Establishing evidence for the Higgs boson dimuon decay using the ATLAS detector

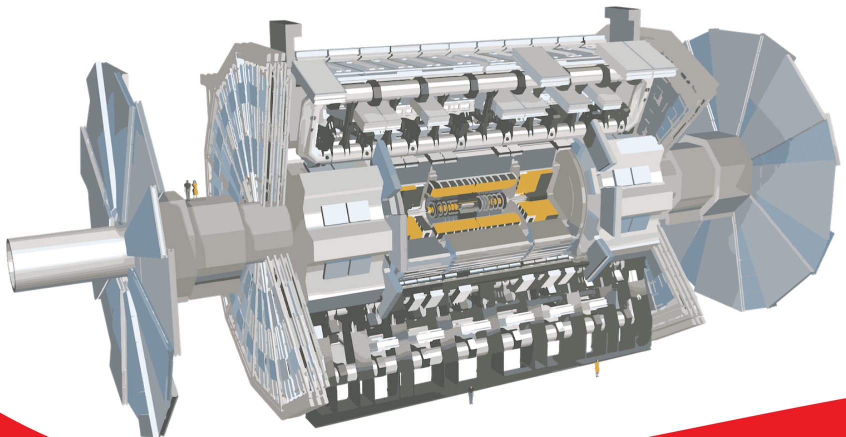
Bryce Norman

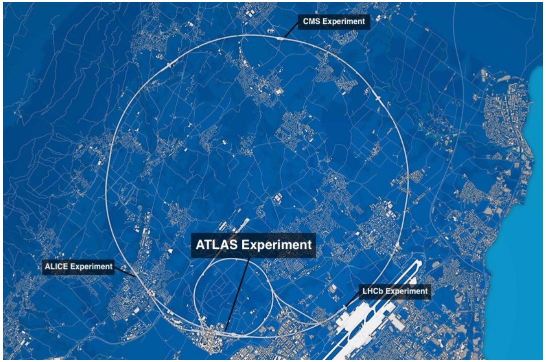
CAP Congress – 2024

Carleton University | Department of Physics

The LHC and ATLAS

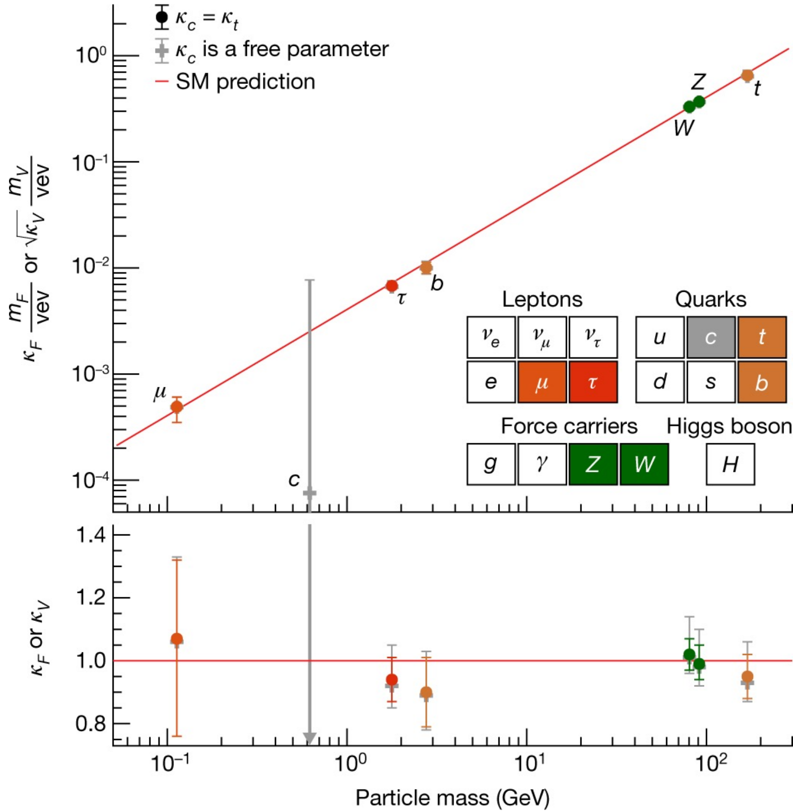
- The Large Hadron Collider (LHC) is the world's largest particle collider
  - Protons accelerated around 27 km ring and collided at centre-of-mass energy of 13.6 TeV
- ATLAS is the largest general-purpose detector on the LHC
  - Helped discover the Higgs boson in 2012
  - Cylindrical detector consisting of many subsystems wrapped in layers
- ATLAS is currently collecting data during LHC Run 3





The Higgs Boson

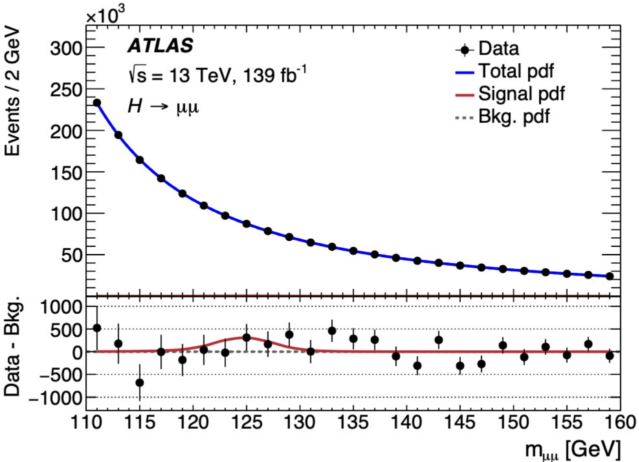
- The Higgs boson was discovered in 2012 by ATLAS and CMS
  - Interactions with the Higgs field give the fundamental particles in the Standard Model mass
  - A particle's mass is proportional to its coupling with the Higgs boson
- We have only observed (>5σ significance) the Higgs boson interacting with very massive Standard Model particles
  - *W* and *Z* bosons, top and bottom quarks, τ lepton

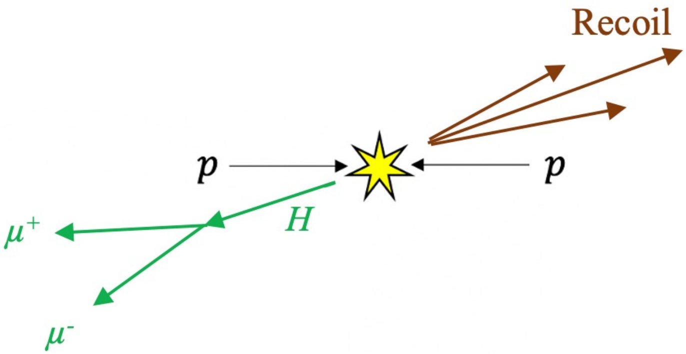


https://arxiv.org/pdf/2207.00092

$H \rightarrow \mu\mu$

- Want to measure a Higgs coupling to a second-generation particle at a much lower, untested mass scale
    - The Higgs to dimuon decay provides the best opportunity
    - Due to the small muon mass, this is a very rare process, and a Higgs boson will only decay to two muons 0.022% of the time
- Previous ATLAS result for Higgs to dimuon decay observed 2σ significance from full run-2 dataset (2015-2018 data)
    - Very low signal-to-background ratio due to low branching ratio

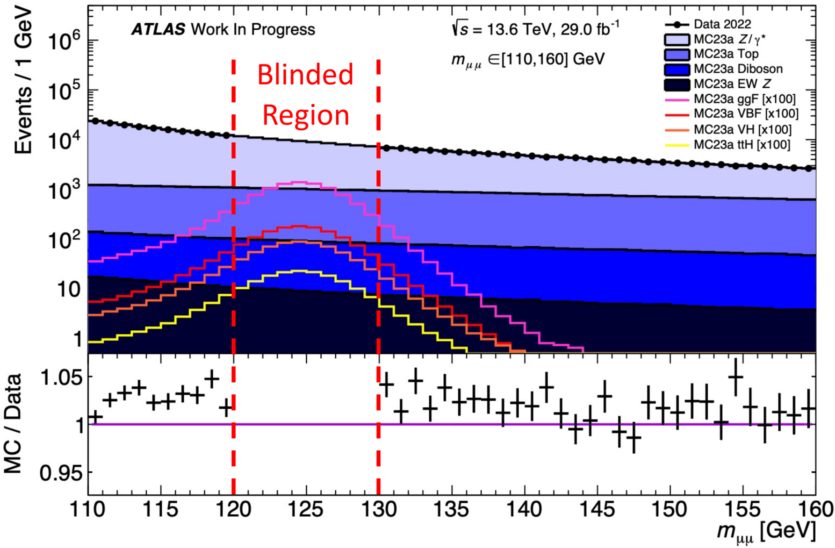




https://arxiv.org/pdf/2007.07830

Establishing Evidence for $H \rightarrow \mu\mu$

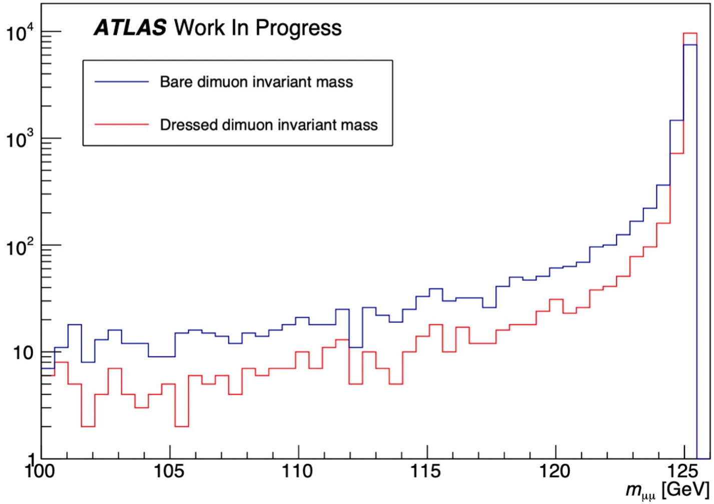
- Very challenging to measure $H \rightarrow \mu\mu$
  - Need to improve analysis to establish evidence (3σ significance)
- Possible improvements include:
  - Improved Final State Radiation (FSR) recovery
  - Implement deep learning
    - Density reweighting of MC to better model background processes
    - Splitting data into optimized categories
  - Increase statistics with data taken during LHC Run-3



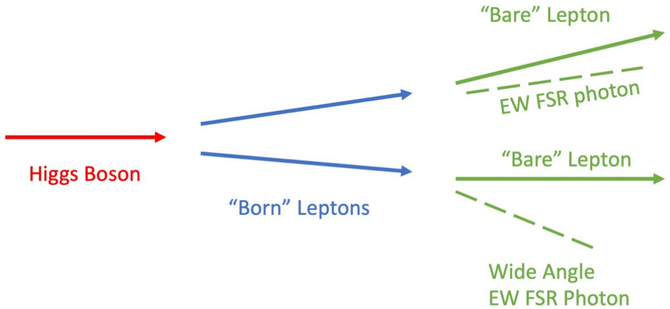
Dimuon invariant mass spectrum for 2022 ATLAS data and Monte Carlo (MC) simulations. $H \rightarrow \mu\mu$ signal is scaled by x100.

FSR Recovery

- After the Higgs boson decays to two “Born” muons, one or both may emit a Final State Radiation (FSR) photon
  - If we only reconstruct the Higgs candidate using the “bare” muons after FSR emission, we are missing the energy carried away by the photon
  - We want to recover this FSR and add it to our muons (“dressed” muon includes FSR photon) before reconstructing Higgs candidate



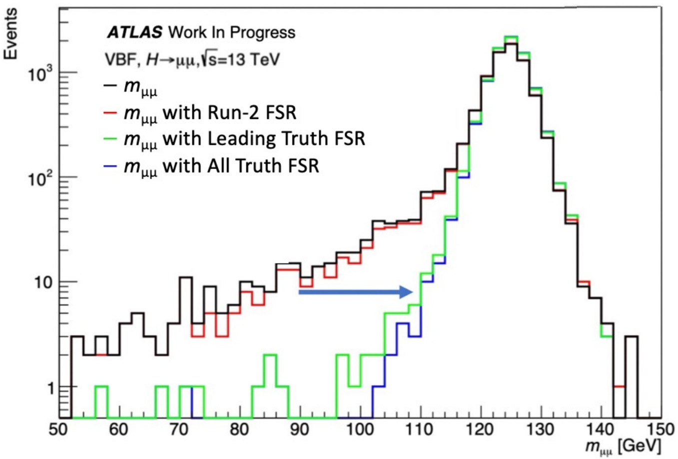
Dimuon invariant mass distribution of truth-level particles (what MC generator produces) after correcting for FSR. A lower tail persists in spectrum due to Dalitz decay



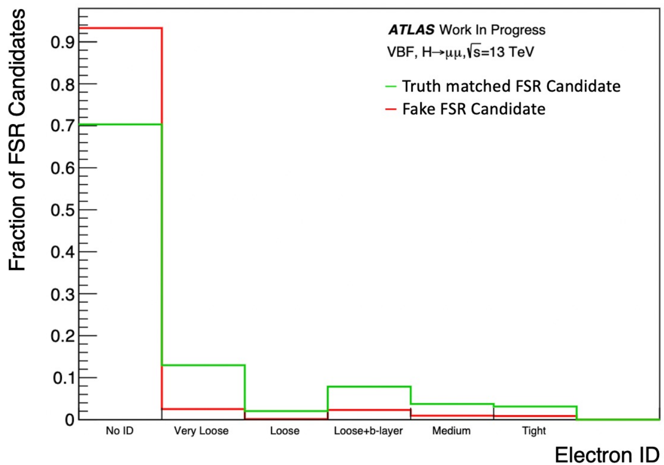
FSR photons can carry significant fraction of Born lepton energy.

FSR Recovery Improvements

- FSR recovery was performed during run-2 ATLAS analysis
  - FSR candidates were selected based on p_T, angle from muon (ΔR) and energy deposited by candidate in EM calorimeter
  - If we add in the truth FSR correction in MC we see that there is room for improvement
  - Idea: Try phase space dependent cuts (p_T and η of FSR) and include new parameters such as predefined identification working points (WP)



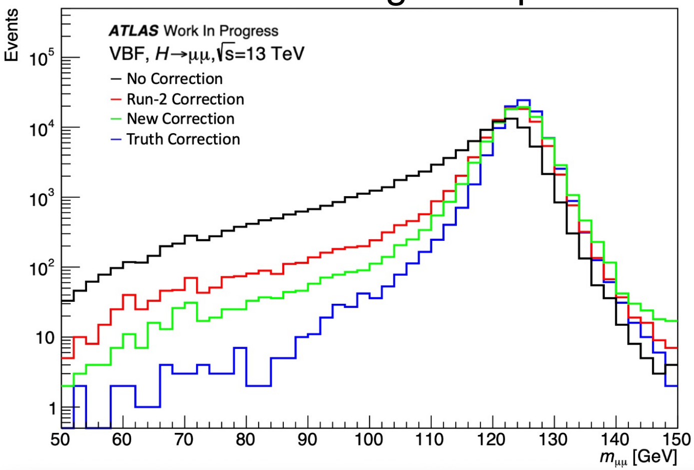
We want to move from the uncorrected mass (black) to the best possible correction (blue). Red shows improvement using run-2 FSR correction.



Photon and electron candidates have predefined ATLAS IDs. Candidates with tighter WPs are more likely to match a truth FSR candidate from a muon

FSR Recovery Improvements

- Improved cuts were found that get FSR recovery closer to the truth correction
  - It is possible to achieve a lower fake rate and higher efficiency than previous run-2 correction in signal events
  - New method moves 1.44% more signal events to 120-130 GeV window
  - Results are based on run-2 MC, need to investigate for run-3
  - Need to investigate impact of cuts on background

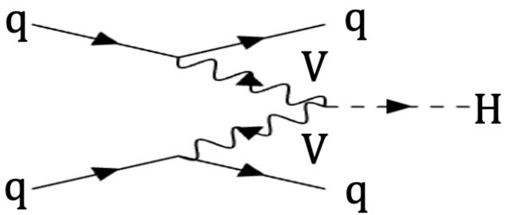


Green is invariant mass with new "V0" FSR correction cuts. Plot only includes events which have truth collinear ($\Delta R<0.2$) FSR with $p_T > 0.5$ GeV.

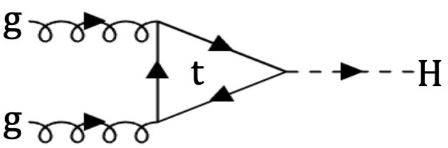
Truth FSR p_T	**Run-2 Cuts** **Fake rate / Efficiency**	**"V0" Cuts** **Fake rate / Efficiency**	**"V1" Cuts** **Fake rate / Efficiency**
3.0-7.5 GeV	**44.92% / 66.66%**	**34.18% / 55.03%**	**37.55% / 59.66%**
7.5-30.0 GeV	**15.06% / 92.80%**	**11.25% / 92.73%**	**14.38% / 95.56%**
30.0-125.0 GeV	**6.20% / 85.50%**	**4.77% / 95.55%**	**5.69% / 95.67%**

Event Categorization

- Data events that pass selection for the analysis are split into mutually exclusive categories for analysis
  - Categories are based on the properties and kinematics of the events measured by the ATLAS detector
  - Some of these categories will have better signal-to-background ratios
  - By extracting the Higgs signal from these categories separately we will see a large increase in the overall statistical significance
  - Run-2 categories were defined based on different Higgs bosons production modes (E.g., gluon-gluon fusion, vector boson fusion)
  - If we can improve categorization it would lead to an increase in significance



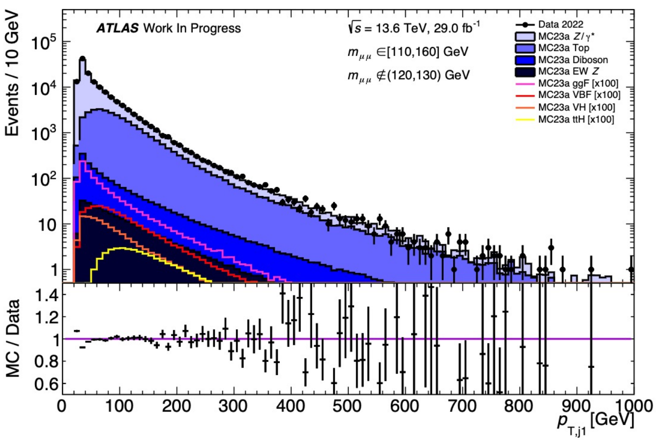
Vector Boson Fusion (VBF)

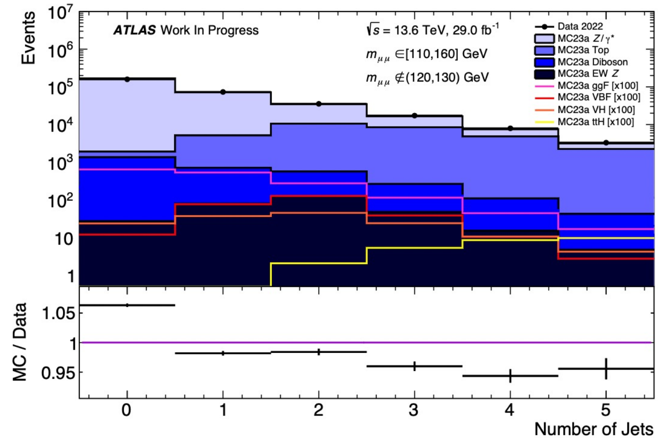


Gluon-Gluon Fusion (ggF)

Using NN to Determine Categories

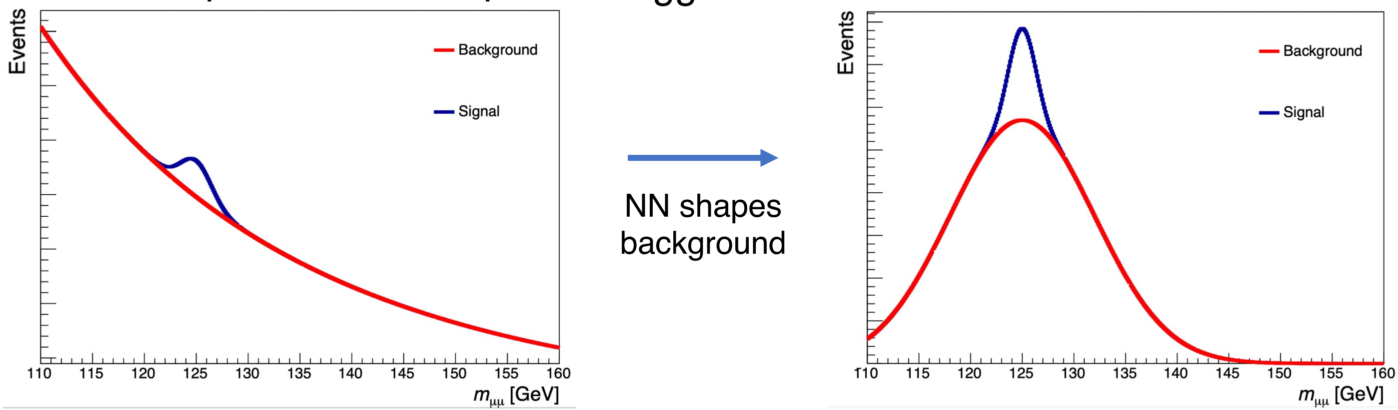
- What if we use machine learning to select optimal categories?
  - Variables from each event will be provided to a neural network (NN)
  - Want to use variables with separation between signal and background
  - Modern deep NNs are very powerful and should be able to differentiate between Higgs signal events and background events
  - NN will develop a classifier which can be used to determine if data events are “signal-like” or “background-like”





Using NN to Determine Categories

- The dimuon invariant mass is an important variable that we want to use to in our final fit to extract the Higgs signal
  - We ideally want a smooth, flat background the we can model and subtract
- Problem: The NN could learn the Higgs mass and shape the background to look like a Higgs peak in $m_{\mu\mu}$
  - Special care needs to be taken to make sure the NN doesn't shape a bump around the expected Higgs mass



Distance Correlation (DisCo)

- Problem: We do not want the output of the NN network classifier to be correlated with the dimuon invariant mass ($m_{\mu\mu}$)
- Solution: Add an additional term to the loss function of the NN which penalizes it for being correlated with $m_{\mu\mu}$
  - Loss function is a measure of how well the NN models the training data
  - The NN will aim to minimize this loss during training

$$L_{\text{total}} = L_{\text{classification}}(\vec{y}, \vec{y}_{\text{true}}) + \lambda \text{dCorr}^2_{y_{\text{true}}=0}(\vec{m_{\mu\mu}}, \vec{y})$$

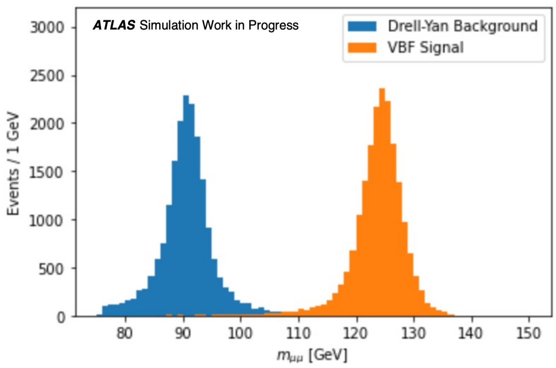
- Want to use a metric that can capture non-linear dependence between distributions
  - Use the distance correlation metric (DisCo)

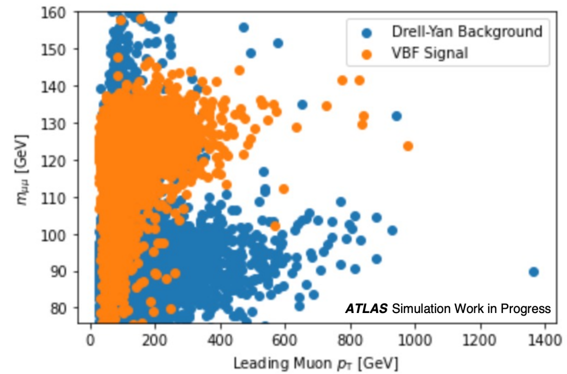
$$\text{dCorr}^2(X,Y) = \frac{\text{dCov}^2(X,Y)}{\text{dCov}(X,X)\text{dCov}(Y,Y)}$$

$$\text{dCov}^2(X,Y) = \langle |X-X'||Y-Y'| \rangle + \langle |X-X'| \rangle \langle |Y-Y'| \rangle - 2\langle |X-X'||Y-Y''| \rangle$$

Training the NN

- Let's read in a sample signal and background dataset
  - VBF $H \rightarrow \mu\mu$ signal MC
  - Drell-Yan $Z \rightarrow \mu\mu$ background MC (main background process)
  - Give the NN the kinematics of the muons and jets for training
- Train one NN with a common loss function (binary cross entropy) and a second NN with an additional DisCo term
  - Want to compare the performance of the two NNs and their correlation with the dimuon invariant mass

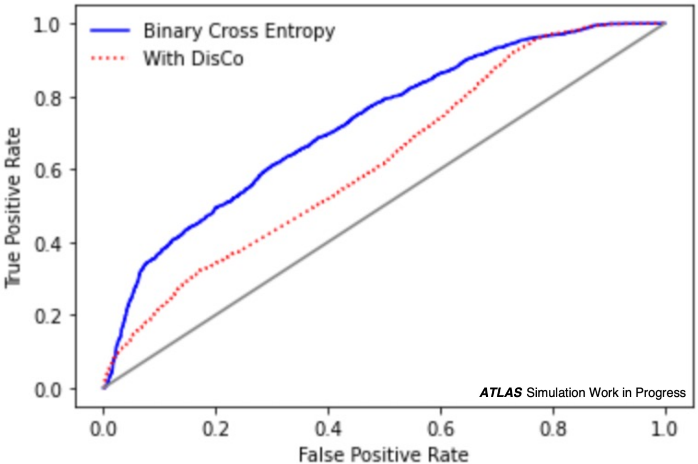




Mass appears to have good discriminating power. Background has mean around Z boson mass, signal has mean around Higgs mass

Results

- NN with DisCo performs worse but is less correlated with $m_{\mu\mu}$
  - Receiver operating characteristic (ROC) curve shows the performance of the NN at different thresholds
  - NN using DisCo is penalized for any correlation with $m_{\mu\mu}$, so it performs worse (which is expected)
  - Both NNs have highest correlation with jet momentum
  - NN using DisCo has less correlation with $m_{\mu\mu}$ (DisCo is working!)



ROC curve shows NN performance. A perfect classifier would have true positive rate of 1 and false positive rate of 0

	Binary Cross Entropy	Binary Cross Entropy + DisCo
DisCo($p_T^{\mu 1}$, NN output)	0.0256	0.0547
DisCo($p_T^{\mu 2}$, NN output)	0.0769	0.0737
DisCo(p_T^{j1}, NN output)	0.3220	0.3462
DisCo(p_T^{j2}, NN output)	0.2813	0.4154
DisCo($m_{\mu\mu}$, NN output)	0.1570	0.0718

Conclusion

- The Higgs to dimuon decay provides the best opportunity to measure a Higgs coupling to a second-generation fermion
- Due to the small branching ratio of $H \rightarrow \mu\mu$, this is a very difficult process to measure
- Improvements in analysis and more statistics are required to establish evidence for this process with the ATLAS detector
- FSR recovery improvements could move more Higgs signal events into our signal window, increasing the signal-to-background ratio
- Optimized categories could result in increased statistical significance of this measurement
  - NNs can be used to develop categories
  - DisCo can be used to prevent the NN from shaping the background

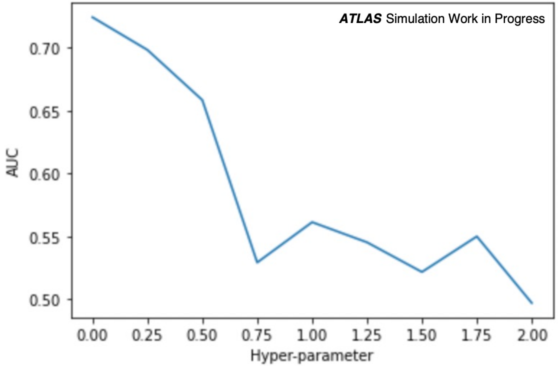
Backup

Effect of Hyperparameter

- The amount of loss contributed by the correlation between the NN output and $m_{\mu\mu}$ is controlled by a hyperparameter lambda

$$L_{\text{total}} = L_{\text{classification}}(\vec{y}, \vec{y}_{\text{true}}) + \lambda \text{dCorr}^2_{y_{\text{true}}=0}(\vec{m_{\mu\mu}}, \vec{y})$$

- What happens when we vary this hyperparameter?
  - As we increase lambda, correlation with $m_{\mu\mu}$ decreases
  - As we increase lambda the performance of the NN also decreases
  - Note that output of NN can vary between trainings, findings are still very preliminary



Area under ROC curve (AUC) is a measure of NN performance. A perfect classifier would have AUC=1.0