ICHEP2018 SEOUL
XXXIX INTERNATIONAL CONFERENCE ON high energy PHYSICS
JULY 4 - 11, 2018 COEX, SEOUL

Calorimeter Fast Simulation Using Different Approaches

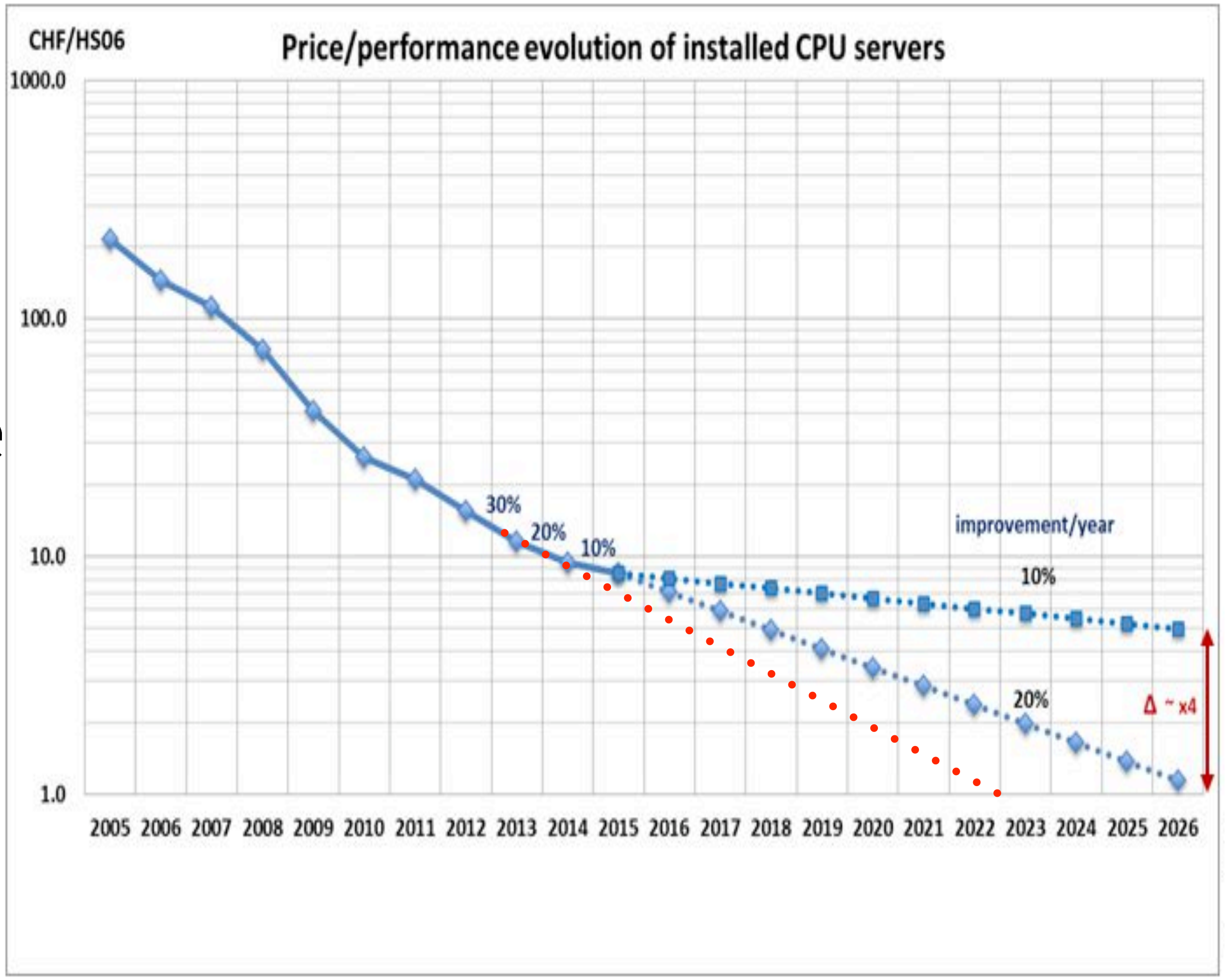
Fedor Ratnikov on behalf of the LHCb
with assistance from:
Egor Zakharov and Elena Orlova

SCHOOL OF DATA ANALYSIS

NRU Higher School of Economics,
Yandex School of Data Analysis

Context

- About 80% of computing resources are used for MC simulation in HEP experiments
- Calorimeter simulation is one of bottlenecks
  - CPU usage for LHCb simulation for Run1 and Run2 conditions
    - calorimeters - 38%
    - RICH - 47%
- Can not expect exponential rise CPU performance
- Need work around for Run3 and HL-LHC



Works Around the Problem

- Optimisation
- Vectorisation
- GPU
- Bypass detailed shower simulation

Bypass Detailed Shower Development

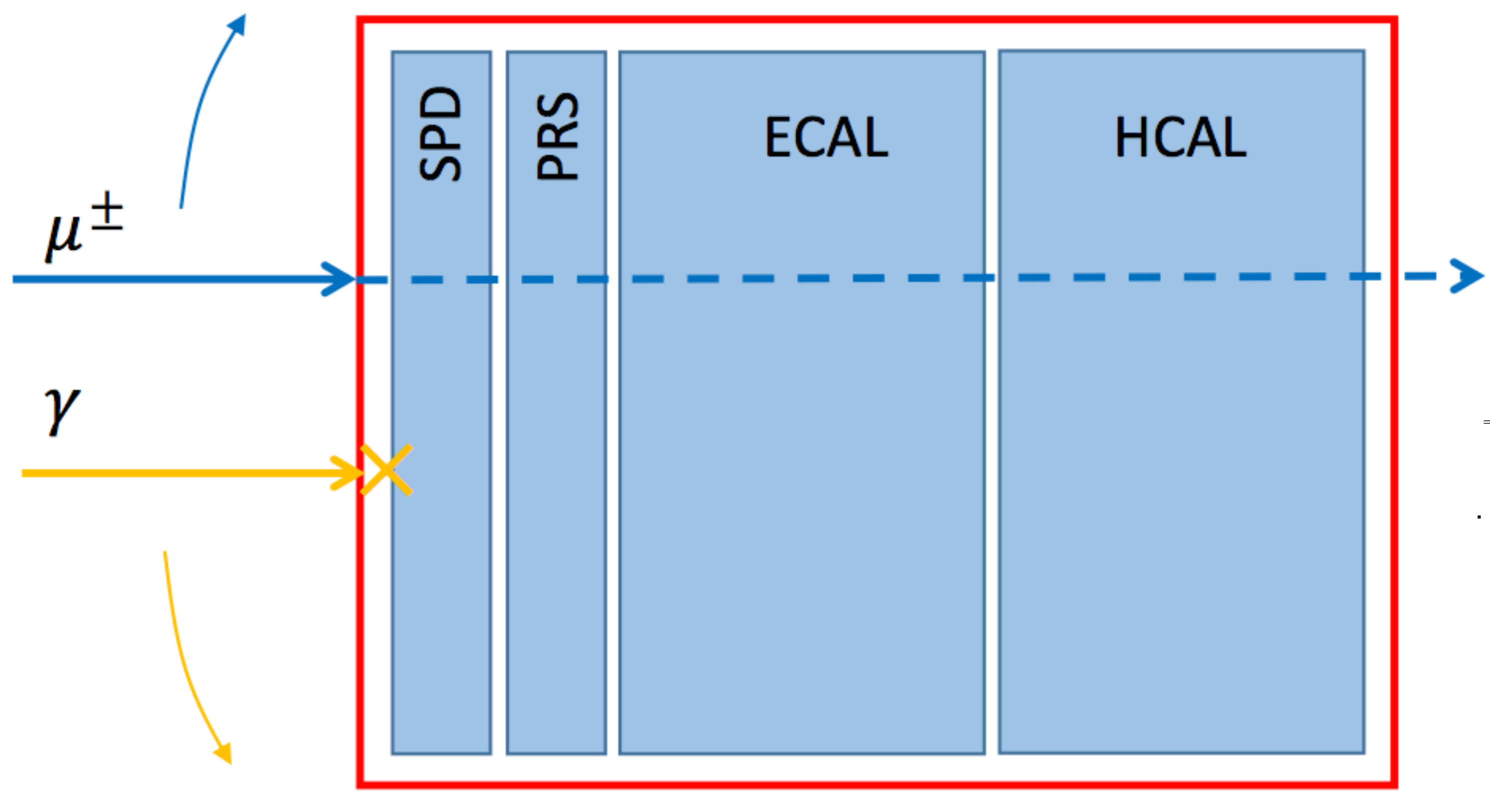
- Why this could work?
  - Shower in calorimeters typically consists of huge amount of microscopic interactions and particles, which behaviours are described by GEANT in details
  - However readout system has pretty coarse granularity
  - The calorimeter response to the individual impact particle has much less degrees of freedom comparing to the full shower produced by the particle
    - law of large numbers is applicable

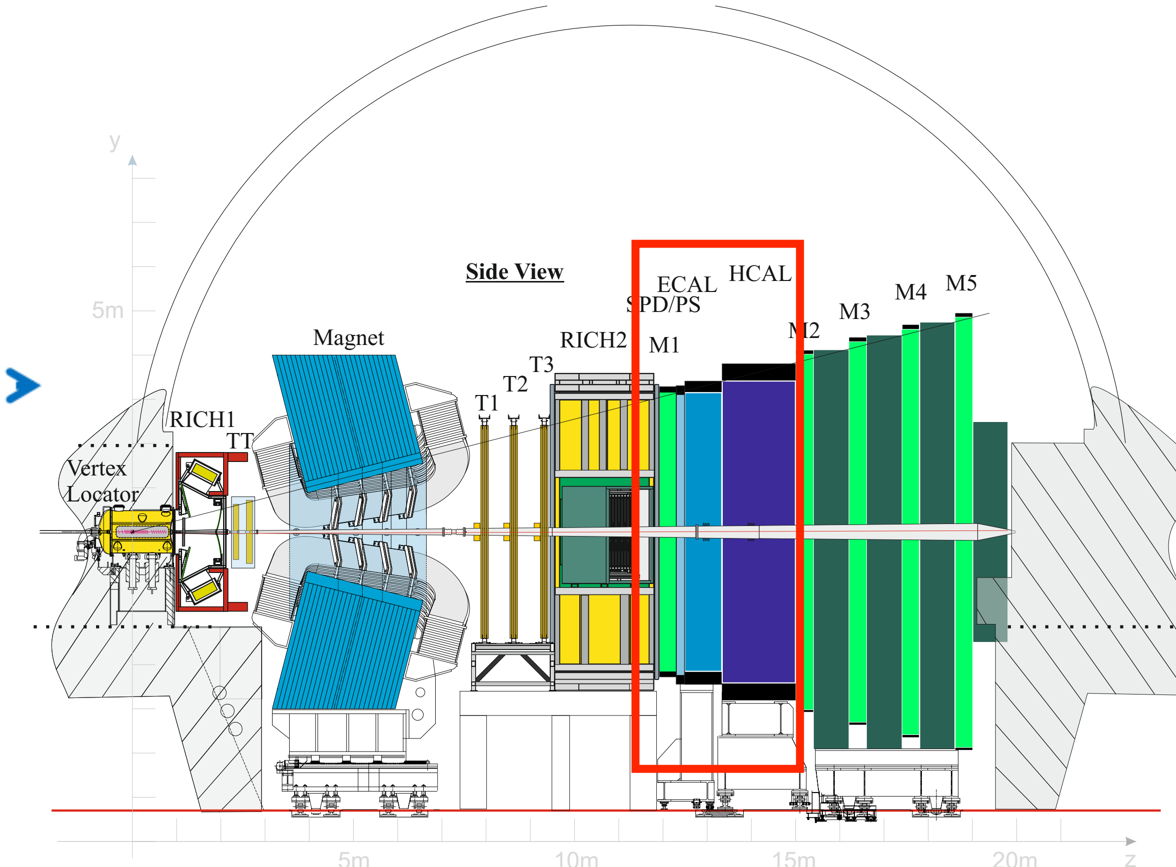
Approaches

- Shower library
  - build a library of calorimeter responses to impact particle in corresponding 5D phase space using GEANT4
    - 3D momentum + 2D coordinate for every particle type
  - use library to construct calorimeter response to the particle with given parameters
- Generative models
  - train a model to reproduce calorimeter response to impact particle in corresponding 5D phase space using GEANT4
  - use model to generate calorimeter response to the particle with given parameters
  - Two modern approaches are popular in Machine Learning community
    - Generative Adversarial Networks (GAN)
    - Variational Autoencoders (VAE)
- We explore all these approaches for LHCb needs

Generic Interface

- The G4track is propagated by Geant4
- The CaloSubHits are created from G4Hits





- The G4track is killed at the calo entrance
- The CaloSubHits are created by a fast calo hit generator

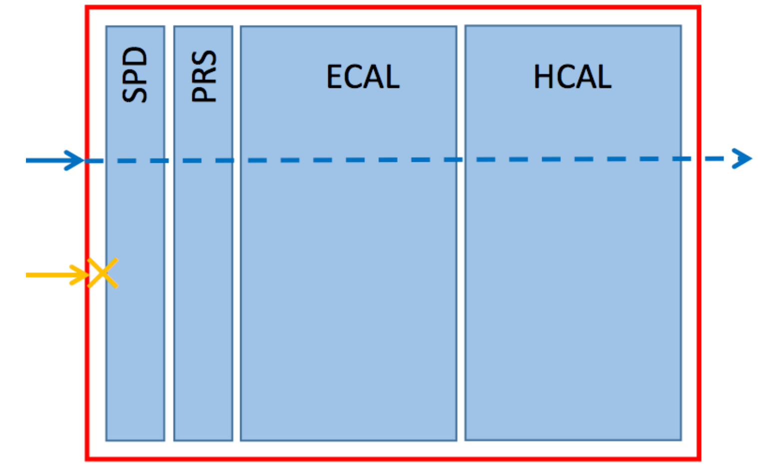
◇ For each particle, possible to choose fast or full mode based on

  ◇ particle type

  ◇ kinematic properties

Shower Library

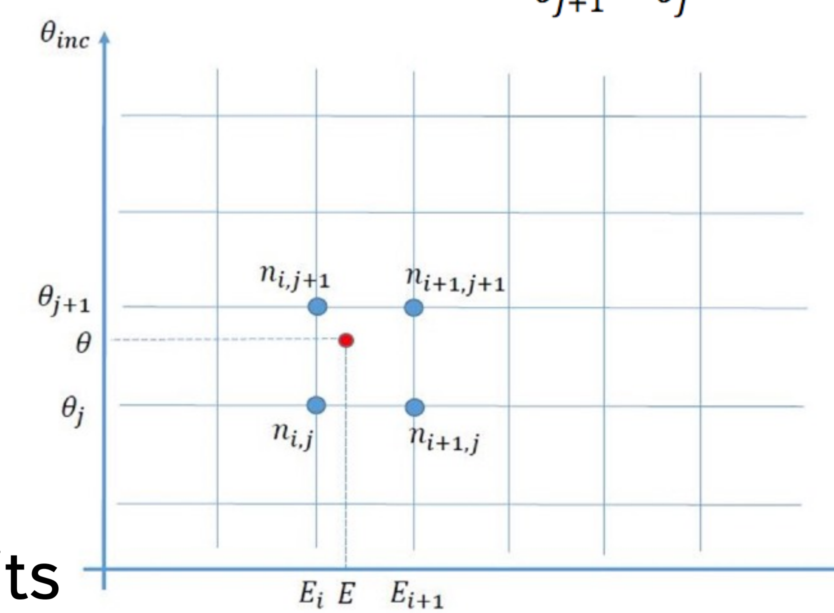
- **Concept**
  - build a library of GEANT4 full-simulated particles on regular grid in (E, θ, φ) phase space on the SPD entrance
    - store hits from all 4 detectors
  - for fast simulation replace GEANT4 hit simulation with hits collection from the library



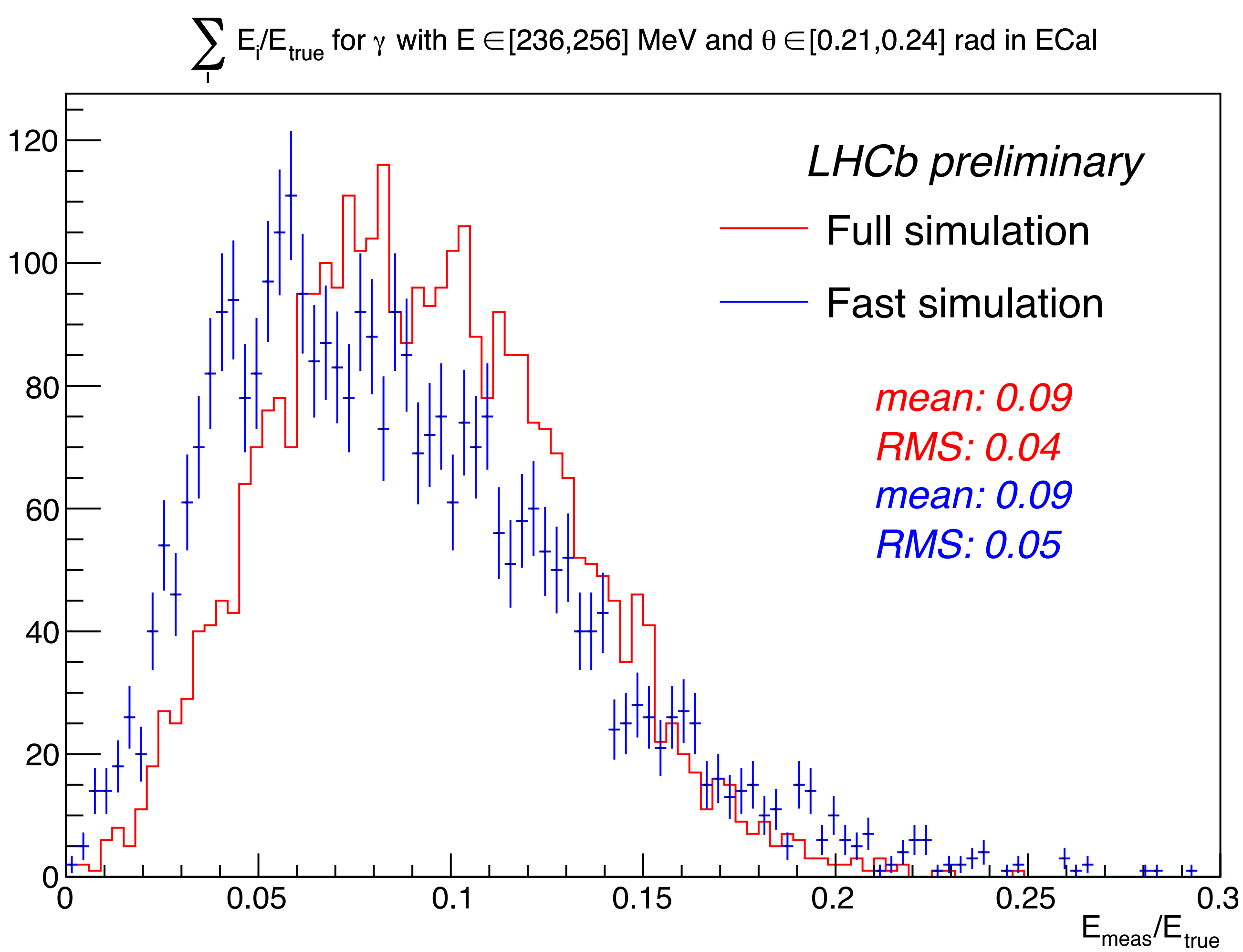
- **Implementation**
  - pick randomly one collection of hits in nodes around the actual parameter point
  - from each collection pick randomly a fraction of hits
    - fraction is weighted according to point proximity
  - interpolated collection is defined as a sum of these hits

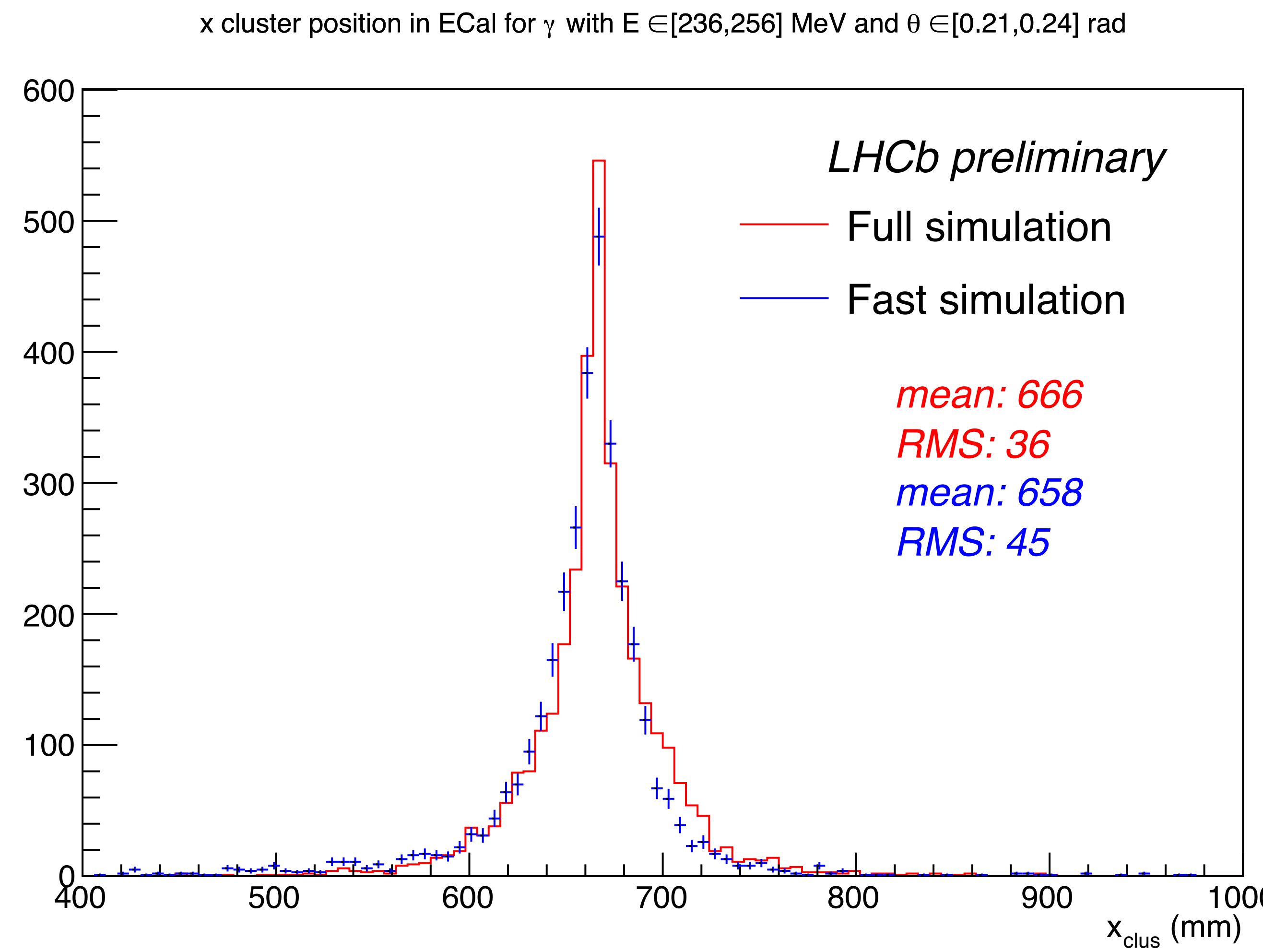
$$f_E = \frac{E - E_i}{E_{i+1} - E_i}$$

$$f_\theta = \frac{\theta - \theta_j}{\theta_{j+1} - \theta_j}$$



Encouraging Results



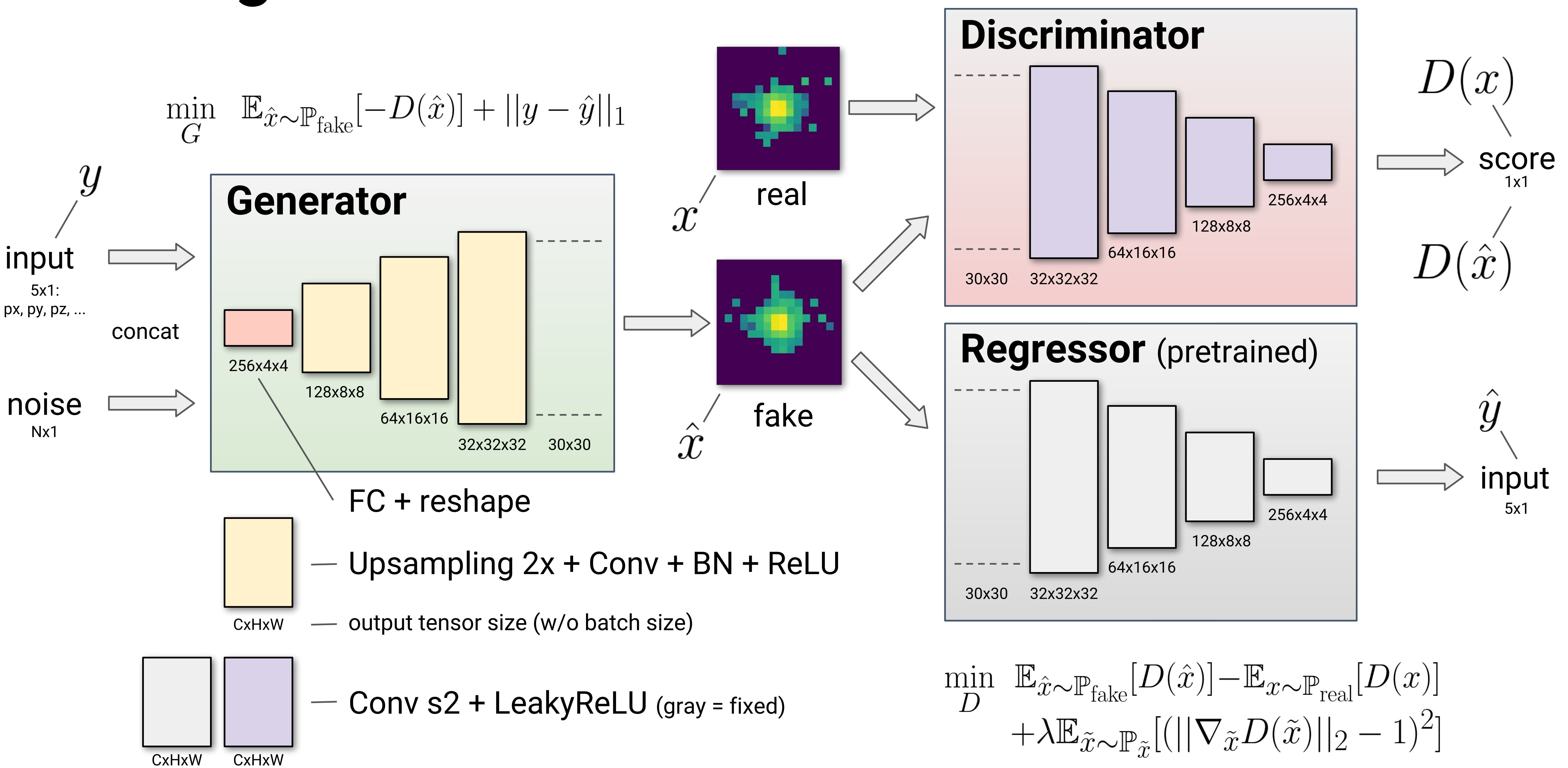


- Reasonable reproduction of energy resolution
- Good reproduction of 2nd and 3rd moments of spatial resolution

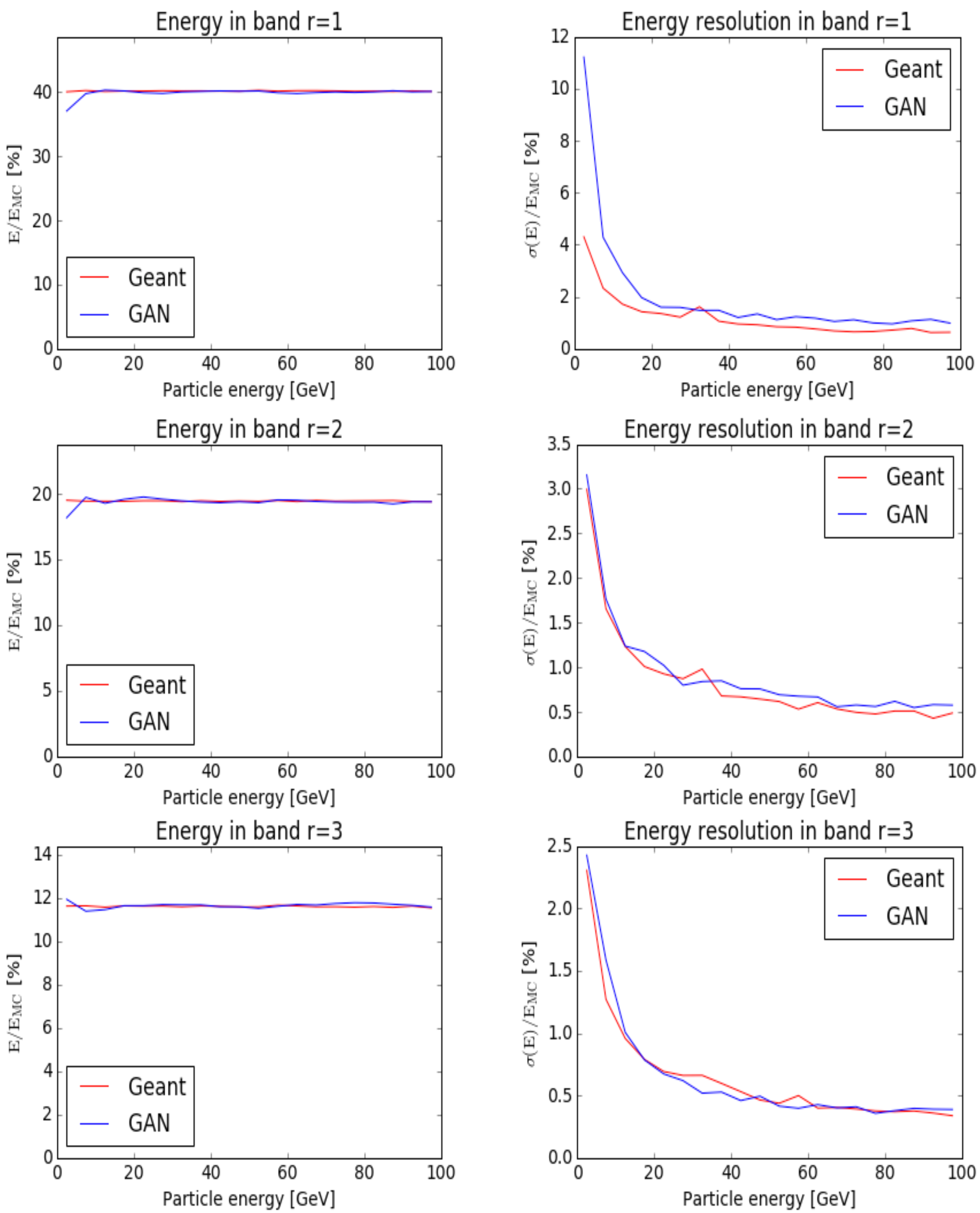
GAN Approach

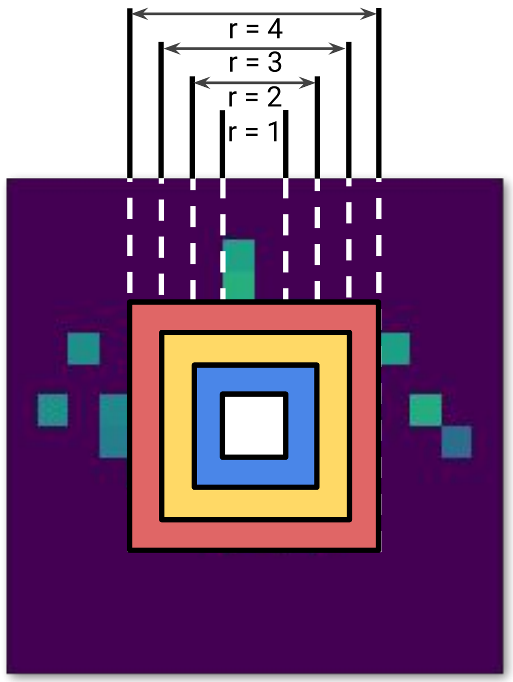
- Use stand-alone LHCb-like calorimeter GEANT4 setup to produce reference train and test samples
  - consider 30×30 ECAL cells of 2×2 cm^2 each
- Use Wasserstein GAN as a baseline
  - most popular, well trainable
- Add conditions (energy, position, angle)
  - add conditions as inputs to Generator
  - build and train Regressor
    - reconstruct impact particle parameter from CALO response
    - add distance between true input and reconstructed conditions into Generator Loss
- This makes generator to produce clusters which
  - look like true clusters
  - are reconstructable like true clusters

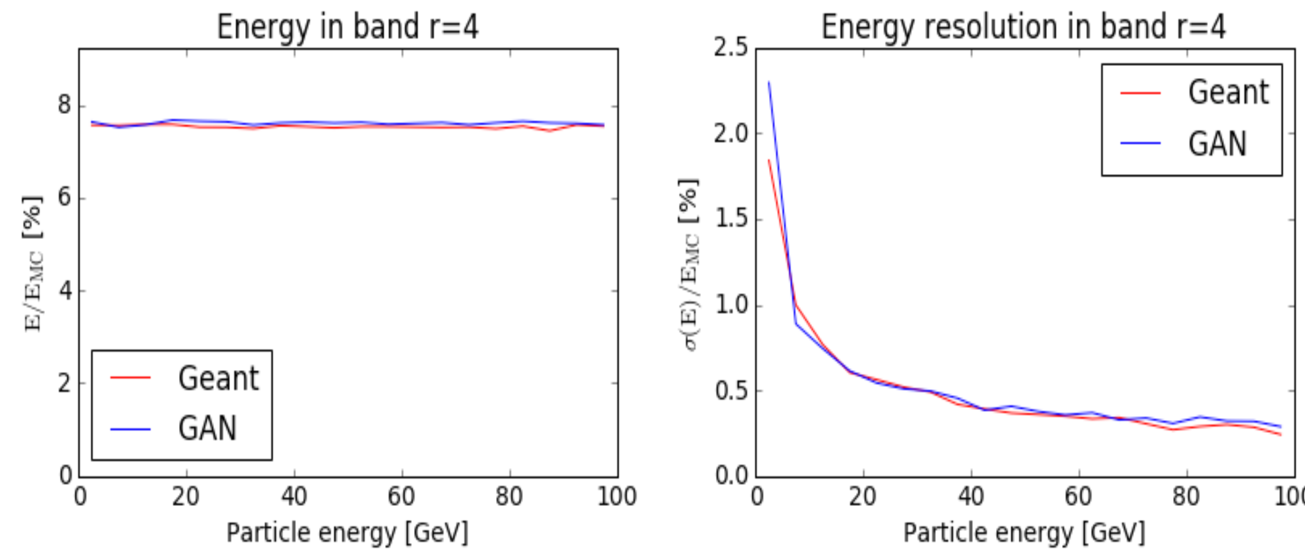
Training scheme



Performance in 1D (Energy)







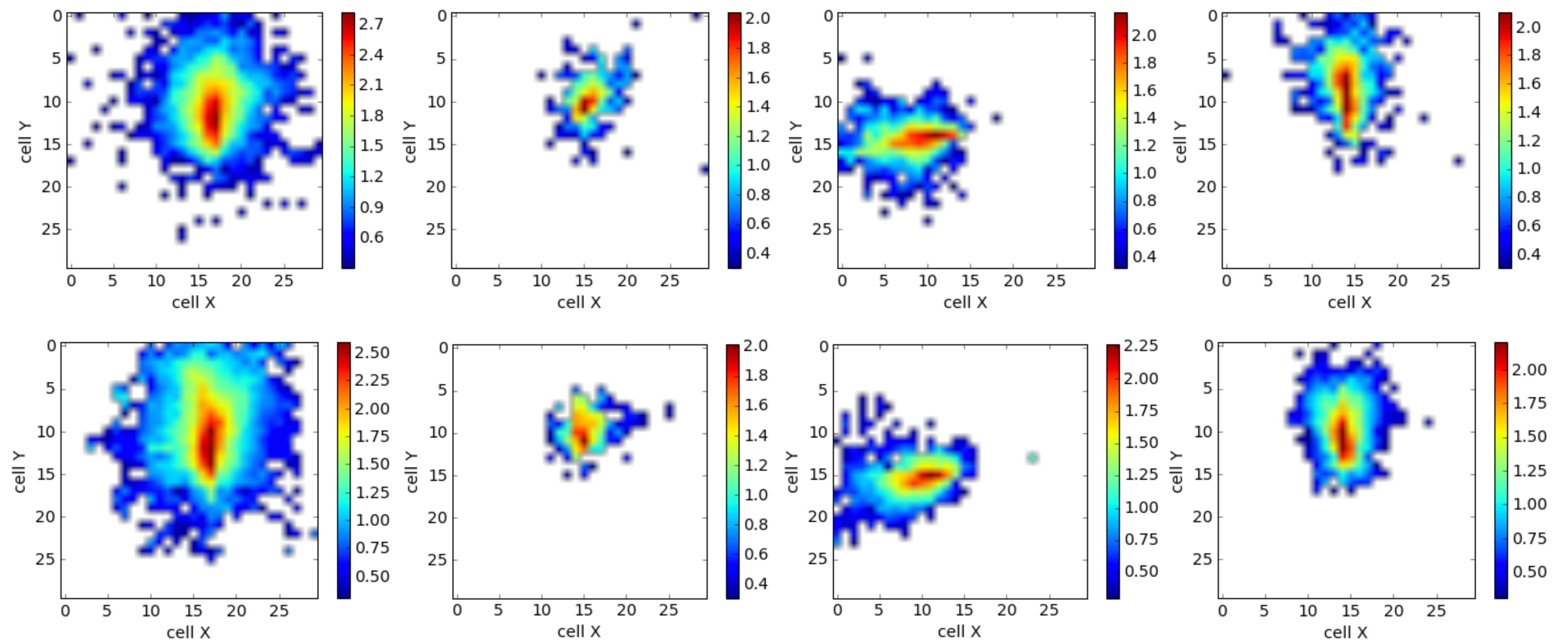
- Good reproduction of first and second moments for cluster shape

Generator in Full 5D

GEANT Simulated

log$_{10}$(cell energy)

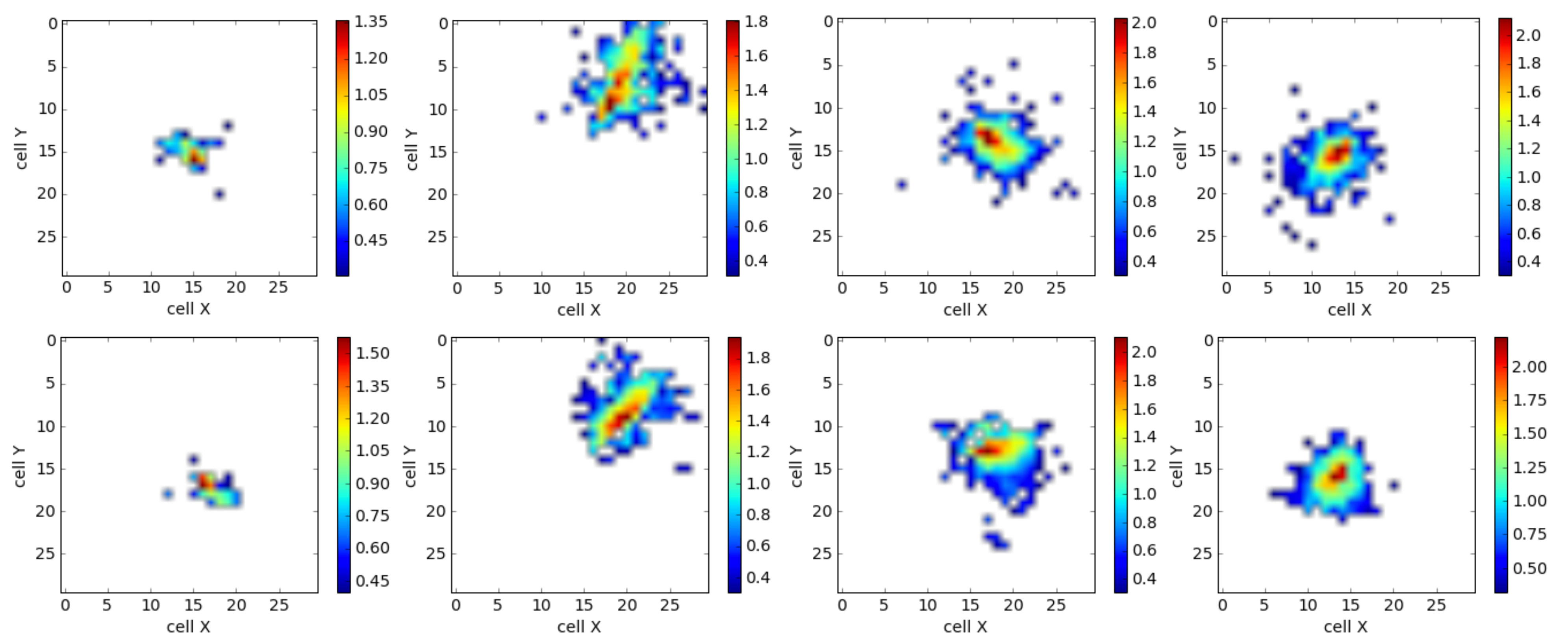
GAN Generated



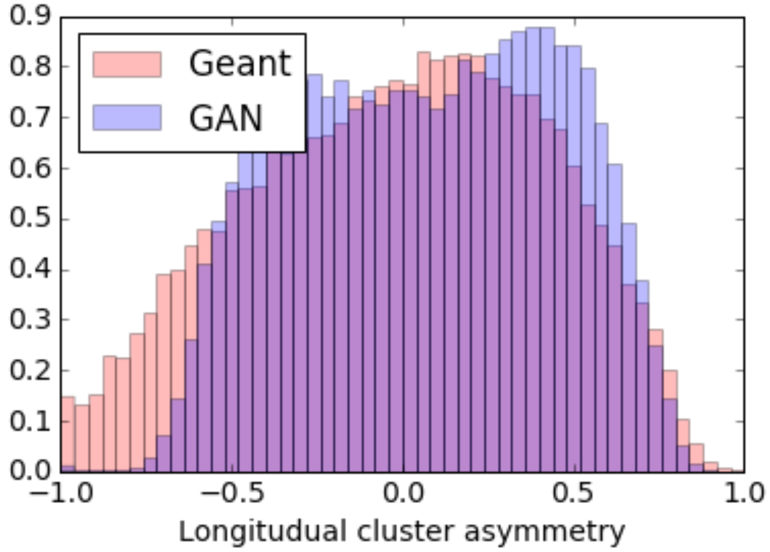
GEANT Simulated

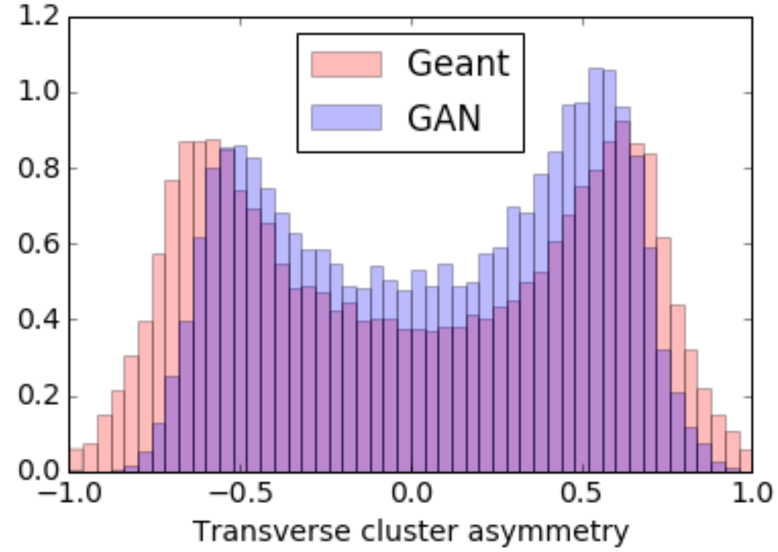
log$_{10}$(cell energy)

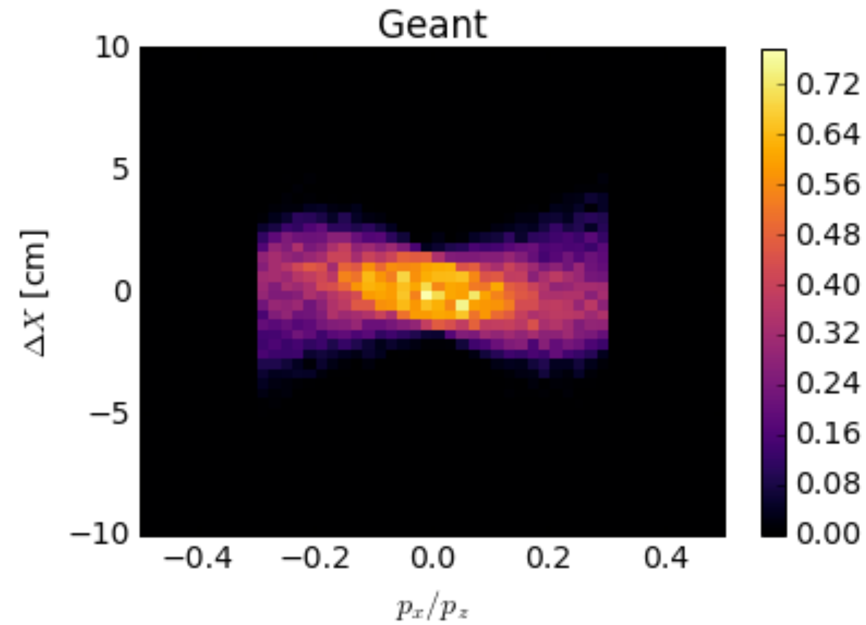
GAN Generated

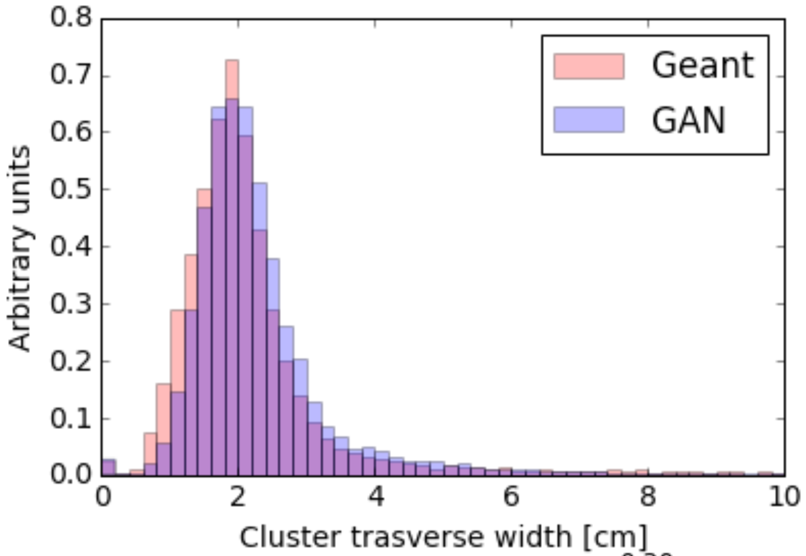


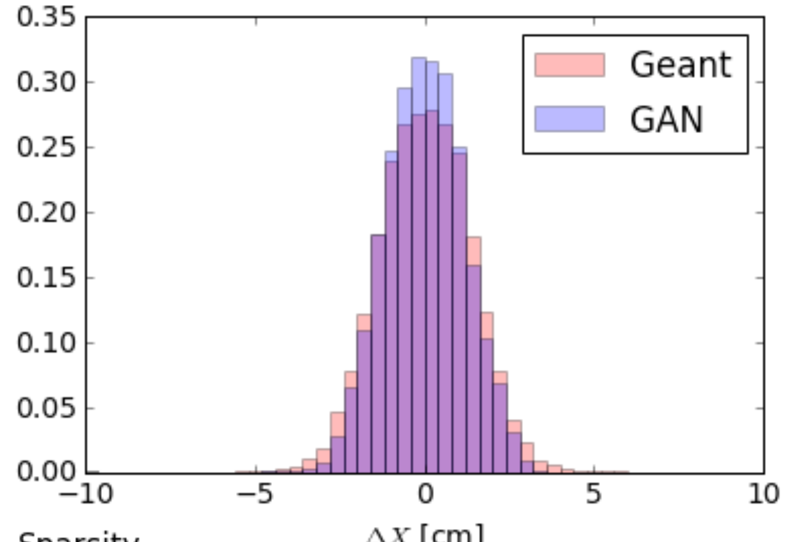
5D Performance Plots

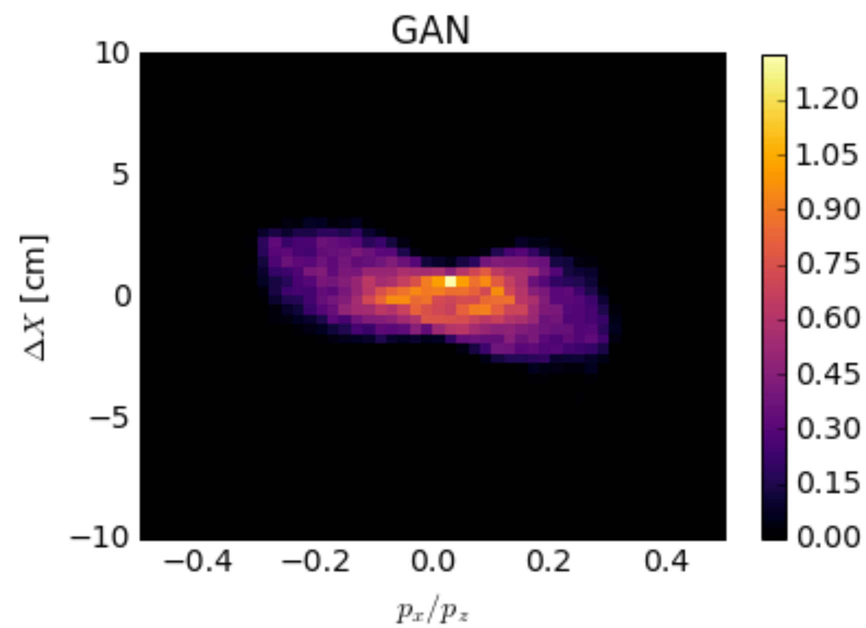


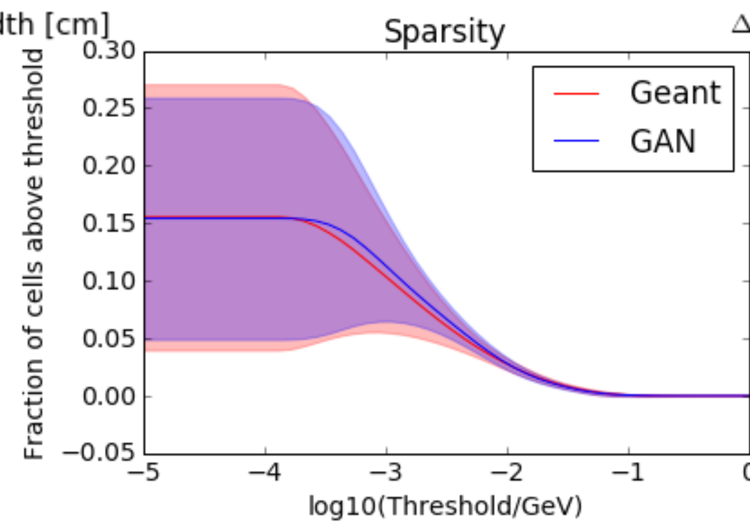












Conclusions

- Two fast simulators to generate calorimeter hits are are being developed
  - Shower library
    - developed infrastructure to plug fast simulation into LHCb simulation stack
    - demonstrated reasonable accuracy in reproducing simulated response
  - ML trained generative models
    - developed GAN architecture to generate conditional distributions for calorimeter response
    - GAN approach generally reproduce properties of electromagnetic showers
      - @ ~5 ms / shower on CPU, ~0.05 ms / shower on GPU (batch size 64)
  - Feasible approaches to speed up MC production for HEP experiments