Search for heavy charged Higgs bosons in final states with two charged leptons and neutrinos at $\sqrt{s} = 13$ TeV with the ATLAS detector

Searching for new particles outside the standard model with the help of machine learning methods

Lisa Hofbauer, 13th July 2020

1. Background and Motivation

1.1 Reasons to search for Higgs bosons outside the standard model

There are phenomena in our universe which the current Standard Model of Particle Physics (SM) cannot explain

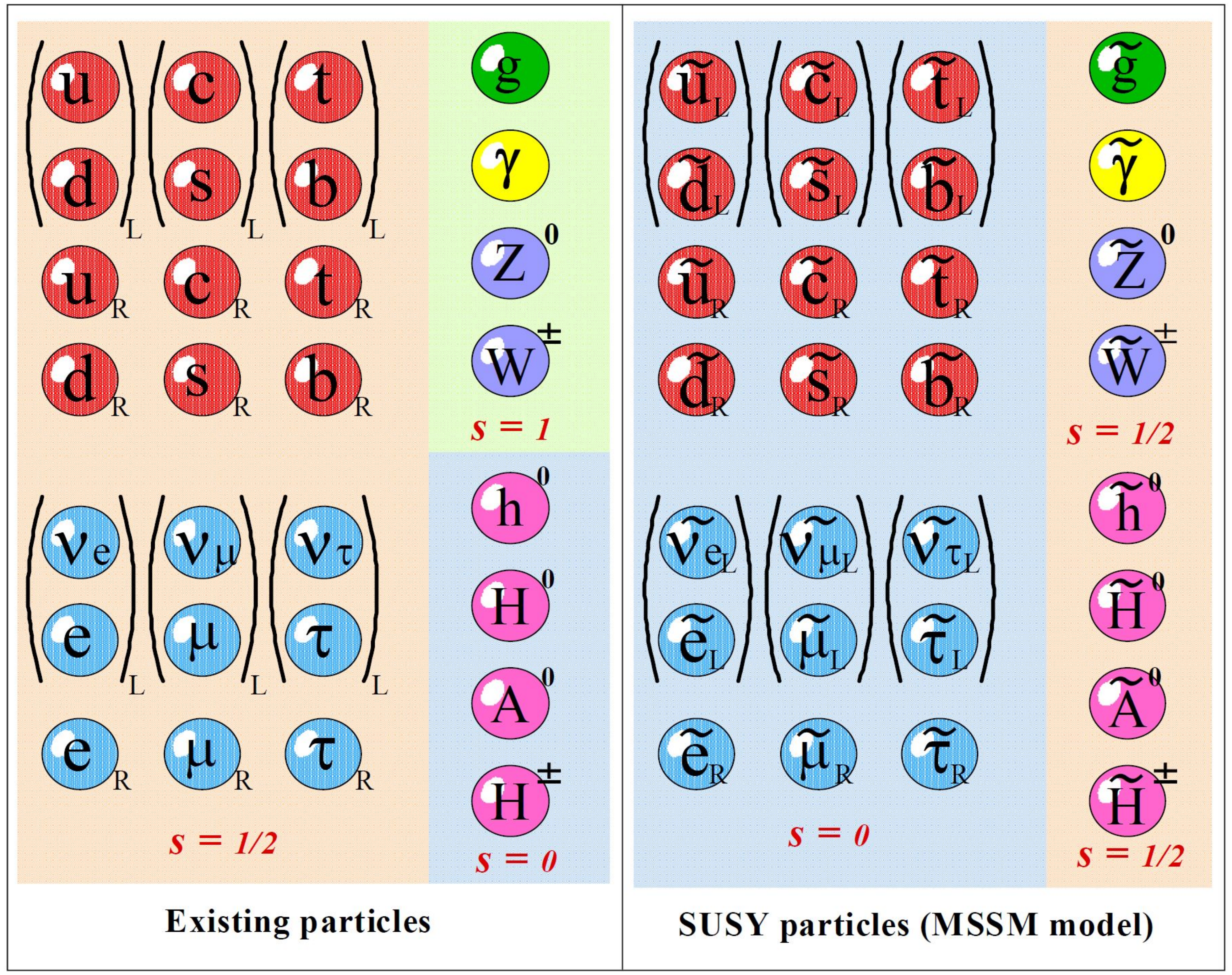
- Dark Matter & Dark Energy
- Neutrino mass
- Baryon asymmetry
- Hierarchy problem
- Gravity

Many theories have been suggested on how the Standard Model can be <u>extended</u> to explain some/all of these properties

Probably the most well known theory:

- Minimal Supersymmetric Standard Model (MSSM)

Contains a Two Higgs Doublet Model (2HDM) and therefore heavy charged Higgs bosons



2HDM needed for Higgsino anomalies and electroweak symmetry breaking

2HDM has equal fermion content and $SU(3)_C \times SU(2)_L \times U(1)_Y$ as current SM. But Higgs sector has two $SU(2)$ doublets.

Problems MSSM would solve:

- Hierarchy problem
- Has a candidate for a dark matter particle (LSP)
- Neutrino mass (lepton number violating MSSM)
- Baryon asymmetry (in very constrained parameter space)

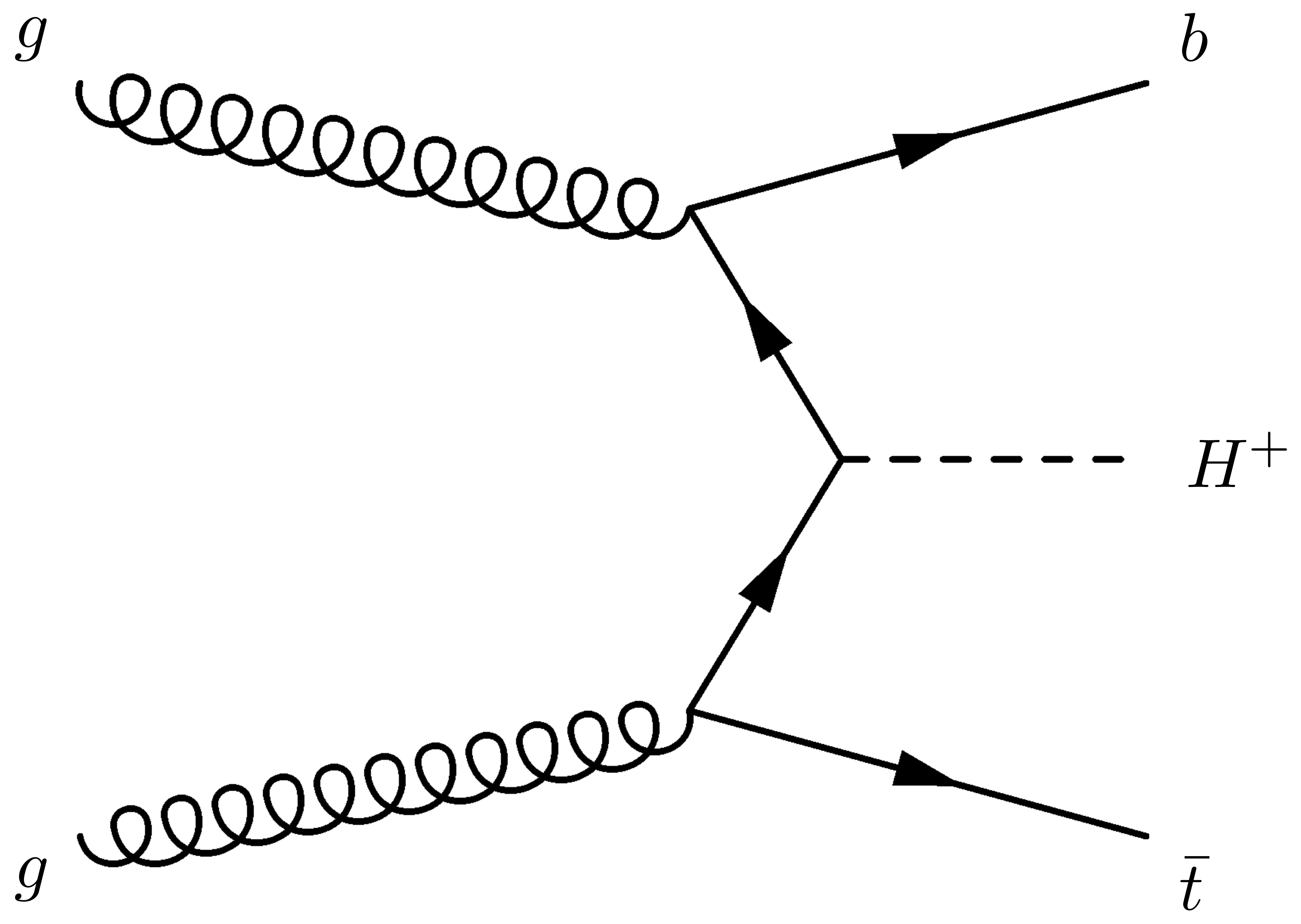
There are extensions to the MSSM which incorporate gravity and/or put weaker constraints on parameters

2. The specific analysis

2.1 Main production mode

Production mode depends on m_{H^+}

For heavy ($m_{H^+} > 200$ GeV) charged Higgs bosons the Feynman Diagram (right) shows the dominant production mode



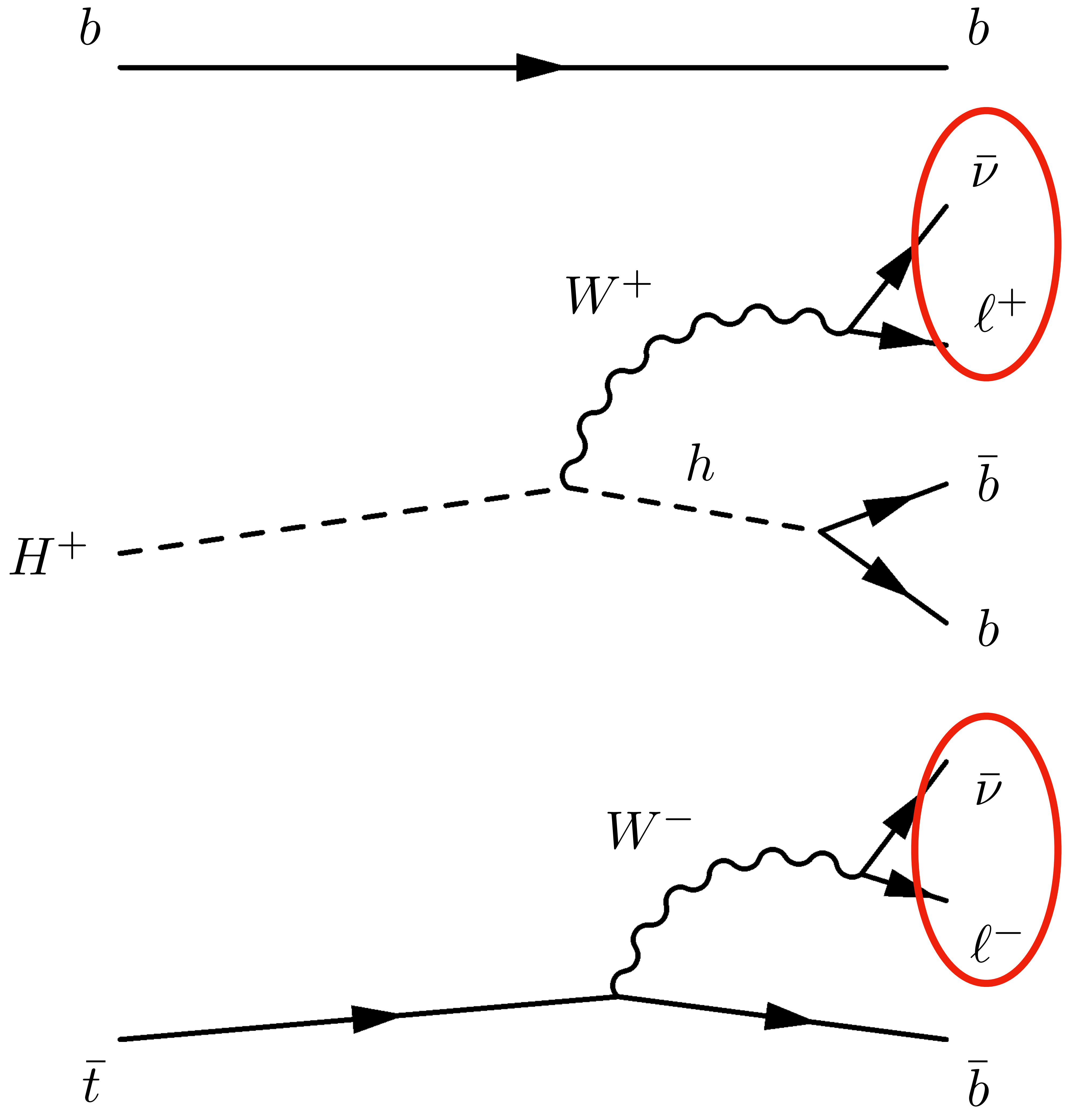
2.2 Involved particles and decays

- h: $BR(h \to b\bar{b}) \approx 56\,\%$,
  $BR(h \to c\bar{c}) \approx 3\,\%$

- W: $BR(W \to \ell\nu) \approx 30\,\%$,
  $BR(W \to qq) \approx 70\,\%$

- t: $BR(\bar{t} \to \bar{b}W^{-}) \approx 91\,\%$

ℓ = electron or muon, no tau



1.2 Properties and difficulties of this decay process

Main background:

- $t\bar{t}$ and $Z + jet$ background

Difficulties:

- two lepton channel: Can't identify which lepton comes from top and which comes from charged Higgs decay
- two neutrinos: Can't identify what part of the missing energy comes from top and which comes from charged Higgs decay
- no general reconstruction possible (only for limit of high energy charged higgs)
- only reconstruction for standard model higgs possible

1.3 Method and goal of this data analysis

Main Goal: Use machine learning algorithms to separate hypothetical charged Higgs boson signal form relevant background processes

Method: Train an algorithm with sample data to best distinct background from signal events.

Order of analysis:

- Event Selection & setting useful variables
- Train alogrithm with Monte Carlo simulated data
- Put real LHC Run 2 data into the tree and look at resulting signal-background-separation

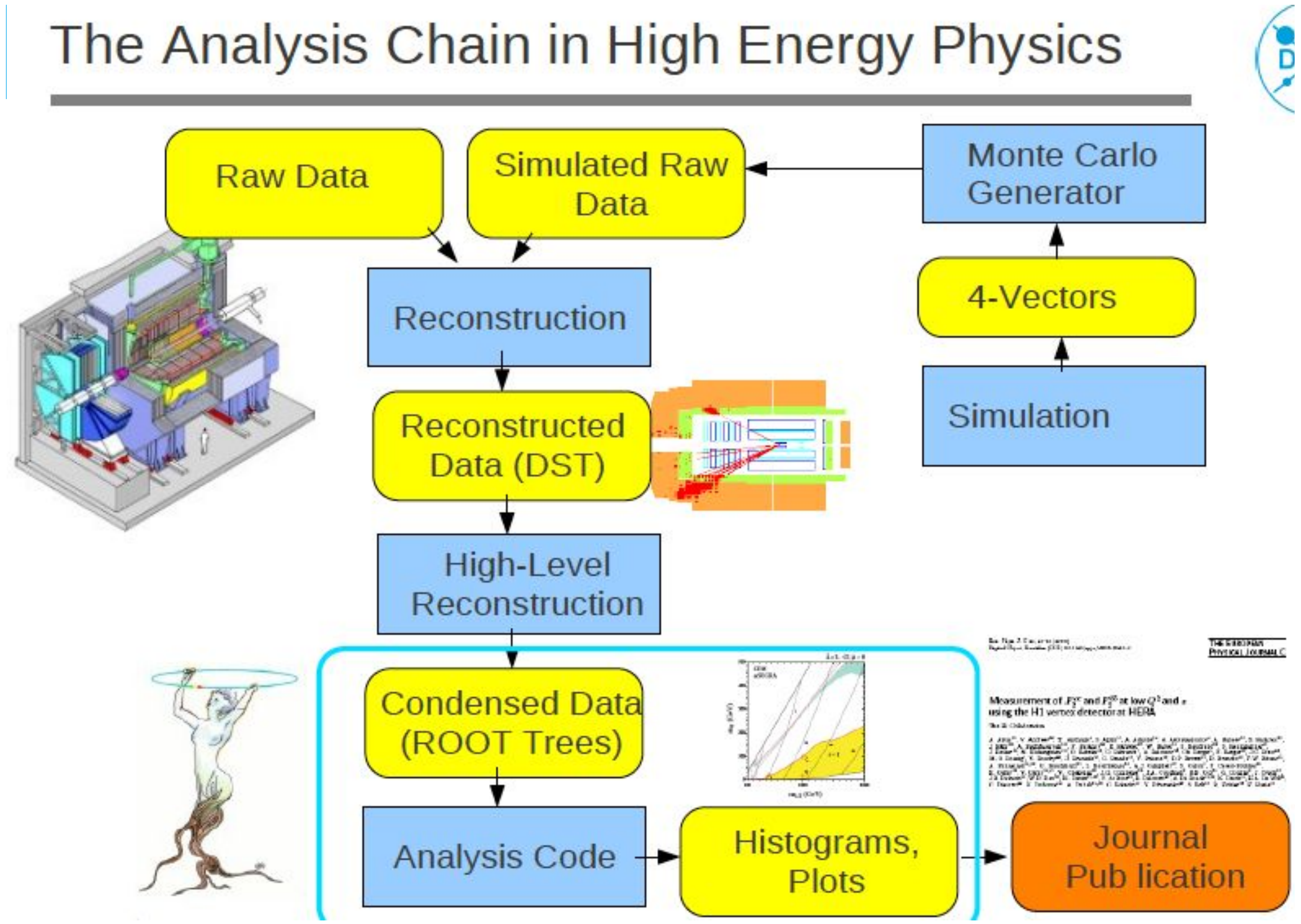
3. Data analysis using Machine Learning

3.1 The ROOT Framework

ROOT used in every aspect of data analysis at LHC (generally High Energy Physics)

<u>ROOT outside of physics:</u>

- Neuroscience
- Commercial Data Mining
- Industrial Control Systems
- Defence
- Finance



Most important ROOT properties:

- TFile: In- and output medium. Can store all basic data types (int, float, double), but also „ROOT data types“ like TH1, TGraph, TTree.
- TTree: Main way LHC data is stored. Every entry of the Tree is an event (either real data or Monte Carlo Simulation). The Tree has „branches“ in which a number of different data can be stored for each event.
- TLeaf: Data for a single variable
- TChain: TTree spread over different TFiles

3.2 TMVA 4

What is TMVA 4?

„Toolkit for Multivariate Data Analysis“

Includes Neural Networks, Likelihood Functions, **Boosted Decision Trees (BDT),** …

What is Multivariate Data Analysis ?

Way of finding patterns in multidimensional data by analysing several data variables at the same time

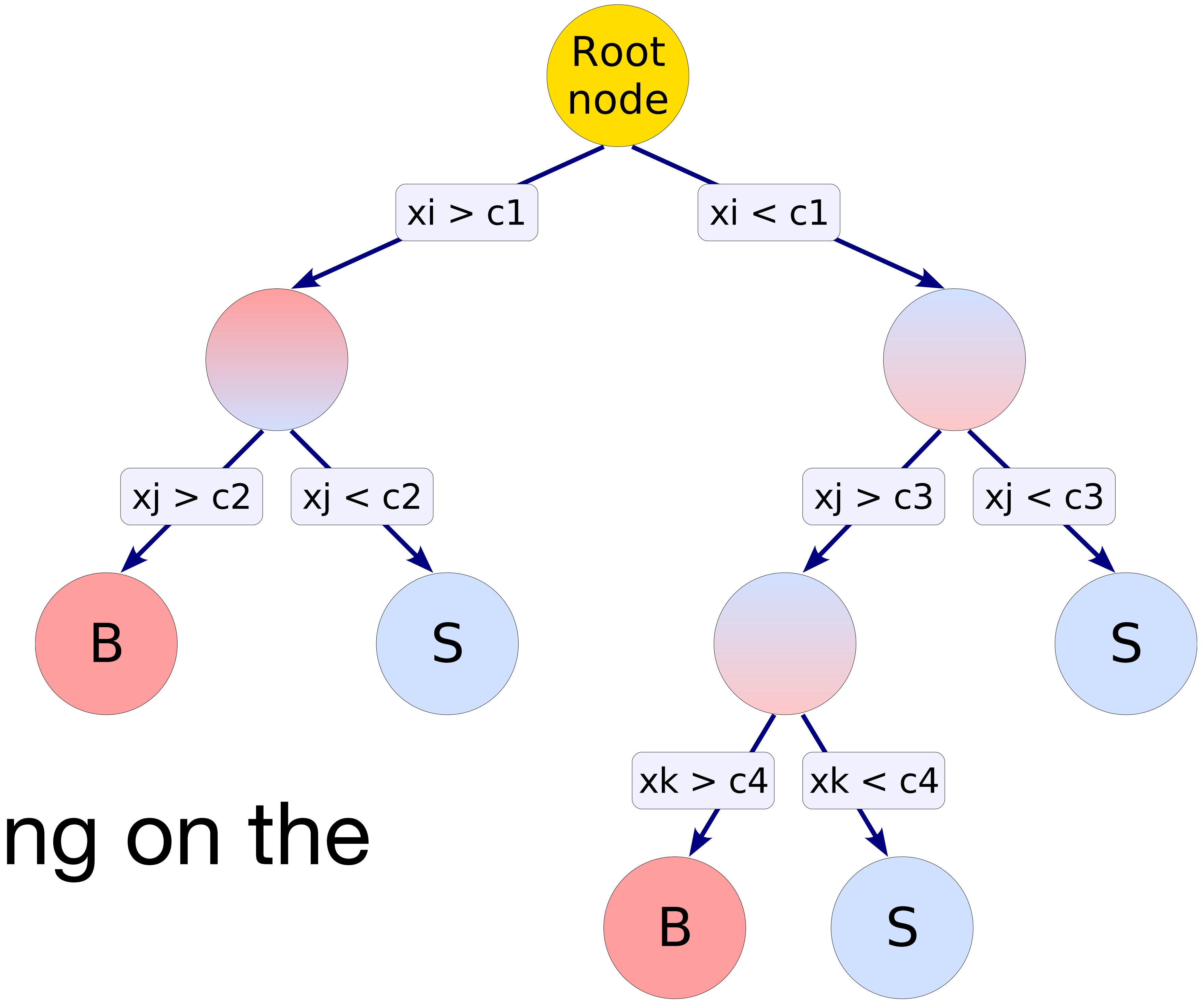
Why do we use Multivariate Data Analysis?

- Correlations between variables are visible
- More dimensions → more accurate/meaningful results

3.3 Boosted Decision Trees

<u>Properties:</u>

- Successive binary structure
- Splits data for one variable at a time until a stop criterion is fulfilled



<u>Boosting:</u>

Input weight gets in- or decreased depending on the correctness of previous classification

Classification false: weight is increased

Classification correct: weight is decreased

Goal: Find best splitting criterion for each variable

Training the algorithm:

- Input data is split randomly into two categories: training and test data
- Supervised learning: Since we put simulated data into the tree, we know for every event if it is signal or background. The tree outcomes can be supervised with this already known information
- See which variables are good for splitting
- Train until happy with classification efficiency

Overtraining

The algorithm splits training data with very high accuracy, but cannot generalize to new data leading to poor performance

Causes for overtraining:

- Too little data
- Data too homogenous

Solutions to prevent overtraining:

- Splitting available data in training and test data
- Performance on the training sample should not be better than on the test sample
- Pruning: Cutting away unnecessary variables in the tree

4. My analysis

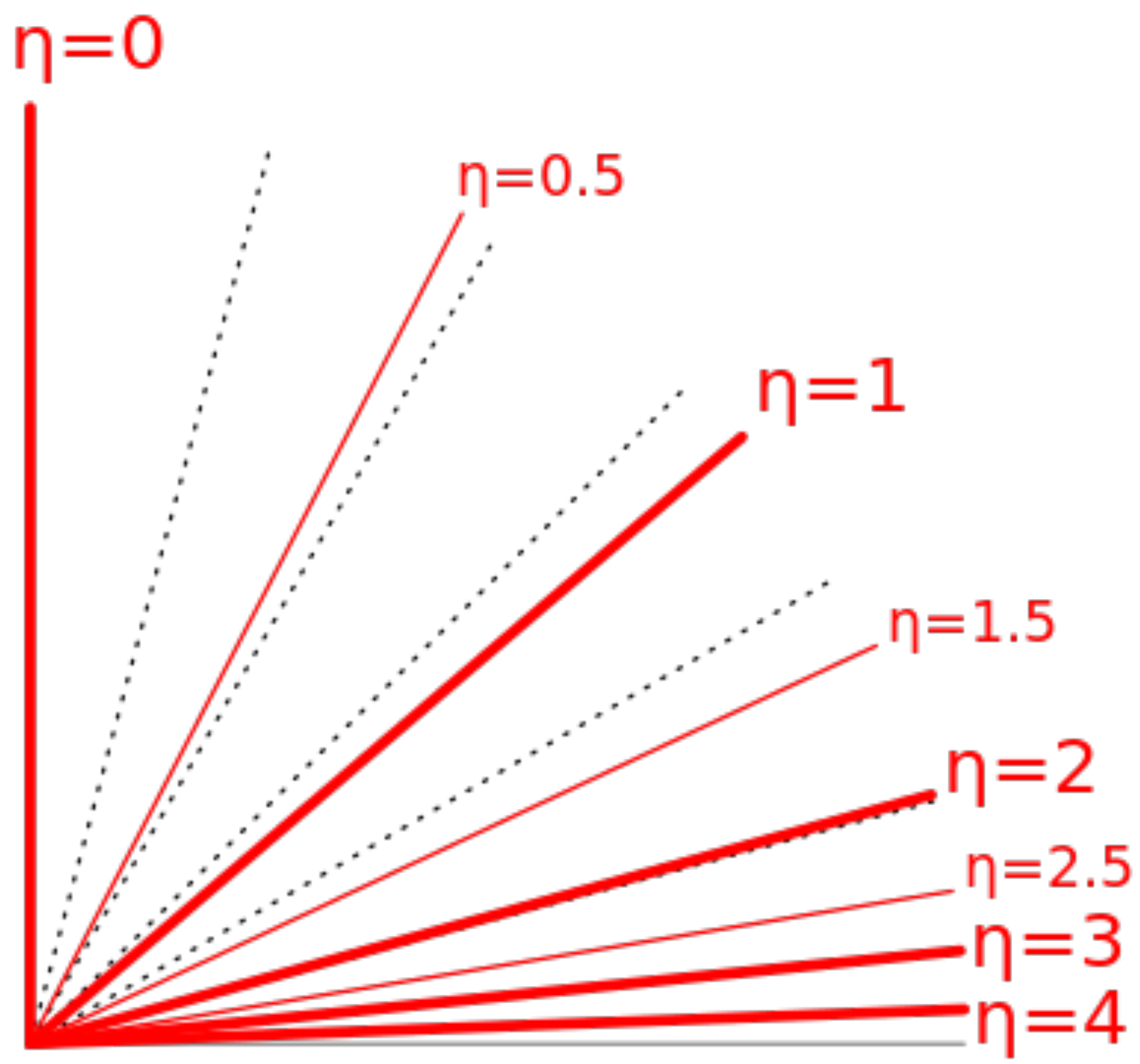
4.1 Event Selection and BDT variables

<u>Selectioncriteria:</u>

- Number of Jets higher than two
- Number of b-tagged Jets more than two
- Invariant mass of two leptons higher than 15 GeV
- One negatively and one positively charged lepton
- Leading lepton with $p_T > 27$ GeV
- Subleading lepton with $p_T > 10$ GeV

Selectioncriteria:

- Leptons outside the Z-Boson mass window 83 GeV $< m_{\ell\ell} < 99$ GeV
- Electrons need to have η outside of $1.37 < |\eta| < 1.52$
- If the subleading lepton is an electron: $p_T > 15$ GeV
- Jets with $|\eta| < 2.5$ are required to have $p_T > 20$ GeV
- Jets with $2.5 \leq |\eta| < 4.5$ have to have $p_T > 30$ GeV



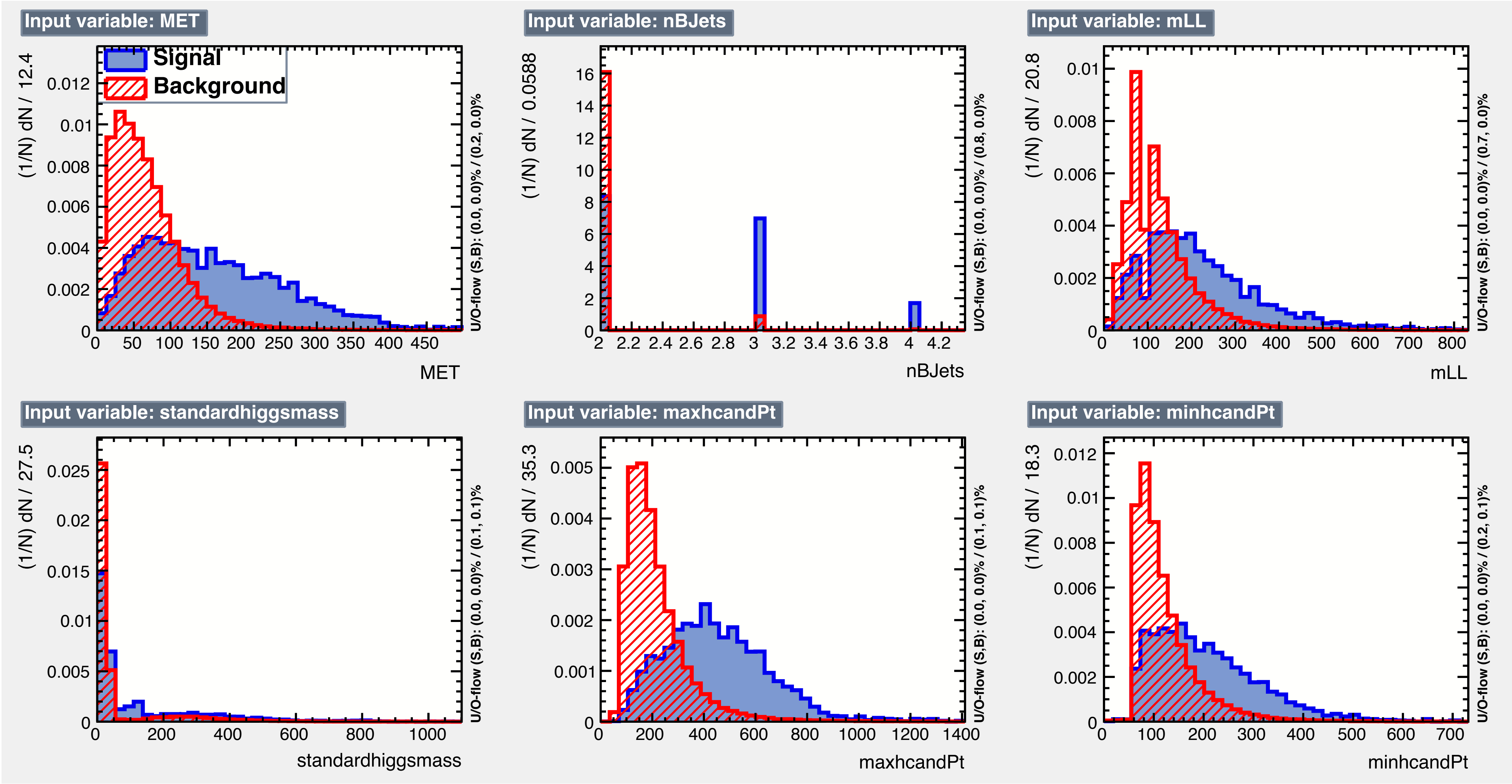
Variables:

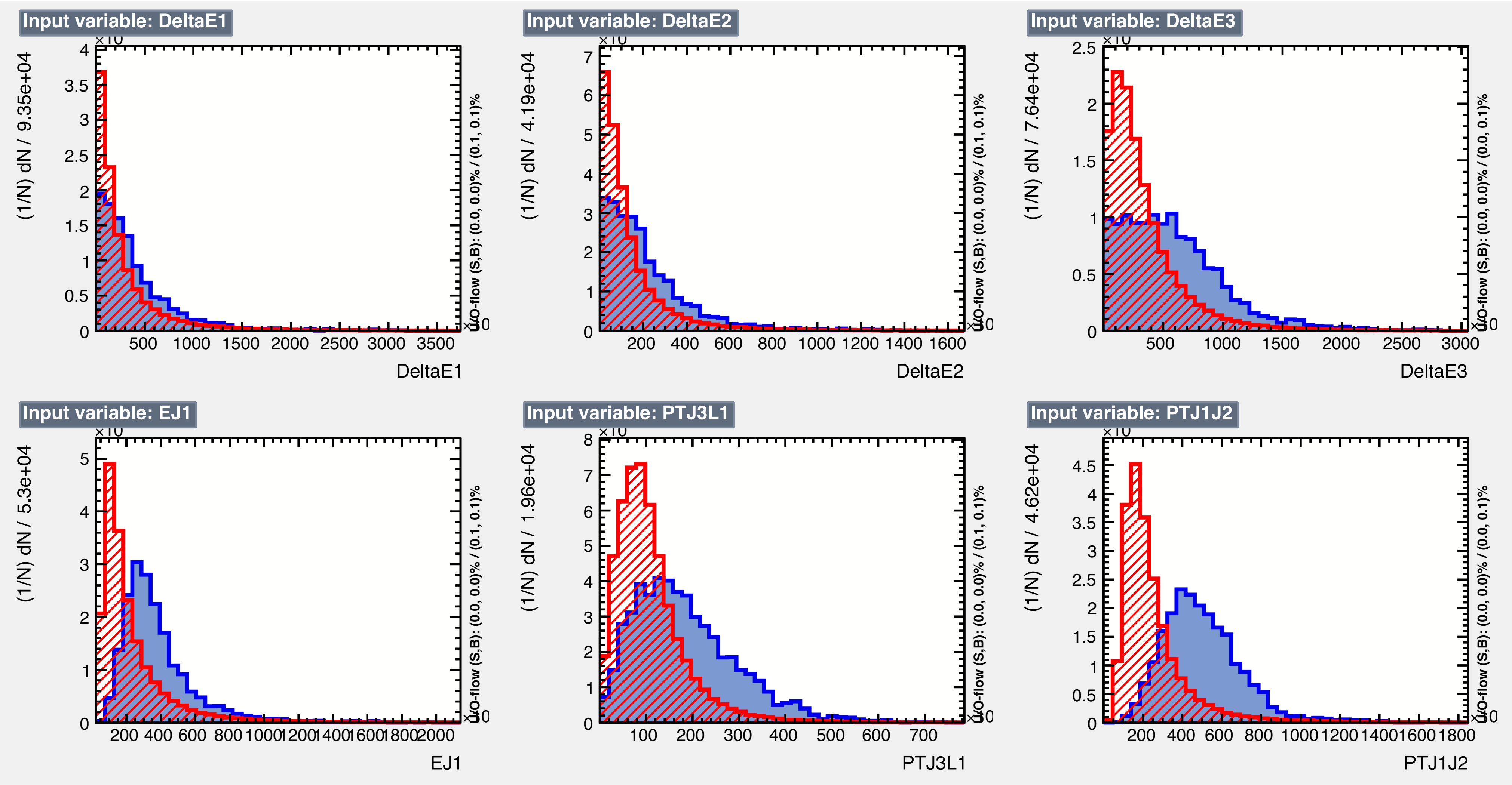
- $m(b_1 + b_2 + \ell_1 + \ell_2 + E_T^{miss})$
- $\Delta m(j_2 + \ell_1, j_1 + j_3 + \ell_1)$
- $p_T(j_1 + j_2)$
- $p_T(j_3 + \ell_1)$
- $\Delta E(\ell_1 + E_T^{miss}, j_1 + j_2)$
- $E(j_1)$
- $\Delta E(\ell_1, E_T^{miss} + j_2)$
- $\Delta E(j_2 + \ell_1 + \ell_2 + E_T^{miss}, j_3)$
- $p_T^{max}(h_{cand})$
- E_T^{miss}
- $m(\ell_1 + \ell_2)$
- $\Delta p_T(j_1, j_3)$
- $\Delta p_T(\ell_2, \ell_1)$
- $\Delta p_T(b_2, b_1 + \ell_2)$
- $m(h_{cand})$
- m_{max}(j-pair)
- m_{min}(j-pair)
- $p_T^{min}(h_{cand})$
- $\Delta p_T(j_1, j_3 + \ell_1)$
- m_{max}(b-pair)
- m_{min}(b-pair)
- $\Delta p_T(j_2, j_3 + \ell_1)$
- nBJets

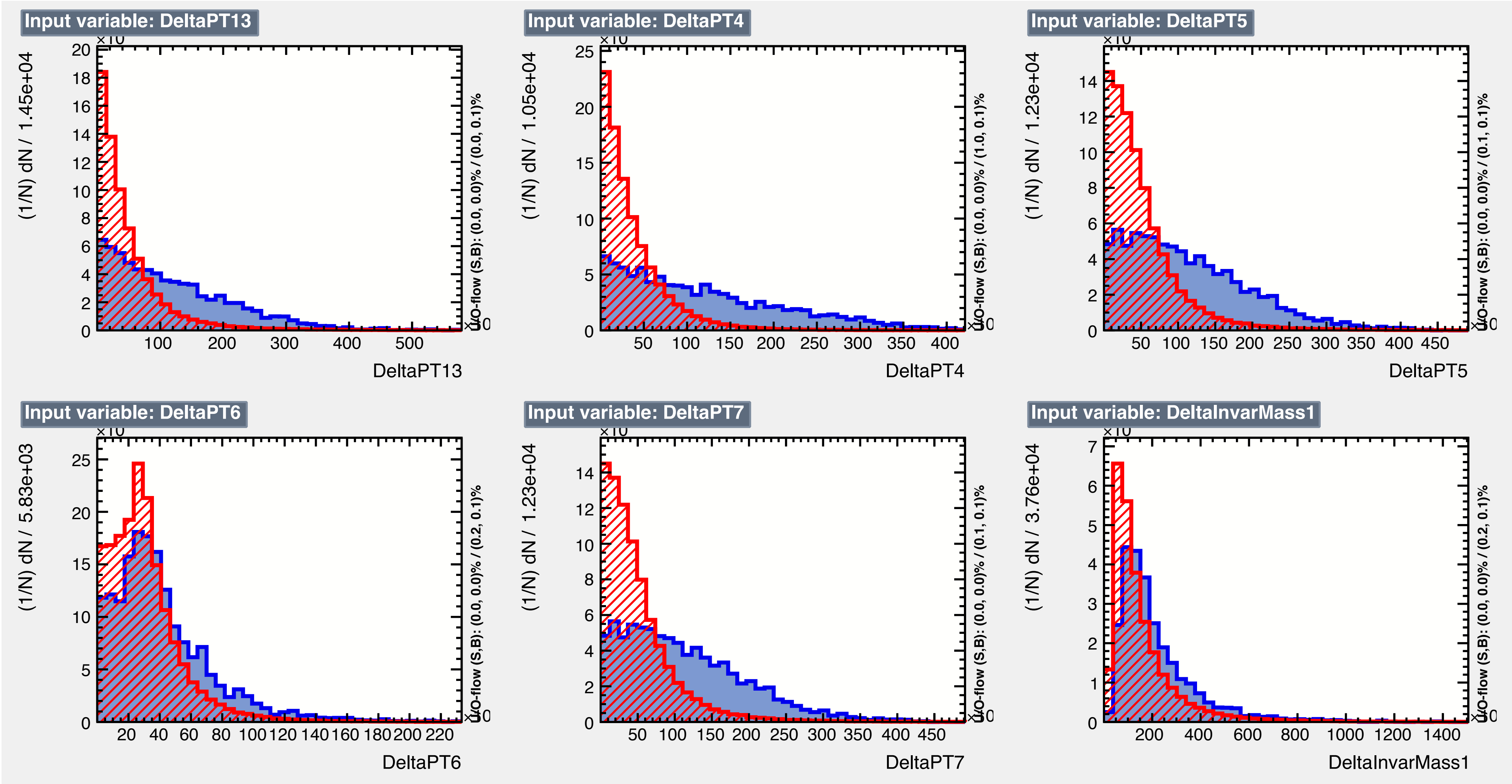
Best Separation: 6.785e-01

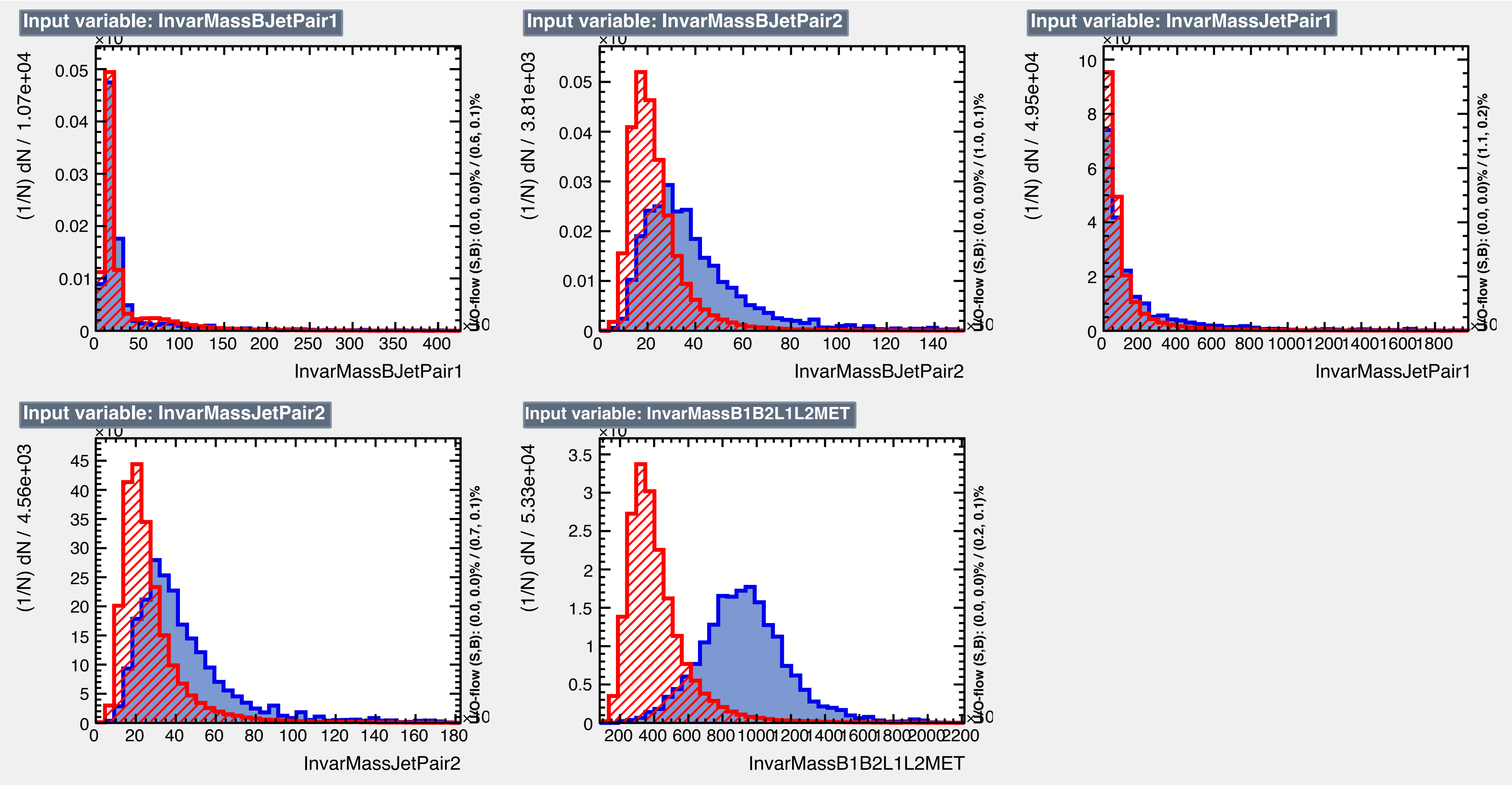
Worst Separation: 1.942e-02

4.2 BDT Training







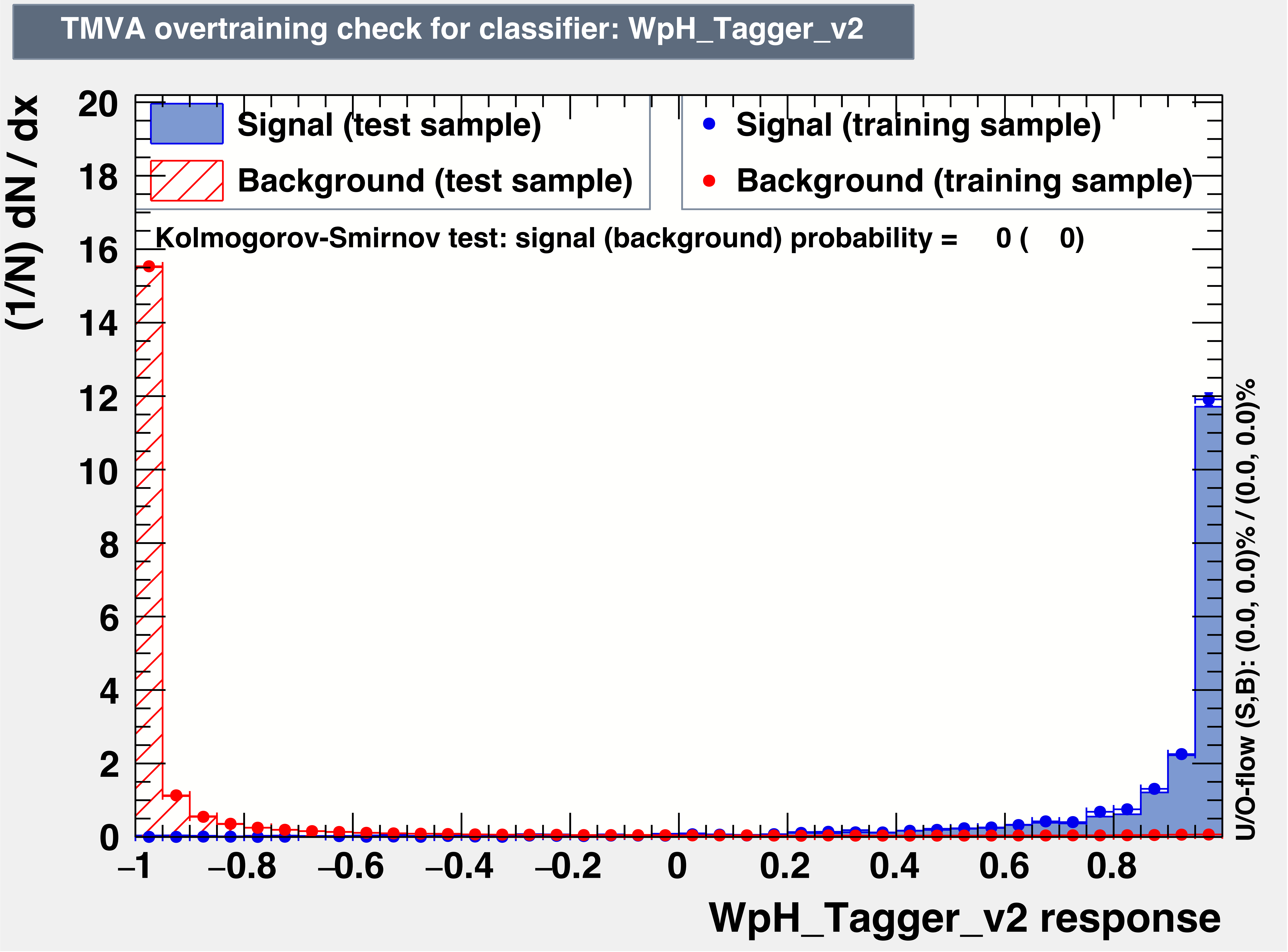


Signal correlation

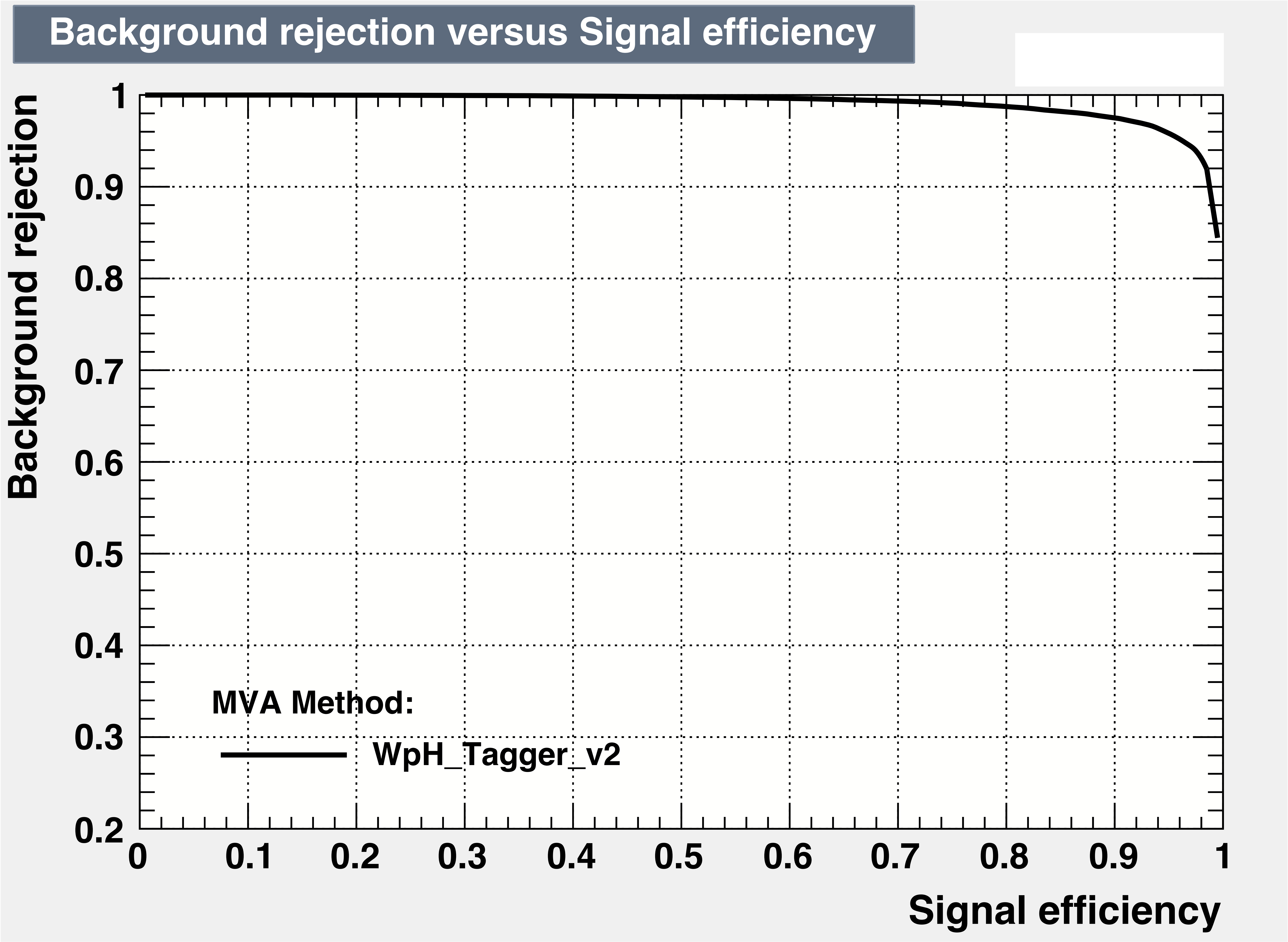
	MET	nBJets	mLL	hmass	maxhPt	minhPt	DPT13	DPT4	DPT5	DPT6	DPT7	DM1	DE1	DE2	DE3	EJ1	PTJ3L1	PTJ1J2	MBJP1	MBJP2	MJP1	MJP2	M1
MET:	+1.000	+0.021	-0.208	0.068	+0.131	+0.072	+0.103	-0.251	+0.248	-0.031	+0.248	+0.016	+0.029	+0.012	-0.292	+0.094	-0.140	+0.167	+0.005	+0.091	+0.020	+0.100	+0.060
nBJets:	+0.021	+1.000	-0.016	+0.068	+0.102	+0.276	-0.073	+0.005	-0.019	+0.014	-0.019	+0.080	-0.005	-0.000	-0.038	-0.000	+0.047	+0.017	+0.073	+0.021	+0.000	-0.058	+0.077
mLL:	-0.208	-0.016	+1.000	+0.000	-0.043	-0.021	-0.055	+0.263	+0.020	+0.305	+0.020	+0.097	-0.052	+0.176	+0.315	-0.035	+0.312	-0.039	+0.015	-0.019	-0.006	-0.017	+0.378
hmass:	-0.068	+0.068	+0.000	+1.000	+0.127	+0.221	-0.027	-0.026	-0.075	+0.020	-0.075	+0.171	+0.059	+0.003	-0.015	+0.061	+0.129	+0.076	+0.002	+0.084	+0.016	+0.055	+0.053
maxhPt:	+0.131	+0.102	-0.043	+0.127	+1.000	+0.326	+0.581	-0.003	+0.319	+0.035	+0.319	+0.166	+0.221	-0.083	+0.034	+0.329	+0.230	+0.758	+0.017	+0.570	+0.025	+0.398	+0.517
minhPt:	+0.072	+0.276	-0.021	+0.221	+0.326	+1.000	-0.182	+0.028	+0.001	+0.023	+0.001	+0.214	+0.049	+0.001	-0.064	+0.088	+0.132	+0.200	+0.022	+0.187	+0.015	+0.100	+0.190
DPT13:	+0.103	-0.073	-0.055	-0.027	+0.581	-0.182	+1.000	-0.028	+0.448	+0.031	+0.448	-0.079	+0.285	-0.095	+0.150	+0.388	+0.114	+0.800	+0.033	+0.391	+0.011	+0.470	+0.266
DPT4:	-0.251	+0.005	+0.263	-0.026	-0.003	+0.028	-0.028	+1.000	-0.017	-0.035	-0.017	+0.025	-0.105	+0.176	+0.531	-0.051	+0.556	-0.002	-0.034	-0.003	+0.006	-0.004	+0.310
DPT5:	+0.248	-0.019	+0.020	-0.075	+0.319	+0.001	+0.448	-0.017	+1.000	+0.131	+1.000	+0.034	+0.194	+0.084	-0.012	+0.207	-0.047	+0.444	+0.026	+0.213	+0.032	+0.246	+0.210
DPT6:	-0.031	+0.014	+0.305	+0.020	+0.035	+0.023	+0.031	-0.035	+0.131	+1.000	+0.131	+0.078	+0.037	+0.105	+0.145	+0.043	+0.184	+0.062	+0.019	+0.047	+0.003	+0.065	+0.267
DPT7:	+0.248	-0.019	+0.020	-0.075	+0.319	+0.001	+0.448	-0.017	+1.000	+0.131	+1.000	+0.034	+0.194	+0.084	-0.012	+0.207	-0.047	+0.444	+0.026	+0.213	+0.032	+0.246	+0.210
DM1:	+0.016	+0.080	+0.097	+0.171	+0.166	+0.214	-0.079	+0.025	+0.034	+0.078	+0.034	+1.000	+0.264	+0.175	+0.029	+0.274	+0.284	+0.251	+0.025	+0.122	+0.020	+0.142	+0.157
DE1:	+0.029	-0.005	-0.052	+0.059	+0.221	+0.049	+0.285	-0.105	+0.194	+0.037	+0.194	+0.264	+1.000	+0.522	+0.330	+0.922	+0.109	+0.388	+0.022	+0.188	-0.008	+0.248	+0.159
DE2:	+0.012	-0.000	+0.176	+0.003	-0.083	+0.001	-0.095	+0.176	+0.084	+0.105	+0.084	+0.175	+0.522	+1.000	+0.529	+0.372	+0.218	-0.040	+0.020	-0.006	-0.002	+0.016	+0.169
DE3:	-0.292	-0.038	+0.315	-0.015	+0.034	-0.064	+0.150	+0.531	-0.012	+0.145	-0.012	+0.029	+0.330	+0.529	+1.000	+0.360	+0.387	+0.106	-0.002	+0.045	-0.012	+0.073	+0.360
EJ1:	+0.094	-0.000	-0.035	+0.061	+0.329	+0.088	+0.388	-0.051	+0.207	+0.043	+0.207	+0.274	+0.922	+0.372	+0.360	+1.000	+0.134	+0.528	+0.024	+0.257	-0.001	+0.323	+0.272
PTJ3L1:	-0.140	+0.047	+0.312	+0.129	+0.230	+0.132	+0.114	+0.556	-0.047	+0.184	-0.047	+0.284	+0.109	+0.218	+0.387	+0.134	+1.000	+0.288	+0.018	+0.164	+0.010	+0.188	+0.397
PTJ1J2:	+0.167	+0.017	-0.039	+0.076	+0.758	+0.200	+0.800	-0.002	+0.444	+0.062	+0.444	+0.251	+0.388	-0.040	+0.106	+0.528	+0.288	+1.000	+0.024	+0.510	+0.016	+0.579	+0.401
MBJP1:	+0.005	+0.073	+0.015	+0.002	+0.017	+0.022	+0.033	-0.034	+0.026	+0.019	+0.026	+0.025	+0.022	+0.020	-0.002	+0.024	+0.018	+0.024	+1.000	+0.021	+0.034	+0.029	+0.020
MBJP2:	+0.091	+0.021	-0.019	+0.084	+0.570	+0.187	+0.391	-0.003	+0.213	+0.047	+0.213	+0.122	+0.188	-0.006	+0.045	+0.257	+0.164	+0.510	+0.021	+1.000	+0.072	+0.804	+0.365
MJP1:	+0.020	+0.000	-0.006	+0.016	+0.025	+0.015	+0.011	+0.006	+0.032	+0.003	+0.032	+0.020	-0.008	-0.002	-0.012	-0.001	+0.010	+0.016	+0.034	+0.072	+1.000	+0.052	+0.020
MJP2:	+0.100	-0.058	-0.017	+0.055	+0.398	+0.100	+0.470	-0.004	+0.246	+0.065	+0.246	+0.142	+0.248	+0.016	+0.073	+0.323	+0.188	+0.579	+0.029	+0.804	+0.052	+1.000	+0.216
M1:	+0.060	+0.077	+0.378	+0.053	+0.517	+0.190	+0.266	+0.310	+0.210	+0.267	+0.210	+0.157	+0.159	+0.169	+0.360	+0.272	+0.397	+0.401	+0.020	+0.365	+0.020	+0.216	+1.000

Background correlation

	MET	nBJets	mLL	hmass	maxHPt	minhPt	DPT13	DPT4	DPT5	DPT6	DPT7	DM1	DE1	DE2	DE3	EJ1	PTJ3L1	PTJ1J2	MBJP1	MBJP2	MJP1	MJP2	M
MET:	+1.000	+0.027	+0.115	+0.111	+0.133	+0.079	+0.248	+0.068	+0.308	+0.067	+0.308	+0.110	+0.114	+0.122	+0.010	+0.177	+0.095	+0.240	-0.018	+0.080	+0.037	+0.189	+0.222
nBJets:	+0.027	+1.000	+0.004	+0.054	+0.072	+0.080	+0.033	+0.024	+0.038	+0.011	+0.038	+0.049	+0.036	+0.026	+0.019	+0.043	+0.041	+0.060	+0.065	+0.063	+0.016	+0.051	+0.087
mLL:	+0.115	+0.004	+1.000	+0.119	+0.077	+0.060	+0.064	+0.226	+0.151	+0.372	+0.151	+0.048	+0.010	+0.108	+0.242	+0.047	+0.240	+0.065	+0.014	+0.040	+0.019	+0.042	+0.400
hmass:	+0.111	+0.054	+0.119	+1.000	+0.493	+0.504	+0.161	+0.087	+0.150	+0.112	+0.150	+0.258	+0.172	+0.114	+0.099	+0.197	+0.252	+0.297	+0.035	+0.265	+0.027	+0.216	+0.259
maxhPt:	+0.133	+0.072	+0.077	+0.493	+1.000	+0.598	+0.429	+0.231	+0.328	+0.201	+0.328	+0.461	+0.353	+0.252	+0.248	+0.408	+0.526	+0.657	+0.049	+0.552	+0.033	+0.472	+0.539
minhPt:	+0.079	+0.080	+0.060	+0.504	+0.598	+1.000	+0.107	+0.095	+0.195	+0.119	+0.195	+0.350	+0.205	+0.135	+0.088	+0.234	+0.359	+0.390	+0.080	+0.331	+0.036	+0.279	+0.279
DPT13:	+0.248	+0.033	+0.064	+0.161	+0.429	+0.107	+1.000	+0.334	+0.600	+0.206	+0.600	+0.242	+0.508	+0.367	+0.461	+0.587	+0.444	+0.821	+0.048	+0.293	+0.023	+0.618	+0.407
DPT4:	+0.068	+0.024	+0.226	+0.087	+0.231	+0.095	+0.334	+1.000	+0.241	+0.117	+0.241	+0.107	+0.137	+0.217	+0.509	+0.226	+0.512	+0.334	+0.025	+0.152	+0.015	+0.257	+0.463
DPT5:	+0.308	+0.038	+0.151	+0.150	+0.328	+0.195	+0.600	+0.241	+1.000	+0.259	+1.000	+0.322	+0.388	+0.337	+0.305	+0.431	+0.270	+0.604	+0.004	+0.213	+0.040	+0.450	+0.350
DPT6:	+0.067	+0.011	+0.372	+0.112	+0.201	+0.119	+0.206	+0.117	+0.259	+1.000	+0.259	+0.092	+0.092	+0.121	+0.249	+0.136	+0.298	+0.220	+0.011	+0.133	+0.015	+0.153	+0.389
DPT7:	+0.308	+0.038	+0.151	+0.150	+0.328	+0.195	+0.600	+0.241	+1.000	+0.259	+1.000	+0.322	+0.388	+0.337	+0.305	+0.431	+0.270	+0.604	+0.005	+0.213	+0.040	+0.451	+0.350
DM1:	+0.110	+0.049	+0.048	+0.258	+0.461	+0.350	+0.242	+0.107	+0.322	+0.092	+0.322	+1.000	+0.477	+0.388	+0.211	+0.503	+0.399	+0.585	+0.025	+0.308	+0.034	+0.429	+0.309
DE1:	+0.114	+0.036	+0.010	+0.172	+0.353	+0.205	+0.508	+0.137	+0.388	+0.092	+0.388	+0.477	+1.000	+0.847	+0.600	+0.965	+0.421	+0.631	+0.028	+0.236	+0.019	+0.469	+0.274
DE2:	+0.122	+0.026	+0.108	+0.114	+0.252	+0.135	+0.367	+0.217	+0.337	+0.121	+0.337	+0.388	+0.847	+1.000	+0.667	+0.784	+0.398	+0.469	+0.028	+0.167	+0.018	+0.350	+0.295
DE3:	+0.010	+0.019	+0.242	+0.099	+0.248	+0.088	+0.461	+0.509	+0.305	+0.249	+0.305	+0.211	+0.600	+0.667	+1.000	+0.669	+0.446	+0.448	+0.040	+0.170	+0.010	+0.335	+0.470
EJ1:	+0.177	+0.043	+0.047	+0.197	+0.408	+0.234	+0.587	+0.226	+0.431	+0.136	+0.431	+0.503	+0.965	+0.784	+0.669	+1.000	+0.477	+0.709	+0.029	+0.275	+0.024	+0.529	+0.368
PTJ3L1:	+0.095	+0.041	+0.240	+0.252	+0.526	+0.359	+0.444	+0.512	+0.270	+0.298	+0.270	+0.399	+0.421	+0.398	+0.446	+0.477	+1.000	+0.689	+0.026	+0.355	+0.026	+0.525	+0.509
PTJ1J2:	+0.240	+0.060	+0.065	+0.297	+0.657	+0.390	+0.821	+0.334	+0.604	+0.220	+0.604	+0.585	+0.631	+0.469	+0.448	+0.709	+0.689	+1.000	+0.023	+0.444	+0.032	+0.754	+0.495
MBJP1:	-0.018	+0.065	+0.014	+0.035	+0.049	+0.080	+0.048	+0.025	+0.004	+0.011	+0.005	+0.025	+0.028	+0.028	+0.040	+0.029	+0.026	+0.023	+1.000	+0.088	+0.062	+0.028	+0.062
MBJP2:	+0.080	+0.063	+0.040	+0.265	+0.552	+0.331	+0.293	+0.152	+0.213	+0.133	+0.213	+0.308	+0.236	+0.167	+0.170	+0.275	+0.355	+0.444	+0.088	+1.000	+0.035	+0.563	+0.408
MJP1:	+0.037	+0.016	+0.019	+0.027	+0.033	+0.036	+0.023	+0.015	+0.040	+0.015	+0.040	+0.034	+0.019	+0.018	+0.010	+0.024	+0.026	+0.032	+0.062	+0.035	+1.000	+0.034	+0.034
MJP2:	+0.189	+0.051	+0.042	+0.216	+0.472	+0.279	+0.618	+0.257	+0.450	+0.153	+0.451	+0.429	+0.469	+0.350	+0.335	+0.529	+0.525	+0.754	+0.028	+0.563	+0.034	+1.000	+0.370
M1:	+0.222	+0.087	+0.400	+0.259	+0.539	+0.279	+0.407	+0.463	+0.350	+0.389	+0.350	+0.309	+0.274	+0.295	+0.470	+0.368	+0.509	+0.495	+0.062	+0.408	+0.034	+0.370	+1.000



TMVA overtraining check for classifier: WpH_Tagger_v2
Signal (test sample)
Background (test sample)
Signal (training sample)
Background (training sample)
Kolmogorov-Smirnov test: signal (background) probability = 0 (0)
(1/N) dN / dx
20
18
16
14
12
10
8
6
4
2
0
–1
–0.8
–0.6
–0.4
–0.2
0
0.2
0.4
0.6
0.8
WpH_Tagger_v2 response
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%


Background rejection versus Signal efficiency
Background rejection
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Signal efficiency
MVA Method:
WpH_Tagger_v2

4.3 Future of the analysis

- Remove variables that don't benefit the tree → Look at correlations and try to reduce them as good as possible
- Remove variables that don't sufficiently separate signal and background
- Try out new variables and see if adding them helps the analysis
- Right now analysis only runs for $m_{H^+} = 800$ GeV, but we also want to do it for $m_{H^+} = 400$ GeV and $m_{H^+} = 1600$ GeV