SQuIRELS

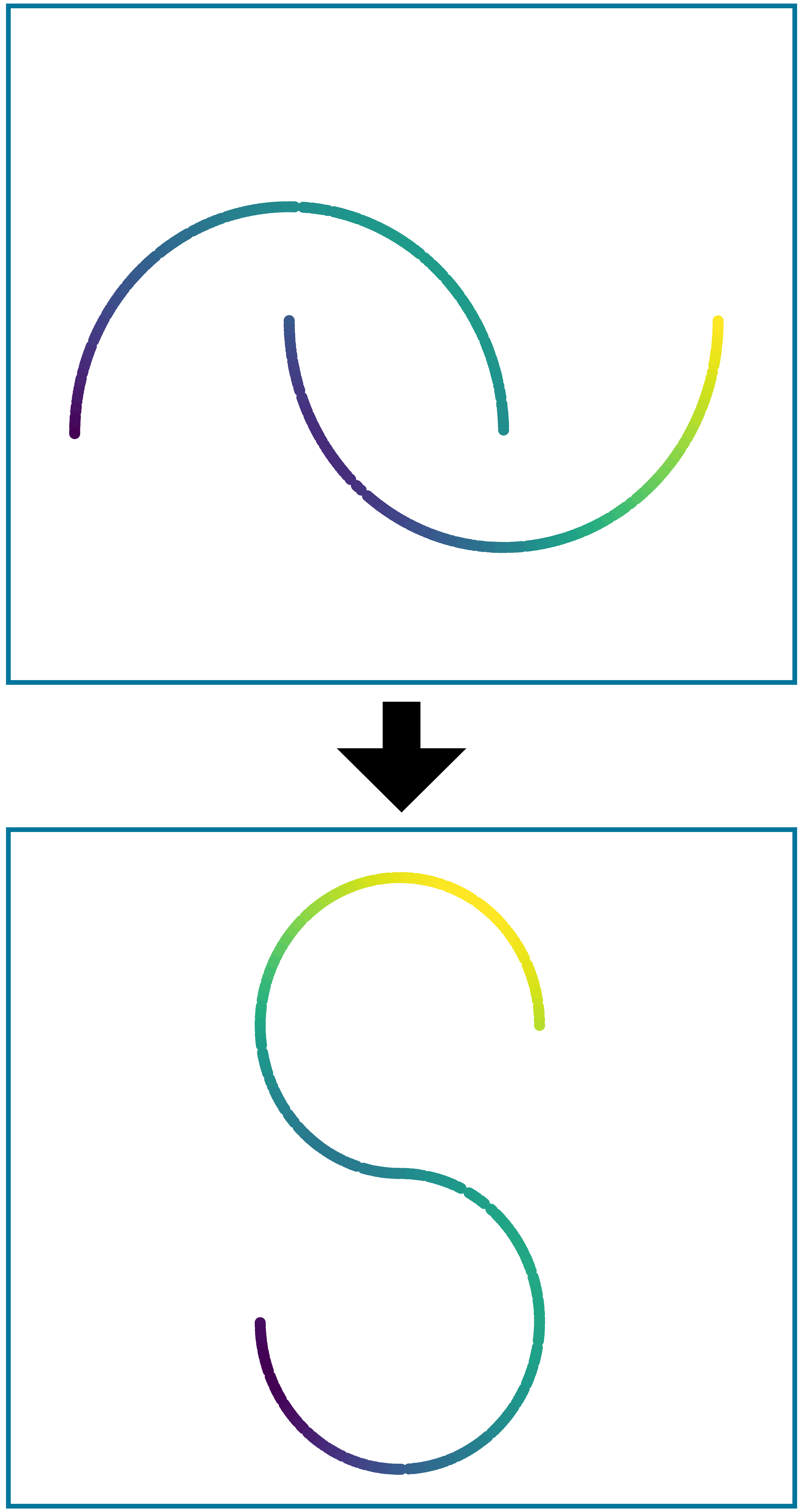
Sascha Diefenbacher, Vinicius Mikuni, Benjamin Nachman

ML4jets, Hamburg 2023

BERKELEY LAB

NERSC

ATLAS EXPERIMENT



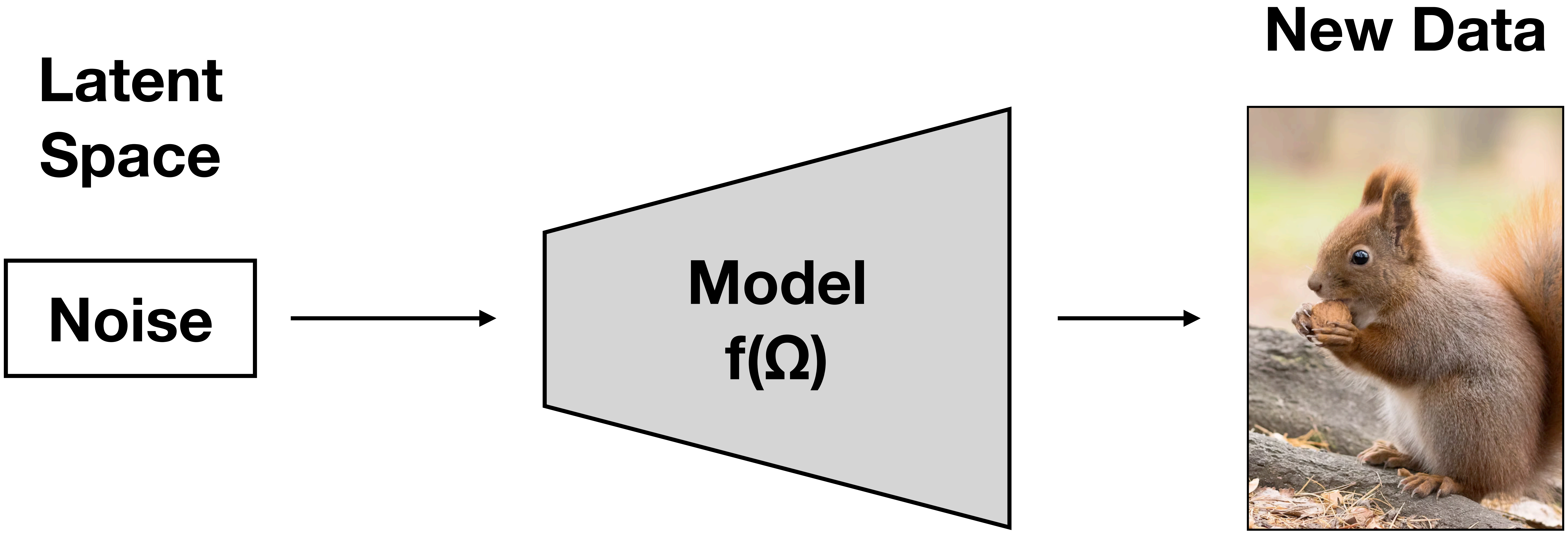
Refinement

- Fundamentally: Refinement = mapping arbitrary distribution to other arbitrary distribution
  - Simulation refinement
  - Unfolding
  - Background estimation
  - Anomaly detection
  - Many more…

Generative Models

First step:

- Standard generative models:
- Map random noise to new data



Generative Models

Now: generalize to arbitrary starting distribution

- VAE:

Generative Models

Now: generalize to arbitrary starting distribution

- VAE: ✗
  - Latent regularization needs latent space with tractable KLD
  - Starting distribution cannot be arbitrary
  - Needs tractable base distribution

Generative Models

Now: generalize to arbitrary starting distribution

- VAE: ✗ Needs tractable base distribution
- NF:

Generative Models

Now: generalize to arbitrary starting distribution

- VAE: ✗ Needs tractable base distribution
- NF: ✗
  - NLL calculation needs tractable base
  - Two-Flow trick possible
  - Requires learning full generative model

Generative Models

Now: generalize to arbitrary starting distribution

- VAE: ✗ Needs tractable base distribution
- NF: ✗ NLL calculation needs tractable base
- Diffusion:

Generative Models

Now: generalize to arbitrary starting distribution

- VAE: ✗ Needs tractable base distribution
- NF: ✗ NLL calculation needs tractable base
- Diffusion: ✗
  - Diffusion process leads to Gaussian
  - Needs Gaussian base distribution

Generative Models

Now: generalize to arbitrary starting distribution

- VAE: ✗ Needs tractable base distribution
- NF: ✗ NLL calculation needs tractable base
- Diffusion: ✗ Needs Gaussian base distribution
- GAN:

Generative Models

Now: generalize to arbitrary starting distribution

- VAE: ✗ Needs tractable base distribution
- NF: ✗ NLL calculation needs tractable base
- Diffusion: ✗ Needs Gaussian base distribution
- GAN: ✓
  - No point in loss calculation depends on starting distribution
  - Can have arbitrate starting distribution

Generative Models

Now: generalize to arbitrary starting distribution

- VAE: ✗ Needs tractable base distribution
- NF: ✗ NLL calculation needs tractable base
- Diffusion: ✗ Needs Gaussian base distribution
- GAN: ✓ Can have arbitrate starting distribution

  ✗ •Low performance (compared to SotA)

  •Difficult training

Generative Models

Now: generalize to arbitrary starting distribution

- VAE: ✗ Needs tractable base distribution
- NF: ✗ NLL calculation needs tractable base
- Diffusion: ✗ Needs Gaussian base distribution
- GAN: ✓ Can have arbitrate starting distribution
  ✗ Low performance/difficult training

Generative Models

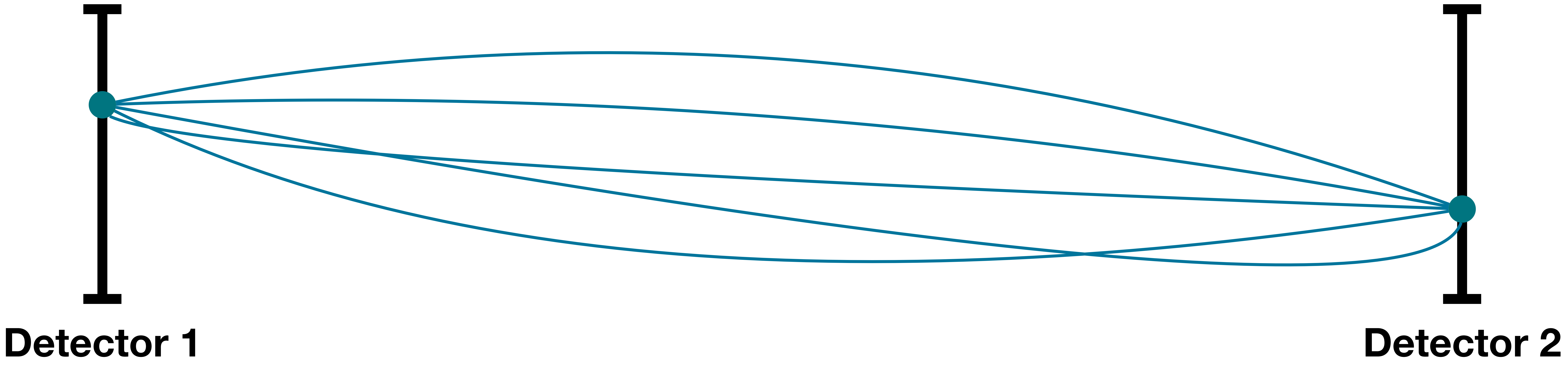
Now: generalize to arbitrary starting distribution

- VAE: ✗ Needs tractable base distribution
- NF: ✗ NLL calculation needs tractable base
- Diffusion: ✗ Needs Gaussian base distribution
- GAN: ✓ Can have arbitrate starting distribution
  ✗ Low performance/difficult training

→ **New model**

Schrödinger Bridges

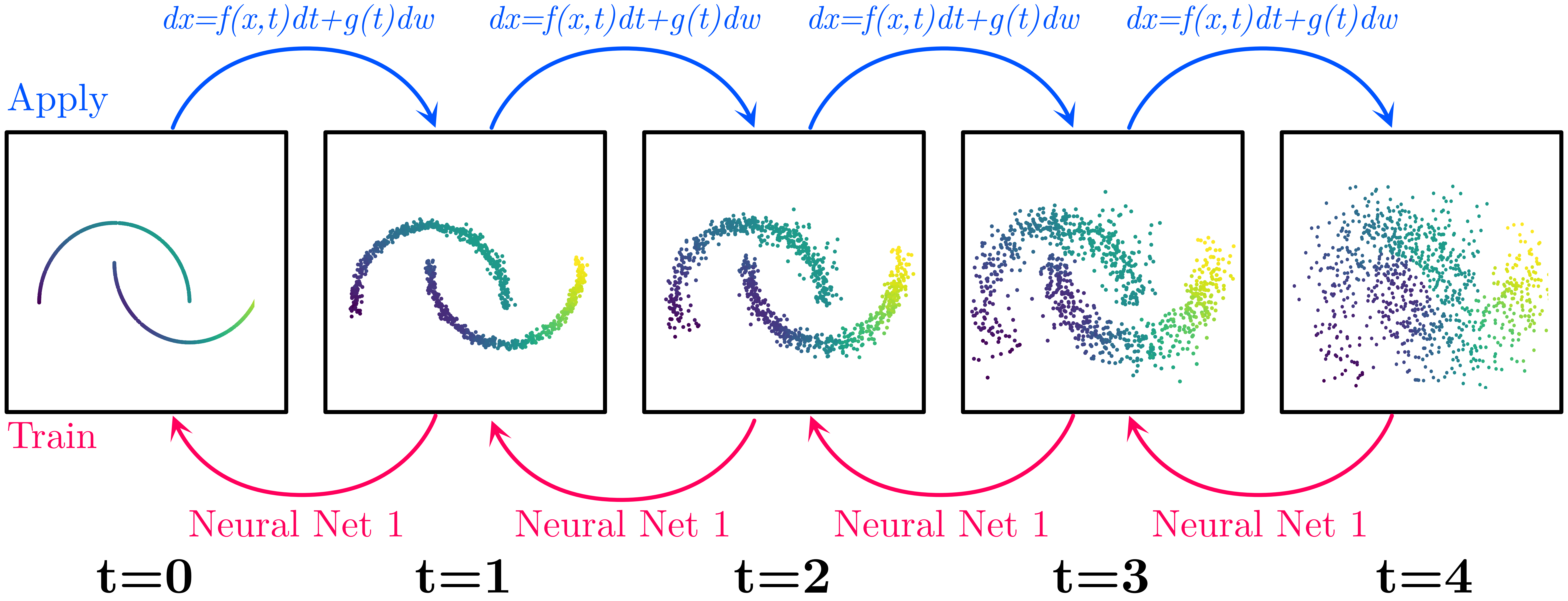
- Based on optimal transport problem



- What is the best path between the two observations
- Boundary condition: $x(t = 0) = x_0$, $x(t = 1) = x_1$

Schrödinger Bridges

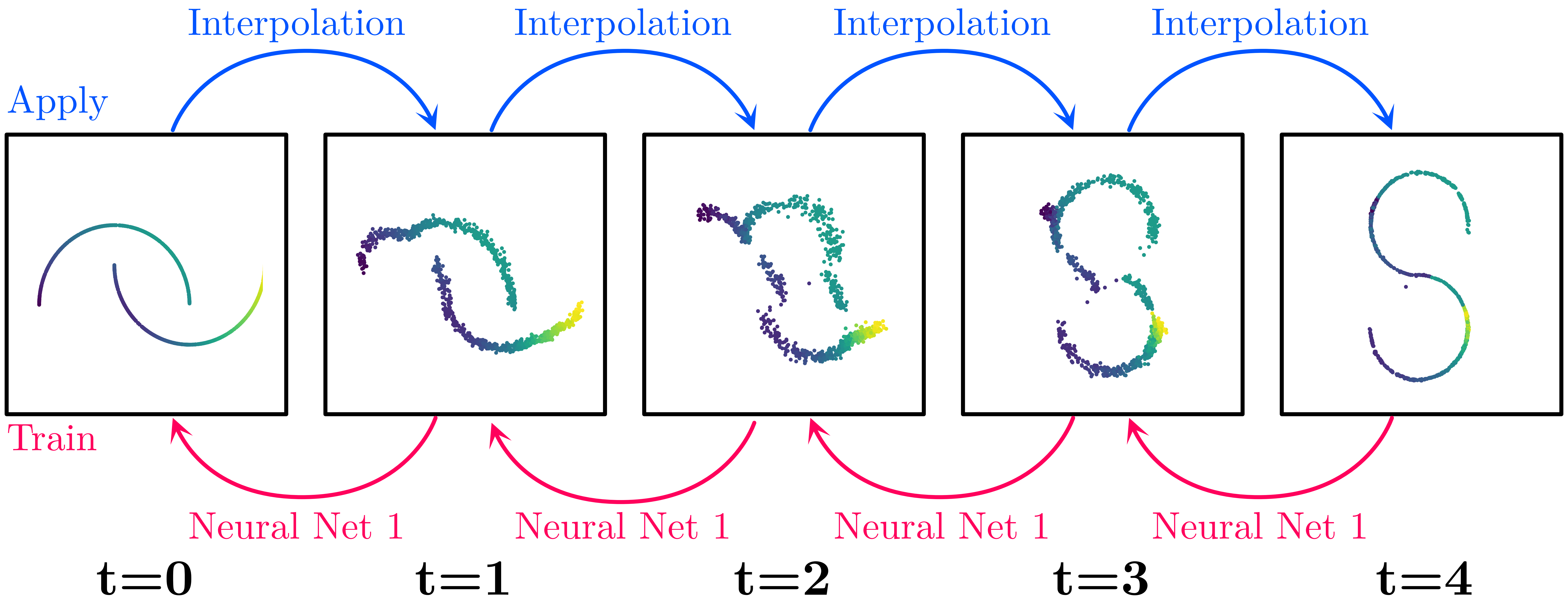
Start from standard Diffusion Models



- Usually: f(x,t) diffuses to normal distribution
- Modify f(x,t) to map between start and target distribution, ensure boundary conditions are matched

Schrödinger Bridges

Approach 1: Stochastic interpolation



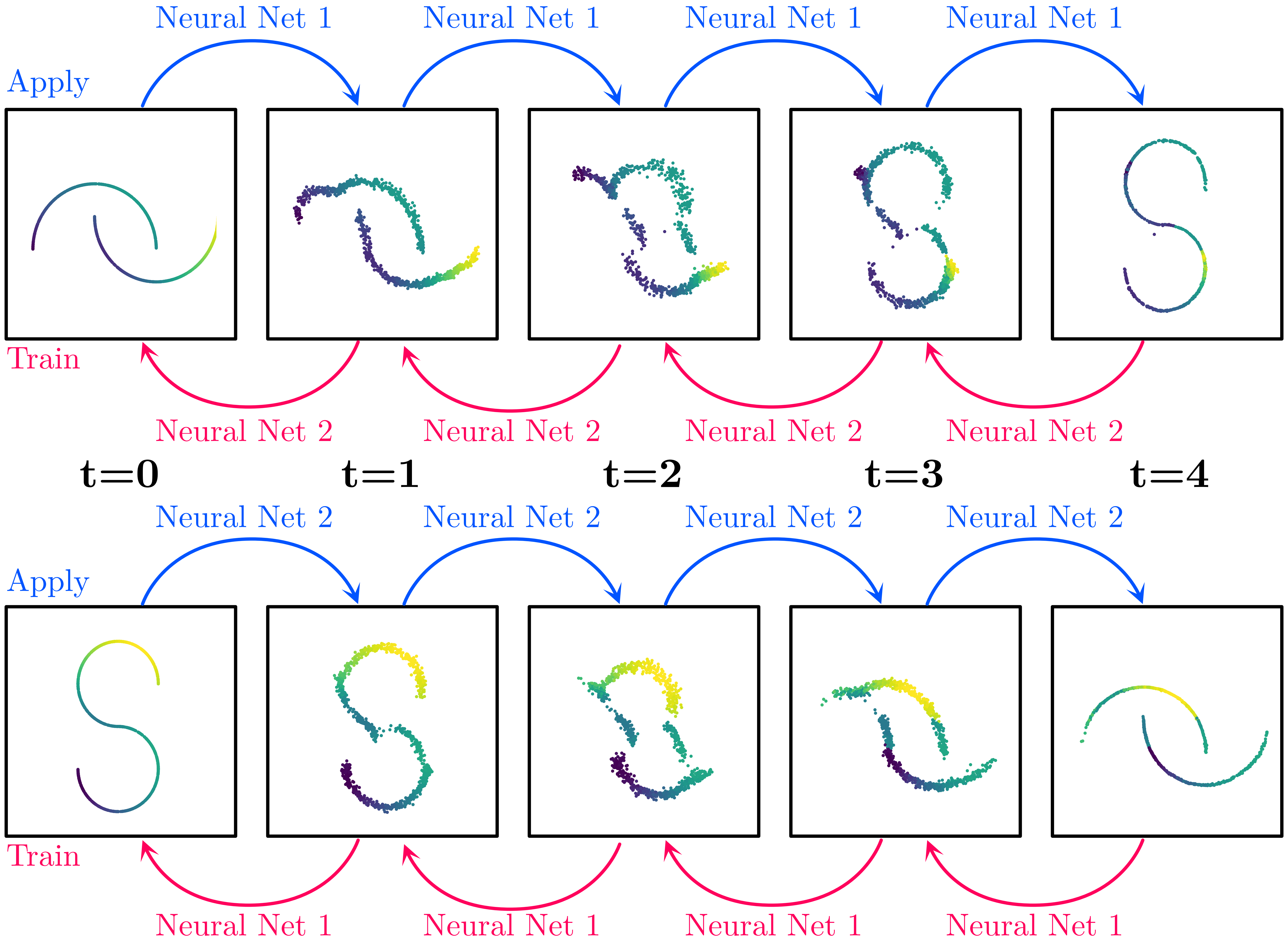
- Define random shift that ends at boundary conditions
- Requires pairs of data points

Guan-Horng Liu et. al. **I2SB: Image-to-Image Schrödinger Bridge,** 2302.05872

Schrödinger Bridges

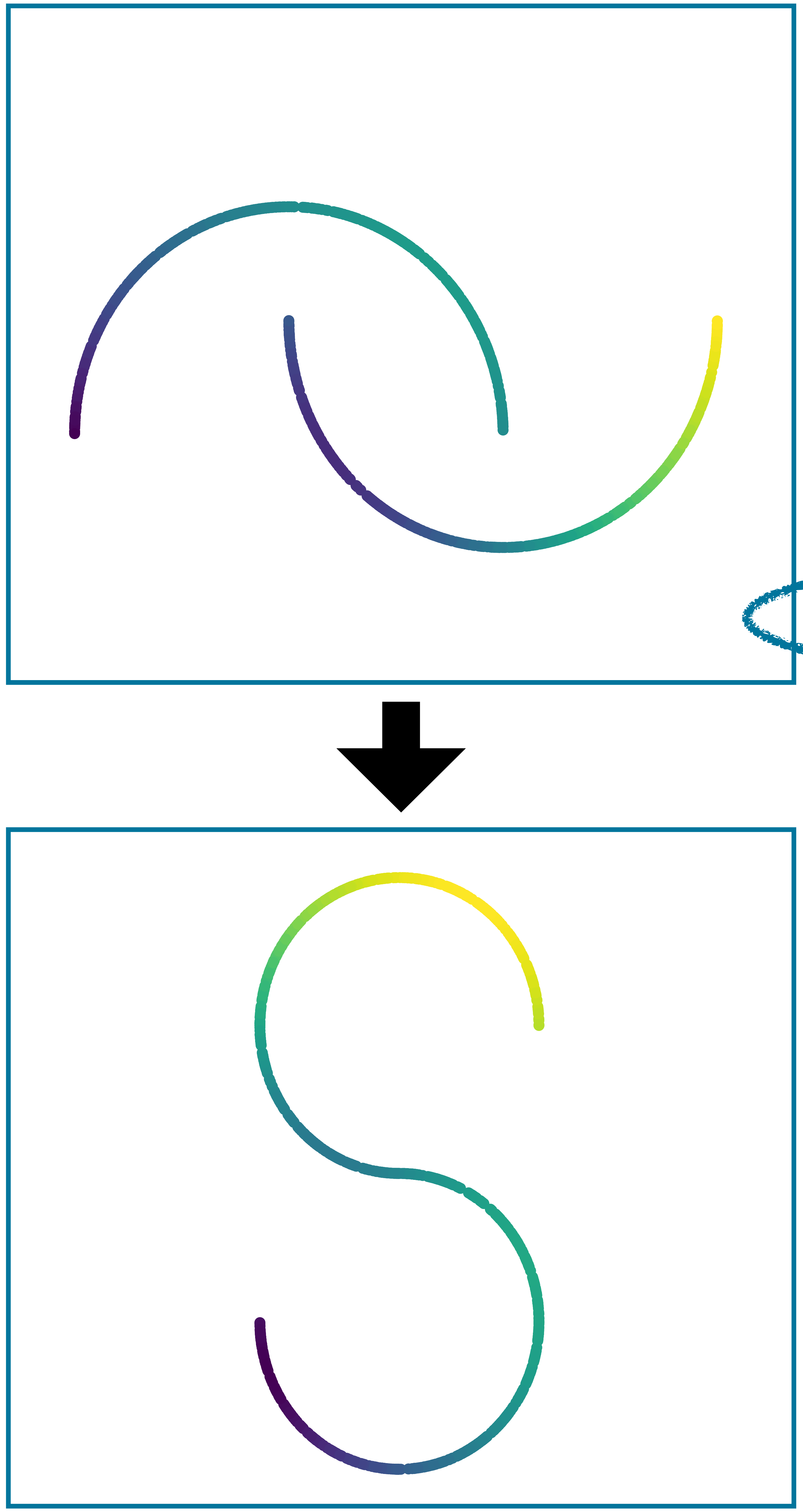
Approach 2: 2 Models with Iterative Fitting

- Train NN 1 from start to target distribution
- Train NN 2 to undo NN 1
- Reverse and repeat
- No paired data required



Valentin De Bortoli et. al. **Diffusion Schrödinger Bridge with Applications to Score-Based Generative Modeling,** 2106.01357

Refinement



- Fundamentally: Refinement = mapping arbitrary distribution to other arbitrary distribution
  - Simulation refinement
  - Unfolding
  - Background estimation
  - Anomaly detection
  - Many more…

Diefenbacher et. al. **Refining Fast Calorimeter Simulations with a Schrödinger Bridge,** 2308.12339

SQuIRELS

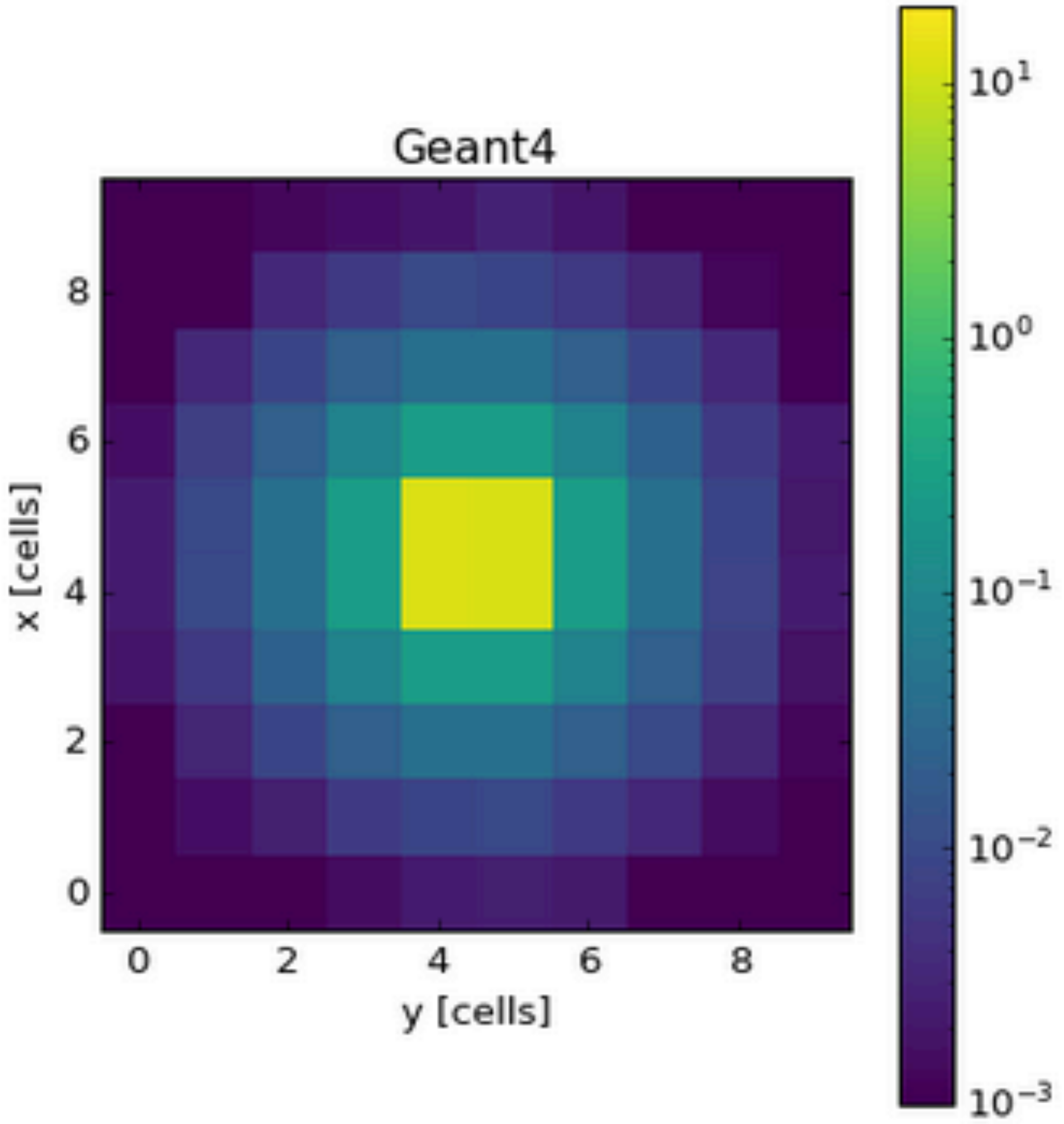
SQuIRELS

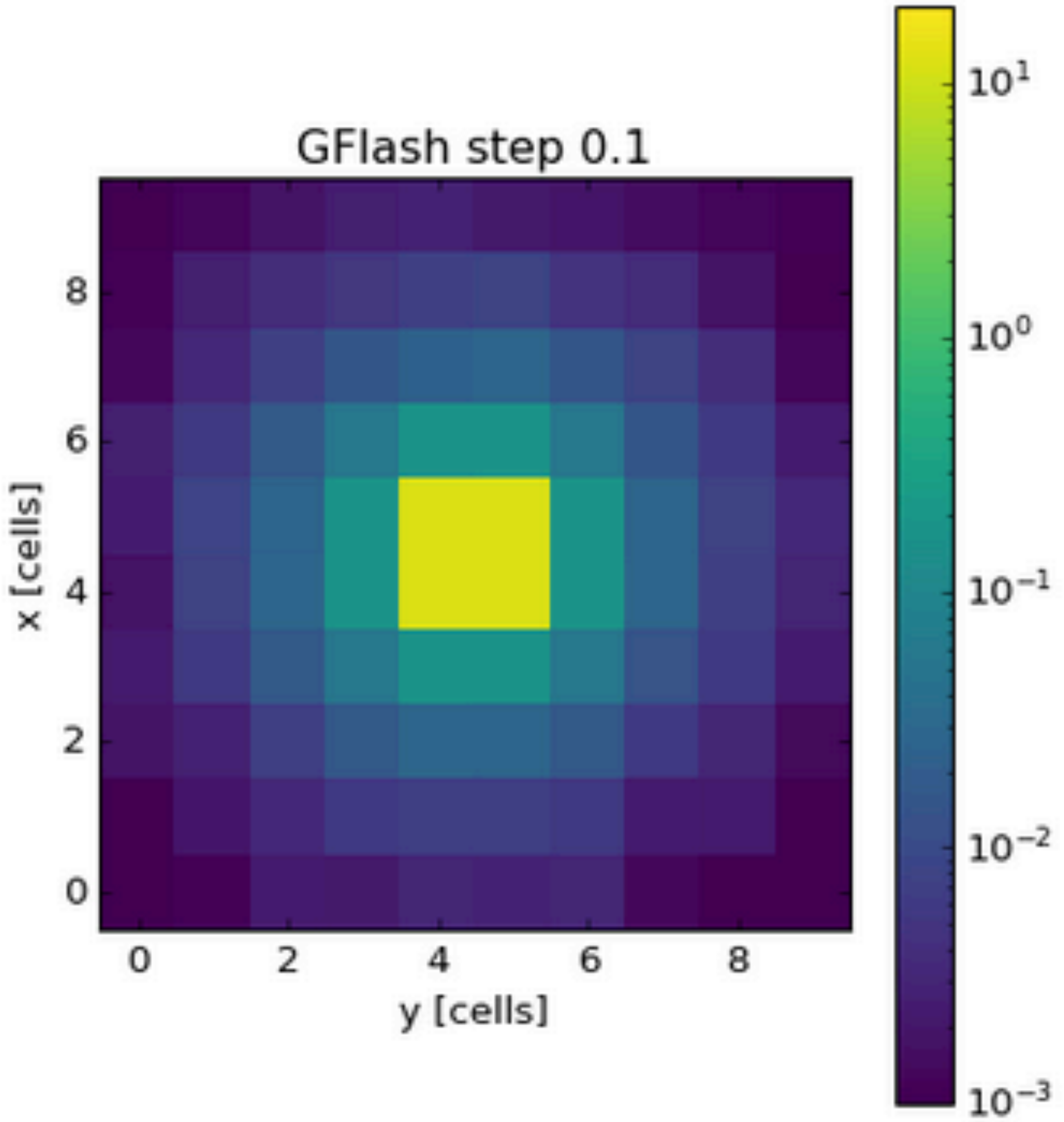
Schrödinger bridge Quality Improvement via Refinement of Existing Lightweight Simulations

- 2308.12339

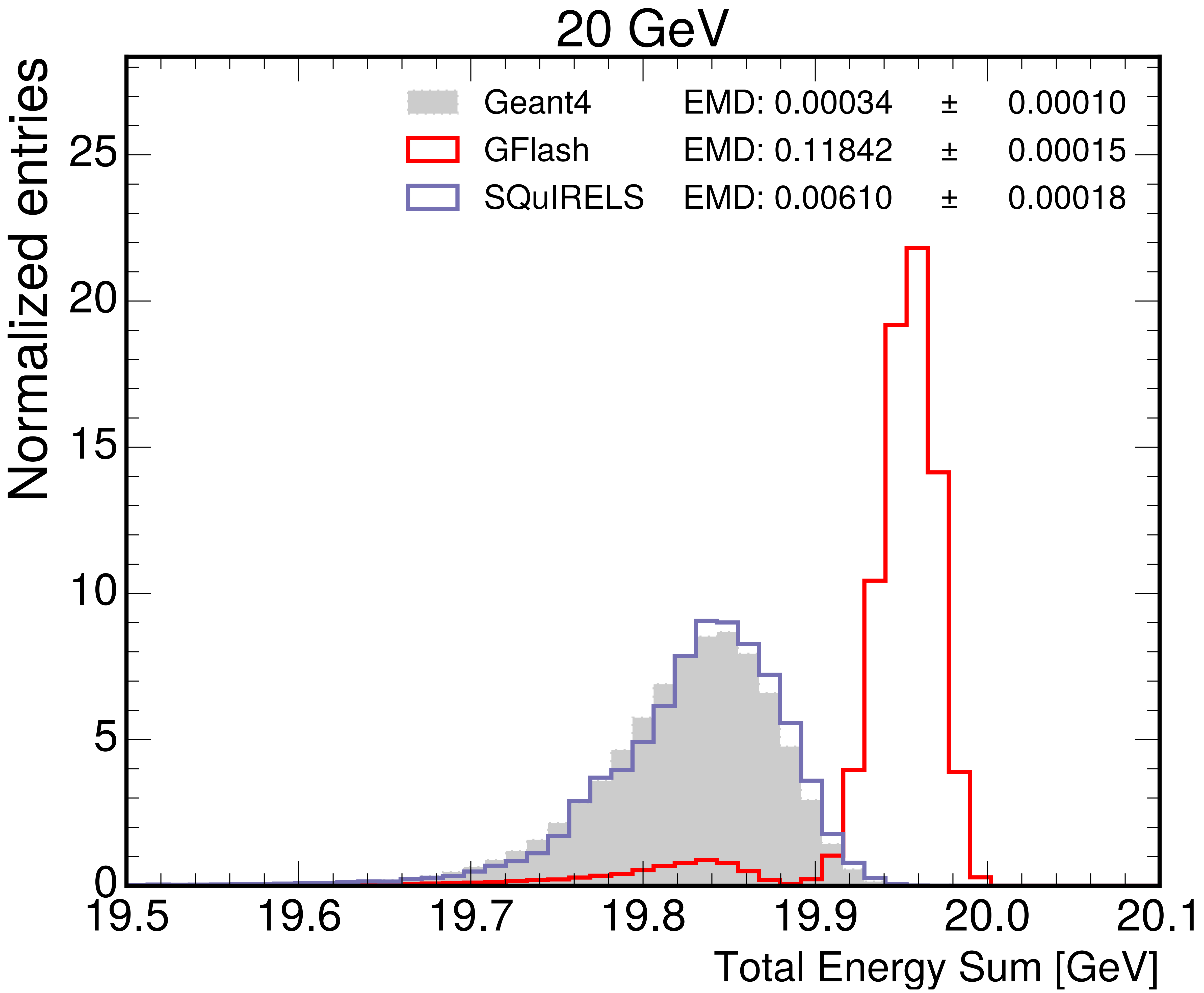
Data

- 10x10 homogeneous crystal calorimeter
- Simulate electron showers for particle energies 10-100 GeV
- Starting distribution: GFlash fast simulation
- Target distribution: Geant4 full simulation

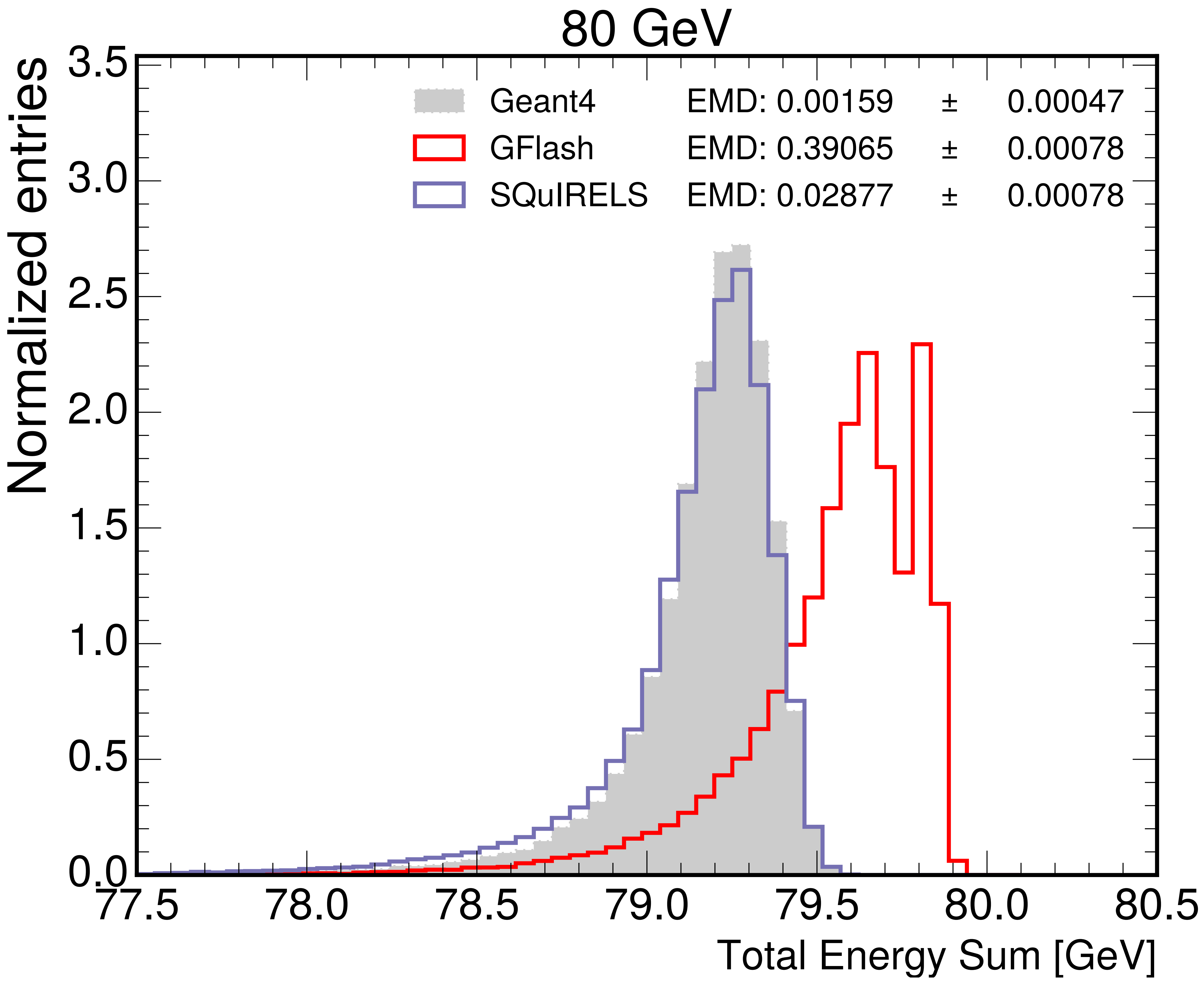




Energy sum

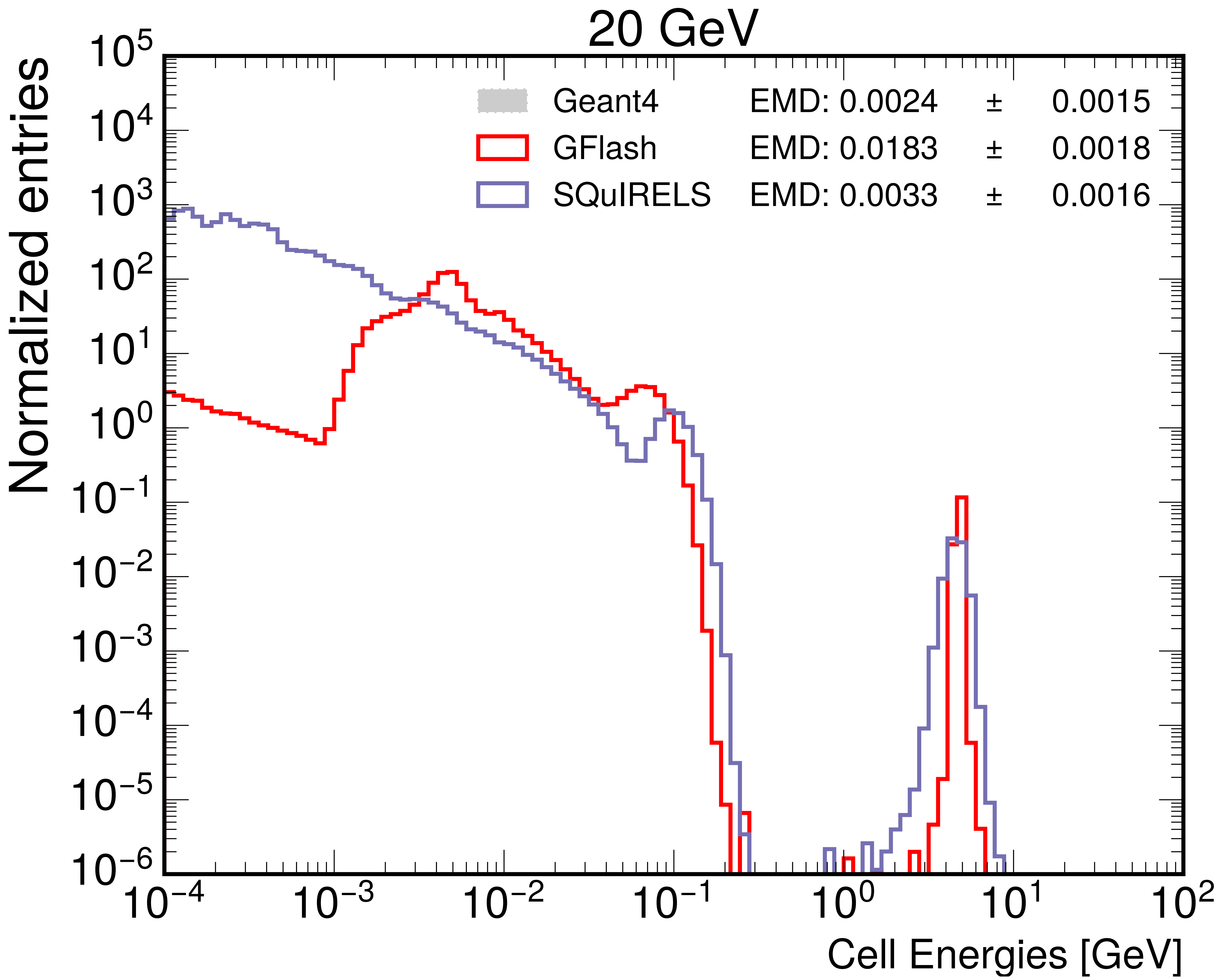


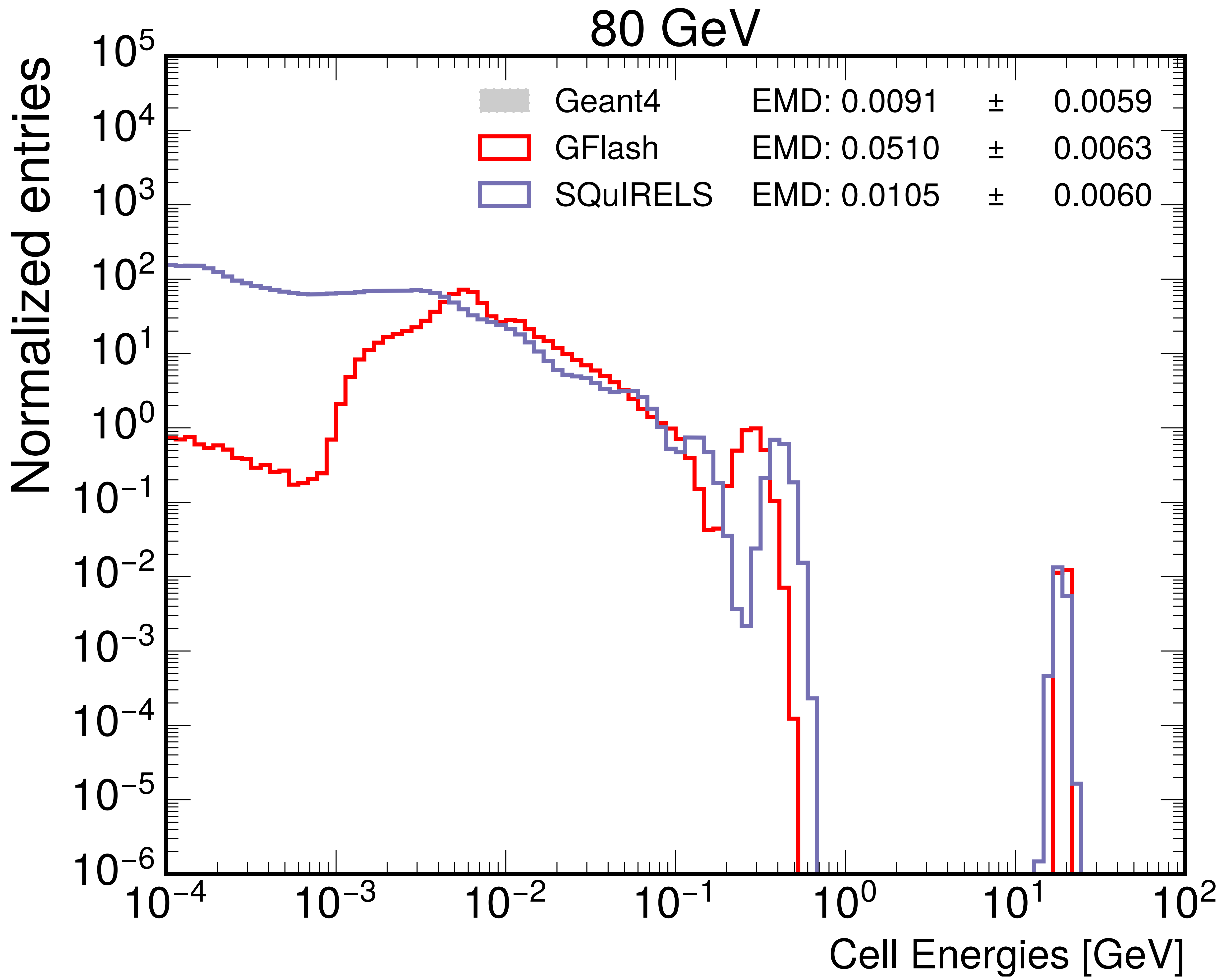
20 GeV
Normalized entries
Geant4 EMD: 0.00034 ± 0.00010
GFlash EMD: 0.11842 ± 0.00015
SQuIRELS EMD: 0.00610 ± 0.00018
0
5
10
15
20
25
19.5
19.6
19.7
19.8
19.9
20.0
20.1
Total Energy Sum [GeV]



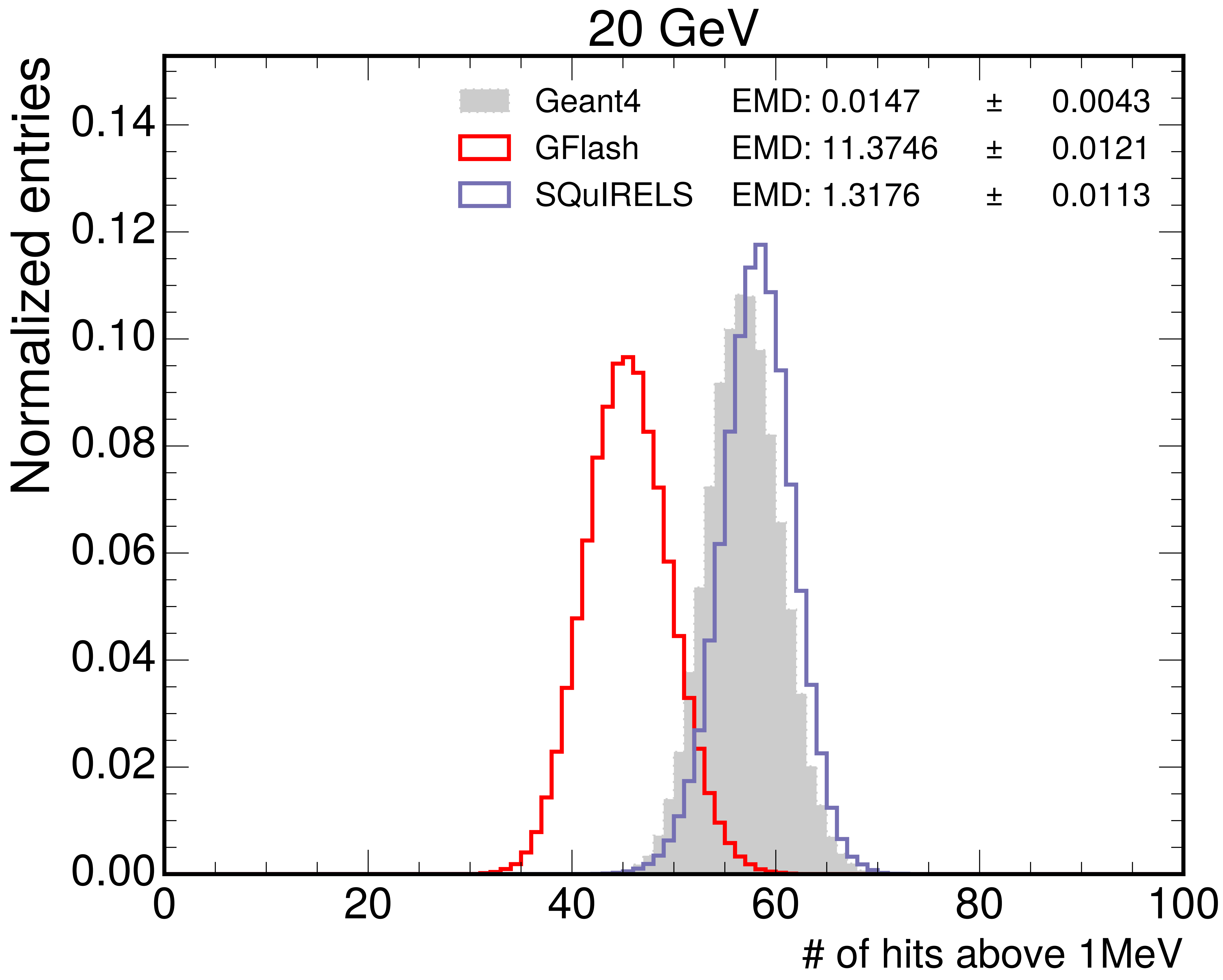
80 GeV
Normalized entries
Geant4 EMD: 0.00159 ± 0.00047
GFlash EMD: 0.39065 ± 0.00078
SQuIRELS EMD: 0.02877 ± 0.00078
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
77.5
78.0
78.5
79.0
79.5
80.0
80.5
Total Energy Sum [GeV]

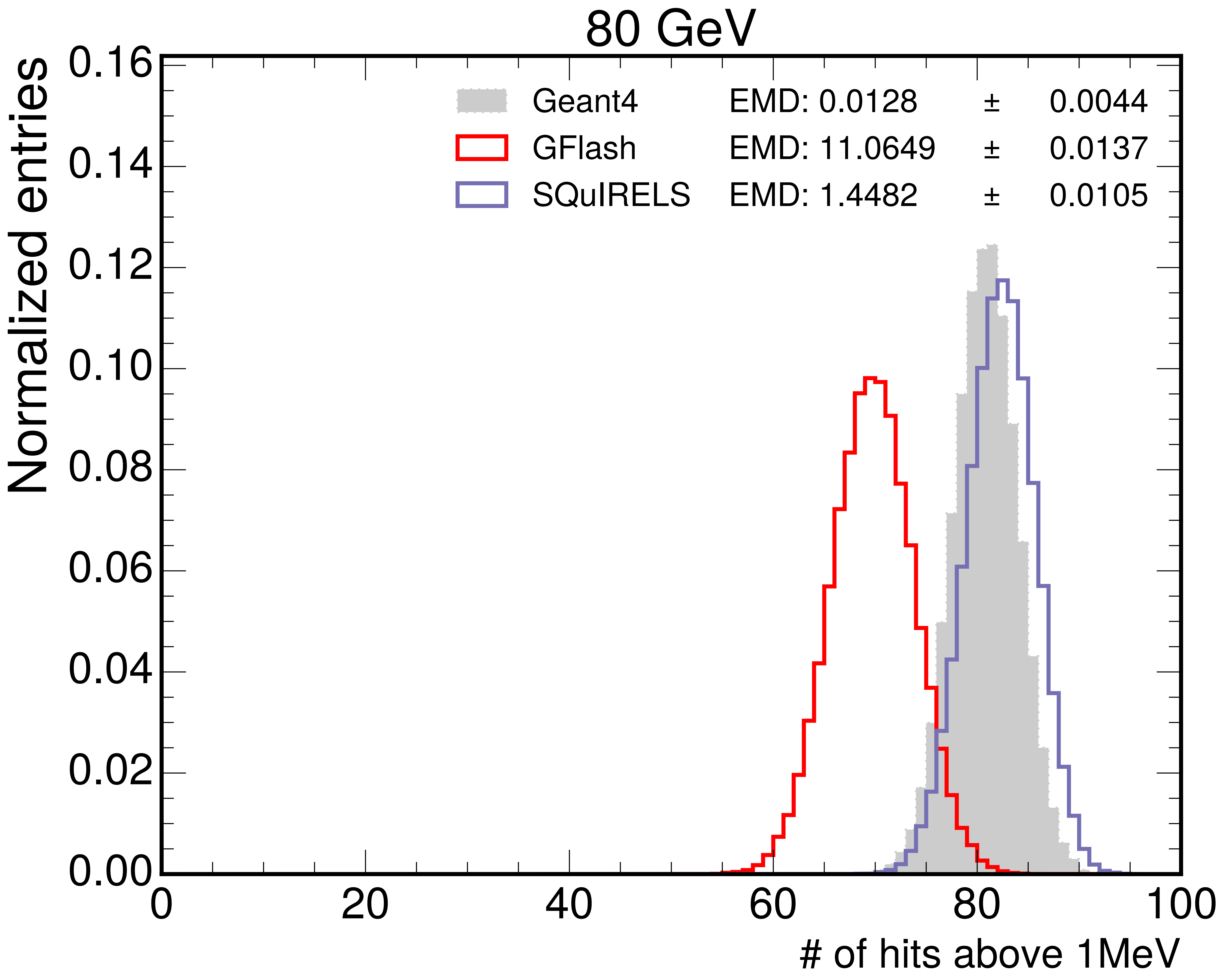
Energy Spectrum





Number of Hits





Timing Results

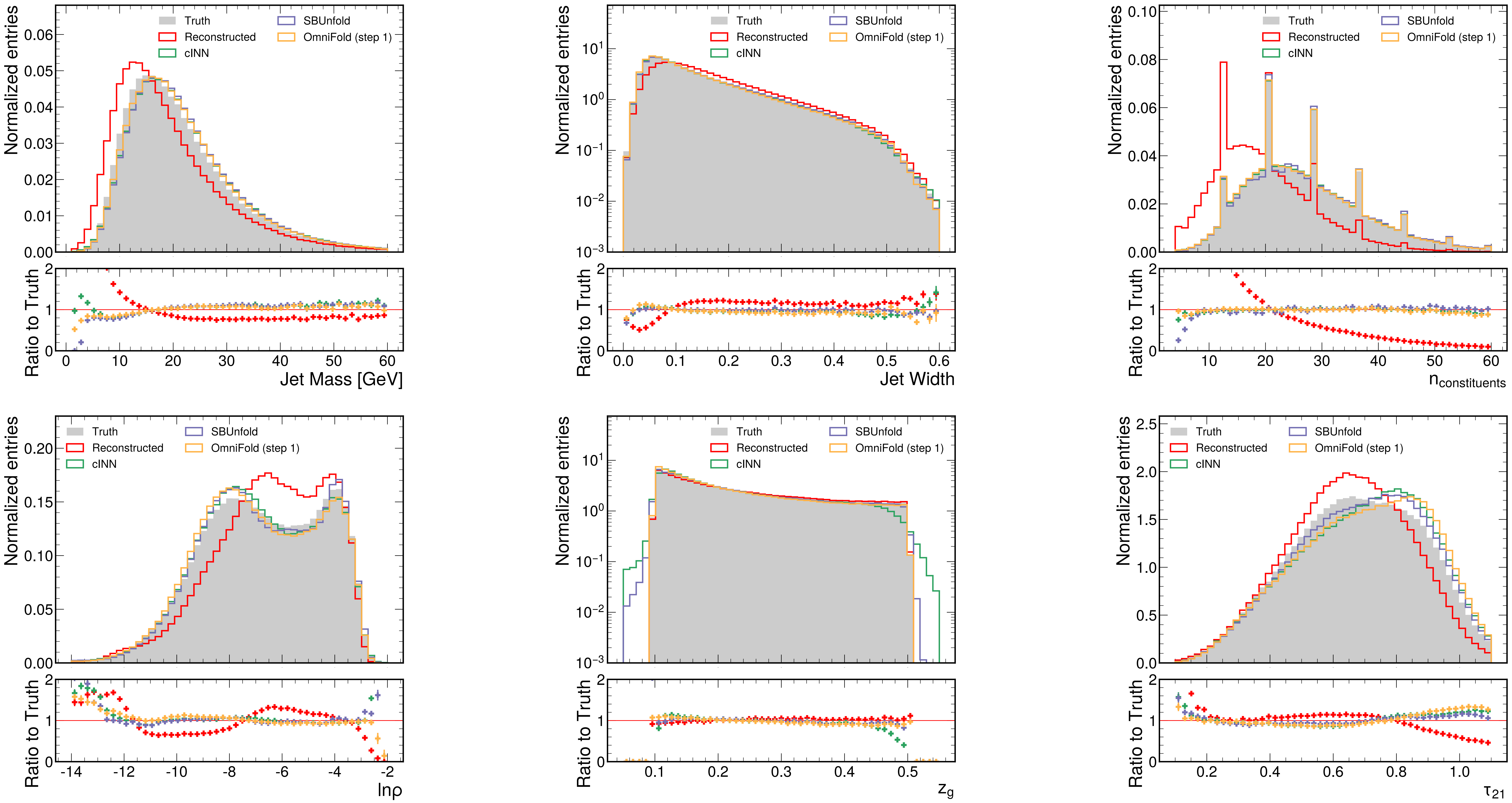
Simulator	CPU [ms/shower]	GPU [ms/shower]
GEANT4	404.8 ± 8.5	/
GFLASH	8.5 ± 0.4	/
SQuIRELS (refine)	7.21 ± 0.04	0.0522 ± 0.0002
SQuIRELS (full)	15.7 ± 0.4	8.5 ± 0.4

Refinement

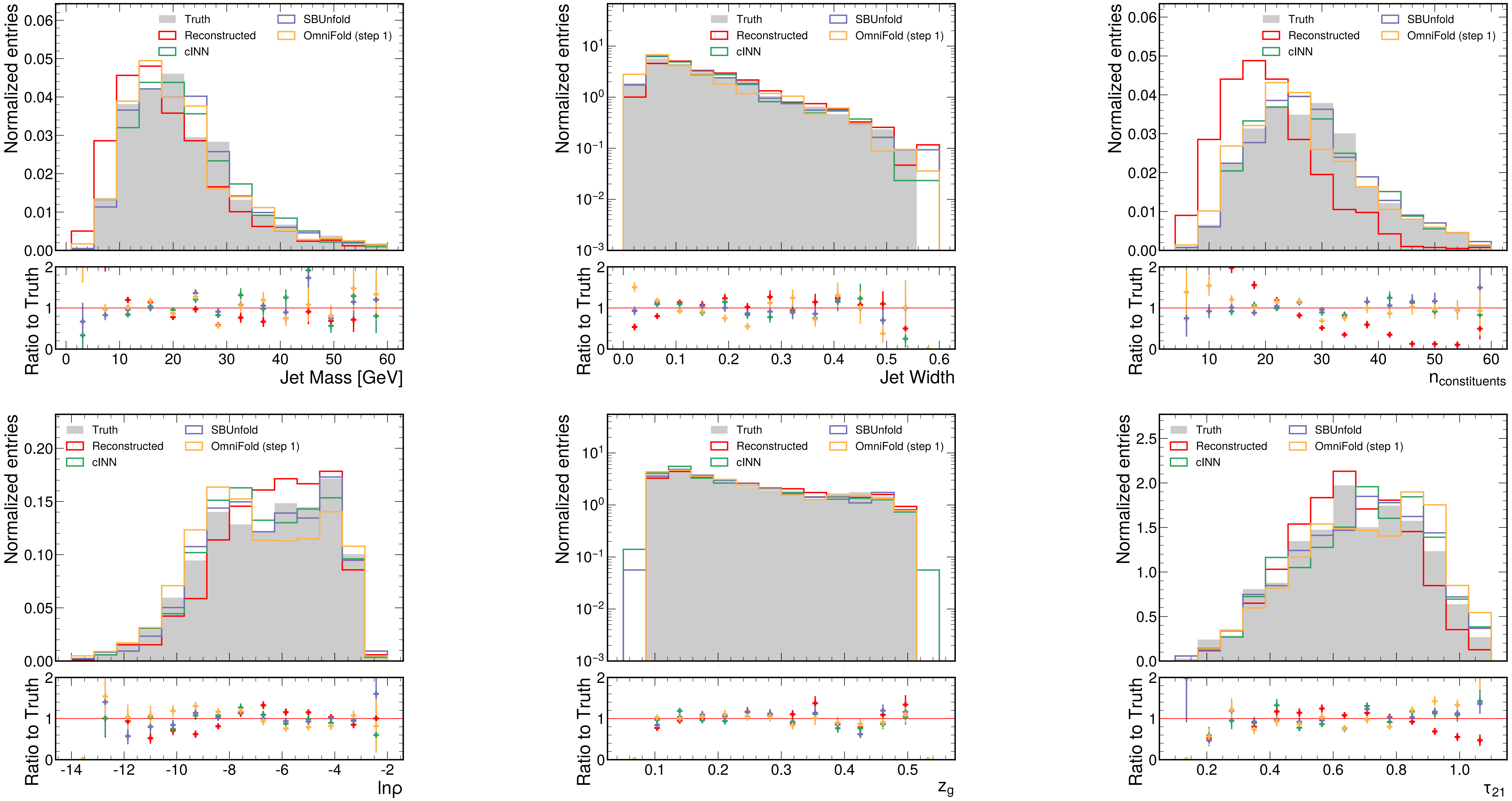
- Fundamentally: Refinement = mapping arbitrary distribution to other arbitrary distribution
  - Simulation refinement
  - Unfolding
  - Background estimation
  - Anomaly detection
  - Many more…

Diefenbacher et. al. **Improving Generative Model-based Unfolding with Schrödinger Bridges,** 2308.12351

Herwig → Pythia (High Stat.)



Herwig → Pythia (Low Stat.)



Herwig → Pythia

High Stat:

Model	EMD(×10)/Triangular Discriminator(×10³)		
	OmniFold Step 1	cINN	SBUnfold
Jet mass	**6.1±0.1/1.5**	10.1±1.2/2.4	9.0±0.1/3.1
Jet Width	0.06±0.001/1.1	0.05±0.003/0.7	**0.02±0.001/0.2**
N	**1.7±0.1/0.2**	6.1±4.0/0.2	3.0±0.1/0.6
$\log \rho$	1.35±0.03/1.1	3.1±2.1/0.8	**0.4±0.1/0.7**
z_g	0.086±0.001/**1.2**	0.3±0.1/12.7	**0.049±0.001**/3.5
τ_{21}	0.23±0.02/4.6	0.7±0.4/3.5	**0.12±0.02/1.4**

Low Stat:

Model	EMD(×10)/Triangular Discriminator(×10³)		
	OmniFold Step 1	cINN	SBUnfold
Jet mass	**8.7±1.8**/13.6	9.2±3.0/8.4	**7.7±2.5/6.9**
Jet Width	0.14±0.02/18	0.07±0.02/5.7	**0.05±0.02/4.6**
N	12±3/10.9	**5.4±1.3/3.8**	**5.8 ±1.6**/3.7
$\log \rho$	4.0±0.8/11	1.6±0.5/6.2	**1.2±0.3/4.4**
z_g	0.08±0.02/**1.5**	0.08±0.03/7.2	**0.06±0.01**/7.1
τ_{21}	0.4±0.07/16	0.2±0.05/12	**0.1±0.04/8**

Conclusion

- Schrödinger Bridges promising for any refinement
- Calorimeter refinement
  - Faster than full simulation
  - More accurate than fast sim.
- Unfolding
  - Beats state of the art unfolding methods

Backup

Schrödinger Bridges

- Start from initial distribution $p_0(x_0) = p_\alpha$

$$p(x_0, N) = p_0(x_0) \prod_{k=0}^{N-1} p_{k+1|k}(x_{k+1}|x_k)$$

- Apply transition kernels $p_{k+1|k}$
- This defines path π
- Now demand that path fulfills boundary condition:

$$\pi^* = \operatorname{argmin}\{\mathrm{KL}(\pi|p) : \pi_0 = p_\alpha, \pi_N = p_\beta\}$$

Schrödinger Bridges

- Can be approximated using iterative fitting procedure:

$$\pi^{2n+1} = \arg\min \left\{ \mathrm{KL}(\pi|\pi^{2n}) : \pi_N = p_\beta \right\}$$
$$\pi^{2n+2} = \arg\min \left\{ \mathrm{KL}(\pi|\pi^{2n+1}) : \pi_0 = p_\alpha \right\}$$

- Gradually brings endpoints of path closer to target distributions
- For generative application, distr. unknown, KLD hard to calculate
- Instead: define Gaussian transition kernels in both directions

$$p^n_{k+1|k}(x_{k+1}|x_k) = \mathcal{N}(x_{k+1}; F^n_k(x_k), 2\gamma_{k+1})$$
$$q^n_{k|k+1}(x_k|x_{k+1}) = \mathcal{N}(x_k; B^n_{k+1}(x_{k+1}), 2\gamma_{k+1})$$

Schrödinger Bridges

- Instead: define Gaussian transition kernels in both directions

$$p^n_{k+1|k}(x_{k+1}|x_k) = \mathcal{N}(x_{k+1}; F^n_k(x_k), 2\gamma_{k+1})$$

$$q^n_{k|k+1}(x_k|x_{k+1}) = \mathcal{N}(x_k; B^n_{k+1}(x_{k+1}), 2\gamma_{k+1})$$

- New recursive updating method:

$$B^n_{k+1} = \text{argmin}_B \mathbb{E} \left\| B(x_{k+1}) - x_{k+1} - F^n_k(x_k) + F^n_k(x_{k+1}) \right\|^2$$

$$F^{n+1}_k = \text{argmin}_F \mathbb{E} \left\| F(x_k) - x_k - B^n_{k+1}(x_{k+1}) + B^n_{k+1}(x_k) \right\|^2$$

- Approximate F and B as neural networks, and define forward steps:

$$x_{k+1} = F^n_\sigma(k, x_k) + \sqrt{2\gamma_{k+1}} Z,$$

$$x_{k-1} = B^n_\theta(k, x_k) + \sqrt{2\gamma_k} \tilde{Z},$$

- Alternatingly fix F/B and update B/F

Generative Fast Sim

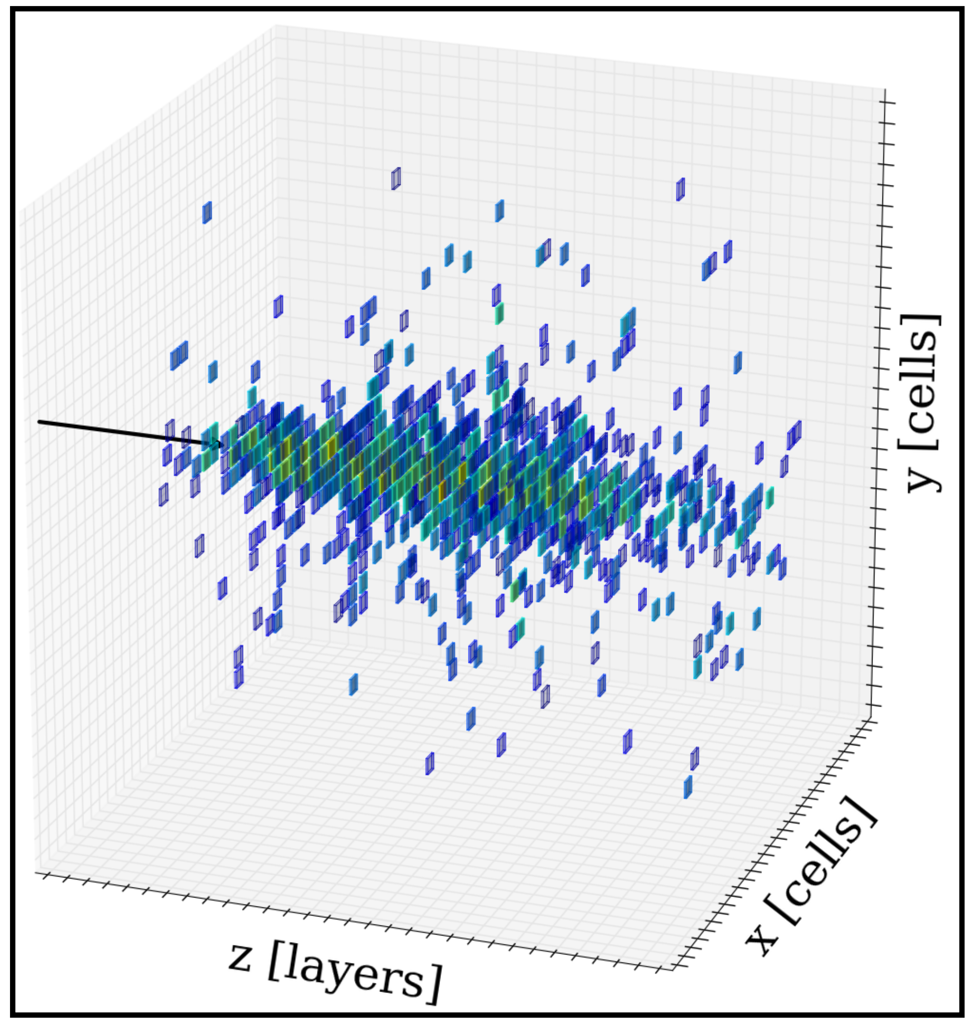
Standard generative approach:

Map random noise to new data

Latent Space

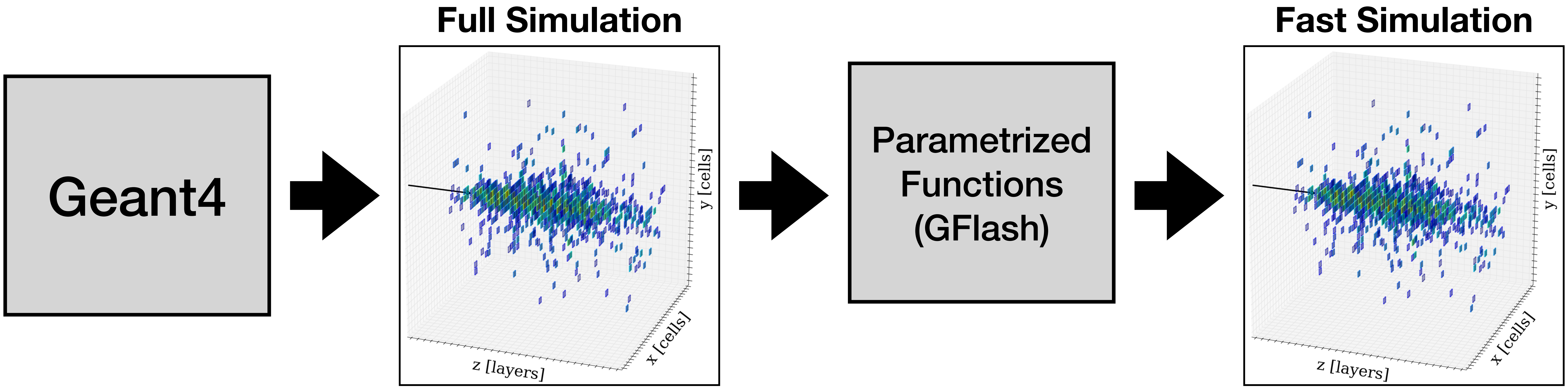
Noise → **Model f(Ω)** →

High Dim. Data

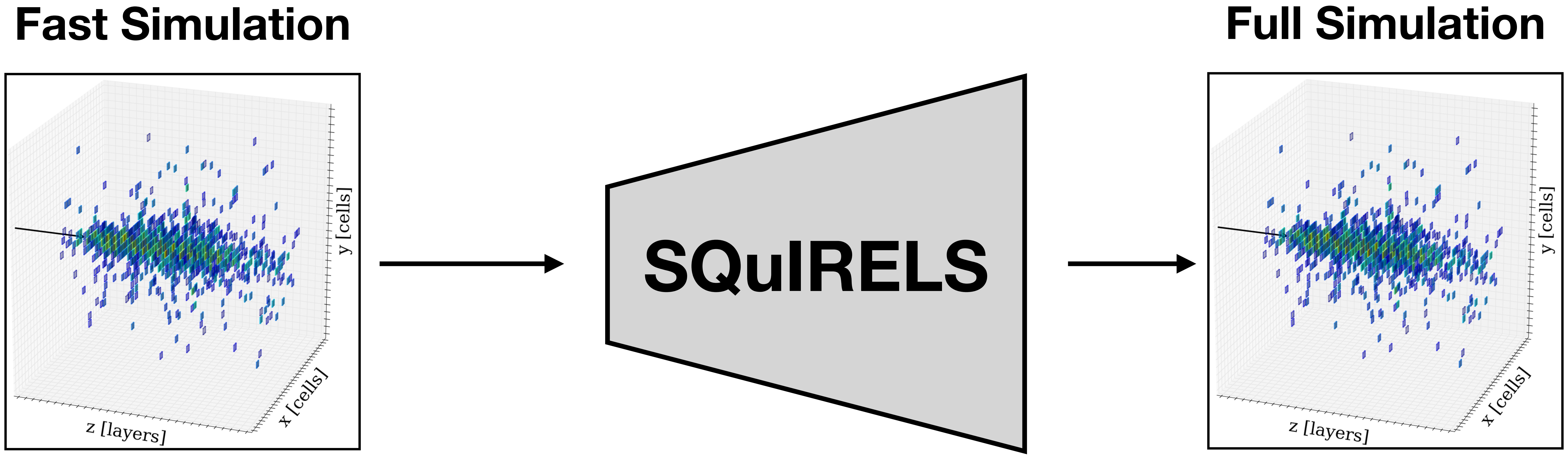


Classical Fast Sim

- Build reference shower set using full sim
- Define parametrized functions on full sim showers
- Quickly sample from parametrized model

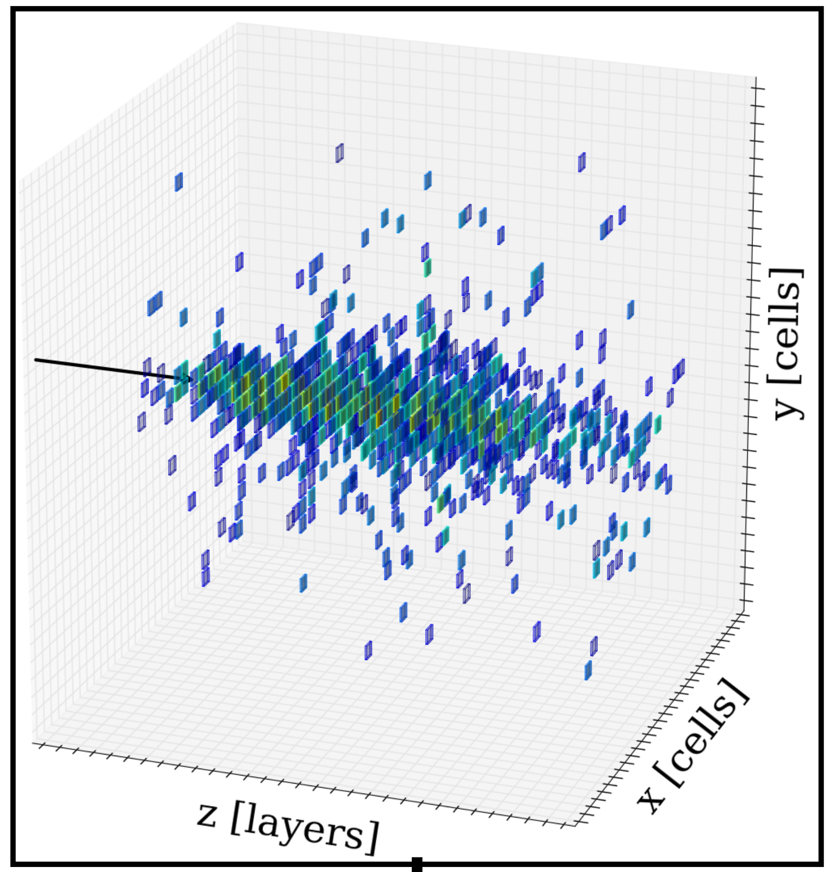


SQuIRELS Setup



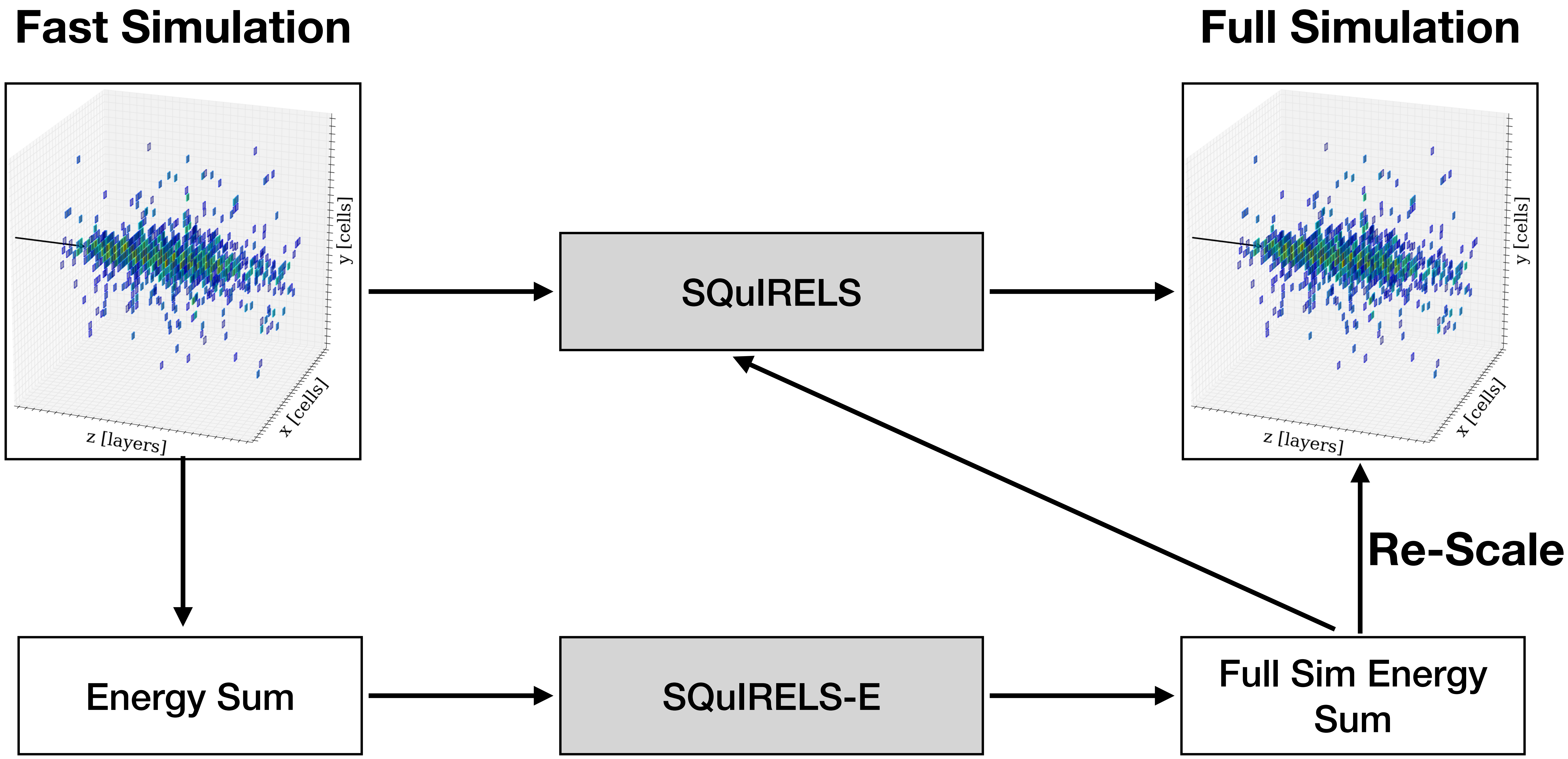
SQuIRELS Setup

Fast Simulation



Energy Sum → SQuIRELS-E → Full Sim Energy Sum

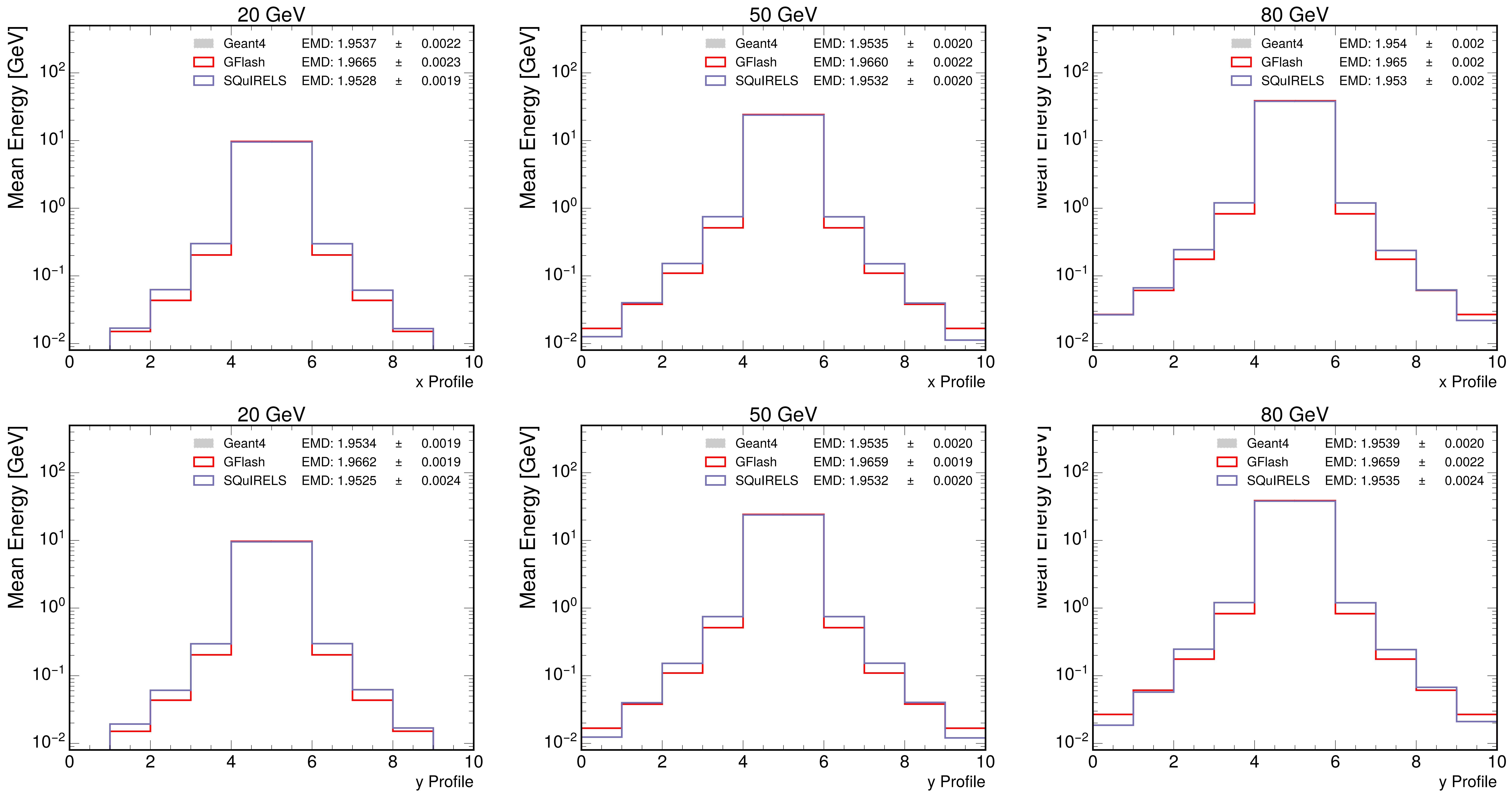
SQuIRELS Setup



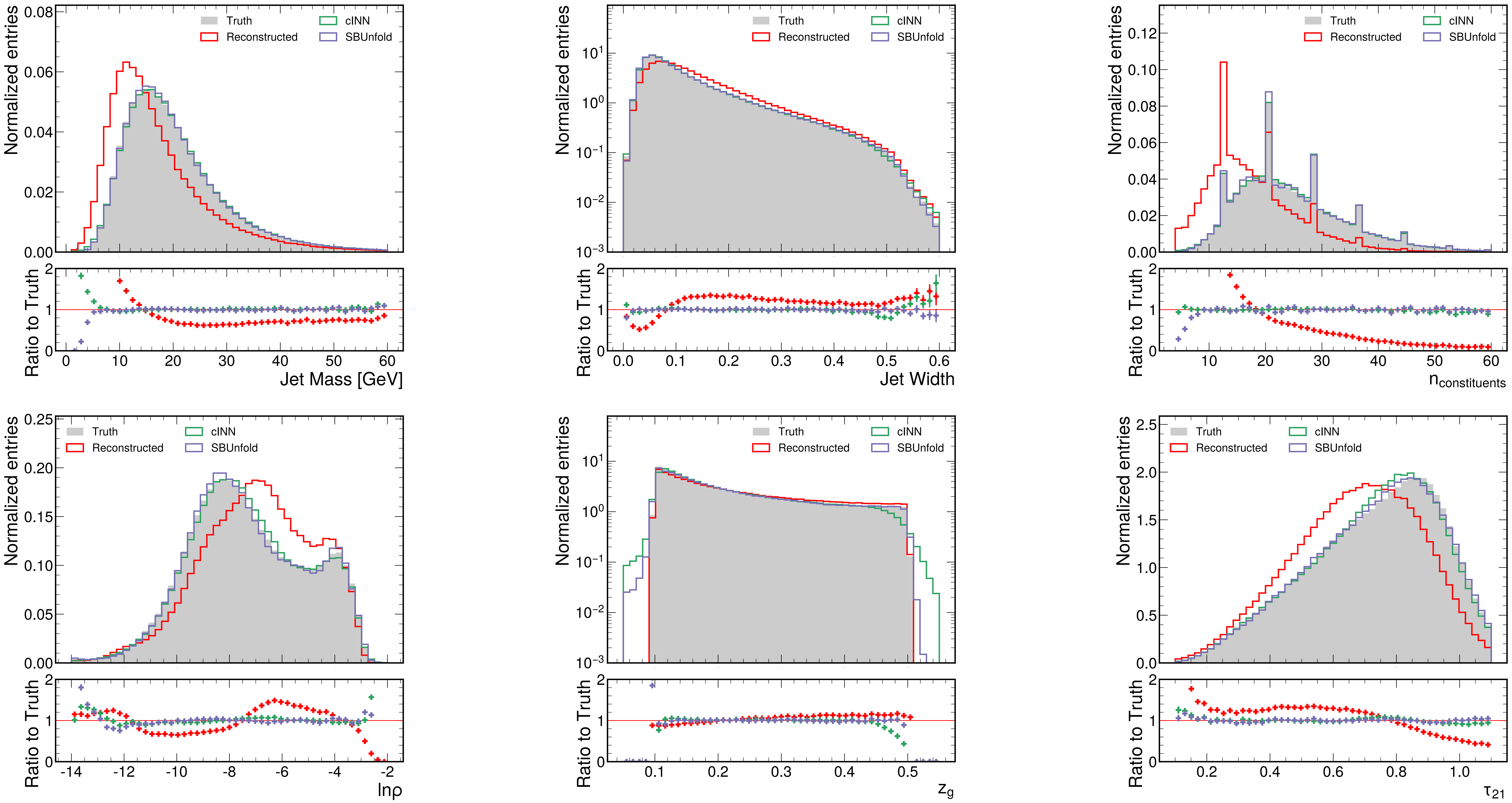
Results

Observalbe	EMD		
	GEANT4	GFLASH	SQuIRELS
E_{sum} 20 GeV	0.0003(1)	0.1184(2)	0.0061(2)
E_{sum} 50 GeV	0.0009(2)	0.2599(6)	0.0147(4)
E_{sum} 80 GeV	0.0016(5)	0.3906(8)	0.0288(8)
E_{spec} 20 GeV	0.002(1)	0.018(2)	0.003(2)
E_{spec} 50 GeV	0.006(4)	0.034(3)	0.007(4)
E_{spec} 80 GeV	0.009(6)	0.051(6)	0.011(6)
N_{hit} 20 GeV	0.015(4)	11.37(1)	1.32(1)
N_{hit} 50 GeV	0.015(5)	11.84(1)	0.085(7)
N_{hit} 80 GeV	0.013(4)	11.06(1)	1.45(1)
E_{max} 20 GeV	0.0014(5)	0.3379(2)	0.0490(9)
E_{max} 50 GeV	0.0021(7)	0.4342(4)	0.083(1)
E_{max} 80 GeV	0.0027(9)	0.4757(5)	0.106(2)
$profile_x$ 20 GeV	1.954(2)	1.967(2)	1.953(2)
$profile_x$ 50 GeV	1.954(2)	1.966(2)	1.953(2)
$profile_x$ 80 GeV	1.954(2)	1.965(2)	1.953(2)
$profile_y$ 20 GeV	1.953(2)	1.966(2)	1.953(2)
$profile_y$ 50 GeV	1.954(2)	1.966(2)	1.953(2)
$profile_y$ 80 GeV	1.954(2)	1.966(2)	1.953(2)

Profiles


20 GeV
Geant4 EMD: 1.9537 ± 0.0022
GFlash EMD: 1.9665 ± 0.0023
SQuIRELS EMD: 1.9528 ± 0.0019
Mean Energy [GeV]
x Profile
50 GeV
Geant4 EMD: 1.9535 ± 0.0020
GFlash EMD: 1.9660 ± 0.0022
SQuIRELS EMD: 1.9532 ± 0.0020
Mean Energy [GeV]
x Profile
80 GeV
Geant4 EMD: 1.954 ± 0.002
GFlash EMD: 1.965 ± 0.002
SQuIRELS EMD: 1.953 ± 0.002
x Profile
20 GeV
Geant4 EMD: 1.9534 ± 0.0019
GFlash EMD: 1.9662 ± 0.0019
SQuIRELS EMD: 1.9525 ± 0.0024
Mean Energy [GeV]
y Profile
50 GeV
Geant4 EMD: 1.9535 ± 0.0020
GFlash EMD: 1.9659 ± 0.0019
SQuIRELS EMD: 1.9532 ± 0.0020
Mean Energy [GeV]
y Profile
80 GeV
Geant4 EMD: 1.9539 ± 0.0020
GFlash EMD: 1.9659 ± 0.0022
SQuIRELS EMD: 1.9535 ± 0.0024
y Profile

Pythia → Pythia (vs Diffusion)



Pythia → Pythia (vs Diffusion)

Model	EMD(×10)/Triangular Discriminator(×10^3)		
	FPCD	cINN	SBUnfold
Jet mass	0.74±0.08/**0.19**	1.4±0.2/0.29	**0.70±0.06**/0.30
Jet Width	0.0087±0.0006/0.9	0.013±0.002/0.25	**0.0029±0.0005/0.04**
N	0.81±0.06/0.1	2.3±0.8/**0.09**	**0.57±0.04**/0.9
$\log \rho$	0.34±0.01/0.77	1.1±0.3/**0.64**	**0.27±0.01**/0.68
z_g	0.035±0.007/12.4	0.095±0.003/10.9	**0.009±0.001/3.1**
τ_{21}	0.024±0.002/0.3	0.2±0.1/0.6	**0.016±0.001/0.2**