Machine learning-based methods in quantum chromodynamics

Felix Ringer

EINN 2023, Paphos Cyprus

Old Dominion University

Jefferson Lab

Recent progress in AI & machine learning

Natural language processing
- Transformers

Dall-E

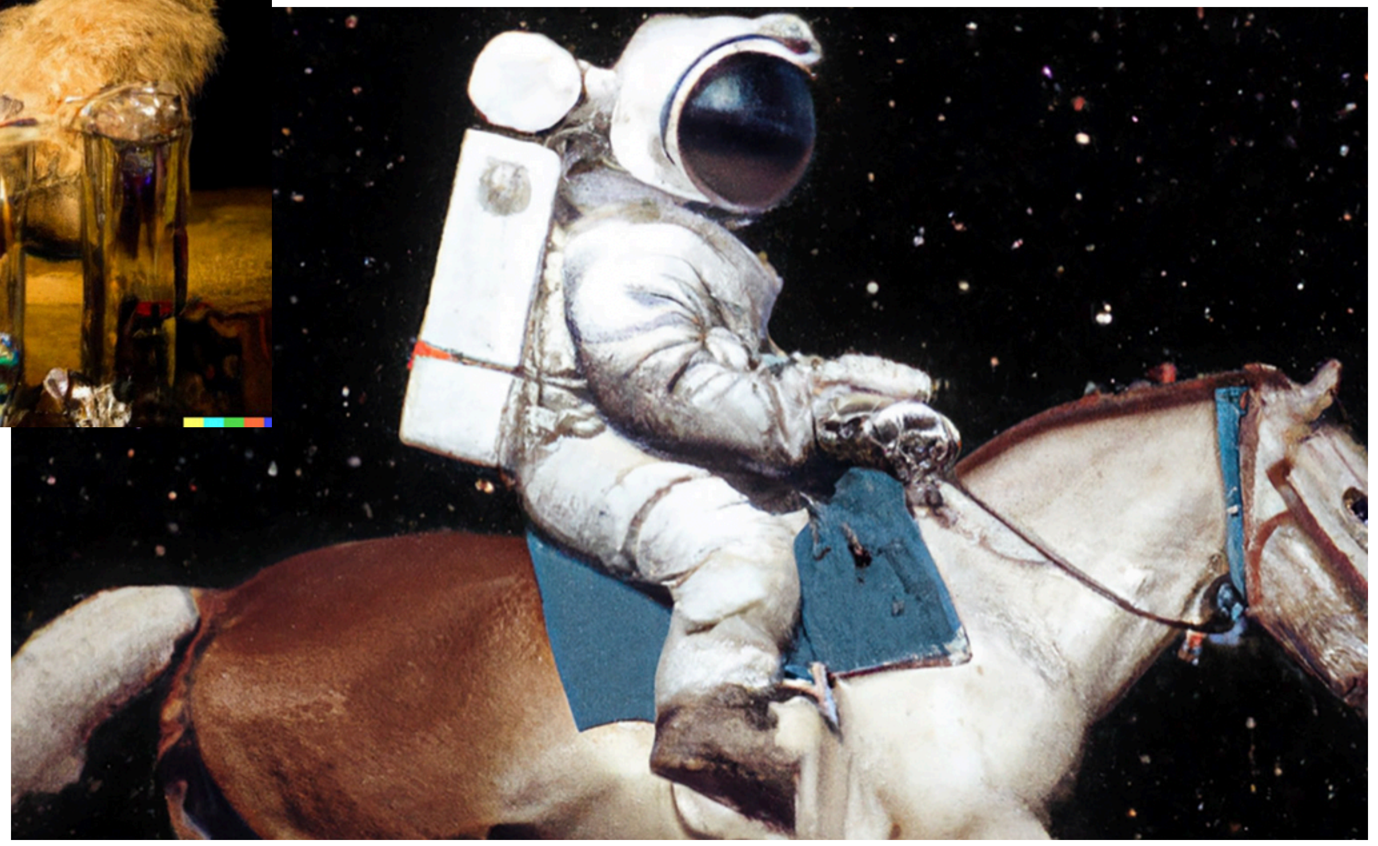


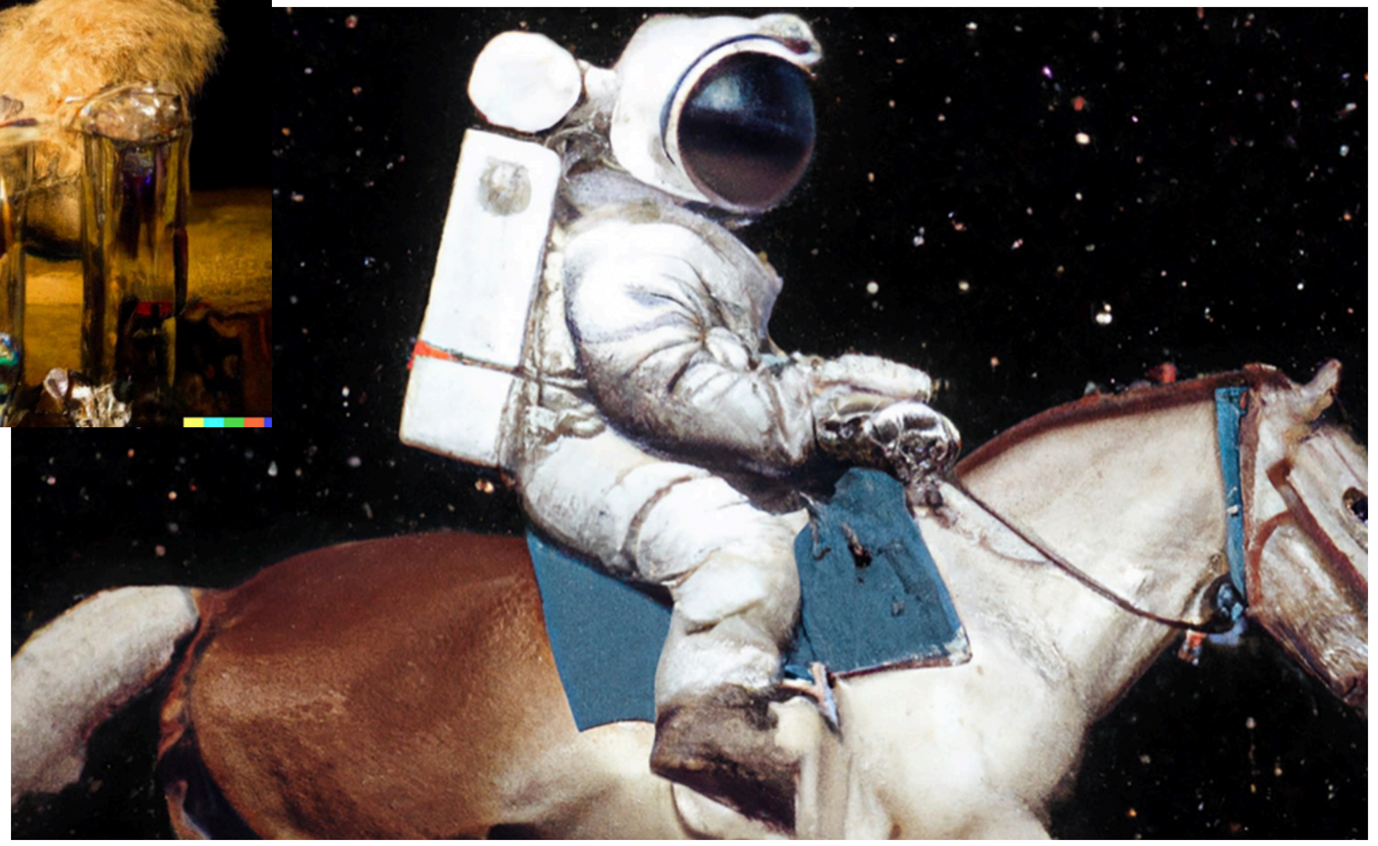
Image generation
- Diffusion models

Recent progress in AI & machine learning

Natural language processing
- Transformers

Improve constraints on PDFs

EIC simulations

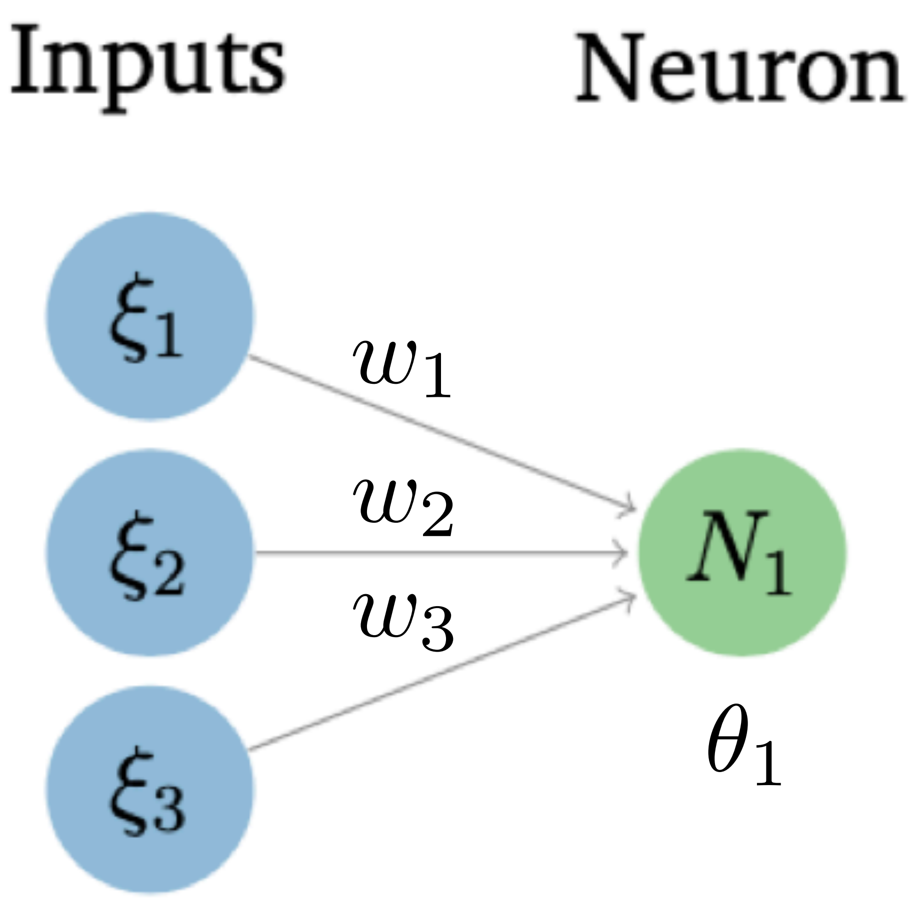


Dall-E

Image generation
- Diffusion models

Neural networks

Feed forward neural network



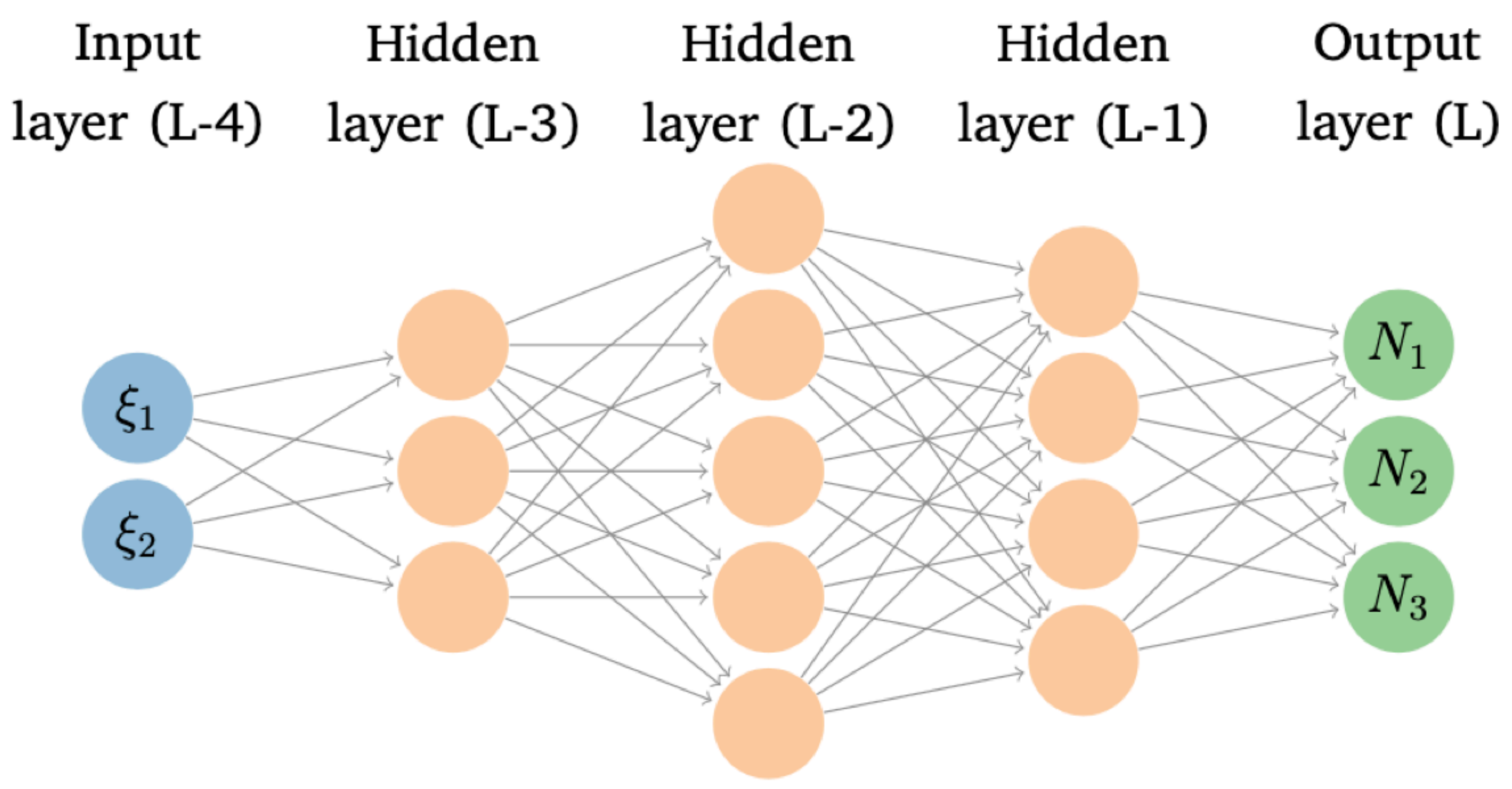
Nonlinear activation function

$$N_1(\vec{\xi}; \vec{w}, \theta_1) \; = \; \phi(\vec{\xi} \cdot \vec{w} + \theta_1)$$

4 parameters: weights bias

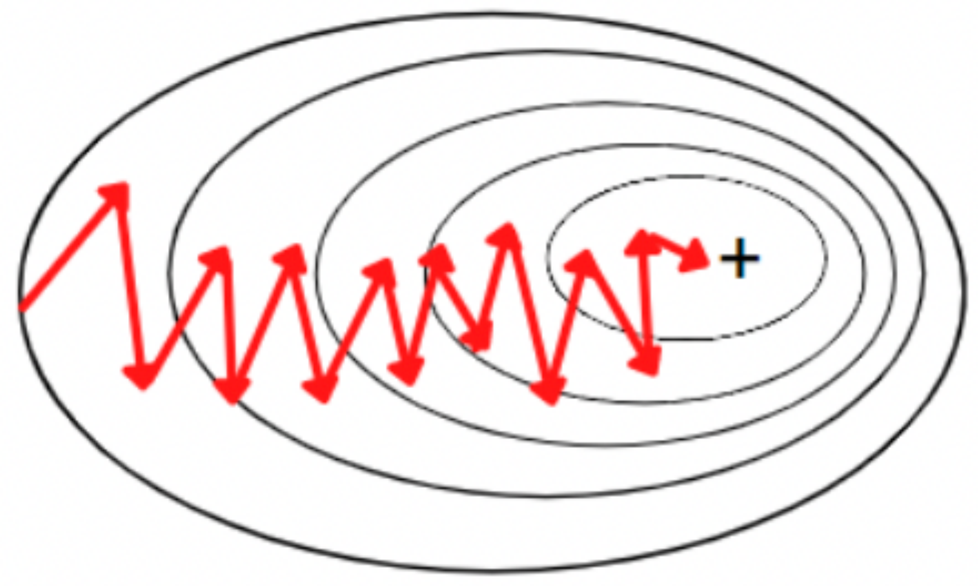
Build up a complex function by adding many neurons & layers

Neural networks

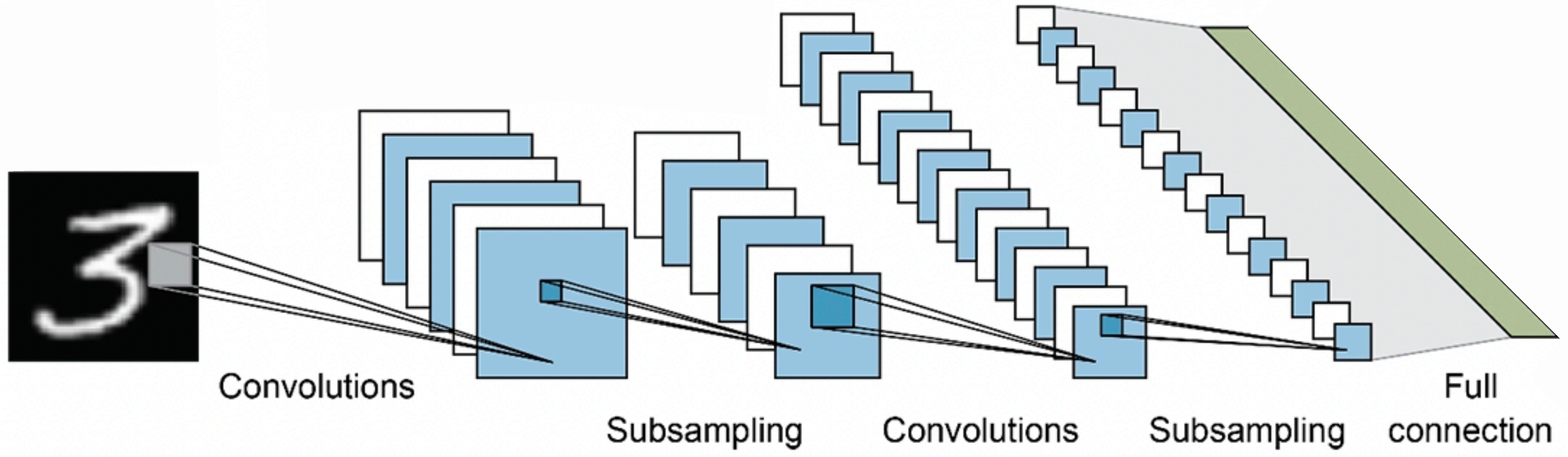


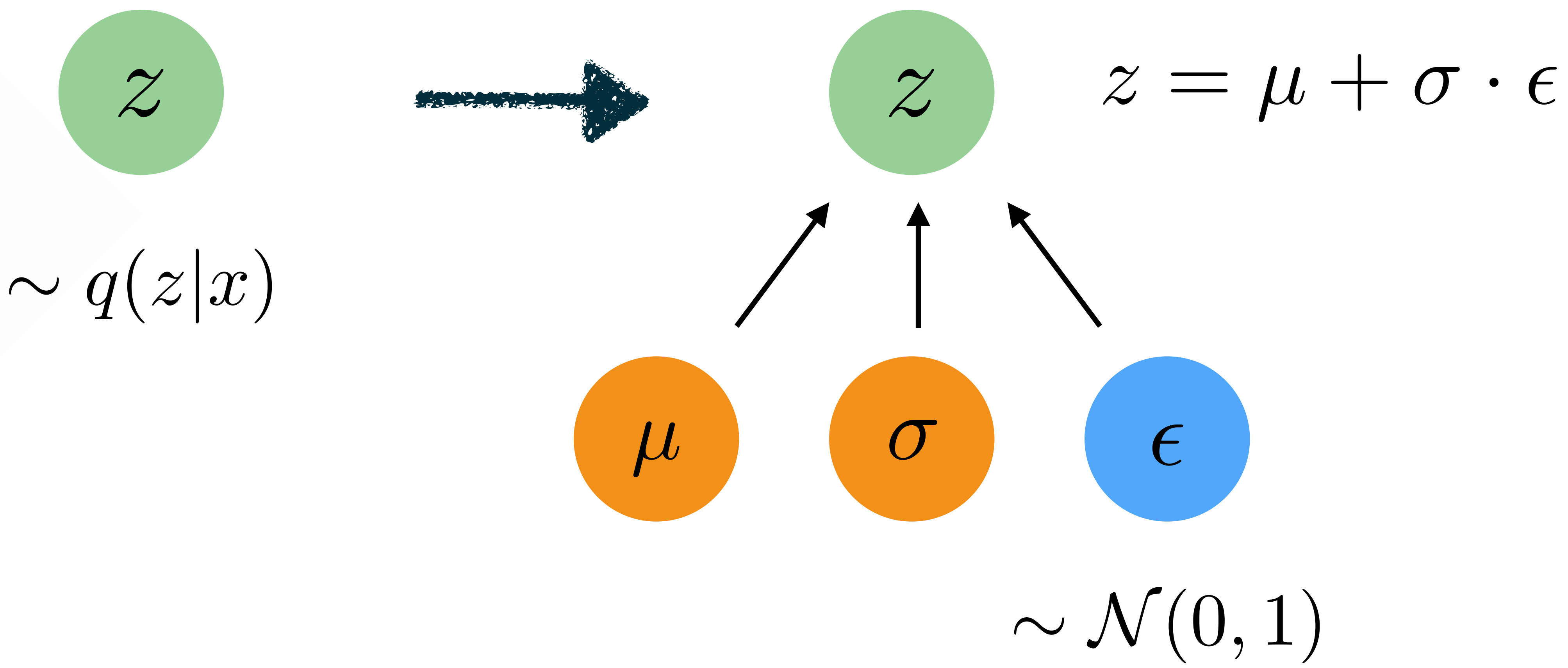
Can be trained efficiently with

- Backpropagation (chain rule)
- Stochastic gradient descent
- Up to ~billion parameters



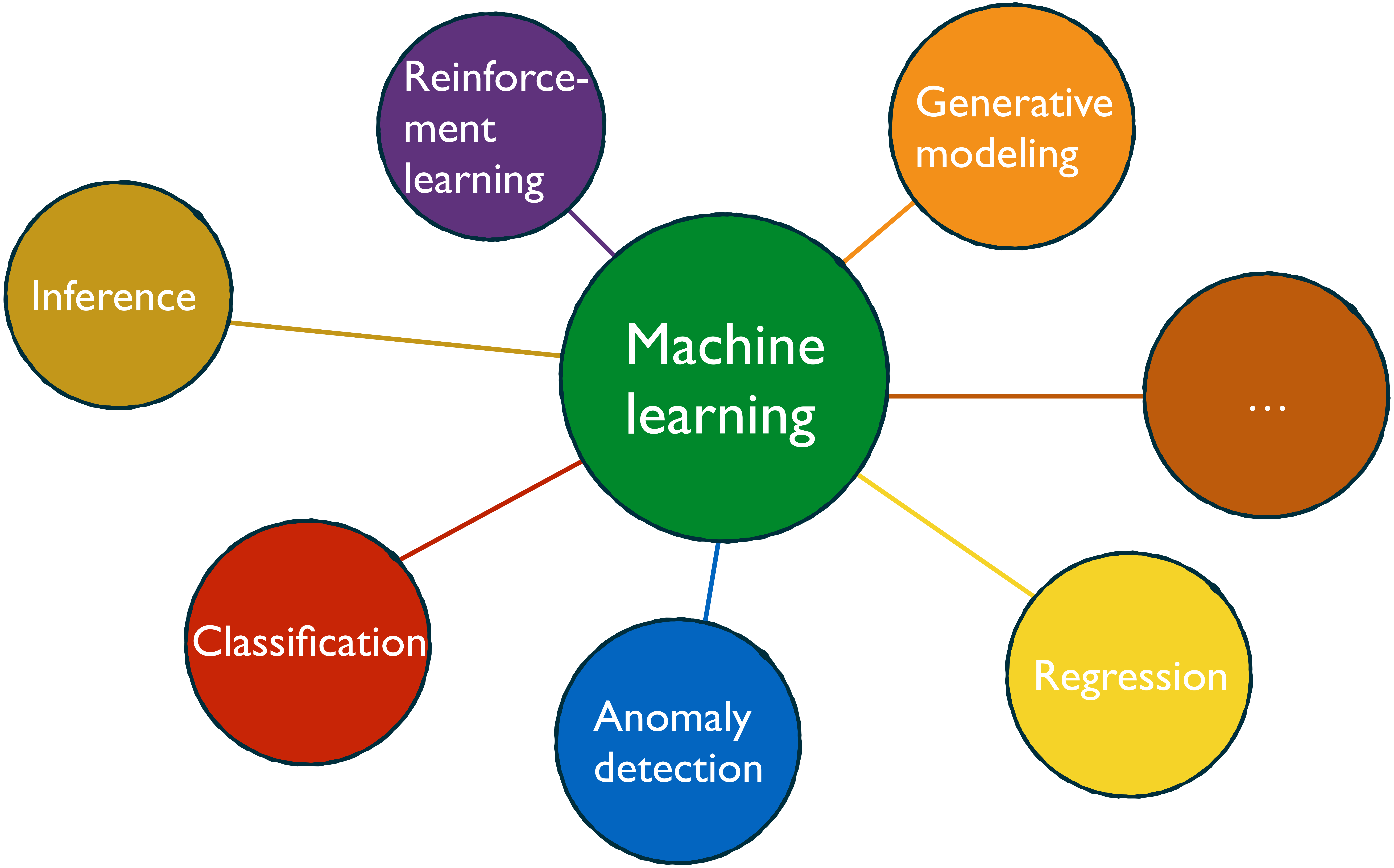
Neural networks





- Convolutional neural networks
- RG flow

- Stochastic nodes e.g. Bayesian NNs, diffusion models

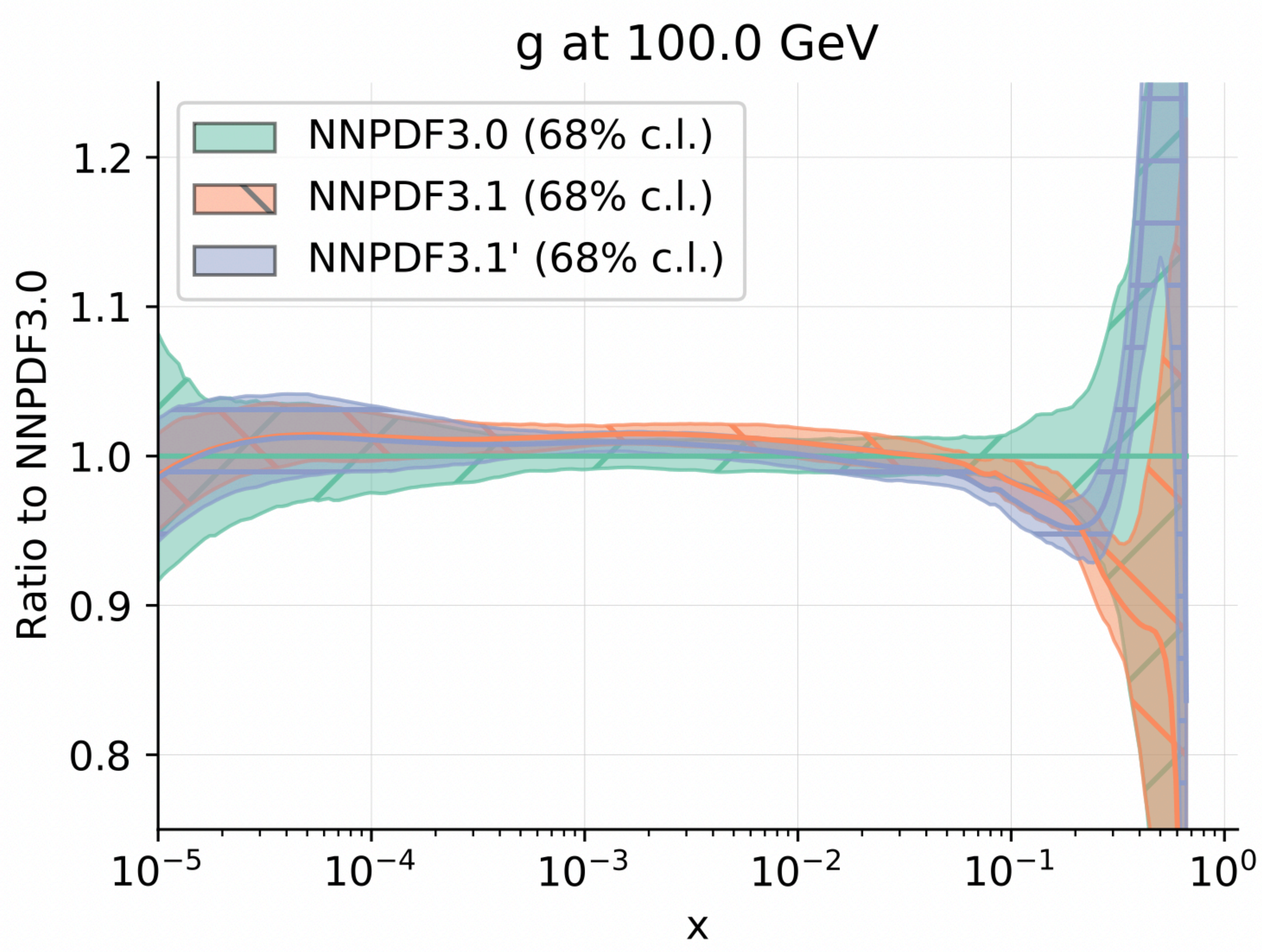

Reinforce-
ment
learning
Generative
modeling
Inference
Machine
learning
...
Classification
Anomaly
detection
Regression

Parametrization of PDFs

2203.05506

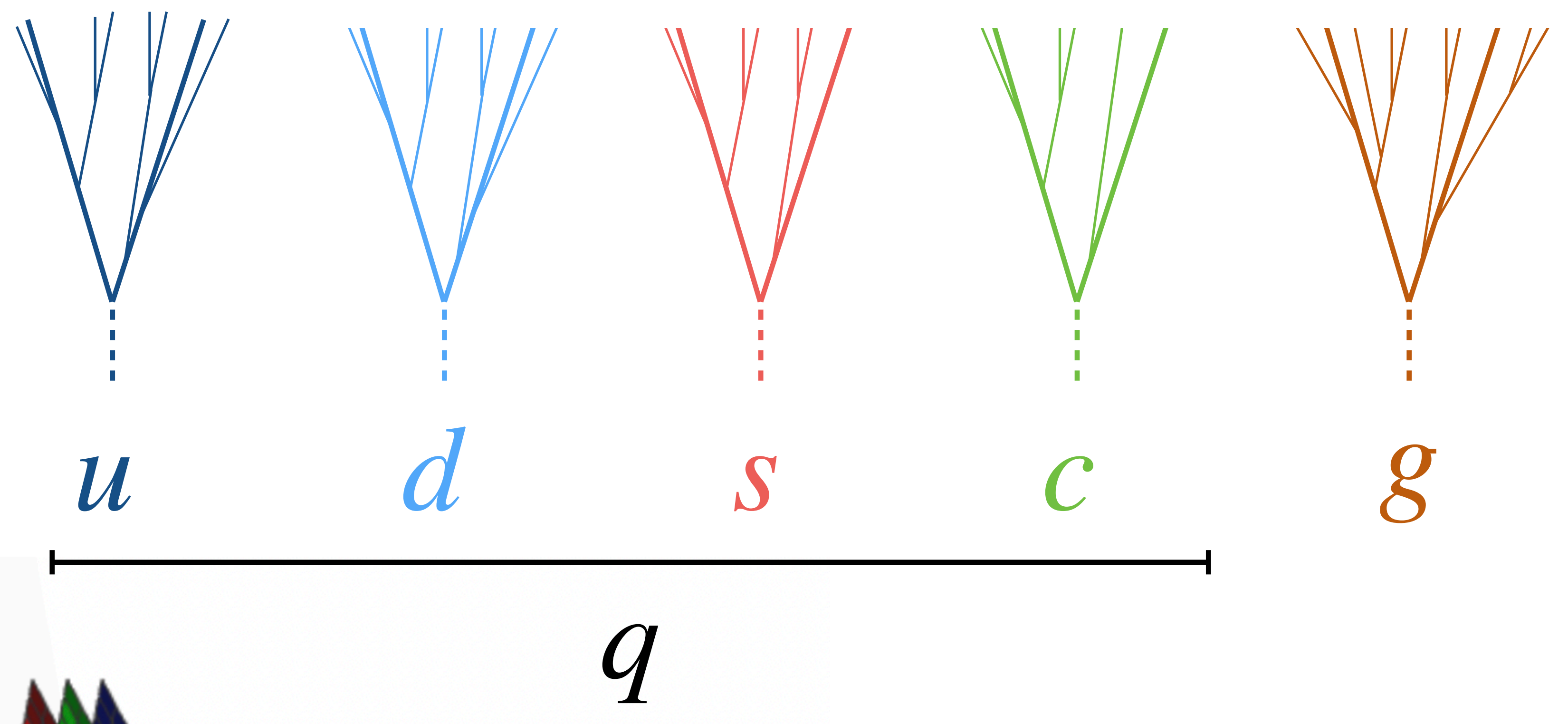
- Flexible functional form
- Minimize associated biases
- NNPDF Collaboration
- Extractions from lattice QCD

→ *see Pavel Nadolsky's talk*



Jet physics & Machine learning

- Various jet classifiers have been developed
  - Typically ML significantly outperformed traditional observables
  - Event-by-event information vs. low-dimensional observables



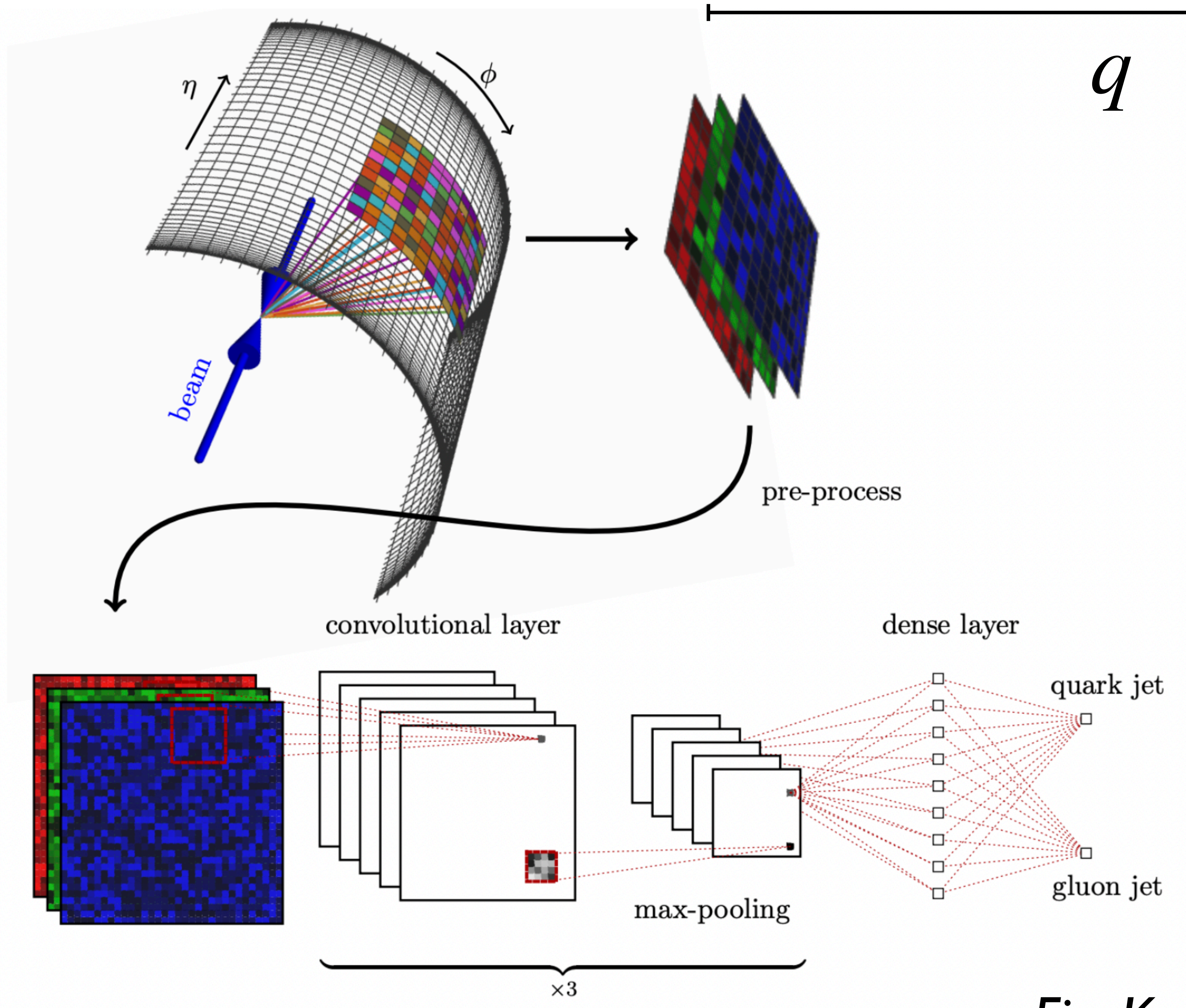
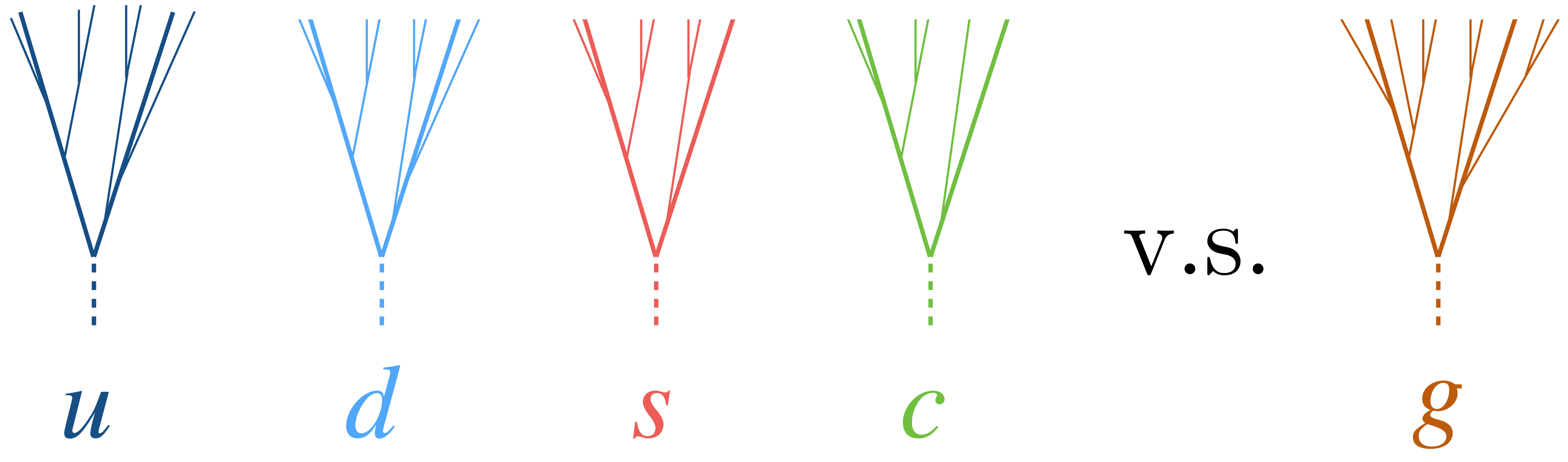
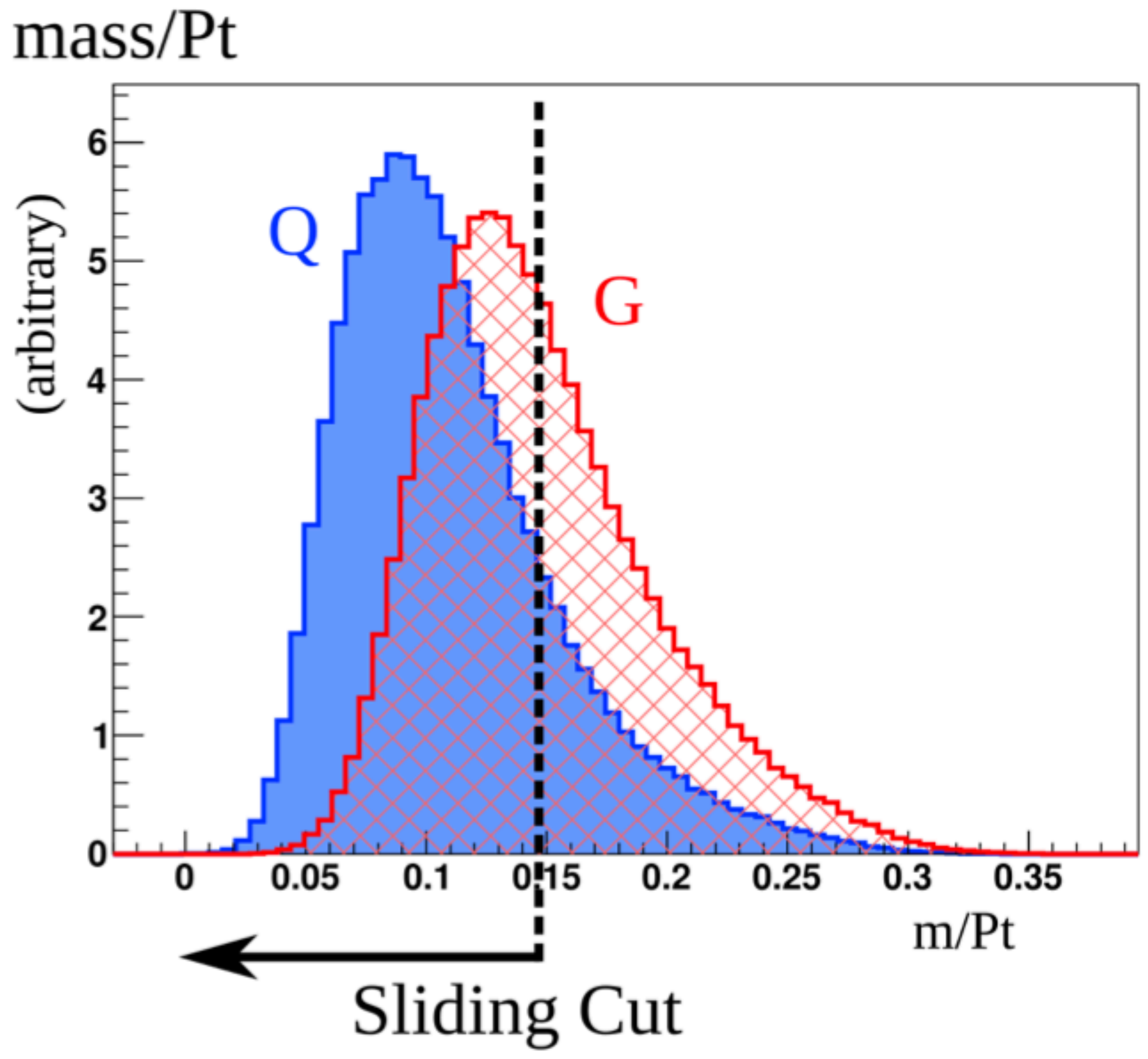


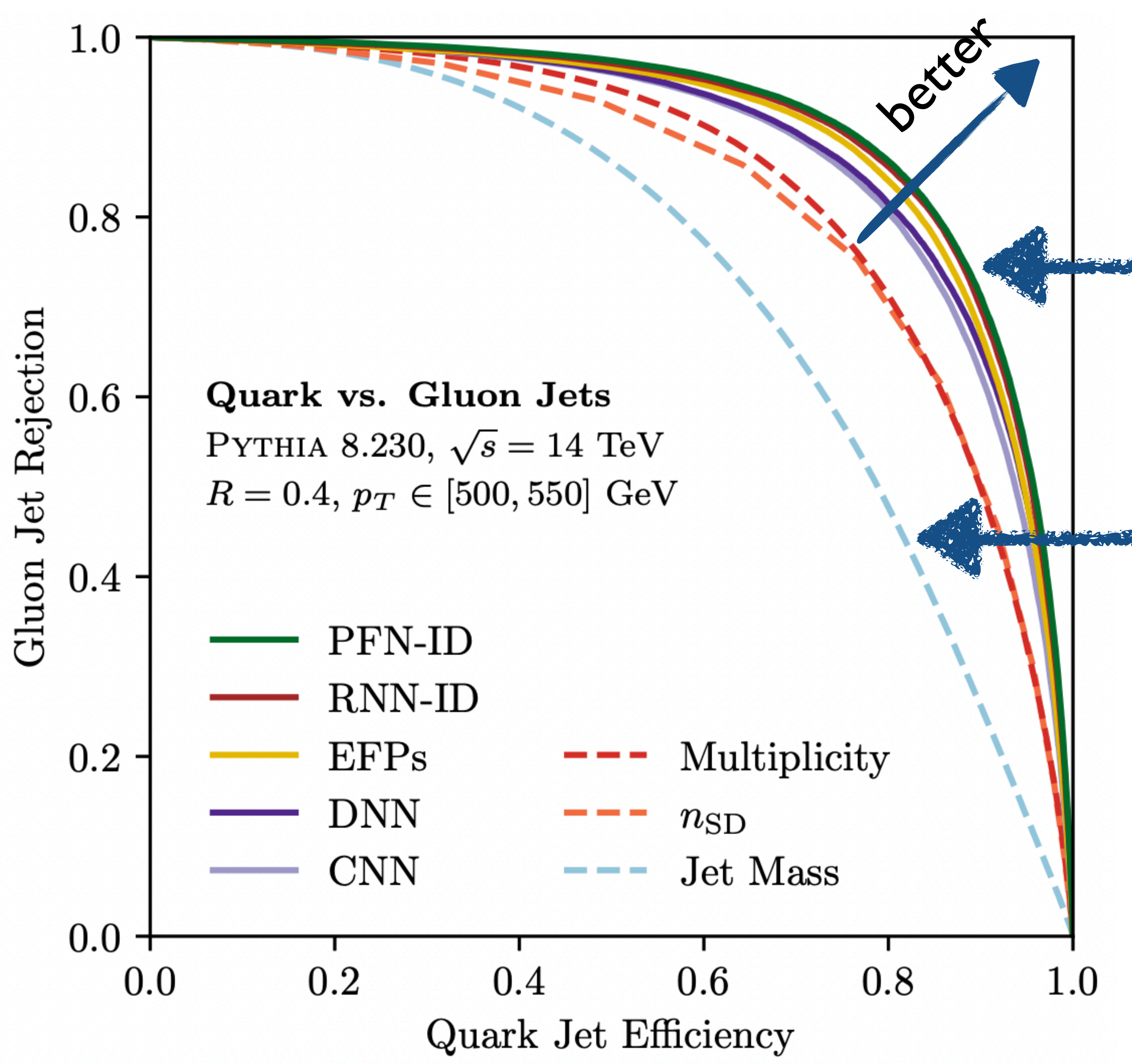
Fig. Komiske, Metodiev, Schwartz

Jet physics & Machine learning

- Various jet classifiers have been developed
  - Example: Quark vs. gluon jet classification
  - Quantify using a ROC curve







AI/ML

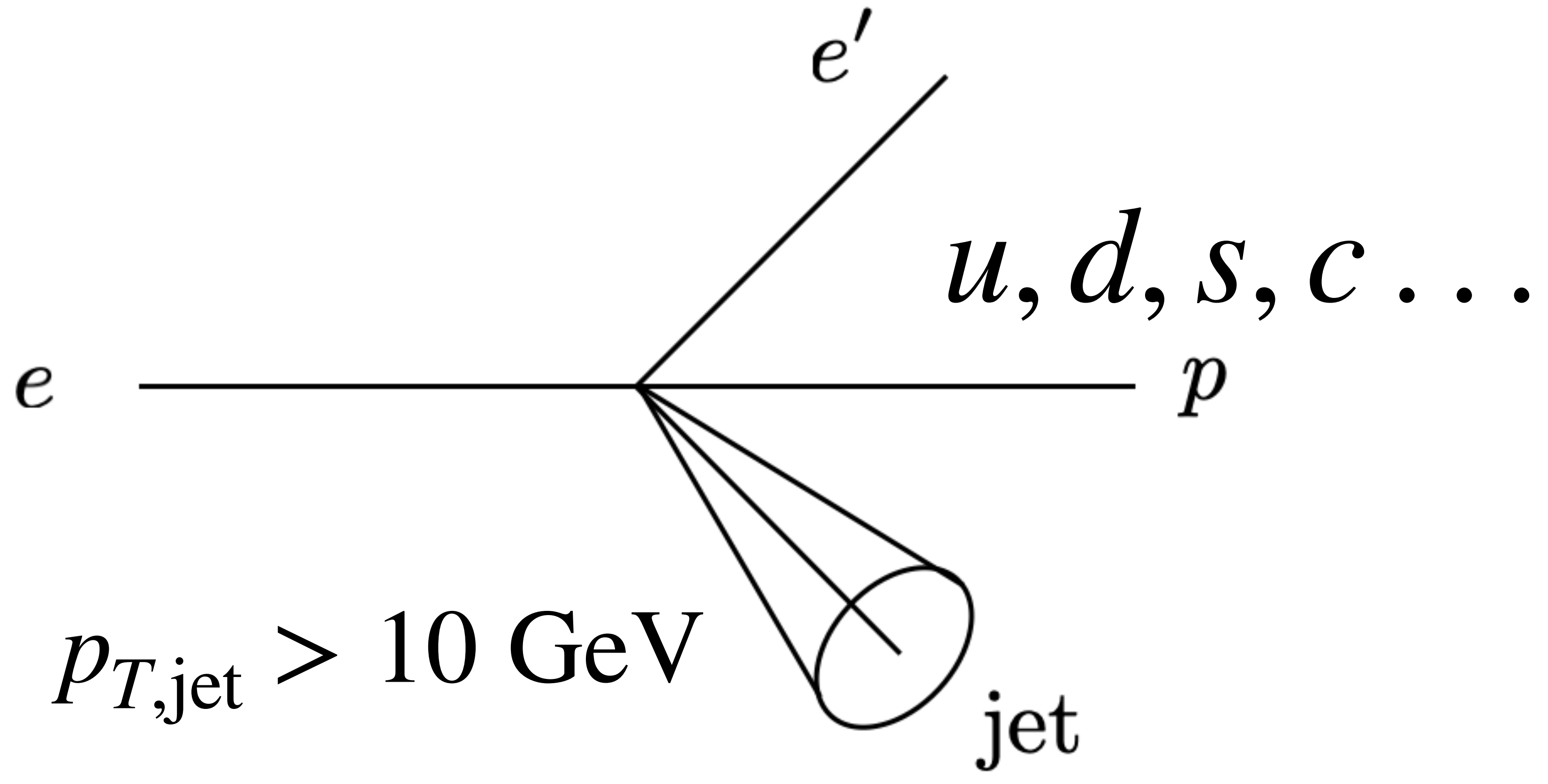
Traditional observable

Gallicchio, Schwartz
Komiske, Metodiev, Thaler `19

EIC & RHIC jets

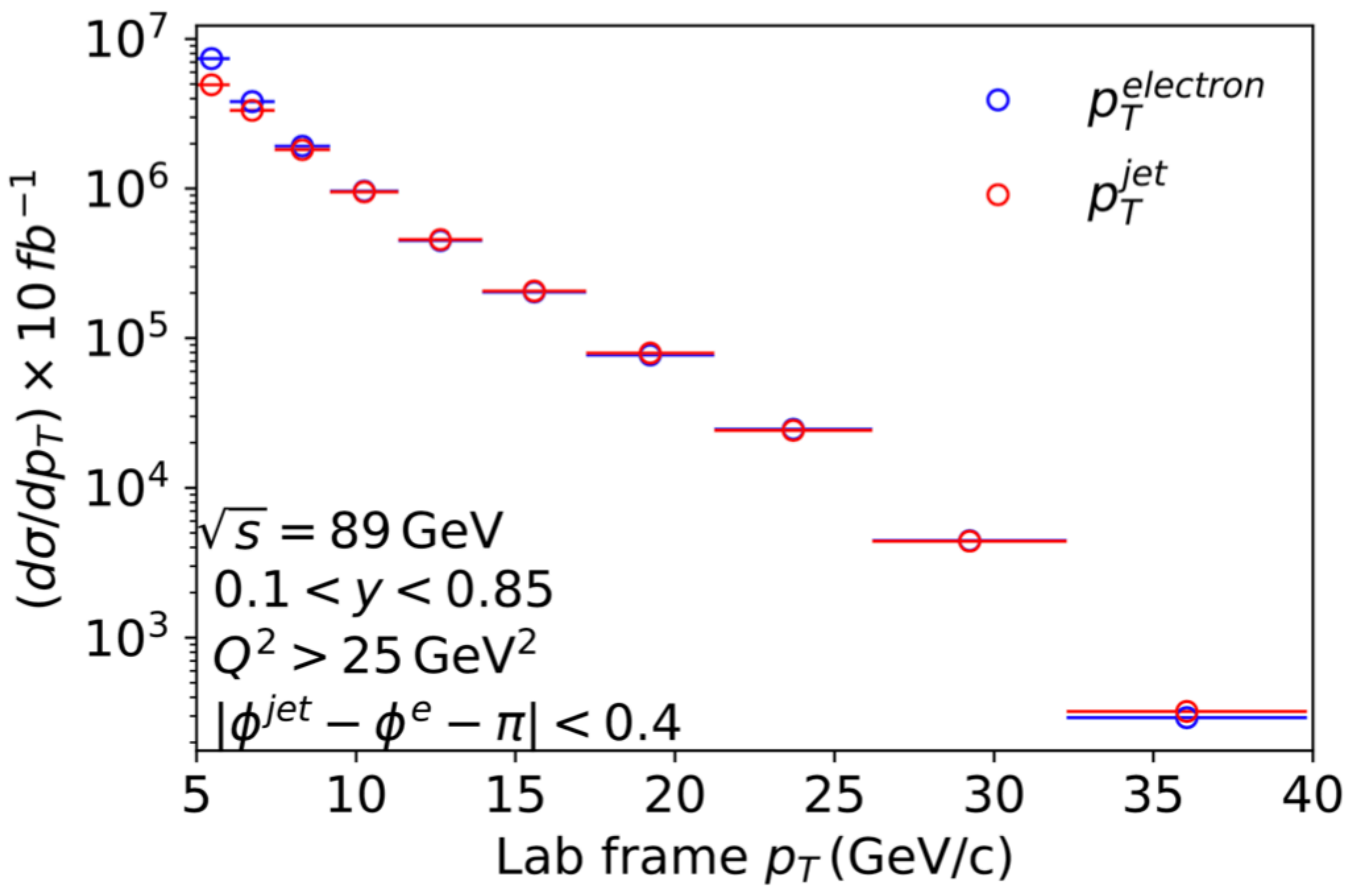
- Relatively low particle multiplicities at the EIC

- PYTHIA6
  - Particle $(p_{Ti}, \eta_i, \phi_i, \mathrm{PID}_i)$

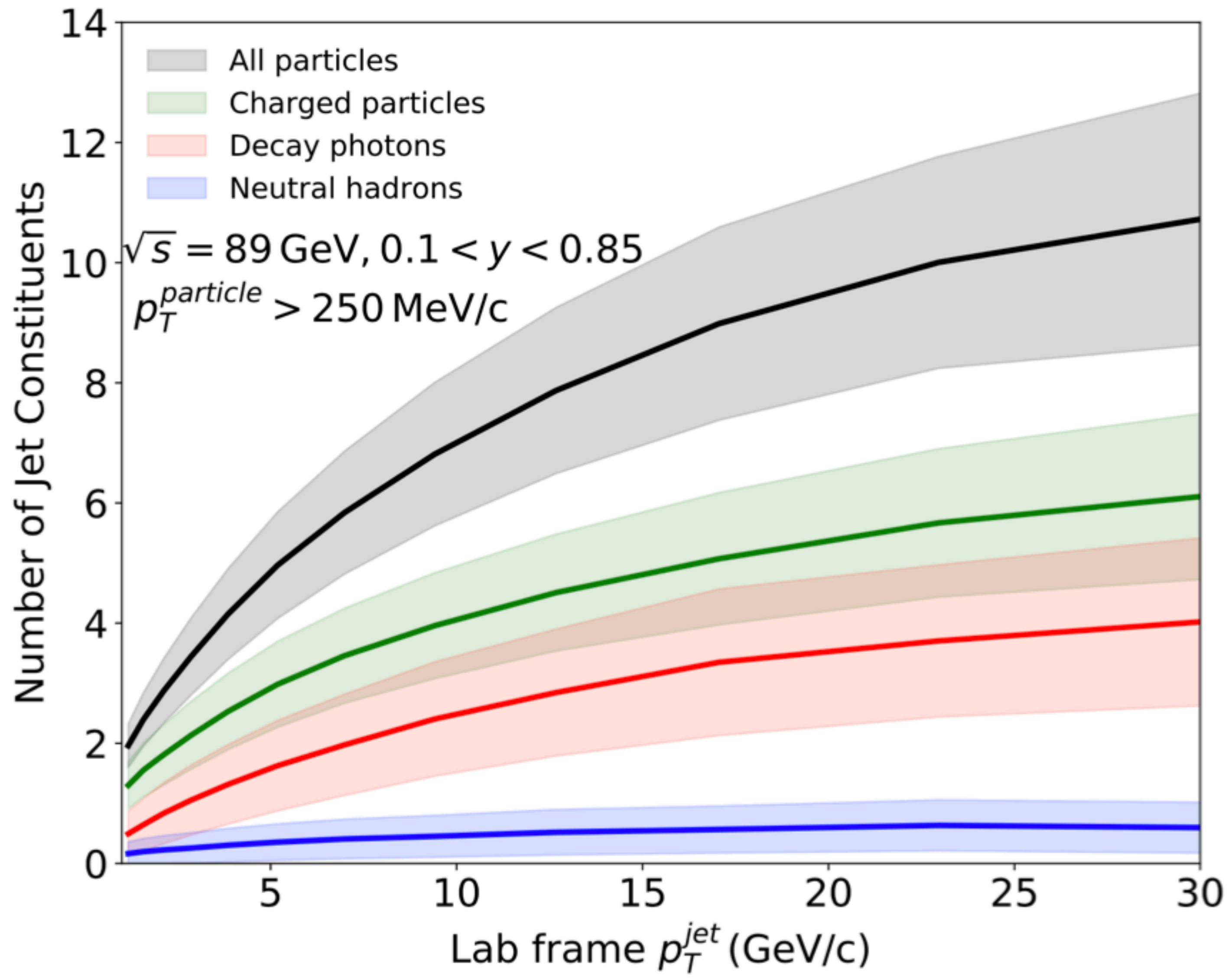


- Note: Not limited to jets

Transverse momentum



Particle #



Machine learning setup

Lee, Mulligan, Ploskon, FR, Yuan `22

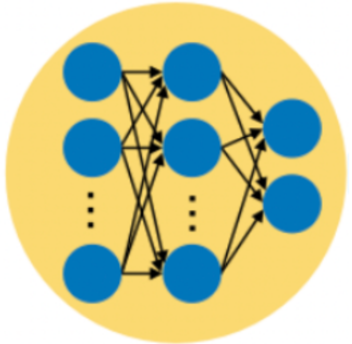
- Binary classification tasks

- Particle Flow Networks

see Komiske, Metodiev, Thaler JHEP 01 (2019) 121
Permutation invariant Deep Sets

$$f(p_1, \ldots, p_M) = F\left(\sum_{i=1}^{M} \Phi(p_i)\right)$$

Classifier

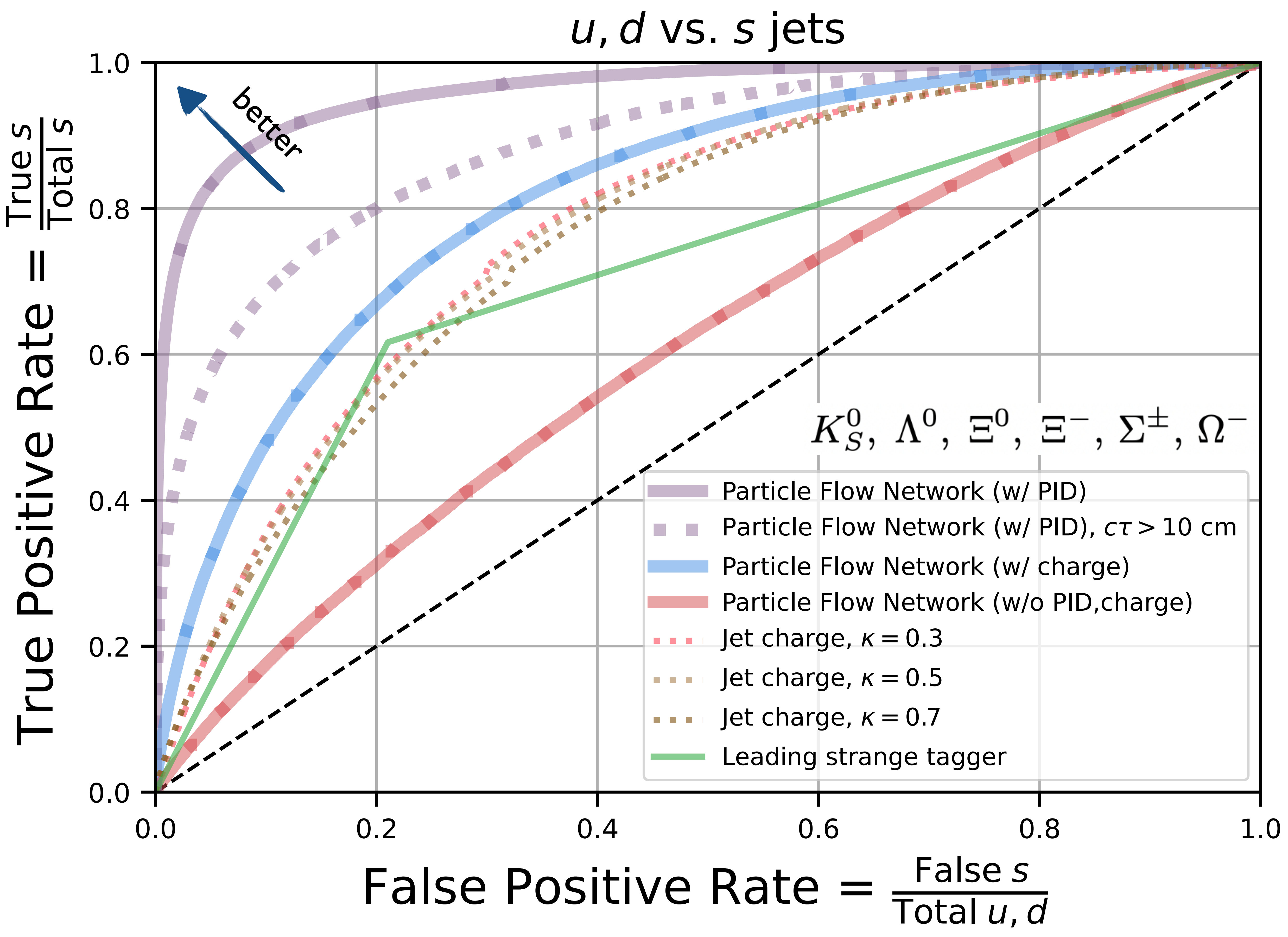


Neural networks

- Possible improvements: Graph neural networks, transformers

Example: strange jet identification

Lee, Mulligan, Ploskon, FR, Yuan `22



Significant gain with machine learning!

- Quantifies total information content
- Motivates further theory efforts
- Soft particles, tracking & PID important
- Impact on EIC detector?

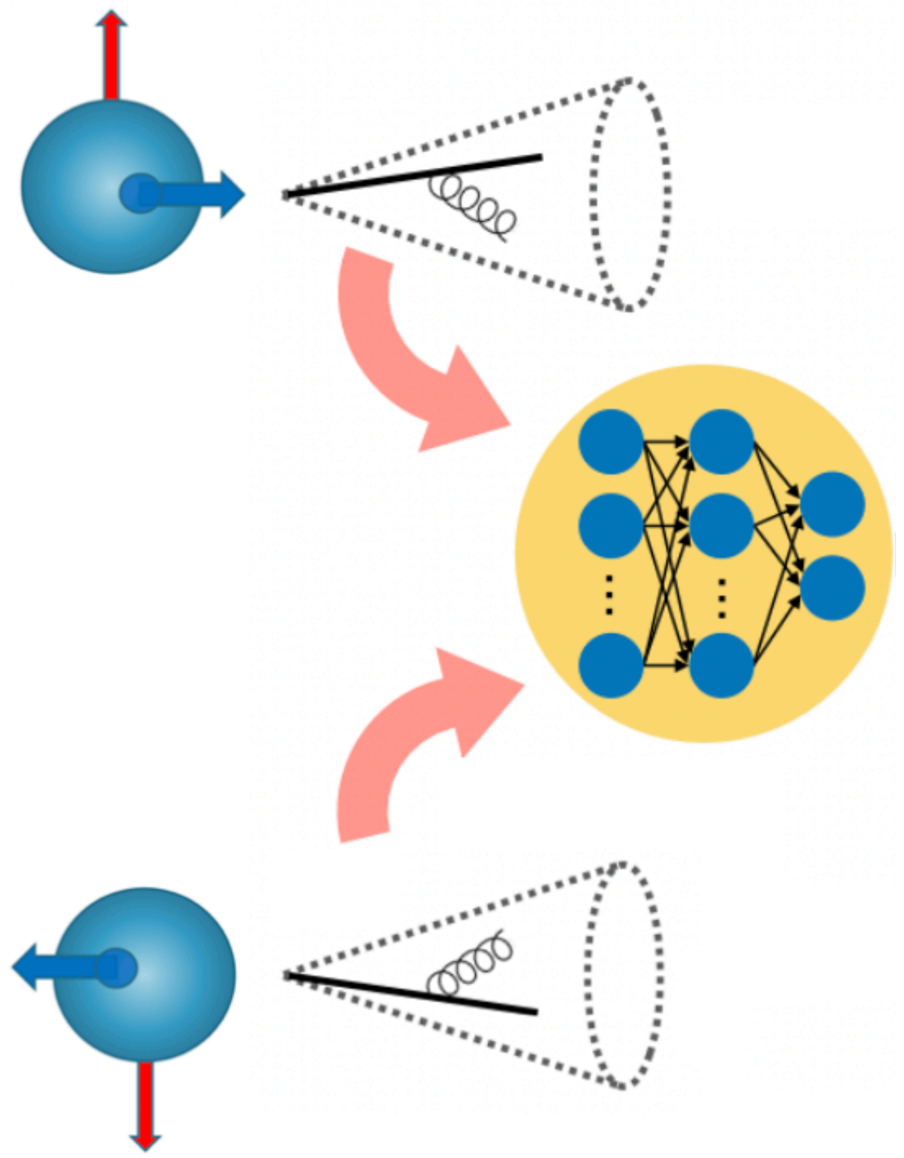
Data & code available

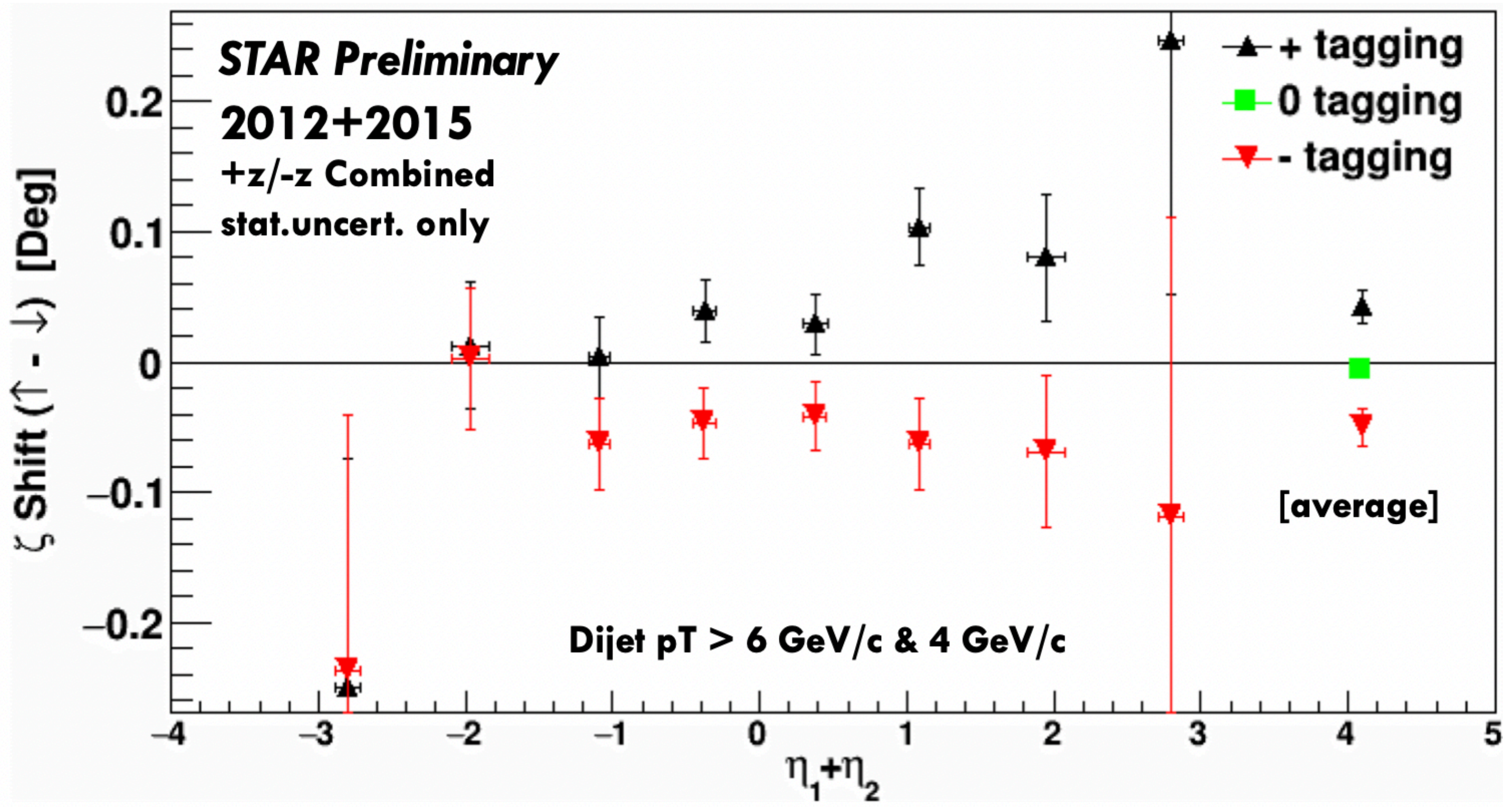
https://zenodo.org/record/7538810#.Y8RcaS-B2gQ

ML trained on event-by-event data

- Can we directly train on data?
- Enhance spin asymmetries

Fatemi EINN `19, Liu DNP `19
Lee, Mulligan, Ploskon, FR, Yuan `22





Burkardt sum rule `04

$$\sum_{a=q,\bar{q},g} \int_0^1 \mathrm{d}x\, f_{1T}^{\perp(1)a}(x) = 0$$

- Train classifier on jets in collisions with different initial proton spin → effectively $\max_\theta |A_{UT}(\theta)|$

Can potentially obtain better constraints on spin PDFs

Generative modeling

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

High quality samples

Mode coverage

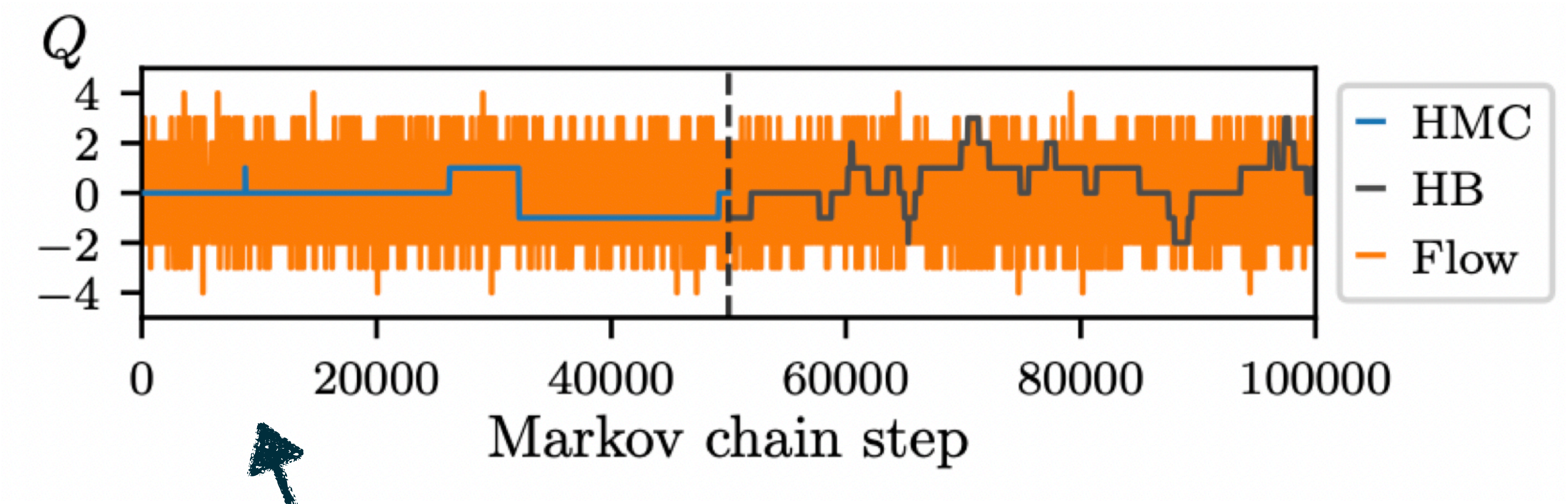
Fast sampling

Stable training

Normalizing flows for lattice field theory

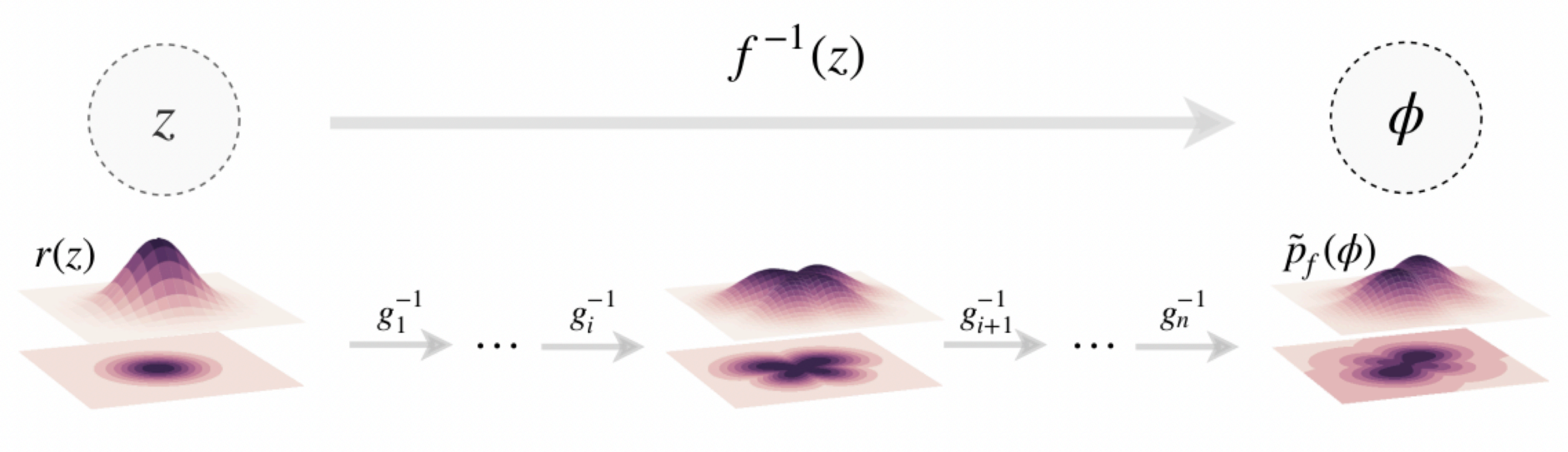
Shanahan et al

- Sampling of lattice gauge field configurations
- Multi-modal distributions
- Asymptotically exact with additional accept/reject step



U(1) & extended to QCD

Invertible map with tractable Jacobian



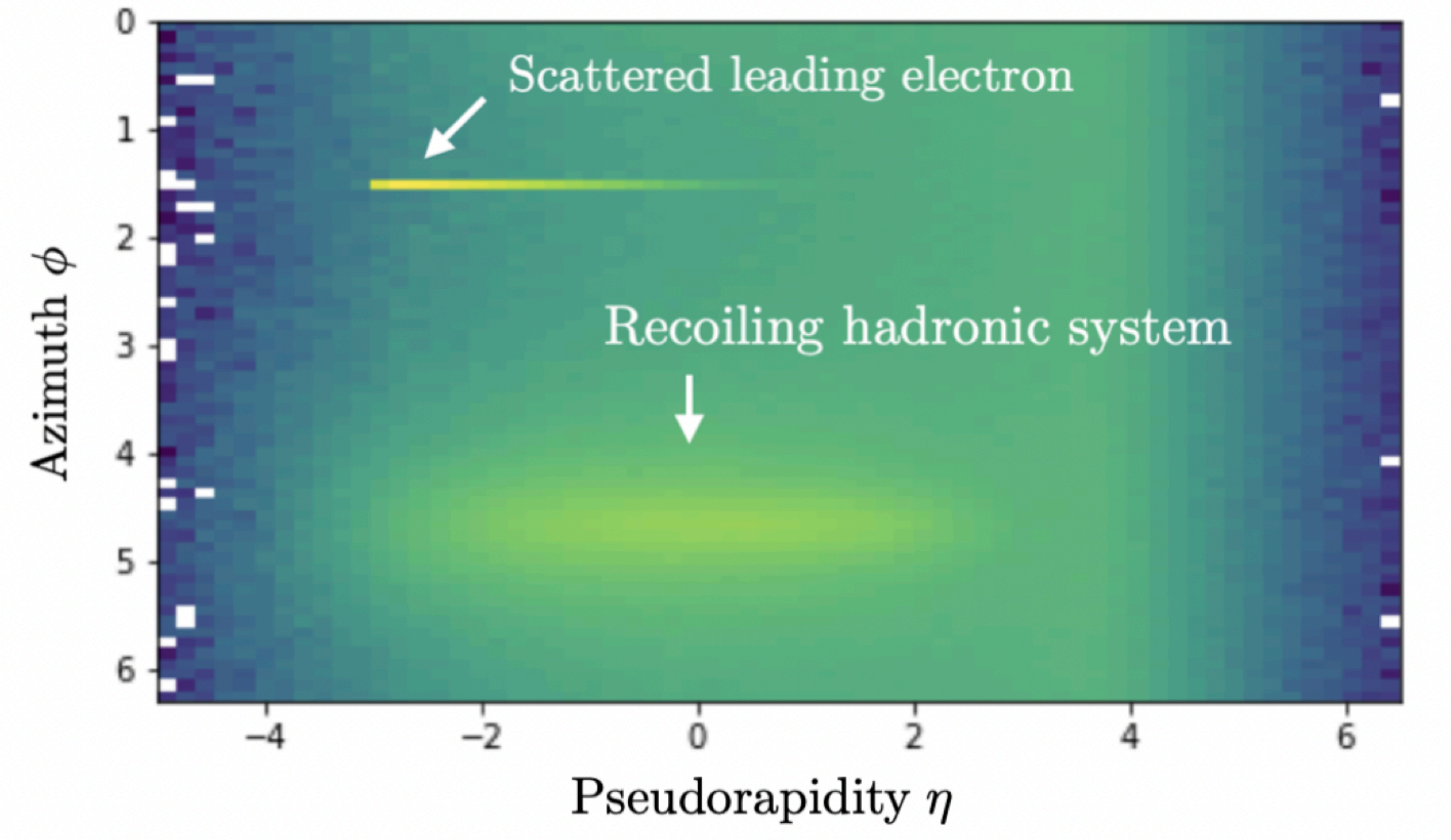
Applications to multi-loop integrals see Butter et al.

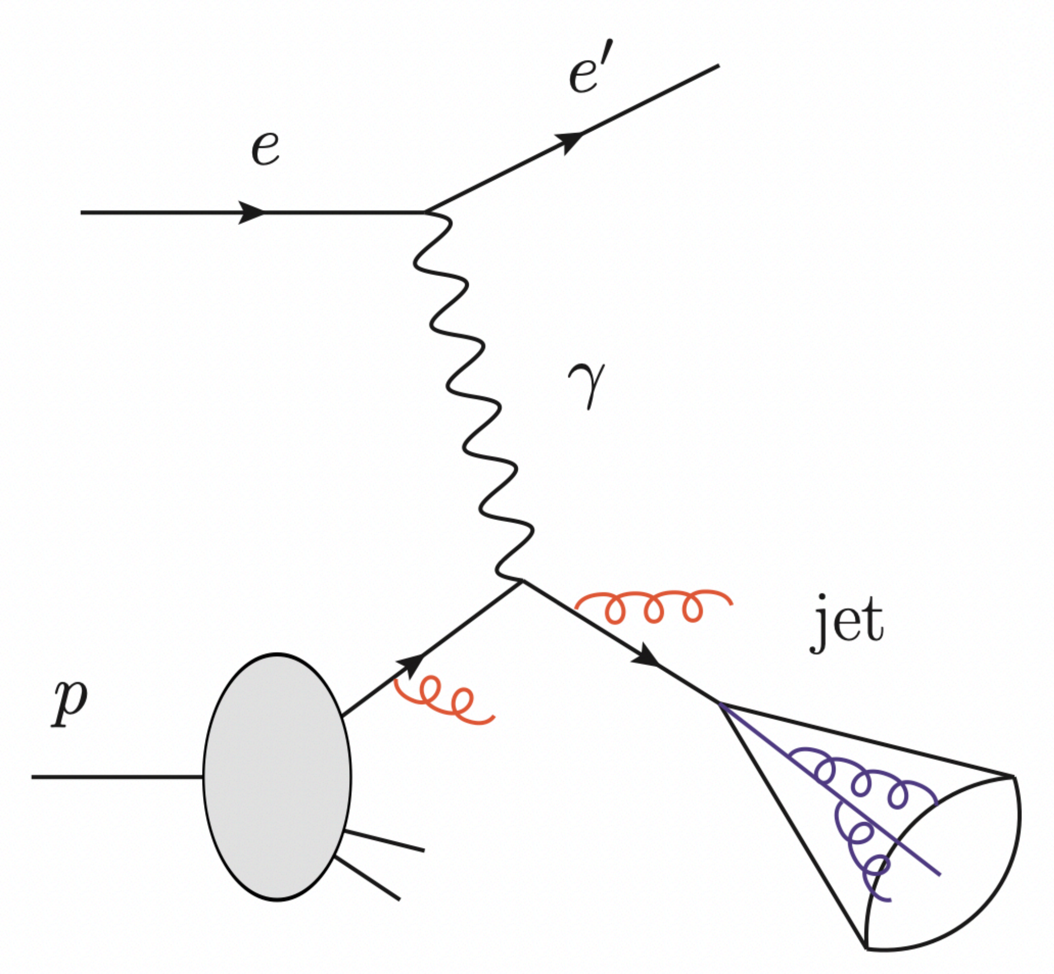
Simulating electron-proton scattering events

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

  → Develop a generative model

- Development of MC event generators
- Searches of BSM physics
- Event-level data analysis



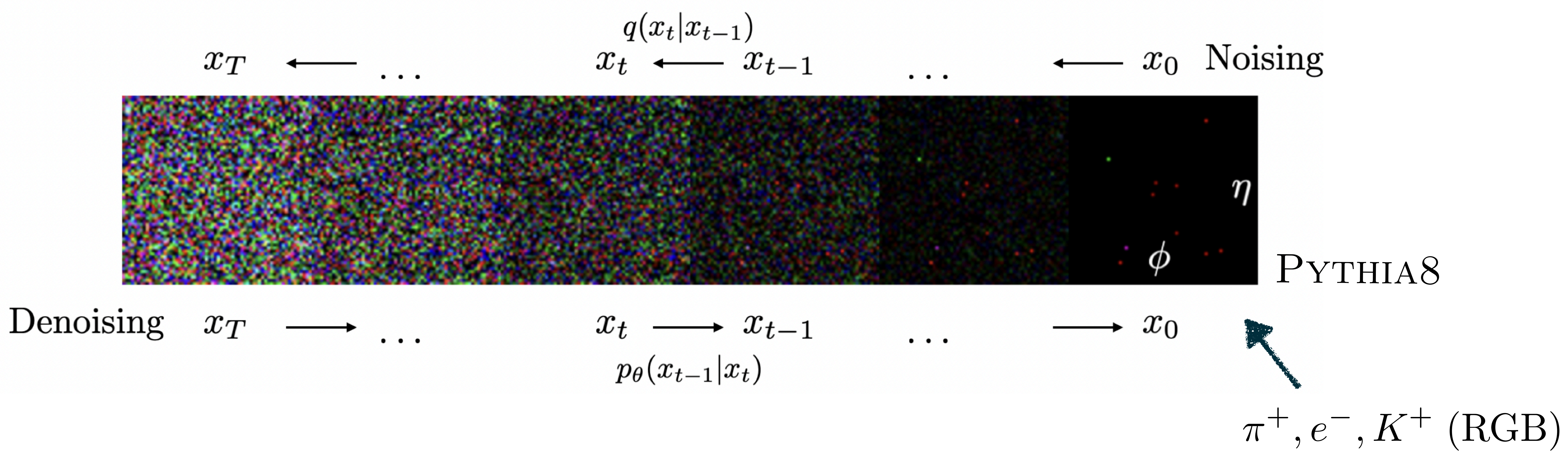


$\textsc{Pythia}8,\ Q > 10\ \mathrm{GeV}$

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

Diffusion models

- Represent events as images (pixelated)



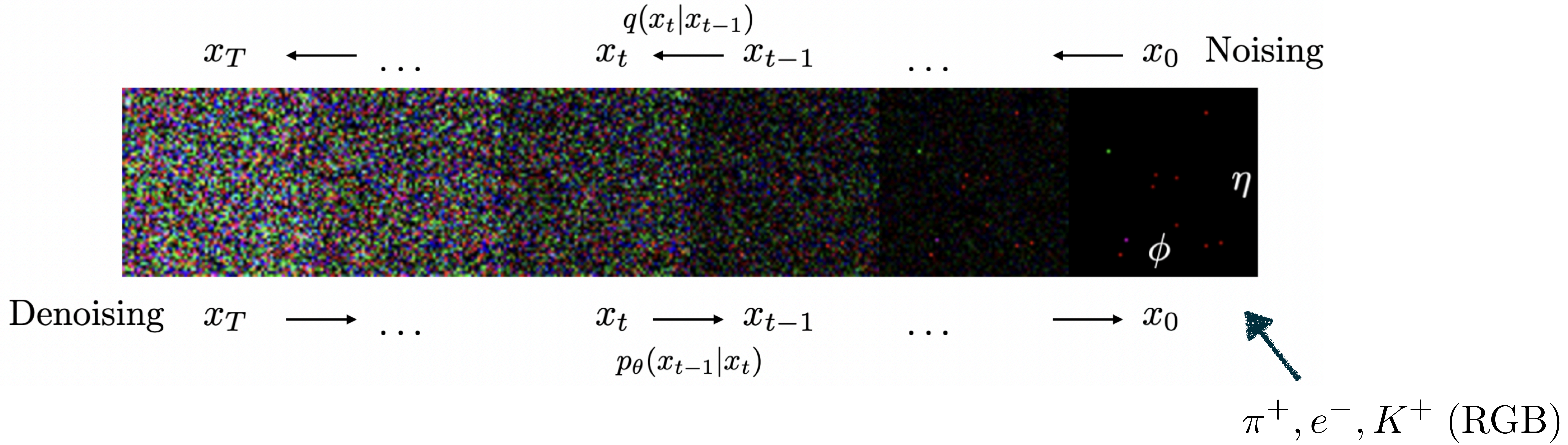
- Markovian noising process $\quad q(x_1, \dots, x_T|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1})$

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

Devlin, Qiu, FR, Sato `23

Diffusion models

- Represent events as images (pixelated)



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

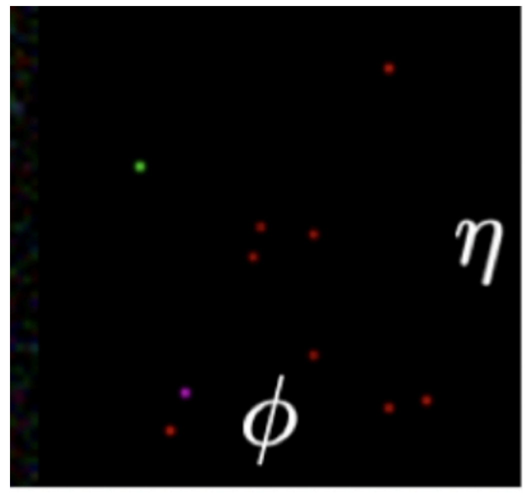
Stochastic differential equation in continuum limit

Train convolutional U-Net

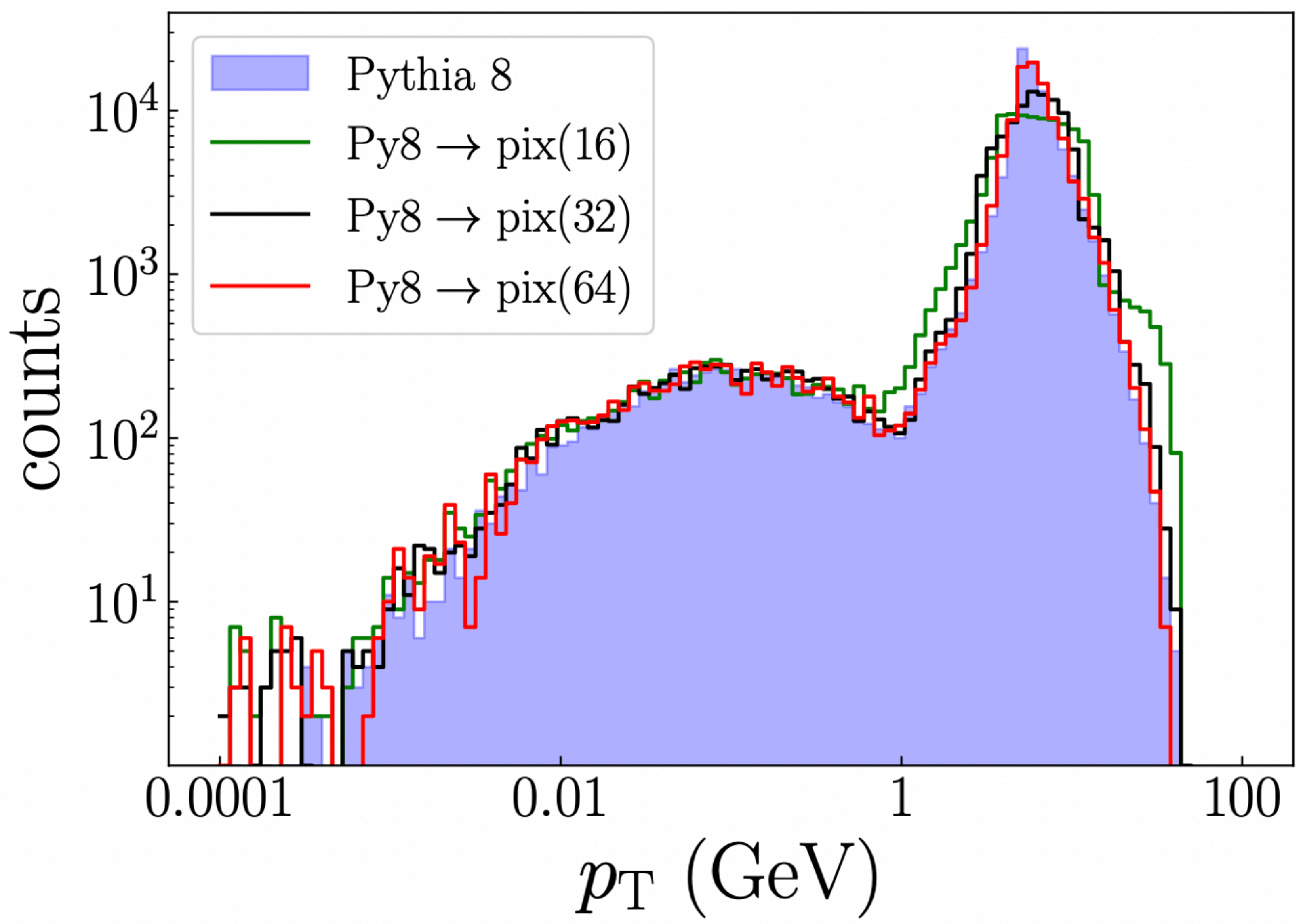
Devlin, Qiu, FR, Sato `23

Simulating electron-proton scattering events

- Sparse data



- Steeply falling distributions



→ Use suitable rescaling

- Momentum variable $z_i = \frac{2p_{Ti}}{\sqrt{s}} \cosh \eta_i$

Event-wide momentum conservation

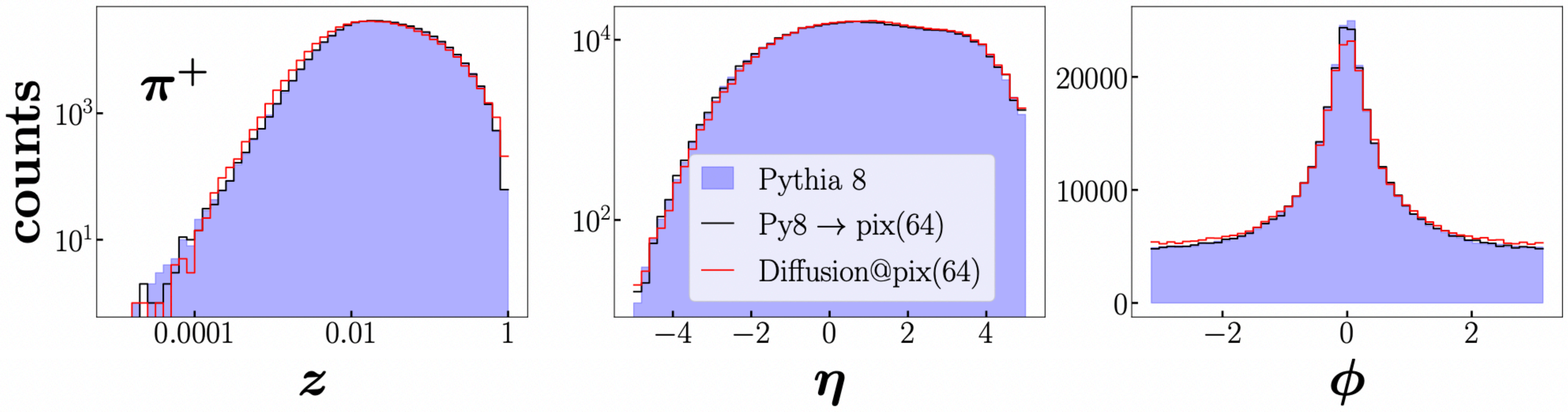
with $\sum_{i \in \text{event}} z_i \leq 2$

Devlin, Qiu, FR, Sato `23

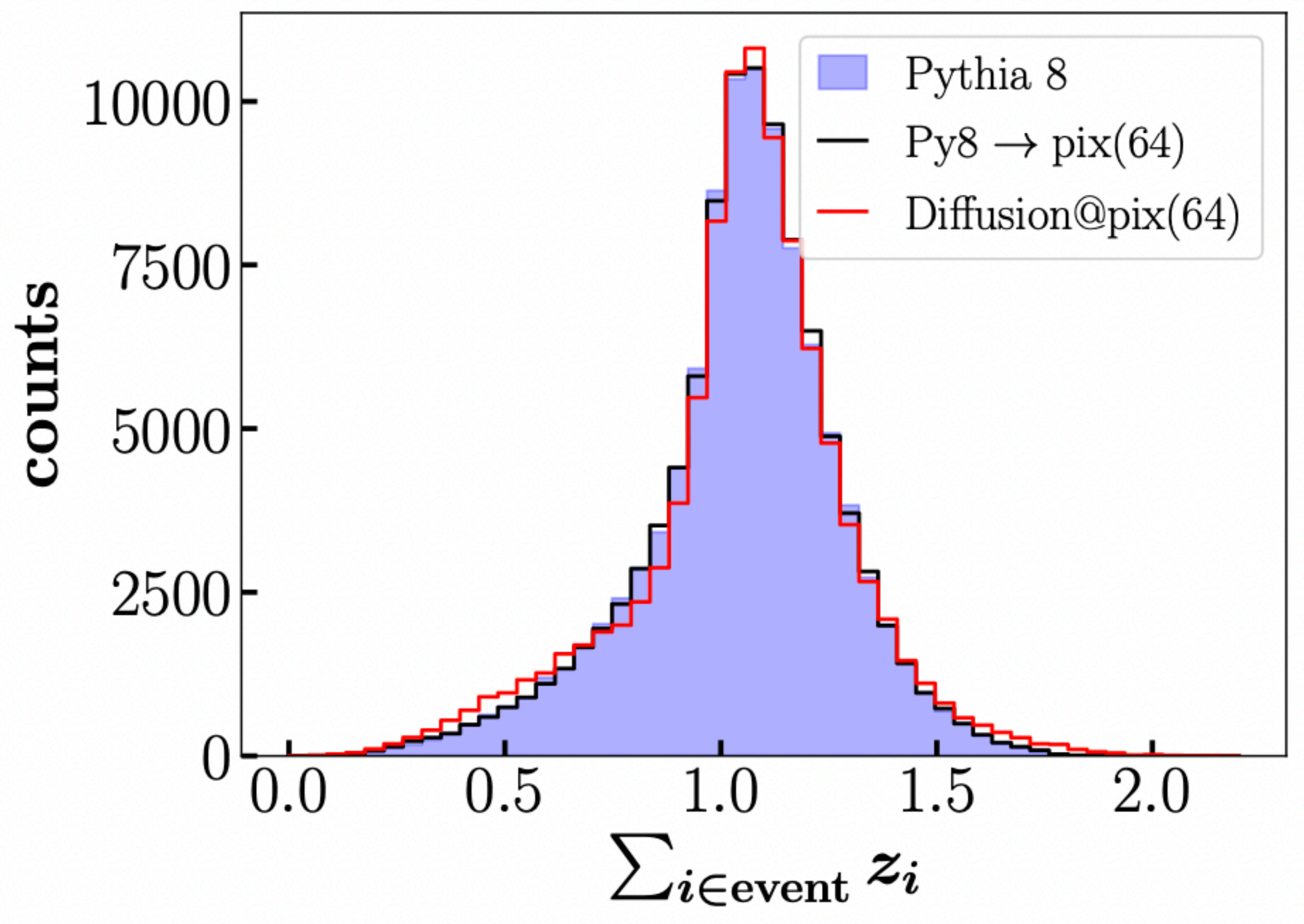
Simulating electron-proton scattering events

Devlin, Qiu, FR, Sato `23

- Momentum & angular distributions



- Momentum sum rule



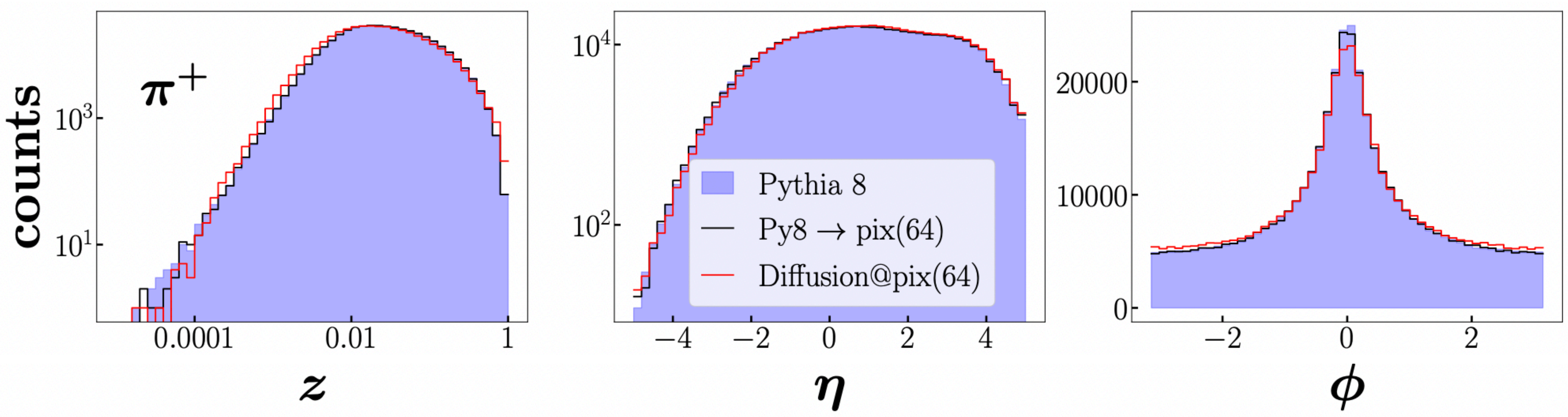
Event-wide constraint learned by the model

$$\sum_{i\in\text{event}} z_i \leq 2$$

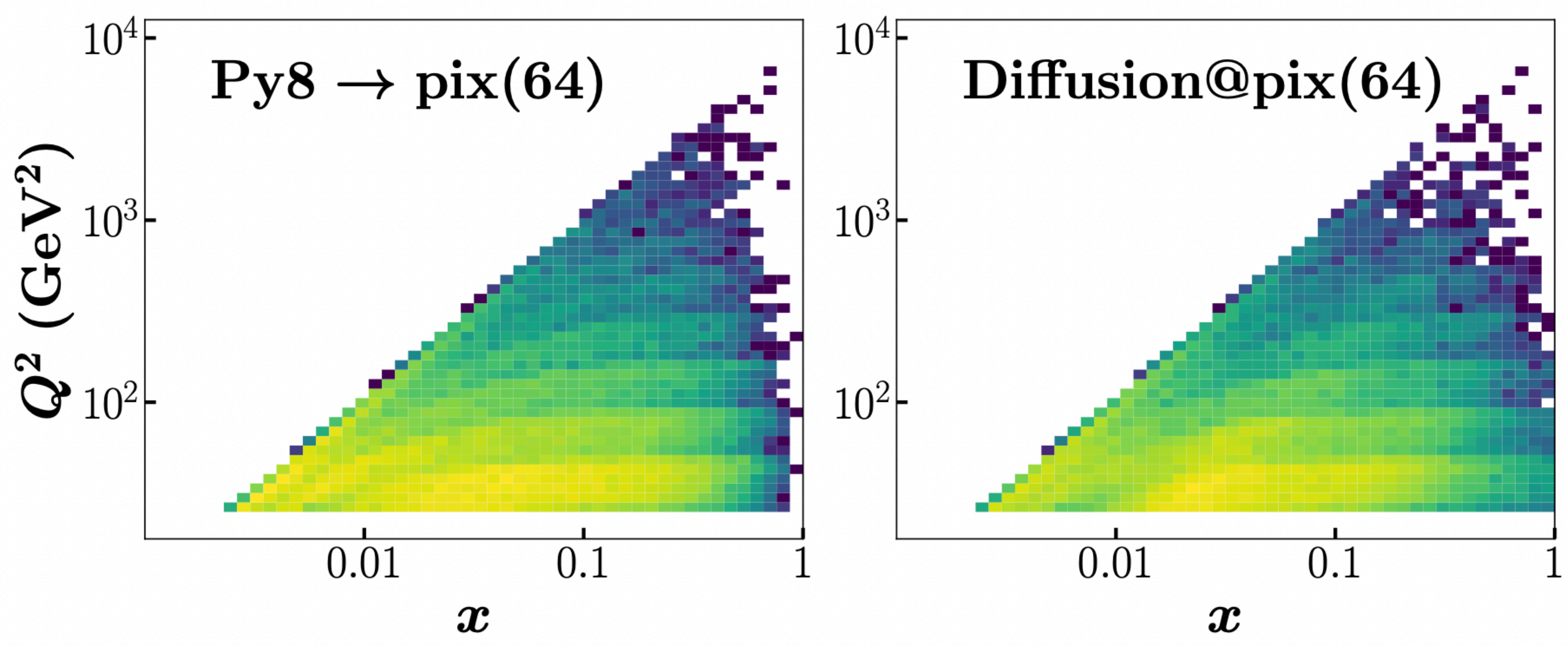
Simulating electron-proton scattering events

Devlin, Qiu, FR, Sato `23

- Momentum & angular distributions

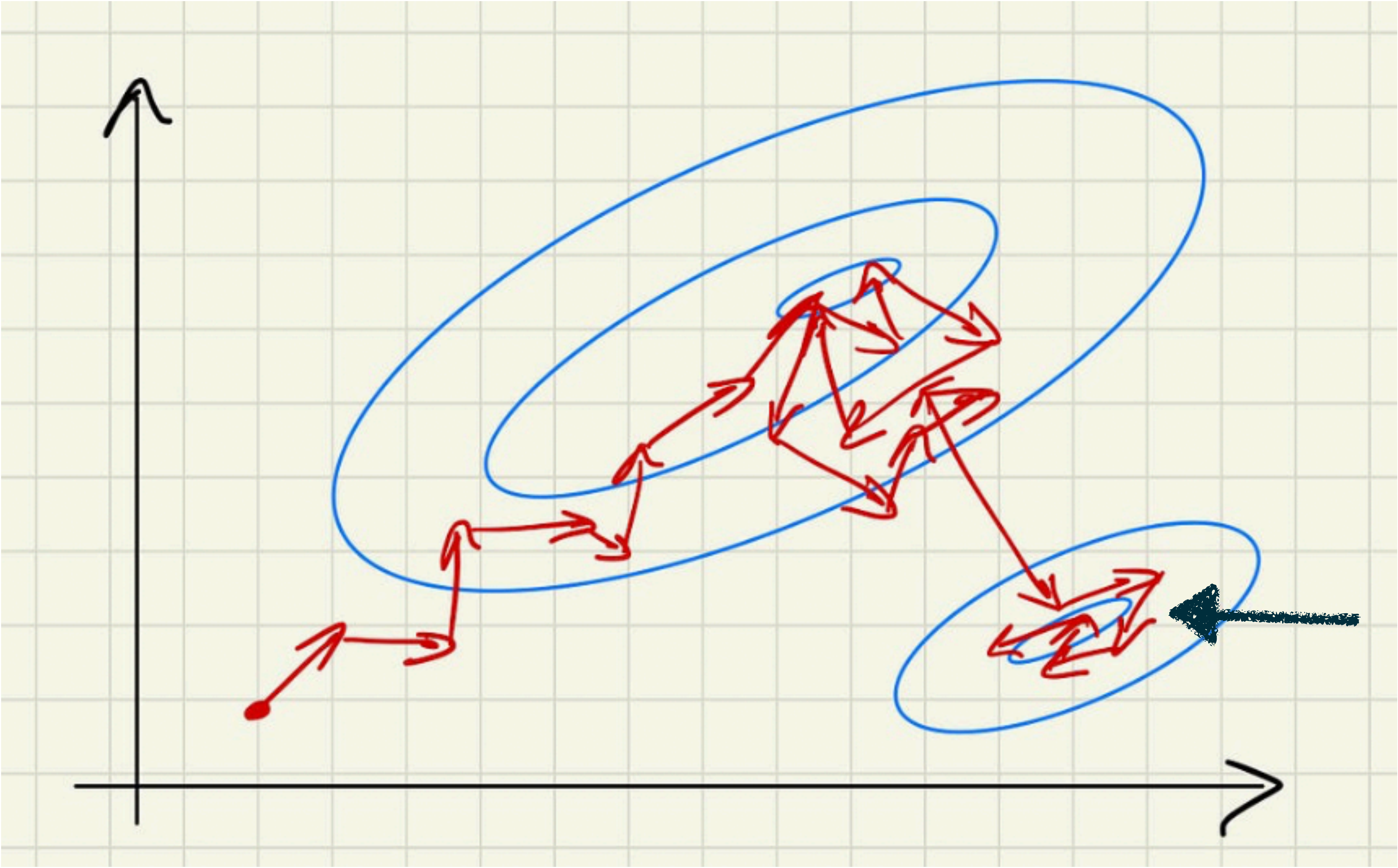


- DIS kinematics



MCMC sampling with diffusion models

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

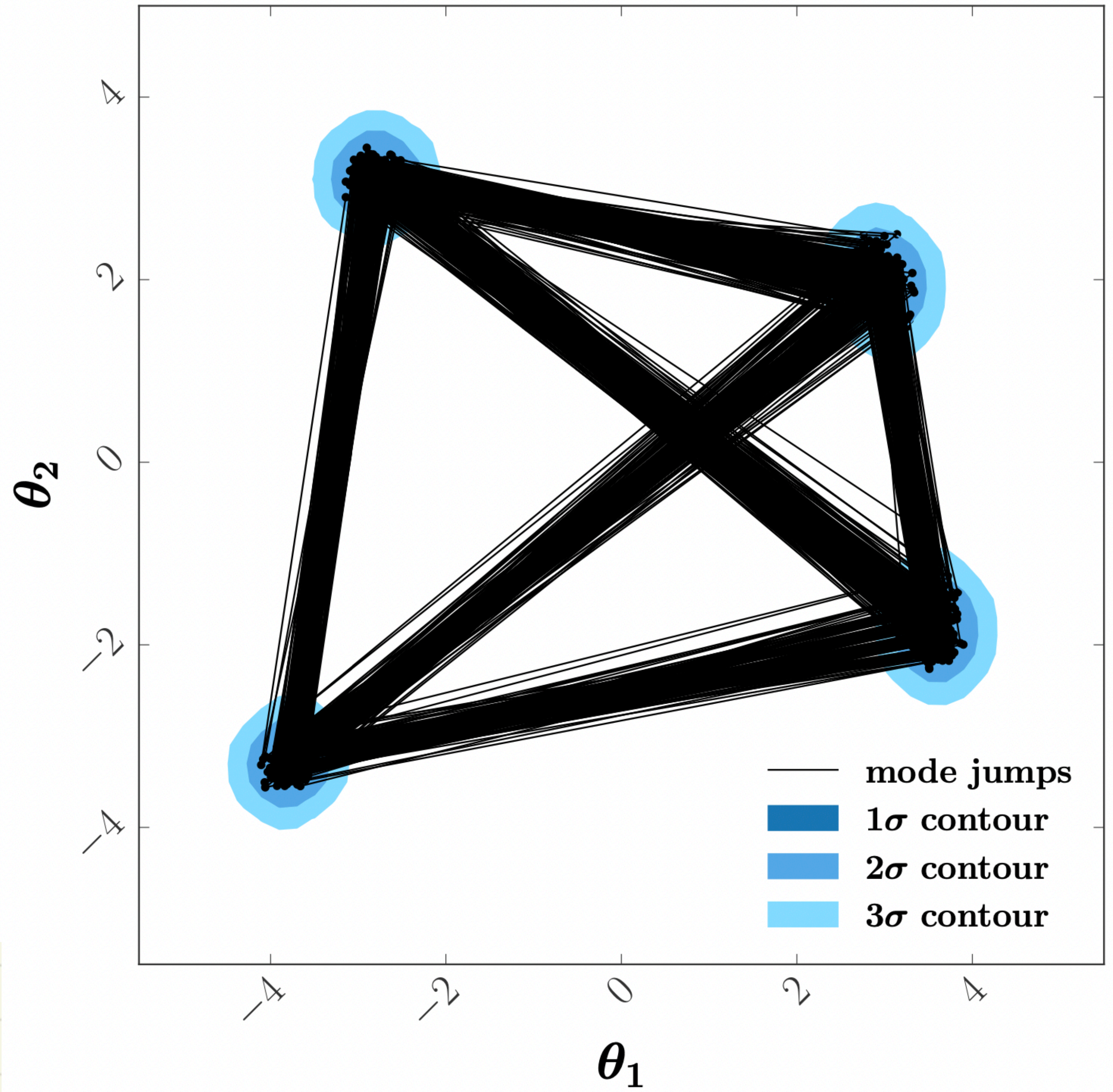


Gaussian proposal function

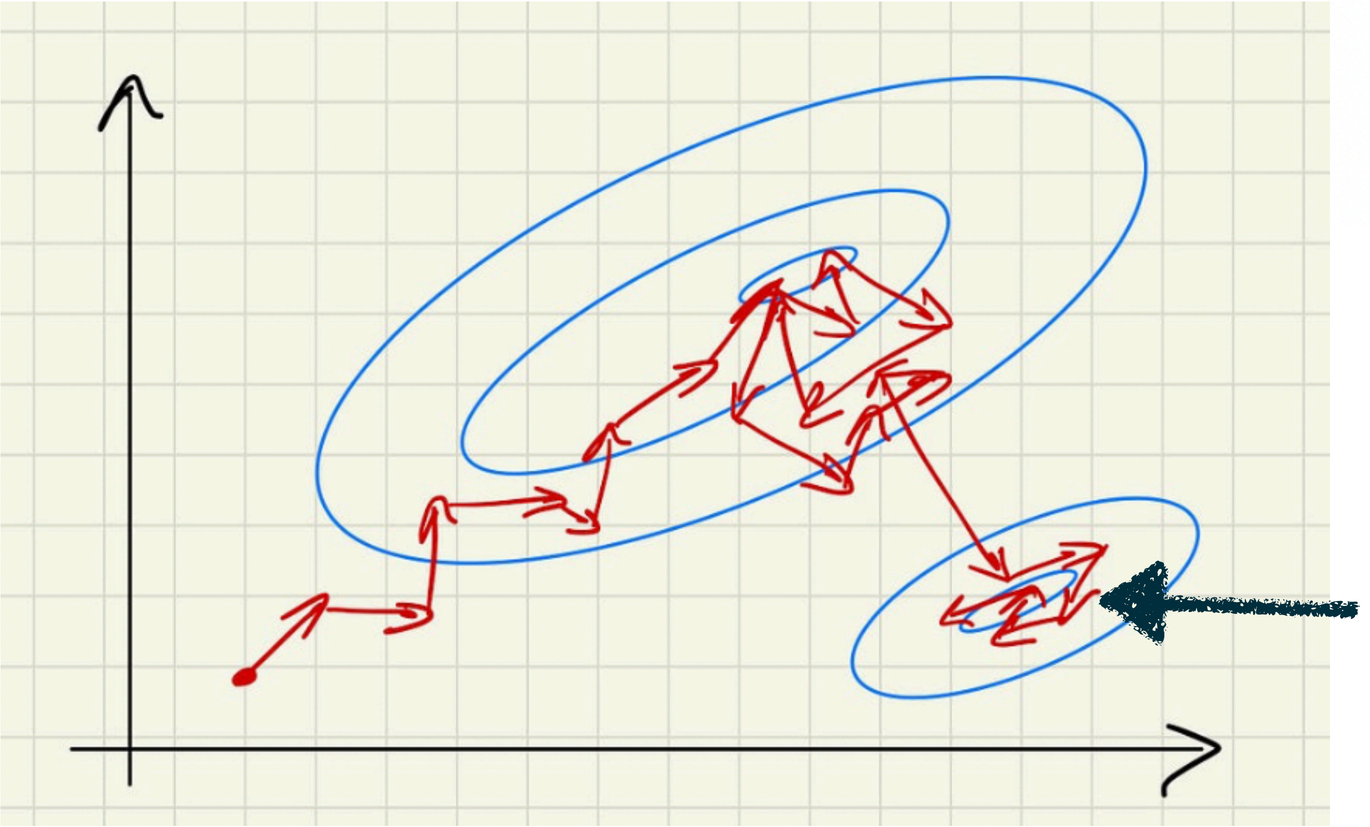
Hunt-Smith, Melnitchouk, FR, Sato, Thomas, White `23

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

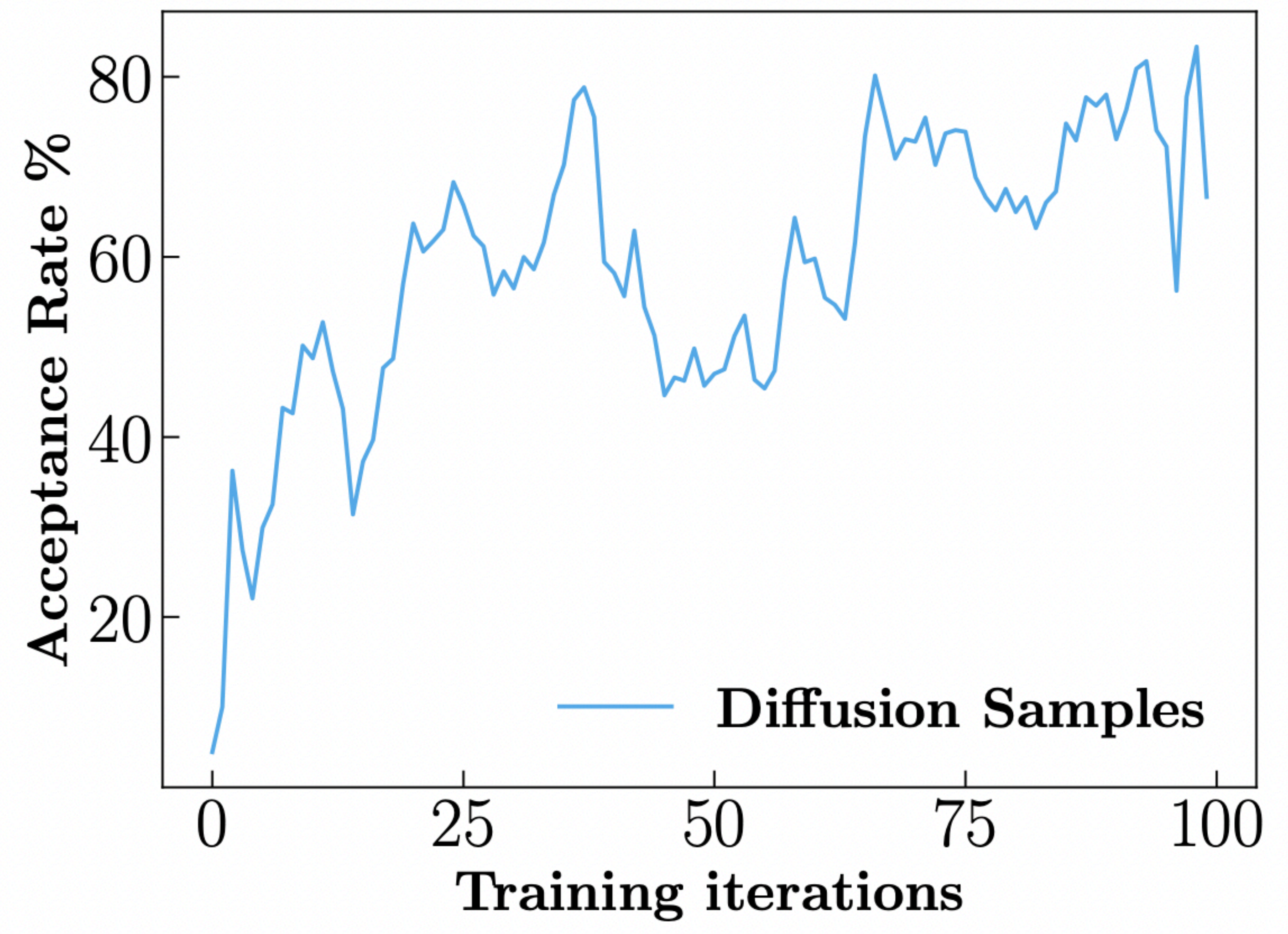
MCMC sampling with diffusion models

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

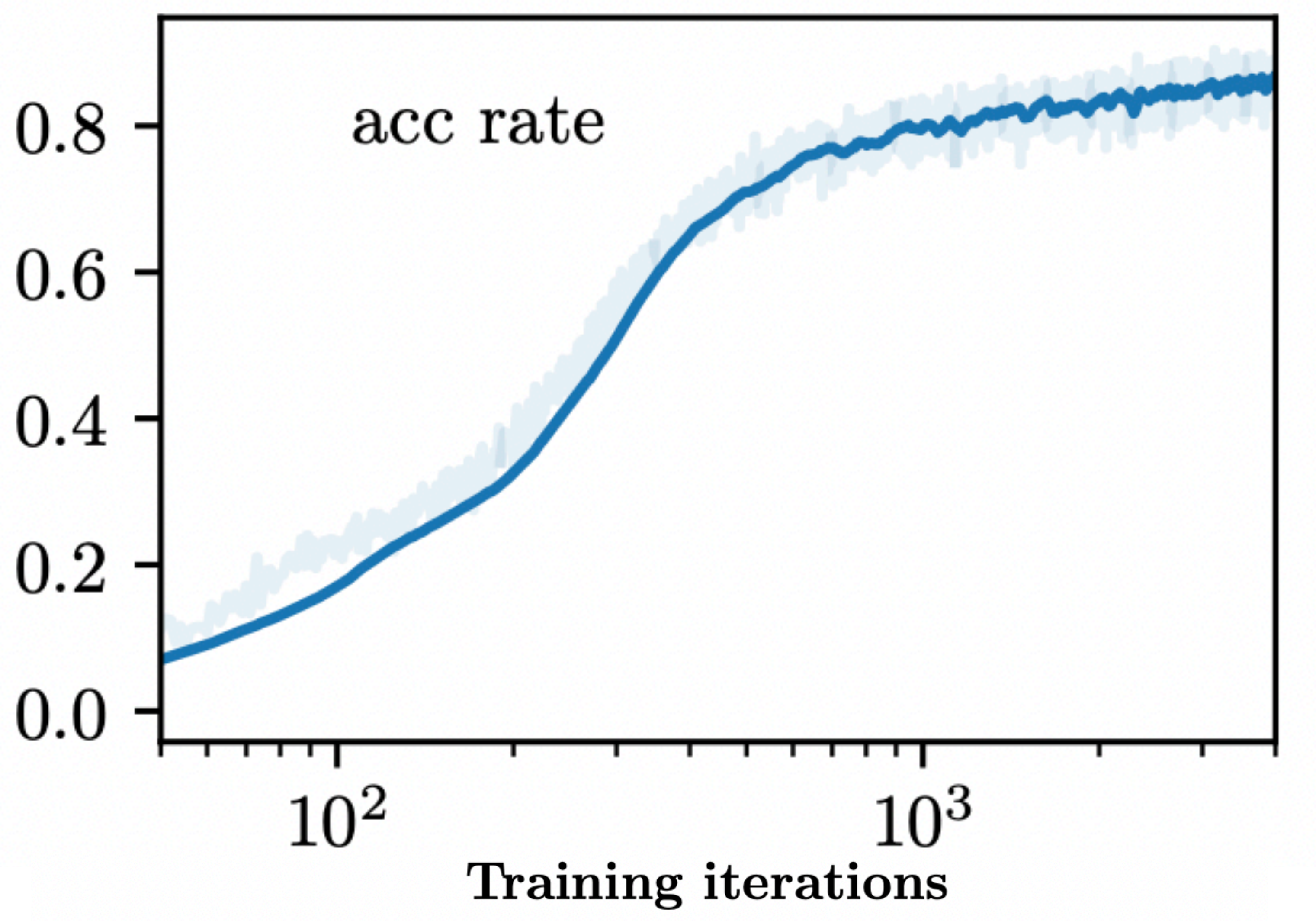
10d Gaussian mixture

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

Acceptance rate



Diffusion model



Normalizing flow

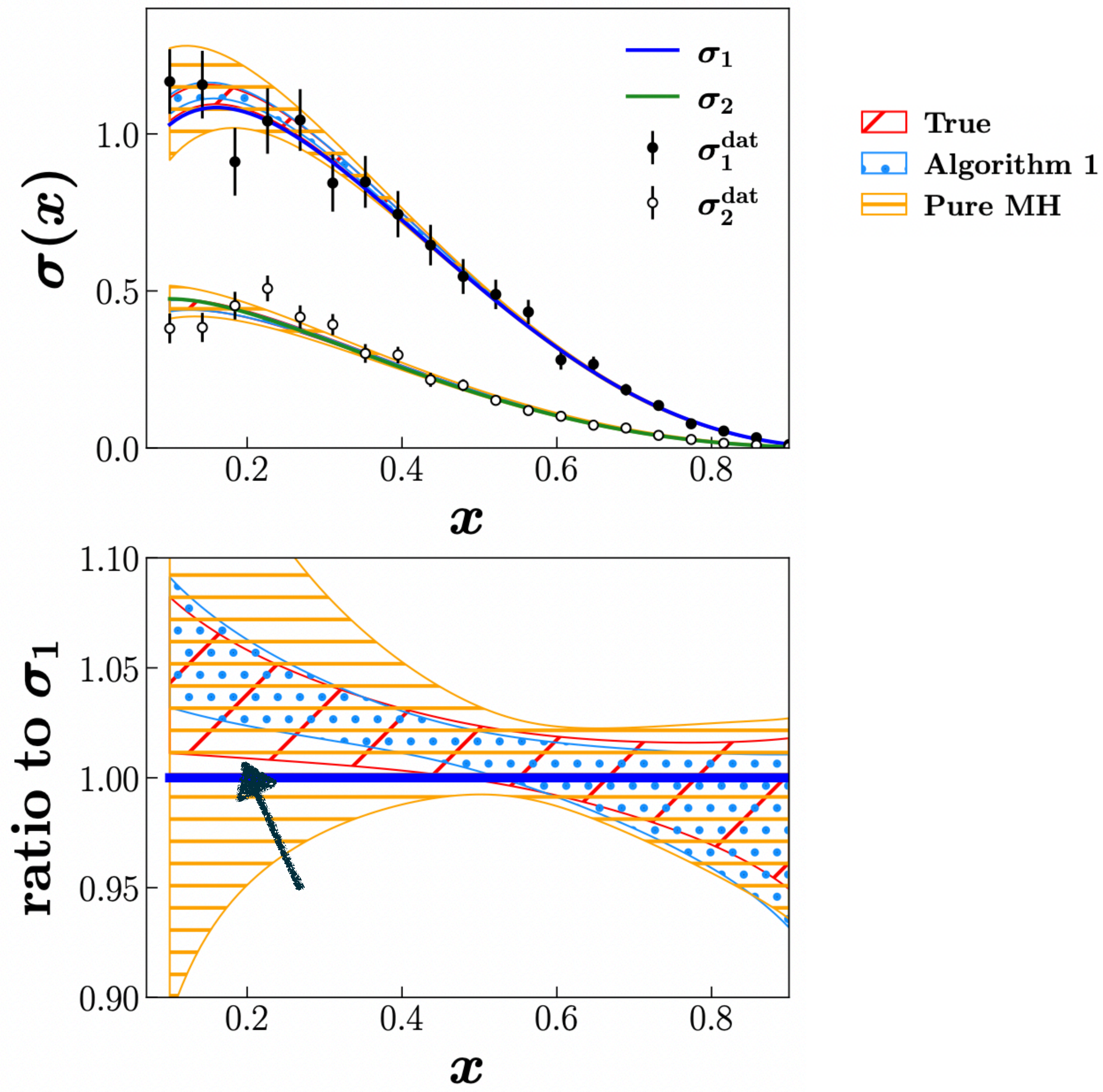
Additionally, it requires probability for accept/reject step

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



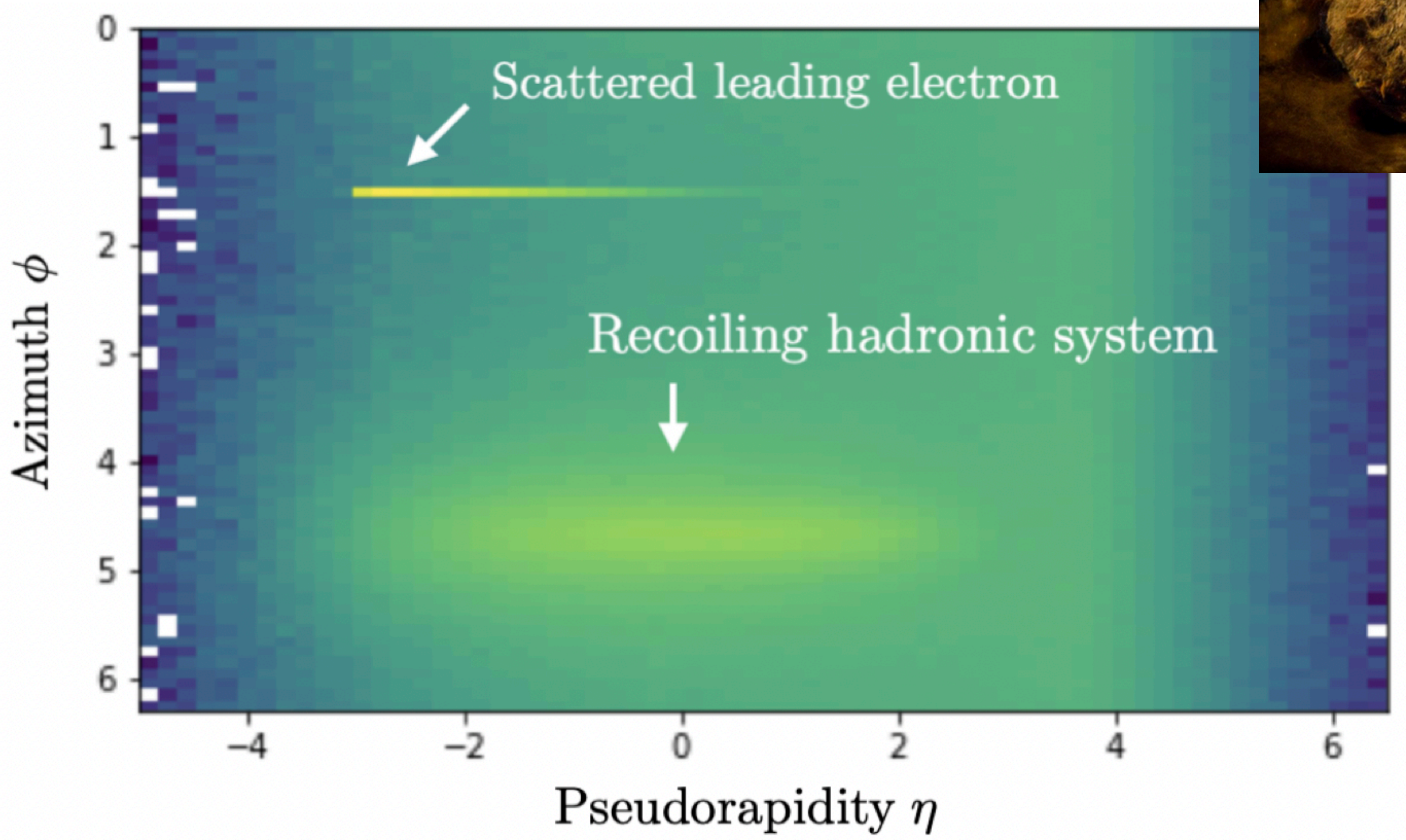
Fast convergence of diffusion model-assisted MCMC

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

Conclusions & outlook

- Various new applications of AI in fundamental physics
- Classification, regression, generative modeling
- Also multi-loop integrals, sign problems, nuclear structure etc.

- Physics-inspired learning theory
- EFTs for neural networks

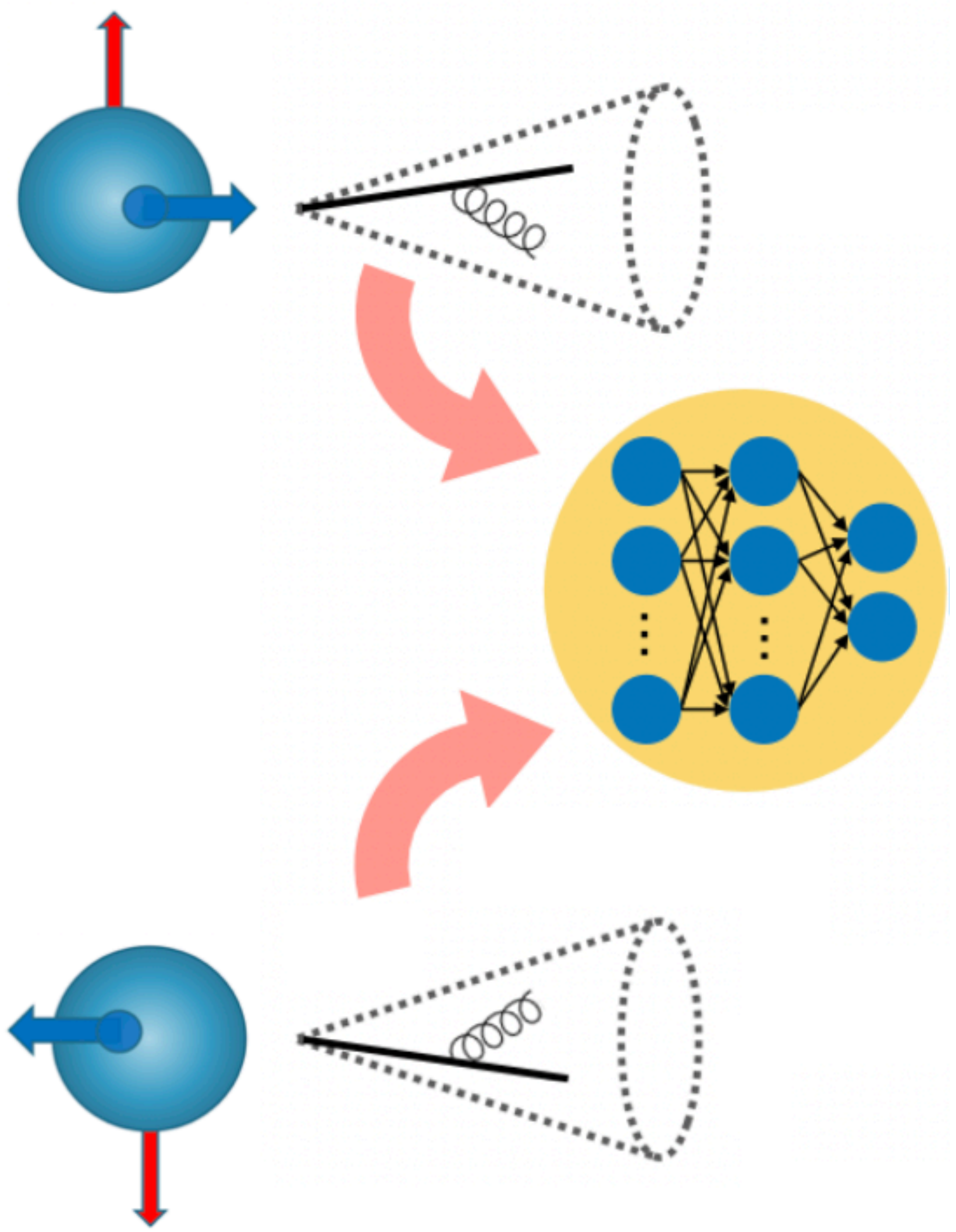


Hadron structure & spin physics

Lee, Mulligan, Ploskon, FR, Yuan `22

- How can we apply these techniques to hadron structure & spin physics?

1. Supervised machine learning
2. Train on data e.g. $A_{UT} = \dfrac{\mathrm{d}\sigma^{\uparrow} - \mathrm{d}\sigma^{\downarrow}}{\mathrm{d}\sigma^{\uparrow} + \mathrm{d}\sigma^{\downarrow}}$



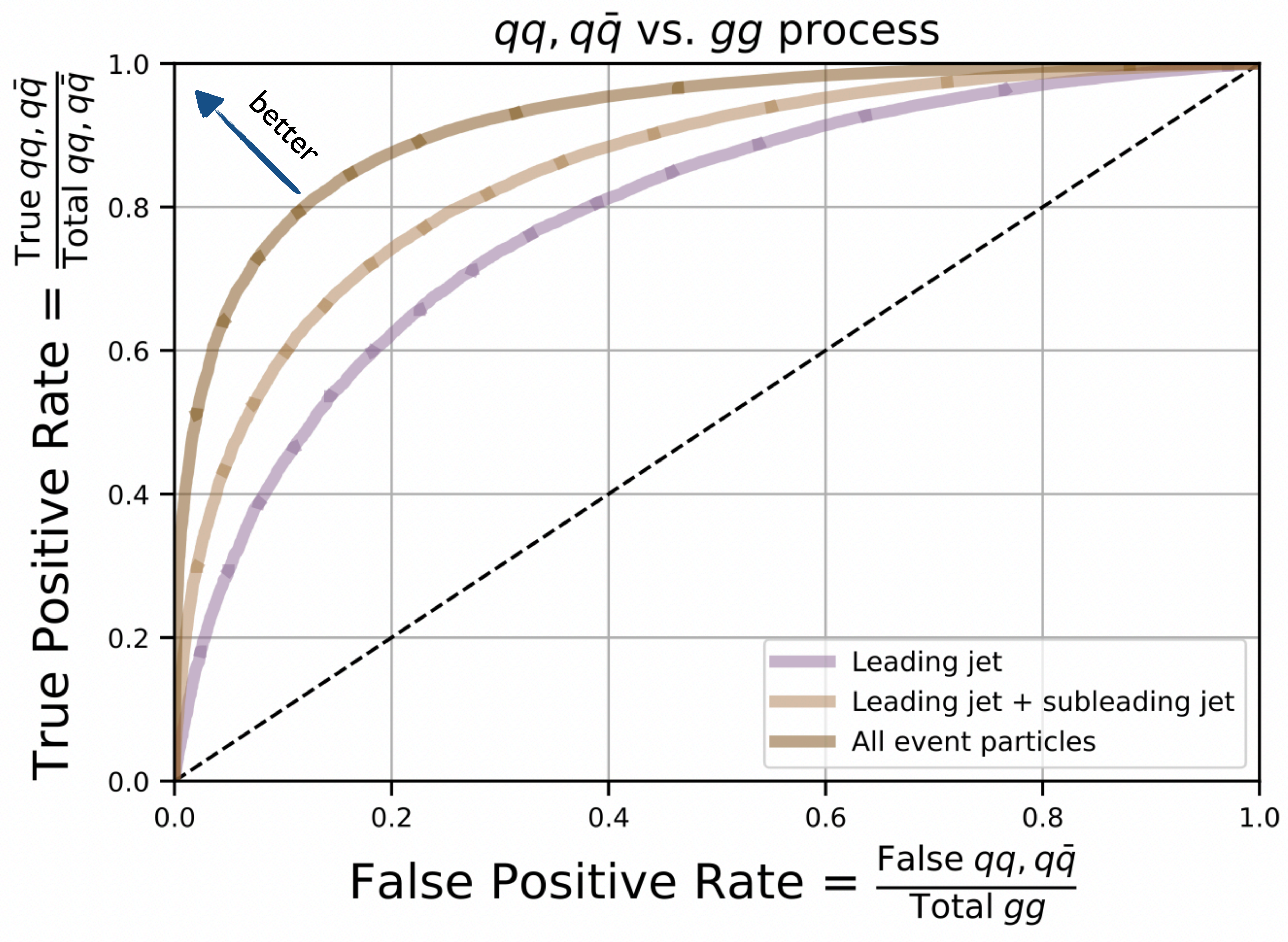
- Reformulate regression task as classification problem $\max_{\theta} |A_{UT}(\theta)|$

→ Upper limit on what can possibly be achieved

→ Identify new observables

Example: quark vs. gluon scattering

Lee, Mulligan, Ploskon, FR, Yuan `22



Photoproduction region

Significant gain with machine learning!

- Quantifies total information content
- Motivates further theory efforts
- Soft particles, tracking & PID important
- Impact on EIC detector?

Data & code available

https://zenodo.org/record/7538810#.Y8RcaS-B2gQ

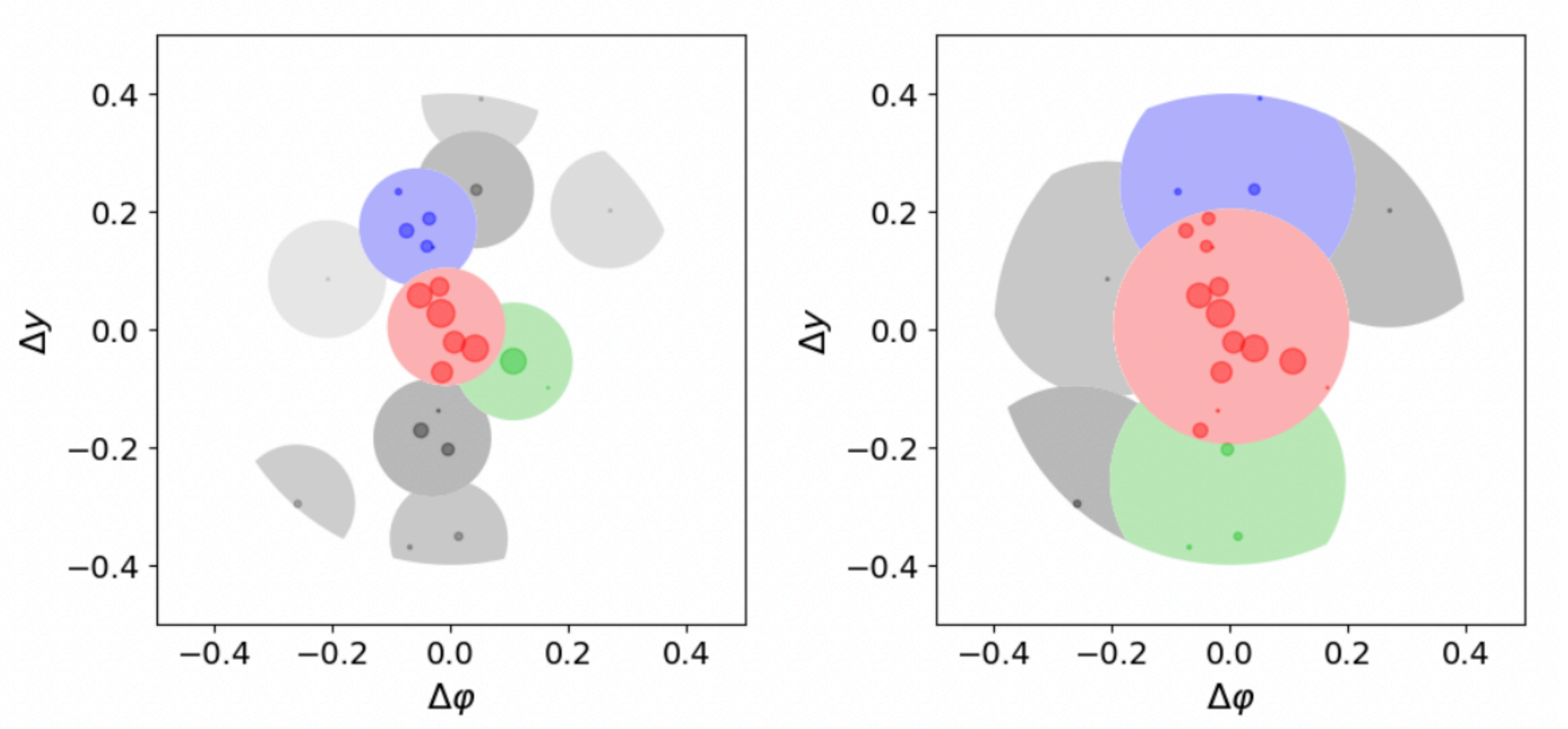
Jet classification & IRC safety

- Can we make use of all this additional information?

- Several jet classification tasks are IRC safe → we can find tractable observables in pQCD

- Recluster particles into IRC-safe subjets before training ML algorithms

Athanasakos, Larkoski, Mulligan, Ploskon, FR `23
Metodiev, Larkoski `19



Matches IRC-unsafe ML algorithm