HMC Overview

A Revolutionary Approach to System Memory
Agenda

▶ Inception
▶ Architecture
▶ Reliability
▶ Comparisons to other Memories
▶ Industry Adoption
  ▪ HMC Consortium
▶ Summary
Memory Challenges By Application

Higher Performance for Server Applications

Reduced Latency for Networking Applications

Lower Power for Mobile Applications
Needed: A Memory Revolution

- Continued global demand for mobility: connected anytime, anywhere, any device
- Device proliferation fueling exponential data growth
- Cloud services stretching current networking, storage and server capabilities
- Big data analytics challenge: information rich but data poor
Hybrid Memory Cube (HMC)

Fast process logic and advanced DRAM design in one optimized package

- Power Efficient
- Smaller Footprint
- Increased Bandwidth
- Reduced Latency
Agenda

▶ Inception – what is it
▶ Architecture
▶ Reliability
▶ Comparisons to other Memories
▶ Industry Adoption
  ▪ HMC Consortium
▶ Summary
HMC Architecture
Enabling Technologies

Abstracted Memory Management
Memory Vaults Versus DRAM Arrays
Logic Base Controller

Through-Silicon Via (TSV) Assembly
Innovative Design & Process Flow
Advanced Package Assembly
HMC Architecture

Start with a clean slate

DRAM
Re-partition the DRAM and strip away the common logic
HMC Architecture

Stack multiple DRAMs
HMC Architecture

Re-insert common logic on to the Logic Base die

3D I & TSV Technology

- DRAM7
- DRAM6
- DRAM5
- DRAM4
- DRAM3
- DRAM2
- DRAM1
- DRAM0

Logic Chip
HMC Architecture

Add advanced switching, optimized memory control and simple interface to host processor(s)...

3D & TSV Technology

Vault Control

 Logic Base

 Memory Control

 Crossbar Switch

 Link Interface Controller

 Processor Links

 Link Interface Controller

 Link Interface Controller

 Link Interface Controller

 Vault

 DRAM

 Logic Base
HMC Architecture
Link Controller Interface

Example:
- 16 Transmit Lanes @ 10Gb/s Each
- 16 Receive Lanes @ 10Gb/s Each
The Package
The Stack-up
HMC Near Memory

- All links between host CPU and HMC logic layer

- Maximum bandwidth per GB capacity
  - HPC/Server – CPU/GPU
  - Graphics
  - Networking systems
  - Test equipment
HMC Far Memory

- HMC links connect to host or other cubes
  - Links form networks of cubes
  - Scalable to meet system requirements

- Future interfaces
  - Higher speed electrical
  - Optical
  - Whatever the most appropriate interface for the job!
Agenda

- Inception – what is it
- Architecture
- Reliability
- Comparisons to other Memories
- Industry Adoption
  - HMC Consortium
- Summary
## RAS Feature Comparison

<table>
<thead>
<tr>
<th>FEATURE</th>
<th>DRAM</th>
<th>RDIMM</th>
<th>HMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extensive Test Flow</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Data ECC</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Address/Command Parity</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Mirroring (back-up memory)</td>
<td></td>
<td></td>
<td>✓✓</td>
</tr>
<tr>
<td>Sparing (Chipkill)</td>
<td>✓✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lockstep (redundancy w/better ECC)</td>
<td>✓✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRC Coding</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self Repair</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIST</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error Status and Debug Registers</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIMM Isolation (flags faulty DIMM)</td>
<td>✓✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Memory Scrubbing</td>
<td>✓</td>
<td>✓✓</td>
<td></td>
</tr>
</tbody>
</table>

✓ Supported  ✓✓ Redundant or not needed
Agenda

▶ Inception – what is it
▶ Architecture
▶ Reliability
▶ Comparisons to other Memories
▶ Industry Adoption
  ▪ HMC Consortium
▶ Summary
Extreme Performance Comparison

- What does it take to support **1.28TB/s** of performance?
  - Comparison of HMC to DDR3L-1600 and DDR4-3200

**Active Signals**
- DDR3 requires ~14,300
- DDR4 requires ~7,400
- HMC only needs ~2,160, HMC is ~85% less than DDR3

**Operating Power**
- DDR3 requires ~2.25KW
- DDR4 requires ~1.23KW
- HMC only needs ~350W, HMC is ~72% less than DDR4

**Board Space**
- DDR3 requires ~165,000 sq mm
- DDR4 requires ~82,500 sq mm
- HMC only needs ~8,712 sq mm, HMC is ~90% less than DDR4

**Assumptions:**
1DPC, (SR x4) RDIMMs, 6.2W/channel for DDR3 @ 12.8GB/s, 8.4W/channel for DDR4 @ 25.6GB/s 5W per Link for HMC @ 160GB/s, 143 pins/channel for DDR3, 148 pins for DDR4, 270 per HMC, RDIMM area equals 10mm pitch x 165mm long, HMC w/keep outs equal 1089 sq mm, CPU for RDIMMS = 65W, CPU for HMC = 95W, each CPU supports up to 4 channels.
High Performance Comparison (single link)

• What does it take to support **60GB/s** of performance?
  • Comparison of HMC to DDR3L-1600 and DDR4-3200

**Channels**
- DDR3 requires 5 channels
- DDR4 requires 3 channels
- **HMC only needs 1 Link**

**Board Area**
- DDR3 requires ~7,734 sq mm
- DDR4 requires ~3,843 sq mm
- **HMC only needs ~1,089 sq mm**

**Active Pins**
- DDR3 requires 670 pins
- DDR4 requires 345 pins
- **HMC only needs 72 pins**

**BW/pin**
- DDR3 ~90MB/pin
- DDR4 ~174MB/pin
- **HMC ~833MB/pin**

Assumptions: Same as previous example of 1.28TB/s Bandwidth
Agenda

- Inception – what is it
- Architecture
- Reliability
- Comparisons to other Memories
- Industry Adoption
  - HMC Consortium
- Summary
HMC Consortium (HMCC)

HMCC Mission
Promote widespread adoption and acceptance of an industry standard serial interface and protocol for Hybrid Memory Cube

- HMCC specification and technical detail confidential
- Adopters get early access to HMCC specification

Log into hybridmemorycube.org to become an adopter and get into the game!
HMC Consortium Launches to Positive Reviews

“HMC could lead to unprecedented levels of memory performance …”
– Electronics News

“HMC offers the potential to alleviate the [memory] bottleneck ….”
– EE Times

“…the Hybrid Memory Cube guys are solving a huge problem that’s been a pain point for the industry for a few years…
– GigaOm

“Micron created an entirely new category of memory…”
– Tom’s Hardware

http://www.hybridmemorycube.org/
Broad Industry Adoption

- Developer group led by industry giants:
  - More than 150 organizations pursuing adopter status
  - 49 fully registered Adopters to date:

  APIC Corporation
  Cadence Design Systems, Inc.
  Convey Computer Corporation
  Cray Inc.
  DAVE Srl
  Design Magnitude Inc.
  eSilicon Corporation
  Exablate Corporation
  Galaxy Computer System Co., Ltd.
  GDA Technologies
  GLOBALFOUNDRIES
  GraphStream Incorporated
  Huawei Technologies
  Infinera Corporation
  Inphi
  ISI/Nallatech
  LeCroy Corporation
  Luxtera Inc.
  Marvell
  Maxeler Technologies Ltd.
  Montage Technology, Inc.
  Netronome
  Northwest Logic
  Oregon Synthesis
  Science & Technology Innovations
  Suitcase TV Ltd
  Tongji University
  University of Heidelberg ZITI
  Arira Design
  Dream Chip Technologies GmbH
  Engineering Physics Center of MSU
  Ezchip Semiconductor
  Fujitsu Advanced Technologies Limited
  Juniper Networks
  LogicLink Design, Inc.
  New Global Technology
  OmniPhy
  SEAKR Engineering
  Tabula
  Teradyne, Inc
  UMC
  USC Information Sciences Institute
Agenda

- Inception – what is it
- Architecture
- Reliability
- Comparisons to other Memories
- Industry Adoption
  - HMC Consortium
- Summary
HMC - A Revolutionary Memory Shift

- Increased Bandwidth
- Power Efficiency
- Smaller Size
- Scalability
- Reduced Latency