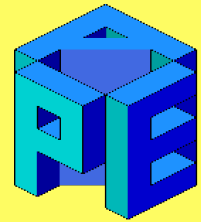


The APEnet interconnect

3D toroidal GPU-optimized network



The N-Dimensional torus network is a well established solution to interconnect parallel computing system dedicated to a broad class of scientific applications (domain decomposition, stencil computation, ...).

The APEnet is an interconnection architecture implementing a 3D Torus network optimized for scientific computation on GPU-accelerated clusters.

APEnet+, the current release, is an FPGA-based full-length double-slot PCI-Express card.

APEnet+ is optimized for high bandwidth and low latency:

- support for RDMA communication paradigm allowing for zero-copy approach;
- support for "NVidia GPUDirect" P2P communications avoiding buffer copies between GPU and host and excellent GPU-to-GPU latency.

APEnet+ is scalable and cost effective interconnection solution:

- up to 32k computing nodes in current implementation
- no external switching hardware required, only card and cables.

APEnet+ Advanced features:

- Hardware support for system fault tolerance
- Hardware acceleration of specific software tasks

APEnet 2012 development roadmap:

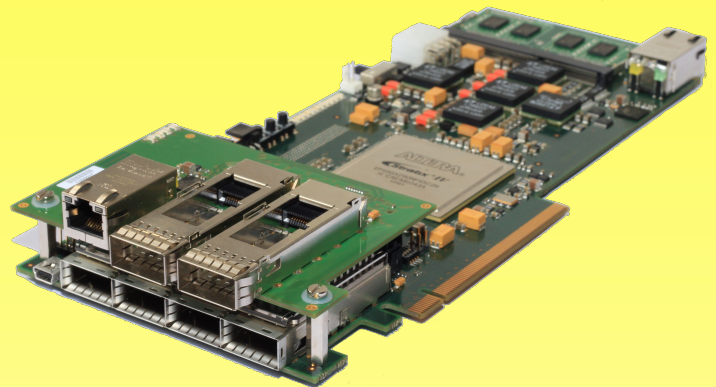
- PCI-Express Gen3 x16
- Link speed enhancement (up to 56 Gb/s)

The APE Group

The APE research group has been designing and developing in the area of HPC and Embedded systems for more than 25 years.

Web site:

<http://apegate.roma1.infn.it>



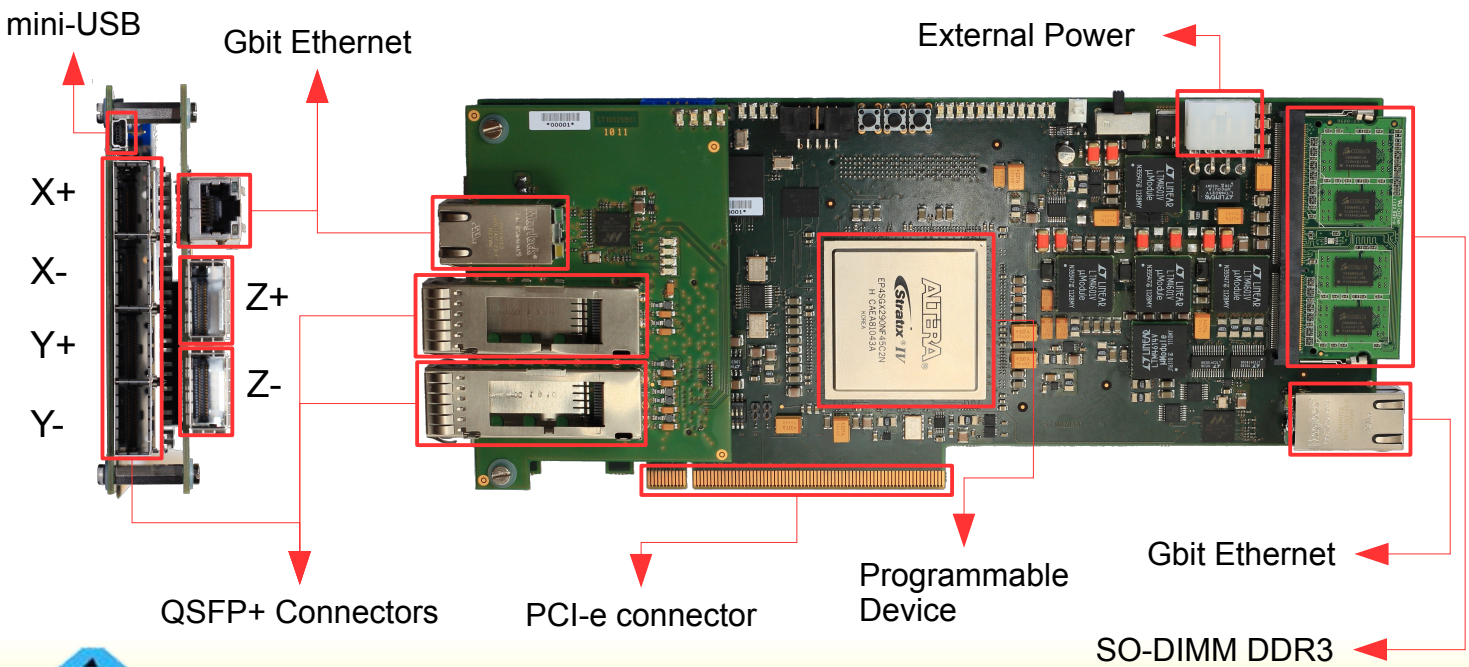
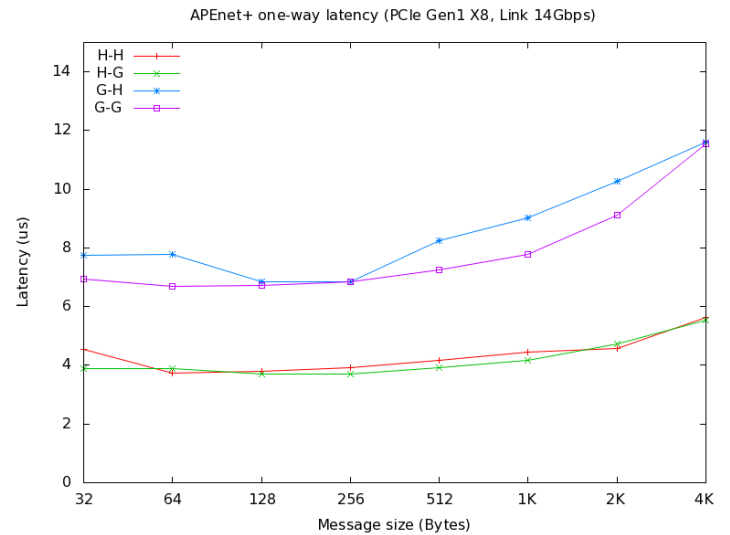
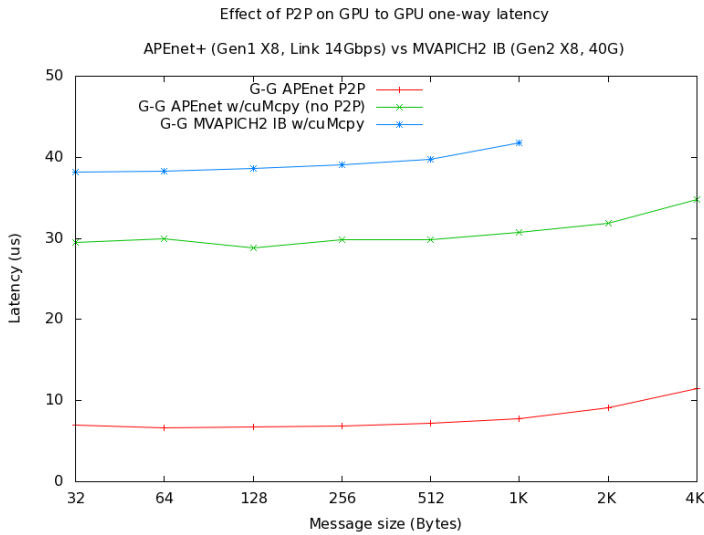
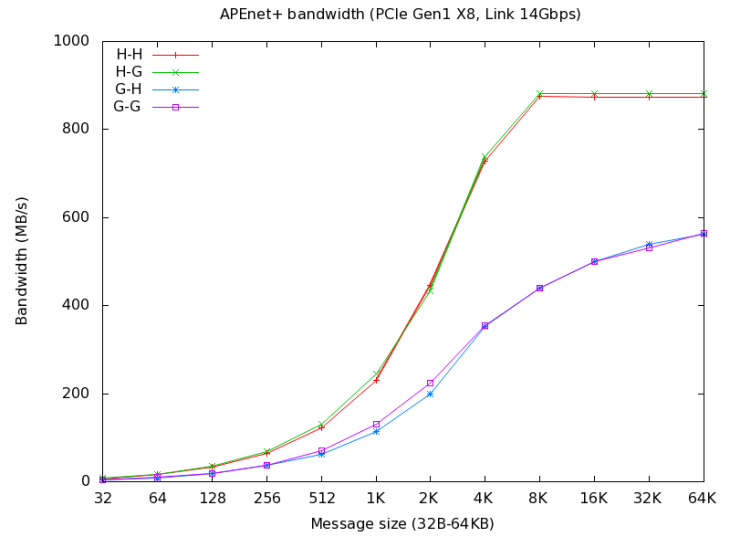
Feature Summary

- ◆ Host Bus Interface Specification:
 - ➔ PCI Express Gen 2 x8
 - ➔ 2.5 or 5.0 GT/s link rate
 - ➔ Autonegotiate x8 x4 x2
- ◆ Connectivity:
 - ➔ 6 Bi-Directional 34 Gbps Links
 - ➔ Passive Copper Cables
 - ➔ Optional Optical active Cables
 - ➔ QSFP+ connectors compliance
- ◆ Physical Specification:
 - ➔ 11.0 x 26.5 cm
 - ➔ 2 PCI I/O slot wide in 6 links configuration
 - ➔ 1 PCI I/O slot wide in 4 links configuration
- ◆ Tools:
 - ➔ Linux Host Driver
 - ➔ API Library
 - ➔ OpenMPI module
- ◆ Advanced Features:
 - ➔ RDMA Communication Paradigm
 - ➔ NVidia GPUDirect P2P Communications
 - ➔ up to 32³ supported toroidal mesh
 - ➔ Auto-routing of packets
 - ➔ 2 Virtual Channel per Link
 - ➔ FPGA based
 - ➔ Tested on Linux x86_64
- ◆ Working on:
 - ➔ Support for System fault tolerance
 - ➔ Collective offloading
 - ➔ Collective re-routing on on-board Ethernet

Benchmarking

- ➔ Preliminary performance test measures has been performed on early prototype hardware.
- ➔ 2 Xeon-based computing nodes each with
 - ➔ NVIDIA C2050
 - ➔ Apenet+
- ➔ Link speed is limited to 14 Gb/s
- ➔ Gen 1 speed PCIe sockets

We demonstrated a minimum **GPU to GPU latency** for small messages of **6.6 us**.



Contact info:
<http://apegate.roma1.infn.it>
 piero.vicini@roma1.infn.it
 Tel/Fax: +39 0649914423

INFN Sezione di Roma
 Piazzale Aldo Moro 2
 00185 Roma