Hot Chips 2017

Xilinx 16nm Datacenter Device Family with In-Package HBM and CCIX Interconnect

Gaurav Singh
Sagheer Ahmad, Ralph Wittig, Millind Mittal, Ygal Arbel, Arun VR, Suresh Ramalingam, Kiran Puranik, Gamal Refai-Ahmed, Rafe Camarota, Mike Wissolik
Virtex® UltraScale+™ HBM Family (VU3xP)

4th Gen 3D IC
- TSMC CoWoS
- 3 16nm FPGA die
- 2 HBM2 Stacks
- Lidless Package w/ Stiffener
- 55 mm Package (Die Area: Not Disclosed)

16nm TSMC FF+ FPGA
- HBM enabled with hard memory controller + hard switch
- 2.8M System Logic Cells
- 9024 DSP Blocks (18x27 MACs) @ 891 MHz
- 341 Mbit FPGA On-Die SRAM
- 4 DDR4-2666 x72 Channels
- 96 32.75Gbps Serdes
- 8 100G Ethernet MACs w/ RS-FEC
- 4 150G Interlaken MACs
- 6 PCIe Gen4 x8 Controllers (4 w/ CCIX)

2 HBM2 In-Package DRAM Stacks
- 1024 Bits @ 1.8 Gbps + ECC
- 8 Gbyte
### Virtex® UltraScale+™ HBM Family

<table>
<thead>
<tr>
<th>Device Name (Name)</th>
<th>VU31P</th>
<th>VU33P</th>
<th>VU35P</th>
<th>VU37P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Logic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System Logic Cells (K)</td>
<td>970</td>
<td>970</td>
<td>1,915</td>
<td>2,860</td>
</tr>
<tr>
<td>CLB Flip-Flops (K)</td>
<td>887</td>
<td>887</td>
<td>1,751</td>
<td>2,615</td>
</tr>
<tr>
<td>CLB LUTs (K)</td>
<td>444</td>
<td>444</td>
<td>876</td>
<td>1,308</td>
</tr>
<tr>
<td><strong>Memory</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max. Distributed RAM (Mb)</td>
<td>12.5</td>
<td>12.5</td>
<td>24.6</td>
<td>36.7</td>
</tr>
<tr>
<td>Total Block RAM (Mb)</td>
<td>23.6</td>
<td>23.6</td>
<td>47.3</td>
<td>70.9</td>
</tr>
<tr>
<td>UltraRAM (Mb)</td>
<td>90</td>
<td>90</td>
<td>180</td>
<td>270</td>
</tr>
<tr>
<td>HBM DRAM (Gb)</td>
<td>32</td>
<td>64</td>
<td>64</td>
<td>64</td>
</tr>
<tr>
<td>HBM AXI Ports</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td><strong>Clocking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clock Management Tiles (CMTs)</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td><strong>Integrated IP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSP Slices</td>
<td>2,880</td>
<td>2,880</td>
<td>5,952</td>
<td>9,024</td>
</tr>
<tr>
<td>PCIe® Gen3 x16 / Gen4 x8</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>CCIX Ports(2)</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>100G Ethernet w/ RS-FEC</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td><strong>I/O</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max. Single-Ended HP I/Os</td>
<td>208</td>
<td>208</td>
<td>416</td>
<td>624</td>
</tr>
<tr>
<td>GTY 32.75Gb/s Transceivers</td>
<td>32</td>
<td>32</td>
<td>64</td>
<td>96</td>
</tr>
<tr>
<td><strong>Speed Grades</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extended (1)</td>
<td>-1, -2L, -3</td>
<td>-1, -2L, -3</td>
<td>-1, -2L, -3</td>
<td>-1, -2L, -3</td>
</tr>
<tr>
<td><strong>Footprint</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimensions (mm)</td>
<td>HP I/O, GTY 32.75Gb/s</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H1924</td>
<td>45x45</td>
<td>208, 32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H2104</td>
<td>47.5x47.5</td>
<td>208, 32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H2892</td>
<td>55x55</td>
<td>416, 64</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

© Copyright 2017 Xilinx
Agenda

- Application Drivers
- HBM: Design Changes
- HBM: Package/Thermal Consideration
- CCIX: What is CCIX
- CCIX: How CCIX is supported
Datacenter

- Vision Processing (CNN/DNN)
  - Higher compute density (2.8MLCs, 9024 DSPs – 32 TOPs INT8)
- Natural Language Processing (LSTM, RNN)
  - Memory bandwidth (weights, fully-connected layers) 3.6Tbps
- Efficient Host interface
  - Multiple PCIe Gen4/CCIX ports
- Seamless heterogenous nodes
  - SVM with CCIX
- Memory expansion (CCIX)

400G Networking

- N ports @400G
  - x96 high bandwidth interfaces - 32.75Gbps
  - x8 100G MACs, 4x Interlaken MACs
  - 2.8M LCs for P4 packet processing
  - 3.6Tbps HBM2 packet buffering
Virtex® UltraScale+™ HBM (VU+HBM): Key Features

- Base 16nm FPGA Platform (GTY, DDR4, URAM, CMAC)
- Hard Memory Controller for HBM
- 230 GB/s Bandwidth per HBM
  1024 IO @ 1.8 GTps
- PCIe Hard IP with CCIX TL
- Hard AXI Switch for Unified and Flexible Addressing
- 4GB Density per HBM
  (4H x 8Gb)
## HBM Integration Benefits

### Bandwidth

<table>
<thead>
<tr>
<th></th>
<th>VU9P + DDR4</th>
<th>VU37P + HBM2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth</td>
<td>85 GB/s</td>
<td>460 GB/s</td>
</tr>
</tbody>
</table>

### Power

<table>
<thead>
<tr>
<th></th>
<th>VU9P + DDR4</th>
<th>VU37P + HBM2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power</td>
<td>~27 pJ/bit (memory + IO power)</td>
<td>~7 pJ/bit (memory + 2.5D I/O)</td>
</tr>
</tbody>
</table>

### Advantages

<table>
<thead>
<tr>
<th></th>
<th>VU9P + DDR4</th>
<th>VU37P + HBM2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advantages</td>
<td>High density, low cost/bit</td>
<td>Highest performance, lowest power/bit</td>
</tr>
</tbody>
</table>

### Disadvantages

<table>
<thead>
<tr>
<th></th>
<th>VU9P + DDR4</th>
<th>VU37P + HBM2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disadvantages</td>
<td>PCB footprint, signal integrity, power</td>
<td>Lower DRAM Capacity</td>
</tr>
</tbody>
</table>

- **5X** (Higher bandwidth)
- **4X** (Higher power efficiency)
HBM Integration
Virtex® UltraScale+™ HBM: HBM Subsystem

- **Xilinx Virtex UltraScale+ HBM**
  - Hardened memory controllers
  - Hardened switch interconnect w/ 32 AXI ports
  - Option to bypass memory controllers and/or switch interconnect

- **Pseudo Channels**
  - A pair share command and address bus
  - Separate data bus that switches at full frequency
  - 16 independent pseudo-channels per HBM
  - An HBM pseudo-channel can only access 1/16th of HBM device address space

- **Virtex UltraScale+ HBM Interface**
  - AXI interfaces to PL to provide unified access across HBM channels
  - AXI provides simultaneous Rd and Wr
Bandwidth considerations

- **HBM Subsystem Interface to Programmable Logic (PL) Fabric**
  - 16 256-bit AXI ports per HBM stack (32 ports per FPGA)
  - 20,000+ signals @ 450Mhz

- **HBM Bandwidth Distributed Throughout FPGA PL Fabric**
  - “Fingers” into the programmable fabric help distribute bandwidth
Hard HBM Memory Controller (HBM MC)

- **8x Hard MC per HBM Memory Stack**
  - One controller per 2x 64b (72b) pseudo channels
  - Two AXI ports per controller

- **HBM MC: Controller Features**
  - Half-rate (e.g. 450MHz for 1.8GT/s)
  - AXI interface: 256bit
  - Reordering support
  - 32Byte access granularity
    - 64DQ * BL4
  - ECC support (SECDED 64b/8b)
  - Parity protection on Data and Command with retry support
HBM Interface Performance Results

Example with 4 Masters and 4 HBM channels

Uniform random:
- Every master to all channels, with uniform random distribution
- Channels can be grouped to a ‘local’ group of 2, 4, 8, 16 or all 32

Point to point
- Each master to one channel, but can be any of the channels
- Linear or random addressing within a channel
- Channels can be grouped to a ‘local’ group of 2, 4, 8, 16 or all 32

Legends:
- UNR = Uniform Random
- LIN = Linear
- RND = Random
- PTP = Point to Point
- t256B = Transaction size of 256Byte
- PTPW = farthest neighbor
- RW/RO/WO = Read/Write/Read-only/Write-only

Typical results, synthetic access patterns show higher performance
Packaging
Test Chip addresses HBM integration challenges – some examples
- Incoming HBM residue on micro-bump addressed by IQC and process tuning
- 55x55mm package co-planarity improved to < 12 mil by appropriate substrate material selection and stiffener design
- Reliability challenge such as underfill crack addressed by stress tuning and process improvement – passing 1200 hour HTS and 1200 cycles TCB
- HBM max junction of 95C for long term operation is a challenge for package thermal budget and system level cooling

Passed HTS & TCB Stress
HBM Integration – Thermal Design

HBM can be 97°C Tj and HBM I/F 95°C Tj @30°C A
HBM gradient ~10°C (~2°C/Layer)

Air cooling requires attention to heat-sink design
HBM 8-Hi will be a challenge

PCI-e card: Full length/full height
Card power (4x VU3xP): 320W
Airflow: 15CFM
Typical ambient 30°C

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

@ 30°C Inlet ambient to PCI-e card
Why CCIX?

➤ Moore’s law is slowing down
➤ Heterogenous computing is the solution
  – CPU + FPGA
  – CPU + GPU
  – CPU + Intelligent NIC
➤ There is a need for an efficient interconnect for this heterogenous system
➤ Characteristics of this interconnect
  – High bandwidth: 25G → 32G→56G → 100G per lane
  – Low latency
  – Leverage existing ecosystem where possible
  – Optimized for short transfers as well
➤ But why coherency?
  – Simplified programming and data sharing model
  – Lower latency (no-driver)
  – Accelerator thread has same access to memory as CPU thread (Democratize memory access)
CCIX Summary

- High bandwidth IO
  - 25Gbps Gen1 (specification complete)
  - Backward compatible to 16Gbps and lower speeds

- Full capability in the accelerator
  - Accelerator-processor peer processing (homenode)
  - Caching capability
  - Memory expansion

- Flexible topology
  - 1 CPU to 1 accelerator
  - Option to connect multiple accelerators

- Optimized for multi-chip transfers
  - Low overhead header format
  - Message packing and simplified messaging
  - Request and Snoop chaining
  - Port Aggregation

- Full Ecosystem support
  - Interface IP available from Cadence, Synopsys
  - Coherent controllers from ARM, Netspeed, ArterisIP
  - Verification IP from Cadence, Synopsys, Avery Design Systems
  - How to join: www.ccixconsortium.com (33 members and counting)
System Topologies

Switched-Meshed-Direct Attach Accelerator- Direct Attached Memory
New CCIX Capable UltraScale+ PCIe Hard Block

Extends Xilinx 16nm UltraScale+ Hard Block for PCI Express 4.0
- Up to Gen4 8 Lanes or Gen3 16 Lanes
  - Compliant to PCIe Base Spec 4.0
- Feature Rich Transaction Layer
- SR-IOV, ATS, PRI Supported

New Supported Features
- Data Link Layer
  - Support for CCIX VC Initialization
- 2 VC CCIX Transaction Layer
  - CCIX Optimized TLP Mode supported
- CFG Space Module
  - 2 Virtual Channels, WRR based VC Arbitration
  - ATS, PRI Capabilities Structures
  - CCIX DVSECs
New CCIX Capable UltraScale+ PCIe Hard Block

➢ Extends Xilinx 16nm UltraScale+ Hard Block for PCI Express 4.0
  • Up to Gen4 8 Lanes or Gen3 16 Lanes
    – Compliant to PCIe Base Spec 4.0
  • Feature Rich Transaction Layer
  • SR-IOV, ATS, PRI Supported

➢ New Supported Features
  – Data Link Layer
    • Support for CCIX VC Initialization
  – 2 VC CCIX Transaction Layer
    • CCIX Optimized TLP Mode supported
  – CFG Space Module
    • 2 Virtual Channels, WRR based VC Arbitration
    • ATS, PRI Capabilities Structures
    • CCIX DVSECs

CCIX Transport Latency

CCIX Protocol Packet Efficiency
Virtex® UltraScale+™ HBM: Summary

- Scalable Family: 4 Devices 1-3 FPGA die, 1-2 HBM2 Stacks
- 4 Tbps (HBM2 + DDR4-2666): Weight Bandwidth for ML
- 32 TOPs INT8: Machine Learning Operations
- 3.6 Tbps HBM2: Packet Buffering for 400G Networking
- Coherent Low Latency Host Interface: CCIX
- Switchless Peer 2 Peer SVM: CCIX Heterogeneous Scale-Up
- 96 lanes of PCIe G4: 6 PCIe controllers, 4 CCIX controllers