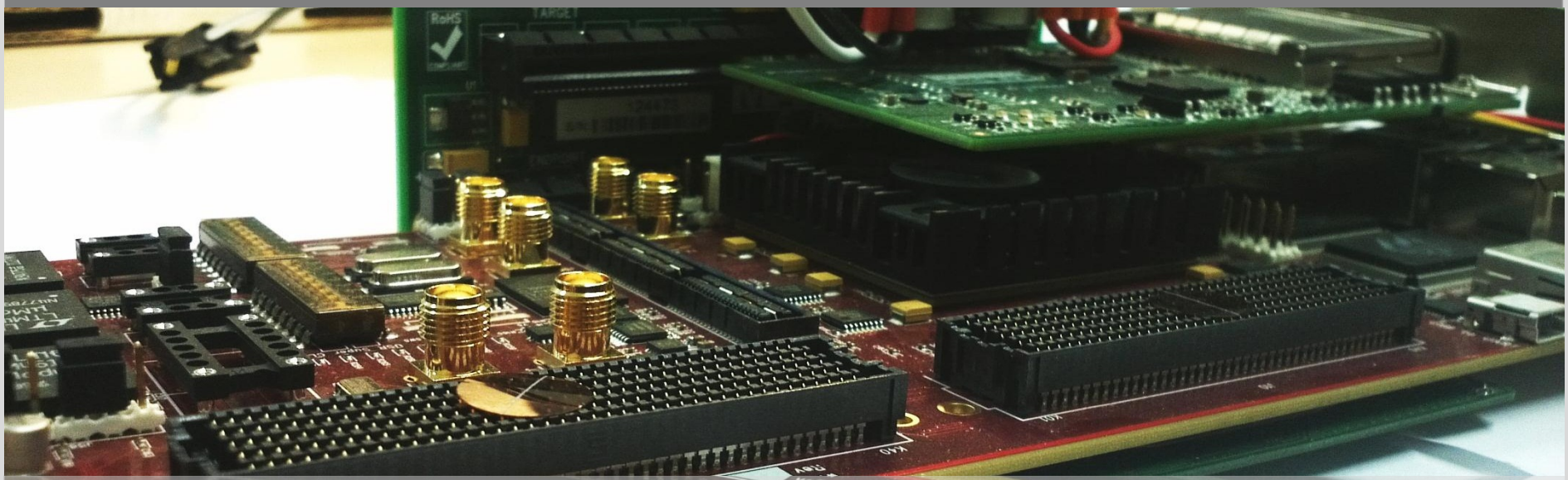
Karlsruher Institut für Technologie

An FPGA platform for ultra-fast data acquisition

M. Caselle, M. Balzer, S. Chilingaryan, A. Kopmann, U. Stevanovic, M. Vogelgesang

FPGAs in Research - Applications, Technologies and Tools, Forschungszentrum Jülich, 3-4 December 2012

KIT, Institut für Prozessdatenverarbeitung und Elektronik
UFO project group



Ultra Fast X-ray Imaging (ANKA/UFO experimental station)

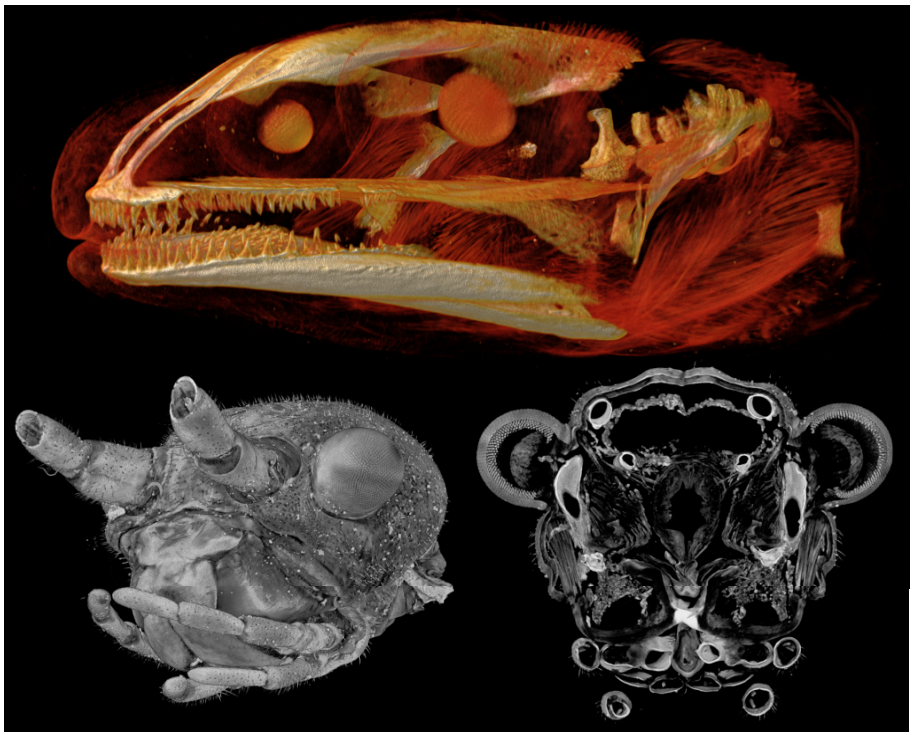
UFO → **U**ltra-**F**ast X-ray Imaging of Scientific Processes with **O**n-line Assessment and Data-driven Process Control

High spatial resolution (<1 µm) included 2D and 3D visualizations

+

Time resolution (2D: ≈10kHz, 3D: ≈10Hz) to give insight in the temporal structure evolution and thus access to dynamics of processes

Main application fields: medical diagnostics, biology, non-destructive testing, materials research and etc.

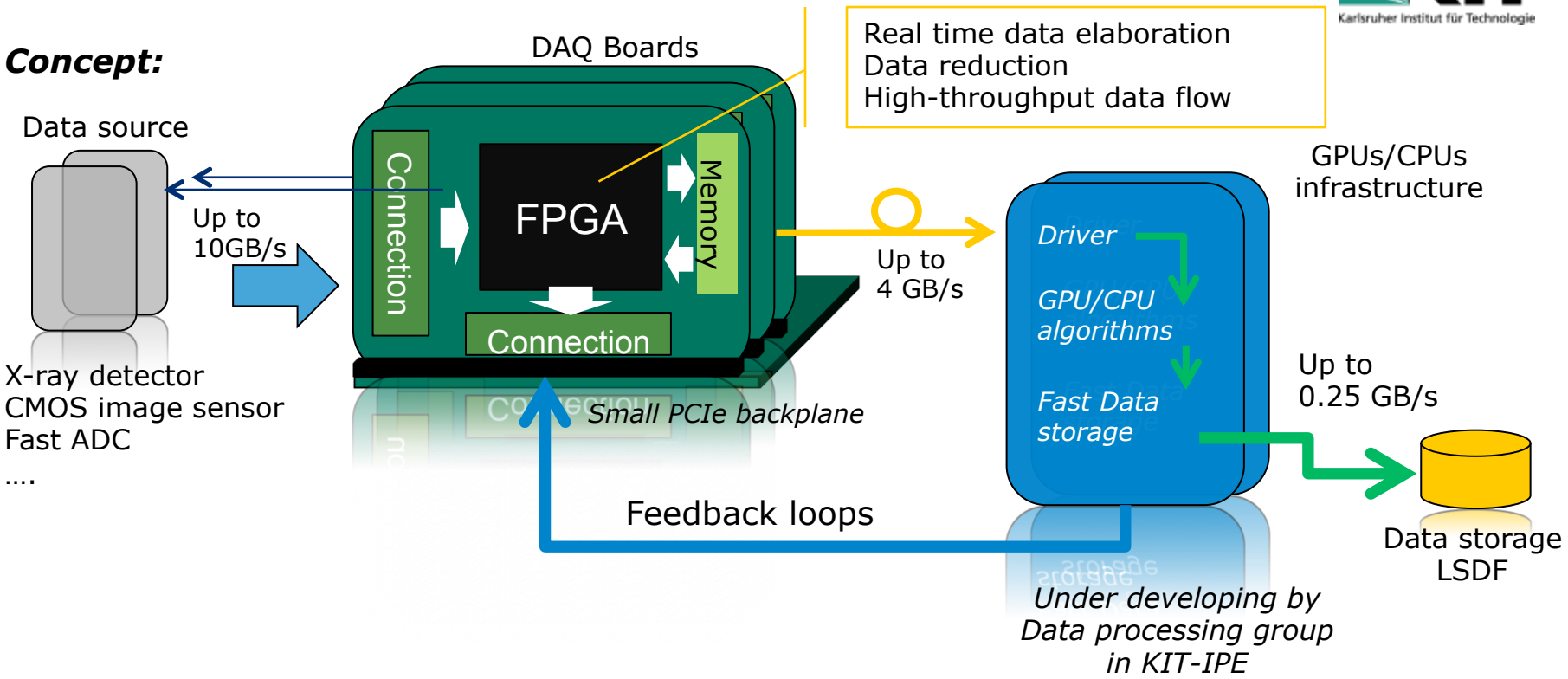


Requirements:

High granularity and **low noise** monolithic silicon pixel detector, few µm pixel pitch, several MPixel matrix operating at several **kframes/sec**

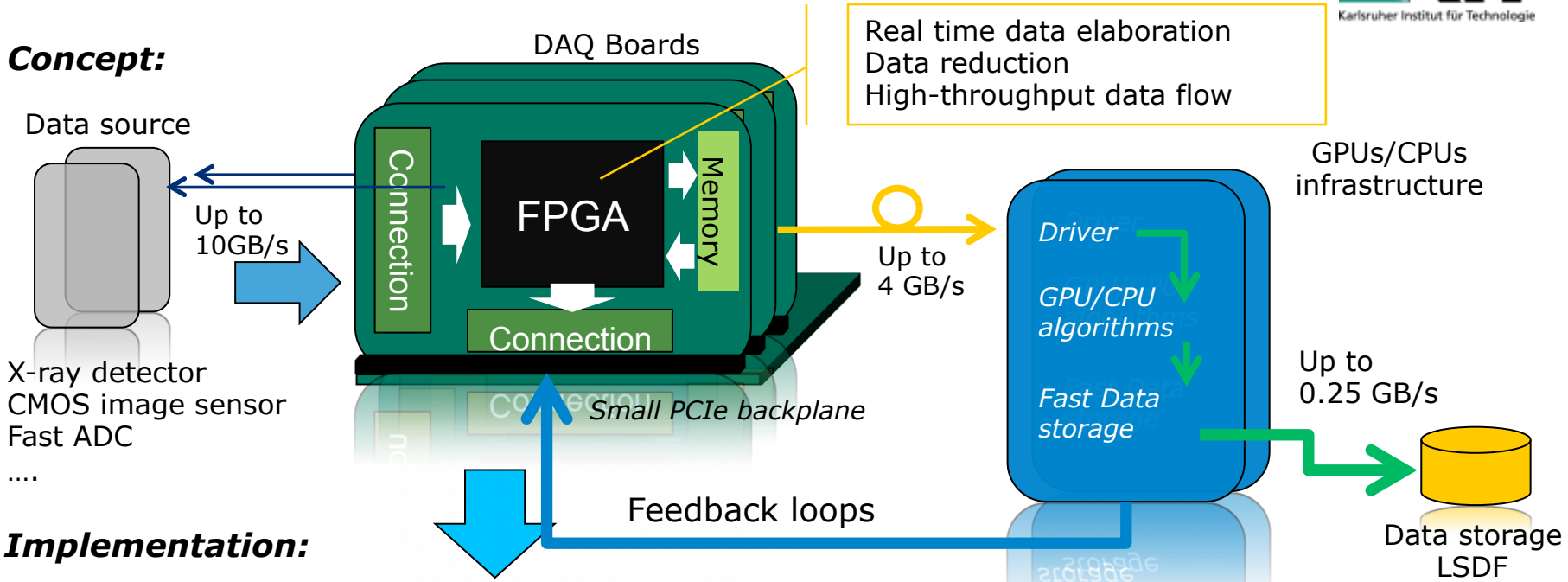
*High readout bandwidth up to **50Gb/s with GPU** (3D-tomography reconstruction)*

KIT-IPE – Readout concept of high data throughput for scientific applications

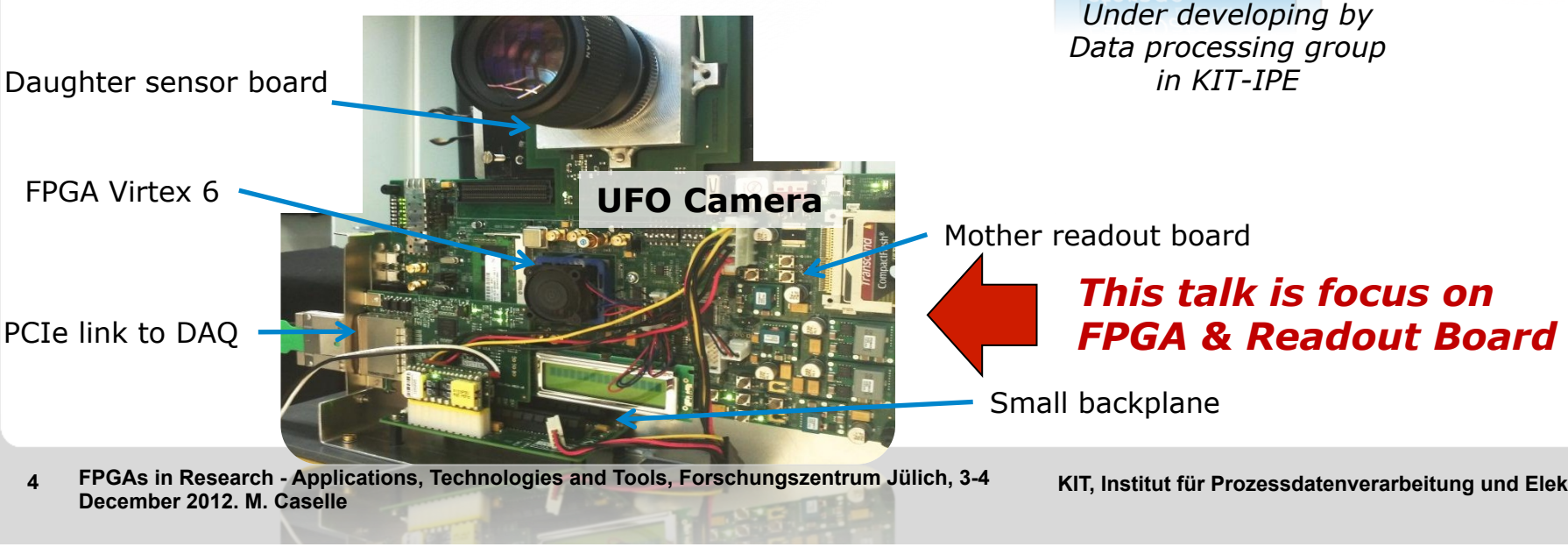


KIT-IPE – Readout concept of high data throughput for scientific applications

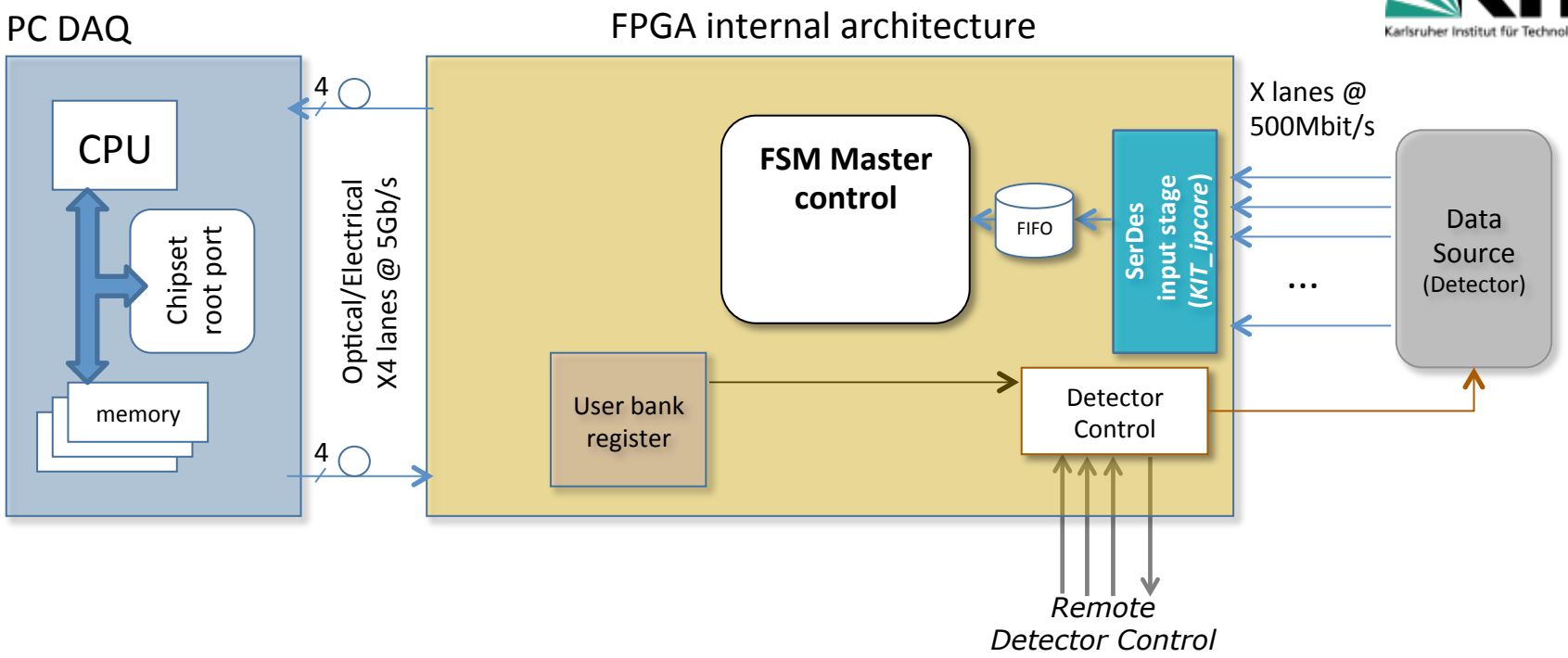
Concept:



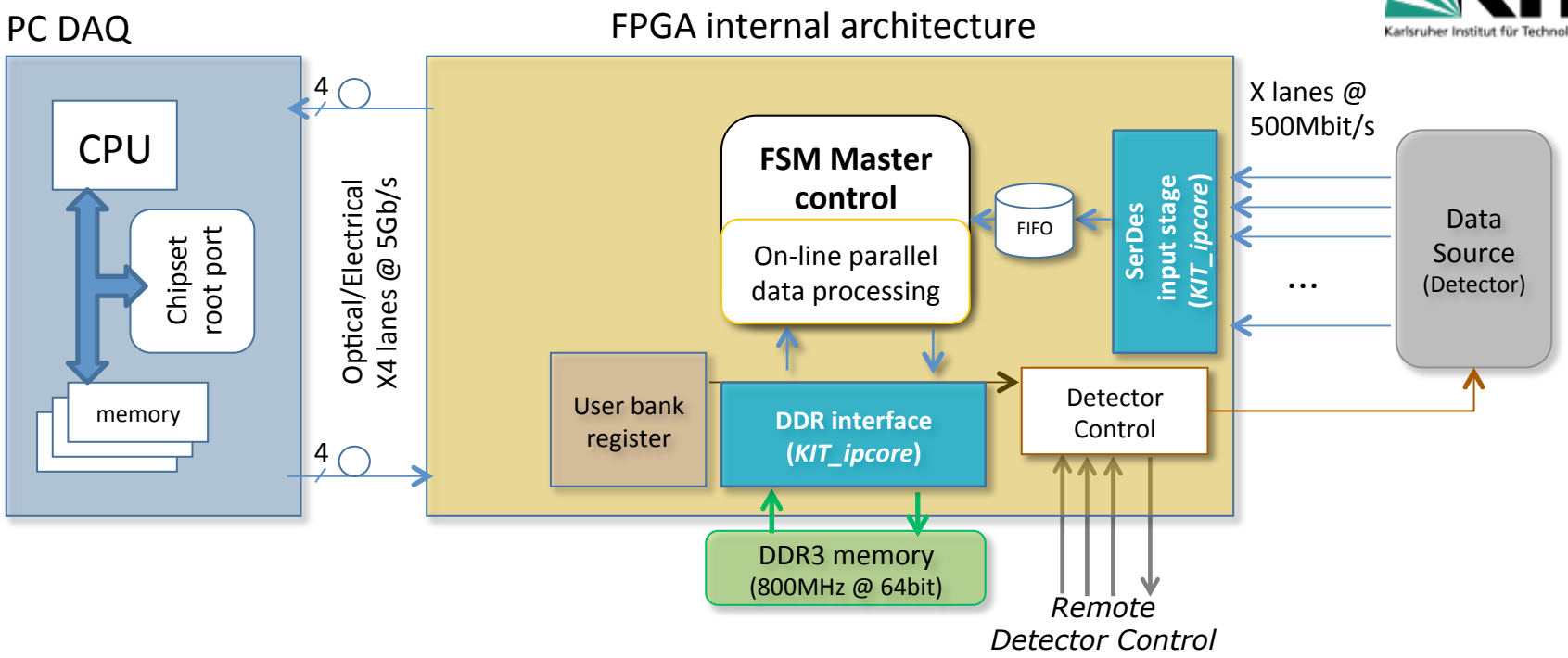
Implementation:



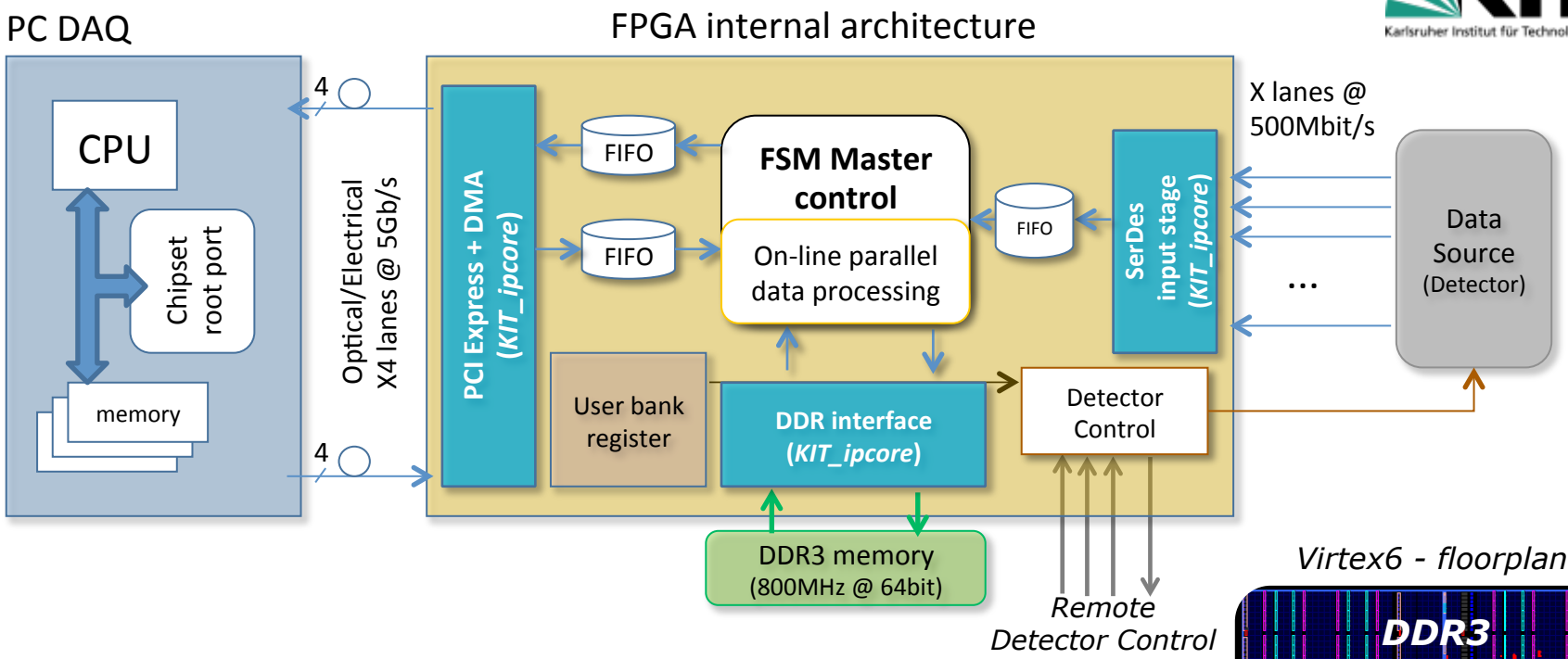
Flexible high-throughput FPGA platform



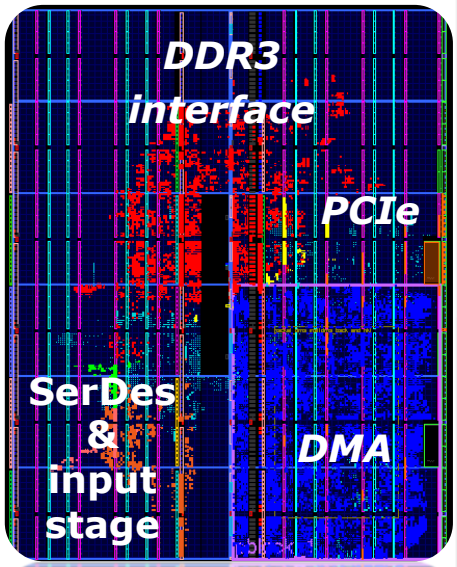
Flexible high-throughput FPGA platform



Flexible high-throughput FPGA platform

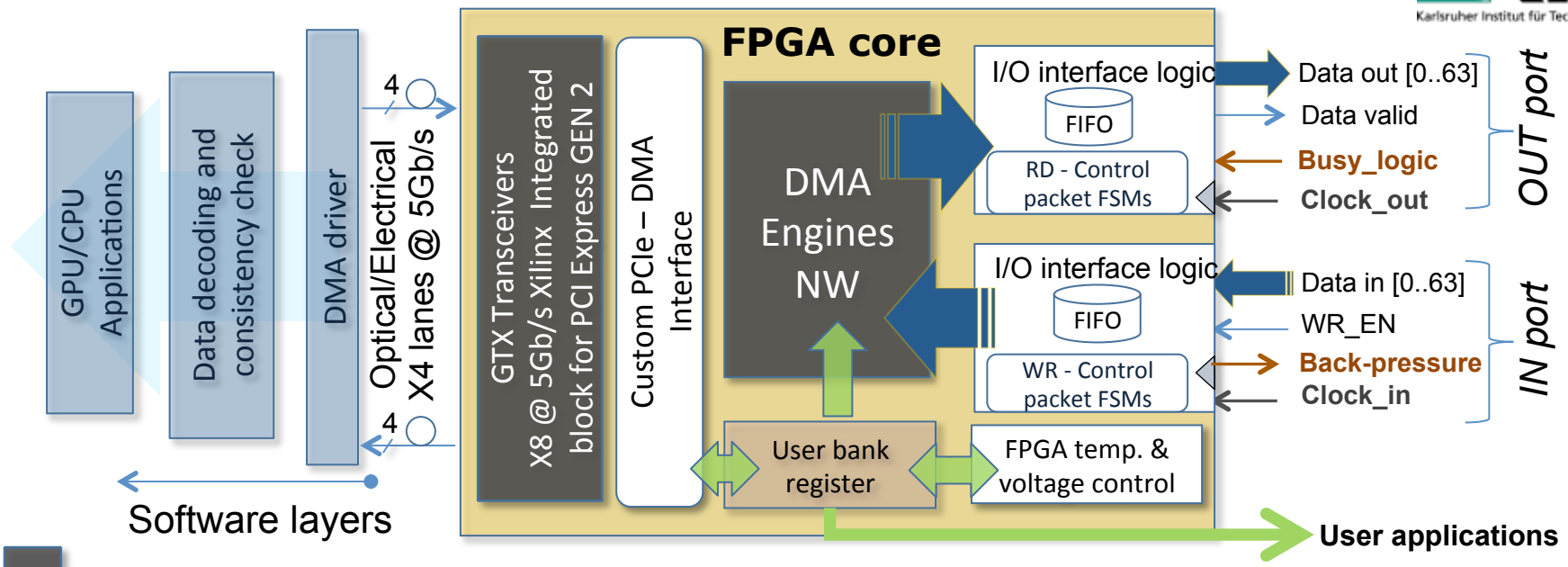


Virtex6 - floorplan



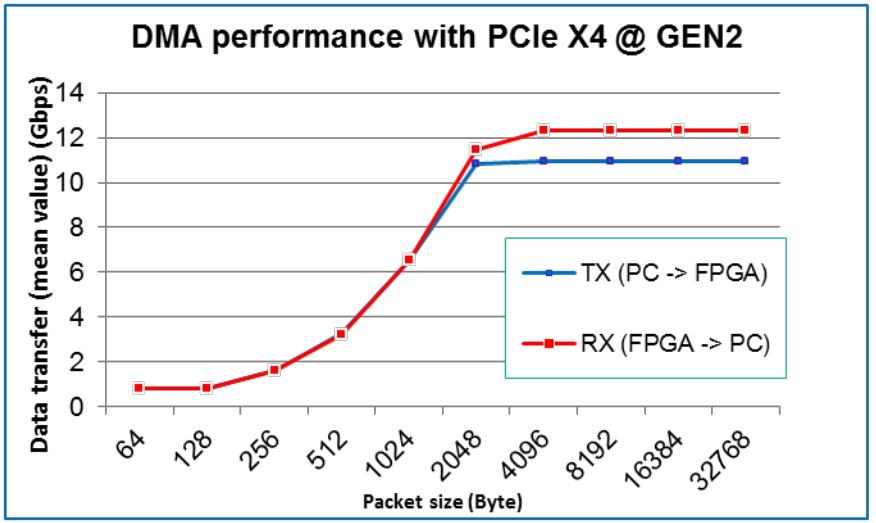
- ✓ Three logic cores have been developed for a flexible high-throughput platform
  - ✓ PCIe-Bus Master DMA readout architecture
  - ✓ Multi-port high speed DDR3 interface
  - ✓ Configurable 2..16 bits “SerDes” (Serializers /Deserializers) architecture
- ✓ PCI Express/DMA Linux 32-64 bits driver with ring buffer data management
- ✓ Integration in the parallel GPU/CPU computing framework

PCIe-Bus Master DMA readout architecture



Xilinx / North-West IP-core

- ✓ Bus Master DMA operating with 4lanes PCIe @ Gen2 (250MHz)
- ✓ Two individual engines for write/read from FPGA (User logic) to PC centre memory
- ✓ IN and OUT FIFO-like interface (for User logic)
- ✓ FIFO used to decouple the time domain between DMA and User custom logic



Preliminary, PCIe-Bus Master DMA new architecture

Disadvantage of IP-cores from external vendors, are:

1) expensive (35k€ for North-West DMA and 10-60k€ for EZDMA/QuickPCIe-IP by PLDA)
2) for unique FPGA family (Virtex 6, speed grade -2)

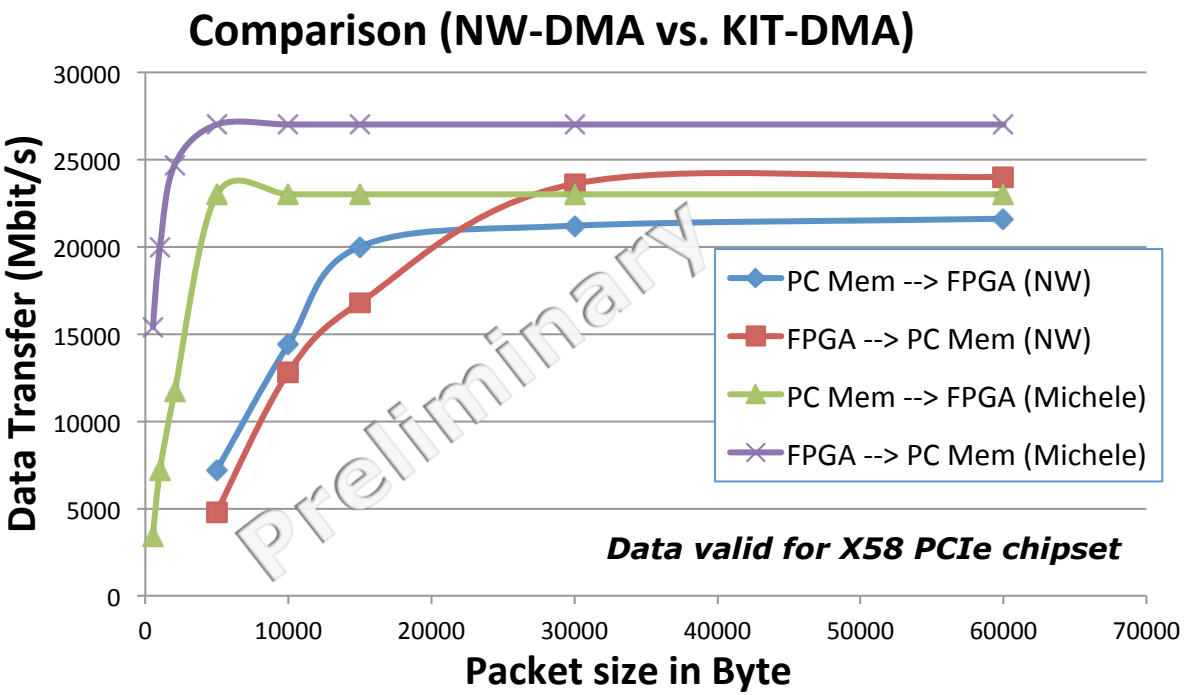
Preliminary, PCIe-Bus Master DMA new architecture

Disadvantage of IP-cores from external vendors, are:

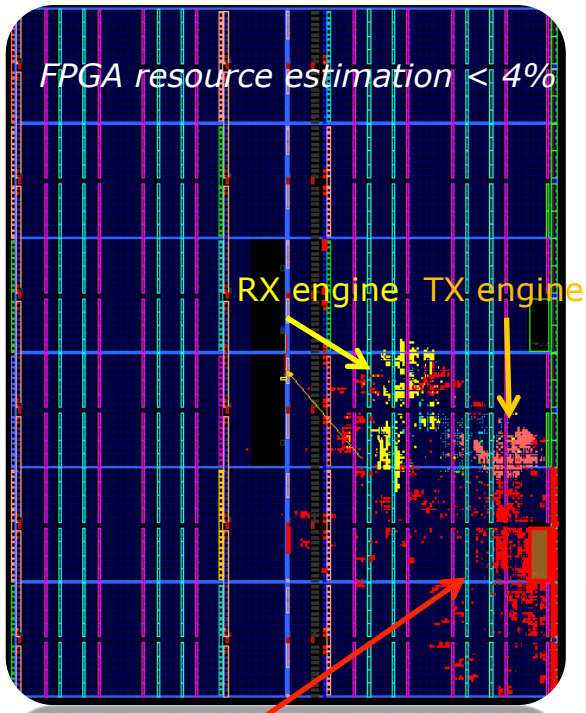
1) expensive (35k€ for North-West DMA and 10-60k€ for EZDMA/QuickPCIe-IP by PLDA)
2) for unique FPGA family (Virtex 6, speed grade -2)

New, KIT-IPE Bus Master DMA engines operating with x8 lanes PCI Express @ GEN 2

IN/OUT data at 128 bit @ 250MHz → internal bandwidth of 32 Gb/s in Read/Write



Virtex6 - floorplan



FPGA ring buffer management → on-going

Software 64bit@linux driver → under optimization (~ 32Gb/s)

Two-ports DDR3 memory interface architecture

Why a two-ports DDR3 memory controller .. ?

The Xilinx Multi-port Memory Controller (IP-Core) is limited in the maximum data throughput (less than 2GB/s for each port) & complex user interface.

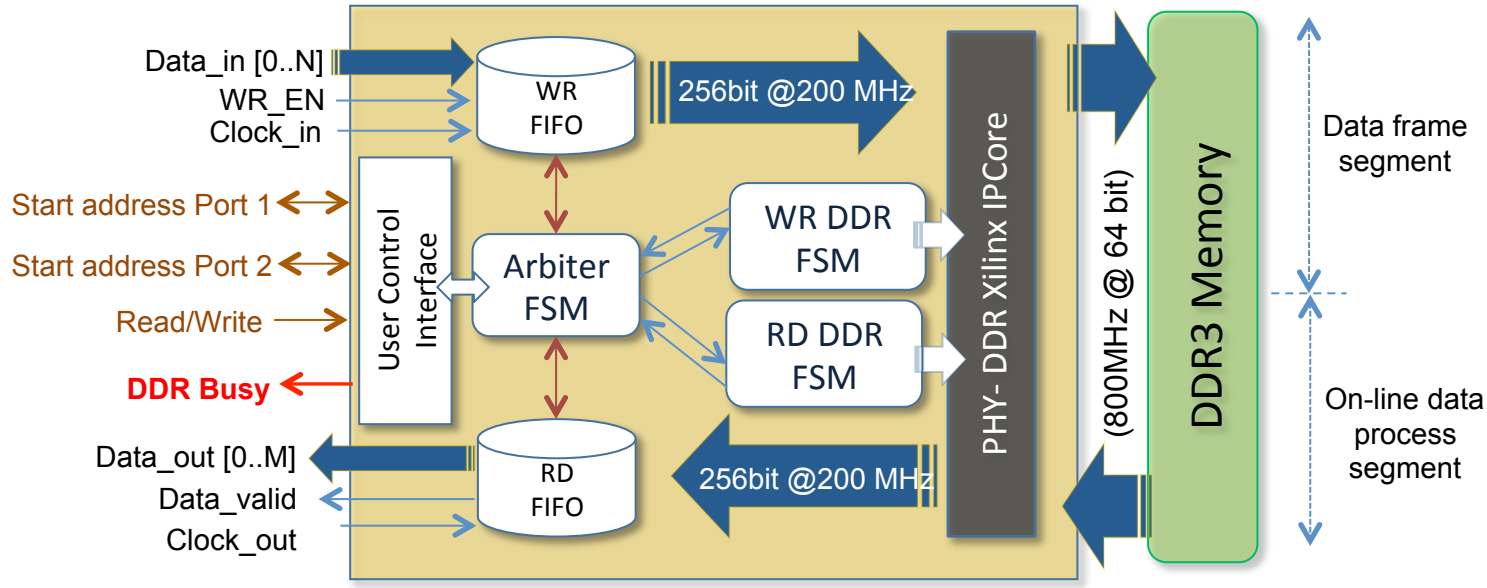
Ref. LogiCORE IP Multi-Port Memory Controller (MPMC) (v6.03.a), DS643 March 1, 2011

Two-ports DDR3 memory interface architecture

Why a two-ports DDR3 memory controller .. ?

The Xilinx Multi-port Memory Controller (IP-Core) is limited in the maximum data throughput (less than 2GB/s for each port) & complex user interface.

Ref. LogiCORE IP Multi-Port Memory Controller (MPMC) (v6.03.a), DS643 March 1, 2011



- ✓ Bandwidth 51Gb/s, limited by FPGA speed grade (Virtex 6, speed grade -1)
  - ✓ Two operations are possible in same/different segmentation/s (each operation ~ 25Gb/s)
- ✓ Data interface FIFO-like, minimum control signals are required
  - ✓ FIFO used to decouple the time domain between Memory Controller and custom User logic
- ✓ Configurable user define data width N and M → 32/64/128/512 bits

A configurable “SerDes” input stage architecture

Why not a Xilinx ISERDERSE stage .. ?

- Limited parallel data width (output) not more than 10bits (for two ISEDERSE in cascade configuration) and not dynamically configurable. The FSM Alignment in not included in the Xilinx tools.
  - *Ref. Virtex-6 FPGA Select IO resources user guide. ug361 (v1.3) august 16, 2010.*

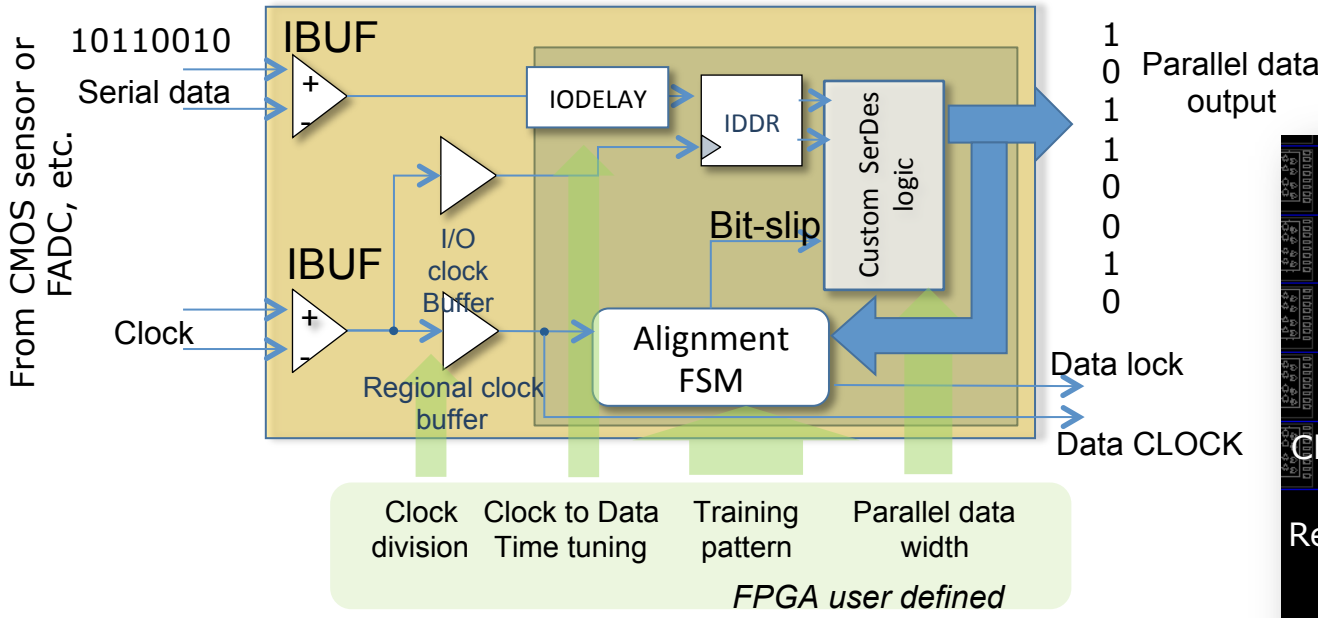
A configurable “SerDes” input stage architecture

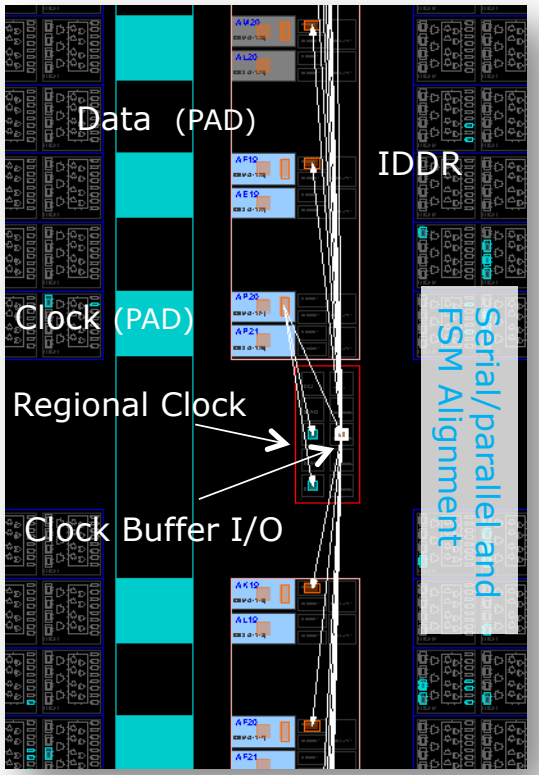
Why not a Xilinx ISERDERSE stage .. ?

Limited parallel data width (output) not more than 10bits (for two ISEDERSE in cascade configuration) and not dynamically configurable. The FSM Alignment in not included in the Xilinx tools.

➡ *Ref. Virtex-6 FPGA Select IO resources user guide. ug361 (v1.3) august 16, 2010.*

Configurable 2 to 16bit parallel data output “SerDes “ logic with MSB Alignment State Machine





✓ Individual clock-to-data time tuning by IODELAY (time step of 75psec)

✓ I/O clock buffer located in the centre of the FPGA bank

✓ Regional buffer synchronous to parallel data out

✓ “SerDes” input stage fully configurable by User

Future developments for high speed readout systems

Requirements:

- Real-time FPGA + GPU data elaboration → high data throughput (range of 64Gb/s)
- Data source and FPGA readout board located far from DAQ system
- Using commercial/well-known protocol for ease interface with commercial devices/boards

Two differents approaches are possible:

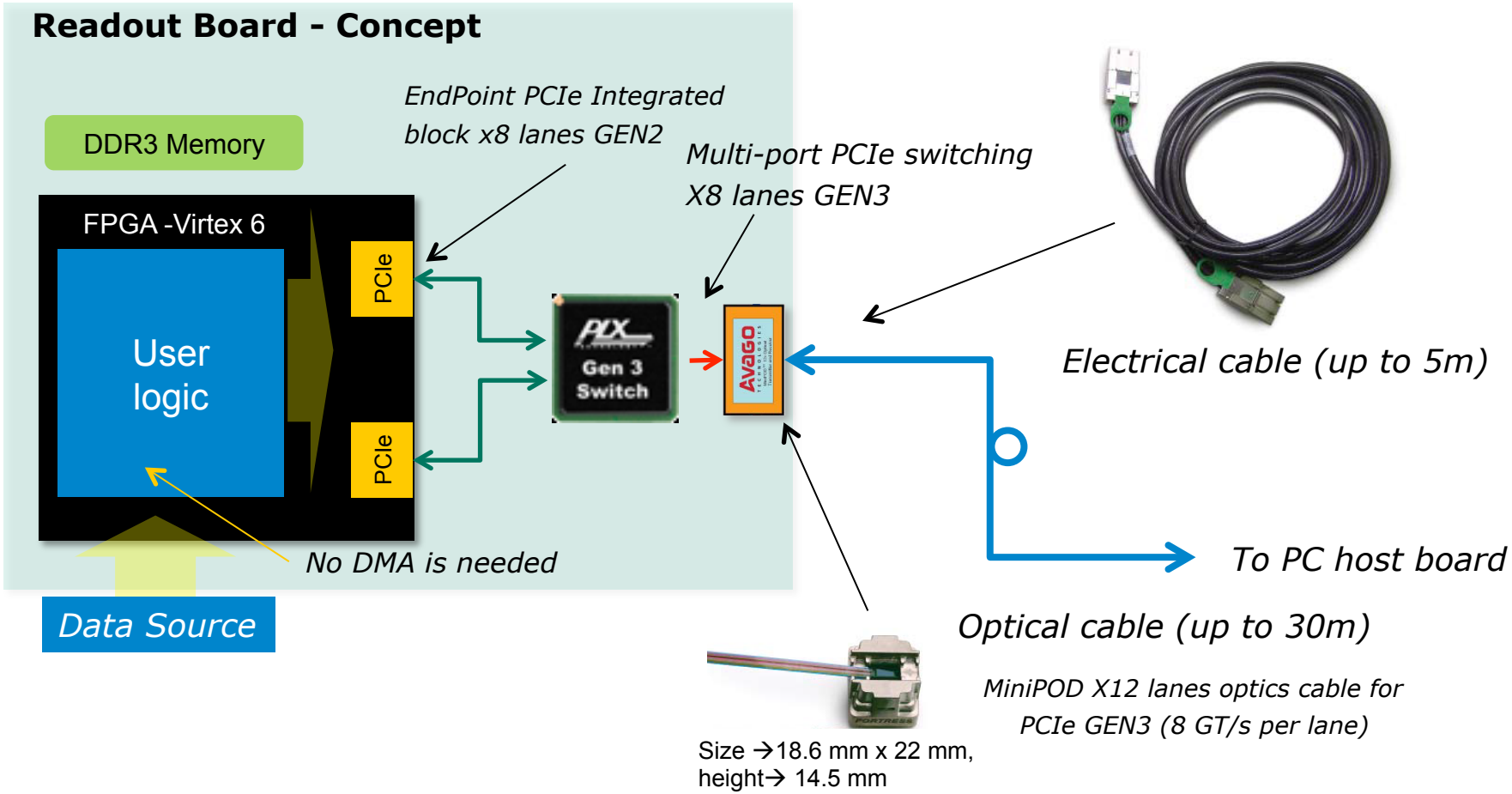
- ***Peer – to – peer (P2P) streaming data transfer***

(based on new generation of PCI express protocol)

- ***Point– to – node (net) for distributed GPU/CPU High Performance Computing (HPC) clusters***

IPE - PCI Express Readout card - Overview

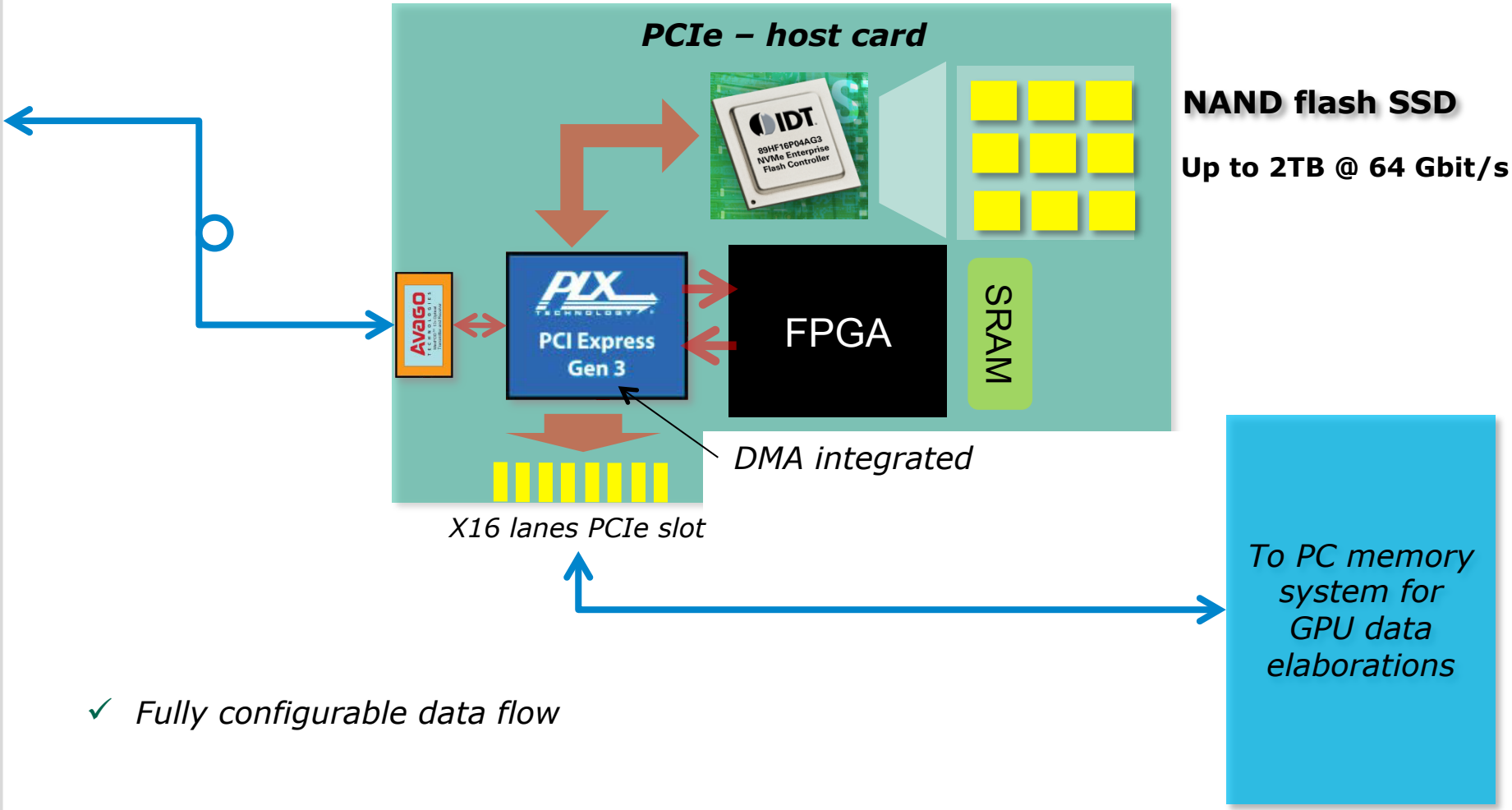
- ✓ *PCIe GEN3 optical/electrical data transmission (8 lanes x 8GT/s)*



- ✓ *64 Gb/s (W) + 64Gb/s (R) → full-duplex mode*
- ✓ *FPGA Real Time process → close to data source*

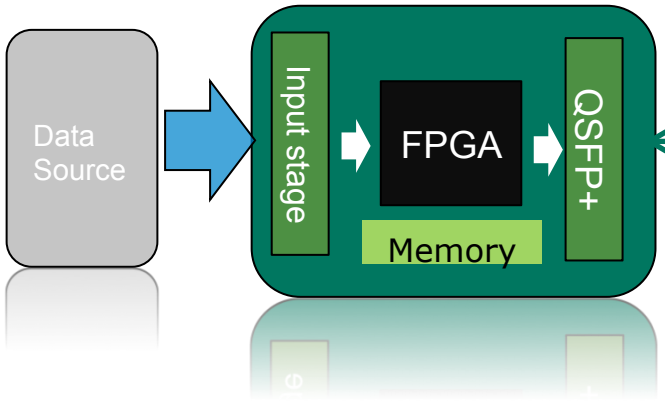
IPE - PCI Express Host card- Overview

✓ *PCIe host board with high speed data recording*



✓ *Fully configurable data flow*

High bandwidth readout system based by InfiniBand



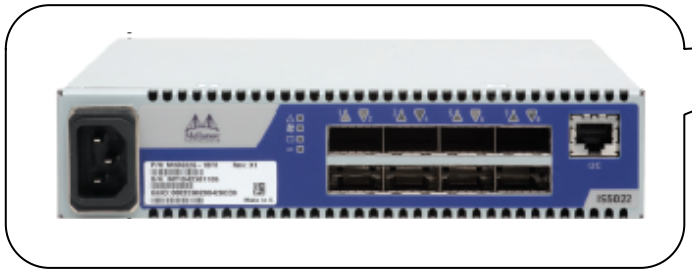
Optical or electrical data link up to 100m

40Gb/s → InfiniBand, in house

120Gb/s → InfiniBand available soon

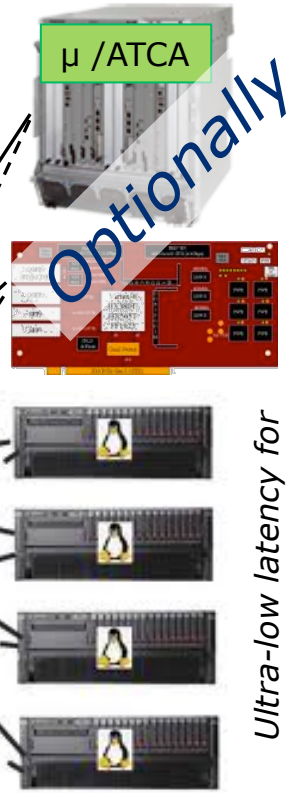
384Gb/s → in the next two years

QDR 40Gbps InfiniBand protocol



8-port switch capable of up to 640Gb/s

Infiniband Router



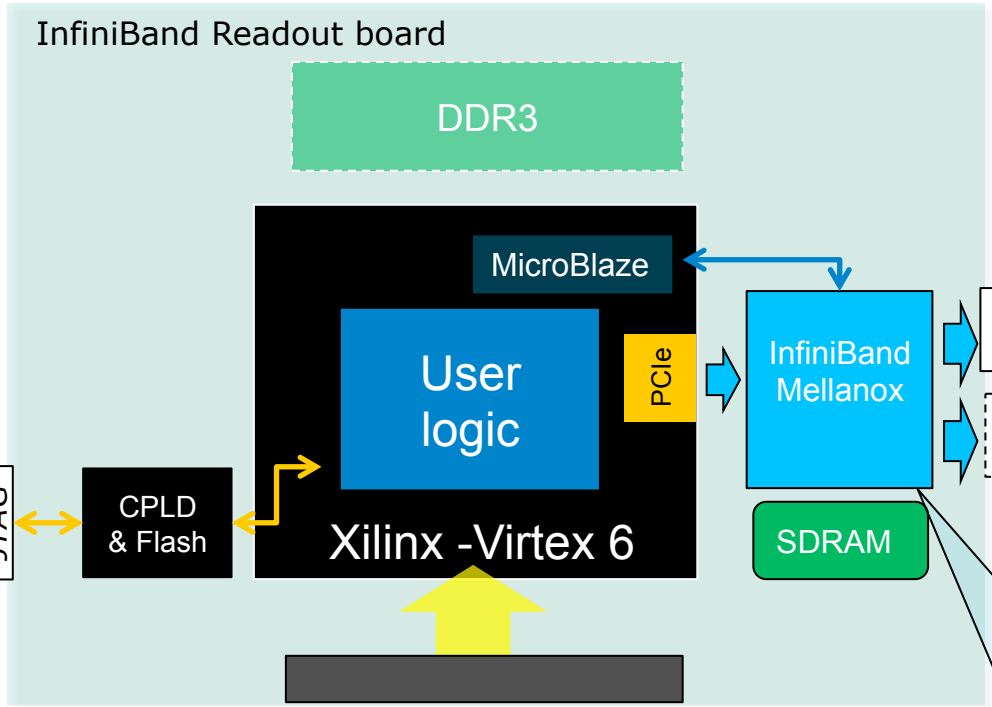
Ultra-low latency for high cluster performance

InfiniBand DAQ cluster

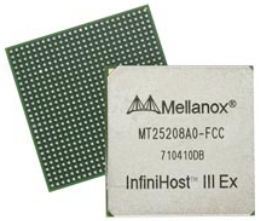
InfiniBand GPU cluster under developing in KIT-IPE by Data processing group IPE-KIT

Heterogeneous FPGA + CPU + GPU

InfiniBand readout Board - Overview

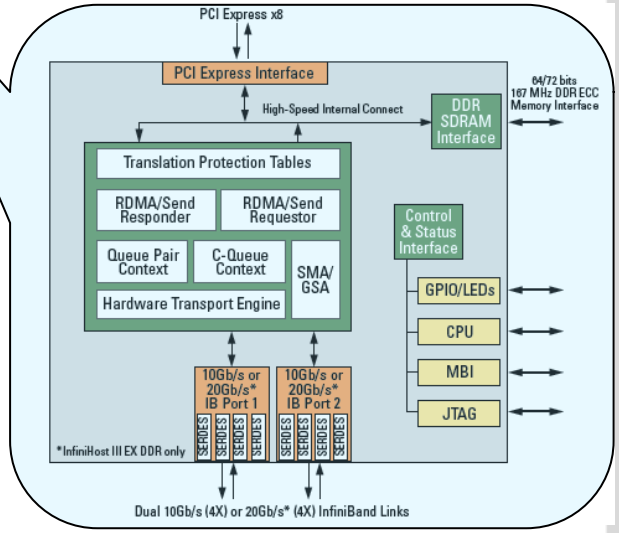


JTAG for programming and BSDL (VHDL) test



InfiniHost III Ex InfiniBand silicon device

InfiniBand switch



High Speed connectors ***From data Source***

(HPC Samtec or similar)

- IP based application layer → possible (i.e. TCP, UDP, SSH, FTP ..)
- The InfiniHost provide the **PHY**, **Link** and **Transaction** layers for InfiniBand
- **Remote DMA** for fast data transfer → intranet communication

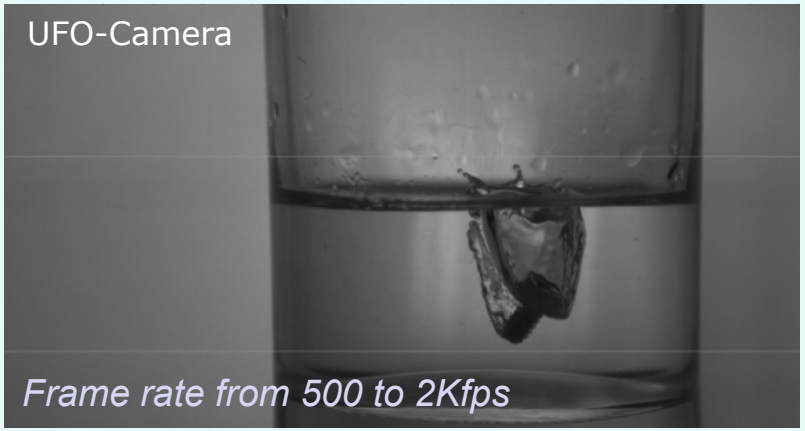
Conclusion and What's next

- *Logic cores for high data throughput platform → employed in several scientific applications:*
  - **TeraHz detector + readout system for CSR** (M.Caselle, V. Judin, A.S. Müller, M. Siegel, N. Smale, P. Thoma, M. Weber, S. Wünsch). KIT departments IPE-IMS and ANKA
  - **A X-ray camera for phase contrast tomography** (M. Caselle, A. Kopmann, Felix Beckmann (HZG), Joerg Burmester(HZG) KIT and HZG
  - **A X-ray camera for high spatial resolution tomography** (M. Caselle, M. Balzer, A. Kopmann, V. E. Asadchikova) Shubnikov Institute of Crystallography, Russian Academy of Sciences, Moscow, Russia
  - **A readout electronics for Ultrafast electron beam X-ray tomography system "ROFEX" in HZDR** (*proposal under discussion*)
- *New KIT-DMA (32Gb/s) engines → developed and tested*
- *Driver 64bit@Linux → under optimization*

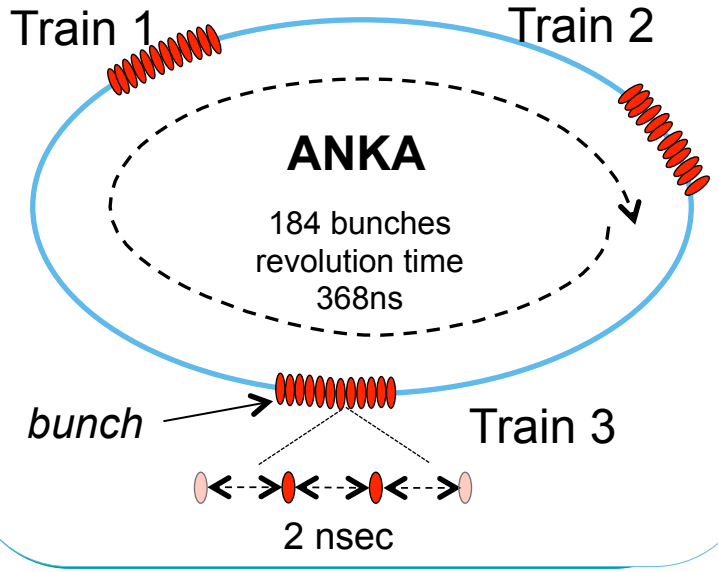
What's next

- *Design & production of readout board based by:*
  - *PCIe GEN3 optical communication*
  - *InfiniBand protocol*
- *Integration in the GPU/CPU compute infrastructure*

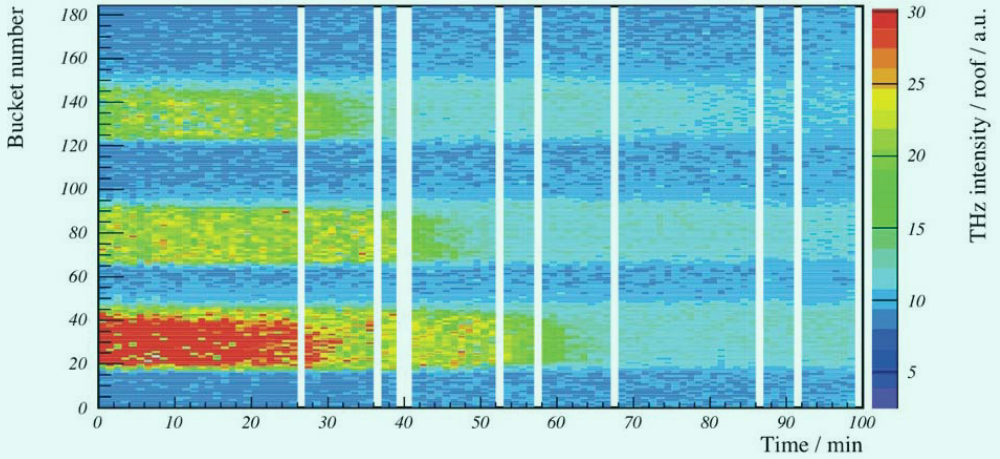
Bandwidth: 8 Gb/s. Future upgrade: 50Gb/s



Thank you for your attention



Bandwidth: 6Gb/s: Future upgrade to 24Gb/s

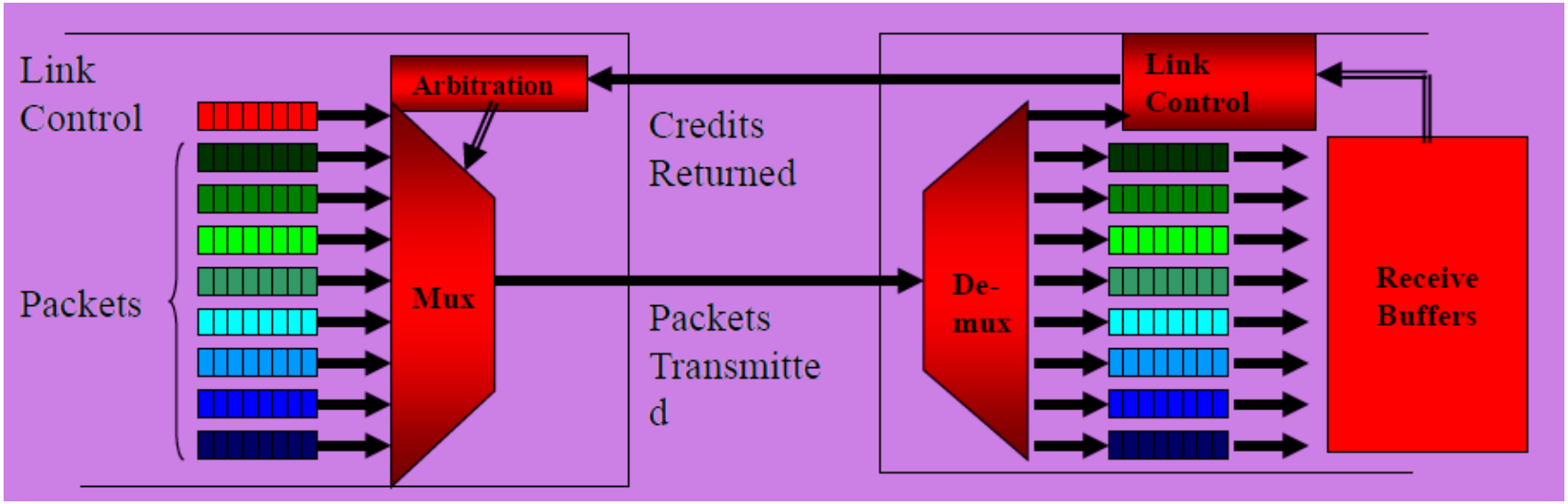


Recording & analysis of time evolution of each bunch in a multi-bunches accelerator filling-scheme

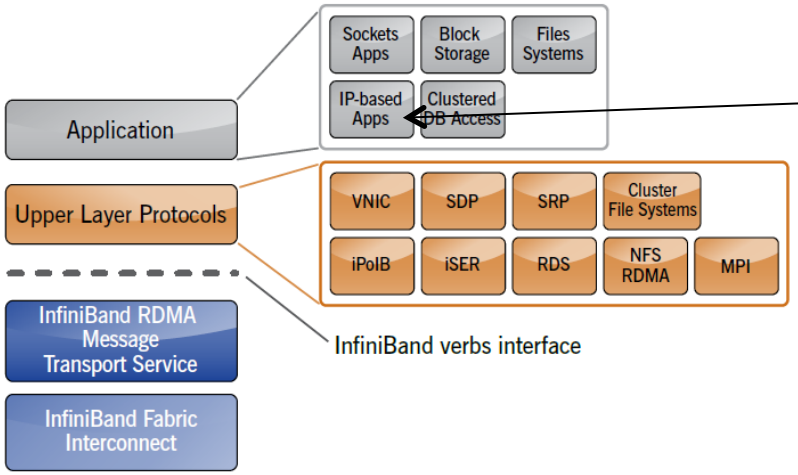
Backup slides ..

InfiniBand: Link layer Flow Control

Credit-based link-level flow control
• Link Flow control assures NO packet loss within fabric even in the presence of congestion
• Link Receivers grant packet receive buffer space credits per Virtual Lane
• Flow control credits are issued in 64 byte units

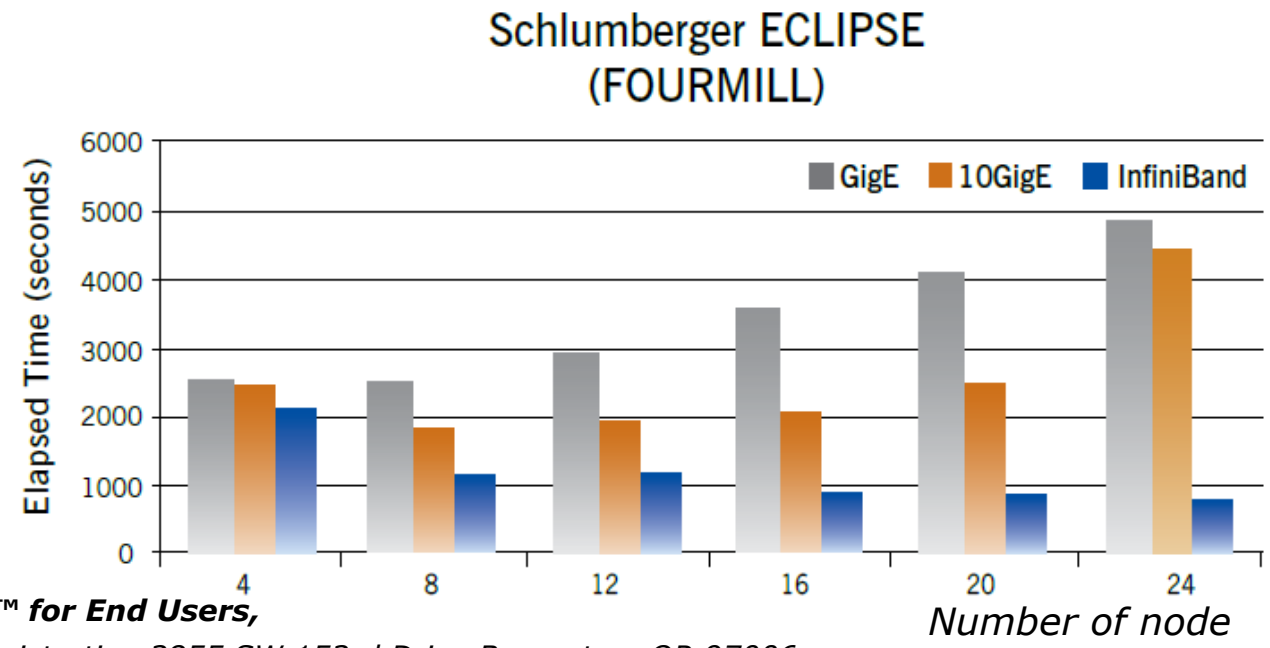


InfiniBand: application layers and latency



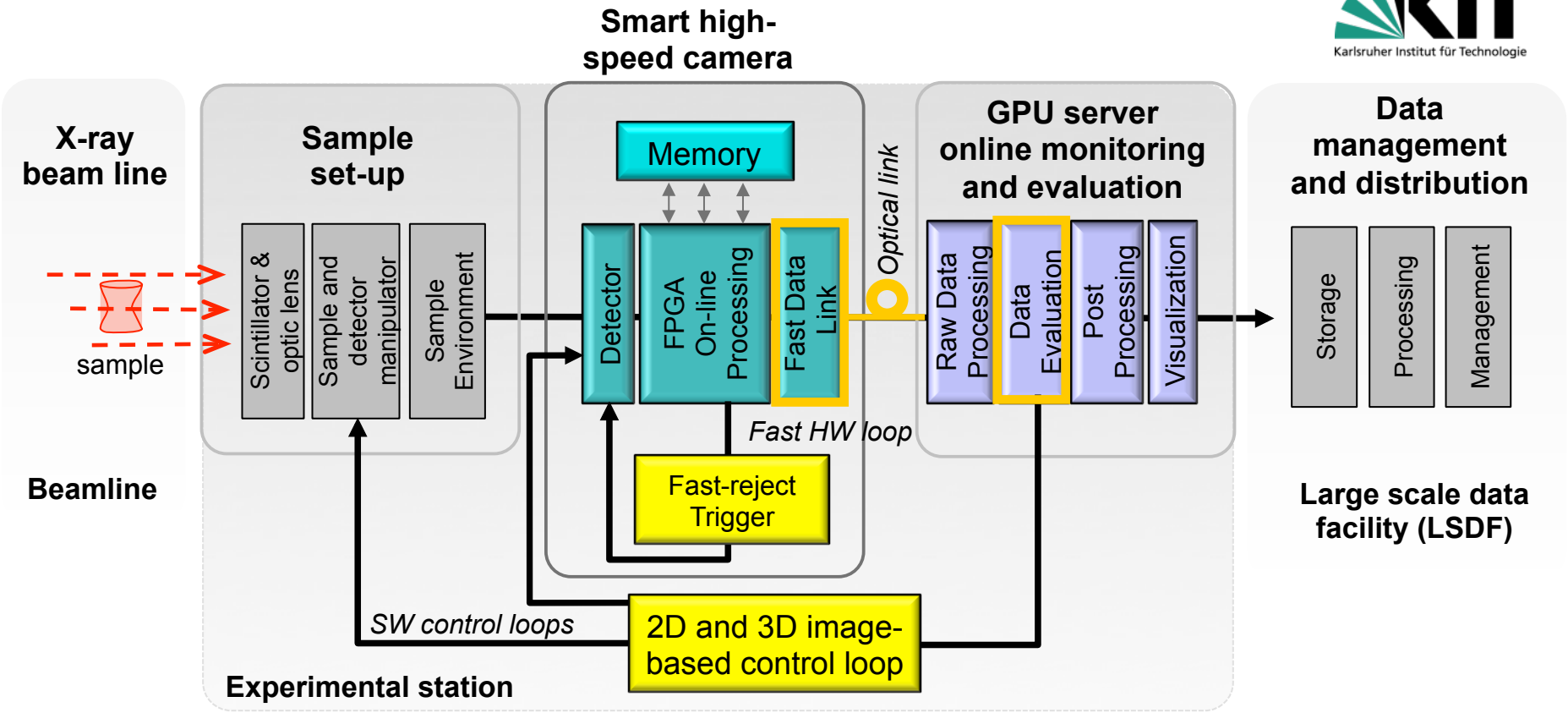
UDP or TCP, FTP, ssh …

Designing with InfiniBand



Ref: ***Introduction to InfiniBand™ for End Users,***
InfiniBand Trade Association Administration 3855 SW 153rd Drive Beaverton, OR 97006

UFO architecture - overview

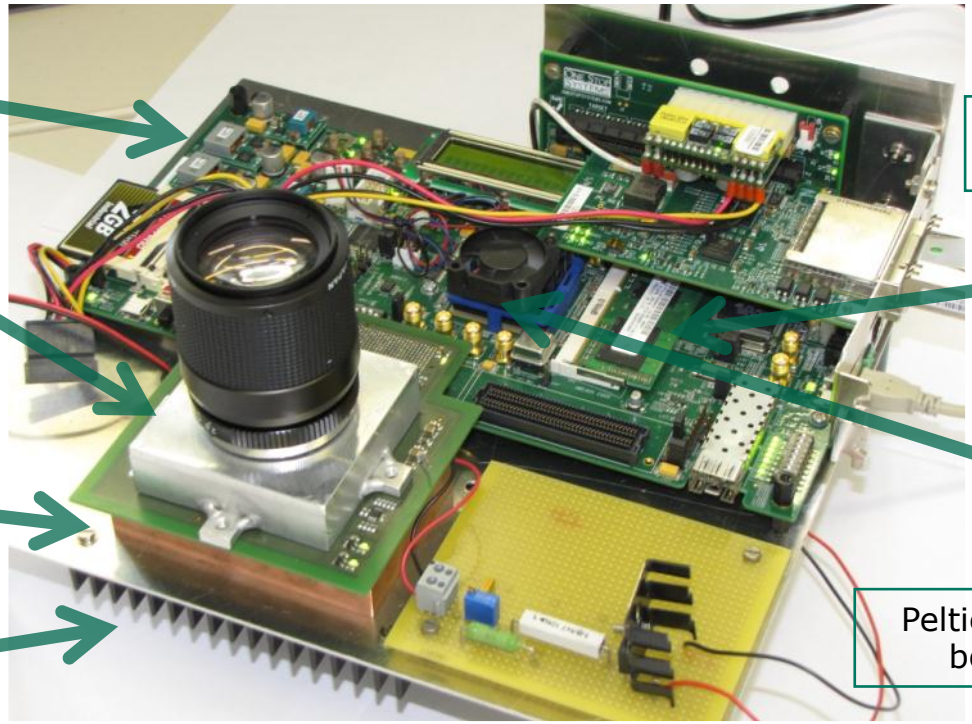


- ✓ High speed & bandwidth, full programmable camera (continuous data acquisition at full speed)
- ✓ Optimized image processing algorithm using GPU computing
- ✓ Fast HW loop: On-line image-based self-event trigger architecture (Fast reject)
- ✓ SW control loops: based on 2D and 3D data evaluation:

  2D data → camera calibration, autofocus, self-alignment & etc..

  3D data reconstructed → like optical flow, etc

UFO Camera - overview



The main features already implemented and tested, include:

✓**Fully configurable camera** → adjustable image exposure time and dynamic range, analog and digital pixel features as pixel threshold, mask, analog gain, etc.

✓**Continuous data acquisition at full speed**

✓**On-line image-based self-event trigger architecture (Fast reject)**

✓**Region-of-interest readout strategy using self-event trigger information**

✓**Easily extendable to any available CMOS image sensor**

Readout electronics for Coherent Synchrotron Radiation

- Measure of the *peak amplitude* of each bunch *(resolution few mV)*
- Measure of the *pulse width of* each bunch *(resolution few psec)*
- Measure of the relative *time jitter* between electron bunches *(res. few psec)*

Strategy:

Digitalize each pulse with 4 samples + pulse reconstruction & Constant Fraction Discriminator (CFD) for precise pulse timestamp.