UCLouvain

Université catholique de Louvain
Secteur des Sciences et Technologies
Institut de Recherche en Mathématique et Physique
Centre for Cosmology, Particle Physics and Phenomenology

Application of deep learning techniques in CMS : from matrix element regression to the search for Higgs boson pair production

Doctoral dissertation presented by

Florian Bury

in fulfilment of the requirements for the degree of Doctor in Sciences

Thesis support committee

Prof. Céline Degrande	UCLouvain, Belgium
Prof. Christophe Delaere (Supervisor)	UCLouvain, Belgium
Prof. Vincent Lemaitre (Chair)	UCLouvain, Belgium
Dr. Rainer Mankel	DESY, Germany
Prof. Pascal Vanlaer	ULB, Belgium

December, 2022

Remerciements

"No cause is lost, if there is but one fool left to fight for it."
— Captain Jack Sparrow

J'aimerais tout d'abord remercier Christophe pour m'avoir encadré et formé pendant les cinq années qui m'ont mené de mon mémoire de master jusqu'à la présentation de cette thèse. Merci pour tes conseils, pour m'avoir progressivement laissé prendre mon indépendance, et permis d'apprendre de ta vaste expérience.

I want to express my deepest gratitude to my other mentor Agni. For all the things I have learnt, not only about the double Higgs searches or analysis techniques, but also about collaboration, personal and career developments. I could not have dreamt of a better colleague to work through all the challenges and shenanigans that were foolish enough to get in our way.

Je voudrais aussi remercier les autres membres de mon jury, Céline, Vincent, Pascal et Rainer, pour leurs commentaires éclairés pendant et après la défense privée. Ils m'ont poussé à m'intéresser à des aspects que je n'avais pas considérés, ainsi qu'à perfectionner ce manuscrit et mes connaissances.

Cette thèse n'en serait pas à ce stade non plus sans de nombreux collègues que j'ai eus le plaisir de côtoyer dans les couloirs et les bureaux du CP3. Pieter pour son aide technique toujours précieuse et quasi instantanée. Martin pour tout ce qu'il m'a appris avec son mélange atypique de cynisme et de bonne humeur, ainsi que nos aventures vécues à transmettre notre passion aux étudiants. Julien pour les années que nous avons passées depuis les bancs du bac jusqu'aux bureaux de thésards, et pour m'avoir offert à moi, expérimentateur, l'autre vision du monde de la physique vu par un phénoménologiste. Sébastien pour tes conseils et ton expertise toujours avisée. Pietro pour ta bonne humeur communicative et ton point de vue toujours intéressant sur la physique et les physiciens. Anna pour les discussions et corrections sur les jets, la physique et tout autre sujet. Alessia, Suat et Khawla pour avoir partagé avec moi les péripéties de la thèse et m'avoir permis de m'amuser sur quelques tâches de vos analyses. Maxime pour m'avoir fait prendre goût au monde de la muographie. À mes collègues de bureau qui m'ont supporté pendant toutes ces années, en particulier Sandhya pour m'avoir invité dans son petit groupe d'analyse dans lequel je me suis beaucoup amusé. Merci à tous les participants du frag, ce fut pour moi un honneur d'avoir été votre referee, et un plaisir de vous avoir poussé aux challenges les plus absurdes, néanmoins relevés avec panache.

Évidemment, rien de tout cela n'aurait été possible sans le soutien indéfectible de ma famille. Merci à mes parents qui je crois n'ont jamais vraiment compris sur quoi je travaillais, mais ont toujours cru en moi et m'ont soutenu sans parfois que je m'en rende compte sur le moment. J'espère qu'ils seront fiers du travail accompli dans cette thèse. Merci aussi à mes petites sœurs pour m'avoir aidé à garder les pieds sur terre.

Merci à tous mes *vrais* amis qui m'ont aidé à préserver une certaine santé mentale aux travers de cette thèse et que je ne saurais tous citer sans une page de plus. Merci à tous les membres de la Route™ et plus généralement ceux que j'ai côtoyés via le scoutisme pour les bouffées d'air frais qu'ont été les moments passés en votre compagnie, à mes coéquipiers hockeyeurs de la meute avec qui j'ai partagé victoires et défaites sans que j'en regrette une seule seconde, à mes colocs, officiels ou non, de Louvain-la-Neuve qui ont vécu aux premières loges mon aventure. Un merci tout particulier à Simon, qui est probablement le non-physicien le plus au courant de tous mes accomplissements et déboires durant cette thèse, un véritable ami sur lequel j'ai toujours pu compter.

Finally, many thanks to my colleagues abroad in the HH $\rightarrow$ bbWW analysis : Gourab, Minuk, Dennis, Peter, Benjamin, Christian, Karl, Saswati and Tao. It has been an intensive journey but I learnt a lot from and with you.

J'aimerais finalement remercier le FRS-FNRS pour l'octroi d'une bourse d'aspirant m'ayant permis de réaliser cette thèse, et à l'UCLouvain pour m'avoir accueilli dans ses locaux et son institution scientifique.

Ces quatre dernières années me semblent à la fois être passées en un éclair et avoir duré une vie entière. L'écriture de cette thèse, espacée sur cette dernière année avec des périodes d'écriture plus ou moins productives, me donne en même temps l'impression d'une tâche herculéenne accomplie, et celle de n'avoir qu'égratigné la surface. Car si je mesure les changements que cette thèse a apportés dans ma vie, et j'en remercie encore toutes les personnes mentionnées ici pour y avoir participé d'une façon ou d'une autre, ce document me semble ironiquement bien maigre pour résumer ces quatre années d'apprentissage. Faute d'avoir de meilleurs mots pour l'exprimer, je laisse une personne plus sage clore ces remerciements.

"Know what's weird ? Day by day nothing seems to change, but pretty soon, everything is different"

— Calvin and Hobbes

Contents

Introduction

Over the last decades, particle physicists have taken the burden of developing the most comprehensive model of the subatomic world and the laws that govern the interactions of the infinitely small. The outcome of this effort consists in the standard model (SM) that encapsulates the theoretical insights of brilliant minds, validated through meticulous observation of various complex experiments by generations of experimental physicists, carefully assembled as a self-consistent theory. The pinnacle of this piece by piece endeavour resides in the discovery of the Higgs boson in 2012 by the CMS and ATLAS collaborations, granting for the first time access to a new scalar sector and a mechanism to provide a mass to all elementary particles, about half a century after its prediction. Despite the many achievements of this theory, there is little doubt that this is not the end of the road. While as a predictive model it can be tested to a very high level of accuracy, several fundamental pieces of the puzzle are still missing. The absence of the gravity, one of the four fundamental interactions, the absence of a candidate for dark matter or the lack of explanation for the matter-antimatter asymmetry are very strong indications that the theory is incomplete.

Particle physicists are currently at the crossroads. In the last ten years since the discovery of the Higgs boson, no major excess has been observed that could provide a direction towards new physics, despite an extensive program of research and large efforts carried out in the experiments located on the Large Hadron Collider (LHC). As we are approaching the major upgrade of the LHC that marks half the lifetime of the accelerator, there is an urge to provide at the very least a significant enough hint as to the direction of research to which efforts should be attributed. In that context, the scalar sector appears as one of the most promising paths, as any massive new particles should manifest themselves through interactions with the Higgs boson. There are however concerns that the energy scale achievable by the accelerator might not be high enough to reach these new states, which could nonetheless be probed through indirect searches and precision measurements of the parameters of the theory. In particular, the shape of the scalar field potential is of utmost interest. One of the parameters governing its shape, the trilinear Higgs coupling, is predicted, but its measure represents an ongoing challenge. Accessible directly through the production of a pair of Higgs bosons (HH), a phenomenon a thousand times fainter than the already ambitious measurement of the single Higgs production, the wide variety of decay channels requires an equally wide collaboration to make the most

out of every promising channel, each coming with its own hurdles and handles over the parameters' measurement. This work will focus on a search carried out using the CMS detector in the bbWW final state at a centre of energy of 13 TeV in which the author was involved, both in terms of a test of the SM measurement and a direct search for an unknown heavy resonance. Additionally, a growing trend arose in particle physics in the development of Effective Field Theory (EFT) allowing for a relatively assumption-free extension of the SM, serving as an intermediate framework between the observations made by experimentalists, and the various models produced by theorists. This direction was also followed in this work in the context of the final state at hand.

The major challenge of precision measurements, and to a lesser extent to direct bump searches, is the control over the measurement uncertainty sources. On one hand, statistical uncertainties can only be alleviated through the collection of more data. On the other hand, making the most of the collected data requires a deeper look at the methods employed in the measurement and their systematic uncertainties. The recent rise of Machine learning (ML) algorithms and in particular the deep learning techniques represents the common thread of the work performed in this thesis. These cutting-edge methods have proved to be valuable companions to understand the flow of data coming from the detector and to reach higher sensitivity than what was deemed possible not so long ago. Through the use of an advanced classification scheme with DNNs, the sensitivity to the double Higgs production in the bbWW channel was improved compared to the previous search. In parallel, a method called the Matrix Element Method (MEM) promises high levels of sensitivity taking into account both our knowledge of the SM and of the detector, with the setback of an expensive computation time that is prohibitive for the amount of data generated by LHC experiments. In this context, a hybrid method using both the MEM and DNNs has been developed to tackle this issue, opening the way for its application into a real-life analysis.

The text is organised as follows. First the theoretical framework of the QFT on which the SM is build is presented, then its application to the physics of the elementary particles, with an emphasis on the scalar sector and more specifically on the double Higgs production mechanism. Different resonant interpretations are listed together with the EFT development particularised to that mechanism. The second chapter is dedicated to the experimental setup, with a brief description of the LHC and a more in depth characterisation of the CMS detector, the various reconstruction techniques and statistical tools. The third chapter will address the ML methods with an emphasis on DNNs, the MEM and the hybrid method that I proposed in a publication [1]. Finally, the fourth chapter starts by a brief overview of the current public HH searches, the rest being dedicated to the CMS data analysis in which I was involved in the bbWW final state and its results. The conclusion and outlook follow with an eye on the horizon of double Higgs searches and methods.

Chapter

1 The standard model of particle physics

Particle physics – also referred to as high-energy physics (HEP) – deals with the most fundamental constituents of matter and their interactions. Although the study presented in this document is strongly oriented towards the experimental research in this field, using large particle accelerators and complicated detector apparatus, a brief overview of the theoretical framework is essential to understand the context of this research.

This chapter will briefly address the current understanding that we have of the world of elementary particles within the Standard model (SM), starting from a brief overview of Quantum field theory (QFT) then the introduction of the SM with its successes and shortcomings, along with an effective way of simulating its interactions. Next we will focus on the double Higgs production (HH) mechanisms, its current understanding within the theory, and potential extensions within EFTs and their limitations.

The first sections are mostly based on Peskin and Schroeder's introduction of Quantum Field Theory [2], Srednicki's book on Quantum Field Theory [3], Introduction to High Energy physics by Perkins [4] and Tilman Plehn's lectures on LHC physics [5].

In this chapter, we will adhere to the following conventions. The metric tensor $g_{\mu\nu}$ signature is $(+,-,-,-)$, such that the four-vectors of position $x^\mu = (t, \boldsymbol{x})$, momentum $p^\mu = (E, \boldsymbol{p})$ and derivatives $\partial^\mu = \partial/\partial x_\mu = (\partial/\partial_t, \nabla)$ transform as $x_\mu = g_{\mu\nu}x^\nu = (t, -\boldsymbol{x})$ and a massive particle has $p^2 = p_\mu p^\mu = E^2 - |\boldsymbol{p}|^2$. The scalar fields will typically be denoted by ϕ, and fermion Dirac spinor with ψ. Symbols relative to particles can be attached to them and inferred from context. γ^μ are the Dirac matrices following the Weyl convention $\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3$, used to define the fermion conjugate $\bar{\psi} \equiv \psi^\dagger\gamma^0$. The totally antisymmetric tensor $\epsilon^{\mu\nu\rho\sigma}$ is determined by $\epsilon^{0123} = +1$.

Two unit systems are used throughout this document. The international system (SI) is the system commonly used in macroscopic structures and its units (kg, s, m, K, etc) will be quoted explicitly. Another system more suited for HEP and elementary particles is based on so-called *natural units*. Equalising $\hbar = c = k_B = \epsilon_0 = \mu_0 = 1$ results in

$$[\text{energy}] = [\text{mass}] = [\text{temperature}] = [\text{length}]^{-1} = [\text{time}]^{-1} = \text{eV}, \tag{1.1}$$

which allows for example to express the mass of the electron as $m_{\text{electron}} = 9.109 \times 10^{-28}\,\text{g} = 0.511\,\text{MeV}$. By default, the natural system is used for all equations in this document and units are expressed in eV ($1\ \text{eV} = 1.602\ 10^{-19}$ J).

1.1 Quantum Field Theory and observables in a nutshell

In classical mechanics, the fundamental element is the action S which is an integral of the Lagrangian, or more specifically in the context of a local field theory the Lagrangian density $\mathcal{L}$ which needs to be a scalar for the action to be Lorentz invariant. If we consider free fields $\phi(x)$, where $x = (t, \mathbf{x})$ represents the space-time coordinates, the action can be written as a functional of the form

$$S = \int \mathcal{L}(\phi, \partial_\mu \phi)\ d^4x. \tag{1.2}$$

The principle of least action specifies that when the states evolve from one configuration at time t_1 to another at time t_2, it must follow a path for which S is an extremum, otherwise expressed as $\partial S = 0$ which yields the Euler-Lagrange equation of motion

$$\partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} - \frac{\partial \mathcal{L}}{\partial \phi} = 0. \tag{1.3}$$

From there one can define the Hamiltonian $H = \int \mathcal{H}\ d^3x$ from the Hamiltonian density defined as

$$\mathcal{H} = \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \partial_\mu \phi - \mathcal{L}. \tag{1.4}$$

The fields were so far defined as free, considering now interactions requires working in the *interaction picture* where the Hamiltonian above is split into a free and an interacting parts $H = H_0 + H_{int}$. The time dependence in operators can be arbitrarily chosen to be governed by H_0 while the time dependence in states would be governed by H_{int}, from which follows in the Heisenberg perspective

$$H_I(t) = e^{iH_0(t-t0)}(H_{int})e^{-iH_0(t-t0)} \tag{1.5}$$

$$|\psi(t)\rangle_I = e^{iH_0(t-t_0)}\,|\psi\rangle\,, \tag{1.6}$$

where $|\psi\rangle_I$ and H_I are the state and interacting Hamiltonian in the interactive picture.

The Schrödinger equation can be developed to yield

$$i\frac{d\left|\psi\right\rangle}{dt} = H\left|\psi\right\rangle \Rightarrow \frac{d\left|\psi\right\rangle_I}{dt} = iH_I\left|\psi\right\rangle, \tag{1.7}$$

which is the equation to solve for the interacting state.

The solution can be written as

$$\left|\psi(t)\right\rangle = U(t,t_0)\left|\psi(t_0)\right\rangle, \tag{1.8}$$

where $U(t,t_0)$ is a unitary time evolution operator such that $U(t_1,t_2)U(t_2,t_3) = U(t_1,t_3)$ and $U(t_0,t_0) = 1$. Eq. (1.7) then becomes

$$i\frac{dU}{dt} = H_I(t)U. \tag{1.9}$$

H_I being an operator, the solution of Eq. (1.9) is given by the *Dyson's formula*

$$U(t,t_0) = T\ \exp\left(-i\int_{t_0}^{t} H_I(t')dt'\right), \tag{1.10}$$

where T is the time-ordering symbol.

In collider experiments we are only interested in cross-section or decay rate computations for which we need to extract the transition probability between an initial state $|i\rangle$ and a final state $|f\rangle$, corresponding to collections of particles with well-defined momenta P_j^{initial} and P_k^{final} ($j \in [1,N]$, $k \in [1,M]$) respectively. An important assumption is then made that these initial and final states are *asymptotically free*, which means that $|i\rangle$ in $t \to -\infty$ and $|f\rangle$ in $t \to +\infty$ are eigenstates of the free Hamiltonian H_0, and their interaction as they get closer is only coming from H_{int}. Using these limits with Eq. (1.8) yields

$$\left|f\right\rangle = \lim_{t_\pm \to \pm\infty} U(t_+,t_-)\left|i\right\rangle, \tag{1.11}$$

and the amplitude of the interaction is obtained via projection and the definition of the unitary operator S called the S-matrix

$$\lim_{t_\pm \to \pm\infty} \left\langle f\right| U(t_+,t_-)\left|i\right\rangle \equiv \left\langle f\right| S\left|i\right\rangle. \tag{1.12}$$

Provided the interaction part is "small" compared to the free Lagrangian and using Eq. (1.10), the S-matrix can be developed as a perturbative series

$$S = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \int d^4x_1 ... \int d^4x_n T\{\mathcal{H}_I(x_1)...\mathcal{H}_I(x_n)\}. \tag{1.13}$$

The Wick's theorem ensures that each expression containing the time-ordering symbol can be developed in a finite sum of products, and therefore each term of the series can be computed as an amplitude that can be represented by a Feynman diagram, allowing an easy interpretation of the terms contributing to a certain

process at a given order. The first order of the series can be factored out, and it is customary to write $S \equiv 1 + iT$ where only the second part refers to the scattering with the transition matrix – or T-*matrix* – that we are mostly interested in.

The particular case of scattering experiments involve either two beams of particles meeting head-on or a single beam hitting a target. The measurable observable is the cross-section, which is the rate of the scattering of a certain process divided by the incident particle flux, and can be determined from the interaction probability of the particles. The amplitude associated to the scattering interaction is computed from the T-matrix, factoring out the 4-momentum conservation to separate the "kinematics" that do not depend on the Hamiltonian from the "dynamics" that do

$$\langle \{p_f^{out}\} | iT | p_A^{in} p_B^{in} \rangle = (2\pi)^4 \delta^4 \left(p_A^{in} + p_B^{in} - \{p_f^{out}\} \right) \; i\mathcal{M}(p_{A,B}^{in} \to \{p_f^{out}\}), \tag{1.14}$$

where p_A^{in}, p_B^{in} and $\{p_f^{out}\}$ are the incident and outcoming particles of the scattering momenta, and $\mathcal{M}$ is called the Matrix element (ME).

The probability of interaction can be computed in a small region of the phase-space $d^3p_1 ... d^3p_n$ with

$$\mathcal{P}(\mathcal{A}\ \mathcal{B} \to 1\ 2\ ...\ n) = \left(\prod_{f=1}^{n} \frac{d^3 p_f}{(2\pi)^3} \frac{1}{2E_f} \right) \left| \langle \{p_f^{out}\} | iT | p_A^{in} p_B^{in} \rangle \right|^2 \tag{1.15}$$

that combined with Eq. (1.14) allows to compute the differential cross-section

$$\frac{d\sigma}{d\Phi} = \frac{\left| \mathcal{M} \left(p_{A,B}^{in} \to \{p_f^{out}\} \right) \right|^2}{4 E_A^{in} E_B^{in} \Delta v}, \tag{1.16}$$

where the phase-space measure is given by

$$d\Phi = (2\pi)^4 \delta^4 \left(p_A^{in} + p_B^{in} - \{p_f^{out}\} \right) \prod_f \frac{d^3 p_f}{(2\pi)^3}. \tag{1.17}$$

Note that in Eq. (1.16) $|\mathcal{M}|^2$ is understood as summed over final spins and averaged over initial spin degrees of freedom. Δv is the relative velocity between the incident particles, and that the whole denominator is invariant under Lorentz boost.

Total or partial cross-section can be obtained from Eq. (1.16) by integration over the phase-space, while the differential decay rate (and its total or partial value obtained also by integration) defined as the rate at which unstable particles of mass M at rest decays to a given final state can be obtained similarly with

$$\frac{d\Gamma}{d\Phi} = \frac{\left| \mathcal{M} \left(p_A \to \{p_f^{out}\} \right) \right|^2}{2M}. \tag{1.18}$$

Similarly to the S-matrix, the cross-section can also be expanded in a series depend-

ing on the parameter $\alpha = g^2/4\pi$, supposedly small enough to allow the perturbation theory to apply

$$\sigma = \sigma_{LO} \cdot \left(1 + \frac{\alpha}{2\pi}\sigma_1 + \left(\frac{\alpha}{2\pi}\right)^2 \sigma_2 + ...\right). \tag{1.19}$$

While the lowest-order computation – also called leading order (LO) – is relatively easy, adding higher-level correction can improve the accuracy of the theory predictions. The first term of Eq. (1.19) is called next-to-leading order (NLO), the second Next-to-next-to-leading order (NNLO) and so forth. Their computation however involve solving diagrams which contain loops through which arbitrary high momenta can flow through an infinite integral. These divergences – called ultraviolet divergences – are taken care of by redefining the field and constants with counter-terms. This is referred to as *renormalisation*. During this procedure, the "bare" couplings of Eq. (1.19) are replaced by "constants" that acquired a dependence on an unphysical energy, the *renormalisation scale* μ_R^2.

The "running" of the coupling constant with the scale is given by the Beta function $\beta(\alpha) = d\alpha/d(\ln \mu_R^2)$ that arises from the Callan-Symanzik equation. While the exact cross-section σ (hypothetically obtained non-perturbatively) is independent of the renormalisation scale, the fact that the infinite series has to be truncated to allow computations introduces a dependence on this parameter. Including higher orders allows mitigating this effect. The typical procedure to estimate this uncertainty is to fluctuate this scale by a factor of two around some nominal value linked to the process at hand (the mass of a resonance, for example), which is deemed sufficient most of the time but remains an arbitrary choice.

1.2 Standard Model and Symmetries

1.2.1 Fundamental particles

In the beginning of the previous century, only a handful of particles were known, namely the electron and the proton, soon followed by the neutron. It was then discovered that the picture was far from complete, with the muon and neutrino, and much later that the proton and neutron were actually composite particles made up of quarks bound by gluons.

All matter as far as we know is made up of fundamental spin-$\frac{1}{2}$ particles that are called fermions. We further differentiate leptons that can be neutral or carry an integer electric charge, from the quarks that carry a fractional electric charge. The interactions between these particles are mediated by exchange of integer spin particles called gauge bosons, carriers of the fundamental interactions. *Gravity* is the weakest of them by a large account and is still not accounted for in the current formulation of the SM, the *electromagnetic* interaction between charged particles

is mediated by the photon (γ), the *weak force* notably responsible for the β-decay of the nuclei is carried by the W and Z bosons, finally the *strong force* acts on the quarks to bind them inside the nucleus and is carried by the gluons (see Table 1.1). All the fermions and bosons have antiparticle counterparts, possessing the exact same properties except for an opposite charge.

The quarks are classically divided into three generations with similar features. Each consists of two quarks with electric charges $+\frac{2}{3}$ and $-\frac{1}{3}$. They are also charged under the strong interaction, this charge is called the *colour*. They go as follows : up (u) and down (d); charm (c) and strange (s); top (t) and bottom (b) quarks. Leptons are defined using the same convention, each generation consisting in two leptons where one exhibits no electric charge – a neutrino – while the other has a charge of -1 : electron neutrino (ν_{e}) and electron (e); muon neutrino (ν_{μ}) and muon (μ); tau neutrino (ν_{τ}) and tau (τ). A summary of these particles is on Table 1.2.

The weak force has been shown to couple to pairs of leptons or quarks through the mediation of the $\mathrm{W}^{\pm}$ and Z bosons, leading to a source of flavour violation in the SM. In particular, the $\mathrm{W}^{\pm}$ bosons explain the decay of the second and third generations of fermions into the first, and the instability of the neutron. The Z boson on the other hand is involved in what are called neutral currents, similar to the ones obtained from the electromagnetic interaction except that the charge is replaced by the flavour, allowing its interaction with neutrinos. In addition, these decays have been demonstrated to only apply to left-handed particles, leading to the following decomposition (hypothetical right-handed neutrinos were omitted in this description).

$$L^1 = \begin{pmatrix} \nu_{\mathrm{e}\,L} \\ \mathrm{e}_L \end{pmatrix} \quad L^2 = \begin{pmatrix} \nu_{\mu\,L} \\ \mu_L \end{pmatrix} \quad L^3 = \begin{pmatrix} \nu_{\tau\,L} \\ \tau_L \end{pmatrix} \tag{1.20}$$

$$Q^1 = \begin{pmatrix} \mathrm{u}_L \\ \mathrm{d}_L \end{pmatrix} \quad Q^2 = \begin{pmatrix} \mathrm{c}_L \\ \mathrm{s}_L \end{pmatrix} \quad Q^3 = \begin{pmatrix} \mathrm{t}_L \\ \mathrm{b}_L \end{pmatrix} \tag{1.21}$$

$$E^{1,2,3} = \mathrm{e}_R, \mu_R, \tau_R \quad U^{1,2,3} = \mathrm{u}_R, \mathrm{c}_R, \mathrm{t}_R \quad D^{1,2,3} = \mathrm{d}_R, \mathrm{s}_R, \mathrm{b}_R \tag{1.22}$$

Table 1.1 | List of fundamental interactions and their associated boson mediators in the SM, their masses and what they act on. The mass values are taken from the world average measurement average [6]. The gluons, photon and graviton (which is still hypothetical) are supposed massless and the experimental upper limit is quoted between parentheses.

Interactions	Gravitational	Electromagnetic	Weak		Strong
Mediator particle	Graviton	Photon (γ)	$\mathrm{W}^{\pm}$ bosons	Z boson	Gluons (g)
Mediator mass	0 ($< 6 \times 10^{-32}\,\mathrm{eV}$)	0 ($< 1 \times 10^{-18}\,\mathrm{eV}$)	$80.4\,\mathrm{GeV}$	$91.2\,\mathrm{GeV}$	0 ($< 1.3\,\mathrm{MeV}$)
Acts on	Mass-energy	Electric charge	Flavour charge		Colour charge

Table 1.2 | List of fermions in the SM with their masses and charges, split per generation. The exact mass values and neutrino upper limits are taken from the world measurement average provided by the particle data group [6].

Generation	Leptons			Quarks		
	Flavour	Mass	Electric charge	Flavour	Mass	Electric charge
First	electron neutrino (ν_e)	$<1.1\,\mathrm{eV}$	0	up (u)	$2.16^{+0.49}_{-0.26}$ MeV	$+2/3$
	electron (e)	$511\,\mathrm{keV}$	-1	down (d)	$4.67^{+0.48}_{-0.17}$ MeV	$-1/3$
Second	muon neutrino (ν_μ)	$<0.19\,\mathrm{MeV}$	0	charm (c)	1.27 ± 0.02 MeV	$+2/3$
	muon (μ)	$106\,\mathrm{MeV}$	-1	strange (s)	93^{+1}_{-5} MeV	$-1/3$
Third	tau neutrino (ν_τ)	$<18.2\,\mathrm{MeV}$	0	top (t)	$172.9 \pm 0.4\,\mathrm{GeV}$	$+2/3$
	tau (τ)	$1.78\,\mathrm{GeV}$	-1	bottom (b)	$4.18^{+0.03}_{-0.02}$ GeV	$-1/3$

1.2.2 Gauge invariance

The Dirac Lagrangian for free massive fermions can be written as

$$\mathcal{L}_{Dirac} = \bar{\psi}(i\partial\!\!\!/ - m)\psi, \tag{1.23}$$

where ψ is the fermion field of mass m and $\partial\!\!\!/ = \gamma_\mu \partial^\mu$. In this equation, the first part is the kinetic term and the second is the mass term.

If we require the Lagrangian to be invariant under a local $U(1)$ transformation of the field

$$\psi(x) \to e^{i\alpha(x)}\psi, \tag{1.24}$$

we need to assume the existence of a spin-1 vector field $A_\mu(x)$ that transforms as

$$A_\mu(x) \to A_\mu(x) - \frac{1}{g}\partial_\mu(x), \tag{1.25}$$

and define the covariant derivative

$$D_\mu = \partial_\mu + igA_\mu(x). \tag{1.26}$$

In addition, the field strength tensor

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{1.27}$$

is also invariant under this transformation, and the interaction term for spin-1 particles can be used in the Lagrangian of quantum electrodynamics (QED)

$$\begin{aligned}\mathcal{L}_{QED} &= -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \bar{\psi}(iD\!\!\!/ - m)\psi, \\ &= -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \bar{\psi}i\gamma^\mu(\partial_\mu + igA_\mu)\psi - m\bar{\psi}\psi,\end{aligned} \tag{1.28}$$

which is effectively invariant under $U(1)$ transformations, and required the introduction of a bosonic field A_μ.

1.2.3 Standard Model Lagrangian

In a more general way, a local gauge invariant Lagrangian under a certain group requires a covariant derivative $D_\mu = \partial_\mu - igA^a_\mu T^a$, where g is the gauge coupling, A^a_μ is the connection and T^a is the group generator. A gauge field tensor can consequently be built as

$$igF^a_{\mu\nu}T^a = [D_\mu, D_\nu]. \tag{1.29}$$

The Standard model (SM) Lagrangian is based on this gauge invariance principle, requiring local invariance on $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, where C refers to the colour, L to the left-handed fermions and Y to the weak hypercharge.

The group $SU(3)_C$ describes Quantum chromodynamics (QCD) and is responsible for the strong force. Its eight generators, given by the Gell-Mann matrices λ_a (with $\mathrm{Tr}(\lambda^a\lambda^b) = 2\delta^{ab}$, $a = 1, ..., 8$), give rise to eight gauge bosons G^a_μ. The covariant derivative is given by

$$D^{ij}_\mu = \partial_\mu - ig_s G^a_\mu \frac{\lambda^{ij}_a}{2}, \tag{1.30}$$

and the field tensor can be derived from Eq. (1.29), which in this case yields

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f^{abc} G^b_\mu G^c_\nu, \tag{1.31}$$

where f^{abc} is the structure constant of the group and g_s the coupling constant of the strong interaction.

The QCD Lagrangian can therefore be written

$$\mathcal{L}_{QCD} = -\frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu} + \bar{\psi}_i \left(i \not{D}^{ij} \right) \psi_j. \tag{1.32}$$

Similarly, the $SU(2)_L \otimes U(1)_Y$ groups describe the weak and electromagnetic interactions, their covariant derivatives acting differently on left- and right-handed fermions (based on chirality)

$$D_\mu = \partial_\mu - ig\frac{\sigma_i}{2} W^{\mu i} - ig'\frac{Y}{2} B_\mu \tag{1.33}$$

$$D'_\mu = \partial_\mu - ig'\frac{Y}{2} B_\mu, \tag{1.34}$$

where g and g' are the two coupling constants, Y the hypercharge, σ_i the Pauli matrices generators of the $SU(2)$ group (with $\mathrm{Tr}(\sigma_i\sigma_j) = 2\delta^{ij}$, $i = 1, 2, 3$), while $W^{\mu i}$ and B_μ are the three and one spin-1 bosons associated to the $SU(2)_L$ and $U(1)_Y$ groups respectively. Their field tensors can be built accordingly, following Eq. (1.29) :

$$W^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g\epsilon^{ijk} W^j_\mu W^i_\nu \quad , \quad B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{1.35}$$

where ϵ is the fully antisymmetric structure constant of the $SU(2)$ group.

The gauge and fermionic part of the SM Lagrangian can be written as

$$\mathcal{L}_{gauge} = -\frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a - \frac{1}{4}W^i_{\mu\nu}W^{\mu\nu}_i - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \tag{1.36}$$

$$\mathcal{L}_{fermion} = \sum_{\psi,I} \bar{\psi^I}(i\bar{D})\psi^I, \tag{1.37}$$

with the various fields $\psi = L, Q, E, U, D$ from Eqs. (1.20) to (1.22) and I representing the three generations as defined on Table 1.2.

Contrary to Eq. (1.28) however, a mass term of the type $m\bar{\psi}\psi$ would break gauge invariance under $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$. Thus, so far all the particles are massless and have to acquire their mass through a different mechanism.

1.2.4 Electro-weak symmetry breaking mechanism

It has been experimentally known for a long time that most fermions – first the leptons and later the quarks – and some bosons – $W^{\pm}$ and Z – are massive particles while the current Lagrangian density that was built up to now only allows massless particles due to gauge invariance. To reconcile the Lagrangian with the observation, a complex scalar field $\phi = (\varphi^+ \ \varphi^0)^T$, doublet under $SU(2)$, with an hypercharge $Y = 1$ and no colour, has been postulated simultaneously in Refs. [7–9] and later in Ref. [10]. The mechanism through which the leptons, quarks, and the bosons acquire their mass is therefore called the *Brout-Englert-Higgs* mechanism. The Lagrangian associated to this scalar field is

$$\mathcal{L}_{scalar} = D_\mu\phi^\dagger D^\mu\phi - V(\phi^\dagger\phi). \tag{1.38}$$

The potential $V(\phi^\dagger\phi)$ was parameterised in the most general renormalisable formulation respecting gauge symmetry

$$V(\phi^\dagger\phi) = -\mu^2(\phi^\dagger\phi) + \lambda(\phi^\dagger\phi)^2, \tag{1.39}$$

with $\mu^2 > 0$. The minimum of this potential is displaced to $\langle 0|\,\psi^\dagger\psi\,|0\rangle = \mu^2/\lambda \equiv v^2$, where v is the vacuum expectation value (VEV) of the scalar field and its non-zero value breaks the electro-weak group $SU(2)_L \otimes U(1)_Y$. Expanding the field around the VEV with a polar parameterisation yields

$$\phi(x) = \frac{1}{\sqrt{2}}\exp\left(i\sigma^i\xi_i(x)\right)\begin{pmatrix}0\\ v+h(x)\end{pmatrix} \xrightarrow{EWSB} \phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v+h(x)\end{pmatrix}, \tag{1.40}$$

where the $\xi_i(x)$ are Goldstone bosons, and the Electroweak symmetry breaking

(EWSB) has been performed. The covariant derivative of that field is

$$D_\mu = \partial_\mu - ig\frac{\sigma_i}{2}W^{i\mu} - i\frac{g'}{2}B_\mu. \tag{1.41}$$

Developing further the Lagrangian yields

$$\begin{aligned}\mathcal{L}_{scalar} &= D_\mu\phi^\dagger D^\mu\phi + \mu^2(\phi^\dagger\phi) - \lambda(\phi^\dagger\phi)^2 &(1.42)\\ &= \frac{v^2}{8}\left(g^2W_\mu^iW^{i\mu} + g'^2B_\mu B_\nu - 2g'gB_\mu W^{3\mu}\right)\left(1+\frac{h}{v}\right)^2 \\ &\quad + \frac{1}{2}\left(\partial_\mu h\partial^\mu h\right) - \lambda v^2h^2 - \lambda vh^3 - \frac{\lambda}{4}h^4 - \frac{\lambda v^4}{4}. &(1.43)\end{aligned}$$

The first line of Eq. (1.43) can be interpreted as a mass term for the $\mathsf{W}^{1,2,3}$ and B gauge bosons, its last term shows a mixing between the W^3 and the B fields. The physical bosons can be defined from the gauge bosons :

$$\mathsf{W}_\mu^\pm = \tfrac{1}{\sqrt{2}}(\mathsf{W}_\mu^1 \mp i\mathsf{W}_\mu^2) \tag{1.44}$$

$$\begin{pmatrix}\mathsf{Z}_\mu\\ \mathsf{A}_\mu\end{pmatrix} = \begin{pmatrix}\cos\theta_W & -\sin\theta_W\\ \sin\theta_W & \cos\theta_W\end{pmatrix}\begin{pmatrix}\mathsf{W}_\mu^3\\ \mathsf{B}_\mu\end{pmatrix} \tag{1.45}$$

where θ_W is the Weinberg angle defined as $\tan\theta_W = g'/g$. In the end, the four degrees of freedom of this new scalar field ϕ have been absorbed as a mass term for the scalar boson and the following mass terms for the weak bosons, while the photon remains massless,

$$\begin{aligned}m_\mathsf{W} &= \tfrac{gv}{2},\\ m_\mathsf{Z} &= \sqrt{g^2+g'^2}\,\tfrac{v}{2},\\ m_\mathsf{A} &= 0.\end{aligned} \tag{1.46}$$

The weak boson masses are related through the $m_\mathsf{Z} = m_\mathsf{W}/\cos\theta_W$. Using these definitions, the first line of Eq. (1.43) can be rewritten as

$$\mathcal{L}_{scalar} \ni \left(m_\mathsf{W}^2\mathsf{W}_\mu^+\mathsf{W}^{-\mu} + \tfrac{1}{2}m_\mathsf{Z}^2\mathsf{Z}_\mu\mathsf{Z}^\mu\right)\left(1+\frac{h}{v}\right)^2. \tag{1.47}$$

Not only does the part of the scalar Lagrangian in Eq. (1.47) provides a prediction for the masses of the weak bosons, it also describes their interaction with the Higgs boson.

The second line of Eq. (1.43) can be interpreted as an extra scalar particle that uses the remaining degree of freedom to obtain a mass given by $m_\mathsf{H} = \sqrt{2\lambda}v$, this particle is called the Higgs boson. The experimental value for its mass is $m_\mathsf{H} = 125.09 \pm 0.24$ GeV [11] and the VEV can be accessed through the measurement of

the weak bosons masses in Eq. (1.46) with a value of $v \simeq 246$ [6]. Together they allow a prediction of the value of λ and consequently the following terms implying trilinear ($\lambda v h^3$) and quartic Higgs couplings ($\frac{\lambda}{4} h^4$).

A mass term for the fermions would break gauge invariance, but they have been observed as massive particles. To reconcile the theory with the experiment, one can let the scalar field ϕ couple to the fermions via Yukawa terms

$$\begin{aligned}\mathcal{L}_{Yukawa} &= -\sum_i Y_\ell^{ij} \left(\bar{L}^i.\phi\right) \ell^j + \text{h.c.} \qquad \text{(leptons)} \\ &= -\sum_{ij} Y_{\mathsf{u}}^{ij} \left(\bar{Q}^i.\phi^C\right) \mathsf{u}^j - \sum_{ij} Y_{\mathsf{d}}^{ij} \left(\bar{Q}^i.\phi^C\right) \mathsf{d}^j + \text{h.c.} \quad \text{(quarks)},\end{aligned} \tag{1.48}$$

where $\phi^C = i\sigma_2\phi^*$. $Y_\ell, Y_{\mathsf{u}}, Y_{\mathsf{d}}$ are the Yukawa couplings of the leptons (electron, muon, and tau), the up-type quarks (up, charm, top) and down-type quarks (down, strange, bottom). More specifically, they are matrices in the flavour space.

After EWSB in Eq. (1.40), the lepton part of the Yukawa interaction becomes

$$\mathcal{L}_{Yukawa} \in -\frac{Y_\ell v}{\sqrt{2}} \left(\bar{\ell}_L \ell_R + \bar{\ell}_R \ell_L\right) \left(1 + \frac{h}{v}\right). \tag{1.49}$$

This part of the Lagrangian displays a new mass term for the charged leptons given by $m_\ell = Y_\ell v/\sqrt{2}$ where interestingly $Y_t \sim 1$, and an interaction term between pairs of lepton-antilepton and the Higgs boson. Fig. 1.1 provides an experimental validation of the scalar field couplings. Because for leptons there is a single Yukawa matrix used to rotate the left-handed leptons, it can be diagonalised in the flavour space to provide a unique Yukawa coupling value for each flavour. While the expression is similar for quarks, since there are two matrices Y_{u} and Y_{d} they cannot be both diagonalised and contrary to leptons the mass and flavour eigenstates cannot all be associated. By convention the matrix for the up-type quarks is rotated and therefore the relations between the mass and flavour eigenstates of the down-type quarks are given by a non-diagonal matrix called the Cabibbo–Kobayashi–Maskawa (CKM) matrix. The elements of this matrix can be described with 4 parameters, one of them being a CP-violating phase.

Additionally, to the interaction with the Higgs boson in the type of Eq. (1.49), the EWSB implies that leptons interact with the W and Z bosons and the photon with flavour diagonal terms. While the same applies for quark, the impossibility to diagonalise both Yukawa matrices implies the modification of the charged currents with the W boson and apparition of the transitions between up- and down-type of quarks.

In total, the SM contains a total of 17 free parameters for which the model does not provide any prediction, such that they need to be determined experimentally. Additionally, to the local symmetries that have been described before, there are

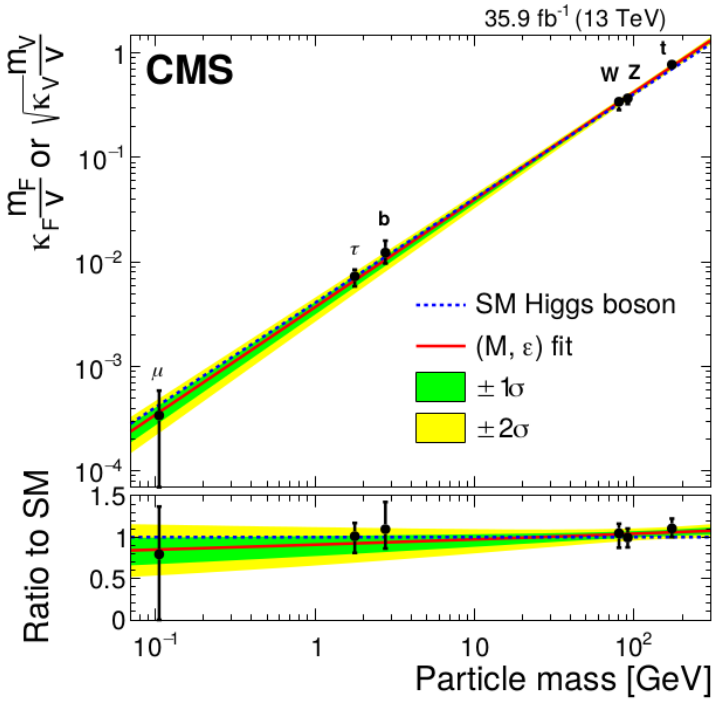


Figure 1.1 | Measurement of the coupling of the Higgs with several fermions through Yukawa coupling – namely the top and bottom quarks, and the tau and muon leptons – and the weak bosons W and Z, performed by the CMS collaboration [12]. The agreement with the SM predictions of Eqs. (1.47) and (1.49) validates the inclusion of the scalar field into the Lagrangian to provide a mass to the elementary particles.

global symmetries that have been determined fairly early by experiments, such as *lepton number* conservation (one per generation) and the *baryon number*. However, these were not imposed from the beginning, so they are commonly referred to as *accidental* symmetries.

The absence of experimental evidence of right-handed neutrinos implies that a Yukawa term cannot be included in the Lagrangian and that no interaction can happen between neutrinos and the scalar field. In the context of the SM the neutrinos are therefore massless after EWSB, which contradicts experimental evidence of neutrino oscillations that have the prerequisite of massive neutrinos. One possible way to reconcile the theory with the experiment is to include the right-handed neutrino in Eq. (1.48), but similarly to the quarks case, one ends up with an additional Yukawa matrix Y_ν that cannot be diagonalised conjointly with Y_ℓ. A new matrix, the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix, needs to account for the mixing between the neutrino flavour and mass eigenstates. While the CKM matrix is almost diagonal, experimental evidence shows the mixing between

neutrinos is almost maximal. The non-diagonal terms allow significant transition probabilities between neutrino flavours, an effect observed in neutrino experiments. If the PMNS matrix was to be included in the SM, seven additional parameters need to added to the free parameters of the model, one of which is the CP-violating phase of the matrix. It is also possible that neutrinos are Majorana particles, since they could potentially be their own antiparticles.

1.3 Event modelling and generation

One of the main challenge in collider experiments from the theory point of view is the computation of inclusive observables such as the total cross-section or the decay rate of not only the *hard process* computed from the Feynman diagrams but also all the additional phenomena attached to the collision that need to be accounted for. For example, the hadronisation and the decays of short-lived particles that cannot be measured in a detector. In hadron colliders such as the LHC where protons or heavy ions are used it is a very challenging task.

For the strong interaction, the perturbation expansion assumptions break down at low energies. This is illustrated in the running of the strong coupling constant $\alpha_s = g_s/4\pi$ which can be expressed at first order as

$$\alpha_s(\mu^2) = \frac{\alpha_s(\mu_0^2)}{1 - \alpha_s(\mu_0^2) b_0 \log \frac{\mu^2}{\mu_0^2}}, \quad b_0 = \frac{1}{12\pi}(2n_f - 33), \tag{1.50}$$

where $n_f = 5$ is the number of massless quark flavours in a good approximation, although $n_f = 6$ yields the same conclusion. Consequently, $b_0 < 0$ and the strong coupling decreases as the renormalisation scale μ increases. At high energies, the coupling is small enough to warrants the use of the perturbation theory for QCD and more importantly validates the approximation of the truncation of the infinite series. In the limit where the energy becomes infinite, the strong force does not apply and the quarks and gluons are free, these particles are therefore defined as *asymptotically-free* [13, 14]. On the other hand, when the energy is below a certain scale, at around $\Lambda_{QCD} \simeq 200$ MeV, the coupling becomes infinite, and one can no longer assume the high order terms in the perturbation series are small enough to be discarded. For example, the current measurement around the mass of the Z boson is $\alpha_s(m_Z) = 0.1181 \pm 0.0011$ [6].

When quarks or gluons are produced in the hard process, they emit additional partons, a phenomenon called Final State Radiation (FSR). The process of emitting these additional particles is called *parton shower*, it does not depend on the initial hard process and how the initial partons are generated. This property of universality and the fact that the energy scales between the high-momentum transfers of the hard process and the lower energy processes occurring inside a hadron allow its

factorisation in the computations of the cross-section. During the development of the shower, the renormalisation factor decreases and fewer partons are generated until the cut-off value Λ_{QCD} is reached, and the quarks are confined within a hadron of neutral colour charge. The hadrons and potential decay products can then be measured in the detector as a collimated cluster called a *jet*. The challenge within the experiment is to infer the properties of the initial quark or gluon from the collection of energy deposits of the hadrons within the jet inside the detector.

To achieve this, it is useful to produce simulations and compare them to data. In proton-proton collisions, only one parton in each of the protons interact. To simulate that, parton distribution functions (PDFs) represent the probability $f_a(x)$ of having an approximately free parton with flavour a carrying a fraction $x \in [0, 1]$ of the initial proton energy. The flavour a can correspond to gluons, valence quarks as well virtual quarks (also called *sea* quarks) inside the hadron, although at large momentum fraction the latter have a lower probability because of their virtualness, as illustrated on Fig. 1.2. These probability distributions are independent of the ME. Going to higher order, the partons can undergo Initial State Radiation (ISR) and generate soft particles before going into the hard process. This entails a renormalisation of the PDF and similarly to the coupling constant the introduction of a *factorisation scale* μ_F whose dependence is also an artefact due to the truncation of the perturbative series. The measured values of the PDFs as well as associated uncertainties (both theoretical and experimental) and the dependence on the factorisation scale can be obtained from collider data, for example the NNPDF collaboration [15].

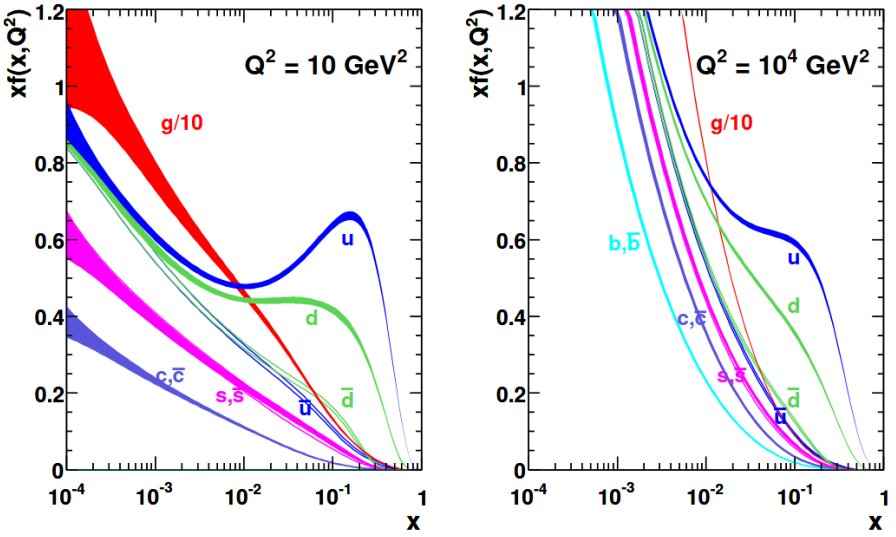


Figure 1.2 | Parton distribution functions (PDFs) at two different scales Q^2 and NLO in the strong coupling α_s, from Ref. [16].

Uncertainties coming from the renormalisation and factorisation scales μ_R and μ_F require proper care by the experimentalists. The *ad-hoc* procedure consists in varying both scales independently by a factor two around their nominal values, yielding six variations (excluding the extreme cases where one is varied up and the other down) plus the nominal case. They can be either considered separate nuisance parameters or considered together as an envelope where the maximal variations in both directions is considered as the only nuisance parameter.

The factorisation assumptions allow computing the microscopic cross-section of the collision between two hadrons integrated over the PDFs

$$d\sigma_{hh\to X} = \sum_{a,b} \int dx_a dx_b \; d\Phi \; f_a(x_a, \mu_f^2) f_b(x_b, \mu_f^2) \; d\sigma_{a,b\to X}(x_a, x_b, \mu_R^2), \quad (1.51)$$

where X is an arbitrary final-state and $d\sigma_{a,b\to X}(x_a, x_b, \mu_R^2)$ is the scattering cross-section of two partons a and b coming from the two hadrons, whose fraction of momentum and flavour are integrated and summed over respectively.

A collision in a hadron-hadron collider (referred to as an *event*) from its interaction to the detection of the final state products can involve up to hundreds of different particles. The properties of these collision final states are of utmost importance for an accurate measurement of the inclusive observables, their simulation thus requires a step-by-step iteration handled by Monte-Carlo generators to produce a full history of the event, as represented on Fig. 1.3.

The first step in the simulation of a collision event is the computation of the parton-level differential cross-section in Eq. (1.16) at a given perturbation order and particle multiplicity, although typically generators are limited to $O(10)$ particles due to combinatorial challenges. Computing the ME at a given order includes a certain number of diagrams which features a factorial growth with the multiplicity of the final states. Then the interaction probability can be obtained by summing and averaging over several degrees of freedom such as the helicity and the colour states, and squaring by taking into account the interactions between all the diagrams. Some burden of the computation can be alleviated by using helicity amplitude methods that bring down the complexity from a quadratic to a linear dependence on the number of diagrams. This is typically done numerically with libraries such as HELAS [18] for SM particles with a few extensions. More automatic ways exist for any physics model with ALOHA [19]. Physics models are numerical equivalents containing fields, parameters, and a Lagrangian, using FeynRules [20] or Sarah [21]. The integration over the phase space in Eq. (1.17) needs to be done numerically. Some techniques have a very good convergence rate in a few dimensions, but it degrades very fast in higher-dimension integrations, which is typically the case of ME. Monte-Carlo integration techniques do not suffer from this issue and are the tool of choice. Adaptive methods such as importance sampling in numerical integrators like Vegas [22] allow further improvement on the convergence, which

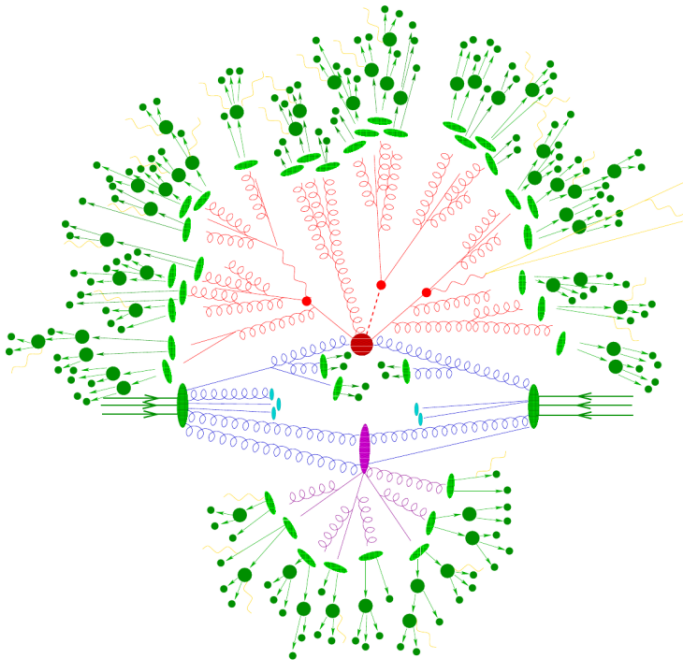


Figure 1.3 | Summary representation of the modelling of a collision event in a hadron-hadron collider, illustrating the complexity of the different mechanisms at play that need to be simulated in Monte-Carlo generators, from Ref. [17]. The hard process is represented by the red blob and its subsequent decays and FSR are in red. Initial state partons and ISR are in blue, while the light green ellipses represent the point where the hadronisation process begins. From there, the produced hadrons and their decays are in dark green. Additionally, the MPI and UE are in pink and light blue.

can however be rather hard to achieve given the complexity of the functions that can present multiple sharp peaks due to propagator enhancements. This can be solved by efficiently reparameterising the phase space over intermediate invariants, and in the case of multiple propagators via a multichannel method, this will be discussed more in depth in Section 3.2.2. The sampling points can then either be used to estimate the inclusive cross-section using the classical numerical integration technique, or be interpreted as the simulated hard scattering of two partons. In the latter case, the PDF needs to be taken into account to simulate the partons probability according to Eq. (1.51). Each event has been sampled with a measure that the algorithm adaptively tried to get as close as possible to the function to integrate to reduce the variance, but is nonetheless not identical. Therefore, each event comes with a certain weight that needs to be taken into account when summing to get the total cross-section. These *weighted* events can then be *unweighted* using a Monte-Carlo acceptance method, so they all have σ/N weights.

After the hard process has been simulated, the parton shower needs to be accounted for. Under the factorisation assumption, both mechanisms can be separated, and the parton shower is assumed to be totally independent of the underlying hard process. Each parton emerging from either the initial or final states will be dressed by a certain number of radiative corrections under the form of additional partons to emerge as a collimated jet, and intermediate resonances decays also need a proper simulation. These emissions and decays are treated as random occurrences, dictated numerically by a Markov-chain Monte-Carlo algorithm. The order of the emissions is based on an *evolution variable* that can differ between the different available libraries, for example p_{T} ordering for PYTHIA [23, 24] and angular ordering for HERWIG [25, 26]. The shower starts right after the hard scattering with an energy scale relative to the process (e.g. the mass of the resonance) and evolves all the way down until a cut-off value is reached. At each iteration, a random number is pulled and compared to the decay probability given by a Sudakov form factor [27]. This factor provides the probability that a parton evolves between two scales without a resolved radiation. Based on this trial, the shower is either stopped, or an additional parton is emitted with a fraction of the initial energy. This fraction is governed by splitting functions that are spin and colour dependent, such as the Altarelli–Parisi functions [28]. They govern the soft and collinear divergences of the emission process, and allow generating the values allocated to the degrees of freedom of the split parton products. This ensures that the parton shower is unitary and does not change the value of the cross-section in the computation. Additional requirements of parton shower algorithms are infrared- and collinear-safety, namely that splitting a parton into a set of partons or adding a parton with vanishing energy, such that the total energy is conserved, should not impact the identified jet configuration.

Additional partons, whether they come from ISR or FSR, need to follow a few conditions such as colour and four-momentum conservation. A classical strategy of *global-recoil* consists in using the large number of partons produced during the parton shower to recoil off the additional radiation of the gluon, such as to satisfy the required conservation. Another approach consists in a *dipole-recoil* [29] where a pair of matching colour–anticolour partons, one of them acting as the radiator and the other as recoiler, are used to preserve the dipole 4-momentum. Depending on the affiliation of the radiator and recoiler to the initial (I) or final (F) states, this defines four types of dipoles : II, IF, FI, and FF. In the IF and FI cases, a colour line is drawn from an initial state to a final state. This dipole setup is common in most parton shower generators, in particular in PYTHIA for the FSR states (therefore FF and FI). The ISR however uses by default global-recoil, which is fine for II but poses a problem of consistency between IF and FI. A recent update of PYTHIA includes the dipole-recoil strategy for the initial-final state colour flow that takes into account the colour connection between the incoming and outgoing partons. This is however not the default strategy, a point that will become relevant further in this document.

At the end of the parton shower, no additional radiation can be created and the partons energy does not allow them to be considered free any more. The simulation must therefore transition to colour neutral hadrons and mesons. As there is no microscopic description, only effective non-perturbative models can be used, whose parameters need to be tuned by dedicated experiments and can be used in many other cases thanks to the universality of the hadronisation. Several of those models exist : the string fragmentation, the cluster fragmentation models and the independent fragmentation models, along with many variants and hybrids. The string model – also referred to as the "Lund" model [30] – is used in PYTHIA and based on the linear confinement assumption, where the energy stored in a colour dipole field increases with the separation distance. The separation between the quark and antiquark in the dipole can be seen as a colour flux tube or "string" that is massless and without transverse degree of freedom. As the separation between the pair increases, the energy stored in the string increases until the system is split into two pairs of colour singlets via quantum mechanical tunnelling. The process is repeated until the hadrons are on shell and only mesons made up of quark-antiquark pairs remain. The model becomes more complex when gluons are included and modelled as having two strings attached, or when baryons need to be accounted for and a description with multiple strings is required. The cluster model [31], used in HERWIG, is based on the colour pre-confinement property [32] that states that partons at the end of the parton shower tend to form colour-singlet clusters in a limited part of the momentum and coordinate phase space. These clusters can be interpreted as excited hadron resonances and decayed into observed hadrons if their mass is small enough. In case of heavy resonances, they are first split into lighter clusters before the decay.

In parallel to the hard scattering, the remaining partons of the colliding hadrons that did not undergo the hard scattering can interact and the products of these multiple parton interactions (MPIs) are propagated through hadron shower, hadronisation, and the colour interplay with the initial and final states (colour reconnection). We denote the result of all these processes occurring in parallel to the hard scattering as the Underlying Event (UE). Similarly to hadronisation and PDF, they need to be tuned by experimental data.

In many experimental cases, the leading order predictions might not be precise enough, either because the uncertainties of the renormalisation scale are dominant, or because the LO cannot properly address the whole kinematic range and collinear products. In this case, a computation of the cross-section at NLO is required. It now consists in three parts : the so-called Born contribution that we used for the LO, the virtual and real corrections. The virtual corrections come with loops that introduce ultraviolet divergences, they can however be dealt with by using counterterms and by renormalising the theory. Infrared divergences are guaranteed to cancel in the summation of the virtual and real corrections, but they only do so after the

integration of the phase-space. A n-body Born contribution causes a real emission of $n+1$ particles. The parton shower on the other end already includes all the higher order corrections to the LO of the ME. When higher orders are to be included in the ME, double counting needs to be taken care of and several automatic approaches are being used for this so-called *matching*. The method used in MG5_aMC@NLO [33] consists in adding a suppression term to the real emission contribution of the ME that was already included in the parton shower, and a correction to the Born contribution to conserve the total cross-section. In that manner, the parton shower needs no correction. The downside is that this suppression gives rise to events with negative weights that degrade the statistical precision, as more simulated NLO events need to be generated compared to LO to compensate the cancellation between positive and negative weights. Direct unweighting is impossible due to the different kinematics of the real and virtual contributions, and on-the-fly histograms must be filled by weighted events to perform the summation. On the other hand, POWHEG [34–36] avoids this issue by replacing the first and hardest order of the parton shower by the exact one from the ME including the real emission. For completeness, it is important to point out that several other event generators exist either at LO or NLO, including parton shower and other features, for example WHIZARD [37] or SHERPA [38].

Similarly, the *merging* method refers to the complementarity of the ME and parton shower, since hard and spatially separated partons are well reproduced by the ME and the parton shower is better suited for lower energy and collinear partons. It is however necessary to avoid double counting and to make sure that the distributions are smooth at the interface of the parton level. This is handled by phase space cuts applied after the shower and hadronisation that define regions where one or the other methods is used, a method also called *slicing*. Several schemes for this purpose have been developed. At LO, the CKKW(-L) [39, 40] and MLM [41] methods reject the components of the ME that produce two very close partons, with some intermediate checks for the former during the shower propagation and the rejection of harder jets from the parton shower. UMEPS [42, 43] is another method derived from CKKWL, with subtraction schemes to ensure the stability of the merging and the inclusive cross-section. At NLO, an additional jet emerges from the ME, the methods developed at this order are thus improvements of the ones at LO, with the rejection of the additional jet configurations of the parton shower, replaced by the ones of the ME. To name a few (without being exhaustive) : NL^3 [44] (based on CKKWL), UNLOPS [45] (based on UMEPS), FxFx [46] (implemented in MG5_aMC@NLO) or MiNLO [47, 48] (implemented in POWHEG).

It is important to note that generating events at higher order quickly becomes expensive in computation time and while this brings an improvement in the modelling of the observable distributions it comes with the price of a degradation of the statistical precision, especially if the subtraction scheme is employed and events

can have negative weights. This trade-off can be a dilemma for experiments that rely heavily on simulations. One common approach is to compute the total cross-section at a high order N^aLO, generate events at a lower order N^bLO (with $b < a$ and usually $b = 0, 1$), and reweight events by a *K-factor* defined as $K = \sigma^{N^aLO}/\sigma^{N^bLO}$. Technically at higher orders the renormalisation scale uncertainty should be reduced, but with the K-factor method there is no way to know how it is affected in the different regions selected from the generated events. This method can however only work when the low order used for the event generation produces differential distributions with a relatively good accuracy and the K-factor only acts on the normalisation. If not, then differential K-factors based on appropriate event kinematics need to be used.

Another computationally too expensive task is the treatment of the decay of heavy resonances via its ME at higher orders. In some cases, the decay width is much smaller than the mass of the resonance ($\Gamma \ll m$) and the narrow width approximation (NWA) can be used to factorise the production cross-section with an on-shell particle from its decay branching ratio (BR). This approximation is based on the fact that with a small width, the on-shell production is strongly enhanced over its off-shell counterpart. When the approximation does not hold, part of the off-shell production mechanism can be recovered as well as the spin correlations by dedicated methods, e.g. MADSPIN [49], or improvements of the approximation [50]. This can happen when the width is large enough that off-shell production becomes significant or when intermediate resonances have a mass gap smaller than their total width such that their interference has a large effect. Although the latter rarely occurs within the SM, it can be the case for its potential extensions. Additionally, kinematic constraints must be taken into account. For example, while most decays of the Higgs boson can be treated with the NWA, the decay into a pair of on-shell Z or W bosons is heavily suppressed since the mass of the Higgs boson is much smaller than the production threshold, and off-shell production cannot be neglected any more.

After event generation, it is possible to modify the event to change its signature with a Matrix element *reweighting*. This method will become relevant in this document and is used to reweight an event initially generated with a process α into a new process β, provided they have the same final state. At LO the formula is

$$w \to \tilde{w} = w \times \frac{|\mathcal{M}_\beta(\{\mathbf{x}\})|^2}{|\mathcal{M}_\alpha(\{\mathbf{x}\})|^2}, \tag{1.52}$$

where $\mathcal{M}(\{\mathbf{x}\})$ is the ME evaluated on the same set of initial partons 4-momenta $\{\mathbf{x}\}$.

The main drawback of this method is the degradation of the statistical precision of observables extracted from these reweighted events, as the integration algorithm no longer has an efficient phase space parameterisation. Events produced in larger numbers in the enhanced regions of the process α to reduce statistical uncertainties

can reside in a suppressed region of the process β. While this is not a modelling problem, since Eq. (1.52) will reduce their weights accordingly, it is not optimal any more. This is especially true in the opposite situation where enhanced regions of the process β can be under-populated and the statistical precision suffer from this. It can become a major issue when the two processes coverage shrinks, as many events will become unphysical and obtain a zero weight, while other regions of the non-zero phase-space are not populated at all. Other issues can arise after the hard scattering in the later parts of the event generation that may depend on the spin and colour states of the produced partons (implicitly summed in the ME of Eq. (1.52)), potentially not coinciding with what is supposedly produced in the β sample. Note that a correction factor $\sigma_\beta^{tot}/\sigma_\alpha^{tot}$ also needs to be included on the overall sample.

1.4 Beyond the standard model

The standard model (SM) is a tightly contained theory, it depends on a certain set of free parameters that can only be determined experimentally but then allows precise predictions of a wide variety of processes. The success of this theory lies in the astonishingly good agreement between its predictions and the measurements performed at collider experiments such as the Compact Muon Solenoid (CMS) collaboration, whose cross-section summary is on Fig. 1.4. The theory is renormalisable and therefore supposed to be consistent up to the Planck scale ($M_{\text{Planck}} \simeq 1.2 \times 10^{19}$ GeV) where gravity is not negligible any more and a unified quantum gravity theory is required. There are nonetheless several issues in the current model that indicate a potential smaller energy scale Λ where new physics might appear. A non-exhaustive list is presented here-under.

- **Hierarchy problem:** The renormalisation of the bare mass of the Higgs boson from loop corrections introduces quadratic divergences in the cut-off scale Λ for the free parameter of the Higgs boson mass. Since nothing prevents this scale to be large, even up to the Planck scale, it is surprising that this mass is so small in comparison as it would require a very fine-tuning of the bare parameters. A more likely explanation would be additional symmetries or degrees of freedom that could protect the Higgs mass appearing above the electroweak scale $\Lambda \gtrsim v$, though they have yet to be unveiled.
- **Dark Matter:** Astronomical evidence and cosmological considerations show that about 84% of the mass content in the Universe is unaccounted for when a comparison is made between the expected content in particles and the gravitational observations [52]. The missing content could be explained by an invisible massive particle that would be electrically neutral to avoid direct detection, and likely weakly-interacting.

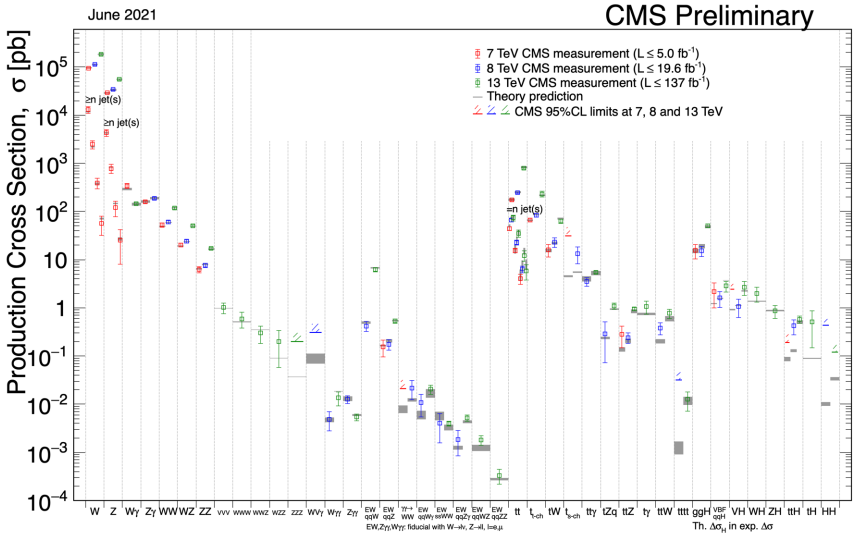


Figure 1.4 | Summary of the cross-section measurements performed by the CMS collaboration. For each process and potential channel, the theoretical prediction is shown in grey with uncertainties, while the different measurements performed at different energy and luminosity levels are shown in red, blue, and green. In case of a measurement the observed value and uncertainties from statistical and systematic sources are shown using a box plot, otherwise an upper limit is shown at 95 % CL. From Ref. [51].

- **Neutrino mass:** As mentioned in Section 1.2.3 the neutrinos are assumed massless in the SM. However, neutrino observatories have showed evidence of neutrino oscillations, which can be explained by a mixing between mass and flavour eigenstates. The introduction of a right-handed neutrino can introduce a mass term from a Yukawa interaction, but the very small couplings compared to other fermions could be the result of new physics. Possibly with new degrees of freedom appearing at high energies, unfortunately unreachable by current colliders, for example a see-saw mechanism that could produce very high mass neutrino counterparts.
- **Matter-antimatter asymmetry:** Astronomical observations have so far failed at detecting large traces of antimatter in the Universe. This asymmetry in the production of matter compared to antimatter would require CP, baryon and lepton number violation, and while CP violation terms are present in the SM, their experimental values are much too small compared to what would be required to explain the current asymmetry.
- **Fermion mass hierarchy:** While the Yukawa couplings are free parameters dictating the masses of the fermions and nothing governs their values in the current model, several patterns seem to exist. There is a mass hierarchy between the different generations (leptons, up- and down-type quarks), and

also within a generation, the down-type being heavier than their up-type counterparts. A similar hierarchy can also be observed in the quark mixing angles in the CKM matrix. These hierarchies could be accidental or could come from a more general mechanism with fewer free parameters, potentially a spontaneous breaking of a larger group.

- **Strong CP problem:** No evidence has been found of CP violation in the strong sector so far, yet the QCD part of the SM Lagrangian contains a violating CP term θ. Therefore, this parameter should have a value so small it would require an unlikely fine-tuning between this parameter in the Lagrangian and the diagonalisation of the Yukawa quark matrices.

There exists a multitude of theoretical models that solves one or several of the aforementioned issues. These models often predict Beyond the standard model (BSM) dynamic, and the task of the experimentalists is to track down every possible signature proposed by these models in the hope of discovering New Physics. This model-dependent approach, performed in a wide variety of sectors by several collider experiments, has to date provided a number of not very significant deviations, some of which due to statistical fluctuations as the accumulation of larger datasets showed, some of which waiting for more data to be further looked into. The opposite approach of inferring these signs in higher energy scales through indirect effects they can have on SM observables is an alternative that is becoming more and more attractive. And while in this more model-independent approach it would be harder to interpret any deviation, at the very least it could provide a general direction in which to direct further experimental efforts.

1.5 Double Higgs production mechanism in the Standard Model

The Higgs boson discovered in 2012 by the ATLAS and CMS collaborations [53, 54] allowed for the first measurement of the Higgs boson mass, and together with the value of the VEV the prediction of the λ parameter in Eq. (1.43). This in turn allowed the prediction of the trilinear and quartic couplings and determine the cross-sections of the LO associated processes, respectively the double and triple Higgs boson production. While the latter is likely to be out of reach of the current and near-future detectors due to tiny cross-section [55–58], the characterization of double Higgs (HH) production is under study and could lead to insightful discoveries about the upper scale of validity of the SM in an effective theory, or provide a glimpse of a potentially richer Higgs sector. The value of the trilinear Higgs coupling is of particular interest as it represents a test of the SM through the measurement of λ. This can potentially lead to a deeper understanding of the Higgs potential and how the Higgs boson acquires the VEV and gives mass to all elementary particles.

1.5.1 Single Higgs production and decay mechanisms

This section is mostly inspired from the Handbook of LHC Higgs Cross-Sections [59–61] and intends to illustrate the fundamentals of the production and decay modes of the Higgs boson that are prerequisite for the understanding of the double Higgs production mechanisms.

The dominant production modes for the Higgs boson are shown on Fig. 1.5 while their value centred around the Higgs mass is displayed on Fig. 1.6a. The main production mode is via gluon-gluon fusion (GGF) through a heavy quark loop, its dominance over the other modes is why it is the most studied one over a wide range of energies. As the dynamics of this mode are driven by strong interactions, the computation of its cross-section at NNLO is currently the state of the art, as it is subject to radiative QCD corrections. Since the loop is dominated by heavy quarks some considerations must be taken into account regarding the uncertainties on the top mass. The vector-boson fusion (VBF) is the second most important mode and provides access to bounds on couplings with the EW gauge bosons. The presence of two additional jets in the backward and forward regions of the detector can serve as an additional feature for background rejection, provided they are within the detector acceptance, although separating them from PU jets is not trivial. The WH and ZH production modes, dubbed Higgs-strahlung, are even fainter but can be observed in particular channels where the backgrounds contamination can be kept reasonable. They are however very sensitive to a low-mass Higgs boson. Higgs radiating from a pair of top quarks is one of the least prominent decay channel currently studied, and its computation is plagued by large uncertainties due to the large dependency of the renormalisation scale on the strong coupling constant and on the factorisation scale of the parton density functions inside the proton. While its relatively low cross-section represents a challenge for its study, it is nonetheless helped by the fact that the background contamination can be very well mitigated.

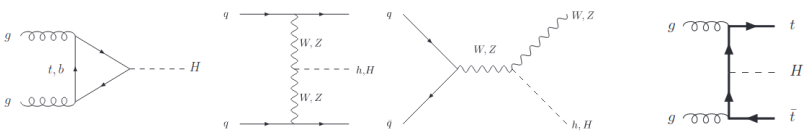
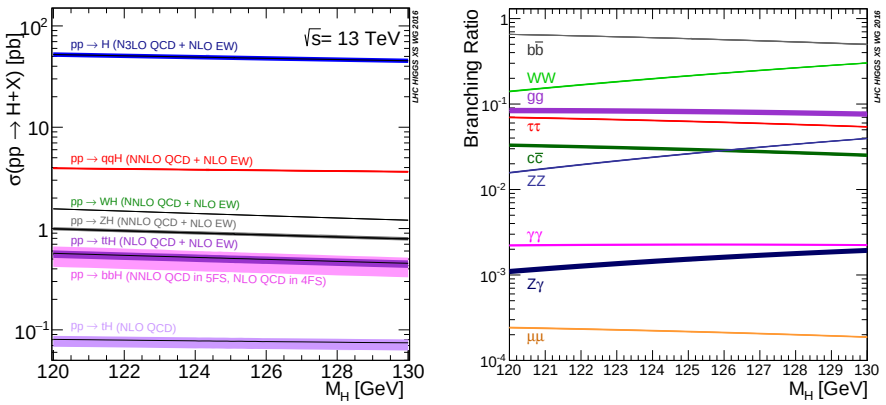


Figure 1.5 | Main Higgs boson production Feynman diagrams at LO. From left to right (by order or importance): gluon-gluon fusion, vector-boson fusion, Higgs-strahlung and top quark pair associated production. As the coupling between fermions and the Higgs boson are proportional to the Yukawa coupling of the fermion and therefore its mass, the diagrams that include an interaction between quarks and the Higgs boson are mostly dominated by top quark contributions, sometimes also including the smaller bottom quark contribution.



(a) *(b)*

Figure 1.6 | Predictions from the SM of the Higgs production mechanism around the measured value of $125\,\mathrm{GeV}$ from Refs. [59–61]. Left : main production mechanisms at different orders and centre of masse energy of $13\,\mathrm{TeV}$. Right : BRs of the Higgs decay.

The branching ratio (BR) values extracted from the ratio of partial and inclusive decay rate, computed from Eq. (1.18), are displayed on Fig. 1.6b, centred around the measured value of the Higgs mass (although such computation has also been derived for a wider range of masses). The coupling to fermions is solely determined by their mass, as the Yukawa coupling enters linearly in the Lagrangian of Eq. (1.49). While the top quark pair decay of the Higgs boson is kinematically impossible, the decay to a pair of bottom quark pair is favoured due to the large mass of the bottom quark comparatively to the other fermions. On the other side of the mass range, the decay into a pair of electrons is extremely disfavoured, and likely unreachable in current experiments. The coupling to the weak bosons depend on the Electroweak (EW) coupling constants and Weinberg angle in Eq. (1.43) that favours a decay into W bosons rather than Z bosons. Energy conservation laws also force one of the weak bosons to be produced off-shell. The Higgs boson does not couple directly with the gluons nor the photons, however the decay is allowed through a fermion loop of mostly top and bottom quarks (and weak boson loop for the photon case).

1.5.2 Double Higgs production and coupling modifiers

This section focuses on the phenomenology of the double Higgs production mechanisms with an emphasis on collider experiments, it is based on Refs. [62, 63]. The main focus of current non-resonant HH studies is the measurement of the trilinear Higgs coupling λ and its comparison with its expected value from the knowledge

of the Higgs mass and the VEV. Contrary to a possibly naive interpretation, the cross-section does not necessarily increase with the value of the coupling, due to interference effects. Around the SM it scales as $\Delta\sigma/\sigma \sim -\Delta\lambda/\lambda$, the full dependence is displayed on Fig. 1.7 and expressed as a function of the *coupling modifier* $\kappa_\lambda = \lambda/\lambda_{\text{SM}}$. This illustrates how much theoretical cross-section uncertainties need to be kept under control to allow for a precise measurement, as well as the non-trivial dependence of both the cross-section and its uncertainties on the coupling.

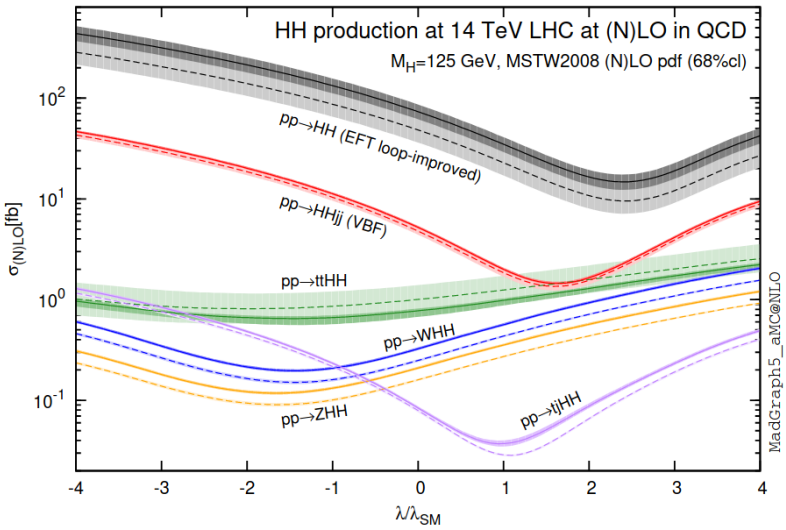
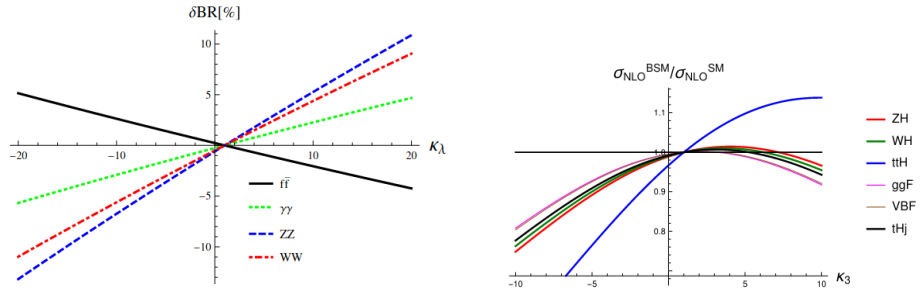


Figure 1.7 | Evolution of the HH cross-section for the different production modes as a function of the trilinear coupling modifier $\kappa_\lambda = \lambda/\lambda_{SM}$, the dashed (solid) lines and light-(dark-)colour bands correspond to the LO (NLO) computations and to the scale and PDF uncertainties added linearly. From Ref [64].

An additional complexity arises with the observation that both the Higgs decay on Fig. 1.8a and single Higgs production cross-section on Fig. 1.8b also show a dependence on the trilinear coupling. This comes from NLO contributions including a κ_λ dependency, as illustrated on Fig. 1.9. The dependency from the decay and the production mode depends on the process and kinematics, and scales linearly with the κ_λ coupling, while a universal contribution from a pure Higgs boson loop includes a quadratic scaling with κ_λ. These effects must be taken into account when performing a measurement on the HH production cross-section, especially since two Higgs boson decays occur in potentially different channels with a non-trivial interplay of κ_λ, and the fact that single Higgs processes will contribute as backgrounds with distributions that are κ_λ-dependent.



(a) *(b)*

Figure 1.8 | Evolution of the BR modification of the Higgs decay (left) and the single Higgs production cross-section (right) as a function of the coupling modifier κ_λ, from Refs. [65, 66].

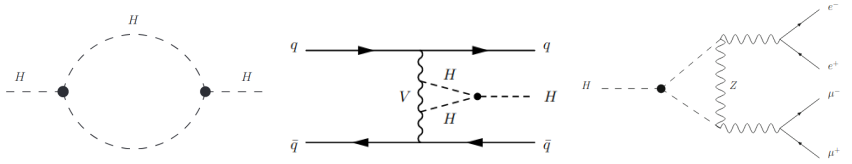


Figure 1.9 | Diagram contributions containing a κ_λ dependency and contributing to the single Higgs production : universal $\mathcal{O}(\kappa_\lambda^2)$ loop dependency (left), process and kinematic dependent $\mathcal{O}(\kappa_\lambda)$ production dependency (middle), and decay dependency (right). From Ref. [65]

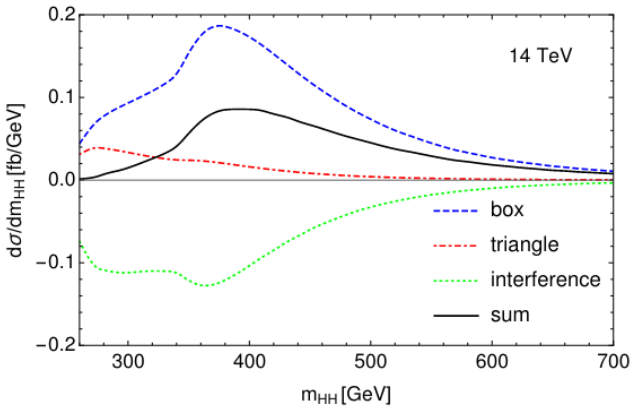


Figure 1.10 | LO differential cross-section of the HH pair as a function of its invariant mass for the different contributions of the gluon-gluon fusion production mode, their interference, and sum which is the only observable. From Ref. [62].

The production of a pair of Higgs boson is a $2 \to 2$ process, thanks to the 4-momentum conservation in Eq. (1.15) and the scalar nature of the final state that allows the removal of an arbitrary global azimuthal phase, the system can be characterized by three variable, at LO and before parton shower. One relates to the boost in the longitudinal direction, already determined in the PDFs, and therefore the HH system depends on two variables, namely the invariant mass of the Higgs pair system m_{HH} and the polar angle of one of the bosons with respect to the beam line in the HH centre of mass frame (also called Collins-Soper frame [67]) encoded in the variable $\cos\theta^*_{\mathsf{HH}}$. It can be shown that in the heavy top quark mass limit $m_{\mathsf{t}} \to \infty$ [68] that the distribution of $\cos\theta^*_{\mathsf{HH}}$ is almost flat, therefore the only variable of interest regarding differential cross-section is the invariant mass m_{HH}. Additionally, due to its scalar nature, its relatively small decay width means that the Narrow width approximation (NWA) can be used to yield an accurate prediction, provided the decay particles are kinematically allowed for an on-shell Higgs boson.

Similarly to the single Higgs, the main mode of HH production is though GGF via a heavy quark loop (Fig. 1.11). The two main diagrams – the so-called box and triangle diagrams, only the latter depending on the trilinear coupling – interact destructively. This interaction explains the non-trivial dependency on κ_λ in Fig. 1.7, also illustrated for the invariant mass spectrum on Fig. 1.10, and reduces the cross-section by about 50% compared to the box diagram only case. On the other hand, any deviation of the value of the trilinear coupling compared to its expected SM value will result in major deviations of the total cross-section, as it only affects the triangle diagram. Interestingly enough, at LO, before parton shower and in the heavy top quark mass limit, the two contributions completely cancel each other in the total amplitude when $m_{\mathsf{HH}} = 2m_{\mathsf{H}}$, although higher order corrections will mitigate this effect. The second most important production mode is VBF, where both Higgs emerge from the t-channel exchange of virtual weak bosons or via a single Higgs (off-shell) production in the similar mode that then splits into a pair. The remaining modes, including double Higgs-strahlung and associated top quark pair production, are also very similar to their single Higgs counterparts but are not expected to be detectable until the next generation of experiments. Generally, any single Higgs production diagram can lead to double Higgs production at the cost of adding a trilinear coupling interaction to the diagram and therefore a large suppression of amplitude, as visible on Fig. 1.7. The HH production cross-sections with QCD corrections, assuming $m_{\mathsf{H}} = 125$ GeV and $\sqrt{s} = 13$ TeV, are given by

$$
\begin{aligned}
\sigma_{\mathrm{GGF}} &= 31.05\,\mathrm{fb} \; {}^{+2.2\%}_{-5.0\%} \; \pm 3.0\ \% \\
\sigma_{\mathrm{VBF}} &= 1.73\,\mathrm{fb} \; {}^{+0.03\%}_{-0.04\%} \; \pm 2.1\ \% \\
\sigma_{\mathrm{ZHH}} &= 0.363\,\mathrm{fb} \; {}^{+3.4\%}_{-2.7\%} \; \pm 1.9\ \% \\
\sigma_{\mathrm{W^+HH}} &= 0.329\,\mathrm{fb} \; {}^{+0.32\%}_{-0.41\%} \; \pm 2.2\ \% \\
\sigma_{\mathrm{W^-HH}} &= 0.173\,\mathrm{fb} \; {}^{+1.2\%}_{-1.3\%} \; \pm 2.8\ \% \\
\sigma_{\mathrm{t\bar{t}HH}} &= 0.775\,\mathrm{fb} \; {}^{+1.5\%}_{-4.3\%} \; \pm 3.2\ \%
\end{aligned}
\tag{1.53}
$$

from Ref. [62]. The first set of uncertainties represents the scale uncertainties in subscripts where the nominal renormalisation and factorisation scales have been varied up and down by a factor of two, the second uncertainties represent the strong coupling α_s and PDF uncertainties. These values have been computed at different order of QCD and EW corrections, at NNLO for GGF (with an approximation on the top mass) and double Higgs-strahlung, N^{3}LO for VBF and NLO otherwise.

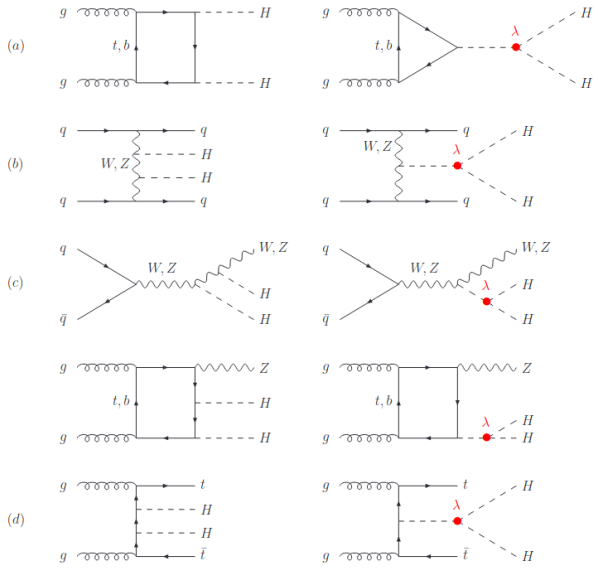


Figure 1.11 | Main diagrams contributing to the double Higgs production at LO : (a) gluon-gluon fusion, (b) vector-boson fusion, (c) double Higgs-strahlung and (d) top quark pair associated production. The trilinear coupling $\lambda \equiv \lambda_{HHH}$ is marked in red. From Ref. [62].

The concept of coupling modifiers can be extended to all interactions vertices in the context of the κ-framework [61, 69] . In addition to κ_λ, one can for example define $\kappa_t = Y_t/Y_t^{SM}$ that modifies the top interaction with the Higgs field and therefore its mass. In general, they modify the differential cross-section behaviour,

as visible for κ_λ on Fig. 1.12. At $\kappa_\lambda = 2.4$ the destructive interference between the triangle and box diagrams is maximal, hence the dip in terms of cross-sections. On the other hand, in the regions where $\kappa_\lambda < -1$ and $\kappa_\lambda > 5$ the triangle-type dominates and the low-m_{HH} region – where this diagram contribution peaks on Fig. 1.10 – is enhanced.

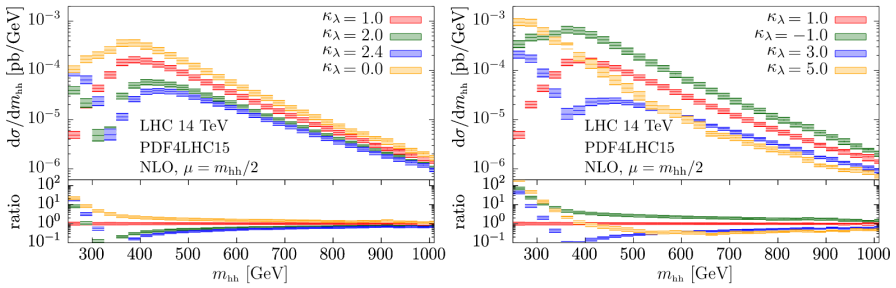


Figure 1.12 | Effect of the value of $\kappa_\lambda = \lambda/\lambda_{SM}$ on the invariant mass plot of the Higgs boson pair for small positive values (left) and larger or negative values (right), from Ref. [62].

The κ-framework is interesting from the experimental point of view, as any deviation from $\kappa = 1$ can be interpreted as a deviation of the SM. While derived from pure LO considerations, the dependency on the coupling modifiers remains at NLO. This framework is however limited to a detection of deviations, as it is not related to any field theory embedding these deviations cannot be interpreted in any BSM scenario. It only represents an "ad-hoc" construction that can only work around the SM without a way to generalise it, and is in addition scale dependent and non-gauge invariant. A proper generalisation of the SM rather lies in the concept of EFTs, in which more motivated versions of κ_λ and κ_t will be defined.

1.6 Resonant enhancement of double Higgs production

Several BSM scenarios of heavy states decaying to pairs of SM Higgs bosons can lead to an enhancement of the double Higgs sensitivity and allow an early discovery. This section will briefly describe a few of them, especially the ones relevant to this work, and is based on more complete reviews in Refs. [62, 63].

The simplest extension of the SM consists in the addition of a new gauge singlet scalar S [70–74] to the already existing scalar doublet ϕ, that can then decay into a pair of Higgs bosons [75–94]. This way, the only interactions between SM and BSM fields is through the Higgs sector. This model is called the *Higgs singlet/portal* and its potential is parameterised as

$$V(\psi, S) = -\mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2 + \frac{a_1}{2}\phi^\dagger\phi\ S + \frac{a_2}{2}\phi^\dagger\phi\ S^2 + b_1 S + \frac{b_2}{2}S^2 + \frac{b_3}{3}S^3 + \frac{b_4}{4}S^4. \tag{1.54}$$

One can write these two scalar fields as

$$\phi = \frac{1}{\sqrt{2}}(0\ ,\ v+h)^T \quad , \quad S = \frac{1}{\sqrt{2}}(v_s + s), \tag{1.55}$$

where the v and v_s are the VEV of the two fields respectively, and while h is the usual SM Higgs boson, the new field introduces a new scalar boson s. After EWSB the two fields mix and produce two mass eigenstates h_1, h_2 with associated masses m_1, m_2 such as

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} h \\ s \end{pmatrix}, \tag{1.56}$$

where θ is the coupling angle. The first eigenstate can be associated to the SM Higgs boson such that $m_1 = 125$ GeV $< m_2$. With this parameterisation, the value of the mixing angle controls the couplings of each boson with the fermion and gauges bosons, and their production cross-sections. As no experimental evidence shows a large deviation of the light Higgs boson h_1 with SM particles, this angle should be small ($\theta \simeq 0$) such that the mixing matrix is almost diagonal. Note that if $m_2 > 2m_1$, the decay $h_2 \to h_1 h_1$ is on-shell and could contribute to the double Higgs production as a cross-section enhancement, but even if this condition is not met the trilinear Higgs coupling could be impacted by this singlet scalar field and parameterised in a EFT framework.

In case there is no symmetry associated to S, the VEV is non-physical, one can set $v_s = 0$ and there are five physical parameters : $v = 246$ GeV, $v_s = 0$, $m_1 = 125$ GeV, m_2 with assumption $m_2 > 2m_1$ and θ. All the other parameters can be determined analytically from that point. On the other hand, if a Z_2 symmetry is assumed, such that $S \to -S$, then $a_1 = b_1 = b_3 = 0$, and if S acquires a VEV ($v_s \neq 0$) this symmetry is softly broken. In that case the only physical parameters are m_1, v, m_2, $\sin\theta$ and $\tanh\beta \equiv v/v_s$, out of which only the first two have experimental values and the rest are free parameters.

Another model aiming at an extension of the SM consist in considering from the start two scalar field doublets under $SU(2)_L$ ϕ_1 and ϕ_2. This model is named Two Higgs Doublet Model (2HDM) and the most general way to parameterise it is [95, 96]

$$V(\phi_1, \phi_2) = m_{11}^2\phi_1^\dagger\phi_1 + m_{22}^2\phi_2^\dagger\phi_2 - \left(m_{12}^2\phi_1^\dagger\phi_2 + \text{h.c.}\right) + \frac{\lambda_1}{2}(\phi_1^\dagger\phi_1)^2 + \frac{\lambda_2}{2}(\phi_2^\dagger\phi_2)^2 + \lambda_3(\phi_1^\dagger\phi_1)(\phi_2^\dagger\phi_2) + \lambda_4(\phi_1^\dagger\phi_2)(\phi_2^\dagger\phi_1)$$

$$+\left[\frac{\lambda_5}{2}(\phi_1^\dagger\phi_2)^2+\lambda_6(\phi_1^\dagger\phi_1)(\phi_1^\dagger\phi_2)+\lambda_7(\phi_1^\dagger\phi_2)(\phi_2^\dagger\phi_2)+\text{h.c.}\right]. \quad (1.57)$$

Under the hypothesis that there are no flavour changing neutral currents and that the $\mathcal{Z}_2$ symmetry is softly broken, each fermion couples to a single doublet. This defines the different schemes of the 2HDM which can be *type-I*, *type-II*, *lepton-specific* or *flipped*. While the original scalar sector of the SM only had two free parameters m_h and μ, this new model has 14, out of which some are real due to the hermicity of the Lagrangian – m_{11}^2, m_{22}^2, λ_1, λ_2, λ_3 and λ_4 — while others are complex – m_{12}^2, λ_5, λ_6 and λ_7. This set of free parameters can be reduced by imposing CP-conservation in the scalar sector such that all parameters are real, and Z_2 symmetry ($\phi_1 \to -\phi_1$, $\phi_2 \to \phi_2$) which cancels $\phi_1 \to \phi_2$ transitions which would otherwise cause Flavour Changing Neutral Current (FCNC). This has the consequence that $\lambda_6 = \lambda_7 = 0$.

With two doublets of $SU(2)_L$ there are eight fields, and one can parameterise the doublets as

$$\phi_i = \begin{pmatrix} \phi^+ \\ (v_i + \rho_i + i\eta_i)/\sqrt{2} \end{pmatrix} \quad , \quad i = 1, 2, \quad (1.58)$$

where ϕ_i^+ is complex while ρ_i and η_i are real fields.

Out of these eight degrees of freedom, three are needed to give the mass to the weak gauge bosons. The remaining five degrees of freedom correspond to the scalar Higgs bosons after EWSB : two charged bosons $\mathsf{H}^+, \mathsf{H}^-$, one pseudoscalar A and two scalar h and H. The couplings between the two scalars are given by $\tanh\beta \equiv v_2/v_1$ and α which are rotation angles of the mixing matrices for the charged and pseudoscalar, and the neutral scalars, respectively. In the alignment limit where $\beta - \alpha = \pi/2$ the light scalar boson is identified as the SM Higgs boson and the decay of the heavy boson $\mathsf{H} \to \mathsf{hh}$ can be significant in the double Higgs production [97, 98, 98], although other decay modes such as $\mathsf{H} \to \mathsf{ZA}$ or $\mathsf{H} \to \mathsf{AA}$ can dominate.

Other variations of the model do not necessarily use all the above assumptions, for example the complex or CP-violating 2HDM (C2HDM) [99], or the combination of the 2HDM with the Higgs singlet model [100, 101]. There also exists supersymmetric models such as the MSSM [102, 103] that feature the same doublets initial assumption.

Going further, the Georgi-Machacek (GM) model [104, 105] considers the classic doublet ϕ and adds to the scalar sector a real and complex triplets under $SU(2)_L$. After EWSB this models gives rise to a custodial fiveplet (H_5^{++},H_5^{+},H_5^{0},H_5^{-},H_5^{--}) degenerate in mass m_5, a custodial triplet (H_3^{+},H_3^{0},H_3^{-}) degenerate in mass m_3 and a heavy custodial singlet H with mass m_{H}. Double Higgs production is only concerned about neutral bosons, and the custodial symmetry forbids the decay of H_3^0 and H_5^0 into a Higgs pair, therefore the only resonant enhancement would be from $\mathsf{H} \to \mathsf{hh}$ that could impact trilinear coupling and modify the production

cross-section provided $m_{\mathsf{H}} > 2m_{\mathsf{h}}$.

From the experimental point of view, provided scalar resonances decaying to pairs of SM Higgs bosons have a narrow width, their production and decay mechanism only depend on the mass of the scalar state, and not the rest of the model specificities. Therefore, experimental constraints can be interpreted in a model-independent way, if the BR of such states into the Higgs boson pair is assumed. On the contrary, spin-2 states cross-section enhancements would depend on the production mode. In proton collisions, each model will have a different experimental sensitivity based on its prescription for the coupling with either quarks or gluons.

Examples of a spin-2 resonant state that will be employed in this work are the extradimensional models, where the space-time is assumed to $4+n$-dimensional and massive spin-2 states – the Kaluza-Klein (KK) *gravitons* – are generated. These models postulate that gravity is allowed to propagate through the bulk of the extra dimensions, while the SM fields are only localised to the four usual space-time dimensions called the 3-brane. This could explain the vast difference between the scales of the three fundamental forces inside the SM and the much fainter gravitational force that is yet to be included. While models with several additional dimensions $n \geq 2$ [106] produce light KK states that have not been observed experimentally, a subcategory of Warped extra dimension (WED) with $n = 1$ in the case of the Randall–Sundrum (RS) models [107–112] gives rise to KK states above the TeV scale, therefore displaying signatures that have not yet been fully explored but could be within reach of current collider experiments. WED models typically also generate scalar excitations of the extra dimension of the metric in the form of a lighter spin-0 state called the *radion* [113, 114] that can decay to heavy bosons, including the Higgs boson.

The RS models also have the advantage of providing solutions to other problems of the SM, namely the electroweak and lepton flavour hierarchies. The distinction is made between the IR brane – also called TeV brane – where the usual SM fields live and the UV brane – also called Planck brane – where gravity extends. The metric of such space would be the product of the four-dimensional metric by a *wrap* factor, which is a rapidly changing function of an additional dimension. If this function is parameterised as an exponential, it could explain the large difference between the electroweak and Planck scales, naturally solving the electroweak hierarchy problem. The parameters of the theory are the warp ratio k and the volume of the extra dimension L, it has been determined that the observed electroweak hierarchy can be explained when $kL \simeq 35$. It is also convenient to define $\tilde{k} = k/M_{pl}$, normalised to the Planck Mass. There exists several versions of the model, in the RS1 model for example all the SM fields are restricted on the IR brane, but there also exists a RS2 version where the IR brane is pushed to infinity. The former version is usually favoured, but suffers from the problematic that the light fermions have large couplings with the KK states and would lead to unobserved flavour violation

processes. This also means the coupling to the Higgs boson are suppressed and these models are not very relevant to the double Higgs production mechanism

In so-called *Bulk* models there is no dichotomy and the SM fields are allowed to extend towards the five-dimensional bulk while the Higgs field stays localized on the IR brane. This formulation can the flavour hierarchy problem : heavy fermions are located closer to the IR brane and the Higgs field, while lighter fermions are located farther and their interaction with the Higgs field is suppressed exponentially. In addition, the KK couplings are hierarchical and the most important decay modes are with heavier SM particles, including the decay to a pair of Higgs bosons which would enhance its production rate.

The production cross-section of these spin-2 KK gravitons and to a smaller extent spin-0 radions depends on the model and its parameters. Such priors need to be set to some motivated values to be simulated and to allow the measurement of experimental constraints that cannot therefore be interpreted strictly as model-independent. Nonetheless, in case the experimental sensitivity does not show large variations to the model parameters, and especially if the width is narrow, there is room for other interpretations.

If the masses of the spin-0 or spin-2 states are too large to allow a direct detection, their effect can still be detected in an indirect search. Indirect effects of heavy resonances can be detected through the prism of an Effective Field Theory (EFT) approach [115]. And while the condition for a resonant decay to a pair of Higgs bosons is strictly $m_{\mathrm{X}} > 2m_{\mathrm{h}}$, a resonance that would not fulfil this condition could still impact the production rate via loop circulation or other means [116–118]. All these additional potential contributions, along with the ones that are not listed here, would however need to be evaluated on an individual basis which would be in opposition to the model-independent aim of the experimental enterprise. They will briefly be discussed in a following section.

1.7 Effective field theory formalism

The idea behind an Effective Field Theory (EFT) is to approximate arbitrarily complex interactions into a set of parameters in a systematic and consistent way. The precise calculations of these parameters can be translated into observables that provide experimental sensitivity. This framework can be used to extend the SM in the most general way. The addition of a new term in the Lagrangian requires modifications outside the initial sphere of influence of this term to respect gauge symmetries. In the particular example of the double Higgs production, an additional term of the form $\Delta\mathcal{L} = -ch^3$ to modify the triple Higgs boson vertex would result in additional multi-Higgs vertices and higher EW corrections that would impact the other terms in the SM Lagrangian. All these interplays can be fully

described by the parametric description of an EFT to arbitrary precision. Two EFT descriptions aiming at extending the SM are currently used in HEP : the Standard Model Effective Field Theory (SMEFT) and Higgs Effective Field Theory (HEFT) – also sometimes referred to as Electroweak Chiral Lagrangian (EWChL). They both lead to a self-consistent, gauge-invariant Lagrangian, the only differences are their definition of the Higgs field as a $SU(2)_L \times U(1)_Y$ doublet and singlet respectively, as well as the ordering of the set of operators added into the Lagrangian. HEFT is more general, as the scalar particle is not assumed to be the SM Higgs. It actually predates and comprises the SMEFT.

The principle behind SMEFT lies in the addition of fields that act on a short range or at high energy scales to the SM Lagrangian, such that the new physics is coupled to the SM with a cut-off scale Λ, well separated from the electroweak scale $\Lambda \gg m_{\mathsf{Z}}$, and the indirect effects on low-energy observables can be parameterised in a model-independent way. In order for this effective Lagrangian to be renormalisable, all terms are operators that must be of dimension four or less, so these additional operators of higher order d must have coefficients proportional to Λ^{4-d}. This parameterisation is therefore renormalisable for each order. The most generic way to integrate such operators into a EFT Lagrangian [119–121] is

$$\mathcal{L}_{\mathsf{SMEFT}} = \mathcal{L}_{SM} + \sum_{d>4}^{\infty} \sum_{i} \frac{c_i^{(d)}}{\Lambda^{d-4}} \mathcal{O}_i^{(d)}, \tag{1.59}$$

where $\mathcal{O}_i^{(d)}$ are the effective operators with dimension $[E]^d$ that respect gauge symmetries, $c_i^{(d)}$ are called the Wilson coefficients and are the parameters that can be translated to actual observables. In the weakly coupled assumption, the coefficients $c_i^{(d)}$ are naturally small and the theory can be treated with small deviations on the couplings, although not all BSM deviations are allowed.

More general, and therefore more complex, HEFT considers new physics with heavy particles of mass M coupled to the scalar sector. The couplings are of order $\mathcal{O}(1)$ and loop expansion is used as power counting [122] in the Lagrangian

$$\mathcal{L}_{\mathsf{HEFT}} = \mathcal{L}_{LO} + \sum_{L=1}^{\infty} \left(\frac{1}{16\pi^2} \right)^L c_i^{(L)} \mathcal{O}_i^{(L)}, \tag{1.60}$$

with similar definition of the Wilson coefficients and operators as Eq. (1.59). A naive dimensional analysis argument [123] states that $\Lambda = 4\pi v \simeq M$ such that $v^2/M^2 \simeq 1/16\pi^2$. Contrary to SMEFT, some couplings can only be constrained by HH measurements and the combinations with single Higgs measurement is not straightforward. Since the discovery of the SM Higgs boson and the relatively good agreement with its expectation, HEFT has declined in popularity over the last decade.

Both these EFT frameworks are discussed in the following sections.

1.7.1 Linear EFT

The effective Lagrangian of Eq. (1.59) contains an infinity of orders increasingly suppressed orders of $E_{exp}/\Lambda \ll 1$, however for each order the number of contributing operators is finite, allowing to truncate the sum at a certain order to keep only the dominant operators. The SM is recovered when either all the Wilson coefficients are null ($c_i^{(d)} = 0$ for all i and d) or when $\Lambda \to \infty$. The parameterisation of Eq. (1.59) is bound to several assumptions. The first one is that its validity is only guaranteed if $E \ll \Lambda$, otherwise the additional operators contribute equally and the truncation of the higher-order terms breaks down. Outside that range a *UV-complete* theory needs to be developed. In addition, this hypothetical UV theory needs to have couplings that can be developed in a perturbative expansion into the Wilson coefficients. Dimensional arguments lead to the constraint $|c_i^{(d)}| < (4\pi)^{n_i-2}$ for n_i fields [124]. Although by themselves the operators are not renormalisable, the introduction of the cut-off Λ allows the Lagrangian to be renormalisable order by order, allowing for higher-order correction calculations [125–127].

Obtaining an observable from this formalism requires computation of the amplitude and therefore the ME from Eq. (1.14). Both the SM and EFT parts of the effective Lagrangian, truncated at dimension-6 operators, can be factorized in the ME

$$|\mathcal{M}|^2 = |\mathcal{M}_{SM}|^2 + 2\sum_i \frac{c_i}{\Lambda^2}\mathrm{Re}(\mathcal{M}_{SM}^*\mathcal{M}_i) + \sum_{i,j}\frac{c_i c_j}{\Lambda^4}\mathrm{Re}(\mathcal{M}_i^*\mathcal{M}_j), \qquad (1.61)$$

where $\mathcal{M}_{SM}$ is the ME relative to the SM part of the Lagrangian, while $\mathcal{M}_i$ is the one relative to one EFT operator.

The first term in Eq. (1.61) refers to the actual SM predictions, the third to the pure EFT contributions – either operator squared or a mixed product of them – while the second is the interference term between the two. It would be tempting to omit the pure Λ^{-4} part of Eq. (1.61), similarly to the perturbation truncation. This would however be ill-advised, as this contribution could be non-negligible in the cases where either the Wilson coefficients are large, the interference term would be suppressed [128], or the SM amplitude vanishes (for example flavour-violating processes).

The objective of the experimental approach to EFT is to put constraints on the Wilson coefficients or in the best case scenario prove their non-zero value, which will therefore indicate a sign of new physics potentially at higher energy than what is achievable with current accelerators. The constraints can be obtained from deviations of SM observables and can then be translated to bounds on the scale at which this new physics could appear, a procedure known as a *bottom-up* approach. On the other hand, the *top-down* approach would be to start from a BSM scenario, work out the associated Wilson coefficients and compare them with the current experimental constraints or deviations to interpret them back into the parameters

of the new model.

The operators of dimension 5,7,etc involve lepton number violation which has not been experimentally observed so far, in particular order 5 operators give rise to a Majorana mass term for neutrinos [129] which could solve one of the current problems of the SM (cf. Section 1.4). Lack of experimental evidence supporting these predictions suggests their abandon henceforth. Despite each order having a finite number of operators, this number is very large. There are however many linear combinations of operators and a basis of independent operators based on arbitrary choices can be determined. There exists two bases in the literature, the Warsaw [130] and the Strong Interaction Light Higgs (SILH) [131, 132] bases, that only differ in the operator definitions but should yield exact same physics conclusion. Further simplifications can also be applied, for example assuming strongest form of flavour universality – that the electroweak coupling of the gauge bosons to leptons is independent of the lepton flavour – and baryon number conservation [130, 133]. The SILH basis was used in this section for its shorter coefficient comparison with HEFT.

The number of parameters remains problematic to perform global fits, therefore most studies restrict themselves to modifications of small sets of operators that are relevant for the case at hand. In the context of the double Higgs production, the subset of operators that give LO contributions to the gluon-gluon fusion production mechanism lies in this additional part of the Lagrangian :

$$\Delta\mathcal{L}_6 = \frac{c_H}{2\Lambda^2}\mathcal{O}_H - \frac{c_6}{\Lambda^2}\mathcal{O}_6 - \frac{c_{t\phi}}{\Lambda^2}Y_t\mathcal{O}_{t\phi} + \frac{c_{tG}}{\Lambda^2}Y_t g_s\mathcal{O}_{tG} + \frac{c_{\phi G}}{\Lambda^2}\frac{\alpha_s}{4\pi}\mathcal{O}_{\phi G} \tag{1.62}$$

$$\begin{aligned}
\mathcal{O}_H &\qquad \partial_\mu(\phi^\dagger\phi)\partial^\mu(\phi^\dagger\phi) \\
\mathcal{O}_6 &\qquad (\phi^\dagger\phi)^3 \\
\mathcal{O}_{t\phi} &\qquad (\phi^\dagger\phi)\ \bar{Q}_L\ \phi^c\ t_R + \text{h.c.} \\
\mathcal{O}_{tG} &\qquad \bar{Q}_L\sigma^{\mu\nu}G_{\mu\nu}\phi^c\ t_R + \text{h.c.} \\
\mathcal{O}_{\phi G} &\qquad (\phi^\dagger\phi)\ G^a_{\mu\nu}G^{a,\mu\nu}
\end{aligned} \tag{1.63}$$

where Q_L is the $(t,b)_L$ doublet (other flavours are negligible), Y_t the top quark Yukawa coupling, $\phi^c = i\sigma_2\phi$ is the charge conjugate Higgs doublet, and $\sigma^{\mu\nu} = [\gamma^\mu, \gamma^\nu]$.

Similarly to $\mathcal{O}_{t\phi}$, any fermion can be used with a similar expression, although usually only top quarks, bottom quarks and tau leptons are considered given the Yukawa factor that is particularly small for other flavours (cf. Fig. 1.1). However, in the context of HH production, the measurement is challenging enough that only top quarks can be considered. In future more precise measurements these operators would become relevant, especially when including the decay of the Higgs bosons,

as well as new operators connecting the vector bosons, related to the leading decay modes in Fig. 1.6b. Additional operators would also need to be accounted for in higher-order QCD and EW corrections. This goes however beyond the scope of this section and will not be discussed. CP symmetry has been assumed (although it is still an open research topic), such that CP-violating operators can be neglected, and all other Wilson coefficients are forced to have real values.

The operators $\mathcal{O}_H$ and $\mathcal{O}_6$ affect the Higgs self interaction terms directly, though while $\mathcal{O}_6$ only affects the HH production, $\mathcal{O}_H$ will also include a universal rescaling of single Higgs production cross-sections. The operators $\mathcal{O}_{t\phi}$ and $\mathcal{O}_{tG}$ both affect the Higgs coupling to the top quark, the former shifts the Yukawa coupling of the top quark and the latter causes a new contact interaction between the Higgs, a gluon and the top quark, and is usually referred to as a *dipole operator*. Both are induced within the top quark loop in the HH production diagrams. $\mathcal{O}_{\phi G}$ produces a tree level contact interaction between two gluons and two Higgs bosons. The strong dependence between the HH cross-section and these operators is illustrated on Fig. 1.13.

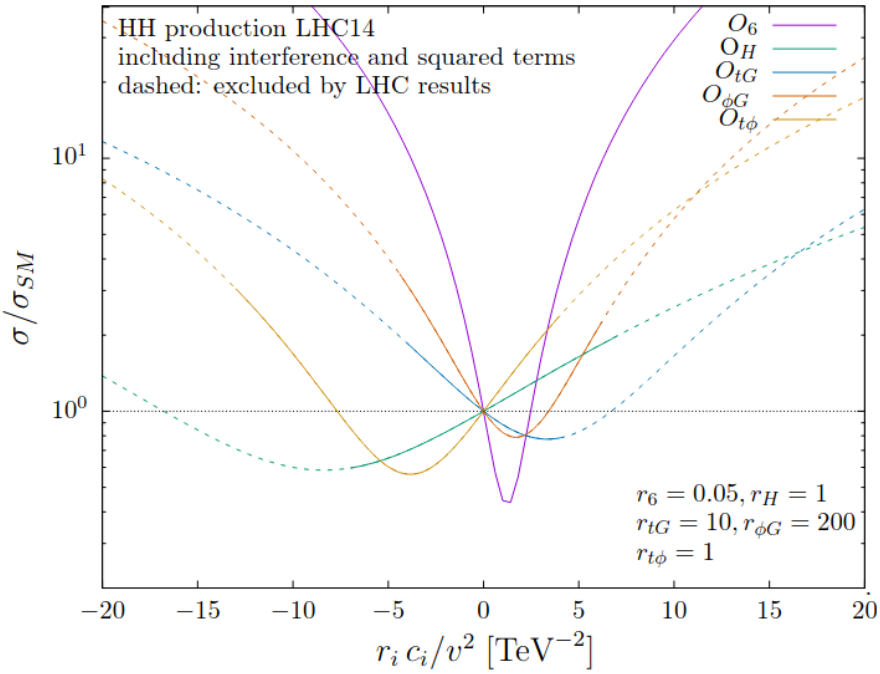


Figure 1.13 | Effects of several Wilson coefficients of dimension-6 relevant to double Higgs production on its cross-section. The dashed lines are the already excluded values from LCH Run 1 Higgs and top quark measurements. A multiplicative factor r_i has been included for ease of comparison, and in practice with $r_6 = 0.05$, the effect of the $\mathcal{O}(6)$ is negligible compared to the others. From Ref. [62].

In many UV models, the dipole operator $\mathcal{O}_{tG}$ appears at loop-level such that applied to the context of HH production (where it appears in a loop) it is loop-suppressed compared to the other operators. It is often excluded from the searches for the sake of simplification, at the cost of introducing a theory bias [134]. In any case, if the UVs theory couples strongly to the heavy sector, then SMEFT would not suitable any more at low energies. The $\mathcal{O}_{\phi G}$ operator, although also typically suppressed by a loop factor, act at tree-level in the HH diagrams and have sizeable effects, on the level of the remaining operators, and is therefore not disregarded. The remaining four operators $\mathcal{O}_H$, $\mathcal{O}_6$, $\mathcal{O}_{t\phi}$ and $\mathcal{O}_{\phi G}$ are investigated in several SMEFT analyses [135, 136].

After redefinitions of the field to remove c_H from the kinetic term [62, 135] and the Wilson coefficients to absorb the suppression factor Λ

$$\begin{aligned} h &\to (1 - \frac{c_H v^2}{2\Lambda^2})h - \frac{c_H v}{2\Lambda^2}h^2 - \frac{c_H}{6\Lambda^2}h^3 \\ c_i &\to c_i \frac{\Lambda^2}{v^2}, \end{aligned} \tag{1.64}$$

the EWSB can be performed to yield the Lagrangian of gluon-gluon fusion at LO :

$$\begin{aligned} \mathcal{L}_{\text{HH}} = &- \frac{m_h^2}{2v}\left(1 - \frac{3}{2}c_H + c_6\right)h^3 - \frac{m_h^2}{8v^2}\left(1 - \frac{25}{3}c_H + 6c_6\right)h^4 \\ &+ \frac{\alpha_s c_{\phi G}}{4\pi}\left(\frac{h}{v} + \frac{h^2}{2v^2}\right)G^a_{\mu\nu}G^{a,\mu\nu} \\ &- \left[\frac{m_t}{v}\left(1 - \frac{c_H}{2} + c_{t\phi}\right)\bar{t}_L t_R h + h.c.\right] \\ &- \left[\frac{m_t}{v^2}\left(\frac{3c_{t\phi}}{2} - \frac{c_H}{2}\right)\bar{t}_L t_R h^2 + h.c.\right]. \end{aligned} \tag{1.65}$$

From there, the SMEFT corrections to the trilinear and quartic couplings are given by

$$\lambda_{H^3}/\lambda_{H^3}^{SM} \equiv \kappa_\lambda = 1 - \frac{3}{2}c_H + c_6, \qquad \lambda_{H^4}/\lambda_{H^4}^{SM} \equiv 1 - \frac{25}{3}c_H + 6c_6. \tag{1.66}$$

The EFT deviations of the trilinear and quartic couplings are different, and a measurement of both could be a powerful probe of new physics in the Higgs sector. The measurement of these deviations from the SM could however prove challenging in a global fit given the complex interaction between the different operators (as visible on Fig. 1.13), especially if their Wilson coefficients are small. This can be alleviated by either considering a global fit of one or a combination of the operators using different LHC measurements, or to investigate the different impacts that the operators have on the invariant mass distributions of the Higgs pair, the most sensitive of the two independent variables characterising double Higgs production.

1.7.2 Anomalous couplings and shape benchmark analysis

Contrary to the linear SMEFT, the non-linear HEFT such as in Eq. (1.60) displays terms of the type $c_n(h/v)^n$. As these terms leave the electroweak gauge sector unchanged, this parameterisation cannot be constrained by electroweak measurements. It is however a natural formulation for strongly coupled new physics, and would be more sensitive to large deviations in the scalar sector at scales above the VEV.

Gluon-gluon fusion

The GGF production mode contribution to the non-linear EFT is given by [63, 135–137] :

$$\begin{aligned}\mathcal{L}_h = &\frac{1}{2}\partial_\mu h \partial^\mu h - \frac{1}{2}m_H^2 h^2 - \kappa_\lambda \frac{m_H^2}{2v} h^3 \\ &- m_t \left(v + \kappa_t \frac{h}{v} + c_2 \frac{h^2}{v^2} \right) \left(\bar{t}_L t_R + h.c. \right) \\ &+ \frac{\alpha_s}{8\pi} \left(c_g \frac{h}{v} + c_{2g} \frac{h^2}{v^2} \right) G^a_{\mu\nu} G^{a,\mu\nu}, \end{aligned} \tag{1.67}$$

where only the top quark as been considered in the fermion term given its dominance, and the weak gauge bosons contributions neglected due to their low contribution to the gluon-gluon fusion process.

Five couplings have been defined in Eq. (1.67), their contributions are highlighted in the diagrams of Fig. 1.14. Similarities can be found between the operators defined in Eq. (1.67) and their SMEFT counterparts in Eq. (1.63), and simple relations can be worked out. The dipole operator $\mathcal{O}_{tG}$ is however missing in this HEFT formulation due to its higher order, as well as the link with single Higgs interactions. The two coefficients previously introduced as *coupling modifiers* κ_λ and κ_t responsible respectively for the Higgs trilinear interaction and the interaction between the Higgs boson and the top quark have a unit value in the SM hypothesis $\kappa_\lambda^{SM} = \kappa_t^{SM} = 1$, while the remaining couplings c_2, c_g and c_{2g} represent interactions that do not exist in the SM at all and in that context are expected to be null.

The ME can be computed from the Lagrangian of Eq. (1.67), and be used to derive a parametric expression of the cross-section :

$$\begin{aligned} R_{\text{HH}} = {}^{\sigma_{LO}}/_{\sigma_{LO}^{SM}} = &A_1\kappa_t^4 + A_2 c_2^2 + (A_3\kappa_t^2 + A_4 c_g^2)\kappa_\lambda^2 + A_5 c_{2g}^2 + (A_6 c_2 + A_7\kappa_t\kappa_\lambda)\kappa_t^2 \\ &+ (A_8\kappa_t\kappa_\lambda + A_9 c_g\kappa_\lambda)c_2 + A_{10}c_2c_{2g} + (A_{11}c_g\kappa_\lambda + A_{12}c_{2g})\kappa_t^2 \\ &+ (A_{13}\kappa_\lambda c_g + A_{14}c_{2g})\kappa_t\kappa_\lambda + A_{15}c_g c_{2g}\kappa_\lambda, \end{aligned} \tag{1.68}$$

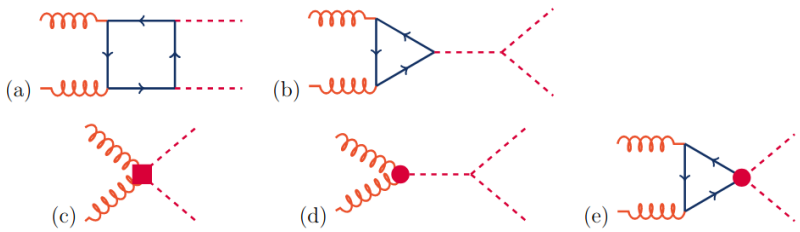


Figure 1.14 | Diagrams of the main contributions for the gluon-gluon fusion production in the context of the non-linear Lagrangian in Eq. (1.67). To the SM processes in (a) and (b) are added pure BSM interactions : (c) and (d) represent contact interactions between one or two Higgs bosons with couplings c_g and, c_{2g} respectively, while (e) represents the contact interaction with top quark pairs of coupling c_2. From Ref. [137].

where the A_i can be determined from a fit using samples generated at different points of this 5-dimensional space, see Ref. [137] for example.

The additional terms corresponding to NLO contributions were derived in Ref. [138]

$$\begin{aligned}{}^{\Delta\sigma}/_{\sigma^{SM}} = {} & A_{16}\kappa_t^3 c_g + A_{17}\kappa_t c_2 c_g + A_{18}\kappa_t c_g^2 \kappa_\lambda + A_{19}\kappa_t c_g c_{2g} \\ & + A_{20}\kappa_t^2 c_g^2 + A_{21} c_2 c_g^2 + A_{22} c_g^3 \kappa_\lambda + A_{23} c_g^2 c_{2g}. \end{aligned} \tag{1.69}$$

Covering this whole space of five parameters would however turn out too costly from an experimental point of view, both from event generation and the statistical estimation of the constraints on each parameter. For this reason, it is very useful to define a finite set of benchmarks that are representative enough of the various regions of this BSM scenario and their associated kinematic description. Two sets of benchmark points were derived using clustering methods. A first approach based on statistical Two-sample tests resulted in 12 points [137], with an additional point in a later publication [139]. A second approach using unsupervised ML methods produced 7 additional points [140]. Both sets of couplings for the benchmarks that were identified are on Table 1.3 and Table 1.4 respectively. The invariant mass distribution of the double Higgs mass from the first method for the different defined benchmarks is on Fig. 1.15, it illustrates how couplings variations can vastly impact this variable. The variability can be explained by the cancellation between the triangle- and box-type diagrams contributing to the GGF at LO. The $\cos\theta^*_{\mathrm{HH}}$ distribution on the other hand is relatively flat, as expected from the dominance of the s-wave diagram [97].

Table 1.3 | Values obtained from the clustering of the 12 benchmarks defined in Ref. [137], with the additional "8a" point from Ref. [139].

Benchmark number	1	2	3	4	5	6	7	8	9	10	11	12	8a
κ_λ	7.5	1.0	1.0	-3.5	1.0	2.4	5.0	15.0	1.0	10.0	2.4	15.0	1.0
κ_t	1.0	1.0	1.0	1.5	1.0	1.0	1.0	1.0	1.0	1.5	1.0	1.0	1.0
c_2	-1.0	0.5	-1.5	-3.0	0.0	0.0	0.0	0.0	1.0	-1.0	0.0	1.0	0.5
c_g	0.0	-0.8	0.0	0.0	0.8	0.2	0.2	-1.0	-0.6	0.0	1.0	0.0	0.8/3
c_{2g}	0.0	0.6	-0.8	0.0	-1.0	-0.2	-0.2	1.0	0.6	0.0	-1.0	0.0	0.0

Table 1.4 | Values obtained from the clustering of the 7 benchmarks defined in Ref. [140].

Benchmark number	1	2	3	4	5	6	7
κ_λ	3.94	6.84	2.21	2.79	3.95	5.68	-0.10
κ_t	0.94	0.61	1.05	0.61	1.17	0.83	0.94
c_2	-1./3.	1./3.	-1./3.	1./3.	-1./3.	1./3.	1.
c_g	0.5x1.5	0.0x1.5	0.5x1.5	-0.5x1.5	1./6.x1.5	-0.5x1.5	1./6.x1.5
c_{2g}	1./3.x(-3.)	-1./3.x(-3.)	0.5 x(-3.)	1./6.x(-3.)	-0.5 x(-3.)	1./3.x(-3.)	-1./6.x(-3.)

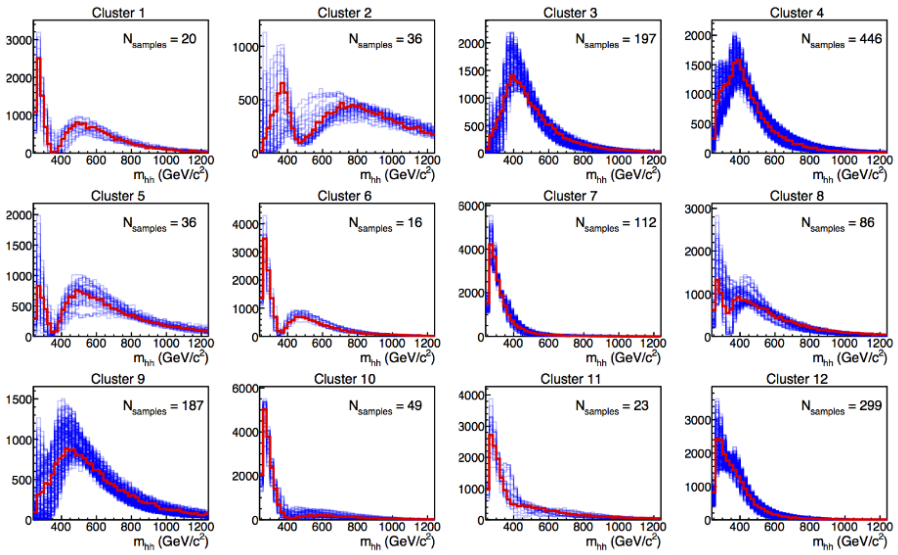


Figure 1.15 | Generator-level distributions of the double Higgs invariant mass m_{HH} for the different benchmarks defined in Ref. [137]. The blue lines are all the distributions of the identified cluster, while the red line is for the benchmark that was identified as the most representative within the cluster.

While this parameterisation allows for a beneficial reduction of the number of points on which experimental constraints can be optimized, together with the fact that each benchmark was designed as the most representative of the cluster, the

generalisation of the inference from the benchmarks to a specific set of couplings is generally not guaranteed. First because the relationship between the clusters and the couplings is not bijective (one can compute limits at specific points, but not exclusions over a wide range), second because of the variability within a cluster (blue lines of Fig. 1.15).

This effect can however be somewhat circumvented by using Eqs. (1.68) and (1.69) to reweight from one set of couplings to another based on a parameterisation in the $(m_{\text{HH}}, \cos\theta^*_{\text{HH}})$ plane. This allows to generate events at a handful of points in the coupling space to produce distributions that are then morphed at any desired point. This procedure is similar to the matrix element reweighting in Eq. (1.52) with the additional benefit that the spin and colour final states remain unchanged and the integration phase-spaces are similar. The statistical uncertainties from the reweighting itself should therefore be minimal, and can be further reduced by using several samples at different coupling points reweighted to a single point. An unavoidable source of uncertainties however comes from the non-analytical nature of the procedure based on a statistically limited binned weight values.

Vector-boson fusion

The VBF production mechanism represents a valuable source of information about the coupling between the scalar and electroweak fields that is inaccessible from the dominant GGF mechanism. Comparatively, although its cross-section is an order of magnitude smaller than that of GGF, the presence of additional quarks with large angle separation can be used to reduce the background contamination. The additional term affecting the VBF diagrams to the Lagrangian of Eq. (1.67) can be written in a general way as

$$\mathcal{L}_{VBF} = \frac{v^2}{4} Tr\left(D_\mu \Sigma^\dagger D^\mu \Sigma\right)\left[1 + 2\kappa_V \frac{h}{v} + \kappa_{2V}\frac{h^2}{v^2} + ...\right], \tag{1.70}$$

where $\Sigma = e^{i\sigma_a \pi^a / v}$ are 2×2 matrices of the coset $SO(4)/SO(3)$ [141].

This parameterisation introduces two new couplings κ_V and κ_{2V} that govern the HHV and HHVV couplings, respectively, as illustrated on the diagrams of Fig. 1.16. It is interesting to note that the parabola case where $\kappa_{2V} = \kappa_V^2$ results in a cancellation between the two diagrams in the left of Fig. 1.16, as the longitudinal vector bosons scattering cross-section scales as $\mathcal{A}(\text{V}_L\text{V}_L \to \text{HH}) \propto (\kappa_{2V} - \kappa_V^2)$ [142]. Outside that regime, a more significant fraction of events presents a boosted HH spectrum. The collimated decay products that can emerge in that case – especially the large radius merged jets – can have a unique signature against most backgrounds. In that scenario, the sensitivity of the measurement of these two couplings could be very competitive compared to the GGF mechanism in the detection of BSM deviations.

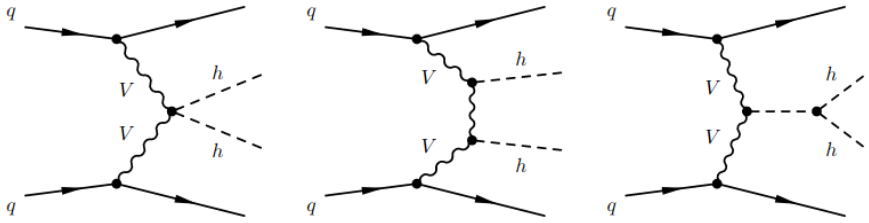


Figure 1.16 | Main diagrams contributing to the double Higgs production via the VBF mechanism, from Ref. [142]. The three diagrams scale with κ_{2V}, κ_V^2 and $\kappa_V \kappa_\lambda$ respectively.

1.7.3 Coupling modelling

Both the GGF and VBF have a non-trivial dependence on the coupling parameters, both in terms of cross-section and shape effects. Producing scans of couplings would be computationally intensive, especially in more than one dimension. This can be in parts circumvented assuming EFT description in terms of anomalous couplings, which allows an interpolation using only a handful of measured cross-sections.

For example, in the case of the pure GGF production in the SM part of Eq. (1.67), the two contributions are the box (denoted $\Box$) and triangle (denoted $\triangle$) diagrams on Fig. 1.14 ((a) and (b)) at LO. The total amplitude is given by $A = \kappa_t \kappa_\lambda \triangle + \kappa_t^2 \Box$, and therefore the cross-section reads

$$\begin{aligned} \sigma(\kappa_t, \kappa_\lambda) \sim |A|^2 = \; & \kappa_t^2 \kappa_\lambda^2 |\triangle|^2 + \kappa_t^4 |\Box|^2 + \kappa_t^3 \kappa_\lambda |\triangle^* \Box + \triangle \Box^*| \\ = \; & \boldsymbol{c}(\kappa_t, \kappa_\lambda)^T . \boldsymbol{v}, \end{aligned} \tag{1.71}$$

where $\boldsymbol{c}(\kappa_t, \kappa_\lambda) = (\kappa_t^2 \kappa_\lambda^2, \kappa_t^4, \kappa_t^3 \kappa_\lambda)$ is the coupling vector and $\boldsymbol{v} = (|\triangle|^2, |\Box|^2, |\triangle^*.\Box + \triangle.\Box^*|)$ is the vector of components that include the squared diagrams and their interference. Although this formulation is based on a LO interpretation of diagrams, at higher orders the cross-section would still scale with the power of the couplings and the discussion remains valid even if the content of $\boldsymbol{v}$ changes, and is also valid in a differential form.

Only three sets of couplings ($(c_1, c_2, c_3)^T \equiv \boldsymbol{C}$) and their cross-section ($(\sigma_1, \sigma_2, \sigma_3)^T \equiv \boldsymbol{\sigma}$) are necessary to determine fully $\boldsymbol{v} = \boldsymbol{C}^{-1} \boldsymbol{\sigma}$. This allows to compute the cross-section for any other set of couplings

$$\sigma(\kappa_t, \kappa_\lambda) = \boldsymbol{c}(\kappa_t, \kappa_\lambda)^T \boldsymbol{C}^{-1} \boldsymbol{\sigma}. \tag{1.72}$$

This can be extended to the c_2 coupling in the diagram (c) of Fig. 1.14 such that

$$\boldsymbol{c}(\kappa_t, \kappa_\lambda, c_2) = (\kappa_t^2 \kappa_\lambda^2, \kappa_t^4, c_2^2, c_2 \kappa_t \kappa_\lambda, c_2 \kappa_t^2, \kappa_t^3 \kappa_\lambda), \tag{1.73}$$

in which case six cross-sections must be known to solve the system. In case only

the c_2 scan is of interest, the other couplings can be set to their SM values and $\boldsymbol{c}(c_2) = (1, c_2^2, c_2)$, which reduces the number of required cross-sections back to three.

Similarly, for the VBF production using Eq. (1.70) and the three diagrams in Fig. 1.16 (denoted A, B and C in order)

$$\sigma(\kappa_\lambda, \kappa_V, \kappa_{2V}) \sim |A|^2 = \kappa_{2V}^2|A|^2 + \kappa_V^4|B|^2 + \kappa_V^2\kappa_\lambda^2|C|^2 + \kappa_{2V}\kappa_V^2|A^*B + AB^*| \\ + \ \kappa_{2V}\kappa_V\kappa_\lambda|A^*C + AC^*| + \kappa_V^3\kappa_\lambda|B^*C + BC^*|, \tag{1.74}$$

in which case six cross-sections of the VBF process must be used to fully determine the system.

Since the formulation of Eq. (1.72) can also be differential, any variable x can be used to obtain $d\sigma/dx$ at any coupling set from several distributions $d\boldsymbol{\sigma}/dx$ for which the couplings are known. The caveat is that the interpolation cannot reproduce exactly the shape obtained directly from the generation at one coupling point, the effect is fortunately minor unless the requested point shows more extreme features than the ones for which the shapes are available. For this reason, the initial coupling points must be chosen carefully to ensure maximal kinematic difference so that events populate all the corners of the phase space.

1.7.4 EFT validity

EFTs are very useful tools that allow extending the SM by providing a consistent picture of its potential deviations. Once experimental data has been used to put constraint on the EFT parameters, they can be reinterpreted in any type of BSM scenario. This is especially useful given the wide variety of models aiming at solving the issues listed in Section 1.4. The question of their validity yet comes to mind, with an emphasis on the indetermination that the cut-off value Λ introduces. This section, inspired by Refs. [124, 136, 143–146], will be dedicated to addressing the following questions with an emphasis on the scalar sector :

- What are the conditions under which the EFT is valid as a low-energy approximation of a BSM theory ?
- When is it justified to truncate the series to dimension-6 operators ($c_i^{(6)}$), and when is the effect of dimension-8 operators ($c_i^{(8)}$) non-negligible ?
- When is it relevant to consider only tree-level, and when should loop corrections or real emissions be included ?

It is important to note that answering these questions cannot be performed in a completely model-independent way. A massive particle above the cut-off scale can be integrated out into one or several Wilson coefficients, but its mass is not the only parameter that gives the Wilson coefficients their value, for example a coupling

can also enter the formulas. Constraining the Wilson coefficients can therefore only provide indications on the heavy resonance if assumptions (usually broad) are made. The range of validity might also depend on the degree of coupling of the new particle to the SM. Weaker couplings imply lower scales, and this assumption has an effect on the range of energy for which the EFT interpretation can be considered valid. Especially since the cut-off value cannot be directly determined in low-energy experiments, as low-energy observables only depend on $c_i^{(6)}/\Lambda$. In general, while the measurement of EFT parameters is performed in a model-independent way, their interpretation is not.

The assumptions are related to what is called *power counting*, the set of rules by which a UV theory coupling and mass scales are translated into coefficients that scale as Λ^{4-D}, where D is the dimension of the operator they relate to. Once the UV theory is known with its relevant energy scale Λ, then the validity of the EFT expansion can be recast as

$$c_i^{(6)}(g_*) = \frac{\tilde{c}_i^{(6)}(g_*)}{\Lambda^2} < \delta_i^{exp}(\kappa\Lambda), \tag{1.75}$$

where $c_i^{(6)}$ is a dimension-6 operator on a specific UV coupling g_* and all SM couplings (denoted g_{SM} further), $\tilde{c}_i^{(6)}$ its dimensionless equivalent, δ_i^{exp} its associated experimental upper limit constraint, and $0 < \kappa < 1$ the tolerance on the error caused by neglecting higher order operators. Eq. (1.75) then defines the space of validity in (g_*,Λ) of the EFT expansion for the UV in question. Note that while this description is a particular case of a single coupling, which still fits many theories, it can be generalised with the same conclusion.

This power counting is also necessary to assert the importance of each order of operator. So far, it was implied that the same underlying strength was applied on all orders, such that a dimension-6 operator scales as $g_*^2(E^2/\Lambda^2)$. But in the special case where $g_* \gg g_{SM}$, which can happen due to some selection rule or symmetry, then at a certain energy range ($\Lambda(g_{SM}/g_*) < E < \Lambda$) the dimension-6 operators are enhanced by a factor $(g_*/g_{SM})^2$. In that context, as already briefly discussed for Eq. (1.61), the squared dimension-6 contribution $(c_6^i)^2$ to the ME is dominant compared to the pure SM contribution. Comparatively, the dimension-8 operators interfering with the SM couplings to yield the same $1/\Lambda^4$ order, are still suppressed by a $(g_{SM}/g_*)^2$ factor and can still be neglected. If however, the stronger interaction only occurs at $D = 8$, these operators would dominate and the truncation at $D = 6$ would cease to hold, with the EFT expansion being still valid. Several special cases are discussed further in the following, with an emphasis on the ones relevant to this document.

For example, due to some symmetries, the operators at $D = 4$ (the SM ones) and $D = 6$ might be weaker due to some symmetry at low-energy, even if approximate. There is a specific example related to the GGF HH production mode. The operator

$\mathcal{O}_{\phi G}$ of Eq. (1.59) is suppressed if the Higgs is a composite Nambu-Goldstone [147], since it is not invariant of the Higgs shift symmetry $\phi \to \phi + a(\phi^\dagger\phi)\phi/f^2$ [148] (where a an appropriate constant and f an intermediate energy scale). Some dimension-8 operators, introducing scattering of two gluons and two Higgs bosons respect this shift-symmetry, and are therefore dominant [136]. For this specific model, it would therefore be required to include higher-order coefficients in the development of Eq. (1.68).

It is also possible that the dimension-6 operators vanish for reasons different from symmetry, but related to the UV model. For example, in triple-Higgs production the energy-growing piece of type $(h/v)^3$ has a non-zero coefficient that only comes from two dimension-8 operators [149, 150]. Another case, directly related to the resonant double Higgs production, are the light KK graviton and radion of Section 1.6. Since they are linked to the stress-energy tensor of $D = 4$, the operators of $D = 6$ can only come from loop-correction and are therefore subdominant compared to the tree-level $D = 8$ operators. Finally, this vanishing can be a simple accident within a UV theory. A fine-tuning that yields coefficients $\tilde{c}_i^{(6)} \ll \tilde{c}_i^{(8)}$, a rather "unfortunate" occurrence that can nonetheless happen.

Similarly to the dimension-6 truncation, often times only the tree-level contributions are considered. This usually makes sense as the loop corrections are suppressed by the expansion parameter $g_{SM}^2/16\pi^2$, as in Eq. (1.19). While going beyond tree-level in the SM is motivated by the usually better experimental precision and therefore the need for higher order corrections, as no leading order deviations from $D > 4$ operators has been observed they do not require such precision. This is nonetheless not always the case, and NLO corrections are sometimes required. For example, NLO QCD corrections might lead to a large k-factor in EFT predictions. As was already discussed in Section 1.7.1, tree- and loop-level contributions to dimension-6 operators might appear at different orders in the UV dynamics, leading to the same 1-loop level in effect. Another example related to the theme of this document is the decay modes of the Higgs boson. Due to their high precision at the per-mille level and compared to the $\mathcal{O}(10\ \%)$ effect of dimension-6 operators, 1-loop effects may be probed, for example four-fermion contact interactions. Similarly, real-emissions may sometimes be worth including, especially when additional jets have a direct impact on some observed quantities, for example the HH VBF production mode.

All the various points raised so far could cause the reader to understand that including higher order effects such as $D = 8$ and loop corrections should be considered from the start. This would however be imprudent, as not only will this complicate greatly the framework and interpretations, but the experimental sensitivity is nowhere near enough to be able to disentangle their effects from the tree-level dimension-6 operators. This is especially true for HH searches where the precision is particularly poor given their limited sensitivity. A safer approach would be to focus on the latter only, which in itself presents enough challenges, and only move further in complexity

in the particular cases where the $D \leq 6$ truncation validity breaks.

The question of the validity is of vital interest for any EFT interpretation to make sense. Although experimental constraints are model-independent, some considerations must be taken into account. Whether the constraints are upper limits as in Eq. (1.75) or two-sided intervals with a similar description, they should be provided together with the upper value of the kinematic variables that define the typical energy scale $M_{cut} = \kappa\Lambda$, where κ is defined as in Eq. (1.75). This would typically be the centre-of-mass energy $\sqrt{s}$ in a particle collider, although it may be challenging in proton-proton collisions such as at the LHC where that value varies for each event depending on the partons fraction of momentum. In the particular case of the VBF Higgs production mode, the total momentum transfer of the two additional quarks can serve as such scale. However, in the Higgs decay modes where neutrinos appear, this is impossible to estimate. In Ref. [144] it is suggested to use the p_{T} of the VBF jets as its estimator, which is valid as long as the Higgs is produced near the threshold.

Only by providing this M_{cut} value can the EFT range of validity be determined, similarly to what is done in Ref. [151] for example. A parallel method would be to provide the constraints both with the quadratic components on the coefficients. If the constraints are similar, then the truncation can be deemed valid. This could give a glimpse whether the constraints apply also to strongly coupled UV theories (for which the squared coefficients are not subleading) or only the weakly coupled ones. Additionally, truncation of the EFT series introduces errors that are model-dependent, and should be provided separately from the usual signal uncertainties with the assumptions associated to them. This is especially important for pre-LHC precision experiments, where the precent or per-mille level of precision does not translate directly to the same precision level for coefficients constraints. Setting these very constrained coefficients to zero in LHC analysis may be over optimistic [143].

1.7.5 SMEFT versus HEFT

The type of expected UV completion also determines the EFT framework. SMEFT assumes that the EWSB is linearly realised through the Higgs boson. HEFT is more general in the sense that the non-linear realisation is equivalent to the absence of symmetry and no relation is assumed between the Higgs and Goldstone bosons. Since the discovery of the Higgs boson SMEFT has gained in popularity, aided by a more restricted symmetry structure. The downside is that the range of UV theories that can be mapped to SMEFT is limited compared to HEFT. This begs the following question : which UV theories are precluded when working in SMEFT instead of HEFT, and when should the latter be solely used ?

As HEFT is more general, this question can be addressed by looking at the types of UV theories that cannot be translated into the SMEFT framework, which is a

matter of decoupling. Obviously, in case new physics arises at a higher energy scale that cannot be cast on the SM if such scale is taken to infinity, then only HEFT can be used. On the other extreme case, if the scale is close to the weak scale instead, both frameworks are allowed but HEFT can be used to resum into a series expansion of gv/Λ, effectively providing a better description.

In Ref. [152], a geometrical approach was followed to distinguish both frameworks. This approach is outside the scope of this document, nonetheless instead of a general discussion it is worth considering the particular cases where HEFT is either preferred or the only possibility. Two cases can be distinguished, the first one consists of particles that acquire all their mass from EWSB, the second when additional sources of EWSB are introduced.

Light BSM states such as the ones listed in Section 1.6 can be integrated out into Wilson coefficients within an EFT formulation. In the singlet scalar extension of Eq. (1.54), it is possible to do so using either the linear or non-linear, using respectively a weakly- or strongly-coupled regime, depending on the mixing angle of Eq. (1.56) [153]. In case that mixing is small, the two scalar states are decoupled, the linear expansion holds and dimension-six operators provide a good description of the single- and multiple-Higgs couplings. However, when the two states are strongly-coupled and the mixing angle is large, only the non-linear expansion holds and the linear and quadratic Higgs couplings to the massive vector bosons are decorrelated. As a side-note, the effect of this singlet extension could be detectable in double Higgs production through the modification of the $\mathcal{O}_{\phi G}$ operator in Eq. (1.63). This particular case illustrates a general observation that theories exhibiting non-decoupling effects can only be mapped to a non-linear EFT, as long as there is a well-defined decoupling limit. As the coupling depends on the parameters of the UV theory, they determine which of the two frameworks is more suited, and the transition between the two should be smooth.

On the contrary, when new sources of EWSB are introduced whose heavy states are integrated out, only HEFT can provide a reliable parametrisation. An example lies in the 2HDM also described in Section 1.6. While some parameterisation of the Lagrangian of Eq. (1.57) are done in SMEFT [154, 155], a study has shown that the introduction of a *v-improved* matching to include dimension-8 operators related to the non-zero electroweak VEV [156] improved the agreement with the full UV model. This is equivalent to a matching into HEFT [154].

Finally, SMEFT is built on the assumption that $E \ll \Lambda$, in which case the series converge quickly. Because of the infinite tower of operators, if $E \sim \Lambda$, this assumption breaks down and HEFT must be used. But even in the intermediate case where $E \lesssim \Lambda$, HEFT converges faster in terms of the number of orders to include, even if the SMEFT expansion exists [152, 156]. In case there is suspicion on the SMEFT convergence, which could indicate that the new physics might be located close to the weak scale, then HEFT is more transparent.

The question of the transition between HEFT and SMEFT is also more complex than it seems. It can be shown when developing Eq. (1.65) that the naive relations are on Table 1.5. Similar relations can be derived for the Warsaw basis, as done in Ref. [157], with the caveat that the gluon couplings contain a dependency to α_s, meaning they are scale dependent, something that is not a problem in the SILH basis. Assuming that these relations always stand is yet ill-advised, as it can only stand in case all dimension-6 operators are considered in interaction with the standard model, as well as their double-insertion in HH diagrams. However, in practice a truncation at the order Λ^{-2} is performed, either at cross-section level (assuming the SM is unsuppressed), or at the amplitude level (assuming the SM is suppressed). The former is more suited for fitting, but the Wilson coefficients are not restricted enough to the SM and the cross-section can appear negative. The latter can only provide qualitative results, but serves as a truncation validity test when its results are compared to the pure linear case. In any case, the relations of Table 1.5 should not be assumed as exact, and the question of the truncation of the SMEFT dimension-6 operators remains an open debate [157–159].

Table 1.5 | Relation between HEFT and SMEFT Wilson coefficients in Section 1.7.2 and Section 1.7.1 (with field redefinitions of Eq. (1.64)), from Ref. [138].

HEFT	SMEFT (SILH)
κ_λ	$1 - \frac{3}{2}c_H + c_6$
κ_t	$1 - \frac{1}{2}c_H - c_{t\phi}$
c_2	$-\left(\frac{3}{2}c_{t\phi} + \frac{1}{2}c_H\right)$
c_g	$128\pi^2 c_{\phi G}$
c_{2g}	$64\pi^2 c_{\phi G}$

In conclusion, there is currently no consensus as to which framework is better suited, and no experimental evidence that one is favoured. Although there is a preference for SMEFT given what we know of the Higgs boson, the arguments developed in this section show that it is too soon to dismiss HEFT. On the experimental point of view, their treatment is very similar as constraints can be put on each coefficient no matter the framework. There is however one major distinction in the relations between the EFT parameters, which is important in the HH searches. In SMEFT they are coupled, such that a global fit should be performed using both single and double Higgs measurements. This requires a larger collaborative effort but allows for stronger constraints, using the more sensitive single Higgs measurements of κ_t for example to further constrain c_2 (cf. Eq. (1.67)). In HEFT, the terms with order h^2 are not constrained by electroweak measurements, and only HH measurements can provide bounds for c_2, κ_{2V} and c_{2g}, although they are less sensitive to gluon couplings that are of higher order. While this limits the precision, it could prove a valuable test to select one of the frameworks.

Chapter 2

The CMS experiment at the LHC

This chapter is dedicated to the experimental side of high-energy physics and how discoveries and measurements are made in collider experiments in European Organization for Nuclear Research (CERN) facilities. After a brief overview of the history and scientific experiments carried out in this laboratory, this chapter will then focus on the Large Hadron Collider (LHC) and more specifically the Compact Muon Solenoid (CMS) experiment located on one of its interaction points (IP). Finally, this chapter will address the different reconstruction techniques used to detect and characterize the different particles produced from the collisions, and how they are treated in an offline analysis to study the rare phenomena physicists are after.

2.1 CERN and the Large Hadron Collider

From the aftermath of the second world war emerged the idea of an international laboratory that would allow cooperation in the fields of nuclear and high energy physics to cope with the ever-increasing scale of the experiments, and promote worldwide peace through scientific collaboration. This led to the official foundation of what became known as CERN by twelve European countries in 1954, followed shortly by the construction of the first buildings of the laboratory bearing the same name at the frontier between Switzerland and France near Geneva, in a town called Meyrin.

While initially focused on nuclear physics, over the years the laboratory oriented itself towards high energy physics and particle accelerators, and while nuclear physics is still the focus of a few experiments it is not dominant in the CERN landscape any more. Within the accelerators complex, the Large Hadron Collider (LHC) stands out as the largest particle accelerator in the world. Built in the tunnel that originally hosted the previous Large Electron–Positron Collider (LEP), this circular collider of 26.65 km is able to accelerate protons to an energy of up to 13 TeV in the

centre-of-mass, and 13.6 TeV recently. A series of smaller accelerators is used to accelerate the protons to the minimal energy at which the LHC can operate. The whole accelerator chain on Fig. 2.1 has been improved since 2018 in preparation for the ongoing Run-3 during the writing of this document to reach 13.6 TeV. However, the work performed in the following sections targeted the Run-2 data, and this section will therefore focus on the state of the laboratory at that period.

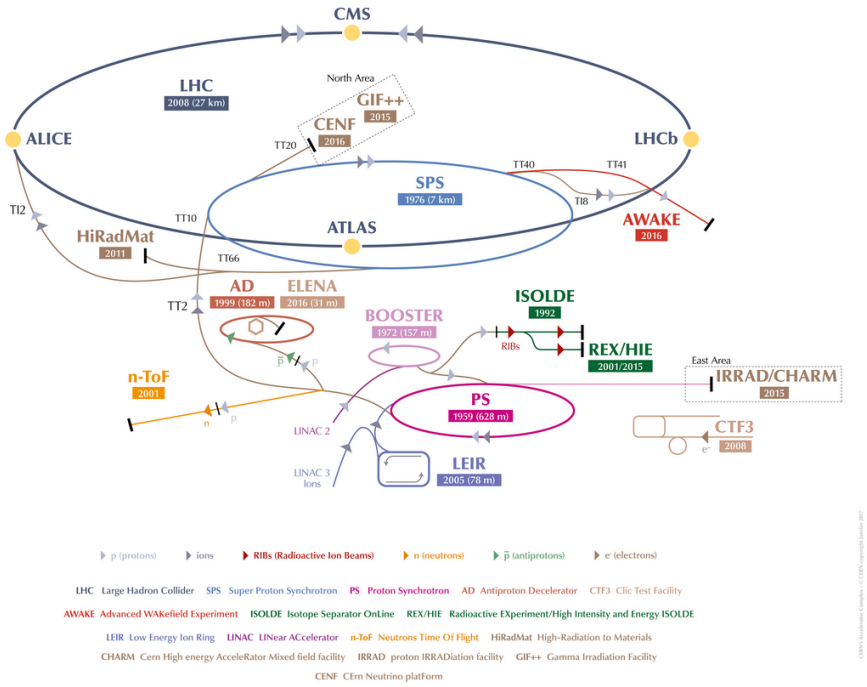


Figure 2.1 | Schematic view of CERN and the accelerator complex with the experiments attached, from 2016 to 2018 in Ref. [160].

Linear accelerators are particularly useful when the particles start at rest and are able to provide large accelerations with relatively low voltages. A linear accelerator (LINAC2) is therefore responsible for accelerating the protons extracted from a plasma of H_2 gas obtained from a duoplasmotron up to 50 MeV, while the LINAC3 is used together with the LEIR for heavy ion collisions that are out of scope for this document. First the proton beam is extracted from the plasma with a charged grid, providing a continuous flow of 90 keV protons that is packed into bunches by a quadrupole radio-frequency with an additional focus and gain of energy to 750 keV. Finally, the main part of the LINAC2, the Alvarez Drift-Tube Linac, brings the beam to the output 50 MeV energy level.

After a certain energy level, the linear accelerators however become impractical as the accelerating segments need to reach extensive lengths while maintaining a

sufficient gradient. At this point it becomes advantageous to use circular accelerators instead with a single radio-frequency cavity, the inconvenient being that magnetic dipoles need to be used to keep the beam within the pipe, the intensity of the magnetic field increasing as the beam is accelerated at every turn. Focusing the beam is also of high concern as a more focused beam will produce more collisions, for such a task several quadrupoles are aligned in a so-called drift section to focus and defocus in alternative directions in order to squeeze the beam.

The first circular accelerator following the LINAC2 is the Proton Synchrotron Booster (PSB), a superposition of four 127 m long rings accelerating the beam to 1.4 GeV. The bunches provided by each ring are concatenated, together with two additional bunches 1.2 s later and an additional gap, evenly spaced, and provided to the next stage.

The Proton Synchrotron (PS) has a length of 628 m and its role is to accelerate the beam up to 25 GeV and most importantly to perform the so-called *Radio-Frequency (RF) gymnastics*. The bunches are split in several steps until the beam is organised in the *train* that is expected by the LHC where the bunches are separated by 25 ns (7.5 m as the beam is very close to the speed of light). At the end of the PS accelerating procedure of 3.6 s, there are 72 bunches.

The stage before the LHC is the Super Proton Synchrotron (SPS) that now accelerates the beam to 450 GeV in its 6.9 km tunnel. The machine also performs operations on the bunches called a *super-cycle* of 21.6 s to fill the two counterclockwise beams of the LHC in several iterations. This requires precise synchronisation with the LHC to insert the train at the correct place after shortening the bunches. While the initial design was to provide in total 2808 bunches to the LHC, technical issues in the SPS limited this number to 2208 in 2016, later increased to 2556, and forced the introduction of a new filling scheme to allow more intense bunch and not degrade the amount of collision data, the BCMS scheme [161]. In that scenario the PSB provides eight bunches instead in two cycles followed by an empty space. These bunches are merged into four in several steps to double proton density, and the beam continues to the usual accelerator steps.

The two LHC beams circulate in 26.65 km long parallel pipes located around 100 m below the French-Swiss border. Due to geological reasons associated with the civil engineering of the tunnel, the plane of the machine is built with a 14 mrad angle with respect to the horizontal. To minimize beam losses coming from collisions between the proton beam and air molecules, the pressure is kept at ultra-low levels between 10^{-6} to 10^{-11} mbar. The vertical magnetic field required to curve the beam trajectory is performed by 1232 superconducting NbTi dipoles cooled down to 1.9 K with superfluid helium, allowing the magnet to perform the ramp up of the magnetic field from 0.54 to 7.7 T as the beam is accelerated. This design choice at very low temperature requires the two pipes to be thermally insulated and enclosed in an iron flux return yoke embedded in a common cold mass. Additionally, 540

quadrupole magnets are used to focus the beam with higher order fields such as 688 sextupoles for correction of the energy dependence on the magnetic field and 168 octupoles for beam stability. The beam is accelerated by eight superconducting radio-frequency cavities, that provide an energy increase of 60.6 keV per turn during the 20 min ramp-up time, and afterwards 7 keV per turn to keep up with the small energy loss encountered by the beam. To achieve such an acceleration the peak field strength of the RF cavities is 5.5 MV/m with a precise frequency of 400.79 MHz, slight variations are however needed to match the speed increase of the beam, of the order of 1 kHz.

The LHC is divided into eight arcs of 40° and eight straight sections of 528 m each, this originated from the LEP design aimed at minimizing the synchrotron radiation loss at each turn, a factor not much relevant any more due to the larger mass of the proton. Each section is associated with its surface and underground installation, referred to as *LHC points*, illustrated on Fig. 2.2. Four experiments are located on Interaction points (IP) where the beams are exchanged between outside and inside the ring to ensure the same path length : ATLAS [162] and CMS [163] are multipurpose high-luminosity experiments featuring a main toroidal and solenoidal magnets respectively, LHCb [164] is optimised for the study of flavour and b-physics and ALICE [165] who specialised in quark-gluon plasma studies in heavy ion collisions. Additional smaller experiments such as LHCf [166], MoEDAL [167] and TOTEM [168] are also located around the four main experiments and benefit from the same collision point, albeit with a much smaller fraction of scattered particles. Around each collision point, dipoles are used to merge the two beams into a single pipe, and closer quadrupoles provide the final focusing before the collision point.

Beam losses are a major concern in hadron colliders, as major losses of the 250 MJ energy stored in the beam could result in catastrophic damage to equipment. Even minor losses of the beam could provide enough energy for the dipoles to heat up and lose the superconductive state, a phenomenon known as a *quench*. Large dump resistors are used to absorb the 1 GJ energy of the 13 kA current flowing through each of the arc cryostats, a situation that would interrupt LHC operations and must be avoided as much as possible. Active monitoring of the beam and passive protection are handled by the Machine Protection System (MPS). Many sensors are scattered around different constituents of the accelerator – experiment IP and beam position and losses monitoring equipment, RF cavities, vacuum valves, electrical system, etc – to perform early detection of a potentially dangerous event, while passive protection in the form of collimation systems in LHC points 3 and 7 are used to block particles exiting the beam path due to large transverse oscillation amplitudes or energy errors. Each of the 100 collimators is made up of four parts in a 1.2 m long jaws with a 2 mm aperture used to absorb primary beam halo and following hadronic showers.

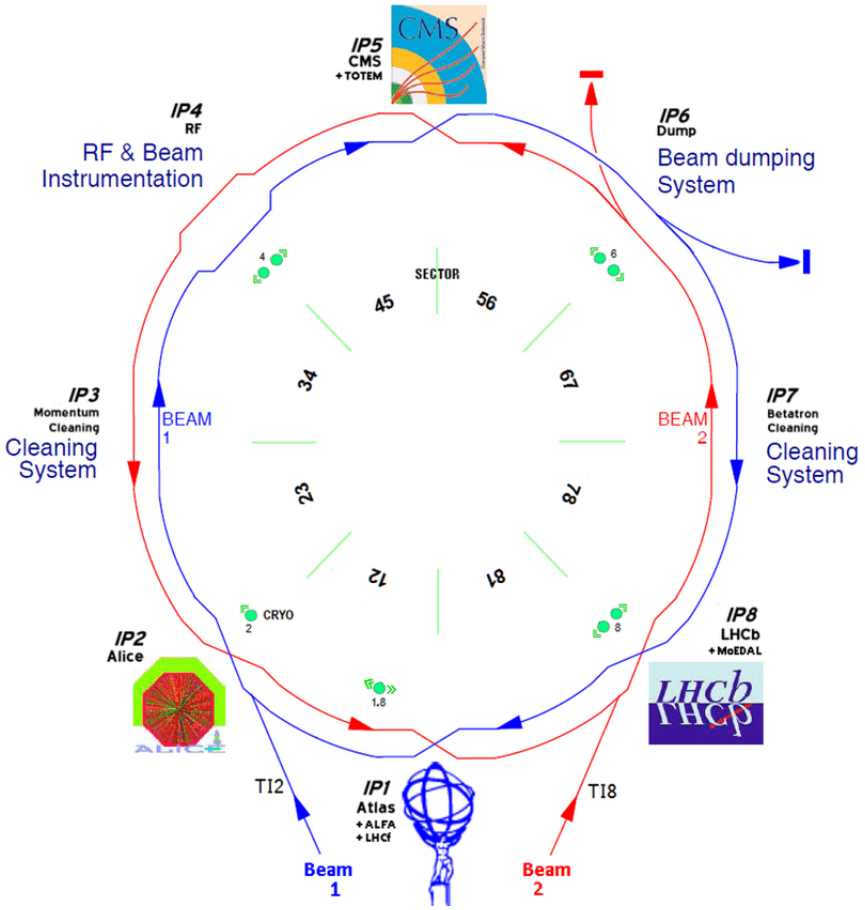


Figure 2.2 | LHC layout with its eight Interaction pointss (IPs) labelled IP1 to IP 8, from Ref. [169].

In case of failure of the beam, the LHC Beam Dumping System (LBDS) is used at point 6 to dump both beams into a 10 m long graphite blocks. 15 fast so-called *kicker* magnets are used to deflect the beam out of the ring into a set of 15 other septum magnets for deflection in the vertical axis. A particle-free gap is left in the two LHC beams with a length of 3 µs called the *abort gap*, enough for the kicker magnets to ramp up. Any failure in the synchronicity of this procedure or residual beam in the abort gap is handled by absorbers around the magnets. The intensity of the beam needs to be spread out to avoid damaging the dump block before impacting it, this is performed by 10 kicker magnets that effectively sweep the beam over the target. The beam dump procedure is also regularly performed at the end of each cycle of the LHC filling.

The LHC cycle consists first in the injection of the *probe beam*, a low intensity beam used to perform safety measurements and checks of the accelerator subsystems. If the conditions are deemed safe and stable, the probe beam is dumped, and the physics beam is injected at point 2 and point 8 for the clockwise and counter-clockwise pipes, respectively. Once the filling phase is completed, the RF cavities start accelerating the beam from the SPS injection energy of 450 GeV to the nominal Run-2 energy of 6.5 TeV, a procedure that takes about 20 minutes. Next, the *squeezing* step is performed for about 30 minutes to focus the beam and yield a larger amount of collisions for the experiments. Then the beam is adjusted into collision mode, declared stable and the data taking period starts. This period lasts on average 8 hours but can be kept for much longer, although the rate of collisions unavoidably declines due to the proton density reduction from the collisions. The beam is finally dumped, and the whole cycle can start again about two hours later.

2.2 Luminosity and pileup

A factor as important as the centre-of-mass energy of a particle collider is its luminosity. This parameter directly impacts the amount of data collected in a detector and should be maximized as much as possible. The instantaneous luminosity defines the fraction between the scattering process rate and its cross-section, when the beam has a Gaussian profile it can be expressed as [170] :

$$\mathcal{L} = \frac{f_r N_b N_p^2}{4\pi\sigma_x\sigma_y} = \frac{f_r N_b N_p^2}{4\pi\epsilon\sqrt{\beta_x^*\beta_y^*}} : \tag{2.1}$$

where the parameters and their values along the data taking years between 2016 and 2018 are :

- $f_r = 11.245$ kHz is the revolution frequency,
- $N_b =$[2208,2546] is the number of colliding bunches (variable per data-taking year),
- $N_p = 1.1\ 10^{11}$ is the proton density per bunch,
- $\sigma_{x,y} = 11$ µm is the standard deviation of the assumed Gaussian shaped beam density profile at the IP,
- $\epsilon = 0.3$ nm is the un-normalized beam emittance,
- $\beta_{x,y}^* = 40$ cm is the minimum of the betatron function $\beta_{x,y}(s)$ – as a function of the s position along the ring – at the IP, by definition $\sigma_{x,y} = \sqrt{\epsilon\beta_{x,y}^*}$.

Some assumptions of Eq. (2.1) might not be entirely fulfilled in the following cases :

- Head-on collisions : in a real-life collider, the two beams interact with a small cross angle to avoid multiple bunch collisions and parasitic effects such as

long-range electromagnetic interactions. The following reduction factor needs to be taken into account :

$$F = \frac{1}{\sqrt{1 + \left(\frac{\sigma_z \theta_C}{2\sigma_x}\right)^2}}, \tag{2.2}$$

where

- $\theta_C = 260$ µrad is the crossing angle,
- $\sigma_z = 70$ mm is the longitudinal RMS bunch length,
- $\sigma_x = 11$ µm is the transverse RMS bunch size under the assumption that $\sigma_x = \sigma_y$.

This reduces the luminosity by about 15 %.

- The betatron minimum is expected to be at the IP, its waist profile created by the squeeze of the quadrupole magnets – called *minibeta insertion* – might not however be at the point where the beams cross.
- Non-Gaussian characteristics and misalignments between the beams are not covered by Eq. (2.1).

The luminosity computed from the beam parameters is often referred to as *instantaneous luminosity* $\mathcal{L}_{\text{inst}}$, and the integral over time yields the *integrated luminosity* $\mathcal{L}_{\text{int}} = \int \mathcal{L}_{\text{inst}} dt$. The former is however not a constant over time, between cycles when the parameters of the beam have changed but also within a cycle when the protons are being "burned off" by either losses within the ring or in the collisions themselves. This leads to an exponential decay of the luminosity $\mathcal{L}(t) \propto exp(-t/\tau)$ where τ is typically of the order of hours or days. Its precise measurement is therefore of utmost importance in any analysis, since it usually represents one of the largest sources of uncertainty.

Several methods exist to provide a real-time value of the instantaneous luminosity, one would be to measure each parameter of Eq. (2.1) but that would require very precisely measured parameters and could be impacted by the assumption failures detailed above. Another method, independent of the formula but detector-dependent, is to rely on reactions for which the cross-section is known to a high degree of precision and using luminometers where the rate of a quantity is given by

$$R = \sigma_{\text{ref}} \mathcal{L}, \tag{2.3}$$

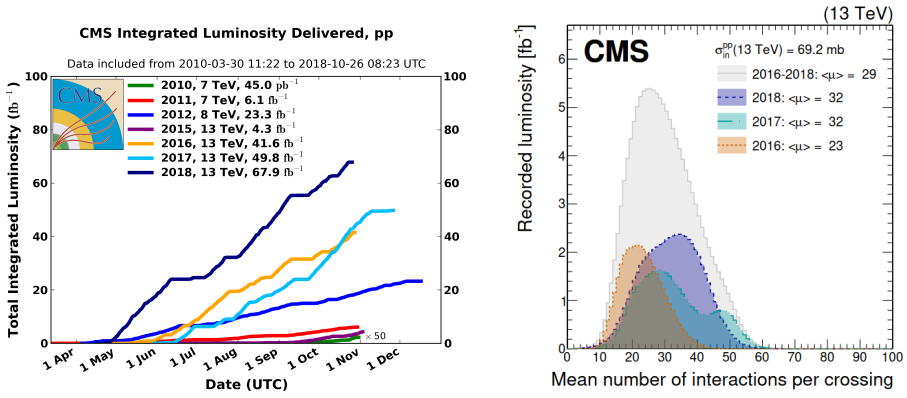
where σ_{ref} is the cross-section of the reference process, which for a hadron collider can be chosen as the inelastic scattering of a proton. In this case, one can define the average number of collisions per Bunch Crossing (BX) called *pileup* μ and the luminosity can be obtained with

$$\mathcal{L} = \frac{\mu f_r}{\sigma_{\text{inel}}} = \frac{\mu_{\text{vis}} f_r}{\epsilon \sigma_{\text{inel}}} = \frac{\mu_{\text{vis}} f_r}{\sigma_{\text{vis}}}, \tag{2.4}$$

where the efficiency for the inelastic scattering to have detectable effect has been introduced, so the formula can relate to the different counting methods available for measurements.

Several parts of the detector are used independently to perform the same measurement and provide an accurate value, yet this method needs to be calibrated to the value of σ_{vis}. This is usually done once a year in a specific data-taking period through a *van der Meer* [171] scan where the two beams are varied transversely. This scan allows for a precise evaluation of the beam spot parameters $\sigma_{x,y}$ and through Eq. (2.1) the instantaneous luminosity during this run, where all the parameters are carefully selected and the beam is close to the Gaussian shape. This measure then allows for a very precise measurement of σ_{vis} whose value can be used conjointly with the luminometers even if the accelerator parameters vary during data-taking. The evolution of the recorded luminosity is shown on Fig. 2.3a and the improvements brought by the CERN beams department in terms of the filling scheme yielded an ever-increasing trend towards larger amount of data collected.

A large pileup such as the one recorded by CMS on Fig. 2.3b poses serious challenges in terms of event reconstruction when more vertices of interaction and many more particles in the detector are included in a single collision. This can be kept under a manageable level through *luminosity levelling*, where some parameters impacting the luminosity are modified at the beginning of a LHC cycle, so the pileup is not too extreme with a still high luminosity. Then, as collision go by, the proton density decreases in the bunches and with it the pileup, allowing the restoration of the beam parameters that maximize the luminosity.



(a) *(b)*

Figure 2.3 | Evolution of the integrated luminosity (left) and average pileup (right) along the data-taking years for the CMS experiment between 2010 and 2018. From Ref. [172] and Ref. [173].

Data-taking periods at the LHC are grouped into *runs* of a few years, between which long shutdowns of a couple of years are dedicated to major LHC and experiments maintenance and upgrades. A year of data-taking is split into roughly ten months of proton collisions, about a month of heavy ion collisions, and a technical shutdown during Winter.

During the first LHC run, referred to as Run-1, the Compact Muon Solenoid (CMS) collaboration recorded nearly 30 fb^{-1} at centre-of-mass energy of 7 and 8 TeV. The energy was then raised to 13 TeV after the long shutdown and CMS recorded a total of 150.76 fb^{-1} during three data-taking years of Run-2 – namely 2016, 2017 and 2018. As of the writing of this document, the Run-3 has just finished its collection of data for 2022 at the centre-of-mass energy of 13.6 TeV.

2.3 The Compact Muon Solenoid experiment

The CMS experiment is a 21 m long cylinder, with diameter of 15 m and a weight of approximately 14000 T, located inside the experimental cavern of the LHC IP5 a hundred meters underground of the town of Cessy, France. It acts as a multipurpose fully enclosed particle detector able to detect a wide variety of particles, identifying them and measuring their kinematic properties with 10 different subdetectors illustrated on Fig. 2.4.

Two coordinate systems exists within the CMS detector. First the Cartesian coordinate system has its x-axis directed toward the centre of the LHC ring, the y-axis is vertical with respect to the plane and pointing upwards while the z-axis follows the anti-clockwise beam trajectory when seen from the top. Due to the slight angle of the LHC machine, the x-axis and y-axis are not perfectly horizontal and vertical compared to sea level. The x-y plane is commonly denoted as the *transverse* plane. To take advantage of the detector symmetry, a spherical coordinate system is defined with the centre of the detector – where the collisions are supposed to happen – as the origin. The three coordinates are the distance from the centre, the azimuthal angle ϕ in the transverse plane relative to the x-axis and the polar angle in the longitudinal z-y plane relative to the z-axis. It is however common to define the pseudorapidity as

$$\eta = -\ln\left(\tan(\theta/2)\right), \tag{2.5}$$

which becomes useful considering that for massless particles – or any relativistic particle – this expression is equivalent to the rapidity, and that a rapidity difference between two particles is a Lorentz invariant. From there, the angular distance can be defined as

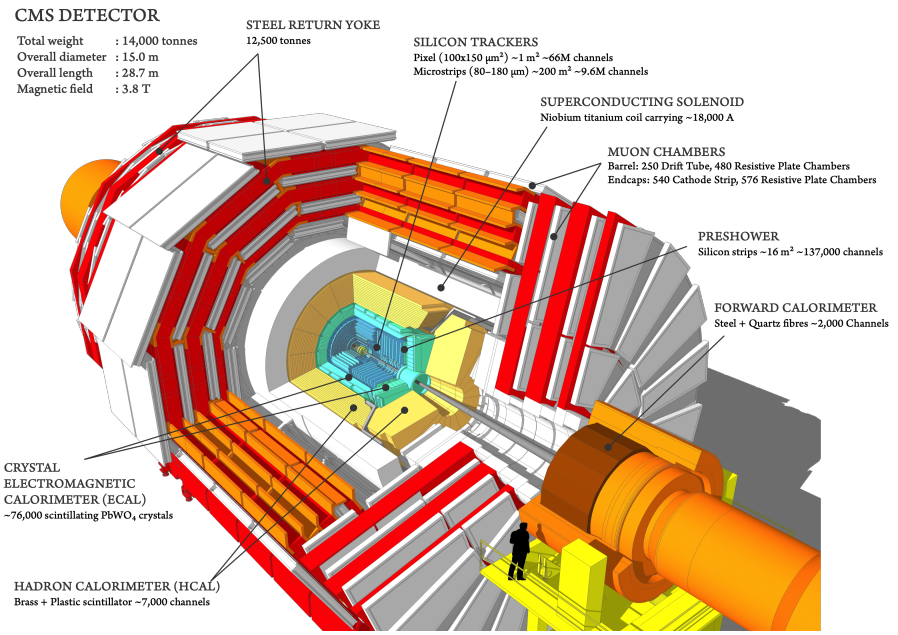


Figure 2.4 | Schematic view of the CMS experiment with the various subdetector constituents, from Ref. [163].

$$R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \tag{2.6}$$

The detector geometry defines a *central* region inside the barrel, and a *forward* region in the two endcaps. Two types of detection methods are used for position and energy measurement : tracking where the position is recorded with the smallest impact on the particle as possible which allows measuring the bending of the trajectory inside a uniform magnetic field and infer its energy, and calorimetry where the particle is stopped and all its kinetic energy converted in another form of measurable energy. One major component of the detector is its 13 m long and 6 m diameter solenoid magnet, generating the internal 3.8 T uniform magnetic field required for tracking with a 18000 A current thanks to its superconductive Nb-Ti material cooled down to 5 K. Within its volume is located the silicon tracker, the closest detector to the beam with by order of distance to the IP the pixel then strip trackers. The electromagnetic and hadronic calorimeters surround the tracker and extend all the way to the magnet, they are responsible for measuring photons and electrons, and the heavier elements respectively. Outside the magnet inner volume are the muon chambers, that use different gas detection methods for tracking the muons escaping from the detector and taking advantage of the closed magnetic loops captured by the steel yoke that provides a 2 T magnetic field.

With a collision rate of 40 MHz, and the quantity of the signal channels coming from each of the listed subdetectors generating of the order of one megabyte of data per collision, the CMS detector would generate about 40 terabytes of data per second. Most of this data consists in inelastic scattering of the initial protons, a QCD process that is well understood and therefore not interesting to be saved on tape, nor would it be feasible to do so given the amount of bandwidth required. The trigger system has been designed to cut down on this rate to about 1 kHz by only selecting collisions that display certain features, determined a priori based on physics considerations and the search for exotic signatures. This is performed through a two-stage procedure : the first stage called the Level 1 Trigger (L1) is hardware based for a fast but crude decision-making and leaves a 100 kHz rate to the second level, the High-Level Trigger (HLT) that performs event reconstruction on a farm of CPUs to make more educated decisions and yield the desired rate. The events passing both stages can then be recorded and used in offline reconstruction phases.

The next sections will be dedicated to a more in depth description of each subdetector and trigger system. More details can be found in Ref. [163].

2.3.1 The silicon tracker

The silicon tracker has a diameter of 2.5 m and a length of 5.8 m, with a pseudorapidity extension up to $|\eta| < 2.5$. It is made up of the pixel and the strip detectors whose main difference resides in the dimension of each n-type semiconductor junction, a choice dictated by the distance from the IP as illustrated on Fig. 2.5. When a charged particle crosses the silicon wafer, charges are generated and collected by the read-out chip and stored in an internal buffer if the collected charge exceeds a certain threshold. If the event passes the first level of the trigger, the stored information is digitized and sent to the data acquisition system for reconstruction.

The silicon tracker is the closest subdetector to the IP, its goal is to measure the trajectory of all charged particles emerging from the collision to allow measurement of their momentum through the bending in the magnetic field, as well as the identification of the different interaction vertices emerging from the hard process and the pileup. Given the high track multiplicity, to achieve this goal the silicon tracker needs to have a very fine granularity – especially the closest layers to the IP – and response time, but also a high radiation tolerance to cope with the harsh environment. It in turns requires an efficient cooling of the sensor to evacuate the heat generated by the electronics and minimise the effect of thermal noise, but also to lessen leakage current and extend the life of the sensors that could undergo thermal runaway. This happens when the cooling is not sufficient any more, or is interrupted due to technical problems, and a positive feedback loop between leakage currents and temperature leads to the sensor becoming irreversibly unusable. At

the end of the Run-2, it already happened for several modules – without hindering the very good performance of the tracking efficiency – but could have more impact during the Run-3 with the integrated luminosity increase. While the pixel tracker is kept at -5 °C, the strip tracker cooling has been brought down from -15 °C to -20 °C after the 2016 data taking and is expected to eventually run at -25 °C during Run-3. Cooling also comes with the cost of an increased material budget on top of the support, electric and optic cabling, which can be an inconvenience as the interactions between the particles with the material in the tracker should be minimized to avoid energy loss or emission of secondary particles that make the reconstruction more complex. This has been studied and optimized, the entirety of the silicon tracker has a radiation length X_0 ranging from 0.4 to 1.8 and a nuclear interaction length λ_i between 0.1 and 0.5, each respectively characterising the energy loss from electromagnetic and nuclear inelastic interactions with the material.

The pixel tracker initially consisted in 2016 of three layers in the barrel and two disks in each forward region. To cope with the increasing luminosity, the Phase-1 upgrade [174] in 2017 saw the introduction of an additional layer and disk that brought the pixel tracker to four layers spanning from 2.9 to 16 cm from the IP in the BPIX and three disks on each side in the FPIX with a distance between 29.1 and 51.6 cm from the centre of the detector. The 1856 modules consist of a lattice of 160×416 pixel cells in a rectangular active area of 16.2×64.8 mm^2, each n-in-n cell with a size of 100 × 150 μm^2. Thanks to this very small granularity the pixel tracker achieves a very good efficiency of 99 % or above, and the hit resolution of about 10 µm × 25 µm in the transverse and longitudinal directions provides a precise measurement in two directions – z-ϕ for the barrel, r-ϕ in the endcaps – while the third one comes for the position of the sensor from alignment calibration studies.

The strip tracker has a more complex geometry : in the barrel the four layers of the Tracker Inner Barrel (TIB) and the six layers of the Tracker Outer Barrel (TOB) are completed by the 3 disks forming each of the Tracker Inner Disk (TID), in the endcaps there are nine disks for each Tracker End Cap (TEC). This brings the strip tracker to an area of 198 m^2 covered with 15148 modules split into 9.3 million channels. Further, from the IP the granularity requirements can be relaxed and the z position in the barrel – r in the endcaps – becomes less relevant for the measurement of the magnetic bending of the trajectory. Therefore, the strip detector consists of elongated rectangular p-on-n single sided silicon sensors, with a pitch from 85 µm to 205 µm, a much larger length of up to 20 cm, and a varying width ranging from 320 µm to 500 µm the further the sensor is from the IP. A precise position measurement can only be determined in two directions given the position of the sensor, r-ϕ in the barrel, z-ϕ in the endcap. Some so-called *stereo*-modules were built with a second strip module tilted by 100 mrad, with the objective to

correlate the hits between the two sensors and recover the measurement in the remaining direction. Because of alignment considerations, it was however found to be suboptimal and stereo-modules are considered as two separate strip sensors. With a hit efficiency above 99 % for the TIB, TID and TEC, and above 98% for the TOB – its inner layers suffering at high luminosity – the strip tracker shows excellent performance despite many years of operation.

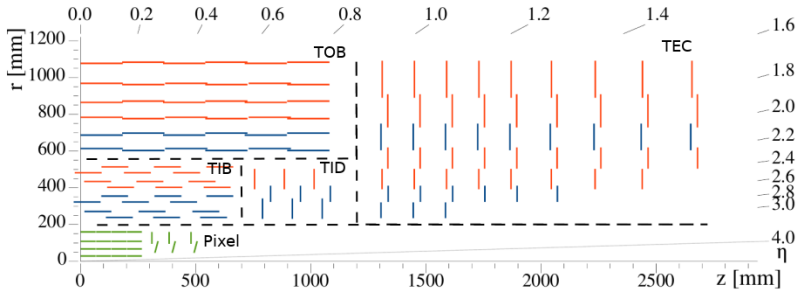


Figure 2.5 | cross-section of one quarter of the CMS silicon tracker in the r-z plane. The silicon pixel tracker modules are in green, the silicon strip stereo modules in blue and the single sided silicon strip are in orange. From Ref. [175], with the additional delimitations of the different subcomponents.

2.3.2 Electromagnetic calorimeter

The Electromagnetic calorimeter (ECAL) is responsible for measuring the electromagnetic shower produced by charged and neutral particles – mostly photons and electrons – thereby measuring their energy. It is composed of the barrel (EB) that covers the pseudorapidity range up to $|\eta| = 1.479$ and the endcap (EE) extending the range to $|\eta| = 3.0$, as depicted on Fig. 2.6.

The ECAL is expected to contain the electromagnetic radiations in a short and tightly contained shower to have good energy and position resolutions, and to be radiation hard. For its information to be used in the first trigger level, it also needs to have a fast response time. The selected sensitive material is a lead-tungstate ($PbWO_4$) crystal, its high density (8.23 g cm^{-3}) and short radiation length allow for a compact volume yet opaque to electromagnetic showers since the total thickness accounts for 25.8 X_0 Is Moliere radius, that describes the transverse scale of the electromagnetic shower, is also relatively small (2.2 cm), and with a short scintillation decay time about 80% of the light is collected in the bunch crossing window of 25 ns. While radiation hard, the crystal's transparency can be affected and a continuous calibration of the ECAL is required through the injection of two laser pulses during the LHC abort gap, the effect of which can be measured and compared to the expected value to perform corrections. The transparency is also highly dependent

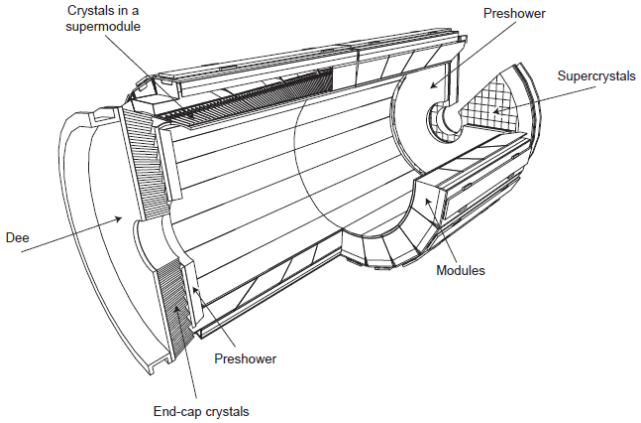


Figure 2.6 | Schematic view of the CMS Electromagnetic calorimeter (ECAL), from Ref. [176].

of the temperature, with a light output decreasing by 2.1 % per degree, hence the ECAL is kept at a constant 18 °C.

The EB consists of 61200 crystals with a rectangular cross-section of 26×26 mm^2, which corresponds to an aperture of 0.0174×0.0174 in the (η,ϕ) plane, and a length of 230 mm. To ensure proper coverage and avoid the issue of the crystal neighbour interface, the crystals are not pointed directly at the IP but rather with a slight angle. The EE on the other end is made up of 7324 crystals of 28.62×28.62 mm^2 with a length of 220 mm, also slightly angled towards the IP. The scintillation light emitted per MeV of deposited energy amounts to 4.5 photoelectrons that are readout by a photodetector, whose type differs because of the different levels of radiation and magnetic field orientation : Avalanche photodiode (APD) for the EB and Vacuum phototriode (VPT) for the EE.

The signals from the photodetectors are preamplified, shaped and digitized before being buffered. The energy sums and shower profiles of *supercrystals* made up of 5×5 crystals are sent to the L1 at every bunch crossing. If the read-out signal is brought back then the information about the 10 previous bunch crossings is sent to the read-out system in the experimental cavern, the contamination from the previous bunch crossings collected energy is worked out and suppressed from the processed signal.

The EE being in the forward region is subject to more radiation than the EB, and has a worse angular resolution, especially to high energy neutral pions decaying to two photons that might be too collimated. To alleviate this issue, an extra detector called the *preshower* is placed prior to EE in the particles paths. Two lead layers with radiation length of 2 and 1 X_0 are used as sampling calorimeter to pre-emptively initiate the shower in combination of a silicon strip detector for both layers.

2.3.3 Hadronic calorimeter

The Hadronic calorimeter (HCAL) is a sampling calorimeter whose aim is to measure the energy of hadrons such as pions, kaons, protons, and neutrons. The absorbing material must have a small nuclear interaction length λ_i such that the depth of the detector is enough to stop most of the hadrons. For that purpose, a mix of steel and brass plates has been chosen. In order to use the information from the detector in the L1, a fast response time is crucial, and plastic scintillators serve the purpose well. Similarly to the ECAL, the signal is read out and digitized and sent to the service cavern in a pipeline, awaiting a potential reconstruction, while only the more basic information contained in the longitudinal information is translated into transverse deposit via a look-up table in an approximate but fast process, then sent to the L1. Calibration is performed through ultraviolet flashes injection into the sensitive material to extract the corrections to the radiation damage, additionally some radioactive sources can be introduced in the steel tubes for further calibration.

The HCAL is divided into four sections, as illustrated on Fig. 2.7. The barrel (HB) and endcap (HE) are responsible for the coverage of the pseudorapidity range of $|\eta| < 1.3$ and $1.3 < |\eta| < 3.0$ respectively. Limited by the space left between the HB and the solenoid, it only accounts for 5.82 λ_i at $\eta = 0$ and 10.6 λ_i at $\eta = 1.3$, which in the most central region is not enough to catch the most energetic hadronic shower, despite the ECAL adding another 1.1 λ_i on average. The HO was added outside the solenoid magnet enclosure to extend the detector to longer showers, taking advantage of the solenoid volume and the magnetic return yoke first iron layer as additional absorption material. This ensures the entire HCAL is not shorter than 11.8 λ_i except at the barrel-endcap transition region. The final subdetector, the HF is located in the very forward region, 11.2 m from the IP to allow coverage up to $|\eta| = 5.2$.

The HB absorbers consist of a 40 mm steel front plate, eight 50.5 mm and six 56.5 mm brass layers, and a 75 mm steel backplate. They are used to develop the hadronic shower, while the active part of the energy measurement is undertaken by 70000 plastic scintillator tiles that transform the deposited energy into photons collected by wavelength shifting fibre (WLS). These are in turn connected to hybrid photodiodes (HPDs) that amplify the photoelectrons of a photodiode into an electrical signal. The scintillators are made in tiles of cross-section 0.087×0.087 in the (η,ϕ) plane in a single longitudinal section. The HO granularity matches roughly the one of the HB with five rings of scintillator sections and an additional section around $\eta = 0$ together with a 19.5 cm thick iron plate to compensate the minimal absorber length in the central region.

The HE absorbers and scintillators are very similar for the HB apart from thicker brass plates at 79 mm and the granularity of 0.17×0.17 at $|\eta| > 1.6$. One major difference however is the longitudinal segmentation of the scintillators so that each

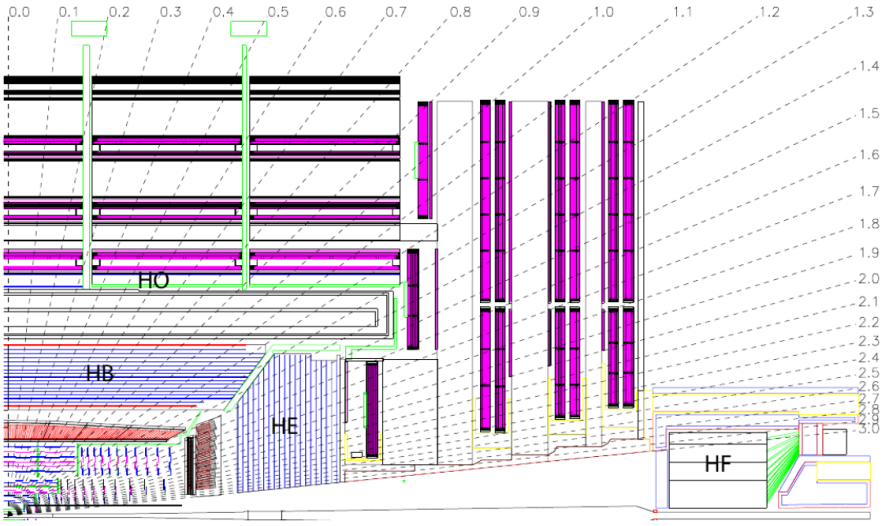


Figure 2.7 | Schematic view of a quarter of the CMS hadronic calorimeter (HCAL), from Ref. [177].

layer can be calibrated separately and the more significant radiation damage in the forward region handled in a more precise manner. While this segmentation was in two to three parts, it was increased to seven in 2017 for a more precise calibration. Similarly, the HPD were replaced by Silicon Photomultiplier (SiPM) that had better radiation hardness.

The very forward region at the edges or the detector and very close to the beam line is covered by the HF. Given the heavy particle fluxes, the sensitive material must have high radiation hardness. As the environment would be too hostile for scintillators, quartz fibre detectors are much more resistant and used instead, alternating with steel plates as absorber. While the former allow for 68 % of the energy to be measured in the 25 ns bunch crossing window, the latter has an excellent timing resolution with pulses 10 ns wide. Inside the sensitive material the shower produces a Cherenkov light channelled through the fibre into a Photomultiplier tube (PMT) protected by a steel and concrete shielding. Discrimination against electromagnetic shower is performed between fibres that span the whole length of the HF and fibres that do not cover the first 22 cm such that a shower originating from an electron or a photon would only be detected in the longest fibre contrary to hadrons that produce a deeper shower.

2.3.4 Muons chambers

Muons are often present in the final states of many physics processes of interest, and thanks to their low energy loss through the material, their tracks can extend outside the calorimeters and the solenoid, which represents a very clear signature. An accurate estimation of their position and energy is therefore crucial for precision measurements. Muons are charged particles detected as such in the inner tracker. This detector nevertheless lacks muon identification capabilities and does not contribute to L1 decisions. Instead, muon chambers installed in the outer layers of the detector are protected from much of the hadrons that could punch through the calorimeters, solenoid magnet and the steel return yoke whose gaps they are embedded in, while still benefiting from the closed magnetic loops for momentum measurement. The much lower track multiplicity compared to the inner tracker allows for less granular and cheaper gas detectors that can cover a larger area with a reasonable response time. The barrel region, where neutron-induced background and muons rate are low, is occupied by drift tube (DT) up to $|\eta| = 1.2$, while the endcap that is exposed to less uniform magnetic field and higher radiation contamination uses cathode strip chamber (CSC) in the pseudorapidity range $0.9 < |\eta| < 2.4$, as illustrated on Fig. 2.8. While these types of detectors have a relatively long response time yet with excellent time resolution of about 5 ns, additional very fast with poorer position resolution resistive plate chamber (RPC) are installed in both the barrel and the endcap up to $|\eta| < 1.9$.

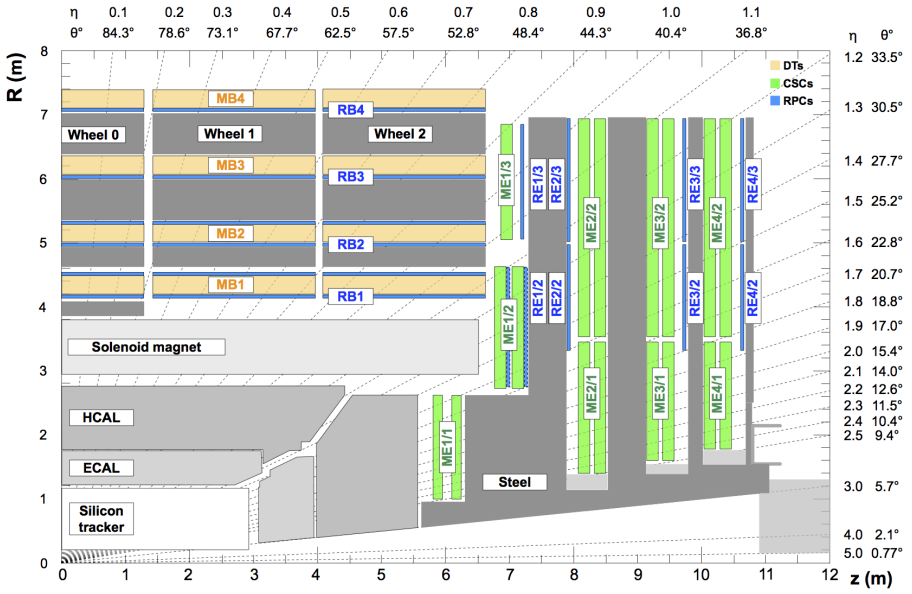


Figure 2.8 | Schematic view of a quarter of the CMS muon chambers, from Ref. [178].

The DT installed in the barrel are chambers of 44 $\times$ 13 mm^2 cross-section containing a 50 µm gold-plated stainless steel wire, with a length between 2 and 3 m. A constant voltage between the electrode strips and this wire acting as the anode allows for the collection of charges generated through ionisation by a charged particle traversing the CO_2-Ar gas volume. The drift time provides a transversal measurement of the position with respect to the wire, with a maximum drift time of 400 ns. Four layers of chambers form a *superlayer*, and two z-aligned superlayers form a *station* in the outer layers, providing a r-ϕ measurement with a 200 µm resolution, while in the inner layers a perpendicular superlayer is added perpendicularly to provide additional z measurement with a resolution between 200 and 600 µm. Four barrel layers amount to 250 chambers and 172000 sensitive wires.

In the endcap the CSC are trapezoidal chambers where 6 radial cathode strip planes supplement 6 tangential anode wires. This perpendicular layout allows for a precise ϕ measurement from the strips and a coarser measurement or r by the wires of the charges produced in the Ar-CO_2-CF_4 gas mixture. Four wheels of these chambers are located in each endcap, for a total of 480000 readout channels.

The RPC modules supplementing both the barrel and the endcap are filled by a mixture of $C_2H_2F_4$, C_4H_{10} and SF_6 gases in a 2 mm gap. Read-out strips are used to detect the charge avalanche produced by the energy deposit from the high voltage. These strips are aligned with the beam in the barrel and are radial in the endcap for coarse measurements of ϕ due to their low position resolution of the order of the centimetre. Their fast reaction time is below the nanosecond, albeit with a pulse shape of 100 ns that effectively blinds the chamber and makes it more suited for triggering rather than precise measurement. The barrel contains six stations while each endcap contains four, amounting to 130000 channels.

The signals of each subdetector are amplified and shaped through appropriate methods before being sent to their read-out systems, including the previous bunch crossings information. In parallel, some basic information is sent to L1 from correlated tracks in several chambers that serve as input to a track searching to be used as trigger primitives. If the read-out signal is received, the signal shape accumulated over several bunch crossing windows is used in the reconstruction of the muon trajectory and together with the bending of the magnetic field its momentum.

2.3.5 Trigger and data acquisition systems

Level 1 trigger

The L1 has to cope with a very large rate of events and must make a decision in less than 3.8 µs to avoid the overflow of the buffers containing the whole detector data. For this purpose it is built with hardware components, reprogrammable FPGAs or custom build ASICs, able to provide very fast yet coarse decisions. It receives data

from ECAL, HCAL and the muon system as depicted on Fig. 2.9, the inner tracker being too complex for the task. The decision lies on whether some potential particle candidates are present in an event, such as electrons or photons (indistinguishable without the inner tracker), muons, jets, hadronic decay of a τ, missing transverse energy from a non-detected neutrino, and more importantly if these candidates meet the requirements to be deemed fit enough to go through to the second stage. For example a minimal p_{T}, isolation, track, or calorimeter quality. If these requirements are met, the L1 Accept (L1A) signal is sent to the Trigger Control and Distribution System (TCDS) to all the subdetectors to initiate the full read-out within the HLT. The different subsystems of the L1 trigger are illustrated on Fig. 2.9.

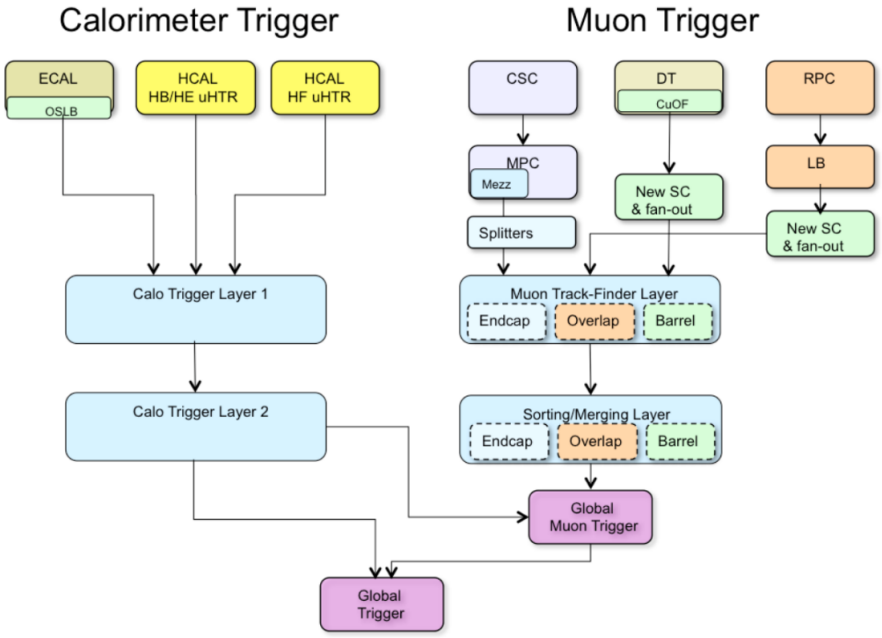


Figure 2.9 | Level 1 Trigger (L1) organigram, starting from the primitives emerging from the calorimeters and muon chambers, down to the final trigger decision after treatment through the various algorithms, from Ref. [179].

The calorimeter trigger combines the information from both the ECAL and the HCAL. For the purpose of triggering, and in order to limit the amount of information to be treated by the hardware, the calorimeter cells are grouped into Trigger Tower (TT) that define the granularity of the first level reconstruction, a matching between one HCAL and a 5×5 square of ECAL cells. The calibration is performed by the first layer, then time-multiplexed to the second layer where jet reconstruction is performed. From there the reconstruction phase attempts at identifying the different objects that could have left significant deposits in terms of TT, by first determining

the *seed* of the tower clusters, its local maximum.

Jets candidates are built from a 9×9 lattice of TT around the seed in the (η,ϕ) plane. The contamination from pileup is evaluated from the neighbouring clusters with rectangular 3×9 shape, out of which the three lowest contributions are removed and the jet energy is calculated from the TT, before being further calibrated based on its p_{T} and η values.

The energy deposits of e/γ will be most significant in the ECAL, and candidates are dynamically build by adding seed neighbouring TT into a cluster. Because of bremsstrahlung or photon conversion when the electron trajectory is bent in the magnetic field, the shower can have a larger profile in the ϕ direction and the shape in the calorimeter will be asymmetric, the algorithm is therefore allowed to build a lattice up to a rectangular shape of 2×5 in (η,ϕ) on which a shape selection is performed. To avoid a jet being associated with a e/γ candidate, the corresponding deposits in the HCAL are compared and a veto above a certain ratio of energy is performed. Finally, an isolation selection is performed by summing the energy inside a 6×9 lattice around the candidate but excluding it, and based on the level of pileup a dynamic cut is performed. Candidates of hadronic decays of τ follow a similar procedure in the HCAL, with the addition that two neighbouring clusters can be merged. The twelve leading e/γ candidates are passed to the next L1 stage.

The muon system is processed in parallel. The hits from both the DTs and CSCs – the two subdetectors with the best position resolution – are combined into track segments, while the hits from the RPCs are clustered together. These serve as trigger primitives and fed to the track-finder after being separated into three categories : the barrel tracks with $|\eta| < 0.83$ and handled by the Barrel Muon Track Finder (BMTF), the endcap tracks with 1.24 $|\eta| < 2.4$ by the Endcap Muon Track Finder (EMTF), and the overlap between the two is treated by the Overlap Muon Track Finder (OMTF) in the 0.83 $|\eta| < 1.23$ region to allow for an optimal treatment of the barrel-endcap interface. The lower occupancy of the BMTF allows for a road-search algorithm to be used for the definition of muon candidates, while the EMTF and OMTF use a more simple and robust look-up table method with patterns derived from a prior offline analysis. Finally, the candidates from the three systems are merged, sorted and potential duplicates are removed before being passed to the Global Muon Trigger (μGMT). This last step performs the comparison with the information emerging at the end of the calorimeter branch to evaluate the pileup corrections to the muon tracks and apply quality cuts. Then eight best candidates are sent to the final layer.

The Global Trigger (μGT) is responsible for the final decision if the event is to be read out or not. Based on the information from all the candidates, a wide list of about 300 pre-programmed rules, known as *trigger paths*, are evaluated on the candidates. If any rule of this *trigger menu* returns a positive value, the L1A signal is sent out. Some algorithms can be very basics, such as for calibration or monitoring

studies where a random selection is performed to limit their bandwidths, a process known as *prescaling*. Other algorithms are quite complex and rely on the presence and combined kinematic properties of several candidates that are characteristic of an exotic physics signature worth recording. In this case the probability for such rules to be fired is small enough that no prescaling is necessary.

Data acquisition and High-Level trigger

Thanks to the advance of processing power and network transfer speed, the second level of the trigger system composed of a farm of about 22000 CPU cores is able to process the whole read-out of the detector taking advantage of the finest granularity. When the L1A signal is propagated back to the detector buffers, the stored information is sent to the Front-End Drivers (FEDs) located in the service cavern. There, the optical analogue signal is digitized and the first steps of event reconstruction is performed. Next it is passed down to Front-End Readout Optical Links (FEROLs) that assemble the data from a certain number of FEDs and sent to the surface with 10 GB optical links. Commercial computing cores assemble multiple fragments into a *super-fragment*. The HLT farm then handles the reconstruction of the full event from the super-fragments where each event is assembled by a Builder Unit (BU) and stored on RAM, the whole chain is on Fig. 2.10.

Despite the processing power of the HLT, full reconstruction of each event from the 100 kHz rate provided by the L1 would need a decision to be made in a few hundred micro-seconds. However, several algorithms performing the reconstruction of high-level physics objects with similar methods as the offline reconstruction, albeit simplified for faster results, can make the total reconstruction take as much as a few seconds. The solution to this paradox is to process many consecutive events in parallel, such that there is always a core available to treat an event coming from the L1. Optimisation of the HLT decision flow allows keeping the number of cores to a reasonable level without risking full occupancy and data loss. It was designed as a succession of multiple independent steps of reconstruction and filtering in increasing order of complexity, so that an event that would not fulfil any of the roughly 500 trigger paths of the HLT menu can be rejected as soon as possible, even before its final reconstruction, with an average of 50 ms per event. The HLT is the basis on which the multiple offline analyses are constructed and its constitution is constantly tuned with respect to the luminosity and the physics search strategies, while keeping the rate budget into careful consideration. Monitoring and calibration streams are kept with a prescaling that evolves during a fill, as the instantaneous luminosity and pileup are time dependent. The lowest granularity of a time segment of data is called a Luminosity section (LS) and represents about 23 s of data-taking. They are then merged and sent to the CERN computing centre called *Tier0* for long term storage or offline reconstruction. The different reconstructed quantities built by the trigger system are kept in the stored data for potential further analysis, although

offline reconstruction uses more accurate and complex algorithms.

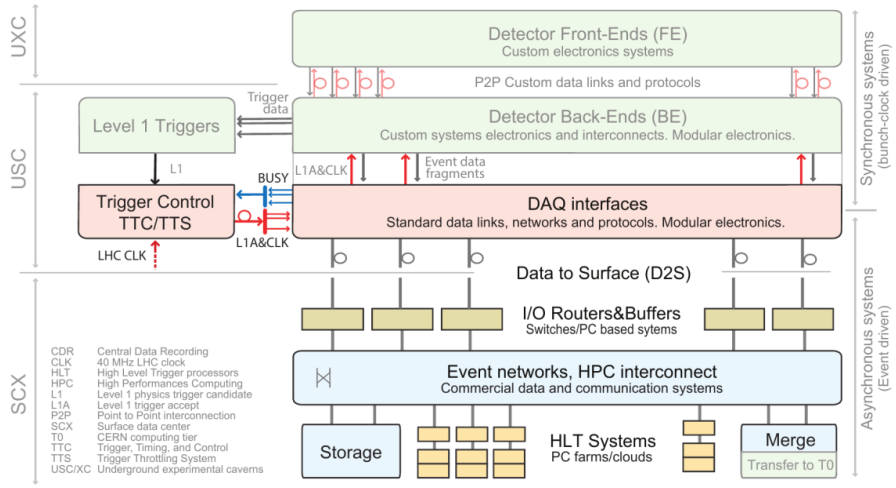


Figure 2.10 | Representation of the different stages of the High-Level Trigger (HLT). While the entire detector information awaits in the backs-ends, the trigger data is being processed in the Level 1 Trigger (L1). If the L1 Accept (L1A) signal (in red) is sent to the Data Acquisition (DAQ) interface, the data fragments (in grey) are received, aggregated and transferred to the surface data centres for full reconstruction and the filtering through High-Level Trigger (HLT) before being stored. From Ref. [180].

2.4 Data reconstruction and enhancement techniques

At the end of the online data-taking, the events are encoded into a raw format that contains the low level information produced by each sensor. To perform an offline analysis and extract meaningful interpretations, higher level physics objects associated to particles need to be reconstructed, calibrated and their quality assessed. This multistep procedure involves a collaborative effort that can be reproduced multiple times as reconstruction techniques and calibration constants are refined. From these objects, complex selections can be performed to enhance the sensitivity of a specific search for a phenomenon of interest. These multiple steps are the focus of this section, starting from tracking and its link with calorimetry, then the reconstruction of the different objects and finally the methods used in offline analyses.

2.4.1 Tracking and calorimetry

The information obtained from the silicon tracker consists in collections of hits into the different sensors, either pixels or strips. The first step towards producing tracks is the grouping into *clusters*. As a charged particle crosses the strip or pixel junction, the geometrical effect and the magnetic field causes the energy deposit to extend to several neighbours, a phenomenon known as *charge-sharing*. *Zero-suppression* artificially sets to zero the signal of cells below the noise threshold, and the obtained charge profile can then be used to determine the hit position with a precision below the simple width of the sensor. In the pixel detector, this is done by comparing the shape of the cluster with templates obtained from complex simulations, while the strip detector bases its clusters on the signal-to-noise ratio. In both cases, the Lorentz drift due to the magnetic field in the barrel and potential radiation damage are taken into account.

The particle tracks can be built from the positions extracted from the cluster, this requires notably a very precise determination of the sensor position through alignment studies [181]. In the first step, the tracks can be approximated as helices developing in parallel to the beam axis. The track curvature R allows for a measurement of the transverse momentum of the particle

$$p_{\mathrm{T}}\,[\mathrm{GeV}] = 0.3\ B\,[\mathrm{T}]\ \ R\,[\mathrm{m}]\,. \tag{2.7}$$

However, this naive approximation is spoiled by the energy losses, non-uniform magnetic field, multiple scattering and secondary emissions sustained by the particles through the material. Electrons have for example a probability of 85% to emit a bremsstrahlung photon and hadrons have a 20% chance of undergoing nuclear interaction before reaching the calorimeter. In addition, the high track multiplicity of the inner tracker creates many hits that could lead to fake tracks, just by combinatorial effect between hits in different layers. To minimize this, while keeping the efficiency high, a Combinatorial Track Finder (CTF) [182] method is performed through several iterations. Each iteration is a combination of four steps whose parameters can evolve depending on the iteration level, and the hits from each confirmed track can be removed in the next iterations to reduce the complexity. First a *seed* is generated most often in the pixel tracker, then a Kalman Filter (KF) [183] method extrapolates the track to the outer layers, a more complex fit is applied on the entire track and finally some quality criteria are evaluated for the track.

The track seeding uses hits from the pixel detector, its fine granularity allowing a lower occupancy compared to the strip tracker. At first pairs of hits in adjacent layers matching a track passing by the beam location are formed, then associated into groups of three to four hits, a number necessary to build a rough curved trajectory and are considered as seed. In further iterations, hits from the strip

tracker and the muon systems can be used once the tracks initiated by the pixel have been removed as seeds, in order to reconstruct potentially displaced secondary vertices corresponding to long-lived particles emanating from the IP. Additional requirements are introduced in these iterations regarding the transverse momentum and impact parameter.

The so-called *track finding* step starts from the seed parameters and extrapolates the position of the next hit in the outer layers, including the uncertainties on each parameter. In the first iteration, perfect conditions are assumed and the trajectory follows a helix, while in further iterations the experimental inaccuracies are included. The track information is updated with each hit until the last layer is reached, or if the track does not meet enough quality criteria, such as too many missing hits or a transverse momentum below the threshold. The procedure is then performed inwards, updating the tracks with hits that might have been missed in the first pass.

Once all the hit candidates for a track are selected, the whole track parameters are refit using a KF with all the inhomogeneities of the magnetic field and potential material interaction nuisances, and then smoothed. This provides a final track candidate whose quality is assessed through a multi-variate analysis (MVA) of the number of missing hits and intercepted layers, the χ^2 score of the fit, and the compatibility with the beam spot, these requirements being iteration level dependent. Multiple tracks can be obtained from the same hits with the KF and therefore duplicates needs to be cleaned out.

Muon tracking [184, 185] benefits a lot from the muon chambers information, given the lower track multiplicity. The drift time of the DTs can provide a position resolution smaller than the chamber side profile, while CSCs use the combined wire and strip information to assign hits. Both detectors have stacked chambers which allow a quasi straight segment to be defined and provided as seed to the Kalman Filter, with the additional information coming from RPCs strip hits that are clustered, their charge-weighted average defining the cluster position. The tracks built from these three subdetectors are called *standalone* muons. The combination between the muon tracks from the inner and outer tracker can be done in either directions. *Tracker muons* refer to tracks in the inner tracker that are extrapolated to the muon systems and matched to one of their identified segments. *Global muons* are propagated "outside-in", and the matching of the tracks from the inner and outer tracker is checked with a KF filter in a combined fit. Tracker muons were introduced to have high efficiency in regions with less instrumentation and low p_{T} tracks where no match is recorded in the outer stations, at the cost of a higher misidentification rate. Global muons benefit from the entire information of the inner and outer tracker and suffer less from misidentification, especially above 200 GeV. Standalone muons only use the outer tracker, hence their momentum resolution is worse, and they are less distinguishable from cosmic muons.

Electrons on the other hand suffer more from bremsstrahlung effect in the bent trajectory and large energy losses. Electron tracking [186] therefore either uses as seed the ECAL energy deposits before being propagated through the tracker, or uses the usual track finding step but requiring a match in the ECAL. However, the KF was built with the assumption that the uncertainties covering the energy losses were Gaussian in nature, which is not always true for electrons that can suffer from catastrophic energy loss, for example when emitting a high energy photon. This would cause a high number of missing hits, or a particularly bad χ^2 rendering the KF method less effective. To account for these effect, an alternative filter has been developed as a Gaussian Sum Filter (GSF) [187]. In this method, the Bethe-Heitler formula that describes the fractional energy loss of electrons through bremsstrahlung is approximated using a Gaussian mixture model, where Gaussians of different amplitude are propagated through the different layers with their associated energy loss. The criteria for including hits into a track are looser than those of the KF. The matching or seeding with the ECAL deposits requires a cluster position and uncertainties to be defined. Such cluster is first defined as a seed with higher than threshold and neighbours energy deposit, it is then expanded if the energy of the adjacent cells are above another threshold based on the noise. These define topological clusters for which a Gaussian shape is attached with variable amplitude and position, and a fixed width. The resulting clusters are defined as a Gaussian mixture of these topological clusters, and the former position and amplitudes are used in combination of the tracks to define a GSF track that goes through an adapted track quality evaluation.

Reconstructed tracks allow for the estimation of position of the proton-proton collisions within a bunch crossing. That position where the hard process occurred is called the Primary Vertex (PV), by opposition to the other vertices emerging from pileup interactions. The PV provides, among other things, a handle to mitigate the pileup contribution but also a comparative point from where secondary vertices associated to long-lived particles can be identified. The vertex reconstruction consists in selecting a subset of tracks passing some quality criteria and using the z position of their point of closest approach (PCA) to the expected centre of the beam position. Once the clustering is performed, the position of the vertex is estimated from at least two tracks and the PV associated with the hard process is the one with the highest $\sum p_{\mathrm{T}}^2$ from all the objects that the clustering algorithm associated to the vertex. For each LS, the fit of the primary vertices positions allows for a measurement of the beam position.

2.4.2 Object reconstruction

In this section, we will describe how the Particle-Flow (PF) [188] method combines the tracks and calorimeter deposits to generate particle candidates available for offline physics analyses. The PF algorithm, having access to the whole detector

reconstructed low level objects, can correlate measurements to define the particles present in the event along with their kinematic properties. A *PF* block is defined as a group of objects of different types that can share the same origin, for example when there is a calorimeter deposit whose centre is aligned with the extrapolation of a track in the inner tracker.

For each of the PF block defined in an event, a sequence of reconstruction is performed to identify physics objects, starting with muons, then electrons, isolated photons, charged and neutral hadrons with their associated photons, and finally the Missing Transverse Energy (MET). At each iteration of this sequence, the PF elements used to identify a particle are removed from the PF block. After identification by the PF an object can be subject to multiple selections to classify it into different categories with varying purity and efficiency, referred to as a Working Points (WPs).

This section will provide an overview of the different object offline reconstructions. The HLT reconstructed steps performed online during data-taking will not be described, but they only differ through simplifications and a more regional application of the methods described in this section.

Muon reconstruction

The first step in the PF reconstruction of the muons is the evaluation of their isolation. The scalar sum of the p_{T} of all charged particles and calorimeter deposit E_T is summed within a cone of radius ΔR, usually $\Delta R = 0.4$. This sum should not exceed 10 % of the muon candidate p_{T}, otherwise the muon is considered non isolated. In that case, the muon track could actually be a high-p_{T} hadron that was able to *punch through* the solenoid and calorimeters to leave a distinctive trace in the muon chambers.

Muon identification relies on tracks reconstructed with tight quality criteria in one of the trackers. In the case of the inner tracker, the energy deposit shape in the calorimeter cells is used to differentiate it from hadrons. This is to avoid wrongly identifying muons as coming from the hard process, while they actually come from leptonic decays of heavy or long-lived hadrons. These muons are defined as passing the loose WP. They can originate from either the PV or from a hadron decay, suppressing the contamination from cosmic muons. On the other hand, muons passing the tight WP can only originate from the PV, they have to be reconstructed in both the inner tracker and muon chambers, with at least six hits in the former, including one in the pixel detector, and a matching segment in at least two muon stations of the latter. On top of that, additional cuts on the χ^2 track score, transverse and longitudinal IP cuts are also applied. The medium WP was developed as an intermediate set of criteria leading to higher efficiencies (Fig. 2.11). Similarly, a loose and tight WP exist regarding the lepton isolation values.

The measurement of the muon momentum requires a fit of its trajectory, four algorithms are used. For muons with $p_{\mathrm{T}} < 200$ GeV, the bending is sufficient for a precise measurement of the momentum by the inner tracker that has a better position resolution than the muon chambers, and the *inner tracker fit* method uses only the information from the latter. For higher values the track becomes straighter, and the small reach of the inner tracker makes it less optimal compared to the volume of the muon chambers, the combined information is therefore used in the *Tracker-Plus-First-Muon-Station fit* method. The two last methods are the *picky fit* employed when there is shower in the muon stations and the fit only considers the most compatible hits, and the *dynamic-truncation fit* is more oriented towards large energy losses on the muon trajectory by performing the fit starting from the closest layer in the inner tracker and looking for close segments in the muon stations. The Tune-P algorithm performs the choice between the four algorithms based on the results of each fit, their goodness of fit and associated uncertainties.

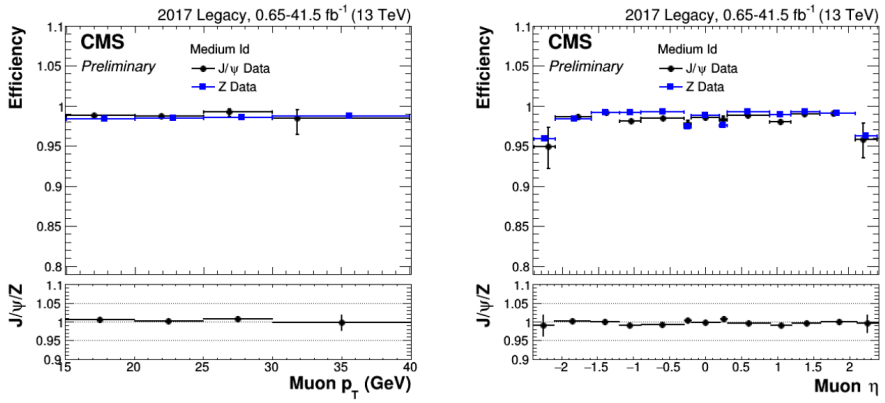


Figure 2.11 | Muon medium id Working Point (WP) evaluated on 2017 dataset using a Tag-and-Probe (T&P) method, as a function of the transverse momentum (left) and pseudorapidity (right). From Ref. [184].

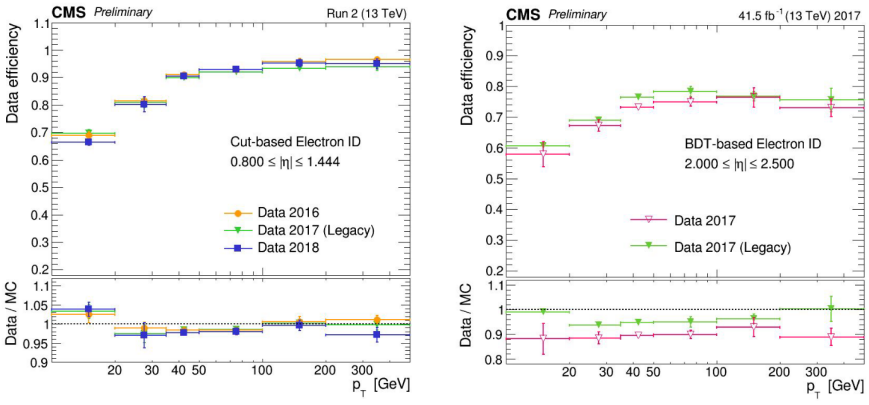
Electron and photon reconstruction

Electrons and photons are related through bremsstrahlung and electron-positron pair creation, they are therefore treated together. Their isolation is defined based on the transverse energy sum of a cone located around the object with angular spacing of $\Delta R = 0.3$, similarly to the definition of the muon isolation. The identification of electrons resides in the connection between a GSF track and an ECAL energy deposit, to which maximum two tracks can be associated for the electron identification [186] to be performed. The electron momentum is obtained through both its track curvature and the calorimeter measurement, although above $p_{\mathrm{T}} \simeq 20$ GeV the bending radius in the tracker becomes large enough that the energy resolution of

the ECAL dominates. On the other hand, photons are reconstructed solely from ECAL deposits, with the requirement that there are no linked tracks. Additionally, there should not be any large HCAL clusters associated to either electron and photon PF candidate, otherwise a corrective factor depending on the E_T and η is applied to the ECAL measurement.

Several backgrounds can yield non-prompt leptons, for example converted photons, misinterpreted hadrons, or electrons emitted by the weak decay of hadrons inside a jet. Several selections are therefore performed to provide WPs for prompt electrons. Several parameters of the track and ECAL cluster are used in either a cut-based method optimized on $t\bar{t}$ simulated events (Fig. 2.12a) or a Multi-variate analysis (MVA) method using a Boosted Decision Tree (BDT) trained on Drell-Yan+jets simulations (Fig. 2.12b).

Photons are considered if they are isolated from other tracks and if the cluster profile is compatible with the expected shape. Similarly to the electron, both a cut-based and MVA identifications are defined, except that since photons do not produce tracks, the only variables are defined on the shape inside the two calorimeters. Non isolated photons however cannot be considered by the PF independently of hadrons.



(a) *(b)*

Figure 2.12 | Performance of the cut-based (left) and BDT (right) approaches as a function of the transverse momentum. From Ref. [189].

Jet reconstruction

Once muons, electrons and isolated photons are removed from consideration in the PF algorithm, the remaining objects can be reconstructed as hadrons, either charged or neutral, and non isolated photons, for example from the decay of a π^0. These can originate from either a hadronic tau, or a parton that went through parton

shower. The latter is reconstructed as an object called a *jet*. While charged hadrons carry on average two thirds of the jet energy, the remaining energy is carried by neutral particles from which about three quarters are photons with good energy resolution provided by the ECAL and one quarter is neutral hadrons that can only be measured in the HCAL with a lower resolution. Thanks to the PF access to the whole detector information, the three detectors can be combined to perform a global measurement of the total jet energy and avoid double counting between tracks and calorimeter deposits, despite the complexity of hadronic interaction within the tracker and missing neutral tracks.

ECAL and HCAL clusters not associated to any tracks are interpreted as neutral particles, respectively photons and neutral hadrons. Inside tracker acceptance, clusters associated to tracks can be attributed to charged hadrons. The compatibility between the cluster energy and the different tracks p_{T} is assessed in a new fit and any excess in the calorimeters is attributed to neutral particles. Outside tracker acceptance, only the correlation between ECAL and HCAL can be used to interpret clusters as single hadrons, while ECAL-only clusters can be attributed to photons. A clustering algorithm is then performed on all the identified PF objects.

The aim of a clustering algorithm is to identify a jet from such cluster, with a total sum of particles energy as close as possible to the initial parton. On top of the potential presence of multiple jets whose contributions need to be disentangled, clustering algorithms needs to follow additional requirements to ensure stability in regard to perturbative QCD, namely to be infrared- and collinear-safe, such that adding a soft radiation or splitting a particle into two contributions should not change the jet collection produced by the algorithm. The anti-k_T algorithm [190] satisfies these conditions, it performs iterative steps following the distance

$$d_{ij} = min\left(p_{T,i}^{-2}, p_{T,j}^{-2}\right)\frac{\left(\Delta R_{ij}\right)^2}{R^2} \quad , \quad d_{iB} = p_{Ti}^{-2}, \tag{2.8}$$

where ΔR_{ij} is the angular distance between particle i and j, $p_{T,i}$ and $p_{T,j}$ their transverse momentum, and R the expected angular distance for the object. In CMS, during Run-2 two types of jets were defined : *AK4 jets* with $R = 0.4$ for jets emerging from single partons, and *AK8 jets* with $R = 0.8$ for boosted topologies where two very collimated partons form a single large cluster, for example when a high-p_{T} heavy resonance decays into two quarks. In each iteration, the two particles with minimal distance are combined, usually by adding their 4-vectors, unless the minimal distance is the one between a particle and the beam d_{iB}, in which case a jet is declared and removed from the set of particles. The specific definition of the anti-k_T algorithm is such that collinear radiations that have a small distance are aggregated first, while soft radiations with large distances are combined last, and the final jets have a real cone-like structure that facilitates experimental treatment of the corrections. Several improvements can be brought in terms of limiting the number of

distances to be computed at each iteration, reducing the overall computing time, for example the ones implemented in the FastJet [191] package. Additional selections can be applied on the jets to provide several quality criteria, guaranteeing within some efficiency that the jets originate from prompt partons.

Jet substructure plays an important role in the search for heavy BSM particles decaying into SM ones, the large mass difference producing a significant boost of the latter from which further decays could lead to several very collimated jets with merged clusters, reconstructed as AK8 jets. To separate the hard parts of the jet from the soft contamination and the pileup, grooming techniques [192–194] were developed without being able to remove all the soft contributions. Several algorithms were proposed to solve this issue, one of which is the Soft drop (SD) method [195] which consists in *declustering* the AK8 then re-applying a clustering procedure. The clustering history is then scanned backwards, starting from the last merging of two jets and checking the condition

$$\frac{min(p_{Ti}, p_{Tj})}{p_{Ti} + p_{Tj}} > z_{cut} \left(\frac{\Delta R_{ij}}{R_0} \right)^{\beta}, \tag{2.9}$$

where β and z_{cut} control the removal of soft radiation. If the condition is met, the two jets are considered as *subjets* of the AK8, otherwise the softest of the two is removed and the condition is evaluated again on the previous clustering step of the hardest jet, repeated until the condition is met, or the singleton jet is reached. In CMS the value of β has been set to zero, thus no need to fix R_0, while the z_{cut} value is set to 0.1 as in Ref. [195].

Hadronic taus

The tau lepton can decay into either a lepton – muon or electron – with associated neutrino, or a pair of quarks, in both cases with a tau neutrino. While the leptonic decay amounts to around 35.2 % [6] of the decays, it is hardly distinguishable from a prompt lepton. The other hadronic decay mode is more probable and the resulting hadrons produce a jet that can however be distinguished from the quark and gluons jets by the multiplicity, collimation, fraction of electromagnetic jet constituents, and isolation of the jet and the associated tracks [188].

The Hadron-plus-strips (HPS) [196, 197] algorithm is tasked to reconstruct the hadronic taus τ_h, starting from the reconstructed hadronic jets, with the anti-k_T algorithm and distance $\Delta R = 0.4$ as described in Section 2.4.2. All the objects within a cone of $\Delta R = 0.5$ around the jet axes are considered as seed regions. Then π_0 candidates are reconstructed using "strips" in the (η, ϕ) space, adding the electrons and photons momenta, and charged hadrons $h^{\pm}$ are reconstructed using charged particles from the PF. The hadronic taus decay can follow several chains containing one or several π_0 and/or $h^{\pm}$. In the algorithm, seven of those are defined and each scenario is tested to find the most suitable decay candidate matching the

scenario with additional constraints. The resonance mass should loosely match one of the mesons to which the decay chain is attached, the total charge should be $\pm$ 1 unless of missing hadrons mode in which case only the highest-p_{T} hadrons are considered, and all the reconstructed product of the tau hadronic decay should be within a *signal cone* of radius $\Delta R = 3.0/p_{\mathrm{T}}^{\tau_h}$ GeV, limited to the values between 0.05 and 0.1, such that the boost of the tau is taken into account. Among all τ_h candidates, only the highest-p_{T} one is kept, and its momentum is set from the sum of all its constituent.

After reconstruction with the HPS algorithm, the hadronic tau candidates are contaminated by several backgrounds : collimated jets, muons, and electrons that can mimic a single $\pi^{\pm}$, the electrons can even mimic an additional π^0 decay into photons through bremsstrahlung radiation. During Run-2, the task of identifying genuine hadronic taus is performed by a multiclassification MVA algorithm called DeepTau [198]. It takes as inputs high-level variables from the properties of the tau candidate, such as isolation or PV compatibility, or potential SVs, as well as the spatial information from all the subdetectors arranged in a two-dimensional distribution in the (η, ϕ) plane. The latter is built from two grids, a fine 11×11 grid of cells with size $\eta \times \phi = 0.02 \times 0.02$ to cover the "signal cone" with $\Delta\mathrm{R} < 0.1$, and a larger 21×21 grid of cells with size $\eta \times \phi = 0.05 \times 0.05$ to cover the "isolation cone" with $\Delta\mathrm{R} < 0.5$. The subdetectors are split into three sets of grids ($\mathrm{e}^{\pm}/\gamma$, $\mu^{\pm}$ and hadrons), that are processed and combined within a complex Deep Neural Network (DNN) architecture, incorporating further the high-level inputs. The final classification probabilities are then used in the definition of a final discriminator.

Missing transverse energy

Neutrinos are sometimes produced in the final state of a hard process of interest, yet they practically never interact within the detector and their direct measurement is impossible. There is nonetheless a way to indirectly infer some approximate kinematic properties through energy conservation. Since the colliding protons have a mostly longitudinal momentum – the collision angle effect in the transverse plan is close to negligible – energy conservation dictates that the products of the collision should have a total momentum that is cancelled over the transverse plane. Although drawing conclusions on the longitudinal momentum is hard given the non hermicity close to the beam line, despite the HF, and the fact that the energy fraction carried out by each of the partons is unknown, the summed transverse momentum of all the particle in the final states should be close to zero if it were not for the neutrinos. Therefore, only the neutrino total transverse momentum $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$– also referred to as Missing Transverse Energy (MET) $E_{\mathrm{T}}^{\mathrm{miss}}$– can be recovered as

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = -\sum_i \vec{p}_{T,i}, \tag{2.10}$$

where i runs over all the PF candidates.

It would however be naive to use the $E_{\mathrm{T}}^{\mathrm{miss}}$ as a direct estimation of the neutrino transverse momentum from the hard process. Firstly because its assumptions are based on the fact that all particles other than the neutrinos are detected, while in reality some can escape outside the acceptance region and through reconstruction inefficiencies. Secondly, the definition of the $E_{\mathrm{T}}^{\mathrm{miss}}$ depends on the measurement of all the other PF candidates, some of which can be subject to mismeasurements, especially jets due to the clustering algorithm and the limited calorimeter resolution, or from more general instrumental effects, calorimeter noise or parasitic interaction between protons and the residual gas in the beam pipe, called *beam halo*. Lastly, neutrinos can be produced from hadron weak decays and would not necessarily come from the hard process, although they should be relatively soft, and $E_{\mathrm{T}}^{\mathrm{miss}}$ does not allow to decouple the measurement from several neutrinos, that could even cancel on the transverse plane in particular configurations.

B-tagging

Bottom (anti)quarks are an essential part of several physics processes of interest and constitute very specific signatures. For example, they are present in the almost only decay mode of the top quark, and are therefore always present in $\mathrm{t\bar{t}}$ processes, but also constitute the main decay channel of the Higgs boson. The identification through the *tagging* of the jet as originating from a b quark – referred to as *b-tagging* – is therefore crucial. Luckily, the hadronisation of these quarks yields B mesons such as B^0 and $\mathrm{B}^{\pm}$ whole unusually large lifetime of about 15 ps, when boosted by a large momentum, produces a displaced vertex – also called Secondary Vertex (SV) – up to one centimetre from the PV, as illustrated on Fig. 2.13. Thanks to the very good position resolution from the tracker, this SV can be reconstructed and represents a quite clear signature of a jet originating from a b quark, commonly called a *b-jet*. Event with the absence of a displaced vertex, the b-jet will have different features compared to jets originating from gluons or light quarks, and these differences can be leveraged. While this section is only dedicated to the b-tagging, similar methods to some extent can be applied for tagging of charm quarks.

Several methods and tagging algorithms exist, but they all start from the same basis of reconstructed tracks with following selection criteria [199].

- Tracks must have a $p_{\mathrm{T}} > 1$ GeV, fit $\chi^2 < 5$ and at least one pixel hit. This is to ensure good momentum and Impact parameter (IP) resolutions.
- The displaced vertices from long-lived K^0 or Λ hadrons (with larger lifetimes than B mesons or material interactions) are suppressed by requiring that their distance from the PV is below 5 cm.
- To reduce contamination from pileup tracks, the transverse (longitudinal) IP of the track must be smaller than 0.2 (17) cm, additionally the distance

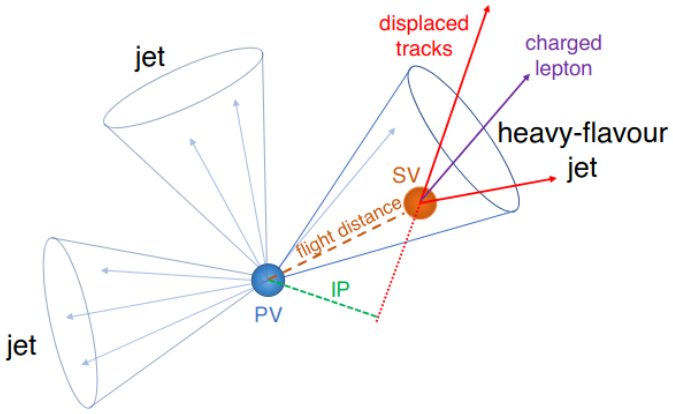


Figure 2.13 | Example of a hard process leading to several jets including a b-jet decaying within some distance from the Primary Vertex (PV), illustrating the displaced vertex and tracks with sizeable Impact parameter (IP). From Ref. [199].

between the track and the jet axis at their PCA must be less than 0.7 mm.

From these tracks, several features can be used to distinguish b-jets from light flavour jets, independently of the presence of a reconstructed displaced vertex. The IP plays a major role, its sign is defined from the angle between the IP vector and the jet axis. If the jet is produced "upstream", the angle will be below $\pi/2$ and defined as a positive IP, otherwise it will be negative. Light flavour jets will have a symmetric distribution of the IP sign, while b-jets will tend towards the positive only. An associated variable is called the impact parameter significance (IPS) and is defined by the ratio of the IP and its uncertainty. Heavy flavour jets like b-jets tend to have larger masses and undergo a harder fragmentation, therefore their decay products will have on average larger p_{T} relative to the jet axis. Relatively often, these heavy flavour jets can also produce a muon or electron within the decay chain, for b-jets this happens about 20% of the time, and these leptons can be used as an additional signature.

When the displaced vertex can be reconstructed, it represents a striking signature from which several variables can be used in the b-tagging algorithms. This includes the SV mass, corrected from the PV by the flight direction and momentum of the tracks originating from the SV, or the 2D or 3D distance between the two vertices. Failure to reconstruct the displaced vertex can come from a too short flight distance that is below the resolution achievable by the tracker, or that fewer than two tracks passing selection criteria were identified, in which case there is not enough information to pinpoint the location of a SV.

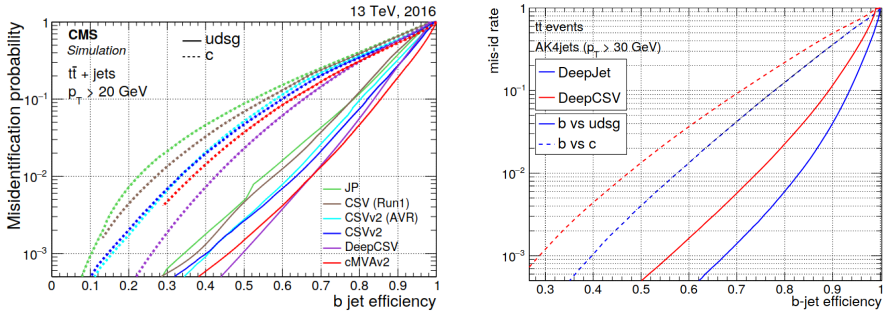
Two algorithms exist for the task of identifying a displaced vertex. The Adaptive Vertex Reconstruction (AVR) algorithm was used during Run-1, it uses the tracks

with the above criteria, and iteratively removes tracks that originate from the PV based on several criteria and using an adaptive vertex fitter. Reduction of material interaction or long-lived hadrons is performed through cuts on the SV mass and tracks that cannot originate from a resonance whose mass is close to these hadrons. Finally, the remaining tracks are used to build the SV that is associated with the jet. For Run-2 this algorithm was supplanted by the Inclusive Vertex Finder (IVF) algorithm, that uses the same tracks with looser requirements. Several tracks matching some IP requirements are used as seeds, and the same adaptive method as the AVR is used to clean tracks of the PV. Arbitration is needed when a track is associated to several vertices, in which case a compatibility criterion is used, before a final refitting of the SV is performed.

Once the tracks and potential SV are determined, several algorithms have been developed to exploit this information for b-jet identification. Some of these algorithms used in Run-1 and Run-2 are described in the following.

- Jet probability taggers : the signed IPS is used to compute a likelihood.
  Examples : Jet Probability (JP), Jet B Probability (JBP)
- Lepton taggers : lepton taggers for the special cases where a lepton emerges from the jet (but only applicable to these cases).
  Examples : Soft Electron Tagger (SET), Soft Muon Tagger (SMT)
- Global multi-variate analysis (MVA) discriminants : DNN algorithms trained at classification between the different jets flavours from a set of input features defined on the jets and possibly the SV, applicable to all events (the norm during Run-2).
  Examples :
  - Combined Secondary Vertex (CSVv2) : Multilayer perceptron (MLP) with separate training with and without SV, this second version is specifically trained with vertexing from IVF, which is the algorithm used throughout Run-2.
  - Deep Combined Secondary Vertex (DeepCSV) : improvement of the CSVv2 with deeper MLP, category inclusive training and more low-level variables.
  - Combined Multivariate Algorithm (cMVAv2) : BDT combining the output of the JP, JBP, SMT, SET, and the two vertex reconstruction variants of the CSVv2 algorithms (Fig. 2.14a).
  - DeepJet [200] (or DeepFlavour in tagging) : more complex architecture and using approximately 650 input variables from different sources (Fig. 2.14b).

In this document both DeepCSV and the DeepJet algorithm are used, for AK8 subjets and AK4 jets respectively.



(a) *(b)*

Figure 2.14 | Performance of the different taggers in CMS, the x axis represents the efficiency at selecting a true b-jet, while the y-axis gives the misidentification probability when a light-jet or a c-jet is selected as a b-jet. Each curve represents a tagger, the points constituting the curve are the varying thresholds applied on the tagger score. Left is from Ref. [199], right is from Ref. [200].

Pileup mitigation

Multiple interactions can occur within a single bunch crossing, adding to the hard process many inelastic proton scatterings. This pileup can reach up to 60 interactions in the tails of the distribution as illustrated on Fig. 2.3b, introducing low-p_{T} particles to the event. In addition, Out-of-time (OOT) pileup can occur when particles travel through the detector in a longer time period than the bunch crossing interval of 25 ns, also considering the delay introduced by detector signal processing, causing the particle to register hits or calorimeter deposits in the next events.

These pileup contributions have an approximate uniform distribution over the detector and cause several losses of performance. The energy deposits from all the soft contributions impacts the energy resolution of the jets, consequently the precision of the missing transverse energy is impaired. Additional charged particles in the tracker increases the combinatorial background, decreasing the track finding efficiency and increasing the fake rate, other track-related computations can suffer from this, for example vertexing, b-tagging or jet substructure. Large track multiplicity can reduce lepton isolation by random occurrences of collinear tracks, as well as complexifying the search for hadronic decays of the tau lepton.

In the CMS experiment, two methods are dedicated to mitigating this effect for PF jets. In this work, Charged-Hadron Subtraction (CHS) [201] was used for AK4 jets and Pile Up Per Particle Identification (PUPPI) [173, 202] for AK8 jets. The CHS is solely based on the identification of the different vertices in the event using tracking information. Tracks used in the reconstruction of a pileup vertex are removed, the remaining tracks are kept for jet clustering. This method is limited to charged

particles in the tracker acceptance region, and even though a bias has been observed at pseudorapidity values close to 2.5, no effect was observed on its resolution. A crude way to correct the p_{T} of neutral hadrons, photons, and hadrons outside of acceptance – objects for which a track cannot allow for vertex identification – is to subtract the uniform contribution of the pileup from each object. This density ρ can be estimated and is of the order of 1 GeV per pileup interaction and per unit area in the (η, ϕ) plane [188], and the subtracted p_{T} quantity depends on the *effective area* A_{eff} of the object, for jets it can be their *catchment area* [203, 204].

PUPPI also works on the particle level before clustering is performed, also taking into account tracking information but with the addition of the local particle distribution and the pileup properties of the event. For each neutral particle, i a shape is computed with all the other particles j as such

$$\alpha_i = log \sum_{j \in \text{event} \neq i, \Delta R_{ij} \leq R_0} \frac{p_{T,j}}{\Delta R_{ij}}, \tag{2.11}$$

where R_0 is set to 0.4, a compromise between isolation and jet reconstruction performance requiring large and small cones respectively. The index j runs over all charged particles within tracker acceptance $|\eta| < 2.5$, and all charged and neutral particles outside of it, in case no particle is located within R_0 the value is set to $\alpha_i = 0$. The shape α is then compared to the charged pileup distribution, using its median and RMS to provide a weight in the range [0,1] from an approximate signed χ^2 distribution cumulative function. Outside tracker acceptance, a transfer factor is used to obtain the distribution parameters. The larger the weight, the more likely the track is coming from the PV, for two main reasons. First because particles emerging from the hard process are more likely to be with a large p_{T}. Second, thanks to the collinearity of the parton shower, they tend to be relatively close to other particles within the considered cone, while pileup particles usually do not come from parton showers and only correlate by chance. For charged particles where the vertexing information is available, the weight is set to 1 for particles coming from a PV as well as charged particles not attached to any vertex and for which the PCA has a longitudinal distance compared to the PV of $d_z < 0.3$ cm, and 0 otherwise.

This weight is used to rescale the 4-momenta of all particles in the jet clustering algorithm after a cut on the weights at 0.01 is applied, effectively removing jets with a 99 % probability to be originating from pileup. An additional cut on neutral particles is applied based on their p_{T} and number of vertices to reduce the residual dependence of jet energies on the number of interactions. PUPPI has the advantage of being applicable outside the tracker acceptance region and to neutral particles, and also has better performance than CHS at low p_{T} and large pileup, both methods converging at high p_{T} where pileup tracks have negligible contribution.

The MET is not expected to be significantly affected by pileup particles since they practically do not bring any $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, though its resolution can be impacted due to the uniform energy density over the detector coverage. With the Charged-Hadron Subtraction algorithm a correction called *Type-0* is applied assuming that neutral and charged total transverse energy sum are equal and that the latter is exactly measured, it is therefore computed based on the total charged energy sum of the pileup particles. This correction however introduces a bias and in this work the direct uncorrected MET is used.

Lepton isolation is defined as the sum of transverse momentum of all the tracks within a cone around the lepton. Since pileup can enter this cone by random occurrences, this effect also has to be corrected. While charged particle tracks are supposedly already corrected, only the neutral contribution of the pileup needs to be subtracted. The correction is computed as

$$\mathrm{Iso}_{PF}^{\mathrm{abs}} = \sum_{i \in h^+, PV} p_T^i \; + \; \max\left[0, \sum_{i \in h^0} p_T^i + \sum_{i \in \gamma} p_T^i - \Delta\beta \sum_{i \in h^+, PU} p_{\mathrm{T}}^i\right], \quad (2.12)$$

where h^+ corresponds to pileup corrected charged PF candidates, while h^0 and γ to neutral and photon candidates. The last term corresponds to the pileup correction from neutral particles, whose total transverse energy is assumed to be the same as the charged one, similarly to the MET, and the $\Delta\beta \simeq 0.5$ comes from the observation that charged and neutral contributions to the pileup represent $2/3$ and $1/3$ of the total transverse energy respectively. Alternatively, this term can be replaced by ρA_{eff}. Relative isolation is computed as $\mathrm{Iso}_{PF}^{\mathrm{rel}} = \mathrm{Iso}_{PF}^{\mathrm{abs}}/p_{\mathrm{T}}^{l}$.

2.4.3 Analysis methods

Once the data has been stored and the different physic objects are reconstructed, the search for a specific physics process or a range of processes can begin. In most analyses performed by CMS the procedure follows the same scheme, a specific model, or more generally a specific signature, is assumed and used to determine a set of selections that define a *signal region (SR)*. These selections are determined and optimised based on simulations designed to be as similar as data as possible. This requires a precise theoretical modelling of the hard interactions in the event generation, both the SM contributions that are expected to contribute to the SR and the description of the expected signal which can extend to BSM scenarios, already covered in Section 1.3. It also requires precise modelling of the detector response.

Despite the high complexity of experimental event simulation, there are limitations to the degree of precision with which it can reproduce the different processes appearing in the data of the SR, either because of the limited amount of events

that can be generated – event generation with full detector simulation can be quite computationally intensive – or because of the different approximations used through the simulation from both the theoretical and experimental sides. For example, mismodelling of the parton shower and the hadronisation or imprecisions in the material budget cannot be avoided. Simulations are also carried out in parallel to the data-taking, and some assumptions have to be made about the status of the detector and the accelerator that may not fully reproduce the reality of the conditions (alignment, calibrations, ageing of the detector, etc), especially given the variations of these conditions. It is therefore necessary to employ various corrections on the different objects such as jet and lepton energy scales, assign Scale factors (SFs) on the different selections employed to correct for potential biases compared to data, and scale the different simulated processes to the data by use of the luminosity, cross-sections and associated efficiencies, all of which require careful considerations and an estimation of the associated uncertainties. Another parallel avenue is the estimation of some processes through data in control regions (CRs), either used because of the low accuracy of these processes simulation or their complete absence, for example charge mismeasurement of leptons or jets faking leptons. These methods are known as *data-driven*.

Data enhancement can be used to improve further the sensitivity of an analysis, for example using multi-variate analysis methods. The data then needs to be interpreted by means of statistical tools to allow for a quantitative estimation of the process in consideration. This section is intended at describing each of these steps in general.

Luminosity measurement

The principles behind the measurement of the luminosity through luminometers and Van der Meer (VdM) scans has been described in Section 2.2. This section will provide the experimental side of this measurement, particularised to CMS.

The VdM scans are performed once a year, with lower luminosity and about 50 times fewer bunch crossings. This is to allow the beams to be variably separated and the beam shape measured with great accuracy thanks to the lower hit rate. The measured parameters come with systematic uncertainties that will impact the instantaneous luminosity for the rest of the year.

During physics data-taking at nominal luminosity several parts of the detectors are used to provide an instantaneous luminosity measurement compared to the value measured during VdM scans. The subdetectors need to pass a few requirements : they must be able to perform measurement bunch-by-bunch for the calibration and to do so at high rates, their response must be linear with the instantaneous luminosity so that the value measured during calibration can be extrapolated during data-taking, and have minimal Out-of-time (OOT) effect. In the last item, one differentiates *Type-1* OOT effect due to a slow response or dead time of the

detector which causes a different response between the first bunch and the following ones, and *Type-2* that consists in particles coming from a previous bunch crossing, slow particles or short-term material activation for example, which can however be parameterised.

The detectors fulfilling these requirements and used as luminometers are given below, based on [205]. The luminosity measurement is based on Eq. (2.3), the rate R being the measured quantity. Some detectors are triggered on some quantity that flags when a collision occurred within a bunch crossing, allowing for a value of R, while other systems can have multiple particles hitting the system at any time, necessitating a so-called *zero-counting* algorithm to correct for pileup effect. Once R is measured, the cross-section of a hit in the detector or in other words of the inelastic scattering of protons $\sigma_{vis} = 71.3 \pm 3.5$ mb [206] can be used to obtain the instantaneous luminosity.

- *Hadron foward calorimeter zero counting (HFOC)* : The HF can be used with a hit counting method in the different calorimeter cells.
- *Hadron foward calorimeter transverse energy (HFET)* : A measurement of energy deposit in the Forward hadronic calorimeter (HF) is a trigger for a collision happening within a bunch crossing, the measurement is made within the same detector as the HFOC and would suffer from the same systematics, but their correlations can help improve the measurement.
- *Fast Beam Condition Monitor (BCM1F)* : The BCM1F [207] is a sensor located around the beam pipe on both sides of the pixel tracker. It was made of single-crystalline diamond sensors in 2016, replaced in 2017 by poly-crystalline diamond and silicon sensors with a fast response time to avoid *Type-1* OOT effects.
- *Pixel Luminosity Telescope (PLT)* : The 16 telescopes made up of 3 layers of pixel tracker can serve as a very good measurement due to their low occupancy and therefore good linearity. Their location very close to the IP ensures virtually no *Type-2* OOT effects and that any collision would be recorded with at least one hit, although zero-counting must be applied.
- *Pixel cluster counting (PCC)* : the entire pixel tracker can be used by counting the number of pixel clusters, which needs prompt reconstruction.
- *Cross-measurements* : Other independent parts of the detectors can be used to evaluate long-term stability and linearity of the other subdetector measurements, for example the DT of the muon chambers or the cavern radiation monitoring system called RAMSES [208].

Each of these subdetectors takes part in the overall luminosity measurement, some systematic uncertainties are shared and some are exclusive to certain systems, potential non-linearity must also be parameterized and its effect added to the list. The measured luminosities and total uncertainties for Run-2 are 36.3 $\text{fb}^{-1} \pm$ 1.2 %

in 2016 [209], 41.5 $fb^{-1}\pm$ 2.3 % in 2017 [210], and 59.7 $fb^{-1}\pm$ 2.5 % in 2018 [211], although some effects are correlated between data-taking years and this must be taken into account when performing a fit over the entire Run-2 dataset.

Detector simulation

Converting final state particles after parton shower described in Section 1.3 into a complete detector response in the CMS detector is a task handled by a simulation software based on GEANT4 [212], referred to as *full simulation*. This method is however very CPU intensive, and may be challenging for heavily suppressed backgrounds with large cross-sections for which many events need to be generated but few can pass selections. Although not used in this present document, it is important to note that alternatives exist, such as a simpler and faster version of the detector simulation, referred to as *fast simulation*, or the use of DELPHES [213] where several parametrisations are employed to simplify and accelerate the computations even further, commonly known as *parametric simulation*.

The full simulation uses the knowledge of not only the sensitive material, but also all the supports, cooling pipes, electronics and any other additional material. Particles are propagated from the IP through the detector via Monte Carlo techniques, simulating energy deposits, interactions with the material, bending in the magnetic field, potential decays, and finally the simulation of the electronic response they produce in each detector readout electronic channel. These simulated hits, known as *simhits* are then fed to the usual reconstruction algorithms, except that compared to data they also include simulation history.

Pileup needs to be included in the simulation. This is done through the prior generation of *minimum bias* events – inelastic proton scatterings – by PYTHIA and GEANT4. These collections of events are then sampled randomly following a Poisson distribution based on the expected mean number of interactions and added to the hard scattering collision in the same bunch crossing but also emulating Out-of-time effects. To better model the pileup distribution measured on Fig. 2.3b a reweighting based on the actual average number of interactions needs to be performed. The uncertainty of the pileup modelling from its estimated cross-section of 69.2 mb is used as a systematic uncertainty template in the statistical analysis.

Lepton efficiency

Lepton efficiencies are computed via a Tag-and-Probe (T&P) method [185, 186], which can be applied on any object produced in pairs. For leptons, $Z \to ee$ / $\mu\mu$ events are used, but the following is not restricted to them. A *tag* object is required to pass tight criteria $\mathcal{S}$, so to obtain high purity events, while the associated object in the pair called the *probe* is required to pass looser selection cuts $\mathcal{B}$. The probe

serves as a baseline on which the efficiency of the criterion $\mathcal{S}$ can be evaluated as $\epsilon_S^{data/MC} = N_\mathcal{S}/N_\mathcal{B}$, where $N_{\mathcal{S}(\mathcal{B})}$ represents the number of events for which the probe passed the tight $\mathcal{S}$ (looser $\mathcal{B}$ only) selection, for both data and simulations. While $N_{\mathcal{S}/\mathcal{B}}$ can be directly obtained from simulation, their derivation in data requires the subtraction from contaminations. This is achieved by fitting the invariant mass distribution and the numbers extracted from the Z peak, as illustrated on Fig. 2.15. An associated SF is computed as $\text{SF}_\mathcal{S} = \epsilon_S^{data}/\epsilon_S^{MC}$, in different p_T and η regions to better represent the kinematic dependence of the identification efficiency. Combinatorial effects, especially at low p_T where background contamination is large, can increase the statistical uncertainties. This can be mitigated by using only events where a single probe is associated to a tag. Specific care must be taken to make sure the HLT paths do not bias the T&P selections, which sometimes requires trigger matching.

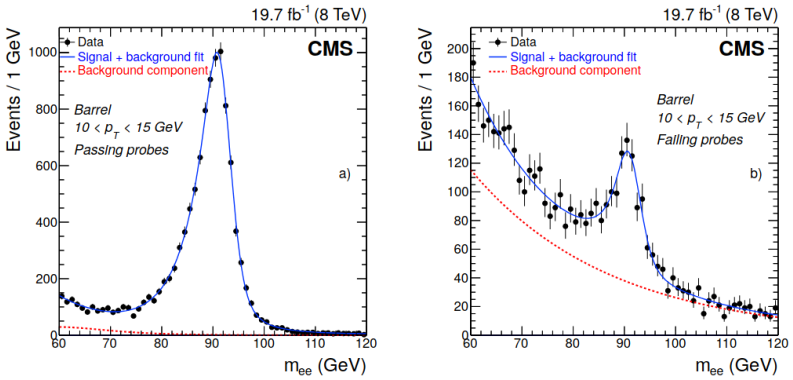


Figure 2.15 | Invariant mass of Z $\rightarrow$ ee system for passing (left) and failing (right) probes, superimposed with the fit of the signal Z peak with a monotonically falling background contamination. From Ref. [186]

Lepton energy scale and resolution

Reconstructing lepton energy from the ECAL SC can suffer from inefficiencies due to leakages outside the acceptance, the crystal gaps or between barrel and endcap, and the energy that is missed and intercepted in the HCAL. Additionally, energy losses can appear due to prior material interaction, for example in the tracker, or due to pileup interactions [186]. A MVA is trained using SC information, taking into account pileup through its estimated density, and with the target as the ratio of true electron energy divided by the energy measure from the SC. This regression is then used to correct the bias and improve resolution when applied to both data and simulations. While this correction accounts for most of the resolution disagreement with data, a remaining correction needs to be applied to account for imperfect

description of the tracker material interactions or the ECAL geometry, and the evolution of the transparency and noise within the ECAL crystals. This is performed by comparing the invariant mass of the $Z \to ee$ events in both data and simulations, in different regions and categories, but also in different run periods to account for temporal effects. The shift in the Z peak is used as SF, and a smearing with a Gaussian fluctuation is added to simulation to match the resolution observed in data, based on the same regions.

Muon energy is purely measured from tracks, which measurement can be affected by any tracker misalignment [185]. Low- or intermediate-p_{T} muons are mostly measured with the inner tracker, as it dominates in terms of performance. A similar method as for the electrons is used to estimate the corrections using $Z \to \mu\mu$ events in both data and simulations, although other resonant decays can be used. Additionally, cosmic muons can be measured in each half of the detector and their momentum measurement compared. High-p_{T} muons can suffer from a close to straight curve that makes both the transverse momentum and charge q measurement more challenging. An *endpoint method* is used on both cosmic muons – unusable for endcaps though as they are mostly vertical – and muons emerging from collisions to perform measurements of q/p_{T} and determine a bias shift.

Jet energy scale and resolution

Mismeasurement of the jet energy and transverse momentum are unavoidable due to the non-linearity of the calorimeter response and to the complex treatment from the PF algorithm and the event generation, in fragmentation modelling for example. Jet energy needs to be properly calibrated and the simulation should match the spectrum observed in data [214]. The simulated jets must go through the same procedure as in data, namely their reconstruction via calorimeter clustering with the anti-k_T algorithm as described in Section 2.4.2. The convention followed by CMS is to exclude neutrinos emerging from the jet as they would not be measured in the detector, and although the corrections are computed in samples with no major neutrino contribution, so they do not really impact the measurement, the energy fraction they can represent in heavy flavour jets should lead to an additional systematic for measurements sensitive to it. The Jet Energy Correction (JEC) are performed in four steps, described in the following. On the simulation side the samples used are multijets, Z+jets, γ+jet and single-neutrino events to account for pure pileup events, while the selections in data are targeted at the same processes with several criteria, *zero-bias* events that are not triggered but kept with large prescaling are selected to account for detector noise and pileup.

The first step consists in pileup removal, similarly to the description of Section 2.4.2 the diffuse offset energy density ρ can be multiplied by the jet area and subtracted from the jet energy sum. During Run-2, ρ was computed using a random cone method where many jets are reconstructed in each zero-bias events in randomly

placed cones to map the whole (η,ϕ) plane prior to the clustering algorithm. ρ also evolves rather linearly with the pileup, that should therefore be generated accordingly to data, whereas OOT pileup effect is a small contribution. The values of the associated SF are on Fig. 2.16a.

The response corrections are then applied on both the data and simulation, they are however only computed from the latter to ensure no bias from data-based methods, as corners of the phase-space might not be well populated. Generator level jets are searched within half the radius used by the clustering algorithm around the jet axis, and the response is parameterised as the ratio between measured and true transverse momentum, averaged over the bins of true p_{T} and η. This response correction is applied on both jets from data and simulations.

While the bulk of the jet corrections were performed in the previous step, residuals can remain between the energy scales of data and simulations. These are measured from data with two complementary methods. The p_{T}-balance method uses a reference object to estimate the Jet Energy Scale (JES) – similar to the T&P method – while the Missing transverse momentum Projection Fraction (MPF) uses the whole hadronic activity in the event.

Finally, flavour corrections can be applied to take into account the differences in jet fragmentation energy and composition. Acceptance issues and neutral hadron contribution can have sizeable effects on the response, although they both become insignificant at high-p_{T}. In simulation, the flavour is estimated from the matched generator level jet, and used together with the p_{T} and η as binning for the flavour response corrections.

Additionally, simulation tend to be too much optimistic on the Jet Energy Resolution (JER), and a smearing of the jet p_{T} is performed based on the ratio of resolutions between data and simulations s_{JER} as a function of η. When the reconstructed jet is matched to a generator-level jet within the half clustering radius, the correction factor to the jet p_{T} is computed as

$$c_{\mathrm{JER}} = \mathrm{max}\left(0, 1 + (s_{JER} - 1)\frac{p_{\mathrm{T}} - p_{\mathrm{T}}^{true}}{p_{\mathrm{T}}}\right). \tag{2.13}$$

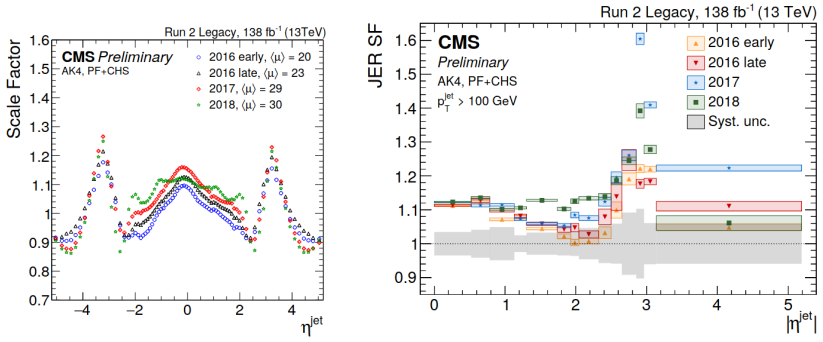
When no matching can be performed, a Gaussian smearing is performed as

$$c_{\mathrm{JER}} = \mathrm{max}\left(0, 1 + \mathcal{N}(0, \sigma_{JER}\sqrt{max(0, s_{JER}^2 - 1)}\right), \tag{2.14}$$

where the relative p_{T} resolution σ_{JER} is used as the Gaussian standard deviation with zero mean to sample a random number. The associated SF is on Fig. 2.16b.

Each uncertainty on the measurement of the JEC produces a variation in both direction of the correction by 1 sigma, propagated to the jets as two alternate jet collections. In total, there are 21 sources of uncertainties for the JEC, plus the one coming from the JER. While the uncertainty sources are uncorrelated,

their variations within one source are not, which complexifies their treatment. In analyses relying heavily on jets, and therefore very sensitive to JEC, the whole list of uncertainties must be used, while analyses less affected by them can use a reduced set where a grouping per detector region is performed, so the amount of variations is reduced to 11. Furthermore, the corrections of the jets transverse momentum need to be propagated to the MET, called *MET Type-1 correction*.



(a) (b)

Figure 2.16 | Jet Energy Correction relative to the pileup removal (left) and Jet Energy Resolution (right) for the three different data-taking years of the Run-2 as a function of the jet pseudorapidity. From Ref. [215].

Jet tagging efficiency

Mismodelling between the distributions of the tagger input variables and by extension their score between data and simulation may occur because of global imprecisions during the event generation. Taggers are especially dependent on tracker alignment and the uncertainty in the track parameters, and any simulation inaccuracies can generate large discrepancies in the tagger score. A Scale factor (SF) can be designed to correct for this effect in each flavour category F with $\mathrm{SF}_F = \epsilon_F^{data}/\epsilon_F^{MC}$, where ϵ_F^{data} and ϵ_F^{MC} refer to the tagging efficiency in data and Monte Carlo simulations respectively, usually as functions of jet p_T and η. Their computation require different procedures, and the resulting SFs are provided centrally by the CMS collaboration [199].

While computing the tagging efficiency on simulations is rather straightforward (flagging jets containing B or D mesons, signatures of b-jets and c-jets, only requires to look into the event generation history), the tagging efficiency measurement in data is more complex and performed on several datasets selected online, often with prescaling. A mix of samples is used to provide a sufficient number of events for all kinds of jets. Multijet samples provide light flavour jets, except when accompanied

by a muon in which case it contains more heavy flavour hadrons. $t\bar{t}$ samples, for the b-flavour jets of the top quark decays, are selected in either the double or single lepton channels. In the latter case, due to the decay of the W boson, there is a higher proportion of c-flavour jets. Additional samples where a W boson is produced in association to a c-quark are included to enhance their proportions.

For light flavour jets, a negative-tag method [216] is performed based on the assumption of symmetry around the zero of the IP and signed SV flight distance, purely from resolution effect. Light flavour jets will dominate the region with negative values, which allows determining the misidentification rate of the tagger, after a correction by the simulations.

Several methods are employed in the case of the b-tagging efficiencies. In muon enriched samples, the muon is required to be within the jet and can be used to identify real b-jets in several ways. The SF is measured either through a fit or the b-tagging correlation information. Fully-leptonic $t\bar{t}$ can be easily selected with the requirement of two isolated muons and a few more selections to remove potential contaminations. From there a kinematic discriminant, with variables independent of the vertexing information used in the tagger, is trained to discriminate b-jets from other jets and a fit can then be performed to obtain the SF. This *Kin* method can be cross-checked by a two-tag counting method that gives a rougher but more robust estimate of the SF, where the efficiency is extracted from counting the number of events with two b-tagged jets. A T&P method can also be used on semileptonic $t\bar{t}$ events where the W mass hypothesis can be tested on the hadronic branch, and the two b-jets after jet-quark assignment are used for the tag and the probe tests, a fit is then performed before and after applying the tagger. Similarly to c-jets, a weighted average is used to combine the different measurements.

The c-tagging efficiencies are computed with two methods, one for each sample used to obtain c-flavour jets. In events where a c-quark is produced in association with a leptonically decaying W boson, the main production mode consists in the interaction of a s-quark and a gluon. In that configuration, the c-quark and the W can only be of opposite sign, determined from the soft muon from their decays. The main background contaminations are from W events with jets from gluon splitting, with equal parts of opposite-sign and same-sign signatures. The subtraction between the two categories provides an estimate of real c-jets, for which the efficiency of passing the tagger can then be evaluated. In the semileptonic category of the $t\bar{t}$ process, the leptonic branch with a muon allows for a selection, with additional constraints to ensure maximal selection of $t\bar{t}$ events. A jet-quark assignment is performed using a mass discriminant by forming pairs of jets as W candidates, and triplet of jets as t-quark candidates. The b-tagging is only performed afterwards to this step to avoid any bias, then the selected b-jets are removed from consideration. An energy difference between the different types of quarks produced in the W decay is used in a fit on a mass discriminant variable to extract the SF. Both measurements can

then be combined in a weighted average.

While so far the discussion was limited to the applications of the tagger score as a fixed WPs, they can also be applied in the statistical analysis, most often within a high-level variable or as input of a MVA. A SF must therefore be computed on the entire range of the tagger score, a method known as *IterativeFit* [217], which will then depend on the p_{T}, η and tagger score of the jet. For this, a T&P method is again performed on two sets of events, dileptonic $\mathrm{t\bar{t}}$ and Z+jets events. In the former, b-jets are expected and the b-tagging is used to select the tag jet, while in the latter, they are vetoed by the same algorithm to select light flavour jets. Then an iterative method is used to extract a SF for one flavour, using the subtraction from the other and its SF computed in the previous iteration, a process repeated until both SF values stabilize. The non-dominant contribution from c-jets is assumed with a SF set to unity.

Although the values of the SF are provided for all the CMS collaboration, the phase space of each analysis is different, and the efficiency of the tagger in simulations must be determined case-by-case based on the analysis backgrounds and signals in order to predict correct event yield compared to data. Several methods exist for WP-based selections, some manage to omit the efficiencies, but the most recommended one consists in applying a event-level weight based on all the jets on which the tagging algorithm is applied, whether they pass the selection or not, based on the simulation efficiencies ϵ_i and scale factor SF_i for each jet i :

$$\begin{aligned} w = \frac{P(\text{data})}{P(\text{MC})} &= \frac{\prod_{i \text{ tagged}} \epsilon_i \mathsf{SF}_i \prod_{j \text{ not tagged}} (1 - \mathsf{SF}_j \epsilon_j)}{\prod_{i \text{ tagged}} \epsilon_i \prod_{j \text{ not tagged}} (1 - \epsilon_j)} \\ &= \prod_{i \text{ tagged}} \mathsf{SF}_i \prod_{j \text{ not tagged}} \frac{1 - \mathsf{SF}_j \epsilon_j}{1 - \epsilon_j}. \end{aligned} \tag{2.15}$$

For the IterativeFit method, the expected yield must be preserved, which is not guaranteed given the difference in phase space of an analysis compared to the T&P region of the measurement. To ensure this, the event weight is compared before and after application of the tagger weight composed of the product of the SF for all jets on which the tagger is applied. The ratio of these weights

$$r = \sum w_{\text{before}} / \sum w_{\text{after}} \tag{2.16}$$

is applied as an additional event-level correction to the tagger weight, usually as a function of the jet multiplicity.

Related to the tagging of jets, the distinction between jets coming from the hard process and from the pileup can benefit from a dedicated selection criterion. In addition to the CHS in Section 2.4.2, an orthogonal jet selection technique as been developed as a *pileup jet id* criterion [218], a single discriminator combining both the vertex and shape information through a MVA. Its aim is to solve the problem of

overlapping jets that grows quadratically with the number of pileup interactions at low-p_{T}, producing jets with an overestimated p_{T}. Within the tracker acceptance the vertexing information can be used on tracks (from charged particles) and contributes as 4 variables to the classifying BDT, the rest of the 12 input variables consist in shape variables to target the diffuseness of a jet and are available even outside the tracker acceptance. The training is done separately for each region of the calorimeter and uses Z $+$jets simulated events, and the efficiency is compared with a data sample of Z $\rightarrow \mu\mu+$jets using a T&P method and the jet recoiling against the Z boson as a probe. This selection criterion and associated SF can then be applied on all soft jets, typically with $p_{\mathrm{T}} < 50$ GeV. Similarly to the tagging correction for a fixed WP, Eq. (2.15) can be used on the pileup jet identification method, using efficiencies and mistag rates provided globally by the CMS collaboration.

Analysis strategy

Based on the topology of the signal in question, basic selections can be first applied on the objects appearing in the expected final states. These pre-selection steps are used to determine the different backgrounds that will enter the analysis and cut down on the amount of simulations and processed data that will have to be considered, while at the same time keeping a high signal efficiency.

These pre-selections are however relatively basic and higher level selections need to be implemented to exploit the data, with the goal of reducing the background contamination in the SR while keeping the signal sensitivity as high as possible. To that end, reconstructed high-level variables such as combined invariant masses or angular variables can be determined from the knowledge of the signal kinematics obtained through its simulation. The optimization of the SR remains one of the main challenge of an analysis, and while the data in this region is not looked into to avoid any bias, several CRs can be used to assess the background modelling of the data. Once the agreement has been established and all the systematic uncertainties are well understood, the SR can be unblinded and the statistical analysis can take place.

While carefully crafted high-level variables are a valuable aid to improve the analysis sensitivity, a trend towards the use of MVA tools (for example Boosted Decision Trees (BDTs) or Deep Neural Networks (DNNs)) has been growing over the years. These methods are particularly good at classification tasks. At set of input variables, can be provided to the algorithm, which is then trained at discriminating between the various backgrounds and the expected signal on simulations. The score returned for each event can be used in the statistical analysis instead of the handcrafted variables and typically yields better performance, even when only defined on low level variables, which can save up some time during the exploratory step. This however comes with a cost of complexity, the backward inference from the score being tedious at best and often impossible. To ensure proper understanding of the

behaviour of the algorithm requires many indirect verifications before it can be validated for further steps. A more in depth discussion of these MVA techniques is given in Section 3.1.

Statistical analysis

The interpretation of the data involves the definition of a statistical model based on a likelihood ratio method [219, 220]. The parameter of interest (POI) is the *signal strength* μ. The following development is based on mutually exclusive regions, in this context the different bins of a histogram. We denote the observed count of events as d_i, the expected number of signal and backgrounds events s_i and b_i respectively, with i the index of the bins. While the signal expected count s_i can depend on several parameters – a mass or set of couplings – it will be assumed implicitly. Multiple POIs can be used inside the signal strength, which will be implicit as well.

Each process p entering s_i and b_i need to be scaled to data by

$$\mathcal{L}_{int}\ \sigma_p\ \epsilon_p = \mathcal{L}_{int}\ \sigma_p\ \frac{\sum_{j\in p} c_j w_j}{\sum_{j\in p} w_j}, \tag{2.17}$$

where $\mathcal{L}_{int}$ is the integrated luminosity, σ_p and ϵ_p the cross-section and efficiency of the process to pass the selections of the SR. In case the events have weights w_j, such as the ones coming from the unweighting procedure described in Section 1.3, the efficiency can be rewritten as in the second part of the equation where c_j are the event-level corrections described in the previous sections.

The likelihood can be expressed as a product of Poisson likelihood over each independent region [220] :

$$\begin{aligned}\mathcal{L}_{s+b}(\mu) &= \prod_i p\left(d_i|\mu s_i + b_i\right) \\ &= \prod_i \frac{e^{-(\mu s_i + b_i)}(\mu s_i + b_i)^{d_i}}{d_i!}.\end{aligned} \tag{2.18}$$

A statistical test for the presence of a signal in the defined analysis region consists in designing two hypotheses : the null H_0 hypothesis where $\mu = 0$ and the signal is absent, also called the b-only hypothesis, and the alternative $s+b$ or H_1 hypothesis where the data is described by both the background and the signal. The most discriminating test statistic is the likelihood ratio according to the Neyman-Pearson lemma

$$\lambda = \frac{\mathcal{L}(\mu = 1)}{\mathcal{L}(\mu = 0)} \equiv \frac{\mathcal{L}_{s+b}}{\mathcal{L}_b}. \tag{2.19}$$

From there, a p-value can be obtained by comparing toys extracted from the known

likelihoods and the value when using the actual data. This p-value can then be converted into a significance as a number standard deviation from the null hypothesis from the cumulative distribution of a one-sided Gaussian [221]

$$p = \int_Z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx. \tag{2.20}$$

In case no excess is expected, instead of quantifying that excess, one would want to produce an upper limit. It is however important to note that the decision to produce a confidence interval or an upper limit should be set a priori. Deciding on the type of result to quote based on the obtained results could result in over- or under-coverage, a phenomenon known as *flip-flopping* [222]. In addition, this discussion has so far ignored the different sources of systematic uncertainties that take into account imperfect knowledge of the theoretical predictions and experimental measurements. These are treated in the fit as *Nuisance parameters (NPs)* and denoted by θ.

While in Eq. (2.18) the signal strength is allowed to float freely, for NPs one may want to set priors to their values based on experimental constraints – also called ancillary measurements. Simple multiplicative factors can be used as a *log-normal* constraint, with the uncertainty of the NP is set as its width. More complicated sources of systematic uncertainty are object dependent and built as *shape* templates from the 1-sigma variations on both sides vertically around the *nominal* histogram [223]. Each shape NP results for each bin in an *up* and *down* variations with Gaussian priors.

Statistical uncertainties arise from the finite number of events in each bin. A procedure known as the Barlow-Beeston lite method [223, 224] consists in assigning for each bin a NP with variations relative to the statistical uncertainty. One caveat is that this can introduce many NPs. To mitigate that, in each bin, several processes statistical NPs are added in quadrature into a single NP, if their total yield uncertainty does not exceed half of the bin content.

Although some NPs can be measured together with the signal strength as parameters of interest, most often analyses are not sensitive to them enough. One can therefore include them rather as constraints into the likelihood of Eq. (2.18) [225]

$$\mathcal{L}_{s+b}(\mu, \theta) = \prod_i p\left(d_i | \mu s_i + b_i; \theta_n\right) \prod_n f\left(a_n | \theta_n\right), \tag{2.21}$$

where each NP with index n is associated a measured value of a_n.

Naturally, the simple likelihood ratio of Eq. (2.19) is not suitable any more and its extension to the introduction of the NPs is called a *profile likelihood ratio* [226, 227] :

$$\lambda(\mu) = \begin{cases} \frac{\mathcal{L}(\mu,\hat{\hat{\theta}})}{\mathcal{L}(\hat{\mu},\hat{\theta})} & \text{if } 0 \leq \hat{\mu} \leq \mu \\ \frac{\mathcal{L}(\mu,\hat{\hat{\theta}})}{\mathcal{L}(0,\hat{\hat{\theta}}(0))} & \text{if } \hat{\mu} < 0 \\ 1 & \text{if } \hat{\mu} > \mu \end{cases} \tag{2.22}$$

where $\hat{\mu}$ and $\hat{\theta}$ are the values of the NPs that maximize the likelihood, and $\hat{\hat{\theta}}$ are the values that maximize it for the specified μ. The first line is the usual profile likelihood ratio, the second represents the case when $\mu < 0$ which is unphysical and therefore the value is artificially set to $\mu = 0$ when $\hat{\mu} < 0$, and the last case is the region where $\hat{\mu} > \mu$ that is excluded as we usually like to set upper limits and reject large values of μ.

The profile likelihood ratio of Eq. (2.22) is useful to provide a test statistics that is independent of the NPs, which comes at the cost of a broader profile likelihood for any given μ than if the NPs were fixed, illustrating how important is the control of the associated systematic uncertainties for a precise measurement. From there a test statistic can be derived as

$$t_\mu = -2 \ \ln \lambda(\mu), \tag{2.23}$$

since $0 \leq \lambda(\mu) \leq 1$ by definition in Eq. (2.22). Assuming that the test statistic t_μ follows the Pdf $f(t_\mu|\mu)$ under assumption of signal strength μ, the p-values can be computed as

$$p_\mu = \int_{t_{\mu,obs}}^{\infty} f(t_\mu|\mu) \equiv CL_{s+b} \tag{2.24}$$

$$p_0 = \int_{t_{\mu,obs}}^{\infty} f(t_\mu|\mu = 0) \equiv CL_b, \tag{2.25}$$

where $t_{\mu,obs}$ is the test statistics obtained from data. Equation (2.24) is used in the case of an upper limit, while Eq. (2.25) can be used in the case of discovery, as discussed previously. The pdfs in Eqs. (2.24) and (2.25) can be determined by generating pseudo-experiments – called *toy Monte Carlo* – to produce the test statistic values from the likelihood Eq. (2.21) with d_i and a_n generated from the simulations. In these toys, a value for μ is assumed and the NPs are set to the maximum likelihood values while still allowed to float.

In the case of upper limits, they can be set on the signal strength μ_{up} at 95% confidence level in the frequentist interpretation, such that $CL_{s+b} \leq 0.05$. This might however cause issues in case no signal is present, yet a fluctuation of the expected events from the background-only hypothesis could result in a signal exclusion region. The CL_b on the other hand quantifies the confidence of a potential discovery and can be used to define the CL_s criterion [228, 229]

$$CL_s = \frac{CL_{s+b}}{CL_b}, \tag{2.26}$$

and the upper limit can be determined from all the values of μ such that $CL_s \leq 0.05$, usually referred to as *observed limits*.

While the analysis is blinded, one usually relies on the *expected sensitivity* to quantify the performance of the analysis. This is estimated by replacing $t_{\mu,obs}$ by the median upper limit $t_{\mu,med}$ under hypothesis that $\mu = 0$, and similarly to Eq. (2.25) determining the values of μ such that

$$\int_{t_{\mu,med}}^{\infty} f(t_\mu|\mu = 0) = 50\%. \tag{2.27}$$

Uncertainties on this expected upper limit can be represented by 1-sigma and 2-sigma bands, when the value in Eq. (2.27) is replaced by the bands containing 68% and 95% of the normal distribution content around the median respectively. This however requires large computation time, as enough toys need to be generated to have a good modelling of the Pdfs for each value of μ. This can be alleviated by the fact that the toys in the background-only hypothesis do not depend on μ, and the pseudo-data can be recycled [230, 231]. An alternative also resides in using the Wilks' [232] and Wald's [233] theorems to compute the expected limits using an analytical formula in the approximation of the large data samples.

In this document, the COMBINE [225] software was used to derive statistical interpretation, it is based on ROOFIT [234] and ROOSTATS [235], while the minimization of the $-\ln \mathcal{L}$ is performed by MINUIT [236]. For completeness, there exist alternatives to COMBINE that use the same libraries, such as HISTFACTORY [237] and its standalone implementation in Python PYHF [238, 239].

Chapter 3

Machine learning and the Matrix Element Method

A major limitation on the discovery of potential signals of new physics resides in the amount of data that can be recorded and analysed in offline analyses within an experiment such as CMS. Major efforts are being made on several sectors : improvement of the instantaneous luminosity from the LHC teams, fine-tuning of the trigger menus and overall reconstruction efficiencies of the different subdetectors of the experiments, and finally the application of more sensitive methods of analysis to extract the most information from the data and simulations. This chapter will be dedicated to the latter, focusing on machine learning (ML) techniques and the Matrix Element Method (MEM).

Machine learning (ML) methods – sometimes also referred to as Multi-variate analysis (MVA) methods in Section 2.3 – are a family of algorithms that learn to solve a non-analytical problem by learning patterns from examples and observation. Historically they were divided into two categories, supervised where the data is labelled and unsupervised when it is not. More recent trends have nonetheless showed that a continuum exists between these categories and new branches such as reinforcement learning or generative models have appeared, some will be discussed at the end of this chapter. While unsupervised learning has many applications, such as clustering or dimensionality reduction, they represent a minor part of all the ML methods used in High-energy physics (HEP). Supervised learning on the other hand has been used for decades in experiments such as CMS. Since they are trained on Monte Carlo simulations, the process and simulation history attached to each event can be used within the algorithm. Supervised learning consists mostly in two types of task : classification where the data is separated into several categories, inclusive or not, is widely used for signal versus background separation in offline analyses, while regression where a numerical value is produced from a set of input variables is useful for predictions or corrections of a known variable. Both cases have been illustrated in several occasions in Section 2.3, and have also been used in the work presented in this document. The discussion over the different algorithms

will be restricted to Boosted Decision Trees (BDTs) and Neural Networks (NNs), in Sections 3.1.2 and 3.1.3 respectively.

The Matrix Element Method (MEM) is a powerful tool to extract from a real or simulated event its compatibility with a theoretical hypothesis. As it benefits from both the knowledge of the physics process within the SM and the detector response through a parameterised Transfer Function (TF), this probability of compatibility possesses a high discriminating power that can be used as an input variable of the statistical analysis. Its dependence on the hypothetical process and potential associated parameters also allows its use as a likelihood, and its interpretation is relatively straightforward. The method will be discussed further in Section 3.2.

Both techniques however have several drawbacks. The MEM comes with both a cost in terms of complexity and computation time, as it is based on the integration of a non-trivial function over a large and multidimensional phase space. The ML algorithms require a training set to be built, and their accuracy usually depends on the size of this training set. Although in HEP it is often not a major inconvenient due to the sheer size of the simulations that have been processed, it is a factor to be kept under consideration. In addition, their performance come from a very high non-linearity that impairs the interpretation of their prediction, which is why they are often nicknamed as *black box*. In this section Section 3.3 will be discussed a hybrid method that combines both NNs and the MEM, in an attempt to cancel out some of their drawbacks. This method will be illustrated using an analysis that was carried out by CMS as a practical example. This work was produced by the author of this document and published in Ref. [1].

3.1 Machine learning

This section will focus solely on supervised learning, the most used category of ML algorithms in HEP by a large extent, and represents the only category that is present in this work. This section is mostly based on Haykin's book on neural networks [240], lecture notes from Michel Verleysen on machine learning [241] and a few other references [242–245].

3.1.1 Machine learning in a nutshell

ML models in supervised learning consist in a combination of three concepts : the data, the model, and the loss.

The data consists in pairs of m labelled data points

$$(x_1, y_1), ..., (x_m, y_m), \tag{3.1}$$

where x_i $(i = 1, ..., m)$ are vectors or numeric values called the *features* and y_i are the *labels* associated to each data point.

Features represent properties of a data point, containing low or high level information. There are obviously many ways on how the information about the data point can be expressed, and certain properties will prove more valuable than other, the challenging task of performing ML often resides on their definition and selection. These features are numeric real values, assembled as a vector of n-dimension residing in a feature space $\mathcal{X}$. While these features represent the input of the ML algorithm, the labels represent its output. Typically, they are defined with the help of human experts, and belong to the label space $\mathcal{Y}$ which can be one-dimensional or made up of label vectors. Their form can be categorical in the case of classification, or numerical values in the case of regression.

The model can be seen as a *hypothesis map* $h : \mathcal{X} \to \mathcal{Y}$. The best models are the ones that produce an estimation of the label $\hat{y}_i = h(x_i)$ that is as close as possible to the true label y_i, they are also often called *predictor maps*. The space of the hypothesis maps is practically endless, in practice it must be limited to a subset around one type of models, containing all the intrinsic parameters that govern its behaviour when presented with data, often referred to as *hyperparameters*. The space of all considered hypothesis maps is denoted the hypothesis space $\mathcal{H}$. The so-called *model selection* procedure consists in selecting one of its element, not only the type of model but also performing scans and optimisation in its hyperparameter space. This step is often the lengthiest as many configurations are tested, although with some expertise it can be reduced.

Selecting the best model from the hypothesis space $\mathcal{H}$ is done through the loss function

$$L : \mathcal{X} \times \mathcal{Y} \times \mathcal{H} \to \mathbb{R} : ((x, y), h)) \mapsto L\left((x, y), h)\right), \tag{3.2}$$

which assigns to each pair of feature x and label y, given the hypothesis map h, a real positive number quantifying the discrepancy between the prediction and true label.

The objective in the model selection is to find the hypothesis map h that will minimize the loss function for any data point. The form of the loss function depends on the problem at hand, for regression the typically used loss function is the squared error

$$L\left((x, y), h)\right) = \left(y - h(x)\right)^2. \tag{3.3}$$

When estimated over the whole dataset the squared error is averaged over all the data points, the loss function is called the Mean Squared Error (MSE). Its advantages are a convex and differentiable form, which will become handy in the next sections, at the cost of a higher impact of outliers. In contrast, the Mean Absolute Error (MAE) consisting in the average of $|y - h(x)|$ is less impacted by outliers, but is not

differentiable at every point (although this can be solved numerically) and causes additional complexity, a reason for which the MSE is usually preferred.

In classification tasks however, the MSE performs poorly due to its incapability to quantify the confidence of the classification hypothesis. A good loss function should punish (reward) very confident classification with high $|h(x)|$ that are wrong (right). Several loss functions exist with that goal in mind, such as the $0/1$ loss, or the hinge loss, that however suffer from non-differentiability, which is crucial in the following and implies complex optimization methods. One widely used alternative is through the cross-entropy between two distributions p and q

$$H(p,q) = -\sum_x p(x)\log q(x). \tag{3.4}$$

In the case of a binary classification $p \in \{y, 1-y\}$ and $q \in \{h(x), 1-h(x)\}$ and therefore the binary cross-entropy can be used as a loss function

$$L\left((x,y),h)\right) = -y\log h(x) - (1-y)\log\left(1-h(x)\right), \tag{3.5}$$

often called the logistic loss. This definition can be extended to multi-label classification using a categorical cross-entropy.

Bias-variance trade-off

In a regression task – although the following conclusion can be generalized to classification – the true labels y often come from a measurement that can be affected by noise. This noise can be parameterised by a Gaussian distribution with zero mean such that $y(x) = f(x) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ is the noise and f the function that the regression tries to reproduce with $h(x)$.

The expectation and variance of the labels are therefore

$$\begin{aligned} E[y(x)] &= E[f(x)+\epsilon] = E[f(x)] = f(x), \\ \mathsf{Var}[y(x)] &= E[(y(x) - E[y(x)])^2] = E[(y(x)-f(x))^2] = E[\epsilon^2] = \sigma^2, \end{aligned} \tag{3.6}$$

since the noise has zero mean and the function f is independent of the data.

Therefore, the expectation of the squared error loss can be computed

$$\begin{aligned} E[(y(x)-h(x))^2] &= \mathsf{Var}[\epsilon] + (f(x) - E[h(x)])^2 + \mathsf{Var}[h(x)] \\ &= \sigma^2 + \mathsf{Bias}[h(x)]^2 + \mathsf{Var}[h(x)]. \end{aligned} \tag{3.7}$$

The MSE generalization error consists in the intrinsic error of the measurement in the first term, then the sum of the regression bias and variance. These two last terms represent the bias-variance trade-off, a simple model will have a sizeable bias due to an imperfect modelling of the label based on the features, while a more complex model might start modelling the noise in the label and suffer from a large

variance, as illustrated in Fig. 3.1. Depending on the amount of data at hand for the training, selecting the correct amount of complexity is the entire point of the model selection, but there is a threshold that cannot be solely solved by a complexity optimization.

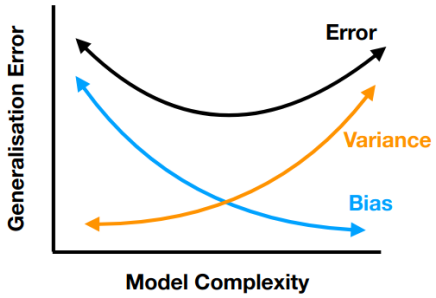


Figure 3.1 | Schematic representation of the bias-variance trade-off, as a function of the model complexity. From Ref. [245].

Regularisation and overfitting

There are several ways to play with the trade-off in Eq. (3.7). One of which is to reduce the dimensionality of the problem, namely to remove some features which have poor discrimination power. This will increase the bias but the reduction in the variance of the prediction can produce a global reduction of the generalization error, this is called *regularisation*. This however becomes unpractical when the number of features is large and the combinatorial effect of evaluating the error for each subset of features is unfeasible.

Other techniques consist in the introduction of additional terms into the loss function that are only based on the parameters of the model. They will be designated by θ such that $\hat{y} = h(x|\theta)$ and can for example represent the parameters of a polynomial fit. These are the parameters that are adapted during the training on the data points, by contrast with the hyperparameters that are fixed for a given model.

In the so-called *ridge regression*, the loss function becomes

$$L_{ridge} = \sum_{i}^{m} \left(y_i - h(x,\theta)\right)^2 + \lambda \sum_{j}^{n} \theta_j^2. \tag{3.8}$$

To the MSE in Eq. (3.8) is added a factor that penalizes large values of the model parameters θ, while the $\lambda \geq 0$ is a new hyperparameter that fixes the trade-off between the model performance and that penalization term. It can be shown that this factor will emerge in both the bias and variance in Eq. (3.7), in such a way

that increasing its value will increase the bias and reduce the variance, potentially decreasing the generalization loss for optimal values of λ.

The *lasso* regularisation technique works in a very similar way, replacing the squared parameters by an absolute value

$$L_{lasso} = \sum_{i}^{m} \left(y_i - h(x, \theta)\right)^2 + \lambda \sum_{j}^{n} |\theta_j|. \tag{3.9}$$

Given their similar form, the lasso and ridge regularisation are often referred to as L1 and L2 regularisation schemes. They perform equally well, with some minor differences such as the fact that with L1 some parameters can be optimized to zero, effectively suppressing some features, which makes it suitable when only a few features are meaningful, if not then the L2 will perform better. As this can however not be predicted beforehand, usually these two hyperparameters are chosen based on a hyperparameter scan or used both in a convex combination.

Overfitting – also called overtraining – occurs when a ML algorithms lose its generalisation capabilities, by learning the specificities of the data rather than patterns. This can happen when the model has too many parameters compared to the amount of data, or that there were too many cycles over the training data, such that the variance over a dataset not seen during the training is large.

Detecting overfitting requires keeping some data out of the training set to act as a validation set, on which the performance is regularly evaluated. When the model starts overfitting, the validation loss will start to increase, contrarily to the loss evaluated on the training set, which by definition can only decrease. This however reduces the amount of data available for training, and the comparison between several models might not be on the same footing if outliers are present in one set or the other in different trainings. Additionally, in the context of the statistical inference of Section 2.4.3 in physics analyses, the events used during the training of a model cannot be used in the likelihood ratio because there will be a mismodelling compared to the real data which has not been seen during the training. Ignoring them on the other hand can reduce by a large margin the statistical power of the simulations.

Cross-validation – also called cross-training – can be used to perform several trainings on the same dataset, each model having access to different parts of the data and the fluctuations are averaged. One version called k-*fold* cross-validation consists in splitting the dataset into k subsets, each model being trained on $k-1$ subsets and evaluated on a single one. In a physics analysis, it can also be useful to keep all the simulation events in such a way that each event is processed by a model that has not seen the event during the training.

Once overfitting has been detected, it can be solved by either reducing the model complexity, or more automatically restrict its expressiveness with regularisation

techniques. In the case of NN in particular, *dropout* [246–248] layers can also be used to randomly ignore a fraction of neurons during the training so general features are learnt without fixating on smaller effects, effectively replacing a large NN into a random ensemble of smaller ones that can be loosely compared to bagging methods.

Gradient descent

The optimization of the model parameters based on the loss function is a highly non-trivial problem, especially considering the potentially high dimensionality of the parameter space. Among the optimization methods, gradient-based methods are well suited for this kind of task given their robustness in many dimensions. The aim is to perform iterative modifications of the model parameters by the local optimization of linear approximations of the objective function.

Gradient Descent (GD) methods are performed in steps whose size is dictated by a hyperparameter call the *learning rate* η, and whose direction is the opposite of the local gradient of the loss function. In the linear approximation, the loss function can be written as

$$L(x|\theta(t)) \simeq L(x|\theta(t-1)) + (\theta(t) - \theta(t-1))^T \nabla_\theta L(x|\theta)|_{\theta=\theta(t-1)}, \qquad (3.10)$$

where $\theta(t)$ represents the set of parameter at the end of the iteration, and $\theta(t-1)$ the ones at the previous iteration. The gradient is computed locally in the parameter space, at the place of the last iteration. This linear approximation is valid provided the two sets are relatively close to each other.

The objective is to minimize the loss function, so in order to have $L(x|\theta(t)) < L(x|\theta(t-1))$, a GD step can be computed as such

$$\theta(t) = \theta(t-1) - \eta \nabla_\theta L(x|\theta)|_{\theta=\theta(t-1)}. \qquad (3.11)$$

The learning rate η dictates the size of each GD step, and its values are important for the training. Too small values and the convergence will become extremely slow, too large values and the algorithm will diverge. In addition, the loss function might exhibit multiple local minima, and a too small learning rate might get the GD trapped inside one, rather than the aimed-at global minimum.

The learning rate is one of the few hyperparameters whose variation through the training might produce improvements. When approaching convergence, the learning rate might be too large to reach the point of global minimum and the GD always overshoots. Several methods exist, such as an analytical decay of the learning rate following a monotonically decreasing function, or to decrease the learning rate when the loss reaches a plateau, indicating a potential back and forth around the minimum. Early stopping [249], or the possibility to finish the training once no gain is observed on the loss function values, is also widely used.

So far, the loss function was defined over the whole set of data points. This is however computationally intensive and might point the GD directly to a local minimum. Dividing the total set into *batches* of data points for which the loss is computed can speed up the convergence and introduce stochastic variations in the gradient direction that have the potential to avoid trapping the training inside a local minimum. An *epoch* is then defined as the number of times the model has seen the entire dataset. While the batch size is an important variable in the hyperparameter set, the learning rate usually has more impact.

The GD, and the Stochastic Gradient Descent (SGD) on batches of the total dataset, can present erratic directions that impede the training In addition saddle points can become numerous in deep models. Several improvements in the GD have been developed in the literature, one of which consists in the addition of a *momentum* [250] to the GD, such that the direction of the gradient does not present sudden large variations over the different steps.

The GD of Eq. (3.11) becomes

$$\begin{aligned} v_t &= \beta v_{t-1} + (1-\beta)\ \nabla_\theta L(x|\theta)|_{\theta=\theta(t-1)}, \\ \theta(t) &= \theta(t-1) - \eta v_t, \end{aligned} \tag{3.12}$$

where v_t is anexponentially weighted average of past gradients and β the parameter that dictates how much momentum the gradient step should keep, $\beta = 0$ returns to the usual stochastic gradient descent.

Similarly, RMSProp uses the memory of squares of the past gradient

$$\begin{aligned} v_t &= \beta v_{t-1} + (1-\beta)\ \nabla_\theta \left(L(x|\theta)|_{\theta=\theta(t-1)}\right)^2, \\ \theta(t) &= \theta(t-1) - \eta \frac{L(x|\theta)|_{\theta=\theta(t-1)}}{\sqrt{v_t} + \epsilon}, \end{aligned} \tag{3.13}$$

where ϵ is for numerical stability. RMSProp is an improvement over AdaGrad [251] that only considered the sum of gradients and not its exponential average. On the other hand, AdaDelta [252] does not consider accumulation of gradient but rather computes the RMS of the gradient.

Adam [253] combines both the momentum with order one and RMSProp of order two, each with its own momentum parameter, to compute the exponentially weighted average of past gradients. Several other algorithms exist, some take advantage of the second order derivative with the Hessian matrix, however it is often difficult to compute and several approximations have to be used.

An alternative to the classic momentum method is to use *Nesterov* [254] accelerated gradients. Several interpretations [255, 256] exist, but the main idea is to compute the gradient at the next iteration step, a form of look-ahead formulation of the gradient descent.

Metrics

There are several ways to estimate the performance of a ML model, depending on the objective. This section will be restricted to classification tasks as the MSE is globally used for regression, with minor improvements in some methods.

The *confusion matrix* is a generalization of the $0/1$ loss for multi-label classification, for each pair of labels $c, c' \in 1, ..., k$ among k labels, the loss can be written as

$$L^{c\to c'}\left((x,y),h)\right) \equiv \begin{cases} 1 \text{ if } y = c \text{ and } h(x) = c' \\ 0 \text{ otherwise} \end{cases} .$$

Over the whole dataset $\mathcal{D}$, the loss can then be computed for each c and c' as

$$\hat{L}^{c\to c'}(h|\mathcal{D}) = \sum_i L^{c\to c'}\left((x_i,y_i),h)\right), \tag{3.14}$$

such that $\hat{L}^{c\to c'}(h|\mathcal{D})$ represents a matrix with row c as true label and column c' as predicted label. The perfect classifier would have a diagonal confusion matrix $\hat{L}^{c\to c'}(h|\mathcal{D}) = N_{c,c'}\delta_{c,c'}$.

Although also defined in a multi-classification, the following metrics are easier to understand in the context of a binary classification, as illustrated on the confusion matrix of Fig. 3.2. In HEP both the TPR and FPR have meaning in a signal versus background classification, as the signal efficiency and background contamination rate respectively. In an ideal classifier the former would be maximal and the latter minimal, a visualisation of the dependence between the two quantities as a function of the threshold applied on the ML score is the Receiver Operating Characteristic (ROC) curve, as illustrated on Fig. 3.3. Better classifiers will have a larger Area Under Curve (AUC), and the decision boundary can be set on one point of the ROC curve.

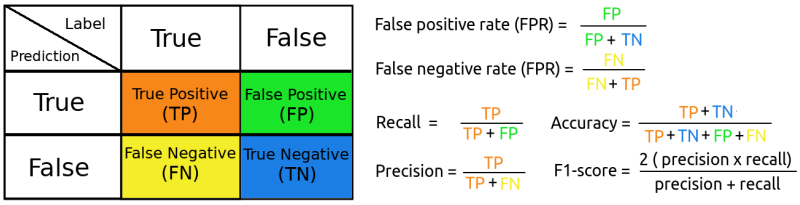


Figure 3.2 | Binary classification confusion matrix, with several metrics definitions.

The *precision* quantifies the fraction of signal events out of all the events passing the ML selection, while the *recall* quantifies the fraction of signal events that were selected from the initial distribution. Ideally they should both be maximized, but whether the focus is on one or the other depends on the objective, another common metrics combining the two is called the *F1-score*, as defined on Fig. 3.2.

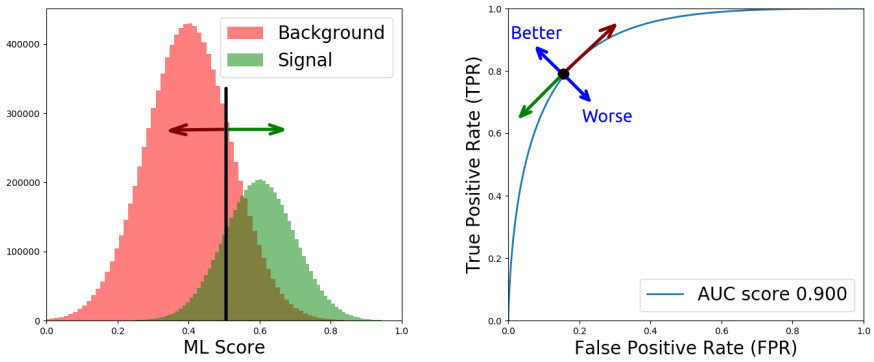


Figure 3.3 | Classification score from the ML algorithm (left) and associated ROC curve with its AUC score. The effect of moving the decision boundary on the distribution on the ROC curve with the arrows.

Binary classification and likelihood ratio

As described in Section 2.4.3, in case of a density $p_X(x|\theta)$ of a random vector X with parameterisation θ, the most powerful test statistic between two hypotheses θ_0 and θ_1 over a dataset $\mathcal{D} = \{x_1, ..., x_n\}$ is given by

$$\lambda(\mathcal{D}; \theta_0, \theta_1) = \prod_{x \in \mathcal{D}} \frac{p_X(x|\theta_0)}{p_X(x|\theta_1)} = \prod_{x \in \mathcal{D}} r(x; \theta_0, \theta_1). \tag{3.15}$$

It can be shown that the likelihood ratio in Eq. (3.15) is strictly equivalent to the one obtained by a change of variable $u = s(x)$, where s is a strictly monotonic function over the density ratio $r(x; \theta_0, \theta_1)$ [257, 258]. It is possible to produce this monotonic function using supervised classification algorithms, with a method called the *likelihood ratio trick* [259, 260]. If a classifier is trained to separate two equal-sized samples $\{x_i\} \sim p_X(x|\theta_0)$ and $\{x_i\} \sim p_X(x|\theta_1)$, the binary cross-entropy loss

$$L_{BCE} = -\frac{1}{N} \sum_{i=0}^{N} \delta(\theta = \theta_1) \log(\hat{s}(x|\theta_0, \theta_1)) + \delta(\theta = \theta_0) \log(1 - \hat{s}(x|\theta_0, \theta_1)), \tag{3.16}$$

will be minimised by the optimal decision boundary

$$s(x|\theta_0, \theta_1) = \frac{p_X(x|\theta_1)}{p_X(x|\theta_0) + p_X(x|\theta_1)} = \frac{1}{1 + r(x|\theta_0, \theta_1)}. \tag{3.17}$$

In the case where the classifier reaches this optimal decision boundary ($\hat{s} \to s$), the monotonic relation between s and r in Eq. (3.17) ensures that the classifier can be used as a surrogate for the likelihood ratio in Eq. (3.15) and has reached

the best classification performance. While in practice hard to achieve, a calibration method can restore the relation as long as s and r follow a monotonic relation. This development further assumes a smooth variation around x of the approximate map $\hat{s}(x)$, and is therefore more suited for a neural network architecture than a BDT that employs steps functions which might not be monotonic.

3.1.2 Boosted Decision Trees

A *decision tree* is a flowchart-like hypothesis map, consisting of nodes connected by edges, and the decision $h(x)$ results from an iterative division of the feature space into piece-wise constant subsets of this space. Starting from a *root* node where all features pass through, multiple nodes apply a decision function on one feature each, until no further decision is applied and the category for the classification is defined in a *leaf*, as illustrated on Fig. 3.4. In the case of a regression, the leaf represents a numerical value instead of a class label, and the hypothesis map can be interpreted as a piece-wise constant function. The depth of the decision tree is defined by the number of decision nodes passed before arriving to a leaf.

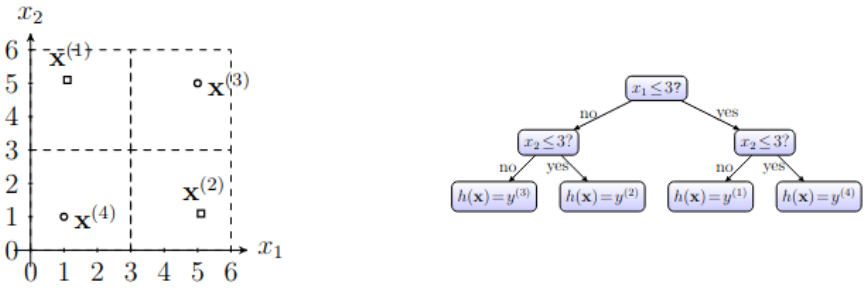


Figure 3.4 | Example of a decision tree with a depth of two (right), illustrating a hypothesis map $h : \mathbb{R}^2 \to \mathbb{R}$ that splits the feature space into four rectangular subregions (left).

At each step of the algorithm, starting from the root node, a decision node has to be chosen to maximize the information, or express differently to produce two branches with the homogeneous distinction. The information gain G for a specific feature f between the two branches emerging from a parent node is defined by the *exact greedy algorithm*

$$G(D_p, f) = I(D_p) - \frac{N_{left}}{N_p} I(D_{left}) - \frac{N_{right}}{N_p} I(D_{right}), \tag{3.18}$$

where D_p, D_{left}, D_{right} are the datasets of the parent, left and right branches respectively, with their associated sizes N_p, N_{left}, N_{right}.

The function I quantifies the impurity of a branch, that can either be computed with the entropy $E(D) = -\sum_i p_i \log(p_i)$ or the Gini index $G(D) = 1 - \sum_i (p_i)^2$,

where p_i are the probabilities of each class. The information gain can be seen as the decrease in entropy from the splitting of the node, the feature that produces the largest gain is selected and the decision node is created. The algorithm continues downstream until the maximum depth is reached, or the node becomes pure. When the feature is not categorical, its numeric values can be ordered and the splitting threshold is found for the value of the feature that maximizes the information gain.

Decision trees tend to overfit very fast with the depth, since deep decision trees have a high variance. To overcome this, the decision trees are usually shallow and combined. *Bagging* consists in selecting several of these *weak* learners, and the decision is drawn from the most selected class for classification or the average for regression. This produces a *strong* learner that is less prone to overfitting. A variation of this procedure in the context of decision trees is a process known as a *random forest*, where only a random subset of features is used for each tree, a process called *feature bagging*. Similarly, *boosting* consists in training weak learners in series, where each data point that is misclassified is assigned a higher weight for the next weak learner. The final decision consists in a weighted average of each of the T weak learner's decision $H_T(x) = \sum_t^T \alpha_t h_t(x)$, where h_t is the weak learner at the iteration time t. In the context of decision trees this is known as a Boosted Decision Tree (BDT). Both bagging and boosting are not restricted to decision trees, though they are the current standard for that type of ML algorithm.

In AdaBoost [261] each boosting step works as follows :

- a weak learner tries to minimize the misclassification error $\epsilon_t = \sum_{i,h(x_i)\neq y_i}^n w_{i,t}$ where $w_{i,t}$ is the weight of misclassified data point i at iteration step t,
- the weight function is defined as $\alpha_t = \frac{1}{2}\log\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$,
- the weak learner is added to the ensemble as $F_t(x) = F_{t-1}(x) + \alpha_t h_t(x)$, such that relatively good learner h_t participate more to the global decision,
- the data point weights are updated as $w_{i,t+1} = w_{i,t} e^{-y_i \alpha_t h_t(x_i)}$, such that misclassified points have an increased weight, and then each weight is normalized so that $\sum_i^n w_{i,t+1} = 1$.

Another widely used algorithm is XGBoost [262] that uses gradient boosting. The idea of this method is still to aggregate weak learners as in classical boosting, but the addition of each weak learner aims at minimizing the loss function by GD steps, that works as follows :

- the *pseudo-residuals* are computed as the local gradient of the loss function

$$r_{i,t} = \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right] |_{F(x)=F_{t-1}(x)}$$

  for each data point (x_i,y_i) at iteration step t,
- train a weak learner $h_{t+1}(x)$ on training set $\{x_i, r_{i,t}\}$,
- find the weight that minimizes the loss function

$$\alpha_t = \min_\alpha \sum_i L(y_i, F_t(x_i) + \alpha h_t(x_i),$$

- update the model $F_t(x) = F_{t-1}(x) + \alpha_t h_t(x)$.

In particular, XGBoost uses both the first and second derivative to perform the steps above.

BDTs are particularly useful when the amount of training data is not large and. They are especially robust, for example against missing values, imbalanced datasets, and redundant attributes.

3.1.3 Neural Networks

Neuron model

The basic element of a NN is the *neuron* – sometimes called *perceptron* [263] in the literature – illustrated on Fig. 3.5a. It takes a certain number of input signals from a vector of features through a set of *synapses*, each characterized by a weight that parameterises the strength of the connection. A bias can be introduced and represented as an additional weight applied on a unit fixed input. The signals are then summed, at which point the response is still linear, and the nonlinearity is obtained by an activation function ϕ. The whole chain reads

$$y_k = \phi(v_k) \quad , \quad v_k = \sum_{i=0}^{p} w_{ki} x_i \quad (x_0 = 1) \tag{3.19}$$

for a neuron indexed by k.

A NN is then built by associating several neurons into a layer and associating several layers into a network, as depicted on Fig. 3.5b. Each neuron in a layer takes as input signal the outputs of the neuron in the previous layer. *Hidden* layers refer to the layers that are neither receiving the input signals, nor producing the outputs. When several of them are present the NN is commonly referred to as a Deep Neural Network (DNN), allowing more complicated patterns to be learnt, one of the benefit of the advance of deep learning in general.

The activation function is the only hyperparameter attached to the neuron, and plays a crucial role in introducing some non-linearity. While the logistic function $1/(1 + e^{-x})$ or the hyperbolic tangent $\tanh(x)$ were mostly used in the past, illustrated on Fig. 3.6, they lead to *vanishing gradient* issues. This happens when, during the training, the inputs of the neurons have large values and the output is located on the flat region of the activation function. The derivative values then become vanishingly small and the learning stops, which became a showstopper for deep models as the gradients inevitably vanished in the early layers. This was circumvented by the introduction of the Rectified Linear Unit (ReLU) activation

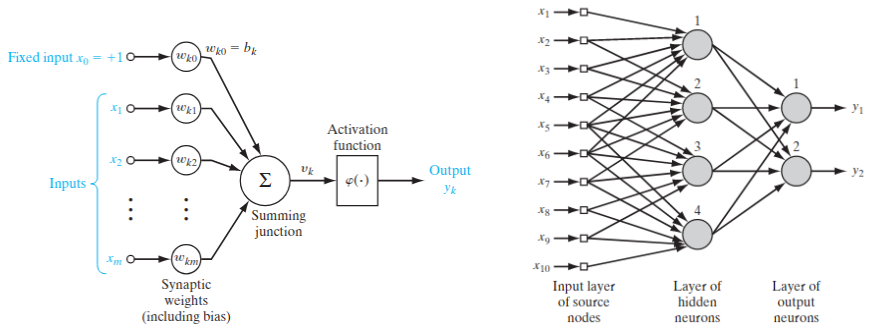


Figure 3.5 | Nonlinear model of a neuron (left), following Eq. (3.19), and combination of several neurons into a network (right), from Ref. [240].

function and its variations that introduced a constant gradient over the whole range of positive values. Although there is a discontinuity around zero in the derivative, this can be solved numerically and is the source of non-linearity in each neuron. In addition to avoiding the saturation issue, ReLU activation functions are suspected to allow for a similar effect to the dropout where negative inputs produce zero gradient, effectively shutting down the neuron.

$$\phi_{\text{ReLU}} = \begin{cases} x \text{ if } x > 0, \\ 0 \text{ else} \end{cases} , \phi_{\text{Leaky ReLU}} = \begin{cases} x \text{ if } x > 0, \\ \alpha x \text{ else} \end{cases} , \phi_{\text{ELU}} = \begin{cases} x \text{ if } x > 0, \\ \alpha(e^x - 1) \text{ else} \end{cases} . \tag{3.20}$$

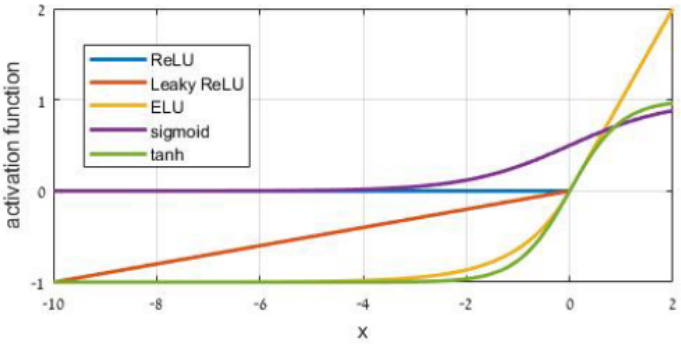


Figure 3.6 | Various activation functions used in the literature, notably the ones defined in Eq. (3.20), from Ref. [264]. For Leaky ReLU and ELU, $\alpha = 0.1$.

Back-propagation

GD from Section 3.1.1 can be performed by computing the gradient of the loss function with respect to the parameters of the model, in this case the neuron weights. For the output neurons, this poses no issues, but the neurons in the hidden and input layers do not have direct access to the loss function. The gradient in these layers is then obtained from the *back-propagation* [265] of the gradient from the output layer, going backward by contrast to the forward propagation of the input signals.

The loss function over the dataset (or over a batch) of N events is $L(x|W) = \frac{1}{N}\sum_{n=1}^{N} E(n)$, where $E(n)$ is the error computed on each output neuron, and $W = \{w_{ij}^k\}$ is the set of weights of the NN (k is the index of the layer, i the neutron and j index the weights inside the neuron). For a single data point, the gradient for a specific neuron i at layer k can be computed as such

$$\frac{\partial E(n)}{\partial w_{ij}^k} = \frac{\partial E(n)}{\partial v_j^k}\frac{\partial v_j^k}{\partial w_{ij}^k}, \tag{3.21}$$

using the chain-rule on Eq. (3.19).

If we denote by $\delta_j^k \equiv \frac{\partial E(n)}{\partial v_j^k}$ the local gradient, then

$$\frac{\partial E(n)}{\partial w_{ij}^k} = \delta_j^k y_i^{k-1}, \tag{3.22}$$

and the gradient $\nabla_W L(x|W)$ from Eq. (3.11) at layer k can be computed from the local gradient at layer k and output of the previous layer associated to the weight.

The GD can then be applied on each weight – or any of the amelioration in Section 3.1.1 – as

$$w_{ij}^k(t+1) = w_{ij}^k(t) - \eta\frac{\partial E(n)}{\partial w_{ij}^k}, \tag{3.23}$$

and the back-propagation only depends on δ_j^k.

For output neurons the chain rule can be used as the loss function and labels t_j – also called *targets* – are known, for example in the case of the MSE

$$\delta_j = \frac{\partial L(y_j, t_j)}{\partial y_j}\frac{\partial y_j}{\partial v_j} = -2(t_j - y_j)\phi'(v_j). \tag{3.24}$$

For hidden layers it can be computed from the errors of the next layer

$$\delta_j^k = \sum_m \frac{\partial E(n)}{\partial v_m^{k+1}}\frac{\partial v_m^{k+1}}{\partial v_j^k} = \phi'(v_j)\sum_m \delta_j^{k+1} w_{ij}^k \tag{3.25}$$

The learning steps then go as follows : a batch of event is propagated through the network to compute all the v_j^k and y_j^k, the local gradient on the output layer is

computed from the loss function and expected targets, then back propagated to the previous layers, and finally for each neuron the weights are adapted with GD following Eq. (3.23).

Initialisation

Given the complexity of the loss space that has to be minimised, the initial conditions on the parameters play an important role. The initialisation procedure can depend on the activation functions, and the weight initialization can be non-intuitive. For example, constant initialisation for each neuron will have them produce the same predictions, go through the same gradient and identical evolution. Too small initial weights will produce a gradient vanishing and low convergence, and too large values will make the learning diverge with exploding gradients. The best scenario is when the mean of the activations are around zero, and their variance should not vary too much between layers, several schemes going in that direction have been suggested [249, 266, 267]. Through the training the weights are updated and will not follow this condition beyond initialisation, leading to more and more different layer input distributions as the training goes, this is called *internal covariate shift*. Normalising the activation functions during the training can circumvent this issue and is done through a technique called *batch normalisation* [268].

Similarly, the features provided to the NN should go through a *pre-processing* phase. If they are provided as such to the algorithm, they may have different impacts over the training simply through their different ranges and mean values. A single layer NN will fully linear activation function would not be impacted by this, as the weights will adapt to the feature inputs, but deeper and non-linear models do not have this advantage. Therefore, one common practice is to rescale each feature by $x' = (x - \mu)/\sigma$ so their distributions have zero mean and unit variance.

3.2 Matrix Element Method

3.2.1 Description

The MEM originated from the Tevatron experiments DØ and CDF for the measurement of the top quark mass [269–276], and has been used in a variety of analyses in both CMS [277–280] and ATLAS [281–285] collaborations. The aim of the method is to compute $P(x|\alpha)$, i.e. the probability to observe an experimental event x, seen as a collection of detected particles 4-momenta $\{P_1, P_2, ..., P_n\}$, from a theoretical hypothesis α, that can represent a specific physics process, a mass, a coupling, or a combination of them.

The hard scattering likelihood from Eq. (1.16) can be rearranged for the particular case of a hadron collider in such a way

$$d\sigma(q_1 q_2 \to y;\alpha) = \frac{(2\pi)^4 |\mathcal{M}(q_1 q_2 \to y;\alpha)|^2}{q_1 q_2 s} d\Phi(y), \tag{3.26}$$

to take into account the partons fraction of momentum q_1 and q_2, and the center-of-mass energy s. The phase space measure $d\Phi(y)$ is the same as in Eq. (1.17).

To obtain the differential cross-section in pp collisions, Eq. (3.26) must be convoluted through the flavour dependent PDF $f_a(q)$ for flavour a

$$d\sigma(pp \to y;\alpha) = \int_{q_1,q_2} \sum_{a_1,a_2} f_{a_1}(q_1) f_{a_2}(q_2) d\sigma(q_1 q_2 \to y;\alpha) \tag{3.27}$$

The final-state particles contained in y differ from the ones in x, as they are the particles that emerged from the hard scattering before hadronisation and parton shower. These effects need to be convoluted with Eq. (3.27) but are untraceable and would not allow an integration over the phase space. To solve this, a parameterisation of all the detector effects can be included in a TF $T(x|y)$, that can either be analytical or a simulated-based binned histogram. This approximation represents several highly non-trivial phenomena, and it is necessary to use several factorisation assumptions to allow for their determination, namely between the different particles in the final-state and their different variables

$$T(x|y) = \prod_{i=1}^{n} T_i^E(x^i|y^i) T_i^\eta(x^i|y^i) T_i^\phi(x^i|y^i), \tag{3.28}$$

where E, ϕ and η represents the energy and angular dependencies of the TF for each particle i in the final-state. In practice, only the energy resolution matters. The directions of the detected leptons is known with such high precision that their resolutions can be assumed as delta functions. This is not entirely true for jets, however the angular resolution is much better than the one of the energy. A reasonable approximation would be to consider that angular variations are already covered in the energy TF, and that their resolution can be modelled by delta functions, which will be the case in the following. Care must however be taken for objects with small separation angle, for which the delta or factorisation assumptions might break down. Neutrinos, particles outside the acceptance and initial-state partons can represent a broad volume on which to integrate, negatively impacting the integration unless kinematic constraints are used to remove these degrees of freedom.

It is also assumed that the TF is entirely decoupled from α, since in general the detector reconstruction is independent of the hard scattering. This is not always true, although it very often is, as it may impact the angular separation of the particles in the final-state. For example, a lighter resonance can decay into a more collinear pair of jets and the TF would need to be adapted to the resonant mass. This must be evaluated case-by-case, but the effect is likely to be negligible.

The MEM can then be summarised in the following form

$$P(x|\alpha) = \frac{1}{\sigma_\alpha^{vis}} \int_{q_1,q_2} \sum_{a_1,a_2} \int_y d\Phi(y)\ dq_1\ dq_2\ f_{a_1}(q_1)\ f_{a_2}(q_2) \frac{(2\pi)^4 |\mathcal{M}(q_1 q_2 \to y)|^2}{q_1 q_2 s}\ T(x|y)\ dy. \quad (3.29)$$

The visible cross-section σ_α^{vis} acts as normalisation factor and is there to make sure $P(x|\alpha)$ can be treated as a likelihood :

$$\int_x P(x|\alpha)\ \epsilon(x)\ dx = 1 \Leftrightarrow \sigma_\alpha^{vis} = \int_{x,y} d\sigma(pp \to y;\alpha)\ T(x|y)\ \epsilon(x)\ dx\ dy, \quad (3.30)$$

where $\epsilon(x)$ represents the selection efficiency of event x that is used in the measurement, representing the detector acceptance, trigger and offline selections. In practice $\epsilon(x) = 1$ for selected events and $\epsilon(x) = 0$ otherwise. For other applications that do not necessarily require a normalised quantity, for example as input to a MVA, the *MEM weight* is defined as $w(x|\alpha) = \sigma_\alpha^{vis} \times P(x|\alpha)$.

As visible on Eq. (3.30), the normalisation of σ_α^{vis} is directly related to the normalisation of the TF :

$$\int_x T(x|y)\epsilon(x)\ dx = 1. \quad (3.31)$$

There are two ways to handle the efficiency in Eq. (3.31), mostly related to the particle energy as most often the angular TFs will be represented as delta functions. The *process-based* normalisation scheme considers that a generated parton will produce a jet of any possible energy, or not reconstructed at all. In this case, the TF is independent of the selections and $\epsilon(x) = 1$ for the events x that are evaluated in Eq. (3.31). This scheme is useful to accommodate the fact that in some events not all partons have produced a jet. Another possible scheme is dubbed *selection-based*, it is based on the concept that all objects that enter the integration are supposed to have passed the selections. Since the acceptance $\epsilon(x)$ is a binary condition, Eq. (3.31) can be interpreted as normalising the TF with an integration lower bound that may depend on other quantities, for example the pseudo-rapidity. An additional upper bound of the integration would typically be included for leptonically decaying tau, identified by the subsequent lepton whose energy cannot exceed the initial tau. This new TF $T'(x|y)$ is then such that $\epsilon(x) = 0 \Rightarrow T'(x|y) = 0$, i.e. that an event not passing the selections will have a TF value of zero. Therefore $\int_x T'(x|y)\epsilon(x)\ dx = 1$, and factorising Eq. (3.30) shows that σ_α^{vis} does not depend on the TF, which is the advantage of this normalisation scheme. For the work developed in this chapter, the events used to derive the values of $w(x|\alpha)$ have passed the selections and belong to the selection-based scheme.

3.2.2 Adaptative numerical integration and optimisations

The integral in Eq. (3.29) cannot be solved analytically and must be resolved through numerical integration. Several methods exist, but only Monte Carlo techniques scale reasonably well with the number of dimensions. As we are considering high-dimension integration, this section will be restricted to such methods. To obtain the integral value of a function f with a Monte Carlo method, N points $\{x_n\}$ are generated randomly in the integration space Ω and evaluated by the function

$$I = \int_\Omega f(x)dx \simeq \frac{1}{N}\sum_{n=1}^{N} f(x_n). \tag{3.32}$$

This method error typically scales as $\sqrt{V_N/N}$, where V_N represents the variance of f over the N points.

In practice, to decrease the error for a fixed number of points, the variance must be reduced. This is achieved through adaptative Monte Carlo technique, where the points are not draw uniformly over the whole space Ω. Instead, they are drawn from a distribution that emphasises the sampling of points where f is large, with a weight that counteracts the bias in this function. This is called *importance sampling* and can be expressed with a sampling function g such that the integral is computed as

$$I = \int_\Omega f(x)dx = \int_\Omega \frac{f(x)}{g(x)} g(x)dx \simeq \frac{1}{N}\sum_{n=1}^{N} \frac{f(x_n)}{g(x_n)}, \tag{3.33}$$

where now the points x_n are generated according to g. Ideally, if $g = f/I$ then the variance is minimal. However, the integral value is not known a priori, and there is in general no easy way to generate points from the function f. Therefore, the goal is to use an approximation such that $g \sim f$, making sure that $g(x_n) \geq f(x_n)(\forall x_n \in \Omega)$ to not under populate the tails of f and underestimate the integral. Adaptative methods attempt to iteratively modify the sampling functions during the evaluation of the sum in Eq. (3.33) to get close to the optimum. These methods were already mentioned in Section 1.3 for event generation, the one used throughout this chapter is VEGAS [22].

An alternative to importance sampling is called *stratified sampling*. It consists in subdividing the whole space into subregions $\Omega = \bigcup_{m=1}^{M} \Omega_m$, in each of which a number of events N_m are generated, such that $\sum_{m=1}^{M} N_m = N$. In this case, the integral can be computed as

$$I = \int_\Omega f(x)dx \simeq \sum_{m=1}^{M} \frac{v_m}{N_m} \sum_{n=1}^{N_m} f(x_n^{(m)}) \quad (x_n^{(m)} \in \Omega_m), \tag{3.34}$$

where v_n is the volume of the subregion Ω_m. It can be shown that the variance

of this estimator is always less than or equal to the variance of the estimator over the whole phase-space Ω. Therefore, by iteratively modifying the division of the phase-space such that regions where the values of the functions are large have a smaller volume and more points are generated in them.

Both case rely on the factorisation assumption, which assumes that the sampling function can be factorised between each variable $g(x_i) = g_1(x_i^{(1)}) \times ... \times g_1(x_i^{(|\Omega|}))$. This means that the algorithm will optimise each variable independently, which prevents a factorial growth in terms of complexity. An issue arises when a peak appears in f that is mapped to more than one variable of integration. In this case, the factorisation assumption prevents the algorithm to converge to an optimal point generation, as it cannot take into account the correlation. This can lead to a larger variance, or the plain non convergence of the integration.

In Eq. (3.29), narrow peaks can come from two origins : narrow resolutions in the transfer function or propagator enhancements in the ME. The former is already dealt with in Eq. (1.17) as the variables of integration $d^3p = p^2 dp \sin\theta \; d\theta \; d\phi$ are the ones on which the resolutions are expressed. The ME on the other hand can present narrow Breit-Wigner resonances that need to be inverted. In addition, the delta function for the energy-momentum conservation and missing particles such as neutrinos have to be integrated out. Usually one can however combine the two to improve convergence of the integration.

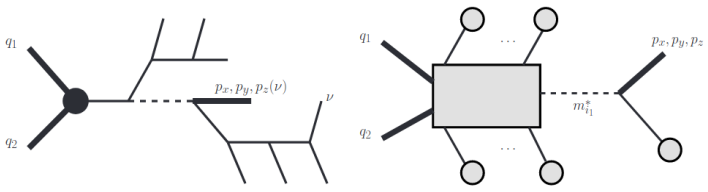


Figure 3.7 | Left : arbitrary decay chain, with several resonances and final state particles, including missing particles as neutrinos. Right : representation of the decay chain as seen during the change of variable. In this particular case, the initial variables of the parton fraction of momentum q_1 and q_2, as well as the three components of one of the neutrino (thick lines), represent five degrees of freedom. The conservation of momentum delta function in Eq. (1.17) is used to remove four of them, the last one is replaced by the invariant mass of the resonance $m^*_{i_1}$ (dotted line), such that its Breit-Wigner contribution in the ME is represented by that sole integration variable. From Ref . [286].

To circumvent this issue, a general algorithm was proposed in Ref. [286], where several analytical changes of variables associated to various diagram topologies have been proposed to create a more efficient mapping of the phase space. After these changes of variables, each peak is mapped to a single variable of integration, as illustrated on Fig. 3.7. Later on, a modular toolkit called MoMEMta [287] (used in this document) was developed to encompass all the required parts in the integration of Eq. (3.29), and perform the integration.

3.3 Hybrid MEM+DNN

3.3.1 The method in a nutshell

The method that was presented in Ref. [1] is based on the observation that out of the two main drawbacks of the MEM, the complexity of the integration and the computation time, only the former was addresses by MoMEMta. Depending on the complexity of the integration phase-space, the computation of the MEM weight might take from a few seconds to a few minutes. This is already expensive for the amount of data and simulations generated in LHC experiments, and becomes downright unusable when considering that it has to be repeated for each hypothesis α, that can be several processes, several simulation parameters or both. And this scales exponentially with the dimensionality of α.

On the other hand, while training a DNN takes some time and expertise, once it is trained the evaluation time is very short, especially considering vectorization and batch processing. In addition, it was realized that Eq. (3.29) is merely a function of the visible particles 4-momenta, albeit non-analytical. Provided this function has a relatively smooth shape, which we expect from a likelihood-like quantity, it can be learnt by a DNN as a regression on a small but statistically significant enough subset of the simulated processes at detector-level via MoMEMta. Then it can be applied on the whole simulated dataset as well as real data events using the fast inference of the DNN. This gain of time over the whole dataset should counterbalance the overhead time spent generating the training sample and training the DNN.

The practical use case for the validation of this method is the resonant $\mathrm{H} \to \mathrm{Z}(\to \ell^+\ell^-)\mathrm{A}(\to \mathrm{b\bar{b}})$ process that arises in the context of 2HDM model from Section 1.6, that was studied in both ATLAS [288] and CMS [289], the latter being the framework in which the work of this section was performed. This analysis consists in several processes treated very differently in the MEM computation. The signal $\mathrm{H} \to \mathrm{ZA}$ features multiple resonances that represent a challenge in terms of efficient mapping of the phase space, and the dependence on two unknown parameters – the masses of the hypothetical extra A and H Higgs bosons – illustrate the struggle of the classical method with multidimensional parameter space. The main backgrounds consist of fully leptonic $\mathrm{t\bar{t}}$ events with the presence of two undetected neutrinos that complicates the MEM integration, and Drell-Yan $\mathrm{Z}^*/\gamma \to \ell^+\ell^-$ + b-jets (DY) events with a large final-state particles kinematic range. The TFs for the leptons and b-jets were taken from Ref. [290] and some of them are depicted on Fig. 3.8.

The event used in this section are the ones passing the selections defined in Ref. [289], reconstructed with the full CMS simulation. Events are first required to pass dilepton triggers, then are further selected if they contain a pair of opposite-sign leptons. Electrons must have $|\eta| < 2.5$ and muons $|\eta| < 2.4$. Asymmetric p_{T} requirements are applied based on the pair of leptons, the leading electron (muon) must be

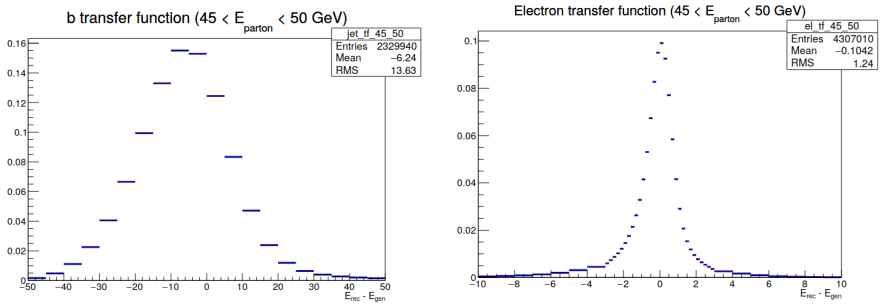


Figure 3.8 | Transfer function for electrons and b-jets as a function of the true generated energy E_{gen} and its reconstructed value E_{rec}, for specific ranges of the former. From Ref . [290].

above 25 (20) GeV and subleading 15 (10) GeV. Jets are required to have $p_{\mathrm{T}} >$ 20 GeV, $|\eta| < 2.4$ and be isolated from the leptons by $\Delta\mathrm{R} > 0.3$. The DeepCSV algorithm is used to identify candidate b-jets using the medium WP. Events must have at least two b-tagged jets, the pair of b-jets is selected as the two jets with the highest b-tag scores. Some final selections are performed before signal extraction, $70 < m_{ll} < 110$ GeV to enhance the presence of $\mathrm{Z} \to \ell\ell$ and $E_{\mathrm{T}}^{\mathrm{miss}} < 80$ GeV to reduce the $\mathrm{t\bar{t}}$ background contamination.

The input variables to the DNN are the 4-momenta of the visible particles in the final-state, namely the two b-jet candidates and the two opposite-sign leptons. In fully connected DNN architectures such as the one used in this section, the order of the inputs has no importance, however it needs to remain consistent throughout training and evaluation. To order leptons and b-jets, a p_{T}-ordering was used. The MET is also included in both the MEM computation and DNN training for the $\mathrm{t\bar{t}}$ weight, as it is used to constrain the neutrinos phase-space.

To provide relatively good invariants the (p_{T},η,ϕ,M) frame has been chosen for the momenta instead of its Cartesian equivalent, the mass being further dropped as it is hard to evaluate for jets for example. Given the cylindrical symmetry, the ϕ angle of each particle has been given as relative to the leading lepton, to remove an arbitrary reference in the azimuthal angle. In that frame of reference, any boost along the beam direction will conserve the ϕ and η differences between objects, in the relativistic approximation, and the p_{T} is not impacted. This parameterisation has shown better performance, as the DNN does not have to learn about the Lorentz boost originating from the parton initial difference in momentum. Finally, a standard preprocessing has been performed by removing the mean and dividing by the standard deviation of each input variables in the training set. The MEM weights can span several orders of magnitude, it is therefore useful to regress on the so-called *event information* $I(x|\alpha) = -\log w(x|\alpha)$.

The training sample consists of a few hundred thousand events in each of the three

categories defined previously, the H $\rightarrow$ ZA sample being an aggregation of signal events in 23 different mass configurations (m_{A},m_{H}). Each event had its weight computed by MoMEMta and then used to train one DNN per process hypothesis. The special case of the H $\rightarrow$ ZA where the weight had to be computed for each mass configuration used a parametric DNN [291]. During that training for H $\rightarrow$ ZA events the values of m_{A} and m_{H} were provided as inputs, while for $\mathrm{t\bar{t}}$ and DY events they were provided randomly, in the same proportions as the signal. During inference, the values of the masses can be set to any given set of parameters, with the advantage that these parameter values do not necessarily need to have been seen during the training, taking full advantage of the interpolation capability of the DNN and already illustrating one benefit of this hybrid method. The dependence of the weight – through the more understandable event information – on the m_{A} and m_{H} parameters is illustrated on Fig. 3.9. When the parameters used in the event generation are the same as the ones used for the MEM weight computation, its value is higher than for different parameters, the difference growing with the gap between the real and tested parameters. Even when only one of the tested parameters matches the true parameter, the weight is higher than when both parameters are different from the true values.

The DNNs that were selected for this work consist of fully connected dense networks with several layers of many neurons, with ReLU activation functions in the hidden layers and SELU [292] – a variant of the ELU activation function – in the output node. More details concerning the training hyperparameters can be found in Ref. [1]. The DNNs were trained with Keras [293] and a Tensorflow [294] backend, the entire dataset was separated into different datasets. The training set was used in the actual update of the model weights, while in parallel the validation set was used to detect any overfitting during the training. The evaluation set was used in the model selection over the hyperparameter space to select the model with the lowest regression loss (mean squared error). The events in the test set were not used at any prior stage and have been used in the illustrations of this chapter to ensure no bias.

3.3.2 MEM weight regression results

In this section we compare the regression of the weights in the test set – not seen at any step, either training or model selection – for each of the weight and its associated regressive DNN. The results of the regression are on Figs. 3.10 to 3.12, compared with the predicted weight by MoMEMta. In all scenarios, despite some unavoidable regression errors, the DNNs perform reasonably well.

The DY process is the easiest to compute from the side of the MEM given that the kinematic range of the final-state particles is large and the association of the DY hypothesis with any event allows fast convergence in a few seconds. On the

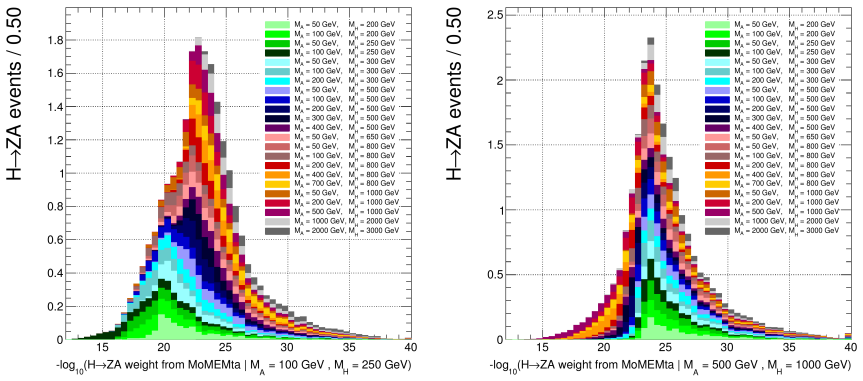
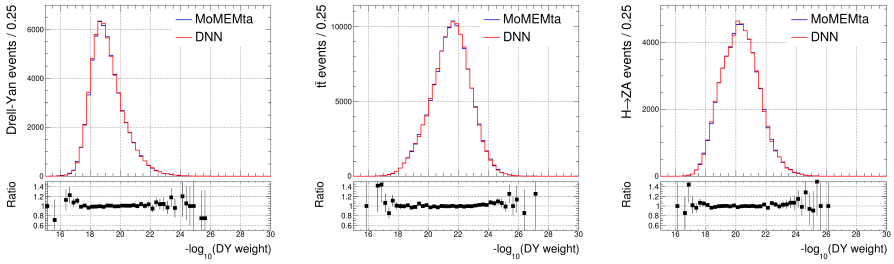


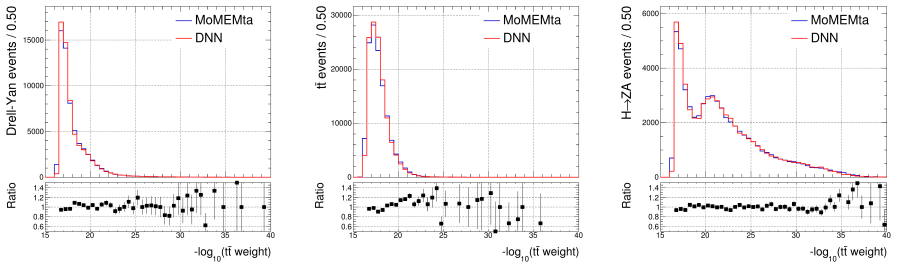
Figure 3.9 | Event information $I(x|\alpha) = -\log w(x|\alpha)$ of $H \to ZA$ events generated with various values of the parameters m_A and m_H but evaluated for specific cases : $m_A = 100$ GeV and $m_H = 250$ GeV (left), $m_A = 500$ GeV and $m_H = 1000$ GeV (right). The histograms have been stacked for easier visualisation.



(a) *(b)* *(c)*

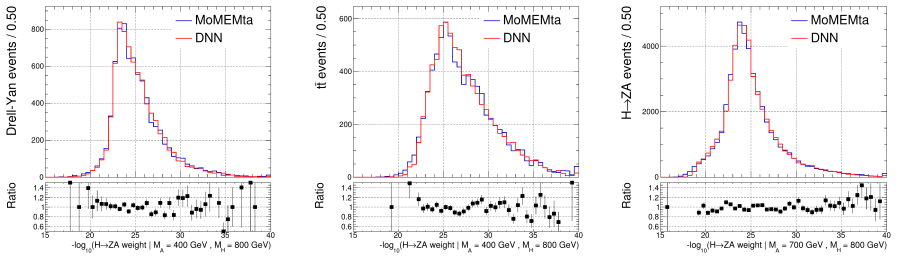
Figure 3.10 | Event information of events from the three different processes, DY (left), $t\bar{t}$ (middle) and $H \to ZA$ (right) events, evaluated on the DY hypothesis.

other hand the $t\bar{t}$ process features resonances of the top quark, their inversion can be tackled using appropriate changes of variables, however this precursor at 173 GeV makes the integration grid less suited to process that do not display such a resonance, impacting the computation time that can be several times longer than for DY weights. As expected, DY events have higher weights with the DY hypothesis than the $t\bar{t}$ events as visible on Figs. 3.10a and 3.10b, and inversely for the $t\bar{t}$ hypothesis on Figs. 3.11a and 3.11b. It is however not straightforward to compare Fig. 3.10a and Fig. 3.11a, as the σ_α^{vis} is absent in the MEM weight. The $H \to ZA$ events have a higher weight for the DY hypothesis than the $t\bar{t}$ events in Fig. 3.10c because of the larger mass resonances, while the two peak structure under the $t\bar{t}$ hypothesis in Fig. 3.11c comes from low values of the parameters that are more



(a) *(b)* *(c)*

Figure 3.11 | Event information of events from the three different processes, DY (left), $t\bar{t}$ (middle) and $H \to ZA$ (right) events, evaluated on the $t\bar{t}$ hypothesis.



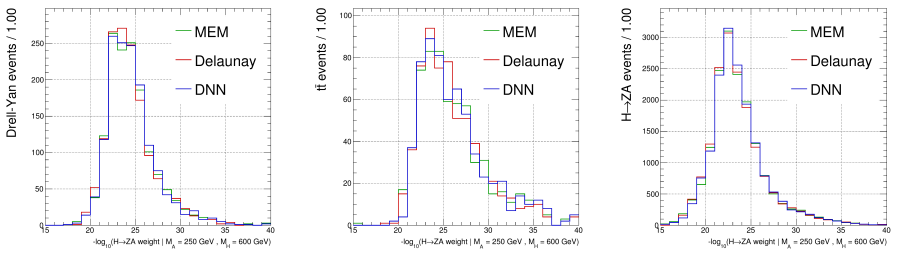
(a) *(b)* *(c)*

Figure 3.12 | Event information of events from the three different processes, DY (left), $t\bar{t}$ (middle) and $H \to ZA$ (right) events, evaluated on the $H \to ZA$ hypothesis at the specific mass point of $m_A = 400$ GeV and $m_H = 800$ GeV.

easily matched to the top quark resonance and therefore obtain a higher weight, while heavy resonances produce the second peak. The computation of the weights under $H \to ZA$ hypothesis is more complex from the point of view of the integration because of the multiple resonances in the decay chain of $H \to ZA \to \ell\ell bb$. While the delta function of the energy-momentum conservation allowed in the removal of the neutrinos or b-jets degrees of freedom in the other weight integrations, this is not enough simplifications for the $H \to ZA$ hypothesis and the computation time can rise to several minutes per event.

One key feature of the parametric DNN used to regress the MEM weight is its ability to provide the weight at mass configurations not seen during the training. Provided the requested parameters used do not exit the interpolation region defined by the parameter training space, since DNNs are known to perform poorly in extrapolatio, they could be obtained from the regression without the need to perform the MEM integration any more. This was confirmed by computing the weights of events at a

new mass point on Fig. 3.13, and comparing the values obtained with the regressive DNN already trained on the other 23 points. In addition, the piece-wise linear interpolation using the Delaunay triangle method was shown as comparison, also using as input the other 23 weights of the event. The good agreement, especially in terms of the position of the peak and to a lower extent the description of the distribution tails, illustrates *a posteriori* that the smoothness assumption of the MEM weights in the parameter space is valid. However, while this interpolation could be used without the hybrid method presented in this section, any new event would be required to pass by several MEM integration steps again, the DNN on the other hand can be used on any new event and completely short-circuit the need for MoMEMta onwards.



(a) *(b)* *(c)*

Figure 3.13 | Event information of events from the three different processes, DY (left), $t\bar{t}$ (middle) and $H \to ZA$ (right) events, evaluated on the $H \to ZA$ hypothesis at the specific mass point of $m_A = 250$ GeV and $m_H = 600$ GeV. This mass configuration was not seen during the training, thereby the regression from the DNN (in blue) results from its interpolation capabilities, compared to the true MEM weights computed separately with MoMEMta (green). In addition, the Delaunay triangulation (red) using the previously computed points is shown for comparison.

3.3.3 Applications

The DNNs described in Section 3.3.2 show good performance in the prediction accuracy of the weights computed by MoMEMta, yet the impact of the regression errors has not been quantified. In addition, MEM weights are rarely used directly in a physics analysis, and the effect of the prediction errors have to be propagated through their application. This section addresses several of them used throughout the literature.

Discrimination

A common use of the MEM consists in their use as a discriminating variable to either perform a cut-and-count or a shape analysis. An analytical discriminant between two hypotheses α and β can be defined as

$$\mathcal{D}(x) = \frac{P(x|\alpha)}{P(x|\alpha) + P(x|\beta)} = \frac{W(x|\alpha)}{W(x|\alpha) + \gamma W(x|\beta)} \text{ where } \gamma = \frac{\sigma_\beta^{vis}}{\sigma_\alpha^{vis}}. \quad (3.35)$$

This definition follows Eq. (3.17) and should therefore provide the best classification test. While in practice the visible cross-section should be used for both hypotheses in the γ factor, its effect is similar to the application of a monotonous function that does not alter the performance of the analytic discriminant, as described in Section 3.1.1. For convenience, it has been set to one.

As case study α and β will be set to the $t\bar{t}$ and DY hypotheses. The evaluation of the performance of the discriminant will be done through the ROC curve, on Fig. 3.14. The similar performance shows that the regression errors do not have a significant impact. The slight advantage of the DNN resides in the lower proportions of outliers compared to the MEM.

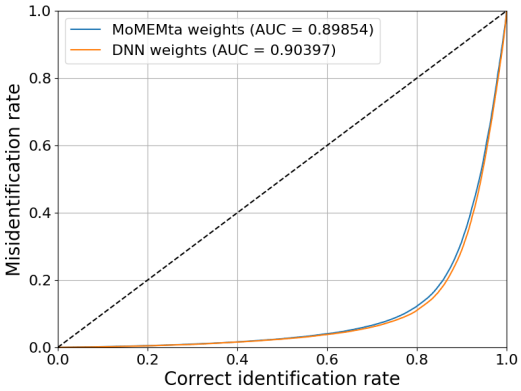


Figure 3.14 | ROC curve of the discriminant build from Eq. (3.35) applied on $t\bar{t}$ and DY hypotheses, as a function of the correct identification and misidentification rates of the $t\bar{t}$ events. The comparison is done between the discriminant built from weights computed by MoMEMta (blue) and the regressive DNNs (red), associated Area Under Curves (AUCs) are shown in the legend.

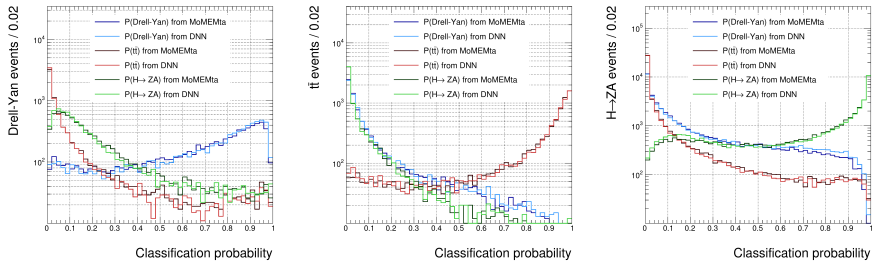
Theoretically, a DNN classifier trained to discriminate the two processes of Fig. 3.14 could produce a boundary decision that is close to Eq. (3.35), it could not however outperform it based on the discussion of Section 3.1.1 and the Neaman-Pearson lemma without including additional information that is not already contained within the MEM weights. The case of multiple hypotheses is however more complicated from the statistical point of view. In that case, it might be worth to bypass the com-

plication by using a multiclassification DNN to approximate the decision boundary. This training was solely based on the MEM weights obtained from MoMEMta. The training is done through a categorical cross-entropy loss function to maximise the probability of correct identification in each of the three output nodes, corresponding to the hypotheses. After the training, it can be evaluated on weights produced by the regressive DNNs to quantify the potential performance loss.

Two types of classifier can be defined, a *global* classifier that is trained to classify regardless of the parameters and a *parametric* classifier that is trained with the knowledge of the parameters. While the former can allow the quantification of an excess over the whole mass plane, without the possibility to pinpoint its location, the latter can be evaluated for specific points and should yield better performance, albeit its need for the look-elsewhere effect correction.

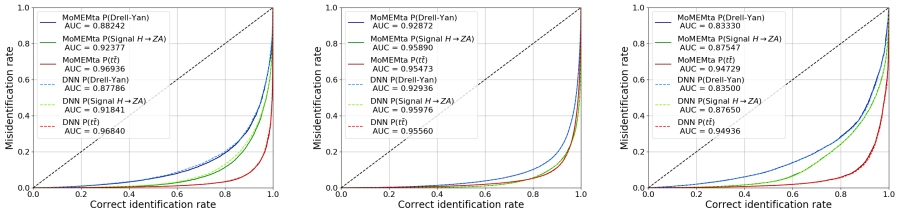
Both classifiers are provided with the DY and $t\bar{t}$ weights, only the treatment with the $H \to ZA$ weights differ. The global classifier receives the 23 weights at the different mass points for each event. This coverage should be enough that if the resonance was to exist in the area defined by these points, at least one of them would be sensitive enough to provide discrimination power. The parametric classifier is trained with a single $H \to ZA$ weight and given its associated masses m_A and m_H parameters, repeated for each mass point, effectively multiplying the training set by the number of mass points available.

The classification probabilities distributions obtained from the global classifier are on Fig. 3.15, showing very good agreement between the predictions from the weights of MoMEMta and its equivalent with the regressive DNNs. The associated ROC curves are on Fig. 3.16a, while the ones of the parametric classifier, averaged over all mass points, are on Fig. 3.16b. As comparison, the same exercise as with the discriminant on Eq. (3.35) was performed using only the $t\bar{t}$ and DY weights, but this time with the DNN classification. While the performance of the $H \to ZA$ classification as expected decreases, it is nonetheless unexpectedly reasonable given that the weights used for that process are suboptimal. Similarly, the comparison of Fig. 3.16c and Fig. 3.16a shows that introducing $H \to ZA$ weights in the classification helps the discrimination of the two other processes, confirming that weights obtained from another hypothesis can still provide discrimination power, even though in theory suboptimal. The value of the AUC for the parametric classifier as a function of the mass parameters is on Fig. 3.17, illustrating no bias in the performance of the classifier when using weights from either MoMEMta or the DNNs. The lower performance in the low mass region can be explained by the kinematic of the final-states being more compatible with the $t\bar{t}$ and DY hypotheses, and weights which are much closer in values, resulting in lower discriminating power.



(a) *(b)* *(c)*

Figure 3.15 | Classification probabilities using the global classifier for each hypothesis from the three different processes, DY (left), $t\bar{t}$ (middle) and H $\rightarrow$ ZA (right) events.



(a) Global classifier *(b) Parametric classifier* *(c) $t\bar{t}$ and DY weights only*

Figure 3.16 | ROC curve illustrating the identification performance of the three hypotheses as three separate binary classifications (one for each output node), using both the weights from MoMEMta (solid line) and the ones obtained by DNN regression (dotted line), for the global classifier (left), parametric classifier (middle) and a simpler classifier using only the $t\bar{t}$ and DY weights (right).

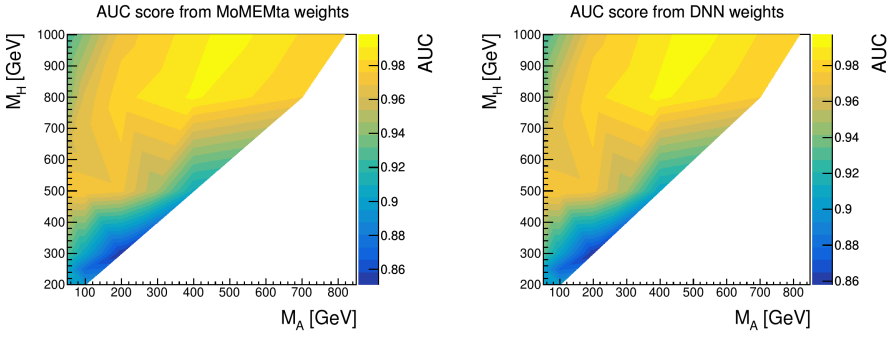


Figure 3.17 | Area Under Curve (AUC) of the parametric classifier in the parameter mass plane for the weights obtained from MoMEMta (left) and the regressive DNNs (right). The 23 mass points were used in this mapping, the interpolation was performed with the Delaunay piece-wise linear interpolation.

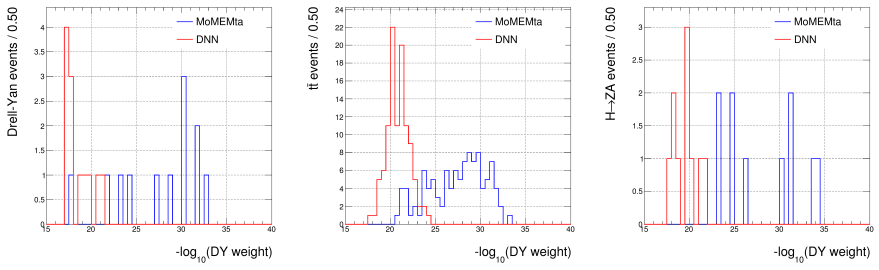
MEM integration failures

In some cases, the numerical MEM integration fails to produce a meaningful value before the threshold in the number of iterations or number of points has been reached, and the weights produced have infinitesimal values. They will be referred to as *invalid weights* in the following and were excluded from the previous studies, so they can be analysed in details in this section.

Slower convergence usually comes from a difficulty to adapt the integration grid assuming the hypothesis on which it is based on, and the associated changes of integration variables that have been selected. For example, using a change of variable over a resonance at a given mass – e.g. the top resonance in the $t\bar{t}$ process – for an event that does not exhibit such a resonance will result in a suboptimal integration grid and the convergence will be slower, if it happens at all. Two cases can however be distinguished : slightly more complicated event topologies compared to the hypothesis in the computation and for which looser thresholds would allow for a later convergence, and incompatible topologies that will not converge even when adding many more points and iterations. To disentangle the two cases, the invalid weights for each hypothesis were recomputed with more points and iterations. The invalid DY weights were rare, and all of them succeeded on the second more intensive attempt, they likely belong to the first category of weights that had barely missed the convergence before end of computation, probably because the DY hypothesis is permissive in terms of the integration grid. The $t\bar{t}$ hypothesis is harder to match and results in slightly higher rate of invalid weights, especially for signal events that fail about 20 % of the time, mostly in the high mass region. In addition, a significant portion of those weights did not converge in the second step either, they can therefore not be investigated further. This is even more the case in the $H \rightarrow ZA$ hypothesis that has more intricate topology from the integration perspective, although the second computation as not been performed due to the variety of parameters that would have to go through it. The rest of the section focuses on the DY and $t\bar{t}$ invalid weights. A summary of the numbers of weights failing the computation is in Ref. [1].

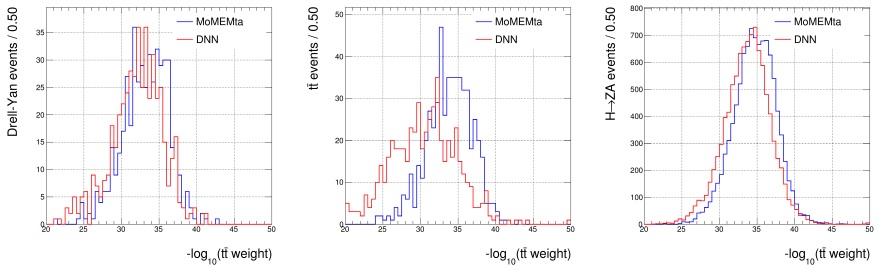
The comparison between what was computed by MoMEMta in the second iteration and the prediction by the regressive DNN is on Figs. 3.18 and 3.19. The DNNs tend to predict significantly higher values than what was obtained with the classic integration, especially for the DY hypothesis. For those weights, when compared to the bulk of the distribution of Fig. 3.10a, the DNN does not predict abnormal weight values, so it is possible that these events weights are underestimated in the phase-space integration. On the other hand, the DNN predictions of the $t\bar{t}$ weights is much lower than on Fig. 3.11b, and coincides more with the MEM prediction. For these events, it is more likely that exotic kinematic features are present, which would both explain the smaller weight and the more difficult integration convergence. Given this non-convergence there is however no way to know at this point whether

the result obtained by MoMEMta can be trusted, as some parts of the phase-space might be underpopulated, and whether these events have entered the extrapolation region of the DNN.



(a) *(b)* *(c)*

Figure 3.18 | Event information of events from the three different processes, DY (left), $t\bar{t}$ (middle) and $H \rightarrow ZA$ (right) events, evaluated on the DY hypothesis for invalid weights that failed the first computation but succeeded in the more intensive second attempt (blue) and their reconstruction using the regressive DNN.



(a) *(b)* *(c)*

Figure 3.19 | Event information of events from the three different processes, DY (left), $t\bar{t}$ (middle) and $H \rightarrow ZA$ (right) events, evaluated on the $t\bar{t}$ hypothesis for invalid weights that failed the first computation but succeeded in the more intensive second attempt (blue) and their reconstruction using the regressive DNN.

Regardless, these invalid weights pose a major concern for real-life analyses. First because any failure of computation requires specific error handling in the analysis workflow. Then because their treatment downstream is not clear, especially considering the simulation versus data agreement that could be impacted. Finally, their ill-defined values could skew the discrimination performance. The DNNs however do not suffer from convergence issues, and can provide a weight for every event in the same amount of time. While the validity of the predictions over invalid weights is still in question, their impact on the application of the weights can still be

quantified. This was studied in the context of the simple discriminant of Eq. (3.35), its ROC curves for both the DY and $t\bar{t}$ hypotheses are on Fig. 3.20. The very small values of the invalid weights even after the second computation step produce a very poor classification, while the more regular values produced by the DNNs produce more well-behaved discriminant values that yield better performance. Even if the performance is poorer compared to Fig. 3.14, their subdominant contribution in the whole dataset should not have much impact and illustrates another practical advantage of this hybrid method in an analysis workflow.

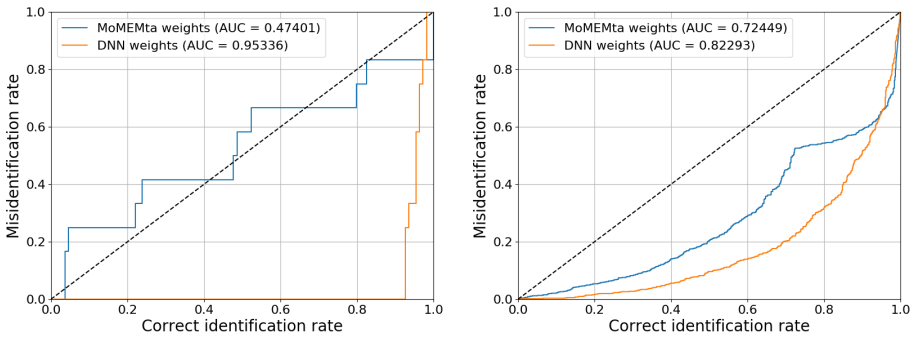
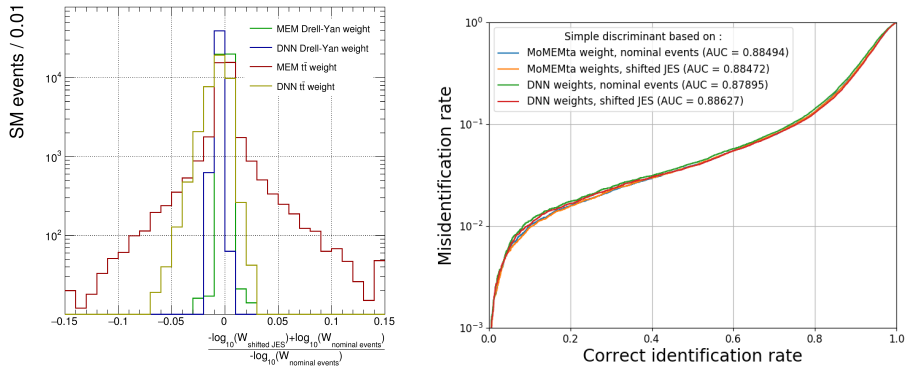


Figure 3.20 | ROC curves of the discriminant of Eq. (3.35) for the correct identification of $t\bar{t}$ events over their misidentification as DY events, in cases where the DY weights are invalid (left) and when the $t\bar{t}$ ones are (right). In blue the curve of the discriminant built from weights extracted from MoMEMta after a second more intensive computation step, and in orange the predicted value from the regressive DNNs.

Effect of nuisance parameters

Nuisance parameters (NPs) are a very important part in physics analysis for the role they play in the statistical model in Section 2.4.3. Their understanding and proper treatment is one of the driving factor in the sensitivity of an analysis. Some of these NP not only introduce a systematic uncertainty in the form of up- and down-shifted templates, but they also modify variables that are used in the statistical evaluation, for example JES. If the jets p_{T} is modified, then in principle the MEM weight has to computed as well for each variation, further increasing the number of calls to the MEM integration per event, which becomes very soon unpractical. On the other end, if the DNN is able to properly interpolate the MEM weight over the final-states momentum space, each variation could be treated directly without the need for new MEM integration.

This was tested by emulating a JES correction with an upward scaling of the two b-jets p_{T} by a factor 10 % – in itself, already a quite extreme variation. The MEM weights were computed with MoMEMta and the already trained DNNs, that have therefore not seen these *JES-shifted* events. This was only done for the DY and $\mathrm{t\bar{t}}$ hypotheses to avoid the complication of the parameterised $\mathrm{H} \rightarrow \mathrm{ZA}$ hypothesis. The comparison between the weights of the regular and JES-shifted event is on Fig. 3.21a, while the performance of the discriminant in Eq. (3.35) can be compared in the ROC curves on Fig. 3.21b. The difference in the weight computed with and without this artificial JES correction can be interpreted as a systematic template for the likelihood fit. This systematic shift does not coincide perfectly between what MoMEMta and the DNNs predict, the latter tends to underestimate the systematic uncertainty that would be attached to this variation. This may be related to the large up-scaling of the b-jets p_{T} that moved the events to unseen regions of the phase-space during the training. However, as no regression bias was observed in the predictions, the effect is small and could just be calibrated based on simulations, for example using Fig. 3.21a directly. A concern would be if this nuisance shift introduced a large difference in selection efficiency. This is not observed on Fig. 3.21b given the close performance of the ROC curves when using either MoMEMta or the DNNs to produce both the regular weight and the weight for the JES-shifted event. Although not a complete validation, there is evidence to support the robustness of the DNN regression approach with respect to nuisance parameters.



(a) *(b)*

Figure 3.21 | Comparison of the effect on a simplified JES correction of the two b-jets in the events, by scaling their p_{T} by 10 % upwards. Left : relative difference in the weight computed for the nominal event with the one computed for the JES-shifted event, using both MoMEMta and the DNNs for the DY and $\mathrm{t\bar{t}}$ hypotheses. Right : ROC curve of the discriminant of Eq. (3.35) applied on the to sets of regular and JES-shifted events.

Likelihood scan

The MEM probability can be used as a likelihood

$$\mathcal{L}(x|\alpha) = \prod_{i=1}^{n} P(x_i|\alpha), \tag{3.36}$$

where the product runs over the n measured events. This likelihood will be maximal around the parameter α that is present in events, in the H $\rightarrow$ ZA case the two-dimensional parameter space of m_{A} and m_{H}. If a single event is observed, the width of the likelihood will correspond to the experimental resolution of the resonances, and decreases as events accumulate.

For simulated events, the log-likelihood that will be considered is

$$-\log(L(x|\alpha)) = \frac{1}{n}\sum_{i=1}^{n} -\log(W(x_i|\alpha)) + \log\Big(\sigma_{\alpha}^{vis}\Big), \tag{3.37}$$

where the average over the n simulated events is to emulate the resolution observed for one single H $\rightarrow$ ZA event without any background. In Eq. (3.37) the factor $-\log(W(x_i|\alpha))$ corresponds to the event information that is directly regressed by the DNN, while the normalization from the visible cross-section needs special care. In theory, it encompasses the acceptance, cross-sections and all related BRs associated to the final-state of interest here

$$\sigma_{\alpha}^{vis} = \langle\epsilon\rangle \times \sigma(pp \rightarrow H) \times \mathsf{BR}(\mathsf{H} \rightarrow \mathsf{ZA}) \times \mathsf{BR}(A \rightarrow b\bar{b}) \times \mathsf{BR}(Z \rightarrow l^{+}l^{-}). \tag{3.38}$$

The acceptance $\langle\epsilon\rangle$ can be computed from simulations, the BR BR(Z $\rightarrow \ell^{+}\ell^{-}$) taken from the SM values, while the other values are the expected 2HDM values extracted from the theory. As the total cross-section and BR of A $\rightarrow$ bb are very model dependent they have been ignored for the purpose of this section, on the other hand the BR of H $\rightarrow$ ZA is mostly kinematic and has been kept. The transition region at $m_{\mathsf{H}} = 2 \times m_{\mathsf{A}}$ where the other decay channel H $\rightarrow$ AA becomes dominant can however be observed.

The likelihood scans in two dimensions are on Fig. 3.22. The elliptical shape of the likelihood is expected as the two resonances are correlated through the di-jet system. The profile likelihoods on Fig. 3.23 show similar values for the resolution than what is expected from the experimental point of view. While this represents one of the powerful application of the MEM, the point emphasized in this section resides in the observation that the likelihood scans would not have been possible with MoMEMta, as each event would have had to be computed for each point in this two-dimensional parameter space, which quickly becomes prohibitive. In general, parameter scans become increasingly expensive, scaling linearly with the requested granularity, and exponentially with the dimension of the phase space. The very fast computation time of the parametric DNN and its ability to interpolate in the

parameter space allows this hybrid method to perform parameter scans that would otherwise be highly unpractical if not impossible from a real-life analysis standpoint.

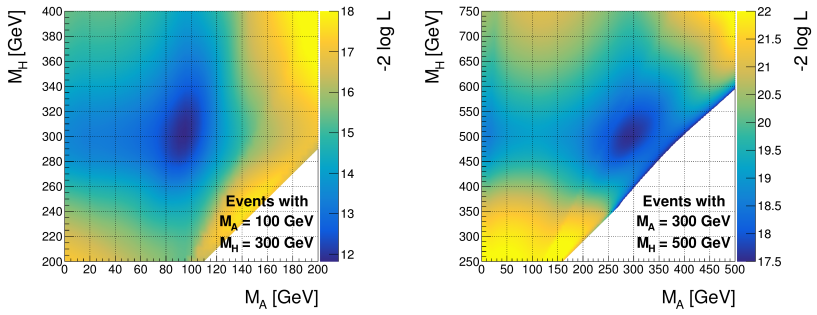


Figure 3.22 | Log-likelihood scan in the (m_{A},m_{H}) parameter space for two sets of H $\rightarrow$ ZA events.

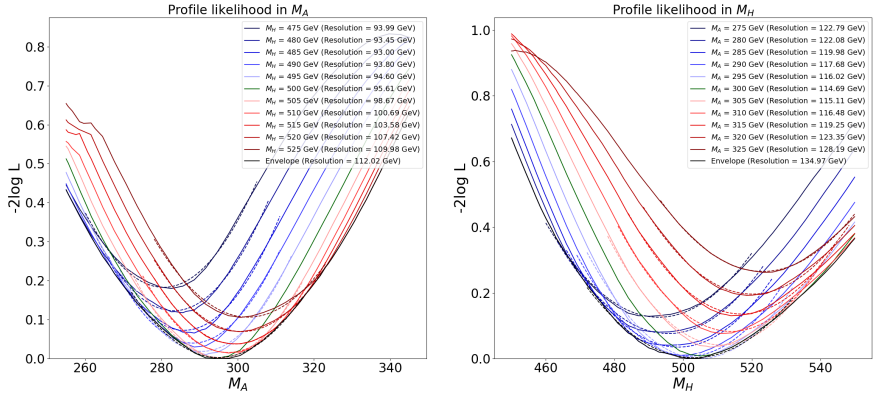


Figure 3.23 | Profile log-likelihood of resonant events with $m_{\mathrm{A}} = 300$ GeV and $m_{\mathrm{H}} = 500$ GeV, with minimum shifted to zero. Each profile in one direction has been produced for several values of the other parameter : the value of the parameter in event (green) and up (red) and down (blue) variation along this parameter are shown, resulting in an envelope that defines the global elliptical profile. The dotted lines represent the quadratic fit performed to obtain the value of the resolution expressed in the legend for each curve.

3.3.4 Real-life analysis comparison

The procedure followed in the CMS H $\rightarrow$ ZA $\rightarrow \ell^+\ell^-\mathrm{b\bar{b}}$ analysis of Ref. [289] consists in the reconstruction of both A and H bosons through the two- and four-body invariant masses m_{jj} and m_{lljj}. The signal region defined by these two quantities is positively correlated through the di-jet pair and is parameterised as an ellipse, the centre of which is defined by one-dimensional fits of the two distributions

assuming Gaussian shapes. The tilt and sizes of the axes are determined from the diagonalisation of the covariance matrix in the mass plane. The simulations and data are then binned as a function of the size ellipse, as depicted on Fig. 3.24.

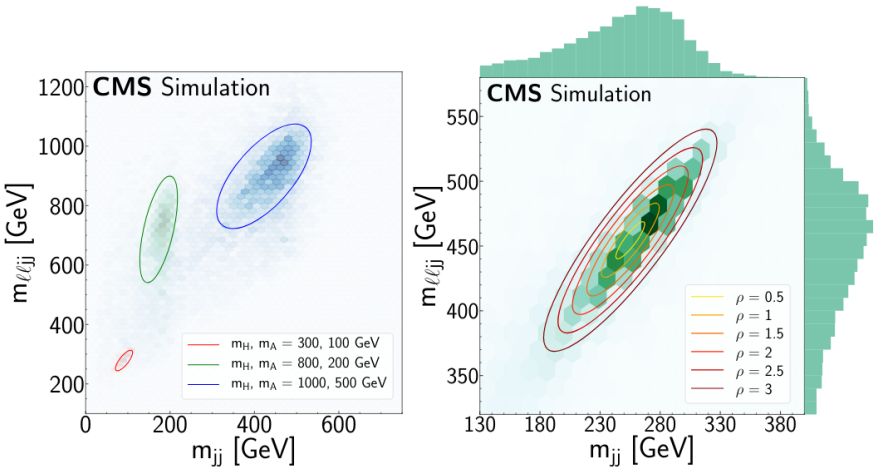


Figure 3.24 | Signal distribution in the mass plane (m_{jj}, m_{lljj}) for different mass configurations (left) and a specific case at m_A = 300 GeV and m_H = 500 GeV, for which several contours have been laid out parameterised by ρ, to be compared with the two-dimensional signal shape. From Ref. [289].

In this analysis 207 points were used over the whole (m_A,m_H) plane. To reproduce the results using the hybrid method presented in the previous sections, the same set of points will be used in this section and their weights produced using the DNNs trained in Section 3.3.2. Since these are new events, and no re-training has been performed, the whole dataset can be used without fear of overfitting and the new parameters should remain within the interpolation region of the H $\rightarrow$ ZA DNN.

The three different hypotheses weights were reproduced by the DNNs for each event, and propagated through the classifiers described in Section 3.3.3, both the global and parametric. The ellipse selection efficiencies were computed for different values of the size parameter ρ and displayed in the ROC curves on Fig. 3.25, together with the ones of the global and parametric classifiers, the latter applied to the specific mass point on which the ellipses were drawn. The discrimination provided by the MEM did not outperform the ellipse method, as expected given its very good optimization. Only at high mass only does the parametric classifier slightly outperform the ellipse method, as the underlying Gaussian assumptions break down.

On the other hand, once the ellipse selection has been performed, the events that are still selected can go through both classifiers. Although the combination of the ellipse and the global classifier is not successful at improving the efficiency, which is expected given that the global classifier is trained on the whole mass plane and not

specifically to this point. The combination of the ellipse and the parametric classifier can yield up to one order of magnitude background contamination reduction for practically the same signal efficiency, especially at high mass and in the boosted region where the ellipses are suboptimal, but even to a smaller extent in the low mass region.

The main point of this section however is not the marginal gain obtained, even though it illustrates that there is some motivation to look into the MEM even for already well-defined analyses, but rather the fact that this application to a real-life CMS analysis would not have been possible at all with MoMEMta. Given the average computation times in Table 3.1, it has been estimated that around 1450 (3050) CPU years would have been needed to produce the results of Fig. 3.25 using the global (parametric) classifier, the detailed summary is in Ref. [1]. The DNNs on the other end can produce weights in about 150 µs when using vectorisation in large enough batches, necessitating about 10 hours of pure computing, notwithstanding the processing of the data, RAM allocation and I/O data streams. Adding the time to pass through the classifiers and the ellipse method, on a cluster of CPUs it can take less than a day. The pure weight computation has been reduced by six orders of magnitude.

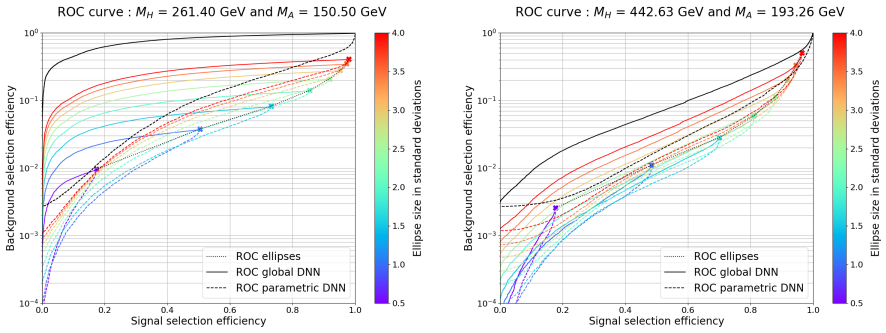


Figure 3.25 | ROC curves for two different mass points including the working points of the ellipse at different sizes given by the colour scale, the global (solid line) and parametric (dotted line) classifiers, together with the combination of the ellipses and each classifier in coloured lines.

3.3.5 Future improvements

The method developed in this chapter, although promising, is relatively simple and was only applied to a case study, with several simplifications and not tackling the statistical interpretation. The next step would be to apply it to a complete CMS analysis, with all its complexity, and see its effect on the final fit. An example would be the HH $\rightarrow$ bbWW analysis of the next chapter, although it was not feasible

Table 3.1 | Average computation time in MoMEMta for each event and hypotheses. DY and $t\bar{t}$ weights peak around the average, while the $H \to ZA$ weights have a broader distribution and it is not uncommon to have a computation time up to two or three times the average value. The size of the each samples is shown, for $H \to ZA$ events it consists in the 207 mass points.

	DY events	$t\bar{t}$ events	$H \to ZA$ events
DY weights	$3.6\,\mathrm{s}$ / event	$3.8\,\mathrm{s}$ / event	$4\,\mathrm{s}$ / event
$t\bar{t}$ weights	$12\,\mathrm{s}$ / event	$10\,\mathrm{s}$/ event	$20\,\mathrm{s}$ / event
$H \to ZA$ wiights	$\sim 600\,\mathrm{s}$ / event	$\sim 600\,\mathrm{s}$ / event	$\sim 600\,\mathrm{s}$ / event
Sample size	280 K events	460 K events	2.5 M events

given the time constraints of the thesis. Regardless, several improvements can be considered, in a few different areas : the integration itself (both its components and method), how the MEM can be used in an analysis, and finally the ML regression. Although only the latter resulted from genuine work in this thesis, any improvements in the previous areas will likely have to be propagated to it. A list of potential challenges is enumerated in Refs. [295–297], a more in depth perspective will be given in the following sections.

Components of the method

There are three components that are being integrated over in Eq. (3.29). The PDFs are provided externally from LHAPDF, which is mature enough for major changes to be unlikely. The TF represents an approximation of the full detector response to the final-state, which is needed because of the intractability of the different steps that happen before the reconstructed particles are recorded : the parton shower, the hadronisation and detector effects. This approximation is usually parameterised from simulations as they have access to the truth-level information, and can either be embedded in a continuous function or discretised in a binned histogram. There is no major breakthrough that can be brought to this method, however there might be developments in the future that would circumvent the need for this representation. Probabilistic programming applied to HEP [298, 299] promises the generation of numbers from complex distributions, as is typically used in event generation (cf. Section 1.3), together with the conditional probability $p(x|z)$ of an observed variable x as a function of the latent space variable z that is tracked down from all the branching. There is currently no such alternative for Geant4 or Pythia, to the knowledge of the author of this document, but there is an ongoing work in that direction, most likely towards frameworks that embed those simulators without the need to rewrite them. If they become widespread, then technically

part of or the whole TF could be replaced by the full now tractable likelihood. An alternative for the specific case of the parton shower lies in the *shower/event reconstruction* techniques [300–303] which use approximate clustering methods to assign a likelihood ratio between two hypotheses based on the jet substructure. This approximately tractable probability could replace part of the parton shower approximation in the TF, although this would mean that the latter will be partitioned, which might not be easily doable.

Another limitation of the method lies in its ME that is currently limited at LO in MoMEMta. This severely impedes the method when precision is required, typically the level that is reached at the LHC. However, moving towards NLO is not so straightforward for the MEM, as the calculations include loops and additional radiations. Therefore, the clear mapping between an event containing several well-defined objects that was present at LO disappears. One of the main change in the NLO ME comes from the presence of additional partons producing jets. There are several ways to deal with these additional jets. The easiest is to apply tight selections to remove these extra radiation jets, and the LO method can be used, at the cost of losing a significant part of the signal efficiency. Another way is to use a ME that includes additional partons in the final-state. This method is more accurate, but the more complicated structure of the ME and the combinatorial challenge make the computation time even more expensive and unpractical.

The most significant effect of the ISR is to change the global transverse momentum balance of the event, the sequence of the radiations having no impact. Therefore, to keep the same integration as the LO case, it is kinematically enough to perform the boost of the final-state (or the incoming partons, as the ME is a Lorentz invariant) based on the total ISR to balance the total momentum. In practice, since the longitudinal component is integrated over, this only needs be performed in the transverse plane. While not including the QCD vertices and propagators, this serves as a quick solution to the extra radiation ISR problem [304, 305]. An improvement consists in including the Sudakov factor (c.f. Section 1.3) to take into account the probability of branching $j \to i + X$. Because several branching can occur, it is usually enough to assume that a single branching is responsible for most of the p_{T} of the total ISR. This improvement typically decreases the potential bias in the parameter estimation. An alternative consist in including an additional TF related to the transverse momentum of the incident particles, even when the ISR falls outside the detector acceptance, and integrate over the transverse momentum of each of the two initial parton legs. This increases the number of integration variables, with a larger statistical uncertainty for the final integral, largely compensated by the reduction of the systematic uncertainty from QCD effects. It is worth pointing out that MoMEMta allows the use of the ISR as an additional branch in the change of variable to include this non-zero transverse sum of momentum. This was actually done in the computation of the DY weights in Figs. 3.10a to 3.10c, since ISR

represents one source of the candidate b-jets.

A more systematic approach at extending the MEM to NLO is complex due to the fact that the two contributions (virtual and real emission) exists in a different phase-space. They are only considered together through a subtraction scheme. While the virtual correction lies in the same phase-space as the Born contribution of the LO case, the real emission consists in an additional radiation. This radiation can be integrated out, while maintaining the kinematics of the Born contribution, which is important for the change of variables necessary for the integral to converge. The subtraction scheme must also be taken into account in the MEM, as well as the aforementioned boost [306–310]. This method is however not implemented in MoMEMta, or any other library for MEM weight computations (to date, and to the knowledge of the author of this document).

As a sidenote, during the development of the method in Section 3.3, it was observed that the ME represented a significant and often times dominant contribution to the overall computation time. An attempt was therefore made to replace the ME by a surrogate DNN in the same spirit as the weight regression. If successful, then this surrogate could be used throughout the integration to accelerate the computation, while allowing the possibility to modify the TF at any time. As the PDFs are also independent of the TF they were also included in the target. During the integration of several events of a $t\bar{t}$ process, the integrand values were saved and used to train a DNN. Despite a relatively good agreement (illustrated on Fig. 3.26), there is a significant proportion of outliers. It was observed that replacing the ME by this surrogate increased the convergence time, the outliers potentially creating "fake" peaks. The computation time gain was therefore not deemed sufficient. This regression was also attempted in Refs. [311–313] for various and sometimes more complicated topologies, with various degrees of success. If accurate enough, this kind of fast surrogate ME could be used in the MEM.

Integration techniques

Several adaptive numerical integration algorithms are included within the Cuba [314] library that MoMEMta is based on. Suave uses importance sampling like Vegas, with the subregion sampling. This technique provides a better description when the substructures are not aligned with the coordinate axes, by slicing the region into two, then iterating cycles of sampling and subdivision on the region with the highest error. Divonne is based on a pure stratified sampling. The objective of the algorithm is to divide the whole integration phase-space into regions that contain the same "spread" value between the maximum and minimum of the function, using numerical optimisation methods. While more powerful for some integration cases, if narrow peaks are present the performance can decline, although some hints

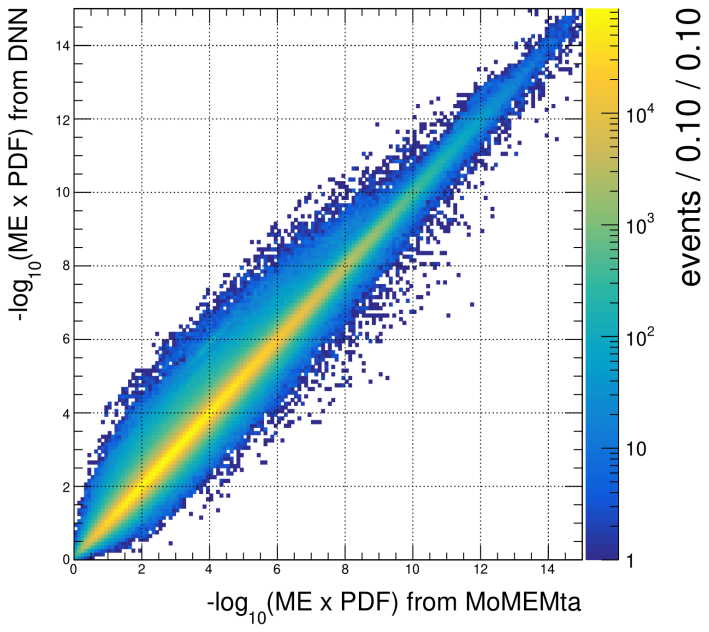


Figure 3.26 | Regression of the ME multiplied by the PDF of the $t\bar{t}$ fully leptonic process, compared between the values used in MoMEMta and a regressive DNN.

can be passed to the algorithm about the location of such peaks, if known from analytical considerations. Finally, Cuhre differs as it is not technically a Monte Carlo generator, rather a deterministic one. A *cubature* rule, that defines how the sampled points evaluated by the function are summed, with a polynomial shape is used to divide the phase-space into subregions, the axis and region with the highest error are selected at each step. This algorithm is particularly useful in a moderate number of dimensions and if the integrand is suited for polynomial interpretation. In the study of Section 3.3, only Vegas was used, a more comprehensive future study should consider all options and potentially provide a comparison in terms of accuracy and computation time. A recent improvement of Vegas [315] implemented additional adaptive stratified sampling and could be worth trying. An alternative improvement consists in parallelising the sampling of numbers that are used to evaluate the integral on a GPU [316–319].

The main limitation of most Monte Carlo adaptive numerical integration methods as is that they rely on the factorisation assumption described in Section 3.2.2. This assumption breaks down for non-separable functions for which the precision is suboptimal, or in worse cases where the coverage is not sufficient and the integral value is wrong. If this assumption can be relaxed, then the sampling function could

better approximate the high-dimensional shape of the function to approximate, and potentially circumvent the requirement that each peak is mapped to a single variable of integration, although it would still help convergence. This requires a non-linear mapping of the phase-space, which is something for which ML methods are particularly suited. The methods described in the rest of this section fall into the *generative model* category of ML. The problem they aim to solve is to approximate an unknown probability density from which drawn samples are known. Typically, one of the first application was in image generation. Many methods exist, some pre-dating the ML developments, but in this section we will focus on the three main ML algorithms as deep learning allows for fast inference and can capture the highly non-linear behaviours. They will be detailed in the following paragraphs, and are summarised on Fig. 3.27.

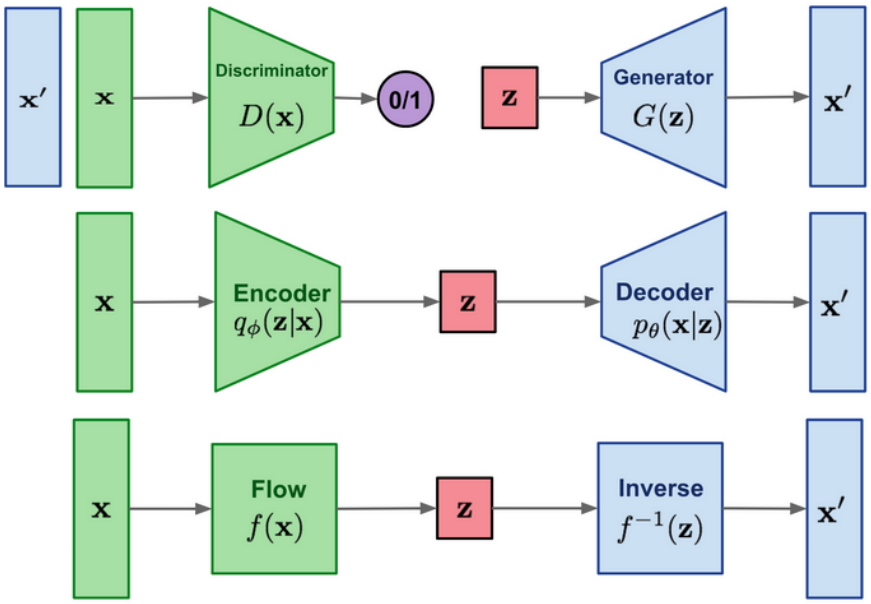


Figure 3.27 | Comparison of the three main ML algorithms : generative adversarial network (GAN) (top), variational auto-encoder (VAE) (middle) and normalising flow (NF) (bottom). From Ref. [320].

Generative adversarial network (GAN) [260] was one of the first algorithms dedicated to image generation. A GAN consists in pitting two NNs against each other, the *generator* takes as input random noise and is trained to reproduce a certain distribution, and the *discriminator* that tries to separate the "fake" predictions of the generator (in this case images) from samples that belong to the unknown distribution (in this case an image database). By training these two models sequentially, the generator is able to produce samples with very good fidelity, yet the density is only

implicit and they perform poorly to reproduce it (although variations exist for density estimation [321]). Another drawback is the *min-max* problem of two antagonistic NNs. The optimal point in the loss function space represents a saddle point that is not guaranteed to be achieved, for example if the discriminator becomes too performant, the generator cannot learn how to "fool" it. This is not desirable for a numerical integration, where failure to cover the tails of the distribution (e.g. $g < f$ in Eq. (3.33)) will result in an underestimation of the integral. Additionally, training both NNs is time-consuming. While still valuable for event generation in HEP [322–327], it becomes expensive for numerical integration, which limits its improvement over classical numerical integration methods.

Variational auto-encoders (VAEs) [328, 329] are another tool to approximate an intractable posterior distribution function. They are based on the concept of *autoencoder*, which is a particular example of a NN with hidden layers that have fewer neurons than the input and output layers. Typically, they are interpreted as an *encoder* that reduces the dimensionality of the problem into a latent space z, and a decoder that tries to reproduce the inputs x given that latent space. The training is then performed by trying to minimise a distance function between the input of the encoder and the output of the decoder. This type of unsupervised training is especially useful in the context of dimensionality or noise reduction, and in clustering tasks, among many other applications. Another application more relevant for data generation is through a probabilistic autoencoder, where the encoder and decoder are seen as probability functions $q_\phi(z|x)$ and $p_\theta(x|z)$. Once the auto-encoder is trained to reproduce samples, the decoder can be used separately to generate new samples of the posterior distribution by applying it on sampled points of the latent space, such that

$$p_\theta(x) = \int_z p_\theta(x|z)p_\theta(z)\ dz. \tag{3.39}$$

The problem with a classical autoencoder is that the latent space is not regular, i.e. that the structure of the phase-space is conserved. Two close points x_i and x_j must have nearby equivalents in the latent space, otherwise the sampling will not produce the correct coverage. To ensure that, VAEs include an additional term to the loss function that forces the distribution on the latent space to have a certain distribution, typically a diagonal Gaussian to ensure computational efficiency. To generate samples in the latent space, the decoder is used to generate mean and variance values of a Gaussian $z \sim \mathcal{N}((\mu_\phi(x), \sigma_\phi(x)))$. A problem however appears as the sampling means that the backpropagation is not possible any more. To overcome this, the *reparametrisation trick* consists in sampling a new number $\tilde{z} \sim \mathcal{N}(0,1)$ such that $z = \sigma_\phi(x) \times \tilde{z} + \mu_\phi(x) \sim \mathcal{N}((\mu_\phi(x), \sigma_\phi(x)))$, and the backpropagation can flow through μ and σ. During the training the additional term to the loss function that forces the latent space to have $z \sim \mathcal{N}(0,1)$ (a special case of regularisation) is given by the Kullback-Leibler (KL) divergence [330]

$D_{KL}(p_\phi(z|x), \mathcal{N}(0,1))$, where

$$D_{KL}(g(y), f(y)) = \int g(y) \log \frac{g(y)}{f(y)} \, dy. \tag{3.40}$$

There is now a trade-off between the auto-encoder goal of minimising the distance loss of the reconstruction of x, and the requirement that the latent space is distributed as a Gaussian. After the training, z can be sampled from the Gaussian distribution directly, and the decoder be used to generate samples. This kind of technique was notably used for event generation [331]. While VAEs are typically more stable in the training that GANs, the Gaussian posterior $q_\phi(x)$ assumption inevitably causes some loss in the density estimation, such that in numerical integration the sampling is suboptimal. The diagonal representation of the latent space can also cause poor sampling quality in regions of the phase-space.

Flow-based models [332], more often referred to as normalising flows (NFs), represent another way to use the latent space in order to sample a posterior distribution, focusing on the exact density estimation rather than generating synthetic data. In this context, a complex mapping can be defined between two n-dimensional spaces : the input space $\mathcal{I}$ (so far referred to as the latent space) from which points z are drawn according to a prior p_z (arbitrary, typically uniform or Gaussian), and a target space $\mathcal{T}$ on which the function is integrated using points x of the unknown density. The mapping h with set of weights w is such that $x_i = h_w(z_i)$ (in the following, the dependency on w will be implicit). The initial distribution p_z is known, as well as the Jacobian determinant of h, then p_X can be determined by the change of variable formula

$$p_X(x) = p_Z(z) \left| \det \left(\frac{\partial h^{-1}(x)}{\partial x} \right) \right| = p_Z(z) \left| \det \left(\frac{\partial h(z)}{\partial z} \right) \right|^{-1}. \tag{3.41}$$

For Eq. (3.41) to apply, h must be a function that is monotonous, so it keeps the ordering of points between the two spaces, and bijective, such that for every point in either space there is a unique equivalent in the other space. The last condition is required, so the inverse transformation h^{-1} exists, and the mapping can be done in both directions. Compared to the VAEs, h can be interpreted as the decoder, and h^{-1} the encoder. The advantage of NFs over VAEs reside in the exactness of Eq. (3.41) that does not produce a reconstruction loss. And since there is no need for a posterior $p(z|x)$, they completely avoid the problem of forcing the regularity of the VAE to avoid disjoint or overlapping regions of the phase-space. Both methods are stable and allow for diverse sampling compared to GANs, but with better quality sampling.

In Refs. [333, 334], a simple NN with different activation functions was used to represent the mapping $h : z \in \mathcal{I} \to x \in \mathcal{T}$, with reasonable performances. The idea was to train on the KL loss between the prediction of the DNN and the function

to integrate, from which a loss function can be derive. However, it was shown in Ref. [335] that the definition of the training on bounded spaces and through the activation functions mean that h is not surjective, i.e. that not all points in target space can be assigned a point in the input space if it is located far enough in the tails of the function. This non-perfect phase-space coverage means that the integral will be underestimated. Additionally, the evaluation of the Jacobian determinant in Eq. (3.41) grows as the cube of the dimension. As it has to be estimated for every point in the loss function, it becomes very expensive in high-dimension. NFs offer a solution to this problem by decomposing h into a series of less complicated bijective transformations $x = h_1 \circ ... \circ h_n(z)$, such that

$$p_X(x) = p_Z(z) \prod_{i=1}^{N} \left| \det \left(\frac{\partial h_i(z_i)}{\partial z_i} \right) \right|^{-1}, \tag{3.42}$$

where $z_i = h_{i-1}(z_{i-1})$ and $z_0 = z$. Additionally, the evaluation of the Jacobian determinant should be fast enough. Typically, a determinant scales as a cubic power of the dimension, and the aim is to reduce that to a linear scaling. In the initial proposition for NF [332], the transformation consisted in a single hidden layer NN returning $d_i(z_i) = u_i \phi(w_i^T z_i + b)$, where u_i are the weights output layer, ϕ, w_i and b the hidden layer activation function, weights, and bias. The invertible operation is then defined as $h_i(z_i) = z_i + d_i(z_i)$, such that

$$\begin{array}{rcl} \frac{\partial h_i(z_i)}{\partial z_i} & = & \mathbb{1} + u \left(\phi'(w_i^T z_i + b) w \right)^T, \\ \det \left(\frac{\partial h_i(z_i)}{\partial z_i} \right) & = & 1 + u^T \phi'(w_i^T z_i + b) w, \end{array} \tag{3.43}$$

where the matrix determinant lemma has been used in the second line. This determinant can be computed linearly and is therefore fast. A limiting factor however lies in the limitations of the families shallow NN for the transformation, which requires many of them in sequence to represent the high-dimensional dependencies. There are several alternatives that include deeper NN, but they all use the same trick to ensure that the determinant computes linearly, which is that the Jacobian matrix is made triangular, i.e. that $(\partial h / \partial z)_j^i = 0$ for $i \leq j$ (where i is the row and j the column). Therefore, the determinant must only be evaluated on the diagonal elements, the cost growing only linearly with the dimension, and not as a third power.

One way to guarantee this is through *autoregressive* models, that exploit the chain rule on a D-dimensional variable $z = (z_1, ..., z_D)$ that states that $p(z_1, ..., z_D) = p(z_D | z1, ..., z_{D-1})\ p(z1, ..., z_{D-1})$. Applied recursively, that rule yields

$$p_Z(z) = \prod_{i=i}^{D} p_{Z_i | Z_{<i}}(z_i | z_{<i}), \tag{3.44}$$

where the subscript $_{<i}$ refers to all the variables with index less than i, i.e. $z_{<d} =$

$(z_1, ..., z_{d-1})$, and $p_{Z_1|Z_{<1}}(z_1|z_{<1}) = p_{Z_1}(z_1)$. If a model can be defined that respects the property that the output h_d only depends on the previous $z_{<d}$ and not the others, then the Jacobian matrix associated to each output must be triangular, as $\partial h_d/\partial z_{i>d} = 0$. In MADE [336] an autoencoder is trained classically, but with masks on both the encoder and decoder to ensure this ordered dependency, this is usually referred to as an *autoregressive model*. This kind of architecture is much deeper than IN Eq. (3.43), and allows the representation of higher-dimensional latent space. Alternatives exist, although more focused to specific generation, such as PixelCNN [337] and WaveNet [338] for image and audio generation respectively. The problem of MADE is that the choice of the ordering matters, two different orderings might represent the density differently. The natural extensions to NFs, with a principle inspired from VAEs, are divided into two categories : Masked Autoregressive Flow (MAF) [339] and Inverse Autoregressive Flow (IAF) [340].

MAF is a model designed for quick density estimation. Each of the N autoregressive mode sequentially considered, following Eq. (3.42), is associated with a random vector $z^{(i)}$. In the first model, the vector is initialised with a Gaussian prior $z^{(1)} \sim \mathcal{N}(0, 1)$ which serves to determine the next vector $z^{(2)}$, until the target space $x = z^{(N+1)}$ is reached. For each model i, the forward pass produces $z^{(i+1)} = \{z_j^{(i+1)}\}$, determined by

$$z_j^{(i+1)} = z_j^{(i)} e^{\alpha_i} + \mu_i \text{ with } \mu_i = f_{\mu_i}^{(i)}\left(z_{<j}^{(i+1)}\right) \text{ and } \alpha_i = f_{\alpha_i}^{(i)}\left(z_{<j}^{(i+1)}\right), \quad (3.45)$$

where $f_{\mu_i}^{(i)}$ and $f_{\alpha_i}^{(i)}$ are both produced by an autoregressive model $f^{(i)}$, such that z_j depends on the previous components $z_{<j}$ only. This is similar to what was done in the reparametrisation trick of VAEs, except σ has been replaced by e^{α} to ensure strict positiveness. The advantage of that formulation is that in the inverse propagation

$$z_j^{(i)} = (z_j^{(i+1)} - \mu_i) e^{-\alpha_i} \text{ with } \mu_i = f_{\mu_i}^{(i)}\left(z_{<j}^{(i+1)}\right) \text{ and } \alpha_i = f_{\alpha_i}^{(i)}\left(z_{<j}^{(i+1)}\right), \quad (3.46)$$

there is no need to invert $f^{(i)}$. And thanks to its autoregressive form, the Jacobian of f^{-1} (the model index will be implicit here for clarity) is triangular and

$$\left|\det\left(\frac{\partial f^{-1}}{\partial z_j}\right)\right| = \exp\left(-\sum_j \alpha_j\right). \quad (3.47)$$

It is important to note that in the forward pass in Eq. (3.45), the components of $z^{(i+1)}$ need to be computed sequentially, with cost of D operations. MAF is therefore not suited for sampling the posterior $p_X(x)$. On the other hand, the backward pass in Eq. (3.46) can be performed in parallel, and MAF provides a fast density estimation of $p_X(x)$.

IAF was designed with the opposite goal in mind (and actually predates MAF),

which is to provide a fast sampling procedure. The method differ slightly, in the sense that the definition of $z^{(i+1)}$ is based on its previous value, such that the forward and backward passes are respectively

$$\begin{aligned} z_j^{(i+1)} &= z_j^{(i)} e^{\alpha_i} + \mu_i \\ z_j^{(i)} &= (z_j^{(i+1)} - \mu_i) e^{-\alpha_i} \\ \text{with } \mu_i = f_{\mu_i}^{(i)}\left(z_{<j}^{(i)}\right) &\text{ and } \alpha_i = f_{\alpha_i}^{(i)}\left(z_{<j}^{(i)}\right). \end{aligned} \tag{3.48}$$

In this case, the sampling can be done in one pass, while the density estimation through the inverse needs to be done sequentially. For numerical stability, in Ref. [340], the authors have included a sigmoid updating to avoid the exponential growth of the parameters during the initial cycles of the training. MAF and IAF are therefore two similar method with opposite goals (as illustrated on Fig. 3.28), and have been used for event generation in HEP [341]. However, for numerical integration, both the sampling (to evaluate the integral) and the inference of the density (to train the KL divergence in the loss function) need to be fast.

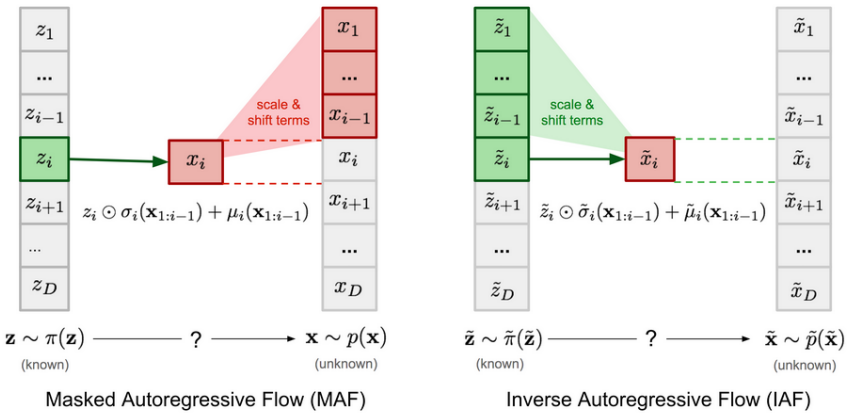


Figure 3.28 | Comparison of the flow between a MAF (left) and IAF, in Eqs. (3.45), (3.46) and (3.48) σ was replaced by e^{α} to ensure positiveness. From Ref. [320].

A parallel endeavour is through the use of *coupling layers*, that was first proposed in Refs. [342, 343], and illustrated on Fig. 3.29. The combination of several coupling layers is sometimes referred to as an Invertible Neural Network (INN) [344]. Each coupling layer partitions the input z into two disjoint partitions A and B. The operations applied to these partitions are

$$\begin{aligned} y^A &= x^A \\ y^B &= C(x^B; m(x^A)), \end{aligned} \tag{3.49}$$

where C is a separable and invertible map, and m any type of function. This ensures that the inverse of the coupling layer exists

$$\begin{aligned} x^A &= y^A \\ x^B &= C^{-1}(y^B; m(x^A)) = C^{-1}(y^B; m(y^A)). \end{aligned} \tag{3.50}$$

The advantage of that formulation is that the Jacobian is directly triangular and therefore easy to compute

$$\det\left(\frac{\partial y}{\partial x}\right) = \begin{vmatrix} I_d & 0 \\ \frac{\partial y^B}{\partial x^A} & \frac{\partial y^B}{\partial x^B} \end{vmatrix} = \frac{\partial y^B}{\partial x^B} = \frac{C}{\partial x^B}, \tag{3.51}$$

and as a bonus does not depend at all on m which can be as complex as possible, typically using a DNN. At first, C was represented as an affine function in Refs. [342, 343] (which has been demonstrated as a special case of both MAF and IAF [339]), a specific image-oriented inverse 1x1 convolution with Glow [345], then with more complicated splines [346, 347]. Multiple coupling layers can be concatenated with varying partitions to fully express the components dependency, as displayed on Fig. 3.29.

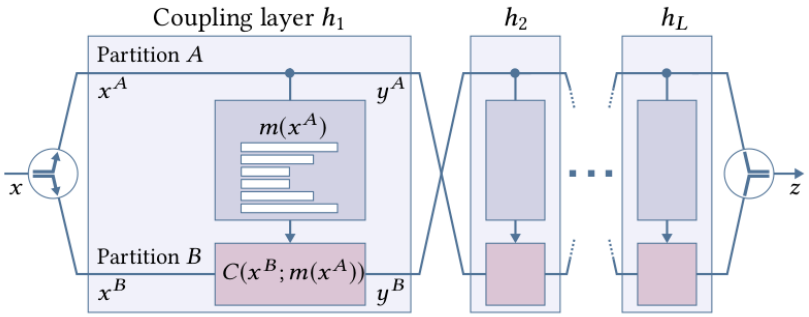


Figure 3.29 | Schematic representation of a coupling layer that splits the components into two partitions A and B, on which a separate map C and a complex NN is used for the connexion m. From Ref. [346].

The fact that this kind of normalising flow can be used both forward (for sampling) and backward (for inference) exactly and without heavy computation cost makes them perfectly suited for numerical integration, as was further demonstrated in Refs. [335, 348]. Applied to the MEM, these methods could improve the accuracy by providing a non-linear mapping that is more suited for the integration. A first attempt at normalising flows was made by the author of this document for regular functions, such as the "camel" function of Fig. 3.30 for the 2-dimensional case. This illustrates the ability of NFs to deal with correlated peaks, while VEGAS generates points in non-present peaks. The better coverage produces a smaller variance for

the same number of points, as validated for several correlated functions and higher dimensions.

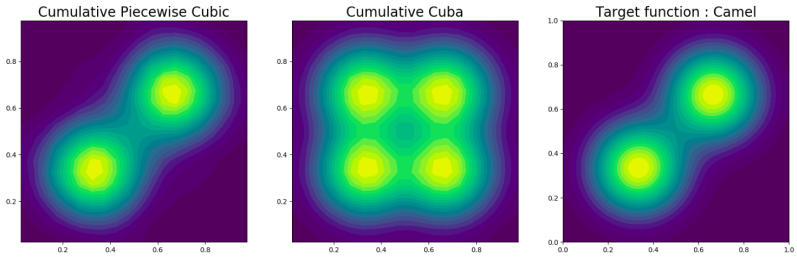


Figure 3.30 | Target "camel" function (right) composed of two Gaussian functions generated as $\mathcal{N}(1/3, 0.2)$ and $\mathcal{N}(2/3, 0.2)$, compared to the histogram of points generated by a NF, following a cubic spline of Ref. [348] (left) and the VEGAS algorithm from CUBA [314] (middle).

For completeness, let us mention one inherent problem of generative models. While they can provide very accurate samples, the stochastic process means that it is not possible to generate them for a more specific region. One example would be the request to generate images of specific objects from a model trained on a larger collection of images. One solution to that is similar to the concept of parametric NN, a specific label associated to the sample is attached to it during the training. After the training, this label can be used to steer the sample generation. In CGAN [349], both the discriminator and generator take as input this label, and in CVAE [350] both the encoder and decoder posterior distribution are now conditioned on this label. Similarly, in the context of NFs, CINN [351] use coupling layers whose NN coupling distribution is conditioned on the label, effectively transforming it into a parametric NN. When the condition is complex, can also be replaced by an intermediate representation, for example a pretrained NN as was used in Ref. [351] for greyscale image colourisation. Finally, while this section developed the three main algorithms as separate classes, it is worth mentioning that they can be combined. For example, Flow-GAN [352] represents an application of a GAN as a NF, a middle ground between the two methods where the generator uses a coupling layer model with affine function from Refs. [342, 343].

Application of the method

Many of the developments to the method by the Tevatron experiment [269–276] can be used to further improve the way the MEM was used in this chapter [305]. They are mostly related to the experimental interpretation, through the detector effects in $T(x|y)$ and the way the likelihood $P(w|\alpha)$ is applied.

The first would be the introduction of a parameter β in the TF $T(x|y;\beta)$, representing all the possible detector effects (for example the JES, as was done at the Tevatron) so they can be measured simultaneously with the theory parameters in $P(x;\alpha,\beta)$. This may impact the normalisation of the TF in Eq. (3.31), as now the equation must hold for any value of β, which can vary its shape. In case of selection-based normalisation scheme of the TF in Eq. (3.31), the independence of σ_α^{vis} from the transfer function ensures that it does not depend on β. In practice, this only holds if the selections applied on the integration bounds of the TF depend on the particle quantities. If not, for example if selections are made on $E_{\mathrm{T}}^{\mathrm{miss}}$, a residual dependence on β might have to be taken into account, or shown negligible depending on its observed impact on σ_α^{vis}.

Another improvement to the detector approximation through its TF would be the inclusion of the b-tagging information. Although the identification is imperfect, it can still be leveraged in the integration to distinguish b-quark jets from the ones of other flavours or gluons through additional terms in the TF for each of the partons in the final-state y that produce a jet :

$$W_b = \begin{cases} \epsilon_b \text{ if jet is assigned to b-quark} \\ \epsilon_{\not b} \text{ otherwise} \end{cases}, W_{\not b} = \begin{cases} 1-\epsilon_b \text{ if jet is assigned to b-quark} \\ 1-\epsilon_{\not b} \text{ otherwise} \end{cases}, \tag{3.52}$$

where ϵ_b and $\epsilon_{\not b}$ are the b-tagging and mistag rates, W_b and $W_{\not b}$ are additional TF components associated to b-tagged and untagged jets. This equation applies to a binary classifier using a fixed WP. This can be extended to a multi-classification (such as DeepJet) by using the tag and mistag rates of each class (typically b-jet, c-jet and lightjet), and to make use of the full range of the score in bins on which $W_{b,\not b}$ and $\epsilon_{b,\not b}$ depend. Additional dependency can be included, for example the transverse momentum or pseudo-rapidity of each jet, typically what is done to calculate b-tagging SF.

This is especially useful in case there are assignment ambiguities. In Section 3.3, the event selection required at least two b-jets, and the selected candidates were based their b-tag score. In the rare cases where an extra b-tagged jet is identified, or in the more frequent cases where one of the b-jet failed to pass the b-tag and the decision is made to lower the requirement on the number of b-tagged jets to recover these events, the different combinations must be tested. This is even more important when light flavour quarks are present in the final-state, typically in the HH $\rightarrow$ bbWW decay channel, when one of the W boson decays hadronically. This can also be extended to the cases of leptons. The TF is then expressed as

$$T(x|y) = \sum_{i}^{n_{comb}} \prod_{l}^{n_{leptons}} T_l(E_l^{rec}|E_l^{gen})\ \delta^{(2)}(\vec{u}_l^{rec} - \vec{u}_l^{gen}) \prod_{j}^{n_{jets}} T_j(E_j^{rec}|E_j^{gen})\ \delta^{(2)}(\vec{u}_j^{rec} - \vec{u}_j^{gen})\ W_f(\text{b-tag}, \eta, p_{\mathrm{T}}), \quad (3.53)$$

where the sum goes over the n_{comb} possible combinations of leptons and jets to which they are assigned their energy TF, the resolution on the direction $\vec{u}$ is assumed perfect, and the subscript "rec" and "gen" denote the reconstructed quantities from the detector and at generator-level from the ME. The jets are also assigned the b-tagging information in a potentially flavour-dependent TF, that may or may not depend on variables such as the b-tag score, transverse momentum and pseudo-rapidity. The dependence on detector effects β was made implicit, and can impact each term of Eq. (3.53) differently, or even selectively.

As a side-technical note, while permutations of indistinguishable final-state particles are already included in MoMEMta, the possible various combinations of Eq. (3.53) are not. Although this can be achieved by resolving the integral for each combination, this is suboptimal. By integrating over the entire product, the number ME evaluations, which can be costly, is minimised. This argument is also valid for β, such that several values can be computed together with the same calls to the ME, with the added benefits that fluctuations of the results are reduced.

On the interpretation side, the analyses performed at the LHC differ a lot compared to the ones at the Tevatron accelerator. The LHC has to deal with very high-dimensional and intractable simulations. This means that physicists have to project the raw data to reduce the dimensionality using a *summary statistic* [353], which is problem-dependent, and does not typically capture all the information of the event, especially not indirect effects from EFT parameters. This used to be a man-crafted variable, for example the invariant mass of a resonance (e.g. the discovery of the Higgs boson at 125 GeV [53, 54]). Although this method provides a solid understanding of results, it does not make use of the full information contained in the data. Nowadays, more and more ML algorithms are trained on this high-dimensional space, and used as a summary statistics. Still, both methods rely on a binned likelihood estimation, as described in Section 2.4.3, which limits the potential of discovery as individual events are aggregated into bins of a distribution. Tevatron analyses that use the MEM can directly use the likelihood defined by the method given the much lower amount of events compared to the LHC for which this is computationally impossible. However, the method developed in this chapter promises to open the door for such interpretations, which promises to make better use of all the information.

The likelihood over a sample of measured events $\{x_1, ...x_n\}$ can be written as

$$L_{sample}(x_1, ...x_n; \vec{\alpha}, \vec{\beta}, \vec{f}) = \prod_i^N L_{event}(x_i; \vec{\alpha}, \vec{\beta}, \vec{f}), \tag{3.54}$$

where the vectorial definition of α and β has been highlighted. $\vec{f}$ characterises the mixture of processes in all the events of the sample, such that

$$L_{event}(x_i; \vec{\alpha}, \vec{\beta}, \vec{f}) = \sum_{\text{process } p} f_p L_p(x_i; \vec{\alpha}, \vec{\beta}), \tag{3.55}$$

where $L_p(x_i; \vec{\alpha}, \vec{\beta})$ corresponds to what was referred to as $P(x|\alpha, \beta)$ so far, with the formula of Eq. (3.29). It is important to note, just like it was the case for the DY, $t\bar{t}$ and H $\rightarrow$ ZA examples developed in Section 3.3, that not all processes depend on the α parameters, which can decrease the amount of computations. This may also be true of the experimental parameters β, though the absence of bias when neglecting the dependency has to be checked case by case.

By comparison, the $\vec{\alpha}$ parameters can be understood as the parameter of interests of Section 2.4.3 and $\vec{\beta}$ as the Nuisance parameters. The treatment of the likelihood can then be the same as with a binned likelihood : it can be marginalised into one parameter, the best-fit and 1σ (2σ) CL level intervals are extracted from the minimum of $-\ln(L)$ and the intersections with the minimum elevated by 1/2 (2).

Including an additional component to the TF was already discussed in Section 3.3.5 for the case of ISR. This could be further developed in the context of this section, where the fraction of events for a given process that give rise to an additional jet could be included in the parameter $\vec{f_p}$. This would be a mean to control the systematic uncertainties related to the factorisation scale.

The sum in Eq. (3.55) can span a lot of different processes, and evaluating all of their weights can be limiting. In practice, only the dominant ones can be considered explicitly. The effect of neglecting some background processes can shift the measurement of the parameters, that will depend on the fraction of events that are neglected, and an increased statistical uncertainty. However, this only represents a quantitative effect that can be corrected, not a qualitatively different interpretation. To account for this and the various approximations of the detector response, parton shower and hadronisation in the TF, fully simulated events can be used as a means of calibration.

This calibration is performed similarly to the likelihood evaluation on data, with the same conditions (ME, PDFs, etc). Pseudo-experiments are simulated, producing likelihood values for a certain number of events and comparing the measurements in Eq. (3.54) from the true parameter values $\vec{\alpha}$, $\vec{\beta}$ and $\vec{f}$. In practice, *bootstrapping* can be used to sample events from a pool with redrawing to reduce computation time, with an effect to take into account when assigning uncertainties. Repeating the

pseudo-experiments to produce a distribution then allows to estimate the potential bias, assign uncertainties and assert their pulls during the fit. A calibration curve can be obtained by varying the parameters. For processes with similar kinematic features, the computation can be limited to one in the likelihood, with the correct proportion f_p in Eq. (3.55).

The advantage of using Eq. (3.54) is that it replaces completely the binned likelihood method typically used in LHC analysis. As there is no need to accumulate the event information into distributions, the method can fully use the information within each data events. The simulated events only serve as a calibration to the method, and a way to estimate the uncertainties. Alternatively, this method can also provide a *differential* likelihood $(\partial L(x)/\partial Z)|_{Z=Z_0}$ over some variable Z evaluated at Z_0 by including a delta function $\delta(Z(y) - Z_0)$ in the integration Eq. (3.29).. This can in practice be used to obtain a differential distribution of the MEM weights that can be useful to disentangle ambiguities in the reconstruction of kinematic variables in decays that involved invisible particles such as neutrinos [295, 354].

ML regression

Several improvements can be brought to the regression of the MEM weights, either from the training or the architecture of the NN.

In the inputs of the method developed in this chapter, the cylindrical symmetry information was removed from the training by only using a relative ϕ information. This makes sense, as the prediction should be independent of the rotation symmetry. An alternative path could reside in *data augmentation*, which consists in expanding the training sample based solely on the information contained in the training dataset. In the context of image classification, this takes the shape of image flip, rotations, scaling, cropping, translations or random noise modification. Applied to the case at hand, it could consist in rotations over the azimuthal axis, which would expand the training set and allow the network to learn about the symmetry. Flipping along the beam axis could also be including, as the physics of the observed final-states is independent of the reflection on the transverse plane. This would however also invert the longitudinal boost from the difference in the two partons momenta, and although they are integrated over the integral value is not guaranteed to be conserved. After the training, *test-time data augmentation* can be used to average the prediction over different rotations, such that any potential bias is smoothed out.

Another case of data augmentation concerns the systematic shifts of the particle variables. While the regressive DNN model was partially validated on shifted b-jets p_{T} in Section 3.3.3, this could be better modelled by including in the training set shifted events from the start. Although this would increase the number of evaluations with MoMEMta, and therefore the computation time required to build the training

set, only a fraction of events could be required, and for maximum shifted variables. This could be enough to extend enough the interpolation region of the DNN such that every systematic variation is well reproduced. This is related to the inclusion of the β parameter representing detector modelling in Section 3.3.5, whether β is included in a parametric learning or not.

Regarding the architecture, the fully connected architecture is relatively simple and could use a more state-of-the-art model. Recently, a team has reproduced the results of Section 3.3 with a Residual Neural Network (ResNet) [355]. This kind of NN uses connexions between distant layers in order to mitigate the problem of vanishing gradients and improve the performance of deep models. The idea was to better reproduce the tails of the distributions of the MEM weights in Figs. 3.10 to 3.12, and the authors showed some improvement there. Going further, the potential developments of the use of the MEM such as in Eq. (3.53) need to be reproduced in the regression, in this case the various combinations between jets and final-states. This is further complicated when the number of particles may vary per event, or when ISR is taken into account. This is hard to associate with a fully connected network, apart from zero-padding the absent inputs, which is suboptimal. Graph Neural Networks (GNNs) [356] represent a recently emerged branch of ML that could leverage this shortcoming, they have been used more and more in HEP in recent years [357]. By defining the inputs in a graph to take into account the geometrical correlations between the different 4-momenta, and allowing a variable number of inputs, GNNs seem like the best suited candidates for a better regression of the MEM weights and adaptation to the constraints of the experimental side. They could further include in the node information the b-tagging score, such as introduced in Eq. (3.52), or the underlying Feynman diagram hypothesis in the same spirit as MoMEMta. Another road, going in the opposite direction as the cylindrical data augmentation, would be to use rotation-invariant architectures [358,359], such that the model can focus solely on the relevant physics. Similarly, the DNN may be decorrelated from the NP using the already proposed data augmentation with systematic variations, or pivoting within a GAN [360]. However, it was advocated in Ref. [361] that the DNN is then agnostic on the NPs, which may result in worse separation power (or in this case regression), and a better approach would consist in a parameterised learning which was already considered for the detector parameters β. Finally, it is worth mentioning a very recent development in Ref. [362] that uses CINNs for the MEM integration.

Application to the HH $\rightarrow$ bbWW case

There are several benefits to using the MEM in the HH searches. In analyses looking for a particularly elusive signal signature, the potential gain of any method that allows to reduce the background contributions while maintaining the signal efficiency can be large in terms of sensitivity. In that context, the MEM could prove

extremely valuable by taking into account correlations between variables optimally.

All current HH analyses use ML methods in one way or the other. Most commonly, a DNN is used as a classification task to separate the HH process from the different backgrounds, with different strategies but the same philosophy. The analysis performed in the HH $\rightarrow$ bbWW channel is no exception. This is motivated by two aspects typical to HH searches : the faintness of the signal compared to the amount of background contamination, and the high-dimensional parameter search of a four particle final-states, notwithstanding the further decays, for example of the W bosons in the bbWW channel. DNNs and their variations, are particularly suited to make use of that amount of this high-dimensional data. No current alternative has come close to the amount of sensitivity reached with these techniques, which explains their now almost omnipresence.

This also includes the physicist inspiration, and explains why these methods are always under scrutiny. As there is no direct way to assess whether the correlations between variables are taken into account properly during the training, and whether the model has correctly "understood" the physics behind, convincing others (and ourselves) of the validity of the predictions requires a lot of work. Additionally, there is no way to make sure that the maximum sensitivity has been achieved with the amount of data at hand. As this data was collected thanks to an amazing effort from the LHC and its experiments personal, with great cost in money and man-hours, everything should be done to maximise the physics reach.

The MEM on the other hand, offers an interesting counter-offer. All the kinematic information is included with correct correlations, including invisible particles whose degrees of freedom are considered based on the process hypothesis. Although this information is implicitly included in the training set of a DNN, there is no certitude that it is fully considered. Another advantage lies in the high final-state multiplicity that produces an exponentially increasing number of kinematic variables. There are two trends in HEP regarding the inputs of an DNN. Some analyses try to minimize this number of inputs by removing variables, either before the training by looking at the correlation matrix, or after the training by evaluating their importance on the prediction. Recent developments, notably in particle taggers developed within the CMS experiment, show that with deep enough architectures it is beneficial to include as many meaningful variables as possible. The algorithm then has the liberty to pay more attention to certain high-level variables and even reconstruct its own from the low-level ones. There is no real consensus as to which approach is better suited, one could limit the potential for mismodelling in overrepresented parts of the phase-space, the other could make better use of the non-linear relations between the inputs. In contrast, the information contained in the MEM weights is close to maximal, given the Neyman-Pearson lemma. Technically, including them in the training would render most other kinematic variables redundant. Therefore, the number of variables could be reduced without removing information.

Finally, the MEM has a direct physical meaning with the framework of QFT, which can give a handle for physicists to properly understand their data. So far, the MEM has been too prohibitive for LHC analyses, leaving the entire floor for MVA methods, which are still motivated and a very good tool for the problem at hand. Yet, there would be a clear benefit in bringing back the MEM on the front stage. The method that was described in this chapter represents a first step into that direction, and hopefully can pave the way for a better synergy between these two different techniques.

Applying this method to the $HH \to bbWW$ analysis was considered but not realised in time for the deadline of this thesis. Still, for the reasons mentioned above, there could be a clear gain. This was shown in Ref. [363], using a discriminant as in Eq. (3.35) between the HH and $t\bar{t}$ hypothesis. It was however clearly mentioned that the computation time was a limiting factor, especially when there is a large background contamination, as is typical of the bbWW channel. Additionally, only one signal and one background hypotheses were included. Several other background hypotheses could be considered, despite the dominance of the $t\bar{t}$ process, and more importantly on the signal hypothesis. Several resonant scenarios can be investigated with a parameterised mass, as was done in Section 3.3 for the $H \to ZA$ case, or by using EFT parameters in the ME. The case where one of the W boson decays hadronically was not considered in Ref. [363] given the combinatorial challenge. This is a limitation that the MEM still suffers from, but the problem of jet-parton assignment is not restricted to it. The MEM has the advantage of considering the correlations between parton final-state, such that all jet assignment configurations in Eq. (3.53) are considered with their physical probablity, and the weights could help identify genuine HH events.

Chapter

4 Search for Higgs boson pair production in the decay channel with two bottom quarks and two W bosons using the Run-2 dataset of the CMS experiment

This chapter is dedicated to the experimental search of the Higgs boson pair production. The first section will address the current published searches and their results from both the CMS and ATLAS experiments, with an emphasis on the recent improvements performed in each channel by the former.

In the next section is reported the search performed in the decay channel of two bottom quarks and two W bosons. This search was performed on proton collision data recorded at $\sqrt{s} = 13$ TeV center-of-mass energy by the CMS collaboration during Run-2 data taking years between 2016 and 2018, corresponding to an integrated luminosity of 138 fb^{-1}.

The phenomenology of the double Higgs production was outlined in Section 1.5. The motivations for the study of the non-resonant production through the SM measurement and parameterised deviations in terms of the coupling modifiers (notably κ_λ and κ_{2V}) were discussed in Section 1.5.2, and its EFT interpretations in Section 1.7.2. The resonant search will focus on narrow-width spin-0 and spin-2 resonances decaying to HH with masses between 250 GeV and 900 GeV, following the description in Section 1.6. As both parts of the analysis follow the same procedure the following sections will address them both, their differences will be outlined when necessary. The presence of neutrinos prevents the reconstruction of the m_{HH} invariant mass, the strategy of the resonant analysis thus follows the one of the non-resonant.

4.1 Review of current experimental results

This section aims at giving a non-exhaustive summary of the current experimental constraints on the double Higgs production. The production cross-sections and self coupling constraints were set at $\sqrt{s} = 8$ TeV (Run 1) and 13 TeV (2016 and Run 2). The Higgs boson has a wide range of possible decay channels, each presenting its own challenges and variable sensitivity over the production cross-section and coupling modifiers. This is further diversified when two Higgs bosons are considered, and strategic thinking must be adopted to elect what channels are the most competitive in the measurement of the observables, their respective BRs are on Fig. 4.1. Naively, one would only look at the channels with highest BR, as they will have the largest amount of HH events. Yet, this does not include the background contamination, which can be substantial depending on the channel.

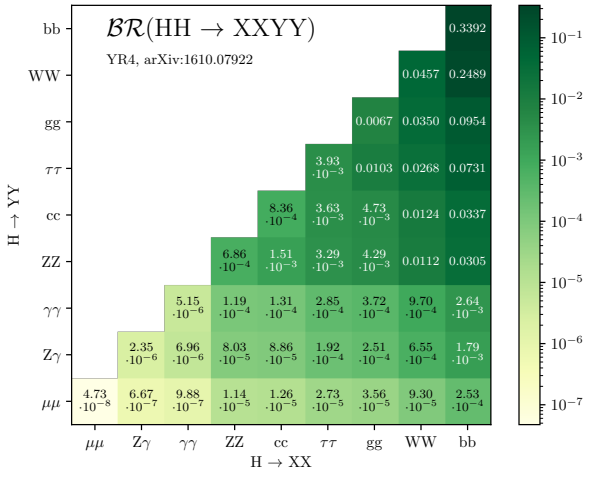
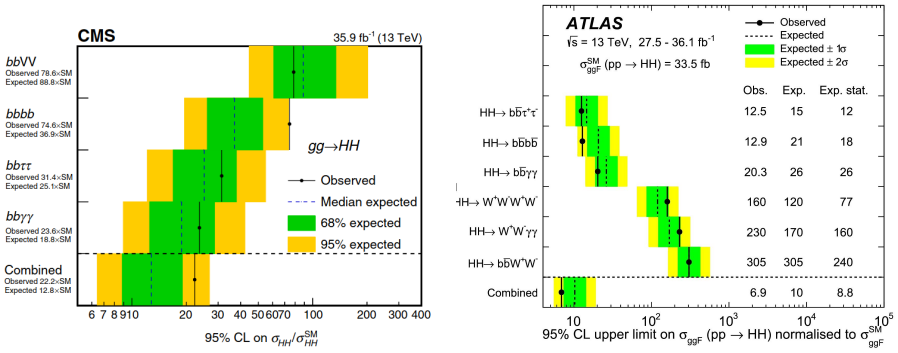


Figure 4.1 | Branching ratios of the HH decay channels, with values from [63].

For example, the bbbb channel has the largest BR but a large multi-jet background, the bb$\gamma\gamma$ one of the smallest BR but benefits from an excellent selection efficiency and measurement of photon energy, and the bb$\tau\tau$ lies in the middle with reasonable BR but also a clear signature from leptons and hardonically decaying taus. It is often said that there is no *golden channel* in HH searches, i.e. a channel whose sensitivity surpasses all others. Therefore, as many channels as possible need to be combined to extract the best sensitivity. Still, comparatively the three channels mentioned in this paragraph have been observed as the most sensitive, which is why

they are usually referred to as the *silver bullets*. This can be observed in the searches performed by CMS and ATLAS using the 2016 dataset, illustrated on Fig. 4.2 and focused on these three dominant channels with a few additions. ATLAS had an advantage over CMS b-tagging efficiency because of different tracking performances, which directly translates into better sensitivity over channels that contain pairs of b-quarks. The very good photon selection efficiency and resolution of photon energy in CMS explains why the bb$\gamma\gamma$ performs better than the other channels.



(a) *(b)*

Figure 4.2 | Combination of HH searches in CMS (left, from Ref. [364]) and ATLAS (right, from Ref. [365]) using data taken in 2016 in various channels and their combinations, with the expected and observed limits on the inclusive cross-section as a function of the theoretical prediction.

In Table 4.1, a comparison has been made between all the currently available HH results from both experiments at the time of writing of this document. Already between Run-1 and 2016 results there is a major improvement in the HH inclusive cross-section measurement, which can be explained by several complementary factors related to the upgrades of the accelerator and the detectors during the long shutdown that followed Run-1. The accelerator upgrades affected the luminosity and energy, which alsmot both doubled in 2016 compared to Run-1. This translates to an increase of statistical precision of about 40 %, and a higher HH cross-section (a factor three between 8 TeV and 13 TeV [62]) mostly due to the GGF production mode, respectively. The major detector upgrades consist in the pixel tracker and HCAL of CMS that are crucial for better jet reconstruction and potential SV localisation for b-tagging, with similar developments for ATLAS. A better trigger system (also helped by the smaller granularity of the subdetectors) also allowed to increase the statistical precision by allowing more data to be recorded. Both muon detection systems were improved, to the benefit of some HH decay channels (the ones including W or Z leptonic decays). Finally, the better reconstruction techniques of the different final-state particles were improved after Run-1, most notably the more systematic

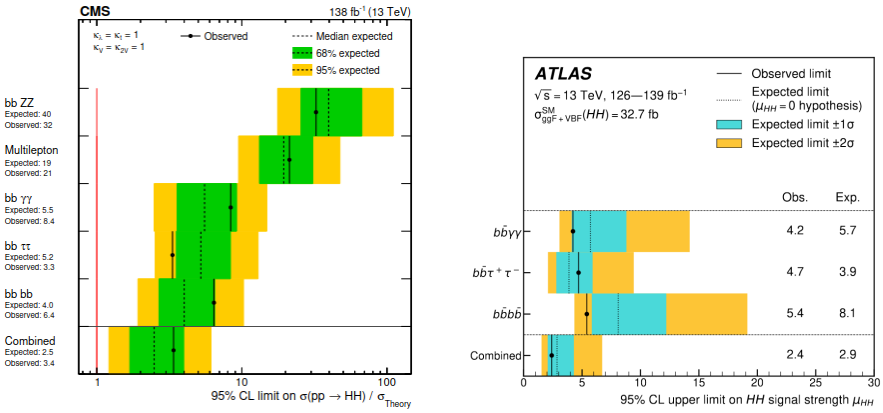
use of MVA techniques, some of which were highlighted in Section 2.4 for CMS. In parallel to the improvements related to data-taking and reconstruction, the analyses techniques also improved and included additional channels to the three dominant bbbb, bb$\tau\tau$ and bb$\gamma\gamma$: bbWW, WWWW for ATLAS (WW$\gamma\gamma$ was already in the Run-1 combination), and bbVV for CMS. The spectrum of analysis results was widened by not only measuring the inclusive cross-section, but also the the coupling κ_λ.

Table 4.1 | Summary of double Higgs production measurements from the two main collaborations in several decay channels with their associated integrated luminosity. Upper limits observed (expected) are expressed as a number of times the SM prediction for both the inclusive (mostly GGF) and the VBF production processes, as well as constraints at 95 % CL on κ_λ and κ_{2V} assuming all other coupling modifiers at their SM value. ATLAS Run 1 results were obtained by combining bbbb, bb$\gamma\gamma$, bb$\tau\tau$ and WW$\gamma\gamma$ (at different integrated luminosity values), its 2016 combination result (with a small amount of data from 2015) was obtained with the same channels with the addition of bbWW and WWWW. The CMS Run 1 results were obtained by combining bbbb, bb$\gamma\gamma$ and bb$\tau\tau$, the 2016 combination was with the addition of bbVV. The BRs used in each decay channels are taken from the SM alone.

Experiment	Channel	$\int \mathcal{L}dt$ [fb^{-1}]	$\sigma_{\mathrm{HH}}/\sigma_{\mathrm{HH}}^{SM}$	$\sigma_{VBF}/\sigma_{VBF}^{SM}$	κ_λ constraints	κ_{2V} constraints
ATLAS (Run 1)	Combined [366]	20.3	70 (48)	-	-	-
ATLAS (2015-2016)	Combined [365]	36.1	6.9 (10)	-	-5 (-5.8) $< \kappa_\lambda <$ 12 (12)	-
ATLAS (Run 2)	bbbb (VBF) [367]	126	-	840 (550)	-	-0.43 (-0.55) $< \kappa_{2V} <$ 2.56 (2.72)
	bb$\ell\nu\ell\nu$ [368]	139	40 (29)	-	-	-
	bbbb [369]	126	5.4 (8.1)	130.5 (133.4)	-3.9 (-4.6) $< \kappa_\lambda <$ 11.1 (10.8)	-0.03 (-0.05) $< \kappa_{2V} <$ 2.11 (2.12)
	bb$\gamma\gamma$ [370]	139	4.2 (5.7)	-	-1.5 (-2.4) $< \kappa_\lambda <$ 6.7 (7.7)	-
	bb$\tau\tau$ [371]	139	4.7 (3.9)	-	-	-
	Combined [372]	139	2.4 (2.9)	-	-0.6 (-2.1) $< \kappa_\lambda <$ 6.6 (7.8)	0.1 (0.0) $< \kappa_{2V} <$ 2.0 (2.1)
CMS (Run 1)	Combined [373–375]	[17.9,19.7]	43 (47)	-	-	-
CMS (2016)	Combined [364]	35.9	22.2 (12.8)	-	-11.8 (-7.1) $< \kappa_\lambda <$ 18.8 (1.6)	-
CMS (Run 2)	bbZZ [376]	138	32.4 (39.6)	-	-8.8 (-9.8) $< \kappa_\lambda <$ 13.4 (15)	-
	WWWW,WW$\tau\tau$,$\tau\tau\tau\tau$ [377]	138	21.3 (19.4)	-	-6.9 (-6.9) $< \kappa_\lambda <$ 11.1 (11.7)	-
	bb$\gamma\gamma$ [378]	138	7.7 (5.2)	225 (208)	-3.3 (-2.5) $< \kappa_\lambda <$ 8.5 (8.2)	-1.3 (-0.9) $< \kappa_{2V} <$ 3.5 (3.1)
	bbbb (resolved) [379]	138	3.6 (7.3)	226 (412)	-2.3 (-5.0) $< \kappa_\lambda <$ 9.4 (12)	-0.1 (-0.4) $< \kappa_{2V} <$ 2.2 (2.5)
	bbbb (boosted) [380]	138	9.9 (5.1)	-	-	0.62 (0.66) $< \kappa_{2V} <$ 1.41 (1.37)
	bb$\tau\tau$ [381]	138	3.3 (5.2)	124 (154)	-1.7 (-2.9) $< \kappa_\lambda <$ 8.7 (9.8)	-0.4 (-0.6) $< \kappa_{2V} <$ 2.6 (2.8)
	Combined [382]	138	3.4 (2.5)	-	-1.24 $< \kappa_\lambda <$ 6.49	0.67 $< \kappa_{2V} <$ 1.38

While several analyses are still ongoing over the whole Run-2 dataset, including the bbWW analysis presented in this document, the current public results have shown a significant improvement compared to their 2016 equivalent, as illustrated on Fig. 4.3 and summarised in Table 4.1. As the shutdown between 2016 and the rest of the Run-2 was significantly shorter than the long shutdown after the Run-1, no major upgrade was performed, although noticeably the inclusion in the CMS tracker of a fourth pixel layer (cf. Section 2.3.1) has improved the tracking performance and partly explains the closer sensitivities between CMS and ATLAS. Beside the luminosity increase of around a factor four (which would roughly translate as a factor two in statistical precision), the main improvements lie in the reconstruction and tagging techniques used in both experiments. In terms of results, the better

sensitivity has motivated the different analyses to provide more experimental results, most notably the inclusion of the VBF cross-section and the κ_{2V} coupling, and more recently the c_2 coupling (cf. Section 1.7.2). In the rest of this section, the improvements of the pre-existing channels and the challenges faced in the new ones will be outlined for the case of CMS, using the Run-2 dataset. Unless specifically mentioned, their references can be found in Table 4.1.



(a) *(b)*

Figure 4.3 | Combination of HH searches in CMS (left, from Ref. [382]) and ATLAS (right, from Ref. [372]) during entire Run-2 data taking in various channels and their combinations, with the expected and observed limits on the inclusive cross-section as a multiple of the theoretical prediction.

The bbbb channel contains around a third of all expected HH events, but suffers from a large contamination of multi-jet events resulting from QCD processes, along with additional backgrounds with a certain number of b-quarks in the final-state. In CMS this final-state is studied in two analyses targeting a different part of the m_{HH} spectrum : the *resolved* analysis in low-m_{HH} region where the four b-jets are resolved individually in AK4 jets, and the *boosted* analysis where each H $\rightarrow$ bb decay is contained within a large-radius AK8 jet. The main challenge in both analyses is the reduction and precise modelling of the multi-jet background, it also represents the main improvement compared to the 2016 analysis. The DeepJet algorithm (cf. Section 2.4.2) is used in the resolved analysis, together with a distance measure to correctly associate each H $\rightarrow$ bb decays, and a b-tagging algorithm specifically designed to explore the substructure of the merged jet based on a GNN architecture (called ParticleNet [383]) is used in the boosted analysis. The consideration of two production modes (GGF and VBF), compared to the 2016 analysis that only considered an inclusive cross-section, as well as several coupling modifiers such as κ_{2V}, requires the optimisation of the sensitivity with respect to several parameter of interests within the same analyses. Both use the presence of forward jets, a signature

of the VBF topology, and a BDT trained to further discriminate VBF from GGF. Then several categories, are defined using additional BDTs trained to discriminate HH events from the backgrounds and kinematic variables to improve the sensitivity to the various couplings. The low-m_{HH} region of the resolved analysis makes it more sensitive to κ_λ (the triangle diagram on top of Fig. 1.11 that depends on κ_λ typically has a low m_{HH} spectrum, as illustrated on Fig. 1.10). The boosted analysis provides the best sensitivity to κ_{2V} compared to all current measurements, excluding the $\kappa_{2V} = 0$ hypothesis by more than 6 standard deviation (cf. Table 4.1). This is due to the large boost of the HH pair when $\kappa_{2V} \neq \kappa_V^2$, as discussed in Section 1.7.2, that helps the signal extraction. Both analyses depend on a reliable estimation of the QCD backgrounds, hard to reach with simulations, and therefore use a standard data-driven method (similar to the one pursued in bb$\gamma\gamma$ and bb$\tau\tau$ analyses). Compared to the previous 2016 results, the resolved and boosted analysis claim a factor of 5 and 30 times in terms of upper limits respectively, mostly thanks to reconstruction techniques and b-tagging performance, especially in the substructure of the merged jets. Some difficulties still encountered in these analyses are the trigger selection that is complicated at low m_{HH} as pileup limits the ability to map the SVs for b-tagging purposes, and the limited amount of data in the control regions used to estimate the QCD background.

The bb$\gamma\gamma$ channel only has a BR of 0.3 %, yet the presence of the photon pair provides a selection criterion with very high signal-to-background ratio. In addition, by relying on photon triggers this channel has access to a lower range of m_{HH} not accessible to the bbbb channel for example. A multiclassification BDT is used to define GGF and VBF categories, with for each several background classes. Together with a robust and more suited to anomalous couplings m_{HH} variable estimator, they define multiple categories in each of which the distributions of the invariant masses $m_{\gamma\gamma}$ and m_{jj} are used in the likelihood fit. While the former allows a good background rejection, the latter is useful to discriminate from the single Higgs processes that are indistinguishable using the diphoton peak only. Only the ttH process is irreducible from this method, hence why a DNN is trained to reject it. The main improvements compared to the 2016 analysis consists in better b-tagging, improved classification strategy from the BDT, but more importantly a better reconstruction of the dijet peak, using both a b-jet energy regression (that was also used in the bbbb analysis) and a regression of the m_{jj} to reduce the bias and improve the resolution. Overall, the Run-2 analysis claims a factor three in improvement of the upper limits set by the analysis compared to the 2016 result.

The bb$\tau\tau$ analysis focuses on the final states where at least one tau decays hadronically, therefore representing 88 % of the pair of tau decays, with a BR of HH decay of around 7 %. The Run-2 analysis keeps the three channels separated ($\tau_\mu\tau_{\text{h}}$, $\tau_{\text{e}}\tau_{\text{h}}$ and $\tau_{\text{h}}\tau_{\text{h}}$), and performs selections based on the b-tagging and kinematic considerations. Similar to the bbbb analysis, resolved and boosted topologies of

the H $\rightarrow$ bb decay, as well as GGF and VBF production modes, are considered separately. A DNN is trained between HH and background in each categories, with an additional DNN trained at discriminating VBF events from several classes of backgrounds. The main improvement in the analysis compared to the 2016 one resides in the more complex DNN classification applied on all categories, while a more classic BDT was used only in some categories in 2016. This was enabled by better b-tagging efficiencies that increase the statistics and allowed a sufficient training, something that as not feasible before. In addition, the Run-2 results benefits from better reconstruction efficiencies of the hadronic taus, especially considering the DeepTau algorithm (cf. Section 2.4.2). All these improvements have allowed CMS to catch up with the sensitivity of ATLAS, with upper limits five times better than they were in 2016.

The last channel with a H $\rightarrow$ bb decay that has a large enough BR for consideration is the bbVV channel, where the decay of the Higgs boson to vector bosons can either be to a pair of W or Z bosons. In the 2016 analysis, only the $\mathrm{b\bar{b}}\ell^{+}\nu\ell^{-}\bar{\nu}$ final-state was considered : either the two W bosons decay leptonically, or one Z boson decays into leptons and the other into invisible neutrinos. This analysis used state-of-the art parametric DNN, parameterised in functions of κ_λ and κ_t (and similar for the resonant interpretation), but only considered the resolved H $\rightarrow$ bb decay and not the case when one W decays to hadrons. After Run-2, the bbVV process was studied separately as bbWW and bbZZ. The latter focused on the fully leptonic decay of both Z bosons and can therefore be interpreted as a new analysis. Although the BR is relatively small (3 %), the four-lepton final state is a strong signature, easily distinguishable from the backgrounds. After kinematic selections and the definition of categories for each final-state, a BDT was trained for each of them. This newly opened channel benefits from the b-tagging improvements that benefitted the other channels, and some improvements in the lepton efficiencies as well due to the better tracking reconstruction. The improvement of the bbWW analysis is the goal of the analysis described in this document, it includes among other things a more complex multi-classification architecture, the addition of the $\mathrm{b\bar{b}}\ell\nu q\bar{q}$ final-state and the boosted topologies. The main challenges that this analysis has to face is the much larger background contamination compared to most channels, the added complexity of the combinatorial jet-parton assignment when one W boson decays hadronically, and the difficulty to model fake leptons (i.e. jets being interpreted as leptons). Although a data-driven method was used in the bbZZ analysis, the amount of events with fake leptons is much higher in that decay channel of the bbWW analysis and still represents a challenge. This channel however represents the second largest BR with about one fourth of all HH events, which makes the challenge of its measurement worth the endeavour.

The channels that do not display a decay of the type H $\rightarrow$ bb typically have a smaller BR and at first sight do not seem particularly relevant. They are however

less affected by QCD background contaminations than channels with b-jets, and therefore are rather limited by statistics, which makes them especially interesting for Run-2 analyses (although some were already in the 2016 results of ATLAS). In particular, three different channels are included into the so-called *multilepton* analysis : WWWW, WW$\tau\tau$ and $\tau\tau\tau\tau$. The challenge of this channel consists in the many leptonic and hadronic tau final-states. In the CMS analysis seven categories are defined, with kinematic requirements tuned for each of them, some are combined with jets requirements associated to hadronic W decay. Finally, and similarly to the bbZZ analysis, a BDT is trained on each category. The other challenge of this channel consists in the various backgrounds that are hard to simulate. The fake leptons is a similar problem to the bbZZ analysis, and its data-driven method was employed in the bbWW analysis presented in this document. Additionally, some categories are based on same-sign leptons, which represents a much smaller proportion of events in data than opposite-sign leptons. Any mismeasurement of the charge of a lepton, called a *charge flip*, can result in a source of background not accounted for. To overcome this, an additional data-driven method has been designed to incorporate it, with a procedure similar to the fake lepton estimation. Despite these challenges and the small signal yield, the channel provides a reasonable sensitivity, that can only get better as statistics accumulate, which will benefit both the statistical precision and the data-driven estimations. Another considered channel without b-jets is the WW$\gamma\gamma$ channel, that also benefits from the precise resolution of the diphoton mass peak, but without the possibility to reconstruct the resonance H → WW, except in the fully hadronic decay, but the light quark jets are harder to distinguish from the background. The $m_{\gamma\gamma}$ mass distribution is used in the likelihood fit, and the background rejection is based on kinematic requirements, with additional more cutting-edge binary or multiclassification DNN when a sufficient training set can be assembled. The independence of the fully hadronic decay from the bb$\gamma\gamma$ analysis is done through a rejection DNN.

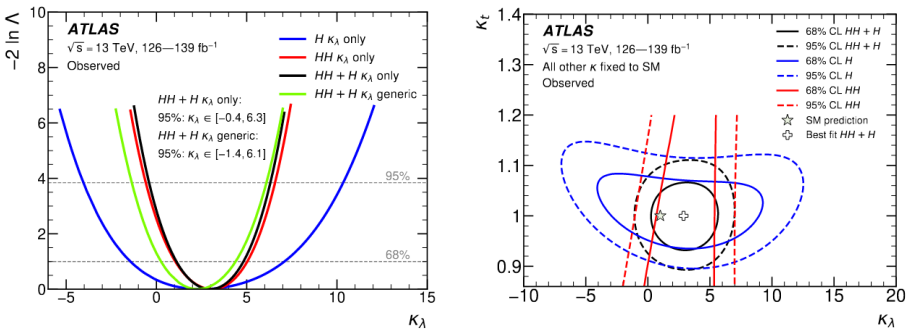
Searches in additional channels are being carried out in the CMS collaboration, for example the aforementioned WW$\gamma\gamma$, fully hadronic bbWW, semileptonic bbZZ, and $\gamma\gamma\tau\tau$ channels, as well as other production modes such as VHH and t$\bar{\mathrm{t}}$HH. Currently, only one measurement targetting the VHH mode specifically has been made public by the ATLAS collaboration, with 95 % CL observed (expected) exclusion limits of $-34.4\ (-24.1) < \kappa_\lambda < 33.3\ (22.9)$ and $-8.6\ (-5.7) < \kappa_{2V} < 10.0\ (7.0)$ [384]. While much less sensitive than the aforementioned searches targeting mostly the GGF production mode (that has a two orders of magnitude larger cross-section, cf. Eq. (1.53)), the VHH searches have access to new VVHH couplings that the former cannot measure.

When comparing the combined results in Fig. 4.3 and the three dominant channels (bbbb, bb$\gamma\gamma$ and bb$\tau\tau$), it would naively seem like the additional channels are not worth the effort as they do not bring much in terms of sensitivity. Dismissing

them would nonetheless be ill-advised for several reasons. First, as the searches for HH production are currently limited by the amount of data that has been collected so far, the only way to provide the best constraints with the data at hand is to include as many channels as possible. They can serve as very useful validations in case an excess is observed in one of them. If a BSM scenario would cause an unexpected increase in the HH cross-section for example, something that would not be known prior to developing the analyses (which can take years). Then, the inclusive cross-section is not the only quantity that is looked at in Run-2. Several production modes and EFT couplings are considered, and the sensitivity to each of them can vary between channels, some being more sensitive to a specific coupling while not ranking in the most sensitive channel based solely on the inclusion cross-section. For example, the boosted topologies are more sensitive to anomalous values of κ_{2V}, and the bbWW channel studied in this document is expected to show good sensitivity as well for this coupling (and to a lesser degree κ_V) thanks to the decay H $\rightarrow$ WW. In the absence of any deviation so far, there is a very strong motivation to look in every possible direction. Finally, the evolution of the sensitivity of each channel as a function of the collected luminosity and time is not trivial to predict. The channels that display a H $\rightarrow$ bb decay will benefit from b-tagging performance improvements and the boosted topologies from jet substructure reconstructions. Both depend on the skills developed by the physicists that design the algorithms, and the increased data can help improve the training of MVA methods. On the other hand, the more marginal channels without H $\rightarrow$ bb are mostly limited by statistical uncertainties, and will benefit directly from an increased luminosity. As most of these channels are new, there are many lessons learned during Run-2, and there is no doubt that they will be put to good use in future searches to extract even more information from the data, just like what happened for the dominant channels between the different runs.

As discussed already in Section 1.5.2, while HH searches allow a direct estimation of κ_λ, single Higgs production are also dependent on κ_λ, albeit only at NLO. While this means a reduced experimental sensitivity, there are three orders of magnitude difference in terms of production cross-section, such that single Higgs measurements can become competitive with the double Higgs ones. Using the Run-2 data and combining several decay channels, the CMS and ATLAS experiments have been able to establish constraints at 95 %CL observed (expected) exclusion limits of $-3.5\ (-5.1) < \kappa_\lambda < 14.5\ (13.7)$ [385] and $-3.2\ (-6.2) < \kappa_\lambda < 11.9\ (14.4)$ [386] respectively. While not to the level of most Run-2 HH results of Table 4.1, they equal the ones of the least sensitive channels. It is however important to point out that all other couplings were set to their SM value, and the sensitivity is degraded a lot when the couplings to fermions and especially the weak boson couplings are determined together with κ_λ in a multi-dimensional fit. This effect however is expected to be similar for HH searches, for which the much lower sensitivity to κ_t compared to single Higgs measurements should have a significant impact in a κ_λ-κ_t two-dimensional

fit. This can be alleviated considering a combination of single and double Higgs measurements, such as was performed by the ATLAS collaboration using the Run-2 data, resulting in a 95 % CL of $-0.4\ (-1.9) < \kappa_\lambda < 6.3\ (7.6)$ [372], as illustrated on Fig. 4.4a. This result was however considering all other couplings set to their SM values, and when the coupling modifiers to heavy quarks, leptons and weak bosons are relaxed, the 95 % CL becomes $-1.4\ (-2.2) < \kappa_\lambda < 6.1(7.7)$. One advantage of including single Higgs measurements is that the constraints on κ_λ remain almost untouched when relaxing the other couplings, thanks to the correlation between the couplings, as was already observed in a previous combination with fewer data [387]. As illustrated on Fig. 4.4b, when a global fit of κ_λ and κ_t is performed the double Higgs measurements alone are not able to constrain κ_t, while combining them with the single Higgs measurements allows recovering the constraint to the level of the latter (although the position of the best fit value is such that it is slightly less stringent in values below unity).



(a) (b)

Figure 4.4 | Observed limits on the combination of single and double Higgs measurements performed by the ATLAS experiment using Run-2 data, from Ref. [372]. Left : observed values of the test statistics $-2\ln\Lambda$ as a function of κ_λ using pure single Higgs (blue) and pure double Higgs (red) measurements, as well as their combination assuming modifications of κ_λ only (black), or allowing the fit to determine κ_t, κ_b, κ_V and κ_τ (green). Right : 68 % and 95 % observed CL on both the κ_λ and κ_t coupling modifiers, using pure single Higgs (blue) or pure double Higgs (red) measurements, and their combination (black).

Similarly, the CMS collaboration performed a combined measurement of HH and $t\bar{t}H$ searches using Run-2 data, albeit only in the $bb\gamma\gamma$ final-state [378] for now, which confirmed the better κ_t constraints observation and the improvement of the κ_λ 95 % CL constraint to $-2.7\ (-3.3) < \kappa_\lambda < 8.6(8.6)$, compared to the result in Table 4.1. While the combination of single and double Higgs searches are challenging, since their orthogonality must be ensured, these results validate the motivation to pursue such endeavours. Further improvements of single Higgs measurements could include differential measurements [388], off-shell Higgs production [389], or

a global EFT fit [390]. The latter could prove more difficult than in HH searches, as the contribution of κ_λ enter at NLO and might be overshadowed by even small contributions from LO operators. Compared to the five-dimensional coupling space of HH GGF production described in Section 1.7.2, single Higgs production requires nine coefficients on top of the Higgs self-interaction [391], which would make the latter unconstrained in a global fit (although it would help in the measurement of the other coefficients), unless differential measurements are included [62].

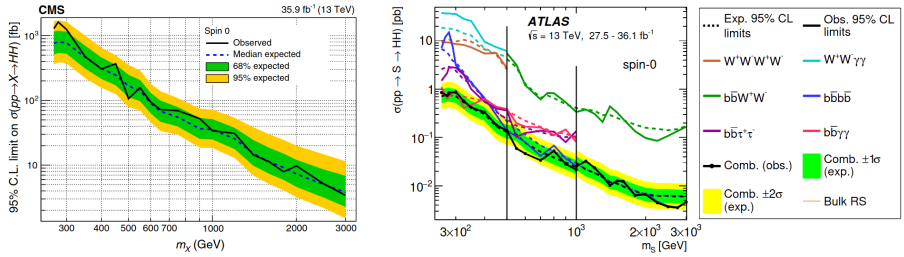


Figure 4.5 | Combination of HH searches in CMS (left, from Ref. [382]) and ATLAS (right, from Ref. [372]) during 2016 data taking year, combined over the different channels considered, with the expected and observed limits on the inclusive cross-section of a spin-0 resonance as a function of its hypothetical mass.

So far, the discussion of the results was focused on non-resonant and EFT parameter interpretation. However, many of the aforementioned analyses have either produced upper limits on a resonant scenario using part of the workflow dedicated to the non-resonant interpretation (this will be the case for the bbWW analysis described in this document), or they were produced in a parallel analysis that followed the same guidelines, hence they will not be discussed in details. The analyses focus on spin-0 and spin-2 resonant models, although only the former can really be interpreted as a model-independent search (if its width is narrow, cf. Section 1.6). Hence, in the comparison of the 2016 results on Fig. 4.5, only the spin-0 results are displayed, as spin-2 results depend on some parameters that make comparison more complicated. The resonance decay is typically a $X \to$ HH decay type, with mass m_{X} ranging from 250 GeV to several TeVs. Analyses focused on the resolved signatures of Higgs boson decay are more sensitive to the lower mass range, while boosted topologies perform better in the higher mass range. They usually deal with the range they are more suited to, the boundary is usually around 1 TeV. For example, the boosted bbWW search [392] focused on the range $0.8 \leq m_{\mathrm{X}} \leq 4.5$ TeV. In case a potential spin-0 scalar particle Y mixes with the SM Higgs boson, but its mass is below the threshold of $2m_{\mathrm{H}}$ for on-shell production of a pair of Higgs boson, then the type of decay $\mathrm{X} \to \mathrm{YH}$ could become dominant in certain channels (notably the bbWW one). Several of these searches are being carried out (for example the CMS bb$\tau\tau$ [393] and bb$\gamma\gamma$ [394] analyses), but are out of the scope of this document.

4.2 Analysis setup and event selection

The analysis presented in this chapter focuses on the bbWW final state, the second-highest BR of the Higgs pair decay, where at least one W boson decays leptonically. The different decays of the W boson pair define two main channels : the Double Lepton (DL) channel where both decay leptonically with final state $\mathrm{b\bar{b}}\ell^+\nu\ell^-\bar{\nu}$ and the Single Lepton (SL) channel where one of them decays hadronically with final state $\mathrm{b\bar{b}}\ell\nu q\bar{q}$. Additionally, to the bbWW final state, the bbZZ and $\mathrm{bb}\tau\tau$ final states can end up in our SR, albeit being a subdominant contribution. The total BR is computed as such

$$\begin{aligned}
\mathrm{BR(HH \to bbWW \to b\bar{b}}\ell^+\nu\ell^-\bar{\nu}) &= 2.\mathrm{BR(H \to b\bar{b}).BR(H \to WW).}[\mathrm{BR(W} \to \ell\nu_\ell)]^2 &= 2.64\ \% \\
\mathrm{BR(HH \to bbZZ \to b\bar{b}}\ell^+\nu\ell^-\bar{\nu}) &= 2.\mathrm{BR(H \to b\bar{b}).BR(H \to ZZ).2.BR(Z} \to \ell\bar{\ell}).\mathrm{BR}(Z \to \nu\bar{\nu}) &= 0.12\ \% \\
\mathrm{BR(HH \to bb}\tau\tau \to \mathrm{b\bar{b}}\ell^+\nu\ell^-\bar{\nu}) &= 2.\mathrm{BR(H \to b\bar{b}).BR(H} \to \tau^-\tau^+).\mathrm{BR}(\tau \to \ell\nu_\ell\nu_\tau)^2 &= 0.89\ \% \\
\mathrm{BR(HH \to bbWW \to b\bar{b}}\ell\nu q\bar{q}) &= 2.\mathrm{BR(H \to b\bar{b}).BR(H \to WW).2.BR(W} \to \ell\nu).\mathrm{BR(W} \to q\bar{q}) &= 10.94\ \%
\end{aligned} \tag{4.1}$$

where the Higgs BRs were taken from Ref. [63], and the other SM particles from Ref. [6]. While $\mathrm{bb}\tau\tau$ final state samples contain both taus decaying leptonically and hadronically, in order to remain exclusive with respect to the $\mathrm{bb}\tau\tau$ final state in Ref. [381], a hadronic decay veto is applied. The corresponding BR in Eq. (4.1) therefore uses the fact that $\mathrm{BR}(\tau \to q\bar{q}) \simeq 4 \times \mathrm{BR}(\tau \to \ell\nu_\ell\nu_\tau)$, while in practice, the inclusive value is used and the events are filtered out.

The total BRs of the DL and SL channels amount to 3.65 % and 10.94 %, but the latter suffers from a larger background contamination. The expected number of GGF signal events, notwithstanding detector acceptance, is then

$$\begin{aligned}
N_{\mathrm{HH(GGF)}} &= \sigma_{\mathrm{HH(GGF)}} \times \mathrm{BR(HH} \to \mathrm{b\bar{b}}\ell^+\nu\ell^-\bar{\nu} \text{ or } \mathrm{b\bar{b}}\ell\nu q\bar{q}) \times L \\
&= 31.05\ \mathrm{fb}^{-1} \times 14.6\ \% \times 138\ \mathrm{fb} \simeq 625 \text{ events.}
\end{aligned} \tag{4.2}$$

In contrast to the expected background contributions of $\sim 10^9$ events, the signal-to-background ratio is tiny and the choice of the analysis strategy is crucial to achieve decent sensitivity. Particles are identified in their final state, and assigned if possible to parent particles in the decay chain, signal extraction is performed through using ML methods and the final evaluation is performed via a profile-likelihood as described in Section 2.4.3. The strategy benefits from a pair of b-jets with an invariant mass peaking around the Higgs mass compared to a smooth background distribution, and on which b-tagging can drastically improve the signal sensitivity. The leptons provide a clear signature to trigger the recording of data events, yet they typically have relatively low-p_T and invariant mass due to the spin-0 signature of the Higgs boson and the 4-body decay. Based on this, the event selection described in Section 4.2.4 reduces the background contamination by two to three orders of magnitude (cf. Table 4.2).

Compared to the previous analysis in this decay channel [395] that was using only 2016 data, in this analysis the whole Run-2 dataset was used, the SL channel has been introduced and a specific study of boosted topologies (when the $H \to bb$ is produced with a large transverse momentum resulting in a single merged jet) has been included to benefit the resonant production at high mass.

The main SM processes contributing to the considered event topology are top quark pair production ($t\bar{t}$), $Z^*/\gamma \to \ell^+\ell^-$ (Drell-Yan) and W decay, both with jets associated production. The $t\bar{t}$ process shows the exact same final state as the signals and is denoted as *irreducible*, only using clever requirements on the event kinematics can its contribution be reduced. Drell-Yan (DY) and W boson decay (W+jets) with associated jets are *reducible* as their contributions can be reduced by applying b-tagging cuts, except when the associated jets are from b quarks, in which case the same procedure as for the $t\bar{t}$ must be used. The DY+jets is a major background in the DL channel, yet very few simulated events pass the requirements of the SR and the statistical precision of the analysis is largely impacted. To circumvent that, a data-driven method has been developed. Additional backgrounds consist in single-top quark in the t- and s-channels (ST), single top with W or Z associated production (tW, tZ, denoted as tV in the following), diboson production (ZZ, ZW, WW, Wγ and Zγ, denoted VV), triboson production (similarly denoted VVV), $t\bar{t}$ and associated vector boson production ($t\bar{t}W$ and $t\bar{t}Z$ with or without associated Higgs, W or jet, denoted $t\bar{t}VX$), various single Higgs production processes (through GGF (ggH), VBF (qqH), $t\bar{t}H$, WH, ZH and tH with associated W or jet), and minor exotic processes. Experimental backgrounds due to jets misidentified as leptons is non-negligible in the SL channel and has been derived using a data-driven approach.

Table 4.2 | Cross-sections and their references of the main processes involved in the analysis with their expected yield, assuming 138 fb^{-1} integrated luminosity. The total number of events before and after analysis selections in the DL and SL channels are given with the associated efficiencies ϵ. For the signal GGF and VBF SM processes, the total yield is split between the $bbWW \to b\bar{b}\ell^+\nu\ell^-\bar{\nu}$ and $b\bar{b}\ell\nu q\bar{q}$ decays (cf. Eq. (4.1)).

Process	Cross-section [pb]	Total yield	DL channel		SL channel		Reference
			Yield	ϵ [%]	Yield	ϵ [%]	
$t\bar{t}$ (fully leptonic)	88.4	$1.2\ 10^7$	$1.5\ 10^6$	12.5	$1.5\ 10^6$	12.5	[6, 396]
$t\bar{t}$ (semi leptonic)	365.52	$5.0\ 10^7$	$1.9\ 10^2$	<0.01	$1.1\ 10^7$	21.8	[6, 396]
$t\bar{t}$ (fully hadronic)	377.85	$5.2\ 10^7$	0	0.0	$2.9\ 10^2$	<0.01	[6, 396]
Drell-Yan	6077.22	$8.4\ 10^8$	$2.0\ 10^5$	0.02	$4.2\ 10^5$	0.05	[397, 397]
W+jets	61526.7	$8.5\ 10^9$	$2.1\ 10^2$	<0.01	$2.8\ 10^6$	0.03	[398]
ST	292.04	$3.9\ 10^7$	$7.9\ 10^4$	0.2	$1.3\ 10^6$	3.4	[399]
WW	62.87	$8.7\ 10^6$	$2.2\ 10^3$	0.03	$5.5\ 10^4$	0.63	[398, 400]
ZW	10.03	$1.4\ 10^6$	$1.2\ 10^3$	0.09	$2.2\ 10^3$	0.16	[398, 400]
ZZ	6.78	$9.4\ 10^5$	$2.5\ 10^3$	0.27	$5.5\ 10^3$	0.58	[398]
$t\bar{t}W$	0.60	$8.3\ 10^4$	$2.1\ 10^3$	2.51	$1.3\ 10^4$	15.36	[6, 63, 399]
$t\bar{t}Z$	0.95	$1.3\ 10^5$	$26\ 10^3$	2.00	$1.6\ 10^4$	12.40	[6, 63, 399]
$HH \to bbWW$ (GGF)	0.03105	113.1 / 468.8	15.0	9.56	96.1	20.50	Eq. (1.53)
$HH \to bbWW$ (VBF)	0.00173	6.3 / 26.1	0.55	8.78	3.83	14.69	Eq. (1.53)

4.2.1 Samples

The data consists in pp collisions at $\sqrt{s} = 13$ TeV center-of-mass energy and 25 ns bunch crossing period. Only the luminosity sections of data that have been certified as good to be used in analysis have been considered. The integrated luminosities amount to 36.3 fb^{-1} in 2016, 41.5 fb^{-1} in 2017 and 59.7 fb^{-1} in 2018, amounting to 138 fb^{-1} for the full Run-2 dataset with an average pileup of around 30 inelastic pp collisions per bunch crossing.

Data events are selected through the HLT triggers on Table 4.3. SL events are required to pass the single-electron or muon triggers, while the sensitivity of the DL is increase by including a combination of single- and double-lepton triggers, the latter being programmed with lower p_{T} thresholds, beneficial for the soft leptons produced from the Higgs decay chain. Events containing two offline reconstructed electrons (muons) are required to pass either the single-electron (single-muon) or the double-electron (double-muon) trigger, while events containing both offline reconstructed electron and muon are required to pass either the single-electron, the single-muon, or the electron + muon trigger.

In some triggers, a combination of HLT paths was used, with and without the d_z requirement for the leptons, and for the case of the double muon triggers with and without the threshold on the invariant mass $m_{\mu\mu} > 3.8$ GeV. The motivation for this combination is to use the triggers with the highest efficiency when available, and lower efficiency triggers when the former are not available or prescaled. These effects are taken into account using dedicated SFs.

Data events are separated into several Primary Dataset (PD) : DoubleMuon, MuonEG, SingleMuon, DoubleEG, SingleElectron for data recorded in 2016 and 2017, and DoubleMuon, MuonEG, SingleMuon, EGamma for data recorded in 2018. Each event will enter one or several of the PDs if one of its trigger path was fired. To remove the overlap, a ranking is performed following a priority list, requiring the event to be selected from the dataset with the highest priority.

Background events were modelled with either LO or NLO using matrix elements modelled using MG5_aMC@NLO, together with POWHEG. To enhance the statistics of the DY, inclusive samples covering the whole phase space and complementary samples binned in jet multiplicity at generator level were combined using a *stitching* [401] method.

Parton shower, hadronisation, τ decays and polarization effects were modelled using the generator PYTHIA as described in Section 1.3. The UE tunes *CUETP8M1*, *CUETP8M2*, *CUETP8M2T4*, or *CP5* [402–404] were used for the samples covering the 2016 data-taking period, and the tune *CP5* was used solely for those covering the 2017 and 2018 data-taking periods. The PDFs used in 2016 data-taking period is mostly NNPDF3.0 [405–409], while in the 2017 and 2018 data-taking eras the NNPDF3.1 set was used. All the samples are produced assuming $m_{\mathrm{t}} = 172.5$ GeV

Table 4.3 | List of HLT triggers used in the analysis. The channels are denoted by either "Mu" or "Ele" for muons and electrons respectively, or their combination, while the numbers refer to the p_{T} cuts applied on the objects. Additional requirements might be added as isolation ("Iso"), pseudorapidity cuts ("eta"), longitudinal d_z requirements ("DZ") or invariant mass ("mass"). Checkmarks (✓) and hyphens (−) are added after each HLT path to indicate whether the path was enabled or disabled during the 2016, 2017, and 2018 data-taking periods.

Single-electron triggers	HLT_Ele25_eta2p1_WPTight_Gsf (✓/−/−)
	HLT_Ele27_WPTight_Gsf (✓/−/−)
	HLT_Ele27_eta2p1_WPLoose_Gsf (✓/−/−)
	HLT_Ele32_WPTight_Gsf (−/✓/✓)
	HLT_Ele35_WPTight_Gsf (−/✓/−)
Single-muon triggers	HLT_IsoMu22 (✓/−/−)
	HLT_IsoTkMu22 (✓/−/−)
	HLT_IsoMu22_eta2p1 (✓/−/−)
	HLT_IsoTkMu22_eta2p1 (✓/−/−)
	HLT_IsoMu24 (✓/✓/✓)
	HLT_IsoTkMu24 (✓/−/−)
	HLT_IsoMu27 (−/✓/✓)
Double-electron triggers	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ (✓/−/−)
	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL (−/✓/✓)
Double-muon triggers	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL (✓/−/−)
	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ (✓/−/−)
	HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL (✓/−/−)
	HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ (✓/−/−)
	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8 (−/✓/−)
	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8 (−/✓/✓)
Electron + muon triggers	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL (✓/−/−)
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ (✓/✓/✓)
	HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL (✓/−/−)
	HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_DZ (✓/−/−)
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ (✓/✓/−)
	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ (✓/✓/−)
	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL (✓/✓/−)
	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ (−/✓/−)

and $m_{\mathrm{H}} = 125$ GeV.

The same procedure was applied to generate signal events, using NLO for non-resonant GGF production and LO for non-resonant VBF production and resonant spin-0 and spin-2 samples. Each decay channel is represented by a distinct signal sample with its associated BR in Eq. (4.1), except for HH → bbWW → $\mathrm{b\bar{b}}\ell^+\nu\ell^-\bar{\nu}$ and HH → bbZZ → $\mathrm{b\bar{b}}\ell^+\nu\ell^-\bar{\nu}$ that are merged as their final state is indistinguishable.

For the resonant analysis, 18 signal samples, with masses ranging from 250 to 900 GeV are used. The low mass region corresponds to narrow width resonance decay to a pair of Higgs, while close to the TeV scale the products of H → bb are very collimated and require the consideration of merged jets topology. The non-resonant analysis uses 4 GGF at NLO and 6 VBF at LO samples, with different values of the couplings defined in Sections 1.5.2 and 1.7.2. Their cross-sections are listed in Table 4.4 and Table 4.5 respectively. Assuming SM behaviour, this limited set of points allows for a continuous search over the spectrum of the coupling modifiers, as described in Section 1.7.3. Additionally, a matrix element reweighting can be applied to the GGF samples, following Sections 1.3 and 1.7.2, to access any combination in the non-linear EFT $(\kappa_t, \kappa_\lambda, c_g, c_{2g}, c_2)$ parameter space.

Table 4.4 | The production cross-section of the non-resonant double Higgs GGF signal samples at NLO precision with a k-factor correction to achieve NNLO-level prediction, as a function of the trilinear coupling modifier defined in Section 1.5.2. Uncertainties are included as relative corrections split between contributions that are constant (PDF and strong coupling α_S variations, from Ref. [410]) and those that vary with κ_λ (scale and top mass approximation uncertainties, from Ref. [411]).

κ_λ	σ [fb]	PDF	α_S	Scale	m_{t}
1	31.0	±2.1 %	±2.1 %	$^{+2.2\ \%}_{-5.0\ \%}$	$^{+4\ \%}_{-18\ \%}$
0	69.7	±2.1 %	±2.1 %	$^{+2.4\ \%}_{-6.1\ \%}$	$^{+6\ \%}_{-12\ \%}$
2.45	13.1	±2.1 %	±2.1 %	$^{+2.3\ \%}_{-5.1\ \%}$	$^{+4\ \%}_{-22\ \%}$
5	91.2	±2.1 %	±2.1 %	$^{+4.9\ \%}_{-8.8\ \%}$	$^{+13\ \%}_{-4\ \%}$

Pileup is simulated through overlaying minimum bias events simulated with PYTHIA, calibrated with the luminosity and with a pp inelastic cross-section of 69.2 mb, as described in Section 2.4.3 together with the discussion about the detector simulation.

4.2.2 Object selection

Leptons

The definition of the electrons and muons in the analysis follows three stages : "loose", "fakeable" and "tight" selections. The loose leptons are used to veto pairs containing low masses, the fakeable leptons are used to remove the overlap between

Table 4.5 | The production cross-section of the non-resonant double Higgs VBF signal samples that were produced at LO, with a LO-to-N^{3}LO k-factor (with a data-taking year dependency), as a function of the coupling modifier in the non-linear EFT framework described in Section 1.7.2. The uncertainties are included independently of the couplings, with a value of $\pm$2.1% for the combination of the PDF and strong coupling α_S, and $^{+0.03\ \%}_{-0.04\ \%}$ for the scale uncertainties (from Ref. [62]).

κ_V	κ_{2V}	κ_λ	σ (2016) [fb]	σ (2017, 2018) [fb]
1	1	1	1.73	
1	1	0	4.59	4.61
1	1	2	1.43	1.42
1	2	1	14.4	14.2
0.5	1	1	10.9	10.8
1.5	1	1	66.3	66.0
1	0	1	27.2	27.1

objects – also known as *cleaning* – and to perform the estimation of the fake lepton background from CRs of the data, and the tight leptons are similar to the fakeable with the addition of more strict selection criteria to be used as candidate finale states of a Higgs decay in the SR. These lepton criteria were defined in the CMS $t\bar{t}H$ multilepton analysis [412], using a MVA discriminant trained on charged and neutral particles reconstructed in a cone around the lepton direction and the lepton itself. Several observables related to the lepton and the reconstructed jets within the cone allow for the differentiation between prompt leptons originating from the hard process, from leptons originating from b hadron decays. This MVA will be denoted as the *lepton MVA*, and is flavour dependent.

To avoid any potential bias effect in the estimation of the fake lepton background, a modified p_{T} variable called cone-p_{T} is used in the definition of the electrons and muons. This variable resumes to the classical p_{T} – that will also be referred to as reco-p_{T} in case of ambiguity – if the electron (muon) passes the electron (muon) MVA tight WP. If not, it is defined as $0.9 \times (p_{\mathrm{T}}^{\mathrm{lep}} + I^{\mathrm{lep}})$ as function of the lepton p_{T} and relative isolation if there are no jets closer than $\Delta\mathrm{R} = 0.4$ to the lepton, and as to 0.9 times the lepton-subtracted jet p_{T} otherwise. This quantity, that only differs from the reco-p_{T} at the fakeable level, was introduced for the case where a bottom (anti)quark hadronises and one of the subsequent hadron decays gives rise to a lepton. In that case, the lepton will not carry the same reco-p_{T} as the initial b quark, while the cone-p_{T} is a more correct estimate and allows the correct weight assignment when performing the fake lepton estimation.

The selections in Tables 4.6 and 4.7 encompass cone-p_{T} and η of the lepton, its isolation, the transverse ($|d_{xy}|$) and longitudinal ($|d_z|$) impact parameters of the lepton track with respect to the event vertex as well as the significance of the impact parameter in three dimensions (d/σ_d), the b-tagging discriminant of the jet

Table 4.6 | Loose, fakeable and tight selection criteria for electrons. The electron WPs discussed in Section 2.4.2, as well as the observables $\sigma_{i\eta i\eta}$, H/E, and 1/E - 1/p are varied as function of the electron candidate pseudorapidity η. The conditions on the electron WP, the Deep Jet discriminant and the relative isolation of the nearest jet to the electron are tightened (relaxed) for fakeable electrons that fail (pass) the requirement on the electron MVA > 0.30, in order to reduce the systematic uncertainty on the fake lepton background estimate on the jet flavour composition, following Ref. [412]. A hyphen ($-$) indicates selection criteria that are not applied.

Electrons			
Observable	Loose	Fakeable	Tight
Cone-p_T	$>$ 7 GeV	$>$ 10 GeV	$>$ 10 GeV
$\lvert\eta\rvert$	< 2.5	< 2.5	< 2.5
$\lvert d_{xy}\rvert$	< 0.05 cm	< 0.05 cm	< 0.05 cm
$\lvert d_z\rvert$	< 0.1 cm	< 0.1 cm	< 0.1 cm
d/σ_d	< 8	< 8	< 8
I_{e}	$< 0.4 \times p_T$	$< 0.4 \times p_T$	$< 0.4 \times p_T$
$\sigma_{i\eta i\eta}$	–	$<$ { 0.011 / 0.030 }[1]	$<$ { 0.011 / 0.030 }[1]
H/E	–	< 0.10	< 0.10
1/E - 1/p	–	> -0.04	> -0.04
Conversion rejection	–	✓	✓
Missing hits	≤ 1	$= 0$	$= 0$
EGamma POG MVA	>WP-loose[2]	>WP-90 (>WP-loose) [2]†	>WP-loose[2]
Deep Jet of nearby jet	–	<WP-tight (<WP-medium) [3]	<WP-medium[3]
Jet relative isolation[4]	–	< 0.7 ($-$) †	–
Electron MVA	–	–	> 0.30

[1] Barrel / endcaps.
[2] WPs as discussed in Section 2.4.2.
[3] WPs as discussed in Section 2.4.2.
[4] Defined as $p_{\text{T}}^{\text{jet}}/p_{\text{T}}^{\text{e}} - 1$ if the electron is matched to a jet within $\Delta\text{R} < 0.4$ or as the PF-relative isolation computed within a cone of fixed size $\Delta\text{R} = 0.4$ otherwise.
† Fails (passes) the requirement prompt-e MVA > 0.30.

nearest to the lepton, and the lepton MVA score. Additionally, electrons passing the fakeable selection are further required to satisfy a set of conditions on the width of the electron cluster in η-direction ($\sigma_{i\eta i\eta}$), the ratio of energy associated to the electron in the HCAL to the energy in the ECAL (H/E), and the difference between the reciprocal of the electron cluster energy and the reciprocal of its track momentum (1/E-1/p). These conditions mimic the electron identification criteria applied at trigger level.

Hadronic taus are reconstructed as described in Section 2.4.2, candidates are considered if they have $p_{\text{T}} > 20$ GeV, $|\eta| < 2.3$, and pass the medium WP of the DeepTau algorithm. They are further cleaned with respect to leptons passing fakeable selection with an overlap inside the cone $\Delta\text{R} < 0.3$. A hadronic tau veto is used in the SL channel to ensure orthogonality with the other HH analyses.

Table 4.7 | Loose, fakeable and tight selection criteria for muons. The conditions on the Deep Jet discriminant and the relative isolation of the nearest jet to the muon are tightened (relaxed) for fakeable muons that fail (pass) the requirement on the muon MVA > 0.50, in order to reduce the systematic uncertainty on the fake lepton background estimate on the jet flavour composition, following Ref. [412]. A hyphen ($-$) indicates selection criteria that are not applied.

Muons			
Observable	Loose	Fakeable	Tight
p_T	> 5 GeV	> 10 GeV	> 10 GeV
$\|\eta\|$	< 2.4	< 2.4	< 2.4
$\|d_{xy}\|$	< 0.05 cm	< 0.05 cm	< 0.05 cm
$\|d_z\|$	< 0.1 cm	< 0.1 cm	< 0.1 cm
d/σ_d	< 8	< 8	< 8
I_μ	$< 0.4 \times p_T$	$< 0.4 \times p_T$	$< 0.4 \times p_T$
PF muon	>WP-loose[1]	>WP-loose [1]	>WP-medium[1]
Deep Jet of nearby jet	–	<WP-interp. (<WP-medium) [2]†	<WP-medium[2]
Jet relative isolation[3]	–	< 0.8 ($-$) †	–
Muon MVA	–	–	> 0.5

[1] WPs as discussed in Section 2.4.2.
[2] Upper cut on the Deep Jet score defined with a linear interpolation from Deep Jet WP-medium at cone-p_{T} 20 GeV to Deep Jet WP-loose at cone-p_{T} 45 GeV, taking the Deep Jet WPs as discussed in Section 2.4.2.
[3] Defined as $p_{\mathrm{T}}^{\mathrm{jet}}/p_{\mathrm{T}}^{\mu} - 1$ if the muon is matched to a jet within $\Delta\mathrm{R} < 0.4$ or as the PF-relative isolation computed within a cone of fixed size $\Delta\mathrm{R} = 0.4$ otherwise.
† Fails (passes) the requirement prompt-μ MVA > 0.5.

Jets

AK4 and AK8 jets considered in the analysis follow the reconstruction methods described in Section 2.4.2 and their corrections in Section 2.4.3.

Both AK4 and AK8 jets are required to pass the loose WP in 2016, and tight in 2017 and 2018 data-taking eras, with $|\eta| < 2.4$ and $p_{\mathrm{T}} > 25$ GeV for the AK4 and $p_{\mathrm{T}} > 25$ GeV for the AK8. The latter also has to contain two subjets with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.4$, a SD mass in the range $30 < m_{SD} < 210$ GeV. The *N-subjettiness* [413] variables that quantify the AK8 jet composition must follow $\tau_2/\tau_1 < 0.75$, to ensure with good confidence its "two-prong" structure.

A cleaning procedure is applied, consisting in removing the jets from consideration that overlap with fakeable electrons or muons within a cone $\Delta\mathrm{R} < 0.4$ for AK4 jets, and $\Delta\mathrm{R} < 0.8$ for AK8 jets. The VBF production mode presents an important feature in the form of two hadronic additional jets aligned along the beam line, therefore at high pseudorapidity with large angular deviation from each other. To increase the sensitivity to that production mode in the non-resonant analysis, the

signature of these jets can be used to define another category of jets called VBF jets. They are identical to the previously defined jets with the exception that the pseudorapidity requirement is relaxed to $|\eta| < 4.7$ and p_{T} threshold raised to $p_{\mathrm{T}} >$ 30 GeV, except in the $2.7 < |\eta| < 3.0$ region where it is raised further to 60 GeV to avoid spurious jets arising from noise in the ECAL endcaps. The VBF jets are cleaned with respect to the $\mathrm{H} \to \mathrm{bb}$ jet candidates. They are removed from consideration if they are located within a cone defined by $\Delta\mathrm{R} < 0.8$ from the two AK4 jet candidates, or by $\Delta\mathrm{R} < 1.2$ in case of a merged AK8. In addition, in the SL, they are further cleaned with respect to candidate $\mathrm{W} \to q\bar{q}$ jets fulfilling $|m_{\mathrm{jj}} - 80.4\ \mathrm{GeV}| < 15$ GeV, also with $\Delta\mathrm{R} < 0.8$. VBF jet pairs are only considered if they fulfil $m_{\mathrm{jj}} > 500$ GeV and $\Delta\eta_{jj} > 3$.

To remove potential soft pileup jet effects, the pileup identification criterion defined in Section 2.4.3 is applied to any jet with $p_{\mathrm{T}} < 50$ GeV, including VBF jet candidates.

The DeepJet algorithm described in Section 2.4.2 is used to identify b-jet candidates among the AK4 jets, and the DeepCSV is used on subjets of the AK8 jets. For a AK8 jet to be considered b-tagged, at least one of its subjet must be b-tagged and have $p_{\mathrm{T}} > 30$ GeV. In the analysis, only the medium b-tagging WP is used for both jet types.

4.2.3 Event level quantities

Missing energy

The Missing Transverse Energy (MET) has been described in Section 2.4.2 and its Type-0 and Type-1 corrections in Sections 2.4.2 and 2.4.3, although only the latter was used. The magnitude of the MET is referred to as E_T^{miss}. In 2017 a correction had to be introduced to mitigate the effect of the ECAL noise in high-η regions.

To mitigate the impact of the pileup, a new variable H_{T}^{miss} is introduced and only defined on the selected jets and leptons passing the fakeable selection, such that any soft hadrons are not taken into account. While its resolution is worse compared to E_T^{miss}, it has the benefit of being less biased at high pileup. Similarly, these two variables are linearly combined in $E_T^{miss}LD$ (for Linear Discriminant) defined as

$$E_T^{miss}LD = 0.6 \times E_T^{miss} + 0.4 \times H_{\mathrm{T}}^{miss}. \tag{4.3}$$

While the two quantities E_T^{miss} and H_{T}^{miss} are correlated, their degree of correlation depends on the origin of the missing energy in the transverse plane. If the missing energy comes from a genuine contribution such neutrinos, they will be highly correlated, while instrumental effects will solely be visible in E_T^{miss}, therefore this correlated variable can contain meaningful information.

Heavy mass estimator

In the DL channel, the decay into two neutrinos of $\mathrm{H} \to \mathrm{WW}^* \to \ell^+\nu_\ell\ell^-\bar{\nu}_\ell$ prevents the full reconstruction of the Higgs boson resonance even when considering the MET, a major handicap in the resonant analysis. A novel technique called the Heavy Mass Estimator (HME) [414] attempts at such reconstruction using kinematic constraints and a probabilistic interpretation.

In the topology $X \to HH \to b_1b_2W_1W_2 \to b_1b_2l_1\nu_{l1}l_2\nu_{l2}$, W_1 and W_2 denote the on- and off-shell bosons. Event kinematics are described by

$$E_{T,x}^{miss} = p_x(\nu_{\ell_1}) + p_x(\nu_{\ell_2}), \tag{4.4}$$

$$E_{T,y}^{miss} = p_y(\nu_{\ell 1}) + p_y(\nu_{\ell_2}), \tag{4.5}$$

$$\sqrt{p^2(\ell_1, \nu_{\ell_1})} = m_{\mathrm{W}}, \tag{4.6}$$

$$20 < \sqrt{p^2(\ell_2, \nu_{\ell_2})} < 45 \text{ GeV}, \tag{4.7}$$

$$\Big(p(\ell_1) + p(\ell_2) + p(\nu_{\ell_1}) + p(\nu_{\ell_2})\Big)^2 = m_{\mathrm{H}}^2, \tag{4.8}$$

$$\left(p(\mathrm{b}_1) + p(\mathrm{b}_2)\right)^2 = m_{\mathrm{H}}^2, \tag{4.9}$$

$$m_{\mathrm{X}} = \Big(p(\ell_1) + p(\ell_2) + p(\nu_{\ell_1}) + p(\nu_{\ell_2}) + p(\mathrm{b}_1) + p(\mathrm{b}_2)\Big)^2 \tag{4.10}$$

where $m_{\mathrm{H}} = 125$ GeV, b_1 and b_2 are two AK4 jets in the resolved categories, or the two subjets of one AK8 jets in the boosted category as they provide a better estimation than when the whole AK8 is used. Each neutrino accounts for three unknown momentum projections, and the problem then has six degrees of freedom. Eqs. (4.4) to (4.6) and (4.8) allow reducing it to two degrees of freedom.

These degrees of freedom are arbitrarily selected as the pseudorapidity η_ν and azimuthal angle ϕ_ν of one of the neutrino. At each iteration of the algorithm, the two values are sampled from a flat distribution, and Eq. (4.10) is used to estimate the heavy resonance mass, assuming both its origin from the on-shell W boson following Eq. (4.6) where m_{W} is sampled from a wide mass probability density function determined from Monte Carlo simulations, or off-shell W boson with constrain from Eq. (4.7). The other neutrino kinematics can be inferred from the other equations, although only solutions for each configuration that follow kinematic constraints are kept. Thousands of iteration of the algorithm for an event yields a likelihood function, and the most probable value is defined as the HME. The distribution in data and simulations of the HME is on Fig. 4.6.

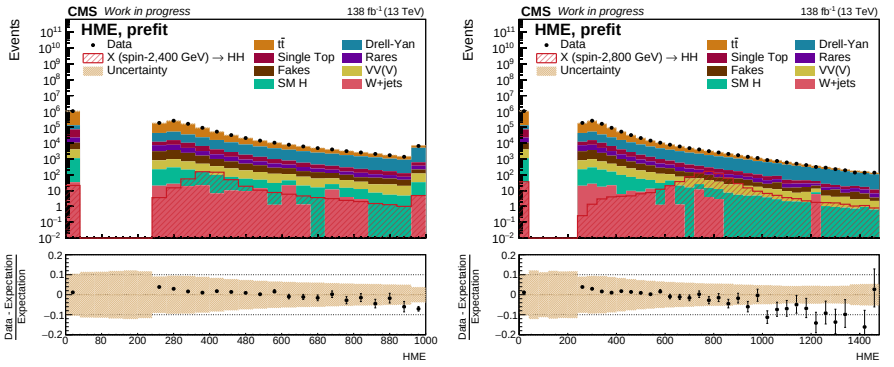


Figure 4.6 | HME distributions using the 138 fb^{-1} of the Run-2 dataset, including two resonant signal samples with spin-0 and mass of 400 and 800 GeV. The events for which no solution was found in any iteration are assigned an arbitrary value of zero.

The double b-jet resonance in Eq. (4.9) is supposed to have an invariance mass at the Higgs boson mass whose value is sampled from a narrow Gaussian centred around 125 GeV. The leading jet transverse momentum is corrected for the reconstruction resolution by sampling over a calibrated pdf, the other jet being rescaled to match the sampled Higgs mass. Momentum corrections are propagated back to the MET, who is further smeared based on simulations. All the corrections and random samplings are illustrated in Appendix A.4, they ensure that no JEC can have an impact large enough to produce a significant change when many iterations are included. The expensive per-event computation can be fixed for any correction, preventing the need to perform it several times per event. Still, the algorithm is very time-consuming, therefore its evaluation was performed pre-emptively and fetched in a look-up table in the analysis workflow.

Parton-level center-of-mass energy

In the context of the search for new physics mass scale in missing energy events, a model-independent variable S_{min} [415–417] was defined as

$$S_{\text{min}} = \sqrt{2\ E_{\text{T}}^{\text{miss}} \times \left(E_{\text{T}}^{vis} - p_{\text{T}}^{vis} \cos(\Delta\phi) \right) + m_{vis}^{2}}, \tag{4.11}$$

where E_{T}^{vis}, p_{T}^{vis} and m_{vis}^{2} are the transverse energy, momentum and invariance mass of the visible particles system, while $\cos(\Delta\phi)$ refers to the angle in the transverse plane between this system and the vector of the MET. Applied in the SL channel, this variable represents the 5-particle system composed of the two b-jets, the two jets from the W boson hadronic decay and the lepton.

4.2.4 Event selection

After trigger selection, a few additional safety selections are applied on the MET quality criteria to exclude events polluted by beam halo effects, detector noise, and PV requirements to ensure good vertexing capabilities.

Pairs of electrons and muons passing loose selections are defined and events for which any pair of same flavour leptons have $m_{\ell\ell} < 12$ GeV are rejected to avoid quarkonia resonances that are not well reproduced in simulations. In addition, to remove the Drell-Yan contributions, events with a pair of loose and opposite-charge leptons with $|m_{\ell\ell} - m_{\mathrm{Z}}(= 91.2\text{ GeV})| < 10$ GeV are rejected.

In the DL channel the event must have two leptons passing the tight selection with opposite charge, the leading lepton must have cone-$p_{\mathrm{T}} > 25$ GeV and the subleading lepton cone-$p_{\mathrm{T}} > 15$ GeV, their pseudorapidity acceptance being $|\eta| <$ 2.5 (2.4) for electrons (muons). The event category is based on the H $\rightarrow$ bb decay. In the boosted category there must be at least one AK8 jet that is b-tagged, while in the resolved category there must be at least two selected AK4 jets, of which at least one must be b-tagged as defined in Section 4.2.2. The categorisation is further refined into a resolved category with only one and at least two b-tagged jets. The boosted category takes priority over the resolved one in case both signatures are present in the event.

In the SL channel the lepton must pass the tight selection, electrons must have cone-$p_{\mathrm{T}} > 32$ GeV and $|\eta| < 2.5$, while for the muon cone-$p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.4$. Events with more than one tight lepton are vetoed to be independent of the DL. The categories are then defined similarly to the DL with the additional hadronic decay of one W boson $W \rightarrow jj$. The resolved category must have at least three selected AK4 jets, out of which at least one is b-tagged, with the same subcategorisation as the DL. In the boosted category, at least one AK4 jet must be outside the cone defined by $\Delta\mathrm{R} < 1.2$ around the b-tagged AK8 jet.

This analysis targets the bbWW decay channel, together with the bbZZ and bb$\tau\tau$ decay channels that lead to the same final-states (cf. Eq. (4.1)). In order to be included in a future HH combination, several additional selections must be included to respect orthogonality : the lepton requirements for the HH $\rightarrow$ bbbb analyses [379, 380]; not more than two tight leptons in the event for the HH $\rightarrow$ bbZZ [376] (in the four-lepton final states) and multi-leptonic channels of the HH $\rightarrow$ WWWW/WW$\tau\tau$/$\tau\tau\tau\tau$ [377] analyses; and the hadronic tau veto of the SL channel for the HH $\rightarrow$ bb$\tau\tau$ [381] and hadronic tau channels of the HH $\rightarrow$ WWWW/WW$\tau\tau$/$\tau\tau\tau\tau$ [377] analyses.

4.2.5 Signal extraction

A DNN performing multi-classification over the events passing all selection criteria is used to define physics motivated categories. Associated distributions are used in the statistical inference, as described in Section 2.4.3. In addition to separating the signal from the various backgrounds, the advantage of a multi-classification is to isolate distinguishable background processes and improve their constraints during the likelihood minimisation. The signal and main background processes have their own categories. Minor backgrounds can be grouped, typically with processes of similar origin or topology, to obtain statistically significant contribution from each class. The categories are detailed in Table 4.8 for the DL channel, and in Table 4.9 for the SL channel.

Table 4.8 | DNN output nodes of the DL channel and their description, for both resonant and non-resonant analyses.

Process	Description	Resonant	Non-resonant
HH(GGF)	Gluon fusion Higgs boson pair	✓	✓
HH(VBF)	Vector boson fusion Higgs boson pair	X	✓
$t\bar{t}$	Top quark pair	✓	✓
ST	Single top quark	✓	✓
DY	Drell-Yan	✓	✓
H	Single Higgs boson	✓	✓
$t\bar{t}VX$	Top quark pair associate vector boson with possible additional vector or Higgs boson ($t\bar{t}V$, $t\bar{t}VV$, $t\bar{t}VH$)	✓	✓
VV(V)	Multiple vector boson (WW, ZW, ZZ, WWW, WWZ, WZZ, ZZZ)	✓	✓
Other DL	All other, among them W boson with associated jets	✓	✓

Table 4.9 | DNN output nodes of the SL channel and their description, for both resonant and non-resonant analyses.

Process	Description	Resonant	Non-resonant
HH(GGF)	Gluon-gluon fusion Higgs boson pair	✓	✓
HH(VBF)	Vector-boson fusion Higgs boson pair	X	✓
$t\bar{t}$	Top quark pair	✓	✓
ST	Single top quark	✓	✓
WJets	W boson with associated jets	✓	✓
H	Single Higgs boson	✓	✓
Other SL	All other, among them Drell-Yan	✓	✓

The input of the DNNs consists in a mixture of low-level variables taken directly from the input objects – leptons and b-jets – momenta, and high-level variables consisting in more complex hand-crafted variables. The multiclass output uses a softmax activation function – a generalisation of the sigmoid function for several output nodes – such that each event produces N_{classes} output values in the range [0,1] with a total sum of 1. Consequently, it can be interpreted as a probability for an event to belong to a certain class. The *winner-takes-all* procedure is applied such that the highest output value determines in which class the event will go on, and its score (also referred to as *DNN score*) is used to fill the associated distribution. Some of these variables distributions are on Figs. 4.7 and 4.8.

During the training each event is associated a weight similar to the one that is filled in the distribution, namely the Monte Carlo generator weight multiplied by all the corrections applied in the analysis chain, the cross-section, luminosity, and efficiency required to scale the simulations to data. Thereby it provides a measure of the "importance" of each event for the DNN, allowing for a better agreement with the data. However, negative event weights have no meaning for the training with respect to the loss function, especially if the batch size is small enough that the cancellation with positive weight events is not fulfilled. These negative weight events are thus ignored during the training, stabilising the training at the cost of a potential exaggerate representation of parts of the phase space that are suppressed in the ME. A final class reweighting must be performed so to not unbalance the training towards classes with large statistics. All the background categories are set to the same sum of weights, while the two signal categories are set to a lower value as downgrading their importance – with a factor ranging from $1/8$ to $1/20$ depending on the cases – showed to lower both signal and background yields such that the sensitivity is improved. The only difference between the resonant and non-resonant analysis is that the former lacks the HH(VBF) category and is built as a parametric DNN [291] as a function of the resonant mass m_{X}.

Each DNN is added another input block in the form of a Lorentz Boosted Network (LBN) [418], a neural network architecture using particles 4-momenta to automatically produce combinations and higher level features. It performs so by producing two sets of particles and rest frames momenta, by boosting the particles in the different rest frames, and computing high level features, all of which from trainable weights such that the gradient descent can optimize this automatic feature construction. The input particles to the LBN are the (two) leading lepton(s) and six (four) p_{T}-leading jets for the SL (DL), while the output features are the 4-momenta, p_{T}, total scalar momentum and angle between the output particles which number from 10 to 16 depending on the channel and analysis type. These 4-momenta are also provided to the DNN as low level features, on top of a list of high level features, listed in Appendices A.1 and A.2. In the resonant SL channel one DNN is trained separately for the resolved and boosted categories, per era and for spin-0 and spin-2

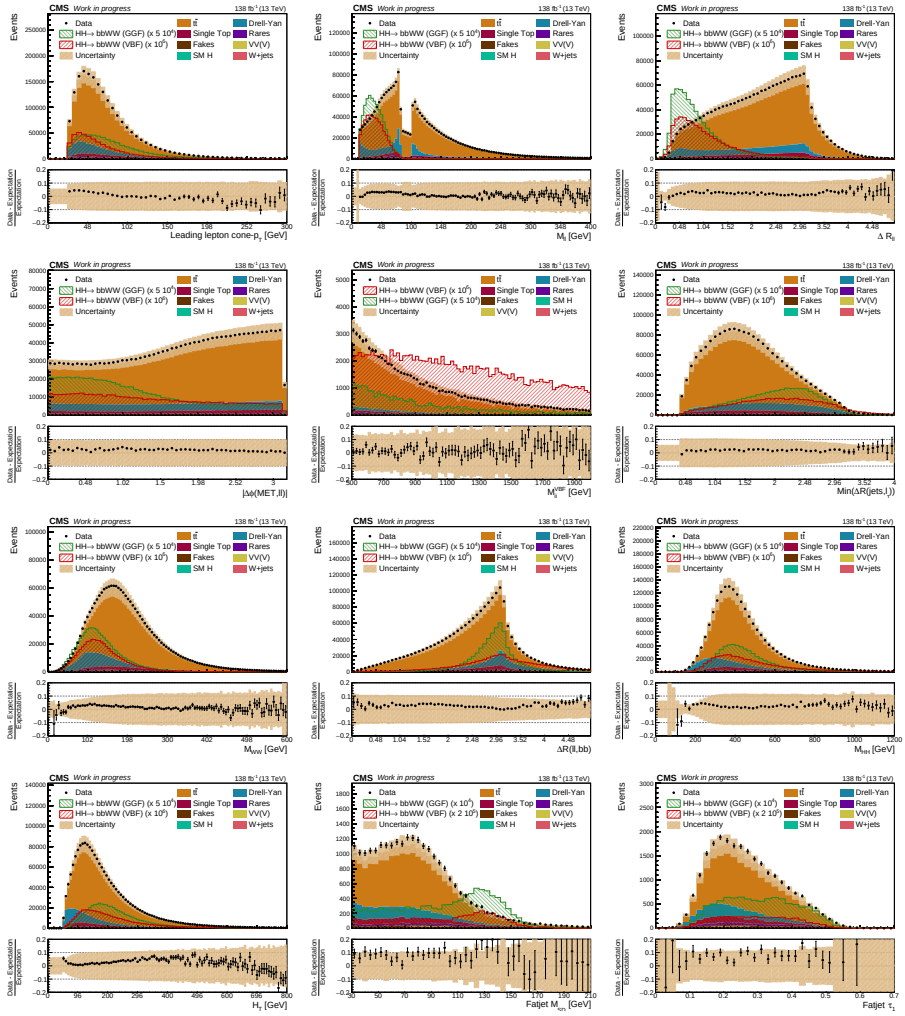


Figure 4.7 | Distributions of some DL input variables using the 138 fb^{-1} of the Run-2 dataset combining the different lepton channels (e^{-}e^{+}, $\mu^{-}\mu^{+}$ and $\text{e}^{\pm}\mu^{\mp}$). Some variables are shown using both resolved and boosted topologies (ℓ_1 cone-p_{T}, M_{ll}, ΔR_{ll}, $\Delta\phi(\text{MET}, ll)$,$M_{jj}^{\text{VBF pair}}$), using resolved jets only ($\min(\Delta\text{R}(\text{jets}, \ell_1))$, M_{WW}, $\Delta\text{R}(ll, bb)$,M_{HH}, H_{T}), or specific to the boosted AK8 – or fatjet – H → bb (M_{softdrop}, sub-jettiness τ_1). The GGF and VBF signal samples overlaid have SM coupling values and are scaled based on their expected cross-sections and BRs, an additional multiplicative factor mentioned in the legend is included for visualisation purposes.

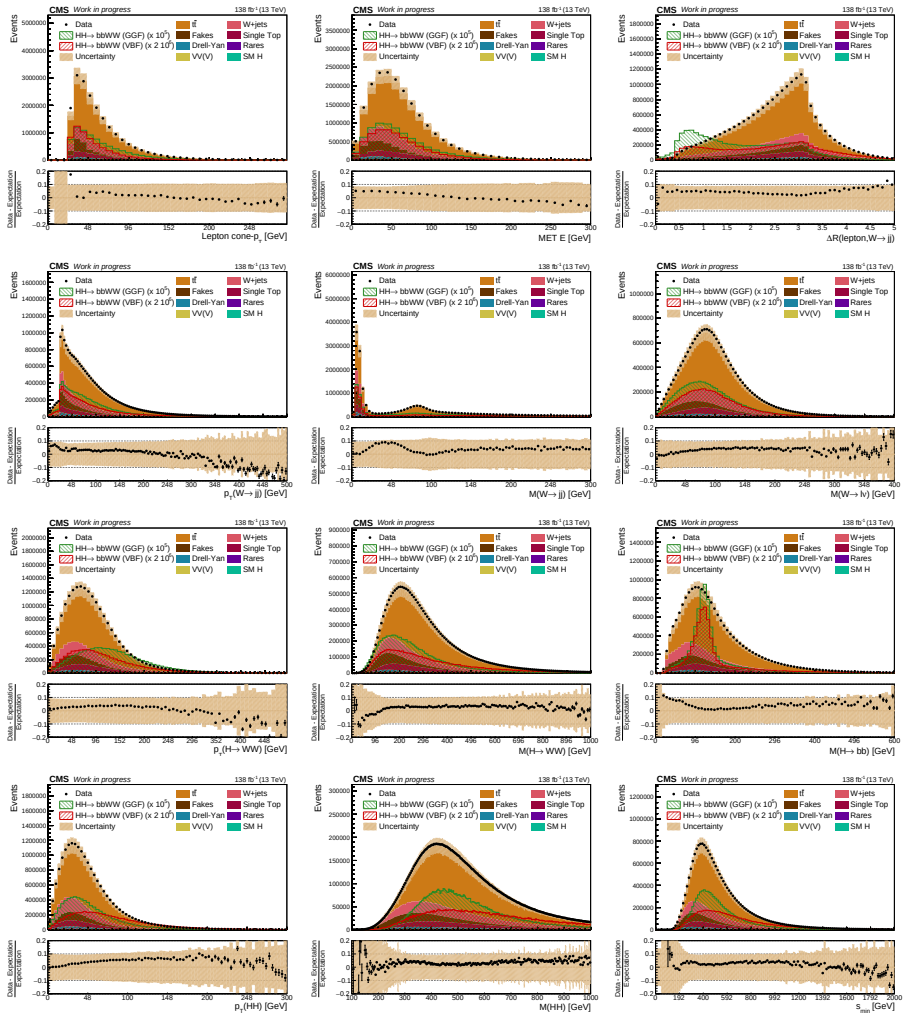


Figure 4.8 | Distributions of some SL input variables using the 138 fb^{-1} of the Run-2 dataset combining the different lepton channels (e$^{\pm}$, $\mu^{\pm}$). Some variables are shown using both resolved and boosted topologies (lepton cone-p_{T}, MET energy, $\Delta\mathrm{R}(l, \mathrm{W} \to jj$, $p_{\mathrm{T}}(\mathrm{W} \to jj)$, $M(\mathrm{W} \to jj)$, $M(\mathrm{W} \to \ell\nu)$, $p_{\mathrm{T}}(\mathrm{H} \to \mathrm{WW})$, $M(\mathrm{H} \to \mathrm{WW})$) or using resolved jets only ($M(\mathrm{H} \to \mathrm{bb})$, $p_{\mathrm{T}}(\mathrm{HH})$, $M(\mathrm{HH})$, S_{min}). The GGF and VBF signal samples overlaid have SM coupling values and are scaled based on their expected cross-sections and BRs, an additional multiplicative factor mentioned in the legend is included for visualisation purposes.

separately, for each case the inputs with more than 80 % correlation are dropped. For the other trainings all the eras, resolved and boosted categories are trained together, zero-padding is used when an input is absent, for example when no AK8 jet is present in the resolved category.

The non-resonant DNNs use a feed-forward ResNet [419] architecture with three blocks of two layers of 235 (229) neurons with ReLU activation functions, with a last layer of 7 (9) output nodes with a softmax activation function for the SL (DL) channel. The resonant analysis use a parametric fully connected DNN with five layers of 212 (512) neurons with Softplus Linear Unit (SLU) [420] – a smoother version of the ReLU – (ReLU) activation functions, with 6 (8) output softmax nodes. The resonant DL DNN training range over m_{HH} was split between the low and high mass regions, respectively in the ranges [250,450] and [400,900] GeV. The high mass DNN also used a few points beyond the TeV scale (only in the training) as it improved the sensitivity in the highest mass points considered in the statistical inference. The range was not split for the SL channel, but the training for spin-0 and spin-2 signals were done separately.

The *correlation matrix* for the DNN trained in the resonant DL channel is on Fig. 4.9. Some variables are highly correlated, and this could motivate a reduction of input variables to only keep uncorrelated variables. However, this decision would overlook two important shortcomings of the correlation matrix. First, the correlation can only measure linear dependence between variables while some might show very non-linear differences, a feature that DNNs are perfectly suited to leverage. Second, the correlation matrix is computed on the entire training sample that consists of very different processes. While some variables might be very correlated for background events, this correlation may vanish in signal events. The information contained between such variables can be helpful in the classification, and would be lost if one of them would be removed solely based on the correlation matrix. This is especially visible in the case of the reconstructed H_{T} and m_{HH} variables on Fig. 4.10 for two signal samples. The correlation coefficient specific to signal events is much lower than on Fig. 4.9, and the additional non-diagonal features can provide additional discrimination power.

N-fold cross-training is performed as described in Section 3.1.1 based on the modulo of the event number to obtain several uncorrelated sets, on both simulations and data. Each of the DNN within the same fold is trained and validated on $N-1$ sets, and used on the events of the remaining set to produce the distributions passed to the statistical inference, each with the same set of hyperparameters but different weight and bias initialisation. Two folds are used for resonant SL channel and five for the other cases. Early stopping and the reduction of the learning rate every time the validation loss reaches a plateau were employed, with an initial learning rate of the Adam optimizer varying between 0.01 and 0.001. Regularization techniques such as L2 regularization and dropout rates used parameters (when applied) in the

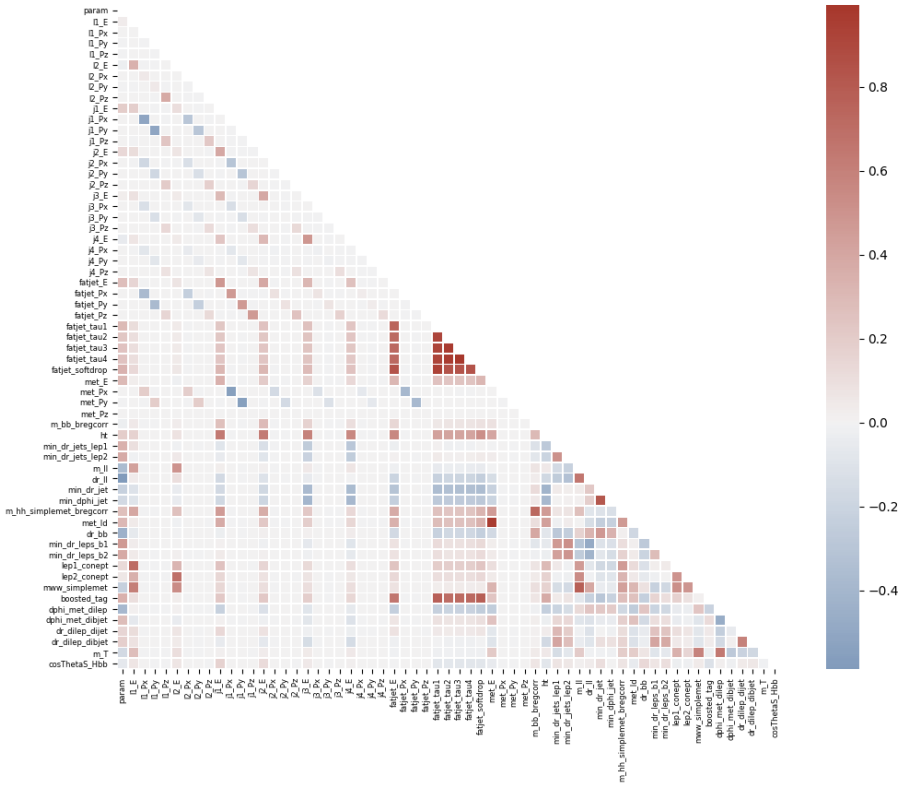


Figure 4.9 | Correlation matrix of the inputs to the DNN in the resonant DL channel, *param* corresponds to the parametric input m_{HH}^{gen}

ranges $[10^{-8},10^{-2}]$ and 0.01 respectively. Batch size ranged between 256 and 50000 depending on the channel and the amount of statistics, and batch normalisation is enabled.

To evaluate the performance of a multiclassification several tools can be used, for example the ROC curve. However, in the context of *winner-takes-all* method in this analysis, a most suited metric is the *confusion matrix*. It compares the true label of the event, and the one that was assigned by the DNN, summing over all the evaluation sample. The performance of the classification can then be assessed for each category, a diagonal confusion matrix being the ideal case. For the resonant DL DNN, the confusion matrices are on Fig. 4.11, the remaining ones are in Appendix A.3. The most important category is the GGF (and VBF for non-resonant searches) as it provides the discrimination between signal and background events, the remaining categories relative purity can be helpful in the likelihood fit to better constrain the different backgrounds.

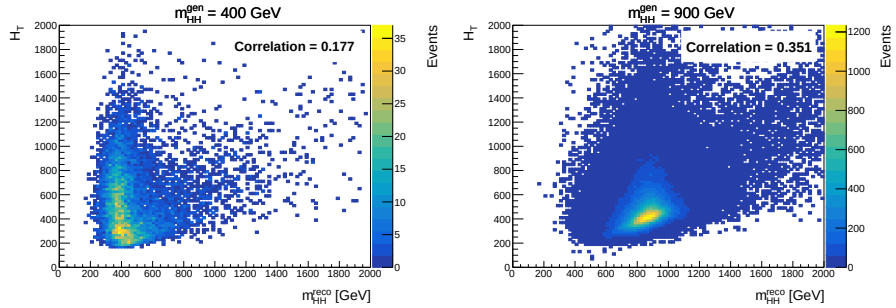


Figure 4.10 | Comparative distribution of two DNN reconstructed input variables H_{T} and m_{HH} for two sets of resonant signal events generated at 400 (left) and 900 GeV (right) along with their correlation.

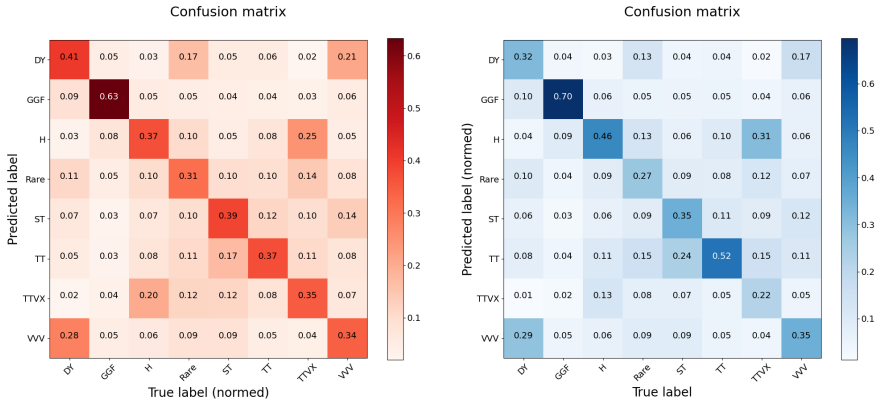


Figure 4.11 | Confusion matrices of the resonant DL channel, normed over predicted label (left) and true label (right).

4.2.6 Categorisation

To achieve sufficient statistics and maintain a good signal sensitivity, the processes from the DNNs are merged in groups and split into sub-categories, as illustrated on Fig. 4.12. This strategy was motivated by their similar shapes in the discriminator distribution, specifically for the categories that show different features depending on the H → bb topology. The categorisation of the DL and SL are defined in Tables 4.10 and 4.11 respectively, some categories are split between the resolved and boosted categories, the former being refined further when one or two AK4 jets are b-tagged.

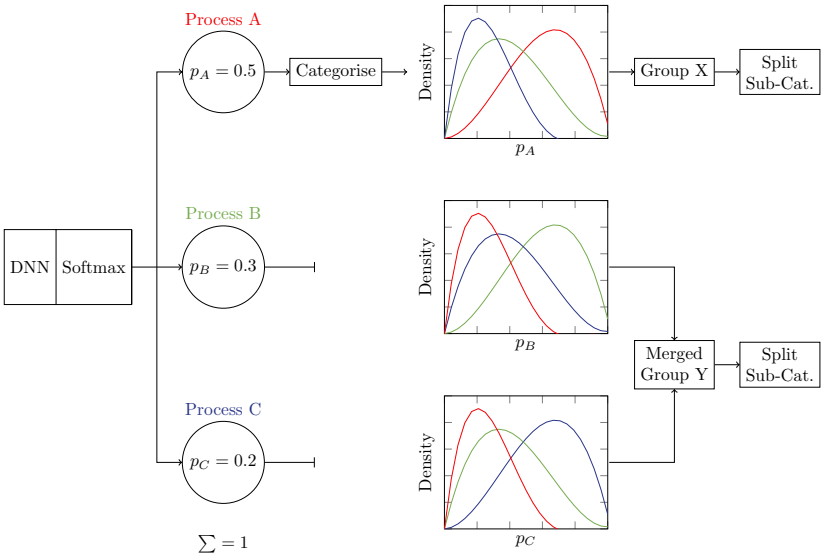


Figure 4.12 | Scheme of the strategy followed in this analysis. First, a DNN with a softmax activation function produces probabilities p_i for different processes ($i \in [A, B, C]$) that are used to categorize events in one of the process classes. Then the classes are merged into groups and some are further split depending on the H $\rightarrow$ bb topology.

Table 4.10 | Sub-categorisations in the DL channel.

Process Group	Sub-Categories		
HH(GGF)	Resolved 1b	Resolved 2b	Boosted
HH(VBF)	Resolved 1b	Resolved 2b	Boosted
Top + Other	Resolved		Boosted
DY + VV(V)	Inclusive		

Table 4.11 | Sub-categorisations in the SL channel.

Process Group	Sub-Categories		
HH(GGF)	Resolved 1b	Resolved 2b	Boosted
HH(VBF)	Resolved 1b	Resolved 2b	Boosted
Top + Higgs	Resolved		Boosted
WJets + Other	Inclusive		

4.2.7 Binning strategy

The histograms are produced initially with 400 bins, resulting in a 0.0025 bin width over the DNN score, to allow for further re-binning. The final binning strategy results from a compromise between generating a fine enough binning so that specific features of the signal can be reproduced, while keeping it coarse enough so that the bins are filled with enough background content to allow fit stability, and at the same time limit the number of bins for an affordable computation time of the fit. In addition, in each category the attached processes will have a distribution that peaks around the rightmost part of the spectrum, which is especially important for signal categories.

The background categories are mostly deprived of signal events, but their inclusion in the fit allows putting stringent constraints on the different background processes normalisation, thereby improving the precision of the excess quantification in the signal categories. Therefore, background categories are only using a quantile binning where the background processes distributions are summed, and their cumulative distribution is split between pre-defined quantiles. With constant value quantiles, this ensures a flat expected shape with different contribution of the different background processes in each bin, 3 bins with $33\ \%$ are used in the boosted categories and 5 bins with $20\ \%$ in the resolved and inclusive categories. The distributions of the DNN score of the non-resonant background categories defined in Table 4.10 using quantile binning is on Fig. 4.13.

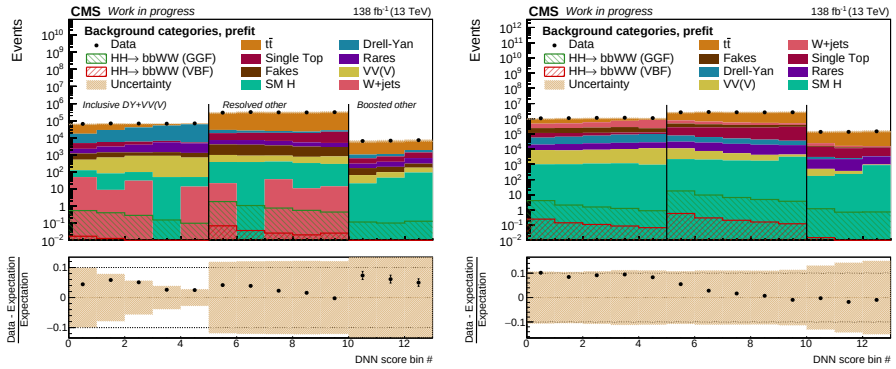


Figure 4.13 | Distributions of the DNN score in the background non-resonant categories of the DL (left) and SL (right) channels using the 138 fb^{-1} of the Run-2 dataset combining the different lepton channels, as a function of the bin number for clarity. The GGF and VBF signal samples overlaid have SM coupling values and are scaled based on their expected cross-sections and BRs, an additional multiplicative factor mentioned in the legend is included for visualisation purposes.

The binning strategy in the signal categories have a large impact on the sensitivity of the analysis, especially at large values of the DNN score where the signal-to-background ratio is the largest. However, a too large number of bins will cause low content and fit instability. This non-convex optimisation problem is addressed by aggregating bins starting from the rightmost bin, until the following criterion is met, in which case a new bin edge is defined and the aggregation content is reset and continues towards the left of the distribution.

$$\sum_P \text{Bin content}(P) - \sqrt{\sum_P \sigma_P^2} > t, \text{ with } \sigma_P = \begin{cases} \sigma_{P,const} \text{ if empty bin} \\ \sigma_{P,stat} \text{ otherwise} \end{cases} \tag{4.12}$$

where the sum runs over each process P, $\sigma_{P,stat}$ is the statistical variance of the bin content, and t is a pre-defined threshold value. When a bin is empty $\sigma_{P,const} = \sigma_0 \bar{w}_P^{event}$, where σ_0 is the constant Poisson uncertainty for the rate of $+1$ sigma confidence level under an observation of 0 events $\sigma_0 = \sqrt{\lambda_0} = \sqrt{1.84}$ and $\bar{w_P^{event}}$ the average event weight of the process. The threshold values are set with quadratically increasing values, such that the background distributions display a monotonously decreasing shape with enough contributions in each bin to ensure fit stability, while the signal-over background ratio increases in the rightmost bins.

In the non-resonant analysis, both channels use a set of thresholds with fixed values and the number of bins is determined based on the convergence of the algorithm and the content and variance of the rightmost bin. The distributions of the DNN score in the signal categories using this binning method is on Fig. 4.14.

In the resonant analysis, the DL channel uses a two-dimensional distribution of the DNN score and the HME defined in Section 4.2.3, that is then re-binned on the HME projected distribution using a quantile binning based on the signal distribution to ensure enough signal yield in each bin, then unrolled and each DNN score distribution re-binned using thresholds in each HME bin. This strategy yielded better performance than a pure MVA-based distribution, especially in the high mass region. Optimizing manually the values of the threshold for each HME bin given the wide range of yields across the m_{HH} spectrum would be a lengthy procedure. To circumvent that the number of bins of the DNN score distribution is fixed, the initial threshold values are set to large values and iteratively decreased until the condition Eq. (4.12) is met, with the addition that each bin must have enough content, otherwise the threshold values are adapted. The SL channel uses quantile bins similarly with the background categories, but the bin edges are based on the signal distribution, such that the bin width will decrease towards the rightmost part of the distribution, the signal content staying constant and the background content decreasing. The distributions are on Fig. 4.15 for two mass points.

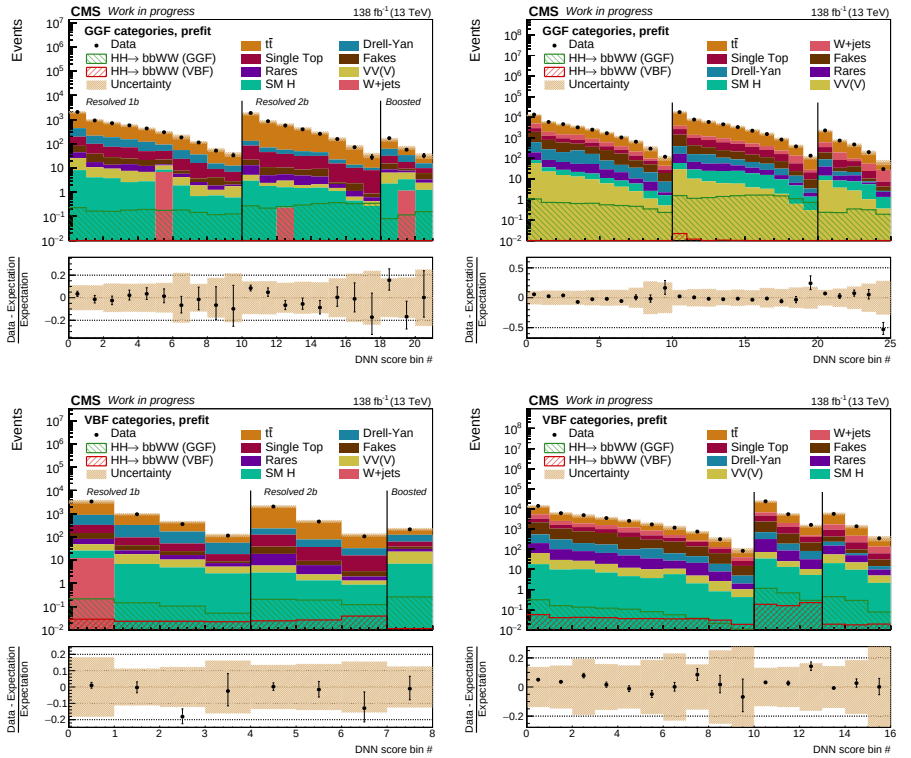


Figure 4.14 | Distributions of the DNN score in the GGF (top) and VBF (bottom) categories of the DL (left) and SL (right) channels using the 138 fb^{-1} of the Run-2 dataset combining the different lepton channels, as a function of the bin number for clarity. The GGF and VBF signal samples overlaid have SM coupling values and are scaled based on their expected cross-sections and BRs, an additional multiplicative factor mentioned in the legend is included for visualisation purposes.

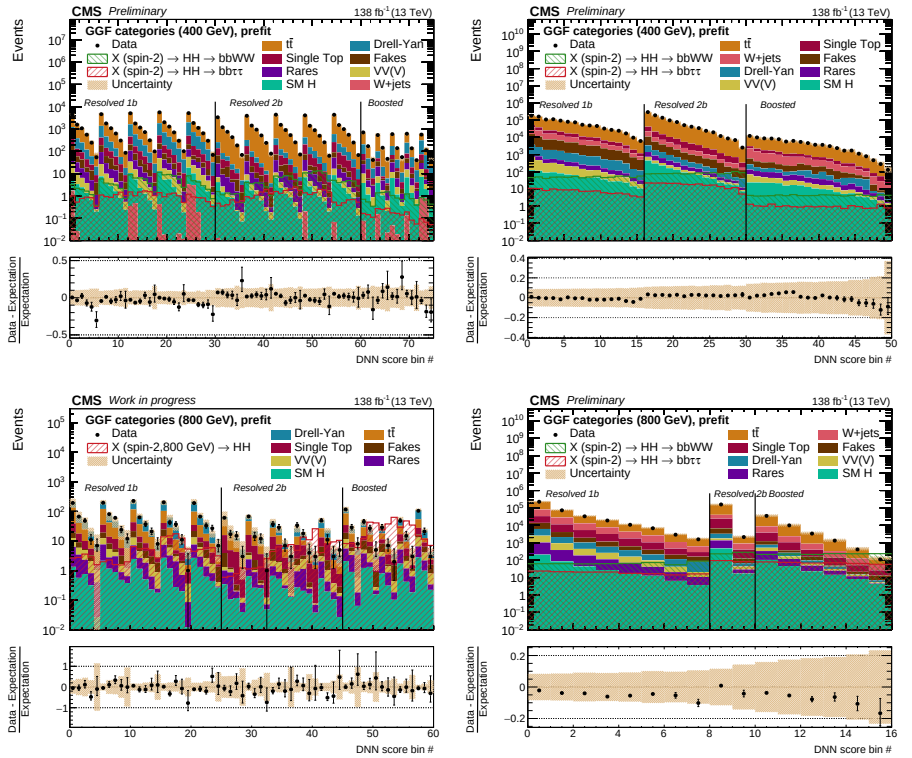


Figure 4.15 | Distributions of the DNN score in the DL channel (left) and SL (right) of the GGF resonant categories using the 138 fb^{-1} of the Run-2 dataset combining the different lepton channels, as a function of the bin number for clarity. The signal samples overlaid were produced assuming a resonance of 400 GeV (top) and 800 GeV (bottom), a cross-section of 1 pb and SM-expected BR values.

4.3 Monte Carlo corrections

4.3.1 pileup reweighting

The number of pileup interaction can be measured from Eq. (2.4) to yield

$$\mu = \frac{\mathcal{L}_{inst}\sigma_{inel}}{f_r}, \tag{4.13}$$

based on the instantaneous luminosity, inelastic proton scattering cross-section and the LHC frequency as described in Section 2.2. The value of the pileup μ represents the average of a Poisson distribution that is different in data and simulations. A reweighting based on the two distributions on Fig. 4.16 is applied to the latter.

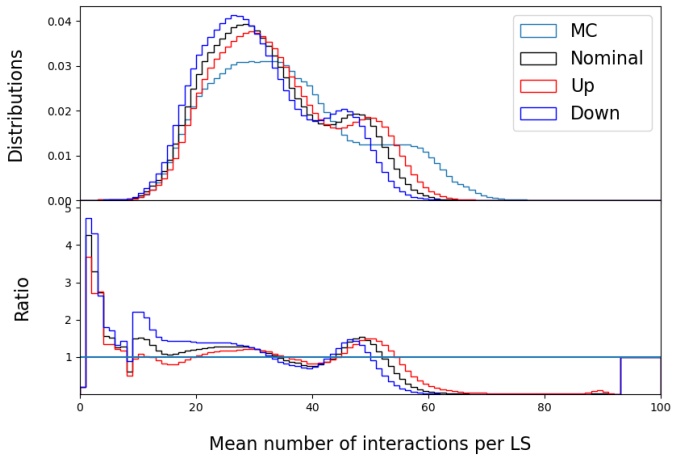


Figure 4.16 | Distributions of the mean number of interactions per LS, for a Monte Carlo sample of a fully leptonic $t\bar{t}$ process in 2017 data-taking year (green), the distributions extracted from data under nominal minimum bias cross-section (black) with its up (red) and down (blue) 1 sigma variations. The ratios can be used as a correction and estimation of uncertainty on the Monte Carlo sample.

4.3.2 Trigger efficiency

The trigger efficiency depends on the kinematic of the event and may not be modelled well by the simulation, a channel dependent SF must be included to correct for this effect. In the DL channel, an alternative trigger is used to select events, based on $E_{\mathrm{T}}^{\mathrm{miss}}$ and H_{T} to ensure orthogonality with the SR triggers. The events are further selected requiring two opposite sign leptons with $p_{\mathrm{T}} > 25$ (15) GeV to mimic the selections in Section 4.2.4. The data and simulation efficiencies are compared to estimate the trigger SF, the leading lepton cone-p_{T} variable was chosen as the efficiencies showed a strong dependence to it. In the SL channel, a T&P method – as described in Section 2.4.3 – was applied on $\mathrm{Z} \to \mathrm{e}^{-}\mathrm{e}^{+}$ and $\mathrm{Z} \to \mu^{-}\mu^{+}$ events, binned as a function of the lepton p_{T} and η. Both the tag and the probe must pass the identification and isolation requirements of Section 4.2.2. In addition, the tag must pass the SL triggers of Table 4.3 and the probe has to match a trigger object within $\Delta\mathrm{R} < 0.5$.

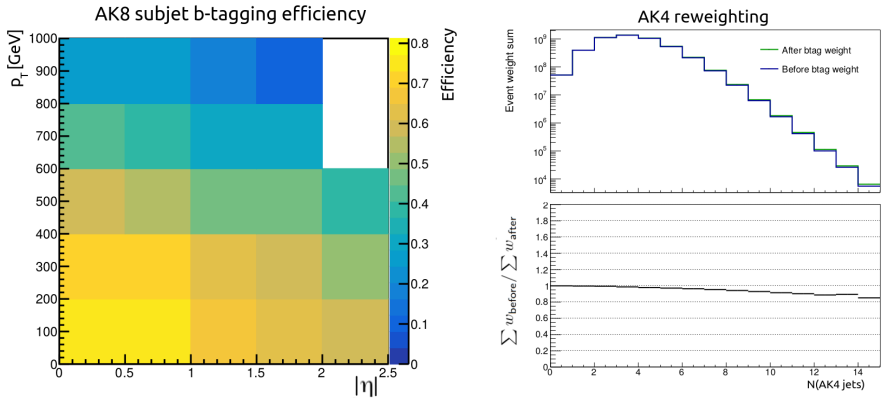
4.3.3 Lepton selection efficiency

The difference in lepton selection efficiency between simulation and data is corrected by a SF obtained from a T&P method in a two stage process : first the efficiency for leptons to pass the loose selection criteria, then from those leptons the efficiency to

pass the tight selection criteria. The leptons entering the data driven estimation of the fake leptons are only applied the former, justifying the implementation of these two stages consecutively. For technical reasons the first stage of the correction for the loose electrons is split into two steps, first the SF associated to the standard electron identification loose WP provided by the CMS collaboration, then the rest of the loose selection criteria that are specific to this analysis. Each SF is applied as a function of the lepton p_{T} and η.

4.3.4 b-tagging corrections

The b-jet identification efficiency and the misidentification probability are not entirely modelled by simulation, and a SF must be applied per event. The methods were discussed in Section 2.4.3. The fixed WP method of Eq. (2.15) was used for AK8 jets using b-tagging efficiencies (Fig. 4.17a). The same principle was used for the pileup jet id, using efficiencies provided by the CMS collaboration. For AK4 jets, the method of Eq. (2.16) was applied, together with the reweighting illustrated on Fig. 4.17b.



(a) *(b)*

Figure 4.17 | Left : AK8 subjet b-tagging efficiency as a function of the subjet p_{T} and η. Right : reweighting of the AK4 b-tag event weight based on Eq. (2.16), after and before the weight application.

4.3.5 Pre-firing of Level-1 ECAL trigger

In 2016 and 2017 data taking years a gradual timing shift was observed in the ECAL providing L1 trigger primitives at high pseudorapidity $2 < |\eta| < 3$. Given the L1 rule that prevents two consecutive BX to be triggered due to technical constraints, events with energy deposits at high η could end up being vetoed, an effect not taken

into account in the simulations. A correction is applied based on the probability for an event not to pre-fire, multiplied for each object in the event to obtain a weight and provided centrally by the CMS collaboration.

4.3.6 $E_{\mathrm{T}}^{\mathrm{miss}}$ modulation corrections

On top of the corrections coming from pileup and jet corrections in Sections 2.4.2 and 2.4.3 applied to the MET, a modulation effect has been observed in the azimuthal angle distribution that follows a sinusoidal curve with a 2π period, increasing as a function of the number of PVs in the event. This can be explained by inhomogeneities of the detector over the detector coverage, due to anisotropic energy response, inactive calorimeter cells, detector misalignment or beam spot displacement.

Linear fits of the x and y distributions of the MET in the transverse plane as a function of the number of vertices in the event have been performed using $\mathrm{Z} \to \mu^{-}\mu^{+}$ and appeared to fully account for the modulation effect, and a correction is provided centrally by the CMS collaboration.

4.3.7 HEM issue

All the corrections to jets from Section 2.4.3 are applied, but an issue appeared while taking data in 2018. Two HCAL modules in sections HE minus 15 and 16 – also dubbed HEM(15-16) – saw their power supply stop functioning, affecting the coverage and therefore jet energy measurements. This occurred during a low voltage power-up where a transient voltage surged to 22 V while the hardware limit was at 11 V, an effect that was never produced during testing and motivated further safety systems to be incorporated in the rest of the data-taking.

The correction consists in scaling down the energy of jets in the $p_{\mathrm{T}} > 15$ GeV and -1.57 $< \phi <$ -0.85 region where the failing sectors are located, by 20 % if -2.5 $< \eta <$ -1.3 and by 35 % if -3.0 $< \eta <$ -2.5. This correction is further propagated to the MET and other event-level variables.

4.3.8 Top reweighting

During Run-1 it was observed that the p_{T} spectra of the top quarks in $\mathrm{t\bar{t}}$ simulated samples were harder than in data [421]. As the samples used are typically at NLO in QCD, one partial explanation might come from higher order QCD or EW corrections, but non-resonant production of $\mathrm{t\bar{t}}$-like final states and other unknown effects might come into play.

To resolve the missing higher order issues, the cross-section was computed with NNLO order in QCD (and NLO in EW) [422] and serves as the numerator of a

correction where the denominator is the nominal value used in the event generation. This ratio is then fitted as a function of the top (anti)quark p_{T} to obtain a SF, the event weight being computed as

$$w = \sqrt{\mathsf{SF}(p_{\mathrm{T}}^{\mathrm{t}}) \times \mathsf{SF}(p_{\mathrm{T}}^{\bar{\mathrm{t}}})}$$
$$\mathsf{SF}(x) = \exp\Big(a + bx + cx^2 + d/(x+e)\Big). \tag{4.14}$$

The advantage of this correction is that it is only based on theory predictions, and it would not suffer from potential contamination of BSM signals in the data. Additionally, as it is purely simulation-driven, there is no concern about data being used twice in the statistical analysis.

4.3.9 VBF dipole recoil

The VBF signal samples were produced at LO and the potential emission of soft radiations during the parton shower was handled by PYTHIA using the default model, namely the global-recoil over all other partons of the shower. This mode assumes no colour connection between the two quark lines in the initial and final states, as described in Section 1.3. However, internal studies within CMS have shown the significant impact of the choice of recoil scheme, with the non-default dipole-recoil mode most closely matching the fixed order NNLO prediction.

A specific systematic uncertainty has been derived to take this effect into account as a normalisation effect given the limited amount of statistics and few number of coupling points generated with the correct configuration and not in all data taking years. It is computed as the ratio of yields between the configuration with the "dipole on" and "dipole off" settings if the statistical precision allows it, per category and data taking year. As only a few samples were produced with the "dipole on" setting, the missing couplings are assigned an extrapolated value, taken as the maximum of the available values. As there is little sensitivity in the categories other than the VBF one in Table 4.10, the systematic is only applied in the latter. A summary of these values are on Table 4.12.

Table 4.12 | Relative yield between the "dipole on" and "dipole off" settings for the cases where the two samples have been generated and the statistical uncertainties allow it, for the distribution of each category entering the likelihood fit, decay channel and set of couplings (κ_V,κ_{2V},κ_λ).

Category	bb$\tau\tau$ (1, 1, 1)	bbVV (1, 1, 1)	bb$\tau\tau$ (1, 2, 1)	bbVV (1, 2, 1)
Boosted HH(VBF)	1.13	1.00	1.06	1.06
Resolved 1b HH(VBF)	1.00	1.07	1.00	1.03
Resolved 2b HH(VBF)	1.08	1.04	1.04	1.04

4.4 Background estimation

The main background contributions to the SR are from $t\bar{t}$, ST and W+jets events in the SL channel, and $t\bar{t}$, ST and DY events in the DL channel. $t\bar{t}$ events are well reproduced from simulation and as an irreducible background have a high efficiency of being selected in the SR so the statistical uncertainties are kept below manageable level. Similarly, the W+jets process can be taken from simulations as its contribution in the SL is lower and depends on the associated jets to be either from a genuine bottom quark or a light-jet misidentified as a b-jet. On the contrary, the efficiency of selection of DY events in the DL is very low and the statistical precision will be impaired, therefore a data-driven method has been developed and presented in Section 4.4.1. Misidentified leptons from jets consists of a background that cannot be modelled well in simulations, and while their contribution is close to negligible in the DL channel, it is far from it in the SL channel. Hence an estimation from data is derived for both channels and presented in Section 4.4.2. To provide the statistical fit some flexibility, each of the main simulated backgrounds ($t\bar{t}$, ST and W+jets) is allowed to vary by a factor of 20 % of their expected yield, and an additional 20% in the boosted regions where the normalisation showed some issues. The minor backgrounds are assigned a 50% factor to account for potential incorrect cross-section, the same factor is also applied to the fakes uncorrelated per data-taking year to account for any wrong normalisation. The CRs defined by the background categories in Section 4.2.5 allow the fit to determine the correct normalisations. Other simulated minor backgrounds are grouped together to improve convergence.

4.4.1 Drell-Yan background

The data-driven method used to estimate the DY distribution in the SR uses data events in a DY Control region (CR). The selection on the dilepton invariant mass in the SR ($\left|m_{\ell\ell} - m_{\mathrm{Z}}\right| > 10$ GeV, denoted "Z veto") is inverted in the DY CR ($\left|m_{\ell\ell} - m_{\mathrm{Z}}\right| \leq 10$ GeV, denoted "Z peak"). The high purity of the DY process in the "Z peak" and "Z veto" regions with no b-tagged jets (Table 4.13) motivate their use in a method called the *ABCD* method to extract transfer weights in the former and apply them in the latter as an estimation of the DY process from data in the SR, as illustrated on Fig. 4.18. The transfer weights are computed as

$$N(x_i|\text{1b/2b, Z veto}) = N(x_i|\text{0b, Z veto})\frac{N(x_i|\text{1b/2b, Z peak})}{N(x_i|\text{0b, Z peak})}, \qquad (4.15)$$

where x_i refers to the bin i of a certain variable x distribution whose DY process content is given by N.

Table 4.13 | DY purity for 2016 data-taking era in the "Z peak" ($\left|m_{\ell\ell} - m_{\mathrm{Z}}\right| \leq 10$ GeV, top) and "Z veto" ($\left|m_{\ell\ell} - m_{\mathrm{Z}}\right| > 10$ GeV, bottom) regions in the resolved – further split per number of b-tagged jets – and boosted categories.

Z peak

Channel	Resolved 0b	Resolved 1b	Resolved 2b	Boosted
e^-e^+	98.55%	91.69%	68.12%	84.64 %
$\mu^-\mu^+$	98.64%	92.37%	69.41%	85.78%

Z veto

Channel	Resolved 0b	Resolved 1b	Resolved 2b	Boosted
e^-e^+	90.15%	27.71%	5.37%	24.89%
$\mu^-\mu^+$	91.29%	30.64%	5.95%	25.23%

The variable x must be decoupled from the two leptons to avoid any bias, and must therefore be defined on the jets only. In the resolved category it was chosen as H_{T} the scalar sum of selected AK4 jets p_{T}, and in the boosted category as the AK8 softdrop mass, as illustrated on Figs. 4.19 and 4.20. The good agreement between distributions from similar b-tagging regions but different dilepton mass regions confirms the independence of the chosen variables in the horizontal direction of application of the weights in Fig. 4.18. Given the high purity of the DY process in the CR the transfer weights in Eq. (4.15) can be obtained by subtracting the minor backgrounds from the data.

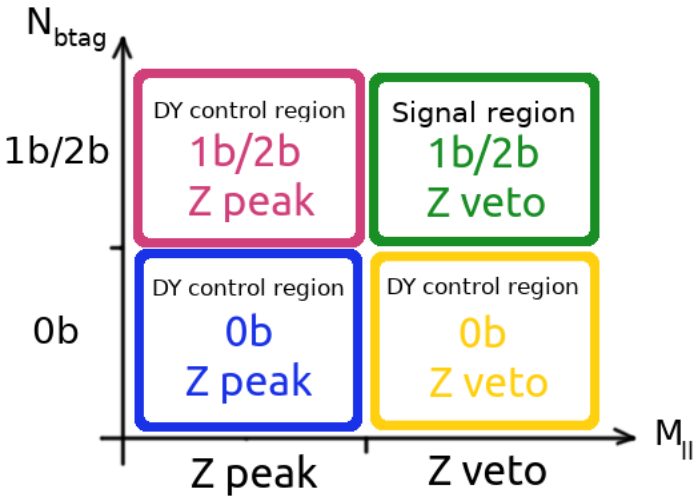


Figure 4.18 | Scheme of the ABCD method used in the estimation of the DY process from data, using three DY CR to estimate transfer factors and apply them in the same dilepton mass region as the SR.

As no difference was observed between the weights obtained from the e^-e^+ and $\mu^-\mu^+$ channels, their contribution is added to decrease the systematic error and the transfer weights obtained used in both channels. The contribution of the DY process $e^{\pm}\mu^{\mp}$ channel is mainly due to the production of two leptonically decaying taus, which after selections is only a minor contribution in this channel. As within the large error bars there was no deviation from the same flavour channels, the DY contribution is estimated in the same way using the already defined transfer weights.

The resolved category is split between the region with one and two b-tagged jets, as defined in Table 4.10. However, the data events used to estimate the DY contribution only come from the region without any b-tagged jets. To avoid using data events twice in the resolved region of the SR, they are randomly split in two contributions based on the event number, defined orthogonal with the DNN selection from the cross-training.

4.4.2 Misidentified lepton background

To estimate the contribution of jets faking leptons in our SR, the Fake Factor (FF) [412, 423] method is used. In this method, a sample of events satisfying all the selection in Section 4.2.4, with the exception that one of the fakeable electrons or muons fail the tight selection. This region is referred to as the Application region (AR) of the FF method, and the fakeable leptons in this region are the lepton candidates. To avoid overlap with the SR, the number of tight leptons is limited to zero (one) in the SL (DL). To avoid double-counting of prompt leptons in the AR, their contributions are estimated on simulations and subtracted from the data as illustrated on Fig. 4.21, the signal contributions in the AR being negligible.

The identification of prompt leptons consists in the possibility to associate them with generator level partons, originating from either a W, Z or H boson decay, or from a decaying τ. Any matched candidate must fulfil $\Delta\mathrm{R} < 0.3$ and $|p_{\mathrm{T}}^{reco} - p_{\mathrm{T}}^{gen}|/p_{\mathrm{T}}^{gen} < 0.5$. Only one generator level lepton can be associated to each reconstructed lepton, the muons are associated first then the electrons, and in case of ambiguity the association is done with a decreasing p_{T}-ordering of the reconstructed leptons.

The estimation of the fake contribution in the SR is done through the application of weights to events in the AR, given by

$$w = (-1)^{n+1} \prod_{i=1}^{n} \frac{f_i}{1 - f_i}, \tag{4.16}$$

where the products extend over all fakeable leptons failing the tight selection, the sign factor is there to aoid double counting. In case the event contains more than one (two) of those leptons in the SL (DL) only the leading (leading and subleading) in terms of cone-p_{T} is considered for the weights, therefore $n \leq 1$ ($n \leq 2$). The

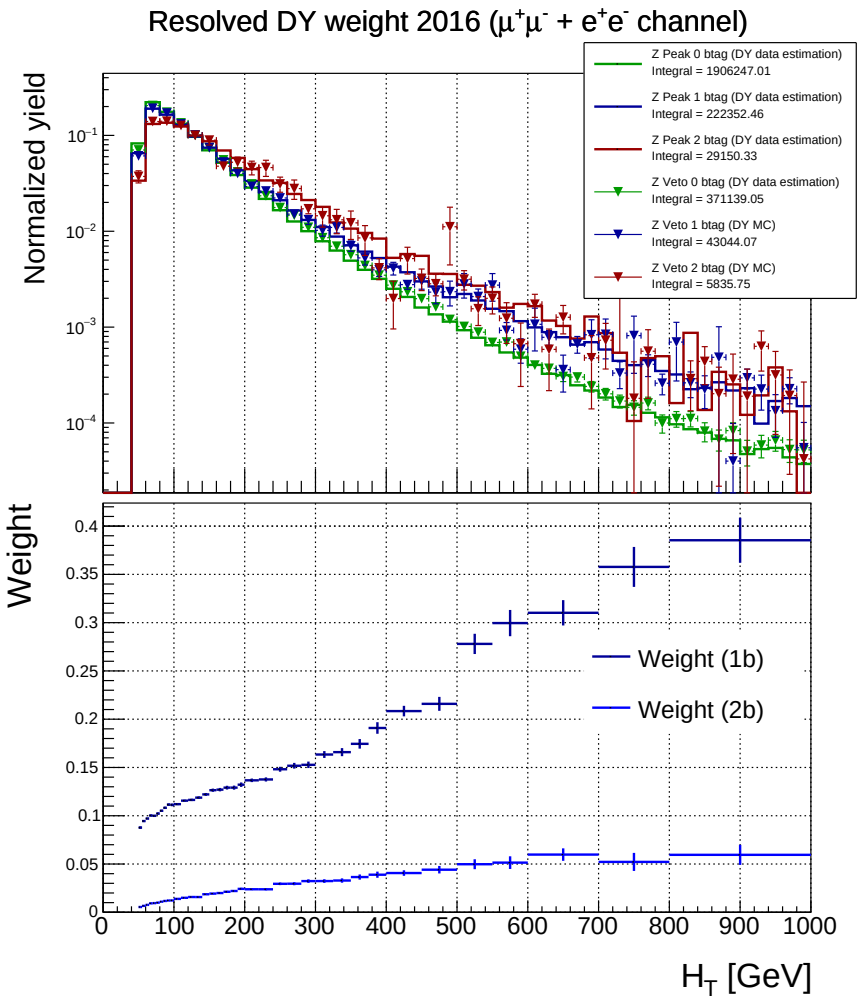


Figure 4.19 | Distributions of the variable used to derive transfer weights in the resolved category using the scalar p_{T} sum of the jets H_{T}. In each category the distributions of the variable in each region of number of b-tagged jets are shown with integral normalized to unity for comparison (top), and the transfer weight(s) (bottom) are based on Eq. (4.15). The solid lines represent the distributions in the "Z peak" region from which the transfer weights are extracted, the DY contribution is estimated by subtraction of the minor background from the data. By comparison, the marked lines illustrate the same distributions in the "Z veto" region from Monte Carlo simulations.

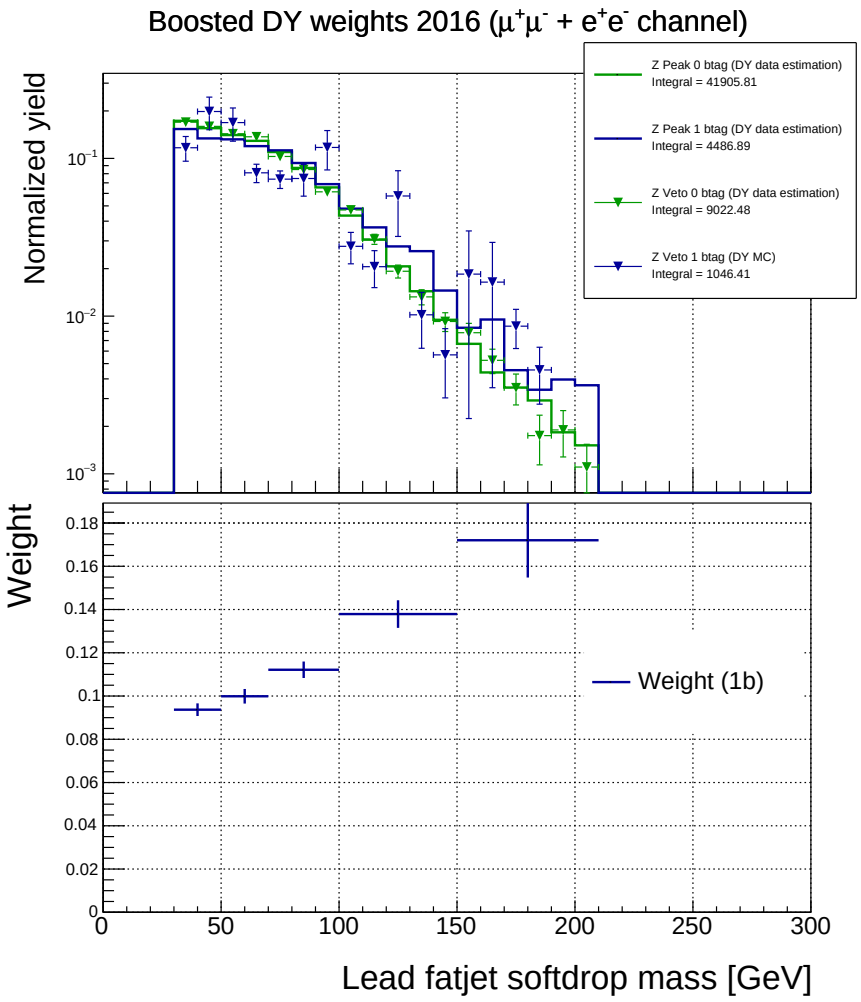


Figure 4.20 | Distributions of the variable used to derive transfer weights in the boosted category using leading AK8 jet softdrop mass. In each category the distributions of the variable in each region of number of b-tagged jets are shown with integral normalized to unity for comparison (top), and the transfer weight(s) (bottom) are based on Eq. (4.15). The solid lines represent the distributions in the "Z peak" region from which the transfer weights are extracted, the DY contribution is estimated by subtraction of the minor background from the data. By comparison, the marked lines illustrate the same distributions in the "Z veto" region from Monte Carlo simulations.

f_i denote the probability for a fakeable lepton to also pass the tight selection, and are parameterised as a function of the cone-p_{T} and $|\eta|$. They are measured in a Measurement region (MR), performed by the $\mathrm{t\bar{t}H}$ analysis of Ref. [412]. In the MR, the events must have exactly one muon or electron passing the fakeable selection, and at least one jet separated from the lepton by $\Delta\mathrm{R} > 0.7$. Then events are classified into a "pass" or "fail" categories depending on them passing the tight selection or not. The f_i probabilities are defined as

$$f_i = \frac{N_{pass}}{N_{pass} + N_{fail}}, \tag{4.17}$$

where N_{pass} and N_{fail} are the number of events in both categories from multijet events, the main contribution to the fake background in the SR. The fail region is mostly populated by multijet events, and its contribution is estimated by data after subtraction of the other processes (W+jets, DY, VV and $\mathrm{t\bar{t}}$) from simulation. The contribution from multijet in the "pass" categories is extracted from a ML fit as in Section 2.4.3 using data and simulations of the other backgrounds. The variable for the fit must be relatively independent of the lepton cone-p_{T} and is a modified version of the transverse mass

$$m_T^{fix} = \sqrt{2p_{\mathrm{T}}^{fix} E_{\mathrm{T}}^{\mathrm{miss}}(1 - \cos\Delta\phi)}, \tag{4.18}$$

where $\Delta\phi$ is the angle between the lepton and the MET, while $p_{\mathrm{T}}^{fix} = 35$ GeV is fixed at the typical transverse momentum of the electrons in W+jets events. This variable is independent of the lepton p_{T}, and the multijet will peak at lower values than the other backgrounds because the MET mostly comes from resolution effects and not from a real missing neutrino.

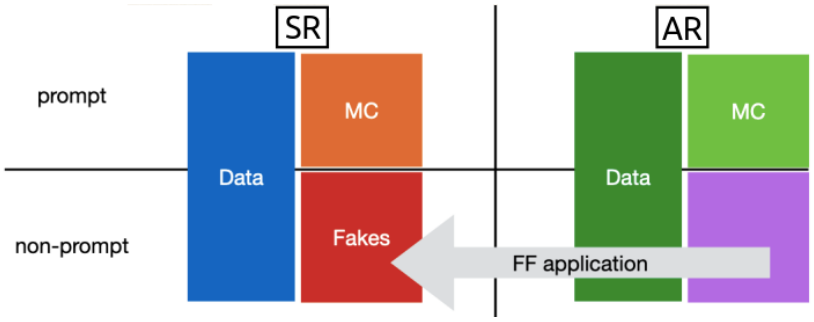


Figure 4.21 | Scheme of the Fake Factor method, with the application of the FF weight in the AR on both data and background Monte Carlo simulations that are then subtracted to obtain a data-driven estimation of the Fakes in the SR.

4.4.3 Non-closure corrections

Multiple biases can appear in the data-driven estimation of the Fakes and DY backgrounds. In the former, they can be caused by differences between the SR and in the MR of the lepton candidates p_{T} spectrum and in the latter from the projection of the transfer weights on a single variable that might not be able to catch the phase space modification between b-tagging regions. In both cases, the effect is minor for low- and high-level variables, but the correlations introduced in their inclusion within the DNN classifiers can produce a sizeable mismodelling compared to data. A non-closure correction has been derived for both cases, in the DL channel for the DY, and in the Single Lepton channels for the Fakes.

A closure test is performed for the DY estimation by comparing the distributions in the SR from the Monte Carlo simulations and in the *closure* distributions from DY events in the CR to which transfer weights specifically computed using simulations are applied. The comparison is on Fig. 4.22a. As this non-closure test is distribution dependent, a correction must be defined for the distribution of each category entering the likelihood fit. The type of correction for a given distribution then depends on its attached category.

For background categories, a linear trend is observed from both the non-closure test and the data-driven mismodelling. Therefore, a linear fit is performed and applied as a correction on the histogram, based on the slope and intercept from the fit, prior to the re-binning of Section 4.2.7. However, the systematic uncertainties of these two parameters are correlated in a non-trivial way. To disentangle them, a diagonalisation of the covariance matrix returned by the fit can be performed. If the fit is parameterised by

$$y(x) = c_0 + c_1\ x, \tag{4.19}$$

where $\vec{C} = (c_0, c_1)^T$ is the parameter vector whose covariance matrix is given by

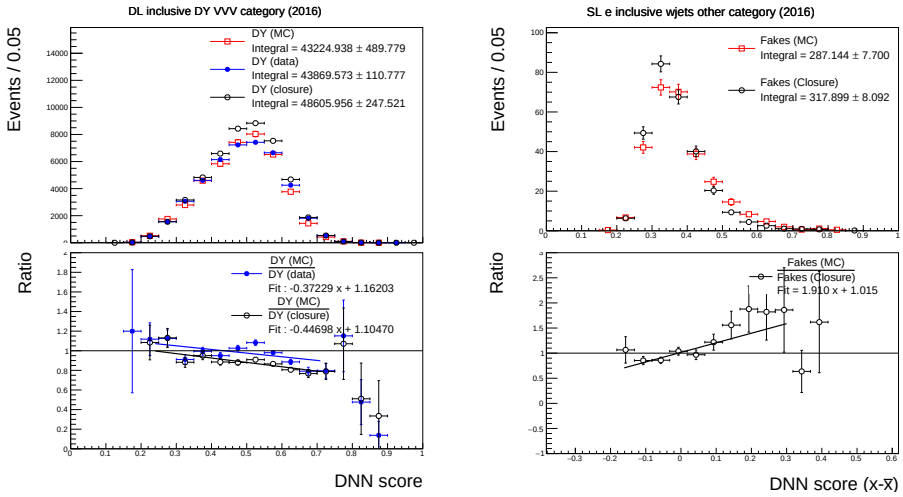
$$\Sigma(\vec{C}, \vec{C}) = \begin{pmatrix} \sigma^2_{c_0} & cov(c_0, c_1) \\ cov(c_1, c_0) & \sigma^2_{c_1} \end{pmatrix}. \tag{4.20}$$

The eigenvalues and associated eigenvectors λ_0, λ_1 and their eigenvectors $\vec{v_0}$,$\vec{v_0}$ of the covariant matrix allow diagonalising it as such

$$V^T \Sigma(\vec{C}, \vec{C}) V = \begin{pmatrix} \lambda_0 & 0 \\ 0 & \lambda_1 \end{pmatrix} = \Lambda, \tag{4.21}$$

where $V = (\vec{v_0}, \vec{v_1})$.

This decomposition in the eigenvector basis can be used to determine two uncorrelated systematic shapes



(a) *(b)*

Figure 4.22 | Non-closure corrections for the DY and Fakes data-driven estimations in the DL and SL respectively. Left : comparison of the distributions in the DY+VVV category and 2016 data taking year between the DY from simulations (top, red), data-driven (top, blue) and non-closure (top, black), and the ratios used to produce the non-closure correction from a linear fit (bottom, black) and the comparison with the data-driven (bottom, blue) to illustrate the need for this correction. Right : comparison of the distributions in the WJets+Other category in 2016 data taking year in the electron channel between the Fakes from simulations (top, red) and closure (top, black), with ratio used in the non-closure corrections (bottom) as a function of the shifted DNN score by the mean of the histogram.

$$\begin{aligned} y_0^{\pm} &= (c_0 \pm \sqrt{\lambda_0} v_{00}) + (c_1 \pm \sqrt{\lambda_0} v_{10})\ x, \\ y_1^{\pm} &= (c_0 \pm \sqrt{\lambda_1} v_{01}) + (c_1 \pm \sqrt{\lambda_1} v_{11})\ x, \end{aligned} \tag{4.22}$$

whose up and down variations can be used as templates for shape systematic uncertainty, that is now decorrelated for the two parameters.

For signal categories, however, the DY contribution is too small to allow any trend to appear within the statistical uncertainties. In that case, only a normalisation uncertainty is assigned and computed from the ratio of integrals between the distribution from simulation and the closure test.

The residual non-closure for the Fakes works similarly, with some differences. Two regions are defined using leptons that are non-prompt, the MC estimation of Fakes in the SR and its closure in the Fake CR, respectively by requiring that the lepton is tight or not. As in the SL channel, the majority of the Fakes are coming from the hadronic $t\bar{t}$ process, the distributions are only drawn from that single process, illustrated on Fig. 4.22b. Similarly to the DY non-closure, a linear fit is performed,

albeit the two degrees of freedom are measured in a two-step procedure, and they only serve to assign two shape systematic uncertainties without any corrections. First the ratio of integral fixed in a range $n \in [0, 2]$ is used to obtain a normalisation uncertainty of the form $1 \pm |1 - n|$ and the two distributions are scaled to the integral of the numerator. Then the bin content is shifted by the mean of the distribution $\bar{x}$ to measure solely the slope parameter s and applied as an additional systematic $1 \pm s'$ where $s' = s.(x - \bar{x}) \in [-1, 1]$ is restricted to a certain range for numerical stability. This procedure has to be performed for each distribution of the categories entering the fit in Table 4.11, and for the electron and muon channels separately. Given the small contribution of the Fakes in the signal regions however, only the normalisation uncertainty is kept.

4.5 Systematic uncertainties

Several sources of systematic uncertainties are considered in this analysis, included as NPs in the maximum likelihood fit, affecting both the normalisation and shape of the distributions. They are categorised into experimental sources applied correlated for each simulated process, data-driven sources and theory sources correlated across data-taking years years. Additionally, statistical uncertainties are taken into account following the description in Section 2.4.3, and normalisation uncertainties are assigned to the main backgrounds to allow the fit to constrain them based on the background categories serving as CRs.

- Experimental uncertainties
  - **Luminosity** : normalisation uncertainty assigned to each simulated process, except the ones estimated from data, split into several contributions to disentangle correlated and uncorrelated sources for each data-taking years, with a total value of 1.2 % in 2016, 2.3 % in 2017 and 2.5 % in 2018.
  - **Pre-firing** : shape uncertainty to cover the correction in Section 4.3.5 applied in 2016 and 2017 data-taking years.
  - **Pileup** : shape uncertainty to cover the correction in Section 4.3.1.
  - **Trigger** : the uncertainties of trigger efficiency corrections in Section 4.3.2 are used as a shape uncertainty.
  - **Lepton identification** : the different sources of uncertainties associated to the different levels of lepton identification of Section 4.2.2 are used as shape uncertainties, as described in Section 4.3.3.
  - **Top p_{T} reweighting** : the shape systematic assigned to the reweighting in Section 4.3.8 is estimated by using the squared weight – a 100% variation – and unit value as up and down fluctuations, respectively.

  - **Jet PU ID** : shape uncertainties associated to the jet PU ID criterion applied on jets, corresponding to the selection efficiency and mistag rate of jets as originating from pileup, as described in Section 2.4.3.
  - **Jet Energy Scale (JES)** : corrections described in Section 2.4.3 for the energy scale of the jets using the merged scheme to reduce the number of shape uncertainties to 11 NPs.
  - **Jet Energy Resolution (JER)** : shape uncertainty associated to the smearing of the jets, as described in Section 2.4.3.
  - **B-tagging** : shapes uncertainties associated to the b-jet identification, coming from six various sources and flavour dependence for AK4, and a single source for the fixed WP identification of AK8 subjets.
  - **HEM** : shape uncertainty attached to the issue described in Section 4.3.7 of turned off HCAL modules.
  - **Unclustered $\boldsymbol{E_{\mathrm{T}}^{\mathrm{miss}}}$** : energies of unclustered PF candidates that are varied to provide a shape uncertainty for the $E_{\mathrm{T}}^{\mathrm{miss}}$ before type-1 corrections.
- Data-driven uncertainties
  - **DY estimation** : the uncertainties of the weights used to estimate the DY from data in Section 4.4.1 are used as shape uncertainties decorrelated for each data taking year and jet categories, while the associated non-closure corrections described in Section 4.4.3 are decorrelated per fit categories, between resolved and boosted categories and per data-taking year except for the boosted categories.
  - **Fake estimation** : the uncertainties of the fake factors in Section 4.4.2 are split into three contributions and decorrelated per data-taking year, while the non-closure in Section 4.4.3 is only applied in the SL channel and for which shape and normalisation uncertainties are applied decorrelated per category and year.
- Theoretical uncertainties
  - **Branching ratios** : the H $\rightarrow$ bb, H $\rightarrow$ WW (same as H $\rightarrow$ ZZ) and H $\rightarrow \tau\tau$ BRs are assigned a normalisation uncertainty of $\pm$1.25 %, $\pm$1.52 %, $\pm$1.64 % respectively.
  - **Event generation** : normalisation uncertainties associated to QCD scale, PDF and α_S are assigned to the simulated processes including the signals.
  - **Mass uncertainties** : top mass corrections for the $\mathrm{t\bar{t}}$ process and on the HH total mass for the signals.
  - **EW corrections** : normalisations attached to $\mathrm{t\bar{t}Z}$ and $\mathrm{t\bar{t}W}$ processes.
  - **Parton shower** : shape uncertainties attached to the variations of α_S

from FSR and ISR.

- **Scales** : shape uncertainties attached to the variations of the renormalisation (Section 1.1) and factorisation (Section 1.3) scales. The shapes are considered independent between the different groups of background processes and the signal samples, so they remain consistent with the order of event generation.
- **VBF dipole recoil** : normalisation uncertainty attached to the dipole recoil issue in Section 4.3.9.
- **$t\bar{t}$ generation** : several shape uncertainties attached to the theoretical settings during the event generation are taken into account for the $t\bar{t}$ process, namely from the damping variable h_{damp} in POWHEG which controls the merging between the ME and the parton shower to regulate high-p_{T} radiation, the colour reconnection between quarks and gluons in the parton shower in PYTHIA split into three sources, and the UE describing the interaction of the partons that do not enter the hard scattering (all described in Section 1.3).

▪ Statistical uncertainties : due to the limited size of the simulated samples used in the statistical inference, additional systematics are included based on the histogram bin errors and following the Beeston-Barlow lite method described in Section 2.4.3.

Estimating the impact of each systematic on the total uncertainties is especially important to improve the analysis in future iterations. One easy way is to compute the up and down variation effect on the nominal histogram yield, as is summarised in the left columns of Table 4.14, with a more detailed summary in Tables A.1 and A.2. While this can give an estimation of how much each NP is allowed to vary during the likelihood fit, it does not take into account shape effect. A more complicated but correct way is to incrementally freeze groups of systematics and determine their importance from the effect on the POI uncertainties over a likelihood scan, using a quadrature subtraction. These likelihood scans are illustrated on Fig. 4.23, and summarised in the right columns of Table 4.14. From the table, it can be seen that most of the uncertainties come from pure statistical uncertainties and the limited size of the Monte Carlo simulated samples. These can only be improved by increasing the amount of collected data and simulated samples respectively, both of which are fixed at this level. Only the remaining systematic nuisances can be improved upon in the analysis methods.

Out of the systematics that affect all simulated samples, one can distinguish between the ones related to the pure data-taking (luminosity, pile-up, etc, grouped inside the "Experimental" denomination) and the ones related to the reconstruction techniques. The former is fixed once the samples are simulated, and they do not have a significant impact anyway, whereas the latter can be further improved, although most are provided centrally by the CMS collaboration. Due to the jet

selection of the decays H → bb and of the hadronic W boson in the SL channel, the jet-related uncertainties are the largest of the experimental side. Similarly, the b-tagging uncertainties of the former decay also represent a sizeable contribution, compared to the one from the other object selections (trigger and leptons, for example). Most of these uncertainties result from a combination of statistical and systematic error, and only the latter depends on the technique used. This limited margin of improvement would however directly be translated into better precision in the HH analyses, but is out of the hands of the people pursuing them. Similarly, the various theoretical uncertainties on the HH signal and the dominant background $t\bar{t}$ samples are sizeable, but can only be improved by the theory community, through a better understanding of generation uncertainties, both in the parton shower and ME. Fortunately, this is an active area.

On the side of systematics directly affected by the techniques used in this analysis, ad-hoc uncertainties, used to stabilise the fit and discussed in Section 4.4, have a significant impact. This could be averted by using a less complicated multi-classification scheme, although the sensitivity might be affected. Finally, while uncertainties associated to the weights used to derive both data-driven estimations of the DY and the Fakes have a low impact, their associated non-closure uncertainties impacts are of one order of magnitude larger, which could motivate a search for more elaborate methods. This is especially true for the Fakes, for which uncertainties have been pessimistically estimated in order to cover any mismodelling.

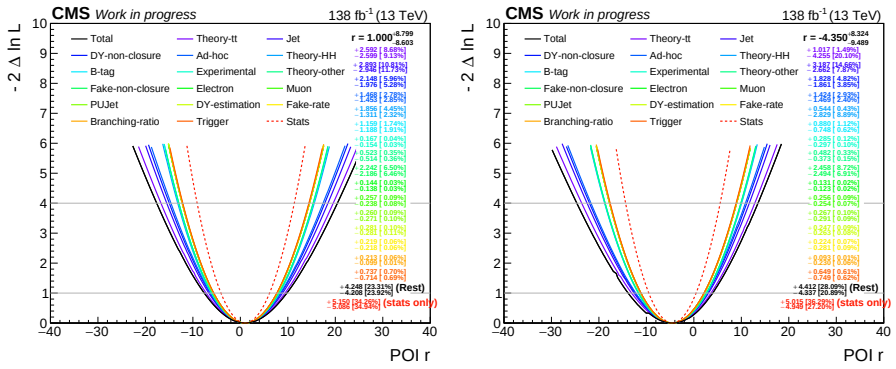


Figure 4.23 | Expected (left) and observed (right) $-2\ln L$ likelihood scan over the inclusive POI when incrementally freezing groups of systematics. *Total* and *Stats* refer to the likelihood when all nuisances are left floating and all frozen respectively. From the former, the groups of systematics defined in Appendix A.8 are frozen one by one, in the order of the legend, until reaching the latter. *Rest* refers to the nuisances related to the Monte Carlo limited sample uncertainties, using the Beeston-Barlow method of Section 2.4.3. The impact of each group of systematic is determined from the quadratic difference from the reduced uncertainties around the best-fit value, illustrated in the right legend (along with their relative contribution to the whole uncertainty range), and summarised on Table 4.14.

Table 4.14 | Effects of systematic uncertainties on the prefit yield variations, and on the POI observed and expected uncertainty range. The former is provided as a range or single value (in percents), determined from the more extensive Tables A.1 and A.2, while the latter are determined when incrementally freezing groups of systematics defined in Appendix A.8 and measuring the 1-σ variations on the $-2\ln L$ around the POI best fit.

Nuisance group	Prefit yield variations [%]		Postfit variations on POI μ	
	Background	Signal	Observed : μ = -4.35 $\Delta\mu = +8.32/-9.49$	Expected : $\mu = 1.00$ $\Delta\mu = +8.80/-8.60$
Nuisances applied to all samples				
Jet + MET	3.1-19.0	3.5-16.0	+3.19 / -2.66	+2.89 / -2.95
Experimental	3.6-4.8	3.6-5.5	+0.28 / -0.30	+0.17 / -0.15
Electron	7.2-8.0	2.8-8.1	+0.13 / -0.12	+0.14 / -0.14
Muon	3.0-3.8	2.0-5.0	+0.26 / -0.25	+0.26 / -0.24
PU jet ID	0.8-4.8	0.8-2.5	+0.27 / -0.29	+0.26 / -0.27
Higgs BR	0.01	2.5	+0.09 / -0.23	+0.21 / -0.10
Trigger	1.2-3.4	1.3-3.5	+0.65 / -0.75	+0.74 / -0.71
B-tagging	14.5-18.0	9.6-16.15	+0.88 / -0.75	+1.16 / -1.19
Theory nuisances				
Minor backgrounds	3.9-5.0		+0.48 / -0.37	+0.52 / -0.51
$t\bar{t}$	7.6-10.0		+1.02 / -4.25	+2.59 / -2.60
HH	4.0-21.8		+0.54 / -2.83	+1.86 / -1.31
DY data-driven nuisances				
ABCD weights	9.8-13.2		+0.25 / -0.26	+0.28 / -0.28
Non-closure	9.8-100.1		+1.83 / -1.86	+2.15 / -1.98
Fake nuisances				
Fake-rates	5.6-27.3		+0.22 / -0.28	+0.22 / -0.22
Non-closure	63-160		+2.46 / -2.49	+2.24 / -2.19
Fit related nuisances				
Ad-hoc norm	20-25		+1.42 / -1.47	+1.47 / -1.45
Limited MC	/		+4.41 / -4.34	+4.25 / -4.21
Statistical uncertainties				
Statistics	/		+5.01 / -4.95	+5.15 / -5.09

4.6 EFT reweighting

To study the HEFT couplings (κ_λ,κ_t,c_2,c_g,c_{2g}) defined in Section 1.7.2, the different sets of couplings (called *benchmarks*) of Tables 1.3 and 1.4 were defined as representative points of volumes in the coupling space that showed similar features in terms of m_{HH} and $\cos\theta^*_{\text{HH}}$. A reweighting procedure similar to Eq. (1.52), particularised to the case of HH searches using Eqs. (1.68) and (1.69), can be used on any non-resonant sample. The weight to be applied to signal events generated with the benchmark a to project them to a new benchmark b can be computed as

$$w(BM_a \to BM_b) = \frac{\sigma(m_{\text{HH}}, \cos\theta^*_{\text{HH}})|BM_b)}{\sigma(m_{\text{HH}}, \cos\theta^*_{\text{HH}})|BM_a)} \frac{\sigma_{tot}(BM_a)}{\sigma_{tot}(BM_b)}, \tag{4.23}$$

where the total cross-section σ_{tot} and differential cross-section as a function of the two variables of the HH system (m_{HH} and $\cos\theta^*_{\text{HH}}$) of both benchmarks can be obtained using the analytical form of Eqs. (1.68) and (1.69), after a fit of the A_i coefficients (following Ref. [137]) provided at the HH group level within the CMS collaboration. Including the NLO contribution of Eq. (1.69) can be done to further improve the modelling, even when starting from LO generated events. In addition, several samples can be reweighted and averaged to the sample benchmark to benefit from a better statistical precision, a procedure known as *many-to-many*. One technical inconvenient of the method is that the couplings used to generate the events of benchmark a must be known, and bugs that occurred during the production of several LO signal samples caused some uncertainty in that aspect. This can be alleviated by evaluating the total and differential cross-sections of the initial benchmark a from a histogram of (m_{HH}, $\cos\theta^*_{\text{HH}}$) filled with generator weights so that Eq. (4.23) is agnostic of the couplings used during the event generation. The effect of these weights on a generator-level variable can be evaluated on Fig. 4.24, showing good agreement even compared to a NLO distribution.

To compare the effect of starting from either LO or NLO samples for the reweighting, it was applied to both cases (on 13 and 4 samples respectively) to 3 sets of couplings used to derive the non-resonant results with the interpolation method described in Section 1.7.3. The comparison with some reconstructed variables and DNN signal node distributions is on Figs. 4.25 and 4.26 respectively. The LO reweighted samples have smaller statistical uncertainties given the much larger number of events and samples considered, yet show some mismodelling, especially in the tails of the distributions where the NLO reweighted perform better. Although the agreement is reasonable in the GGF node of the DNN, the effect of this modelling is large in the VBF node that is however not the most sensitive category for these signal samples, as illustrated by the yield in the legend. Upper limits and scans in κ_λ were performed for the three cases and showed similar sensitivity. This served as a cross-check of the reweighting, but also a motivation to rather use the NLO samples in the reweighting method.

A potential shortcoming of the reweighting method that was illustrated by this comparison is the limited amount of events used in the determination of the weights and their associated statistical fluctuations. The largest deviations from unity occur in low populated regions of high m_{HH} and should have limited impact, unless applied to events of a benchmark with very different kinematic properties. This was deemed unlikely given the spectrum observed in Fig. 1.15. Some distributions of the weights are in Appendix A.6.

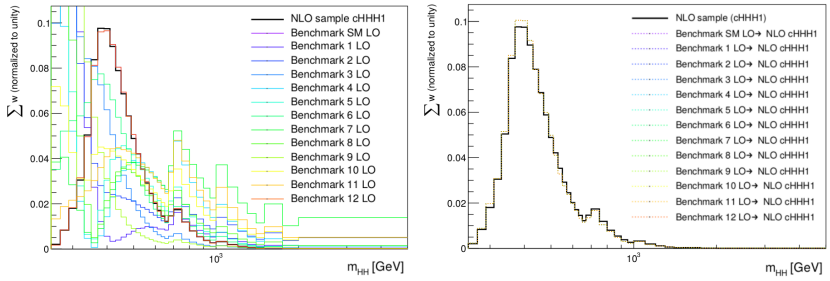


Figure 4.24 | Distribution of the invariant HH mass using the generator-level Higgs bosons before (left) and after (right) application of the weight defined in Eq. (4.23). Several LO HH → bbWW samples corresponding to the benchmarks defined in Table 1.3 are compared and reweighted to a NLO sample generated with SM couplings.

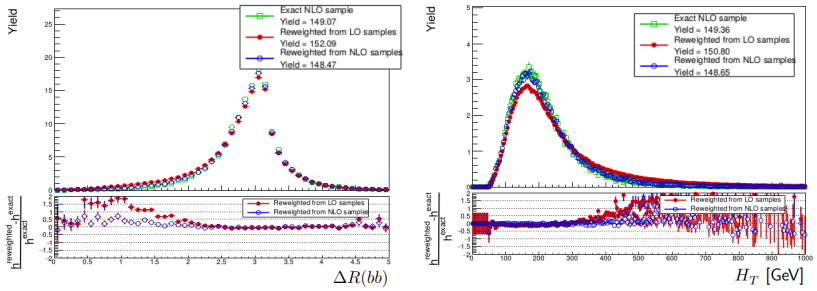


Figure 4.25 | Comparison of a HH → bbWW NLO sample generated with $\kappa_\lambda = 0$ (rest of the couplings at their SM value) with the ensemble of LO and NLO samples reweighted to that specific set of couplings. Left : ΔR angle between the two jets with highest b-tag score in the resolved category. Right : scalar p_{T} sum of all jets.

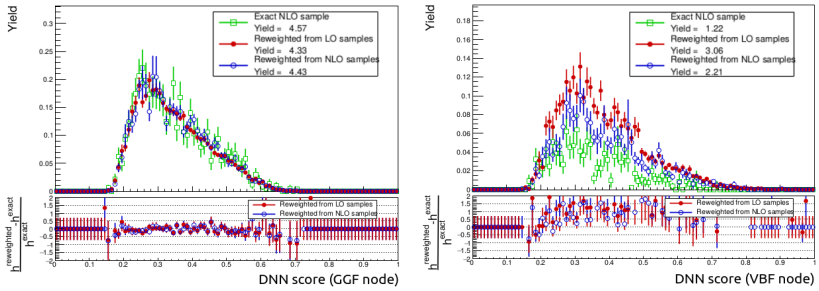


Figure 4.26 | Comparison of a HH → bbWW NLO sample generated with $\kappa_\lambda = 5$ (rest of the couplings at their SM value) with the ensemble of LO and NLO samples reweighted to that specific set of couplings. Left : DNN score distribution in the GGF node in the resolved 1b category. Right : DNN score distribution in the VBF node in the resolved 1b category.

4.7 Results

The rates for the resonant and non-resonant HH production are determined through a binned simultaneous likelihood fit of the DNN score in each category defined in Section 4.2.6, the statistical procedure is the one described in Section 2.4.3.

4.7.1 Resonant analysis

The expected and observed limits for the spin-0 and spin-2 resonances as a function of the m_{HH} is presented on Fig. 4.27. The bulk radion hypothesis defined in Section 1.6 with $\Lambda = 3$ GeV and $kL = 35$ can be excluded above 440 GeV. No exclusion limits can be set on the bulk graviton however for lack of enough sensitivity. A comparison with other CMS results is discussed in Section 5.1, and the results split per channel and data-taking year as well as the effect of including the HME information are in Appendix A.5.

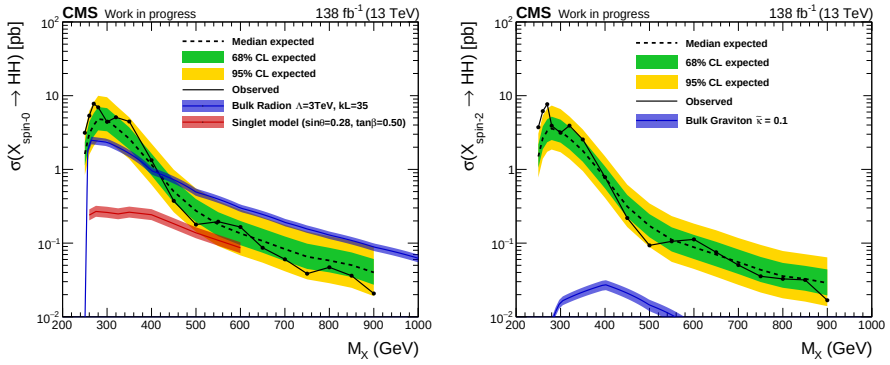


Figure 4.27 | Expected (dotted line) and observed (solid line) limits as a function of m_{HH} for the spin-0 (left) and spin-2 (right) resonances, with theoretical predictions and their associated uncertainties (PDF, α_S and scale uncertainties combined).

4.7.2 Non-resonant analysis

The results of the non-resonant analysis are interpreted through the anomalous coupling framework defined in Section 1.7.2, which represents a large parameter space to scan. The formalism developed in Section 1.7.3 can be used in the statistical analysis by redefining the signal strength as

$$\text{GGF samples} \ : \mu.\mu_{\text{ggHH}}.\sum_i f_i^{\text{ggHH}}(\kappa_t, \kappa_\lambda)$$
$$\text{VBF samples} \ : \mu.\mu_{qq\text{HH}}.\sum_i f_i^{qq\text{HH}}(\kappa_\lambda, \kappa_V, \kappa_{2V}), \tag{4.24}$$

where μ_{ggHH} and μ_{qqHH} are new POIs included with the couplings that can be varied or kept fixed. In this context the f_i correspond to the $\boldsymbol{c}(\alpha)^T(\boldsymbol{C}^{-1})_i \boldsymbol{d\sigma_i/dx}$ in Eq. (1.72) where α is the set of couplings for which the limit is extracted, $d\sigma_i/dx$ the shape of one of the distribution provided for the interpolation – x being the DNN score distribution – and i runs over the 3 (6) samples used in the interpolation for the GGF (VBF) case. This interpolation is solely applied on signal shape assuming constant background shapes, preventing the use of a parametric DNN as was done in the resonant analysis.

The physics model [424] that describes this parameterisation of the likelihood fit was included in the INFERENCE package [425], developed for all HH analysis within CMS. In addition to the model, the package also included the non-trivial effects of the coupling modifier κ_λ over the measurement by its impact on both the BR of the Higgs boson decay and the single Higgs processes treated as background, κ_λ-dependent NNLO/NLO K-factor described in Section 1.5.2, and additional cross-section uncertainties that may also depend on κ_λ.

Standard model limits

The observed and 95% CL expected limits on the SM cross-sections are on Fig. 4.28, expressed as a multiple of the theoretical cross-sections. The SM limit is recovered as a special case of Eq. (4.24) when only the f_i associated to the SM sample is non-zero. The upper limits of the full Run-2 statistics amounting to 138 fb^{-1} are set at the observed (expected) values of 15 (18) times the SM theoretical prediction for the inclusive cross-section, and 386 (381) for the pure VBF production mode.

EFT benchmark limits

The observed and expected limits on the benchmarks points defined in Section 1.7.2 with the reweighting described in Section 4.6 are on Fig. 4.29. Upper limits on the benchmarks defined in Table 1.3 range from 134 to 1297 fb, while for the benchmarks of Table 1.4 they range from 271 to 725 fb, at 95% CL. The differences in sensitivity can be linked to the differences of the m_{HH} spectra in Fig. 1.15, the main variable characterising the HH production. The most sensitive benchmark labelled 2 has a two-peak structure with a very large m_{HH} tail at high values, which makes it easy to distinguish from the backgrounds, contrary to benchmark 7 for which the sensitivity is thus lower. This can also be observed, to a lesser extent, in benchmarks 10 and 12 that are next in terms of low sensitivity.

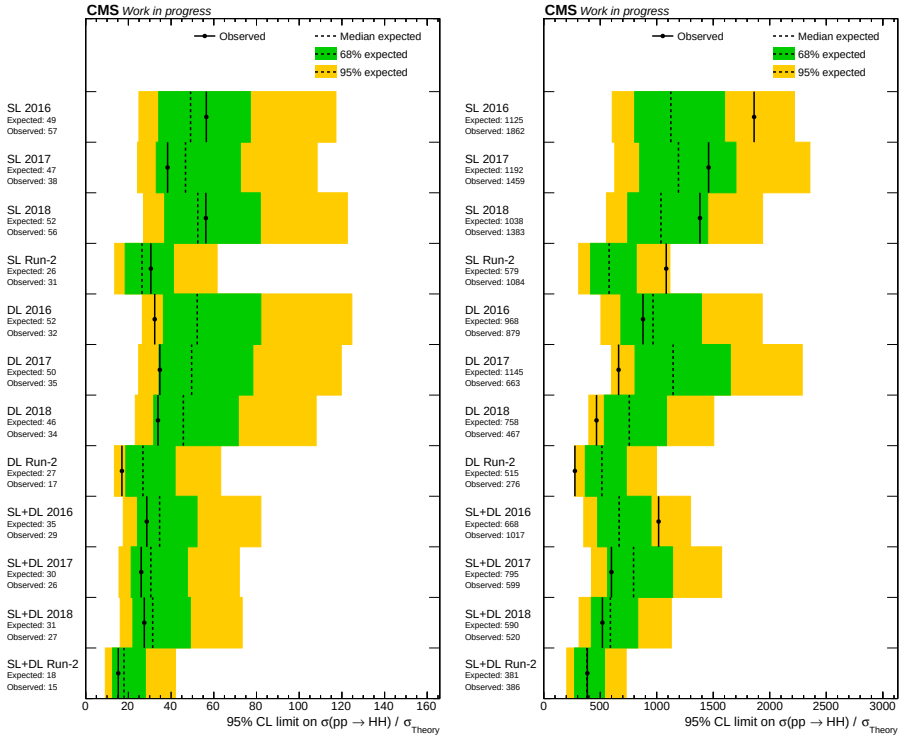


Figure 4.28 | Observed (solid line) and 95% CL expected (dotted line) upper limits as a function of the SM ($\kappa_\lambda = 1$ and $\kappa_{2V} = 1$) theoretical cross-sections of the inclusive (left) and VBF (right) cross-sections. The limits are shown for each channel and data-taking year, as well as their combination.

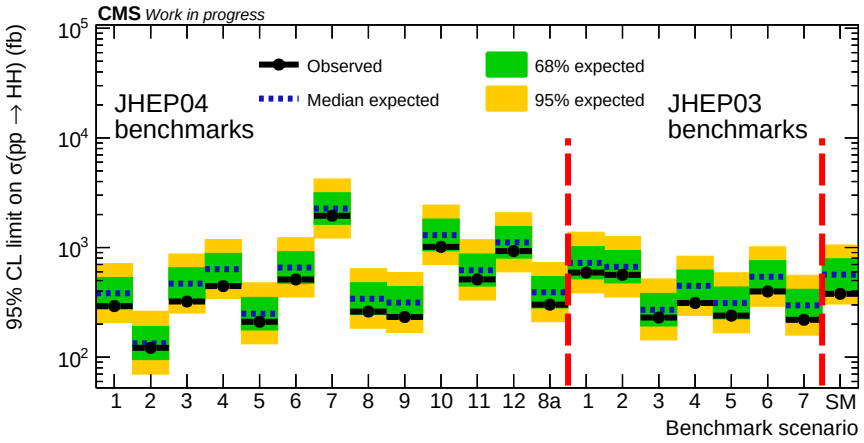


Figure 4.29 | Observed (solid line) and 95% CL expected (dotted line) limits on the cross-section expressed in fb for the different benchmarks points defined in Table 1.3 (JHEP04) and Table 1.4 (JHEP3) (right). The SM point was also obtained through the reweighting as a validation of the method, with comparable upper limit to the actual SM measurement.

Coupling scans

Assuming that the HH signal exists, including the VBF production mode, with the properties predicted by the EFT extension of the SM, it is possible to measure the values of the different coupling modifiers. Their constraints assuming all other couplings at SM values as well as best-fit intervals are on Table 4.15. Two-dimensional likelihood coupling scans assuming the other couplings with SM values are on Fig. 4.31. Additional results, including one-dimensional likelihood scans and exclusion regions in two-dimensional scans, are in Appendix A.7. All couplings are measured within their SM expectation values.

The κ_λ scan of Fig. 4.30a illustrates the effect of the destructive interference between the triangle and box diagrams of Fig. 1.11. Since only the triangle diagram depends on κ_λ, and the fact that the invariant m_{HH} distribution as well as other kinematic distributions depend on these two diagrams interaction (as visible on Fig. 1.10), the acceptance and efficiencies have a non-trivial dependency on κ_λ around the maximum interference at $\kappa_\lambda = 2.4$. Outside the non-monotonous region, positive values of κ_λ involve a suppression between the two diagrams, such that the sensitivity will be lower (therefore a higher upper limit). On the other hand, negative values mean that both contributions of the two diagrams in the m_{HH} invariance distribution are kept and the selection efficiency is higher. At extreme values of $|\kappa_\lambda|$, the triangle diagram heavily dominates and both sides of the upper limits asymptotically converge. Aside from the pure HH effects, the non-trivial dependency of the single Higgs processes and decay BR (cf. Figs. 1.8a and 1.8b) on κ_λ can complicate the interpretation of the sensitivity shape, since they have a symmetric and asymmetric effects, respectively. This maximal interference effect at $\kappa_\lambda = 2.4$ can also be observed in the two-dimensional likelihood scans of Figs. 4.31a and 4.31b, but not on the one in Fig. 4.31c that only considers the VBF production mode on which the trilinear coupling has a weaker impact than in GGF. A similar non-trivial interaction can be observed in both upper limits as a function of κ_{2V} and c_2 on Figs. 4.30b and 4.30e. The same explanation can be given for those two cases, where small values of the coupling imply non-trivial interactions between the diagrams of Fig. 1.16 and Fig. 1.14 respectively, while higher values cause their attached diagrams to dominate over the others and produce a more uniform upper limit.

As expected the sensitivity on κ_t and κ_V is relatively poor, and in the latter case there are indications that the VBF model used in the interpolation might break down for values higher than $|\kappa_V| > 3$, values that are excluded by single Higgs searches anyway. Their one-dimensional upper limits scan were included for completeness, but the interest is rather in the two-dimensional likelihood scans when their measurement is combined with the double Higgs couplings. When the constraints on κ_t are relaxed on Fig. 4.31a, the measurement of κ_λ loses in sensitivity, as expected from Fig. 4.4b. The disjoint intervals come from the stronger quadratic and quartic dependencies

on κ_t from the square of the box and triangle diagrams in Eq. (1.71), compared to the cubic dependency in the interference term that is not sufficient to resolve the degeneracy of the sign. This effect is also visible on some one-dimensional likelihood scans in Appendix A.7. Comparatively, when either κ_{2V} or c_2 are relaxed in Fig. 4.31c and Fig. 4.31b respectively, the κ_λ measurement also suffers. This is especially true in the latter case where an anti-correlation appears, since high values of κ_λ and c_2 tend to produce a m_{HH} spectrum with lower and higher values respectively, their effect on the signal selection efficiency is therefore opposite. The parabola shape in the κ_V-κ_{2V} scan of Fig. 4.31d can be explained by the boost of the Higgs boson pair outside $\kappa_{2V} = \kappa_V^2$, providing a very specific signature that increases the sensitivity of the analysis in this region. This effect can also be observed in the one-dimensional upper limits of Fig. 4.30d, where the two peaks of the degeneracy between $\kappa_V = +1$ and $\kappa_V = -1$ result from the quadratic and quartic dependency on the diagrams in Eq. (1.74), as well as the linear and cubic dependencies in the diagrams that are combined with the one originating from the H $\rightarrow$ WW decay.

Table 4.15 | Best fit estimation of the values of the various EFT couplings, with one and two standard deviations obtained from the likelihood scans (with the two-peak structure of the κ_t likelihood)), along with the observed (expected) 95 % CLs exclusion limits extracted from the upper limit scans, assuming all other couplings set to SM values.

Coupling	Best fit	1-σ errors	2-σ errors	Exclusion regions
κ_λ	4.61	+5.78 / -6.96	+9.55 / -11.45	-7.86 (-8.48) $< \kappa_\lambda <$ 14.34 (15.03)
κ_{2V}	1.05	+1.62 / -1.61	+2.61 / -2.60	-1.64 (-1.86) $< \kappa_{2V} <$ 3.77 (3.99)
κ_t	-1.31	+0.41 / -0.33	+3.52 / -0.70	-1.84 (-1.96) $< \kappa_t <$ 2.09 (2.28)
	1.48	+0.39 / -0.48	+0.74 / -3.48	
κ_V	0.01	+1.05 / -1.23	+1.51 / -1.51	-1.65 (-1.79) $< \kappa_V <$ 1.74 (1.86)
c_2	0.21	+0.50 / -0.50	+0.92 / -0.90	-0.74 (-0.97) $< c_2 <$ 1.18 (1.42)

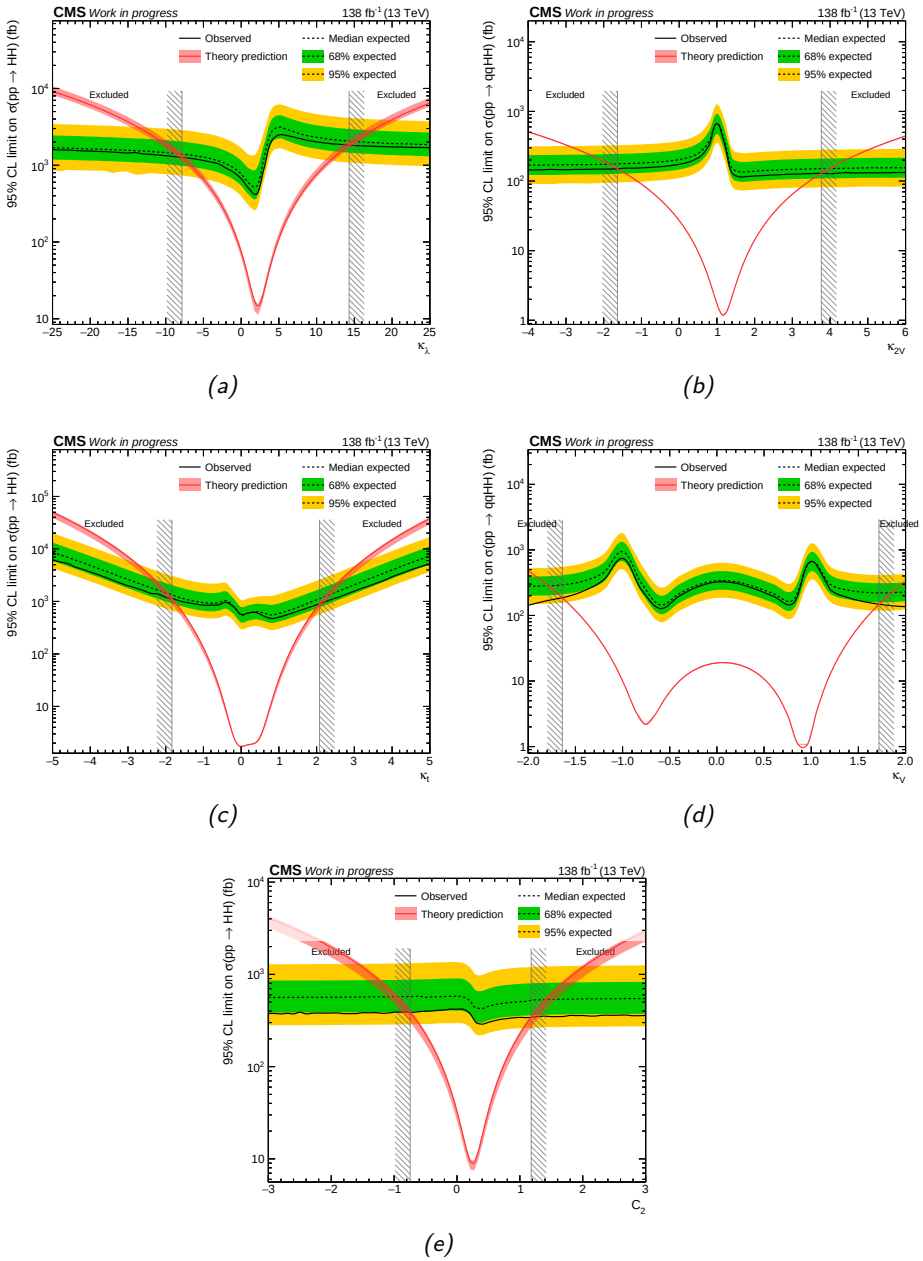


Figure 4.30 | Expected (dotted) and observed (solid line) upper limits on the non-resonant cross-section as a function of the different couplings : κ_λ the trilinear Higgs coupling and κ_{2V} coupling between two weak and two Higgs bosons that are the only couplings accessible at tree-level in double Higgs measurements (top); κ_t and κ_V the top quark and weak boson couplings to the Higgs boson for which single Higgs measurements are expected to provide much better sensitivity (middle); and c_2 the anomalous coupling between two top quarks and two Higgs bosons which is the only pure BSM coupling investigated here (bottom). In all cases, the expected theoretical cross-section from HEFT is shown in red with uncertainty bands, and the observed exclusion regions using dashed delimitations.

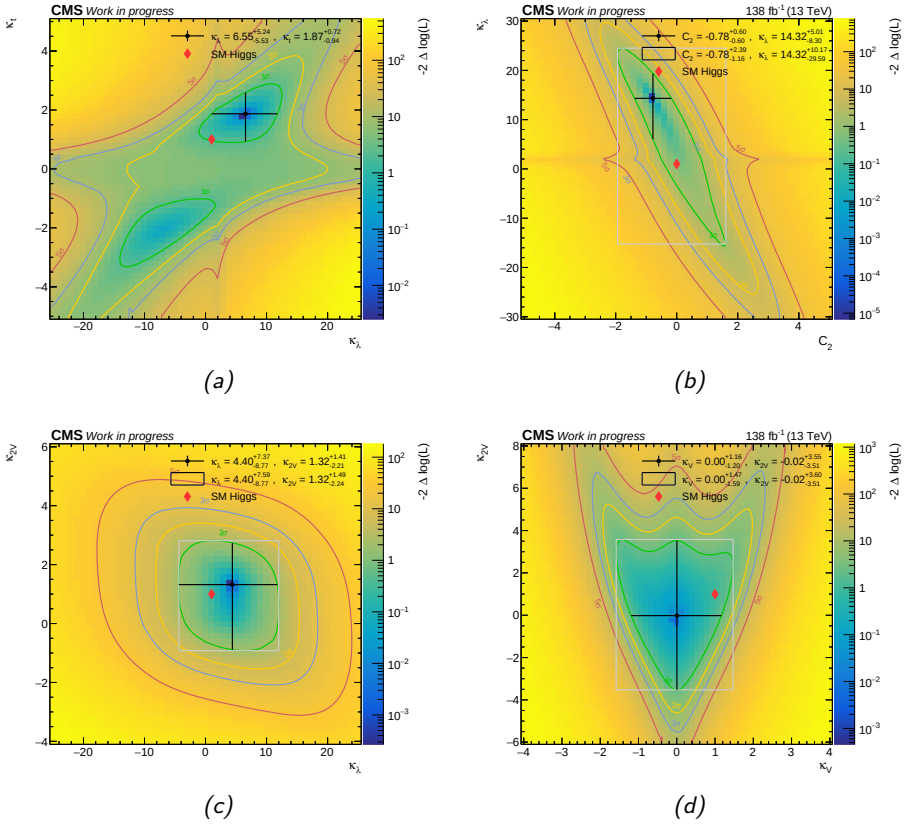


Figure 4.31 | Likelihood scans as a function of the different couplings κ_λ, κ_t, κ_V, κ_{2V} and c_2 in two-dimensional spaces. The uncertainties at different levels on both POIs are displayed as contours and the best-fit value is displayed marginalised, considering each coupling set to its best-fit value or allowed to vary.

4.8 Personal contributions

The analysis performed in this chapter is the result of a collaboration between different groups from the following universities : UCLouvain[1], RWTH[2], DAE[3], SKKU[4], Texas A&M [5] and KBFI[6].

The tasks in the analysis I was responsible for were the DY data-driven estimation of the DL channel in both resonant and non-resonant searches; the implementation, validation, and cross-checks of the EFT reweighting into the different benchmarks; the resonant DL channel including the HME implementation and the parametric DNN used in the likelihood fit; and the pre-approval talk that I presented on behalf of the group. I also implemented the entire analysis workflow in the BAMBOO [426] framework for both searches and channels, including the development of the tools used in UCLouvain to produce the datacards, and performed the statistical analysis with COMBINE and the HH inference tools to obtain the results showed in Section 4.7. I have trained DNNs in the non-resonant searches as a cross-check of the ones used officially in the analysis with similar performance, performed various synchronisation exercises with the other groups workflows that shed lights on bugs from each side, as well as all the statistical tests ensuring the reliability of our measurement and the likelihood fit. I was involved in the study of various alternatives during the conception of the analysis strategy : the comparison of the EFT reweighting using LO and NLO signal samples in Section 4.6, confirming in the general HH group the motivation to use the latter; several methods for the estimation of the DY process from data and the development of the decorrelation script for its non-closure; the various ways of linearising the 2D histograms of the resonant DL and the optimal binning choice; the different ways to properly weight each event in the training of the DNN, including the downscaling of the signal categories; and a comparative study of the two sets of selections that were developed in parallel by RWTH and KBFI that required simultaneous synchronisation with both analysis workflows and the evaluation of their sensitivity.

[1] Université catholique de Louvain, Belgium
[2] Aachen University, III. Physikalisches Institut (A)
[3] Saha Institute of Nuclear Physics, HBNI, India
[4] Sungkyunkwan University, Suwon, South Korea
[5] Texas A&M University, USA
[6] National Institute of Chemical Physics and Biophysics, Tallinn, Estonia

Conclusion

The work presented in this document illustrated the analysis of 138 fb^{-1} of data collected during Run-2 by the CMS detector at the LHC at $\sqrt{s} = 13$ TeV. The data was used to put constraints on the SM predictions of the double Higgs production mechanism (HH) and on BSM scenarios. In particular, the trilinear Higgs self-coupling is a parameter that can only be accessed via HH measurement and several other couplings were probed, such as the coupling to top quarks and to vector bosons. The non-resonant search was extended to EFT scenarios using a reweighting technique, and a resonant search of a heavy state decaying to a pair of Higgs bosons was also carried out. This analysis was performed in one particular channel where one Higgs decays as $H \to bb$ and the other into $H \to WW$, such that the two final states $b\bar{b}\ell^{+}\nu\ell^{-}\bar{\nu}$ and $b\bar{b}\ell\nu q\bar{q}$ take into account the various decays of the W boson, without excluding additional HH decay channels resulting in the same final state.

The sample of data investigated in this work consists of events matching the signature of the decay of each Higgs boson. For $H \to bb$ two topologies are considered, one where there are at least two distinguishable jets and at least one is b-tagged, and one where the two jets are merged into a larger jet that has at least one b-tagged subjet. For the other Higgs decay into $H \to WW$ at least one W boson is required to decay leptonically to allow for triggering, the other can produce an additional lepton or a pair of jets out of which at least one must be reconstructed. The leptons considered in both channels are electrons, muons and leptonically decaying taus. The contamination from SM backgrounds were simulated using the CMS detector description, mostly top quark pair production and W boson with associated jets. Data-driven techniques were used to estimate the Drell-Yan with heavy flavour jets and the misidentified jets as leptons, using appropriate control regions. A multi-classification approach was followed with Deep Neural Networks (DNNs) to enhance the sensitivity on all the signatures considered in the analysis.

The background categories in the classification were used as control regions to check the agreement between simulations and data and to constraint the nuisance parameters in a likelihood fit. The good agreement prompted us to produce observed (expected) 95 % CL upper limits on the inclusive cross-section at 15 (18) times the expected SM non-resonant cross-section, and 386 (381) times for the VBF production mode. An interpolation method assuming EFT deviations was used with

a physics model including the single Higgs processes scaling to produce likelihood scans in the various couplings considered. The 95% CL constraints were set on the two couplings that can be accessed by HH production at tree level as -7.86 (-8.48) $< \kappa_\lambda <$ 14.34 (15.03) and -1.64 (-1.86) $< \kappa_{2V} <$ 3.77 (3.99), and a BSM coupling as -0.74 (-0.97) $< c_2 <$ 1.18 (1.42). Upper limits were placed on several benchmarks corresponding to points in a five-dimension BSM coupling space, without any significant excess observed. The resonant search was performed assuming spin-0 and spin-2 resonances with the narrow width approximation to provide upper limits on the production cross-section on a mass range between 250 and 900 GeV, also without any significant excess.

In parallel, a novel method to combine the MEM with regressive DNNs in an attempt to accelerate the computation has been presented. The MEM provides a way to interpret the data using both the knowledge of the SM and the detector effects, allowing for a very efficient use of the limited amount of data. Its computation time was nonetheless too prohibitive for LHC analyses, and the hybrid method described is this document represents an attempt to circumvent this issue. It was applied to three different processes, the 2HDM scenario of a decay H → ZA, a top quark pair production and a Drell-Yan process with heavy flavour jets. Its regression error was showcased in several applications, out of which it was determined that no loss of performances was observed. When applied as a benchmark on a real-life analysis, the hybrid method showed very large gains in computation time, allowing it to be used in future LHC analyses. Several potential improvements on the application of the methods, the techniques used to obtain the integration values and tools to perform the integration were summarised.

Applied to the HH → bbWW search, the MEM, through its acceleration by DNN regression, promises to improve the sensitivity of the analysis, through a better consideration of the correlation between the different visible particles, as well as considering physical configurations of neutrinos and the combinatorial challenge of parton-jet assignment. At the very least, the MEM weights could be used as input to the multiclassification to reduce the complexity of the learning task and the framework which proved to be a limiting factor of the analysis. Any mismodelling of the simulations in the training set compared to data in small regions of the phase space can be exacerbated in the classification predictions because of the correlations between the large amount of input variables. The ambitiously large number of classes defined in the DNN had to be reduced to a smaller number of categories to absorb the low statistics categories that started showing some mismodelling. The bbWW decay channel, with its large background contamination and high dimensional parameter dependence of the resonant and non-resonant signals, represents the exact conditions in which the hybrid MEM method described in this document could leverage the gains of a DNN regression.

Chapter 5 Outlook

5.1 Results comparison

The results of the non-resonant HH searched carried out with the data of the CMS and ATLAS experiments during the 2016 data taking year are on Fig. 4.2, they report an observed (expected) 95 % CL upper limit of 22.2 (12.8) and 6.9 (10) times the theoretical predictions respectively. Using the 138 fb^{-1} of data collected during the entirety of Run-2, the HH decay channel reported in this document reaches the combined sensitivity of all other channels in 2016, with the 18 (15) upper limit reported in Fig. 4.28. Obviously, the almost factor five in luminosity increase between this analysis and the previous one using the same channel and only the 35.9 fb^{-1} data of 2016, that quoted limits at 78.6 (88.8) times the SM prediction, only accounts for part of the sensitivity gain. A fair comparison would rather be between DL channel of this work in the same year with an upper limit of 32 (52) times the SM prediction. Aside from the improvements on the b-tagging such described in similar channels in Section 4.1, the better sensitivity can be explained by several factors. First, the larger phase-space probed with the relaxation of the number of b-tagged jets to one, and the consideration of the boosted region. Then the more complex multiclassification and advanced architecture allowed for a better extraction of information from the data. The difference is further increased when considering the SL channel that the 2016 analysis had not considered, improving the sensitivity to 35 (29), more than a factor two improvement using the same set of data. In comparison, the same analysis was performed by the ATLAS experiment, albeit only in the DL channel, that reported a limit of 40 (29) times the inclusive theoretical HH cross-section with a luminosity of 139 fb^{-1} [427].

The comparison of the results from both CMS and ATLAS for the Run-2 datasets are displayed on Fig. 4.3. Although not one of the most sensitive channel, the one considered here is nevertheless important to take into account for a combination of results, as it currently holds the first place outside the three most sensitive channels (the "silver bullets") in terms of inclusive cross-section. The constraints on κ_λ however are of the same order as the less sensitive channels.

Aside from the inclusive and κ_λ results, analyses performed on Run-2 data included

additional experimental results, most notably the VBF cross-section and other EFT couplings. With an observed (expected) upper limit on the former of 386 (381) times the SM expectation, which puts the bbWW channel in the fourth position in terms of VBF sensitivity compared to the three dominant channels listed in Table 4.1. Given the poorer sensitivity of the remaining channels, they did not perform this measurement. Similarly, the constraints on κ_{2V} are on the same position compared to the other channels, yet getting close to the sensitivity of the bb$\gamma\gamma$ analysis.

Regarding the c_2 coupling, only the multilepton CMS analysis [377] has published constraints at 95 % CL of $-1.05\ (-0.96) < c_2 < 1.49\ (1.37)$, slightly worse than the one measured in this bbWW analysis, as expected given the lower sensitivity. Finally, the constraints on the fermion and weak boson couplings κ_t and κ_V listed in Table 4.15 are particularly poor compared to the single Higgs measurements on Fig. 5.1 (κ_f referring to the fermion coupling modifier, which is mostly driven by κ_t). Their current measurement are in agreement with the SM value $\kappa_f = \kappa_V = 1$ within 10 %, which is about ten times better than this search in the HH $\rightarrow$ bbWW channel as anticipated.

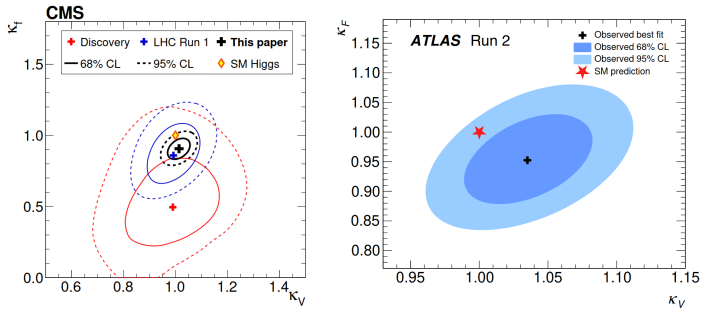


Figure 5.1 | CL contours at 68 % and 95 % levels of the fermion and vector boson coupling modifier (κ_f,κ_V) obtained in single Higgs measurements performed over Run-2 dataset by both CMS (left, from [382]) and ATLAS (right, from [428]) collaborations.

Given the current combined sensitivity of Fig. 5.2, there is a strong belief that during Run-3 the upper limit will be able to reach the theoretical cross-section, especially in the case of a combination between CMS and ATLAS. As described in Section 4.1, a combination between single and double Higgs measurements could improve the κ_λ measurement by including the indirect effect of the coupling in loop corrections, but most importantly include the much tighter constraints that single Higgs analyses have on κ_t. Although challenging by its coordination aspect, these combinations could open the road to the first attempt at a discovery of the HH process, and possibly prove the existence of the trilinear Higgs coupling. Since all Run-2 analyses have exceeded their expectations, compared to a naive extrapolation of a factor two based on the increased luminosity, there is reason to believe that the expectations for Run-3 are likewise undervalued.

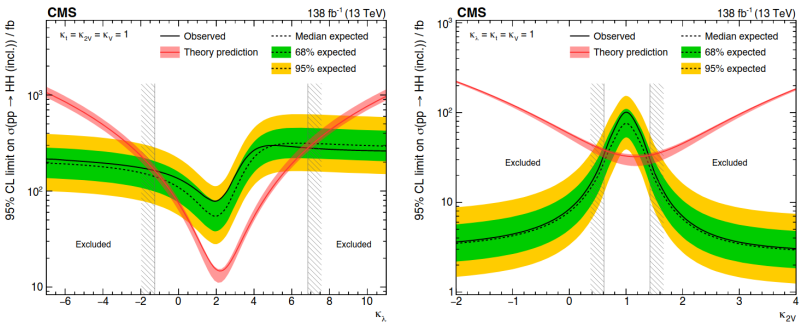


Figure 5.2 | Observed and expected upper limits of the two couplings κ_λ and κ_{2V} using double Higgs measurement of the CMS collaboration in the same channels as Fig. 4.3a [382], together with exclusion regions.

The comparison can also be done for the resonant analysis on Fig. 5.3 with the 2016 previous result, as well as the addition of the boosted analysis targeting the same final state [392]. Compared to the 2016 analysis, the improvement in sensitivity in the DL (both DL and SL) channel ranges from around 10% (50%) for low m_{HH} region, up to around 150% (200%) for higher masses. This also illustrates the gain solely attributed to analysis techniques and the inclusion of more sensitive categories. The boosted analysis uses the same amount of data as in this work, except that boosted topologies of both Higgs decays are considered, making it more sensitive to higher resonant masses where the background contamination is lower, about a factor two better than this analysis in the coverage region around 900 GeV.

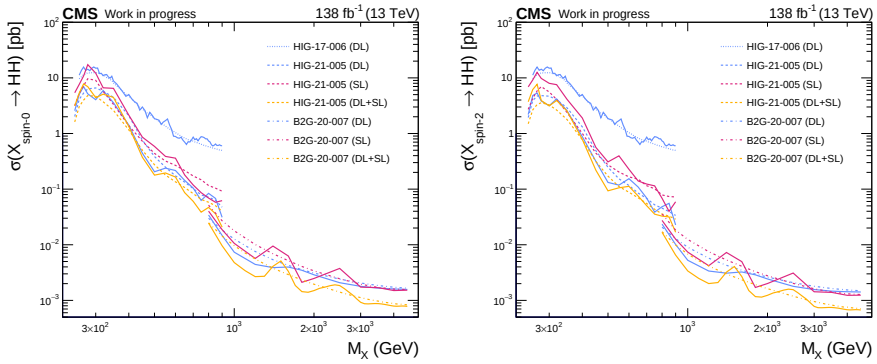


Figure 5.3 | Resonant observed (solid line) and expected (dotted line) upper limits for the analysis presented in this work (HIG-21-005), the bbVV analysis applied on 2016 data (HIG-17-006, from Ref. [395]) and the analysis targeting the same final state with Run-2 dataset but looking for heavier masses and boosted topologies for both the H → bb and H → WW decays (B2G-20-007, from Ref. [392]), for spin-0 (left) and spin-2 (right) resonant states.

5.2 Double Higgs production in future searches

5.2.1 HL-LHC

After Run-3 the LHC will start the so-called *Phase-2* upgrade where both the accelerator and the detectors will be improved to sustain the conditions of the HL-LHC accelerator expected to last until 2040, namely a higher luminosity, expected to reach 3000 fb^{-1} at the end of the three runs planned, and pile-up peaking at around 200 interactions per bunch crossing. Most of the detectors improvements are aimed at providing a higher granularity to cope with the increased number of particles, but also to cope with the higher radiation levels.

Several prospective studies have been carried out in the double Higgs sector both in CMS [429] and ATLAS [430] in an attempt to predict the discoveries and measurements that could be unveiled at the end of the HL-LHC data-taking, using appropriate simulations, similar methods as Run-2 and extrapolations of the performances of the detector. CMS and ATLAS report an expected 2.6 σ [163] and 3 σ detection of HH using five and three different channels respectively, and a combination of both experiments is expected to reach at least 4 σ [431, 432]. As interesting is the measurement of the trilinear coupling, whose projections are illustrated on Fig. 5.4. 95% CL intervals of $-0.18 < \kappa_\lambda < 3.6$ and $-0.4 < \kappa_\lambda < 7.6$ could be reached by CMS and ATLAS respectively, short of excluding $\kappa_\lambda = 0$.

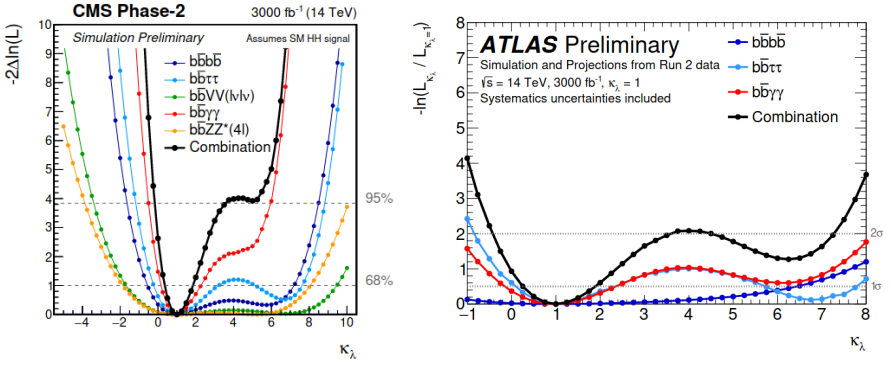


Figure 5.4 | Expected likelihood scan as a function of κ_λ for CMS (left, from Ref. [429]) and ATLAS (right, from Ref. [430]) for the different channels considered and their combinations. In all curves both the statistical and systematic uncertainties have been considered, assuming 3000 fb^{-1} of data and some assumptions on the performances of the future detectors, and the likelihood is built assuming SM parameters, in particular $\kappa_\lambda = 1$.

However since all Run-2 analyses exceeded the expectations by a large margin thanks to a lot of efforts from both collaborations, there is reason to believe that these prospective studies have been at the very least conservative. It is therefore very likely that the HH process could be observed at the end of the HL-LHC and the trilinear coupling confirmed positive to some confidence level, especially considering the very plausible breakthroughs that experimental physics could see emerge in the coming two decades.

5.2.2 Future colliders

Several future colliders are on the table, but they can be loosely classified based on two considerations : what particles they use in collisions and whether they use linear or circular architecture. The different types will be described in this section with a few of the most concrete examples (without any attempt at exhaustiveness) and their impact on HH searches will be highlighted.

Lepton colliders – mostly e^+e^- colliders – are viewed as *precision machines* through their ability to fix the center-of-mass energy to a precise value, contrary to hadron colliders using composite protons. The absence of substructure reduces the amount of theoretical uncertainties (for example, from UE and PDFs) and allows the constraint of event kinematics through the knowledge of the center-of-mass energy. The cleaner experimental environment, without pile-up, high radiation level and most importantly a trigger system vastly reduces experimental hurdle and the amount of systematic uncertainties. Hadron colliders – mostly pp colliders – tend to be seen as *discovery machines*. With a wide range of center-of-mass energies, it can scan for resonances and a larger range of phenomena. This was typically the case for the Higgs discovery in 2012. The luminosity and pile-up can also induce a much higher amount of data, at the cost of more complexity. Contrary to the lepton colliders, hadron colliders can have access to gluon induced production mechanisms. Other types of accelerated particles have been proposed, such as in muon colliders which have some qualities over electron colliders, yet are still in their infancy and several issues have to be ironed out. They will not be tackled in the rest of the section. Hybrid accelerators with electron-hadron collisions also exist, they could prove valuable in studying QCD effects in deep inelastic scattering, the structure of the proton and untested QCD effects, precise PDF measurements and could have a relatively rich program in EW precision studies. As hybrid colliders, they benefit from the advantages of both lepton and hadron colliders for discoveries, however also suffer from the issues of both.

Circular colliders typically display higher luminosities than their linear counterparts, and the possibility to install several detectors at different IP can multiply the physics yield. Their major limitation comes from synchrotron radiations that can become a significant energy loss factor for electrons, reducing the luminosity as the energy

increases. On the other hand, the energy limitation in linear colliders only depends on the accelerating technology and the length of each arm. The trade-off therefore lies between the energy and the luminosity, as illustrated on Fig. 5.5 in the case of e^+e^- collisions. Several projects are being studied for linear colliders : mostly the ILC [433, 434] in Japan and CLIC [435, 436] at CERN, both up to 50 km. Circular colliders include mainly the FCC-ee [437] as a 100 km extension of the LHC and CEPC [438, 439] in China with similar size. Regarding hadron colliders, only circular colliders are really considered as the heavier protons or hadrons are less subject to synchrotron radiation, the most likely candidate is the FCC-hh [440], though a possible extension of the LHC called HE-LHC with stronger magnets and accelerator complex could reach up to 27 TeV. In terms of hybrid ep colliders, the extension program of the LHeC [441] or the future FCC-he are considered. Heavy ion programs should also be considered in the programs of hadron colliders, but they have little impact over searches in the scalar sector and will not be covered.

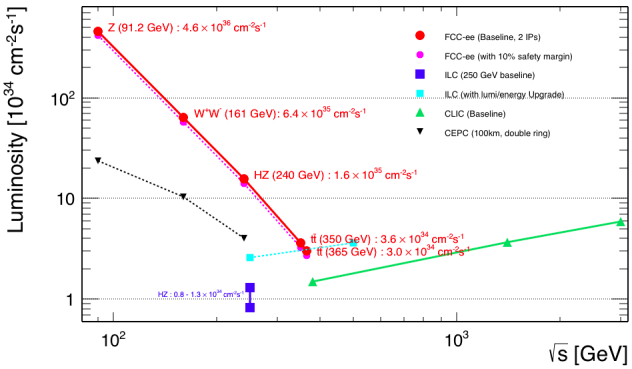


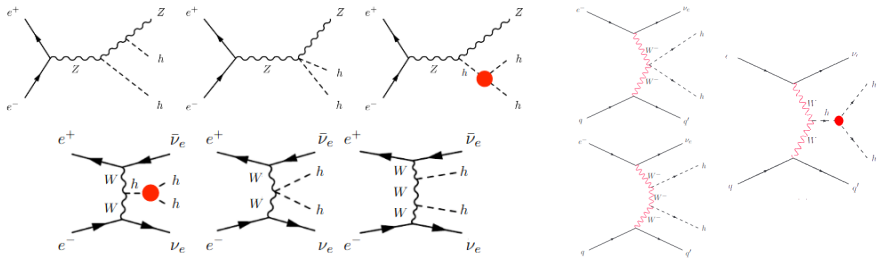
Figure 5.5 | Luminosity as a function of the energy for different proposed colliders : circular (FCC-ee, CEPC) and linear (ILC, CLIC) colliders. From Ref. [442].

In the context of the HH searches the main difference between each collider boils down to the production mechanisms and associated cross-sections. Compared to the main four diagrams of pp colliders in Fig. 1.11, e^+e^- colliders can produce a pair of Higgs bosons using two main modes as illustrated on Fig. 5.6a and a single main mode for ep colliders on Fig. 5.6b. The comparison of the different characteristics of each of the collider described in this section is on Table 5.1, with the expected total HH cross-sections from these diagrams. The absence of gluon-initiated production in e^+e^- colliders is one reason for the cleaner environment, but also the absence of any GGF production mode that unfortunately impedes the total HH cross-section. Similarly, VBF, and more minor production mechanisms cross-sections depend on the amount of energy that can be injected in the center-of-mass, and hadron colliders are vastly superior in that sense.

Table 5.1 | Summary of the main expected characteristics of the future colliders described in this section as of 2019 (compiled from Ref. [443]). Center-of-mass energies can vary in the scans performed in e^+e^- colliders, similarly with the instantaneous luminosity. Integrated luminosity is expected at the end of the running planned years, given as a figure of fair comparison. HH cross-section includes all production modes and is computed at LO, variations in colliders including at least one lepton are due to the possible state of polarisation that have an effect on the cross-section.

Collider	Type		$\sqrt{s}$	$\mathcal{L}_{inst}$	$\mathcal{L}_{int}$	Time	σ_{HH}
	Content	Design		[10^{34} cm^{-2}s^{-1}]	[ab^{-1}]	[years]	[pb]
HL-LHC	pp	Circular	14 TeV	5	6	12	62
HE-LHC	pp	Circular	24 TeV	16	15	20	168
FCC-hh	pp	Circular	100 TeV	30	30	25	936
FCC-ee	e^+e^-	Circular	m_Z	100-200	150	4	0.199
	e^+e^-	Circular	$2m_W$	25	10	1-2	-
	e^+e^-	Circular	240 GeV	7	5	3	-
CEPC	e^+e^-	Circular	m_Z	17-32	2	2	0.199
	e^+e^-	Circular	$2m_W$	10	2.6	1	-
	e^+e^-	Circular	240 GeV	3	5.6	7	-
ILC	e^+e^-	Linear	250 GeV	1.35-2.7	2	11.5	0.313
	e^+e^-	Linear	500 GeV	1.8-3.6	4	8.5	0.262
CLIC	e^+e^-	Linear	380 GeV	1.5	1	8	0.112-0.209
	e^+e^-	Linear	1.5 TeV	3.7	2.5	7	0.01-0.574
	e^+e^-	Linear	3 TeV	5	6	8	0.138-0.921
LHeC	ep	Circular	1.3 TeV	0.8	1	15	0.130-0.221
FCC-he	ep	Circular	3.5 TeV	1.5	2	25	0.674-1.189

Apart from the cross-section, the trilinear coupling measurement is also of high interest and its impact on the cross-section can be different between colliders. For example, pp colliders are more constraining for negative values of κ_λ, while e^+e^- colliders can provide better constraints at positive values using the $\nu\bar{\nu}$HH channel. In terms of analysis methods, a e^+e^- collider is obviously cleaner and the background contaminations is much lower, so technically all decay modes would be accessible. The cross-section and coupling measurement would also be model independent by having access to the Higgs boson decay width from the ZH process. On the other hand, hadron colliders benefit from a higher energy and therefore higher cross-sections, with the caveats of larger backgrounds and more model dependent measurements since the Higgs boson width has to be assumed. The expected precision on the trilinear coupling depends highly on the method used. Colliders with $\sqrt{s} \gtrsim 400$ GeV can use direct HH measurement, while others have to rely on indirect loop effect in single Higgs processes, both measurements also depend on whether the deformations of the single Higgs are considered. This makes predictions



(a) (b)

Figure 5.6 | Left : leading contributions to the double Higgs production mechanisms available at an e^+e^- collider, with Higgs-strahlung VHH (top) and VBF $\nu\bar{\nu}$HH (bottom), from Ref. [443]. Right : leading contributions to the double Higgs production mechanisms available at an ep collider, almost entirely through VBF, from Ref. [444]. In both cases, in the VBF production mechanism the charged current is dominant compared to neutral current, and the trilinear Higgs coupling is highlighted in red.

harder to compare on the same basis, but it is becoming clear that the hadron collider FCC-hh will provide a more precise measurement of the trilinear coupling, down to 3.4-8.5 % precision at 68 % CL based on preliminary studies extrapolating Run-2 results [445, 446]. The FCC-ee does not have access to HH production and from single Higgs measurement alone can only account to at best around 20 % at highest energy, and up to 45 % [445, 447, 448] in the first runs, slightly higher if considering a higher dimension fit. Linear e^+e^- colliders report similar numbers in the single Higgs measurement, about 50 % and 40 % for CLIC [62, 449] and ILC [443], although they have access to HH production and could go as low as 10% in their highest energy runs. Hybrid collider such as FCC-he report around 20% precision from the HH process alone.

There is therefore a strong incentive to move directly to the FCC-hh at the end of life of the LHC. This is further supported by the recent results of the Run-2 analyses of LHC data, illustrating that many of the burdens of hadron colliders – pile-up, UE, triggering, etc – can be mitigated with enough effort and ingenuity, and that the next generation of analyses will likely outperform the predictions. This choice however only considers the scalar sector measurement with an emphasis on HH production and associated coupling, while the decision for a future collider also depends on other sectors of research as well as practicality, since e^+e^- colliders are more affordable than their pp counterparts. For these reasons, current plans regarding FCC envision a first step as an e^+e^- collider, followed by a potential ep intermediate stage and finally the pp collider. This would allow achieving down to a 5% or better precision on the trilinear coupling, as illustrated on Fig. 5.7, and benefit from the Higgs decay width measurement of the FCC-ee to perform model-independent searches with the later stage FCC-hh.

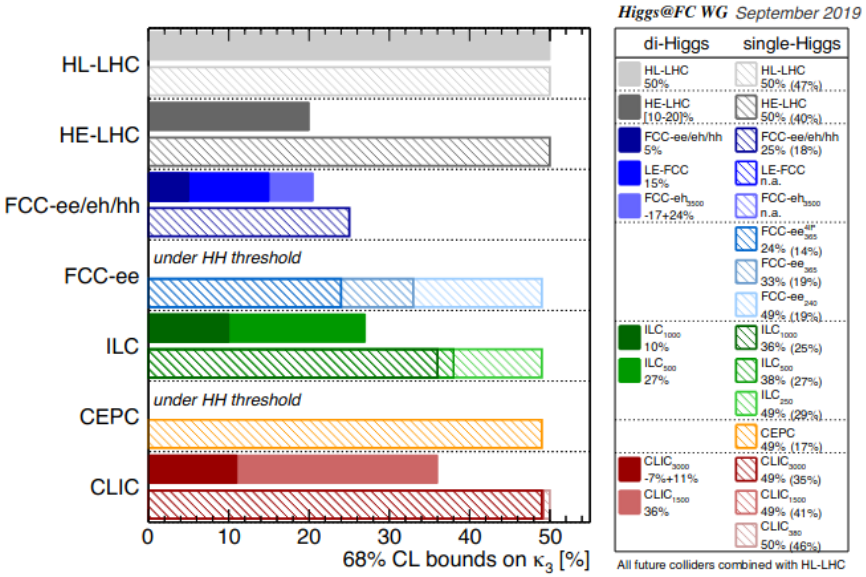


Figure 5.7 | Expected sensitivity on the trilinear coupling at 68 % CL for various future colliders using both double and single Higgs measurement, the former being only accessible to accelerators with $\sqrt{s} \gtrsim 400$ GeV. Double Higgs results are presented with solid lines, using pure HH measurements without taking into account single Higgs deformations, while single Higgs in dashed lines consider them in the bands and numbers, the values without are within parenthesis in the legend. Linear colliders are presented at their different level or energies, using results at the best beam polarisation state for that measurement. From Ref. [443].

5.3 Reinterpretability of results

Long gone is the time when experimental and theoretical physicists could work in their separate worlds. Since the Higgs boson discovery a decade ago, there has not been any convincing deviations from the SM that could indicate a preferred direction to solve the problems listed in Section 1.4. This is especially striking considering that the Run-2, with its impressive amount of collected data, was completed four years ago. Yet, despise the tremendous efforts carried out by the LHC experiments, we have come up short so far. This observation can cause both excitement, as this leaves many door opens for new theories to emerge, and an overwhelming feeling given the task of analysing the very high-dimensional space of the data. Given the wide range of BSM models, it is impossible for experimental physicists to produce searches for all of them, even for large collaborations such as CMS. An analysis can span years between inception and final results, and a back and forth between theory and experiments is obviously very inefficient.

The advent of EFTs has provided an alternative, as they serve as a very useful intermediate step to "parallelise" this process : several experimental analyses can set constraints on some parameters, while validation or rejection of BSM scenarios can start from these constraints. Therefore, there is an ever-growing necessity for a dialogue between experimentalists and phenomenologists. The former need to define a subset of operators that have a significant impact on the process they are considering, while the latter demand enough information from the analysed data to be able to interpret the different BSM scenarios at the UV scale. The question of the validity of EFTs their limitations was already addressed in Section 1.7.4, and highlighted the need for additional information from the experiments, notably the energy scale that the constraints are referring to. In the context of HH searches, considering the EFT results of Fig. 4.30, precision is still pretty poor and it is then too early to discuss higher-order coefficients or loop-corrections. This is partly true for the typical energy scale, but it would be good practice to initiate this procurement with the constraints. However, given the variety of decay channels and the different production modes (GGF and VBF in this analysis, though other channels might include more, such as VHH or $\mathrm{t\bar{t}}$HH), it might be non-trivial. Given the limited sensitivity, it is more likely that an interpretation of the EFT constraints is carried out over the combined channels, in which case the energy scale should be studied and provided at the level of the HH group.

There is another alternative to bridge the gap between theories and experiments. Instead of using an intermediate step, like in EFTs, by giving access to the experimental results and a means to recast an already existing analysis into any given model, a more efficient and direct exchange could start taking place. This is known as *reinterpretability*. The spectrum of information sharing is broad, with a direct relation between the amount of work required on the experimental side and the potential prospects for reinterpretability. The rest of this section will attempt to provide a brief, albeit non-exhaustive, summary of these methods [450].

The most common results and easier to share are upper limits or constraints on a specific model or parameters. In that case, for another scenario to be reinterpreted from the results, it must match the one used in their extraction to a certain degree. This is a major shortcoming, although some reinterpretation has been made possible in the context of EFT and HH production using a similar development to Eq. (1.68) [451], despite the extraction of some non-explicit results from the literature, notably some information about the HH analysis selection efficiencies. On the other side of the range, complete open access to both the data and the frameworks (for example RIVET [452] or MADANALYSIS [453]) used to derive the results, would prove the most versatile options for phenomenologists to recast the analyses. While this strategy has some practical benefits, it also makes sense that publicly funded experiments publicise their data. It however comes with a lot of work and time needed on both sides (due to the full chain of the framework

that needs to be handled), and the intensive computing cost, although surrogate models based on ML exist [454–456]. While experimental collaborations are not too reluctant to publish their data (for example the CMS open data initiative), experimentalists prefer to have the exclusivity on their own data for a given amount of time before releasing them (a somewhat valid opinion). Together with the additional burden of an open access framework, these are both massive showstoppers of this procedure. Alternatively, if enough details of the analysis selections are shared in the literature, reinterpretation can be carried out using approximations and fast parametric simulations (with DELPHES for example). In this case, auxiliary data should also be provided, such as the parameters of the full simulation event generation, the resolutions, efficiencies, and in general any corrections needed to calibrate the detector approximate simulations. A pure selection-based analysis can easily be replicated from scratch, but the more and more systematic use of multivariate methods in the analyses makes it more complicated. As is it impossible to reproduce a given trained model, it is the duty of the experimentalists to share and document them. Documentation is especially important given the possibly large number of input features, and the non-linear response. Any incorrectly defined variable could change the output drastically. The question of the framework and version can also make things cumbersome, although an alternative like ONNX [457] allows for interoperability.

There exists a variety of possibilities within these two extremes. The Simplified Template Cross Sections (STXS) [458] for example is an alternative to provide a centralised template for differential measurements, instead of only inclusive ones. Some analyses also perform searches over a parameterised signal, for example in the search of a resonance approximated by a Gaussian with a variable position and width. A step further resides in providing the full likelihood used to derive the results [459], which can be very complex for the layman. Alternatives such as the use of a simplified likelihood [460] where simplifying assumptions are applied and the background model is reduced to a covariance matrix such that any theoretical model can be plugged in, or ML surrogates for the likelihood [461].

For any of these methods to be successful, a stable and flexible enough platform must be used to share the information. Luckily, such platforms already exist : HEPDATA [462] that is specific for HEP, and ZENODO [463] that, though supported by CERN, is multidisciplinary. The current plan for the analysis developed in Chapter 4 is to provide the full likelihoods on HEPDATA, together with the limits and constrained they are attached to.

It is worth mentioning that in the context of reinterpretability, the method developed in Section 3.3 provides an additional unexpected benefit. Contrary to the classic method where the auxiliary files and code would need to be shared to allow recasting the analysis to another type of signal, the DNNs are self-contained. This means that the MEM weight can be obtained using the DNN inference without

setting up the whole integration chain. One downside would however be that the transfer function is encoded within the DNN, such that there is no way to alter the reconstruction efficiencies, although this may not necessarily a problem for a recast of the results. Typically, using these regressive DNNs is equivalent to sharing a MLs surrogate likelihood, albeit with a more direct physics-oriented interpretation. Another alternative was followed in Refs. [257–259, 464, 465], where a DNN is used to approximate the likelihood ratio directly, with an application to constrain EFTs with ML [466, 467]. By circumventing the need for expensive ME integration, they represent a strong concurrent to the method developed in this document. However, from the admission of its authors, this method has a more complicated convergence and must be tweaked correctly, while the hybrid MEM computation defended here is more robust and physics-oriented, therefore easier to understand and solve.

Going further, there is the concept of *reproducibility* and *reusability*, which are embedded in a program of conservation of the knowledge and data within CERN [468, 469]. The idea behind is not only the preservation of data and analysis chains, but also to allow new interpretations of the data already at hand. This could multiply the yield of scientific results with no need for a new input from the experiments. Several frameworks for such endeavour are already available : RECAST [470] for theorists to propose a new signal for an analysis to process; CHECKMATE [471] that performs a scan of existing analyses, and provides constraints for a given model; GAMBIT [472] with a similar goal but using both HEP and astrophysical data; SMODELS [473] for searches of new physics using simplified models; REANA [474] with the goal of containing an entire analysis workflow that can be run with a single command. None of the aforementioned frameworks and methods are currently or expected to be of use in HH searches. but they have reached a state of maturity that would be worth considering in further analyses.

Appendix

A Additional Material : bbWW results

A.1 SL channel DNN inputs

The following definitions are of interest to reconstruct motivated combinations of particles :

- *bjets* are the two DeepJet-leading AK4 jets
- *lightjets* the two p_T-leading AK4 jets not being the *bjets*
- *h-bb* is the 4-momenta sum of the two *bjets*
- *w-jj* is the 4-momenta sum of the *lightjets*
- *w-lep* is the 4-momenta sum of the lepton and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$
- *h-ww* is the 4-momenta sum of *w-jj* and *w-lep*
- *hh* is the 4-momenta sum *h-bb* and *h-ww*

The low-level features are :

- The leading lepton 4-momenta with PDG-ID $\in \{\pm 11, \pm 13\}$ and charge (both one-hot encoded, but only in non-resonant analysis).
- 6 AK4 jets 4-momenta (first two being the the two DeepJet-leading, last four being the p_{T}-leading which are not among the two DeepJet-leading), with DeepJet discriminator value in the non-resonant analysis.
- Leading AK8 jet 4-momentas, with: τ_1, τ_2, τ_3, τ_4, and the softdrop mass.
- $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ 4-momenta, where: invariant mass and p_z are set to 0.

The high-level features dependent on the specific search.

A.1.1 Non-resonant analysis

The high-level features are :

- invariant mass of *hh*, *h-bb*, *h-ww*, *w-jj* and *w-lep*
- transverse momentum of *hh*, *h-bb*, *h-ww*, *w-jj* and *w-lep*

- $|\Delta\phi|$ between *h-bb* and *h-ww*
- $|\Delta\phi|$ between *h-bb* and the visual part (without $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$) of *h-ww*
- $|\Delta\phi|$ between $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ and the lepton, *h-bb* and *w-jj*
- ΔR between the lepton and *h-bb*
- ΔR between the lepton and *w-jj*
- minimum ΔR between the lepton and a bjet
- ΔR between the two *light-jets*
- $|\Delta\phi|$ between the two *light-jets*
- ΔR between the two *bjets*
- $|\Delta\phi|$ between the two *bjets*
- cone-p_{T} of the lepton
- H_{T} scalar sum of the transverse momentum of all AK4-Jets
- S_{min} as discussed in Section 4.2.3 and defined in Eq. (4.11)
- VBF-tag: existence of a pair of VBF-jets with invariant mass $>$ 500 GeV and $|\Delta\eta| > 3$
- invariant mass of the AK4-jets fulfilling the VBF-tag
- $|\Delta\eta|$ of the AK4-jets fulfilling the VBF-tag
- boosted-tag: existence of a AK8-jet with ≥ 1 btagged subjet
- n-btag: number of btagged AK4 jets
- the Run year $\in [2016, 2017, 2018]$ (one-hot encoded)

A.1.2 Resonant analysis

The high-level features are :

- number of selected AK4-jets
- DeepJet score (DeepCSV score) of selected *bjets* (subjets of AK8-jet in boosted category)
- minimum ΔR between selected lepton and selected AK4-jets
- p_{T} of *h-bb*
- invariant mass of *h-bb*, with bjet energy regression applied
- ΔR between the two *bjets*
- $m_T^W = \sqrt{m_1^2 + m_2^2 + 2(E_{T,1}E_{T,2} - \vec{p_{T,1}}\vec{p_{T,2}})}$ transverse mass of the W decaying leptonically, between the lepton and neutrino, reducing to $\sqrt{p_{\mathrm{T}}^l E_{\mathrm{T}}^{\mathrm{miss}}(1 - cos\Delta\phi)}$
- HT p_{T} scalar sum of AK4-jets with $p_T > 50$

- type of first pair of loose lepton, ordered in cone-p_T
- total charge of first two loose lepton, ordered in cone-p_T
- p_{T} of selected lepton
- p_{T} of *hh* system
- invariant mass of first pair of loose lepton, ordered in cone-p_T
- average ΔR between selected AK4-jets
- $E_{\mathrm{T}}^{\mathrm{miss}}$
- DeepJet score of selected w-jets
- number of AK4 jets satisfying loose DeepJet score
- number of AK4 jets satisfying medium DeepJet score
- invariant mass of two leading AK4 jets, passing medium DeepJet score, ordered in p_T
- ΔR between leading bjet and selected lepton
- ΔR between sub-leading bjet and selected lepton
- $m_T^{\mathrm{t,2\text{-}particles}}$ transverse mass of a top from a 2-particles system made up of a bjet – testing both bjets and selecting the lowest transverse mass – and a composite W boson summing the lepton and the $E_{\mathrm{T}}^{\mathrm{miss}}$, assuming W boson mass
- $m_T^{\mathrm{t,3\text{-}particles}}$ similar to $m_T^{\mathrm{t,2\text{-}particles}}$ except the leptons and $E_{\mathrm{T}}^{\mathrm{miss}}$ and considered separate
- invariant mass of leading and sub-leading p_{T} AK4 jets (subjets of leading p_T AK8 jet in boosted category)
- number of AK4 jets with $|\eta| > 2.4$
- p_{T} of *hh-vis* system
- $E_T^{miss}LD$, as discussed in Section 4.2.3 and defined in Eq. (4.3)
- $\Delta\phi$ between two reconstructed Higgs bosons in the *hhvis* system
- type (electron or muon) of selected lepton
- τ_1 of leading AK8 jet
- invariant mass of *hh* system where z component of neutrino is reconstructed solving the analytic formula where the mass of the system consisting of *w-jets*, lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$ is constrained to Higgs boson mass

A.2 DL channel DNN inputs

The following definitions are of interest to reconstruct motivated combinations of particles :

- *bjets* are the two DeepJet-leading AK4-jets
- *di-bjet* is the 4-momentum sum of the two DeepJet-leading AK4-jets
- *di-jet* is the 4-momentum sum of the two p_{T}-leading AK4-jets
- *di-lepton* is the 4-momentum sum of the two p_{T}-leading leptons

The low-level features are :

- 2 leading leptons 4-momenta, with: PDG-ID $\in \{\pm 11, \pm 13\}$ and charge (both one-hot encoded).
- 4 p_{T}-leading AK4 jets 4-momenta
- Leading AK8 jet 4-momenta with: τ_1, τ_2, τ_3, τ_4, and the softdrop mass.
- $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ 4-momenta, where: invariant mass and p_z are set to 0.

The high-level features dependent on the specific search.

A.2.1 Non-resonant analysis

The high-level features are :

- ΔR between the leading and subleading lepton
- ΔR between the two *bjets*
- ΔR between *di-lepton* and *di-jet*
- ΔR between *di-lepton* and *di-bjet*
- $|\Delta\phi|$ between $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ and *di-lepton*
- $|\Delta\phi|$ between $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ and *di-bjet*
- minimum ΔR between the leading lepton and all AK4 jets
- minimum ΔR between the subleading lepton and all AK4 jets
- minimum ΔR between the leading *bjets* and all leading two leptons
- minimum ΔR between the subleading *bjets* and all leading two leptons
- minimum ΔR among all AK4 jets
- minimum $|\Delta\phi|$ among all AK4 jets
- *di-bjet* invariant mass, with bjet energy regression applied
- *di-lepton* invariant mass
- invariant mass of the 4-momentum sum of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ and *di-lepton*

- invariant mass of the 4-momentum sum of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, *di-lepton*, and *di-jets*, with bjet energy regression applied
- $E_T^{miss}LD$, see Section 4.2.3 and Eq. (4.3)
- H_{T} scalar sum of the transverse momentum of all AK4-Jets
- cone-p_{T} of the leading lepton
- cone-p_{T} of the subleading lepton
- VBF-tag: existence of a pair of VBF-jets with invariant mass $>$ 500 GeV and $|\Delta\eta| > 3$
- boosted-tag: existence of a AK8 jet wit ≥ 1 btagged subjet
- the Run year $\in [2016, 2017, 2018]$ (one-hot encoded)

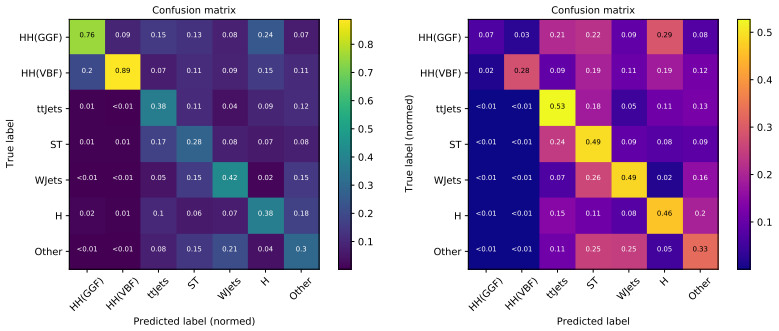
A.2.2 Resonant analysis

The high-level features are :

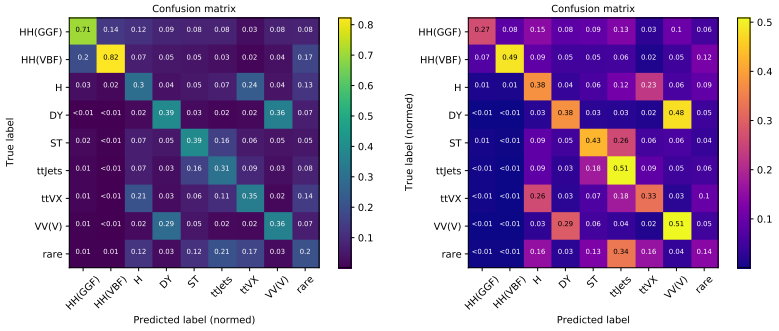
- ΔR between the leading and subleading lepton
- ΔR between the two *bjets*
- ΔR between *di-lepton* and *di-jet*
- ΔR between *di-lepton* and *di-bjet*
- $|\Delta\phi|$ between the leading and subleading lepton
- $|\Delta\phi|$ between the two *bjets*
- $|\Delta\phi|$ between $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ and *di-lepton*
- $|\Delta\phi|$ between $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ and *di-bjet*
- minimum ΔR between the leading lepton and all AK4 jets
- minimum ΔR between the subleading lepton and all AK4 jets
- minimum ΔR between the leading *bjets* and all leading two leptons
- minimum ΔR between the subleading *bjets* and all leading two leptons
- minimum ΔR among all AK4-jets
- minimum $|\Delta\phi|$ among all AK4 jets
- *di-bjet* invariant mass
- *di-bjet* invariant mass, with bjet energy regression applied
- *di-lepton* invariant mass
- invariant mass of the 4-vector sum of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ and *di-lepton*
- invariant mass of the 4-vector sum of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, *di-lepton*, and *di-jets*
- invariant mass of the 4-vector sum of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, *di-lepton*, and *di-jets*, with bjet energy regression applied

- $E_T^{miss}LD$, as discussed in Section 4.2.3 and defined in Eq. (4.3)
- H_{T} scalar sum of the transverse momentum of all AK4-Jets
- cone-p_{T} of the leading lepton
- cone-p_{T} of the subleading lepton
- the number of btagged AK4 jets
- boosted-tag: existence of a AK8 jet wit ≥ 1 btagged subjet
- M_T^{ll} : transverse mass of the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ and *di-lepton* system
- $\cos^*(\theta_{bb})$: helicity angle between the two *bjets* in their center of mass frame (i.e. the Higgs boson rest frame)
- the Run year $\in [2016, 2017, 2018]$ (one-hot encoded)

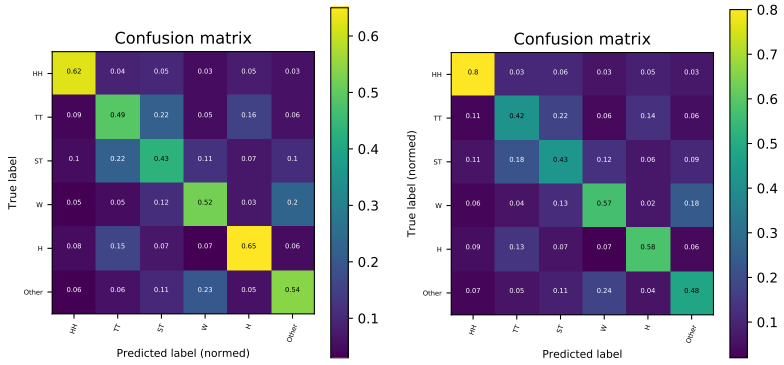
A.3 Confusion matrices



(a) Non-resonant SL



(b) Non-resonant DL



(c) Resonant SL

Figure A.1 | Confusion matrices of the non-resonant SL channel (top), non-resonant DL channel (middle), and resonant SL (bottom), normed over predicted label (left) and true label (right).

A.4 Heavy mass estimator

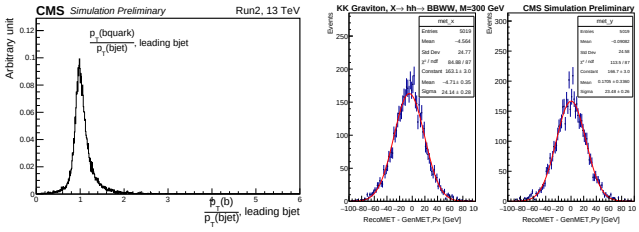


Figure A.2 | Left : transverse momentum ratio of b-quark and reconstructed leading jet derived from a combination of a mix of signal samples, used as a pdf in the HME computations to sample the correct jet energy, the subleading jet energy being fixed to the Higgs mass peak at $m_{\mathrm{H}} = 125$ GeV. Right : MET reconstruction between generator- and reconstruction-level using a specific signal sample (although resolution is consistent between different mass points), used as a pdf in the HME computations to smear the MET.

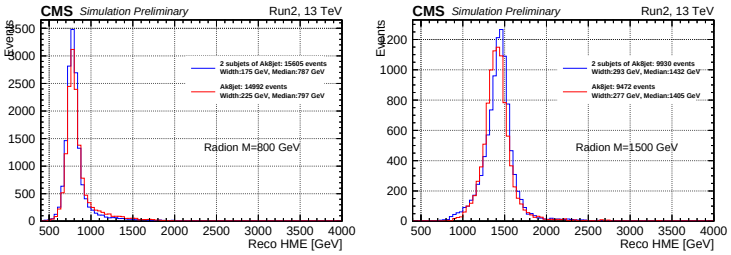


Figure A.3 | HME distributions of two signal samples at different mass points, using either the AK8 jet 4-momentum, or its two subjets, favouring the latter case that minimises the reconstruction bias

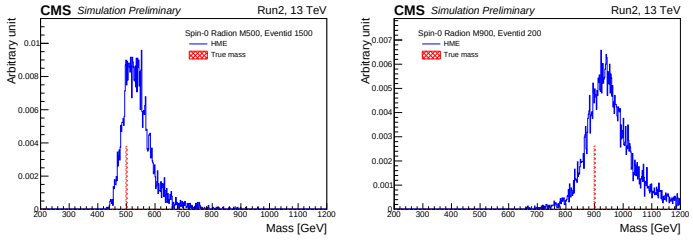


Figure A.4 | Results of the HME computations for single events from different mass points : several iterations of the algorithm are used to yield the distribution from which the HME value is evaluated from the most probable value, shown with the theoretical resonant mass.

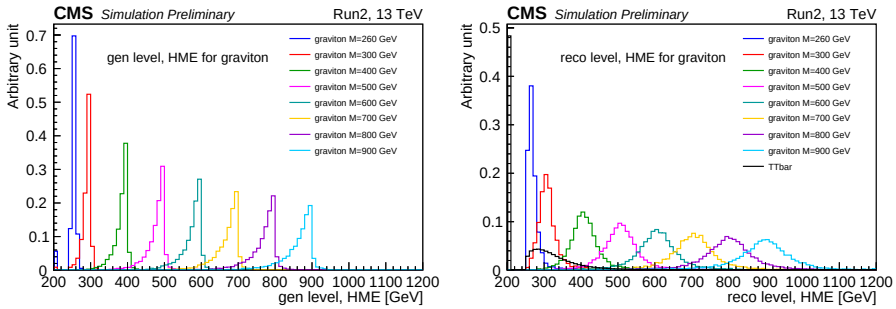


Figure A.5 | HME distribution of the several KK spin-2 graviton signal peaks, using either generator-level (left) or reconstruction-level (right) inputs.

A.5 Resonant results

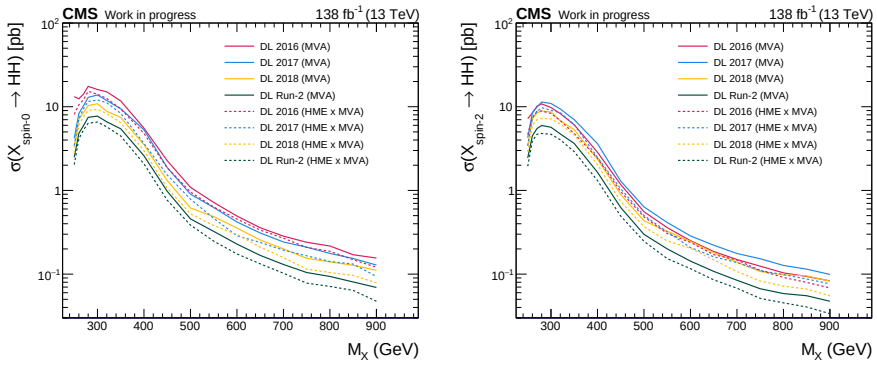


Figure A.6 | Observed upper limits as a function of m_{HH} for the spin-0 (left) and spin-2 (right) resonances, when the distribution of the parametric DNN is used alone in the likelihood fit ("MVA", solid line) or when the HME distribution is included in a two-dimensional distribution ("HME x MVA", dotted line).

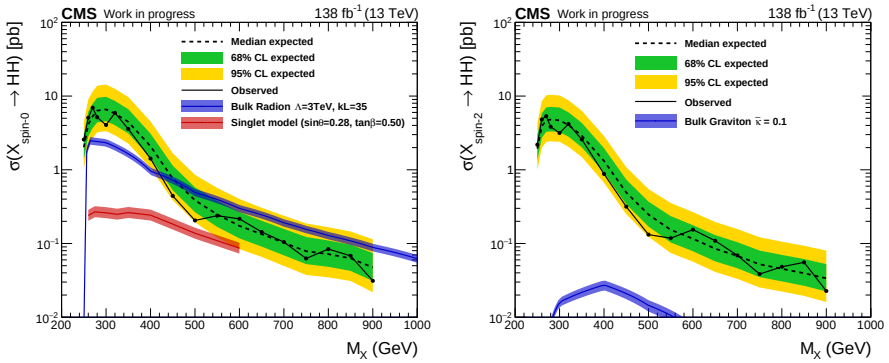


Figure A.7 | Observed and expected upper limits as a function of m_{HH} for the spin-0 (left) and spin-2 (right) resonances in the DL channel.

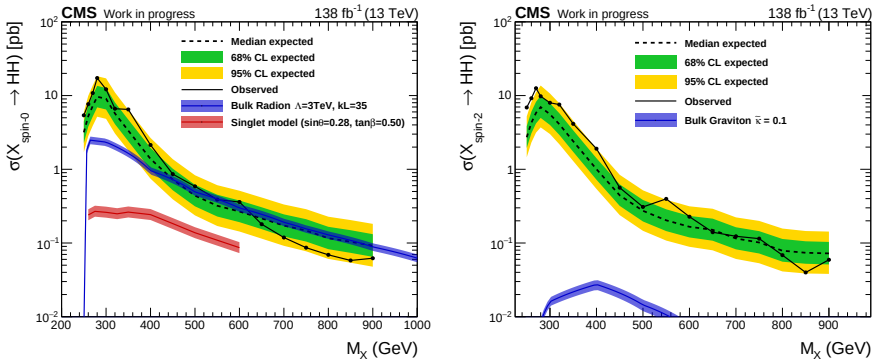


Figure A.8 | Observed and expected upper limits as a function of m_{HH} for the spin-0 (left) and spin-2 (right) resonances in the SL channel.

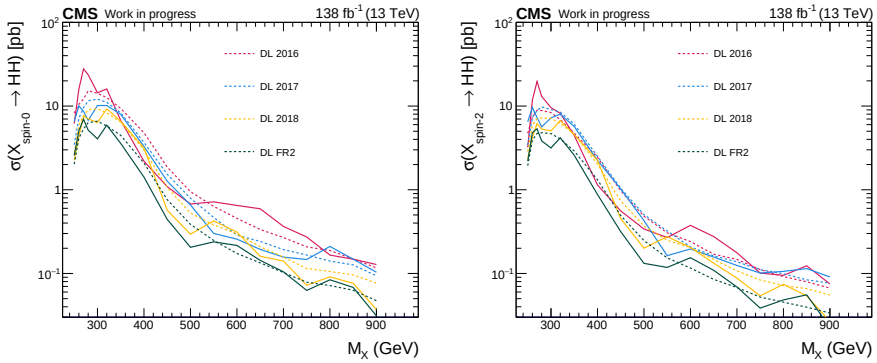


Figure A.9 | Observed (solid line) and expected (dotted line) upper limits as a function of m_{HH} for the spin-0 (left) and spin-2 (right) resonances in the DL channel, split per year and combined over the full Run-2 dataset (FR2).

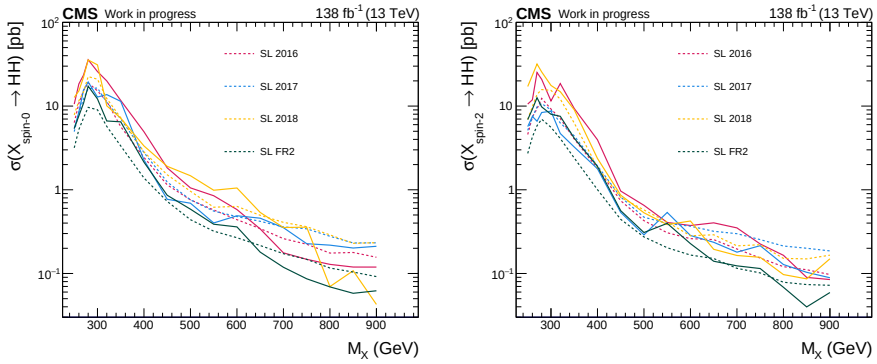


Figure A.10 | Observed (solid line) and expected (dotted line) upper limits as a function of m_{HH} for the spin-0 (left) and spin-2 (right) resonances in the SL channel, split per year and combined over the full Run-2 dataset (FR2).

A.6 EFT reweighting

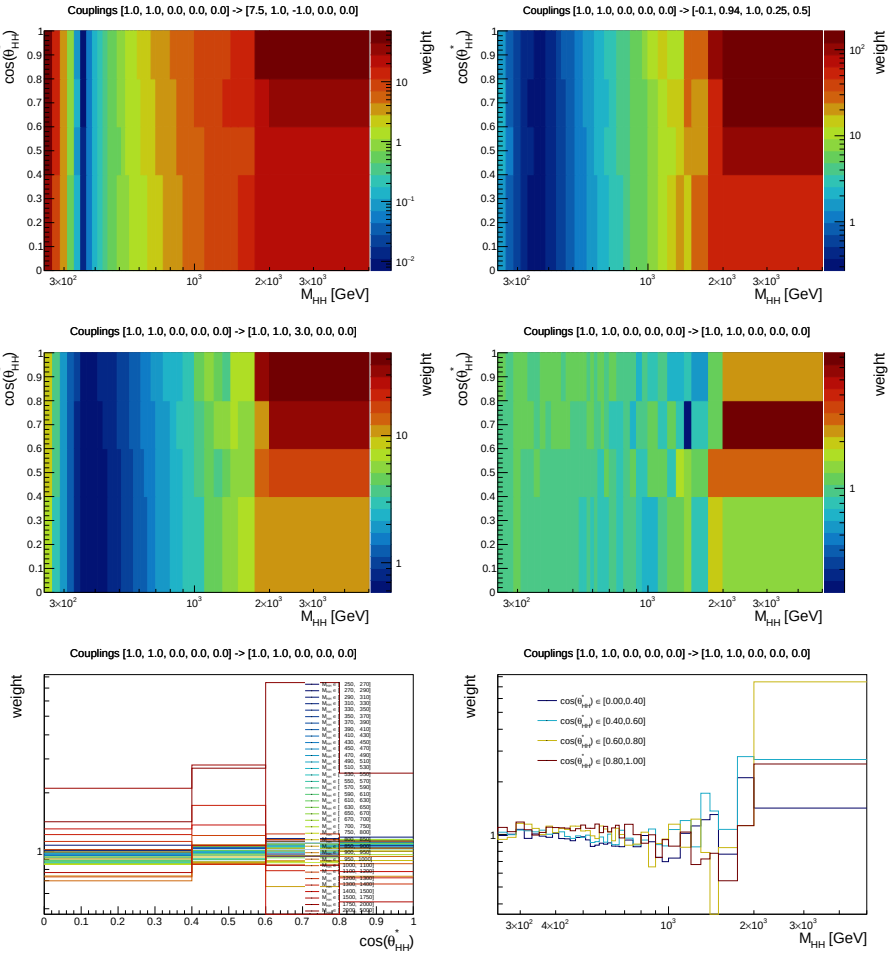


Figure A.11 | Weights used in the EFT reweighting, in the (κ_λ,κ_t,c_2,c_g,c_{2g}) coupling plane. The initial benchmark is the SM point, with as final benchmarks the number 1 of Table 1.3 (top left), number 7 from Table 1.4 (top right), a modified SM point with a non-zero c_2 value used in the c_2 scan interpolation (middle left), and the SM point itself (middle right). Due to the increasing bin width in the m_{HH} axis, used to mitigate the statistical uncertainties in these low-populated bins, a log scale is set on the axis. Most of the weights dependency is on that variable, whereas the effect of $\cos\theta^*_{\text{HH}}$ is minimal. When the SM point is reweighted to itself, in the ideal case all bins should have unit value, and should be also observed in the one-dimensional projections on both axes (bottom). The variations are likely due to statistical variations from the limited amount of generated events used in their determination.

A.7 Non-resonant results

A.7.1 EFT benchmark limits

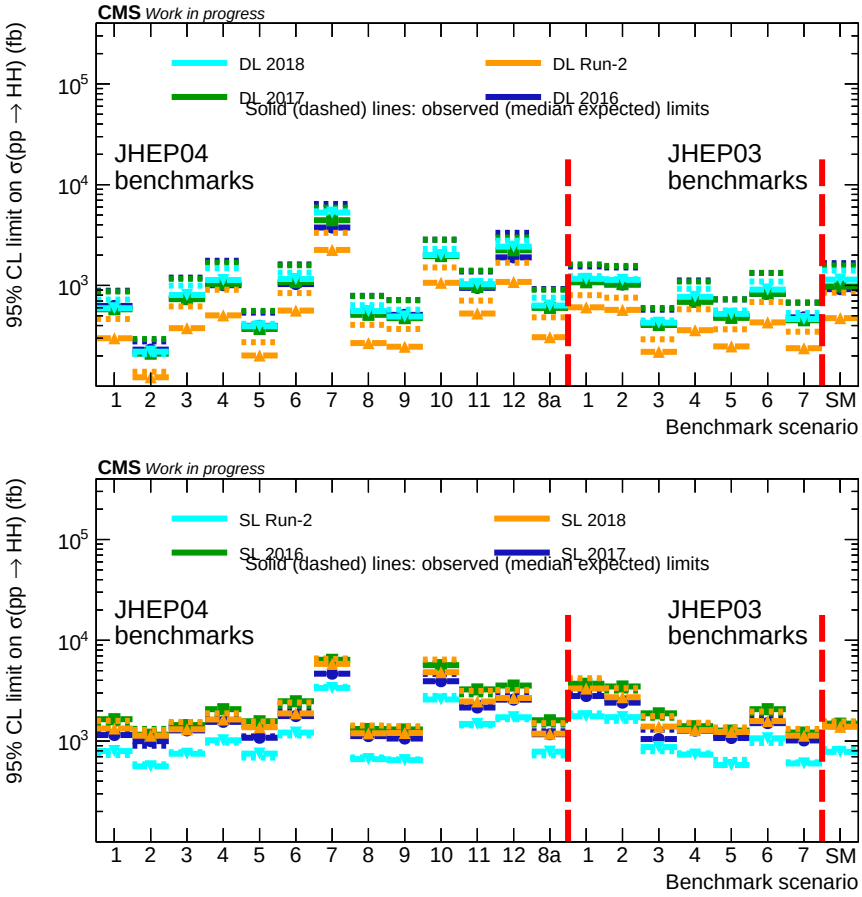
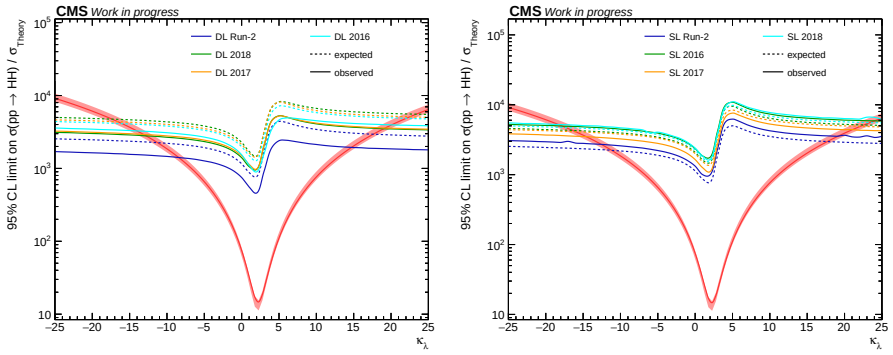
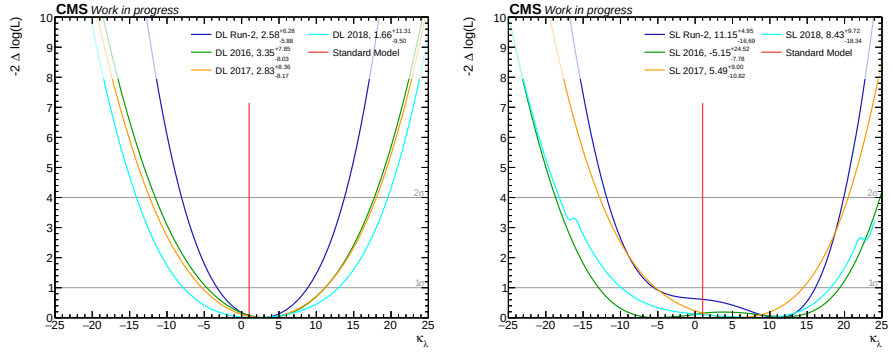


Figure A.12 | Observed (solid line) and 95% CL expected (dotted line) limits on the cross-section expressed in fb for the different benchmarks points defined in Table 1.3 (JHEP04) and Table 1.4 (JHEP3) (right), split between data-taking years and combined Run-2 results in the DL (top) and SL (bottom) channels.

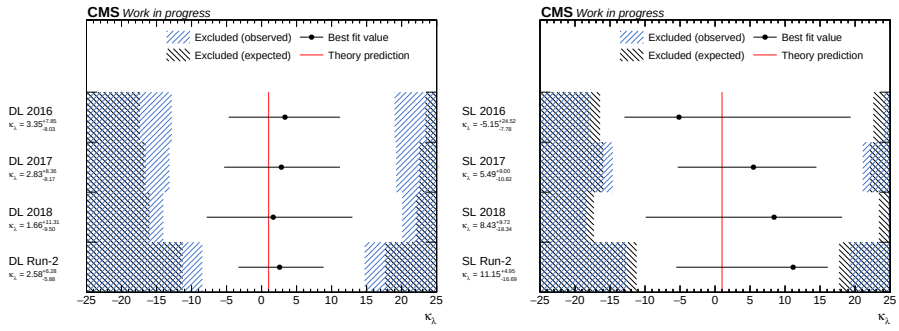
A.7.2 Coupling scans and exclusion regions



(a) Upper limit scan

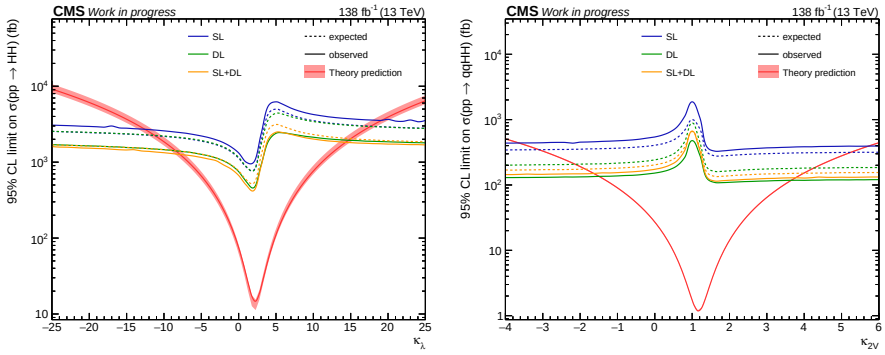


(b) Likelihood scan

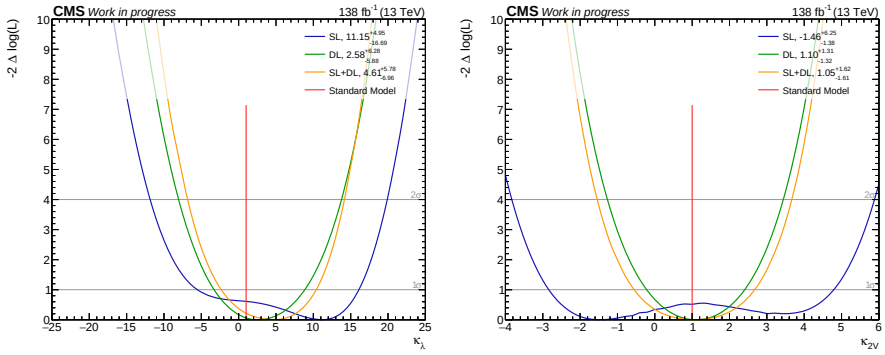


(c) Best fit and exclusion regions

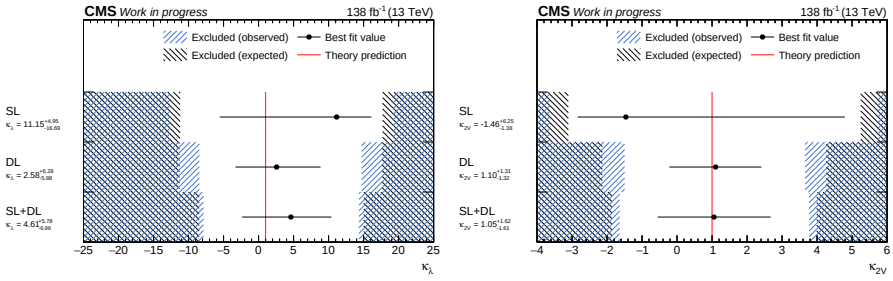
Figure A.13 | Results of the coupling κ_λ split per year and combined over Run-2 for both DL (left) and SL (right) : upper limit scans (top), likelihood scans (middle), and best fit and exclusion regions obtained from the above scans.



(a) Upper limit scans

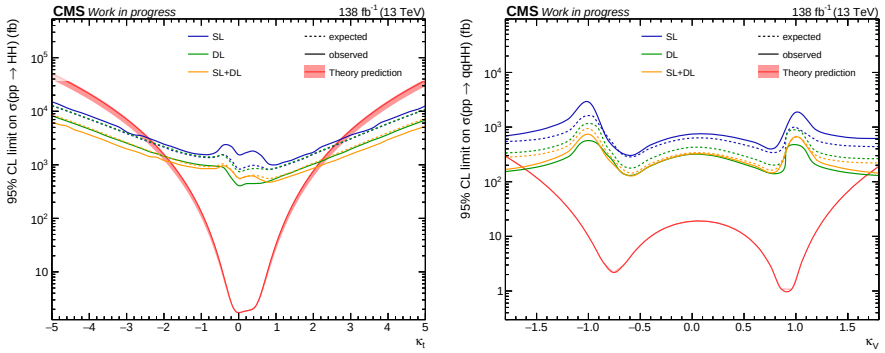


(b) Likelihood scans

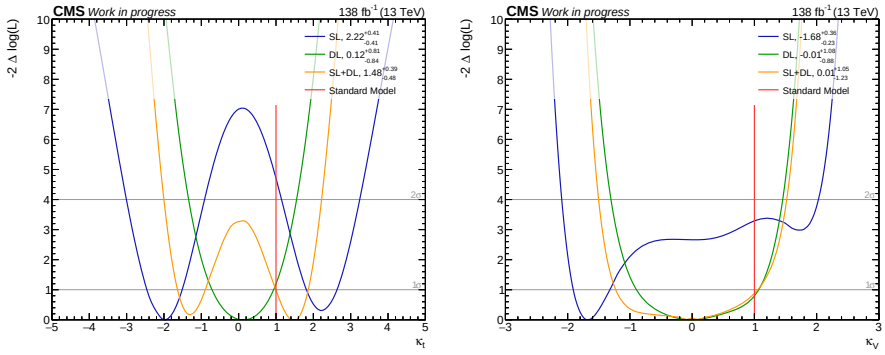


(c) Best fits and exclusion regions

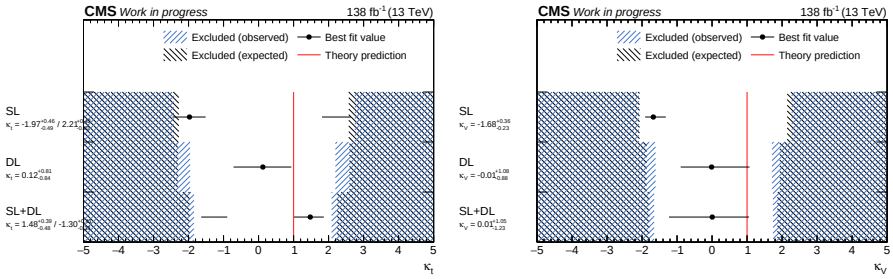
Figure A.14 | Results of the couplings directly measured through HH measurements κ_λ (left) and κ_{2V} (right) split per channel and both combined : upper limit scans (top), likelihood scans (middle), and best fit and exclusion regions obtained from the above scans.



(a) Upper limit scans

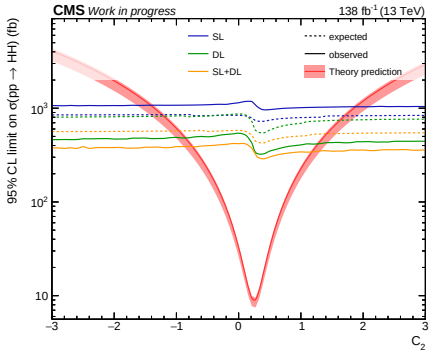


(b) Likelihood scans

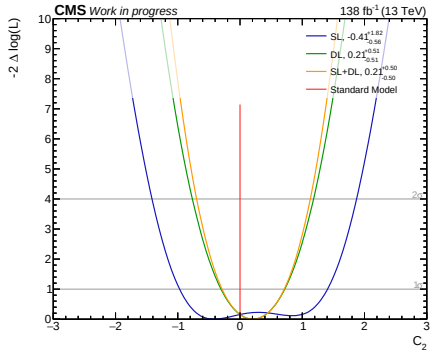


(c) Best fits and exclusion regions

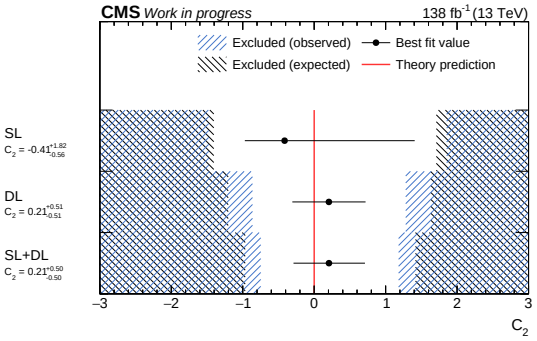
Figure A.15 | Results of the single Higgs couplings κ_t (left) and κ_V (right) split per channel and both combined : upper limit scans (top), likelihood scans (middle), and best fit and exclusion regions obtained from the above scans.



(a) Upper limit scan



(b) Likelihood scan



(c) Best fits and exclusion regions

Figure A.16 | Results of the BSM coupling c_2 split per channel and both combined : upper limit scans (top), likelihood scans (middle), and best fit and exclusion regions obtained from the above scans.

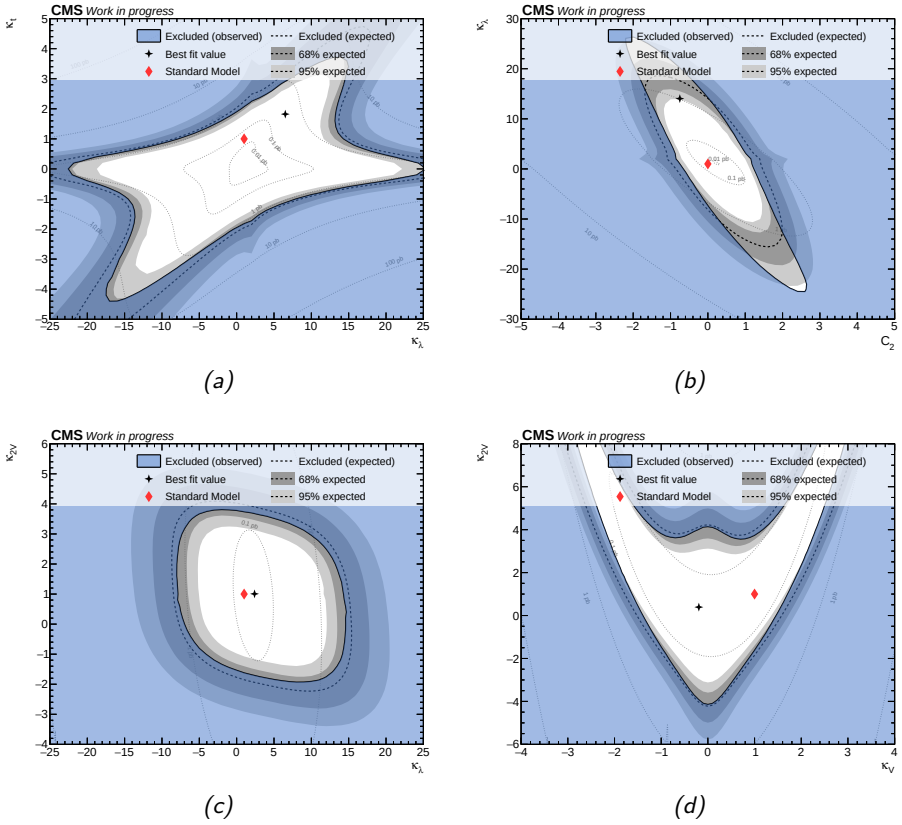


Figure A.17 | Observed and expected exclusion regions as a function of the different couplings κ_λ, κ_t, κ_V, κ_{2V} and c_2 in two-dimensional spaces.

A.8 Prefit nuisance yields

The nuisances groups in Table A.1 are grouped as follows :

- *Jet + MET* : JES, JER and unclustered MET energy sources combined
- *Experimental* : L1 pre-firing, pile-up, luminosity, HEM
- *Electron* and *Muon* identification uncertainties
- *Higgs BR* uncertainties
- *Trigger* selection uncertainties
- *B-tagging* uncertainties combined between the different sources
- *Theory* uncertainties :
  - Minor backgrounds : QCD and EW corrections, PDF and scale uncertainties
  - $\mathrm{t\bar{t}}$: from generation (colour reconnection, h_{damp}, UE), p_{T} reweighting, QCD scale, parton shower ISR and FSR, scales and mass uncertainties
  - HH(GGF) and HH(VBF) : QCD corrections, scales, parton shower ISR and FSR, VBF dipole, and top mass uncertainties. In addition, κ_λ-dependent contributions to QCD scale and top mass uncertainties are included.
- Uncertainties related to data-driven estimations :
  - *DY weights*
  - *Fake-rates*

  And their non-closure uncertainties
- Ad-hoc norm : log-normal uncertainties on main backgrounds to allow them to vary during the likelihood fit, combined over the whole set of backgrounds, described in Section 4.4.

Table A.1 | Effect of the up and down variations of each group of nuisances on the nominal yield. The variations are provided as the maximum for a set of categories (listed in Table 4.10) : the background (inclusive DY+VV, resolved and boosted other), GGF (resolved 1b, 2b and boosted) and VBF (resolved 1b, 2b and boosted) categories. For each of these combined categories the variations are given for the background processes (considered correlated among processes, the different nuisances within a group are combined as uncorrelated), GGF and VBF samples (only SM with $\kappa_\lambda = 1$). The variations are given as relative (in %) to the nominal yield, when a single number is provided the up and down values are symmetric, and when the nuisance is related to a specific process the number refer to that process and not the sum of backgrounds.

Nuisance group	Background categories			GGF categories			VBF categories		
	Background	GGF (SM)	VBF (SM)	Background	GGF (SM)	VBF (SM)	Background	GGF (SM)	VBF (SM)
Nuisances applied to all samples									
Jet + MET	+3.19/-3.10	+6.42/-6.94	+7.75/-6.97	+9.07/-9.33	+4.78/-3.55	+15.14/-15.96	+16.78/-15.67	+12.49/-11.58	+7.19/-6.74
Experimental	+3.61/-3.65	+3.88/-4.06	+4.76/-5.00	+4.29/-4.34	+4.08/-4.11	+4.77/-5.16	+4.33/-4.63	+4.20/-4.43	+5.29/-5.43
Electron	+7.64/-7.54	+7.56/-7.44	+7.99/-7.83	+7.68/-7.58	+7.22/-7.12	+8.18/-8.03	+7.75/-7.65	+7.41/-7.31	+7.86/-7.74
Muon	+3.98/-3.94	+4.65/-4.58	+4.97/-4.89	+3.83/-3.80	+4.27/-4.22	+4.22/-4.18	+3.89/-3.85	+4.66/-4.60	+4.60/-4.54
PU jet	+1.21/-1.13	+1.70/-1.78	+1.54/-1.47	+0.85/-0.84	+0.82/-0.82	+1.23/-1.22	+1.63/-1.55	+1.61/-1.56	+1.16/-1.18
Higgs BR	+0.01/-0.01	+2.49/-2.49	+2.49/-2.49	+0.03/-0.03	+2.49/-2.49	+2.49/-2.49	+0.05/-0.05	+2.49/-2.49	+2.49/-2.49
Trigger	+1.37/-1.37	+1.84/-1.84	+1.79/-1.79	+1.25/-1.25	+1.34/-1.34	+1.35/-1.35	+1.25/-1.25	+1.44/-1.44	+1.34/-1.34
B-tagging	+14.69/-15.99	+10.17/-11.98	+9.64/-12.00	+14.47/-15.57	+12.83/-14.27	+13.20/-14.67	+14.73/-15.88	+11.10/-12.71	+11.30/-12.96
Theory nuisances									
Minor backgrounds	+3.89/-3.90			+3.84/-3.85		+3.77/-3.76			
$t\bar{t}$	+9.60/-9.47			+21.43/-40.32		+25.59/-33.82			
Signal (constant)		+5.54/-11.52	+5.55/-5.92		+4.57/-9.07	+8.64/-18.05		+5.90/-8.29	+8.29/-9.89
Signal (κ_λ)		+5.51/-21.82	+5.51/-21.82		+5.51/-21.82	+5.51/-21.82		+5.51/-21.82	+5.51/-21.82
DY data-driven nuisances									
ABCD weights	9.77			11.28			13.19		
Non-closure	9.75			100.10			72.75		
Fake nuisances									
Fake-rates	+11.31/-9.74			+27.33/-24.70			+21.41/-18.89		
Non-closure	+63.06/-36.74			+75.16/-43.82			+65.48/-39.21		
Fit related nuisances									
Ad-hoc norm	+24.71/-24.71			+21.72/-21.72			+20.03/-20.03		

Table A.2 | Effect of the up and down variations of each group of nuisances on the nominal yield. The variations are provided as the maximum for a set of categories (listed in Table 4.10) : the background (inclusive DY+VV, resolved and boosted other), GGF (resolved 1b, 2b and boosted) and VBF (resolved 1b, 2b and boosted) categories. For each of these combined categories the variations are given for the background processes (considered correlated among processes, the different nuisances within a group are combined as uncorrelated), GGF and VBF samples (only SM with $\kappa_\lambda = 1$). The variations are given as relative (in %) to the nominal yield, when a single number is provided the up and down values are symmetric, and when the nuisance is related to a specific process the number refer to that process and not the sum of backgrounds.

Nuisance group	Background categories			GGF categories			VBF categories		
	Background	GGF (SM)	VBF (SM)	Background	GGF (SM)	VBF (SM)	Background	GGF (SM)	VBF (SM)
Nuisances applied to all samples									
Jet + MET	+7.82/-7.61	+5.28/-4.99	+4.63/-4.15	+9.86/-10.11	+4.87/-5.67	+15.30/-14.84	+19.01/-16.27	+14.03/-12.76	+8.24/-8.60
Experimental	+4.08/-4.04	+3.82/-3.86	+4.62/-4.68	+4.53/-4.63	+3.64/-3.68	+4.63/-5.73	+4.47/-5.03	+4.29/-4.40	+5.19/-5.40
Electron	+3.28/-3.28	+3.35/-3.35	+3.28/-3.28	+3.05/-3.05	+2.82/-2.81	+2.89/-2.89	+3.30/-3.30	+3.03/-3.03	+3.08/-3.08
Muon	+2.16/-2.16	+2.25/-2.25	+2.25/-2.25	+1.87/-1.87	+2.06/-2.06	+2.30/-2.30	+2.20/-2.20	+2.22/-2.22	+2.28/-2.28
PU jet	+4.88/-4.18	+2.00/-1.93	+1.74/-1.67	+2.17/-2.17	+0.98/-1.02	+2.47/-2.44	+3.16/-2.80	+1.49/-1.44	+1.50/-1.47
Higgs BR	+0.00/-0.00	+2.49/-2.49	+2.49/-2.49	+0.01/-0.01	+2.49/-2.49	+2.49/-2.49	+0.01/-0.01	+2.49/-2.49	+2.49/-2.49
Trigger	+3.44/-3.44	+3.40/-3.40	+3.16/-3.16	+2.91/-2.91	+3.13/-3.13	+3.19/-3.19	+3.45/-3.45	+3.43/-3.43	+3.20/-3.20
B-tagging	+15.45/-18.02	+11.60/-13.49	+12.42/-15.27	+16.03/-18.45	+13.84/-15.29	+14.81/-16.15	+15.53/-18.31	+12.41/-13.89	+13.30/-14.76
Theory nuisances									
Minor backgrounds	+4.64/-5.02			+3.65/-3.87		+3.09/-3.14			
$t\bar{t}$	+7.58/-10.06			+8.20/-11.70		+10.91/-13.52			
Signal (constant)		+4.06/-5.17	+4.87/-6.33		+4.46/-8.57	+4.87/-7.54		+6.53/-12.11	+8.51/-11.42
Signal (κ_λ)		+5.51/-21.82	+5.51/-21.82		+5.51/-21.82	+5.51/-21.82		+5.51/-21.82	+5.51/-21.82
Fake nuisances									
Fake-rates	+7.82/-6.91			+6.31/-5.65			+6.54/-5.85		
Non-closure	+62.18/-62.18			+163.61/-163.61			+157.38/-157.38		
Fit related nuisances									
Ad-hoc norm	+28.14/-28.14			+23.60/-23.60			+22.97/-22.97		

Acronyms

μGMT Global Muon Trigger 80

μGT Global Trigger 80

ALOHA Automatic Libraries Of Helicity Amplitudes 25

AdaBoost Decision trees boosting with iterative weak learners 124

AdaDelta Gradient Descent (GD) with second order momentum and RMS of the gradient 120

AdaGrad Gradient Descent (GD) with second order momentum and gradient sum 120

Adam Gradient Descent (GD) with first and second order momentum 120, 196

Bamboo A high-level HEP analysis library for ROOT::RDataFrame 230

CheckMATE Check Models At Terascale Energies 244

Combine Command line interface to many different statistical techniques available inside RooFit and RooStats used widely inside CMS 111, 230

Cuhre Integration by a Deterministic Iterative Adaptive Algorithm 153

Delphes Framework performing a fast multipurpose detector response simulation 100, 243

Divonne Integration by Stratified Sampling for Variance Reduction 152

FastJet Package that provides a broad range of jet finding and analysis tools 90

FeynRules Mathematica-based package which addresses the implementation of particle physics models 25

GAMBIT Global and Modular Beyond-Standard Model Inference Tool 244

Geant4 GEometry ANd Tracking 100, 150

HELAS HELicity Amplitude Subroutines 25

HEPData Repository for publication-related High-Energy Physics data 243

Herwig General-purpose Monte Carlo event generator for the simulation of hard lepton-lepton, lepton-hadron and hadron-hadron collisions 27, 28

HistFactory Tool to build parametrized probability density functions in RooFit/RooStats 111

Keras Python interface for artificial neural networks 135

LHAPDF Standard tool for evaluating PDFs in HEP 150

MadAnalysis User-friendly framework for collider phenomenology 242

Madspin Spin-correlation and finite width extension to `MG5_aMC@NLO` 30

Minuit Numerical minimization software library 111

MoMEMta Modular toolkit for the Matrix Element Method at the LHC 132, 133, 135, 138–146, 149–153, 163, 165, 166

ONNX Open Neural Network Exchange 243

Powheg Positive Weight Hardest Emission Generator 29, 182, 218

PyHF Pure-Python implementation of HistFactory statistical models 111

Pythia Standard tool for the generation of events in high-energy collisions 27, 28, 100, 150, 182, 184, 207, 218

REANA Reproducible research data analysis platform 244

RMSProp Gradient Descent (GD) with second order momentum and exponential average 120

Recast Framework for extending the impact of existing analyses 244

ResNet Residual Neural Network 166, 196

Rivet Robust Independent Validation of Experiment and Theory 242

RooFit Data modelling library 111, 267

RooStats Software tools for statistical methods commonly used in HEP 111, 267

SModelS A tool for interpreting simplified-model results from the LHC 244

Sarah Mathematica package for building and analyzing supersymmetric models 25

Sherpa General-purpose tool for the simulation of particle collisions at high-energy colliders 29

Suave Subregion-adaptive VEGAS 152

Tensorflow Interface for expressing machine learning algorithms 135

Vegas Adaptive multidimensional Monte Carlo integration tool 25, 131, 152, 153, 160, 161, 268

Whizard Multi-purpose event generator for collider physics 29

XGBoost Decision trees boosting with Gradient Descent (GD) 124, 125

Zenodo Multidisciplinary open repository for research papers, data sets, research software, reports, and any other research related digital artefacts 243

2HDM Two Higgs Doublet Model 41, 42, 59, 133, 146, 232

AK4 Anti-$\boldsymbol{k_T}$ algorithm clustered jets with $\boldsymbol{R} = 0.4$ 89, 94, 95, 173, 187–189, 191, 198, 205, 209, 217, 245–250

AK8 Anti-$\boldsymbol{k_T}$ algorithm clustered jets with $\boldsymbol{R} = 0.8$ 89, 90, 94, 95, 173, 187–189, 191, 194, 196, 205, 209, 212, 217, 245–250, 252

ALICE A Large Ion Collider Experiment 64

APD avalanche photodiode 74

AR Application region 210, 213

ASIC Application-Specific Integrated Circuit 78

ATLAS A Toroidal LHC ApparatuS 9, 64, 128, 133, 169, 171–173, 175–179, 233, 234, 236

AUC Area Under Curve 121, 122, 139–141

AVR Adaptive Vertex Reconstruction 93, 94

BCM1F Fast Beam Condition Monitor 99

BCMS Batch Compression Merging and Splitting 63

BDT Boosted Decision Tree 88, 94, 107, 114, 123–125, 174–176

BMTF Barrel Muon Track Finder 80

BPIX Barrel Pixel detector 72

BR branching ratio 30, 35, 37, 43, 146, 170, 172, 174, 175, 180, 184, 194, 195, 200, 202, 203, 217, 224, 226

BSM beyond the standard model 33, 40, 45, 46, 51, 53, 55, 59, 90, 97, 177, 207, 228, 231, 232, 241, 242, 261

BU Builder Unit 81

BX Bunch Crossing 67, 205

C2HDM Complex Two Higgs Doublet Model 42

CEPC Circular Electron Positron Collider 238, 239

CERN European Organization for Nuclear Research 61, 68, 81, 238, 243, 244

CGAN conditional generative adversarial network 161

CHS Charged-Hadron Subtraction 95–97, 106

CINN conditional invertible neural network 161, 166

CKM Cabibbo–Kobayashi–Maskawa 21, 22, 33

CL Confidence Level 164, 172, 176–178, 224, 225, 227, 231–234, 236, 240, 241, 257

CLIC Compact Linear Collider 238–240

CMS Compact Muon Solenoid 5, 9, 10, 31, 32, 61, 64, 68–71, 73, 74, 76, 77, 89, 90, 95, 97, 98, 100, 102, 104, 106, 107, 113, 114, 128, 133, 147, 149, 167, 169, 171–173, 175–179, 185, 205–207, 218, 221, 223, 224, 231, 233–236, 241, 243, 267

cMVAv2 Combined Multivariate Algorithm 94

CR control region 98, 107, 174, 185, 208, 209, 214–216

CSC cathode strip chamber 77, 78, 80, 84

CSVv2 Combined Secondary Vertex 94

CTF Combinatorial Track Finder 83

CVAE conditional variational autoencoder 161

DAQ Data Acquisition 82

DeepCSV Deep Combined Secondary Vertex 94, 134, 188

DL $HH \to bbWW \to bbl\nu l\nu$ (double lepton channel) 180–182, 189, 191–194, 196–204, 208, 210, 214, 215, 230, 233, 235, 251, 254, 255, 257, 258

DNN Deep Neural Network 10, 91, 94, 107, 125, 133–135, 137–146, 148, 149, 152, 153, 156, 160, 165–167, 174–176, 192, 193, 196–203, 210, 214, 215, 221, 223, 224, 230–232, 243, 244, 253

DT drift tube 77, 78, 80, 84, 99

EB electromagnetic calorimeter barrel 73, 74

ECAL electromagnetic calorimeter 73–75, 79, 80, 85, 87–89, 101, 102, 188, 205

EE electromagnetic calorimeter endcap 73, 74

EFT Effective Field Theory 10, 11, 40, 41, 44–46, 49, 50, 54–60, 163, 168, 169, 177, 179, 184, 185, 226, 227, 230, 231, 234, 242, 244, 256

ELU Exponential Linear Unit 126, 135

EMTF Endcap Muon Track Finder 80

EW electroweak 34, 35, 39, 44, 48, 206, 217, 237, 263

EWChL Electroweak Chiral Lagrangian 45

EWSB electroweak symmetry breaking 19, 21, 22, 41, 42, 49, 58, 59

FCC Future Circular Collider 240

FCC-ee Future Circular Collider $\mathbf{e^+e^-}$ 238–240

FCC-he Future Circular Collider **ep** 238–240

FCC-hh Future Circular Collider **pp** 238–240

FCNC Flavour Changing Neutral Current 42

FED Front-End Driver 81

FEROL Front-End Readout Optical Link 81

FF Fake Factor 210, 213

FPGA Field-Programmable Gate Array 78

FPIX Forward Pixel detector 72

FPR False Positive Rate 121

FSR Final State Radiation 23, 26, 27, 218, 263

GAN generative adversarial network 154, 156, 161, 166

GD Gradient Descent 119, 120, 124, 127, 128, 267, 268

GGF gluon-gluon fusion 34, 37–39, 50, 51, 53, 54, 56, 171–176, 179–181, 184, 192, 194, 195, 197, 200, 202, 203, 221, 222, 224, 226, 238, 242, 264, 265

GM Georgi-Machacek 42

GNN Graph Neural Network 166, 173

GSF Gaussian Sum Filter 85, 87

HB hadronic calorimeter barrel 75

HCAL hadronic calorimeter 75, 76, 79, 80, 88, 89, 101, 171, 206, 217

HE hadronic calorimeter endcap 75

HE-LHC High-Energy Large Hadron Collider 238, 239

HEFT Higgs Effective Field Theory 45, 47, 50, 58–60, 220, 228

HEP high-energy physics 11, 45, 61, 113, 114, 121, 150, 155, 159, 166, 167, 243, 244, 267, 268, 271, 274

HF forward hadronic calorimeter 75, 76, 91, 99

HFET Hadron foward calorimeter transverse energy 99

HFOC Hadron foward calorimeter zero counting 99

HL-LHC High-Luminosity Large Hadron Collider 7, 236, 237, 239

HLT High-Level Trigger 71, 79, 81, 82, 86, 101, 182, 183

HME Heavy Mass Estimator 189, 190, 201, 223, 230, 252, 253

HO outer hadronic calorimeter 75

HPD hybrid photodiode 75, 76

HPS hadron-plus-strips 90, 91

IAF Inverse Autoregressive Flow 158–160

ILC International Linear Collider 238–240

INN Invertible Neural Network 159

IP impact parameter 92, 93

IP interaction points 61, 64–67, 69–72, 74, 75, 84, 86, 92–94, 99, 100, 105, 237

IPS impact parameter significance 93, 94

IR In the context of HEP : low-energy or long-distance 43, 44

ISR Initial State Radiation 24, 26, 27, 151, 164, 166, 218, 263

IVF Inclusive Vertex Finder 94

JBP Jet B Probability 94

JEC Jet Energy Correction 102–104, 190

JER Jet Energy Resolution 103, 104, 217, 263

JES Jet Energy Scale 103, 144, 145, 162, 217, 263

JP Jet Probability 94

KF Kalman Filter 83–85

KK Kaluza-Klein 43, 44, 57, 253

KL Kullback-Leibler 155, 156, 159

L1 Level 1 Trigger 71, 74, 75, 77–82, 205

L1A L1 Accept 79–82

LBDS LHC Beam Dumping System 65

LBN Lorentz Boosted Network 193

leaky ReLU Leaky Rectified Linear Unit 126

LEP Large Electron–Positron Collider 61, 64

LHC Large Hadron Collider 9–11, 23, 49, 58, 61–66, 68, 69, 73, 113, 133, 151, 163, 165, 167, 168, 203, 231, 232, 236, 238, 240, 241

LHCb Large Hadron Collider beauty 64

LHeC High-Luminosity Large Hadron Collider 238, 239

LINAC2 Linear accelerator 2 62, 63

LO leading order 15, 28–30, 33, 34, 36–40, 47, 49, 51, 54, 151, 152, 179, 182, 184, 185, 207, 221, 222, 230, 239

LS luminosity section 81, 85, 182, 204

MADE Masked Autoencoder for Distribution Estimation 158

MAE Mean Absolute Error 115

MAF Masked Autoregressive Flow 158–160

ME matrix element 14, 24, 25, 29–31, 46, 50, 56, 132, 151–153, 163, 164, 168, 193, 218, 219, 244

MEM Matrix Element Method 10, 113, 114, 128, 130, 133–140, 142, 144–146, 148–152, 160, 161, 163, 165–168, 232, 243, 244

MET Missing Transverse Energy 86, 91, 97, 104, 134, 188–191, 195, 206, 213, 252, 263

MG5_aMC@NLO MG5_aMC@NLO 29, 182, 268

ML machine learning 10, 51, 113–115, 118, 121, 122, 124, 150, 154, 163, 166, 167, 180, 213, 243, 244

MLP multilayer perceptron 94

MPF Missing transverse momentum Projection Fraction 103

MPI multiple parton interaction 26, 28

MPS Machine Protection System 64

MR Measurement region 213, 214

MSE Mean Squared Error 115–117, 121, 127

MSSM Minimal Supersymmetric Model 42

MVA multi-variate analysis 84, 88, 91, 94, 98, 101, 106–108, 113, 130, 168, 172, 177, 185, 201

N^{3}LO next-to-next-to-next-to-leading order 39

NF normalising flow 154, 156–158, 160, 161

NLO next-to-leading order 15, 24, 28, 29, 36, 39, 40, 51, 57, 151, 152, 177, 179, 182, 184, 206, 221, 222, 230

NN Neural Network 114, 119, 125, 127, 128, 154–157, 160, 161, 165, 166

NNLO next-to-next-to-leading order 15, 34, 39, 184, 206, 207

NP Nuisance parameter 109, 110, 144, 164, 166, 216–218

NWA narrow width approximation 30, 38

OMTF Overlap Muon Track Finder 80

OOT out-of-time 95, 98–100, 103

PCA point of closest approach 85, 93, 96

PCC Pixel cluster counting 99

PD Primary Dataset 182

PDF parton distribution function 24–26, 28, 36, 38, 39, 129, 150, 152, 153, 164, 182, 184, 185, 217, 223, 237, 263, 268

pdf Probability density function 110, 111, 190, 252

PF Particle-Flow 85, 86, 88–90, 92, 95, 97, 102, 217

PLT Pixel Luminosity Telescope 99

PMNS Pontecorvo–Maki–Nakagawa–Sakata 22, 23

PMT photomultiplier tube 76

POI parameter of interest 108, 164, 173, 218–220, 224, 229

PS Proton Synchrotron 63

PSB Proton Synchrotron Booster 63

PU Pile Up 34

PUPPI Pile Up Per Particle Identification 95, 96

PV Primary Vertex 85, 86, 91–94, 96, 191, 206

QCD quantum chromodynamics 18, 23, 33, 34, 38, 39, 48, 57, 71, 89, 151, 173, 174, 176, 206, 217, 237, 263

QED quantum electrodynamics 17

QFT quantum field theory 10, 11, 168

ReLU Rectified Linear Unit 125, 126, 135, 196

RF Radio-Frequency 63, 64, 66

RMS Root Mean Squard 96, 120, 267

ROC Receiver Operating Characteristic 121, 122, 139–141, 144, 145, 148, 149, 197

RPC resistive plate chamber 77, 78, 80, 84

RS Randall–Sundrum 43

SC Super Cluster 101

SD soft drop 90, 187

SELU Scaled Exponential Linear Unit 135

SET Soft Electron Tagger 94

SF scale factor 98, 101–107, 162, 182, 204, 205, 207

SGD Stochastic Gradient Descent 120

SI international system 11

SILH Strong Interaction Light Higgs 47, 60

SiPM Silicon Photomultiplier 76

SL $HH \to bbWW \to bbl\nu qq$ (single lepton channel) 180–182, 186, 188, 190–193, 195, 196, 198–204, 208, 210, 214, 215, 217, 219, 233, 235, 251, 254, 255, 257, 258

SLU Softplus Linear Unit 196

SM standard model 5, 9–11, 15–19, 21–23, 25, 30–33, 35, 36, 38, 40–47, 49–51, 54–57, 59, 60, 90, 97, 114, 146, 169, 172, 177–181, 184, 194, 195, 200, 202, 203, 222, 224–227, 231–234, 236, 241, 256, 264, 265, 269

SMEFT Standard Model Effective Field Theory 45, 49, 50, 58–60

SMT Soft Muon Tagger 94

SPS Super Proton Synchrotron 63, 66

SR signal region 97, 107, 108, 180, 181, 185, 204, 208–210, 213–215

STXS Simplified Template Cross Sections 243

SV Secondary Vertex 91–94, 105, 171, 174

T&P Tag-and-Probe 87, 100, 101, 103, 105–107, 204

TCDS Trigger Control and Distribution System 79

TEC Tracker End Cap 72, 73

TF Transfer Function 114, 129, 130, 133, 150–152, 162–164

TIB Tracker Inner Barrel 72, 73

TID Tracker Inner Disk 72, 73

TOB Tracker Outer Barrel 72, 73

TPR True Positive Rate 121

TT Trigger Tower 79, 80

UE Underlying Event 26, 28, 182, 218, 237, 240, 263

UV In the context of HEP : high-energy or small-distance 43, 46, 49, 56–59, 242

VAE variational auto-encoder 154–156, 158

VBF vector-boson fusion 34, 38, 39, 53–55, 57, 58, 172–175, 181, 184, 185, 187, 188, 192, 194, 195, 197, 200, 202, 207, 218, 221, 222, 224–226, 231, 234, 238, 240, 242, 246, 249, 263–265

VdM Luminosity scan 98

VEV vacuum expectation value 19, 20, 33, 36, 41, 50, 59

VPT vacuum phototriode 74

WED Warped extra dimension 43

WLS wavelength shifting fibre 75

WP Working Point 86–88, 106, 107, 134, 162, 185–188, 205, 217

References

[1] F. Bury and C. Delaere, *Matrix element regression with deep neural networks — Breaking the CPU barrier*, *Journal of High Energy Physics* **2021** (2021) 20 [2008.10949].

[2] M.E. Peskin and D.V. Schroeder, *An Introduction to Quantum Field Theory*, Westview Press (2016).

[3] M.A. Srednicki, *Quantum field theory*, Cambridge University Press (2007).

[4] D.H. Perkins, *"Introduction to High Energy Physics"*, Cambridge University Press, 3rd ed. (1987), "10.1017/CBO9780511809040".

[5] T. Plehn, *Lectures on LHC Physics*, vol. 844 (2012), 10.1007/978-3-642-24040-9.

[6] Particle Data Group collaboration, *Review of particle physics*, *Phys. Rev. D* **98** (2018) 030001.

[7] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, *Phys. Rev. Lett.* **13** (1964) 321.

[8] P.W. Higgs, *Broken symmetries, massless particles and gauge fields*, *Phys. Lett.* **12** (1964) 132.

[9] P.W. Higgs, *Broken symmetries and the masses of gauge bosons*, *Phys. Rev. Lett.* **13** (1964) 508.

[10] G.S. Guralnik, C.R. Hagen and T.W.B. Kibble, *Global conservation laws and massless particles*, *Phys. Rev. Lett.* **13** (1964) 585.

[11] ATLAS & CMS collaboration, *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*, 1503.07589.

[12] CMS collaboration, *Combined measurements of Higgs boson couplings in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *arXiv e-prints* (2018) arXiv:1809.10733 [1809.10733].

[13] D.J. Gross and F. Wilczek, *Ultraviolet behavior of non-abelian gauge theories*, *Phys. Rev. Lett.* **30** (1973) 1343.

[14] H.D. Politzer, *Reliable perturbative results for strong interactions?*, *Phys. Rev. Lett.* **30** (1973) 1346.

[15] R.D. Ball, V. Bertone, S. Carrazza, C.S. Deans, L. Del Debbio, S. Forte

et al., *Parton distributions for the LHC run II*, *Journal of High Energy Physics* **2015** (2015) 40 [1410.8849].

[16] A.D. Martin, W.J. Stirling, R.S. Thorne and G. Watt, *Parton distributions for the LHC*, *Eur. Phys. J. C* **63** (2009) 189 [0901.0002].

[17] S. Höche, *Introduction to parton-shower event generators*, *arXiv e-prints* (2014) arXiv:1411.4085 [1411.4085].

[18] H. Murayama, I. Watanabe and K. Hagiwara, *Helas: Helicity amplitude subroutines for feynman diagram evaluations*, 1992.

[19] P. de Aquino, W. Link, F. Maltoni, O. Mattelaer and T. Stelzer, *Aloha: Automatic libraries of helicity amplitudes for feynman diagram computations*, *Computer Physics Communications* **183** (2012) 2254.

[20] A. Alloul, N.D. Christensen, C. Degrande, C. Duhr and B. Fuks, *FEYNRULES 2.0 - A complete toolbox for tree-level phenomenology*, *Computer Physics Communications* **185** (2014) 2250 [1310.1921].

[21] F. Staub, *SARAH 4: A tool for (not only SUSY) model builders*, *Computer Physics Communications* **185** (2014) 1773 [1309.7223].

[22] G. Peter Lepage, *A new algorithm for adaptive multidimensional integration*, *Journal of Computational Physics* **27** (1978) 192.

[23] T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An introduction to PYTHIA 8.2*, *Computer Physics Communications* **191** (2015) 159 [1410.3012].

[24] T. Sjöstrand, S. Mrenna and P. Skands, *PYTHIA 6.4 physics and manual*, *Journal of High Energy Physics* **2006** (2006) 026 [hep-ph/0603175].

[25] M. Bähr, S. Gieseke, M.A. Gigg, D. Grellscheid, K. Hamilton, O. Latunde-Dada et al., *Herwig++ physics and manual*, *European Physical Journal C* **58** (2008) 639 [0803.0883].

[26] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, *Eur. Phys. J. C* **76** (2016) 196 [1512.01178].

[27] V.V. Sudakov, *Vertex parts at very high-energies in quantum electrodynamics*, *Sov. Phys. JETP* **3** (1956) 65.

[28] G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*, *Nuclear Physics B* **126** (1977) 298.

[29] B. Cabouat and T. Sjöstrand, *Some dipole shower studies*, *European Physical Journal C* **78** (2018) 226 [1710.00391].

[30] B. Andersson, G. Gustafson, G. Ingelman and T. Sjostrand, *Parton Fragmentation and String Dynamics*, *Phys. Rept.* **97** (1983) 31.

[31] B.R. Webber, *A QCD Model for Jet Fragmentation Including Soft Gluon Interference*, *Nucl. Phys. B* **238** (1984) 492.

[32] D. Amati, A. Bassetto, M. Ciafaloni, G. Marchesini and G. Veneziano, *A Treatment of Hard Processes Sensitive to the Infrared Structure of QCD*, *Nucl. Phys. B* **173** (1980) 429.

[33] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *Journal of High Energy Physics* **2014** (2014) 79 [1405.0301].

[34] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040 [hep-ph/0409146].

[35] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070 [0709.2092].

[36] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) 043 [1002.2581].

[37] W. Kilian, T. Ohl and J. Reuter, *WHIZARD—simulating multi-particle processes at LHC and ILC*, *European Physical Journal C* **71** (2011) 1742 [0708.4233].

[38] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert et al., *Event generation with SHERPA 1.1*, *Journal of High Energy Physics* **2009** (2009) 007 [0811.4622].

[39] S. Catani, F. Krauss, B.R. Webber and R. Kuhn, *QCD matrix elements parton showers*, *Journal of High Energy Physics* **2001** (2001) 063.

[40] F. Krauss, *Matrix Elements and Parton Showers in Hadronic Interactions*, *Journal of High Energy Physics* **2002** (2002) 015 [hep-ph/0205283].

[41] M.L. Mangano, M. Moretti, F. Piccinini and M. Treccani, *Matching matrix elements and shower evolution for top-pair production in hadronic collisions*, *Journal of High Energy Physics* **2007** (2007) 013 [hep-ph/0611129].

[42] L. Lonnblad and S. Prestel, *Unitarising Matrix Element + Parton Shower merging*, *JHEP* **02** (2013) 094 [1211.4827].

[43] S. Plätzer, *Controlling inclusive cross sections in parton shower + matrix element merging*, *JHEP* **08** (2013) 114 [1211.5467].

[44] N. Lavesson and L. Lönnblad, *Extending ckkw-merging to one-loop matrix elements*, *Journal of High Energy Physics* **2008** (2008) 070.

[45] L. Lönnblad and S. Prestel, *Merging Multi-leg NLO Matrix Elements with Parton Showers*, *JHEP* **03** (2013) 166 [1211.7278].

[46] N.D. Christensen and C. Duhr, *FeynRules - Feynman rules made easy*, *Computer Physics Communications* **180** (2009) 1614 [0806.4194].

[47] K. Hamilton, P. Nason and G. Zanderighi, *MINLO: Multi-Scale Improved NLO*, *JHEP* **10** (2012) 155 [1206.3572].

[48] R. Frederix and K. Hamilton, *Extending the MINLO method*, *JHEP* **05** (2016) 042 [1512.02663].

[49] P. Artoisenet, R. Frederix, O. Mattelaer and R. Rietkerk, *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, *Journal of High Energy Physics* **2013** (2013) 15 [1212.3460].

[50] E. Fuchs, S. Thewes and G. Weiglein, *Interference effects in BSM processes with a generalised narrow-width approximation*, *European Physical Journal C* **75** (2015) 254 [1411.4652].

[51] CMS collaboration, *Summaries of cms cross section measurements*, 2022.

[52] Planck collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [1807.06209].

[53] ATLAS collaboration, *Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc*, *Physics Letters B* **716** (2012) 1 [1207.7214].

[54] CMS collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Physics Letters B* **716** (2012) 30 [1207.7235].

[55] T. Plehn and M. Rauch, *Quartic Higgs coupling at hadron colliders*, Phys. Rev. D **72** (2005) 053008 [hep-ph/0507321].

[56] T. Binoth, S. Karg, N. Kauer and R. Rückl, *Multi-Higgs boson production in the standard model and beyond*, Phys. Rev. D **74** (2006) 113008 [hep-ph/0608057].

[57] B. Fuks, J.H. Kim and S.J. Lee, *Probing Higgs boson self-interactions in proton-proton collisions at a center-of-mass energy of 100 TeV*, 1510.07697.

[58] F. Maltoni, E. Vryonidou and M. Zaro, *Top-quark mass effects in double and triple Higgs production in gluon-gluon fusion at NLO*, *Journal of High Energy Physics* **2014** (2014) 79 [1408.6542].

[59] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino and R. Tanaka (Eds.), *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, *CERN-2011-002* (CERN, Geneva, 2011) [1101.0593].

[60] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino and R. Tanaka (Eds.), *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions*, *CERN-2012-002* (CERN, Geneva, 2012) [1201.3084].

[61] LHC Higgs Cross Section Working Group, S. Heinemeyer, C. Mariotti,

G. Passarino and R. Tanaka (Eds.), *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, *CERN-2013-004* (CERN, Geneva, 2013) [1307.1347].

[62] B. Di Micco, M. Gouzevitch, J. Mazzitelli, C. Vernieri, J. Alison, K. Androsov et al., *Higgs boson potential at colliders: status and perspectives*, *arXiv e-prints* (2019) [1910.00012].

[63] LHC Higgs Cross Section Working Group, S. Heinemeyer, C. Mariotti, G. Passarino and R. Tanaka (Eds.), *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, 1610.07922.

[64] R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, P. Torrielli et al., *Higgs pair production at the LHC with NLO and parton-shower effects*, *Physics Letters B* **732** (2014) 142 [1401.7340].

[65] G. Degrassi, P.P. Giardino, F. Maltoni and D. Pagani, *Probing the Higgs self coupling via single Higgs production at the LHC*, *Journal of High Energy Physics* **2016** (2016) 80 [1607.04251].

[66] F. Maltoni, D. Pagani, A. Shivaji and X. Zhao, *Trilinear Higgs coupling determination via single-Higgs differential measurements at the LHC*, *European Physical Journal C* **77** (2017) 887 [1709.08649].

[67] J.C. Collins and D.E. Soper, *Angular distribution of dileptons in high-energy hadron collisions*, *Phys. Rev. D* **16** (1977) 2219.

[68] Q.-H. Cao, G. Li, B. Yan, D.-M. Zhang and H. Zhang, *Double Higgs production at the 14 TeV LHC and a 100 TeV p p collider*, Phys. Rev. D **96** (2017) 095031 [1611.09336].

[69] LHC Higgs Cross Section Working Group collaboration, *LHC HXSWG interim recommendations to explore the coupling structure of a Higgs-like particle*, 1209.0040.

[70] R. Schabinger and J.D. Wells, *Minimal spontaneously broken hidden sector and its impact on higgs boson physics at the cern large hadron collider*, *Phys. Rev. D* **72** (2005) 093007.

[71] B. Patt and F. Wilczek, *Higgs-field Portal into Hidden Sectors*, *arXiv e-prints* (2006) hep [hep-ph/0605188].

[72] V. Barger, P. Langacker, M. McCaskey, M.J. Ramsey-Musolf and G. Shaughnessy, *CERN LHC phenomenology of an extended standard model with a real scalar singlet*, Phys. Rev. D **77** (2008) 035005 [0706.4311].

[73] M.T. Bowen, Y. Cui and J.D. Wells, *Narrow trans-TeV Higgs bosons and H→hh decays: two LHC search paths for a hidden sector Higgs boson*, *Journal of High Energy Physics* **2007** (2007) 036 [hep-ph/0701035].

[74] H. Davoudiasl, R. Kitano, T. Li and H. Murayama, *The new minimal standard model*, *Physics Letters B* **609** (2005) 117.

[75] M.T. Bowen, Y. Cui and J.D. Wells, *Narrow trans-TeV Higgs bosons and*

H→hh decays: two LHC search paths for a hidden sector Higgs boson, *Journal of High Energy Physics* **2007** (2007) 036 [hep-ph/0701035].

[76] M.J. Dolan, C. Englert and M. Spannowsky, *New physics in LHC Higgs boson pair production*, Phys. Rev. D **87** (2013) 055002 [1210.8166].

[77] G.M. Pruna and T. Robens, *Higgs singlet extension parameter space in the light of the LHC discovery*, Phys. Rev. D **88** (2013) 115012 [1303.1150].

[78] B. Cooper, N. Konstantinidis, L. Lambourne and D. Wardrope, *Boosted hh→bb¯bb¯: A new topology in searches for TeV-scale resonances at the LHC*, Phys. Rev. D **88** (2013) 114005 [1307.0407].

[79] J.M. No and M. Ramsey-Musolf, *Probing the Higgs portal at the LHC through resonant di-Higgs production*, Phys. Rev. D **89** (2014) 095031 [1310.6035].

[80] C.-Y. Chen, S. Dawson and I.M. Lewis, *Exploring resonant di-Higgs boson production in the Higgs singlet model*, Phys. Rev. D **91** (2015) 035015 [1410.5488].

[81] V. Martin-Lozano, J.M. Moreno and C.B. Park, *Resonant Higgs boson pair production in the $hh \to b\bar{b}\,WW \to b\bar{b}\ell^+\nu\ell^-\bar{\nu}$ decay channel*, *arXiv e-prints* (2015) arXiv:1501.03799 [1501.03799].

[82] S. Dawson and I.M. Lewis, *NLO corrections to double Higgs boson production in the Higgs singlet model*, Phys. Rev. D **92** (2015) 094023 [1508.05397].

[83] S.I. Godunov, A.N. Rozanov, M.I. Vysotsky and E.V. Zhemchugov, *Extending the Higgs sector: an extra singlet*, *European Physical Journal C* **76** (2016) 1 [1503.01618].

[84] L.-C. Lü, C. Du, Y. Fang, H.-J. He and H. Zhang, *Searching heavier Higgs boson via di-Higgs production at LHC Run-2*, *Physics Letters B* **755** (2016) 509 [1507.02644].

[85] F. Bojarski, G. Chalons, D. López-Val and T. Robens, *Heavy to light Higgs boson decays at NLO in the singlet extension of the Standard Model*, *Journal of High Energy Physics* **2016** (2016) 147 [1511.08120].

[86] T. Robens and T. Stefaniak, *LHC benchmark scenarios for the real Higgs singlet extension of the standard model*, *European Physical Journal C* **76** (2016) 268 [1601.07880].

[87] K. Nakamura, K. Nishiwaki, K.-y. Oda, S.C. Park and Y. Yamamoto, *Di-Higgs enhancement by neutral scalar as probe of new colored sector*, *European Physical Journal C* **77** (2017) 273 [1701.06137].

[88] T. Huang, J.M. No, L. Pernié, M. Ramsey-Musolf, A. Safonov, M. Spannowsky et al., *Resonant di-Higgs boson production in the b b ¯ W*

channel: Probing the electroweak phase transition at the LHC, Phys. Rev. D **96** (2017) 035007 [1701.04442].

[89] J. Chang, C.-R. Chen and C.-W. Chiang, *Higgs boson pair productions in the Georgi-Machacek model at the LHC*, *Journal of High Energy Physics* **2017** (2017) 137 [1701.06291].

[90] I. Lewis and M. Sullivan, *Benchmarks for double Higgs production in the singlet-extended standard model at the LHC*, Phys. Rev. D **96** (2017) 035037 [1701.08774].

[91] S. Dawson and M. Sullivan, *Enhanced di-Higgs boson production in the complex Higgs singlet model*, Phys. Rev. D **97** (2018) 015022 [1711.06683].

[92] A. Alves, T. Ghosh, H.-K. Guo, K. Sinha and D. Vagie, *Collider and gravitational wave complementarity in exploring the singlet extension of the standard model*, *Journal of High Energy Physics* **2019** (2019) 52 [1812.09333].

[93] A. Alves, T. Ghosh, H.-K. Guo and K. Sinha, *Resonant di-Higgs production at gravitational wave benchmarks: a collider study using machine learning*, *Journal of High Energy Physics* **2018** (2018) 70 [1808.08974].

[94] M. Carena, Z. Liu and M. Riembau, *Probing the electroweak phase transition via enhanced di-Higgs boson production*, Phys. Rev. D **97** (2018) 095032 [1801.00794].

[95] G.C. Branco, P.M. Ferreira, L. Lavoura, M.N. Rebelo, M. Sher and J.P. Silva, *Theory and phenomenology of two-Higgs-doublet models*, Phys. Rep. **516** (2012) 1 [1106.0034].

[96] J.F. Gunion, D. S., H.E. Habber and G.L. Kane, *The Higgs hunter's guide*, Brookhaven Nat. Lab. (1989).

[97] T. Plehn, M. Spira and P.M. Zerwas, *Pair production of neutral Higgs particles in gluon-gluon collisions*, *Nuclear Physics B* **479** (1996) 46 [hep-ph/9603205].

[98] A. Arhrib, R. Benbrik, C.-H. Chen, R. Guedes and R. Santos, *Double neutral Higgs production in the two-Higgs doublet model at the LHC*, *Journal of High Energy Physics* **2009** (2009) 035 [0906.0387].

[99] D. Fontes, J.C. Romão and J.P. Silva, *$h \to Z\gamma$ in the complex two Higgs doublet model*, *Journal of High Energy Physics* **2014** (2014) 43 [1408.2534].

[100] S. Baum and N.R. Shah, *Two Higgs doublets and a complex singlet: disentangling the decay topologies and associated phenomenology*, *Journal of High Energy Physics* **2018** (2018) 44 [1808.02667].

[101] M. Carena, H.E. Haber, I. Low, N.R. Shah and C.E.M. Wagner, *Alignment*

limit of the NMSSM Higgs sector, Phys. Rev. D **93** (2016) 035013 [1510.09137].

[102] A. Djouadi, L. Maiani, G. Moreau, A. Polosa, J. Quevillon and V. Riquer, *The post-Higgs MSSM scenario: habemus MSSM?*, *European Physical Journal C* **73** (2013) 2650 [1307.5205].

[103] A. Djouadi, L. Maiani, A. Polosa, J. Quevillon and V. Riquer, *Fully covering the MSSM Higgs sector at the LHC*, *Journal of High Energy Physics* **2015** (2015) 168 [1502.05653].

[104] H. Georgi and M. Machacek, *Doubly charged higgs bosons*, *Nuclear Physics B* **262** (1985) 463.

[105] M.S. Chanowitz and M. Golden, *Higgs boson triplets with mw=mzcos w*, *Physics Letters B* **165** (1985) 105.

[106] I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos and G. Dvali, *New dimensions at a millimeter to a fermi and superstrings at a TeV*, *Physics Letters B* **436** (1998) 257 [hep-ph/9804398].

[107] L. Randall and R. Sundrum, *Large Mass Hierarchy from a Small Extra Dimension*, Phys. Rev. Lett. **83** (1999) 3370 [hep-ph/9905221].

[108] R. Sundrum, *Compactification for a three-brane universe*, Phys. Rev. D **59** (1999) 085010 [hep-ph/9807348].

[109] H. Davoudiasl, J. Hewett and T. Rizzo, *Bulk gauge fields in the randall–sundrum model1work supported by the department of energy, contract de-ac03-76sf00515.1*, *Physics Letters B* **473** (2000) 43.

[110] L. Fitzpatrick, J. Kaplan, L. Randall and L.-T. Wang, *Searching for the kaluza-klein graviton in bulk RS models*, *Journal of High Energy Physics* **2007** (2007) 013.

[111] H. Davoudiasl, S. Gopalakrishna, E. Pontón and J. Santiago, *Warped five-dimensional models: phenomenological status and experimental prospects*, *New Journal of Physics* **12** (2010) 075011 [0908.1968].

[112] A. Carvalho, *Gravity particles from Warped Extra Dimensions, predictions for LHC*, *arXiv e-prints* (2014) arXiv:1404.0102 [1404.0102].

[113] C. Csáki, M.L. Graesser and G.D. Kribs, *Radion dynamics and electroweak physics*, *Phys. Rev. D* **63** (2001) 065002.

[114] C. Csáki, J. Hubisz and S.J. Lee, *Radion phenomenology in realistic warped space models*, *Phys. Rev. D* **76** (2007) 125015.

[115] S. Dawson and C.W. Murphy, *Standard model eft and extended scalar sectors*, *Phys. Rev. D* **96** (2017) 015041.

[116] A.C.A. Oliveira and R. Rosenfeld, *Hidden sector effects on double Higgs production near threshold at the LHC*, *Physics Letters B* **702** (2011) 201 [1009.4497].

[117] S. Dawson, A. Ismail and I. Low, *What's in the loop? the anatomy of double higgs production*, *Phys. Rev. D* **91** (2015) 115008.

[118] G. Cacciapaglia, H. Cai, A. Carvalho, A. Deandrea, T. Flacke, B. Fuks et al., *Probing vector-like quark models with Higgs-boson pair production*, *Journal of High Energy Physics* **2017** (2017) 5 [1703.10614].

[119] S. Weinberg, *Baryon- and lepton-nonconserving processes*, *Phys. Rev. Lett.* **43** (1979) 1566.

[120] W. Buchmüller and D. Wyler, *Effective lagrangian analysis of new interactions and flavour conservation*, *Nuclear Physics B* **268** (1986) 621.

[121] C. Love and S. Rao, *Low-energy manifestations of a new interactions scale: Operator analysis.*, *Z. Phys. C - Particles and Fields* **31** (1986) 433.

[122] C. Krause, *Higgs Effective Field Theories - Systematics and Applications*, *arXiv e-prints* (2016) arXiv:1610.08537 [1610.08537].

[123] A. Manohar and H. Georgi, *Chiral quarks and the non-relativistic quark model*, *Nuclear Physics B* **234** (1984) 189.

[124] R. Contino, A. Falkowski, F. Goertz, C. Grojean and F. Riva, *On the Validity of the Effective Field Theory Approach to SM Precision Tests*, *JHEP* **07** (2016) 144 [1604.06444].

[125] E.E. Jenkins, A.V. Manohar and M. Trott, *Renormalization group evolution of the standard model dimension six operators. I: formalism and λ dependence*, *Journal of High Energy Physics* **2013** (2013) 87 [1308.2627].

[126] E.E. Jenkins, A.V. Manohar and M. Trott, *Renormalization group evolution of the Standard Model dimension six operators II: Yukawa dependence*, *Journal of High Energy Physics* **2014** (2014) 35 [1310.4838].

[127] R. Alonso, E.E. Jenkins, A.V. Manohar and M. Trott, *Renormalization group evolution of the Standard Model dimension six operators III: gauge coupling dependence and phenomenology*, *Journal of High Energy Physics* **2014** (2014) 159 [1312.2014].

[128] A. Azatov, R. Contino, C.S. Machado and F. Riva, *Helicity selection rules and noninterference for bsm amplitudes*, *Phys. Rev. D* **95** (2017) 065014.

[129] S. Weinberg, *Baryon- and lepton-nonconserving processes*, *Phys. Rev. Lett.* **43** (1979) 1566.

[130] B. Grzadkowski, M. Iskrzyński, M. Misiak and J. Rosiek, *Dimension-six terms in the Standard Model Lagrangian*, *Journal of High Energy Physics* **2010** (2010) 85 [1008.4884].

[131] G.F. Giudice, C. Grojean, A. Pomarol and R. Rattazzi, *The strongly-interacting light Higgs*, *Journal of High Energy Physics* **2007** (2007) 045 [hep-ph/0703164].

[132] R. Contino, M. Ghezzi, C. Grojean, M. Mühlleitner and M. Spira, *Effective Lagrangian for a light Higgs-like scalar*, *Journal of High Energy Physics* **2013** (2013) 35 [1303.3876].

[133] W. Buchmüller and D. Wyler, *Effective lagrangian analysis of new interactions and flavour conservation*, *Nuclear Physics B* **268** (1986) 621.

[134] F. Maltoni, E. Vryonidou and C. Zhang, *Higgs production in association with a top-antitop pair in the Standard Model Effective Field Theory at NLO in QCD*, *Journal of High Energy Physics* **2016** (2016) 123 [1607.05330].

[135] F. Goertz, A. Papaefstathiou, L.L. Yang and J. Zurita, *Higgs boson pair production in the D = 6 extension of the SM*, *Journal of High Energy Physics* **2015** (2015) 167 [1410.3471].

[136] A. Azatov, R. Contino, G. Panico and M. Son, *Effective field theory analysis of double Higgs boson production via gluon fusion*, Phys. Rev. D **92** (2015) 035001 [1502.00539].

[137] A. Carvalho, M. Dall'Osso, T. Dorigo, F. Goertz, C.A. Gottardo and M. Tosi, *Higgs pair production: choosing benchmarks with cluster analysis*, *Journal of High Energy Physics* **2016** (2016) 126 [1507.02245].

[138] G. Buchalla, M. Capozi, A. Celis, G. Heinrich and L. Scyboz, *Higgs boson pair production in non-linear Effective Field Theory with full m $_t$-dependence at NLO QCD*, *Journal of High Energy Physics* **2018** (2018) 57 [1806.05162].

[139] G. Buchalla, M. Capozi, A. Celis, G. Heinrich and L. Scyboz, *Higgs boson pair production in non-linear Effective Field Theory with full m $_t$-dependence at NLO QCD*, *Journal of High Energy Physics* **2018** (2018) 57 [1806.05162].

[140] M. Capozi and G. Heinrich, *Exploring anomalous couplings in Higgs boson pair production through shape analysis*, *Journal of High Energy Physics* **2020** (2020) 91 [1908.08923].

[141] R. Contino, C. Grojean, M. Moretti, F. Piccinini and R. Rattazzi, *Strong Double Higgs Production at the LHC*, *JHEP* **05** (2010) 089 [1002.1011].

[142] F. Bishara, R. Contino and J. Rojo, *Higgs pair production in vector-boson fusion at the LHC and beyond*, *European Physical Journal C* **77** (2017) 481 [1611.03860].

[143] L. Berthier and M. Trott, *Consistent constraints on the Standard Model Effective Field Theory*, *JHEP* **02** (2016) 069 [1508.05060].

[144] A. Greljo, G. Isidori, J.M. Lindert and D. Marzocca, *Pseudo-observables in electroweak Higgs production*, *Eur. Phys. J. C* **76** (2016) 158 [1512.06135].

[145] C. Englert and M. Spannowsky, *Effective Theories and Measurements at Colliders*, *Phys. Lett. B* **740** (2015) 8 [1408.5147].

[146] A. Biekötter, A. Knochel, M. Krämer, D. Liu and F. Riva, *Vices and virtues of Higgs effective field theories at large energy*, *Phys. Rev. D* **91** (2015) 055029 [1406.7320].

[147] D.B. Kaplan and H. Georgi, *Su(2) $\times$ u(1) breaking by vacuum misalignment*, *Physics Letters B* **136** (1984) 183.

[148] G.F. Giudice, C. Grojean, A. Pomarol and R. Rattazzi, *The strongly-interacting light higgs*, *Journal of High Energy Physics* **2007** (2007) 045.

[149] A. Belyaev, A.C.A. Oliveira, R. Rosenfeld and M.C. Thomas, *Multi Higgs and Vector boson production beyond the Standard Model*, *JHEP* **05** (2013) 005 [1212.3860].

[150] R. Contino, C. Grojean, D. Pappadopulo, R. Rattazzi and A. Thamm, *Strong Higgs Interactions at a Linear Collider*, *JHEP* **02** (2014) 006 [1309.7038].

[151] J. D'Hondt, A. Mariotti, K. Mimasu, S. Moortgat and C. Zhang, *Learning to pinpoint effective operators at the LHC: a study of the $\mathrm{t\bar{t}b\bar{b}}$ signature*, *JHEP* **11** (2018) 131 [1807.02130].

[152] T. Cohen, N. Craig, X. Lu and D. Sutherland, *Is SMEFT Enough?*, *JHEP* **03** (2021) 237 [2008.08597].

[153] G. Buchalla, O. Cata, A. Celis and C. Krause, *Standard Model Extended by a Heavy Singlet: Linear vs. Nonlinear EFT*, *Nucl. Phys. B* **917** (2017) 209 [1608.03564].

[154] B. Henning, X. Lu and H. Murayama, *How to use the Standard Model effective field theory*, *JHEP* **01** (2016) 023 [1412.1837].

[155] M. Gorbahn, J.M. No and V. Sanz, *Benchmarks for Higgs Effective Theory: Extended Higgs Sectors*, *JHEP* **10** (2015) 036 [1502.07352].

[156] J. Brehmer, A. Freitas, D. Lopez-Val and T. Plehn, *Pushing Higgs Effective Theory to its Limits*, *Phys. Rev. D* **93** (2016) 075014 [1510.03443].

[157] G. Heinrich, J. Lang and L. Scyboz, *SMEFT predictions for gg $\to$ hh at full NLO QCD and truncation uncertainties*, *JHEP* **08** (2022) 079 [2204.13045].

[158] I. Brivio et al., *Truncation, validity, uncertainties*, 2201.04974.

[159] M. Trott, *A methodology for theory uncertainties in the SMEFT*, *arXiv e-prints* (2021) arXiv:2106.13794 [2106.13794].

[160] E. Mobs, *The CERN accelerator complex. Complexe des accélérateurs du CERN*, .

[161] H. Bartosik, C. Carli, H. Damerau, R. Garoby, S. Gilardoni, B. Goddard et al., *Performance potential of the injectors after LS1*, .

[162] ATLAS collaboration, *The ATLAS experiment at the CERN large hadron collider*, *Journal of Instrumentation* **3** (2008) S08003.

[163] CMS collaboration, *The CMS experiment at the CERN LHC*, *Journal of Instrumentation* **3** (2008) S08004.

[164] LHCb collaboration, *The LHCb detector at the LHC*, *Journal of Instrumentation* **3** (2008) S08005.

[165] ALICE collaboration, *The ALICE experiment at the CERN LHC*, *Journal of Instrumentation* **3** (2008) S08002.

[166] LHCf collaboration, *The LHCf detector at the CERN large hadron collider*, *Journal of Instrumentation* **3** (2008) S08006.

[167] MoEDAL collaboration, *The physics programme of the MoEDAL experiment at the LHC*, *International Journal of Modern Physics A* **29** (2014) 1430050 [1405.7662].

[168] TOTEM collaboration, *The TOTEM experiment at the CERN large hadron collider*, *Journal of Instrumentation* **3** (2008) S08007.

[169] J. Wenninger, *Machine protection and operation for lhc*, .

[170] P. Grafström and W. Kozanecki, *Luminosity determination at proton colliders*, *Progress in Particle and Nuclear Physics* **81** (2015) 97.

[171] S. van der Meer, *Calibration of the effective beam height in the ISR*, Tech. Rep. , CERN, Geneva (1968).

[172] CMS collaboration, "Cms luminosity - public results." `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults`, 2022.

[173] CMS collaboration, *Pileup mitigation at CMS in 13 TeV data*, *Journal of Instrumentation* **15** (2020) P09018 [2003.00503].

[174] C.T.G. of the CMS Collaboration, *The CMS Phase-1 Pixel Detector Upgrade*, *JINST* **16** (2020) P02027. 84 p [2012.14304].

[175] CMS collaboration, *The Phase-2 Upgrade of the CMS Tracker*, Tech. Rep. , CERN, Geneva (Jun, 2017).

[176] A. Bartoloni, S. Baccaro, L.M. Barone, F. Cavallari, I. Dafinei, D.D. Re et al., *The CMS ECAL barrel HV system*, *Journal of Instrumentation* **8** (2013) C02039.

[177] B.S. Acharya, T. Aziz, S. Banerjee, S. Banerjee, H.S. Bawa, S.B. Beri et al., *The CMS Outer Hadron Calorimeter*, Tech. Rep. , CERN, Geneva (Jun, 2006).

[178] CMS collaboration, *Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, *JINST* **13** (2018) P06015. 53 p [1804.04528].

[179] CMS collaboration, *CMS Technical Design Report for the Level-1 Trigger Upgrade*, Tech. Rep. (Jun, 2013).

[180] CMS collaboration, *The Phase-2 Upgrade of the CMS DAQ Interim Technical Design Report*, Tech. Rep. , CERN, Geneva (Sep, 2017).

[181] CMS collaboration, *Alignment of the CMS tracker with LHC and cosmic ray data*, *Journal of Instrumentation* **9** (2014) P06009.

[182] T. Speer, W. Adam, R. Frühwirth, A. Strandlie, T. Todorov and M. Winkler, *Track reconstruction in the cms tracker*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **559** (2006) 143.

[183] R.E. Kalman, *A New Approach to Linear Filtering and Prediction Problems*, *Journal of Basic Engineering* **82** (1960) 35.

[184] CMS collaboration, *Muon Reconstruction and Identification Performance with Run-2 data*, .

[185] CMS collaboration, *Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$=13 TeV*, *Journal of Instrumentation* **13** (2018) P06015 [1804.04528].

[186] CMS collaboration, *Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{}$ is/i= 8 TeV*, *Journal of Instrumentation* **10** (2015) P06005.

[187] W. Adam, R. Frühwirth, A. Strandlie and T. Todorov, *Reconstruction of electrons with the gaussian-sum filter in the CMS tracker at the LHC*, *Journal of Physics G: Nuclear and Particle Physics* **31** (2005) N9.

[188] CMS collaboration, *Particle-flow reconstruction and global event description with the CMS detector*, *Journal of Instrumentation* **12** (2017) P10003.

[189] CMS collaboration, *Performance of electron and photon reconstruction in Run 2 with the CMS experiment*, .

[190] M. Cacciari, G.P. Salam and G. Soyez, *The anti-ik/isubit/i /subjet clustering algorithm*, *Journal of High Energy Physics* **2008** (2008) 063.

[191] M. Cacciari, G.P. Salam and G. Soyez, *FastJet user manual. (for version 3.0.2)*, *European Physical Journal C* **72** (2012) 1896 [1111.6097].

[192] *Boosted objects: a probe of beyond the standard model physics*, *European Physical Journal C* **71** (2011) 1661 [1012.5412].

[193] *Boosted objects and jet substructure at the LHC*, *arXiv e-prints* (2013) arXiv:1311.2708 [1311.2708].

[194] *Towards an Understanding of the Correlations in Jet Substructure*, *arXiv e-prints* (2015) arXiv:1504.00679 [1504.00679].

[195] A.J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft drop*, *Journal of High Energy Physics* **2014** (2014) 146 [1402.2657].

[196] CMS, *Performance of τ-lepton reconstruction and identification in CMS*, *Journal of Instrumentation* **7** (2012) 1001 [1109.6034].

[197] CMS collaboration, *Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2*, Tech. Rep. , CERN, Geneva (2016).

[198] CMS, *Identification of hadronic tau lepton decays using a deep neural network*, *arXiv e-prints* (2022) arXiv:2201.08458 [2201.08458].

[199] CMS collaboration, *Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV*, *Journal of Instrumentation* **13** (2018) P05011 [1712.07158].

[200] E. Bols, J. Kieseler, M. Verzetti, M. Stoye and A. Stakia, *Jet flavour classification using DeepJet*, *Journal of Instrumentation* **15** (2020) P12012 [2008.10519].

[201] CMS collaboration, *Pileup Removal Algorithms*, Tech. Rep. , CERN, Geneva (2014).

[202] D. Bertolini, P. Harris, M. Low and N. Tran, *Pileup per particle identification*, *Journal of High Energy Physics* **2014** (2014) 59 [1407.6013].

[203] M. Cacciari and G.P. Salam, *Pileup subtraction using jet areas*, *Physics Letters B* **659** (2008) 119.

[204] M. Cacciari, G.P. Salam and G. Soyez, *The catchment area of jets*, *Journal of High Energy Physics* **2008** (2008) 005 [0802.1188].

[205] G. M., *Overview of the cms run-2 luminosity determination and calibration methodology*, 2019, DOI.

[206] CMS collaboration, *Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV*, Tech. Rep. , CERN, Geneva (2016).

[207] CMS collaboration, *The new fast beam condition monitor using poly-crystalline diamond sensors for luminosity measurement at cms*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **936** (2019) 717.

[208] A. Ledeul, A. Savulescu, G. Segura, B. Styczen and D.V. Rivera, *CERN Supervision, Control and Data Acquisition System for Radiation and Environmental Protection*, in *Proc. 12th International Workshop on Emerging Technologies and Scientific Facilities Controls (PCaPAC'18), Hsinchu, Taiwan, 16-19 October 2018*, no. 12 in International Workshop on Emerging Technologies and Scientific Facilities Controls, (Geneva, Switzerland), pp. 248–252, JACoW Publishing, Jan., 2019, DOI.

[209] CMS collaboration, *Precision luminosity measurement in proton-proton*

collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS, *arXiv e-prints* (2021) arXiv:2104.01927 [2104.01927].

[210] CMS collaboration, *CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV*, Tech. Rep. , CERN, Geneva (2018).

[211] CMS collaboration, *CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV*, Tech. Rep. , CERN, Geneva (2019).

[212] *Geant4—a simulation toolkit*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003) 250.

[213] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens et al., *DELPHES 3: a modular framework for fast simulation of a generic collider experiment*, *Journal of High Energy Physics* **2014** (2014) 57 [1307.6346].

[214] CMS collaboration, *Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV*, *Journal of Instrumentation* **12** (2017) P02014 [1607.03663].

[215] CMS collaboration, *Jet energy scale and resolution measurement with Run 2 Legacy Data Collected by CMS at 13 TeV*, .

[216] CMS collaboration, *Identification of b-quark jets with the CMS experiment*, *Journal of Instrumentation* **8** (2013) P04013 [1211.4462].

[217] CMS collaboration, *Search for the associated production of the Higgs boson with a top-quark pair*, *arXiv e-prints* (2014) arXiv:1408.1682 [1408.1682].

[218] CMS collaboration, *Pileup Jet Identification*, Tech. Rep. , CERN, Geneva (2013).

[219] T. Junk, *Confidence level computation for combining searches with small statistics*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **434** (1999) 435.

[220] K. Cranmer, *Practical Statistics for the LHC*, *arXiv e-prints* (2015) arXiv:1503.07622 [1503.07622].

[221] L. Demortier, *P values and nuisance parameters*, in *Statistical issues for LHC physics. Proceedings, Workshop, PHYSTAT-LHC, Geneva, Switzerland, June 27-29, 2007*, p. 23, 2008, DOI.

[222] G.J. Feldman and R.D. Cousins, *Unified approach to the classical statistical analysis of small signals*, Phys. Rev. D **57** (1998) 3873 [physics/9711021].

[223] J.S. Conway, *Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra*, *arXiv e-prints* (2011) arXiv:1103.0354 [1103.0354].

[224] R. Barlow and C. Beeston, *Fitting using finite monte carlo samples*, *Computer Physics Communications* **77** (1993) 219.

[225] T.L.H.C.G. The ATLAS Collaboration, The CMS Collaboration, *Procedure for the LHC Higgs boson search combination in Summer 2011*, Tech. Rep. , CERN, Geneva (Aug, 2011).

[226] R.D. Cousins, J.T. Linnemann and J. Tucker, *Evaluation of three methods for calculating statistical significance when incorporating a systematic uncertainty into a test of the background-only hypothesis for a poisson process*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **595** (2008) 480.

[227] W.A. Rolke, A.M. López and J. Conrad, *Limits and confidence intervals in the presence of nuisance parameters*, *Nuclear Instruments and Methods in Physics Research A* **551** (2005) 493 [physics/0403059].

[228] A.L. Read, *Presentation of search results: The CL_s technique*, *J. Phys. G* **28** (2002) 2693.

[229] T. Junk, *Confidence level computation for combining searches with small statistics*, *Nucl. Instrum. Meth. A* **434** (1999) 435 [hep-ex/9902006].

[230] T.L.H.C.G. The ATLAS Collaboration, The CMS Collaboration, *Procedure for the LHC Higgs boson search combination in Summer 2011*, Tech. Rep. , CERN, Geneva (Aug, 2011).

[231] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554 [1007.1727].

[232] S.S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, *The Annals of Mathematical Statistics* **9** (1938) 60 .

[233] A. Wald, *Tests of statistical hypotheses concerning several parameters when the number of observations is large*, *Transactions of the American Mathematical Society* **54** (1943) 426.

[234] W. Verkerke and D. Kirkby, *The RooFit toolkit for data modeling*, *arXiv e-prints* (2003) physics/0306116 [physics/0306116].

[235] L. Moneta, K. Cranmer, G. Schott and W. Verkerke, *The RooStats project*, in *Proceedings of the 13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research. February 22-27*, p. 57, Jan., 2010 [1009.1003].

[236] F. James and M. Roos, *Minuit - a system for function minimization and analysis of the parameter errors and correlations*, *Computer Physics Communications* **10** (1975) 343.

[237] ROOT collaboration, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, .

[238] L. Heinrich, M. Feickert and G. Stark, "pyhf: v0.7.0." 10.5281/zenodo.1169739.

[239] L. Heinrich, M. Feickert, G. Stark and K. Cranmer, *pyhf: pure-python implementation of histfactory statistical models*, *Journal of Open Source Software* **6** (2021) 2823.

[240] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall (1999).

[241] M. Verleysen, *Machine learning: regression and dimensionality reduction*, 2018.

[242] A. Jung, *Machine Learning: The Basics*, *arXiv e-prints* (2018) arXiv:1805.05052 [1805.05052].

[243] T. Neupert, M.H. Fischer, E. Greplova, K. Choo and M. Denner, *Introduction to Machine Learning for the Sciences*, *arXiv e-prints* (2021) arXiv:2102.04883 [2102.04883].

[244] A.S. Cornell, W. Doorsamy, B. Fuks, G. Harmsen and L. Mason, *Boosted decision trees in the era of new physics: a smuon analysis case study*, *JHEP* **04** (2022) 015 [2109.11815].

[245] Y. LeCun, Y. Bengio and G. Hinton, *Deep learning*, *nature* **521** (2015) 436.

[246] S. Wager, S. Wang and P. Liang, *Dropout Training as Adaptive Regularization*, *ArXiv e-prints* (2013) [1307.1493].

[247] G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors." 2012.

[248] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, *Dropout: A simple way to prevent neural networks from overfitting*, *Journal of Machine Learning Research* **15** (2014) 1929.

[249] G. Montavon, G.B. Orr and K. Müller, eds., *Neural Networks: Tricks of the Trade - Second Edition*, vol. 7700 of *Lecture Notes in Computer Science*, Springer (2012), 10.1007/978-3-642-35289-8.

[250] N. Qian, *On the momentum term in gradient descent learning algorithms*, *Neural Networks* **12** (1999) 145.

[251] J.C. Duchi, E. Hazan and Y. Singer, *Adaptive subgradient methods for online learning and stochastic optimization.*, *J. Mach. Learn. Res.* **12** (2011) 2121.

[252] M.D. Zeiler, *ADADELTA: An Adaptive Learning Rate Method*, *arXiv e-prints* (2012) arXiv:1212.5701 [1212.5701].

[253] D.P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, *arXiv e-prints* (2014) arXiv:1412.6980 [1412.6980].

[254] Y. Nesterov, *A method for solving the convex programming problem with convergence rate o($1/k^2$)*, *Proceedings of the USSR Academy of Sciences* **269** (1983) 543.

[255] I. Sutskever, J. Martens, G. Dahl and G. Hinton, *On the importance of initialization and momentum in deep learning*, in *Proceedings of the 30th International Conference on Machine Learning*, S. Dasgupta and D. McAllester, eds., vol. 28 of *Proceedings of Machine Learning Research*, (Atlanta, Georgia, USA), pp. 1139–1147, PMLR, 17–19 Jun, 2013.

[256] Y. Bengio, N. Boulanger-Lewandowski and R. Pascanu, *Advances in Optimizing Recurrent Networks*, *arXiv e-prints* (2012) arXiv:1212.0901 [1212.0901].

[257] J. Brehmer, G. Louppe, J. Pavez and K. Cranmer, *Mining gold from implicit models to improve likelihood-free inference*, *arXiv e-prints* (2018) arXiv:1805.12244 [1805.12244].

[258] J. Brehmer, F. Kling, I. Espejo and K. Cranmer, *MadMiner: Machine learning-based inference for particle physics*, *arXiv e-prints* (2019) arXiv:1907.10621 [1907.10621].

[259] K. Cranmer, J. Pavez and G. Louppe, *Approximating Likelihood Ratios with Calibrated Discriminative Classifiers*, *arXiv e-prints* (2015) arXiv:1506.02169 [1506.02169].

[260] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair et al., *Generative Adversarial Networks*, *arXiv e-prints* (2014) arXiv:1406.2661 [1406.2661].

[261] Y. Freund and R.E. Schapire, *A decision-theoretic generalization of on-line learning and an application to boosting*, *Journal of Computer and System Sciences* **55** (1997) 119.

[262] T. Chen and C. Guestrin, *XGBoost: A scalable tree boosting system*, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, (New York, NY, USA), pp. 785–794, ACM, 2016, DOI.

[263] F. Rosenblatt, *The perceptron: A probabilistic model for information storage and organization in the brain.*, *Psychological Review* **65** (1958) 386.

[264] L. Shiloh-Perl and R. Giryes, *Introduction to deep learning*, *arXiv e-prints* (2020) arXiv:2003.03253 [2003.03253].

[265] D.E. Rumelhart, G.E. Hinton and R.J. Williams, *Learning Representations by Back-propagating Errors*, *Nature* **323** (1986) 533.

[266] X. Glorot and Y. Bengio, *Understanding the difficulty of training deep feedforward neural networks*, in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Y.W. Teh and

M. Titterington, eds., vol. 9 of *Proceedings of Machine Learning Research*, (Chia Laguna Resort, Sardinia, Italy), pp. 249–256, PMLR, 13–15 May, 2010, https://proceedings.mlr.press/v9/glorot10a.html.

[267] K. He, X. Zhang, S. Ren and J. Sun, *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*, *arXiv e-prints* (2015) arXiv:1502.01852 [1502.01852].

[268] S. Ioffe and C. Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, *arXiv e-prints* (2015) arXiv:1502.03167 [1502.03167].

[269] R. Dalitz and G.R. Goldstein, *Test of analysis method for top-antitop production and decay events*, *Proc. Roy. Soc. Lond. A* **A455** (1999) 2803 [hep-ph/9802249].

[270] CDF Collaboration, *Top Quark Mass Measurement in the $t\bar{t}$ All Hadronic Channel using a Matrix Element Technique in $p\bar{p}$ Collisions at $\sqrt{s} =$ 1.96-TeV*, *Phys. Rev. D* **79** (2009) 072010 [0811.1062].

[271] CDF Collaboration, *Measurement of the top quark mass with dilepton events selected using neuroevolution at CDF*, *Phys. Rev. Lett.* **102** (2009) 152001 [0807.4652].

[272] CDF Collaboration, *Measurement of the top-quark mass in the lepton+jets channel using a matrix element technique with the CDF II detector*, *Phys. Rev. D* **84** (2011) 071105 [1108.1601].

[273] CDF Collaboration, *Measurements of the Top-quark Mass and the $t\bar{t}$ Cross Section in the Hadronic $\tau+$ Jets Decay Channel at $\sqrt{s} = 1.96$ TeV*, *Phys. Rev. Lett.* **109** (2012) 192001 [1208.5720].

[274] D0 Collaboration, *Precision measurement of the top-quark mass in lepton+jets final states*, *Physical Review D* **91** (2015) [1501.07912].

[275] D0 Collaboration, *Measurement of the top quark mass using the matrix element technique in dilepton final states*, *Physical Review D* **94** (2016) [1606.02814].

[276] $\mathrm{D0}$ collaboration, *Measurement of the top quark mass in the lepton + jets final state with the matrix element method*, *Phys. Rev. D* **74** (2006) 092005.

[277] CMS collaboration, *Search for $\mathrm{t\bar{t}}H$ production in the $H \to \mathrm{b\bar{b}}$ decay channel with leptonic $\mathrm{t\bar{t}}$ decays in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *arXiv e-prints* (2018) arXiv:1804.03682 [1804.03682].

[278] CMS collaboration, *Search for $\mathrm{t\bar{t}}H$ production in the all-jet final state in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *arXiv e-prints* (2018) arXiv:1803.06986 [1803.06986].

[279] CMS collaboration, *Evidence for associated production of a Higgs boson with a top quark pair in final states with electrons, muons, and hadronically*

decaying τ leptons at $\sqrt{s} = 13$ TeV, *arXiv e-prints* (2018) arXiv:1803.05485 [1803.05485].

[280] CMS collaboration, *Measurement of spin correlations in t t-bar production using the matrix element method in the muon + jets final state in pp collisions at sqrt(s) = 8 TeV*, *arXiv e-prints* (2015) arXiv:1511.06170 [1511.06170].

[281] ATLAS, *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into in collisions at with the ATLAS detector*, *European Physical Journal C* **75** (2015) 349 [1503.05066].

[282] ATLAS collaboration, *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *arXiv e-prints* (2017) arXiv:1712.08895 [1712.08895].

[283] ATLAS collaboration, *Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector*, *arXiv e-prints* (2017) arXiv:1712.08891 [1712.08891].

[284] ATLAS collaboration, *Measurement of the Higgs boson coupling properties in the $H \rightarrow ZZ^* \rightarrow 4\ell$ decay channel at $\sqrt{s}$=13 TeV with the ATLAS detector*, *Journal of High Energy Physics* **2018** (2018) 95 [1712.02304].

[285] ATLAS collaboration, *Evidence for single top-quark production in the s-channel in proton-proton collisions at $\sqrt{s}$ =8 TeV with the ATLAS detector using the Matrix Element Method*, *arXiv e-prints* (2015) arXiv:1511.05980 [1511.05980].

[286] P. Artoisenet, V. Lemaître, F. Maltoni and O. Mattelaer, *Automation of the matrix element reweighting method*, *Journal of High Energy Physics* **2010** (2010) 68 [1007.3300].

[287] S. Brochet, C. Delaere, B. François, V. Lemaître, A. Mertens, A. Saggio et al., *MoMEMta, a modular toolkit for the Matrix Element Method at the LHC*, *European Physical Journal C* **79** (2019) 126 [1805.08555].

[288] ATLAS, *Search for a heavy Higgs boson decaying into a Z boson and another heavy Higgs boson in the $\ell\ell bb$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett. B* **783** (2018) 392 [1804.01126].

[289] CMS collaboration, *Search for new neutral Higgs bosons through the H$\rightarrow$ ZA $\rightarrow \ell^+\ell^-\mathrm{b\bar{b}}$ process in pp collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **03** (2020) 055 [1911.03781].

[290] B. François, "Search for resonant di-higgs production in cms and development of a model independent approach to look for new physics at the lhc." http://hdl.handle.net/2078.1/191754, 2017.

[291] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski and D. Whiteson,

Parameterized neural networks for high-energy physics, *Eur. Phys. J. C* **76** (2016) 235 [1601.07913].

[292] G. Klambauer, T. Unterthiner, A. Mayr and S. Hochreiter, *Self-Normalizing Neural Networks*, *arXiv e-prints* (2017) arXiv:1706.02515 [1706.02515].

[293] F. Chollet et al., "Keras." https://keras.io, 2015.

[294] M.A. et al, "TensorFlow: Large-scale machine learning on heterogeneous systems." http://tensorflow.org/, 2015.

[295] S. Wertz, *The matrix element method at the lhc: status and prospects for run ii*, *Journal of Physics: Conference Series* **762** (2016) 012053.

[296] J.S. Gainer, J. Lykken, K.T. Matchev, S. Mrenna and M. Park, *The Matrix Element Method: Past, Present, and Future*, in *Community Summer Study 2013: Snowmass on the Mississippi*, 7, 2013 [1307.3546].

[297] Y. Philip, G. Sergeu, N. Mark and D. Zhong, "Sustainable matrix element method through deep learning." https://hepsoftwarefoundation.org/cwp/hsf-cwp-018-CWP_sustainable_matrix_element_method.pdf, 2017.

[298] A. Güneş Baydin, L. Shao, W. Bhimji, L. Heinrich, L. Meadows, J. Liu et al., *Etalumis: Bringing Probabilistic Programming to Scientific Simulators at Scale*, *arXiv e-prints* (2019) arXiv:1907.03382 [1907.03382].

[299] A.G. Baydin et al., *Efficient Probabilistic Inference in the Quest for Physics Beyond the Standard Model*, 1807.07706.

[300] D.E. Soper and M. Spannowsky, *Finding physics signals with shower deconstruction*, *Phys. Rev. D* **84** (2011) 074002 [1102.3480].

[301] D.E. Soper and M. Spannowsky, *Finding top quarks with shower deconstruction*, *Phys. Rev. D* **87** (2013) 054012 [1211.3140].

[302] D.E. Soper and M. Spannowsky, *Finding physics signals with event deconstruction*, *Phys. Rev. D* **89** (2014) 094005 [1402.1189].

[303] C. Englert, O. Mattelaer and M. Spannowsky, *Measuring the Higgs-bottom coupling in weak boson fusion*, *Phys. Lett. B* **756** (2016) 103 [1512.03429].

[304] J. Alwall, A. Freitas and O. Mattelaer, *The Matrix Element Method and QCD Radiation*, *Phys. Rev. D* **83** (2011) 074010 [1010.2263].

[305] F. Fiedler, A. Grohsjean, P. Haefner and P. Schieferdecker, *The Matrix Element Method and its Application in Measurements of the Top Quark Mass*, *Nucl. Instrum. Meth. A* **624** (2010) 203 [1003.1316].

[306] J.M. Campbell, W.T. Giele and C. Williams, *The Matrix Element Method at Next-to-Leading Order*, *JHEP* **11** (2012) 043 [1204.4424].

[307] J.M. Campbell, R.K. Ellis, W.T. Giele and C. Williams, *Finding the Higgs*

boson in decays to $Z\gamma$ using the matrix element method at Next-to-Leading Order, *Phys. Rev. D* **87** (2013) 073005 [1301.7086].

[308] M. Kraus, T. Martini and P. Uwer, *Matrix Element Method at NLO for (anti-)$\mathbf{k_t}$-jet algorithms*, *Phys. Rev. D* **100** (2019) 076010 [1901.08008].

[309] T. Martini and P. Uwer, *Extending the Matrix Element Method beyond the Born approximation: Calculating event weights at next-to-leading order accuracy*, *JHEP* **09** (2015) 083 [1506.08798].

[310] T. Martini, M. Kraus, S. Peitzsch and P. Uwer, *The Matrix Element Method as a tool for precision and accuracy*, *PoS* **EPS-HEP2019** (2020) 673 [2009.14642].

[311] J. Bamber, *Machine Learning techniques to approximate Matrix Element Method calculations*, .

[312] K. Danziger, T. Janßen, S. Schumann and F. Siegert, *Accelerating Monte Carlo event generation – rejection sampling using neural network event-weight estimates*, *SciPost Phys.* **12** (2022) 164 [2109.11964].

[313] D. Maître and H. Truong, *A factorisation-aware Matrix element emulator*, *JHEP* **11** (2021) 066 [2107.06625].

[314] T. Hahn, *CUBA: A Library for multidimensional numerical integration*, *Comput. Phys. Commun.* **168** (2005) 78 [hep-ph/0404043].

[315] G.P. Lepage, *Adaptive multidimensional integration: VEGAS enhanced*, *J. Comput. Phys.* **439** (2021) 110386 [2009.05112].

[316] G. Grasseau, F. Beaudette, C. Perez, A. Zabi, A. Chiron, T. Strebler et al., *Deployment of a matrix element method code for the tth channel analysis on gpu's platform*, *EPJ Web of Conferences* **214** (2019) 06028.

[317] D. Schouten, A. DeAbreu and B. Stelzer, *Accelerated Matrix Element Method with Parallel Computing*, *Comput. Phys. Commun.* **192** (2015) 54 [1407.7595].

[318] S. Carrazza and J.M. Cruz-Martinez, *VegasFlow: accelerating Monte Carlo simulation across multiple hardware platforms*, *Comput. Phys. Commun.* **254** (2020) 107376 [2002.12921].

[319] J. Kanzaki, *Monte Carlo integration on GPU*, *Eur. Phys. J. C* **71** (2011) 1559 [1010.2107].

[320] L. Weng, *Flow-based deep generative models*, *lilianweng.github.io* (2018) .

[321] M.E. Abbasnejad, Q. Shi, A. van den Hengel and L. Liu, *A generative adversarial density estimator*, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10774–10783, 2019, DOI.

[322] B. Hashemi, N. Amin, K. Datta, D. Olivito and M. Pierini, *LHC analysis-specific datasets with Generative Adversarial Networks*, 1901.05282.

[323] R. Di Sipio, M. Faucci Giannelli, S. Ketabchi Haghighat and S. Palazzo, *DijetGAN: A Generative-Adversarial Network Approach for the Simulation of QCD Dijet Events at the LHC*, *JHEP* **08** (2019) 110 [1903.02433].

[324] A. Butter, T. Plehn and R. Winterhalder, *How to GAN LHC Events*, *SciPost Phys.* **7** (2019) 075 [1907.03764].

[325] S. Carrazza and F.A. Dreyer, *Lund jet images from generative and cycle-consistent adversarial networks*, *Eur. Phys. J. C* **79** (2019) 979 [1909.01359].

[326] SHiP collaboration, *Fast simulation of muons produced at the SHiP experiment using Generative Adversarial Networks*, *JINST* **14** (2019) P11028 [1909.04451].

[327] A. Butter, T. Plehn and R. Winterhalder, *How to GAN Event Subtraction*, 1912.08824.

[328] D.P. Kingma and M. Welling, *Auto-Encoding Variational Bayes*, *arXiv e-prints* (2013) arXiv:1312.6114 [1312.6114].

[329] D. Jimenez Rezende, S. Mohamed and D. Wierstra, *Stochastic Backpropagation and Approximate Inference in Deep Generative Models*, *arXiv e-prints* (2014) arXiv:1401.4082 [1401.4082].

[330] S. Kullback and R.A. Leibler, *On Information and Sufficiency*, *The Annals of Mathematical Statistics* **22** (1951) 79 .

[331] S. Otten, S. Caron, W. de Swart, M. van Beekveld, L. Hendriks, C. van Leeuwen et al., *Event Generation and Statistical Sampling for Physics with Deep Generative Models and a Density Information Buffer*, *Nature Commun.* **12** (2021) 2985 [1901.00875].

[332] D. Jimenez Rezende and S. Mohamed, *Variational Inference with Normalizing Flows*, *arXiv e-prints* (2015) arXiv:1505.05770 [1505.05770].

[333] J. Bendavid, *Efficient Monte Carlo Integration Using Boosted Decision Trees and Generative Deep Neural Networks*, 1707.00028.

[334] M.D. Klimek and M. Perelstein, *Neural Network-Based Approach to Phase Space Integration*, *SciPost Phys.* **9** (2020) 053 [1810.11509].

[335] E. Bothmann, T. Janßen, M. Knobbe, T. Schmale and S. Schumann, *Exploring phase space with Neural Importance Sampling*, *SciPost Phys.* **8** (2020) 069 [2001.05478].

[336] M. Germain, K. Gregor, I. Murray and H. Larochelle, *MADE: Masked Autoencoder for Distribution Estimation*, *arXiv e-prints* (2015) arXiv:1502.03509 [1502.03509].

[337] A. van den Oord, N. Kalchbrenner, O. Vinyals, L. Espeholt, A. Graves and K. Kavukcuoglu, *Conditional Image Generation with PixelCNN Decoders*, *arXiv e-prints* (2016) arXiv:1606.05328 [1606.05328].

[338] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves et al., *WaveNet: A Generative Model for Raw Audio*, *arXiv e-prints* (2016) arXiv:1609.03499 [1609.03499].

[339] G. Papamakarios, T. Pavlakou and I. Murray, *Masked Autoregressive Flow for Density Estimation*, *arXiv e-prints* (2017) arXiv:1705.07057 [1705.07057].

[340] D.P. Kingma, T. Salimans, R. Jozefowicz, X. Chen, I. Sutskever and M. Welling, *Improving Variational Inference with Inverse Autoregressive Flow*, *arXiv e-prints* (2016) arXiv:1606.04934 [1606.04934].

[341] B. Stienen and R. Verheyen, *Phase space sampling and inference from weighted events with autoregressive flows*, *SciPost Phys.* **10** (2021) 038 [2011.13445].

[342] L. Dinh, D. Krueger and Y. Bengio, *NICE: Non-linear Independent Components Estimation*, *arXiv e-prints* (2014) arXiv:1410.8516 [1410.8516].

[343] L. Dinh, J. Sohl-Dickstein and S. Bengio, *Density estimation using Real NVP*, *arXiv e-prints* (2016) arXiv:1605.08803 [1605.08803].

[344] L. Ardizzone, J. Kruse, S. Wirkert, D. Rahner, E.W. Pellegrini, R.S. Klessen et al., *Analyzing Inverse Problems with Invertible Neural Networks*, *arXiv e-prints* (2018) arXiv:1808.04730 [1808.04730].

[345] D.P. Kingma and P. Dhariwal, *Glow: Generative Flow with Invertible 1x1 Convolutions*, *arXiv e-prints* (2018) arXiv:1807.03039 [1807.03039].

[346] T. Müller, B. McWilliams, F. Rousselle, M. Gross and J. Novák, *Neural Importance Sampling*, *arXiv e-prints* (2018) arXiv:1808.03856 [1808.03856].

[347] C. Durkan, A. Bekasov, I. Murray and G. Papamakarios, *Neural Spline Flows*, *arXiv e-prints* (2019) arXiv:1906.04032 [1906.04032].

[348] C. Gao, J. Isaacson and C. Krause, *i-flow: High-dimensional Integration and Sampling with Normalizing Flows*, *Mach. Learn. Sci. Tech.* **1** (2020) 045023 [2001.05486].

[349] M. Mirza and S. Osindero, *Conditional Generative Adversarial Nets*, *arXiv e-prints* (2014) arXiv:1411.1784 [1411.1784].

[350] A. Pagnoni, K. Liu and S. Li, *Conditional Variational Autoencoder for Neural Machine Translation*, *arXiv e-prints* (2018) arXiv:1812.04405 [1812.04405].

[351] L. Ardizzone, C. Lüth, J. Kruse, C. Rother and U. Köthe, *Guided Image Generation with Conditional Invertible Neural Networks*, *arXiv e-prints* (2019) arXiv:1907.02392 [1907.02392].

[352] A. Grover, M. Dhar and S. Ermon, *Flow-GAN: Combining Maximum*

Likelihood and Adversarial Learning in Generative Models, *arXiv e-prints* (2017) arXiv:1705.08868 [1705.08868].

[353] N. Soybelman, A. Butter, T. Plehn and J. Brehmer, *Stronger symbolic summary statistics for the lhc*, *Proc. of Workshop at the 35th Conference on Neural Information Processing Systems (NeurIPS) (December 13, 2021)* (2021) .

[354] A. Pin and O. Mattelaer, *Determination of differential cross sections from t anti-t fully leptonic, using the matrix element method*, *Nuovo Cim. C* **035N3** (2012) 229.

[355] M. Feickert, M. Katare, M. Neubauer and A. Roy, *Deep Learning for the Matrix Element Method*, in *41st International Conference on High Energy Physics*, 11, 2022 [2211.11910].

[356] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang and P.S. Yu, *A Comprehensive Survey on Graph Neural Networks*, *arXiv e-prints* (2019) arXiv:1901.00596 [1901.00596].

[357] J. Shlomi, P. Battaglia and J.-R. Vlimant, *Graph Neural Networks in Particle Physics*, 2007.13681.

[358] M. Shuaibi, A. Kolluru, A. Das, A. Grover, A. Sriram, Z. Ulissi et al., *Rotation Invariant Graph Neural Networks using Spin Convolutions*, *arXiv e-prints* (2021) arXiv:2106.09575 [2106.09575].

[359] S. Li, *Rotation Invariance Neural Network*, *arXiv e-prints* (2017) arXiv:1706.05534 [1706.05534].

[360] G. Louppe, M. Kagan and K. Cranmer, *Learning to Pivot with Adversarial Networks*, 1611.01046.

[361] A. Ghosh, B. Nachman and D. Whiteson, *Uncertainty-aware machine learning for high energy physics*, *Phys. Rev. D* **104** (2021) 056026 [2105.08742].

[362] A. Butter, T. Heimel, T. Martini, S. Peitzsch and T. Plehn, *Two Invertible Networks for the Matrix Element Method*, 2210.00019.

[363] K. Ehatäht and C. Veelken, *Application of the matrix element method to Higgs boson pair production in the channel HH $\rightarrow$ bbWW* at the LHC*, *Nucl. Instrum. Meth. A* **1028** (2022) 166373 [2108.05267].

[364] CMS collaboration, *Combination of searches for higgs boson pair production in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Phys. Rev. Lett.* **122** (2019) 121803.

[365] ATLAS collaboration, *Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{\{ s \}}$ = 13TeV with the ATLAS detector*, *Physics Letters B* **800** (2020) 135103 [1906.02025].

[366] ATLAS collaboration, *Searches for Higgs boson pair production in the hh*

$\to$*bb* $\tau\tau$*,* $\gamma\gamma$*WW** *,* $\gamma\gamma$*bb , bbbb channels with the ATLAS detector*, Phys. Rev. D **92** (2015) 092004 [1509.04670].

[367] ATLAS collaboration, *Search for the HH* $\to$*bb*$^-$*bb*$^-$ *process via vector-boson fusion production using proton-proton collisions at* $\sqrt{\{}s\ \} = 13$ *TeV with the ATLAS detector*, *Journal of High Energy Physics* **2021** (2021) 145 [2001.05178].

[368] ATLAS collaboration, *Search for non-resonant Higgs boson pair production in the* $bb\ell\nu\ell\nu$ *final state with the ATLAS detector in pp collisions at* $\sqrt{s} = 13$ *TeV*, *Physics Letters B* **801** (2020) 135145 [1908.06765].

[369] ATLAS collaboration, *Search for non-resonant pair production of Higgs bosons in the* $b\bar{b}b\bar{b}$ *final state in* pp *collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Tech. Rep. , CERN, Geneva (Jun, 2022).

[370] ATLAS collaboration, *Search for Higgs boson pair production in the two bottom quarks plus two photons final state in* pp *collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, *arXiv e-prints* (2021) [2112.11876].

[371] ATLAS collaboration, *Search for resonant and non-resonant Higgs boson pair production in the* $b\bar{b}\tau^+\tau^-$ *decay channel using 13 TeV* pp *collision data from the ATLAS detector*, .

[372] ATLAS collaboration, *Constraining the Higgs boson self-coupling from single- and double-Higgs production with the ATLAS detector using* pp *collisions at* $\sqrt{s} = 13$ *TeV*, 2211.01216.

[373] CMS collaboration, *Search for higgs boson pair production in the* $bb\tau\tau$ *final state in proton-proton collisions at* $\sqrt{(s)} = 8$ TeV, *Phys. Rev. D* **96** (2017) 072004.

[374] CMS collaboration, *Search for two higgs bosons in final states containing two photons and two bottom quarks in proton-proton collisions at 8 tev*, *Phys. Rev. D* **94** (2016) 052012.

[375] CMS collaboration, *Search for resonant pair production of higgs bosons decaying to two bottom quark–antiquark pairs in proton–proton collisions at 8 tev*, *Physics Letters B* **749** (2015) 560.

[376] CMS collaboration, *Search for nonresonant Higgs boson pair production in the four leptons plus two b jets final state in proton-proton collisions at* $\sqrt{s}$ *= 13 TeV*, *arXiv e-prints* (2022) arXiv:2206.10657 [2206.10657].

[377] CMS collaboration, *Search for Higgs boson pairs decaying to WWWW,* $WW\tau\tau$*, and* $\tau\tau\tau\tau$ *in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, *arXiv e-prints* (2022) arXiv:2206.10268 [2206.10268].

[378] CMS collaboration, *Search for nonresonant Higgs boson pair production in final states with two bottom quarks and two photons in proton-proton*

collisions at $\sqrt{s} = 13$ TeV, *arXiv e-prints* (2020) arXiv:2011.12373 [2011.12373].

[379] CMS collaboration, *Search for Higgs boson pair production in the four b quark final state in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *arXiv e-prints* (2022) arXiv:2202.09617 [2202.09617].

[380] CMS collaboration, *Search for nonresonant pair production of highly energetic Higgs bosons decaying to bottom quarks*, *arXiv e-prints* (2022) arXiv:2205.06667 [2205.06667].

[381] CMS collaboration, *Search for nonresonant Higgs boson pair production in final states with two bottom quarks and two tau leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV*, .

[382] CMS collaboration, *A portrait of the Higgs boson by the CMS experiment ten years after the discovery*, *Nature* **607** (2022) 60 [2207.00043].

[383] H. Qu and L. Gouskos, *ParticleNet: Jet Tagging via Particle Clouds*, *Phys. Rev. D* **101** (2020) 056019 [1902.08570].

[384] ATLAS collaboration, *Search for Higgs boson pair production in association with a vector boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, 2210.05415.

[385] CMS collaboration, *Combined Higgs boson production and decay measurements with up to 137 fb-1 of proton-proton collision data at sqrts = 13 TeV*, Tech. Rep. , CERN, Geneva (2020).

[386] ATLAS collaboration, *Constraint of the Higgs boson self-coupling from Higgs boson differential production and decay measurements*, Tech. Rep. , CERN, Geneva (2019).

[387] ATLAS collaboration, *Constraints on the Higgs boson self-coupling from the combination of single-Higgs and double-Higgs production analyses performed with the ATLAS experiment*, Tech. Rep. , CERN, Geneva (2019).

[388] F. Monti, C.E. Pandini, F.L. Lucio Alves, H. Yang, Y. Huang, J. Wang et al., *Modelling of the single-Higgs simplified template cross-sections (STXS 1.2) for the determination of the Higgs boson trilinear self-coupling*, Tech. Rep. , CERN, Geneva (2022).

[389] U. Haisch and G. Koole, *Off-shell Higgs production at the LHC as a probe of the trilinear Higgs coupling*, *JHEP* **02** (2022) 030 [2111.12589].

[390] Anisha, S. Das Bakshi, S. Banerjee, A. Biekötter, J. Chakrabortty, S. Kumar Patra et al., *Effective limits on single scalar extensions in the light of recent LHC data*, 2111.05876.

[391] S. Di Vita, C. Grojean, G. Panico, M. Riembau and T. Vantalon, *A global view on the Higgs self-coupling*, *JHEP* **09** (2017) 069 [1704.01953].

[392] CMS collaboration, *Search for heavy resonances decaying to a pair of*

Lorentz-boosted Higgs bosons in final states with leptons and a bottom quark pair at $\sqrt{s} = 13$ TeV, *arXiv e-prints* (2021) arXiv:2112.03161 [2112.03161].

[393] CMS collaboration, *Search for a heavy Higgs boson decaying into two lighter Higgs bosons in the $\tau\tau$bb final state at 13 TeV*, *JHEP* **11** (2021) 057 [2106.10361].

[394] CMS collaboration, *Search for a new resonance decaying to two scalars in the final state with two bottom quarks and two photons in proton-proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$*, Tech. Rep. , CERN, Geneva (2022).

[395] CMS collaboration, *Search for resonant and nonresonant Higgs boson pair production in the b{b}$\ell\ \nu\ \ell\ \nu$ final state in proton-proton collisions at $\sqrt{s}$=13 TeV*, *Journal of High Energy Physics* **2018** (2018) 54 [1708.04188].

[396] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, *Comput. Phys. Commun.* **185** (2014) 2930 [1112.5675].

[397] Y. Li and F. Petriello, *Combining QCD and electroweak corrections to dilepton production in the framework of the FEWZ simulation code*, Phys. Rev. D **86** (2012) 094034 [1208.5967].

[398] CMS collaboration, "Summary table of samples produced for the 1 billion campaign, with 25ns bunch-crossing." https://twiki.cern.ch/twiki/bin/view/CMS/SummaryTable1G25ns, 2022.

[399] M. Brucherseifer, F. Caola and K. Melnikov, *On the NNLO QCD corrections to single-top production at the LHC*, *Phys. Lett. B* **736** (2014) 58 [1404.7116].

[400] CMS collaboration, "Cross sections and validation for wz and zz samples." https://indico.cern.ch/event/448517/contributions/1943045/attachments/1164999/1679225/Long_Generators_WZxsec_05_10_15.pdf, 2019.

[401] K. Ehataht and C. Veelken, *Stitching Monte Carlo samples*, *arXiv e-prints* (2021) arXiv:2106.04360 [2106.04360].

[402] *Event generator tunes obtained from underlying event and multiparton scattering measurements*, *European Physical Journal C* **76** (2016) 155 [1512.00815].

[403] CMS collaboration, *Extraction and validation of a new set of CMS uc(pythia)8 tunes from underlying-event measurements*, *European Physical Journal C* **80** (2020) 4 [1903.12179].

[404] C. Diez and M. Seidel, "Powheg+pythia8 $t\bar{t}$: tuning α_S^{ISR} and h_{damp}."

https://indico.cern.ch/event/567994/contributions/2299854/attachments/1336057/2009649/tt_tuning_mseidel.pdf, 2019.

[405] R.D. Ball, V. Bertone, S. Carrazza, C.S. Deans, L. Del Debbio, S. Forte et al., *Parton distributions for the LHC run II*, *Journal of High Energy Physics* **2015** (2015) 40 [1410.8849].

[406] R.D. Ball, V. Bertone, S. Carrazza, L.D. Debbio, S. Forte, P. Groth-Merrild et al., *Parton distributions from high-precision collider data*, *European Physical Journal C* **77** (2017) 663 [1706.00428].

[407] J. Butterworth, S. Carrazza, A. Cooper-Sarkar, A. De Roeck, J. Feltesse, S. Forte et al., *PDF4LHC recommendations for LHC Run II*, *Journal of Physics G Nuclear Physics* **43** (2016) 023001 [1510.03865].

[408] J. Rojo, A. Accardi, R.D. Ball, A. Cooper-Sarkar, A. de Roeck, S. Farry et al., *The PDF4LHC report on PDFs and LHC data: results from Run I and preparation for Run II*, *Journal of Physics G Nuclear Physics* **42** (2015) 103103 [1507.00556].

[409] A. Accardi, S. Alekhin, J. Blümlein, M.V. Garzelli, K. Lipka, W. Melnitchouk et al., *A critical appraisal and evaluation of modern PDFs*, *European Physical Journal C* **76** (2016) 471 [1603.08906].

[410] J. Baglio, F. Campanario, S. Glaus, M. Mühlleitner, J. Ronca and M. Spira, $gg \to HH$ *: Combined uncertainties*, *Phys. Rev. D* **103** (2021) 056002 [2008.11626].

[411] S. Amoroso et al., *Les Houches 2019: Physics at TeV Colliders: Standard Model Working Group Report*, in *11th Les Houches Workshop on Physics at TeV Colliders: PhysTeV Les Houches*, 3, 2020 [2003.01700].

[412] CMS collaboration, *Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at* $\sqrt{s} = 13$ *TeV*, *arXiv e-prints* (2020) arXiv:2011.03652 [2011.03652].

[413] J. Thaler and K. Van Tilburg, *Identifying boosted objects with N-subjettiness*, *Journal of High Energy Physics* **2011** (2011) 15 [1011.2268].

[414] T. Huang, J.M. No, L. Pernié, M. Ramsey-Musolf, A. Safonov, M. Spannowsky et al., *Resonant di-higgs boson production in the* $b\bar{b}ww$ *channel: Probing the electroweak phase transition at the lhc*, *Phys. Rev. D* **96** (2017) 035007.

[415] P. Konar, K. Kong and K.T. Matchev, *root_s_hat_min: a global inclusive variable for determining the mass scale of new physics in events with missing energy at hadron colliders*, *Journal of High Energy Physics* **2009** (2009) 085 [0812.1042].

[416] P. Konar, K. Kong, K.T. Matchev and M. Park, *RECO level {sqrt {s}_{min }} and subsystem improved {sqrt {s}_{min }} : global inclusive variables for measuring the new physics mass scale in [InlineMediaObject not available: see fulltext.] events at hadron colliders*, *Journal of High Energy Physics* **2011** (2011) 41 [1006.0653].

[417] A.J. Barr, T.J. Khoo, P. Konar, K. Kong, C.G. Lester, K.T. Matchev et al., *Guide to transverse projections and mass-constraining variables*, Phys. Rev. D **84** (2011) 095031 [1105.2977].

[418] M. Erdmann, E. Geiser, Y. Rath and M. Rieger, *Lorentz boost networks: autonomous physics-inspired feature engineering*, *Journal of Instrumentation* **14** (2019) P06006.

[419] K. He, X. Zhang, S. Ren and J. Sun, *Deep residual learning for image recognition*, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016, DOI.

[420] H. Zhao, F. Liu, L. Li and C. Luo, *A novel softplus linear unit for deep convolutional neural networks*, *Applied Intelligence* **48** (2018) 1707–1720.

[421] CMS collaboration, *Measurements of differential cross sections for* $\mathrm{t\bar{t}}$ *production in proton-proton collisions at* $\sqrt{s} = 13$ TeV *using events containing two leptons*, Tech. Rep. , CERN, Geneva (2018).

[422] M. Czakon, D. Heymes, A. Mitov, D. Pagani, I. Tsinikos and M. Zaro, *Top-pair production at the LHC through NNLO QCD and NLO EW*, *Journal of High Energy Physics* **2017** (2017) 186 [1705.04105].

[423] CMS collaboration, *Search for associated production of Higgs bosons and top quarks in multilepton final states at* $\sqrt{s} = 13$ TeV, .

[424] L. Cadamuro, M. Rieger and F. Monti, "Custom HH physics model implementing gluon gluon fusion (ggf / gghh), vector boson fusion (vbf / qqhh) and V boson associated production (vhh) modes." `https://gitlab.cern.ch/hh/tools/inference/-/blob/master/dhi/models/hh_model.py`, 2021.

[425] L. Cadamuro, A. Carvalho, P. Fackeldey, B. Fischer, T. Lange, D. Noll et al., "HH Inference Tools." `https://cms-hh.web.cern.ch/tools/inference/`, 2021.

[426] P. David, *Readable and efficient HEP data analysis with bamboo*, *EPJ Web Conf.* **251** (2021) 03052 [2103.01889].

[427] ATLAS collaboration, *Search for non-resonant higgs boson pair production in the bb final state with the atlas detector in pp collisions at* $\sqrt{s}$ *==13 tev*, *Physics Letters B* **801** (2020) 135145.

[428] ATLAS collaboration, *A detailed map of Higgs boson interactions by the*

ATLAS experiment ten years after the discovery, *Nature* **607** (2022) 52 [2207.00092].

[429] CMS collaboration, *Prospects for HH measurements at the HL-LHC*, Tech. Rep. , CERN, Geneva (2018).

[430] ATLAS collaboration, *Measurement prospects of the pair production and self-coupling of the Higgs boson with the ATLAS experiment at the HL-LHC*, Tech. Rep. , CERN, Geneva (Dec, 2018).

[431] *Higgs Physics at the HL-LHC and HE-LHC*, *arXiv e-prints* (2019) arXiv:1902.00134 [1902.00134].

[432] R. Teixeira De Lima, *Accessing HH at ATLAS and CMS with the HL-LHC*, *PoS* **LHCP2020** (2021) 082. 5 p.

[433] *The International Linear Collider: A Global Project*, *arXiv e-prints* (2019) arXiv:1903.01629 [1903.01629].

[434] *The International Linear Collider. A Global Project*, *arXiv e-prints* (2019) arXiv:1901.09829 [1901.09829].

[435] M. Aicheler, P.N. Burrows, N. Catalan, R. Corsini, M. Draper, J. Osborne et al., *The Compact Linear Collider (CLIC) - Project Implementation Plan*, *arXiv e-prints* (2019) arXiv:1903.08655 [1903.08655].

[436] *The Compact Linear Collider (CLIC) - 2018 Summary Report*, *arXiv e-prints* (2018) arXiv:1812.06018 [1812.06018].

[437] I. Agapov, M. Benedikt, A. Blondel, M. Boscolo, O. Brunner, M. Chamizo Llatas et al., *Future Circular Lepton Collider FCC-ee: Overview and Status*, *arXiv e-prints* (2022) arXiv:2203.08310 [2203.08310].

[438] The CEPC Study Group, *CEPC Conceptual Design Report: Volume 1 - Accelerator*, *arXiv e-prints* (2018) arXiv:1809.00285 [1809.00285].

[439] The CEPC Study Group, *CEPC Conceptual Design Report: Volume 2 - Physics & Detector*, *arXiv e-prints* (2018) arXiv:1811.10545 [1811.10545].

[440] M. Benedikt, A. Chance, B. Dalena, D. Denisov, M. Giovannozzi, J. Gutleber et al., *Future Circular Hadron Collider FCC-hh: Overview and Status*, *arXiv e-prints* (2022) arXiv:2203.07804 [2203.07804].

[441] *A Large Hadron Electron Collider at CERN : Report on the Physics and Design Concepts for Machine and Detector*, *Journal of Physics G Nuclear Physics* **39** (2012) 075001 [1206.2913].

[442] J. Ellis, *The Future of High-Energy Collider Physics*, in *38th International Symposium on Physics in Collision*, 10, 2018 [1810.11263].

[443] J. de Blas, M. Cepeda, J. D'Hondt, R.K. Ellis, C. Grojean, B. Heinemann et al., *Higgs Boson studies at future particle colliders*, *Journal of High Energy Physics* **2020** (2020) 139 [1905.03764].

[444] M. Kumar, X. Ruan, A.S. Cornell, R. Islam and B. Mellado, *Double higgs production at FCC-he and prospects for measurements of self-coupling*, *Journal of Physics: Conference Series* **623** (2015) 012017.

[445] G. Bernardi, E. Brost, D. Denisov, G. Landsberg, M. Aleksa, D. d'Enterria et al., *The Future Circular Collider: a Summary for the US 2021 Snowmass Process*, *arXiv e-prints* (2022) arXiv:2203.06520 [2203.06520].

[446] A. Taliercio, P. Mastrapasqua, C. Caputo, P. Vischia, N. De Filippis and P. Bhat, *Higgs Self Couplings Measurements at Future proton-proton Colliders: a Snowmass White Paper*, *arXiv e-prints* (2022) arXiv:2203.08042 [2203.08042].

[447] S. Di Vita, G. Durieux, C. Grojean, J. Gu, Z. Liu, G. Panico et al., *A global view on the Higgs self-coupling at lepton colliders*, *Journal of High Energy Physics* **2018** (2018) 178 [1711.03978].

[448] A. Blondel and P. Janot, *Future strategies for the discovery and the precise measurement of the Higgs self coupling*, *arXiv e-prints* (2018) arXiv:1809.10041 [1809.10041].

[449] S. Chang and M.A. Luty, *The Higgs trilinear coupling and the scale of new physics*, *Journal of High Energy Physics* **2020** (2020) 140 [1902.05556].

[450] LHC Reinterpretation Forum collaboration, *Reinterpretation of LHC Results for New Physics: Status and Recommendations after Run 2*, *SciPost Phys.* **9** (2020) 022 [2003.07868].

[451] A. Carvalho, F. Goertz, K. Mimasu, M. Gouzevitch and A. Aggarwal, *On the reinterpretation of non-resonant searches for Higgs boson pairs*, *arXiv e-prints* (2017) arXiv:1710.08261 [1710.08261].

[452] A. Buckley, J. Butterworth, D. Grellscheid, H. Hoeth, L. Lonnblad, J. Monk et al., *Rivet user manual*, *Comput. Phys. Commun.* **184** (2013) 2803 [1003.0694].

[453] E. Conte, B. Fuks and G. Serret, *MadAnalysis 5, A User-Friendly Framework for Collider Phenomenology*, *Comput. Phys. Commun.* **184** (2013) 222 [1206.1599].

[454] P.T. Komiske, R. Mastandrea, E.M. Metodiev, P. Naik and J. Thaler, *Exploring the Space of Jets with CMS Open Data*, *Phys. Rev. D* **101** (2020) 034009 [1908.08542].

[455] J. Alison, S. An, P. Bryant, B. Burkle, S. Gleyzer, M. Narain et al., *End-to-end particle and event identification at the Large Hadron Collider with CMS Open Data*, in *Meeting of the Division of Particles and Fields of the American Physical Society*, 10, 2019 [1910.07029].

[456] P. Musella and F. Pandolfi, *Fast and Accurate Simulation of Particle*

Detectors Using Generative Adversarial Networks, *Comput. Softw. Big Sci.* **2** (2018) 8 [1805.00850].

[457] J. Bai, F. Lu, K. Zhang et al., "Onnx: Open neural network exchange." https://github.com/onnx/onnx, 2022.

[458] N. Berger et al., *Simplified Template Cross Sections - Stage 1.1*, 1906.02754.

[459] K. Cranmer, S. Kraml, H.B. Prosper, P. Bechtle, F.U. Bernlochner, I.M. Bloch et al., *Publishing statistical models: Getting the most out of particle physics experiments*, *SciPost Phys.* **12** (2022) 037.

[460] CMS collaboration, *Simplified likelihood for the re-interpretation of public CMS results*, Tech. Rep. , CERN, Geneva (2017).

[461] A. Coccaro, M. Pierini, L. Silvestrini and R. Torre, *The DNNLikelihood: enhancing likelihood distribution with Deep Learning*, *Eur. Phys. J. C* **80** (2020) 664 [1911.03305].

[462] E. Maguire, L. Heinrich and G. Watt, *HEPData: a repository for high energy physics data*, *J. Phys. Conf. Ser.* **898** (2017) 102006 [1704.05473].

[463] European Organization For Nuclear Research and OpenAIRE, *Zenodo*, 2013. 10.25495/7GXK-RD71.

[464] J. Brehmer, K. Cranmer, I. Espejo, F. Kling, G. Louppe and J. Pavez, *Effective LHC measurements with matrix elements and machine learning*, *J. Phys. Conf. Ser.* **1525** (2020) 012022 [1906.01578].

[465] M. Stoye, J. Brehmer, G. Louppe, J. Pavez and K. Cranmer, *Likelihood-free inference with an improved cross-entropy estimator*, 1808.00973.

[466] J. Brehmer, K. Cranmer, G. Louppe and J. Pavez, *Constraining Effective Field Theories with Machine Learning*, *Phys. Rev. Lett.* **121** (2018) 111801 [1805.00013].

[467] J. Brehmer, K. Cranmer, G. Louppe and J. Pavez, *A Guide to Constraining Effective Field Theories with Machine Learning*, *Phys. Rev. D* **98** (2018) 052004 [1805.00020].

[468] X. Chen, S. Dallmeier-Tiessen, R. Dasler, S. Feger, P. Fokianos, J.B. Gonzalez et al., *Open is not enough*, *Nature Physics* **15** (2019) 113.

[469] M.D. Hildreth, A. Boehnlein, K. Cranmer, S. Dallmeier, R. Gardner, T. Hacker et al., *HEP Software Foundation Community White Paper Working Group - Data and Software Preservation to Enable Reuse*, *arXiv e-prints* (2018) arXiv:1810.01191 [1810.01191].

[470] K. Cranmer and I. Yavin, *RECAST: Extending the Impact of Existing Analyses*, *JHEP* **04** (2011) 038 [1010.2506].

[471] J.S. Kim, D. Schmeier, J. Tattersall and K. Rolbiecki, *A framework to create*

customised LHC analyses within CheckMATE, *Comput. Phys. Commun.* **196** (2015) 535 [1503.01123].

[472] A. Kvellestad, P. Scott and M. White, *GAMBIT and its Application in the Search for Physics Beyond the Standard Model*, 1912.04079.

[473] G. Alguero, J. Heisig, C.K. Khosa, S. Kraml, S. Kulkarni, A. Lessa et al., *Constraining new physics with SModelS version 2*, *JHEP* **08** (2022) 068 [2112.00769].

[474] T. Šimko, L.A. Heinrich, C. Lange, A.E. Lintuluoto, D.M. MacDonell, A. Mečionis et al., *Scalable declarative hep analysis workflows for containerised compute clouds*, *Frontiers in Big Data* **4** (2021) .