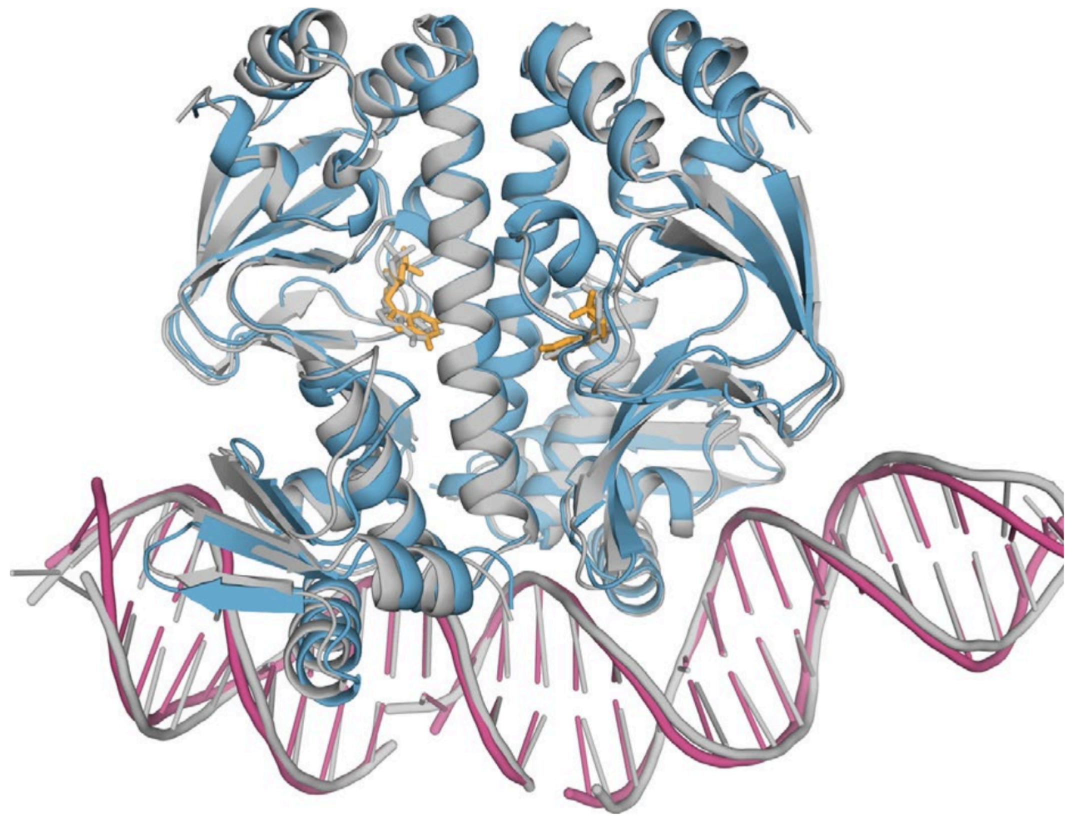
Diffusion models for the Electron-Ion Collider

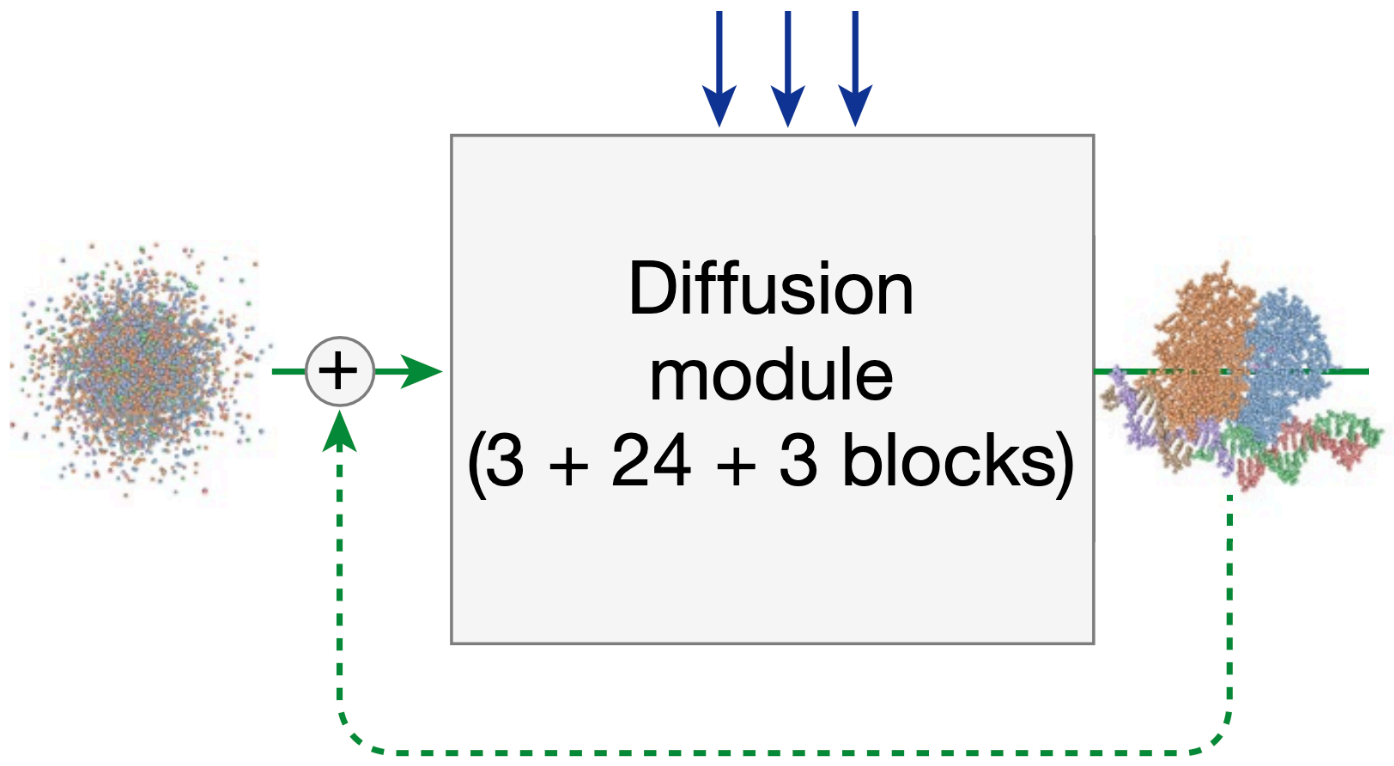
Felix Ringer

QCD at the Femtoscale in the Era of Big Data,
INT, UW, 2024

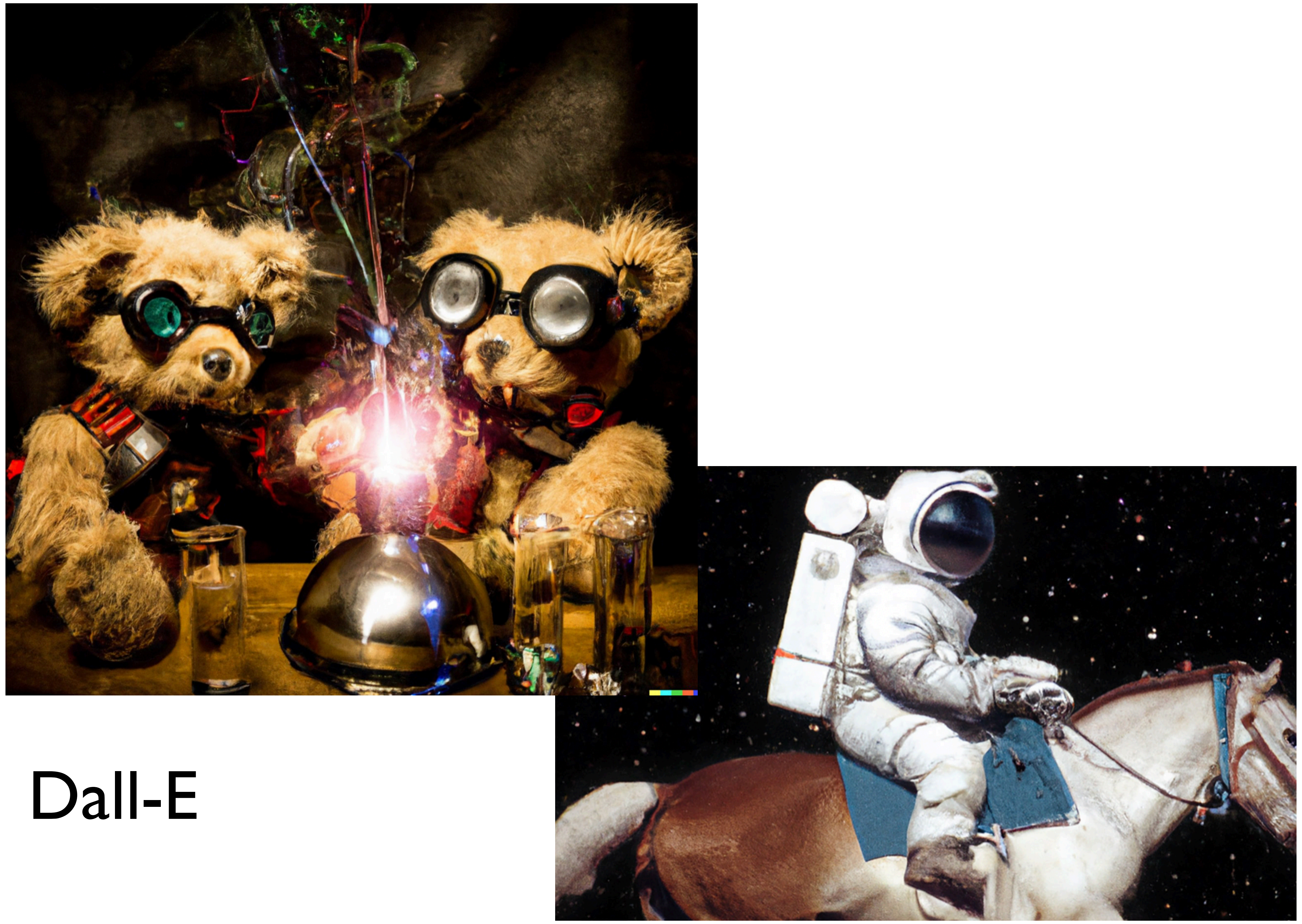
Old Dominion University
Jefferson Lab

Recent progress generative AI





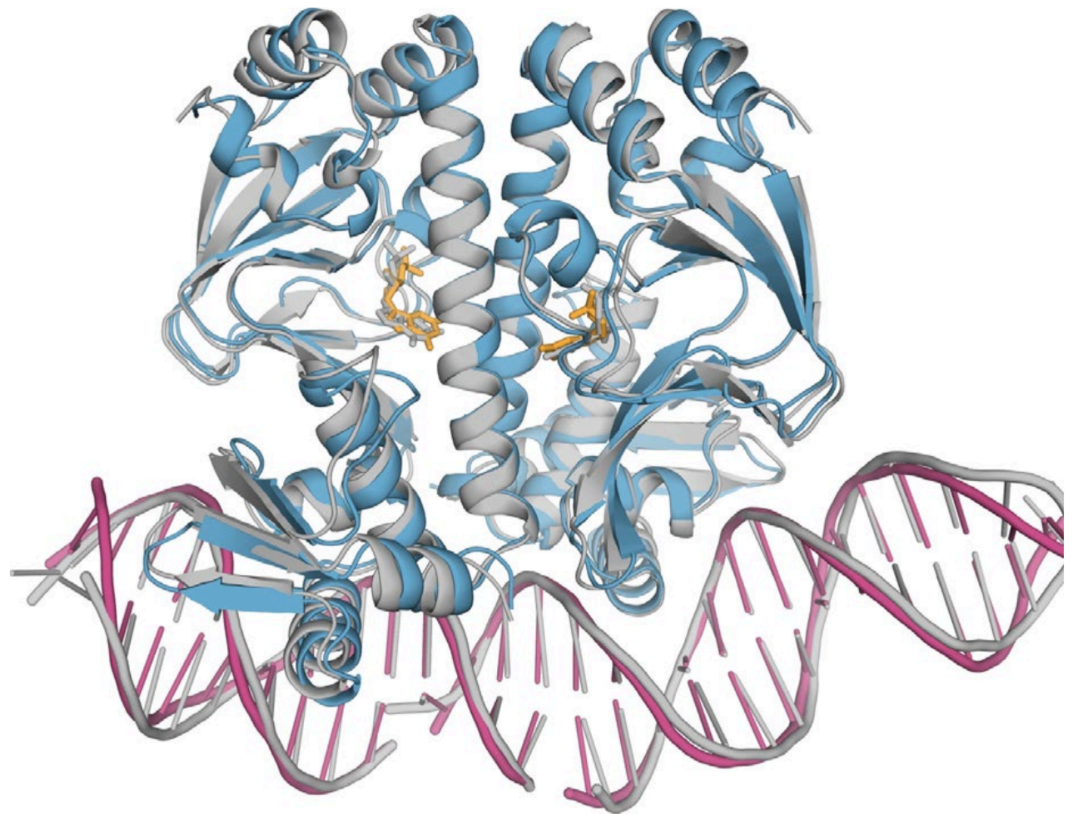
Diffusion-based architecture
in AlphaFold 3 *Abramson `24*

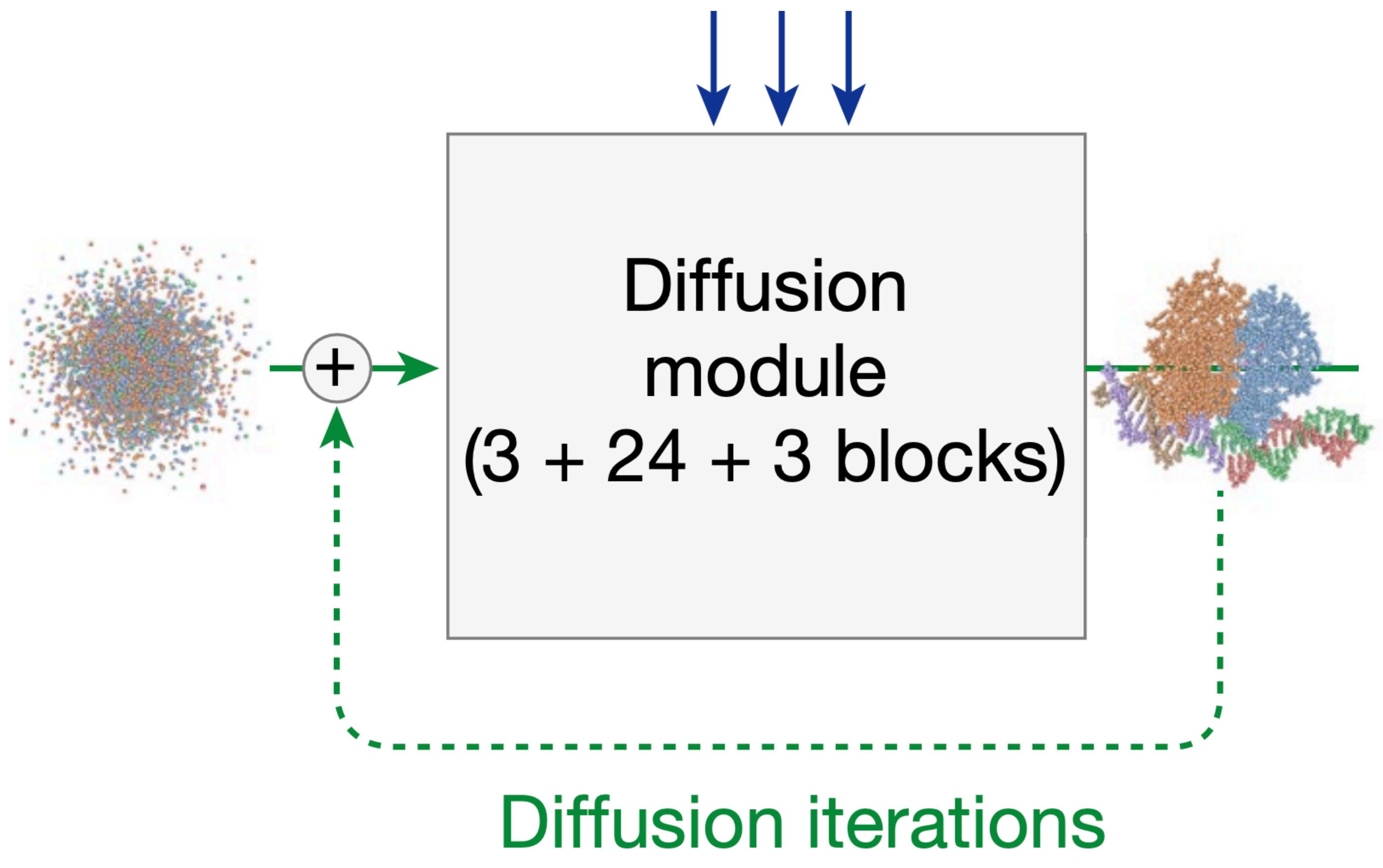


Dall-E

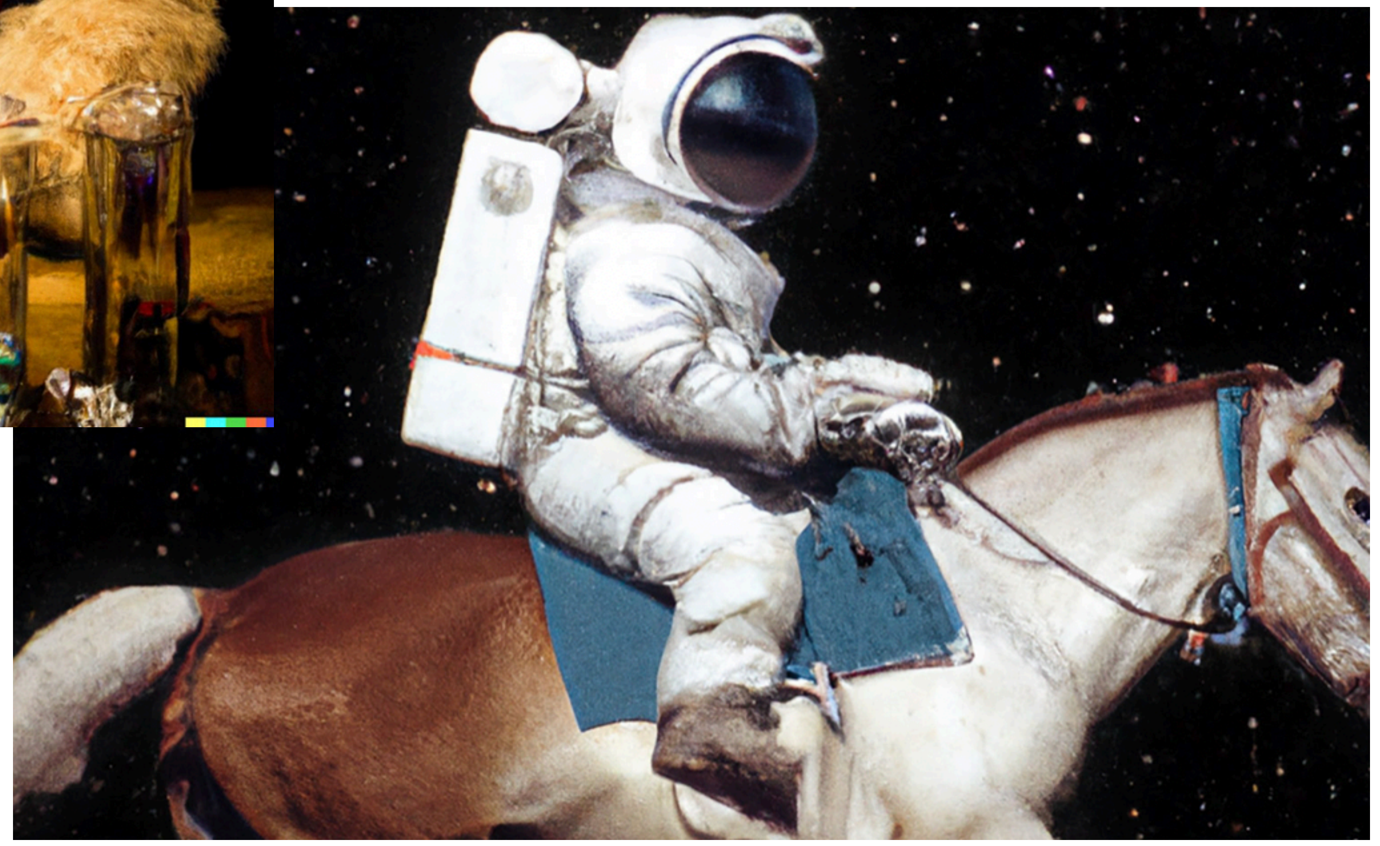
Image generation
- Diffusion models

Recent progress generative AI





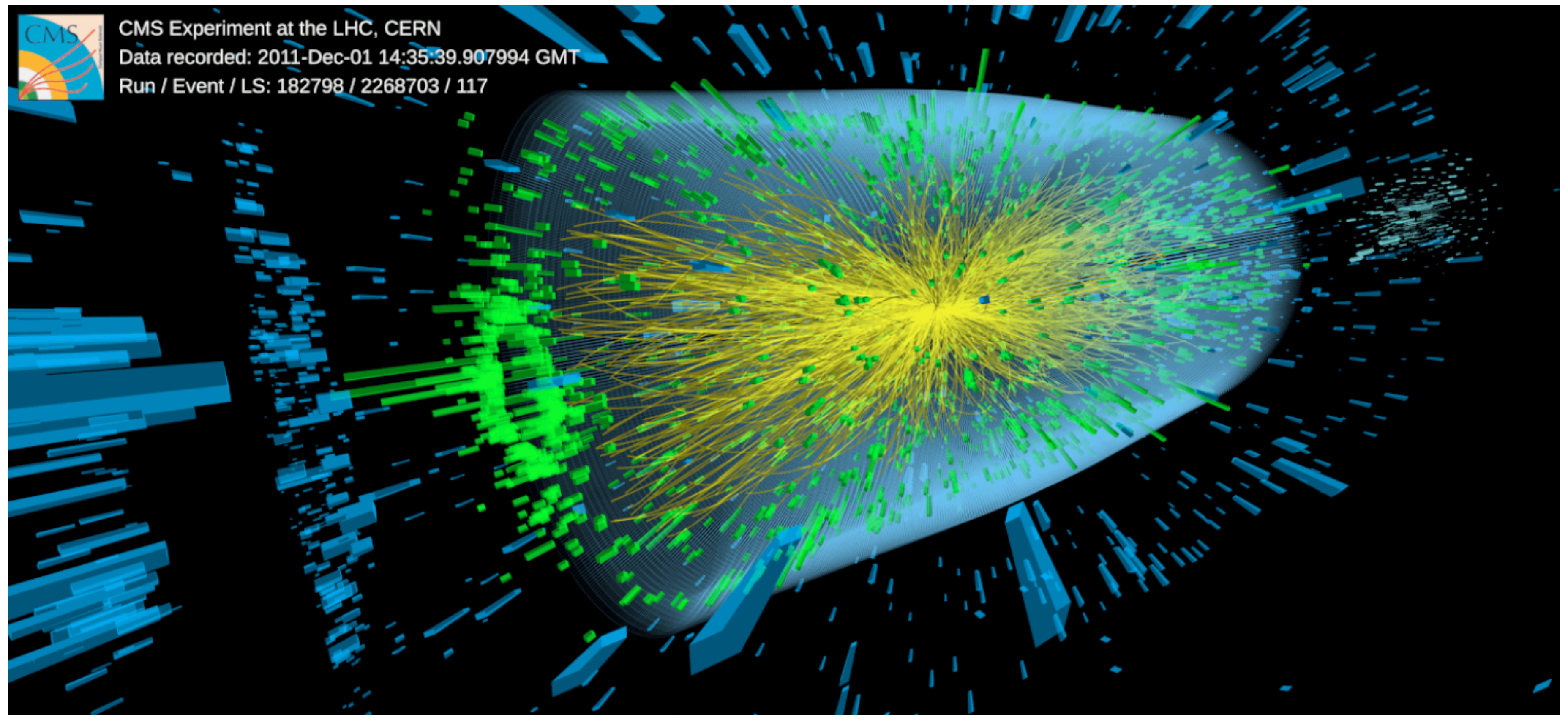
Diffusion-based architecture in AlphaFold 3 *Abramson `24*



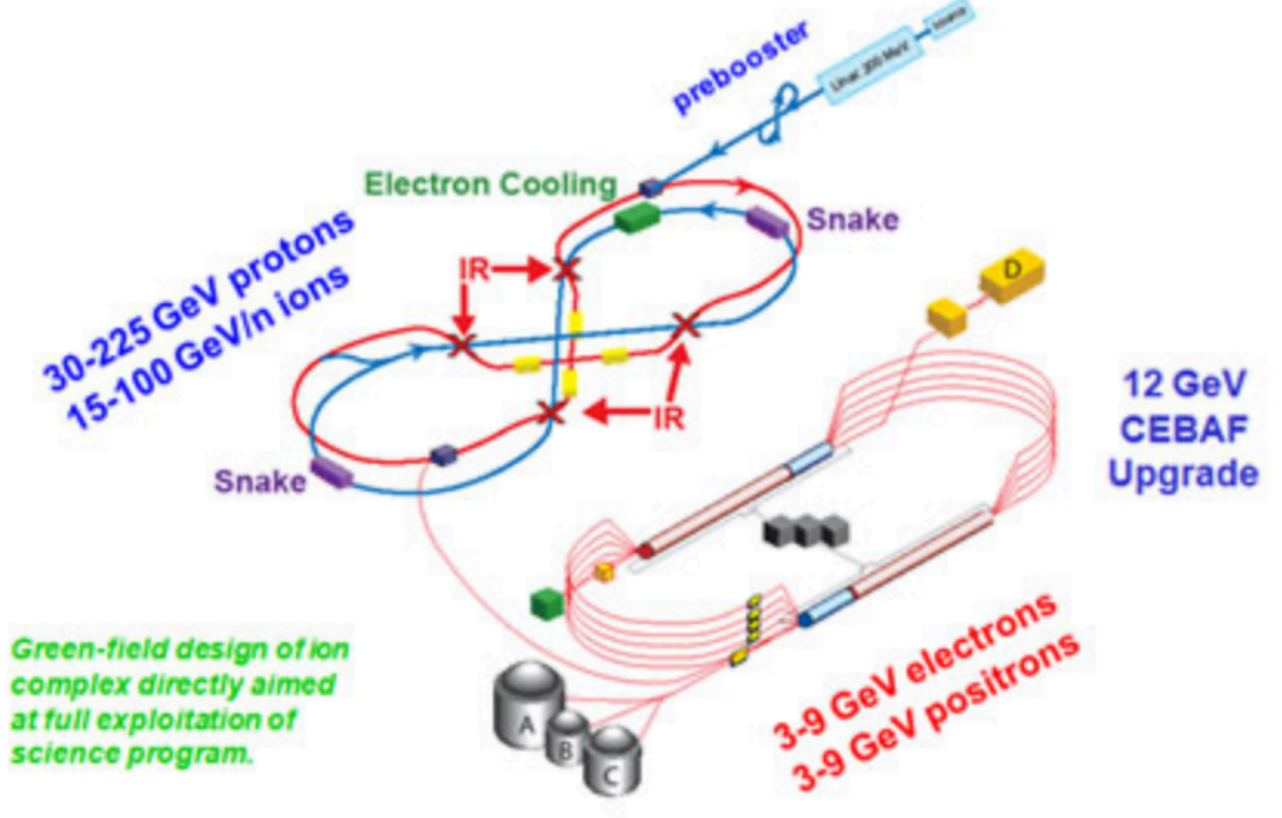
Dall-E

→ Collider simulations?

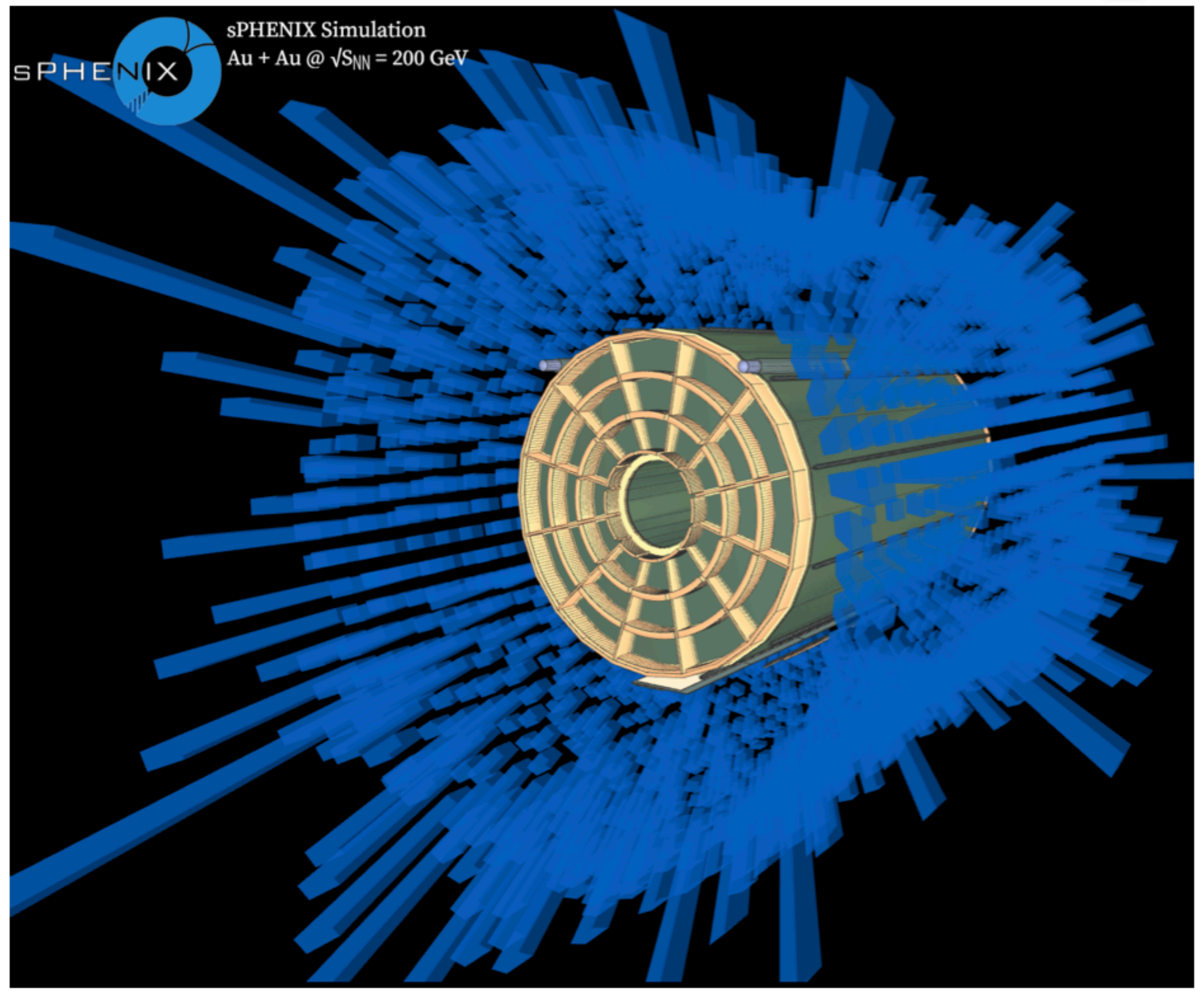
High-energy particle collisions



Large Hadron Collider

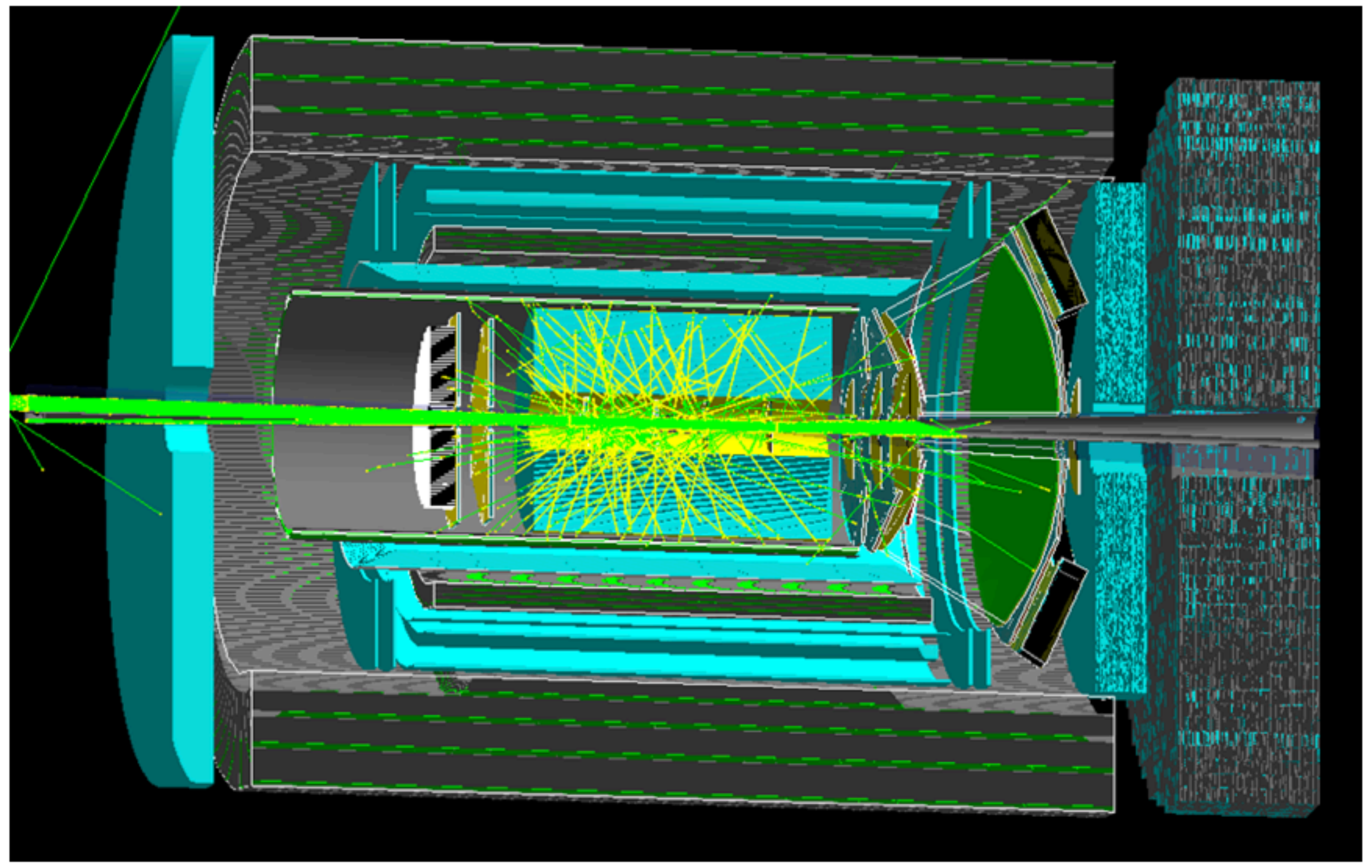


Jefferson Lab



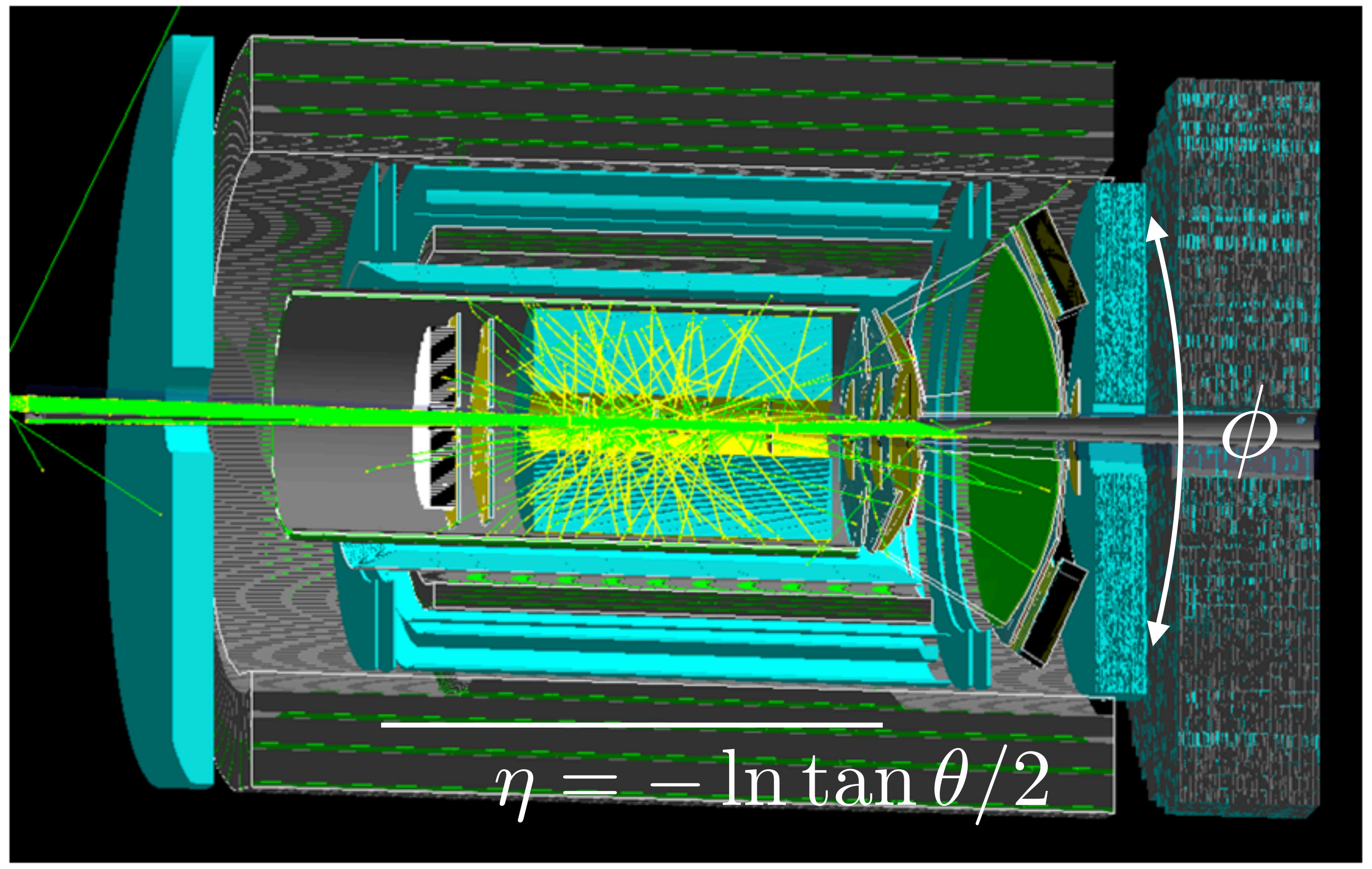
RHIC, sPHENIX

Few to 1000s particles/event



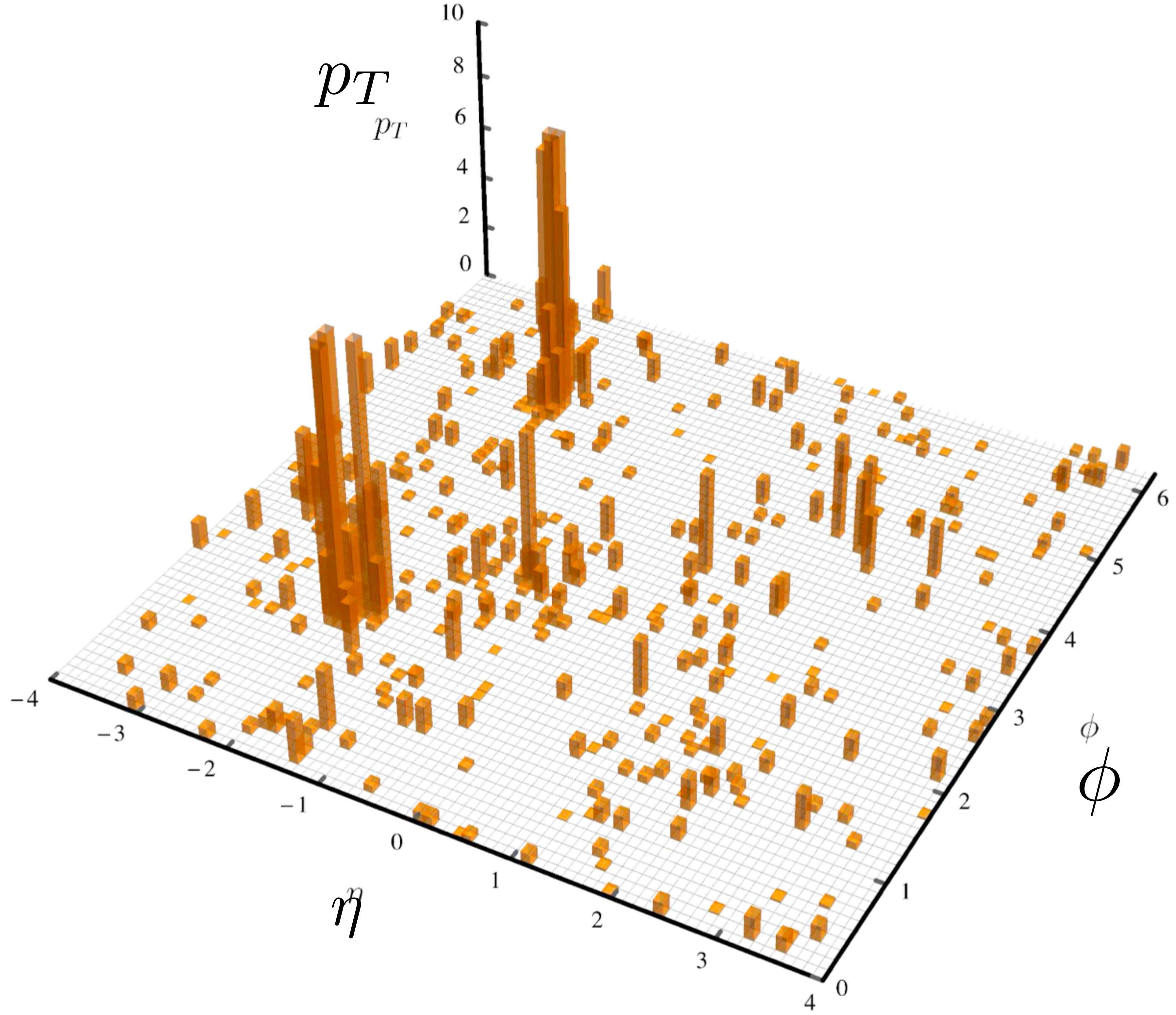
Electron-Ion Collider

High-energy particle collisions

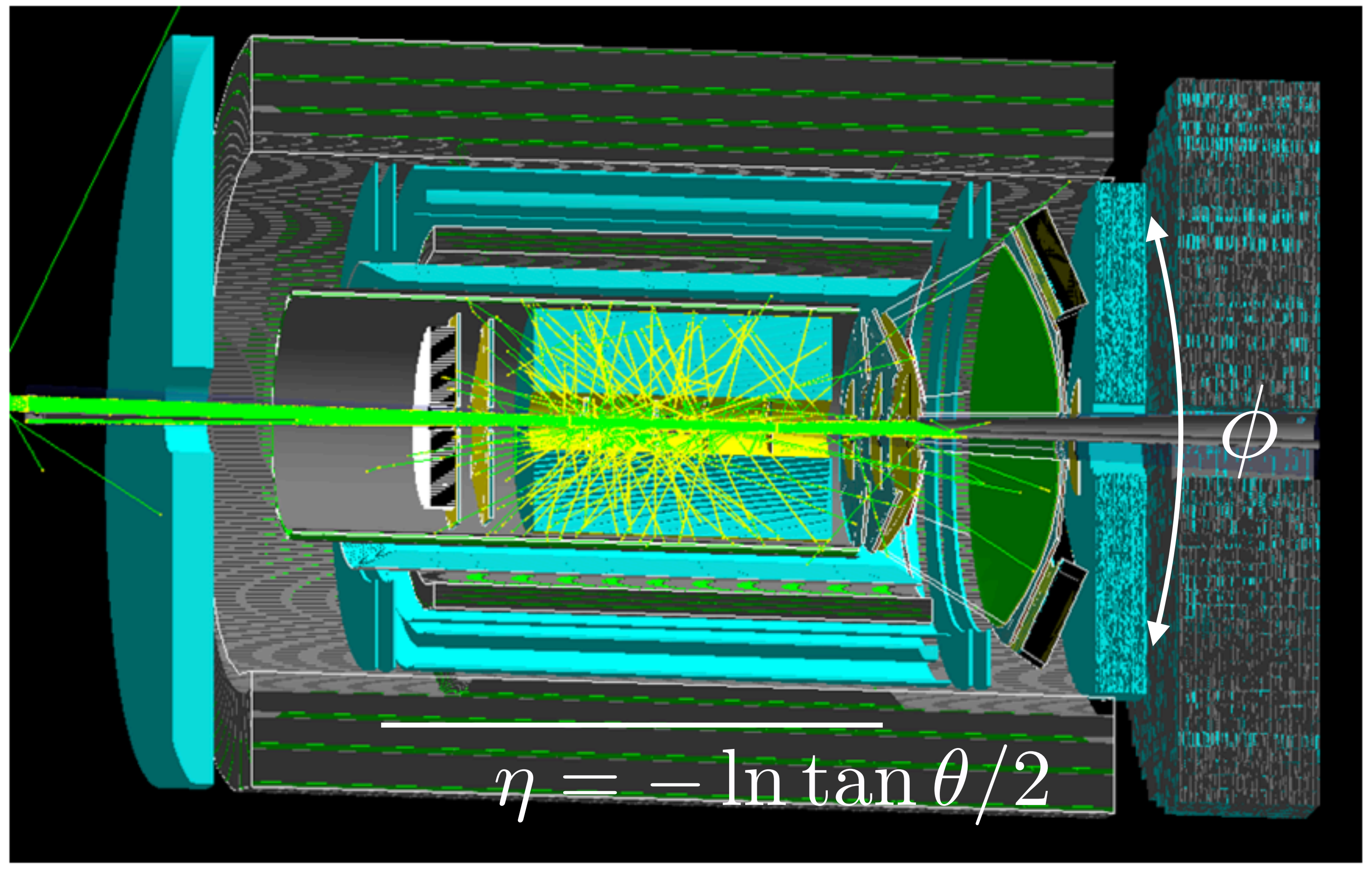


Electron-Ion Collider

Few to several 10s of particles



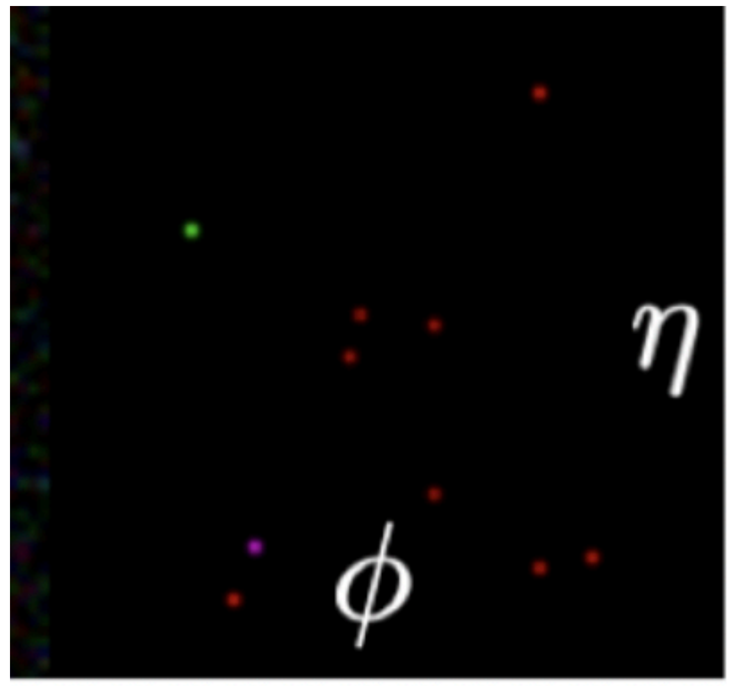
High-energy particle collisions



Electron-Ion Collider

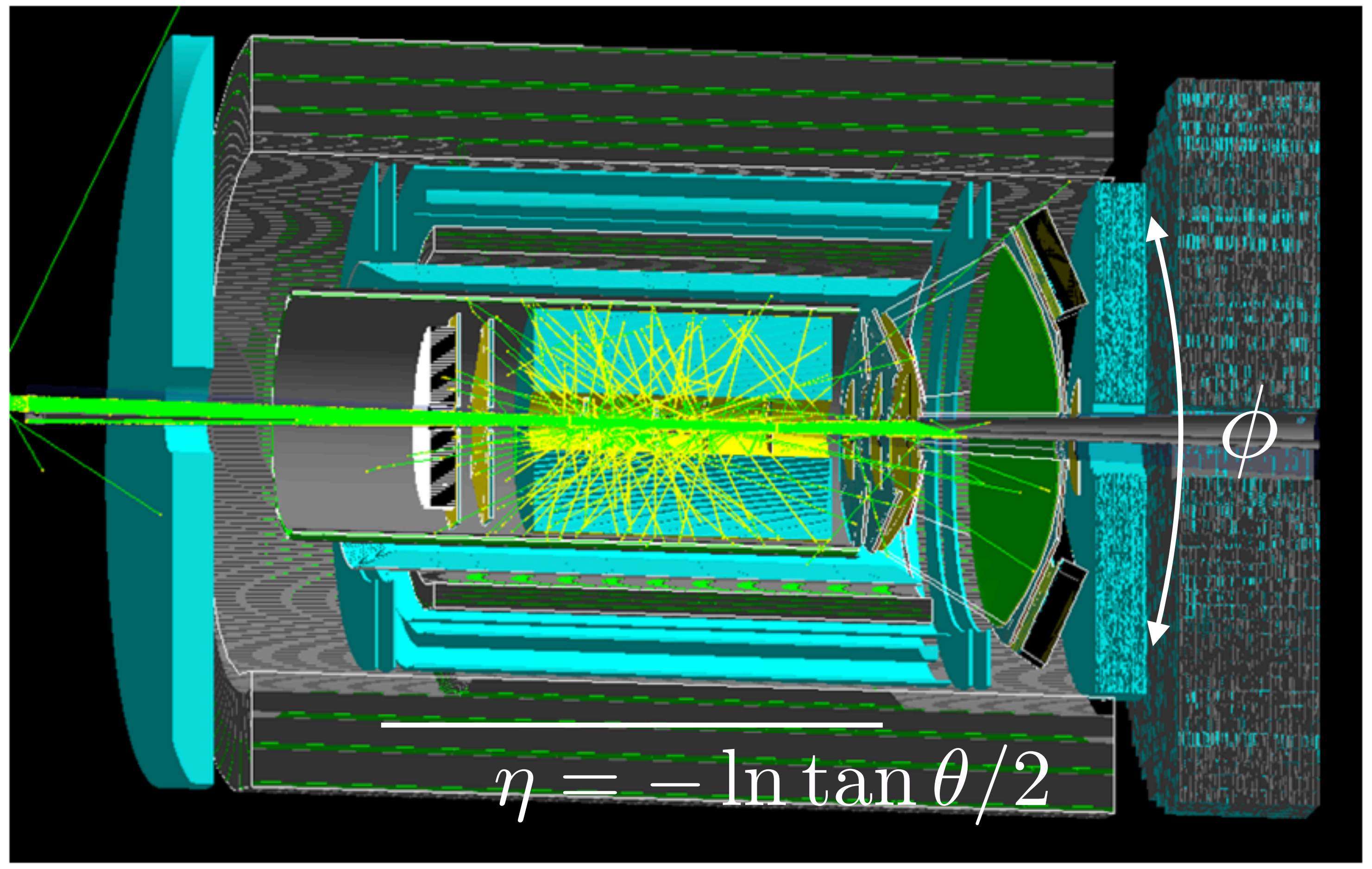
Data representations examples

- Images



May seem natural, pixels
analogous to detector granularity

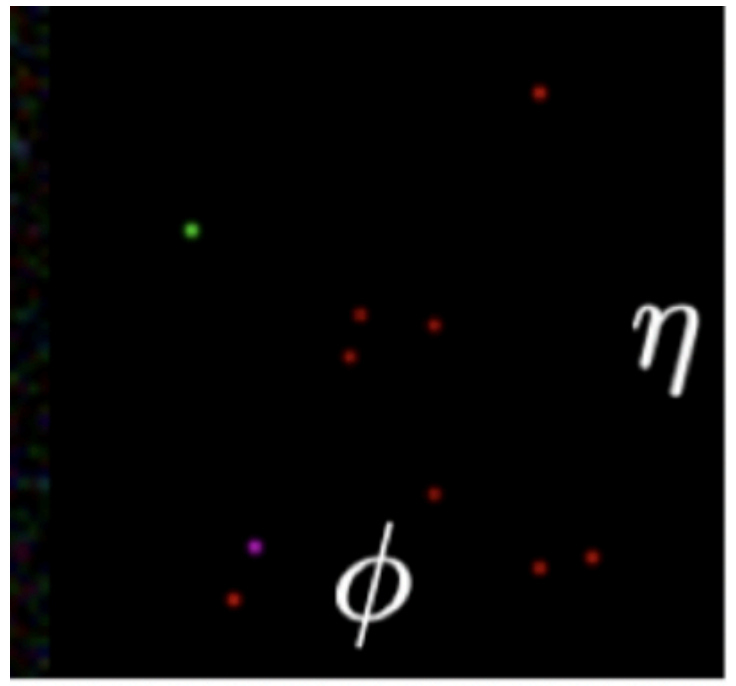
High-energy particle collisions



Data representations examples

- Images

EIC

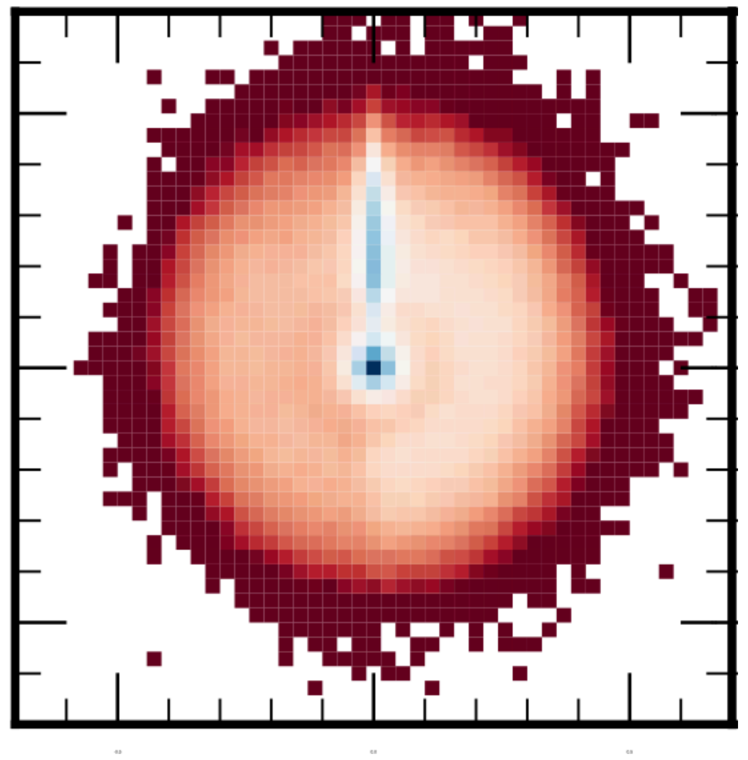


Sparse

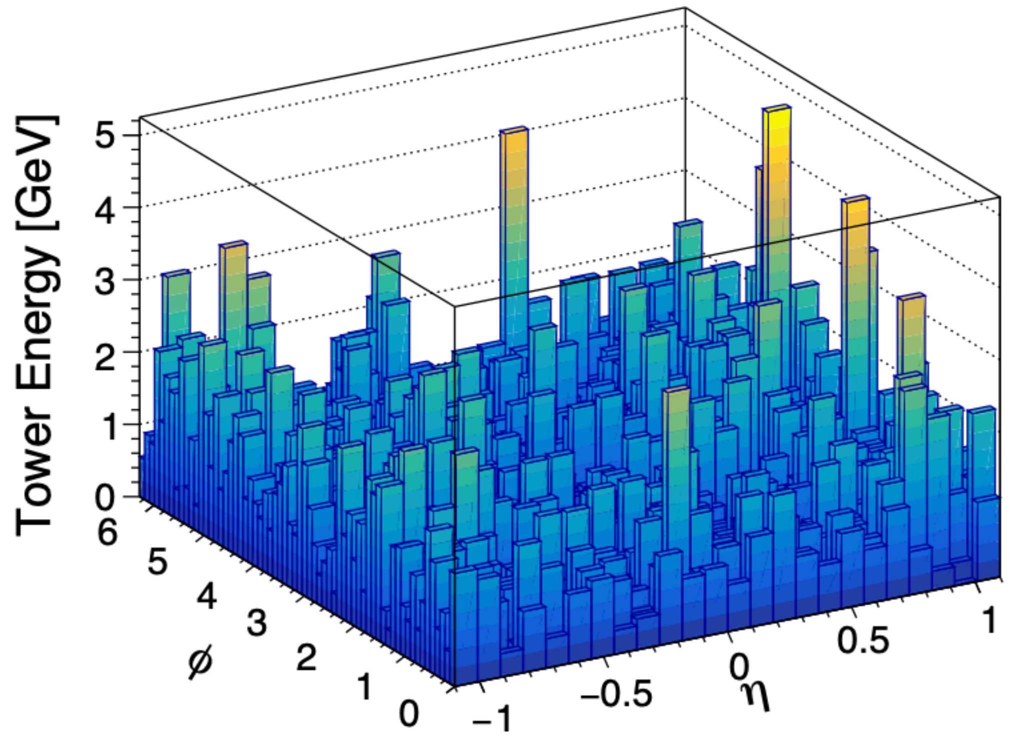
Devlin, Qiu, FR, Sato `23

Jet (averaged)

Mikuni, Nachman, Pettee `23

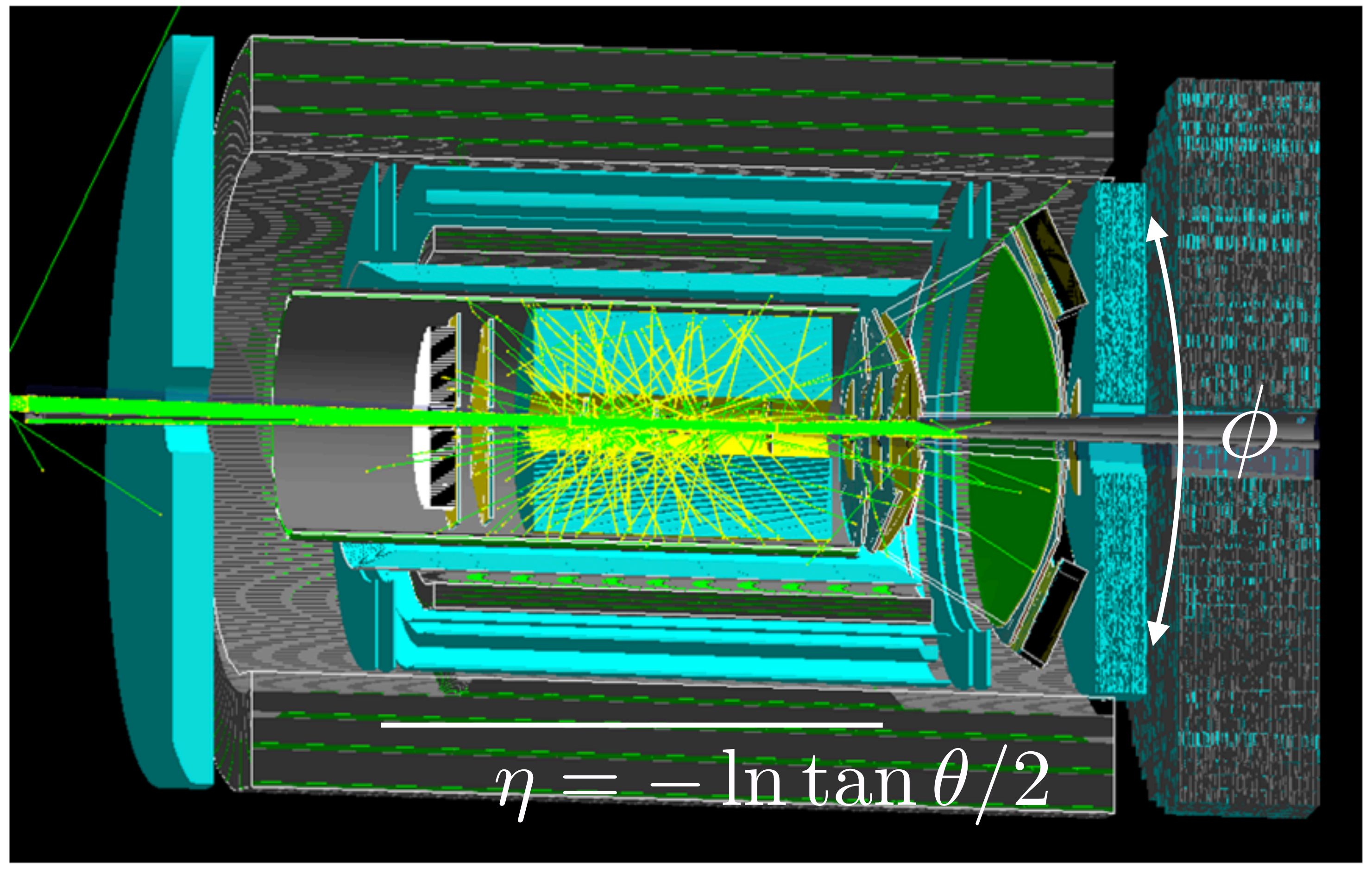


sPHENIX



Torbunov, Huang, Lin, Ren, Go, Rinn, Yu, Viren, Huang `24

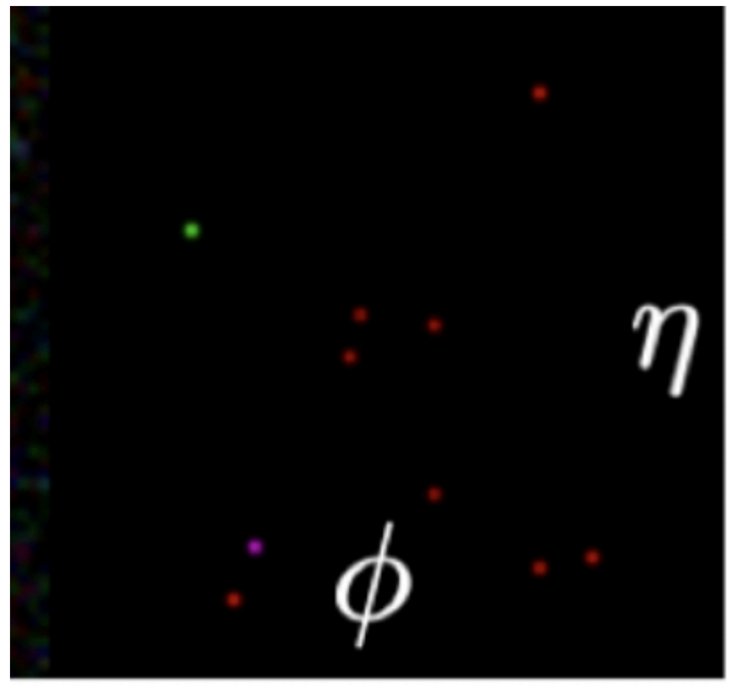
High-energy particle collisions



Data representations examples

- Images

EIC



Sparse

Devlin, Qiu, FR, Sato `23

- Point clouds

$$\text{Event}_1: \; (p_{T1}, \eta_1, \phi_1, \text{PID}_1), \; (p_{T2}, \eta_2, \phi_2, \text{PID}_2), \ldots, \; (p_{Tn}, \eta_n, \phi_n, \text{PID}_n)$$

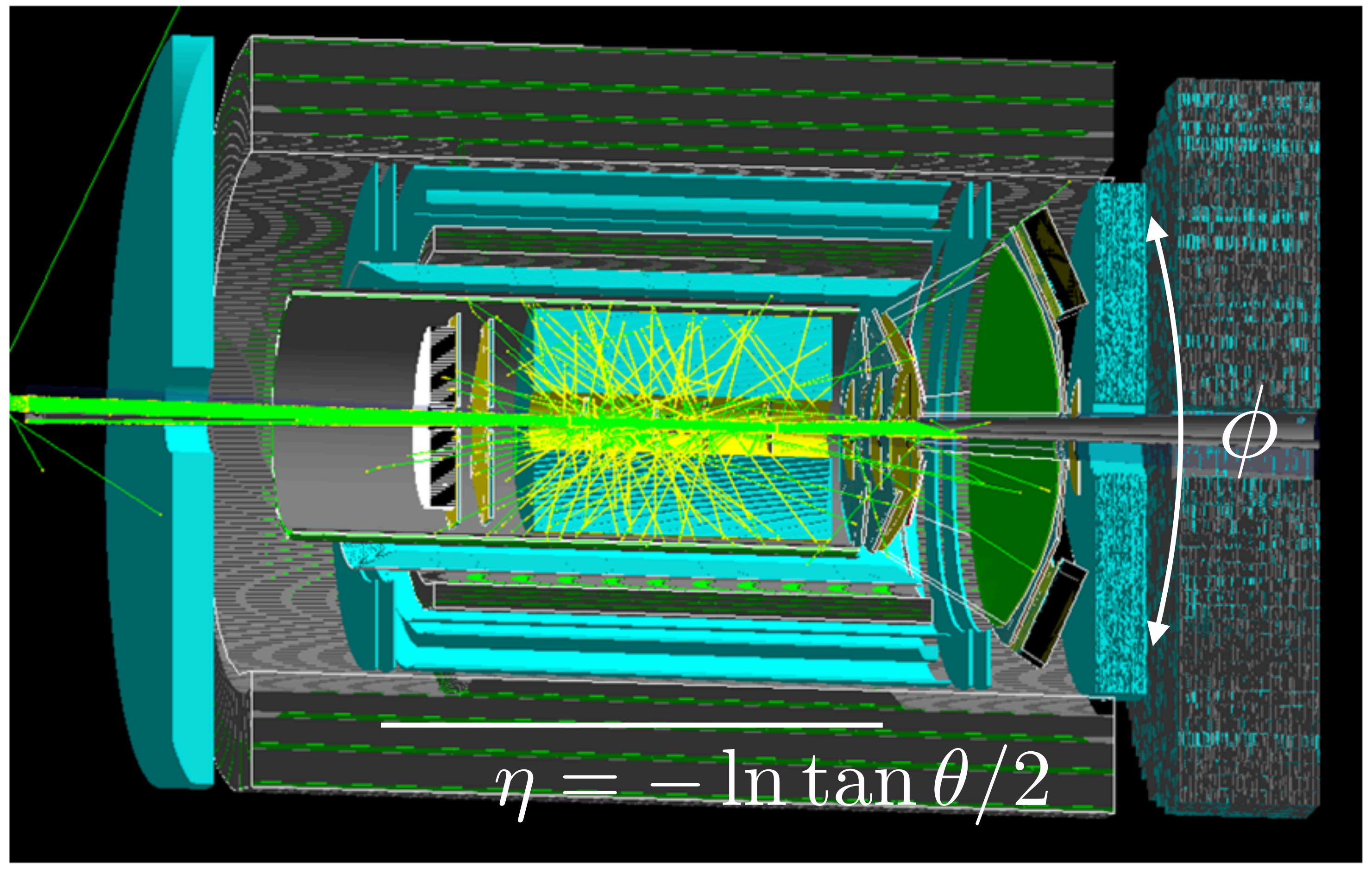
Variable length

$$\text{Event}_2: \; (p_{T1}, \eta_1, \phi_1, \text{PID}_1), \ldots, \; (p_{Tm}, \eta_m, \phi_m, \text{PID}_m)$$

$$\vdots$$

Torales Acosta, Nachman, Mikuni et al, Kasieczka, Thaler et al.,

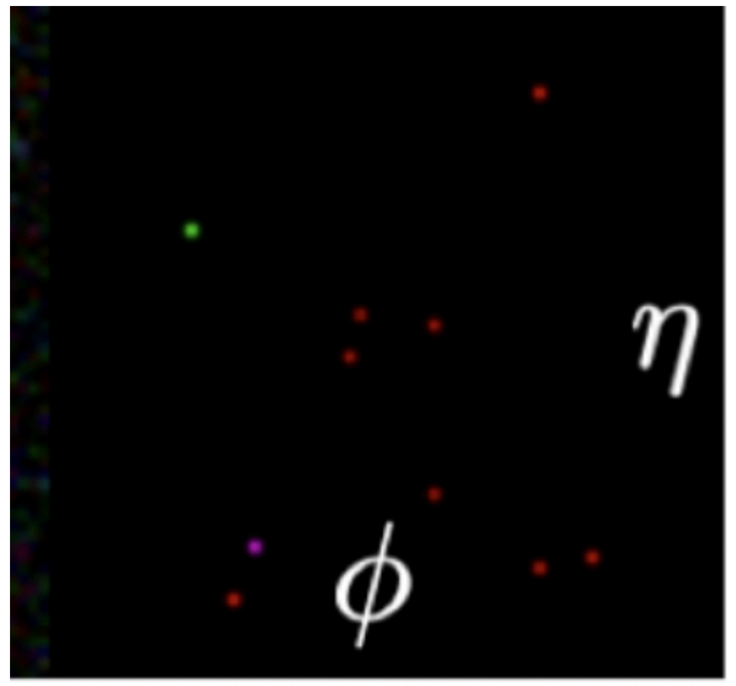
High-energy particle collisions



Data representations examples

- Images

EIC



Sparse

Devlin, Qiu, FR, Sato `23

- Point clouds

$$\text{Event}_1: \; (p_{T1}, \eta_1, \phi_1, \text{PID}_1), \; (p_{T2}, \eta_2, \phi_2, \text{PID}_2), \ldots, \; (p_{Tn}, \eta_n, \phi_n, \text{PID}_n)$$

Variable length

$$\text{Event}_2: \; (p_{T1}, \eta_1, \phi_1, \text{PID}_1), \ldots, \; (p_{Tm}, \eta_m, \phi_m, \text{PID}_m)$$

See Vinicius, Fernando's talk

Torales Acosta, Nachman, Mikuni et al, Kasieczka, Thaler et al.,

Generative modeling

- Generative Adversarial Networks
- Variational Autoencoders
- Normalizing Flows
- Diffusion Models

High quality samples

Mode coverage

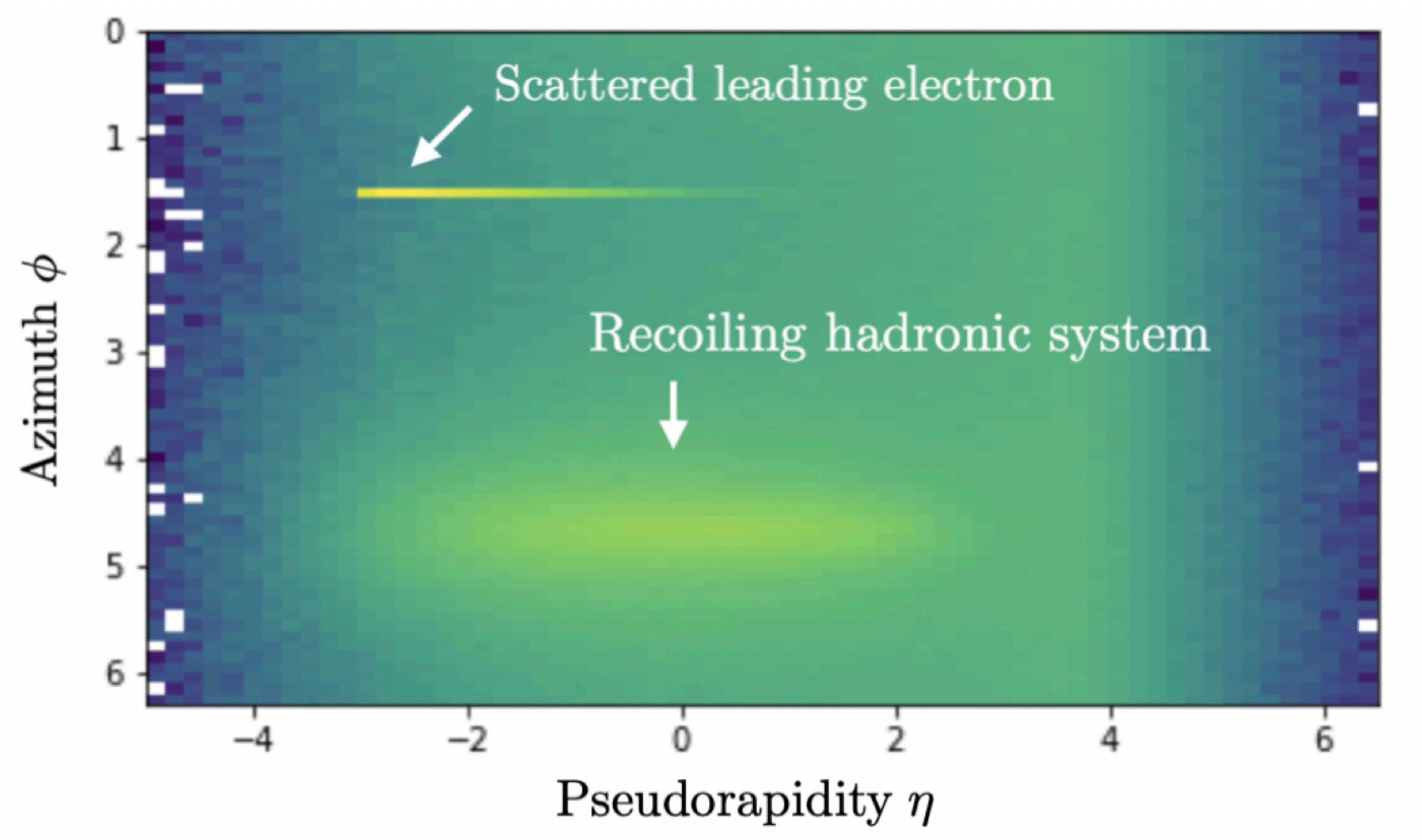
Fast sampling

Stable training

Generative models for EIC events

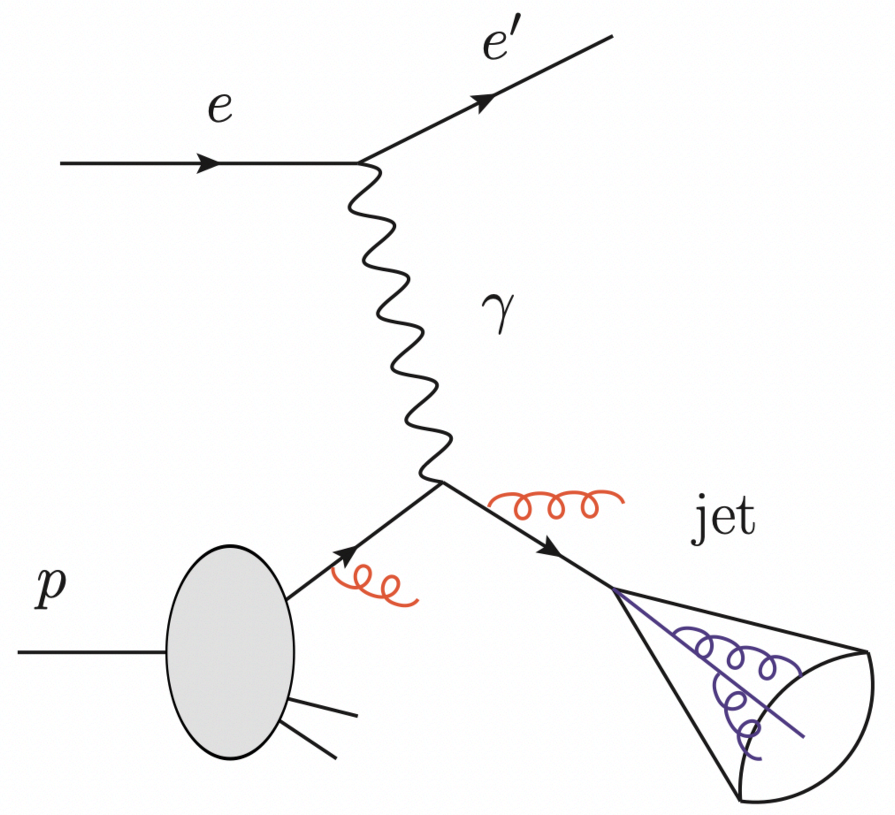
- Surrogate models
- Searches of physics beyond the Standard Model
- Event-level data analysis (differentiable)
- Development of MC event generators

GANs e.g. Lai, Neill, Ploskon, FR `20



- Full ep events at $\sqrt{s} = 105$ GeV

→ Develop a generative model



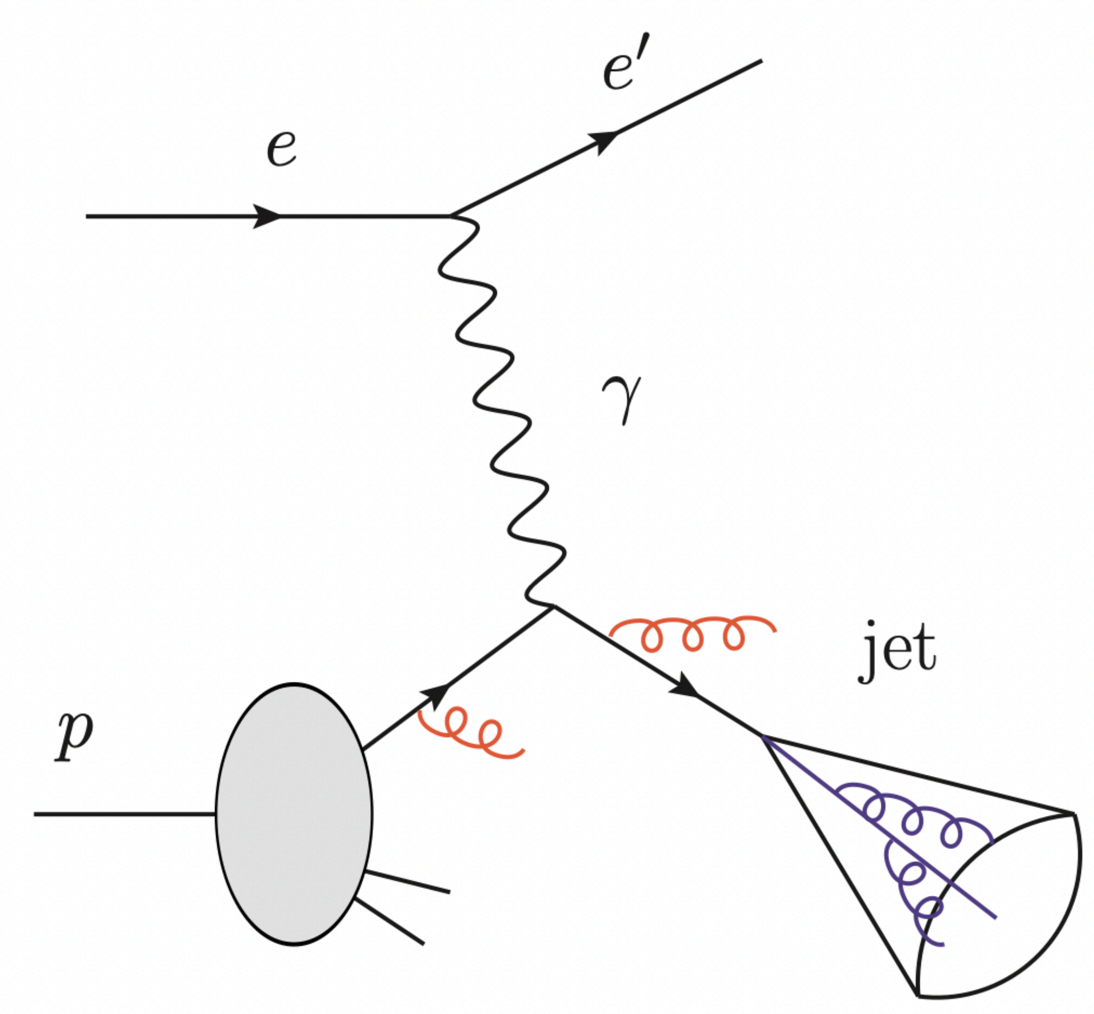
$\textsc{Pythia}8$, $Q > 5$ GeV

w/o photoproduction & CC

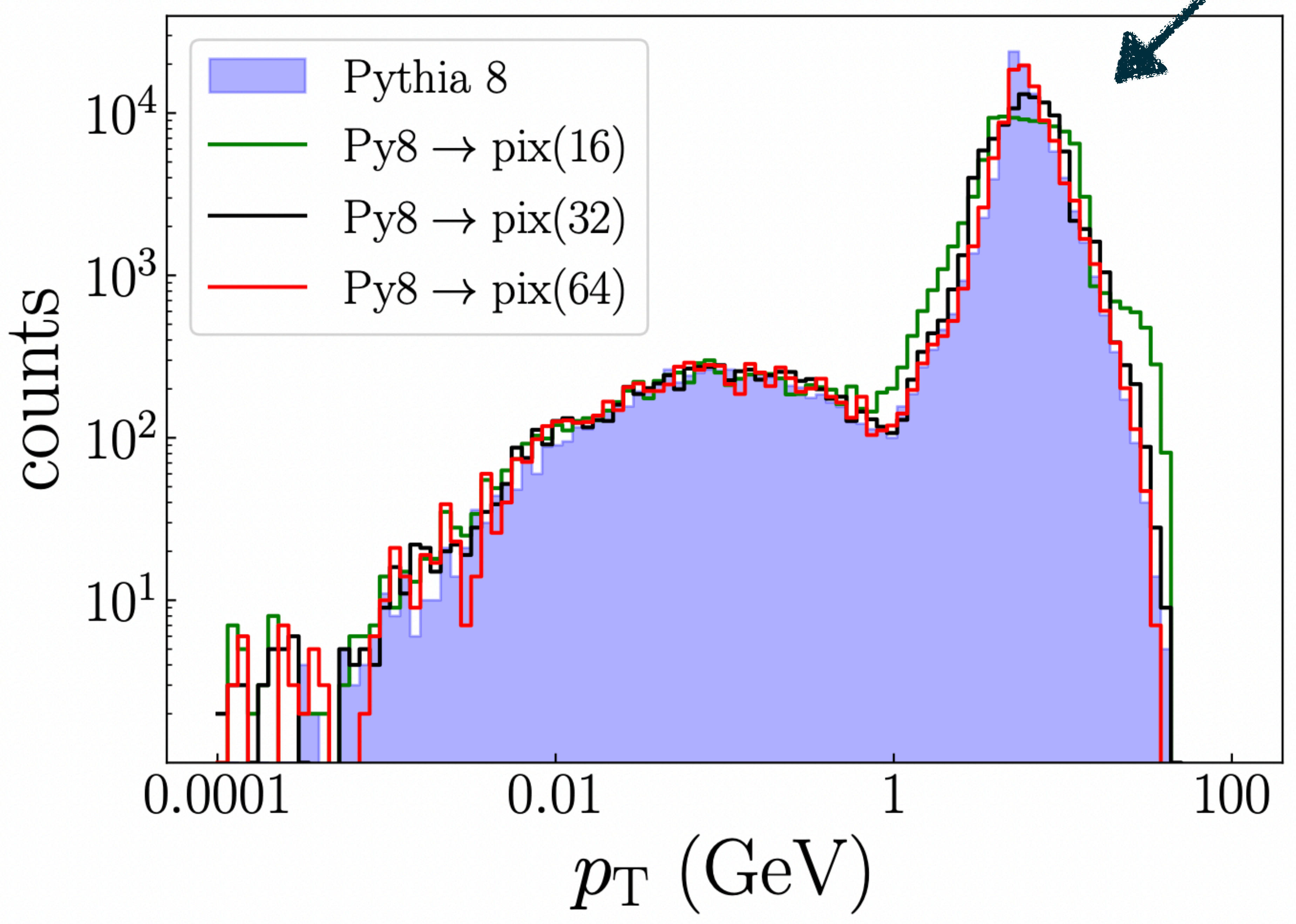
Devlin, Qiu, FR, Sato `23
see also Mikuni, Nachman et al.

Pixelation effects

- Example: Electron distribution



Scattered leading electron



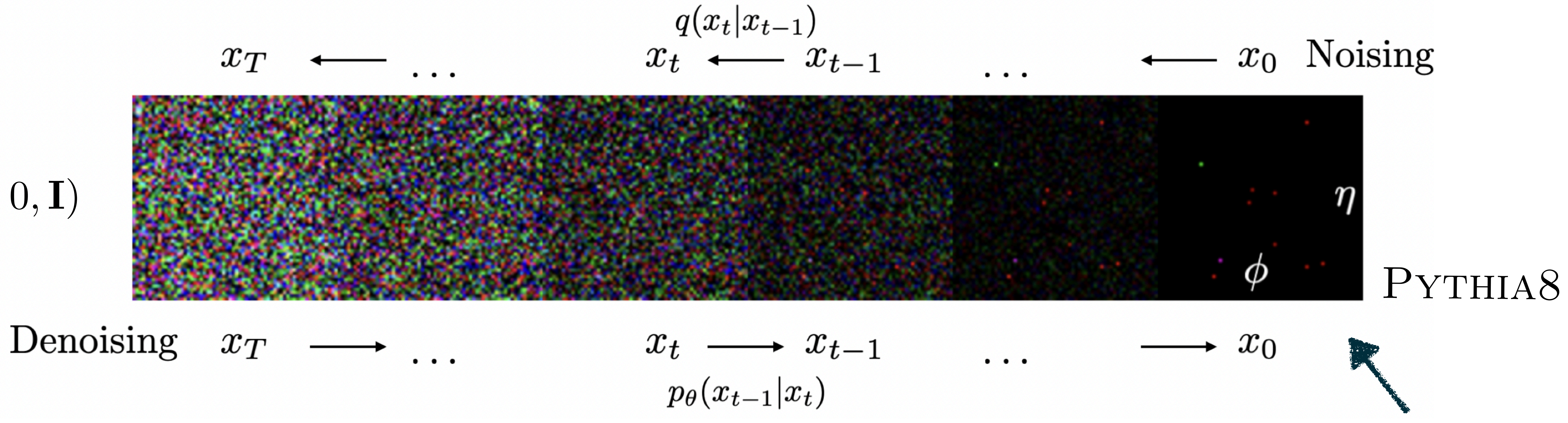
Devlin, Qiu, FR, Sato `23

Diffusion models

Devlin, Qiu, FR, Sato `23

- Represent events as images (pixelated)

$$q(x_T) = \mathcal{N}\left(x_T; 0, \mathbf{I}\right)$$



π^+, e^-, K^+ (RGB)

- Markovian noising process

$$q(x_1, \dots, x_T|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1})$$

  adding Gaussian noise

$$q\left(x_t|x_{t-1}\right) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t\mathbf{I})$$

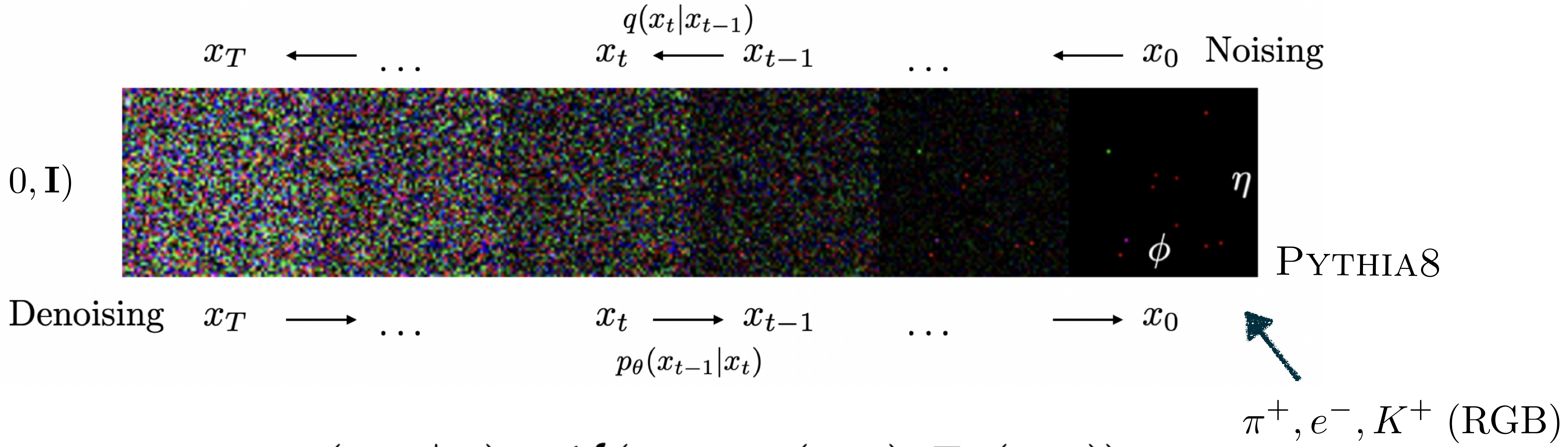
Sohl-Dickstein et al.`15
Ho, Jain, Abbeel `20
Nichol, Dhariwal `21

Diffusion models

Devlin, Qiu, FR, Sato `23

- Represent events as images (pixelated)

$$q(x_T) = \mathcal{N}\left(x_T; 0, \mathbf{I}\right)$$



- Learn denoising process $$p_\theta\left(x_{t-1}|x_t\right) = \mathcal{N}\left(x_{t-1}; \mu_\theta\left(x_t, t\right), \Sigma_\theta\left(x_t, t\right)\right)$$

Stochastic differential equation in continuum limit

Train convolutional U-Net

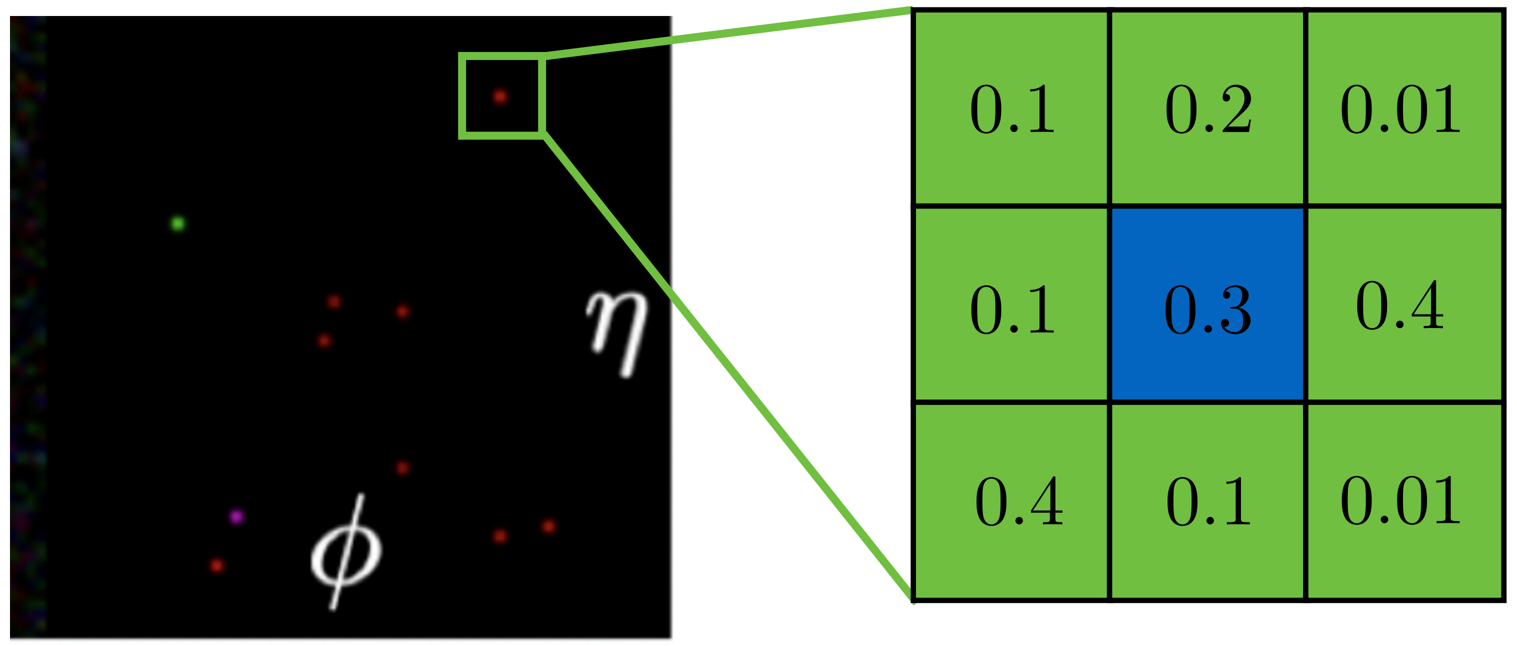
Sohl-Dickstein et al.`15
Ho, Jain, Abbeel `20
Nichol, Dhariwal `21

Diffusion models for the EIC

Devlin, Qiu, FR, Sato `23

- Sparse collider data

  → Need to generate empty & isolated colored pixels



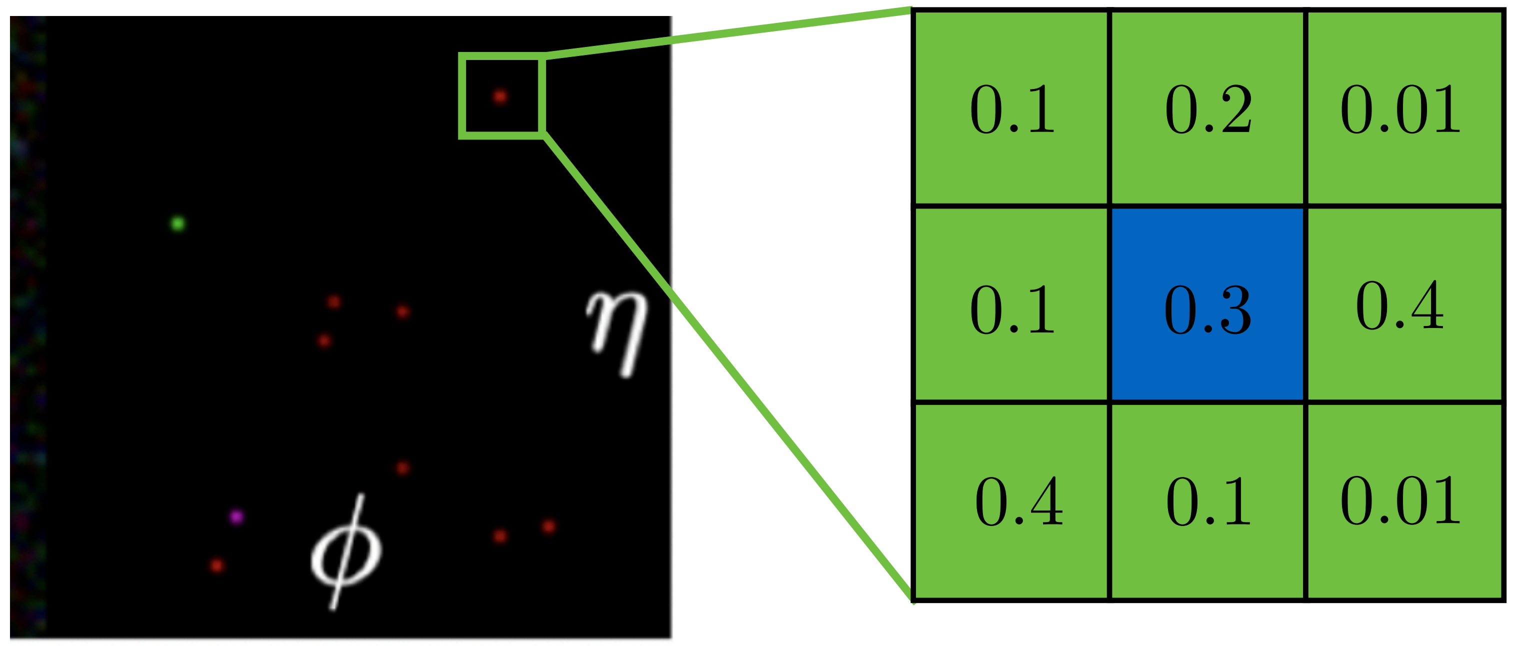
- Empty pixels for the simulated EIC events $99.95 \pm 0.02\%$ for 64×64 pixels

Diffusion models for the EIC

Devlin, Qiu, FR, Sato `23

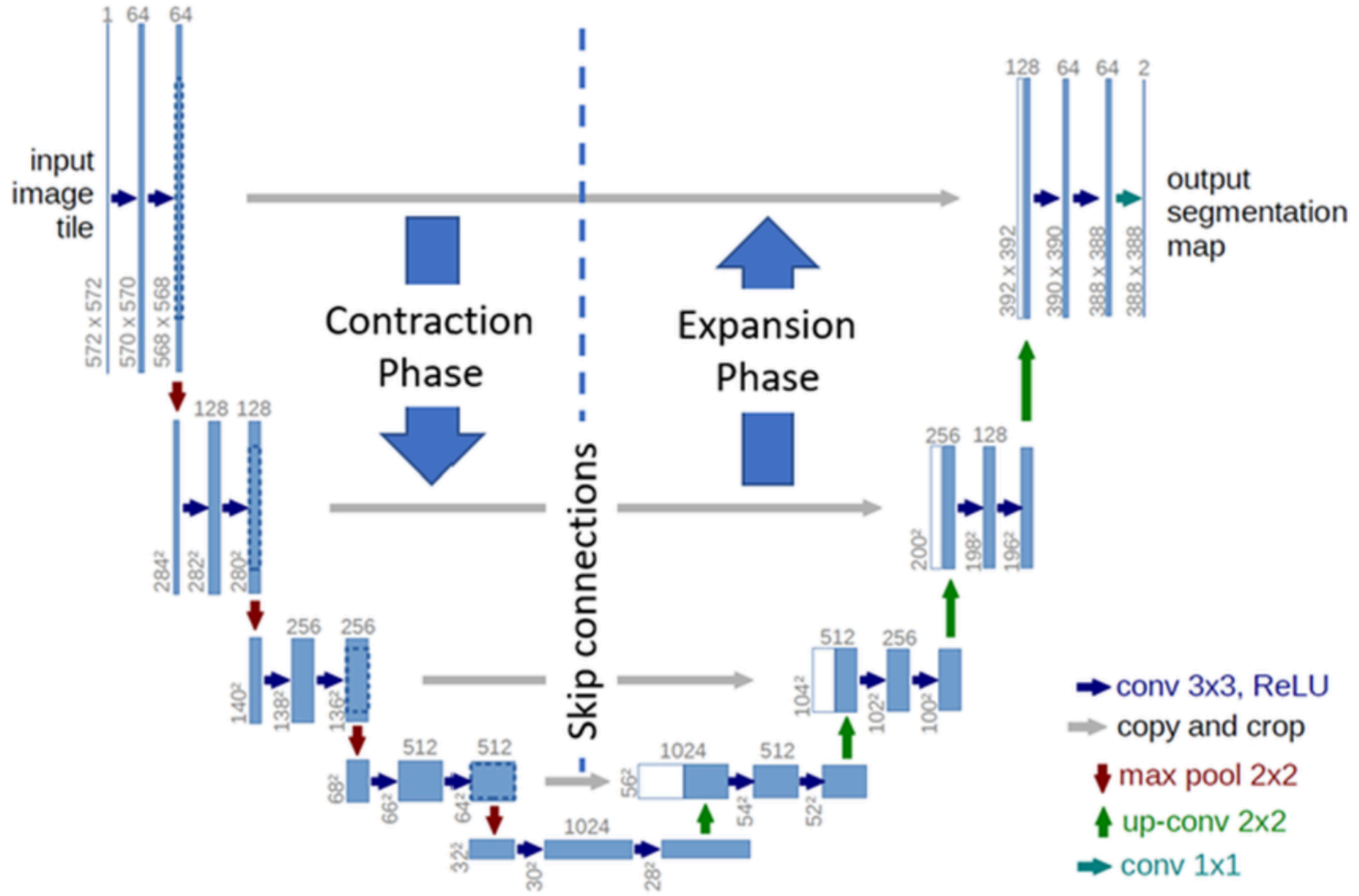
- Sparse collider data

  Need to generate empty & isolated colored pixels



- U-Net
  - 3 layers
  - Circular/periodic convolutional network

Ronneberger et al. `15



(schematic)

Diffusion models for the EIC

Devlin, Qiu, FR, Sato `23

- Steeply falling distributions

Use rescaled momentum variable $$\tilde{z}_i = \frac{2M_{Ti}}{\sqrt{s}} \cosh y_i$$ with $$M_{Ti}^2 = p_{Ti}^2 + m_i^2$$

$$0 \le \tilde{z}_i \le 1$$

Massless limit $$\tilde{z}_i = \frac{2p_{Ti}}{\sqrt{s}} \cosh \eta_i$$

Diffusion models for the EIC

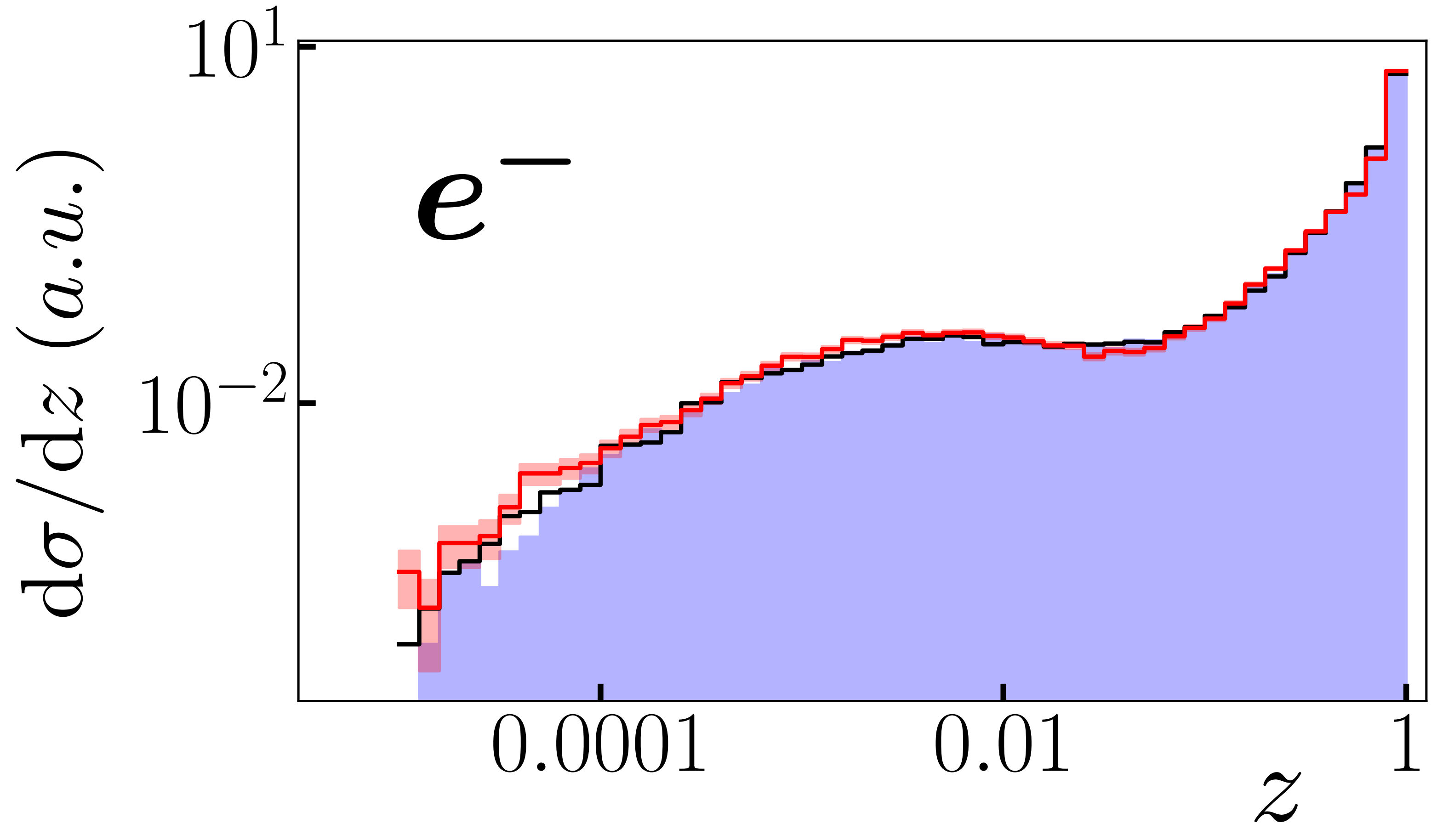
Devlin, Qiu, FR, Sato `23

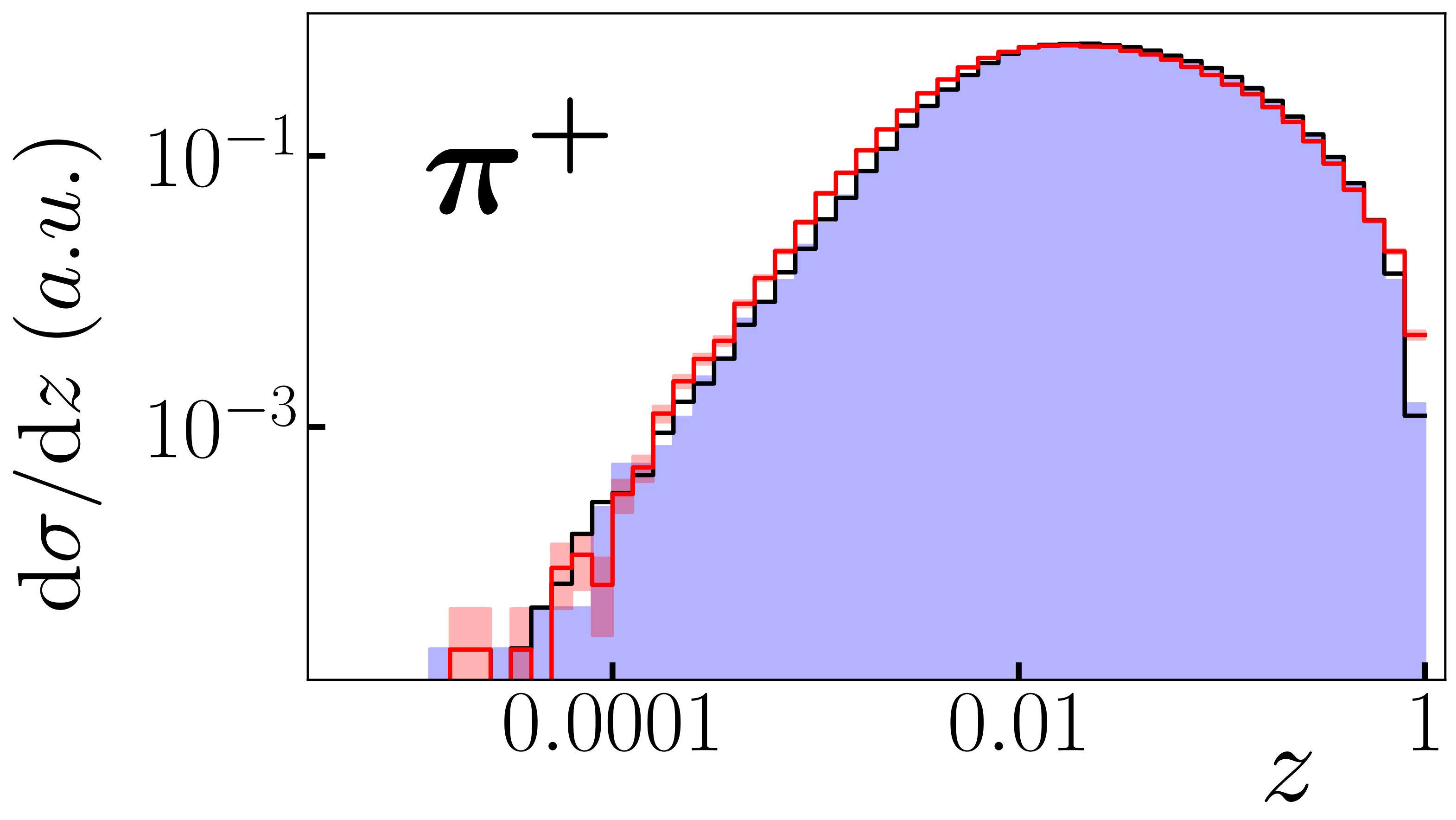
- Steeply falling distributions

Use rescaled momentum variable $\tilde{z}_i = \frac{2M_{Ti}}{\sqrt{s}} \cosh y_i$ with $M_{Ti}^2 = p_{Ti}^2 + m_i^2$

$0 \leq \tilde{z}_i \leq 1$

For example





Large-z peak

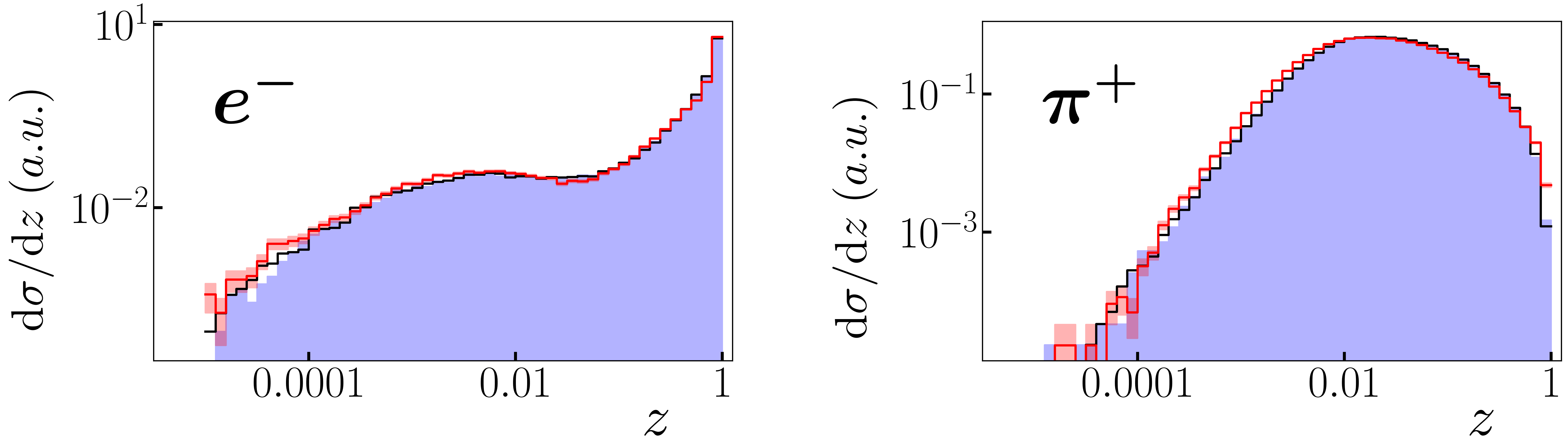
Small-z peak

Diffusion models for the EIC

Devlin, Qiu, FR, Sato `23

- Steeply falling distributions

For example



- Use suitable 1-to-1 remapping

(same for all particles)

$$z \to S(z)E(z) + (1 - S(z))L(z)$$

with

$$E(z) = a_1 e^{c_1 z} + b_1$$
$$L(z) = a_2 \ln(z + c_2) + b_2$$
$$S(z) = \frac{1}{1 + e^{-\alpha(z-\beta)}}$$

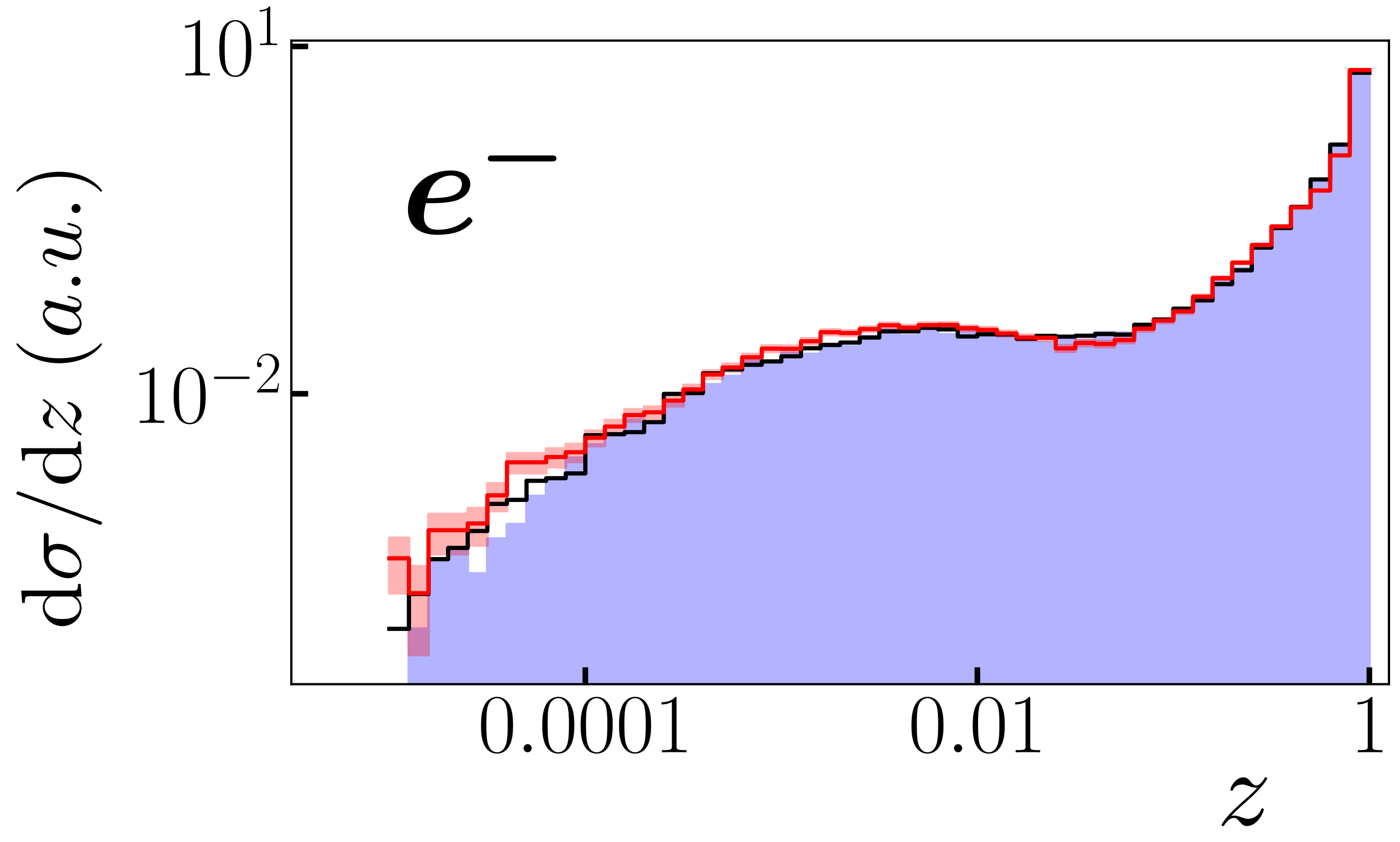
z

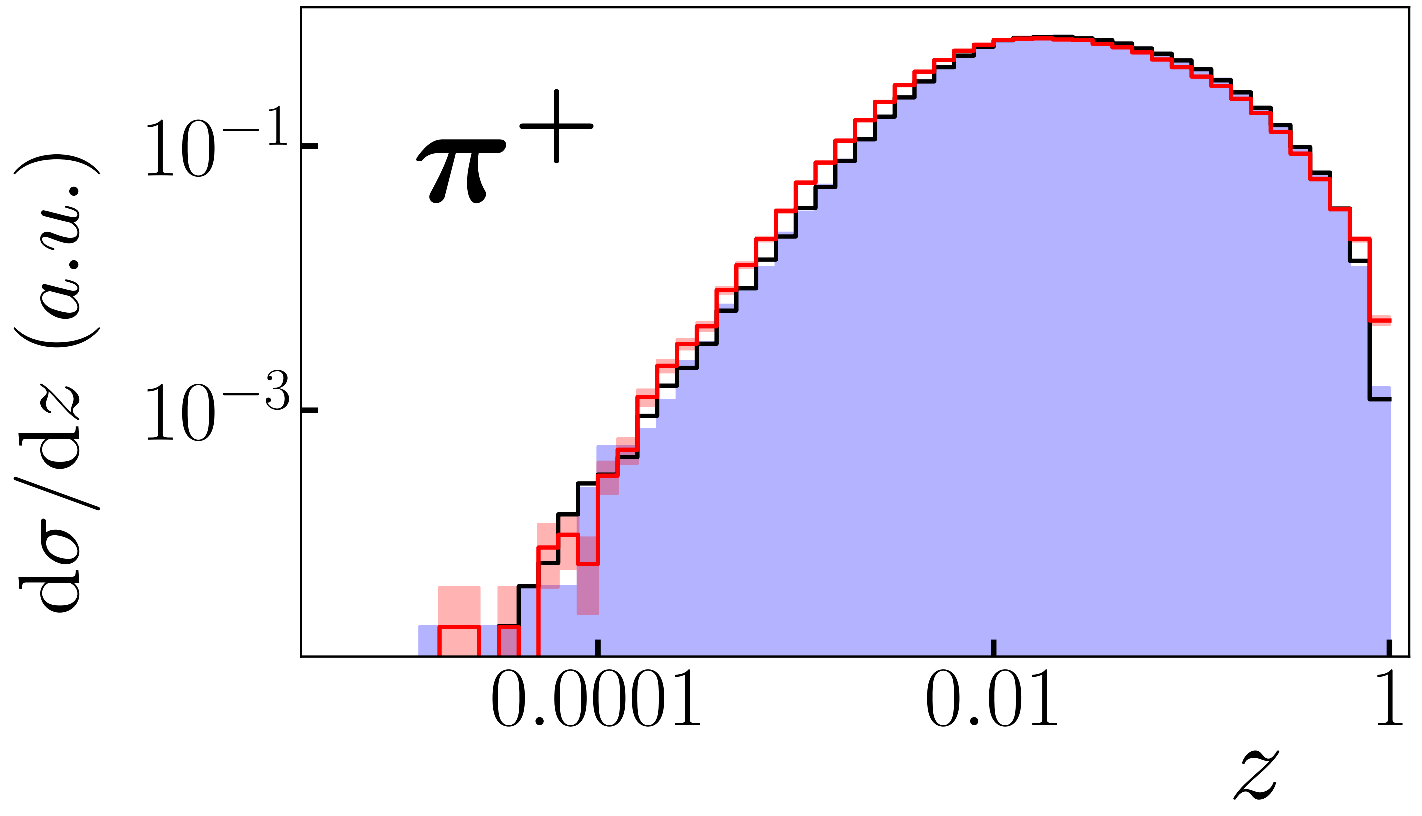
Diffusion models for the EIC

Devlin, Qiu, FR, Sato `23

- Steeply falling distributions

For example





- Use suitable 1-to-1 remapping

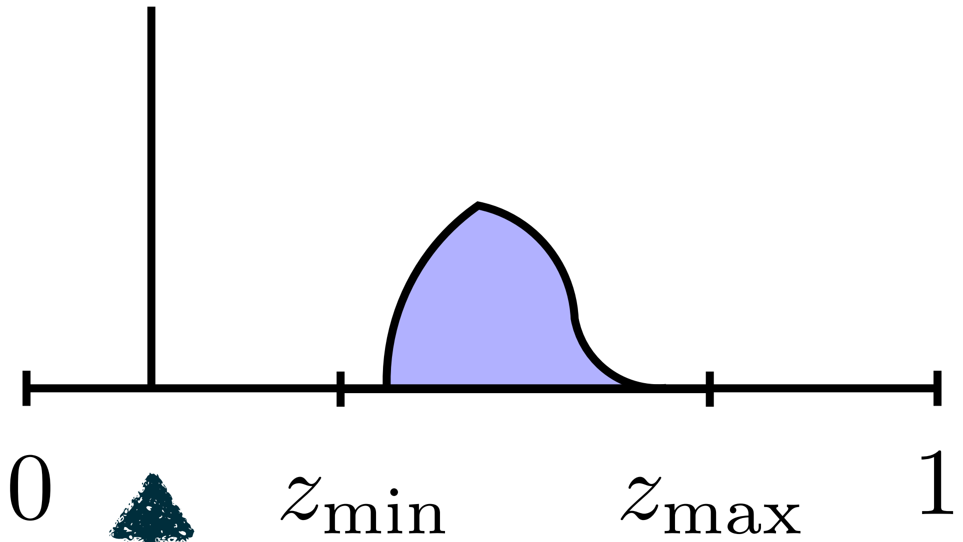
(same for all particles)

$$z \to S(z)E(z) + (1 - S(z))L(z)$$

with

$$E(z) = a_1 e^{c_1 z} + b_1$$
$$L(z) = a_2 \ln(z + c_2) + b_2$$
$$S(z) = \frac{1}{1 + e^{-\alpha(z-\beta)}}$$

linear map →



Empty pixels

Diffusion models for the EIC

Devlin, Qiu, FR, Sato `23

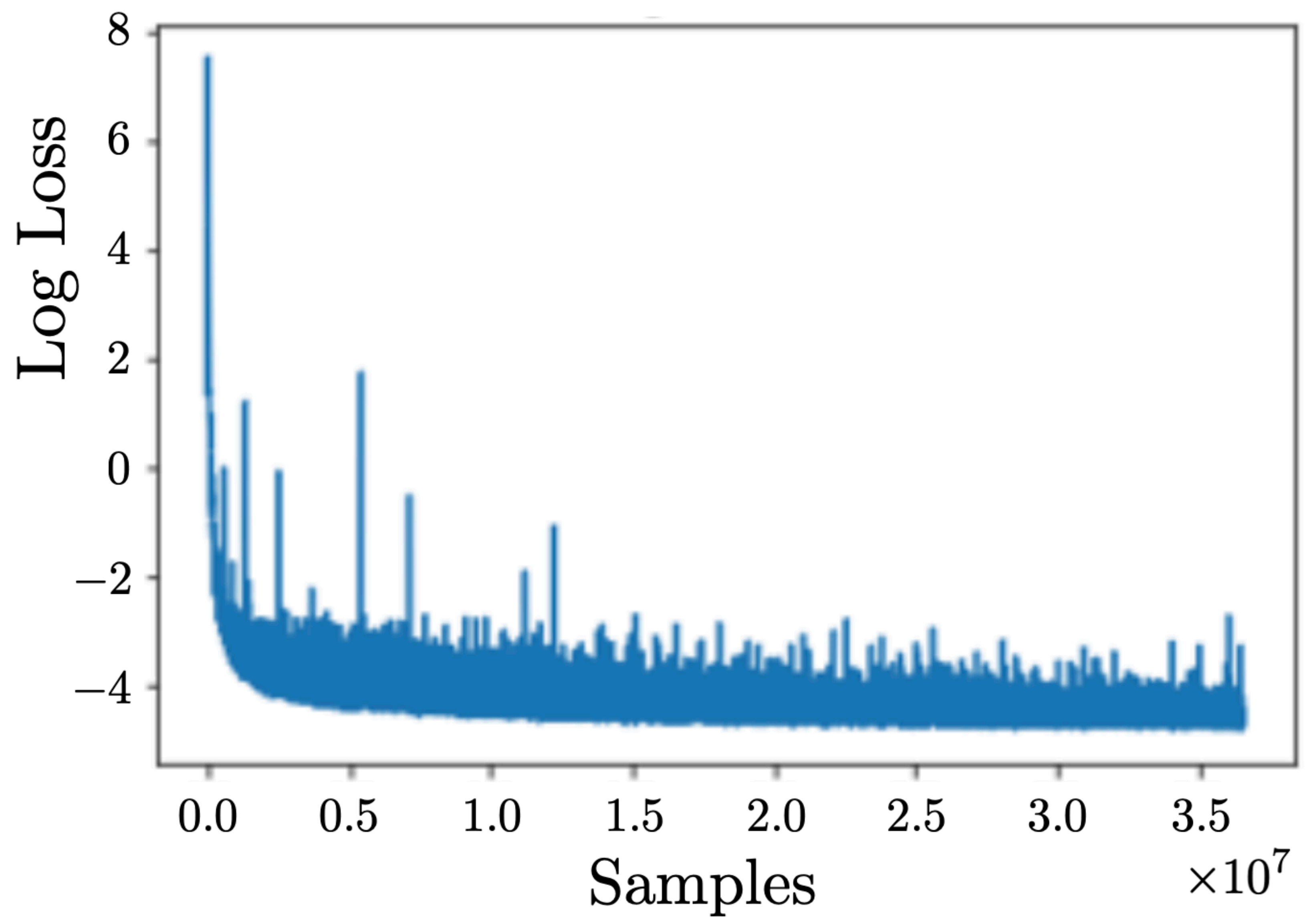
- Training procedure
  - 500 diffusion steps
  - Cosine variance schedule
  - Hybrid loss function
  - Training data set of 10^6 events

$$L_{\mathrm{VLB}} = L_0 + L_1 + \ldots + L_{T-1} + L_T$$

with $$L_{t-1} = D_{\mathrm{KL}}\left(q\left(x_{t-1} \mid x_t, x_0\right) \| p_\theta\left(x_{t-1} \mid x_t\right)\right)$$

$$L_{\text{hybrid}} = L_{\text{simple}} + \lambda L_{\mathrm{VLB}}$$

Training loss

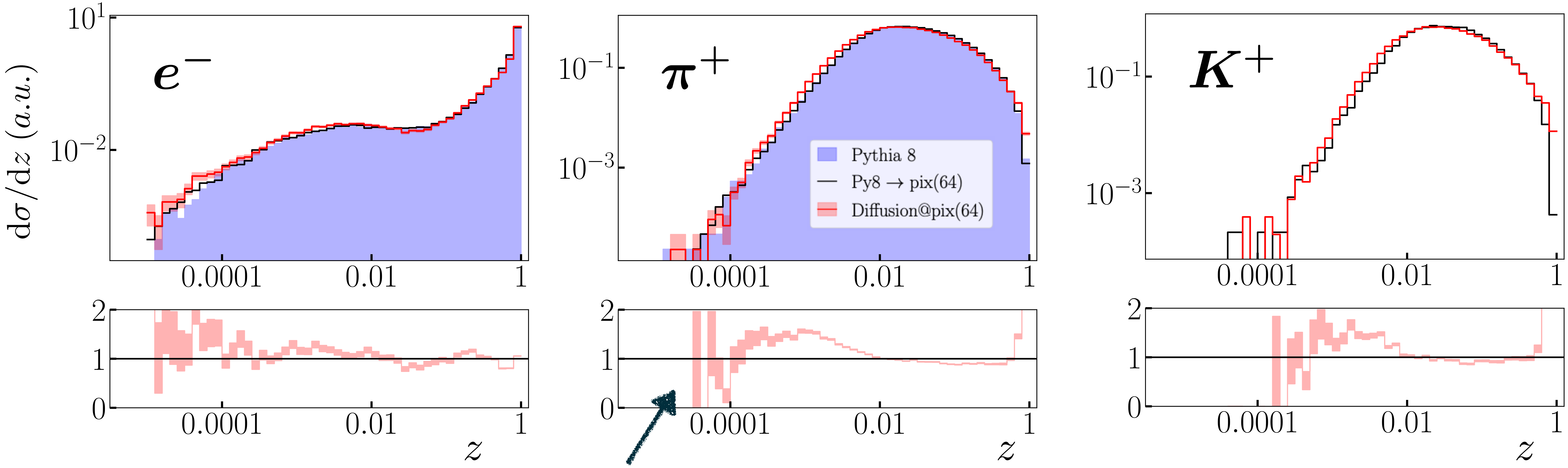


- Sampling $\mathcal{O}(\text{hours})$ for 10^6 events

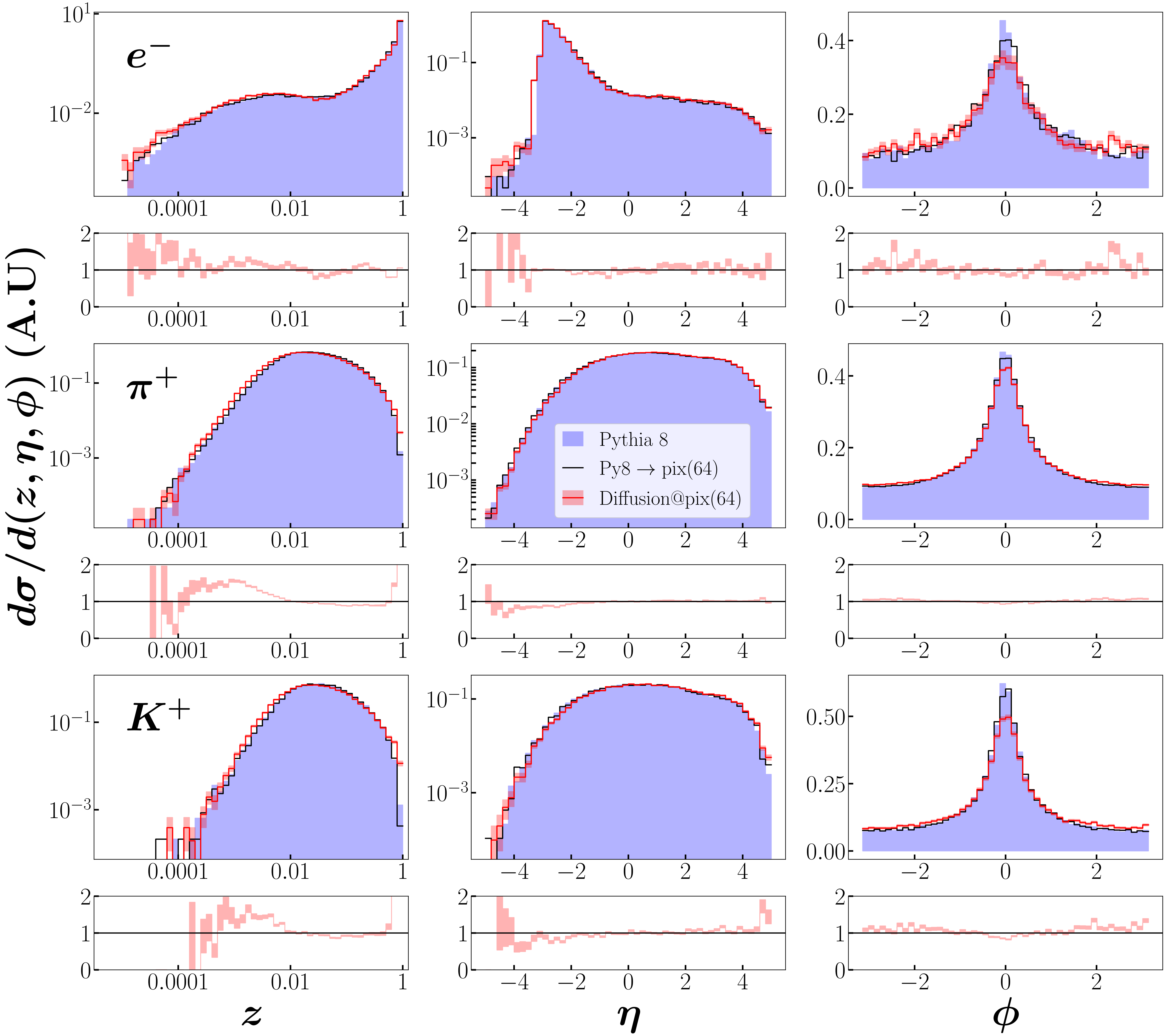
Diffusion models for the EIC

Devlin, Qiu, FR, Sato `23

- Momentum distributions



1σ Statistical uncertainties only



Devlin, Qiu, FR, Sato `23

Momentum, rapidity, azimuthal angle

Back-to-back with leading electron is $\phi = 0$

Diffusion models for the EIC

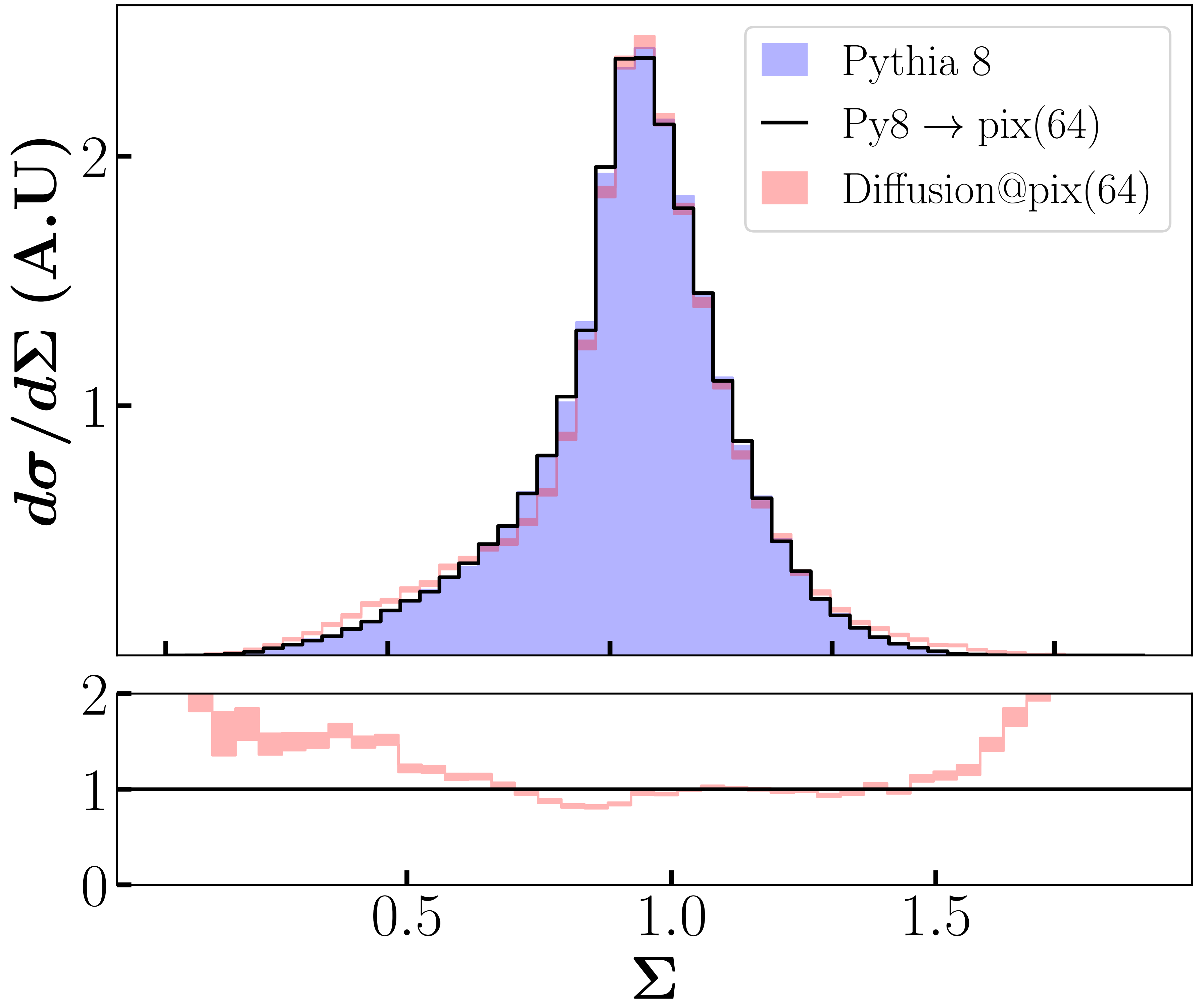
Devlin, Qiu, FR, Sato `23

- Event-wide momentum conservation

Rescaled momentum variable $$\tilde{z}_i = \frac{2M_{Ti}}{\sqrt{s}} \cosh y_i$$

with $$\Sigma \equiv \sum_{i\in\ \text{event}} \tilde{z}_i = 2$$

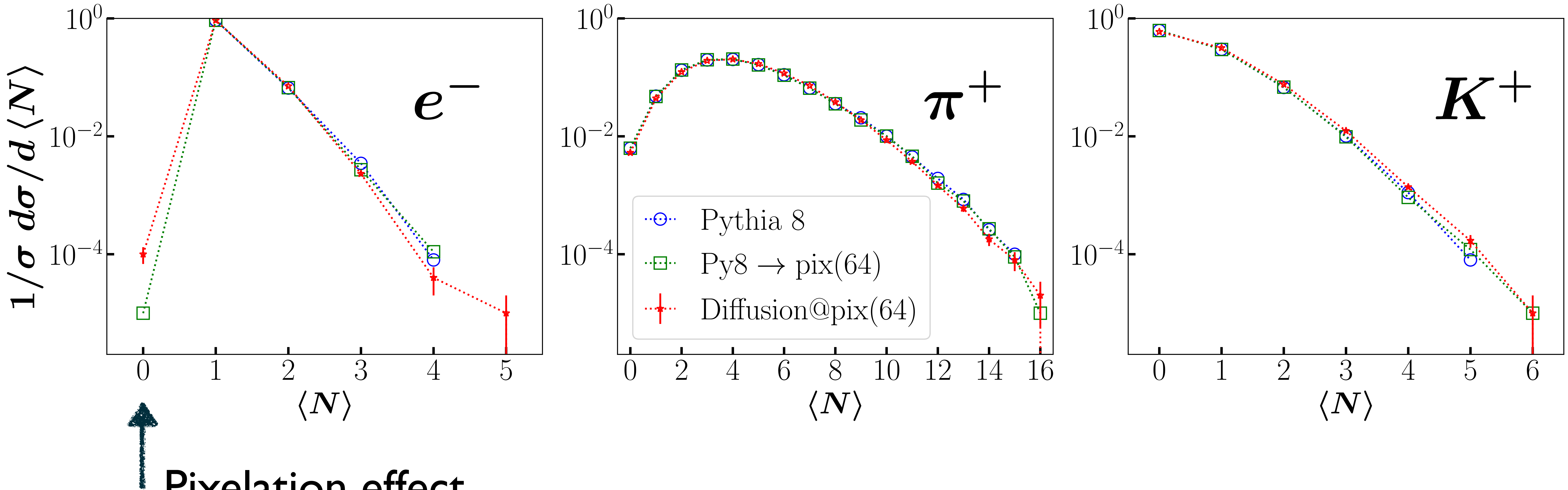
when all particles are included



Diffusion models for the EIC

Devlin, Qiu, FR, Sato `23

- Particle multiplicities



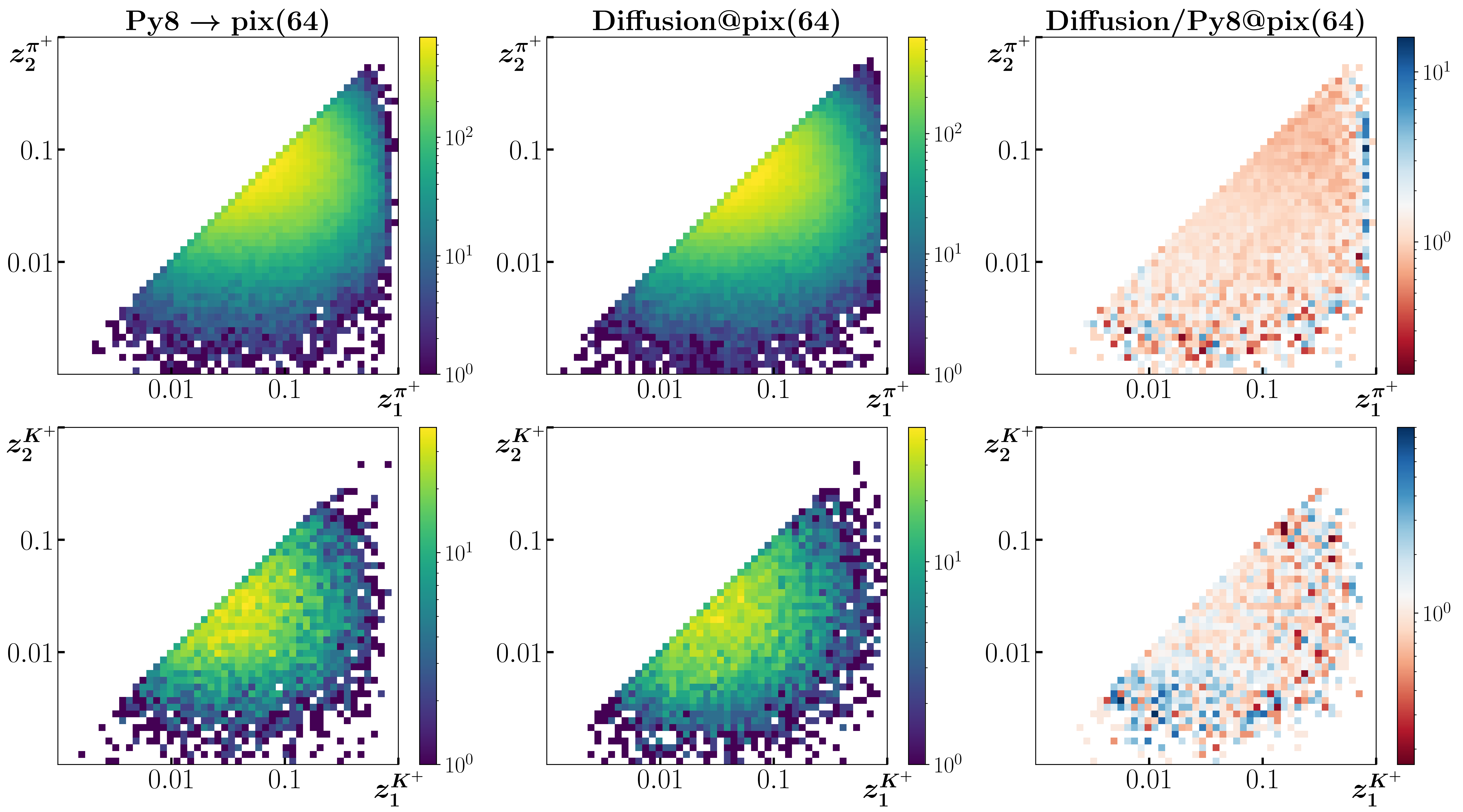
Pixelation effect

cf. e.g. Alanazi et al.

Diffusion models for the EIC

Devlin, Qiu, FR, Sato `23

- Multi-particle correlations

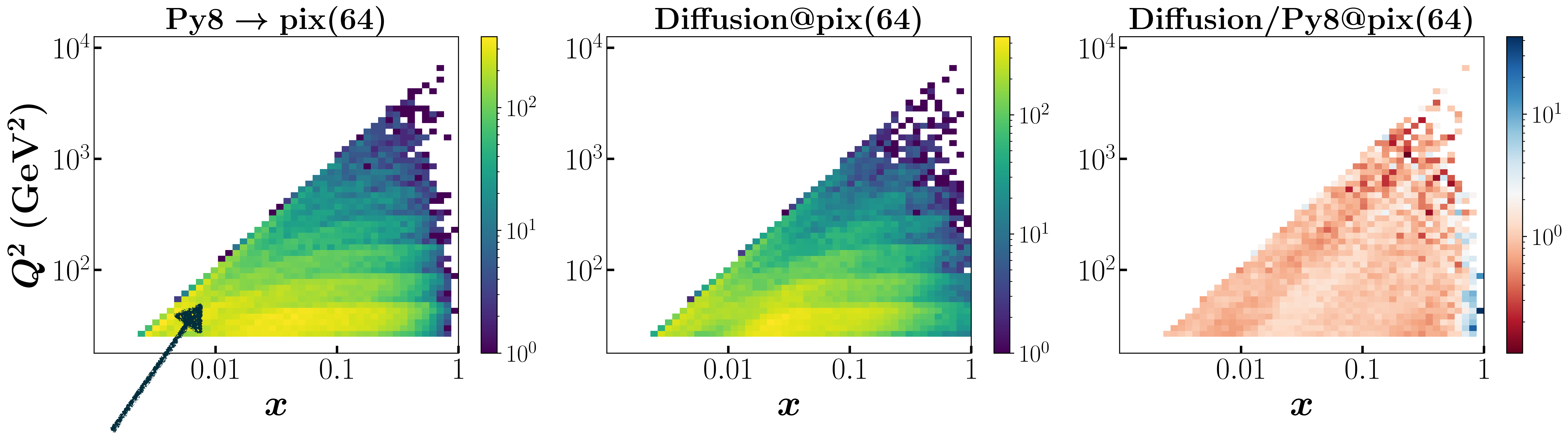


Diffusion models for the EIC

Devlin, Qiu, FR, Sato `23

- Correlation of Deep Inelastic Scattering (DIS) variables

$$x_B = \frac{Q^2}{2p \cdot q} \qquad Q^2 = -(k_e - k_{e'})^2$$



Pixelation effects

Diffusion models for the EIC

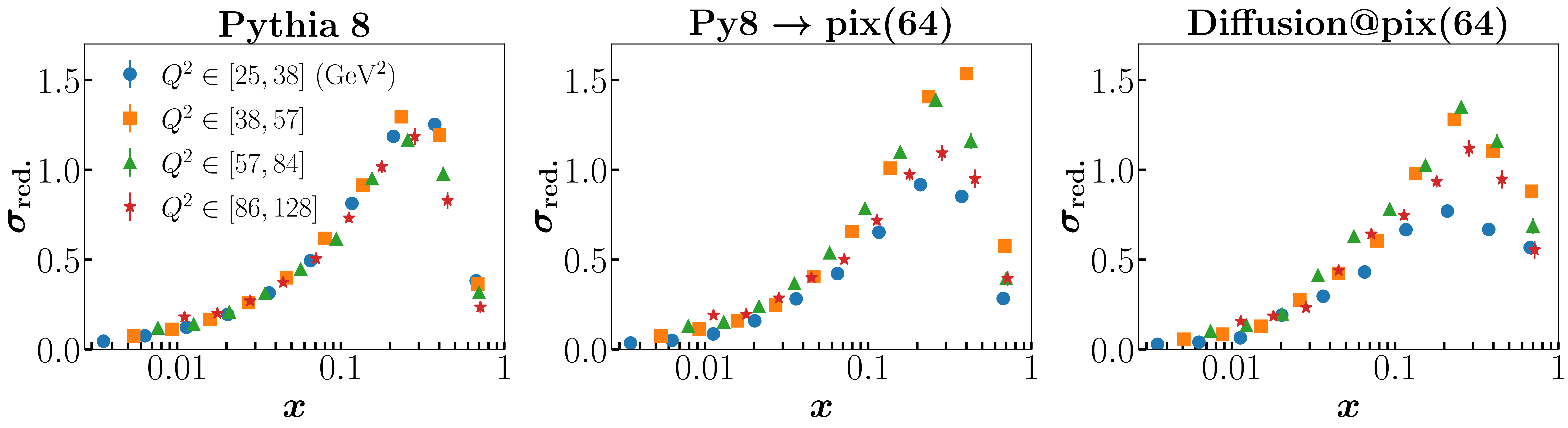
Devlin, Qiu, FR, Sato `23

- DIS reduced cross section

$$\sigma_{\text{red. NC}}^{ep\to e'+X} = \frac{\mathrm{d}\sigma_{\text{NC}}^{ep\to e'+X}}{\mathrm{d}x\mathrm{d}Q^2}\frac{Q^4 x}{2\pi\alpha^2 Y_+}$$
$$= F_2\left(x,Q^2\right) - \frac{y^2}{Y_+}F_L\left(x,Q^2\right) + \frac{Y_-}{Y_+}xF_3\left(x,Q^2\right)$$

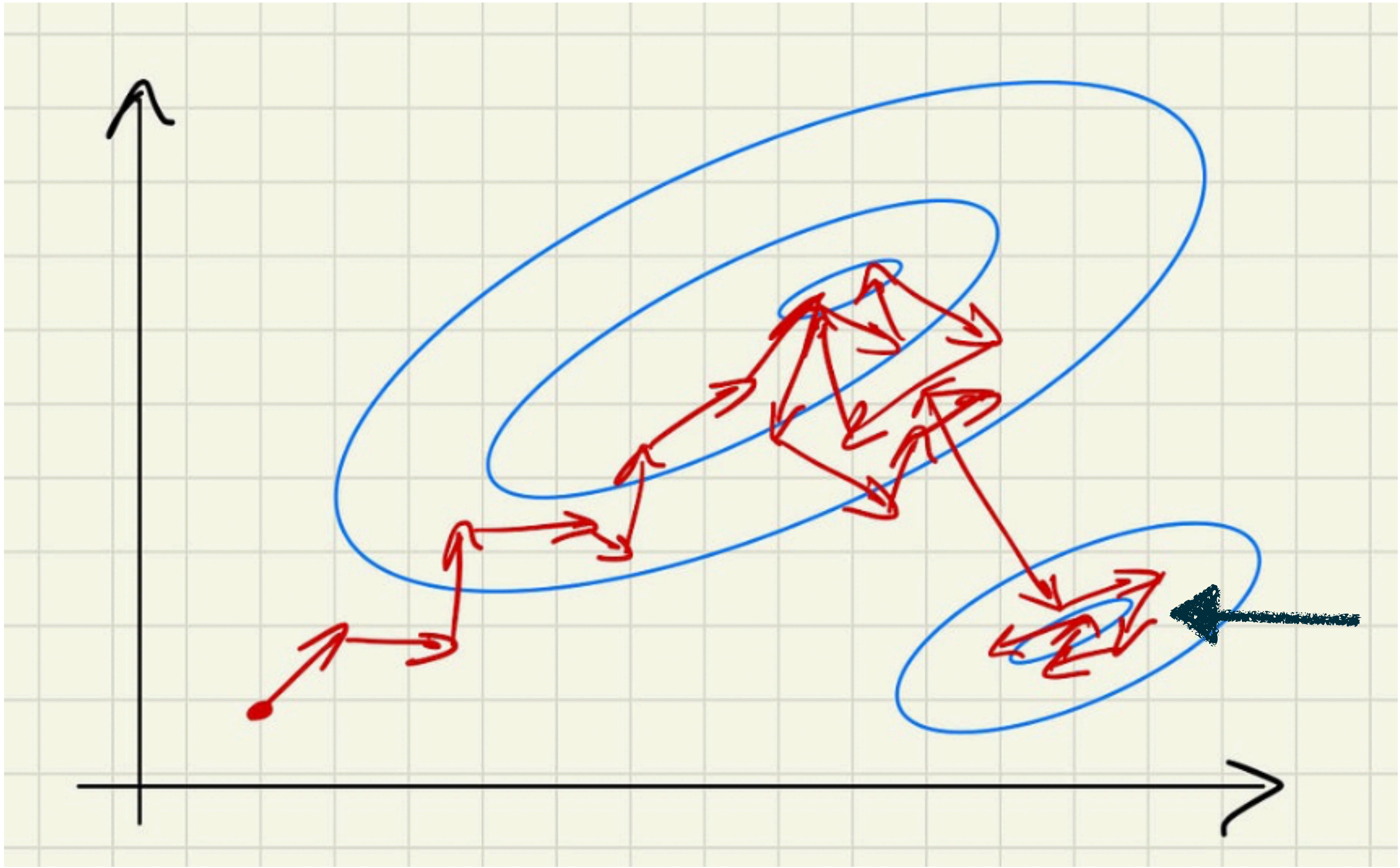
Inelasticity

$$Y_\pm = 1 \pm (1-y)^2$$



MCMC sampling with diffusion models

- Assist Metropolis-Hastings algorithm
- Iteratively train diffusion model on obtained samples



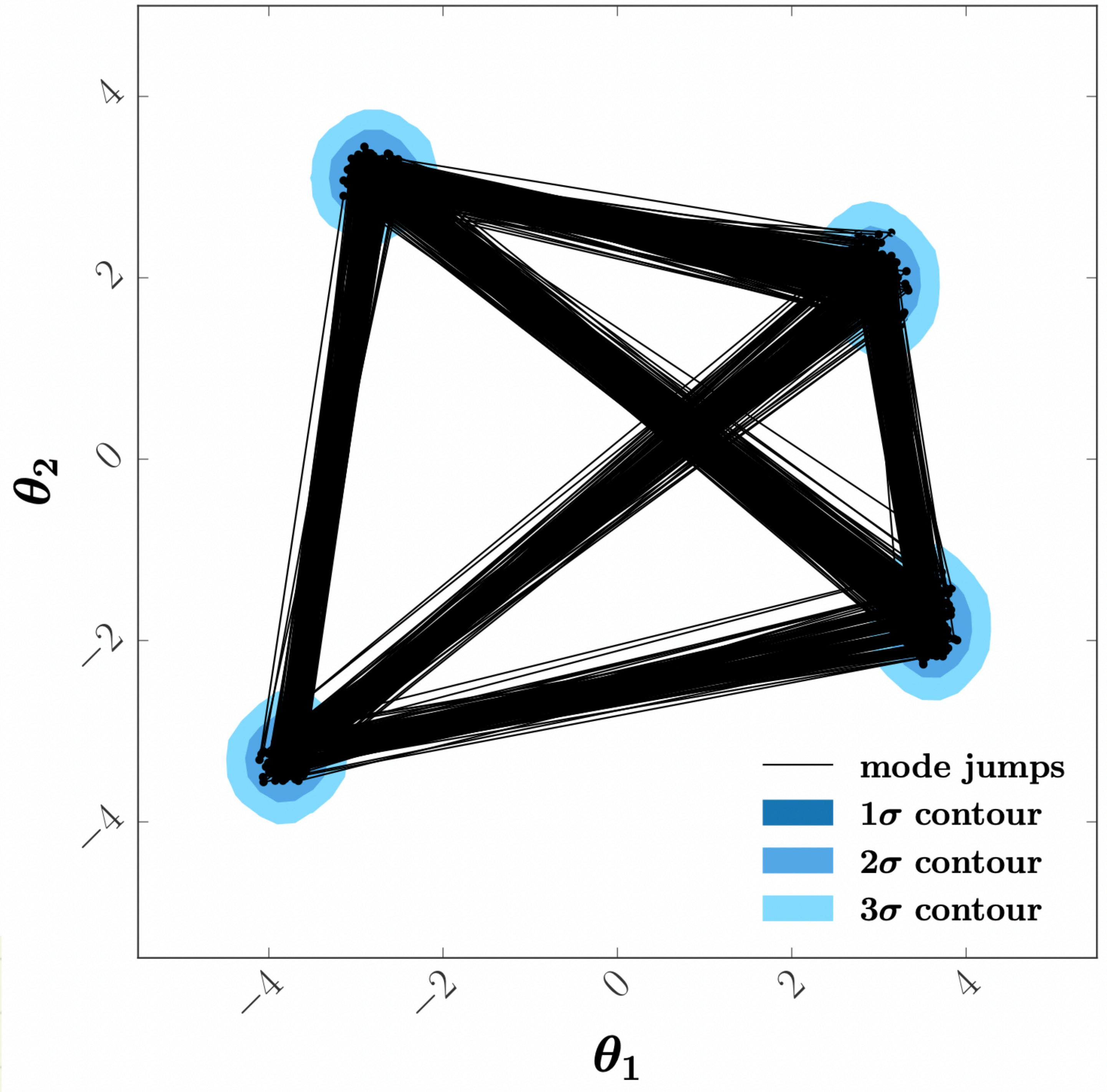
Gaussian proposal function

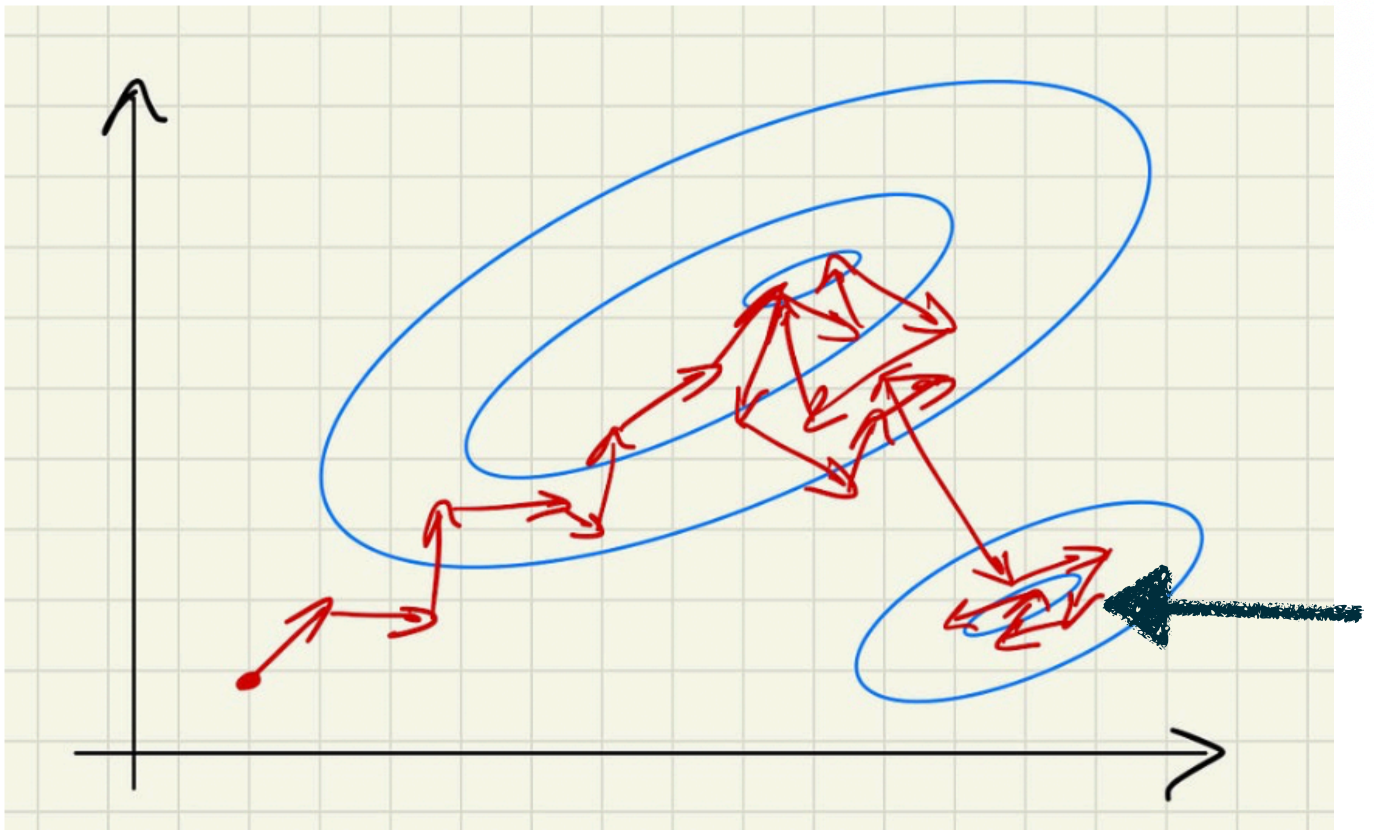
Hunt-Smith, Melnitchouk, FR, Sato, Thomas, White `23
see also Vargas, Grathwohl, Doucet `23

MCMC sampling with diffusion models

- Assist Metropolis-Hastings algorithm
- Iteratively train diffusion model on obtained samples
- Interleave chain with global proposal function from the diffusion model
- Example: 2d Himmelblau function

$$f(\boldsymbol{\theta}) = (\theta_1^2 + \theta_2 - 11)^2 + (\theta_1 + \theta_2^2 - 7)^2$$





Gaussian proposal function

Hunt-Smith, Melnitchouk, FR, Sato, Thomas, White `23
see also Vargas, Grathwohl, Doucet `23

MCMC sampling with diffusion models

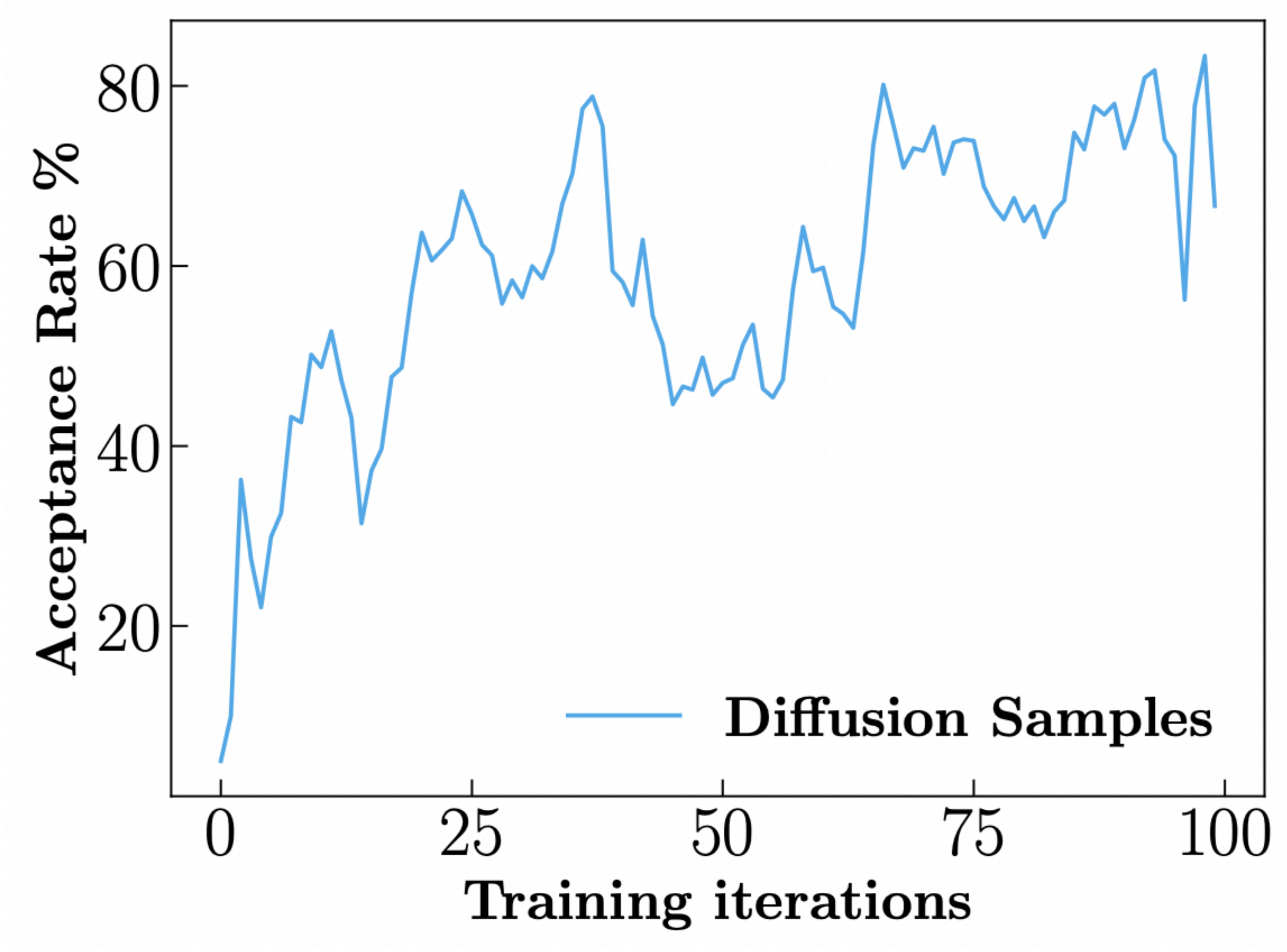
Hunt-Smith, Melnitchouk, FR, Sato, Thomas, White `23
Gabrie, Rotskoff, Vanden-Eijnden `21, `22
see also Yamauchi et al. `23

10d Gaussian mixture

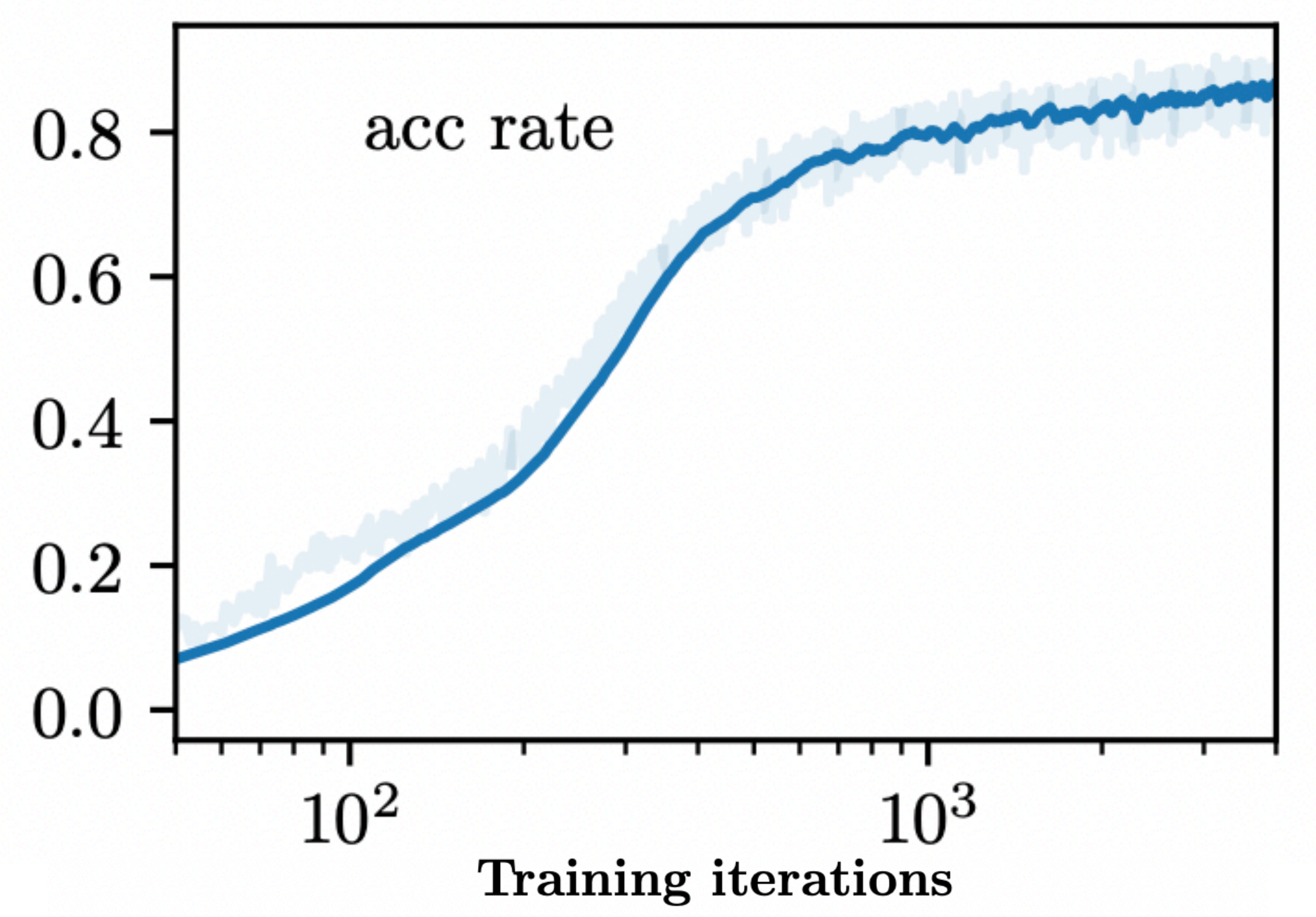
$$f(\theta_i) = w_A \frac{e^{-\frac{1}{2}|\theta_i - \theta_{A,i}|^2}}{(2\pi)^{d/2}} + w_B \frac{e^{-\frac{1}{2}|\theta_i - \theta_{B,i}|^2}}{(2\pi)^{d/2}}$$

Acceptance rate

Additionally, it requires probability for accept/reject step



Diffusion model



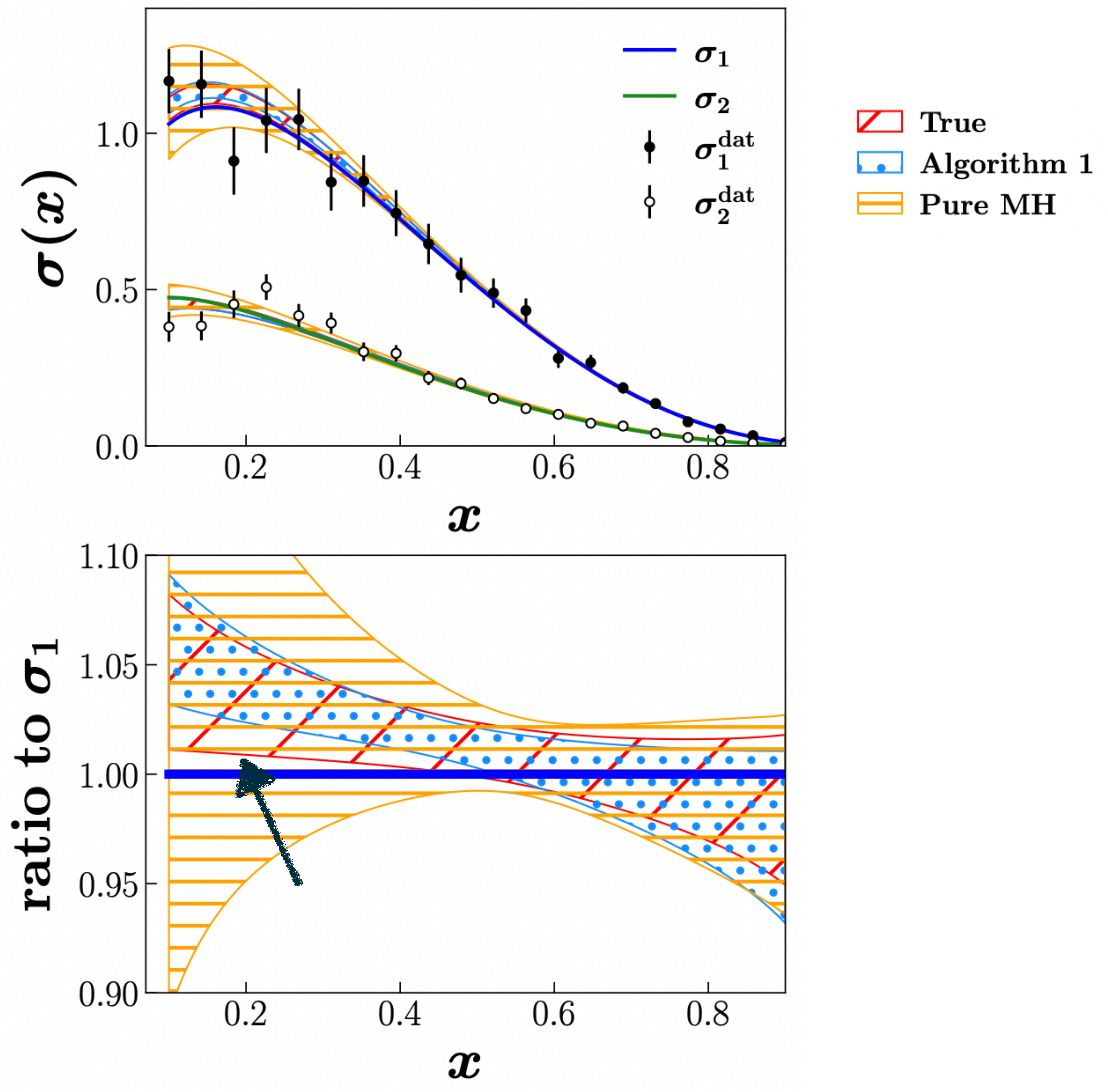
Normalizing flow

Bayesian posterior sampling

- Analysis of (toy) PDFs

$$p(\boldsymbol{\theta}|\boldsymbol{D}) = \frac{\mathcal{L}(\boldsymbol{D}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})}{\int \mathrm{d}\boldsymbol{\theta}\, \mathcal{L}(\boldsymbol{D}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})}$$

- Likelikhood $\quad \mathcal{L} = \exp\left(-\frac{1}{2}\chi^2\right)$



Fast convergence of diffusion model-assisted MCMC

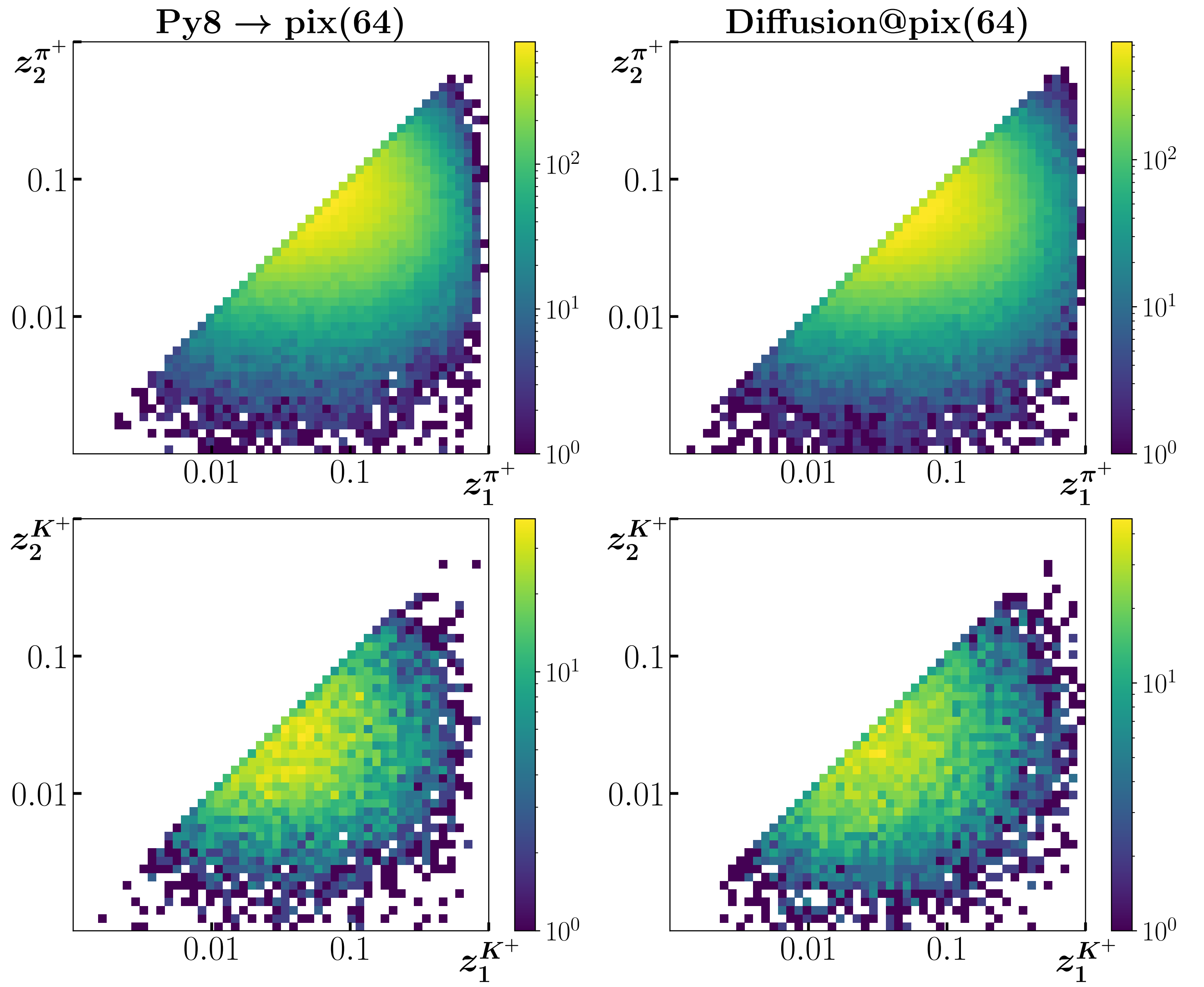
Hunt-Smith, Melnitchouk, FR, Sato, Thomas, White `23

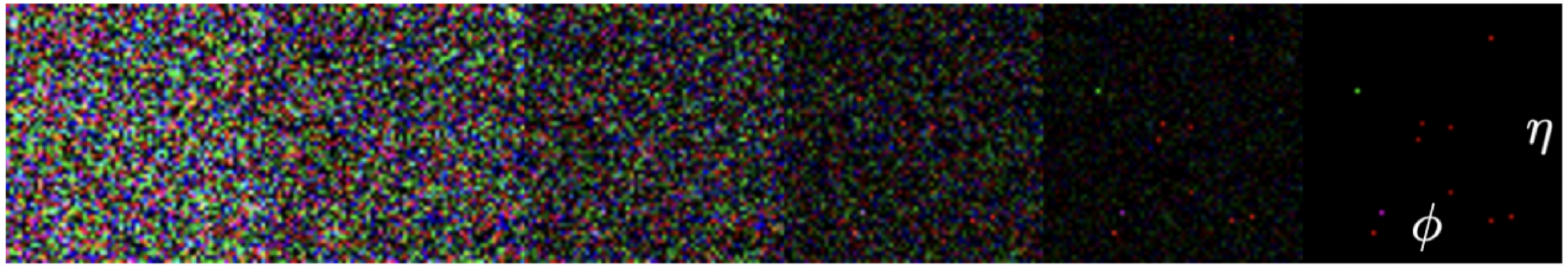
Conclusions & outlook

- Generative models for collider physics
- Simulation of full EIC events with unique challenges
- Various possible future applications

- Point cloud-based diffusion model

→ See Vinicius, Fernando's talk





U.S. DEPARTMENT OF ENERGY | Office of Science