Identification of highly boosted $Z \to ee$ decays with the ATLAS detector

Florian Kiwit
supervised by Dominik Duda

Max Planck Institute for Physics
Munich

LHC Seminar, May 30$^{\text{th}}$ 2022

FSP ATLAS
Erforschung von Universum und Materie

ATLAS EXPERIMENT

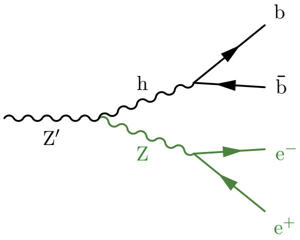
MAX-PLANCK-INSTITUT FÜR PHYSIK

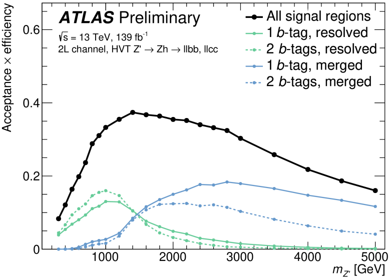
Overview

1. Motivation
2. Standard boosted Z boson reco+id
3. Novel boosted Z boson reco+id
4. Outlook

Why are highly boosted Z bosons interesting?

- Many BSM theories predict new heavy vector bosons with masses at the TeV scale
  - e.g. GUT, Composite Higgs, Extra Dimensions
  - These particles can have large branching ratios to $\mathrm{h/W/Z}$ bosons
  - Heavy BSM particles will lead to high $\mathrm{p_T}$ $\mathrm{h/W/Z}$ bosons
  - Identification of boosted boson decays is crucial



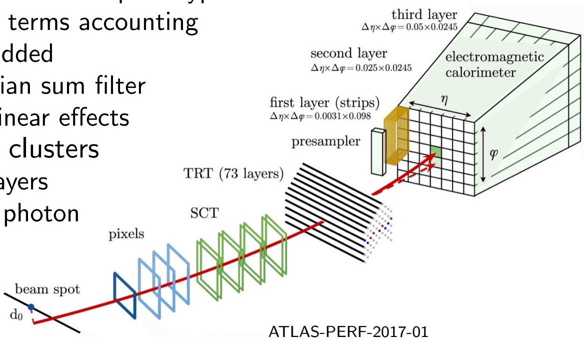


ATLAS-CONF-2020-043

Standard $Z \rightarrow ee$ Reconstruction

Based on standard electron reconstruction

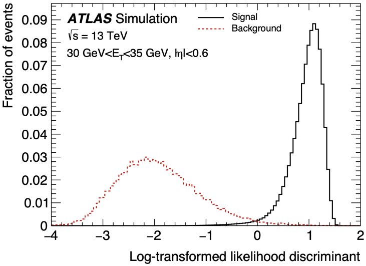
1. Reconstruction of E/γ cluster
   - Window of size 3 x 5 slid over 200 $\times$ 256 towers in $\eta \times \phi$
   - Candidate seeded if $E_{tower} > 2.5$ GeV
2. Reconstruction of track
   - Clusters of hits form space-points
   - Track reconstruction based on pion hypothesis
   - If fit fails, additional terms accounting for bremsstrahlung added
   - Refitting with Gaussian sum filter accounting for non-linear effects
3. Matching of tracks and clusters
   - Four hits in silicon layers
   - Not associated with photon
   - Energy calibration



ATLAS-PERF-2017-01

Standard Electron Identification

- Likelihood based:
  $$L_{S(B)}(\mathbf{x}) = \prod_{i=1}^{n} P_{S(B),i}(x_i)$$
  - S: signal, B: background
  - pdfs P derived from histograms of the simulation samples
  - x_i: Track+Calo information
- Discriminant:
  $$d_L = \frac{L_s}{L_s+L_b} \rightarrow d'_L = -\frac{\ln(d_L^{-1}-1)}{15}$$
- Problem: Ignores correlation between input variables

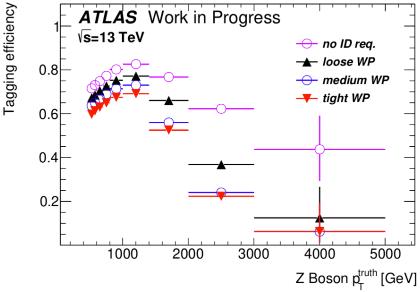


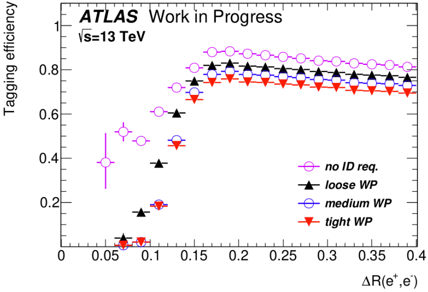
ATLAS-PERF-2017-01

Reconstruction and Identification Degradation

Standard method of electron reconstruction and identification degrades with large $\mathrm{p_T}$ (i.e. large $\mathrm{m_{Z'}}$)

- Due to small angular separation between the leptons from the Z decay



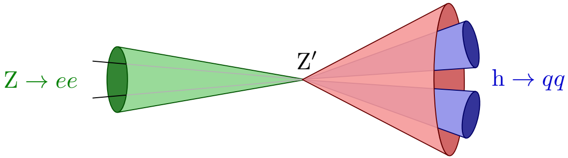


The novel Z Boson Reconstruction and Identification

- Reconstruct $\mathrm{Z} \rightarrow \mathrm{ee}$ as Small-Radius jet (AntiKt4LCTopo)

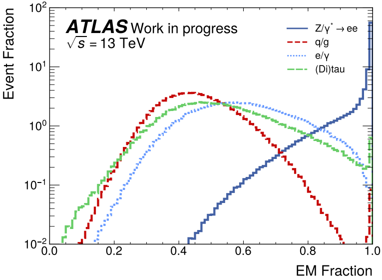
- Develop $\mathrm{Z} \rightarrow \mathrm{ee}$ tagger based on Neural Network to mitigate efficiency loss
- Training based on properties of
  - Small-Radius jet
  - Tracks (Inner Detector)
  - Cluster (Calorimeter)

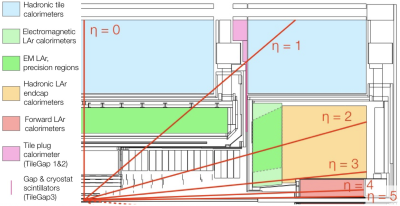
  } Matched to Small-R jet



Boosted Z Boson Reconstruction

- Reconstructing leptonic Z decay as a Small-Radius jet is an unusual approach (high f_{EM})
- AntiKt4LCTopo at constituent level
  - Calibration of EM objects with out-of-cluster and dead material corrections
  - No hadron assumptions
- $\mathrm{m}^{\mathrm{true}}$ are $\mathrm{p}_{\mathrm{t}}^{\mathrm{true}}$ are obtained from 4-vector sum of all relevant generator-level particles ghost-associated to the studied jets (based on "GhostTruth" container)



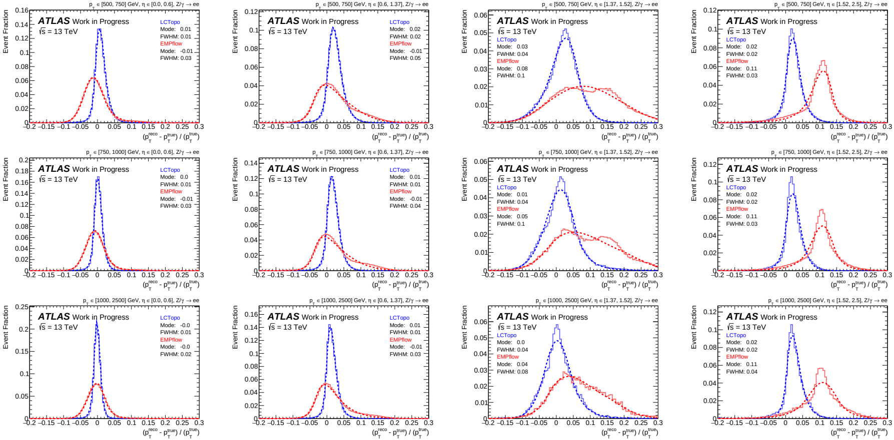


$Z \to ee$ —— P_T Resolution

$p_T \in [500, 750]$

$p_T \in [750, 1000]$

$p_T \in [1000, 2500]$

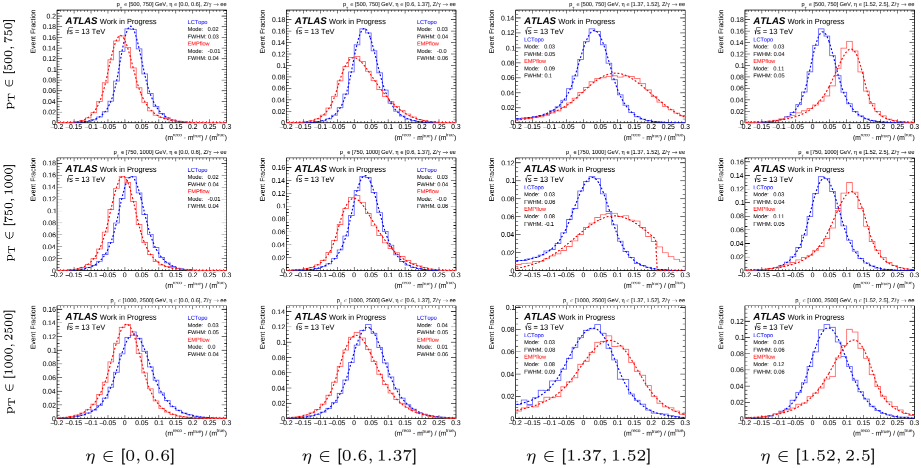


$\eta \in [0, 0.6]$ $\eta \in [0.6, 1.37]$ $\eta \in [1.37, 1.52]$ $\eta \in [1.52, 2.5]$

- Lower resolution in crack region

LCTopo
EMPflow

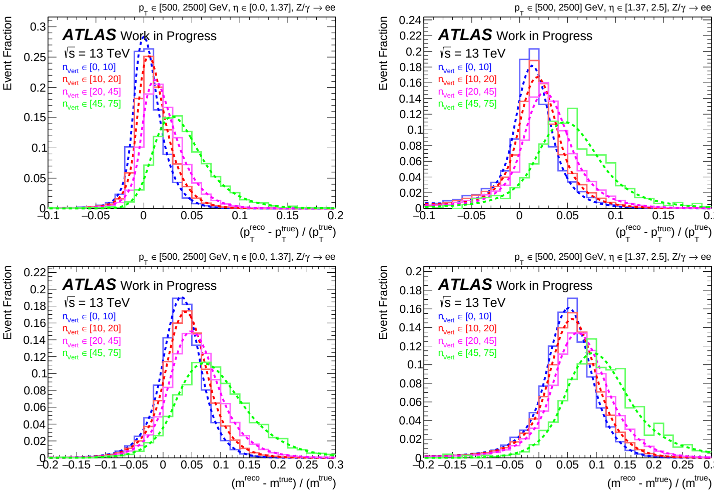
$Z \rightarrow ee$ —— Mass Resolution



- Decent mass resolution

LCTopo
EMPflow

$Z \rightarrow ee$ — P_t/Mass Resolution — n_{Vertices} — LCTopo



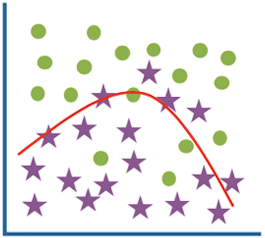
$\eta \in [0.0, 1.37]$ $\eta \in [1.37, 2.5]$

- Overestimation of mass & p_T for high n_{Vertices}

Goal: identify Z $\rightarrow$ ee processes over other background jets

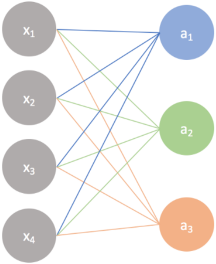
Approach: neural network tagger

1. Dataset (labeled data)
2. Architecture
   - Mathematical function with adjustable parameters
   - Predicts class accordance depending on input vector
3. Adjustment strategy for the parameters (Backpropagation)



A simple Neural Network

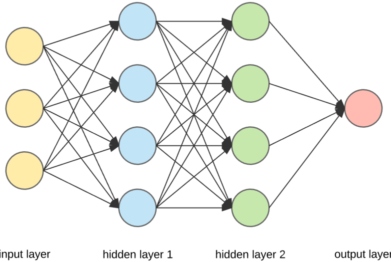
- $f: \mathrm{V}_{\mathrm{input}} \rightarrow \mathrm{V}_{\mathrm{Prediction}}$
- Feedforward Neural Network
- Matrix entries called weights: $w_{l,m,n}$ (l: layer, m: neuron, n: neuron previous layer)



$$\begin{bmatrix} w_1 & w_2 & w_3 & w_4 \\ w_1 & w_2 & w_3 & w_4 \\ w_1 & w_2 & w_3 & w_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} b \\ b \\ b \end{bmatrix} = \begin{bmatrix} w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \\ w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \\ w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \end{bmatrix} \xrightarrow[activation]{} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

Complexity:

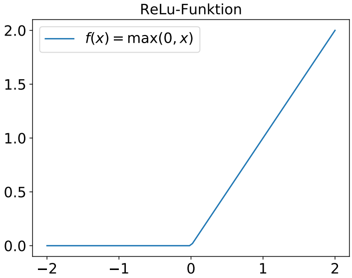
- Modular structure allows for adjusting complexity



$$V_{input} \xrightarrow{M^{(3x4)}} V_{h1} \xrightarrow{M^{(4x4)}} V_{h2} \xrightarrow{M^{(4x1)}} V_{prediction}$$

Activation Function:

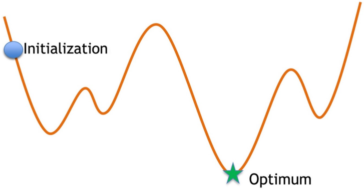
- Non-linear
- Easy to differentiate



Gradient Descent

First-order iterative optimization algorithm for finding a local minimum

- Most optimizers based on gradient descent
- Gradient: Direction of greatest increase of the function
- Gradients could be computed analytically, but not feasible
- Step in direction of negative gradient
- Weights updated with mean gradient of multiple samples (batch)
- Iterate for multiple epochs over dataset



$$L(\mathbf{y}, \hat{\mathbf{y}}, \mathbf{W}) = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

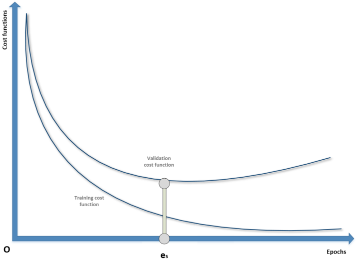
$$\nabla_{\mathbf{W}} L_{\{\mathbf{x},\mathbf{y}\}}(\mathbf{W}) = \begin{bmatrix} \frac{\partial L}{\partial w_{0,0,0}} \\ \vdots \\ \frac{\partial L}{\partial w_{l,m,n}} \end{bmatrix}$$

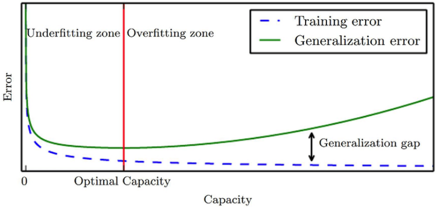
$$\mathbf{W}' = \mathbf{x} - \alpha \nabla_{\mathbf{W}} L(\mathbf{W})$$

L : Loss function
$\mathbf{W}$: Weights
$\hat{\mathbf{y}}$: Prediction
$\mathbf{y}$: Target
$\mathbf{x}$: Input
α : Learning rate

Overfitting

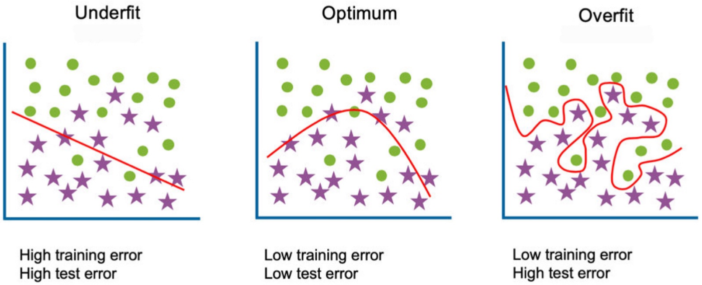
- Training data set used to compute gradients
- Validation data set provides an unbiased evaluation of a model performance
  - Used for hyperparameter optimization (e.g. number of neurons, learning rate, ...)
- Capacity of the network depends on number of layers and nodes





Generalization

- Generalization: performance of the NN on data not used for the training
- Overfitting: learned features only valid for the training data set
- Check for overfitting by splitting the data set into train and validation set



The boosted Z to ee tagger

Network Inputs

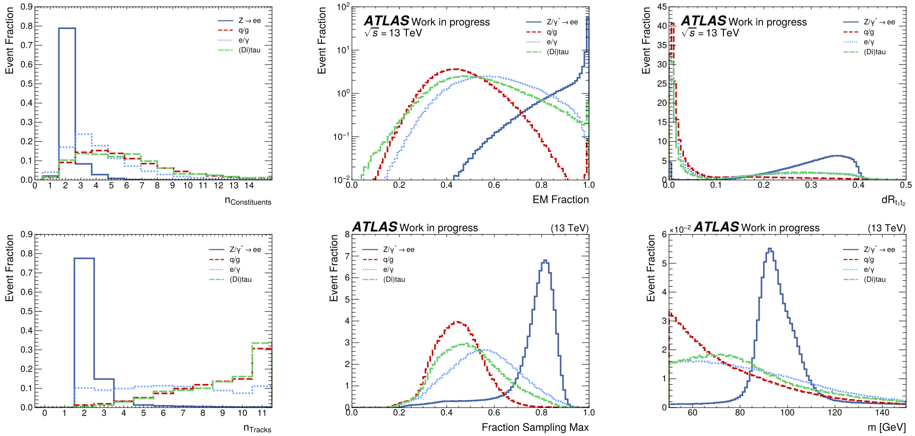
- MC-Simulation with Sherpa 2.2.11
- Selections of $\mathrm{Z} \rightarrow \mathrm{ee}$ candidate Small-R jets:
  - $\mathrm{N}_{\mathrm{Trks}} \geq 2$
  - $\mathrm{m}_{\mathrm{j}} \in [50, 150]$ GeV
  - $\mathrm{p}_{\mathrm{T,\ j}} > 500$ GeV
- ✓: Included in standard electron identification

ATLAS Work in progress

Class	#Events
$\mathrm{Z} \rightarrow \mathrm{ee}$	1,994,086
e/γ	1,663,522
QCD	2,235,635
(Di)tau	372,562

Input parameter	Definition
dR tt	Radial distance between tracks with highest and second highest p_T
Balance	$\frac{p_{T,\mathrm{Trk1}} - p_{T,\mathrm{Trk2}}}{p_{T,\mathrm{Trk1}} + p_{T,\mathrm{Trk2}}}$
Charge	Sum of charge of the two tracks with the highest p_T
EM Fraction	Fraction of energy deposited in the EM calorimeter compared to the total energy of the jet
Eta Bin	Observable that describes in which eta interval the dielectron candidate jet is located
Frac Sampling Max	The ratio of a partial jet energy (contained in the calorimeter layer with the most energy of that jet) to the total jet energy
Mass	Mass of the dielectron candidate jet
N Constituents	Number of calorimeter clusters within the dielectron candidate jet
N Trks	Number of tracks
Width	Width of the dielectron candidate jet
pT Bin	Observable that describes in which p_T interval the dielectron candidate jet is located
✓ Trk1 (Trk2) d0	Transverse impact parameter of the leading (subleading) track
✓ Trk1 (Trk2) d0sig	Significance of the transverse impact parameter of the leading (subleading) track
✓ Trk1 (Trk2) dEta	$\Delta\eta$ between track 1 (2) and closest calorimeter cluster
✓ Trk1 (Trk2) dPhi	$\Delta\phi$ between track 1 (2) and closest calorimeter cluster
Trk1 (Trk2) dRToJet	Radial distance between track with highest (second highest) p_T to the axis of the jet
✓ Trk1 (Trk2) eProbability HT	Electron probability based on transition radiation in the TRT for the leading (subleading) track in the fat electron candidate jet
✓ Trk1 (Trk2) EOverP	$\frac{E_{\mathrm{cluster}}}{p_{T,3\mathrm{Trk1}}} \left(\frac{E_{\mathrm{cluster}}}{p_{T,\mathrm{Trk2}}}\right)$
✓ Trk1 (Trk2) z0	Longitudinal impact parameter of the leading (subleading) track
✓ Trk1 (Trk2) z0sig	Significance of the longitudinal impact parameter of the leading (subleading) track

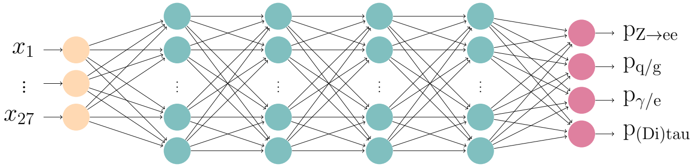
Input distributions



- Significant discrimination power between signal and background due to separation between the e^+e^- pair

Neural Network Architecture

- Feed-forward neural network (PyTorch)
- Input distributions normalized
- Hyperparameter optimized with random and grid search
  - Activation: Sigmoid
  - Hidden layer: 4
  - Nodes per layer: 200
  - Output nodes: 4 (Bkg composition analysis dependent)



Hyper Parameter Optimization

- 144 trainings with random configuration from search space
- Efficiency, accuracy and background rejection plotted in dependence of η and p_T for 20 trainings with highest accuracies

ATLAS Work in progress

Parameter	**Range**	**Mode**
lr	[0.0001, 0.1]	log
Hidden Size	[50, 500]	int
Midlayer	[1, 5]	int
Batch Size	[500, 5000]	int
Activation	ReLU Sigmoid LeakyReLU	item

20 Best Configurations

Latest Set of Hyperparameters:

- 200 Nodes
- 4 midllayer
- Sigmoid activation
- lr = 0.005
- Batchsize: 2048

ATLAS Work in progress

Learning R	Nodes	Layer	Batchsize	Activation	Accuracy	Stable
$8 \cdot 10^{-4}$	325	5	778	Sigmoid	75.24	✗
$5 \cdot 10^{-3}$	356	5	1221	LeakyReLU	75.20	✓
$1 \cdot 10^{-3}$	178	4	2591	Sigmoid	75.19	✓
$3 \cdot 10^{-3}$	214	4	4000	ReLU	75.18	✗
$7 \cdot 10^{-3}$	202	3	2918	ReLU	75.15	✗
$1 \cdot 10^{-3}$	433	5	4565	Sigmoid	75.15	✓
$8 \cdot 10^{-3}$	244	4	1875	LeakyReLU	75.14	✓
$6 \cdot 10^{-3}$	355	3	2324	Sigmoid	75.12	✗
$4 \cdot 10^{-3}$	252	5	4850	ReLU	75.12	✗
$4 \cdot 10^{-3}$	393	3	4286	ReLU	75.12	✗
$3 \cdot 10^{-3}$	199	2	2621	Sigmoid	75.11	✓
$2 \cdot 10^{-3}$	237	5	4405	ReLU	75.11	✗
$3 \cdot 10^{-3}$	138	5	2589	ReLU	75.10	✗
$5 \cdot 10^{-3}$	417	4	1083	Sigmoid	75.09	✗
$2 \cdot 10^{-3}$	460	2	2518	Sigmoid	75.09	✓
$7 \cdot 10^{-3}$	111	5	1907	Sigmoid	75.08	✓
$3 \cdot 10^{-3}$	228	4	2673	Sigmoid	75.07	✗
$5 \cdot 10^{-3}$	369	3	733	ReLU	75.07	✗
$2 \cdot 10^{-3}$	170	2	4603	LeakyReLU	75.06	✓
$2 \cdot 10^{-2}$	251	5	1917	LeakyReLU	75.05	✓

Variable Importance

Mutual Information:
Measure of the mutual dependence between two random variables (Each input and output)

Procedure:

1. Train 10x
2. Determine μ_{Acc} and σ_{Acc}
3. Drop input with lowest importance (Mutual Info)
4. (Repeat 1-3)

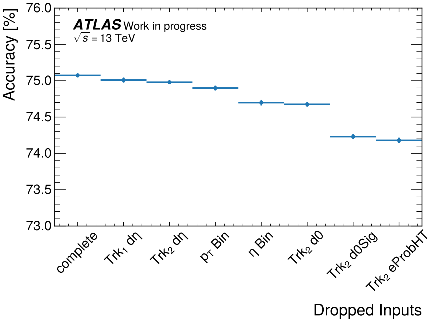
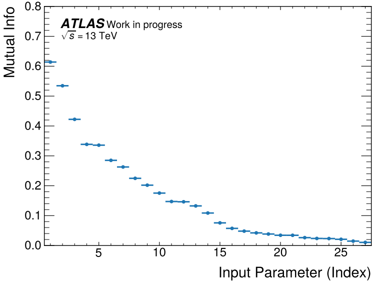


Figure: Left to right: Inputs missing additionally

Latest Set of Input Parameters

- Trk_1 $\mathrm{d}\eta$ and Trk_2 $\mathrm{d}\eta$ removed
- $\mathrm{p_T}$ Bin and η Bin needed for sample weighting
- Number of inputs reduced to 27

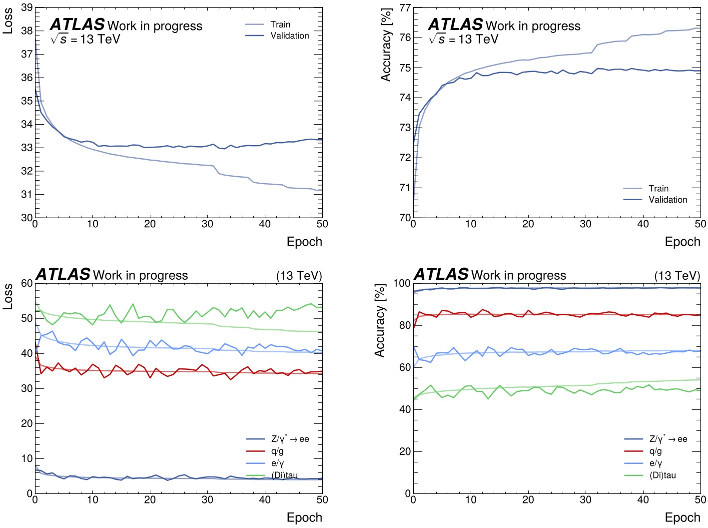


ATLAS Work in progress

Index	Parameter	Importance
1	EM Fraction	0.614
2	N Trks	0.535
3	FracSamplingMax	0.422
4	N Constituents	0.339
5	dR tt	0.336
6	Trk2 dRToJet	0.285
7	Balance	0.263
8	Mass	0.225
9	Trk1 dPhi	0.202
10	Trk1 dRToJet	0.175
11	Trk2 dPhi	0.147
12	Width	0.146
13	Trk2 E/P	0.133
14	Trk1 E/P	0.109
15	ChargeSum	0.076
16	Trk1 d0	0.057
17	Trk1 z0	0.048
18	Trk1 z0Sig	0.042
19	Trk2 z0	0.038
20	Trk2 z0Sig	0.034
21	Trk1 d0Sig	0.034
22	Trk1 eProbabilityHT	0.026
23	Trk2 d0Sig	0.024
24	Trk2 eProbabilityHT	0.023
25	Trk2 d0	0.021
26	Eta Bin	0.015
27	pT Bin	0.010

Loss and Accuracy Curves

- Overfitting for Tau (smallest sample)
- Only small increase of val acc with reduced lr
- Network of epoch with best val acc saved for further analysis

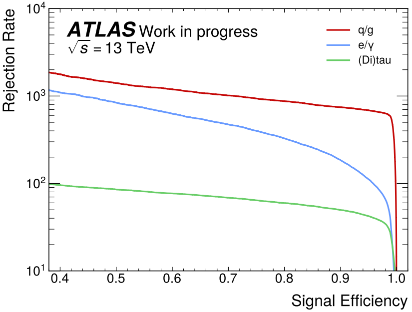


Separation Strength

ATLAS Work in progress (13 TeV)

True Label \ Predicted Label	Signal	EM	QCD	(Di)tau
Signal	98	1.4	0.03	0.79
EM	0.97	68	20	11
QCD	0.15	6.7	85	8.2
(Di)tau	1.6	20	29	49

True and False Rates [%]



- High discrimination power between signal and backgrounds
- Significant separation between the different background classes

Neural Network Discriminant

- Discriminant transforms 4-dim output for binary decision (Sig/Bkg)
- Define WP with signal efficiency of $\varepsilon = 0.95$ in every $\mathrm{p_T}$ bin
- Performance characterized by:

$$\varepsilon = \frac{TP}{TP + FN} \qquad \text{(Efficiency)}$$

$$ACC = \frac{TP}{TP + FP} \qquad \text{(Accuracy)}$$

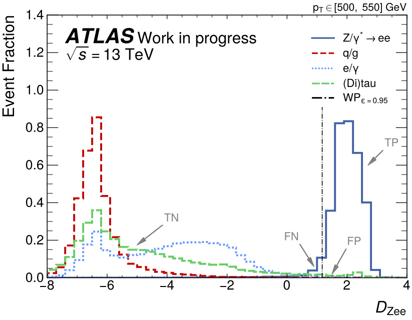
$$REJ = \frac{TP + FP}{FP} \qquad \text{(Rejection Rate)}$$

T : $True$, F : $False$, P : $Positive$, N : $Negative$

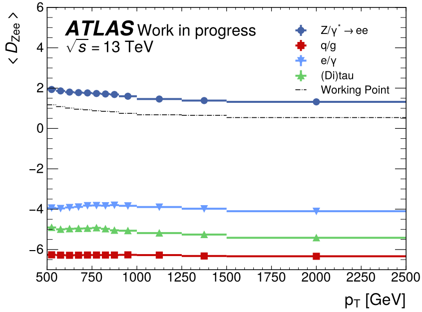
$$D_{\mathrm{Zee}} = \ln\left(\frac{(1 - f_{Zee}) \cdot p_{Zee}}{f_{QCD} \cdot p_{QCD} + f_{e/\gamma} \cdot p_{e/\gamma} + f_{\tau} \cdot p_{\tau}}\right)$$

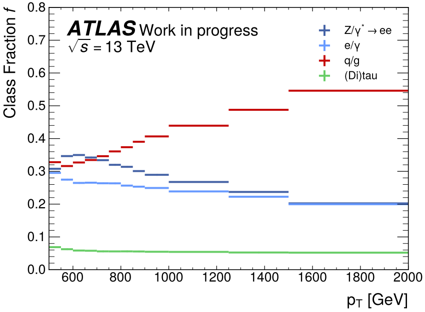
p : Probability

f : Class fraction



Deeplearning Score

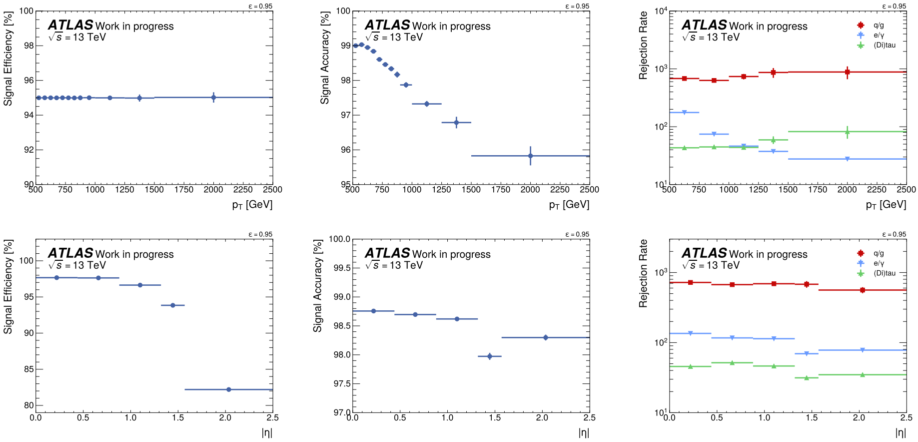




- Mean D_{Zee} of the four classes binned in p_t
- Discrimination between signal and background degrades for high p_t

- q/g events harder than the other events

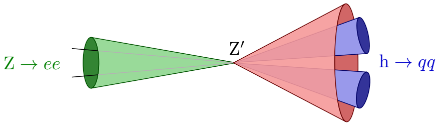
Neural Network Performance

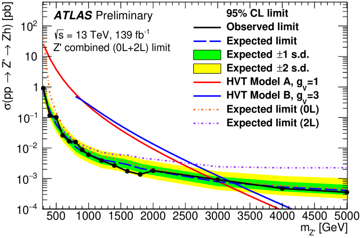


- Accuracy decreased for very high $\mathrm{p_T}$ due to merging of the clusters
- Slightly decreased accuracy for $\eta \in [1.32, 1.57]$ due to supply wiring of the solenoid

Concluding Remarks

- LCTopo at constituent scale better suited than EMPflow
  - Derive JEC (to correct for n_{Vertices} & η dependence)
  - Compare to EMTopo and EMPflow at constituent scale
  - Compare p_T/mass resolution of Z $\rightarrow$ ee jets to resolution of di-electron objects (i.e. use standard E/γ approach)
- Next steps: testing the tagger on analysis level
  - Finalize development of the tagger
  - Search for W'/Z' decay as Small-Radius jet and Large-Radius jet
  - Determine updated $\sigma \times \mathrm{BR}$ at 95% CL exclusion limits and compare with standard approach





ATLAS-CONF-2020-043

Backup

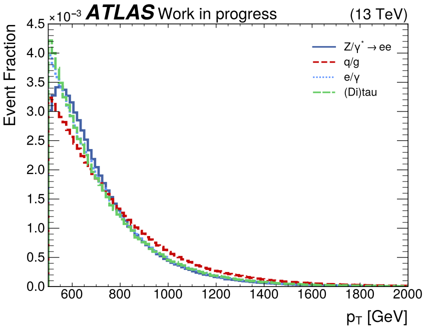
References

CMS Collaboration
Search for heavy resonances that decay into a vector boson and a Higgs boson in hadronic final states at $\sqrt{s} = 13$ TeV
The European Physical Journal C (2017)

ATLAS Collaboration
Search for heavy resonances decaying into a Z boson and a Higgs boson in final states with leptons and b-jets in 139 fb^{-1} of pp collisions at $\sqrt{s} = 13 TeV$ with the ATLAS detector
ATLAS-CONF-2020-043 (2020)

ATLAS Collaboration
ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV
The European Physical Journal C (2019)

Processes



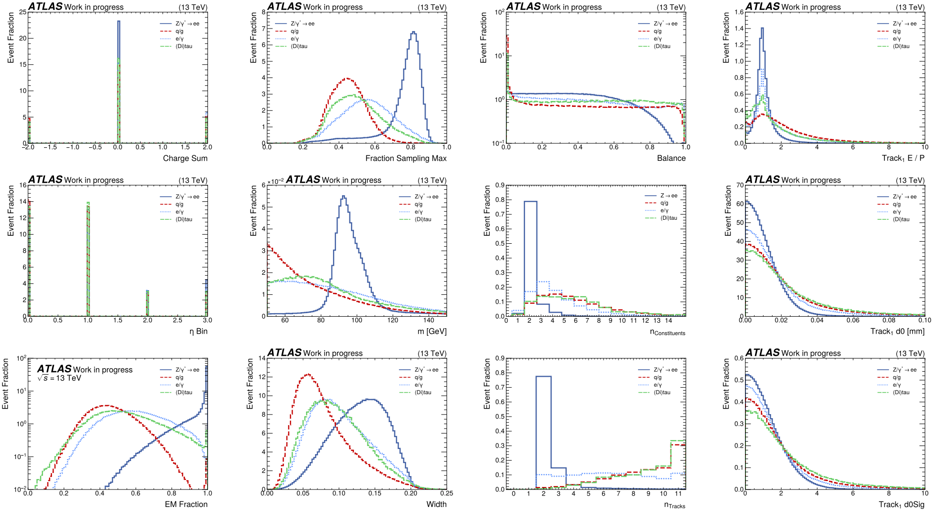
ATLAS Work in progress

Class	#Events
$\mathrm{Z} \rightarrow \mathrm{ee}$	1, 994, 086
e/γ	1, 663, 522
QCD	2, 235, 635
(Di)tau	372, 562

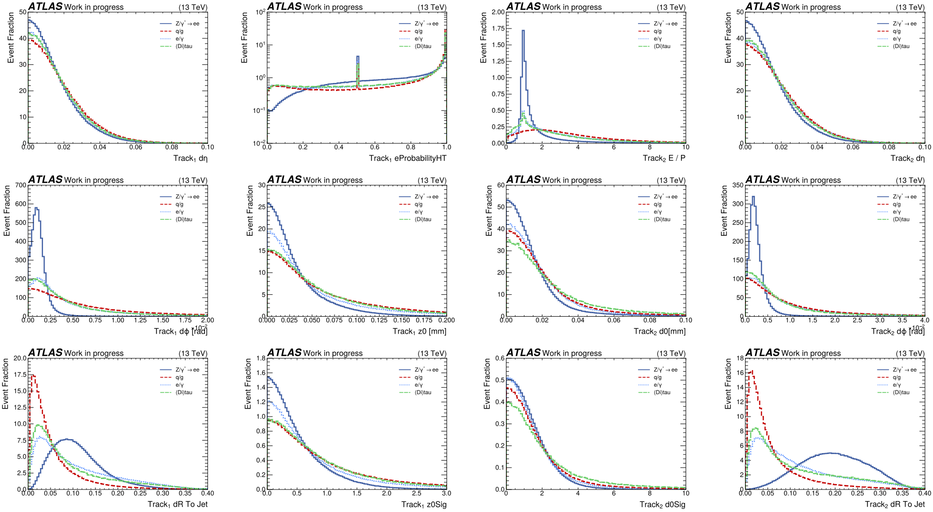
ATLAS Work in progress

Process	Generator
$Z/\gamma^* \rightarrow e^+e^-$	SHERPA2.2.11
$Z/\gamma^* \rightarrow \tau^+\tau^-$	SHERPA2.2.11
$W^\pm \rightarrow e^\pm\nu$	SHERPA2.2.11
$W^\pm \rightarrow \tau^\pm\nu$	SHERPA2.2.11
$Z(\rightarrow \ell^+\ell^-)\gamma$	SHERPA2.2.11
$W^\pm(\rightarrow \ell\nu)\gamma$	SHERPA2.2.11

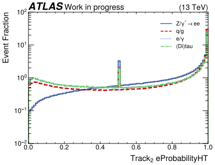
Input Distributions

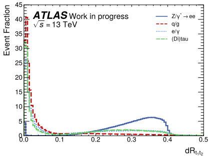


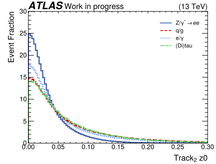
Input Distributions

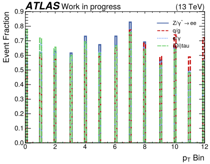


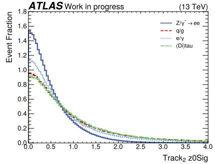
Input Distributions







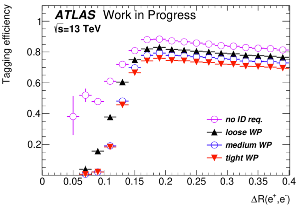


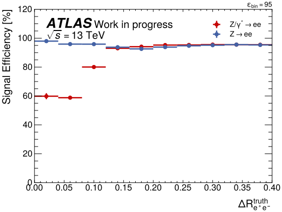


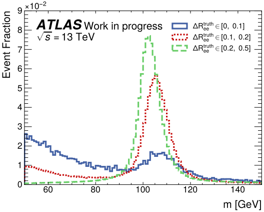
ATLAS Work in progress

Input parameter	Definition
dR tt	Radial distance between tracks with highest and second highest p_T
Balance	$\frac{p_{T,\mathrm{Trk1}}-p_{T,\mathrm{Trk2}}}{p_{T,\mathrm{Trk1}}+p_{T,\mathrm{Trk2}}}$
Charge	Sum of charge of the two tracks with the highest p_T
EM Fraction	Fraction of energy deposited in the EM calorimeter compared to the total energy of the jet
Eta Bin	Observable that describes in which eta interval the dielectron candidate jet is located
Frac Sampling Max	The ratio of a partial jet energy (contained in the calorimeter layer with the most energy of that jet) to the total jet energy
Mass	Mass of the dielectron candidate jet
N Constituents	Number of calorimeter clusters within the dielectron candidate jet
N Trks	Number of tracks
Width	Width of the dielectron candidate jet
pT Bin	Observable that describes in which p_T interval the dielectron candidate jet is located
✓ Trk1 (Trk2) d0	Transverse impact parameter of the leading (subleading) track
✓ Trk1 (Trk2) d0sig	Significance of the transverse impact parameter of the leading (subleading) track
✓ Trk1 (Trk2) dEta	$\Delta\eta$ between track 1 (2) and closest calorimeter cluster
✓ Trk1 (Trk2) dPhi	$\Delta\phi$ between track 1 (2) and closest calorimeter cluster
Trk1 (Trk2) dRToJet	Radial distance between track with highest (second highest) p_T to the axis of the jet
✓ Trk1 (Trk2) eProbability HT	Electron probability based on transition radiation in the TRT for the leading (subleading) track in the fat electron candidate jet
✓ Trk1 (Trk2) EOverP	$\frac{E_{\mathrm{cluster}}}{p_{T,\mathrm{Trk1}}}$ $\left(\frac{E_{\mathrm{cluster}}}{p_{T,\mathrm{Trk2}}}\right)$
✓ Trk1 (Trk2) z0	Longitudinal impact parameter of the leading (subleading) track
✓ Trk1 (Trk2) z0sig	Significance of the longitudinal impact parameter of the leading (subleading) track

Performance $\mathrm{dR}_{\mathrm{ee}}^{\mathrm{truth}}$







- Standard electron reco & ID
- Efficiency degraded for low $\mathrm{dR}_{\mathrm{ee}}^{\mathrm{truth}}$

- NN performance degraded for low $\mathrm{dR}_{\mathrm{ee}}^{\mathrm{truth}}$ for signal including γ^*
- Low efficiency due to rejected γ^* events in signal region

- Mass distribution for signal including γ^*
- Low angular separation for γ^* events
- Mass important input variable