Model-independent differential measurement in HWW at the CMS experiment

CMS-PAS-HIG-24-004

Benedetta Camaiani
Università degli Studi di Firenze e INFN Firenze (Italy)

On behalf of the CMS Collaboration

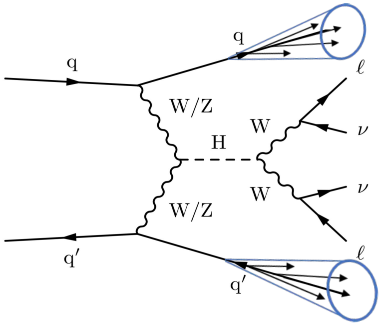
Introduction

- Differential measurement of the **Higgs boson production in association with 2 jets** in the **$H \to W^+W^- \to 2\ell 2\nu$** decay channel using data collected by the CMS detector during the LHC Run 2 ($L = 138 fb^{-1}$, $\sqrt{s} = 13$ TeV)

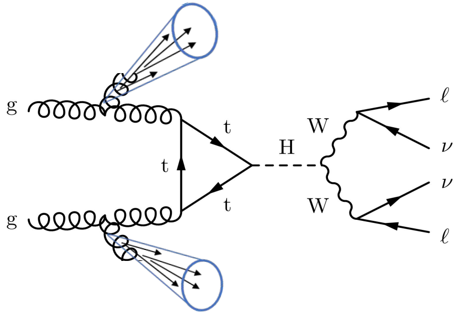
- Cross section as a function of **signed azimuthal angle difference** $\Delta\Phi_{jj}$ **between the two leading jets** in the final state

$$\Delta\Phi_{jj} = \phi_{j_k} - \phi_{j_l} \quad \text{with} \quad \eta_{j_k} > \eta_{j_l}$$

Vector boson fusion (VBF)



Gluon gluon fusion + 2 jets (ggH)

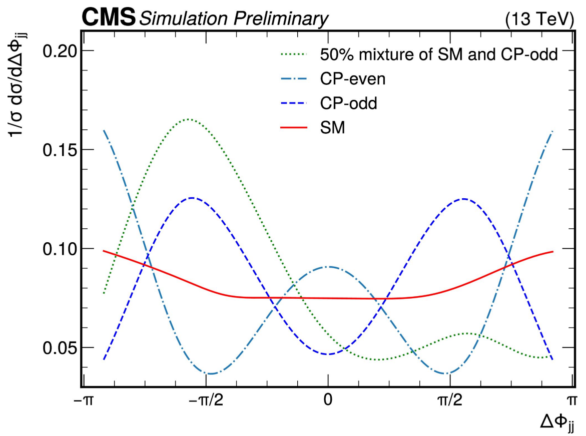


Introduction

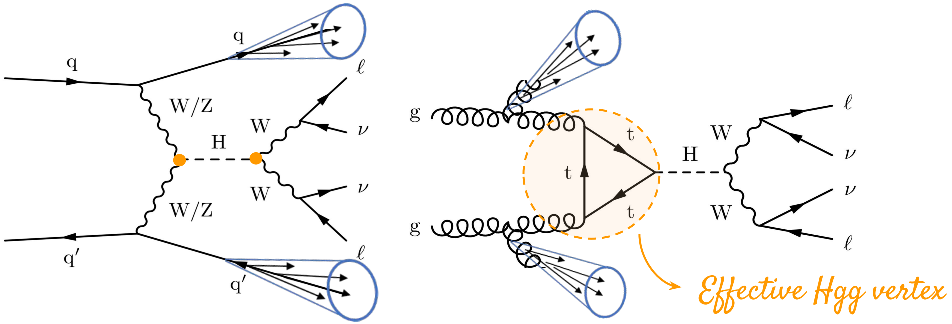
- Differential measurement of the **Higgs boson production in association with 2 jets** in the **$H \to W^+W^- \to 2\ell 2\nu$** decay channel using data collected by the CMS detector during the LHC Run 2 ($L=138 fb^{-1}$, $\sqrt{s}=13$ TeV)

- Cross section as a function of **signed azimuthal angle difference $\Delta\Phi_{jj}$ between the two leading jets** in the final state

$$\Delta\Phi_{jj} = \phi_{j_k} - \phi_{j_l} \quad \text{with} \quad \eta_{j_k} > \eta_{j_l}$$



Vector boson fusion (VBF) **Gluon gluon fusion + 2 jets** (ggH)



BSM theories predict *Anomalous Couplings* (**AC**) in the HVV/Hgg vertex, which may impact the kinematic distribution of the dijet system

➔ **Measuring the signal cross section regardless of the theoretical model governing the HVV/Hgg vertex**

Model dependence

- **Signal extraction** is typically performed by fitting the distribution of one or more observables to data using **templates derived from Monte Carlo** (MC) simulation

- These observables (e.g. m_T, m_{ll}, DNN output) can be **sensitive to the underlying signal model**, leading to **shape differences** in the corresponding MC templates
  - The **cross section may be biased toward the specific model used to generate the template**

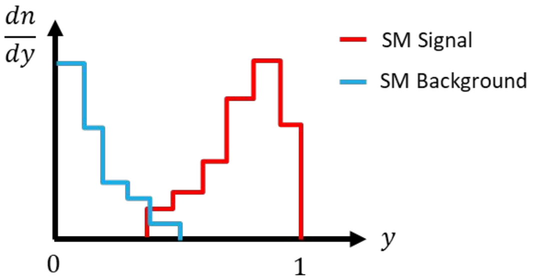
Model dependence

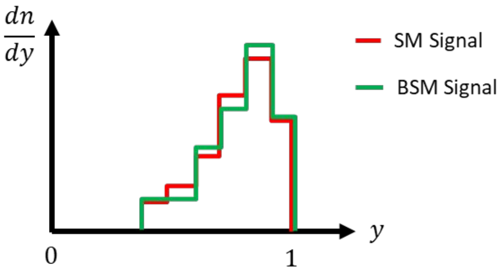
- **Signal extraction** is typically performed by fitting the distribution of one or more observables to data using **templates derived from Monte Carlo** (MC) simulation

- These observables (e.g. m_T, m_{ll}, DNN output) can be **sensitive to the underlying signal model**, leading to **shape differences** in the corresponding MC templates
  - The **cross section may be biased toward the specific model used to generate the template**

What we need

➔ A fit variable that discriminates signal from backgrounds but agnostic with respect to the signal hypothesis





Implemented using an **adversarial deep neural network** (ADNN)

Adversarial deep neural network

Eur. Phys. J. C (2022) 82: 921

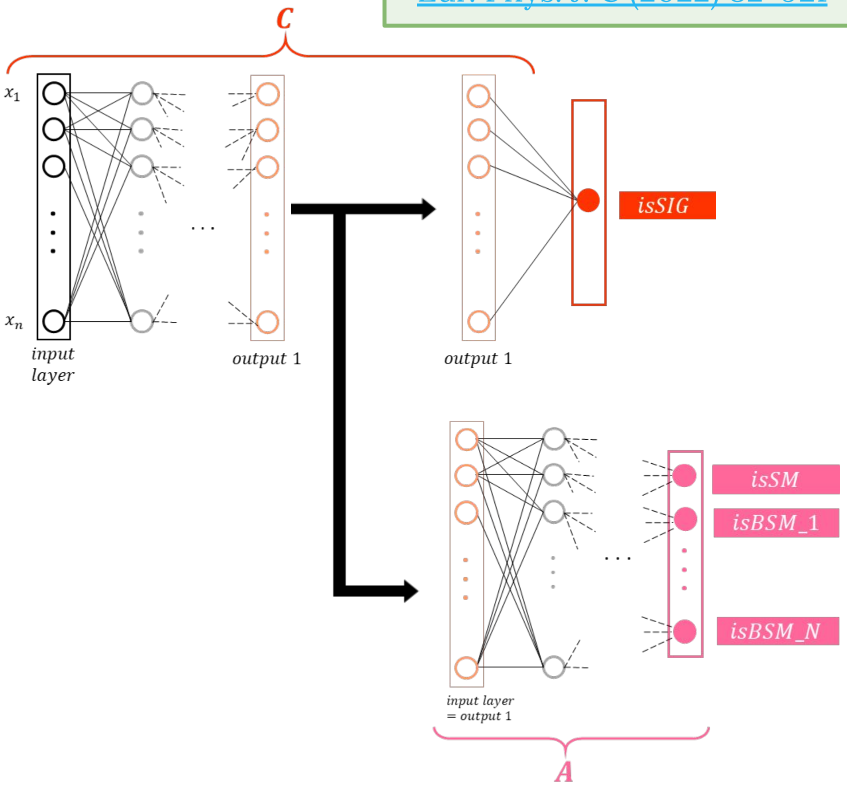
- Adversarial component introduced to suppress the network ability to learn specific features of a given signal hypothesis

  ★ **Classifier** trained to distinguish signal from background events, using samples from **multiple signal hypotheses** (SM+AC)

  ★ **Adversary** trained **only on signal events** to predict the signal model from the classifier internal representation

The adversary penalizes the classifier if its data representation is sensitive to the signal hypothesis

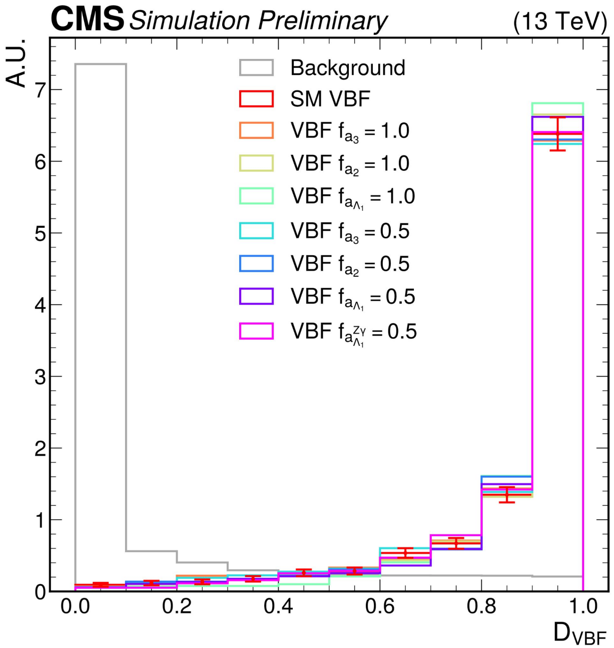
$$\mathcal{L}oss = \mathcal{L}oss(C) - \alpha \cdot \mathcal{L}oss(A)$$



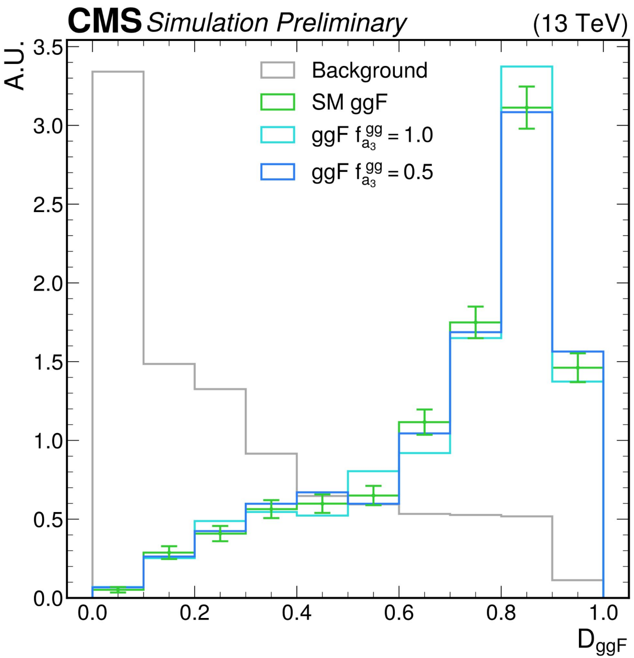
Model-agnostic classification

- ➔ Highly discriminative between signal and background
- ➔ SM and BSM shapes in good agreement, **classifier unable to recognize the physics models of signal events**

VBF-ADNN



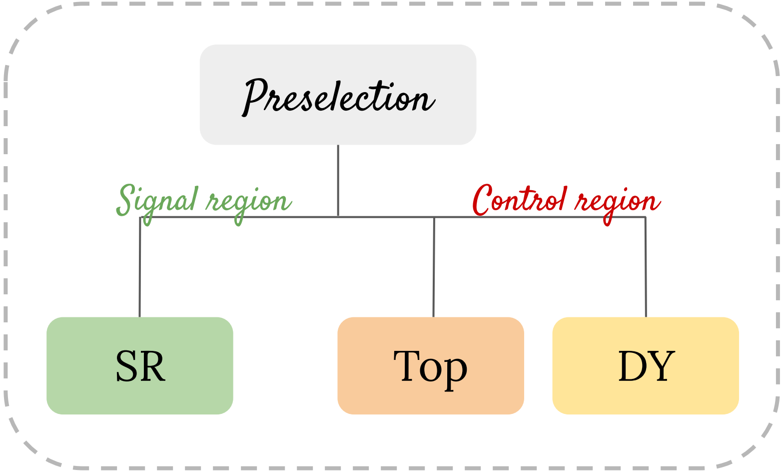
GGH-ADNN



Binary accuracy*: **91%** **79%**

**Fraction of events where predicted labels match the true labels (signal if $D_{sig} \geq 0.5$, background otherwise)*

Analysis strategy in a nutshell



- Different-flavour (eμ) final state with MET and ≥ 2 jets
- **Signal region (SR)** to enhance sensitivity to VBF and ggF production
  - 2 ADNNs trained to extract signal while suppressing model dependence
- **2 control regions (CRs)** orthogonal to the SR to constrain main backgrounds (tt +tW and DY→ττ)

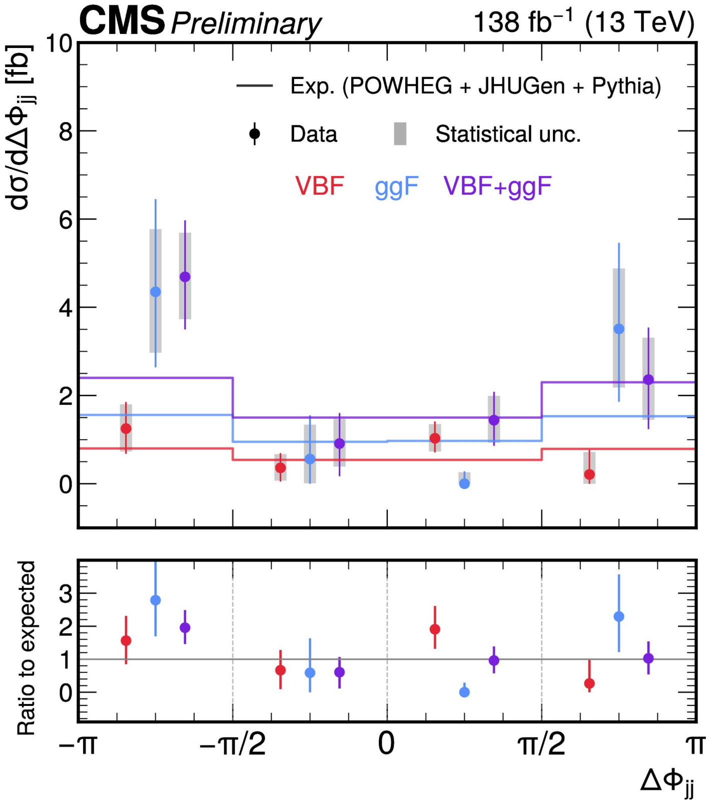
➔ **Each region is divided into four equally-spaced $\Delta\Phi_{jj}$ bins spanning $[-\pi,+\pi]$**

- Signal extracted through a **simultaneous fit of ADNN outputs** across SR and CRs
  - $\mathbf{D_{VBF}}$ (output of the VBF-ADNN) and $\mathbf{D_{VBF,ggF}}$ (2D variable composed of D_{VBF} and D_{ggF})
  - Cross sections measured in a **fiducial phase space** with unfolding embedded in the fit

Differential cross sections

- Signal extracted in three different configurations
  - Overall H+2jets cross section (**$D_{VBF,ggF}$**)
  - Simultaneous measurement of VBF and ggF cross sections (**$D_{VBF,ggF}$**)
  - VBF-only cross section (**D_{VBF}**)
- Results are dominated by the statistical uncertainty

Model dependence (backup) reduced by 30-70% (depending on the $\Delta\Phi_{jj}$ bin) compared to using a standard DNN



SMEFT interpretation

- Differential measurements are used to **constrain extensions of the SM** that introduce anomalous interactions between the Higgs boson and SM particles, using an EFT approach

- SM EFT lagrangian:

$$\mathcal{L}_{\mathrm{SMEFT}} = \mathcal{L}_{\mathrm{SM}} + \mathcal{L}_{\mathrm{EFT}} = \mathcal{L}_{\mathrm{SM}} + \sum_{i=5}^{\infty} \sum_{j}^{N_i} \frac{c_j^{(i)}}{\Lambda^{i-4}} \mathcal{O}_j^{(i)}$$

Wilson coefficients: $c_j^{(i)}$

Higher-dimensional operators: $\mathcal{O}_j^{(i)}$

BSM energy scale: Λ

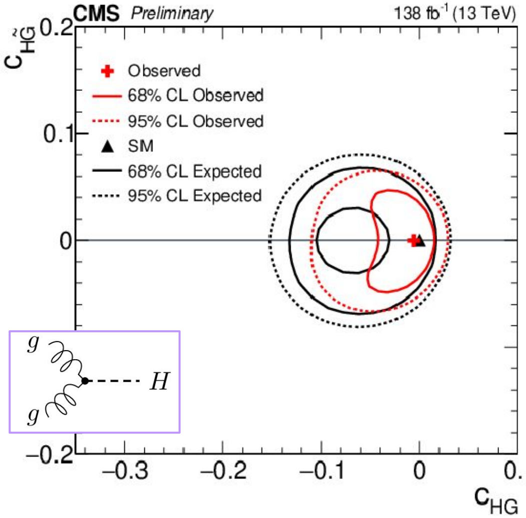
Values of WC different from 0 ➡ **hint of discrepancy from SM**

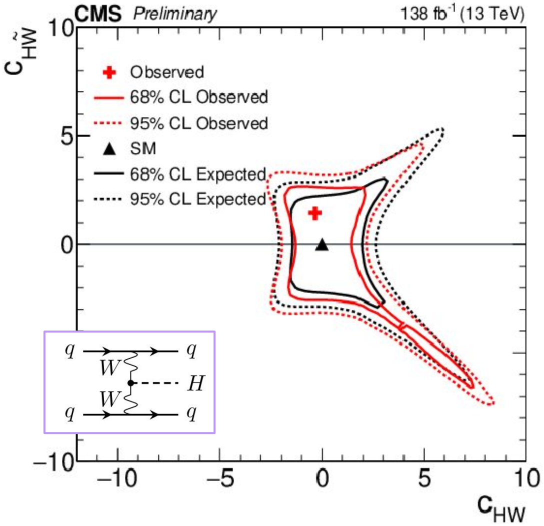
- The goal is to study a subset of i=6 operators and constrain the corresponding Wilson coefficients in the Warsaw basis

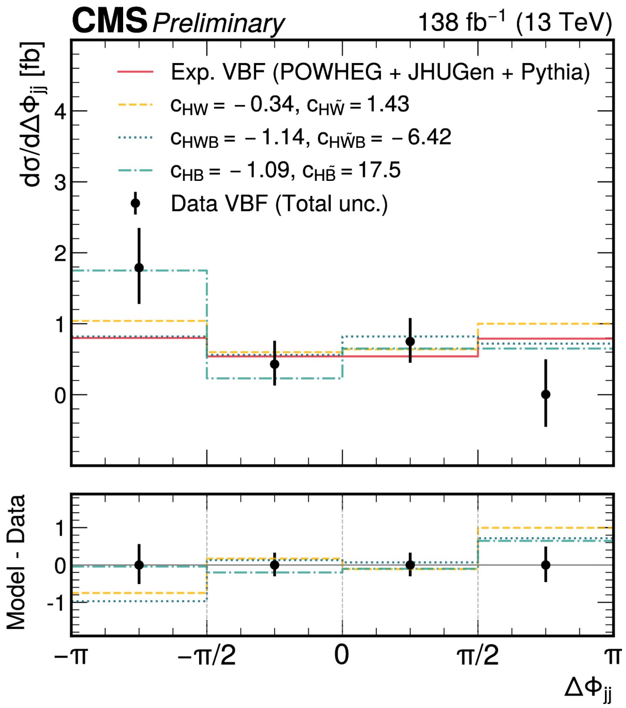
Class	Operator	Wilson Coefficient
$\mathcal{L}_6^{(4)} - X^2H^2$ (CP-even)	$H^\dagger H W^i_{\mu\nu} W^{i\mu\nu}$	c_{HW}
	$H^\dagger H W^i_{\mu\nu} B^{i\mu\nu}$	c_{HWB}
	$H^\dagger H B_{\mu\nu} B^{\mu\nu}$	c_{HB}
	$H^\dagger H G^a_{\mu\nu} G^{a\mu\nu}$	c_{HG}
$\mathcal{L}_6^{(4)} - X^2H^2$ (CP-odd)	$H^\dagger H \tilde{W}^i_{\mu\nu} W^{i\mu\nu}$	$c_{\mathrm{H\tilde{W}}}$
	$H^\dagger H \tilde{W}^i_{\mu\nu} B^{i\mu\nu}$	$c_{\mathrm{H\tilde{W}B}}$
	$H^\dagger H \tilde{B}_{\mu\nu} B^{\mu\nu}$	$c_{\mathrm{H\tilde{B}}}$
	$H^\dagger H \tilde{G}^a_{\mu\nu} G^{a\mu\nu}$	$c_{\mathrm{H\tilde{G}}}$
$\mathcal{L}_6^{(3)} - H^4D^2$	$(H^\dagger H)\Box(H^\dagger H)$	$c_{\mathrm{H\Box}}$
	$(D^\mu H^\dagger H)(H^\dagger D_\mu H)$	c_{HD}

Reinterpreted cross sections

- Tighter constraints are obtained on the CP-even $\mathbf{c_{HW}}$ and $\mathbf{c_{HG}}$ coefficients







- The differential cross section is recalculated using the measured values of the WCs

Conclusions

- Differential measurement of the **Higgs boson production cross section in association with 2 jets** performed in the **$H \to W^+W^- \to 2\ell 2\nu$** decay channel at the CMS experiment
  - Results are consistent with the SM predictions
  - The statistical uncertainty is the dominant component of the total uncertainty

- The use of **an adversarial neural network has reduced the dependence on theoretical models compared to a standard approach**

- Data from **LHC Run 3 will allow for improved precision** in this measurement

Thanks for your attention

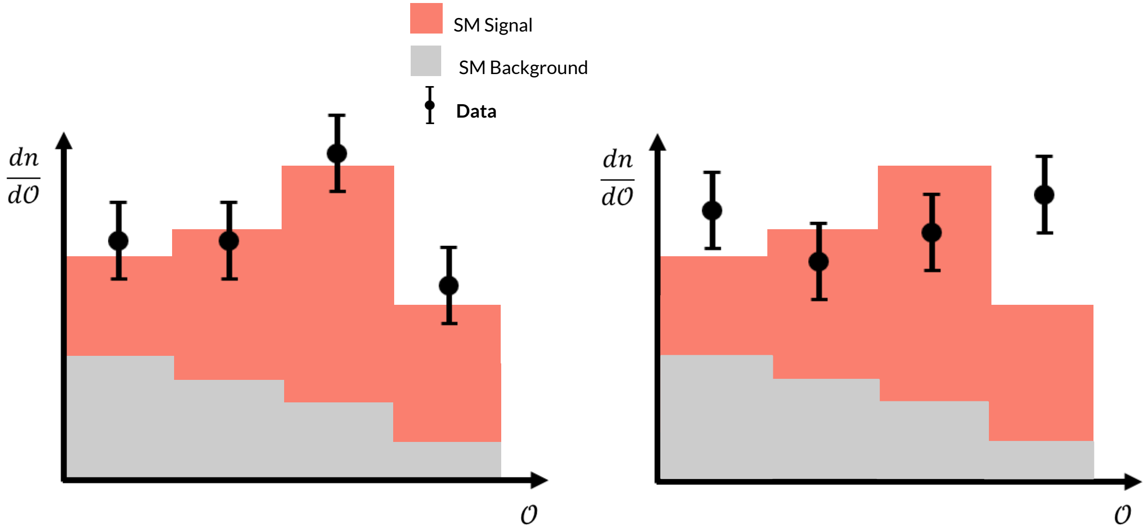
[CMS-PAS-HIG-24-004]

Backup

Template fit

- Extract the PDF of a discriminating observable from Monte Carlo (MC) simulations and then fit it to data
  - Usually, the SM hypothesis is assumed for simulation

- The underlying physics model governing the data is **unknown**
  - Fitting data with the SM hypothesis, will lead to a biased results



The fit results is biased towards the model assumed to generate the MC simulation

Anomalous couplings in HVV/Hgg

Scattering amplitude of one spin-0 Higgs boson (H) and two spin-1 gauge bosons ($V1$ $V2$)

$$A(HVV) \sim \left[a_1^{VV} + \underbrace{\frac{k_1^{VV} q_{V1}^2 + k_2^{VV} q_{V2}^2}{\left(\Lambda_1^{VV}\right)^2}}_{L_1} \right] m_{V1}^2 \epsilon_{V1}^* \epsilon_{V2}^* + a_2^{VV} f_{\mu\nu}^{*(1)} f^{*(2)\mu\nu} + a_3^{VV} f_{\mu\nu}^{*(1)} \tilde{f}^{*(2)\mu\nu}$$

HVV couplings

- $a_1^{VV} = 1$ SM tree level coupling $J^{CP}=0^{++}$
- L_1 CP-even AC
- a_2^{VV} CP-even AC
- a_3^{VV} CP-odd AC

Hgg couplings

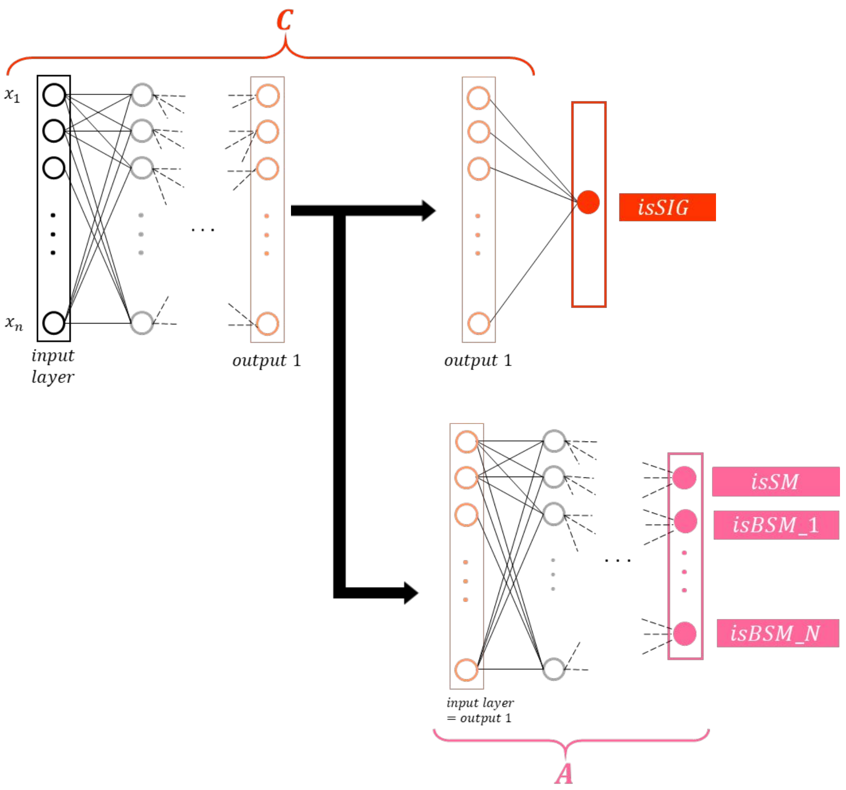
- a_2^{VV} SM loop
- a_3^{VV} CP-odd AC

Events selection

Region	Requirements
Global selection	Oppositely-charged $e\mu$ final state $p_{T}^{\ell 1} > 25\,\mathrm{GeV}$ $p_{T}^{\ell 2} > 13\,\mathrm{GeV}$ (10 GeV for 2016 data) $p_{T}^{\ell 3} < 10\,\mathrm{GeV}$ $m_{\ell\ell} > 12\,\mathrm{GeV}$ $p_{T}^{\ell\ell} > 30\,\mathrm{GeV}$ $E_{T}^{miss} > 20\,\mathrm{GeV}$ at least two jets with $p_T > 30\,\mathrm{GeV}$ and $\lvert\eta\rvert < 4.7$ $m_{jj} > 120\,\mathrm{GeV}$
SR	$m_T > 60\,\mathrm{GeV}$ $m_{T}^{\ell_2} > 30\,\mathrm{GeV}$ no b-tagged jets with $p_T > 20\,\mathrm{GeV}$
Top CR	$m_{\ell\ell} > 50\,\mathrm{GeV}$ $m_{T}^{\ell_2} > 30\,\mathrm{GeV}$ at least one b-tagged jet
DY CR	$40 < m_{\ell\ell} < 80\,\mathrm{GeV}$ $m_T < 60\,\mathrm{GeV}$ no b-tagged jets with $p_T > 20\,\mathrm{GeV}$

Adversarial deep neural network

Eur. Phys. J. C (2022) 82: 921



Classifier

- ***Binary* neural network**, trained on both **signal and background events**
- Aims to determine if the event is signal- or background-like
- Signal sample includes events coming from different *domains*, i.e. **different signal models** (SM and AC)

Adversary

- ***Multiclass* neural network** trained **only on signal events** (SM + AC hypotheses)
- **Tries to guess the physics model of signal events**, regressing the domain from the second-to-last layer of C

The adversary penalizes the classifier if its data representation is sensitive to the signal hypothesis

Competitive learning

- The classifier is penalized if its output contains too much information on the domain of origin of signal events

- If C manages to prevent A from identifying the signal model, then the **classification is independent of** the domains of origin of the events, i.e. **the physics model of signal events**

Compute first the gradient of $\mathcal{L}$ with respect to the C weights. A weights frozen in this step.

The parameter α regulates the interplay between A and C

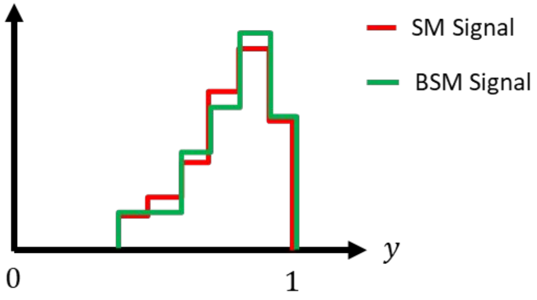
Two-step training procedure

1. $\mathcal{L}oss = \mathcal{L}oss(C) - \underbrace{\alpha \cdot \mathcal{L}oss(A)}_{\text{Penalty term}}$
2. $\mathcal{L}oss(A)$

Compute the gradient of $\mathcal{L}(A)$ with respect to the A weights

Optimization of the ADNN

- Hyperparameter optimization using the Optuna package by:
  - Maximizing the **binary accuracy** of the classifier
  - Minimizing the average of the two-sample **Kolmogorov-Smirnov** (K-S) **score between pairs of signal hypothesis** ("average K-S test statistic")



Optimized hyperparameters

Hyperparameter	Interval
α	$[0, 100]$
n_{nodes}	$[10, 100]$
n_{l}^{C}	$[1, 10]$
n_{l}^{A}	$[1, 10]$

Some technical details

★ **Adam gradient-descent optimized** $\eta^{C} = 10^{-5}$, $\eta^{A} = 10^{-3}$

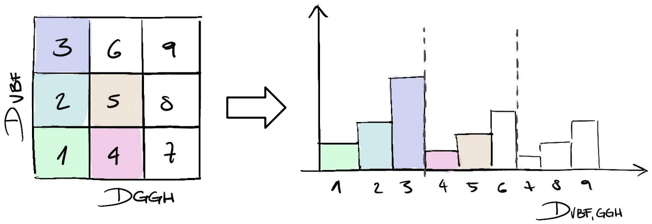
★ Activation function of hidden layers: **ReLU**

★ **Classifier** loss function and activation function of the output layer: **binary cross-entropy** and **sigmoid**

★ **Adversary** loss function and activation function of the output layer: **categorical cross-entropy** and **softmax**

Signal extraction

- Combined **binned maximum likelihood fit** of the ADNN output distribution with signal and background templates, performed simultaneously in all SR bins
  - 1D variable: $\mathbf{D_{VBF}}$ (output of the VBF-ADNN)
  - 2D variable: two-dimensional variable $\mathbf{D_{VBF,ggF}}$ composed of D_{VBF} and D_{ggF} (output of the GGH-ADNN)



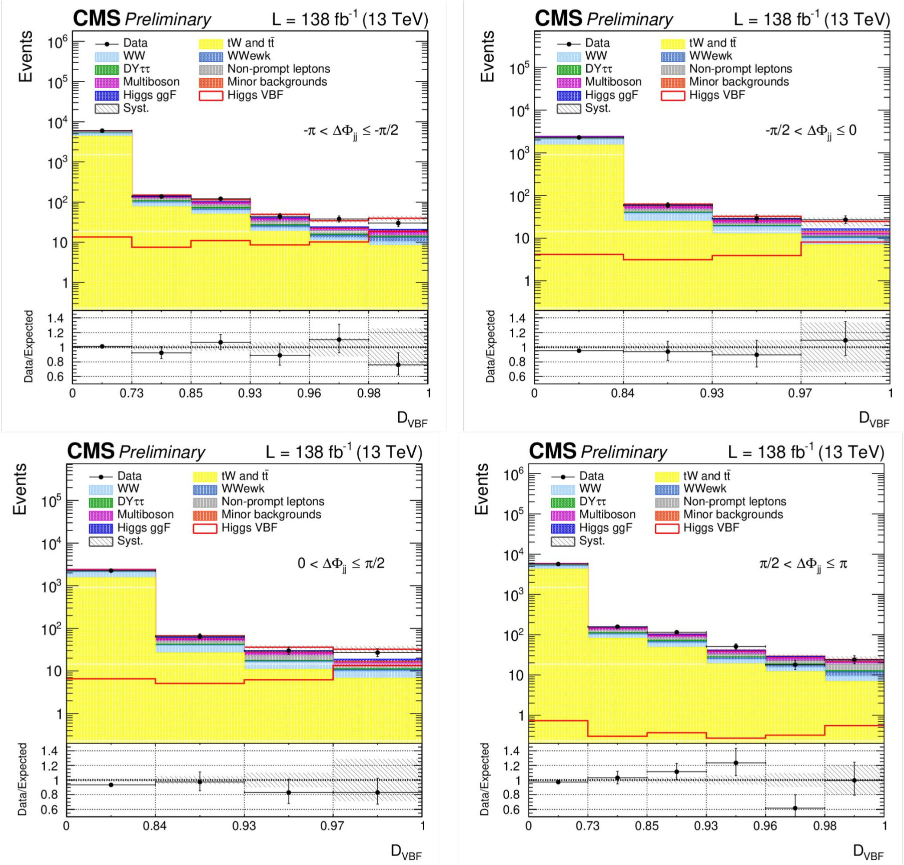
- **Unfolding procedure embedded in the likelihood fit**, treating each gen-level $\Delta\Phi_{jj}$ signal contribution as a separate process

- **Fiducial and out-of-fiducial** signal contributions coming from the same $\Delta\Phi_{jj}$ bin are **scaled together** with the same signal strength parameter

- The **number of events** in each $\Delta\Phi_{jj}$ bin of the Top and DY CRs is also **fitted to data** in order to constrain the normalisation of the corresponding background

D_{VBF}

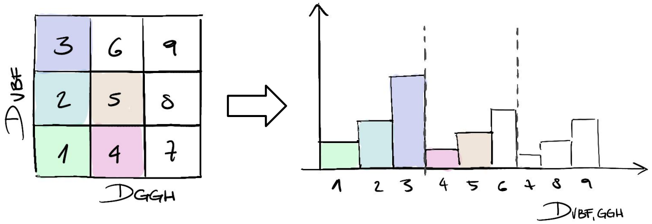
- **Postfit distributions** of D_{VBF} in the $\Delta\Phi_{jj}$ **bins of the SR** (logarithmic scale)

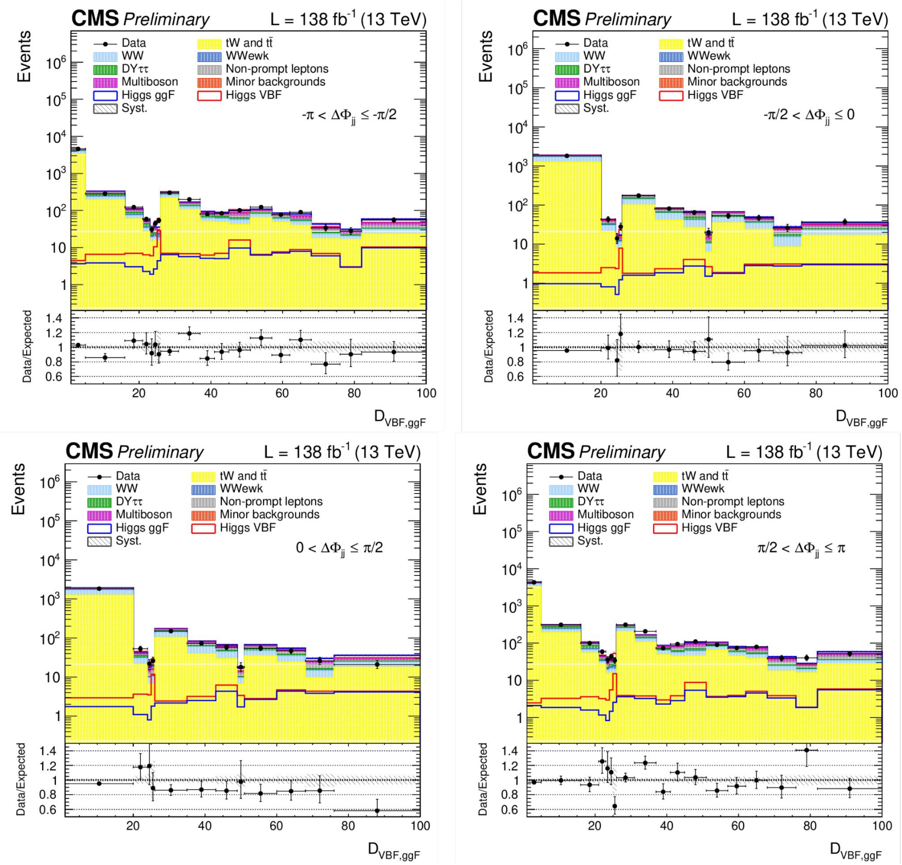


$D_{VBF,ggF}$

- **Postfit distributions** of $D_{VBF,ggF}$ in the $\Delta\Phi_{jj}$ **bins of the SR** (logarithmic scale)
  - Distribution unrolled in one dimension

The content of each bin belonging to the 2D distribution is transposed ("unrolled") to a 1D histogram, where bin labels on the x-axis correspond to the **bin numbering of the 2D map**





Fiducial volume

- Same cuts of the SR are applied on gen-level quantities except for the MET and m_T^{l2} requirements, that have been removed
- Leptons are *dressed* i.e., the momenta of radiated photons within a cone of ΔR < 0.1 around the lepton direction are summed to the lepton momentum
- Signal samples are split in $\Delta\Phi_{jj}$ bins, using the same binning scheme as the analysis categories

Fiducial requirements
oppositely-charged $e\mu$ (not from τ decay) final state,
at least two jets with $p_T > 30$ GeV,
$p_T^{\ell 1} > 25$ GeV, $p_T^{\ell 2} > 13$ GeV, $p_T^{\ell 3} < 10$ GeV,
$p_T^{\ell\ell} > 30$ GeV, $m_{\ell\ell} > 12$ GeV,
$m_{jj} > 120$ GeV, $m_T > 60$ GeV,
$\lvert\eta_{j_1}\rvert < 4.7$, $\lvert\eta_{j_2}\rvert < 4.7$,
$\lvert\eta_{\ell_1}\rvert < 2.5$, $\lvert\eta_{\ell_2}\rvert < 2.5$

Fit configurations

Fit	Measurement	POI	Variable	Motivation
1	VBF and ggF measured together	VBF and ggF scaled using a common POI ($\mu_{i,VBF+ggF}$ **with i=0,..,3**)	$D_{VBF,ggF}$	Overall H+2jets differential cross section
2	VBF and ggF as two independent processes	VBF and ggF scaled using two distinct POIs ($\mu_{i,VBF}$ and $\mu_{i,ggF}$ **with i=0,..,3**)	$D_{VBF,ggF}$	Most model independent measurement
3	VBF cross section	Only VBF scaled ($\mu_{i,VBF}$ **with i=0,..,3**)	D_{VBF}	Allows more stringent constraints on SMEFT couplings

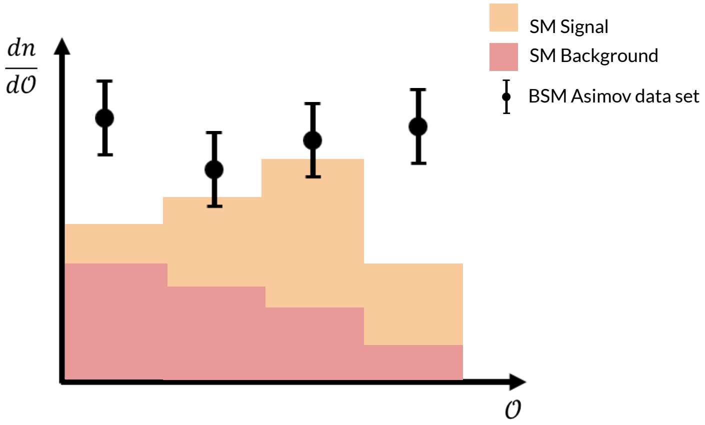
Uncertainties breakdown

- Contributions of different sources of uncertainty, expressed as a percentage of the total uncertainty ($\Delta\sigma_i / \Delta\sigma_{tot}$ x 100)
- For asymmetric errors, the largest of the up and down uncertainties is reported
- The systematic component includes all sources except for background normalization, which is part of the statistical component

Uncertainty source	$-\pi < \Delta\Phi_{jj} \leq -\frac{\pi}{2}$	$-\frac{\pi}{2} < \Delta\Phi_{jj} \leq 0$	$0 < \Delta\Phi_{jj} \leq \frac{\pi}{2}$	$\frac{\pi}{2} < \Delta\Phi_{jj} \leq \pi$
Theory (sig.)	18%	16%	16%	17%
Theory (backg.)	30%	17%	16%	27%
Integrated luminosity	16%	17%	15%	14%
b tagging	16%	14%	15%	13%
Jets	15%	15%	13%	14%
Non-prompt	14%	15%	15%	17%
PileUp	15%	15%	13%	16%
Leptons	17%	15%	13%	15%
MET	14%	15%	13%	14%
Backg. norm.	14%	15%	13%	14%
MC stat.	31%	31%	32%	40%
Stat. uncertainty	91%	93%	93%	86%
Syst. uncertainty	46%	38%	39%	52%

Model dependence estimation

- **Pseudodata set under BSM signal hypothesis** and SM background contribution
- **Fit using the SM signal** and background **templates**



The bias reflects the impact of **assuming the SM hypothesis** when extracting the signal **in scenarios that include BSM contributions**

Fit result

$$\hat{\mu} = \frac{s^{fit}_{BSM}}{s^{exp}_{SM}}$$

- *number of measured BSM signal events*
- *number of expected SM signal events*

Cross sections ratio

$$\tilde{\mu} = \frac{\sigma_{BSM}}{\sigma_{SM}}$$

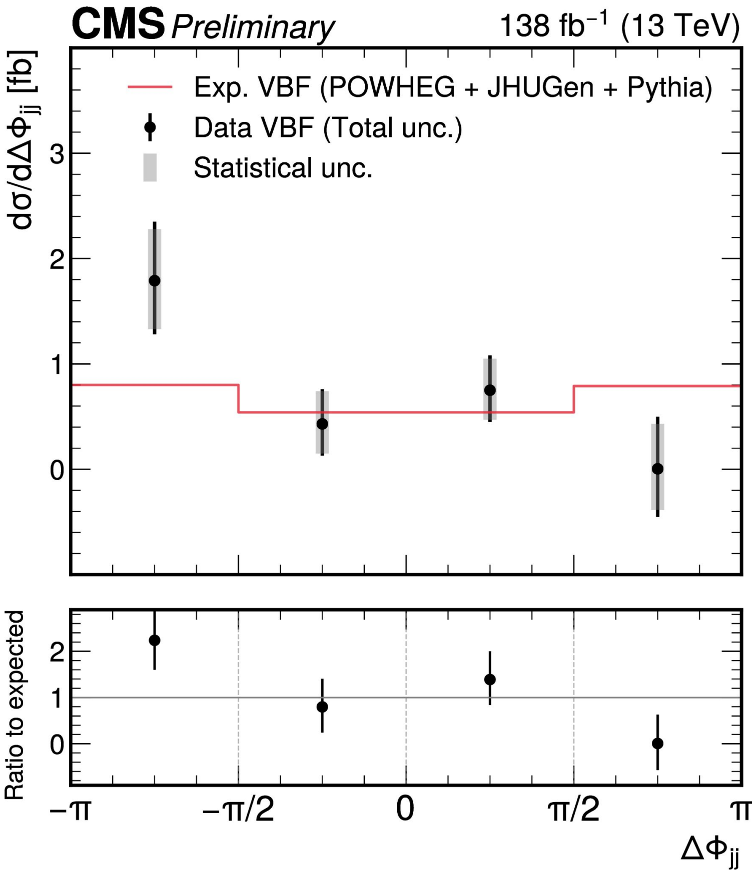
expected value if there is no bias in the analysis

Total bias

$$b = \frac{\hat{\mu} - \tilde{\mu}}{\tilde{\mu}}$$

it includes shapes effects in the fit variable, systematic shifts from unfolding procedure, variations in acceptance and selection efficiencies

VBF differential cross section



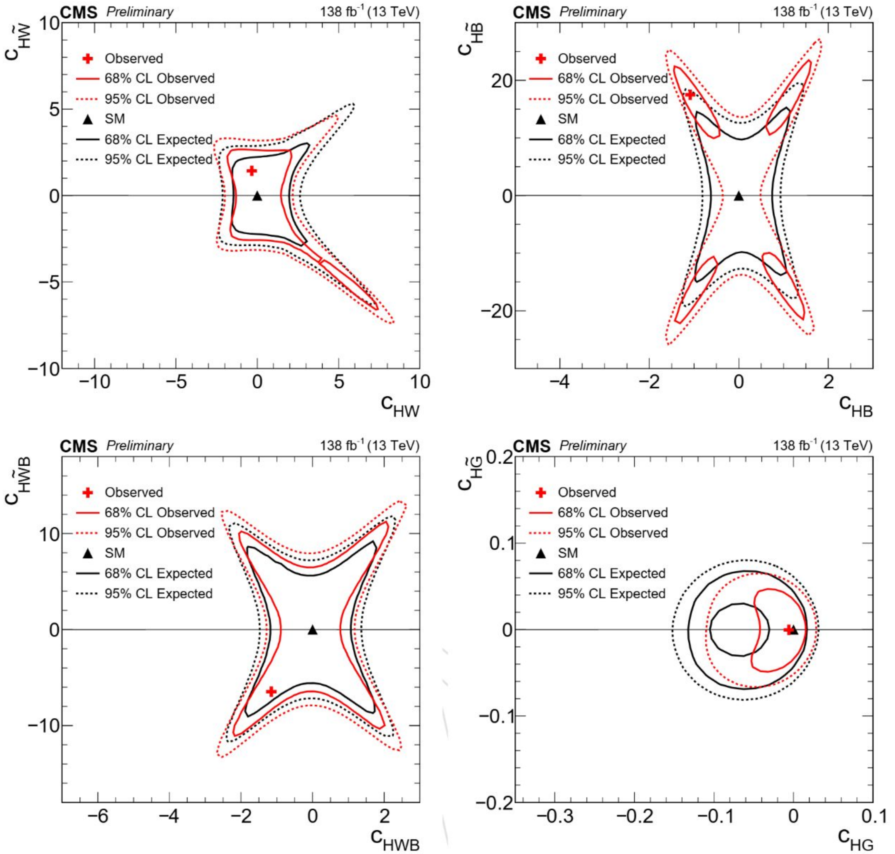
Asimmetria

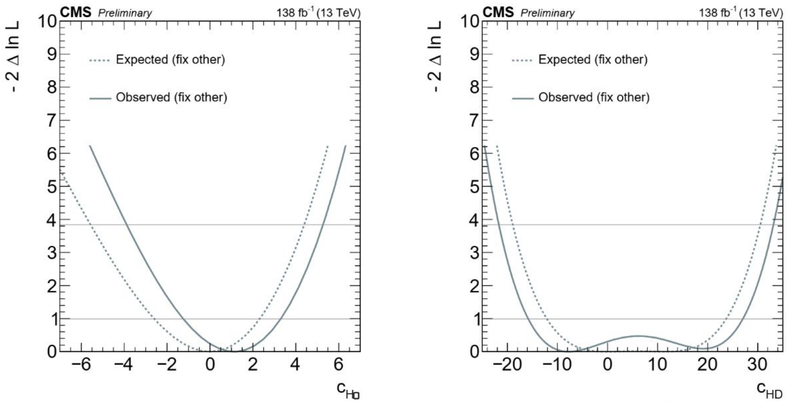
- Asymmetry of the $\Delta\Phi_{jj}$ distribution

$$A = \frac{N(\Delta\Phi_{jj} > 0) - N(\Delta\Phi_{jj} \leq 0)}{N(\Delta\Phi_{jj} > 0) + N(\Delta\Phi_{jj} \leq 0)}$$

A = 0.4 $\pm$ 0.3 (tot. unc.)

Wilson coefficients results (1)





Wilson coefficients results (2)

Wilson Coefficients		Observed	Expected	Significance
$c_{H\Box}$	Best fit	1.14	0.00	0.5
	68% CL	[-1.26,3.31]	[-2.60,2.33]	
	95% CL	[-3.88,5.27]	[-5.56,4.39]	
c_{HD}	Best fit	-7.76	0.00	0.0
	68% CL	[-15.8,27.0]	[-12.1,23.7]	
	95% CL	[-21.7,33.1]	[-19.0,30.5]	
c_{HW}	Best fit	-0.34	0.00	0.0
	68% CL	[-1.46,1.15]	[-1.17,1.76]	
	95% CL	[-2.35,3.57]	[-2.07,6.50]	
c_{HWB}	Best fit	-1.14	0.00	0.5
	68% CL	[-1.92,1.82]	[-1.47,1.50]	
	95% CL	[-2.44,2.34]	[-2.06,2.10]	
c_{HB}	Best fit	-1.09	0.00	1.7
	68% CL	[-1.35,-0.72] ∪ [0.92,1.51]	[-0.80,0.96]	
	95% CL	[-1.56,1.75]	[-1.13,1.27]	
c_{HG}	Best fit	-0.01	0.00	0.0
	68% CL	[-0.03,0.01]	[-0.10,0.02]	
	95% CL	[-0.08,0.02]	[-0.14,0.03]	
$c_{H\tilde{W}}$	Best fit	1.43	0.00	0.7
	68% CL	[-5.95,-5.22] ∪ [-2.21,2.32]	[-1.87,1.91]	
	95% CL	[-7.12,3.00]	[-5.60,4.53]	
$c_{H\tilde{W}B}$	Best fit	-6.42	0.00	0.9
	68% CL	[-9.84,8.90]	[-7.40,7.52]	
	95% CL	[-12.3,11.4]	[-10.51,10.62]	
$c_{H\tilde{B}}$	Best fit	17.5	0.00	2.3
	68% CL	[-20.6,-14.5] ∪ [13.0,21.1]	[-12.2,11.9]	
	95% CL	[-24.5,-5.9] ∪ [6.0,24.1]	[-17.3,16.6]	
$c_{H\tilde{G}}$	Best fit	-0.000	0.00	0.0
	68% CL	[-0.035,0.034]	[-0.062,0.061]	
	95% CL	[-0.058,0.057]	[-0.075,0.073]	

Reinterpreted cross sections