Dedicated On-chip Network Hierarchies for Algorithm Acceleration

Kent Orthner, Travis Johnson

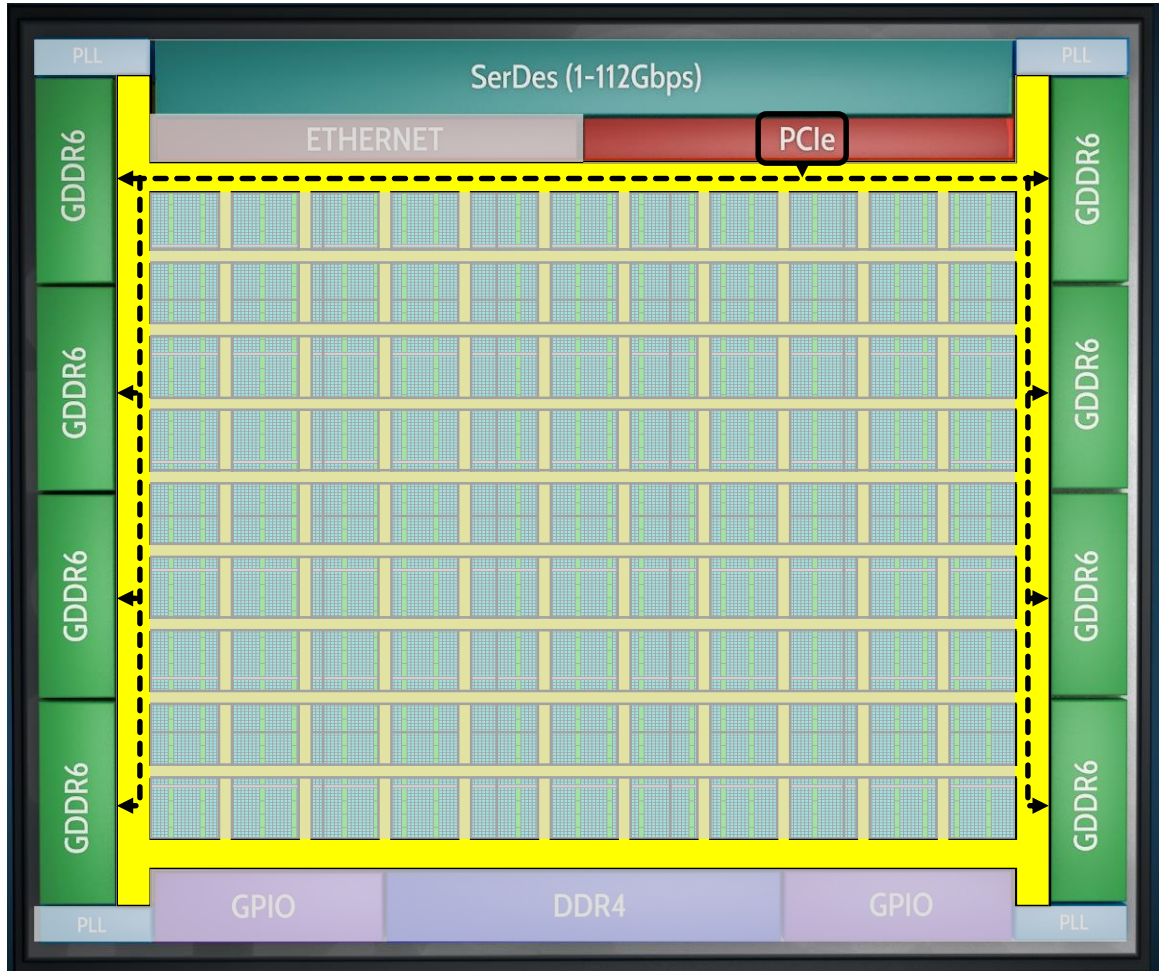
Problem Statement

- Modern acceleration workloads require
  - High speed interfaces
  - Memory bandwidth
  - Efficient data movement inside the device

- Example: Achronix Speedster 7t FPGA
  - 4x 400 Gbps Ethernet connections (1.6 Tbps)
  - 2x 16-lane PCIe Gen5 (512 Gbps)
  - 8x GDDR6 + 2x DDR4 (4.5 Tbps)

Speedster7t: High Bandwidth Network Hierarchy

- Traditional AXI interconnect in a “Ring” around the outside of the FPGA core
  - Master/Slave interfaces with Read & Write transactions
  - Supports connectivity between interfaces
    - E.g. PCIe Gen 5 to GDDR6, DDR4
    - All Configuration and Status interfaces.
    - Support FPGA programming and configuration traffic.
  - Allows any peripheral to access multiple access points into and out of the FPGA core.
- For example
  - PCIe accesses external memory without needing the FPGA fabric core at all.



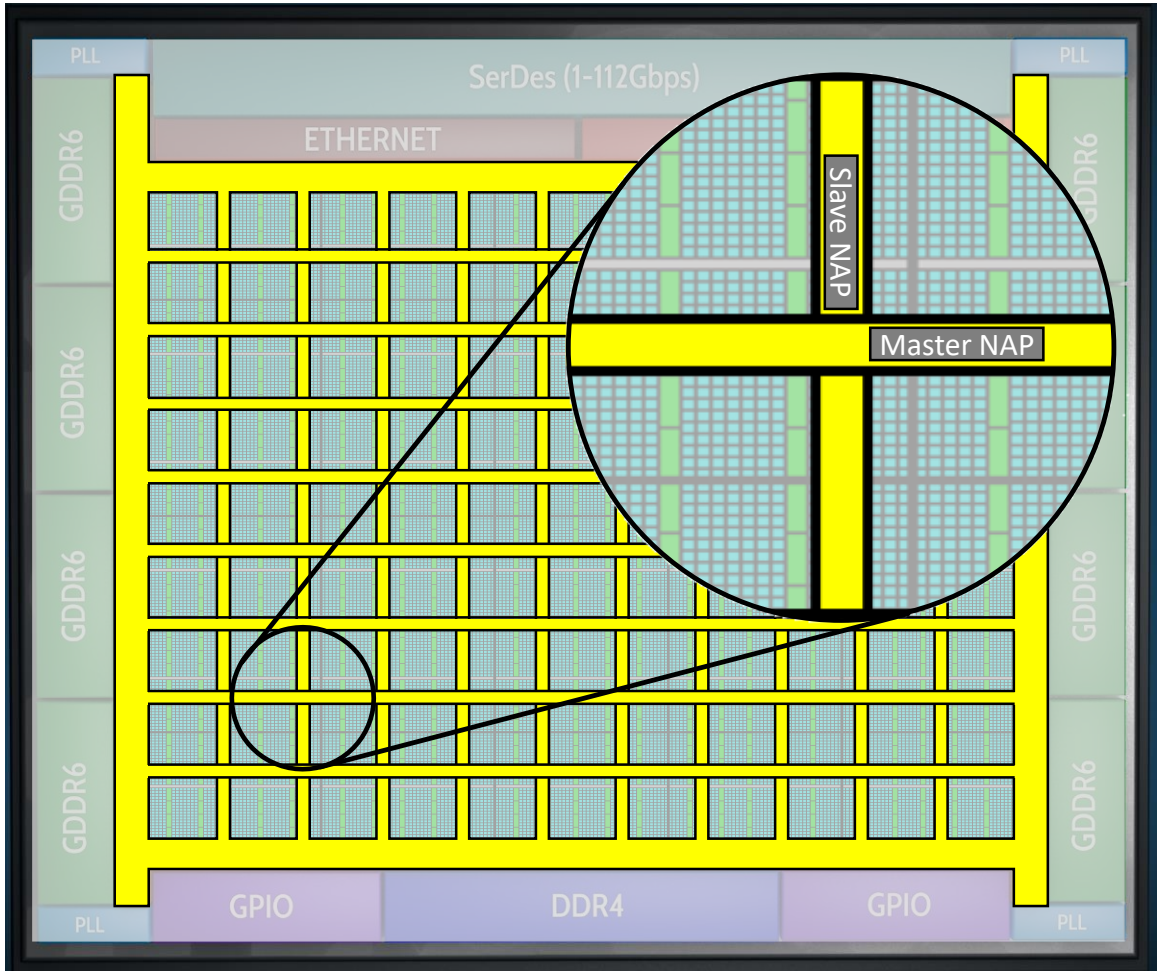
Achronix
Data Acceleration

Speedster7t: High Bandwidth Network Hierarchy

- High-speed specialized mesh that runs through the core of the FPGA

- Performance
  - Each row/column: 256b @ 2 GHz = 512 Gbps in both directions
  - N/S Bandwidth: 10 Tbps bidirectional
  - E/W Bandwidth: 8 Tbps bidirectional
  - External Bandwidth: >2Tbps bidirectional

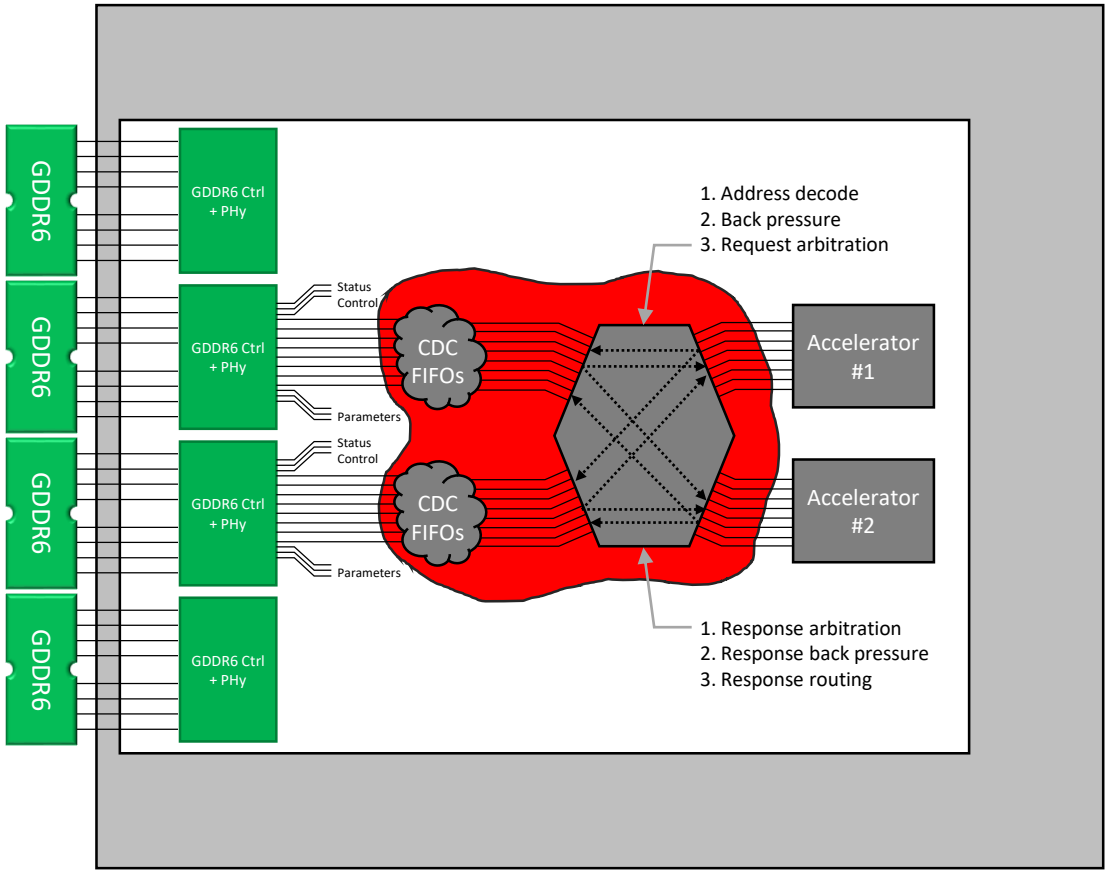
- Connection modes
  - Transactions (AXI)
  - Packets (for Ethernet)
  - Raw data streams within the FPGA core



Achronix
Data Acceleration

Speedster7t NOC: A New Design Paradigm

Traditional FPGA Design

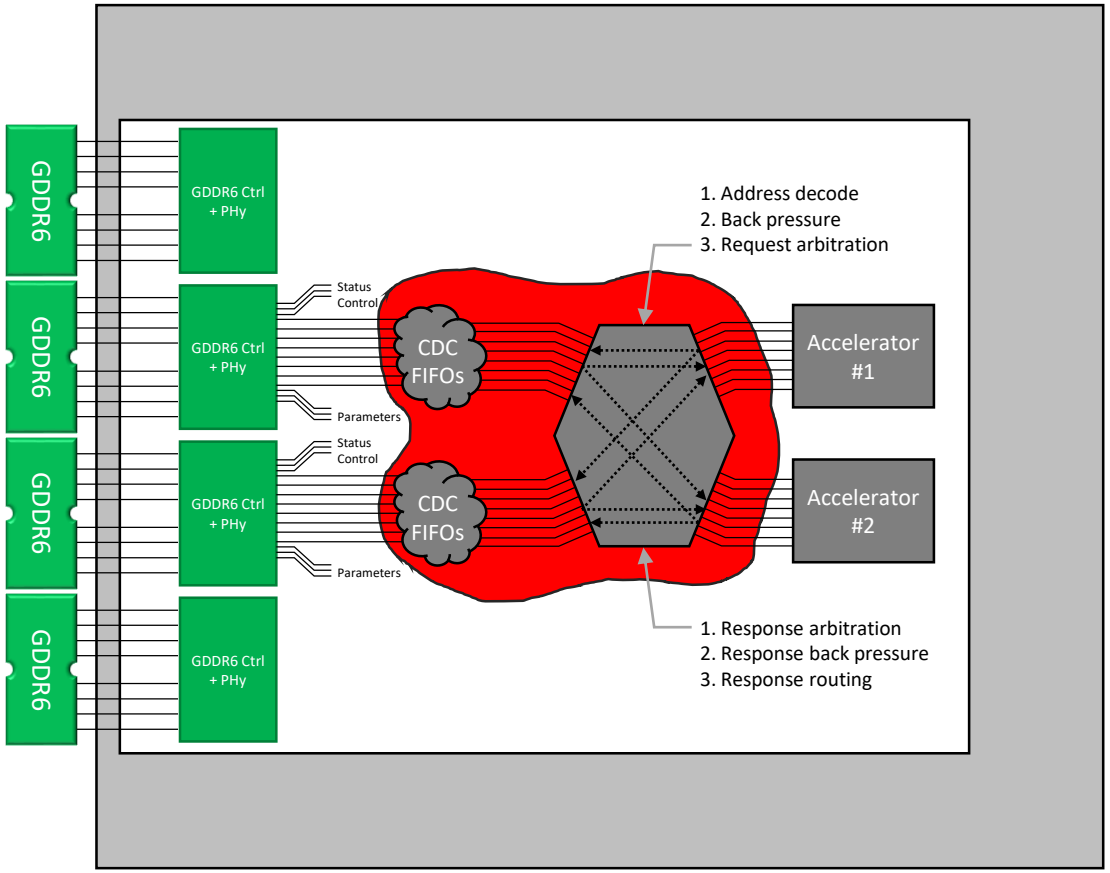


Quadratic Area Growth for Each Additional Accelerator and Interface

Achronix
Data Acceleration

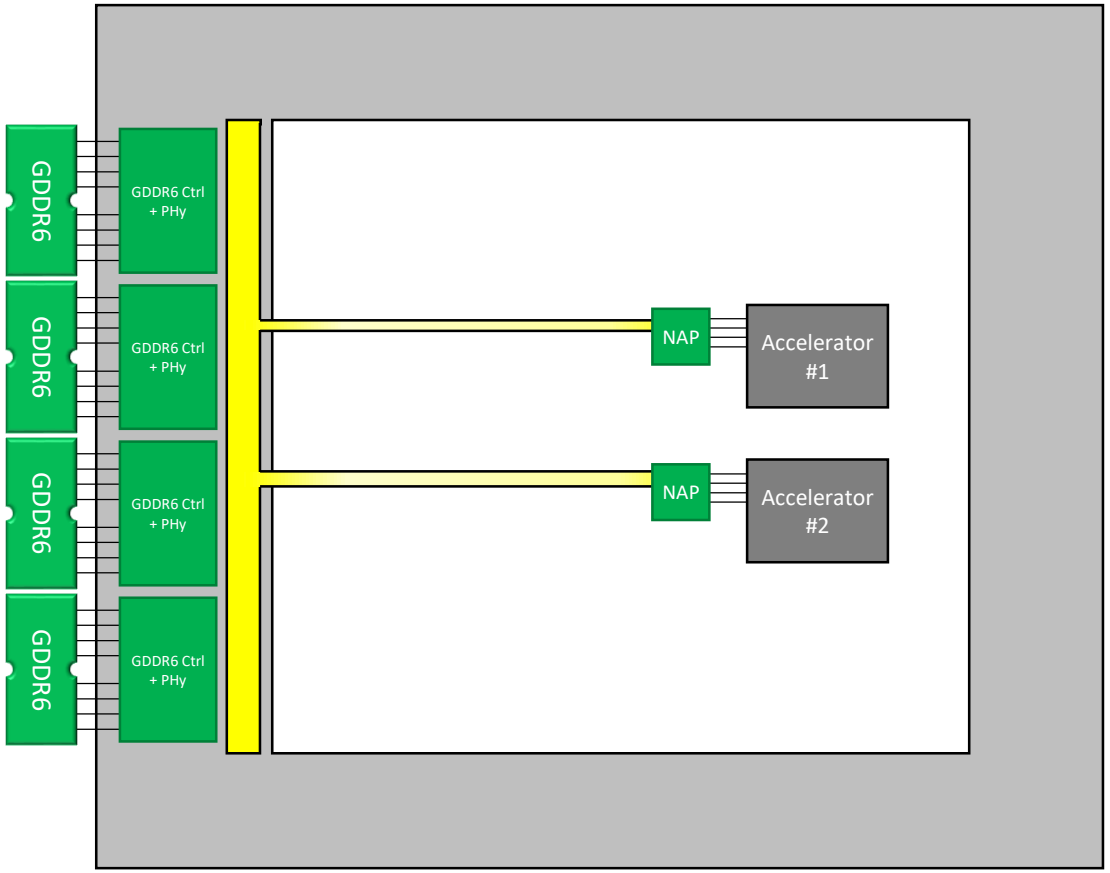
Speedster7t NOC: A New Design Paradigm

Traditional FPGA Design



Quadratic Area Growth for Each Additional Accelerator and Interface

Speedster7t FPGA+ Design

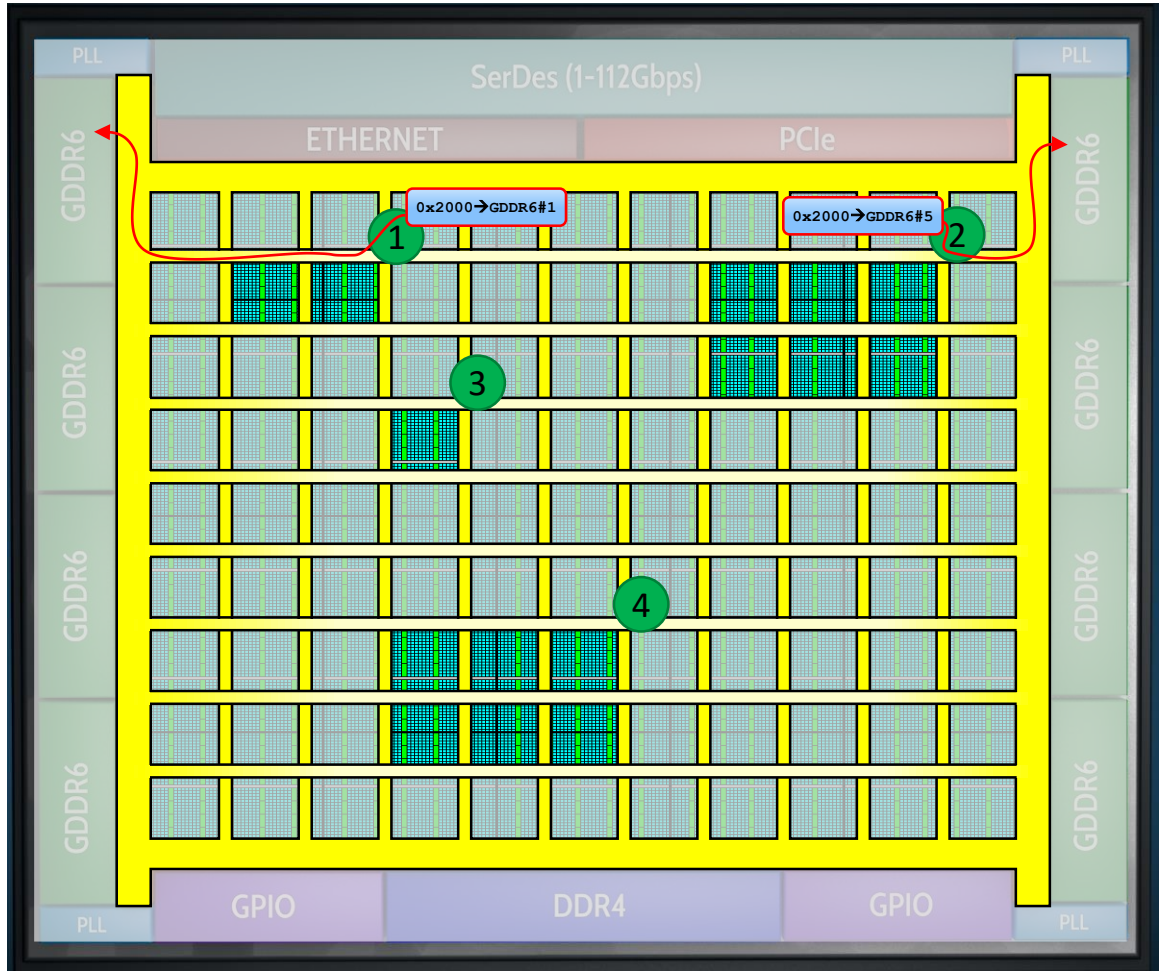


Linear Area Growth for Each Additional Accelerator and Interface

Achronix
Data Acceleration

Virtualization & Security

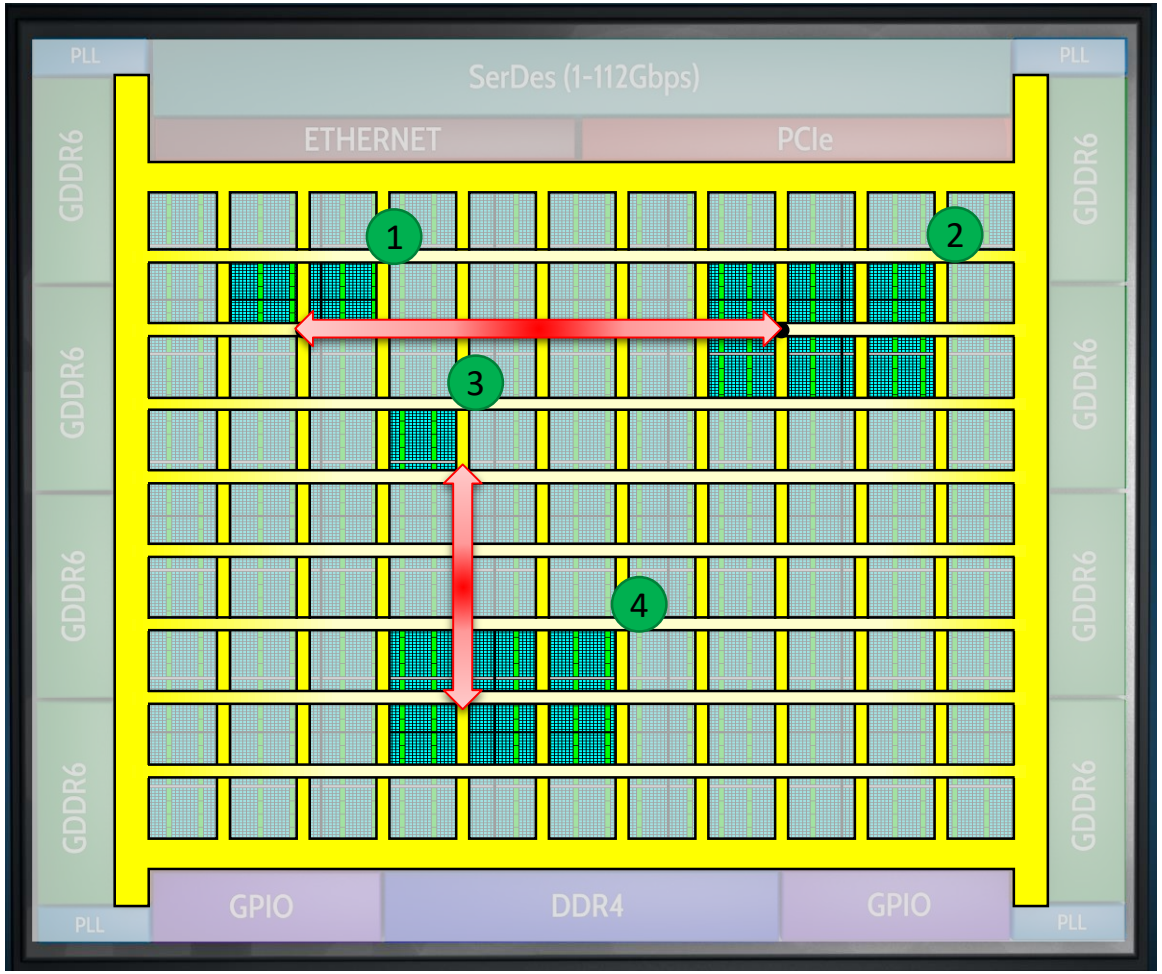
- Every access point requires an **Address Translation Table** (ATT)
  - The memory space seen by each accelerator can be independently controlled.
  - The Host PC (Over PCIe) or bitstream can remap the address space without requiring an RTL change.
  - For example:
    - Accelerator 1 RTL sees it's private memory space at address 0x2000.
    - Accelerator 2 RTL sees it's private memory space at address 0x2000.
    - But they don't share their memory spaces!
    - The per-node ATT allows the host PC control the re-direction of transactions.
- ATT also provides **Security**
  - Host PC or bitstream can configure the ATT to prevent certain individual accelerators from seeing portions of the address space.
  - For Example:
    - Accelerator 2 can be prevented from accessing GDDR6 #1 entirely.



Achronix
Data Acceleration

Speedster7t Internal Connectivity: Flit Transfers

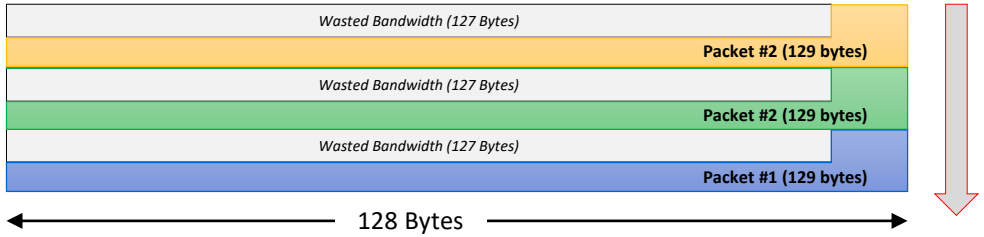
- NOC Connectivity
  - Support data transfer between nodes inside the FPGA fabric.
  - Just push in ~288 bits of data with the destination row or column, and it pops out the other side a few clock cycles later.
  - Superset of AXI Streaming.
- Example use case:
  - Connect accelerators
  - Flexible pipelining
  - Get across the device quickly
- Benefit:
  - Ease-of-use: simple interface
  - No congestion between accelerators
  - Eliminates floorplanning challenges

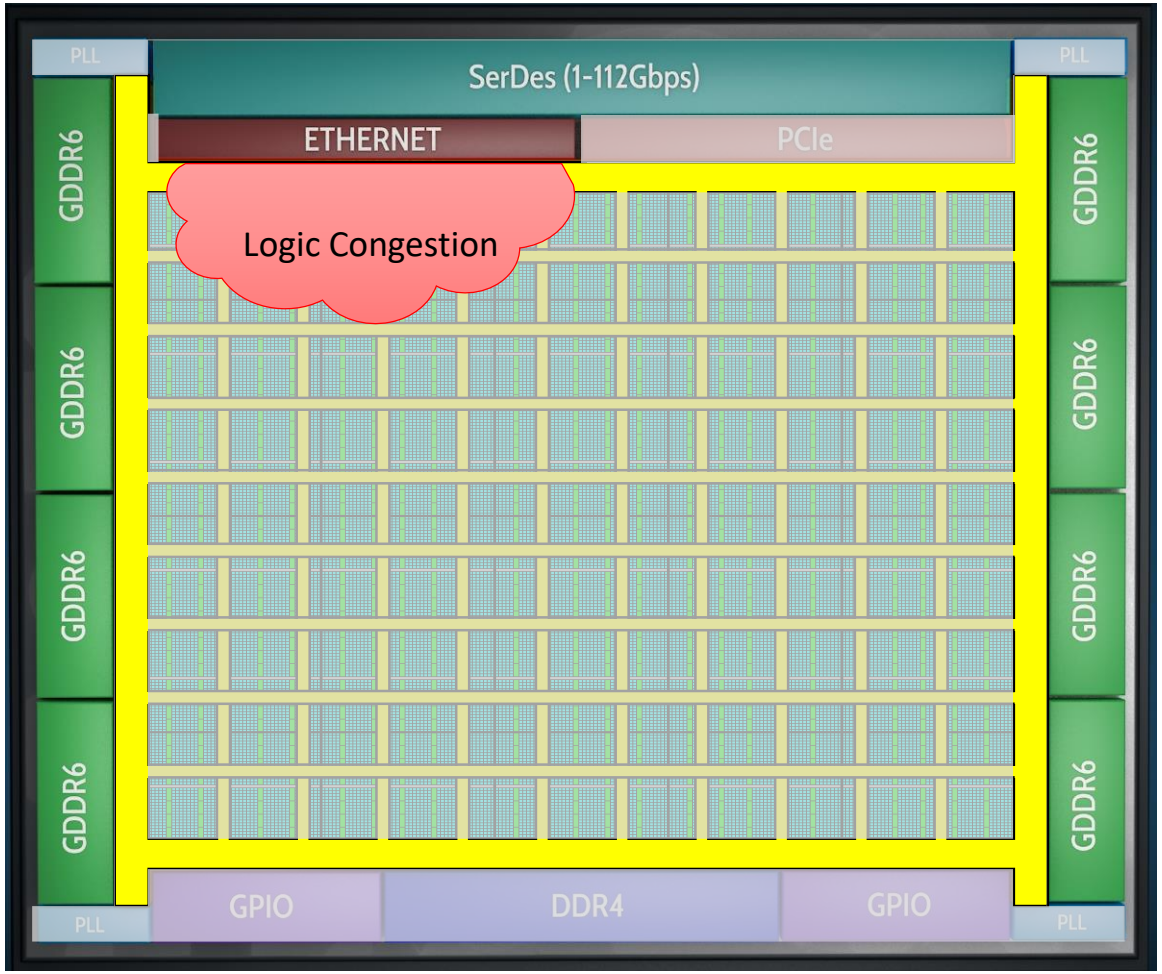


Achronix
Data Acceleration

400 Gbps Ethernet

- Traditional Approach:
  - **1024b** bus at **724 MHz**

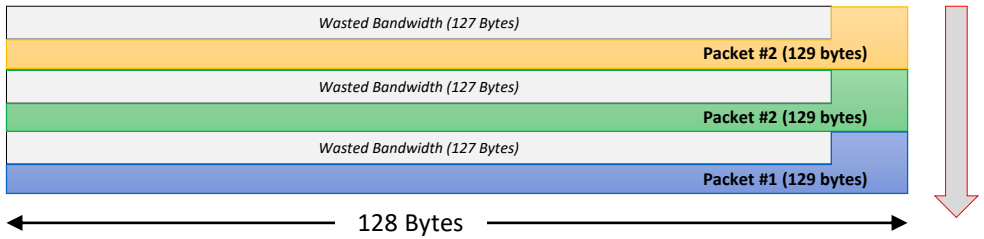




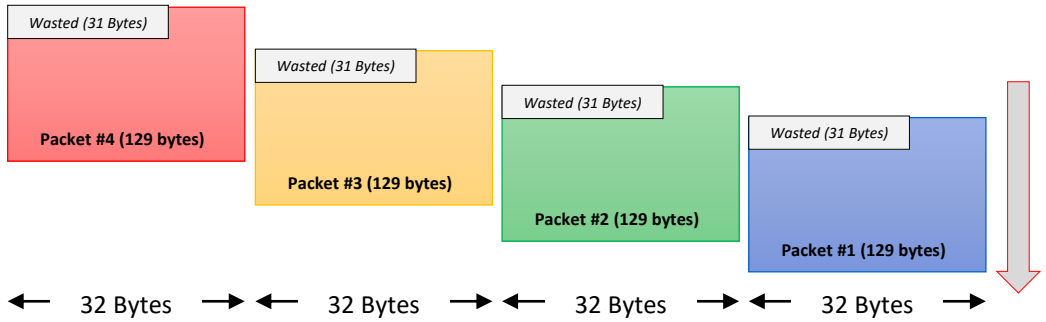
Achronix
Data Acceleration

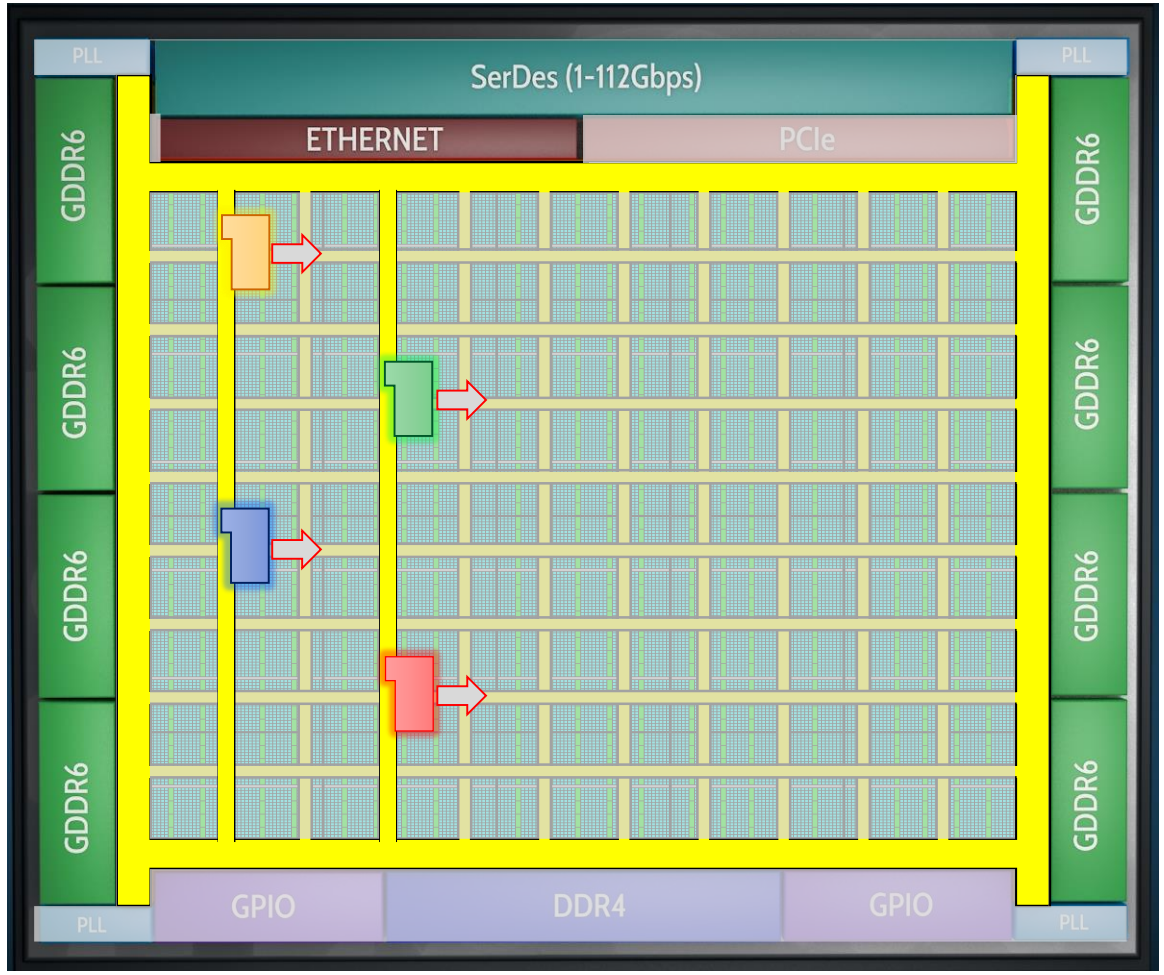
400 Gbps Ethernet

- Traditional Approach:
  - **1024b** bus at **724 MHz**



- Speedster Approach
  - Rearrange the data into multiple narrower packets to make it more efficient and much easier to use.
  - 4x **256b** Busses at **506 MHz**
  - Distribute the data to your choice of location within the FPGA fabric.
  - Alternatively, 1024 bit Quad-Segmented Interface is also supported.





Achronix
Data Acceleration

Results

- **Ease of use**
  - For an FPGA application with 80 Machine Learning acceleration units spread through the core, the on-chip networking hierarchy takes care of memory accesses, including address decoding, arbitration, and routing to 8 memory interfaces.
    - The network consumes no FPGA logic to do so.
    - The interface to each acceleration unit is **256 bits @ 750 MHz**.

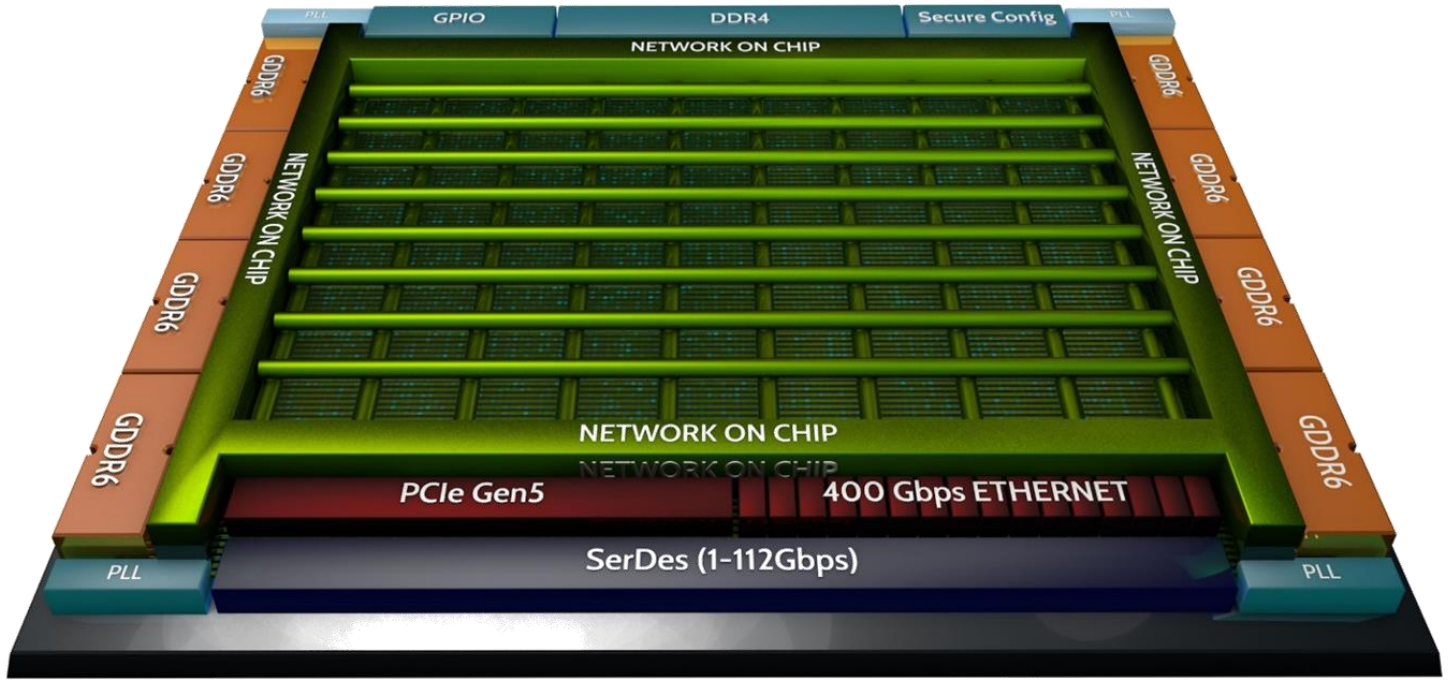
- **Area Reduction**
  - With a pure-FPGA design with **64** accelerator units with 128b connected to 8 memory interfaces
    - The interface to each acceleration unit is **128 bits @ less than 400 MHz**.
    - The network consumes **580k** Look-Up Tables (LUTs). This is a significant portion of the device which would be otherwise available for algorithm acceleration.
    - By comparison, the Speedster NoC supports **80** (vs 64) interfaces, where each is **256b** (vs. 128) at **750MHz** (vs. 400), using **0 LUTs**.

- **Massive throughput**
  - Each of 4 independent columns delivers a 400Gbps Ethernet stream as a 1024-bit wide interface operating at 750 MHz into the FPGA core.
    - Allows an implementation of a **1.6 Tbps** networking switch with the programmability of an FPGA
  - Each of 16 GDDR6 memory channels provides 256 Gbps of throughput, for a total of **4 Tbps**. The Achronix NoC Hierarchy conveniently distributes this throughout the FPGA fabric for ease of use and to reduce congestion.

Summary

- New techniques are required to move data to and from memory and high speed interfaces to on-chip processing units.

- A multi-level on-chip network hierarchy can deliver terabits of data to user logic at arbitrary locations in the processing core.
  - Automated address decoding, routing, and arbitration within the device.
  - Specialized transfer modes allow the delivery of multiple 400Gbps Ethernet packet streams.
  - Unprecedented ease of use and minimal time-to-market.



Achronix
Data Acceleration