ML4Sim @ CMS

Kevin Pedro (FNAL)

April 22, 2021

CMS Simulation

- CMS FullSim is *4–6× faster* than baseline Geant4
  - Numerous technical optimizations & physics-preserving approximations
  - Sustained effort to commission and adopt new Geant4 versions
- CMS FastSim application: *10–20× faster* than FullSim
  - Includes sim- and reco-level optimizations (tracking)
  - Currently used for generation of large supersymmetric model scans, some studies of systematic uncertainties

➢ Well-positioned for Run 3, but further acceleration crucial for Phase 2

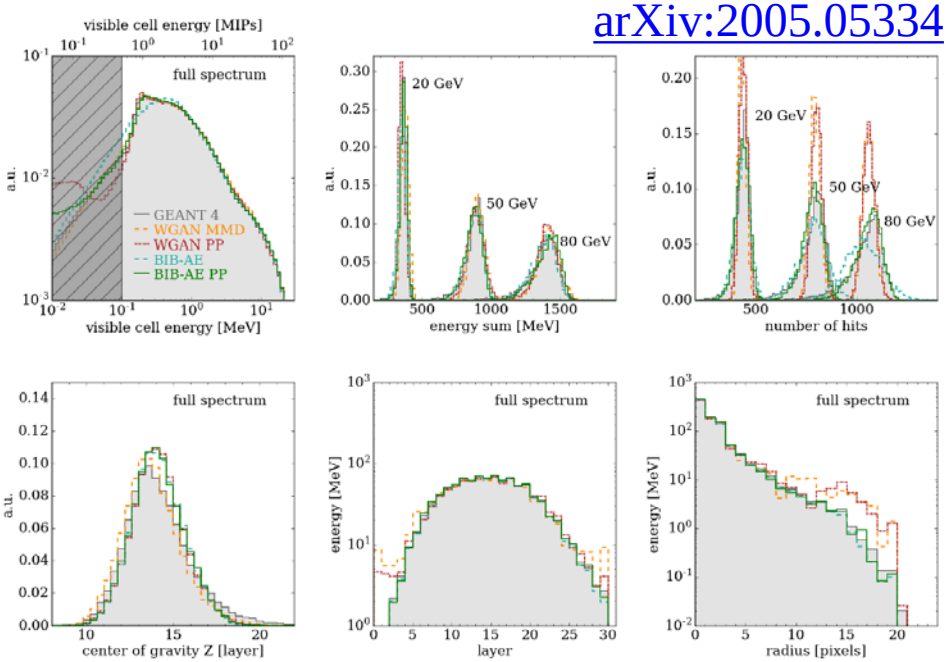
  - May deploy some ML prototypes for specific use cases during Run 3
  - Provides natural avenue to utilize *heterogeneous computing resources* (GPUs, FPGAs, HPCs, etc.)
  - Need to balance *tradeoffs*: ML research vs. implementing, testing, validating in experiment software framework and production

- Next slides highlight several R&D projects from CMS members
  - More R&D projects ongoing internally

Ongoing Efforts (1)

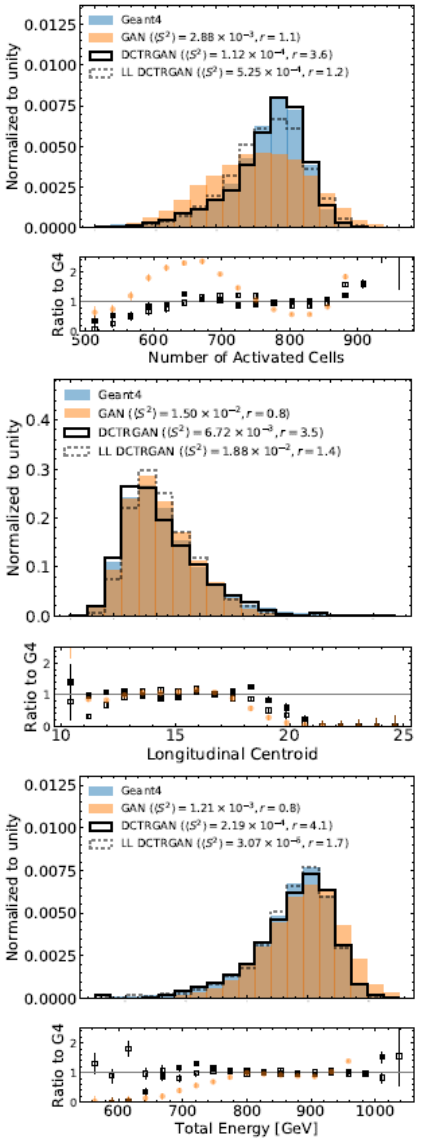
- Bounded Information Bottleneck:
  - Generalization/combination of VAE and GAN
  - Aimed at ILC imaging calorimeters
    - Similar to CMS HGCal
  - Improves on standard GANs

arXiv:2005.05334



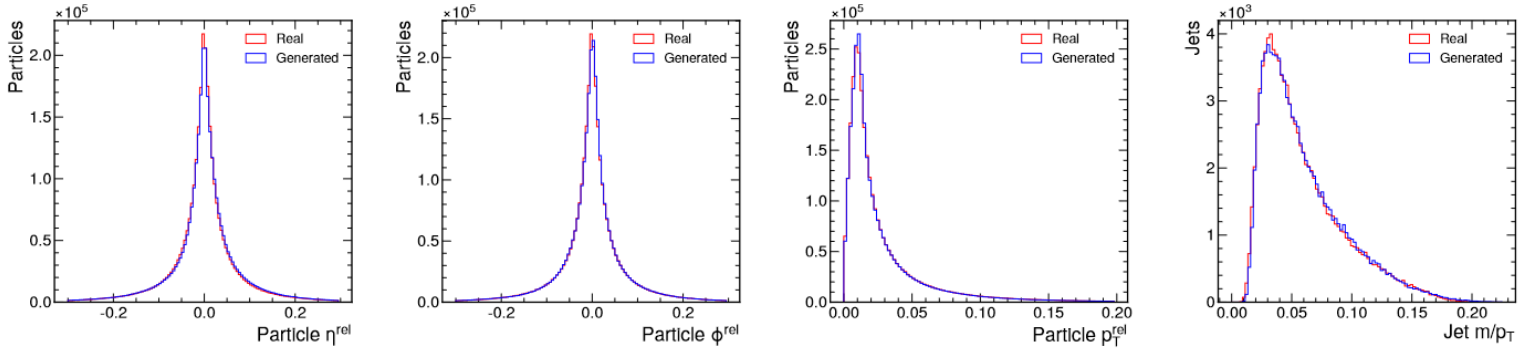
- Possible to improve GAN results with an additional classifier: "DCTRGAN"
  - Trained to reweight events after GAN training finishes

arXiv:2009.03796



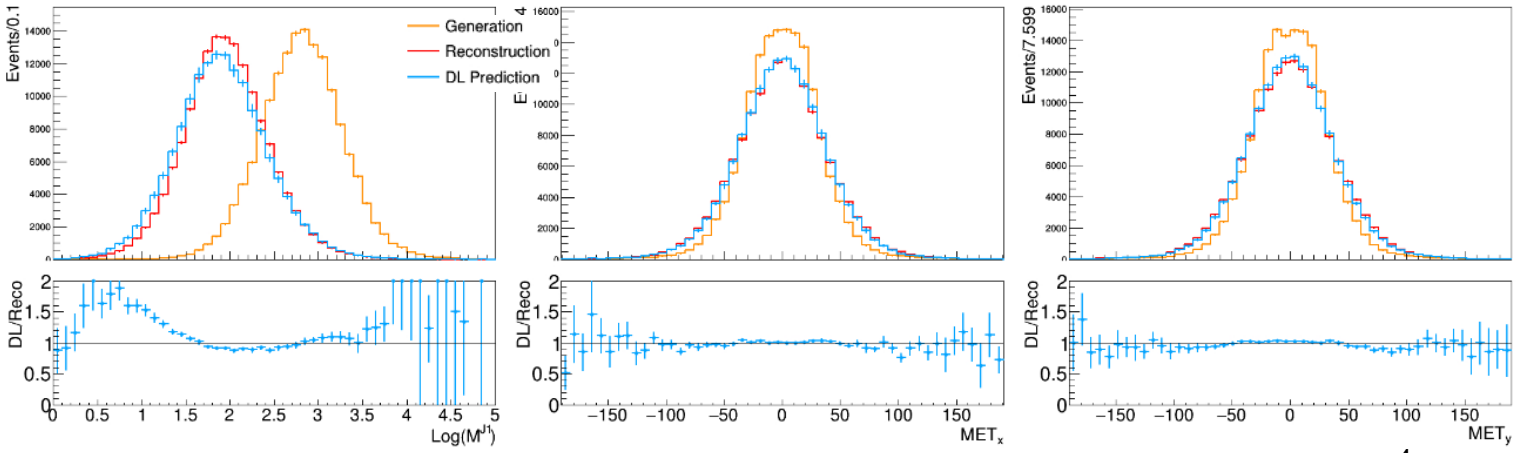
Ongoing Efforts (2)

arXiv:2012.00173



- Sparse data generation (above):
  - Graph-based (MPNN) GAN to handle irregular geometry
  - Also exploring VAE-based architecture and techniques to generate variable-size graphs
- Analysis-specific Fast Simulation (below):
  - Regression-based approach:
    - Learn detector response
    - Apply to generated events
    - Target specific high-level variables

arXiv:2010.01835



Backup