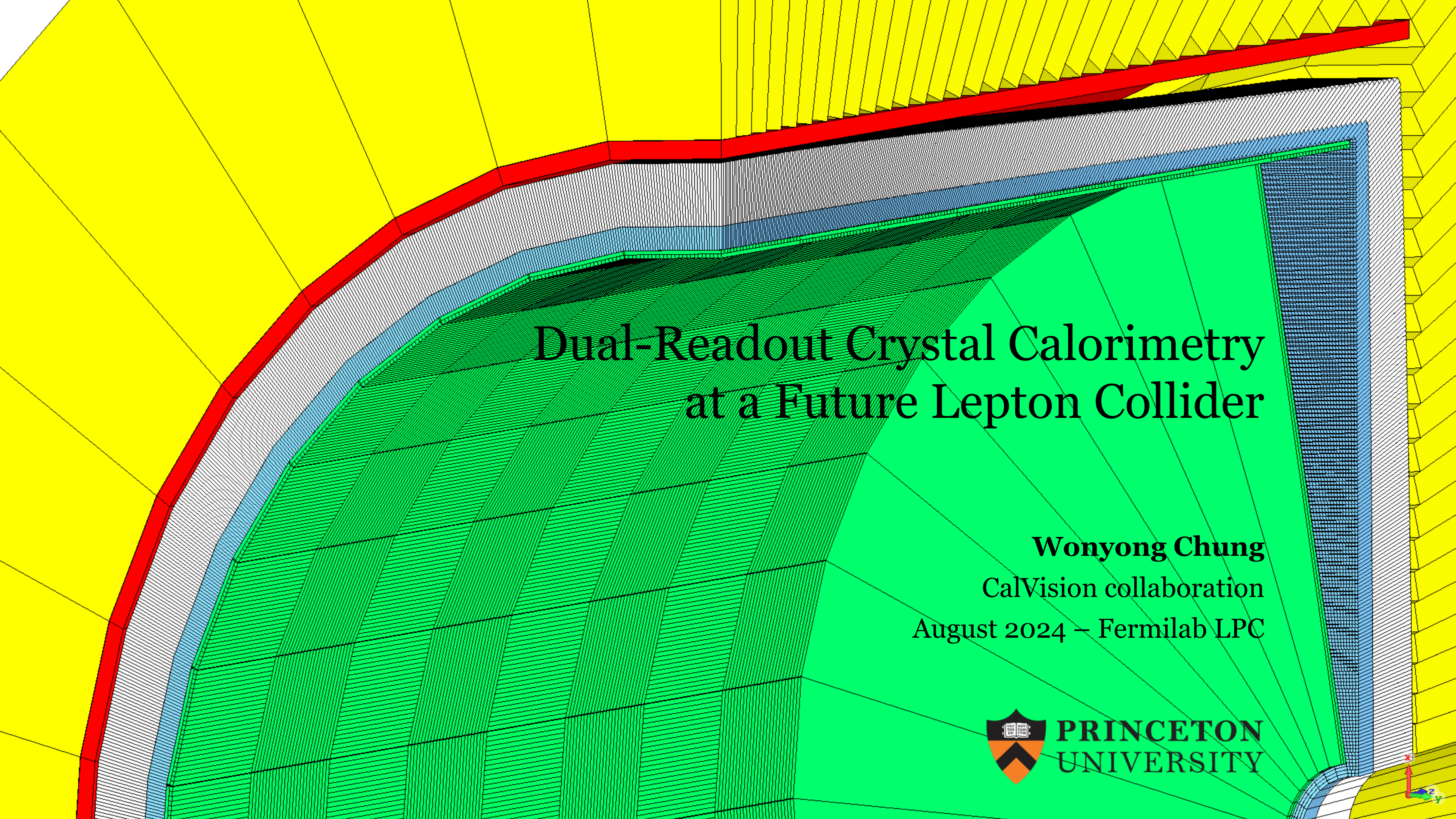


Dual-Readout Crystal Calorimetry at a Future Lepton Collider

Wonyong Chung
CalVision collaboration

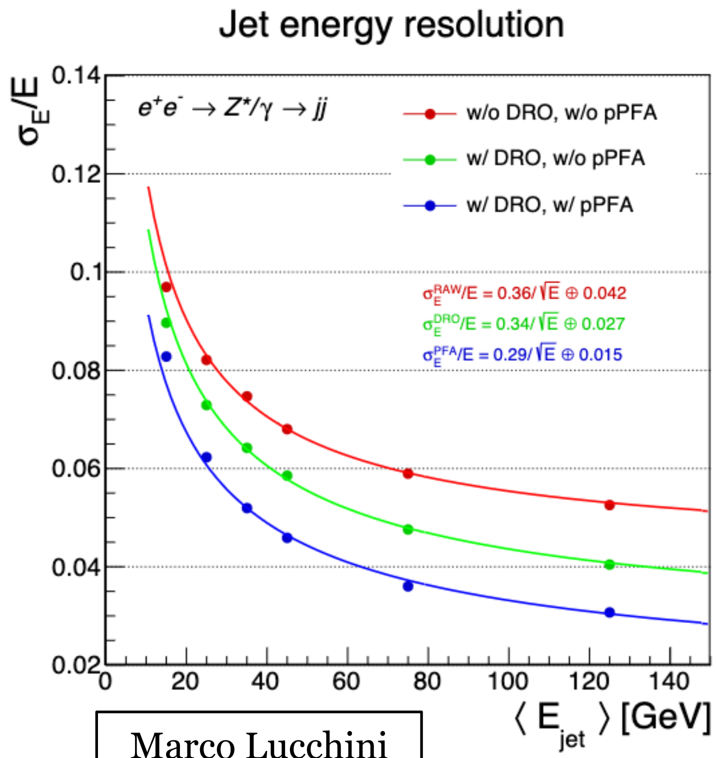
August 2024 – Fermilab LPC

PRINCETON UNIVERSITY

Segmented Crystal EM Precision Calorimeter (SCEPCal)

New perspectives on segmented crystal calorimeters for future colliders
JINST 15 (2020) P11005 [2008.00338]

Particle Flow with a Hybrid Segmented Crystal and Fiber Dual-Readout Calorimeter
JINST 17 (2022) P06008 [2202.01474]



Marco Lucchini

This talk:

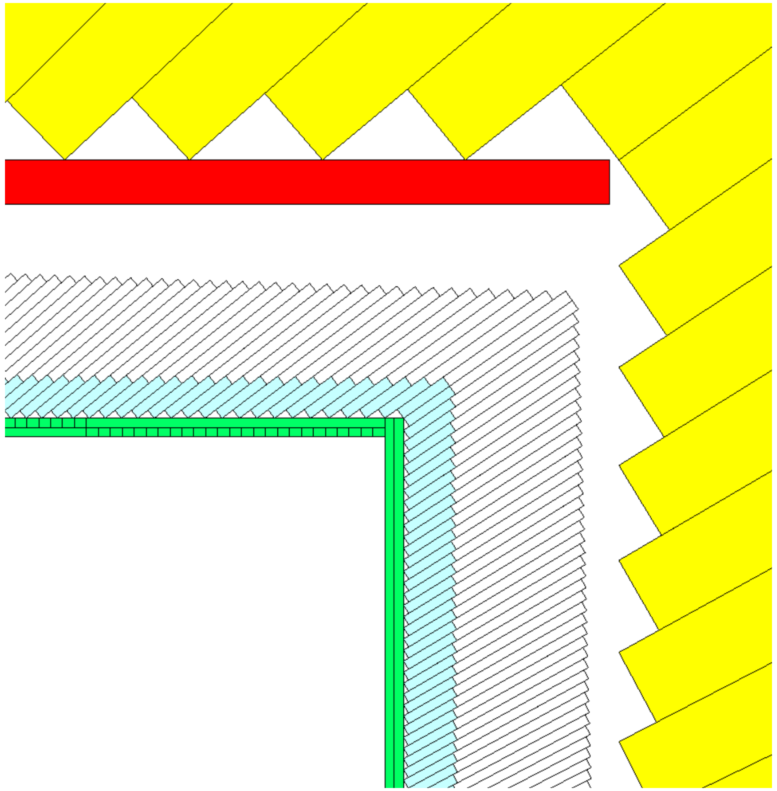
Detector and simulation
Implementation in dd4hep

First full simulation studies of single photons

Event displays

EM pointing resolution studies

Reconstruction strategy in the AI/ML era



Implementing a new detector from scratch in dd4hep

- **What you need:**

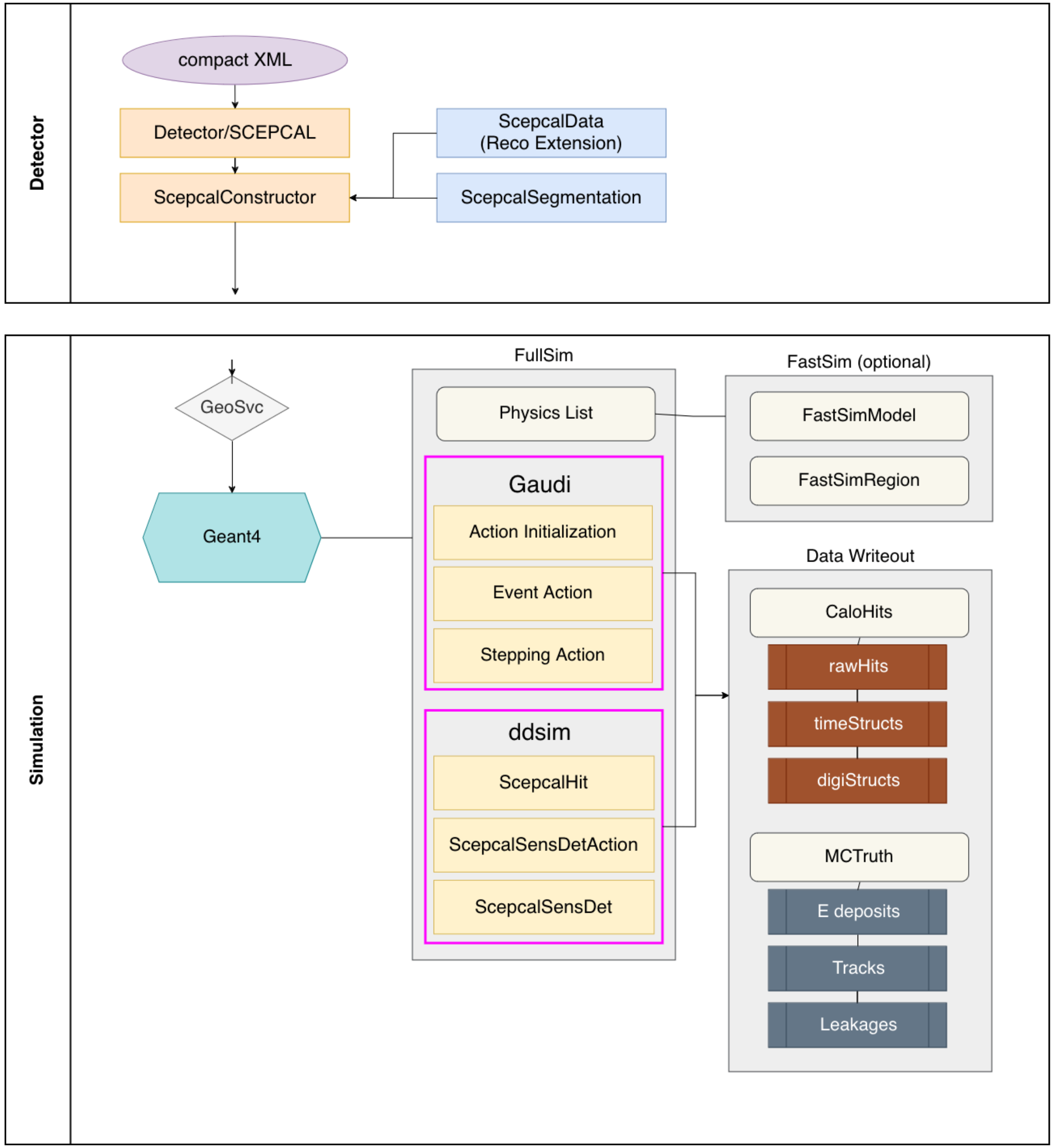
- dd4hep (detector description):
  - Compact XML description
  - Geometry constructor
  - Material definitions
  - Segmentation class, factory, handle
    - Technically optional, but needed if implementing a geometry not already in dd4hep (e.g. projective)
- ddsim:
  - Sensitive Detector (assuming a new type – e.g. DR w/ S&C)
    - Class definition, wrapper
    - Custom SD Hit, Action classes
  - Reco extension structs – metadata attached to geometry
    - Optional, useful for event displays
  - Accepts more variety of hep inputs, no reco chain (yet)

- Gaudi + k4 (FWCore, Geant4Sim, Gen):
  - Geant4 Actions (Event/Stepping), Initializations
  - Rough integration with hep inputs, better reco chain

- Steering (config) files

- Remaining: digitization, wrapping for instrumentation, cooling, etc.

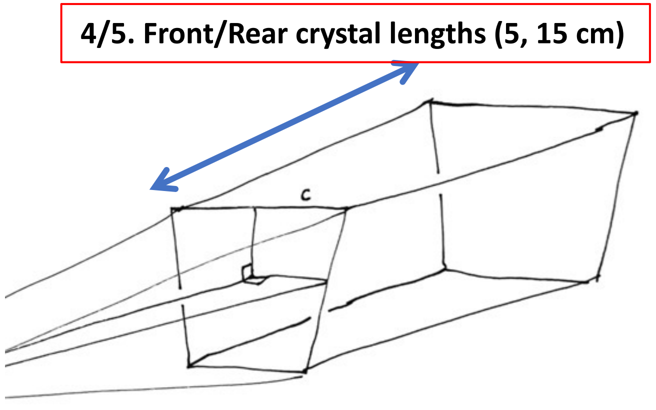
- Special thanks to Sanghyun Ko and dual-readout repository

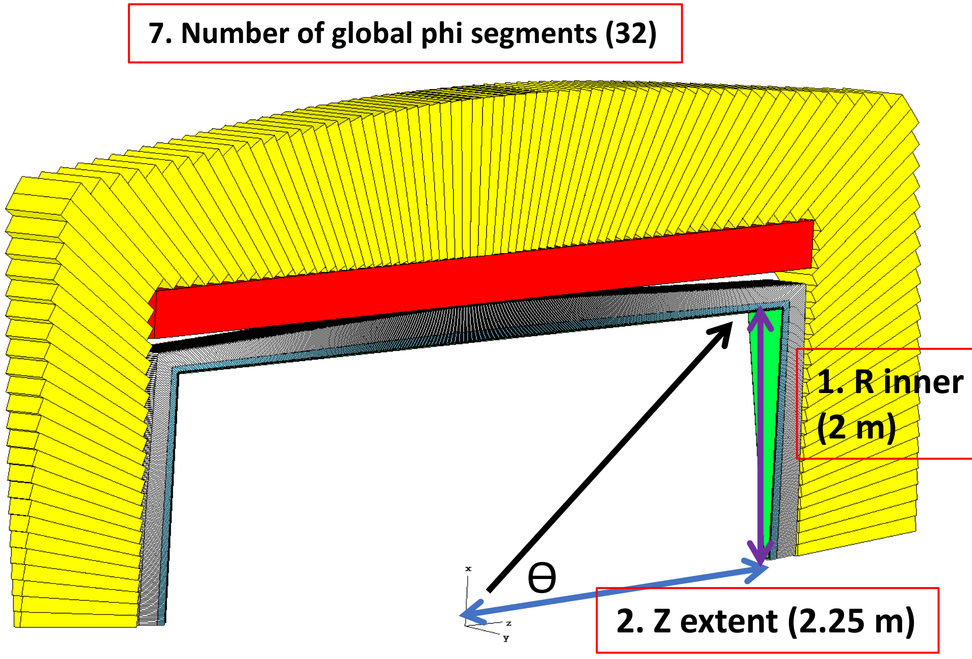


Detector description

- **Fully parametrized construction**

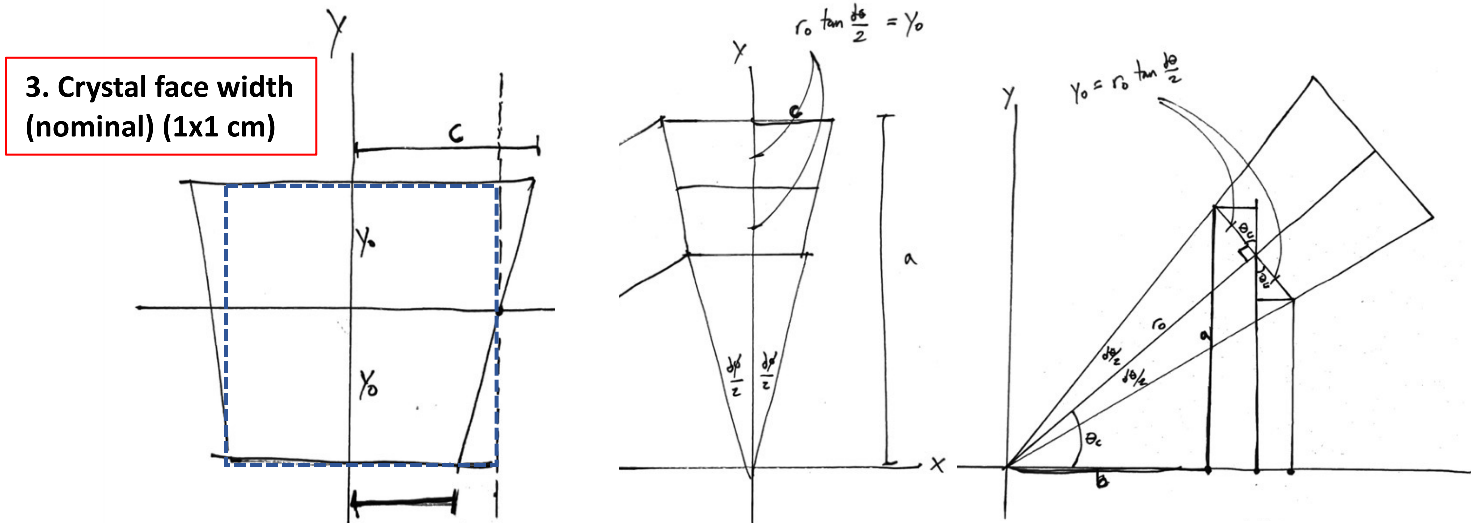
- Only 7 parameters needed:
  - Inner radius
  - Z extent of barrel
  - Crystal face width (nominal)
  - Front crystal length
  - Rear crystal length
  - Timing crystal thickness (nominal)
  - Number of phi segments
    - Ensures hermeticity
    - Enables timing layer
    - Takes care of projective gaps





- Geometry optimizations
  - Intermediate envelope volumes, <1000 volumes each
  - Orange slices (barrel)
  - Rings (endcap)

  - ~10x speed/memory improvement vs. monolithic single container

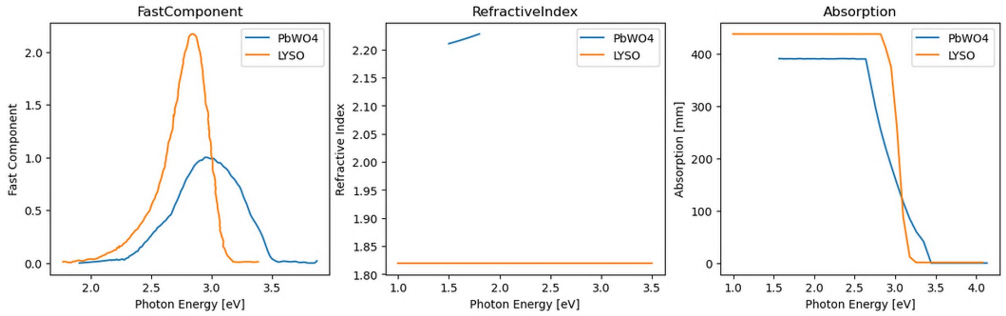


Simulation details

- “SimDRCalorimeterHit” custom edm4hep class
    - Records S/C counts
    - Wavelength/time bins (300-1000 nm, 0-300 ns)

- Material properties have big effect

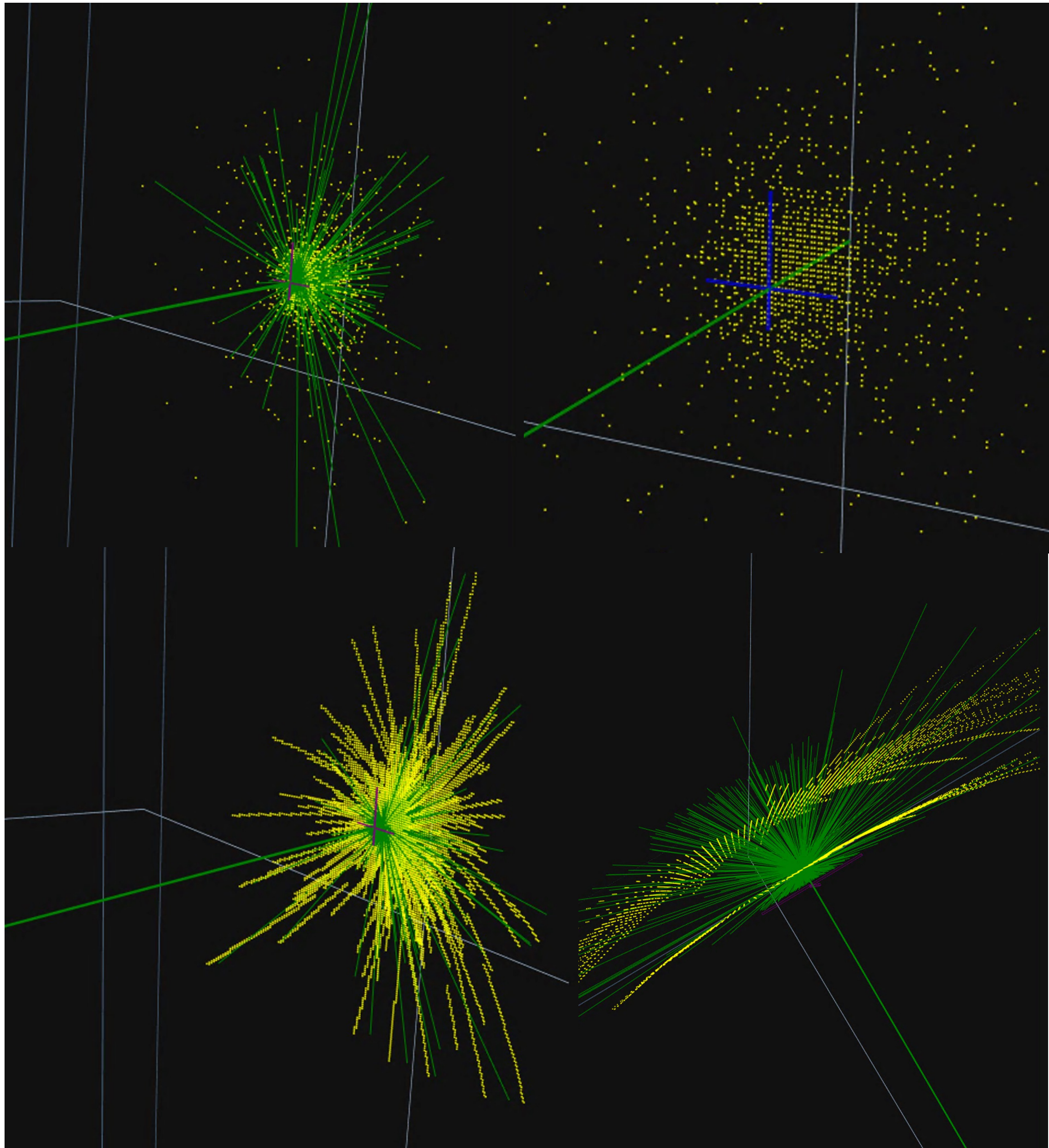
- Can get very different results depending on the “sensitive action” filter used for the subdetector
    - Step-level energy deposit filter
    - Track-level wavelength filter



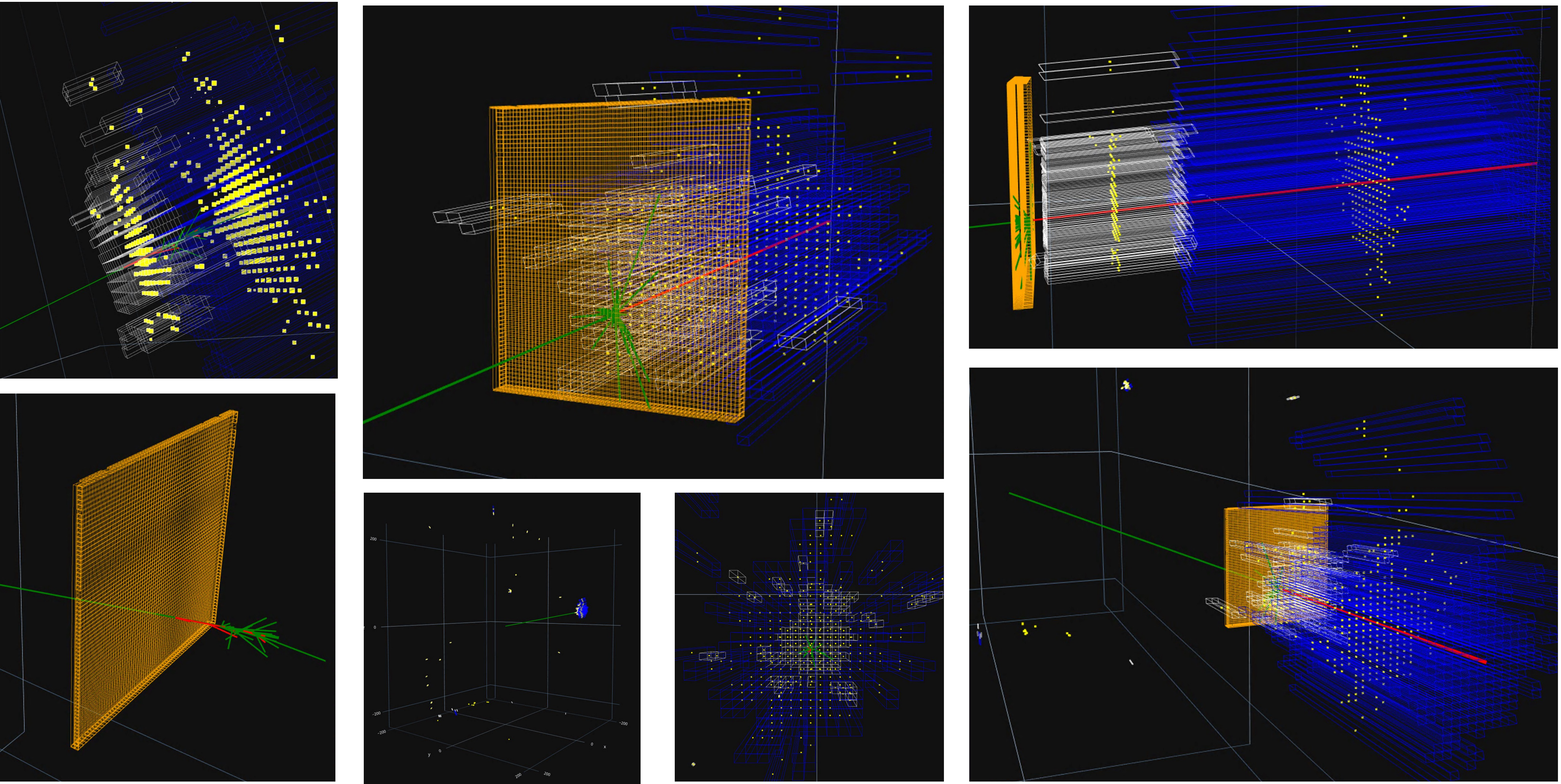
edep0

10 MeV gamma

wavelength (200nm)



Event displays – 10 GeV gamma



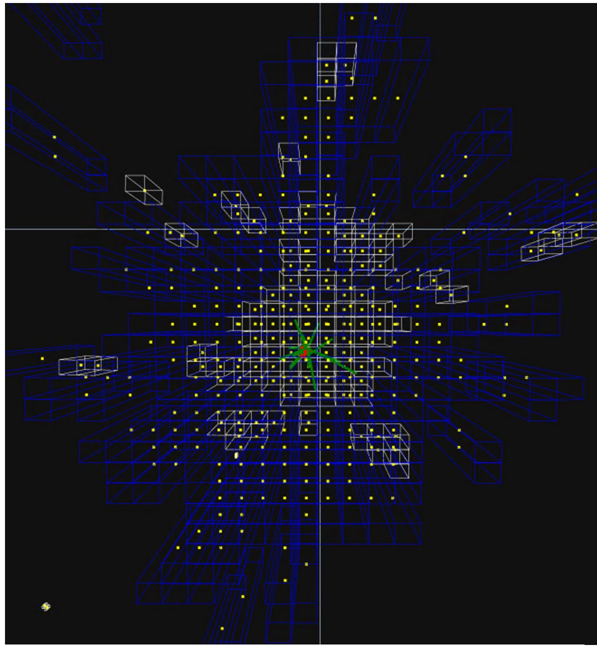
Projective layer pointing studies – 10 GeV gammas

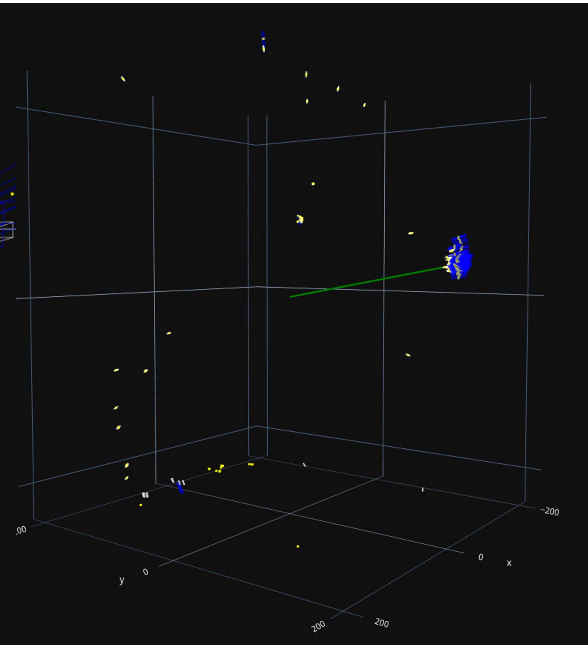
1. Separate hits into front and rear hits

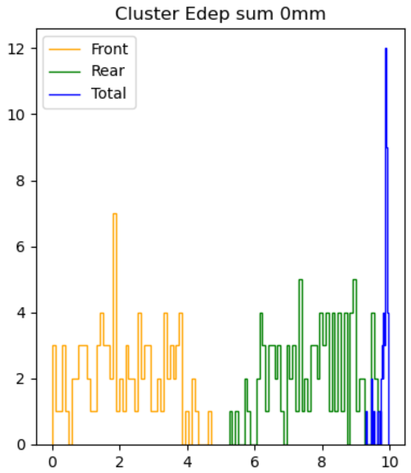
2. Simple clustering:
    - Pick max Edep hit
    - Sum up hits in expanding samples of angular radius around the hit, up to 11.25 deg (pi/16)
    - Pick the sample that contains 95% of E of pi/16

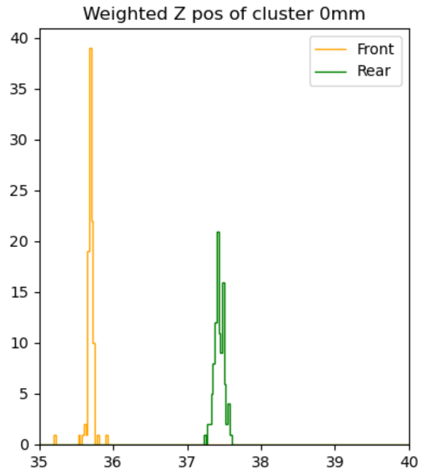
3. Within the cluster:
    - Pick single hits that are >2.12% (e^-3.85) of cluster E
    - Weight the z pos of the hit by the amount of excess over cut
    - Average the weighted z-positions -> z-pos of cluster

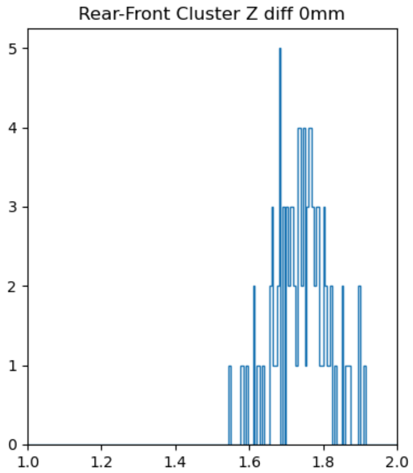
4. Plot difference of rear-front clusters, 100 events

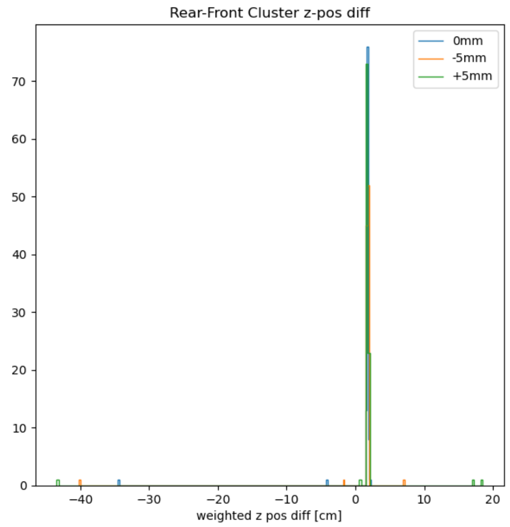


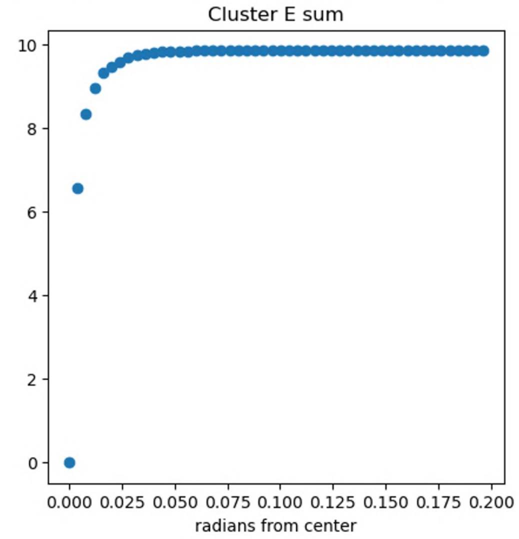






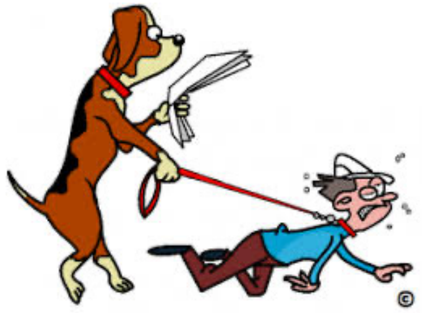






Reconstruction strategy

GravNet [1902.07987] (GNN, 2019, distance-weighted, HGCAL)
MLPF [2101.08578] (GNN, 2021, jets/pileup, CMS, comparable to PFA)
ParticleNet [2309.13231] (GNN, 2019, jet tagging unordered particle clouds, CEPC)
Particle Transformer [2202.03772] (Transformers, 2022, kinematics + pairwise interactions, JetClass 100M open data set)
… and many more

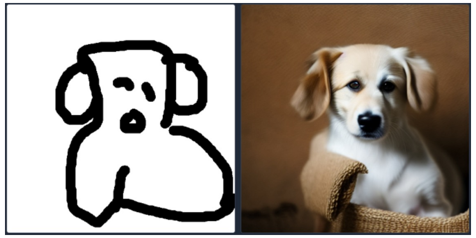


- Trend towards unordered particle clouds
    - **Clustering becomes an inner detail of overall reconstruction**
- **Join the fray?**
    - Could start throwing calo hits at the latest NNs…
    - Ongoing work to converge into full detector sim (tracker + ECAL + HCAL)

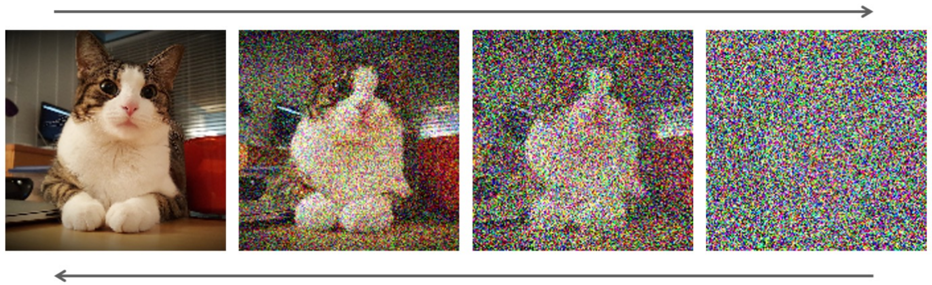
- Meanwhile, need to understand the different neural networks
    - **Unfortunate reality of AI/ML: physics follows computer science**
    - Different types of NNs were originally built, optimized, tuned for a specific task
        - Image recognition (CNNs)
        - Image generation (GANs, Diffusion)
        - Etc.
- **Study how to reframe the physics question into these contexts as much as possible**
    - Understand and integrate the strengths of rule-based PFAs into NNs

Pick the right neural network for the detector

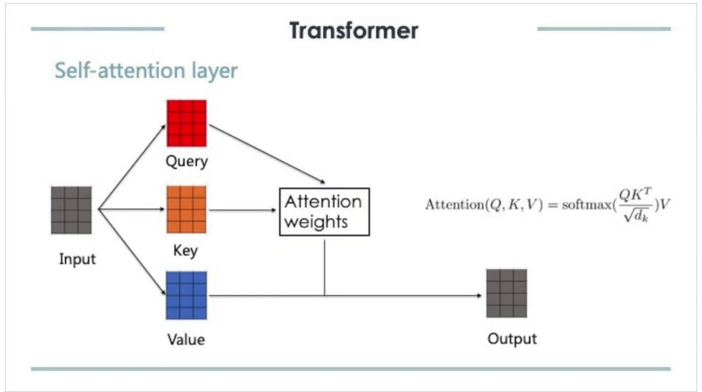
StableDiffusion + ControlNet



- **Tile-based sampling calorimeters:**
  - Diffusion models – **inherently designed to reconstruct from noise**
  - ControlNet/Inpainting – can use **tracks as a guide**



  - Reconstructing from noise – potential to integrate systematics into NNs



- **Multi-segmented/homogeneous calorimeters:**
  - Transformers – **designed to maintain long range context in sequential data, i.e. NLP**
  - Use **tracks for attention, segmentation and timing provide sequence**

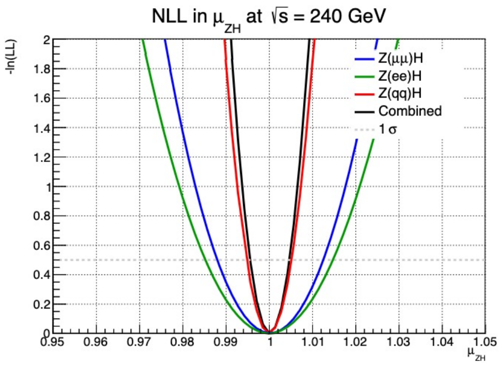
`The quick brown fox jumped over the lazy dog`

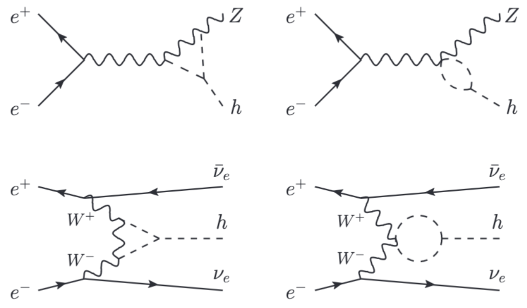
  - GNNs are generally short range and generally use distance, arguably same kind of assumptions as rule-based PFA

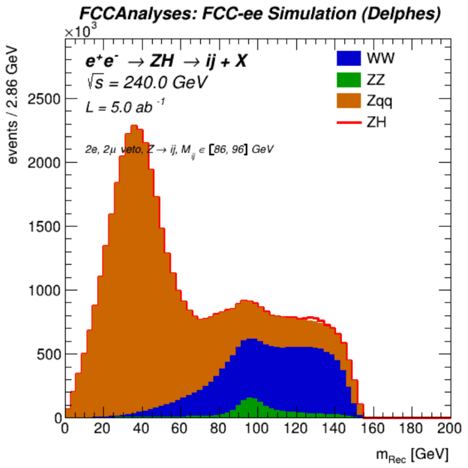
- **Mixed/"Semi-Homogeneous"**:
  - Generative Adversarial Networks – Diffusion model in sampling portion, transformer/GNN in homogeneous, directly leverage the detector design

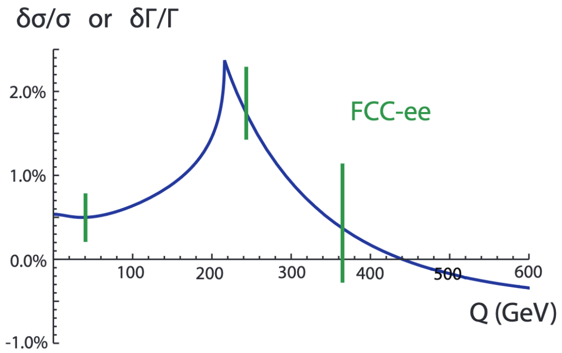
Train according to the physics case

- **Physics case is essential to training approach**
- **For a future lepton collider:**
  - HHH from loop correction to the HZ cross section
  - Z(qq)H apparently dominates
  - Hinges on ability of detector to reduce ZZ background
- Foundational approach? – orders of magnitude more data, unexpected performance gains seen in LLMs (circa ~2016-17)
  - Feed all processes/jets at once instead of training on specific subsets, see what happens?
    - (Not saying we should actually do this)
- Need orders of magnitude better performance



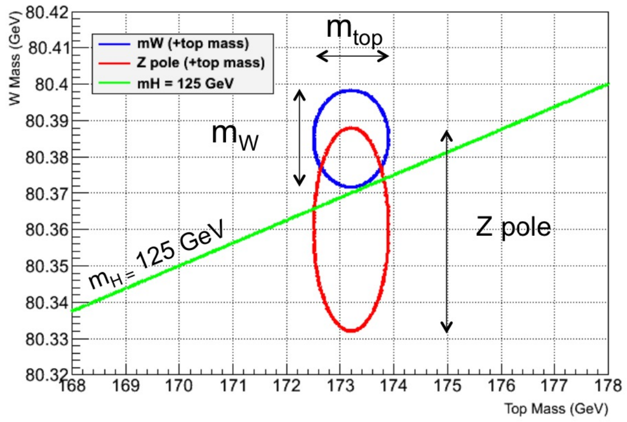


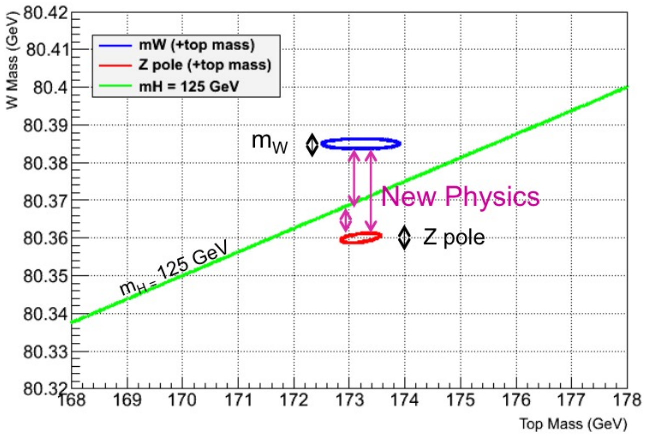




A special Higgs challenge
https://arxiv.org/pdf/2106.15438

Probing the Higgs sector at the FCC-ee
https://cds.cern.ch/record/2835483





Baselines, new technologies, new benchmarks

- Fully parametrized, hot-swappable sub-detector designs in new frameworks
  - Fast iteration and optimization possible
  - A new world of low-level simulation study

- Top-down view of detector geometry and data readout
  - "Triggerless" DAQ, flexible online software (see LHCb)
  - Next-gen ASICs *should* be fast enough to run real-time inference on front-end electronics

- Opportunity for unprecedented vertical integration

- Current designs are arguably the 'final forms' of CMS, ATLAS, etc.
  - We need more new technologies
    - Lattice-oriented crystals
    - Chromatic calorimetry
    - Quantum sensors for HEP [2311.01930]

- Growing understanding of NN maths as relevant to physics
  - "A mathematical perspective on Transformers" [2312.10794]

    We develop a mathematical framework for analyzing Transformers based on their interpretation as interacting particle systems, which reveals that clusters emerge in long time.

Constraints and limitations for FCC detectors

Physics process	Rate (kHz)
Z decays	100
$\gamma\gamma \rightarrow$ hadrons	30
Bhabha	50
Beam background	20
Total	$\sim$ 200

Subdetector	Physics	Background/noise
CLD Vertex Detector	150 MB/s	6 GB/s
CLD Tracker	160 MB/s	10 GB/s
IDEA Drift Chamber	60 GB/s	2 GB/s
IDEA Si Wrapper	32 MB/s	0.5 GB/s
IDEA DR Calorimeter	10 GB/s	1.6 TB/s *
IDEA pre-shower	320 MB/s	820 MB/s
IDEA Muon Detector	4 MB/s	67 MB/s

* Assuming no suppression for isolated counts