Seminar High-Performance Computing with FPGAs

SS 2018

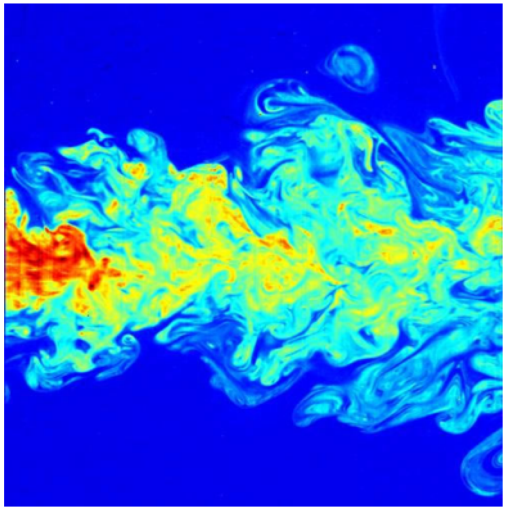
Prof. Dr. Christian Plessl
Paderborn University

Agenda

- Introduction to topic
- Discussion of seminar organization
- Selection of topics
- Agenda of future meetings

Demand for High-Performance Computing

- Numerical computer simulation is standard methods in many areas
  - Molecular dynamics, electromagnetic field propagation, fluid dynamics, ...
  - Increasing computational demand:
    - Higher spatial and temporal resolution
    - Higher fidelity models
    - Iteration for optimization



- Emerging areas with new characteristics
  - E.g. Bioinformatics, image recognition, data mining
  - Increasing demand due to rapidly growing data (Big Data)

Bilder: UCLA, eBO enterprises

Challenges and Developmetns

- HPC is table-stakes or competitive advantage in many areas of science and engineering (user perspective) ...

- ... but energy-efficiency has become a key concern of data centers (operator perspective)
  - Sizing of cooling and power supply
  - Operating cost

- Trends and roadmaps for HPC
  - Parallel computing
  - Heterogeneous computing
  - Custom computing (specialization)

- The Opportunities and Challenges of Exascale Computing. US Department of Energy, 2010.
- Challenges facing HPC and the associated R&D priorities: a roadmap for HPC research in Europe. PlanetHPC, 2013.

Custom Computing

general-purpose CPU

no customization

Application

specialized CPU

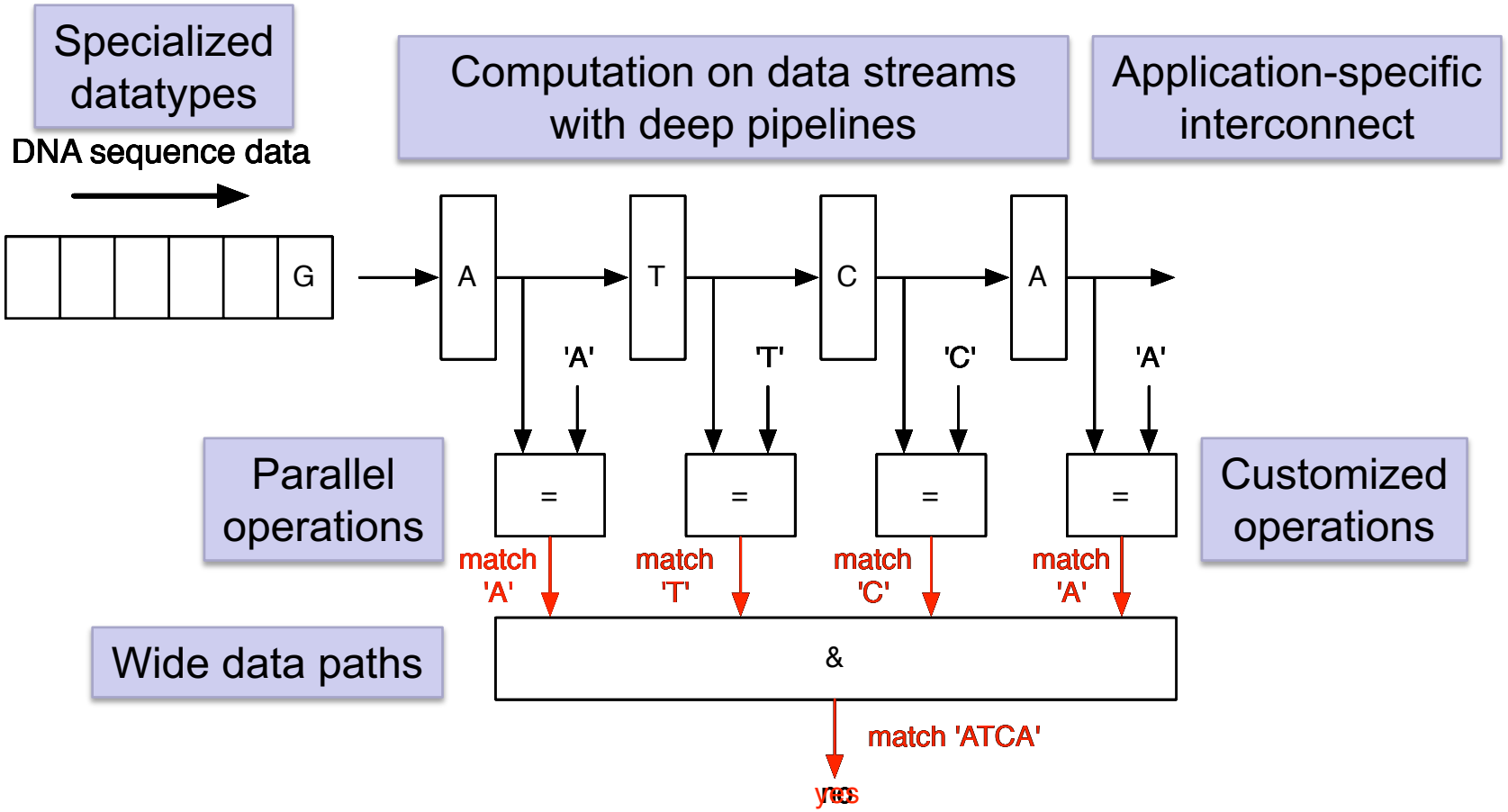
Arithm. operations
Memory subsystem

Custom Computing Engine

Complete customization of computing engine for a specific application

Example for a custom computing engine

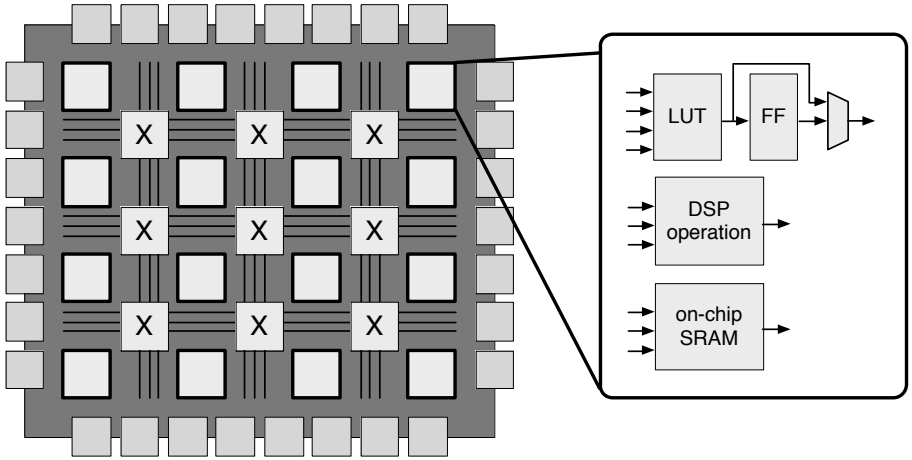
- Bioinformatics: Search in genome databases
  - Find sequence 'ATCA' in DANN database



Custom Computing Hardware

- Field-Programmable Gate Arrays (FPGAs)
  - Software-programmable electronic devices
  - Can implement any digital circuit
  - Widely-used standard component (e.g. Xilinx, Intel/Altera, Lattics, ...)

- Rapid technological improvements
  - From glue logic to HPC
  - Capacity, floating-point support, memory interfaces

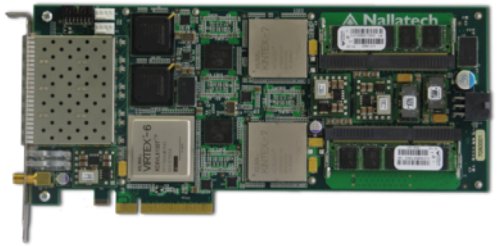


Custom Computing Systems

- Experimental systems
  - Prototypes at universities and industrial research labs (1995-2005)

TKDM, own development at ETH

- FPGA accelerator boards for PCIe interface
  - Predominantly used today



- Custom computing server for HPC
  - Maxeler, Convey, XtremeData, Cray, SRC, SGI, Micron, IBM, ...
  - Tightly coupled CPU and FPGA
  - Integrated HW/SW development tool flow

Maxeler MPC-C System

Success Stories

- Numerous publications show the potential of FPGAs for relevant HPC problems
- Some examples
  - Linear algebra: CG solver for sparse linear equation systems [1]
    - 20-40x faster than CPU
  - Geophysics: 3D convolution [1]
    - 70x faster than CPU, 14x faster than GPU
  - Molecular dynamics [2]
    - 80x faster than NAMD (single core) CPU
  - Bioinformatics (BLAST) [3]
    - 5x faster than optimized, parallel CPU implementation
  - Climate modeling [4]
    - 4 FPGAs 19x faster than two socket CPU, 7x faster than GPU

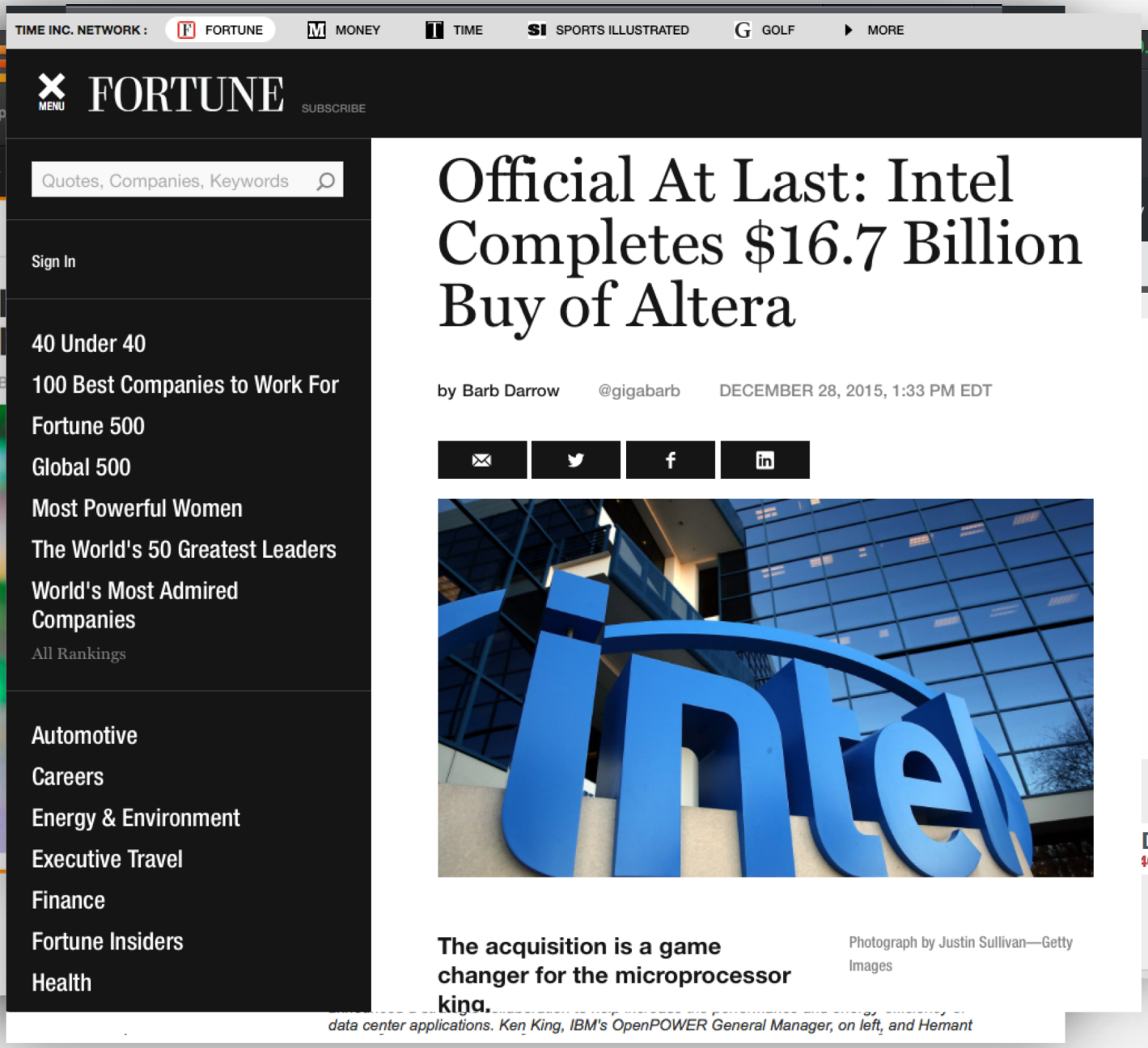
[1] O. Lindtjorn, R. G. Clapp, O. Pell, O. Mencer, M. J. Flynn, and H. Fu. Beyond traditional microprocessors for geoscience high-performance computing applications. IEEE Micro, Mar.–Apr. 2011.

[2] M. Chiu and M. C. Herbordt. Molecular dynamics simulations on high-performance reconfigurable computing systems. ACM TRETS Nov. 2010.

[3] A. Mahram, and M. C. Herbordt. NCBI BLASTP on High-Performance Reconfigurable Computing System. ACM TRETS Jan 2015.

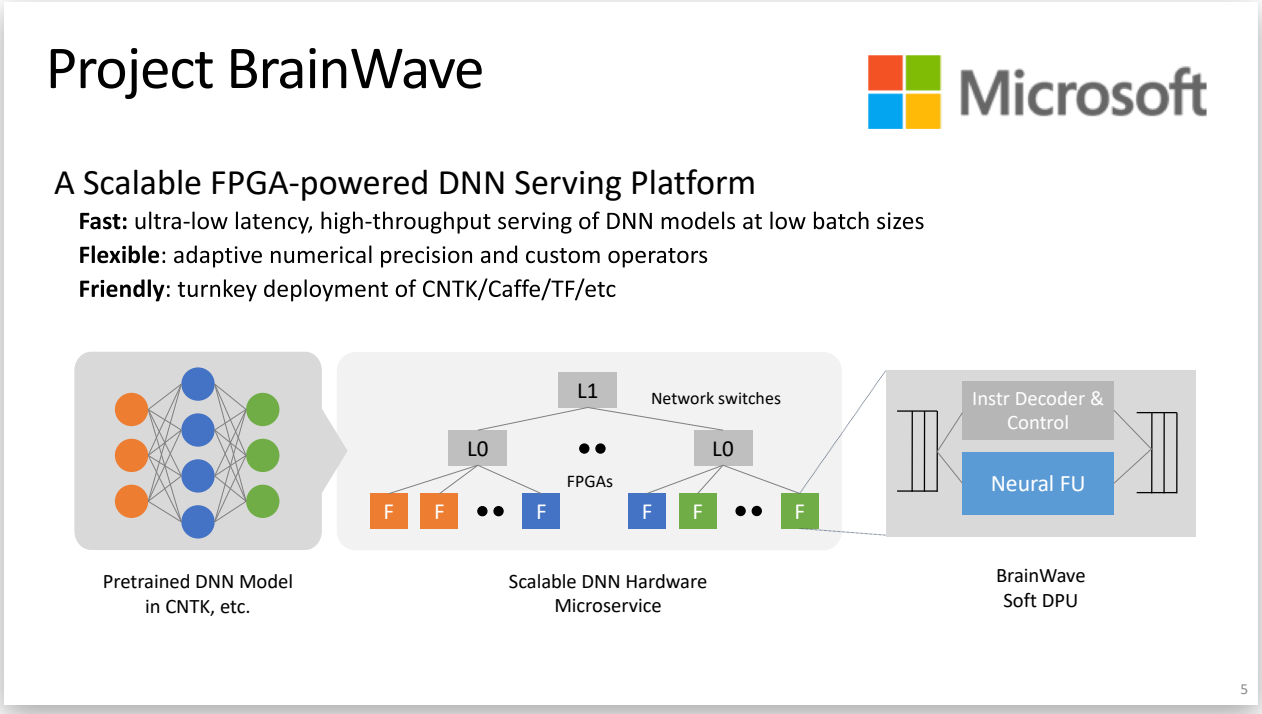
[4] L. Gan, H. Fu, W. Luk et. al. Solving the Global Atmospheric Equations through Heterogeneous Reconfigurable Platforms. ACM TRETS Mar. 2015

Custom Computing: Close to Break Through?



Deep Learning Inference

- Neural Network Acceleration: Killer App in data center
  - DNN training with GPU, double precision, very costly
  - DNN inference works with greatly reduced precisions
  - Significant improvement in latency and throughput at low batch sizes



Custom Computing at PC²

TOP500 The List.

Search Go

HOME | NEWS | LISTS | STATISTICS | RESOURCES | ABOUT | MEDIA KIT | GREEN500

Home / News / Paderborn University Will Offer Intel CPU-FPGA Cluster for Researchers

Paderborn University Will Offer Intel CPU-FPGA Cluster for Researchers

Michael Feldman | September 23, 2017 04:49 CEST

E-mail Tweet Like +1 Share

TOP500 LIST

50TH LIST ANNIVERSARY

NEWSLETTER SIGN UP

Paderborn University, in Germany, has announced it has been selected to host an Intel cluster, powered by Intel Xeon CPUs and Arria 10 FPGAs.

If you've never heard of Paderborn, that's because it's a relatively modest-sized university, located in a relatively modest-sized city of the same name. The institution enrolled just 20,000 students this year, but its Paderborn Center for Parallel Computing (PC²) is engaged in some cutting-edge research that makes it notable in the field of HPC. One area of interest is energy-efficient supercomputing, which, according the PC² website, has an emphasis on "Field-Programmable Gate Arrays (FPGAs) and Manycore architectures."

Currently, the center is working on FPGA-accelerated applications in the areas of theoretical physics,

Research Questions in FPGA-accelerated HPC

- Computer architecture
  - How to optimally integrate FPGAs and CPUs in a server and a parallel computer system?
  - How to expose close-to-metal FPGAs to users in data center in a safe and reliable way?
- Applications
  - What are the most promising application domains?
  - How to balance specialization with required flexibility in application?
  - How do FPGAs with CPUs and GPUs in terms of performance, energy, cost?
- Tooling
  - How to improve productivity of FPGA development without unjustifiable compromises in performance or resource efficiency?
  - How to create reusable and composable building blocks akin to software libraries?

Goals of the Seminar

- Familiarize yourself with a new, non-trivial scientific topics or area
  - Starting point: your assigned topic
  - Find and evaluate background material from different sources (e.g. text books and at least 3 research papers)
  - Work independently
- Present your topic in a talk
  - Presentation skills and rhetoric
  - Explain complex issues in an accessible way to the audience
  - Concentrate on the essential points
  - Preparation for defense of Master's thesis
- present your topic in a paper
  - Structure and summarize knowledge
  - Learn writing structured scientific texts
  - Preparation for writing Master's thesis

Goals of the Seminar (2)

- Additional goal of the seminar: introduction to the academic peer review and publishing process

- Academic publishing process
  - Author submits draft paper about scientific findings to journal
  - Editor distributes submission to several other experts in the field (peers)
  - Each peer writes a detailed review on the submission (numeric ratings, free form text recommending changes)
  - Based on the reviews, the editor decides about the next steps (accept as is, accept with major/minor revision, reject)
  - Several revision/review rounds until editor accepts paper for publication
  - Final version published in print and electronically

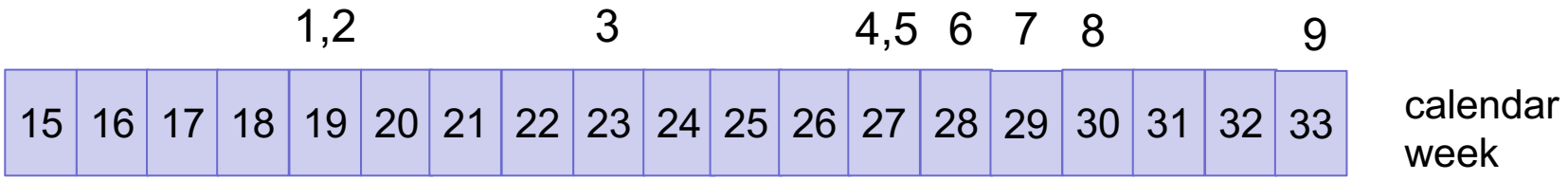
- We will use the same process in this seminar

Tasks of the Participants

- Write paper on the selected topic
  - Length: ~15 pages
  - Information based on multiple research articles
  - Written in English
  - Use of LaTeX and BibTeX text processing system (templates will be provided)
- Review papers from other participants
  - Each participant will review two papers by other students
- Present a talk on the selected topic
  - Presentation time: 30 minutes
  - Discussion and feedback: 10 minutes
  - PowerPoint template will be provided

Schedule

	Activity	Time / Deadline
1	Introduction + topic assignment	9 May 16:00-18:00
2	Lecture: reading + writing	9 May 16:00-18:00
3	Meeting with supervisor: discussion outline and bibliography	4 or 5 June, timeslot scheduled individually
4	Submission of draft paper	2 July 16:00
5	Lecture: reviewing + presenting	2 July 16:00-18:00
6	Meeting with supervisor: discussion draft presentation	12 or 13 July, timeslot scheduled individually
7	Submission of reviews	16 July midnight
8	Seminar presentations	23 and 24 July, 9:00-18:00
9	Submission of final paper	17 August midnight



Rules and Grading

- Conditions for passing the seminar
  - Attendance of all lectures and individual meetings
  - Submission of the seminar paper
  - Presentation of a talk
  - Active participation in the discussions and reviews
  - Adhering to all deadlines

- Grading
  - The grade for the seminar is determined by evaluating the quality of the paper and the presentation, the participation in discussions, and the quality and extent of the feedback given to other participants

- Plagiarism
  - Copying larger sections of text from other work without proper citation is considered plagiarism and will be result in failing the seminar

Sources for scientific literature

- Today digital libraries from major publishers provide access to almost all scientific publications
  - This is the best way to start your search for related work
  - UPB has a site subscription to almost all contents in these libraries
  - Access is free from within the UPB network (or from outside via VPN)

- Particularly relevant for the area of computer science and computer architecture are
  - ACM Portal http://portal.acm.org
  - IEEE Explore http://ieeexplore.ieee.org
  - SpringerLink http://www.springerlink.com
  - Google Scholar http://scholar.google.com

- My favorite is ACM Portal, since it aggregates information from many publishers

Next Steps

- Topics discussed in future meetings
  - Introduction to the scientific publication process
  - Reviewing papers
  - Writing research papers
  - Giving a scientific talk
- Select date for next meetings
  - Find a suitable time slot for all participants
- Start reading your paper and background literature
- Get familiar with LaTeX and BibTeX
  - Good starting point: http://tobi.oetiker.ch/lshort/lshort.pdf

Proposed Seminar Topics

- Area A: Survey and empirical studies of FPGA technology and tools
  - A1. System Architectures for FPGAs in Cloud Computing
  - A2. Customization of interconnects for parallel and distributed applications
  - A3. Technological Trends for FPGAs (empirical research)
  - A4. Survey and empirical study of the use of FPGAs in HPC
  - A5. High-Level Synthesis
  - A6. Domain-specific compilation tools for FPGAs

Proposed Seminar Topics (2)

- Area B: Survey and analysis of the use of FPGAs for tacking problems in HPC
  - B1. Deep neural networks
  - B2. Binarized neural networks
  - B3. Bioinformatics
  - B4. Image Processing
  - B5. Video Encoding
  - B6. Financial Applications
  - B7. Fast Fourier Transformation (FFT)
  - B8. Molecular Dynamics
  - B9. Database Acceleration
  - B10. Regular Expressions
  - B11. Stencil Computations
  - B12. Software-Defined Networking
  - B13. Tree algorithms
  - B14. Linear Algebra with Sparse Matrices