Phenomenology of vector-like leptons with Deep Learning at the Large Hadron Collider

Felipe Freitas[1], **J. P. Pino[1]**, António P. Morais[1], Roman Pasechnik[2]

Workshop on Compact Objects, Gravitational Waves and Deep Learning

[1]Departamento de Física, Universidade de Aveiro e CIDMA [2]Department of Astronomy and Theoretical Physics, Lund University

The need to go beyond the SM

The SM is, at this point in time, the most sucessful theory that describes all subatomic phenomena (S. Chatrchyan 2012, F. Abe 1995). However, it still lacks in some key areas:

- Inability to explain the observed particle spectra (family replication, masses and couplings hierarchies);
- Lack of dark matter candidate;
- Neutrino mass generation;
- Hiearchy problem.

Problems arise from the effective nature of the model. A new model **inspired** on unification and supersymmetry principles is introduced (Camargo-Molina 2019 and 2017, Morais 2020). The model brings with it new states including VLLs, VLQs, new scalars and new neutrinos.

Presentation outline

1. A brief introduction of the 3HDM-SHUT model
   A very brief overview of the particle content in the model and showcase of the potential mass ranges to be studied.

2. Numerical studies of the 3HDM-SHUT low energy limit
   Numerical workflow and analysis is performed with focus in vector-like lepton phenomonology.

3. Conclusions

At the low energy limit, this framework results in 2 new VLQs.

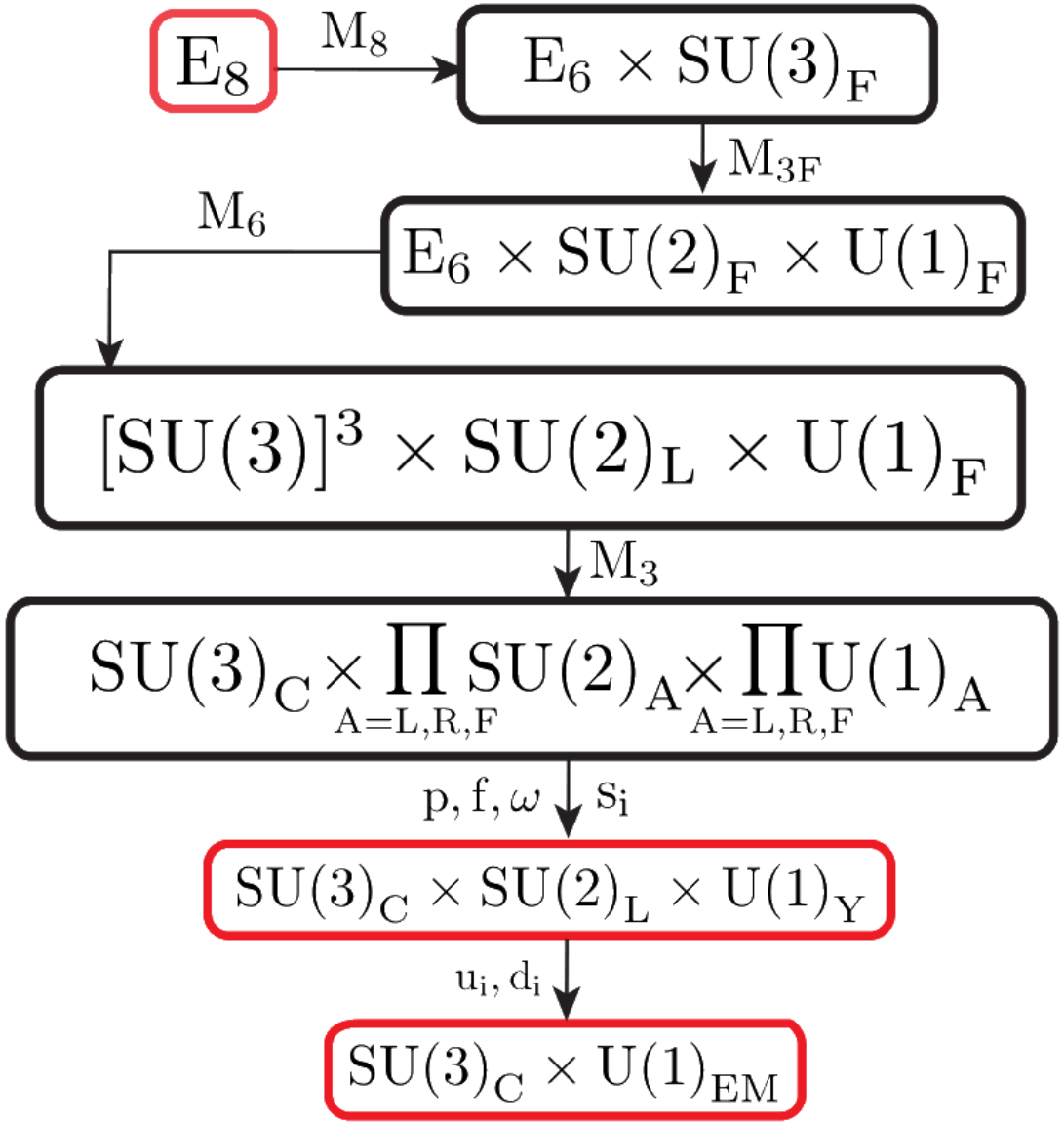
3 generations of VLLs, with two distinct scenarios (e_4, e_5 and e_6)

1. Two light states (m < 1 TeV) and one heavy state (m > 1 TeV);
2. One light state and two heavy states.

A total of 15 neutrino states

1. 9 $SU(2)_L$ doublets (6 BSM and 3 SM-like);
2. 6 $SU(2)_L$ singlets (all BSM).

The framework allows for the existance of sterile neutrinos (in the keV-MeV range).



SARAH model and UFO files (1)

1. Lagrangian Density implementation,
2. Generation of UFO files with all interaction vertices.

MadGraph hard scattering (2)

1. pp collisions at 14 TeV for signals and backgrounds,
2. 250k events for each topology.

PYTHIA8 + Delphes + ROOT (3)

1. Event hadronization,
2. ATLAS detector effects,
3. Angular and kinematic variables.

Deep Learning analysis (4)

1. Significance studies for different statistical models,
2. Evolutive algorithm, with Keras and TensorFlow.

In this work, we consider the main irreducible backgrounds to be:

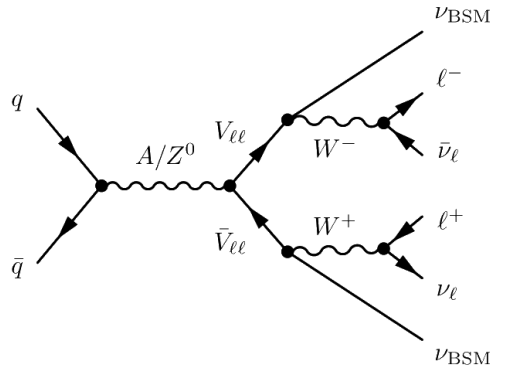
1. For ZA topologies: $t\bar{t}$; $t\bar{t} + Z^0(l^+l^-)$; $t\bar{t} + Z^0(\nu_l\bar{\nu}_l)$ and W^+W^-,
2. For VLBSM topologies: $l\nu_l$; $l\nu_l + (j, jj)$,
3. For VBF topologies: $t\bar{t} + (j, jj)$; W^+W^-.

For simplicity, we consider flavor opposite final states.

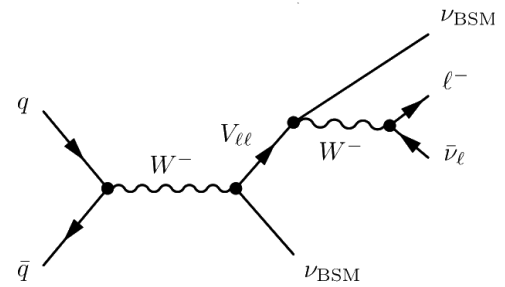
Event selection via simple cuts:

1. Charged leptons with $p_T > 25$ GeV and $|\eta| \leq 5$,
2. Missing transverse energy $\not{E}_T > 15$ GeV,
3. Jet events: $\Delta R = 1.0$, $p_T > 35$ GeV, $|\eta| \leq 5$.

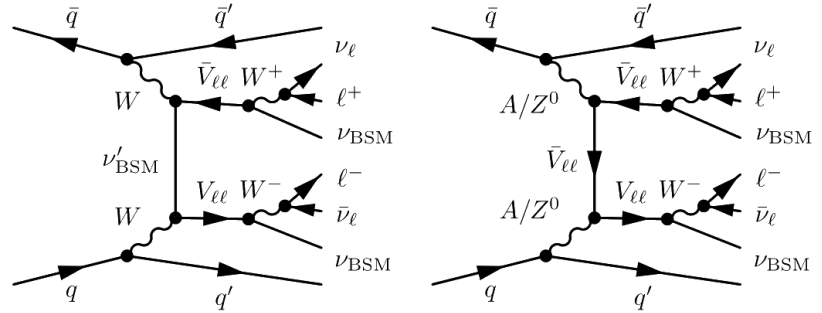
ZA signal



VLBSM signal



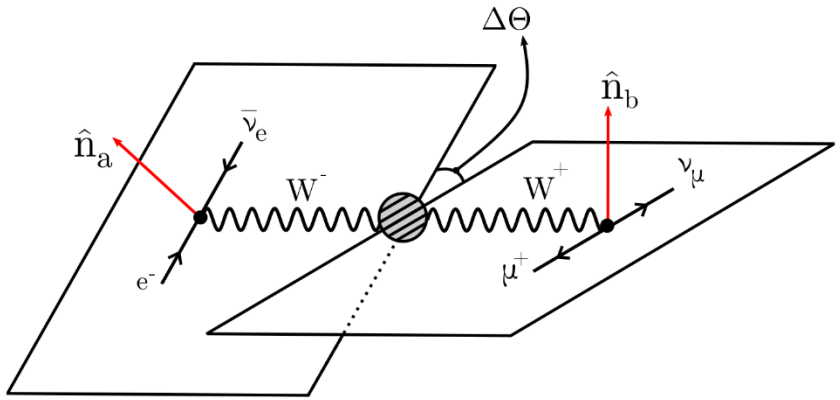
VBF signal



Input data set comes from kinematic/angular features of the topologies in question

	Dimension-full	Dimensionless	
Lab. frame	$p_T(e^-)$, $p_T(\mu^+)$,$p_T(e_4)$ $p_T(\bar{e}_4)$, $M(e_4)$, $M(\bar{e}_4)$ $M_T(W^-)$, $M_T(W^+)$, MET	$\cos(\theta_{\bar{\nu}_e e})$, $\cos(\theta_{\bar{\nu}_\mu \mu^+})$, $\cos(\theta_{W^-W^+})$, $\cos(\Delta\phi)$, $\cos(\Delta\theta)$, η_e, η_{μ^+}, η_{e_4}, $\eta_{\bar{e}_4}$	$\Delta R(e, \bar{\nu}_e)$, $\Delta R(\mu^+, \nu_{\mu^+})$
W^- frame	$p_T(e^-)$, $p_T(e_4)$	$\cos(\theta_{\bar{\nu}_e e})$, η_e, η_{e_4}	
W^+ pair frame	$p_T(\mu^+)$, $p_T(\bar{e}_4)$	$\cos(\theta_{\nu_\mu \mu^+})$, η_{μ^+}, $\eta_{\bar{e}_4}$	
$\ell'\bar{\ell}'$ frame		$\cos(\Delta\phi)$, $\cos(\Delta\theta)$	

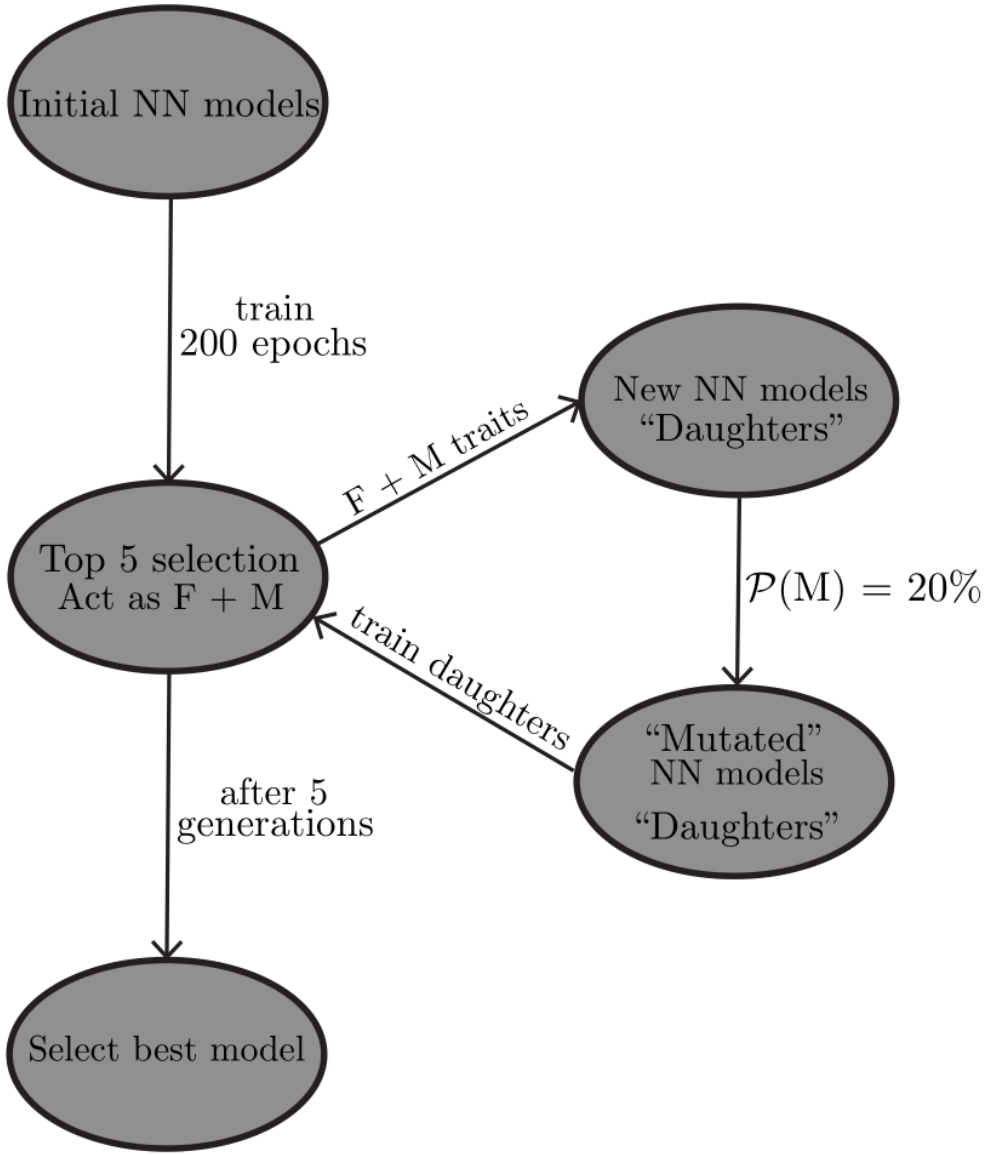
	Dimension-full	Dimensionless
Lab. frame	$p_T(\mu^+)$,$M_T(W)$, $p_T(W)$, MET	$\cos(\theta_{\mu^+})$, $\cos(\theta_{\bar{\nu}_\mu \mu^+})$, $\cos(\theta_W)$, η_{μ^+}, η_W



For VBF and ZA signals, we consider angular distributions between the W decay planes, in the vector-like center of mass frame.

A efficient way to search the best neural is through a evolutive algorithm, with hyperparameters:

1. Hidden Layers: 1 to 5;
2. Number of Neurons: 256, 512, 1024, 2048;
3. Kernel initializer: 'normal', 'he normal' and 'he uniform';
4. L2 regularizer with penalty: 10^{-3}, 10^{-5} and 10^{-7};
5. Activation function: 'relu', 'elu', 'tanh' and 'sigmoid';
6. Optimizer: 'adam', 'sgd', 'adamax' and 'nadam'.



The best model is choosen based on two distinct metrics: accuracy and the Asimov significance

$$\mathcal{Z}_A = \left[2\left((s+b)\ln\left(\frac{(s+b)(b+\sigma_b^2)}{b^2+(s+b)\sigma_b^2}\right) - \frac{b^2}{\sigma_b^2}\ln\left(1+\frac{\sigma_b^2 s}{b(b+\sigma_b^2)}\right)\right)\right]^{1/2}$$

Some characteristics of the model construction are universal to all architectures:

1. Inputs: Standard normalized vectors from ROOT observables. Training set with 80% of data, and test set with 20%;
2. Cyclic learning rate;
3. At output layer, data is transported with probabilities as entries for signal and background (S, B);
4. Batch size of 32 768 entries;
5. 200 epoch training with 5 epoch patience;
6. Accuracy models are found through the binary cross entropy. Asimov models are found via the above equation.

Cut-based analysis drops events and leads to unbalanced data sets

Original dataset:Training (SMOTE) Test	ZA	
	Signal	Backgrounds
$M(e_4) = 200$ GeV	(77405, 36) (65983, 36) (15481, 36)	$t\bar{t}$: (44725,36) **(A)** (65983, 36) **(B)** (8945, 36) **(C)** $t\bar{t}, Z^0(l^+l^-)$: (81825,36) (65983, 36) (16365, 36) $t\bar{t}, Z^0(\nu_\ell\bar{\nu}_\ell)$: (46705,36) (65983, 36) (9341, 36) W^+W^- : (48475,36) (65983, 36) (9695, 36)

(A) Dataset following cuts;
(B) Balanced training (80% + SMOTE) datasets;
(C) Remaining test (20%) datasets.

Synthetic datasets are generated via SMOTE, via oversampling of minority classes.

A more efficient method than having to run more Monte Carlo events.

The initial analysis is performed for the following parameters choice

	e_4	e_5	e_6
Mass (GeV)	677.27	3257.86	4239.84

	$\nu_{10,11}$	$\nu_{12,13}$	$\nu_{14,15}$
Mass (GeV)	m_{e_4}	m_{e_5}	m_{e_6}

	ν_4	ν_5	ν_6	ν_7	ν_8	ν_9
Mass (GeV)	2.1568×10^{-4}	0.138	36.74	2139.85	2536.56	3034.84

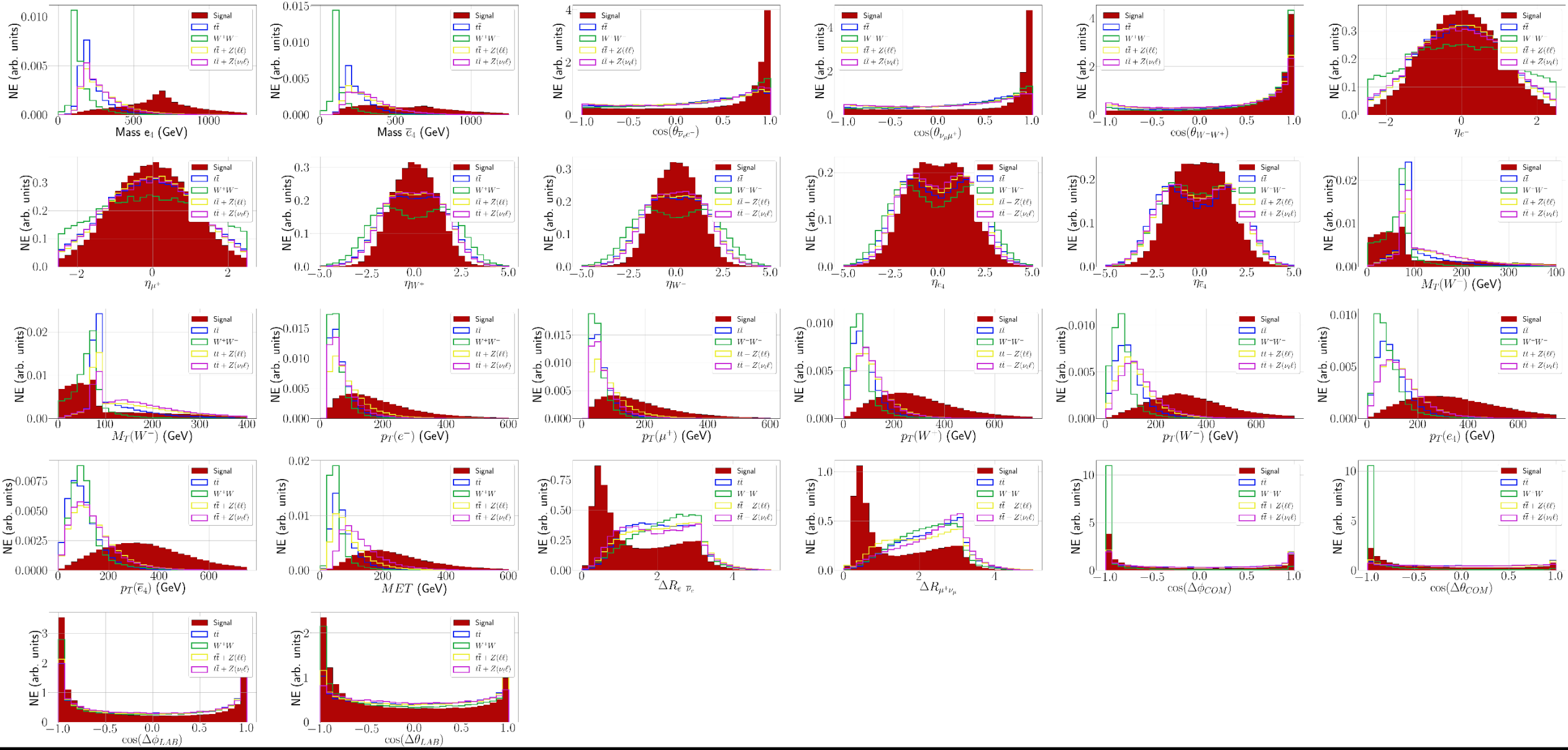
With cross sections,

	ZA topologies	VBF topologies	VLBSM topologies
$\boldsymbol{\sigma}$ (pb)	$4.40 \times 10^{-7} \pm 2.62 \times 10^{-10}$	$8.96 \times 10^{-7} \pm 5.88 \times 10^{-10}$	$7.70 \times 10^{-5} \pm 4.33 \times 10^{-8}$

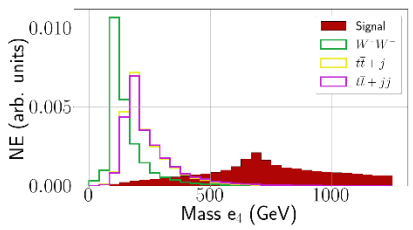
	$t\bar{t}$	$t\bar{t}+j$	$t\bar{t}+jj$	$t\bar{t}+Z^0(\ell^+\ell^-)$	$t\bar{t}+Z^0(\bar{\nu}_\ell\ell)$
$\boldsymbol{\sigma}$ (pb)	$6.72 \pm 3.01 \times 10^{-3}$	$7.85 \pm 5.06 \times 10^{-3}$	$5.99 \pm 3.70 \times 10^{-3}$	$5.36 \times 10^{-4} \pm 3.81 \times 10^{-7}$	$1.06 \times 10^{-3} \pm 6.95 \times 10^{-7}$

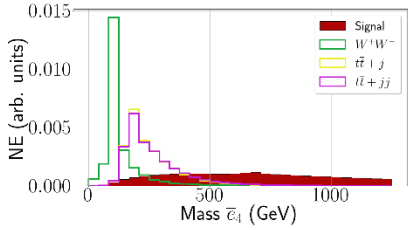
	W^+W^-	$pp \to \ell\nu_\ell$	$pp \to \ell\nu_\ell + j$	$pp \to \ell\nu_\ell + jj$
$\boldsymbol{\sigma}$ (pb)	$0.839 \pm 5.45 \times 10^{-4}$	10309.1 ± 5.4	2943.6 ± 2.1	1233.2 ± 0.7

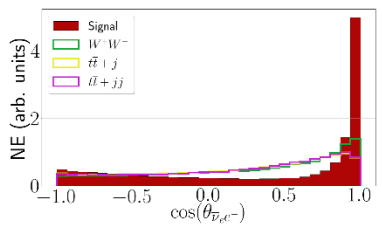
Kinematic variables for ZA topologies with the ATLAS detector

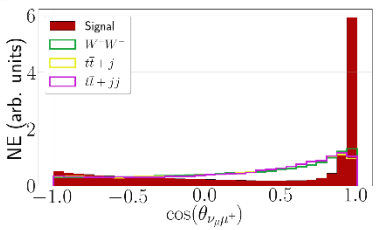


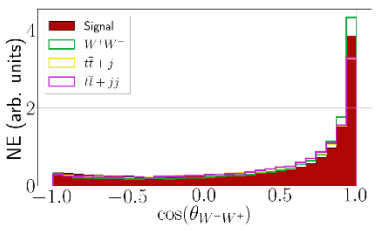
Kinematic variables for VBF topologies with the ATLAS detector

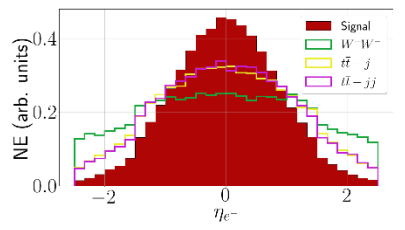


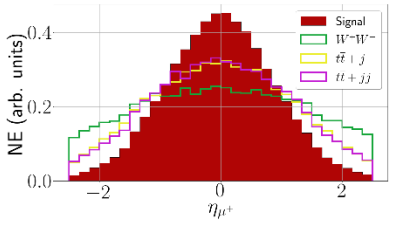


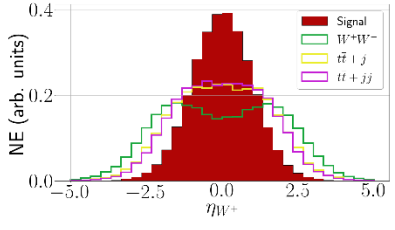


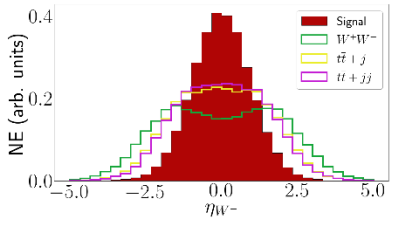


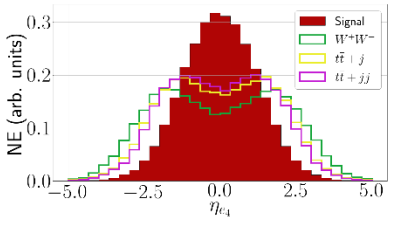


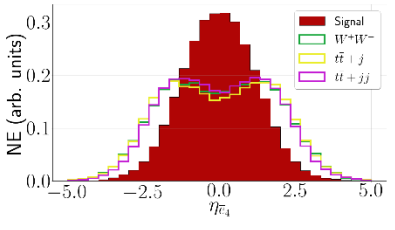


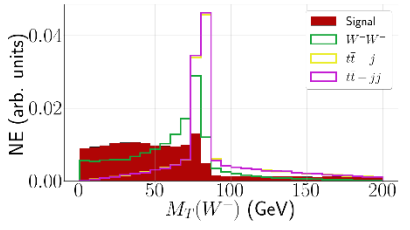


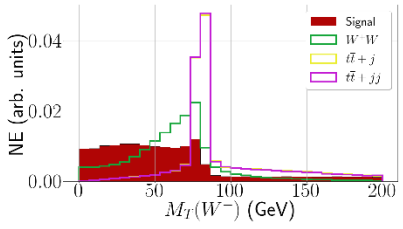


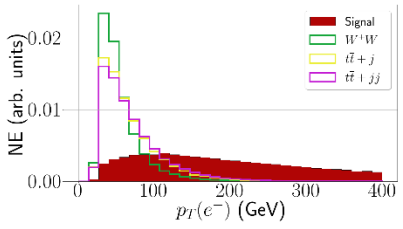


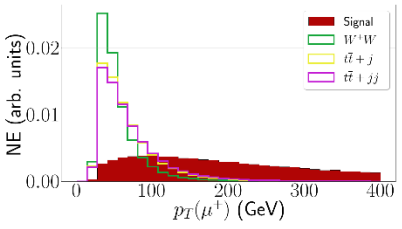


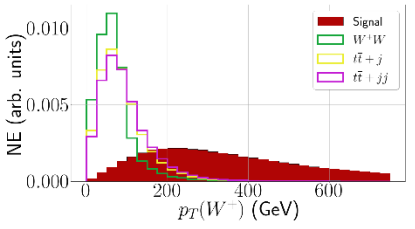


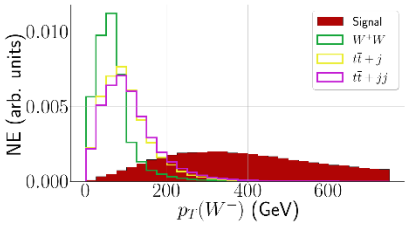


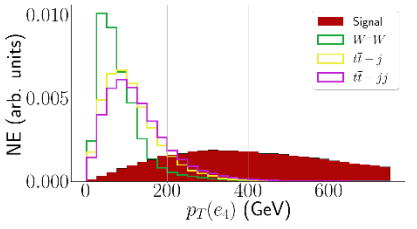


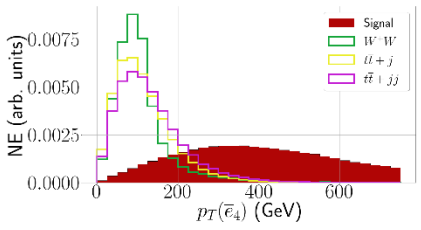


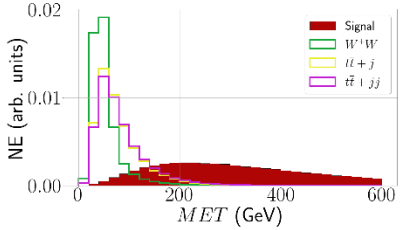


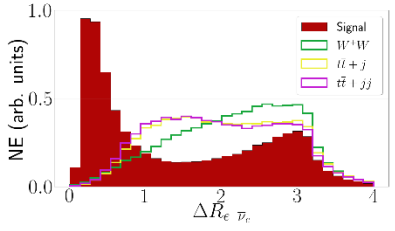


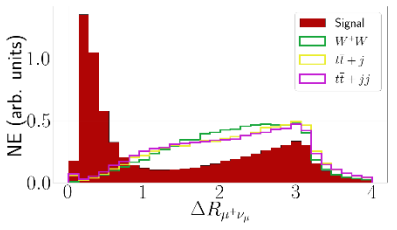


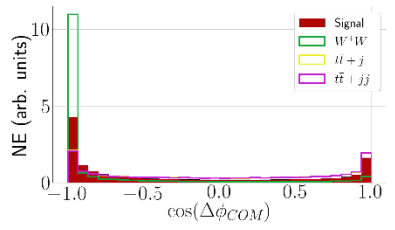


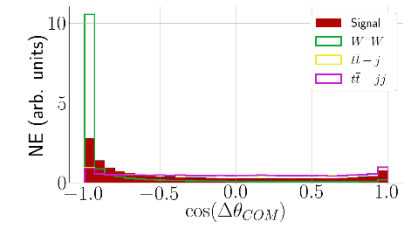


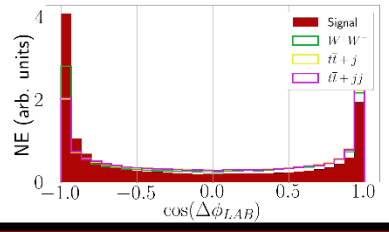


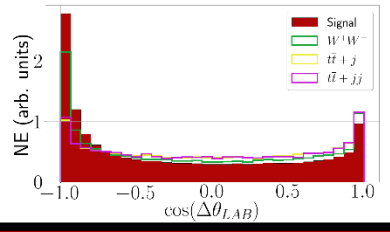




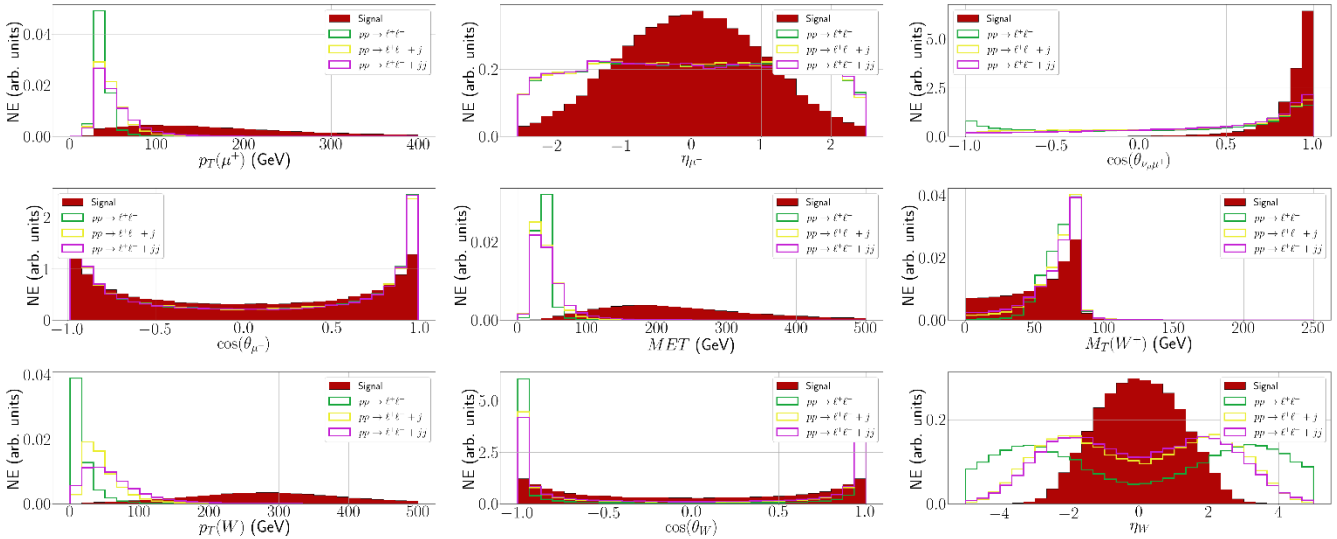




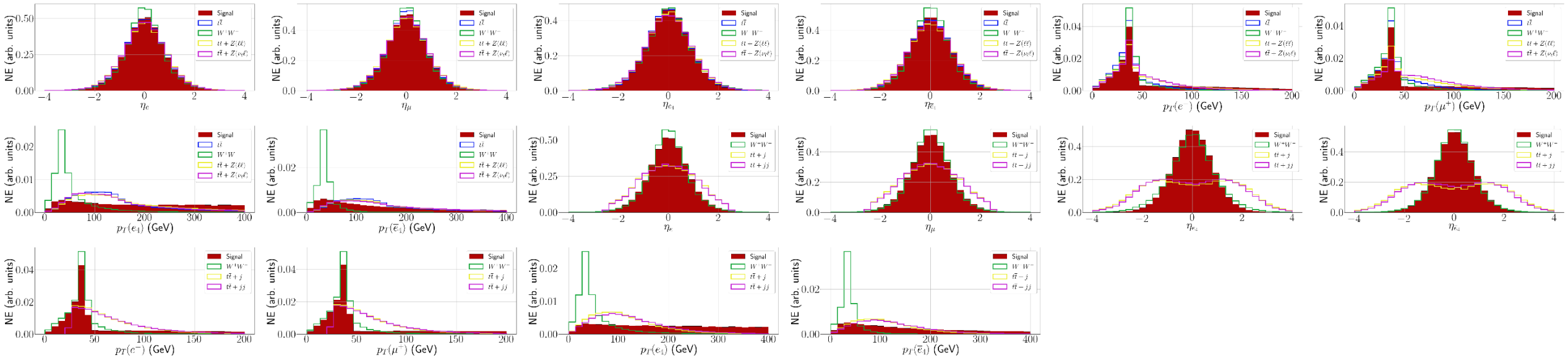


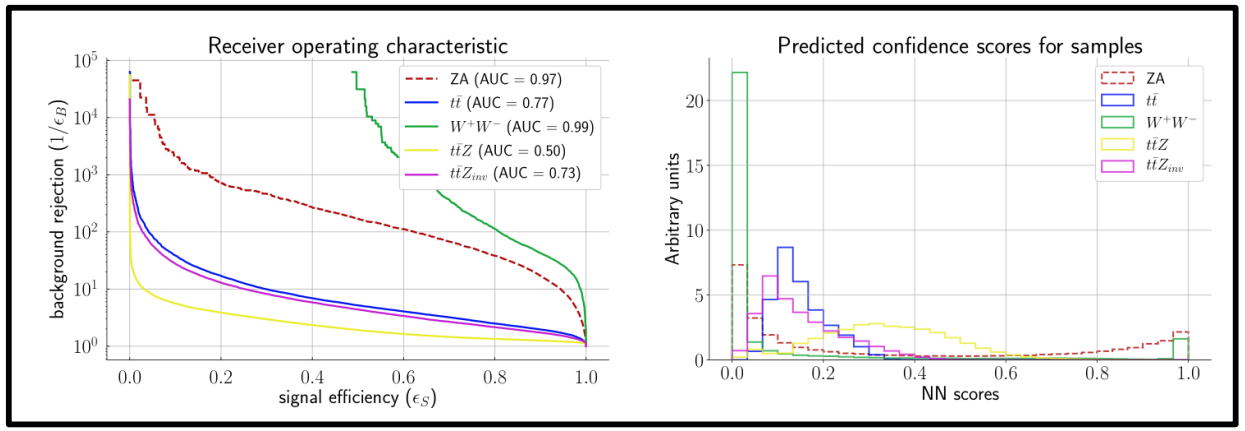


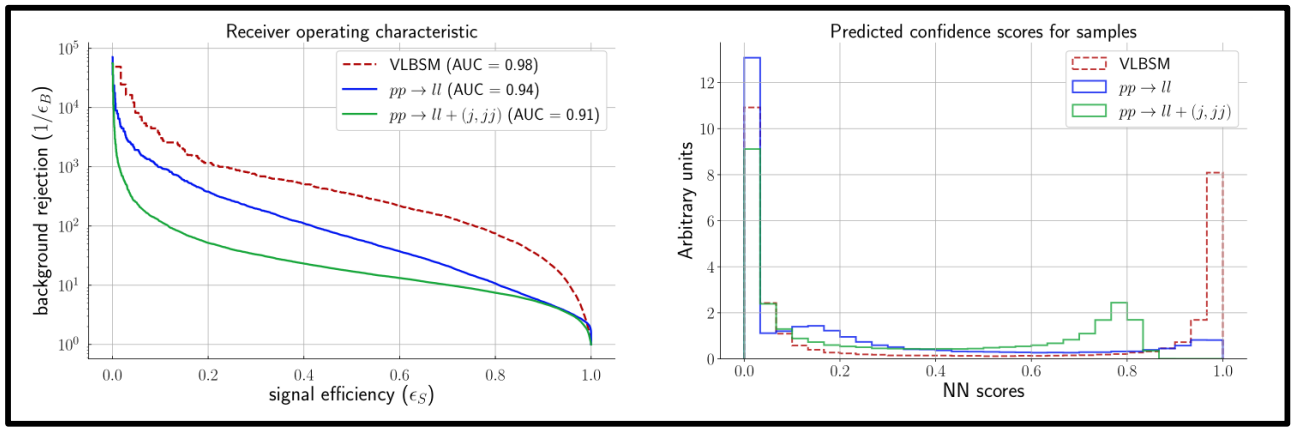
Kinematic variables for VLBSM topologies with the ATLAS detector

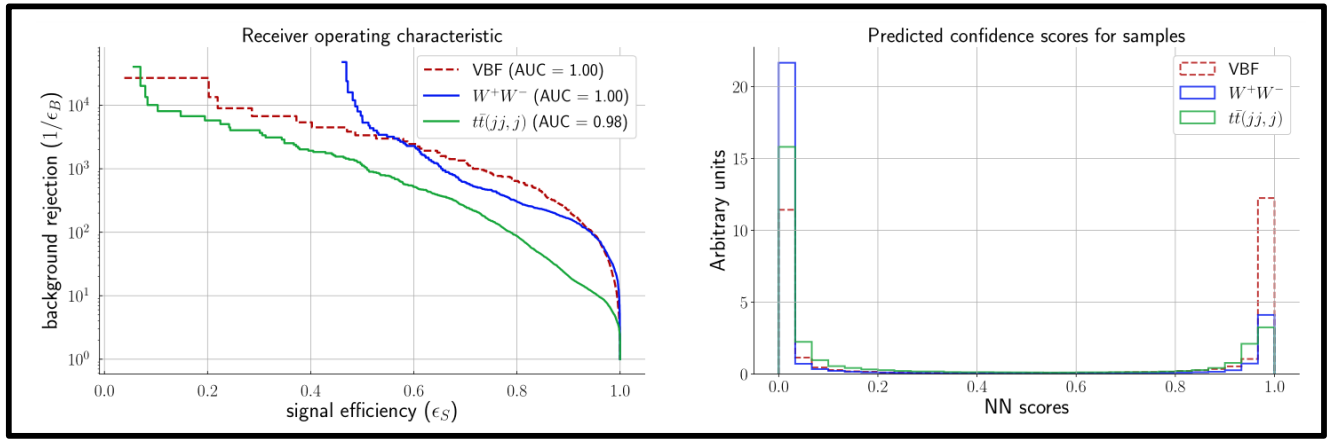


Boosted variables for VBF and ZA topologies with the ATLAS detector

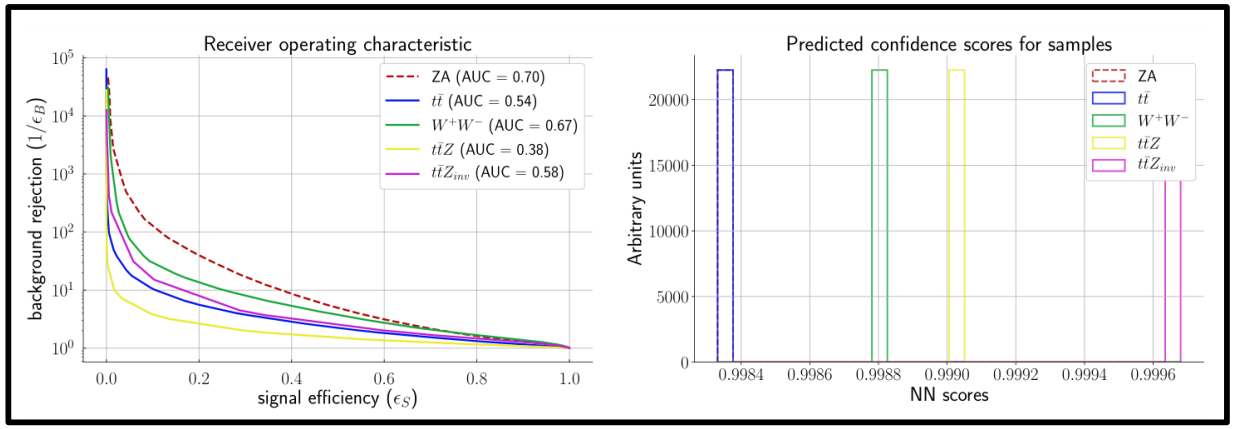


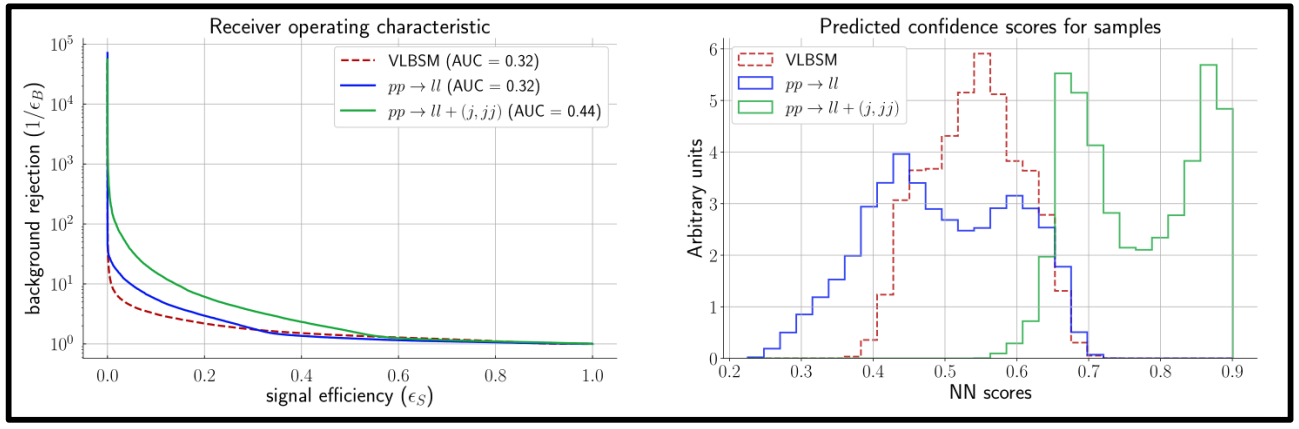


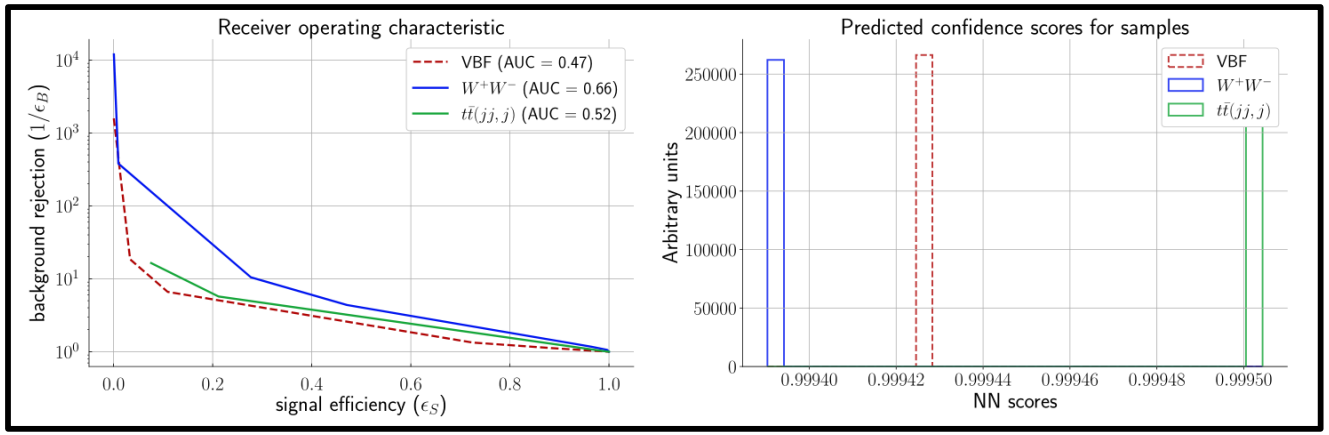




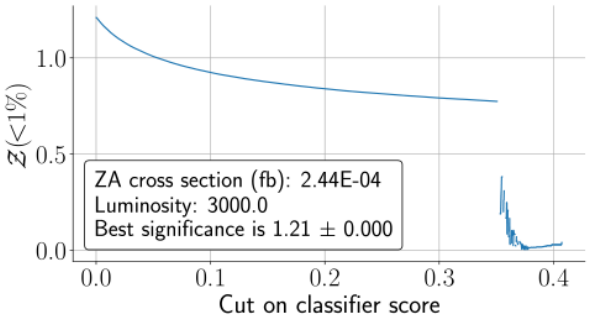
ROC and Predicted confidence scores following evolutive algorithm with the accuracy metric and a beam luminosity of 3000 fb-1.

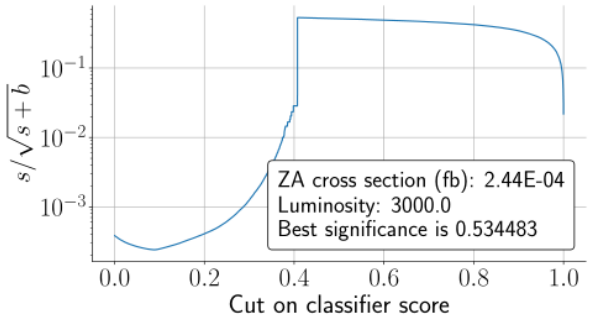


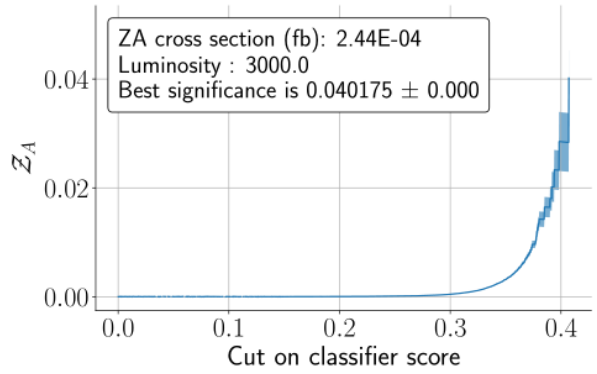


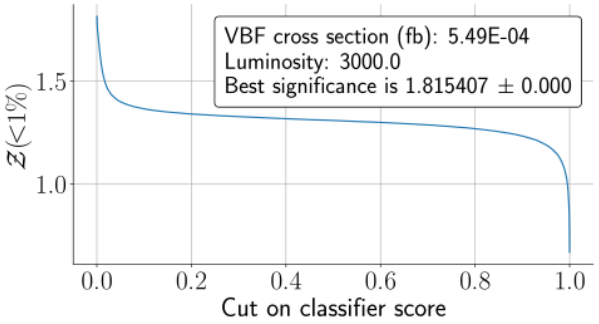


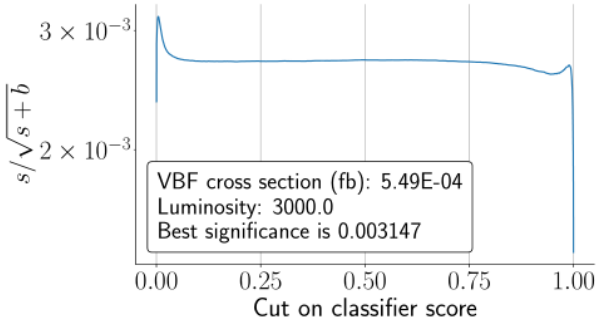
ROC and Predicted confidence scores following evolutive algorithm with the Asimov metric and a beam luminosity of 3000 fb^{-1}.

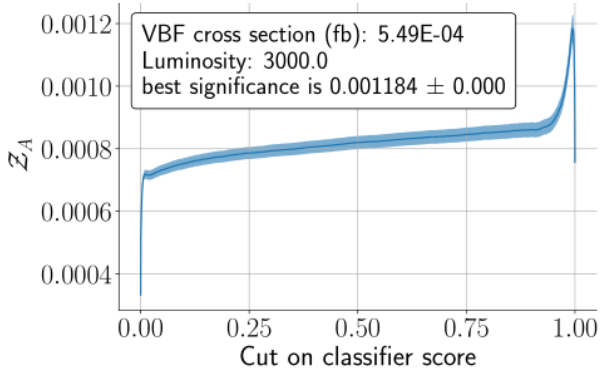


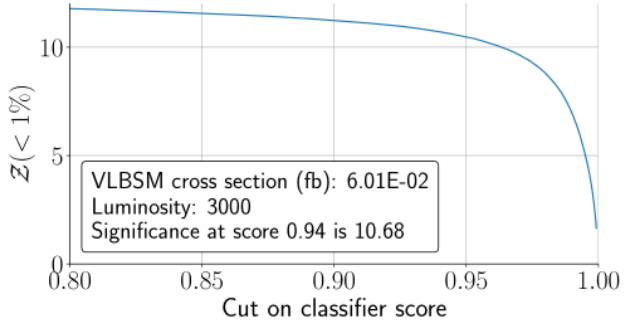


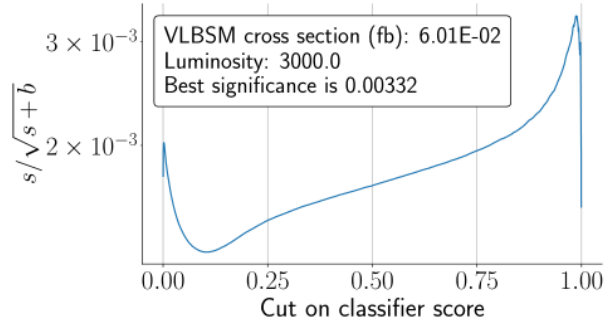


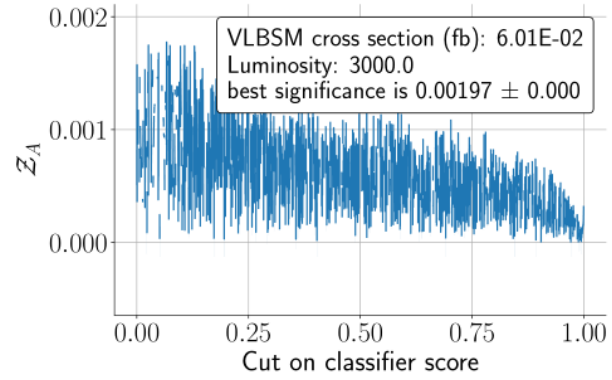








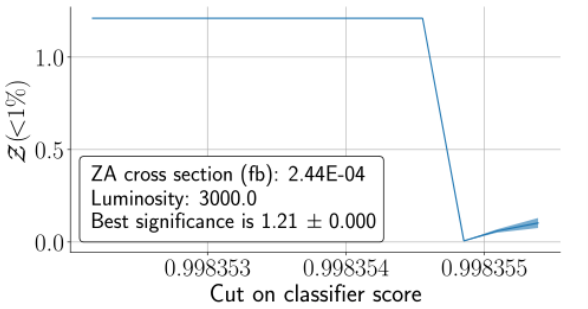


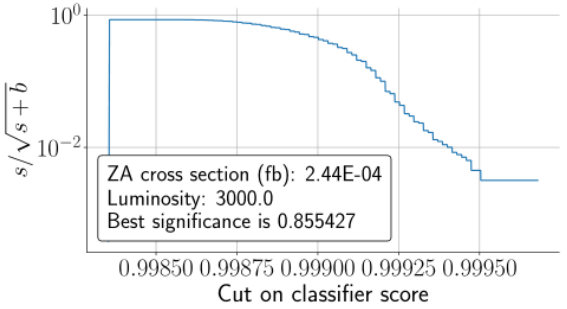


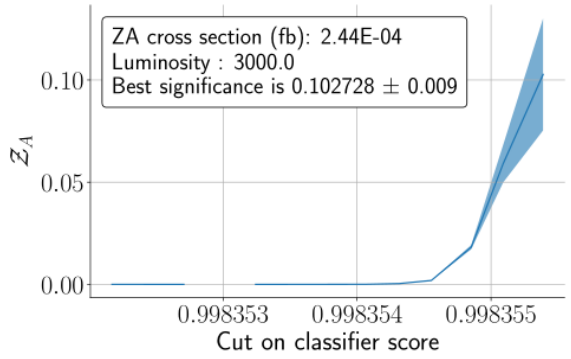
Z_A: $\sigma_C = 0.04\ \sigma$

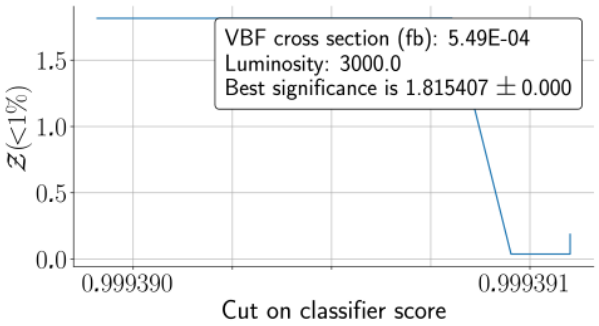
$Z(<1\%)$: $\sigma_C = 13.71\ \sigma$

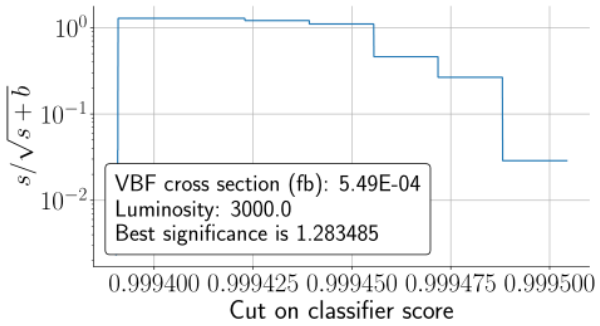
$s/\sqrt{s+b}$: $\sigma_C = 0.55\ \sigma$

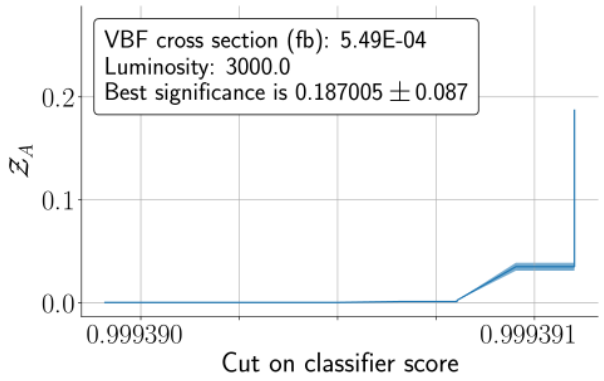


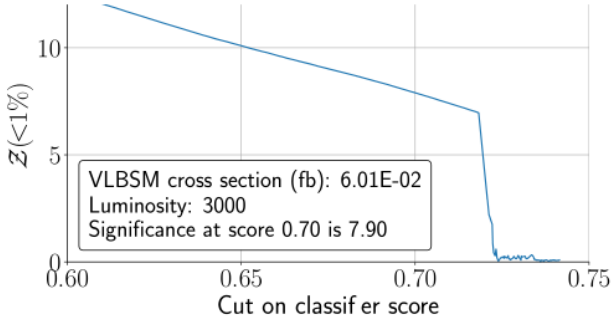


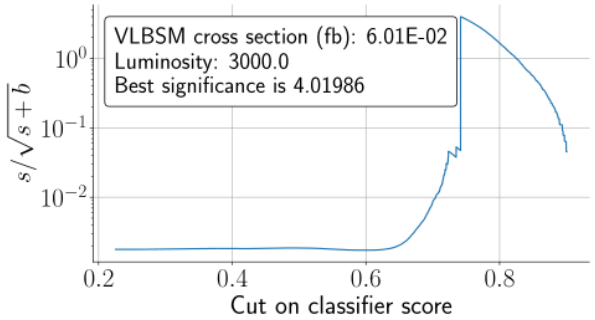


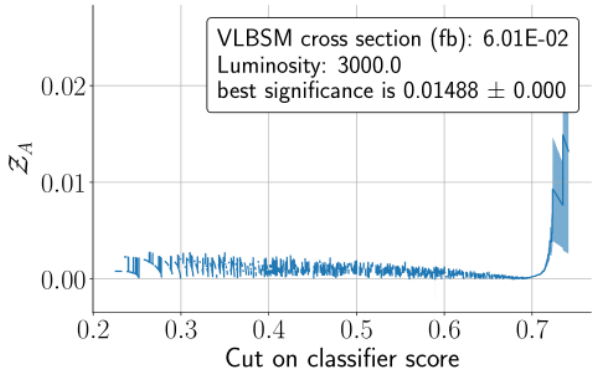










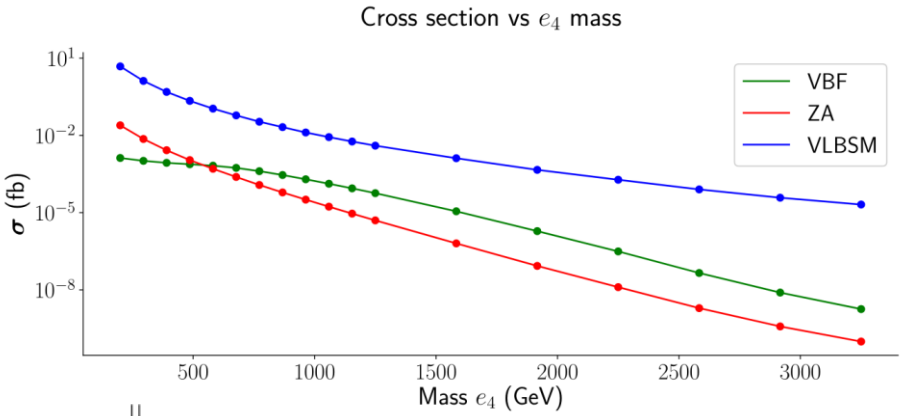


Z_A: $\sigma_C = 0.33\ \sigma$

$Z(<1\%)$: $\sigma_C = 10.93\ \sigma$

$s/\sqrt{s+b}$: $\sigma_C = 6.16\ \sigma$

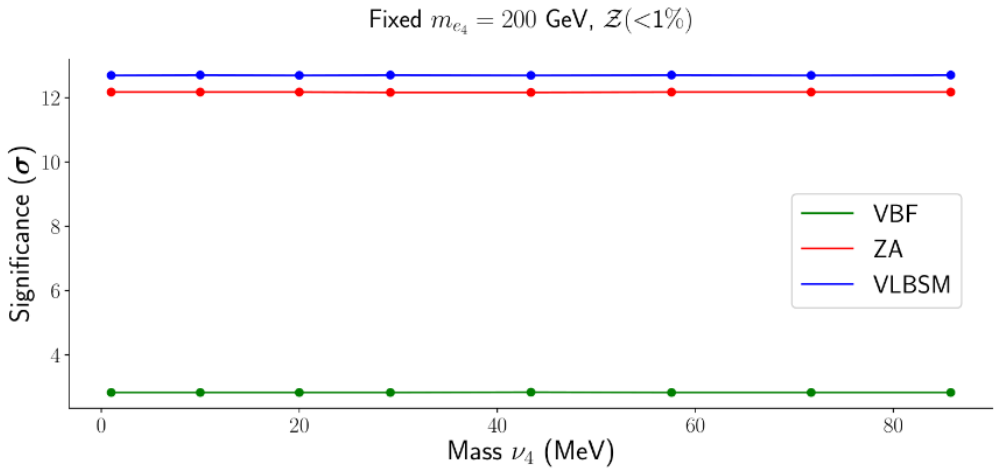
Mass of e_4	$s/\sqrt{s+b}$			$\mathcal{Z}(<1\%)$			$\mathcal{Z}_A$		
	ZA	VBF	VLBSM	ZA	VBF	VLBSM	ZA	VBF	VLBSM
200 GeV	6.10	2.00	12.65	12.18	2.83	12.70	4.44	0.145	4.28
486 GeV	1.77	1.50	11.26	2.60	2.13	8.62	0.30	0.53	0.20
677 GeV	0.86	1.28	4.02	1.21	1.82	7.90	0.11	0.187	0.015

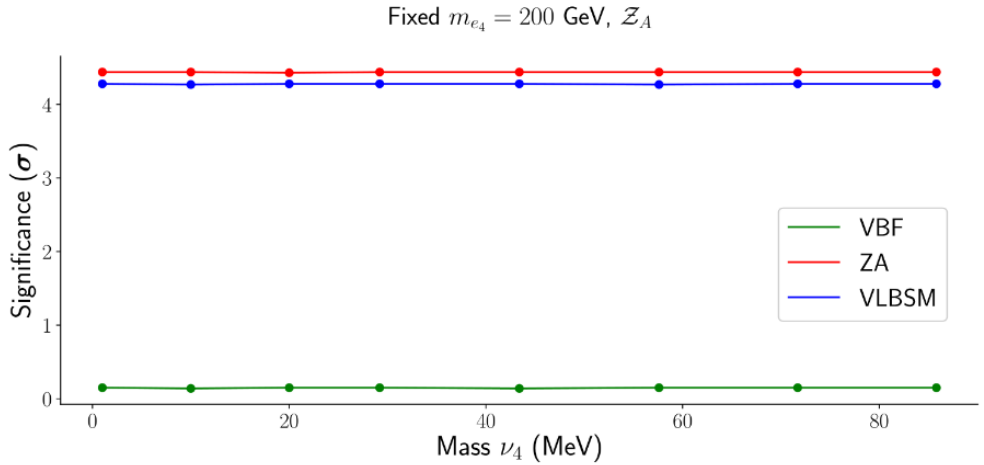


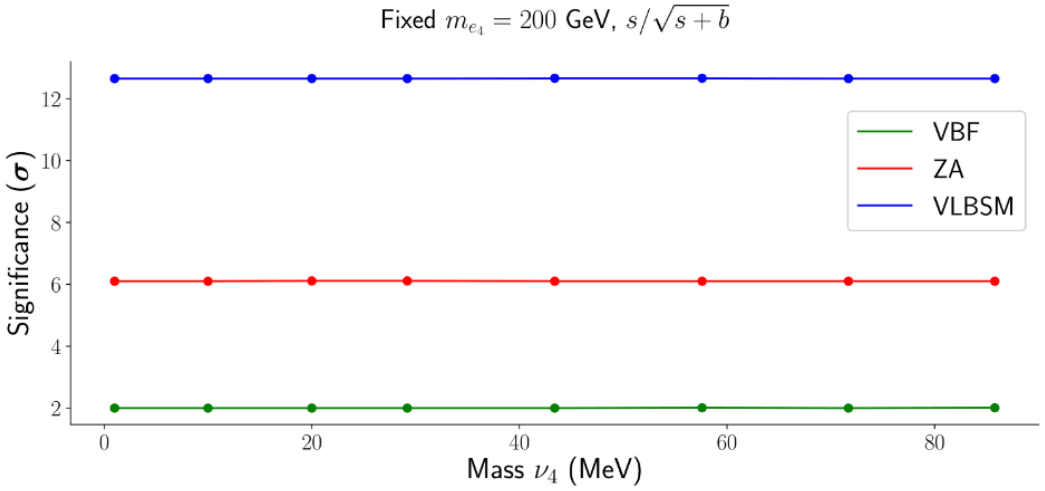
Mass of e_4	$s/\sqrt{s+b}$			$\mathcal{Z}(<1\%)$			$\mathcal{Z}_A$		
	ZA	VBF	VLBSM	ZA	VBF	VLBSM	ZA	VBF	VLBSM
200 GeV	4.01	9.4×10^{-3}	0.31	12.18	2.83	12.95	2.05	2.47×10^{-3}	1.41
486 GeV	0.95	1.51	6.66×10^{-3}	2.59	2.13	7.83	0.12	4.6×10^{-4}	2.15×10^{-4}
677 GeV	0.53	3.15×10^{-3}	3.32×10^{-3}	1.21	1.82	10.68	0.040	1.18×10^{-3}	1.97×10^{-3}
868 GeV	0.26	0.93	6.18×10^{-4}	0.52	1.32	6.60	0.01	0.30	2.47×10^{-4}
1250 GeV	0.05	4.37×10^{-4}	1.20×10^{-4}	0.17	0.59	4.90	4.28×10^{-4}	2.05×10^{-4}	2.65×10^{-3}

Heavy states (m > 1 TeV) have a reduced significance. We do note, however a combined Z(<1%) = 5.66 σ. Signal with jets may improve this, due higher branching ratios.

For light states, higher significances are obtained. A light VLL should have a strong presence in a high luminosity run at the LHC.

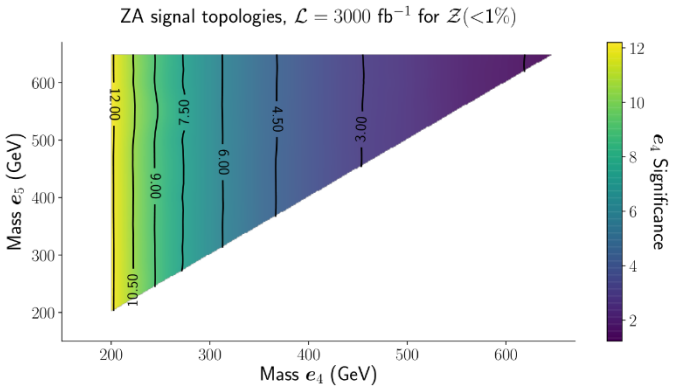


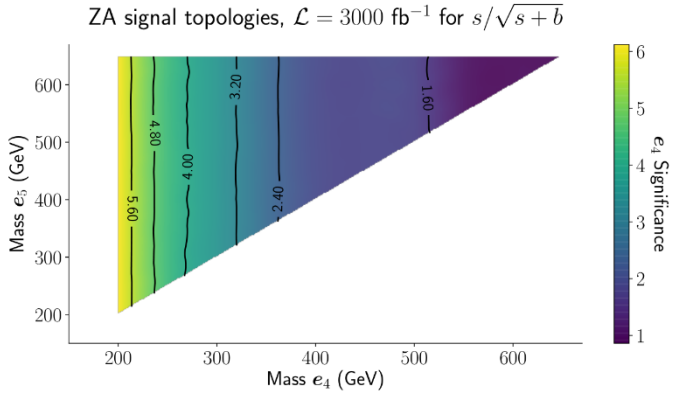


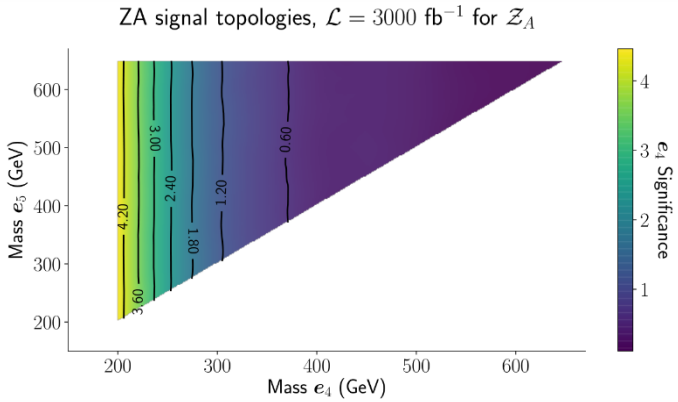


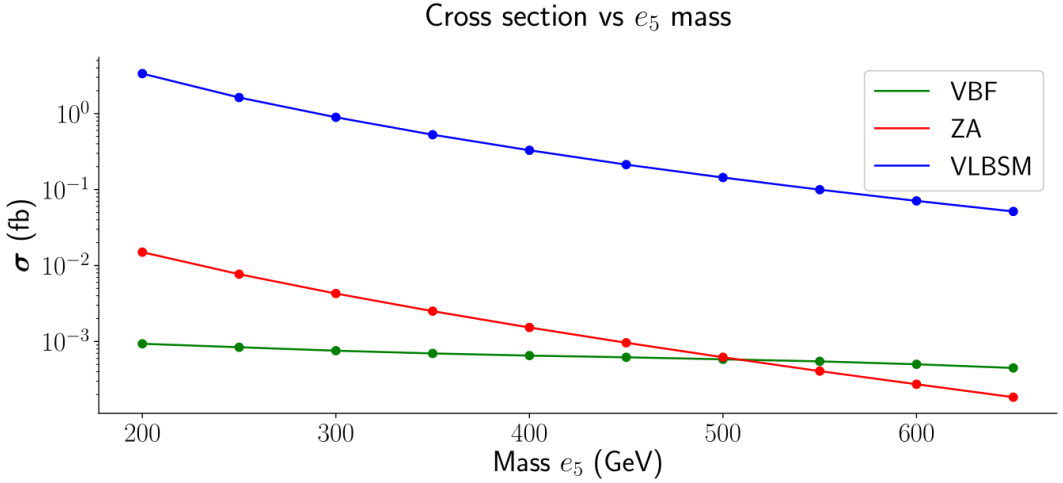
For completeness of information, we show the effects of the BSM light neutrino.

In particular, we note that no significant variations in the significance in the [0.1 100] MeV range









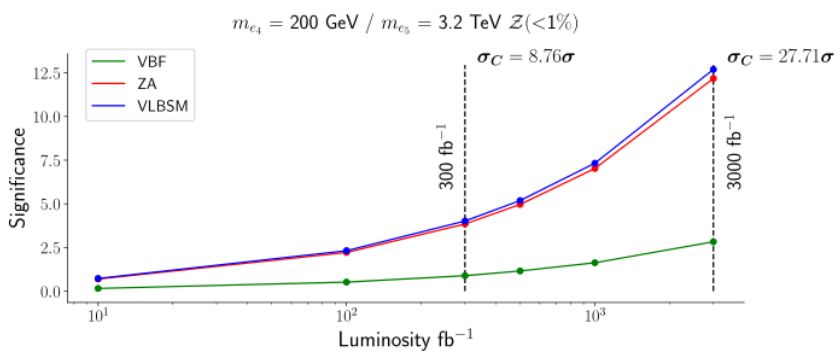
	$m_{e_4} = 200$ GeV $m_{e_5} = 200$ GeV	$m_{e_4} = 200$ GeV $m_{e_5} = 250$ GeV	$m_{e_4} = 200$ GeV $m_{e_5} = 300$ GeV
$e_5 \to W\nu_4$	0.5082513	0.5069336	0.5063337
$e_5 \to W\nu_5$	0.3111182	0.3119451	0.3123246
$e_5 \to W\nu_6$	0.1806305	0.1811213	0.1813417

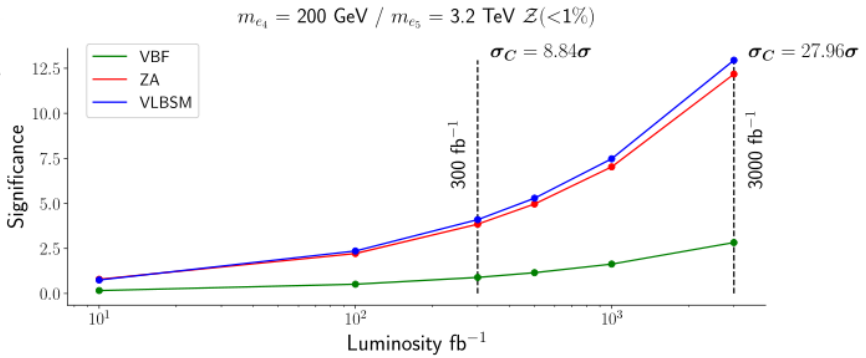
The lightest VLL (e_4) significance does not have a noticeable impact from a varying e_5. Small variations in decay branching ratios of e_5 supports the independence of e_5.

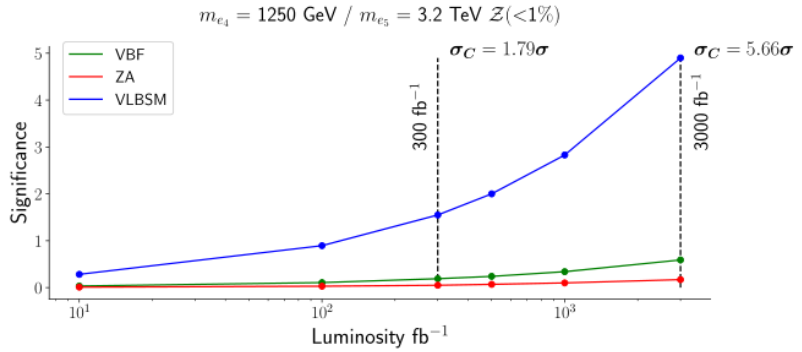
Asimov metric

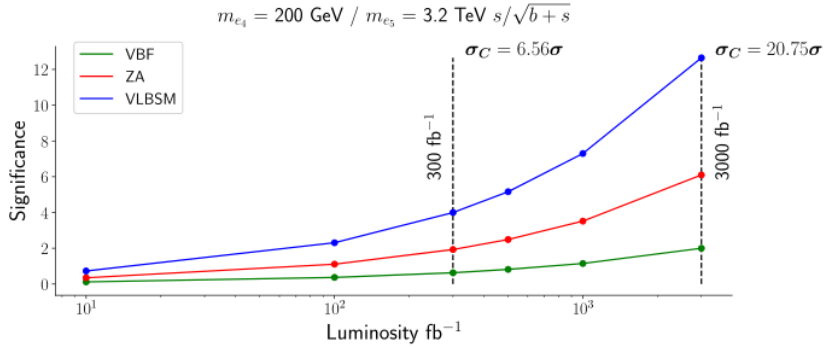
AUC metric

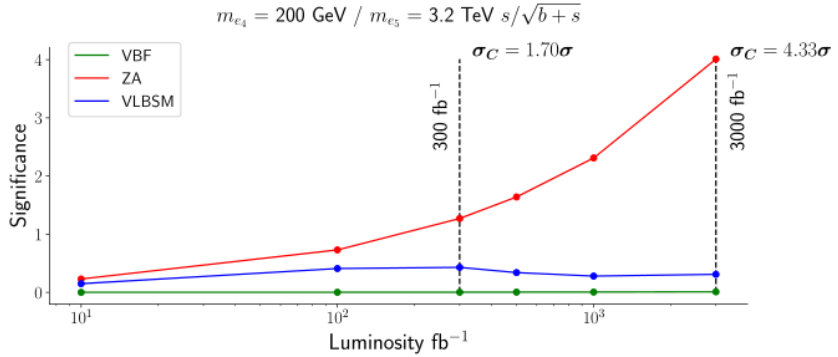
AUC metric

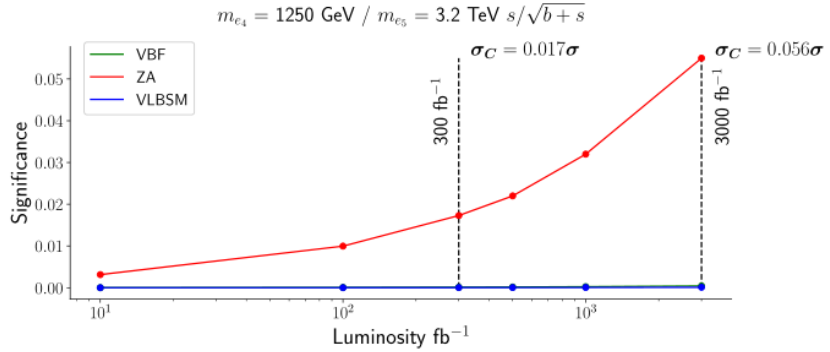


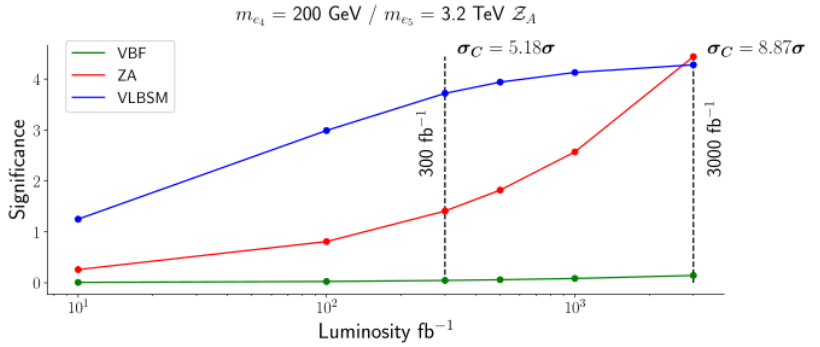


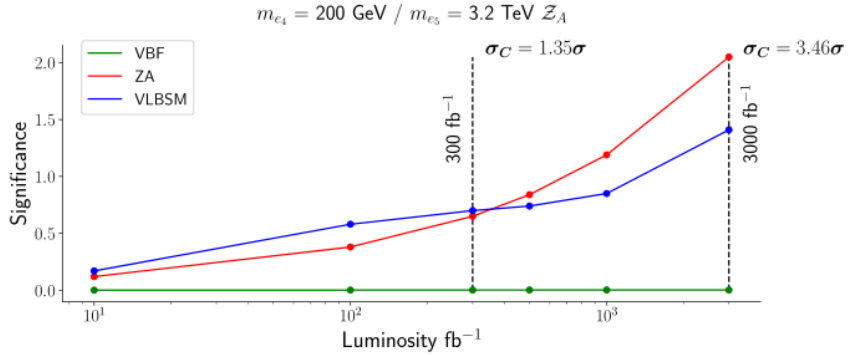


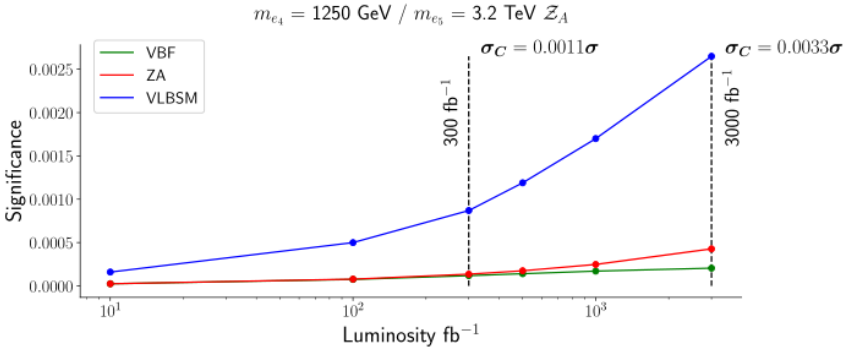












Conclusions

In this work we have studied the phenomenology of VLLs at the LHC, relevant for run III and beyond. Using deep learning techniques we obtain the statistical significance.

1. For a light a VLL, 200 GeV, and L = 3000 fb^{-1} the combined significances:

Asimov metric:	AUC metric:
Z_A: 8.57σ	Z_A: 3.46σ
Z(<1%): 27.71σ	Z(<1%): 27.96σ
s/√s+b: 20.75σ	s/√s+b: 4.33σ

2. For this same scenario, a 5σ significance at run III luminosities (Monica 2019), 300 fb^{-1}, is obtained in the Asimov metric (for all statistics) and the AUC metric (Z(<1%)).
3. For heavy VLL states, a 5σ significance is only achievable at high luminosities.
4. Higgs and flavour studies are also interesting. The presence of a sterile neutrino also provides a potential dark matter candidate.

References

- S. Chatrchyan et al. Physics Letters B 716.1 (2012), pp. 30–61
- F. Abe et al. Phys. Rev. Lett. 74 (1995), pp. 2626–2631
- José E. Camargo-Molina et al., Phys. Rev. D 99 (2019), 035041
- José E. Camargo-Molina et al., Phys. Rev. D 95 (2017), 075031
- António P. Morais, et al., arXiv (2020), 2001.06383
- Monica D'Onofrio. https://indico.cern.ch/event/855882/contributions/3601848/attachments/1930307/3196895/Prospects_Run3_HLLHC_Fermilab.pdf.

Thank you for your attention

universidade de aveiro

FCT Fundação para a Ciência e a Tecnologia

CENTRO 2020

PORTUGAL 2020

UNIÃO EUROPEIA

Fundos Europeus Estruturais e de Investimento

Felipe Freitas[1], **J. P. Pino[1]**, António P. Morais[1], Roman Pasechnik[2]

Workshop on Compact Objects, Gravitational Waves and Deep Learning

[1]Departamento de Física, Universidade de Aveiro e CIDMA [2]Department of Astronomy and Theoretical Physics, Lund University

Auxiliar Slides

Felipe Freitas[1], **J. P. Pino[1]**, António P. Morais[1], Roman Pasechnik[2]

Workshop on Compact Objects, Gravitational Waves and Deep Learning

[1]Departamento de Física, Universidade de Aveiro e CIDMA [2]Department of Astronomy and Theoretical Physics, Lund University

Field	$\mathbf{SU(3)_C}$	$\mathbf{SU(2)_L}$	$\mathbf{U(1)_Y}$	# of generations
A^a	**1**	**2**	0	1
B	**1**	**1**	0	1
G	**8**	**1**	0	1
ϕ	**1**	**2**	1	1

Field	$\mathbf{SU(3)_C}$	$\mathbf{SU(2)_L}$	$\mathbf{U(1)_Y}$	# of generations
Q	**3**	**2**	$1/3$	3
u_R	**3**	**1**	$4/3$	3
d_R	**3**	**1**	$-2/3$	3
L	**1**	**2**	-1	3
e_R	**1**	**1**	-2	3

$$Q^i = \begin{bmatrix} u_L \\ d_L \end{bmatrix}^i \quad L^i = \begin{bmatrix} \nu_L \\ e_L \end{bmatrix}^i$$

The invariant Lagrangian density,

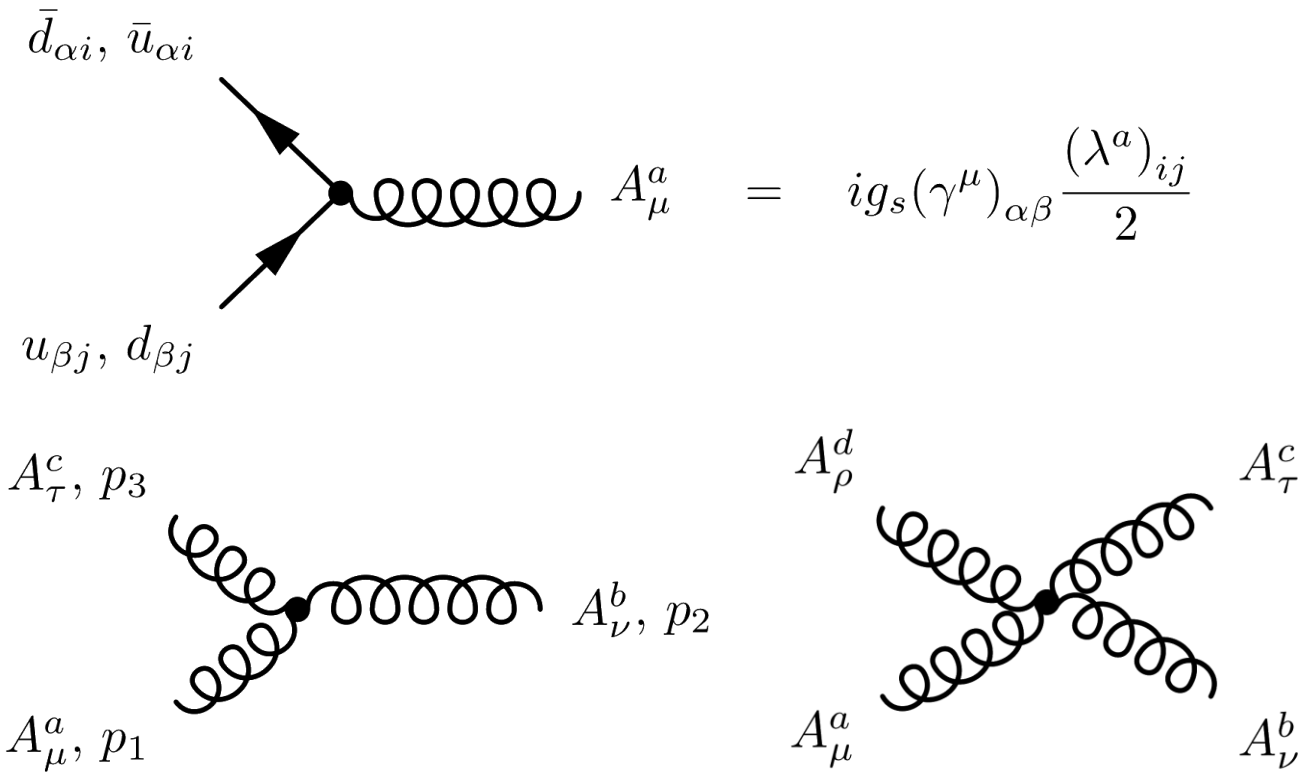
$$\mathcal{L}_{\mathrm{QCD}} = i\left(\bar{Q}\right)^{ij} \not{D}^{k}{}_{i} Q_{kj} + i(\bar{u}_{\mathrm{R}})^{ij} \not{D}^{k}{}_{i} (u_{\mathrm{R}})_{kj} + i\left(\bar{d}_{\mathrm{R}}\right)^{ij} \not{D}^{k}{}_{i} (d_{\mathrm{R}})_{kj} - \frac{1}{4} G^{a}_{\mu\nu} G_{a}^{\mu\nu},$$

with,

$$G^{a}_{\mu\nu} = \partial_\mu A^{a}_{\nu} - \partial_\nu A^{a}_{\mu} - g_s f^{a}{}_{bc} A^{b}_{\mu} A^{c}_{\nu},$$

$$D_\mu = \partial_\mu + i g_s \frac{\lambda_a}{2} A^{a}_{\mu},$$

$$[\lambda_a, \lambda_b] = 2i f_{ab}{}^{c} \lambda_c$$



At high energies, due to asymptotic freedom, QCD can be dealt with great accuracy, with lattice QCD being an example of this. At low energies, the coupling is strong and quarks become confined. Processes that result in quarks (or gluons) as final states lead to formation of jets of particles.

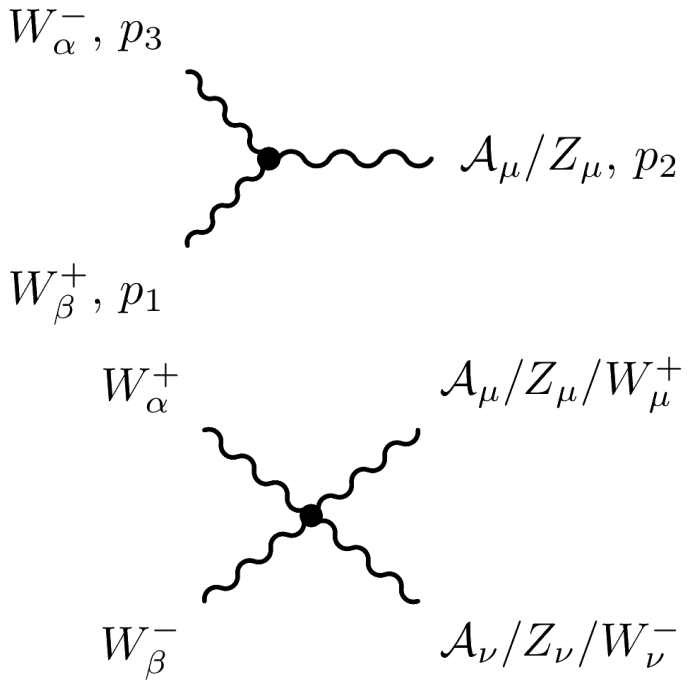
The invariant Lagrangian density,

$$\mathcal{L}_{EW} = -\frac{1}{4}B^{\mu\nu}B_{\mu\nu} - \frac{1}{4}F_a^{\mu\nu}F^a_{\mu\nu} + i\bar{\psi}^j \not{D}\psi_j + (D_\mu\phi)^\dagger(D^\mu\phi) - V\left(\phi^\dagger,\phi\right)$$

with,

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \qquad F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - g'\epsilon^a{}_{bc}A^b_\mu A^c_\nu,$$

$$D_\mu = \partial_\mu + ig'\frac{Y}{2}B_\mu + ig\frac{\sigma_a}{2}A^a_\mu, \qquad V\left(\phi^\dagger,\phi\right) = \mu^2\phi^\dagger\phi + \lambda\left(\phi^\dagger\phi\right)^2,$$



Via the Higgs mechanism, where the scalar gains a VEV, $v^2 = -\mu^2/2\lambda$, mass is generated for the electroweak bosons and the Higgs field

$$m_Z = \frac{v}{\sqrt{2}}\sqrt{g^2+g'^2} \qquad m_{\mathcal{A}} = 0$$

$$m_{W^\pm} = \frac{gv}{\sqrt{2}} \qquad m_h = \sqrt{-2\mu^2}$$

where,

$$\begin{cases} \mathcal{A}_\mu = \dfrac{1}{\sqrt{g^2+g'^2}}\left(g'A^3_\mu + gB_\mu\right) \\ Z_\mu = \dfrac{1}{\sqrt{g^2+g'^2}}\left(gA^3_\mu - g'B_\mu\right) \end{cases} \qquad \begin{cases} W^-_\mu = \dfrac{1}{\sqrt{2}}\left(A^1_\mu + iA^2_\mu\right) \\ W^+_\mu = \dfrac{1}{\sqrt{2}}\left(A^1_\mu - iA^2_\mu\right) \end{cases}$$

Due to gauge invariance, mass terms are not allowed. However, couplings to the scalar field are authorised

$$\mathcal{L}_{Y,l} = (y_l)_{ij}\phi\cdot\bar{L}^i e^j_R + \text{H.c.}$$
$$\mathcal{L}_{Y,q} = (y_d)_{ij}\phi\cdot\bar{Q}^i d^j_R + (y_u)_{ij}\tilde{\phi}\cdot\bar{Q}^i u^j_R + \text{H.c.}$$

$$\xrightarrow{\langle\phi\rangle\neq 0}$$

$$\mathcal{L}_{l,SB} = (y_l)_{ij}\frac{v}{\sqrt{2}}\bar{e}^i_L e^j_R + (y_l)_{ij}\frac{1}{\sqrt{2}}h\bar{e}^i_L e^j_R + \text{H.c.}$$
$$\mathcal{L}_{q,SB} = (y_d)_{ij}\frac{v}{\sqrt{2}}\bar{d}^i_L d^j_R + (y_u)_{ij}\frac{v}{\sqrt{2}}\bar{u}^i_L u^j_R + \text{H.c.}$$

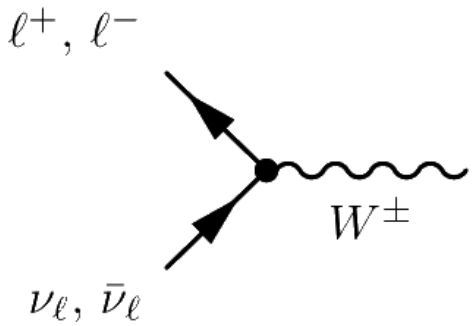
All quarks and charged leptons gain mass, leaving only the neutrinos as massless.
For the quark sector, a generic non diagonal Yukawa matrix is considered, which in turn results in quark mixing in the mass basis.

$$(y_u)_{ij} = (\mathcal{U}^*_L)_{ik}\left(y^D_u\right)^{kk}(\mathcal{U}_R)_{kj},$$
$$(y_d)_{ij} = (\mathcal{D}^*_L)_{ik}\left(y^D_d\right)^{kk}(\mathcal{D}_R)_{kj},$$

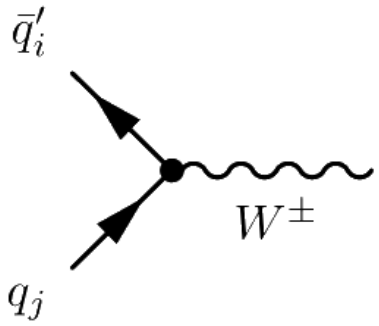
$$\xrightarrow{\text{Mass basis}}$$

$$\left(\bar{d}'_R\right)_k = (\mathcal{D}_R)_{kj}d^j_R, \qquad \left(\bar{d}'_L\right)_k = (\mathcal{D}^*_L)_{kj}d^j_L,$$
$$\left(\bar{u}'_R\right)_k = (\mathcal{U}_R)_{kj}u^j_R, \qquad \left(\bar{u}'_L\right)_k = (\mathcal{U}^*_L)_{kj}u^j_L,$$

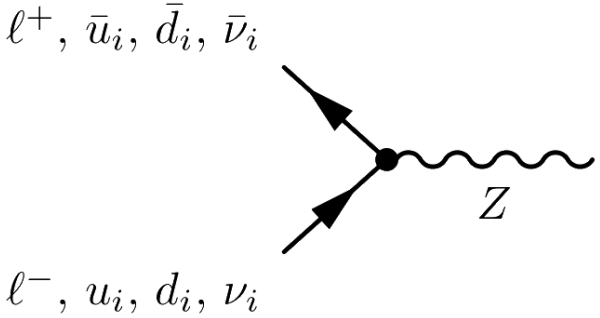
This has direct implications in observable physics, since it implies that quarks mix. In fact, while interactions via Z boson and photon remain diagonal both in the mass and interacting basis, for charged currents with W, they do not



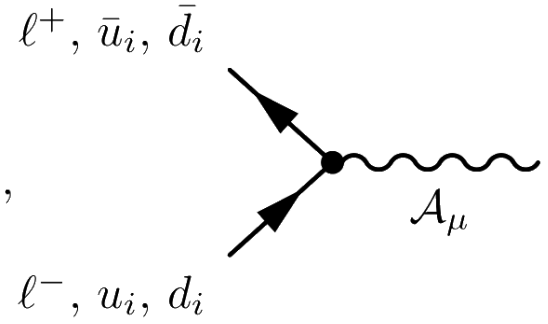
$$= -i\frac{g}{\sqrt{2}}\gamma^\mu P_{\mathrm{L}},$$



$$= -i\frac{g}{\sqrt{2}}(V_{\mathrm{CKM}})_{ij}\gamma^\mu P_{\mathrm{L}},$$



$$= -i\frac{g}{\cos(\theta_w)}\gamma^\mu\left(g_V^i - g_A^i\gamma^5\right),$$



$$= -ieQ_i\gamma^\mu$$

The mixing effects are encoded in the Cabibbo–Kobayashi–Maskawa matrix $V_{\mathrm{CKM}} = \mathcal{U}_L\mathcal{D}_L^\dagger$

The lack of right-handed neutrinos means no mass terms can be created, even with higher order corrections. Assuming a potential existence of right-handed neutrinos and only lepton generation, the Lagragian density post EWSB

$$\mathcal{L}_N = \frac{y_\nu v}{\sqrt{2}}\bar{\nu}_e N_R + M_N N_R N_R + \text{H.c.} \qquad \Longrightarrow \qquad M_\nu = \begin{bmatrix} 0 & \frac{y_\nu v}{\sqrt{2}} \\ \frac{y_\nu^* v}{\sqrt{2}} & M_N \end{bmatrix}$$

In the limit $M_N \gg v$ we get eigenvalues $m_\nu \approx \frac{1}{2}\frac{v^2 y_\nu^2}{M_N}$ and $m_N \approx M_N$. Considering now some numerical cases, with $M_N = 10^{16}$ GeV, $y_\nu = 1$ and $v = 246$ GeV we get

$m_\nu = 0.003$ eV, $\quad m_N = 10^{16}$ GeV

With the light neutrino in line with the experimental measured values for neutrinos. The generation of this mass came at a cost of a heavy right handed neutrino (in the GUT scale)

VLQs and SM-like quarks gain mass at tree-level,

	VLQs masses	SM-like quarks masses	
1st Gen.	$m^2_{S/D} = \frac{1}{2}(f^2+p^2)\mathcal{Y}_2^2$	$m_u = 0$	$m_d = 0$
2nd Gen.	$m^2_{S/D} = \frac{\omega^2(f^2+p^2+\omega^2)}{2(f^2+\omega^2)}\mathcal{Y}_2^2$	$m_c^2 = \frac{1}{2}\mathcal{Y}_2^2(u_1^2+u_2^2)$	$m_s^2 = \frac{1}{2}\mathcal{Y}_2^2\frac{d_2^2p^2}{(f^2+p^2+\omega^2)}$
3rd Gen.	$m_B^2 = \frac{1}{2}(f^2+\omega^2)\mathcal{Y}_1^2 + \frac{f^2p^2}{2(f^2+\omega^2)}\mathcal{Y}_2^2$	$m_t^2 = \frac{1}{2}\mathcal{Y}_1^2(u_1^2+u_2^2)$	$m_b^2 = \frac{1}{2}\mathcal{Y}_1^2 d_2^2$

Even with the most generic VEV setting, the first generation remains massless at tree-level,

$$\omega \sim f \sim p \longrightarrow \frac{\mathcal{Y}_1}{\mathcal{Y}_2} = \frac{m_{\mathrm{t}}}{m_{\mathrm{c}}} \approx \frac{m_{\mathrm{b}}}{m_{\mathrm{s}}} \approx \frac{m_{\mathrm{B}}}{m_{\mathrm{D,S}}} \sim \mathcal{O}(100)$$

We present now some physical viable scenarios for lepton sector, as it is the focus of analysis. After EWSB, the lepton mass matrix

$$[M_L] = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \kappa_7\omega & \kappa_5\omega \\ 0 & 0 & 0 & 0 & \kappa_6\omega & \kappa_8\omega \\ 0 & 0 & 0 & 0 & \kappa_1 p & \kappa_3 f \\ 0 & 0 & 0 & \kappa_2 p & 0 & 0 \\ 0 & 0 & 0 & \kappa_4 f & 0 & 0 \end{bmatrix} \xrightarrow{\omega \sim f \ll p} \begin{cases} m_{e_4} \approx \omega\sqrt{\kappa_5^2 + \kappa_8^2}. \\ m_{e_5} \approx p\kappa_1, \\ m_{e_6} \approx p\kappa_2 \end{cases}$$

Considering possible numerical cases

	Couplings	VEVs (TeV)	Masses (TeV)
e_4	$\kappa_{5,8} \sim \mathcal{O}(10^{-3} - 10^{-2})$	$p \sim \mathcal{O}(500 - 1000)$ $\omega \sim f \sim \mathcal{O}(1000)$	$\mathcal{O}(0.1 - 1)$
e_5	$\kappa_1 \sim \mathcal{O}(10^{-3.5} - 10^{-2})$		$\mathcal{O}(0.15 - 10)$
e_6	$\kappa_2 \sim \mathcal{O}(10^{-2})$		$\mathcal{O}(5 - 10)$

The neutrino sector is composed of 6 $\mathrm{SU}(2)_\mathrm{L}$ doublets and 6 $\mathrm{SU}(2)_\mathrm{L}$ singlets. Before EWSB, the neutrino matrix is block diagonal

$$M_\nu = \begin{bmatrix} \bar{M}_{9\times 9} & 0 \\ 0 & M_{6\times 6} \end{bmatrix}$$

For this case, we have eigenvalues $m^2_{\nu_{1,2,3}} = 0, \quad m^2_{\nu_{4,5}} = m^2_{e_6} \quad m^2_{\nu_{6,7}} = m^2_{e_5} \quad m^2_{\nu_{8,9}} = m^2_{e_4}$

Following EWSB, the matrix can be recast in a seesaw form

$$m_\nu = \begin{bmatrix} \mathbf{0}_{3\times 3} & \dfrac{v_\mathrm{EW}}{\sqrt{2}}(\boldsymbol{y}_\nu)_{3\times 12} \\ \dfrac{v_\mathrm{EW}}{\sqrt{2}}(\boldsymbol{y}_\nu^\mathrm{T})_{3\times 12} & (\boldsymbol{\mu}_N)_{12\times 12} \end{bmatrix}$$

This structure allows for 3 sub-eV neutrinos, as well as states in the keV-MeV range.

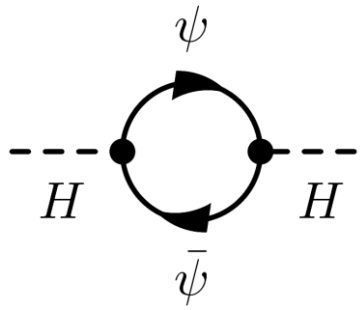
Interaction vertices for fermion are sensitive to elements of mixing matrices

$$U_L^e \cdot M_L \cdot U_R^{e\,\dagger} = m_e^{\rm diag},$$
$$U_\nu \cdot m_\nu \cdot U_\nu^\dagger = m_\nu^{\rm diag}.$$

	Lepton Mixings	Neutrino Mixings
SM-like sector	$U_\ell^{SM} = \mathbb{1}_{3\times3}$	$U_\nu = (U_{\rm PMNS})_{3\times3}$
BSM sector	$(U_{L,R}^{\rm VLL})_{3\times3} \sim \mathcal{O}(1)$	$U_{1,2} \sim \mathcal{O}(1)$ $D_{1,2} \sim \mathcal{O}(10^{-3} - 10^{-8})$

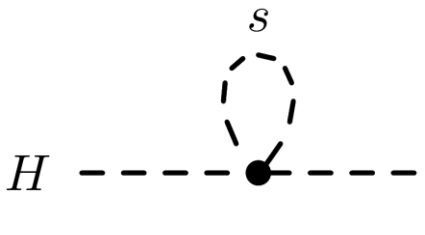
$D_{1,2}$ represents doublet/singlet mixing. $U_{1,2}$ represents doublet/doublet and singlet/singlet mixing.

One of the main problems in the SM is the hiearchy problem. Computation of corrections to the Higgs propagator at one loop order gives



$$\Longrightarrow \quad -i\Sigma(p^2=0) = -\frac{\lambda_\psi^2}{8\pi^2}\Lambda_{\mathrm{GUT}}^2 + \dots$$

Considering top physics, one sees that this contribution is 30 orders of magnitude higher than the experimental value for the Higgs mass (m $\sim$ 125 GeV). A possible fix is an introduction of a new scalar, s, with the same quantum numbers as the fermion, except spin



$$\Longrightarrow \quad -i\Sigma(p^2=0) = \frac{\lambda_s}{8\pi^2}\Lambda_{\mathrm{GUT}}^2 + \dots$$

If $\lambda_s = \lambda_\psi^2$ then the quadratic divergences cancel out, leaving only small logarithmic corrections. This one of the main arguments to SUSY.

Field	$\mathbf{SU(3)_C}$	$\mathbf{SU(2)_L}$	$\mathbf{U(1)_Y}$	# of generations
Q_L	**3**	**2**	$1/3$	3
L	**1**	**2**	-1	3
d_R	**3**	**1**	$-2/3$	3
u_R	**3**	**1**	$4/3$	3
e_R	**1**	**1**	-2	3

Table 3.1

Field	$\mathbf{SU(3)_C}$	$\mathbf{SU(2)_L}$	$\mathbf{U(1)_Y}$	# of generations
$E_{L,R}$	**1**	**2**	-1	3
$D_{L,R}$	**3**	**1**	$-2/3$	2
ν_R	**1**	**1**	0	6

Table 3.2

Field	$\mathbf{SU(3)_C}$	$\mathbf{SU(2)_L}$	$\mathbf{U(1)_Y}$	# of generations
ϕ	**1**	**2**	1	3

Table 3.3

With the doublets defined as $Q_L^i = \begin{bmatrix} u_L \\ d_L \end{bmatrix}^i$, $L^i = \begin{bmatrix} \nu_{e_L} \\ e_L \end{bmatrix}^i$ and $E_{L,R}^i = \begin{bmatrix} \nu'_{e_L,R} \\ e'_{L,R} \end{bmatrix}^i$

In Tab. 3.1, we showcase the SM-like fermions. In Tab. 3.2 we include the 3 new VLL generations denoted as $E_{L,R}$ and two light VLQ generations denoted as $D_{L,R}$. This last sector also offers a rich neutrino sector, with six of them originating from the $E_{L,R}$ doublets and six right-handed Majorana neutrinos, ν_R.

The Lagrangian density is determined by writing all renormalisable Lorentz and gauge invariant terms,

Bosonic sector: $\mathcal{L}_{\text{bos,f}} = -\frac{1}{4}B^{\mu\nu}B_{\mu\nu} - \frac{1}{4}A_b^{\mu\nu}A^b_{\mu\nu} - \frac{1}{4}G_c^{\mu\nu}G^c_{\mu\nu} + \frac{1}{2}(D_\mu\phi_a)(D^\mu\phi^a)^\dagger,$

Fermionic sector: $\mathcal{L}_{\text{kin,f}} = i\left(\bar{Q}_L\right)^i \not{D}(Q_L)_i + i\left(\bar{L}\right)^i \not{D}(L)_i + i\left(\bar{d}_R\right)^i \not{D}(d_R)_i + i(\bar{u}_R)^i \not{D}(u_R)_i + i(\bar{e}_R)^i \not{D}(e_R)_i +$
$+ i\left(\bar{E}_L\right)^i \not{D}(E_L)_i + i\left(\bar{E}_R\right)^i \not{D}(E_R)_i + i\left(\bar{D}_L\right)^i \not{D}(D_L)_i + i\left(\bar{D}_R\right)^i \not{D}(D_R)_i + i(\bar{\nu}_R)^i \not{D}(\nu_R)_i,$

Yukawa sector: $\mathcal{L}_{\text{y}} = (Y^a)_{ij}\left(\bar{Q}_L\right)^i (D_R)^j \phi_a + (\Gamma^a)_{ij}\left(\bar{Q}_L\right)^i (d_R)^j \phi_a + (\Delta^a)_{ij}\left(\bar{Q}_L\right)^i (u_R)^j \tilde{\phi}_a +$
$+ (\Theta^a)_{ij}\left(\bar{E}_L\right)^i (e_R)^j \phi_a + (\Upsilon^a)_{ij}\left(\bar{E}_L\right)^i (\nu_R)^j \tilde{\phi}_a + (\Sigma^a)_{ij}\left(\bar{L}\right)^i (\nu_R)^j \tilde{\phi}_a +$
$+ (\Pi^a)_{ij}\left(\bar{L}\right)^i (e_R)^j \phi_a + (\Omega^a)_{ij}\left(\bar{E}_R\right)^i (\nu_R)^j \tilde{\phi}_a + \text{H.c.},$

Bilinear terms: $\mathcal{L}_{\text{bil}} = (M_D)_{ij}\left(\bar{D}_L\right)^i (D_R)^j + (M_E)_{ij}\left(\bar{E}_L\right)^i (E_R)^j + \frac{1}{2}(M_{\nu_R})_{ij}(\bar{\nu}_R)^i (\nu_R)^j +$
$+ (M_{LE})_{ij}\left(\bar{L}\right)^i (E_R)^j + (\Xi)_{ij}\left(\bar{D}_L\right)^i (d_R)^j.$

Scalar potential: $V\left(\phi, \phi^\dagger\right) = (m_i)^2\left|\phi^i\right|^2 + \left(m_{ij}^2 \phi^i \left(\phi^j\right)^\dagger + \text{H.c.}\right) + \lambda_{ijkl}\left(\phi^i \left(\phi^j\right)^\dagger \phi^k \left(\phi^l\right)^\dagger + \text{H.c.}\right)$

Model implementation using SARAH, taking as input all parameters necessary in to the definition of the Lagrangian density.

SARAH model and UFO files

$$\begin{aligned}\mathcal{L} = &-\frac{1}{4}C^{\mu\nu}C_{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} \\ &+ i\bar{\psi}\gamma^{\mu}d_{\mu}\psi + i\bar{\chi}\gamma^{\mu}D_{\mu}\chi \\ &+ [\mathcal{D}^{\mu}\phi]^{*}\mathcal{D}_{\mu}\phi - V(\phi,\phi^{*}) \\ &+ y\bar{\psi}\phi\chi\end{aligned}$$

Field	$U(1)_B$	$U(1)_C$
ϕ	1	1
ψ	1	0
χ	1	-1

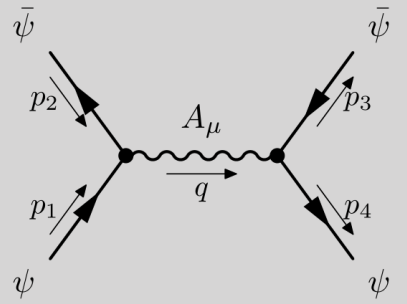
Hard scattering Monte Carlo simulator to compute matrix elements for signal and background events.

We compute cross sections and decay widths up to leading order processes.

MadGraph hard scattering

$$i\mathcal{M} = \bar{v}(\mathbf{p_2})(-ie\gamma^{\mu})u(\mathbf{p_1})\frac{-ig_{\mu\nu}}{q^2+i\epsilon}\bar{u}(\mathbf{p_4})(-ie\gamma^{\nu})v(\mathbf{p_3})$$

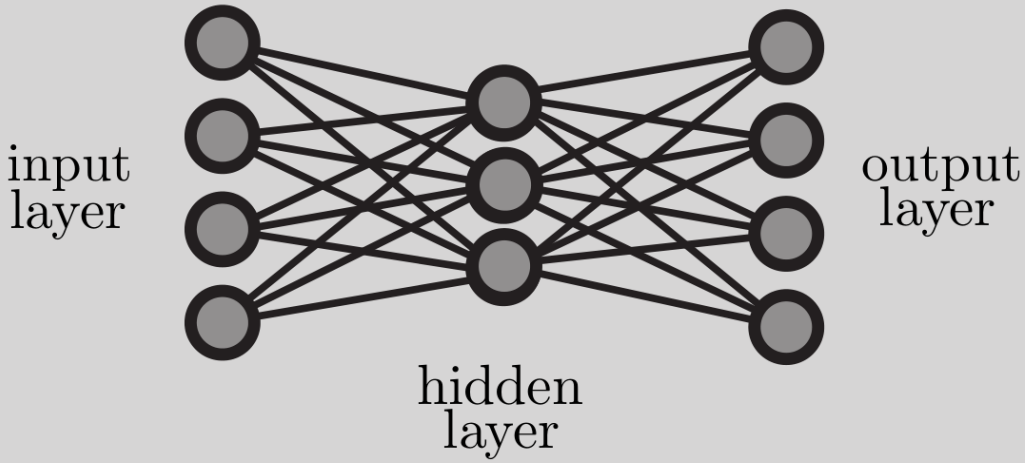
$$\sigma_i = \frac{1}{64\pi^2 s}\int d\Omega(\hat{\mathbf{p}}_3)|i\mathcal{M}|^2$$



The final step is compute the significance of signal events over background using a simple neural network.

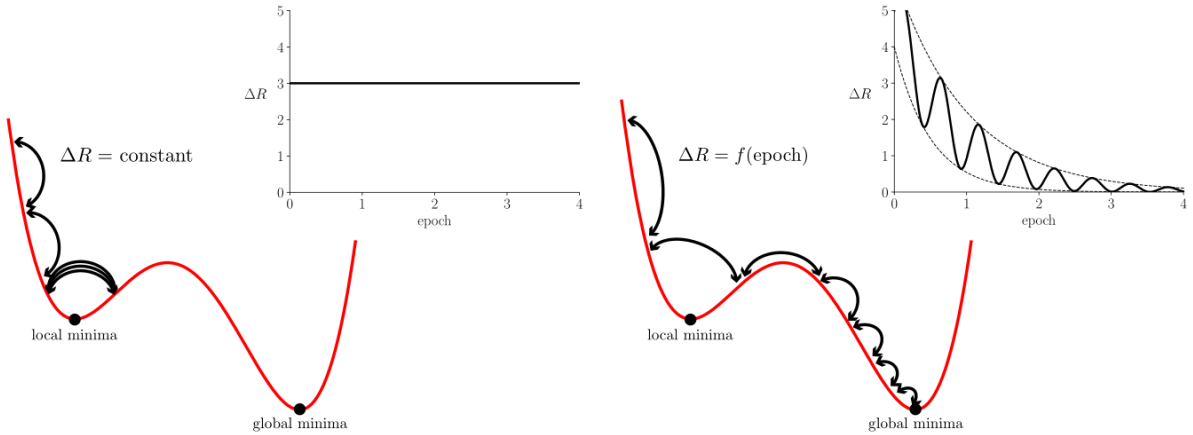
Machine Learning analysis

$$Z_A = \left[2\left((s+b)\ln\left(\frac{(s+b)(b+\sigma_b^2)}{b^2+(s+b)\sigma_b^2}\right) - \frac{b^2}{\sigma_b^2}\ln\left(1+\frac{\sigma_b^2 s}{b(b+\sigma_b^2)}\right)\right)\right]^{1/2}$$



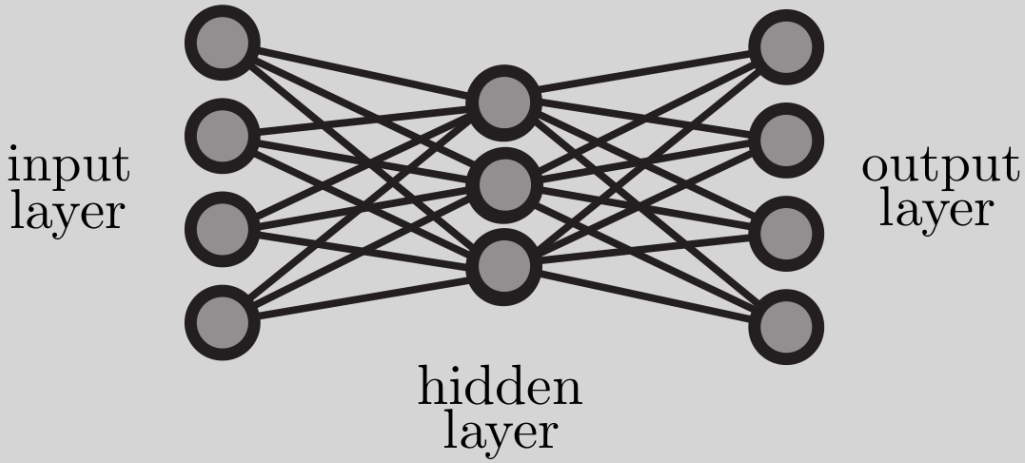
Neural networks are rooted in the universal approximation theorem, that is, for given number of hidden layers and finite number of hidden units, one can approximate any function on compact subsets of R^n.

However, that approximation is only possible when neural networks are provided with appropriate weights and bias. So, the determination of such parameters are resolved by optimisation algorithms, that minimise a given set of parameters.



Machine Learning analysis

$$Z_A = \left[2\left((s+b)\ln\left(\frac{(s+b)(b+\sigma_b^2)}{b^2+(s+b)\sigma_b^2}\right) - \frac{b^2}{\sigma_b^2}\ln\left(1+\frac{\sigma_b^2 s}{b(b+\sigma_b^2)}\right)\right)\right]^{1/2}$$



The model may end up stuck in local minimas, as opposed to global ones. Utilisation of adaptive learning rates which vary with epoch addresses this problem.
Other problems, such as possible overfitting of data and oversampling are commonly known in data science and algorithms that address these problems are readily available via current deep learning packages.

Deep learning differentiates itself from regular optimisation problems via the activation function.

When a neuron receives an input

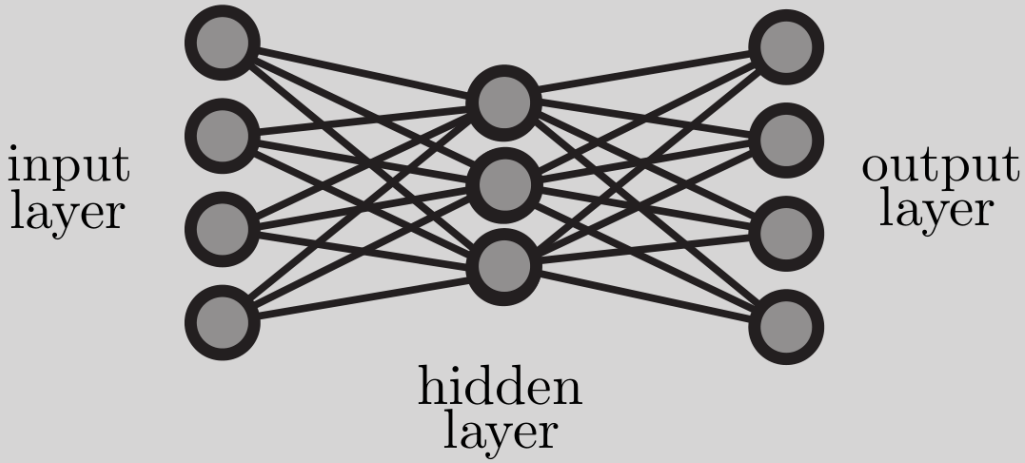
$$y = Wx + b$$

An activation function acts on y

$$f \circ y = f(Wx + b)$$

Machine Learning analysis

$$Z_A = \left[2\left((s+b)\ln\left(\frac{(s+b)(b+\sigma_b^2)}{b^2+(s+b)\sigma_b^2}\right) - \frac{b^2}{\sigma_b^2}\ln\left(1+\frac{\sigma_b^2 s}{b(b+\sigma_b^2)}\right)\right)\right]^{1/2}$$



In the subsequent neurons, composition continues $y_1 \to y_2 = f(y_1) \to y_3 = g(f(y_1)) \to \dots$

Usually one applies back-propagation algorithms to fine tune the weights of the model, to ensure lower errors. The use of differentiable activation functions is at times preferred.

There is other important functions/parameters such as regularisers, loss functions and initialisers.

Let us dicuss the main properties of the model, starting with the superpotential

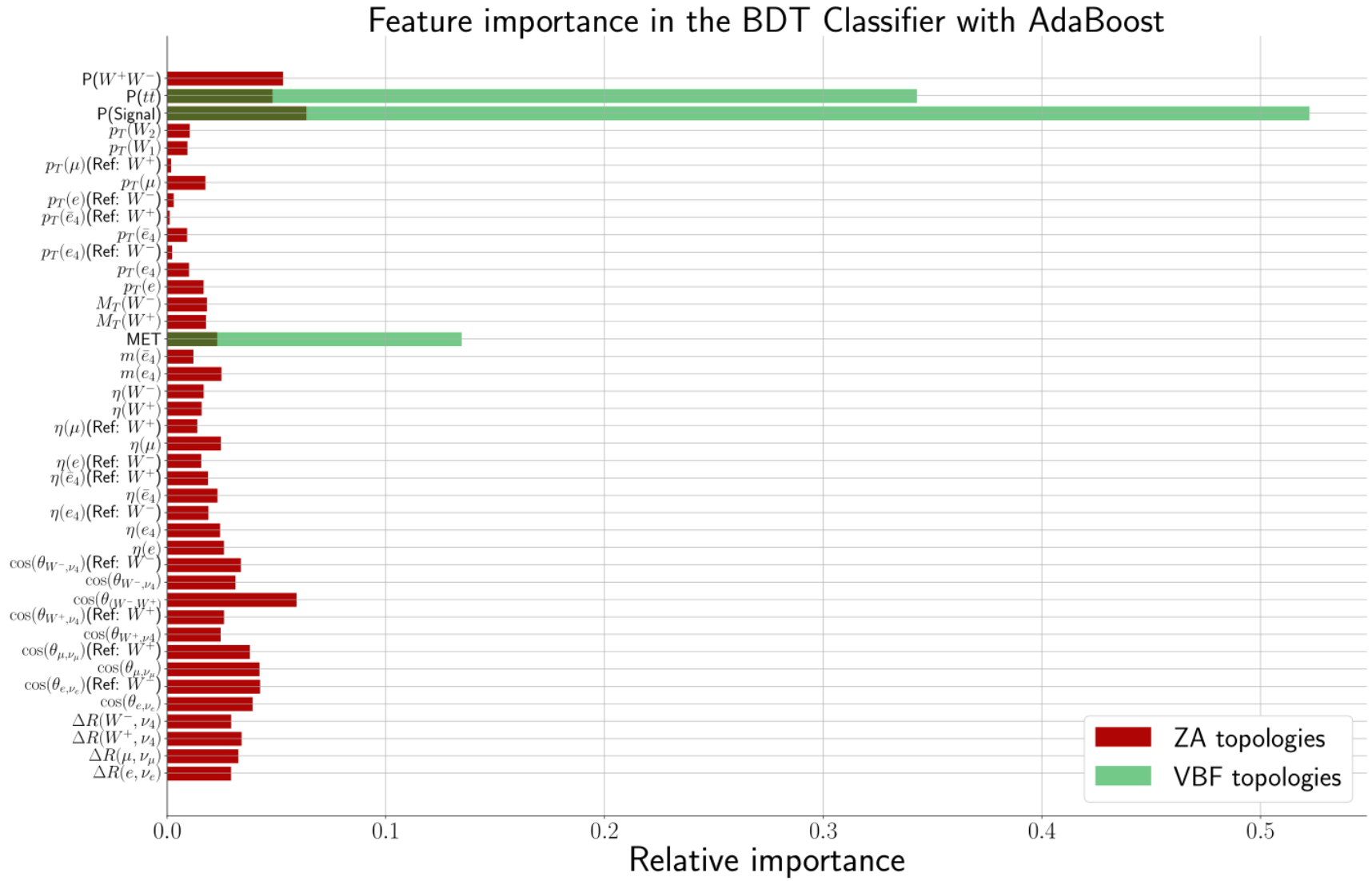
$$
\begin{aligned}
W =& \mathcal{Y}_1 \varepsilon_{ij} \left(\boldsymbol{\chi}^i \boldsymbol{q}_L^3 \boldsymbol{q}_R^j + \boldsymbol{\ell}_R^i \boldsymbol{D}_L^3 \boldsymbol{q}_R^j + \boldsymbol{\ell}_L^i \boldsymbol{q}_L^3 \boldsymbol{D}_R^j + \boldsymbol{\phi}^i \boldsymbol{D}_L^3 \boldsymbol{D}_R^j \right) \\
& - \mathcal{Y}_2 \varepsilon_{ij} \left(\boldsymbol{\chi}^i \boldsymbol{q}_L^j \boldsymbol{q}_R^3 + \boldsymbol{\ell}_R^i \boldsymbol{D}_L^j \boldsymbol{q}_R^3 + \boldsymbol{\ell}_L^i \boldsymbol{q}_L^j \boldsymbol{D}_R^3 + \boldsymbol{\phi}^i \boldsymbol{D}_L^j \boldsymbol{D}_R^3 \right) \\
& + \mathcal{Y}_2 \varepsilon_{ij} \left(\boldsymbol{\chi}^3 \boldsymbol{q}_L^i \boldsymbol{q}_R^j + \boldsymbol{\ell}_R^3 \boldsymbol{D}_L^i \boldsymbol{q}_R^j + \boldsymbol{\ell}_L^3 \boldsymbol{q}_L^i \boldsymbol{D}_R^j + \boldsymbol{\phi}^3 \boldsymbol{D}_L^i \boldsymbol{D}_R^j \right)
\end{aligned}
$$

SM Leptons	SM Quarks	VLLs	VLQs	Right-handed neutrinos
$\boldsymbol{\ell}_{L,R}$	$\boldsymbol{q}_{L,R}$	$\boldsymbol{\chi}$	$\boldsymbol{D}_{L,R}$	$\boldsymbol{\chi}, \boldsymbol{\phi}$

Note that only quarks gain mass at tree-level. For example,

$$\boldsymbol{\phi} \boldsymbol{D}_L \boldsymbol{D}_R \to \left\langle \tilde{\phi} \right\rangle \boldsymbol{D}_L \boldsymbol{D}_R$$

Terms of the form $\langle \tilde{\chi} \rangle \, \boldsymbol{\ell}_L \boldsymbol{\ell}_R$ aren't allowed by symmetry.



Deeper architecture for the ZA BDT, then VBF BDT. VBF then uses less features overall.

There is no unique configuration for the BDT as well

Some random notes:

All new states aren't arbitrarily introduced. They all belong to representations of the E6 group. They are a direct consequence of the unification picture.

The W likes to decay into light jets more then leptons (Branching ratios of 68%). New signals with jets. ttbar background is now important for VBF signals with jets, since now we have jets in the central region.

VBF signals are characterized with jets in the forward-backward region (eta -> inf). ttbar has jets in the central region (eta -> 0). ttbar background is minimized for VBF + lepton final states.

Sterile neutrino. Even if nu4 is not DM, it still is long lived enough that it escapes the detector. Meaning it is still valid the approximation of missing energy

FNCs not properly analysed. There are very tight constraints in flavor observables

Some random notes:

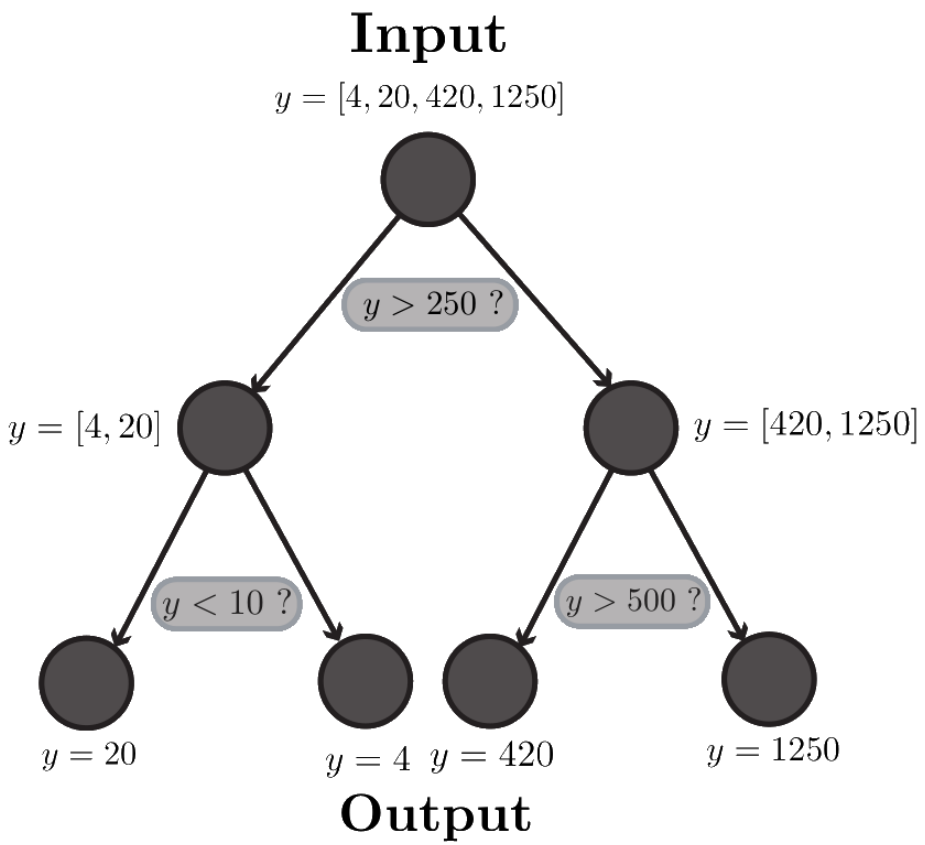
Systematics of 1%. Low significance errors (computed by the percentage of background events). BDT separates the events very well, with very low errors. If we increase the systematics we will increase the error.

In a nutshell, we can interpret Deep learning as an fitting algorithm, if you will. The L2 regularizer works by penalizing higher order expansions in the polynomial fitting to avoid overfitting.

A bigger batch size implies a more accurate and bigger representation of the data. The size of the the batch size works here, since we deal with numerical datasets, and as such we do not require extensive use of memory. Image datasets will require lower batch sizes.

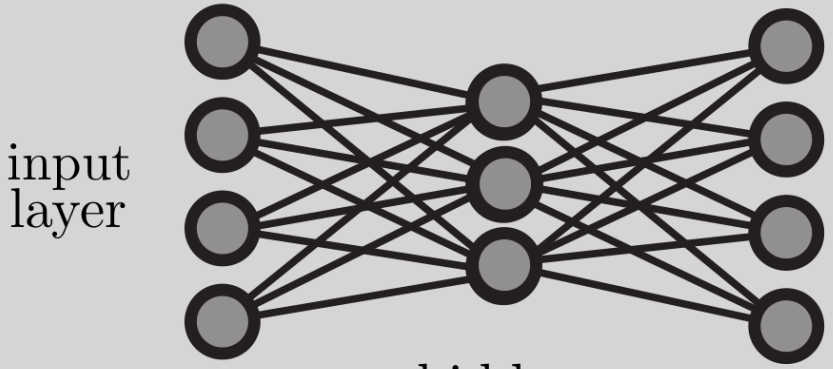
Some random notes:

200 epochs with patience of 5 epochs. If, after 5 epochs, the model does not improve its accuracy, then it stops.



Machine Learning analysis

$$Z_A = \left[2\left((s+b)\ln\left(\frac{(s+b)(b+\sigma_b^2)}{b^2+(s+b)\sigma_b^2}\right) - \frac{b^2}{\sigma_b^2}\ln\left(1+\frac{\sigma_b^2 s}{b(b+\sigma_b^2)}\right)\right)\right]^{1/2}$$



While the deep learning algorithm provides great accuracies, as we will see later on. An extra step will be required, where a BDT is considered. The BDT will take kinematic features and the deep learning predictions as input. The process finishes with the computation of the Asimov significance.