DIE STACKING IS HAPPENING!

BRYAN BLACK
12-9-2013
Let’s make this keynote as interactive as possible
AGENDA

1. Sample technology flow for die stacking
2. Why die stacking is finally happening?
3. How will it impact the market?
BUT FIRST... THERE ARE TWO FLAVORS OF DIE STACKING

VERTICAL STACKING (3D)

INTERPOSER STACKING (2.5D)
WHAT IS A THROUGH SILICON VIA (TSV)?

• TSV is a copper via that electrically connects the front and back sides of the silicon
• The TSV is an old concept
• Long ago TSVs were a 5:1 ratio and more recently have become 10:1
• These diameters are mostly useless
THE INTERESTING INNOVATION OCCURRED 10+ YEARS AGO...

- Wafer thinning and thin wafer handling are the recent innovations that make die stacking manufacturing feasible.

![Diagram showing TSV (5 to 10um diameter) and wafer sub with dimensions 50-100µm]
COMPLEXITY OF DIE STACKING

• **Resources:** Internal and external limitations
• **Business:** New business partners, new business models, new roles for existing partners, new capital investments, no history for forecasting
• **Product:** How does a design take advantage of die stacking?, CAD infrastructure, Define requirements, Mitigate high risk
• **Test:** Wafer probe, KGD, Manufacturing thermal heads, Impact of gimbaling, component test & repair
• **Assembly manufacturing:** Big die challenges, Thin wafer handling, Die to die bonding thermal compression or reflow, Assembly order, New mbump technologies, EM, ESD, Underfill
• **Wafer Manufacturing:** TSVs, Carrier attach/detach, Thinning, mbump
• **Platform:** System characterization changes with integration
• **Packaging:** Large and new materials to reduce warpage
• **Thermals:** System solutions, Package solutions, Materials selection
• **Reliability:** How is reliability proven with multiple vendor components with different requirements?
• **Quality:** How is quality distributed among partners? Who does burn in?
• **Yield & Cost:** DFY, Require time for maturity

No single challenge is an issue but there are many, they are diverse, and they interact
It would take weeks to walk through the nuances of die stacking technologies

This is a cartoon illustrating one instance of the technology at 10,000 feet
THE BASIC STRUCTURE OF THIS CARTOON

- Two dies attached to each other using μBumps
- TSVs in the bottom die provide external I/O access and power delivery to the top die
INTERFACES

• Die to die interfaces have a PHY on each die with a TSV and μBump in the path
• The interface on both dies is exposed during assembly but not after
**STEP 1 (FRONT-END COMPLETE)**

- This cartoon is TSV processing after Contact (TSV Middle)
- Process the wafer as usual through the FE layers and Contact

![Diagram](image-url)

- **Contact**
- **FE Layers**
- **Wafer Sub**
- **Transistor layer**
- **750μm**
STEP 2 (DEEP ETCH TSV CAVITY)

Contact
FE Layers
Wafer Sub

750μm
STEP 3 (TSV SIDEWALL PROTECTION)

• Sidewall protection (Nitride or Oxide) is required to prevent Cu TSV from diffusing into the Si substrate
STEP 4 (TSV FILL)

- TSV diameter influences power delivery capability and EM limitations
- Two small can be a challenge for high power devices

Wafer Sub

FE Layers

Contact

Dia > 5μm

Pitch > 15μm

750μm
STEP 5 (METALLIZATION AND BUMPING)

• Add metal layers starting with M1; TSVs land on M1
• Then add UBM and C4 bump interface
TEST POINT 1 (WAFER PROBE)

- Perform typical wafer probe
- Cannot test TSVs
**STEP 6 (CARRIER WAFER)**

- Using Adhesive attach a carrier wafer to the front side of the wafer for the back side processing

- The carrier provides a Si base for tool chucks and support for processing thin wafers
STEP 7 (WAFER THINNING)

Carrier Wafer

Metal Layers

Contact

FE Layers

Wafer Sub

PHY

PHY
STEP 8 (BACKSIDE PROCESSING)

• Add BS RDL with passivation openings for die to die interface
STEP 9 (MICRO-BUMP BACKSIDE INTERFACE)

Carrier Wafer

Metal Layers

Contact

FE Layers

PHY

PHY

Wafer Sub

BS RDL

Passivation

Cu pad for μBump landing

PHY
STEP 10 (BACKSIDE TAPE)

- Carrier Wafer
- Metal Layers
- Contact
- FE Layers
- Wafer Sub
- BS RDL
- Passivation
- Dicing Tape
STEP 11 (REMOVE CARRIER)

- Remove Carrier wafer, clean, and singulate
STEP 12 (REMOVE DIE FROM TAPE)

- Bare die handling

Very important
STEP 13 (PICK & PLACE INSERTION THEN UNDERFILL)
RISK HANDLING ALL DIES IN THE SYSTEM

Components are shipped and stored with exposed μBump interface.

Die #2

PHY

μbump

Package Substrate

Passivation
BS RDL
Wafer Sub
FE Layers
Contact
Metal Layers

PHY

PHY
STEP 15 (ATTACH OTHER DIES)

• Die to die interfaces are no longer exposed
There is research to be done for cooling irregular shapes.
ESD RESEARCH (CHALLENGE #1)

• The TSV, PHY, Control, Test, and ESD must fit within the \(\mu\)Bump pitch

• The \(\mu\)Bump pitch is going to scale quickly but ESD doesn’t scale well

• If ESD doesn’t scale, the \(\mu\)Bump pitch increases to accommodate and BW is reduced

Die to Die Interface Pattern
ESD RESEARCH (CHALLENGE #2)

• What if die #2 is a DRAM?
• DRAM has a tremendous amount of capacitance
• How does it discharge without damaging the die to die interface?
THE TWO INTERESTING QUESTIONS

Why is die stacking finally happening?
&
How will it impact the market?
Die stacking is catching on in FPGAs, Power Devices, and MEMs

But...

There is nothing in mainstream computing CPUs, GPUs, and APUs
SYSTEM TRENDS IN THE INDUSTRY

• The industry is driving performance density

• Improvements in performance density lead to new form factors

Increasing Performance Density

• New form factors enable new usage models

• Without new usage models the industry stagnates
MOORE’S LAW IS THE BACK BONE OF INTEGRATION

µProcessor

Intel 4004
Source: proyectoyautja.proboards.com

Intel 4004
Source: proyectoyautja.proboards.com

Intel P5
Source: gecko54000.free.fr

Intel P5
Source: gecko54000.free.fr

Multi-Media

Graphics

AMD “Ontario”
Source: AMD

Graphics

North Bridge

AMD K8
Source: barcelona.com

North Bridge

South Bridge

AMD “Kabini”
Source: bytesandbits.it

South Bridge

IVR

Intel “Haswell”
Source: hardware.slashdot.org

IVR

Cache & FPU

Intel 486
Source: gecko54000.free.fr

Cache & FPU

Intel 486
Source: gecko54000.free.fr


MOORE’S LAW IS THE BACK BONE OF INTEGRATION
MICROARCHITECTURE MOTIVATION FOR INTEGRATION

• Communication is overhead consuming power, latency, and footprint

• Interface power is proportional to bandwidth and the link RCs

• And... BW is limited by the off die interface which doesn’t scale quickly

• But... off die BW demand increases with transistor density
COST IS THE PRIMARY MOTIVATION FOR INTEGRATION

- Increased transistor density and lower power each generation is great
- But lower cost is the driving force of the industry and integration
- Improvement in cost per transistor is why the industry adopts new process nodes

Source: AMD
SI INTEGRATION IS RUNNING OUT OF GAS!

• Moore’s Law will continue but there is a limitation

• All similar technology components have been integrated such as Cache, FPU, Multi-Media, NB, GPU, SB, etc...

• Only disparate technologies such as DRAM, MEMS, True IVR, Storage, Optics are left
SI INTEGRATION IS RUNNING OUT OF GAS!

• Moore’s Law will continue but there is a problem

• Process scaling will to stop supporting diverse functionalities on a single die such as fast logic, low power logic, analog, and cache

• The single die will want to break into specialized components to maximize the value of new and existing process nodes
MOORE’S LAW MAY NOT REDUCE COST

- Moore’s Law will continue but there are long term cost challenges
- Process node complexity is climbing due to feature size scaling and diversity of supported functionalities

Source: AMD
DIE STACKING IS IDEAL FOR INTEGRATION

• All they do is reduce metal interconnect by improving proximity of disparate technologies

VERTICAL STACKING (3D)

INTERPOSER STACKING (2.5D)
• Process complexity is increasing and yield is dropping as mask count increases
• Large die sizes will continue to have yield challenges
• Die partitioning is challenging and there is significant microarchitecture research
For example, thick oxide gates are useful for high performance high efficiency IO devices.

Separating functionalities such as Analog reduces node complexity.

Interesting research to determine which process nodes fit best with each functionality.
DIE STACKING MOTIVATION (MEMORY INTEGRATION)

- System power is fixed in all platforms
- Compute performance in all platforms is proportional to memory BW
- Memory BW power increases with demand

Lower performance will obviously never happen
A new DRAM solution is required to prevent performance from flattening out

Coming Soon!
### 1st Generation HBM is Sampling!

[HTTP://WWW.JEDEC.ORG/STANDARDS-DOCUMENTS/DOCS/JESD235](HTTP://WWW.JEDEC.ORG/STANDARDS-DOCUMENTS/DOCS/JESD235)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voltage (I/O, VPP)</td>
<td>1.2V; 2.5V</td>
</tr>
<tr>
<td>Data Rate per pin</td>
<td>800 – 1.2 MT/s</td>
</tr>
<tr>
<td>DRAM Die Density</td>
<td>2Gb</td>
</tr>
<tr>
<td>DRAM Dies per Stack</td>
<td>4</td>
</tr>
<tr>
<td>Interface Width</td>
<td>1024</td>
</tr>
<tr>
<td>DQs/Channel</td>
<td>128</td>
</tr>
<tr>
<td>Channels/Die; Channels/Stack</td>
<td>2; 8</td>
</tr>
<tr>
<td>Address/Command Interface</td>
<td>DDR</td>
</tr>
<tr>
<td>Data Bus Clocking</td>
<td>Source synchronous; uni-directional/4byte</td>
</tr>
<tr>
<td>On Die DLL</td>
<td>No</td>
</tr>
<tr>
<td>Data Bus Termination – On die</td>
<td>No</td>
</tr>
<tr>
<td>Address/Cmd Termination – On die</td>
<td>No</td>
</tr>
<tr>
<td>DQ Training Mode</td>
<td>N/A</td>
</tr>
<tr>
<td>Page Size</td>
<td>2KB</td>
</tr>
<tr>
<td>Banks/Channel</td>
<td>8</td>
</tr>
<tr>
<td>DRAM Prefetch; Burst Length</td>
<td>256 bits; 2</td>
</tr>
</tbody>
</table>

**Diagram:**
- **Logic Die:**
  - uBump interface for 2.5D or 3D
  - 8 x 128b Channels

- **4 DRAMs:**
Die stacking improves the proximity of the DRAM to Compute

Dense and fine pitch interconnect enables simple low power interfaces as well as fine grain power control of the DRAM
• Dramatically improved memory BW/W rolls back the impact of recent memory system power growth and yields years of future scaling

This part of the curve is successfully avoided again.

We can be here!
INTERPOSER WILL BE THE SOC WITH MULTIPLE 3D COMPONENTS

• Focus process node development on specific application functionalities
  – Reduce complexity and mask layer count
  – Reduce process node TTM
  – Reduce wafer runtime
  – Reduce wafer start cost

• Yield improves
• Functionalities scale at their own pace
• Improve performance, power, area, and cost of each functionality
• IP sharing includes test, reliability, and yield learning
• How is the SoC partitioned and integrated with the system?
• What are the interfaces and how do they get standardized?
• How is this tested?
• How is power managed?
SOCS EVOLVE AND COST CHANGES OVER TIME

- SoCs are 100% internal to orgs
- ARM creates an industry eco-system reducing cost with greater sharing
- SoC complexity drives cost up and the industry consolidated
AS ALWAYS COST WILL DRIVE THE NEXT EVOLUTION OF SOCS

SoCs are 100% internal to orgs

ARM creates an industry eco-system reducing cost with greater sharing

SoC complexity drives cost up and the industry consolidated

Die Stacking will reduce cost by sharing at the die level and driving customization
SOMETHING EVEN MORE INTERESTING IS GOING TO HAPPEN

- Socket ownership will shift from traditional SoC companies to Platform companies
- Market penetration will be easier for custom accelerators sparking a growth area rich with research
TAKEAWAYS

Die stacking is happening in the mainstream

It is happening now because we need it

&

It is going to change who and how we build sockets in the future
DISCLAIMER & ATTRIBUTION

The information presented in this document is for informational purposes only and may contain technical inaccuracies, omissions and typographical errors.

The information contained herein is subject to change and may be rendered inaccurate for many reasons, including but not limited to product and roadmap changes, component and motherboard version changes, new model and/or product releases, product differences between differing manufacturers, software changes, BIOS flashes, firmware upgrades, or the like. AMD assumes no obligation to update or otherwise correct or revise this information. However, AMD reserves the right to revise this information and to make changes from time to time to the content hereof without obligation of AMD to notify any person of such revisions or changes.

AMD MAKES NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE CONTENTS HEREOF AND ASSUMES NO RESPONSIBILITY FOR ANY INACCURACIES, ERRORS OR OMISSIONS THAT MAY APPEAR IN THIS INFORMATION.

AMD SPECIFICALLY DISCLAIMS ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR ANY PARTICULAR PURPOSE. IN NO EVENT WILL AMD BE LIABLE TO ANY PERSON FOR ANY DIRECT, INDIRECT, SPECIAL OR OTHER CONSEQUENTIAL DAMAGES ARISING FROM THE USE OF ANY INFORMATION CONTAINED HEREIN, EVEN IF AMD IS EXPRESSLY ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

ATTRIBUTION

© 2013 Advanced Micro Devices, Inc. All rights reserved. AMD, the AMD Arrow logo and combinations thereof are trademarks of Advanced Micro Devices, Inc. in the United States and/or other jurisdictions. Other names are for informational purposes only and may be trademarks of their respective owners.