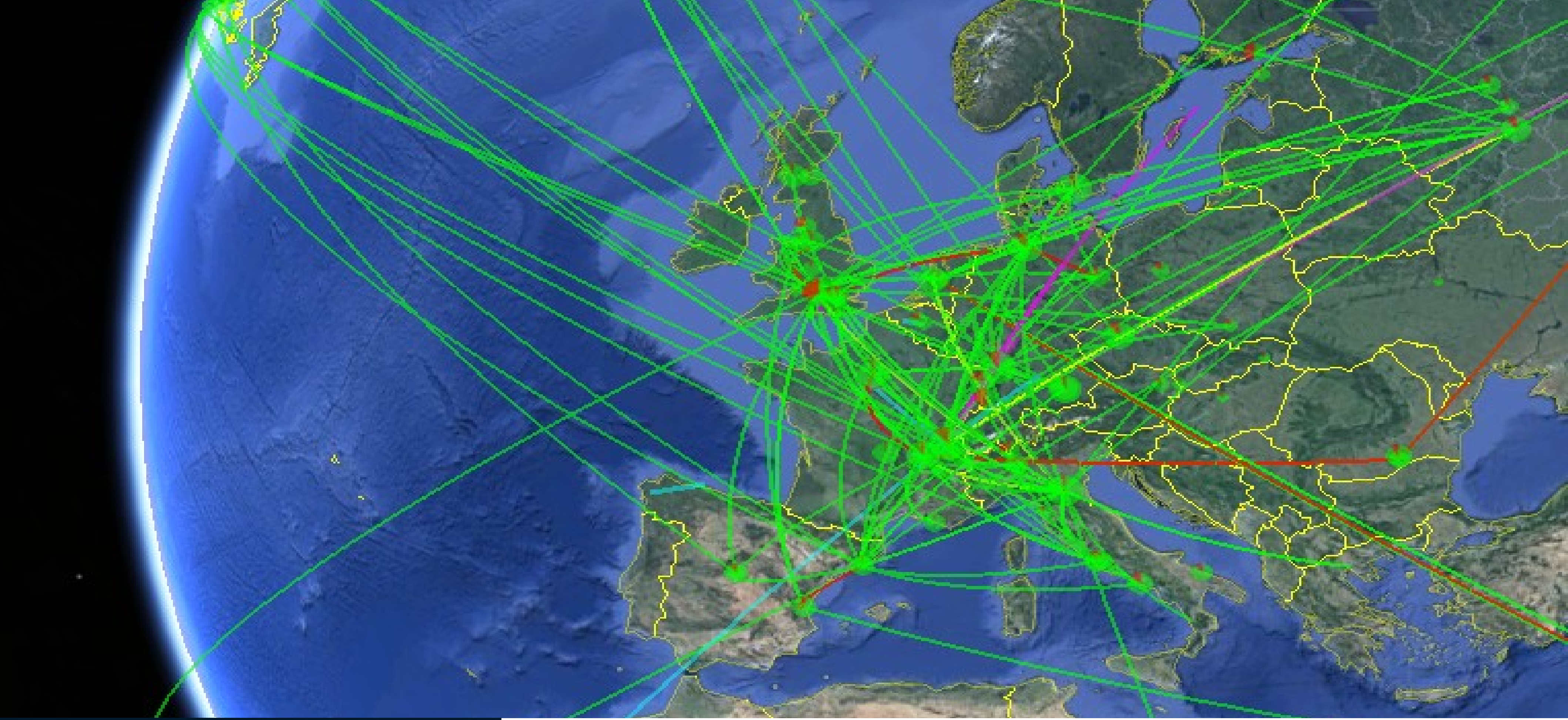


4th COMCHA School

Zaragoza, 8 – 15 April 2026

CAPA COMCHA Computing challenges

Financiado por la Unión Europea NextGenerationEU

XUNTA DE GALICIA

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

AGENCIA ESTATAL DE INVESTIGACIÓN

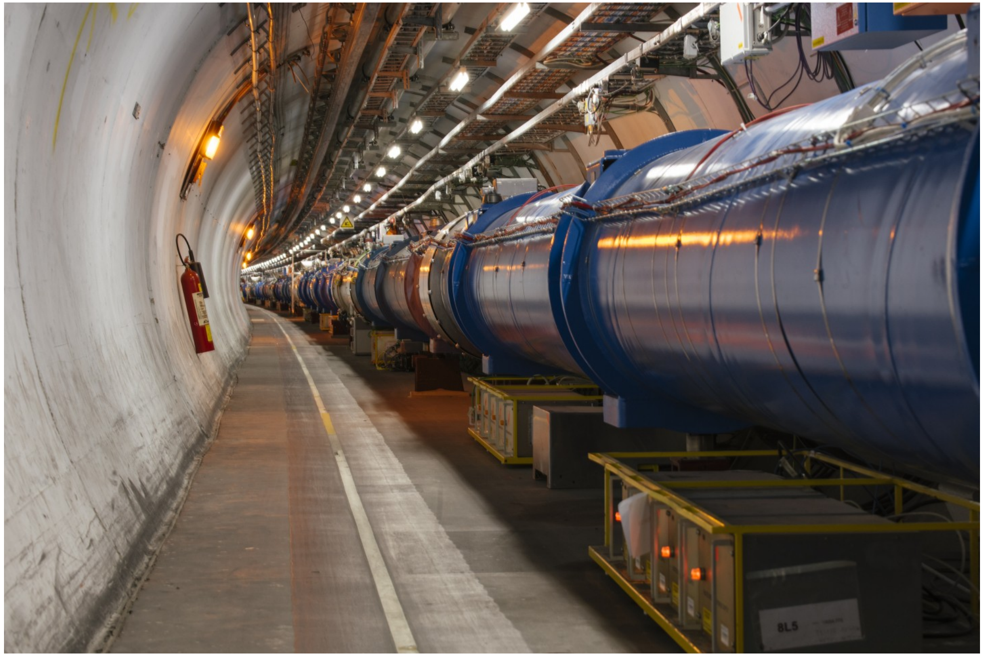
INTALENT UDC - INDITEX

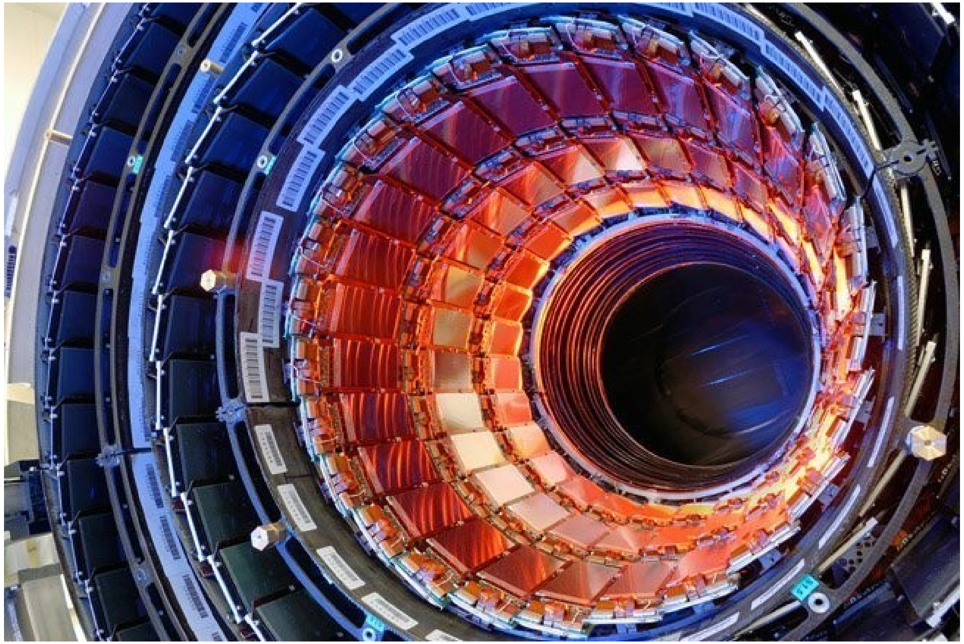
UNIVERSIDADE DA CORUÑA

A new computing era

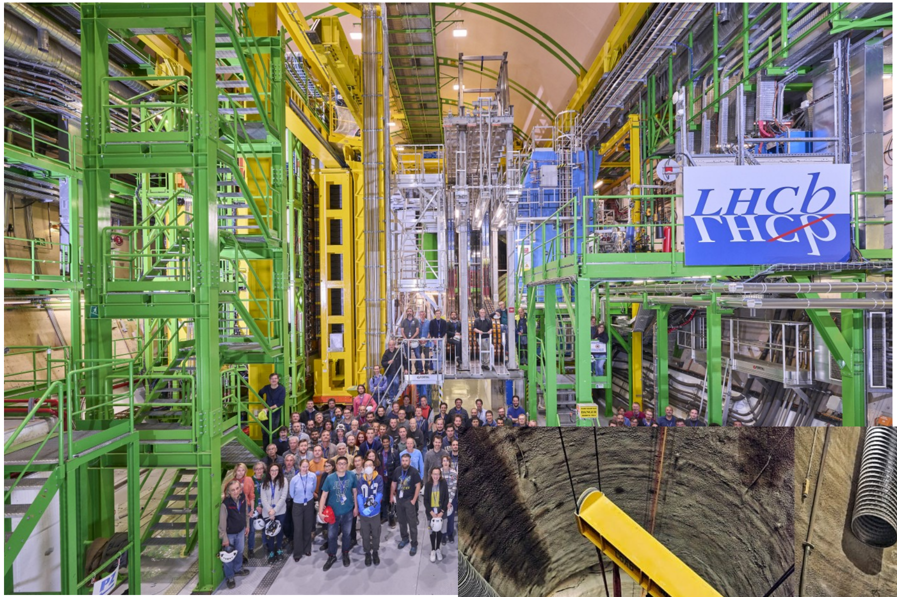
Veronika Chobanova

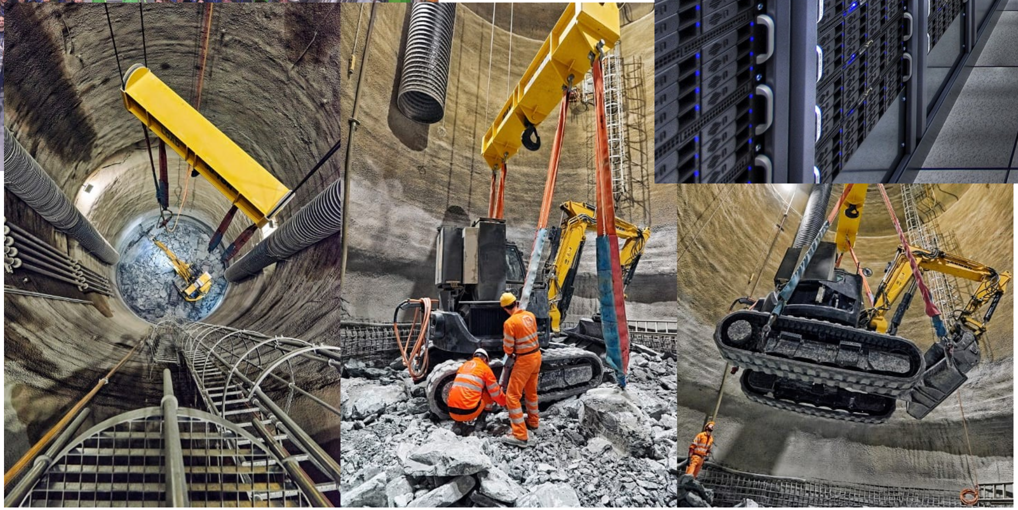
Particle physics is about hardware...

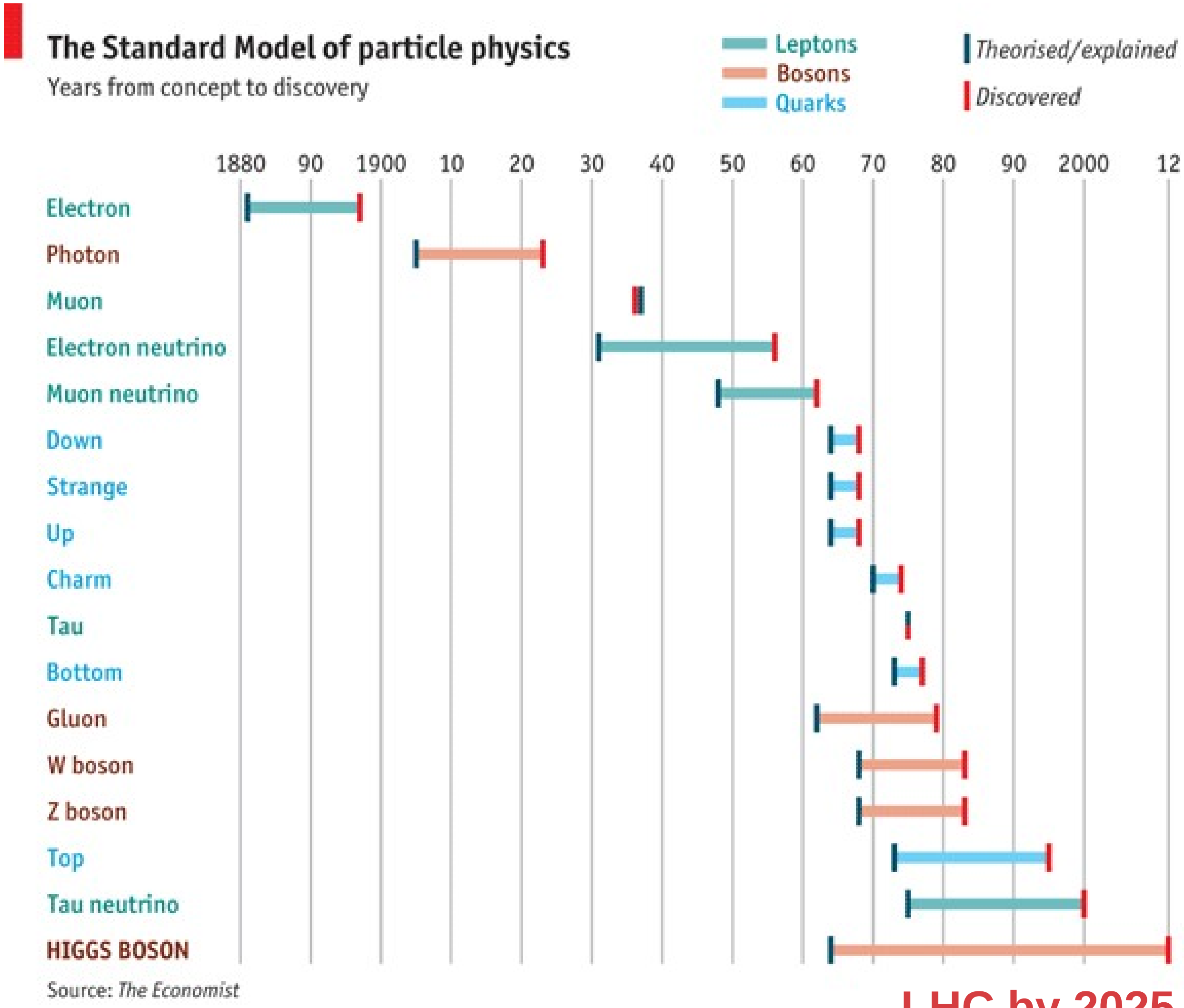


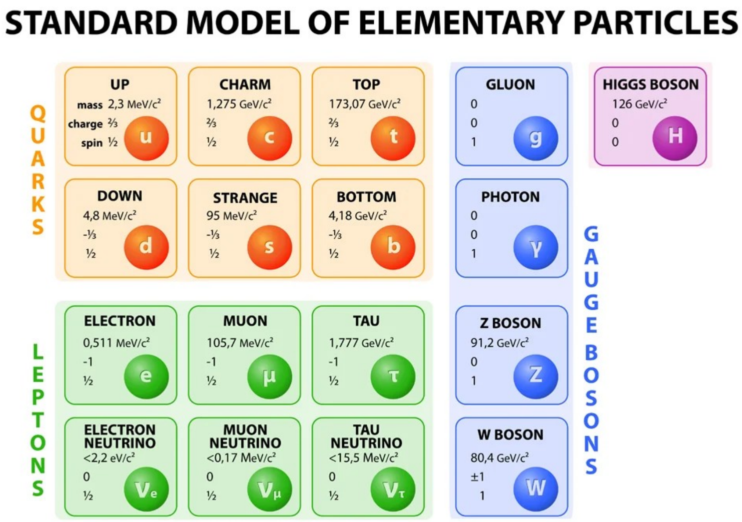


… and so much more









~10 nb^{-1} UA1 & UA2

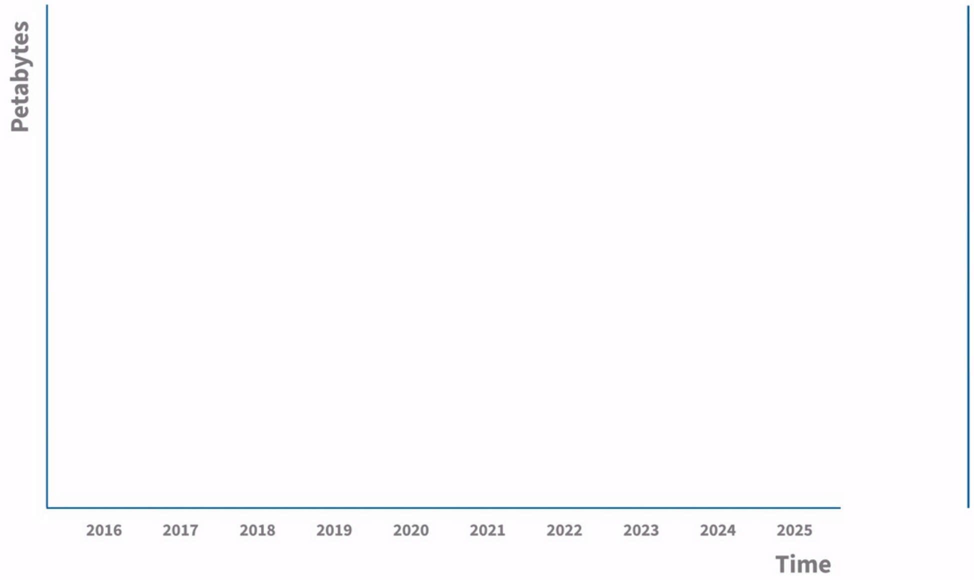
~100 pb^{-1} (x 10^4) D0 & CDF

~10 fb^{-1} (x 10^5) ATLAS & CMS

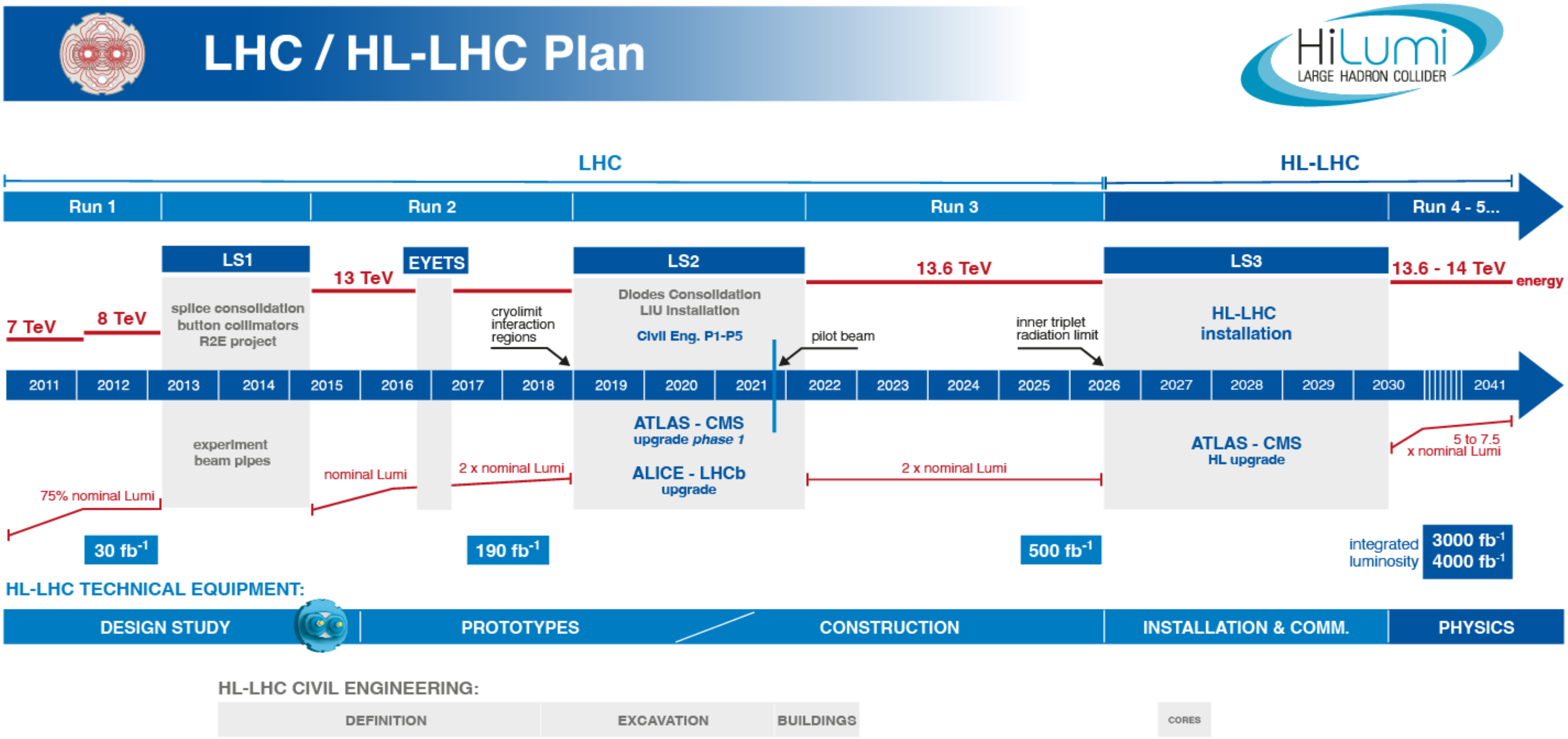
LHC by 2025 ~500 fb^{-1}

LHC by 2041 ~3000-4000 fb^{-1}

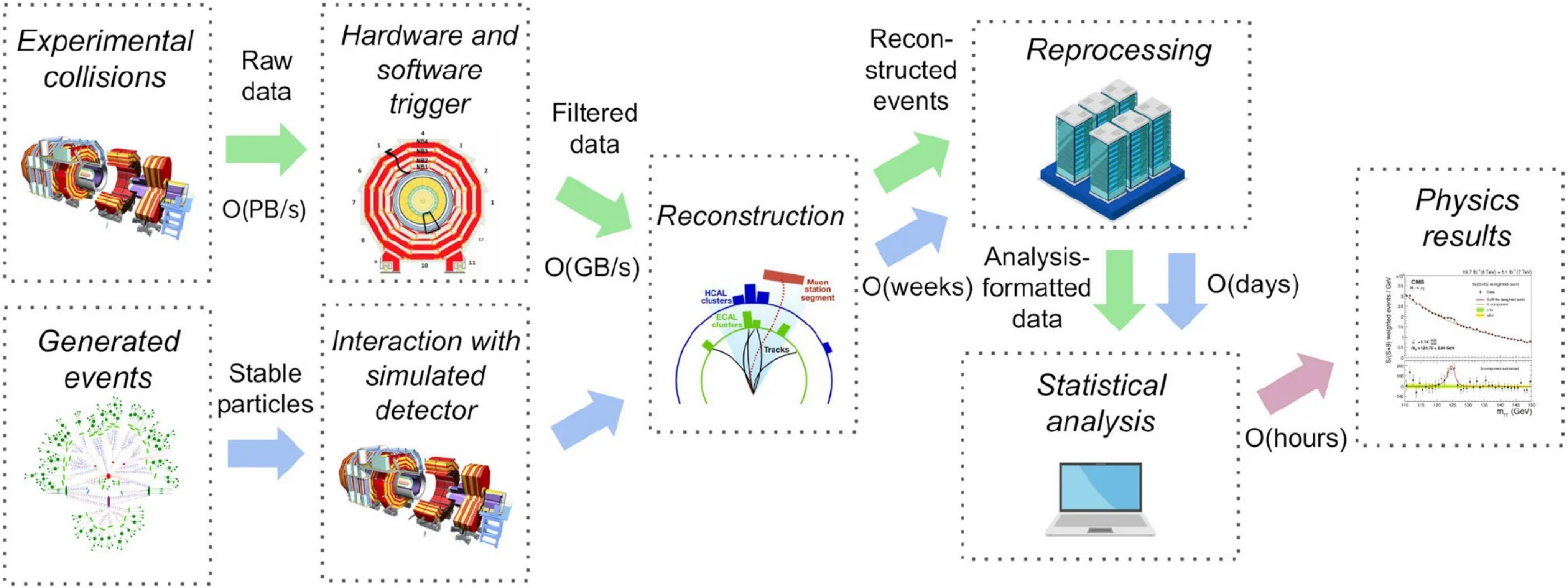
LHC has stored 1 EB of data



x10 in the next 15 years

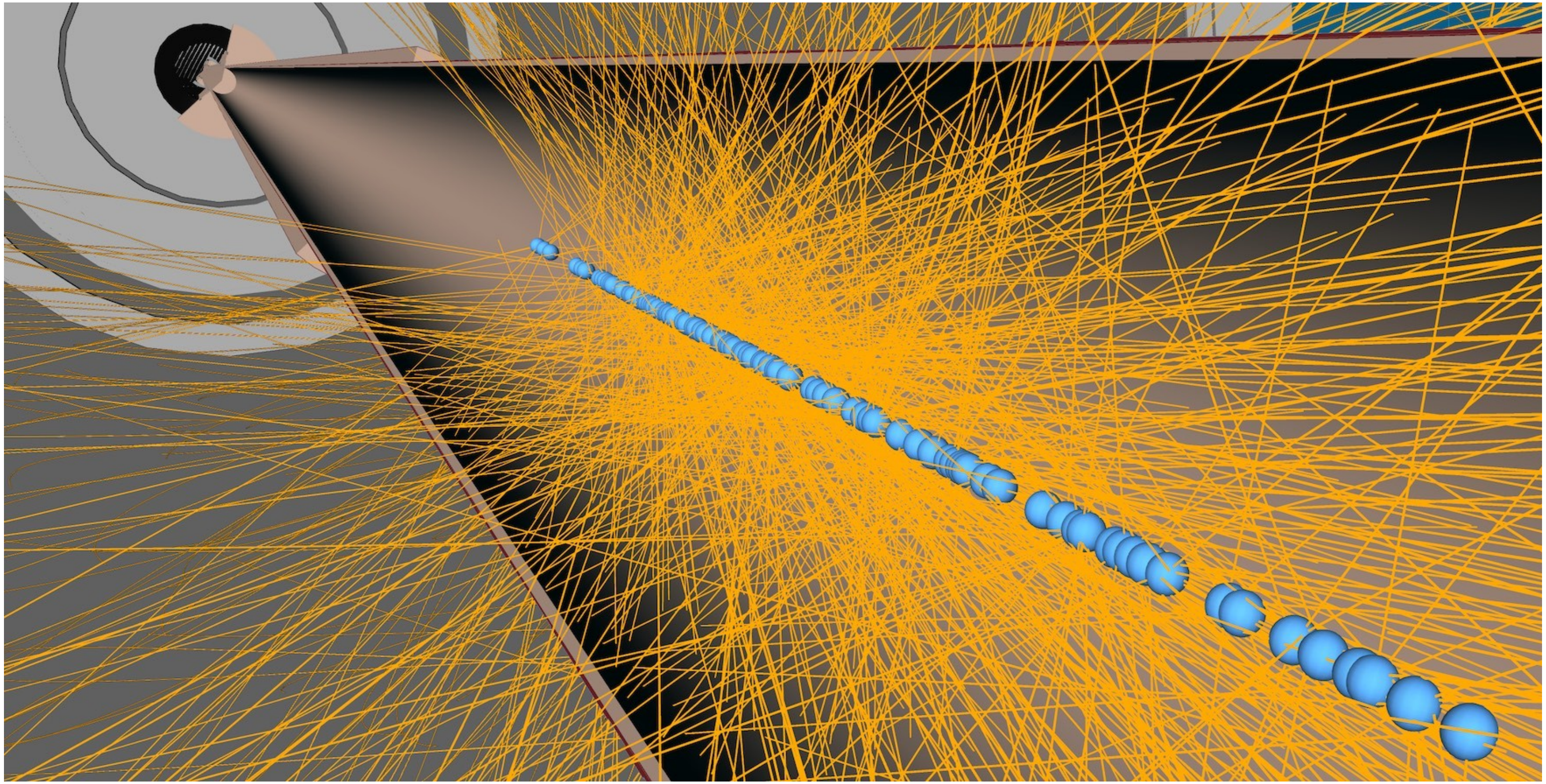


A typical processing work flow at a HEP experiment



Front. Big Data 4, 2021

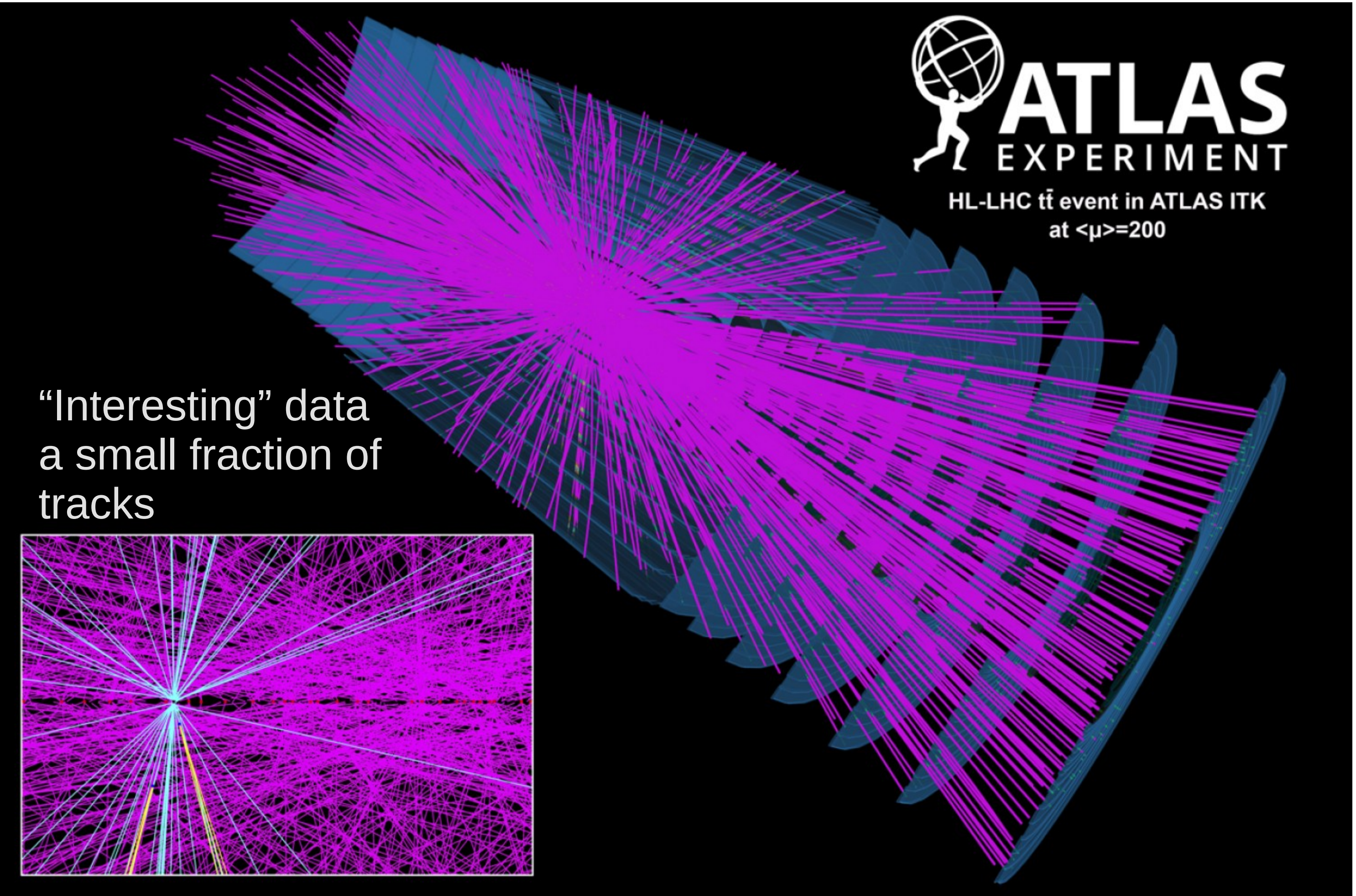
Challenge: Data acquisition and storage



HL-LHC inst. luminosity Run 2 x10

Billions of hits forming $O(10^3)$ tracks from $O(10^2)$ primary vertices

Challenge: Data acquisition and storage

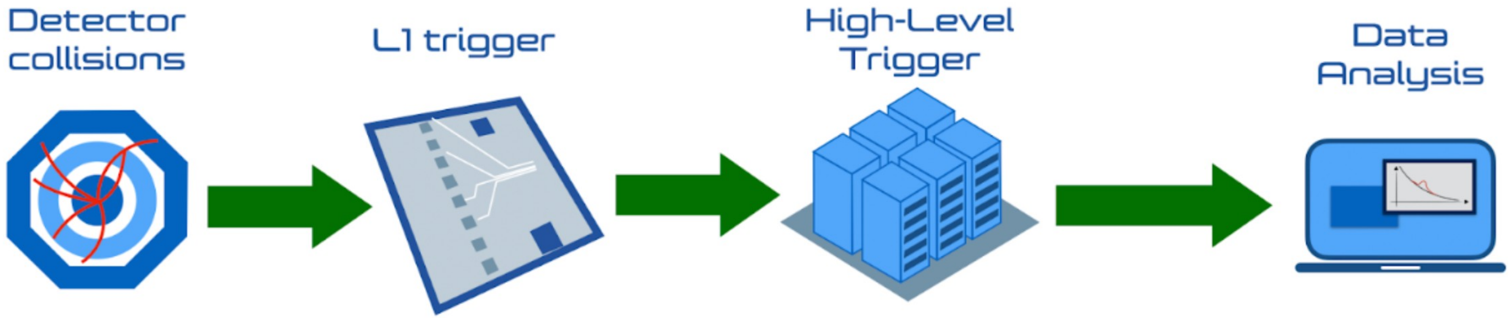


- Challenge to both hardware and software
- Not possible to store all collisions on disk and tape forever
- Filter systems reconstruct objects in the detector and discard “uninteresting” collisions

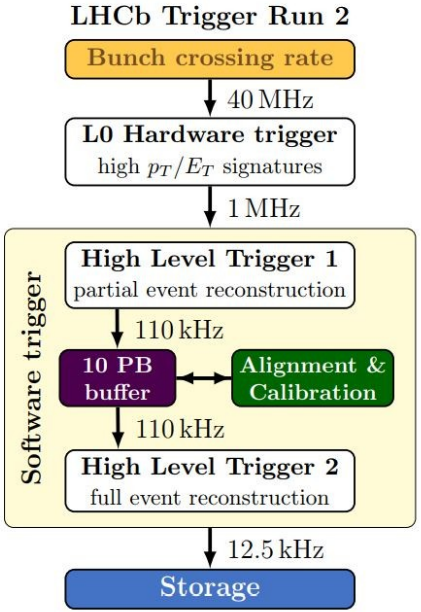
Challenge: Data acquisition and storage

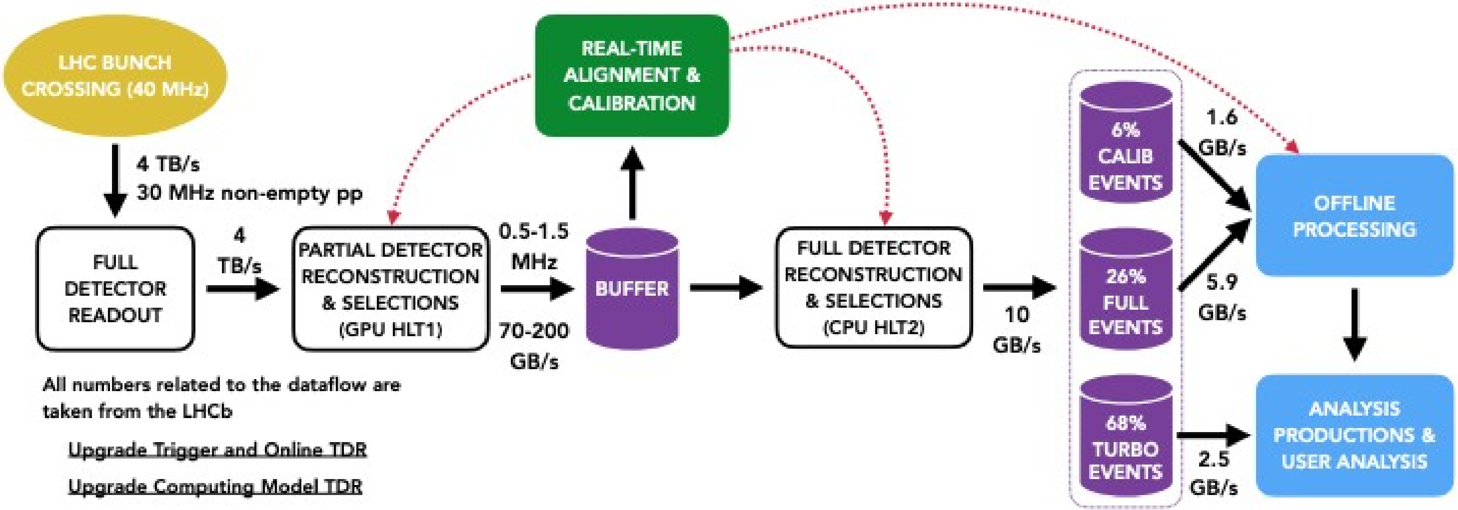
- Detector upgrades (e.g. add timing): new hardware platforms (FPGA, GPU)
- (Partly) reconstruct event at trigger level instead of saving raw data
- Apply ML techniques to improve efficiency

Experiment	Data size	Rate before trigger	Rate after trigger	DAQ throughput	
ALICE now	12000 kB	50 kHz	50 kHz	700 GB/s	
ATLAS now	1500 kB	27000 kHz	90 kHz	55 GB/s	90x
ATLAS 2030	5000 kB	27000 kHz	1 MHz	5000 GB/s	
CMS now	2000 kB	27000 kHz	100 kHz	200 GB/s	32x
CMS 2030	8400 kB	27000 kHz	750 kHz	6300 GB/s	
LHCb now	80 kB	24000 kHz	24000 kHz	2400 GB/s	5x
LHCb 2036	200 kB	24000 kHz	24000 kHz	12000 GB/s	



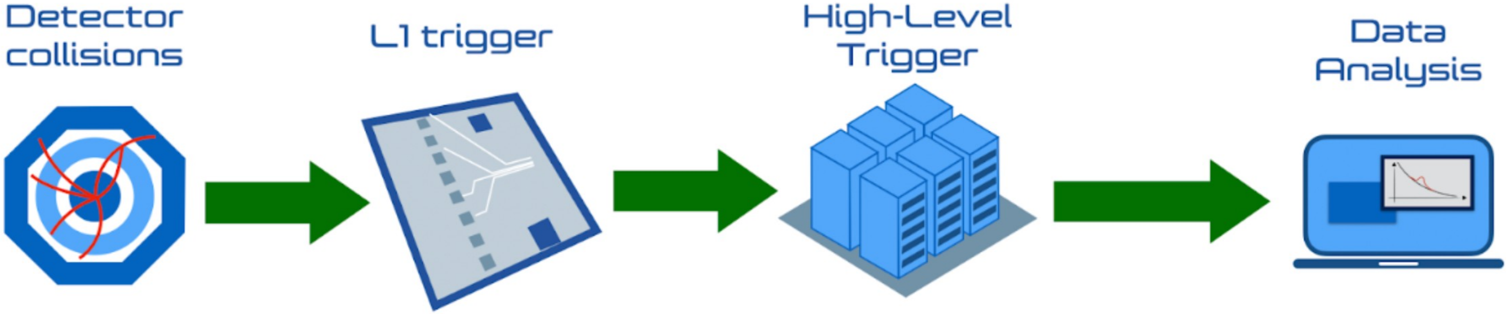
Challenge: Data acquisition and storage



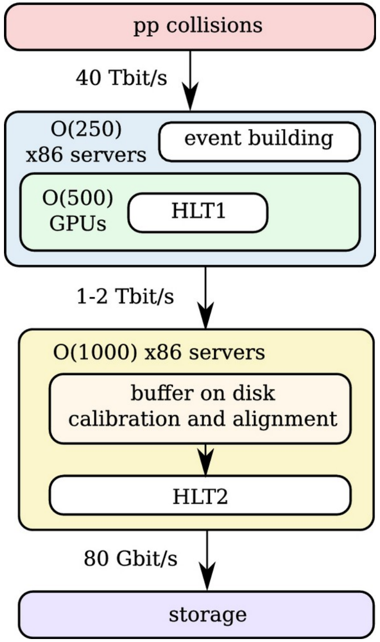


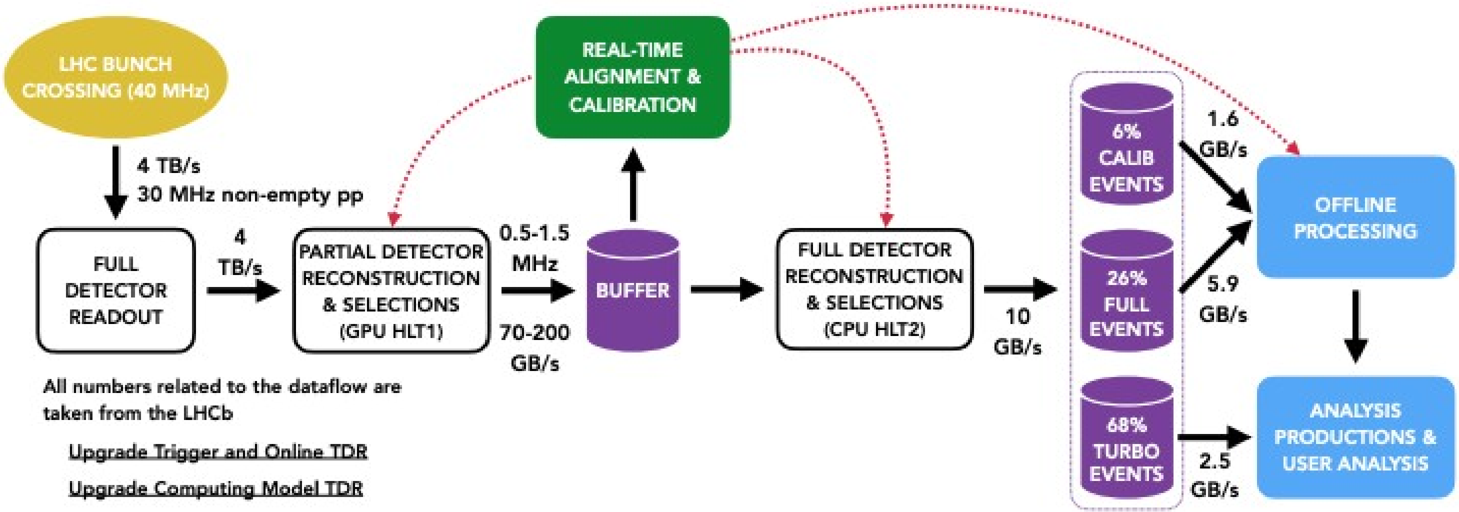
Allen: A high level trigger on GPUs for LHCb

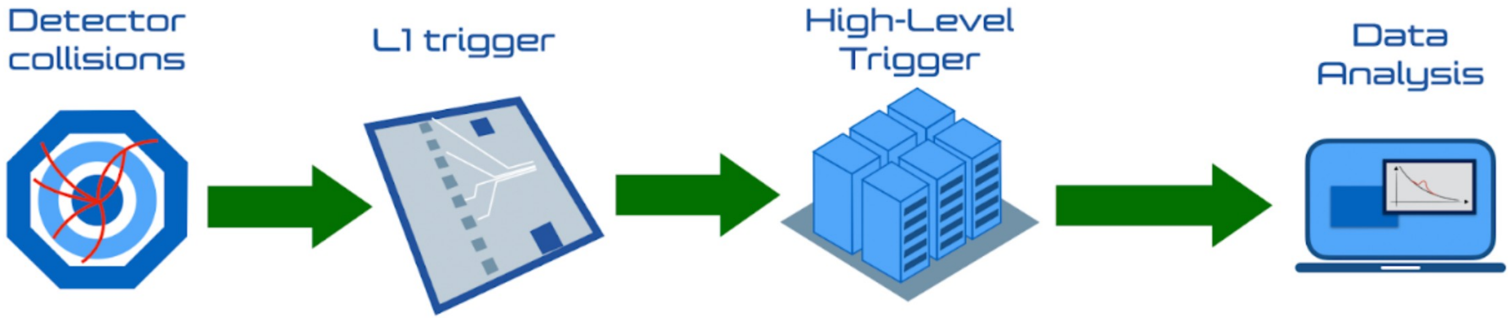
Comput Softw Big Sci 4, 7 (2020)



Challenge: Data acquisition and storage



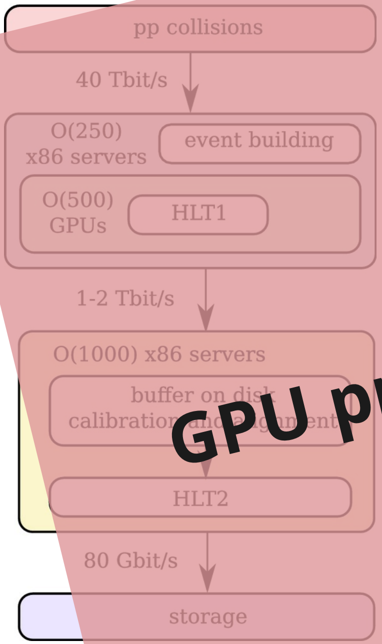


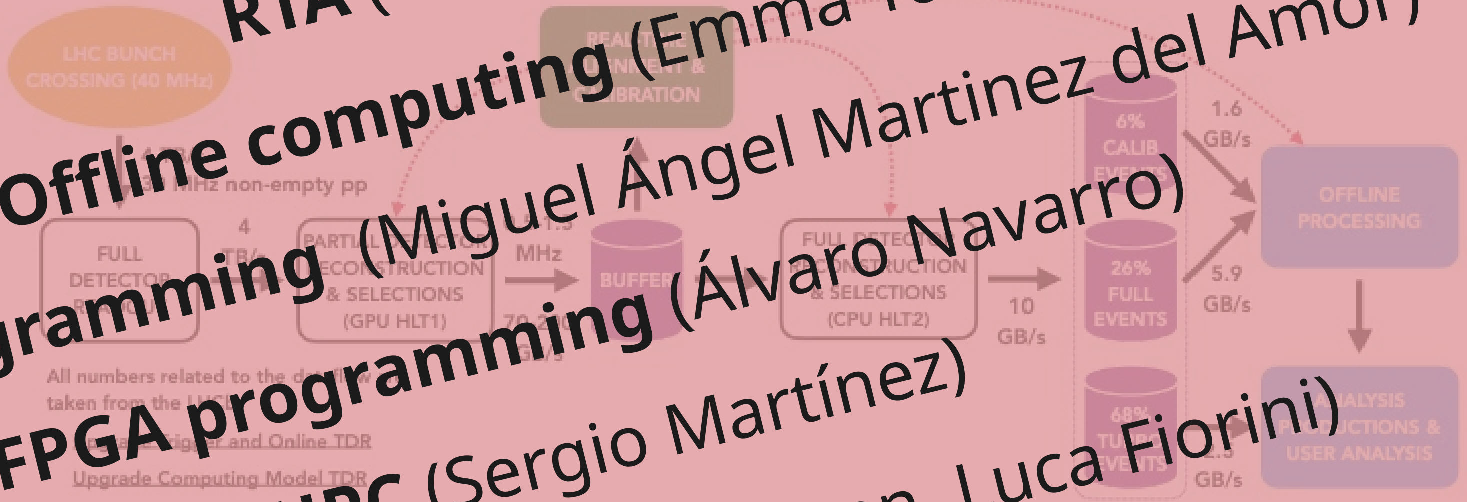


Allen: A high level trigger on GPUs for LHCb

Comput Softw Big Sci 4, 7 (2020)

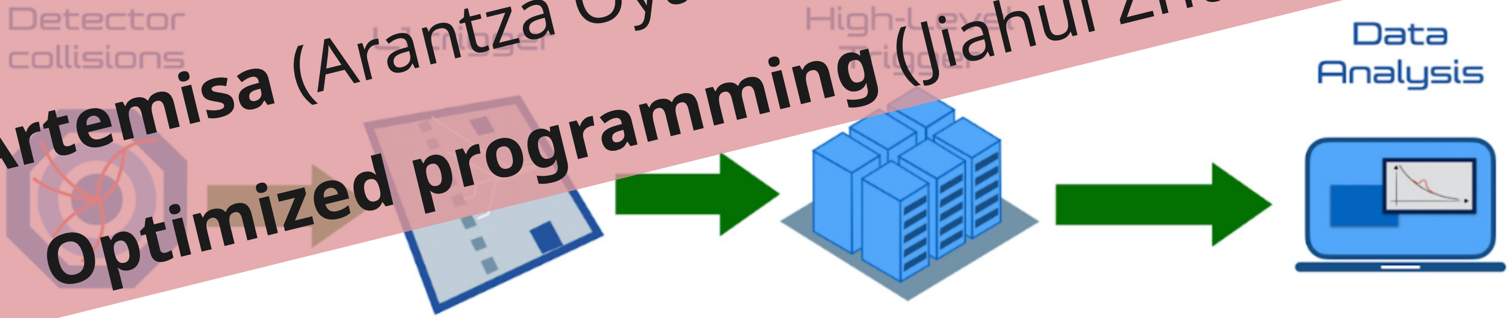
Challenge: Data acquisition and storage





Allen: A high level trigger on GPUs for LHCb

Comput Softw Big Sci 4, 7 (2020)



Lectures & hands-on

RTA (Carla Marin)

Offline computing (Emma Torró)

GPU programming (Miguel Ángel Martinez del Amor)

FPGA programming (Álvaro Navarro)

HPC (Sergio Martínez)

Artemisa (Arantza Oyanguren, Luca Fiorini)

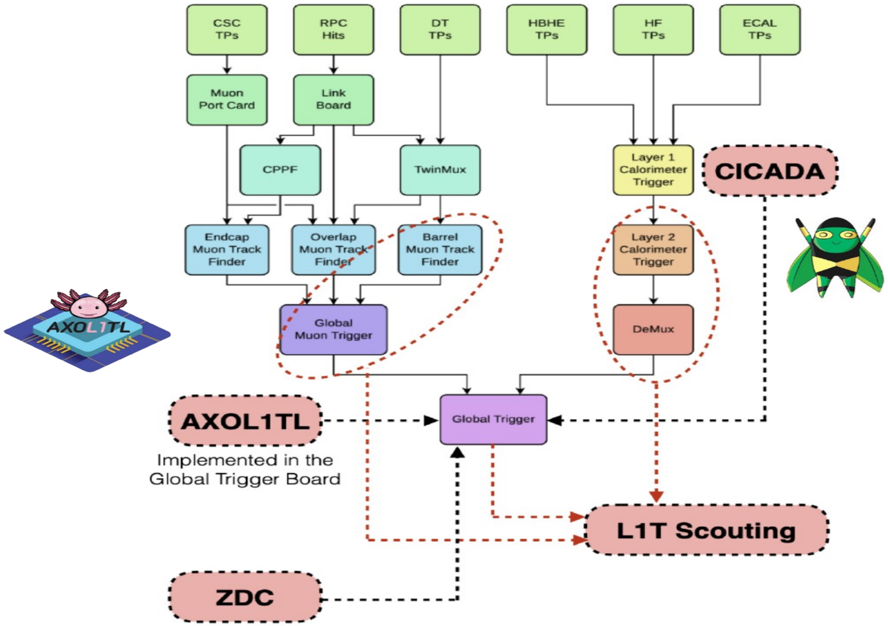
Optimized programming (Jiahui Zhuo)

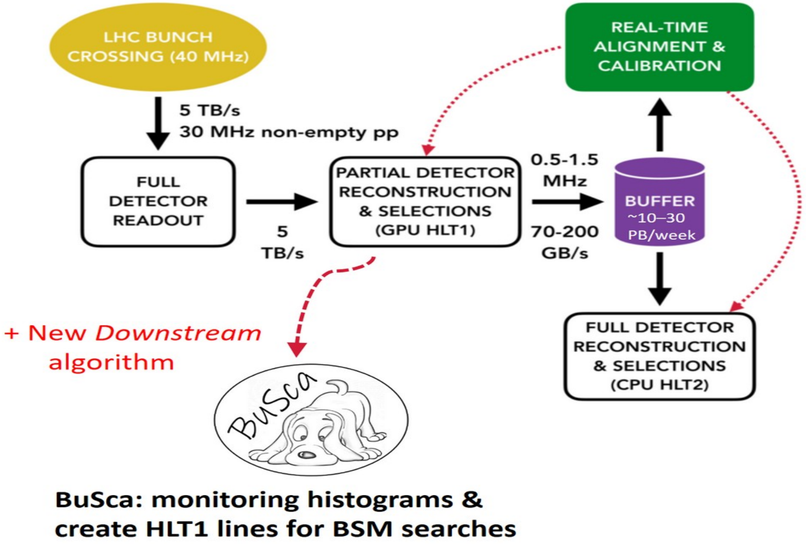
Challenge: data acquisition and storage

Anomaly detection in Real Time (at trigger level) using ML:

CMS: AXOL1TL (global trigger)
CICADA (calorimeter trigger) at L1

LHCb: BuSca (prior to trigger decision, no hardware trigger)





arXiv:2411.19506

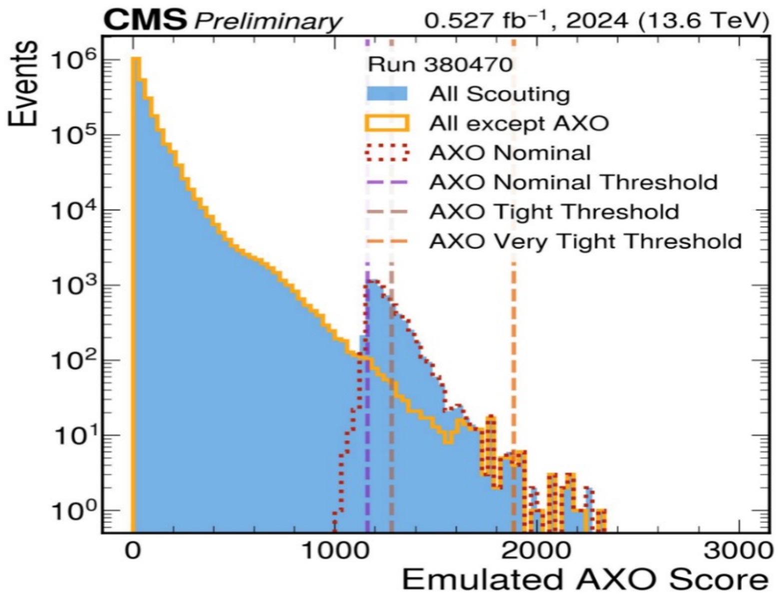
https://doi.org/10.1051/epjconf/202533701161

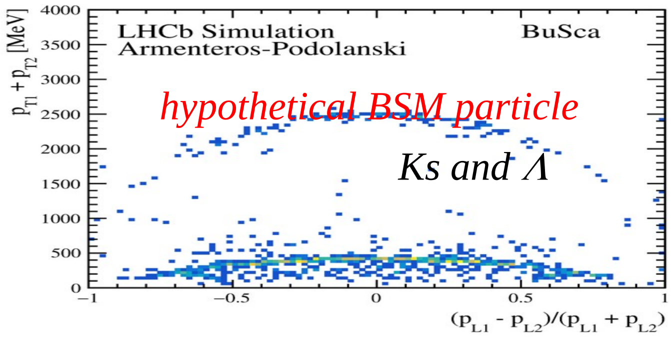
Challenge: data acquisition and storage

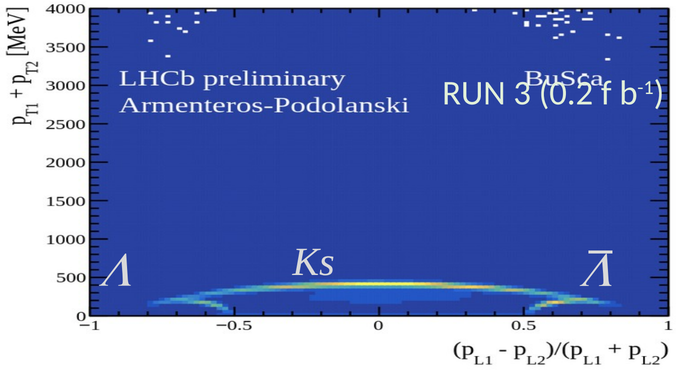
Anomaly detection in Real Time (at trigger level) using ML:

CMS: AXOL1TL (global trigger)
CICADA (calorimeter trigger) at L1

LHCb: BuSca (prior to trigger decision, no hardware trigger)





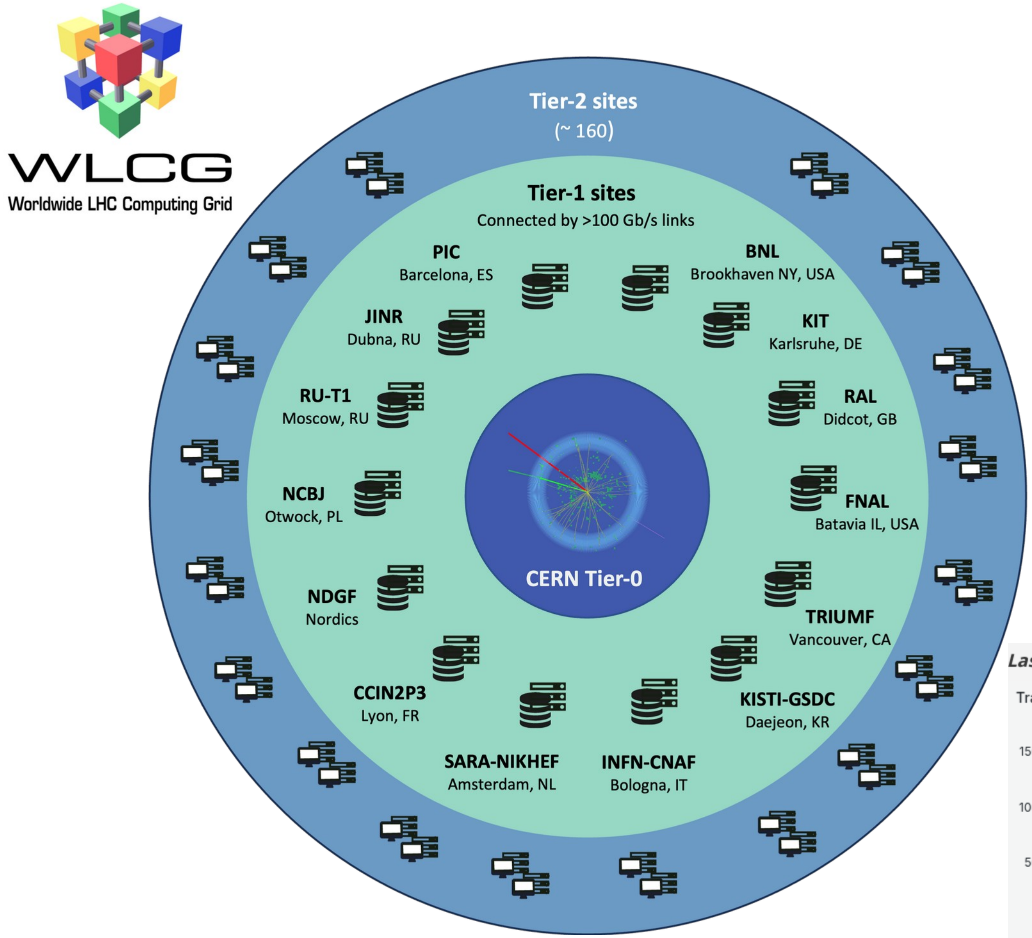


[LHCB-FIGURE-2024-018]

arXiv:2411.19506

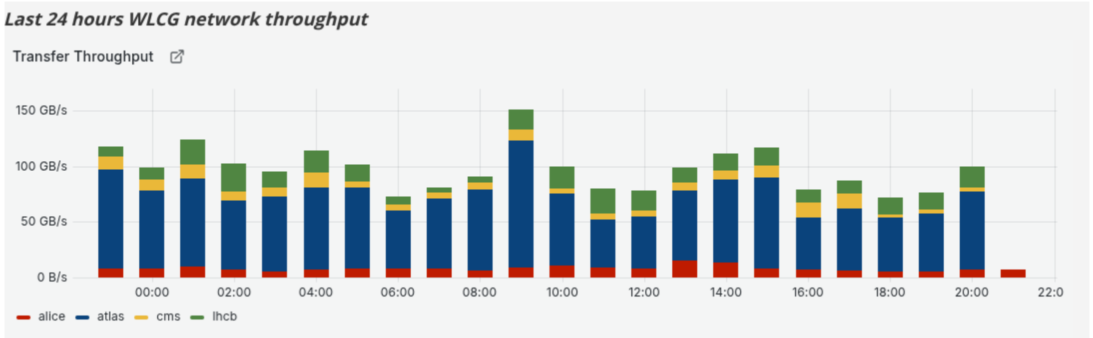
https://doi.org/10.1051/epjconf/202533701161

Challenge: data storage, distribution and analysis



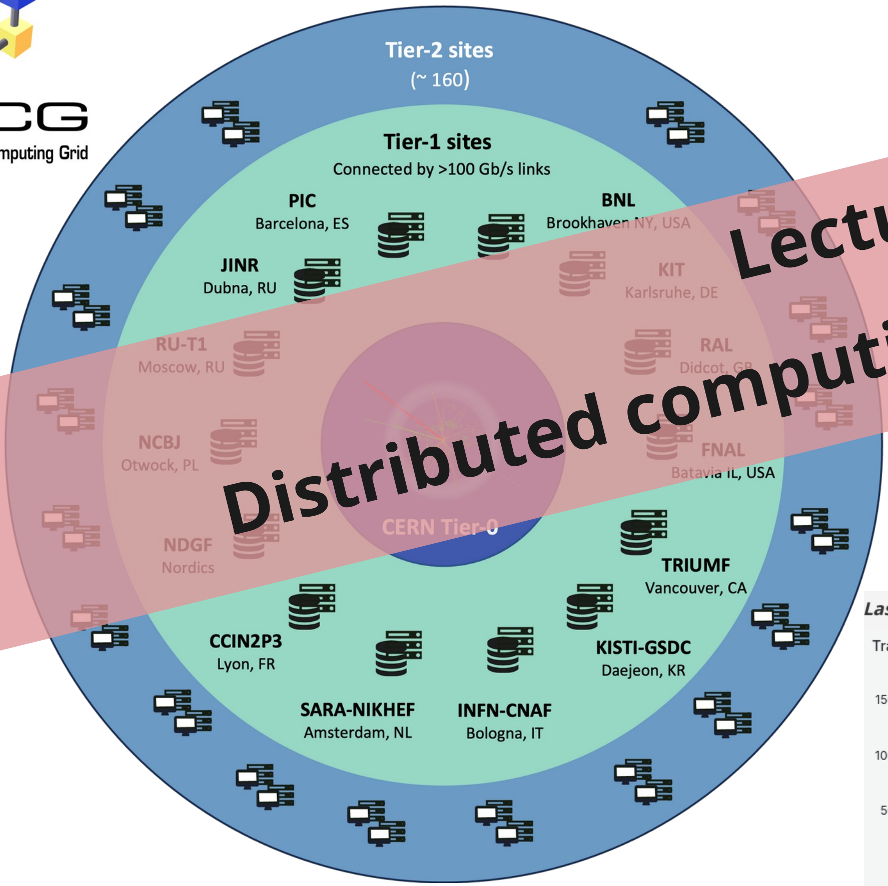
- Provides 1.4M cores and 1.5 EB of storage
- At 170 sites in 42 countries worldwide
- Users: 12000

- CERN Tier 0 sends data to Tier-1 sites via optical-fibre links
- Tier 1 to ~160 Tier-2 locations, usually in the same geographical region
- Tier 3 local resources



Challenge: data storage, distribution and analysis

Worldwide LHC Computing Grid

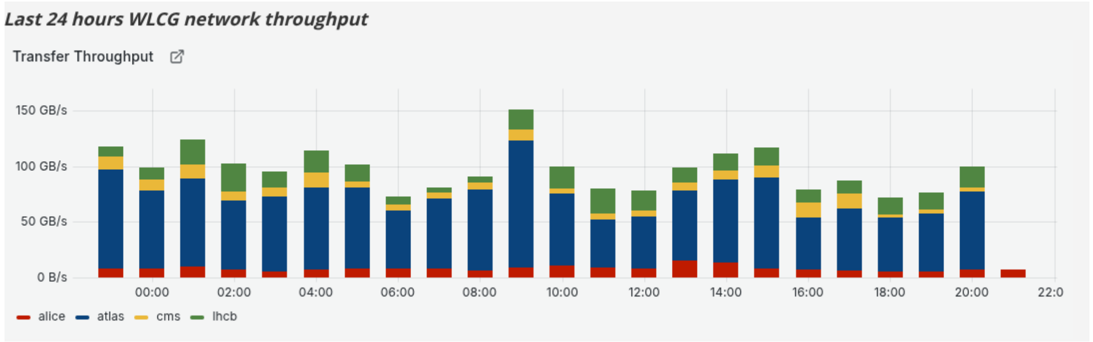


- Provides 1.4M cores and 1.5 EB of storage
- At 170 sites in 42 countries worldwide
- Users: 12000

Lecture

Distributed computing (Gonzalo Merino)

- CERN Tier-0 sends data to Tier-1 sites via optical-fibre links
- Tier 1 to ~160 Tier-2 locations, usually in the same geographical region
- Tier 3 local resources

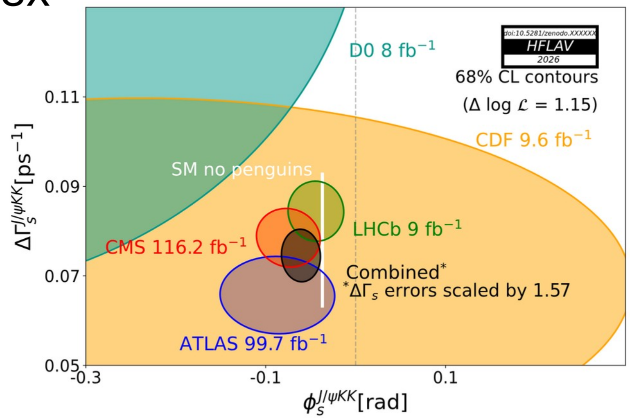


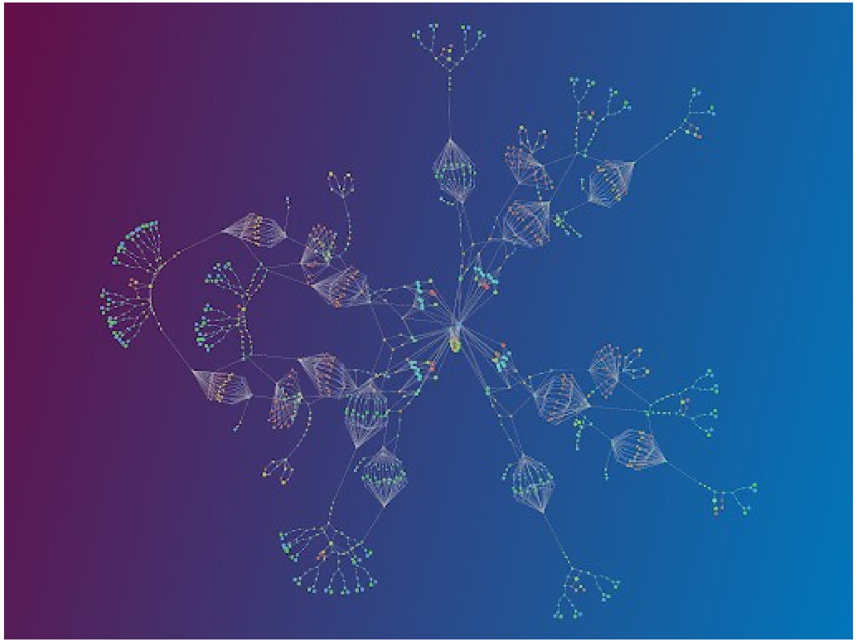
Challenge: Data analysis

GPU-based data analysis tool
Ipanema-β [arXiv:1706.01420]

Increased data introduces multiple challenges

- Size of data samples processed (both data and simulation)
- Speed of data processing
- Modelling, including ML training
- Higher precision requires careful control of systematic uncertainties – tends to make analyses more complex

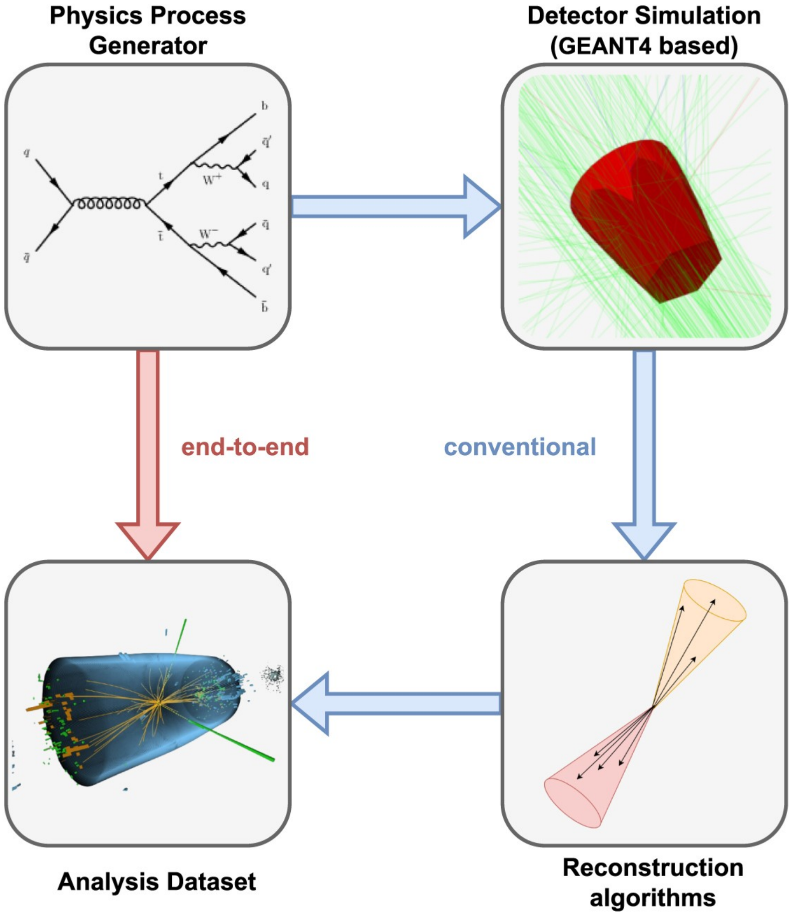




snakemake [1000Research 2021, 10:33] analysis pipeline for Run 2 $B_s^0 \rightarrow J/\psi KK$ at LHCb with over 1200 steps! [Phys. Rev. Lett. 132 (2024) 051802]

Challenge: Simulation

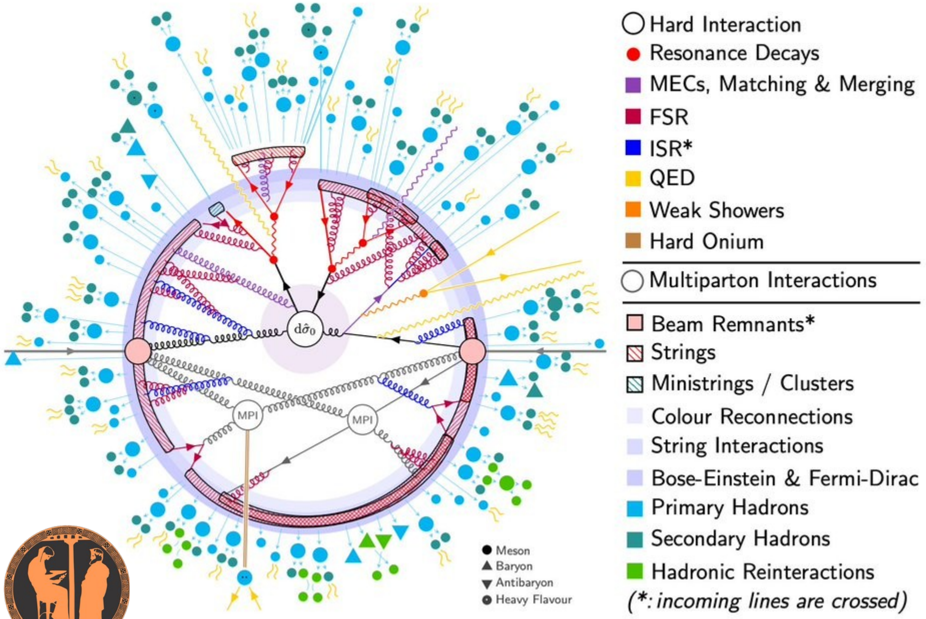
- Data models based on simulation, typically need ~ x10 data size to understand them
- Simulation is a complex and computationally expensive process

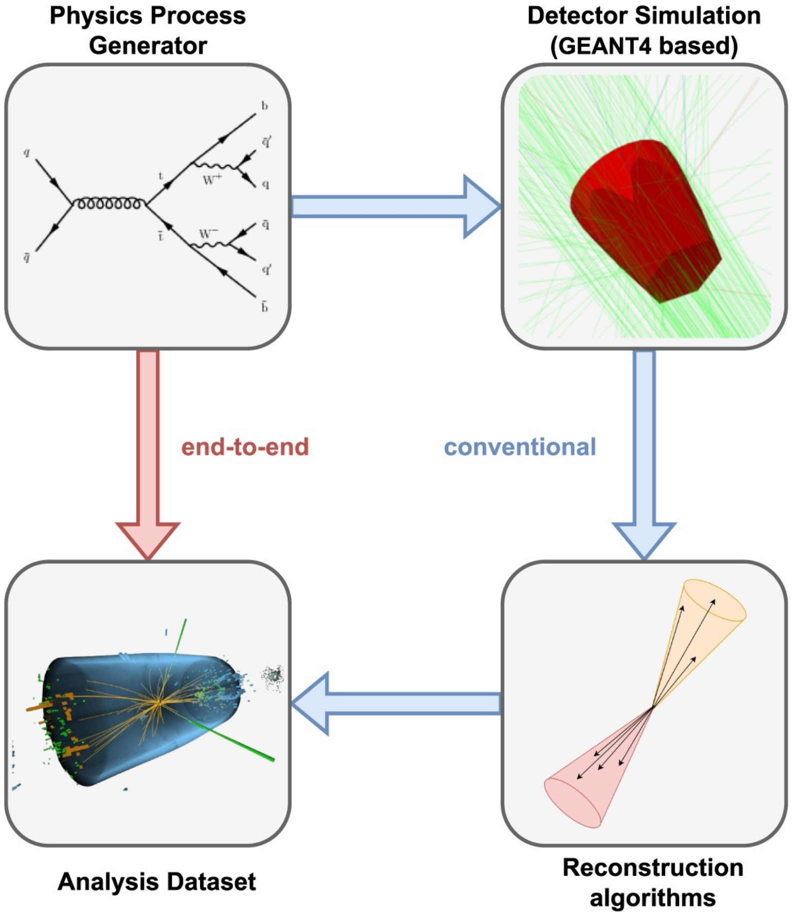


Mach. Learn.: Sci. Technol. 5 035007

Challenge: Simulation

- Data models based on simulation, typically need ~ x10 data size to understand them
- Simulation is a complex and computationally expensive process



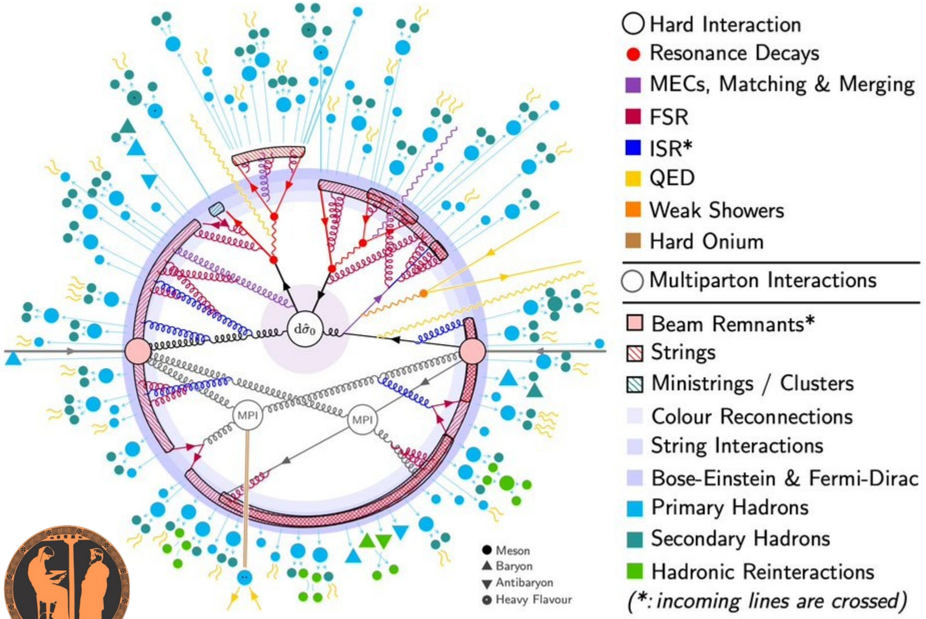


Mach. Learn.: Sci. Technol. 5 035007

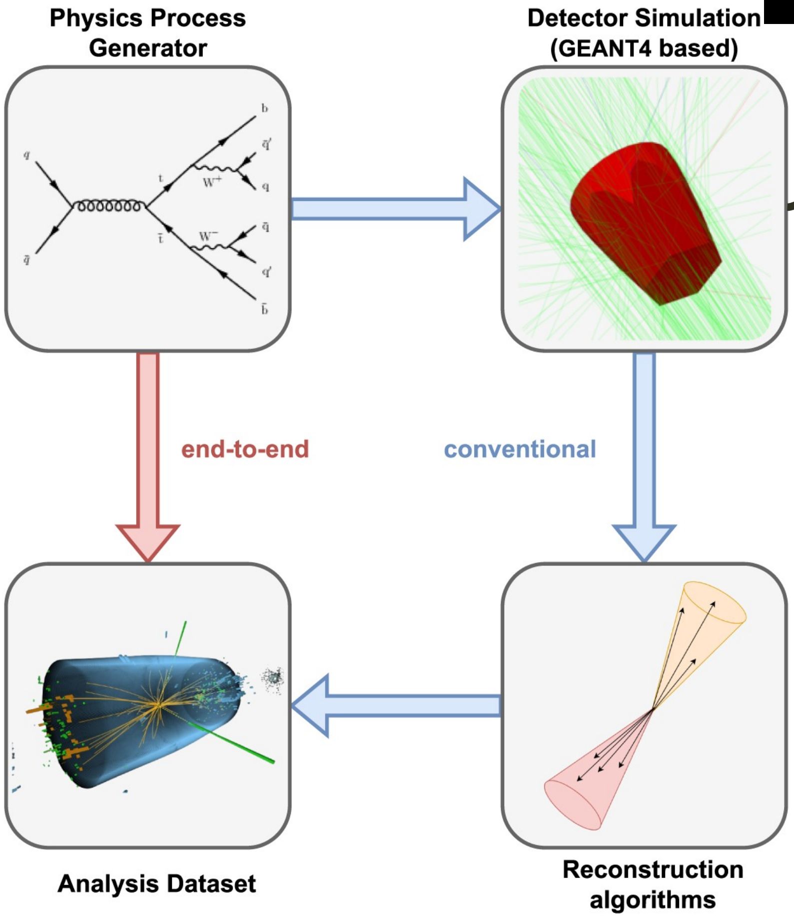
PYTHIA

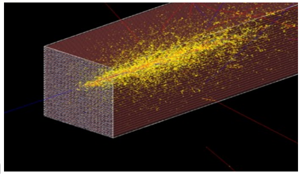
Challenge: Simulation

- Data models based on simulation, typically need ~ x10 data size to understand them
- Simulation is a complex and computationally expensive process



PYTHIA





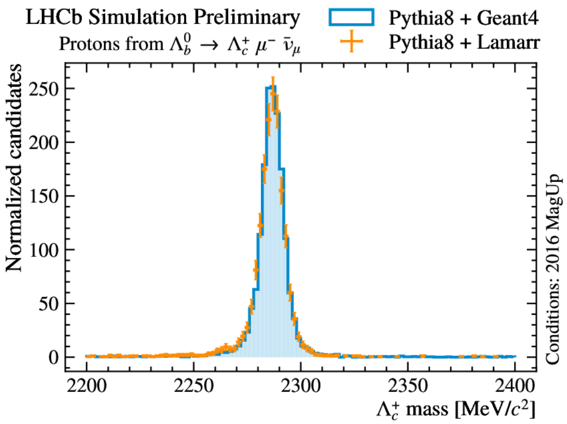
Mach. Learn.: Sci. Technol. 5 035007

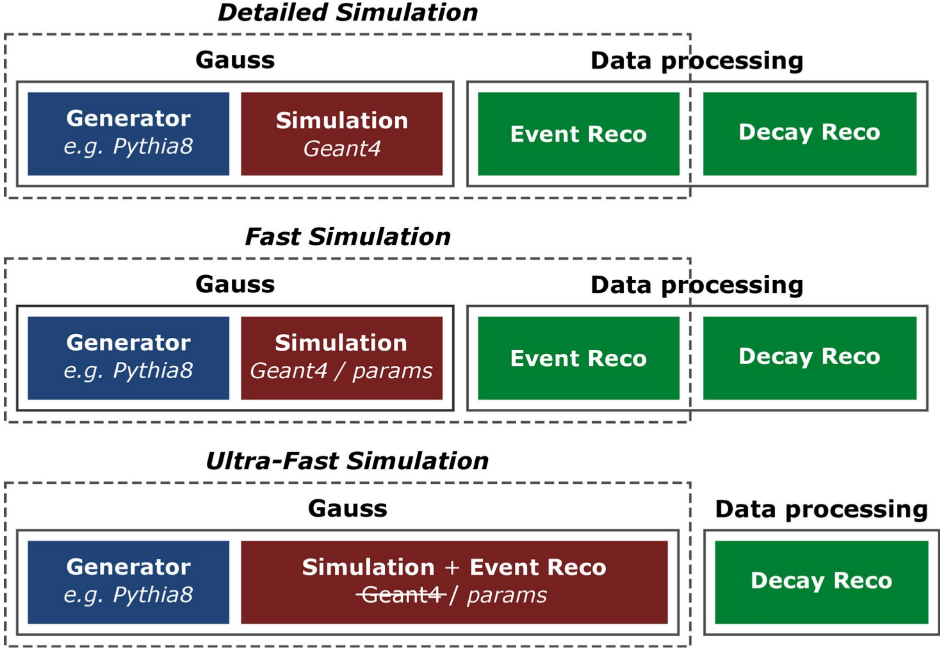
Challenge: Simulation

Increased data requires more simulation

- Data models based on simulation, need ~ x10 data size to understand them
- Data-simulation differences introduce systematic uncertainties
  $\rightarrow$ Need faster simulation

- Already exploiting ML tools to introduce orders of magniture speed-up
  - ATLAS CaloGAN
    [Phys. Rev. D 97, 014021 (2018)]
  - CMS FastSim
    [J. Phys.: Conf. Ser. 513 022012]
  - LHCb Lamarr
    [arXiv:2303.11428]





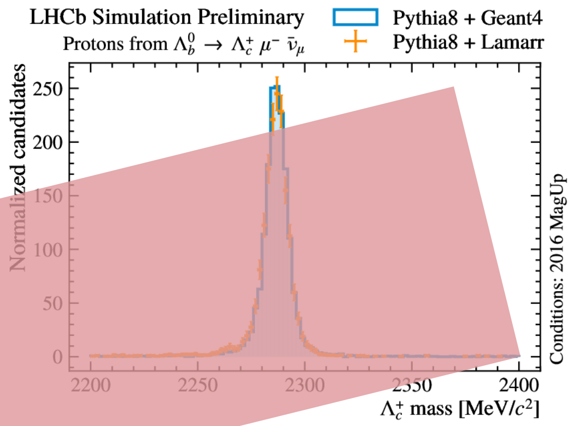
Challenge: Simulation

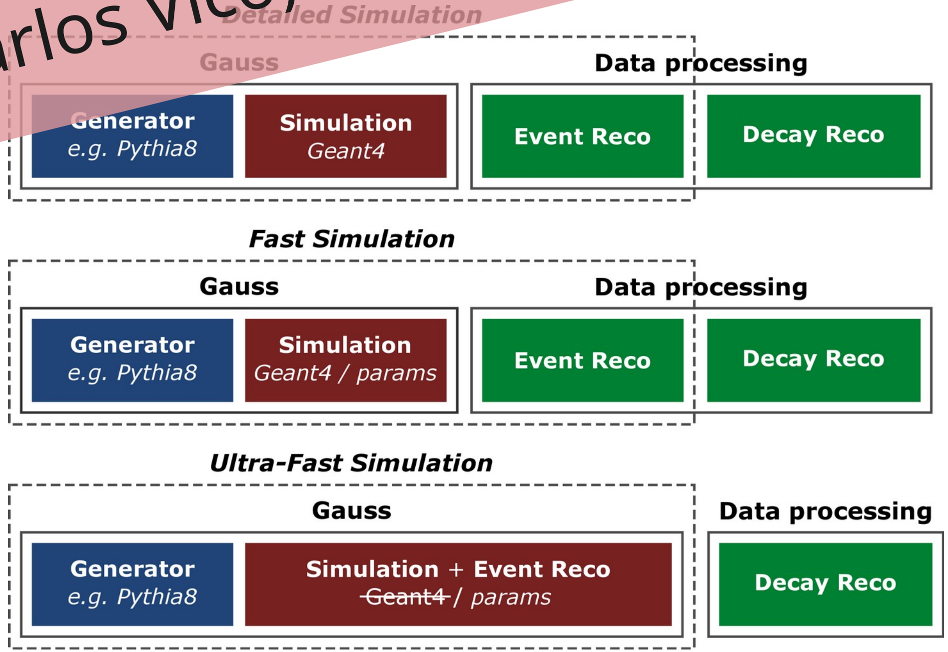
Increased data requires more simulation

- Data models based on simulation, need ~ x10 data size to understand them
- Data-simulation differences introduce systematic uncertainties
  → Need faster simulation

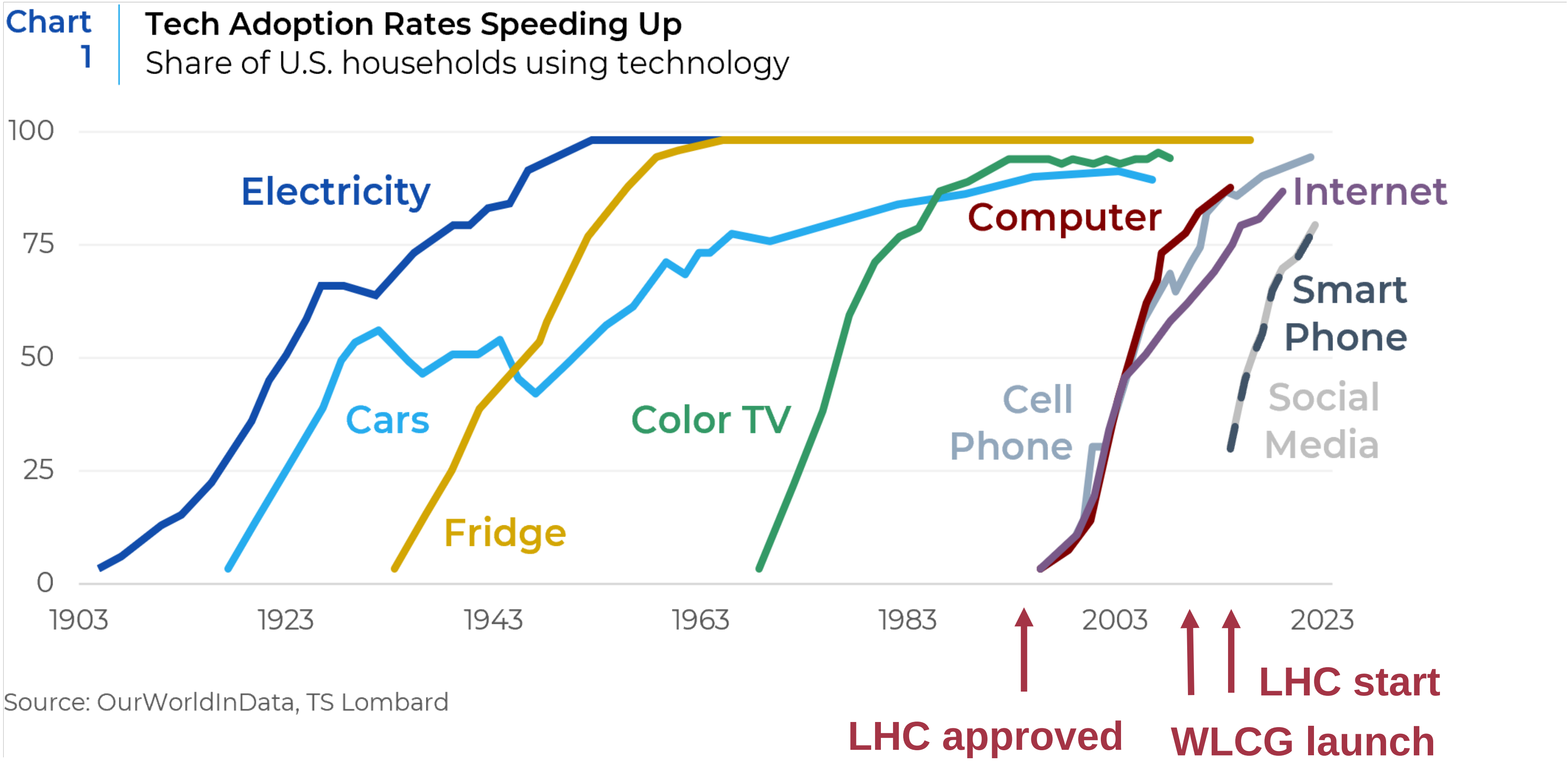
- Already exploiting ML tools to introduce orders of magniture speed-up
  - ATLAS CaloGAN [Phys. Rev. D 97, 014021 (2018)]
  - CMS FastSim [J. Phys.: Conf. Ser. 513 022012]
  - LHCb Lamarr [arXiv:2303.11428]

Lecture & hands-on Simulations (Carlos Vico)

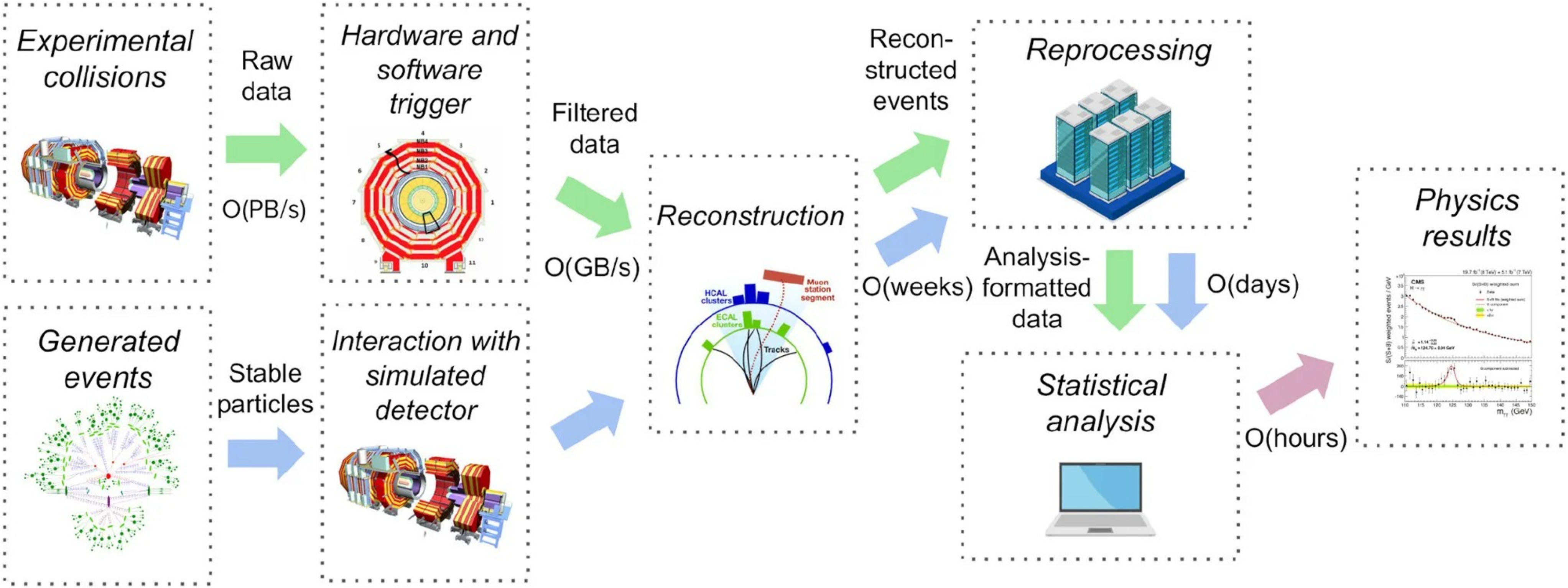




The current era

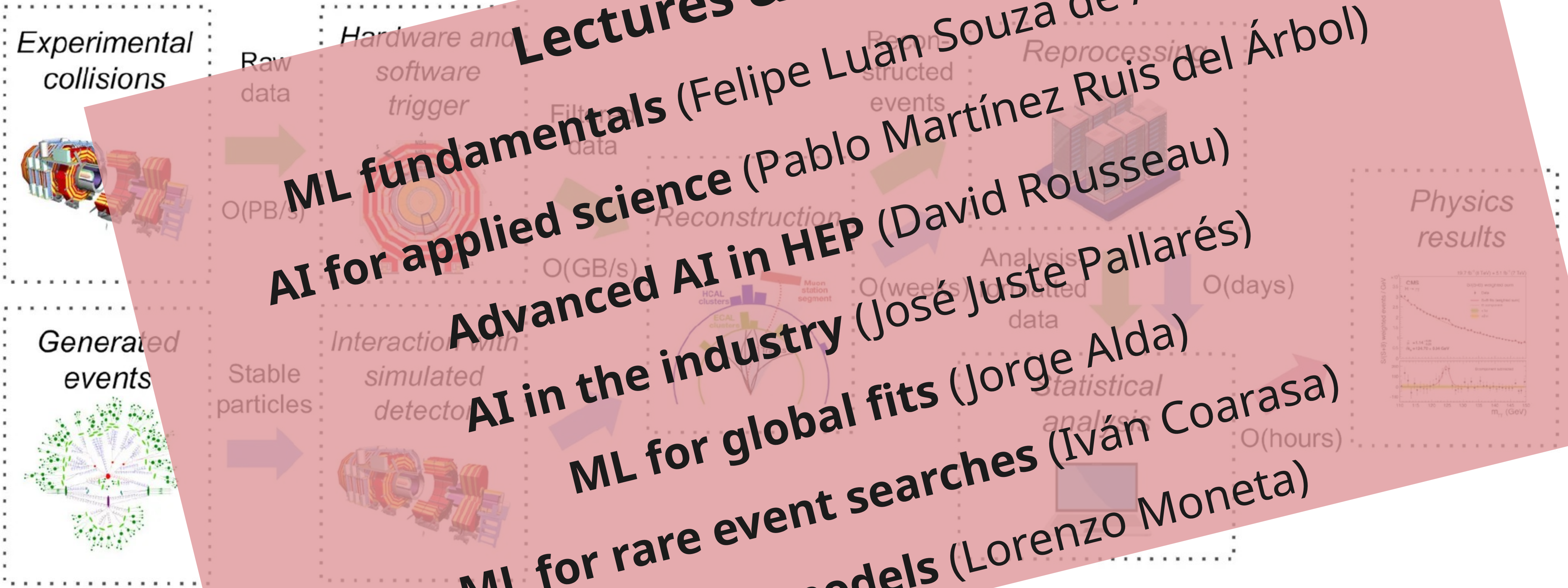


AI in every step of the way



Front. Big Data 4, 2021

AI in every step of the way



Lectures & hands-on

ML fundamentals (Felipe Luan Souza de Almeida)

AI for applied science (Pablo Martínez Ruis del Árbol)

Advanced AI in HEP (David Rousseau)

AI in the industry (José Juste Pallarés)

ML for global fits (Jorge Alda)

ML for rare event searches (Iván Coarasa)

Foundation models (Lorenzo Moneta)

Front. Big Data 4, 2021

LLM

- In large projects knowledge gets easily fragmented across gitlab, TWiki, internal notes...
- LLM as a productivity boost
  - chATLAS to help write code, search documentation, answer questions on ATLAS code

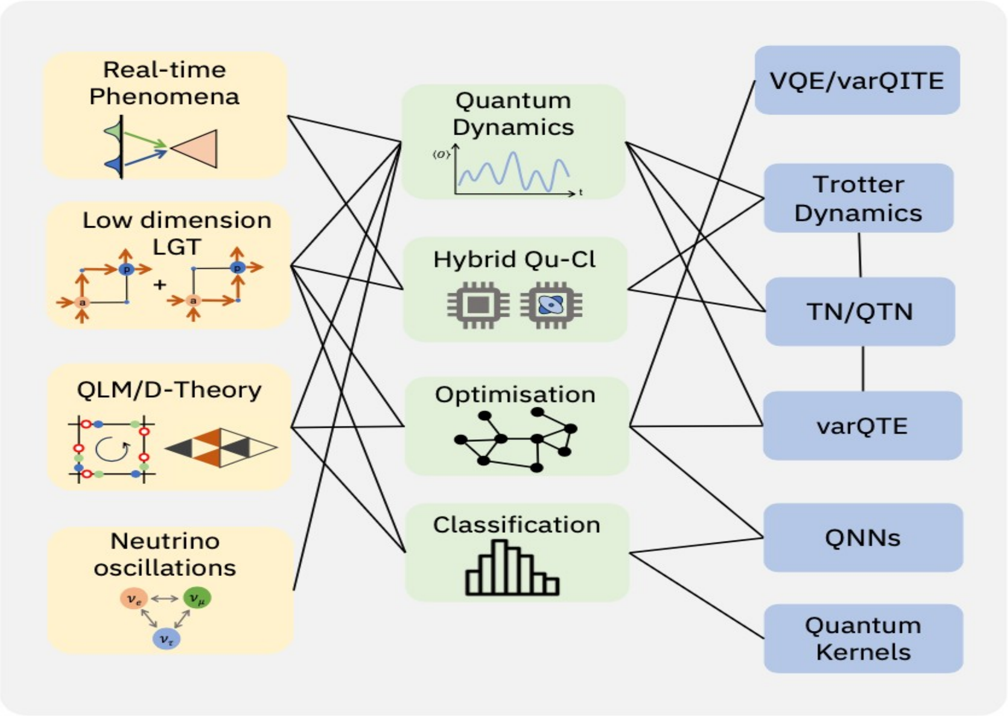
chatlas

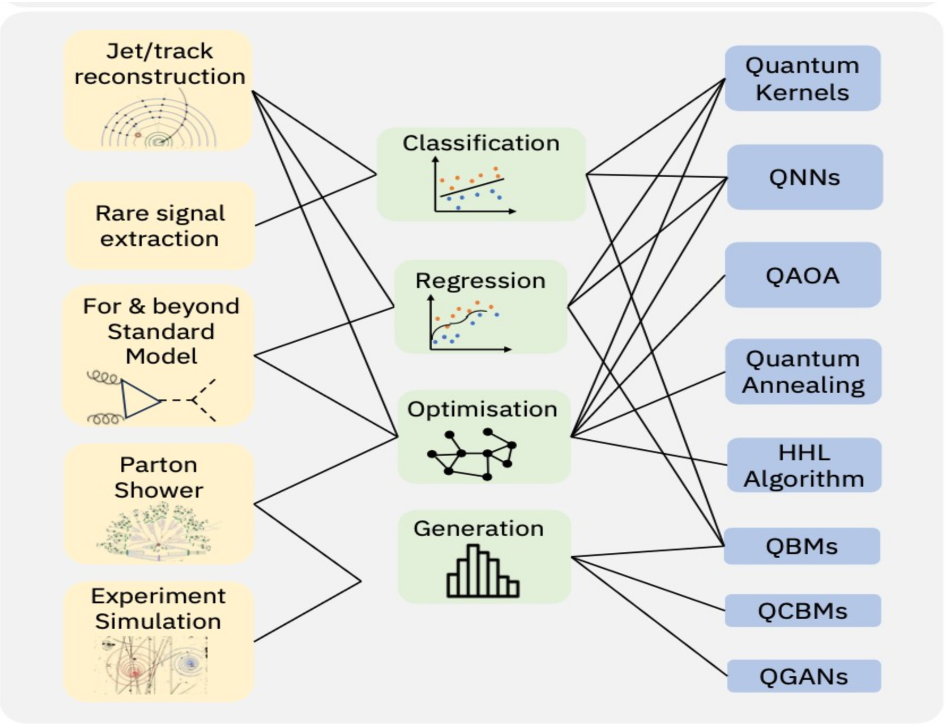
LLM

- In large projects knowledge gets easily fragmented across gitlab, TWiki, internal notes...
- LLM as a productivity boost
  - chATLAS to help write code, search documentation, answer questions on ATLAS code

Lectures & hands-on
LLMs (Cilicia Uzziel Pérez)

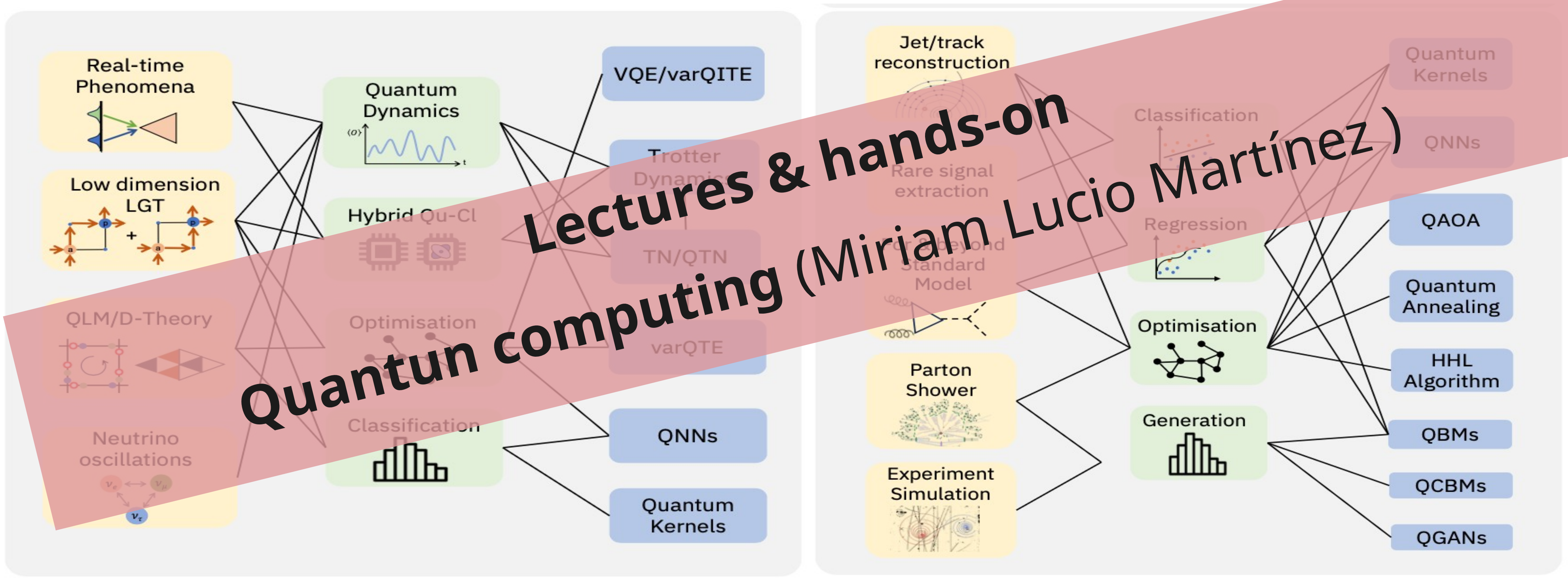
Quantum Computing





Roadmap QC4HEP [PRX Quantum 5, 037001]

Quantum Computing



Lectures & hands-on

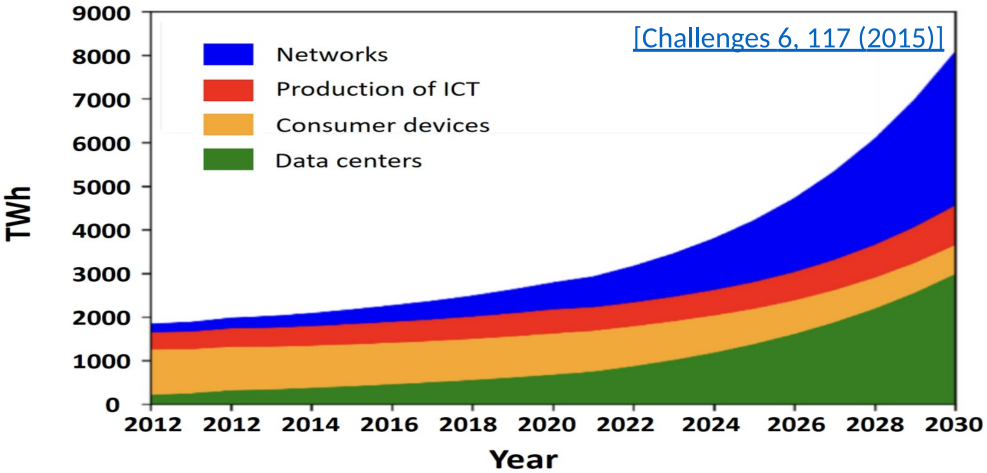
Quantun computing (Miriam Lucio Martínez)

Roadmap QC4HEP [PRX Quantum 5, 037001]

Challenge: sustainability

BE RESPONSIBLE

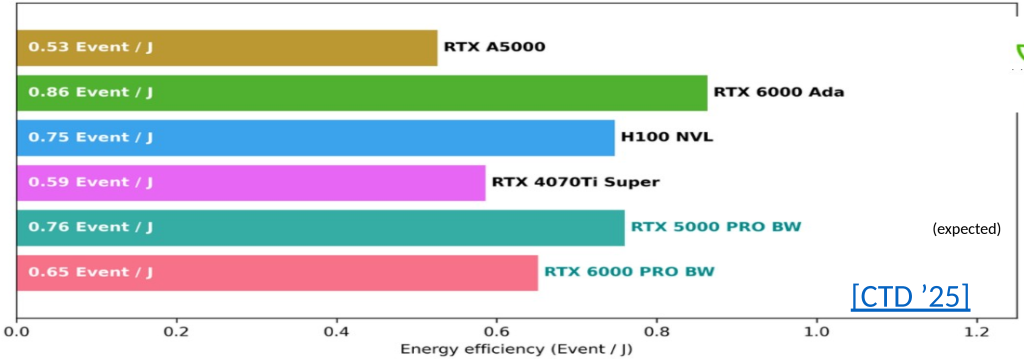
Huawei report expects up to 50% of global electricity to be used by communication



Accelerator	Alveo U50	Alveo U250	RTX 3090
Implementation	<8,3>	<8,3>	TRT INT8
Throughput (Events/s $\times 10^6$)	0.55	1.10	0.82
Active Power Draw (W)	75	230	350
Energy per Event (μJ)	140	210	430
Energy Gain	3.1x	2.0x	1.0x
Price (USD)	3000	~ 10 000	1500

arXiv:2502.02304

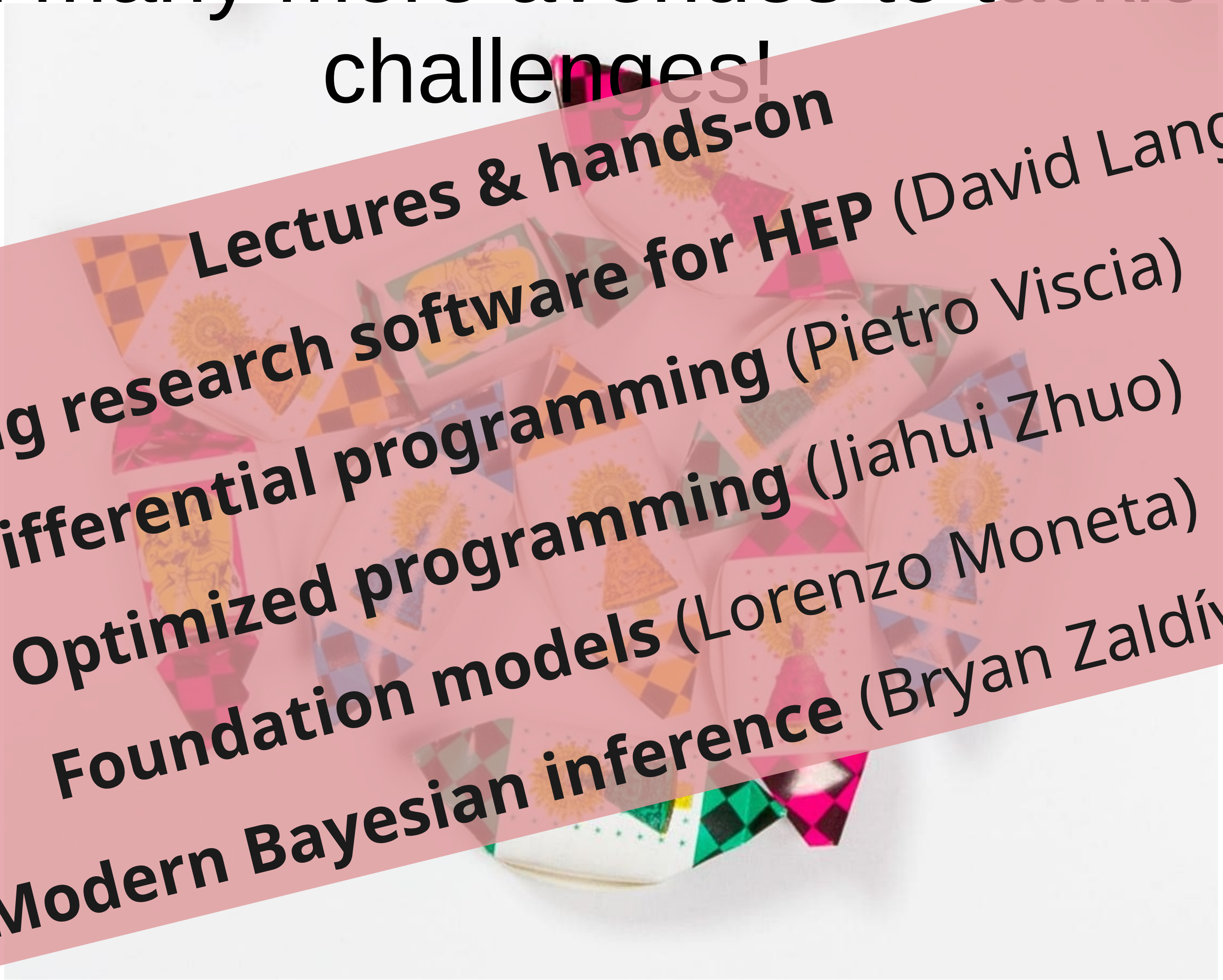
EE = # real time processed events in LHCb / J



HL

Throughput ↑
Energy ↓

… and many more avenues to tackle the challenges!



Lectures & hands-on

Building research software for HEP (David Lange)

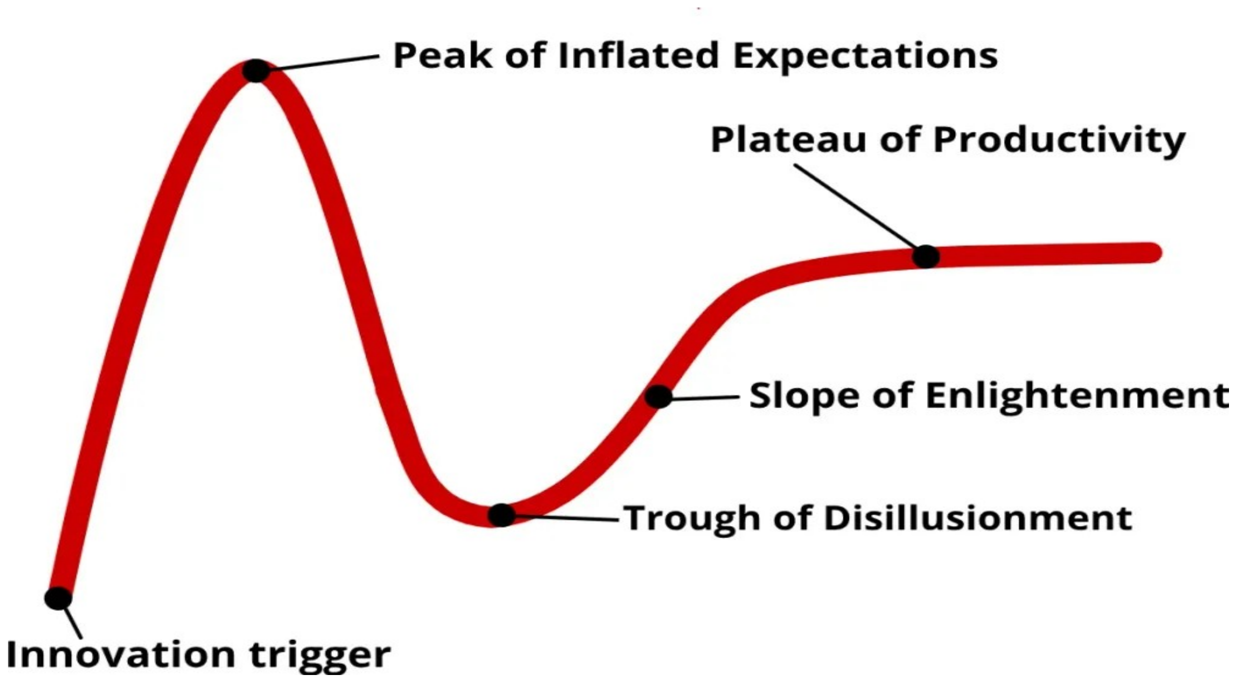
Differential programming (Pietro Viscia)

Optimized programming (Jiahui Zhuo)

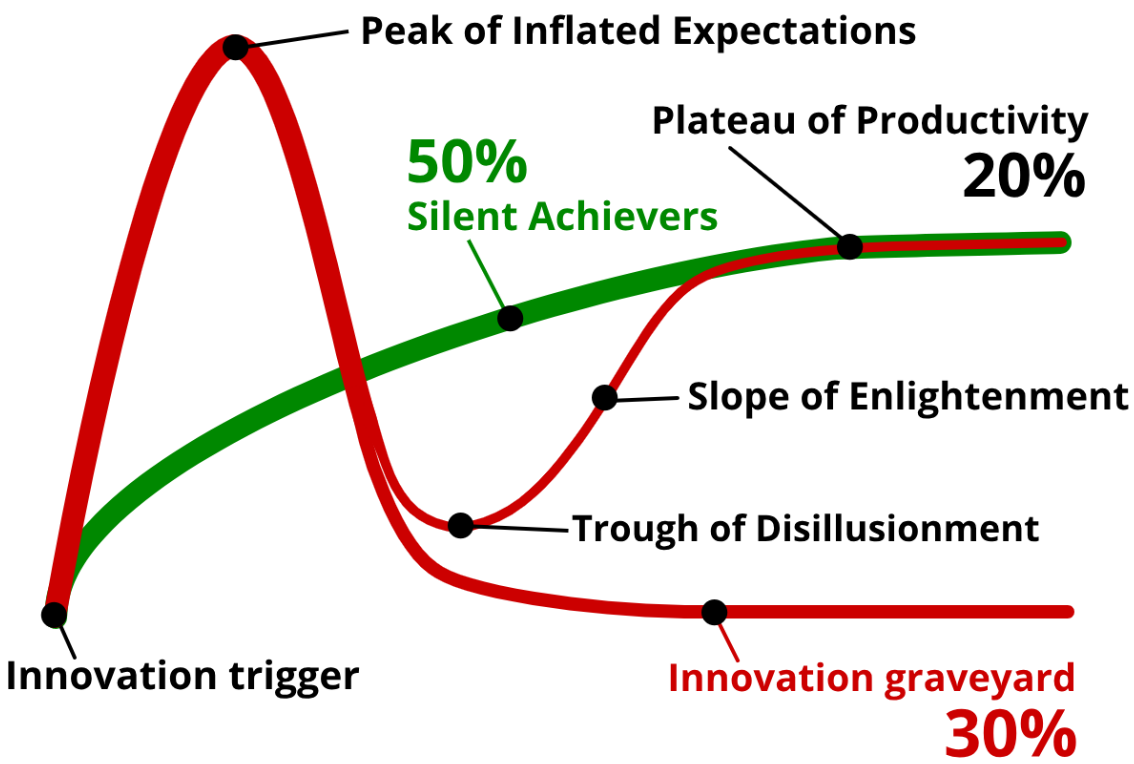
Foundation models (Lorenzo Moneta)

Modern Bayesian inference (Bryan Zaldívar)

Gartner hype cycle



Gartner hype cycle: hidden curves



Thanks for your attention!

Special thanks to Arantza for valuable inputs into this talk!