Flow Annealed Importance Sampling Bootstrap meets Differentiable Particle Physics

Nachman Group @Berkley, 28.01.2025



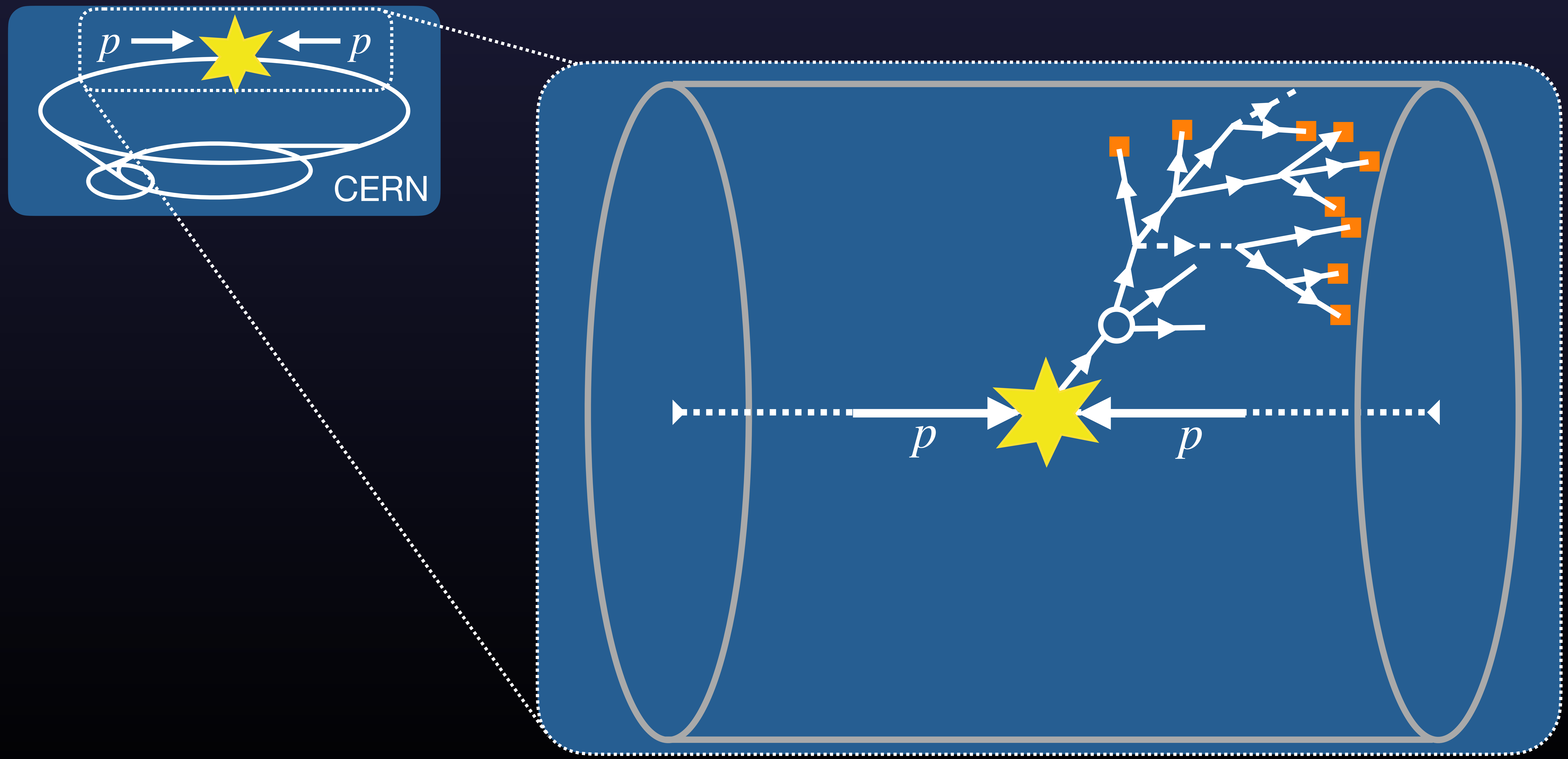


Annalena Kofler[1,2,3] Vincent Stimper[1,4,5] Mikhail Mikhasenko[6,7] Michael Kagan[8] Lukas Heinrich[3]

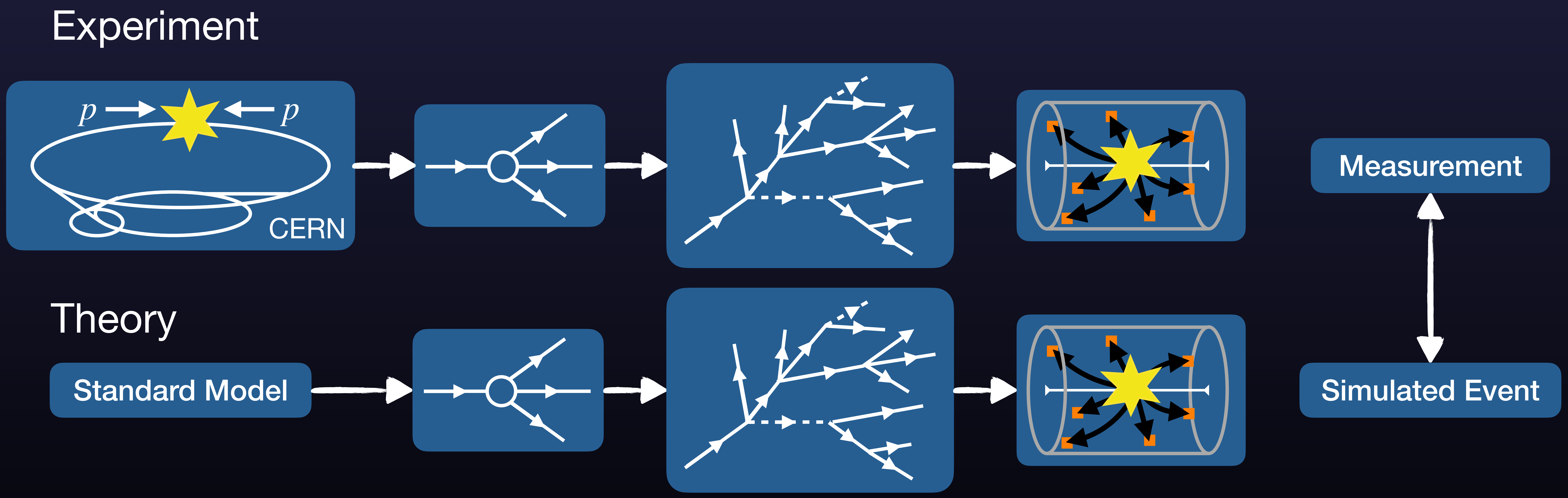
[1]Max Planck Institute for Intelligent Systems Tübingen, [2]Max Planck Institute for Gravitational Physics Potsdam, [3]Technical University of Munich, [4]Isomorphic Labs, [5]University of Cambridge, [6]University of Bochum, [7]ORIGINS Excellence Cluster Munich, [8]SLAC National Accelerator Laboratory

How can we learn more about particles?

Experiment



How can we analyze data?



→ Find deviations from theory

How to improve the precision of measurements?

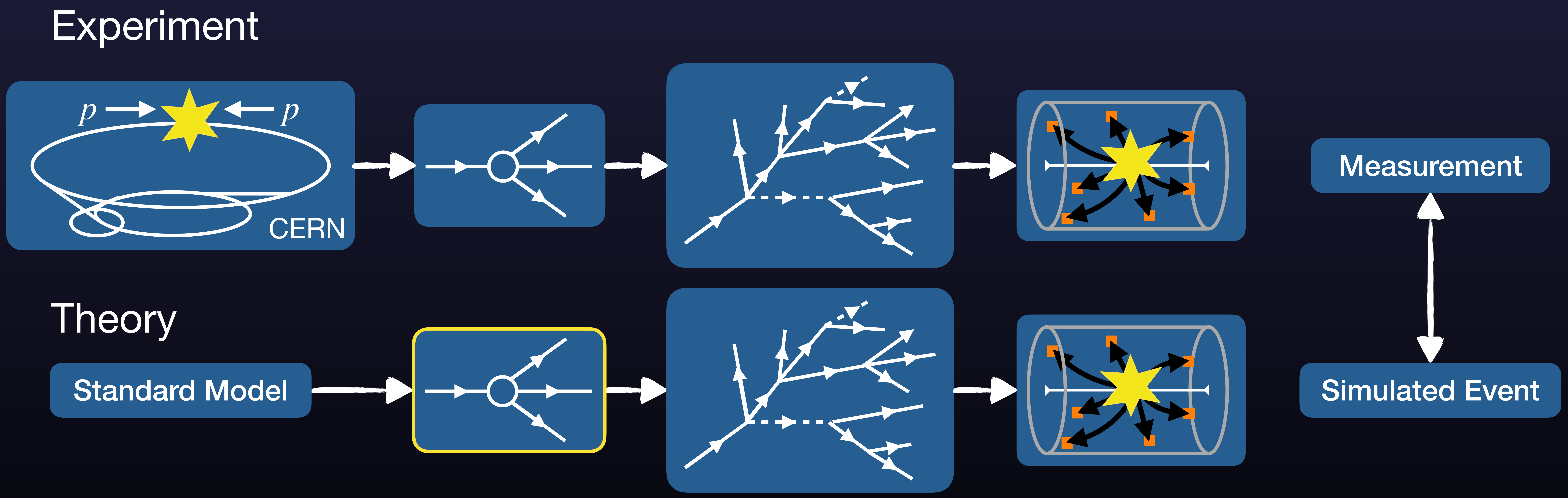
→ Take more data!

Number of events until now

Number of events until 2030

ATLAS Software and Computing HL-LHC Roadmap, 2022.

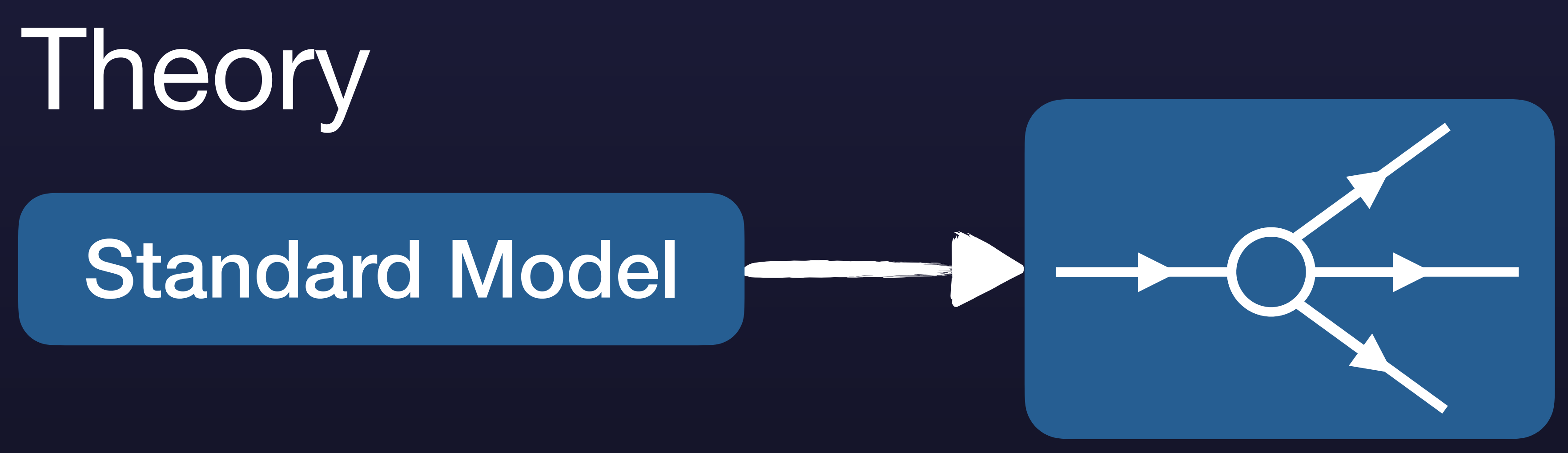
What does this mean for the simulations?



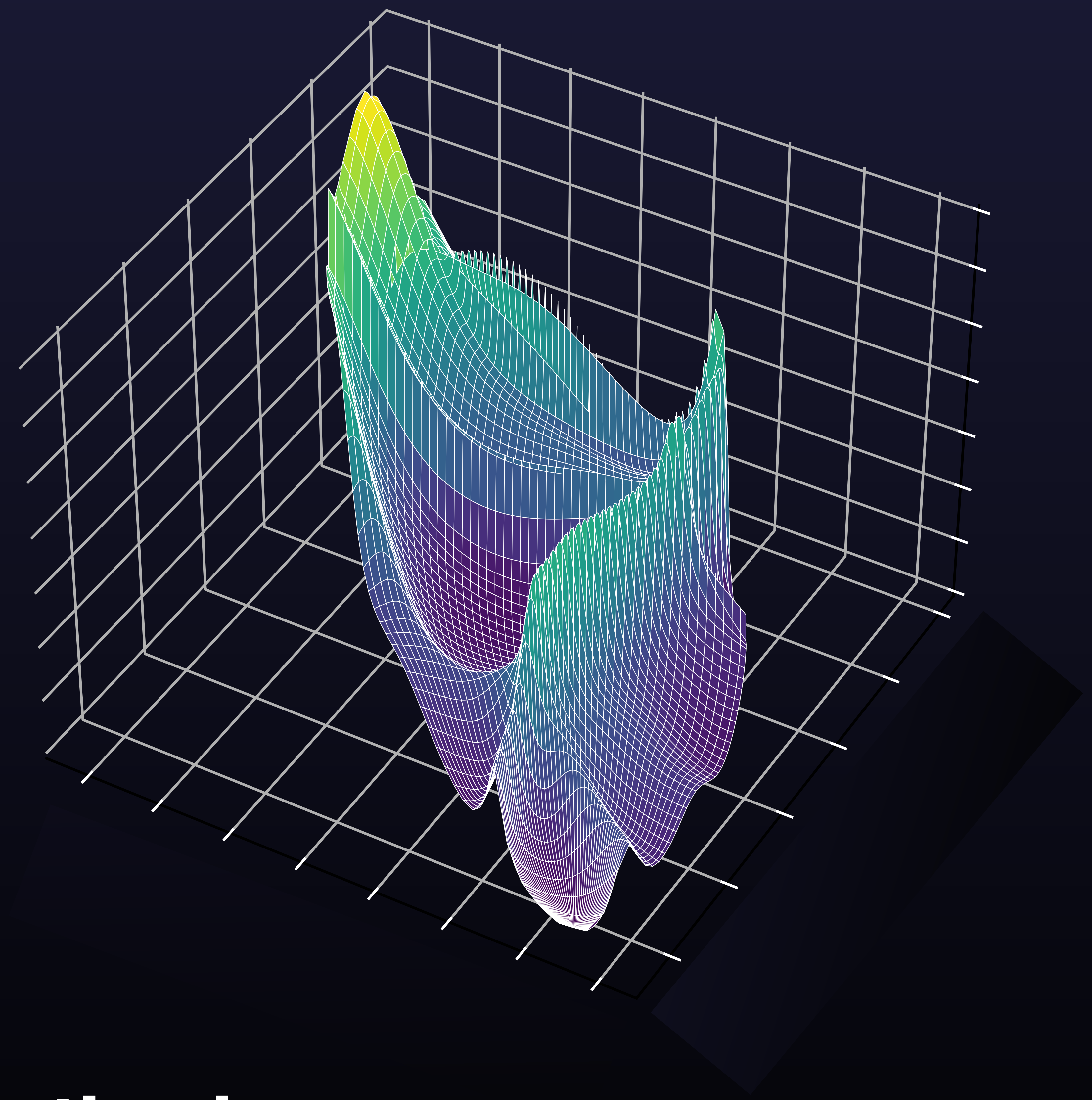
→ We need to speed up the simulation!

→ In this work: Event generation

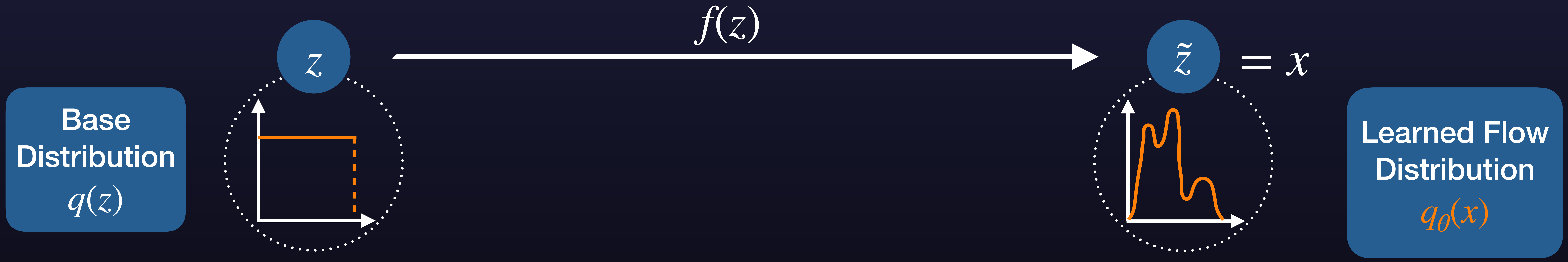
What is event generation from a ML perspective?



- Analytic formula for unnormalized distribution $p(x)$ ("matrix element")
- $p(x)$ describes the outgoing particles
- Sample from this distribution
- **Normalizing flow instead of standard methods**

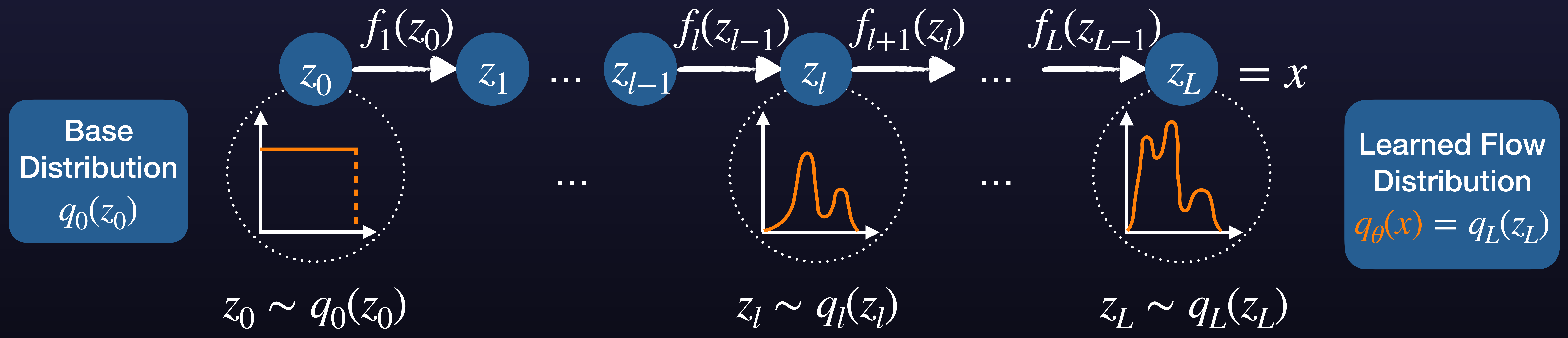


What is a normalizing flow?



Rezende and Mohamed, “Variational Inference with Normalizing Flows.” ICML’15.
Durkan et al., “Neural Spline Flows.”, NeurIPS’19.

What is a normalizing flow?

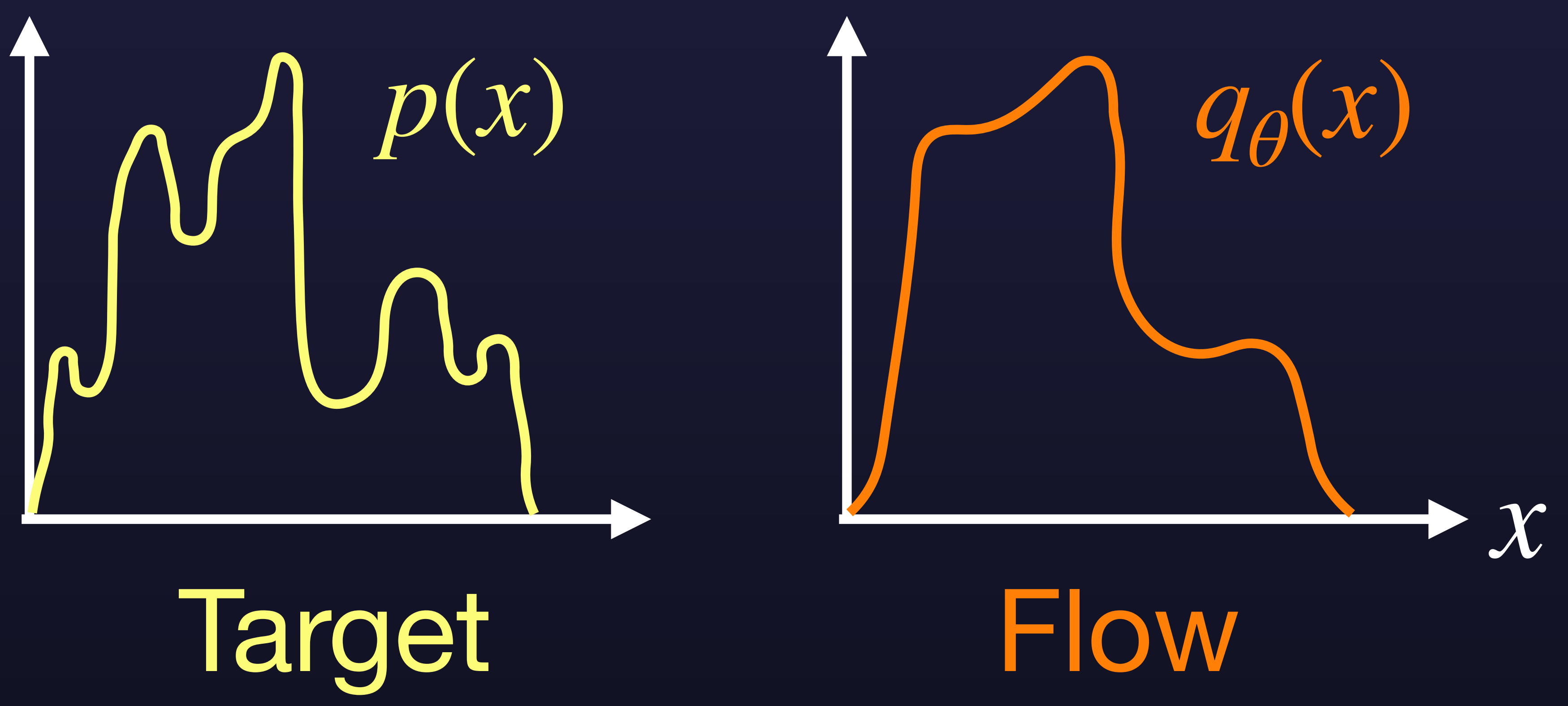


→ Generate samples x and evaluate density $q_\theta(x)$

Rezende and Mohamed, “Variational Inference with Normalizing Flows.” ICML’15.
Durkan et al., “Neural Spline Flows.”, NeurIPS’19.

How to train a normalizing flow?

3 Methods



Samples from target $x \sim p(x)$

(1) Forward KL Divergence (fKLD)

Samples from flow $x \sim q_\theta(x)$ and density evaluation of $p(x)$

(2) Reverse KL Divergence (rKLD)

(3) Flow Annealed Importance Sampling Bootstrap (FAB)

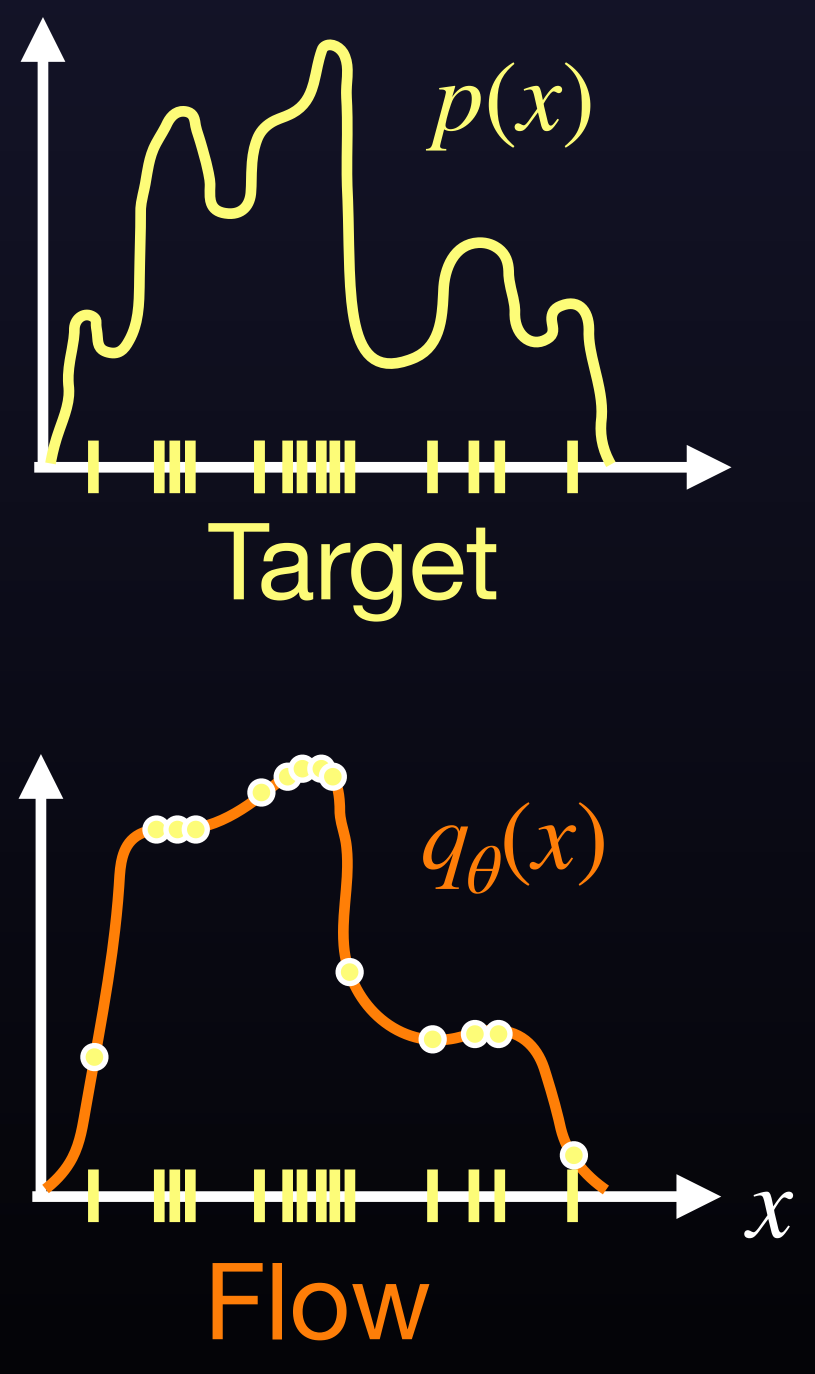
Midgley, Stimper, et al., “Flow Annealed Importance Sampling Bootstrap”, ICLR, 2023.

Training with Samples

Forward KL Divergence (fKLD)

$$L_{\text{fKLD}} = D_{\text{KL}}(p \,\|\, q_\theta) = \mathbb{E}_{x\sim p(x)}\left[\log \frac{p(x)}{q_\theta(x)}\right] = -\sum_{i=1}^{N} \log q_\theta(x_i) + \text{const.}$$

Generate $x \sim p(x)$ → Evaluate $q_\theta(x)$ $\underset{\nabla_\theta}{\rightleftarrows}$ L_{fKLD}

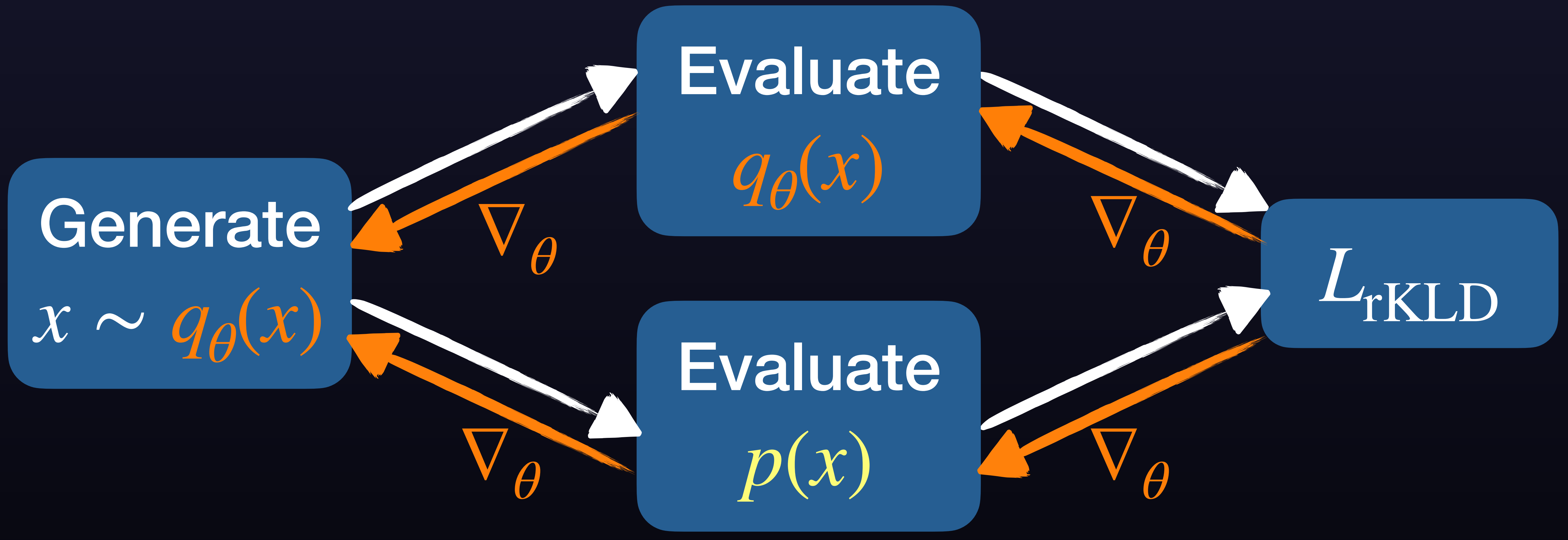


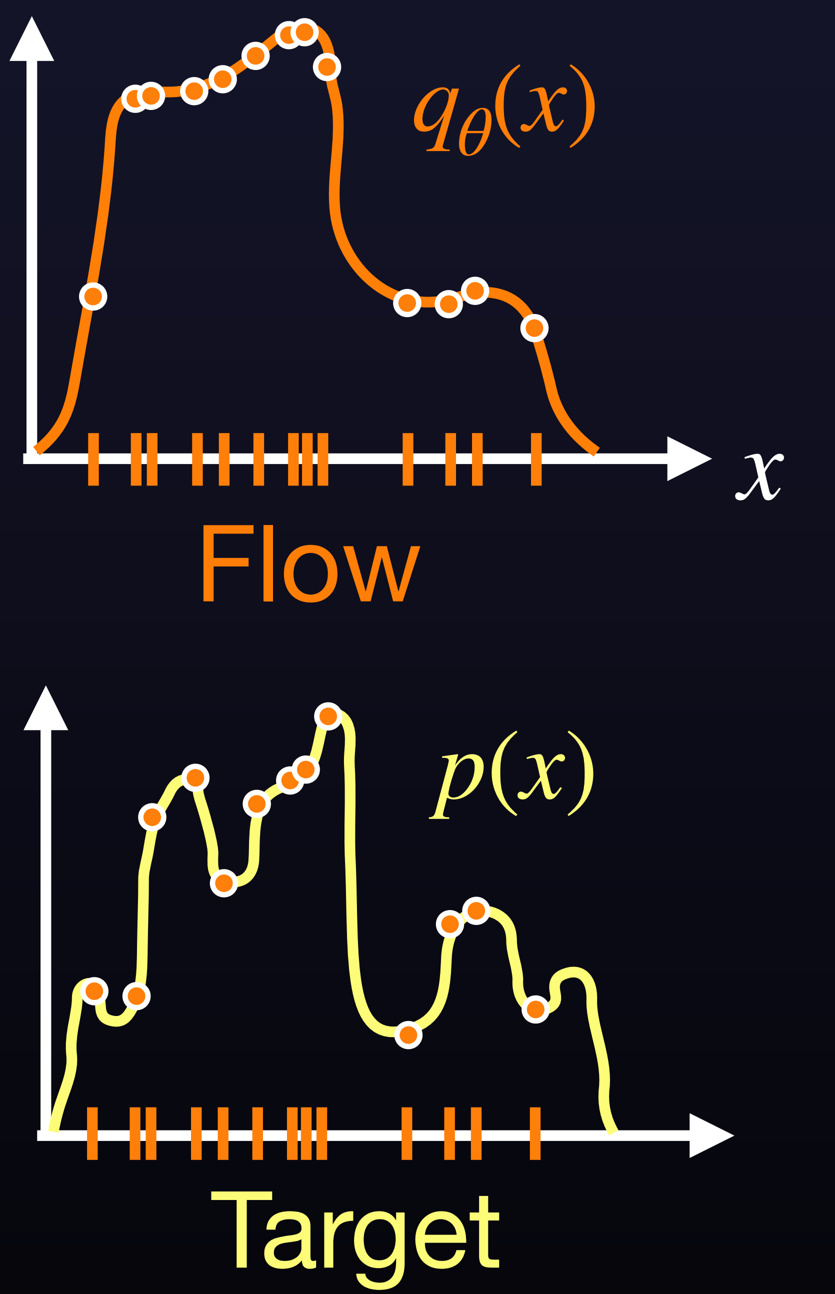
→ Expensive to generate training data

Training with Density Evaluation

Reverse KL Divergence (rKLD)

$$L_{\mathrm{rKLD}} = D_{\mathrm{KL}}(q_\theta \,\|\, p) = \mathbb{E}_{x \sim q_\theta(x)}\left[\log \frac{q_\theta(x)}{p(x)}\right]$$



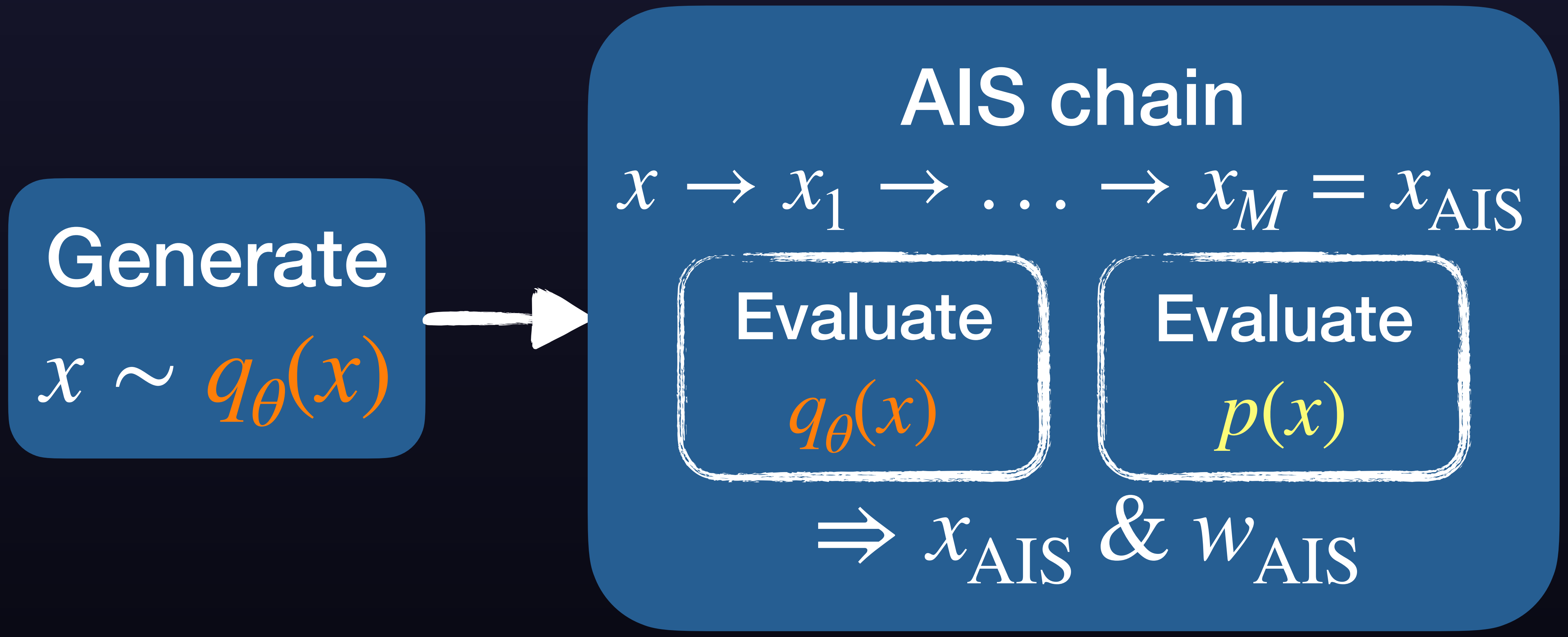


→ Requires differentiable target distribution $p(x)$

Training with Density Evaluation

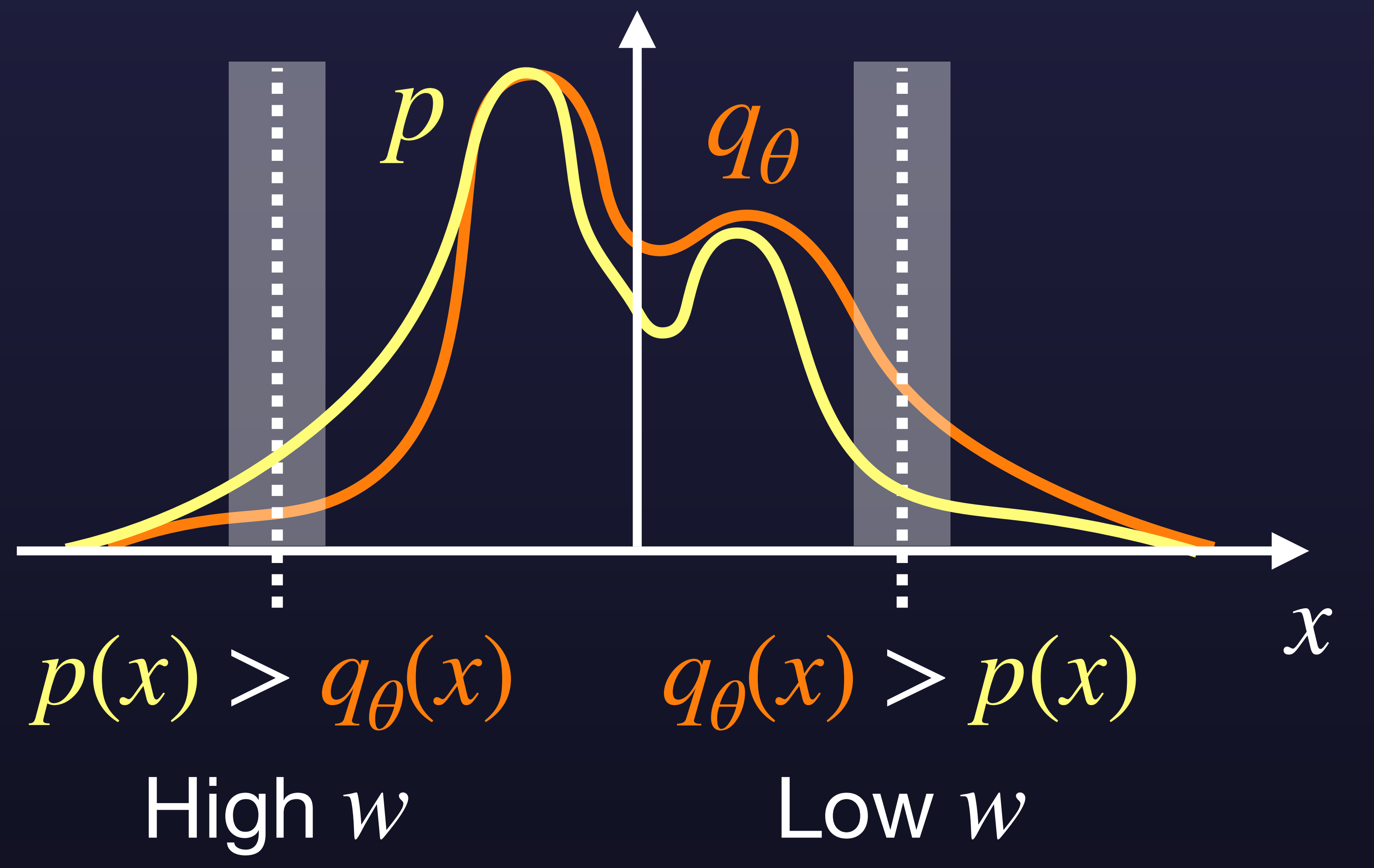
Flow Annealed Importance Sampling Bootstrap (FAB)

Improve flow samples $x \sim q_\theta(x)$ with Annealed Importance Sampling (AIS)



Midgley, Stimper, et al., “Flow Annealed Importance Sampling Bootstrap”, ICLR, 2023.

What is importance sampling?



- Compare two distributions via importance weight:

$$w = \frac{p(x)}{q_\theta(x)}$$

- Reweight samples from $q_\theta(x)$ towards $p(x)$ with weights

- Problem: $p(x)$ and $q_\theta(x)$ are very different
  $\rightarrow$ Samples from $q_\theta(x)$ not representative of $p(x)$
  $\rightarrow$ Importance sampling unstable

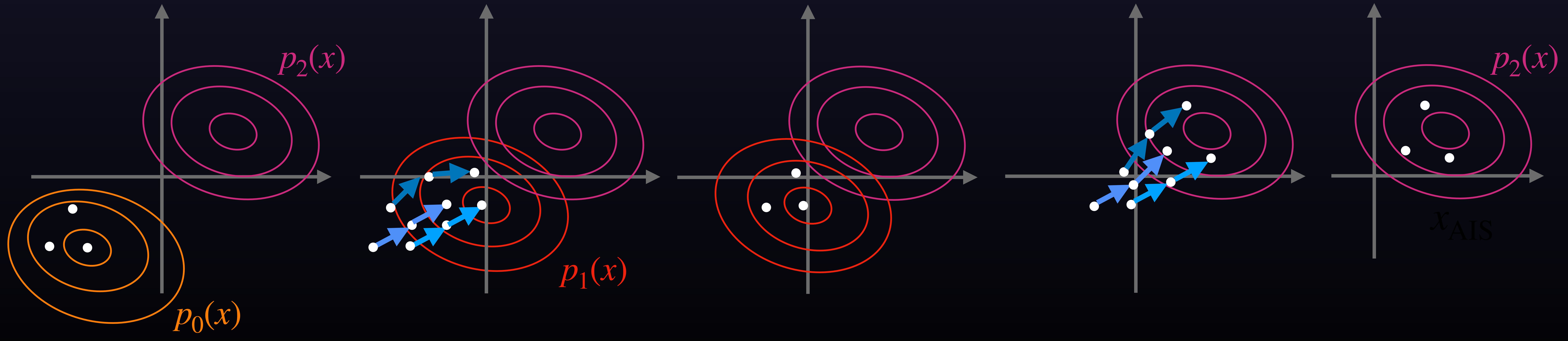
How does Annealed Importance Sampling (AIS) work?

- Goal: Move samples from initial $p_0(x) = q_\theta(x)$ toward final distribution $p_2(x)$
- Combine IS with MCMC chain of intermediate distribution $p_1(x)$

$$\log p_1(x) = 0.5 \cdot \log p_0(x) + 0.5 \cdot \log p_2(x)$$

- Use HMC to move samples between distributions

$$w_{\mathrm{AIS}} = \frac{p_1(x_1)}{p_0(x_1)} \frac{p_2(x_2)}{p_1(x_2)}$$



Neal, “Annealed importance sampling”, 2001.

Why does HMC require a differentiable target?

- Goal: Move samples between AIS intermediate distributions with HMC
- HMC based on Hamiltonian equations:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\partial H}{\partial p} \qquad\qquad H = \frac{p^2}{2m} + \underbrace{V(x)}_{\log p(x)}$$

$$\frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{\partial H}{\partial x} = -\frac{\partial}{\partial x} \log p(x)$$

$\rightarrow$ $p(x)$ needs to be differentiable

 Betancourt, "A Conceptual Introduction to Hamiltonian Monte Carlo", 2018

Training with Density Evaluation

Flow Annealed Importance Sampling Bootstrap (FAB)

Improve flow samples $x \sim q_\theta(x)$ with Annealed Importance Sampling (AIS)

Generate
$x \sim q_\theta(x)$

$\rightarrow$

AIS chain
$x \rightarrow x_1 \rightarrow \ldots \rightarrow x_M = x_{\text{AIS}}$

Evaluate $q_\theta(x)$ | **Evaluate** $p(x)$

$\Rightarrow x_{\text{AIS}}$ & w_{AIS}

AIS chain implemented via Hamiltonian Monte Carlo (HMC)
$\rightarrow$ Requires differentiable target $p(x)$

Midgley, Stimper, et al., “Flow Annealed Importance Sampling Bootstrap”, ICLR, 2023.

Training with Density Evaluation

Flow Annealed Importance Sampling Bootstrap (FAB)

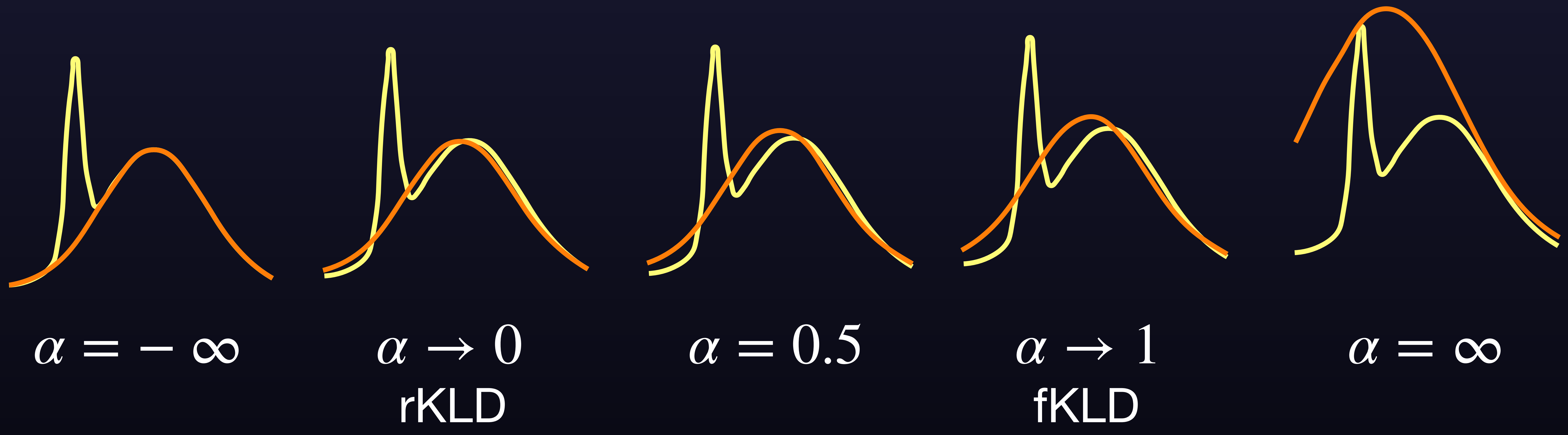
Improve flow samples $x \sim q_\theta(x)$ with Annealed Importance Sampling (AIS)

Generate $x \sim q_\theta(x)$ → **AIS chain** $x \rightarrow x_1 \rightarrow \ldots \rightarrow x_M = x_{\text{AIS}}$ [Evaluate $q_\theta(x)$] [Evaluate $p(x)$] $\Rightarrow x_{\text{AIS}}$ & w_{AIS} - - → Store x_{AIS} & w_{AIS} in replay buffer → Evaluate $q_\theta(\bar{x}_{\text{AIS}})$ → L_{FAB}

Midgley, Stimper, et al., “Flow Annealed Importance Sampling Bootstrap”, ICLR, 2023.

What is special about the loss function of FAB?

α-divergence: $$D_\alpha(p||q_\theta) = -\frac{1}{\alpha(1-\alpha)}\int p(x)^\alpha q_\theta(x)^{1-\alpha}\,\mathrm{d}x$$

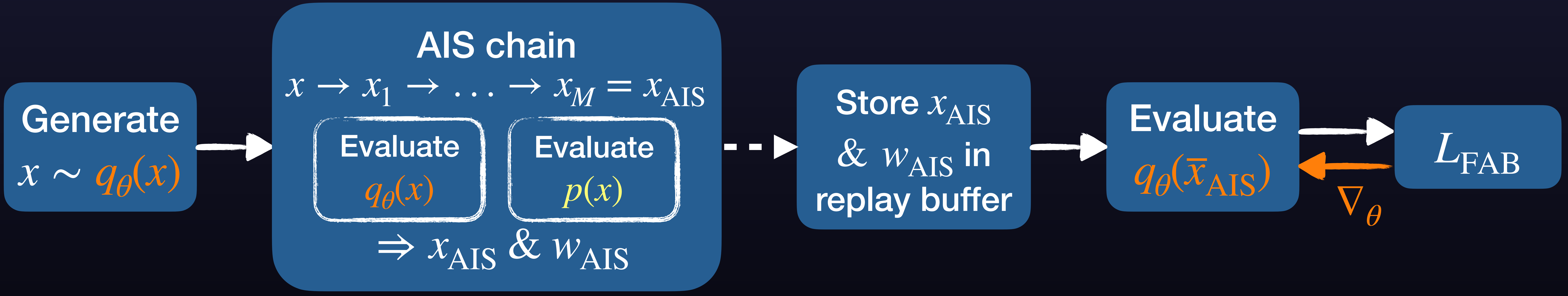


$$D_{\alpha=2}(p||q_\theta) \propto \int \frac{p(x)^2}{q_\theta(x)^2}\, q_\theta(x)\,\mathrm{d}x \xrightarrow{\text{Approximate}} L_{\mathrm{FAB}} = -\sum_{i=1}^{N} \frac{\bar{w}_{\mathrm{AIS}}^{(i)}}{\sum_j \bar{w}_{\mathrm{AIS}}^{(j)}}\, q_\theta\left(\bar{x}_{\mathrm{AIS}}^{(i)}\right)$$

Training with Density Evaluation

Flow Annealed Importance Sampling Bootstrap (FAB)

Improve flow samples $x \sim q_\theta(x)$ with Annealed Importance Sampling (AIS)

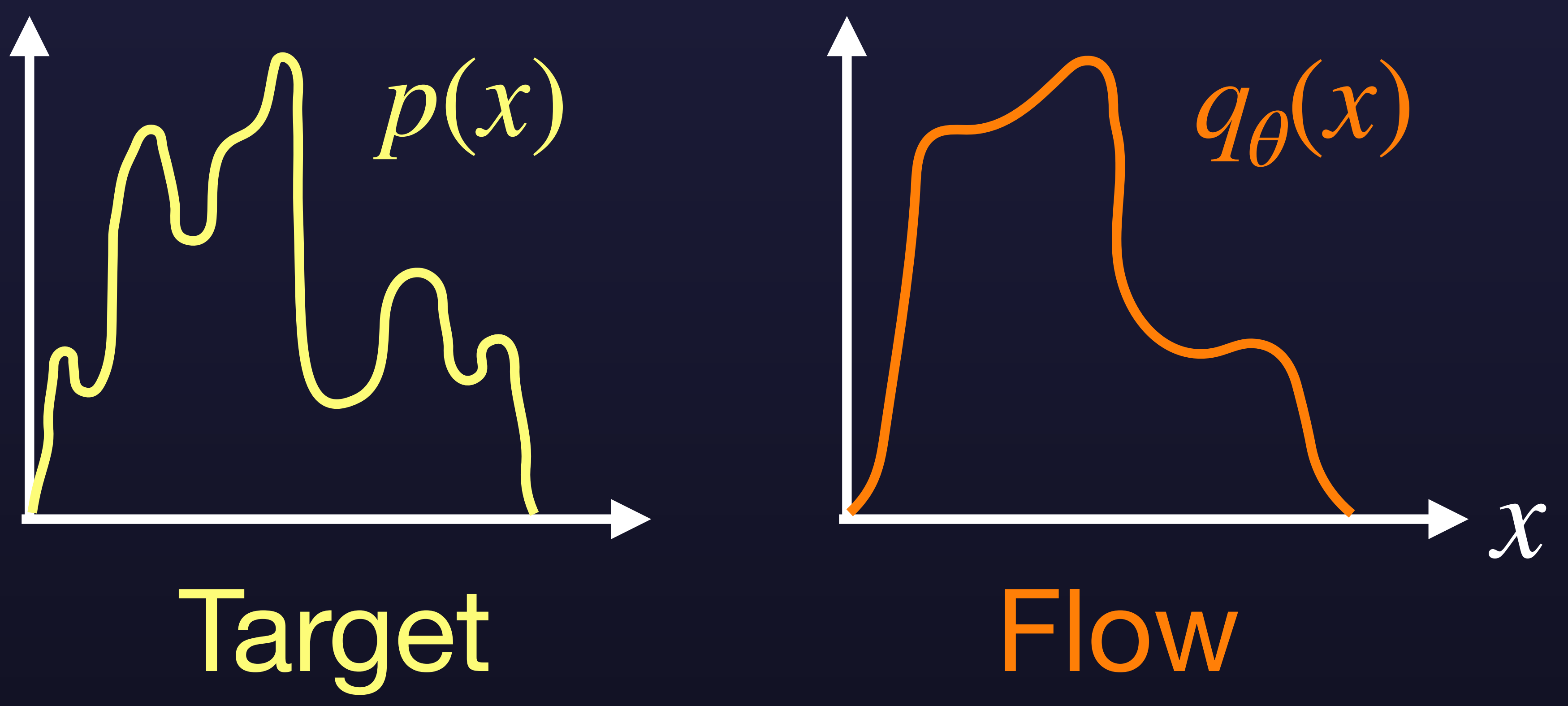


$$L_{\mathrm{FAB}} = -\sum_{i=1}^{N} \frac{\bar{w}_{\mathrm{AIS}}^{(i)}}{\sum_j \bar{w}_{\mathrm{AIS}}^{(j)}} q_\theta\left(\bar{x}_{\mathrm{AIS}}^{(i)}\right) \to$$ minimizes variance of importance weights

Midgley, Stimper, et al., “Flow Annealed Importance Sampling Bootstrap”, ICLR, 2023.

How to train a normalizing flow?

3 Methods



Samples from target $x \sim p(x)$

(1) Forward KL Divergence (fKLD)

Samples from flow $x \sim q_\theta(x)$ and density evaluation of $p(x)$

$\boxed{\text{(2) Reverse KL Divergence (rKLD)} \quad \text{(3) Flow Annealed Importance Sampling Bootstrap (FAB)}}$

$\rightarrow$ differentiable target $p(x)$?

Midgley, Stimper, et al., "Flow Annealed Importance Sampling Bootstrap", ICLR, 2023.

Differentiable target distributions

Examples

Recent work in particle physics: differentiable implementations of $p(x)$

- **ComPWA:** $\Lambda_c^+ \to pK^-\pi^+$ (2D)
- **MadJAX:** $e^+e^- \to t\bar{t}, t\bar{t} \to W^+b, \bar{t} \to W^-\bar{b}$ (8D)

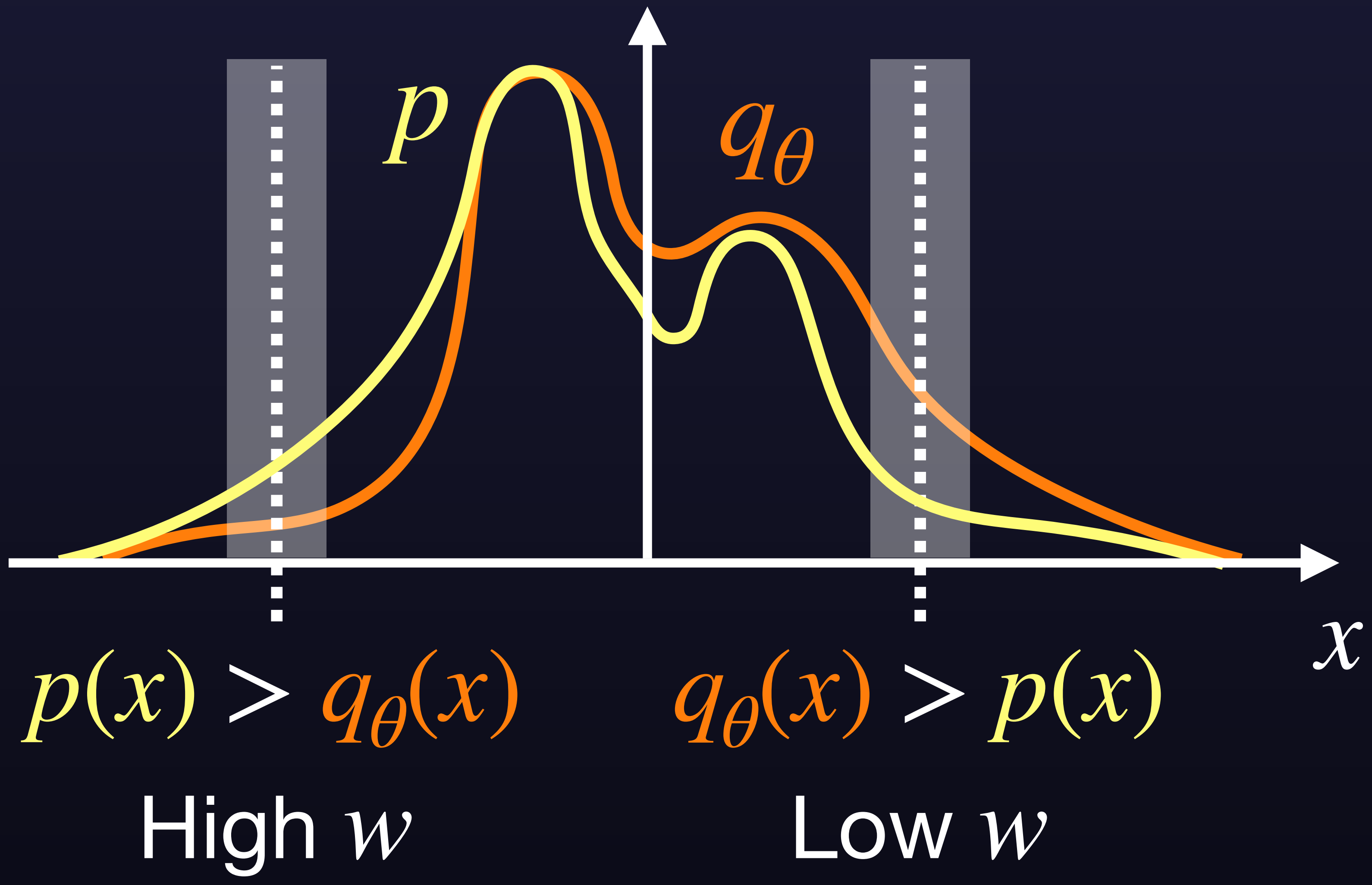
Common Partial Wave Analysis: A collaboration-independent organisation for amplitude analysis software, doi 10.5281/zenodo.6908150

Heinrich and Kagan, “Differentiable Matrix Elements with MadJax.” J. Phys. Conf., 2023.

How to compare the approaches?

- Importance sampling efficiency ϵ

$$\epsilon = \frac{1}{N}\frac{\left(\sum_i w_i\right)^2}{\sum_i w_i^2} \in [0,1] \quad \text{with} \quad w = \frac{p(x)}{q_\theta(x)}$$



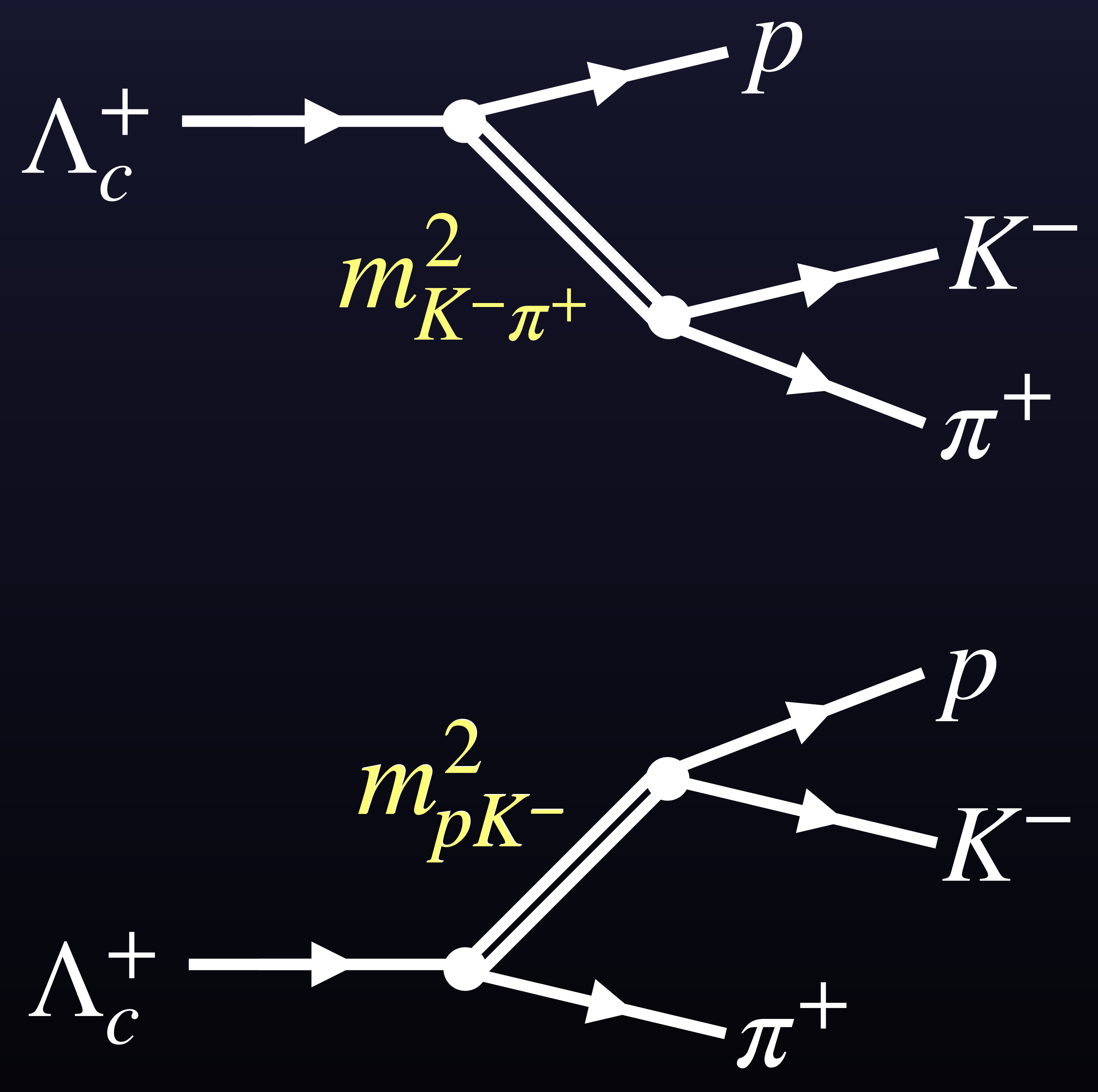
$$p(x) \neq q_\theta(x) \rightarrow \epsilon \ll 1$$

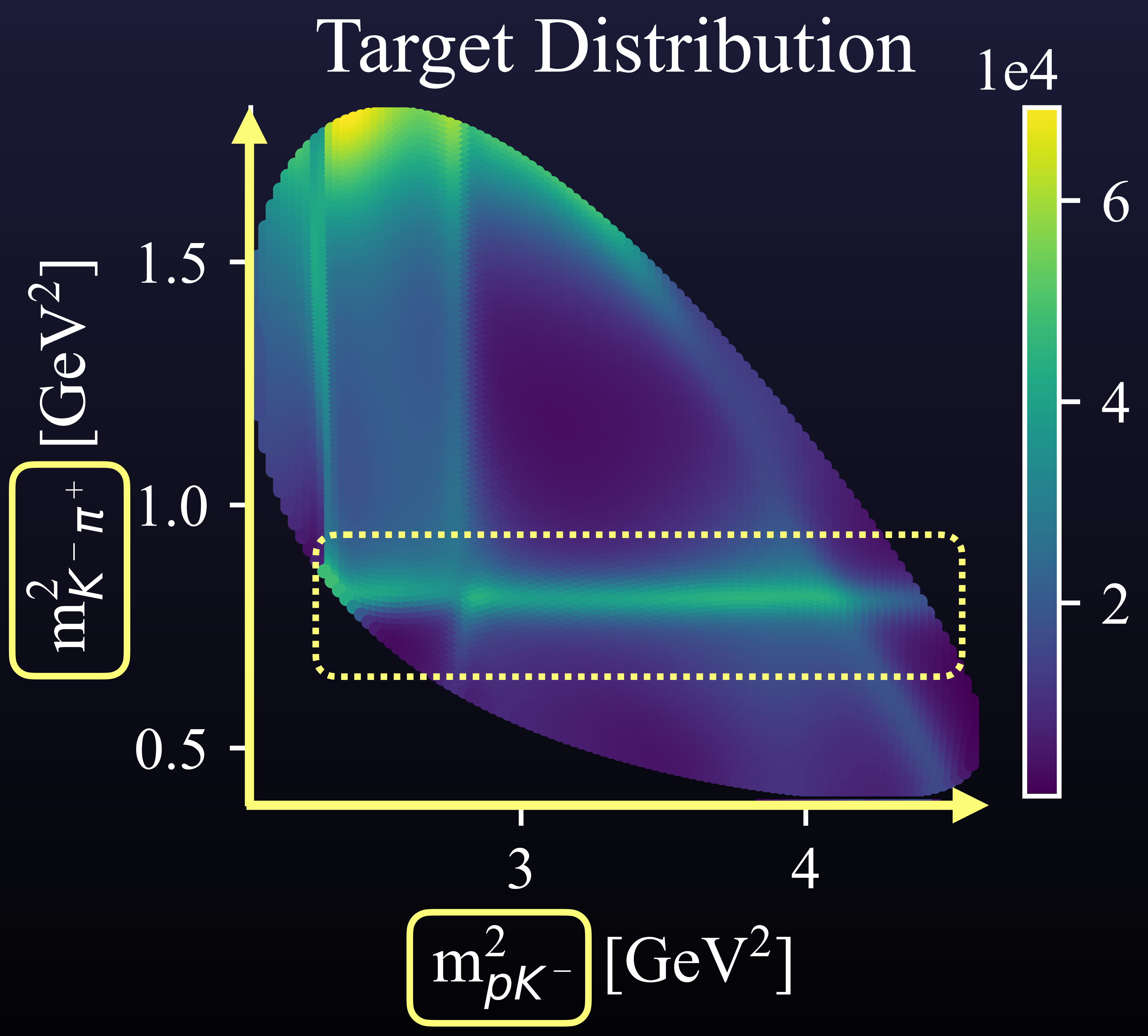
- More in the paper:
  fKLD on test data set, integral estimate & unweighting efficiency

Results: $\Lambda_c^+ \to pK^-\pi^+$ (2D)

2D: $\Lambda_c^+ \to pK^-\pi^+$

What do we see?

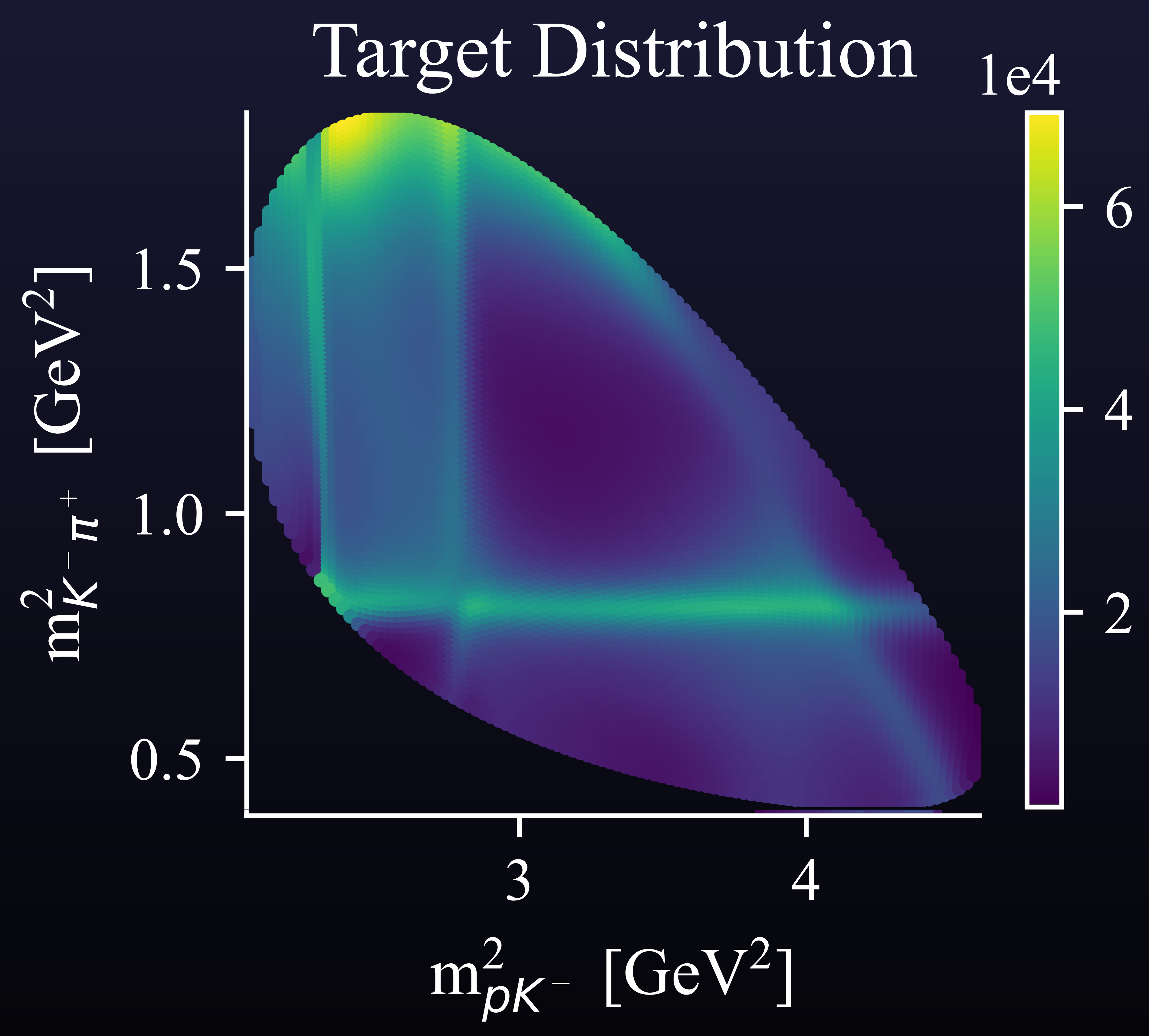




LHCb, "Amplitude analysis of the $\Lambda_c^+ \to pK^-\pi^+$ decay and Λ_c^+ baryon polarization measurement in semileptonic beauty hadron decays.", Phys. Rev. D 108, 2023.

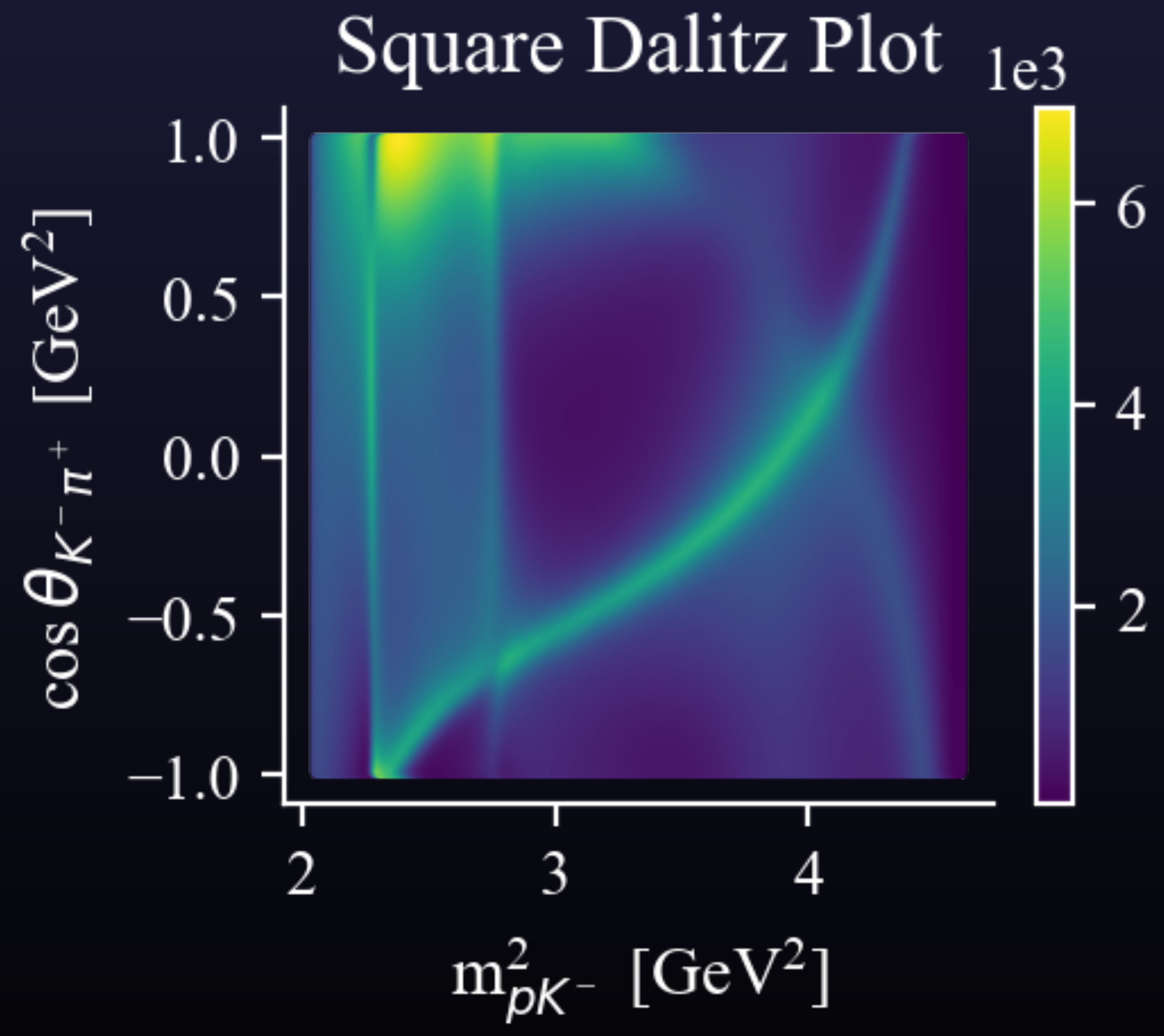
2D: $\Lambda_c^+ \to pK^-\pi^+$

What is the input to the normalizing flow?



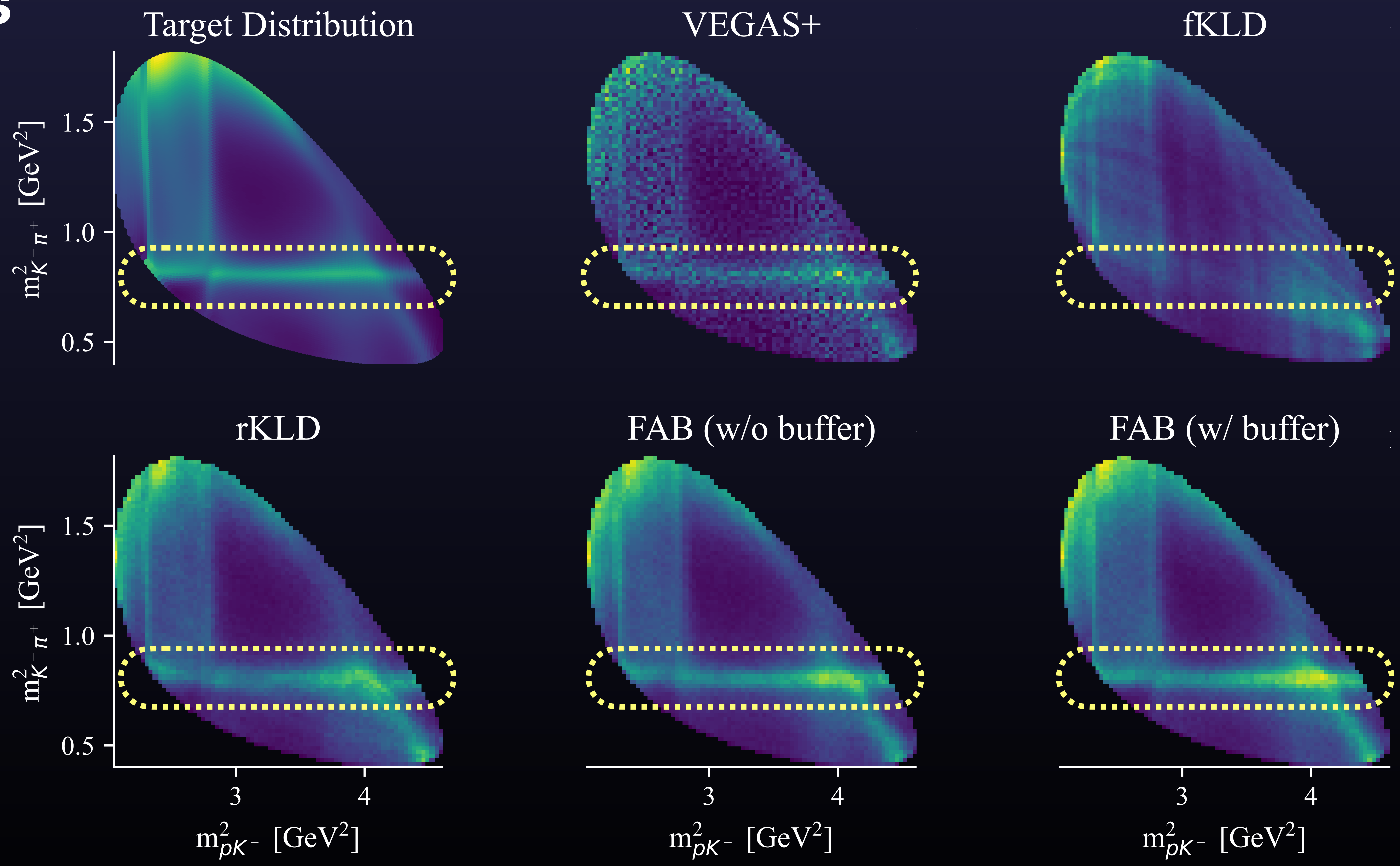
Invertible

Transformation



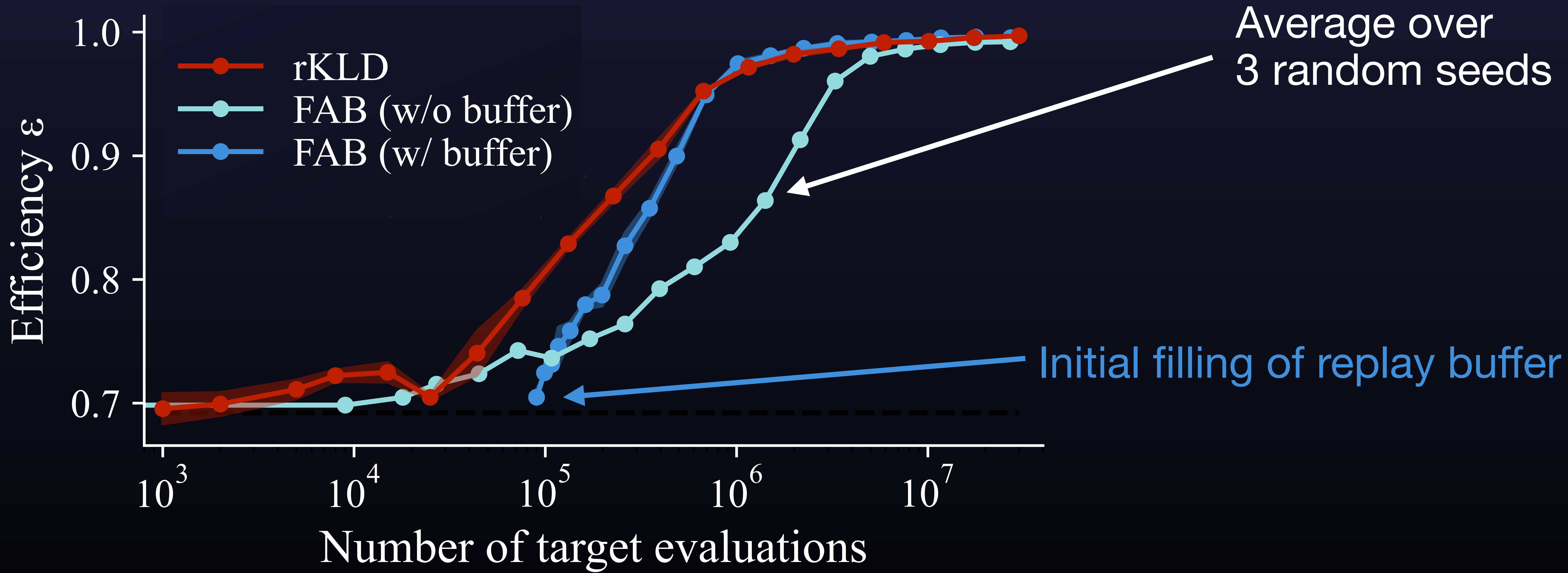
2D: $\Lambda_c^+ \to pK^-\pi^+$

Histograms

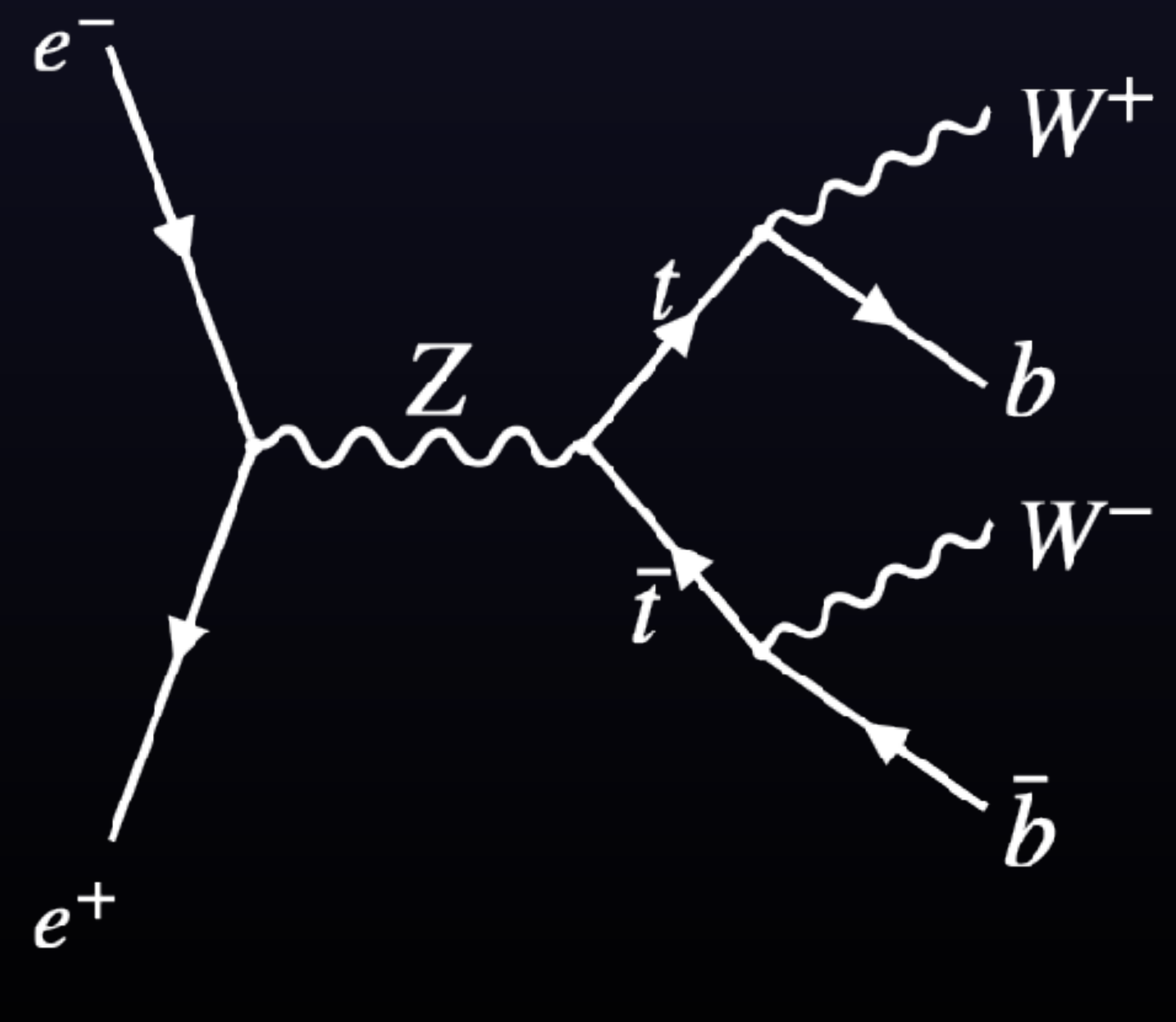


2D: $\Lambda_c^+ \to pK^-\pi^+$

Efficiency vs. Number of target evaluations

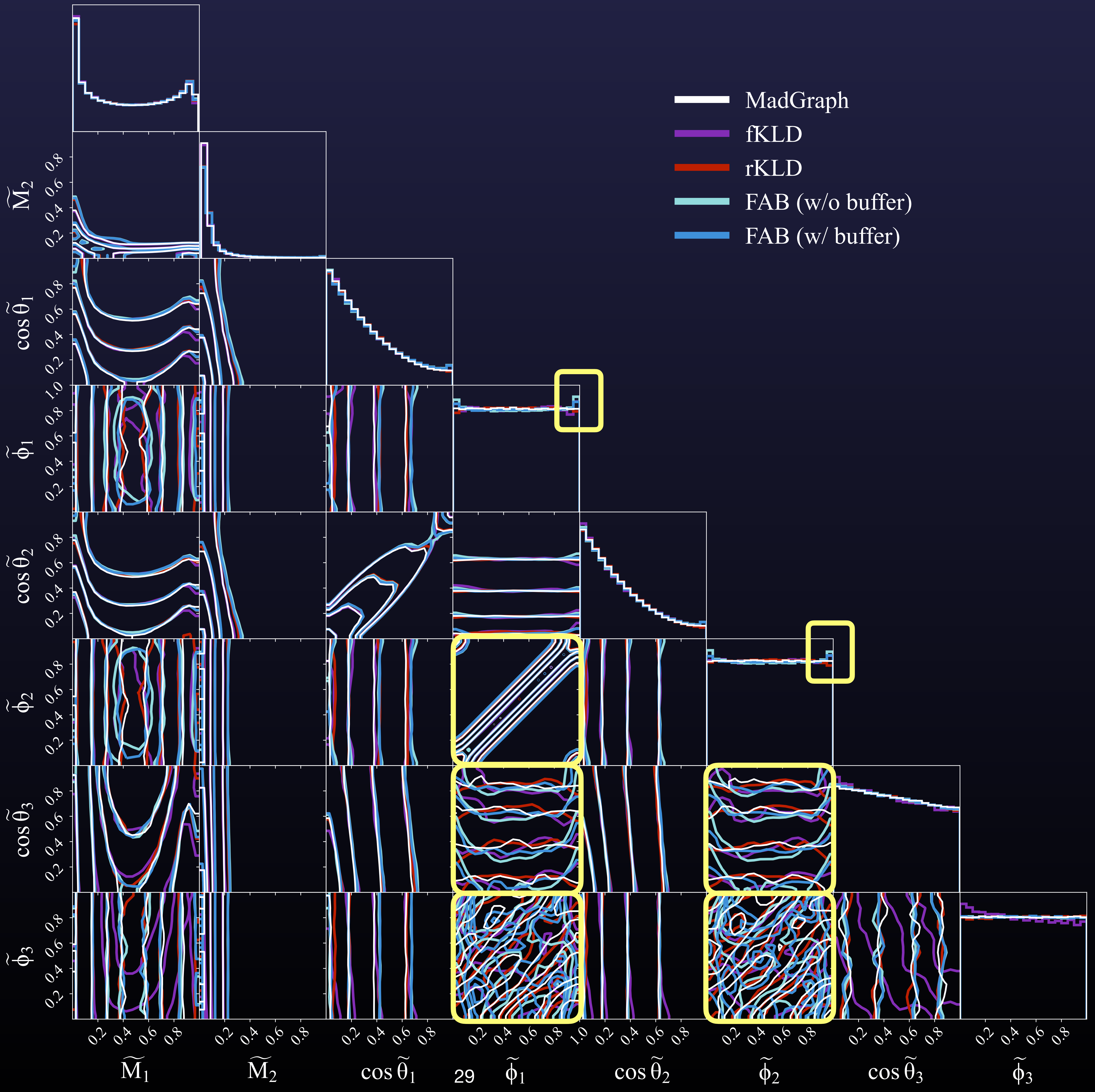


Results: $e^+e^- \to t\bar{t}$, $t \to W^+b, \bar{t} \to W^-\bar{b}$ (8D)



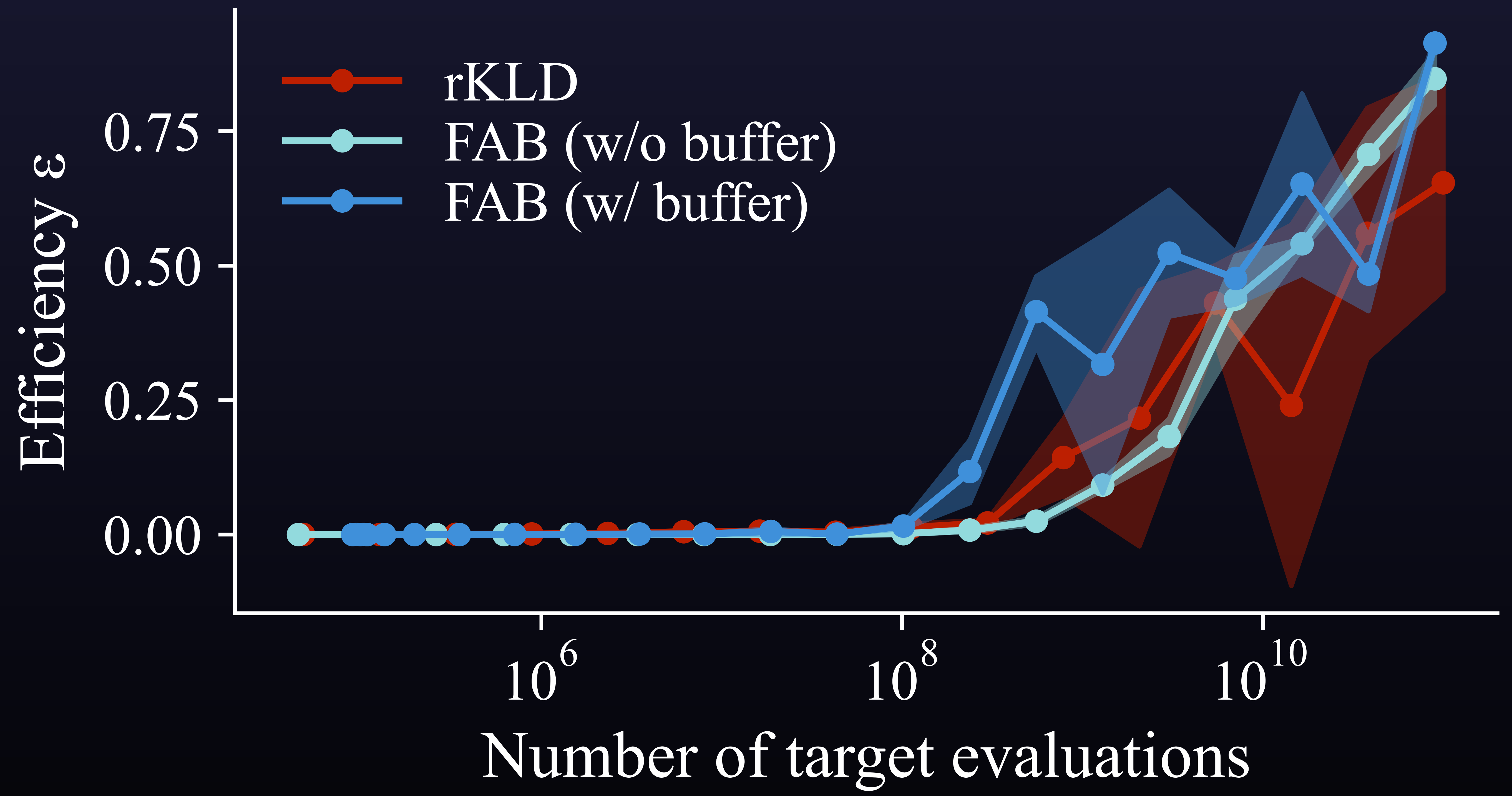
8D: $e^+e^- \to t\bar{t}$,
$t\bar{t} \to W^+b, \bar{t} \to W^-\bar{b}$

Corner Plot

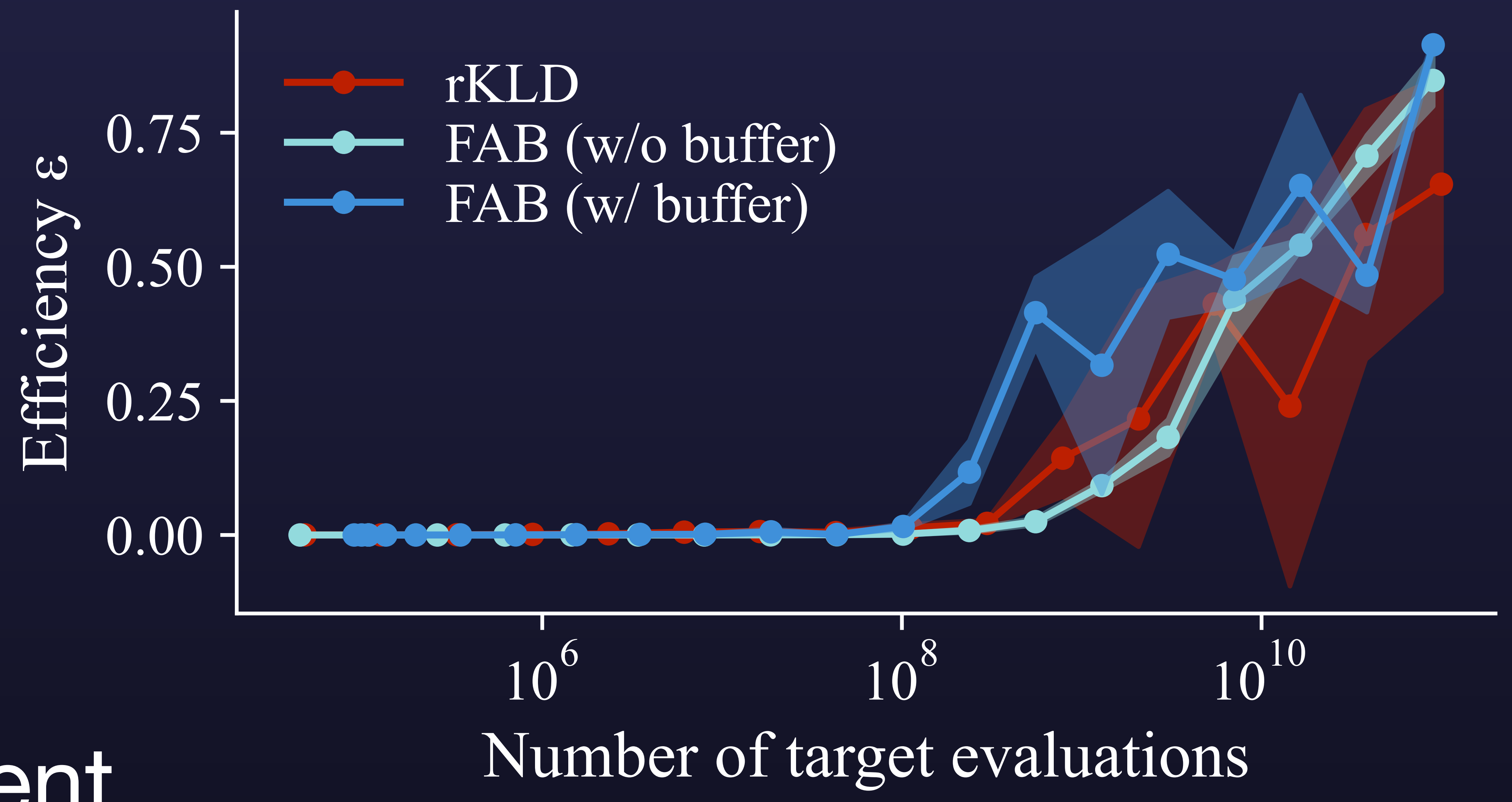


8D: $e^+e^- \to t\bar{t}, t\bar{t} \to W^+b, \bar{t} \to W^-\bar{b}$

Efficiency vs. Number of target evaluations



Take-Aways



- Different ways to train a normalizing flow
- Training with density evaluation requires differentiable implementation of matrix element
- Flow Annealed Importance Sampling Bootstrap outperforms others in high dimensions

Do you have any questions?

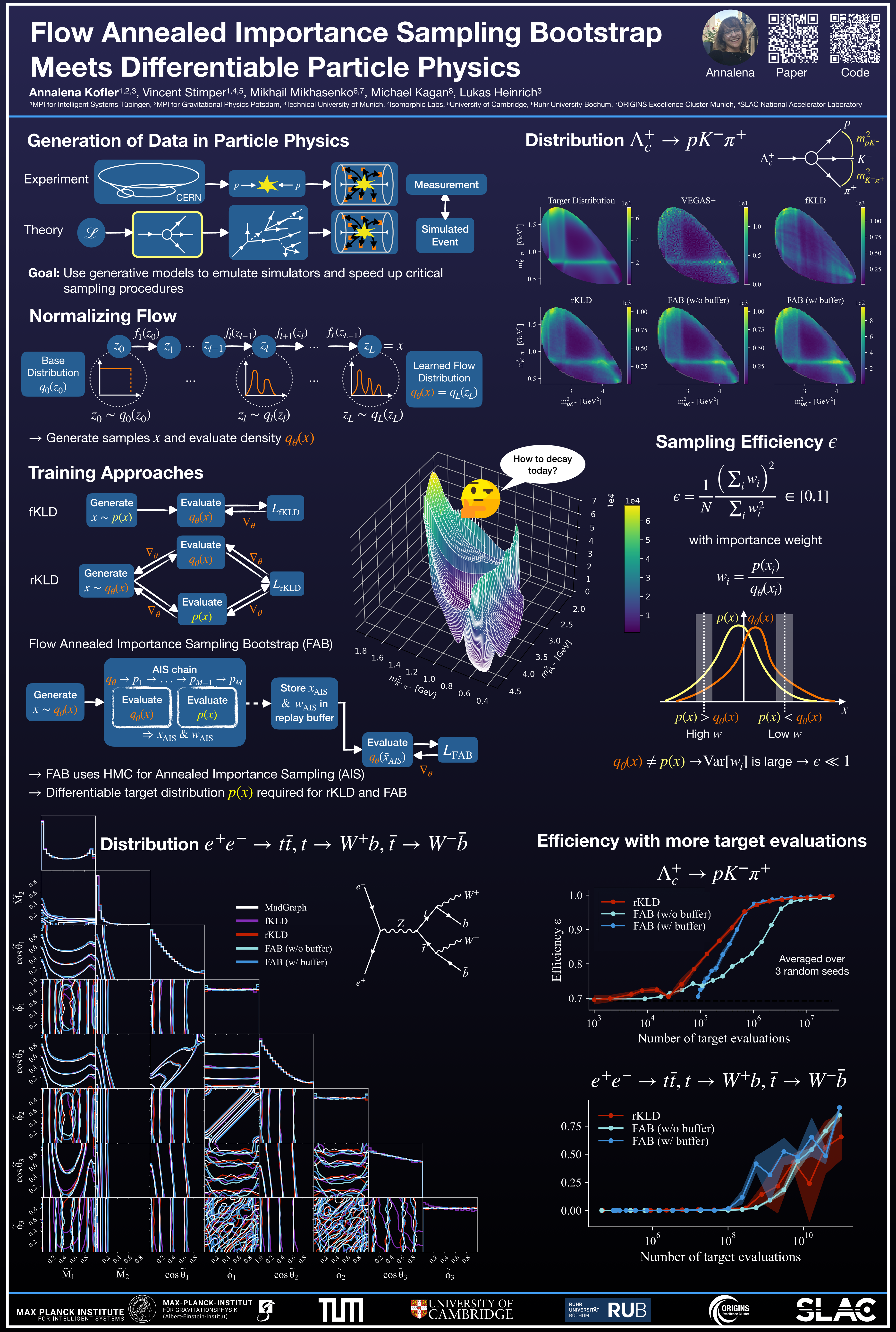
Paper

Code

Do you have any questions?

Paper

Code



References

- ATLAS Software and Computing HL-LHC Roadmap, 2022.
- Rezende and Mohamed, "Variational Inference with Normalizing Flows." ICML'15.
- Durkan et al., "Neural Spline Flows.", NeurIPS'19.
- LHCb, "Amplitude analysis of the $\Lambda_c^+ \to pK^-\pi^+$ decay and Λ_c^+ baryon polarization measurement in semileptonic beauty hadron decays.", Phys. Rev. D 108, 2023.
- Heinrich and Kagan, "Differentiable Matrix Elements with MadJax." J. Phys. Conf., 2023.
- Midgley, Stimper, et al., "Flow Annealed Importance Sampling Bootstrap", ICLR, 2023.

Other Evaluation Metrics

- Importance sampling efficiency ϵ

$$\epsilon = \frac{1}{N}\frac{\left(\sum_i w_i\right)^2}{\sum_i w_i^2}$$

- fKLD on test data set

$$\mathbb{E}_{x\sim p(x)}\left[\log \frac{p(x)}{q_\theta(x)}\right]$$

- Integral estimate

$$\bar{I} = \int p(x)\;\mathrm{d}x = \int \frac{p(x)}{q_\theta(x)}\;q_\theta(x)\;\mathrm{d}x \approx \frac{1}{N}\sum_i w_i$$

- Unweighting efficiency

$$\epsilon_{\mathrm{uw}} = \frac{1}{N}\frac{\sum_i w_i}{w_{\max}}$$