Data Acceleration: Challenges & Opportunities

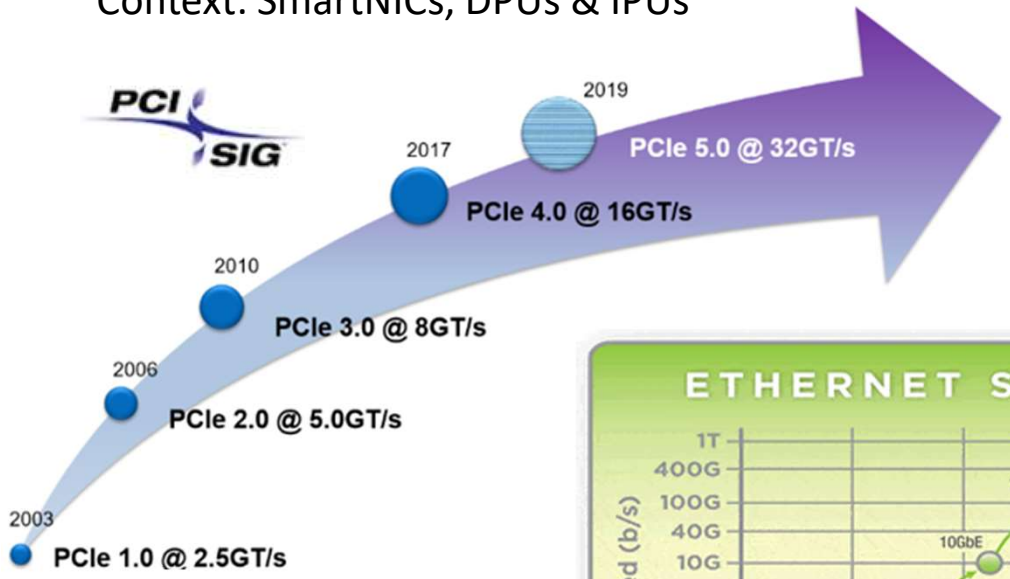
Scott Schweitzer
Achronix Semiconductors

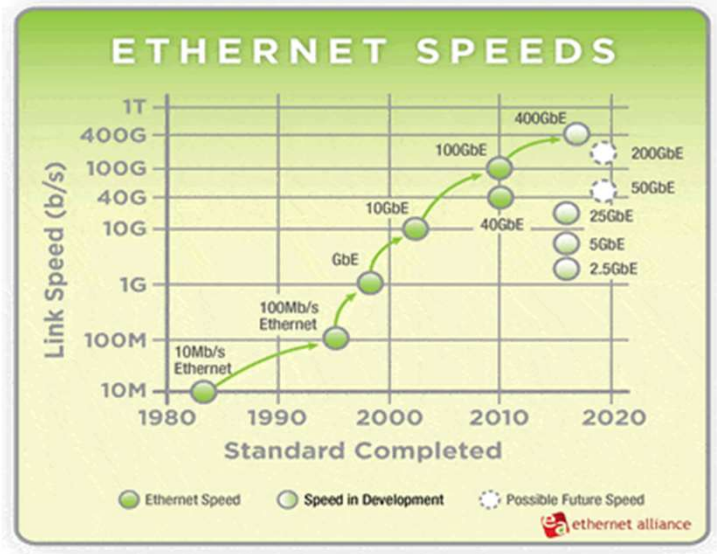
Agenda

- **Challenges**
  - Hardware
    - Bandwidth
    - Queueing
    - Available On Chip B/W
  - Software
    - Programmability
    - Security
- **Opportunities**
  - Hardware
    - Computational Scaling
    - Memory I/O
    - Network on Chip Bandwidth
  - Software
    - Programmability
    - Security

Challenge is Bandwidth

Context: SmartNICs, DPUs & IPUs





NIC Interface	Year*	Speed (GB/sec)
10GbE x 2	2005	2.5
PCIe Gen 1x8	2005	2.0
PCIe Gen 2x8	2006	4.0
40GbE x 2	2010	10
PCIe Gen 3x8	2010	8
25GbE x 2	2015	6.25
PCIe Gen 3x16	2016	16
100GbE x 2	2017	25
PCIe Gen 4x16	2020	32
400GbE x 2	2022	100
PCIe Gen 5x16	2022	64

OPEN Compute Project®

Opportunity – Computational Scaling

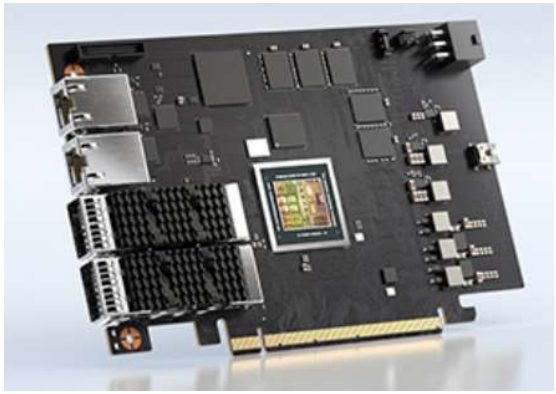
GPU

NVIDIA GEFORCE RTX 3080
8,704 CUDA Cores @ 1.44-1.71GHz
272 Tensor Cores
10GB GDDR6X RAM
320 bit Internal Bus Width
PCIe Gen 4x16

FPGA

Bittware VectorPath
692K 6-input lookup tables (LUTs)
2,560 Machine Learning Processors
16B GDDR6 RAM
256 bit Internal Bus Width
PCIe Gen 4x16

DPU



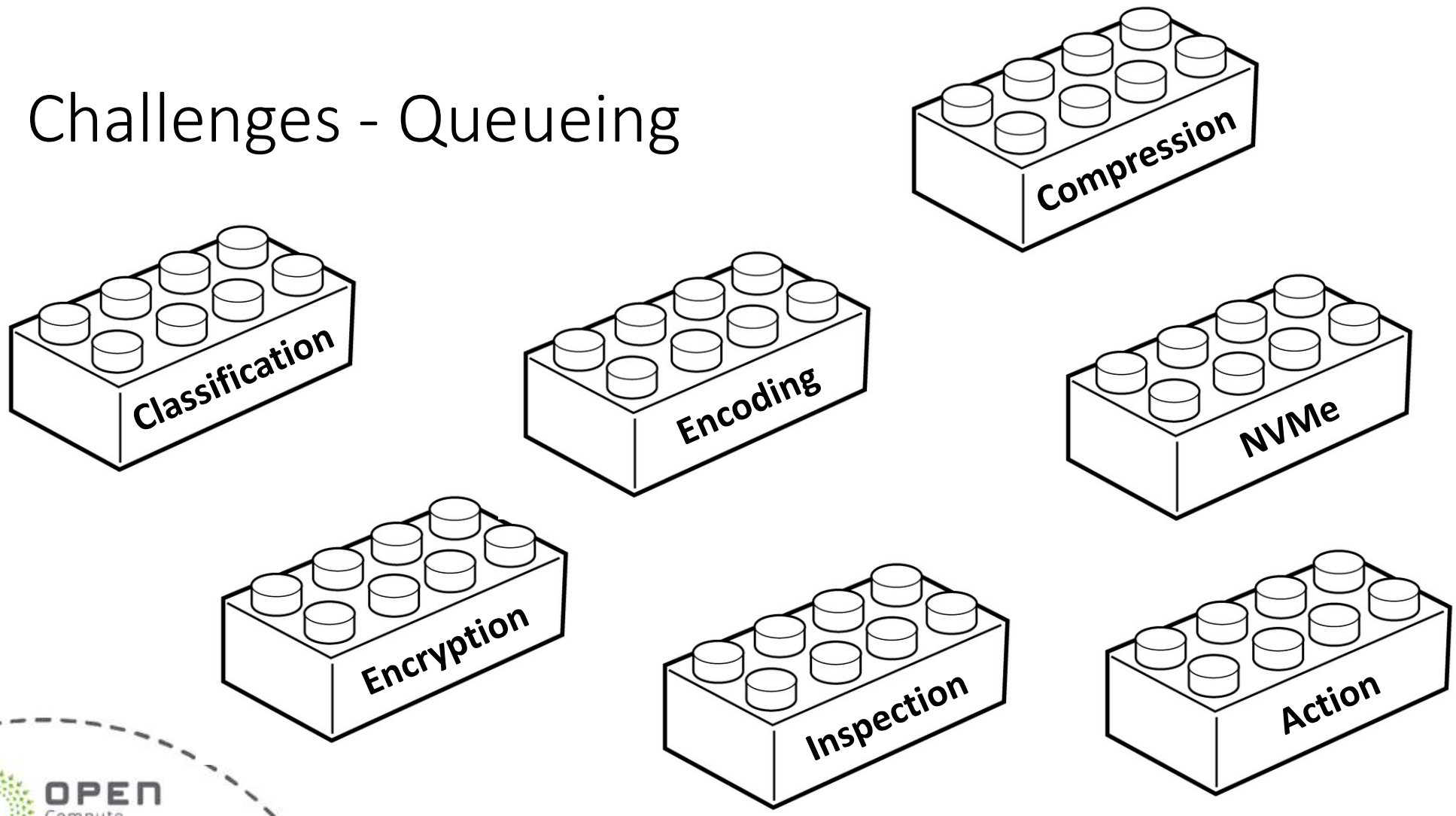
NVIDIA BlueField-2 DPU
16 Arm Cores, 256 Threads @ 3GHz
16GB DDR5 RAM
PCIe Gen 4x32

OPEN Compute Project

CXL Compute Express Link

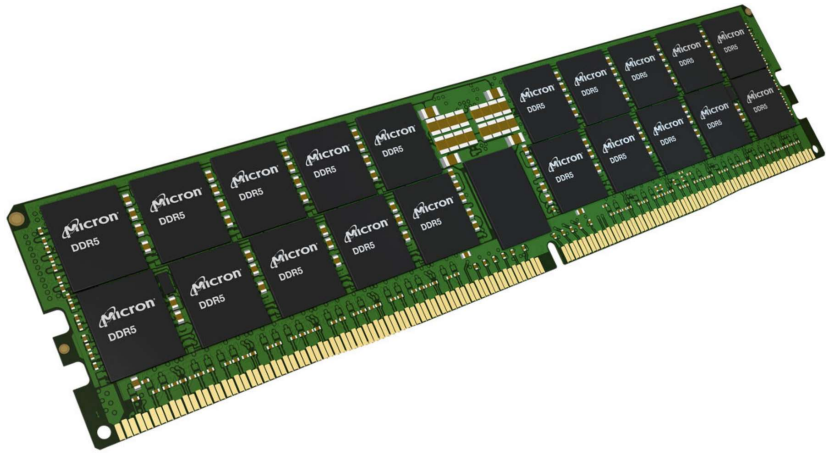
ODSA Workshop 2021

Challenges - Queueing

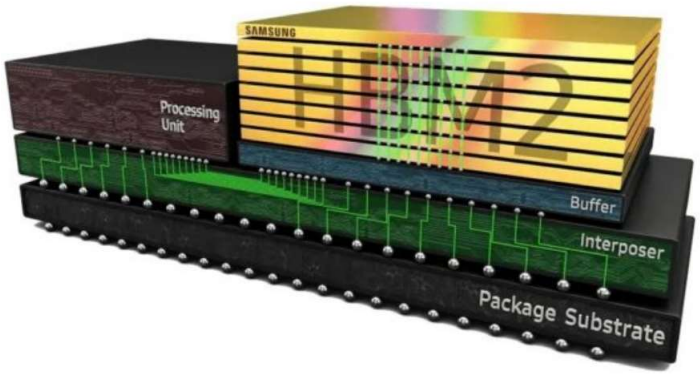


OPEN Compute Project®

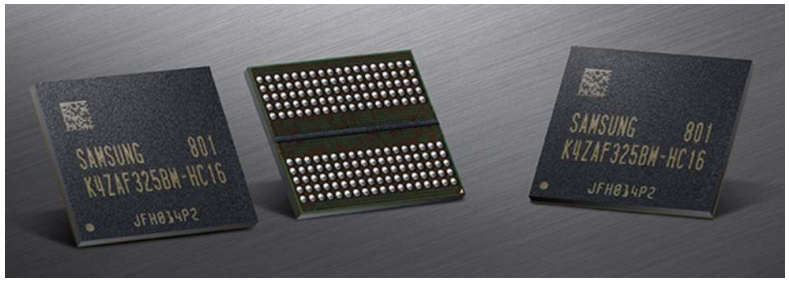
Opportunity – Memory I/O



DDR5



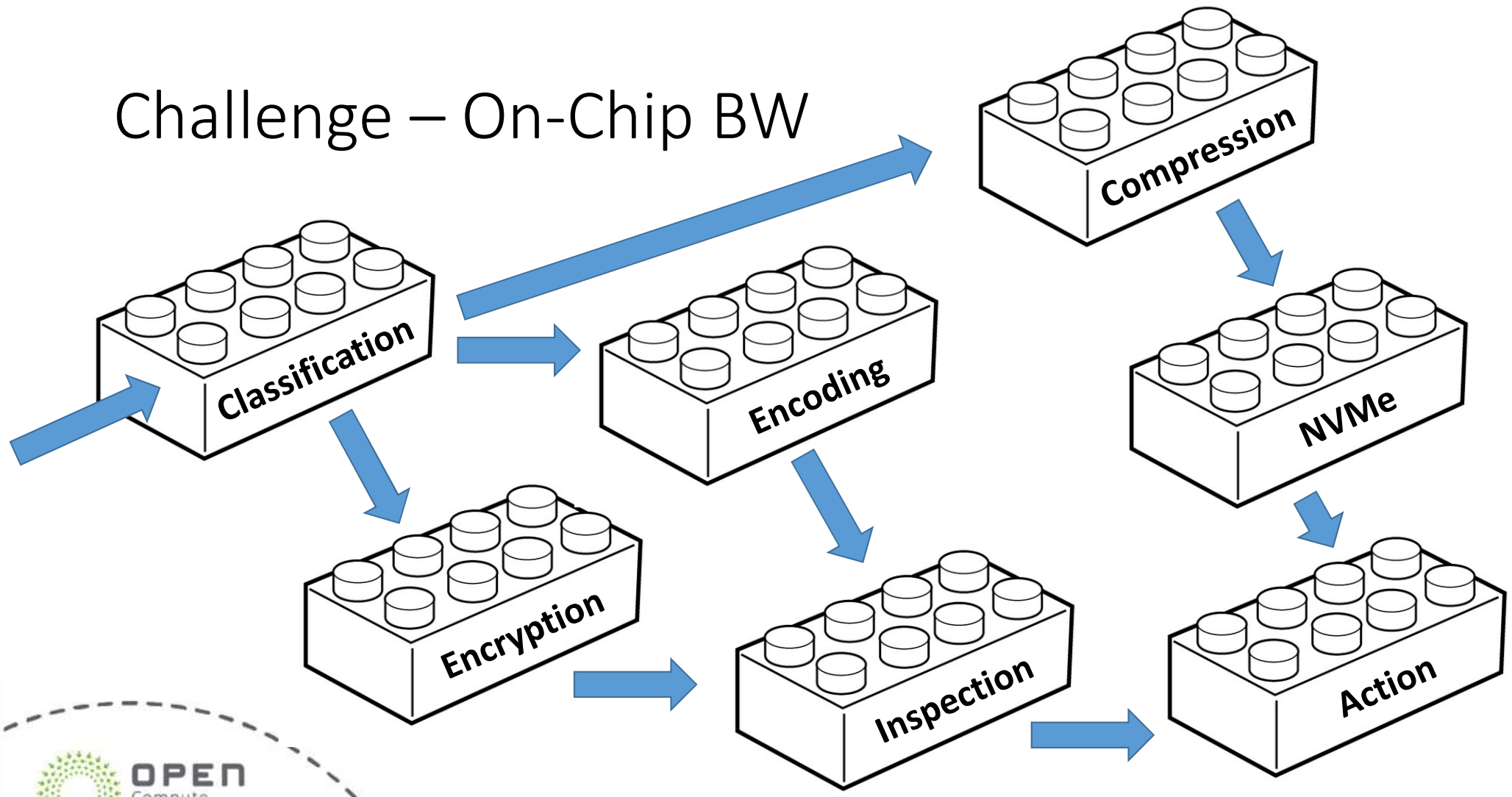
HBM2



GDDR6

OPEN Compute Project®

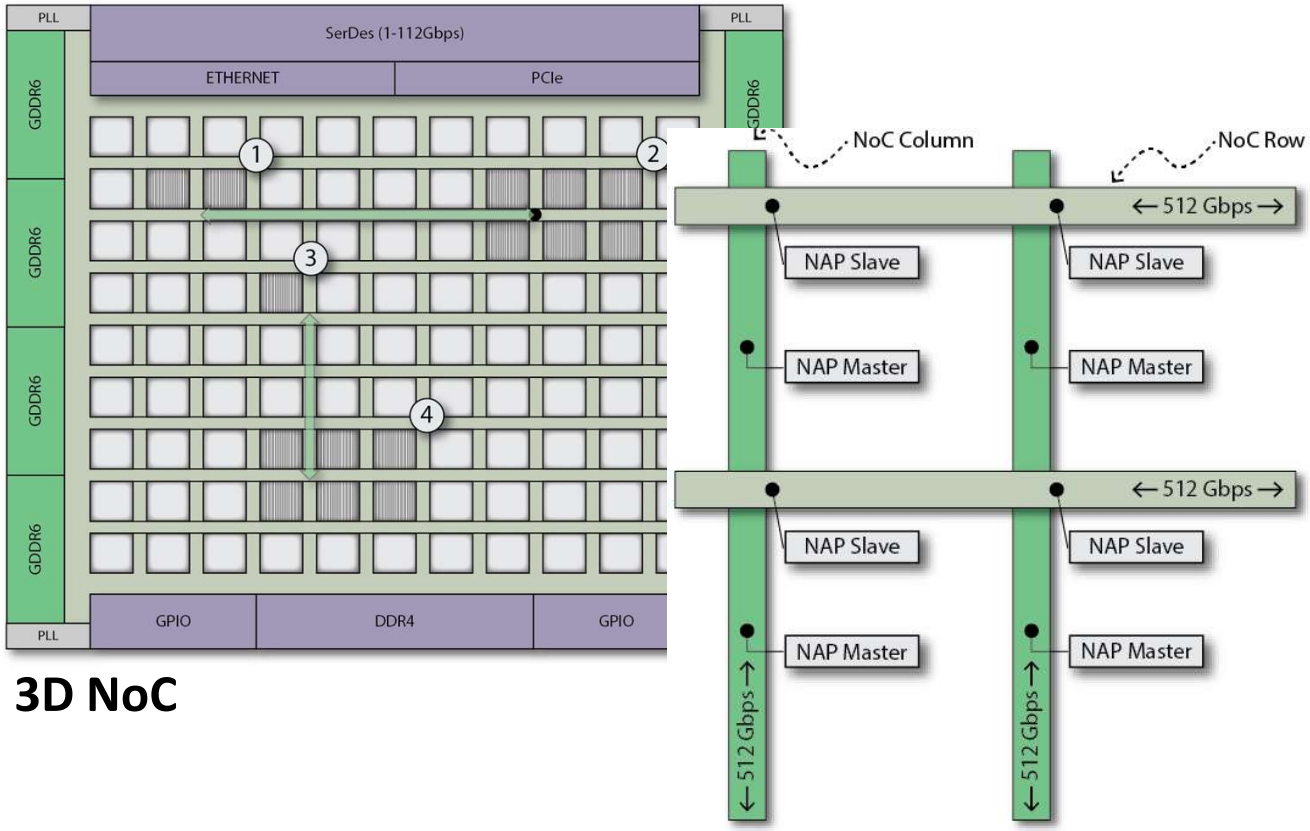
Challenge – On-Chip BW



OPEN Compute Project®

Opportunity – High Speed Network on Chip

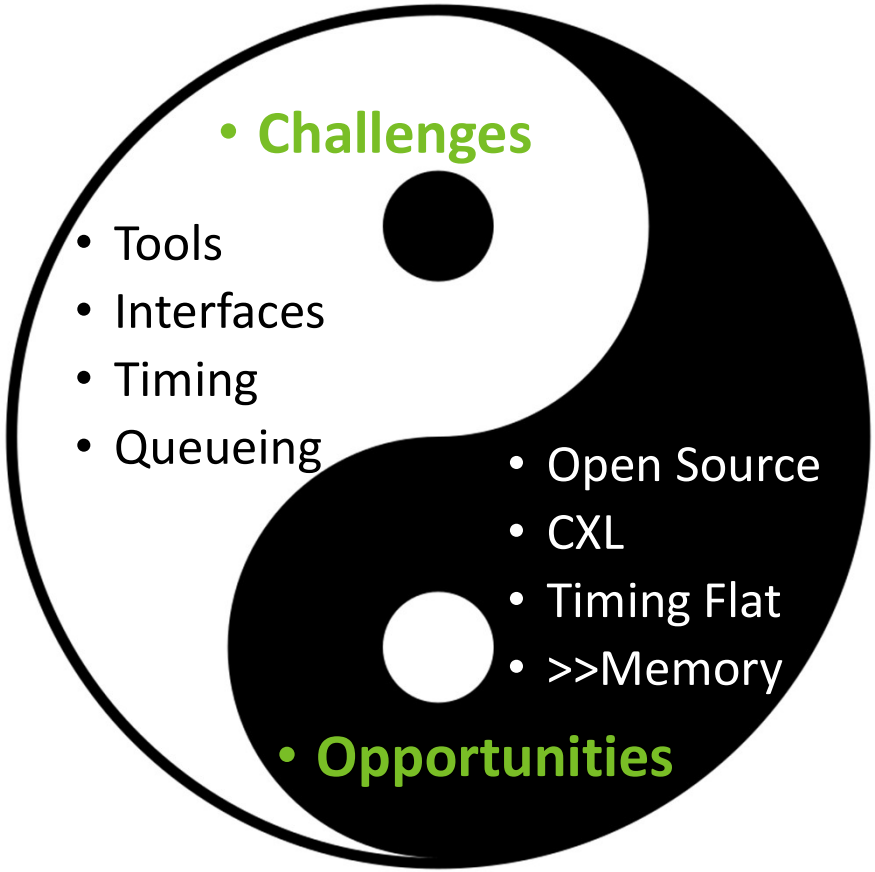
2D Block



3D NoC

OPEN Compute Project

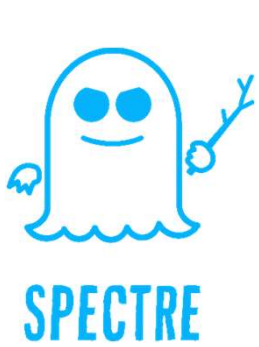
Programmability



OPEN Compute Project®

Security – Challenges & Opportunities

- Where and how to store secrets
  - Trusted Platform Module (TPM)
  - Onboard Secure Root of Trust
  - Physically Uncloneable Functions (PUF)
- Which accelerated libraries to include and at what speeds
- Side Channel Attacks (SCA)
  - SCA Counter Measures



The Exciting Future of Chiplets & Accelerators

- Standardization
  - Packaging issues: dimensions, thermal, connection and pin placement
  - Interface issues: signaling, common control plane or security context
  - Protocols between chiplets and to the host
  - Documentation
  - Tools
- Ecosystem, market place ➔ tools and flows, simulation, analysis
  - Imagine issues stacking several complex accelerators in the same package
  - Thermal, magnetic, signal interference