Challenges and Opportunities in Software & Computing for Future Colliders

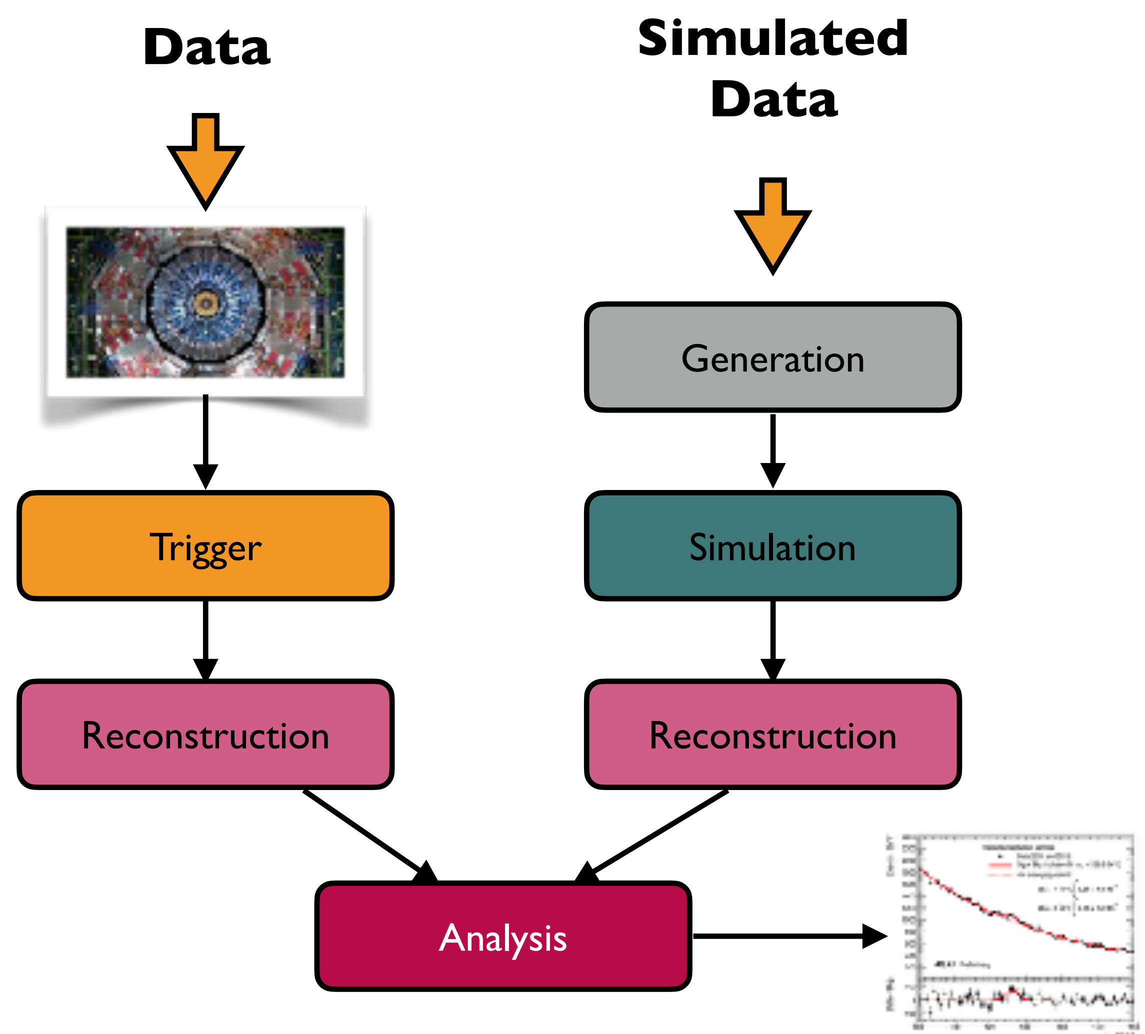
Heather M. Gray

Introduction

- Software and computing are used ever increasingly in high-energy physics during **every step** of the **data processing chain**
  - From detector control, through trigger, to reconstruction and analysis
- The (offline) **code base** is enormous
  - ~50M lines of C++
  - Also large (but size unknown) python code base

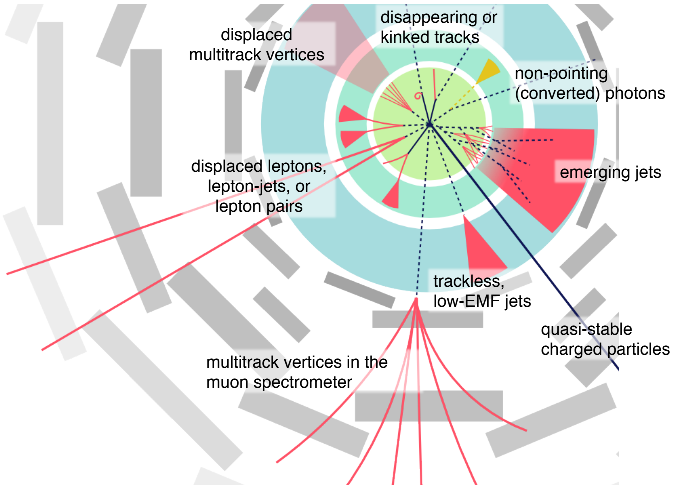
Typical data processing chain at the LHC

Also used to control accelerators & detectors



How may Future Colliders Differ

- Can group future colliders into two groups
  - Near-term: LHC upgrades (including HL-LHC)
  - Long-term: Future lepton colliders, potential hadron and muon colliders
- A number of features of these colliders induce challenges and opportunities for software and computing
  - Backgrounds: Increased pile up, beam-induced background
  - Increasingly sophisticated detectors
    - More channels, additional information
  - Higher data rates: better triggers (or no triggers)
  - Increasing demands in physics precision
  - Need to explore unconventional signatures



H. Russell

Challenges and Opportunities

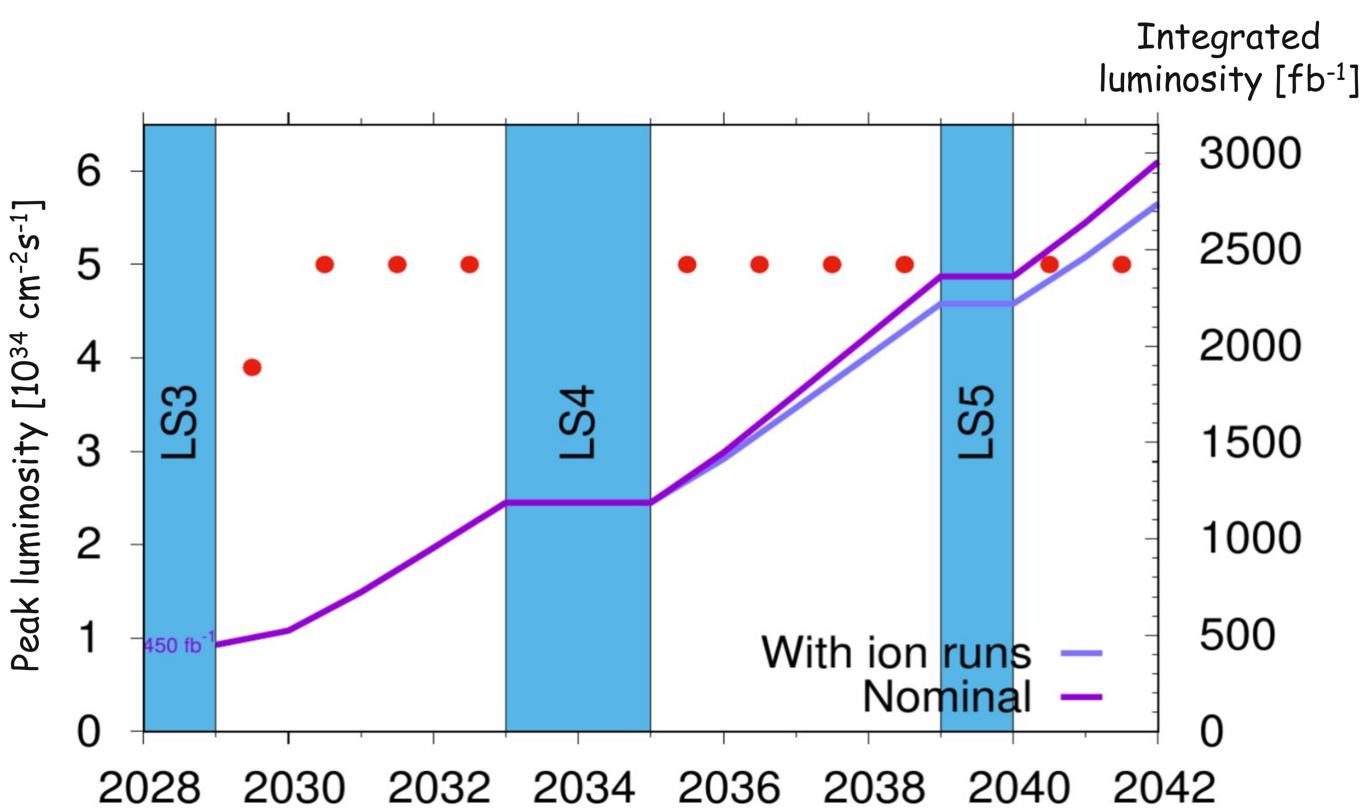
- Computing technology evolution
    - Increased concurrency
    - Increasingly diverse architectures
- Machine learning
- Data science, including python for scientific computing
- Open Source Software
- Funding constraints

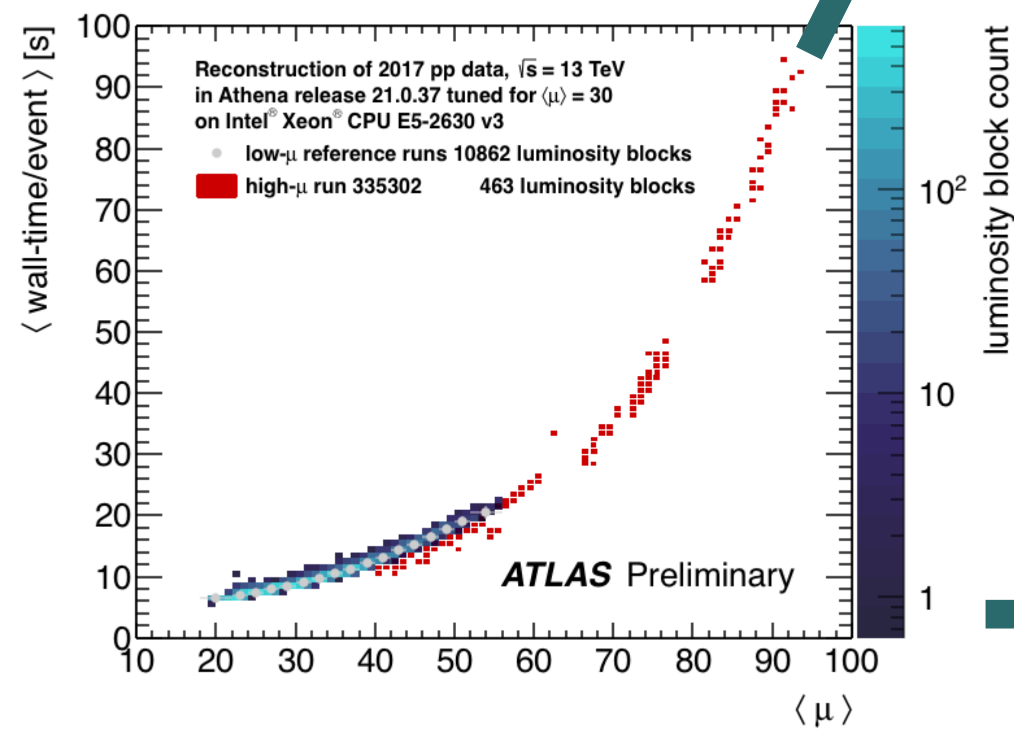
The goal of this talk is to explain the impact on these factors on software and computing to highlight the challenges and also provide some ideas about the opportunities

Characteristics

Backgrounds: Additional Interactions

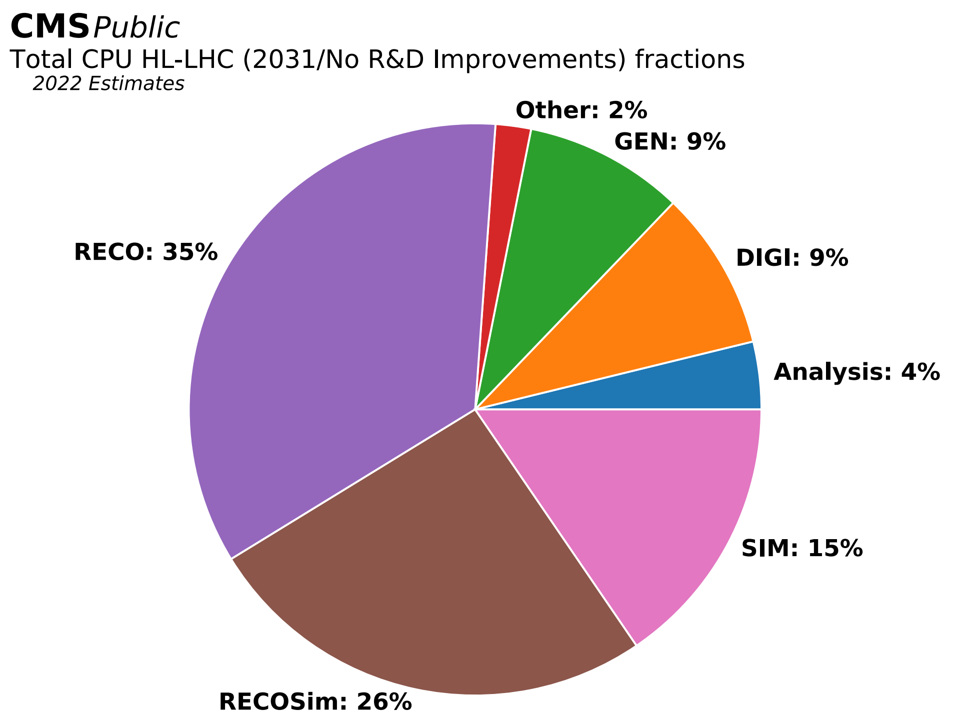
- At hadron colliders, each time two bunches of cross (or collide), multiple pairs of protons undergo inelastic collisions
- Mean number of interactions per bunch crossing or pile up (μ) is given by the following formula
  - $$\langle \mu \rangle = \frac{L \cdot \sigma_{\text{inel}}}{N_{\text{bunch}} \cdot f_{\text{acc}}}$$ depends linearly on the luminosity
- Track reconstruction algorithms scale quadratically with pile up

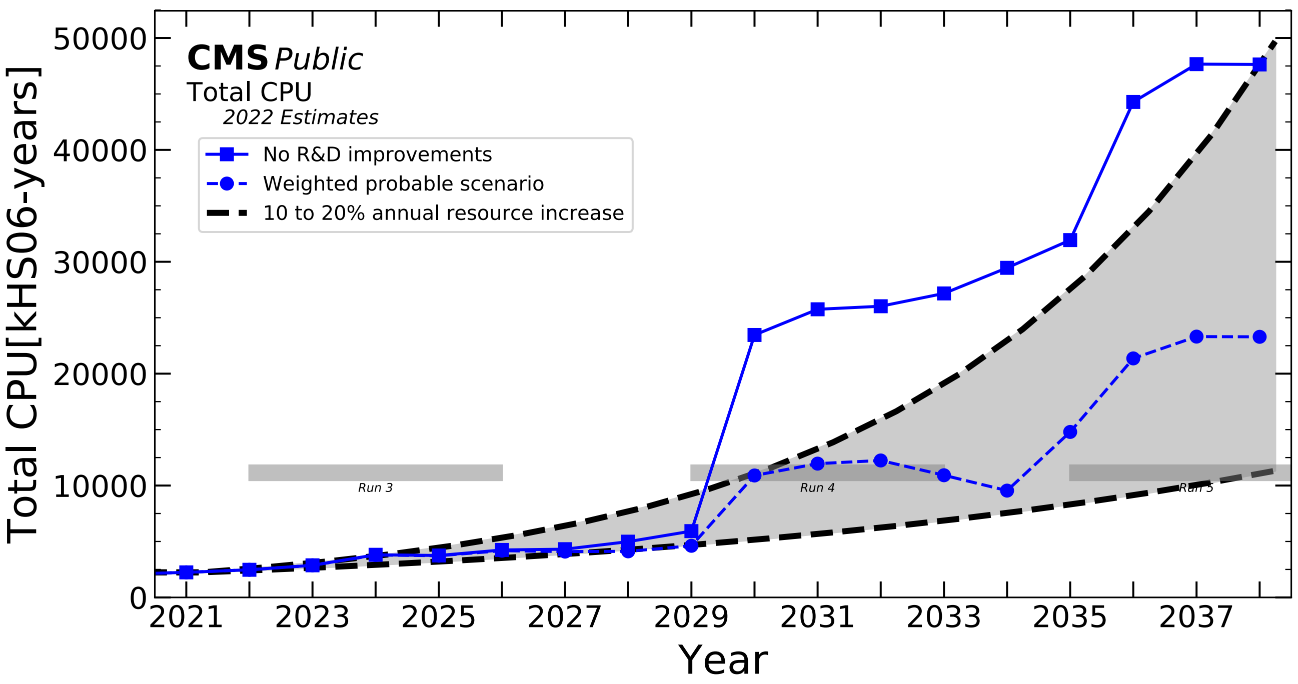




Tracking is a CPU Hog

- CPU demands of tracking are **significant**
  - **Largest** component of reconstruction
  - **Largest** component of CPU needs
- One component of the so-called "**LHC Computing Challenge**"
  - Mismatch between computing needs and resources
    - Depends strongly on **assumptions**
    - Target of **aggressive** software developments



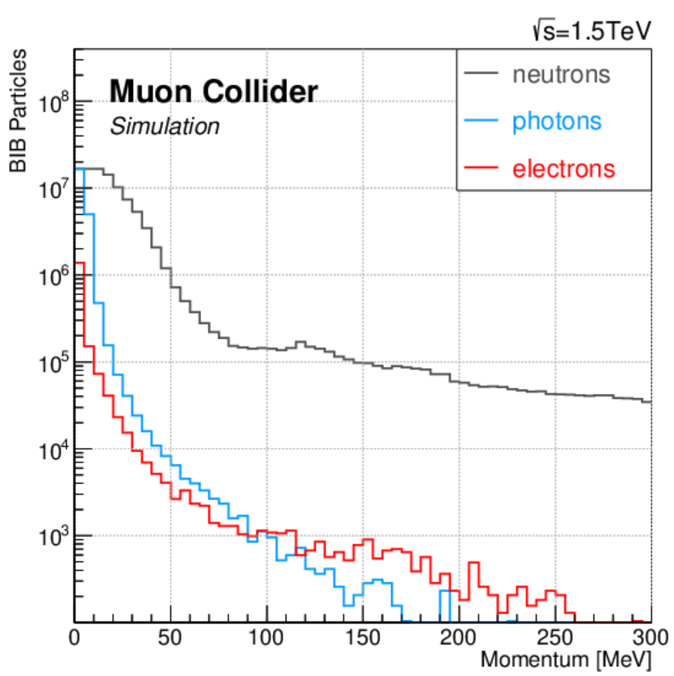


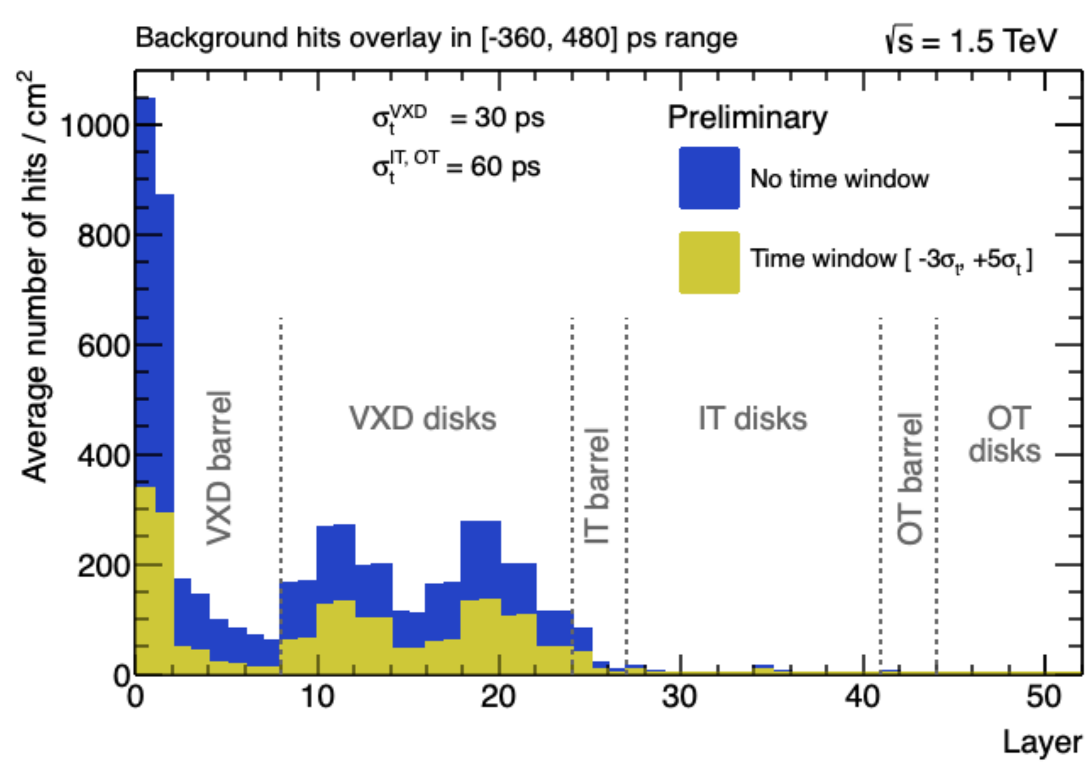
CMS O&C Public Results

Similar results for ATLAS

Beam-induced Background at Muon Colliders

- Muon colliders are susceptible to the background from the secondary and tertiary **muon decay** products
  - Reduced several orders of magnitude by the Machine Detector Interface (**MDI**)
- **10x hit density** from BIB at muon colliders in tracking detectors compared to pile-up at the HL-LHC
  - Similar impact on algorithms as from pile up





Credit

BIB drives MC Resource Needs

CPU

Approximate Size / Event [MB]			
heavily-filtered	trimmed	full truth	full truth (low threshold)
80	400	8,400	36,000

Disk

Full simulation	Approximate time/event [min]
Physics + BIB Overlay	1-2
BIB simulation	1500

Slide Credit

Sophisticated Tracking Detectors

- We'll discuss Moore's Law later, but one result is the increasing miniaturization of silicon components
  - Up 65x increase channels in silicon detectors when controlling for size
- More precise measurements, but larger data volume
- Timing adds extra dimension

ATLAS Inner Detector

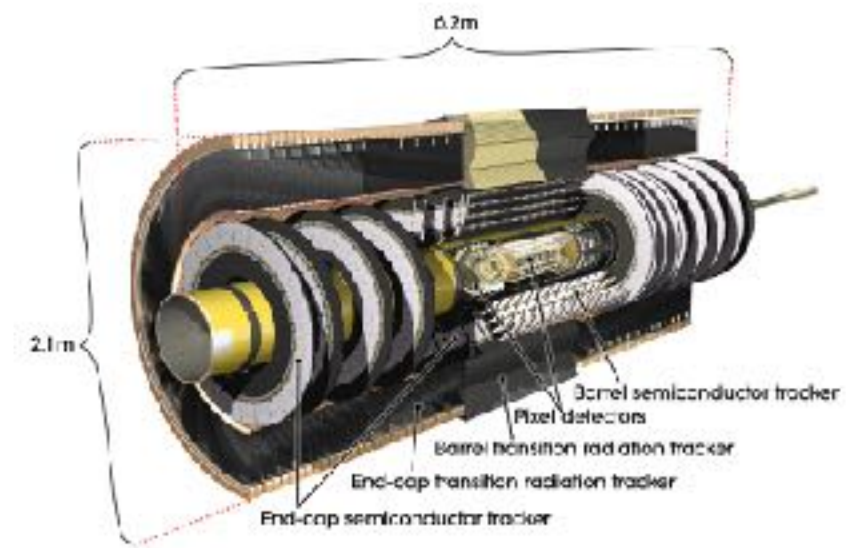


Image Source

ATLAS ITk

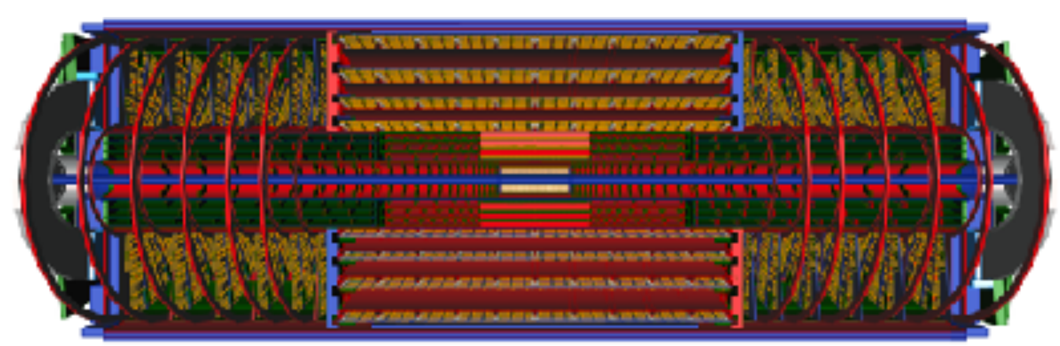


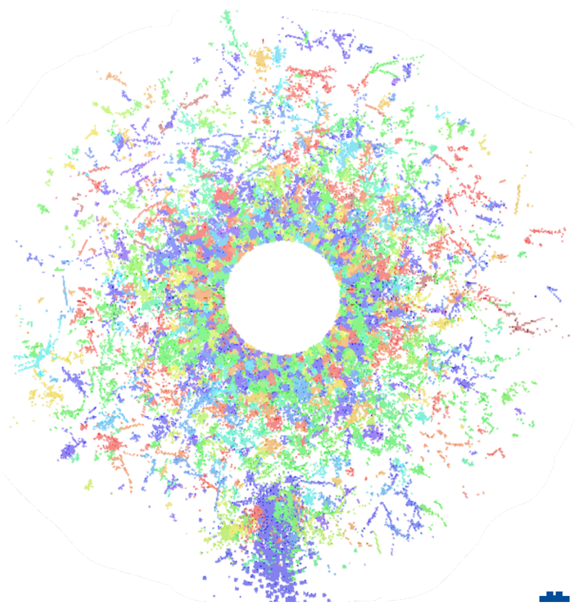
Image Source

	LHC	HL-LHC
ATLAS Pixel	80 (92) million	6 billion
ATLAS Strips	6 million	60 million

180 m^2 of silicon

Detailed Shower Reconstruction

- Another innovative use of silicon is in the CMS High Granularity Calorimeter (HGCAL) end-cap: high readout and high granularity
- 47 layer sampling calorimeter: silicon (26)/plastic scintillator
  - 6M silicon channels; 620m² silicon sensors
- Requires **new reconstruction algorithms** for reasonable computing resources requirement



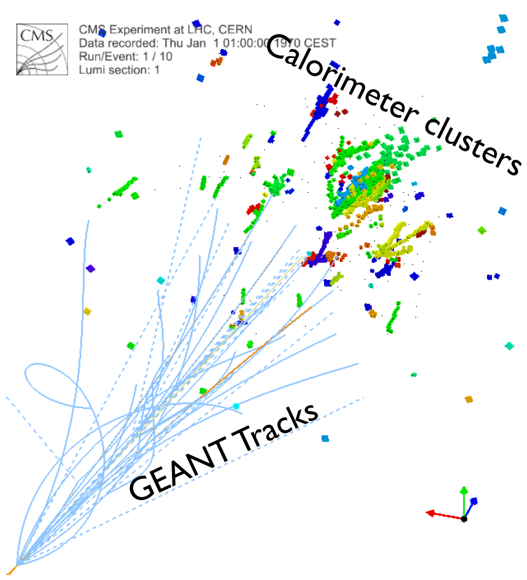


Image Source

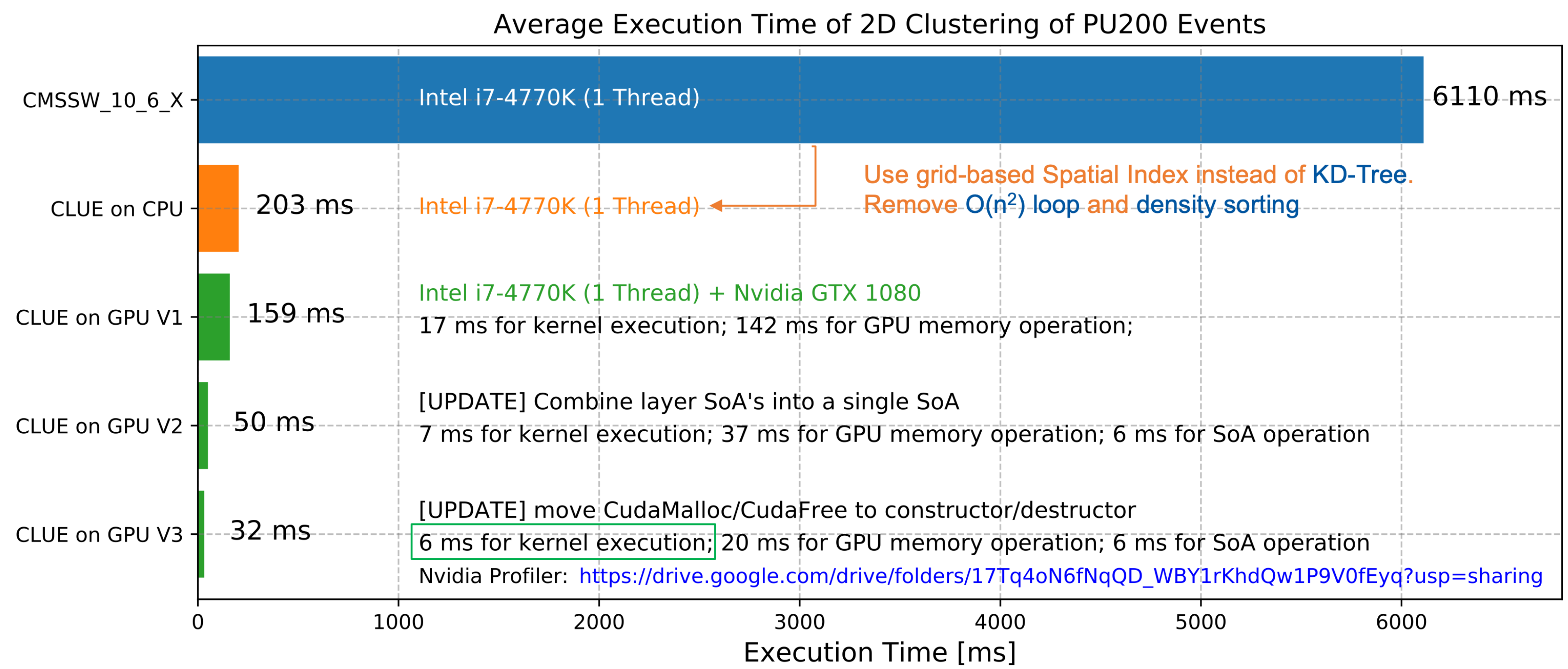


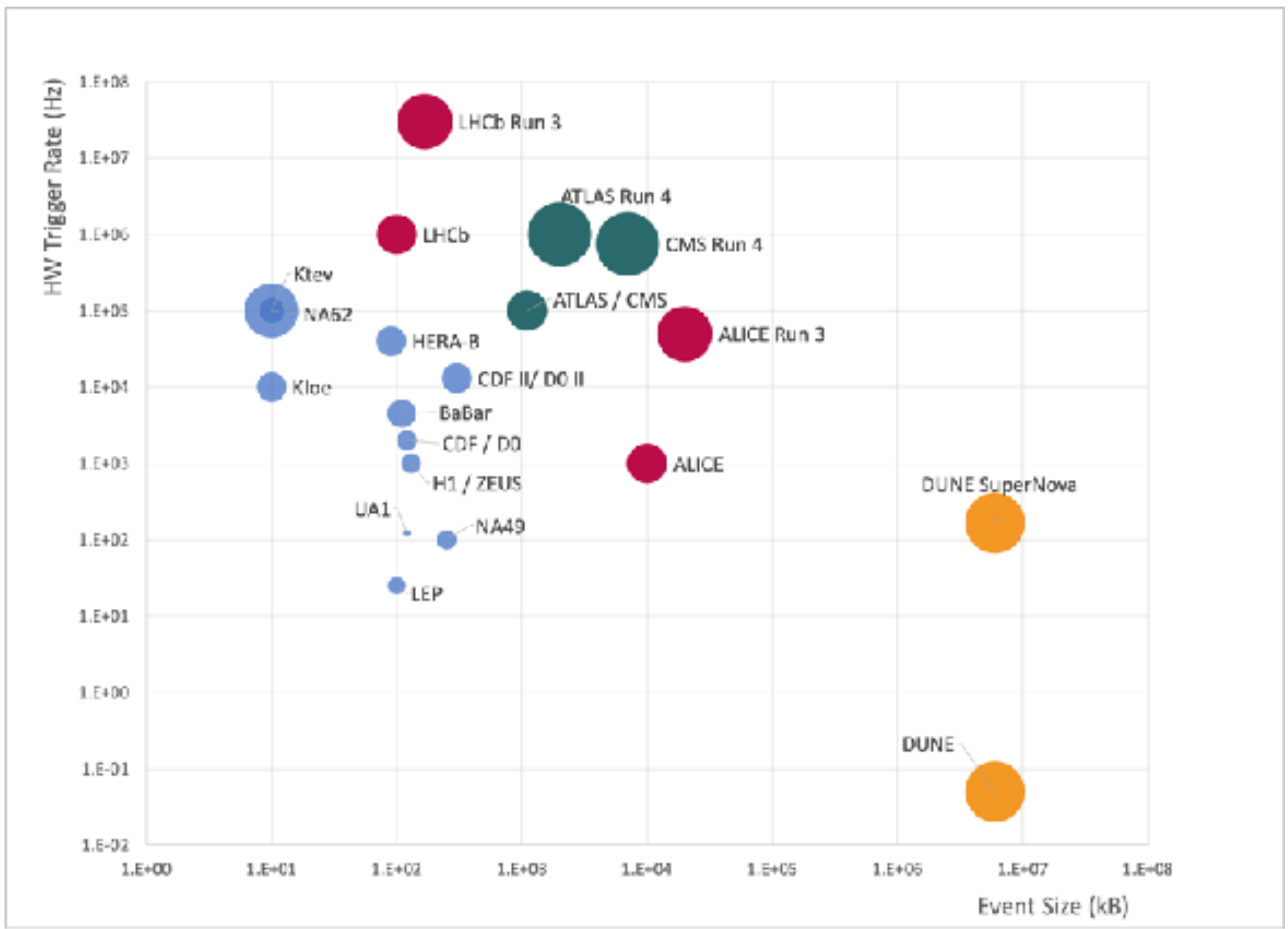
Image Credit

Triggers

Trigger rate increases by more than an **order of magnitude** for **ALICE and LHCb** for Run 3

Trigger rate increase by an **order of magnitude** for **ATLAS and CMS** for Run 4

Even **larger event** sizes for DUNE but lower rate



Not shown, potential LHCb and ALICE upgrades

after A. Cerri

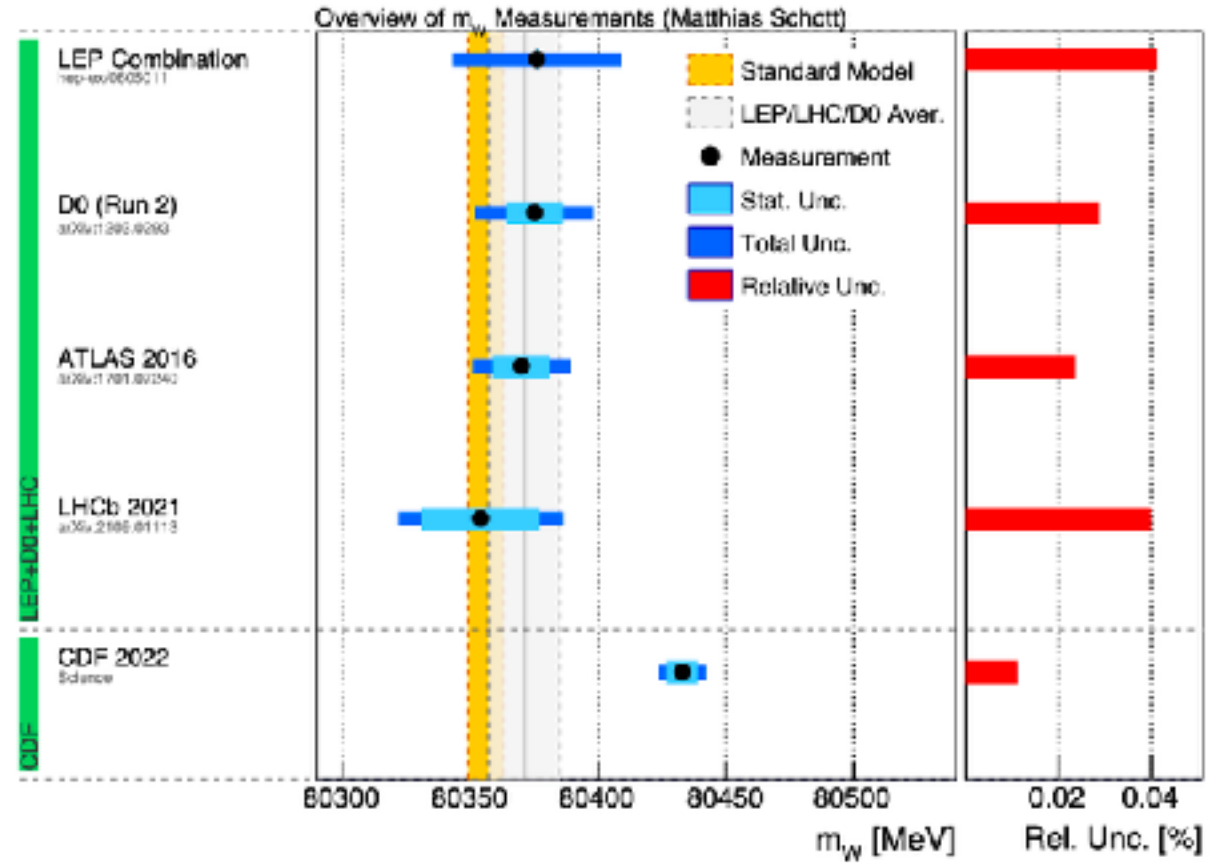
Trigger Evolution

- Triggers have extremely **low latency** requirements
  - Track reconstruction can be a challenge
- Algorithms are evolving in two primary directions
  - **More computation** and **more complex** algorithms (close to offline physics performance) for the hardware trigger
  - **Triggerless read-out**: no hardware trigger and the software trigger processes all events
- Can mix approaches hardware accelerators in the software trigger
- We can expect these trends to accelerate for future accelerators

Source

Physics Precision

- Future colliders aim to make increasingly precise measurements
- e.g. W mass measurements today
- Extremely high precision for future lepton colliders
- Require:
  - precise theory
  - precise calibration
- Result in large computational needs



M. Schott

Quantity	current	ILC250	ILC-GigaZ	FCC-ee
$\Delta\alpha(m_Z)^{-1}$ ($\times 10^3$)	17.8*	17.8*		3.8 (1.2)
Δm_W (MeV)	12*	0.5 (2.4)		0.25 (0.3)
Δm_Z (MeV)	2.1*	0.7 (0.2)	0.2	0.004 (0.1)
Δm_H (MeV)	170*	14		2.5 (2)
$\Delta\Gamma_W$ (MeV)	42*	2		1.2 (0.3)
$\Delta\Gamma_Z$ (MeV)	2.3*	1.5 (0.2)	0.12	0.004 (0.025)
ΔA_e ($\times 10^5$)	190*	14 (4.5)	1.5 (8)	0.7 (2)
ΔA_μ ($\times 10^5$)	1500*	82 (4.5)	3 (8)	2.3 (2.2)
ΔA_τ ($\times 10^5$)	400*	86 (4.5)	3 (8)	0.5 (20)
ΔA_b ($\times 10^5$)	2000*	53 (35)	9 (50)	2.4 (21)
ΔA_c ($\times 10^5$)	2700*	140 (25)	20 (37)	20 (15)

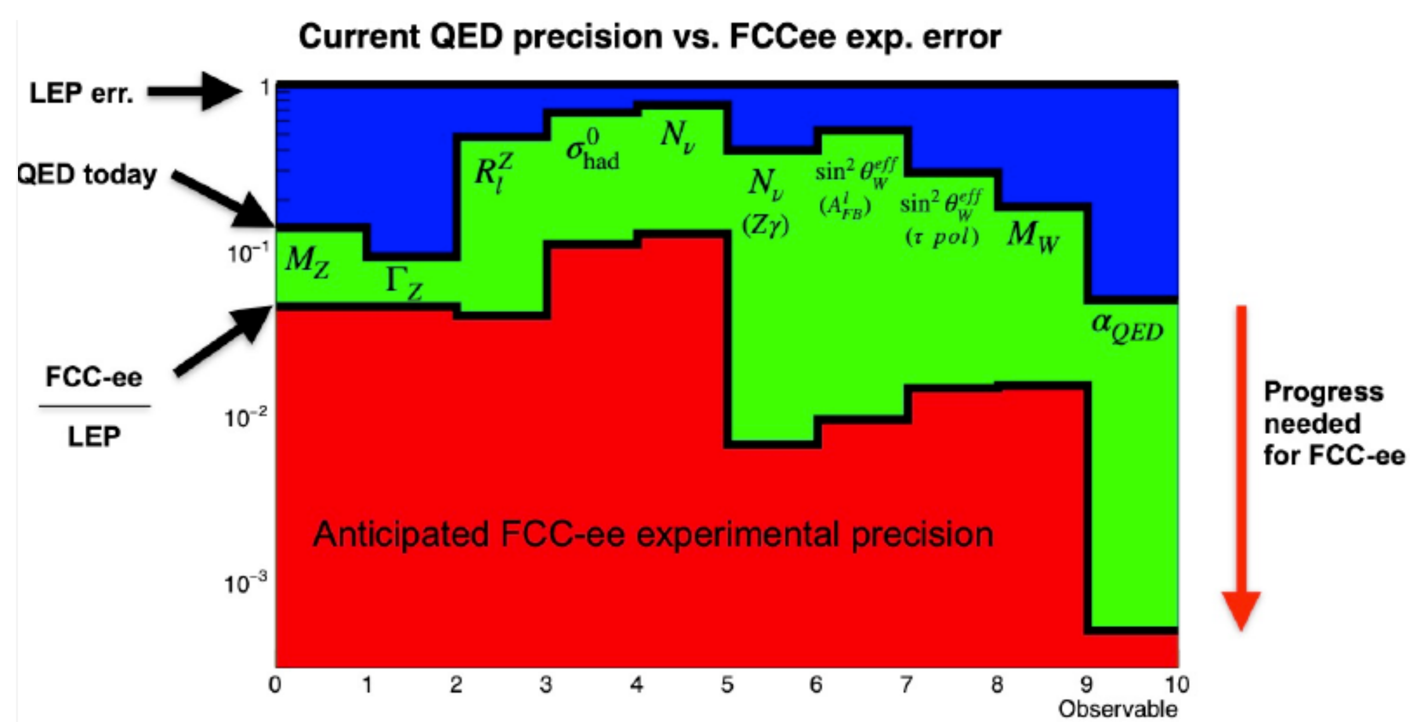


Image Source

Example: FCC-ee

- Z-pole running requirements driving computing needs
- Multiple ways of event reconstruction and simulation to address systematic
- Current LHC-scale computing is sufficient for simulation needed

Using LHC-scale computing is nearly sufficient (eg, within 10x) for all the simulation needed for the Z-pole run of a FCC-ee detector

	Generation	Simulation	Reconstruction	DELPHES
Computing unit	$3.5–5.2 \cdot 10^{10}$	$2.6–3.9 \cdot 10^{6}$	$5.2–7.8 \cdot 10^{6}$	$2.4–3.6 \cdot 10^{10}$
ATLAS equivalent	$3.5–5.2 \cdot 10^{13}$	$2.6–3.9 \cdot 10^{9}$	$5.2–7.8 \cdot 10^{9}$	$2.4–3.6 \cdot 10^{13}$

Events simulated per day using the equivalent of the ATLAS computing facilities

RAW storage similar to the full HL-LHC

Run	$\sqrt{s}$ (GeV)	Statistics	RAW data
Z	91.2	$3 \cdot 10^{12}$ Z decays (visible)	3–6 EB
WW	160	10^{8} $W^{+}W^{-}$ events	0.1–0.2 PB
ZH	240	10^{6} ZH events	1–2 TB
$t\bar{t}$	350, 365	10^{6} $t\bar{t}$ events	1–2 TB

Analysis level data similar to LHC Run 2

Run	$\sqrt{s}$ (GeV)	Statistics	AOD data
Z	91.2	$3 \cdot 10^{12}$ Z decays (visible)	15–30 PB
WW	160	10^{8} $W^{+}W^{-}$ events	0.5–1 TB
ZH	240	10^{6} ZH events	5–10 GB
$t\bar{t}$	350, 365	10^{6} $t\bar{t}$ events	5–10 GB

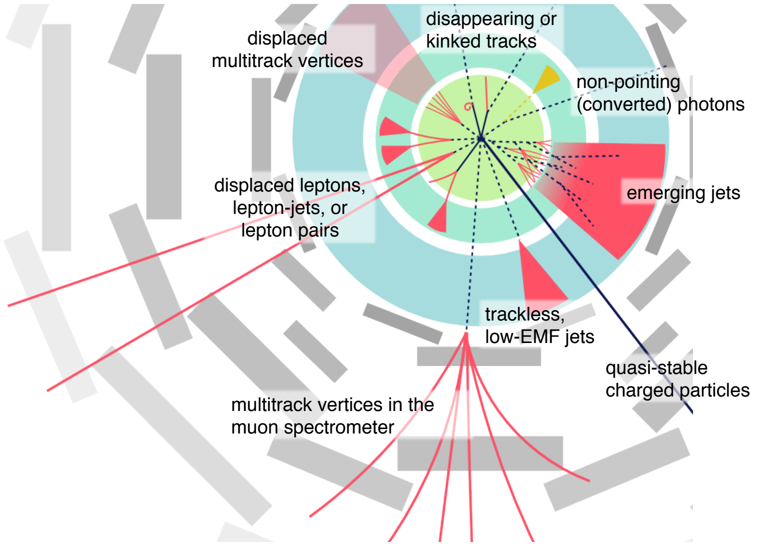
Ganis, Helsens:
https://arxiv.org/abs/2111.10094

7

Slide Credit

Unconventional Signatures

- As we've discovered the Higgs boson at the LHC with no signs of new physics
- Important to ask if there could be signs of new physics that we just don't see
  - Weak (or no) interaction
  - Long lifetimes
  - High mass
- 'Long-lived particles' have become an are of focus
  - Requires new algorithms and additional computing resources
    - e.g. more track reconstruction algorithms

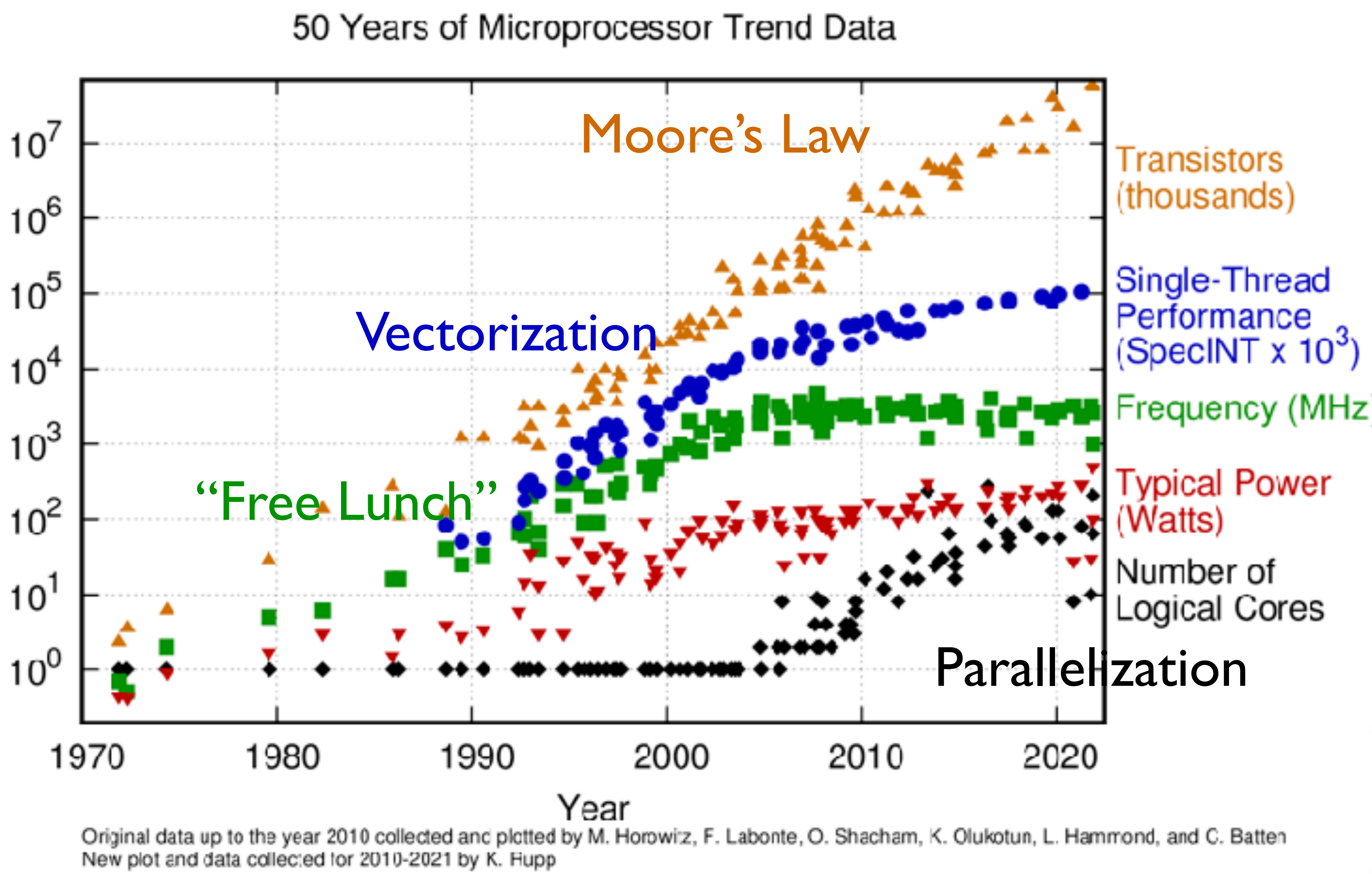


H. Russell

Challenges and Opportunities

Moore’s Law

- Number of transistors in an integrated circuit doubles approximately every two years



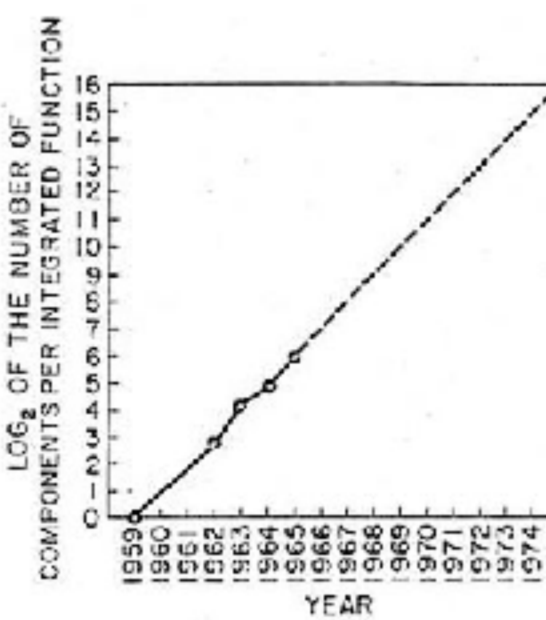


Image Credit Slides after

Source

Beyond CPUs

- **Hardware accelerators** are custom-made hardware designed to perform specific functions more efficiently than CPUs
- Wide variety of hardware accelerators depending on the application
  - e.g **GPU**, **FPGA**, **TPU** ….
- We use hardware accelerators frequently in our daily lives
  - e.g. graphics acceleration, encryption, machine learning, decoding video streams
- A large fraction of the power in High Performance Centers (HPCs) comes from GPUs
- Can also consider "New" computing paradigms
  - **Neuromorphic** computing, **quantum** computing….
- Hardware accelerators are significantly more challenging to program than CPUs

Machine Learning

- Machine learning methods have been used in HEP since the **1990s** [see Bhat, 2011 for a review]
  - Recent advent of **deep learning** has boosted performance
- Classification and regression used in **all steps** of the HEP software pipeline
- Developments in machine learning are often driven by **industry**
  - HEP benefits through the application of these techniques
- In most cases, aim for improved **physics performance** rather than improved speed
- Covered in far more detail in Javier's talk, but has transformed the software landscape
- Also good use case for hardware acc

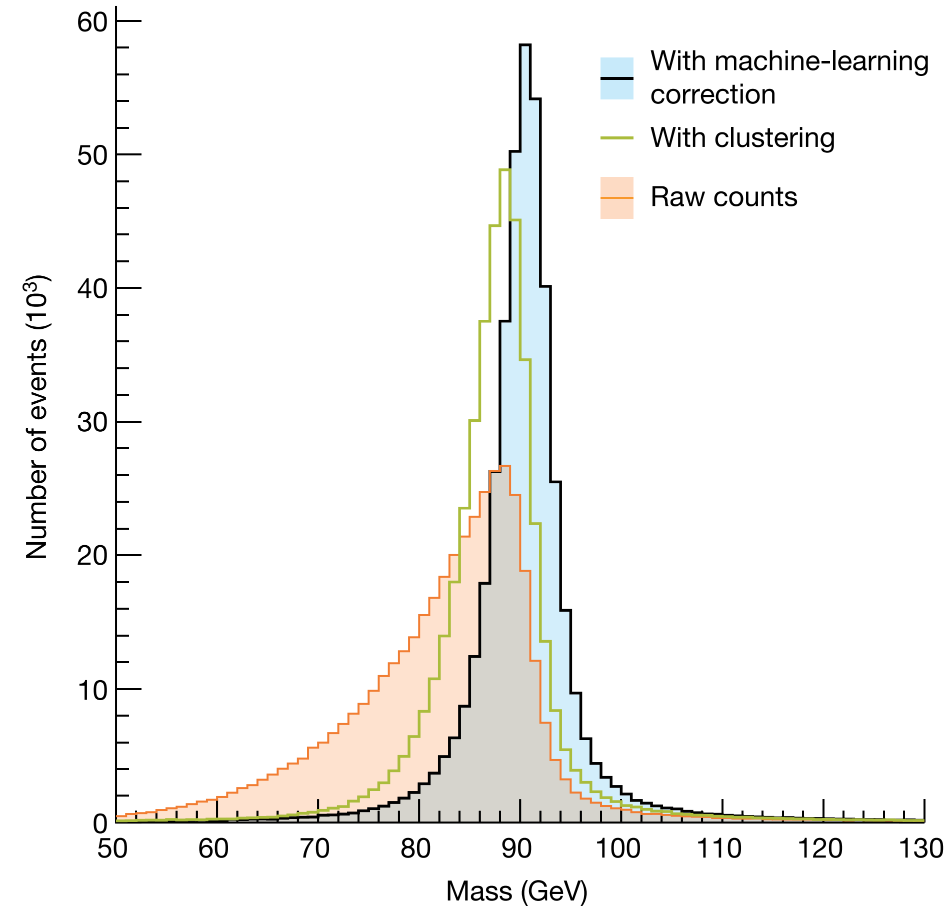


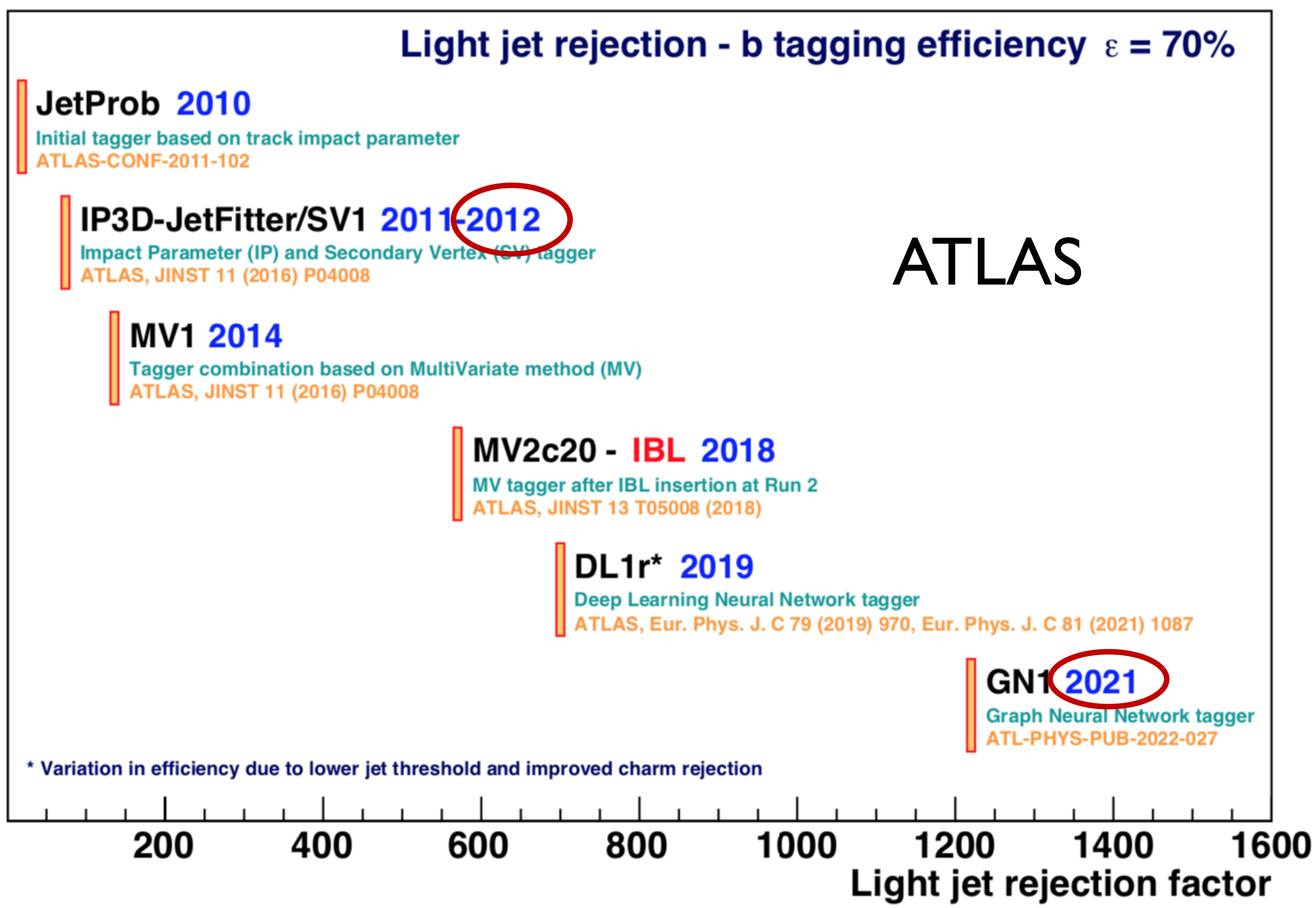
Table 1 | Effect of machine learning on the discovery and study of the Higgs boson

Analysis	Years of data collection	Sensitivity without machine learning	Sensitivity with machine learning	Ratio of P values	Additional data required
CMS[24] $H \to \gamma\gamma$	2011–2012	2.2σ, $P = 0.014$	2.7σ, $P = 0.0035$	4.0	51%
ATLAS[43] $H \to \tau^+\tau^-$	2011–2012	2.5σ, $P = 0.0062$	3.4σ, $P = 0.00034$	18	85%
ATLAS[99] $VH \to bb$	2011–2012	1.9σ, $P = 0.029$	2.5σ, $P = 0.0062$	4.7	73%
ATLAS[41] $VH \to bb$	2015–2016	2.8σ, $P = 0.0026$	3.0σ, $P = 0.00135$	1.9	15%
CMS[100] $VH \to bb$	2011–2012	1.4σ, $P = 0.081$	2.1σ, $P = 0.018$	4.5	125%

https://doi.org/10.1038/s41586-018-0361-2

One Example: Flavor tagging

- Extensive (and exclusive) use of ML for **flavor tagging** for many years
- **Example:** Improvement in light jet rejection for ATLAS over the years
  - Large improvement by the use of deep learning and GNNs
    - Unclear what the limit is here (GN2 under development)



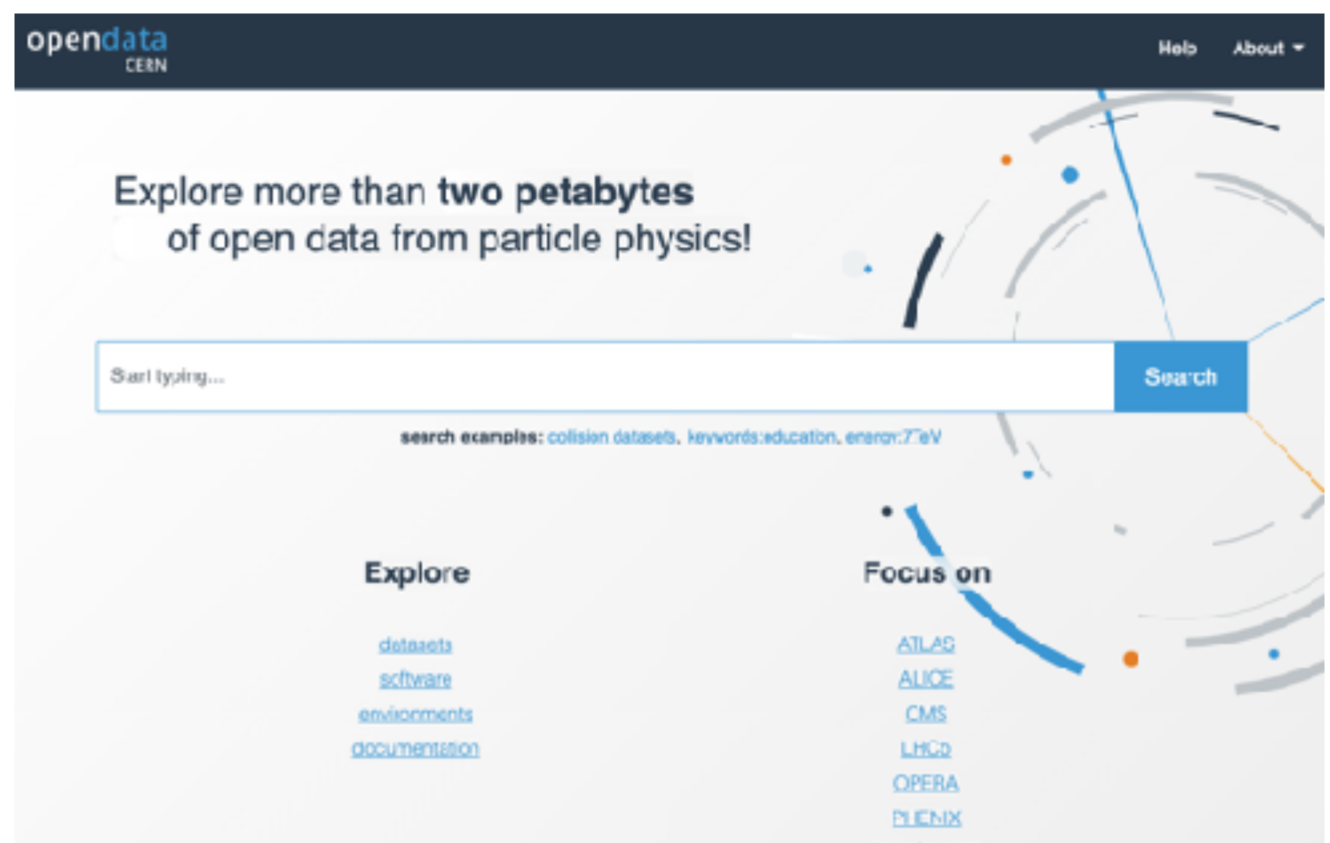
M. Kado

Open [Software, Data]

- Open source philosophy has long played an important role in software development
- At the LHC, first the **results**, then the **software**, then **data** and most recently the **likelihoods** of the LHC experiments have become open
  - **Reinterpretation** can probe additional models
  - However: can be challenging to use our software/data if you don't have direct access to experts and significant hardware resources
- CERN **Open Data Policy**

reana

https://reanahub.io/



Common Software R&D Institutes

- HEP experiments at the LHC and in the future face **similar changes**
  - Formation of the **HEP Software Foundation (HSF)** in 2015
    - Provides a common forum for software for HEP experiments
- Funded R&D efforts in **common software** in a number of countries
- Activity encouraged by the European Strategy
  - "[…] vigorously pursue **common,** coordinated R&D efforts […], to develop **software** […] that exploit the recent advances in information technology and data science […]"
- Common projects can aid software maintainability
  - More likely to have a pool of people available for maintenance
  - Can also be challenging to fund in the long term once beyond the R&D stage

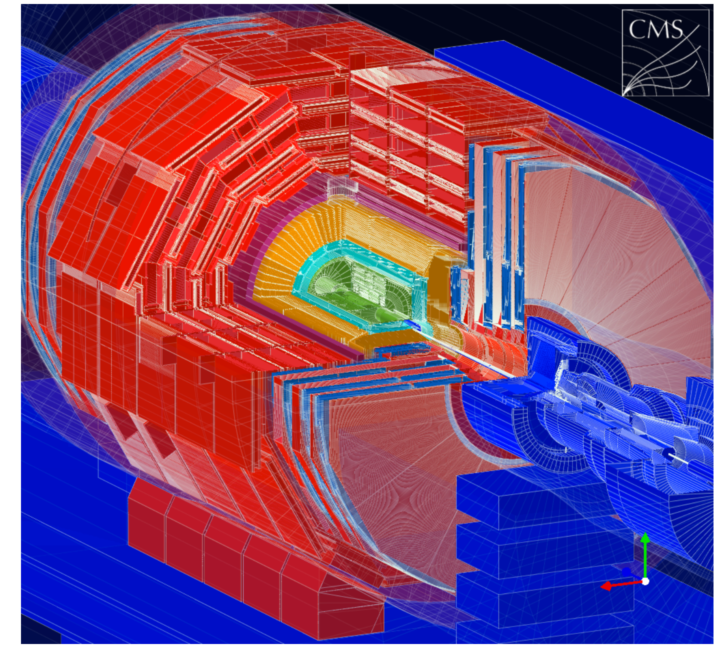
Examples of Software Institutes

- IRIS-HEP, NSF, 2018
  - Analysis systems, innovative algorithms, DOMA, training
- ErUM-DATA, Helmholtz Institute, Germany
  - Heterogeneous computing and virtualized environments, machine learning for reconstruction and simulation
- EP R&D, CERN, Switzerland, 2020
  - Turnkey software systems, faster simulation, track and calo reconstruction, efficient analysis
- HEP-CCE, DOE, USA, 2019
  - Portable Parallelization Strategies, I/O Strategy on HPC, Event generators
- AIDAInnova, European Commission EU, 2021
  - Turnkey software, track reconstruction, particle flow, ML simulation
- SWIFT-HEP STFC, 2021 and ExCALIBUR-HEP, 2020, UKRI UK
  - Exascale data management, Event generators, detector simulation on GPUs, FPGA tracking for HLT

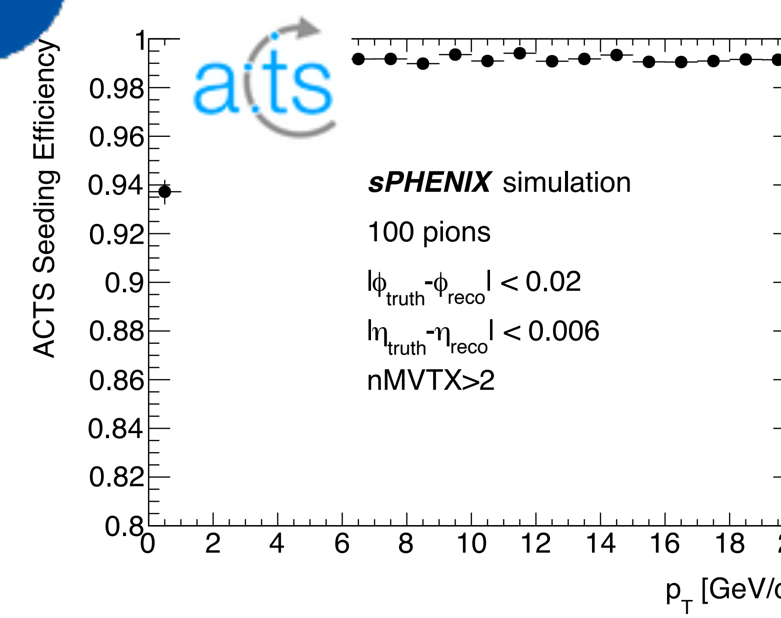
Slide Credit: G. Stewart

Software for Multiple Experiments

- **Common packages** have been used extensively by many experiments over many years including CLHEP, ROOT, Geant4, GAUDI
- For Run-3, ALICE uses **ALFA**, framework developed with GSI (FAIR) as common integration platform for online/offline processing
  - Online reconstruction using heterogeneous farm
  - Enables parallel data processing
- **DD4HEP** is now used by CMS, LHCb among other experiments for the detector description
- **ACTS** has origins in ATLAS tracking software, but currently being explored by different experiments
- LHCb is splitting off **Gaussino** as experiment-independent part of Gauss simulation framework (w. CERN SFT/FCC)
- Can save resources by non re-inventing the wheel



Vuosalo et al.

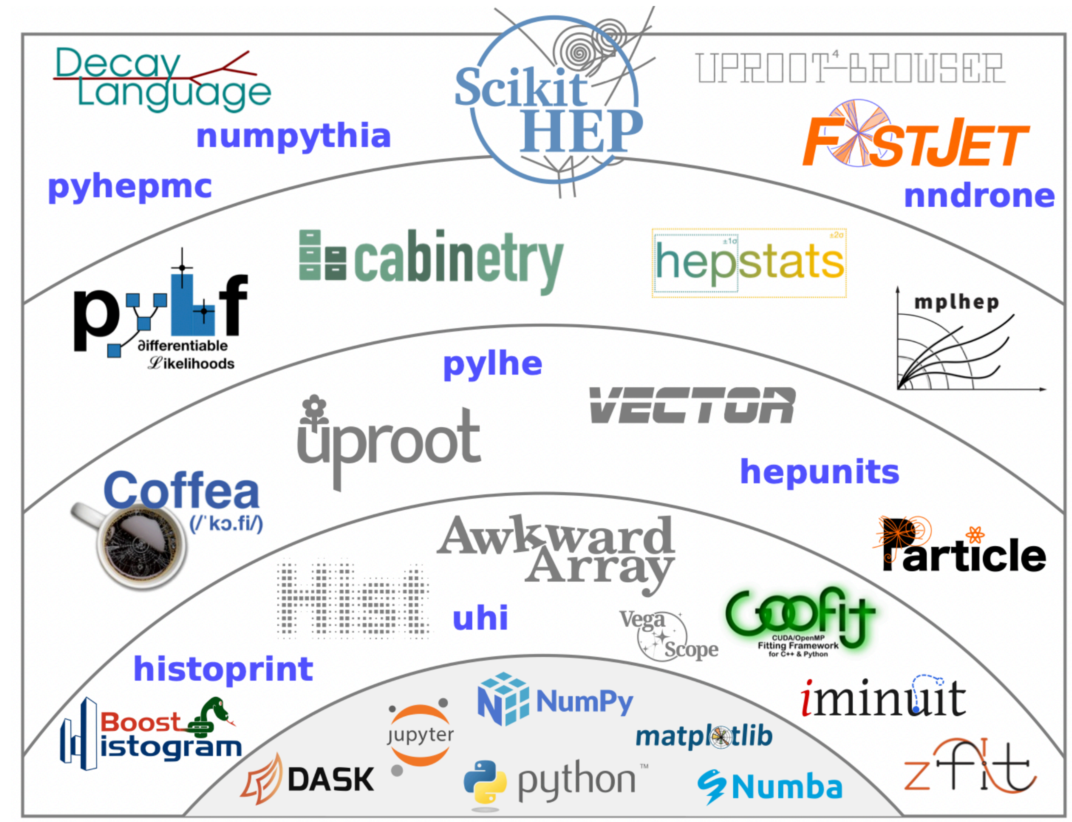


Osborn et al, arXiv:2103.06703

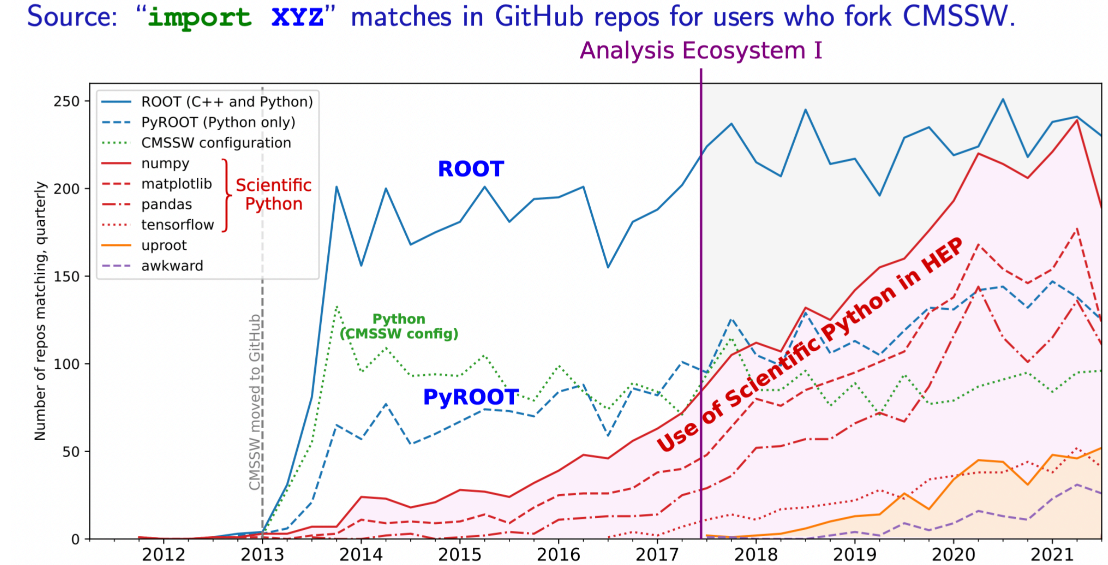
Python for Analysis

- Ongoing **boom** in the field of data science
- **Python** has become the language of choice for data science applications
  - Huge community has developed well-documented tools
  - numpy, matplotlib, pytorch, tensorflow, etc
- Balanced against our **own** designed-to-purpose and customized tools, in particular, ROOT
- **Python** is becoming increasing popular for analysis especially amongst the younger members of our community

ROOT
Data Analysis Framework



Source: "**import XYZ**" matches in GitHub repos for users who fork CMSSW.



J. Pivarski

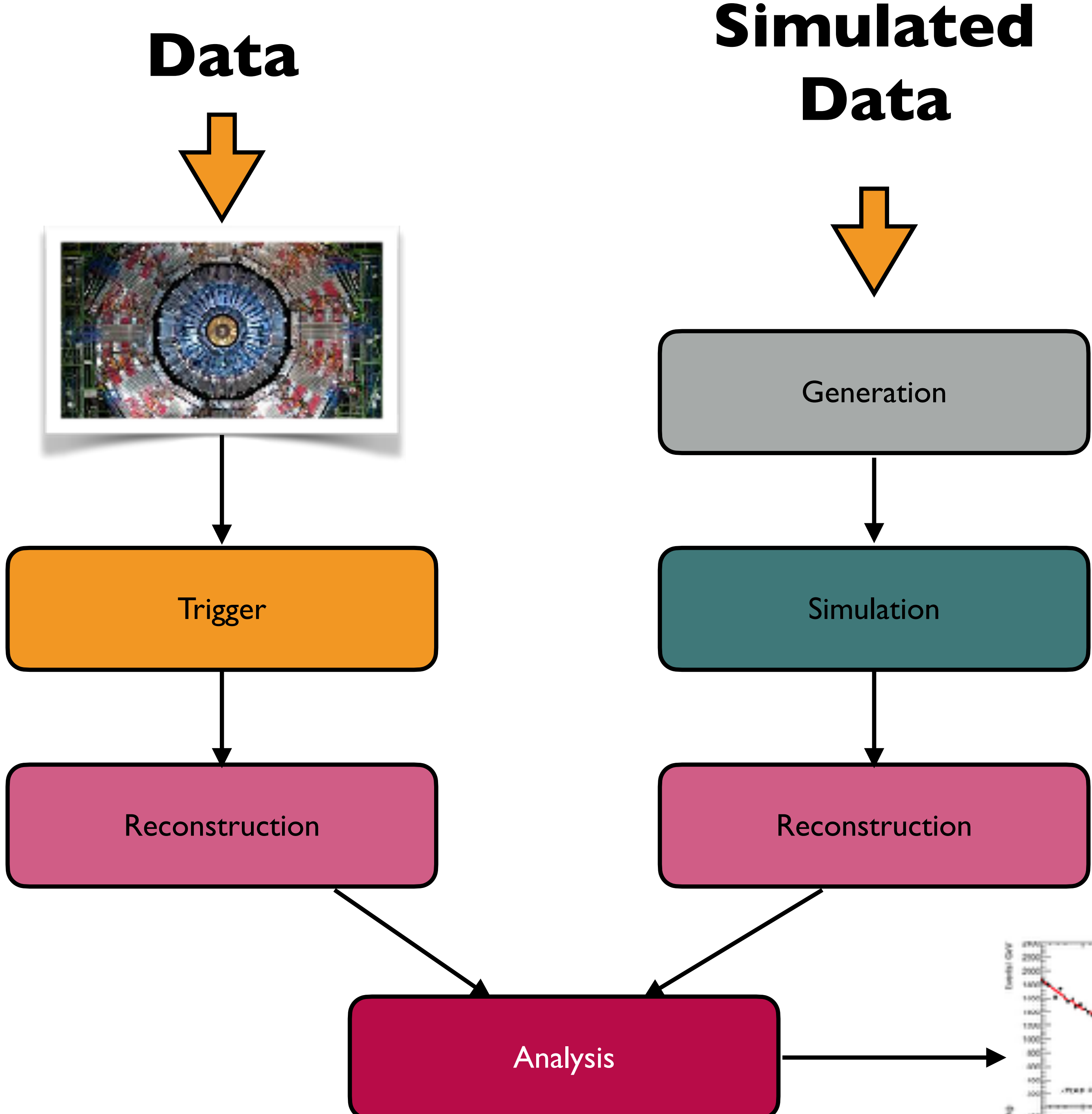
Conclusion

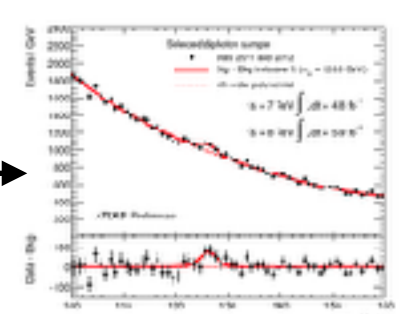
- A taster of current and future challenges and opportunities in software and computing
- A hadron collider would result in significant computational challenges
- Also challenges for electron colliders (precision) and muon colliders (beam background)
- At the same time, the field has been evolving rapidly
  - Many opportunities to think about doing things in a dramatically different way in the future
    - No trigger!
    - Even more machine learning/AI

Back up

Other Challenges

Challenges anticipated at each step of the **data processing** and **simulation** chain





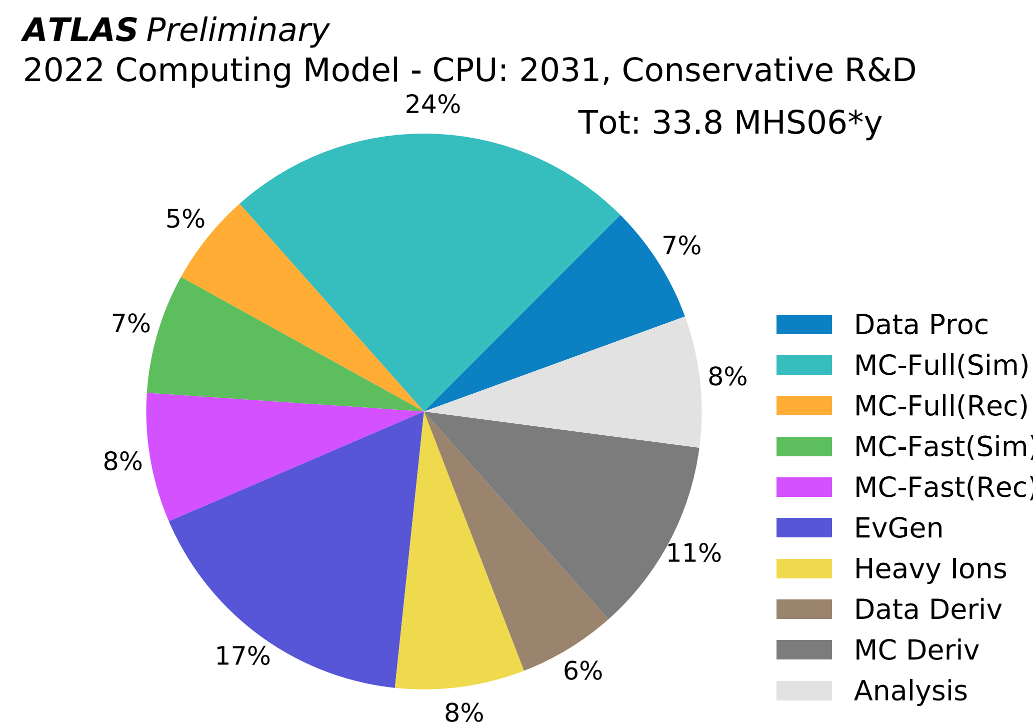


Image Credit

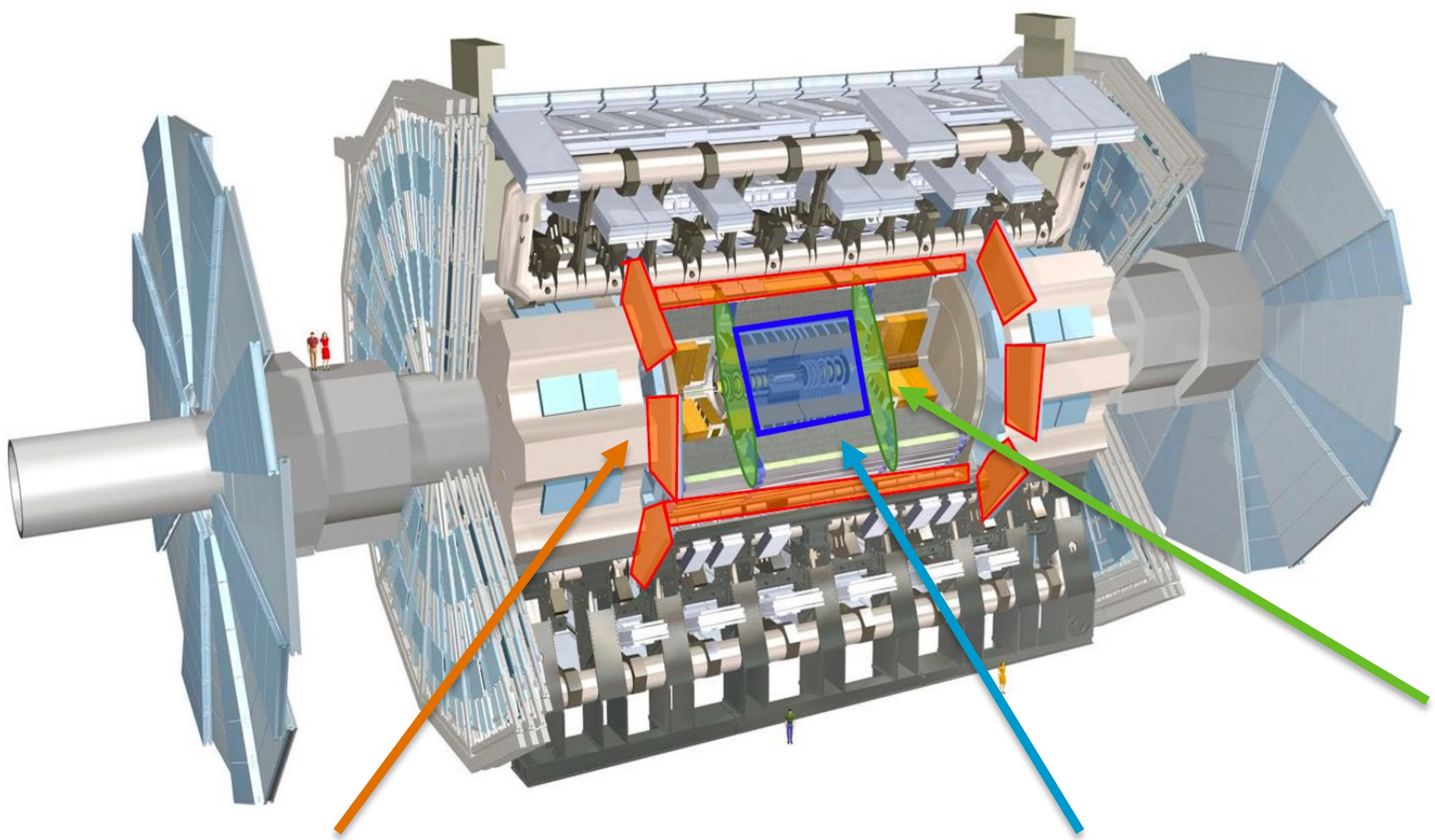
HL-LHC Resources: CMS

Processing Step	Time/evt [HS06s]	
	200 PU	140 PU
Gen+Sim	1900	
Digi+PU mix+Reco	5100	3200

Tier	Event size [MB]	
	200 PU	140 PU
RAW	5.9	4.3
AOD	2	1.4
MiniAOD	0.25	0.18
NanoAOD	0.004	0.004

Source

ATLAS Upgrades



Upgraded Trigger and Data Acquisition System

- Single Level Trigger with 1 MHz output
- Improved 10 kHZ Event Farm

Electronics Upgrades

- On-detector/off-detector electronics upgrades of LAr Calorimeter, Tile Calorimeter & Muon Detectors
- 40 MHz continuous readout with finer segmentation to trigger

High Granularity Timing Detector (HGTD)

- Precision time reconstruction (30 ps) with Low-Gain Avalanche Detectors (LGAD)
- Improved pile-up separation and bunch-by-bunch luminosity

New Muon Chambers

- Inner barrel region with new RPCs, sMDTs, and TGCs
- Improved trigger efficiency/momentum resolution, reduced fake rate

New Inner Tracking Detector (ITk)

- All silicon with at least 9 layers up to $|\eta| = 4$
- Less material, finer segmentation

Additional small upgrades

- Luminosity detectors (1% precision)
- HL-ZDC (Heavy Ion physics)

Not covered in this talk.

CMS Upgrades

Phase 2 Upgrade Under a 

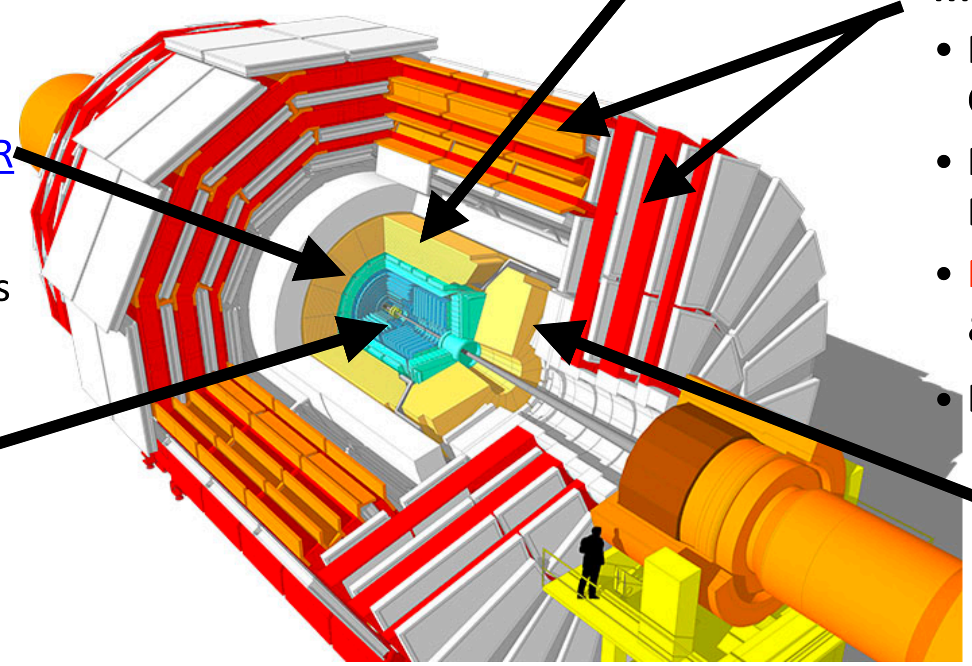
Level 1 Trigger TDR
- New track trigger at 40 MHz
- Particle flow selection
- 750 kHz L1 output
- 40 MHz data scouting (real time analysis)
- L1T latency: 12.5 μs

DAQ & High Level Trigger (HLT) TDR
- Full optical readout
- Heterogeneous architecture
- 60 TB/s event throughput
- 7.5 kHz HLT output

Barrel Calorimeter TDR
- ECAL crystal granularity readout at 40 MHz with precise timing for e/gamma at 30 GeV
- New ECAL and HCAL back-end boards

Muon System TDR
- New Drift Tubes (DTs) & Cathode Strip Chambers (CSCs) FE/BE readout
- New Resistive Plate Chambers (RPCs) BE electronics
- New Gas Electron Multipliers (GEMs) & new iRPCs 1.6 < $|\eta|$ < 2.4
- Extended coverage to $|\eta|$~ 3

New MIP timing detector (MTD) TDR
- Barrel: LYSO crystals + SiPMs
- Endcap: Low-gain avalanche diodes
- 30 ps timing resolution
- Full coverage to $|\eta|$~ 3



Replaced Tracker TDR
- Increased granularity
- Extended coverage to $|\eta|$~ 4
- Designed for tracking in L1T

Beam Radiation Instrumentation and Luminosity (BRIL) TDR
- Target 1% offline (2% online) luminosity uncertainty

New High-Granularity Endcap Calorimeter (HGCAL) TDR
- Imaging calorimeter
- Si, Scint+SiPM in Pb/Cu-W/SS
- 3D showers and precise timing