Scientific Computing at MPP

Oliver Schulz

MAX-PLANCK-GESELLSCHAFT

Max-Planck-Institut für Physik
(Werner-Heisenberg-Institut)

oschulz@mpp.mpg.de

MPP Project Review, December 18, 2018

Outline

Available computing resources

- In-house batch-system with CephFS storage system
- MPP Linux-cluster at MPCDF
- MPG supercomputers Cobra and Draco at MPCDF
- Experiment-specific resources (Grid, . . .)

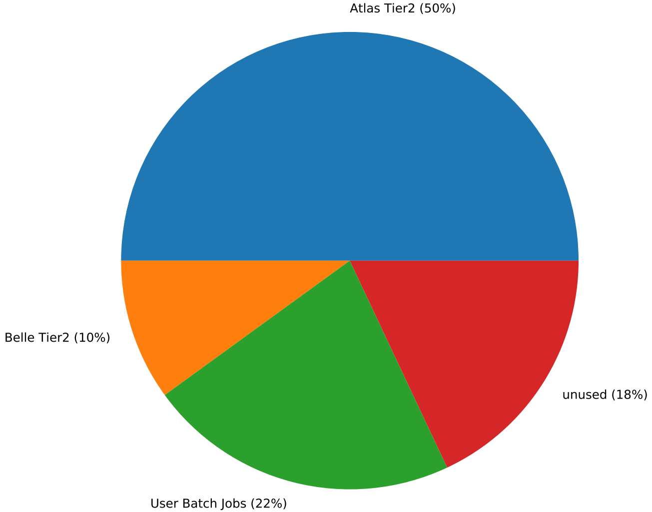
In-house batch-system

- Condor batch system, utilizes spare computing capacity on user workstations (Ubuntu and SUSE Linux)
- Computing capacity (Condor):
  About 120 nodes, 800 cores, 7TB RAM
- Storage capacity (CephFS): 500 TB, 90% full. 180 TB extension delivered, commissioning in January.

MPP Linux-cluster at MPCDF

- Computing capacity: 130 nodes, 3544 cores, 3.5 TB RAM
- Storage capacity: 600 TB Storage (GPFS), 3 PB dCache (2PB used, mainly Atlas)
  - GPFS: 600 TB
  - dCace: 3 PB (2PB used, mainly my ATLAS)
- Operating system: SLC-6, update to CentOS-7 in January
- New: Two high-RAM nodes for new department J. Henn, 192 cores with 6TB RAM and 36 cores with 3TB RAM

MPP Linux-Cluster Utilization 2018



MPG supercomputer COBRA at MPCDF

- New supercomputer COBRA installed at MPCDF in February and April 2018
- Total: 3188 nodes, 127520 cores, 483 TB RAM
- Storage capacity: 5 PB (GPFS, 0.75 PB perm., rest temp.)
- Peak performance 10 PetaFlop/s (Hydra had 1.7)
- Fast OmniPath interconnect,
  5 domains (islands) with fat-tree topology
- Old supercomputer Hydra decommissioned Nov. 14, 2018,
  Hydra extension Draco continues to be available

Software projects at MPP (and beyond)

- Quality of software more than ever a crucial factor for successful research (theory and experiment)
- Avoid re-inventing the wheel, pool resources, release as open source for the scientific community
- Problem: Software development still often treated as a second-class scientific activity, reflected in the resulting software
- Selected success stories: CUBA, BAT, Ploughshare, SecDec, CRESST software trigger, new LEGEND-group software
- Many other important software projects driven by MPP

Cuba

Multidimensional numerical integration

- Motivation: Very common problem, but efficient and stable solutions highly non-trivial
- Developers: Thomas Hahn et al.
- Four different integration algorithms, all with C/C++, Fortran, Mathematica and Julia interface
- Multi-purpose, used in many physics software projects
- Also non-physics / industry users
- New release 4.2 in September 2018
- Homepage: http://www.feynarts.de/cuba/

BAT: Bayesian Analysis Toolkit

- Motivation: Bayes' theorem simple on paper, but numerics are hard
- Allen Caldwell et al., multi-institute involvement: at MPP, TUM, TU-Dortmund
- BAT v1.0 release in May 2018
- Successor BAT.jl (a.k.a BAT-2) under development, completely rewritten in Julia, release planned for Spring 2019.
- New Adaptive harmonic mean integration algorithm (AHMI) calculation of Bayes factors in higher dimensions.
- BAT.jl will also offer increased parallelization, and multiple MCMC algorithms (e.g. HMC with automatic differentiation)

Ploughshare → fastNLO/APPLfast

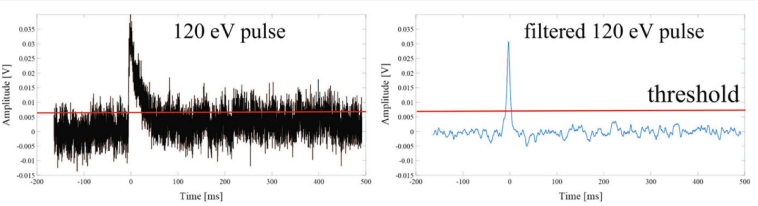
- Higher order QCD calculations (e.g. NNLO QCD) computationally intense, makes phenomenological studies, such as PDF, α_s, etc. challenging
- Software tools fastNLO and APPLgrid can store calculations in a small files
- HEP community highly interested in interpolation grids for NNLO calculations (already standard method for NLO)
- Work at MPP: Interfacing of fastNLO and APPLgrid with generators, calculation of grids at MPCDF.
- Example ATLAS inclusive jets: 400k CPU hours, 40 cores, 180GB RAM per node, ≈10TB temp. disk usage, results in final files ≈20MB
- Lead by Daniel Britzger (joined MPP in 2018)
- Close collaboration between experimentalists and theoreticians
- Web: ploughshare.web.cern.ch, fastnlo.hepforge.org

SecDec

Numerical evaluation of dimensionally regulated parameter integrals (see Talk by G. Heinrich)

- Motivation: Discovery of BSM physics without "smoking gun" needs precision calculations
- Developed by group around G. Heinrich
- Calculations with high precision can be very costly: Use multiple GPUs
- Recent publication: "A GPU compatible quasi-Monte Carlo integrator interfaced to pySecDec" (arXiv:1811.11720, Nov.2018), coupled with release of qmc v1.0.0
- Homepage: http://secdec.hepforge.org/
- Development and performance testing in-house on MPP GPU servers

CRESST software filter and trigger



- CRESST introduced software processing of full, non-triggered detector signal data stream
- Software optimum filter maximizes signal-to-noise ratio, individually for each detector
- Software trigger and energy reconstruction can use same algorithm
- Artificial pulses at random positions in the data stream allow for precise measurement of efficiency
- Selected CRESST detector: Hardware trigger threshold 48 eV, software trigger threshold 30 eV

New LEGEND-group software stack in julia

- MPP LEGEND group is developing a completely new software stack for Germanium-detector experiments
- Complete rewrite of many scattered C++ software packages in Julia, with substantial additions
- Project fairly advanced, fully open-source, first preview demonstrated to LEGEND collaboration in Dec. 2018
- Aims:
  - Become (mostly) feature complete during first half of 2019
  - High performance, support multi-threading, GPUs, etc.
  - Fully integrated: Support DAQ-components, slow-control, detector simulation, signal analysis, spectrum analysis, . . .
- See talk by A. Zsigmond (new detector simulation software)

Hardware Landscape

- Single-core performance has reached physical limit, still small incremental improvements, but no major gains any more, no solutions for much higher serial performance in sight
- Unbroken trend to higher parallelization:
  rising CPU core count
  and heterogeneous architectures (CPU + X)
- Current heterogeneous architecture of choice: CPU + GPU
- Future: CPU + FPGA,
  CPU + data-flow hardware (Intel CSA)?

Software Landscape

- Current software typically does not really utilize current hardware: SIMD, multi-threading, SIMT(GPU), mixed precision, tiered I/O, ...
- Not a new problem, but gap increasing
- Exceptions like machine learning with DNNs and simulations on grids: Excellent software stacks and tooling, e.g. users get GPU support etc. without effort
- Funding agencies are beginning to require that software fully utilize available hardware
- Need to re-think how to write software: Algorithms, implementations, programming languages, everything
- No easy solutions, but new software technologies available: C++ SYCL, Julia CUDAnative, ...
- Software development becomes more challenging, increased skill set required - we need more training!

Summary

- Substantial computing resources available at MPP and MPCDF - choose the right set of resources for your task
- New resource since 2018: MPP supercomputer Cobra
- MPP very active in various software projects with high visibility and broad applicability
- Hardware landscape continues to changes, software is not keeping up, situation is getting critical
- Unrelated idea: Some form of regular in-house informal get-together on scientific software development?