Semiconductor**

02/07/2012
# Wafer Level Stacking Approaches

<table>
<thead>
<tr>
<th>Infineon/IBM</th>
<th>RPI/Ziptronix/ZyCube</th>
<th>Tezzaron</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Infineon</strong>: W deep via</td>
<td><strong>RPI</strong>: Dielectric bonding</td>
<td><strong>Tezzaron</strong>: Copper bonding</td>
</tr>
<tr>
<td><strong>IBM</strong>: SOI wafer thinning</td>
<td><strong>Ziptronix</strong>: Covalent bond</td>
<td>Backside of the stacked wafer</td>
</tr>
<tr>
<td><strong>ZyCube</strong>: Injection glue bonding</td>
<td></td>
<td>3 wafer stack</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Tezzaron</strong>: 3D Sensor</td>
</tr>
</tbody>
</table>
## Market Drivers: 3D

<table>
<thead>
<tr>
<th>Driver</th>
<th>Functionality</th>
<th>Technical Parameter # 1</th>
<th>Technical Parameter # 2</th>
<th>Value Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stacked NAND Flash</td>
<td>Cell Phones</td>
<td>Memory density</td>
<td></td>
<td>High packing density</td>
</tr>
<tr>
<td>Hard Drives</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flash Drives</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Micro-processor + Memory</td>
<td>Workstations</td>
<td>Latency bandwidth</td>
<td>Power</td>
<td>Execution time</td>
</tr>
<tr>
<td>Memory</td>
<td>Multiple</td>
<td>Density</td>
<td>Latency</td>
<td>Varies</td>
</tr>
<tr>
<td>Image Sensor</td>
<td>Cell Phones</td>
<td>Quantum Efficiency</td>
<td>Number of pixels</td>
<td>Image Quality</td>
</tr>
<tr>
<td>Automotive</td>
<td>Camera</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
There is a lot of 3D today

Samsung
16Gb NAND flash (2Gx8 chips), Wide Bus DRAM

Micron
Wide Bus DRAM

Intel
CPU + memory

OKI
CMOS Sensor

Xilinx
4 die 65nm interposer

Raytheon/Ziptronix
PIN Detector Device

IBM
RF Silicon Circuit Board / TSV Logic & Analog

Toshiba
3D NAND
Through-Silicon Via (TSV)

- Via First
- Via Last
- Via at Front end (FEOL)
- Via at Mid line (MOL?)
- Via at Back end (BEOL)

“SuperContact”

Dr. J.Q. Lu
RPI
Span of 3D Integration

Packaging

3D Through Via Chip Stack

IBM/Samsung

Wafer Fab

3D-ICs

100-1,000,000/sqmm

1000-10M Interconnects/device

IBM

100,000,000s/sqmm

Transistor to Transistor

- Ultimate goal

1s/sqmm

Peripheral I/O

- Flash, DRAM

- CMOS Sensors

Peripheral I/O

- Flash, DRAM

- CMOS Sensors
# 3D Interconnect Characteristics

<table>
<thead>
<tr>
<th></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><strong>Size</strong></td>
<td>L X W X D Material</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.2 μ X 1.2 μ X 6.0μ W in Bulk</td>
<td>0.85 μ X 0.85 μ X 10μ W in Bulk</td>
<td>0.60 μ X 0.60 μ X 2μ W in SOI</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><strong>Minimum Pitch</strong></td>
<td>&lt;2.5 μ</td>
<td>1.75 μ</td>
<td>0.8 μ</td>
<td>30/120 μ</td>
<td>2.4 μ</td>
<td>5 μ</td>
</tr>
<tr>
<td><strong>Feedthrough Capacitance</strong></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><strong>Series Resistance</strong></td>
<td>&lt;1.5 Ω</td>
<td>&lt;3 Ω</td>
<td>&lt;1.5 Ω</td>
<td>&lt;0.5 Ω</td>
<td>&lt;</td>
<td>&lt;</td>
</tr>
</tbody>
</table>
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

---

10um TSV
20um Pitch

1um TSV
2um Pitch

1024 bit bus
Single layer interconnect
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:
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
## DRAM wants 2 different processes!

<table>
<thead>
<tr>
<th>Bit cells</th>
<th>Low leakage</th>
<th>High Vt Devices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-slow refresh</td>
<td>Vneg Well</td>
</tr>
<tr>
<td></td>
<td>-low power</td>
<td>Thick Oxide</td>
</tr>
<tr>
<td></td>
<td>-low GIDL</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sense Amps</th>
<th>High speed</th>
<th>Low Vt Devices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word line drivers</td>
<td>-better sensitivity</td>
<td>Copper interconnect</td>
</tr>
<tr>
<td>Device I/O</td>
<td>-better bandwidth</td>
<td>Thin Oxides</td>
</tr>
<tr>
<td></td>
<td>-lower voltage</td>
<td></td>
</tr>
</tbody>
</table>
“Dis-Integrated” 3D Memory

Memory Layers

Controller Layer

Memory Cells

Bitlines

Wordlines

Wordline Drivers

Senseamps

I/O Drivers

Power, Ground, VBB, VDH

BiSTAR

Tezzaron Semiconductor
Octopus Stack

2 Layer Stacked Device (SEM)

DRAM Control/Logic

DRAM Memory Cells
Octopus DRAM

- 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

- Power -40%
- Density x4++
- Performance +300%
- Cost -50%
Main Memory Power Cliff

DDR3 ~40mW per pin
1024 Data pins → 40W
4096 Data pins → 160W
Die on Wafer ~24uW per pin
The Industry Issue

➢ To continue to increase CPU performance, exponential bandwidth growth required.

➢ More than 200 CPU cycles of delay to memory results in cycle for cycle CPU stalls.

➢ 16 to 64 Mbytes per thread required to hide CPU memory system accesses.

➢ No current extension of existing IC technology can address requirements.

➢ Memory I/O power is running away.

Need 50x bandwidth improvement.

Need 10x better cost model than embedded memory.
The “Killer” App: Split-Die

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

Tezzaron 3D DRAM

Customer Host Device

I/O Pad area: Bumping or wire bonding
Die to Wafer With BCB Template

- KGD
- 2um alignment / 5um pitch limit
- Cu-Cu thermo compression bonding
- Multilayer capability

RPI Effort under Dr. James Lu
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

Memory also acts as interposer

Tezzaron Semiconductor
Hyper-Integration

5-9 layer stacks

<table>
<thead>
<tr>
<th>Layer</th>
<th>5 Layer</th>
<th>7 Layer</th>
<th>9 Layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poly</td>
<td>9</td>
<td>11</td>
<td>17</td>
</tr>
<tr>
<td>Copper Wire</td>
<td>21 (25)</td>
<td>32 (38)</td>
<td>34 (42)</td>
</tr>
<tr>
<td>Al/W Wire</td>
<td>7</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>Trans. Count</td>
<td>3B</td>
<td>3.1B</td>
<td>5.5B</td>
</tr>
</tbody>
</table>

Tezzaron Semiconductor
DRAM Die

“extra TSVs”
~100,000 in the core area
~50,000 in gap

“extra TSVs”
~40,000 in the core area
~250,000 on periphery

Customer circuits
Current Memory Split-Die Projects

- Octopus Memory Dies (4 pieces)
- Flip-Chip Solder Balls
- Processor

Tezzaron Semiconductor
R8051/Memory

5X Performance
1/10th Power
Tezzaron 3D Devices June/July 2011
2.5/3D Circuits

FPGA (4Xnm) | 2 Layer Processor
---|---
Active Silicon Circuit Board

3 Layer 3D Memory

μBumps

Die to Wafer Cu Thermal Diffusion Bond

C4 Bumps

Organic Substrate

Solder Bumps
3D MPW

• Complete 3D PDK 8\textsuperscript{th} Release
  – GF 130nm
  – Calibre, Synopsis, Hspice, Cadence
  – MicroMagic 3D physical editor
  – Magma 3D DRC/LVS
  – Artisan standard cell libraries

• MOSIS, CMP, and CMC MPW support
  – 90nm, 150nm SOI
  – Silicon Workbench

• >70 in process
• >400 users
Near End-of-Line

- M8, TM
- M7
- M6
- M5
- M4
- M3
- M2
- M1

2x, 4x, 8x Wiring level ~ 0.2/0.2um S/W

TSV is 1.2µ
Wide and ~10µ deep

5.6µ
New Apps – New Architectures
Relative TSV Size
Pitch and Interconnect

- SuperContact™ is 500f² (including spacing)
- Face to face is 350f² (including spacing)
- Chip on wafer I/O pitch is 35,000f²
- Standard cell gate is 200 to 1000f²
  - 3 connections
- Standard cell flip-flop is 5000f²
  - 5 connections
- 16 bit sync-counter is 125,000f²
  - 20 connections
- Opamp is 300,000f²
  - 4 connections

f² is minimum feature squared
The Applications

• Sensors
  – CMOS cameras
  – POI image processors – very high frame rates
  – High Energy Physics

• Bio Assay
  – SOC++

• Logic Memory

• Logic - Logic
<table>
<thead>
<tr>
<th>OC768 Packet Engine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dual PPC 64x ARM SOC</td>
</tr>
<tr>
<td>FPGA (CPU Augmentation)</td>
</tr>
<tr>
<td>DRAM</td>
</tr>
<tr>
<td>Flash</td>
</tr>
<tr>
<td>CAM</td>
</tr>
<tr>
<td>CAM</td>
</tr>
<tr>
<td>FPGA (Packet Cracker)</td>
</tr>
<tr>
<td>Stack Controller</td>
</tr>
</tbody>
</table>
3D-Routing Node (NOC)
Fault Tolerant Self-configuring/Re-configuring
3D Interconnect
3D Interconnect
Data Paths: On-ramp/Off-ramp
130nm Implemented Node
DRC, LVS, Transistor synthesis, Crossprobing. Multiple tapeouts, 0.35um-45nm >20GB, ~10B devices

- Independent tech files for each tier.
- Saves GDSII as flipped or rotated.
- Custom output streams for 3D DRC / LVS.

MAX-3D by Micro Magic, Inc.
Fully functional 3D layout editor.
3D LVS using QuartzLVS from Magma

- **Key features**
  - LVS each of the 2D designs as well as the 3D interconnections between them in a single run
  - Driven by a 3D “tech file” that specifies the number and order of layers, interconnect material, etc
  - TSV aware LVS extraction
  - Full debug environment to analyze any LVS mismatch

---

```plaintext
# 3D LVS Tech file

WAFER: 1

LAYOUT TOP BLOCK: lvslayer1_1
SCHEMATIC TOP BLOCK: lvslayer1
GDSII FILE: lvslayer1_1.gds
SCHEMATIC NETLIST: lvslayer1.sp

INTERFACE UP METAL: 1;0
INTERFACE UP TEXT: 1;101

...

INTERFACE:
LAYOUT TOP BLOCK: lvstop
SCHEMATIC TOP BLOCK: lvstop
GDSII FILE: lvstop_ALL.gds
SCHEMATIC NETLIST: lvstop.sp
BOND OUT METAL: 5;0
BOND OUT TEXT: 5;101
```
Challenges

• Tools
  – Partitioning tools
  – 3D P&R

• Access

• Heat

• Testing
  – IEEE 1500
  – IEEE 1149

• Standards
  – Die level
    • JEDEC JC-11 Wide bus memory
  – Foundry interface
Summary

• 3D has numerous and vast opportunities!!
  – New design approaches
  – New ways of thinking
  – Best of class integration of
    • Memory
    • Logic
    • RF
    • MEMS