Integrated Device Technology

RISC-V I/O Scale-out Architecture for Distributed Data Analytics

Mohammad Akhter

Integrated Device Technology, Inc.

mohammad.akhter@idt.com

July 12, 2016

Contributors/Acknowledgement

- Henry Cook, SiFive, Inc.
- Howard Mao, SiFive, Inc.
- Krste Asanovic, SiFive, Inc.
- Mohammad Akhter, IDT, Inc.
- Stephen Durr, IDT, Inc.

Massive Data Growth Driving Internet

91%

Social media is driving Mobile Internet Access

More Video uploaded to YouTube than the traditional TV networks

3.6B Photos uploaded in 2014

(whatsapp, facebook, instagram, snapchat …)

807M people watched YouTube video

"Charlie bit my finger again“

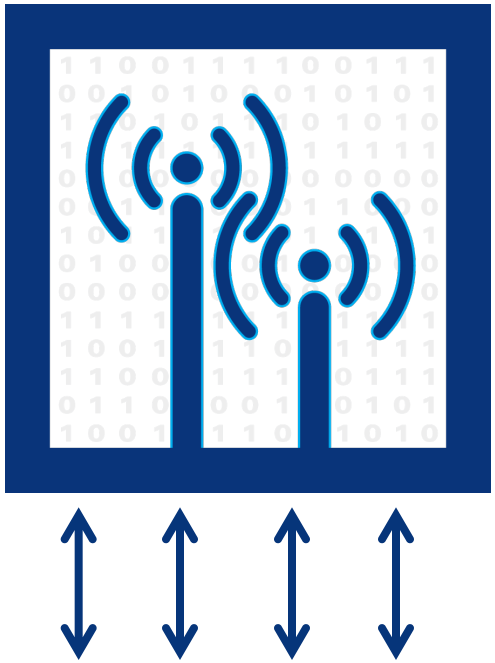
(642M to Disneyland since 1955)

Unstructured Data ….

Machine Data not included ….

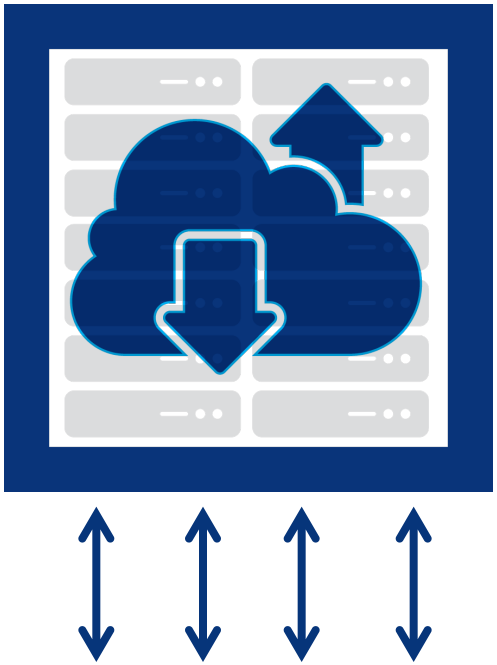
Live Data from Wireless to the Cloud

The Real-Time Usage Rising



Networks and Data Centers must THINK FAST

- Millions of users
- More data per user
- Low Latency
- Data Analytics

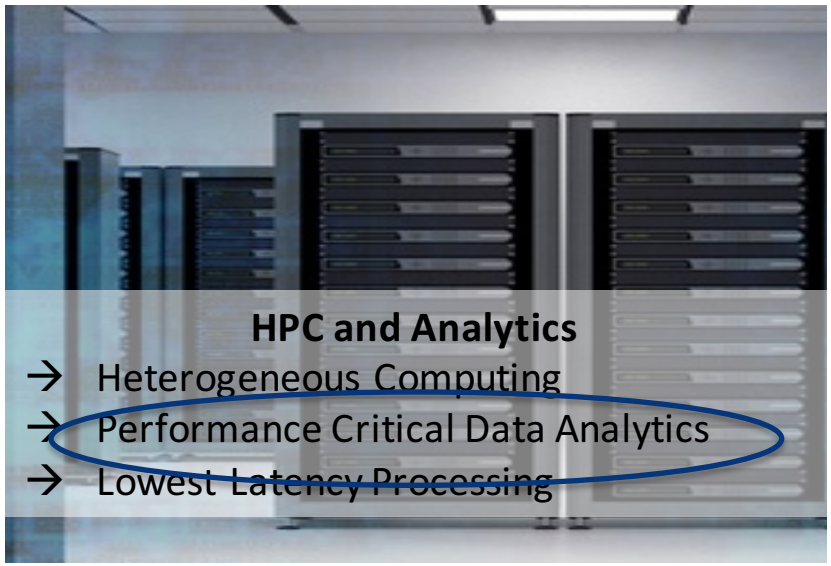


Social Media, Finance, Health, Video, Audio, Cloud Compute, Transportation, Military C3I

Focus on the Use Cases …

Robotics/Sensor Fusion/Autonmous

- → Real-time Video/Image Recognition
- → Time synchronized Processing
- → Deterministic Latency
- → Fault-Tolerant System

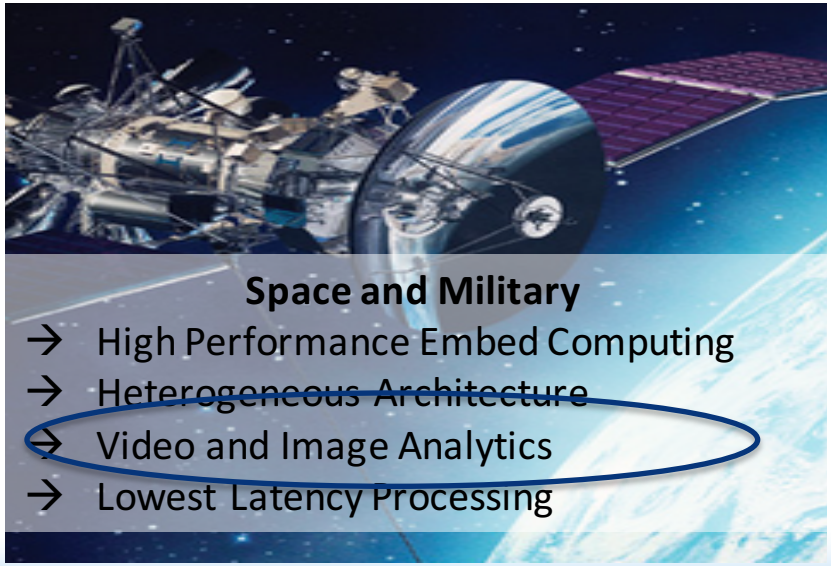


HPC and Analytics

- → Heterogeneous Computing
- → Performance Critical Data Analytics
- → Lowest Latency Processing

Wireless Access Network

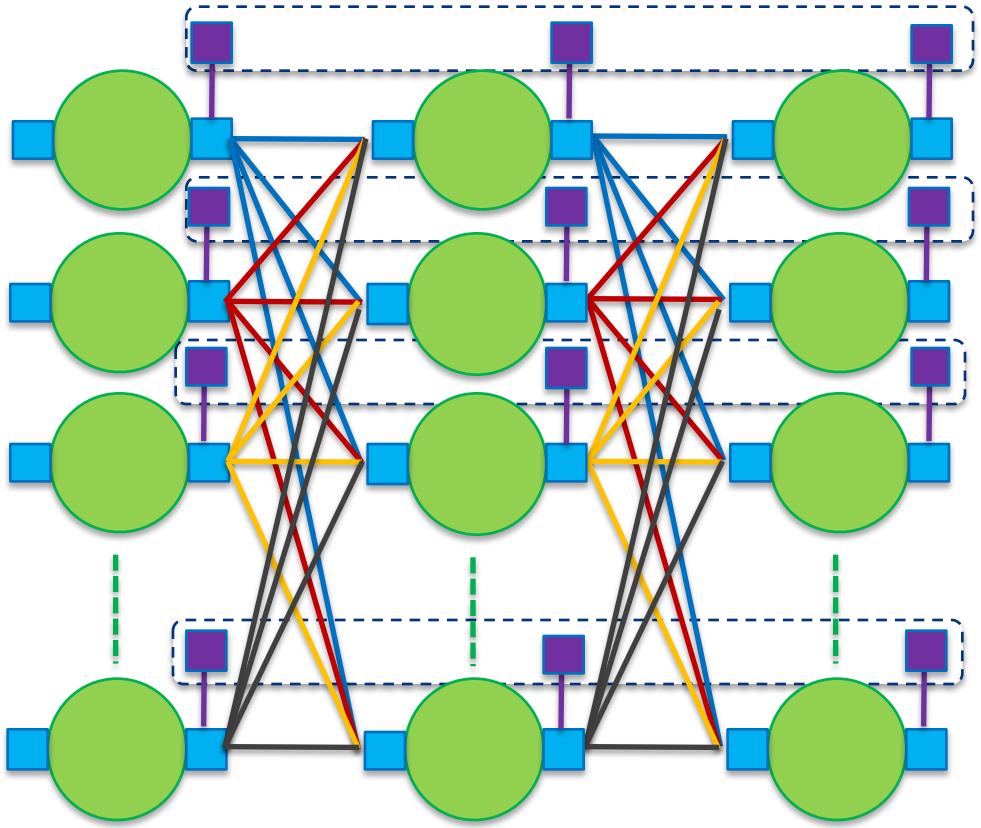
- → Edge high performance computing
- → Deterministic Lowest Latency
- → Fault-Tolerant System
- → Network and Signal Processing



Space and Military

- → High Performance Embed Computing
- → Heterogeneous Architecture
- → Video and Image Analytics
- → Lowest Latency Processing

Underlying Structure for Analytics…

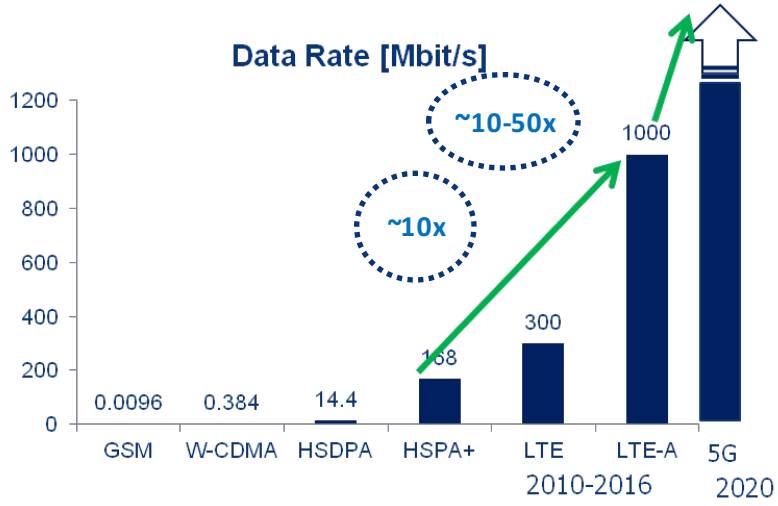


Demands balanced low latency computing, I/O, memory, and storage processing

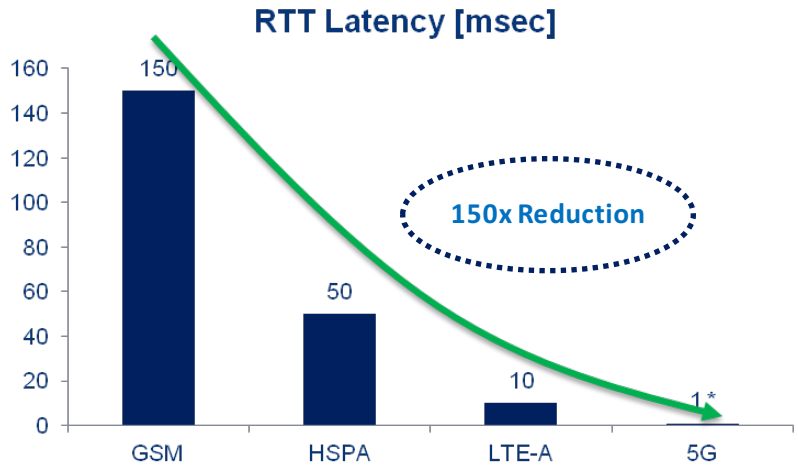
Wireless Network Evolution

Driven by Real-time Data with better QoS

Higher Bandwidth driven by Video/Social Media

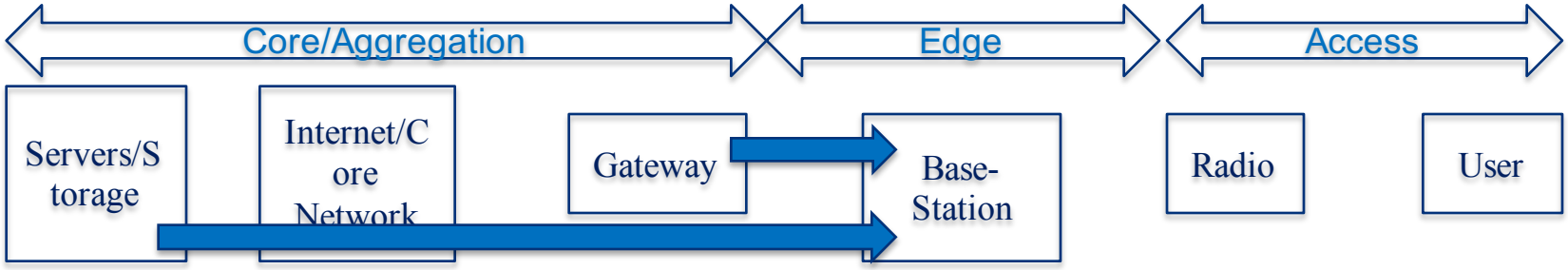


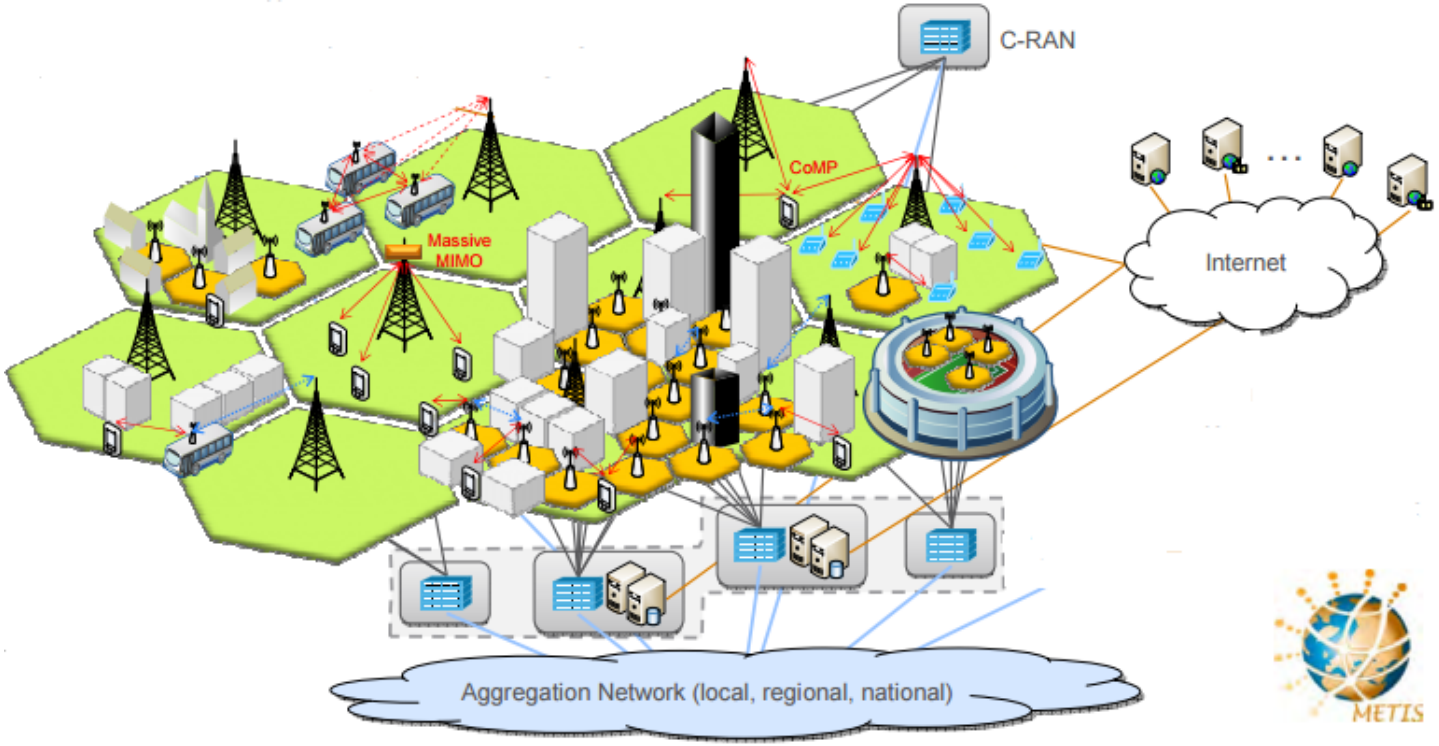
Lower Latency to achieve superior QoE



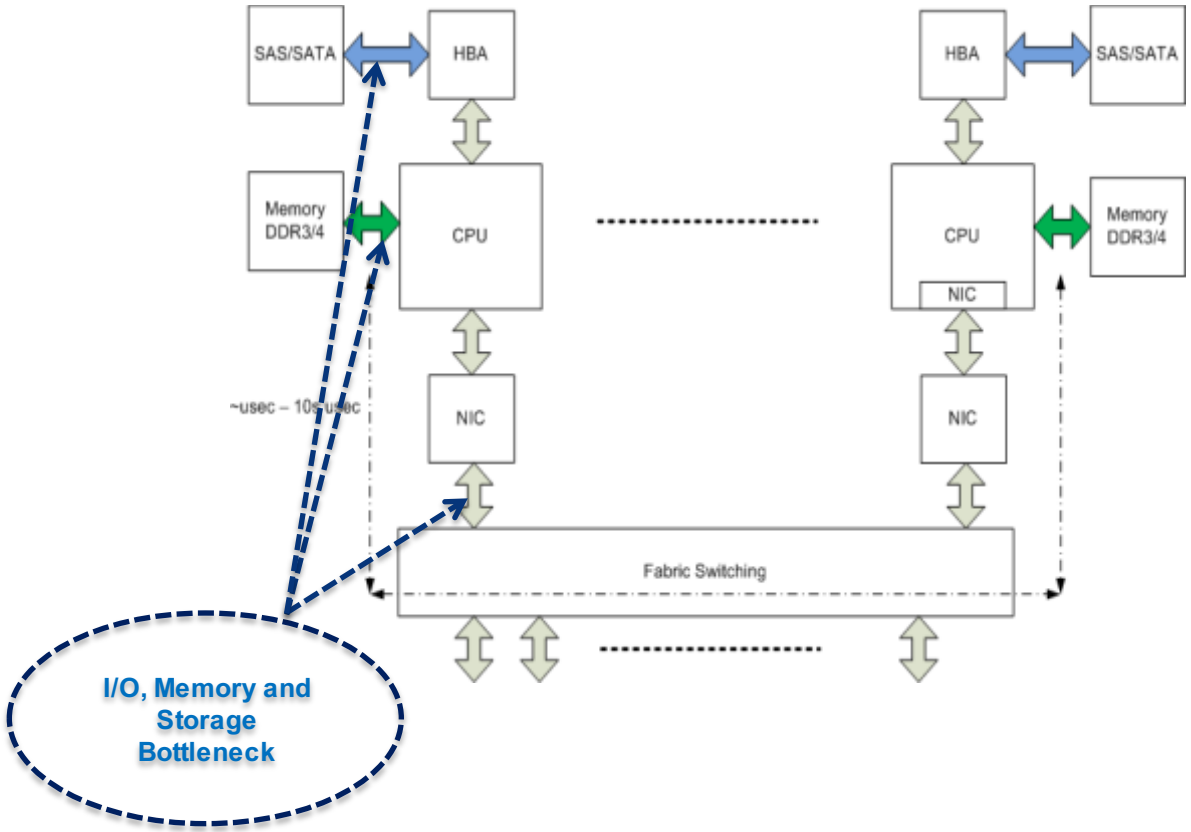
Edge-Core Network Convergence

Race to msec!



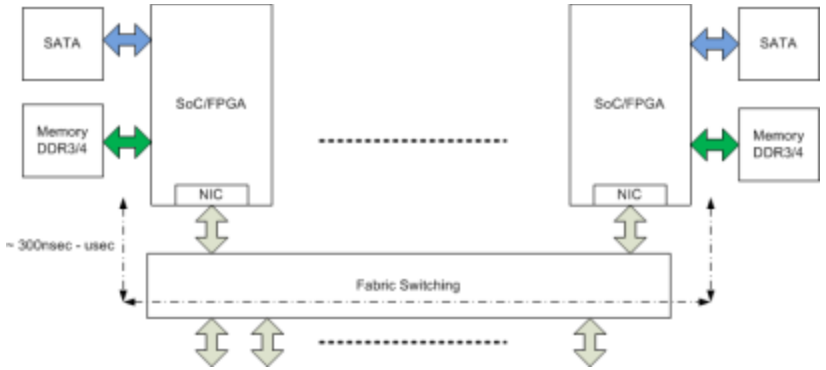


Traditional Computing Model

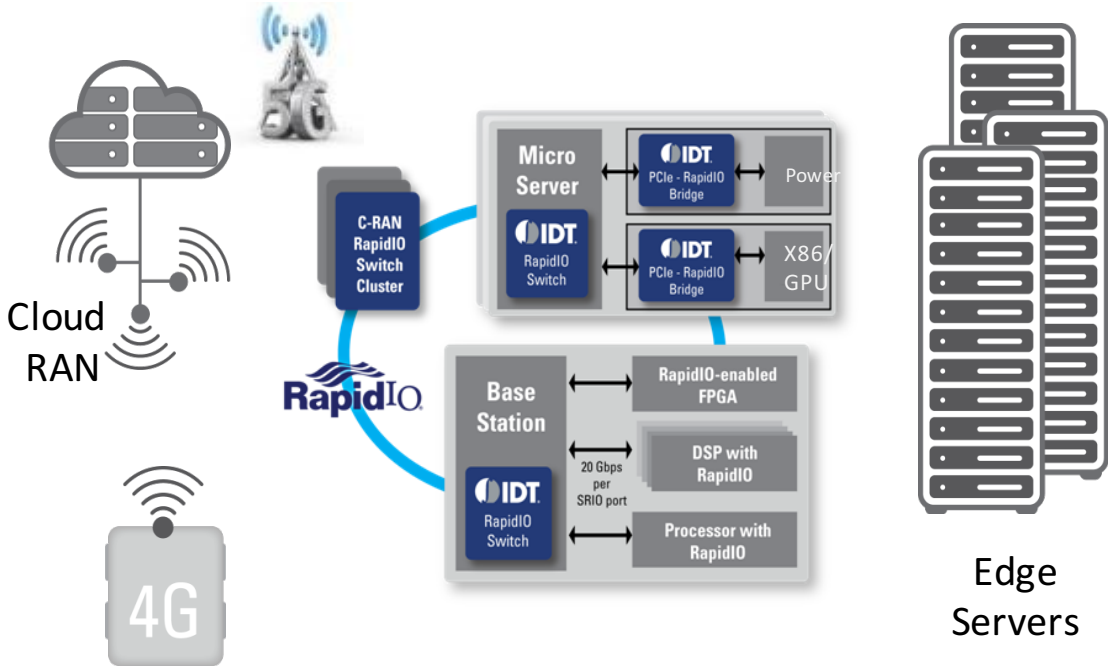


Suffers from inherent mismatch between I/O, Memory, and Storage and Processing (CPU/Accelearators) Performance Evolution

Distributed Edge Network Computing Model

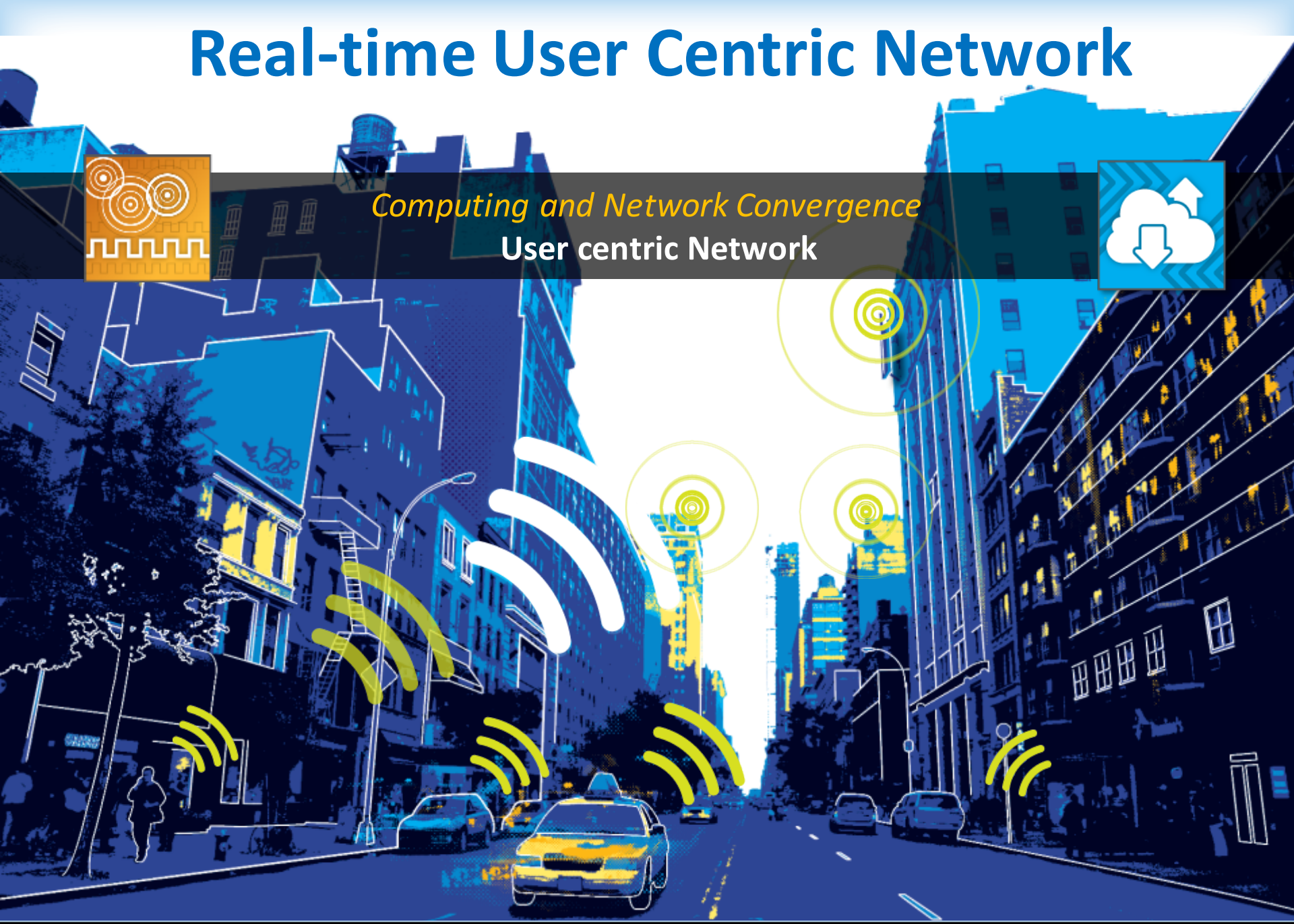


- → Reduces bandwidth demand from Access to Core
- → Provides Real-time predictive decision
- → User Centric optimization of Content delivery



4G BTS with integrated computing

Real-time User Centric Network



Deep Learning Video Example (Hacking a new Computing)

Hacking a new Computer

- Balancing GPU/CPU performance with Memory and I/O

Hacking a new Computer

- Scale-out Computing and storage

38U System
144 TFLops
608 GPU Nodes (4x4x36)
38 Port Switching Fabric

4 GPU Eval Cards
8G I/O NIC

4 GPU (NVIDIA Tegra K1)
56G I/O (4x14G NIC)
200G Switching

Hacking a new Computer

Deep Learning Micro-Cluster

Real-time Video/Social Analytics

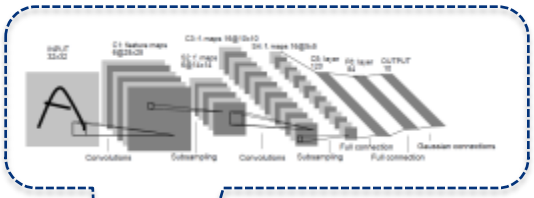
- > TFlops Processing/node
- ~29 frames/sec/node
- ~100 ns RapidIO switching latency
- ~5W-11 W per GPU node

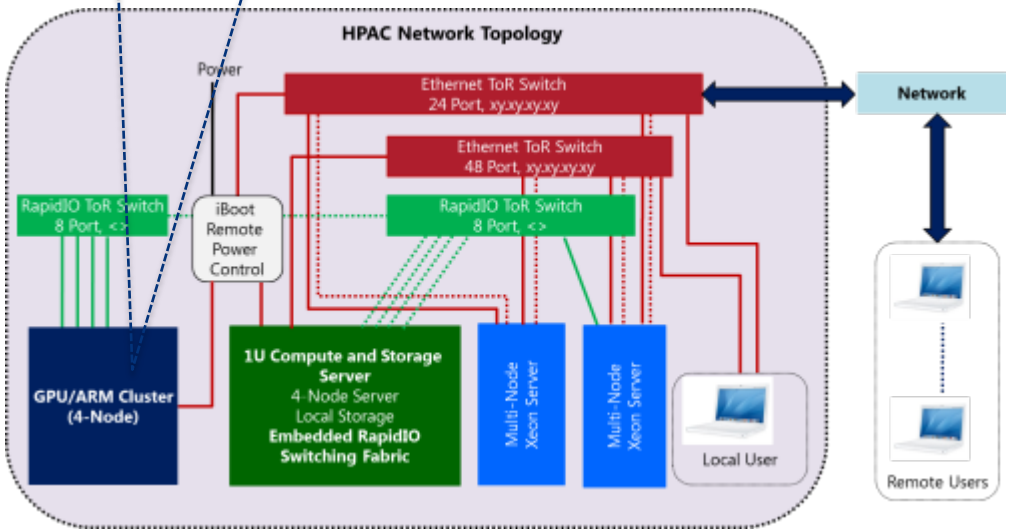
RapidIO Switch Appliance

Low Power ARM+GPU Cluster with RapidIO

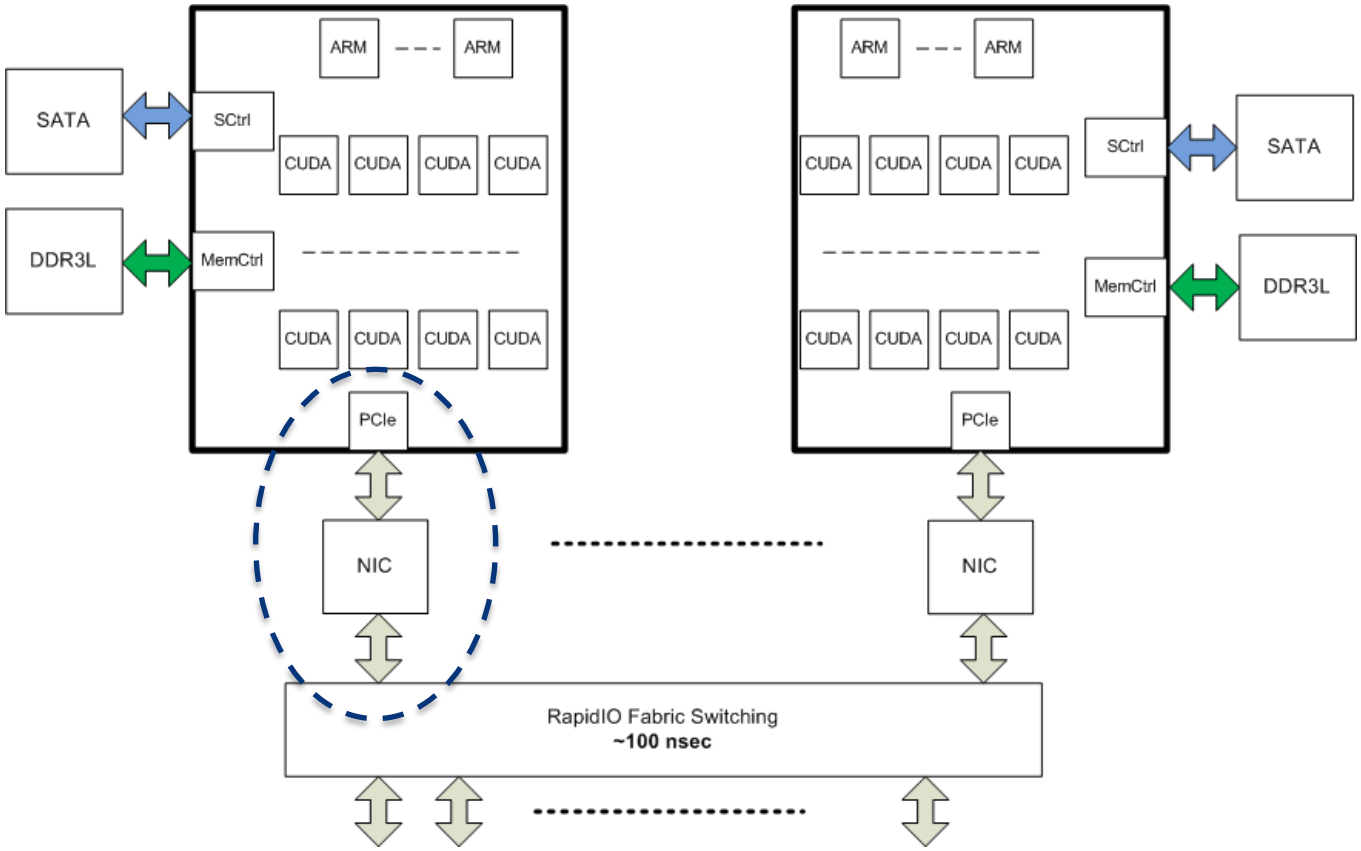
Example Analytics

Deep Learning Model for Image/Video



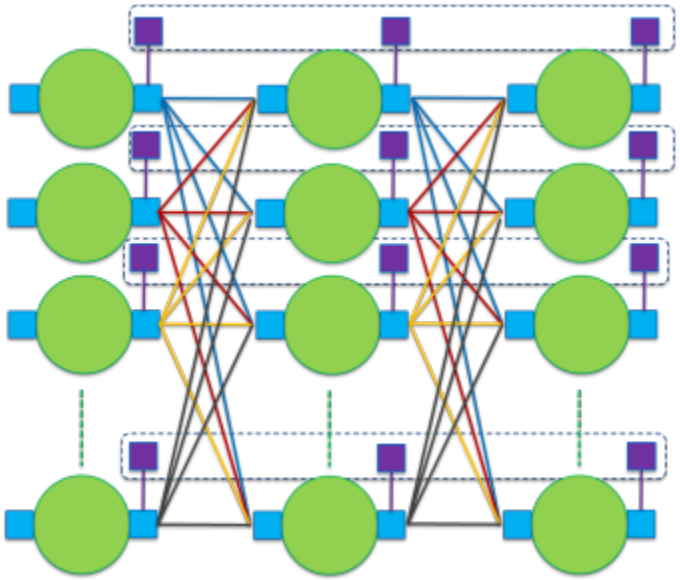


Analytics Computing Model Challenges



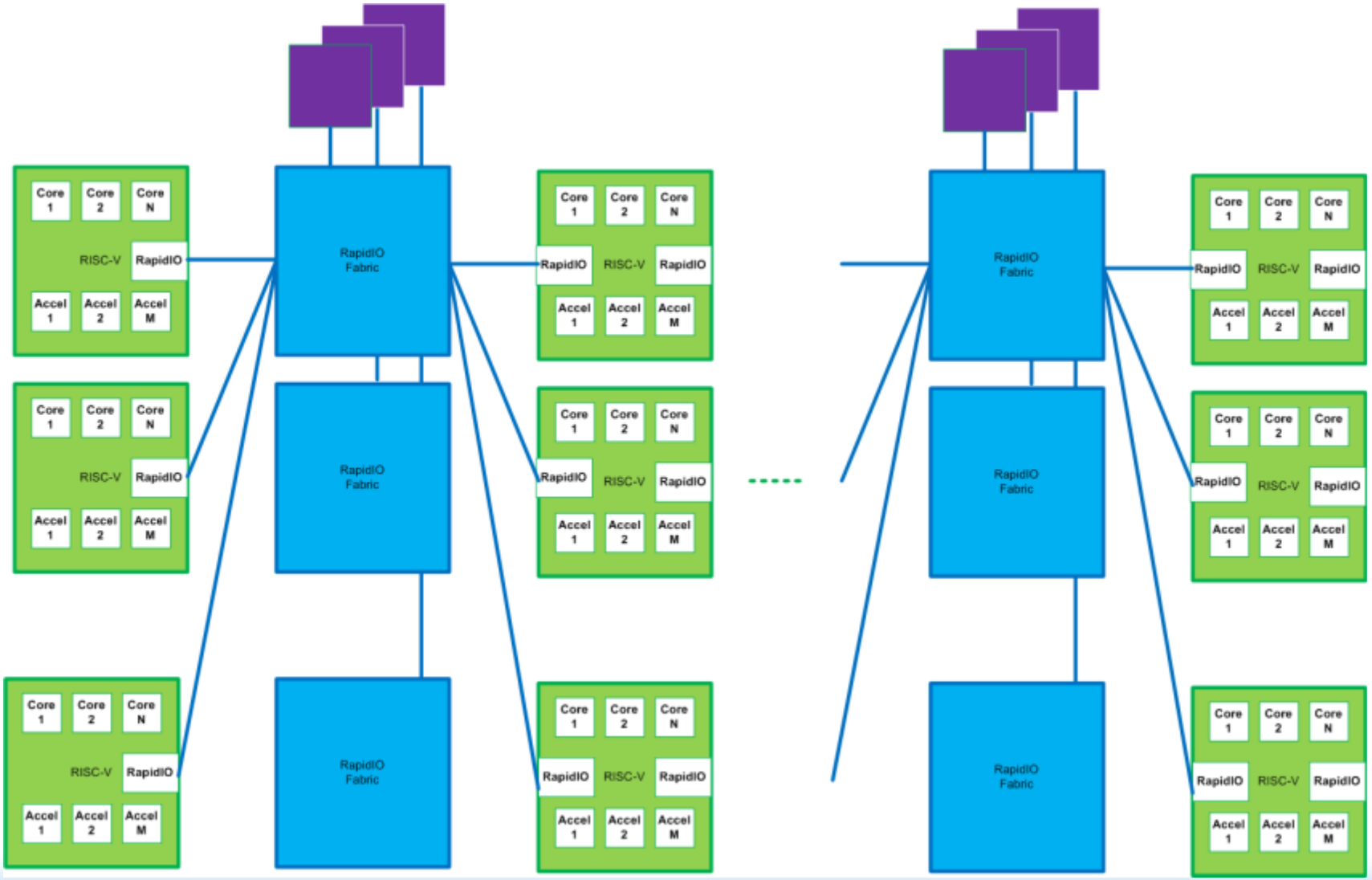
+ Shared memory between CPU and GPU
+ Supports Scale-out System
- Shared Memory and Storage challenging without hardware support
+ Switch latency excellent (~100 nsec)
- Memory to memory latency can be improved

Scale-out Analytics Clustering Model



+ Large Scale-out Analytics Model
+ Low latency < 200 - 300 nsec
+ Low Power
+ Shared Memory/Storage Pool
+ Open ISA Model
+ Standard Fabric

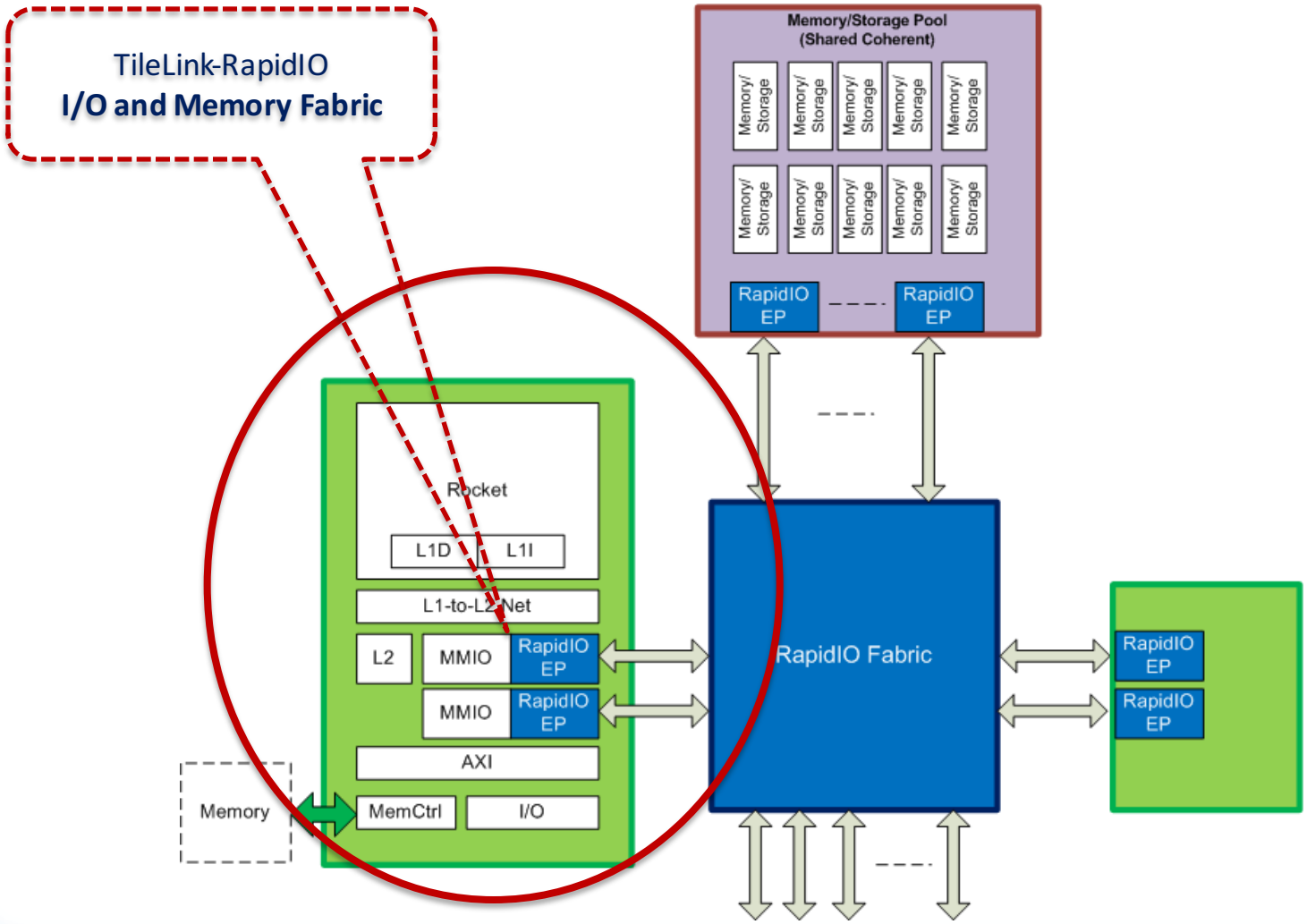
RISC-V Scale-out Analytics Clustering Model



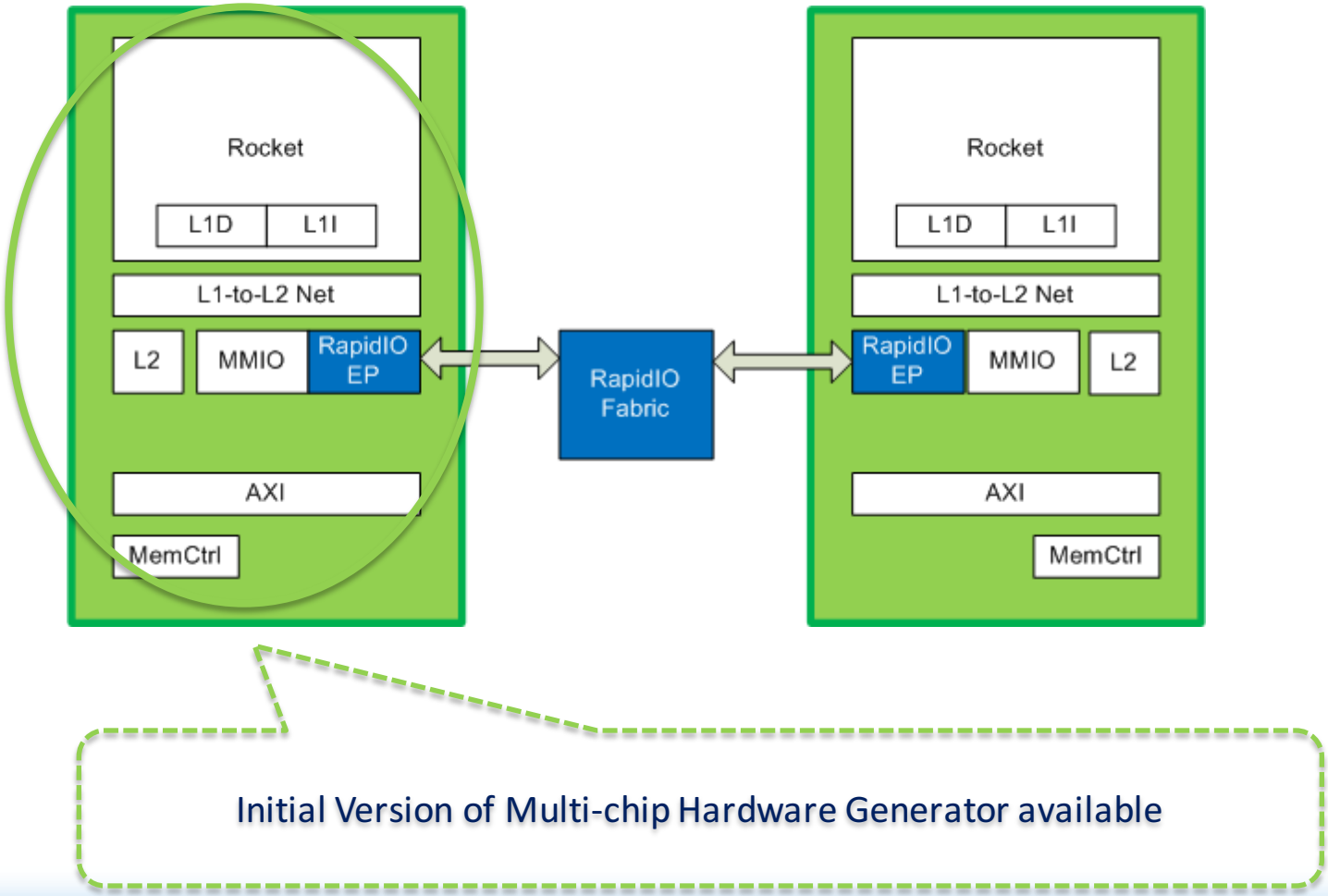
Fabric
TileLink and AMBA

- **Off-chip Fabric**
  - Scale-out and Remote Memory Acceleration with RapidIO
- **On-chip Fabric**
  - **TileLink**
    - UC Berkley
    - Hierarchical/Deadlock Free/Built-in Message types (supports Accelerators)
    - Extensible with Custom Message for Coherency (5-state and others)
  - AMBA
    - ARM Inc.
    - AMBA AXI and ACE Protocol Specification
      - Issue E. 22 Feb, 2013
    - CC-NUMA Architecture, Up-to 5 State Model

RISC-V Architecture with RapidIO

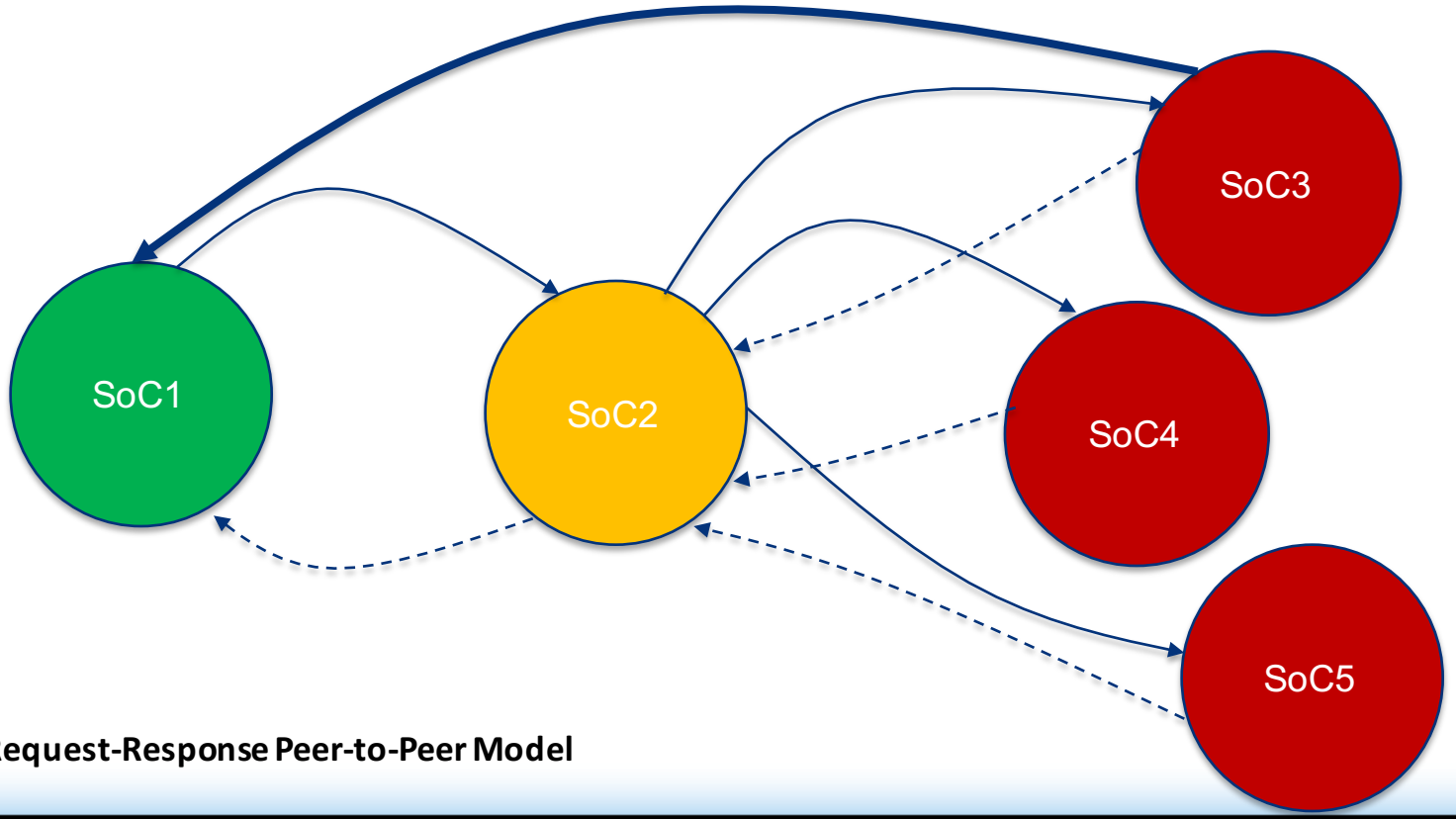


RISC-V Multi-chip Generator



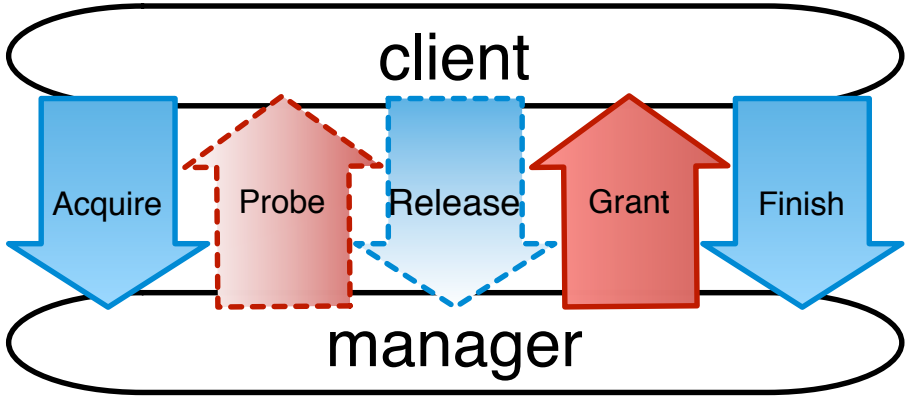
Remote Memory/IO Scale-out Packets

DestinationID	SourceID	FlowControl	Addressing	TYPE	Payload	Protection



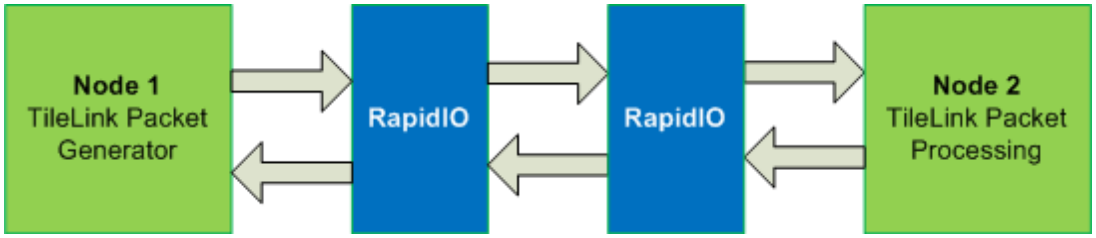
Request-Response Peer-to-Peer Model

TileLink Basic Protocol and RapidIO Packet Structure



	+0								+1								+2								+3									
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31		
Byte 0	ackID					vc	CRF		prio		tt=10		Ftype				destination ID [7:0]								source ID [7:0]								W	0
Byte 4	TType (transaction)								rQoS				rd-/wr-size				srcTID								rSizeBurst								W	1
Byte 8	ud				r B		rsvd		rOffset				rsvd				rUnion																W	2
Byte 12	Address																Address																W	3
Byte 16	Address																Address																W	4
Byte 20	Payload (Word 0)																																W	5
Byte 24	Payload (Word 1)																																W	6
	...																																	
	Payload (Word 15)																																W	20
	CRC																Padding																W	21

Hardware Simulation Model



RapidIO Egress

RapidIO Ingress

Target Latency CPU to CPU ~ 100 – 200 nsec

Summary

- **Scale-out Analytics with balancing Fabric and Computing**
  - Scale-out through distributed Edge Computing Model
    - Reduces round-trip latency
    - Reduces access to core network bandwidth
  - RISC-V with integrated Fabric for I/O and Memory
    - Fabric for Any Topology (Mesh/Hypercube)
    - Low Latency directly from SoC on-chip Fabric
    - Supports TileLink scale-out for multi-node clustering
    - Enables Balanced Coherent and scale-out architecture

RISC-V CPU Generator model with Port for RapidIO available!