New Site: Paderborn Center for Parallel Computing (PC²)

Christian Plessl

Paderborn University, Germany
Paderborn Center for Parallel Computing

Cray User Group Conference 2019 – Montreal – 11 May 2019

Paderborn Center for Parallel Computing

Paderborn: Germany's Chippewa Falls

- Heinz Nixdorf (1925–86)
  - founder of Nixdorf Computer
  - businessman, sportsman, donor
- Major player in business computing
  - headquarter in Paderborn
  - > 30'000 employees worldwide
  - > 20 countries
  - > 5 B DM revenue
- Our local Seymour Cray
  - or Steve Jobs
- Remains of Nixdorf Computer seeded IT industry in our area

NIXDORF COMPUTER

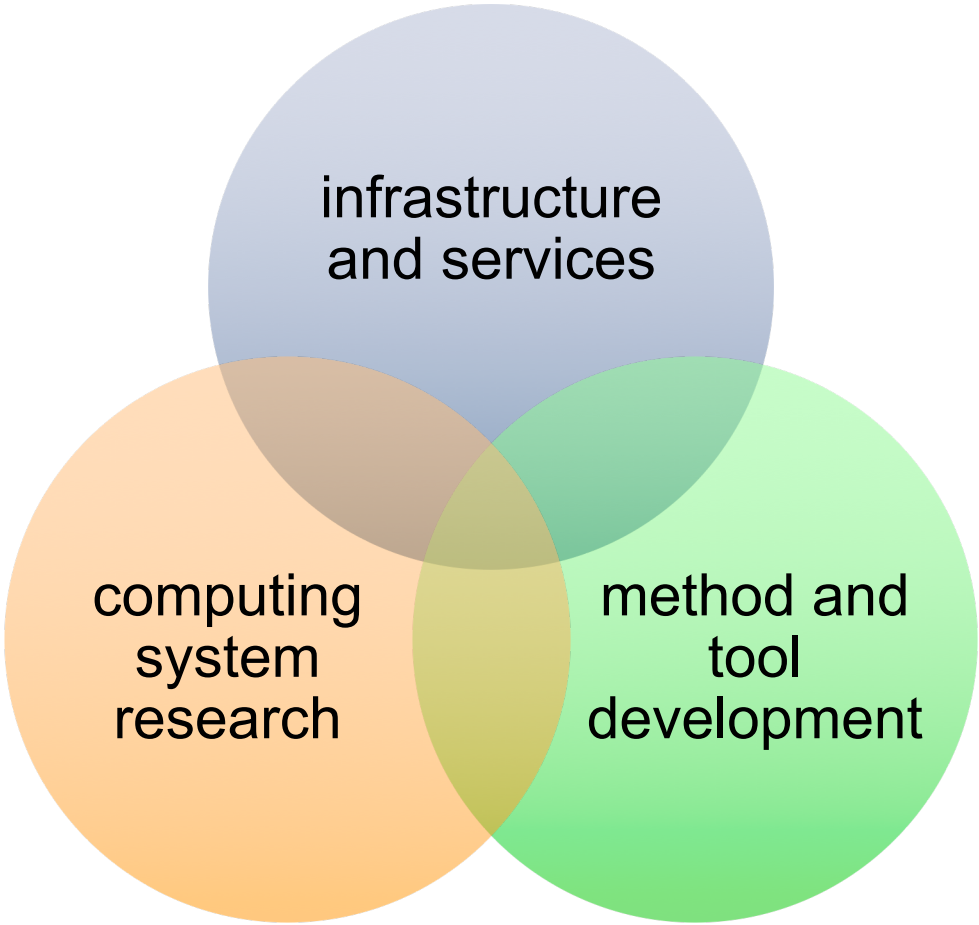
Heinz Nixdorf

Paderborn University

- Founded in 1972
  - 20’000 students
  - 260 faculty members, ~1200 PhD students and postdocs
- Departments
  - Humanities
  - Economics
  - Natural Science
  - Mechanical Engineering
  - Math., CompSci and EE
- Research focus
  - optoelectronics and photonics
  - material science
  - business informatics
  - intelligent technical systems

Paderborn Center for Parallel Computing

- Scientific institute of Paderborn University
  - established in 1992
  - roots in theoretical computer science
- Service provider and research institution
  - provision HPC infrastructure and services for computational sciences
  - develop new methods and tools for HPC simulation in cooperation with domain scientists
  - perform computing systems research for energy-efficient HPC with emphasis on heterogeneous and accelerated computing with FPGAs and manycores
- Long track record in exploring emerging and off the beaten path technologies



PC² History and Innovations

	HPC System	Properties / Innnovation	Research Topics
1991	Parsytec SC320	• system design in Germany (Aachen), Transputer processors developed in UK • largest parallel computer with freely programmable network • parallel programming with OCCAM	• graph partitioning • optimal embedding of graphs of degree 4
1992	Parsytec GCel #262 of Top500	• 1024 processors, largest parallel computer with Transputers in Europe • Solaris Unix and parallel programming environment PARIX • scalable 2D communication network	• general graph embedding, in particular in 2D meshes
1995	Parsytec GC/PP #118 of Top500	• transition to standard technologies (CPU, compiler, operating systems) • innovation through heterogeneous nodes: PowerPC (computing) + Transputer (communication)	• load balancing • HIBRIC-MEM streaming cache
1999	Fujitsu-Siemens hpcLine #351 of Top500	• use of Intel x86 and Solaris/Linux as standard components • innovation in networking: Scalable Coherent Interface (SCI), European development • first large scale SCI-cluster worldwide	• message passing • fault tolerance • start of HPC usage beyond computer science

PC² History and Innovations (2)

	HPC System	Properties / Innovation	Research Topics
2003	Megware FPGA Cluster	• combination of standard CPU/OS technologies with application-specific accelerators (FPGA) • used for powering one of the best Chess computers of the day • Myrinet network with low latency	• distributed game tree search • custom computing
2003	HP PLING	• step towards 64bit CPU technology (Intel Itanium) • first 64bit Linux Cluster with InfiniBand in Europe	• software support for InfiniBand in 64bit Linux
2004	Fujitsu/ICT Arminius #213 of Top500	• Direct water cooling for CPUs • integration of GPU nodes in cluster • PCIe / InfiniBand • x86-64, Linux	• 3D visualization of simulations • immersive control
2007	Fujitsu Siemens BiSGrid	• nodes with high compute power, 4 sockets with AMD processors	• grid computing • workflow management
2013	Clustervision OCuLUS #173 of Top500	• heterogeneous nodes with GPUs, Intel Xeon Phi	• virtualization • multi/many Core
2018	Cray CS500 Noctua	• 16 nodes with FPGA accelerators and dedicated interconnect between FPGAs	• HPC acceleration with FPGAs

Noctua HPC Cluster

- Cray CS500 cluster system
- 256 CPU nodes
  - 256 nodes with 2 x 20-core Xeon Skylake Gold 6148
  - 192 GiB RAM / node
  - 100 Gbit/s Omni-Path interconnect
  - 700 TB Lustre parallel file system
- 16 FPGA nodes
  - same configuration as CPU nodes
  - each with 2 x Nallatech 520N FPGA boards
  - Stratix 10 GX2800, 32GB DDR4, 4 memory channels
  - PCIe 3.0 x16
  - 4 QSFP+ ports
- Operational since Sept 2018

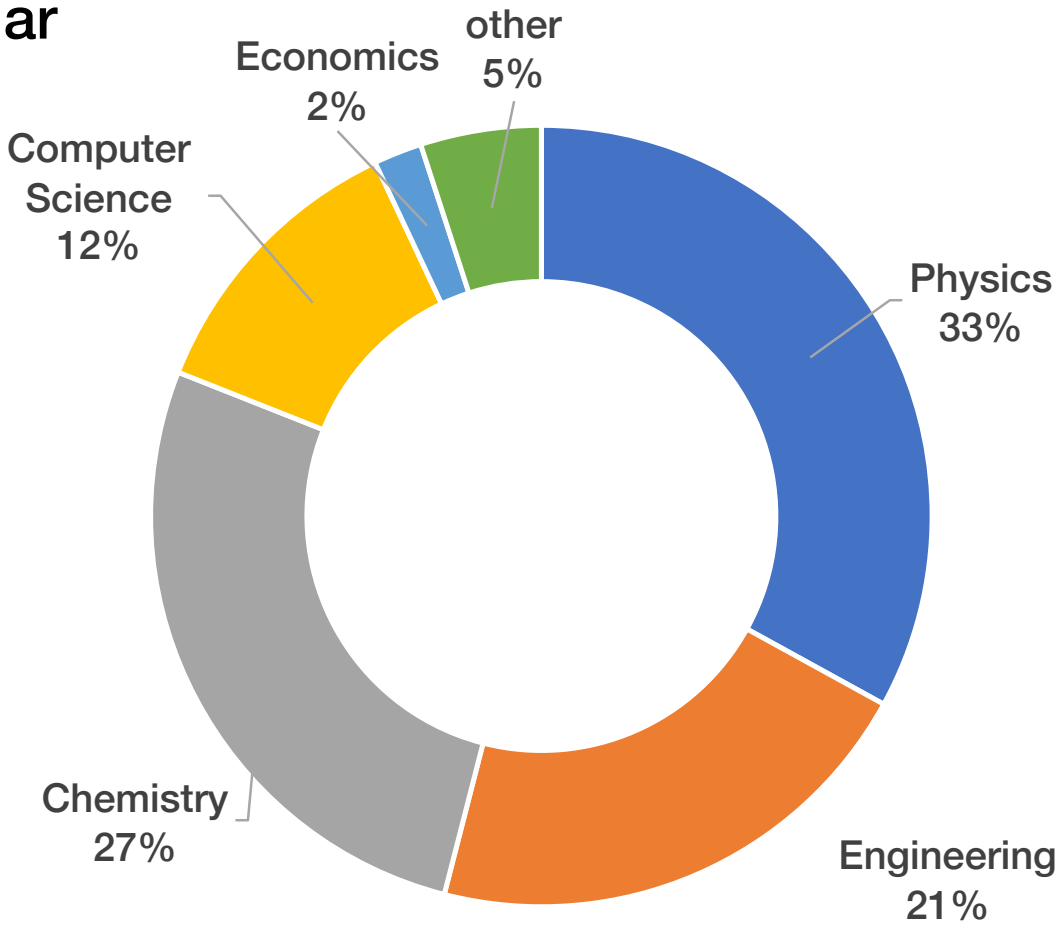
at time of installation largest academic installation of FPGAs in HPC cluster

Noctua Phase 2

- New data center optimized for HPC
  - best of class energy efficiency and flexibility
  - warm water cooling (free cooling)
  - modular design to support concurrent operation and upgrades of multiple generation HPC systems
  - extensibility (power, cooling, office space)
- Specificaitons
  - white space: 300m^2
  - other technical facilities: 1100m^2
  - initial power / cooling capacity: 1.2-2 MW
  - office space for 25+ persons + seminar rooms, labs, ...

Workloads and Users

- Solid state physics and chemistry (in particular DFT codes)
  - CP2K, VASP, QuantumEspresso, Turbomole
- Optoelectronics and photonics
  - CST microwave studio
  - in-house codes
- Engineering
  - Fluent, OpenFOAM
- Computer science
- Statistics
  - 70 active projects
  - 400 active users



data: PC² 2016 (Oculus cluster)

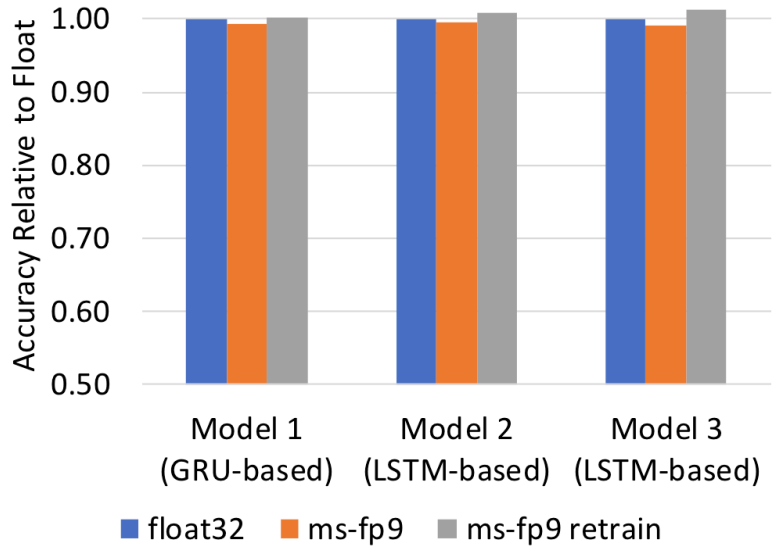
Why FPGAs?

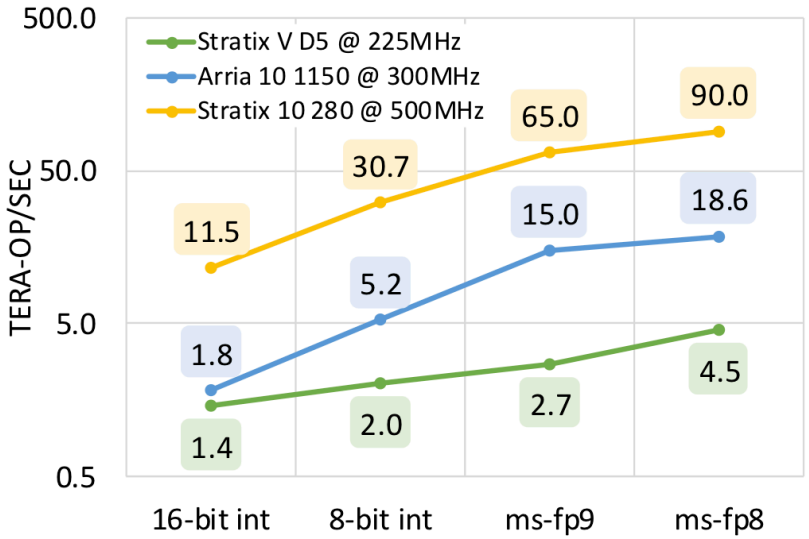
- Status quo
  - end of Dennard scaling and Moore's law is imminent
  - Post-CMOS technologies will not be ready for many years
  - demand for HPC and general data center applications growing rapidly
  - CPUs are fundamentally inefficient due to generality (instructions, caches, OoO)

- What can we do
  - scale out by using ever larger and more costly systems
  - specialization of architectures
  - develop new methods that do not require exact computation and/or high precision
  - method/architecture codesign

sweet spot for FPGAs

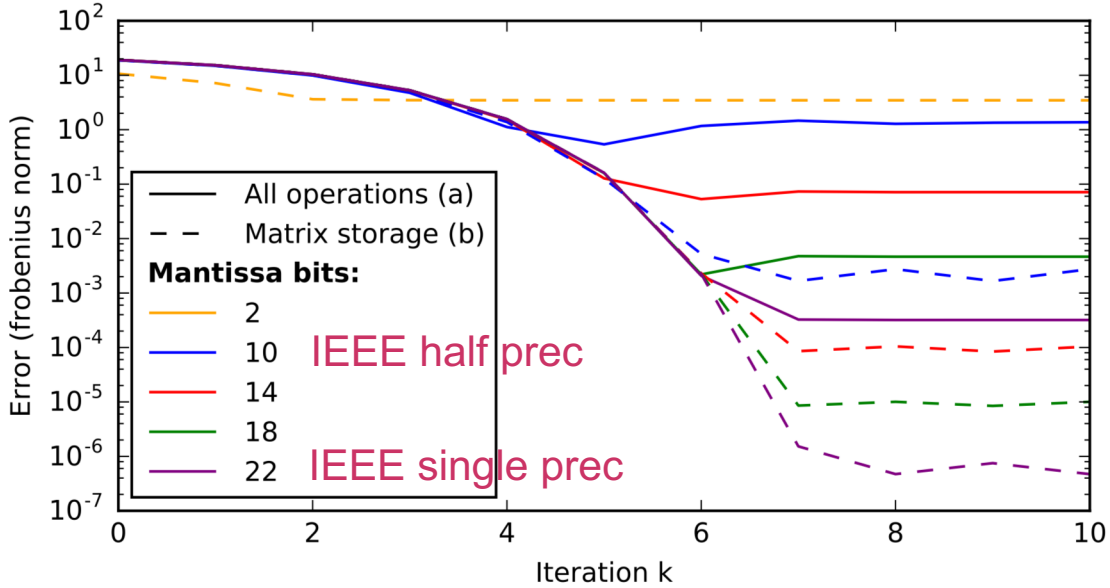
FPGAs are currently the only viable technology for application-specific computing (when ASICs don't pay off)





Approximate Computing

- Exploit performance/energy vs. accuracy trade-offs in computing architectures
- Suitable if:
  - application is inherently tolerant to inaccuracies
  - inaccuracies can be compensated, e.g. iterative methods
- Target applications
  - molecular dynamics, quantum chemistry
- Architectures
  - CPU/GPU: reduce memory bandwidth
  - FPGA: trade area saved for more computing units



iterative computation of $A^{-1/p}$: approximation error for custom floating-point formats

- **A general algorithm to calculate the inverse principal p-th root of symmetric positive definite matrices**. Communications in Computational Physics, 25(2):564--585, Mar. 2019.
- **A massively parallel algorithm for the approximate calculation of inverse p-th roots of large sparse matrices**. In Proc. Platform for Advanced Scientific Computing Conference (PASC). ACM, 2018.
- **Accurate Sampling with Noisy Forces from Approximate Computing**. In preparation.

Capabilities of Todays Top-Of-The-Line FPGAs

Example: Intel Stratix 10 GX2800 (used in Noctua)

- > 900,000 configurable logic blocks
  - up to 4 Boolean functions of 8 inputs
- 5760 hardened arithmetic units (DSP)
  - fixed point and IEEE 754 SP floating-point
- > 11,000 independent SRAM blocks
  - width/depth/ports highly configurable
- integrated DDR4-2666 memory controllers
- 96 serial transceivers, up to 28.3 Gbps
- typically about 300-600MHz
- power consumption 50-225W

100 TERRA-OPS

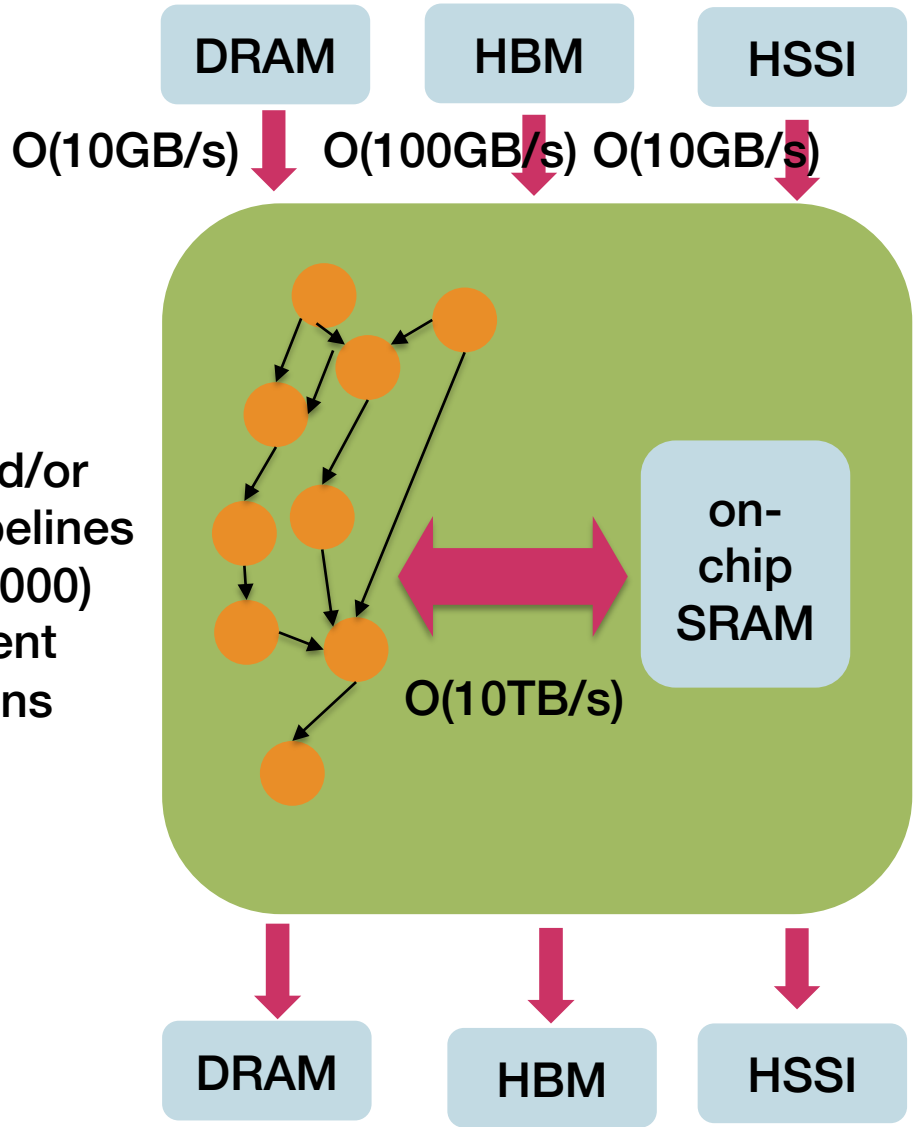
10 single-precision TFLOPS

20 TB/s internal SRAM bandwidth (full duplex)

300 TB/s communication bandwidth (full duplex)

up to 80 GFLOPS/W

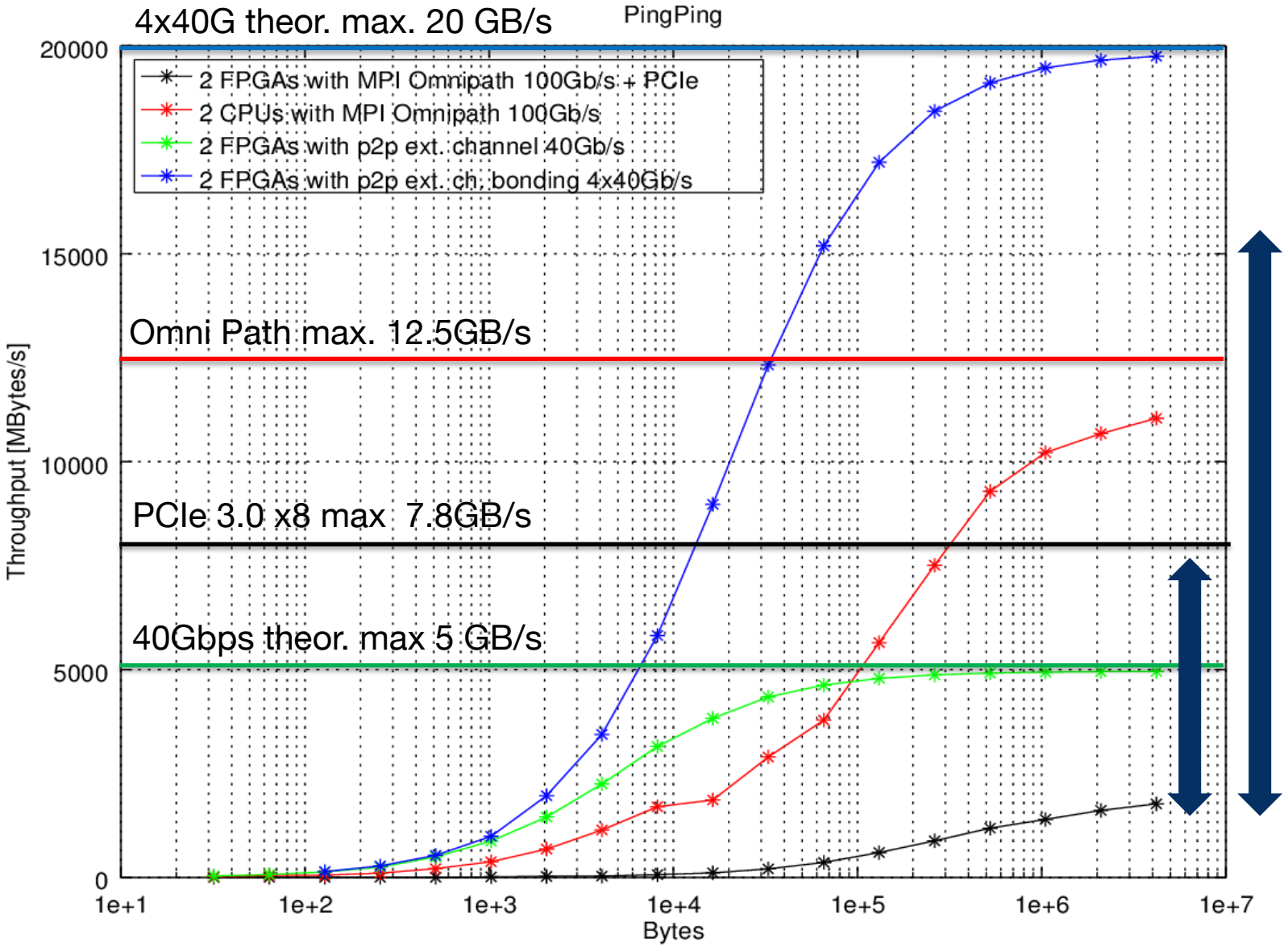
How Can FPGAs Compete with CPUs or GPUs



- Compute-bound applications
  - customization of operations and data formats
  - new methods considering FPGA architecture
- Memory-bound applications
  - unrolling and data flow computing with very deep pipelines
  - application-specific, distributed memory architectures
- Latency-bound applications
  - speculative or redundant execution
- I/O-bound applications
  - on-board network interfaces
  - direct FPGA-to-FPGA communication

HBM: high-bandwidth memory
HSSI: high-speed serial interface, e.g. 100G Ethernet

Direct Integration of FPGAs in Interconnect



- peer-to-peer optical links between FPGAs
  - high throughput
  - low latency (<600ns)
  - even better for streaming
- building application specific networks
  - circuit switched (optical switch)
  - packet switched (Slingshot!)

Ideas for Collaboration within CUG

- What we can share
  - provisioning of FPGA firmware and tool versions with Slurm
  - applications and libraries with FPGA support (CP2K, DBCSR, FFT)
  - integration of optical switches as secondary networks in cluster
  - access to our FPGA partition for research and development


```
srun --partition=fpga \
     --constraint=18.0.1
```

- Possible areas of collaboration
  - integration of FPGAs as network-attached accelerators in Slingshot
  - tools for application analysis to identify suitable functions for offloading
  - numerical methods for approximate computing in linear scaling DFT and molecular dynamics

Contact

- We are looking forward working with the CUG community

Further Information / Feedback

Christian Plessl
Paderborn University
christian.plessl@uni-paderborn.de

Twitter: @plessl @pc2_upb

http://pc2.uni-paderborn.de

Paderborn Center for Parallel Computing