OpenCAPI

OpenCAPI™ Overview

October 14, 2016

Topics

OpenCAPI

- Industry landscape
- Approaches to address challenges

OpenCAPI Consortium

- What is the OpenCAPI Consortium
- Who can join and participate

OpenCAPI Technical Details

- What is OpenCAPI, today and looking forward
- Benefits of OpenCAPI

Industry Landscape

- Two major technology trends will heavily impact the industry
  - Hardware acceleration will become commonplace as microprocessor technology and design continues to deliver far less than the historical rate of cost/performance improvement per generation
  - New advanced memory technologies will change the economics of computing
- Existing system interfaces are insufficient to address these disruptive forces
  - Traditional I/O architecture results in very high CPU overhead when applications communicate with I/O or Accelerator devices at the necessary performance levels
  - Systems must be able to integrate multiple memory technologies with different access methods and performance attributes
- These challenges must be addressed in an open architecture allowing full industry participation
  - Establish sufficient volume base to drive cost down
  - Support broad ecosystem of software and attached devices

OpenCAPI Approach

- **What is OpenCAPI?**
  - OpenCAPI is an Open Interface Architecture that allows any microprocessor to attach to
    - Coherent user-level accelerators and I/O devices
    - Advanced memories accessible via read/write or user-level DMA semantics
    - Agnostic to processor architecture
- **Key Attributes of OpenCAPI**
  - High-bandwidth, low latency interface optimized to enable streamlined implementation of attached devices
    - 25Gbit/sec signaling and protocol built to enable very low latency interface on CPU and attached device
    - Complexities of coherence and virtual addressing implemented on host microprocessor to simplify attached devices and facilitate interoperability across multiple CPU architectures
  - Attached devices operate natively within an application's user space and coherently with processors
    - Allows attached device to fully participate in application without kernel involvement/overhead
  - Supports a wide range of use cases and access semantics
    - Hardware accelerators
    - High-performance I/O devices
    - Advanced memories
  - 100% Open Consortium / All company participants welcome / All ISA participants welcome

OpenCAPI Consortium

- Open Forum to Manage the OpenCAPI Specification and Ecosystem
- Founded by AMD, Google, IBM, Mellanox, and Micron
- Initial deliverables
  - OpenCAPI 3.0 specification
  - OpenCAPI interface reference design for FPGA accelerators
  - Additional Enablement (eg documentation, simulation environment)
- Workgroups will be formed to evolve architecture and ensure compliance
- Technical details and membership info at www.opencapi.org

OpenCAPI Participation

- **Who should join OpenCAPI?**
  - Microprocessor vendors looking for a better way to attach high performance devices
  - Hardware Accelerator developers looking for a better way to attach to systems
  - Software developers looking for an easier and more efficient way to integrate hardware accelerators into applications
  - System Vendors looking for more ability to innovate and differentiate
  - End users looking for choices to exploit new technologies and improve datacenter cost/performance

- **How do I take advantage of OpenCAPI and get involved?**
  - Join the OpenCAPI Consortium
  - Once a member has joined then they can take advantage of all the components that are part of www.opencapi.org toolbox

- **JOIN TODAY!!**

Addressing the Industry Trend

- **Strong industry growth in use of various accelerators**
  - Introduction of device coherency requirements
  - Storage and Memory, Compute, Network
  - Various form factors including GPUs and FPGAs

- **Bottom-up Design with Radical New Capabilities is required**
  - Hyperscale datacenters and HPC are driving need for much higher network bandwidth
  - Deep learning and HPC require more bandwidth between accelerators and memory
  - New storage technologies are increasing the need for bandwidth and CPU efficiency
  - Increased industry dependence on hardware acceleration for performance
  - OpenCAPI addresses these needs by providing higher bandwidth and lower latency

OpenCAPI 3.0 Features

Base Accelerator Support

- Accelerator Reads with no intent to cache, DMA write using Program Addresses
  - The accelerator is working in the same address domain as the host application
    - Pointer chasing, link lists are all now possible without Device Driver involvement
  - Address translation on host (processor) with error response back to the accelerator
    - Very efficient translation latency mechanism using host processor Address Translation Cache (ATC) and MMU
  - Non-posted writes only
  - Ability for Partial Read/Write DMAs
    - Write with byte enables
- Translate touch to warm up address translation caches
  - Allows accelerator to reduce translation latency when using a new page
- Wake Up host thread
  - Very efficient low latency mechanism in lieu of either interrupts or host processor polling mechanism of memory
- Atomic Memory Operations (AMO) to Host Processor Memory
  - Accelerator can now perform atomic operations in the same coherent domain just like any other host processor thread

Virtual Addressing

- **An OpenCAPI device operates in the virtual address spaces of the applications that it supports**
  - Eliminates kernel and device driver software overhead
  - Improves accelerator performance
  - Allows device to operate directly on application memory without kernel-level data copies or pinned pages
  - Simplifies programming effort to integrate accelerators into applications

- **The Virtual-to-Physical Address Translation occurs in the host CPU**
  - Reduces design complexity of OpenCAPI-attached devices
  - Makes it easier to ensure interoperability between an OpenCAPI device and multiple CPU architectures
  - Since the OpenCAPI device never has access to a physical address, this eliminates the possibility of a defective or malicious device accessing memory locations belonging to the kernel or other applications that it is not authorized to access

OpenCAPI 3.0 Features (cont.)

Base Accelerator Support

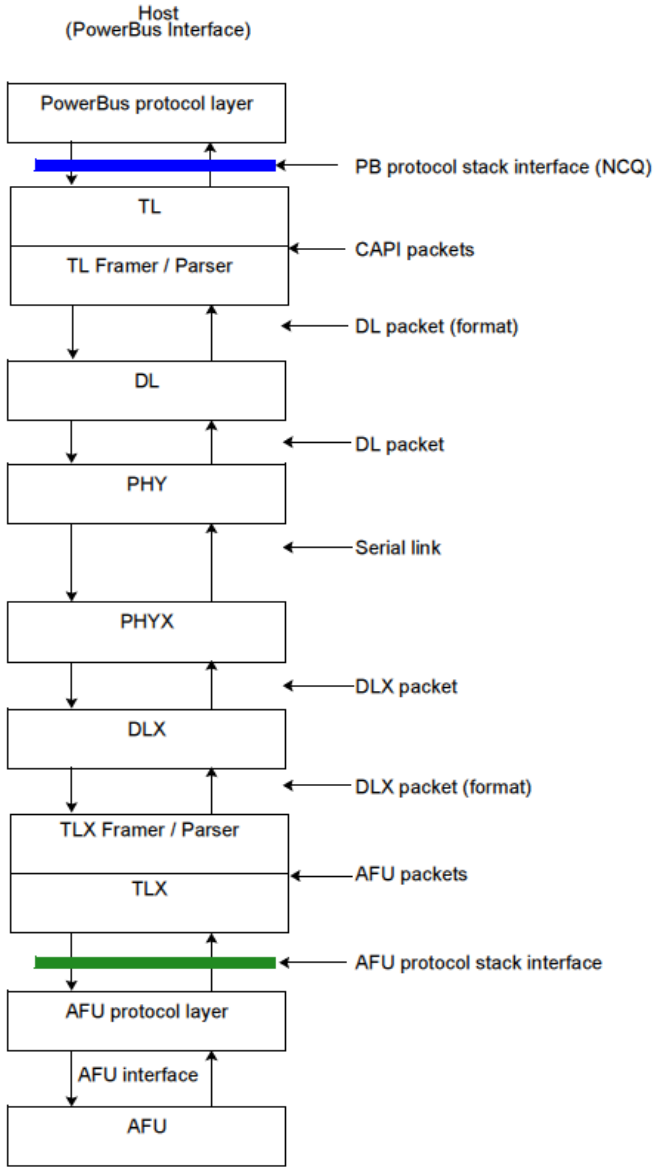
- MMIO slave
  - Accelerators have BAR space that provide MMIO register capability
- Configuration space facility slave
  - Efficient discovery and enumeration of accelerators
- OpenCAPI attached Memory
  - High bandwidth and low latency memory home agent capability
  - Load/Store model access to OpenCAPI attached memory
  - Host Application can access memory attached to OpenCAPI endpoint as part of coherent domain
  - Data resides very close to the consumer with very low latency
  - Atomic Memory Operations (AMO) support toward OpenCAPI attached memory

OpenCAPI 4.0 Features (future)

Full Feature of the Architecture Specification

- Accelerator Caching/Coherence support
  - Enabling application to have a Program Address Cache on accelerator chip with Host Address proxy directory and translation on host processor chip
  - Caching on accelerators provides latency advantage for frequently addressed/modified data
- Improved latency and ordered write performance from accelerator to host memory
- Pinned translations in host processor Address Translation Cache (ATC)
- Enhanced wake host thread with rollover to interrupt
  - Low latency communication method between accelerator and host application.
  - Avoid inefficient interrupts and host processor polling on memory
- Enhanced OpenCAPI attached memory
  - Host/Accelerator sharing of OpenCAPI attached memory
  - Atomic memory operations executed in attached memory controller on accelerator
- Link Width of x4, x8, x16, x32 support (OpenCAPI 3.0 supports x8 only)

OpenCAPI protocol stack



- TL
- DL
- *X remote chip (FPGA or ASIC component)

- DL/TL not symmetrical with Accelerator DLX/TLX
- Accelerator protocol layer optional
  - Advanced accelerators can i/f directly to TLX

FPGA Toolbox

Hardware to enable coherent acceleration

- Architecture Specs
- TLx and DLx Reference Designs

Operating System Enablement

- Reference Driver that partner can use as starting point or develop their own

Custom application and accelerator development

- OpenCAPI Simulation Environment (OCSE) to support TLx and DLx Ref Designs

FPGA TLX and DLX Reference Designs

- TLX and DLX will be provided as reference designs to OpenCAPI consortium members
- TLX and DLX are not symmetric with TL and DL that are on the host processor
- 64B flit flow at 400MHz
- TLX presents Accelerator interface
  - Very efficient thin layer in a packet based format
    - Manages Credits
    - Accelerator master <command, address, data>
    - Accelerator memory slave read and write w/Atomics
    - Accelerator maintenance
      - MMIO slave
      - Configuration, initialization
    - Error handling