

Measurement of Higgs Production Cross Section via Vector Boson Fusion in $H \rightarrow ZZ \rightarrow 4l$ final state at 13 TeV using Artificial Neural Networks

Miquéias Melo de Almeida¹

Advisor: Andre Sznajder¹
Co-advisor: Nicola De Filippis²

¹Universidade do Estado do Rio de Janeiro

²INFN and Politecnico di Bari

April 11th, 2019



Table of Contents

1 CMS AN-18-120

- Introduction
- Theoretical Motivation
- The CMS Experiment at CERN
- Datasets, Triggers and Simulated Samples
- Objects and Events Selections
- Artificial Neural Networks (ANNs)
- Systematic Uncertainties
- Statistical Analysis and Results

2 CMS L1 Tracking Trigger AM+FPGA

- Introduction
- Hardware Work
- Simulation Studies

3 The Fast Matrix Element (FastME)

- Introduction
- Simulation Studies
- The FastME Package

4 Conclusions

The CMS AN-18-120

Introduction

- This analysis is a measurement of the Higgs VBF production XS in the HZZ4L channel using ANN discriminants;

- Analysis characteristics:

- follows similar requirements established in CMS HZZ4L analysis;
- VBF signal region (VBF-SR) defined similarly to CMS VBF category (no MELA);
- proposes the usage of a 3rd jet when available;
- events at VBF-SR divided into two jet-based subcategories;
- Artificial Neural Network (ANN) as a VBF discriminant;

Documentation

Available on the CMS information server

CMS AN-18-120

CMS Draft Analysis Note

The content of this note is intended for CMS internal use and distribution only

2018/11/29
 Head Id: 482729
 Archive Id: 479138:482729MP
 Archive Date: 2018/11/28
 Archive Tag: trunk

Measurement of Higgs Production Cross Section via Vector Boson Fusion in $H \rightarrow ZZ \rightarrow 4l$ final state at 13 TeV using Artificial Neural Networks

M. M. Almeida¹, A. Sznajder¹, N. De Filippis², P. Bhat³, S. Uzunyan⁴, M. Eads⁴, D. Faia⁴, R. Singh⁴, H. Prosper⁵, and K. Johnson⁵

¹Universidade do Estado do Rio de Janeiro

²INFN and Politecnico di Bari

³Fermi National Accelerator Laboratory

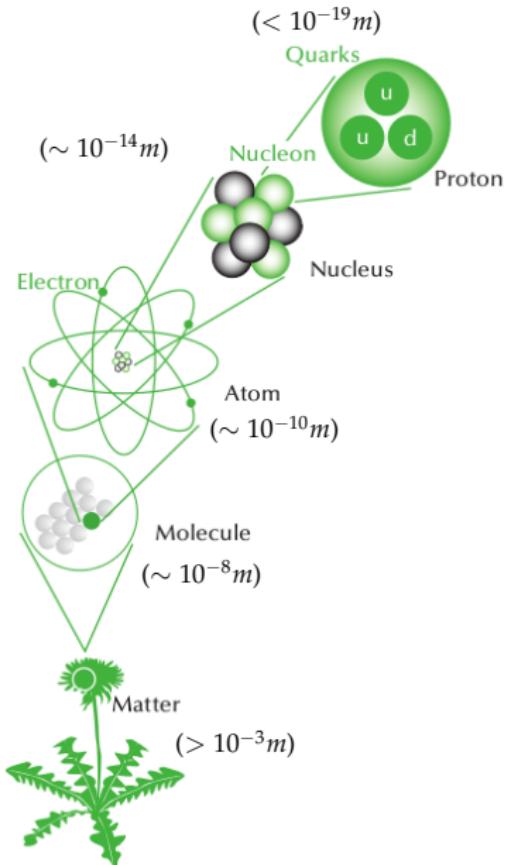
⁴Northern Illinois University

⁵Florida State University

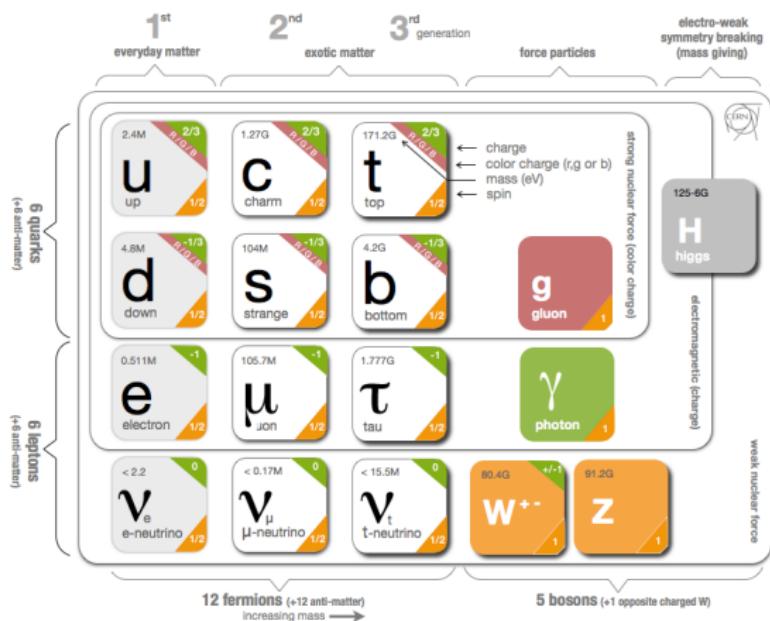
Abstract

A isolated measurement of the Higgs boson production cross section in the Vector Boson Fusion (VBF) production mode is performed in the $H \rightarrow ZZ \rightarrow 4l(l = e, \mu)$ channel. The study is performed using data samples corresponding to an integrated luminosity of 35.9 fb^{-1} from pp collisions at $\sqrt{s} = 13 \text{ TeV}$, which has been collected by the CMS experiment during 2016 at the LHC. A multivariate analysis is performed through the usage of Artificial Neural Networks (ANNs). Statistical shape analysis of ANNs developed for two orthogonal jet-based categories is done by combining the discriminants distribution from each category. The Higgs VBF signal strength modifier is measured to be $\mu_{qql\ell}^{obs} = 1.28^{+0.34}_{-0.88}$ for a expected Higgs boson of $m_H = 125 \text{ GeV}$. This result is comparable with the SM expectation. The observed significance of the present analysis is $\sigma_{qql\ell}^{obs} = 1.9$, while the expected one is $\sigma_{qql\ell}^{exp} = 1.8$. The observed and expected 95%CL limits are estimated as $\mu_{qql\ell}^{exp} < 3.79$ and $\mu_{qql\ell}^{exp} < 1.66$, respectively. A projection for future luminosities is also presented and it is expected that the present analysis will have enough significance for the VBF-Higgs production evidence (3.4σ) at 150 fb^{-1} and the observance (5.1σ) at 359 fb^{-1} .

Particle Physics and the Standard Model



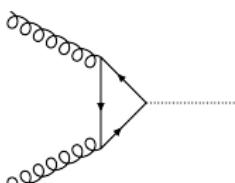
- Elementary Particle Physics is the field of Physics dedicated to the study of fundamental building blocks of matter and their interactions;
- The Standard Model (SM) resumes what physicists know so far:



The Higgs Boson

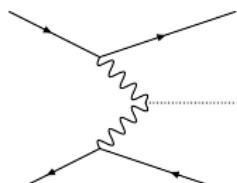
- In Physics symmetry dictates interactions (gauge theories);
- However, gauge bosons and fermions must be massless in the EWK theory;
- Solution: Higgs scalar field to spontaneously break gauge symmetry giving mass to W, Z and fermions;
- Higgs particle (the field quanta) can be produced in some process, such as:

$$gg \rightarrow H$$



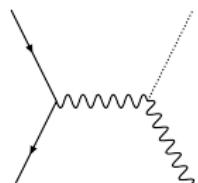
fusion of gluons

$$q\bar{q} \rightarrow H q\bar{q}$$



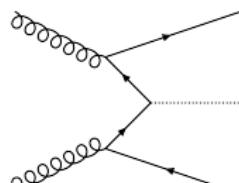
fusion of bosons Z/W
(VBF)

$$q\bar{q} \rightarrow VH$$



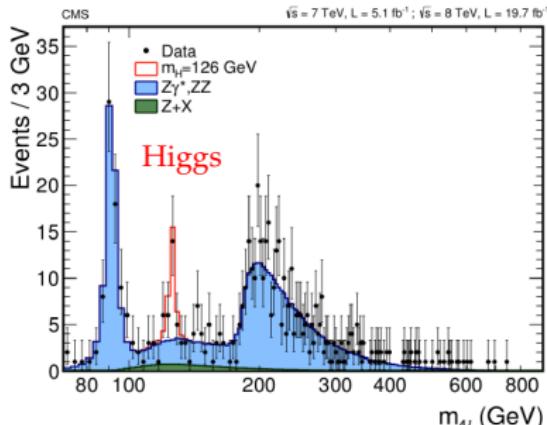
radiation from Z/W

$$gg \rightarrow t\bar{t}H$$



fusion of top quarks

Higgs discovered at the LHC in 2012



Physics Processes in this Analysis

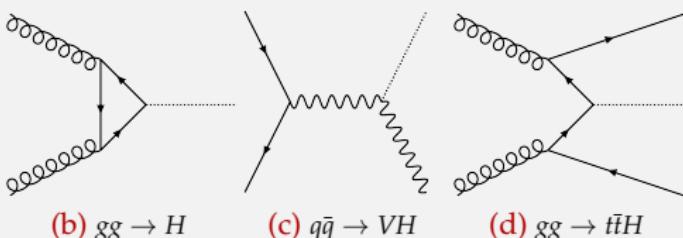
- Signal is Vector Boson Fusion (VBF):

- second largest Higgs production mode;
- tree-level and clean of beyond-SM processes;
- good frame for measurements of Higgs properties;
- estimated from MC normalized by $\sigma.BR$ given by LHC Higgs XS Working Group;

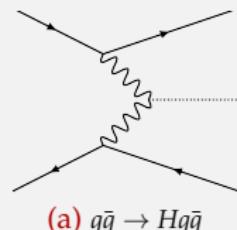
- Backgrounds:

Remaining SM Higgs Production Modes

Estimated from MC normalized by $\sigma.BR$
given by LHC Higgs XS Working Group

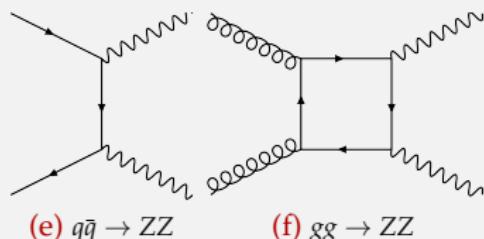


VBF (only)



ZZ's

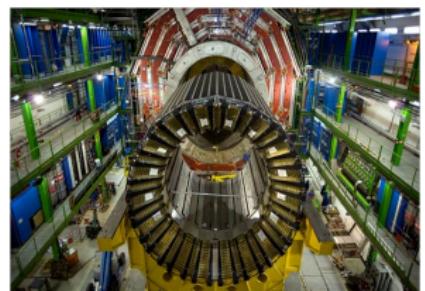
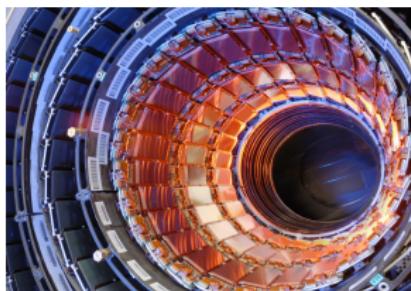
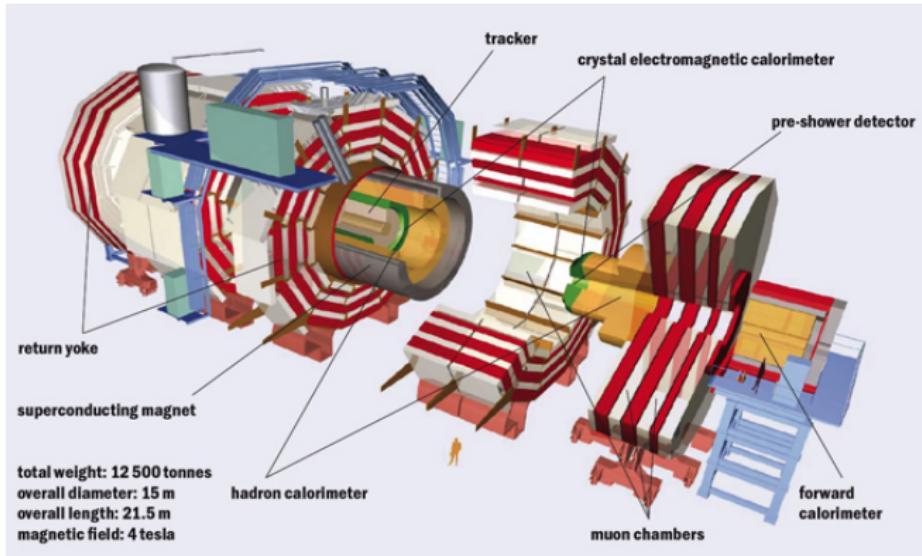
Estimated from MC normalized
by $\sigma.BR$



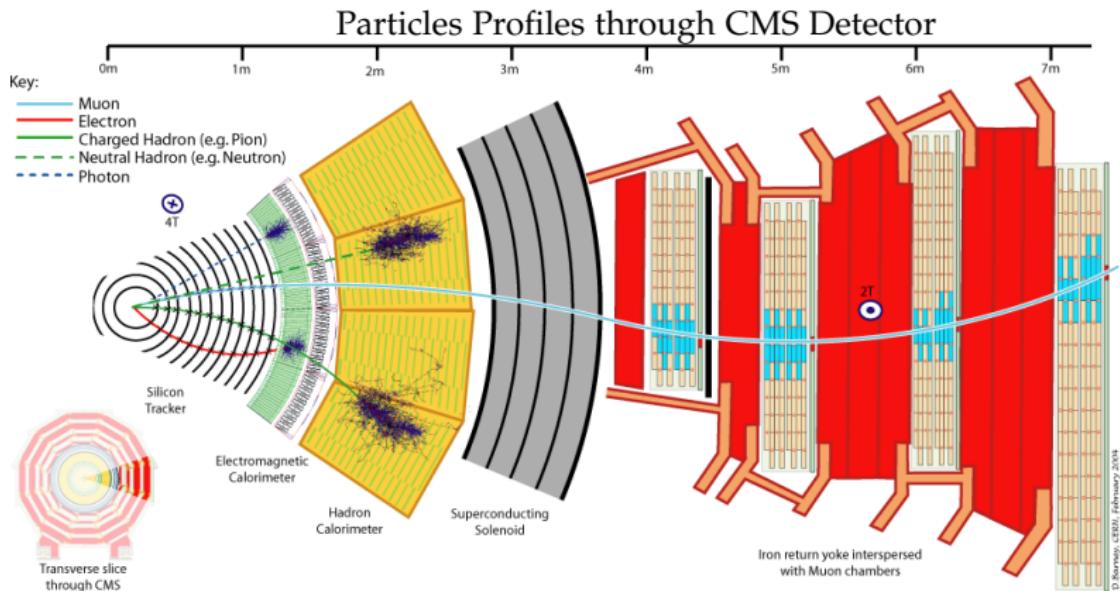
Z + X

Estimated via Fake Rate *data-driven* method

Compact Muon Solenoid in a Nutshell



CMS-Particles Interaction Profile



- CMS has particle-specialized sub-detectors;
- "Long-life" particles (reach detectors) are identified by signal patterns on CMS;
- "Short-life" particles (decay into other particles, not reaching any detector) are identified through the properties of "long-life" particles;

Datasets and Triggers

- Data:

- full 2016 Data: $L = 35.9\text{fb}^{-1}$, 03Feb ReReco (full list in backup);

- Triggers:

- based on multi-lepton HLT paths;
- isolated di-lepton paths + non-isolated tri-lepton paths + single-lepton paths;
- requirements to avoid double-counting is applied;
- overall trigger efficiency is higher than 99% wrt. 4-lepton analysis selection;

HLT path	L1 seed	Prescale	Primary dataset
HLT_Ele17_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	L1_DoubleEG_15_10	1	DoubleEG
HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	L1_DoubleEG_22_10	1	DoubleEG
HLT_DoubleEle33_CaloIdL_GsfTrkIdVL	(Multiple)	1	DoubleEG
HLT_Ele16_Ele12_Ele8_CaloIdL_TrackIdL	L1_TripleEG_14_10_8	1	DoubleEG
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL	L1_DoubleMu_11_4	1	DoubleMuon
HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL	L1_DoubleMu_11_4	1	DoubleMuon
HLT_TripleMu_12_10_5	L1_TripleMu_5_5_3	1	DoubleMuon
HLT_Mu8_TrkIsoVVL_Ele17_CaloIdL_TrackIdL_IsoVL	L1_Mu5_EG15	1	MuonEG
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL	L1_Mu5_EG20	1	MuonEG
HLT_Mu17_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL	L1_Mu12_EG10	1	MuonEG
HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL	L1_Mu20_EG10	1	MuonEG
HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL	L1_SingleMu*	1	MuonEG
HLT_Mu8_DiEle12_CaloIdL_TrackIdL	L1_Mu6_DoubleEG10	1	MuonEG
HLT_DiMu9_Ele9_CaloIdL_TrackIdL	L1_DoubleMu7_EG7	1	MuonEG
HLT_Ele25_eta2p1_WPTight	L1_SingleEG*	1	SingleElectron
HLT_Ele27_WPTight	L1_SingleEG*	1	SingleElectron
HLT_Ele27_eta2p1_WPLoose_Gsf	L1_SingleEG*	1	SingleElectron
HLT_IsoMu20 OR HLT_IsoTkMu20	L1_SingleMu*	1	SingleMuon
HLT_IsoMu22 OR HLT_IsoTkMu22	L1_SingleMu*	1	SingleMuon

Simulated Samples

- Signal: VBF_HToZZTo4L_M125_13TeV_powheg2_JHUGenV6_pythia8;
- Background: remaining samples in the table;
- Special samples for ggH (*) and qqZZ (**) included.

Process	Dataset Name	$\sigma \cdot BR (fb)$
SM Higgs MC Samples		
$* gg \rightarrow H \rightarrow ZZ \rightarrow 4l$	[1] GluGluHToZZTo4L_M125_13TeV_powheg2_minloHJJ_JHUGenV6_pythia8	12.180
$q\bar{q} \rightarrow H q\bar{q} \rightarrow ZZ q\bar{q} \rightarrow 4l q\bar{q}$	[1] VBF_HToZZTo4L_M125_13TeV_powheg2_JHUGenV6_pythia8	1.044
$q\bar{q} \rightarrow W^+ H \rightarrow W^+ ZZ \rightarrow 4l + X$	[1] WplusH_HToZZTo4L_M125_13TeV_powheg2-minlo-HWJ_JHUGenV6_pythia8	0.232
$q\bar{q} \rightarrow W^- H \rightarrow W^- ZZ \rightarrow 4l + X$	[1] WminusH_HToZZTo4L_M125_13TeV_powheg2-minlo-HWJ_JHUGenV6_pythia8	0.147
$q\bar{q} \rightarrow ZH \rightarrow ZZZ \rightarrow 4l + X$	[1] ZH_HToZZ_4LFilter_M125_13TeV_powheg2-minlo-HZJ_JHUGenV6_pythia8	0.668
$gg \rightarrow t\bar{t}H \rightarrow t\bar{t}ZZ \rightarrow 4l + X$	[1] ttH_HToZZ_4LFilter_M125_13TeV_powheg_JHUGen_pythia8	0.393
$gg \rightarrow H \rightarrow WW \rightarrow 2l2v$	[1] GluGluHToWWTo2L2Nu_M125_13TeV_powheg_JHUGen_pythia8	1101.790
$q\bar{q} \rightarrow H q\bar{q} \rightarrow WW q\bar{q} \rightarrow 2l2v$	[1] VBFHToWWTo2L2Nu_M125_13TeV_powheg_JHUGen_pythia8	85.776
$q\bar{q} \rightarrow W^+ H \rightarrow W^+ WW \rightarrow l\nu 2l2v$	[1] HWplusJ_HToWWTo2L2Nu_WToLNu_M125_13TeV_powheg_pythia8	2.138
$q\bar{q} \rightarrow W^- H \rightarrow W^- WW \rightarrow l\nu 2l2v$	[1] HWminusJ_HToWWTo2L2Nu_WToLNu_M125_13TeV_powheg_pythia8	1.357
$q\bar{q} \rightarrow ZH \rightarrow ZWW \rightarrow 2l2l2v$	[1] HZJ_HToWWTo2L2Nu_ZTo2L_M125_13TeV_powheg_pythia8	2.029
$gg \rightarrow b\bar{b}H \rightarrow b\bar{b}WW \rightarrow 2l2v + X$	[1] bbHToWWTo2L2Nu_M-125.4FS_yb2_13TeV_amcatnlo	11.068
SM Backgrounds MC Samples		
$q\bar{q} \rightarrow ZZ \rightarrow 4l$	[1] ZZTo4L_13TeV_powheg_pythia8	1256.000
$q\bar{q} \rightarrow ZZ \rightarrow 4l$	[1] ZZTo4L_13TeV_amcatnlo_FXFX_pythia8	1212.000
$* * q\bar{q} \rightarrow ZZ \rightarrow 4l + jets (EWK)$	[1] ZZJJTo4L_EWK_13TeV-madgraph_pythia8/	4.404
$gg \rightarrow ZZ \rightarrow 4e$	[1] GluGluToContinToZZTo4e_13TeV_MCFM701	1.590
$gg \rightarrow ZZ \rightarrow 4\mu$	[1] GluGluToContinToZZTo4mu_13TeV_MCFM701	1.590
$gg \rightarrow ZZ \rightarrow 4\tau$	[1] GluGluToContinToZZTo4tau_13TeV_MCFM701	1.590
$gg \rightarrow ZZ \rightarrow 2e2\mu$	[1] GluGluToContinToZZTo2e2mu_13TeV_MCFM701	3.190
$gg \rightarrow ZZ \rightarrow 2e2\tau$	[1] GluGluToContinToZZTo2e2tau_13TeV_MCFM701	3.190
$gg \rightarrow ZZ \rightarrow 2\mu2\tau$	[1] GluGluToContinToZZTo2mu2tau_13TeV_MCFM701	3.190
$Z \rightarrow ll + jets$	[1] DYJetsToLL_M-10to50TuneCUETP8M1_13TeV-amcatnloFXFX_pythia8 [1] DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX_pythia8	$6.104e^6$ $1.861e^7$
$t\bar{t}$	[1] TTJets_TuneCUETP8M1_13TeV-amcatnloFXFX_pythia8	$815.96e^3$
$t\bar{t} \rightarrow 2l2v2b$	[1] TTTTo2L2Nu_13TeV_powheg	$8.731e^4$
$WZ \rightarrow 3l\nu$	[1] WZTo3LNu_TuneCUETP8M1_13TeV_powheg_pythia8	4430
ZZZ	[1] ZZZ_TuneCUETP8M1_13TeV-amcatnlo_pythia8	13.980
WWZ	[1] WWZ_TuneCUETP8M1_13TeV-amcatnlo_pythia8	165.100
WZZ	[1] WZZ_TuneCUETP8M1_13TeV-amcatnlo_pythia8	55.650
$t\bar{t}W$	[1] TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin_pythia8	204.300
$t\bar{t}Z$	[1] TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo_pythia8	252.900
[1] RunIIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/		

Objects Selections

Electrons (momentum calibration applied)

Loose

- $p_T > 7 \text{ GeV};$
- $|\eta| < 2.5;$
- $|d_{xy}| < 0.5 \text{ cm};$
- $|d_z| < 1.0 \text{ cm}.$

Tight

- Loose selections plus:
 - $|SIP_{3D}| < 4.0;$
 - Isolation ($\Delta R = 0.3$) $< 0.35;$
 - MVA (BDT) calorimeter-based:

Minimum BDT score	$ \eta < 0.8$	$0.8 < \eta < 1.479$	$ \eta > 1.479$
$5 < p_T < 10 \text{ GeV}$	-0.211	-0.396	-0.215
$p_T > 10 \text{ GeV}$	-0.870	-0.838	-0.763

Muons (momentum calibration and FSR applied)

Loose

- Global/Tracker Muons;
- $p_T > 5 \text{ GeV};$
- $|\eta| < 2.4;$
- $|d_{xy}| < 0.5 \text{ cm};$
- $|d_z| < 1.0 \text{ cm}.$

Tight

- Loose selections plus:
 - PF Muon;
 - $|SIP_{3D}| < 4.0;$
 - Isolation ($\Delta R = 0.3$) $< 0.35;$
 - Ghost-cleaning (single- μ as more).

Objects Selections

Photons and FSR

- $p_T > 2 \text{ GeV}$, $|\eta| < 2.4$;
- Isolation ($\Delta R = 0.3$) < 1.8 ;
- $\Delta R(\gamma, l)/E_{T,\gamma}^2 \geq 0.012$ and $\Delta R(\gamma, l) \geq 0.5$;

Jets (JECs applied)

- anti-kT ($R = 0.4$) PF CHS (*Carged Hadron Subtracted*) jets;
- $p_T > 30 \text{ GeV}$, $|\eta| < 4.7$;
- cleaning $\Delta R(jet, l/\gamma) > 0.4$;
- b-tagging with CSV (*Combined Secondary Vertex*) algorithm;

MET (*Missing Transverse Energy*)

- PF MET with type-1 correction: $\vec{E}_T^{miss} = -(\sum_{jets} \vec{p}_T^{JEC} + \sum_{uncl.} \vec{p}_T)$;
- filters^a from JETMET POG applied (improves signal-to-noise ratio).

^aSee backup slides.

Event Selections

SM Higgs Region

- 1 **Z candidates:** pair of same-flavor, opposite-charge and FSR corrected leptons (e^+e^- , $\mu^+\mu^-$) having invariant mass $12 < m_{ll(\gamma)} < 120$ GeV;
- 2 **ZZ candidates:** pair of non-overlapping (different leptons) Z candidates. The Z with smallest $|m_{ll(\gamma)} - m_Z^{PDG}|$ is identified as Z_1 and the other Z as Z_2 . ZZ candidates must satisfy:
 - any two leptons must have $\Delta R(\eta, \phi) > 0.02$ (**ghost removal**);
 - at least two out of the four leptons must have $p_T > 10$ and 20 GeV;
 - any two leptons must have (without FSR- γ) $m_{ll} > 4$ GeV (**QCD suppression**);
 - $m_{Z_1} > 40$ GeV and $m_{Z_1 Z_2} > 100$ GeV;
 - if more than one ZZ candidate survives previous cuts, the one with highest scalar leptons p_T sum is chosen;

m4l Distributions and Yields

After SM Higgs Selections

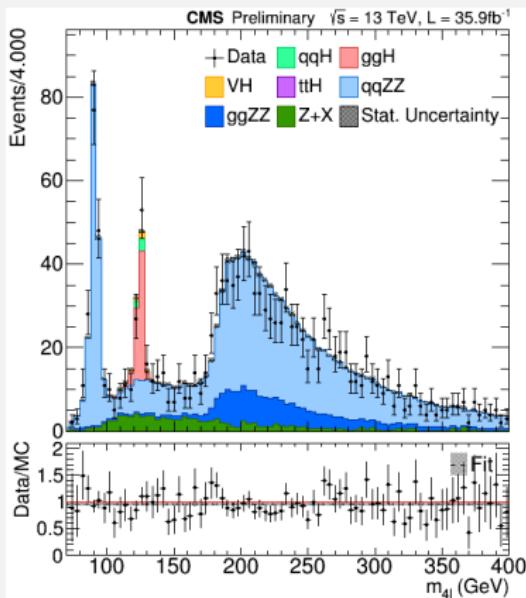


Table 16: Number of expected events from background and signal, with total (statistical+systematic) uncertainty reported, and number of observed events after the SM Higgs selections in the mass range $m_{4l} > 70 \text{ GeV}$.

Process	4 μ	4e	2e2 μ	4l
ggH	19.34 ± 3.73	11.02 ± 2.34	25.99 ± 5.20	56.35 ± 6.81
VH	1.45 ± 0.21	0.92 ± 0.14	2.14 ± 0.32	4.51 ± 0.41
tH	0.36 ± 0.08	0.23 ± 0.05	0.48 ± 0.11	1.07 ± 0.14
qqZZ+ZZJJ	387.01 ± 24.48	234.64 ± 24.81	538.35 ± 43.61	1160.00 ± 55.83
ggZZ	65.81 ± 7.43	43.85 ± 6.14	102.32 ± 12.60	211.98 ± 15.86
Z+X	24.28 ± 7.71	27.80 ± 8.09	56.90 ± 17.09	108.99 ± 20.42
Σ backgrounds	498.25 ± 26.98	318.46 ± 26.91	726.18 ± 48.78	1542.89 ± 61.90
qqH (signal $m_H = 125 \text{ GeV}$)	1.86 ± 0.36	1.10 ± 0.24	2.53 ± 0.52	5.49 ± 0.68
Total expected	500.11 ± 26.98	319.56 ± 26.91	728.71 ± 48.78	1548.38 ± 61.90
Observed	503	287	669	1459

Event Selections

SM Higgs Region

- 1 Z candidates:** pair of same-flavor, opposite-charge and FSR corrected leptons (e^+e^- , $\mu^+\mu^-$) having invariant mass $12 < m_{ll(\gamma)} < 120$ GeV;
- 2 ZZ candidates:** pair of non-overlapping (different leptons) Z candidates. The Z with smallest $|m_{ll(\gamma)} - m_Z^{PDG}|$ is identified as Z_1 and the other Z as Z_2 . ZZ candidates must satisfy:
 - any two leptons must have $\Delta R(\eta, \phi) > 0.02$ (**ghost removal**);
 - at least two out of the four leptons must have $p_T > 10$ and 20 GeV;
 - any two leptons must have (without FSR- γ) $m_{ll} > 4$ GeV (**QCD suppression**);
 - $m_{Z_1} > 40$ GeV and $m_{Z_1Z_2} > 100$ GeV;
 - if more than one ZZ candidate survives previous cuts, the one with highest scalar leptons p_T sum is chosen;



VBF Signal Region (VBF-SR)

- 3** In order to enhance VBF-to-background ratio:
 - Number of jets:
 - EITHER, 2 or 3 jets from which at most one b-tagged jet;
 - OR, more than 3 jets with no b-tagged jet;
 - ZZ candidates must have $118 \leq m_{Z_1Z_2} \leq 130$ GeV;

m_{4l} Distributions and Yields

After VBF-SR Selections

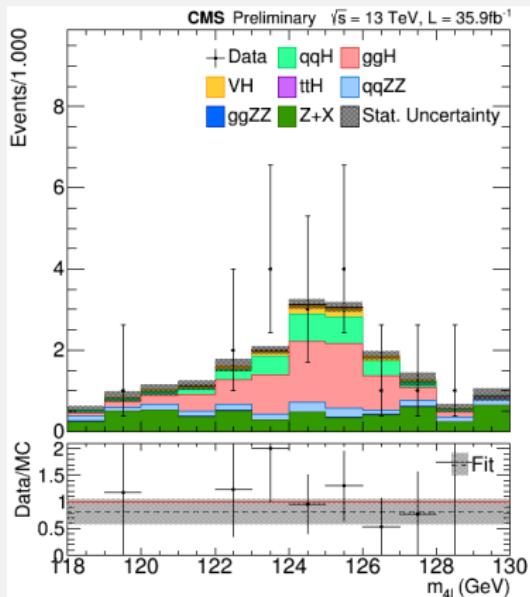


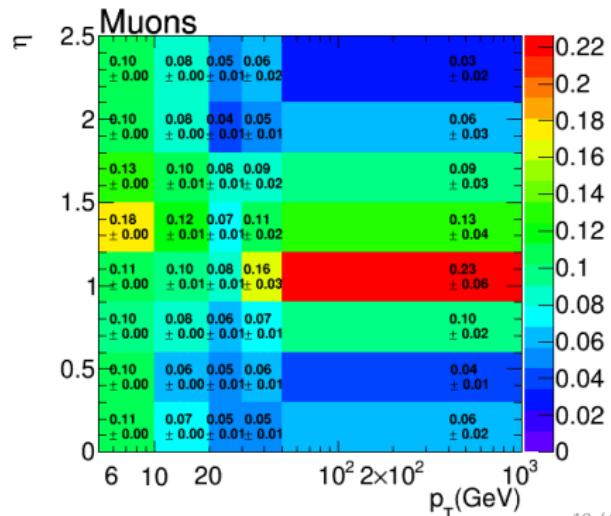
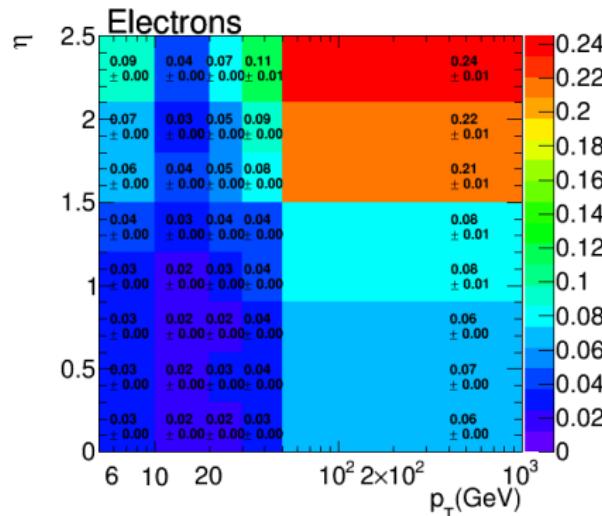
Table 17: Number of expected events from background and signal, with total uncertainty (statistical+systematic), and the number of observed events after the VBF-SR selections (see Sec. 4.3).

Process	4μ	$4e$	$2e2\mu$	$4l$
ggH	2.46 ± 0.53	1.28 ± 0.30	3.10 ± 0.69	6.84 ± 0.92
VH	0.34 ± 0.05	0.20 ± 0.03	0.46 ± 0.07	1.00 ± 0.09
tH	0.05 ± 0.01	0.03 ± 0.01	0.06 ± 0.01	0.14 ± 0.02
qqZZ+ZZJJ	0.67 ± 0.04	0.36 ± 0.04	0.74 ± 0.06	1.77 ± 0.08
ggZZ	0.05 ± 0.01	0.03 ± 0.01	0.06 ± 0.01	0.14 ± 0.02
Z+X	2.03 ± 0.83	0.33 ± 0.04	2.72 ± 0.76	5.08 ± 1.12
Σ backgrounds	5.60 ± 0.99	2.23 ± 0.31	7.14 ± 1.03	14.97 ± 1.46
qqH (signal $m_H = 125\text{GeV}$)	1.05 ± 0.22	0.58 ± 0.13	1.39 ± 0.30	3.02 ± 0.39
Total expected	6.65 ± 1.01	2.81 ± 0.34	8.53 ± 1.07	17.99 ± 1.51
Observed	5	2	10	17

- The MC and Z+X remaining at VBF-SR are the events used in order to train and test the ANN models;

The Z + X Background (Data-Driven) Estimation

- Z + X originates from processes with non-prompt leptons: heavy-flavor meson decays, mis-reconstructed jets and electrons from γ conversions;
- Strategy: measure FR in specific control regions (CRs) and apply it to the SR;
- First step, measuring the FR:
 - samples of $Z l_1 l_2 + l_3$ ($l_{1,2}$ tight leptons, l_3 loose lepton);
 - $p_T^{l_1, l_2} > 10, 20$ GeV, $m_{l_1 l_2} < 4$ GeV, $|m_{l_1 l_2} - m_Z^{\text{PDG}}| < 7$ GeV and $E_T^{\text{miss}} < 25$ GeV;
 - contribution from WZ (with potential 3 real leptons) is subtracted;
- The FR ($N_{\text{tight}} / N_{\text{loose}}$, ie. probability of loose lepton pass tight selections) is mapped wrt. to $p_T^{l_3}$ vs. η^{l_3} :

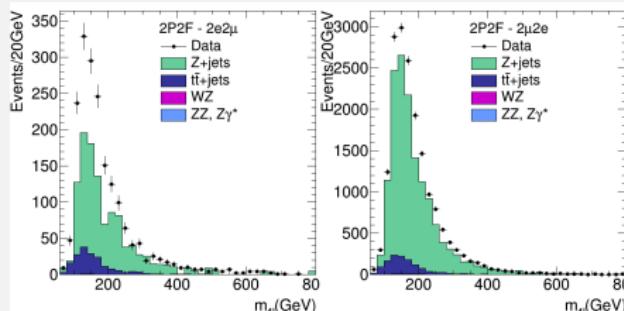
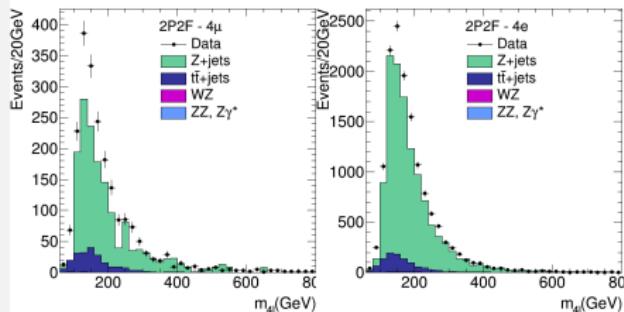


Z + X Background Estimation

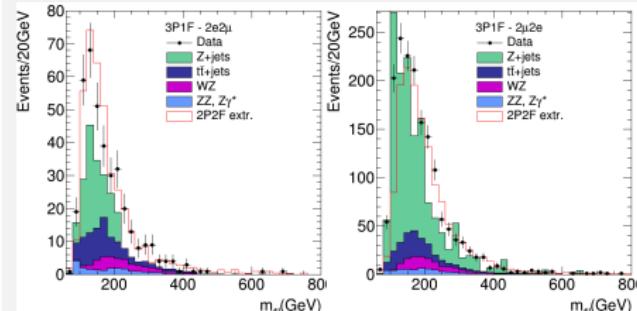
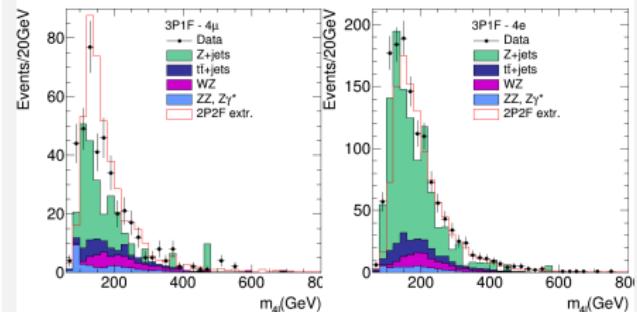
- Second step, building CRs:

- Orthogonal to the SM Higgs selections and enriched by fake-lepton events;
- Require $Z_{l_1 l_2} Z_{l_3 l_4}$ where $l_{1,2}$ are always *tight* leptons and if $l_{3,4}$ are *loose* leptons, define 2P2F while if only l_4 is *loose*, define 3P1F;

$$2\text{P2F CR: } w_{Data} = \frac{f_3}{1-f_3} + \frac{f_4}{1-f_4}$$



$$3\text{P1F CR: } w_{Data} = \frac{f_4}{1-f_4}$$

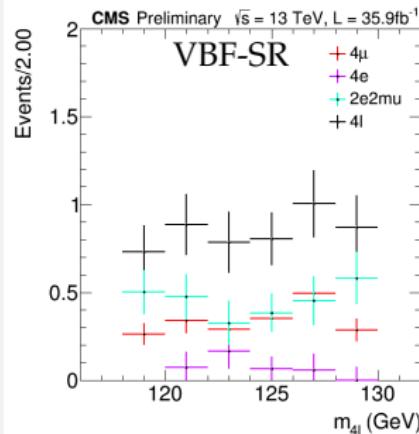
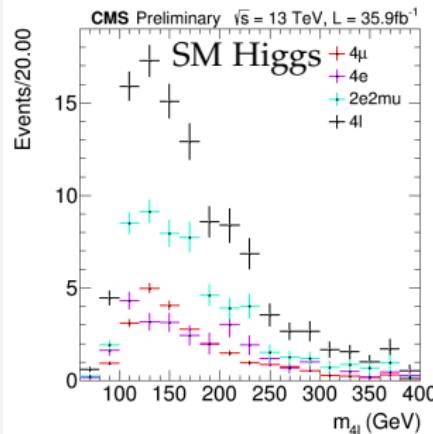


Z + X Background Estimation

- **Third step**, using the measured $\text{FR}(p_T, \eta)$ in order to estimate $Z + X$ yield and shape in the SR;
- At SR, $Z + X$ is given by two components: one from 2P2F and one from 3P1F, via the observed Data and ZZ contribution;
- Events selected in the two CRs were stored for training ANN and derive its $Z + X$ shape;

$Z + X$ at SR:

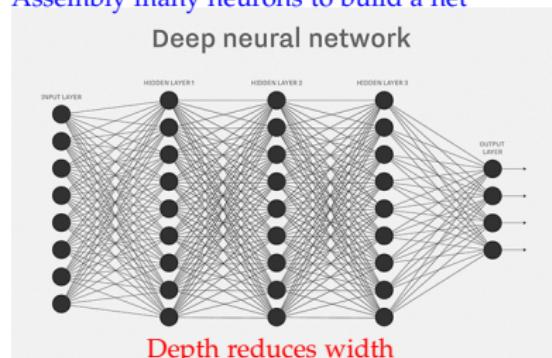
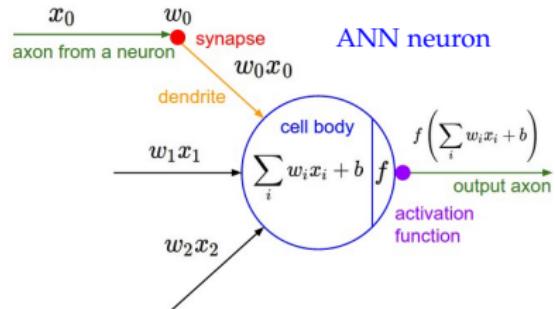
$$N_{SR}^{bkg} = (1 - \frac{N_{3P1F}^{ZZ}}{N_{3P1F}}) \sum_i^{N_{3P1F}} \frac{f_a^i}{(1-f_a^i)} - \sum_j^{N_{2P2F}} \frac{f_b^j}{(1-f_b^j)} \frac{f_c^j}{(1-f_c^j)}$$



Artificial Neural Networks (ANNs)

- An ANN is a composite of multidimensional parameterized functions in terms of weights (w_i) and biases (b_i): $\Omega(x) = \Phi(f_1(w_1, b_1; x), \dots, f_n(w_n, b_n; x))$;

Assembly many neurons to build a net



- It comes from Machine Learning field (closest one to AI's development) in Computer Science;
- ANNs have been successfully applied in many tasks as recognition of handwritten digits, images, sounds, sequences of data, etc;
- Google and Microsoft are investing on them: Deep Mind, Cloud Machine Learning Engine, Cortana, Azure Studio, etc;

ANN Supervised Learning

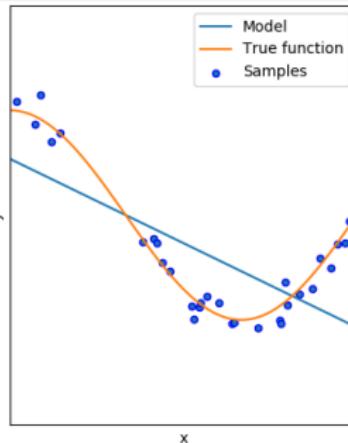
- Training an ANN consists in finding parameters (w_i and b_i) for $\Omega(x)$, such that it models a dataset in the best (possible) way;
- Supervised ANN training:
 - 1 Initialize ANN weights (w_{ji}, b_{ji}) randomly;
 - 2 Input at the ANN an arbitrary set ($\mathbf{x}, \Omega(\mathbf{x})$);
 - 3 Compute the ANN error: (Loss) $\mathcal{E} = f(\Omega(\mathbf{x}) - \Omega(\mathbf{x})_{pred})$;
 - 4 Update the neuron parameters with loss gradient descent: $\Delta w_{ji} = -\eta \frac{\partial \mathcal{E}}{\partial w_{ji}}$;
 - 5 Repeat from 2 to 3 until \mathcal{E} gets smaller enough;
- Here's an ANN learning $f(x) = a.x + b$ (just a linear dataset):



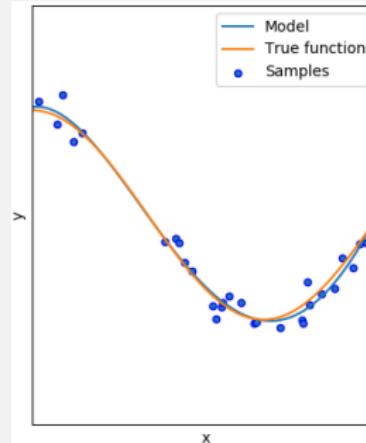
ANN Learning Situations

- An ANN training commonly leads to one of these three main situations:

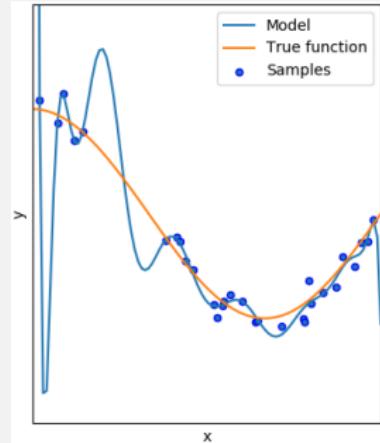
Underfit



Goodfit



Overfit



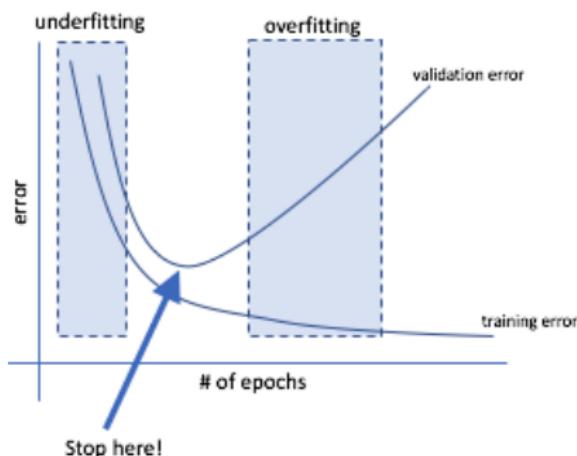
- too few parameters;
- not capable to learn data;
- Solution: more parameters (layers/neurons) or training time;

- enough parameters;
- learns data properly;
- capable of correct predictions for unseen data (generalization);

- too many parameters;
- learns data variance (noise);
- Solution: reduce parameters or training time, dropout, regularizations;

Monitoring the ANN Learning Process

- The previous three cases can be monitored by looking at the so called learning curves;
- One defines at least two sub-datasets:
 - training: which is used for training the ANN (seen data);
 - testing: ANN make predictions on it at every epoch (unseen data);
- This produces the following curves:



- As the model adapts to the seen data, it becomes worse to make correct predictions on new (unseen) data;
- When the validation loss becomes flat or increases while training loss keeps decreasing is a recommended procedure to stop the ANN training in order to avoid the overfitting;

MC Preparation for ANN Training

- For the ANN studies the MCs¹ are prepared in the following way:
- The channels ($4e$, 4μ and $2e2\mu$ selected separately) are merged into just a sample for each MC (the channels are randomized inside the sample);
- Each merged sample are split into 2 independent sets:
 - **Training**: contains 80% of all events (from each sample) and is used to train ANNs;
 - **Testing**: contains remaining 20% of events and is used to test ANN after training;
- Then each set of all MCs are merged (and randomized) to compose the final training/testing input set to train/test ANNs;
- Additionally, two subsets have been defined based on the available number of jets per event:
 - **Njets2**: only events with exactly two jets;
 - **Njets3**: only events with at least three jets;
- ANNs are built and trained via the open-source **Keras**² (standard ML community tool) python package;

¹Data is not used on ANN training.

²<https://keras.io/> (now interfaced with TMVA).

Training Strategy

- It's hard to assure a set of parameters as the best one;
- ANN architecture optimization by scanning over several parameters;

Training parameters (focus on low level variables)

Parameter	Tested options
Inputs	leptons/jets(p_T, η, ϕ), MET
Pre-processing	none, normalization, standardization
Topologies ^a	7:5:3, 21:13:8, 10:10:10:10, 30, 100, ...
Early stop ^b	100, 600, 3000
Minimizer ^c	SGD, Adam, Adagrad, Adadelta, RMSprop
Batch size ^d	1, 5, 32, 64, 128, 786
Neuron	ReLU, SeLU
Loss scaling ^e	XS (process total XS), $\sigma \cdot \epsilon \cdot BR$ (event weight)
Dropout ^f	none, 0.1, 0.3, 0.5, 0.7, 0.9, 0.99, 0.3:0.4:0.2, 0.5:0.25:0.1

^aThat refers to hidden layers. The output layer is always single sigmoid neuron

^bNumber of epochs to stop training if no improvement occurs.

^cMethod used to compute parameters update.

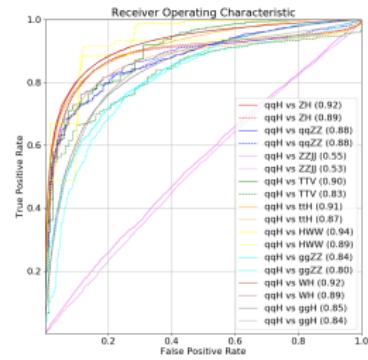
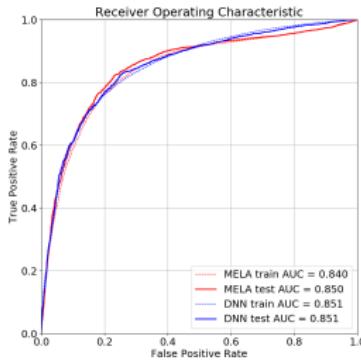
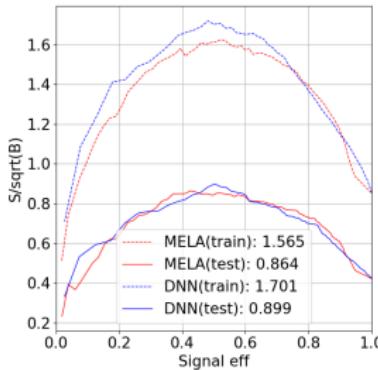
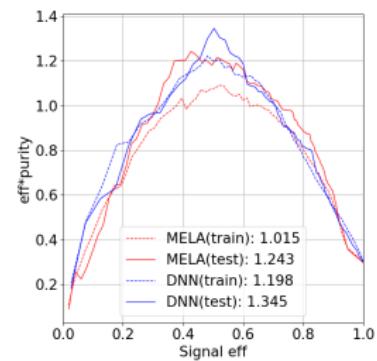
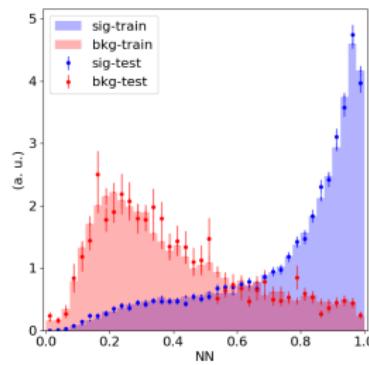
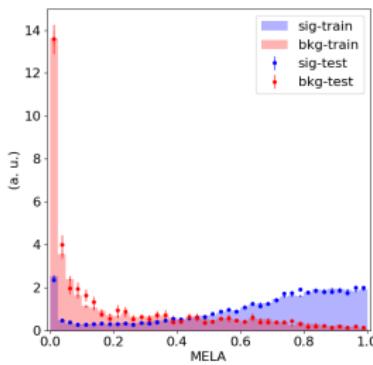
^dSubset from training set used to get parameters updates. N batches = N iterations per epoch.

^eIt's possible to use weights in training to optimize discrimination.

^fFraction of inputs randomly set to zero during training.

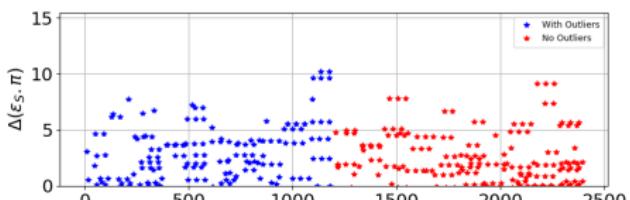
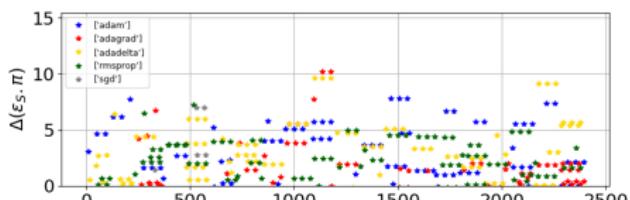
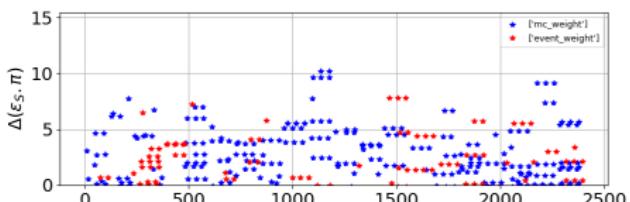
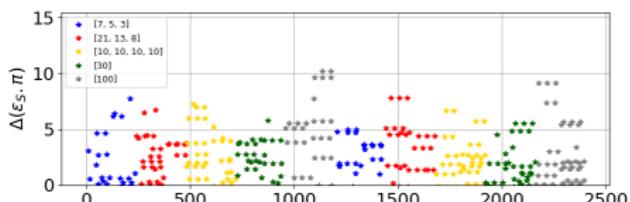
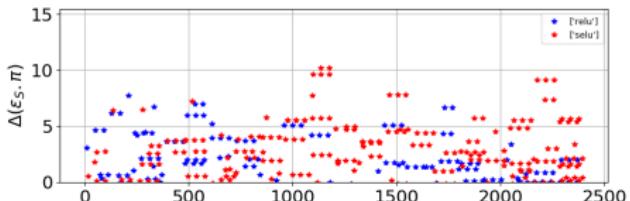
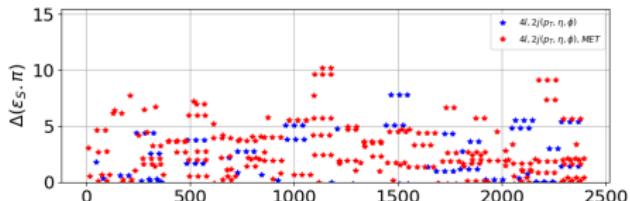
Training Strategy

- Trained ANN models have been validated by checking plots:



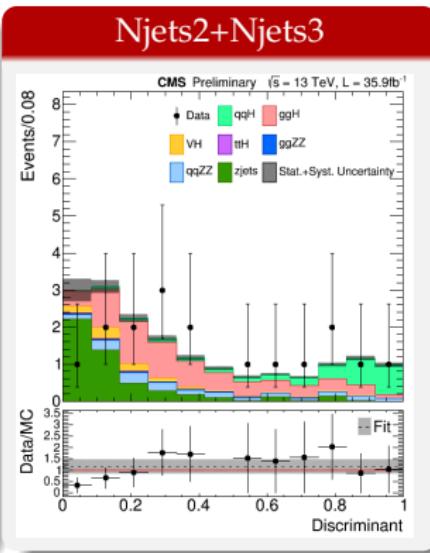
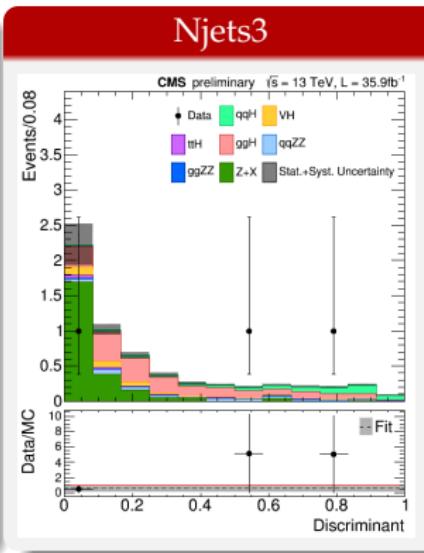
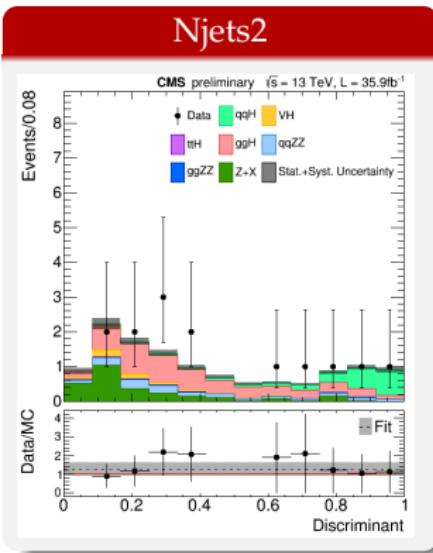
View of Performances for Several ANN Parameters

- Here are some performances of ANN parameters obtained during the analysis;
- This allowed to fix some parameters and make more focused trainings;
- The final ANNs have been chosen based on the metric present in this plots ($\epsilon \cdot \pi$);



Final ANNs for Njets2, Njets3 and their Combination

- Here are the final (best) ANNs, which have been chosen based on the $\max(\epsilon, \pi)$ (ie. efficiency \times purity);



- Z + X shape derived by feeding NNs with observed Data and ZZ MC ³ and, repeating the procedure previously explained for this background estimation;

³Note that, the FR doesn't need to be redone.

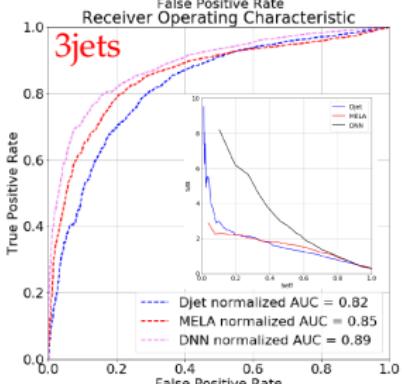
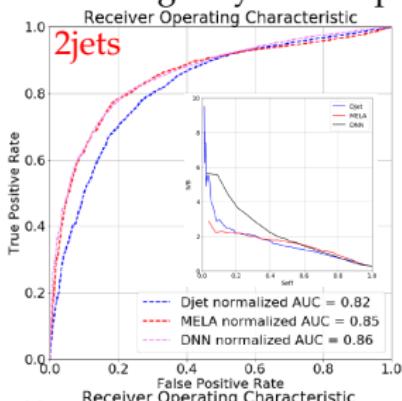
Why to use 3rd Jet?

- VBF topology is commonly tagged by two highly energetic jets with high η separation;
- For the total VBF cross section, though, there are other Feynman diagrams;
- One of them is the case of a 3rd jet in the event which is irradiated from one of the two main jets (tree-level);
- Our studies show the 3rd jet in the events contains valuable information which allows to increase signal/background ratio;

Fraction of events, after the VBF-SR selections, containing the jet_i (i-th jet)

	j_3 (%)	j_4 (%)
qqH	21.3	5.0
ggH	28.8	7.1
$qqZZ$	19.5	2.9
<i>Data</i>	18.0	0.0

VBF discrimination against the backgrounds using only 2 and up to 3 jets



Experimental and Theoretical Systematic Uncertainties

- Experimental and Theoretical systematic uncertainties accounted in this analysis (enter as log-normal nuisance parameters in the statistical analysis):

Experimental Uncertainties

Source	Magnitude (%)
Luminosity	2.6
Lepton $\epsilon_{ID/Reco}$	2.5-9
Lepton energy scale	0.04-0.30
m_{4l} resolution	20
Jet energy scale	3.3
E_T^{miss}	7-26
b-tagging	1
Z + X	6-23

Theoretical Uncertainties

Source	Magnitude (%)
QCD scale (VBF)	+0.4/-0.3
PDF set (VBF)	± 2.1
QCD scale (gg)	± 3.9
PDF set (gg)	± 3.2
Bkg K factor (gg)	± 10.0
QCD scale (WH)	+0.5/-0.7
PDF set (WH)	± 1.9
QCD scale (ZH)	+3.8/-3.1
PDF set (ZH)	± 1.6
QCD scale ($t\bar{t}H$)	+5.8/-9.2
PDF set ($t\bar{t}H$)	± 3.6
QCD scale ($q\bar{q} \rightarrow ZZ$)	+3.2/4.2
PDF set ($q\bar{q} \rightarrow ZZ$)	+3.1/-3.4
Electroweak corrections ($q\bar{q} \rightarrow ZZ$)	± 0.1
$\text{BR}(H \rightarrow ZZ \rightarrow 4l)$	2.0

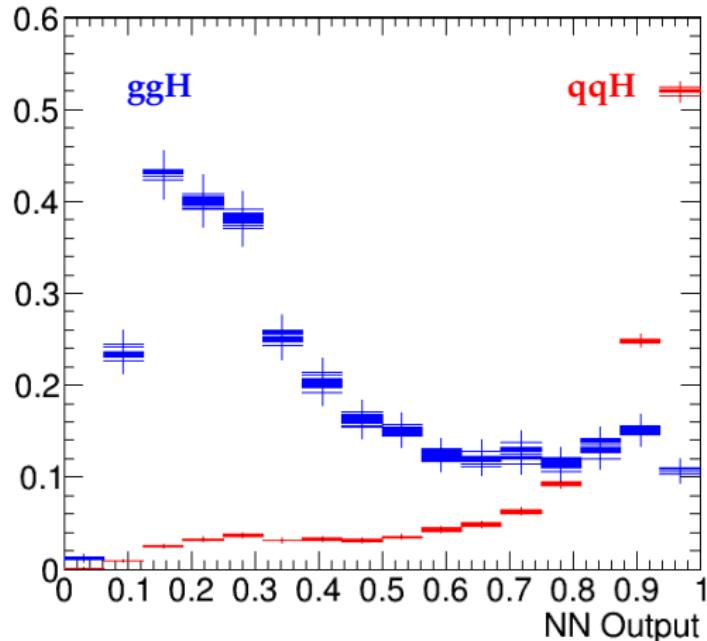
- The systematic uncertainties on the VBF ANN discriminants are added into the statistical analysis via their nominal and shifted shapes;

ANN Discriminants Systematic Uncertainties

- Systematic uncertainties on the ANN discriminants have been estimated from two sources: the systematic uncertainty on their inputs and the systematic uncertainty on the 3rd jet;
- The first case (which affect both jet-based categories) the systematic uncertainty on the ANN discriminants shape and yield has been derived by feeding the discriminants with the $\pm 1\sigma$ shifted value of each input;
- The shifts are produced from one input variable at a time, such that, in the end there are $N_{(Inputs)} \times [1 + 2.N_{(InputsUncertainties)}]$ ANN distributions, including the nominal and shifted shapes;
- This procedure follows similar idea applied in previous CMS analysis:
 - cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2012_141_v9.pdf
 - <https://cds.cern.ch/record/2205282>
 - <https://cds.cern.ch/record/2273847>

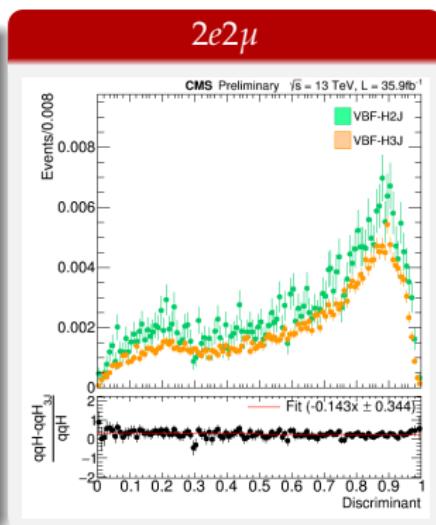
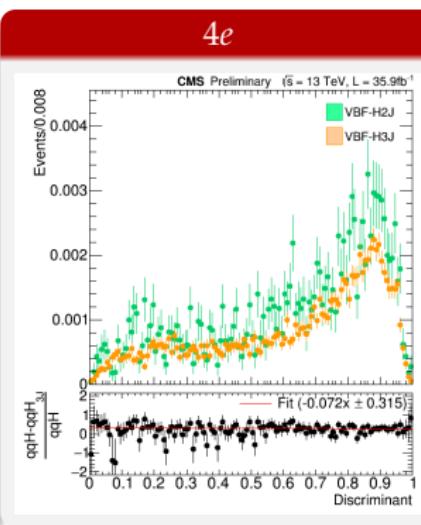
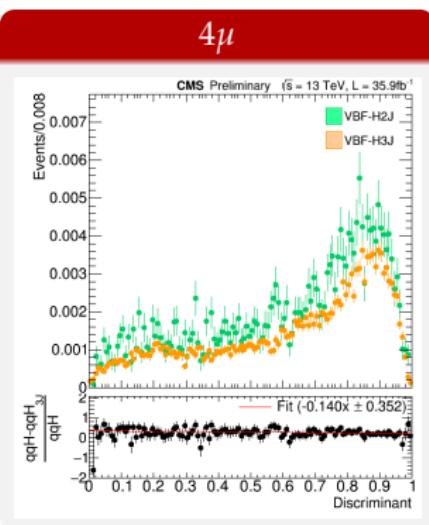
ANN Discriminants Systematic Uncertainties

- Here is an example of nominal and shifted distributions (superimposed) from one ANN for qqH and ggH (largest background) processes;
- The shifts look good and under control (in other words the discriminants are stable), mainly for the signal;



ANN Discriminants Systematic Uncertainties

- Since this analysis is proposing the usage of the 3rd jet and is not possible:
 - to replace the current VBF MC sample (VBF-H2J) by its NLO VBF-H3J version⁴;
 - or merge the two MC samples in suitable way;
- A systematic uncertainty⁵ because of using the 3rd jet from the current VBF MC sample has been estimated by computing the ratio between the ANN distribution using VBF-H2J and VBF-H3J, separately per four-lepton channel:



⁴Process VBF_HJJJ available at PowhegV2 (private generation following 2016 configurations).

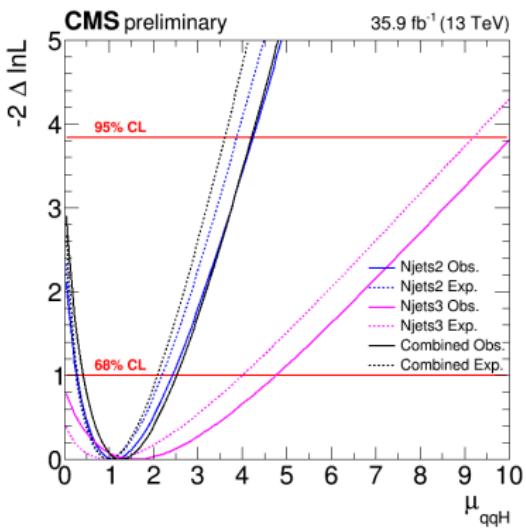
⁵It affects only Njets3 category.

Statistical Analysis

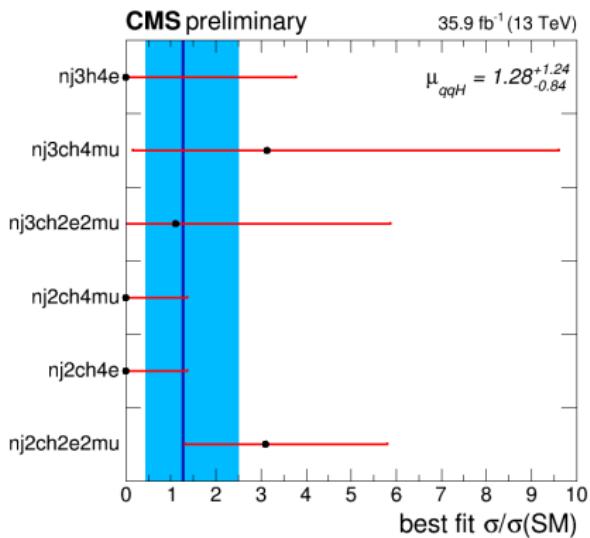
- The statistical analysis was done using the Higgs Combine tool:
 - Higgs Combine tool is a package developed by LHC statistic community;
 - It allows one to compute cross-sections, limits and significances taking into account the statistical and systematic uncertainties;
 - It is used in several analysis in the LHC experiments nowadays;
- Here, a 1D shape analysis has been done:
 - Inputs were ANN discriminants via 1D histograms (with proper signal and background normalizations) and the statistical and systematical uncertainties;
 - Results achieved by combining the ANN discriminants from each jet-based category are highlighted in the next slides;

Statistical Analysis

- The VBF signal strength is measured to be $\mu \equiv \sigma_{qqH}^{Obs}/\sigma_{qqH}^{SM} = 1.28^{+1.24}_{-0.84}$ by combining ANN discriminants of Njets2 and Njets3 categories;



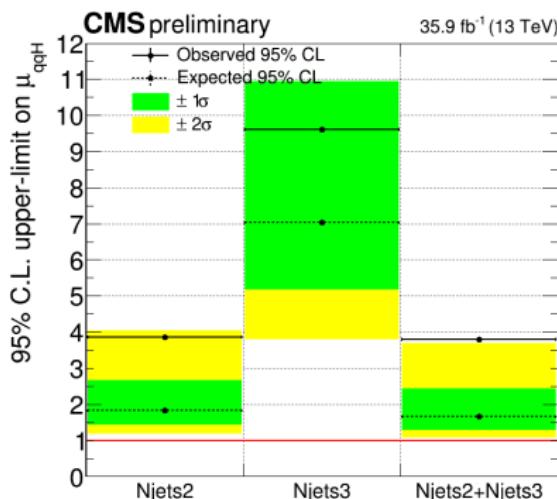
(g) μ_{qqH} likelihood scans.



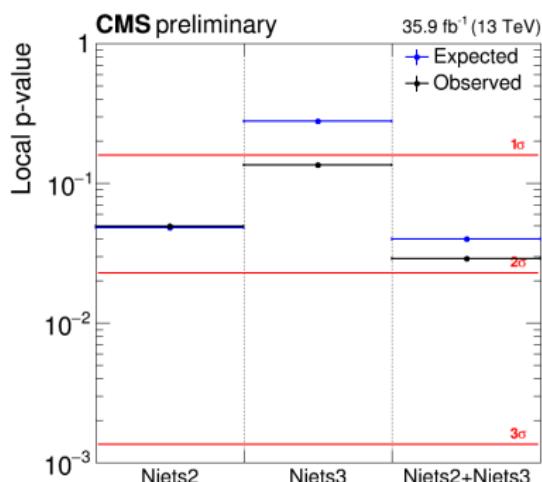
(h) μ_{qqH} best fit on each channel.

Statistical Analysis

- Limits and significances have been computed via the HybridNew method:
 - Limits show that hypothesis of VBF events in the present analysis can't be excluded, setting $\mu_{qqH}^{Obs} < 3.8$ and $\mu_{qqH}^{Exp} < 1.7$ at 95%CL;
 - Significances obtained are $\sigma_{qqH}^{Obs} = 1.9$ and $\sigma_{qqH}^{Exp} = 1.8$;



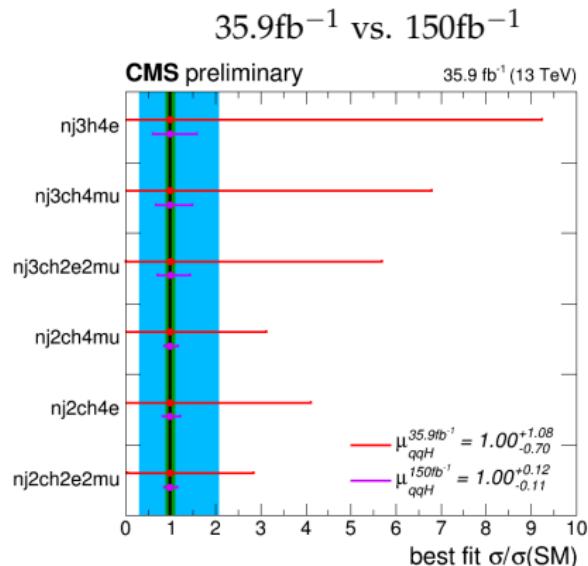
(i) Upper limits on μ_{qqH} .



(j) Significances via the present analysis.

Projections for Future Luminosities

- The VBF signal strength measurement and the significance of the present analysis has been projected for future luminosities scenarios at the LHC;
- Systematic uncertainties have been accounted by scaling the present luminosity;
- The total uncertainty on the measurement of μ_{qqH} is expected to reduce $\sim 87\%$ at $L = 150\text{fb}^{-1}$;



- Significance of 5.1σ is expected for a luminosity 10x larger than the present one. Below is a scan of the expected VBF significance between the present luminosity (35.9fb^{-1}) and future scenarios at the LHC.

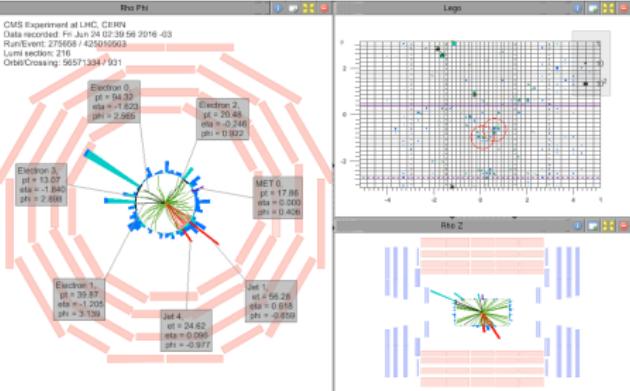
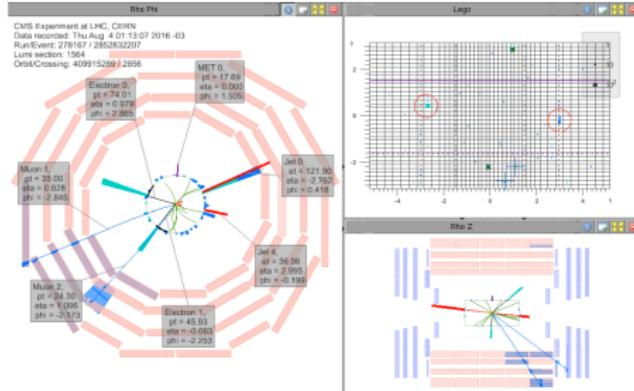
Luminosity (fb^{-1})	35.9	150.0	300.0	359.0	1077.0	1795.0	3000.0
Factor ($\times L^{35.9\text{fb}^{-1}}$)	1.00	4.18	8.36	10.00	30.00	50.00	83.57
Expected significance	1.8	3.4	4.7	5.1	8.6	10.9	14.0

Most & Least VBF-Like Event Display in each Category

ANN score: 0.93

Njets2 Category

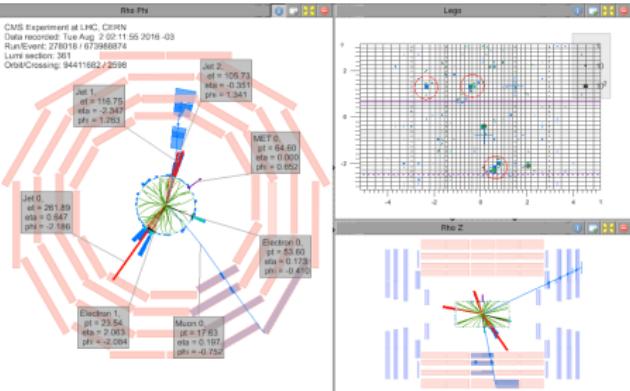
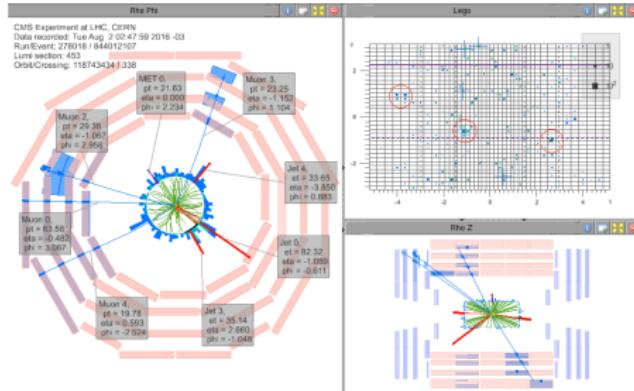
ANN score: 0.15



ANN score: 0.77

Njets3 Category

ANN score: 0.04



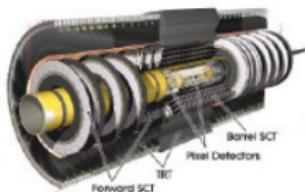
CMS Level 1 Tracking Trigger Associative Memory + FPGA

Introduction

- During his PhD the author was involved into the CMS Level 1 Tracking Trigger AM+FPGA project:
 - one of the three CMS-L1TT projects (CMS upgrade phase II - HLLHC):
 - Inclusion of inner-tracker data as part of L1 trigger;
 - Original tracker designed for $L_{inst.} \sim 10^{34} \text{ cm}^{-2} \cdot \text{s}^{-1}$ and $\text{PU}_{Ave.} \sim 20\text{-}30$;
 - Expected in phase II: $L_{inst.} \sim 10^{34} \text{ cm}^{-2} \cdot \text{s}^{-1}$ and $\text{PU}_{Ave.} \sim 140\text{-}200$;
 - Required decision time: $5\mu\text{s}$;
 - 500-1k Tb/s of data to be processed;
 - led by Fermilab working group;
 - the project aims for the usage of Associative Memories in combination with FPGAs;
 - the author studied and implemented new components on the available software:
 - synthetic match;
 - duplicate removal;
 - stub bending;
 - road and combination truncation
 - track fitter χ^2 adjustment;
 - results have been produced with different high-lumi scenarios (2-10k events):
 - $(\mu/\pi/e) + \text{PU}(140, 200, 300, 400)$;
 - $\nu + \text{PU}(140, 200, 250)$ (simulates pure PU, low p_T particles);
 - $t\bar{t} + \text{PU}200$;
 - jets($p_T = 250\text{GeV}$) + PU200;
 - hardware work: board inspections and tests;

The CMS-L1TT AM+FPGA Approach

Tracker Detector



Data transfer

**Data
formatting**

Partition detector into
trigger towers/sectors

Pick your favorite method:

*Associative Memory (AM) Approach
(proven approach from CDF/SVT)*

*Hough Transformation
tracklet-based*

Adaptive Pattern Recognition

*Biology Inspired ...
your choice here...*

**Pattern
Recognition**

Finer pattern recognition

**Track
Fitting**

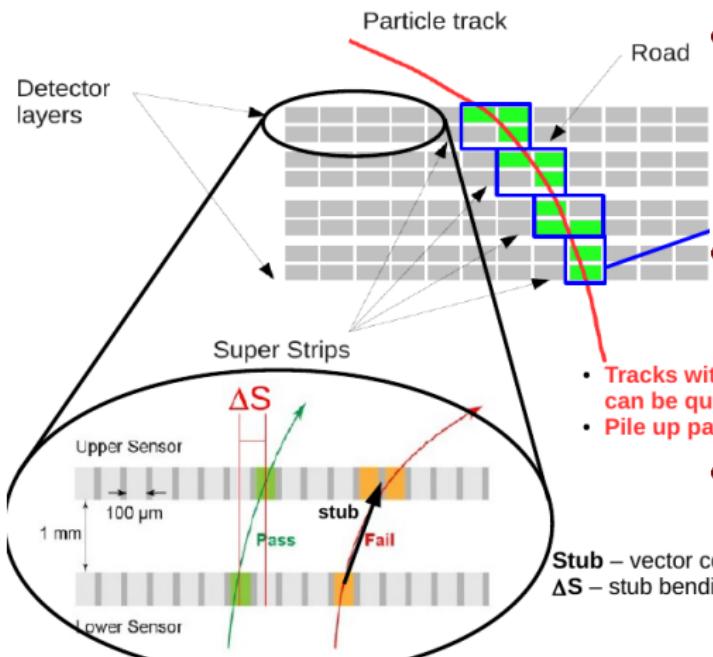
- The three CMS-L1TT projects can be divided into three stages:

- Data Formatting:** fragmentation of the CMS detector in $\eta - \phi$ sectors (trigger towers);
- Pattern Recognition:** selection of coarse hits patterns (potentially real tracks);
- Track Fitting:** extraction of refined track info using all hits from selected patterns;

The CMS-L1TT AM+FPGA Approach

- Here's the main idea and definitions adopted in the CMS L1TT AM+FPGA approach:

- Superstrip (SS):** cluster of hits in the detector layers. They receive an ID based on their $z - \phi$ position;
- Road:** pattern of built from SS's;



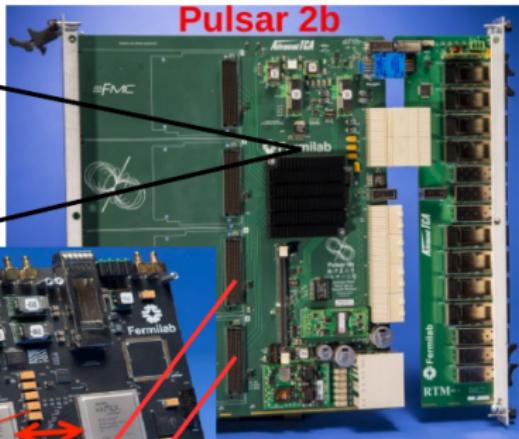
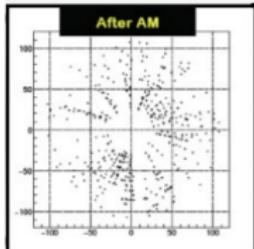
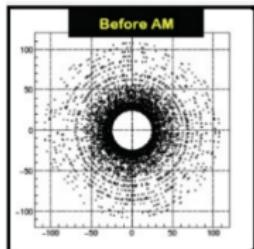
- An AM chip containing a big set of roads simulated via MC triggers the roads observed in real data. The hits from detector layers are processed in parallel;
- Once real roads are triggered, a set of possible hits combinations are built (possible tracks);
 - Tracks with pT below a threshold can be quickly rejected;
 - Pile up particles most $\text{pT} \leq 2\text{GeV}$
- A fit select which combination is a real tracker;

Stub – vector connecting 2 sub-layers

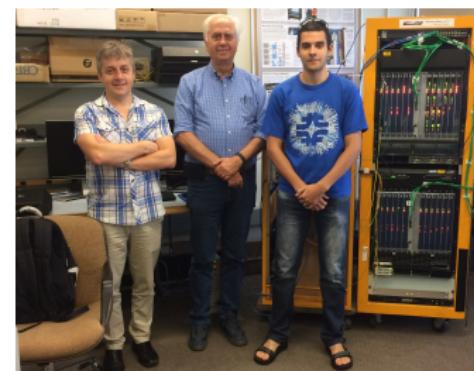
ΔS – stub bending (half strip cluster distance)

The Hardware for the CMS-L1TT AM+FPGA

Input: all hits from a collision



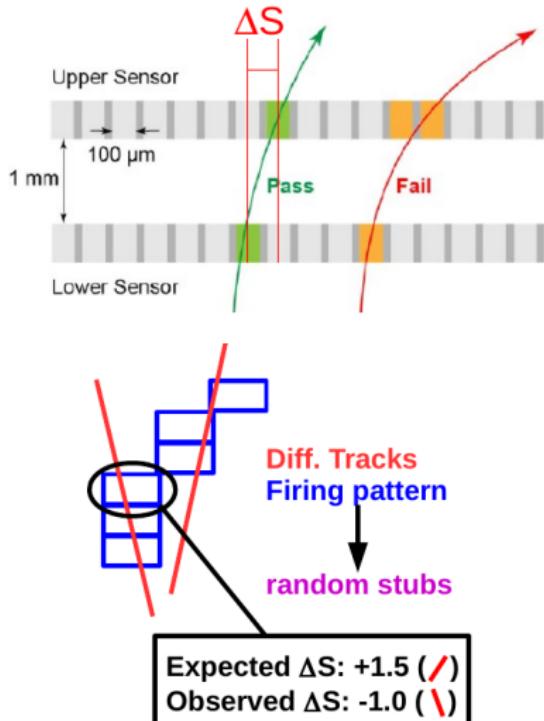
Output: useful tracks



Simulation Studies: Stub Bending (ΔS)

- The stub bending is the core idea behind the CMS L1TT project:
 - It helps to mitigate PU (mainly low p_T particles);
- In the AM+FPGA approach the ΔS prevents random patterns to be fired:
 - Without ΔS an AM pattern can be triggered by hits coming from different real tracks crossing the detector layers in different angles;
 - The ΔS was encoded in the AM framework via the SS ID's. The following formula defines the SS ID when the stub bending is required:

$$ss = i_{\Delta S} * N_\phi + i_\phi$$



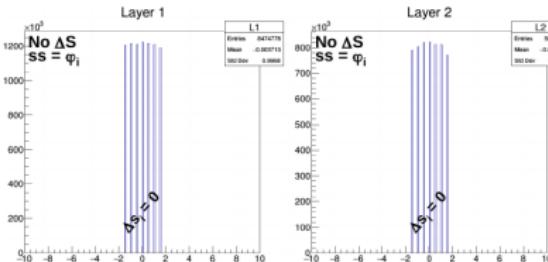
Simulation Studies: Stub Bending (ΔS)

- The SS- ΔS formula:

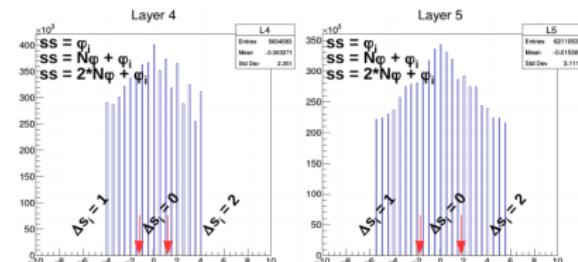
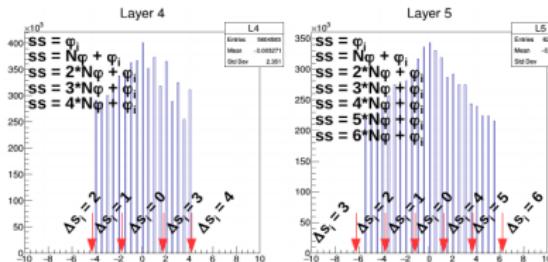
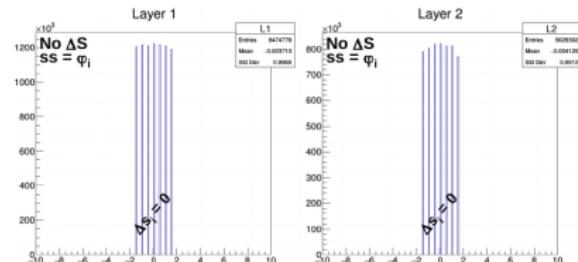
$$ss = i_{\Delta S} * N_{\phi} + i_{\phi}$$

- $i_{\Delta S}$: ΔS value of a given stub (max);
- N_{ϕ} : number of trigger-tower segmentations in ϕ ;
- i_{ϕ} : index of the ϕ segment which the stub belongs;
- Two possibilities of building the SS ID's according to the ΔS values:

Symmetric (eg. SYM115577)



Asymmetric (eg. ASYM115577)

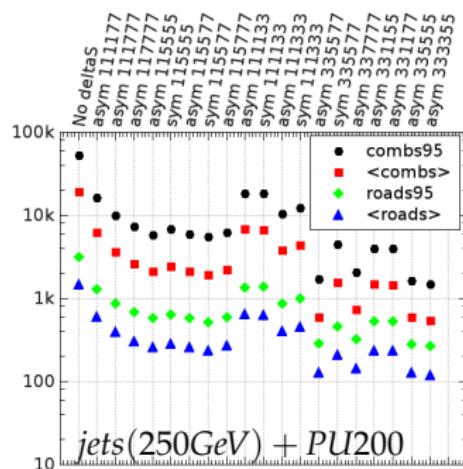
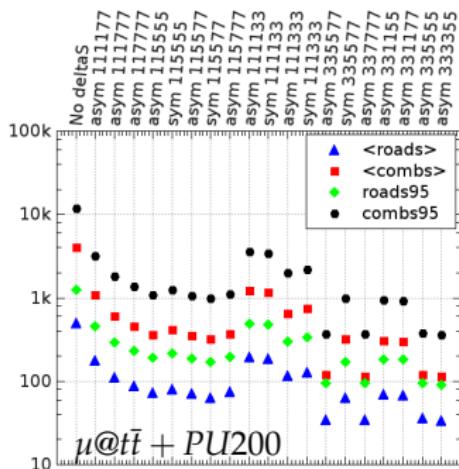


Simulation Studies: Stub Bending (ΔS)

- This ΔS approach allows the following schemes (negative ranges omitted):

#ranges	range width	ΔS values ([] central ranges)
3	9	[-2.0, 2.0], [2.5, ...]
5	7	[-1.5, 1.5], [2.0, 5.5], [6.0, ...]
7	5	[-1.0, 1.0], [1.5, 3.5], [4.0, 6.0], [6.5, ...]
9	3	[-0.5, 0.5], [1.0, 2.0], [2.5, 3.5], [4.0, 5.0], [5.5, ...]

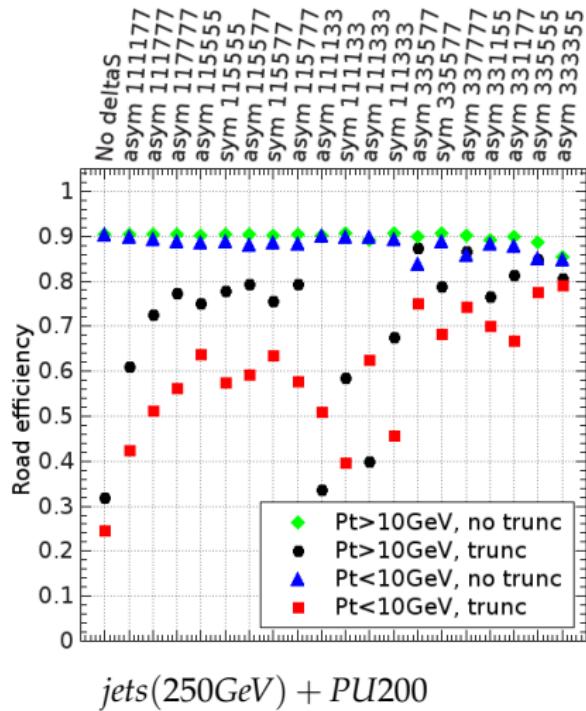
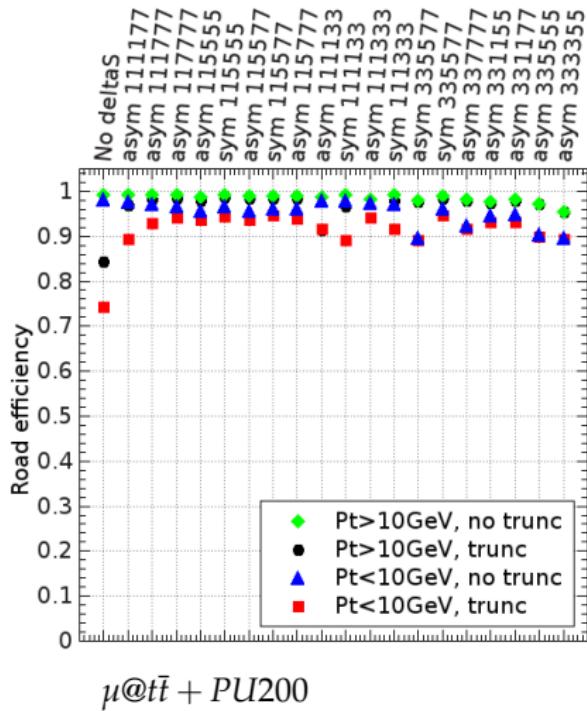
- Effect of ΔS on the number of roads and combinations:
 - Reduction of up to $\sim 10x$ on roads and $\sim 50x$ on combinations;
 - Symmetric method produces few more combs/roads than Asymmetric one;



Simulation Studies: Stub Bending (ΔS)

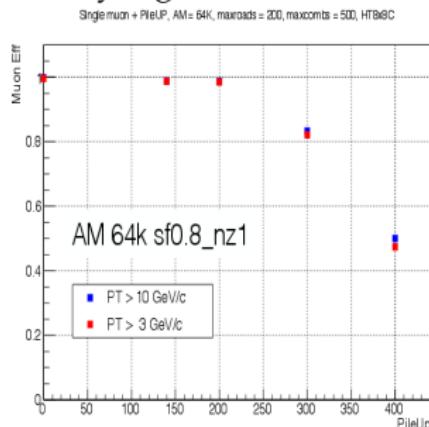
- Effect of ΔS on the road efficiency:

- Up to 20% and 50% of efficiency can be recovered when truncation is applied;

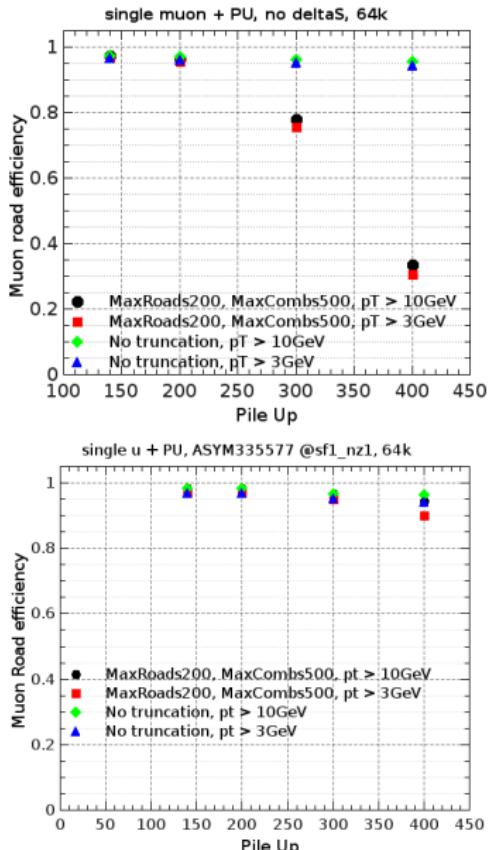


Simulation Studies: Stub Bending (ΔS) - The Edge of the Mountain

- At the ending of author's iteration with the CMS L1TT AM+FPGA there was a worry about PU spikes (as it happened in LHC Run I);
- Studies with single- μ +PU presented in the group showed large efficiency loss at very high PU:



- The author decided to check that and apply the ΔS approach (not considered by them at that time):

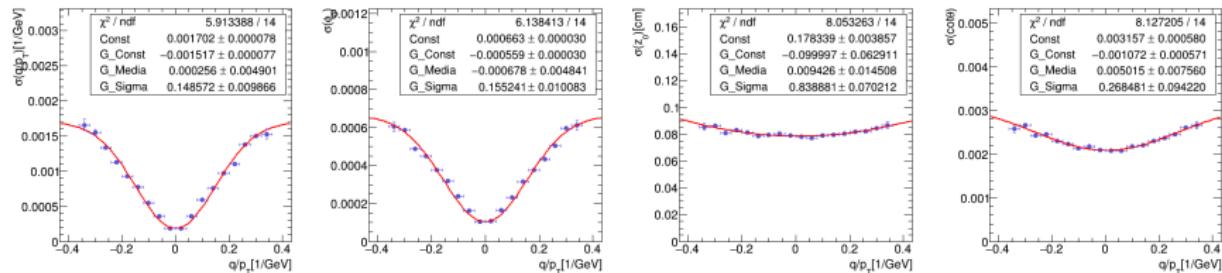


Simulation Studies: Synthetic Efficiency

- Synthetic efficiency is meant to check the efficiency based on the track parameters (q/p_T , ϕ_0 , z_0 and $\cot \theta$);
- Task: match MC and AM reco tracks using their parameters;
- For so, one defines a χ^2 -like function:

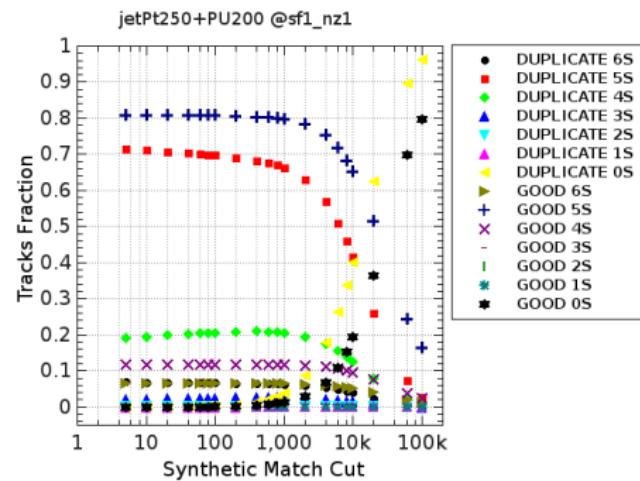
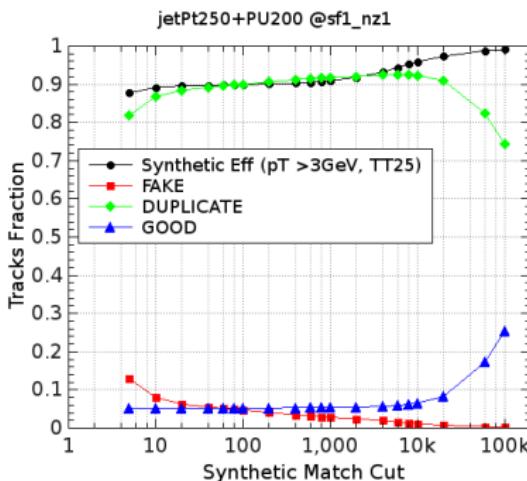
$$\chi^2_{match} = \sum_{i=0}^4 \frac{\delta^2 p_i}{\Omega^2(q/p_T)_i}, \quad \delta p_i = (p_i^{MC} - p_i^{Reco})$$

- The Ω function normalizes the dependence between the resolution in each track parameter and q/p_T :



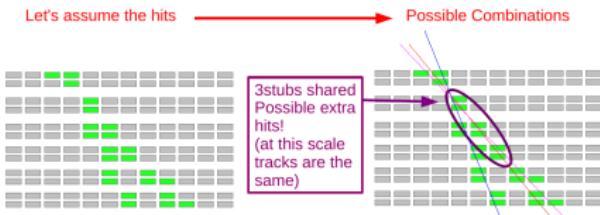
Simulation Studies: Synthetic Efficiency

- The χ_{match} establishes three types of tracks when receives a threshold value \bar{q} :
 - Good:** first reco track with smallest $\chi_{match} < \bar{q}$;
 - Duplicate:** other reco tracks with $\chi_{match} < \bar{q}$;
 - Fake:** any reco track with $\chi_{match} \geq \bar{q}$;
- In order to define the value \bar{q} a tracking match based on stubs and the synthetic efficiency were simultaneously done:
 - Scanning the cut on χ_{match} one checks (on a dedicated MC sample) the number of original stubs composing the reco track;
 - Then, one checks which cut reduces the Fake rate and increases the Good rate as much as possible and, avoiding random stub combinations (GOOD <5S);
 - It was decided to have $\chi_{match} = 40$;



Simulation Studies: Duplicate Removal

- The pattern match based on SS clusters produces several duplicate tracks in the CMS L1TT AM+FPGA;
- For that reason, a procedure to remove such tracks has been developed: the duplicate removal;
- The duplicate removal is a stub-based mechanism with the following algorithm:
 - 1 A reco track is taken from the reco tracks list (**A**) and inserted on a new tracks list (**B**);
 - 2 Then, a loop is done over the remaining tracks on list **A**:
 - If a track is found to share a given number n of stubs with any track on the list **B**, it is removed from list **A**;
 - Otherwise, the track is stored into the list **B**;
 - 3 The tracks remaining in the list **B** are the final tracks;
- The DR mechanism was studied in order to tune the minimum number of stubs which allows massive remotion of duplicated tracks and high synthetic efficiency;



Simulation Studies: Duplicate Removal

- Here are some results obtained for the DR tuning⁶;
- The final DR cut was chosen to be 0 (zero);

$\mu + PU200$

DR option	Goods	Duplicates	Fakes	Track eff	Synthetic eff
None	1.976	25.785	0.614	0.985	0.989
5	1.976	25.785	0.614	0.985	0.989
4	1.976	8.898	0.275	0.98	0.989
3	1.973	0.604	0.095	0.964	0.989
2	1.969	0.065	0.047	0.953	0.989
1	1.967	0.007	0.039	0.951	0.988
0	1.966	0.000	0.038	0.951	0.988

$jet(p_T = 250GeV) + PU200$

DR option	Goods	Duplicates	Fakes	Track eff	Synthetic eff
None	8.506	143.735	8.924	0.89	0.897
5	8.506	143.735	8.924	0.89	0.897
4	8.506	52.935	4.109	0.883	0.897
3	8.481	4.746	1.167	0.823	0.895
2	8.431	0.642	0.597	0.754	0.889
1	8.412	0.067	0.506	0.738	0.887
0	8.406	0.003	0.482	0.74	0.886

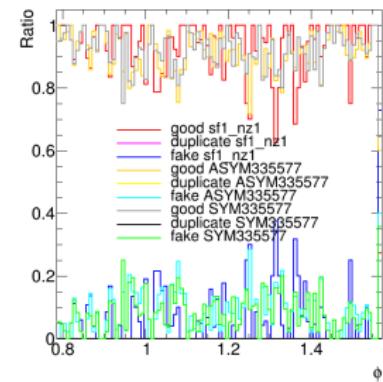
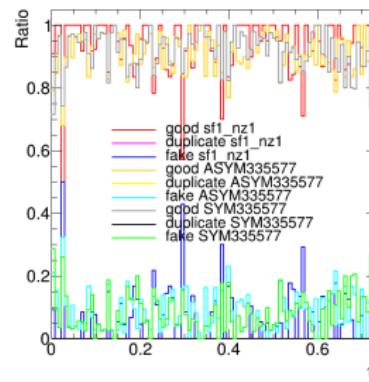
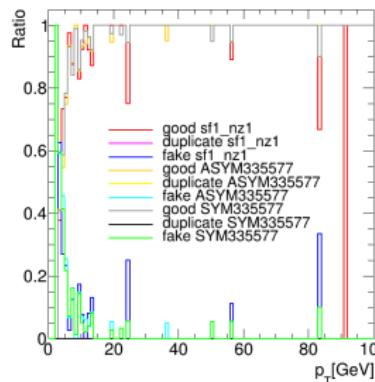
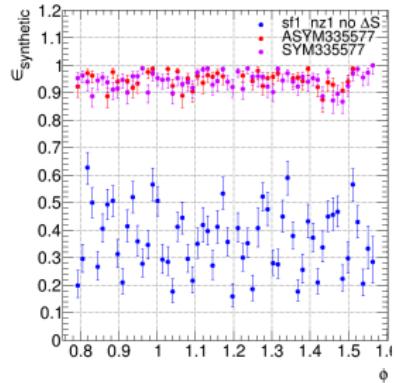
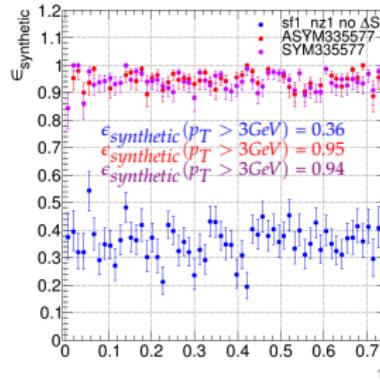
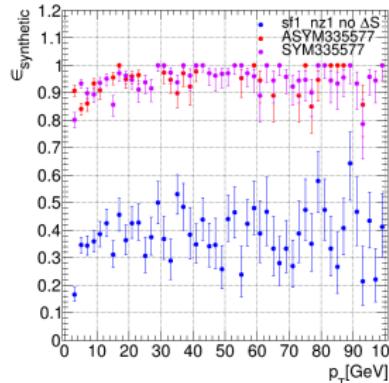
⁶ Notice the gap between the track and synthetic efficiencies: that comes from the extra stubs which builds up a good (stubs) combination for the original track

Simulation Studies: Final FOMs

- The next slides summarizes the final results found via the simulation package adopting the implementations presented here;
- The FOMs (figures of merit) are the common graphs used within the CMS L1TT AM+FPGA approach in order to show the performance of simulation studies;
- The FOMs are the efficiency and track categorization rates versus (p_T, η, ϕ) ;

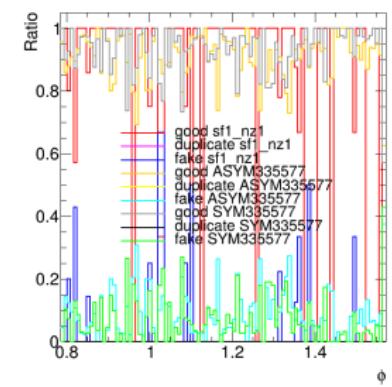
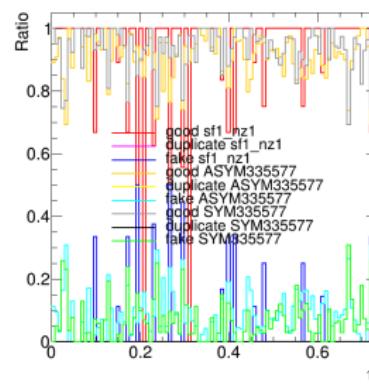
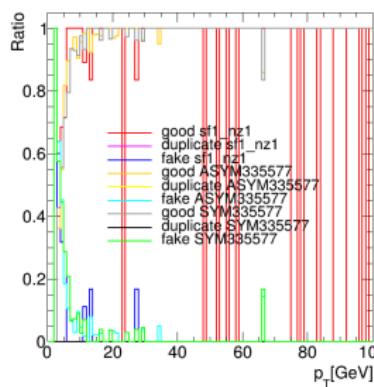
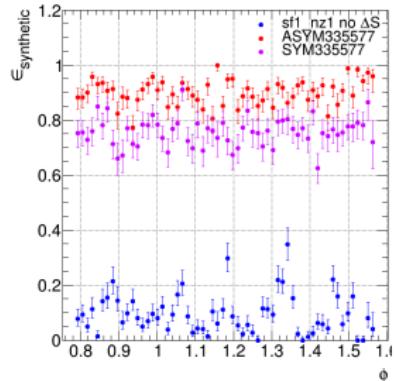
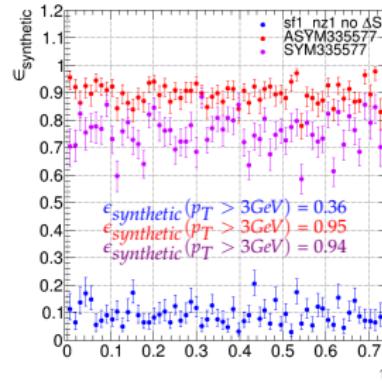
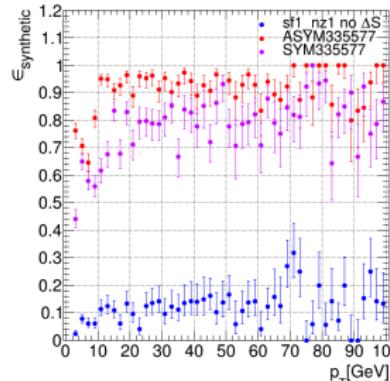
Simulation Studies: Final FOMs - $\mu + PU300$

Track reconstruction efficiency for $\mu + PU300$ sample. The pattern bank used had 64k patterns and truncation at 200 roads and 500 combinations has been applied.
 Duplication removal was applied by requiring DR=0.



Simulation Studies: Final FOMs - $\mu + PU400$

Track reconstruction efficiency for $\mu + PU400$ sample. The pattern bank used had 64k patterns and truncation at 200 roads and 500 combinations has been applied.
 Duplication removal was applied by requiring DR=0.



A Fast Matrix Element

Theoretical Foundation

- Here is presented a procedure called *Fast Matrix Element (FastME)*;
- It was studied in the very beginning of the author's PhD;
- The project was first idealized by prof(s). Andre Sznajder (DFNAE-UERJ) and Stephen Mrenna (CSD - FNAL):
 - A method capable of deriving event weight from MC sampling into a given phase space;
 - It should allow one to get proper normalization of random events, for instance;

Why Fast?

Time to compute the weight per event via *MadWeight5*. For an usual analysis these numbers multiply by thousand.

Process	Time/Event (s)
ZH	<5
t̄t fully-leptonic	10
Zbb	18
t̄t semi-leptonic	41
t̄tH fully-leptonic	60

ME Methods

$$\mathcal{P}(x|\alpha) = \frac{1}{\sigma_\alpha} \int d\omega_1 d\omega_2 f(\omega_1) f(\omega_2) \int d\Phi(y) |\mathcal{M}_\alpha(y)|^2 W(x,y)$$

- $\mathcal{P}(x|\alpha)$ is an event probability;
- $W(x,y)$ handled as approximation;
- $\mathcal{M}_\alpha(y)$ not possible for all physics;
- $\mathcal{M}_\alpha(y)$ in NLO or so on?

Theoretical Foundation

- The original idea of finding event weights didn't lead to promising results: assigned weights didn't model properly the events;
- A new idea appeared, then:
 - Would it be possible to discriminate events based on a match between the particles from a probe event and a MC one?
- *FastME* algorithm:
 - 1 Loop over the particles (i) from a MC event and match them to the particles (j) from a probe event according to

$$R_{(i,j)}^2 = \sum_{k=1}^n \left(\frac{v_k^{(i,MC)} - v_k^{(j,Data)}}{\sigma_{v_k}} \right)^2$$

where, k stands for the kinematic variables (p_T , η , ϕ) and the particles pairs (i, j) are chosen to minimize R_{ij} ;

- 2 A distance between the probe event and the MC event is computed by summing in quadrature the minimum distances (R_{ij}) between their particles:

$$D^2 = \sum_{i=1}^m [R_{(i,j)}^2]_{min}, \text{ with } j(i+1) \neq j(i)$$

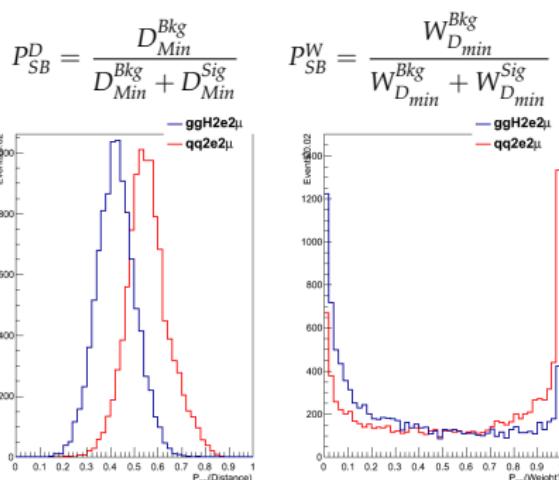
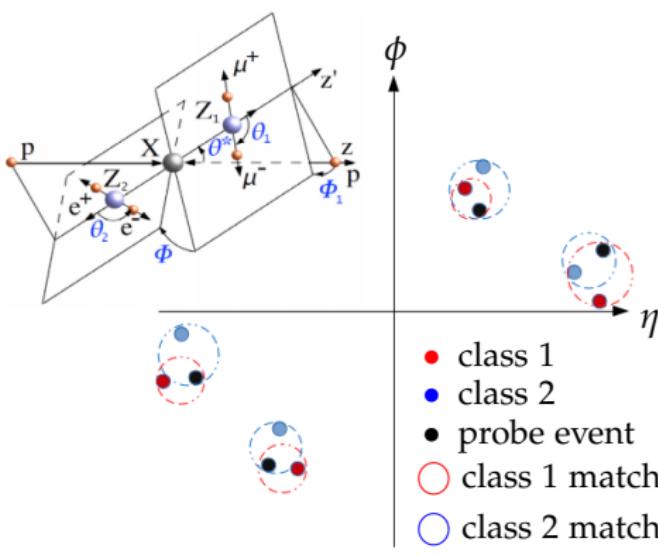
- 3 Finally, a discriminant for the probe event is computed using the closest MC events (from each class) via the formulas

$$P_{SB}^D = \frac{D_{Min}^{Bkg}}{D_{Min}^{Bkg} + D_{Min}^{Sig}}$$

$$P_{SB}^W = \frac{W_{D_{min}}^{Bkg}}{W_{D_{min}}^{Bkg} + W_{D_{min}}^{Sig}}$$

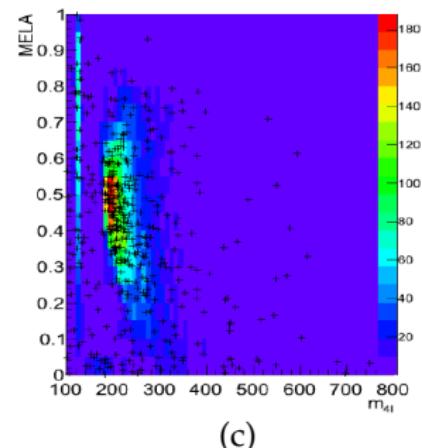
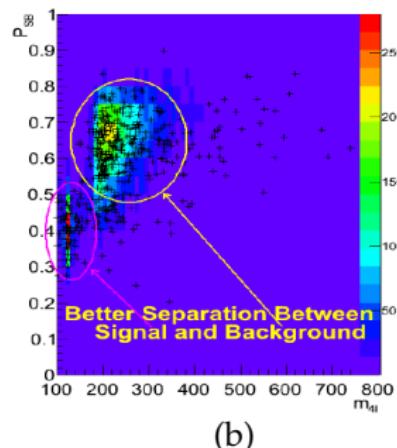
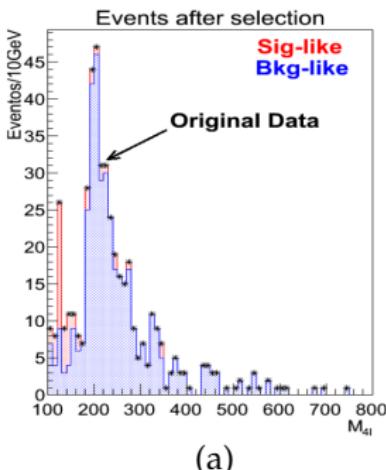
Theoretical Foundation

- Here's an illustration of the method to clarify the algorithm. The MC events present a topology associated to the EM of a given physical process, such as, each particle has a correlation with the others particles in the event. A data event (black points) receives a probability of being from a kind or other (blue and red point) via the correlation of the distances (represented by the blue and red circles) between it and the MC events.



Simulation Studies: *FastME* Applied to HZZ4L CMS Data

- After interesting results with MadGraph/Powheg/Sherpa samples of ggH and qqZZ, it was natural an interest of applying *FastME* to the HZZ4L data collected by CMS on 2015 during the LHC RunI;
- The results have been compared to the formal CMS discriminant, the so called MELA for discriminating SM Higgs against $q\bar{q}ZZ$ background;
- Below: (a) observed events classified as signal and background by the *FastME*, (b) P_{SB}^D and (c) MELA discriminants distribution versus the m_{4l} .



The *FastME* Package

- The success of *FastME* idea on real CMS data encouraged us to move the standalone codes created until that moment into a organized package;
- During the author's first travel to Fermilab, this package started to be maintained on GitHub:

A screenshot of a GitHub repository page for the `FastMatrixElement` project. The repository has 14 commits from `mmeledea`. The commits are:

File	Message	Time
<code>FastMatrixElement</code>	Updating composer structure	9 months ago
<code>tests</code>	Example of root file from event generation feature	10 months ago
<code>README.md</code>	Update README	2 months ago
<code>SetUpEnvironment.sh</code>	Automatic setup	a year ago

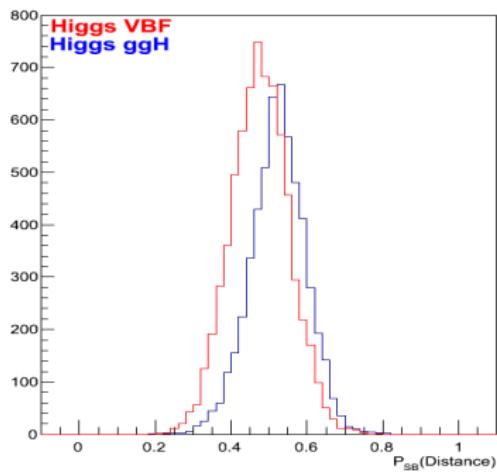
The repository page also displays the `README.md` file content, which includes the *FastME* logo and a brief description of the project's goal.

FastME
Probing and Discriminating Data by MC Topology (stable & updated version of the project)

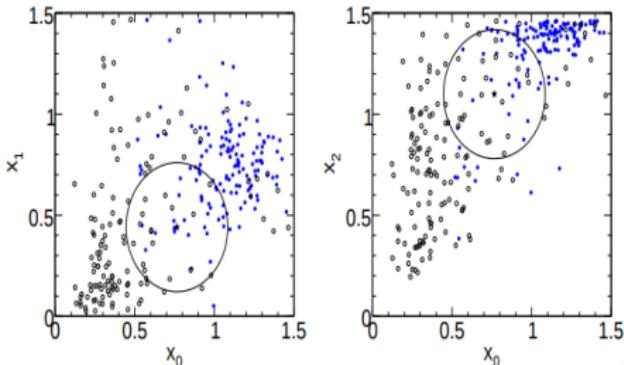
Fast Matrix Element main goal is to achieve similar capability and full physic information as the many types of Matrix Element method available today. However, FastME try to reduce the computing time needed to get such info. The Monte Carlo events already contain the full physic information (that can be in higher orders of parameters correction) about the related event and FastME uses these events as inputs to compute in a fast way the nature (signal/background) of a specific event (a data event, for instance). The process relies into compare de available parameters (pT and η) between the event and the MC. Such comparison allows one to compute the distance between an event and each MC in that phase space.

The End of *FastME* Project

- Although the modified original idea of *FastME* has showed some nice results we faced two issues:
 - It doesn't have a good performance for discriminating VBF against ggH;
 - TMVA has a similar implementation and is more flexible;



8.5 k-Nearest Neighbour (k-NN) Classifier



- Such issues lead to the end of the project:
 - No power enough for the presented physics analysis;
 - Not a new method (we re-invented the wheel);

CONCLUSIONS

Conclusions: CMS-AN-18-120

- ANN approach successfully implemented for an isolated VBF $H \rightarrow ZZ \rightarrow 4l$ XS measurement;
- Reliable procedure implemented for systematic uncertainties;
- Results provided from the combination of our best ANN configurations:
 - best fit for signal strength: $\mu_{qqH}^{Exp} = 1.00^{+1.08}_{-0.70}$ and $\mu_{qqH}^{Obs} = 1.28^{+1.24}_{-0.84}$;
 - 95%CL upper limits on μ_{qqH} : $\mu_{qqH}^{Exp} < 1.66$ and $\mu_{qqH}^{Obs} < 3.79$;
 - significances: $\sigma_{qqH}^{Exp} = 1.8$ and $\sigma_{qqH}^{Obs} = 1.9$;
- Projections provided for future luminosity scenarios at the LHC:
 - Expected to improve signal strength precision up to ~87% at the end of RunII;
 - Significance evolution:

Luminosity (fb^{-1})	35.9	150.0	300.0	359.0	1077.0	1795.0	3000.0
Factor	1.00	4.18	8.36	10.00	30.00	50.00	83.57
Expected significance	1.8	3.4	4.7	5.1	8.6	10.9	14.0
- Analysis documentation is ready and released: [AN-18-120](#);
- No issues raised last meeting with HZZ subgroup (December 7th, 2018) at CERN (green light for a thesis endorsement);

Conclusions: CMS-AN-18-120

- **Ongoing:** analysis of full [2017 MC & Data](#):

- package of macros for future studies (parallel ANN training, Z+X derivation and statistical analysis);
- already in use by a colleague in Bari (Nicola's student).

8 commits 1 branch 0 releases 1 contributor

Branch: [master](#) [New pull request](#) [Find file](#) [Clone or download](#)

File	Commit Message	Date
 mmelodea	updating guide for setting up keras at RECAS	Latest commit c37e05d 14 days ago
 Guide	updating guide for setting up keras at RECAS	14 days ago
 Keras	Updating macros for Bari	17 days ago
 StatisticalAnalysis	Initiating repository for CMS AN-18-120 codes	a month ago
 ZplusX	Initiating repository for CMS AN-18-120 codes	a month ago
 README.md	Initial commit	a month ago

 [README.md](#)

CMS_18_120_ANCodes

Macros from CMS AN-18-120



66 / 88

Conclusions: CMS L1TT AM+FPGA

- Significant contribution has been given to the CMS L1TT AM+FPGA:
 - Several new MC samples generated and made available for the group;
 - Developments, studies and implementations:
 - Synthetic matching and efficiency;
 - Duplicate removal;
 - Stub bending (ΔS);
 - Effects of truncation on roads and/or combinations;
 - Tracking fitter χ^2 cut revision;
- Support during the electronic inspections:
 - Check up of Pulsar BII boards and the PRMs;
 - Check up of optical cables connecting Pulsar boards;
 - Check up of boards connected in the crates;

- Creation of dedicated documentation about nomenclatures, workflow and implementations developed by the author within the package;

Notes on AM Simulation Studies

LUCIANO RISTORI¹, SERGO JINDARIANI¹, MÍQUEIAS M. ALMEIDA²
ON BEHALF OF CMS L1TT AM+FPGA WORKING GROUP

¹Fermi National Accelerator Laboratory (Fermilab), USA

²Universidade do Estado do Rio de Janeiro (UERJ), Brazil

September 21, 2017

Conclusions: Fast Matrix Element

- Although the *FastME* idea didn't get finished, some conclusions can be drawn from it:
 - The results are a re-statement of the KNN method developed by the TMVA team and suggest that it can even be simplified (no need of a volume in the chosen phase space);
 - This *FastME* idea is quite sensitive to the physical process in analysis and can even become useless;
 - The method is also sensitive to the size of the pattern banks used in the analysis and study of bias correction could be needed;
 - The method is easily applicable in several processes, some of which the MEM can not or don't have higher order corrections included;

Thank you for the attention!



Datasets used in the Analysis

Run range	Datasets	Integrated luminosity
273150-275376	Global Tag: 80X_dataRun2_2016SeptRepro_v7 /DoubleMuon/Run2016B-03Feb2017_ver2-v2/MINIAOD /DoubleEG/Run2016B-03Feb2017_ver2-v2/MINIAOD /MuonEG/Run2016B-03Feb2017_ver2-v2/MINIAOD /SingleMuon/Run2016B-03Feb2017_ver2-v2/MINIAOD /SingleElectron/Run2016B-03Feb2017_ver2-v2/MINIAOD	5.892 fb ⁻¹
275656-276283	/DoubleMuon/Run2016C-03Feb2017-v1/MINIAOD /DoubleEG/Run2016C-03Feb2017-v1/MINIAOD /MuonEG/Run2016C-03Feb2017-v1/MINIAOD /SingleMuon/Run2016C-03Feb2017-v1/MINIAOD /SingleElectron/Run2016C-03Feb2017-v1/MINIAOD	2.646 fb ⁻¹
276315-276811	/DoubleMuon/Run2016D-03Feb2017-v1/MINIAOD /DoubleEG/Run2016D-03Feb2017-v1/MINIAOD /MuonEG/Run2016D-03Feb2017-v1/MINIAOD /SingleMuon/Run2016D-03Feb2017-v1/MINIAOD /SingleElectron/Run2016D-03Feb2017-v1/MINIAOD	4.353 fb ⁻¹
276831-277420	/DoubleMuon/Run2016E-03Feb2017-v1/MINIAOD /DoubleEG/Run2016E-03Feb2017-v1/MINIAOD /MuonEG/Run2016E-03Feb2017-v1/MINIAOD /SingleMuon/Run2016E-03Feb2017-v1/MINIAOD /SingleElectron/Run2016E-03Feb2017-v1/MINIAOD	4.117 fb ⁻¹
277932-278808	/DoubleMuon/Run2016F-03Feb2017-v1/MINIAOD /DoubleEG/Run2016F-03Feb2017-v1/MINIAOD /MuonEG/Run2016F-03Feb2017-v1/MINIAOD /SingleMuon/Run2016F-03Feb2017-v1/MINIAOD /SingleElectron/Run2016F-03Feb2017-v1/MINIAOD	3.186 fb ⁻¹
278820-280385	/DoubleMuon/Run2016G-03Feb2017-v1/MINIAOD /DoubleEG/Run2016G-03Feb2017-v1/MINIAOD /MuonEG/Run2016G-03Feb2017-v1/MINIAOD /SingleMuon/Run2016G-03Feb2017-v1/MINIAOD /SingleElectron/Run2016G-03Feb2017-v1/MINIAOD	7.721 fb ⁻¹
281207-284068	Global Tag: 80X_dataRun2_Prompt_v16 /DoubleMuon/Run2016H-03Feb2017_ver2-v1/MINIAOD /DoubleEG/Run2016H-03Feb2017_ver2-v1/MINIAOD /MuonEG/Run2016H-03Feb2017_ver2-v1/MINIAOD /SingleMuon/Run2016H-03Feb2017_ver2-v1/MINIAOD /SingleElectron/Run2016H-03Feb2017_ver2-v1/MINIAOD /DoubleMuon/Run2016H-03Feb2017_ver3-v1/MINIAOD /DoubleEG/Run2016H-03Feb2017_ver3-v1/MINIAOD /MuonEG/Run2016H-03Feb2017_ver3-v1/MINIAOD /SingleMuon/Run2016H-03Feb2017_ver3-v1/MINIAOD /SingleElectron/Run2016H-03Feb2017_ver3-v1/MINIAOD	8.857 fb ⁻¹

JETMET POG Filters

Filter	Description
HBHENoiseFilter HBHENoiseIsoFilter	remove noisy events from the HCAL, where the HBHE scintillator produces anomalous signals with pulse shapes and pixel multiplicities discrepant from those from a clean signal
EcalDeadCellTriggerPrimitiveFilter	removes events with non-functioning ECAL data links, comparing the sum of energy deposited in each supercluster cell to the energy saturation of the trigger primitive
goodVertices	filter events with noisy vertex reconstruction (due to pileup effects) by requiring the reconstruction of at least one good vertex full filling the following criteria: high number of degree of freedom ($NPV > 4$), collisions restricted along the zaxis ($zPV < 24\text{cm}$) and small radius of the PV ($rPV < 2\text{cm}$)
eeBadScFilter	removes events with noisy ECAL endcap superclusters
globalTightHalo2016Filter	removes events with enhanced MET from beam-halo particles which are in time with the beam
BadPFMuonFilter BadChargedCandidateFilter	remove events with mis-reconstructed muon and charged hadron PF candidates

Post-Fit Yields and Distributions

Table 19: Background and signal estimations, with total uncertainty (statistical+systematic), derived from fitting the **s+b** model to the observed data, accounting 35.9fb^{-1} at $\sqrt{s} = 13\text{TeV}$.

Process	4μ	$4e$	$2e2\mu$	$4l$
ggH	2.48 ± 0.34	1.29 ± 0.19	3.14 ± 0.44	6.91 ± 0.59
VH	0.34 ± 0.03	0.20 ± 0.02	0.46 ± 0.04	1.00 ± 0.06
ttH	0.05 ± 0.01	0.03 ± 0.00	0.06 ± 0.01	0.14 ± 0.01
qqZZ+ZZJJ	0.67 ± 0.04	0.36 ± 0.03	0.75 ± 0.05	1.77 ± 0.07
ggZZ	0.05 ± 0.01	0.03 ± 0.00	0.06 ± 0.01	0.14 ± 0.01
Z+X	1.74 ± 0.34	0.29 ± 0.06	2.35 ± 0.44	4.37 ± 1.34
Σ backgrounds	5.34 ± 0.49	2.19 ± 0.20	6.81 ± 0.63	14.34 ± 0.82
qqH (signal $m_H = 125\text{GeV}$)	1.35 ± 0.77	0.76 ± 0.42	1.79 ± 1.01	3.90 ± 1.34
Total estimated	6.69 ± 0.91	2.95 ± 0.47	8.60 ± 1.19	18.24 ± 1.57

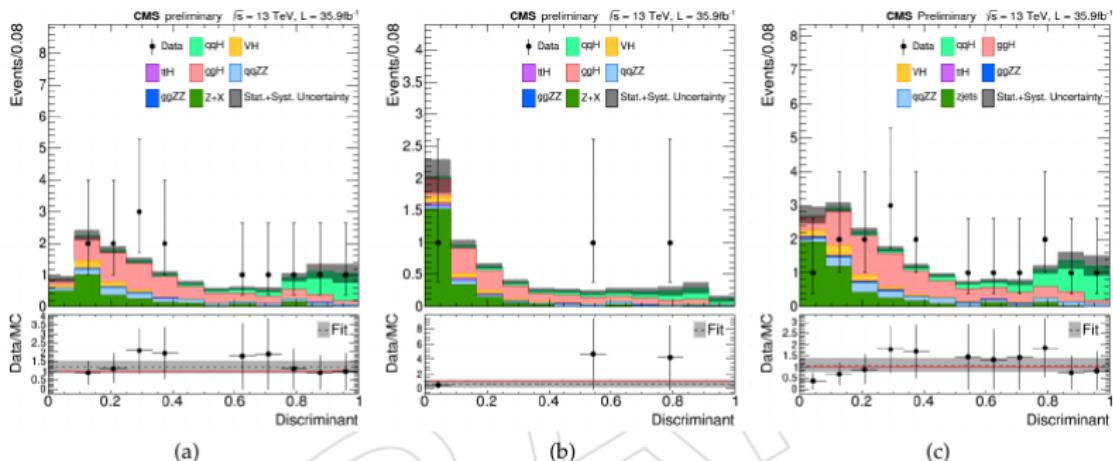
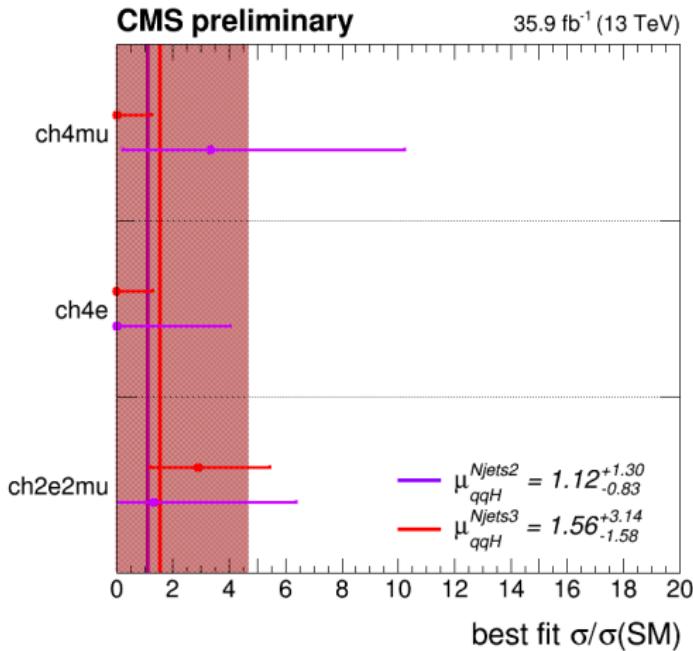


Figure 27: Post-fit NN distribution. The fit is done with the assumption of S+B hypothesis ($\mu_{VBF} = 1$ a priori).



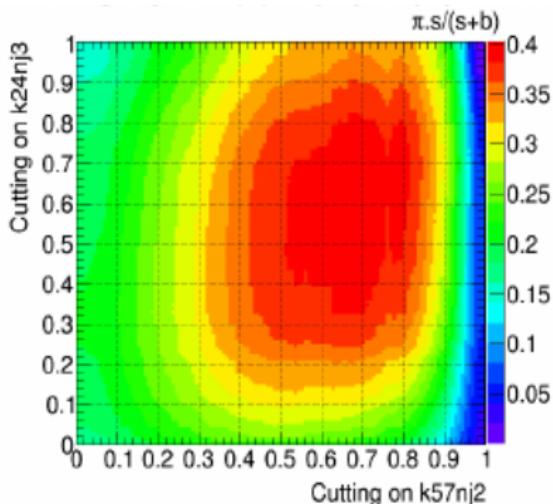
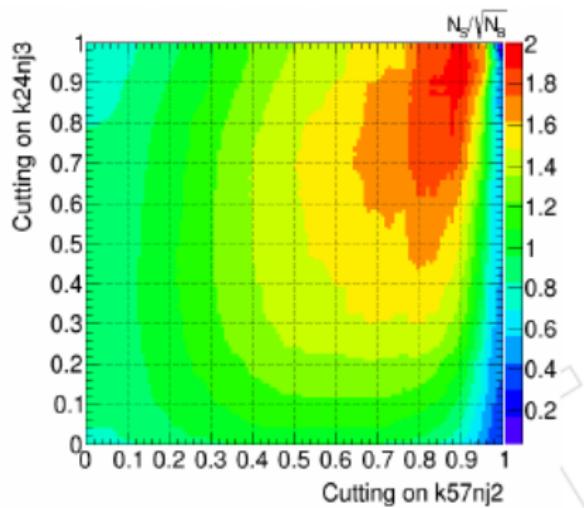
Comparison Between Best Fits of Njets2 and Njets3 Categories

- The best fits from each jet-based category. As it is shown, there's no advantage in using them alone instead of combining as it was done in the analysis;



Sensitivity of Combined ANNs

- Note that, no systematic uncertainty has been accounted here;



Events after SM Higgs Selections in each 4l Channel

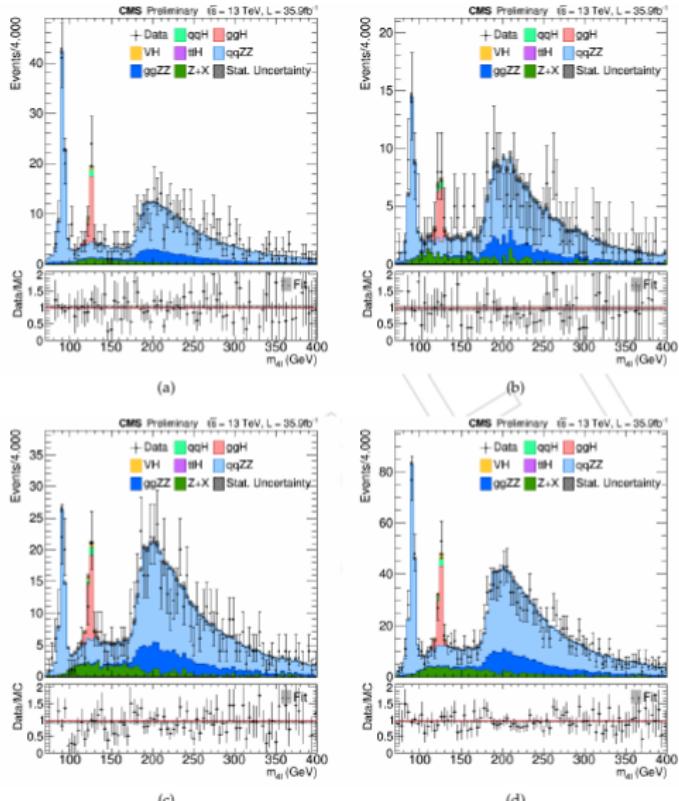
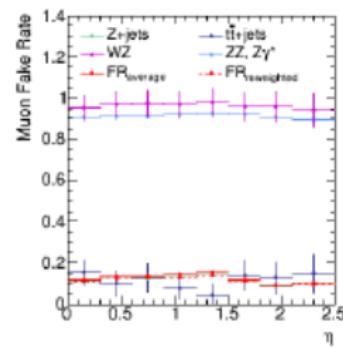
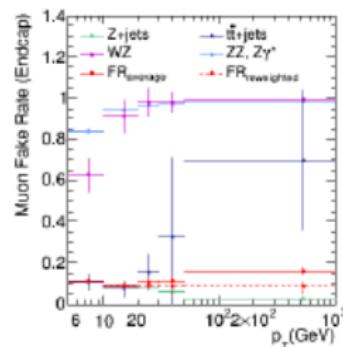
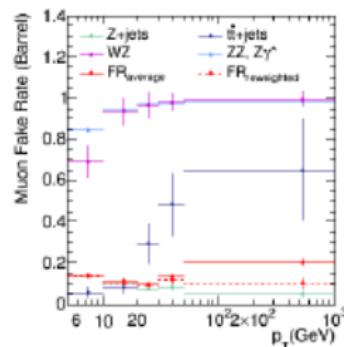
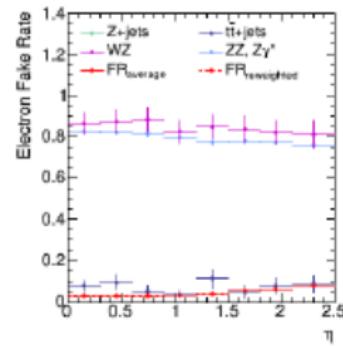
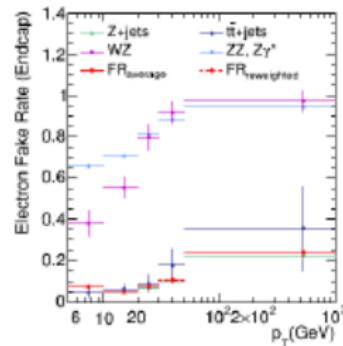
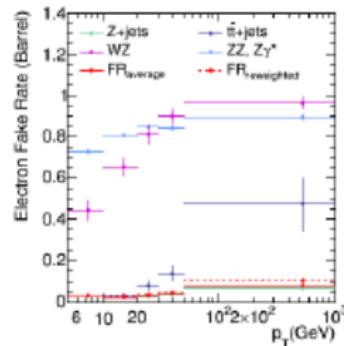


Figure 23: Final four-lepton mass distributions for the SM Higgs separated in the channels (a) 4μ , (b) $4e$ and (c) $2e2\mu$ and, combined into the (d) $4l$ final state. The distributions are shown for $70 < m_{4l} < 400$ GeV and include the $Z + X$ background estimated by the data-driven method explained in Sec. 6.4.

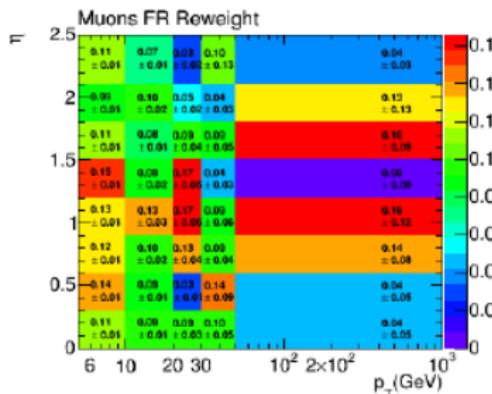
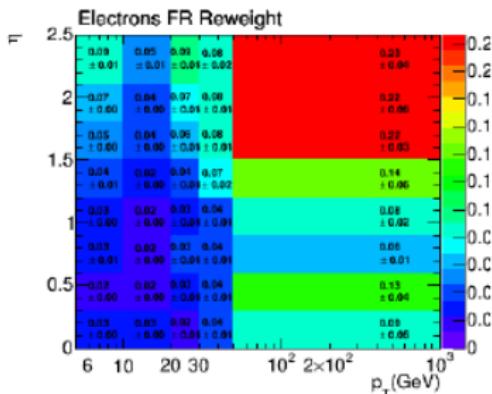
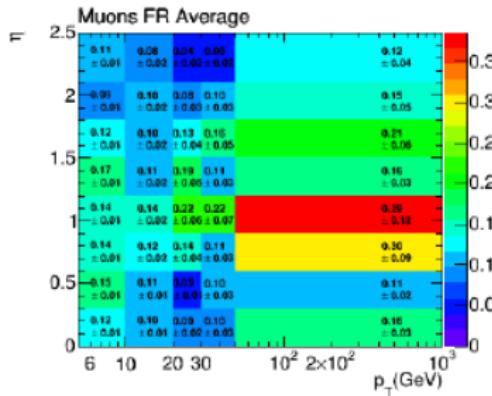
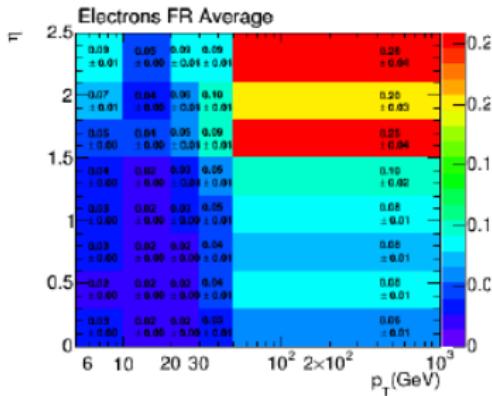
Z+X Systematic Uncertainty and Final Yields

- Z+X systematic uncertainty from FR: compute its variation by averaging MC (DY , $ZZ/Z\gamma$, WZ , $t\bar{t}$) FR and reweighing with 2P2F yields. Then propagate to Z+X estimation;



Z+X Systematic Uncertainty and Final Yields

- The behavior in 2D:

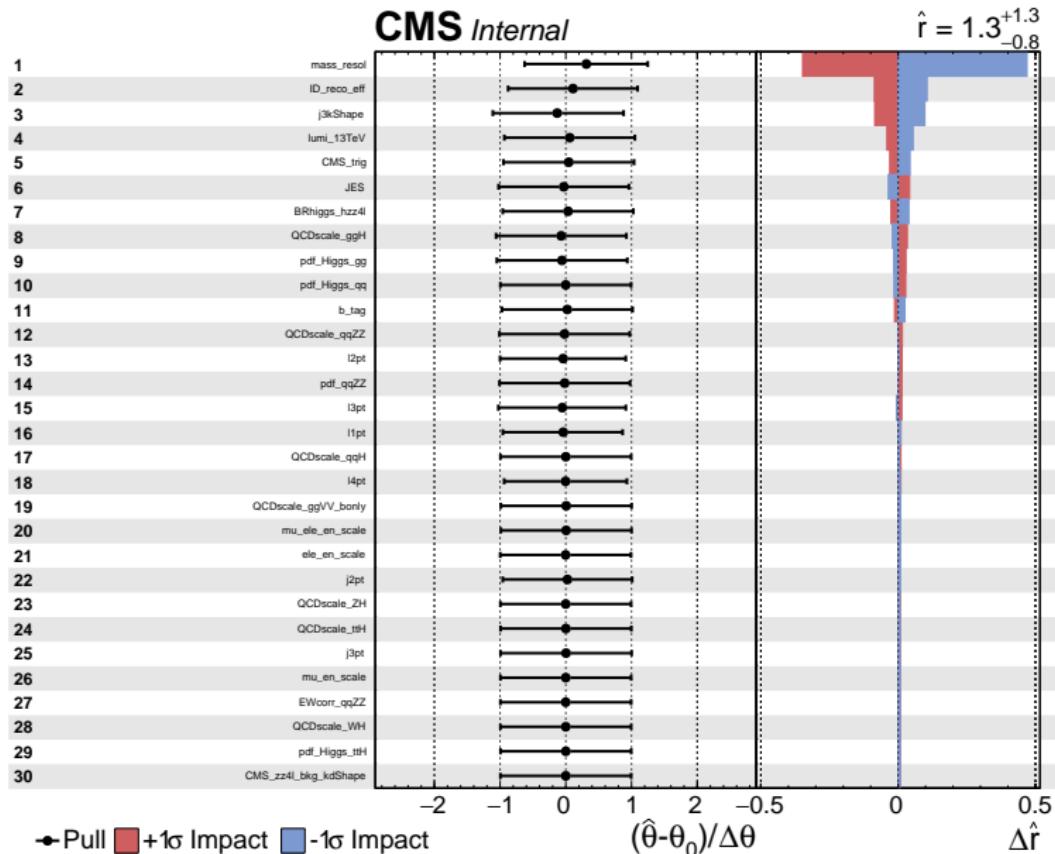


Z+X Systematic Uncertainty and Final Yields

Table 13: Final yields estimated for Z+X in the signal region from the measurements on data using the fake rate methods of opposite-sign (OS-OS) and same-sign (OS-SS) leptons. The estimates are reported with the total uncertainty for each final state after the SM Higgs and VBF-SR selections. The m_{4l}^{OS}/m_{4l}^{SS} is the ratio between the m_{4l} distribution for each final state obtained for Z+X in the OS and SS procedures (it is not computed for VBF-SR due to the very few Z+X events estimated there).

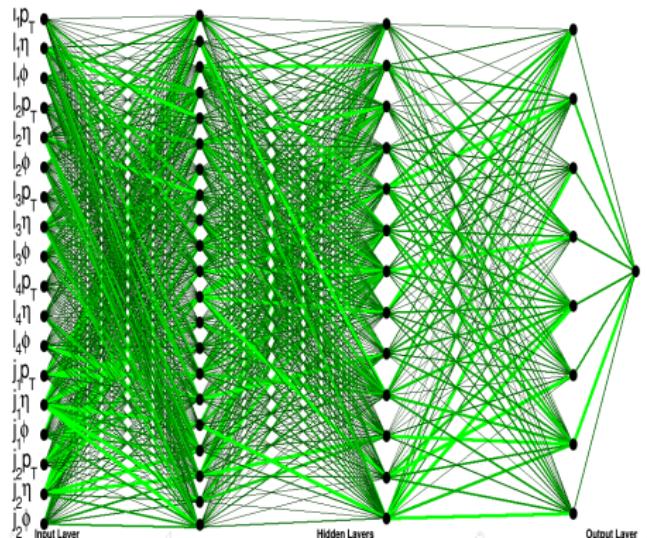
Z+X	4μ	4e	$2e2\mu$	4l
		SM Higgs		
OS-OS	24.28 ± 7.79	27.80 ± 1.84	56.90 ± 6.71	108.99 ± 10.33
OS-SS	24.00 ± 4.90	36.00 ± 6.00	64.00 ± 8.00	124.00 ± 11.14
m_{4l}^{OS}/m_{4l}^{SS}	0.75 ± 0.31	0.88 ± 0.33	0.98 ± 0.35	0.85 ± 0.33
		VBF-SR		
OS-OS	2.03 ± 0.49	0.33 ± 0.20	2.72 ± 0.42	5.08 ± 0.68
OS-SS	1.00 ± 1.00	1.00 ± 1.00	0.00 ± 0.00	2.00 ± 1.41

Systematic Uncertainties Impact on μ_{qqH} Fit

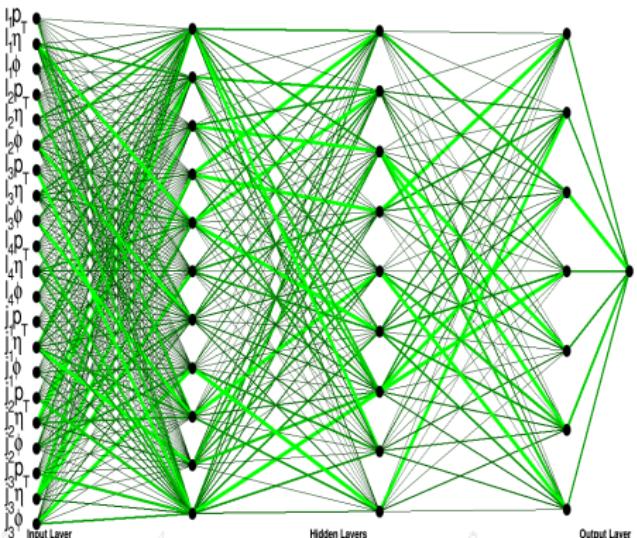


Architecture of Chosen ANNs

Figure: ANNs architecture created in this analysis. Black dots stand for inputs/neurons, while lines stand for the size of the ANN parameters chosen after training them. Wider and brighter lines means the parameter (weight or bias) associated to a given input for a neuron is larger (in other words its contribution is more relevant). ANN for Njets2 has 21:13:8 hidden neurons, while for Njets3 it has 11:9:7 (all neurons of SeLU type).



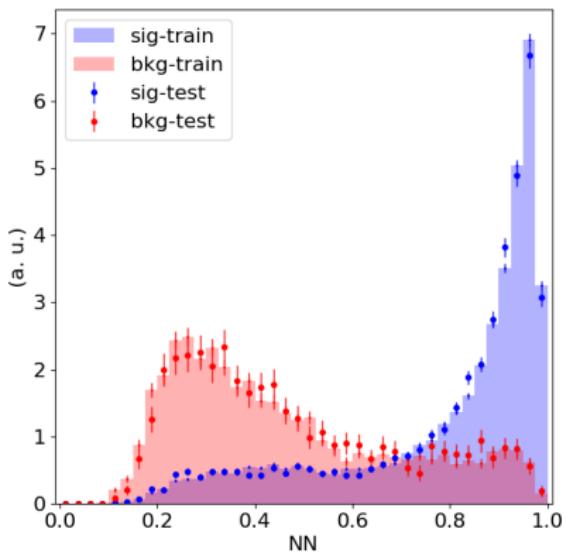
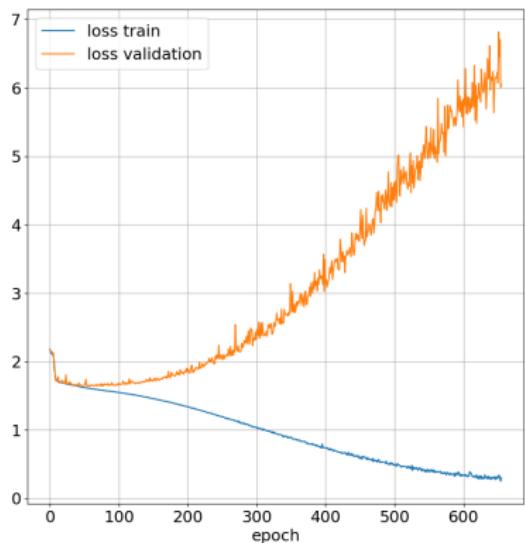
(a) Njets2 - 21:13:8



(b) Njets3 - 11:9:7

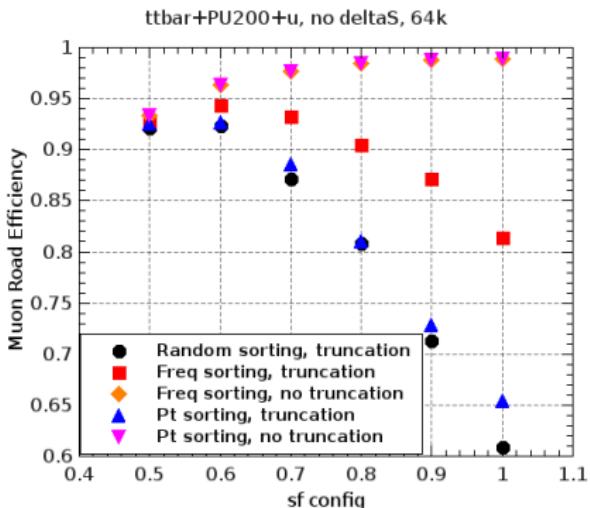
Keras ModelCheckpoint Option Against Overfitting

- Here is an example of overfitting in my case;
- Good solution found: **ModelCheckpoint**, it saves the model every time the validation loss decreases;
- Although the learning curves show overfitting the final ANN shapes from training and testing sub-datasets matches nicely;



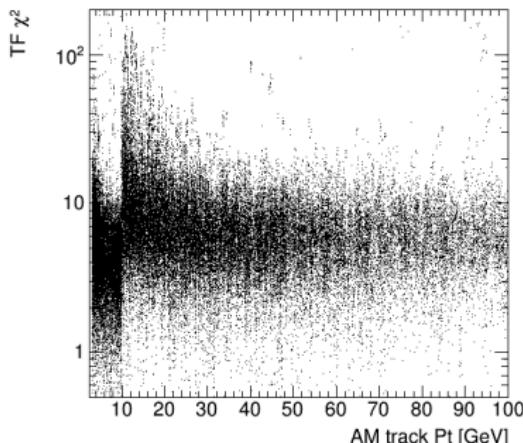
Simulation Studies: Roads and Combinations Truncation

- It was needed to check the approach under truncation of roads and combinations (for reducing latency, for instance);
- Implemented two new flags into the software for controlling the number of roads and combinations to be accepted for further processing;
- A study done by the author showed the smallest impact on the efficiency due to truncation happens when roads are sorted by frequency:



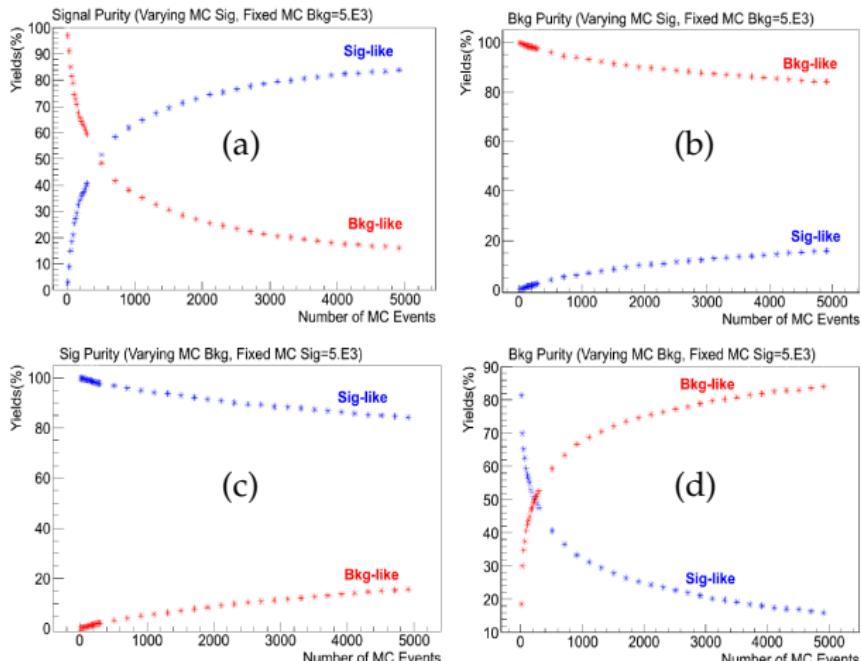
Simulation Studies: χ^2 Revision

- In order to optimize the final results, it was decided to re-study the tracking filter χ^2 value;
- It decides whether a combination of stubs is fine or not;
- Points to address:
 - Just an unique cut applied;
 - Dependence on the tracks p_T ;
- A new set of cuts have been adopted:
 - 6/6 combinations have a tight and unique threshold given by the 99% percentile of the theoretical $\chi^2_{ndof=8}$ curve (equals to 20.2);
 - 5/6 combinations have a p_T -based cut according to 8 ranges, which were defined to guarantee $\epsilon_{tracks}^{6/6+5/6} = 0.99 * \epsilon_{roads}^{6/6+5/6}$



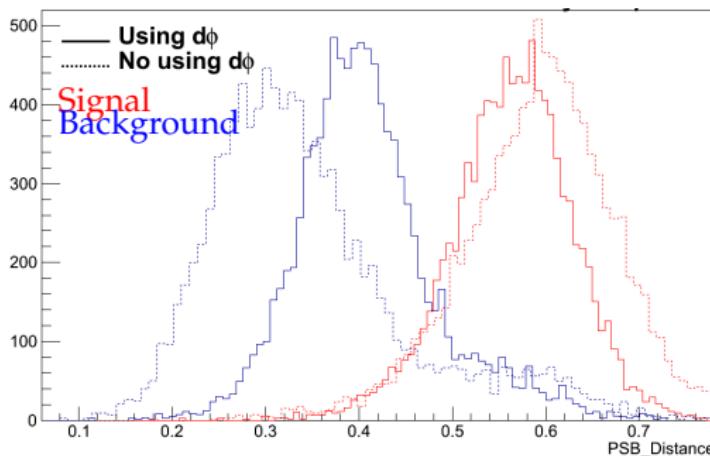
Simulation Studies: Bias from Pattern Bank Size

- The first point addressed during the development of the project was the influence of the pattern banks size. On graphs (a) and (b) the background pattern bank has a fixed size while the signal one is varied. On graphs (c) and (d) the opposite case is shown. Note, here "purity" is computed using the absolute number of events (without normalization).



Simulation Studies: Impact of ϕ Variable

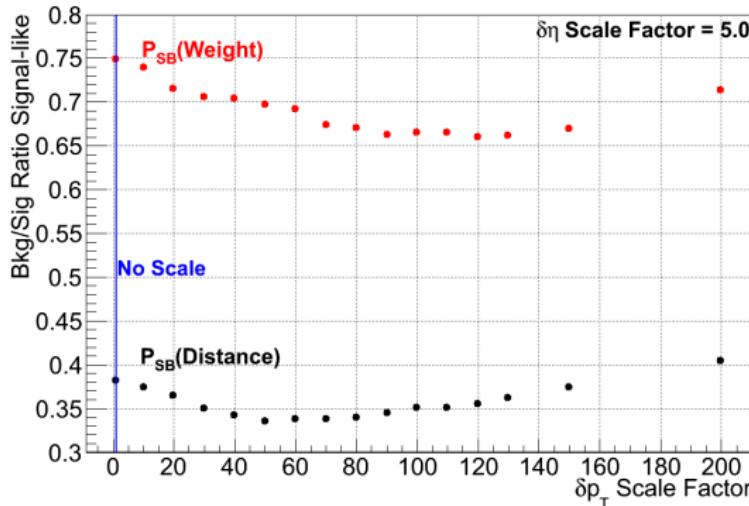
- Some studies have been done in order to optimize the performance of the discriminants;
- In the beginning of the project results showed that ϕ and E (energy) doesn't contribute and can actually worse the discriminant performance;
- Below is a comparison between the P_{SB}^D distribution using ϕ and without it:



- Based on such plot ϕ was removed from the default algorithm within *FastME* and left as an option to the user;

Simulation Studies: Scaling v_k 's Contribution

- Another point of optimization was the scaling of the variables used to compute the $R_{(ij)}$;
- Since the variables v_k present quite different ranges of variation it's important do level them;
- Here's a scan showing the variation of the fraction of signal events being misclassified as background in function of the δp_T . A fixed value of $\delta\eta = 5.0$ was used;



- The latest version of the project has an automated method which assigns the cumulative mean of MC events as the scaling factors;

The FastME Package

- Features:

- User gives to the program a configuration file;
- Original samples are replicated;
- $R_{(j,i)}$ can be simultaneously computed for each class;
- Creates a ROOT file containing D_{min}^{class} and $P_{SB}^{D/W}$;

FastME configuration file example

```
#####
## User config #####
## Path of samples ##
sig_data 0
data_path /eos/user/m/mmelodea/Keras/HiggsAnalysis/run2/KerasVBF/DNN_optimized_results/DataAugment
bkg_data 1
data_path /eos/user/m/mmelodea/Keras/HiggsAnalysis/run2/KerasVBF/DNN_optimized_results/DataAugment
sig_mc 0
mc_path /eos/user/m/mmelodea/Keras/HiggsAnalysis/run2/KerasVBF/DNN_optimized_results/DataAugment
bkg_mc 1
mc_path /eos/user/m/mmelodea/Keras/HiggsAnalysis/run2/KerasVBF/DNN_optimized_results/DataAugment

## Tree info ##
tree_name csv
id_branch_name ParticleId
pt_branch_name ParticlePt
eta_branch_name ParticleEta
phi_branch_name ParticlePhi

#####
## FastME configs #####
##NOTE
#MC quantity is by default based on fraction to mantain cross section relation
#if mc_limit is flaged as large than 1, then all MC, if they have events enough, will be used up
#if a MC doesn't have events enough, then all the available events in that sample is used
#if xs_scale is set to true, then data_limit and mc_limit must be set as -1

outfile_name vbf_gghjj_gen_mc_gf10_qm0p05
outfile_path /eos/user/m/mmelodea/Keras/HiggsAnalysis/run2/KerasVBF/DNN_optimized_results/DataAugment
phs_dr_method mindr
flavor_constraint false
n_cores 2
data_limit 10
mc_limit -1
scale_dPt -2
scale_dEta -2
scale_dPhi -2
```

FastME workflow view

