HBM Package Integration: Technology Trends, Challenges and Applications

Suresh Ramalingam

Hot Chips
Aug 21, 2016
Agenda

- Motivation
- HBM Packaging Options
- Interposer Design
- Supply Chain
- Application and Challenges
- Summary

Stacked Silicon Interconnect Technology Refers to Xilinx 3D solutions
FPGA’s in the Data Center Today

- The Accelerator (FPGA or GPU) is used to offload only certain tasks
  - These tasks are called “Workloads”, and FPGA’s are well suited for many workloads
  - Note: Accelerators *don’t* replace the CPU!

- Hard and Soft is the basic approach
  - **Hard** is the IO, Memory and PCIe interfaces
    - Does not change
  - **Soft** is the workload being accelerated
    - Is configured on the fly using P.R.

- API’s are run on the CPU to reprogram the FPGA to accelerate the workload as needed.
  - Average P.R. happens every 15 minutes!

- Acceleration requires lots of memory BW
Multi-Die Technology for HBM (Side-by-Side)

- MCM: Multi-chip Module
- FO-MCM: Fan-out MCM
- FL-MCM: Fine-line MCM
- NTI: No TSV Interconnection
- SLIM: Silicon-Less Integrated Module
- EMIB: Embedded Multi-die Interconnect Bridge

I/O density (continuous interface)

- >$10^4$
- $10^3$
- $10^2$

Organic Interposer/MCM W/ S 5/5um, ML - 12

Cisco, ECTC 2016

Figure 2. A schematic top view of the 3D SiP designed.

Figure 3. A schematic cross-sectional view of the 3D SiP designed.
Multi-Die Technology for HBM (Side-by-Side)

MCM: Multi-chip Module
FO-MCM: Fan-out MCM
FL-MCM: Fine-line MCM
NTI: No TSV Interconnection
SLIM: Silicon-Less Integrated Module
EMIB: Embedded Multi-die Interconnect Bridge

Grey Zone: Scaling
(multi-die integration & fine line & metal layer)
- EMIB
- Fan-out
- Fine Line Substrate
- NTI/SLIM

Organic Interposer/MCM W/5/5um, ML -12

EMIB (Intel)

Advanced Organic - Ex. PHY W/S<3/3um, ML3-4

I/O density (continuous interface)

>10^4

10^3

10^2
Multi-Die Technology for HBM (Side-by-Side)

MCM: Multi-chip Module
FO-MCM: Fan-out MCM
FL-MCM: Fine-line MCM
NTI: No TSV Interconnection
SLIM: Silicon-Less Integrated Module
EMIB: Embedded Multi-die Interconnect Bridge

I/O density (continuous interface)

>10^4
10^3
10^2

Advanced Organic - Ex. PHY W/S<3/3um, ML3~4
FL-MCM (Shinko/Hitachi/UMTC/Ibiden)

Organic Interposer/MCM W/S 5/5um, ML -12
Multi-Die Technology for HBM (Side-by-Side)

- MCM: Multi-chip Module
- FO-MCM: Fan-out MCM
- FL-MCM: Fine-line MCM
- NTI: No TSV Interconnection
- SLIM: Silicon-Less Integrated Module
- EMIB: Embedded Multi-die Interconnect Bridge

Advanced Organic - Ex. PHY W/S<3/3um, ML3~4

Organic Interposer/MCM W/S 5/5um, ML ~12

FO-MCM (TSMC/ASE/SPIL/Amkor)

NTI (SPIL)

<table>
<thead>
<tr>
<th>Layer</th>
<th>Scheme (thk)</th>
<th>Dielectric</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1: L/S 2/2 um</td>
<td>F5-RDL1</td>
<td>SiO2/SiNx</td>
</tr>
<tr>
<td>M2: L/S 5/5 um</td>
<td>BS-RDL2</td>
<td>PBO</td>
</tr>
<tr>
<td>M3: L/S 10/10 um</td>
<td>BS-RDL3</td>
<td>PBO</td>
</tr>
</tbody>
</table>

Contact to BGA Ball

Figure 16. Hybrid Integration Scheme
Multi-Die Technology for HBM (Side-by-Side)

- MCM: Multi-chip Module
- FO-MCM: Fan-out MCM
- FL-MCM: Fine-line MCM
- NTI: No TSV Interconnection
- SLIM: Silicon-Less Integrated Module
- EMIB: Embedded Multi-die Interconnect Bridge

I/O density (continuous interface)

- >10^4
- 10^3
- 10^2

Organic Interposer/MCM
WS 5/5um, ML -12

Advanced Organic - Ex. PHY
WS <3/3um, ML3-4

2.5D Si interposer
WS <1/1um, ML ≤ 3

SSIT (Xilinx)

Fiji (AMD)

Package Substrate
# Multi-Die Package Design Rule Comparison

<table>
<thead>
<tr>
<th>Design Rules for Die to Die interconnection</th>
<th>MCM (Substrate) Integrated Fine Layers</th>
<th>EMIB (Embedded Multi-die Interconnect bridge)</th>
<th>Silicon Interposer (65 nm BEOL)</th>
<th>WLFO (Wafer Level Fan-out)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum Bump pitch (um)</td>
<td>130 (C4) 40 (u-bump d2d interface)</td>
<td>130 (C4) 40 (u-bump) bridge</td>
<td>&lt; 40 (u-bump)</td>
<td>40 um RDL pad pitch</td>
</tr>
<tr>
<td>Via size / pad size (um)</td>
<td>10 / 25</td>
<td>0.4 / 0.7</td>
<td>0.4 / 0.7</td>
<td>10/25</td>
</tr>
<tr>
<td>Minimum Line &amp; Space (um)</td>
<td>2 / 2</td>
<td>0.4 / 0.4</td>
<td>0.4 / 0.4</td>
<td>2 / 2</td>
</tr>
<tr>
<td>Metal thickness (um)</td>
<td>2-5</td>
<td>1</td>
<td>1</td>
<td>2-5</td>
</tr>
<tr>
<td>Dielectric thickness (um)</td>
<td>~5</td>
<td>1</td>
<td>1</td>
<td>&lt; 5</td>
</tr>
<tr>
<td># of die-to-die connections per layer + GND shield layer (2L)</td>
<td>1000's</td>
<td>1000’s (bridge interface length limited)</td>
<td>10,000’s</td>
<td>1000’s</td>
</tr>
<tr>
<td>Minimum die to die spacing (um)</td>
<td>&lt; 500</td>
<td>&lt;2500</td>
<td>&lt;100</td>
<td>&lt; 250</td>
</tr>
<tr>
<td># of High density layers feasible</td>
<td>Not a limitation</td>
<td>Not a limitation</td>
<td>Not a limitation</td>
<td>1-3L layers</td>
</tr>
<tr>
<td>Die Sizes for assembly and # of assemblies</td>
<td>Not a concern d2d interconnect only</td>
<td>Size &amp; # limitation?</td>
<td>Not a concern</td>
<td>Size limitation?</td>
</tr>
</tbody>
</table>
HBM2 System Overview (Jedec)

- HBM2 system with SOC/DRAM on interposer with 3-6mm length
- 24 signals across 55um u-bump pitch across interface
- Supports 2Gb/s PHY (1Tb/sec bandwidth for 4-Hi)

(Jedec) Interposer Parameters *

<table>
<thead>
<tr>
<th>Width</th>
<th>Space</th>
<th>Thickness</th>
<th>Length</th>
<th>Resistance</th>
</tr>
</thead>
<tbody>
<tr>
<td>2u</td>
<td>2u</td>
<td>0.5-2u</td>
<td>6mm</td>
<td>36 ohms</td>
</tr>
</tbody>
</table>
Interposer Design Tools & Methodology

- **Vertical Routing**
  - A model from ubump to package pin is generated and used by high frequency designs (e.g. GT and IO)

- **Horizontal Routing**
  - **Die LVS and extraction**
    - Standard extraction
    - Ubump is extracted as a subcircuit
  - **Interposer with box die LVS and extraction**
    - Interposer metal extraction
    - TSV is extracted as a subcircuit
  - **Combine the extracted netlists from die and interposer for simulation**

- **Top-down view**
  - Interposer metal extraction
  - TSV is extracted as a subcircuit

- **Vertical route**
  - Vertical routing from Die 1 to Die 2 to Die 3

- **Horizontal route**
  - Die LVS and extraction
  - Interposer with box die LVS and extraction
  - Combine the extracted netlists from die and interposer for simulation
2.5D - uses the same tool sets as single die design with customized interposer / top die PDK

EDA vendor Tools are validated by TSMC design reference flow

PKG - uses same tool sets as Flip chip (C4-to-BGA)
- TSV budget is handled in the Silicon design environment
- Layout and PI tools must be capable to handle large data sets
Supply Chain – Silicon Interposer Approach

- Xilinx in production with 2nd generation of products with TSMC CoWoS

- TSMC CoWoS™
- UMC/Inotera
- SPIL/Amkor/ASE

[Diagram showing the supply chain process involving TSV Si Interposer, Logic IP, Memory, FPGA, TSV Si Interposer thinning/C4/Sorting, Chip-on-Wafer Bonding, Thinning/C4/Sorting, De-carrier & Dicing, Packaging on substrate, Final Test & Shipment, and KGD (1-3) chip stacking with reconfiguration.]
HBM Integration – HPC Application

- HBM Power map provided by vendors
- Thermal model can be done in Flotherm or IcePak environments for example

HBM can be 97C and HBM I/F 96C @30C
HBM gradient ~14C (~2.5C/Layer)

Air cooling can be a challenge! HBM 8-Hi needs to support > 95C $T_j$ ....

PCI-e card: Full Length/Full Height
Card power: 320W
Airflow: 15CFM
Typical ambient 30C

@ 30C Inlet ambient to PCI-e card
Telecom Application with HBM

PKG size: 52.5x52.5 mm
Total Heat Dissipation is 116.3 W

Temperature Rise above ambient @Rhs 0.34

Air cooling is a challenge! HBM needs to support > 95C Tj .....
Summary

- Tb/s low latency bandwidth and lower system power is driving the need for HBM adoption
- Silicon Interposer (2.5D) is the incumbent technology of choice. Potentially lower cost, fine pitch interconnect wafer-level and substrate based technologies are emerging
- To drive broader adoption of HBM applications (cooling limited) and higher performance stacks (8-Hi), higher HBM junction temperature (>95°C) needs to be supported