Machine learning for charged Higgs

References : “Deep-learning in Search of Light Charged Higgs” (1803.01550)

“Search for a charged Higgs boson in pp collision at $\sqrt{s} = 8$ TeV” (1508.07774) JHEP11(2015) 018

1 September @ Konkuk University, Korea

Jubin Park (Chonnam National University)

Contents

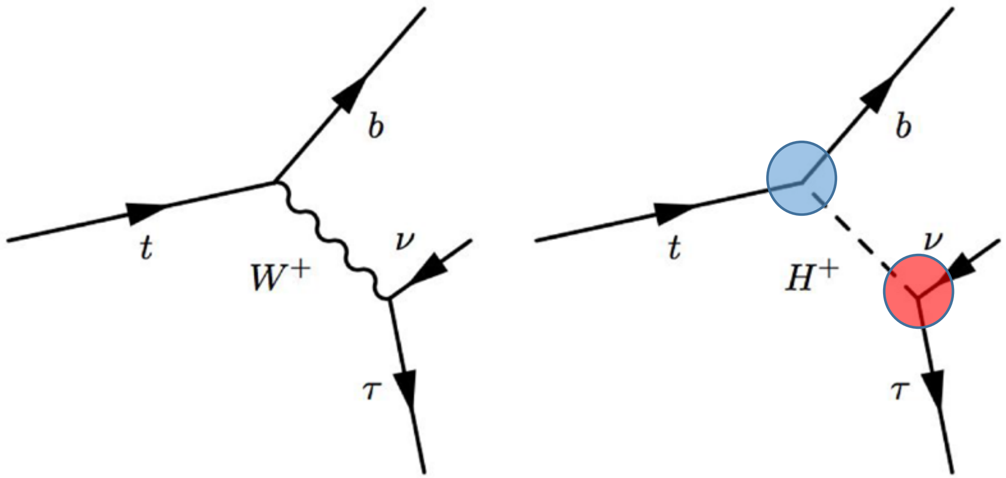
- Motivations
- Goal(s)
- Experimental bounds
- Golden channel and CP-conserving Type-X) 2HDM
- Simulation(s) and Feature selection I and II
- Machine learning methods (DNN, SNN)
- Results
- Conclusion

Motivations

- Charged Higgs boson arise in models with more than one Higgs doublet. It takes part in the charged current interactions.
- For example, 2HDM or MSSM are the simplest models that can have a charged Higgs boson.
- LHC is essentially a top quark factory. It is a natural place to search for the charged Higgs via top quark decay.

Goal(s)

- To disentangle the charged Higgs effects from the background.
- It becomes a challenging task for light ($m_{H^\pm} \gtrsim$ 110 GeV), and weakly-coupled charged Higgs particles.



$-i\sqrt{2}\frac{m_t}{v}\,\zeta_u$, $\zeta_u = \cot\beta$

Type-X 2HDM

$-i\sqrt{2}\frac{m_\tau}{v}\,\zeta_l$, $\zeta_l = -\,tan\,\beta$

Experimental Bounds

- Actually $m_{H^{\pm}}$ can be appreciably lower !!! [22]
- The discovery of the scalar boson (125 GeV) at the ATLAS and CMS experiments forces us to set one of the neutral states as the Higgs boson. (h → the recently observed scalar with a 125 GeV mass → the SM-like Higgs boson !!)
  → Exact alignment limit and the mixing angle between the CP-even Higgs states becomes as $\sin(\beta - \alpha)$ ~ 1.
- In the CP-conserving Type-X 2HDM,

$$100 \text{ GeV} \leq m_{H^{\pm}} \leq 120 \text{ GeV}$$

is allowed by the current experimental bounds.

In the past … The search for $H^{\pm}$

- Through the pair production at the LEP [24-29].
- Via top quarks decays at the Tevatron [30-34].

The golden channel(s)

- Through the top quark decay :
  $W^{\pm}$ (BG : background) and $H^{\pm}$ (signal) contributions are seen to have identical topologies.
  $W^{\pm}$ diagrams generates the irreducible SM background.
  The effect of the low-mass charged Higgs can be shadowed by the strong W boson contamination.
  →It is difficult to disentangle using cut-based methods.
- There are studies that use the cut-based analysis to extract the signal. [3-5]
- For the reference, the single top quark cross section associated either with b-quark or W^+ boson is around 3.04 pb at NLO.

(CP-conserving Type-X) 2HDM (100~120 GeV)

- Goal : to extract a charged Higgs boson information from single top quark decays.
- The model contains a total of five Higgs bosons : the neutral ones (h, H, A), and the charged on ($H^{\pm}$).
- Type-X, I, II and Y .... -> Theory talk(s) !

Usual cuts in the detector

- The number of leptons at the final-state -> One lepton !
- The number of jets at the final-state -> b-jets !
- Kinematical constraints on phase space of reducible and irreducible backgrounds

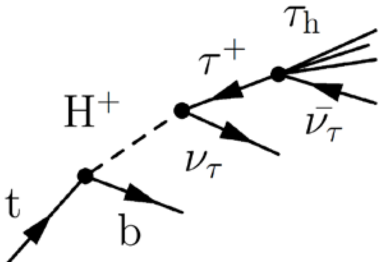
The charged Higgs signal of interest takes the form

$$p\,p \quad \to \text{b-jet} + \text{N-jets} + \ell + \not{E}$$

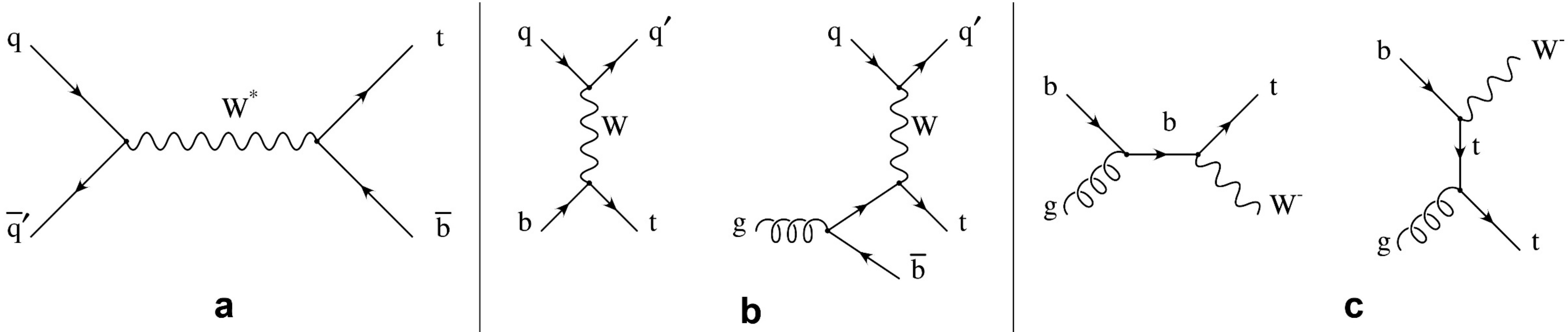
Irreducible background $W^{\pm}$

$H^{\pm}$ **Signal**

Our broad steps to the goal(s)



- Production of top quarks via pp -> t X at the LHC energy.
- Decay of the top quark : $t \rightarrow b\, H^+$
- Extraction of the charged Higgs information from the data.



Single top production !

Event generation and Feature Selection

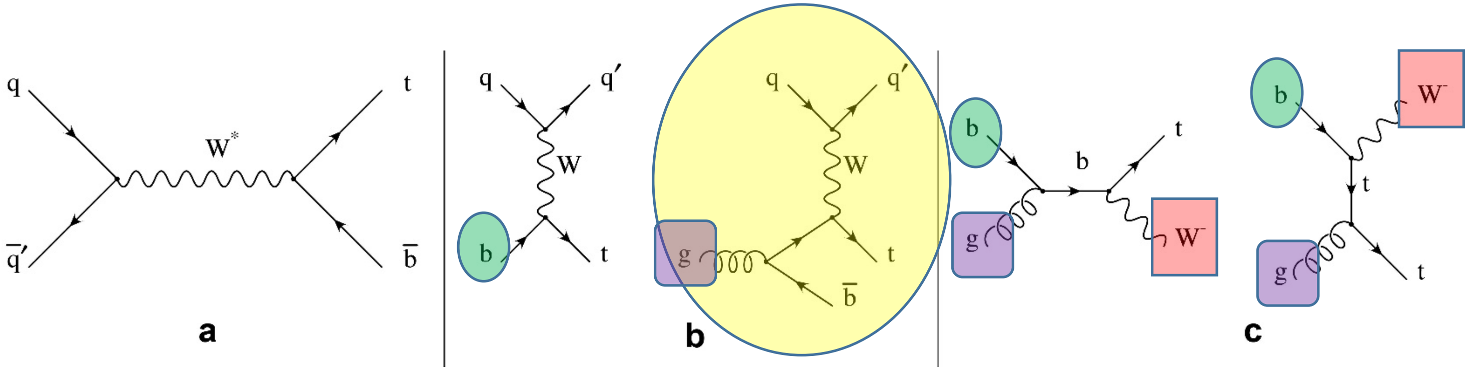
- Analyze the charged Higgs production and its leptonic decay
- We take the charged Higgs boson mass as 90, **100,** 110 and 120 GeV.
- High energy proton-proton (pp) collisions are simulated at the LHC energies ($\sqrt{s}$=13 TeV).
- To simulate events, we use MADGRAPH[40]

Note !!!

- The top quark is allowed decaying with charged Higgs and b-quark with BR($t \rightarrow b\, H^{+}$) = 0.1120.
- Event must have one and only one isolated lepton (coming from the tau decays).
- Preselection cuts
  P_T (jet, b-jet, leptons(e, μ)) > 20 GeV
  |η| (jet, b-jet, leptons(e, μ)) < 2.4
  ΔR (jet, b-jet, leptons(e, μ)) > 0.4
  MET (due to neutrinos) > 20 GeV

Signal
+
Irreduc. BG

Process	s-chan.	tW-chan.	t-chan
single top	0.549 [41]	71.700 [42, 43, 41]	216.99 [41]



Reducible
Backgrounds

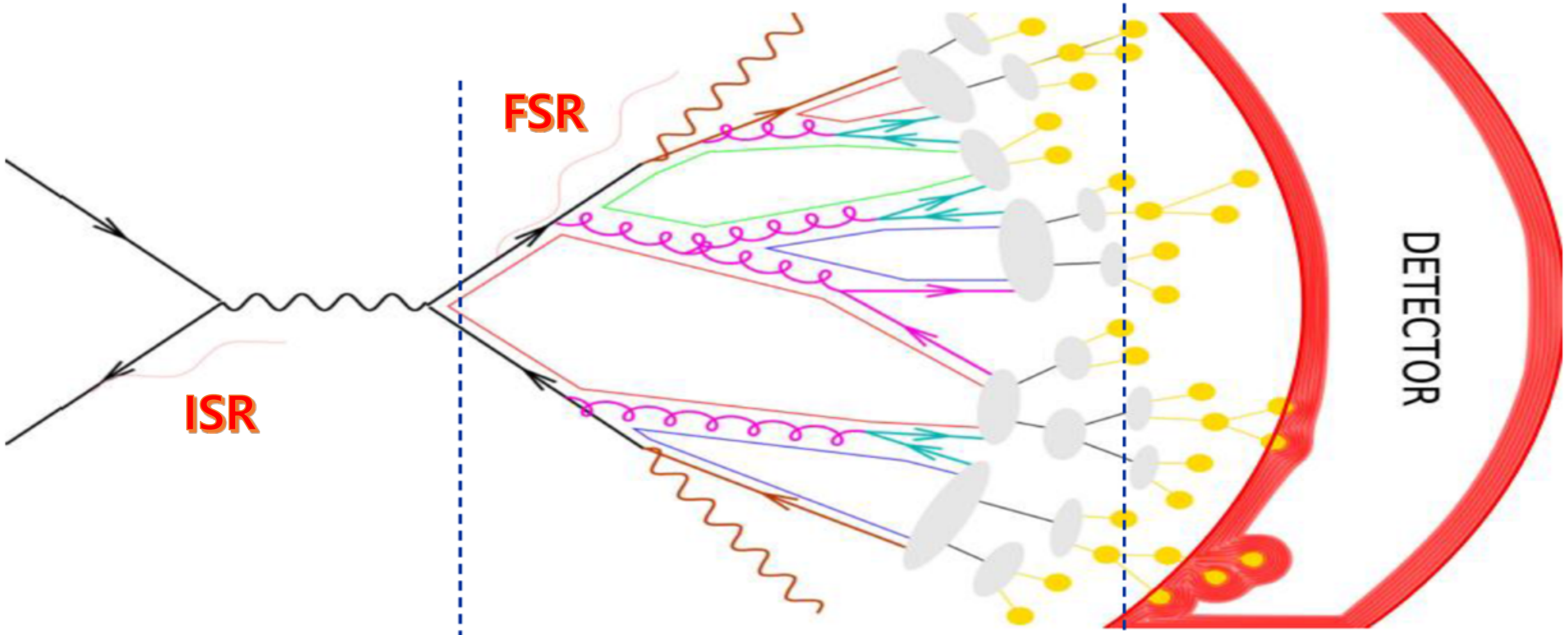
The process	+0 jets	+1 jets	+2 jets
$W^{\pm}$	24390	4288	1637
$W^{\pm}c$	586.0	503.0	269.6
$W^{\pm}bb$	5.01	8.5	9.6
	leptonic	semi-leptonic	hadronic
$t\bar{t}$	22.45	69.78	116.16

Outline of simulations

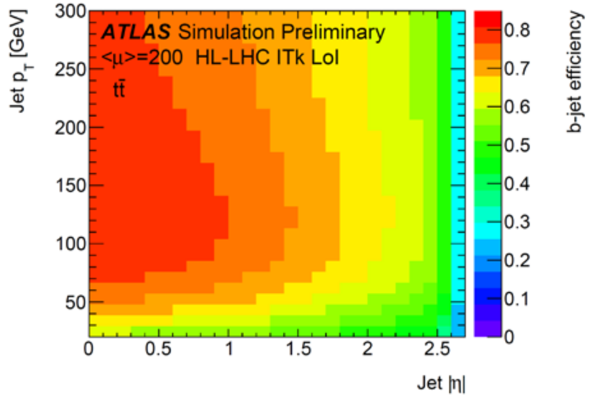
Hard Scattering Part

Hadronization and showering Processes

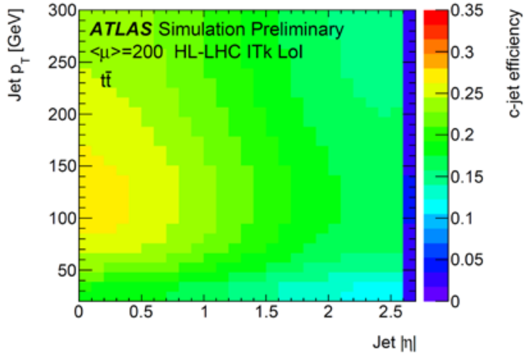
Reconstruction level



b-jet efficiency



c-jet efficiency



Delphes3

Parton level	Hadron level	Reconstructed objects level
Madgraph5_aMC@NLO	**PYTHIA6**	**Delphes3**
LHE files	**STDHEP/HEPMC files**	**LHCO, ROOT files**

ATL-PHYS-PUB-2016-026

2015 MadGraph school on Collider Phenomenology November 23-27 @ Shanghai

Feature Set 1 : Usual kinematic variables

1. p_T^{ℓ} : transverse momentum of the lepton (muon or electron).
2. $\not{E}_T$: Missing Transverse Energy (MET) of neutrinos,
3. N^{jets} : number of jets in an event,
4. N^{b-tag} : number of b-tagged jets in an event,
5. $\Delta R_{\ell b}$: distance between the lepton and the b-tagged jet (distance or angle),
6. $\Delta \eta_{\ell b}$: relative rapidity between the lepton and b-tagged jet (distance or angle),
7. p_T^{jet-1} : transverse momentum of the highest-energy jet,
8. η^{jet-i} : rapidity of the highest-energy jet,
9. ϕ^{jet-i} : azimuthal angle of the highest-energy jet,
10. b-tag: b-tagging information,
11. η_{ℓ} : lepton rapidity,
12. ϕ_{ℓ} : lepton azimuthal angle,
13. $\eta^{\not{E}_T}$: MET rapidity,
14. $\phi^{\not{E}_T}$: MET azimuthal angle,
15. p_T^{b-tag} : transverse momentum of b-tagged jet.

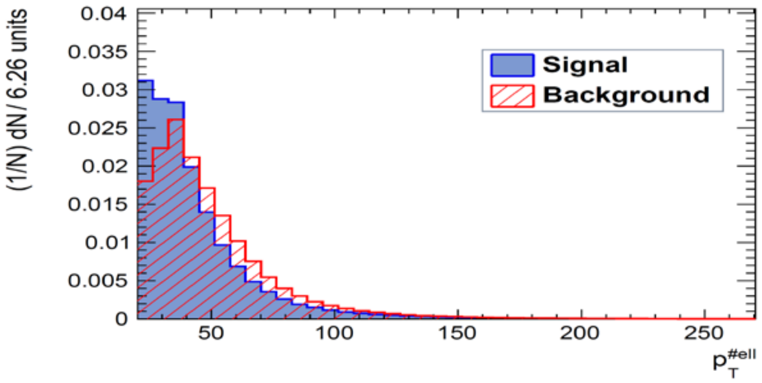
Feature Set 2 : Human engineered features

1. $M_T^{\ell\nu}$: transverse mass of leptons (it can be written as $M_T^{\ell\nu} = \sqrt{2 p_{\ell T} \not{p}_T}$ where $p_{\ell T}$ is the lepton transverse momentum).
2. $M_{\ell\nu b}$: the reconstructed top quark mass,
3. M_{jj} : dijet invariant mass (the highest-energy two jets),
4. M_{jjj} : trijet invariant mass,
5. α_T : parameter controlling the QCD background ($\alpha_T = p_T^{jet-2}/m_{jj}$).

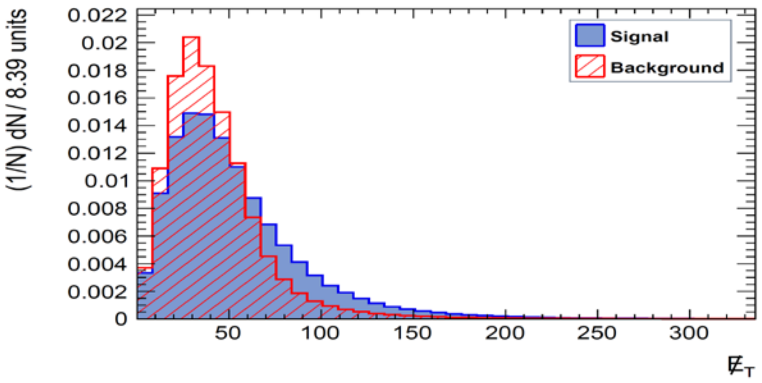
They can be expressed as combinations of the features in FS1

Distributions of some features (FS1)

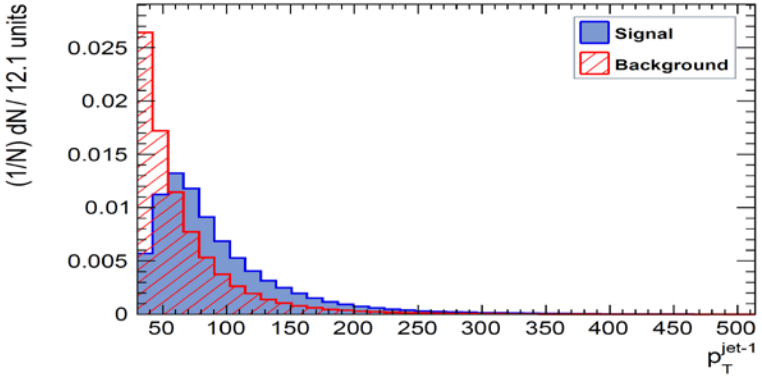
Sometimes, applying cuts on individual features
may not always lead to an effective signal extraction


(1/N) dN / 6.26 units
Signal
Background
p_T^{#ell}

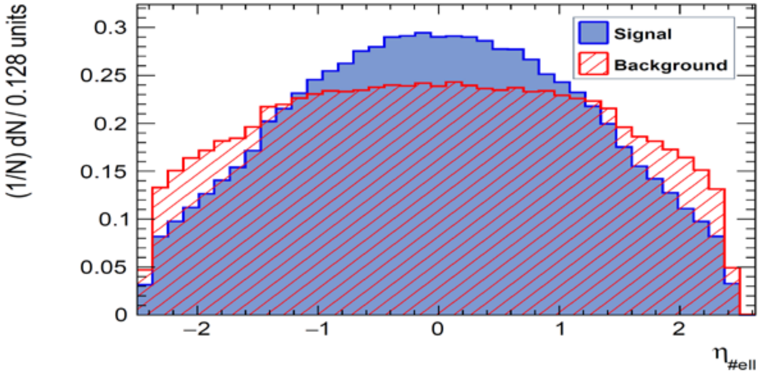
(a)


(1/N) dN / 8.39 units
Signal
Background
E̸_T

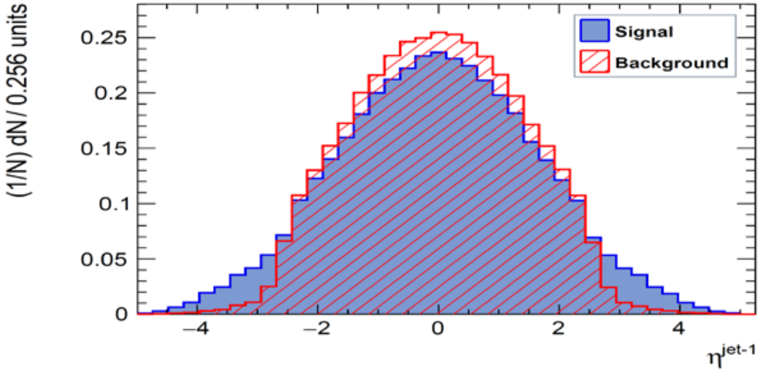
(b)


(1/N) dN / 12.1 units
Signal
Background
p_T^{jet-1}

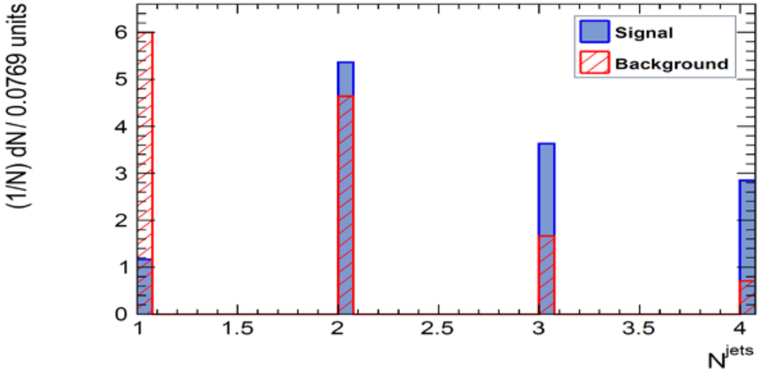
(c)


(1/N) dN / 0.128 units
Signal
Background
η_{#ell}

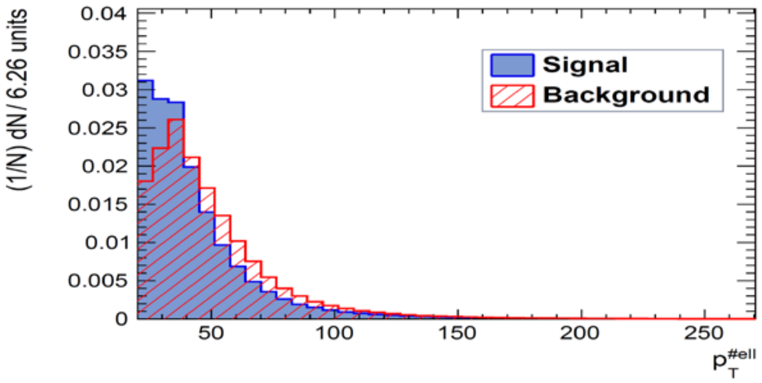
(d)


(1/N) dN / 0.256 units
Signal
Background
η^{jet-1}

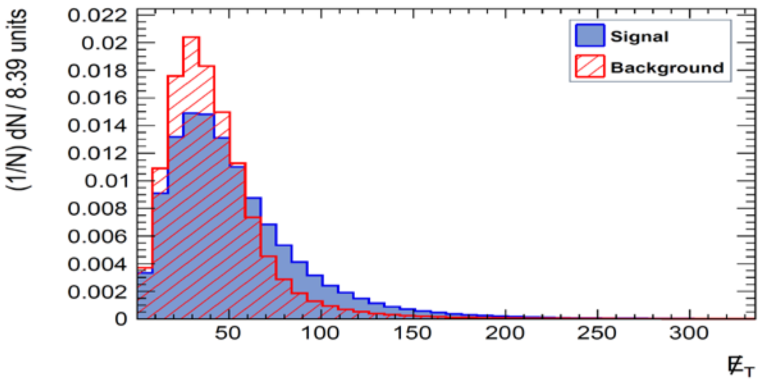
(e)


(1/N) dN / 0.0769 units
Signal
Background
N^{jets}

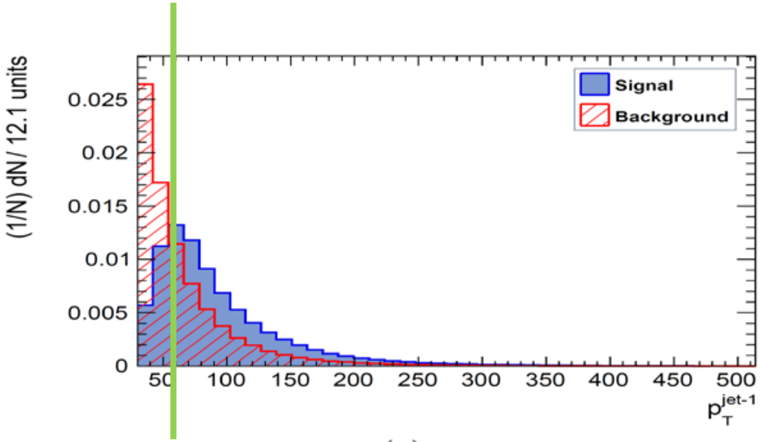
(f)


(1/N) dN / 6.26 units
Signal
Background
$p_T^{\#ell}$

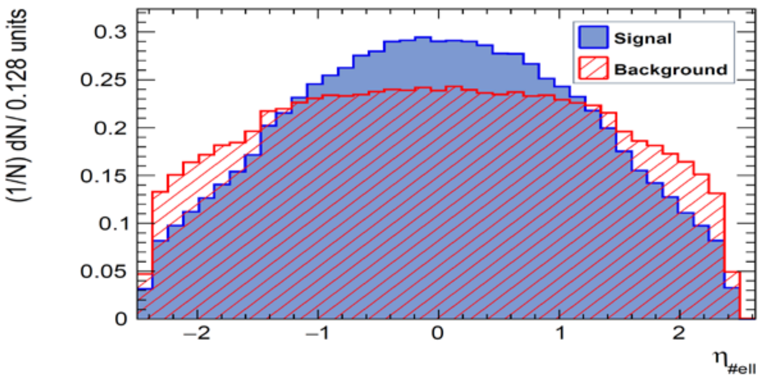
(a)


(1/N) dN / 8.39 units
Signal
Background
$\not{E}_T$

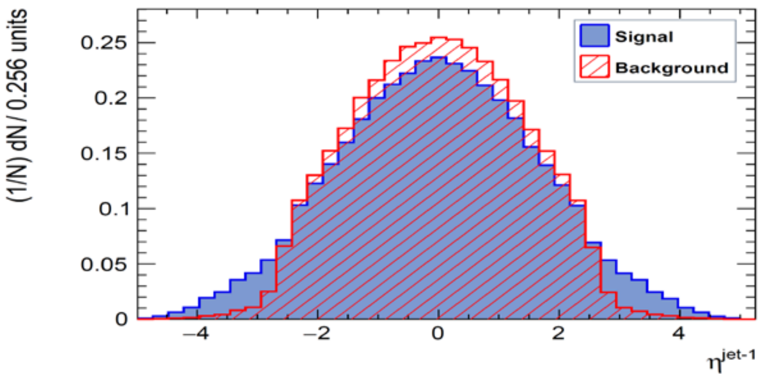
(b)


(1/N) dN / 12.1 units
Signal
Background
p_T^{jet-1}

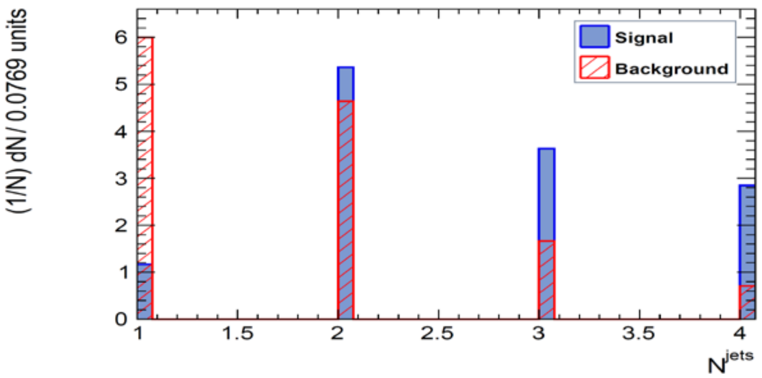
(c)


(1/N) dN / 0.128 units
Signal
Background
$\eta_{\#ell}$

(d)

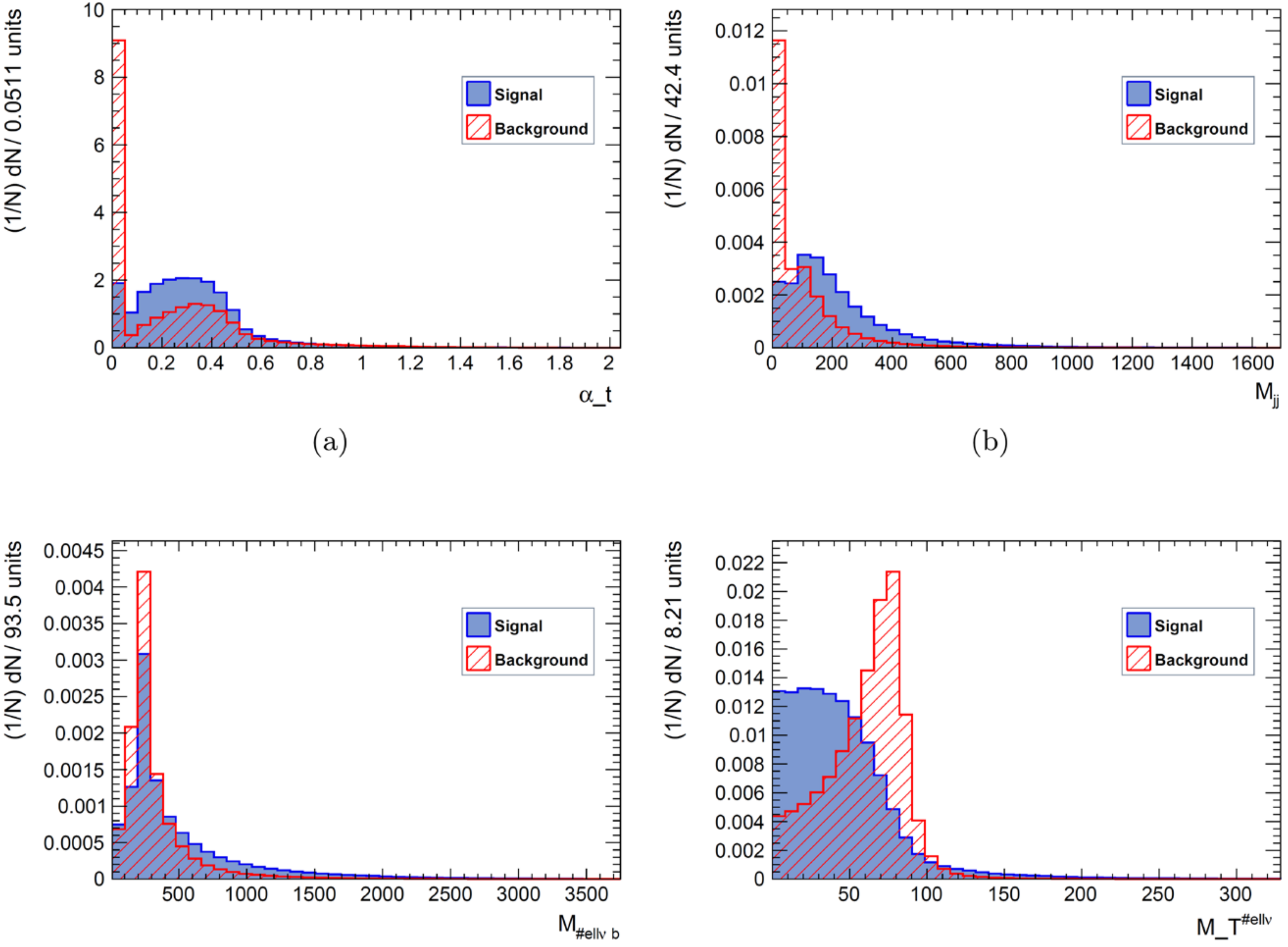

(1/N) dN / 0.256 units
Signal
Background
η^{jet-1}

(e)


(1/N) dN / 0.0769 units
Signal
Background
N^{jets}

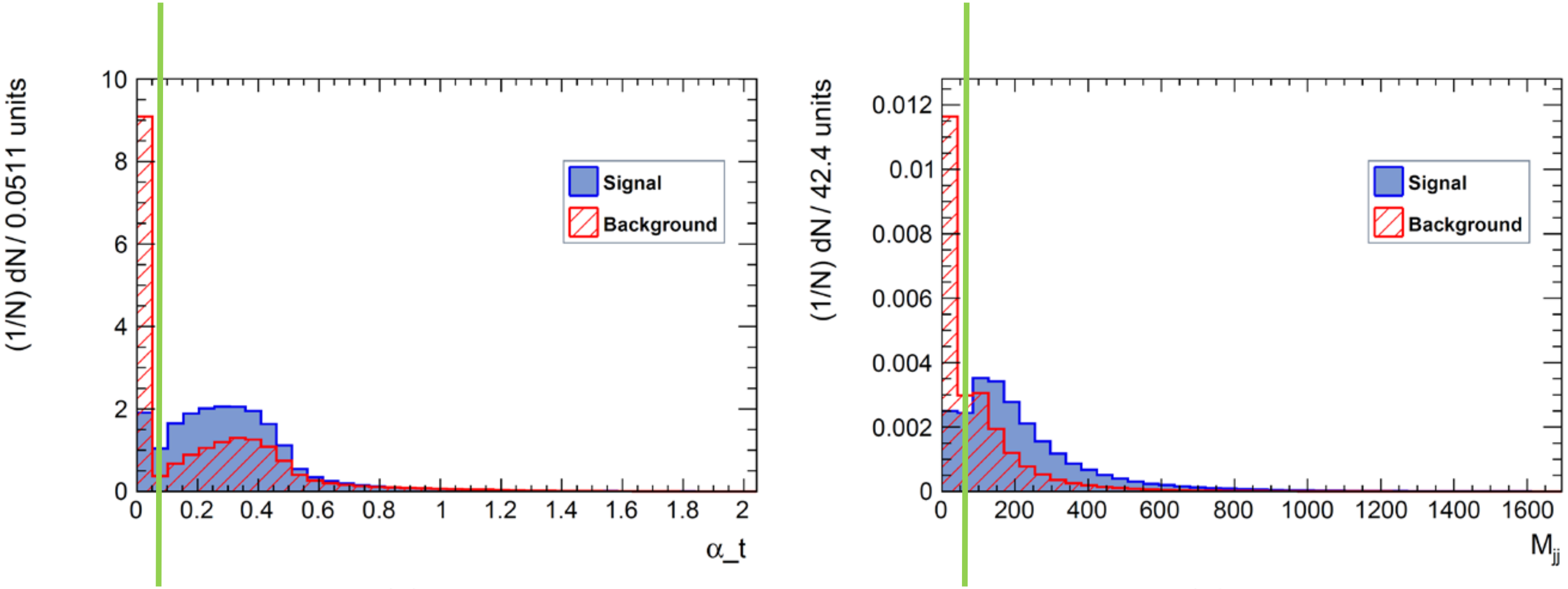
(f)

Distributions of some features (FS2)


(1/N) dN / 0.0511 units
α_t
(1/N) dN / 42.4 units
M_jj
(1/N) dN / 93.5 units
M_#ellν b
(1/N) dN / 8.21 units
M_T^#ellν
Signal
Background

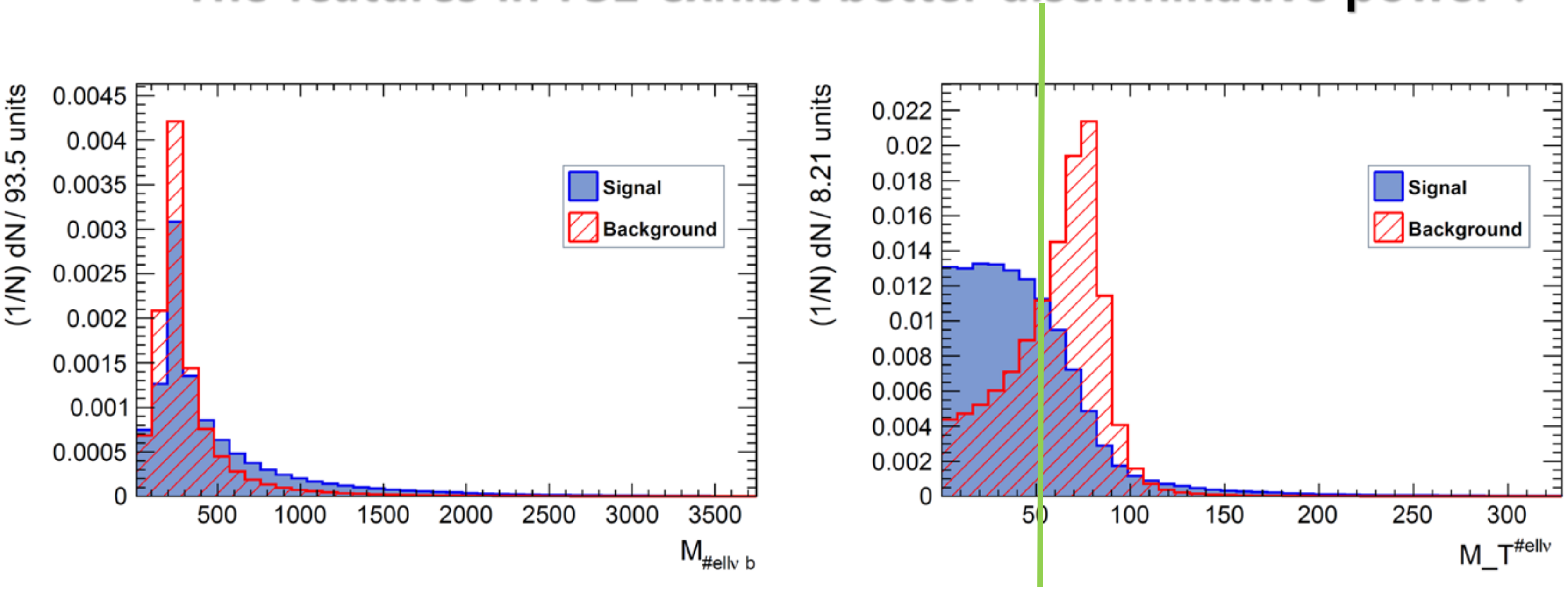
(a) (b)

(c) (d)



(a) (b)

The features in FS2 exhibit better discriminative power !



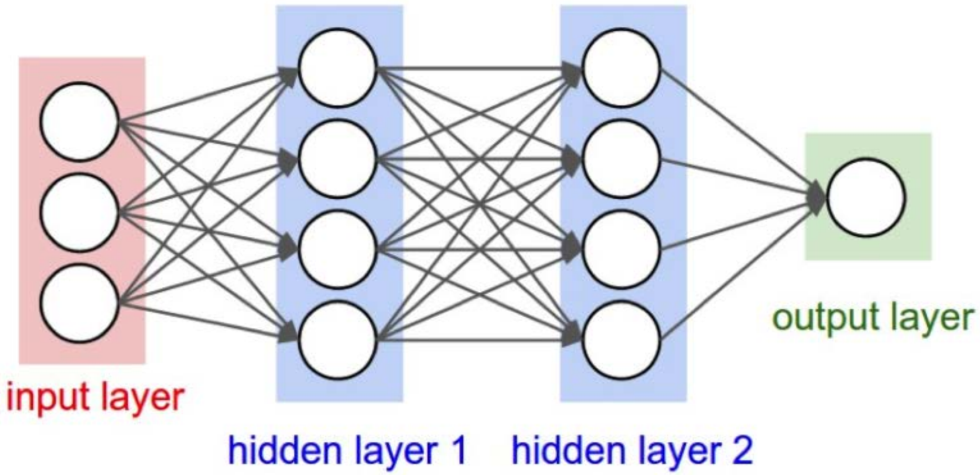
(c) (d)

Machine learning methods

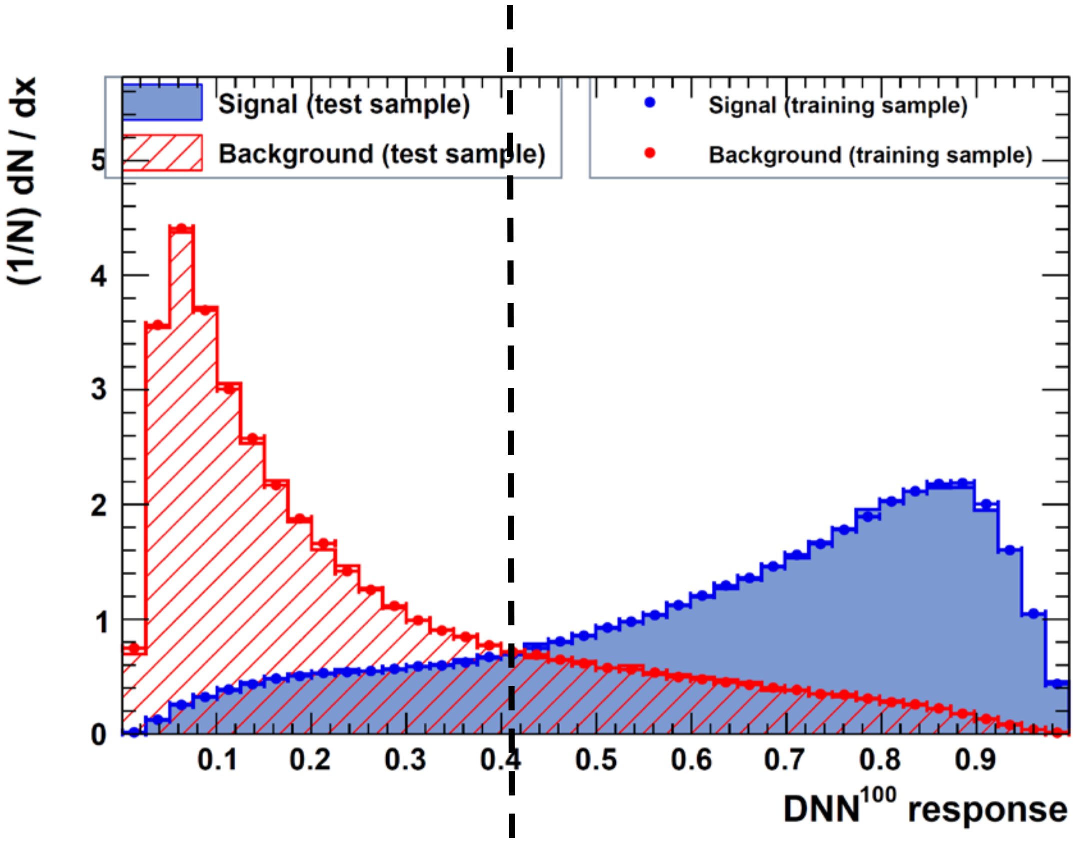
- Artificial neural networks (ANN), biology-inspired computational methods and other methods have been applied.
- Single-layer ANNs exhibit difficulties in learning particle properties.
- Deep neural networks (DNN), endowed with various layers, are more effective alternatives compared to single-layer ANNs.

Two DNN structures

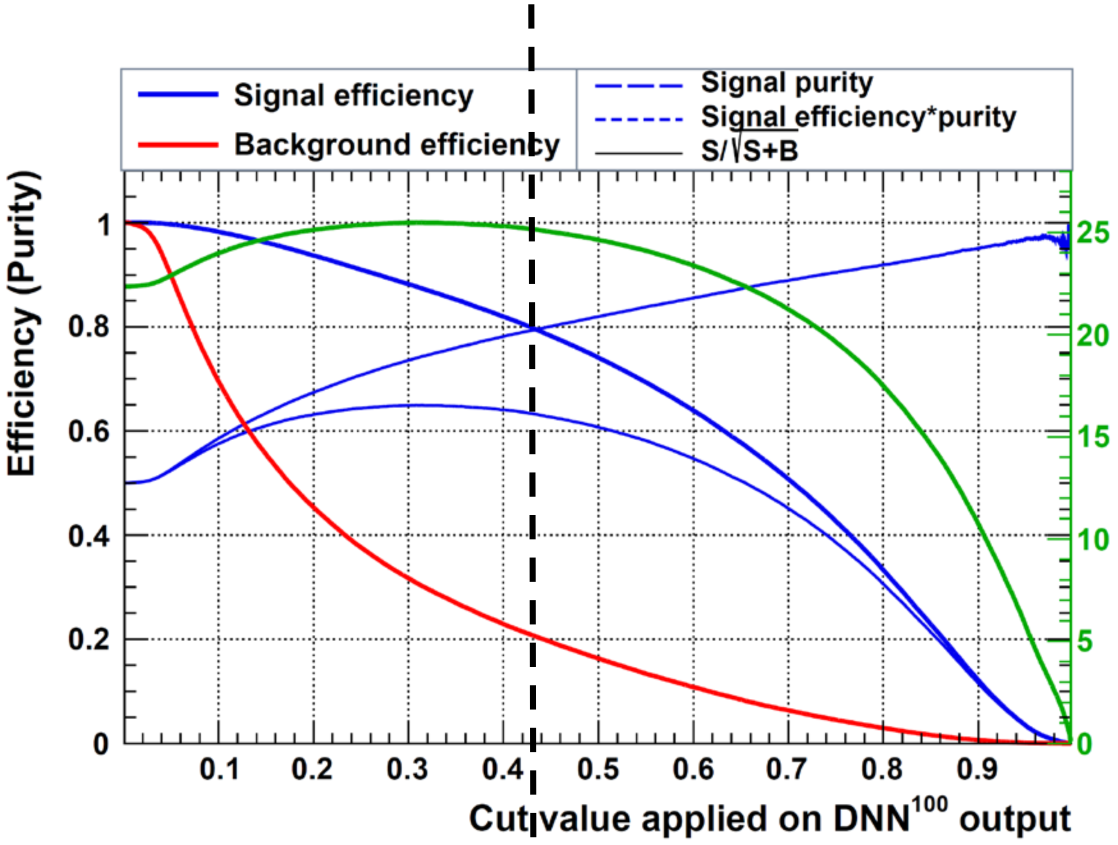
- Three hidden layers N hyperbolic tangent units
- A linear output unit with cross-entropy loss
- DNN with N=100 : DNN^{100}
- DNN with N=300 : DNN^{300}



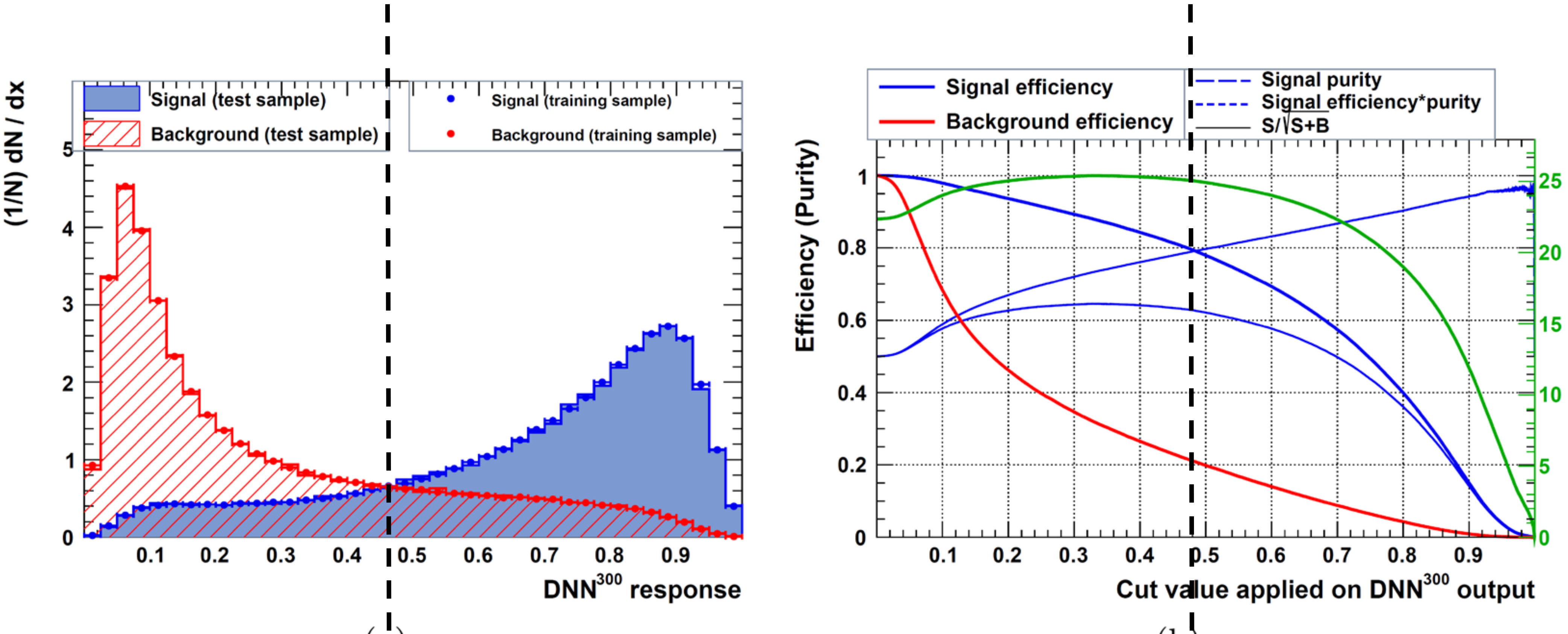
Extracting Charged Higgs with DNN



(a)



(b)

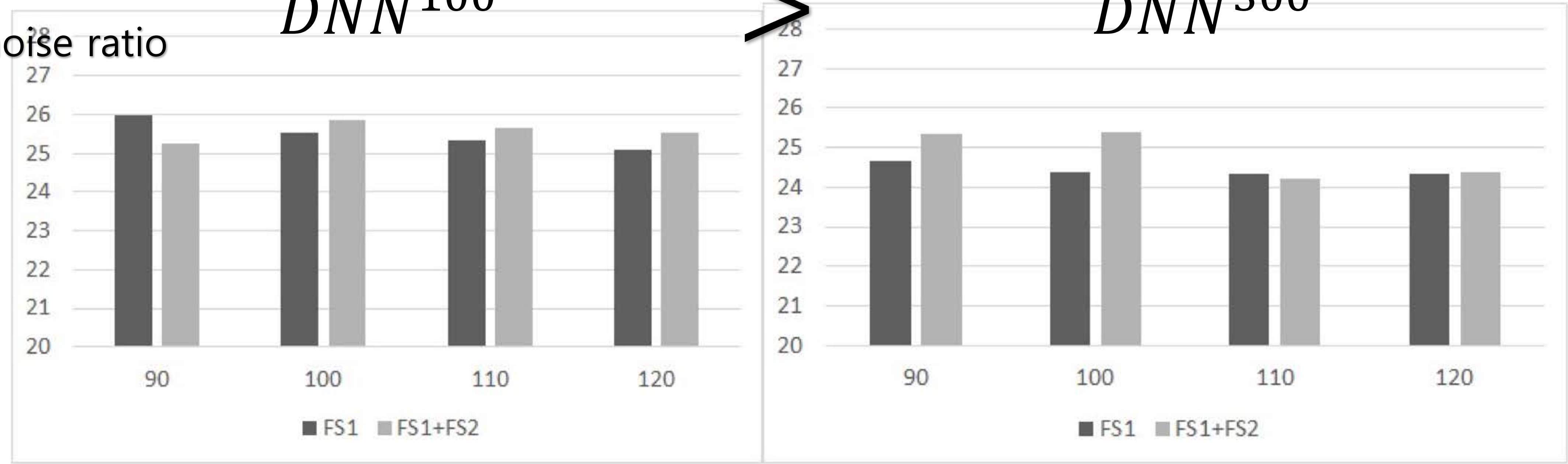


Signal (test sample)
Background (test sample)
Signal (training sample)
Background (training sample)
(1/N) dN / dx
DNN^300 response
Signal efficiency
Background efficiency
Signal purity
Signal efficiency*purity
S/√S+B
Efficiency (Purity)
Cut value applied on DNN^300 output

(a) (b)

Signal-to-noise ratio $\frac{s}{\sqrt{s+B}}$

$DNN^{100} > DNN^{300}$



(a) (b)

1. Increasing the number of processing units in DNN does not proportionally increase efficiency

2. Inclusion of high-level features (FS1 -> FS1+FS2) does not lead to any dramatic increase in $\frac{s}{\sqrt{s+B}}$.

The human engineered features in FS2 are replaced by the learned non-linear relationships among layers of the DNN !

DNN v.s. SNN in FS1

Signal efficiencies

$m_{H^\pm}$	DNN^{100}	DNN^{300}	SNN^{20}	SNN^{100}	SNN^{300}
90	0.880	0.8588	0.886	0.890	0.887
100	0.878	0.8741	0.869	0.878	0.873
110	0.874	0.8778	0.844	0.847	0.842
120	0.897	0.8717	0.838	0.835	0.843

DNN v.s. SNN in FS1+FS2

Signal efficiencies

$m_{H^\pm}$	DNN^{100}	DNN^{300}	SNN^{20}	SNN^{100}	SNN^{300}
90	0.871	0.877	0.863	0.874	0.871
100	0.884	0.881	0.878	0.881	0.879
110	0.886	0.881	0.888	0.877	0.869
120	0.868	0.868	0.897	0.875	0.843

Conclusion

- DNNs themselves find regions of high efficiency.
- Increasing the number of processing units in DNNs does not necessarily cause an increase in efficiency due mainly increased complexity.
- DNNs can be of critical help in extracting the signal information from strong BG contamination.

Thank you for your attention !

Backup slides