Efficiency Parameterisation with Neural Network

4th IML Workshop
21 October, 2020

מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE

SAPIENZA
UNIVERSITÀ DI ROMA

INFN

C. Badiali, F. A. Di Bello, G. Frattari, E. Gross, V. Ippolito, E. Loiacono, I. Luise, M. Kado, N. Kakati, J. Shlomi

Efficiency Parameterization with Neural Networks

Francesco Armando Di Bello[1,2] · Jonathan Shlomi[3] · Chiara Badiali[1,2] · Guglielmo Frattari[1,2] · Eilam Gross[3] · Valerio Ippolito[2] · Marumi Kado[1,2, 4]

the date of receipt and acceptance should be inserted later

Abstract Multidimensional efficiency maps are commonly used in high energy physics experiments to mitigate the limitations in the generation of large samples of simulated events. Binned efficiency maps are however strongly limited by statistics. We propose a neural network approach to learn ratios of local densities to estimate in an optimal fashion efficiencies as a function of a set of parameters. Graph neural network techniques processes in very restricted regions of phase space. To mitigate this difficulty, a commonly used approach is the *event weighting technique* which replace selection cuts with event weights. These weights are typically defined from binned efficiency maps. The difficulty in these methods is the range of applicability of efficiency maps that are limited in the number of dimensions (typically two), and subsequently, fail to capture more sub-

https://arxiv.org/abs/2004.02665v2

Overview

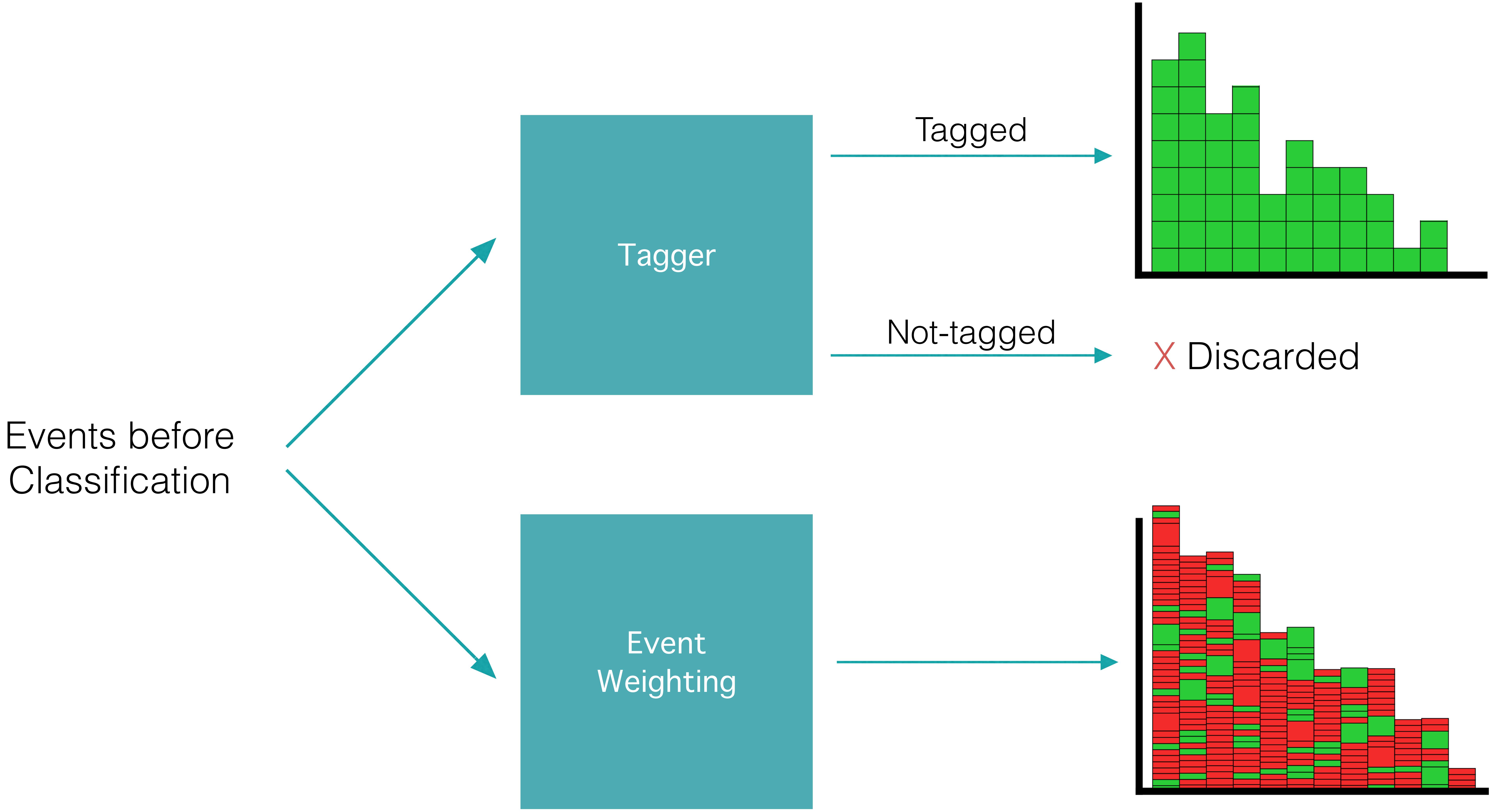
- Efficiency Parameterisation
- The binned approach and its limitations
- The NN based approach
- Case study
- Summary

Event Weighting

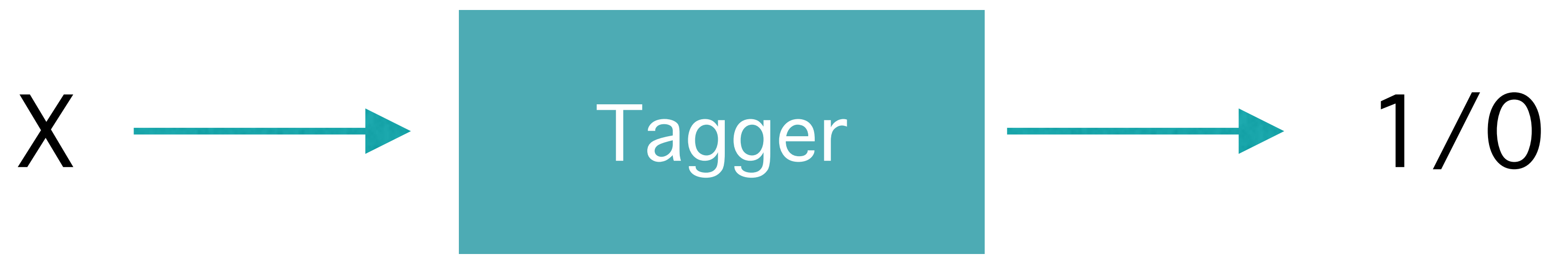
- We often care about events in very restricted regions of phase space
- Trivial solution: Apply a selection cut

- Issue: Low statistics
- Alternate solution: Event Weighting Technique

Event Weighting



Efficiency Parameterisation

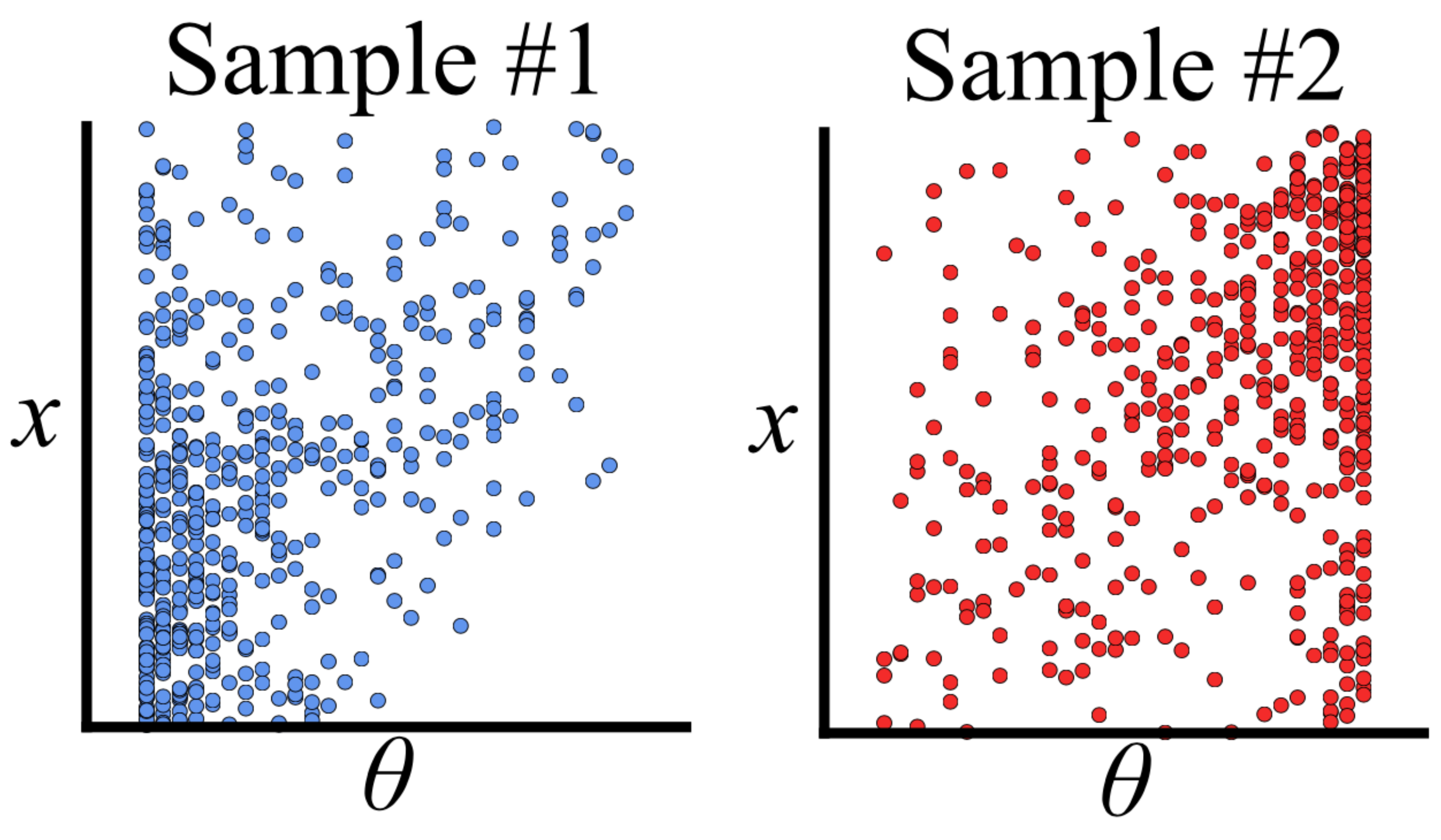


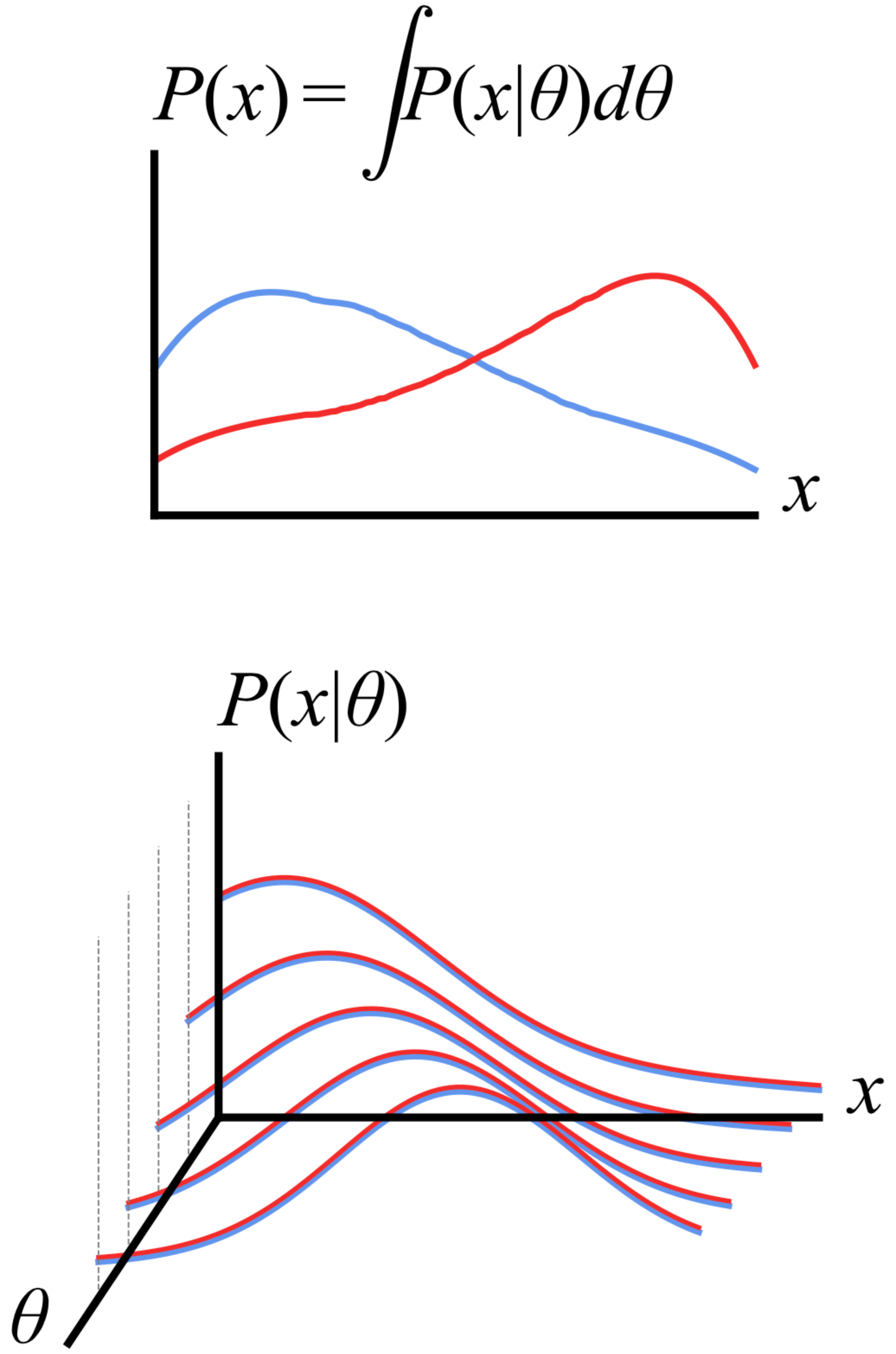
- ‘X’ depends on a set of variables ‘θ’
- Goal 1: We want to know the efficiency, $\epsilon(\theta)$ of the classifier

$$\epsilon(\theta) = \frac{N(tagged\,|\,\theta)}{N(\theta)}$$

- Goal 2: Define θ so that $p(X\,|\,\theta)$ is identical b/w two different samples. (Generalisation)
- We can parameterise the efficiency of the classifier independent of the sample.

Generalisation





What constitutes θ ?

- Typically, ‘θ’ is unknown

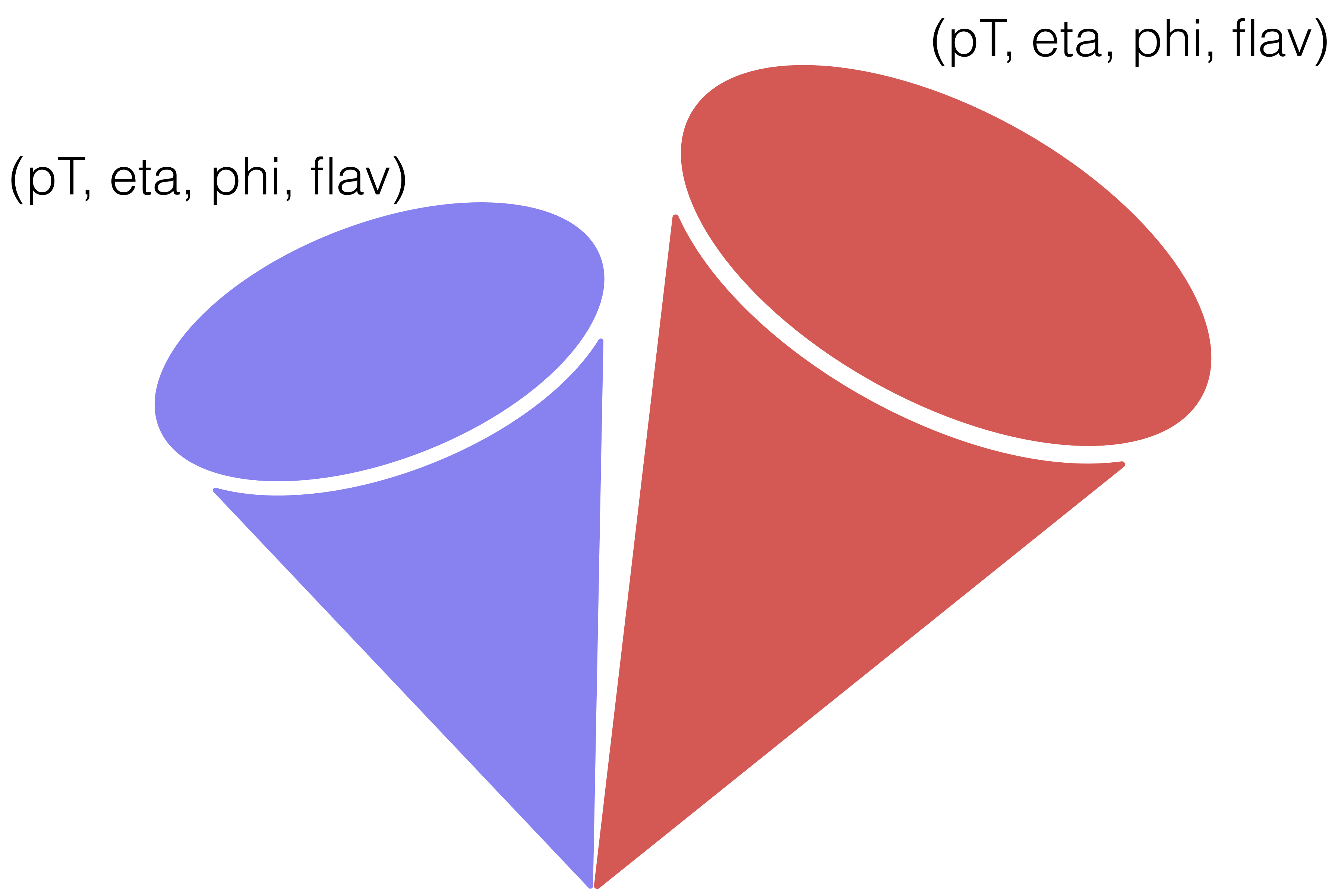
- For b-tagging in ATLAS (input to tagger: ‘a jet’, output : ’is it a b-jet’),

  p_T and η are the most dominant components of θ.

- Common Practice: binned efficiency maps (p_T, η)

- Fails to capture the complete picture

Limitations of the binned maps



- Issues -
  - θ is not known fully
  - Dimension of θ is large
  - Dimension of θ is not constant (influence of neighbouring jet)
- Histogram based maps cannot capture the full dependencies of the classifier efficiency

- We propose an NN based approach to address these issues

The NN approach (background)

- Density Ration Estimation

- Let's assume two distributions - $p(\theta)$ and $q(\theta)$
  - $p(\theta)$ - Distribution of jets that passed the tagger
  - $q(\theta)$ - Distribution of jets that did not pass the tagger

- Since θ is the independent variable here, (not X) the two distributions are not separable

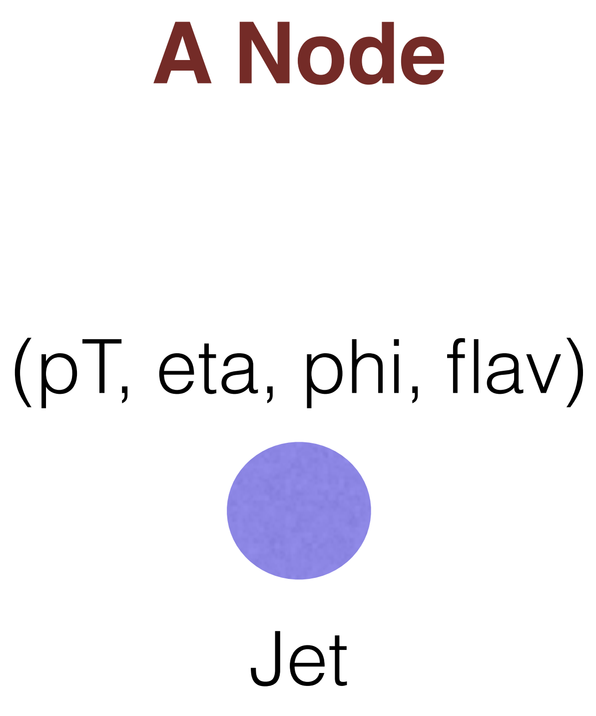
- If we train a binary classifier $g(\theta)$, it converges to -
  - $g(\theta) \approx \dfrac{p(\theta)}{p(\theta) + q(\theta)} = \epsilon(\theta)$ = efficiency

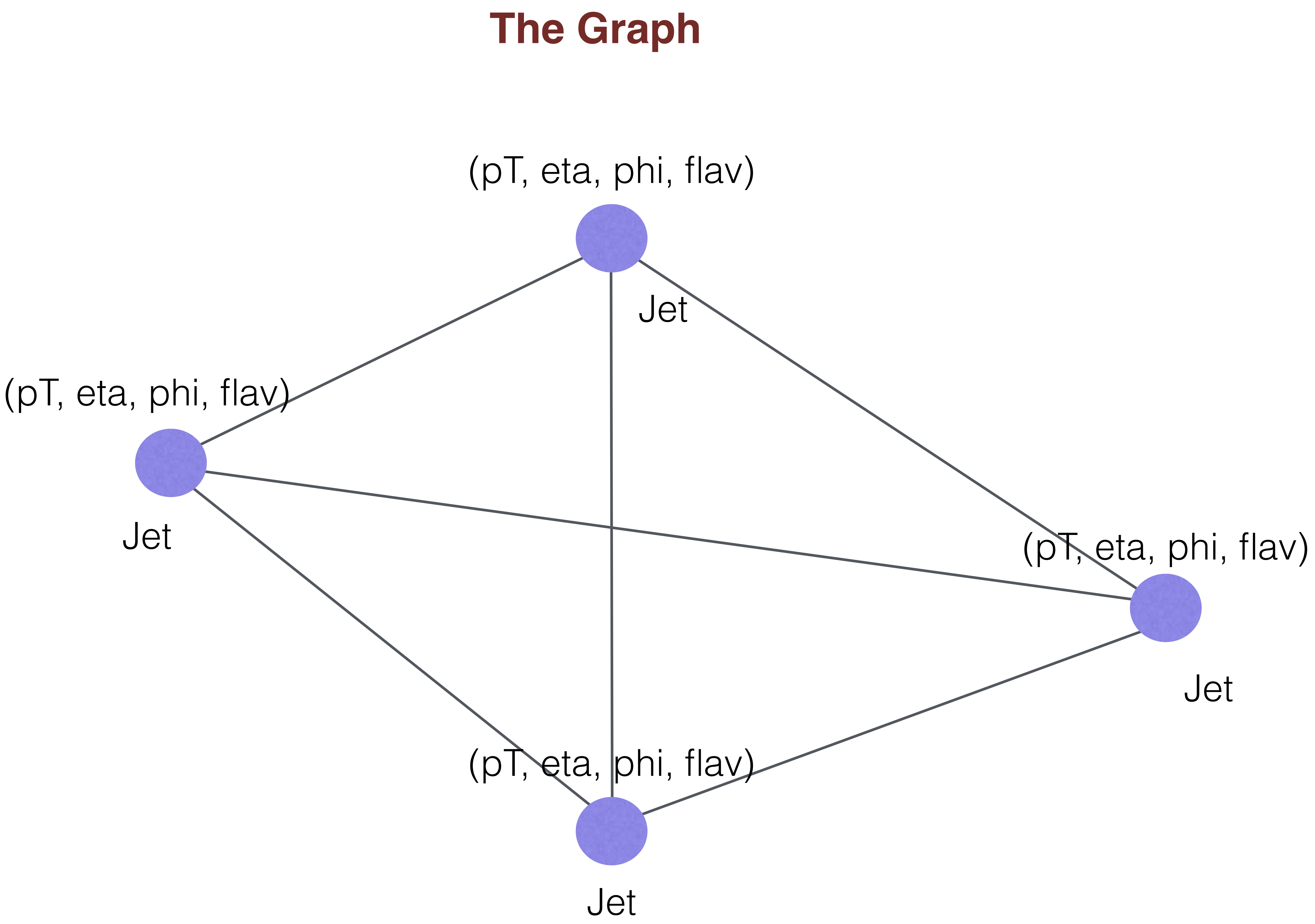
The NN approach

- “ θ is not fully known “
- We want the network to -
  - Infer θ during training
  - Consider the jet-jet dependency
- Need to learn jet-jet dependency + num jets in an event is not fixed

-> GNN

Event Representation

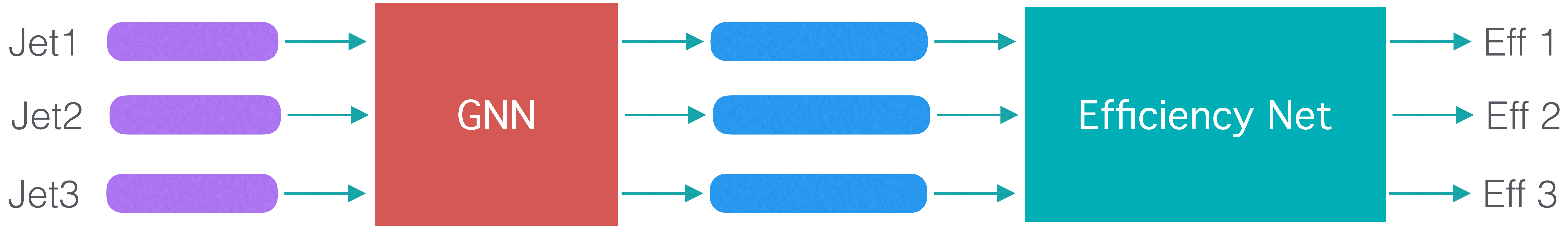




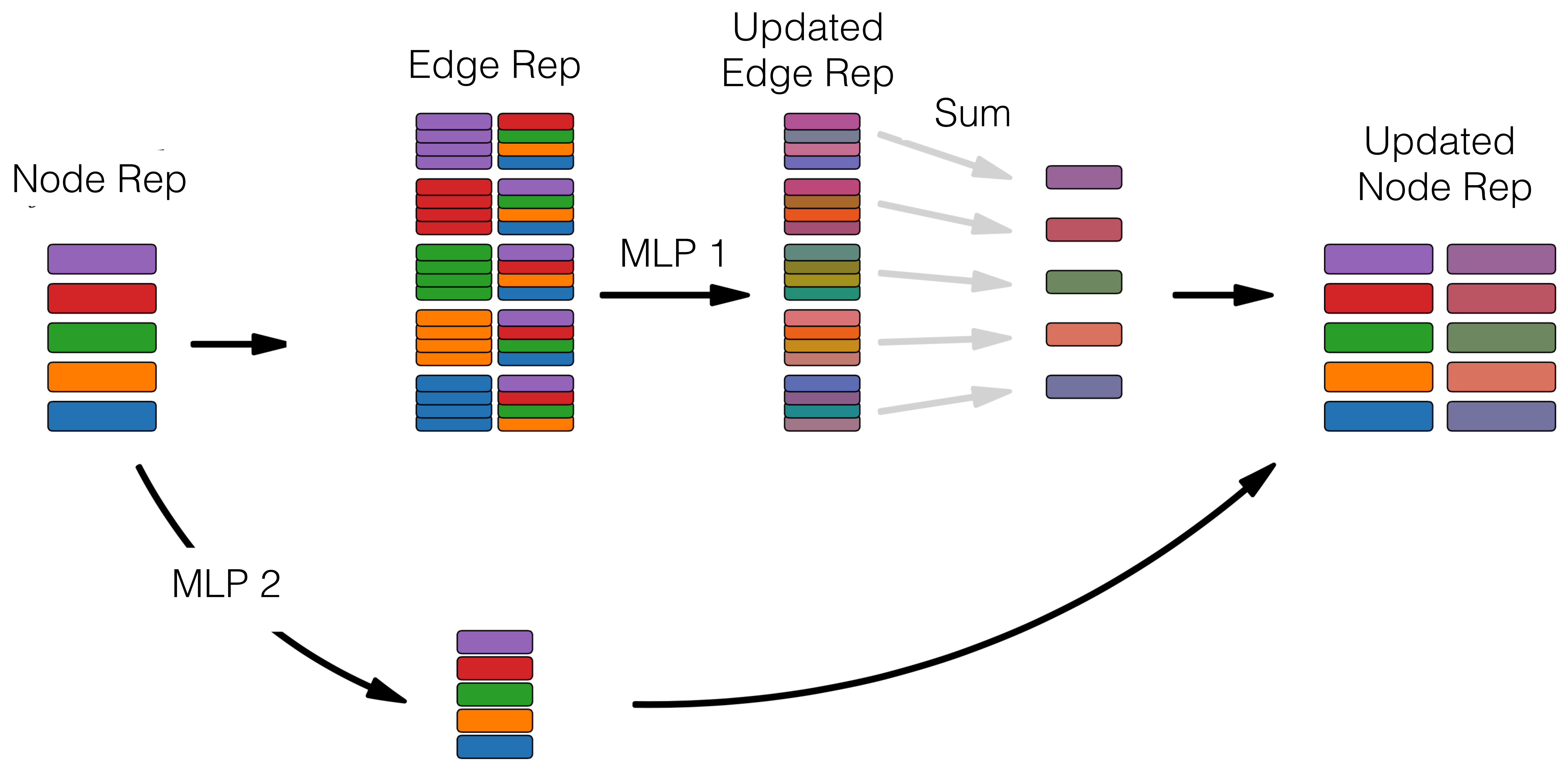
- The network can learn dR (a part of θ) during training

The NN architecture

- Construct a high dimensional, neighbourhood aware representation of the jets
- Pass the jets through a binary classifier



A GNN block

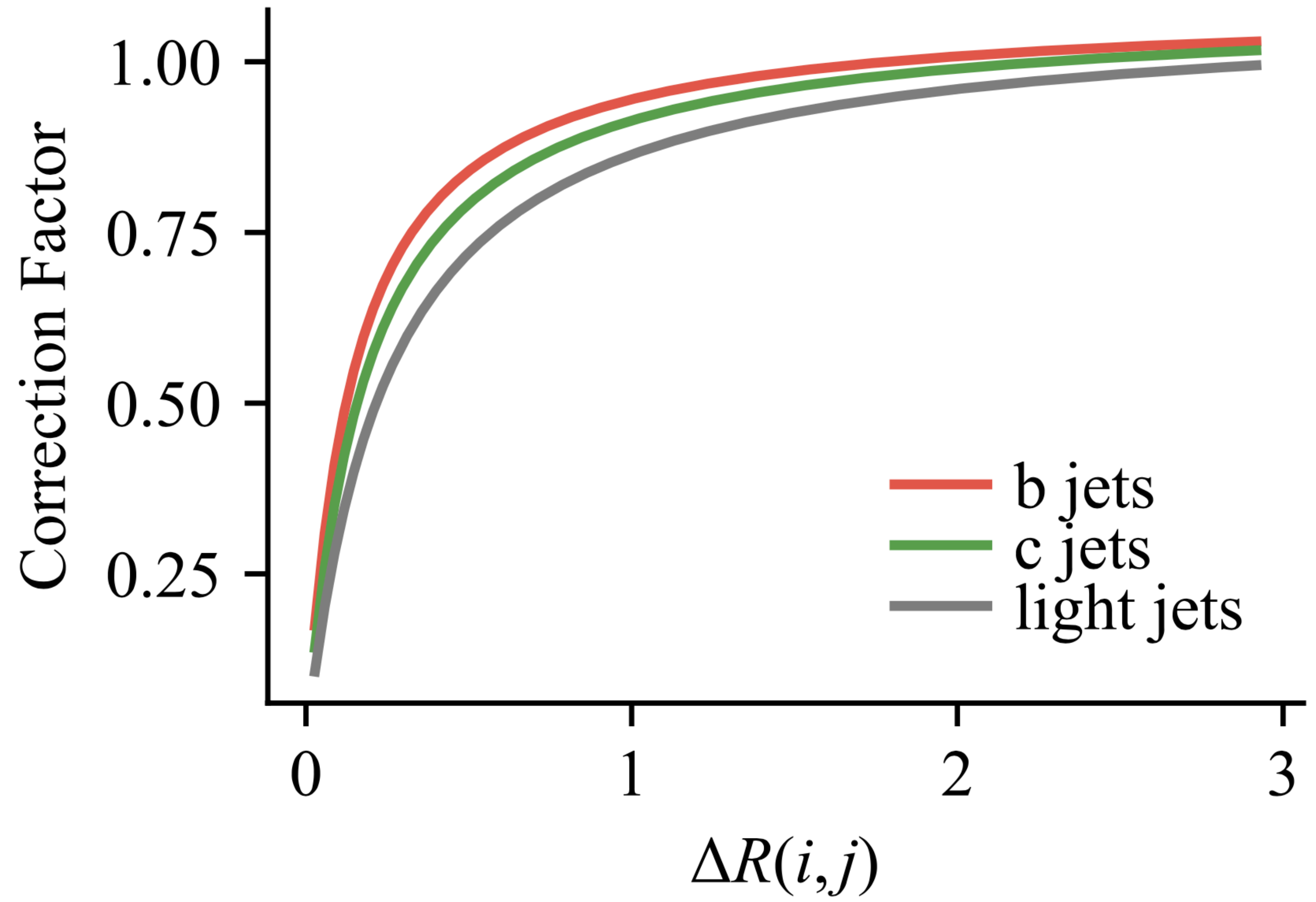


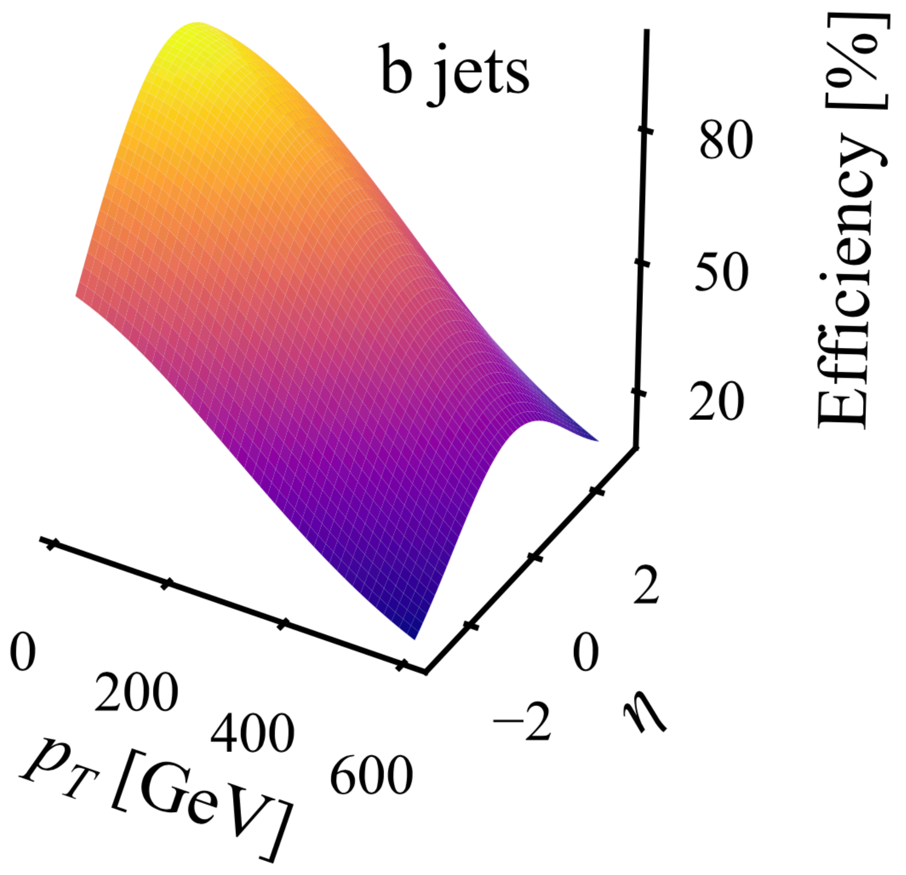
- Efficiency of a jet is conditioned on the near-by jets

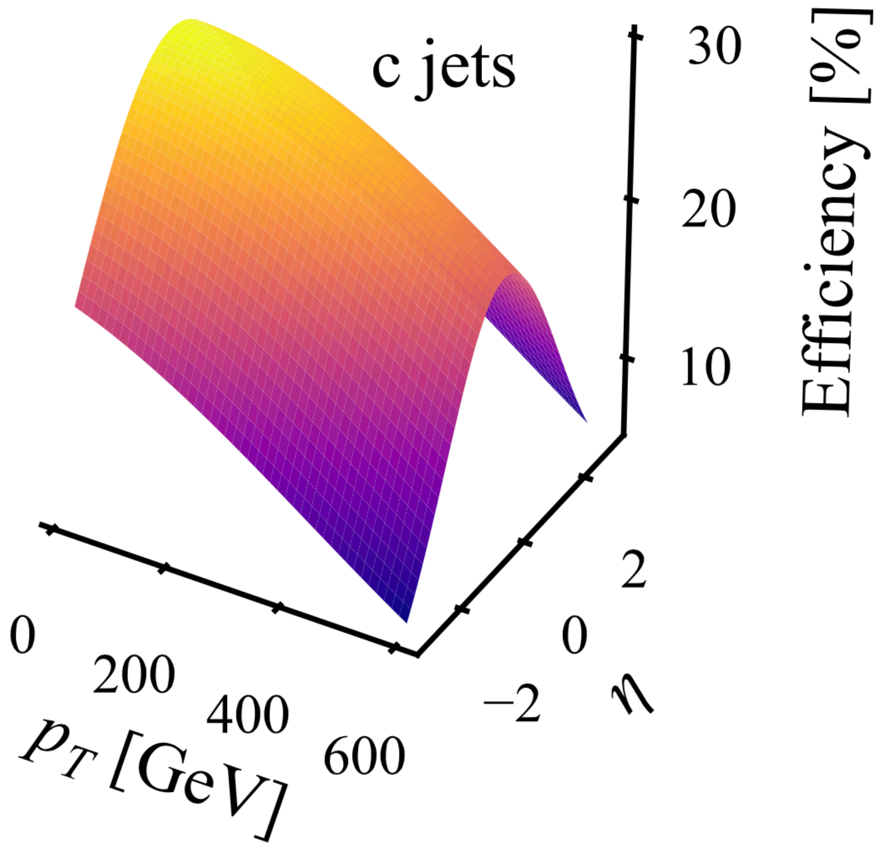
Case Study (Toy model)

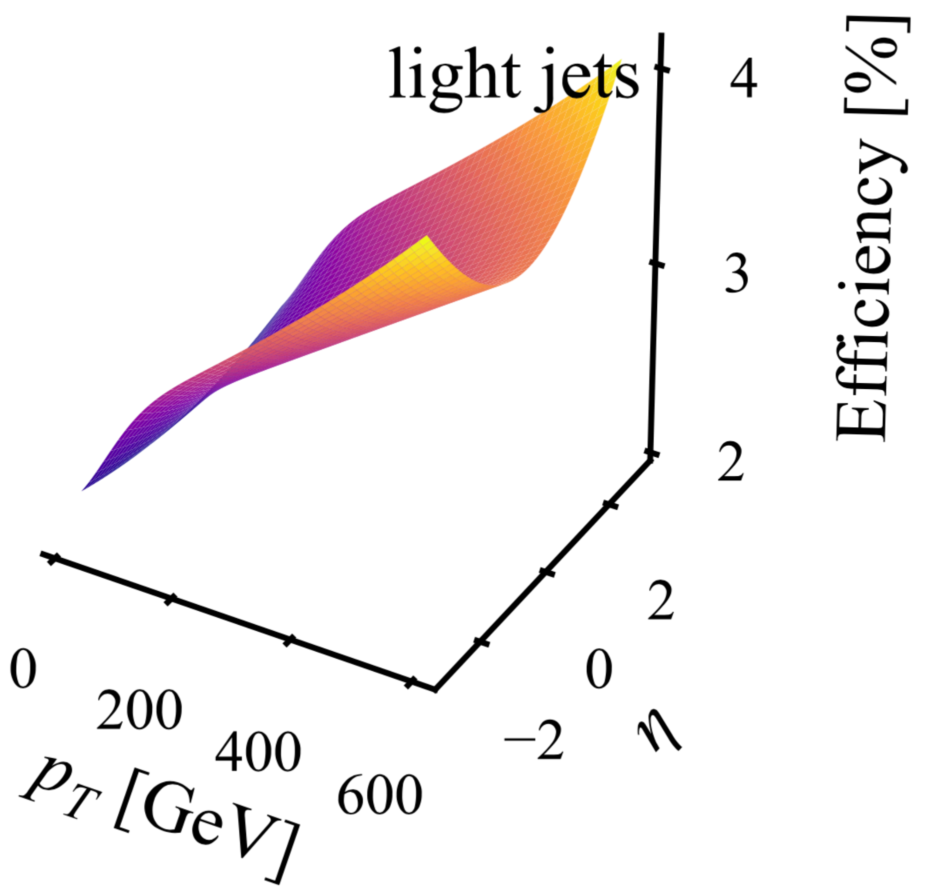
- pT, eta, phi distribution sampled using functions
- True efficiency is given by -

$$\epsilon_{jet_i} = \epsilon_{f_i}(p_T, \eta). \prod_j \hat{\epsilon}_{ij}\left(\Delta R_{ij}, f_j\right)$$



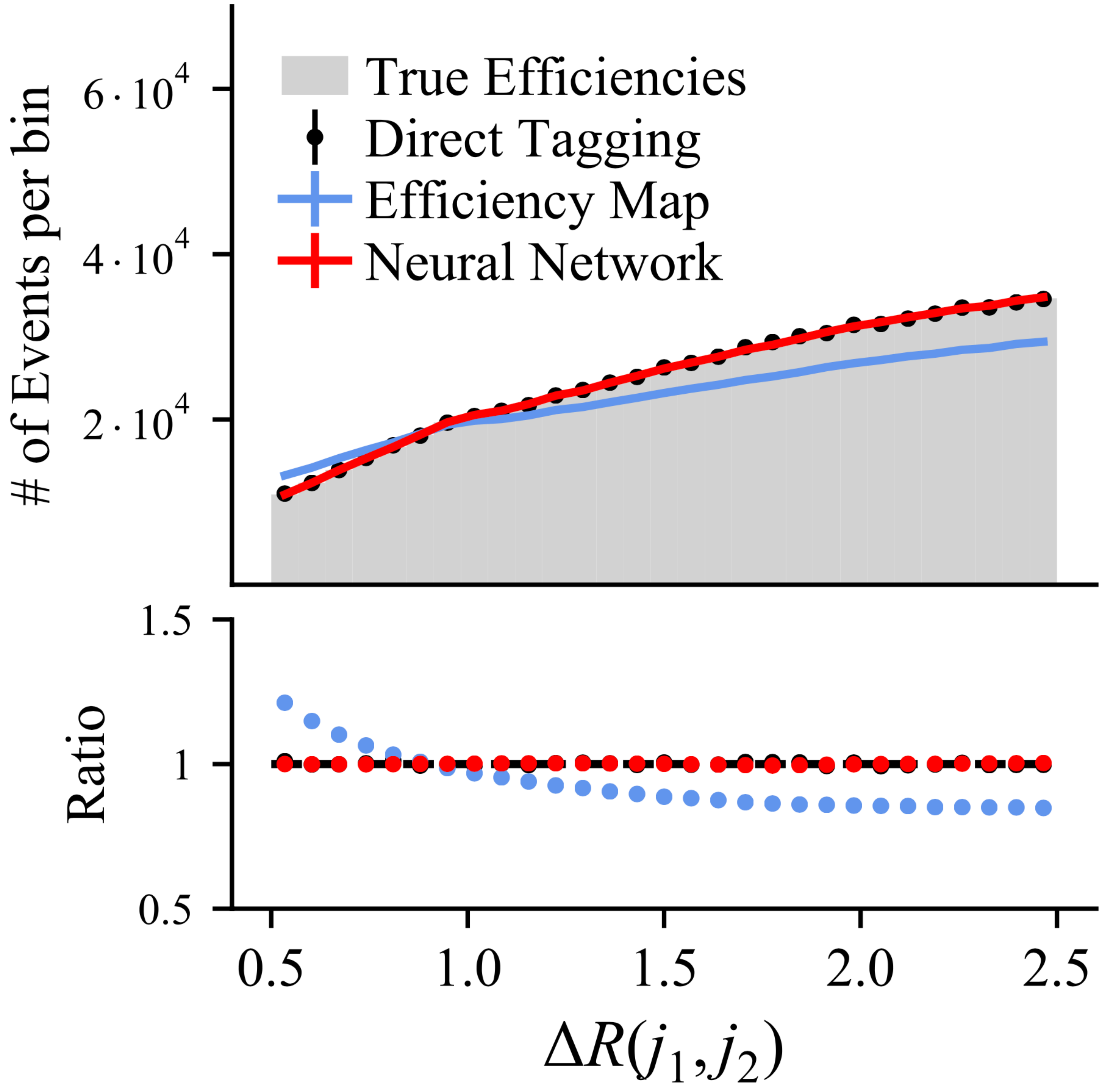


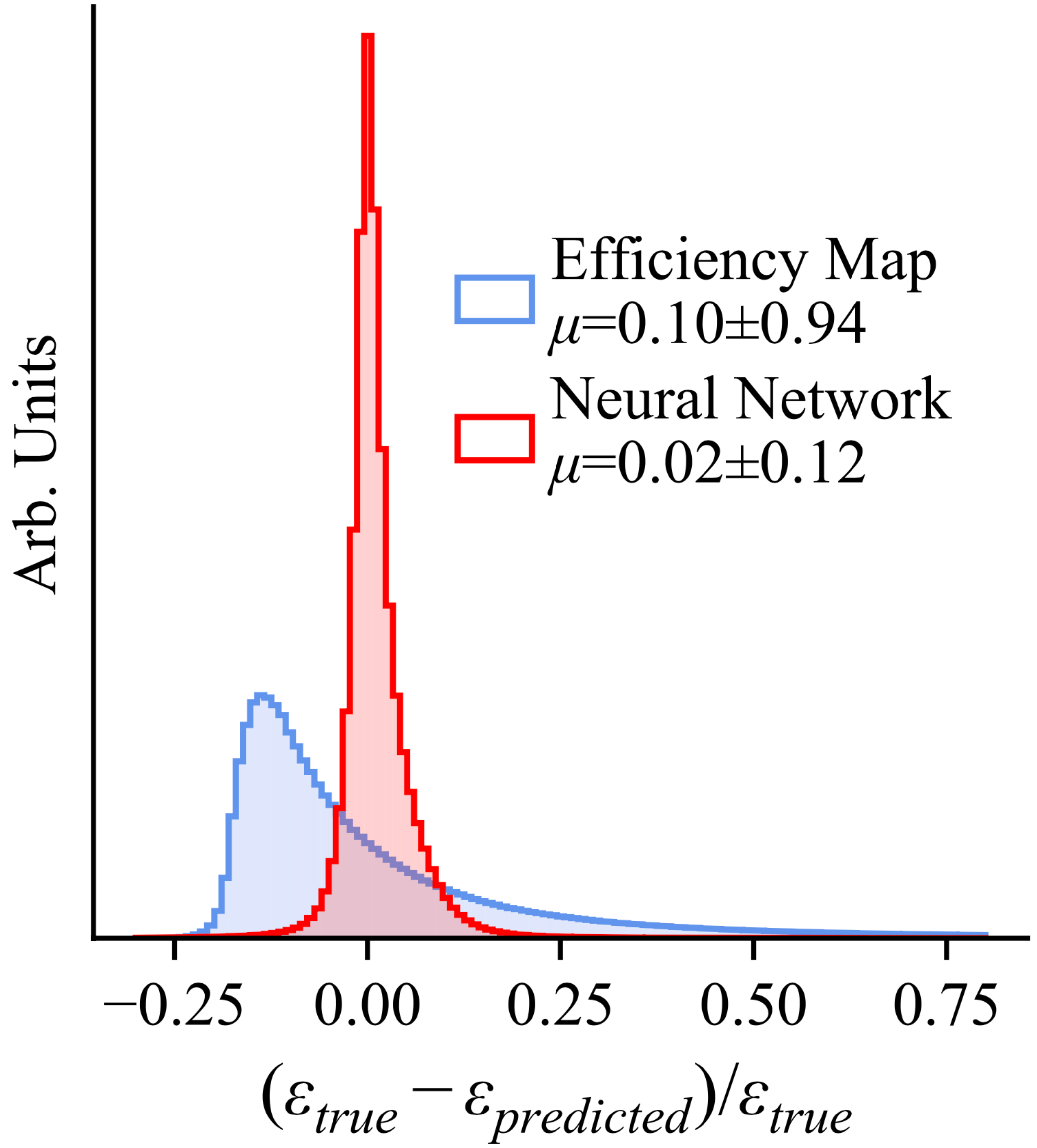




Result I

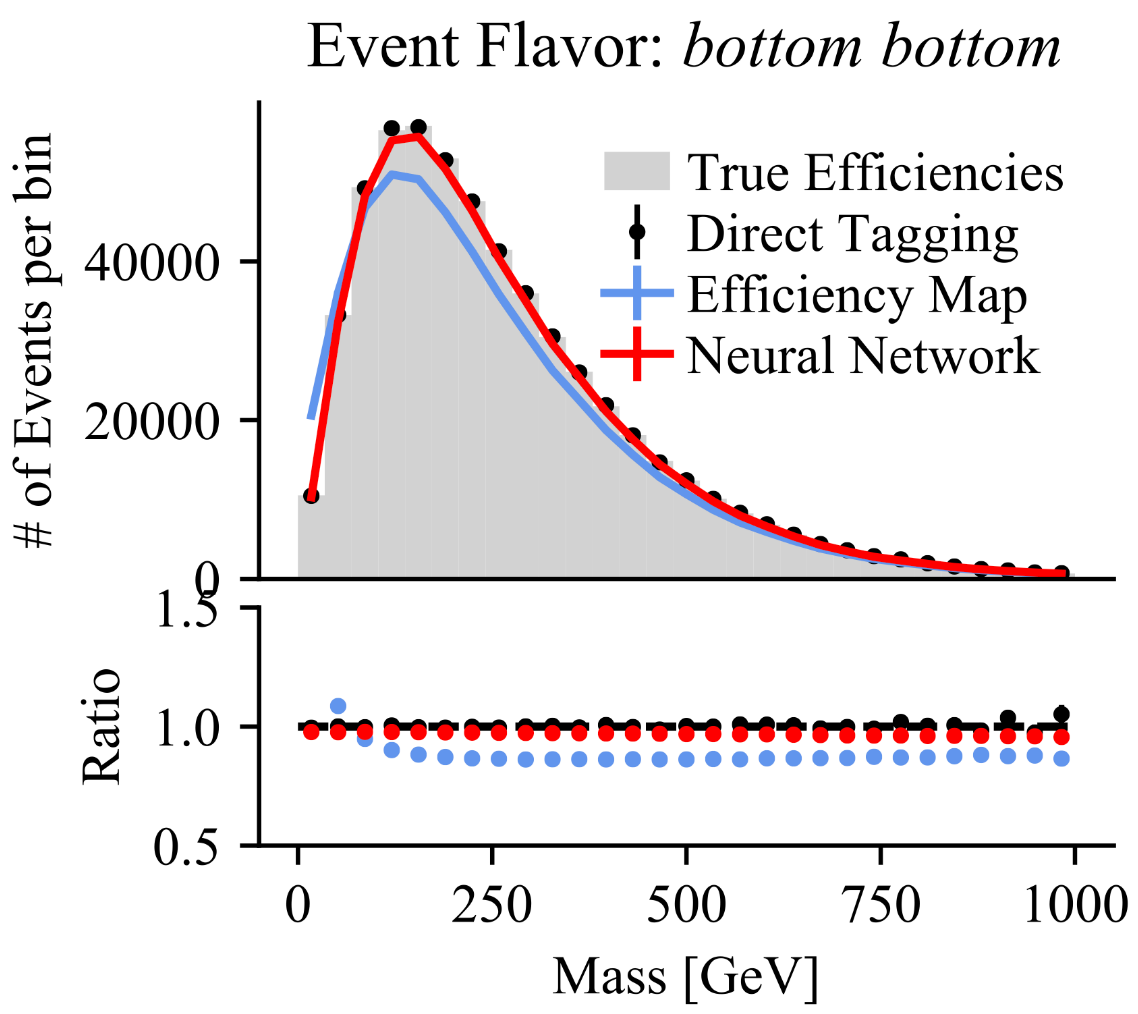
- 1 tag region, leading jet is b-tagged

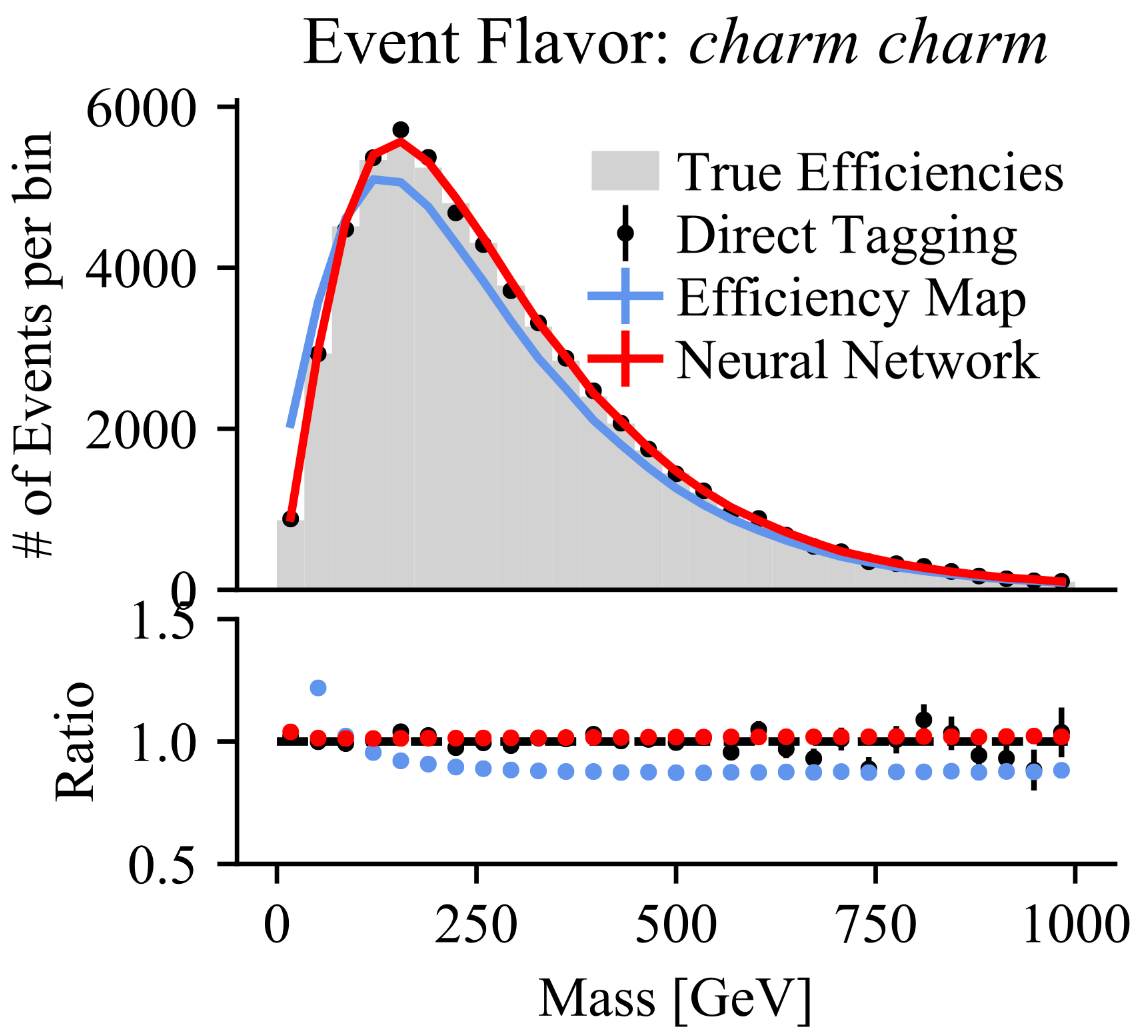


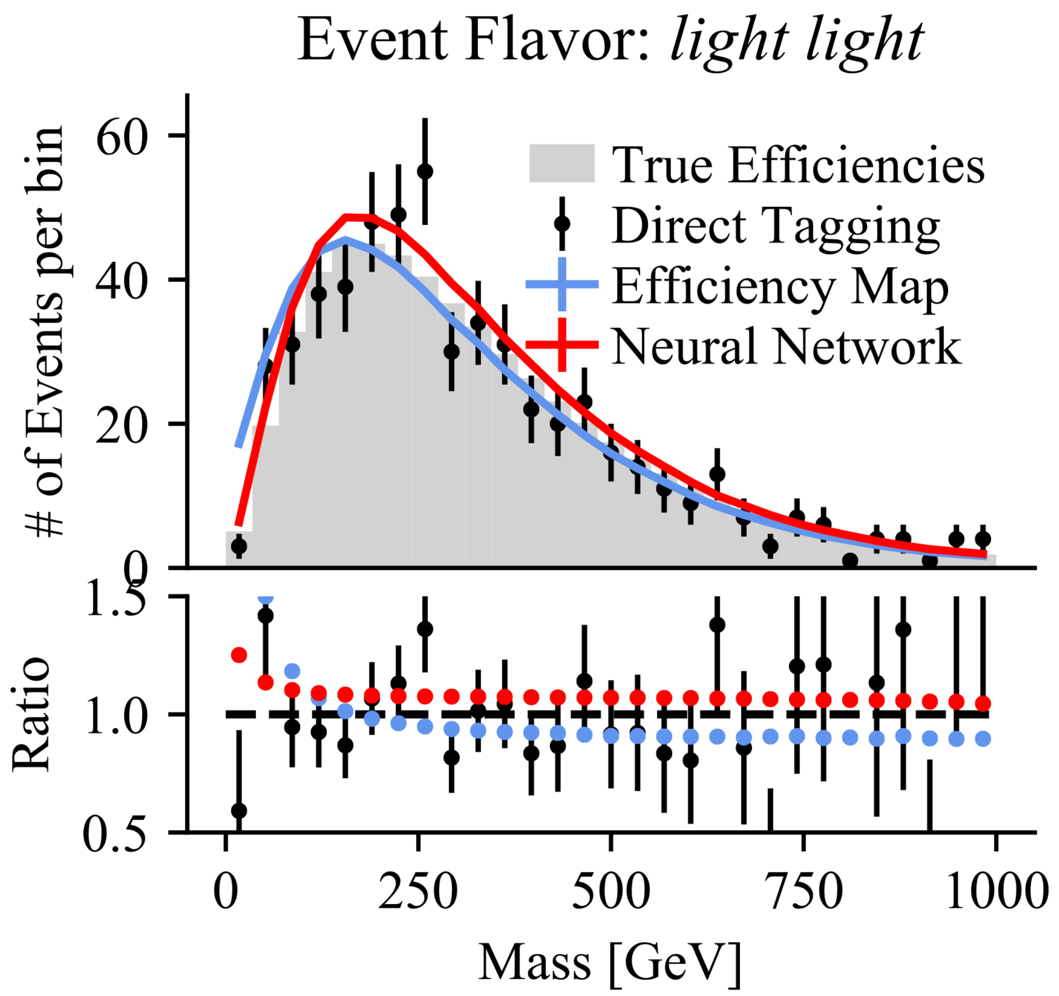


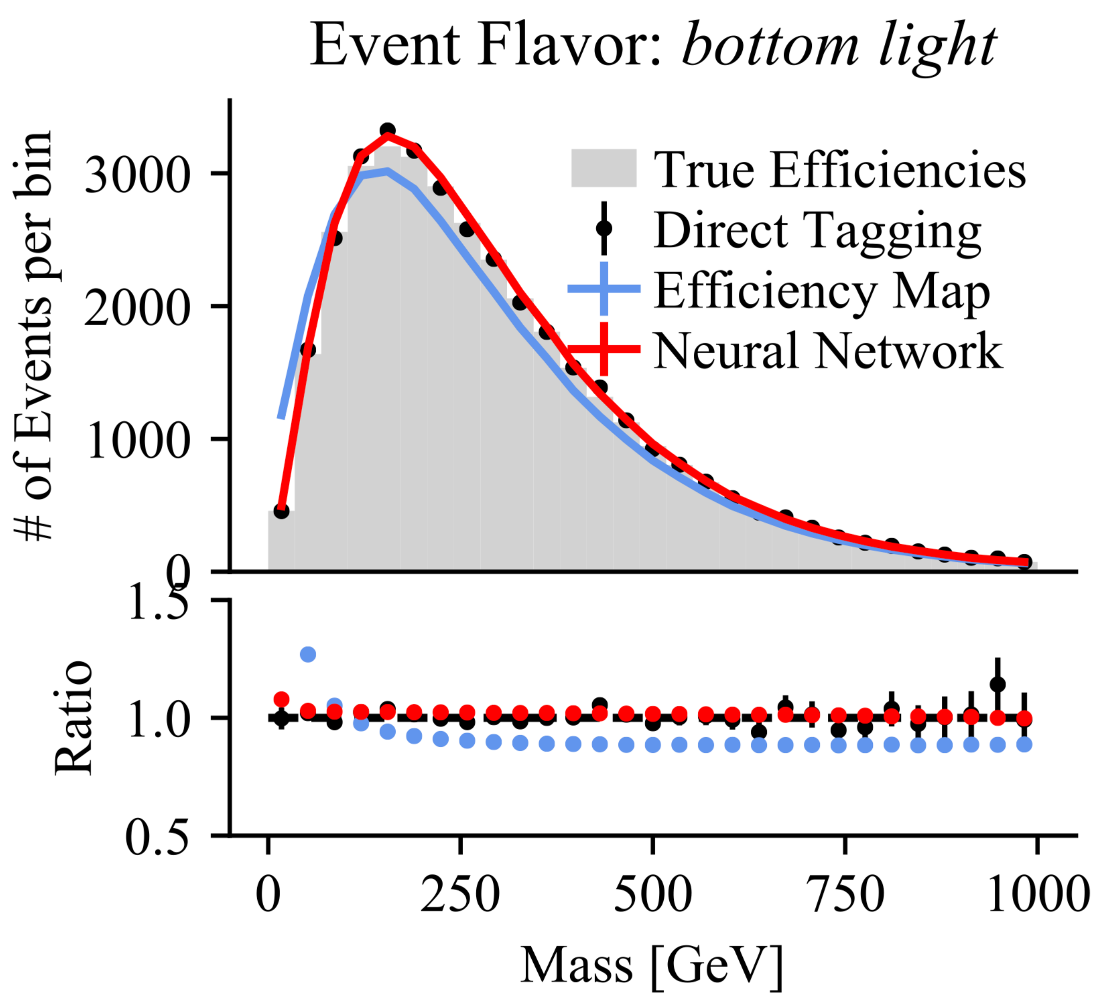
Result II

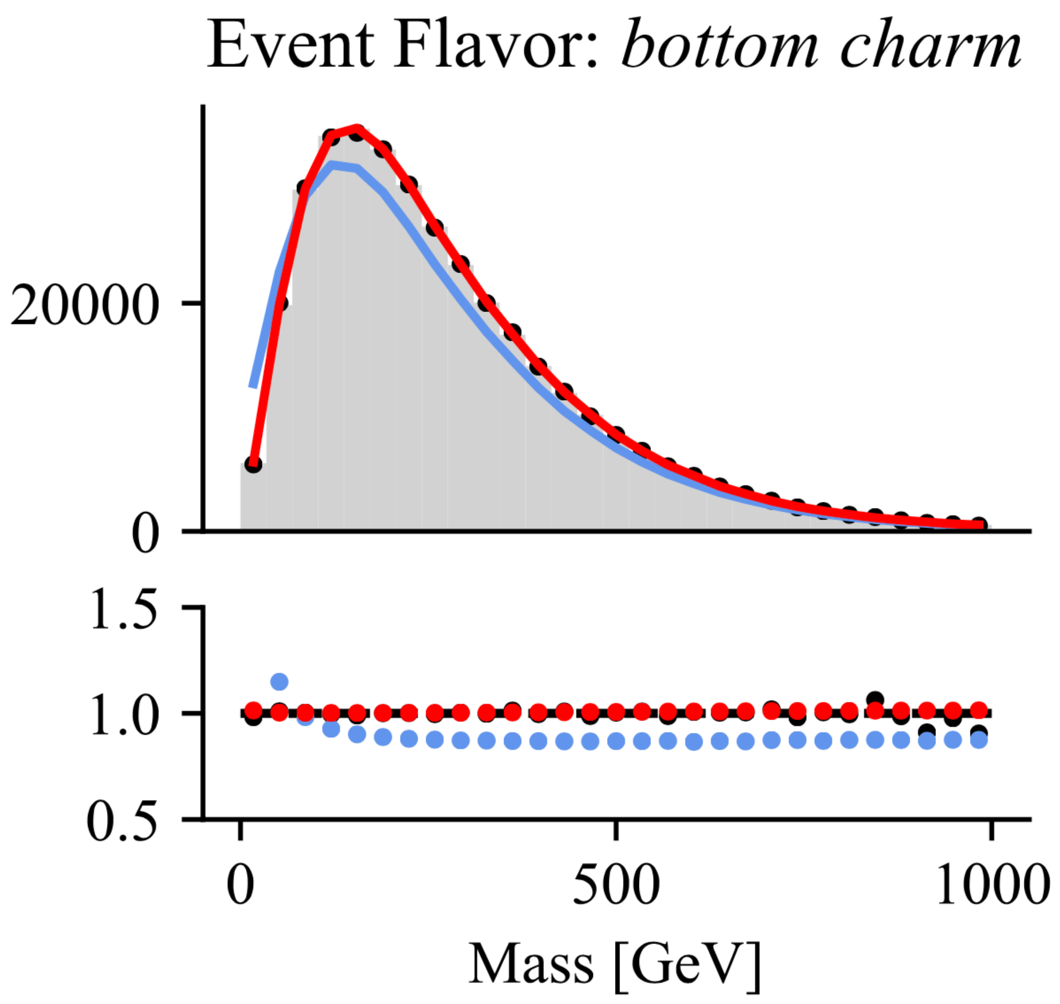
- 2 tag region, leading and sub-leading jet are b-tagged. Event efficiency = $\epsilon_1 \epsilon_2$

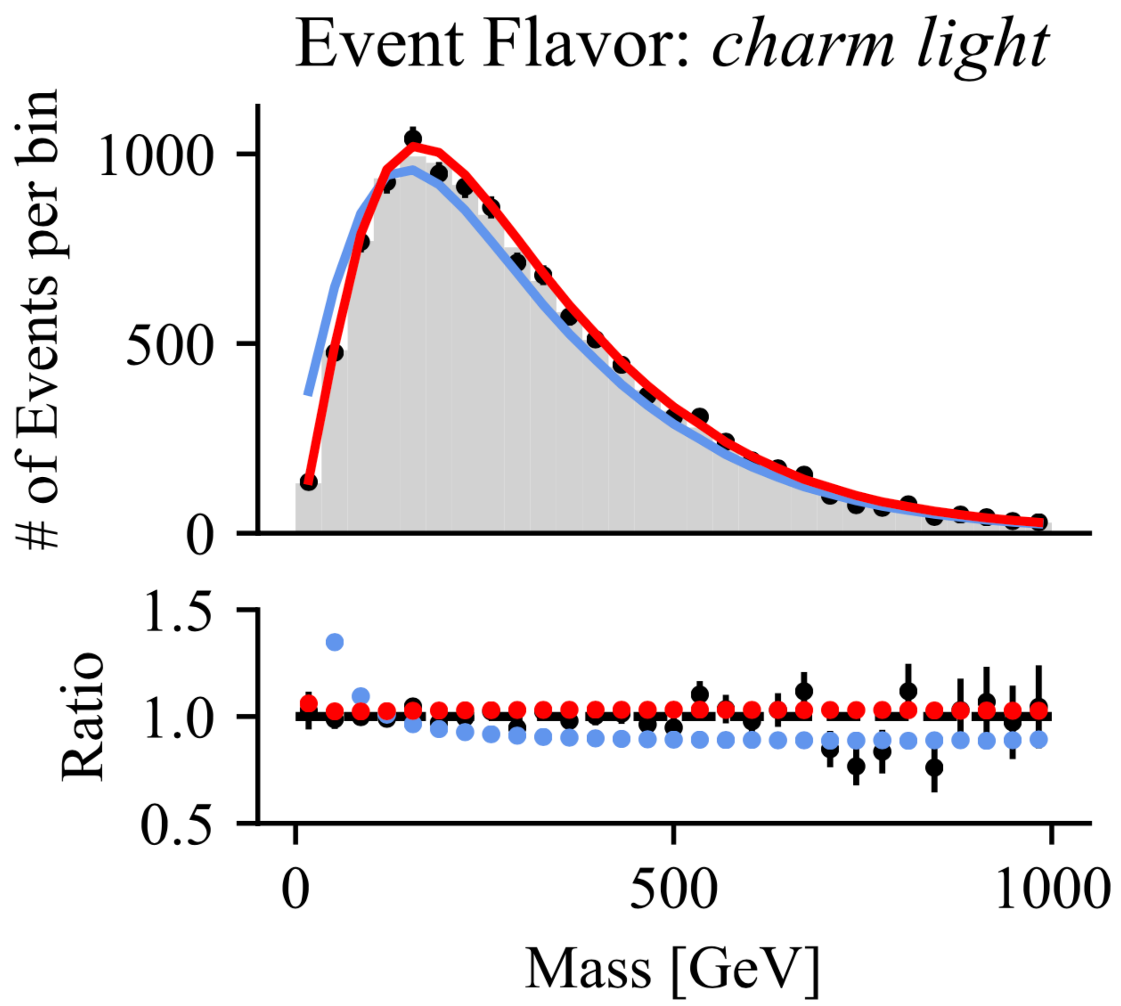






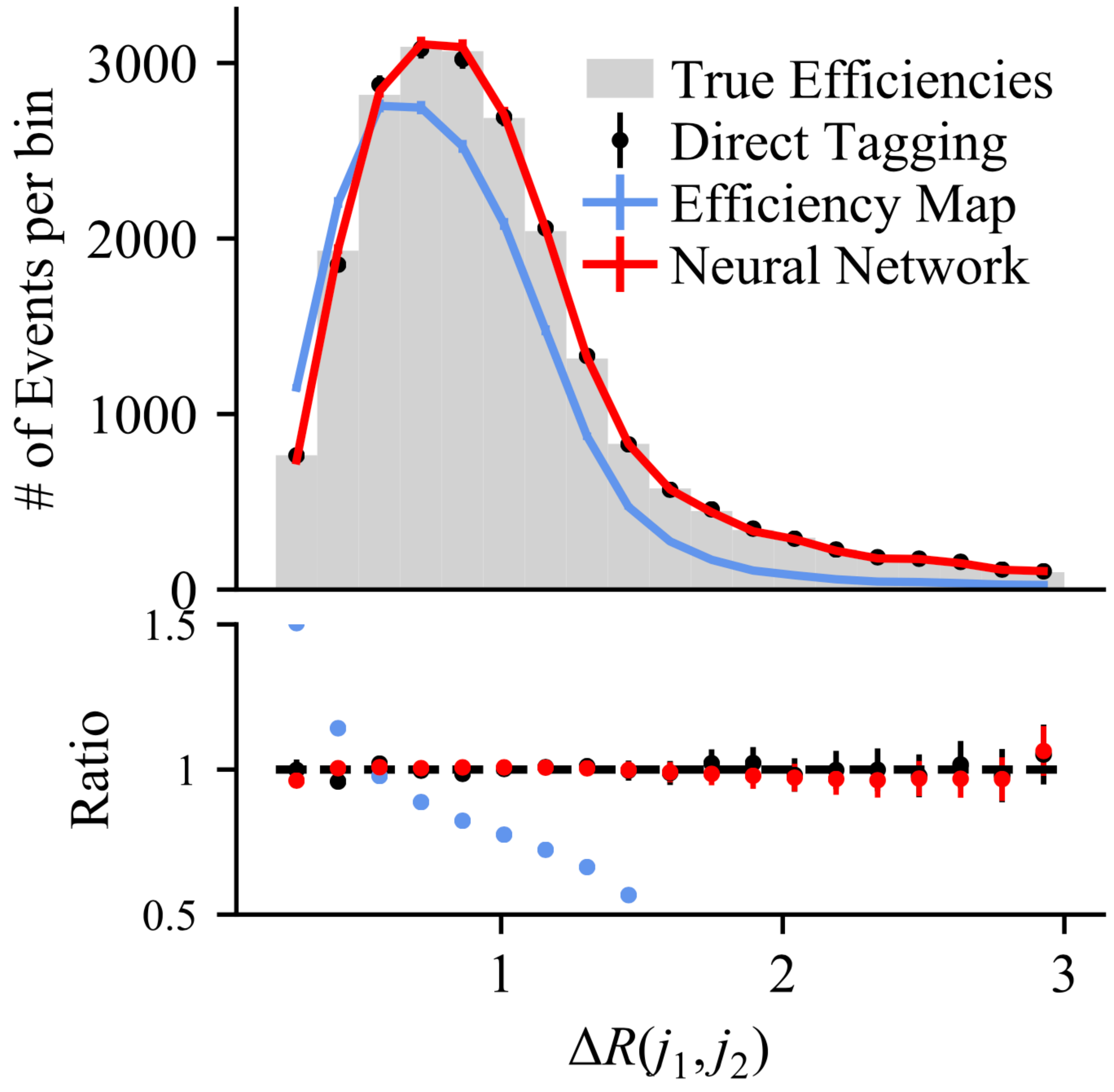


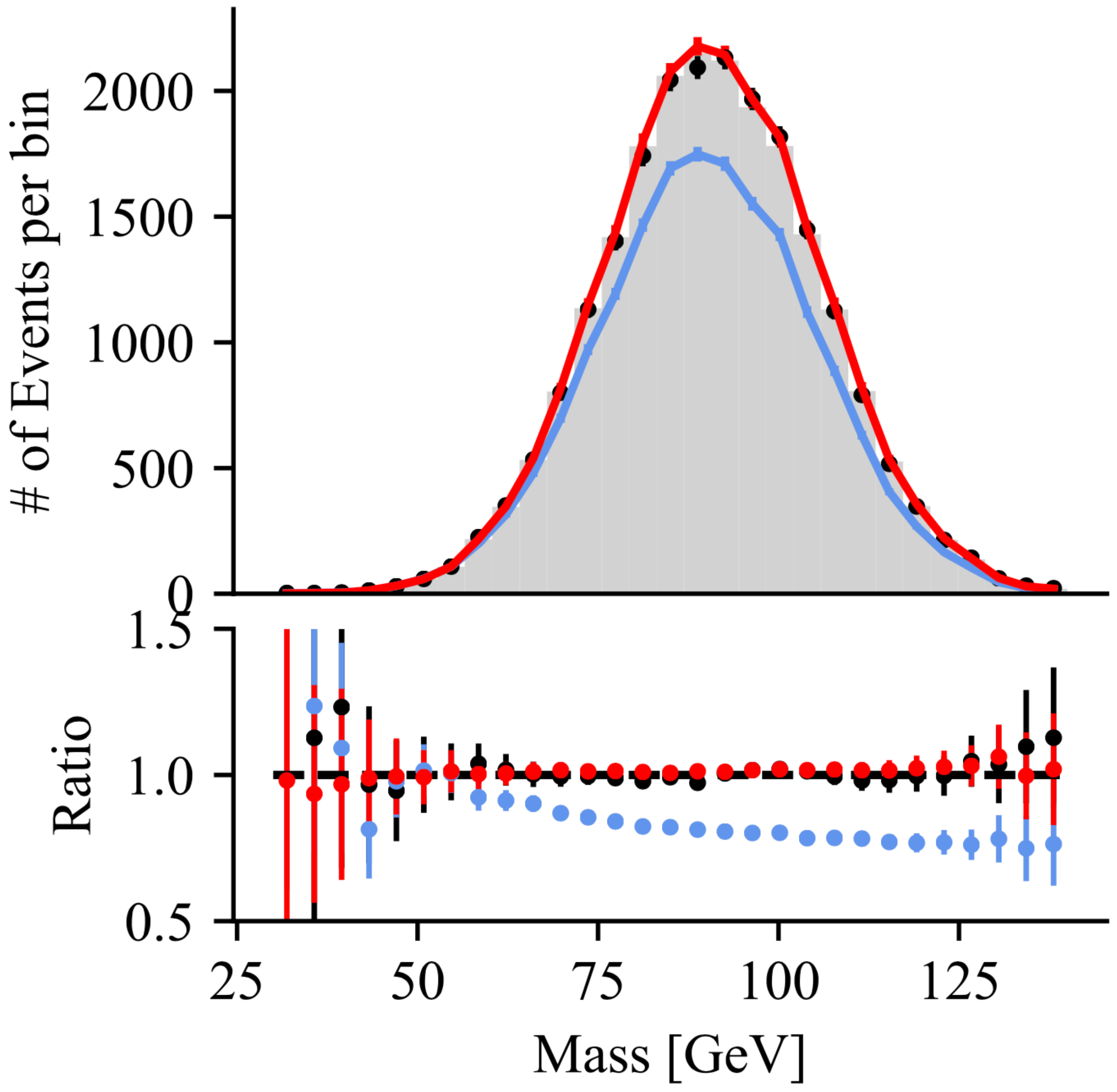




Result III - Generalisation

- New sample with a different theta distribution





Current work - Comments about θ

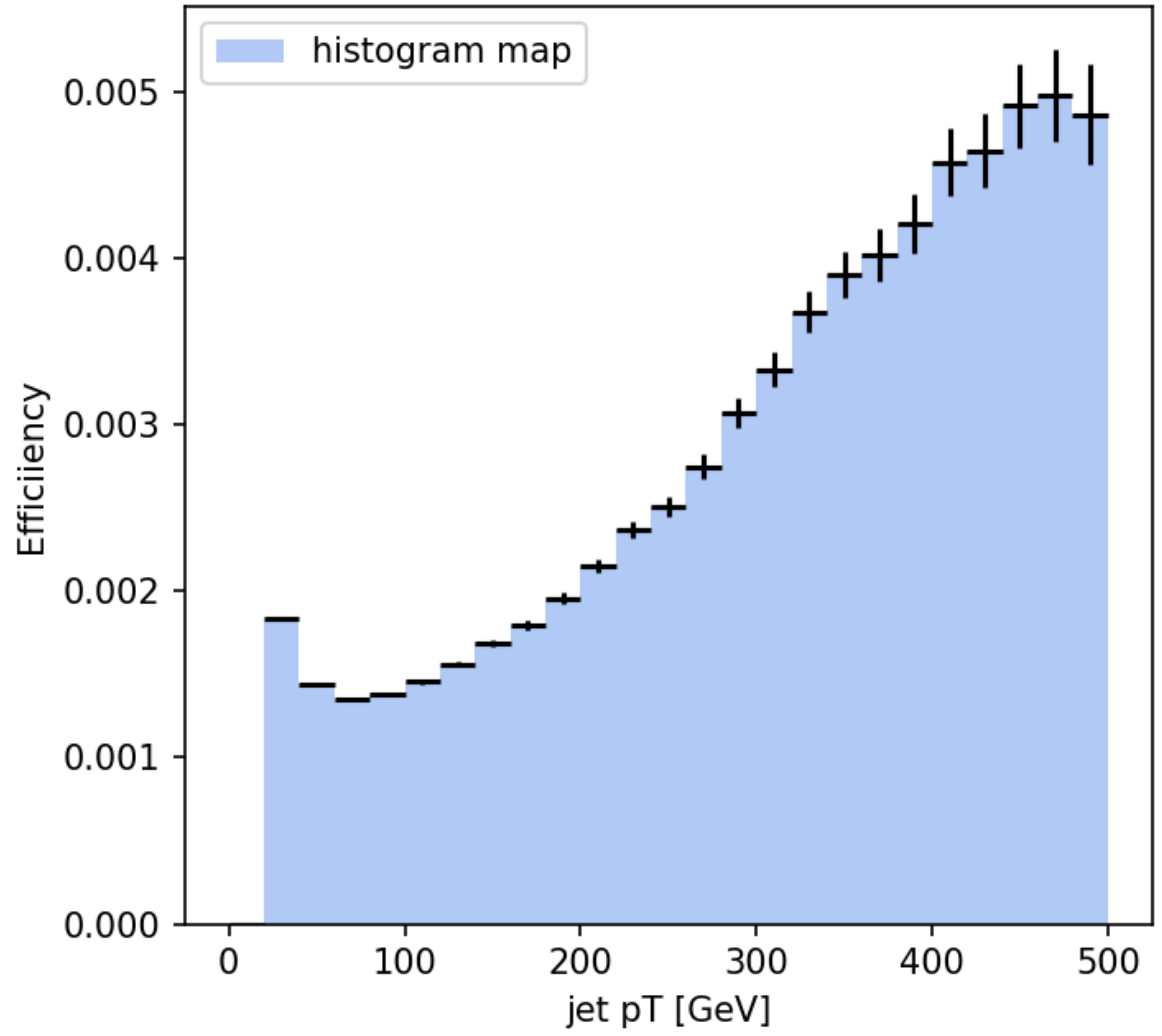
- Results from the current ATLAS specific studies are very promising
- Observation -
    - The pileup info, 'mu' helps improve the closure
    - For b and c jets, the truth hadron info helps
    - For light jets, knowing whether it has a quark or gluon origin also helps in improving the efficiency estimation

Current Work - Uncertainty Estimation

- For histogram, we have limited statistics for each bin
- That helps us construct a Confidence Interval around the estimated efficiency
- Region with less data -> more uncertainty
- How to estimate the uncertainty of the NN estimator??

    -> Bootstrapping

- We would also like to stabilise the model first



Uncertainty Estimation

Model Uncertainty

- Model = Ensemble of NN
  - Stabilise the model prediction. Repeated training should lead to the same predictions

Statistical Uncertainty

- Bootstrapping
  - Emulate multiple dataset by resampling
  - N dataset -> N model
  - Sampling distribution (during inference)
  - Construct Confidence Interval

- Model Uncertainty cannot be decoupled from the stat uncertainty. So we try to reduce its impact during bootstrapping by using ensemble training

Conclusions

- We discussed an NN based approach for efficiency estimation in a multidimensional space
- Advantages -
  - Better Efficiency estimation, as it can account for much larger number of parameters than the binned maps
  - Automatically infers theta during training
  - Learns the jet-jet dependency
  - Generalise well on sample that was not used for training
- The approach can be generalised to other studies with a similar setup.
- Currently we are looking into implementing it in ATLAS

Thank you for listening…