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

# Optimal transport solutions for pileup mitigation at hadron colliders


L. Gouskos [1], **F. Iemmi** [2], S. Liechti [4], B. Maier [1]
V. Mikuni [3], H. Qu [1]

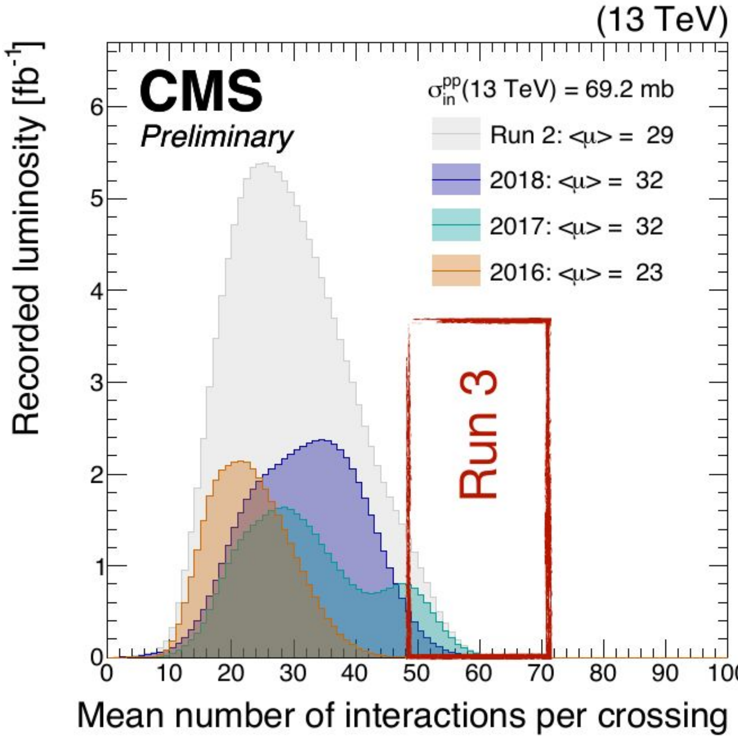

CLHCP, November 19th, 2023

Based on PhysRevD.108.096003

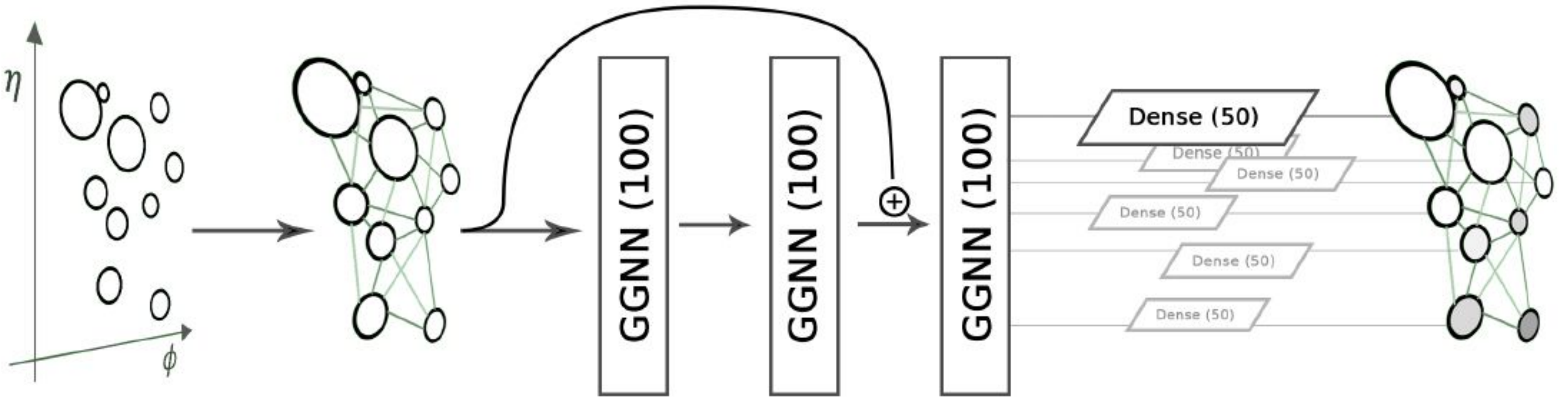

[1] European Organization for Nuclear Research (CERN)
[2] Institute of High Energy Physics (IHEP), Beijing
[3] National Energy Research Scientific Computing Center (NERSC), Berkeley
[4] University of Zurich (UZH), Zurich

# Introduction: state of the art in PU mitigation

- **Pileup is ubiquitous** at hadron colliders
- PU doubled in Run3 wrt Run2
  - Will reach <PU> = 140 at HL-LHC
- **State of the art** PU mitigation in CMS: **PUPPI**
  - **Cut based algorithm**
  - For each particle, check activity in a small cone around it
  - Obtain a per-particle probability to be LV
- Attempting something similar with **machine learning sounds like a natural choice**



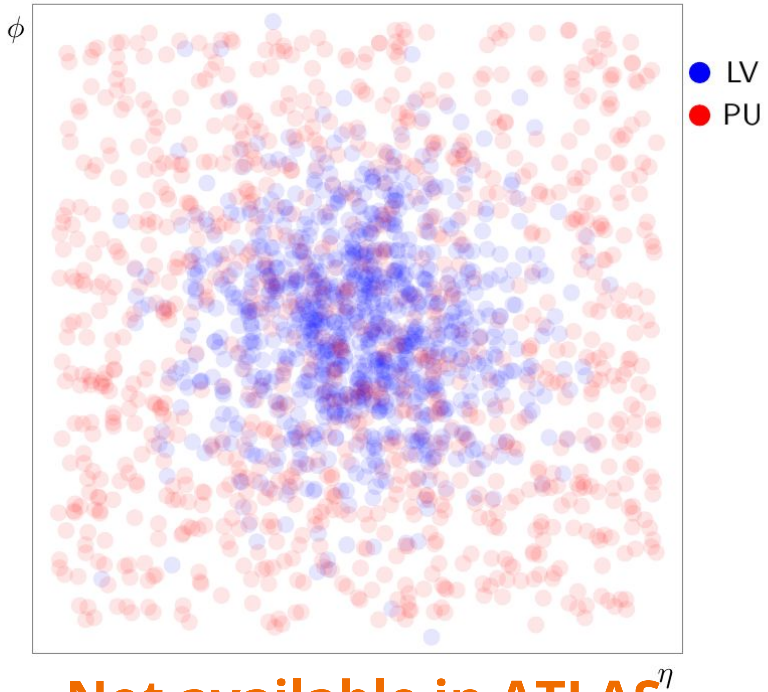
# Introduction: ML for PU mitigation




- Published literature demonstrates **ML can drastically improve over current PU mitigation techniques** [1], [2], [3]
- In particular, **graph neural networks proved to be very effective**
    - For each particle, gather information about its neighbors in a much more expressive way
- **General strategy: fully-supervised models** trained on Delphes simulation using per-particle truth labels

# Introduction: ML for PU mitigation

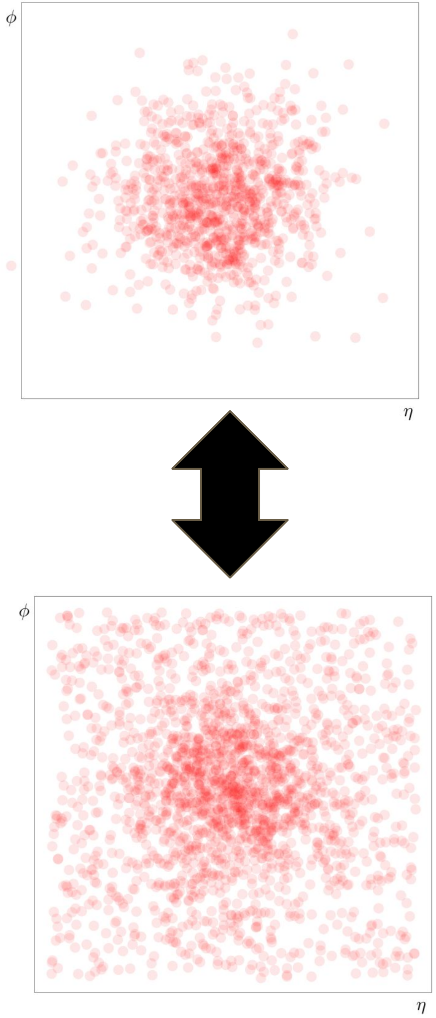
- **Critical issue**: **per-particle labels are not available in Geant4**-based simulations
  - Previous approaches can't be implemented in ATLAS or CMS
- Recently it has been proposed to train on charged and evaluate on neutrals [1]
  - Can be done in Geant4 full-sim
  - Relies on extrapolations
  - Charged–>neutral, central–>forward
- We **developed a ML-based PU mitigation strategy not relying on labels or extrapolations**




**Not available in ATLAS or CMS!**

# Optimal transport for PU mitigation

- Optimal transport (OT) can **measure the “distance” between probability distributions**
- Consider a sample of events, generated twice, with and without PU ($\mathbf{X}_{\mathrm{PU}}$ and $\mathbf{X}_{\text{no-PU}}$)
  - Stores the same events, but one has PU superimposed
- Train a network to **minimize the distance between PU and no-PU samples**
  - Use Attention-Based Cloud network: **ABCNet** [1]
- Output of the network is a collection of per-particle weights **ω**
- Loss function is: OT ($\mathbf{X}_{\text{no-PU}}$, $\boldsymbol{\omega} * \mathbf{X}_{\mathrm{PU}}$)
- **We don’t need any per-particle labels**
- **TOTAL**: training optimal transport with attention learning




No-PU

PU

# Loss function

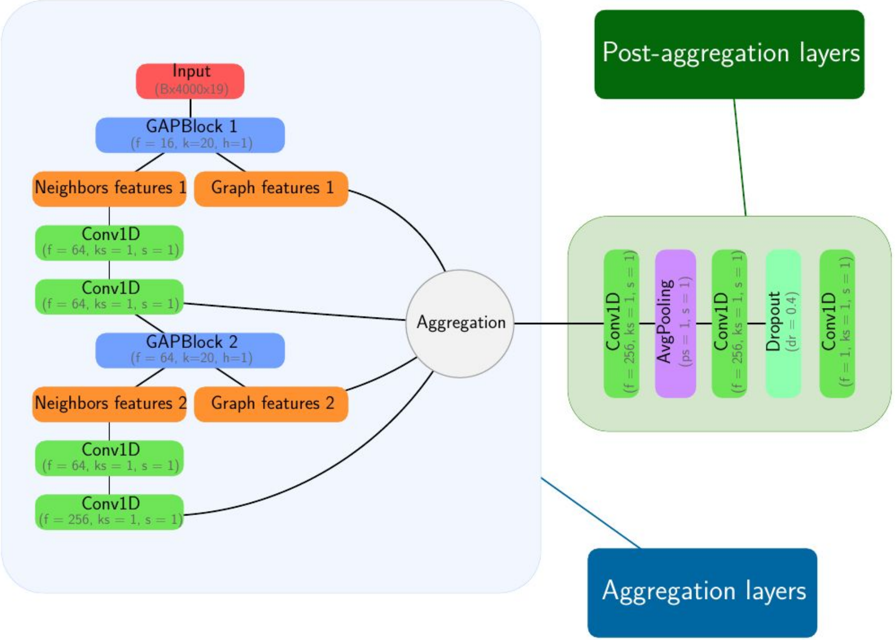
- Optimal transport focuses on optimal matching between individual particles in PU and no-PU samples
  - No guarantee that energy is conserved between the two
- Possible to add an energy constraint term to the loss
  - Enforces energies in the PU and no-PU samples to be similar
- Final loss function

$$\boxed{\mathcal{L} = \mathrm{OT}(\vec{x}_p \cdot \vec{\omega}, \vec{x}_{np}) + \lambda \times \mathrm{MSE}\,(\mathrm{MET}(\vec{x}_p \cdot \vec{\omega}), \mathrm{MET}(\vec{x}_{np}))}$$

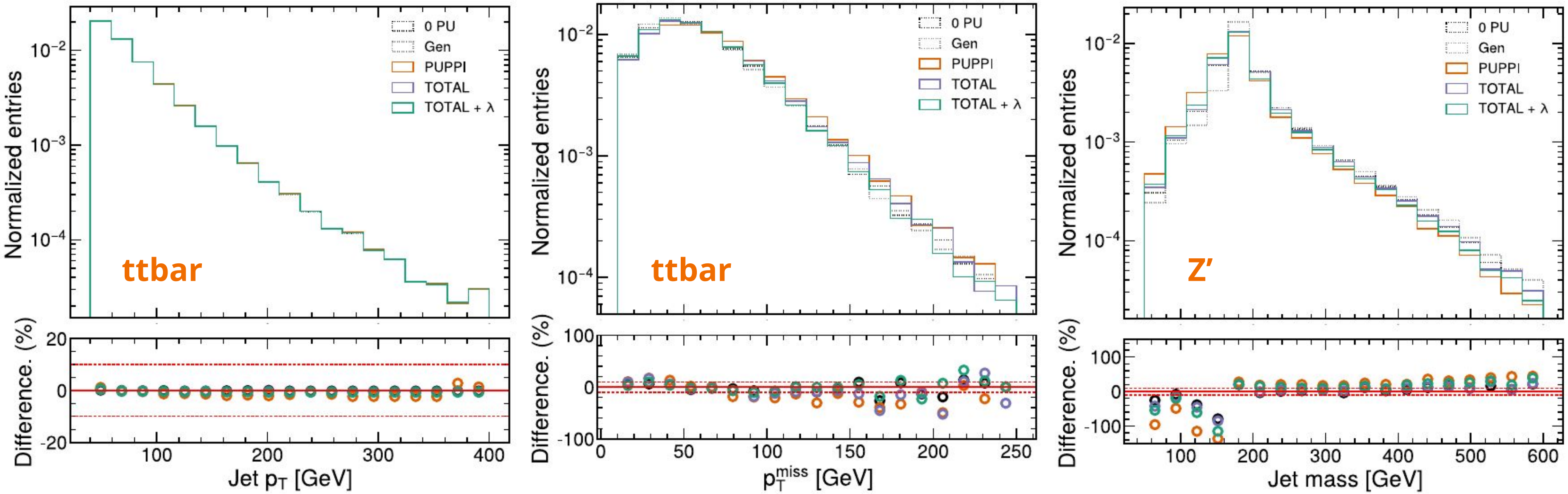
where $\vec{x}_p = \text{PU sample; } \vec{x}_{np} = \text{no-PU sample; } \mathrm{MSE} = \text{mean squared error}$

# The model

- Delphes simulation of Phase2 CMS detector
- 9 input features
  - Detailed list in backup
- Train with mixture of QCD, ttbar and VBF(H->inv.) events
- Train with Phase2 PU profile
  - <PU> = 140
- **Output is per particle weight**
- **Use such output à-la-PUPPI**
  - Reweight each PF candidate 4-momentum
- From weighted PF collection, **cluster jet and MET collections**

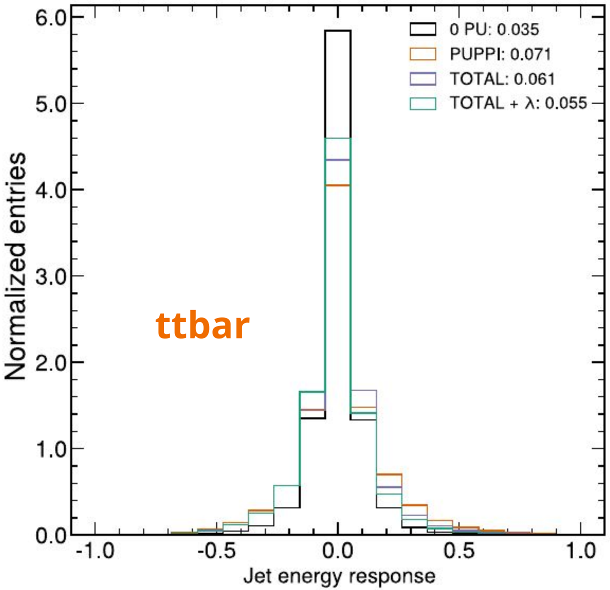


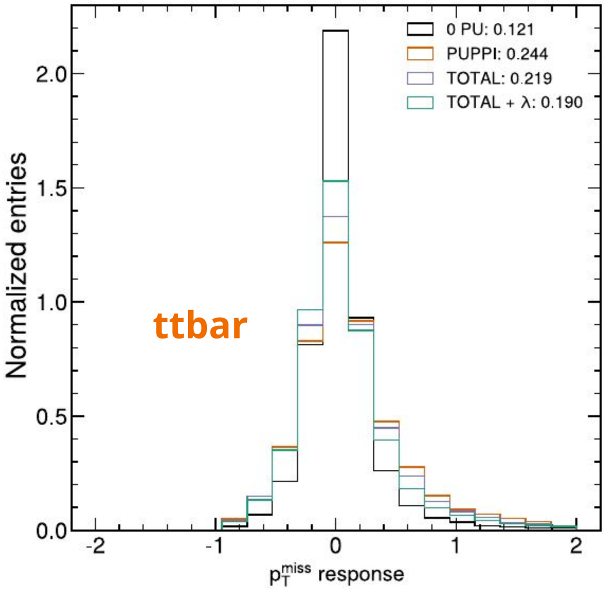
# Event-level distributions

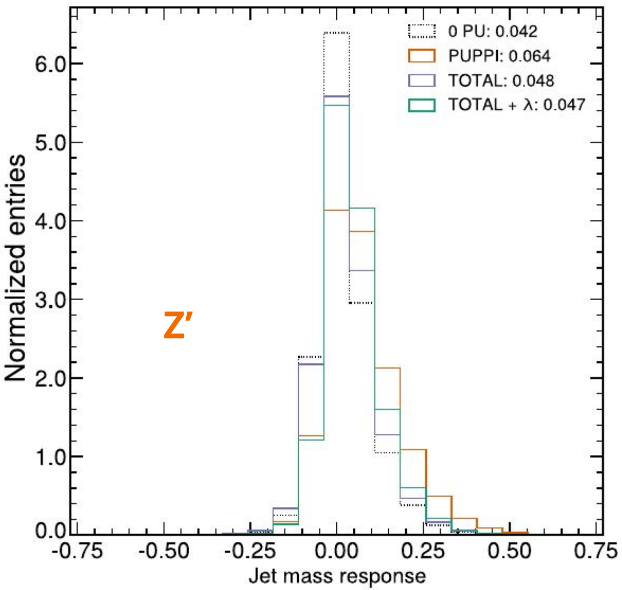



- Event level distributions show no distortions wrt gen level, no-PU and PUPPI

# Inclusive resolutions

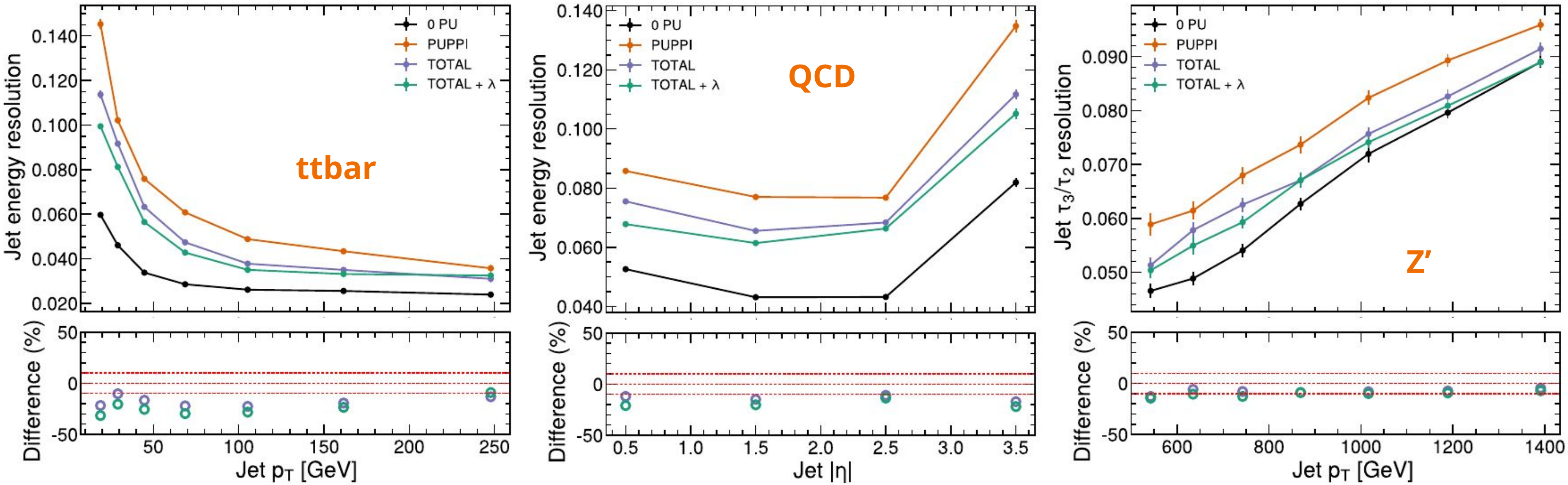


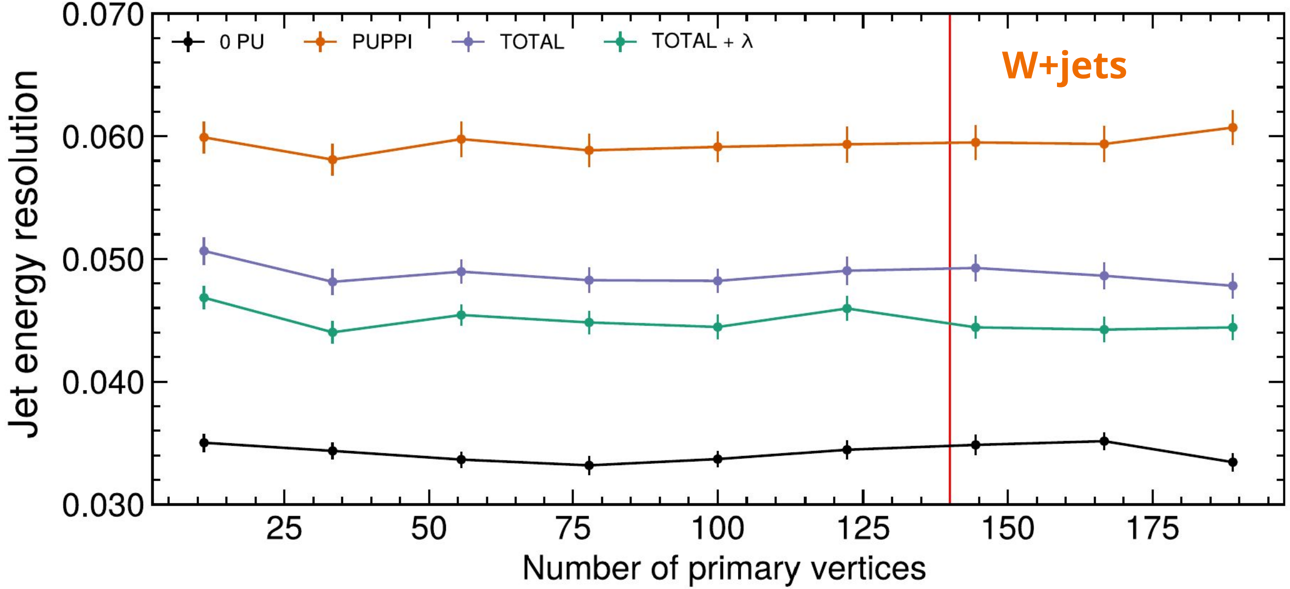

- Inclusive resolutions for variables shown in previous slide
- TOTAL resolution can be up to 25% better than PUPPI

# Differential resolutions




- Resolutions as a function of jet $p_T$ and η
- TOTAL resolution can be up to 30% better than PUPPI

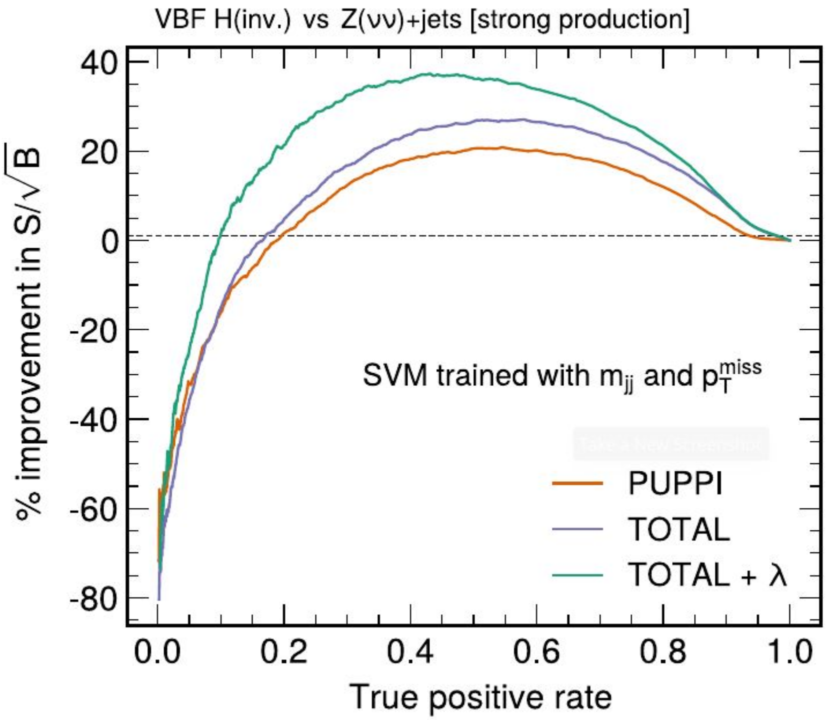
# Robustness




- Evaluate on W+jets with flat PU profile: both conditions not seen during training
- **TOTAL is robust**: correctly generalizes to new processes and PU conditions

# Searches for new physics

- **Search for enhanced Higgs to invisible rates**
- Signal: VBF H(dark matter)
- Background: VBF Z(νν)
- Train linear classifier using MET and dijet mass
- TOTAL results in ~15% improvement in S/sqrt(B) wrt PUPPI



# Conclusions

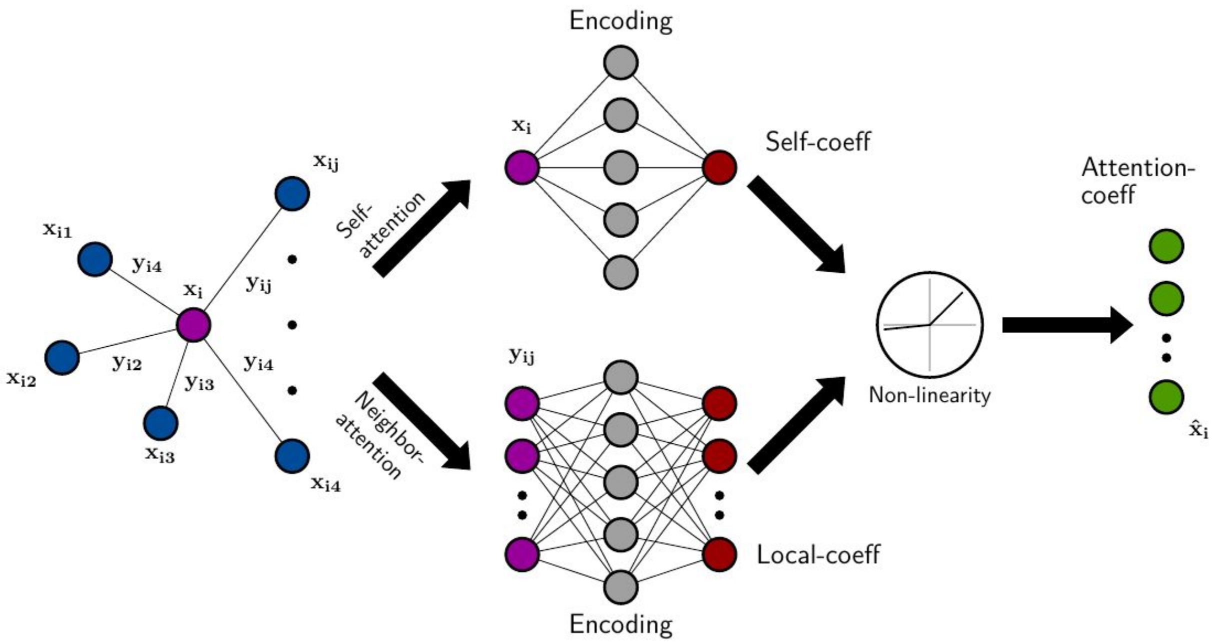
- We developed a **novel algorithm to reject PU particles** that considerably improves over the state-of-the-art
- Unlike competing algorithms, **it does not rely on per-particle truth labels**
    - Per-particle truth not available in Geant4
- **No need for truth labels: TOTAL can be ported to ATLAS/CMS**
- **TOTAL**: training optimal transport with attention learning
- **Learning** happens through optimal transport **in a self-supervised way**
- TOTAL proof of concept can be extended to a wide class of denoising problems
    - Only relies on reliable simulation of signal and noise

# Thank you!

Questions/comments?
You can check the paper PhysRevD.108.096003 and/or get in touch with me (fabio.iemmi@cern.ch)

# Backup slides

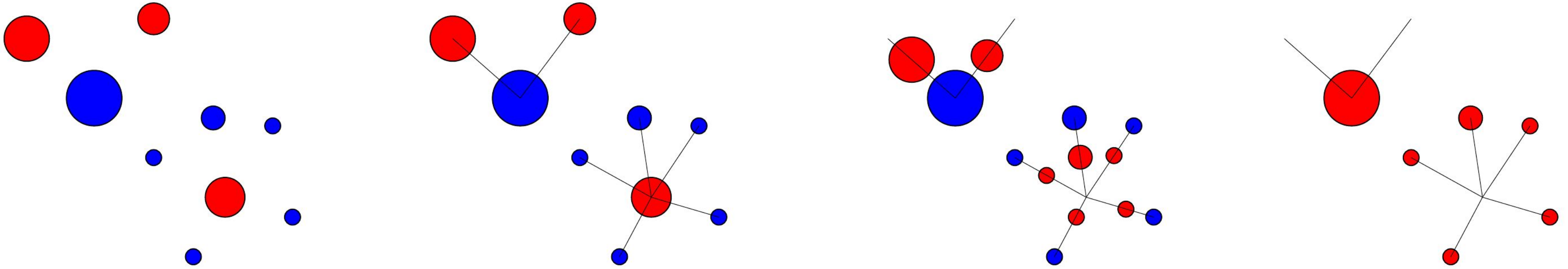
# Attention Based Cloud network – ABCNet




- ABCNet: graph network enhanced with attention mechanisms
- Want to try ABCNet? Code is online on GitLab [1]
    - Custom ABCNet layers are Keras-Functional-API compatible and ready to run

# Optimal-transport-based loss function

- Earth mover's distance (AKA Wasserstein distance): minimum work required to move **earth** to fill **holes**
- With EMD you can match 3D distributions (e.g., earth and holes)
- Abstracting this, we want to match (multi)-D particle distributions
- In particular, we want to match $\boldsymbol{\omega} * \mathbf{X}_{\mathrm{PU}}$ and $\mathbf{X}_{\text{no-PU}}$



# Optimal-transport-based loss function

- EMD can only be computed (approximately) in 3D; want to match multi-D distributions instead
- Solution: OT has a closed mathematical form in 1D
- After some math, we can show that 1D OT can be reduced to a sorting problem
    - Fast and trivial to solve
- Take input n-D space and project (“slice”) it onto n-D, unit-radius sphere
- Solve 1D OT for the projection: **Sliced Wasserstein Distance** (SWD)