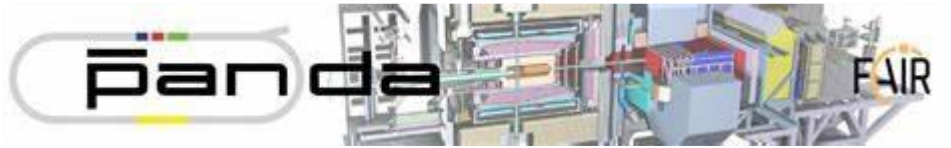
Karlsruhe Institute of Technology

Status of High-flex and novel DAQ architecture for PANDA

Michele Caselle, Suren Chilingaryan, Timo Dritschler, Andreas Kopmann, Weijia Wang



Outline

- Status of High-Flex PCIe card
  - Status and what's next

- Novel heterogenous FPGA-GPUs DAQ architecture based on emerging ethernet protocols

  → *An evolution of the GPUDirect technology over Ethernet / InfiniBand*

- High-throughput, low latency networking with RDMA over InfiniBand and Ethernet, using RoCE extension

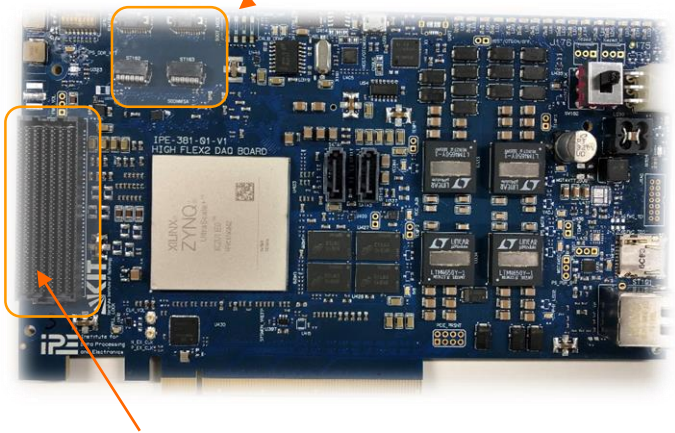
  → Next talk by Timo Dritschler (KIT)

Development and production of High-Flex

- First version of the High-Flex card
  - Successfully produced and High-Flex fully tested
  - Results / performances already presented during the previous PANDA CM 20/1 (DAQ session)
  - Minor layout adaption was necessary

- Second version of the High-Flex
  - Layout completed ✓
  - Two cards produced and partially tested
  - One card sent to external company for FPGA mounting
  - First card with FPGA mounted, will be available at IPE (next week)
    - Final test and characterizations
    - Card potentially available for PANDA

Firefly optical connection, from speed grade 1 to 2
(up to 28 Gb/s per lane)

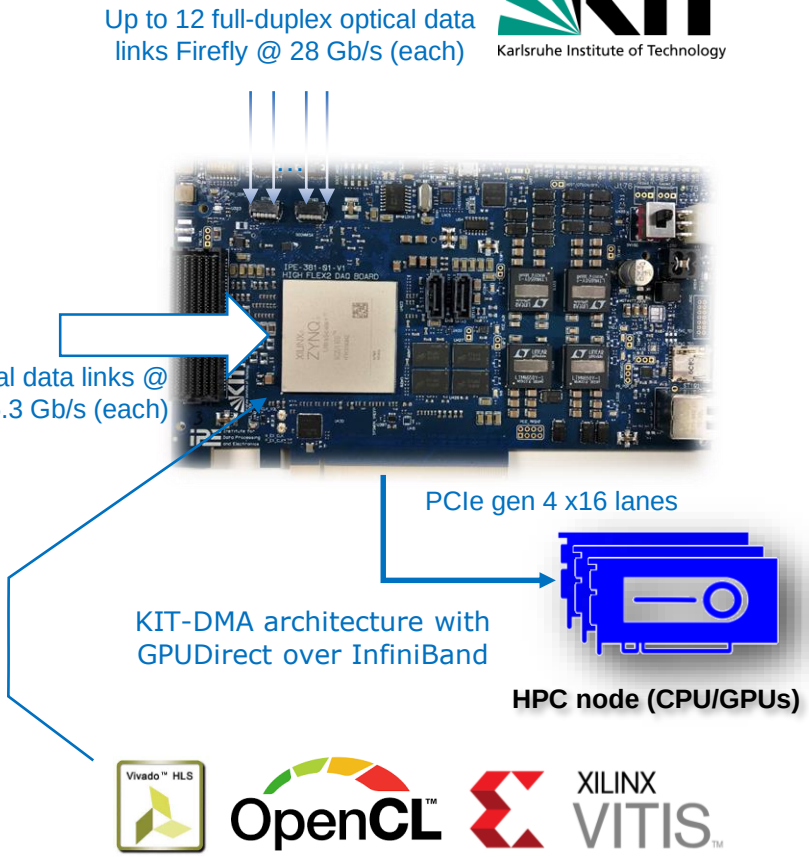


FMC+ improved the layout of few GTH/GTY transmission lines
(not necessary for PANDA)

Just to remind …

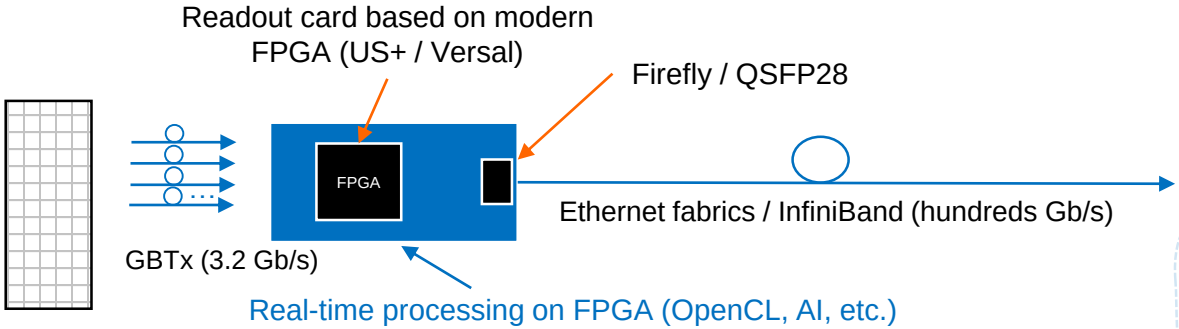
- Developed within the project “Detector Technology and System Platform” of the Helmholtz Association of German Research Centers (HGF) coordinated at KIT (POF III)
- Application fields: HEP, beam diagnostics, superconducting sensors and quantum technologies, AI, many others ...

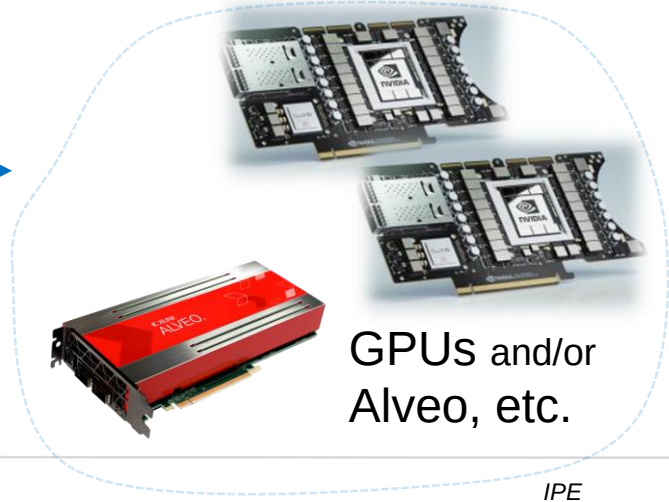
- Features:
  - Developed for massive data throughput (> 210 Gb/s)
  - Firmware based by KIT-DMA for heterogenous FPGA-GPUs
  - Powerful ZYNQ US+ devices (up to 3K DSP slides)
  - Focus on algorithms (Host ↔ Kernel architecture)
  - Data Processing on FPGA: C/C++, OpenCL, Python, MATLAB
  - Fast AI inference on ZYNQ: HLS4ML, DPUs (Deep-neural Processor Units)



What's about the next evolution of DAQ ?

- New challenges: scalability, commercially available hardware, easily upgradeable and easy to maintain, etc.
  - Readout card based on multiple high-throughput emerging ethernet connections (no PCIe)
  - Directly connected to commercial devices: ethernet switches, GPUs, accelerator cards (i.e. Xilinx Alveo)
  - Focus on algorithms, High-level synthesis (C/C++, OpenCL, AI) → to remove major barrier on hardware development and allowing developers *with little or no FPGA expertise* to deploy complex data processing on FPGA

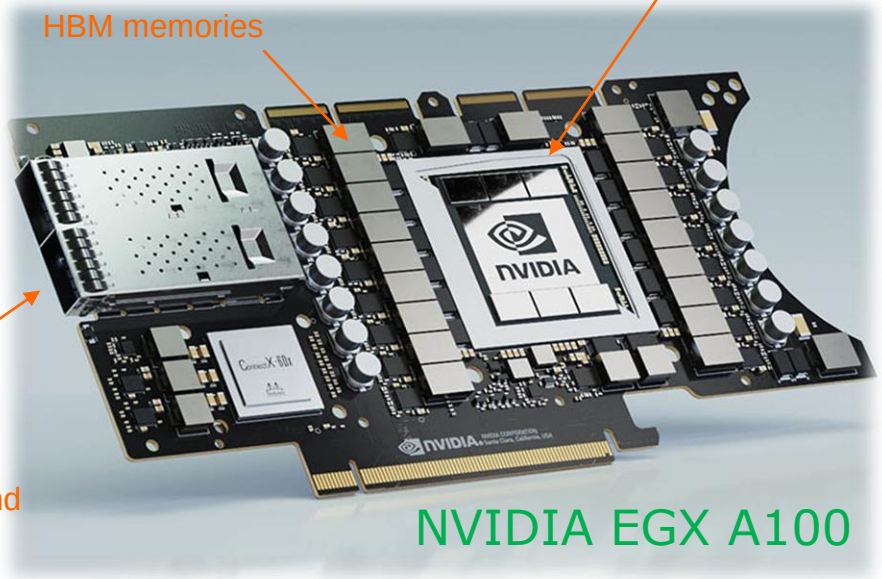




Heterogenous FPGA –GPUs/CPUs

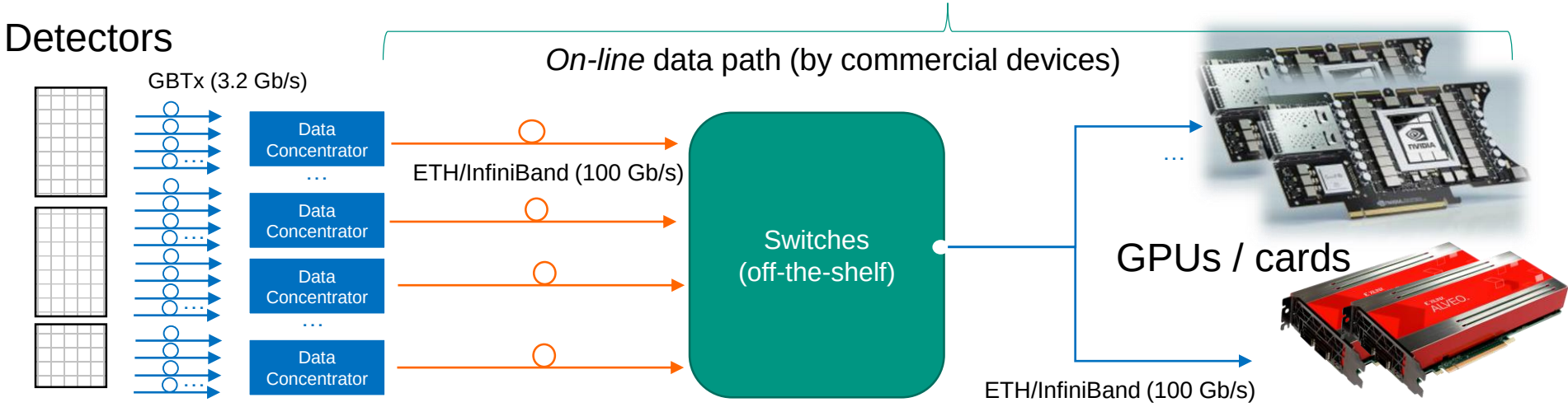
New generation of GPU cards

- New generation combines GPU and Network (NIC) all in a single PCIe card.
- Features:
  - NVIDIA A100 Ampere-based GPU
  - Mellanox ConnectX-6 Dx NIC
  - Up to 200 Gbps on a single card
  - InfiniBand for GPU-to-GPU communication
  - New concept of “GPUDirect” over ETH



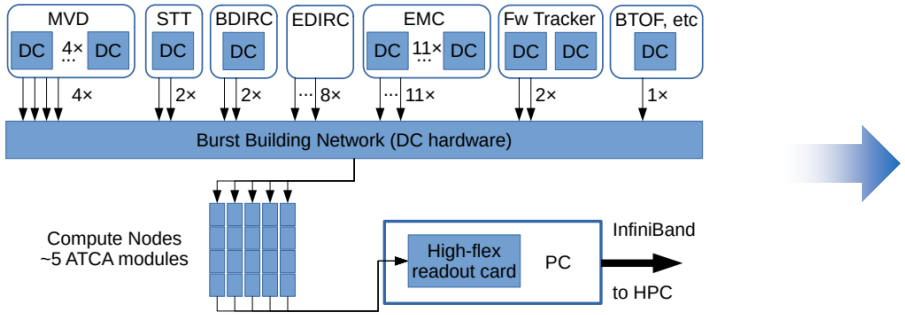
DAQ for detector

- More complex DAQ scheme with many sub-detectors
- Modern FPGA devices provide multiple data-links over 100 Gb/s → dramatic reduction of the number of off-detector cards and optical data links
- On-line data path covered by off-the-shelf devices → dramatic reduction of custom electronics
- Heterogenous architecture



Possible evolution of PANDA - DAQ

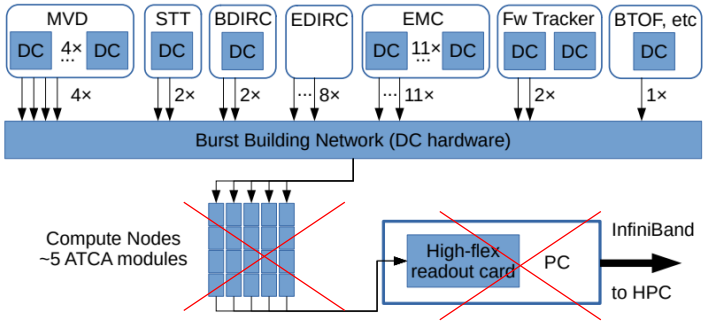
TDR (Figure 4.27)



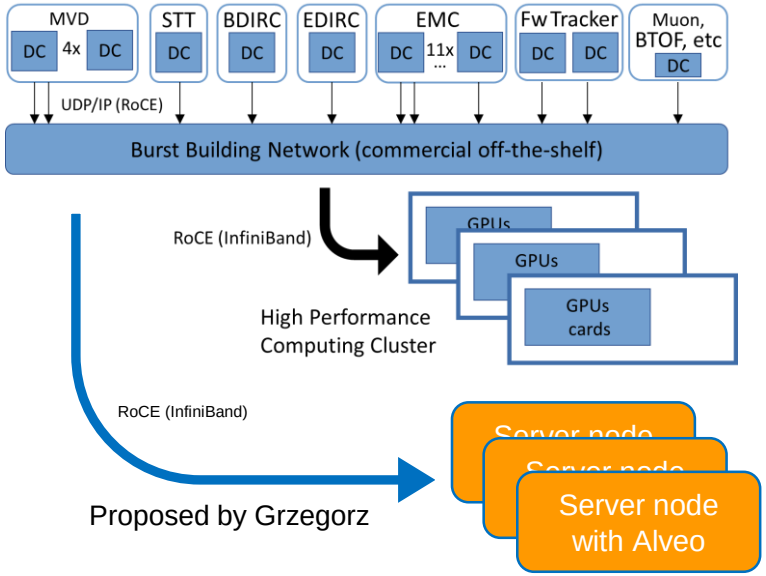
Possible evolution of PANDA - DAQ

- New DAQ architecture very similar to the proposed DAQ scheme from Dr Grzegorz Korcyl (PANDA CM 20/1)

TDR (Figure 4.27)



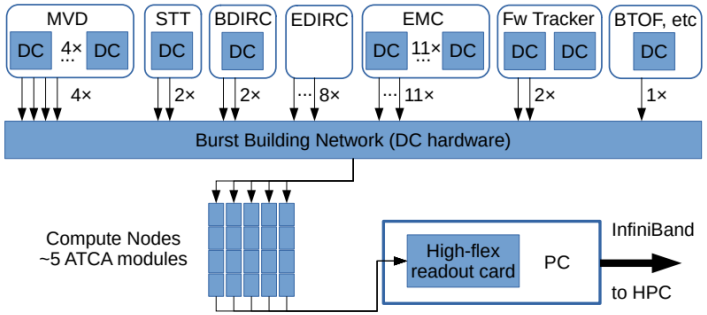
TDR (Figure 4.28)



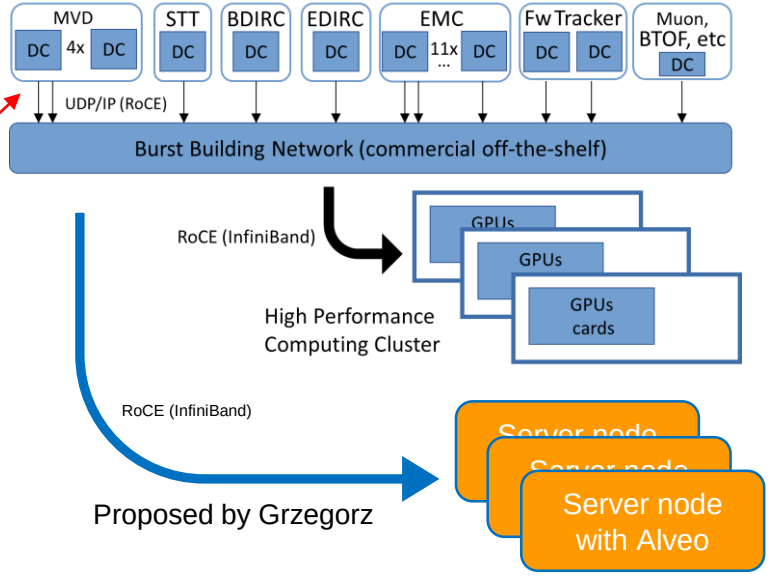
Possible evolution of PANDA - DAQ

- New DAQ architecture very similar to the proposed DAQ scheme from Dr Grzegorz Korcyl (PANDA CM 20/1)

TDR (Figure 4.27)



TDR (Figure 4.28)

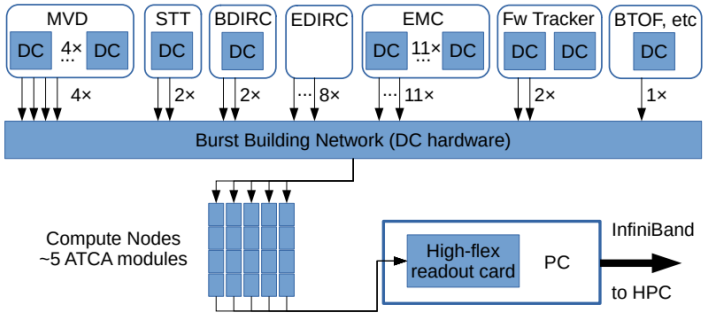


- What's about the FPGA firmware at level DC ?
- How unify the connection between FPGA – "off-the-shelf" devices ?
- InfiniBand (IP-core) on FPGA → extremely expensive
- Review of the DAQ TDR → suggested to explore UDP / simple TCP protocol

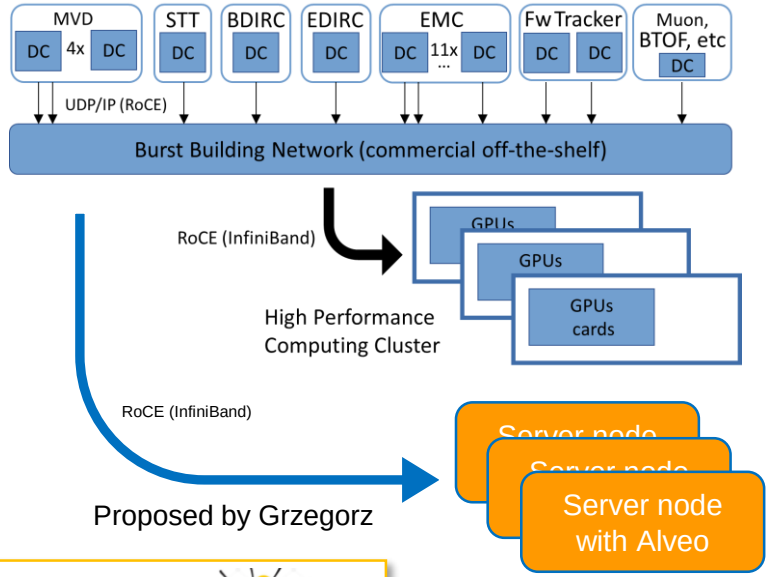
Possible evolution of PANDA - DAQ

- New DAQ architecture very similar to the proposed DAQ scheme from Dr Grzegorz Korcyl (PANDA CM 20/1)

TDR (Figure 4.27)



TDR (Figure 4.28)

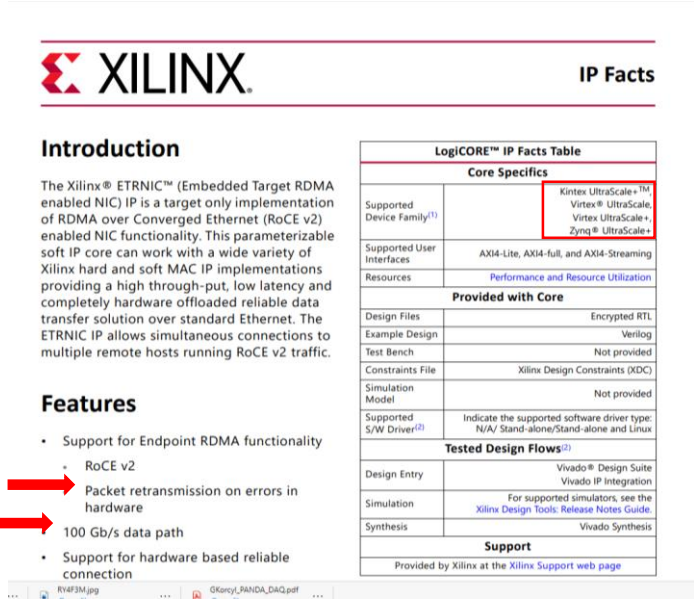


- What's about the FPGA firmware at level DC ?
- How unify the connection between FPGA – "off-the-shelf" devices ?
- InfiniBand (IP-core) on FPGA → extremely expensive
- Review of the DAQ TDR → suggested to explore UDP / simple TCP protocol

Why not RoCE for PANDA ?

RoCE on FPGA

- RoCE (RDMA over Converged Ethernet) is a **standard**, offers a **non-complex UDP/IP protocol** which could incapsulate InfiniBand packets
- Higher RDMA performance over 100G direct implementation on modern FPGA by the Xilinx IP-core (free), ETRNIC (Embedded Target RDMA enabled NIC)
- Fully supported by commercial devices: switches, GPUs, Alveo, etc.
  - ***Toward an unified ETH protocol for PANDA***
- RoCE IP-core should be fully compatible with the current Data Concentrator
- Available a KIRO (KIT InfiniBand Library) which support also RoCE
  - Library potentially ready for PANDA phase 0 and evaluation tests



XILINX

IP Facts

Introduction

The Xilinx® ETRNIC™ (Embedded Target RDMA enabled NIC) IP is a target only implementation of RDMA over Converged Ethernet (RoCE v2) enabled NIC functionality. This parameterizable soft IP core can work with a wide variety of Xilinx hard and soft MAC IP implementations providing a high through-put, low latency and completely hardware offloaded reliable data transfer solution over standard Ethernet. The ETRNIC IP allows simultaneous connections to multiple remote hosts running RoCE v2 traffic.

Features

- Support for Endpoint RDMA functionality
  - RoCE v2
  - Packet retransmission on errors in hardware
  - 100 Gb/s data path
- Support for hardware based reliable connection

LogiCORE™ IP Facts Table	
Core Specifics	
Supported Device Family[1]	Kintex UltraScale+™, Virtex® UltraScale, Virtex UltraScale+, Zynq® UltraScale+
Supported User Interfaces	AXI4-Lite, AXI4-full, and AXI4-Streaming
Resources	Performance and Resource Utilization
Provided with Core	
Design Files	Encrypted RTL
Example Design	Verilog
Test Bench	Not provided
Constraints File	Xilinx Design Constraints (XDC)
Simulation Model	Not provided
Supported S/W Driver[2]	Indicate the supported software driver type: N/A/ Stand-alone/Stand-alone and Linux
Tested Design Flows[2]	
Design Entry	Vivado® Design Suite, Vivado IP Integration
Simulation	For supported simulators, see the Xilinx Design Tools: Release Notes Guide.
Synthesis	Vivado Synthesis
Support	
Provided by Xilinx at the Xilinx Support web page	

More details: Timo's talk

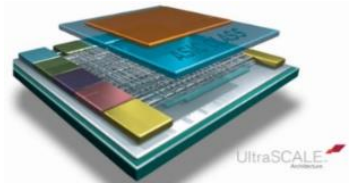
Conclusions & what's next

- High-Flex is potentially available for PANDA (mid-Nov. 2020)
- Proposed DAQ scheme for PANDA based on off-the-shelf devices
  - DAQ scheme well compatible with the proposed DAQ from Dr Grzegorz Korcyl
- Proposed an unified ETH connections for PANDA
  - To connect FPGA cards and commercial devices
- RoCE IP-core could potentially implemented on existing devices and future Data Concentrator cards
  - A FPGA "common readout infrastructure" based on RoCE is under development at KIT
- Library KIRO available for PANDA

Many thanks for your attention

Backup

FPGA - AI platform: HighFlex 2



- Processor System (ARM): User Applications
- Programmable Logic (FPGA): fast and low latency application

Front-End:
KAPTURE
KALYPSO

- 12 lanes @ 28 Gbps
- Data throughput (full-diplex) up to 336 Gbps

- Compatible with standard FMC

PCI EXPRESS

- PCIe Gen 4 (x8 or 16 lanes)
- Data rate up to 240 Gbps

- Two SSD raid Local data storage

HighFlex 2 FPGA: ZYNQ Ultrascale plus

- System Logic Cell: 653 K
- Flip-Flops: 597 K
- LUT: 298 K
- Distributed RAM: 9,1 Mb
- Block RAM: 21.1 Mb
- UltraRAM: 22.5 Mb
- DSP Slices: 2928
- PL-DDR4: 2 GB

AMBA AXI4 interfaces for primary data communication

- Quad-core Arm Cortex-A53
- NEON & Single/Double Precision Floating
- PS-DDR4: 4 GB