HBM: Memory Solution for High Performance Processors

Kevin Tran and June Ahn

SK hynix Inc.
October 2014
Agenda

- “Completion of HBM1 Qualification” Announcement
- Landscape of memory solution challenges
- Why HBM is the suitable solution
- HBM Architecture Review
- Conclusion
### Achievement of HBM1 Qualification

**SK hynix achieved Customer Qualification Level samples in Sep’14**

<table>
<thead>
<tr>
<th>SK hynix TSV chronicle</th>
<th>SK hynix World-First HBM Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘08 4Gb Flash</td>
<td>➢ Worldwide first HBM provider</td>
</tr>
<tr>
<td>‘10 4Gb DRAM DDP-WLP</td>
<td>➢ Customer Qualification Samples shipping today</td>
</tr>
<tr>
<td>‘11 16Gb DRAM 9MCP</td>
<td>➢ Volume production begins Q1’15</td>
</tr>
<tr>
<td>‘11 16/32GB 4hi KGSD DIMM WIO</td>
<td>➢ HBM2 design wins in progress with major SoCs in multiple markets</td>
</tr>
<tr>
<td>‘13 5mKGSD HBM</td>
<td></td>
</tr>
</tbody>
</table>

 bilder

*Bottom View*  *Top View*  *Section View*
Agenda

- “Completion of HBM1 Qualification” Announcement
- Landscape of memory solution challenges
- Why HBM is the suitable solution
- HBM Architecture Review
- Conclusion
Challenges#1) Bandwidth

All DRAMs expect to face immense bandwidth requirement

**Data Rate/Pin**

<table>
<thead>
<tr>
<th>Year</th>
<th>GDDR5</th>
<th>DDR4</th>
<th>LPDDR4</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Challenges**

- DRAM transistor and tester limit over 10Gbps
- Trade off with power consumption
- Severe die overhead

**Gb/s/pin**

<table>
<thead>
<tr>
<th>GDDR5</th>
<th>DDR4</th>
<th>LPDDR4</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.0</td>
<td>3.2</td>
<td>4.2</td>
</tr>
</tbody>
</table>

**IO**

<table>
<thead>
<tr>
<th>GDDR5</th>
<th>DDR4</th>
<th>LPDDR4</th>
</tr>
</thead>
<tbody>
<tr>
<td>X32</td>
<td>X16</td>
<td>X32</td>
</tr>
</tbody>
</table>

(Source: SK hynix)
Challenges#2) Density/Latency

Density has increased by 1000X over the past two decades, Latency has decreased only by 56%

<table>
<thead>
<tr>
<th>Density/Latency</th>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Capacity / Cost Limitation</td>
</tr>
<tr>
<td></td>
<td>• DRAM Scaling Challenges</td>
</tr>
<tr>
<td></td>
<td>• Severe die overhead increase</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>Density (Mb)</th>
<th>Mode</th>
<th>tRC (ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1992</td>
<td>4Mb</td>
<td>Cache DRAM</td>
<td>110</td>
</tr>
<tr>
<td>2002</td>
<td>256Mb</td>
<td>DDR</td>
<td>6X</td>
</tr>
<tr>
<td>2003</td>
<td>1Gb</td>
<td>DDR2</td>
<td>5X</td>
</tr>
<tr>
<td>2012</td>
<td>4Gb</td>
<td>DDR3</td>
<td>4X~5X</td>
</tr>
<tr>
<td>2014</td>
<td>8Gb</td>
<td>DDR4</td>
<td>4X</td>
</tr>
</tbody>
</table>

(Source: SK hynix)
Reducing power and increasing performance are always trade-offs.

**Challenges#3) Power Efficiency**

- Trade-off between performance and power

<table>
<thead>
<tr>
<th>Year</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>100%</td>
</tr>
<tr>
<td>2003</td>
<td>-57%</td>
</tr>
<tr>
<td>2012</td>
<td>-79%</td>
</tr>
<tr>
<td>2014</td>
<td>-70%</td>
</tr>
</tbody>
</table>

(Source: SK hynix)
Challenges#4) Form Factor

System level board design challenges for # of DRAM

<table>
<thead>
<tr>
<th>Mode</th>
<th>2008</th>
<th>2011</th>
<th>2014</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed(Mbps)</td>
<td>800</td>
<td>1600</td>
<td>2133</td>
<td>3200</td>
</tr>
<tr>
<td># of DRAM</td>
<td>4</td>
<td>15</td>
<td>48</td>
<td>78</td>
</tr>
<tr>
<td># of DQ</td>
<td>64</td>
<td>240</td>
<td>768</td>
<td>1248</td>
</tr>
</tbody>
</table>

(Required Memory Bandwidth = 4 x Line rate)

(Source: SK hynix)
Revolutionary Changes with TSV memories

TSV is a revolutionary technology enabling next generation memories

▼ High Speed

▼ Lower Power

▼ High Density

▼ Small Form Factor
Agenda

- “Completion of HBM1 Qualification” Announcement
- Landscape of memory solution challenges
- Why HBM is the suitable solution
- HBM Architecture Review
- Conclusion
Each application has different memory requirement, but most common are high bandwidth and density based on real time random operation.
Lower speed/pin and Cio reduce power consumption by 42% compared to GDDR5

I/O Power Efficiency@IDD4R
Pseudo channel improves tFAW by 60% compared to DDR4

Low Latency
1GB HBM package size is smaller than 1 tablet of aspirin
SiP enhances memory channel conditions

**Memory Channel Conditions**

- **Long Distance**: Loading, Power, C&R
- **Speed, tDV**

**DRAM vs. SiP**

<table>
<thead>
<tr>
<th>Items</th>
<th>DRAM (Off-chip)</th>
<th>SiP (2.5D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical dimension</td>
<td>very large</td>
<td>small</td>
</tr>
<tr>
<td>Signal Distance</td>
<td>long</td>
<td>short</td>
</tr>
<tr>
<td>Signal Loading</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>Driver Strength</td>
<td>large</td>
<td>small</td>
</tr>
<tr>
<td>IO Speed /Pin</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>Termination</td>
<td>need</td>
<td>no</td>
</tr>
<tr>
<td>Power Consumption</td>
<td>high</td>
<td>Low</td>
</tr>
</tbody>
</table>
Agenda

- “Completion of HBM1 Qualification” Announcement
- Landscape of memory solution challenges
- Why HBM is the suitable solution
- HBM Architecture Review
- Conclusion
Through Silicon Via - TSV

TSV is the underlying technology for 3D Stack (High Density / Small size PKG / High speed)
HBM 2.5D SiP Structure

System-in-Package implementation with KGSD*

- **FBGA**
- **KGSD**

* KGSD (Known Good Stacked Die)
HBM Overall specification

- **HBM1**
  - 2Gb Density per DRAM die
  - 1Gbps speed /pin
  - 128GB/s Bandwidth
  - 4 Hi Stack (1GB)
    - x1024 IO
    - 1.2V VDD
    - KGSD w/ μBump

- **HBM2**
  - 8Gb per DRAM die
  - 2Gbps speed/pin
  - 256GBps Bandwidth/Stack
  - 4/8 Hi Stack (4GB/8GB)
HBM Architecture Overview

4 Core DRAM + 1 Base logic die (Chip on Wafer)

[Table]

<table>
<thead>
<tr>
<th>Items</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Stack</td>
<td>4(Core) + 1(Base)</td>
</tr>
<tr>
<td>Ch./Slice</td>
<td>2</td>
</tr>
<tr>
<td>Total Ch. for KGSD</td>
<td>8</td>
</tr>
<tr>
<td>IO/Ch.</td>
<td>128</td>
</tr>
<tr>
<td>Total I/O/KGSD</td>
<td>1024(=128 x 8)</td>
</tr>
<tr>
<td>Address/CMD</td>
<td>Dual CMD</td>
</tr>
<tr>
<td>Data Rate</td>
<td>DDR</td>
</tr>
</tbody>
</table>

Each HBM die has 2 channels
1 channel consists of 128 TSV I/O with 2n prefetch

HBM Base Die Architecture

Base die consists of 3 Areas – PHY, TSV, Test Port Area

HBM ballout area 6,050x3,264 μm

Pseudo Channel Concept

- HBM is comprised of 8ch (2Channel/die) with 128DQs per channel.
- Each channel(CH) consists of 2 Pseudo Channel(PS). Only BL4 is supported.

(Note: Pseudo channel is only applicable to HBM2)
Each pseudo channel share AWORD, but has separated banks & independent 64 I/Os.
Restriction of tFAW in Legacy mode

- For Legacy Mode, Each channel has 2KB page size
- Restriction of Gapless Bank Activation by tFAW (4 activate window)
  - $t_{FAW}=30\text{ns} > 4\text{Bank}*t_{RRD}=16\text{ns}$ ➔ Lower efficiency of Band Width
- Suppose $t_{CK}=2\text{ns}$, $t_{FAW}=30\text{ns}$, $t_{RRD}=4\text{ns}$
Benefit of Pseudo Channel

- Pseudo channel has reduced page size compared to Legacy mode. : 2KB ➔ 1KB
- Lower Active Power(IDD0) by 1K Page size
- Define tEAW (1KB x 8 ACT) instead of tFAW (2KB x 4 ACT)
- Bandwidth improvement by more Activations during tFAW
Base Die Customization – Future HBM Concept

Logic Layer ➔ Host I/F + Memory I/F + Base Logic/IP Block

Customization to meet various requirements

Overcome Memory Scaling
- Timing
- Refresh

- Parallel-to-Serial(P2S)/S2P
- JTAG, PMBIST
- Configuration Registers
- Error Handling
- ...

SoC

Host I/F

Memory I/F

DRAM
Agenda

- “HBM1 Qualification Sample” Announcement
- Landscape of memory solution challenges
- Why HBM is the suitable solution
- HBM Architecture Review
- Conclusion
Complicated TSV Ecosystem requires close collaboration among all stakeholders.
Conclusion

HBM enables new memory subsystem architecture for Next Generation high performance processors. SK hynix is the leader in HBM technology

- SK hynix HBM1 Customer Qualification samples are shipping
- HBM addresses the major memory requirements with lower power, lower latency, and smaller form factor.
- HBM2 Pseudo channel improves bus utilization and system performance
Thank You