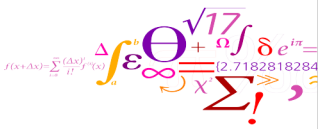
Developer Friendly Generation of Accelerators for Data Centers

Sven Karlsson
DTU Compute
Technical University of Denmark
svea@dtu.dk

Pascal Schleuniger
FNHW
pascal.schleuniger@fhnw.ch

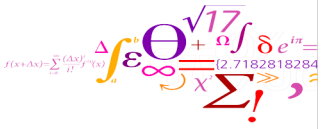


Developer Friendly Generation of Accelerators for ~~Data Centers~~ HPC!

Sven Karlsson
DTU Compute
Technical University of Denmark
svea@dtu.dk

Pascal Schleuniger
FNHW
pascal.schleuniger@fhnw.ch

Do interrupt and ask questions!



Challenges in ~~data centers~~ HPC

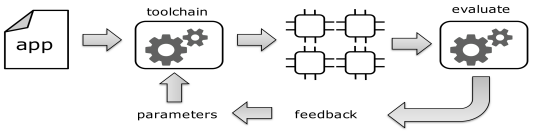
- New workloads
  - More advanced analytics; Image and video analytics.
  - AI and deep learning.
- More and more need for HPC!
  - Larger machines!; Stricter energy efficiency, IO and network requirements.
- Accelerators on FPGAs promise to address some of these challenges:
  - Good power efficiency; Flexible hardware; Quick turn-around.
- But, until now, required expert hardware design knowledge!

Outline

- Challenges in HPC and the role of FPGAs
- Tools for automated design of FPGA accelerators
- Results and experiences
- What is next? Is HPC on FPGAs viable?
- Summary

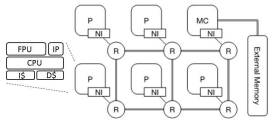
Yup! This is going to be nerdy later on! Bring out your FPGA data sheets...

Accelerator composer and design space explorer



- Tool which automatically can compose custom hardware based on application and configuration.
- Iterative feedback to designer → agile development.
- Can quickly evaluate different strategies.
- What kind of systems can be produced?

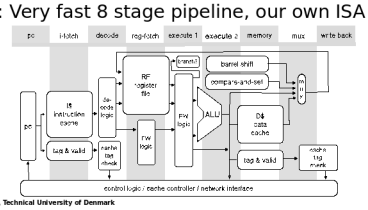
Class of systems addressed



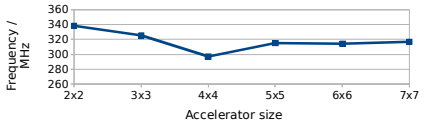
- Tools can compose highly configurable and fully programmable processor arrays
  - Based on our own high performance soft-core: Tinuso.
  - Full GCC based toolset: support for C, C++, more...
  - Full software stack: POSIX software supported; Generally programmable!
  - Technology independent.
  - Many different composable IP blocks, configuration parameters.
  - Can be cache-coherent with host.

Tinuso Architecture

- Principles: move complexity to compiler/software, predication
- Tinuso-1: Very fast 8 stage pipeline, our own ISA



Output from tool: Maximal clock frequency



- Very short accelerator development time: one week.
  - Here mapping to Stratix V, A7, speedgrade 1; Evaluating different accelerator mesh sizes.
- Performance and scale surpassing competing technology.
  - Clock speeds 50% higher!
- No hardware design knowledge needed.
  - Tools automate the process using high level application parameters

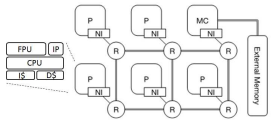
Now what?

- Tinuso-2: Clock frequency >90% of blockRAM frequency
- Double-precision FPU roughly same size as a mesh node
- Tinuso-2 and Tinuso-1 approximately the same size.

- Stratix V: 8x8 Tinuso-1 mesh takes up 83% of a device with 622 k logic elements, LEs

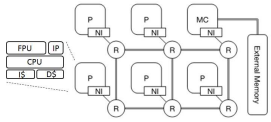
- What we know about Stratix 10
  - 1 GHz blockRAM frequency!
  - 5.5 MLEs!

How fast systems can we have on Stratix 10?



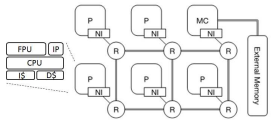
- >900 MHz system clock frequency!

How large systems can we have on Stratix 10?



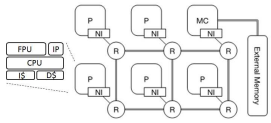
- > 680 Tinuso-1 based mesh nodes
- > 600 GOPS
- NO HLS/OpenCL involved
- Control flow not an issue
- Memory behave as the memory you are used to!
- Memory throughput limiting but no worse than for HLS/OpenCL!

I want double-precision floating point!



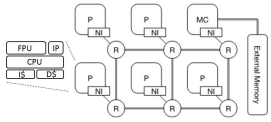
- > 340 Tinuso-1 based mesh nodes with floating point add and mult
- > 300 GFLOPS
- Ouch

Lets try vector or simd units



- Trade-off between parallelism and hardware resource usage
- > 43 Tinuso-1 based mesh nodes with wide floating point units
- > 600 GFLOPS
- A bit better!

Lets try FMA



- > 43 Tinuso-1 based mesh nodes with wide floating point units
- > 1200 GFLOPS
- If we count operations generously…

- But still 10 TFLOPS single precision!

Summary

- Accelerators promise to solve some challenges in HPC.
- So far creating accelerators has been a task for hardware specialists.
- Our tools allows application specialists, not hardware specialists, to create accelerators.
  - Performance is very good!
- Stratic 10 + Tinuso-2 = 1.2 TFLOPS double precision

Disclaimer: Our cores and tools are still research tools. Not a finished product!

Contact me for access!

Have slides, will travel!