Supercomputing at the Mobile Edge

In Supercomputing at the Edge, driven from the desire to improve the management of Big Data, we push portions of data center content to the edge of the access network. This improves the user experience, enables next generation applications and revenue generating services.

RAPIDIO® FOR MOBILE EDGE COMPUTE:

High Performance & Low Power

- Up to ½ pflop per rack GPU system demonstrated
- 100ns cut through latency for switches
- Sub-microsecond memory-to-memory transfers latency with packet termination and reliable transmission
- Typical switch power is approximately 300mW per 10Gbps

Open Scalability

- Open standard processor interconnect
- Open standard 1U server design
- Heterogeneous connectivity for x86, GPU, FPGA, and DSP, etc.
- Peer-to-peer, "any-to-any" topologies from 1Gbps to 40Gbps
- Available top-of-rack (ToR) switches for rapid scale out and deployment
- Performance enables disaggregated, rack scale system architectures essential for NFV / SDN support

Accelerated Development and Ease of Deployment

- "Bolt On" MEC server to existing RapidIO-based base stations
- Large ecosystem of merchant systems, software, and silicon
- Open standard software APIs, drivers and fabric management
- Use x86 processors with IDT Tsi721 PCIe®-to-RapidIO bridge
- RapidIO rides on standard electrical specifications, supporting standard hardware channel designs
- Available copper and optical connectivity options

Benefits of Supercomputing at the Edge:

- Applications can be distributed to the network closer to the users
- Middle mile bottleneck congestion is mitigated
- Real-time application driven analytics are supported
- Augmented Reality becomes reality
- M2M communications enables real-time control such as in autonomous vehicles
- Video compression / decompression reduces network load
- Local caching of hot data such as video reduces network load while improving user experience
- New revenue generation models enabled to the service providers

IDT Develops a New Architecture for a Better Way

IDT has introduced an architecture that can put low-latency supercomputing-caliber computing and interconnect with heterogeneous computing at the edge of the wireless network. The combination of IDT's low-latency RapidIO interconnect, timing, power management and CPRI IQ compression solutions deployed in the field at the macro base station, the central office and the cloud radio access network (C-RAN), enable this new deployment paradigm.

Heterogeneous MEC board block diagram.

40 Gbps S-RIO per port to front panel
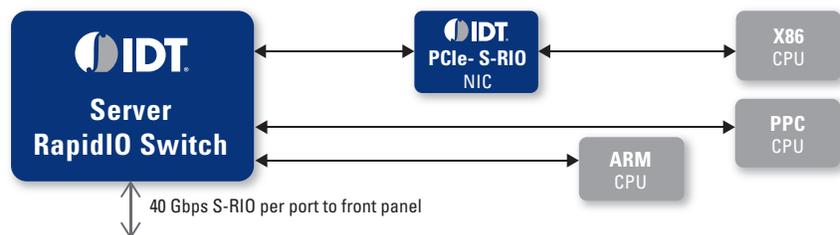
IDT and its partners have leveraged low-power RapidIO interconnect to demonstrate a GPU-based compute architecture which scales to 0.5 petaflops within a single rack. It combines IDT's 100 ns latency RapidIO interconnect technology, IDT timing technology and NVIDIA's Mobile Tegra K1 GPU technology using Prodrive Technologies' RapidIO-enabled servers, and Concurrent Technologies' GPU cards with embedded RapidIO interconnect. Supercompute densities such as these make big network data analysis, video and image streaming and analytics, and many other new applications possible.

Supercomputing at the Mobile Edge
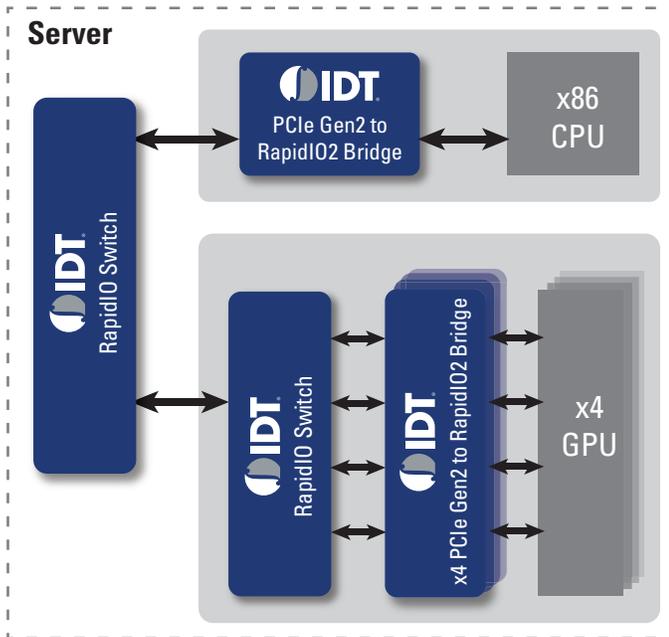
Server

MEC with GPU acceleration: IDT and its partners have demonstrated this system, which scales to 1/2 petaflop per rack.

Accelerating Mobile Edge Compute (MEC) with RapidIO

MEC servers architected on RapidIO take advantage of the 100ns latency and throughputs of up to 40G. With a large ecosystem of standards-based boards such as those based on the Advanced Mezzanine standard and open source software, system and software development is also accelerated. These platforms can quickly be installed and are effectively "plug-n-play" with RapidIO connectivity that is ubiquitous in existing installed mobile network equipment, as well as many future platforms. Such systems are readily scaled over RapidIO, with large scale systems quickly developed based on commercially available top-of-rack switches such as those from Prodrive Technologies.

Paving the Future in 5G

- 40Gbps with roadmap to 100 Gbps+ announced
- Time Synchronization and Distribution for complete network timing requirements similar to IEEE1588
- ARM 64-bit scale out over RapidIO will enable next generation MEC server & compute with up to 100s of coherent processors
- Backward compatibility preserves existing hardware and software infrastructure and investment

Edge Compute Hardware Available Today

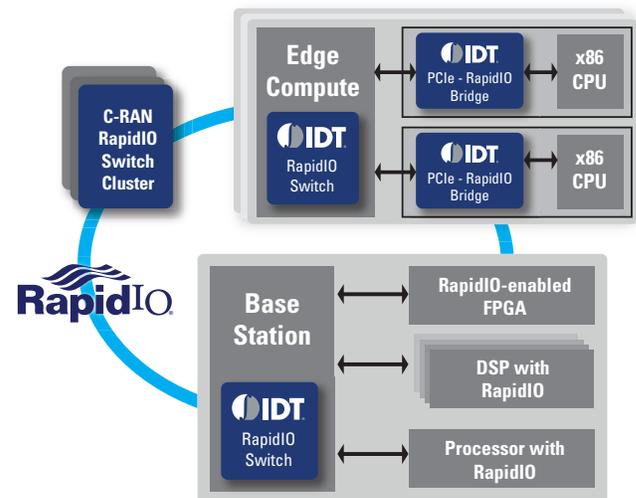1U RapidIO Server

GPU Compute

x86, ARM, PowerPC-based Servers

Top-of-Rack RapidIO Switching

Scale Out with C-RAN

C-RAN platform with Mobile Edge Compute. Edge Compute servers can be plugged into existing C-RAN infrastructure through available RapidIO top-of-rack (ToR) switching, for example.

To request samples, download documentation, or learn more, visit: **idt.com/go/MobileEdge**