Imperial College London

Faculty of Natural Sciences

February Report

Imperial College London

Measuring Higgs boson properties in the diphoton channel and simplified template cross sections

This brief report summarises Higgs boson property measurements at CMS, specialising to the diphoton channel. The CMS apparatus is introduced and photon reconstruction is defined. Following this, a description of the most recent analysis performed by CMS is given, which precisely measures signal strengths for the different Higgs boson production mechanisms. Finally, the report introduces the simplified template cross section framework; a natural progression to signal strength analyses which reduces the theoretical uncertainties directly folded into previous measurements.

Author:
Jonathon Langford

Supervisor:
Dr. Nick Wardle

22nd February, 2018

1 Introduction

In 2012, the CMS and ATLAS experiments at the Large Hadron Collider (LHC) [1, 2] announced the discovery of a new particle consistent with the standard model (SM) Higgs boson [3,4]. The SM of particle physics is an incredibly powerful theory which triumphs in explaining a vast range of high-energy physics data [5–7]. Observing the Higgs, the quanta of a scalar field which mechanises electroweak symmetry breaking [8–10], marked the completion of the theory. However, there are a number of physical phenomena that the SM fails to incorporate, such as gravitation and dark matter. Thus, new physics beyond-the-standard model (BSM) is required. As a consequence, it is of paramount importance to scrutinize the SM in as many channels as possible and measure parameters of the theory with unprecedented accuracy.

Since the Higgs discovery, the CMS experiment has collected a large set of proton-proton collision data at an increased centre-of-mass energy ($\sqrt{s}$ = 13 TeV), and reconstruction and analysis techniques have dramatically improved. Coupling these experimental achievements with the theoretical motivation of fully understanding the dynamics of electroweak symmetry breaking, have led to comprehensive studies on the properties of the Higgs boson. Recent analyses from CMS and ATLAS include limit setting on the Higgs width [11], and measurements of the particle's mass [12], production and decay rates, coupling strengths [13], and CP properties [14, 15]. No significant deviations from the predictions of the SM have been observed. Furthermore, direct searches for BSM physics manifesting as an extended Higgs sector have provided no clear evidence of a signal [16].

The $H \to \gamma\gamma$ decay channel has played a major role in the discovery of the Higgs boson and the subsequent measurements of its properties [17]. Although the decay has a relatively small branching fraction in the SM ($\approx$0.2%), low background and good energy resolution allow for a precise reconstruction of the invariant mass peak. The most recent analysis in the diphoton channel was performed on 36 fb^{-1} of proton-proton collision data collected by the CMS experiment in 2016 [18]. Events are classified according to Higgs production mechanisms to improve sensitivity, and allow measurements of individual production mode signal strengths.

The outline of this report is as follows. A brief description of the CMS detector is provided in section 2, focussing on parts specific to the $H \to \gamma\gamma$ channel. Section 3 covers the most recent analysis in detail, explaining the parameters of interest and how they were measured. Section 4 introduces a novel measurement framework known as simplified template cross-sections (STXS) which allows for a systematic reduction in the theoretical uncertainties associated with the measurements [19]. Finally, a brief remark on how this aids in combining results across channels, and ultimately across both CMS and ATLAS, is given.

2 The CMS detector and photon reconstruction

A brief summary of the CMS detector configuration is provided in this section. Following this, the reconstruction and identification of photon candidates is described. It is important to note that the reconstruction of all other physics objects[1] is required in the $H \to \gamma\gamma$ analysis, but has not been included in this report. A complete description of the CMS experiment, including the reconstruction and identification of other particles can be found in reference [1].

The CMS apparatus was designed to precisely measure and accurately reconstruct physics objects originating from proton-proton collisions at a centre-of-mass energy $\mathcal{O}$(10 TeV) and at luminosities up to 10^{34} cm^{-2} s^{-1}. The detector exhibits a symmetric cylindrical geometry. At the core lies a superconducting solenoid, which immerses various particle detection systems in a 3.8 T axial magnetic field. The steel return yoke outside of the solenoid is large enough to accommodate four stations of gas-ionization muon detectors which provide an almost 4π coverage in solid angle. The bore of the solenoid holds a silicon pixel and strip tracking system, a lead tungstate crystal electromagnetic calorimeter (ECAL) and a brass and scintillator hadron calorimeter (HCAL), in order of increasing radius from the interaction point. Both calorimeters consist of a barrel and two endcap detectors providing full azimuthal coverage within the pseudorapidity region, $|\eta| < 3$. Steel/quartz-fibre forward calorimeters extend this coverage to $|\eta| < 5$.

The particle flow algorithm is implemented for global event reconstruction [20]. This incorporates the information from all sub-detectors in an attempt to reconstruct and identify all individual particles in an event. Photon candidates are identified in the algorithm as clusters of energy in the ECAL with no corresponding trajectory detected in the tracking system. Excellent energy resolution is provided by the ECAL

[1] This includes electrons, muons, light quark and bottom quark jets and missing energy.

via a highly granulated structure consisting of 75,000 2cm × 2cm $PbWO_4$ crystals. The clustering algorithm benefits from this structure, enabling an almost complete determination of photon energy; including those which convert to e^+e^- pairs before reaching the ECAL.

The clustering algorithm consists of three steps. Initially, cluster "seeds" are established when the maximum energy of a local deposit in the ECAL is above a given threshold. Next, topological clusters are grown from the seeds by aggregating crystals that share a common side with a cell already in the cluster. The added crystals are required to have an energy in excess of two standard deviation of the noise in the ECAL (80 MeV in the barrel and 300 MeV in the endcaps). Exploiting the granularity of the ECAL, the energy deposit in each crystal is shared among clusters assuming a Gaussian transverse profile of the electromagnetic shower. Lastly, "superclusters" are built by dynamically merging together clusters.

3 Recent analysis measuring Higgs properties in the $H \to \gamma\gamma$ channel

This section describes the most recent analysis performed by CMS in the diphoton channel [18], which is performed on 35.9 fb^{-1} of proton-proton collision data at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. Events are selected by diphoton triggers with transverse energy thresholds of 30 and 18 GeV for the highest and second highest energy photons respectively. The shower shape, isolation and ratio between hadronic and electromagnetic deposits of the photon candidates must satisfy loose requirements in the trigger selection. Monte-Carlo (MC) event generators are used to simulate both signal and background events [21–24], where the dominant backgrounds to $H \to \gamma\gamma$ are irreducible prompt diphoton production and reducible backgrounds from γ + jet and jet + jet events, such that the jets are misidentified as photons. The GEANT4 package is utilised to simulate the response of the CMS detector [25], and the simulated events are weighted to reproduce the distribution of the number of interactions in data. On top of this, $Z \to e^+e^-$ events in which the electron showers are falsely reconstructed as photons, are used to calibrate the MC samples in terms of the energy scale and resolution.

The final stage analysis employs the following methodology to measure the parameters of interest. Initially, diphoton candidate events must satisfy preselection criteria, and a Boosted Decision Tree (BDT) classifier calculates a photon ID BDT score which isolates prompt photons from mis-identified jet fragments. Events are then classified into discrete regions of phase space, targeting different Higgs production mechanisms and according to their mass resolution and signal-to-background ratio. In each category, a signal model is constructed from simulation to parametrise the peak in the diphoton mass distribution, $m_{\gamma\gamma}$. The $m_{\gamma\gamma}$ value is calculated per event using the following equation,

$$m^2_{\gamma\gamma} = 2E_{\gamma 1}E_{\gamma 2}(1 - \cos\phi), \tag{1}$$

where $E_{\gamma 1}$ and $E_{\gamma 2}$ are the energies of the leading and sub-leading decay photons respectively, and ϕ resembles the angle between the photons in the decay plane. A separate method extracts the background model from data. After incorporating systematic uncertainties, the combined signal plus background models for each category are simultaneously fit to each $m_{\gamma\gamma}$ distribution. Note, equation 1 shows it is imperative to accurately determine the primary vertex from which the photons originate, as this has a direct effect on the ϕ measurement and subsequently, the diphoton invariant mass resolution. If the position of the vertex along the beam axis is known to better than 1 cm, then the $m_{\gamma\gamma}$ resolution is dominated by the photon energy resolution. Thus, to ensure precise vertex assignment, the analysis uses another BDT with input features associated with tracks recoiling against the diphoton system.

The main parameter of interest is the signal strength, μ, defined as the ratio of the measured Higgs cross section to the SM prediction ($\sigma/\sigma_{\mathrm{SM}}$). Furthermore, the signal strengths for each production mechanism, μ_i are measured, where i labels gluon-gluon fusion (ggH), vector boson fusion (VBF), vector boson associated production (VH) or production in association with top quarks (ttH). The best-fit signal strength values along with the associated uncertainties are extracted from a likelihood fit. A separate framework is also employed to interpret the data. This κ framework introduces specific modifications to Higgs boson couplings allowing for deviations from SM predictions [26, 27]. The results of the κ interpretation are not presented in this report.

3.1 Event classification

Event classification is used to improve sensitivity and gain access to individual Higgs boson production mode signal strengths. This entails separating events according to production mechanism, diphoton mass resolution and the signal-to-background ratio. To evaluate the $m_{\gamma\gamma}$ resolution on a per-event basis, a diphoton multivariate classifier is implemented which uses information on the diphoton kinematics, vertex scenario and the individual photon ID BDT scores as input. The diphoton BDT, which gives a high score for events with a signal-like topology and good mass resolution, is used as one of the major elements in event classification and selection.

The classification process uses the other reconstructed physics objects in an event. Each category is defined by a dedicated set of criteria which are based on the expected topology and kinematics of the objects in an event of that type. Furthermore, the diphoton BDT score is used to separate some categories according to the $m_{\gamma\gamma}$ resolution.

Higgs bosons produced in association with a top quark pair (ttH) can be identified via the two b quarks from the decay of the top quarks. Events of this type are further categorised depending on the type of top decay; semi-leptonic (ttHLeptonic) or hadronic (ttHHadronic), where the presence of charged leptons or additional jets are required respectively. For clarity, the semi-leptonic category defines events in which atleast one W boson from the top decay, decays to a lepton + neutrino. Whereas the hadronic mode selects events in which both W bosons decay to jets. Associated production with a vector boson (VH) feature leptons, missing transverse momentum or jets from the W or Z decay. Selection criteria for this production mechanism are again formulated to split into further categories depending on the type of vector boson decay, by exploiting the kinematics of the additional physics objects. For example, leptonic Z decays (ZHLeptonic) demands two same flavour isolated leptons above a p_T threshold, with a dilepton invariant mass in the Z mass region. Other sub-categories, with respective criteria, are leptonic W decays (WHLeptonic), Z or W leptonic decays with a relaxed selection (VHLeptonicLoose), Z or W leptonic decays with at least one missing lepton (VHMET) and vector boson hadronic decays (VHHadronic). Vector boson fusion (VBF) production features two final state jets with large rapidity separation. In addition to standard critera on the jet characteristics, to isolate VBF events from gluon-gluon fusion (ggH), a multivariate classifier is trained on the jet kinematics using ggH + jets as background. The output score is then combined with the diphoton BDT score in an additional multivariate classifier to select VBF events. Moreover, the score of this combined discriminant is used to separate VBF events into three sub-categories according to the purity (VBF 0, 1 and 2).

Events which do not fall into the aforementioned categories are divided into four "untagged" categories using the score of the diphoton BDT. This categorises the remaining ggH events according to the $m_{\gamma\gamma}$ resolution and signal-to-background ratio. The event classification process applies the criteria in order of highest priority; the order in which the categories are defined in this section.

3.2 Signal and background modelling and systematics

To model the signal peak in the diphoton invariant mass distribution of each category, a shape is constructed from simulated events. In practice, the selection criteria described in section 3.1 do not perfectly isolate events of a given type, with some leakage across production modes into the different categories[2]. On top of this, since the primary vertex assignment has a dramatic effect on the $m_{\gamma\gamma}$ distribution width, separate fits are made for events which have correct vertex assignment within 1 cm, and those which do not. Thus, for each process, category and vertex scenario, the diphoton peaks are fitted using a convolution of at most five Gaussian functions, and parametric descriptions of the signal shapes are obtained in terms of the Higgs mass, m_H. The final signal fit in each category is obtained by summing the fits for the production processes and vertex scenarios and normalising to the expected yields in the respective category.

The discrete profiling method is implemented to model the background distribution from data [28]. As the functional form of the background is not known, the discrete profiling method treats the uncertainty associated with the choice of background parametrisation as a nuisance parameter in the fit. Many candidate function families are fit to the background $m_{\gamma\gamma}$ distribution[3] by minimising the value of twice the negative

[2] See table 3 and figure 14 in reference [18], which show the expected number of signal events per category for each production process.

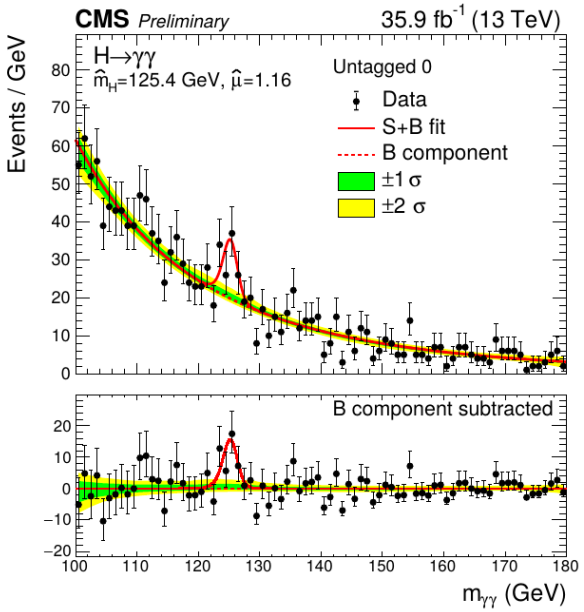
[3] i.e. outside of the signal region: $120 < m_H < 130$ GeV.

logarithm of the likelihood (2NLL). The discrete profiling method then combines the background functions by using the overall best 2NLL when measuring a parameter of interest.

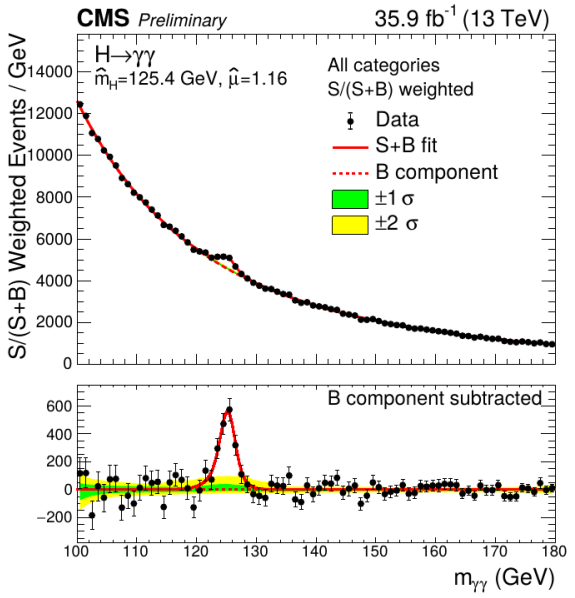
A comprehensive set of systematics are considered in the analysis, including both theory and experimental uncertainties. The systematics are categorised by their effect on the signal distribution, and treated accordingly. Those which directly alter the shape are incorporated into the model as nuisance parameters. Systematics which do not modify the shape, are handled as uncertainties on category yields. Finally, if a systematic disturbs an input in the classification criteria, it is treated as a variation on the category yield to represent migration between categories. The dominant systematics, yielding a few percent impact on the results, include the photon shower shape modelling, photon and jet energy scales, integrated luminosity, and theory uncertainties.

3.3 Results and discussion

A simultaneous fit of the signal plus background models is performed to all categories, allowing the individual signal strengths, μ_i, and m_H to vary. Figures 1a and 1b show the fit to data for the best resolution "untagged" category and the sum of all categories respectively. Note, the categories in the combined $m_{\gamma\gamma}$ distribution have been weighted by their respective sensitivity. The best fit values, extracted from the simultaneous fit, provide an overall signal strength of $\mu = 1.16$, and a Higgs mass, $m_H = 125.4$ GeV.



(a) The best resolution "untagged" category.



(b) All categories combined and weighted by their respective sensitivity.

Figure 1: Signal plus background model fits to data. The one (green) and two (yellow) standard deviation bands represent the uncertainties in the background component of the fit, performed using the discrete profiling method. The residual plots underneath, show the $m_{\gamma\gamma}$ distribution after background subtraction.

Likelihood scans for μ and all μ_i are performed, in which all other nuisances, including the value of m_H, were profiled. Figure 2a shows the scan for the overall signal strength, which provided a value of $\hat{\mu} = 1.16^{+0.15}_{-0.14} = 1.16^{+0.11}_{-0.10}$ (stat.) $^{+0.09}_{-0.08}$ (syst.) $^{+0.06}_{-0.05}$ (theo.). The results of the scans for each production mechanism are presented in figure 2b, where the important results are the following. The best fit signal strength for the ttH production mode relates to a 3.3σ excess over the no signal hypothesis, and is compatible within 1.6σ of the SM prediction. Likewise, the VH and VBF channels were observed to have 2.4σ and 1.1σ excesses respectively, with respect to the absence of that production mode. All results are highly compatible with SM predictions.

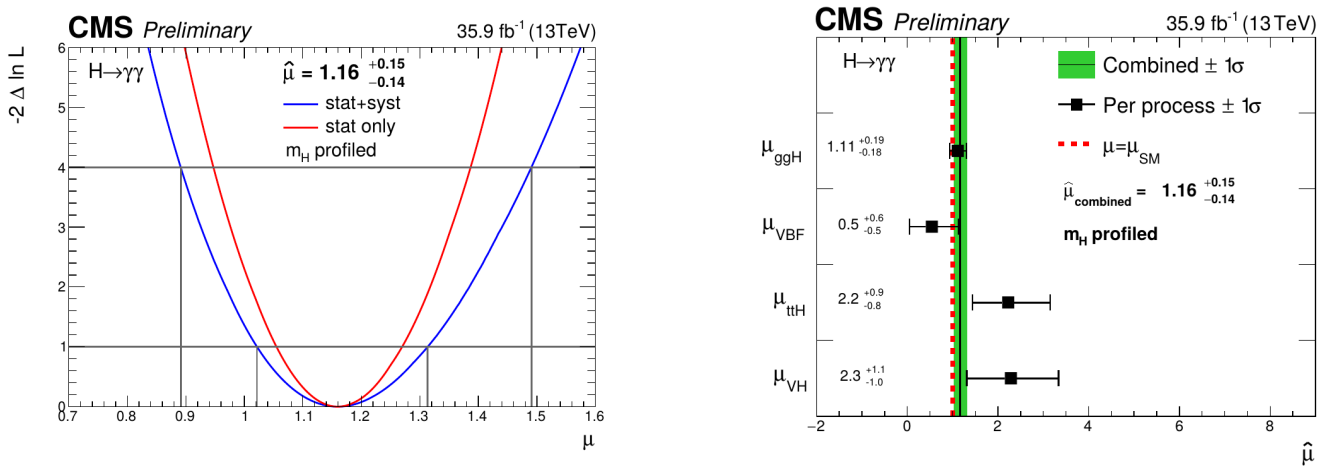


Figure 2: (Left) Likelihood scan for the overall signal strength, μ, where m_H is profiled in the fit. (Right) Individual signal strengths in each production mode from likelihood scans. The green band shows the overall signal strength, and the red dashed line represents the SM prediction.

4 Simplified template cross sections

Measurements of the Higgs boson couplings in different Higgs decay channels[4] [29], follow a similar structure to the $H \to \gamma\gamma$ analysis described in section 3. Such analyses are designed to interpret data via the measurements of signal strengths and multiplicative coupling modifiers in categories designed to be enriched in events of a given Higgs production mode. However, measurements of this type directly incorporate theoretical uncertainties in terms of SM predictions and the underlying physics model.

The simplified template cross-sections (STXS) framework has been developed to progress Higgs boson coupling measurements in the future [19]. STXS aim to maximise experimental sensitivity whilst minimizing the theory dependencies in the measurements. These demands are realised by measuring cross sections instead of signal strengths in fiducial volumes of the inclusive Higgs boson production phase space. Thus, theoretical dependencies from SM predictions are diminished. Figure 3 portrays a diagram of how a volume of phase space is selected.

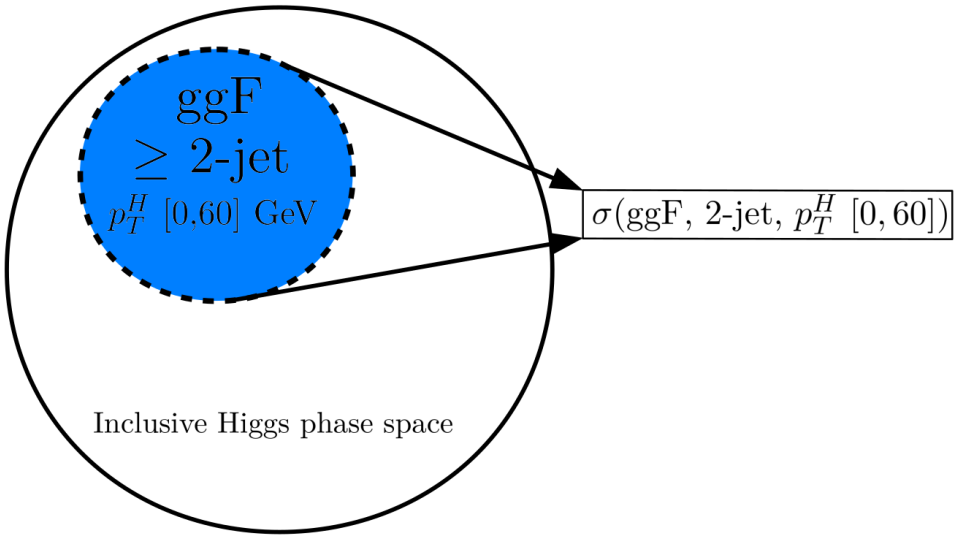


Figure 3: A schematic for choosing a fiducial volume in the inclusive Higgs phase space. The volume selected isolates ggH (labelled here as ggF) production events with at least two jets and Higgs transverse momentum, p_T^H between 0 and 60 GeV. In the STXS framework, the cross section is measured in this mutually exclusive region of phase space.

Whilst optimizing for experimental sensitivity, the fiducial volumes, called "bins" for simplicity, are cho-

[4]These decay channels are namely Higgs boson decaying to ZZ, WW, bb, $\tau\tau$ and $\mu\mu$.

sen with the following principles. Like in previous analyses, Higgs production mechanisms are separated to eliminate production mode dependencies. Bins are aligned with experimental categories to avoid measurements that require extrapolation to larger regions of phase space. STXS also demands a roughly constant experimental acceptance in each bin to obviate direct dependencies on the underlying theoretical distribution in the simulation. Furthermore, a priority on the theory side is to separate bins with the largest theory systematics and to isolate those which may have increased sensitivity to BSM enhancements. In spite of these requirements, to avoid introducing uncontrolled theoretical uncertainties by selecting too specific phase space regions, analyses which employ multivariate techniques are carefully checked.

The project has been divided into several stages which define the way in which the framework evolves with increasing statistics. Employing these techniques allows for a systematic reduction of the theoretical uncertainties in increasingly finely-grained bins. Stage 0 bin definitions have close correspondence to the current event classification described in section 3.1, whereas stage 1 begins to further split into regions of kinematic phase space. For example, the stage 1 binning for ggH production is shown in figure 4, where events are initially split by number of jets, and then further sub-divided by the Higgs p_T. In terms of timescale, stage 1 is designed to be fully implemented by the end of LHC run 2. Following this, Stage 2 is planned to be integrated during run 3, and be fully achieved at the HL-LHC [30]. The bin definitions for stage 2 are to be decided after gaining experience from previous stages.

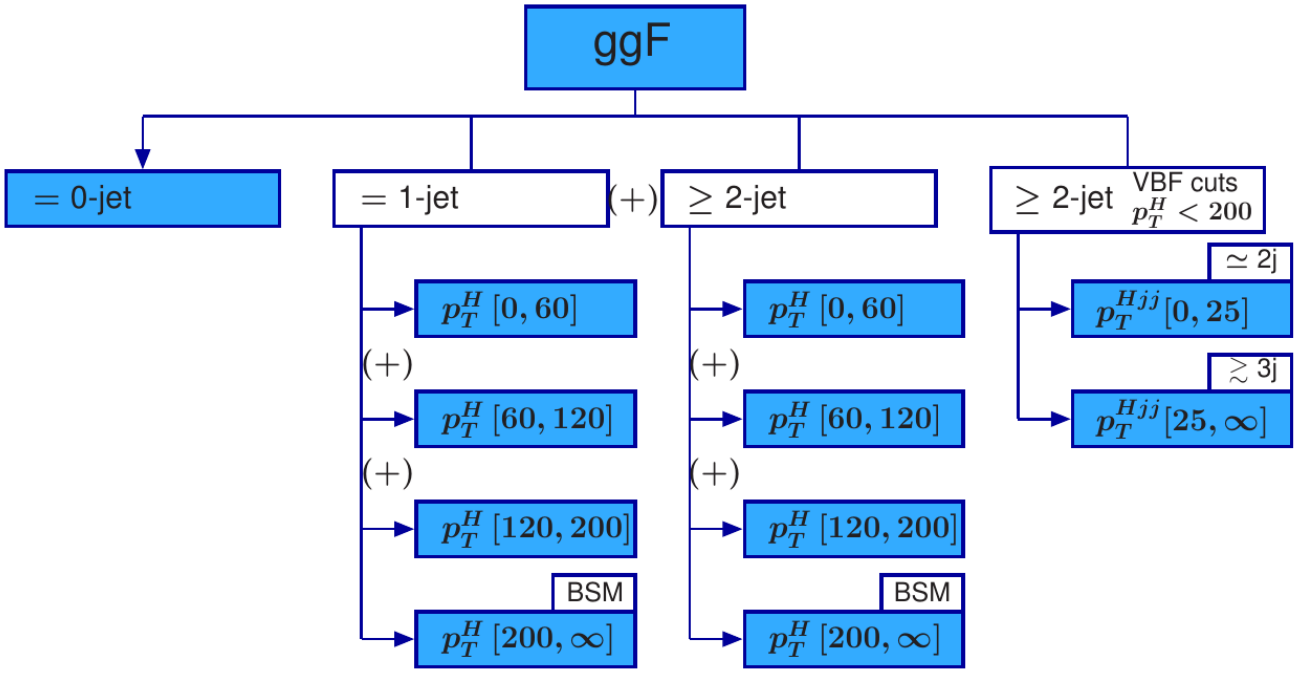


Figure 4: Stage 1 binning for ggH production (labelled here as ggF). The blue boxes represent the final bins in which the cross sections are measured.

The framework effectively combines the advantages of signal strength measurements and differential measurements to provide an interface between measurement and interpretation. The treatment of theoretical uncertainties is shifted to the interpretation step by decoupling the raw measurement. Finely grained bins provide more information for theoretical interpretation, from which one can determine signal strength or coupling modifiers and test specific BSM models. On top of this, by decoupling the measurement, advances in theoretical predictions can be easily propagated directly to the interpretation step. Due to this property, measurements remain long-term useful. Finally, one of the major advantages of the STXS framework is the ability to easily combine the measurements of different decay channels, provided analyses implement the same bin structure. Thus, sensitivity would be greatly improved. Moreover, if employed by both CMS and ATLAS, STXS provides an efficient method for combining measurements across collaborations.

5 References

[1] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[2] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[3] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012.

[4] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012.

[5] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579–588, 1961.

[6] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967.

[7] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc.*, C680519:367–377, 1968.

[8] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321–323, 1964.

[9] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, 1964.

[10] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global Conservation Laws and Massless Particles. *Phys. Rev. Lett.*, 13:585–587, 1964.

[11] Constraints on the Higgs boson width from off-shell production and decay to ZZ to llll and llvv. Technical Report CMS-PAS-HIG-14-002, CERN, Geneva, 2014.

[12] Georges Aad et al. Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments. *Phys. Rev. Lett.*, 114:191803, 2015.

[13] Georges Aad et al. Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV. *JHEP*, 08:045, 2016.

[14] Serguei Chatrchyan et al. Study of the Mass and Spin-Parity of the Higgs Boson Candidate Via Its Decays to Z Boson Pairs. *Phys. Rev. Lett.*, 110(8):081803, 2013.

[15] Georges Aad et al. Evidence for the spin-0 nature of the Higgs boson using ATLAS data. *Phys. Lett.*, B726:120–144, 2013.

[16] Philip Bechtle, Sven Heinemeyer, Oscar Stal, Tim Stefaniak, and Georg Weiglein. Applying Exclusion Likelihoods from LHC Searches to Extended Higgs Sectors. *Eur. Phys. J.*, C75(9):421, 2015.

[17] Vardan Khachatryan et al. Observation of the diphoton decay of the Higgs boson and measurement of its properties. *Eur. Phys. J.*, C74(10):3076, 2014.

[18] Measurements of properties of the Higgs boson in the diphoton decay channel with the full 2016 data set. Technical Report CMS-PAS-HIG-16-040, CERN, Geneva, 2017.

[19] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2016.

[20] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. (CMS-PAS-PFT-09-001), 2009.

[21] J. et al Alwall. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[22] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.

[23] Vardan Khachatryan et al. Event generator tunes obtained from underlying event and multiparton scattering measurements. *Eur. Phys. J.*, C76(3):155, 2016.

[24] T. Gleisberg, Stefan. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter. Event generation with SHERPA 1.1. *JHEP*, 02:007, 2009.

[25] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth.*, A506:250–303, 2003.

[26] Chiara Mariotti and Giampiero Passarino. Higgs boson couplings: measurements and theoretical interpretation. *Int. J. Mod. Phys.*, A32(04):1730003, 2017.

[27] Margherita Ghezzi, Raquel Gomez-Ambrosio, Giampiero Passarino, and Sandro Uccirati. NLO Higgs effective field theory and κ-framework. *JHEP*, 07:175, 2015.

[28] P. D. Dauncey, M. Kenzie, N. Wardle, and G. J. Davies. Handling uncertainties in background shapes. *JINST*, 10(04):P04015, 2015.

[29] Vardan Khachatryan et al. Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV. *Eur. Phys. J.*, C75(5):212, 2015.

[30] G. Apollinari, O. Brüning, T. Nakamoto, and Lucio Rossi. High Luminosity Large Hadron Collider HL-LHC. *CERN Yellow Report*, (5):1–19, 2015.