Optimizing IP for Cost-Effective 2.5D Integration

Lisa Minwell
SR. DIRECTOR MARKETING

January 14, 2014
Next generation high performance computing (HPC), graphics and networking applications require 2.5D/3D integrated memory

- Bandwidth expansion
- Memory enhancement
- Product feature set expansion
- Energy efficiency
- Integration
Memory Cost Scaling Challenge

- Physical and electrical materials challenges
- Equipment capability
- Cost and complexity

Source: Mike Black, Micron, EDPS 2013
Complex systems use many external DRAMs (DDR3/DDR4/LPDDR3/LPDDR4) for packet buffering. Limited bandwidth of DDRx devices now requires very high numbers of DRAM devices connected to the central ASIC.

- To buffer 1Tbps would require 40 DDR3 DRAMs all connected through separate buses.
- ASICs already at >1,500 pins and typically pin-limited. Very difficult to increase bandwidth to memory using current approach.
  - DDR4 does not solve the problem – 1Tbps would still require 20 DRAMs.
- Current approach consumes a lot of power.

Traditional off-package interconnection between CPU and memory chip not going to scale.

Packaging and interconnect technology vital in defining memory sub-system performance.

Power Efficiency Requirement

- Scaling of I/O data rate with constant I/O power dissipation
- Even more challenging for computing memory subsystem memory bandwidth (BW) scaling

Source: ITRS 2011 Roadmap & Dr. Bill Bottoms
Approaches to 3D Integrated Circuits

Chip Level
- TSMC
- Samsung
- Micron
- Xilinx
- SK-Hynix
- Tezzaron
- Intel

Device Level
- Stanford
- Besang
- LETI
- Sandisk

Wafer Level
- Tezzaron
- IMEC
- RPI
- LETI
- Monolithic IC
## Value Proposition for Higher Value Packaging

<table>
<thead>
<tr>
<th>Market / Benefit</th>
<th>Low Power Dissipation</th>
<th>High Bandwidth CPU &lt;-&gt; DRAM</th>
<th>Low Latency IC &lt;-&gt; IC</th>
<th>Heterogeneous Integration</th>
<th>Form Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cellphones</strong> and esp. Smartphones</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td><strong>Compute servers</strong>, Network routers</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td><strong>Tablets</strong> and other Mobile devices</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td><strong>Standard PCs</strong> and Workstations</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td><strong>Automotive</strong> applications</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>

- Extremely valuable
- Very valuable
- Valuable

Other decision factors: unit cost, system cost savings, NRE, time-to-profit, risk, etc.
Stacking Memory Technologies

HMC: Courtesy of Micron

HBM: Courtesy of SK-Hynix

DiRAM: Courtesy of Tezzaron
# Markets for TSV-Based 3D Packaging Technologies

<table>
<thead>
<tr>
<th>Market</th>
<th>Smart Phone</th>
<th>Tablet</th>
<th>Networking</th>
<th>Graphics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>Apps. Processor</td>
<td>Apps. Processor</td>
<td>Networking Processor</td>
<td>Graphic Processor</td>
</tr>
<tr>
<td>Power</td>
<td>1 – 2W</td>
<td>1 – 5W</td>
<td>20W+</td>
<td>20W+</td>
</tr>
<tr>
<td>Memory Type</td>
<td>Wide I/Ox</td>
<td>Wide I/Ox or LPDDRx</td>
<td>HMC/HBM</td>
<td>HBM</td>
</tr>
<tr>
<td>Memory Size</td>
<td>2 –&gt; 4 GB</td>
<td>4-&gt; 8 -&gt; 16GB B</td>
<td>4-&gt; 8 -&gt; 16GB B</td>
<td>4 -&gt; 16GB</td>
</tr>
<tr>
<td>Interface</td>
<td>Widel/O2</td>
<td>Wide I/Ox Or DDR</td>
<td>SerDes or Parallel</td>
<td>Parallel</td>
</tr>
<tr>
<td>I/O</td>
<td>1000</td>
<td>1000 or 500</td>
<td>&lt;500 or &gt;1000</td>
<td>1024 (&gt;1600 total )</td>
</tr>
<tr>
<td>Min. Bump Pitch</td>
<td>40x40um rows</td>
<td>40x50um rows or 80um rows</td>
<td>1mm/95x55um array</td>
<td>55x55um array/ staggered rows</td>
</tr>
<tr>
<td>Packaging</td>
<td>3D</td>
<td>2.5D medium L/S density or 3D with heat management</td>
<td>Off chip memory or 2.5D high density interposers</td>
<td>2.5D high density interposers</td>
</tr>
</tbody>
</table>
Using an interposer allows the integration of highly parallel connections to memory stacks inside the package.

- Much higher total bandwidth
- Significant reduction in power consumption
- Much smaller board footprint

2.5D Increases Bandwidth

Current solution: ASIC plus multiple DRAMs

New solution: single package containing ASIC plus memory stack(s)
2.5D Reduces Cost

- Complex systems require multi-layer PCBs of 20+ layers – very expensive
- Typically, only a small percentage of the board requires the connection density that drives layer count
- Using an interposer removes the high density interconnect from the board
  - Reduces the layers required and cost
  - Increases the manufacturability and signal integrity
  - Increases density

Low connection density only requires 4-6 layer PCB

4-8 complete sub-systems on a single card instead of two
Interposer Technology Scenarios

- 2.5D Silicon Interposer on Organic Substrate
- 2.5D Glass on Organic Substrate
- 2.1D Hybrid Organic Substrate
- 2.5D Low Cost Silicon on Organic Substrate

- High-end Applications
- Mid-range Applications
- Consumer High Volume

- 2015
- 2016
- 2017
According to Yole Développement, the market value of all the devices using TSV packaged in 3D in the 3D-IC or 3D-WLCSP platforms will grow from $2.7 billion in 2011 to $40 billion in 2017 (9 percent of the total semiconductor value).

2015 will be the year for the implementation of 3D TSV technology in high-volume production with silicon interposer.

Organic 2.1D will follow providing a low cost solution.
Three trends for customers adopting 2.5D/3D technology

**Trend #1**
Extend the lifetime of existing packaging technologies

**Trend #2**
2.5D adoption for high-end markets at high price

**Trend #3**
2.5D/3D offered to reduce system cost
eSilicon – Largest Independent SDMS Provider
The Supply Chain Evolution

Vertically Integrated Manufacturing Solution 1958-1975
- Design
- EDA
- IP
- Foundries
- Test
- Package / Assembly

Diverts energy and resources from core competencies

Disaggregated Fabless Solution 1975-2000
- Design
- EDA
- IP
- Foundries
- Test
- Package / Assembly
- Customer

Complex, time-consuming, costly process

eSilicon Solution 2000-present
- Design
- EDA
- Package / Assembly
- Test
- Customer
- Foundries
- IP
- eSilicon

Reduces cost, risk and time to volume production
- Increasing cost per tapeout, tooling coupled with increasing design risk
- Increased price per gate
- A new solution is necessary to keep ASICs and ASSPs moving
Monolithic scaling saved for a very few due to complexity and cost

Effect of Landscape Change

- 180nm
- 130nm
- 65nm
- 45/40nm
- 90nm
- 22/20nm
- 16/14nm
- 32/28nm

Driven mostly by IP availability

Optimum price per gate

Few companies willing to pay more per gate

Root of ASIC tapeout industry slowdown

Relative Tapeouts and Volume
Solving the Semiconductor Problem

Source: Gartner, 2013

SDMS with eSilicon
Adding Capability While Reducing Total Cost of Ownership, Complexity and Risk

Complexity / Cost / Risk
$200M to Design and Scale an IC at 20nm

Cycles of Learning/Design Starts
14% Drop in Design Starts from 2011 to 2016

Source: Gartner, 2013
eSilicon – The “Integrator”

- Manufacturers
  - Foundries concerned by many alternatives on structure and interconnect
  - OSATs may have to deal with multiple incompatible interconnect methodologies

- Product companies
  - Device builders don’t always want to port to new processes
  - IP companies have no way to supply as silicon

- Customers
  - Don’t want to stitch together complex solutions
  - Want one supplier that will take responsibility

- eSilicon
  - Will act as the central integrator
  - Will source the best pieces of the puzzle
  - Will take responsibility for the final product
eSilicon Experience

2.5D Silicon Interposer on Organic Substrate

2.5D Glass on Organic Substrate

2.1D Hybrid Organic Substrate

Planning

2.5D Low Cost Silicon on Organic Substrate

High-end Applications  Mid-range Applications  Consumer High Volume

$$$$$  $$$$  $$  $

2015  2016  2017
eSilicon’s Mainstream 2.5D Integration Experience

- Taped-out silicon interposer in December 2014 for large ASIC surrounded by 4 HBM1 stacks
- Graphics processing application
- Accompanying ASIC to tapeout January 2015
  - eSilicon developed PHY
  - eSilicon developed IO
  - Northwest logic controller
  - Implemented some HBM2 features including bump repair
Industry First Test Vehicle Targeted for Networking Applications

- 50x50mm BGA
- 38x30mm interposer
- 50% larger than largest Si interposer
- 4 HBM daisy chain memories at 5.5x7.7mm
- 1 Massive daisy chain ASIC 24x19mm
- 4 JEDEC HBM1 stacks
- All interconnect at 55um pitch
Focus needs to shift from IP, EDA, foundry and assembly to tile sourcing, architecture, DFT, wafer sort, final test, characterization and yield management.
Architecture & Design

- Fine pitch for more highly parallel interfaces
- Ultra-low power to span 2mm instead of many inches
- What goes in die can be completely different
- Use nodes for what they are best for, and plan on reuse
- Performance bottlenecks removed completely

- With new architecture comes new physical design issues
- Hierarchical timing closure at tile level instead of block level
- Rectangular nature of interposers will influence die floorplan
- Unfamiliar interfaces
  - Greater capability
  - Unproven IP
- Mitigation strategies needed for design with new IP
• Existing I/O solutions assume driving long distances while 2.5D solutions do not have pins leaving the package
  • ESD is 100s of volts rather than 1000’s of volts
• Bump interface types not as spaced apart providing opportunity for reduced interconnect capacitive load
• Finer bump pitches enable smaller I/Os which can be combined into a large I/O macro
• Integrating new memory technologies
New set of challenges

- Test whole packaged system through a master chip
  - One chip must be custom in heterogeneous construction
  - Reduces the signal count needed at test
  - For complex networking chip, you still need about 600 pins for all of the test vectors in order to keep the test time down
  - Needs test interfaces on other tiles to be ready for test interface

- In 2.5D structure, package size is likely gated by:
  - Silicon arrangement
  - Not by ball count
  - And will have a lot of power delivery, but do not forget the test-pin requirements to keep the test time low
Test Issues/ Strategies

Wafer sort

- KGT (known good tiles) and KGI (known good interposers)
- Too many signals for a probe card
- Advanced use of loopback
- Can’t economically access all of the microbumps
  - Depending upon pitch and wafer volume
  - Test coverage may suffer
  - Significant use of test compression
  - Many strategies may exist
    - Sparse bump areas
    - Varying bump geometries
    - Add probe pads for test

Final Test

- Small portion of signals come out to package balls
- Failure modes must include what actually failed
  - Need to know what tile failed for internal interfaces
  - Assignable failure cause instead of failure mode is critical
- Test times will be longer and there may be higher power in these system-tests at package level
Characterization & Yield Challenges

Characterization

- Base characterization on a subset of pins
- Test coverage must be very well understood
- Requires creative DFT modes to be able to exercise and monitor critical aspects in characterization over:
  - Process corner
  - Temperature
  - Voltage
- Per-parameter characterization needs margins with Cpk>2
- Need in-depth understanding of partner tiles regarding specification margin
- System-level testing (SLT) may become more predominant

Yield

- With more integration, yield targets must drop
- Known yield is more important and realistic than perfect yield
- Define who owns what aspects:
  - OSATs
  - Tile manufacturers
  - Interposer manufacturers
  - Integrators
- Need a larger budget for failure analysis
- The board level yield management becomes much simpler, since complexity moves to package.
eSilicon manages the ecosystem and assumes yield risk

Complex TSV Ecosystem

SiP Assembly and Test

OSAT

Memory Vendor

Set Maker

Foundry

IP Enabler

PHY and MC IP controller IP

Business Model for SiP
Where is eSilicon Going Next?

Extended HBM (EHBM) ...

Beyond the interposer

- Developing consortium
  - HBM interface that does not require interposer
  - Target markets
    - Gaming
    - Networking
    - Supercomputer
  - Dramatically drop unit cost with elimination of interposer
  - Speed the adoption of stacked memory