Challenges in the HEP SW Landscape

Jahresversammlung der deutschen Teilchenphysiker 2016
Bad Honnef

18. Nov. 2016

Benedikt Hegner
EP-SFT - CERN

Thanks to Ian Bird for most of the input on resources

Outline

- Where software impacts physics
- HL-LHC a major challenge
- Software Upgrade Project
- Software at Universities
- HEP software in Germany

Why is software a hot topic?

- This year we've seen a remarkable LHC performance
- Computing needs directly impacted by
  - LHC live time (37% ⇒ > 60%)
  - Luminosity ($1.0 \times 10^{34} \Rightarrow 1.2 \times 10^{34}$ or better)
  - Pile-up (CMS, ATLAS) (21 ⇒ 33 on average)
- Tough job to get all the data stored & processed
  - It somehow worked out!
- Situation not yet resolved for 2017

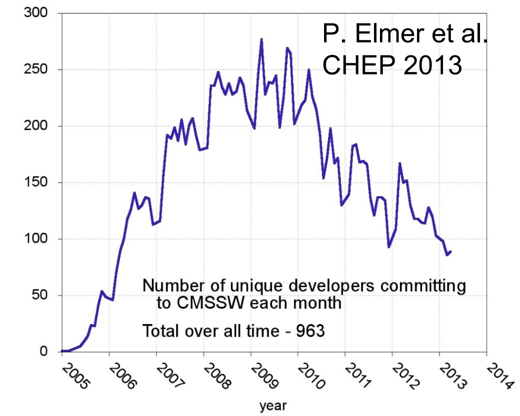
Physics started to be limited by computing and software

Limitation on the Physics Output

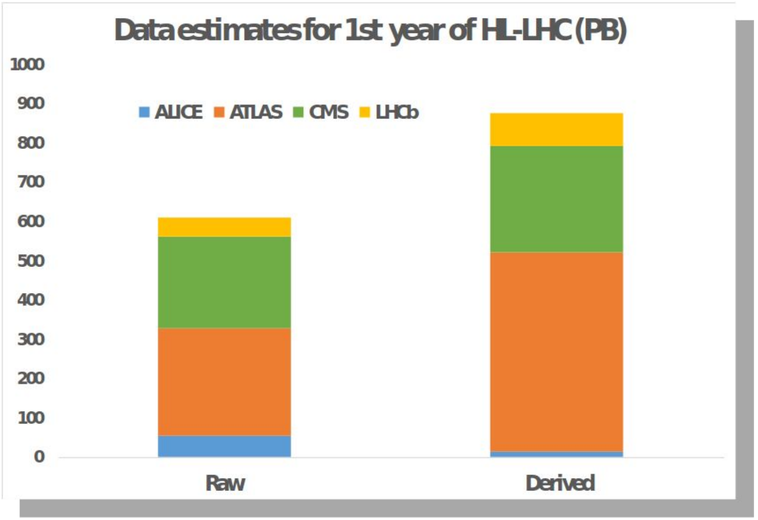
- Trigger systems could deliver much more if offline processing coped with it
  - Instead may need to narrow trigger menu even more
- Production of full simulation doesn’t keep up with demands
  - The move to fast sim costs precision and validation efforts
- Re-processing times a common bottleneck for publishing results
- Preparation of hardware upgrades relies on proper simulation
  - Conflicts with manpower needs for operation

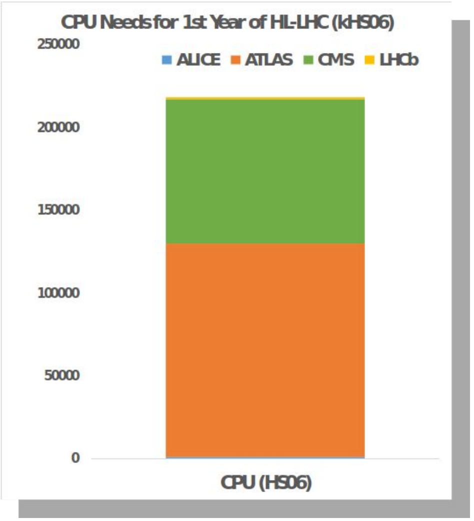
And it is not expected to improve

Quite the contrary…



Estimates of Resource Needs for HL-LHC





Data:

- Raw 2016: 50 PB → 2027: 600 PB
- Derived (1 copy): 2016: 80 PB → 2027: 900 PB

CPU:

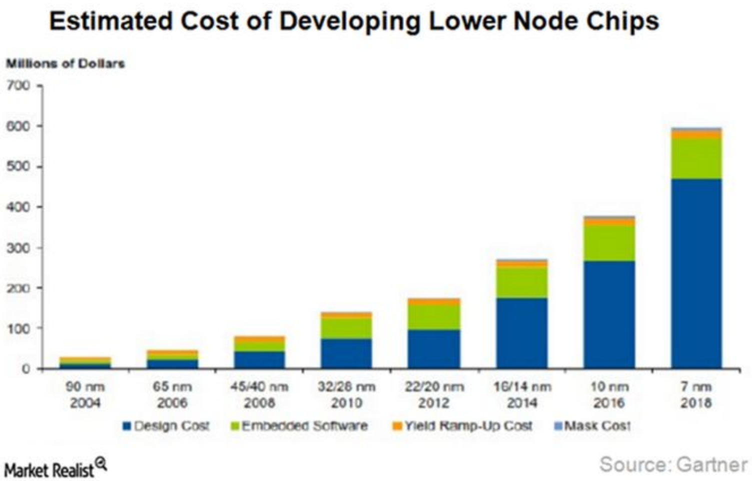
- x60 from 2016 (*)

Technology at ~20%/year will bring x6-10 in 10-11 years (assuming a flat budget)

Missing factor of 10 in CPU has to come from SW innovation

(*) In this number ATLAS and CMS already assume they have to limit themselves on MC statistics

Slow down in processor technology development



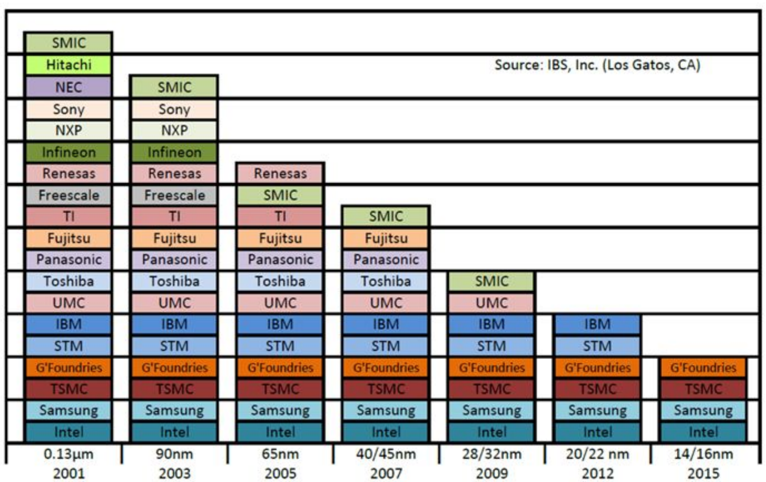
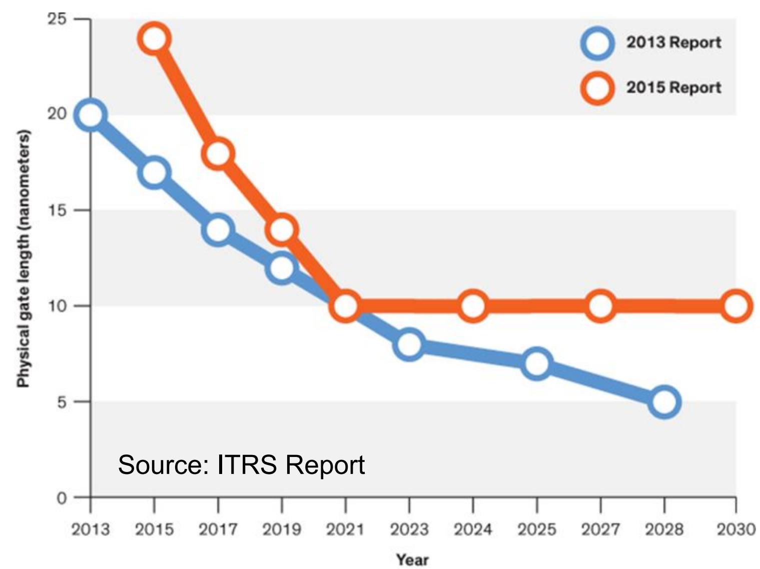


Figure 4. Dramatic Consolidation of state of the art CMOS Fabs. Source: IBS , Inc. (Los Gatos, CA).

Non-linear costs for development

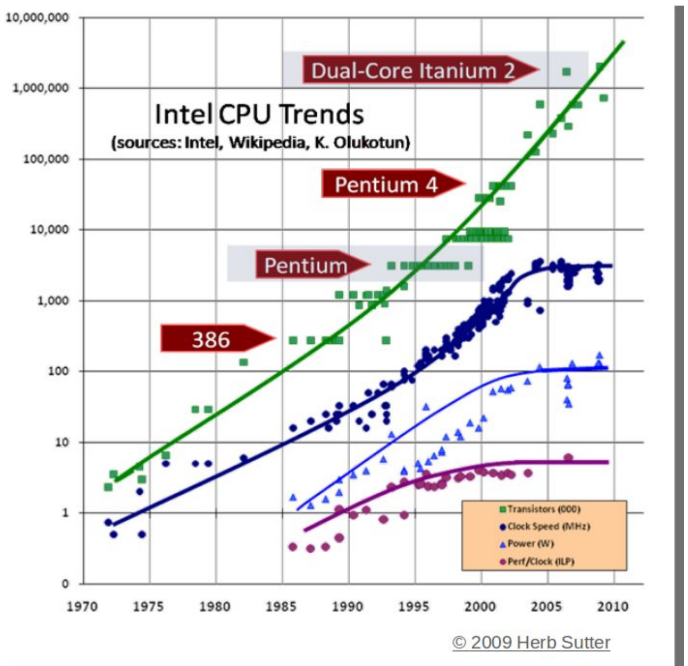
Only four companies able to fabricate 14 nm chips

10 nm Samsung fab costs \$14 B



Slow Software on modern CPUs

- Not just a HEP problem
- Transistors go into many cores, co-processors, vector units, etc.
- Memory access and I/O paths also become problematic
- Getting performant software requires significant investment in skills



The new realism: software runs slowly on supercomputers

No supercomputer runs real applications faster than five percent of its design speed. Robert Roe and Tom Wilkie report on recalibrating expectations of exascale and on efforts to tune software to run faster

director of the HLRS supercomputer centre in Stuttgart, Germany. The HLRS was not targeting exascale per se, he said, because the centre's focus was on the compute power it could actually deliver to its users.

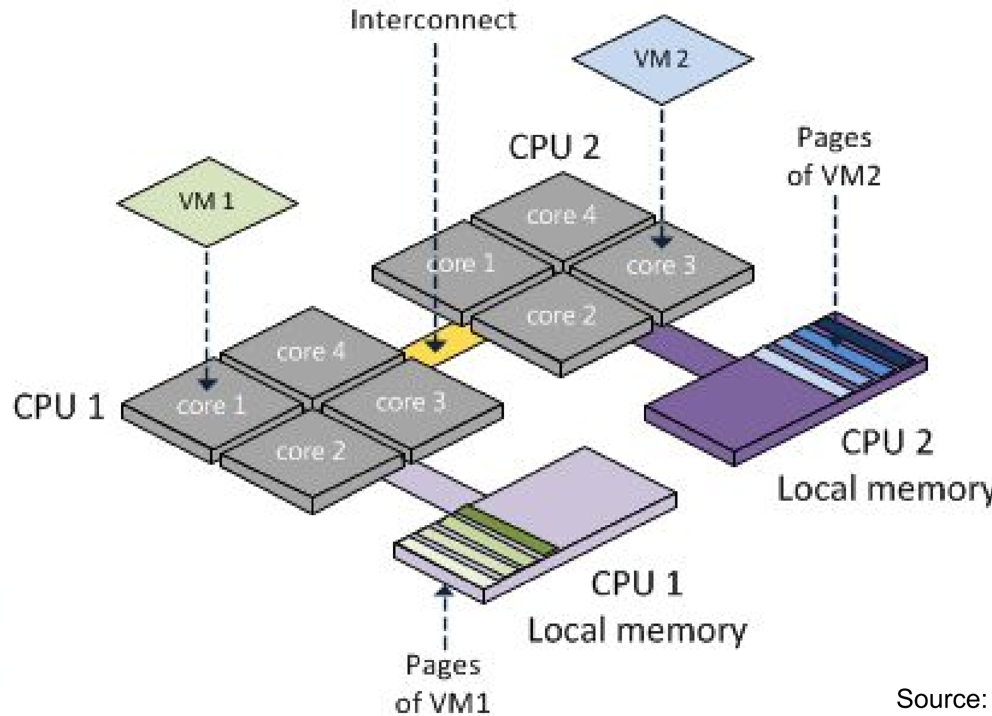
According to Resch, simple arithmetic meant that if an exascale machine achieved a sustained performance of only 1 to 3 percent, then this would deliver 10 to 30 Petaflops. So buying a 100 Petaflop machine that was 30 per cent efficient – which should be achievable, he claimed – would deliver the same compute power, for a much lower capital cost and about one tenth the energy cost of an exascale machine.

The Jülich Supercomputing Centre (JSC)

encourage our users to try and reach exascale readiness.'

At the Lawrence Livermore National Laboratory, effort is going into developing APIs and tools to create applications and effectively optimise how code is run on a cluster. Supinski stated that the LLNL's plan was to use a programming tool developed at Livermore called Raja. 'The idea of Raja is to build on top of new features of the C++ standard. The main thing that we are looking at is improving application performance over what we are getting on Sierra, Sequoia, and Titan. Application performance requirements are what we really care about.'

'The US Coral programme means that



Source:
Scientific Computing World

Additional challenges

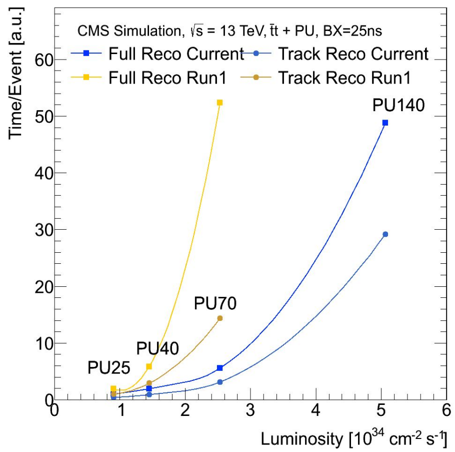
- Experiments are moving towards merging online and offline processing
  - LHCb and ALICE going for triggerless setups

- High pile-up a challenge for track reconstruction
  - Can we take sufficient advantage of our HW upgrades?

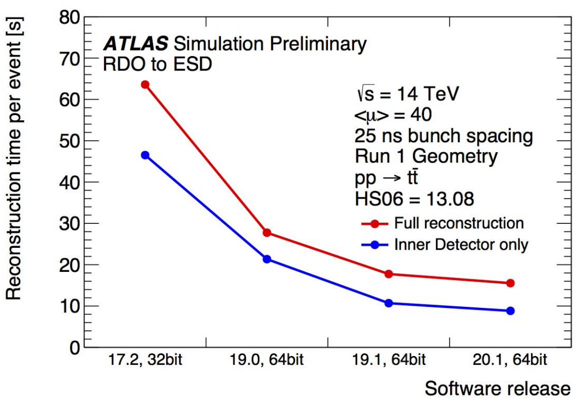
Again software performance has direct impact on physics performance

Physics (code) optimization in the past

- During LS1 all experiments invested heavily in improving their software
  - Required detailed expert knowledge rare
    - Important to invest on the right topic
  - Significant gains achieved



- One example - ATLAS and CMS Tracking
  - Better/vectorized implementations of calculations
  - Addressing hot-spots, e.g. magnetic field in ATLAS
  - Tuned reconstruction strategy for higher pileup, e.g. new seeding algorithms
  - Up to x4 improvements in speed



- Low-hanging fruits mostly gone now

A SW upgrade project for HL-LHC and 2020s

Looking forward to the next 10 years, we see the following challenges:

- **Scale:** The HL-LHC will integrate 100 times the current data, with significantly increased data (pileup) and detector complexity.
- **Performance/cost:** Estimates of computing needs run faster than Moore's Law by factors of 3-30
- **Technology/Market evolution:** the return of heterogeneity; technology change will also make it challenging to exploit Moore's Law without software evolution.
- **Infrastructure:** GRID will be replaced by a mix of GRID, clouds, HPC
- **Sustainability:** Most of the current software was designed 15-20 years ago; there are many software sustainability challenges.

HEP Software Foundation and Whitepaper

The HSF (http://hepsoftwarefoundation.org) was created 1.5 years ago to help addressing all these problems in the long run.

One goal is to create a *Community White Paper (CWP)* on the overall strategy and roadmap for software and computing

- Paper to be delivered by summer 2017
- Base for discussing possible funding scenarios for both the software upgrade and SW projects of Belle-II, ILC, neutrino program, etc.
- "Kick-off" workshop at UC San Diego on 23-26 Jan 2017
    - Individual working groups started already
- For details see **http://hepsoftwarefoundation.org/cwp.html**

Software at Universities

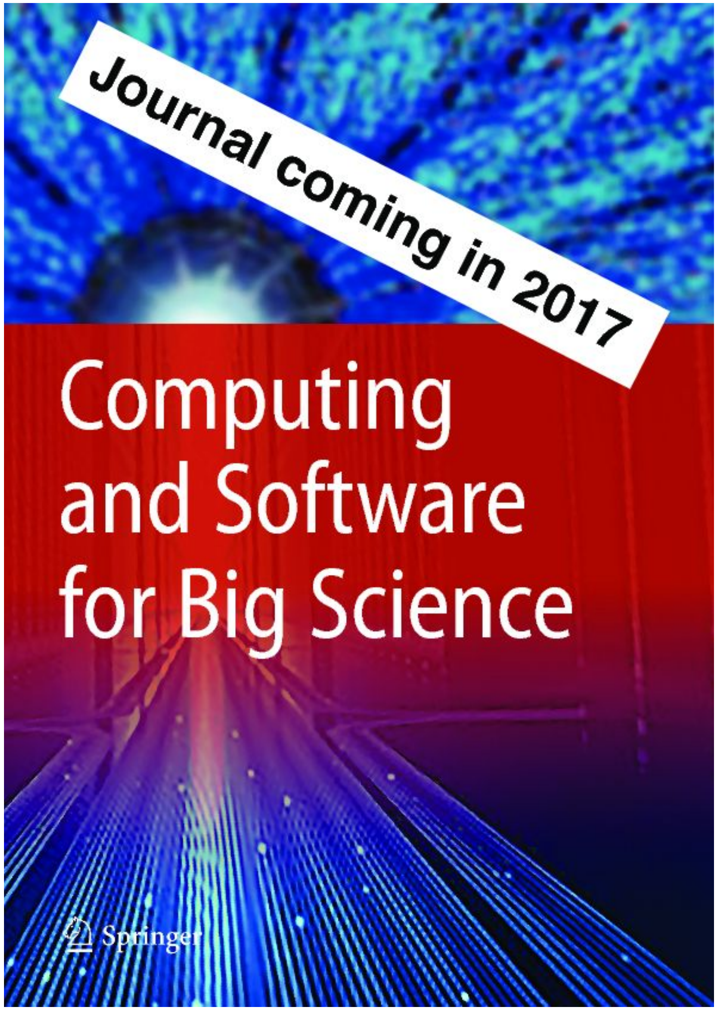
- Nowadays there is no physics without being exposed to software
  - One of the few common denominators between theory and experiment!
- It is by no means an easy topic
  - But proficiency translates directly into productivity
- Unguided students repeat the same mistakes all over again
  - 20 workarounds instead of 1 fix
  - Plenty of n-tuple frameworks

Software as part of a Physics Career

- Software expertise is valuable for almost every project
- It is more than just coding - one needs to know the physics
  - Even ‘technical work’ like ATLAS’ move to xAOD can only be done with physics in mind - and has impact!
- As valuable as software is, software experts aren’t
  - No career track for physicists with strong focus on SW
  - Despite ever increasing complexity and need
- Physics software should be recognized as ***scientific*** contribution

Journal Software in HEP and other fields

- Need to enable software oriented physicists to publish their work in high profile journals
- Efforts in the german community to set up a journal alongside EPJC with springer
  - Will start in 2017



Software as part of Education

- In the past doing well on producing “data scientists” and “data engineers” within HEP
  - Being the only field with real *big data*
- Machine-learning as one of many new exciting topics
  - Interesting R&D to improve on our data analysis methods
  - Students and companies are asking for it!
- Expertise can only be taught by experts
  - High-quality education relies on high-quality research

How to Connect to the Community

- All HEP experiments have software R&D projects waiting for new contributors
  - Plenty of dedicated workshops
  - Huge “return-of-investment”
- Community-White-Paper defining the direction for the next decade
- The SW community tries to actively reach outside
  - Summer schools by Helmholtz Alliance
  - HEP Software Foundation
  - *Connecting the Dots* workshop
  - *CHEP* and *ACAT* however exist since long

German SW contributions to CHEP

- CHEP takes place every 18 months
  - The place to show one's work to the community
- In 2016 about 300 presentations on SW
  - No plenary from Germany
  - 22 parallel talks (with a very wide definition of SW!)
    - 8 Heavy Ion, 4 Belle II, 5 ATLAS/CMS/LHCb
    - A list in backup slides
- Germany activity well below average given its size
  - In particular at the LHC

Existing efforts in Germany

- Important contributions to Belle-II and ILC
  - Framework and tracking
- Online system of ALICE and for FAIR experiments
- Some smaller efforts of (mostly) individuals
  - e.g. alignment tools, deep-learning or web-based analysis

- Can we establish a German-wide strategy?
  - Invest to ensure the physics output of our experiments and the competitiveness of our groups
  - Which are the fields of expertise to focus on?

Backup

German SW talks at CHEP 1/3

ATLAS / CMS / LHCb

- Machine learning and parallelism in the reconstruction of LHCb and its upgrade (Marian Stahl, Heidelberg)
- Alignment of the CMS Tracker: Latest Results from LHC Run-II (Gregor Mittag, DESY)
- Accelerated tracking using GPUs at CMS High Level Trigger for Run 3 (Felice Pantaleo, CERN/Uni Hamburg)
- Modernizing the ATLAS Simulation Infrastructure (Andrea Di Simone, DESY)
- Frozen-shower simulation of electromagnetic showers in the ATLAS forward calorimeters (Ksenia Gasnikova, DESY)

Belle II

- High Level Interface to Conditions Data at Belle II (Martin Ritter, LMU)
- The Simulation Library of the Belle II Experiment (Martin Ritter, LMU)
- Software Quality Control at Belle II (Martin Ritter, LMU)
- An interactive and comprehensive working environment for high-energy physics software with Python and jupyter notebooks (Thomas Hauth, KIT)

German SW talks at CHEP 2/3

Heavy Ion Physics

- A precision device needs precise simulation: Software description of the CBM Silicon Tracking System (Hanna Malygna, GSI)
- The high-rate data challenge: computing for the CBM experiment (Volker Friese, GSI)
- GPU-accelerated track reconstruction in the ALICE High Level Trigger (David Rohr, Frankfurt)
- Reconstruction of Micropattern Detector Signals using Convolutional Neural Networks (Lucie Flekova, Darmstadt)
- Message Queues for Online Reconstruction on the Example of the PANDA Experiment (Tobias Stockmanns, FZJ)
- FPGA based data processing in the ALICE High-Level Trigger in LHC Run 2 (Heiko Engel, Frankfurt)
- Online computing architecture for the CBM experiment at FAIR (Jan de Cuveland, Frankfurt)
- Support for Online Calibration in the ALICE HLT Framework (Mikolaj Krzewicki, Frankfurt)

German SW talks at CHEP 3/3

- RootJS: Node.js Bindings for ROOT 6 (Marek Szuba, KIT)
- JavaScript ROOT v4 (Sergey Linev, GSI)
- Experiment Software and Projects on the Web with VISPA (Robert Fischer, RWTH)
- Experience of Google's latest Deep Learning library, TensorFlow, with Docker in a WLCG cluster (Gerhard Ferdinand Rzehorz, Goettingen)
- PODIO - Applying plain-old-data for defining physics data models (Frank Gaede, DESY)