SLAC NATIONAL ACCELERATOR LABORATORY

NEVIS LABORATORIES COLUMBIA UNIVERSITY

Semi-Supervised Permutation Invariant Particle-Level Anomaly Detection

Gabriel Matos, Elena Busch, Ki Ryeong Park, Julia Gonski
November 7th, 2024
Columbia University
ML4Jets 2024 @ LPNHE, Paris

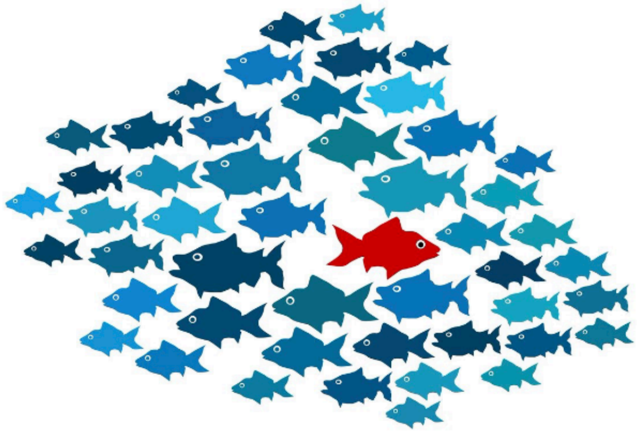
Outline

1. Overview of Anomaly Detection
2. Particle Flow Network (PFN)
3. Anomaly Detection on Particle Flow Latent Space (ANTELOPE)
4. Application to LHC Olympics Dataset
5. Performance Results

Anomaly Detection

- Typically, collider searches follow a recipe
  1. Pick a model for some signature
  2. Determine relevant parameters for the model (mass ranges, etc.)
  3. Design selections on your observables to increase signal over background
- **Problem:** way too many models/signatures to develop independent searches for!
- How can we probe across many signatures w/out heavily relying on particular models?

Anomaly Detection (AD): train a machine learning model to detect anomalous features of a dataset inconsistent with a background only model



Unsupervised
Detect outliers by training directly on data

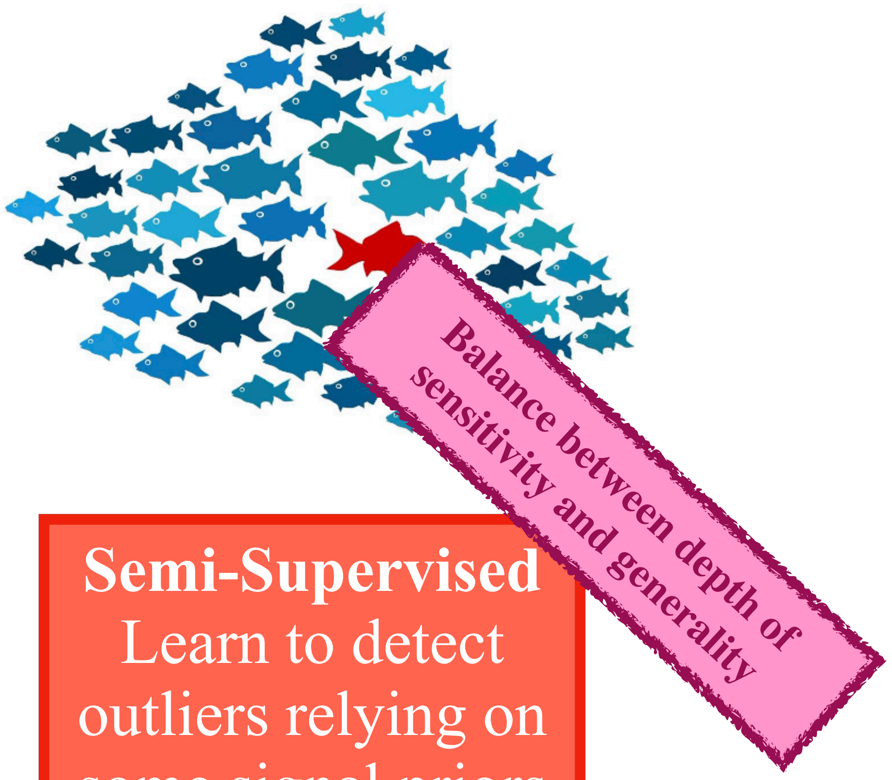
Weakly-Supervised
Detect outliers with combination of labeled and unlabeled data

Semi-Supervised
Learn to detect outliers relying on some signal priors

Anomaly Detection

- Typically, collider searches follow a recipe
  1. Pick a model for some signature
  2. Determine relevant parameters for the model (mass ranges, etc.)
  3. Design selections on your observables to increase signal over background
- **Problem:** way too many models/signatures to develop independent searches for!
- How can we probe across many signatures w/out heavily relying on particular models?

Anomaly Detection (AD): train a machine learning model to detect anomalous features of a dataset inconsistent with a background only model



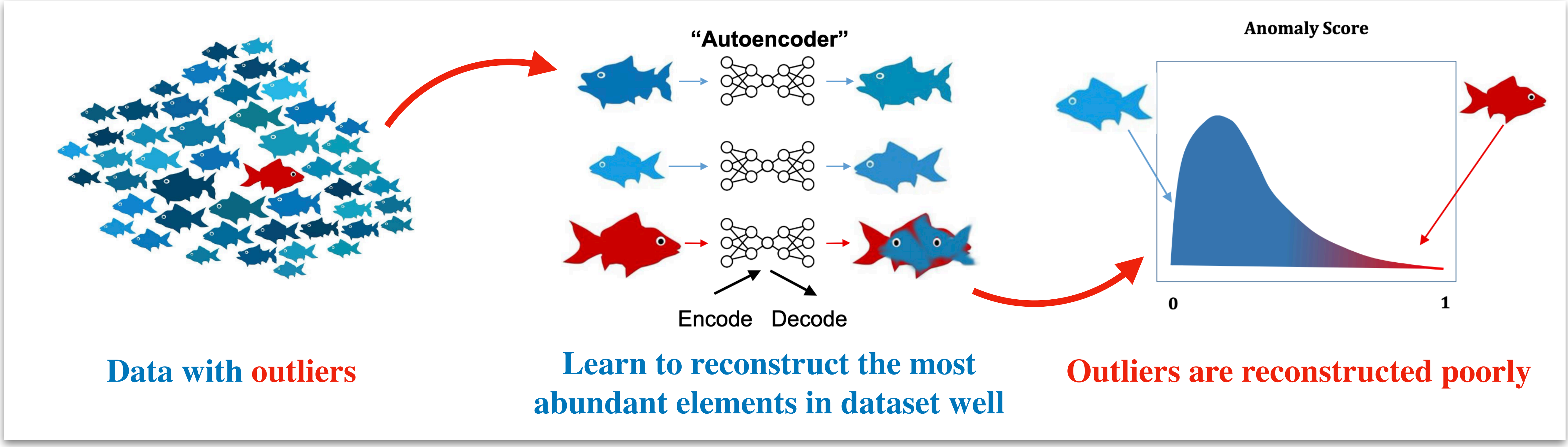
Unsupervised
Detect outliers by training directly on data

Weakly-Supervised
Detect outliers with combination of labeled and unlabeled data

Semi-Supervised
Learn to detect outliers relying on some signal priors

Challenges of AD Models in Collider Physics

- A common model for AD in collider data is the autoencoder (AE) / variational AE (VAE)



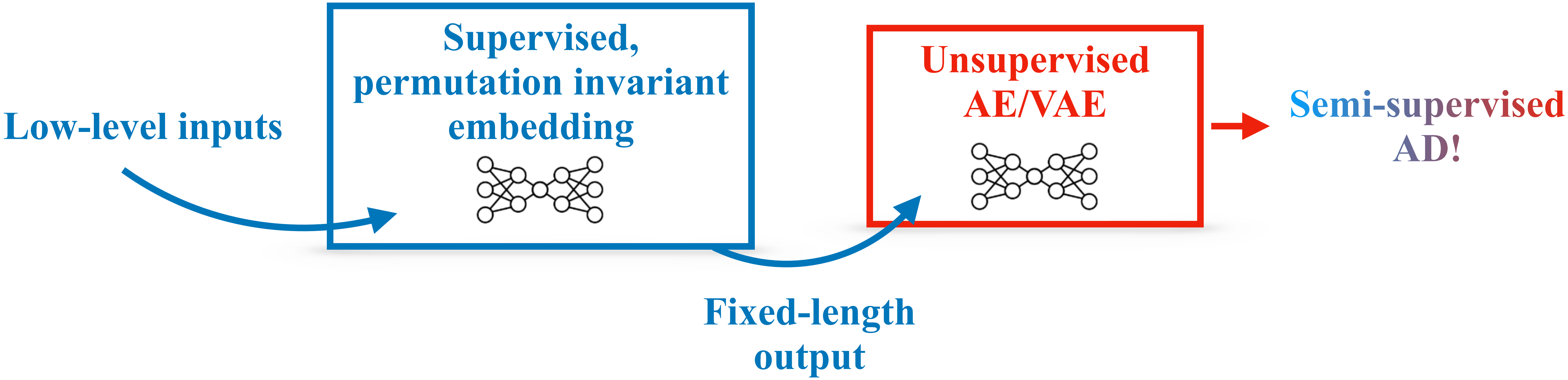
- **Problem:** AEs require a fixed length input → hard to provide low-level inputs
  - Collider events, at a fundamental level, can be described as an unordered collection of particles represented by low-level detector objects (tracks, calorimeter clusters, etc.)
    - This modeling is inherently variable in length and permutation invariant
  - This can be accommodated in AEs by zero-padding the inputs or via an RNN that models inputs as a sequence, but introduces unnatural ordering/padding to the inputs

Permutation Invariant Particle-Level AD

- There are subtle correlations in low-level detector signatures that we should exploit…

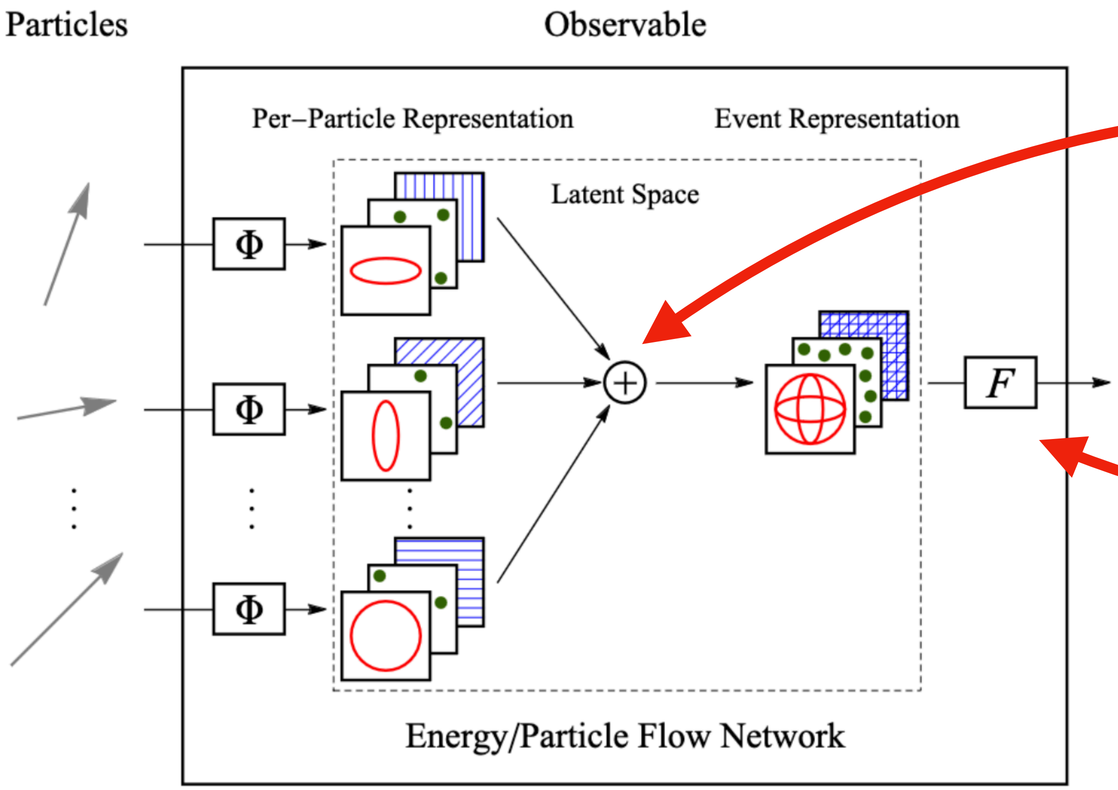
How can we gain the benefits of a semi-supervised approach for AD while using low-level detector objects and maintaining the natural permutation invariance of collider events?

- **We can use a supervised classifier to create an intelligent embedding of data**
  - This embedding should be fixed-length, preserve permutation invariance, and select salient features of our data that can be exploited in an AD task
  - We then use this embedding as input to an AE/VAE architecture



Particle Flow Network (1)

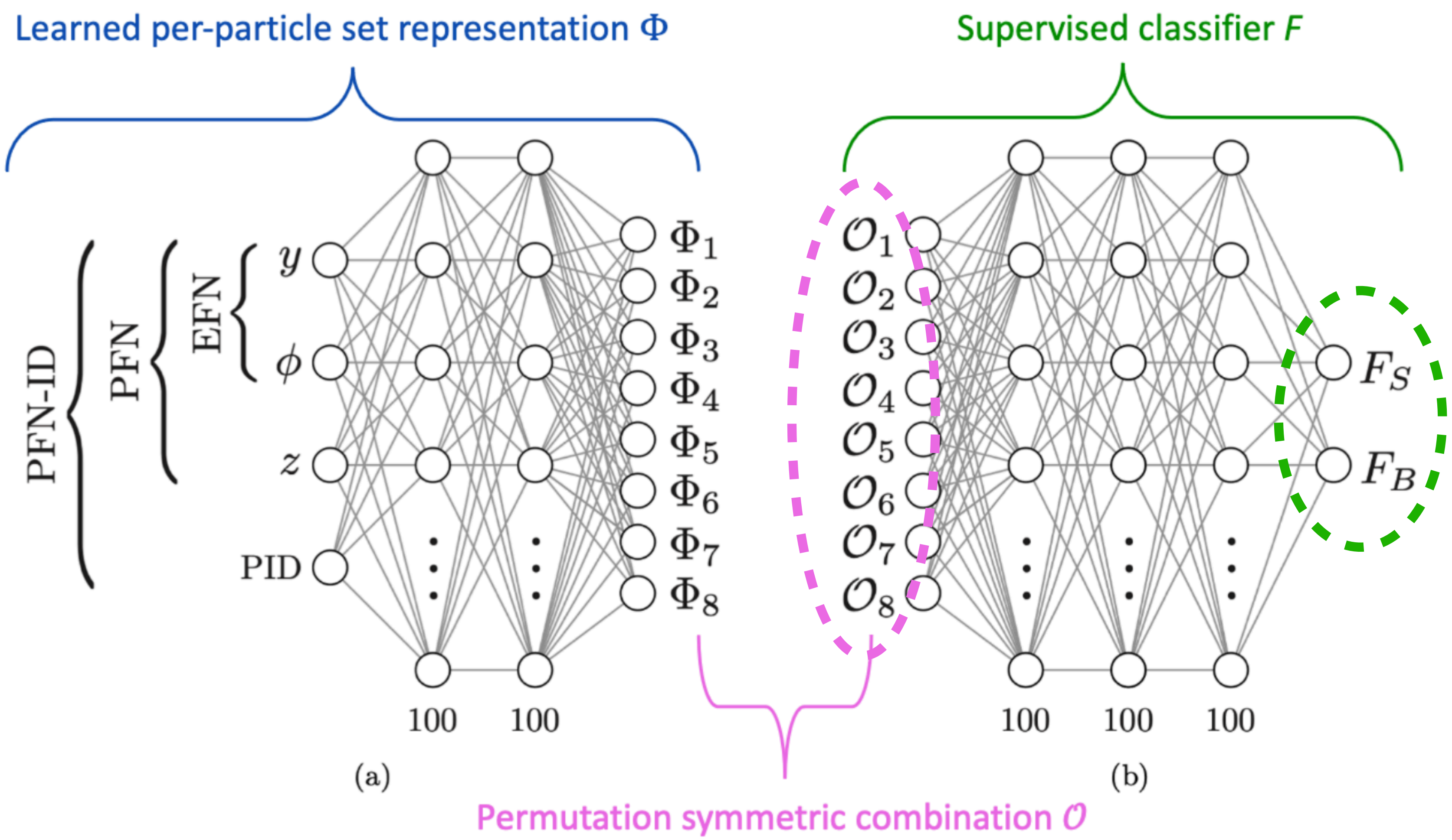
- One way to achieve this embedding is via a **Particle Flow Network** (PFN) architecture
- The PFN is a **supervised** classifier based on the the Deep Sets framework for point clouds
  - The network takes in an arbitrary number of particles with features that are encoded into a latent space, per-particle representation by a set of learned functions Φ_a
  - These per-particle representations are combined into event level observables $\mathcal{O}_a$ that are **inherently permutation invariant** by **summing** over all input particles



$$\mathcal{O}_a = \sum_i \Phi_a(y_i, \phi_i, z_i, \mathrm{PID}_i)$$

- These observables are fed to a network F that is optimized for binary classification

Particle Flow Network (2)



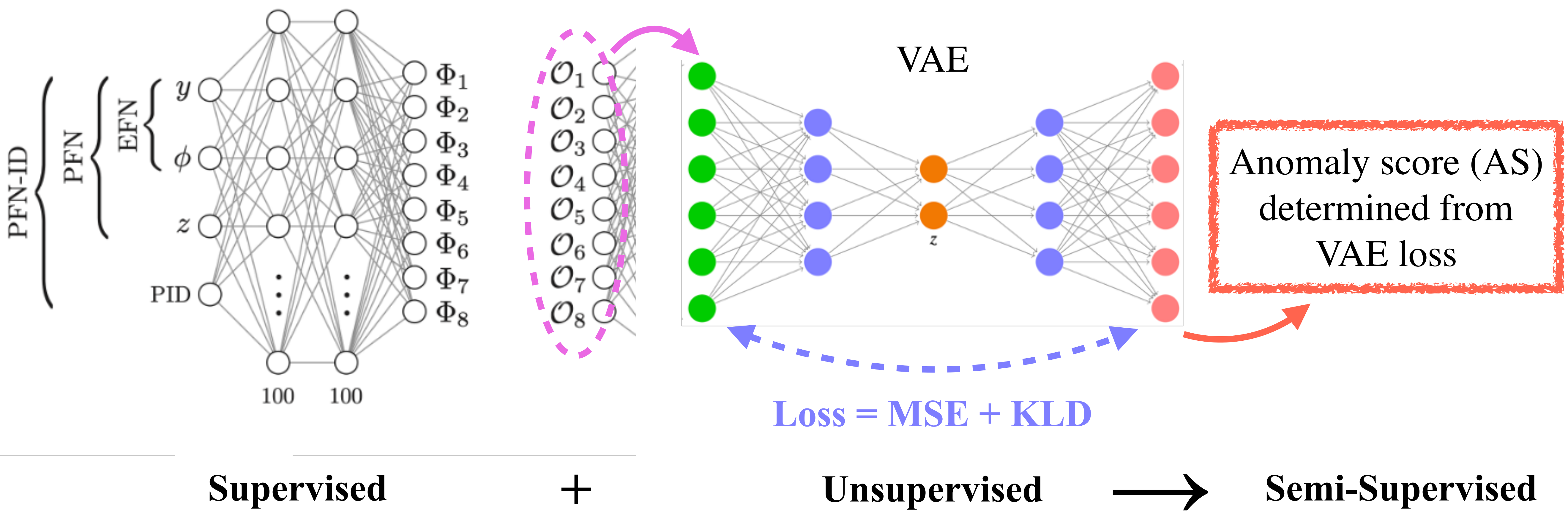
Supervised PFN Score

$$\mathcal{O}_a = \sum_i \Phi_a(y_i, \phi_i, z_i, \mathrm{PID}_i)$$

The PFN solves the problem of representing variable length unordered data by creating a **symmetric observables $\mathcal{O}_a$** of an arbitrary number of particles with d features

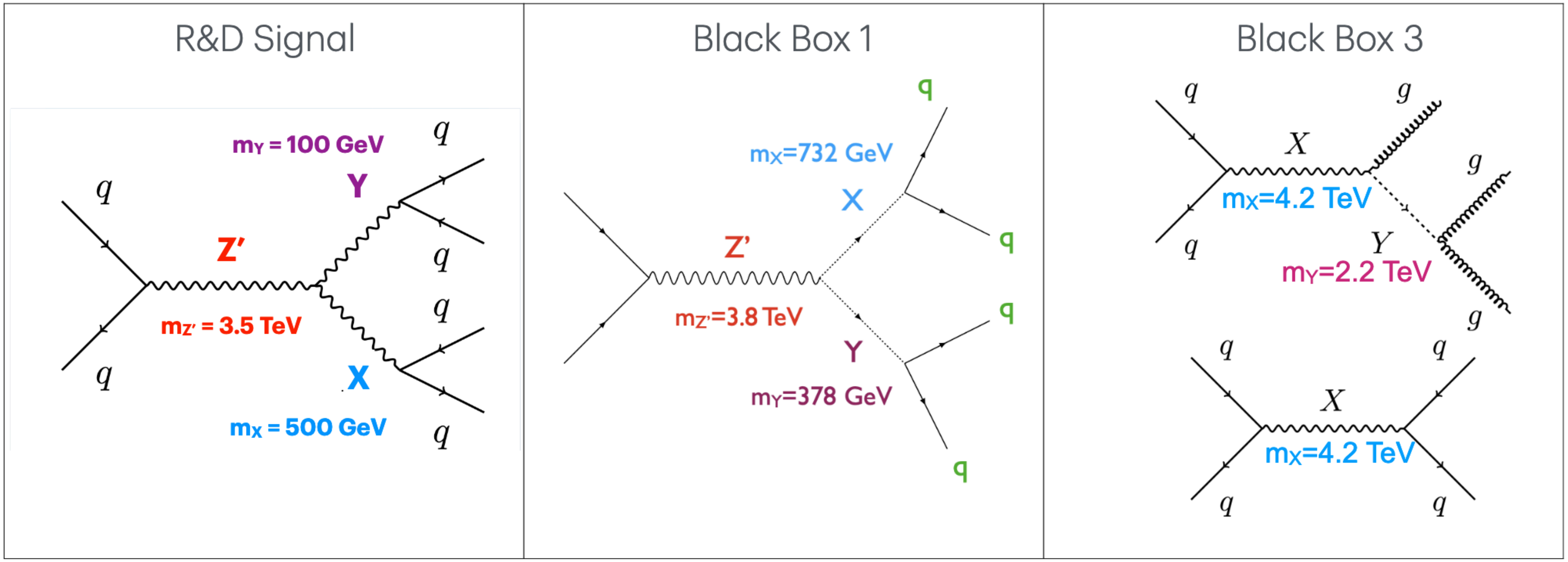
Anomaly Detection on Particle Flow Latent Space

- The PFN embedding encodes key discriminating features learned by the supervised model according to the signal & background priors
  - These can be exploited to broaden the sensitivity beyond the trained signal model by AD while keeping a low-level, symmetric input modeling
- We use the PFN latent variables to design a novel architecture titled **ANTELOPE** that trains an **unsupervised** VAE on the $\mathcal{O}_a$ and performs AD on the PFN's latent space



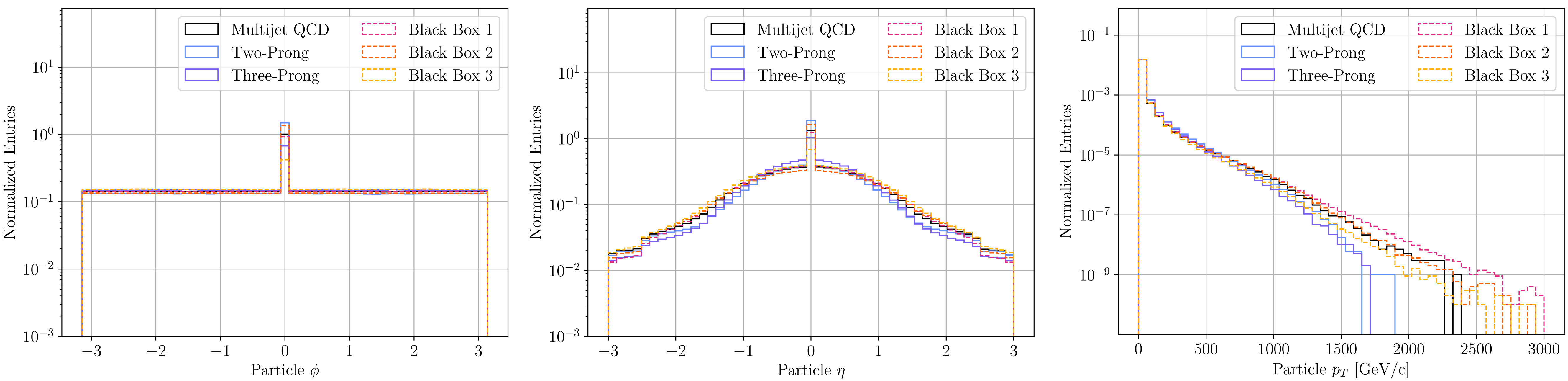
LHC Olympics Dataset

- As proof of concept, we've applied the ANTELOPE model to the LHC Olympics dataset
- The dataset consists of 3 R&D and 3 black box samples
  - Each event described as a set of up to 700 (massless) particle four vectors (p_T, η, ϕ)
  - **R&D:** QCD multijet, 2-prong, and 3-prong
  - **Black boxes:** BB1 2-prong BB2 QCD multijet (No signal) BB3 Resonance → Dijet/Trijet



Inputs & Pre-Processing

- We keep the 160 highest p_T particles as inputs and use their p_T, η, ϕ as training features
  - Each feature is normalized via Min-Max scaling between 0 and 1
  - Events with less than 160 particles are zero-padded
    - By design, the padding does not impact the performance of the PFN/VAE



# of Events	R&D QCD	R&D 2-prong	R&D 3-prong	BB1	BB2	BB3
Signal	N/A	100k	100k	834	N/A	1200 dijet / 2000 trijet
Total	**1M**	**100k**	**100k**	**1M**	**1M**	**1M**

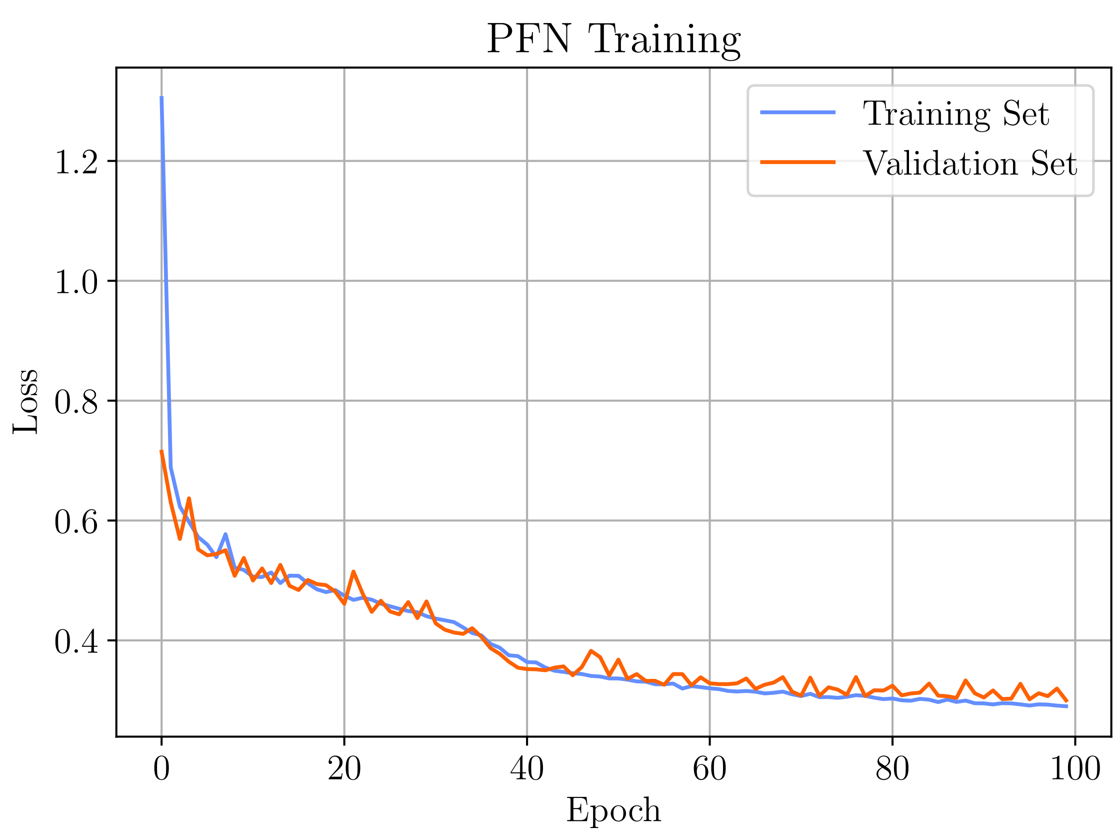
Training

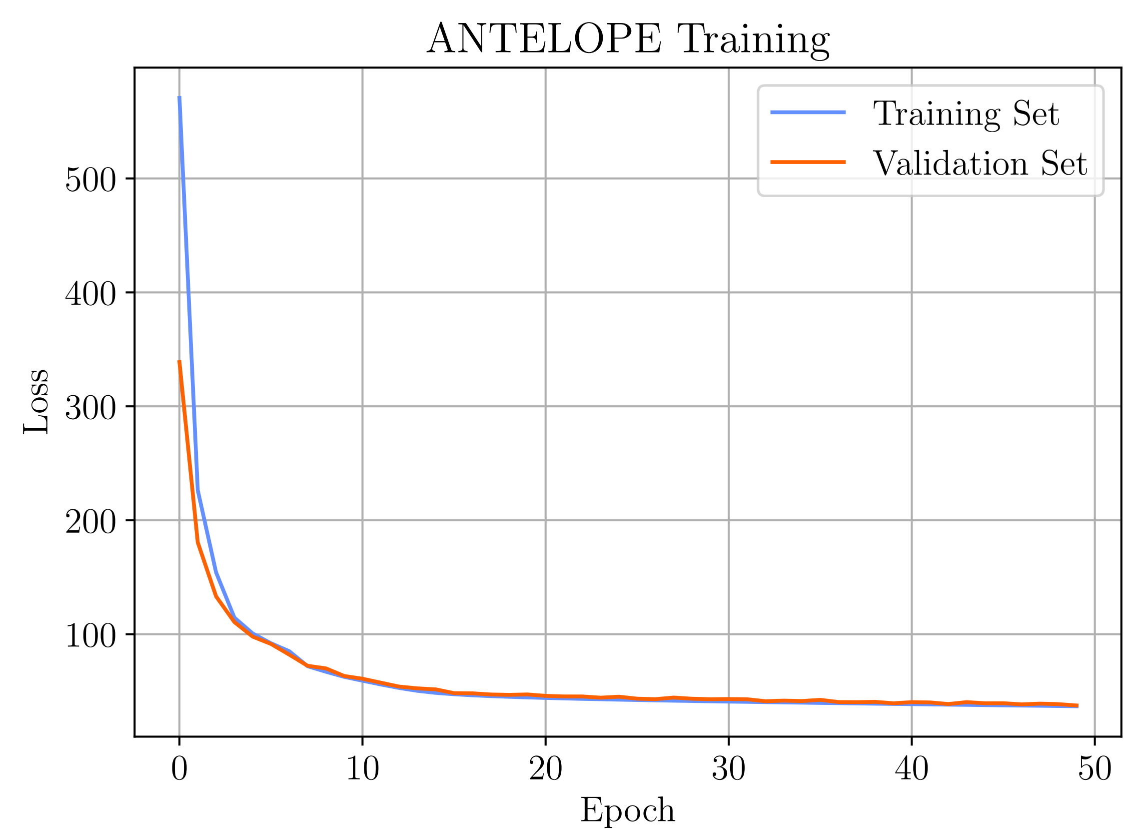
- The PFN used to generate the embedding for ANTELOPE is **trained on the 2-prong and QCD R&D samples**, with 80k events used from training and 20k for validation
- The VAE is trained on an orthogonal 80k events from QCD R&D meant to represent data, where these events are first encoded into the PFN's pre-trained $\mathcal{O}_a$ basis before training

1 **Train supervised PFN with 2-prong and QCD**

2 **Embed events into PFN's latent space**

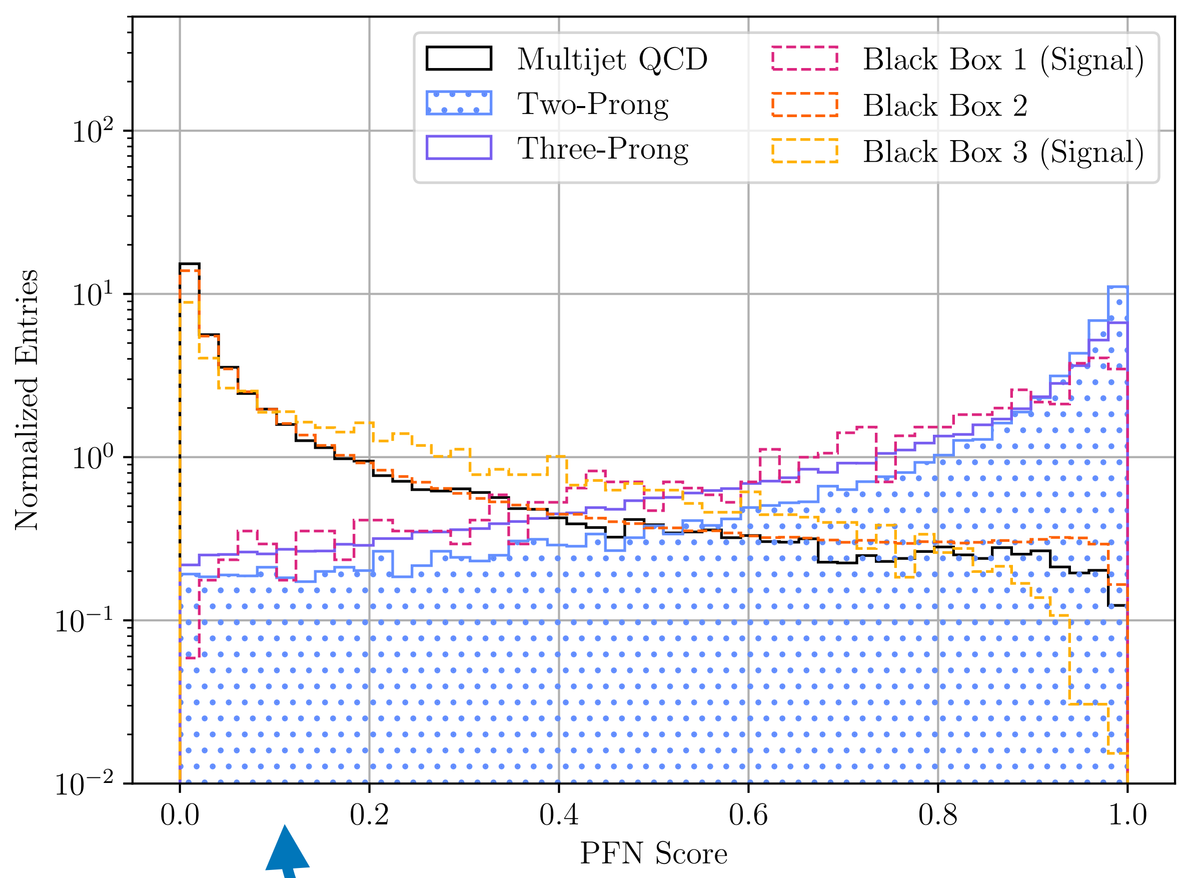
3 **Train VAE with orthogonal QCD events**



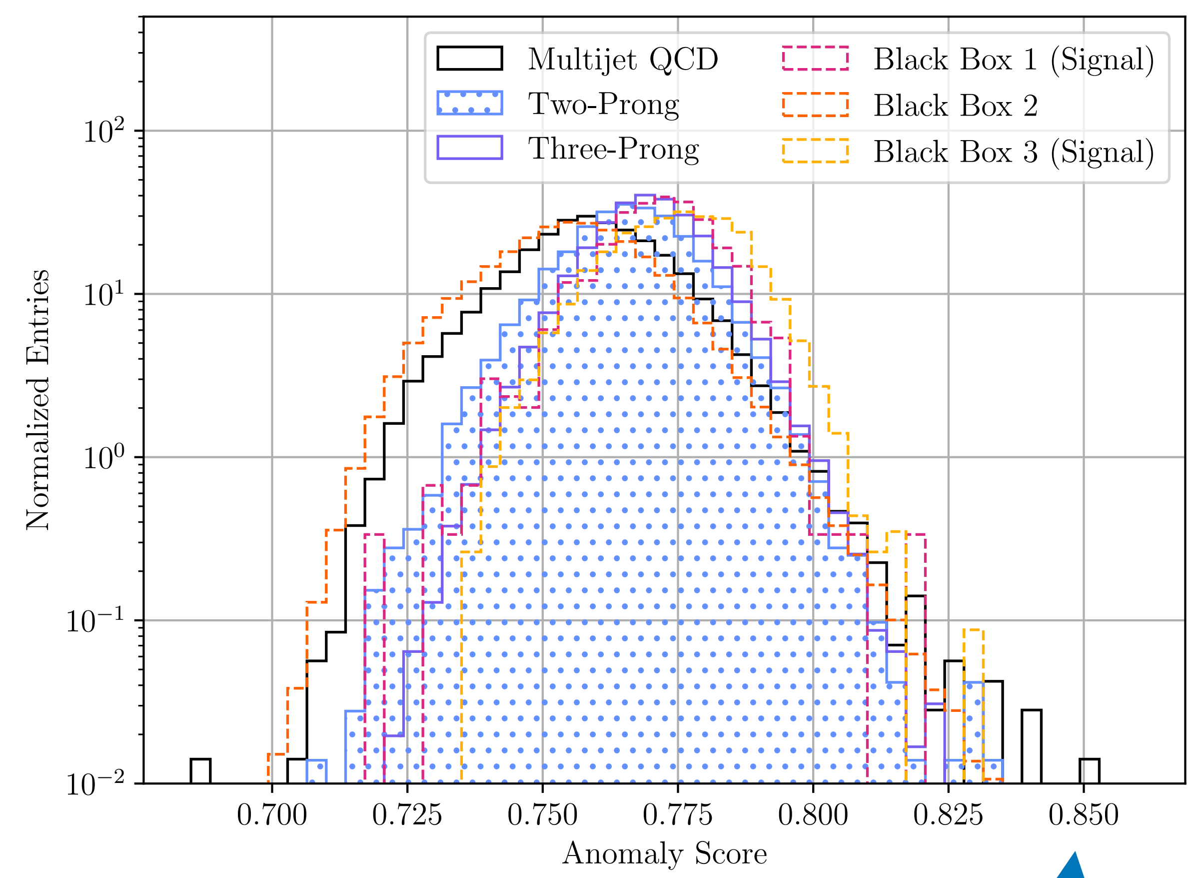


ANTELOPE Performance: Anomaly Score

- We assess ANTELOPE's performance & model independence by comparing it to the standalone, supervised PFN classifier
  - We evaluate both ANTELOPE and the PFN on the validation samples from training, the 3-prong R&D dataset, as well as the *signals* in the black boxes

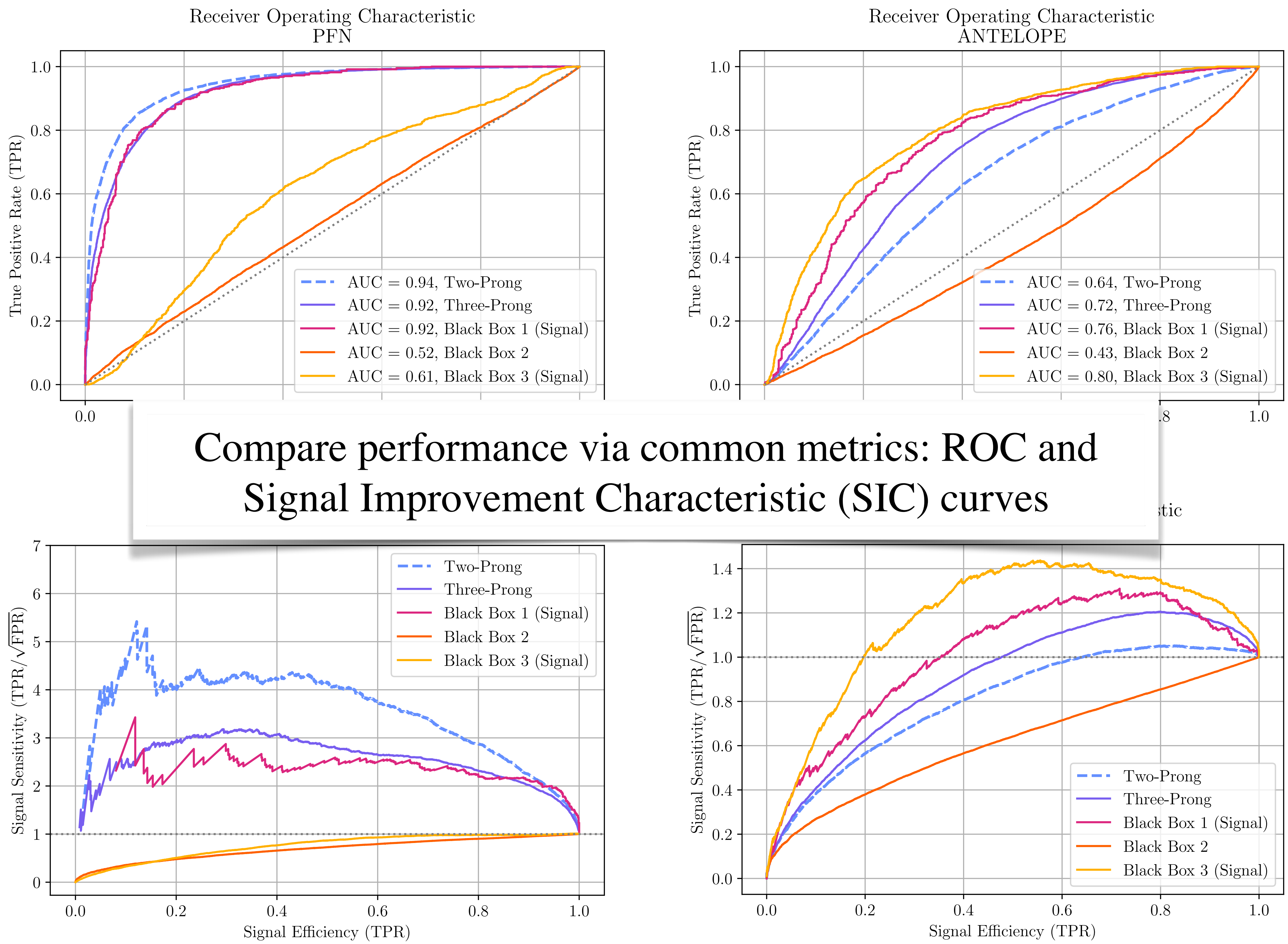


Softmaxed output of PFN classifier



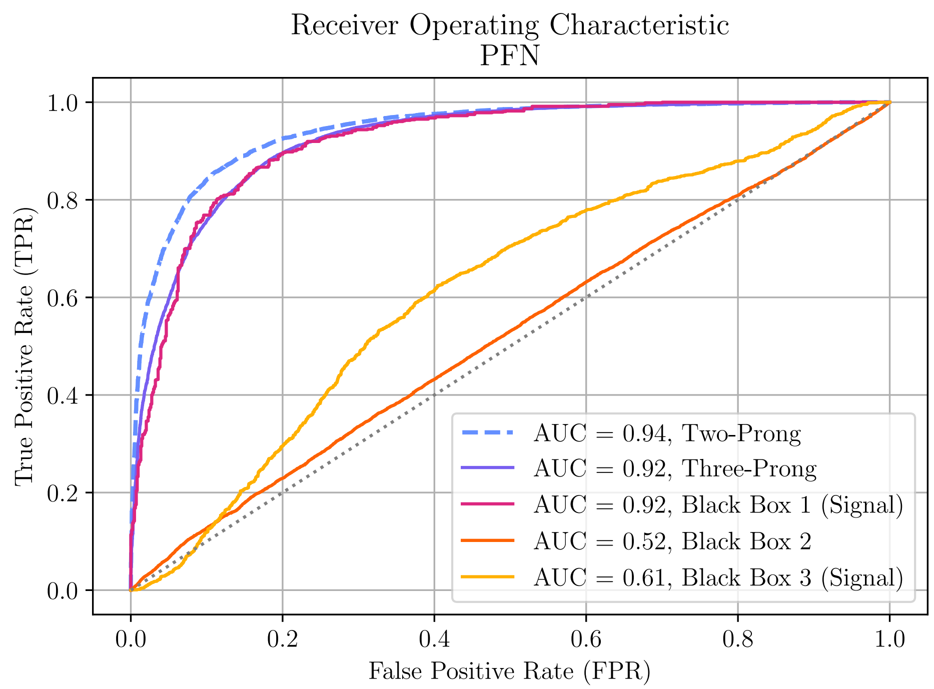
Log of ANTELOPE's output loss transformed by a sigmoid

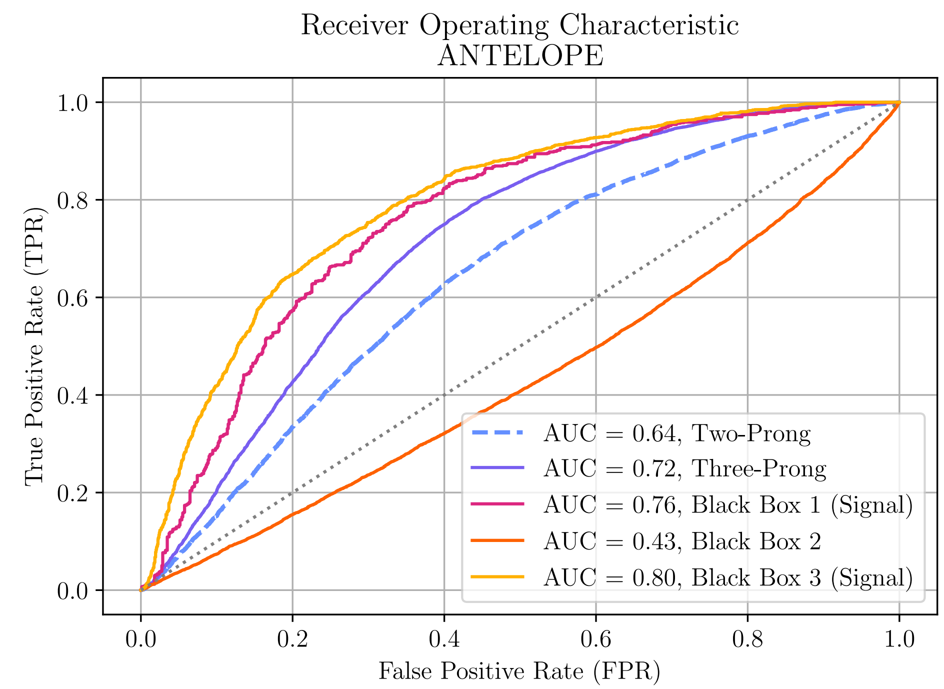
ANTELOPE Performance: ROC & SIC Curves

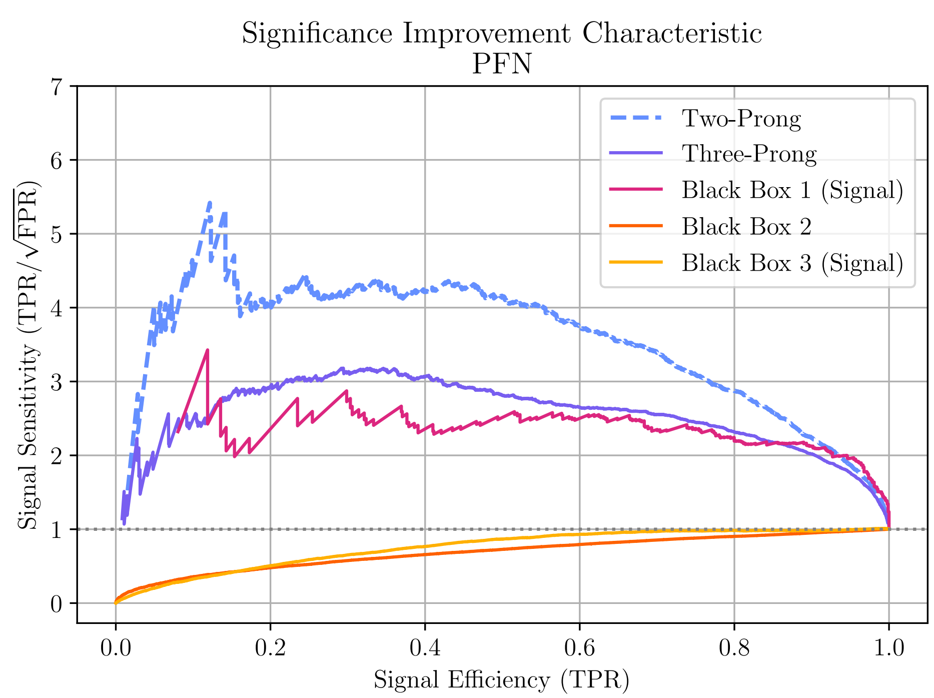


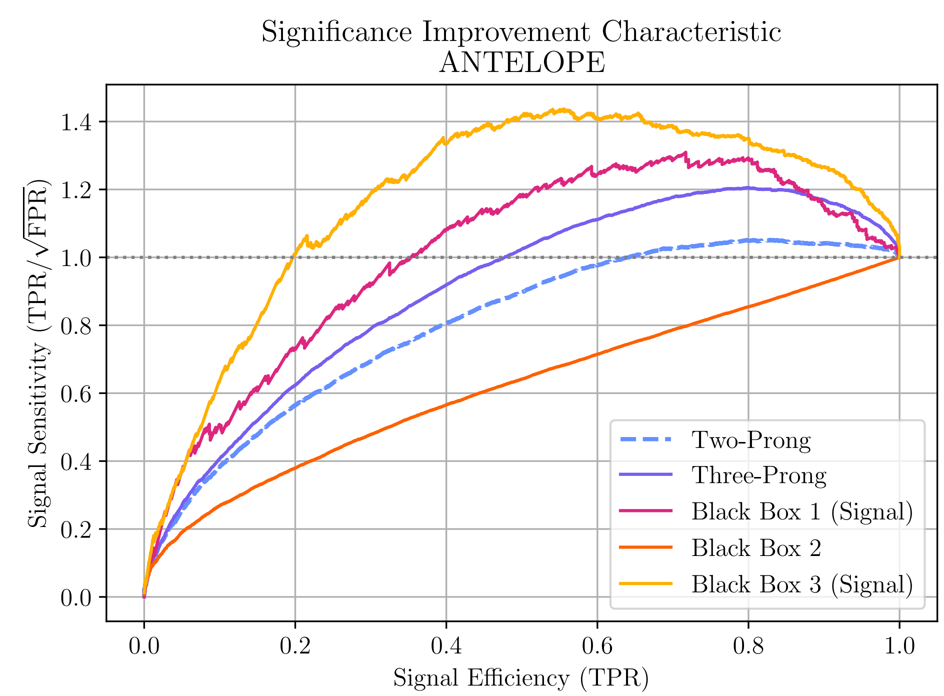
Compare performance via common metrics: ROC and Signal Improvement Characteristic (SIC) curves

ANTELOPE Performance: ROC & SIC Curves

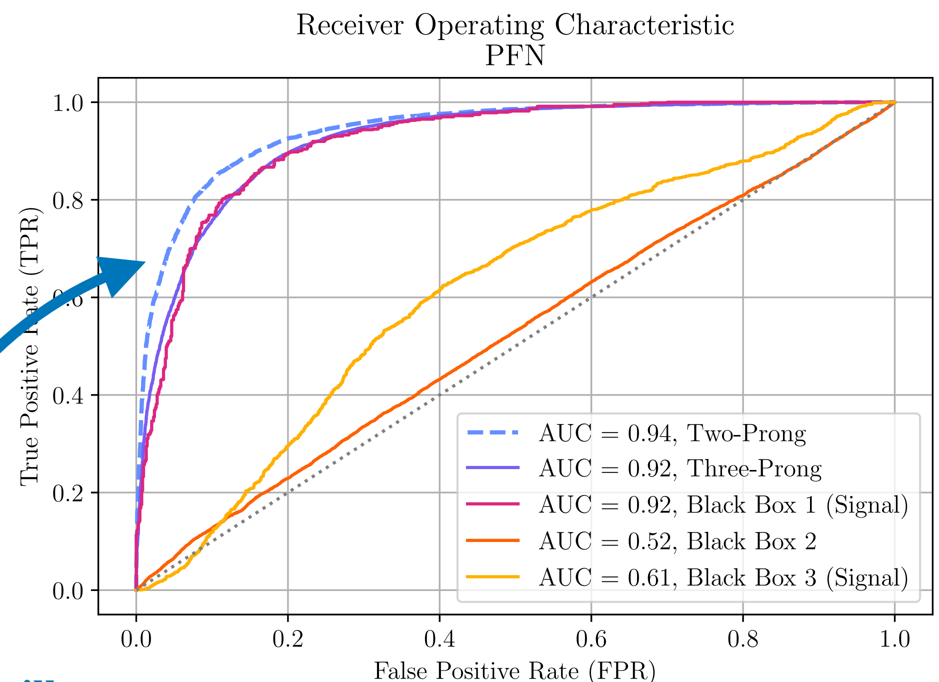


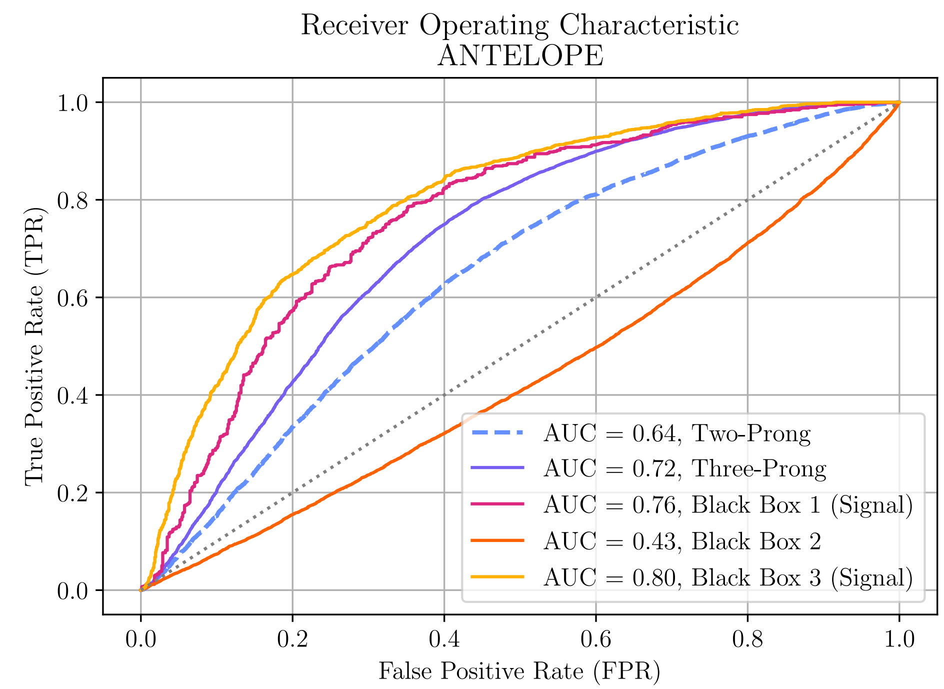




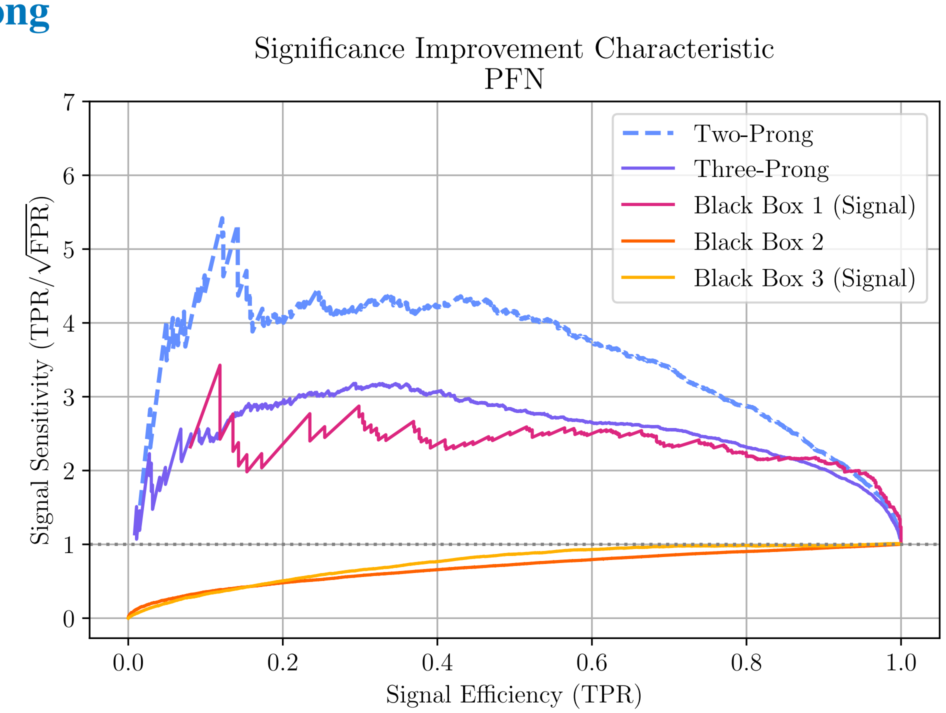


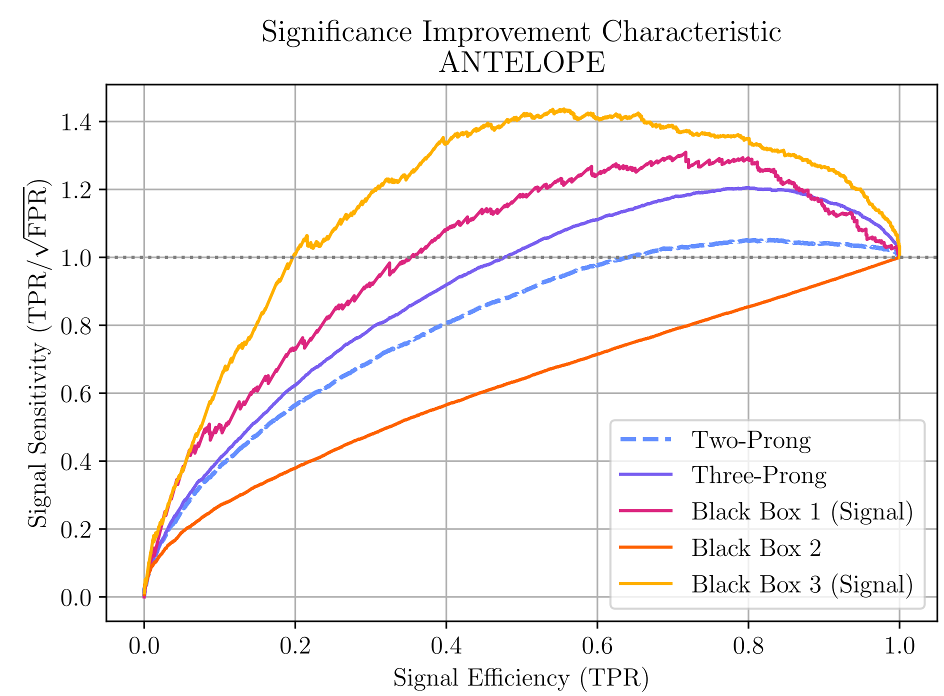
ANTELOPE Performance: ROC & SIC Curves



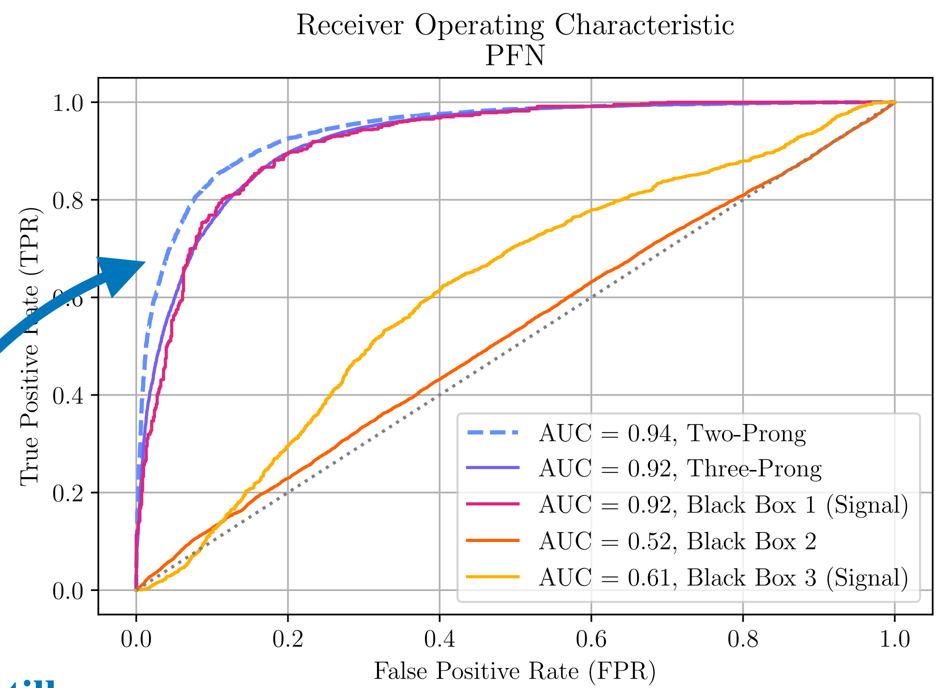


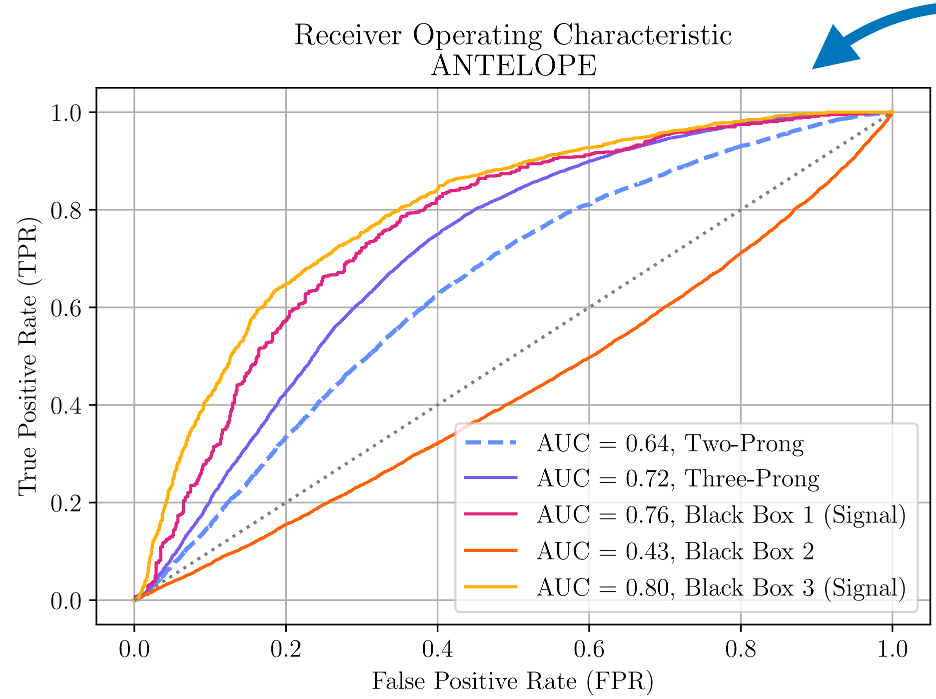
PFN sensitivity still best for 2/3-prong





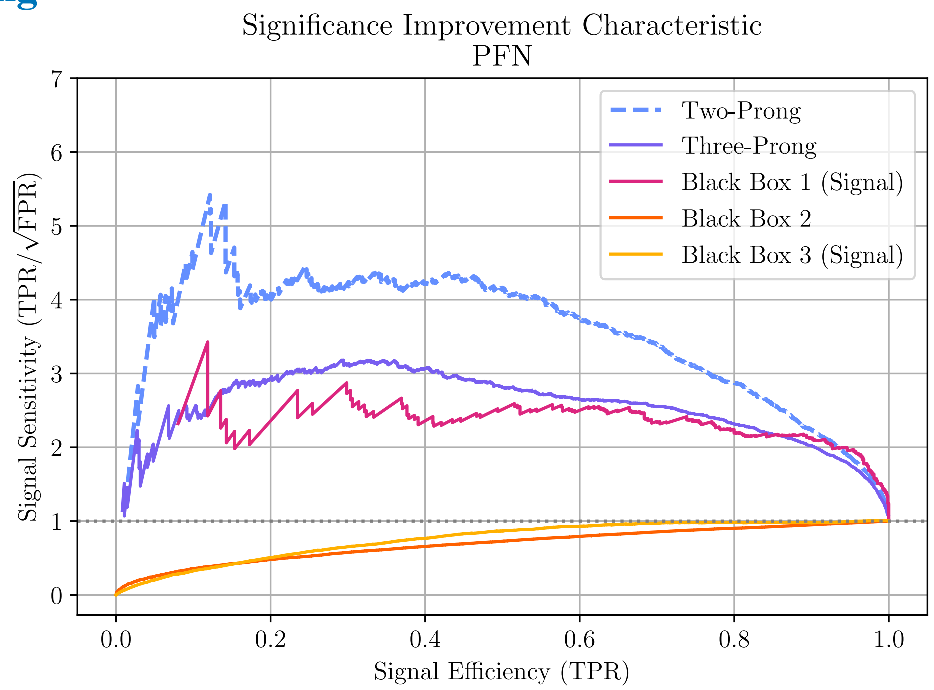
ANTELOPE Performance: ROC & SIC Curves

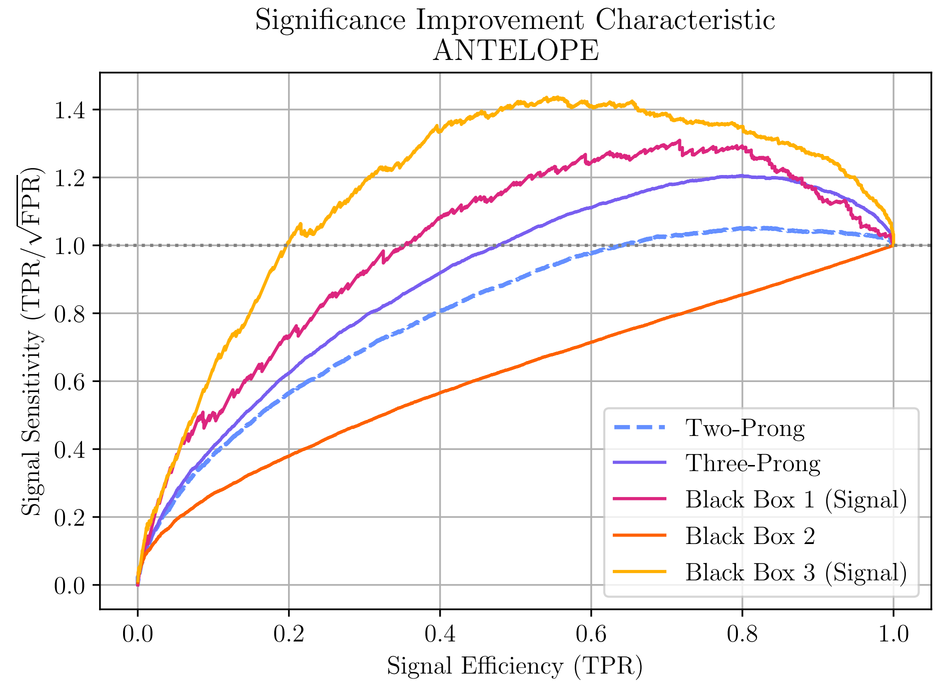




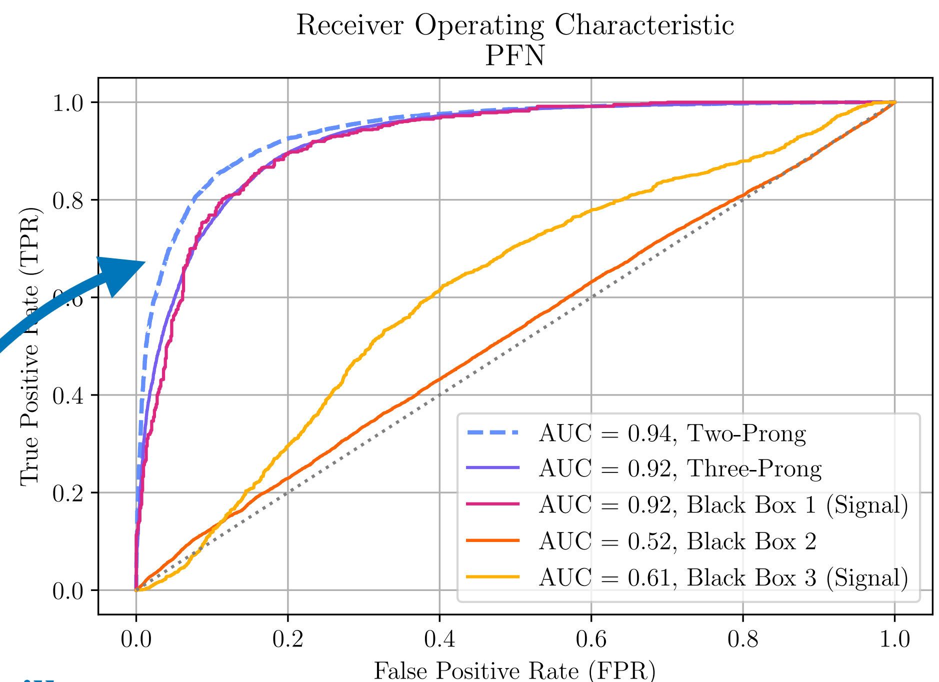
ANTELOPE generalized to all signals

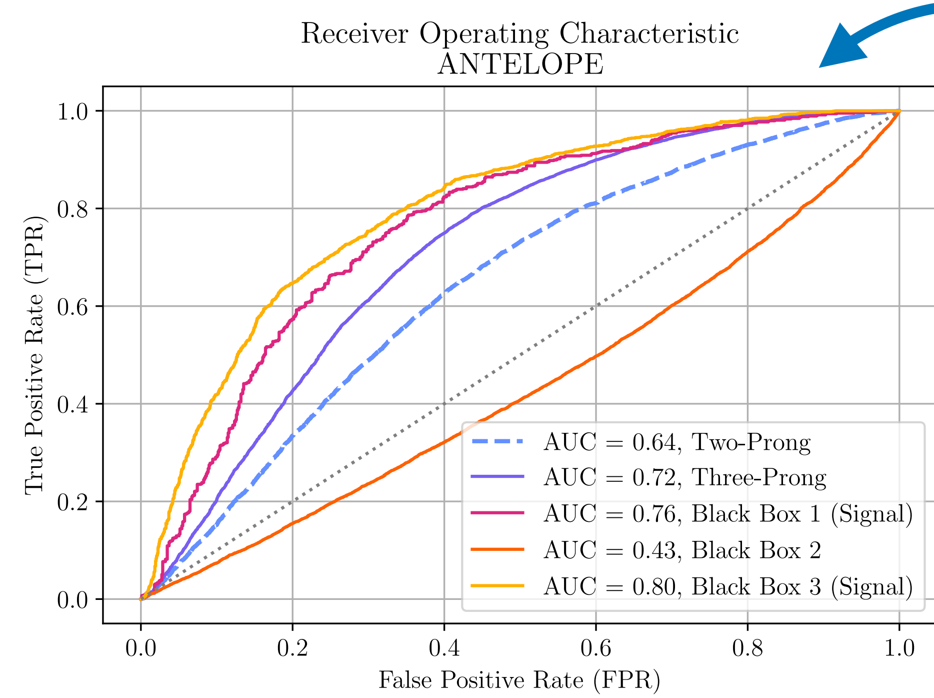
PFN sensitivity still best for 2/3-prong





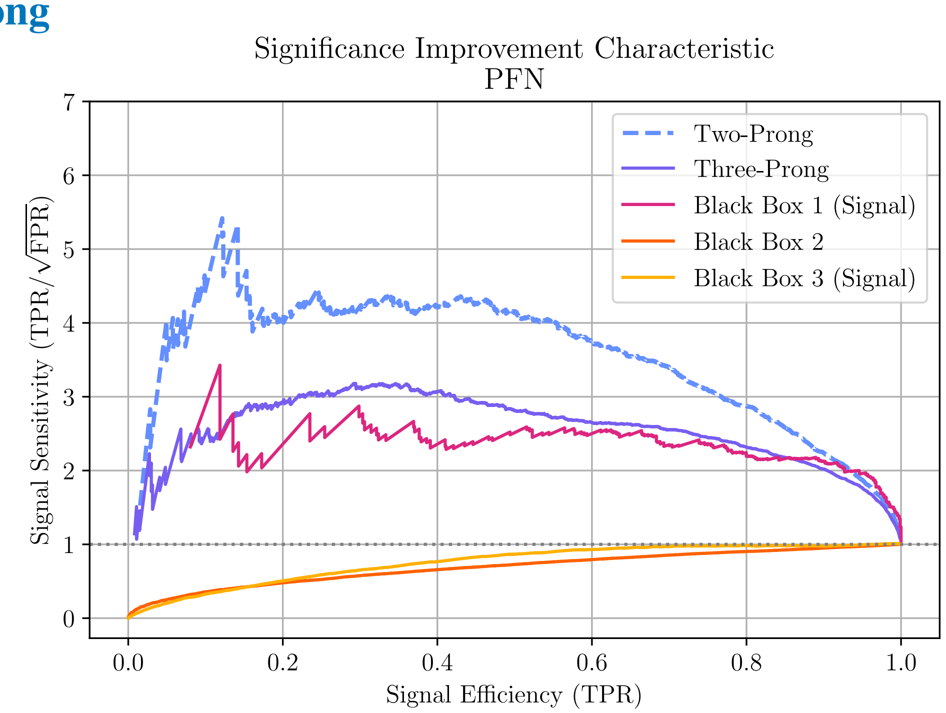
ANTELOPE Performance: ROC & SIC Curves

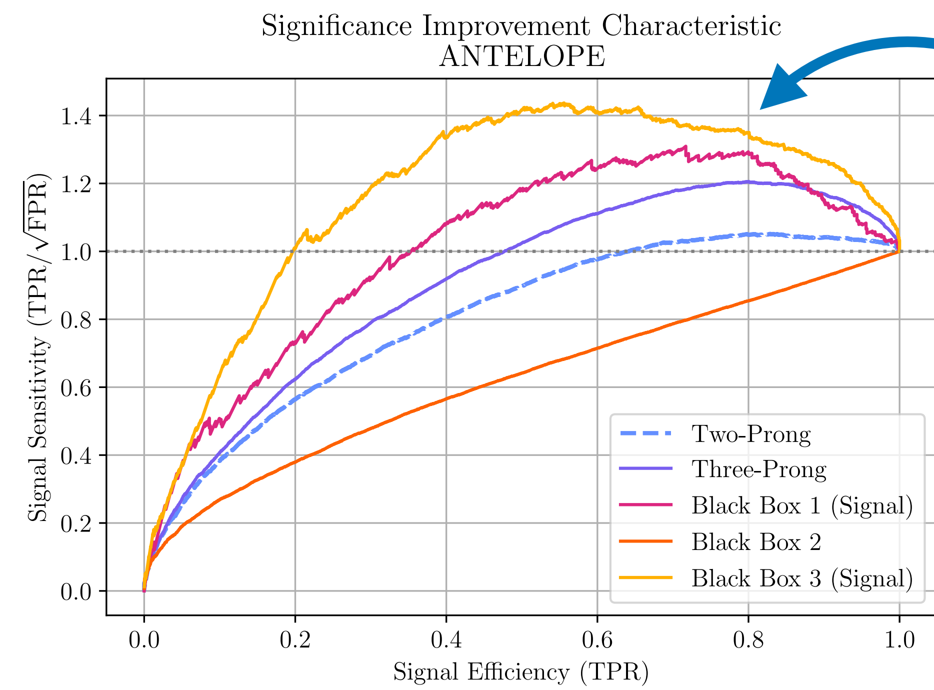




ANTELOPE generalized to all signals

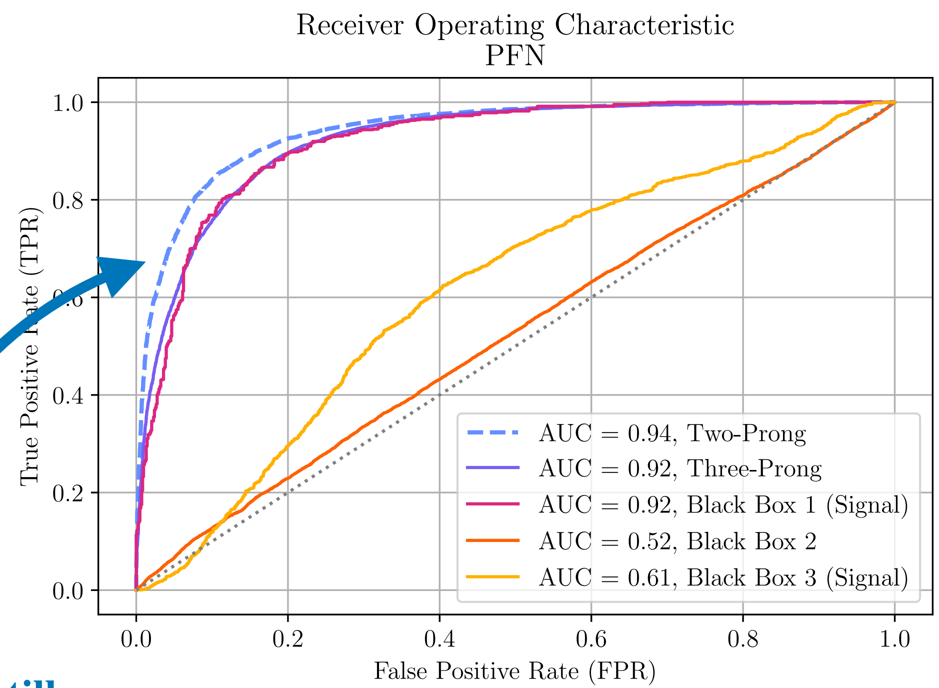
PFN sensitivity still best for 2/3-prong

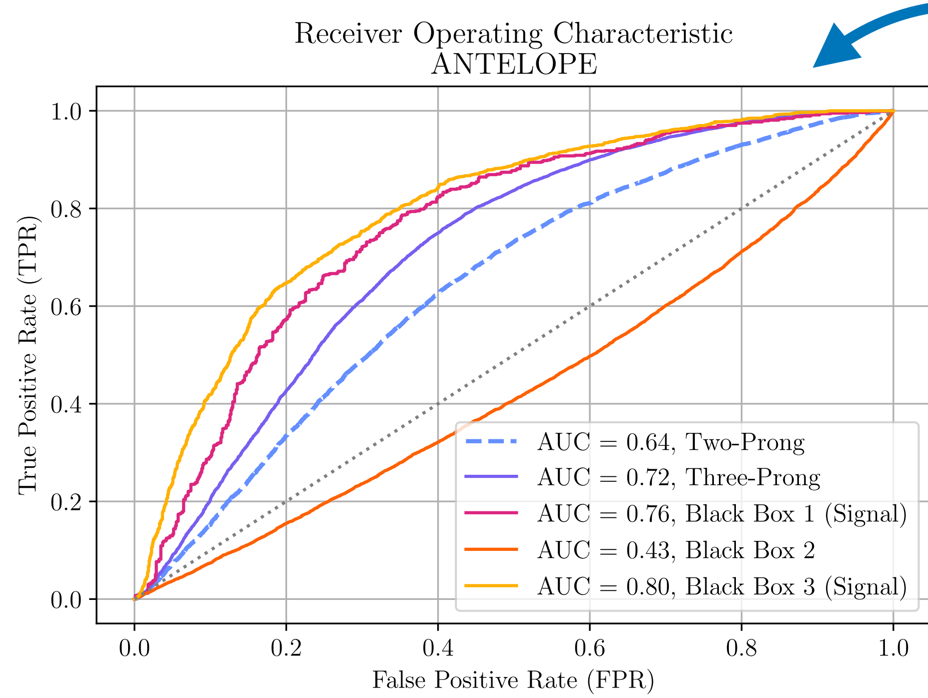




ANTELOPE shows best sensitivity for BB3

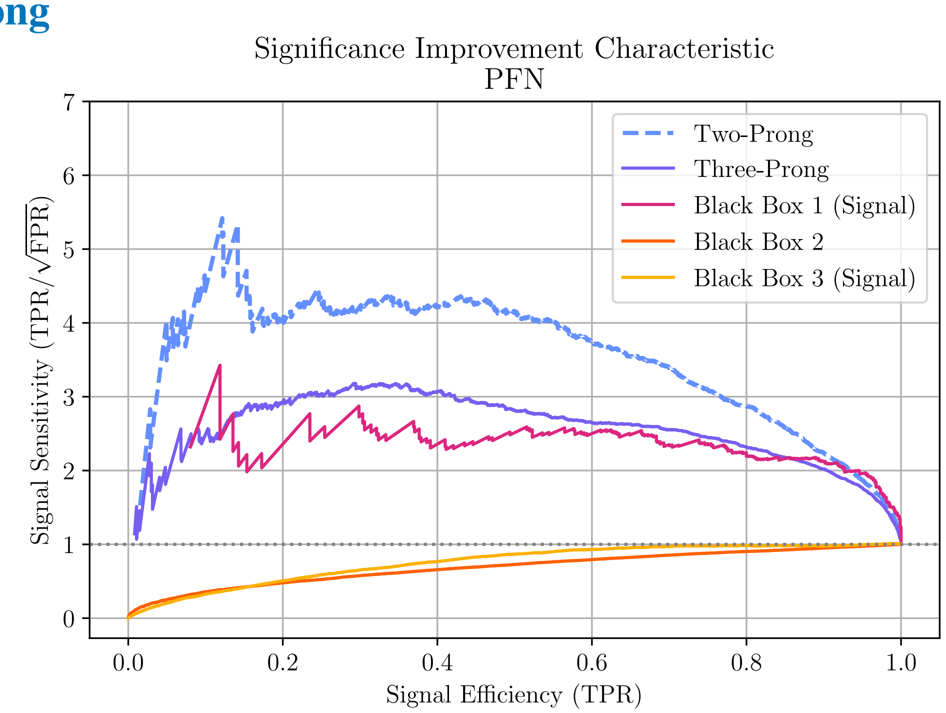
ANTELOPE Performance: ROC & SIC Curves

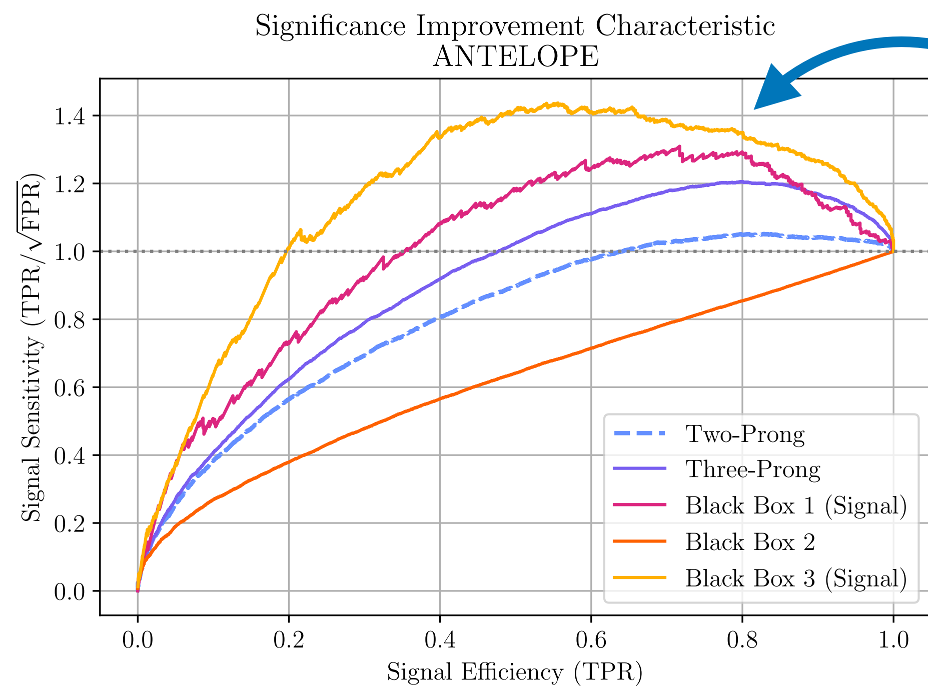




ANTELOPE generalized to all signals

PFN sensitivity still best for 2/3-prong





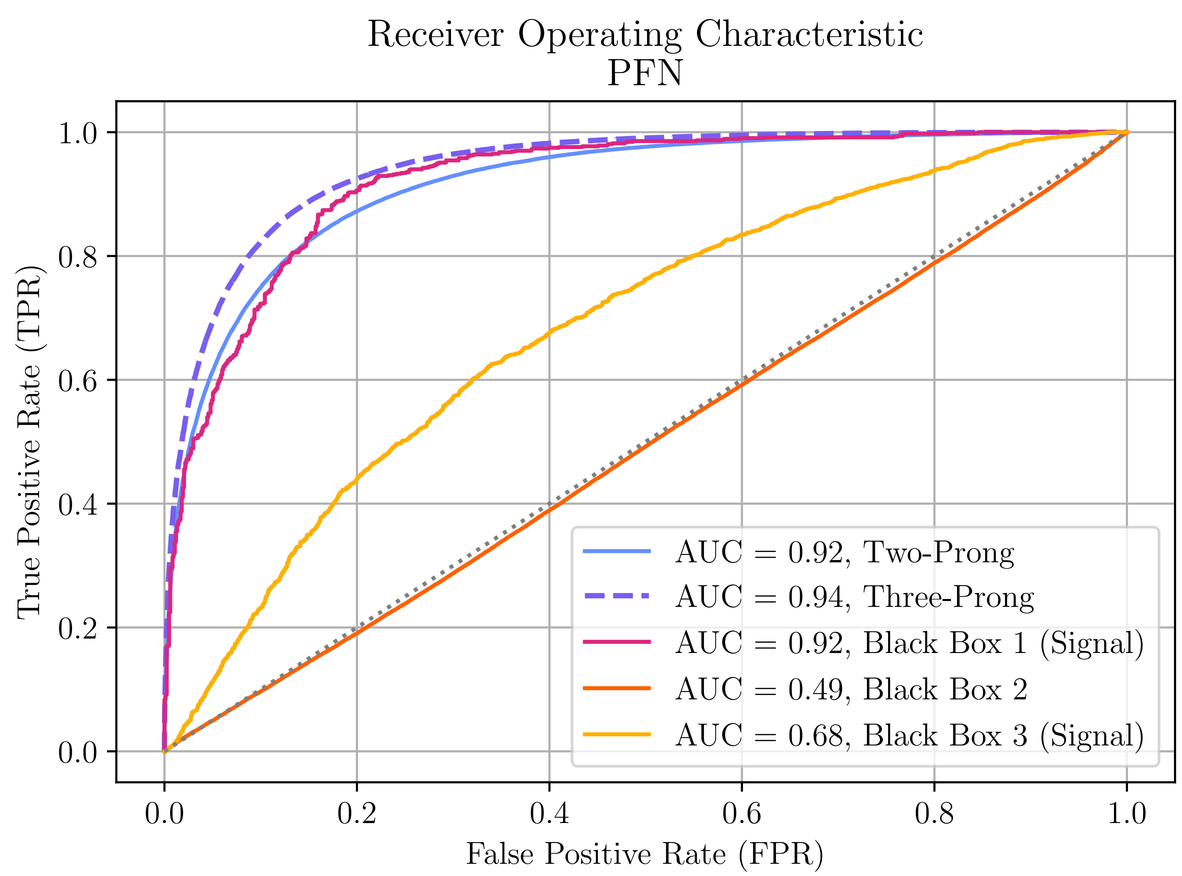
ANTELOPE shows best sensitivity for BB3

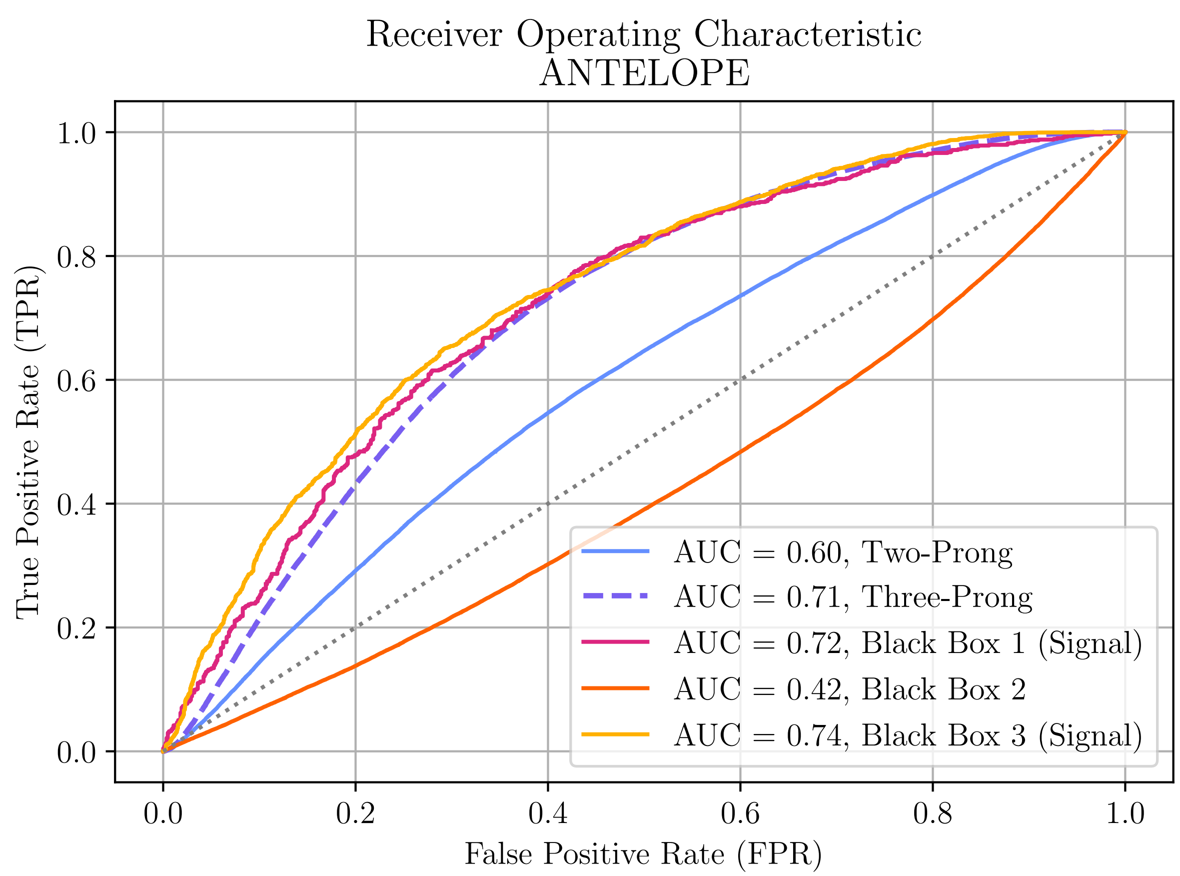
In LHCO, no method was sensitive to BB3!

ANTELOPE Performance: Model Robustness

- We verified the robustness of the ANTELOPE approach by repeating these studies using the **three-prong R&D** sample instead to train the PFN
- We observe similar trends in performance
  - The PFN encoding can be varied to accommodate different signal priors

ANTELOPE still provides model independence and sensitivity beyond these priors





Concluding Remarks

- We've developed a new semi-supervised architecture ANTELOPE that performs anomaly detection on particle-level, unordered data by encoding the features into a PFN latent space
- ANTELOPE shows generalizability across many signal models assessed via LHCO dataset
  - The network is able to exploit low-level correlations to be sensitive to the challenging signal of BB3, where both dijet and trijet decay modes must be found
  - Different PFN embeddings can be employed for broad applications in BSM searches

Thank you!

Semi-supervised permutation invariant particle-level anomaly detection

arXiv: 2408.17409

Gabriel Matos[a] **Elena Busch**[a] **Ki Ryeong Park**[a] **Julia Gonski**[b]

[a] *Columbia University, 538 West 120th St, New York, NY 10027*

[b] *SLAC National Accelerator Laboratory, 2575 Sand Hill Rd, Menlo Park, CA 94025*

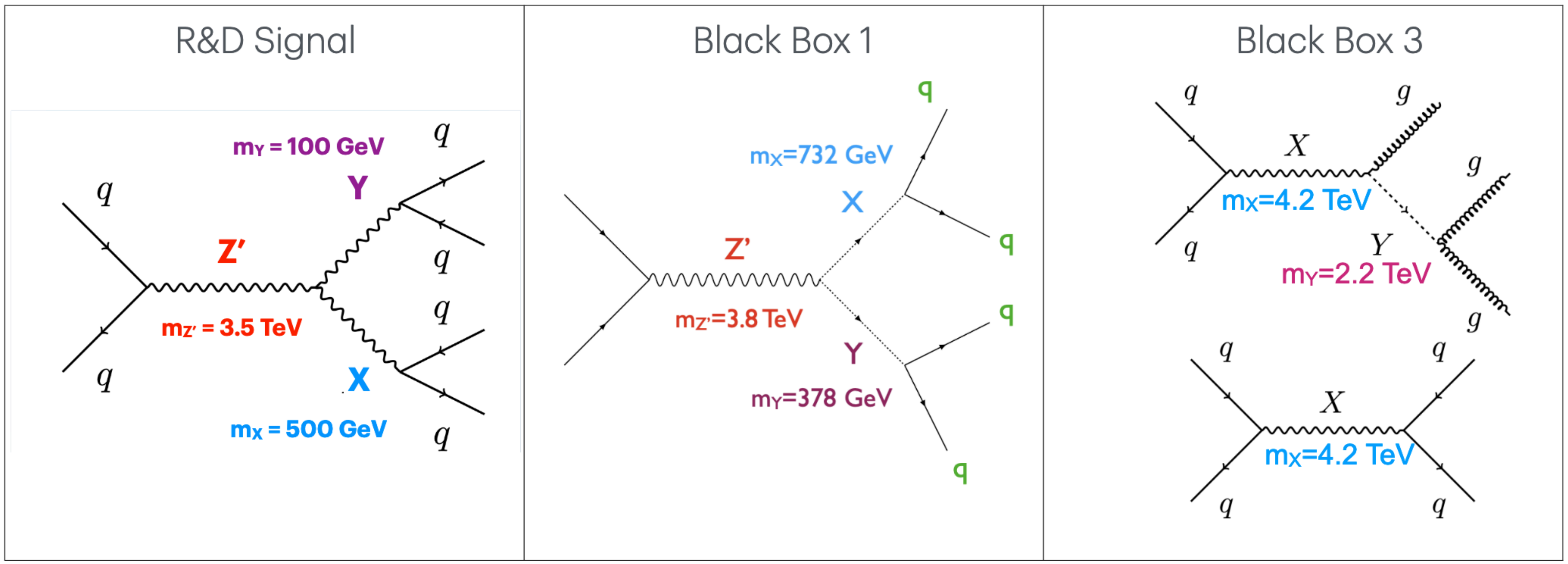
E-mail: gabriel.pinheiro.matos@cern.ch

*ChatGPT's interpretation of ANTELOPE

Backup

LHC Olympics Dataset

- Sample definitions can be found in LHCO paper, R&D and black boxes Zenodo links

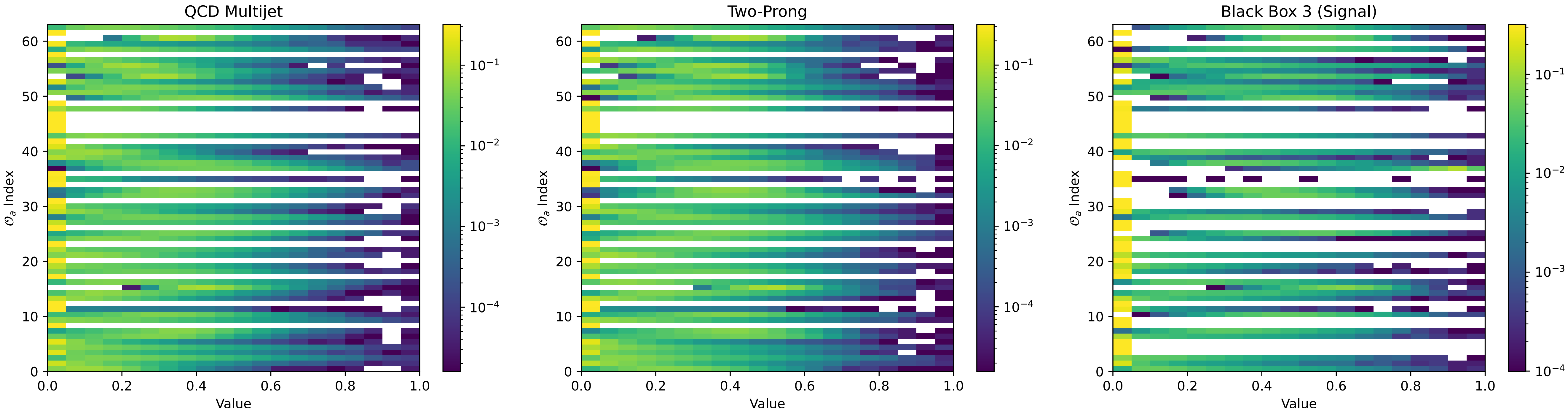


2-prong and 3-prong	2-prong	KK Graviton → R(→gg) + g
$m_{Z'} = 3.5$ TeV	$m_{Z'} = 3.8$ TeV	$m_X = 4.2$ TeV
$m_X = 500$ GeV	$m_X = 732$ GeV	$m_R = 2.2$ TeV
$m_Y = 100$ GeV	$m_Y = 378$ GeV	**KK Graviton → gg**
		$m_X = 4.2$ TeV

- **QCD Multijet** → R&D: generated w/ Pythia, BB2: generated w/ Herwig & modified Delphes card

PFN Latent Space

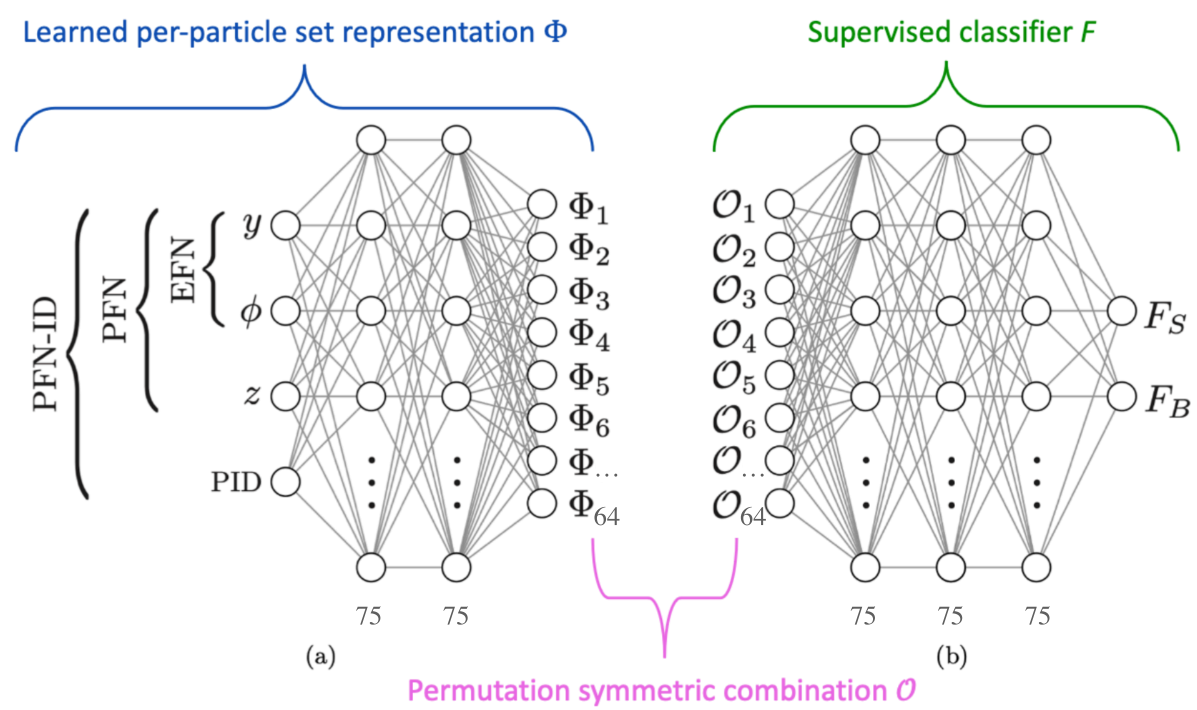
- Another nice feature of the PFN embedding is that we can inspect the latent space vars.
  - The latent space of the PFN is 64-dimensional → 64 individual $\mathcal{O}_a$ distributions
  - These are plotted below, with all $\mathcal{O}_a$ distributions scaled between 0 and 1



Anomalous distributions in latent space w.r.t. both QCD and two-prong samples

PFN Architecture Details

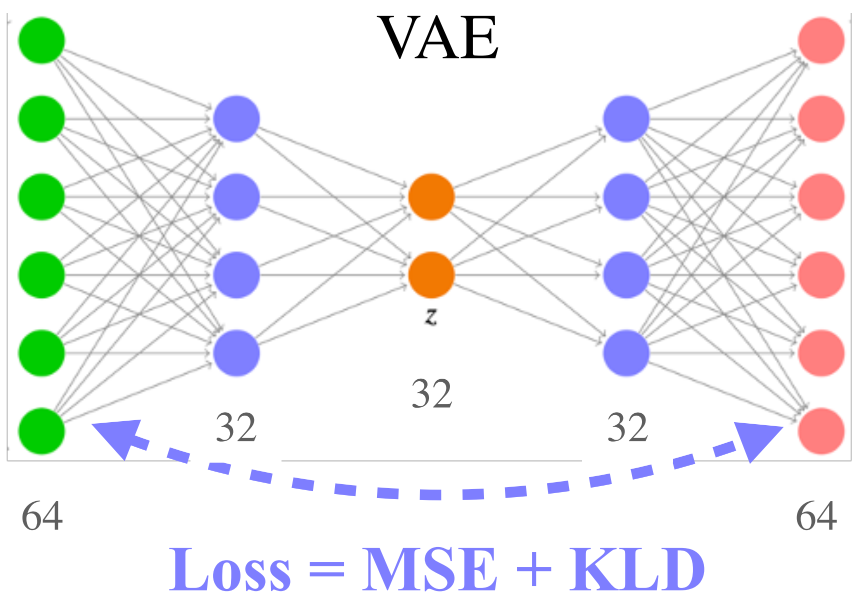
$$160 \times \begin{matrix} p_T \\ \eta \\ \phi \end{matrix} \longrightarrow$$



$$\mathcal{O}_a = \sum_i \Phi_a(y_i, \phi_i, z_i, \mathrm{PID}_i)$$

- First stage of the PFN has a 3-dimensional input of p_T, η, ϕ followed by two dense fully connected layers of dimension 75 with an output dimension of 64 (i.e. the Φ_a)
- Each Φ_a gets summed for the 160 input particles to derive the 64-dimensional $\mathcal{O}_a$
- The second stage of the PFN takes in the 64 $\mathcal{O}_a$ as input, followed by 3 dense fully connected layers of dimension 75, and 2-dimensional binary classifier output
- A cross-entropy loss is used for the training and model optimization

ANTELOPE Architecture Details



- Takes in as input the 64-dimensional $\mathcal{O}_a$ from a pre-trained PFN
- The VAE has a hidden layer of dimension 32, and latent layer of dimension 32
- A combination of the mean square error (MSE) and Kullback-Leibler divergence D_{KL} are used in the definition of the loss function

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} |\mathcal{O}_i - \mathcal{O}_i'| + \lambda D_{\mathrm{KL}}$$