Fast detector simulation and anomaly detection using Graph variational autoencoders

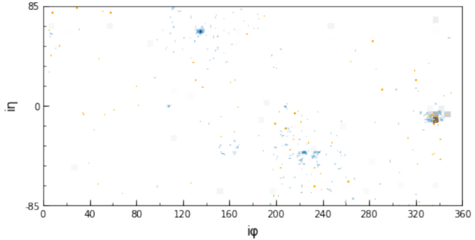
December 7, 2023

Collisions at the LHC

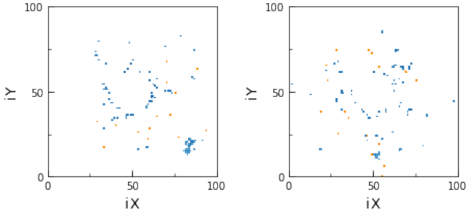
- In the CMS and ATLAS detectors at LHC, opposing proton beams are smashed[3]
- By products are measured using various sub-detectors
- Trackers track the trajectories of photons and charged particles
- ECAL measures the transverse momenta of electrons/photons and cylindrical coordinates
- HCAL measures the transverse momenta of hadrons and cylindrical coordinates

The generated image

- From hits in a subdetector, we get "images" wrt ϕ and η[2]

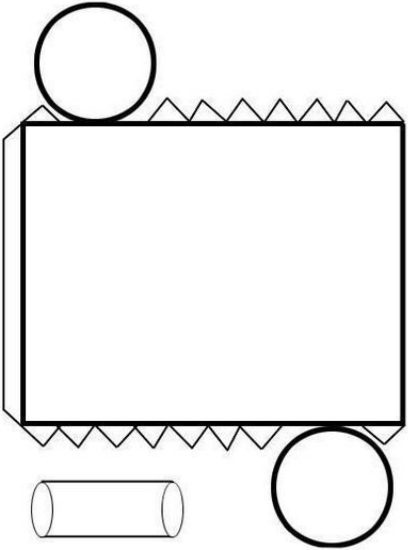


(a) Barrel section of composite image in ECAL-centric geometry. Image resolution: 170 × 360.



(b) Endcap sections of composite image in ECAL-centric geometry. Image resolution: 100 × 100.

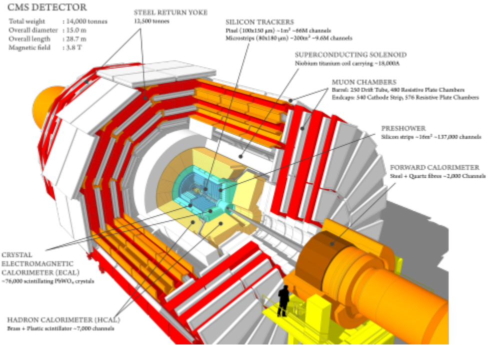
Imagine making a (finite) cylinder out of paper



[1]

[1] https://www.pinterest.co.uk/pin/772437773572199448/

Cross section of the CMS detector



Simulating the collisions

- Hard(core) scattering
- Secondary collisions
- Decays and radiative correction
- Hundreds of different processes, everything Monte Carlo!

Hard scattering

- The moments right after the primary collision...
- Modeled using perturbative QCD- cross sections are given by Feynman diagrams
- The initial/final states are not eigenstates of unperturbed theory, unlike QED
- Higher order loops with IR divergences require very precise Monte Carlo or analytical estimates

Secondary collisions/processes

- Non-relativistic QCD and semi-classical appromations for decay processes
- Highly empirical, sometimes benchmarked by detector calibrations
- Factorization theorems and Parton Distribution functions
- General purpose libraries- PYTHIA, SHERPA[7]

Detector effects and ML models

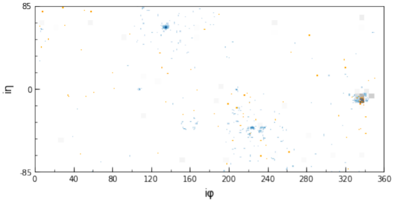
- GEANT4 has extremely precise models of detector interaction with final product states[1]
- The entire simulation pipeline can take days to run on supercomputers, for sufficient hits/samples
- Non ML alternatives to GEANT4 (like DELPHES) tradeoff precision for speed[4]
- ML models fit precise data to well-behaved functions- similar to variational inference of energy eigenstates in quantum systems

Generative ML for end to end simulation

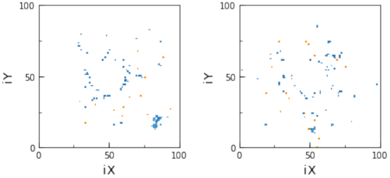
- Particle hit images are inherently probabilistic- so ML models approximate probability distributions on the data
- Dominant Generative ML paradigms- VAE, GAN, NF
- Popular implementations use convolution layers to generate pixel by pixel
- Downside- detector hits are highly sparse which limit applicability of convolutions


Ah [redacted], here we go again.

Sample detector image



(a) Barrel section of composite image in ECAL-centric geometry. Image resolution: 170 × 360.



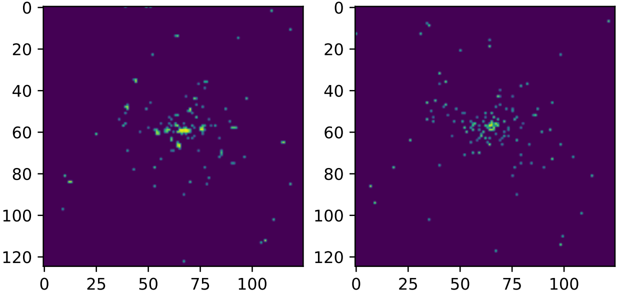
(b) Endcap sections of composite image in ECAL-centric geometry. Image resolution: 100 × 100.

Graph modeling and message passing convolutions

- Number of non-zero pixels << Total number of pixels [6]
- We model non-zero hits alone, and their positions
- Assume each hit depends on nearest neighbor hits
- Graph message passing implements precisely this[5]
- Output invariant of order of nodes or their neighbors- similar to translation invariance in CNNS

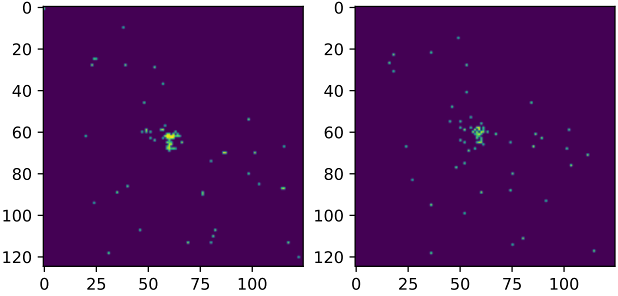
Some initial results

Quark jet



An anomaly

Gluon jet



while questions:
attempt_to_answer()
print("Thank you and peace out!")

Agostinelli, Sea et al. (2003). "GEANT4, a simulation toolkit". In: *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3, pp. 250–303.

Andrews, M et al. (2018). "End-to-end event classification of high-energy physics data". In: *Journal of Physics: Conference Series*. Vol. 1085. 4. IOP Publishing, p. 042022.

Collaboration, CMS et al. (2008). "The CMS experiment at the CERN LHC". In: *Jinst* 3, S08004.

De Favereau, J et al. (2014). "DELPHES 3: a modular framework for fast simulation of a generic collider experiment". In: *Journal of High Energy Physics* 2014.2, pp. 1–26.

Gilmer, Justin et al. (2017). "Neural message passing for quantum chemistry". In: *International conference on machine learning*. PMLR, pp. 1263–1272.

Hariri, Ali, Darya Dyachkova, and Sergei Gleyzer (2021). "Graph generative models for fast detector simulations in high energy physics". In: *arXiv preprint arXiv:2104.01725*.

Sjöstrand, Torbjörn et al. (2001). "High-energy-physics event generation with PYTHIA 6.1". In: *Computer Physics Communications* 135.2, pp. 238–259.