Selected Applications on Data-Intense Research

super computing systems by David Müller and Christof Bühler

Abstract / Rational

- Modern science, engineering, businesses are confronted with massive growth of data (10^{15} – 10^{18} Bytes per day).
- Imposed challenges encompass processing speed, storage capacity, power, cooling, and especially human factors:
  While the power of information systems increases – the amount of information a human can directly digest does not.
- Key to discoveries is data reduction that must aim at providing good clues for the right questions.

Challenges at SwissFEL

- Data rates / volumes: 10 - 100 TeraByte / day / detector
- Throughput (bandwidth) and latency bottlenecks:
  readout ⇨ pre-processing ⇨ storage ⇨ post-processing
- Diverse detectors: BPM, Pixel Detector, CMOS/CCD cameras
- Number of detectors (channels)
- Diverse data formats
- RAM = Reliability, Availability, Maintainability

SCS Company Profile (www.scs.ch)

- founded in 1993 by Prof. Dr. Anton Gunzinger
- provides development services for industrial and academic customers in Switzerland, Europe, and US
- ≈75 employees: electronic and software engineers, physicists, and mathematicians
- Application fields: HPC, embedded computing, intelligent sensors, life science, enterprise applications
- Know How: HW, SW, algorithms, system design

Subject of Research

Experiment or Simulation Code

Data Generation

Readout

Data Pre-Processing

Filesystem Write

Storage

Filesystem Read

Data Post-Processing/ Visualization

Data Analysis

Results

High Energy Physics (HEP)
Climate Science & Weather Forecast
Market Analysis
Social Network Analysis
Life Science
SwissFEL Beam Diagnostics
SwissFEL Experiments

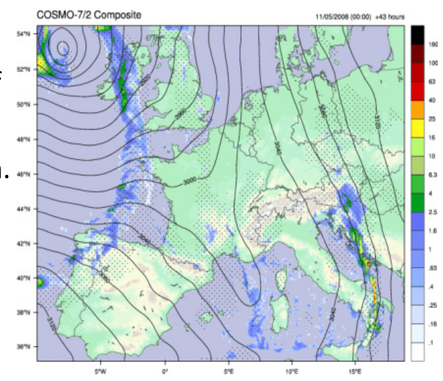
COSMO Dynamical Core

COSMO is a local scale atmospheric model for weather forecast and for climate research. Its dynamical core computes the non-hydrostatic compressible dynamical equations; it is one of the main workloads for supercomputers in Switzerland. The COSMO dynamical core is very data-intense and requires optimized usage of the memory bandwidth available in the system.

SCS is re-writing the dynamical core

- Achieve highest performance
- Achieve performance portability on x86 and GPU architectures
- DSEL (Domain Specific Embedded Language) based on C++

This ongoing work is executed in collaboration with C2SM, MeteoSwiss and CSCS.
Another task focuses on data reduction and enhanced archiving strategies.



Precipitation forecast calculated by the COSMO model on a grid resolution of 7 km x 7 km

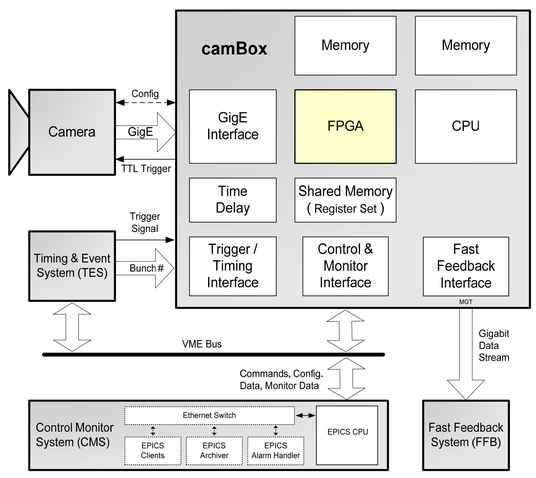
FPGA-accelerated Image Processing for SwissFEL

For the SwissFEL Free Electron Laser at PSI (Villigen) an novel FPGA-accelerated camera system is developed that

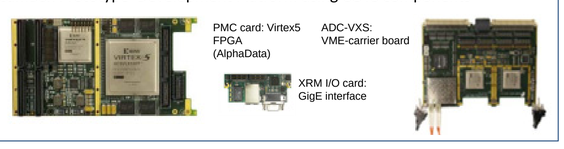
- records and analyzes beam profiles at a frame repetition rate ≤ 100 Hz
- provides a generic camera interface (initially: 2 PSI-selected GigE cameras)
- is fully controlled and operated by EPICS
- is synchronized by the SwissFEL Event & Timing logic
- corrects for salt & pepper noise, cosmics, offset & dark levels, gamma function, camera rotations (≤ 5°,+/- 90°,180°)
- calculates ROI-based intensity profiles along camera's x- and y-axis
- derives key beam parameters: center position, width (FW%M) via different methods (e.g. CoM, curve fitting)
- is implemented as prototype on commercial card (Virtex5 card, AlphaData) and then migrated to the production system.

Joint project of PSI and SCS started in 2011.

Diagram of camBox Prototype System - Design and logic Architecture



camBox Prototype Development Platform using COTS components



Fast Track Trigger (FTT)

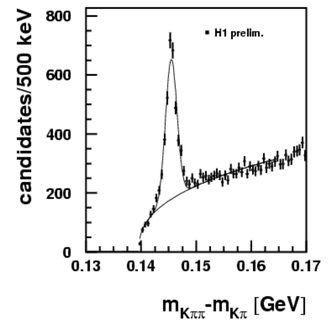
The FTT reduces the event rate in a particle physics experiment before the data is read out of the frontend electronics. The FTT reconstruct particle tracks and invariant masses.

- Used at the H1 detector @ HERA 2000
- Linking of particle track segments using Content Addressable Memory (CAM) at full input rate (10MHz) in FPGA
- Reconstruction of invariant masses in DSP software

Joint project of ETHZ, Universität Dortmund and SCS in 2001.

Trigger Rack

Scientific Target: Highly Enriched Event Selection for D* Decays

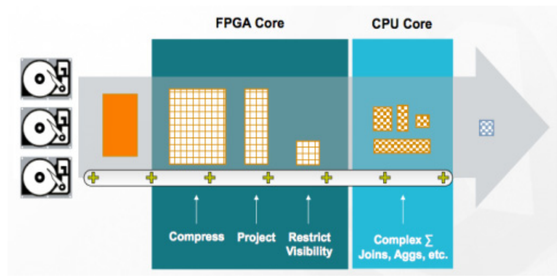


NETEZZA Data Mining Acceleration

The IBM Netezza Data Warehouse Appliance focuses on massive parallel processing of huge datasets. Search-operations (project / 'SELECT' and restrict / 'WHERE') are implemented in FPGA hardware. The appliance therefore can process Petabyte-scale tables as fast as they stream from the underlying parallel disk system.

SCS is a development partner of Netezza.

Netezza – Highest performance data selection engine © by Netezza



MPI- & GPU-accelerated Image Processing for Quantiative Biology

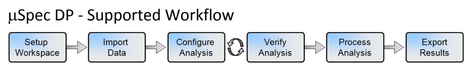
New approaches in drug discovery focus on cellular imaging often denoted as high-content screening (HCS). Current HCS platforms process 50,000 - 200,000 images/day, generate hundreds of GB/day, and require extended processing times (hours). The developed program 'Micro-Spectroscopy Data Processor' (µSpec DP) provides fast and fully automated analysis of cellular structures and molecular signals from thousands of micro-spectroscopic images.

The software µSpec DP fully supports

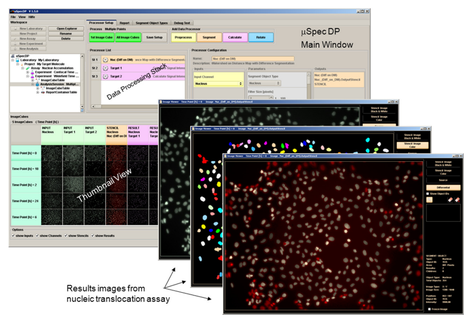
- data processing workflow: import, analysis setup, verification, processing, export
- interactive and automated analysis of cellular structures (nuclei, membranes, speckles, …) and molecular signals from micro-spectroscopic images (vector data)
- intuitive visual tools ensuring seamless data processing
- two-stage parallelization techniques
  1) MPI (MPJexpress) ⇨ multi core and/or cluster mode (SPMD)
  2) NVIDIA's CUDA ⇨ core image algorithms (SIMD)

Joint project of Novartis and SCS started in 2010.

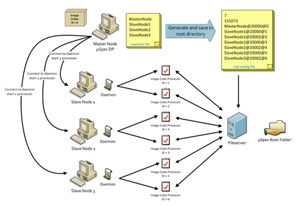
µSpec DP - Supported Workflow



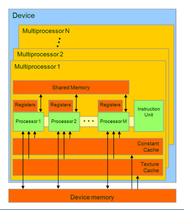
µSpec DP main window and result panels



1) MPI-based data processing on multi-core and/or networked clusters



2) GPU-accelerated image processing (filters, segm.)



Acknowledgments: H. Braun, R. Ischebeck, T. Korhonen, D. Anicic, B. Patterson, R. Abela (Paul Scherrer Institut, Villigen, Switzerland)