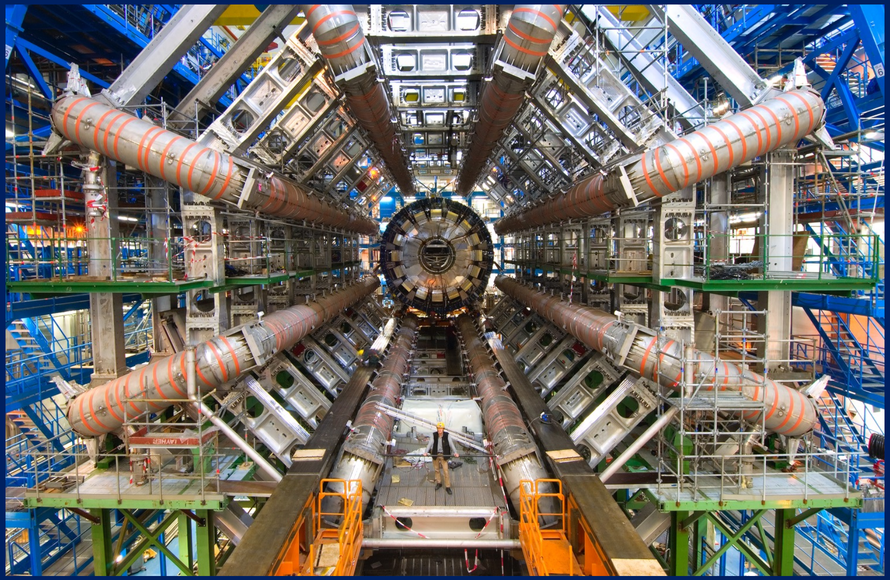
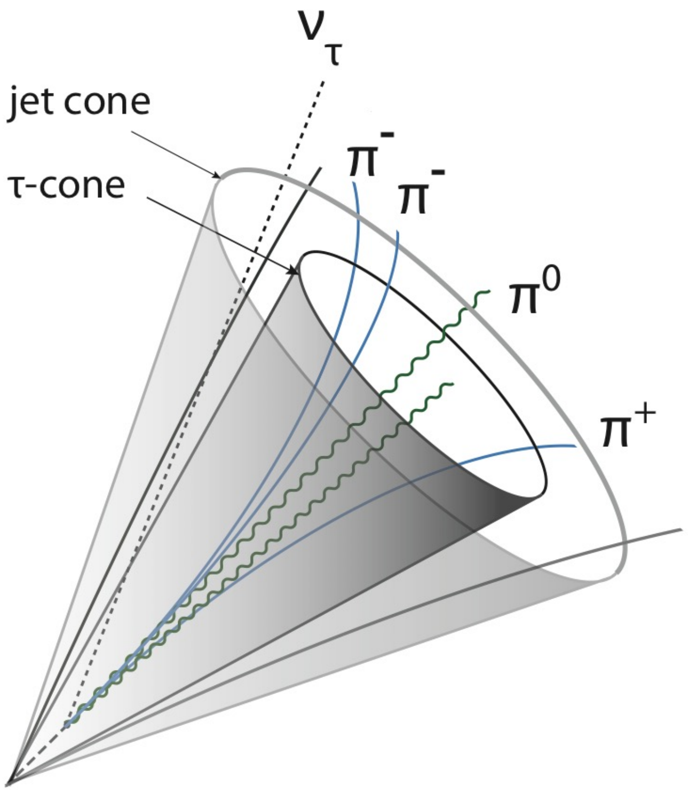


Tau Energy Scale Calibration at ATLAS using Machine Learning

Miles Cochran-Branson
Advisor: Quentin Buat
UW REU Final Presentation
August 17, 2022

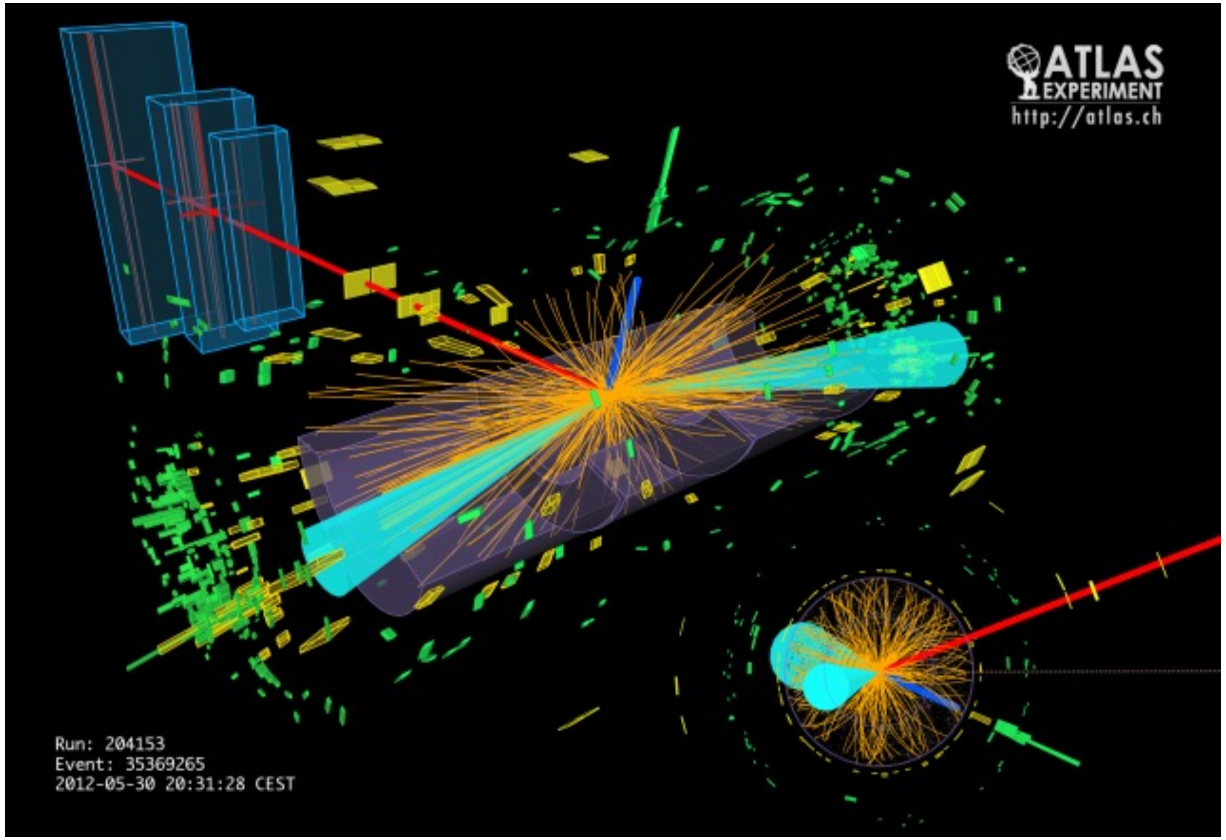
Tau Leptons in the ATLAS detector

- Tau lepton, $m_\tau = 1.777$ GeV
  - Decays ~instantly in detector
  - Tau decays:
    - $\tau \to \nu_T \nu_\ell \ell, \;\; \ell = e, \mu$
    - $\tau \to \nu_T$ hadrons
  - We study hadronic decays (65% BR)
  - $\tau_{had-vis}$: visible products of decays (not ν)
- **This work: quantify tau energy scale**



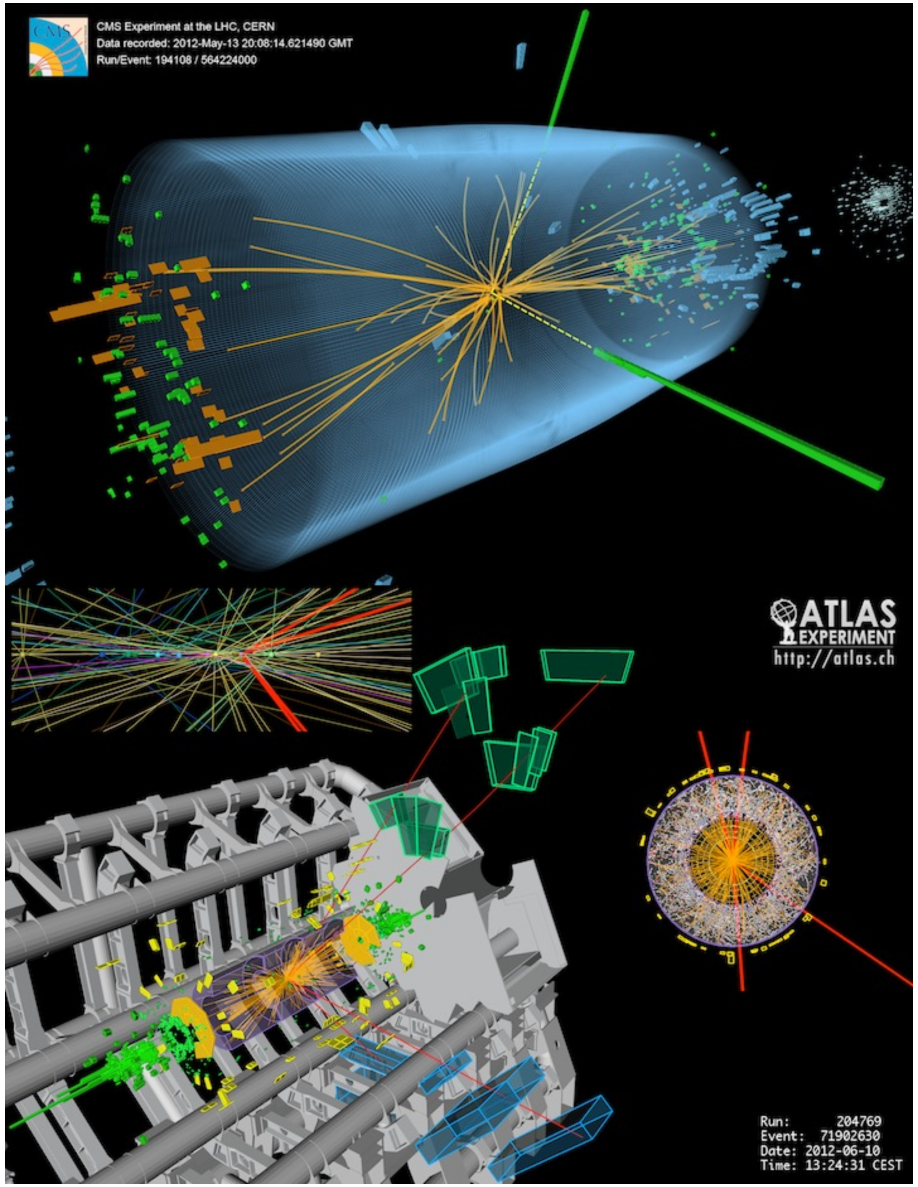
Tau Lepton as probe of Higgs

- Study $H \to \tau\tau$ process; has highest BR of leptonic H decays
- ~10% uncertainties in this decay channel remain
  - Possibilities for BSM physics
- Run 2 produced $\approx 500 \cdot 10^3$ such decays
  - Can afford to be picky



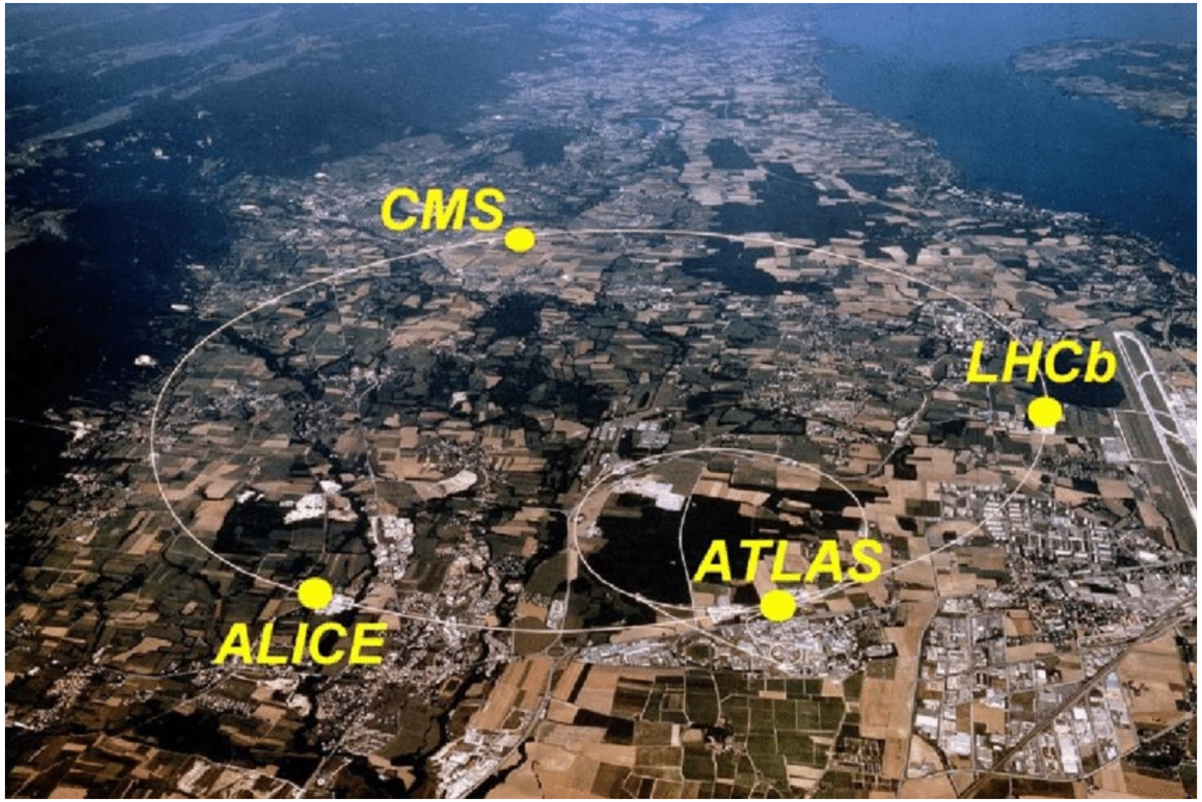
Higgs Boson

- Only scalar (spin-0) particle
- New interactions: self-coupling and Yukawa
  - I.e., Higgs field gives mass to particles
- **More precision measurement required**
- **Many properties still not understood**



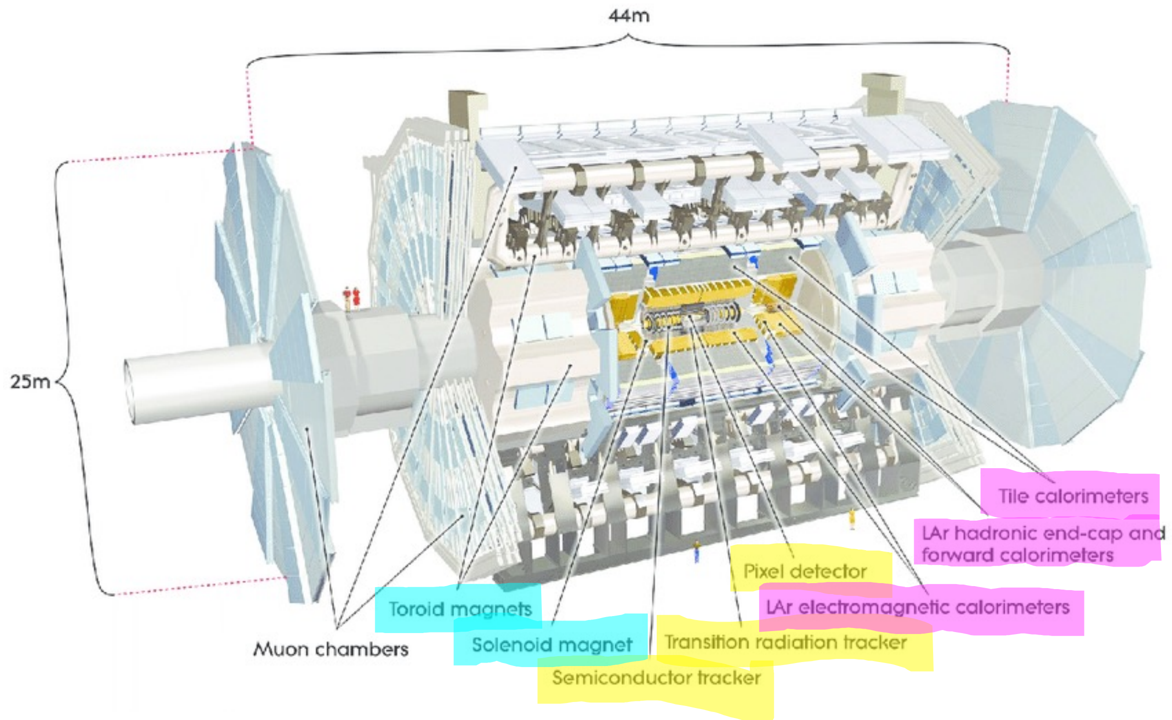
The Large Hadron Collider (LHC)

- Collides protons and heavy ions at 99.99…% the speed of light
- Energy = new particles
- Goal of LHC physics: find new particles
    - Look at kinematics, find anomaly
        - Necessitates finding 4-vector of particle
    - Test validity of SM and look for BSM physics

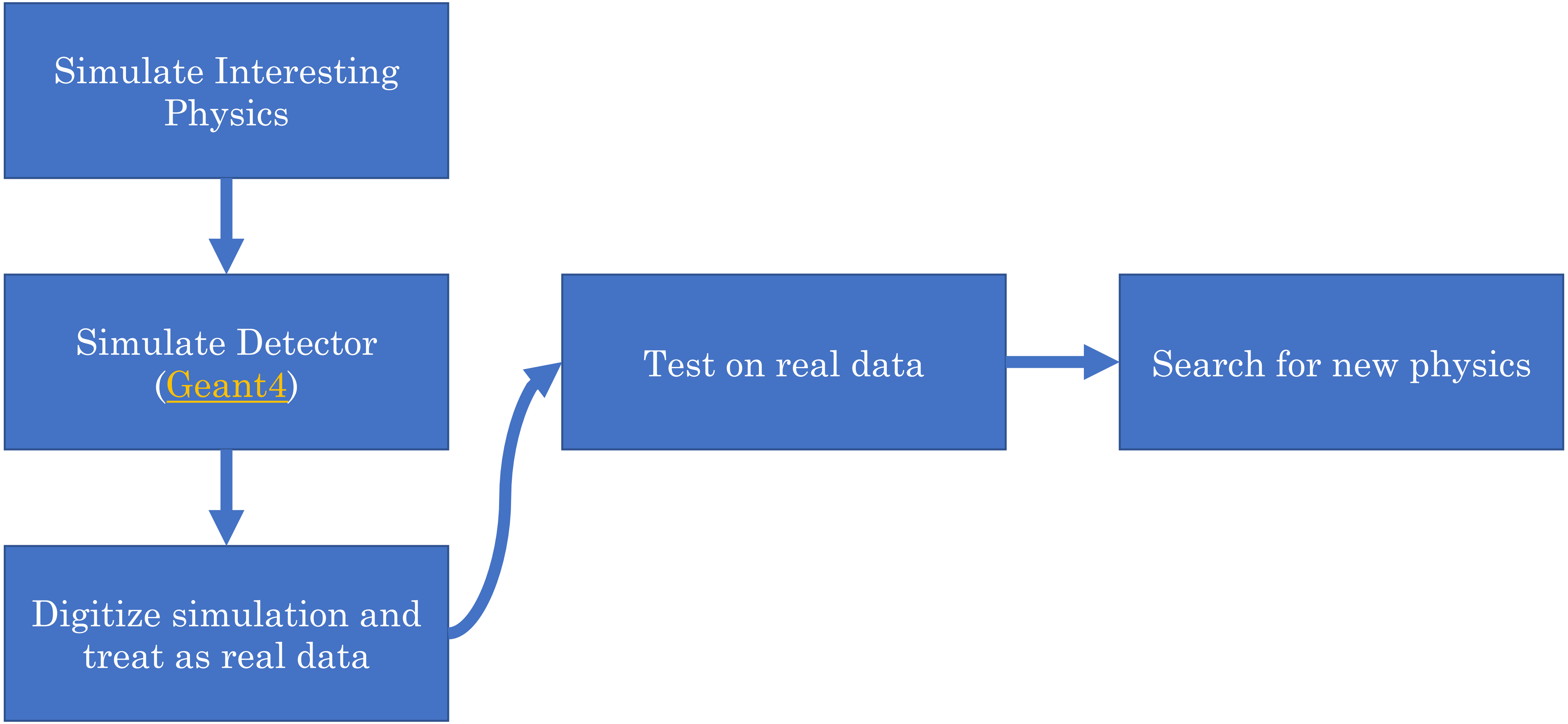


The ATLAS Experiment

- Components of detector: tracker, magnets, calorimeter
- 4-vector the HEP way:
  - p_T: transverse momentum
  - ϕ: azimuthal angle
  - $\eta \equiv -\ln\left(\tan\left(\frac{\theta}{2}\right)\right)$, θ : angle off z-axis
  - m (or E)
- **My work: find p_T of τ lepton = find Tau Energy Scale**

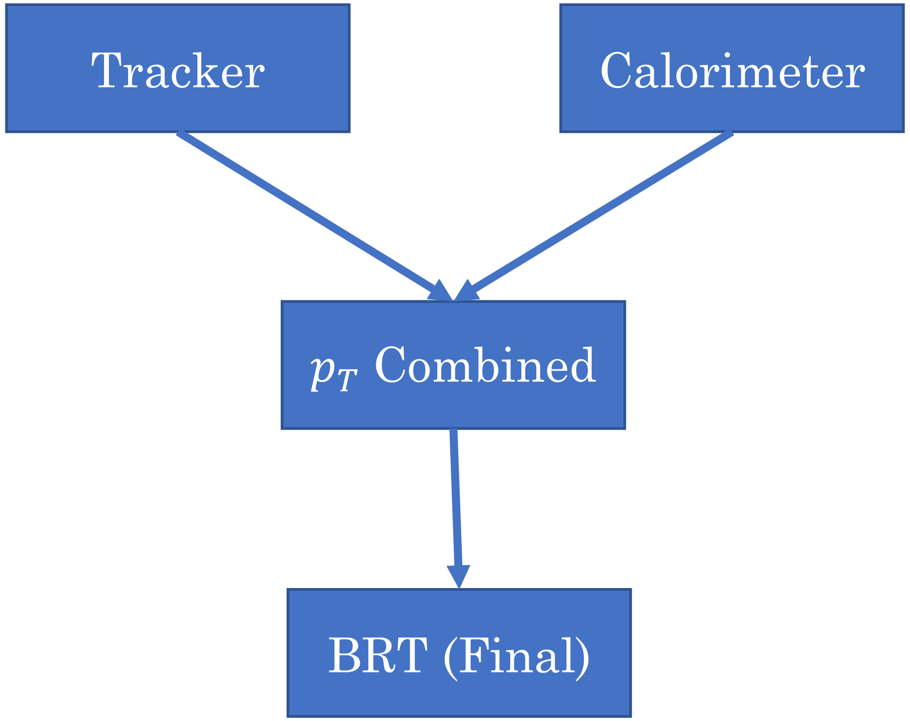


A note on HEP Algorithm Development



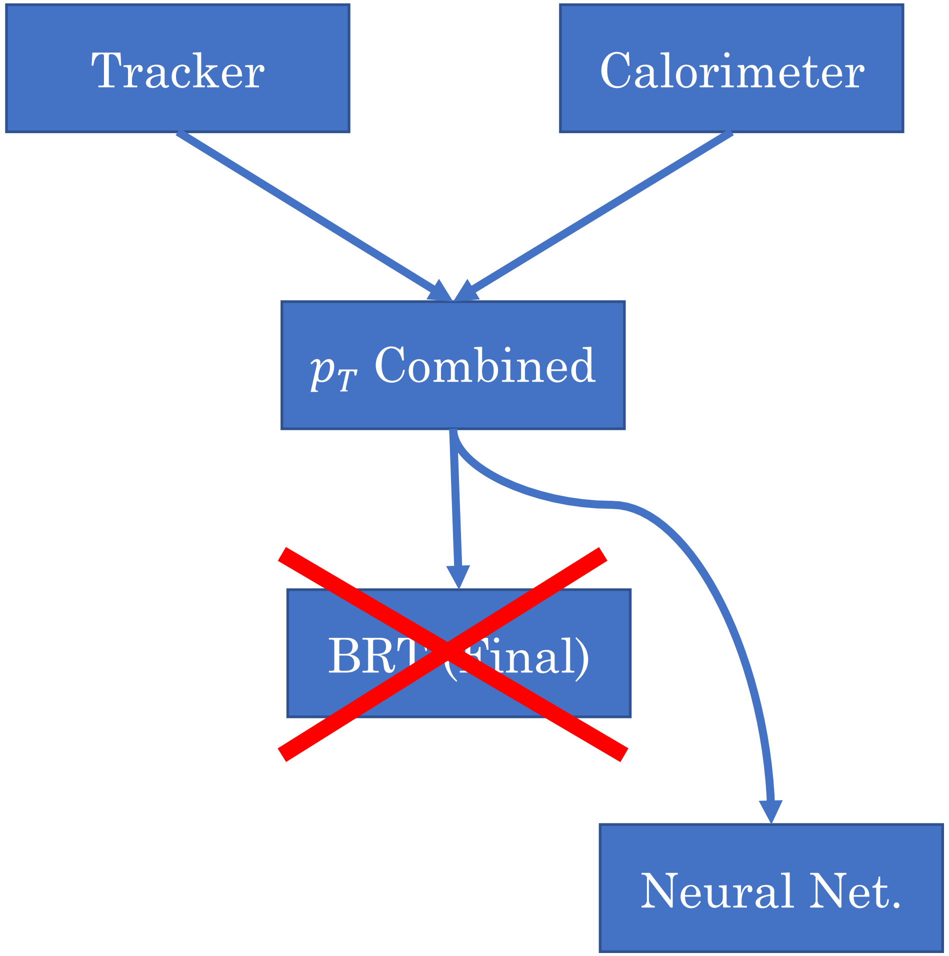
Tau Energy Scale Calibration

- Tracker and Calo data
- Data combined for estimate of p_T: *combined*
- p_T combined to Boosted Regression Tree (BRT): *final*



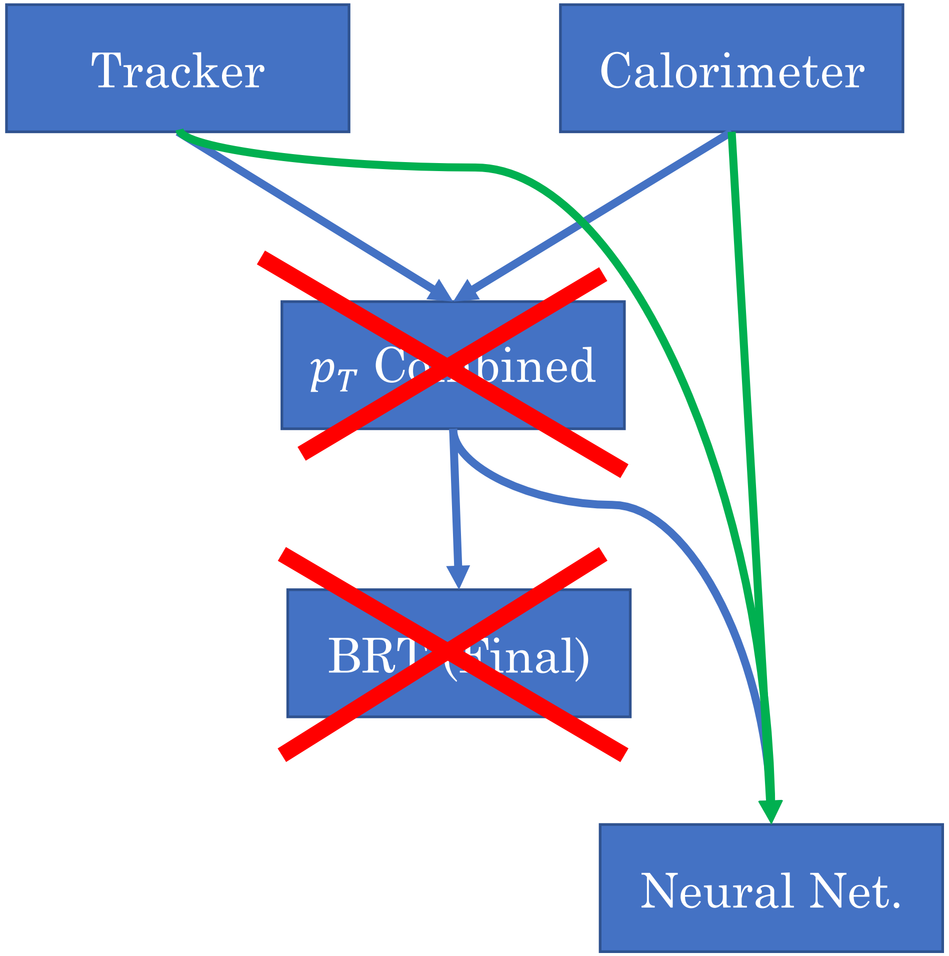
Tau Energy Scale Calibration

- Tracker and Calo data
- Data combined for estimate of p_T: *combined*
- p_T combined to Boosted Regression Tree (BRT): *final*
- This work: replace BRT with NN
- **Modern ML shown to beat old**
- **More diagnostic power**

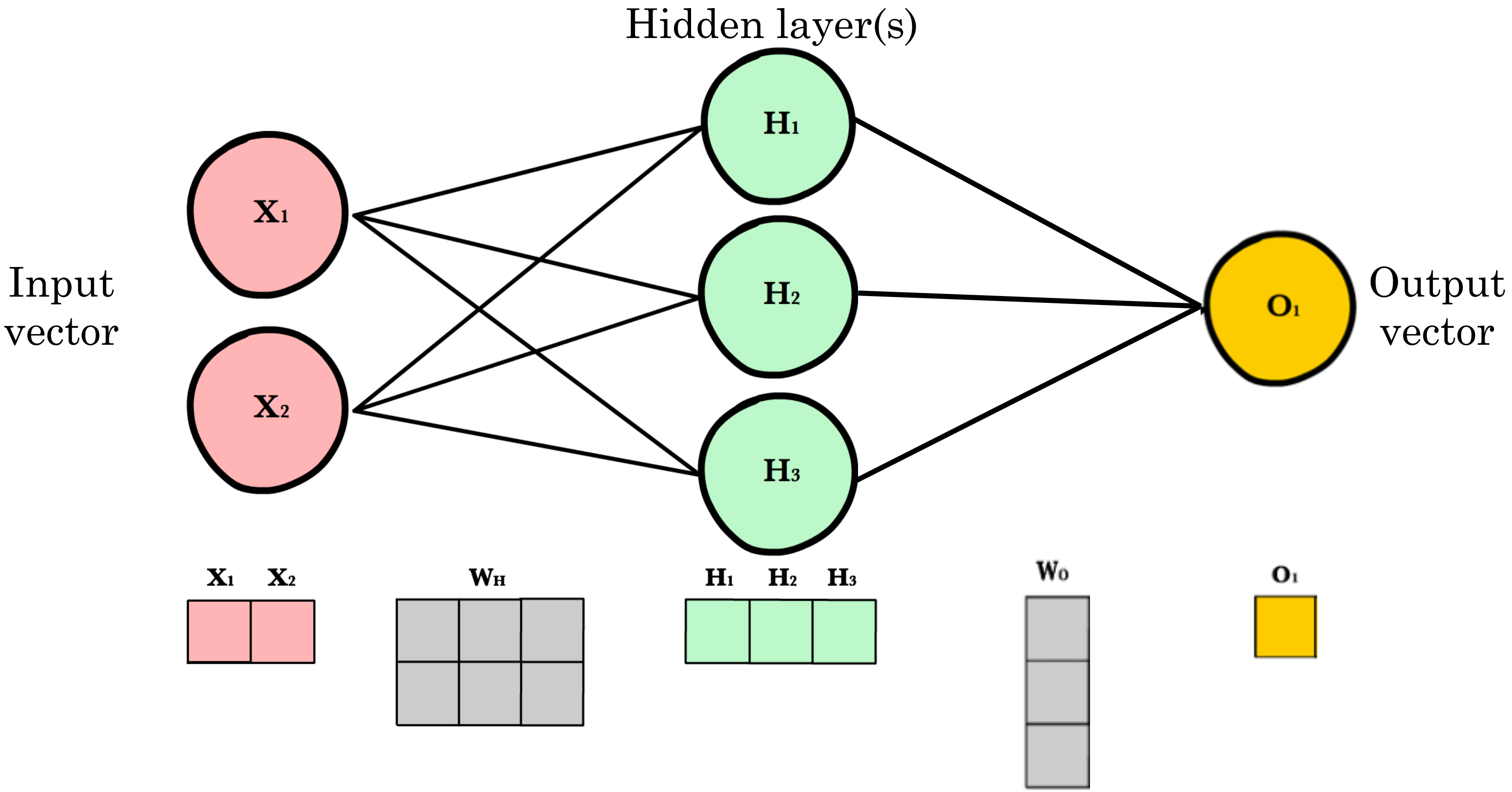


Tau Energy Scale Calibration

- Tracker and Calo data
- Data combined for estimate of p_T: *combined*
- p_T combined to Boosted Regression Tree (BRT): *final*
- This work: replace BRT with NN
- **Modern ML shown to beat old**
- **More diagnostic power**
- **Future: potential to bypass intermediate steps**



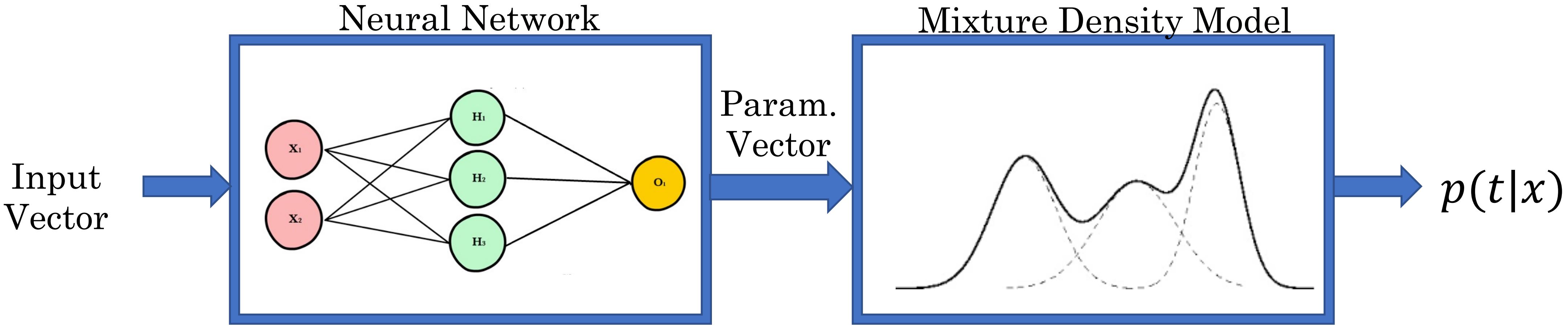
Dense Neural Networks (DNN)



- **Networks are functions composed of layers of matrix multiplication**

Source: ML Glossary, Forwardpropogation

Mixture Density Networks (MDN)



- **MDNs are neural networks with a probabilistic output given by a mixture density model**

Source: Bishop, C., *Mixture Density Networks*, 1994

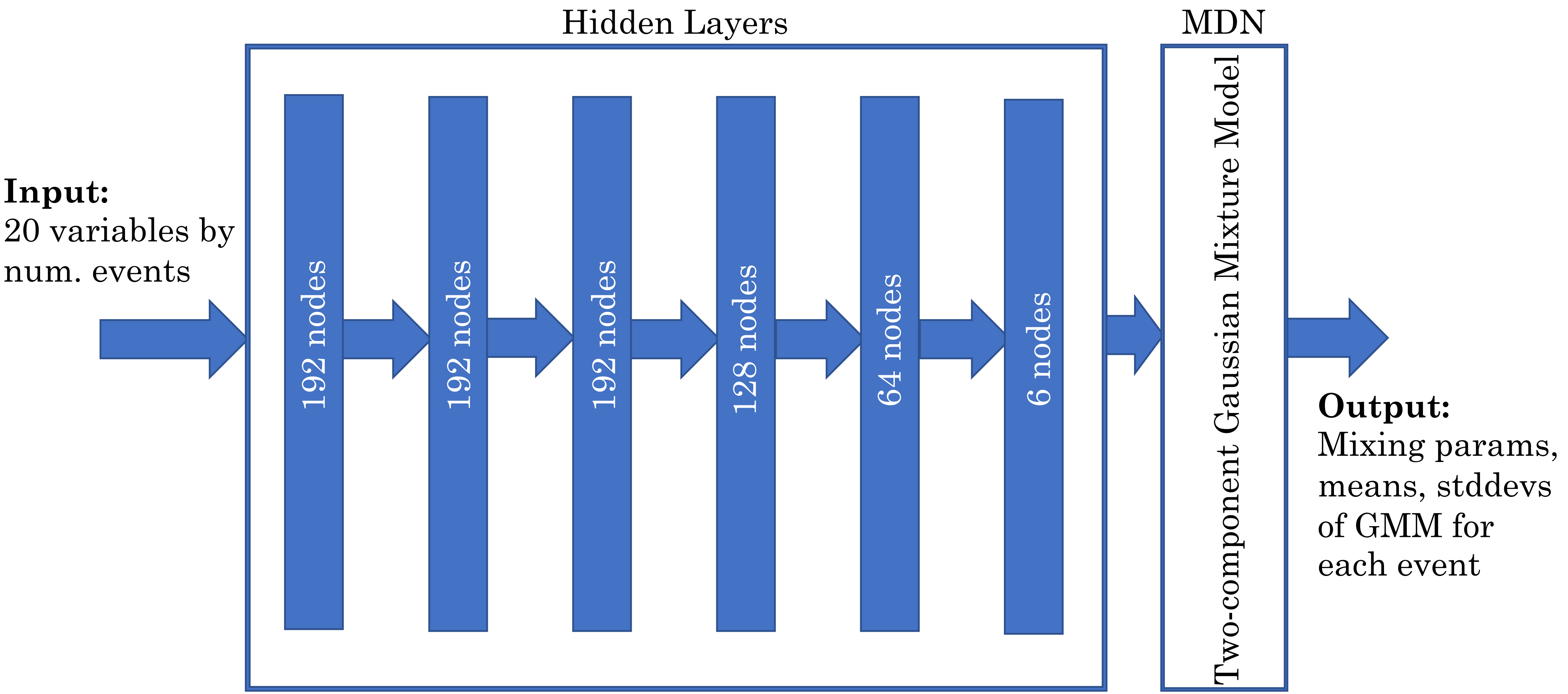
Universal Approximation Theorem

Any function can be approximated by a neural network with a single hidden layer!

The same holds for feed-forward networks with multiple layers

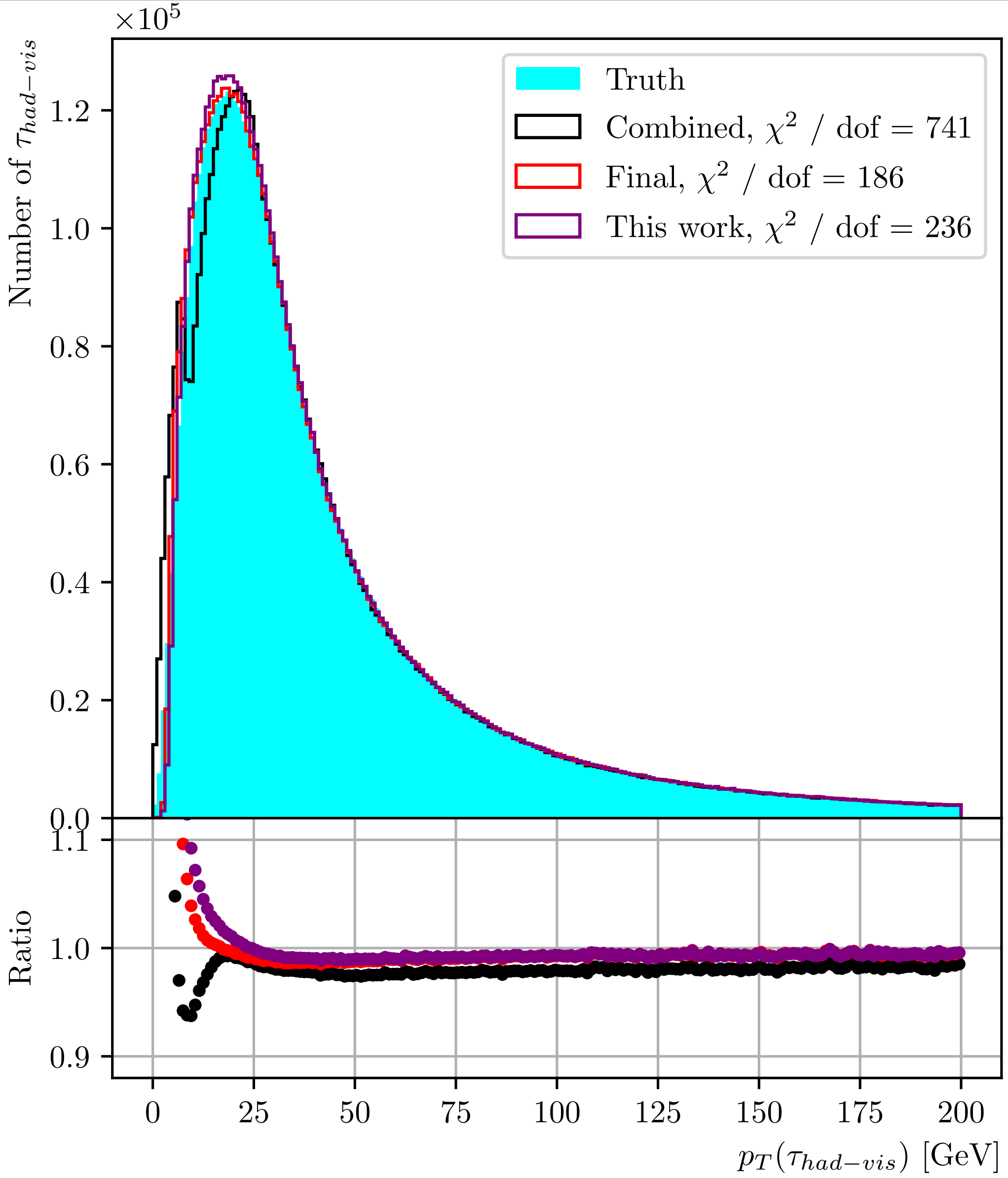
It is not clear, however, how many nodes are needed to accomplish this. It is possible that infinitely many nodes are required…

MDN Structure

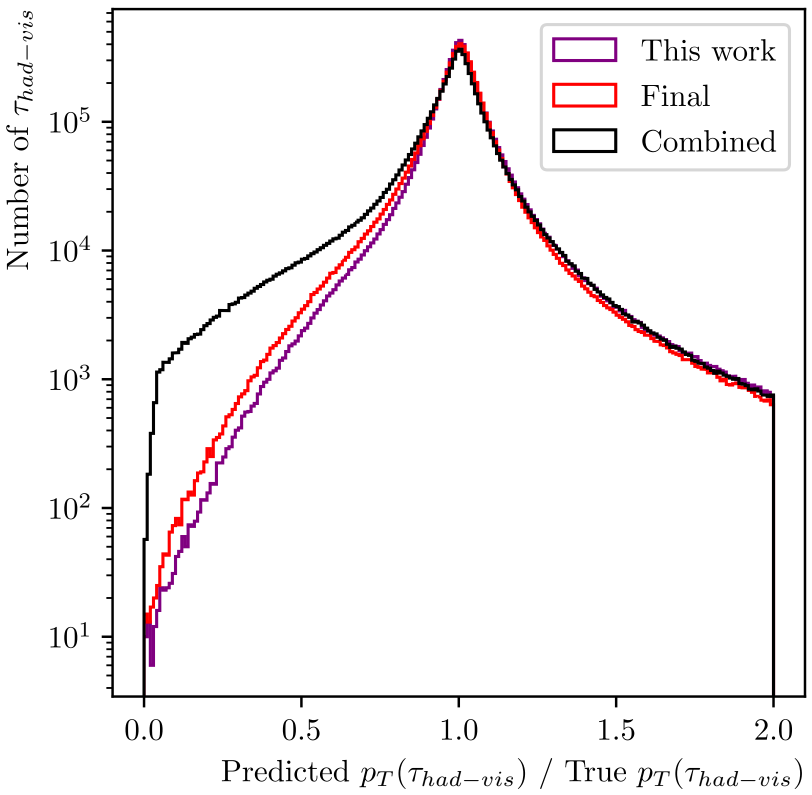


MDN Purpose

- Input from low and high-level variables from detector
- Try to match simulated data (i.e., Truth)
  - Data-aware technique
  - Regression problem
- **Goal: given a tau, find p_T**

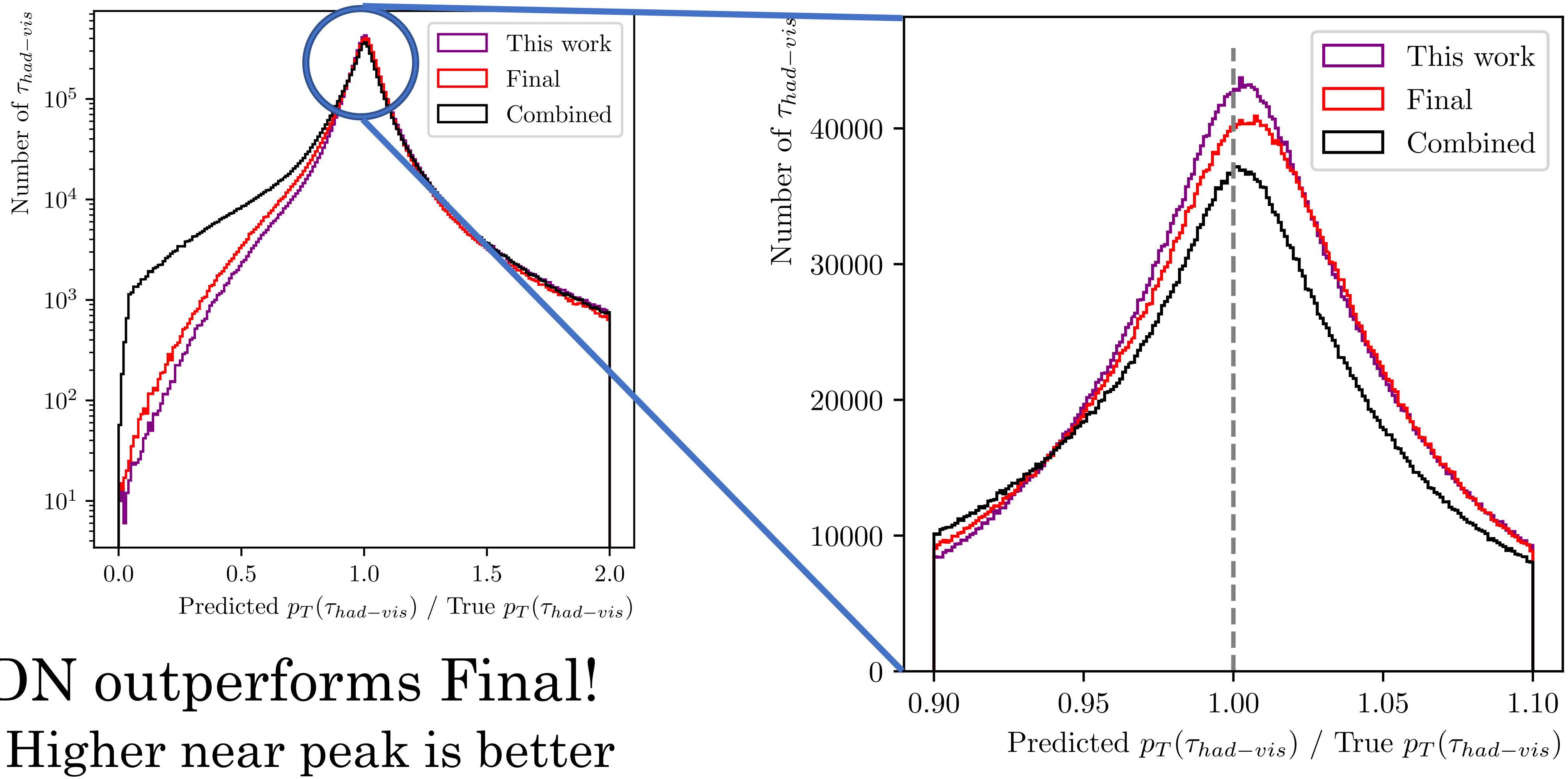


MDN Performance: Line shape



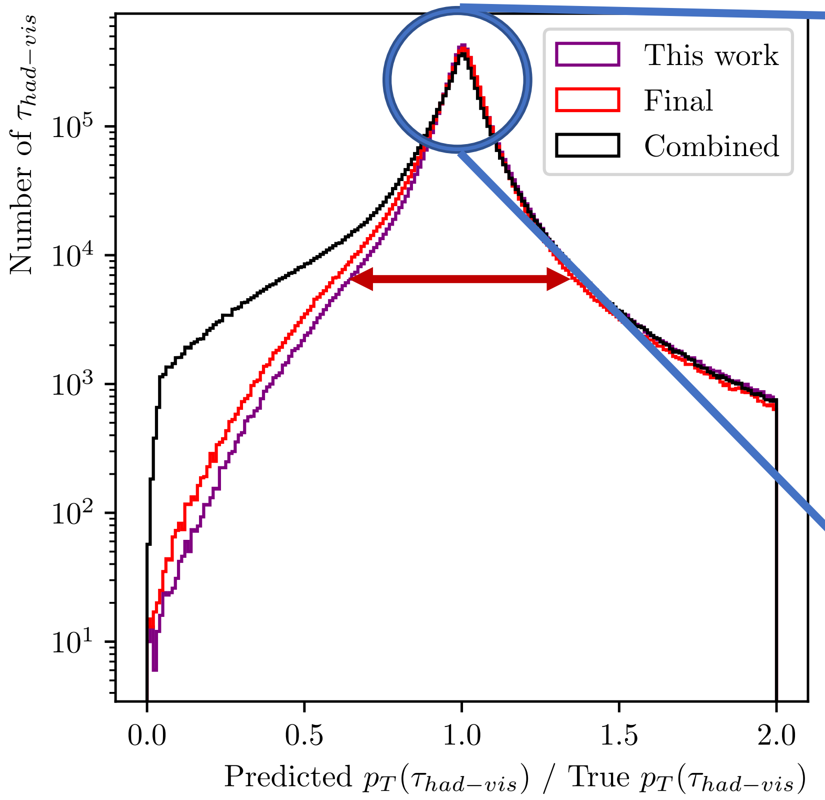
- MDN outperforms Final!
  - Higher near peak is better

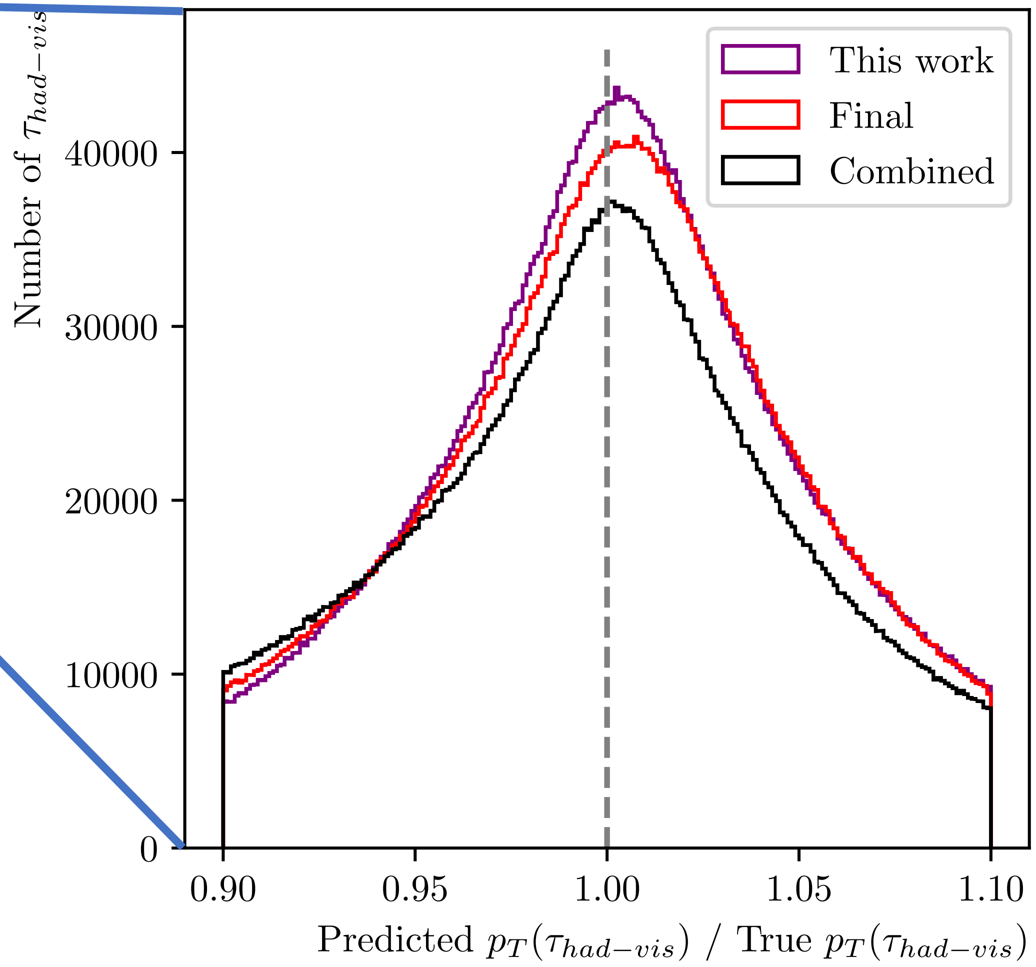
MDN Performance: Line shape



- MDN outperforms Final!
  - Higher near peak is better

MDN Performance: Line shape

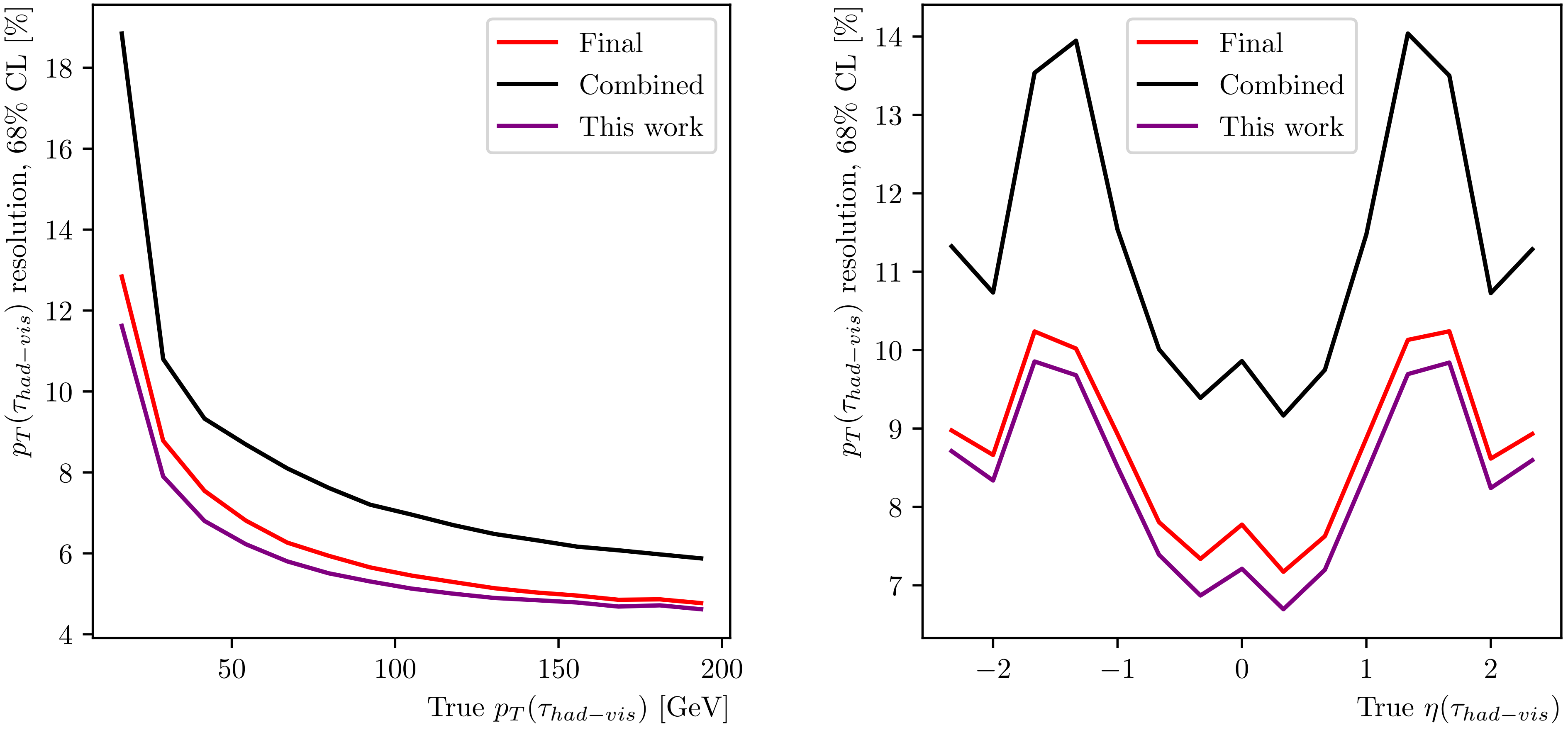




- MDN outperforms Final!
  - Higher near peak is better

Note: **resolution** is the width of the above curves at 68% CL

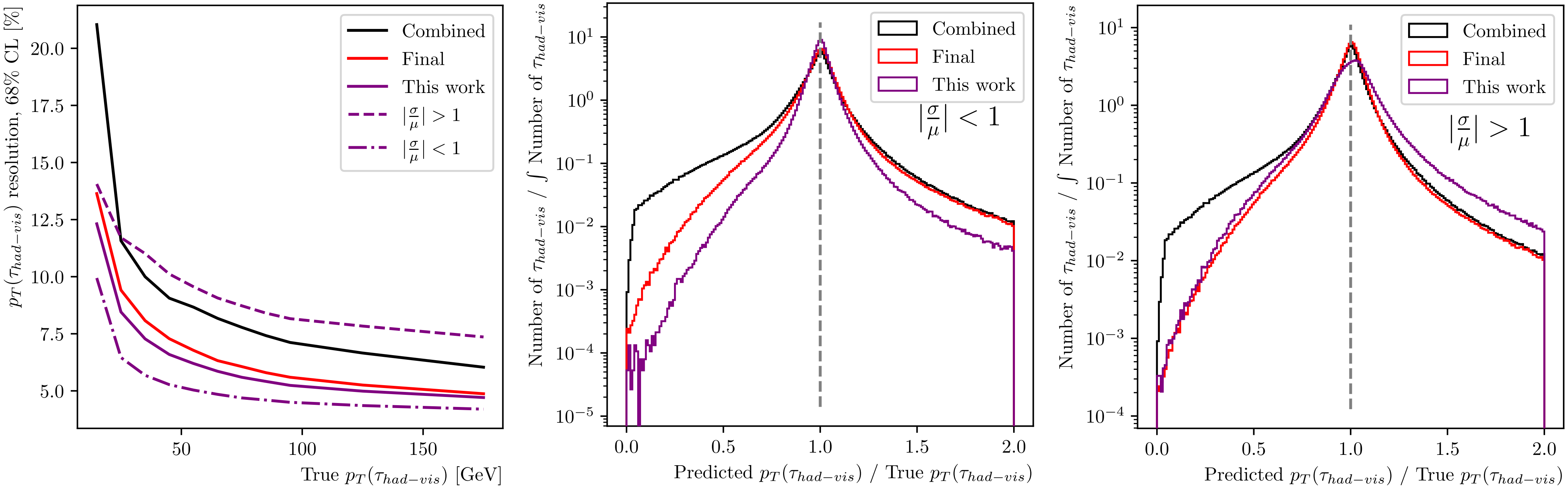
MDN Performance: Resolution



- MDN performs better than final on all metrics
- Have shown that an MDN produces better results than BRT

Event Selection

- With σ and μ from MDN, we can select "better" events



- Has huge diagnostic power
- Can split events on high and low purity; have enough events to even discard some!

Conclusions

- Novel MDN *beats* current method for TES calibration
- Includes extra information to aid selecting taus
- Presented work at Tau working group meeting. Recommendations for next steps in project:
  - Incorporate into ATLAS software
  - Remove high-level variables
  - Train a more complex network on outputs directly from tracker and calorimeter

A big thank you to Quentin for his mentorship throughout this project, the organizers of the REU for their support, and INT and NSF for funding the REU program

Backup slides!

Unknown properties of the Higgs

- Does it interact with lighter particles?
  - $H \rightarrow \mu\mu$ detected at CMS, small BR, other, lighter particles?
- Does it really interact with itself?
  - Decays of higgs to higgs are rare
- What's it's lifetime?
  - Theory predicts *very* small so extremely hard to measure
- BSM higgs?
  - Not scalar?
  - Not fundamental particle?
  - Other exciting things?

Source for e.g.: Grey 2022, *Happy Birthday Higgs Boson! What we do and don't know about the particle.* Nature Vol. 607, pp 220-221 (https://www.nature.com/articles/d41586-022-01834-5)

Mixture Density Networks (MDN)

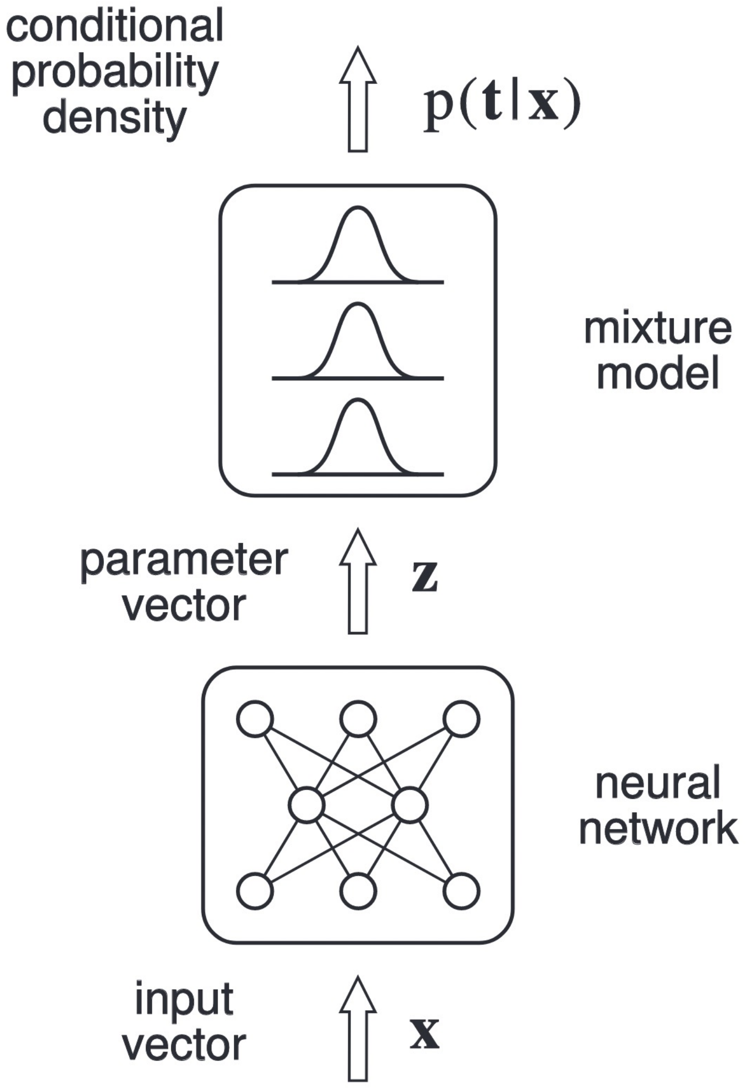
- Basic idea: substitute output layer of DNN with conditional prob. density
- General form of output:

$$p(t|x) = \sum_{i=i}^{k} \pi_i \, \phi_i(t|x)$$

$$\sum_{i=1}^{k} \pi_i \;= 1, \quad \pi_i \geq 0$$

π_i: mixing coefficients, ϕ_i: some pdf, x: input parameters, and t: output parameters

- MDN used in ATLAS pixel cluster splitting (Khoda, 2019)



Source: Bishop, C., *Mixture Density Networks*, 1994

Global μ, σ from GMM output layer

- Gaussian Mixture $f_k(x)$ with k gaussian components ϕ gives $\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\sigma}$
- We want some global μ, σ
- From law of total expectation:

$$\mu_g = \sum_{i=1}^{k} \pi_i \mu_i$$

- From law of total variance:

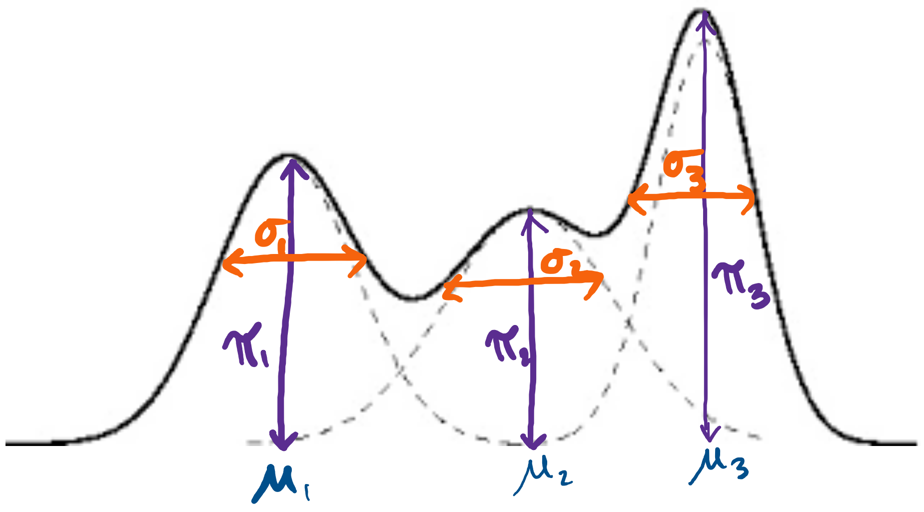
$$\sigma_g^2 = \sum_{i=1}^{k} \pi_i (\sigma_j^2 + \mu_j^2) - m_g^2$$

- Can find 'better' events using

$$\left| \frac{\sigma}{\mu} \right| < 1$$

$$f_k(x) = \sum_{i=1}^{k} \pi_i \, \phi(x; \mu_i, \sigma_i)$$

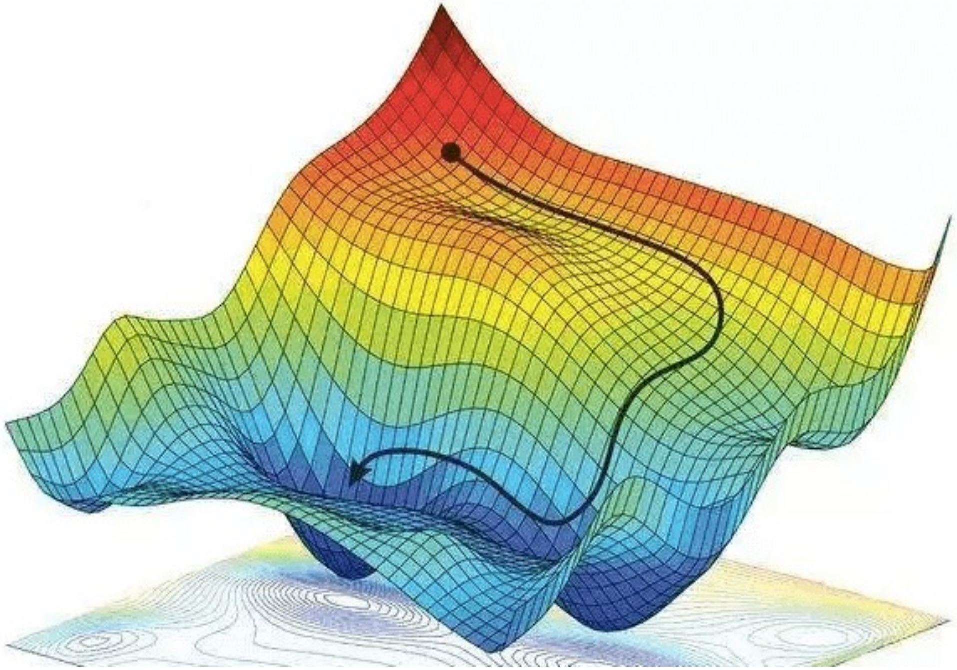
$$\sum_{i=1}^{k} \pi_i = 1, \quad \pi_i \geq 0$$



Source: Trailovic, L., and Pao, L., *Variance Estimation and Ranking of Gaussia Mixture Distributions in Target Tracking Applications*, 2002

Training a ML model

1. Break dataset into three sections: training, testing, validation
2. Define loss function
3. Pass training data to model
4. Evaluate loss function on validation data
5. Move towards minimum of loss
6. Update model
7. Repeat steps 3,4,5 until loss reaches a (local) minimum
8. Evaluate model on testing data



A little more on 4-vectors

- Common 4-vector: (p_x, p_y, p_z, m) or $(\boldsymbol{p}, E)$
- HEP 4-vector: (p_T, η, ϕ, m) or $(p_T, \eta, \phi, \mathrm{E})$
- Transformations:

$$p_x = p_T cos(\phi)$$

$$p_y = p_T \sin(\phi)$$

$$p_z = p_T / \tan(\theta) = p_T \sinh(\eta)$$

$$|p| = p_T \cosh(\eta)$$

$$p_T = \sqrt{p_x^2 + p_y^2}$$

$$\tan(\phi) = p_y / p_x$$

Dataset

- Simulayed $\gamma^* \rightarrow \tau\tau$ data
  - group.perf-tau.MC20d_StreamTES.425200.Pythia8EvtGen_A14NNPDF23LO_Gammatautau_MassWeight_v3_output.root
- Total number of events: 12608274
- Number of training taus: 12498400
- Number of validation taus: 2499680
- Number of testing taus: 6249020
- Cuts applied:
  - ptIntermediateAxisEM/ptIntermediateAxis < 25
  - ptPanTauCellBased/ptCombined < 25
  - ptIntermediateAxis/ptCombined < 25