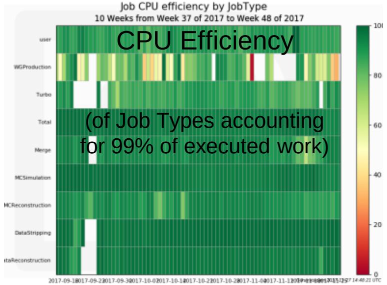
LHCb Software & Computing

Stefan Roiser

LHCC/WLCG Referees

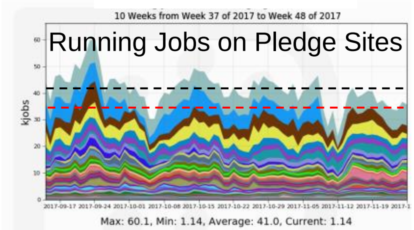
29 November 2017

Operations

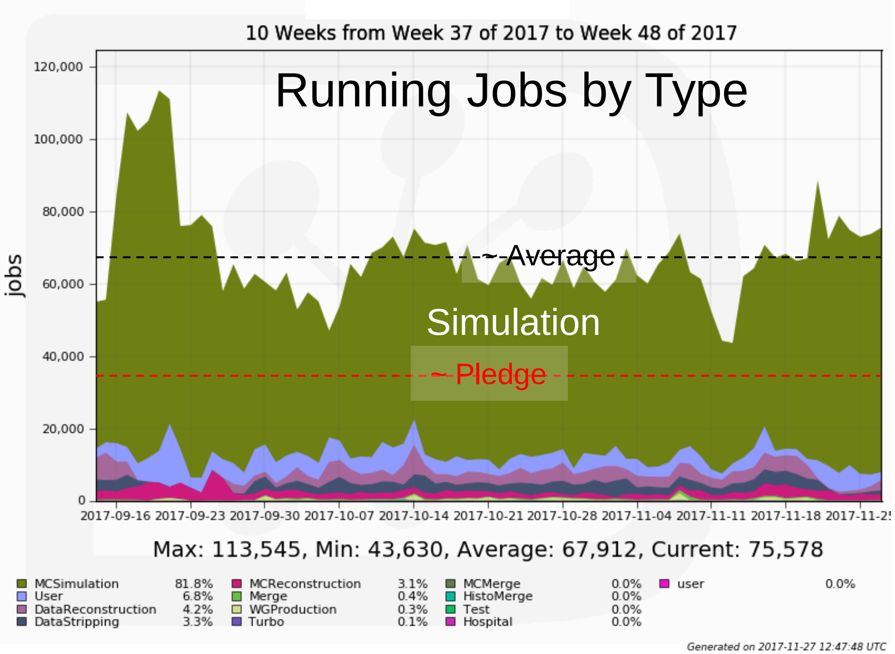


CPU Efficiency

(of Job Types accounting for 99% of executed work)



Running Jobs on Pledge Sites

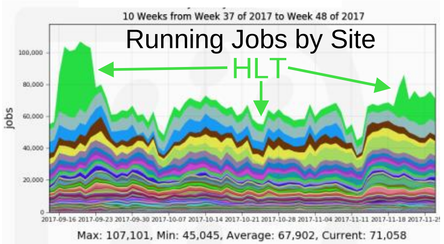


Running Jobs by Type

~ Average

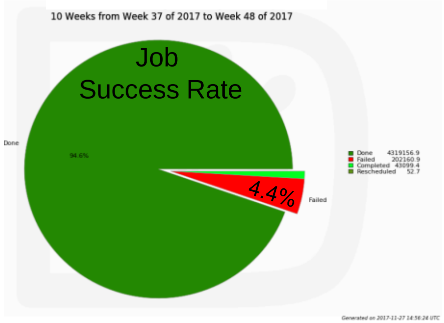
Simulation

~ Pledge



Running Jobs by Site

HLT



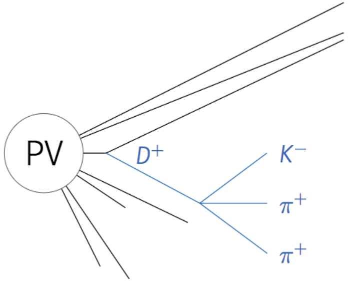
Job Success Rate

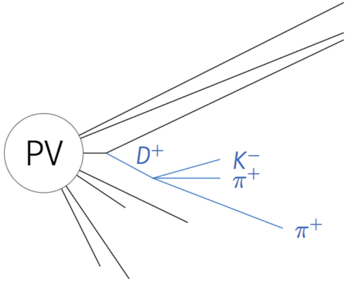
4.4%

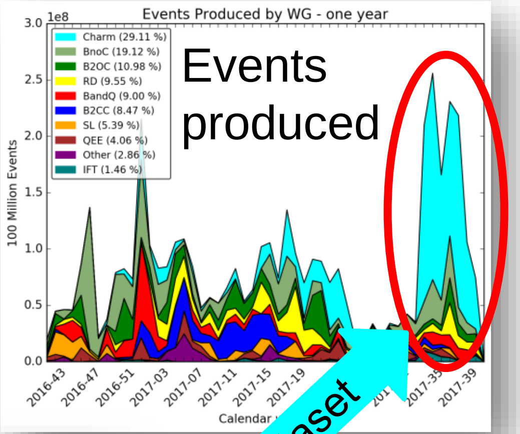
- Work on distributed computing resources dominated by MC Simulation
- Usage on all resources ~ double of pledges. Use of pledged resources ~ +17 %
- Peaks in running jobs b/c of HLT farm available. Full usage again during YETS
- Continuous high CPU efficiency and job success rates also during this period

Further Optimizations of Simulation Workflows

- New Re-decay fast simulation in production
  - Re-use underlying event multiple times with only re-generating the signal information
  - Major CPU savings between a factor 10 and 50 depending on the production configuration
  - More fast simulation options being worked on

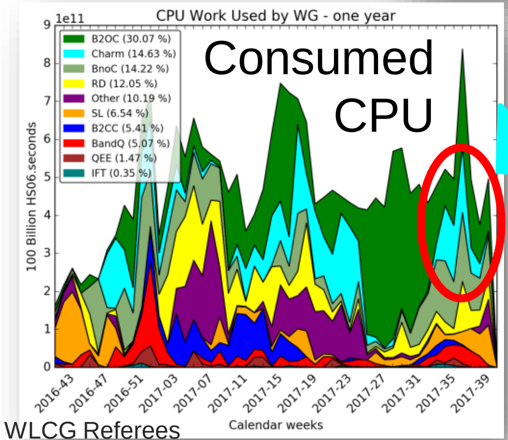






Events produced

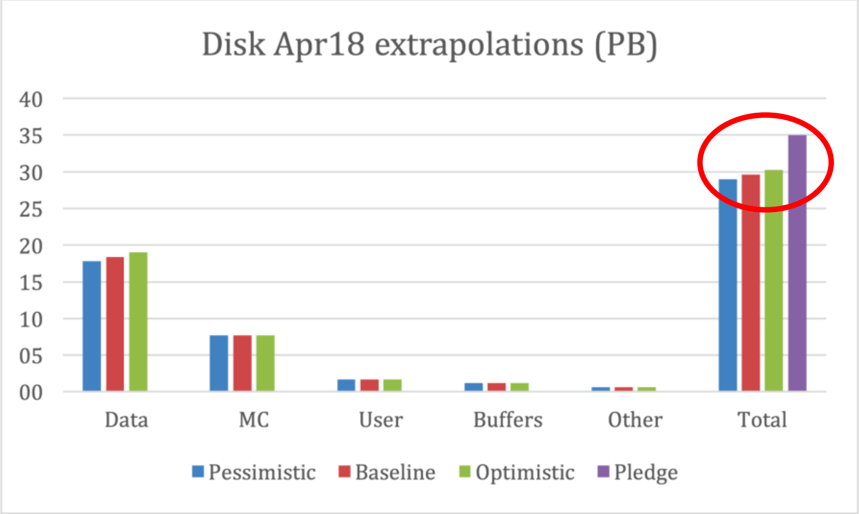
same dataset

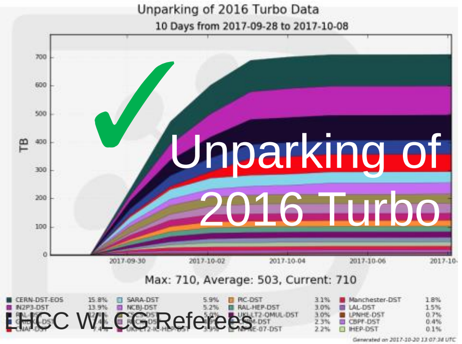


Consumed CPU

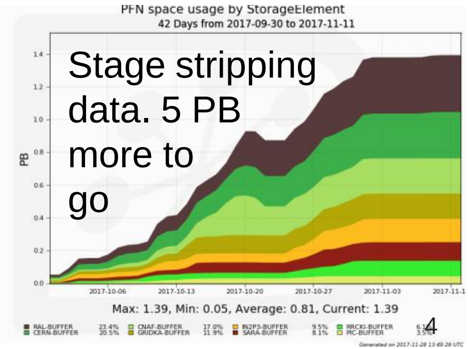
Remaining 2017 Resources and 2018 Forecast

- Lower disk usage in 2017 as effect of less LHC stable beam time before the September technical stop than foreseen. Re-use available disk for:
  - Unparking “Turbo” data to disk that needed to be put on tape in 2016 → done
  - Stage input data for re-stripping of Run 2 data → ongoing
- As consequence reduction of 2018 disk and tape requests


Disk Apr18 extrapolations (PB)


Unparking of 2016 Turbo Data
10 Days from 2017-09-28 to 2017-10-08
Max: 710, Average: 503, Current: 710

✓ Unparking of 2016 Turbo


PFN space usage by StorageElement
42 Days from 2017-09-30 to 2017-11-11
Max: 1.39, Min: 0.05, Average: 0.81, Current: 1.39

Stage stripping data. 5 PB more to go

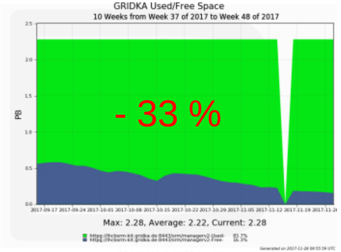
Operational Incidents

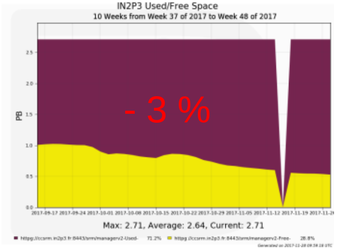
- Unfortunate outage of CNAF coincides with start of re-stripping campaign of Run 2 data
    - Possible loss of reconstruction output, which is single copy
    - Re-stripping campaign finalization delayed until site is back in operation
- Recently notified of damage of 5 tape cartridges at GRIDKA
    - Includes loss of 547 reconstruction output files from both Run1 and Run2

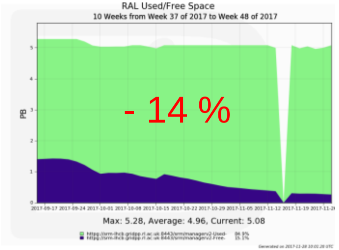
- Reconstruction output data needs to be re-reconstructed
    - Reco output single copy defined in computing model after CRSG/RRB discussion
    - Re-reconstruction of individual files is operationally expensive

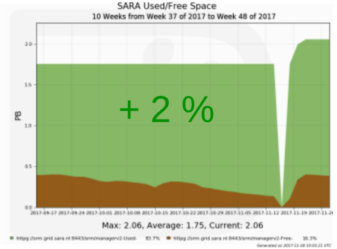
T1 Storage Situation

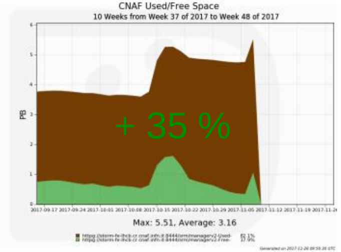
- Comparing total space deployed vs 2017 pledges
  - Percentage is over/under pledging of all disk space tokens per T1 site

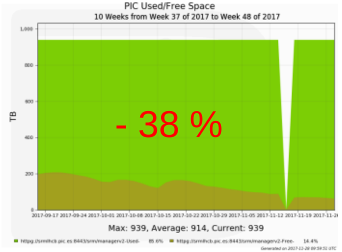


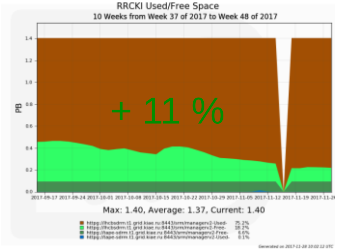










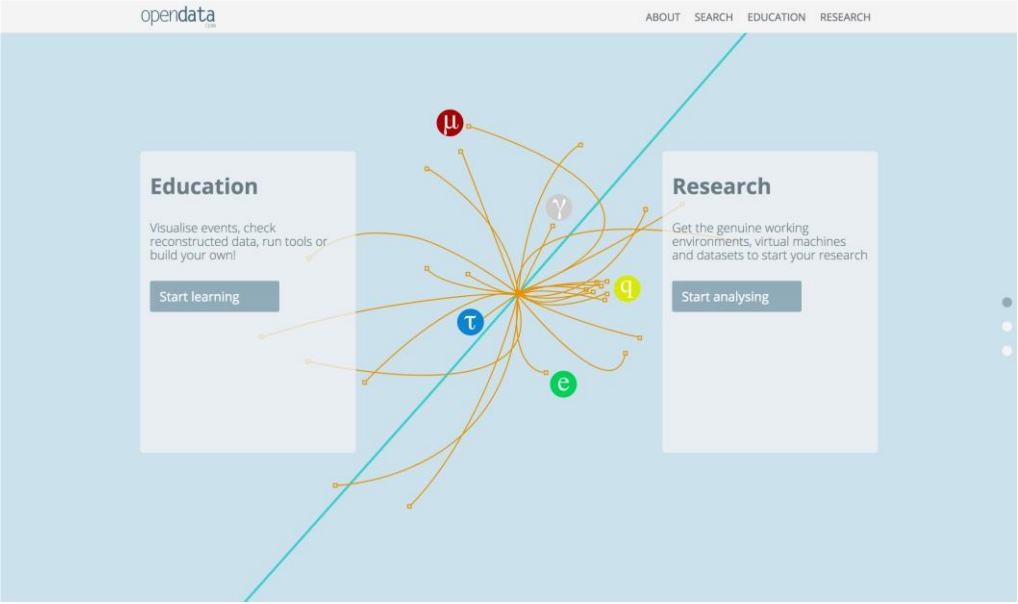


- Overall situation is ok but as data processing is planned on basis of pledged resources problems at individual sites
  - Currently running low on free space
  - Operational mitigations measures needed towards end of the year data processing

Open Data

- Several individual example analysis and outreach data sets have been released via the CERN open data portal

- Next step, plan to release substantial amounts of Run 1 data via the open data portal in the order of several 100 TBs



Education

Visualise events, check reconstructed data, run tools or build your own!

Start learning

Research

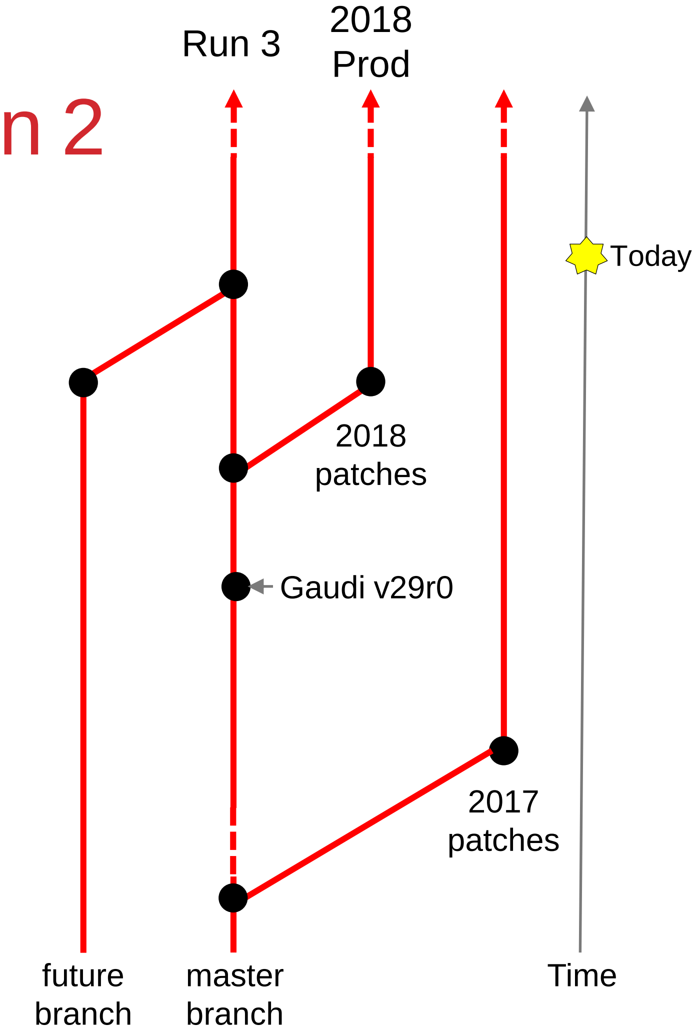
Get the genuine working environments, virtual machines and datasets to start your research

Start analysing

- Understanding that resources needed for these activities will be provided outside the experiment pledged resources

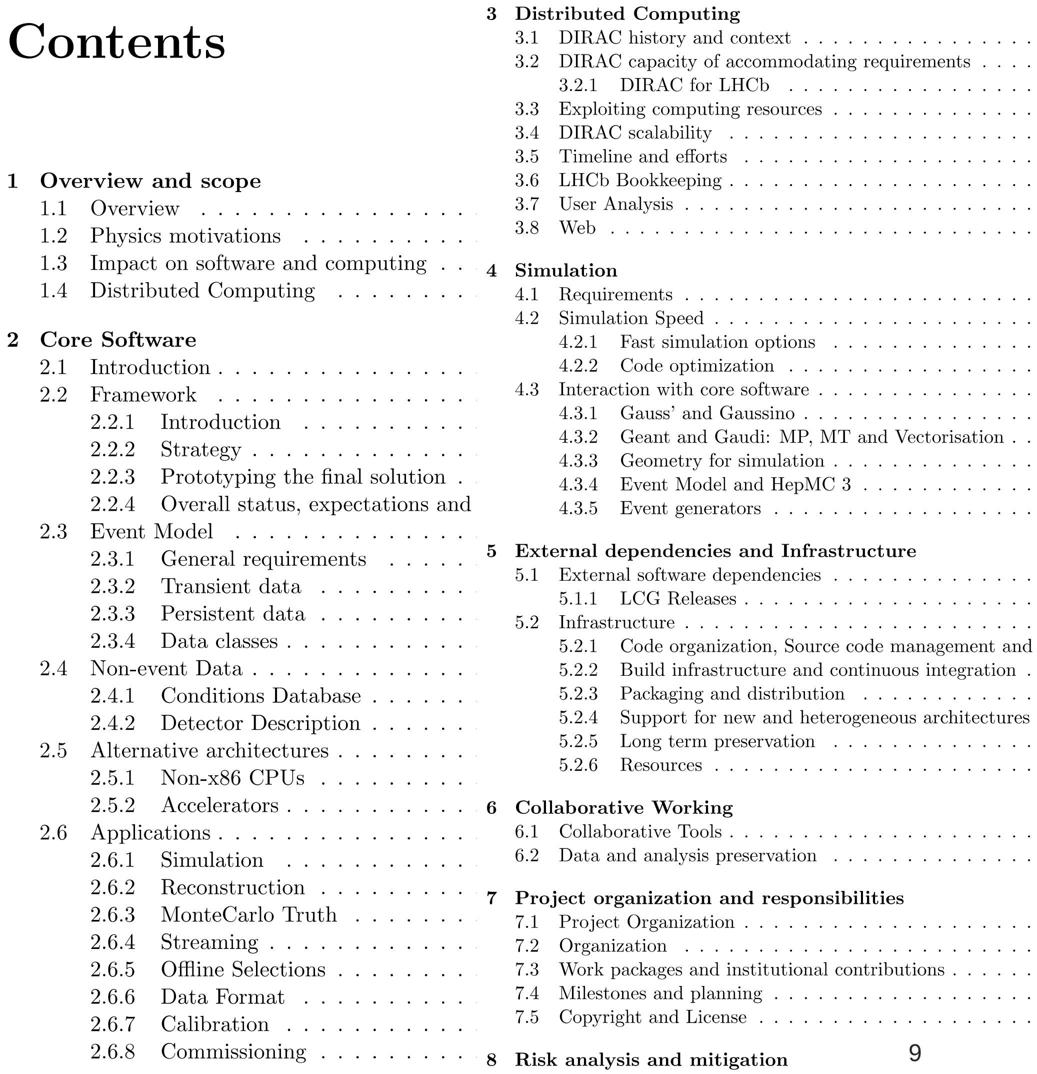
Software Development Beyond Run 2

- Branch for 2018 production has been done
- Currently merging two main branches
  - Future: software engineering work for the upgrade
  - Master: main branch for physics developments
- Master will become common branch for engineering and physics developments towards Run 3
  - Allows better for engineers / physicists pairs working on further code improvements
  - Regular releases to be cut for MC upgrade productions and HLT integration tests



TDR

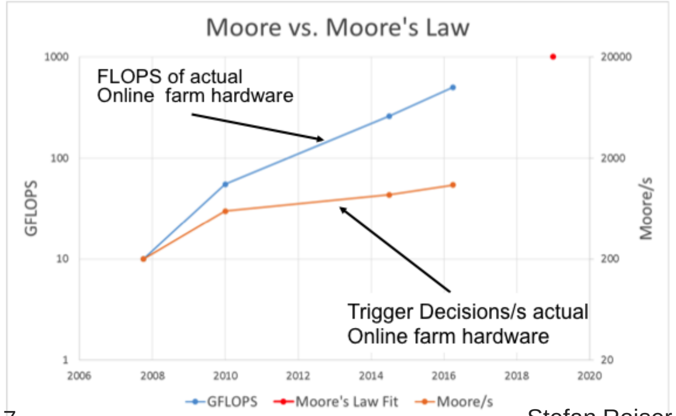
- TDR being finalised
  - Hot phase of writing during September/October
  - Each chapter states relevant issues, discusses technology choices or options, describes implementation plan
- Computing workshop 20-24 Nov
  - Fully dedicated to software & computing upgrade
  - Review and final editing of the TDR
- Deadline for TDR is Q4/2017



Contents

Overall Software Strategy

- Our software is not exploiting current hardware capabilities
  - 15-years-old single-thread, sequential-event processing framework
  - Cache misses are increasingly a problem



© Rainer Schwemmer

- Redesign Gaudi software framework along:
  - Better use of processors: exploit vector registers with SIMD instructions
  - Reduce memory footprint, enabling new architectures: multi-threading, task-parallel execution

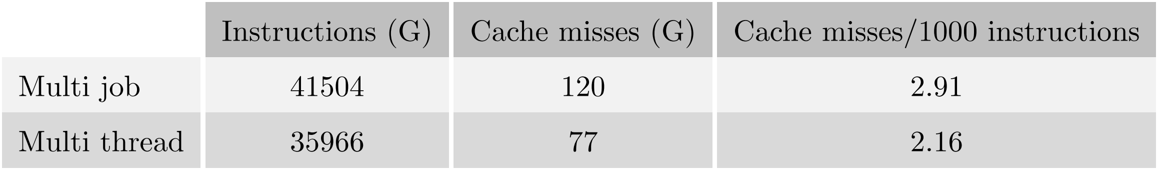
- Also use up-to-date software for currently unmaintained and non thread-safe packages for:
  - Conditions database
  - Detector description

Core Framework Status

- Based on Gaudi (common LHCb+ATLAS project) and its Gaudi-Hive extension

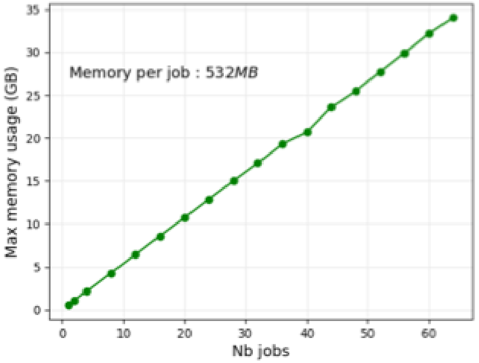
- Multi-threaded framework is ready
  - More than 100 algorithms, including the full HLT1 sequence, have been converted

- Large memory gain with respect to multi-job resulting in less cache misses

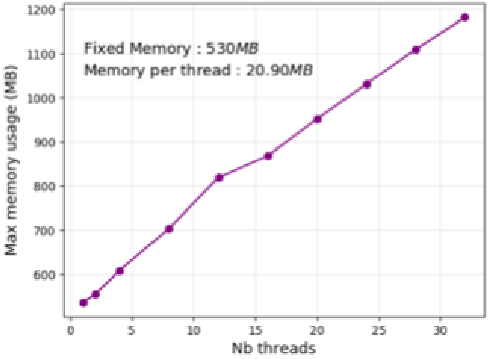


	Instructions (G)	Cache misses (G)	Cache misses/1000 instructions
Multi job	41504	120	2.91
Multi thread	35966	77	2.16

Multi-process

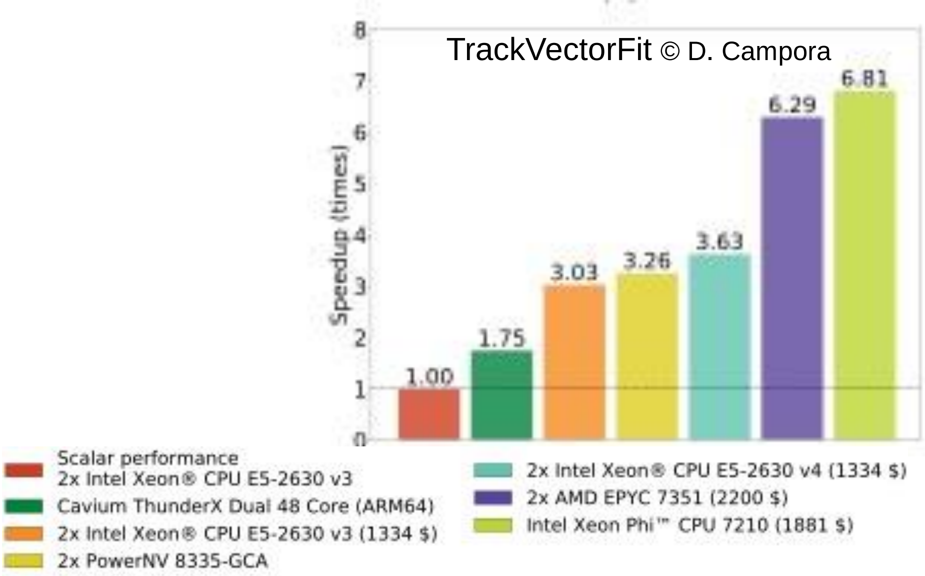


Multi-thread



Code Vectorization

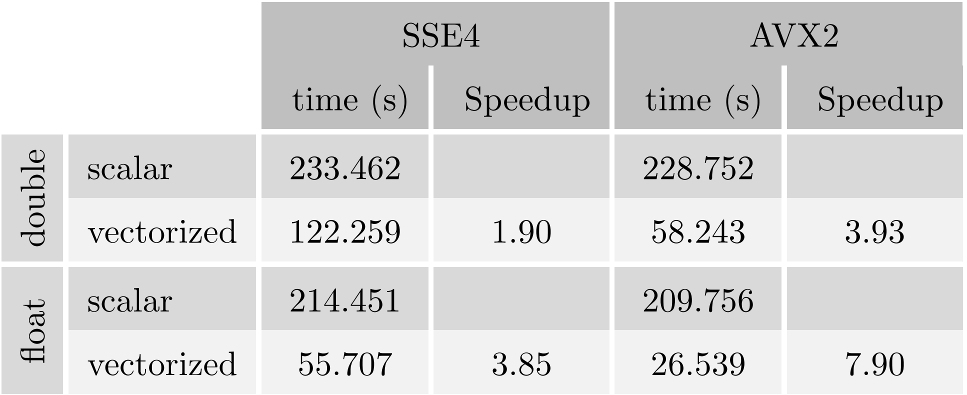
- Example: vectorization of the Kalman fit
  - Large speedup when using many-cores
  - Gain “washed out” by rearrangement of data structures
  - Work in progress



TrackVectorFit © D. Campora

- Example: vectorization of the RICH photon reconstruction
  - Shows perfect speedup via SIMD instructions

Timing and speedup of SIMD instructions © Chris R Jones



		SSE4		AVX2	
		time (s)	Speedup	time (s)	Speedup
double	scalar	233.462		228.752	
	vectorized	122.259	1.90	58.243	3.93
float	scalar	214.451		209.756	
	vectorized	55.707	3.85	26.539	7.90

More Areas of the Upgrade

- Distributed Computing
  - Evolution of the distributed computing environment (DIRAC)
  - Change user analysis model on distributed computing
    - Working Group productions by default but keep possibility of individual analysis jobs
- Simulation
  - Develop more fast simulation options (Delphes, Calo Shower, Particle Gun, …)
  - Experiment agnostic simulation framework based on LHCb/Gauss together with FCC
- Management
  - Defining “computing agreements” with 20 different institutes

- Computing Model for the Upgrade due by Q3/18
  - Difficult to fit within “flat budget” scenario
    - 5 GB/s HLT output * 5 Mio seconds LHC lifetime = 25 PB/year for data only

Summary

- Offline productions progressing well
- Re-decay fast simulation workflow deployed in production
- Additional load on operations team because of tape losses and disk unavailabilities

- Work towards the Run 3 upgrade in full swing
- Software framework ready, concentrating on algorithmic improvements
- Finalizing TDR for submission by the end of the year

Backup