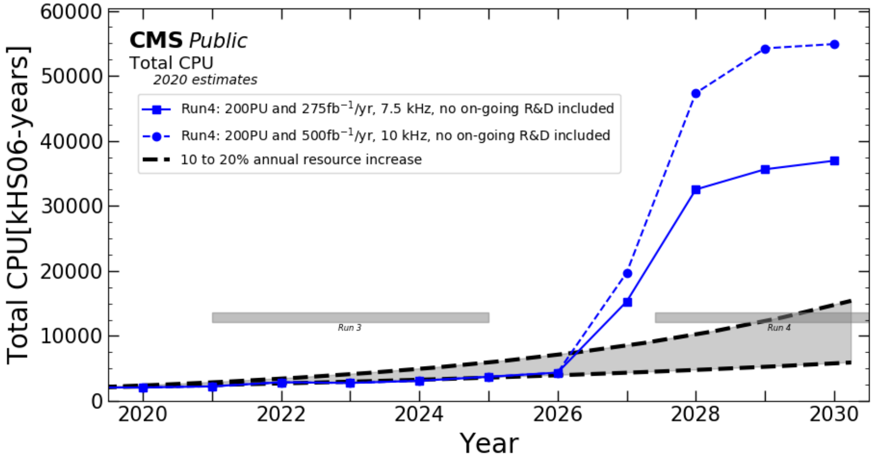
Advances in detector simulation tools

Samuel Bein (Universität Hamburg)
on behalf of ALICE, ATLAS, CMS, LHCb Collaborations
June 9, LHCp 2021

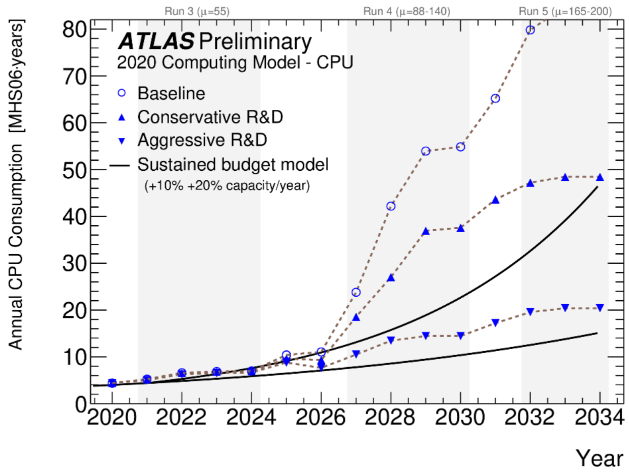
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Motivation

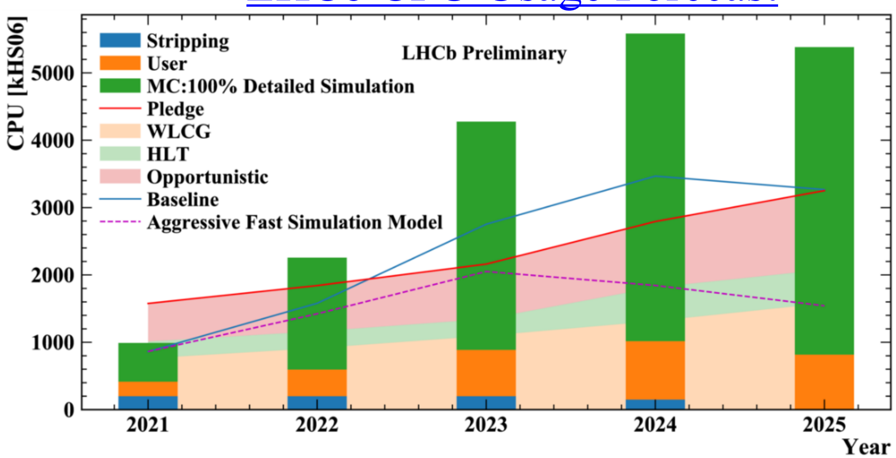
[CMSOfflineComputingResults]



[AtlasComputingandSoftwarePublicResults]



[LHCb CPU Usage Forecast]



- ALICE's CPU needs will scale with 15% year-over-year model

- All the experiments will require fundamental changes to computing model in order to cope with increased data rates

Benevolent Challenges

Increased Luminosity brought by HL-LHC

- Expect 5x increase in instantaneous starting with Run 4
- 20x increase in integrated luminosity with respect to current data set
- Even sooner for LHCb and ATLAS

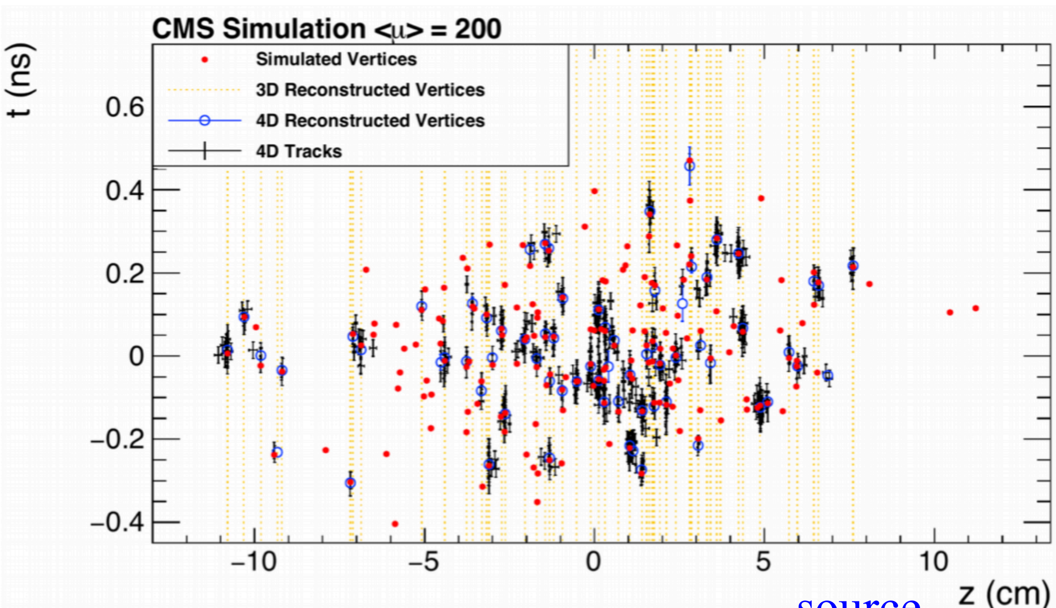
Increased event complexity

- Increased detector complexity for better quality event reconstruction
- Pushing physics analyses out on to the tails of distributions means we require improved simulation accuracy in extreme phase spaces

arXiv:1803.04165

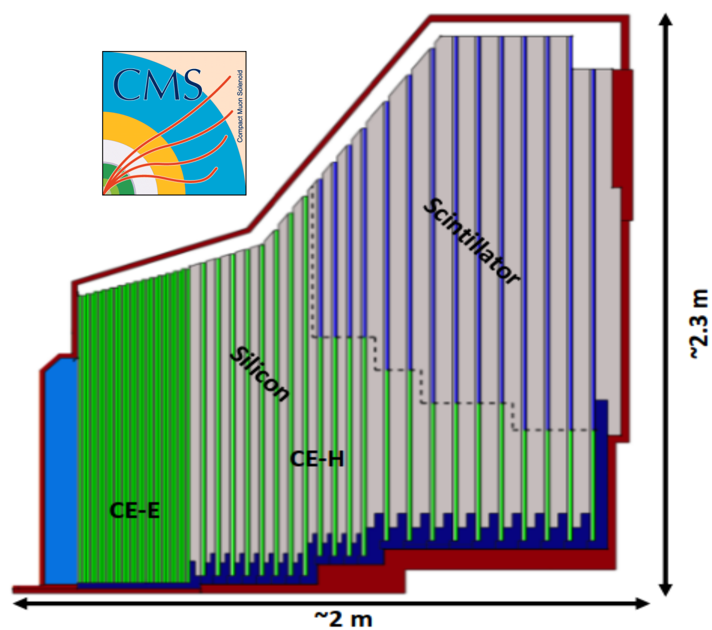
Benevolent Challenges

- CMS HGCal
  - New endcap calorimeter under design and development for Phase 2: High granularity calorimeter
    - 6 million channels, 10x more sensitive volumes than existing calorimeters
    - Possibility for high-quality particle flow analysis in the forward region



source

HGCal



Topics of discussion

1.) fully detailed MC simulation

- Largely based on common software, particularly Geant4
  - State of the art models for material interactions between particles and (detector) material
  - Aggregate of many physics interaction models (Physics Lists) implemented along with particle transport

2.) fast MC simulation

- Largely parametric particle interaction models
- Typically experiment-specific
- Often tuned using fully detailed event samples

3.) machine learning

- Make use of emerging tools to aid in detector simulation:

=> use 3.) to speed up 1.)

=> use 3.) to improve the accuracy of 2.)

=> try to shift event production from 1.) to 2.)

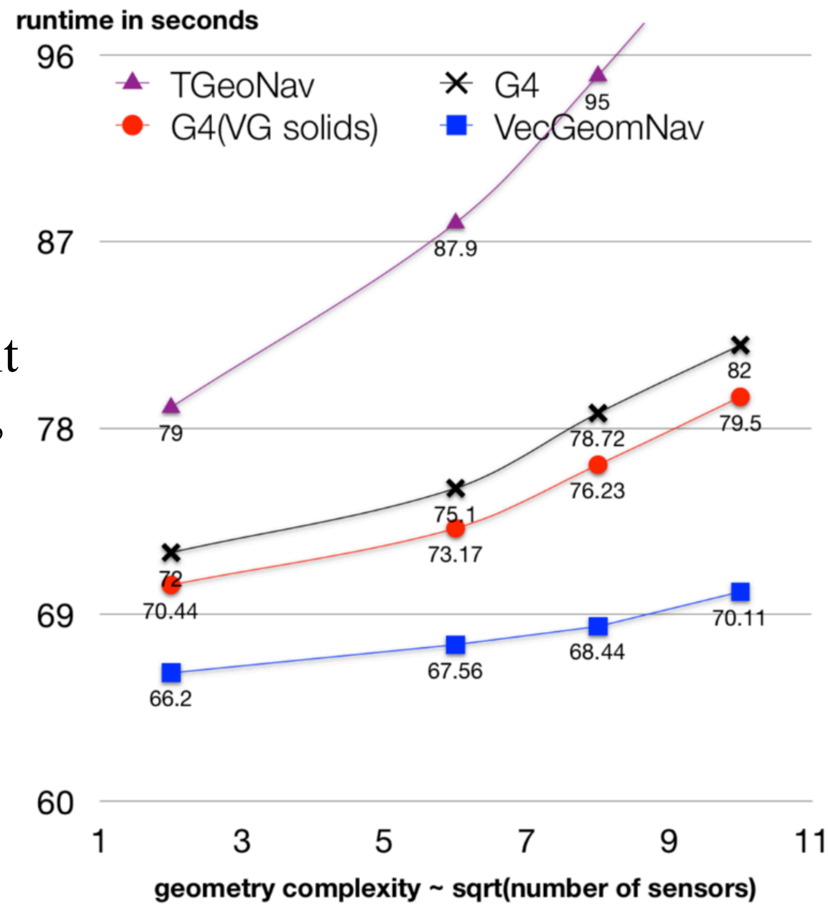
FullSim developments

Geant4 physics R&D

- Increase throughput *(events simulated per second)*
  - non-zero chance of a factor of 2 increase in throughput by Run 4
  - without sacrificing physics accuracy, where possible,
  - allowing limited reduction in accuracy where acceptable.
- Develop new physics lists for increased accuracy:
  - aim to have systematic effects scaling down as 1/sqrt(lumi),
  - serve new uses cases, e.g., high granularity calorimetry.
- Explore the use of accelerators, which may be useful for speeding up the computation associated with repetitive processes, e.g., neutron tracking
- Experiments can provide custom code and feedback:
  - CMS using dedicated physics list, optimised with test beam data
    - Will require increased accuracy to make full use of HGCal
  - ATLAS also planning to use TestBeam data in near future

Parallelization

- Well separated particles do not interact or interfere, factorizable energy loss
  - Much of simulation is parallelizable, but is not yet widely implemented
- R&D ongoing to this end
- Within the Geant4 realm
  - VecGeom
    - efficient propagation of particles through geometry
    - developed and optimized to exploit vectorised transport on CPU, GPU, etc.
    - Interface now provided within Geant4
  - R&D to explore
    - AdePT: GPUs for complete simulation workflow
    - Celeritas, CaTS: GPUs, accelerators, HPC particle transport

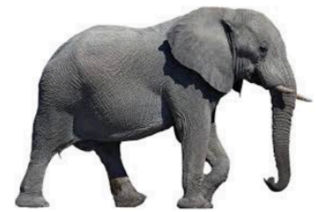


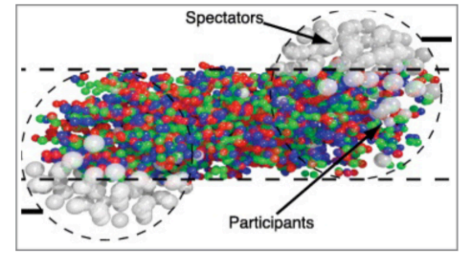
EPJ 2020

Parallelization

ALICE

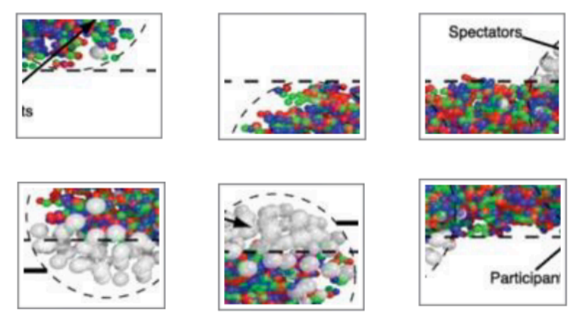
- event splitting and collaborative simulation parallelism
- load balanced parallelism and resource utilization
- opportunistic simulation on computing resources with a small time-window
  - elephant events would normally not finish
- Late forking to limit memory use

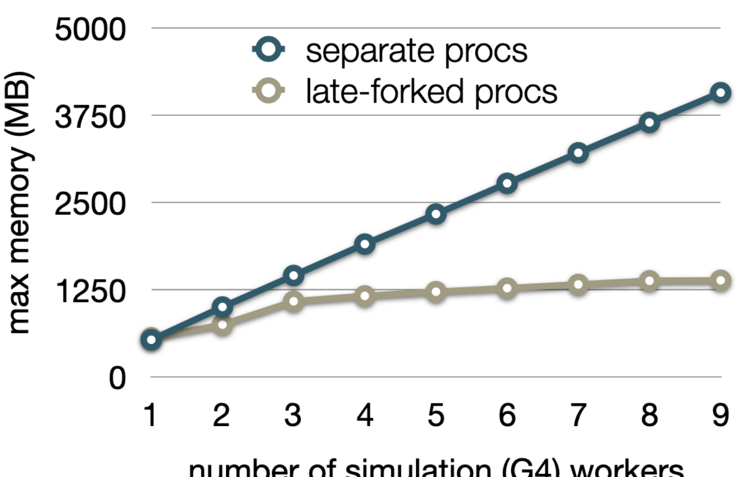




full event

sub-events
(one at a time)





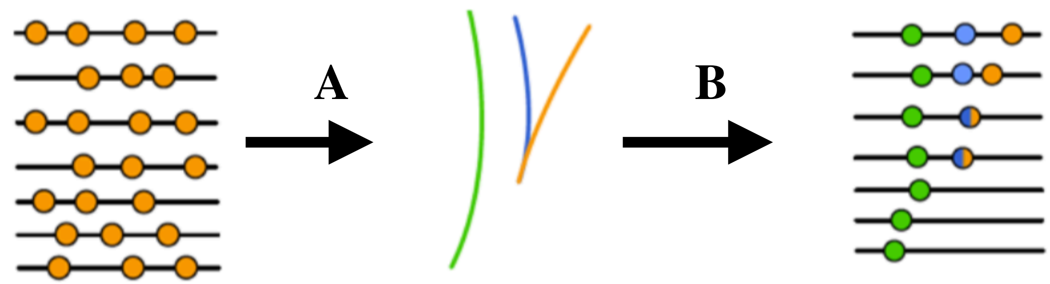
[Wenzel]

LHCb: [Gauss-on-gaussino] multi-threaded particle tranport

Fast MC basics

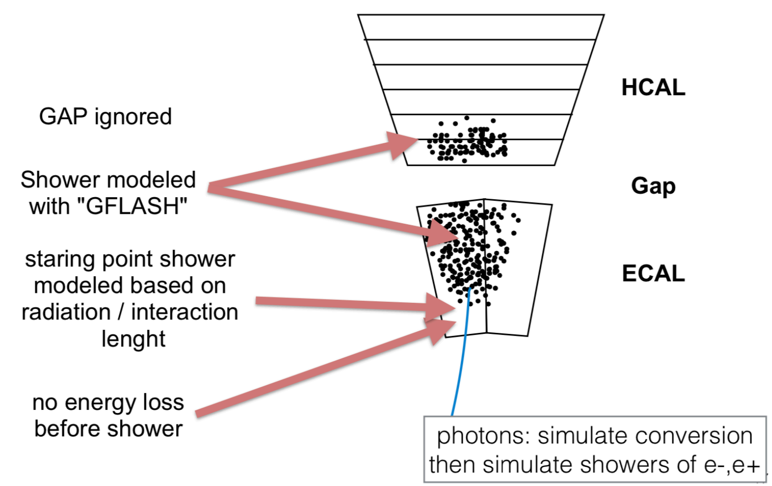
Fast tracking

- Truth-assisted
  - Track seeding
  - Hit association, combinatoric reduction
  - Track fitting



Fast calorimetry

- GFlash *(Grindhammer, et. al., 1990)*
  - Classic parametrisation-based showering model for lateral and longitudinal profiles
  - Shower libraries



Tricks

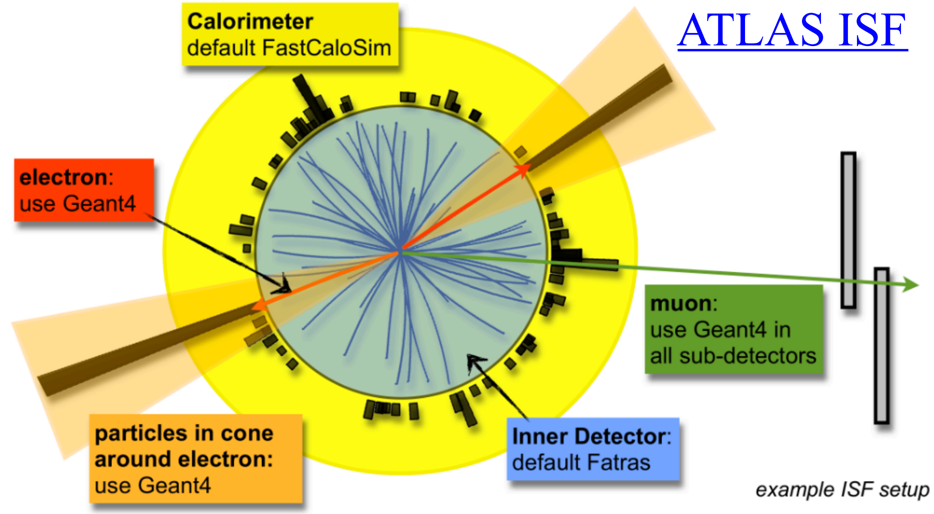
- Pile-up premixing: mix pileup events before introducing signal process
- ReDecay: Separate signal processes from background, re-decay signal
- etc

Parametric Fast MC

ATLAS Fast MC

- ATLFASTII (10-0x speed up)
  - FastCaloSim - parametric showering
  - FATRAS - fast tracking based on thin surfaces
- Option to run different simulation for different parts of an event
- Aim to simulate 75% of events by Run 3

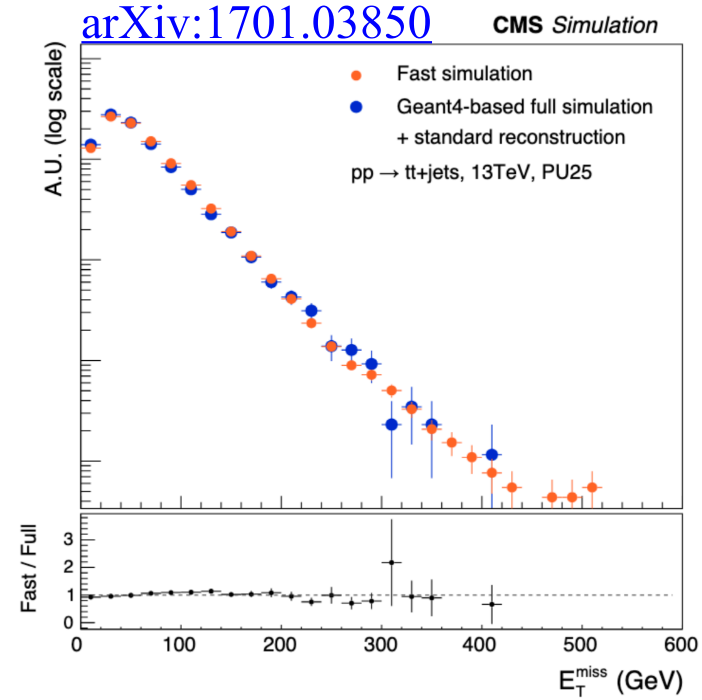
ATLAS ISF



CMS Fast MC

- Simplified geometry, thin surfaces
- Fast calorimetry and tracking: 10x speed-up
- Realistic simulation of low-level objects
  - hits and clusters, tuned to match FullSim
  - apply standard high-level modules
- 15% of Run 2 events, aim to reach 50%

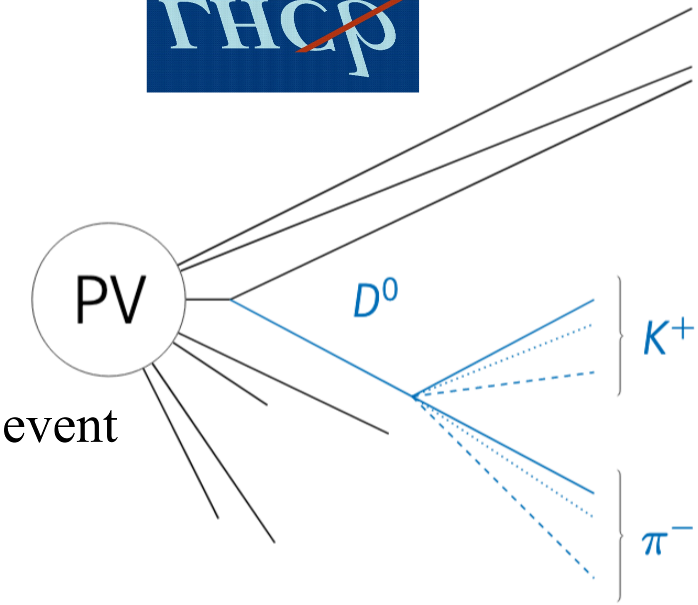
arXiv:1701.03850

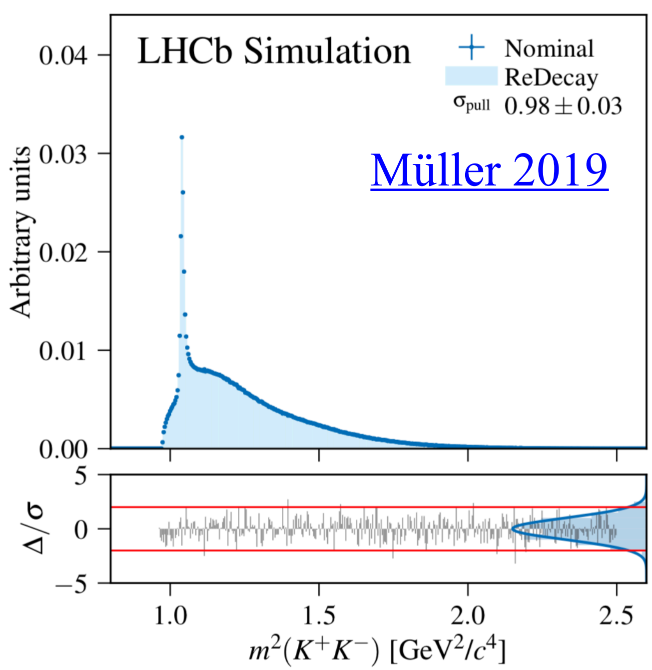


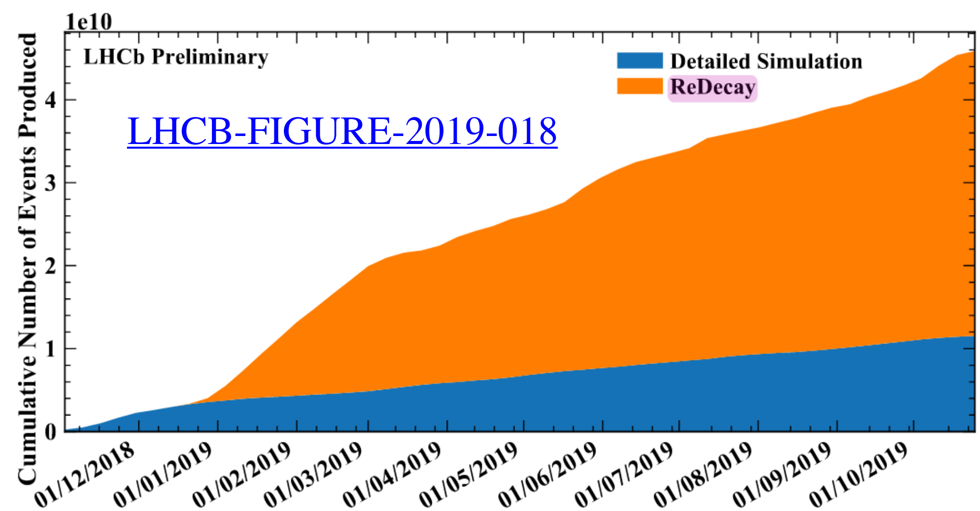
ReDecay

LHCb

- generate full event
- remove signal
- simulate remaining event
- generate/simulate multiple signal decays
- combine simulated signal with stored underlying event







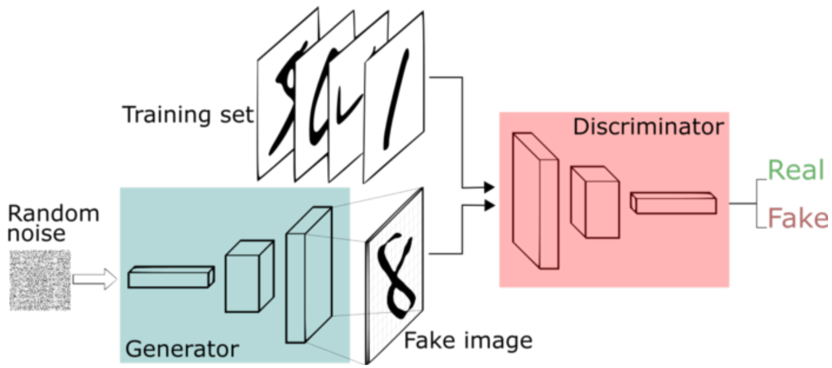
Machine learning for MC

Train neural networks (NNs) to

- Generate physics images of
  - calorimeter showers arXiv:1701.05927 arXiv:2005.05334 arXiv:1807.01954
  - final-state observables arXiv:2010.01835
  - detector hits, etc
- Refine (fast) simulation by
  - modifying collections produced by existing fast simulation
    - such as Delphes, parametric FastSim, GAN, or pure generator-based collections arXiv:1802.03325
  - re-weighting events output by existing simulation arXiv:2009.03796
  - re-weighting **latent space** of a GAN-based fast simulation to produce unweighted, refined fast simulation arXiv:2106.00792
- Accelerate FullSim, Geant-4 based applications by
  - interpolating across coarsely sampled time slices, denoising

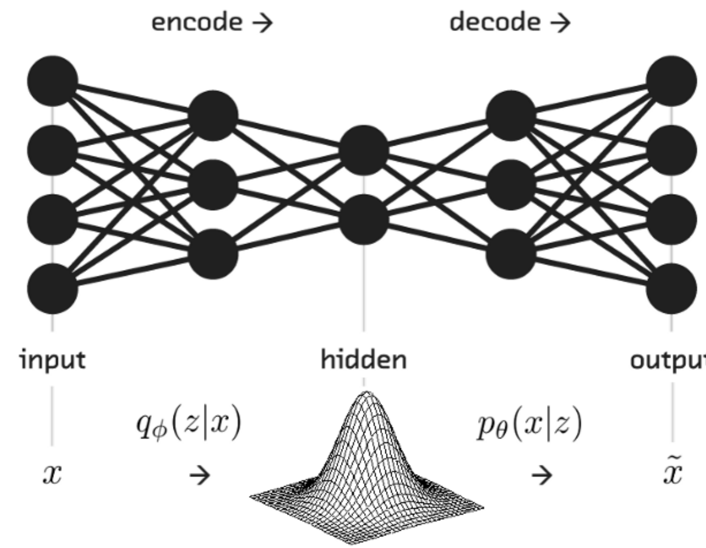
ML Techniques

Generative Adversarial Network (GAN)



- Train NN to generate new images starting from random noise input
- Discriminate between synthetic images and real data
- Wasserstein GAN: we don't care if „this jet" looks ok, we only care if P[x(jet)] looks ok
  - „Earth mover" ensemble-based loss function
  - Greater stability, convergence

Variational Autoencoder (VAE)

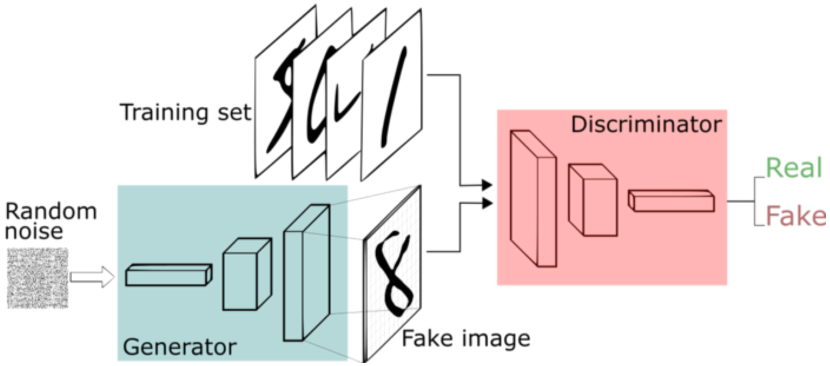


- Learn compressed representation (latent space) of inputs, randomly sample to synthesise new examples
- Variational: learn *probability distribution* of latent space
  - Better for generative output
  - Risks when latent space dimensionality is too small

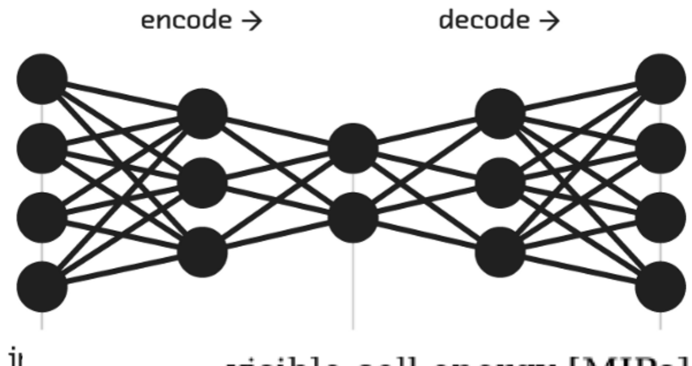
Also CNN, Normalizing flows, others

ML Techniques

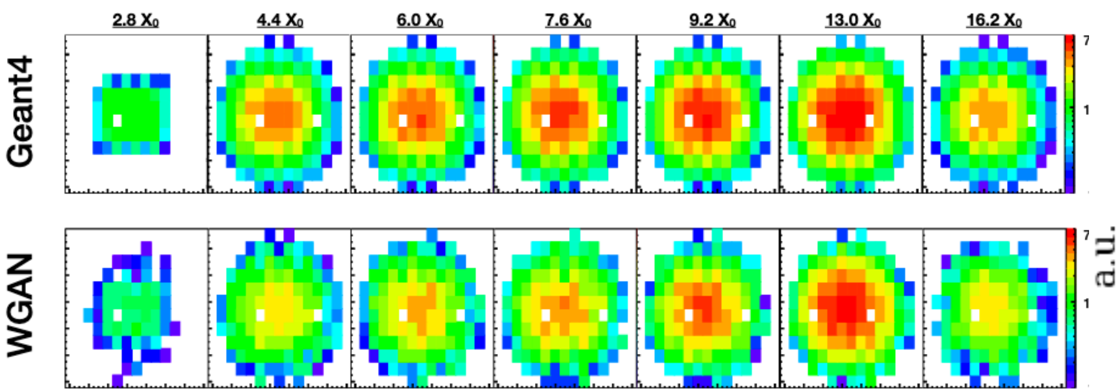
Generative Adversarial Network (GAN)



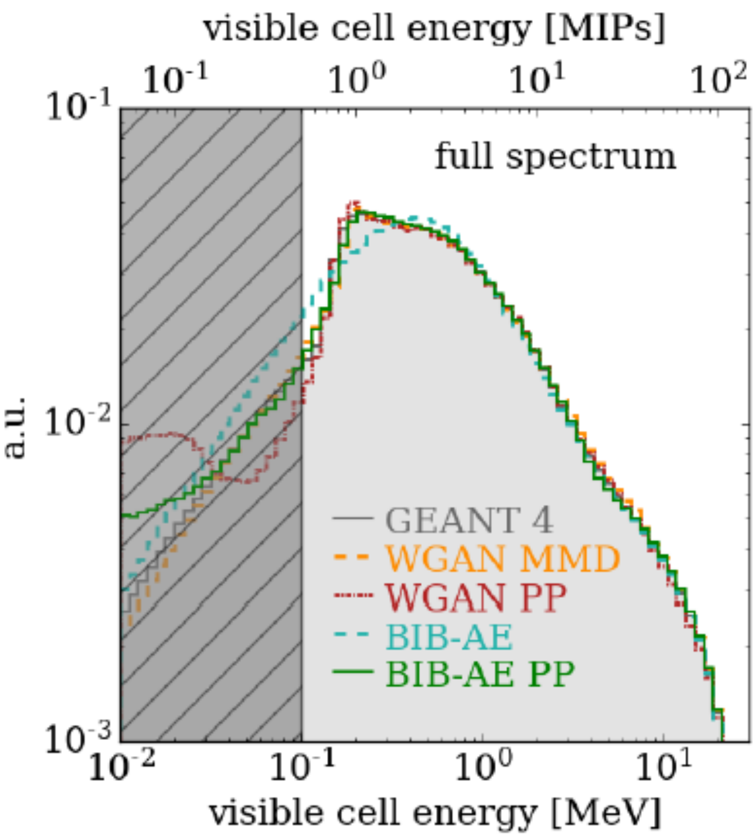
Variational Autoencoder (VAE)



GAN for HGCal calorimeter showers



arXiv:1807.01954

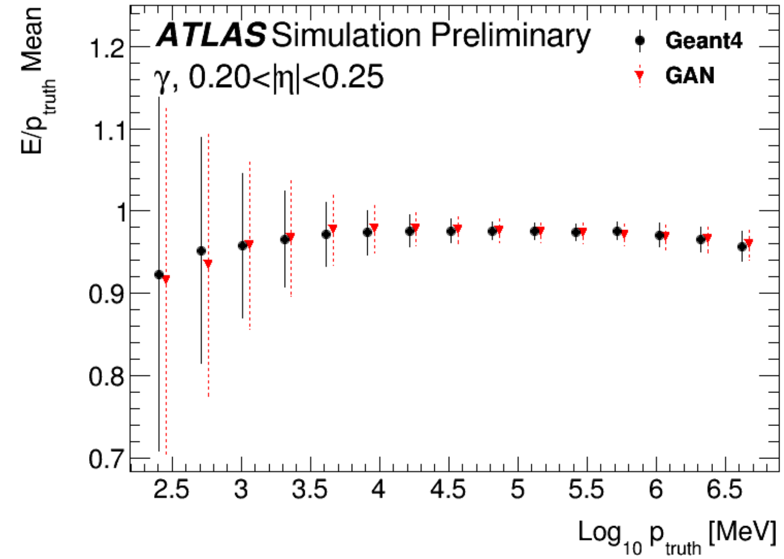


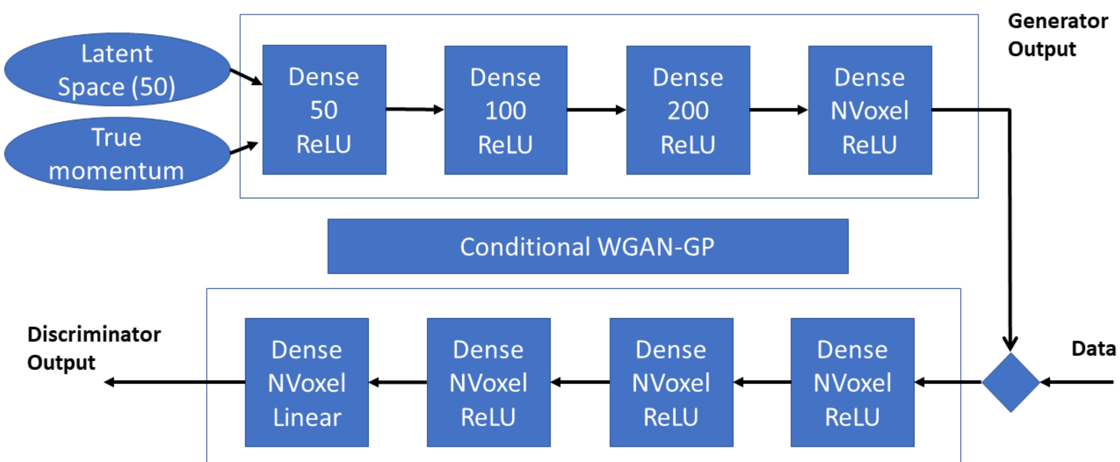
arXiv:2005.05334

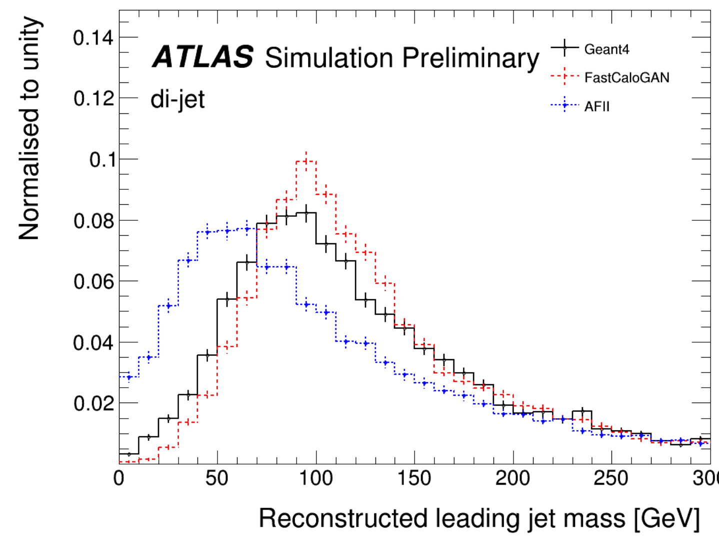
Also CNN, Normalizing flows, others

ATLAS: FastCaloGAN

- Strong majority of SIM CPU spent on calorimeter showering
- Trained 1 WGAN per 100 η slices for each particle type:
  - electron, photon, pion samples (300 GANs
- Improvements to modeling over FastCaloSim
- In Run 3 use for 75% of produced MC events



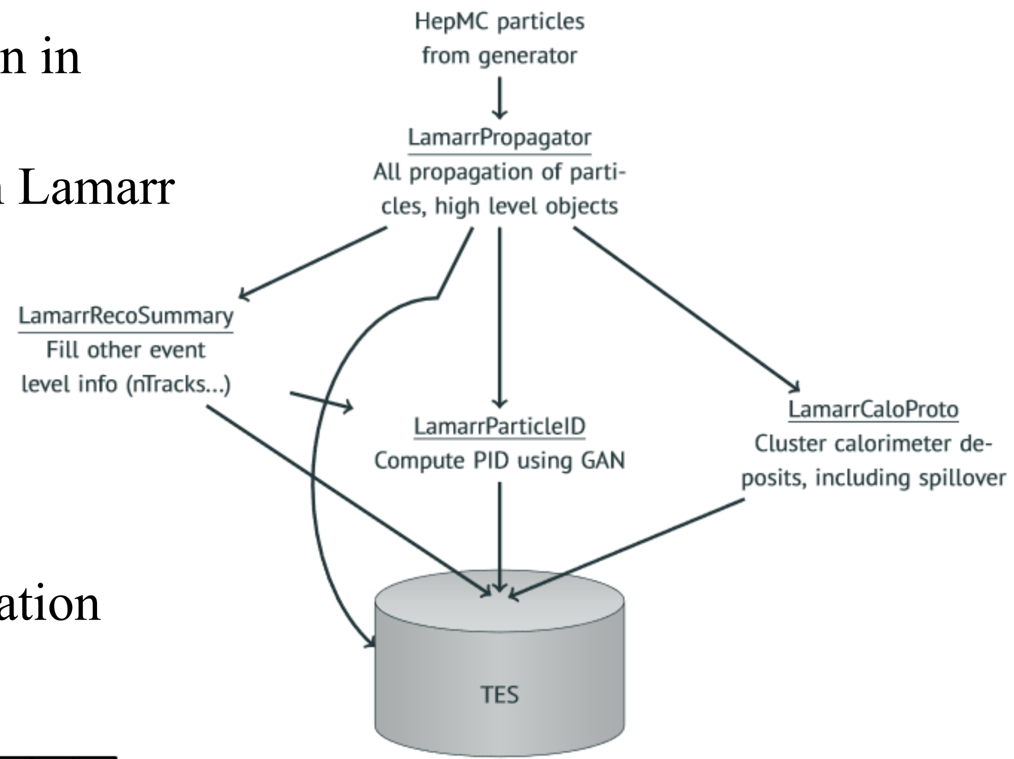


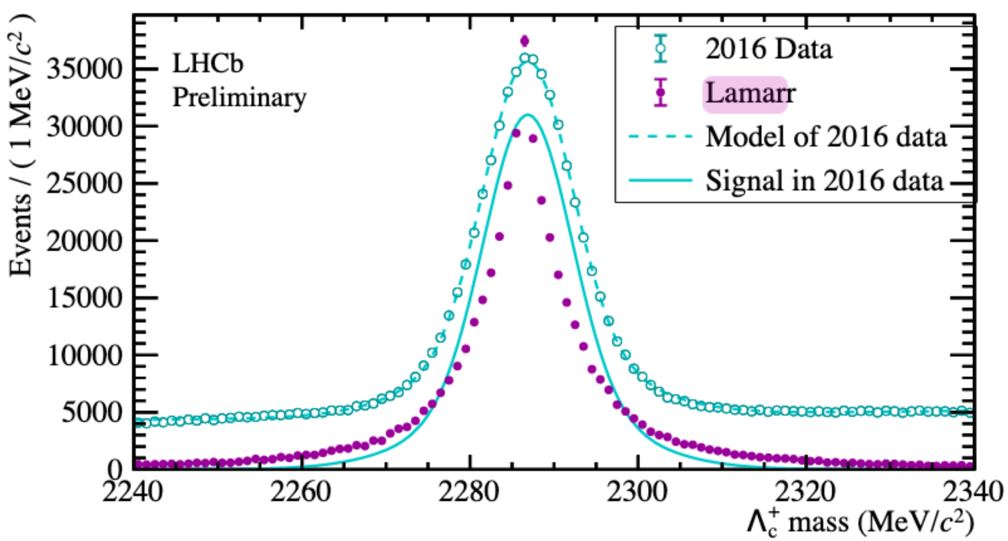


(source)

LHCb: End-to-end with Lamarr

- Imminent need for drastic reduction in simulation time
- Developing custom fast simulation Lamarr
- 4 GANs + 1 NN classifier
  - RICH detector
  - Muon reconstruction
  - Particle ID
  - (future) Calorimeter response
- Input: generator-level truth information
- Output: final-state observables

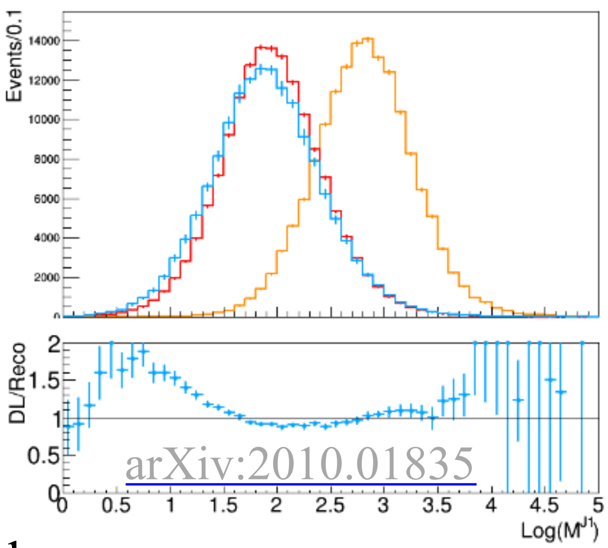




CHEP2019 (1)
CHEP2019 (2)
ICHEP2020

Conclusion

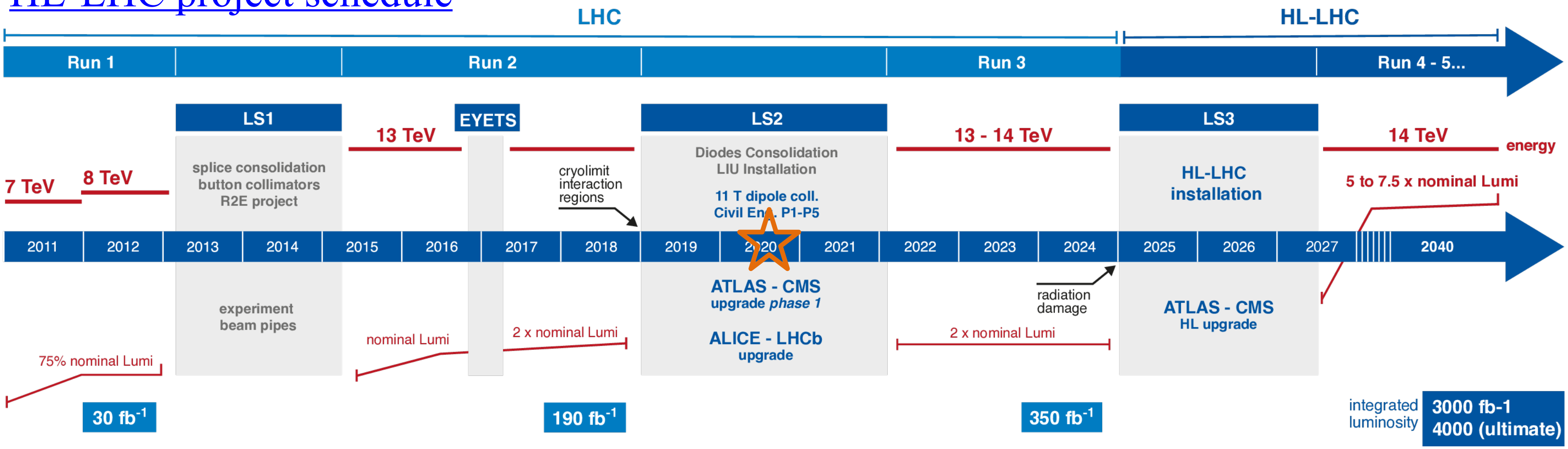
- Event simulation is a pillar of LHC analysis
- Run 4 (Run 3) will bring qualitative increases in data rates for ATLAS and CMS (LHCb and ALICE)
- All experiments must make paradigm-shifting changes to the MC production chain to stay within the pledged computing budgets
- High occupancy and new detector technology increase the need for accurate simulation which is not too slow
- Speed-ups coming with increased parallelisation
  - Geant4, VecGeom, heterogeneous resources
  - Experiment-specific tricks
- Multi-Faceted ML R&D being pursued aggressively
  - Within experiments and independent of
  - Adiabatic, qualitative changes to existing frameworks
  - idea meritocracy with bright future!



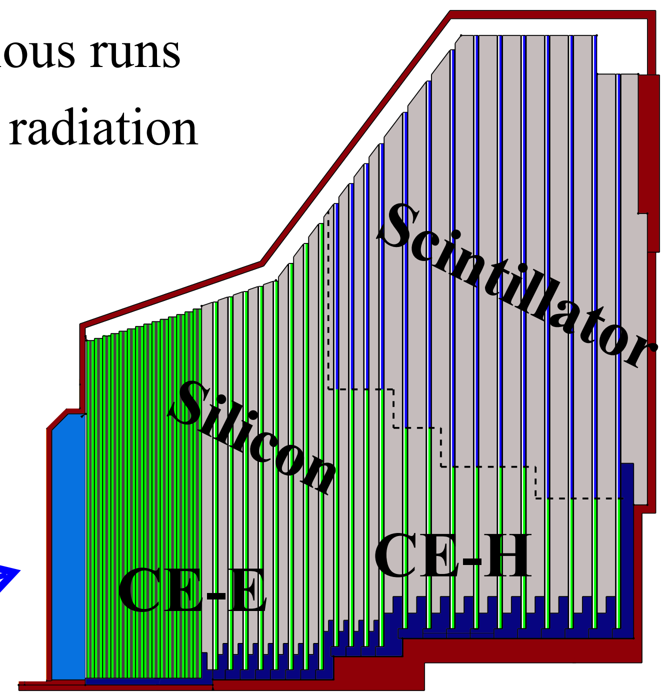
chat: https://cern.zoom.us/j/4478496715

Upgrades

[HL-LHC project schedule]

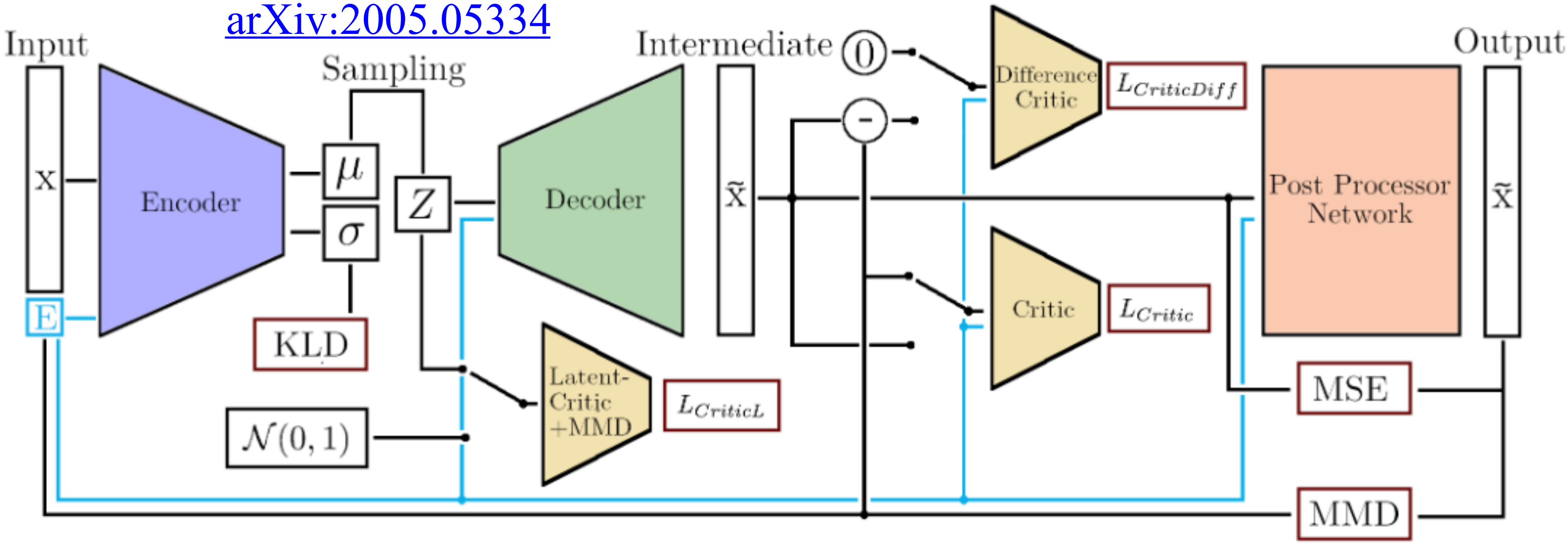


- Run 4+ expected to deliver ~10× data from previous runs
  - Higher luminosity: higher occupancies, higher radiation → need new detectors!
- CMS detector upgrades include:
  - Pixel (inner tracker): 66M → 1947M channels
  - Outer tracker: 9.6M → 215M channels
  - High Granularity Calorimeter (HGCal): 85K → 6M channels



BIB-AE Architecture

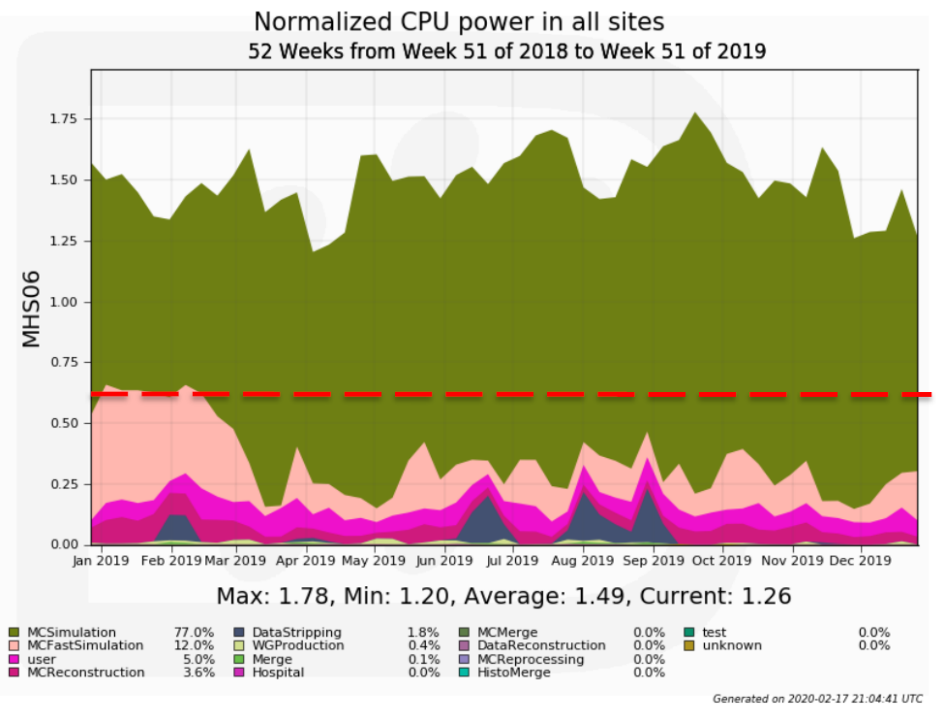
arXiv:2005.05334

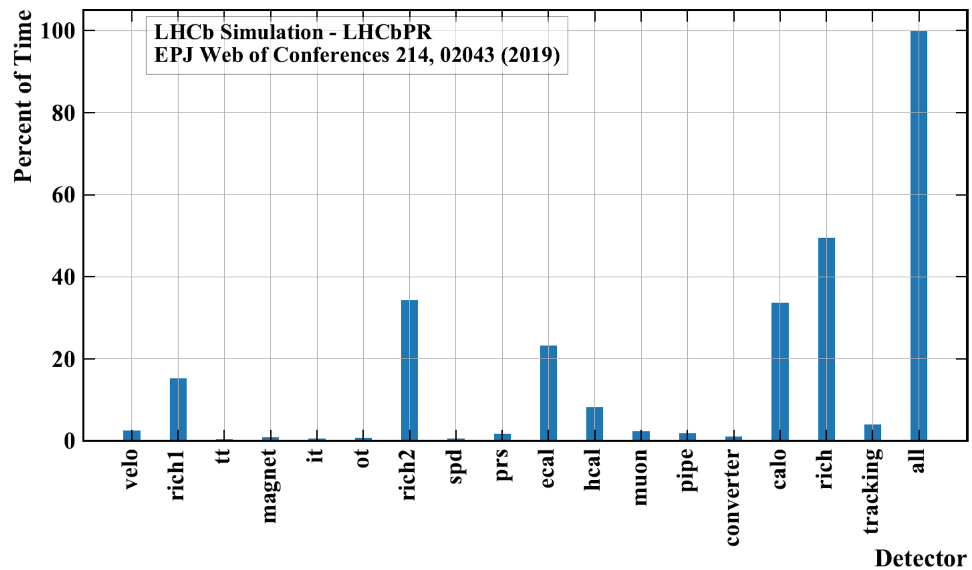


$$
\begin{aligned}
L_{\text{BIB-AE}} = & -\beta_{C_L} \cdot \mathbb{E}[C_L(E(x))] \\
& -\beta_{C} \cdot \mathbb{E}[C(D(E(x)))] \\
& -\beta_{C_D} \cdot \mathbb{E}[C_D(D(E(x)) - x)] \\
& +\beta_{\text{KLD}} \cdot \text{KLD}(E(x)) \\
& +\beta_{\text{MMD}} \cdot \text{MMD}(E(x), \mathcal{N}(0,1))).
\end{aligned}
$$

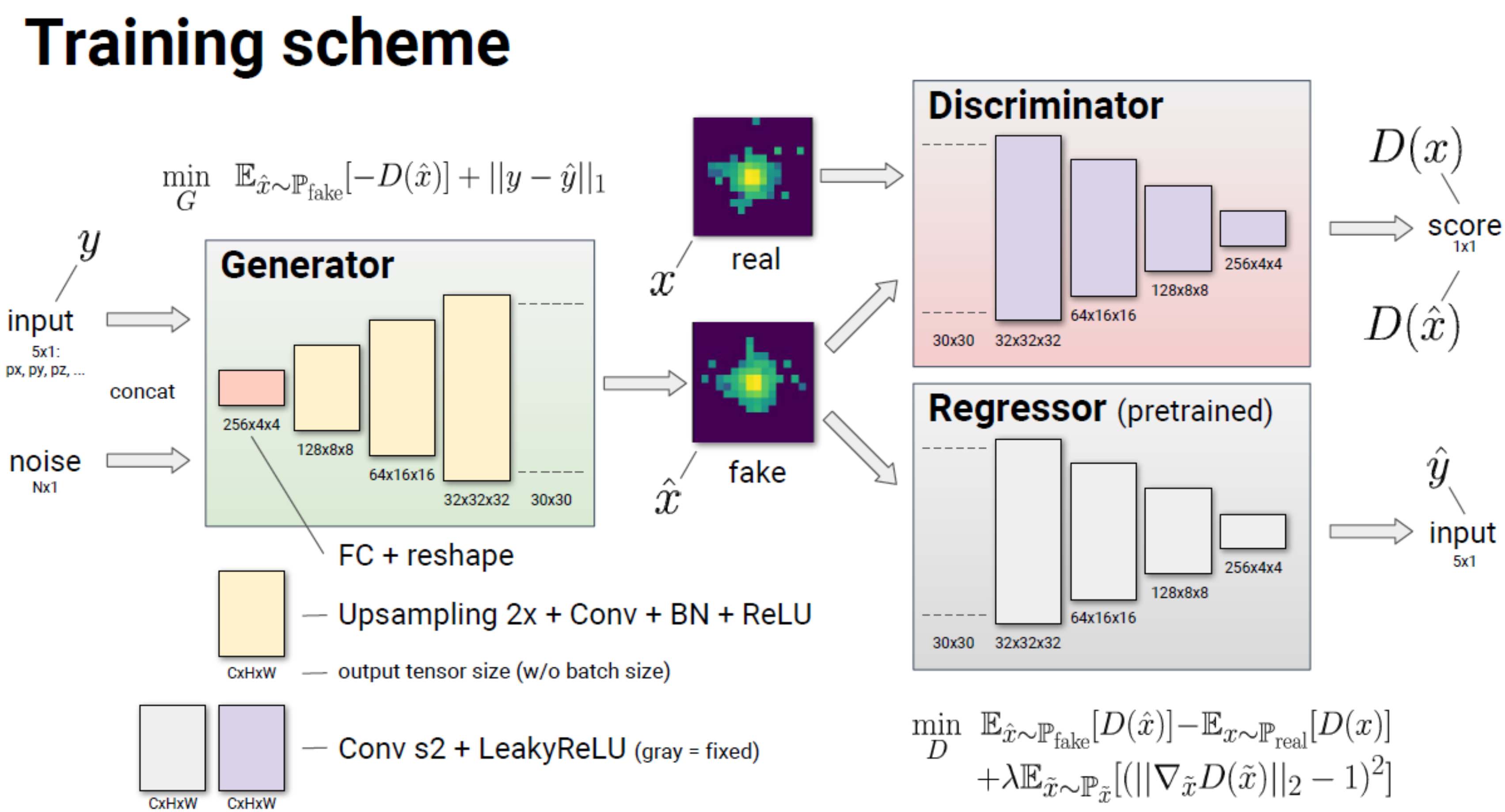
LHCb FullSim CPU Usage

- From [M. Rama, ICHEP2020]
- Also *Eur. Phys. J. Web Conf.* 214 (2019) 02043





LHCb Calorimeter GAN



[Gauss-on-gaussino - multi-threaded prototype of simulation]