中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

Status of TOTAL model for Pileup study

Chu Wang, Fabio Iemmi

Institute of High Energy Physics

Quantum Computing and Machine Learning Workshop

based on arXiv:2211.02029

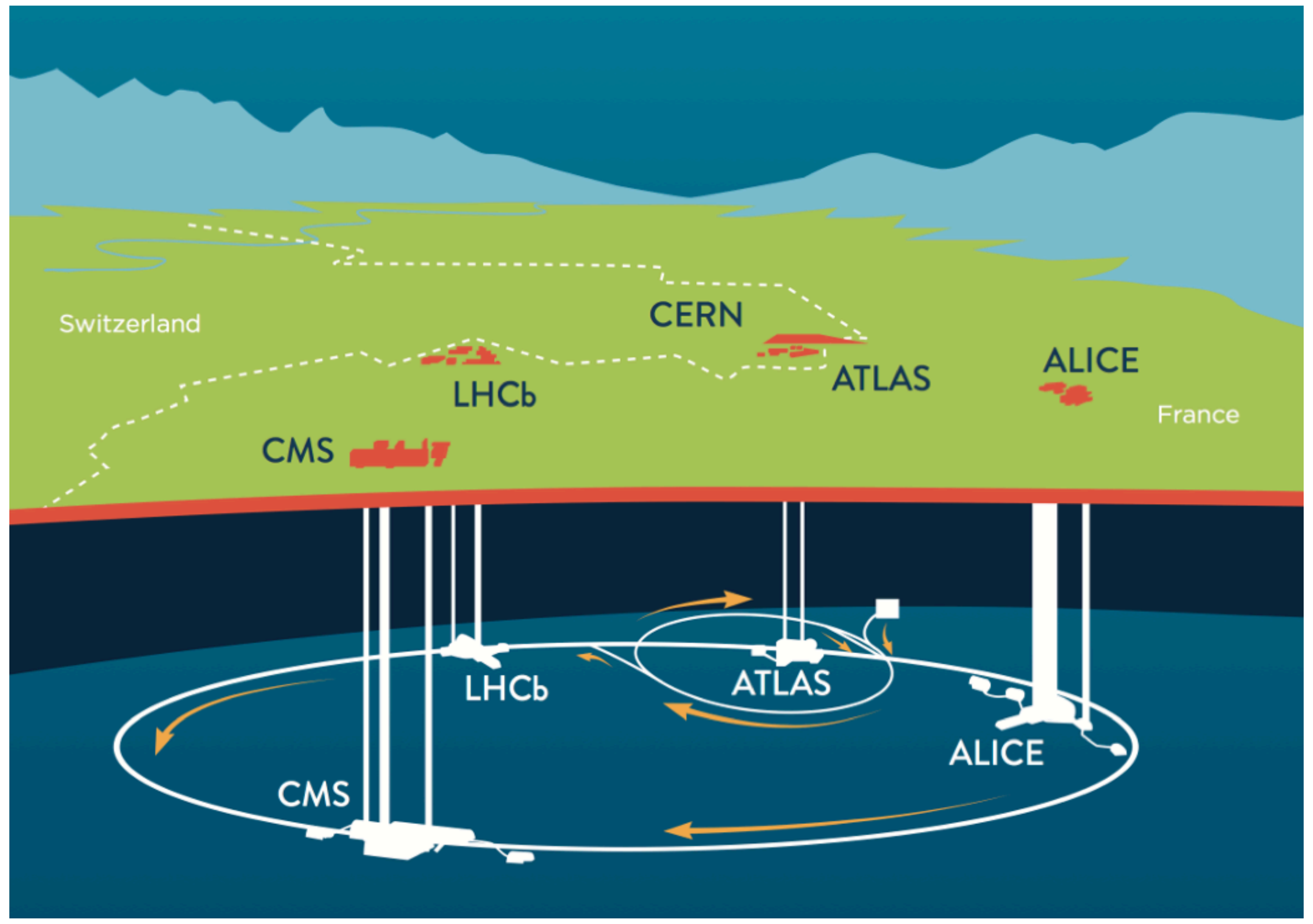
Outline

中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

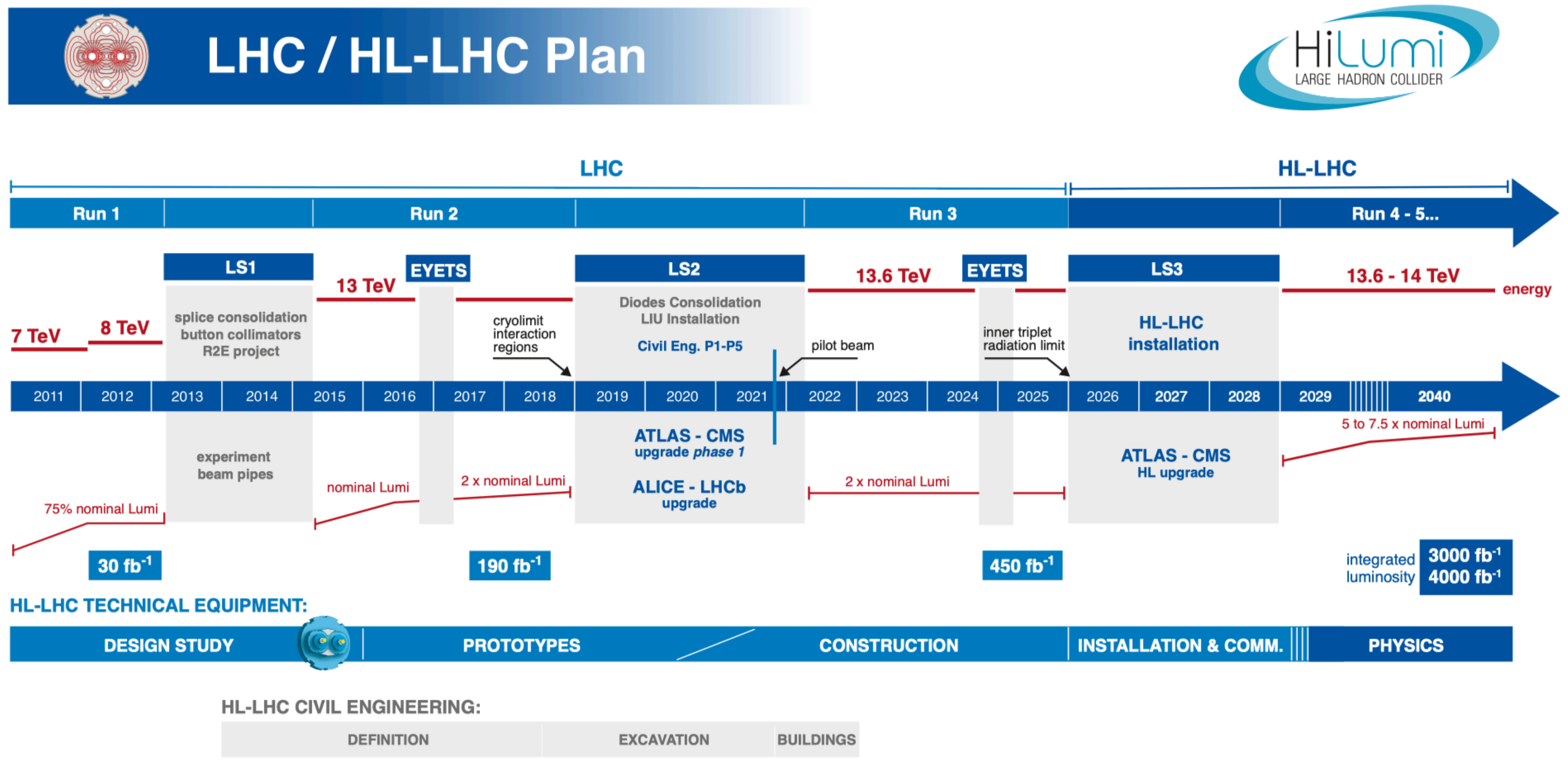
* Pileup in LHC
* TOTAL algorithm
* Performance comparison
* Summary

LHC

- Large Hadron Collider（LHC）：
  - A circular proton collider with a circumference of 27 kilometers, located about 100 meters below the French-Swiss border.
  - 4 Experiments： CMS， ATLAS， LHCb， ALICE
  - Centre of mass energy：
    - First Run（Run1）： 7-8 TeV
    - Second Run（Run2）： 13 TeV
    - Third Run （Run3）： 13.6 TeV
    - Future： 14 TeV
  - Luminosity：
    - First Run（Run1）： 30 fb^{-1}
    - Second Run（Run2）： 190 fb^{-1}
    - Third Run（Run3）： 450 fb^{-1}
    - Future： 3000~4000 fb^{-1}



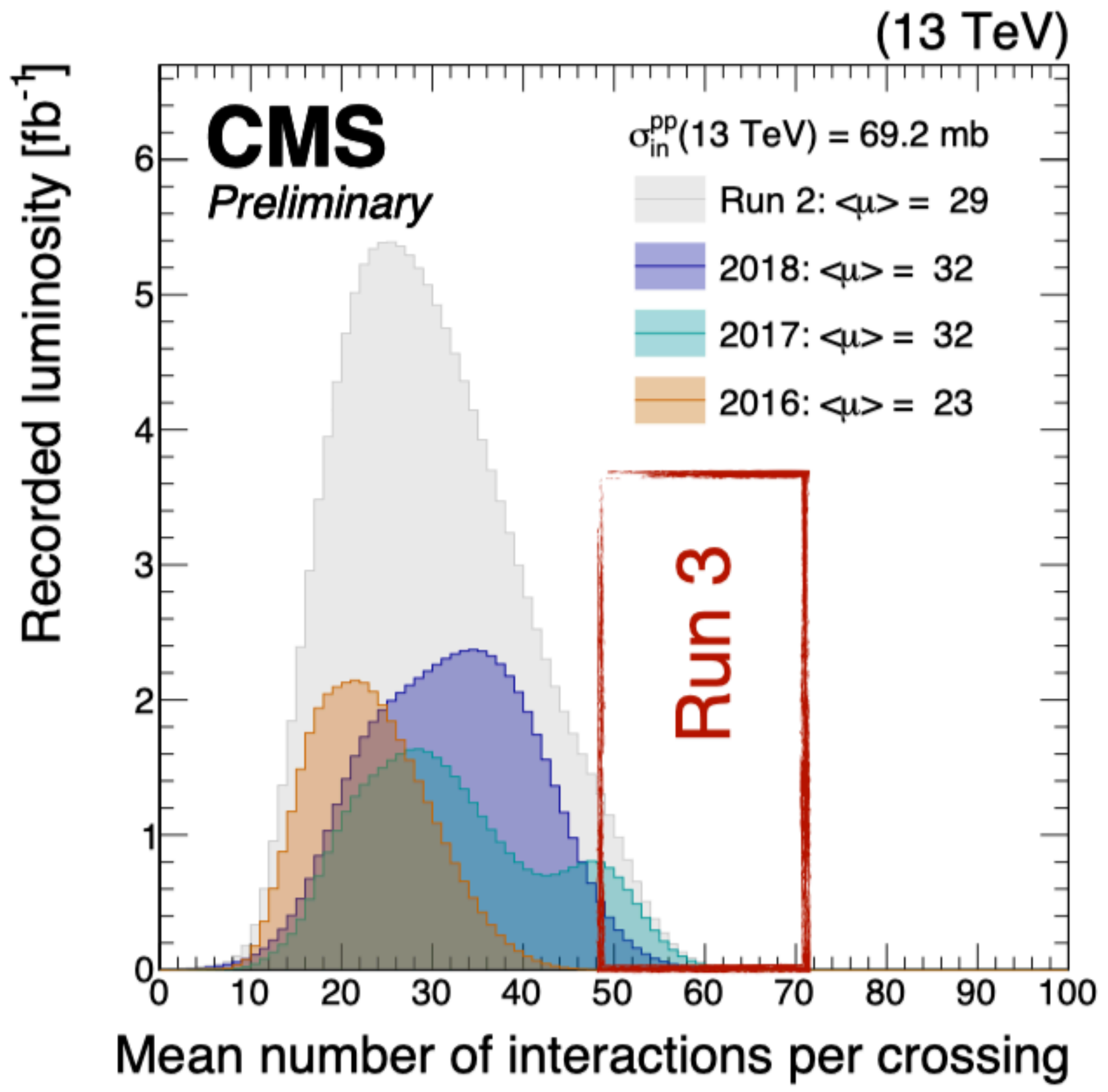
LHC and 4 experiments

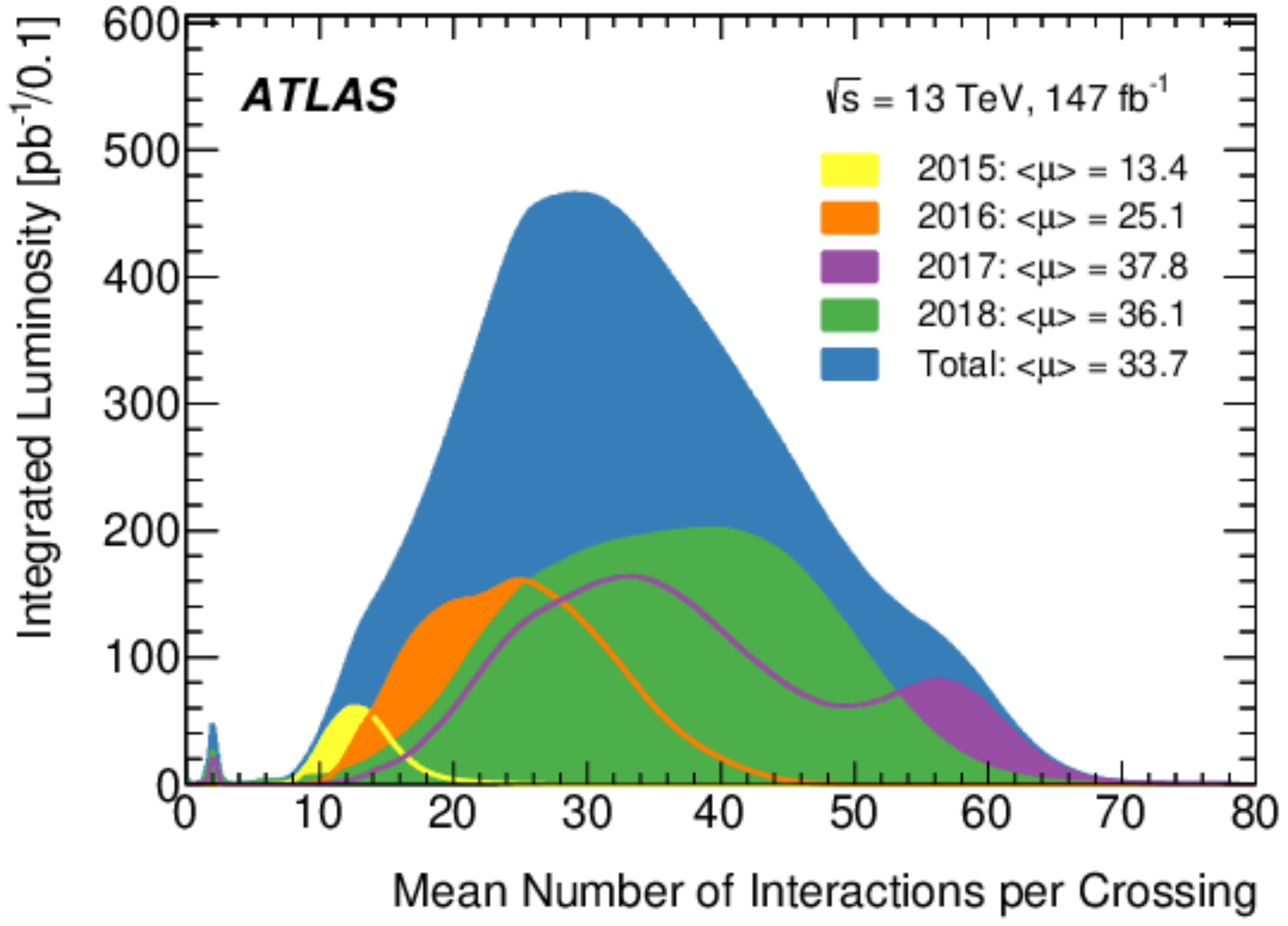


LHC Plan

Pileup in LHC

中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences





- **Pileup: Additional proton proton collisions at the LHC**
  - The average number of pileup during the Run3 was above 50 (CMS)
  - Average pileup on future high-lumi LHC will be 140
- **Will severely degrade quality of observables (jet multiplicity, jet substructure) if not properly treated**
- **PU mitigation is crucial at hadron colliders**
  - For charged particles, use tracking information to disentangle particles from PU
  - Very challenging for neutral particles

Current PU ID in CMS

- **Pileup Per Particle Identification (PUPPI)**
- **PUPPI is Rule-Based algorithm**
  - Calculates a weight $w \in [0, 1]$ for each particle in the event
    - Encodes the probability for a particle to be leading vertex (LV) or not
    - Weight used to reweight the particle 4-momentum before jet clustering
  - For charged: use tracking information and assign 0 or 1
  - For neutrals: build α variable

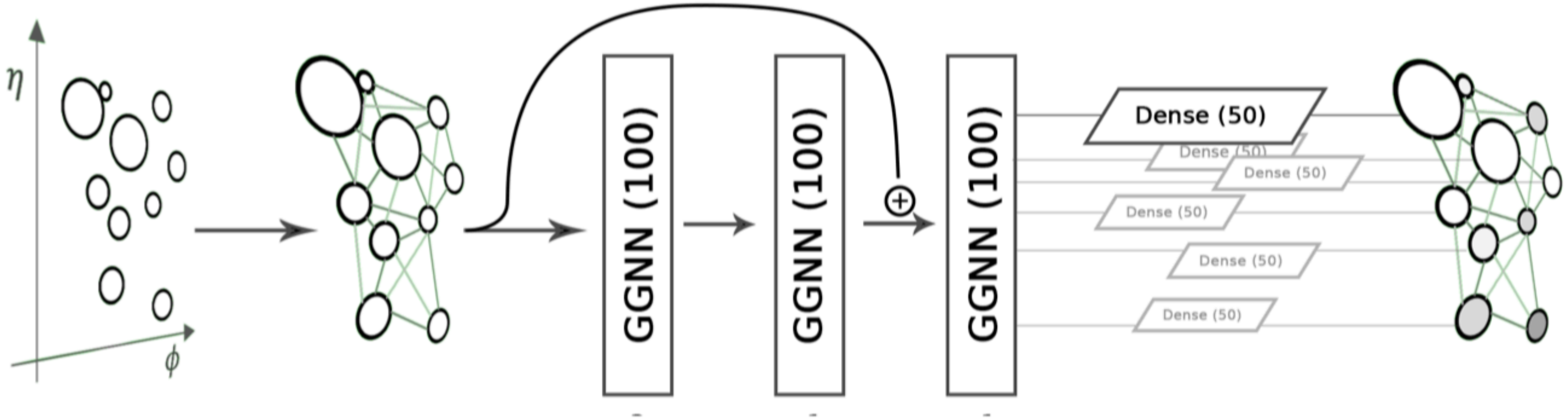
$$\alpha_i = \log \sum_{j \neq i, \Delta R_{ij} < R_0} \left(\frac{p_{T,j}}{\Delta R_{ij}} \right)^2 \begin{cases} |\eta_i| < 2.5 & j \text{ are all charged particles from LV} \\ |\eta_i| > 2.5 & j \text{ are all kinds of particles} \end{cases}$$

  - QCD is harder and more collimated than PU, so that α is much higher
  - After some math and assumptions translate α into weight

ML for pileup mitigation

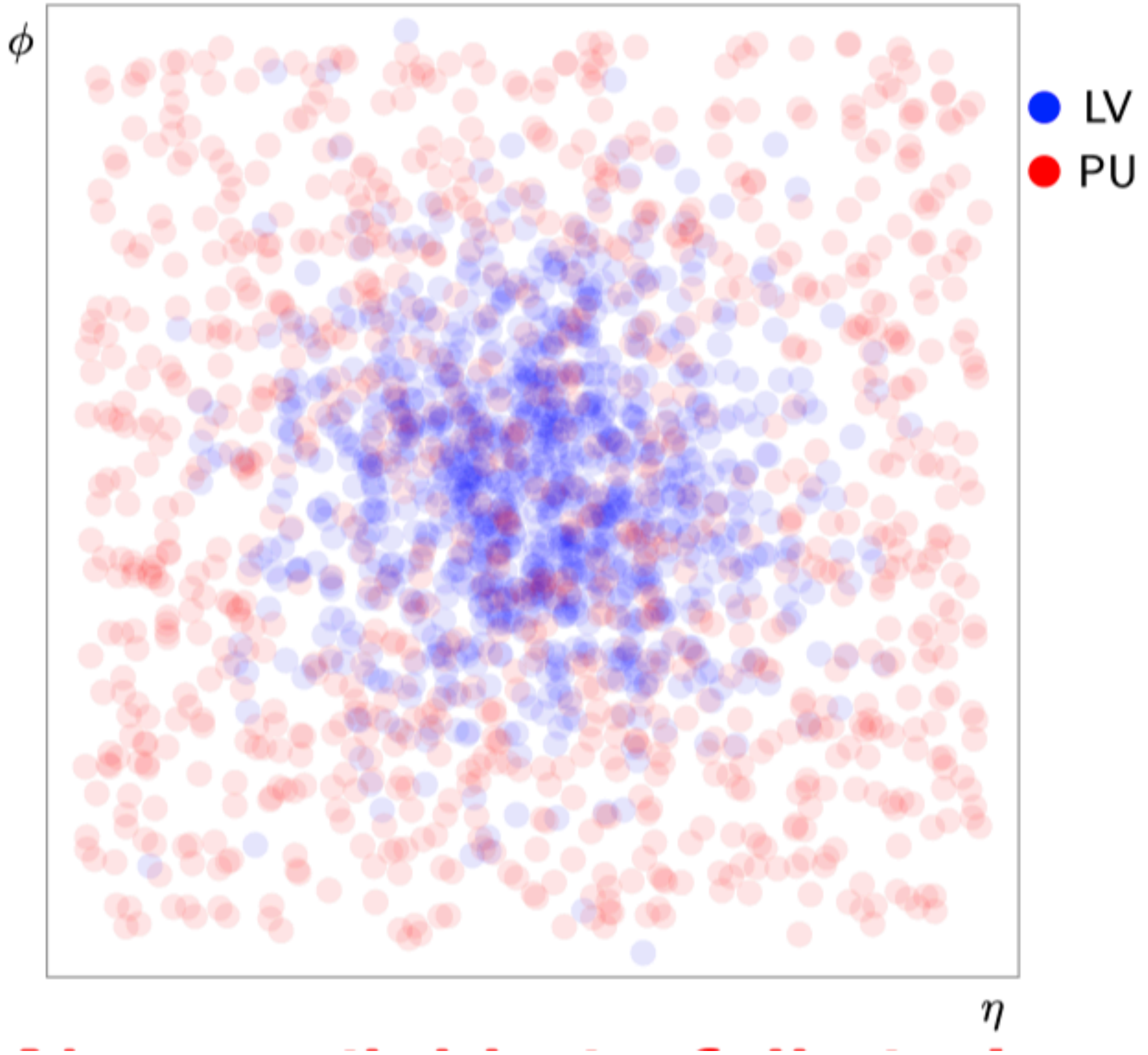
中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

- **Published literature demonstrates that ML can improve over current algorithm [1,2,3]**
- **In particular, GNNs proved to be very effective**
  - Collect info about neighbouring particles in a much more expressive way
- **General strategy in these works: train from a supervised model using Delphes fast-simulation truth labels**
  - A flag set to 1 for charged particles from the LV, -1 for charged pileup and 0 for all neutral particles. This flag provides a simple encoding for weight



Problems in current ML PU algos

- **Critical issue: per-particle labels are not available in Geant4-based full simulations**
  - Marking the source of each particle is not supported
  - Previous approaches can’t be ported to experiments such as ATLAS and CMS
- **Therefore, an algorithm that does not depend on the true labeling of each particle is needed**
  - That's the question this study will solve



Not available in full-sim!

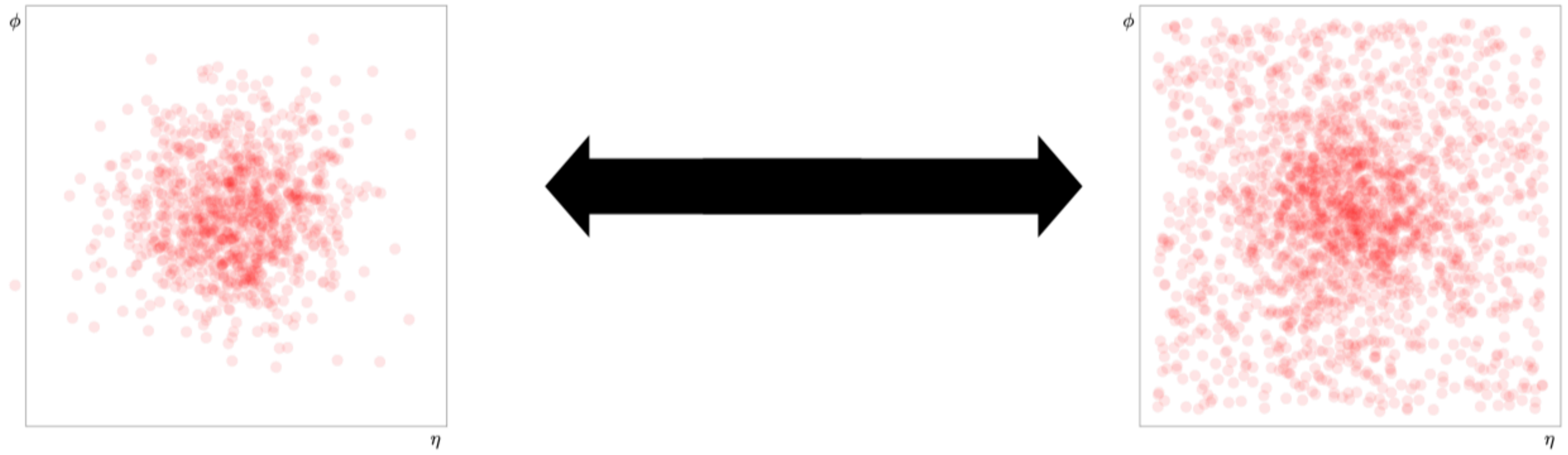
Outline

中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

- Pile Up in LHC
- TOTAL algorithm
- Performance comparison
- Summary

A novel approach in this study

- **At first, we simulate identical proton-proton collisions in two scenarios:**
  - Only the hard interaction is simulated: no-PU sample (X_{no-PU})
  - Pileup is superimposed to the hard interaction: PU sample (X_{PU})
- **Train network to learn differences between the two samples**



  - Network choice: Attention-Based Cloud Network (ABCNet)

TOTAL algorithm

中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

- **Used optimal transport，OT to design loss function**
  - Optimal transport (OT) can measure the “distance” between probability distributions
- **Network output：per-particle weights** w
  - Used to mitigate PU （PU particles are close to 0, LV particles are close to 1)
- **During training, weight PU sample（**X_{PU}**）by the weights** w
- **Tweak weights to minimize the distance between** X_{no-PU} **and** $w{\cdot}X_{PU}$
- **No need for per-particle labels in this setup**

TOTAL algorithm

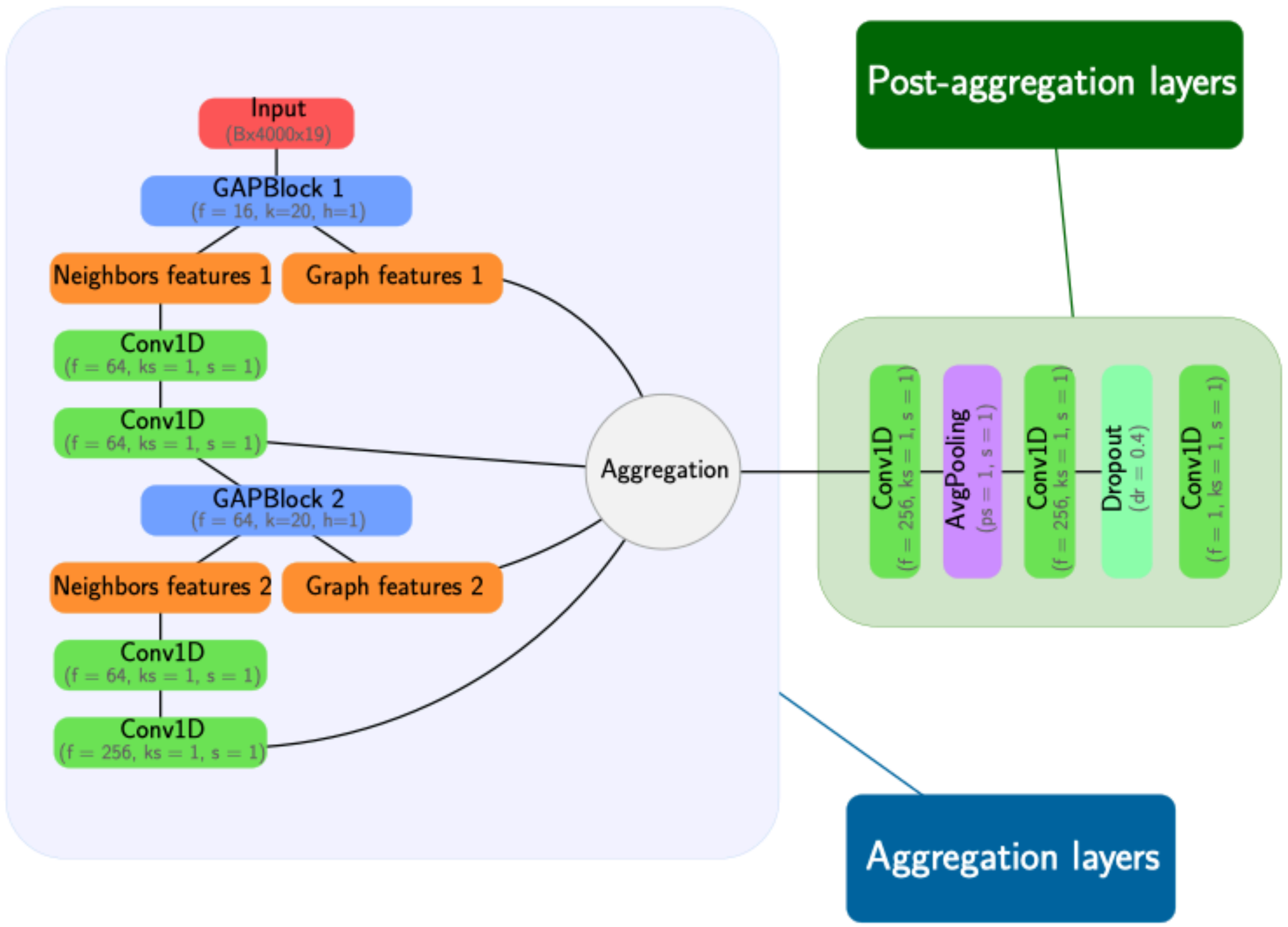
- **Sliced Wasserstein Distance （SWD）**
  - A "distance" measure for multidimensional variable distributions based on optimal transportation
  - No guarantee that energy is conserved between the two
- **Add an event-level MET constraint term to the loss**
  - Enforce energies in no-PU and PU events to be similar
- **Final loss function:**
  - $\mathcal{OT} = \mathrm{SWD}(\omega \cdot \mathbf{X}_{\mathrm{PU}}, \mathbf{X}_{\text{no-PU}}) + \lambda \times \mathrm{MSE}\left(\mathrm{MET}(\omega \cdot \mathbf{X}_{\mathrm{PU}}), \mathrm{MET}(\mathbf{X}_{\text{no-PU}})\right)$
- **Call this Training Optimal Transport with Attention Learning: TOTAL**

Outline

中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

- LHC and CMS experiment
- Pile Up in CMS
- TOTAL algorithm
- **Performance comparison**
- Summary

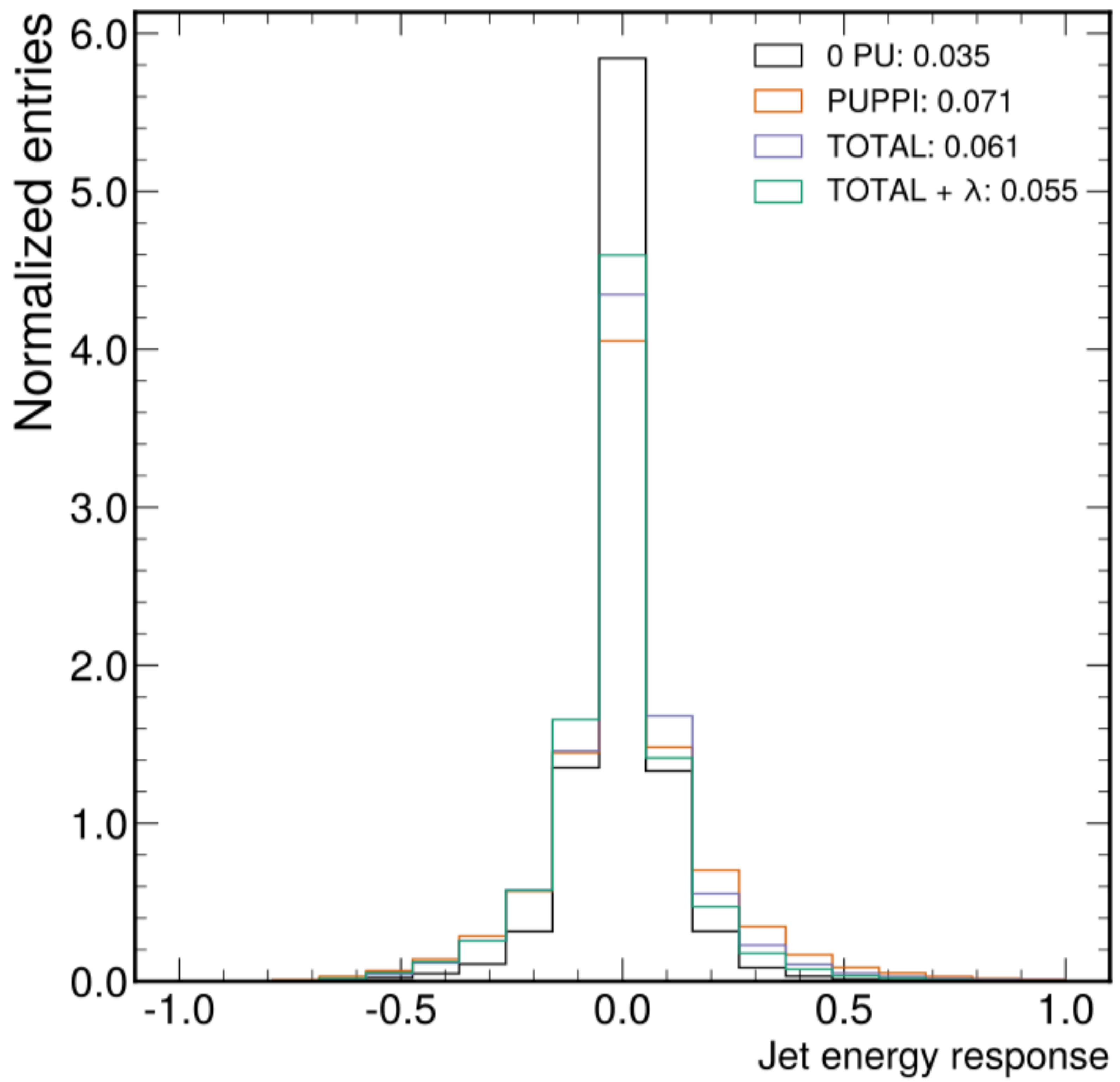
The Model



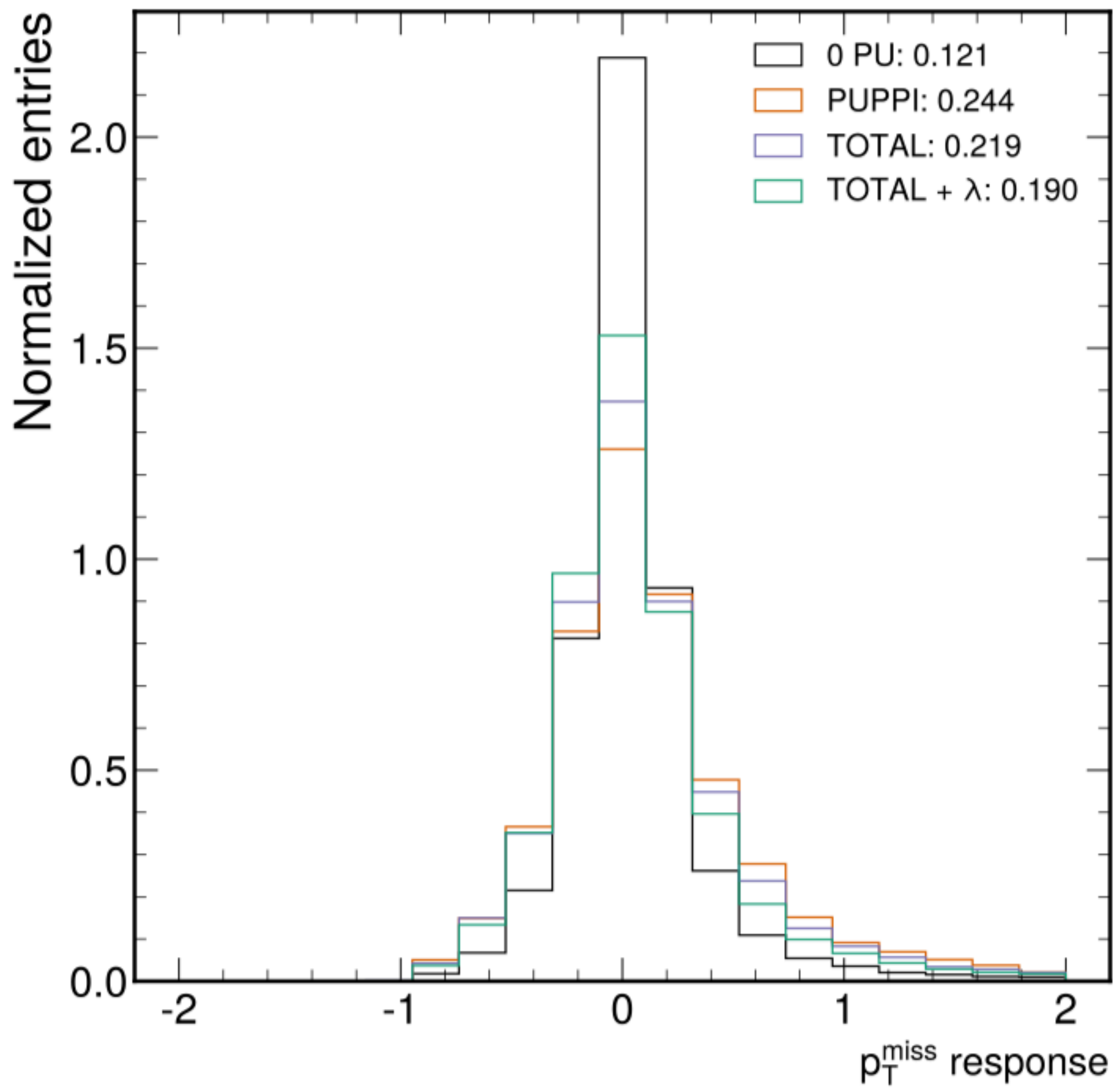
- Compare TOTAL with PUPPI and no-PU scenario
- Reweight each particle's 4-momentum by the network weight
- Cluster TOTAL jets and TOTAL MET
- Define the resolution as:
  - $$\delta = \frac{q_{75\%} - q_{25\%}}{2}$$
  - where q is the X-th quantile of the considered response distribution

Energy response

- The improvement in Jet energy response is about 23% relative to PUPPI and 22% for missing transverse momentum



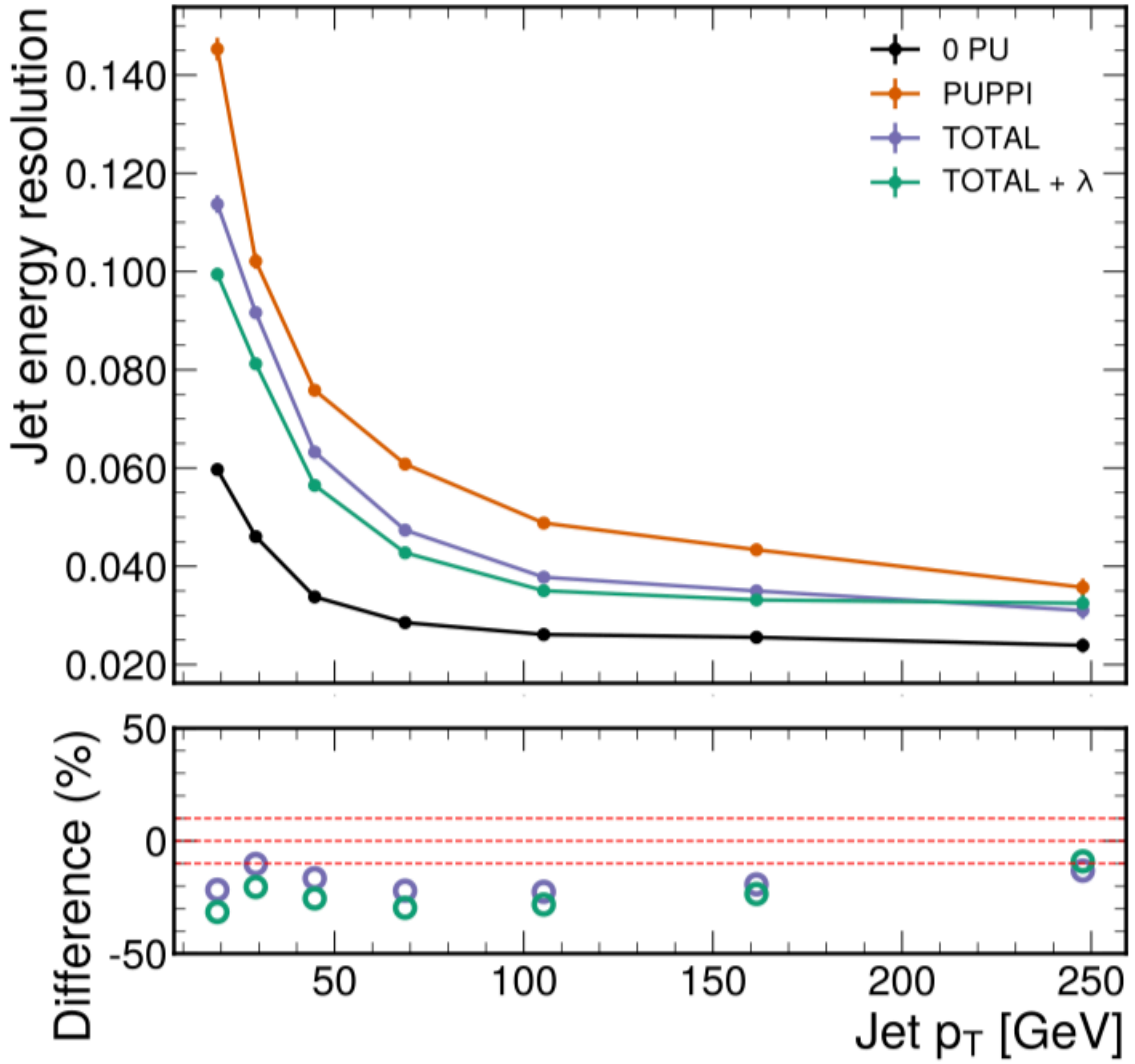
喷注能量响应



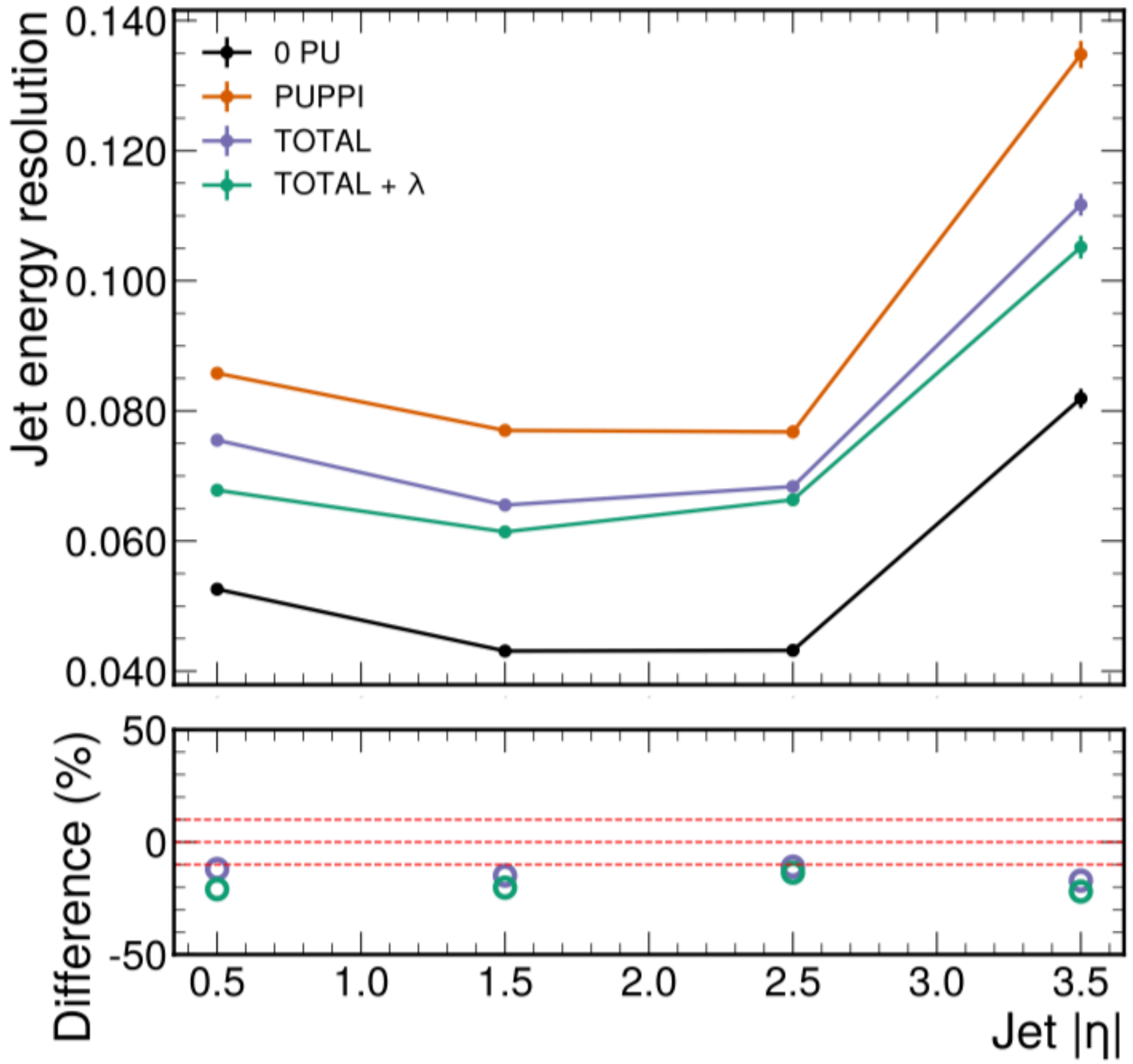
缺失横动量响应

Jet differential resolutions

- **Improvement of Jet energy resolution relative to PUPPI: 20～30%**



Injection energy resolution varies with injection energy

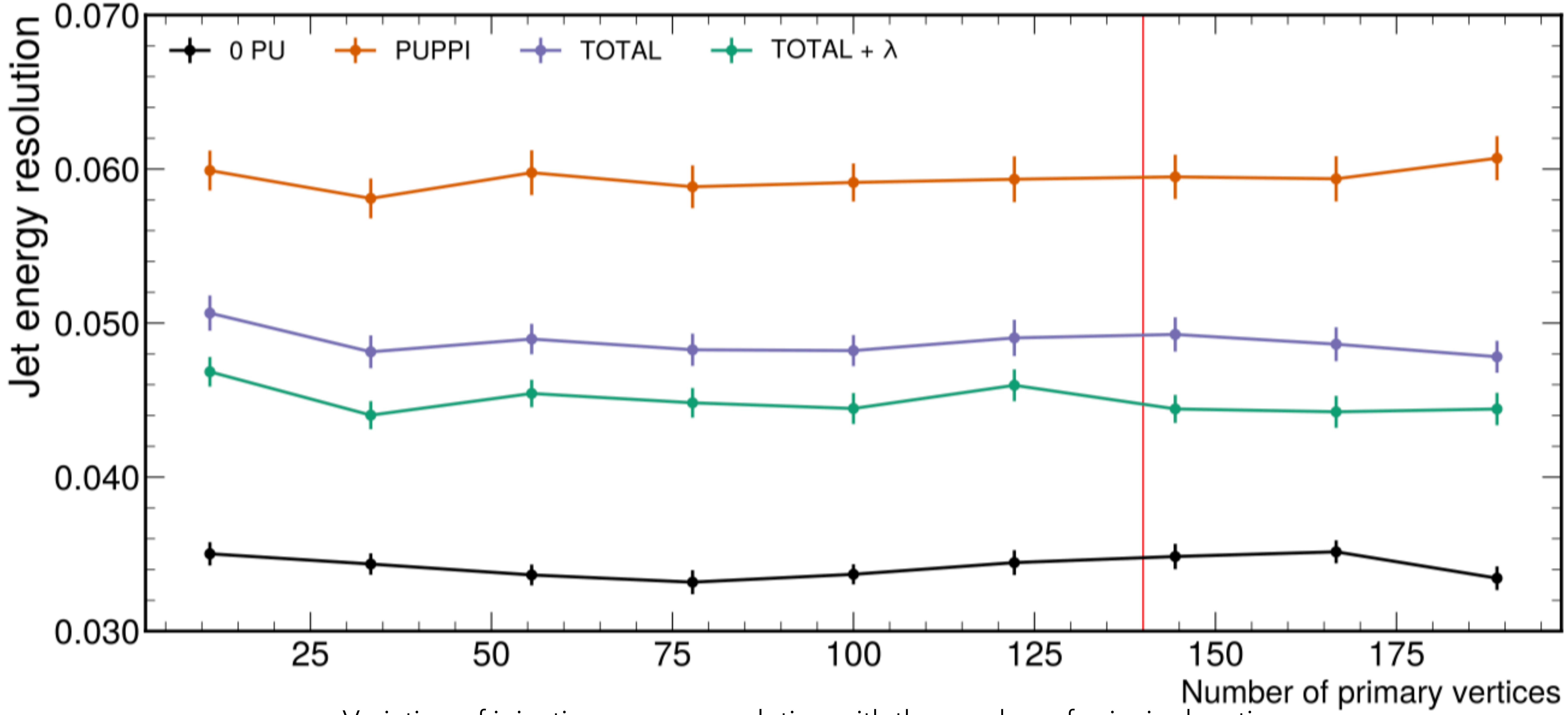


Variation of injection energy resolution with η

Robustness

Institute of High Energy Physics
Chinese Academy of Sciences

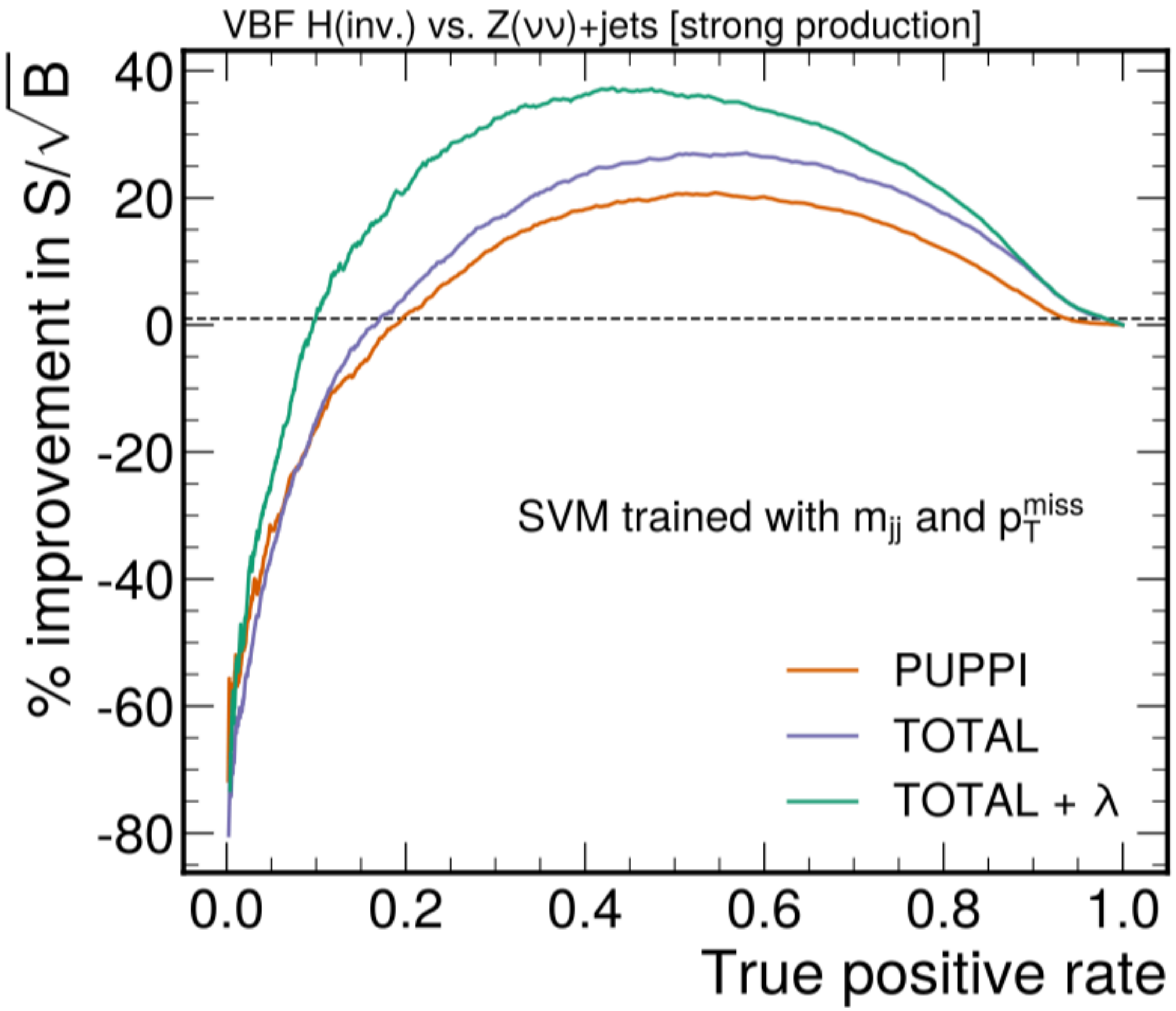
- **The model in this study was trained based on 140 primary vertices, but it still works for 200 vertices with little change**



Variation of injection energy resolution with the number of principal vertices

Physics impact

- Study impact of TOTAL on LHC searches:
  - Search for BSM VBF H(inv.)
- Signal signature: pair of forward jets and MET
- Main background: strongly produced Z($\nu\nu$) + Jets
- Perform toy analysis by training a linear classifier (SVM) using dijet mass and MET
- Observe the change in significance (S/$\sqrt{B}$) with signal efficiency:
  - TOTAL is about a 15% improvement over PUPPI



Outline

中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

- Pileup in LHC
- TOTAL algorithm
- Performance comparison
- Summary

Summary

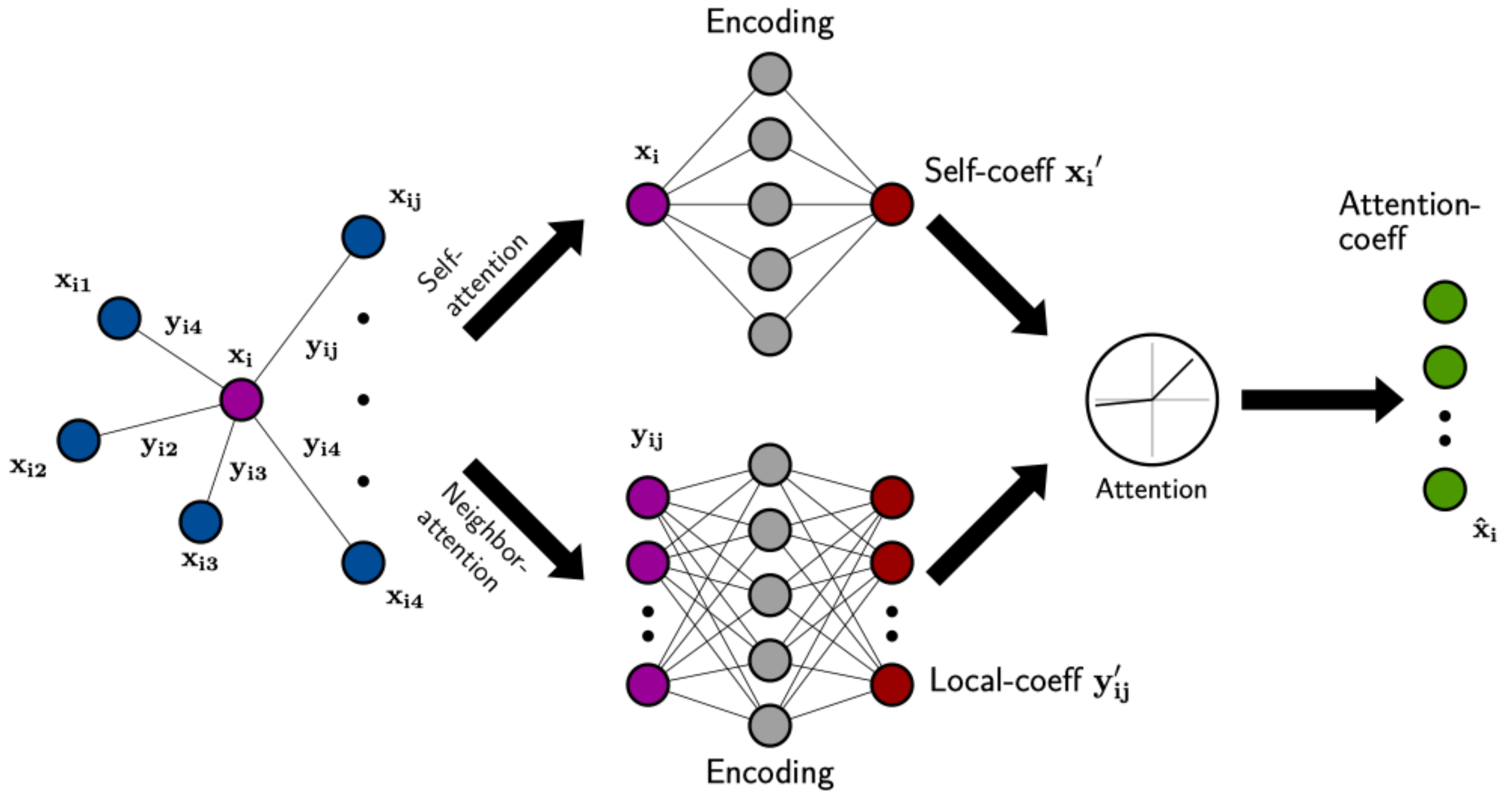
中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

- **We presented novel algorithm to reject PU particles at high-intensity hadron colliders**
  - Trained and tested on Delphes simulation of Phase2 CMS detector
- **We are Training Optimal Transport with Attention Learning: TOTAL**
- **We solved the longstanding problem of neutral labels in PU mitigation**
- **We do not rely on explicit, per-particle labelling**
- **Such an algorithm will be crucial at the High-Luminosity LHC, where much harsher data-taking conditions are expected**
- **Our approach can be generalized to a wide range of denoising problems**

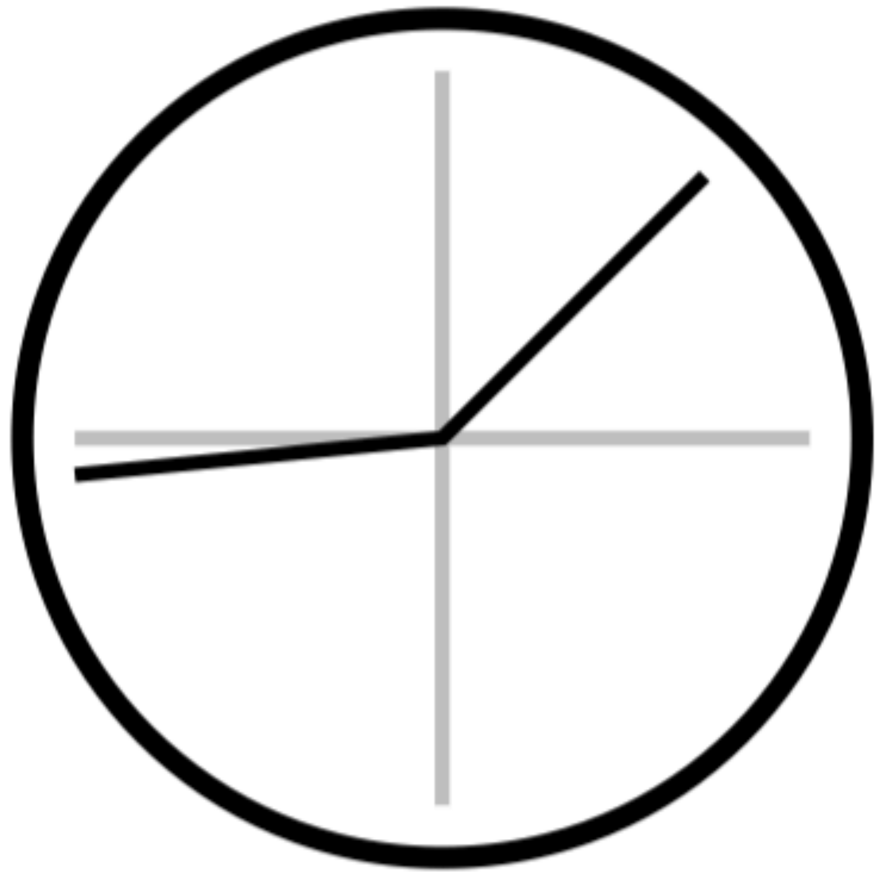
Backup

Attention Based Cloud Network

- **ABCNet is an graph neural network enhanced with attention mechanisms**
  - Treat particle collision data as a set of permutation-invariant objects
  - Attention mechanisms filter out the particles that are not relevant for the learning process
  - Implemented inside custom graph attention pooling layers (GAPLayers)



Attention mechanism



Attention

- Add together self- (x_i') and local- (y_{ij}') coefficients and apply non-linearity

$$c_{ij} = \text{LeakyRelu}(x_i' + y_{ij}')$$

- Align coefficients c_{ij} by applying SoftMax

$$c_{ij}' = \frac{\exp(c_{ij})}{\sum_k \exp(c_{ik})}$$

- Get attention coefficients by multiplying y_{ij}' by c_{ij}'

$$\hat{x}_i = \text{Relu}\left(\sum_j c_{ij}' y_{ij}'\right)$$

Efficient OT

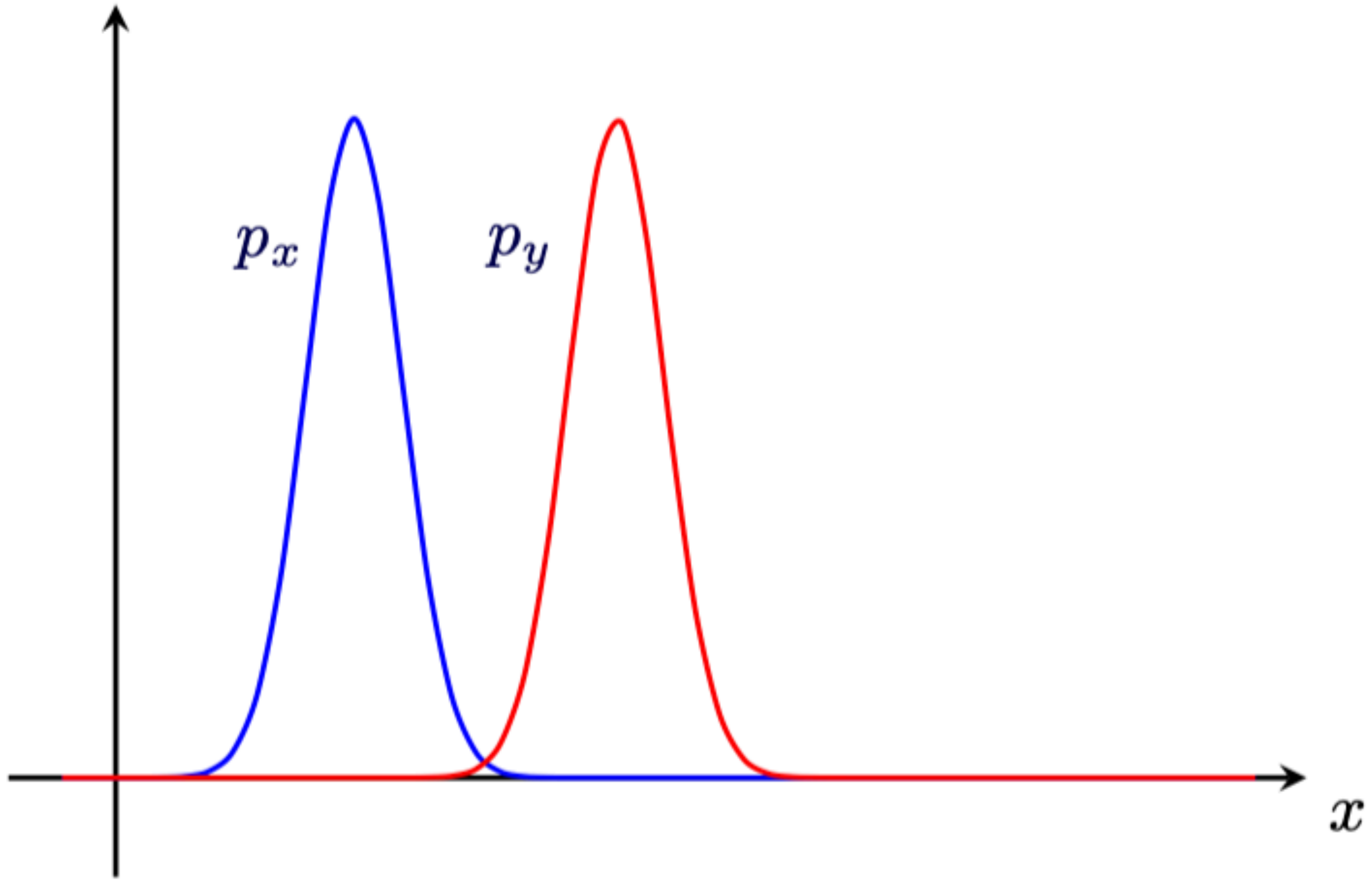
- **The optimal transport problem has a closed form for 1D problems:**

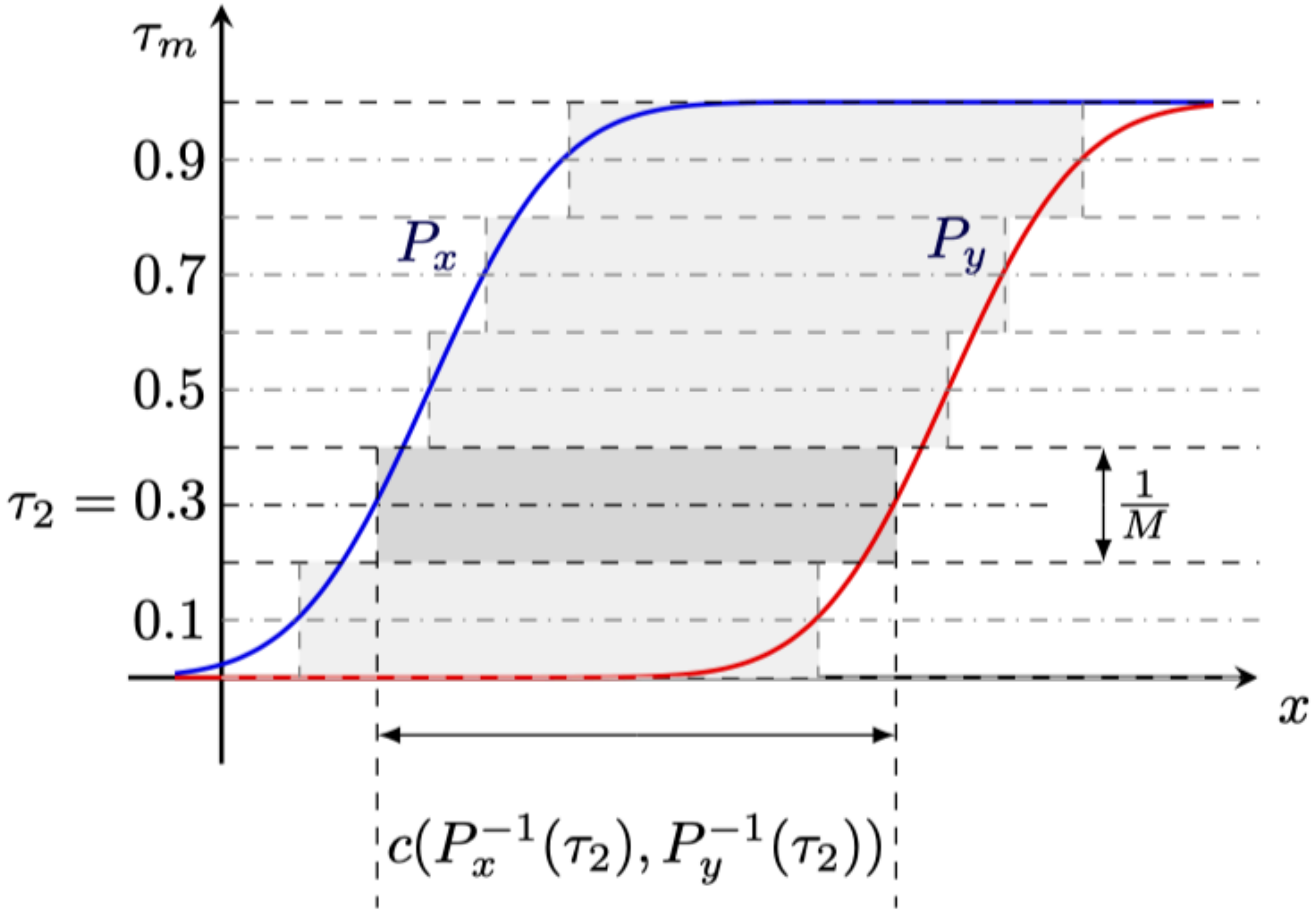
$$W_c(p_X, p_Y) = \int_0^1 c\left(P_X^{-1}(\tau), P_Y^{-1}(\tau)\right) \mathrm{d}\tau$$

- where pX , pY are 1D PDFs, $PX^{-1}(\tau)$, $PY^{-1}(\tau)$ are the respective CDFs and c(·, ·) is the transportation cost function
- No guarantee that the integral is solvable
- The integral can always be approximated by the finite sum

$$\frac{1}{M}\sum_{m=1}^{M} c\left(P_X^{-1}(\tau_m), P_Y^{-1}(\tau_m)\right), \qquad \tau_m = \frac{2m-1}{2M}$$

Example: M=5

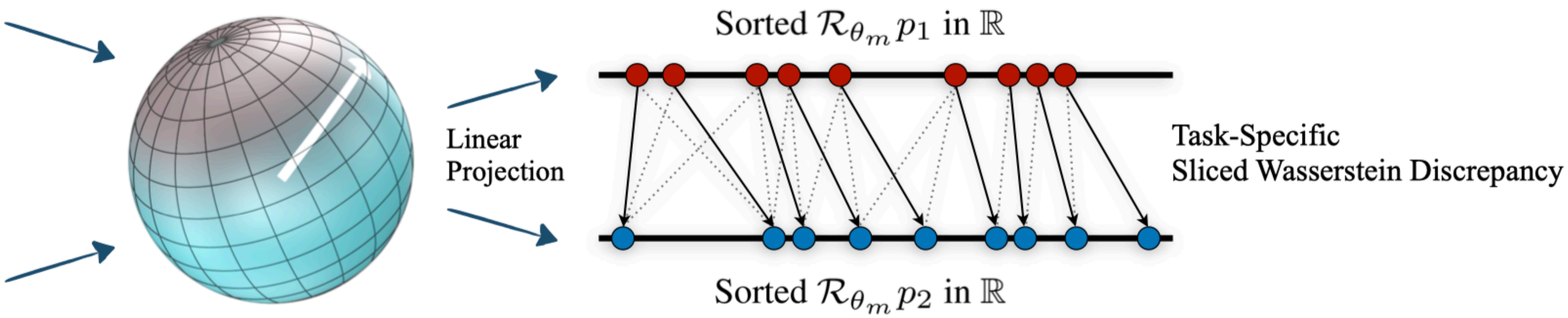




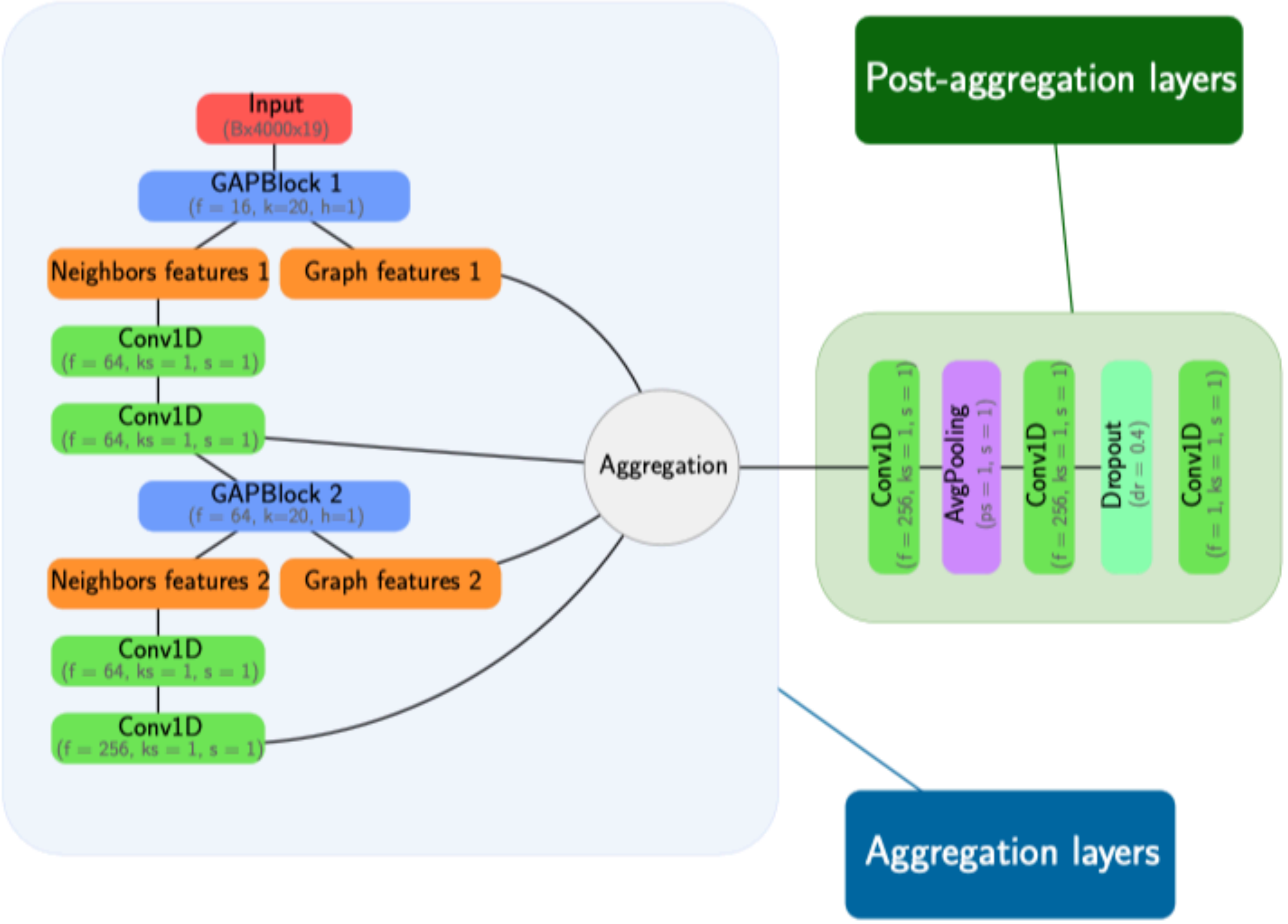
- $m \in \{1,2,3,4,5\} \implies \tau_m = \frac{2m-1}{2M} \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$

sliced Wasserstein distance (SWD)

- Each particle is a sample from a n-D feature space
- **SWD**: take n-D feature space and **project (slice)** it **to 1D**
- Project on a vector belonging to S^{n-1}
- For robustness, take **multiple random slices**
- Now can **solve the 1D OT problem for each slice**
- **Sort particles by slice**
- The **average on all slices and particles** becomes the **loss function**



The model



- **9 input features**:
  - (p_T, η, ϕ, E)
  - Charge
  - PDG ID
  - dXY & dZ impact parameters
  - Vertex association (for charged)
- **Loss**: $\mathrm{SWD}(\vec{x}_p \cdot \vec{\omega}, \vec{x}_{np}) + \mathrm{MET}$ constraint
- **Cost function**: squared distance
- **Sliced features**: (p_T, η, ϕ, E)
- **Output**: per-particle weight $\vec{\omega}$

- Train on **300k events**, equally split between QCD multijet, $t\bar{t}$ dileptonic and VBF Higgs(4ν) processes
- Consider **9000 particles per event** (zero-padding included)
- Gather the **20 *k*-nearest neighbors** for each particle when building graph