Building Heterogeneous Chiplets with AMBA Interconnects

Jeff Defilippi, Senior Director Product Management
Feb 26, 2023
Building Heterogeneous Chiplets with AMBA Interconnects

- The scalable AMBA CHI (Coherent Hub Interface) specification has provided the foundation for building complex SoCs across a broad range of applications including mobile, networking, automotive and data centers. The high-speed, credited, packetized attributes make it well suited to build multi-chip(let) networks.

- This talk will highlight the scope of a new AMBA CHI Chip-to-Chip specification that is currently in development and illustrate how the specification will complement chiplet standards such as UCIe.
Economics Driving Chiplet Investments

- Increasing design cost with less benefit
  - Logic continues to scale, but IO & SRAM only shrinking by 5~10%

- SoC NRE limiting product derivatives
  - Chiplets lowers overall platform cost and barrier to deploy a diverse product portfolio

- Market demands performance and efficiency
  - New chiplet technology offers a 20x speed and power improvement over traditional PCIe SERDES

Source: IBS, as cited in IEEE Heterogeneous Integration Roadmap
SoC Accelerator Framework for Heterogenous Compute

Moving from Monolithic to Chiplet

SINGLE DIE
(AMBA)

MULTI-DIE
(UCIe)

MULTI-CHIP on PCB
(PCle, CXL)
The Path to a Chiplet Ecosystem

- PHY, Transport and Protocol standards drive IP development and broader adoption

Semi-Custom Chiplet Platform
- PHY, Transport and Protocol standards required for interoperability

Multi-Vendor Chiplet Ecosystem
- PHY, Protocol and Transport
- Mechanical, thermal & power
- Pre/post-silicon test & debug
- Software standards
- Silicon qualification and reliability
The Path to a Chiplet Ecosystem

- **Custom Chiplet Platform**
  - PHY, Transport and Protocol standards drive IP development and broader adoption

- **Semi-Custom Chiplet Platform**
  - PHY, Transport and Protocol standards required for interoperability

- **Multi-Vendor Chiplet Ecosystem**
  - PHY, Protocol and Transport
  - Mechanical, thermal & power
  - Pre/post-silicon test & debug
  - Software standards
  - Silicon qualification and reliability

*We are here*
## AMBA: Specifications, Interface and Protocols Diagram

### Key AMBA Specifications: CHI, ACE, AXI, AHB, APB

<table>
<thead>
<tr>
<th>Key AMBA specifications</th>
<th>AMBA</th>
<th>AMBA 2</th>
<th>AMBA 3</th>
<th>AMBA 4</th>
<th>AMBA 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CHI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CHI</td>
</tr>
<tr>
<td>Coherent Hub Interface</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Credited coherent protocol</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Layered architecture for scalability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ACE</strong></td>
<td></td>
<td></td>
<td></td>
<td>ACE</td>
<td></td>
</tr>
<tr>
<td>AXI coherency Extensions</td>
<td></td>
<td></td>
<td></td>
<td>ACE5</td>
<td></td>
</tr>
<tr>
<td>ACE is a superset of AXI – system-wide coherency across multicore clusters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>AXI</strong></td>
<td></td>
<td></td>
<td></td>
<td>AXI3</td>
<td></td>
</tr>
<tr>
<td>Adv. eXtensible Interface</td>
<td></td>
<td></td>
<td></td>
<td>AXI4</td>
<td></td>
</tr>
<tr>
<td>AXI supports separate A/D phases, bursts, multiple outstanding addresses, OoO responses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>AHB</strong></td>
<td></td>
<td></td>
<td></td>
<td>AHB</td>
<td></td>
</tr>
<tr>
<td>Adv. High-performance Bus</td>
<td></td>
<td></td>
<td></td>
<td>AHB+Lite</td>
<td></td>
</tr>
<tr>
<td>AHB supports 64/128 bit multi-managers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AHB-Lite for single managers</td>
<td></td>
<td></td>
<td></td>
<td>AHB+Lite</td>
<td></td>
</tr>
<tr>
<td><strong>APB</strong></td>
<td></td>
<td>APB2</td>
<td>APB3</td>
<td>APB4</td>
<td>APB5</td>
</tr>
<tr>
<td>Advanced Peripheral Bus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System bus for low bandwidth peripherals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Selected AMBA specifications only. Other specifications include: ATB, ATP, CXS, DTI, GFB, LPI and LTI.
## AMBA: Specifications, Interface and Protocols Diagram

### Key AMBA Specifications: CHI, ACE, AXI, AHB, APB

<table>
<thead>
<tr>
<th>Key AMBA specifications</th>
<th>AMBA</th>
<th>AMBA 2</th>
<th>AMBA 3</th>
<th>AMBA 4</th>
<th>AMBA 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CHI</td>
</tr>
<tr>
<td>Coherent Hub Interface</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ACE</td>
</tr>
<tr>
<td>AXI coherency Extensions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ACE5</td>
</tr>
<tr>
<td>AXI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ACE5</td>
</tr>
<tr>
<td>Adv. eXtensible Interface</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AXI3</td>
</tr>
<tr>
<td>AXI3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AXI4</td>
</tr>
<tr>
<td>AXI4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AXI5</td>
</tr>
<tr>
<td>AHB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AHB</td>
</tr>
<tr>
<td>Adv. High-performance Bus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AHB5</td>
</tr>
<tr>
<td>AHB +Lite</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AHB5 +Lite</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>APB</td>
</tr>
<tr>
<td>Advanced Peripheral Bus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>APB2</td>
</tr>
<tr>
<td>APB2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>APB3</td>
</tr>
<tr>
<td>APB3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>APB4</td>
</tr>
<tr>
<td>APB4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>APB5</td>
</tr>
<tr>
<td>APB5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Selected AMBA specifications only. Other specifications include: ATB, ATP, CXS, DTI, GFB, LPI and LTI.
NVIDIA Grace Hopper Superchip

- Combining Grace CPU (72x Arm Neoverse V2 CPUs) with Hopper GPU inside the package
- New 900 GB/s NVIDIA-CTC interconnect using AMBA CHI for the multichip protocol
- Unified virtual memory system with shared paged tables
- 30X Higher System Memory B/W to GPU in a Server
Extending AMBA Architecture to Multichip with CHI C2C

- New AMBA CHI Chip-to-Chip Specification in Development

- A unified interface for device attach
  - Compute, Accelerator and Memory
  - Coherent and IO Coherent accelerator traffic

- Architectural features extended across chip(let) boundaries
  - Virtualization (DVM), Interrupts, Timers
  - Telemetry and Resource Management
  - Security for trust, memory protection and Confidential Compute

- Composable with die re-use across different solutions
  - Appropriate for Multi-Die and Multi-chip through PCB
  - Supports UCIe and bespoke PHY options

Accelerator
Memory (DDR, HBM, etc)
IO (PCIe, CXL, etc)
CHI Protocol Overview

+ AMBA 5 CHI protocol features:
  - Full cache coherency model
  - Supports snoop filter and directory-based systems for snoop scaling
  - Definition for simpler IO and Memory interfaces

+ Other features:
  - Credited, non-blocking, high-speed transports
  - Configurable bus width - 128, 256 or 512 bits
  - Support for end-to-end Quality of Service (QoS)
  - Virtual memory management through Distributed Virtual Memory (DVM) operations

AMBA CHI Specification:
https://developer.arm.com/search?q=amba%205%20chi%20architecture%20specification
AMBA CHI Chip-to-Chip Specification

- Packetization of CHI channels with cross-chip credit exchange
- Link aggregation for bandwidth scalability
- Portability across different PHY and packaging technology
- Clean interface definition between on-chip interconnect and PHY/Link Layer for IP portability between vendors
- Support for point-to-point and chiplet network topologies
Link Layer

Physical channel

Protocol Layer

Packetization Layer

C2C Interface

Data-link Retry Buffer

CRC Check and Data-link Retry

CHI FLIT

Packetization Layer

RSP

DAT

SNP

MISC

Data-link Retry Buffer

Physical channel

CRC Check and Data-link Retry

CHI FLIT

Packetization Layer

RSP

DAT

SNP

MISC

Data-link Retry Buffer

Physical channel

CRC Check and Data-link Retry

CHI FLIT

Packetization Layer

RSP

DAT

SNP

MISC
Packetization Layer → Message class credit exchange → Packetization Layer

Protocol Layer

Link Layer

Packetization Layer

C2C Interface

CHI FLIT

Credit exchange

Physical channel

Data-link Retry Buffer

CRC Check and Data-link Retry
Align Industry around an open platform to enable chiplet based solutions

- Enables construction of SoCs that exceed maximum reticle size
  - Package becomes new System-on-a-Chip (SoC) with same dies (Scale Up)
- Reduces time-to-solution (e.g., enables die reuse)
- Lowers portfolio cost (product & project)
  - Enables optimal process technologies
  - Smaller (better yield)
  - Reduces IP porting costs
  - Lowers product SKU cost
- Enables a customizable, standard-based product for specific use cases (bespoke solutions)
- Scales innovation (manufacturing and process locked IPs)

Heterogeneous Integration Fueled by an Open Chiplet Ecosystem (Mix-and-match chiplets from different process nodes / fabs / companies / assembly)

- 20X I/O Performance at 1/20th Power vs off-package SerDes at Launch
- Gap more prominent with better on-package technologies in future
UCIe Layered Architecture

- UCIe 1.0 provides a complete multichip interface specification
  - Leverages PCIe/CXL protocol for architectural agnostic attach
  - PHY definitions for 2.1D and 2.5D packaging

- AMBA CHI C2C will use the UCIe Streaming Interface
  - Streaming interface with additional flit formats provide link robustness using UCIe defined data-link CRC and retry

- Arm will support both CHI and CXL/PCIe options to provide flexibility for the wide range of potential chiplet integrations

- UCIe defines support for both 2.1D and 2.5D packaging

https://www.uciexpress.org/webinars
AMBA CHI C2C Extends on-chip performance to Multi-chip(lets)

- Economics are driving chiplet investments and momentum toward Semi-Custom Chiplet Platforms
  - Standards investments for PHY, Transport and Protocol IP accessibility and interoperability
  - New PHY and packaging technology offering ‘on-chip’ performance and efficiency; 20x speed and power improvement over traditional PCIe SERDES

- New AMBA CHI Chip-to-Chip specification will extend the on-chip bus to multi-chip
  - Provides a unified memory, security and virtualization architecture across chiplets
  - Built to run over standards based die-to-die interconnects such as UCIe or bespoke solutions

- Stay tuned!
  - AMBA CHI C2C specification is under development, more updates expected throughout the year