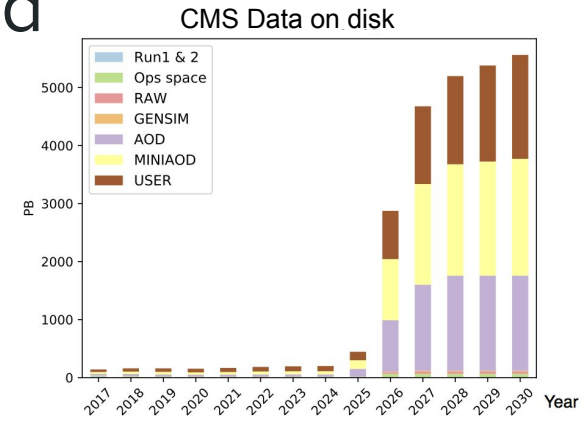
Software Working Group Introduction

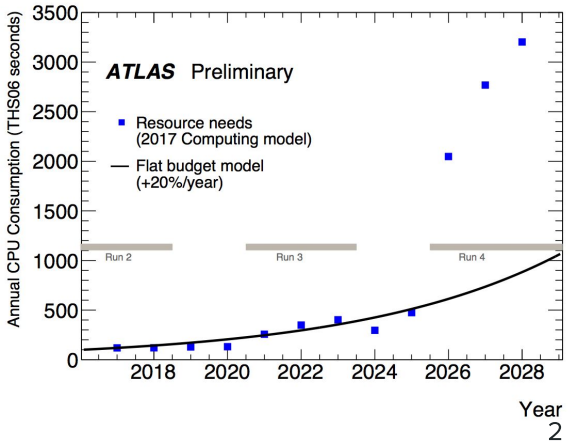
Graeme Stewart (co-convenor Jakob Blomer)

2018-03-16

Challenges to the HL-LHC and beyond

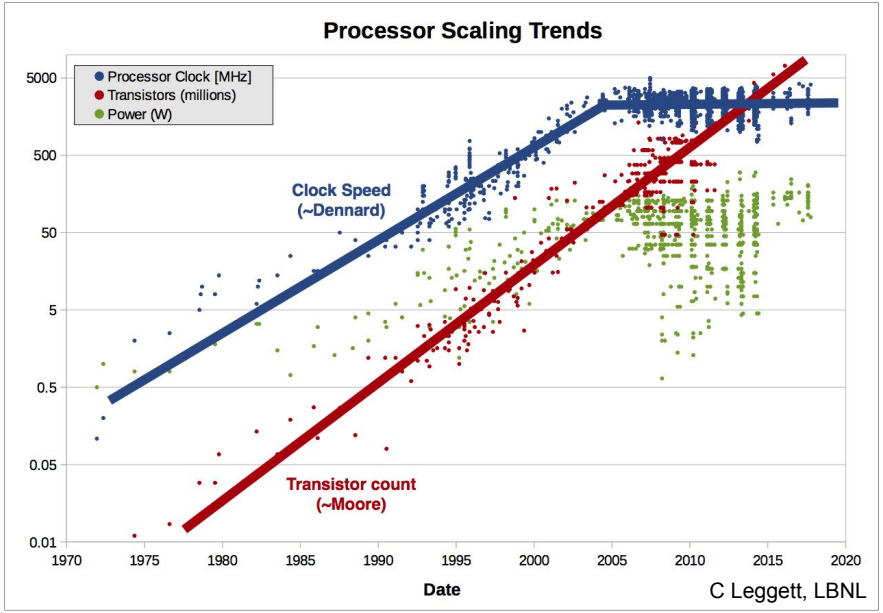
- High-Luminosity LHC is far from being a solved problem for software and computing
  - Naive extrapolation from today is not affordable
- Beyond HL-LHC, there are a number of different options for new machines
  - Lepton colliders (ILC, CLIC, FCC-ee) have overall less serious computing challenges
    - Require performant, robust, easy to use/deploy software
  - Hadron colliders (HE-LHC, FCC-hh) bring a massive data rate and complexity problem
    - Extreme for everything: generators, simulation, reconstruction, analysis
- Whatever the future, we pass through the HL-LHC on the way
  - HEP Software Foundation Community White Paper maps out that path





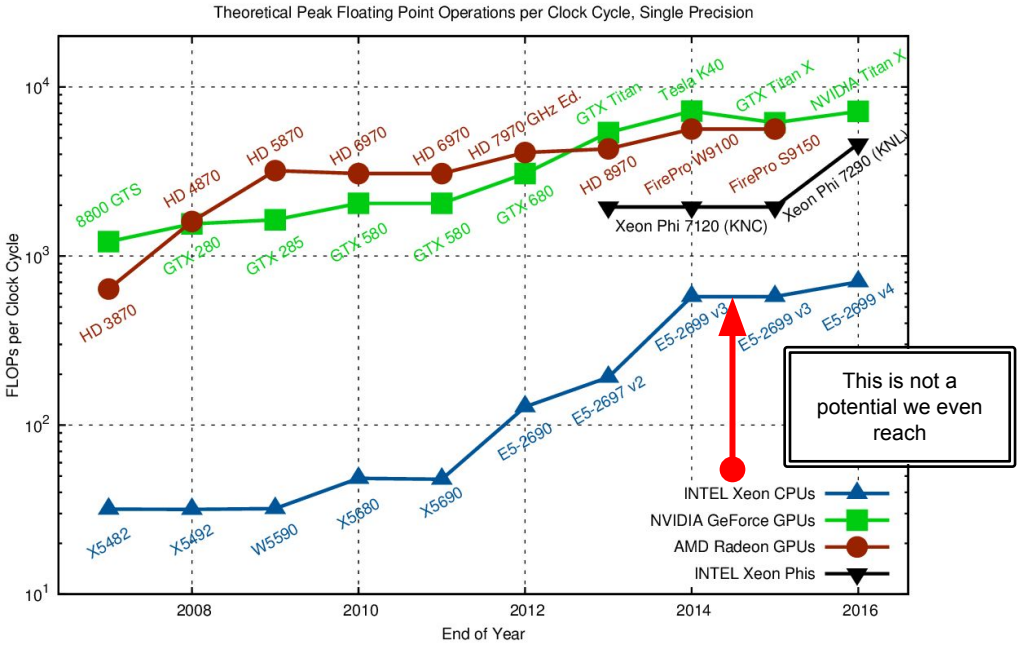
Processor evolution

- Moore's Law continues to deliver increases in transistor density
  - Doubling time is lengthening
- Clock speed increases stopped around 2006
  - No longer possible to ramp the clock speed as process size shrinks (Dennard scaling failed)
- So we are basically stuck at ~3GHz clocks from the underlying Wm^{-2} limit
  - This is the *Power Wall*
  - Limits the capabilities of serial processing
  - CPU based concurrency still in development for Run 3



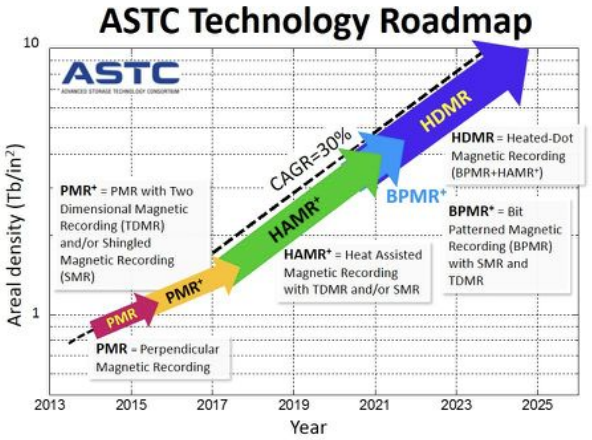
Compute Accelerators

- Most of the CPU die goes to things other than doing maths
  - Even CPU vector registers are hard for us to exploit
- Accelerators have a different model
  - Many cores, high floating point throughput, but lose a lot of ‘ease of use’
- We have to adapt to maintain our ability to use processors effectively

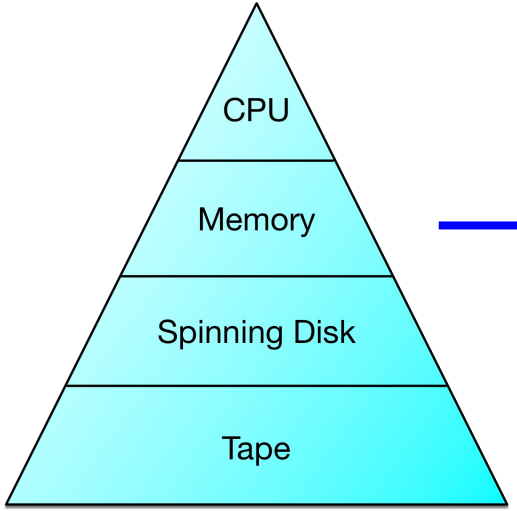


Other Technology Trends

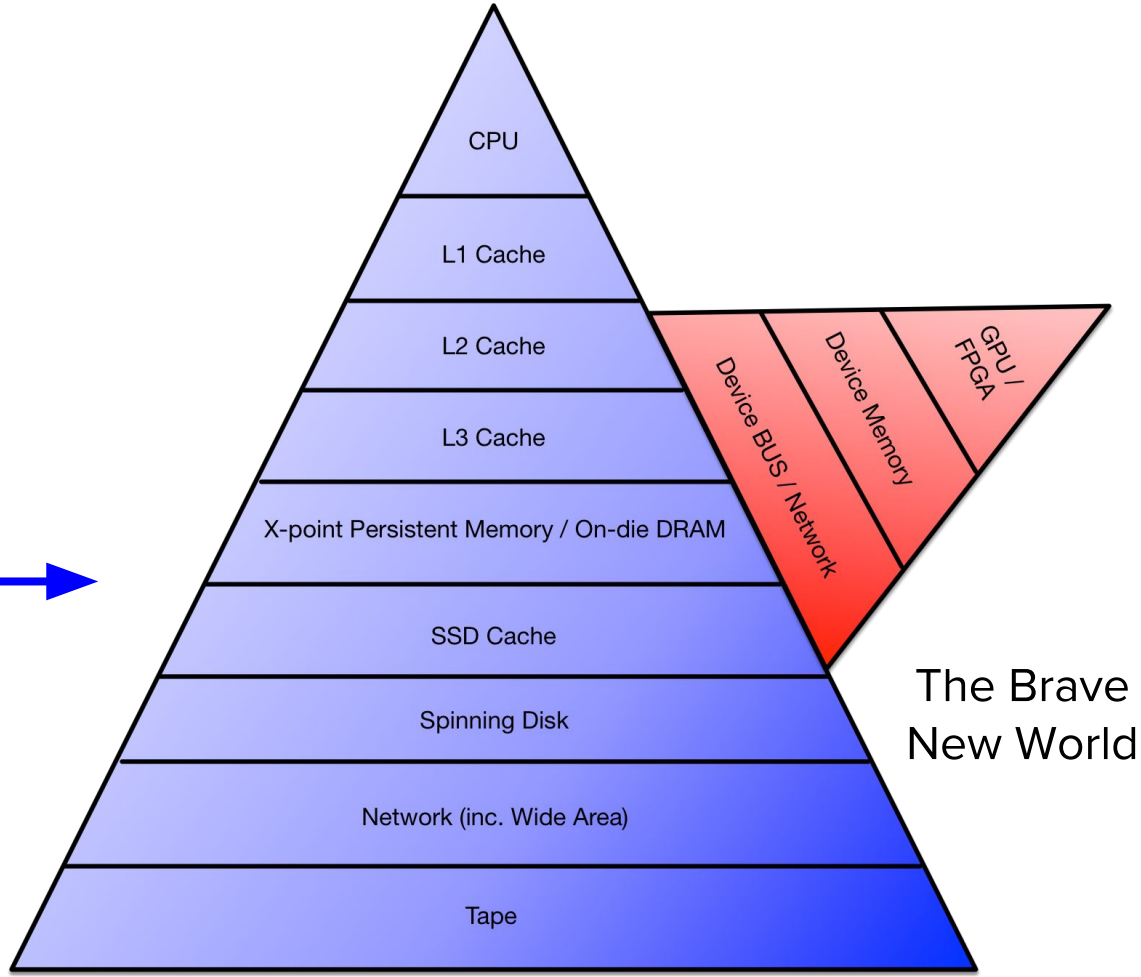
- Memory
  - DRAM improvements now modest
  - Overall, memory ‘landscape’ becomes more complex
    - Memory/storage boundary blurring
- Storage
  - Spinning disk capacity keeps climbing
    - Time to read and cost improves, but slowly
  - SSDs can read much faster, but price remains too high for bulk storage
  - Tape remains cheap to buy, slow to access with few companies left, O(1)
- Networks
  - Capacity increases expected to continue, latency will not change
  - Next generation networks offer capability to open channels between sites on demand
    - Useful, but an additional complexity
- Note: Game changer technologies might appear, but we cannot count on them



Meaning...



The Good Old Days



The Brave New World

Software needs and challenges

Goal: **ambitious** and **focused** work programme with milestones, deliverables and resource estimates

- Evolution and management of massive code bases created over many years
  - Current software is the base from which we design future detectors
- Meet the software challenges of future experiments
  - Very complex events - hard for reconstruction in particular
  - High rates - efficient, high speed data reduction pipelines
  - Huge volume - massive scale data and processing management
- Landscape for software becomes more varied
  - No more ‘free lunch’ from Moore’s Law
  - Harder to exploit hardware - need to adapt to accelerators and deep technology stack for data flow
- Advances from other fields offer promise, but need adapted
  - Data science and concurrency tools
- These are not problems that can be solved without investment
  - Software R&D program, running alongside detector R&D itself
  - Expect 5 years for advanced prototypes, deployment in 10 years

Software working group

- Open process
  - Gather ideas from the whole of the HEP software and computing community
  - Ensure alignment with developments outside CERN EP
  - 100 people on the mailing list
- Lightning Talks
  - Two sessions of lightning talks [1, 2] - open to anyone to propose a topic
  - Total of 28 short talks presented and discussed
    - Speakers from CERN EP and beyond
- Core group
  - Formed to distill these ideas and guide us towards R&D proposals
  - 15 people (LHC Experiments, CLiC, FCC, SFT, CERN IT)

Lightning talks

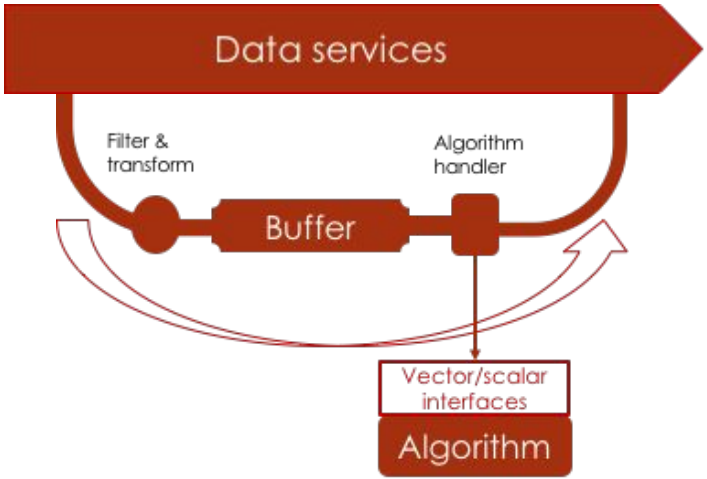
- Simulation for future experiments
- Reconstruction challenges for trackers and calorimeters
- New scalable analysis models
- Applied machine learning
- Tools for concurrency on heterogeneous resources
- Exabyte data flow and data management
- Support for new architectures and SoC systems
- Software integration

Big thanks to all the contributors!

		Challenge Area						Solution Technology						
Title	Author	Simulation	TDAQ	Reconstruction	Analysis	Facility	Other	Concurrency and Accelerators	Machine Learing	Caching Layers	Advanced Networks	Storage	Turnkey	Other
New Approach for ATLAS Detector Geometry Modelling	Alexander Sharmazanashvili						x							
Software Defined and Named Data Networking, and the Consistent Operations Paradigm	Harvey Newman					x					x			
VectorFlow, a subscription-based vector data-shipping service	Andrei Gheata	x		x				x						
Future technologies for efficient and light-weight handling of non event data	Gianluca Cerminara					x	x			x				
Filesystem-less high performance I/O of HEP Data	Danilo Piparo				x	x						x		
Track Reconstruction in a Concurrent world	Andreas Salzburger			x				x						
Machine Learning technologies applied to online event selection	Maurizio Pierini		x						x					
Machine Learning for Fast Simulation	Sofia Vallecorsa	x							x					
Automation of Data quality and certification with Deep Learning	Giovanni Franzoni		x						x					
Software for Detector Optimisation Studies in EP-LCD	Marko Petric						x						x	
Turnkey software solutions	Benedikt Hegner						x						x	
High throughput data analysis on future heterogeneous platforms	Danilo Piparo				x	x		x						
Heterogeneous computing	Felice Pantaleo					x		x						
Browser as a platform / compute device	Jakob Blomer						x							x
Machine Learning to empower physics modeling	Marilena Bandieramonte	x							x					
Handling inference efficiently in online/offline reconstruction software	Vincenzo Innocente			x		x			x					
Preparing for our Exascale data management challenge	Mario Lassnig					x				x	x	x		
Rethinking Data Center Computing	Giulio Eulisse			x	x	x		x						x
Embedded Linux for Run Control and ARM porting	Ralf Spiwoks		x										x	
Example of an embedded Linux Operating System and System-On-Chip based platform in the DAQ use case	Adrian Fiergolski		x										x	
To boldly go… - Application deployment frameworks for new computing environments	Radu Popescu					x								x
Future Distributed Analysis of HEP Data	Enric Tejedor Saavedra				x							x	x	
Machine learning solutions for simulation and reconstruction of highly granular calorimeters	Jan Kieseler			x				x						
APPAVO - Augmented tools for Particle Physics Analysis, Visualisation and Outreach	Eduardo Rodrigues				x		x							x
Development of a histogram library for concurrent programming	Andrea Bocci		x					x						
Clustering and Tracking with GPUs and machine learning	David Rohr			x				x	x					
Browser Based Detector Displays for Outreach & Education	Alexander Sharmazanashvili						x							x
tkLayout	Stefano Mersi						x						x	

One Lightning Example - VectorFlow

- Vectorisation is great when it happens, but difficult to achieve with our codes today
  - Problem is how to gather appropriate data and fill vector registers with it
    - Does not happen naturally for event by event processing
  - Gather data into a processing buffer from many places
  - Process through an algorithm that has a vectorised interface
- R&D on
  - Concurrency and performance effectiveness
  - Using vectorisation primitives in new areas
  - Integration into existing frameworks
  - Offloading into accelerators
- Adapting to new hardware is difficult work and dedicated expertise is needed to bridge between physics and software



EP department - a centre for software excellence

- Very strong software groups in current and future experiments
  - Including Phase II upgrades
  - Frameworks, Tracking, Data Quality, DAQ and Data Flow, Databases, Detector Description
- Key contributor to core HEP libraries
  - ROOT
  - Geant4
- Central role in distributed data management software and operations
  - Data Management and Workload Management
  - Large resource operations management: Trigger Farms, Tier-0s
  - CVMFS
- Close to CERN IT
  - WLCG operations and developments
  - Critical expertise in technology tracking
- Key player in community initiatives through HEP Software Foundation

Many projects now picked up more widely than at LHC - a success!

Today’s talks

- Future Tracking
  - A key unsolved problem for future experiments
- Machine Learning
  - Applied data science and how it can be used in HEP
- End to End Physics
  - Software and data flow solutions for the exabyte era

These are not presented here as concrete R&D proposals, but as samples of some of the most interesting challenges and ideas from the problem and solution space

Backup

Software Working Group Core Team

Jakob Blomer (Convener)	Danilo Piparo (ROOT, Concurrency)
Graeme Stewart (Convener)	Witek Pokorski (Geant, Generators)
Marco Cattaneo (LHCb)	Radu Popescu (Other languages)
Dirk Duellmann (IT expertise and link)	André Sailer (CLiC, LCD)
Benedikt Hegner (FCC)	Andreas Salzburger (ATLAS, FCC, Tracking)
Mario Lassnig (ATLAS, Data Management)	Niko Neufeld (LHCb, DAQ, FPGAs)
Maurizio Pierini (CMS, Machine Learning)	David Rohr (ALICE, GPUs)
Helge Meinhard (IT R&D)	