Heidelberg Institute for Theoretical Studies
HITS

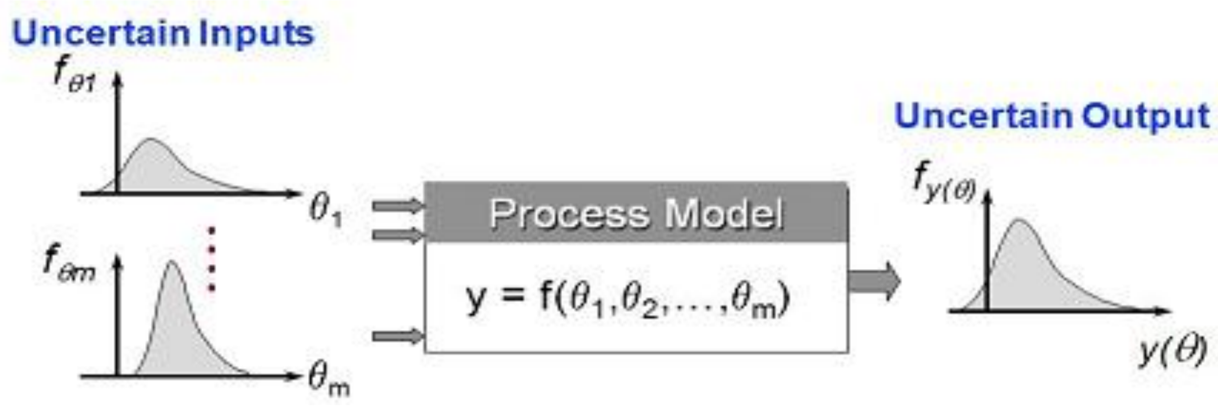
Karlsruhe Institute of Technology

Hardware-aware computing for scientific applications

Michael Bromberger[1,2], Vincent Heuveline[1], Wolfgang Karl[2], Michael Schick[1]

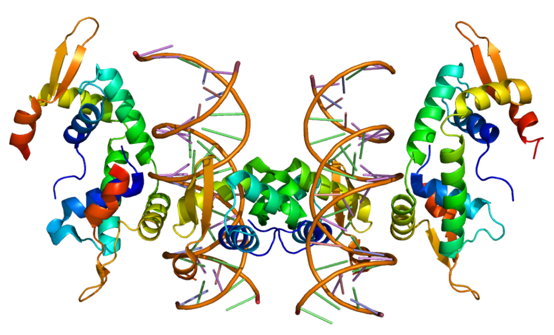
Considered applications

Uncertainty Quantification (UQ)



Galerkin projection with polynomial chaos
Monte Carlo Simulation

Computational biology/Computer vision



Analysis of protein sequences

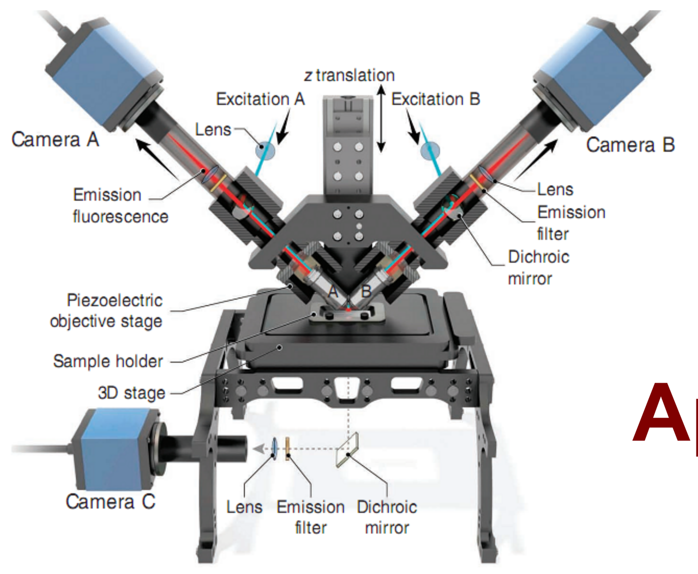


Image processing

Execution time?
Precision?
Accuracy?
Performance?
Energy consumption?

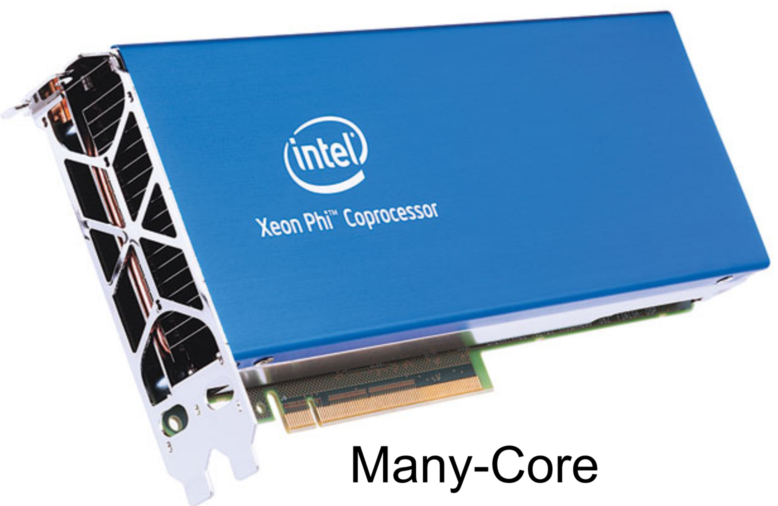
?

Parallelization?
Programmability?
Applicability?
Availability?

Hardware architectures

CPU

GPU



Many-Core

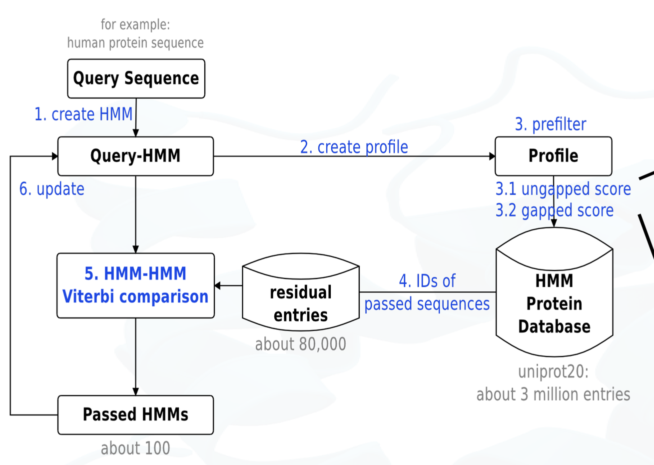
FPGA-based workstation

Server

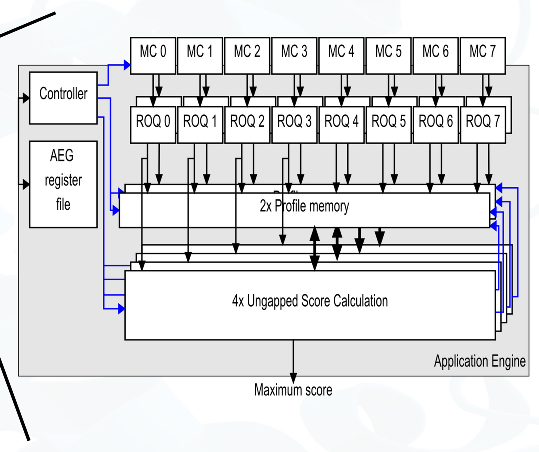
FPGA-based server

Previous and current work

Porting functions to special hardware



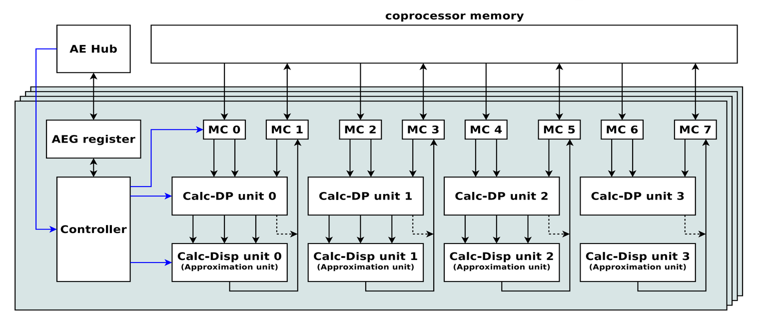
Application HHblits



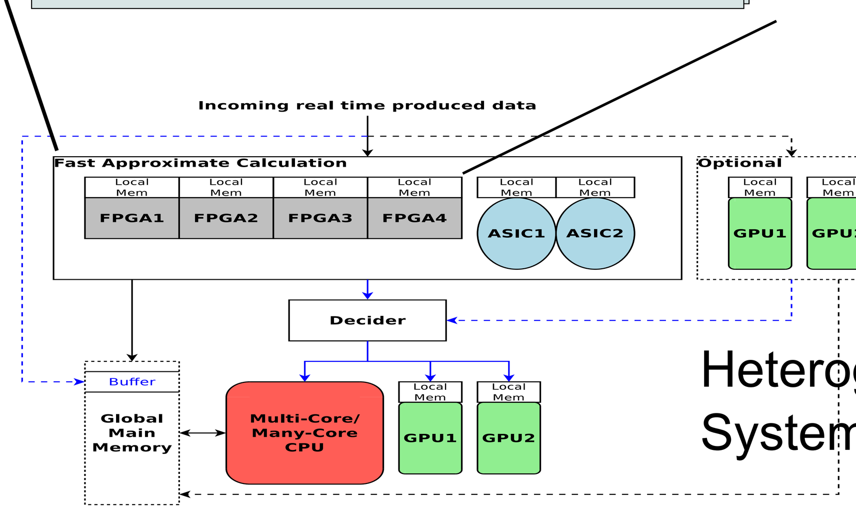
FPGA-based prefilter

FPGA-based prefilter unit is faster than the original SSE execution for larger protein sequences. The Convey HC-1 allows the integration of 16 such units. The first level prefilter is 3.98 times faster against the original implementation for the uniprot20 database. A wise hardware-software pipeline is used for the entire application.

Approximate computing-based accelerators



Approximate Computing Unit (ACU)

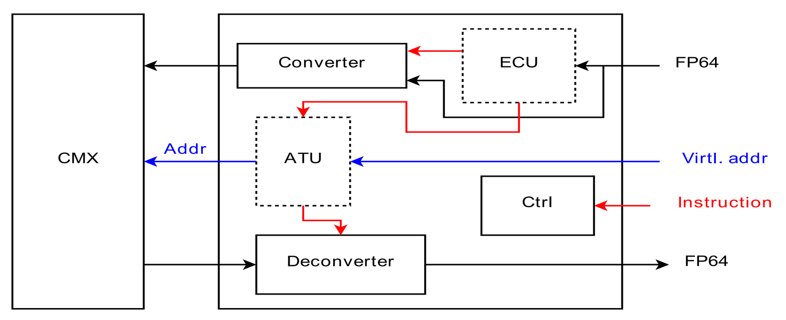


Heterogeneous System

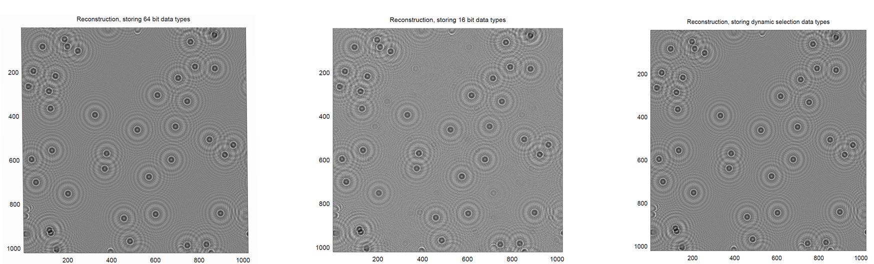
ACU executes the calculation of a depth map 367x faster than the original dynamic programming based algorithm by increasing the mean square root error by a factor of 2. A heterogeneous system can be used to recalculate the results if it is required.

Reduce data transfers to save energy

A memory access consumes 1000 times and transferring data via a wire-less network consumes 1,000,000 times more energy than an integer operation. Therefore, it is very important to reduce the amount of data that has to be transferred. Using a conversion unit between the processor and the main memory allows the programmer to decide between accuracy and energy consumption.



Conversion Unit



64 bit FP 16 bit FP Dynamic FP

Spectral method for a lens-less microscopy

Future and projected work

Porting UQ methods to hardware accelerators

Method: Galerkin projection with polynomial chaos

Step 1:
Investigation of different preprocessing methods to reduce the bandwidth of the matrices that are required for the calculation. This reduces the amount of data that has to be transferred to accelerators as well as the complexity of the matrix-vector product ($O(n^2)$ to $O(n)$).

Step 2:
Developing strategies for implementations on GPGPUs and FPGAs. Furthermore, designing an approximate computing based sparse matrix-vector product.

Comparing different UQ methods in terms of energy consumption

Motivation: Besides the accuracy/precision and the performance of an application the energy consumption is at least equally important nowadays.

Approach: Porting UQ methods based on the galerkin projection or Monte Carlo simulation to different hardware units without optimization. Compare the different implementations in terms of performance and energy consumption.

Optimization: Developing concepts to reduce the energy consumption for the different considered hardware units.
Possible strategies: - increasing accuracy to increase the rate of convergence
- approximate computing (reduction of the accuracy)

1 Heidelberg Institute for Theoretical Studies, Heidelberg, Germany
2 Karlsruher Institute of Technology, Karlsruhe, Germany

Reference
Michael Bromberger, Fabian Nowak, and Wolfgang Karl: **Combined hardware-software multi-parallel prefiltering on the Convey-HC-1 for fast homology detection**, Journal of Parallel Computing (2014), Elsevier, DOI: 10.1016/j.parco.2014.09.006