Reusable Arm Computational Chiplet with 128bit Shared Global Address Space Access

Professor John Goodacre

University of Manchester

12 July 2019

Project ID: 754337

Intro to key terms

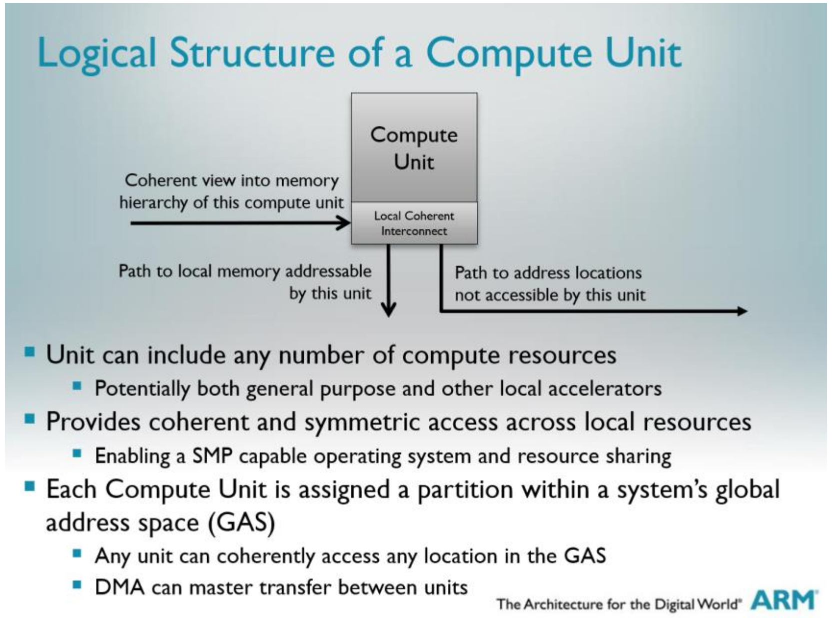
- Compute Unit: The new processing building block
- Chiplet: Silicon module that implements a Compute Unit
- Unimem: The interconnect rules between Compute Units
- This talk: delivering something that is all of the above

Compute Units: circa 2013

13th International Forum on Embedded MPSoC and Multicore
July 15-19, 2013, Otsu, Japan

NEW Slides now available

MPSoC'13 will be held on
July 15-19, 2013
at Biwako Hotel
Otsu, Japan



Logical Structure of a Compute Unit

Compute Unit
Local Coherent Interconnect
Coherent view into memory hierarchy of this compute unit
Path to local memory addressable by this unit
Path to address locations not accessible by this unit

- Unit can include any number of compute resources
  - Potentially both general purpose and other local accelerators
- Provides coherent and symmetric access across local resources
  - Enabling a SMP capable operating system and resource sharing
- Each Compute Unit is assigned a partition within a system's global address space (GAS)
  - Any unit can coherently access any location in the GAS
  - DMA can master transfer between units

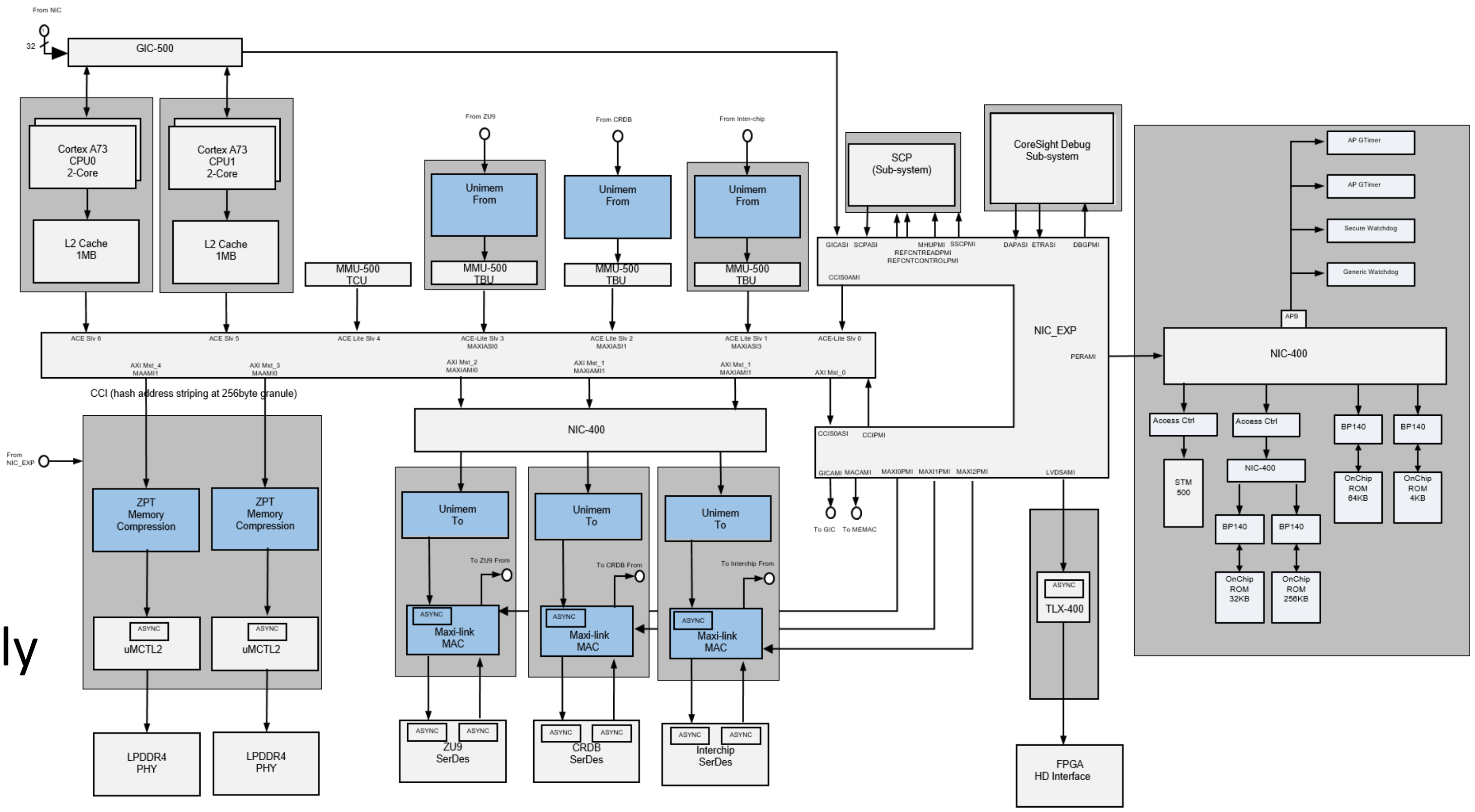
The Architecture for the Digital World® ARM

- Introduced the concept of the Compute Unit
  - Encapsulating a local interconnect and devices
  - Providing the model to link to other units

- The base for the EU FP7 project EuroServer
  - Compute Units implemented as chiplet
  - Multiple chiplets within a package
  - Global address space was a partition of local space
  - Further work across two more generations of H2020 projects

EuroExa Compute Unit Chiplet

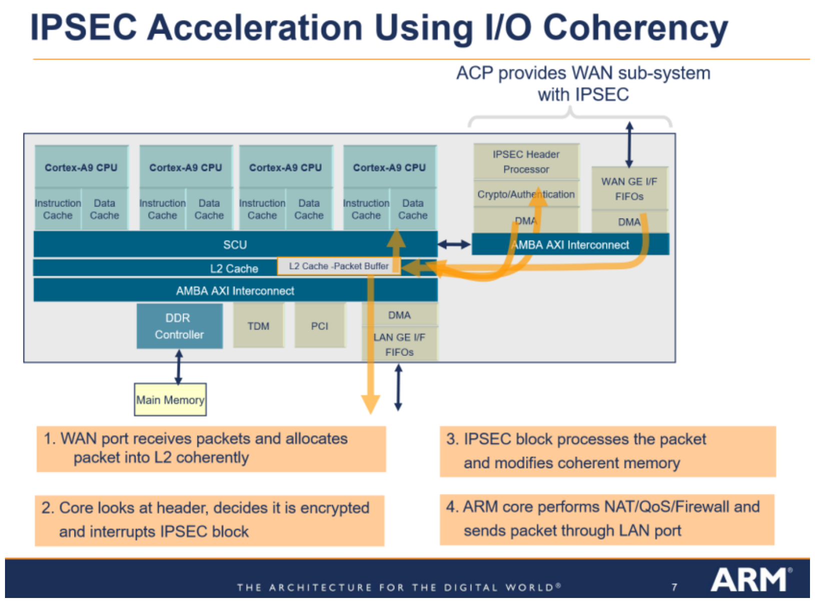
- 4x Cortex A-73 CPUs
  - 1.6 GHz
  - 64 L1 Caches
  - 2MB L2 Caches
  - 8 GB LPDDR4 DRAM
  - 3 MAXILINK Interfaces:
  - 64Gbps ZU9-EG Link
  - 64Gbps VU9 Link (via Firefly)
  - 64Gbps ASIC2ASIC Link (via Firefly
- H/W Support for UNIMEM via bridging devices



Pre-Unimem: circa 2008

8th International Forum on Application-Specific Multi-Processor SoC
23 - 27 June 2008, Aachen, Germany

SLIDES now available PHOTOS now available



IPSEC Acceleration Using I/O Coherency

ACP provides WAN sub-system with IPSEC

1. WAN port receives packets and allocates packet into L2 coherently
2. Core looks at header, decides it is encrypted and interrupts IPSEC block
3. IPSEC block processes the packet and modifies coherent memory
4. ARM core performs NAT/QoS/Firewall and sends packet through LAN port

THE ARCHITECTURE FOR THE DIGITAL WORLD® 7 ARM®

- Introduced the idea of dual interconnects interacting through a single sided coherence scheme
  - “Bridging” was integrated into the CPU (A9’s ACP)
  - There was no translation between the two address spaces

- The ACP “almost coherent port” (thanks Xilinx!)
  - Became key capability of future Arm interconnects
  - Made the A9 effectively the first Compute Unit
  - (and made the Xylinx Zynq great for prototyping)

ExaNest Node (quad units)

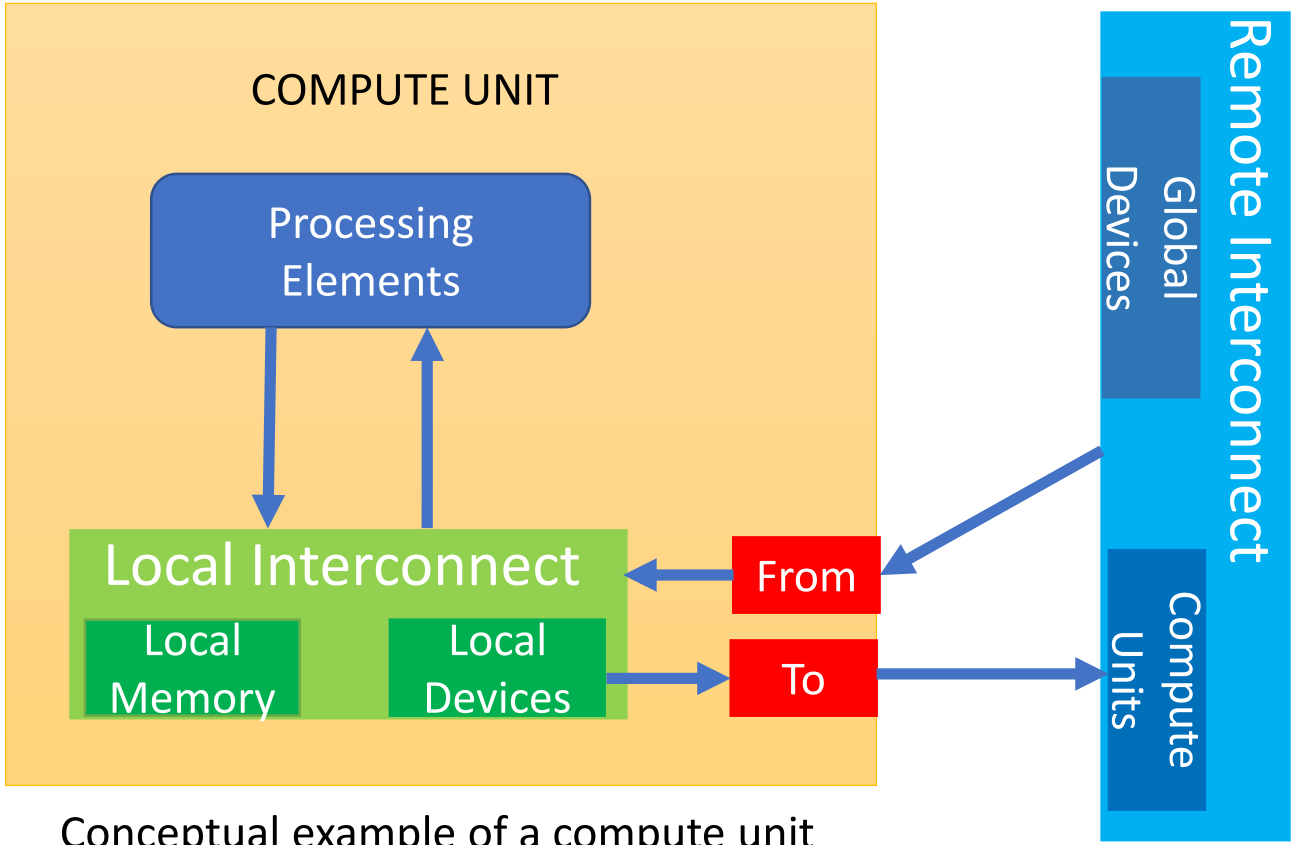
Formalising the Compute Unit

- A compute unit is defined by :
  - Zero of more of Processing Elements (masters that can access Unimem)
  - A local address maps containing
    - Zero or more other masters
    - Zero or more of Devices (slaves)
    - One or more memory regions (Slaves in the bigger address map)
    - A (to remote) Bridging Device to map the local addressed transaction into a remote addressed transaction
    - A (from remote) Bridging device able to map a Remote Address transactions into a local address transaction coherently

- A Unimem System consists of multiple addressable Compute Units

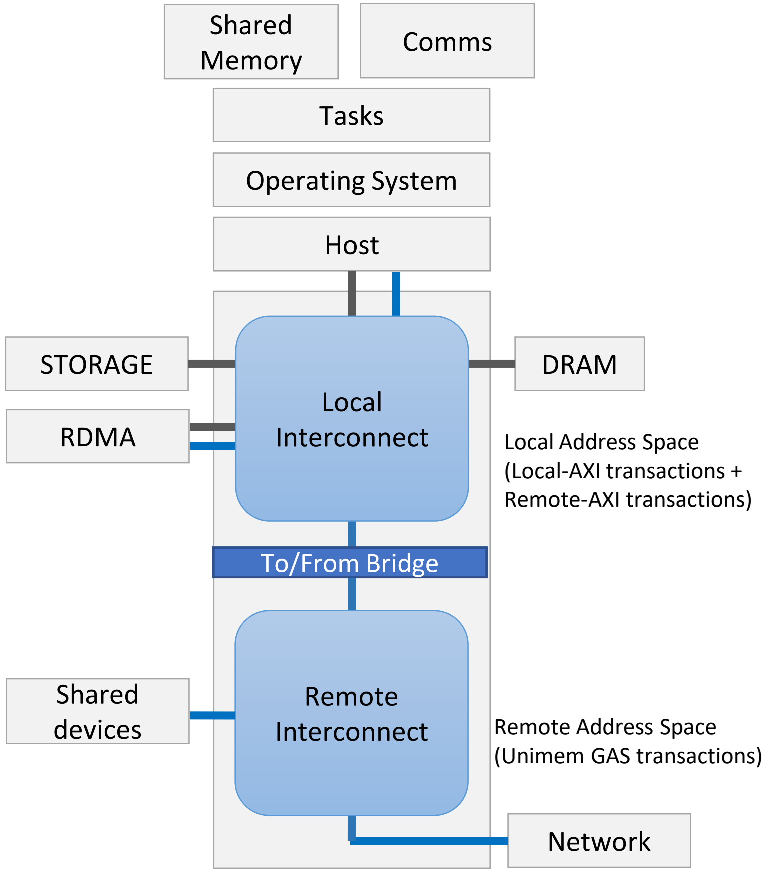
- The remote address space include
  - One or more compute units
  - Zero or more devices (slaves)

- Remote (Unimem) address are identified by a Global Address (GA), consisting of
  - A Global Virtual Address (GVA) within a Global Virtual Address Space (GVAS)
  - a set of progressive coordinates to locate the compute unit owning the GVA
  - a context to provide allocation and security isolation between units accessing regions within the GVAS (protection domain)



Conceptual example of a compute unit

Current implementation



- A processing node (of any size and complexity), that exposes a Unimem bridge is known as a Compute Unit
- Introduces an additional “global” remote address space that can be addressed natively by hardware level r/w transactions
  - Only the data-owner can cache globally shared memory (data locality)
  - Enables nodes to read/write data coherently with the data-owner
  - Provides native hardware level one-sided (read/write/atomic) communication
- Apps can use RDMA to generate both local and remote transactions (prototyped API exposed by ExaNode/ExaNeSt/EuroExa)
  - Block move data between local address space and remote address space
- EuroExa adds support for CPU to natively generate remote transactions
  - Hence apps can also natively access remote address space

AXI is a standard hardware protocol for read/write/atomic transactions published by Arm

Global Address Encoding

	Component/Field	Width	Comment
Global Virtual Address 64 bits)	Offset	30	1GB page sizes
	Page within 48 bit GVAS	18	2^18 256TB of memory per context
	Address Extension (AE)	16	Meaning defined by AE Use field
Location coordinates 16 bits (used for interconnect routing)	L1: A specific Bridge within a Node	2	4 (2 SoCs, 2 FPGAs)
	L2: Which Node within a quadrant	2	4 Nodes per quadrant
	L3: Which quadrant on a blade	2	4 Quads per blade
	L4: Which blade in a NetGroup	2	4 blades per netgroup
	L5: Which netgroup in a rack	3	8 Net-groups per rack
	L6: Which rack within a rackgroup	3	8 racks per rackgroup
	L7: Which rackgroup in a System	2	4 rackgroups per system (rows of containers)
User bits	Context	16	bridge to bridge defined meaning
	Encoding Versioning	4	0 means 48 bit GVA and location bit (AE unused) 1 means full 64 bit GVA 2 means “FORTH encoding in AE” (loc. Bits unused) 3-15 RFU
	Operation	4	Eg atomics within a specific context 0 = r/w, 1 = mailbox, 2 = stream, 3 =- lock, etc…
	Reserved for Future Use	8	…more locations, or bridges within a nodes etc
	Total	**128**	**One day maybe the native of scheme for a PE !**

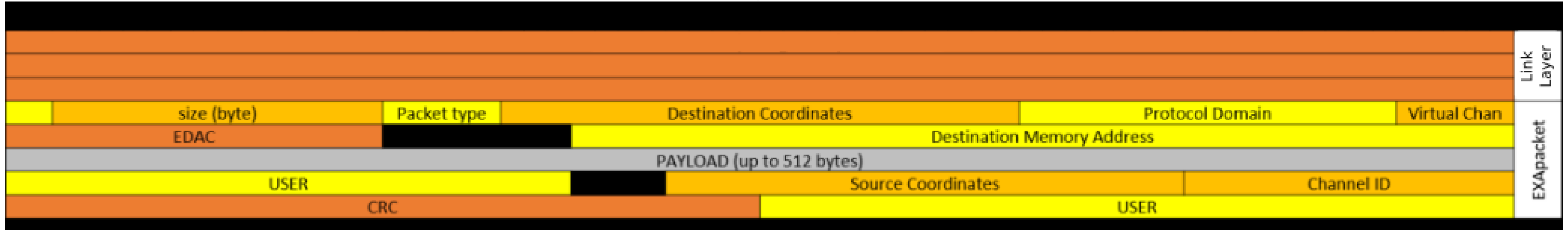
Unimem Bridges

- Each Compute Unit interfaces with the Unimem System via a *Bridges*
- Bridges perform two roles:
  - **Address Translation**
    - Local to Global
  - **Exanet Communication**
    - Exanet Packet Assembly and Disassembly
- Bridges can also be used to directly interface with rDMAs and implement operations directly in H/W
  - E.g. Atomic Operations, Stream transactions,

Unimem-To Bridge

- Translates local virtual-addresses to global Unimem addresses
- Generates Exanet Packets* for a range of operations:
  - Native Load/Store AXI Transactions
  - Command Blocks that are used to create custom operations:
    - Atomics, Fetch and Add etc
  - Interacts with rDMA to perform large memory operations

*Exanet Packet: over the wire encoding and protocol for Unimem



Link Layer

size (byte) | Packet type | Destination Coordinates | Protocol Domain | Virtual Chan

EDAC | Destination Memory Address

PAYLOAD (up to 512 bytes)

USER | Source Coordinates | Channel ID

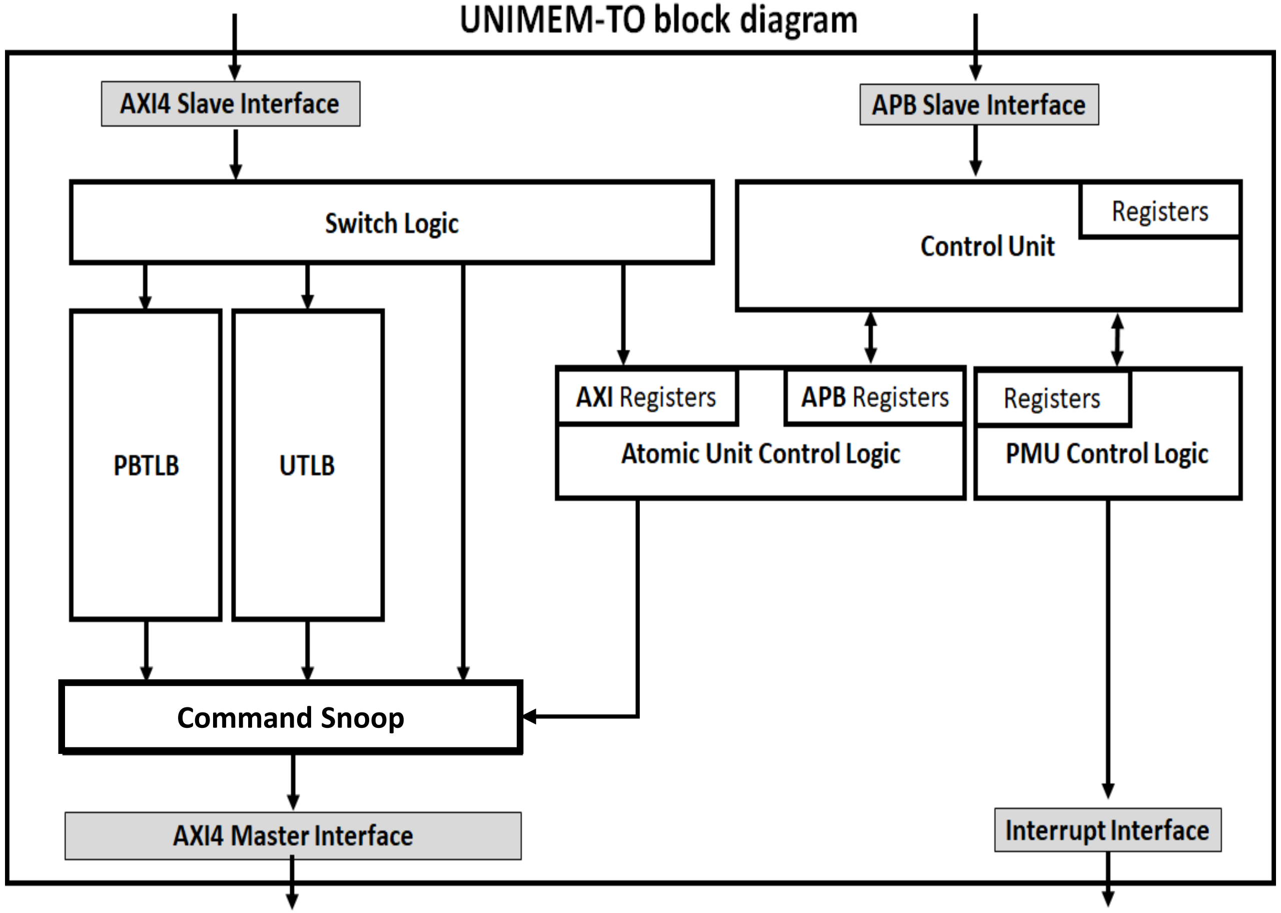
CRC | USER

EXApacket

CPU native UNIMEM-To bridge

- Maps ARMv8 physical addresses to UNIMEM Global Addresses
  - O/S Maps Global UNIMEM pages to a separate 40bit Physical Address Space
- Module Contains two Address Translation Caches
  - Unimem TLB (UTLB).
  - Page Borrowing TLB (PBTLB).
  - Support for 4K, 2MB, 1GB pages.
- Supports address translation for atomic accesses and custom ops via local "Command Block"

UNIMEM-TO block diagram



ASIC Native Unimem-From Bridge

- Disassembles Exanet Packets
- Translates Global Physical Addresses to Local Virtual Addresses
- Can initiate Local Memory-Reads and Writes
- Interacts with Custom Hardware (e.g. rDMA) to perform complex operations
- Generates interrupts for S/W callback functions for operations not handled by hardware

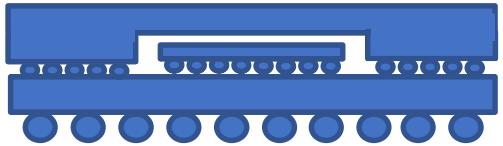
EUROEXA
FUNDED BY THE EUROPEAN UNION

Compute near memory Package

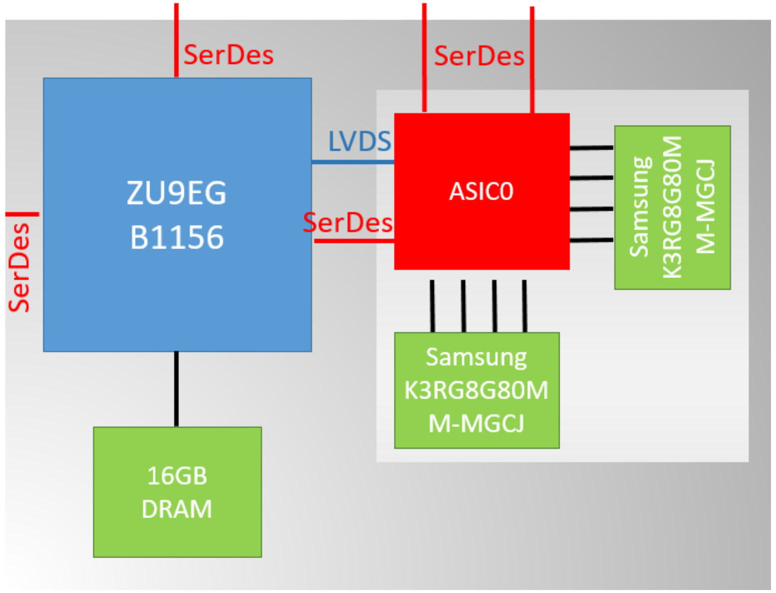
- EuroEXA Compute Unit ASIC:
  - High-end Energy-Efficient Arm CPU
  - Native support for UNIMEM Global Addressing
  - Novel Memory Compression

- Networking and RDMA/UNIMEM resources provided via Xilinx Zynq FPGAs

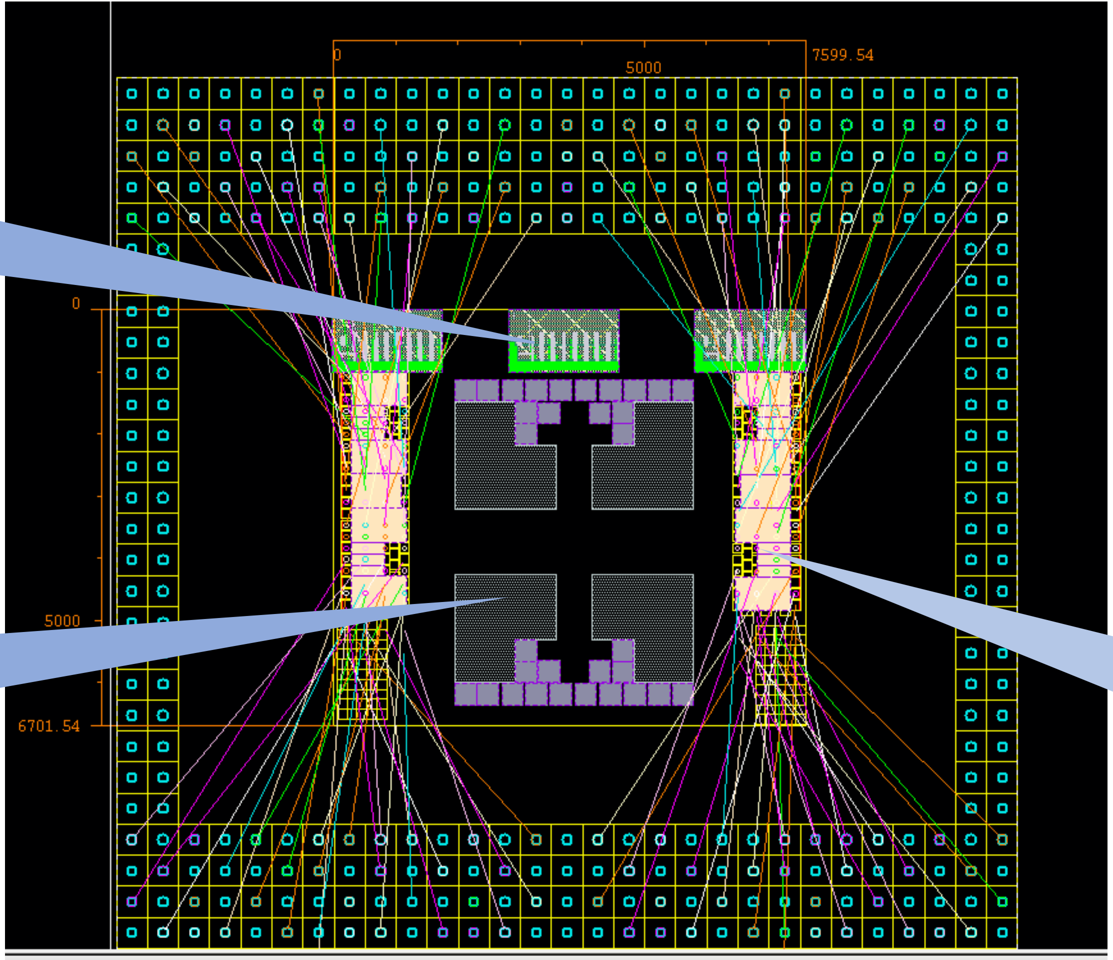
- Accelerator resources provided by Xilinx VU9s
- UNIMEM Capabilities shared between ASIC and FPGA
- Supports novel DRAM inline compress scheme
  - See EuroExa partner https://wp.zptcorp.com/



Stacked DRAM
Native Compute Unit ASIC
Low cost, low
pin count package



Package over die on substrate



SerDes
Used for Unitmem
To/From bridges

Arm Cortex A7x
Arm interconnect
providing required
Coherence

4 x channels of
LPDDR4
Connected within
package

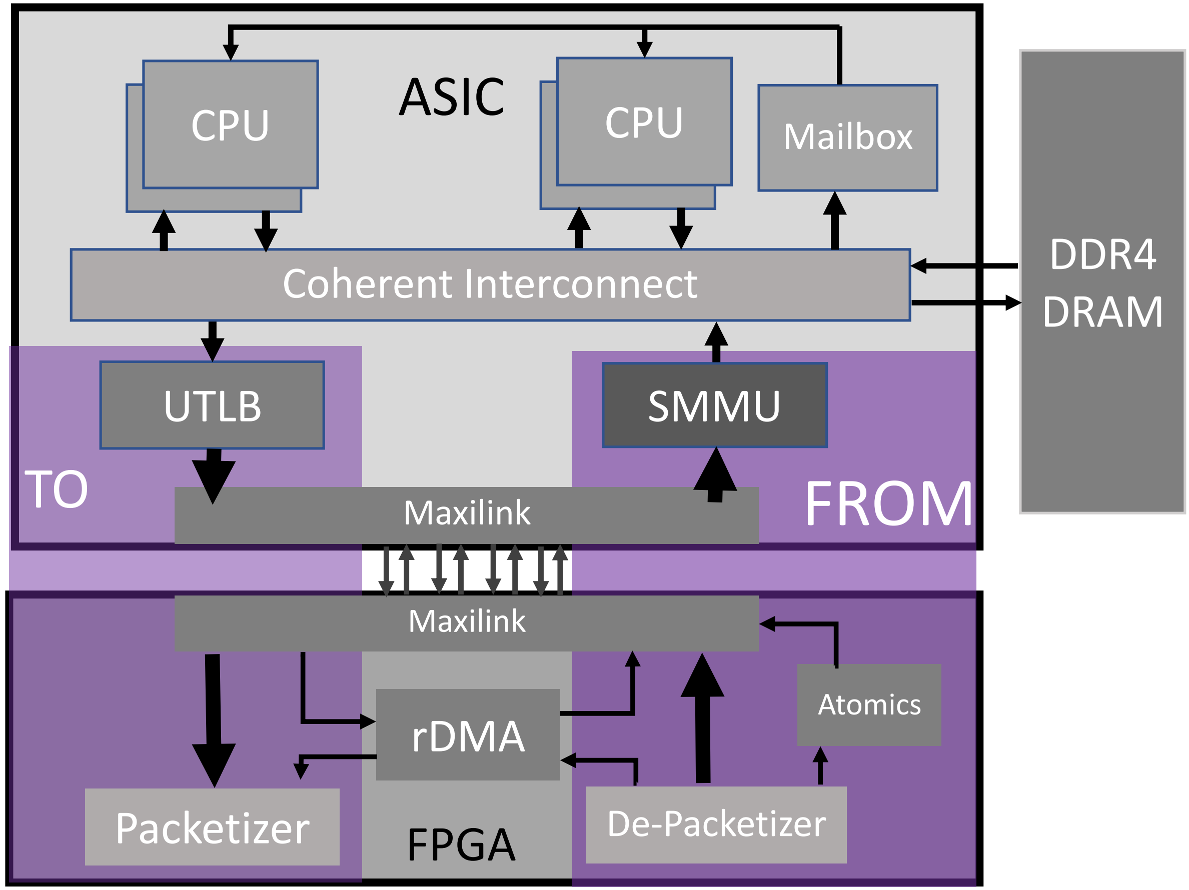
HBM was too expensive

Only ASIC/PHY pins required, memory interconnect is in package

Using PoP DRAM directly on ASIC package substrate

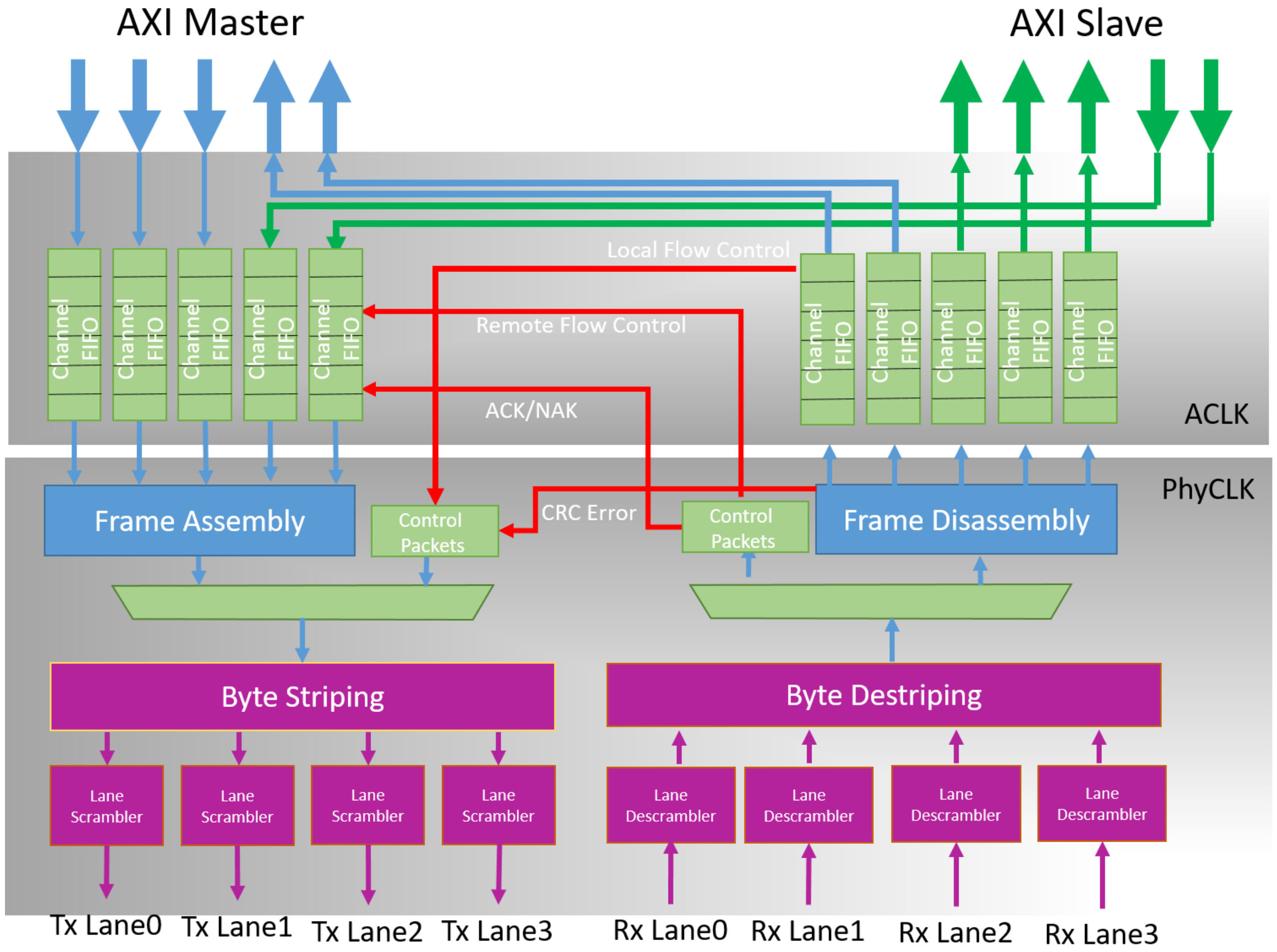
EuroEXA Implementation

- Bridges Split Between ASIC and FPGA
- Address Translation on ASIC
  - Exploits Memory-Management H/W
    - Fast: TLBs, PTW
    - Secure: Protected Access
- Exanet Communication on FPGA
  - Allows H/W to be prototyped easily
- Custom High Speed Link in-between
  - Maxilink: AXI serialised encoding of both local interconnect and remote unimem transactions

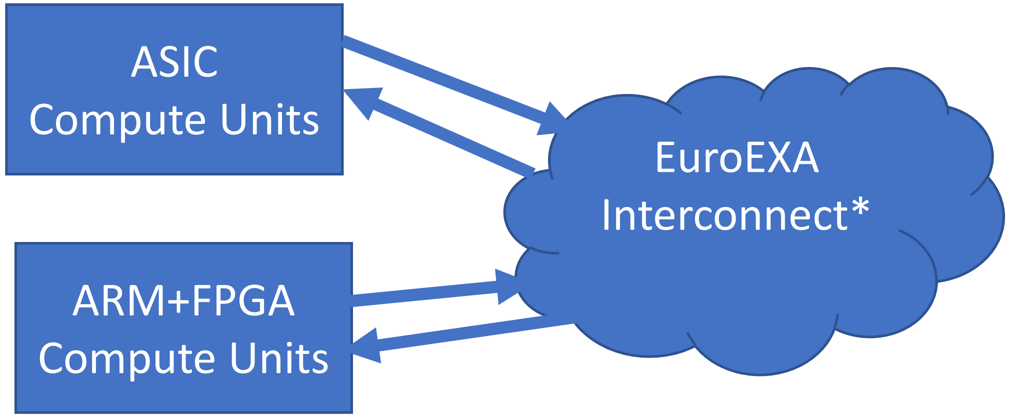


MAXI-LINK MAC

- Low Latency AXI Interconnect
- Serializes Master and Slave interface channels over Serdes
  - 16Gbps per Channel
- In-place Frame Buffers
- Two-stage Synchronization Handshake
  - Error-based Resync
- CRC Error detection and retransmission
- Per-channel Flow Control
- Clock-Correction

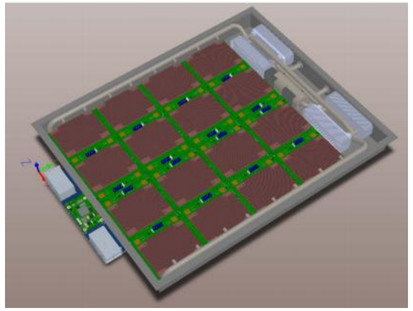


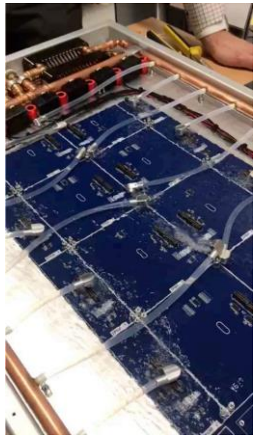
Testbed deployment



100’s of compute units
1000’s of CPU cores
1,000,000’s of FPGA DSP slices
1,000’s of Tera (INT8) OP/s

800 Gb/s per 1 rack unit
16 compute units per 1U
Deploying 2 cabinets
Using modular containers
Liquid cooled





- Deploying at STFC in UK
- Less than 2 yrs until operateration / demos

*Paper461: Design Exploration of Multi-tier interconnects for Exascale systems
ICPP 2019, 48th International Conference on Parallel Processing
August 5-8, 2019, Kyoto Research Park, Kyoto, Japan

Many Thanks

www.euroexa.eu

Project ID: 754337