MACHINE LEARNING FOR FUTURE COLLIDER SIMULATIONS

Raghav Kansal
UC San Diego, Fermilab

US FCC Workshop
26th March 2024

WHY?

WHY?

1. Speed ⇒ Fast simulations

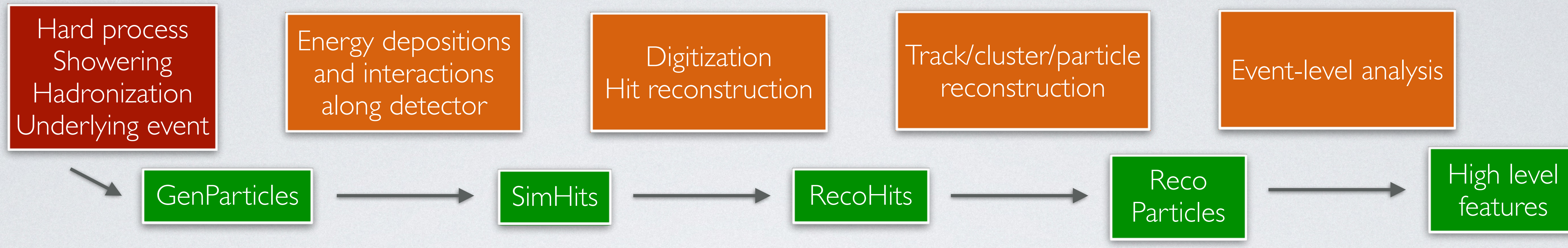
- Needed for HL-LHC
- Useful for preliminary FCC studies, perhaps for operation as well (GPUs!)

WHY?

1. Speed ⇒ Fast simulations
   - Needed for HL-LHC
   - Useful for preliminary FCC studies, perhaps for operation as well (GPUs!)

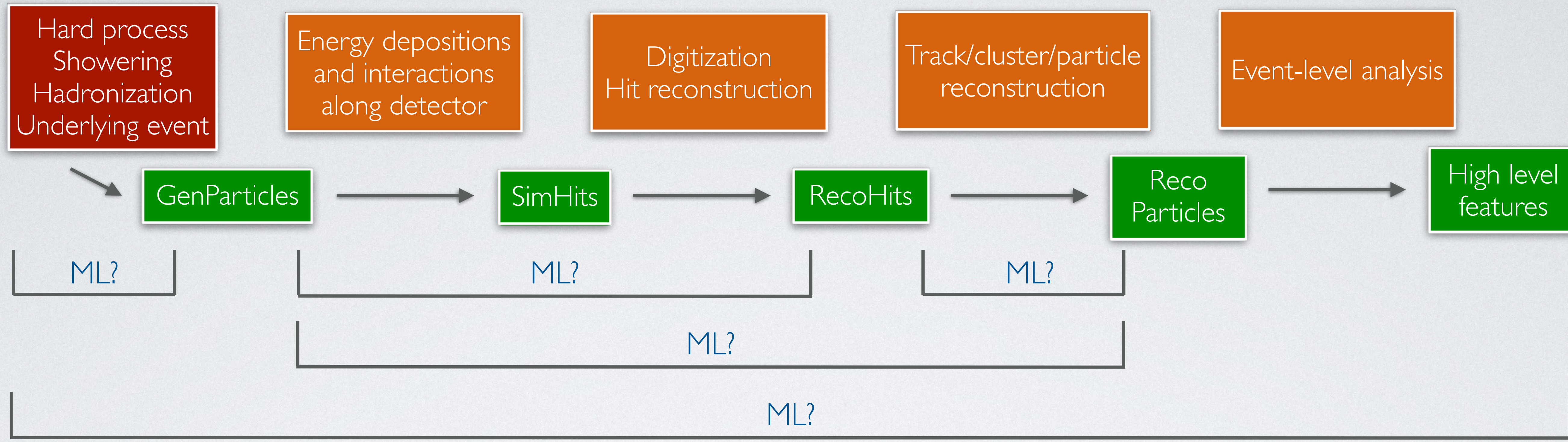
2. Differentiability ⇒ Optimizing detector design
   - O(100-1000s) of detector parameters to tune for ultimate precision
   - Can we do this systematically: optimize simultaneously for maximum H/Z coupling etc. sensitivities + cost

HOW?

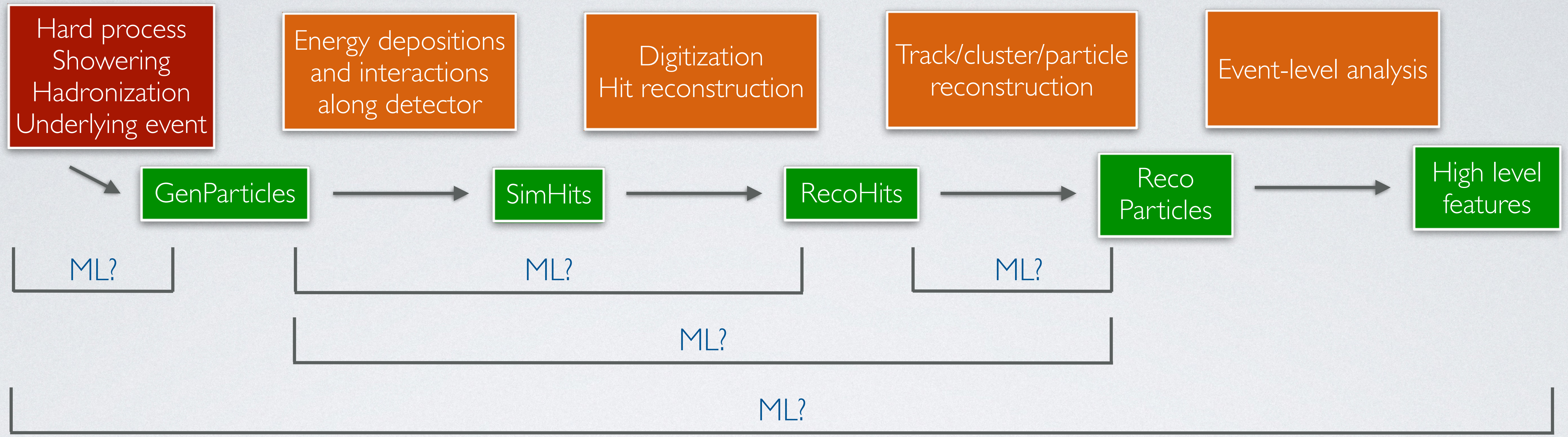


Hard process
Showering
Hadronization
Underlying event
Energy depositions and interactions along detector
Digitization
Hit reconstruction
Track/cluster/particle reconstruction
Event-level analysis
GenParticles
SimHits
RecoHits
Reco Particles
High level features

HOW?

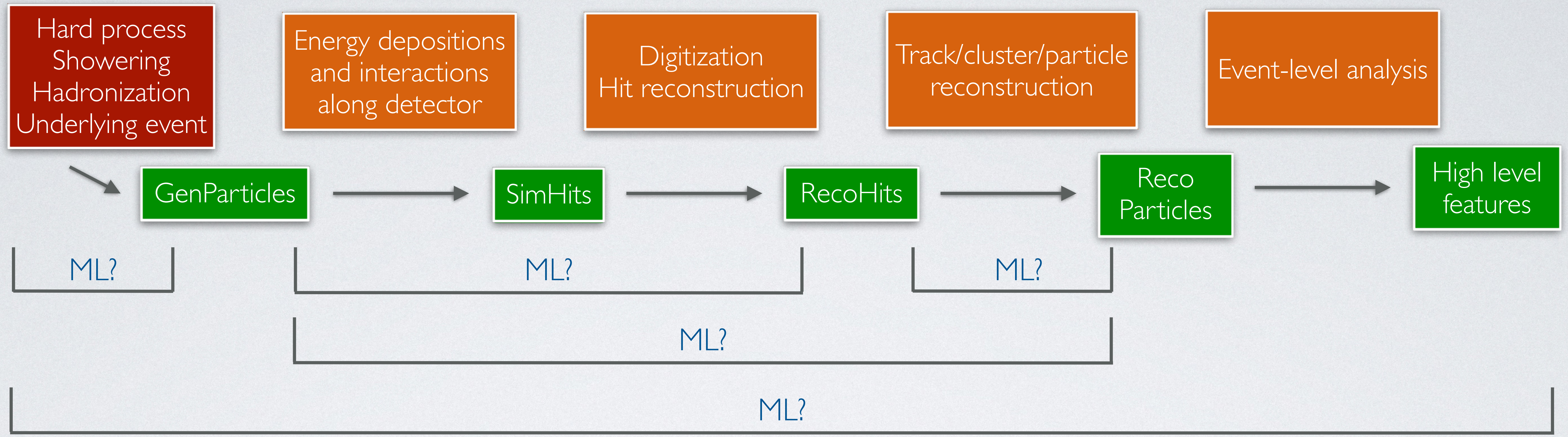

Hard process
Showering
Hadronization
Underlying event
Energy depositions and interactions along detector
Digitization
Hit reconstruction
Track/cluster/particle reconstruction
Event-level analysis
GenParticles
SimHits
RecoHits
Reco Particles
High level features
ML?
ML?
ML?
ML?
ML?

HOW?



1. Rewrite existing code with automatic differentiation (e.g. JAX)
   - Challenges: differentiating through discrete or stochastic steps (e.g. splittings, clustering)

HOW?



1. Rewrite existing code with automatic differentiation (e.g. JAX)
    - Challenges: differentiating through discrete or stochastic steps (e.g. splittings, clustering)
2. Surrogate models: train ML algorithms on inputs and desired outputs
    - Heterogeneous models for each step
    - Challenges: model accuracy, interpretability, validation + incorporating discrete parameters

METHODS AND EXAMPLES

AUTO-DIFFERENTIATION

- Keep track of gradients at each step → repeated chain rule*

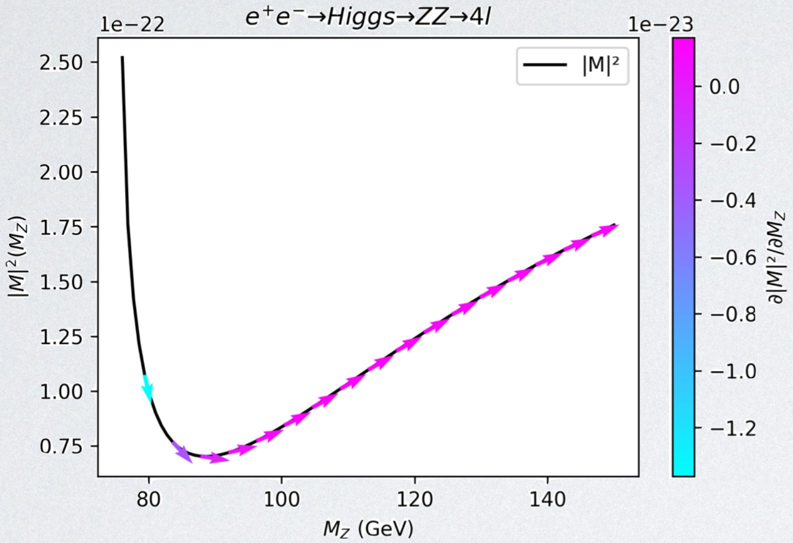
*Review 2403.14606

AUTO-DIFFERENTIATION

- Keep track of gradients at each step → repeated chain rule*

MadJax (Differentiable MadGraph)

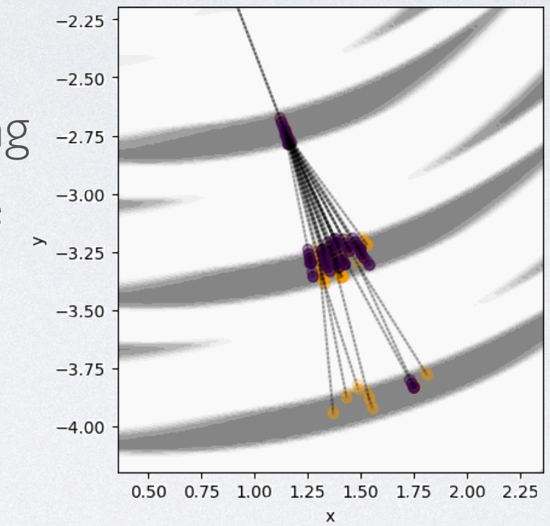
L. Heinrich and M. Kagan (2023)

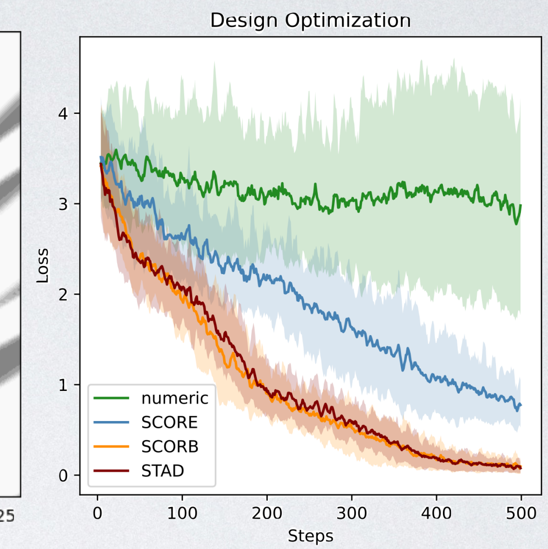


Derivatives of discrete and random processes

L. Heinrich and M. Kagan (2023)

- Estimating gradients using stochastic AD and score
- Optimizing for toy detector inner radius





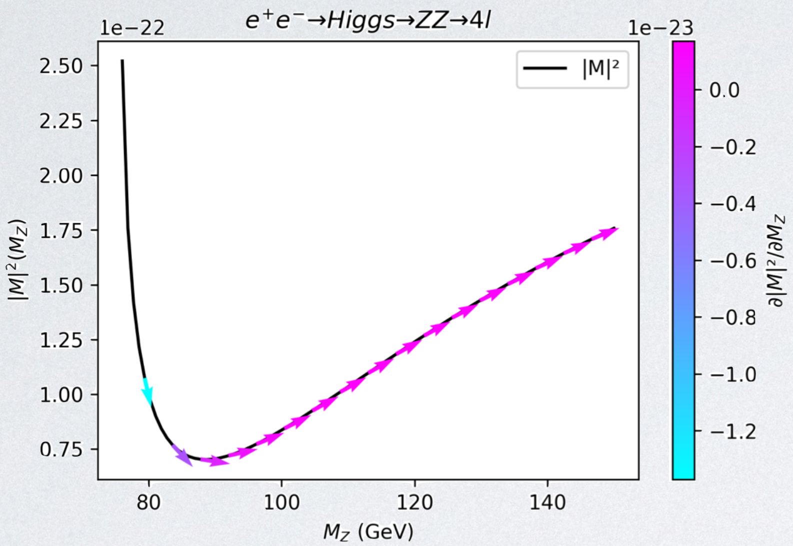
*Review 2403.14606

AUTO-DIFFERENTIATION

- Keep track of gradients at each step → repeated chain rule*

MadJax (Differentiable MadGraph)

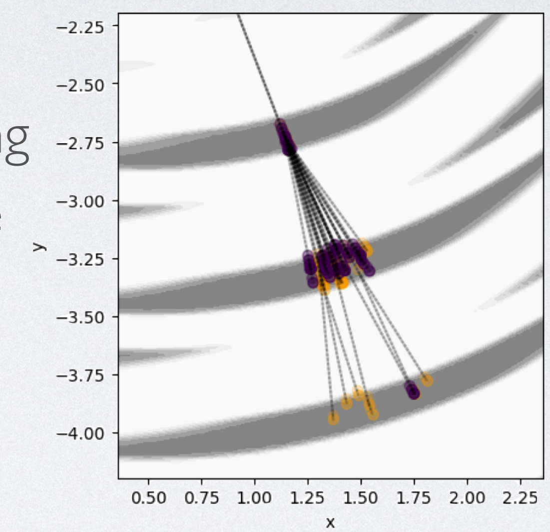
L. Heinrich and M. Kagan (2023)

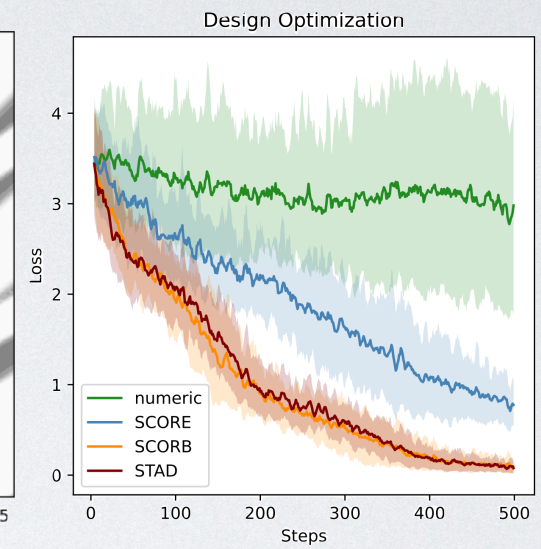


Derivatives of discrete and random processes

L. Heinrich and M. Kagan (2023)

- Estimating gradients using stochastic AD and score
- Optimizing for toy detector inner radius

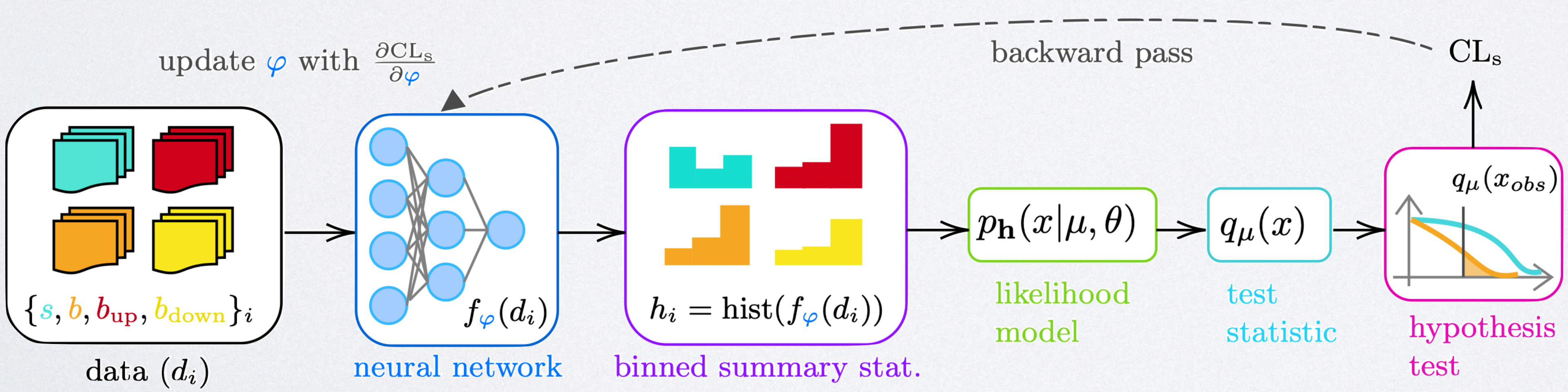


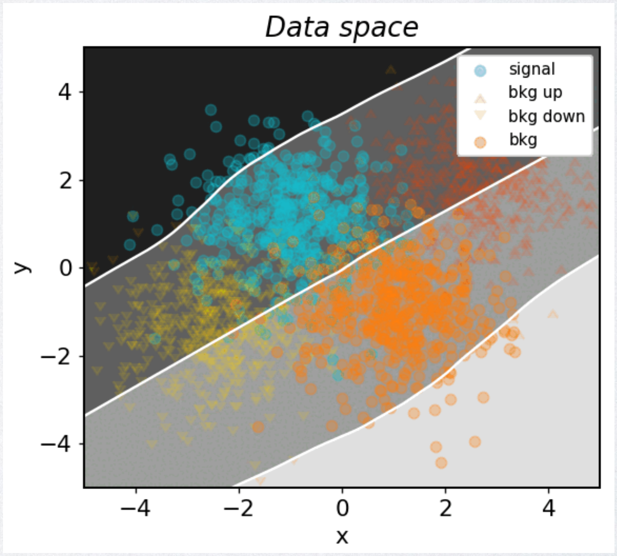


neos: Optimizing directly for p-values / FoMs

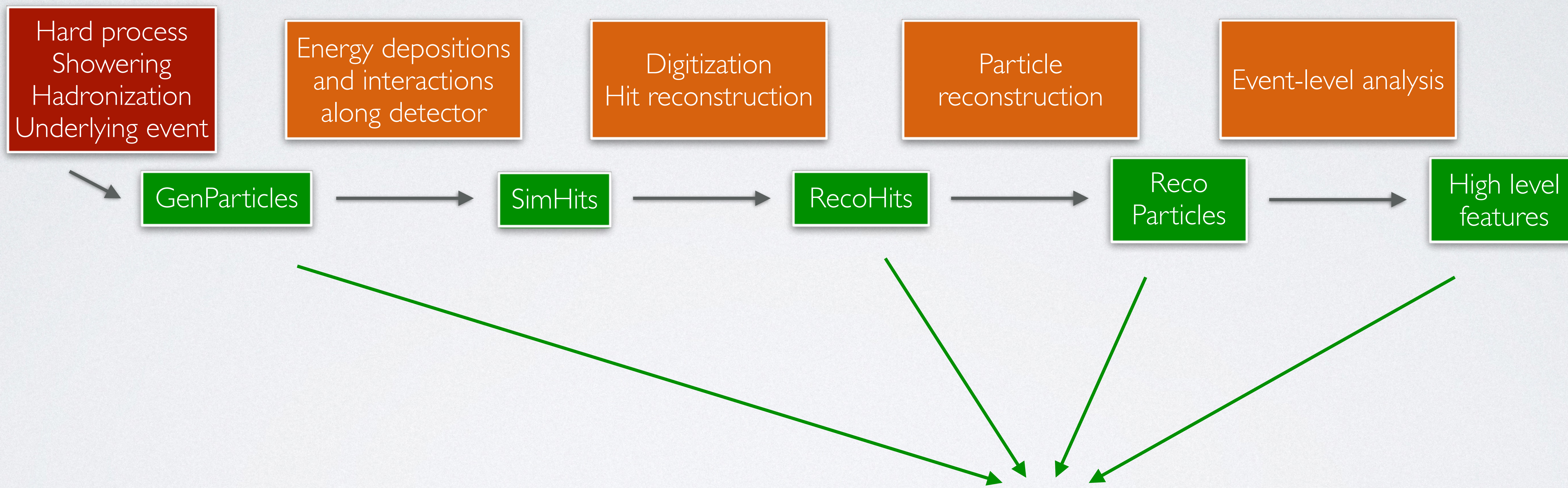
N. Simpson and L. Heinrich (2023)

- Tested on 2D toy problem with uncertainties





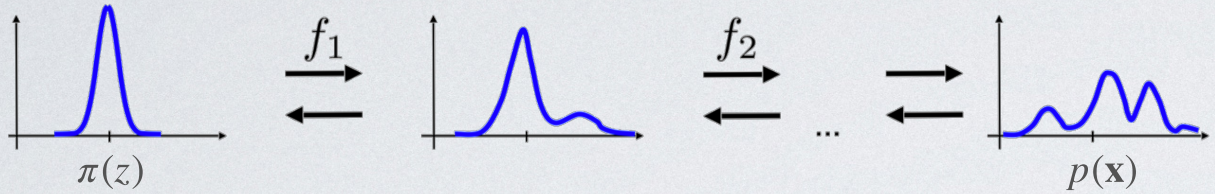
SURROGATE MODELS FOR SIMULATIONS



- Can we learn model $p_\theta(\mathbf{x})$ for the underlying data distribution $p(\mathbf{x})$ by directly training on expected (stochastic) outputs?

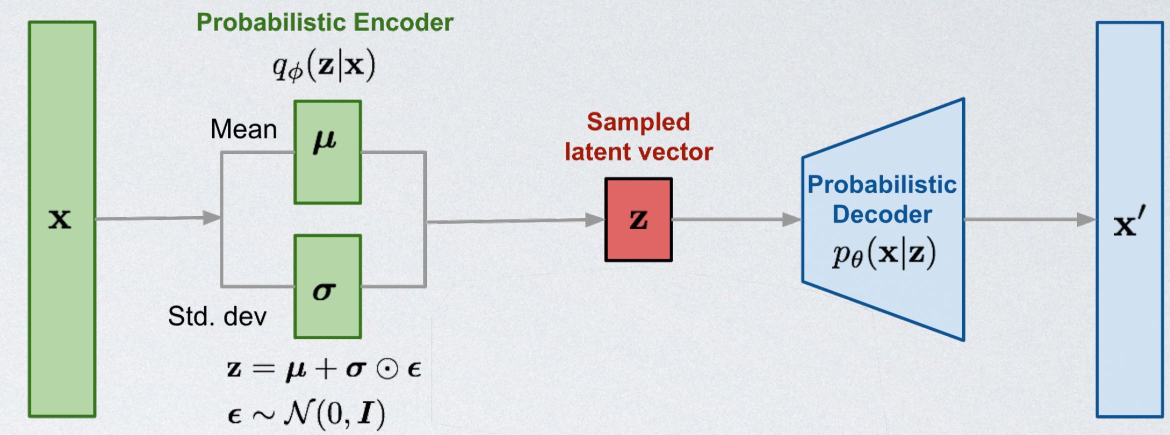
ML LANDSCAPE

Flows



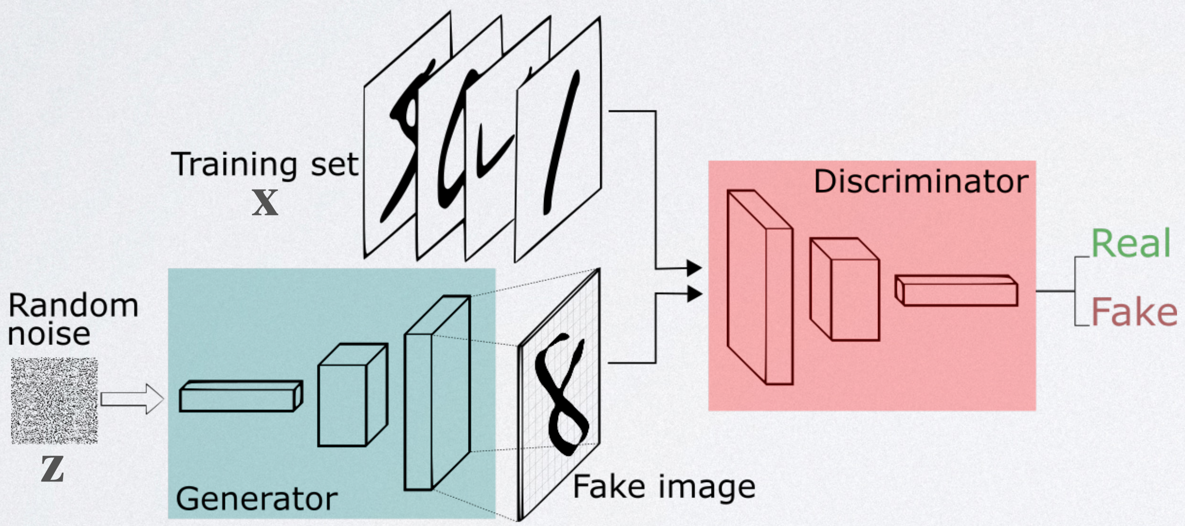
- Learn transformations from simpler $\pi(z)$ to $p(\mathbf{x})$
- Access to exact likelihood
- But restrictive

Variational Autoencoders (VAEs)



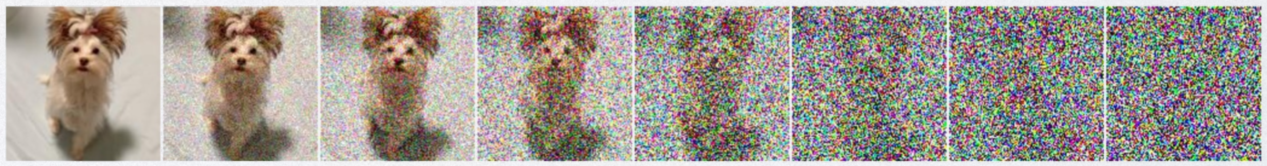
- Maximise an approximation to the likelihood
- Can also be restrictive

Generative Adversarial Networks (GANs)



- Minimise loss wrt to discriminator classifying real or fake
- Less restrictive, generally higher performing
- Difficult to train

(Score-based) Diffusion



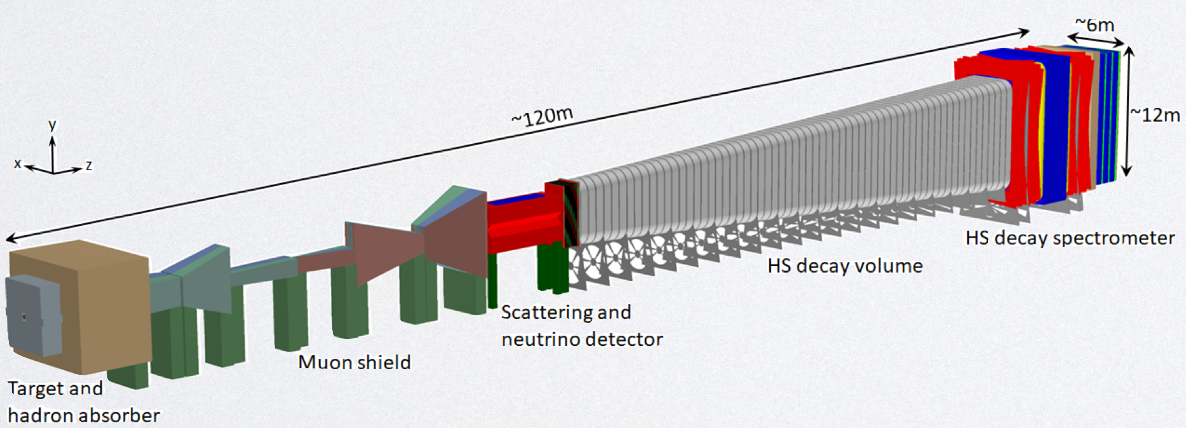
- Learn $-\partial \ln p(\mathbf{x}) / \partial \mathbf{x}$ (score) instead of $p(\mathbf{x})$ directly
- Less restrictive, score doesn't need to be normalised
- Current industry SOTA (DALL-E, StableDiffusion etc.)
- But slow - need O(100)s of steps along the score

OPTIMIZING THE SHIP MAGNET

SHiP Collaboration (2022)

A Baranov et al. (2017)

- "Search for Hidden Particles" (SHiP) experiment proposed at the CERN SPS

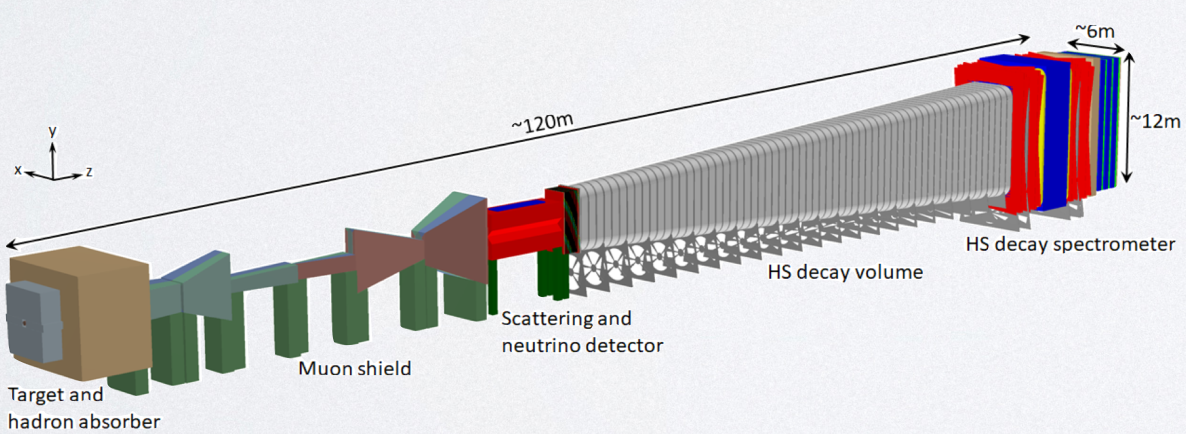


OPTIMIZING THE SHIP MAGNET

- Optimizing 42 parameters of SHiP's muon shield magnets

SHiP Collaboration (2022)
A Baranov et al. (2017)

- "Search for Hidden Particles" (SHiP) experiment proposed at the CERN SPS



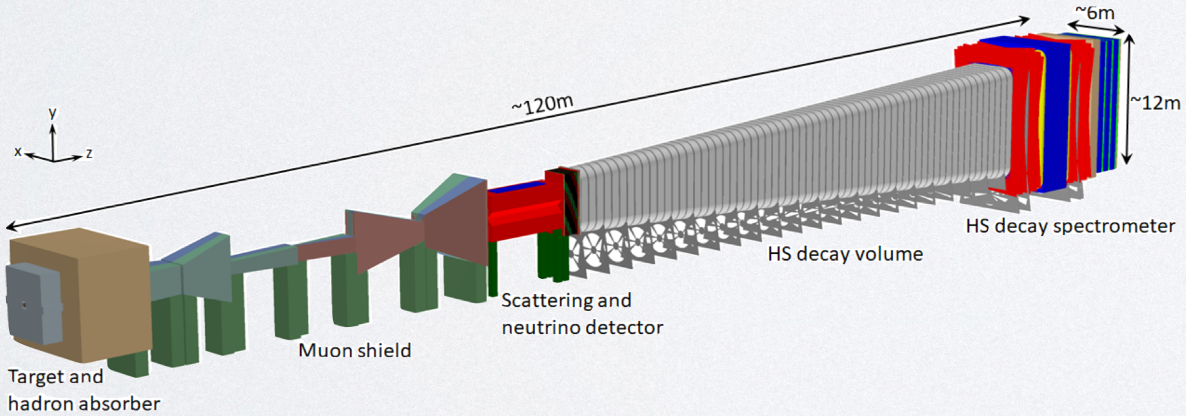
OPTIMIZING THE SHIP MAGNET

- Optimizing 42 parameters of SHiP's muon shield magnets
- GAN, flows and Gaussian process surrogate models used for simulations
  - Used to estimate gradients and optimize for muon shielding + cost

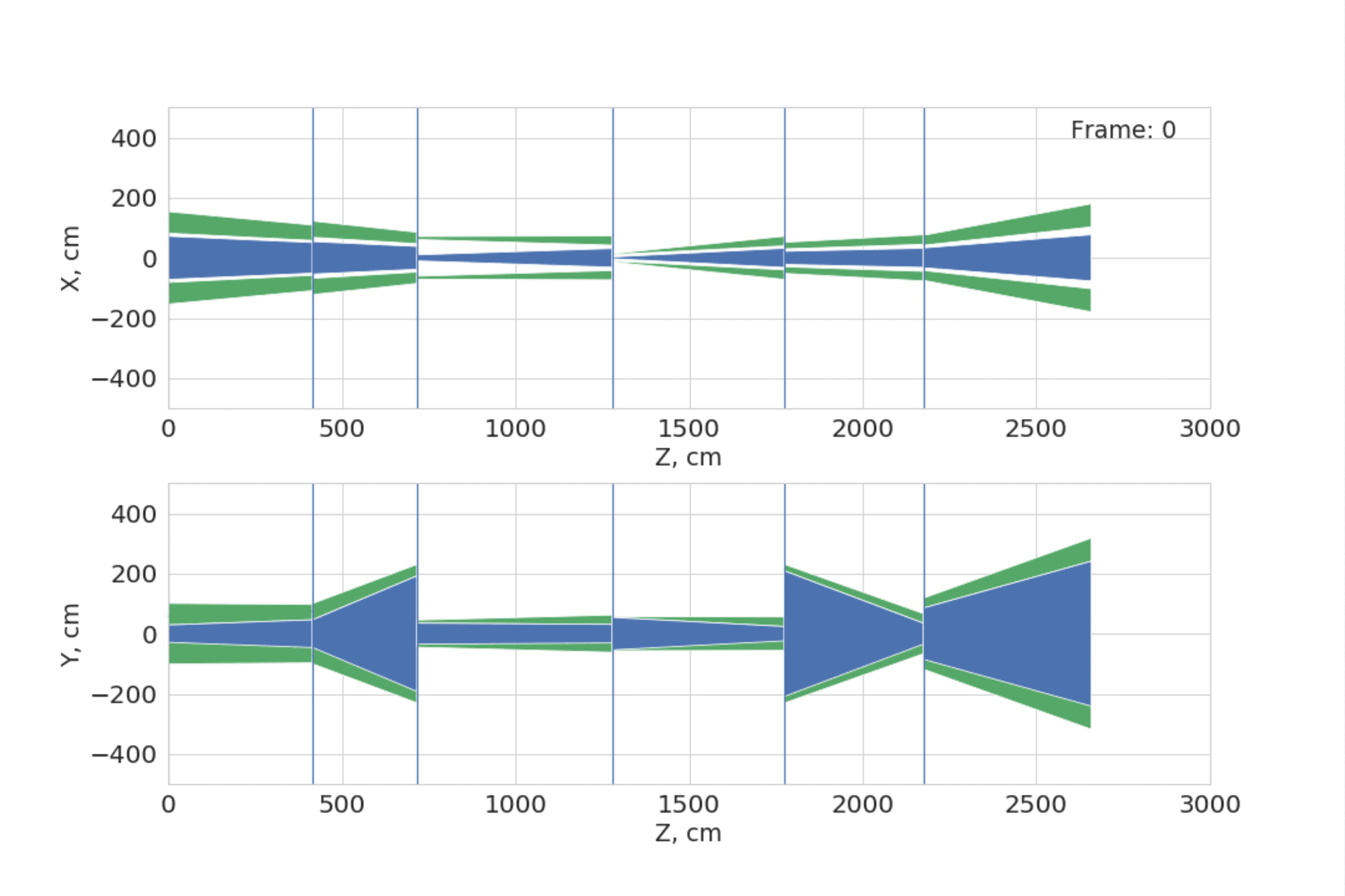
SHiP Collaboration (2022)

A Baranov et al. (2017)

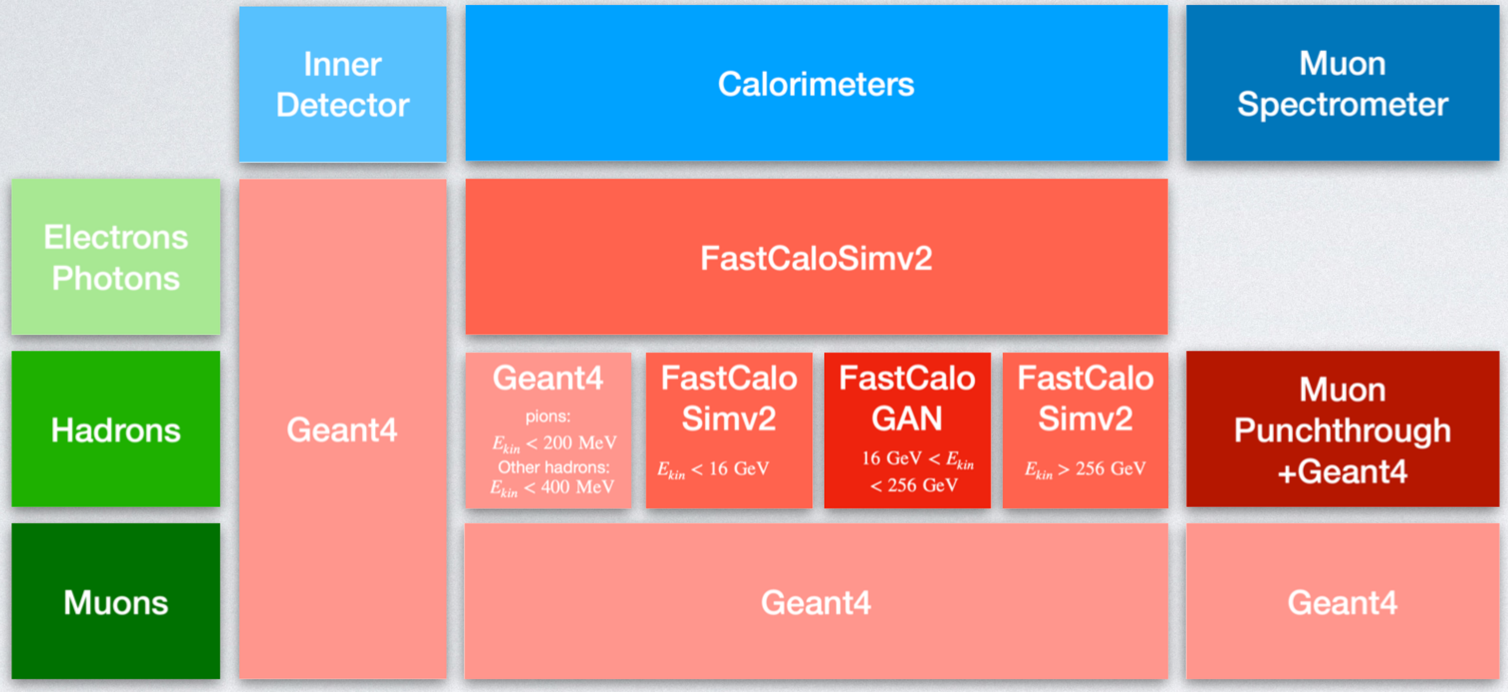
- "Search for Hidden Particles" (SHiP) experiment proposed at the CERN SPS

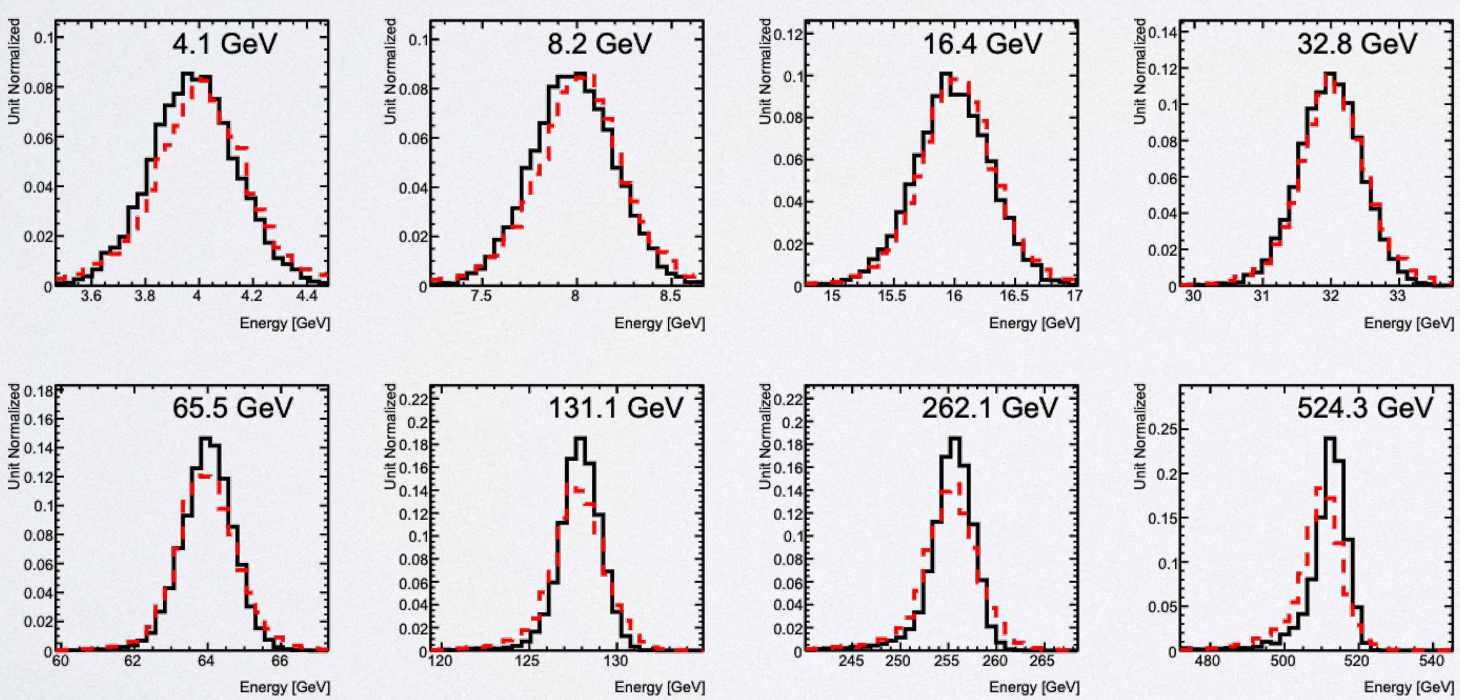


Shirobokov et al. (2020)

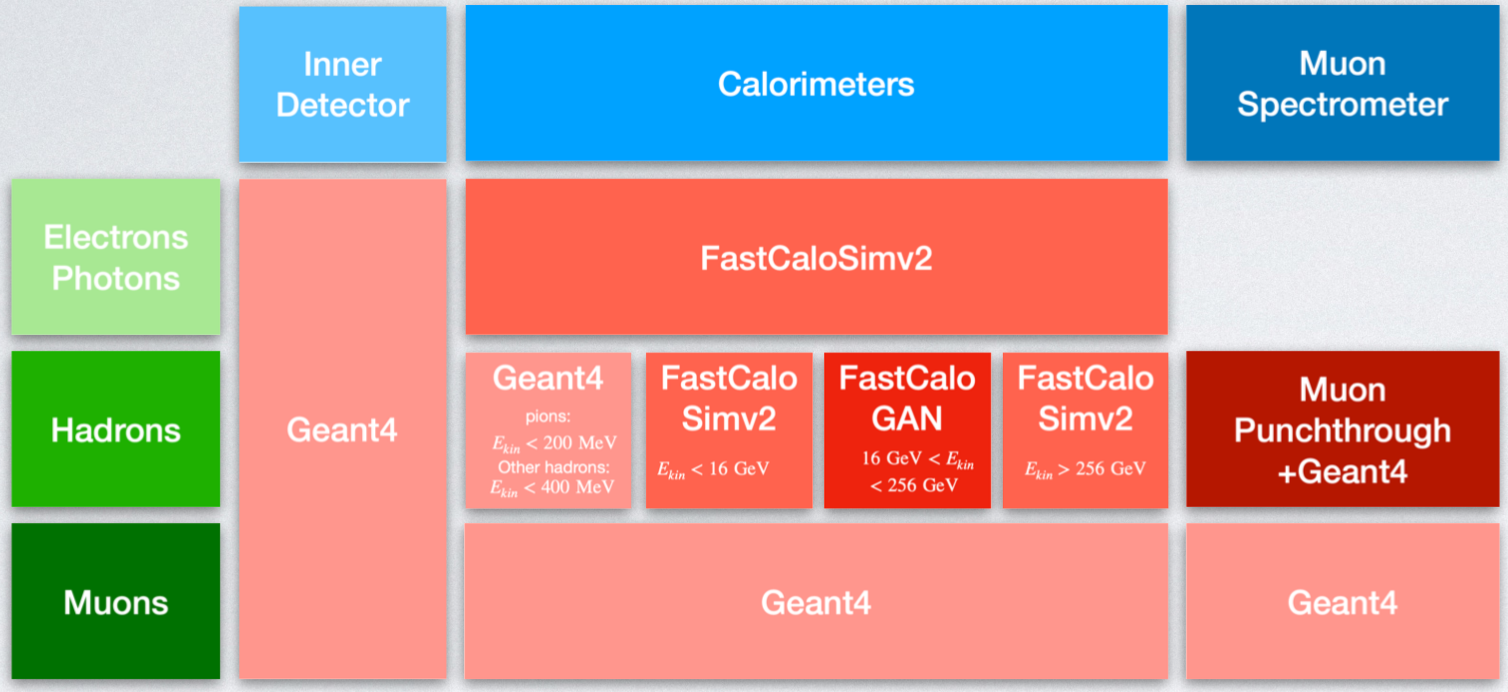


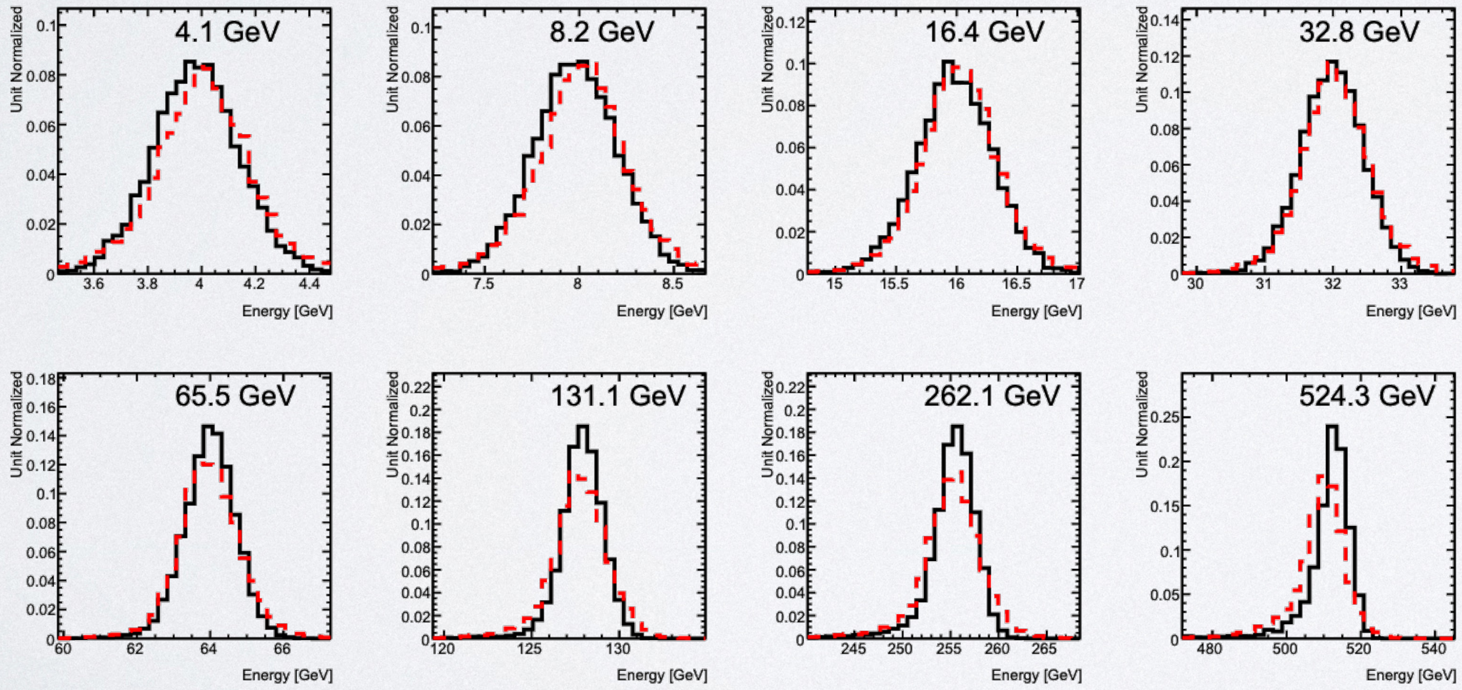
ATLAS FASTSIM





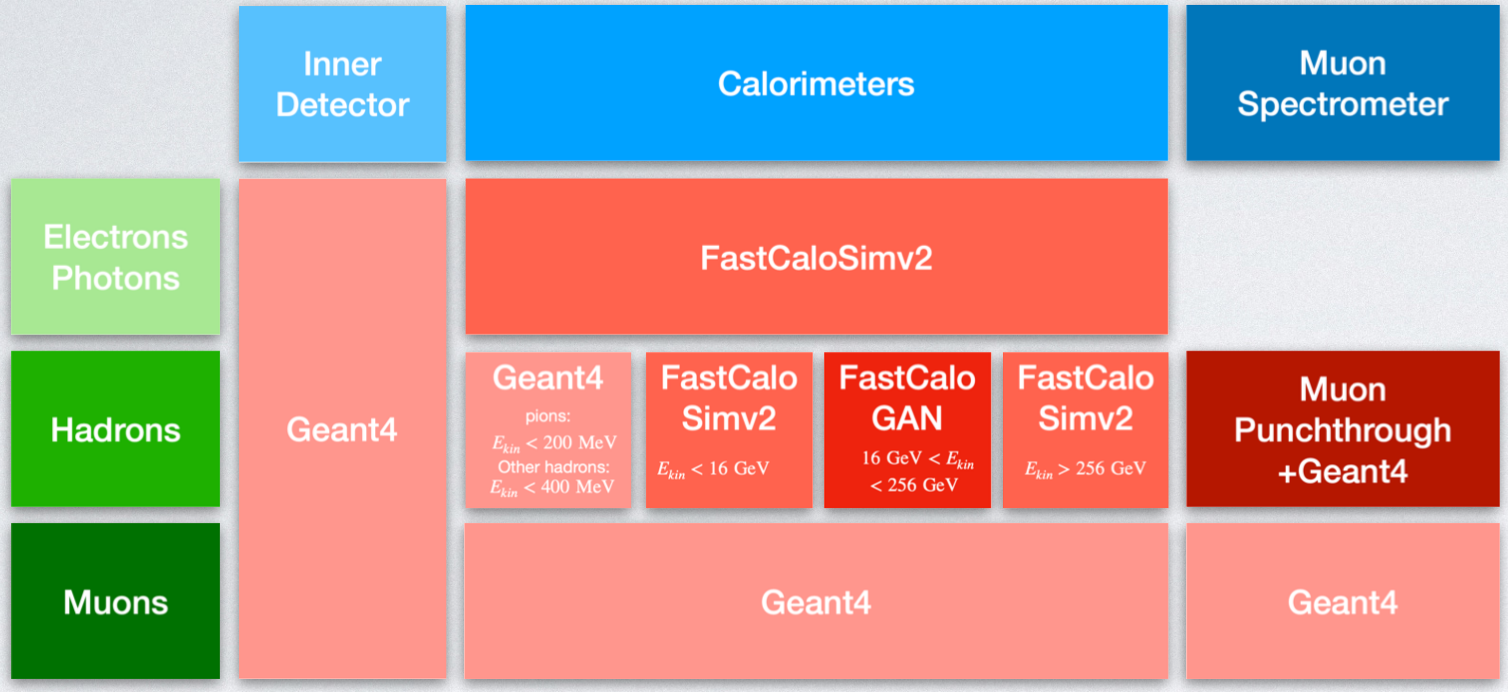
ATLAS FASTSIM



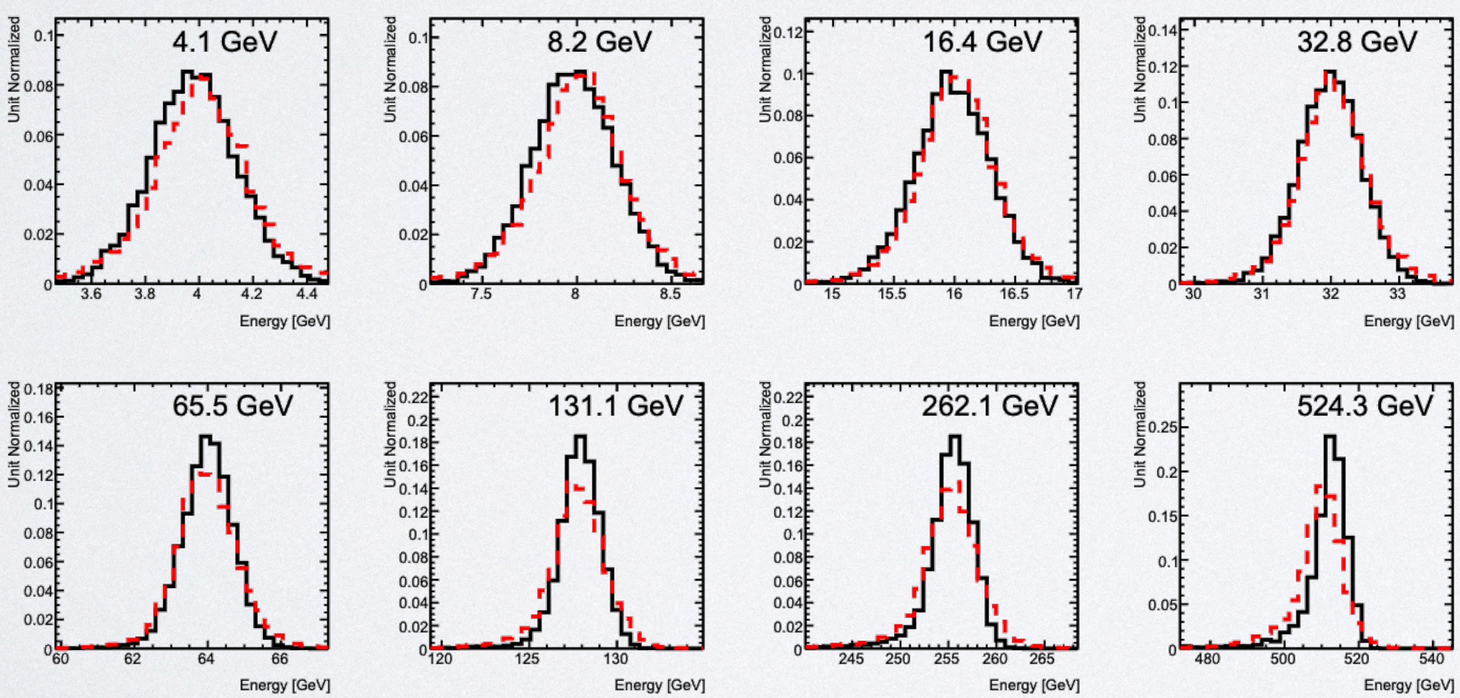


- GANs used already for fast sim
  - One component of "AtlFast3"
  - 7B events for Run 2 analyses!

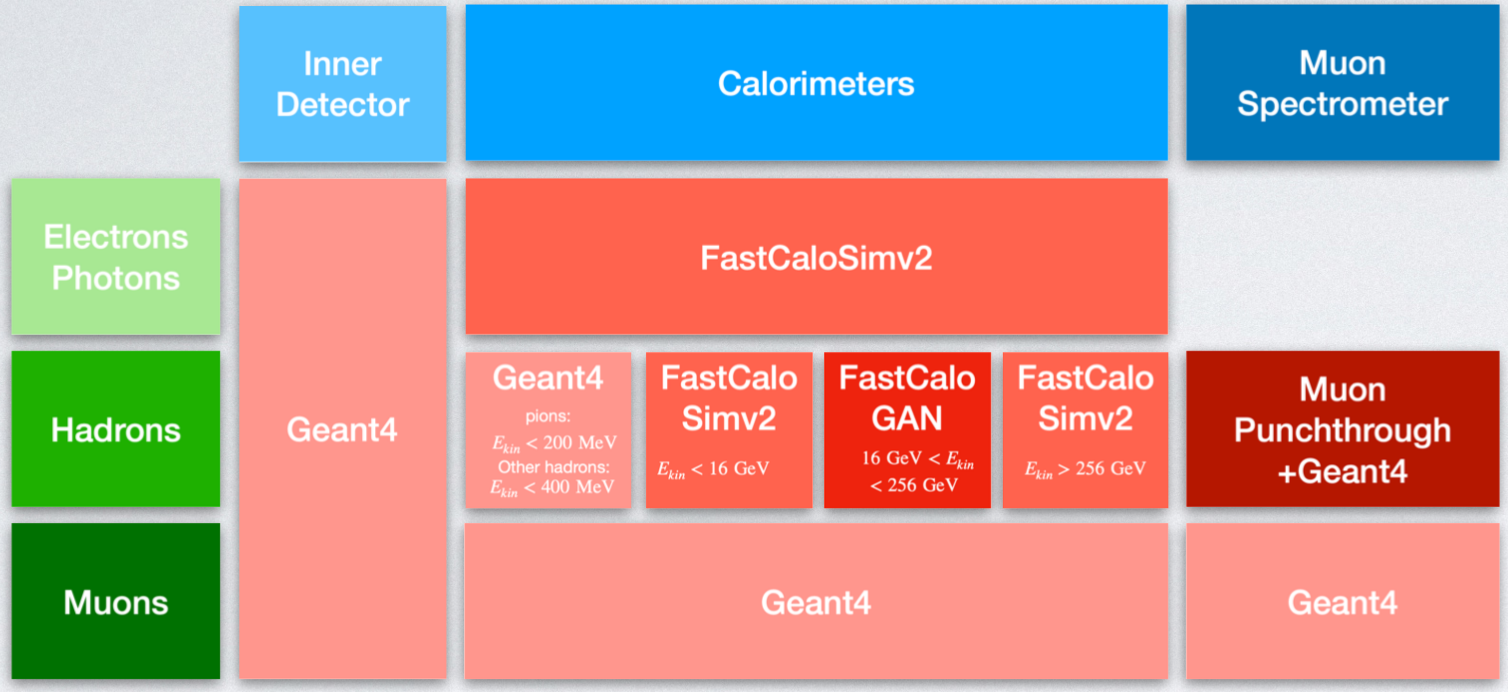
ATLAS FASTSIM

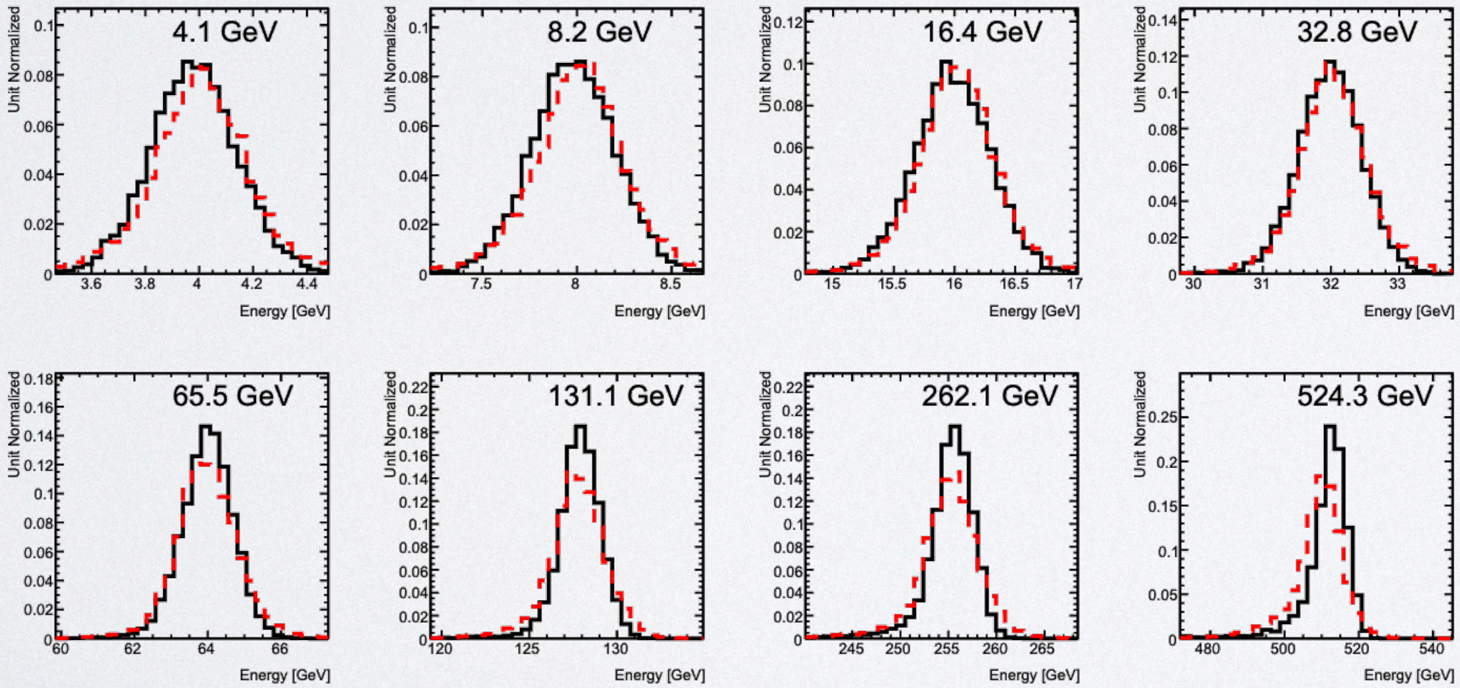


- GANs used already for fast sim
  - One component of "AtlFast3"
  - 7B events for Run 2 analyses!
- Trained on hadron shower images



ATLAS FASTSIM





- GANs used already for fast sim
  - One component of "AtlFast3"
  - 7B events for Run 2 analyses!
- Trained on hadron shower images
- Reasonable performance but:
  - Room for improvement
  - "Voxelisation" to deal with sparsity and high granularity
  - 300 GANs trained for each E, η bin

CALORIMETER SHOWERS

- Idea: learn distribution of hits per gen particle i.e. surrogate for GEANT

CALORIMETER SHOWERS

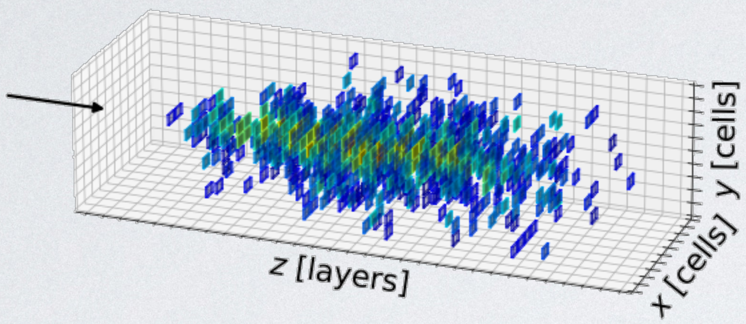
- Idea: learn distribution of hits per gen particle i.e. surrogate for GEANT

*A100 GPU

CALORIMETER SHOWERS

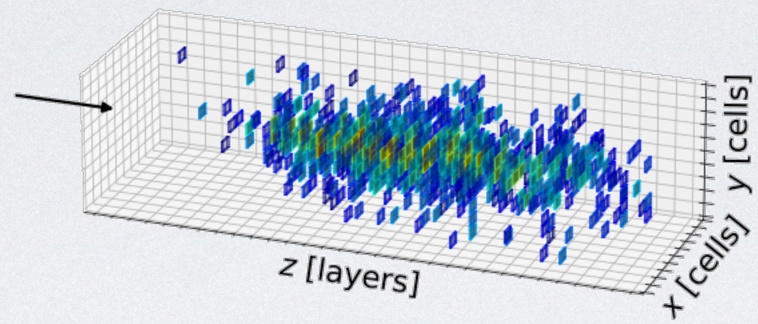
- Idea: learn distribution of hits per gen particle i.e. surrogate for GEANT

BIB-AE (2112.09709)



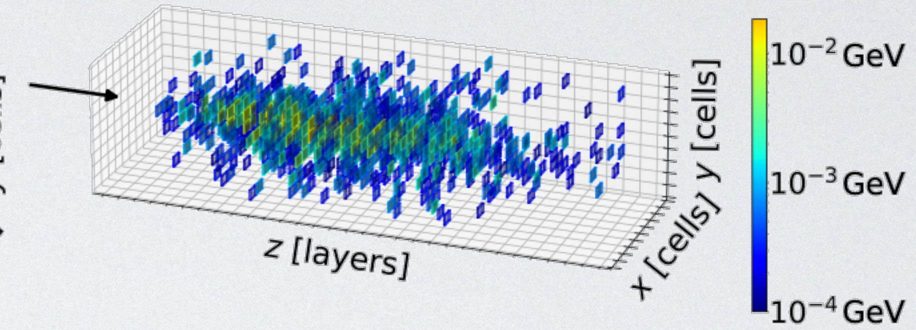
- VAE + GAN + postprocessing
- 0.1 μs/shower*

Sample GEANT shower



- Reproducing shower images
- Good agreement with simplified ILD-like single γ showers

L2LFlows (2302.11594)



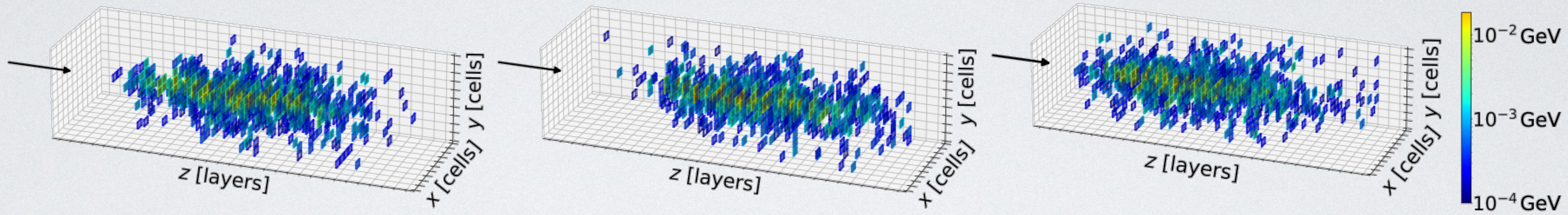
- Series of normalising flows
- 7μs/shower*
- Compared to $O(10s)$/shower with GEANT

*A100 GPU

CALORIMETER SHOWERS

- Idea: learn distribution of hits per gen particle i.e. surrogate for GEANT

BIB-AE (2112.09709) Sample GEANT shower **L2LFlows** (2302.11594)

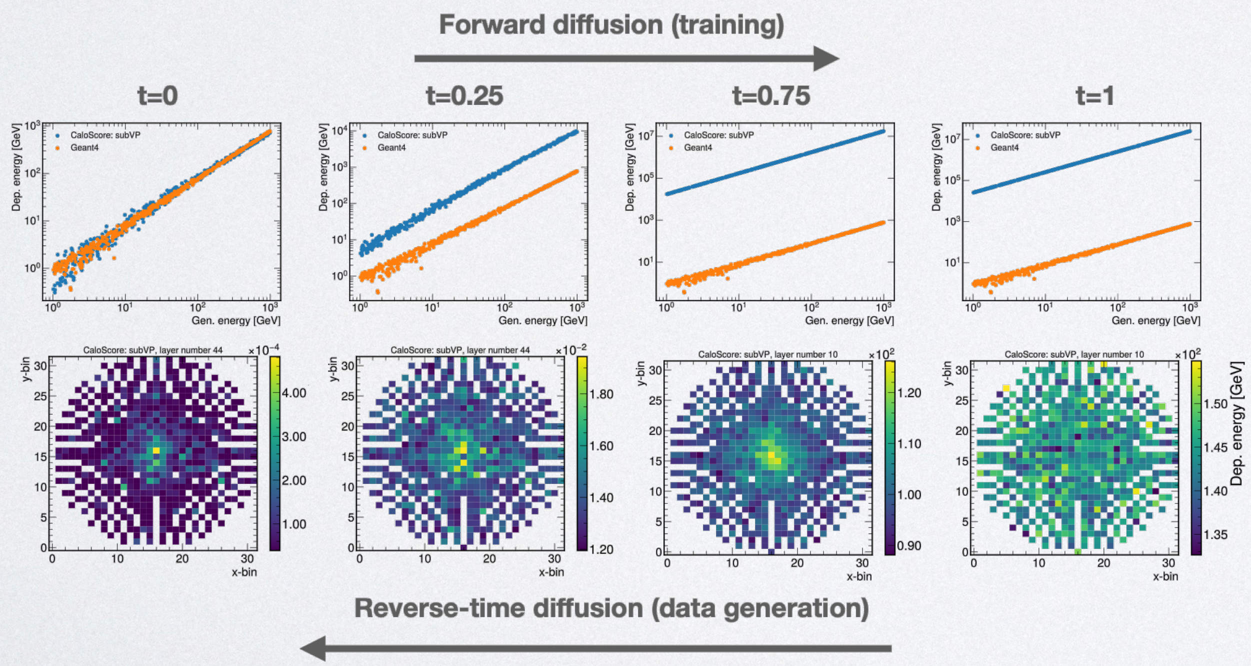


- VAE + GAN + postprocessing
- 0.1 μs/shower*

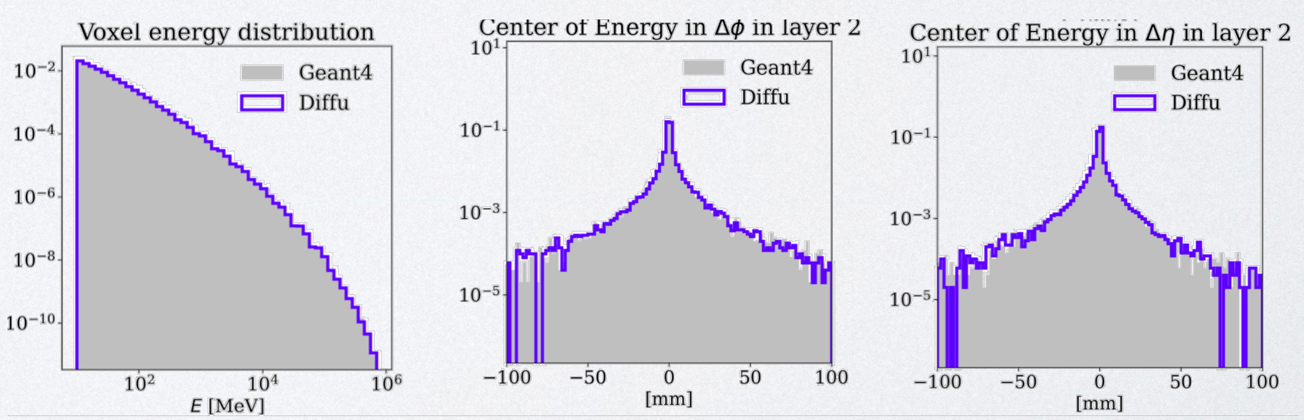
- Reproducing shower images
- Good agreement with simplified ILD-like single γ showers

- Series of normalising flows
- 7μs/shower*
- Compared to $O(10s)$/shower with GEANT

CaloScore (2206.11898) **CaloDiffusion** (CHEP 2023)



- Diffusion based models
- Good agreement with different datasets
- Generally slower than GANs
  - 0.3s/shower*



JETS

- Idea: learn distribution of reco-particles per gen parton i.e., surrogate for Pythia + GEANT + reco

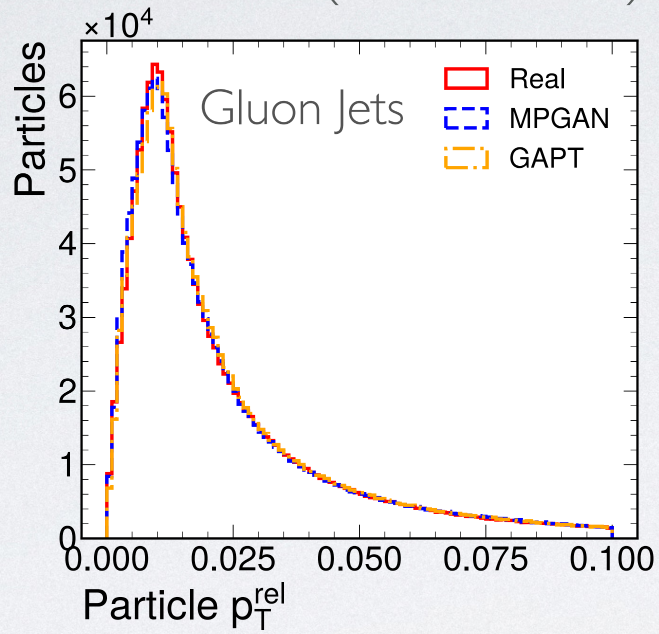
*Single GPU

JETS

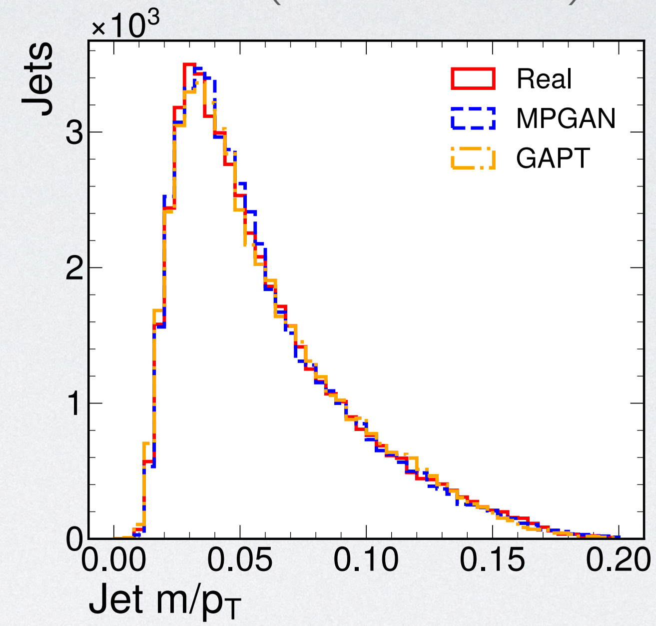
- Idea: learn distribution of reco-particles per gen parton i.e., surrogate for Pythia + GEANT + reco

- GANs
- Point-cloud based
  - Generally better suited to HEP data
- Graph neural networks, transformers, and deep sets
- Good agreement with simplified CMS-like jets
- All $\leq \mathcal{O}(10\mu s)$/jet*

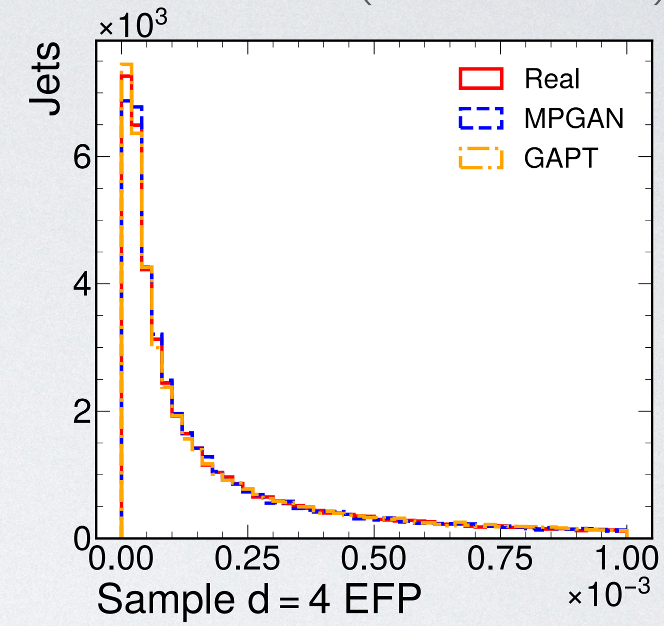
MPGAN (2106.11535)



GAPT (2211.10295)



EPiC-GAN (2301.08128)

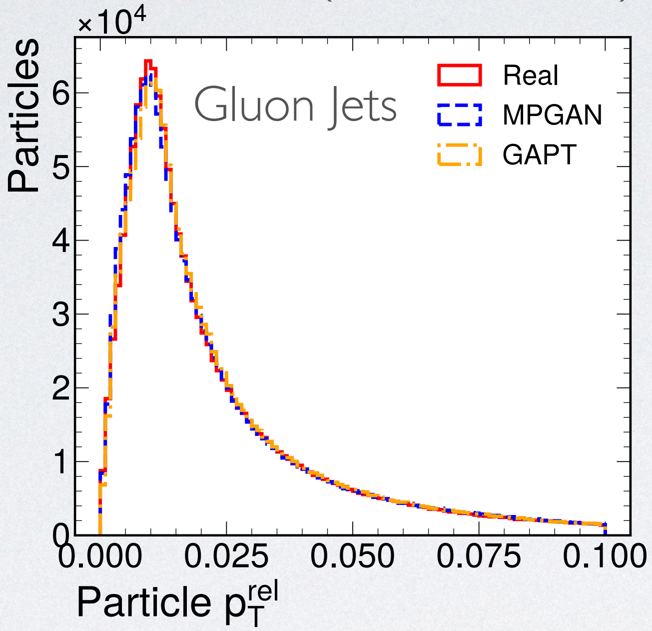


JETS

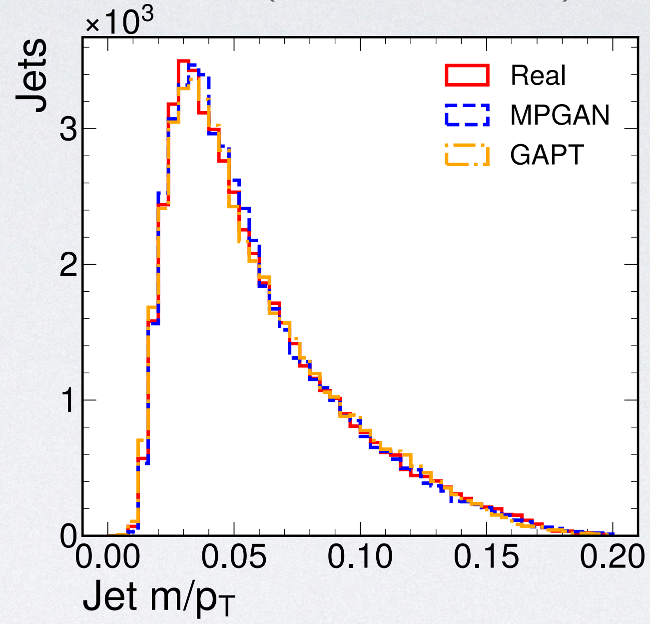
- Idea: learn distribution of reco-particles per gen parton i.e., surrogate for Pythia + GEANT + reco

- GANs
- Point-cloud based
  - Generally better suited to HEP data
- Graph neural networks, transformers, and deep sets
- Good agreement with simplified CMS-like jets
- All $\leq \mathcal{O}(10\mu s)$/jet*

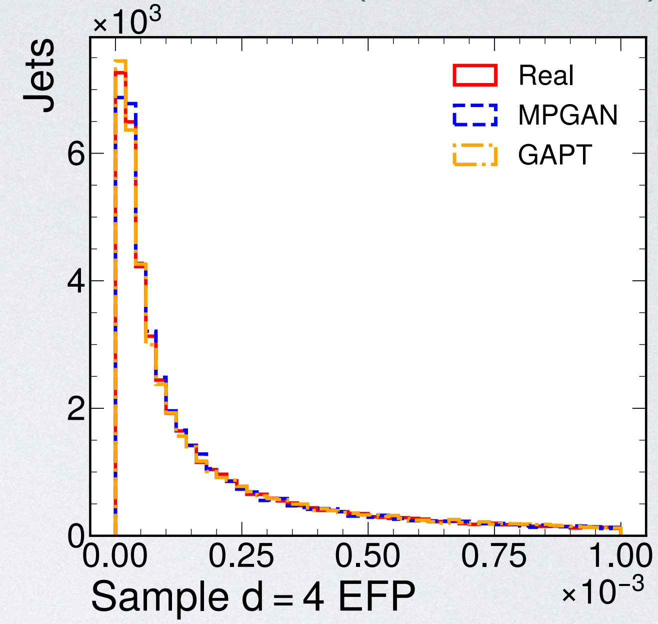
MPGAN (2106.11535)

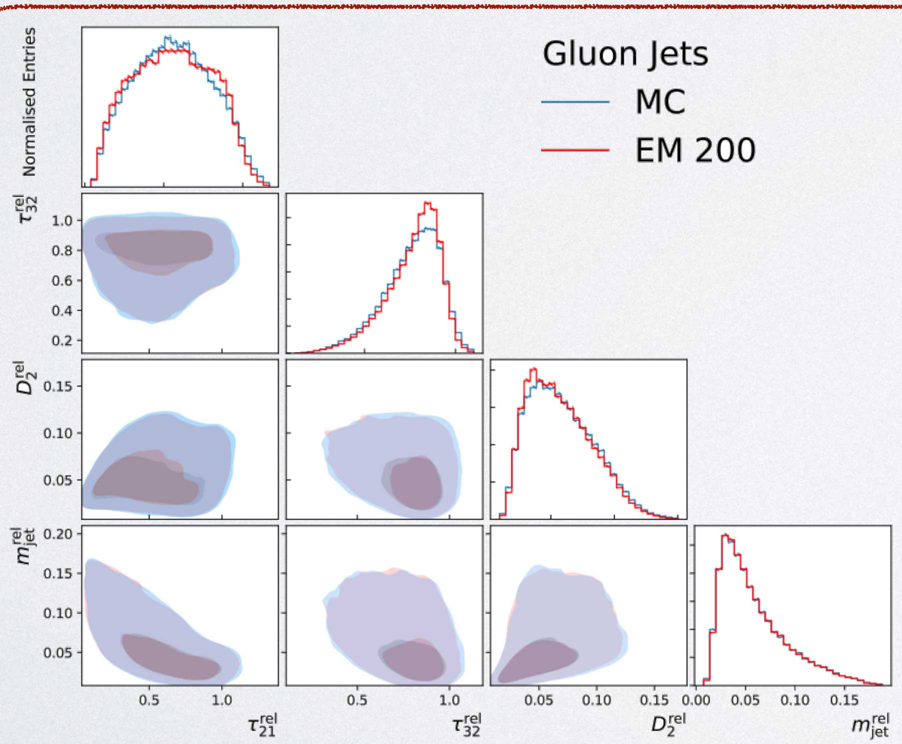


GAPT (2211.10295)



EPiC-GAN (2301.08128)





PC-JeDi (2303.05376)

FPCD (2304.01266)

- Diffusion models
- Also point-cloud based
- Good agreement
- Slower than GANs
  - $\mathcal{O}(1ms)$/jet*
  - But distillation brings this down to $\mathcal{O}(10\mu s)$/jet

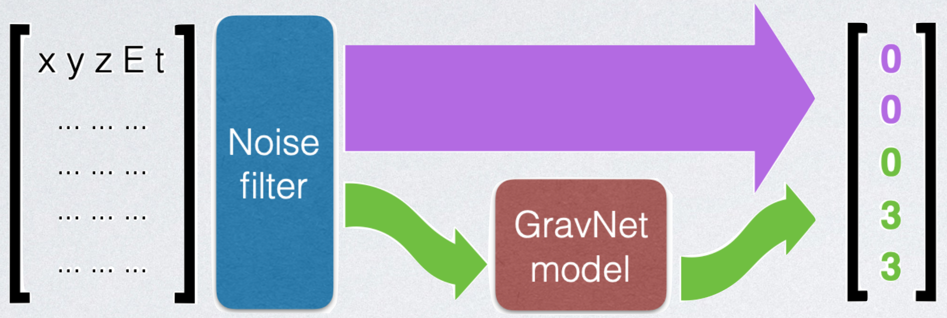
ML FOR RECONSTRUCTION

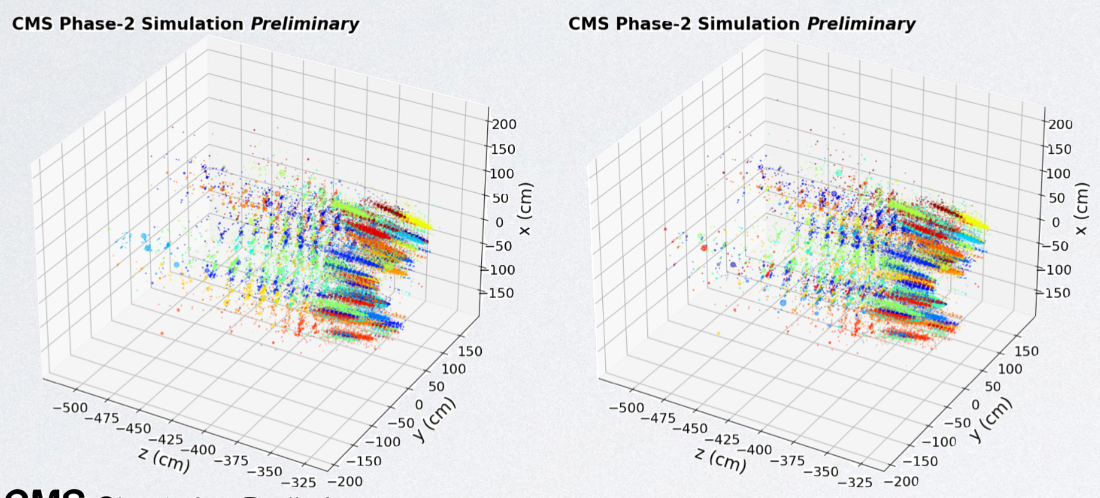
ML FOR RECONSTRUCTION

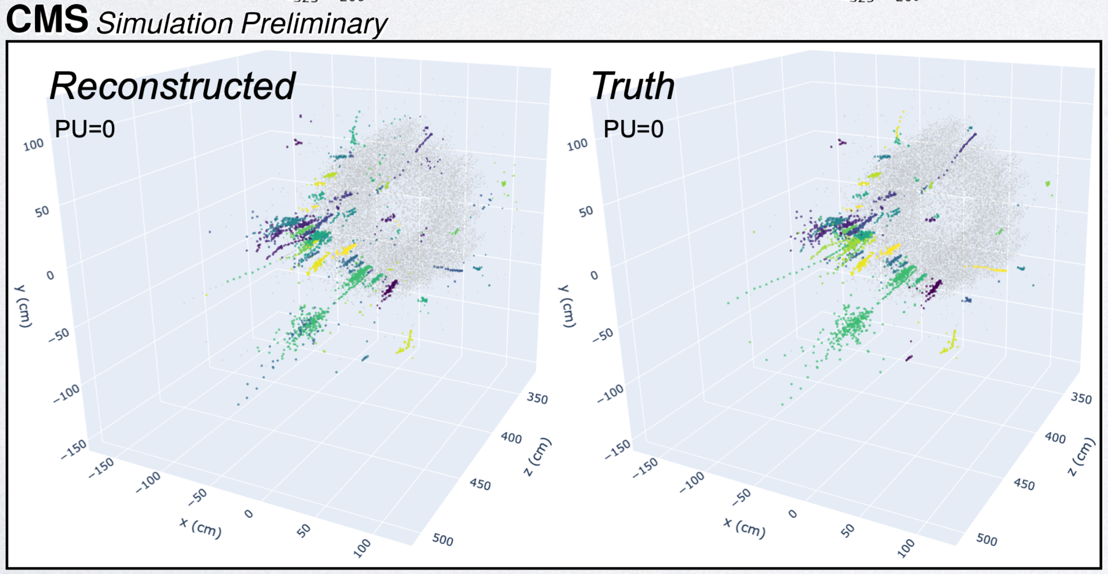
ML for CMS HGCAL

S. Bhattacharya et al. (2023), SR Qasim et al., (2021)

- Hits → tracks + clusters with GNNs





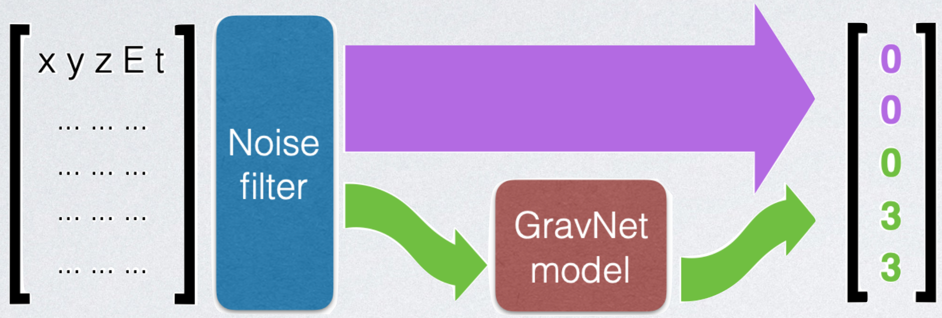


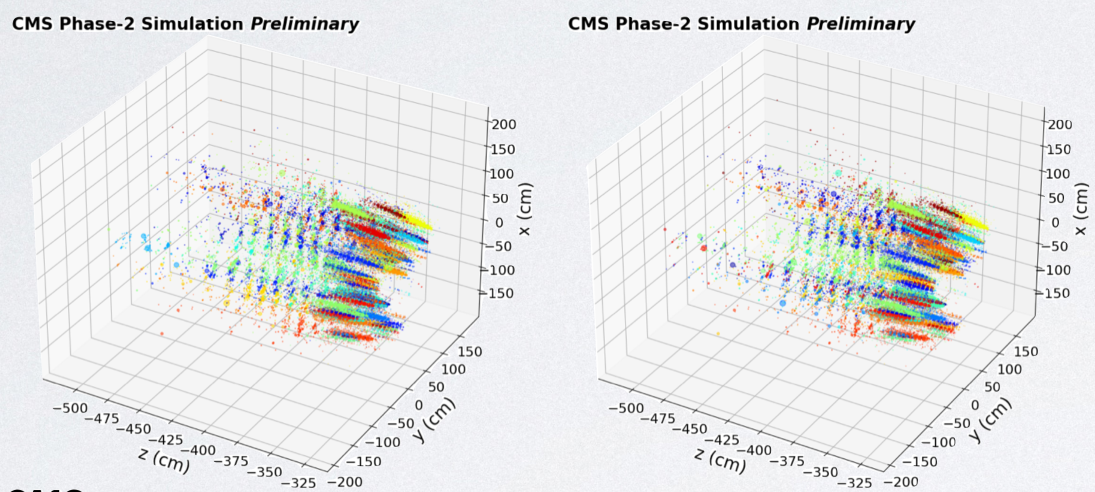
ML FOR RECONSTRUCTION

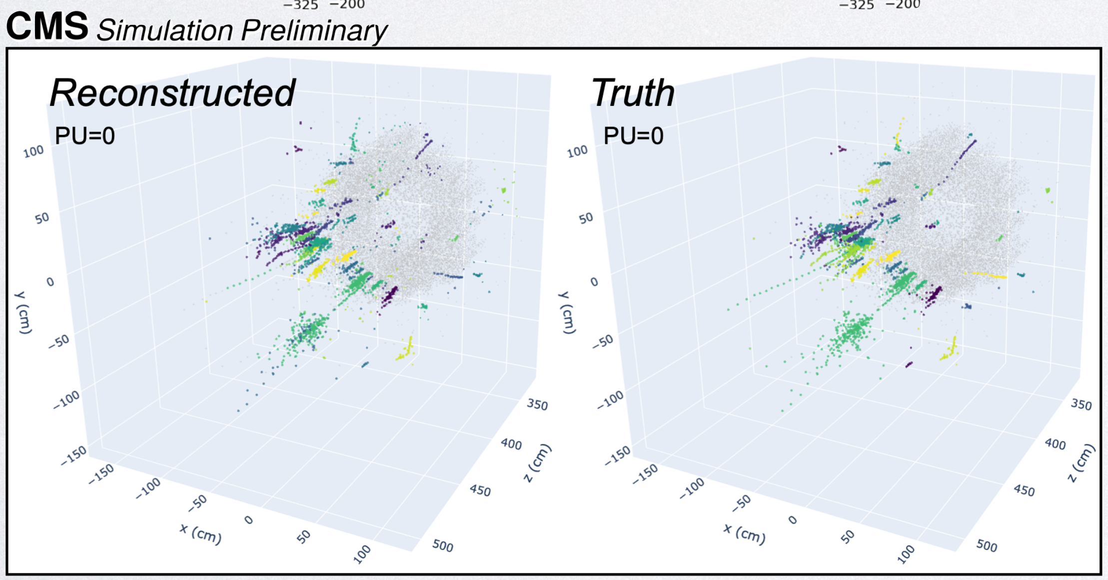
ML for CMS HGCAL

S. Bhattacharya et al. (2023), SR Qasim et al., (2021)

- Hits → tracks + clusters with GNNs



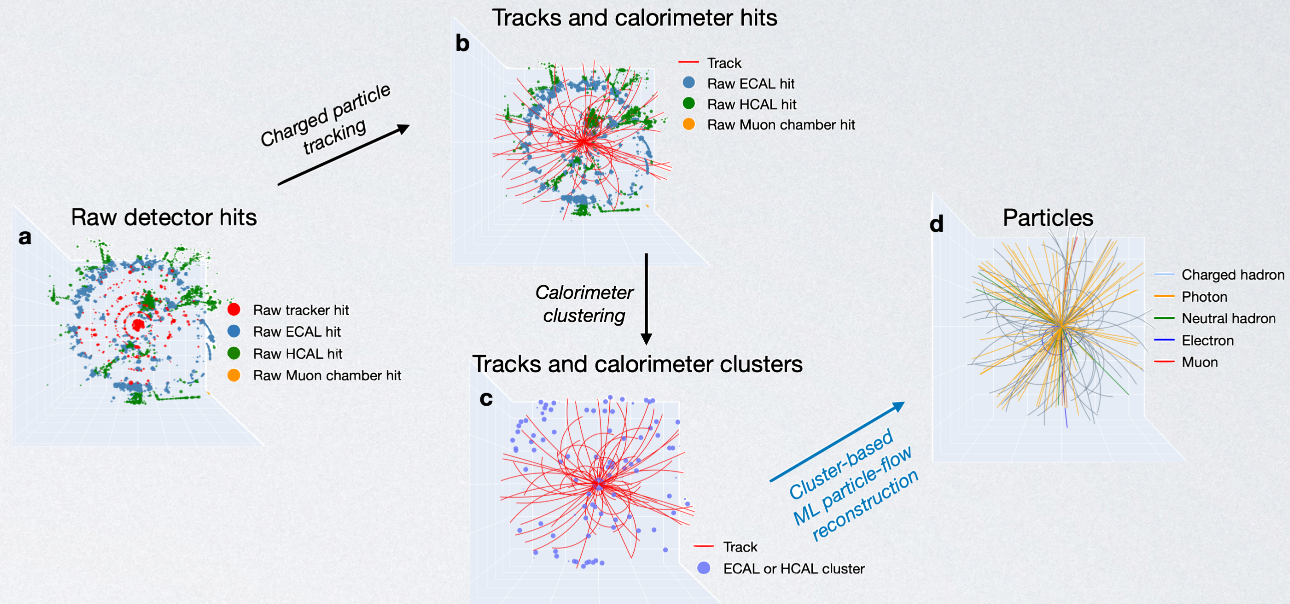




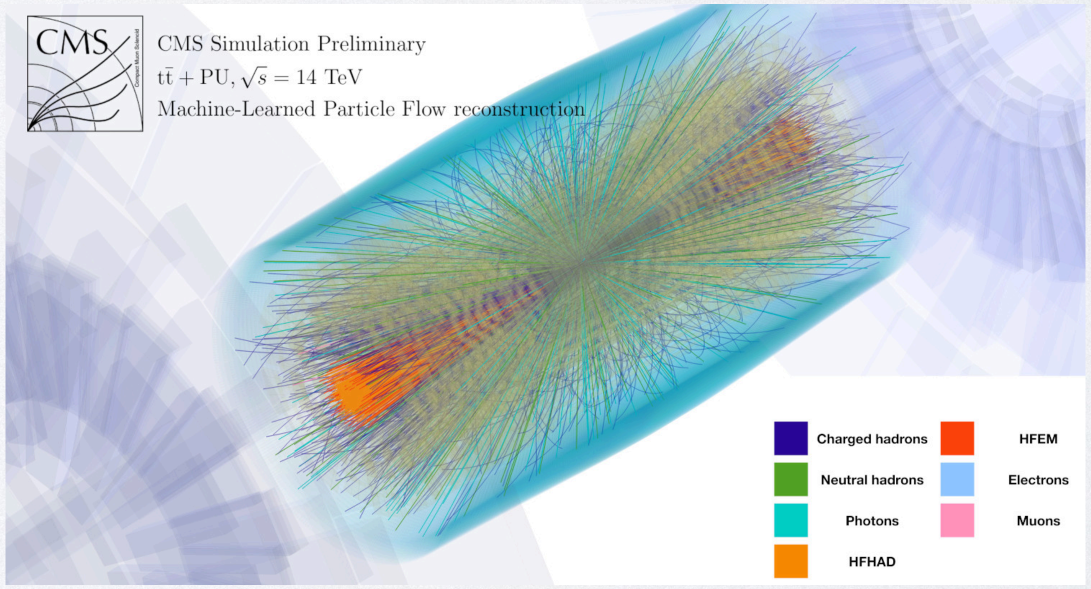
ML for ParticleFlow

Pata et al., 2021-24

- Learning PF (tracks + clusters → particles) with GNNs



- Strong results for CMS and CLIC detectors



NEXT STEPS FOR FCC: HOW DO WE CONVERGE?

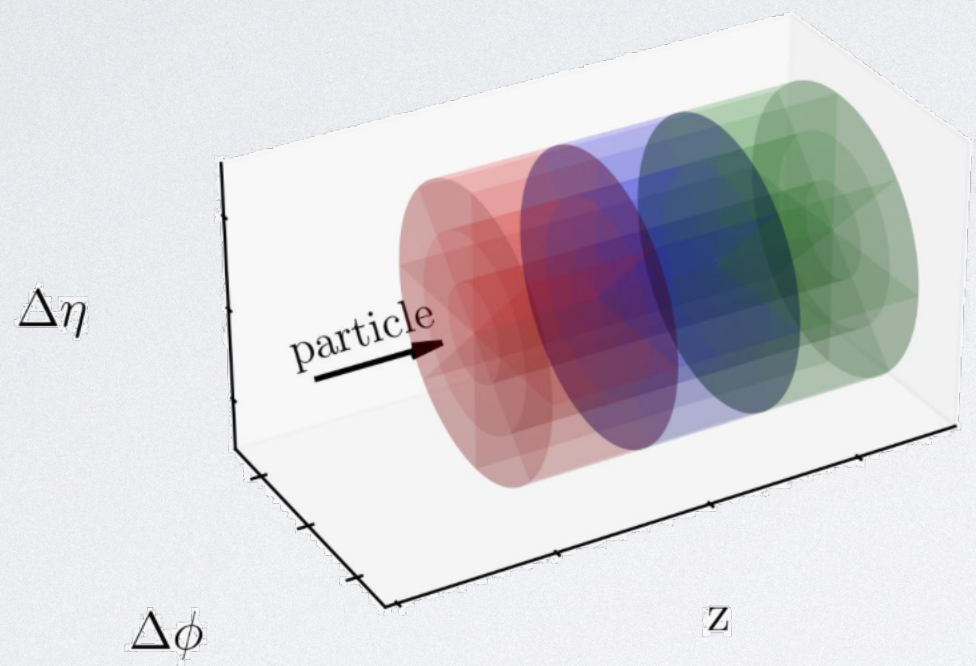
DATASETS

- Need datasets to train / benchmark models

DATASETS

- Need datasets to train / benchmark models

CaloChallenge 2022-23

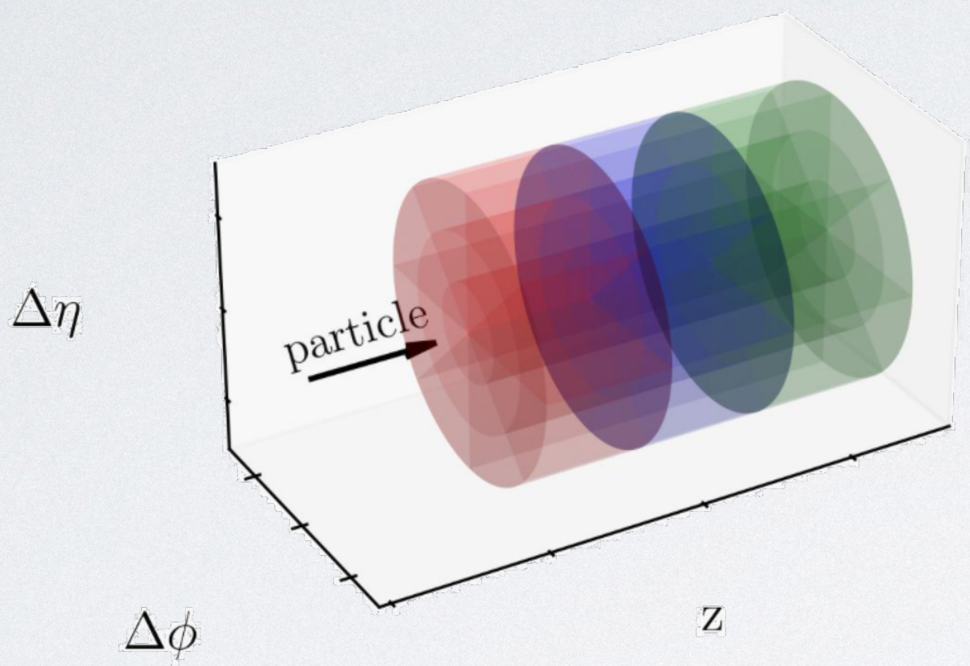


- Public "challenge" for calorimeter simulations
- 3 image-based datasets based on ATLAS-like and general detectors

DATASETS

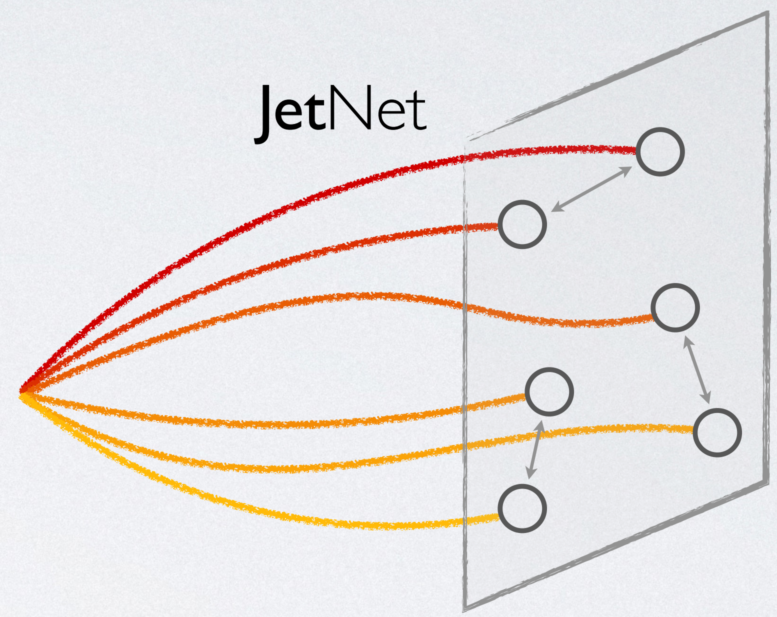
- Need datasets to train / benchmark models

CaloChallenge 2022-23



- Public "challenge" for calorimeter simulations
- 3 image-based datasets based on ATLAS-like and general detectors

JetNet

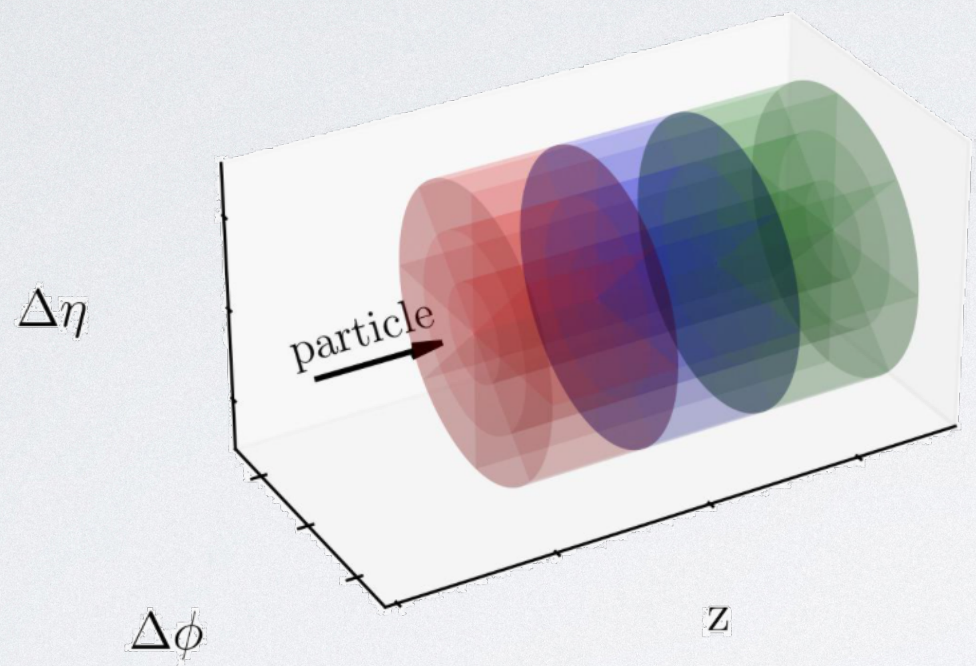


- Public library and (collection of) jet datasets
- All point-cloud based, simplified reco
- Basis for majority of recent work on jets

DATASETS

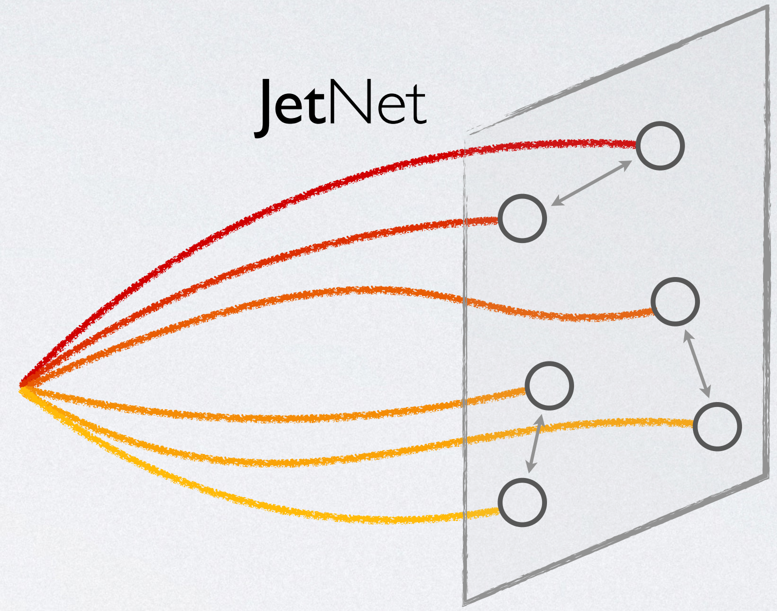
- Need datasets to train / benchmark models

CaloChallenge 2022-23



- Public "challenge" for calorimeter simulations
- 3 image-based datasets based on ATLAS-like and general detectors

JetNet



- Public library and (collection of) jet datasets
- All point-cloud based, simplified reco
- Basis for majority of recent work on jets

- Next step: FCC datasets! FCC Challenge 2025/26?

EVALUATION METRICS

EVALUATION METRICS

- Need standard, quantitative metrics to compare, validate, and trust models

EVALUATION METRICS

- Need standard, quantitative metrics to compare, validate, and trust models
- Studied in detail in 2211.10295 in terms of two-sample GoF tests
  - Traditional method is looking at 1 or 2D histograms
    - Should be quantified, can miss correlations
  - Many multivariate GoF tests studied
  - Fréchet and kernel physics distances found to be most sensitive

EVALUATION METRICS

- Need standard, quantitative metrics to compare, validate, and trust models
- Studied in detail in 2211.10295 in terms of two-sample GoF tests
  - Traditional method is looking at 1 or 2D histograms
    - Should be quantified, can miss correlations
  - Many multivariate GoF tests studied
  - Fréchet and kernel physics distances found to be most sensitive
- Starting to be adopted for jets

	FPD $\times 10^3$	KPD $\times 10^3$	W_1^M $\times 10^3$	$W_{1p}^{p_T^{\mathrm{rel}}}$ $\times 10^3$
Truth	0.08 ± 0.03	-0.006 ± 0.005	0.28 ± 0.05	0.44 ± 0.09
MPGAN	$\mathbf{0.30 \pm 0.06}$	$\mathbf{-0.001 \pm 0.004}$	$\mathbf{0.54 \pm 0.06}$	0.6 ± 0.2
GAPT	0.66 ± 0.09	0.001 ± 0.005	0.56 ± 0.08	$\mathbf{0.51 \pm 0.09}$

EVALUATION METRICS

- Need standard, quantitative metrics to compare, validate, and trust models
- Studied in detail in 2211.10295 in terms of two-sample GoF tests
  - Traditional method is looking at 1 or 2D histograms
    - Should be quantified, can miss correlations
  - Many multivariate GoF tests studied
  - Fréchet and kernel physics distances found to be most sensitive
- Starting to be adopted for jets

	FPD $\times 10^3$	KPD $\times 10^3$	W_1^M $\times 10^3$	$W_{1p}^{p_T^{\mathrm{rel}}}$ $\times 10^3$
Truth	0.08 ± 0.03	-0.006 ± 0.005	0.28 ± 0.05	0.44 ± 0.09
MPGAN	$\mathbf{0.30 \pm 0.06}$	$\mathbf{-0.001 \pm 0.004}$	$\mathbf{0.54 \pm 0.06}$	0.6 ± 0.2
GAPT	0.66 ± 0.09	0.001 ± 0.005	0.56 ± 0.08	$\mathbf{0.51 \pm 0.09}$

- Need to establish recommendations
  - See talks in PHYSTAT

CONCLUSION

- ML-based simulations offer speed and differentiability
- Necessary for optimal detector design
- Many examples and new R&D approaches now in HEP
- Outlook for FCC studies:
  - Need R&D on auto-differentiation and surrogate model accuracy
  - Need to establish datasets
  - Need to validate rigorously
  - Integration with Key4hep