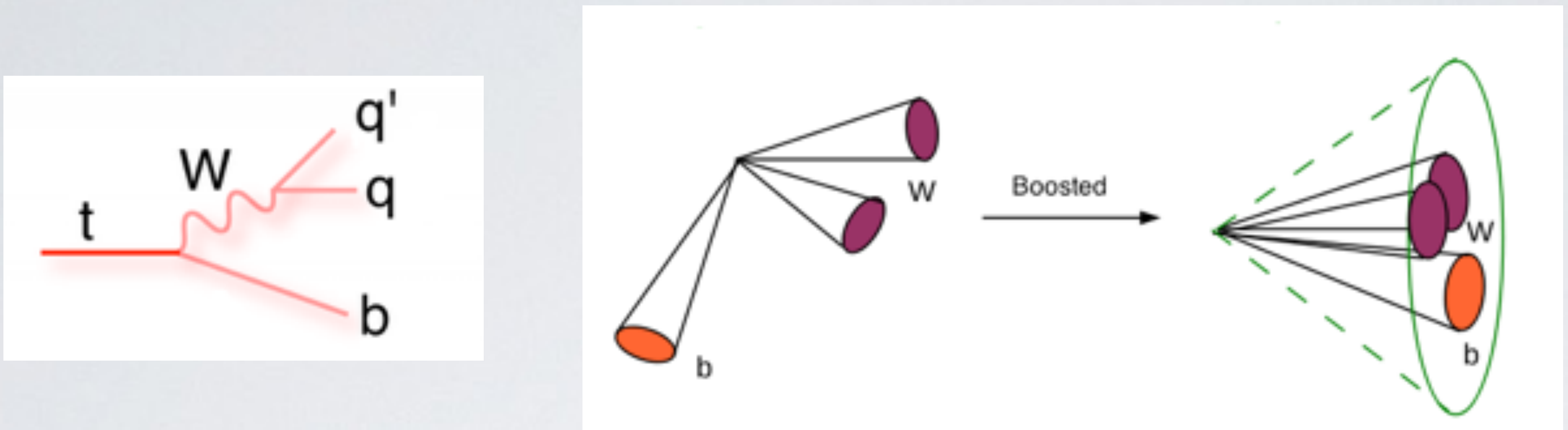
IDENTIFICATION OF BOOSTED W BOSONS AND TOP QUARKS WITH MACHINE LEARNING IN ATLAS

Ece Akıllı
Université de Genève

Joint Annual Meeting of the Swiss Physical Society and Austrian Physical Society
August 23, 2017

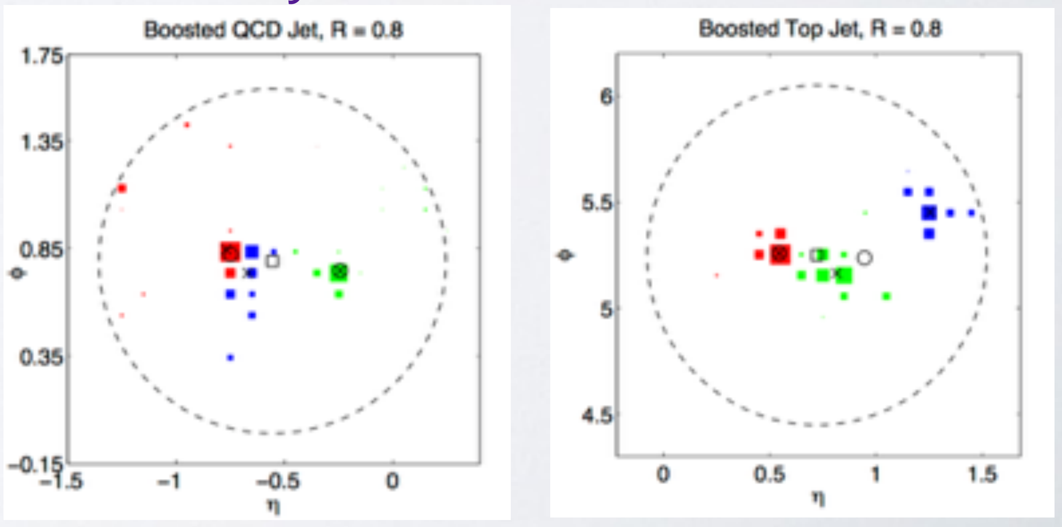
W-BOSON AND TOP-QUARK TAGGING IN ATLAS



Top-Quark Decay

- *W* bosons and top quarks have short lifetime
- Decay products of high-momentum (boosted) hadronically decaying *W* bosons and top quarks are collimated
- Resolved object identification techniques are not successful
- Construct large-radius (large-R) jets
- Use substructure information to identify the W boson and top quark within dijet background

Jet Substructure

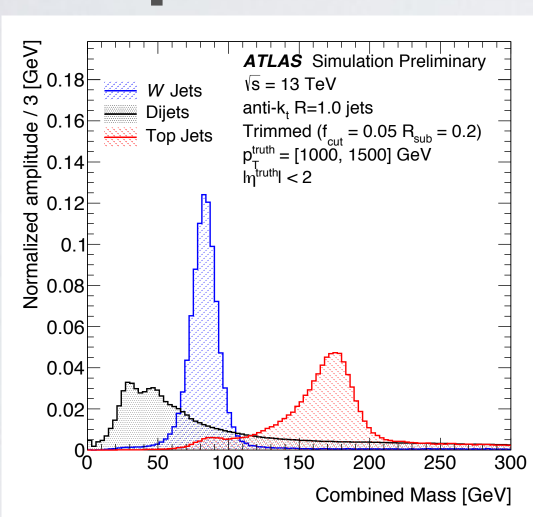


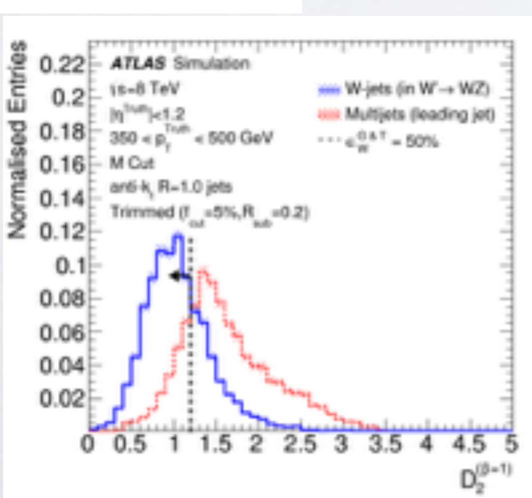
APPLICATION OF BDTS AND DNNS TO W AND TOP TAGGING USING HIGH-LEVEL FEATURES

ATLAS-CONF-2017-064

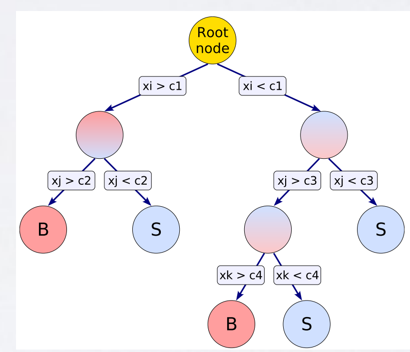
- Construct a classifier by using substructure variables
  - Combine available information to obtain good discrimination
- Two Machine Learning (ML) techniques in parallel
  - Boosted Decision Tree (BDT)
  - Deep Neural Network (DNN)
- Train binary classifiers: W/top vs Dijet

Input variables



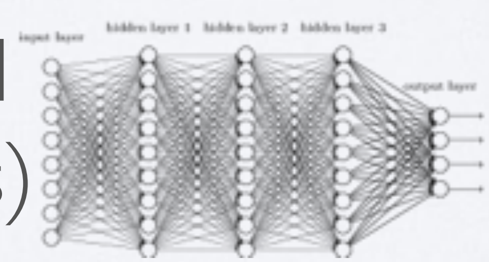


BDT (TMVA)

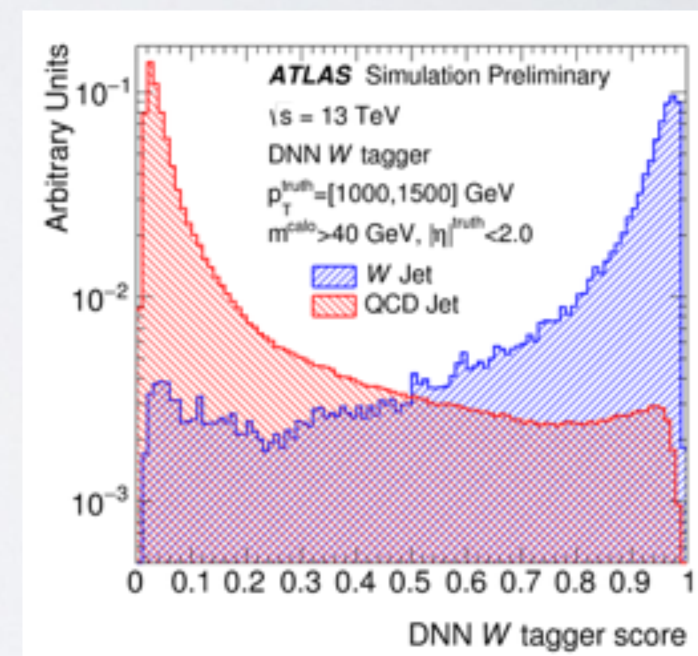


or

DNN (Keras)



Binary classifier



- BDT W
- BDT Top
- DNN W
- DNN Top

STRATEGY

- Split Monte Carlo (MC) simulation samples in training and testing sets

- Optimize BDT, DNN using MC
  - Set of inputs
  - Architecture and training hyper-parameters

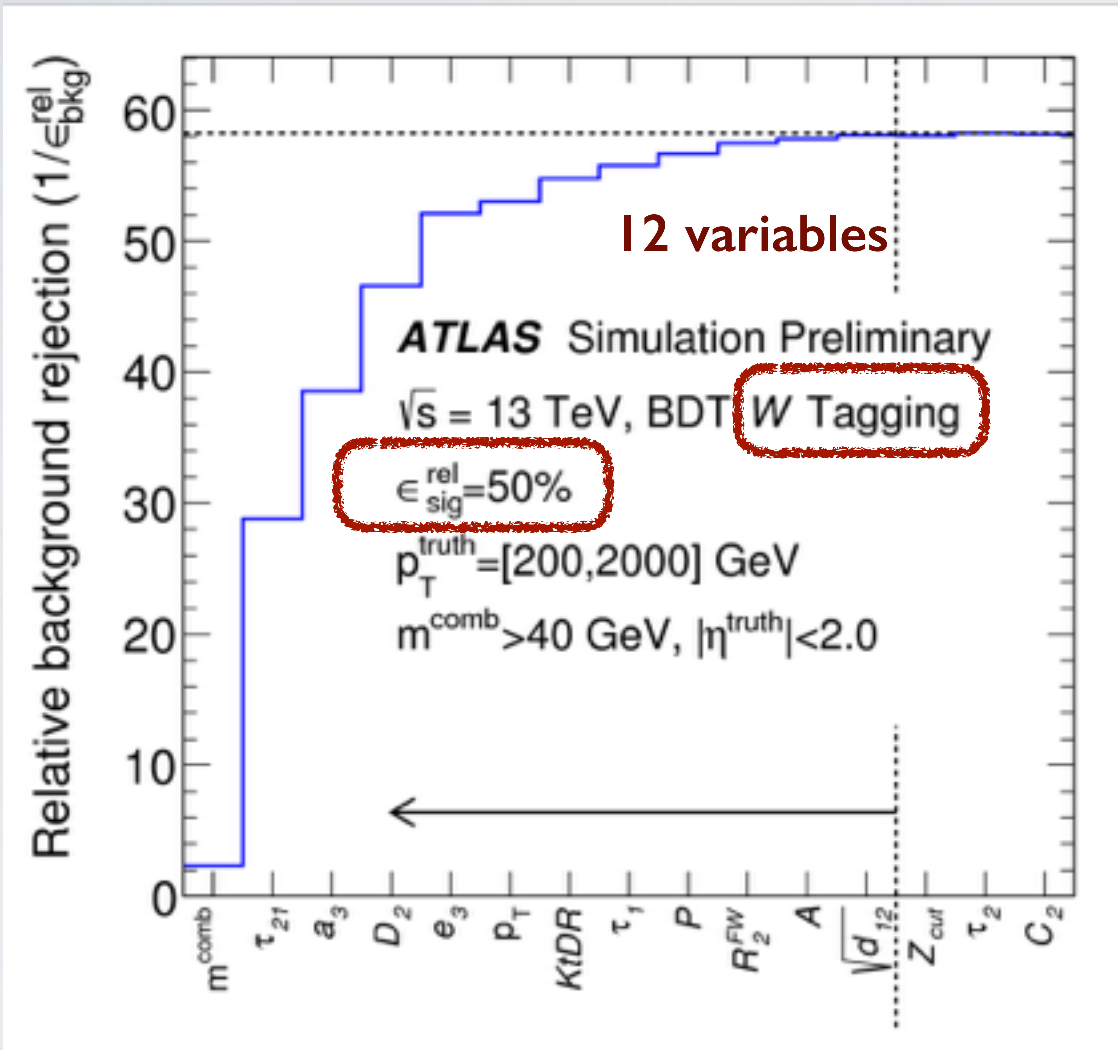
- Performance comparison of ML taggers with reference taggers in MC
  - Current taggers in ATLAS
    - 2-variable taggers
    - HEPTopTagger (Only for tops)
    - Shower Deconstruction (Only for tops)
  - BDT tagger
  - DNN tagger

- Study the performance in data

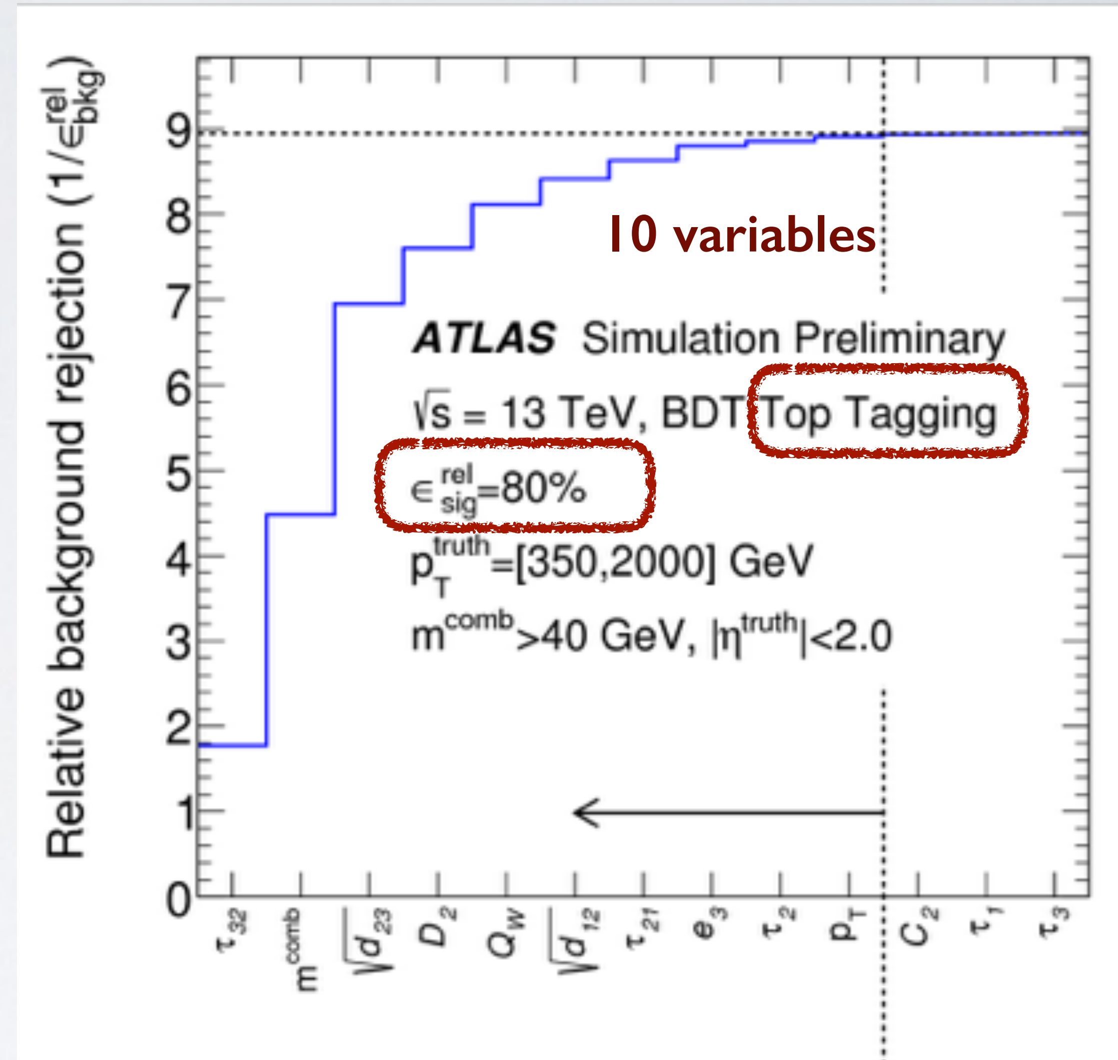
OPTIMIZATION AND PERFORMANCE STUDIES IN MC

BDT TRAINING - INPUTS OPTIMIZATION

W-Boson Tagging



Top-Quark Tagging

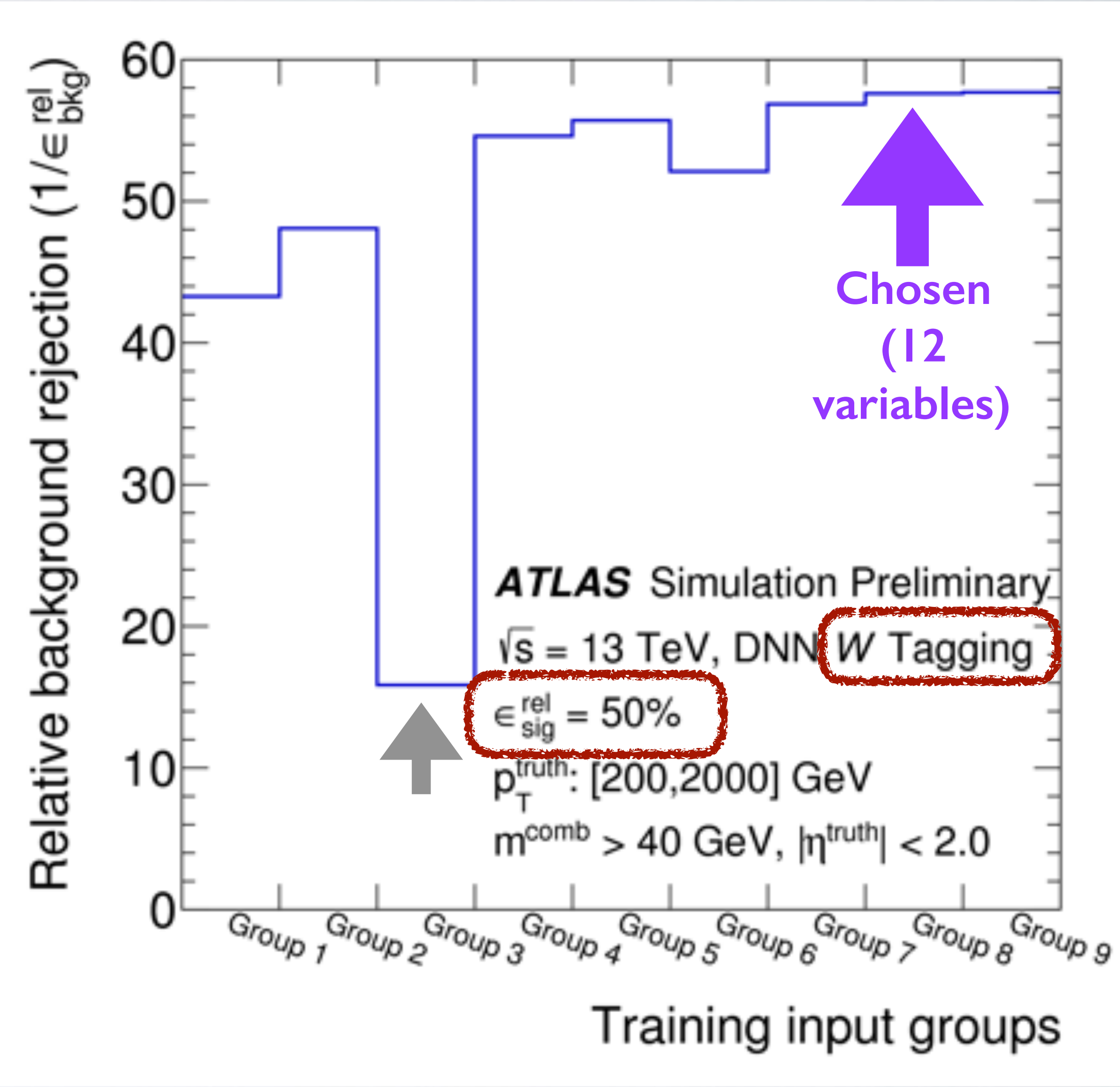


- Add variables in order of importance (improvement in rejection)
- Use a flat p_T spectrum (evaluation)
- Saturation of rejection

DNN TRAINING - INPUTS OPTIMIZATION

W-Boson Tagging

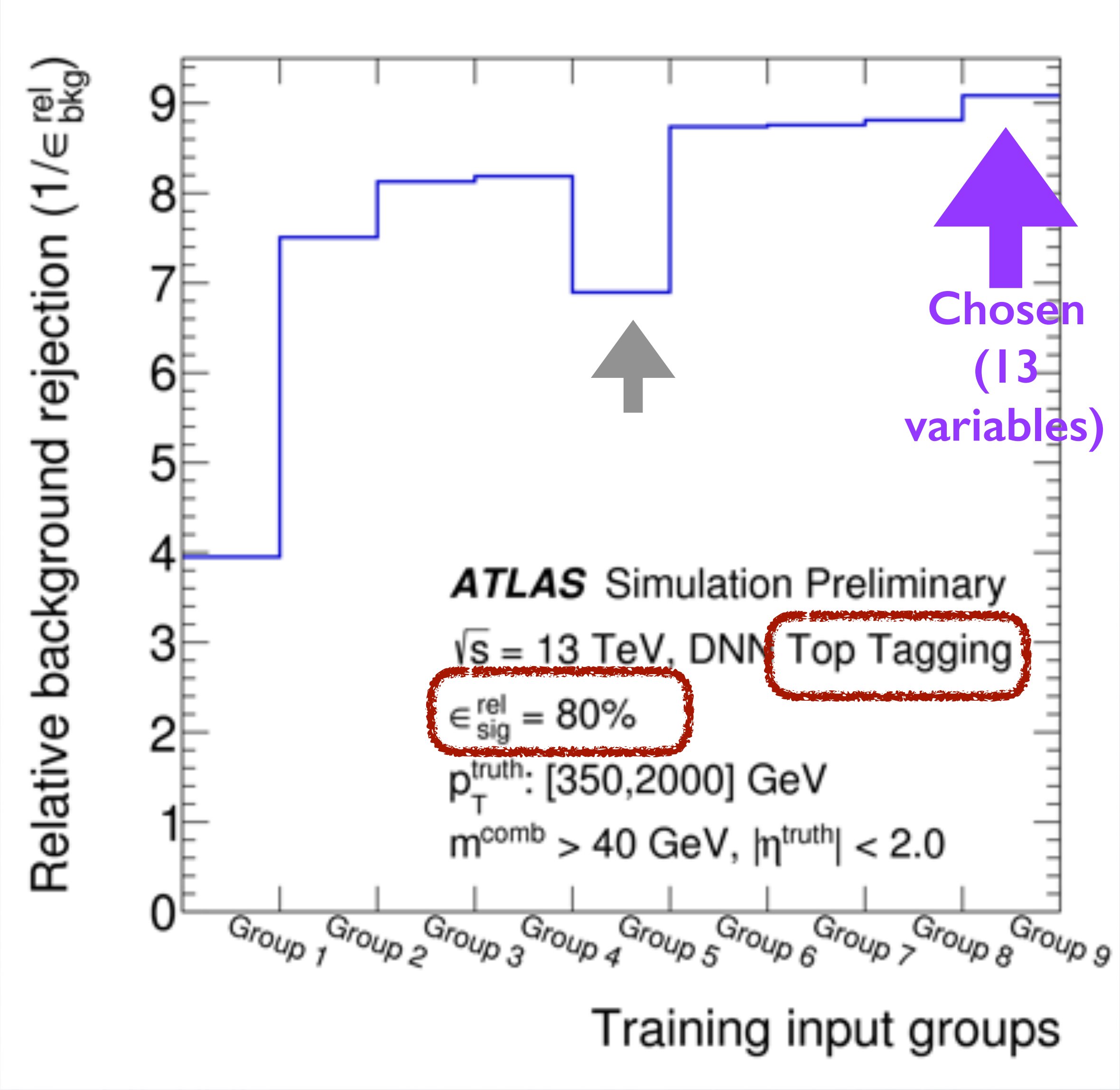
- Study different groups of input variables
- Groups are defined by varying features (scale-dependence, …)
- Use a flat p_T spectrum (evaluation)
- Choose the set with the highest background rejection
- Observed the significance of the scale and jet mass
  - Example: Group 4 = Group 3 + mass



DNN TRAINING - INPUTS OPTIMIZATION

Top-Quark Tagging

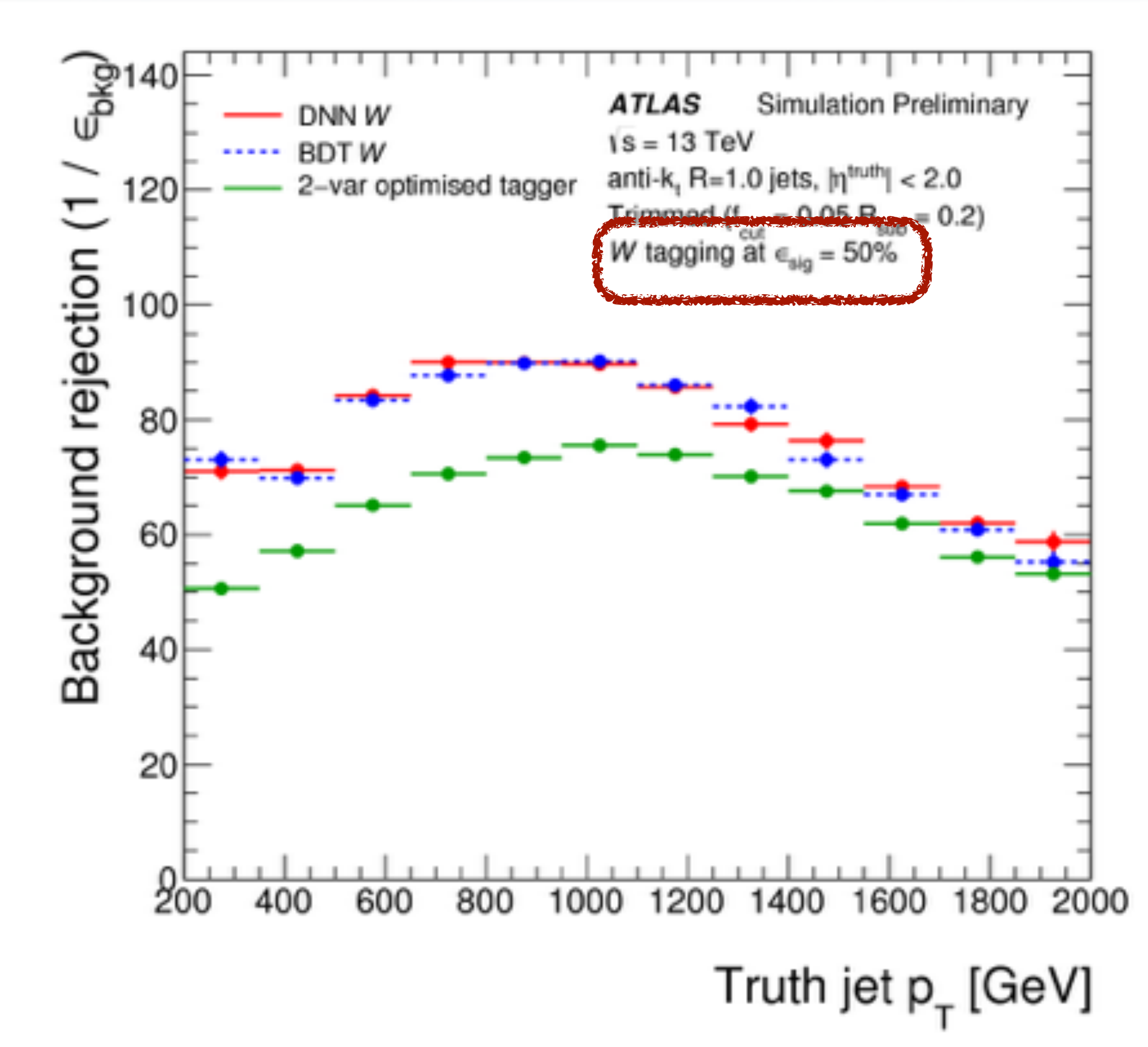
- Study different groups of input variables
- Groups are defined by varying features (scale-dependence, …)
- Use a flat p_T spectrum (evaluation)
- Choose the set with the highest background rejection
- Observed the significance of the scale and jet mass
  - Example: Group 6 = Group 5 + mass



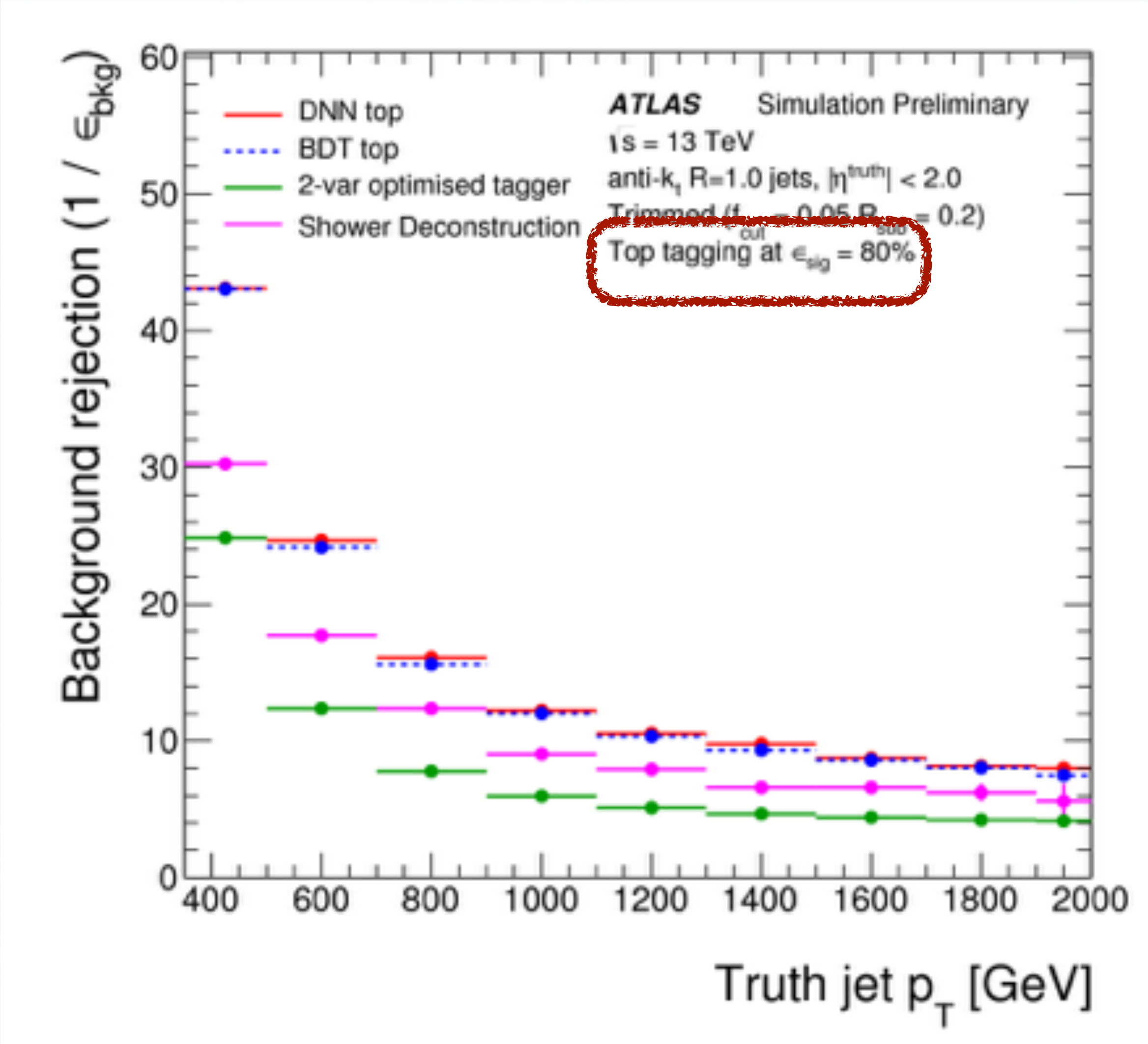
PERFORMANCE COMPARISON

Background Rejection at Fixed-Efficiency Working Point

W-Boson Tagging



Top-Quark Tagging



- BDT & DNN: Improvements observed for both W and top tagging
- Improvement is more significant for top tagging

PERFORMANCE STUDIES IN DATA

- Measure signal efficiency and background rejection in data
- Full ATLAS 2015+2016 dataset: $(36.1 - 36.7)\mathrm{fb}^{-1}$

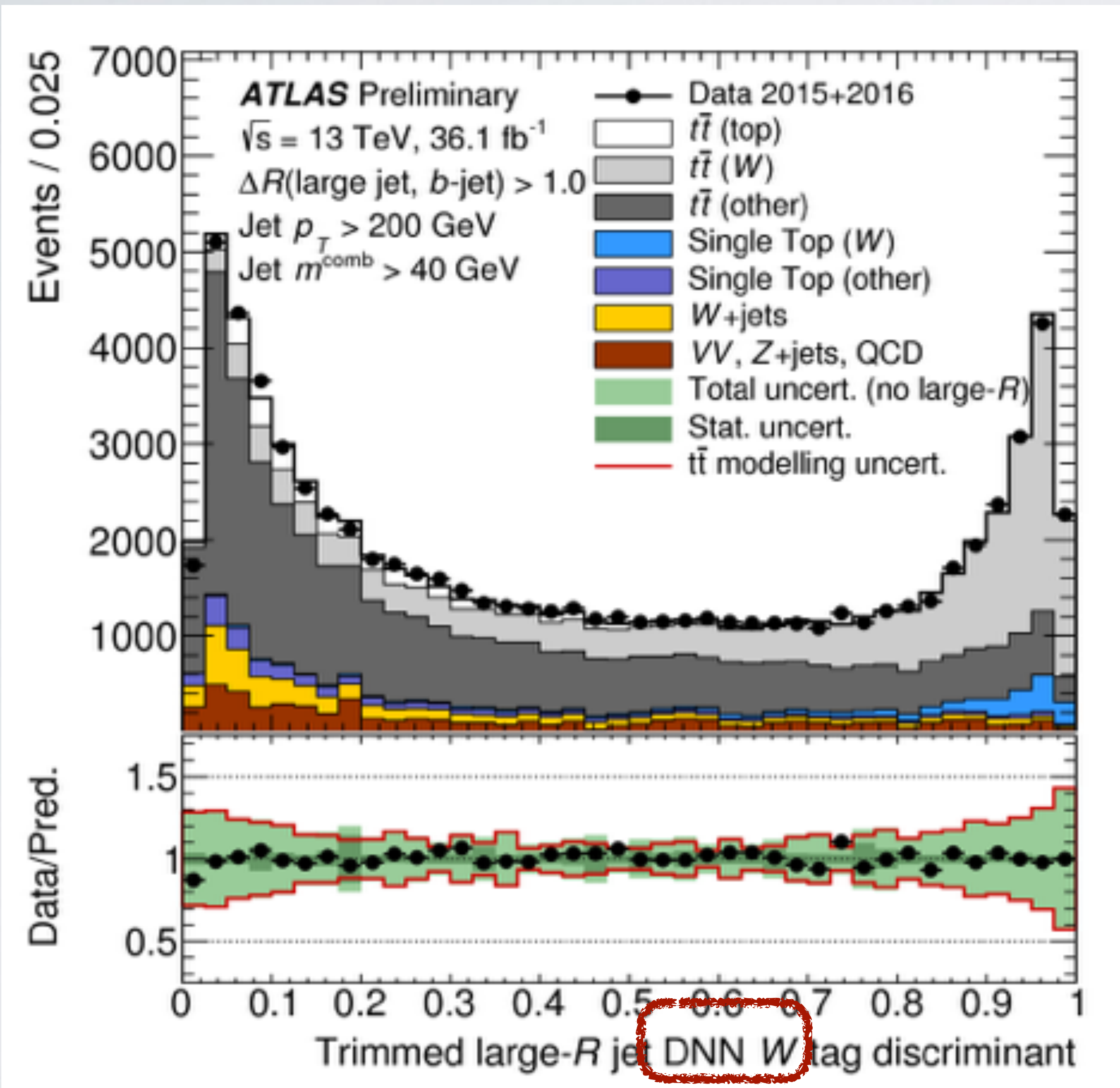
Signal: ttbar with single leptonic top decay

Background: Different background topologies, different features

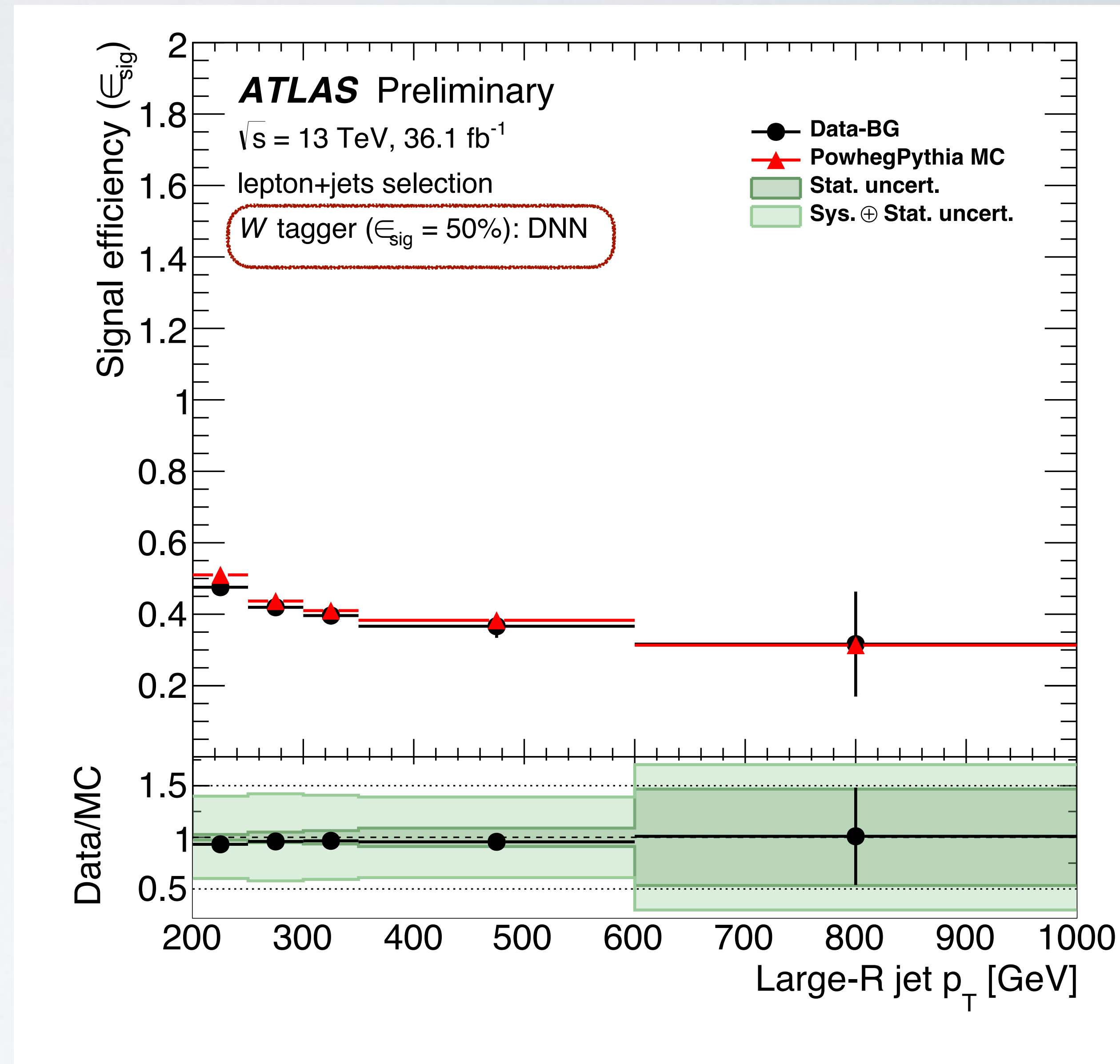
- Dijet events
- Photon + jet events

W TAGGING PERFORMANCE IN DATA - SIGNAL

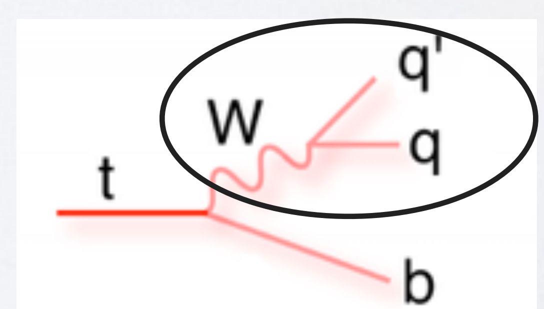
DNN Discriminant



Signal Efficiency

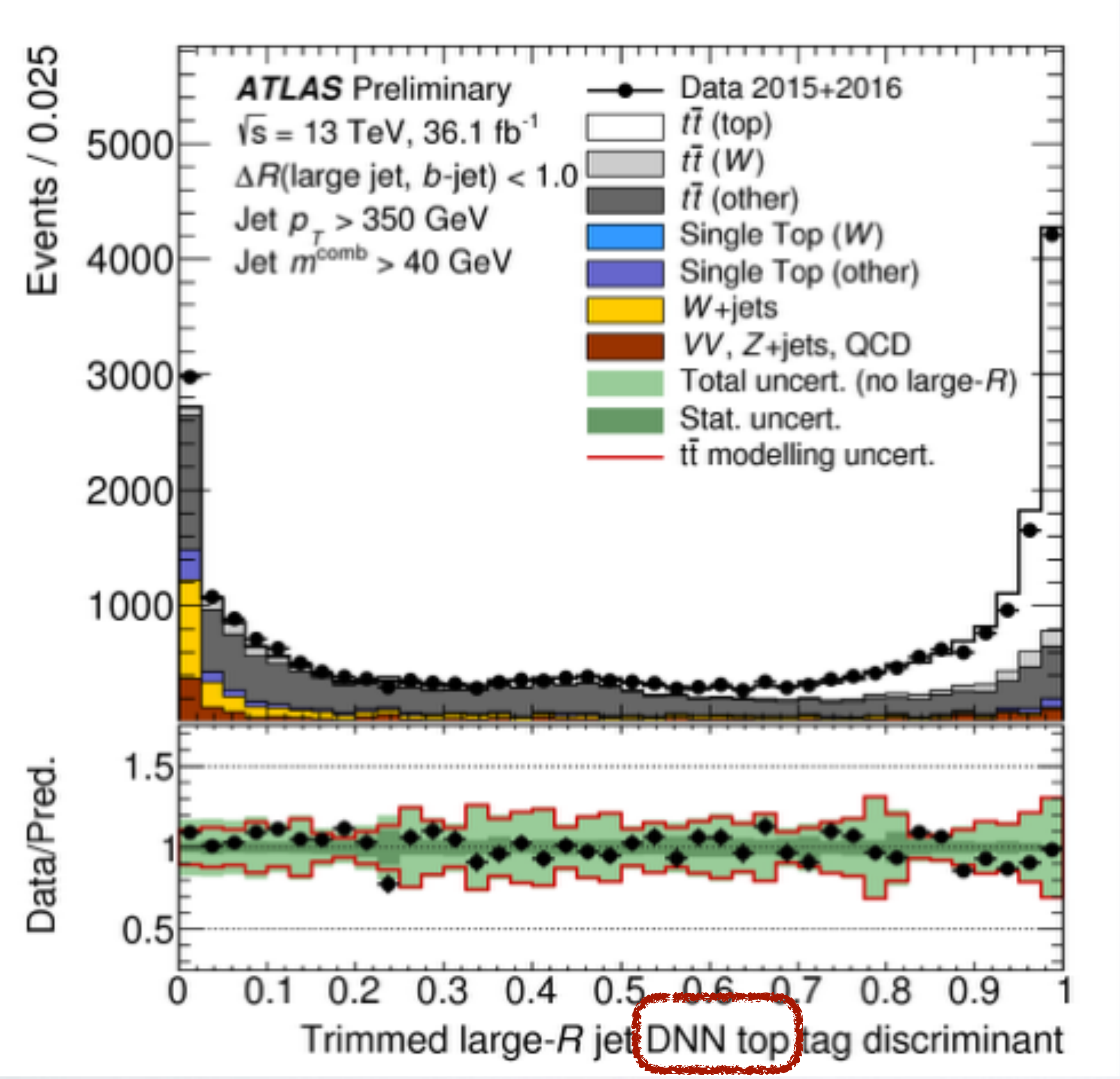


- Well modelled

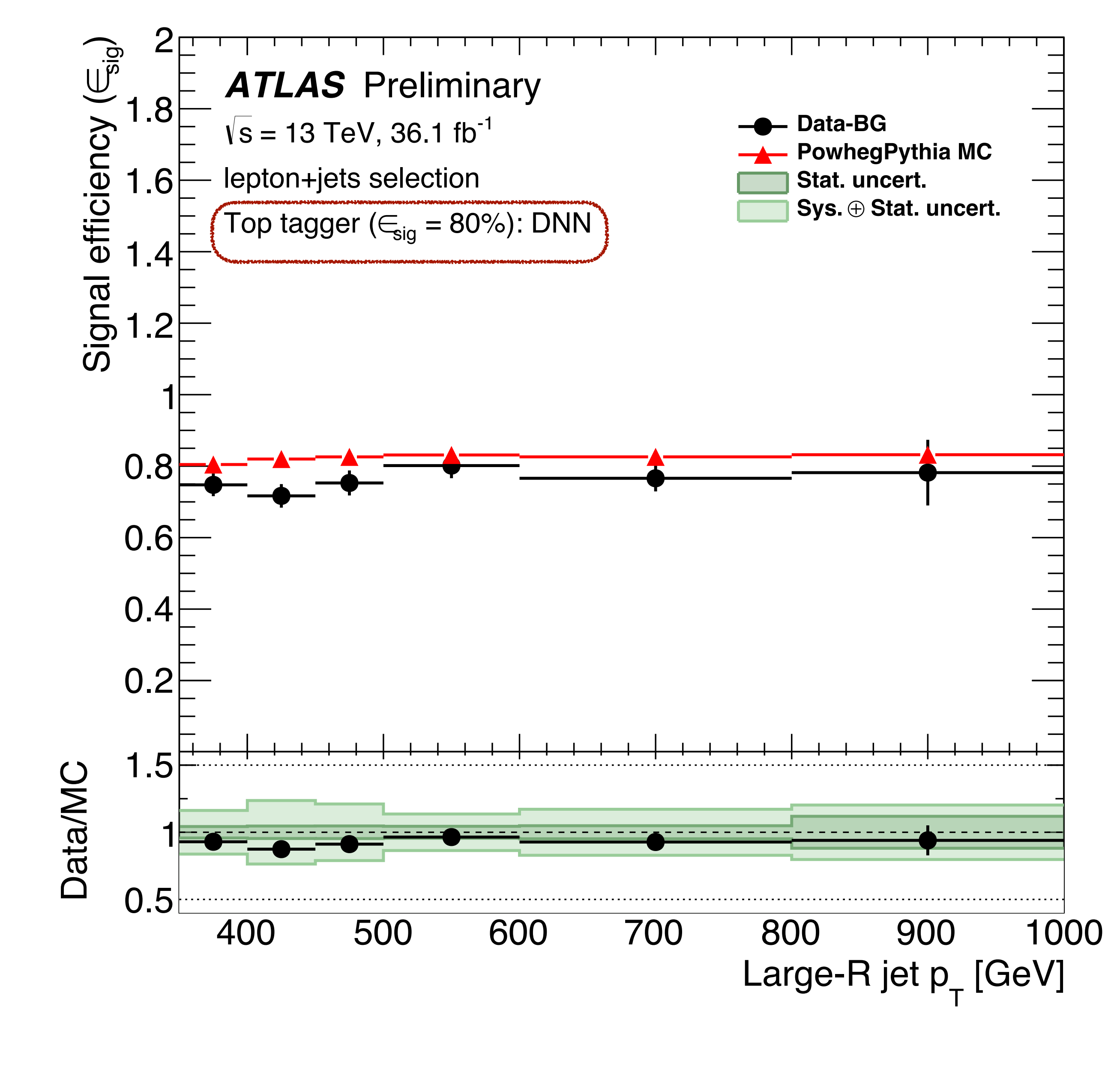


TOP TAGGING PERFORMANCE IN DATA - SIGNAL

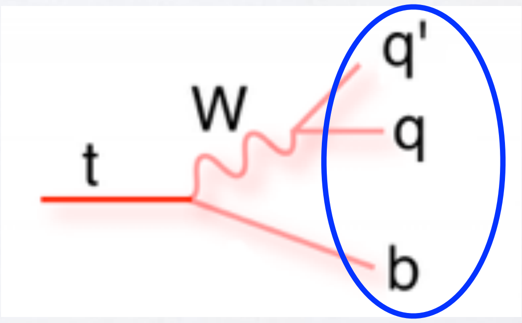
DNN Discriminant



Signal Efficiency

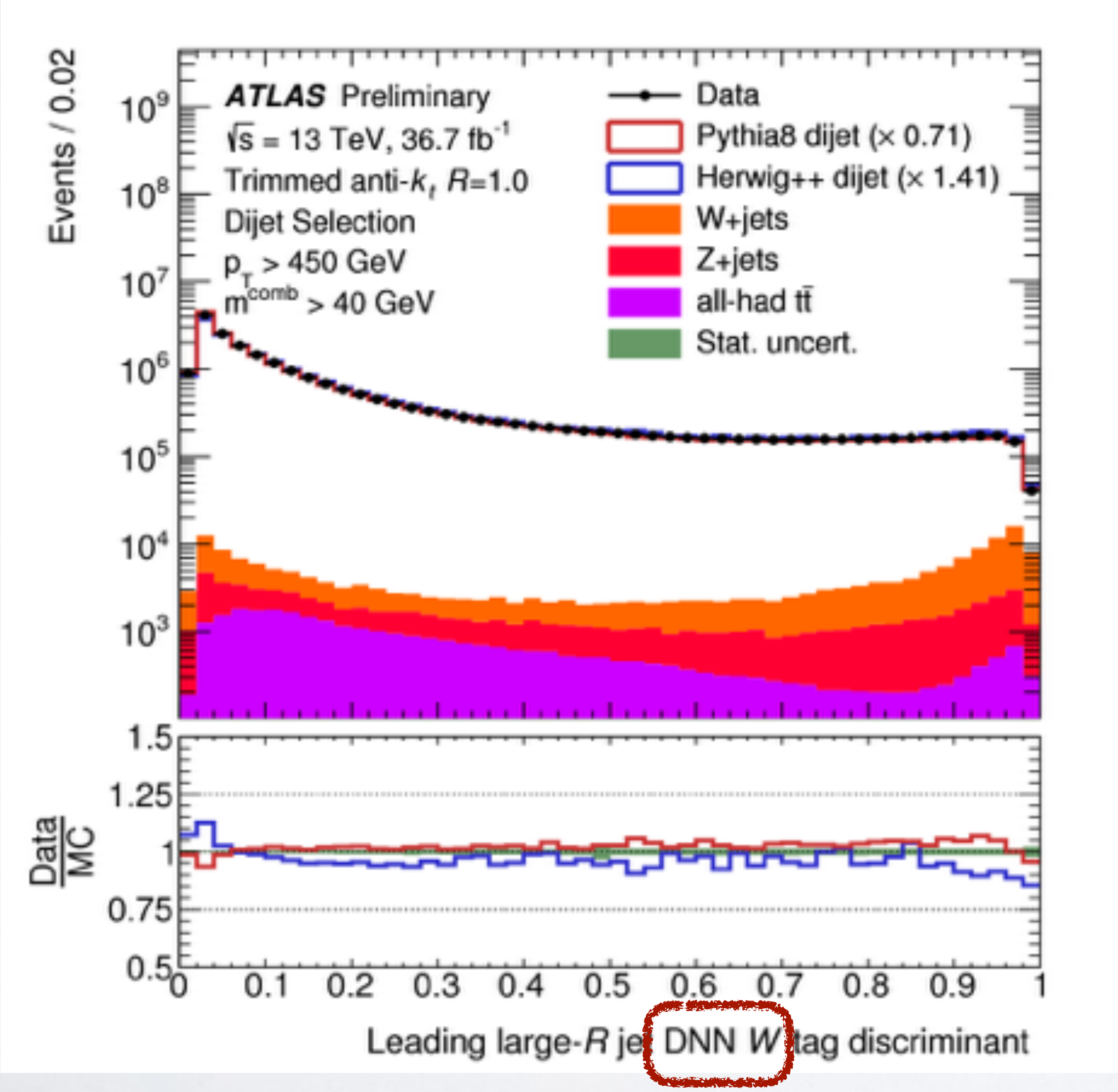


- Well modelled

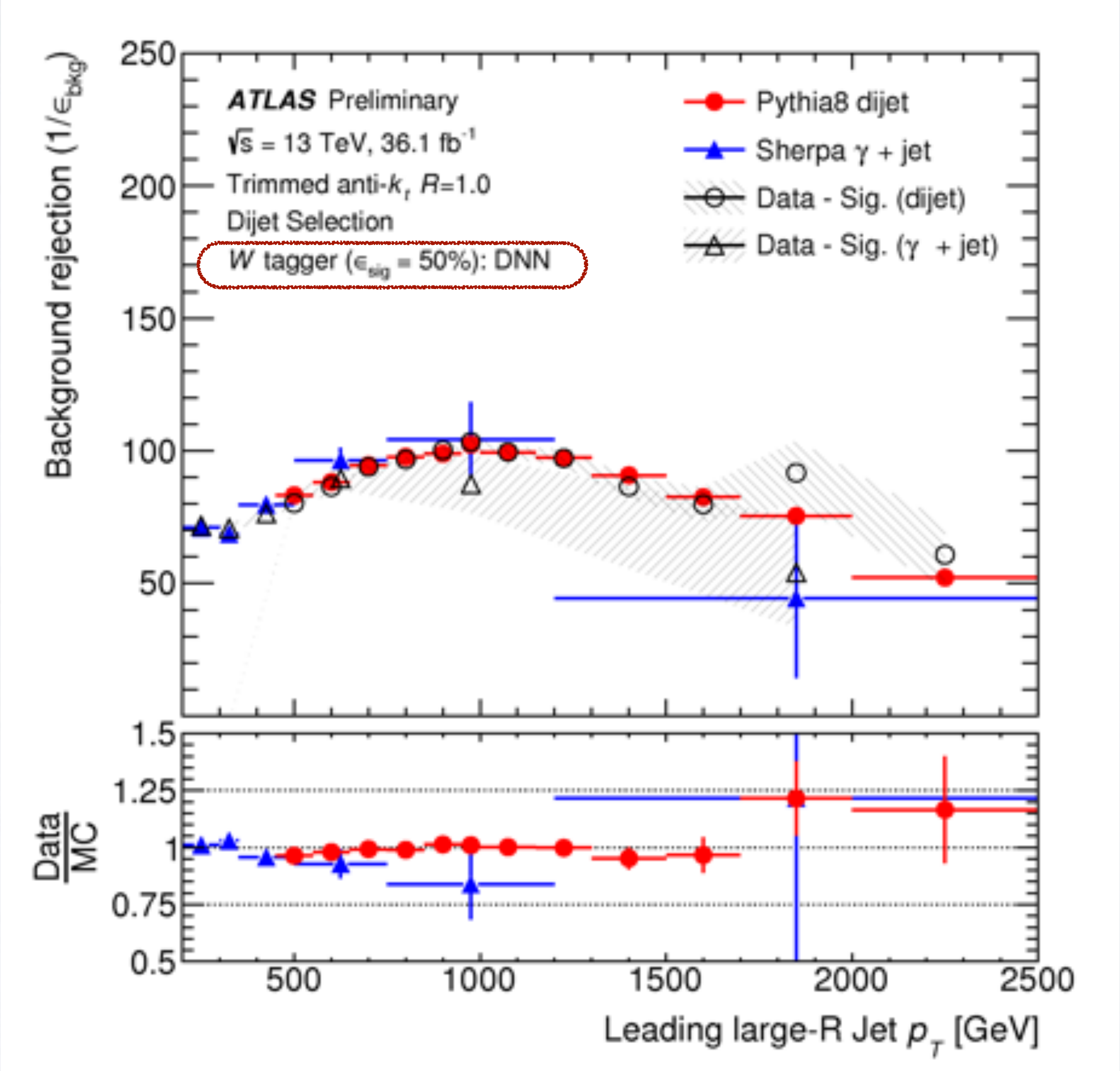


W TAGGING PERFORMANCE IN DATA - BACKGROUND

DNN Discriminant



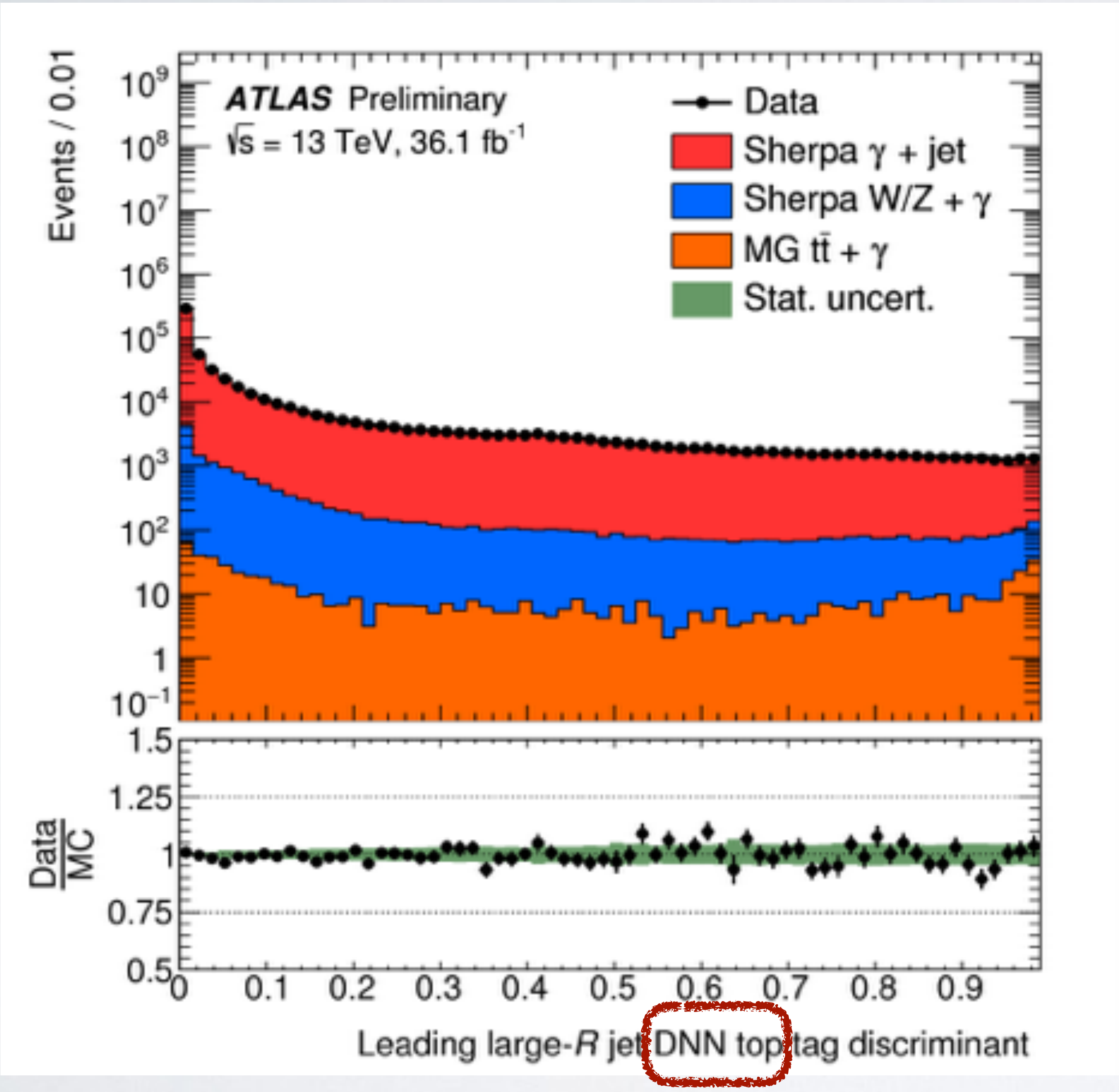
Background Rejection



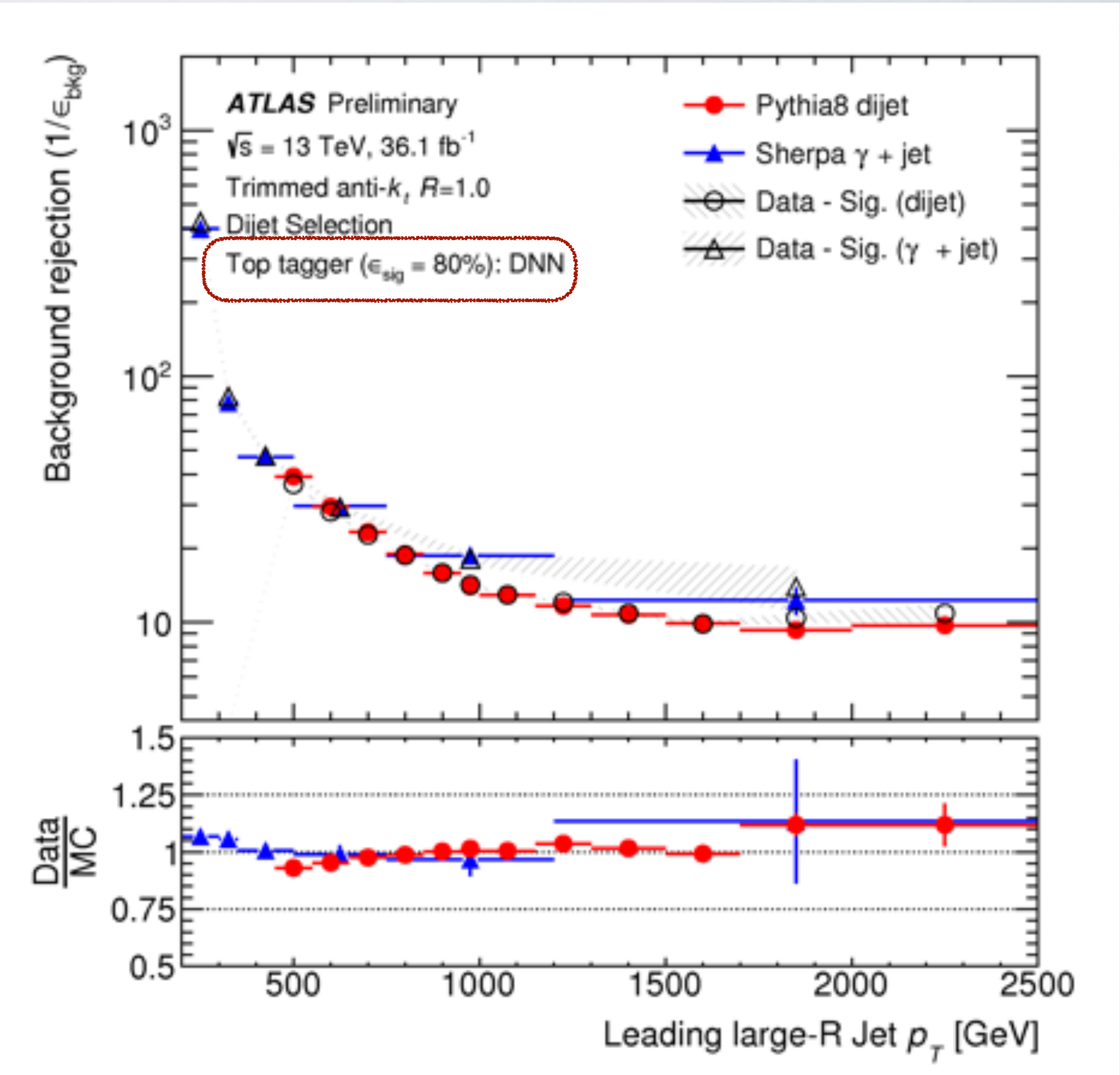
- Well modelled

TOP TAGGING PERFORMANCE IN DATA - BACKGROUND

DNN Discriminant



Background Rejection



- Well modelled

CONCLUSION

Combining high-level inputs in BDT and DNN improves background rejection

- Observed similar performance for BDT and DNN

Signal efficiency measurement in data & MC

- Modelling in agreement with data within uncertainties

Background rejection measurement in data & MC

- Modelling in agreement with data for baseline MC generators
- Similar background rejection in the common region

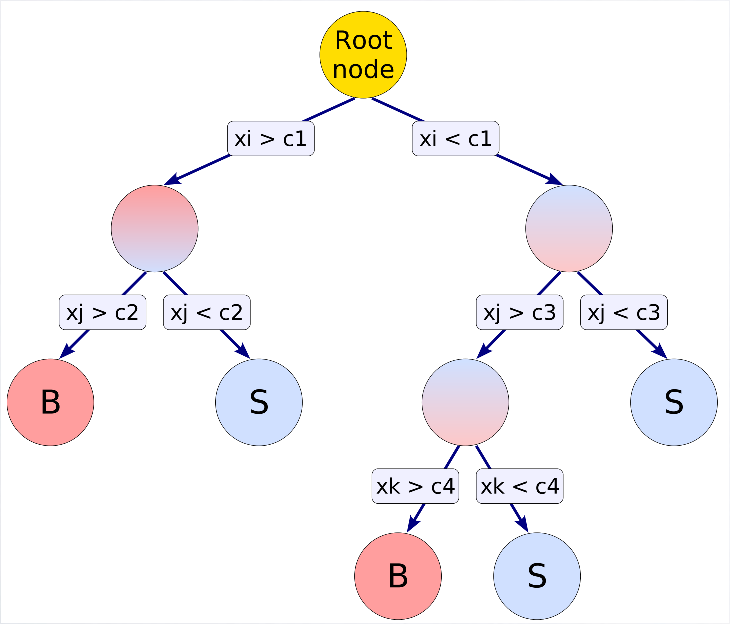
THANK YOU!

BACKUP

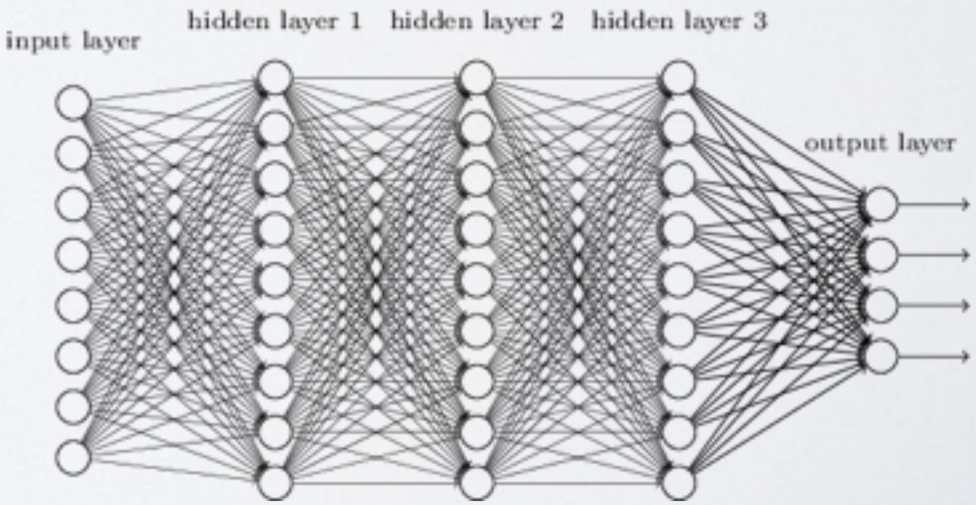
APPLICATION OF BDTS AND DNNS TO W AND TOP TAGGING USING HIGH-LEVEL FEATURES

- Numerous substructure variables are available and are used by ATLAS
- Feed the ML algorithms with jet substructure variables (high-level features)
- Study the performance of *W* and top tagging with two Machine Learning (ML) techniques in parallel

1. Boosted Decision Trees (BDT) using TMVA



2 . Deep Neural Networks (DNN) using Keras with Theano backend



SAMPLES

Training & Testing Samples

- Split signal and background (dijet) in training and testing samples

Training Event Weights: Signal and background samples are weighted to flat truth p_T distribution

Testing Event Weights: Signal samples are weighted to match background (dijet) truth p_T distribution

W Tagging

- p_T = [200, 2000] GeV, η = [-2,2]
- # Training signal jets = 7×10^5
- # Training light jets = 7×10^5

Top Tagging

- p_T = [350, 2000] GeV, η = [-2,2]
- # Training signal jets = 10^6
- # Training light jets = 10^6

INPUTS

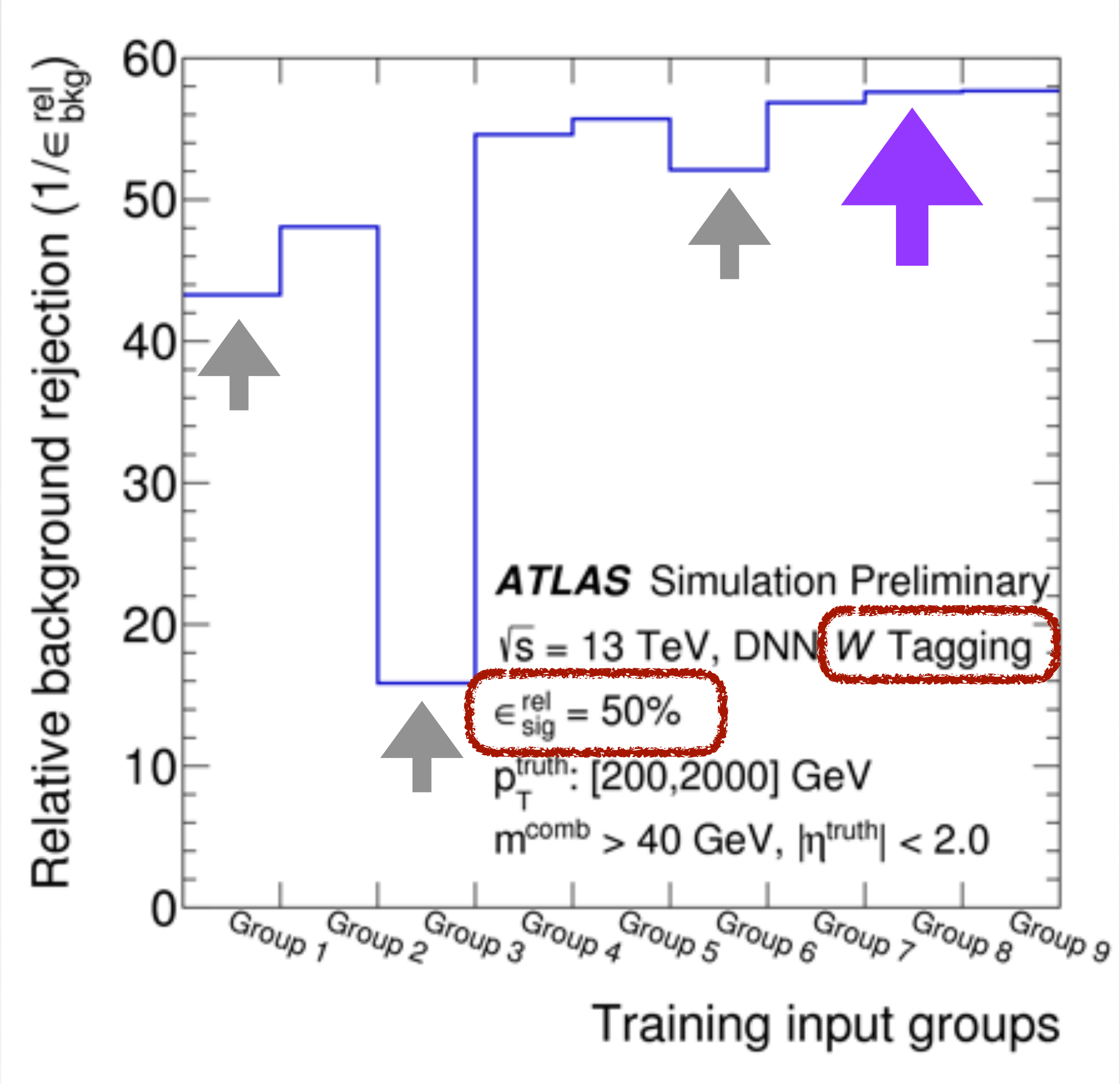
ATLAS-CONF-2017-064

Observable	Variable	Used For	Reference
Jet mass	m^{comb}	top,W	[35]
Energy Correlation Ratios	ECF_1, ECF_2, ECF_3 C_2,D_2	top,W	[41,42]
N-subjettiness	τ_1,τ_2,τ_3 τ_{21},τ_{32}	top,W	[43,44]
Center of Mass Observables	Fox Wolfram (R_2^{FW})	W	[45]
Splitting Measures	Z_{CUT} $\sqrt{d_{12}}$,$\sqrt{d_{23}}$	W top,W	[46] [47]
Planar Flow	$\mathcal{P}$	W	[48]
Angularity	a_3	W	[49]
Aplanarity	A	W	[50]
KtDR	$KtDR$	W	[51]
Qw	Q_w	top	[46]

DNN TRAINING - INPUTS OPTIMIZATION

W-Boson Tagging

- Study different groups of input variables
- Groups are defined by varying features (scale-dependence, …)
- Use a flat p_T spectrum (evaluation)
- Choose the set with the highest background rejection
- Observed the significance of the scale and jet mass

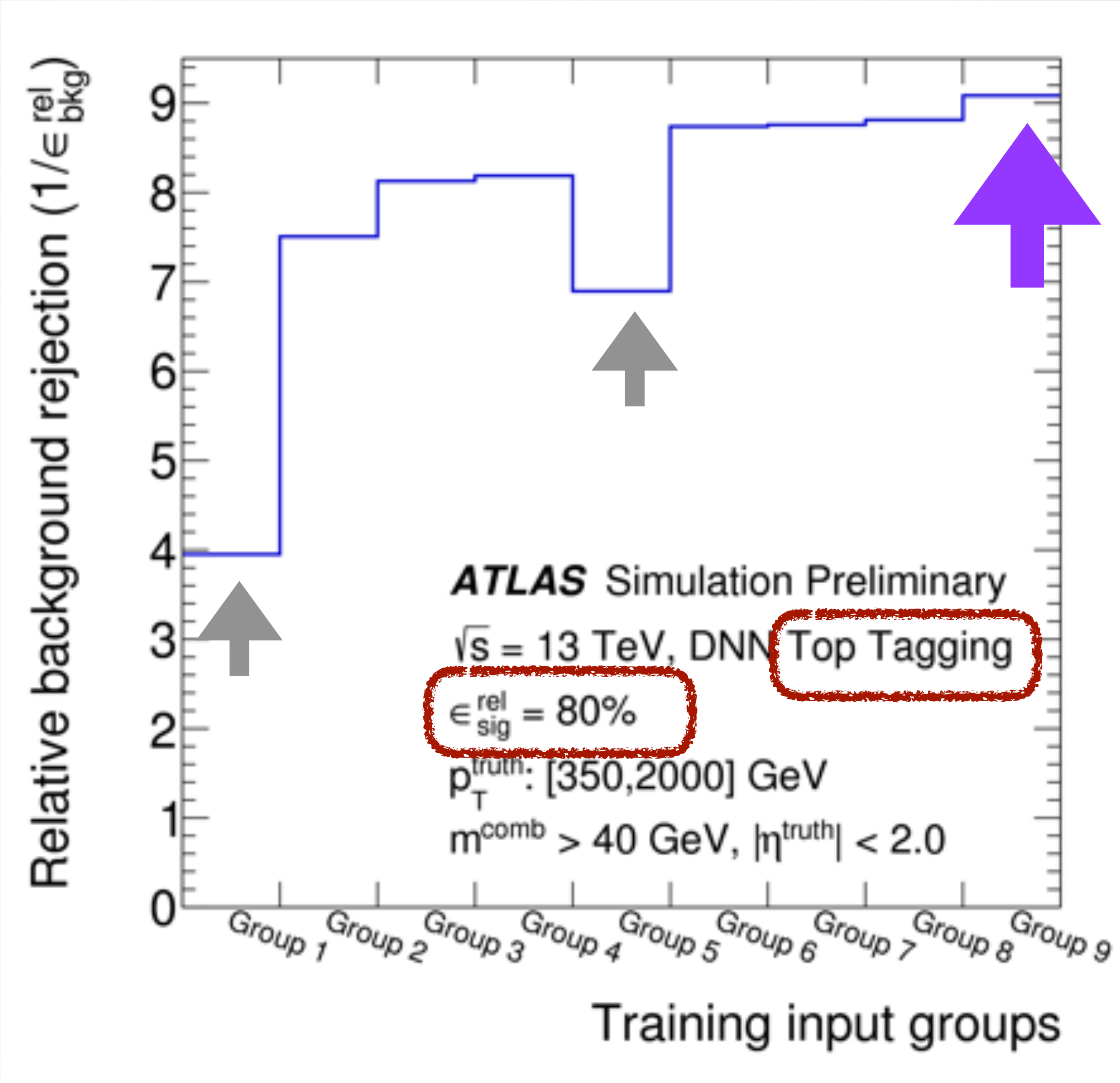


Group 1	τ_1, τ_2, e_3, m^{comb}, p_T
Group 2	τ_1, τ_2, e_3, m^{comb}, p_T, $\sqrt{d_{12}}$, KtDR
Group 3	τ_{21}, C_2, D_2, R_2^{FW}, $\mathcal{P}$, a_3, A, Z_{CUT}
Group 4	τ_{21}, C_2, D_2, R_2^{FW}, $\mathcal{P}$, a_3, A, Z_{CUT}, m^{comb}
Group 5	τ_{21}, C_2, D_2, R_2^{FW}, $\mathcal{P}$, a_3, A, Z_{CUT}, m^{comb}, p_T
Group 6	τ_1, τ_2, e_3, m^{comb}, p_T, R_2^{FW}, $\sqrt{d_{12}}$, KtDR, a_3, A
Group 7	τ_{21}, C_2, D_2, R_2^{FW}, $\mathcal{P}$, a_3, A, Z_{CUT}, m^{comb}, $\sqrt{d_{12}}$, KtDR
Group 8	τ_{21}, C_2, D_2, R_2^{FW}, $\mathcal{P}$, a_3, A, Z_{CUT}, m^{comb}, p_T, $\sqrt{d_{12}}$, KtDR
Group 9	τ_1, τ_2, τ_{21}, $\sqrt{d_{12}}$, C_2, D_2, e_3, m^{comb}, p_T, R_2^{FW}, $\mathcal{P}$, a_3, A, Z_{CUT}, KtDR

DNN TRAINING - INPUTS OPTIMIZATION

Top-Quark Tagging

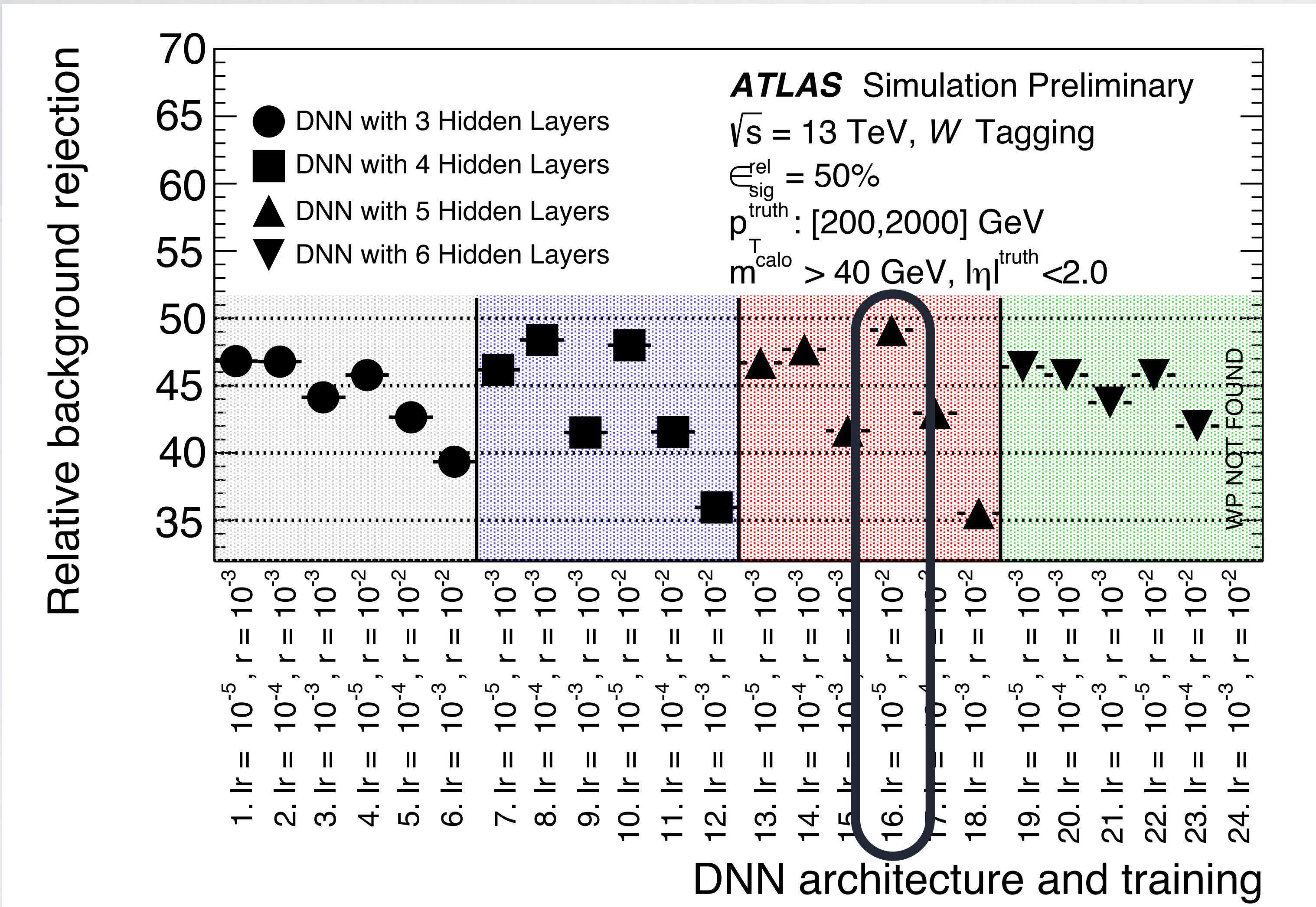
- Study different groups of input variables
- Groups are defined by varying features (scale-dependence, …)
- Use a flat p_T spectrum (evaluation)
- Choose the set with the highest background rejection
- Observed the significance of the scale and jet mass



Group 1	C_2, D_2, τ_{21}, τ_{32},
Group 2	C_2, D_2, τ_{21}, τ_{32}, m^{comb}
Group 3	C_2, D_2, τ_{21}, τ_{32}, m^{comb}, p_T
Group 4	τ_1, τ_2, τ_3, e_3, m^{comb}, p_T
Group 5	C_2, D_2, τ_{21}, τ_{32}, $\sqrt{d_{12}}$, $\sqrt{d_{23}}$, Q_W
Group 6	C_2, D_2, τ_{21}, τ_{32}, $\sqrt{d_{12}}$, $\sqrt{d_{23}}$, Q_W, m^{comb}
Group 7	τ_1, τ_2, τ_3, e_3, m^{comb}, p_T, $\sqrt{d_{12}}$, $\sqrt{d_{23}}$, Q_W
Group 8	C_2, D_2, τ_{21}, τ_{32}, $\sqrt{d_{12}}$, $\sqrt{d_{23}}$, Q_W, m^{comb}, p_T
Group 9	τ_1, τ_2, τ_3, τ_{21}, τ_{32}, $\sqrt{d_{12}}$, $\sqrt{d_{23}}$, Q_W, C_2, D_2, e_3, m^{comb}, p_T

DNN TRAINING - HYPER-PARAMETER OPTIMIZATION

Earlier studies: ATL-PHYS-PUB-2017-004



Grid search for DNN

- Varied: Learning rate (lr), regularizer (r), number of hidden layers

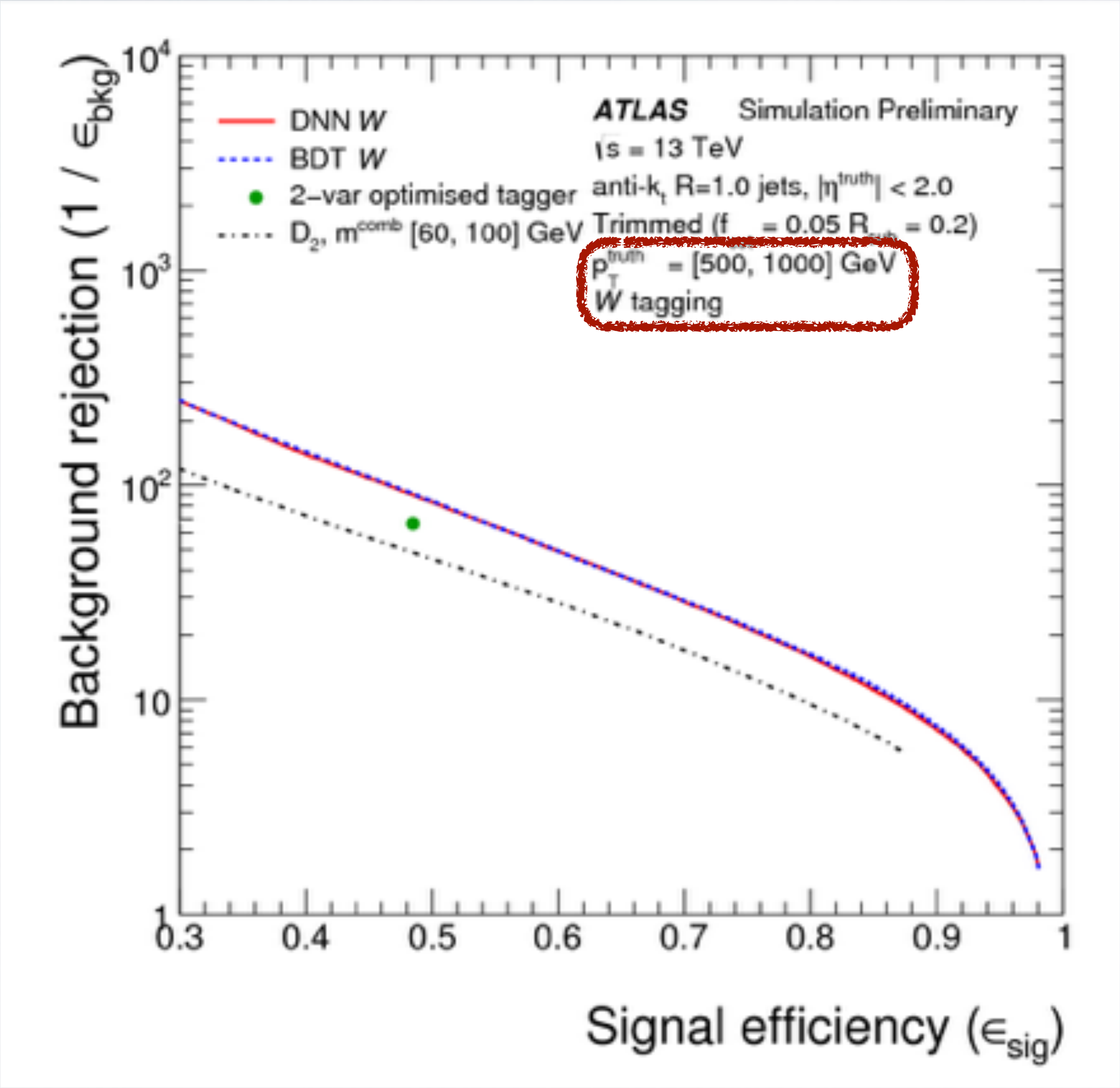
Fixed:

- Layer type = Dense with Batch Normalization
- Activation function = Rectified linear units
- Weight initialization = Glorot uniform

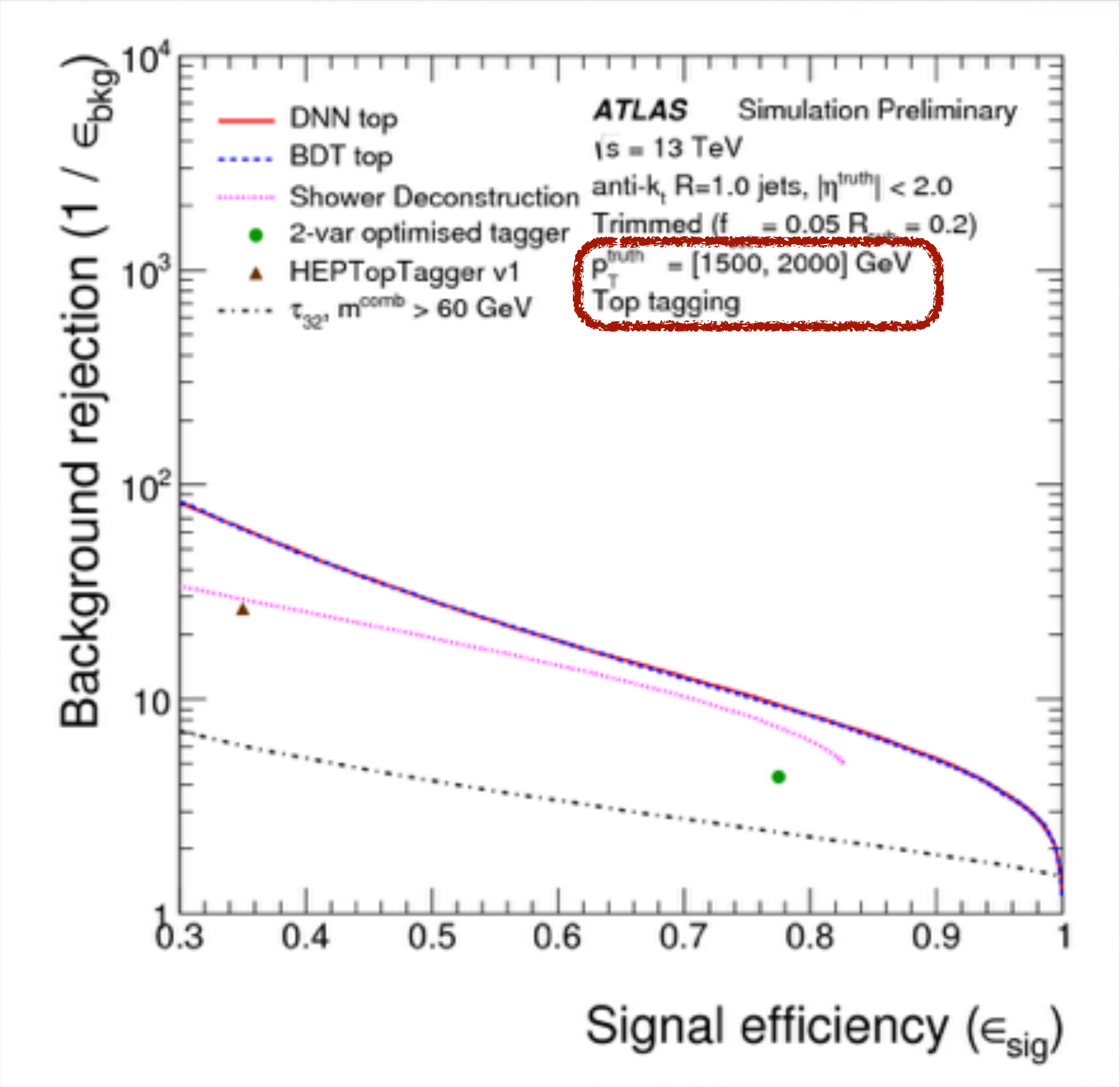
Similar grid search carried out for top-quark tagging and BDT

PERFORMANCE EVALUATION

W-Boson Tagging

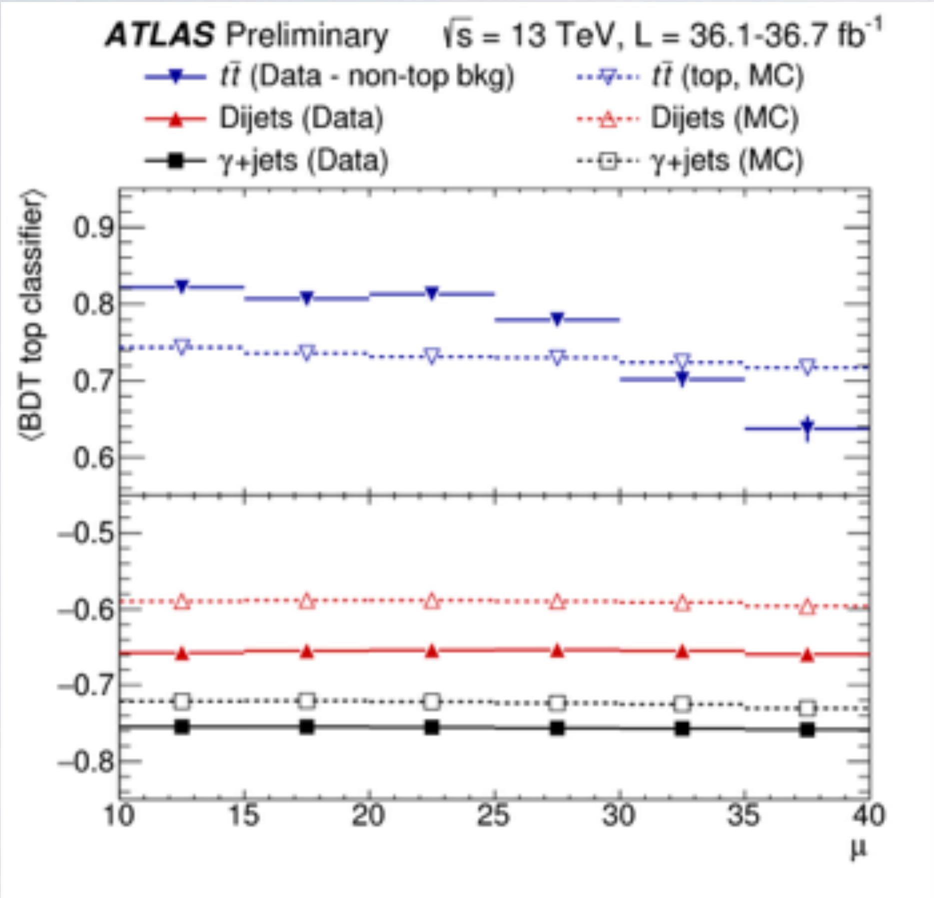


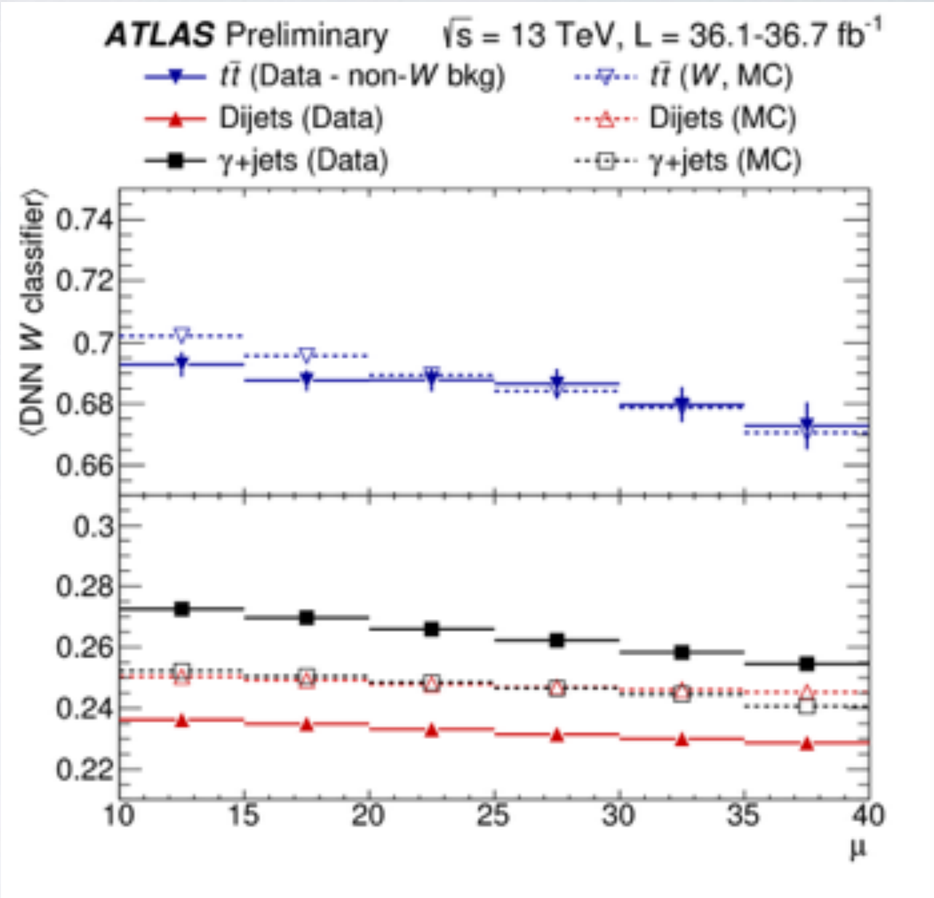
Top-Quark Tagging



- BDT & DNN: Improvements observed for both W and top tagging
- Magnitude of improvement differs for W and top tagging, but not the overall benefit of using a BDT or DNN

PILE-UP ROBUSTNESS





Robustness against pile-up

- Further investigation and evaluation of uncertainties are pending

PILE-UP ROBUSTNESS

Robustness against pile-up

- Further investigation and evaluation of uncertainties are pending