The versatility of flow-based fast calorimeter surrogate models

Ian Pang

Nov 7, 2024

ML4Jets, Paris

RUTGERS

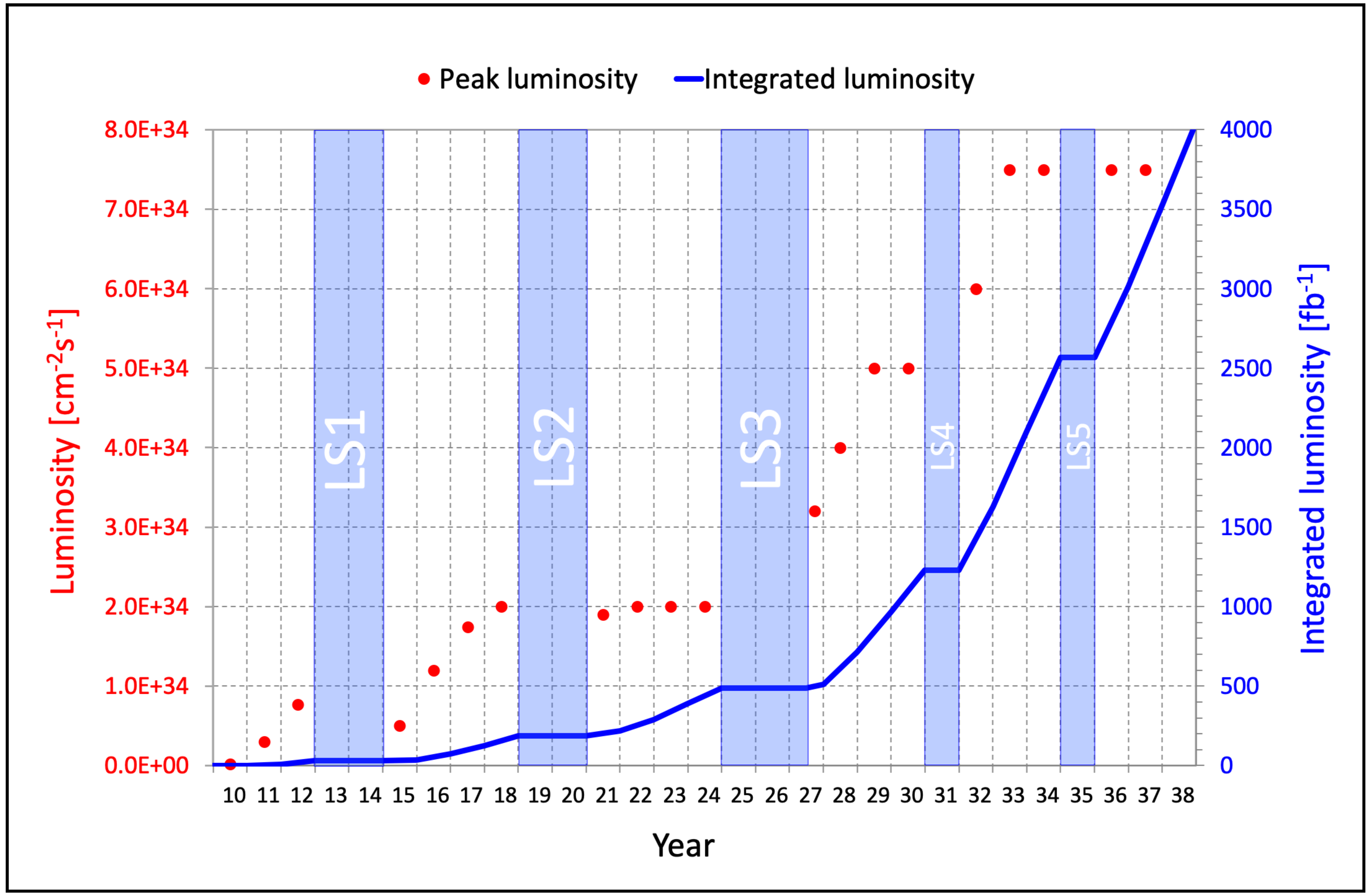
[2312.11618] C. Krause, B. Nachman, **IP**, D. Shih, Y. Zhu

[2404.18992] H. Du, C. Krause, V. Mikuni, B. Nachman, **IP**, D. Shih

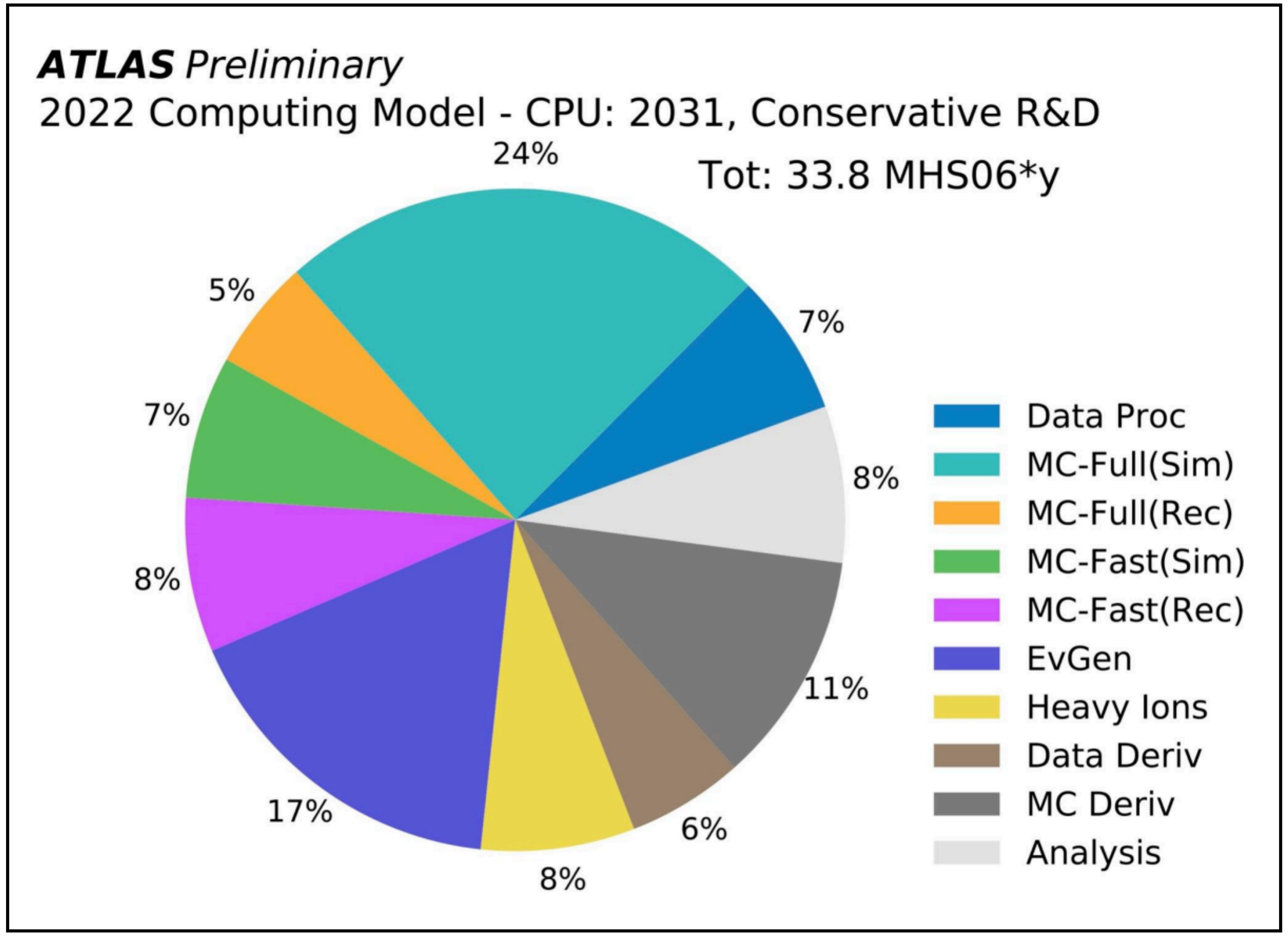
ian.pang@physics.rutgers.edu

Fast Calorimeter Surrogate Modeling

Calorimeter shower simulation is major bottleneck in LHC computational pipeline!



https://lhc-commissioning.web.cern.ch/schedule/images/LHC-ultimate-lumi-projection.png

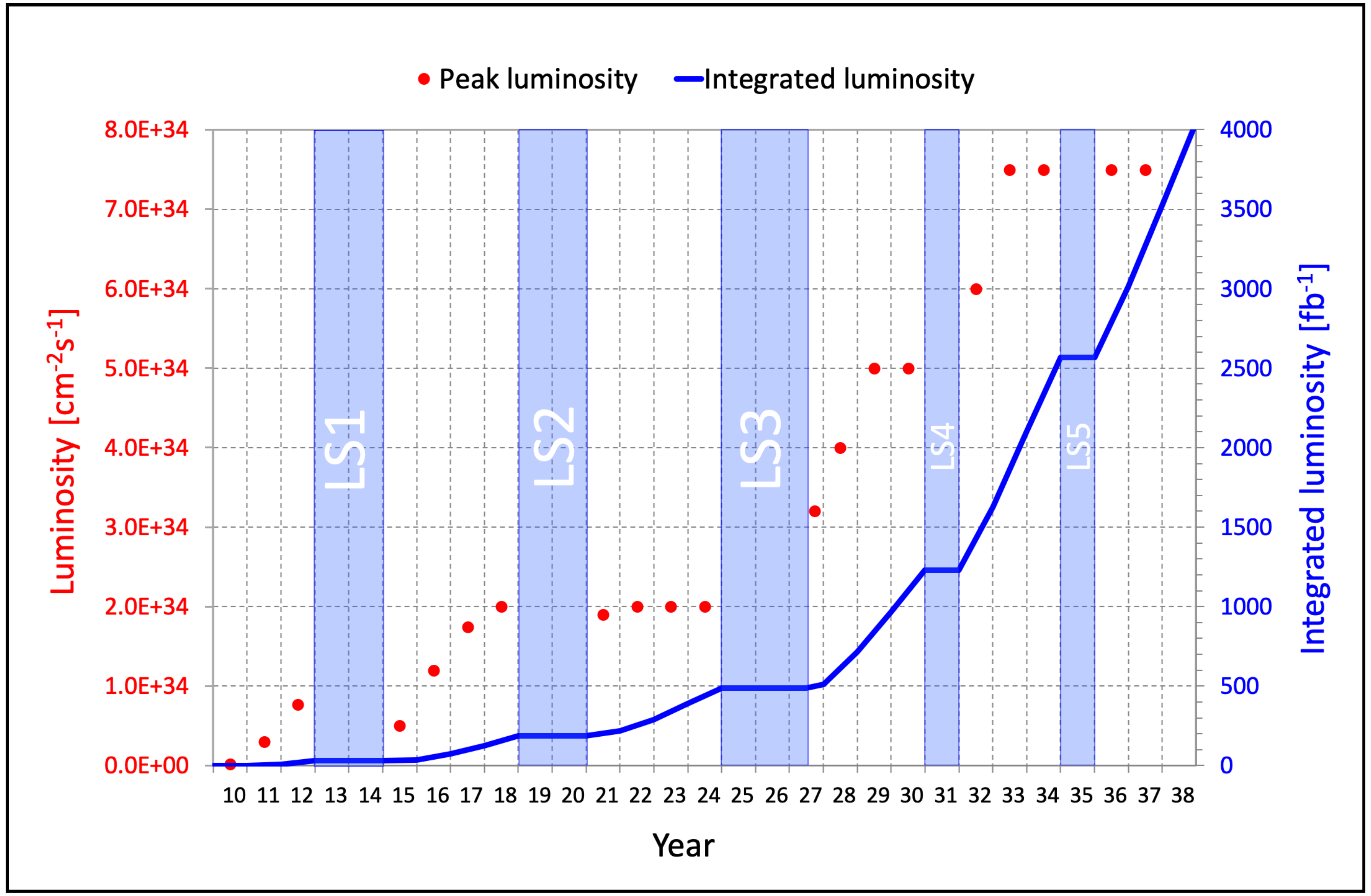


CERN-LHCC-2022-005

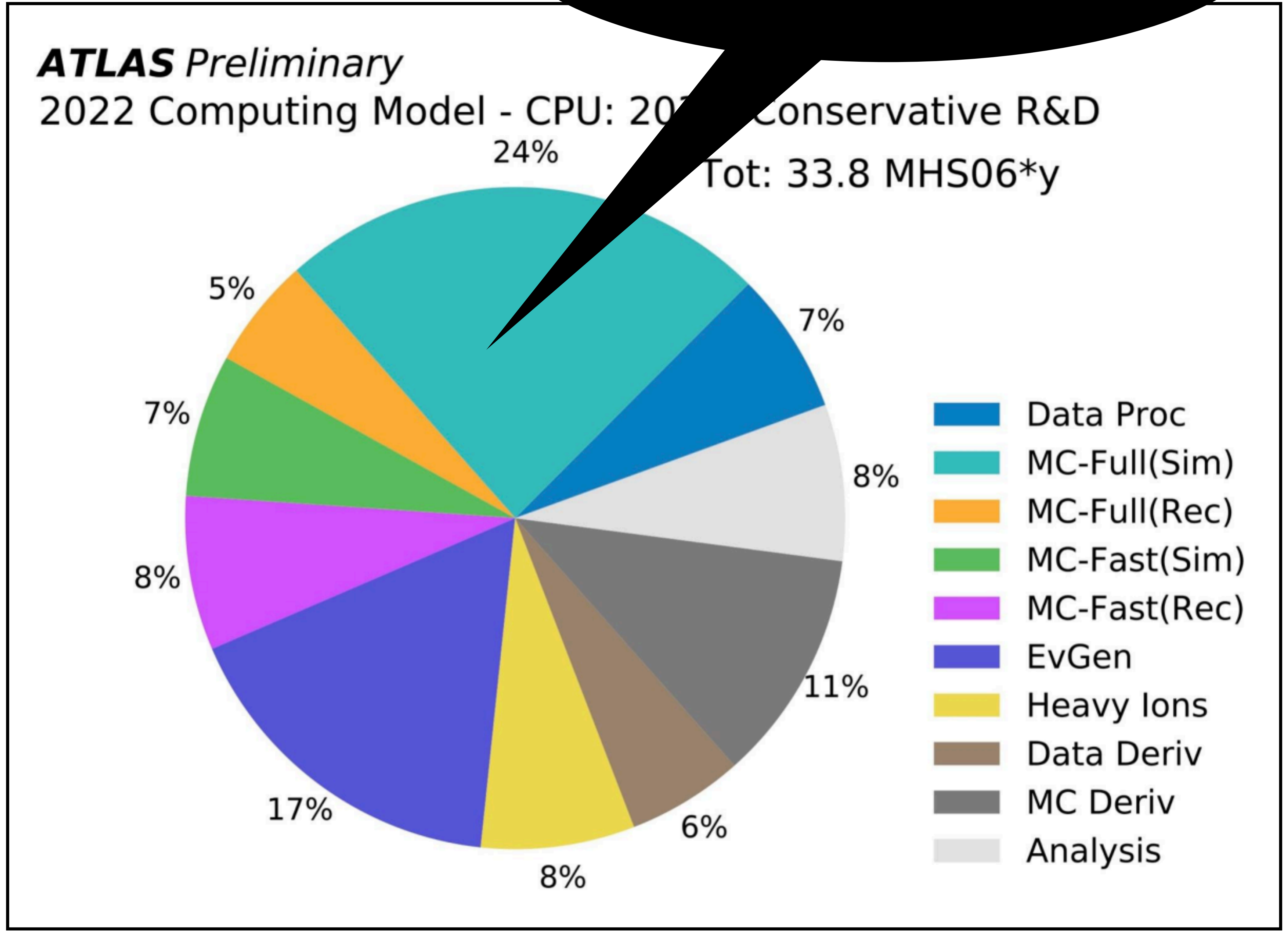
Fast Calorimeter Surrogate Modeling

Calorimeter shower simulation is major bottleneck in LHC computational pipeline!

MC to compare measurement to theory



https://lhc-commissioning.web.cern.ch/schedule/images/LHC-ultimate-lumi-projection.png

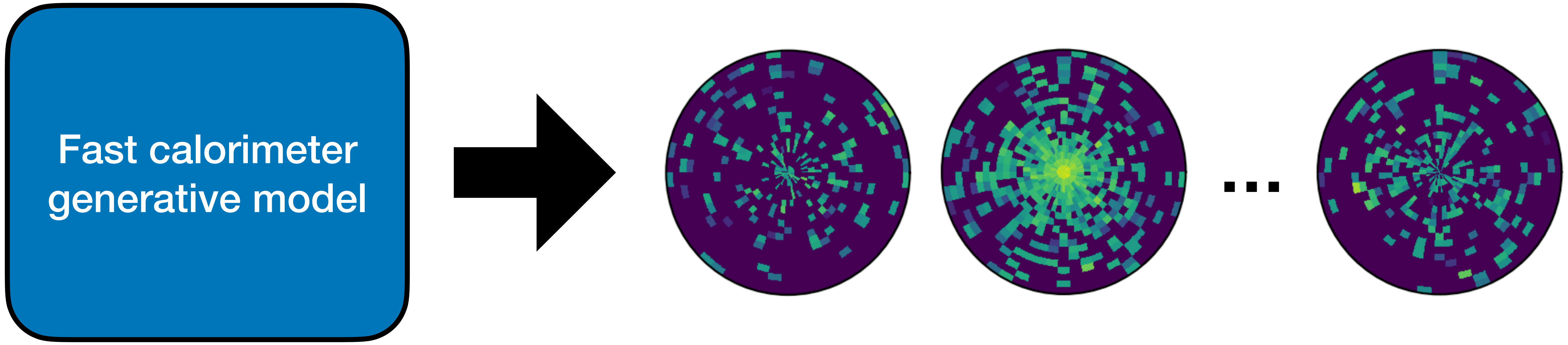


CERN-LHCC-2022-005

Fast Calorimeter Surrogate Modeling

Calorimeter shower simulation is major bottleneck in LHC computational pipeline!

Surrogate modeling to **speed up** generation of expensive GEANT4 calorimeter showers



Fast Calorimeter Surrogate Modeling

Many different approaches tested on this task!

- GANs (e.g. 1712.10321, 2309.06515)
- VAEs (e.g. 2211.15380, 2312.09290)
- Normalizing flows (e.g. 2106.05285, 2302.11594)
- Diffusion (e.g. 2308.03847, 2308.03876)
- Flow matching (2405.09629)

(See CaloChallenge summary paper [2410.21611] which compares the various approaches)

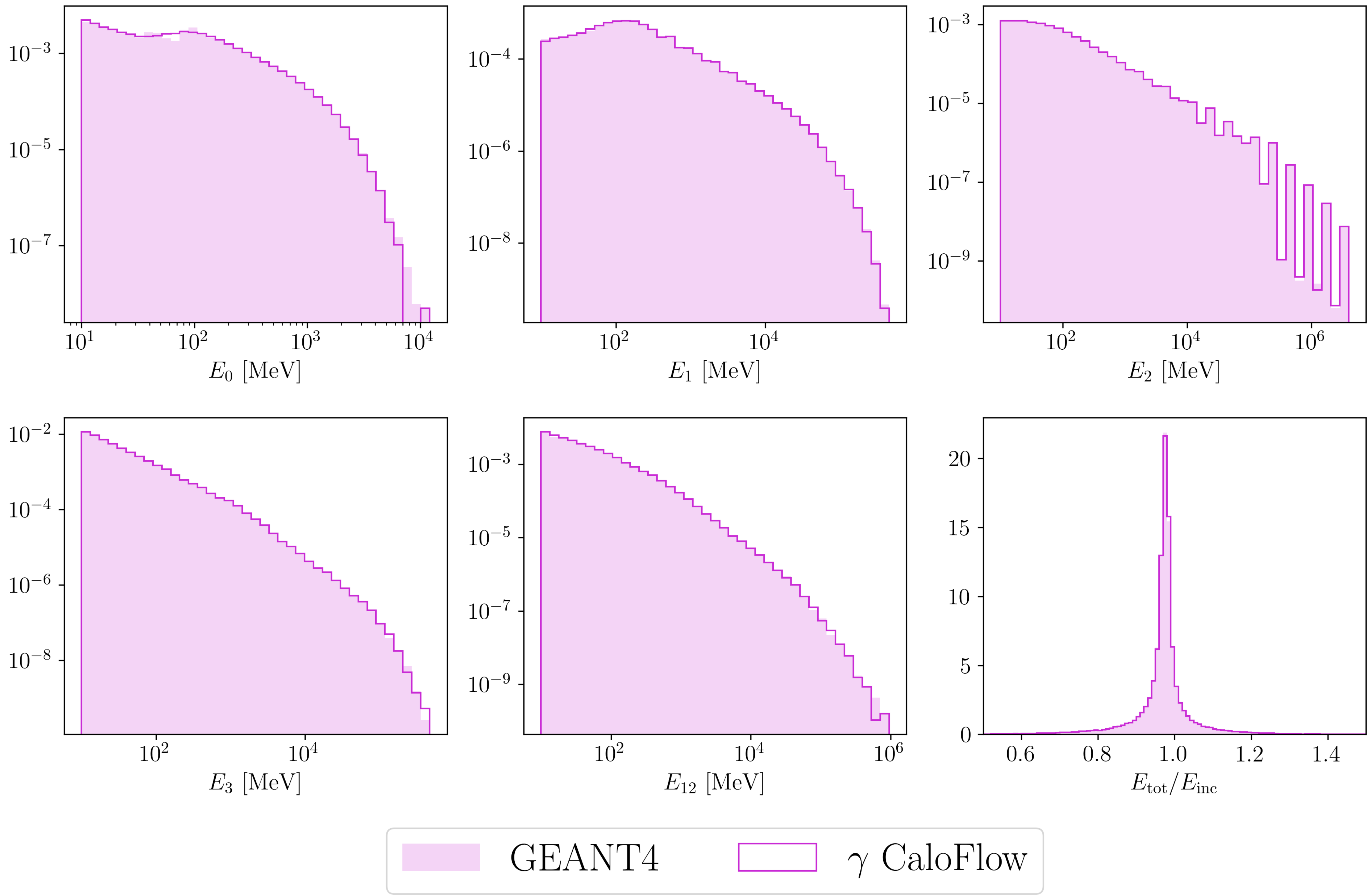
https://calochallenge.github.io/homepage/

Fast Calorimeter Surrogate Modeling

Many different approaches tested on this task!

- GANs
- VAEs
- **Normalizing flows**
- Diffusion/Flow matching*

Access to likelihood!



$\mathcal{O}(10^4) - \mathcal{O}(10^5)$ times faster than GEANT4

[2210.14245] C. Krause, **IP**, D. Shih

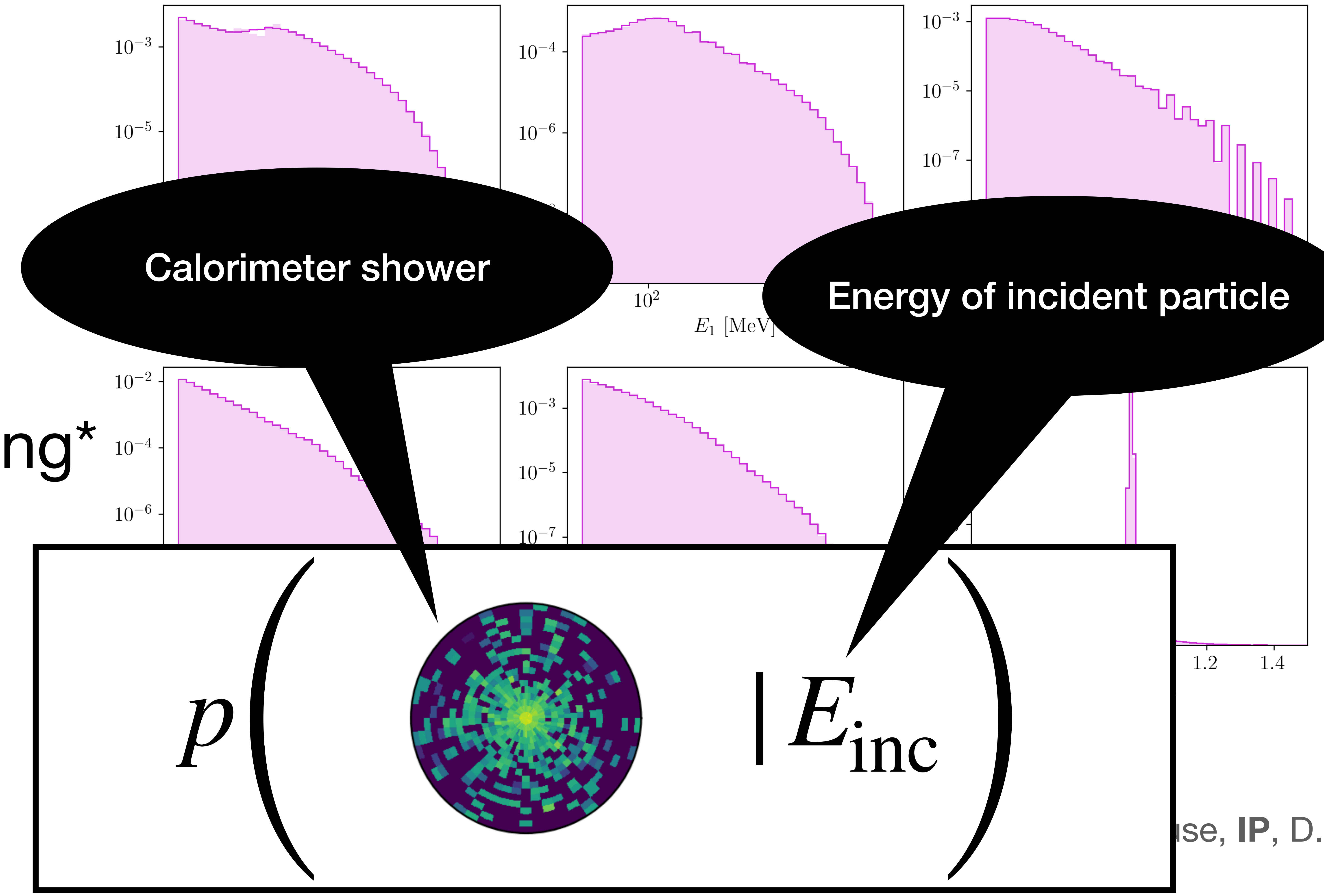
* Likelihood may be obtained from diffusion/flow matching models as well. However, it is often more difficult to do so.

Fast Calorimeter Surrogate Modeling

Many different approaches tested on this task!

- GANs
- VAEs
- **Normalizing flows**
- Diffusion/Flow matching*

Access to likelihood!



se, **IP**, D. Shih

* Likelihood may be obtained from diffusion/flow matching models as well. However, it is often more difficult to do so.

Once we have a trained flow-based fast calorimeter model, we get …

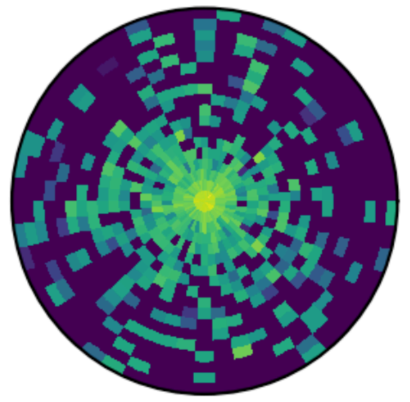
1. A **regression/calibration** model
   [2404.18992] H. Du, C. Krause, V. Mikuni, B. Nachman, **IP**, D. Shih
   - infers the particle incident energy

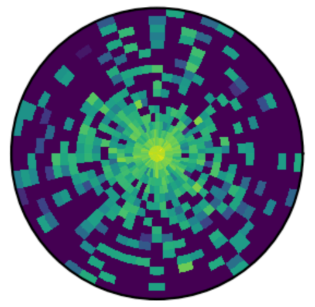
2. An **anomaly detector**
   [2312.11618] C. Krause, B. Nachman, **IP**, D. Shih
   - sensitive to new physics

All for free!

Regression of incident energy

[2404.18992] H. Du, C. Krause, V. Mikuni, B. Nachman, **IP**, D. Shih

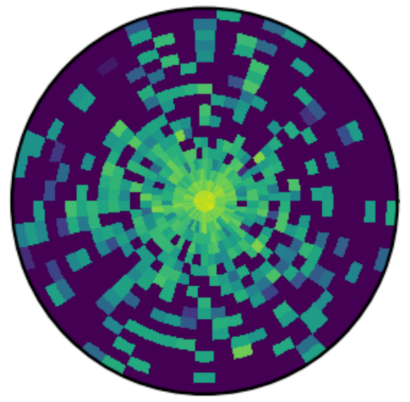
Given  , we want to infer E_{inc}

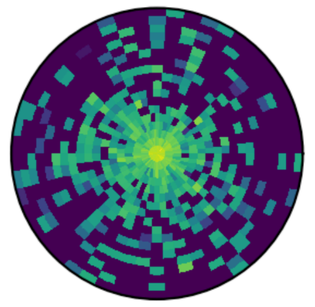
Perform maximum likelihood estimation (MLE) with $p\Big($  $| E_{\text{inc}}\Big)$

Regression of incident energy

[2404.18992] H. Du, C. Krause, V. Mikuni, B. Nachman, **IP**, D. Shih

Here we consider energy of incident π^+

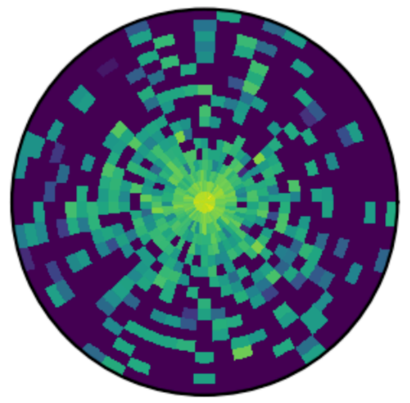
Given  , we want to infer E_{inc}

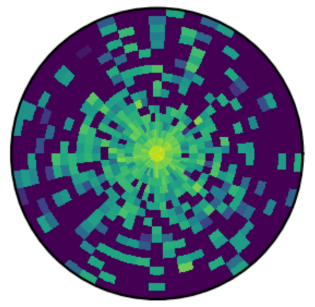
Perform maximum likelihood estimation (MLE) with $p\left(\right.$  $\left.| E_{\text{inc}}\right)$

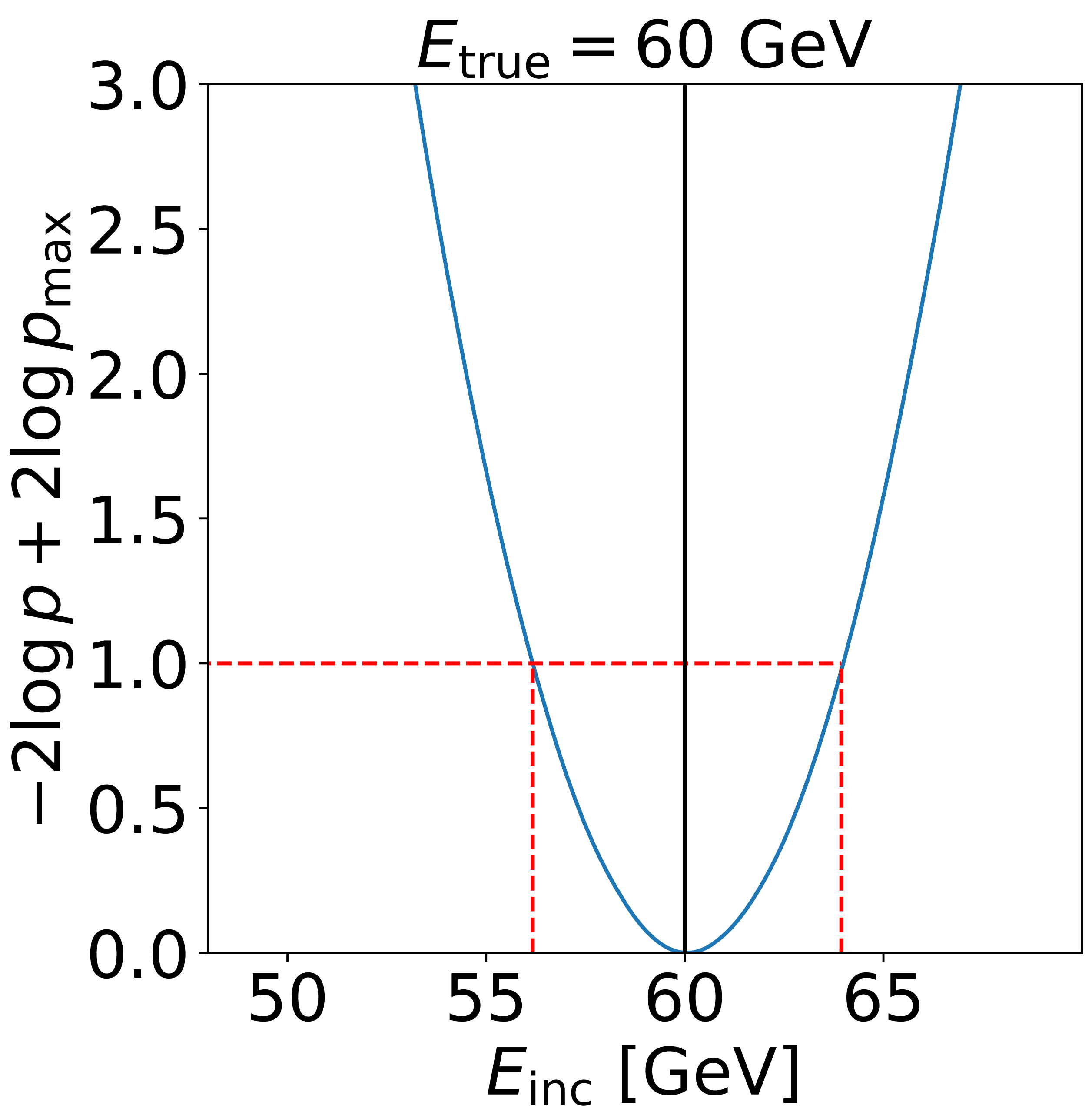
Regression of incident energy

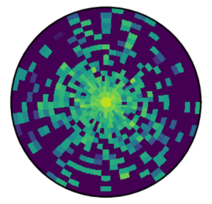
[2404.18992] H. Du, C. Krause, V. Mikuni, B. Nachman, **IP**, D. Shih

Here we consider energy of incident π^+

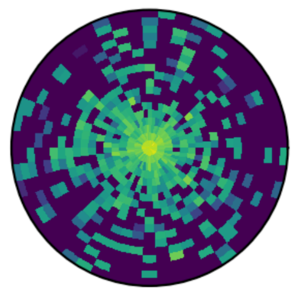
Given  , we want to infer $E_{\rm inc}$

Perform maximum likelihood estimation (MLE) with $p\left(\right.$  $\left. | E_{\rm inc} \right)$



- $p\left(\right.$  $\left. | E_{\rm inc} \right)$: Blue curve
- True $E_{\rm inc}$: Solid vertical line
- Boundary of 68% CI : Red vertical lines

Limitations of mean square error (MSE) calibration

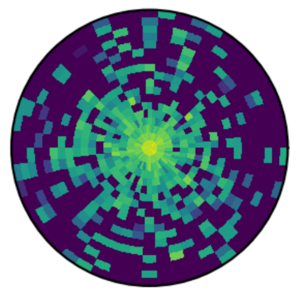
x_i : 

z_i : E_{inc}

Want to regress z_i given x_i

Loss function: $$L[f] = \sum_i (f_{\text{MSE}}(x_i) - z_i)^2 \,,$$

Limitations of mean square error (MSE) calibration

x_i : 

$z_i : E_{\text{inc}}$

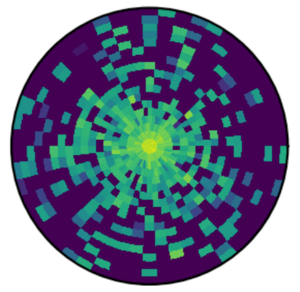
Want to regress z_i given x_i

Loss function: $$L[f] = \sum_i \left(f_{\text{MSE}}(x_i) - z_i\right)^2 ,$$

$$\begin{aligned} f_{\text{MSE}}(x) &= \langle Z \,|\, X = x \rangle \\ &= \int dz\, z\, p^{\text{train}}_{Z|X}(z\,|\,x) \\ &= \int dz\, z\, p^{\text{train}}_{X|Z}(x\,|\,z)\, \frac{p^{\text{train}}_{Z}(z)}{p^{\text{train}}_{X}(x)} \end{aligned}$$

Prior dependent!

Limitations of mean square error (MSE) calibration

x_i : 

$z_i : E_{\text{inc}}$

Want to regress z_i given x_i

Loss function: $$L[f] = \sum_i (f_{\text{MSE}}(x_i) - z_i)^2 \,,$$

$$f_{\text{MSE}}(x) = \langle Z \,|\, X = x \rangle$$
$$= \int dz\, z\, p^{\text{train}}_{Z|X}(z \,|\, x)$$
$$= \int dz\, z\, p^{\text{train}}_{X|Z}(x \,|\, z) \, \frac{p^{\text{train}}_Z(z)}{p^{\text{train}}_X(x)}$$

Prior dependent!

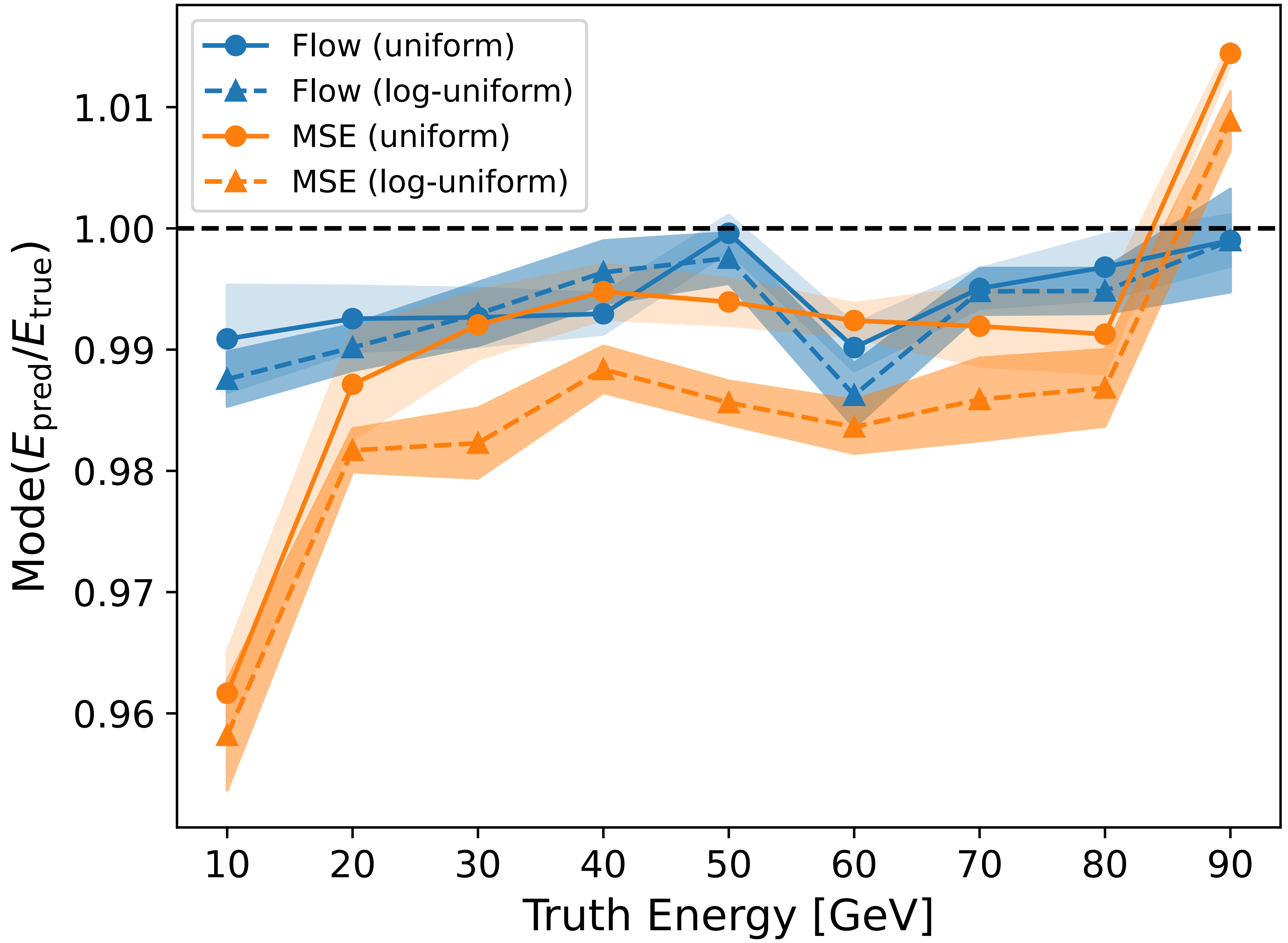
Only point estimate!
(No uncertainty quantification)

Regression of incident energy

[2404.18992] H. Du, C. Krause, V. Mikuni, B. Nachman, **IP**, D. Shih

1. MLE (flow) calibration is independent of the prior $p(E_{\mathrm{inc}})$

- MSE-based calibration depends on $p(E_{\mathrm{inc}})$
- Our calibration is less biased!
  - **Bias**: Deviation of <u>average</u> prediction from true answer

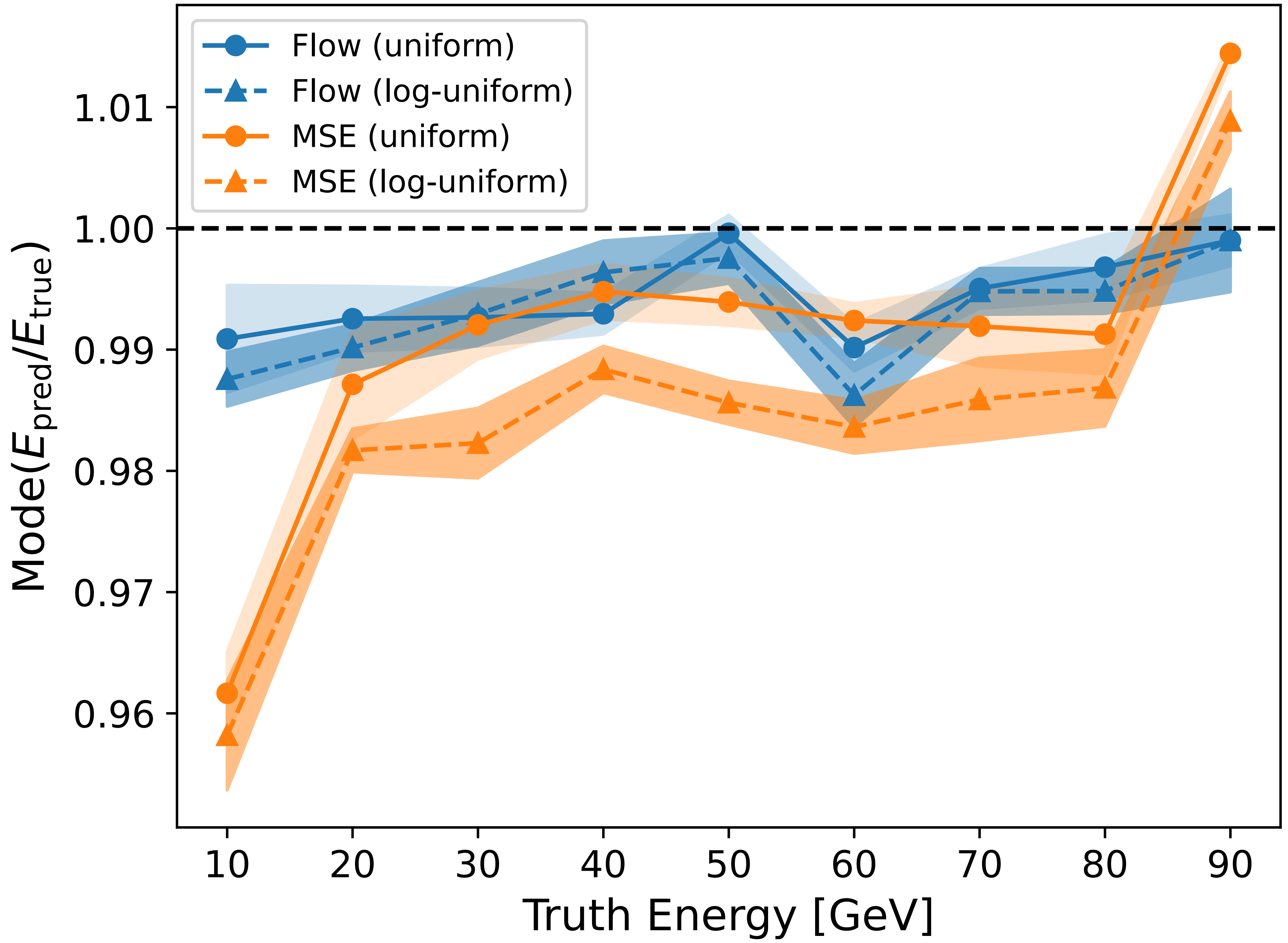


Regression of incident energy

[2404.18992] H. Du, C. Krause, V. Mikuni, B. Nachman, **IP**, D. Shih

1. MLE (flow) calibration is independent of the prior $p(E_{\mathrm{inc}})$

- MSE-based calibration depends on $p(E_{\mathrm{inc}})$
- Our calibration is less biased!
  - **Bias**: Deviation of <u>average</u> prediction from true answer
  - Mode (average) of $p(E_{\mathrm{pred}}/E_{\mathrm{true}})$ at fixed E_{true} closer to 1

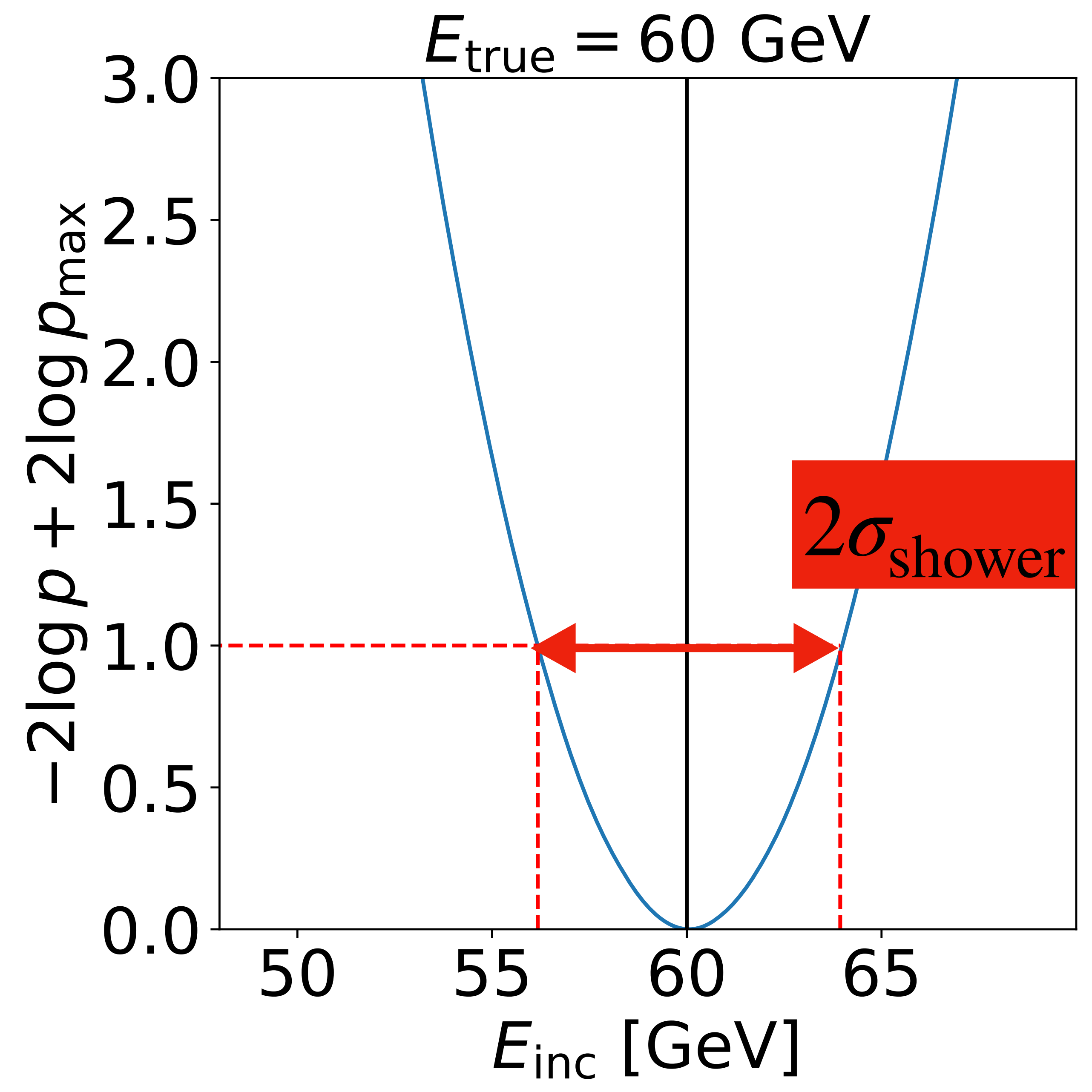


Regression of incident energy

[2404.18992] H. Du, C. Krause, V. Mikuni, B. Nachman, **IP**, D. Shih

2. Access to per-shower resolution σ_{shower}

- MSE-based calibration gives point estimates (no uncertainty quantification)
- Reliable per-shower resolution

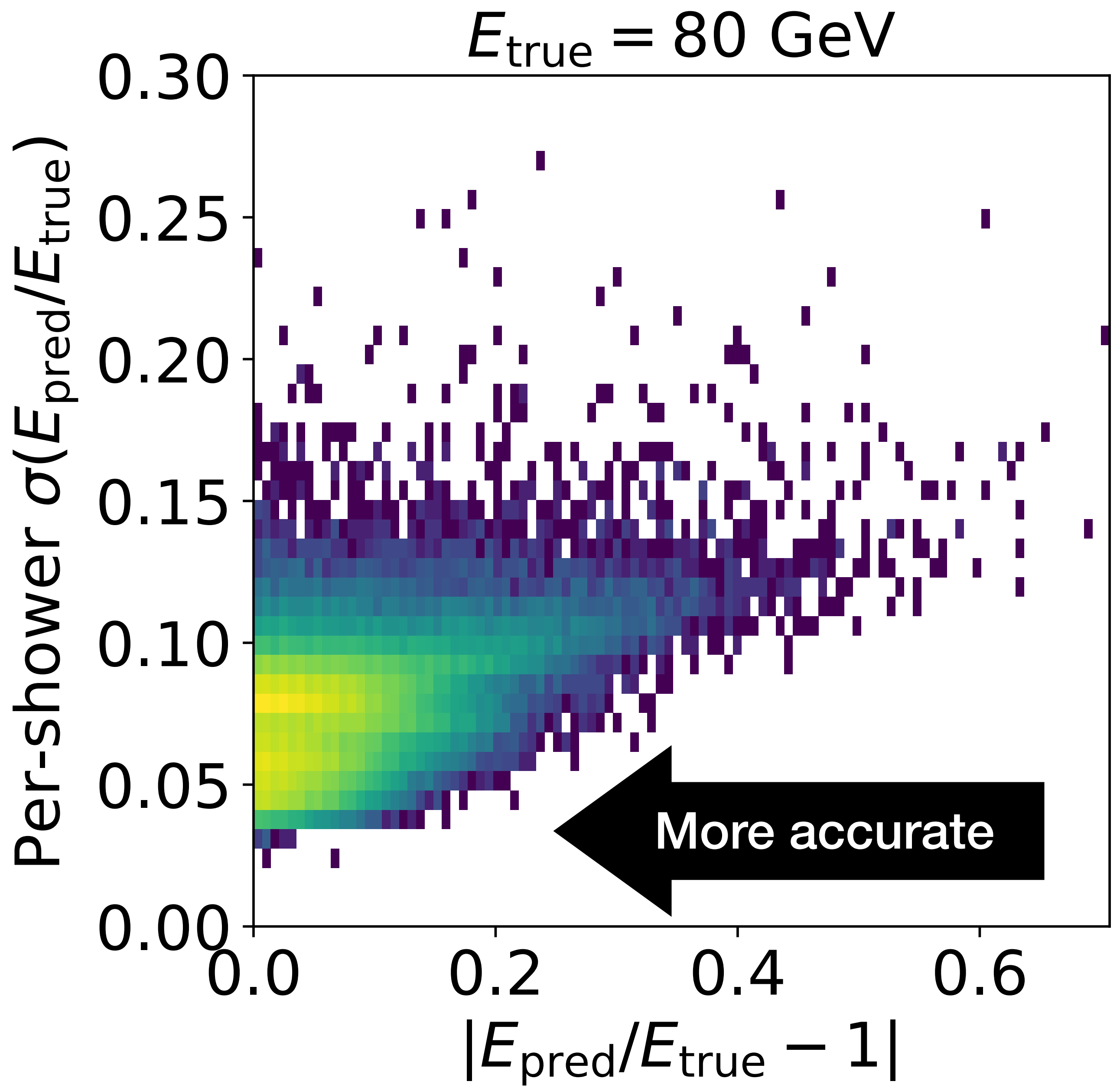


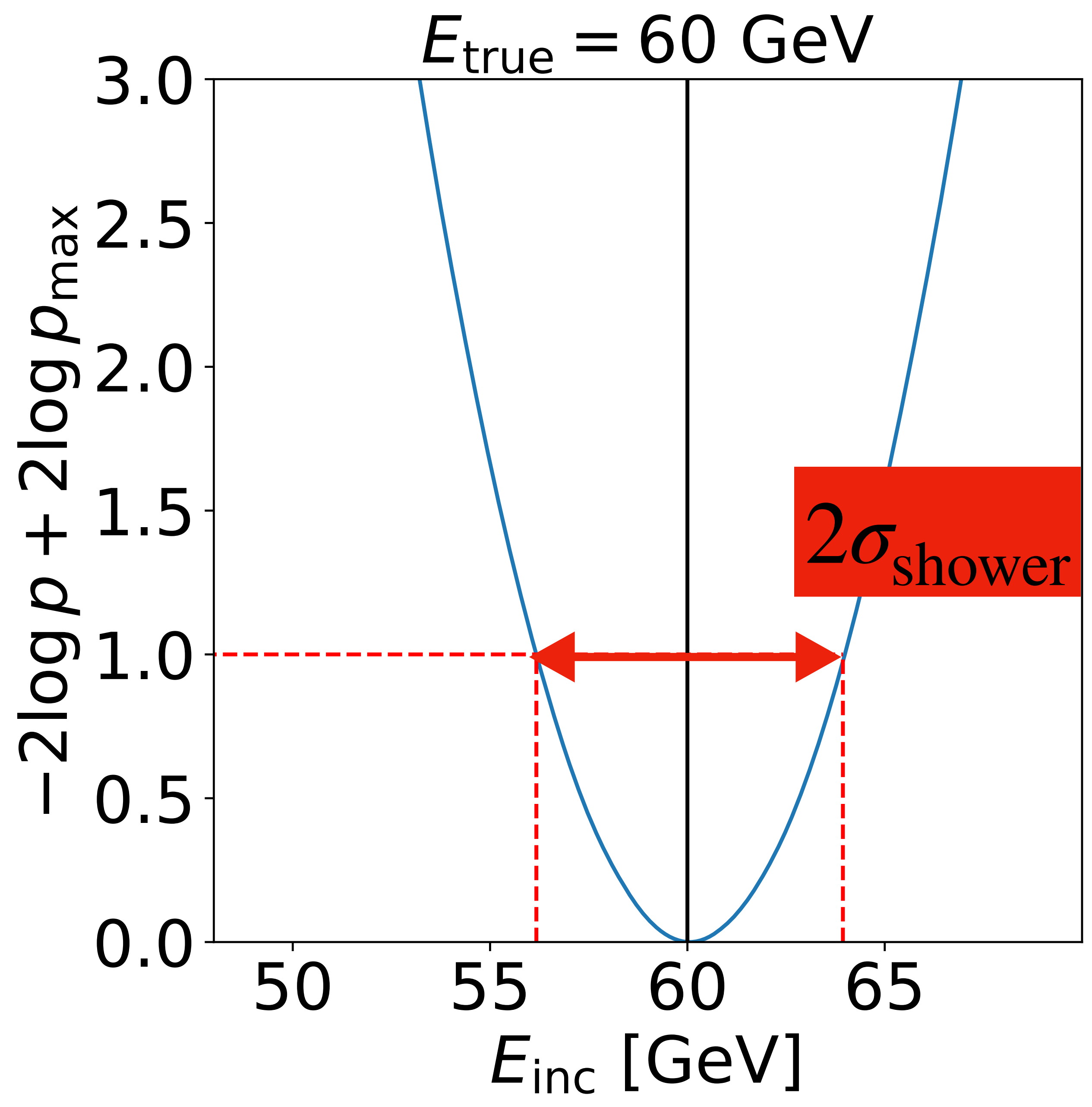
Regression of incident energy

[2404.18992] H. Du, C. Krause, V. Mikuni, B. Nachman, **IP**, D. Shih

2. Access to per-shower resolution σ_{shower}

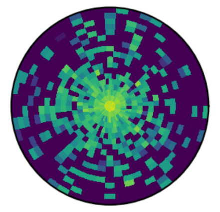
- MSE-based calibration gives point estimates (no uncertainty quantification)
- Reliable per-shower resolution

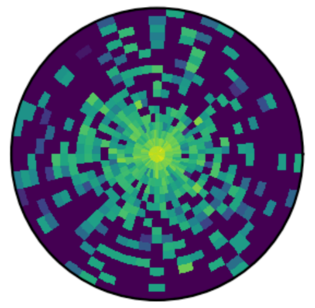




Anomaly detection

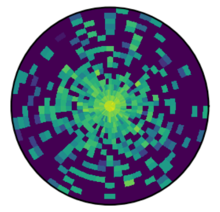
[2312.11618] C. Krause, B. Nachman, **IP**, D. Shih

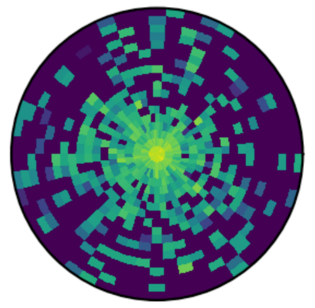
Flow trained to maximize $p\Big($  $|\, E_{\text{inc}}\Big)$ for incident SM particle (e.g. photon)

Detect BSM anomalies by making cut on $p\Bigg($  $|\, E_{\text{inc}}\Bigg)$

Anomaly detection

[2312.11618] C. Krause, B. Nachman, **IP**, D. Shih

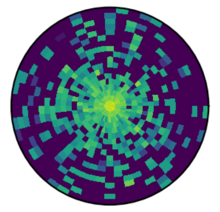
Flow trained to maximize $p\Big($  $\,|\,E_{\text{inc}}\Big)$ for incident SM particle (e.g. photon)

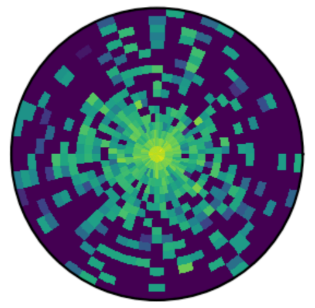
Detect BSM anomalies by making cut on $p\Big($  $\,|\,E_{\text{inc}}\Big)$

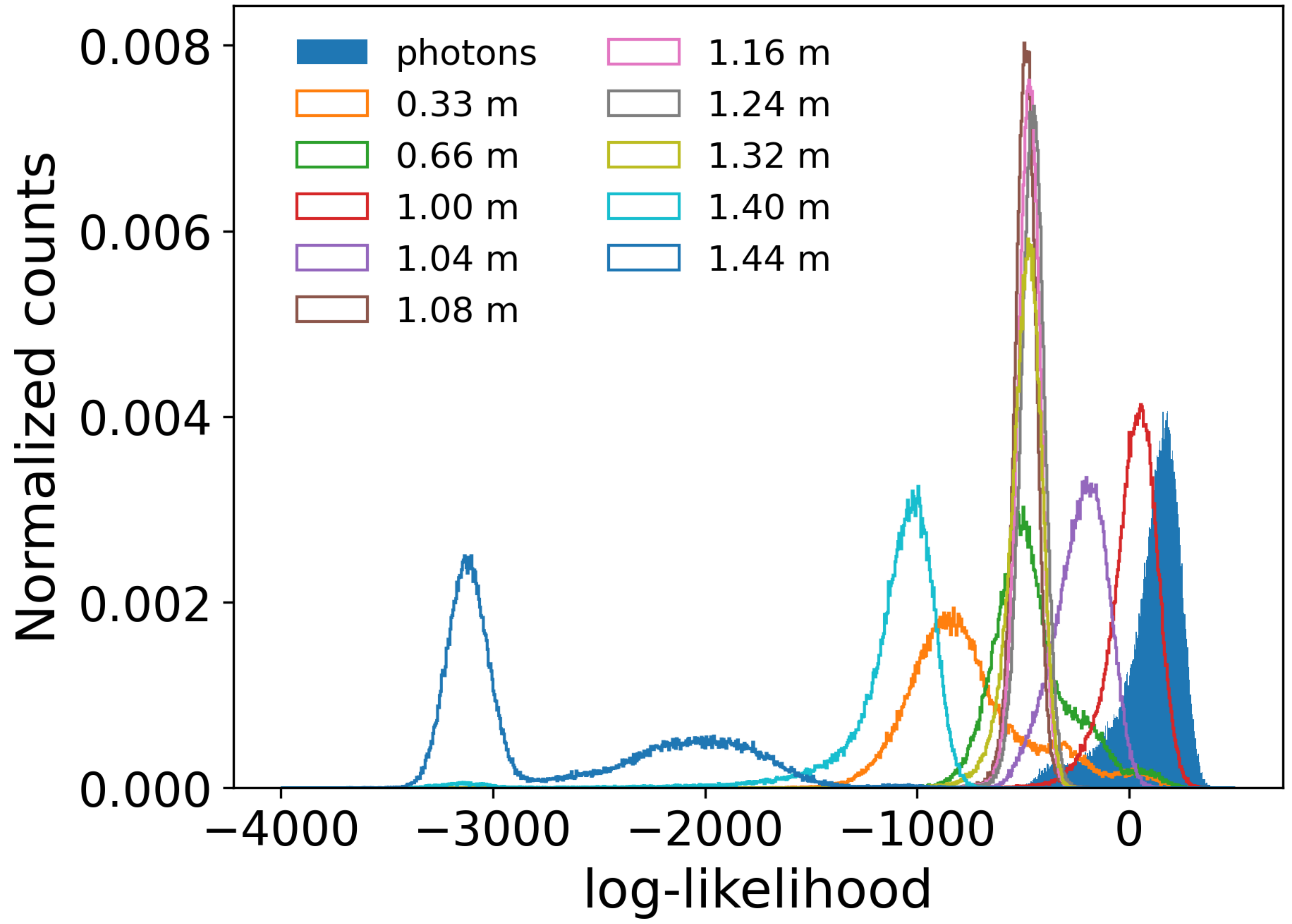
No access to E_{inc}:
Use reconstructed energy $E_{\text{inc}}^{(\text{rec})} = \lambda E_{\text{dep}}$

Anomaly detection

[2312.11618] C. Krause, B. Nachman, **IP**, D. Shih

Flow trained to maximize $p\Big($  $|\,E_{\text{inc}}\Big)$ for <u>incident SM particle</u> (e.g. photon)

Detect BSM anomalies by making cut on $p\Big($  $|\,E_{\text{inc}}\Big)$



- Invisible pseudoscalar particle χ
- $\chi \to \gamma\gamma$ (highly boosted)
- Consider different masses and lifetimes

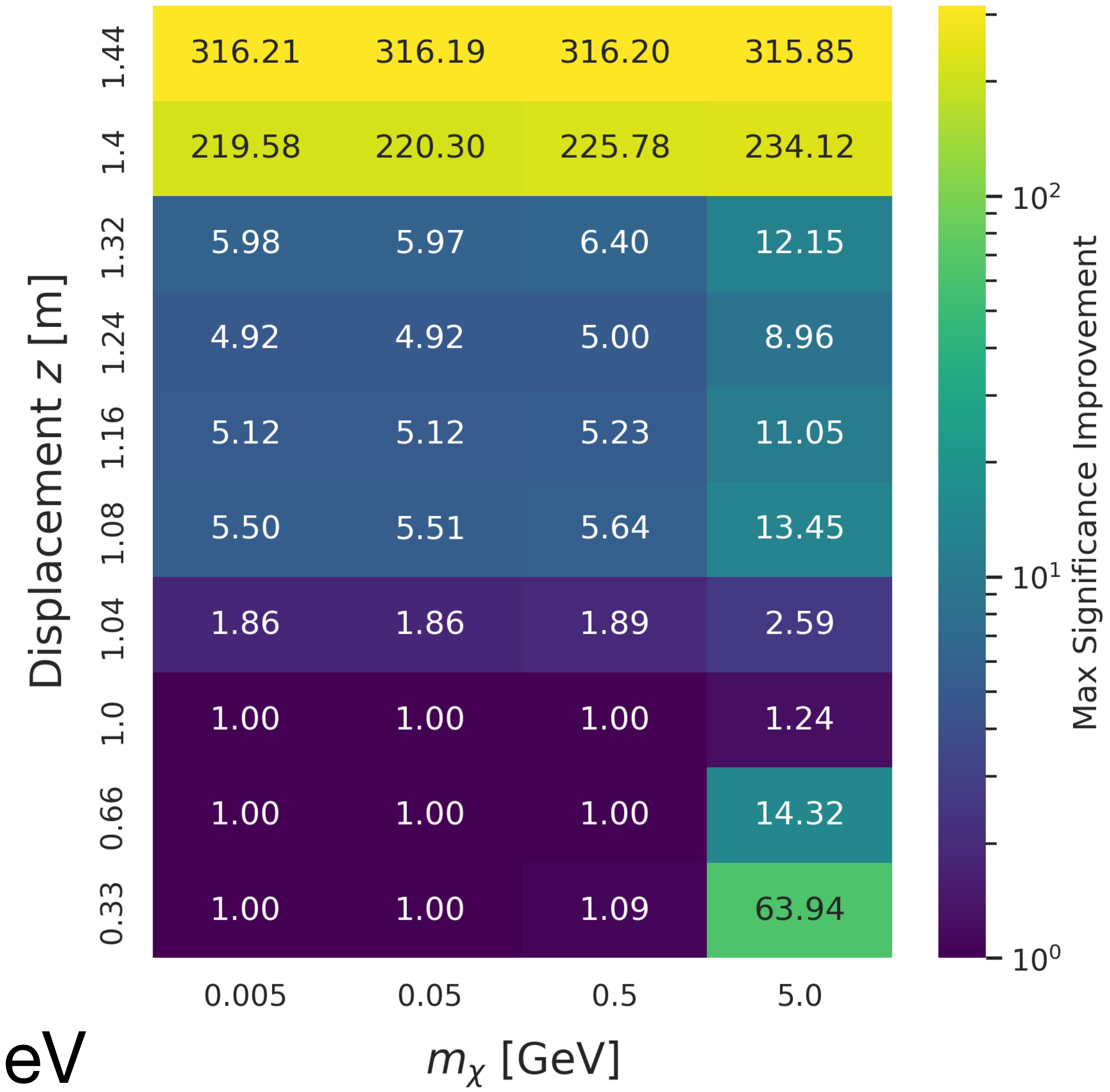
Anomaly detection

[2312.11618] C. Krause, B. Nachman, **IP**, D. Shih

$$\text{Significance improvement} = \frac{\text{True positive rate}}{\sqrt{\text{False positive rate}}}$$

Unsupervised anomaly detection

- Relatively model-agnostic (only assumed photon showers)
- Able to distinguish a variety of anomalous showers from SM showers



Energy of χ = 50 GeV

Anomaly detection

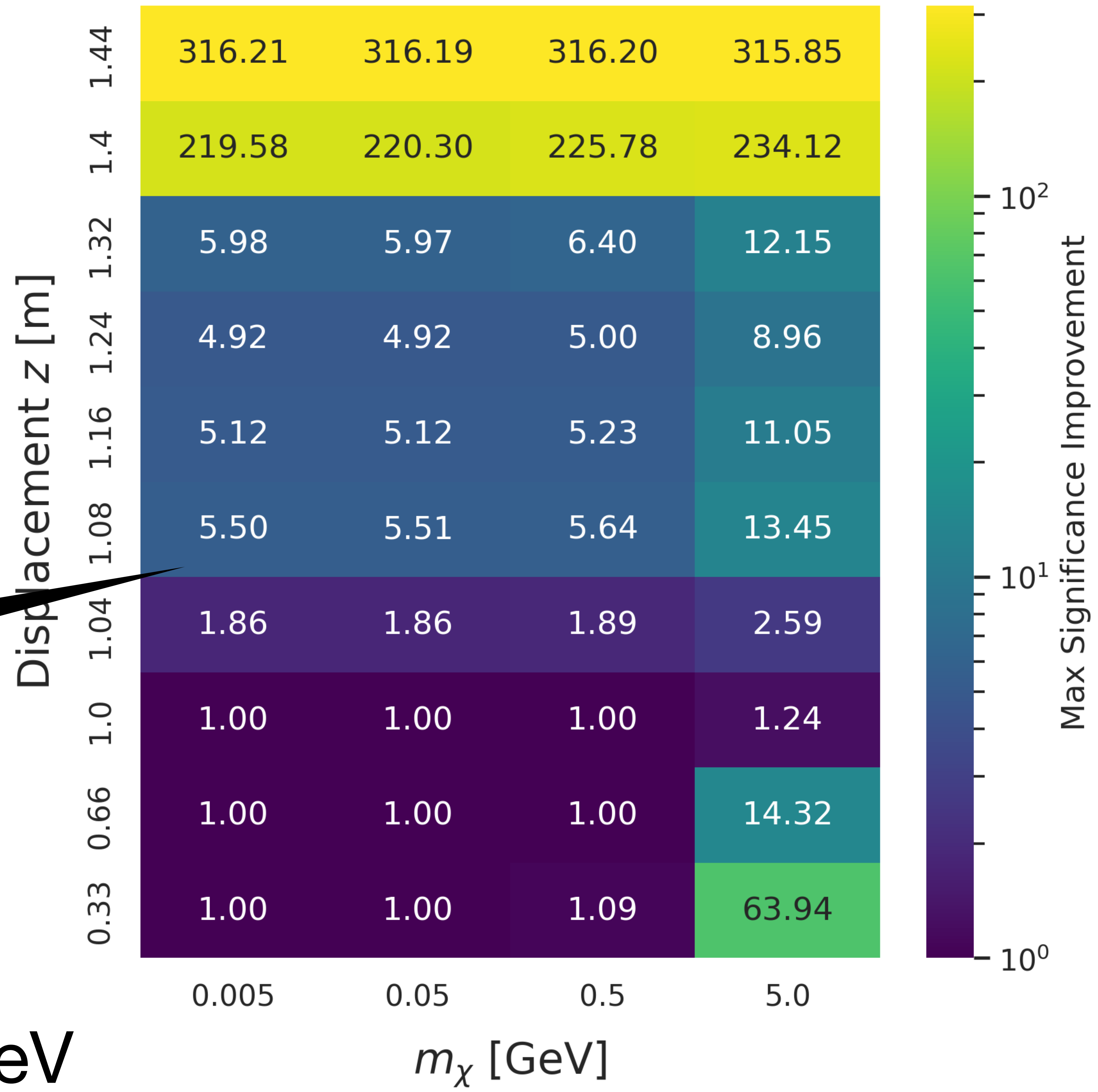
[2312.11618] C. Krause, B. Nachman, **IP**, D. Shih

Significance improvement = $\frac{\text{True positive rate}}{\sqrt{\text{False positive rate}}}$

Unsupervised anomaly detection

- Relatively model-agnostic (only assumed photon showers)
- Able to distinguish a variety of anomalous showers from SM showers

Mostly > 1



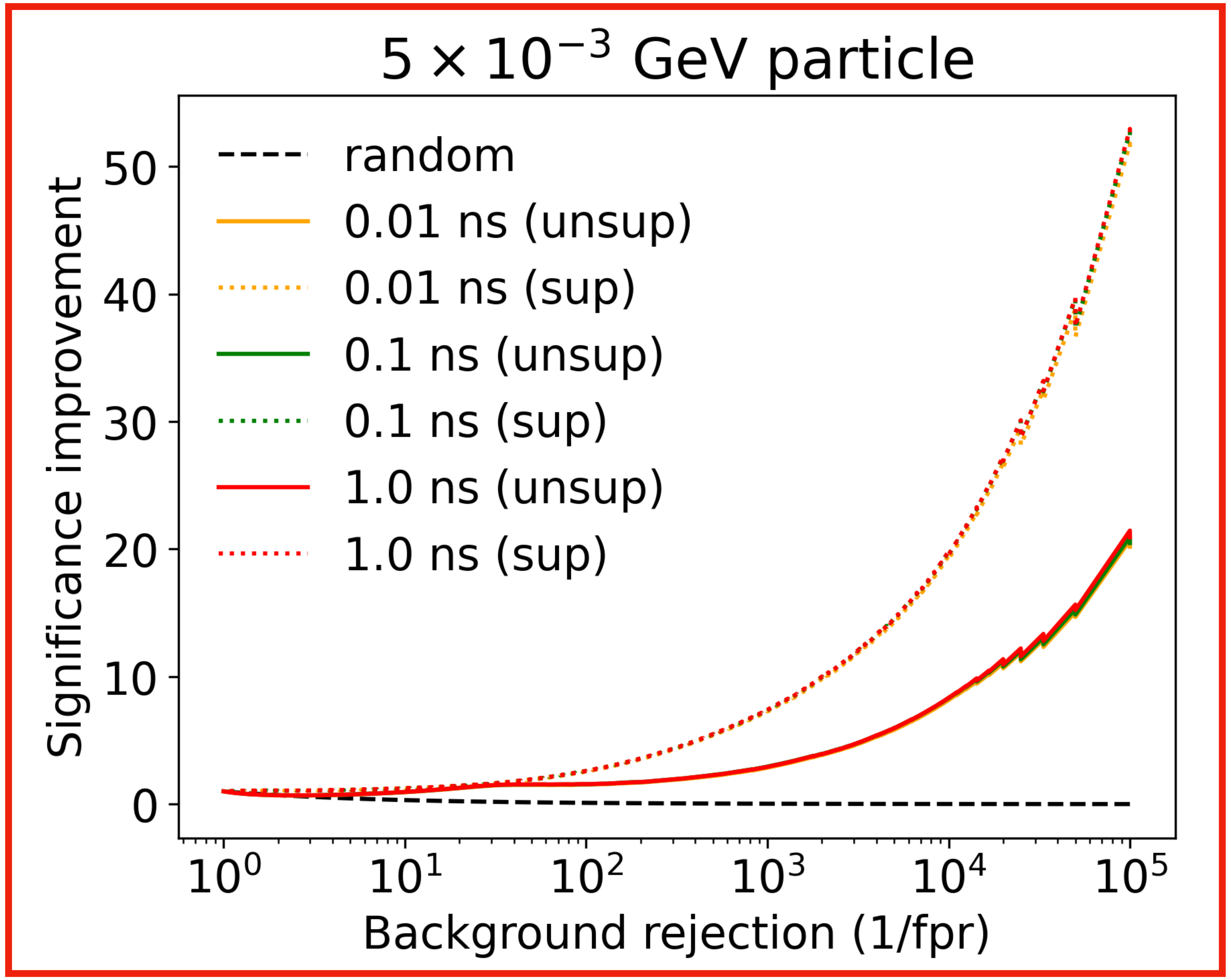
Energy of χ = 50 GeV

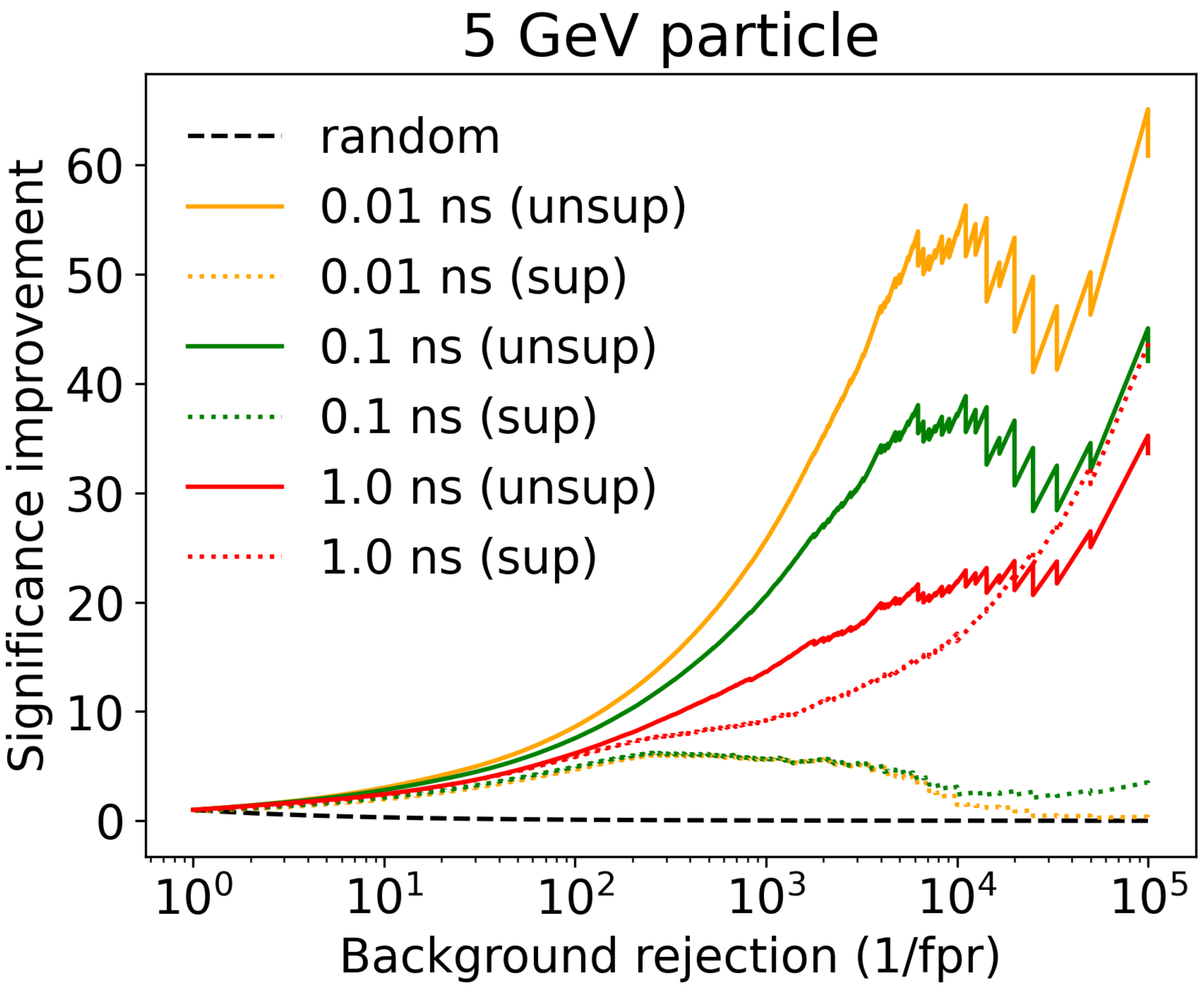
Anomaly detection

$$\text{Significance improvement} = \frac{\text{True positive rate}}{\sqrt{\text{False positive rate}}}$$

[2312.11618] C. Krause, B. Nachman, **IP**, D. Shih

Trained on $m_\chi = 5 \times 10^{-3}$ GeV, lifetime = 1 ns :





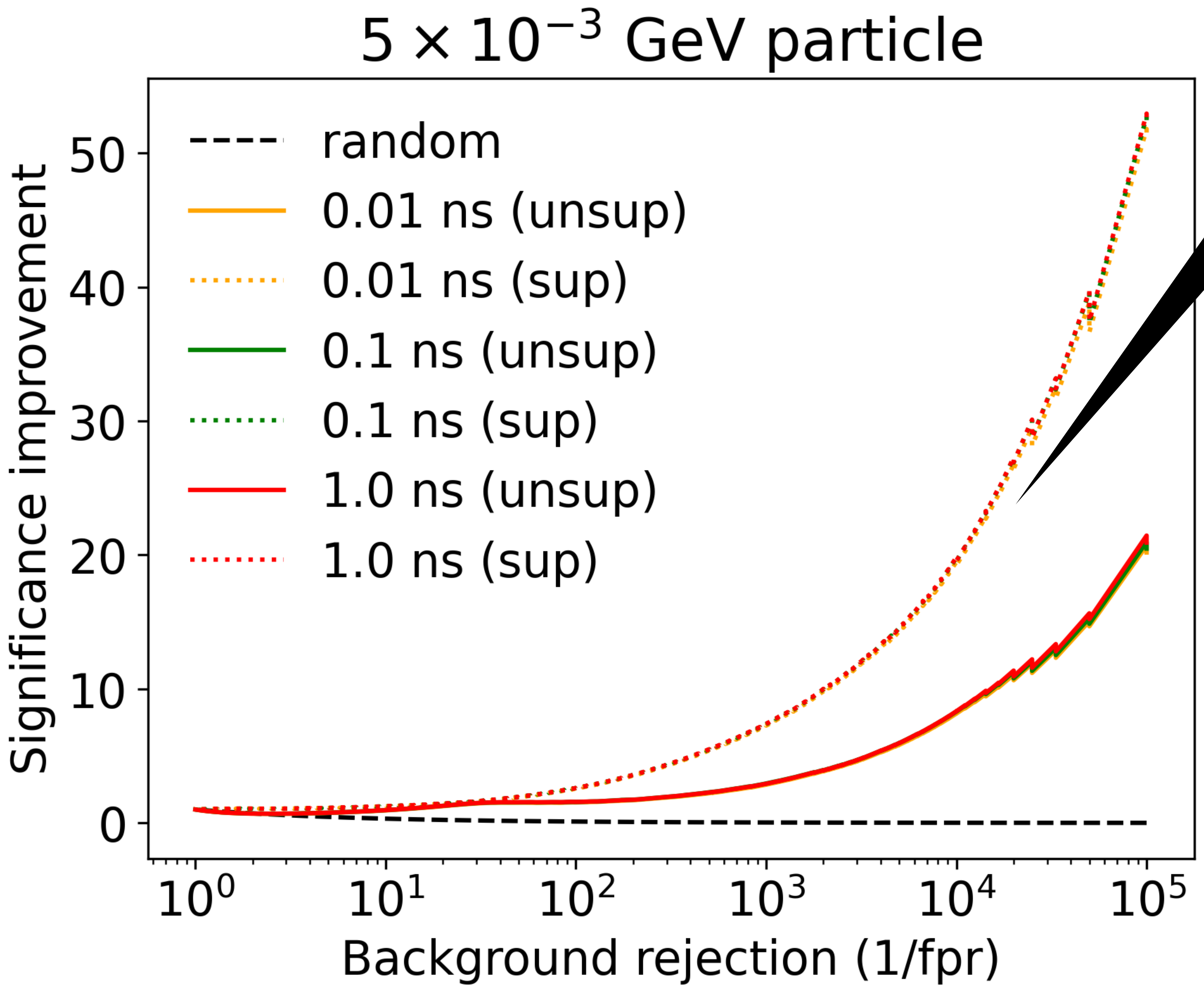
Anomaly detection

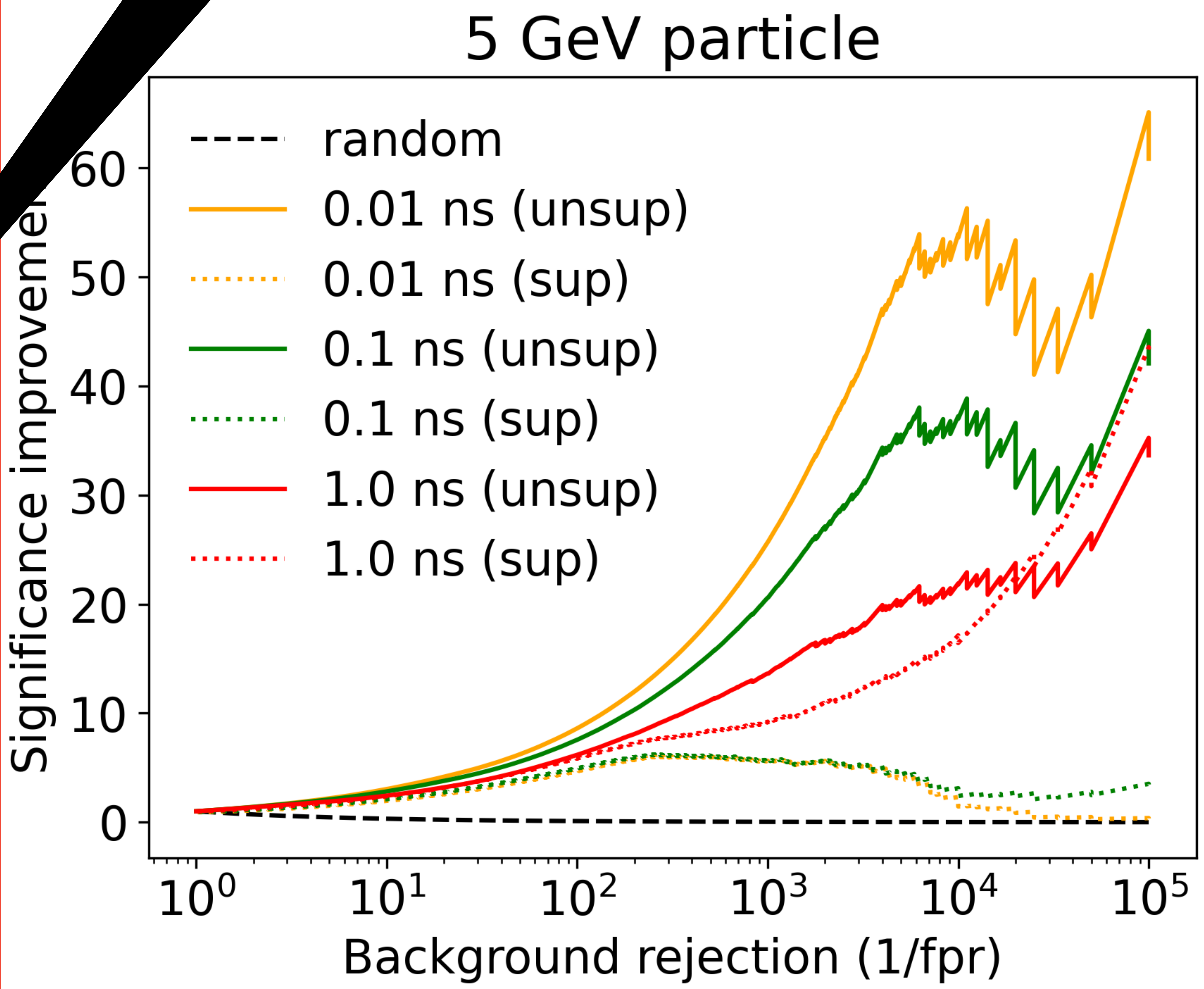
Significance improvement = $\frac{\text{True positive rate}}{\sqrt{\text{False positive rate}}}$

[2312.11618] C. Krause, B. Nachman, **IP**, D. Shih

Trained on $m_\chi = 5 \times 10^{-3}$ GeV, lifetime = 1

Supervised outperforms unsupervised



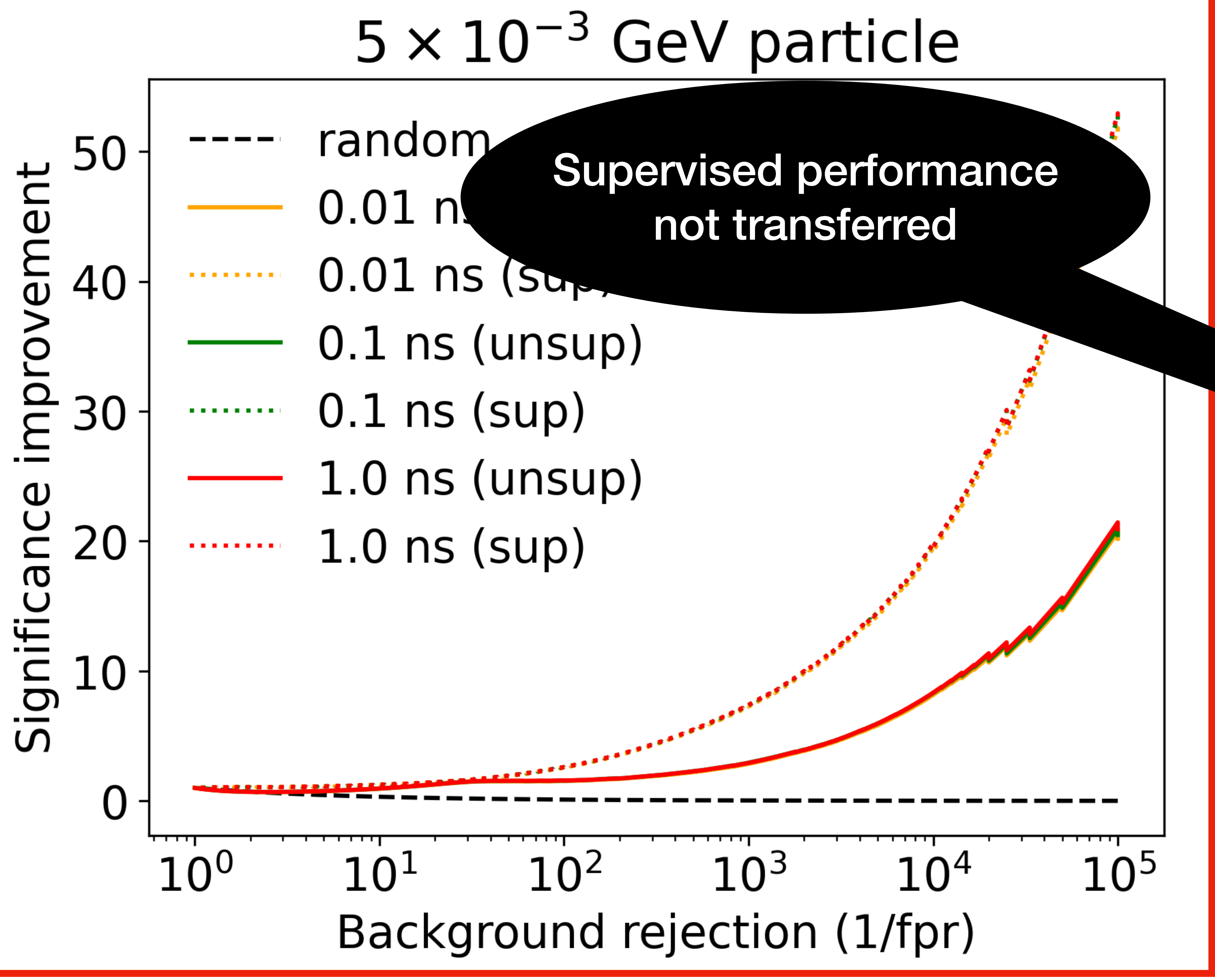


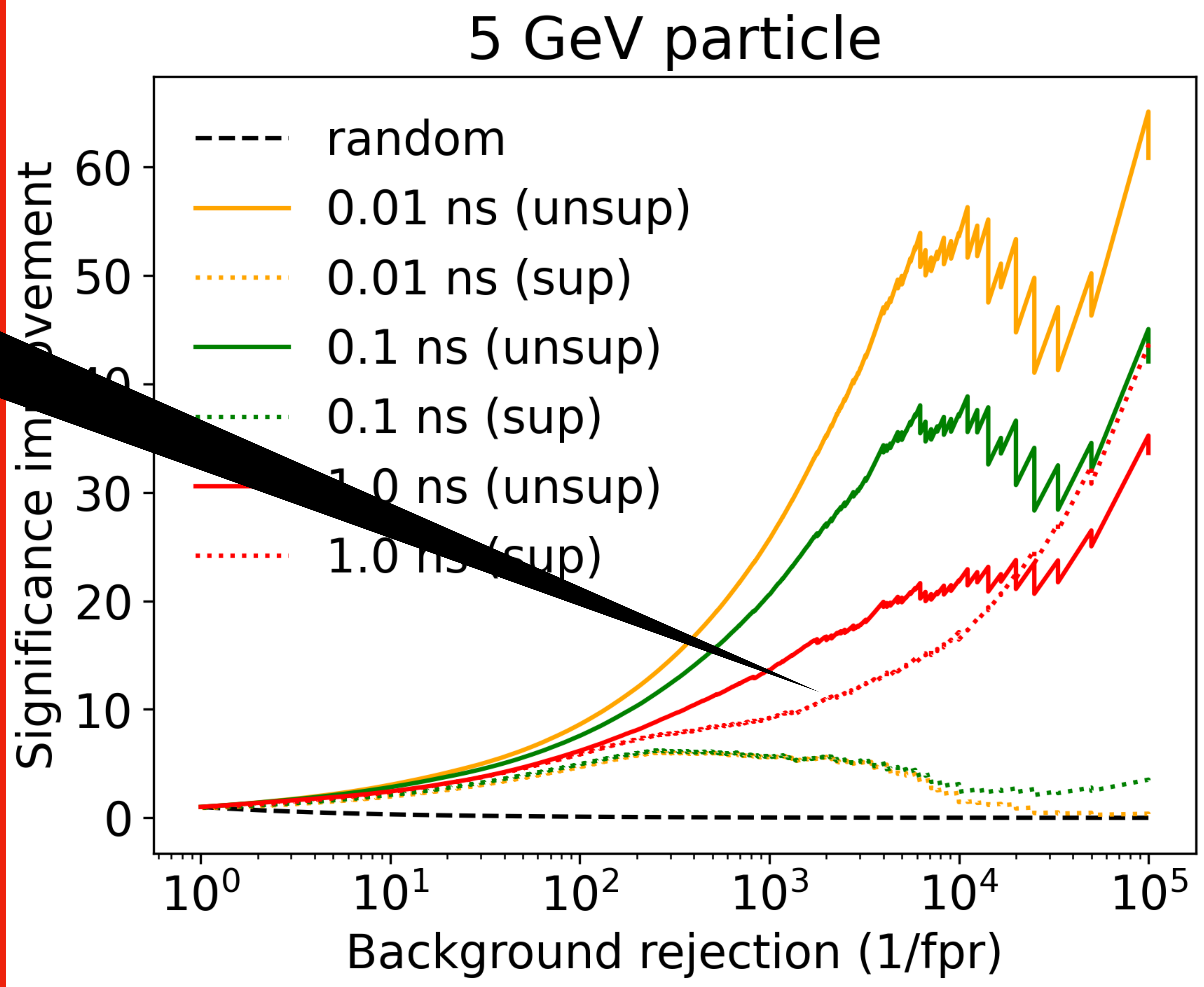
Anomaly detection

$$\text{Significance improvement} = \frac{\text{True positive rate}}{\sqrt{\text{False positive rate}}}$$

[2312.11618] C. Krause, B. Nachman, **IP**, D. Shih

Trained on $m_\chi = 5 \times 10^{-3}$ GeV, lifetime = 1 ns :





Conclusions

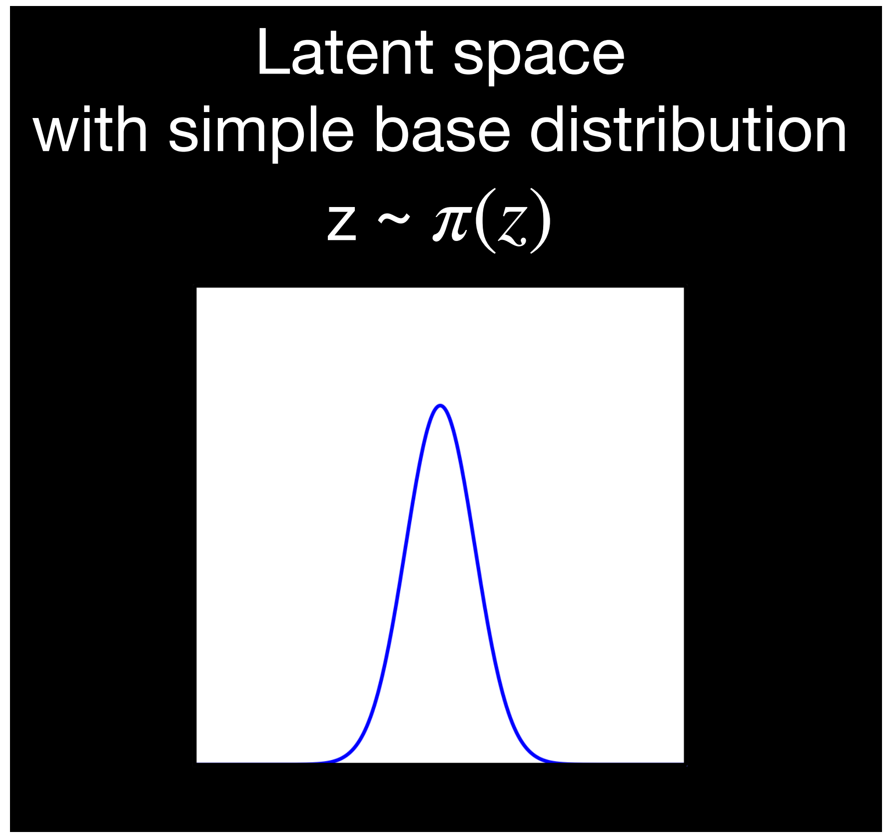
- **Normalizing flows** are state-of-the-art fast calorimeter surrogate models with **access to the likelihood**
- Flow-based fast calorimeter surrogate models can be **repurposed** to do calibration and anomaly detection
- Calibration model is **less biased** than typical direct regression and provides **per-shower resolutions**
- **Unsupervised** anomaly detection that is **model agnostic**

ian.pang@physics.rutgers.edu

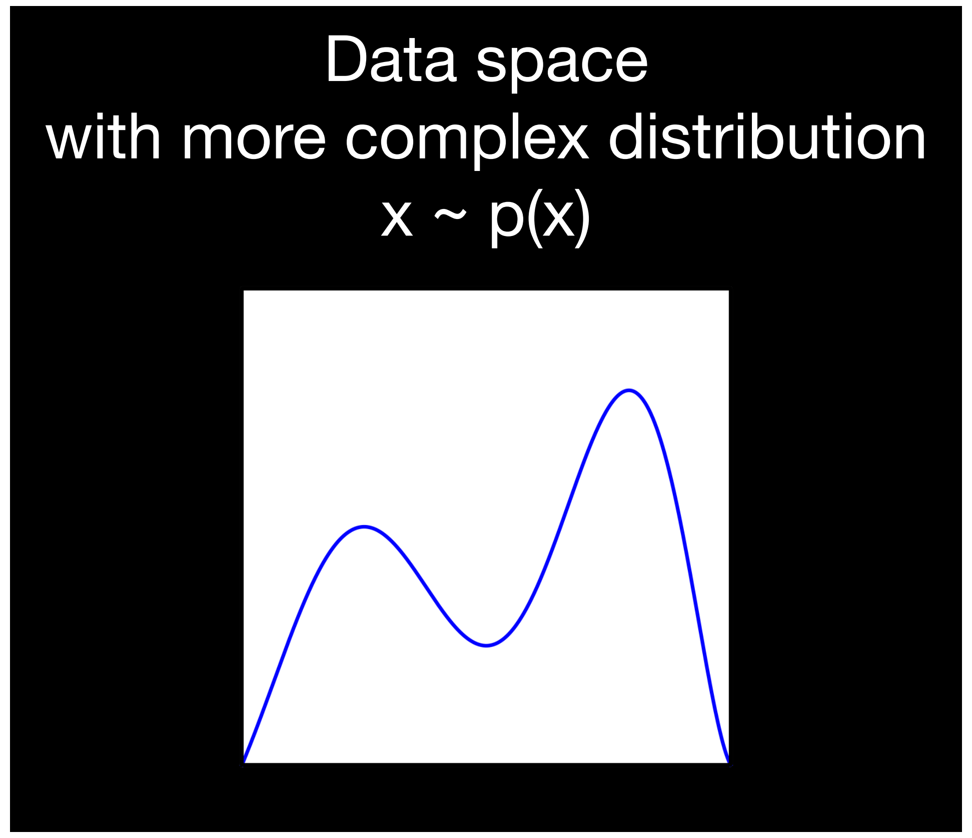
Thank you!

Backup

Normalizing Flows



Bijective transformation f



$$p(x) = \pi(f(x)) \left| \det \frac{\partial f(x)}{\partial x} \right|$$

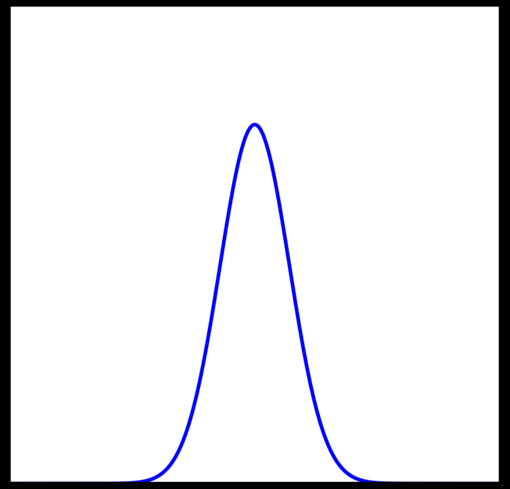
where $f(x) = z$

Density estimation, $p(x)$

Sample generation

Normalizing Flows

Latent space
with simple base distribution
z ~ $\pi(z)$

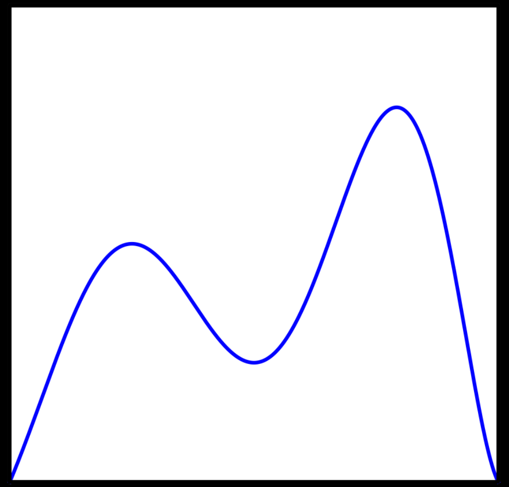


Bijective transformation f

$$p(x \mid c) = \pi(f(x \mid c)) \left| \det \frac{\partial f(x \mid c)}{\partial x} \right|$$

where $f(x \mid c) = z$

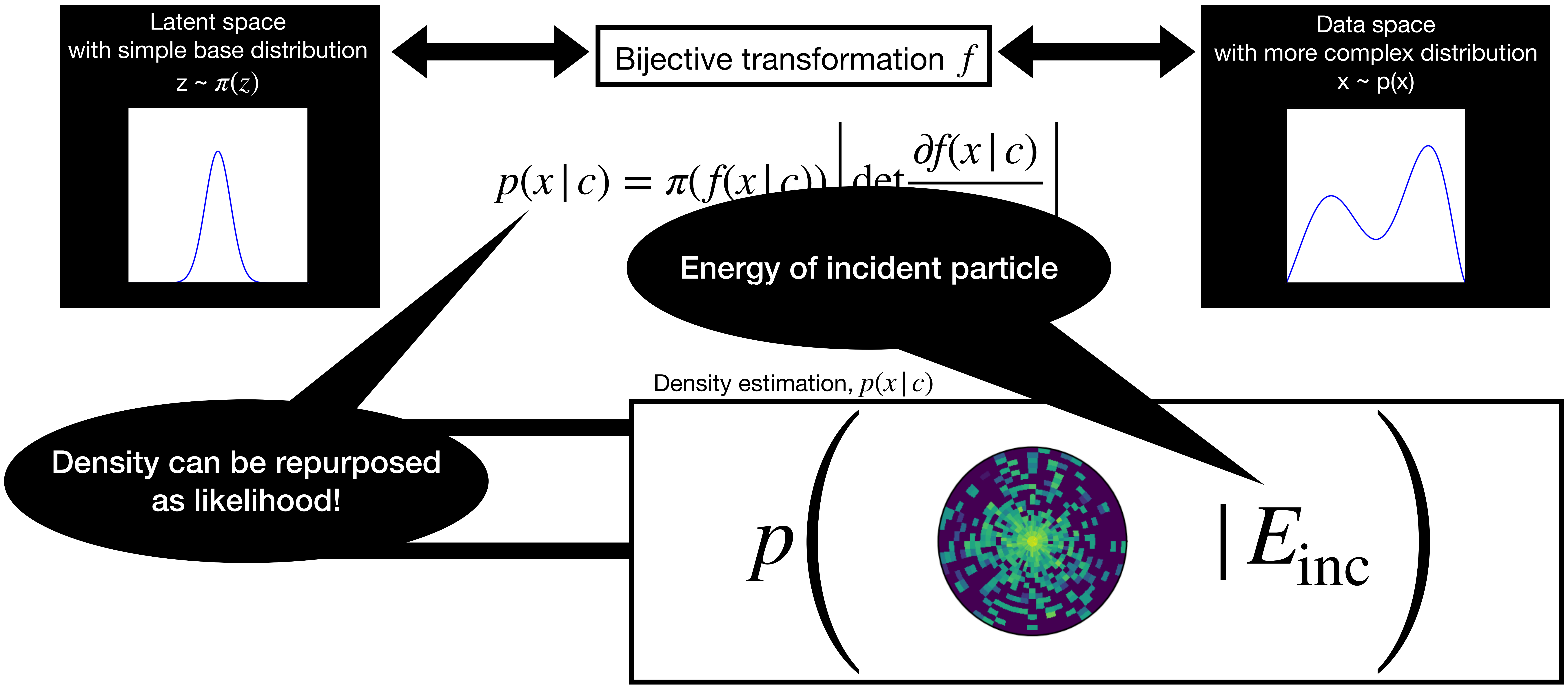
Data space
with more complex distribution
x ~ p(x)



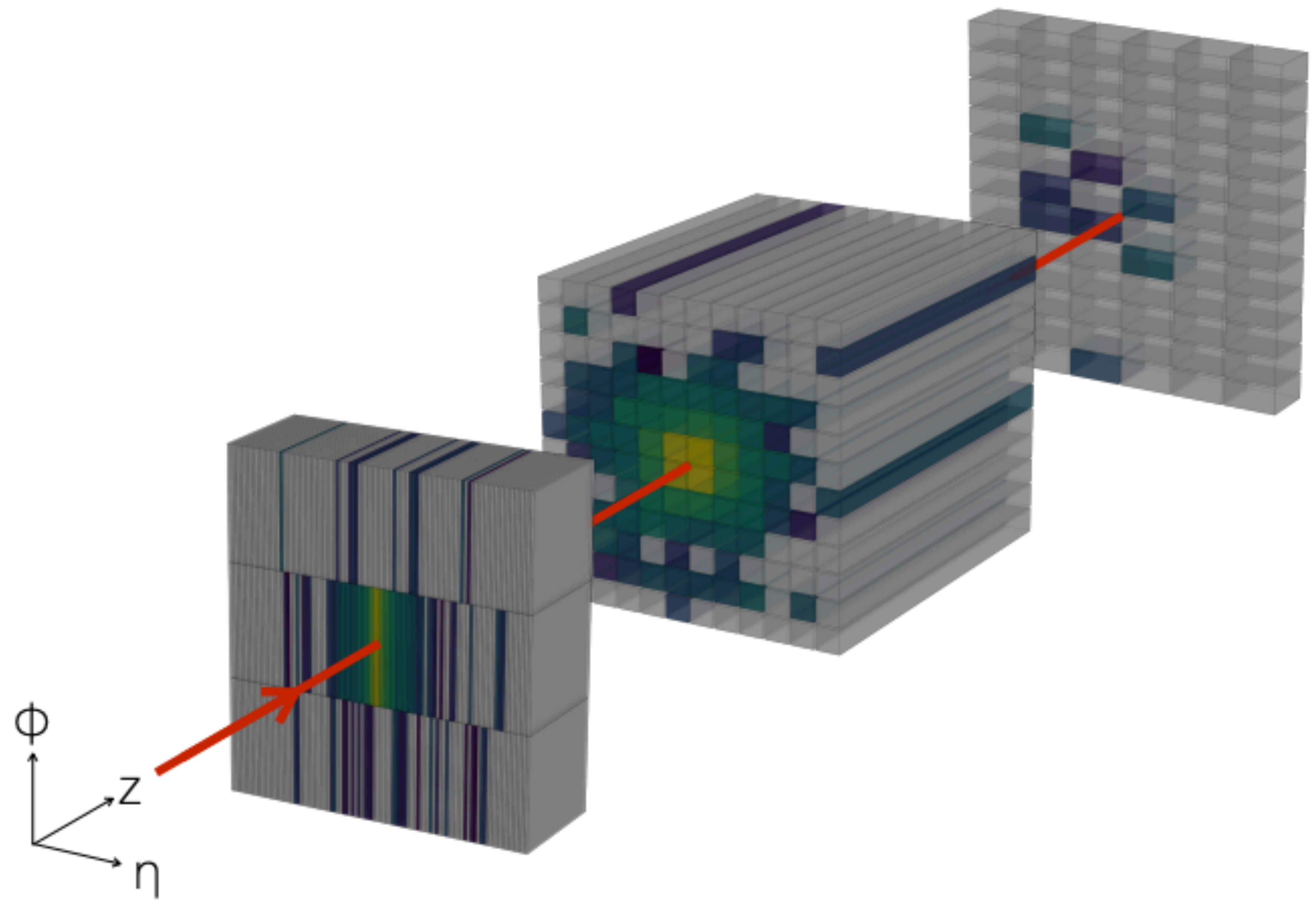
Density estimation, $p(x \mid c)$

Sample generation

Normalizing Flows

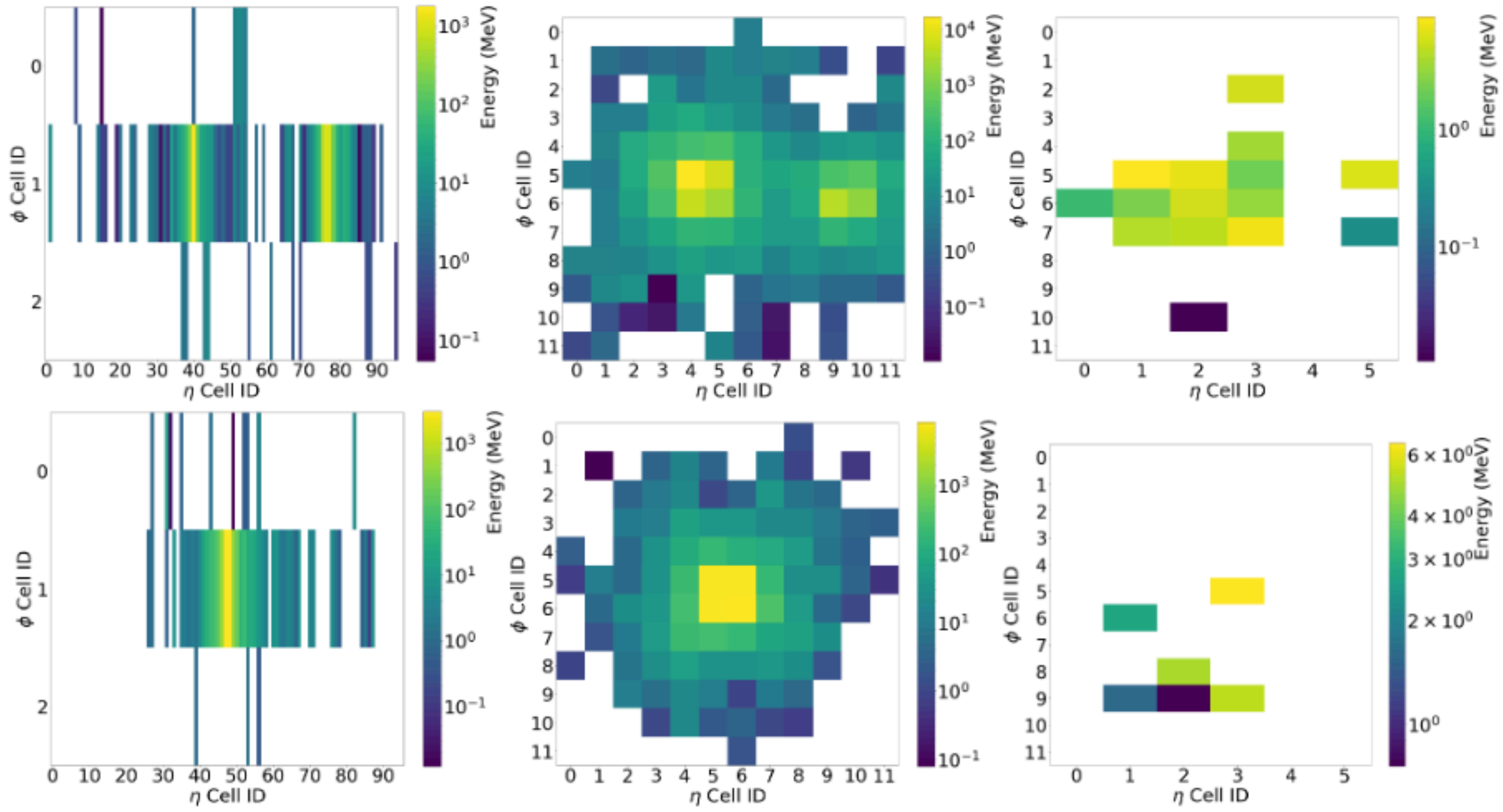


Calorimeter geometry (AD)



Layer index	z length (mm)	η length (mm)	ϕ length (mm)	Number of voxels
0	90	5	160	3×96
1	347	40	40	12×12
2	43	80	40	12×6

Two energy blobs



Reconstructed E_{inc} (AD)

- No a priori access to E_{inc}: Use reconstructed energy $E_{\mathrm{inc}}^{(\mathrm{rec})} = \lambda E_{\mathrm{dep}}$
- Can imagine performing more sophisticated calibration to get $E_{\mathrm{inc}}^{(\mathrm{rec})}$

Calorimeter geometry (calibration)

	Layer index	z length (mm)	η length (mm)	ϕ length (mm)	Number of voxels
ECAL	0	90	5	160	3×96
	1	347	40	40	12×12
	2	43	80	40	12×6
HCAL	3	375	20.83	666.67	3×96
	4	667	166.67	166.67	12×12
	5	958	333.33	166.67	12×6

Mode estimation (calibration)

1. Draw with replacement N samples from N values of E_{pred}, where N is the number of showers in the evaluation dataset for a given fixed E_{true}.

2. Perform kernel density estimation of the drawn samples with kernel bandwith determined using Scott's rule

3. Identify the position of the mode of the estimated density

4. Repeat steps 1-3 for a total of 20 times

5. Compute the mean and standard deviation of the 20 estimated values of the mode

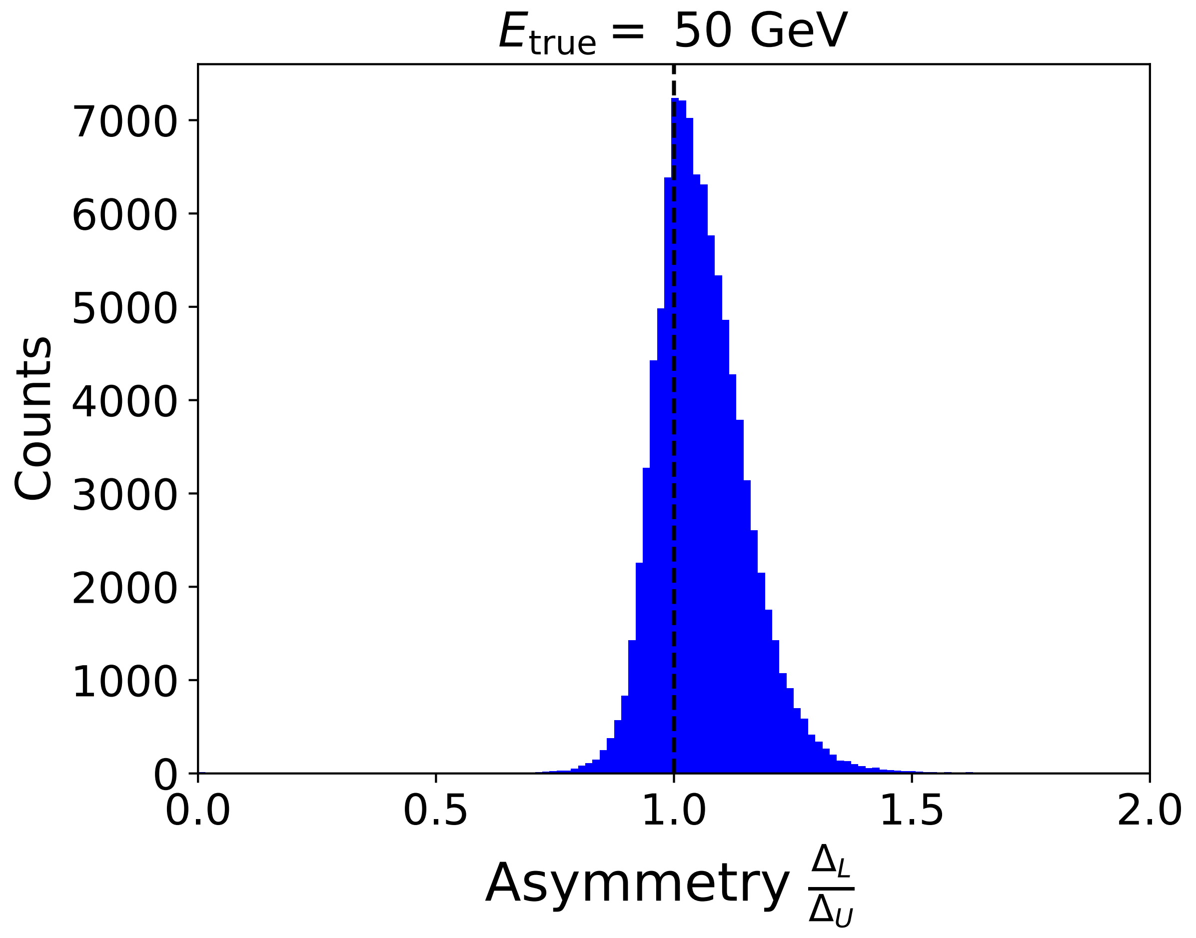
Prior dependence of MSE (calibration)

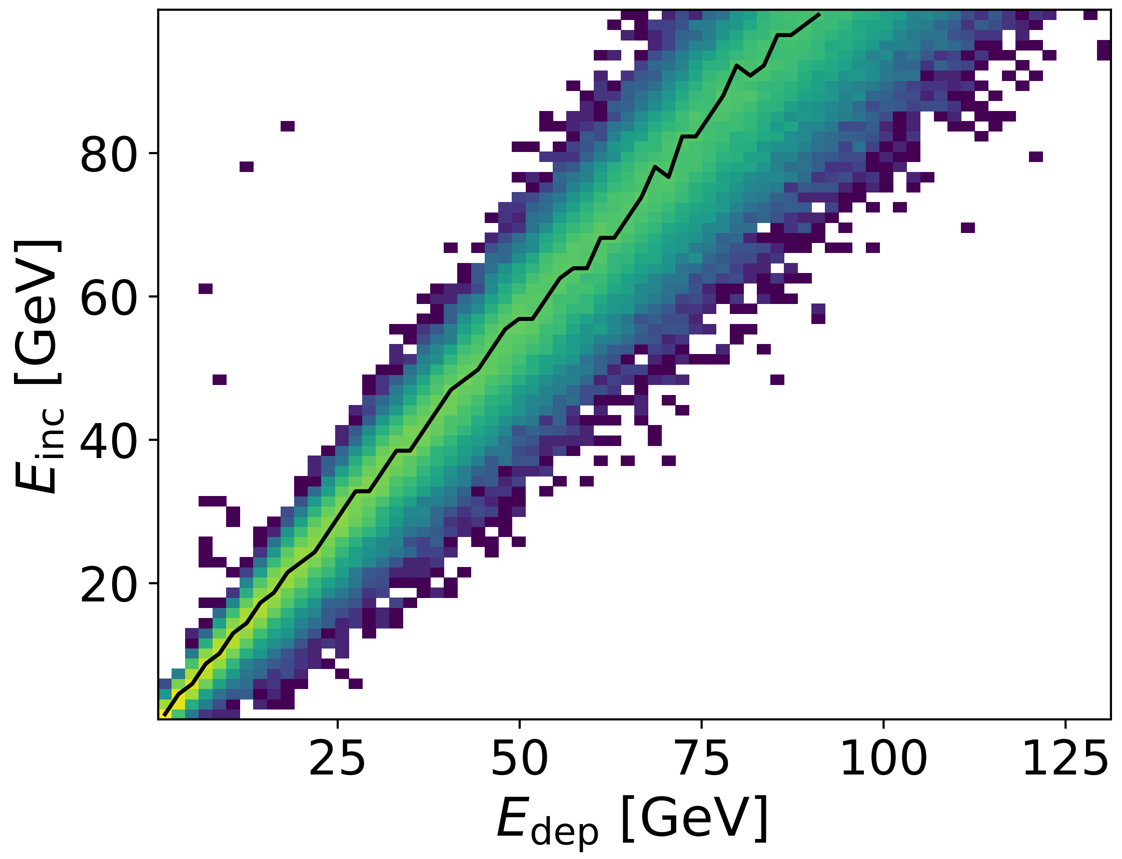
$$L[f] = \sum_i \left(f_{\mathrm{MSE}}(x_i) - z_i\right)^2 ,$$

$$\begin{aligned} f_{\mathrm{MSE}}(x) &= \langle Z \,|\, X = x \rangle \\ &= \int dz\, z\, p_{Z|X}^{\text{train}}(z \,|\, x) \\ &= \int dz\, z\, p_{X|Z}^{\text{train}}(x \,|\, z)\, \frac{p_Z^{\text{train}}(z)}{p_X^{\text{train}}(x)} \end{aligned}$$

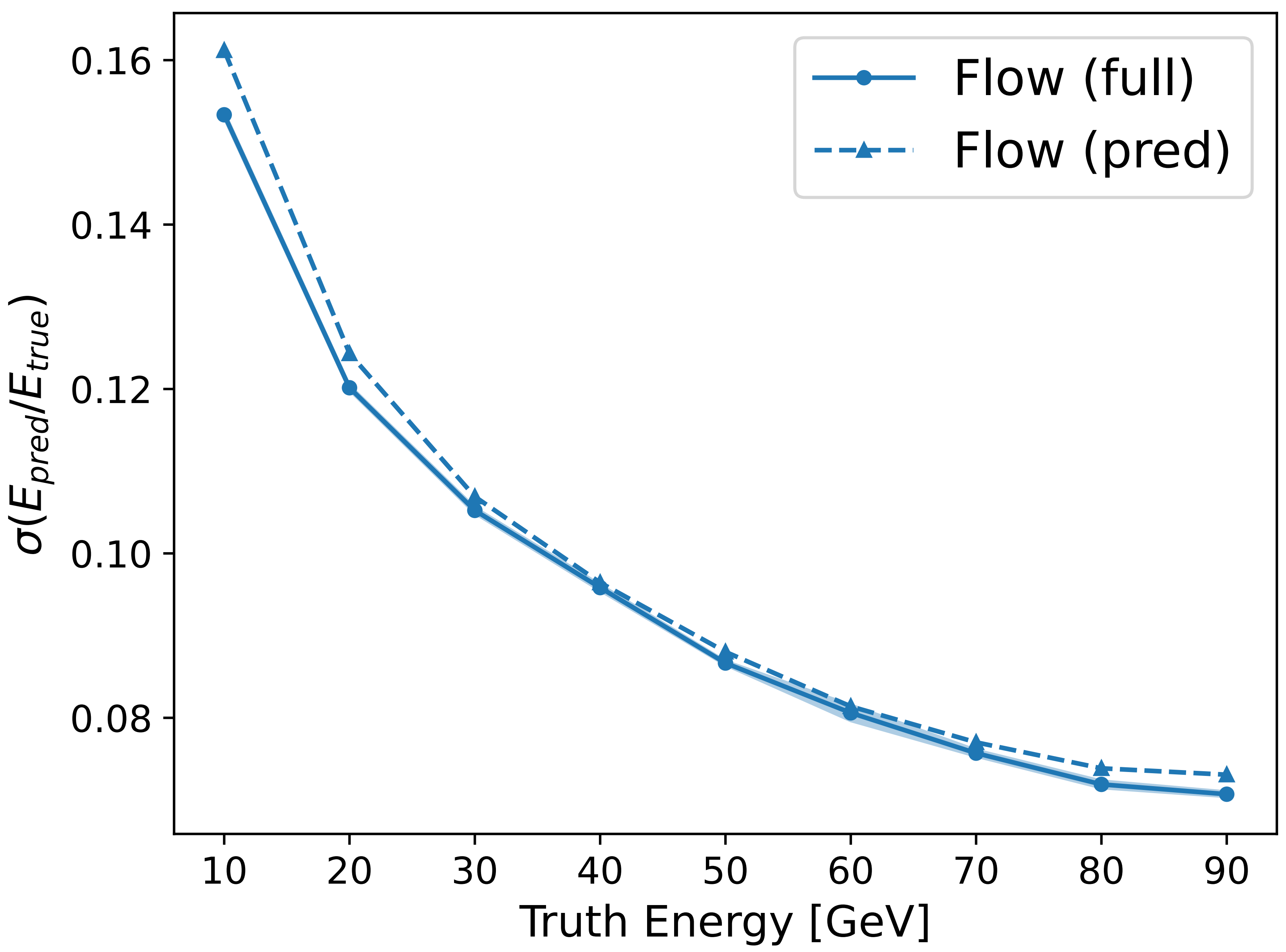
Resolution (calibration)

Able to predict asymmetric resolutions





Mean predicted per-shower resolution agrees with full resolution

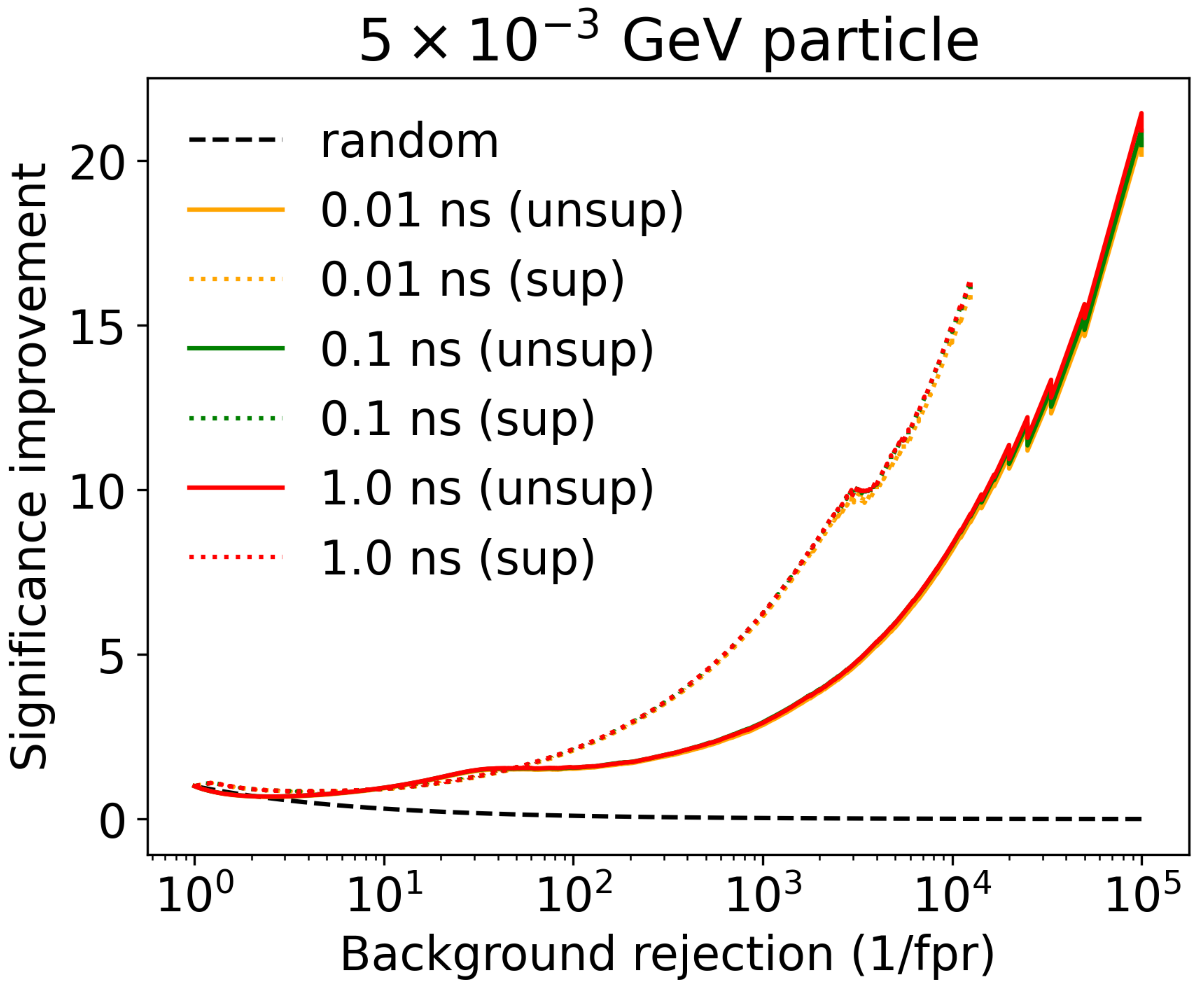


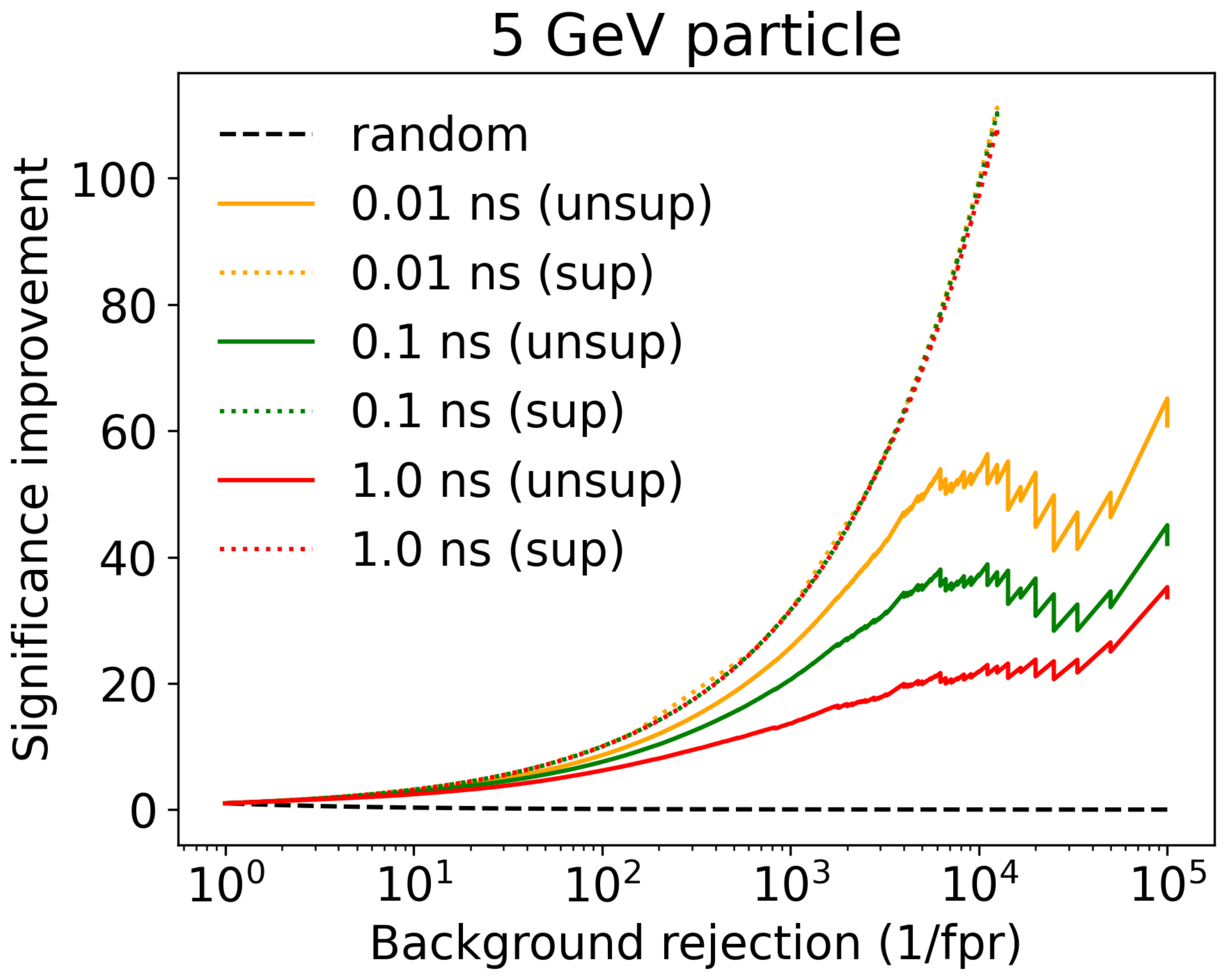
Anomaly detection

$$\text{Significance improvement} = \frac{\text{True positive rate}}{\sqrt{\text{False positive rate}}}$$

[2312.11618] C. Krause, B. Nachman, **IP**, D. Shih

Trained on $m_\chi = 5$ GeV, lifetime = 1 ns :





Anomaly detection

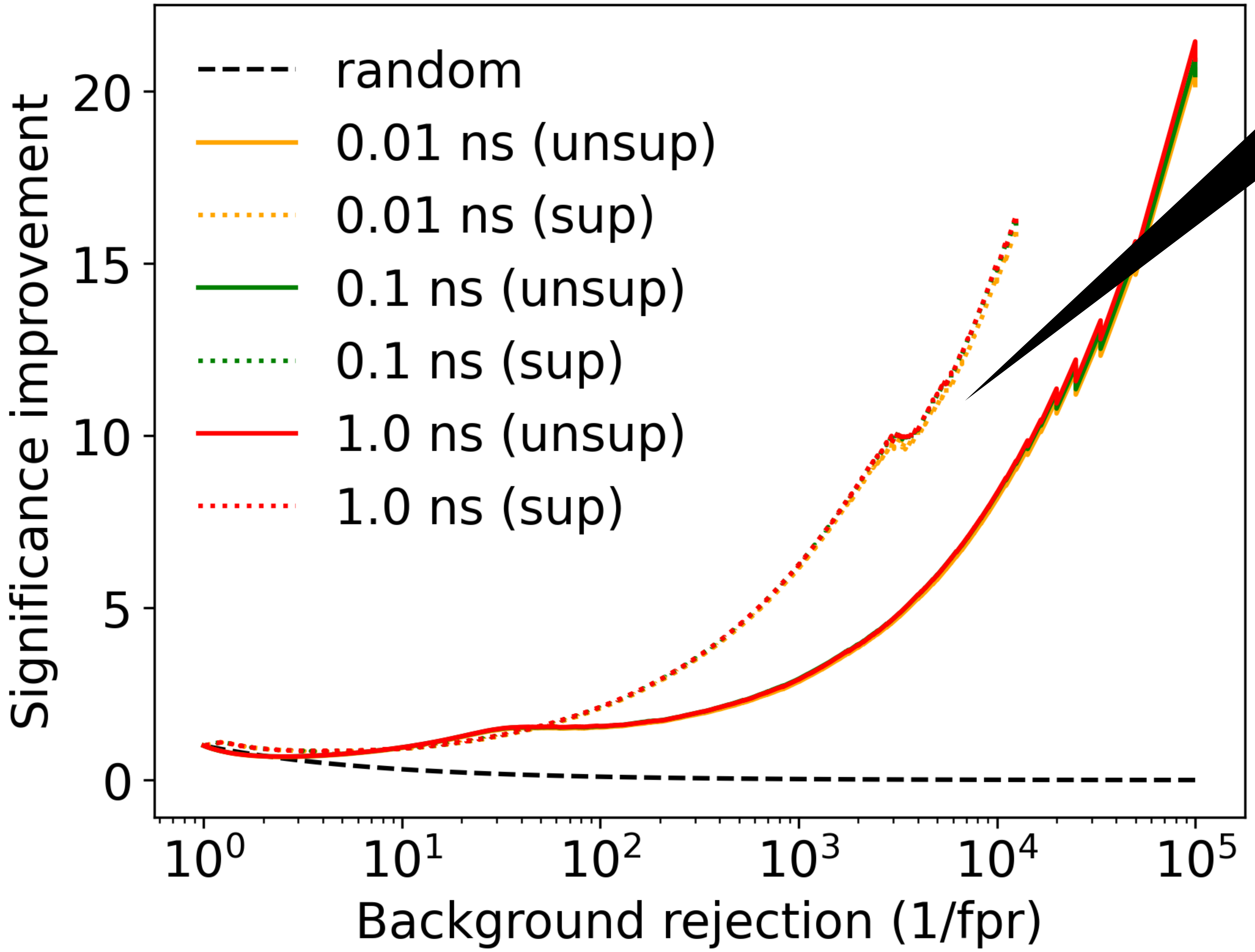
Significance improvement = $\frac{\text{True positive rate}}{\sqrt{\text{False positive rate}}}$

[2312.11618] C. Krause, B. Nachman, **IP**, D. Shih

Trained on $m_\chi = 5$ GeV, lifetime = 1 ns :

Supervised performance transferable in some cases

5×10^{-3} GeV particle



5 GeV particle