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

Search for Higgs boson pair production in the $\mathrm{b\bar{b}}\ell\nu\ell\nu$ final state with the CMS detector

Doctoral dissertation presented by

Sébastien Wertz

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

Thesis evaluation committee

Prof. Florencia Canelli	UZH, Switzerland
Prof. Jorgen D'Hondt	VUB, Belgium
Prof. Christophe Delaere (Secretary)	UCL, Belgium
Prof. Jean-Marc Gérard (Chair)	UCL, Belgium
Prof. Vincent Lemaître (Supervisor)	UCL, Belgium

June, 2018

Remerciements

Le travail mené à bien durant ma thèse n'aurait pu l'être sans la présence et le soutien d'un grand nombre de personnes, et j'aimerais tenter ici d'exprimer ma gratitude envers ces dernières.

Tout d'abord, merci à Vincent, mon promoteur, pour avoir il y a cinq ans déjà accepté de me prendre sous ton aile et de m'initier au monde de la physique des particules. Ton enthousiasme à toute épreuve et ton sens critique aiguisé ont été des exemples à suivre tout au long de cette thèse. Merci également pour ta disponibilité et pour ta patience vis-à-vis d'un doctorant qui n'en fait qu'à sa tête!

Merci à Seb, mon ex-voisin de bureau, pour ton efficacité redoutable et pour ton amicale compagnie. Je remercie chaudement Alex, Brieuc, Christophe et Christophe, Jérôme, Martin, Michele, Miguel, Olivier, Pieter, sans oublier Alessia, Matthias et le petit nouveau Florian, pour vos qualités humaines, pour votre agréable présence durant les temps de midi, au CERN et même en dehors du cadre du travail, et pour cette si bonne ambiance au sein du groupe $\ell\ell b\bar{b}$ en particulier et à CP3 en général.

I would like to thank Florencia, Jorgen, Jean-Marc and again Christophe for accepting to be part of my thesis committee, for the interesting discussions during the private defence, as well as for your suggestions and comments that have improved the quality of this text.

Si les résultats présentés ici ont pu être obtenus en un temps fini, c'est aussi grâce au support technique sans faille fourni par Jérôme, Pavel et Andres, ainsi qu'à la multitude d'outils créés et maintenus par la communauté du logiciel libre. Merci à vous tous. Je n'oublie pas non plus Ginette et Carinne: merci pour votre soutien administratif efficace et compréhensif.

Merci à mes "vrais" amis, pour avoir réussi à garder le contact même après s'être dispersés aux quatre vents à la fin de nos études, et pour tous ces bons moments passés ensemble depuis lors.

Pour leur indéfectible soutien tout au long de mes études, je voudrais sincèrement remercier ma famille, en tout particulier maman, papa, JB, mamy et Tante Anna. À Amélie, merci pour ta patience et ta compréhension durant ces quatre années, et tout simplement pour ta présence dans ma vie.

Enfin, je suis redevable au FRS-FNRS pour l'octroi d'une bourse FRIA ayant permis la réalisation de cette thèse.

Contents

List of Figures

List of Tables

Introduction

The standard model (SM) of particle physics is humanity's best-achieving theory for the description of elementary particles and their interactions. Based on a moderate amount of hypotheses and inputs, it can be used to formulate countless and precise predictions, which have so far shown an impressive agreement with experimental data. The discovery of the Higgs boson by the ATLAS and CMS experiments at the Large Hadron Collider (LHC), almost 6 years ago, has spawned a scientific programme involving a detailed study of the Higgs boson's properties, such as its interactions with other particles of the SM. One of these properties, the interaction of the Higgs boson with itself, constitutes a crucial prediction of the SM that has as yet eluded experimental confirmation. The most direct method to characterise this interaction and thus to test the cornerstone of the SM, the mechanism of spontaneous symmetry breaking, is to observe and measure the simultaneous production of two Higgs bosons (HH) in the high-energy collisions of protons provided by the LHC. Unfortunately, HH production is a process so rare that its discovery is not expected before at least two decades.

Despite its tremendous successes, the SM suffers from a number of shortcomings that have led to the development of candidate models postulating the existence of new particles and interactions. This "New Physics" could lead to noticeable deviations from SM predictions, which can be actively searched for using two complementary approaches. If new states are light enough that they can be frequently produced in LHC collisions and if they then decay to SM particles, they could be discovered through the study of these decay products. Numerous models of New Physics involve resonances that may decay to pairs of Higgs bosons, enhancing the rate of HH production to observable levels. On the other hand, these new states may well be too massive and out of direct reach of the LHC. Their existence could still be visible through indirect, nonresonant effects they imprint on SM processes at lower energies. These effects can be parameterised in a rather generic manner, without having to rely on too numerous assumptions. The HH process is highly sensitive to these effects, so that also in this situation an early discovery of Higgs boson pairs at the LHC is conceivable.

Higgs boson pair production is buried deep inside the data collected by LHC experiments. In order to increase our chances of observing it, our best bet is to consider several experimental channels corresponding to different decay modes of the pair-produced bosons. In this thesis, we have targeted the case where one Higgs boson decays to a pair of bottom quarks, and the other Higgs boson decays to two vector bosons (W or Z bosons) which themselves yield two charged leptons (such as electrons or muons) and two neutrinos. We have analysed a sample of events collected by the CMS detector using proton-proton collisions at a centre-of-mass energy of 13 TeV, with the aim of observing the production of Higgs boson pairs in that final state. The events contain a pair of reconstructed charged leptons (electrons or muons) and a pair of b-tagged jets. To reduce

the contamination due to abundant SM background processes, such as the production and decay of top quark pairs or Z bosons in association with jets, we have constructed and applied for the first time parameterised neural networks trained to recognise signal from background events. The invariant mass of the selected jets as well as the score of these multivariate classifiers provide a powerful signature with which to probe the resonant and nonresonant production of Higgs boson pairs. Based on these observables, we have reported the agreement between the data and SM predictions as a function of the hypothesised resonance mass or as a function of parameters encoding possible deviations from SM predictions of the Higgs boson's couplings, namely the strength of the Higgs boson self-coupling and the strength of the Higgs boson's interaction with the top quark.

The results presented in this thesis have been published in the following paper:

"*Search for resonant and nonresonant Higgs boson pair production in the* $\mathrm{b\bar{b}}\ell\nu\ell\nu$ *final state in proton-proton collisions at* $\sqrt{s} = 13$ *TeV,*" JHEP **1801** (2018) 054,

and have been documented by these internal notes:

- "*Search for resonant production of two Higgs bosons in the* $\mathrm{b\bar{b}}\ell\nu\ell\nu$ *final state in 2016 data,*" CMS AN-2016/444
- "*Search for production of two Higgs bosons in the* $\mathrm{b\bar{b}}\ell\nu\ell\nu$ *final state in 2016 data,*" CMS AN-2016/430

Moreover, intermediate results obtained in the same final state but not described in the present thesis have been described in these public and internal notes:

- "*Search for Higgs boson pair production in the* $\mathrm{b\bar{b}}\ell\nu\ell\nu$ *final state at* $\sqrt{s} = 13$ TeV," CMS-PAS-HIG-16-024
- "*Search for production of two Higgs bosons in the* $\mathrm{b\bar{b}}\ell\nu\ell\nu$ *final state,*" CMS AN-2016/177

We have organised this text as follows. First, the theoretical framework of the SM as well as the methods used to generate quantitative predictions are introduced, and we portray the process of Higgs pair production in the SM. We then outline the effective field theory approach for describing indirect effects on HH production due to New Physics at energies not directly attainable at the LHC, and show how these effects can be efficiently modelled. Different models predicting the resonant production of Higgs boson pairs are also listed. We conclude the chapter by giving a brief overview of the current experimental results on HH production. In the second chapter, we describe the CMS detector and the algorithms used to reconstruct and identify the different particles produced in proton collisions, relevant for the process in which we are interested. The simulation of the detector and its calibration, as well as the machine-learning and statistical tools instrumental in our work are then detailed. The third chapter is dedicated to our analysis of CMS data. We start by describing the event selection procedure and the methods used for background estimation, in particular the dedicated technique developed to model the production of Z bosons in association with b jets. We then report on the multivariate techniques employed to increase our sensitivity to the signal, the systematic uncertainties that affect the interpretation of the data, and the extraction of the results. Finally, in the fourth and last chapter we point to a few ideas that might be pursued in case this work is repeated with the larger amounts of data that are becoming available, and give our conclusions.

Chapter

1. Higgs boson pair production in the Standard Model and Beyond

In this chapter, we describe the standard model (SM) of particle physics, which is based on the general framework of quantum field theory (QFT). Since a theory is not of much use if it cannot generate predictions that can be confronted with experiments, we will review some of the methods and tools used to formulate predictions such as scattering and decay rates of particles. Next, the process of double Higgs production (HH) in proton-proton collisions is introduced. We then show different situations in which HH might deviate from SM predictions, and conclude with a short review of the current experimental results on double Higgs production.

Throughout this chapter, we follow the notations and conventions adopted by Peskin and Schroeder [1]. In particular, we denote space-time coordinates and momentum by $x = x^\mu = (t, \boldsymbol{x})$ and $p = p^\mu = (E, \boldsymbol{p})$, respectively, and define $\partial_\mu \equiv \partial / \partial x^\mu$. The metric signature used is $(+,-,-,-)$, so that a free particle of mass m with momentum p satisfies $p^2 = m^2$. We denote generically by ϕ a scalar field and by ψ a Dirac spinor, with its conjugate written as $\overline{\psi} \equiv \psi^\dagger \gamma^0$. The γ^μ are the Dirac matrices, for which we use the Weyl representation; as usual we have $\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3$. Symbols for specific particles (such as ν, e, H, ...) are either used to denote the particle or the associated (scalar, spinor or vector) field; the meaning can be inferred from context.

We introduce here some geometrical conventions that will be used in this and the next chapters. As our experimental methodology relies on head-on collisions of protons, we choose a right-handed orthonormal coordinate system centered on the collision point, with its z-axis pointing in the direction of one of the incoming protons, i.e. along the beam axis. The momentum $\boldsymbol{p} = (p_x, p_y, p_z)$ of a particle can be written in polar coordinates as $\boldsymbol{p} = |p| \cdot (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$, where θ is a polar angle measured from the z-axis and ϕ is the azimuth of the projection of $\boldsymbol{p}$ on the x-y-plane, measured from the x-axis. The transverse momentum p_T is defined as the magnitude of the projection of $\boldsymbol{p}$ on the x-y-plane, $p_\mathrm{T} = |p| \sin\theta$. The rapidity of a particle is defined as:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right). \tag{1.1.}$$

Differences in y are invariant under Lorentz boosts along the z direction. The quantities p_T and ϕ are also invariant under such transformations, and so is the angular distance ΔR between two particles:

$$\Delta R = \sqrt{(y_1 - y_2)^2 + (\Delta\phi)^2}, \text{ where} \tag{1.2.}$$

$$\Delta\phi = \min\left(|\phi_1 - \phi_2|, 2\pi - |\phi_1 - \phi_2|\right). \tag{1.3.}$$

Two different unit systems are used in this work. The first is the International System of units (SI), and is only used for quantities where SI units, i.e. m, kg, s, A, K and derived units, are quoted explicitly. The second is a natural system based on the rationalized Lorentz-Heaviside system, obtained by defining $c = \epsilon_0 = \mu_0 = k_B = \hbar = 1$. In that system, quantities can be expressed in terms of some powers of the unit of energy, chosen to be the eV:

- [mass] = [momentum] = [energy] = [temperature] = eV
- [length] = [time] = eV^{-1}

Equations written in this chapter are understood to use this latter system, and so are all quantities given in explicit units of eV.

1.1. The Standard Model of particle physics

The content of this section is mostly inspired by Refs. [1–4]. We start by briefly reviewing the general principles of QFTs, defining observable quantities such as cross sections, before describing the particles and interactions in the SM itself.

1.1.1. Quantum Field Theory and observables

Quantum Field Theory is the result of the union between quantum mechanics and special relativity, more specifically the theory of relativistic fields. Given a *free* field $\phi(x)$, where $x = (t, \boldsymbol{x})$ denotes space-time coordinates, we consider the Lagrangian *density* $\mathcal{L}(\phi(x), \partial_\mu\phi(x))$. From the principle of least action $\delta\mathcal{S} = 0$, with $\mathcal{S} = \int \mathrm{d}^4x\,\mathcal{L}$, we obtain the Euler-Lagrange equations of motions for the field:

$$\partial_\mu \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} - \frac{\partial\mathcal{L}}{\partial\phi} = 0 \tag{1.4.}$$

For instance, the Lagrangian density and equation of motion for a free fermionic Dirac field ψ of mass m read:

$$\mathcal{L}_{\text{Dirac}} = \overline{\psi}(i\gamma^\mu\partial_\mu - m)\psi, \tag{1.5.}$$

$$(i\gamma^\mu\partial_\mu - m)\psi = \overline{\psi}(i\gamma^\mu\partial_\mu + m) = 0. \tag{1.6.}$$

The free field can be *quantised* by defining *creation* and *annihilation* operators $a^\dagger_{\boldsymbol{p}}$ and $a_{\boldsymbol{p}}$, function of momentum $\boldsymbol{p}$, in the spirit of the treatment of the quantum harmonic oscillator. Eigenstates of the free particle Hamiltonian, corresponding to a collection of n particles of momentum $\boldsymbol{p}$, are then specified from the number of times $n_{\boldsymbol{p}}$ a creation operator has acted on the vacuum $|0\rangle$: $|n_{\boldsymbol{p}_1} n_{\boldsymbol{p}_2} \ldots\rangle$. The field, thus now promoted to an operator, can be decomposed into Fourier modes, each of which is treated as an independent oscillator:

$$\phi(x) = \int \frac{\mathrm{d}^3p}{(2\pi)^3} \frac{1}{\sqrt{2E(\boldsymbol{p})}} \left(a_{\boldsymbol{p}} e^{-ip\cdot x} + a^\dagger_{\boldsymbol{p}} e^{ip\cdot x}\right) \tag{1.7.}$$

Allowing the fields to interact, we work in the *interaction picture* where the Hamiltonian obtained from the above Lagrangian is divided into the free and interacting parts, $H = H_0 + H_I$, and the time-dependence of the state vector is only due to the latter:

$$\frac{\mathrm{d}\,|\Psi(t)\rangle}{\mathrm{d}t} = iH_I(t)\,|\Psi(t)\rangle\,, \qquad H_I(t) = e^{iH_0(t-t_0)} H_I e^{-iH_0(t-t_0)}. \tag{1.8.}$$

Since we are interested in computing cross sections or decay rates, we need to express the transition probability between an initial state $|i\rangle$ and a final state $\big|f\big\rangle$, corresponding to collections of non-interacting particles with well-defined momenta $\boldsymbol{p}_i^{\text{in}}$ and $\boldsymbol{p}_i^{\text{out}}$ in the far past and future, respectively. These probabilities can be expressed in terms of the $\mathcal{S}$-matrix, which encodes the time-evolution of the initial state: $|\Psi(t \to +\infty)\rangle = \mathcal{S}\,|\Psi(t \to -\infty)\rangle = \mathcal{S}\,|i\rangle$. Hence, the transition amplitude to the considered final state $\big|f\big\rangle$ is obtained by the projection $\big\langle f\big|\mathcal{S}\big|i\big\rangle$. Using (1.8) and provided the interaction H_I is "small", the $\mathcal{S}$-matrix can be written as a perturbative series, the Dyson expansion:

$$\mathcal{S} = \sum_{n=0}^{\infty} \frac{(-i)^n}{n!} \int \cdots \int \mathrm{d}^4x_1 \ldots \mathrm{d}^4x_n\, \mathrm{T}\left\{\mathcal{H}_I(x_1) \ldots \mathcal{H}_I(x_n)\right\}, \tag{1.9.}$$

where $\mathcal{H}_I$ is the interacting Hamiltonian's density and T denotes the time-ordered product. By virtue of Wick's theorem, every term in the series (1.9) can be expressed as a finite sum of normal products, from which amplitudes may now be computed. These amplitudes can be represented by Feynman diagrams, which provide an easy way to identify the terms that contribute to a given process at a given order in the perturbative expansion.

As we will concern ourselves with scattering experiments, we will need to compute scattering cross sections. Given two beams of particles meeting head-on, or a beam of particles meeting a target at rest, the cross section is defined as the rate of scattering *events* of a given type, divided by the incident particle flux. It has dimensions of area and is directly related to the interaction probability between these particles, without reference to the details of the beams. More generally, we will be interested in differential cross sections, $\mathrm{d}\sigma(X)/\mathrm{d}X$, where the counting is carried out as a function of some quantity X that can be expressed in terms of the outcoming momenta, $\{p_f^{\text{out}}\}$. Writing

$$\mathcal{S} = \mathbf{1} + (2\pi)^4 \delta^4\left(p_1^{\text{in}} + p_2^{\text{in}} - \sum p_f^{\text{out}}\right) \cdot i\mathcal{M}\left(p_{1,2}^{\text{in}} \to \left\{p_f^{\text{out}}\right\}\right) \tag{1.10.}$$

to isolate the part of the $\mathcal{S}$-matrix describing the actual interactions between particles into $\mathcal{M}$, the matrix element (ME), we can compute the *fully* differential cross section as function of all final state momenta as:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Phi} = \frac{\left|\mathcal{M}\left(p_{1,2}^{\text{in}} \to \left\{p_f^{\text{out}}\right\}\right)\right|^2}{F}, \tag{1.11.}$$

where the phase-space differential and the flux factor are given by:

$$\mathrm{d}\Phi = (2\pi)^4 \delta^4 \left(p_1^{\mathrm{in}} + p_2^{\mathrm{in}} - \sum p_f^{\mathrm{out}} \right) \prod_f \frac{\mathrm{d}^3 \boldsymbol{p}_f^{\mathrm{out}}}{(2\pi)^3} \frac{1}{2E_f^{\mathrm{out}}}, \tag{1.12.}$$

$$F = 4\sqrt{(p_1^{\mathrm{in}} \cdot p_2^{\mathrm{in}})^2 - (p_1^{\mathrm{in}})^2 (p_2^{\mathrm{in}})^2} = 4E_1^{\mathrm{in}} E_2^{\mathrm{in}} \Delta v. \tag{1.13.}$$

In (1.13), Δv is the relative velocity between the incident particles. This expression for F shows that the cross section is invariant under Lorentz boosts along the direction of the beam(s). Total or differential cross sections can be obtained from (1.11) by integrating over all or part of the phase space. Differential or total decay rates, defined as the rate at which unstable particles at rest decay to a given final state, can be computed in a similar fashion by considering a single initial particle of mass M, and changing the flux factor (1.13) to $F = 2M$.

The $\mathcal{S}$-matrix can be obtained as a perturbative expansion, hence so do cross sections. Assuming the strength of the interaction is governed by some *coupling constant* g, defining $\alpha = g^2/(4\pi)$ the expansion can be written as:

$$\sigma = \sigma_{\mathrm{LO}} \cdot \left(1 + \frac{\alpha}{2\pi} \sigma_1 + \left(\frac{\alpha}{2\pi} \right)^2 \sigma_2 + \dots \right). \tag{1.14.}$$

We see that the interaction had better be weak ($\alpha \ll 1$) if σ is to be computed perturbatively. The lowest-order computation is referred to as leading order (LO); including higher-order terms in (1.14) will result in more accurate predictions but is generally a challenging task. Using one or two additional terms is referred to as next-to-leading order (NLO) or next-to-next-to-leading order (NNLO), and so on.

The calculation of terms beyond LO involves diagrams containing *loops* through which arbitrarily high momenta can flow. This leads to so-called ultraviolet divergences in the amplitudes, which can fortunately be absorbed as redefinitions of the fields and constants present in the Lagrangian, in a procedure dubbed *renormalisation*. The requirement of renormalisability, which guarantees that the theory is insensitive to unknown phenomena at very high energies and that well-defined predictions can be extracted at experimentally accessible scales, imposes strong constraints on the possible interactions we might consider. In particular, the coupling constants need to have dimensions of energy to a power greater or equal than zero.

Through renormalisation, the "bare" coupling constants used in the expansion (1.14) are replaced by "constants" which have acquired a dependence on an unphysical energy, the *renormalisation scale* μ_R: $\alpha \to \alpha(\mu_R)$. The "running" of the coupling constant with the scale is given by $\mathrm{d}\alpha/\mathrm{d}(\log \mu^2) = \beta(\alpha)$, where the Beta function β can be computed perturbatively. The exact (nonperturbative) cross section σ is independent of μ_R, but the truncation of the series (1.14) leads to a dependence of predictions on that scale, creating an intrinsic uncertainty in every perturbative calculation. This uncertainty is generally estimated by varying μ_R by a factor of two around some central value taken to be a typical energy scale of the process at hand, however let us stress here that this procedure is *ad-hoc* and often optimistic in estimating the size of missing higher-order contributions. Including additional terms in the expansion will result in a reduced dependence on μ_R, i.e. more precise predictions.

1.1.2. Particles and symmetries in the Standard Model

In the current state of affairs, we have no way to predetermine the content of the Lagrangian in terms of elementary particles, but have to rely on experimental data. A particle is considered elementary (point-like) as long as is has not exhibited any internal structure. Particles possess a number of properties such as mass, spin, charges, which can be used to classify them. Matter particles have spin $1/2$ and can be divided into *leptons* and *quarks*. Leptons comprise neutrinos, which are electrically neutral and have extremely small but non-zero mass, and "charged leptons" such as the electron and its heavier copies, which have a charge of -1 in units of e. Quarks differ from leptons in that they partake in the strong interaction, and that they have fractional electric charges $2/3$ or $-1/3$.

As it turns out, leptons and quarks come in three *generations* of particles, differing only by their mass. No additional generation has been found so far, and it is unknown why there should be exactly three of them. Only the lightest quarks, "up" and "down", together with the electron make out all ordinary visible matter. The species of the six different leptons and six different quarks is referred to as *flavour*.

All (charged) matter particles can be described as Dirac fermions, and hence come in both particles and *anti-particles* which have opposite charges but appear otherwise identical. Neutrinos being neutral and massive, they could either be Dirac or Majorana fermions, a possibility that is still being investigated.

The known elementary matter particles are listed in Tab. 1.1.

The concept of symmetry plays a central role in the construction of the SM. To begin with, the requirement that the theory obeys special relativity implies that the action should be invariant under translations, rotations and boosts, which together make out the Poincaré symmetry group. Each particle, or field, is thus embedded in a particular irreducible representation of the Poincaré group, characterised by Casimir invariants corresponding to mass and spin.

While the Poincaré group consists of *global* transformations, acting identically at every point of space-time, the description of *interactions* between elementary particles can be achieved by imposing invariance under *local* transformations acting on internal degrees of freedom of the fields. Indeed, the theory can be made invariant under such local *gauge* transformations if we introduce additional fields which themselves transform under the adjoint representation of said gauge symmetry group. Each generator of the group then corresponds to a *gauge boson*, a particle of spin one, which can be seen as a mediator of the corresponding interaction between matter particles. If the group is non-abelian, the gauge bosons will also interact *with themselves* via three- and four-point vertices, whose strengths are all related the same coupling constant thanks to gauge invariance.

The gauge bosons of the SM are listed in Tab. 1.2. The mass of these bosons is directly related with the range of the interaction. For instance, the photon being massless, electromagnetism has infinite range.

Once the symmetries of the theory and the matter content have been specified, every possible combination of fields allowed by these symmetries and leading to a renormalisable theory *must* be included into the Lagrangian. Moreover, finding out which are the underlying symmetries of the theory can only be achieved through experimentation.

Table 1.1. | Summary of leptons and quarks in the SM. Masses are taken from Ref. [5]. Charged lepton masses are quoted with four significant digits, but are known to much better precision (10 significant digits for e and μ; 5 for τ). Electric charges are given as multiples of the absolute charge of the electron. Note that for neutrinos and quarks, mass eigenstates do not correspond to flavour eigenstates. Only upper limits on neutrino masses can be given, reported here using the upper limit on the sum of stable neutrinos obtained from cosmological measurements [6]. At least two neutrino states must have non-zero mass. Quark masses are given as running masses in the $\overline{\text{MS}}$ scheme, except for the top quark, for which the direct measurement is used.

	Generation	1	2	3
Leptons	Name, symbol	electron neutrino, ν_e	muon neutrino, ν_μ	tau neutrino, ν_τ
	Mass	$\longleftarrow \sum \nu_i < 0.0926\,\text{eV}$ @ 90% C.L. $\longrightarrow$		
	Electric charge	0	0	0
	Name, symbol	electron, e	muon, μ	tau, τ
	Mass	511.0 keV	105.7 MeV	1.777 GeV
	Electric charge	-1	-1	-1
Quarks	Name, symbol	up, u	charm, c	top, t
	Mass	$2.2^{+0.6}_{-0.4}$ MeV	1.28(3) GeV	173.5(6) GeV
	Electric charge	$^2/_3$	$^2/_3$	$^2/_3$
	Name, symbol	down, d	strange, s	bottom, b
	Mass	$4.7^{+0.5}_{-0.4}$ MeV	96^{+8}_{-4} MeV	$4.18^{+0.04}_{-0.03}$ GeV
	Electric charge	$^{-1}/_3$	$^{-1}/_3$	$^{-1}/_3$

Table 1.2. | The gauge bosons in the SM and the interaction they carry. Their masses are taken from Ref. [5]; for the gluon only the theoretical mass is given.

Name	Mass	Interaction
Photon, γ	0 ($< 1\times10^{-18}$ eV)	Electromagnetism
Z boson	91.1876(21) GeV	Weak interaction
$W^\pm$ bosons	80.385(15) GeV	
Gluons, g	0	Strong interaction

In the context of symmetries, it is important to mention Noether's theorem, which states that for every continuous symmetry there is a corresponding conservation law. The conserved quantities in the case of the Poincaré group are energy, momentum and angular momentum, whereas for the gauge symmetries these are various charges such as the electric charge.

Quantising a gauge theory leads to subtleties into which we will not enter, such as gauge fixing or the introduction of ghost fields. Suffice to mention that the formulation of interactions in terms of gauge symmetry has desirable consequences, such as the guarantee that the theory is renormalisable. However, it also implies that gauge bosons should be massless, since a mass term would violate gauge invariance. This inconsistency with experimental data, which clearly show that the bosons of the weak interaction are massive, has been resolved thanks to the concept of spontaneous symmetry breaking detailed in the next section.

1.1.3. Electroweak interactions and Higgs mechanism

Electroweak theory has developed from the union of quantum electrodynamics (QED), the quantum field-theoretic description of electromagnetism, and Fermi theory [7], in which weak interactions are modelled using operators of the type

$$\mathcal{L}_{\text{Fermi}} \supset \frac{G_F}{\sqrt{2}} \left(\overline{\nu}\gamma_\mu(1-\gamma_5)\ell\right)\left(\overline{\ell}\gamma^\mu(1-\gamma_5)\nu\right) + \text{h.c.}, \tag{1.15.}$$

where $G_F \approx 1.166 \times 10^{-5}\,\text{GeV}^{-2}$ is the Fermi constant. Similar terms can be written for the quarks but shall not be explicited here. The vector and axial coupling structure "V-A"$= \gamma_\mu(1-\gamma_5)$ accounts for the observation that weak interactions violate parity, as shown by the Wu experiment [8] or by the $\tau - \theta$ puzzle. Fermi theory successfully describes phenomena such as Beta decay of the muon and neutron, low-energy neutrino-electron scattering, and various meson decays. However, such a model is bound to fail since the four-fermion interaction (1.15) is of dimension six (the coupling G_F being of dimension -2). It is hence not renormalisable, and to make things worse it violates unitarity bounds for processes such as $\nu\text{e} \to \nu\text{e}$ scattering at energies of $\approx 100\,\text{GeV}$.

From these and other considerations such as the discovery of neutral-current weak interactions by the Gargamelle experiment [9], the theory can be rephrased by first defining the left- and right-handed projections of the fermion fields:

$$P_{L,R} = \frac{1}{2}\left(1 \mp \gamma_5\right), \qquad \psi_{L,R} \equiv P_{L,R}\,\psi. \tag{1.16.}$$

In the ultra-relativistic limit, these chiral states identify with helicity eigenstates, left- or right-handed fermions having left- or right-handed helicity (and conversely for anti-particles). As can be seen from (1.15), the weak interaction only couples to the left-handed quarks and leptons:

$$\boldsymbol{L} = \left[\begin{pmatrix}\nu_{\text{e}L}\\ \text{e}_L\end{pmatrix}, \begin{pmatrix}\nu_{\mu L}\\ \mu_L\end{pmatrix}, \begin{pmatrix}\nu_{\tau L}\\ \tau_L\end{pmatrix}\right], \; \boldsymbol{Q} = \left[\begin{pmatrix}\text{u}_L\\ \text{d}_L\end{pmatrix}, \begin{pmatrix}\text{c}_L\\ \text{s}_L\end{pmatrix}, \begin{pmatrix}\text{t}_L\\ \text{b}_L\end{pmatrix}\right], \tag{1.17.}$$

which have been placed in doublets of the SU(2) group, while the right-handed fermions

are kept in singlets:

$$\boldsymbol{e} = \left[\mathrm{e}_R, \mu_R, \tau_R\right], \ \boldsymbol{u} = \left[\mathrm{u}_R, \mathrm{c}_R, \mathrm{t}_R\right], \ \boldsymbol{d} = \left[\mathrm{d}_R, \mathrm{s}_R, \mathrm{b}_R\right]. \tag{1.18.}$$

Note that we do not introduce right-handed neutrinos; this possibility will be briefly discussed later in the context of neutrino masses.

The theory can then be gauged through the introduction of mediator fields $\mathbf{W}_\mu$ transforming under the adjoint representation of SU(2). As already mentioned, this construction however contradicts the observation that the weak interaction is short-ranged. Furthermore, fermions are now bound to remain massless, since Dirac mass terms such as $\overline{\mathrm{e}}_R \mathrm{e}_L$ are not invariant under SU(2) rotations of the doublets defined in (1.17).

These issues were resolved in the Glashow–Weinberg–Salam model [10–13], building on the notion of spontaneous breaking of gauge symmetry introduced independently by Brout and Englert [14], Higgs [15,16], and Hagen, Guralnik and Kibble [17]. In this model, the above symmetry group is extended as $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$, where Y is the *weak hypercharge* and $\mathrm{U}(1)_Y$ is gauged using the mediator B_μ. The fermions are coupled to the bosons as:

$$\mathcal{L}_{\text{fermion}} = \sum_{i,\psi} \overline{\psi}_i (i\gamma_\mu D^\mu)\psi_i, \qquad i = 1,2,3;\ \psi = \boldsymbol{L}, \boldsymbol{Q} \tag{1.19.}$$

$$+ \sum_{i,\psi} \overline{\psi}_i (i\gamma_\mu D'^\mu)\psi_i, \qquad i = 1,2,3;\ \psi = \boldsymbol{e}, \boldsymbol{u}, \boldsymbol{d} \tag{1.20.}$$

The covariant derivatives are given by:

$$D^\mu = \partial^\mu - ig\frac{\boldsymbol{\sigma}}{2}\cdot\mathbf{W}^\mu - ig'\frac{Y}{2}B^\mu \tag{1.21.}$$

$$D'^\mu = \partial^\mu - ig'\frac{Y}{2}B^\mu, \tag{1.22.}$$

where two coupling constants g and g' are introduced. The generators of $\mathrm{SU}(2)_L$ are represented as the Pauli matrices $\boldsymbol{\sigma} = (\sigma^1, \sigma^2, \sigma^3)$, satisfying $[\sigma^i, \sigma^j] = i\epsilon^{ijk}\sigma^k$, with ϵ^{ijk} the structure constants of SU(2) (summations over repeating indices are understood). The normalisations of the generators Y are tuned for each field so as to recover the observed electric charges when defining the charge operator as $Q = (\sigma_3 + Y)/2$.

The Lagrangian also contains kinetic and interaction terms for the gauge bosons:

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4}\mathbf{W}_{\mu\nu}\mathbf{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}, \tag{1.23.}$$

where $\mathbf{W}_{\mu\nu}$ and $B_{\mu\nu}$ are the field strength tensors obtained from $\mathbf{W}_\mu$ and B_μ, respectively:

$$\mathrm{W}^i_{\mu\nu} = \partial_\mu \mathrm{W}^i_\nu - \partial_\nu \mathrm{W}^i_\mu - g\epsilon^{ijk}\mathrm{W}^j_\mu \mathrm{W}^k_\nu, \tag{1.24.}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{1.25.}$$

In addition, a complex *scalar* doublet of $\mathrm{SU}(2)_L$, with weak hypercharge $Y = +1$, is introduced:

$$\phi = \begin{pmatrix} \phi^+ & \phi^0 \end{pmatrix}^T \tag{1.26.}$$

$$\mathcal{L}_{\text{scalar}} = D_\mu \phi^\dagger D^\mu \phi - V(\phi^\dagger \phi),\ V(\phi^\dagger \phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2, \tag{1.27.}$$

where V is the most general potential under the requirements of gauge invariance and renormalisability. The scalar is allowed to interact with the fermions *via* Yukawa terms:

$$\mathcal{L}_{\text{Yukawa}} = -\sum_i Y_e^{ij} \left(\overline{L}^i \cdot \phi \right) e^j + \text{h.c.} \qquad \text{(leptons)} \tag{1.28.}$$

$$-\sum_{i,j} Y_u^{ij} \left(\overline{Q}^i \cdot \phi^C \right) u^j - \sum_{i,j} Y_d^{ij} \left(\overline{Q}^i \cdot \phi \right) d^j + \text{h.c.}, \qquad \text{(quarks)} \tag{1.29.}$$

with $\phi^C = i\sigma_2 \phi^*$. The $\boldsymbol{Y}_e$, $\boldsymbol{Y}_u$ and $\boldsymbol{Y}_d$ are the Yukawa couplings for the leptons, up- and down-type quarks, respectively, and are matrices in flavour space.

As long as $\mu^2, \lambda > 0$, the scalar potential has a minimum at 0 and the $\text{SU}(2)_L \times \text{U}(1)_Y$ symmetry is manifest. However, if $\mu^2 < 0$, the potential acquires a non-trivial minimum with a vacuum expectation value (VEV) at $\langle 0|\phi^\dagger \phi|0\rangle = v^2 = -\mu^2/\lambda$, and the vacuum state is then no longer invariant under $\text{SU}(2)_L \times \text{U}(1)_Y$. The symmetry is said to be *spontaneously broken*, but it is rather *hidden* in a delicate interplay between various terms in the Lagrangian. The physical consequences can be made explicit by expanding the field around the new VEV and choosing a polar parameterisation:

$$\phi(x) = \frac{1}{\sqrt{2}} \exp\left(i\boldsymbol{\sigma} \cdot \boldsymbol{\xi}(x)/v\right) \begin{pmatrix} 0 \\ v + \text{H}(x) \end{pmatrix}. \tag{1.30.}$$

In (1.30), the $\boldsymbol{\xi}$ fields are *would-be* Goldstone bosons. These can be removed by performing an $\text{SU}(2)_L$ rotation, so that in *unitary gauge* we have:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + \text{H}(x) \end{pmatrix}. \tag{1.31.}$$

Inserting (1.31) back into the Lagrangian, we obtain i.a. mixing terms between the $\mathbf{W}_\mu$ and B_μ fields. From the chosen vacuum state and the definition of Q, it is clear that there remains an unbroken symmetry group which can be identified with the $\text{U}(1)_{\text{EM}}$ group of QED. This is made manifest by mass-diagonalising the gauge bosons as

$$\begin{pmatrix} \text{Z}_\mu \\ \text{A}_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_w & -\sin\theta_w \\ \sin\theta_w & \cos\theta_w \end{pmatrix} \begin{pmatrix} \text{W}_\mu^3 \\ B_\mu \end{pmatrix},\ \tan\theta_w = \frac{g'}{g},\ \text{W}_\mu^\pm = \frac{1}{\sqrt{2}} \left(\text{W}_\mu^1 \mp i\text{W}_\mu^2 \right), \tag{1.32.}$$

where θ_w is the so-called weak mixing angle. Using (1.31) and (1.32), the various interaction and mass terms can be read from the Lagrangian, featuring:

- charged W bosons with mass $m_{\text{W}} = gv/2$,
- a neutral Z boson with mass $m_{\text{Z}} = m_{\text{W}}/\cos\theta_w$,
- a neutral and massless *photon* A associated with the unbroken group $\text{U}(1)_{\text{EM}}$, whose

coupling constant $e = g\sin\theta_w = g'\cos\theta_w$ is precisely the electromagnetic coupling constant.

Hence, out of the four degrees of freedom of ϕ, three are the pseudo-Goldstone bosons and have been "absorbed" by the W and Z bosons, providing them with the longitudinal polarisation needed to make them massive. The remaining degree of freedom corresponds to a scalar particle, the *Higgs boson* H. The Higgs boson interacts with the massive gauge bosons, with couplings proportional to the squared boson masses, and restores unitarity in $VV \to VV$ scattering reactions ($V = W, Z$). In addition, the Higgs is endowed with a dynamics of its own, given by the Lagrangian:

$$\mathcal{L}_{\text{Higgs}} = \frac{1}{2}\partial_\mu \text{H}\partial^\mu \text{H} - \mu^2\text{H}^2 - \lambda v\text{H}^3 - \frac{\lambda}{4}\text{H}^4 \tag{1.33.}$$

In particular, (1.33) yields:

- a Higgs boson mass, $m_{\text{H}} = \sqrt{-2\mu^2} = \sqrt{2v^2\lambda}$ measured as 125.09(24) GeV [18],
- a cubic or trilinear self-coupling λv, to which we shall return later,
- and a quartic self-coupling λ.

The Fermi interactions introduced in (1.15) can now be retrieved as a low-energy limit of the above model. Indeed, the four-fermion interaction is promoted to a W boson exchange, but if the momentum transfer k in the reaction is small compared to the W mass, the amplitude satisfies:

$$\frac{g}{2\sqrt{2}}\mathcal{J}_\mu^\dagger \frac{g^{\mu\nu} - k^\mu k^\nu/m_{\text{W}}^2}{k^2 - m_{\text{W}}^2}\frac{g}{2\sqrt{2}}\mathcal{J}_\nu \xrightarrow[k^2 \ll m_{\text{W}}^2]{} \frac{g^2}{8m_{\text{W}}^2}\mathcal{J}_\mu^\dagger \mathcal{J}^\mu, \tag{1.34.}$$

with $\mathcal{J}_\mu = \overline{\ell}\gamma_\mu(1-\gamma_5)\nu$. Thus, we have:

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8m_{\text{W}}^2} = \frac{1}{2v^2}, \tag{1.35.}$$

and we deduce $v \approx 246\text{GeV}$. The Fermi interaction (1.15) is called an *effective operator*, and its coupling is inversely proportional to the scale at which the new degrees of freedom (weak and Higgs bosons) appear. Knowing the values of G_F, $\alpha(m_{\text{Z}}) = e^2/(4\pi) \approx 128^{-1}$, m_{Z} and m_{H}, we get:

$$\theta_w \approx 0.5,\ \mu^2 \approx -(88\,\text{GeV})^2,\ \lambda \approx 0.13 \tag{1.36.}$$

Crucially, while the symmetry breaking VEV can be retrieved from the Fermi constant, it is the knowledge of the Higgs boson mass that completely fixes the shape of the scalar potential, through the relations between μ^2, λ and v.

To conclude this section, let us return to the fermion Yukawa coupling matrices introduced above. There is only one such matrix for leptons, and we are free to rotate them in flavour space so as to diagonalise it. After symmetry breaking, by using (1.31) we obtain for each charged lepton ℓ a term:

$$\mathcal{L}_{\text{Yukawa}} \supset -\frac{Y_\ell v}{\sqrt{2}} \left(\overline{\ell}_L \ell_R + \overline{\ell}_R \ell_L \right) \left(1 + \frac{\mathrm{H}}{v} \right) \tag{1.37.}$$

Thus, spontaneous symmetry breaking also provides a mass $m_\ell = Y_\ell v/\sqrt{2}$ to the leptons, as well as residual Yukawa interactions between the leptons and the Higgs boson, with strength proportional to their mass. The story is similar for quarks, with the additional subtlety that there are two quark Yukawa matrices, $\boldsymbol{Y}_u$ and $\boldsymbol{Y}_d$. Since gauge symmetry forces us to rotate $\boldsymbol{Q}$ as a whole, it is impossible to simultaneously diagonalise both matrices, and the mass eigenstates will be different from the flavour states. We can still choose to rotate the up-type quarks to their mass eigenstates, and this fixes the relation between mass and flavour states for the down-type quarks. This relation is given by a unitary matrix, the Cabibbo–Kobayashi–Maskawa (CKM) matrix. By writing the Lagrangian in terms of mass eigenstates for both up- and down-type quarks, we obtain flavour-diagonal terms for the Yukawa interactions with the Higgs boson and the interaction with the Z boson and the photon. However, the charged-current interactions with the W bosons are modified, and now allow transitions between up-type quarks of one generation and down-type quarks of other generations. The strengths of these transitions are determined by the CKM matrix elements. The CKM matrix can be written in terms of only four parameters, one of which is a CP-violating phase. Together with the above discussion, this means the electroweak model contains a total of 17 free parameters that need to be measured. Let us remark that this model features *accidental* (global) symmetries, that are respected by the Lagrangian but were not imposed from the beginning: the conservation of three *lepton numbers* (one per generation) and of one *baryon number*.

The SM as introduced above features massless neutrinos, but the observation of neutrino oscillations implies that at least two out of the three neutrino types are massive. Minimally extending the SM to account for massive neutrinos is certainly feasible, by adding right-handed neutrinos and Yukawa terms for the leptons so as to symmetrise (1.28) and (1.29). Just as with quarks, simultaneously diagonalising the two lepton Yukawa matrices is not possible, and one is left with a mixing matrix between neutrino mass eigenstates, the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix, responsible for neutrino oscillations. Note that as a consequence, only the total lepton number is perfectly conserved. Surprisingly, the mixing between neutrinos turns out to be almost maximal, whereas the CKM matrix is almost diagonal. With this minimal extension, we need to consider seven additional parameters to the SM, one of which is the CP-violating phase of the PMNS matrix.

1.1.4. Strong interaction

The quarks introduced in the last section are never observed in isolation, but are always *bound* inside composite particles, hadrons (except for the top quarks, which decays before a bound state can form). Most hadrons can be classified into mesons and baryons, consisting of two or three "valence" quarks, respectively. To account for the apparent violation of the Pauli principle by the Δ^{++} and Δ^{-} baryons, composed of three up or down quarks with spin $+1/2$, a new quantum number called "colour" taking (at least) three possible values (commonly called red, green and blue) had to be introduced. Quantum

chromodynamics (QCD) is the theory describing the strong interaction between particles carrying colour charge, such as quarks, and is briefly described in this section.

All the quark fields ψ in (1.17) and (1.18) now come in three copies of different colour, ψ_i ($i = r, g, b$). By choosing $\mathrm{SU}(3)_c$ as the gauge group of QCD, it is possible to embed these copies in the three-dimensional fundamental representation of this group. The full gauge group of what we can now call the SM is thus extended as $\mathrm{SU}(3)_c \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$, without affecting the discussion of the previous section. $\mathrm{SU}(3)_c$ possesses eight generators, which can be represented as the Gell-Mann matrices λ_a ($a = 1 \ldots 8$). Each of these generators is associated with a gauge boson G^a_μ, transforming in the adjoint representation of $\mathrm{SU}(3)_c$. They are the mediators of the strong interaction and are called *gluons*. Introducing the new coupling constant g_s, the Lagrangian of QCD is given by:

$$\mathcal{L}_{QCD} = -\frac{1}{4} G_{a,\mu\nu} G_a^{\mu\nu} + \overline{\psi}_i i\gamma^\mu D_\mu^{ij} \psi_j, \qquad D_\mu^{ij} = \partial_\mu - i g_s G_{a,\mu} \frac{\lambda_a^{ij}}{2}. \tag{1.38.}$$

The labels i, j and a are colour indices for quarks and gluons, respectively. Sums over all repeated indices are understood. The interaction term of the covariant derivative in (1.38) is added on top of those introduced in (1.21) and (1.22) for left- and right-handed quarks, respectively. Since $\mathrm{SU}(3)_c$ is non-abelian, the Lagrangian also features self-interaction terms for the gluons, which can be obtained by inserting into (1.38) the definition of the gluon field strength,

$$G_a^{\mu\nu} = \partial^\mu G_a^\nu - \partial^\nu G_a^\mu - g_s f_{abc} G_b^\mu G_c^\nu, \tag{1.39.}$$

where f_{abc} are the structure constants of $\mathrm{SU}(3)_c$: $[\lambda_a, \lambda_b] = i f_{abc} \lambda_c$.

The strong interaction being strong, it is not clear whether perturbative expansions of the type (1.14) are possible. Somewhat luckily, the renormalisation of QCD leads to a running of the strong coupling constant, $\alpha_s = g_s^2/(4\pi)$, that can be written (to first order) as:

$$\alpha_s(\mu^2) = \frac{\alpha_s(\mu_0^2)}{1 - \alpha_s(\mu_0^2) b_0 \log\left(\frac{\mu^2}{\mu_0^2}\right)}, \quad b_0 = \frac{1}{12\pi}(2n_f - 33) \tag{1.40.}$$

With the number of quark flavours[1] $n_f = 6$ we have $b_0 < 0$, and the strong coupling *decreases* when the renormalisation scale increases (in contrast with the behaviour of α). At high energies, the strong coupling becomes small enough that perturbative calculations are tractable, indicating that quarks and gluons are *asymptotically free* [19,20]. Indeed, at $\mu_0 = m_Z$ the coupling is measured to be $\alpha_s(m_Z) = 0.1181(11)$ [5]. Conversely, there exists a scale $\Lambda_{\mathrm{QCD}} \approx 200\,\mathrm{MeV}$ at which the coupling becomes infinite, and the perturbative description breaks down.

When high-energy quarks or gluons (*partons*) are produced in a collision, they emit copious amounts of *final-state* radiation (FSR) in the form of additional gluons and quarks. These new partons are mostly created in the vicinity of the original ones, resulting in collimated clusters of radiation called *jets*. This emission process is called *parton shower*. It

[1] Strictly speaking, here n_f represents the number of *massless* quarks. However, the conclusions do not change if we take $n_f = 5$ and properly account for the top quark mass.

is *unitary* and universal, i.e. it does not depend on how the initial partons were produced. During the shower, the virtuality of the emitting partons decreases, and conversely α_s increases. When reaching a cutoff scale of order Λ_{QCD}, the resulting quarks and gluons become *confined* into objects of zero ("white") net colour charge, hadrons, which propagate freely and which we are able to detect in experiments (either directly or through their decay products). Since the hadronisation of partons is a local process and mostly preserves the jet structure, using the observed hadrons (or some proxy thereof, such as calorimeter towers) and inverting the shower process allows us to infer the kinematics of the original partons created in the "hard" interaction. This inversion is handled by *jet clustering algorithms*.

The definition of jets depends on the clustering algorithm used, the typical size or resolution of the resulting jets, and the recombination scheme. A clustering algorithm should be both infrared- and collinear-safe, i.e. the resulting collection of jets should not change when splitting a particle into two collinear ones, or when adding a soft additional particle to the input collection. The algorithm used in this work is the anti-k_{T} algorithm [21], which satisfies both these properties, ensuring that perturbative QCD predictions in terms of the resulting jets are well-defined. In the anti-k_{T} algorithm, particles are sequentially clustered according to the distances:

$$d_{ij} = \min\left(p_{\text{T}i}^{-2}, p_{\text{T}j}^{-2}\right)\frac{\left(\Delta R_{ij}\right)^2}{R^2},\ d_{iB} = p_{\text{T}i}^{-2}, \tag{1.41.}$$

where ΔR_{ij} is the pseudo-angular distance between particles i and j and R is the typical angular size of the resulting jets. Iteratively, the minimal distance among all particles is searched; if it is a d_{ij}, particles i and j are replaced by new a object according to the recombination scheme chosen. The recombination scheme governs how two 4-momenta should be combined, as this can be done e.g. either by enforcing 4-momentum conservation, as done in this work, or by giving up conservation of energy but requiring that the resulting objects remain massless. If the minimal distance is a d_{iB}, object i is called a "jet" and removed from the input collection. The anti-k_{T} algorithm is fast, thanks to the `FastJet` implementation [22]. In addition, it yields regular, cone-shaped jets, which simplifies the application of experimental corrections. The R parameter should be chosen large enough to ensure most of the radiation from the original partons is clustered inside the jets, but not be so large as to be sensitive to contributions from backgrounds such as additional soft particles present in the event. Its choice thus depends on the typical kinematics of the partons in the problem; in this work, we have used $R = 0.4$.

The inelastic scattering of hadrons (e.g. protons) can be modelled using the hypothesis of *factorisation* between the "hard" scattering of partons inside the hadrons, and the low-energy dynamics within the hadrons. The cross section for the scattering of hadrons $h_{1,2}$ to the final state X is then given as the convolution:

$$\mathrm{d}\sigma_{h_1 h_2 \to X} = \sum_{a,b}\int \mathrm{d}x_a\,\mathrm{d}x_b\ f_a(x_a, \mu_F^2)\, f_b(x_b, \mu_F^2)\ \mathrm{d}\sigma_{a,b\to X}\left(x_a, x_b, X, \mu_R^2\right), \tag{1.42.}$$

where the $\mathrm{d}\sigma_{a,b\to X}$ is the scattering cross section of partons of flavour a and b to the final

state X, and $f_a(x_a, \mu_F^2)$ are the parton distribution functions (PDFs), which correspond (at leading order in QCD) to the probability density to encounter a parton of type a, treated as approximately free, with momentum fraction x_a inside the hadron (i.e., $p_a = x_a p_{h_1}$). The possible flavours of a and b are not only those of the hadrons' valence quarks, but can also be gluons or virtual ("sea") quarks and antiquarks.

Going beyond LO, one encounters infrared divergences due to the emission of collinear *initial-state* radiation (ISR) by the partons, prior to the hard scattering. These lead to a renormalisation of the PDFs, and to a dependence thereof on a new unphysical scale, the *factorisation scale* μ_F. Similarly to the coupling constants, the PDFs cannot be computed from first principles, however, being universal, they can be measured by different means. Measurements at a given scale μ_F can be related to a different scale μ_F' through the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) equations [23–25], obtained using perturbative QCD. As with the renormalisation scale (see Sec. 1.1.1), the dependence of predictions on μ_F can be seen as an artifact of the perturbative calculation method used. Experimental and theoretical uncertainties in PDF estimations, as well as the dependence on the factorisation scale, translate into additional uncertainties in the resulting hadronic cross sections. The PDF sets used in this work are provided by the NNPDF collaboration [26], and are accessed using the `LHAPDF6` library [27].

1.2. Event modelling and generation

In order to compare the predictions obtained using the theory outlined above with experimental measurements, one wishes not only to compute total cross sections and decay rates (inclusive observables), but to compute more fine-grained (exclusive) properties of collision final states, which at high-energy hadron colliders include typically $O(100)$ particles. This task is handled by Monte-Carlo generators, which simulate complete event histories, as shown on Fig. 1.1, distributed according to the theory's probability density. Then, any differential cross section can be estimated by counting the fraction of generated events that satisfy some criterion. One might wonder whether the use of such exclusive quantities would spoil the cancellation of infrared divergences between soft and collinear emissions of real radiation and loops in virtual contributions guaranteed by the Kinoshita–Lee–Nauenberg (KLN) theorem [28, 29]. However, that cancellation extends to quantities expressed in terms of jets defined using infrared- and collinear-safe clustering algorithms, such as the anti-k_T algorithm introduced in the last section. Reviews on the subject of Monte-Carlo event generation can be found in Refs. [30, 31].

Event generation starts with the hard-scattering cross section (1.42) for a given process at a given order in perturbation theory, and the integration over the phase space specified in (1.11)–(1.13). Note that even at LO, a complete description of the hard scattering is only possible for low-multiplicity final states, featuring typically less than 10 particles. The phase-space integration is typically carried out using adaptive Monte-Carlo techniques such as importance sampling, as implemented in Vegas [32], whose convergence rate is independent of the dimensionality of the integral. Efficiently sampling the phase space is made possible by recursively factorising it into two-body components and intermediate invariants, which can be mapped to propagators in the matrix element. If the latter contains different propagator structures, several phase-space parameterisations are combined using the multi-channel method. Each point sampled in the phase space

corresponds to a particular parton-level configuration, or event. At this step, the points are sampled using a different measure than the distribution that is being integrated. Hence, these events are *weighted*, i.e. they each contribute differently to the sum approximating the phase-space integral. To obtain events distributed according to theory predictions, an *unweighting* procedure based on the acceptance-rejection Monte-Carlo method is used, yielding a sample of N events with equal weights $w = \sigma/N$, where σ is the total cross section of the process being generated. The evaluation of the matrix elements requires efficient numerical methods to handle the factorial growth of the number of diagrams with the number of final-state objects, as well as the average and sum over the large number of unobserved degrees of freedom in the initial and final state, due to helicity and color states. The former is achieved by using dedicated parameterisations of the wave functions such as the HELAS routines [33], and the latter with the help of recursion relations [34].

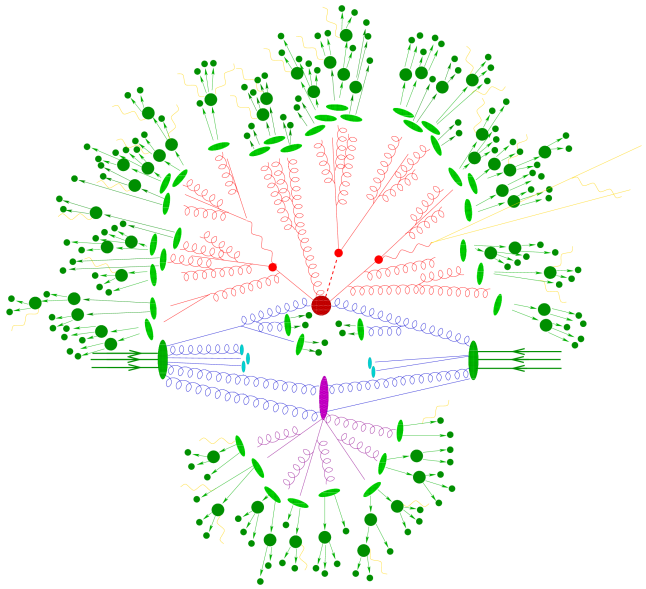

Figure 1.1. | Schematic representation, taken from Ref. [31], of the modelling of a hadron-hadron scattering event by Monte-Carlo generators. The triple green arrows represent the incoming hadrons. The red blob stands for the hard interaction. The outgoing red lines are the decays of resonances produced in the collision, as well as the FSR simulated by the parton shower. The initial-state partons and the ISR are shown as blue lines. Light green ellipses represent the hadronisation process, from where hadrons (dark green circles) emerge and (possibly) decay. Finally, MPI and UE are shown in pink and light blue.

At NLO, the cross section for the production of n particles consists of several parts, namely the LO Born cross section, the NLO virtual correction, and the real emissions. While the first two contributions are to be integrated over the n-body phase space, the last one leads to an $n+1$-particle final-state. A difficulty then arises from the presence of infrared divergences in both the virtual and real corrections, that cancel only when integrating over the different phase spaces. This issue can be cured by including subtraction terms into both contributions, that remove the divergences and themselves

cancel out in the final result [35,36]. Crucially, these subtraction terms can be factorised into the Born contribution and a kernel with *universal* structure, making the automation of the procedure possible [37,38]. However, this is still not sufficient for the generation of Monte-Carlo events at NLO, as will be explained below.

Very often, a high-precision fixed-order computation (say NNLO) of the total cross section for a process is available, but only a lower-precision (say LO) Monte-Carlo sample can be used to obtain more exclusive observables. The sample can then be *normalised* to the best known value by multiplying all event weights by a common K-factor, i.e. $K = \sigma_{NNLO}/\sigma_{LO}$. However, the reduction in the scale uncertainty provided by the precise computation is lost, since there is no general *a priori* way to know how this uncertainty affects exclusive regions of phase space when applying a selection on the Monte Carlo events. In many situations, while the overall K-factor is large, differential distributions are still well modelled by the low-order results. If that is not the case, differential K-factors that depend on event kinematics can be used to correct the modelling of specific distributions.

While decay chains of heavy resonances can in principle be properly modelled by including the full production and decay MEs, this is a computationally intensive task. At NLO, the latter approach is downright impractical, and events have to be generated in the narrow-width approximation (NWA), in which the production and decay amplitudes are factorised. Thus, heavy resonances are produced on-shell and are decayed independently in a separate step. Nevertheless, spin correlation and (partial) off-shell effects can be recovered using e.g. MadSpin [39]. For very narrow or scalar resonances (such as the Higgs boson), the NWA is accurate for most practical purposes and their decay is usually handled by parton shower generators, described below, or by other dedicated tools such as Tauola [40].

Once a sample of hard-scattering events has been created, it is processed by a parton shower generator that simulates the emission of soft and collinear partons in the initial and final states. This can be seen as including all-order corrections due to radiative emissions needed for a complete, exclusive description of many-particle final states. The corresponding approximate virtual corrections are obtained from unitarity arguments, since the shower does not affect the cross section of the process being generated. Parton showers are implemented as Markov-chain Monte-Carlo algorithms: each emission of an additional parton is treated as a random event, depending only on the previous configuration. The emissions are ordered according to an *evolution variable*, which can be e.g. the emission angle, or the virtuality of the parent parton. The shower is started at a large energy scale corresponding to that of the hard scattering, and is evolved all the way down to a cutoff scale, at which point the confinement of partons into hadrons has to be modelled. The probability distribution for the evolution variable at each branching can be computed based on so-called splitting kernels, such as the Altarelli–Parisi functions [23], which also govern the distribution of the remaining degrees of freedom of the partons emitted at each step (e.g. the energy fraction and azimuth w.r.t. the parent partons). These splitting kernels describe the behaviour of the soft and collinear divergences of the emission process. They can be obtained in perturbation theory, e.g. as a limiting case of full matrix elements describing the emission of a parton by a two-parton final state. As several choices are possible for the evolution scale or the parameterisation of

the splitting kernels, that are equivalent only in the strict soft or collinear limit, there exist different parton-shower generators that differ i.a. in this respect.

In the absence of any microscopic description of hadronisation, two classes of "effective", non-perturbative models have been developed: the string ("Lund") and cluster models. The string models, implemented in Pythia [41, 42], start from the observation that the long-distance confining behaviour of QCD is linear, i.e. when a high-energy quark-antiquark pair moves apart, they can be seen as being connected by a flux tube, or "string", with a potential energy proportional to their separation distance. At some point, it then becomes energetically favourable to "break" the string by creating a new quark-antiquark pair. This process iterates until no further breaks are possible, at which point the remaining quark-string systems are replaced by mesons. A Lorentz-covariant description of the process can be achieved by modelling the flux tube as a relativistic string; further refinements to this simple picture are needed to accommodate gluons and the production of baryons and heavy-flavour quarks. String models can be shown to be infrared- and collinear-safe. Cluster models, used e.g. by Herwig [43], are based on "preconfinement", i.e. the propension of showering partons to form colour-singlet "clusters" with universal invariant mass distribution. Light clusters then directly decay to hadrons, while heavy clusters undergo intermediate splittings. Since both string and cluster models involve parameters that cannot be derived from first principles, they have to be *tuned*, i.e. their parameters are fitted from dedicated measurements. Thanks to the universality of hadronisation, these tunes can then be used to formulate predictions for a wide range of processes. A complete event description also requires modelling the underlying event (UE), i.e. the interaction between the remnants of the hadrons that have emitted the partons involved in the hard scattering. Notably, this includes multiple interactions between partons from the initial hadrons (MPI), with subsequent showering and hadronisation, and the interactions between coloured objects in the initial and final states (colour reconnection). As with hadronisation, the UE description has to be tuned to data.

Combining, or *matching* exact fixed-order matrix-element computations with the approximate but all-order parton shower is not straightforward. In particular, it is necessary to avoid any double-counting between configurations handled by both the NLO real-emission matrix element or by the shower, and dead regions handled by neither. Different approaches have been developed to that end, yielding equivalent results up to NLO. In the Powheg method [44, 45], the first (hardest) emission is modified by the real-emission ME, and subsequent showering proceeds as usual. With the alternative MC@NLO method [46], a subtraction term is added to the real-emission ME, to remove contributions corresponding to the first shower emission by Born-level configurations. A correction has then to be added back to the Born term in order to preserve the total cross section. Since the real subtraction term can be larger than the ME itself in some regions of phase space, the generated samples will include negatively weighted events. After the unweighting step, the contribution of each event i to the total cross section σ is then given by $w_i\sigma/\sum_i w_i$, where $w_i = \pm 1$. Note that everything else being equal, the presence of events with negative weights significantly affects the statistical precision of any computation based on those Monte Carlo samples.

Somewhat related to matching is the *merging* of several tree-level ME computations

with different final-state parton multiplicities [47]. In this way, the first few emissions of hard, wide-angle radiation are described by the full MEs, while soft and collinear emissions are left to the shower. Divergences in the hard MEs are regulated by applying resolution cuts at the parton level, whereas any double counting due to the parton shower is avoided using requirements applied after the shower and hadronisation steps. The matching of each hard-multiplicity result with the parton shower, and the merging of the samples with different multiplicities are implemented in different methods such as CKKW(-L) [48,49] and MLM [50] at LO, and FxFx at NLO [51].

The automated generation of LO or NLO matrix elements and parton-level event samples from arbitrary Lagrangian-based models, formulated e.g. with the FeynRules [52] or Sarah [53] packages, has been implemented in programs such as MadGraph5_aMC@NLO [54], Powheg-Box [55], Sherpa [56] or Whizard [57]. Many other Monte-Carlo tools, too numerous to be quoted here, have been developed to model LO and NLO matrix element generation, matching and merging with showers, simulation of hadronisation and underlying event, decaying resonances and final-state hadrons, etc. While most of them are not automatic and implement a limited set of processes, some even achieve NNLO-QCD or NLO-electroweak precision.

Let us conclude this section with the observation that a sample of unweighted events generated under process α can be *reweighted* to obtain a sample for another process β, provided they both yield the same final state. At LO, this can be achieved by modifying the original event weights w, on an event-by-event basis as:

$$w \to \widetilde{w} = w \cdot \frac{|\mathcal{M}_\beta(\boldsymbol{x})|^2}{|\mathcal{M}_\alpha(\boldsymbol{x})|^2}, \tag{1.43.}$$

where $\mathcal{M}(\boldsymbol{x})$ denotes a matrix element evaluated using the parton-level event kinematics $\boldsymbol{x}$. This procedure yields exact predictions for β, with the following caveats:

- Reweighting always degrades the statistical precision of quantities computed from the simulated events, compared to a dedicated unweighted sample with the same number of events. If the two processes have very different kinematical distributions, the variance of the new weights $\widetilde{w}$ will be high, and any quantity computed from the reweighted sample will suffer from a large statistical uncertainty.
- If the new process β contributes to regions of phase space that are not populated by α, the procedure will fail completely (this is just an extreme case of the previous point).
- In practice, the matrix elements in (1.43) are often taken to be spin- and colour-summed MEs. However, the spin and colour of the partons have an impact on other steps of event generation, such as the parton shower. Hence, if the spin and colour structures of α and β are different, this will not be properly taken into account by the reweighting.

If one wishes to conserve the overall normalisation of the sample, the weights have to be multiplied by an extra factor $\sigma_\alpha/\sigma_\beta$. Although more involved than the simple formula (1.43), the reweighting of samples generated at NLO is now also possible within MG5_aMC@NLO [58]; however this is only feasible during the generation of the original sample.

Uncertainties due to the renormalisation and factorisation scales, μ_R and μ_F, are typically estimated by varying these two scales independently around their central value by factors of 0.5, 1 and 2, omitting the two extreme cases where one scale is varied up and the other down. For a given observable, its scale uncertainty is then estimated by taking the envelope, i.e. the range between its minimum and maximum value, of the seven resulting predictions (nominal and six variations). Unfortunately, this *ad-hoc* procedure provides no general guidance on how to correlate the uncertainties on two independent observables. In practice, these scale variations can be computed by reweighting samples of generated events, whereby the matrix element and PDFs are evaluated using the different values of both scales, thus avoiding the need to generate and use alternative samples for each scale variation. This latter reweighting is typically carried out by SysCalc [59] or directly by MG5_aMC@NLO.

Event reweighting is also invaluably useful for the propagation of the uncertainty due to the PDFs, for which we follow the interim recommendations by the PDF4LHC working group [60, 61]. For each event, 100 replicas of the NNPDF 3.0 set [26] are stored. The standard deviation of the event weights computed from this set of replicas, which follow the posterior distribution of the PDF fit, yields up and down variations of the weights that can be propagated to any desired observable to compute the PDF uncertainty. The uncertainty due to the value of α_S in the PDFs is added in quadrature to these previous variations.

1.3. Double Higgs production in the Standard Model

After the discovery of a Higgs boson by the ATLAS and CMS collaborations [62, 63], it is essential to measure all its properties to test the mechanism of spontaneous symmetry breaking as implemented in the SM. Among those properties, the values of the Higgs self-couplings in (1.33) constitute crucial predictions of the SM, since they are fixed by the knowledge of the Higgs boson mass and the symmetry breaking VEV. The most direct way to experimentally access the cubic and quartic self-couplings is to characterise the processes of double and triple Higgs boson production, respectively, since these are the only processes where these couplings appear at LO. While there currently exists no conceivable way to approach triple Higgs production due to its vanishingly small cross section [64], the former is expected to become accessible in the near future. Thus, the characterisation of double Higgs production, or HH, constitutes an active area of research.

At a pp collider, the main production mode of Higgs boson pairs is through the fusion of gluons. This process, first studied in Refs. [65–67], is loop-induced: even at LO, it involves diagrams containing loops of top and bottom quarks, as depicted on Fig. 1.2. A first class of diagrams features a triangular loop yielding an off-shell Higgs, that decays into two on-shell Higgs bosons via a triple-Higgs vertex. However, it is also possible to produce two Higgs bosons through "box" loop diagrams that do not contain the Higgs self-coupling. These two sets of diagrams contribute to the total amplitude with opposite signs, leading to a destructive interference with non-trivial phenomenological consequences.

The total production cross section for $pp \to HH$ is known at NNLO+NNLL accuracy in QCD, in the infinite top quark mass limit (so-called (HEFT) [68, 69]. Under the latter

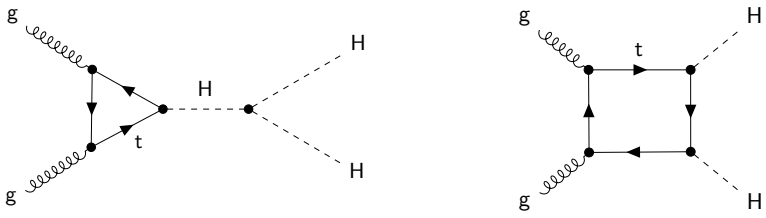

Figure 1.2. | Examples of LO Feynman diagrams contributing to HH production. Only "triangle" diagrams such as the one depicted on the left depend on the Higgs self-coupling λ. Both top and bottom quarks are allowed to circulate in the loop, but contributions from the former dominate the amplitude due to its large Yukawa coupling appearing once (twice) in the triangle (box) diagrams. Diagrams featuring an *s*-channel gluon decaying to two Higgs bosons *via* a triangular loop cancel due to colour conservation.

approximation, justified by the fact that $2m_t > m_H$, the loops in the diagrams 1.2 are collapsed to effective ggH and ggHH vertices, simplifying the calculation of virtual corrections. Results obtained in the HEFT are usually rescaled by the ratio between the exact and HEFT total rates at LO. However, the accuracy of the HEFT approximation is limited by the large invariant mass of the final state, $m_{HH} \gtrsim 2m_t$, leading to sizable corrections due to finite top quark mass effects. The cross section has also been obtained with full top quark mass dependence at NLO in QCD [70,71], and several methods have since been developed to combine the exact NLO with the approximate NNLO results. In Ref. [72] it was recommended to apply the top quark mass correction of $\mathcal{O}(-10\%)$ observed at NLO to the total NNLO+NNLL cross section. With $\sqrt{s} = 13\,\text{TeV}$ and $m_H = 125\,\text{GeV}$ one obtains the value of σ_{HH} used in this work:

$$\sigma_{HH} = 33.45\,\text{fb}\ ^{+4.3\%}_{-6\%}(\text{scale}) \pm 2.1\%\ (\text{PDF}) \pm 2.3\%\ (\alpha_s), \tag{1.44.}$$

with additional uncertainties due to top quark mass effects estimated at $\mathcal{O}(5\%)$ based on an m_t^{-n} expansion at NNLO up to $n = 12$ [73]. However, very recently a more accurate result has been been obtained [74]. In this approximation, dubbed $\text{FT}_{\text{approx}}$ (first employed at NLO in Ref. [64]), parton-level events generated for the NNLO computation in the HEFT are reweighted by the ratio between the exact and HEFT matrix elements at LO for the same partonic configuration. In particular, the contribution with two real emissions is thus modelled with full m_t dependence. This approach yields a slightly lower value for σ_{HH} than the recommendation (1.44):

$$\sigma_{HH} = 31.05\,\text{fb}\ ^{+2.2\%}_{-5\%}(\text{scale}), \tag{1.45.}$$

for which the authors estimate a residual uncertainty from finite m_t effects at ±2.6%.

The HH cross section is notably small, mainly due to the destructive interference between the box and triangle diagrams. For comparison, single H production through gluon fusion is ≈ 500 times more frequent [72]. Note that QCD corrections to the HH process

are important, more than doubling the LO cross section.

Similarly, differential cross sections are available at NNLO+NNLL precision in the HEFT [69,75], at NLO with finite m_t and with matching to parton shower simulation [76, 77], and at NNLO in the FT_{approx} approach [74]. The dependence of higher-order corrections on event kinematics is mild or at the very least covered by scale uncertainties for what concerns the overall normalisation, indicating that until it becomes practical to use NLO Monte-Carlo samples for HH production, samples generated at LO with full top quark mass dependence and rescaled to the cross section (1.44) provide a sufficiently accurate description of that process. Naturally, since scale uncertainties are of the order of 25 % at LO, experimental studies should benefit from the use of higher-order Monte-Carlo simulations and it is planned to switch to NLO in the near future.

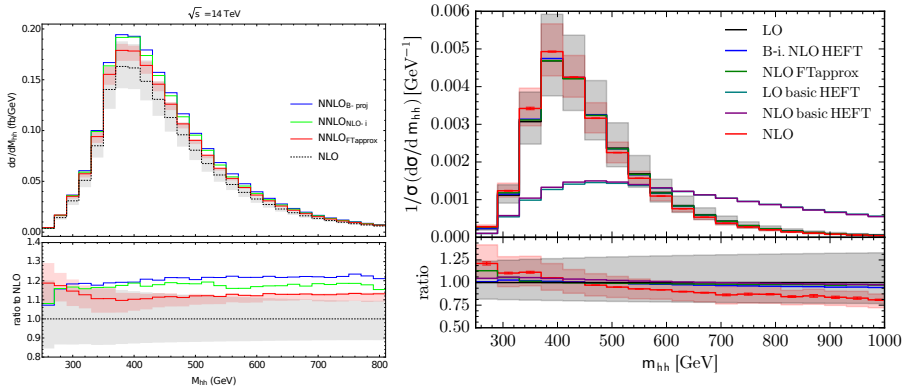

Figure 1.3. | Fixed-order theoretical predictions for the distribution of m_{HH} in gluon-fusion Higgs pair production in the SM. Left: comparison of different approximate NNLO results, taken from Ref. [74]. Right: *normalised* distributions at NLO with full top quark mass dependence, compared to LO and various approximate NLO results, taken from Ref. [70]. Note that the plain HEFT completely fails to model the shape of the distribution. In both figures, shaded bands show scale uncertainties. Differential K-factors, defined as NNLO/NLO on the left and NLO/LO on the right, are displayed on the bottom panels.

At LO and before parton shower, HH production is a $2 \to 2$ process with scalars in the final state, thus once an irrelevant overall azimuthal angle has been removed, the final state is characterised by only three variables. One of these can be related to the longitudinal boost of the Higgs boson pair along the axis of the colliding protons (beam axis), which is determined by the proton PDFs. Hence, the physics of the HH process itself can be isolated in two variables, namely the invariant mass of the boson pair, m_{HH}, and $|\cos\theta^*_{CS}|$, where θ^*_{CS} is the polar angle of one boson w.r.t. the beam axis computed in the centre-of-mass frame of the boson pair (Collins-Soper frame [78]). A partial-wave analysis reveals that the s-wave component is dominant; in fact, the d-wave contribution vanishes in the limit $m_t \to \infty$ [79]. As a consequence, the differential cross section as a function of $|\cos\theta^*_{CS}|$ is almost flat, and the only variable of interest is m_{HH}, shown on Fig. 1.3. Interestingly, in the SM the interference pattern between box and triangle diagrams is such that in the heavy top quark mass limit, the scattering amplitude cancels

at the threshold $m_{\mathrm{HH}} = 2m_{\mathrm{H}}$. Naturally, higher-order terms spoil the simple two-body picture: beyond LO the bosons are not produced back-to-back in azimuth, and non-trivial corrections to the p_{T} of the boson pair ($p_{\mathrm{T}}(\mathrm{HH})$) appear. Finally, note that since Higgs bosons are scalars and have very narrow width, the decay kinematics factorise from the production and the NWA can be accurately employed.

A review of the current state-of-the art modelling of the HH process can be found in Ref. [72].

1.4. Extending the Standard Model

The SM is a tremendously successful theory: based on the measurement of its 18 parameters (three gauge coupling constants, two parameters of the scalar potential, nine fermion masses and four parameters of the CKM matrix)[1], uncountable, precise predictions can and have been formulated. These predictions have shown impressive agreement with experimental data. In principle, the renormalisability of the SM implies that it is a fully consistent theory that can generate predictions for physical processes up to the Planck scale $M_{\mathrm{Planck}} \approx 10^{19}$ GeV, where we know that it should be replaced by a quantum theory of gravity. However, there are serious indications, some listed below, that the SM should be extended or replaced at energies lower than M_{Planck}:

- The **hierarchy problem** appears when taking radiative corrections to bare parameters in the Lagrangian at face value. Regularising the ultraviolet divergences with a cutoff scale Λ yields corrections to most parameters of the type $\log\Lambda$, but the parameters in the scalar potential (1.33) receive quadratically diverging contributions behaving as $\propto \Lambda^2$. Since nothing prevents taking $\Lambda \to M_{\mathrm{Planck}}$, there should be an extraordinarily *fine* tuning of the bare parameters and the corrections so that their difference yields the small observed values for the Higgs boson mass and VEV of order $\mathcal{O}(10^2\,\mathrm{GeV})$. In fact, the corrections become already as large as the measured values when $\Lambda \approx \mathcal{O}(\mathrm{TeV})$. The problem also appears when building extensions to the SM containing new, very heavy scalars, that might mix with the Higgs boson. Either way, protecting the Higgs boson from high-energy phenomena would require new symmetries or degrees of freedom appearing not too far away from the electroweak scale. In addition, one might argue that the mechanism of spontaneous symmetry breaking in the SM is somewhat *ad-hoc*, and that it would be satisfying to be able to explain it as consequence of some underlying dynamics.
- While matter and antimatter are almost symmetric in the SM, the universe contains almost none of the latter. Generating a **matter-antimatter asymmetry** (baryo- and leptogenesis) requires CP, lepton and Baryon number violation, but the SM does not seem to yield enough of either to explain the observed asymmetry. However, one proposed way to obtain electroweak baryogenesis is through modified scalar potentials that would translate into values of the Higgs boson self-coupling larger than in the SM [80].
- The QCD Lagrangian (1.38) should contain an additional, so-called θ term, which violates CP symmetry. Since no **CP violation** has been observed in the **strong sector**, this additional coupling should be extremely close to zero. This is another problem

[1] 25 if we include Dirac neutrino masses.

of *fine-tuning* between the actual θ parameter in the Lagrangian, and a contribution appearing when diagonalising the quark Yukawa matrices, that somehow conspire to yield an almost-zero value.

- The smallness of the neutrino Yukawa couplings introduced in Sec. 1.1.3 hints at the existence of new degrees of freedom at a scale of $\approx 10^9$–10^{13} GeV, unfortunately way out of direct and indirect reach of current collider experiments.
- It would be satisfying to reduce the free parameters of the SM to a smaller, more fundamental set. In particular, it is unknown why there should be exactly three families of fermions, and the irregular pattern of fermion masses and mixings is not understood: this is the **flavour puzzle**. Furthermore, the three gauge couplings appear to (almost) meet when evolved to a high energy scale of $M_{\text{GUT}} \approx 10^{14}$–$10^{16}$ GeV. This might indicate that the SM gauge groups emerges from a spontaneously broken larger gauge group, a grand unified theory (GUT).
- More than 80% of matter in the universe is not composed of the particles listed in Tab. 1.1, but of some other, unknown substance, i.e. **dark matter**. Its presence is inferred indirectly from astronomical observations and from its role in cosmological models. The most popular candidate for dark matter, an electrically neutral, weakly-interacting particle with mass around the electroweak scale, has so far eluded experimental confirmation.

To our knowledge, there is currently no proposed model that would solve all of the above issues, but countless proposals have been put forth to address at least some of them. Most of these require the existence of new degrees of freedom at energies higher than, but close to the weak scale. Some models have been developed not because they solve a specific problem, but because they are *possible* within current experimental constraints and still provide accessible signatures that would be a clear sign of beyond the SM (BSM) dynamics. In this top-down, model-building approach, experimentalists have to specifically search for every proposed signature of new particles. Given our cluelessness as to where New Physics should appear, almost a decade after the LHC was switched on, an alternative bottom-up approach is becoming more and more attractive. In this setting, signs of high-energy degrees of freedom are sought after in the indirect effects they might have on SM observables. This approach is introduced in more details in the next section.

1.5. Effective Field Theory approach to double Higgs production

If the scale Λ at which New Physics appears is well separated from the electroweak scale, it is possible to parameterise in a model-independent way all the indirect effects on lower-energy observables due to the new degrees of freedom. These manifest themselves either as unobservable redefinitions of the SM fields and parameters, or as a tower of additional, nonrenormalisable local interactions among the SM particles [81]. The SM Lagrangian can thus be extended as an effective field theory (EFT) [82–84]:

$$\mathcal{L}_{\text{SM}} \to \mathcal{L}_{\text{eff.}} = \mathcal{L}_{\text{SM}} + \sum_{d>4}^{\infty} \sum_{i} \frac{c_i^{(d)}}{\Lambda^{d-4}} O_i^{(d)}, \tag{1.46.}$$

where $O_i^{(d)}$ are *effective operators* (with dimension of $[\mathrm{E}]^d$), that respect the Poincaré and SM gauge symmetries, and $c_i^{(d)}$ are so-called Wilson coefficients that encode the strength of these new interactions. The SM is recovered when either $c_i^{(d)} = 0$, for all i and d, or $\Lambda \to \infty$. At each order in Λ^{-1} only a finite number of operators contribute, but the expansion (1.46) contains an infinity of terms. Nevertheless, effects of higher-dimensional operators on a given process with typical energy E are increasingly suppressed by powers of E/Λ, so that it is possible to truncate the series and consider only the relevant, dominant operators.

Two conditions should be satisfied so that this approach yields well-defined predictions:

1. The typical energy scale E of the considered processes should be smaller than the new physics scale Λ. At $E \approx \Lambda$, all operators in (1.46) contribute equally, the effective expansion breaks down, and the theory should be replaced with a *UV-complete* description of the new degrees of freedom.
2. The underlying UV theory is required to admit a perturbative expansion in its couplings. Since Wilson coefficients are obtained from these couplings, this can be translated using dimensional arguments into the constraint $|c_i^{(d)}| \lesssim (4\pi)^{n_i - 2}$ for an operator made of n_i fields [85].

While the effective interactions lead in principle to a nonrenormalisable theory, this does not prevent the calculation of higher-order corrections, since there now is a clear cutoff scale Λ [86–88]. Likewise, scattering cross sections will ultimately violate unitarity bounds, but will only do so at energies at which the effective description is expected to break down. A well-known example of effective theory is found in the Fermi description of weak interactions, introduced in Sec. 1.1.3. In that case, the mass of the W bosons played the role of the new physics scale Λ, and the Wilson coefficients could be obtained from the coupling g of the full high-energy, complete theory, i.e. the SM.

In the EFT approach, the experimental goal is to place constraints on, or to prove the non-zero value of the operator coefficients $c_i^{(d)}/\Lambda^{d-4}$. Assuming some reasonable values for the Wilson coefficients, these constraints can be translated into bounds on the scale at which new physics is present. The EFT can also act as a guide, indicating which observables could be most sensitive to deviations from SM predictions. Conversely, starting from a given BSM model, it is generally possible to derive the resulting EFT limit, and experimental constraints or observed deviations on the operators can then be directly reinterpreted in terms of the parameters in the new model. However, one might question the use of an EFT if experimental constraints have poor precision, such that the second condition listed above then pushes the lower bound on the scale Λ below the energies probed by the experiments. For a discussion of the validity of an EFT approach to measurements in the Higgs sector, see e.g. Refs. [85, 89]. Note that even if new particles are discovered, this does not necessarily rule out the use of the EFT.

At order Λ^{-1} there is only one possible operator, which yields a Majorana mass term for neutrinos and violates lepton number conservation [82]. It can thus be neglected in a collider setting. Going to dimension six, it is important to recall that all the possible operators that can be written at a given order Λ^{4-d} do not lead to independent effects. Instead, some operators can be related through Fierz transformations; others can be

shown to be equivalent *up to order* Λ^{4-d} using equations of motion. As a simple example, if we extend QED with the dimension-six operator $(\overline{\Psi}\gamma^\mu\Psi)(\overline{\Psi}\gamma_\mu\Psi)/\Lambda^2$, the equations of motion are $\partial_\mu F^{\mu\nu} = e\overline{\Psi}\gamma^\nu\Psi + O(\Lambda^{-2})$, which means this operator has the same effect as both $(\partial_\mu F^{\mu\nu})(\overline{\Psi}\gamma_\nu\Psi)/\Lambda^2$ and $\partial_\mu F^{\mu\alpha}\partial^\nu F_{\nu\alpha}/\Lambda^2$, modulo terms of order Λ^{-4}. A minimal set of independent operators is called a *basis*, and at dimension six two such sets have been worked out, the so-called Warsaw [90] and SILH [91,92] bases. There are 2499 dimension-six operators that preserve baryon number, in addition to the symmetries mentioned above that the operators need to obey. However, making the common assumption of *minimal flavour violation* (that the structure of flavour violation in the SM, due to the Yukawa couplings, remains the same in the presence of new physics), that set reduces to a more manageable number of 76 parameters. Finally, neglecting the CP-odd operators, we obtain 53 independent operators whose coefficients have to be determined, of which only a subset contributes to any given process.

Linear EFT effects on HH

There are different ways to work out the EFT modifications to HH production, depending on the basis chosen. In the SILH basis, starting from a manifestly gauge-invariant notation the relevant operators are [93–95]:

$$O_6 \qquad -\frac{c_6}{\Lambda^2}\lambda\left(\phi^\dagger\phi\right)^3 \tag{1.47.}$$

$$O_{\rm H} \qquad \frac{c_{\rm H}}{\Lambda^2}\frac{1}{2}\partial_\mu\left(\phi^\dagger\phi\right)\partial^\mu\left(\phi^\dagger\phi\right) \tag{1.48.}$$

$$O_{\phi G} \qquad \frac{c_{\phi G}}{\Lambda^2}\frac{\alpha_S}{4\pi}\left(\phi^\dagger\phi\right)G^a_{\mu\nu}G^{a,\mu\nu} \tag{1.49.}$$

$$O_{{\rm t}\phi} \qquad -\frac{c_{{\rm t}\phi}}{\Lambda^2}y_{\rm t}\left(\phi^\dagger\phi\right)\overline{Q}_3\phi^C{\rm t}_R + {\rm h.c.} \tag{1.50.}$$

$$O_{{\rm t}G} \qquad \frac{c_{{\rm t}G}}{\Lambda^2}y_{\rm t}g_s\left(\overline{Q}_3\sigma^{\mu\nu}T^a{\rm t}_R\right)\phi^C G^a_{\mu\nu} + {\rm h.c.} \tag{1.51.}$$

We have used the same notations as in Sec. 1.1.3, and $\sigma^{\mu\nu} = [\gamma^\mu, \gamma^\nu]$. The normalisation factors in front of the operators, while arbitrary, are taken as in Refs. [93,94]. Operators proportional to the light quark masses are not considered, since those would yield sub-dominant contributions. Given that O_6 modifies the scalar potential, the Higgs VEV will be shifted compared to the treatment of Sec. 1.1.3. The new couplings for mass eigenstates can be read off by working in unitary gauge, expanding the Higgs boson around the shifted minimum of the potential, and performing a redefinition of fields to obtain canonical kinetic terms. Writing the couplings in terms of the measured $m_{\rm H}$, $m_{\rm t}$ and v, the resulting Lagrangian affecting the HH process takes the form:

$$\mathcal{L} \supset -\frac{m_{\rm H}^2}{2v}\left(1 + \underbrace{c_6\frac{v^2}{\Lambda^2} - c_{\rm H}\frac{3}{2}\frac{v^2}{\Lambda^2}}_{\delta_\lambda}\right){\rm H}^3 \tag{1.52.}$$

$$-\frac{m_t}{v}\Bigg(\underbrace{1-c_H\frac{1}{2}\frac{v^2}{\Lambda^2}+c_{t\phi}\frac{v^2}{\Lambda^2}}_{\delta_y}\Bigg)\bar{t}_L t_R H + \text{h.c.} \tag{1.53.}$$

$$-\frac{m_t}{v}\left(-c_H\frac{1}{2}\frac{v}{\Lambda^2}+c_{t\phi}\frac{3}{2}\frac{v}{\Lambda^2}\right)\bar{t}_L t_R H^2 + \text{h.c.} \tag{1.54.}$$

$$+\frac{c_{\phi G}}{\Lambda^2}\frac{\alpha_s}{4\pi}\left(vH+\frac{H^2}{2}\right)G^a_{\mu\nu}G^{a,\mu\nu} \tag{1.55.}$$

$$+\frac{c_{tG}}{\Lambda^2}\frac{m_t g_s}{v}\bar{t}_L\sigma^{\mu\nu}T^a t_R G^a_{\mu\nu}(v+H) + \text{h.c.} \tag{1.56.}$$

We see that O_6 and $O_{t\phi}$ modify the Higgs self-coupling and the top Yukawa coupling, respectively. Both these couplings are already present in the LO Feynman diagrams for HH in the SM (see Fig. 1.2), however $O_{t\phi}$ also yields a new ttHH contact term responsible for additional diagrams, as depicted on Fig. 1.4 (left). The operator O_H contributes to the three latter couplings. From $O_{\phi G}$ we obtain new ggH and ggHH interactions, and as a consequence new *tree-level* diagrams as those shown on Fig. 1.4 (middle and right). Finally, the operator O_{tG} is responsible for the appearance of ggtt, gttH ggttH contact terms, yielding diagrams of the type shown on Fig. 1.5, and also modifies the ggt vertex present in the SM diagrams. This last operator is usually not considered in phenomenological studies of HH, mainly because it is expected to be strongly constrained through measurements of top quark pair production, which it also affects (see e.g. Ref. [96]). However, in Ref. [94] it was argued that even within these constraints, O_{tG} has a strong impact on both single and double Higgs processes, and that its contribution should not be neglected.

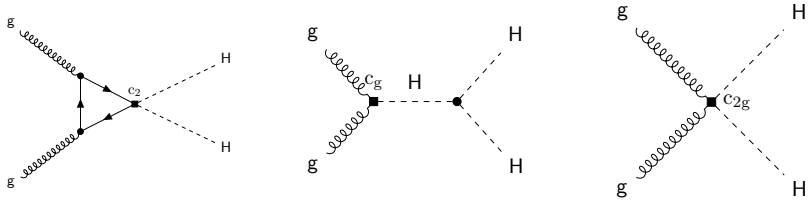

Figure 1.4. Examples of new diagrams contributing to HH production due to the operators $O_{t\phi}$ or O_H (left), and $O_{\phi G}$ (middle and right). The new BSM vertices are shown as black squares, and are labelled according to the couplings (c_2, c_g, c_{2g}) entering the Lagrangian (1.58), introduced later on.

With the $d=6$ EFT Lagrangian in hand, the amplitude for the HH (or any other) process can be expanded around the SM up to $O(\Lambda^{-2})$, and the matrix element written as:

$$|\mathcal{M}|^2 = \left|\mathcal{M}_{SM}\right|^2 + 2\sum_i \frac{c_i}{\Lambda^2}\text{Re}(\mathcal{M}^*_{SM}\mathcal{M}_i) + \sum_{i,j}\frac{c_i c_j}{\Lambda^4}\text{Re}(\mathcal{M}^*_i\mathcal{M}_j), \tag{1.57.}$$

where $\mathcal{M}_{SM}$ and $\mathcal{M}_i$ are the SM amplitude and the amplitude containing one insertion of operator O_i, respectively. The first term in (1.57) corresponds to the SM prediction, the

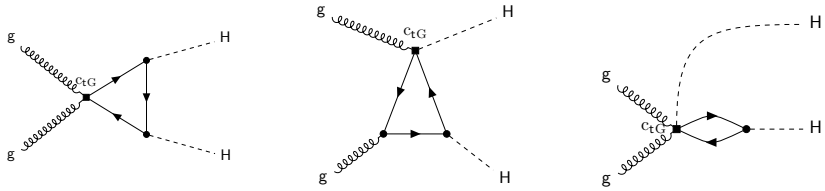

Figure 1.5. | New classes of diagrams entering HH at LO due to the O_{tG} operator defined in (1.51).

second is linear in the parameters c_i/Λ^2 and arises from interferences between the SM amplitude and EFT corrections, and the third term is quadratic and is due to pure EFT contributions (operators squared or interferences between operators). While the latter is formally of order Λ^{-4}, it should be retained in the analysis. Indeed, it might dominate the linear term if the coefficients c_i are large (corresponding to a strongly-coupled UV theory) or if the interference term is suppressed [97], without invalidating the hypotheses of the EFT description since the contributions from dimension-8 operators would still remain sub-dominant as long as $v \ll \Lambda$.

The NLO [98], and even NNLO [99] corrections to HH in the EFT (without O_{tG}) have been worked out in the limit $m_t \to \infty$. For a discussion of the effects of O_{tG} at LO, see Ref. [94]. Due to the characteristic interference pattern between the different diagrams, even moderate values for the operator coefficients lead to important changes to the HH cross section. The effect on the shape of m_{HH} is dramatic, particularly at threshold where the suppression observed in the SM no longer holds [94, 99, 100], as shown on Fig. 1.6. The main observations are the following:

- Given the impact of effective operators on the kinematics, any experimental search for nonresonant HH production must take into account the changes in signal acceptance and analysis sensitivity due to BSM effects.
- The dependence of (global) NNLO/LO K-factors on the operator coefficients is mostly flat when varying one coefficient at a time [98]. This can be understood by observing that the large QCD corrections are dominated by soft and collinear gluon effects, which are relatively insensitive to details of the hard scattering. If several coefficients are allowed to float simultaneously, larger effects are seen, but these are concentrated in regions of parameter space where the cross section is suppressed, and thus far beyond current experimental reach [99].
- Since five operators have to be considered, measuring the total HH cross section does not yield sufficient information to constrain them in a global fit. Using shape information for HH, in particular through m_{HH} at the kinematic threshold, is crucial to disentangle the contributions from the different operators and lift the degeneracy [95, 99, 101] (see e.g. Fig. 1.6, right).
- Out of the five operators contributing to HH, four enter other SM processes at LO, such as single Higgs, $t\bar{t}H$ or $t\bar{t}$ production. The remaining operator, O_6, can only be *directly* probed through HH. Hence, setting meaningful bounds on O_6 will most likely require that the other operators be themselves sufficiently constrained through a global fit of several SM processes [101].

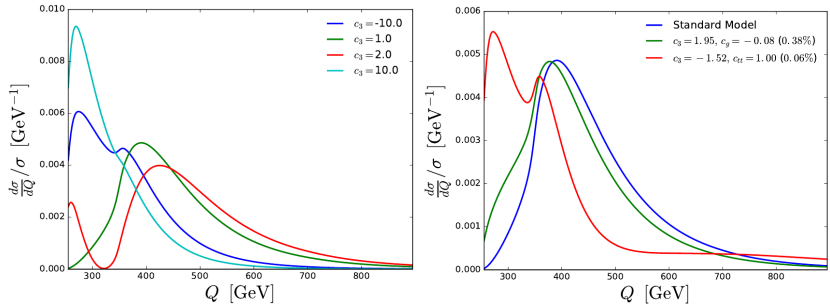

Figure 1.6. | EFT effects on the normalised distribution of m_{HH}, obtained in the HEFT at NNLO with a rescaling prescription to model the effects of finite top quark mass effects. Left: comparison of different values of the Higgs boson self-coupling, parameterised by $c_3 = 1 + \delta_\lambda$ (see (1.52)). Right: distributions obtained with different operator coefficients that yield cross sections almost equal to that in the SM. Deviations from the SM cross section are given in parentheses. The operator coefficients given correspond to $c_3 = 1 + \delta_\lambda$ (see (1.52)), $c_g = c_g/12$ (where the right-hand c_g is the one defined in (1.58)) and $c_{\mathrm{tt}} = 2c_2$ (with c_2 defined in (1.58)). Figures taken from Ref. [99].

Nonlinear EFT or anomalous couplings effects on HH

While above we made the assumption that the Higgs boson was part of ϕ, a scalar doublet of $\mathrm{SU}(2)_L$, the scalar Lagrangian (1.27)–(1.29) can also be written in a different parameterisation, in which H is invariant under $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ (see e.g. Ref. [102]). The SM can then be consistently modified by adding towers of new interactions involving terms of the type $c_n(\mathrm{H}/v)^n$. This so-called nonlinear EFT is suited if new, strong dynamics in the scalar sector appears at a scale $f \gtrsim v$, even if the new states have much larger masses. While the linear EFT is based on the assumption that all BSM effects are small, the nonlinear EFT can accommodate large deviations in the Higgs sector but evades constraints from electroweak precision measurements because it leaves the electroweak gauge sector of the SM unaffected. The use of the nonlinear EFT for parameterising BSM effects in HH production was advocated in Refs. [72, 95, 99]. The relevant part of the nonlinear Lagrangian is:

$$\mathcal{L}_{\mathrm{nonlin}} \supset -\kappa_\lambda \frac{m_{\mathrm{H}}^2}{2v}\mathrm{H}^3 - m_{\mathrm{t}}\bar{\mathrm{t}}\mathrm{t}\left(\kappa_{\mathrm{t}}\frac{\mathrm{H}}{v} + c_2\frac{\mathrm{H}^2}{v^2}\right) + \frac{\alpha_s}{12\pi}\left(c_g\frac{\mathrm{H}}{v} - c_{2g}\frac{\mathrm{H}^2}{2v^2}\right)G^a_{\mu\nu}G^{a,\mu\nu} \tag{1.58.}$$

In (1.58), five couplings have to be determined, but the relationship between contact terms with one or two Higgs bosons (compare with (1.55)), and thus the link with single-Higgs processes, is lost. The SM is recovered when taking $\kappa_\lambda = \kappa_{\mathrm{t}} = 1$ and $c_2 = c_g = c_{2g} = 0$. Here κ_λ and κ_{t} are seen as *anomalous couplings* that modify the Higgs cubic self-coupling and the top quark Yukawa coupling, respectively, while the other couplings lead to diagrams identical to those appearing the linear EFT, shown on Fig. 1.4. Apart from $O_{\mathrm{t}G}$ having no counterpart in (1.58), there are simple relations between the couplings in the two parameterisations.

In the literature, effects on HH arising from (1.58) have been studied by replacing all couplings in the diagrams of Figs. 1.2 and 1.4 by their anomalous values, leading to observables with a more complicated dependence on the parameters than what

is obtained with an expansion of the type (1.57). At LO, the effect of the anomalous couplings on the HH cross section σ are parameterised using the ratio:

$$R_{\mathrm{HH}} = \frac{\sigma^{\mathrm{LO}}}{\sigma^{\mathrm{LO}}_{\mathrm{SM}}} = A_1\kappa_{\mathrm{t}}^4 + A_2 c_2^2 + (A_3\kappa_{\mathrm{t}}^2 + A_4 c_g^2)\kappa_\lambda^2 + A_5 c_{2g}^2 + (A_6 c_2 + A_7\kappa_{\mathrm{t}}\kappa_\lambda)\kappa_{\mathrm{t}}^2 + (A_8\kappa_{\mathrm{t}}\kappa_\lambda + A_9 c_g\kappa_\lambda)c_2 + A_{10}c_2 c_{2g} + (A_{11}c_g\kappa_\lambda + A_{12}c_{2g})\kappa_{\mathrm{t}}^2 + (A_{13}\kappa_\lambda c_g + A_{14}c_{2g})\kappa_{\mathrm{t}}\kappa_\lambda + A_{15}c_g c_{2g}\kappa_\lambda, \quad (1.59.)$$

where the coefficients A_i have been obtained numerically. As observed above, since K-factors show little dependence on the EFT coefficients or on kinematics, the ratio R_{HH} can be applied in a straightforward way to higher-order SM results, so that:

$$\sigma^{(\mathrm{N})\mathrm{NLO}} \approx R_{\mathrm{HH}} \cdot \sigma^{(\mathrm{N})\mathrm{NLO}}_{\mathrm{SM}} \quad (1.60.)$$

In the case where one assumes $c_2 = c_g = c_{2g} = 0$, if we denote by $\mathcal{A}_\triangle$ and $\mathcal{A}_\square$ the (LO) amplitudes corresponding to the triangle and box diagrams of Fig. 1.2, the matrix element for HH production can be written schematically as:

$$|\mathcal{M}|^2 = \left|\kappa_\lambda\kappa_{\mathrm{t}}\mathcal{A}_\triangle + \kappa_{\mathrm{t}}^2\mathcal{A}_\square\right|^2 = \kappa_{\mathrm{t}}^4\left|\frac{\kappa_\lambda}{\kappa_{\mathrm{t}}}\mathcal{A}_\triangle + \mathcal{A}_\square\right|^2 \quad (1.61.)$$

This expression makes the quadratic dependence of the cross section on the Higgs boson self-coupling manifest. Furthermore, since the kinematic behaviour of the matrix element is only determined through the amplitudes $\mathcal{A}_\triangle$ and $\mathcal{A}_\square$, it is clear from (1.61) that any quantity related to the kinematics of HH (ranging from normalised differential cross sections to experimental acceptances) will only depend on the *ratio* $\kappa_\lambda/\kappa_{\mathrm{t}}$. Note that these observations are still valid beyond LO, even if no purely "box" and "triangle" amplitudes can be identified there.

Using (1.61) we can also deduce the relationship between κ_λ, κ_{t} on the one hand, and δ_λ, δ_y in (1.52) – (1.53) on the other. Contrary to what is claimed in Ref. [103], due to the different ways of expanding the observables (compare (1.57) with (1.59)), the relation cannot simply be given by $\kappa_\lambda = 1 + \delta_\lambda$ and $\kappa_{\mathrm{t}} = 1 + \delta_y$. Instead, a measurement of κ_λ and κ_{t} would have to be translated as:

$$\delta_\lambda = \kappa_{\mathrm{t}}(\kappa_\lambda - \kappa_{\mathrm{t}}) \quad (1.62.)$$

$$\delta_y = \kappa_{\mathrm{t}}^2 - 1 \quad (1.63.)$$

However if we allow e.g. for $c_g \neq 0$, there is no *exact* way to relate inferences on the couplings of (1.59) to the coefficients in the linear EFT. In the absence of any clear preference for the linear or nonlinear EFT, if one wishes to probe nonresonant BSM effects on HH production in a consistent way by considering all the relevant operators, it would thus seem necessary to perform two separate fits using either parameterisation.

Efficient modelling of nonresonant BSM effects on HH

While the sensitivity of double Higgs production to new physics effects makes it a tremendous place to look for deviations from SM predictions, it also represents a challenge for experimental analyses. Indeed, optimising an analysis for, setting limits on, or characterising a deviation due to EFT effects requires an efficient modelling of these effects: the naive method of generating Monte-Carlo samples with a fine grid in the five-dimensional space of EFT parameters is clearly not practical. Two approaches have been studied in this work; incidentally, both rely on the idea of matrix-element reweighting introduced in Sec. 1.2.

With the method suggested in Refs. [100, 103], a limited set of points in the parameter space are chosen as *benchmarks* representing the possible types of kinematical behaviours of HH in the nonlinear EFT presented above, with the parameterisation (1.59). Small parton-level samples are generated for a large number ($\mathcal{O}(1500)$) of regularly-spaced points in the five-dimensional parameter space [1]. Each of those samples is used to model the two-dimensional distribution of m_{HH} vs. $|\cos\theta^*_{\mathrm{CS}}|$, which as seen in Sec. 1.3 is sufficient to characterise the physics of the HH process. A measure of the similarity between any pair of points is computed using a two-sample goodness-of-fit (GoF) test of these 2D distributions, which is used by a *clustering* algorithm to form groups of points with homogeneous kinematical behaviours. For each of those groups (clusters), a representative point (benchmark) is chosen as the point which is the most similar to all other samples in the group. In this procedure, the number of resulting clusters has to be specified beforehand, and tuned by hand to yield clusters with satisfyingly similar distributions. It was deemed that 12 clusters represented a good balance between the number of benchmarks and the intra-cluster homogeneity. Figure 1.7 shows the distributions of m_{HH} for the samples contained in each of the 12 clusters. The coupling values corresponding to the benchmarks can be found in Refs. [72, 100, 103].

The small number of representative points given by this procedure means they can each be used to optimise an experimental search, so that the analyses are sensitive to a wide range of signal kinematics. However, results obtained with the benchmarks cannot trivially be used to produce inferences on the couplings of the EFT. First, there is no straightforward relationship between a point in the parameter space and the cluster to which it would be assigned. Second, the variability of the signal kinematics within a given cluster is still too large to allow the generalisation of a result obtained on a single benchmark to the whole cluster to which it belongs. For an attempt at an approximate recasting of experimental results on the benchmarks to arbitrary signal kinematics, based on the similarity measure defined in the clustering procedure, see Ref. [104].

While the chosen benchmarks correspond to discrete points in the EFT parameter space, it is possible to model the kinematics of HH for arbitrary values of the anomalous couplings by *reweighting* the events of the benchmark samples. The weights to be applied, obtained from the fully-differential cross section (i.e. the matrix element) of the process,

[1] The sampling could have been done in four dimensions, yielding better coverage of the parameter space with a same amount of points, since one dimension of the parameter space only affects the overall cross section and is thus irrelevant. What's more, upon close inspection the 1500 points sampled for the clustering procedure do not cover the parameter space consistently.

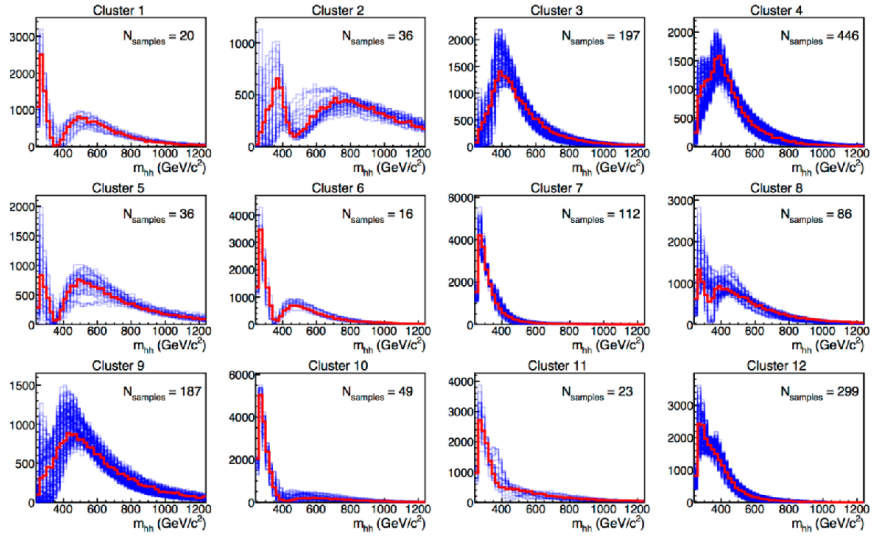

Figure 1.7. | Parton-level distributions of m_{HH} for the samples in each of the 12 kinematical clusters. The benchmark representatives of each cluster are indicated in red. The third cluster contains the SM point. Figure taken from Ref. [100].

are functions of the EFT couplings and can be written in the form (1.59), where the coefficients A_i are now functions of the event kinematics. In Ref. [104] the coefficients are defined in rectangular bins in the plane of m_{HH} vs. $|\cos\theta^*_{\mathrm{CS}}|$. Monte-Carlo samples for a large number of points in the parameter space are used to fit the dependence of the weights on the couplings, in each bin of the plane. This method is not exact, since an histogram-based estimation of a two-dimensional distribution is necessarily biased, and suffers from statistical fluctuations due to the finite size of the simulated samples used for the fit. Furthermore, the whole fitting procedure has to be redone every time a new model is used (for instance, when switching from the nonlinear to the linear EFT), or when changing the parameters used during sample generation (center-of-mass energy, top quark mass, etc.).

We here suggest a different approach, which is exact and can be implemented for any desired model for nonresonant HH production. Instead of fitting the formula (1.59) using Monte-Carlo samples, the *exact* weights can be obtained directly from the matrix element of the process, as described in Sec. 1.2. By "exporting" the matrix elements as a piece of C++ code using a plugin [105] we have written for MG5_aMC@NLO, it is straightforward to reweight the benchmark samples to any chosen hypothesis. Note that since we shall wish the normalisation of the HH samples to be unaffected by the reweighting, we still need to compute the total cross section as a function of the couplings, using (1.59). Another detail turns out to be of some importance in this procedure: when generating the samples used as basis for the reweighting, a *dynamic* renormalisation scale is used, and at LO this implies that the value of α_s used during sample generation depends on event kinematics. When evaluating the matrix elements of (1.43) on a particular event, the α_s value used for generating that event needs to be retrieved so that the reweighted

sample yields identical predictions as if one had generated an unweighted sample for that hypothesis. The effect of using an event-dependent α_s is showed on Fig. 1.8.

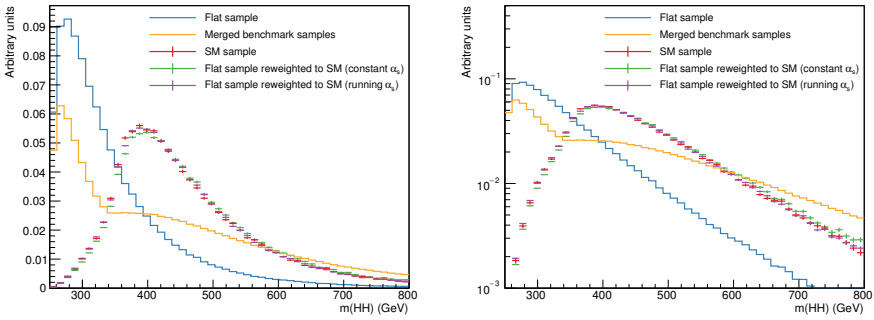

Figure 1.8. Parton-level normalised distributions of m_{HH} for the "flat" sample generated with unit matrix element (blue), the union of benchmark samples defined by the clustering procedure in Ref. [104] (orange), and a dedicated sample for the SM (red). The green sample corresponds to the flat sample reweighted to the SM using event-by-event evaluation of the LO matrix element for HH with full top quark mass dependence, using constant values for α_s. In contrast, the purple curve was obtained by using the same dynamic renormalisation scale as for the generation of the SM sample, and shows perfect agreement with the latter (within statistical uncertainties), whereas the fixed-scale distribution does not. The left and right plot show the same distributions and only differ by the ordinate scale.

Using the freedom provided by event reweighting, the problem of efficiently modelling EFT effects in HH is reduced to that of generating a sample that can be used as a basis for efficient reweighting. With the clustering procedure, that basis sample corresponds to the union of all 12 benchmark samples. Since the benchmarks, by construction, are meant to represent the full variety of possible kinematical behaviours in the nonlinear EFT, they provide a good basis for the subsequent reweighting. However, given that the HH kinematics are essentially characterised by a single quantity (m_{HH}), it would seem that starting from a sample that is *uniform* in m_{HH} should be much simpler. We have generated a so-called "flat" sample by setting $|\mathcal{M}|^2 \equiv 1$ during event generation, so that phase space and PDF factors are retained. Using such a basis, the reweighting ratio 1.43 simplifies to the matrix element in the numerator. As shown on Fig. 1.8, using a flat sample as a basis for reweighting presents no particular difficulty. In the tail of the distribution, for $m_{\mathrm{HH}} \gtrsim 500\,\mathrm{GeV}$, the behaviour of the flat sample is similar to that of the SM. It might seem that the reweighting is not efficient for low m_{HH} values, since a fair share of events receive comparatively low weights in that region. However, the SM is peculiar in being suppressed at threshold, and as can be seen on Fig. 1.7, the EFT leads to distributions with significantly larger contributions close to the threshold (corresponding to an enhancement of diagrams with an s-channel Higgs boson). Hence, it is beneficial for the purpose of reweighting that the flat sample contains a significant fraction of events at low m_{HH}. For the same reason, simply using the SM sample as basis would be ill-advised. Note that there is no way of constructing the "best" basis sample unless one specifies beforehand the fixed set of hypotheses to which one wishes

to reweight. However, choosing a unit matrix element (or another reasonable analytical bias function) certainly constitutes a simple and fast method to obtain a basis sample.

Event reweighting efficiently provides Monte-Carlo samples for any desired point in the EFT parameter space, but does not evade the necessity to iterate over the full samples to obtain predictions for the corresponding hypotheses. For the purpose of fitting or constraining the EFT couplings, it would be desirable that the dependence of any prediction on the couplings be available as a closed-form expression. This is a problem of *morphing* distributions, and has been addressed in Refs. [106,107]. We shall not enter into the details, but briefly show how that works out for the distribution of m_{HH}. We have implemented the reweighting of the flat sample for the linear EFT model of (1.52)–(1.56) provided by the authors of Ref. [94]. By computing the matrix element for specific values of the couplings, with simple arithmetics it is possible to isolate each contribution in the EFT expansion of observables (1.57). Given that five operators contribute to HH, any distribution can be decomposed into:

$$\underbrace{1}_{\text{SM}} \oplus \underbrace{5}_{\text{Interferences SM - operator}} \oplus \underbrace{5}_{\text{Op. squared}} \oplus \underbrace{\frac{5\cdot(5-1)}{2}}_{\text{Interferences op. - op.}} = 21 \text{ terms.} \quad (1.64.)$$

Once a distribution for each of those terms has been obtained, they can be combined analytically for arbitrary values of the coefficients c_i/Λ^2 using an expression of the form (1.57). Figure 1.9 shows the distribution of m_{HH} for each of the terms in (1.64), and for combinations of those terms assuming illustrative values of the EFT couplings. The morphing technique can also be employed with no reference to event reweighting, as it only requires as input a set of distributions obtained with specific values of the couplings. In the case of HH, that would require the generation of 21 different samples, which represents a reasonable increase compared to the 12 benchmark samples.

1.6. Resonant enhancement of double Higgs production

The existence of new, heavy states decaying to pairs of SM Higgs bosons is predicted by numerous BSM models [108–110]. Current experimental constraints on these models do not rule out the possibility of an early discovery of Higgs boson pair production. In this section we briefly describe some of these model families; a more complete review can be found in Ref. [72].

- The **Higgs singlet** or **Higgs portal** models are the simplest extension of the SM leading to resonant production of Higgs pairs [110–118]. They are based on the realisation that $(\phi^\dagger\phi)$ is a gauge singlet, and that a scalar, gauge-singlet field S can thus be added to the SM so that the only interactions between SM and BSM fields happen through the Higgs boson. Different versions of the singlet model have been developed, as the field S can be taken as real or complex and different conditions on the potential for S can be imposed. In the simplest version, the singlet is a real scalar and the full scalar potential, imposed to be invariant under $S \to -S$, is given by:

$$V(\phi, S) = \mu_1^2(\phi^\dagger\phi) + \mu_2^2 S^2 + \lambda_1(\phi^\dagger\phi)^2 + \lambda_2 S^4 + \lambda_3(\phi^\dagger\phi)S^2 \quad (1.65.)$$

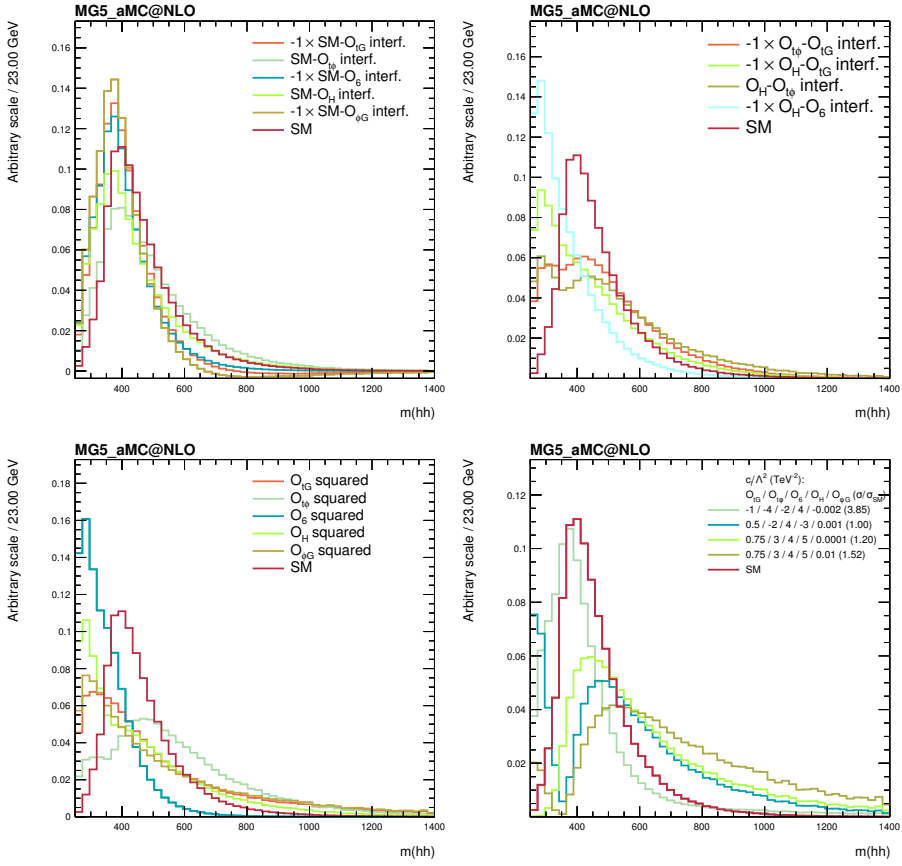

Figure 1.9. Parton-level normalised distributions of m_{HH} obtained through event-by-event matrix-element reweighting of a flat sample. The top left plot shows the interferences between each operator in the linear EFT and the SM; the top right plot shows a subset of all possible interference terms between pairs of operators, and the bottom left plot shows the contributions from the squaring of the operators. Some interference terms contribute with a negative sign to the total observable (using the conventions of (1.47)–(1.51), but with a switched sign for $O_{\mathrm{t}\phi}$), and are indicated with a factor "−1×" in the legend. Distribution obtained when combining the different terms using various values for the EFT couplings are shown on the bottom right plot. The values used for each operator coupling are indicated in the legend, along with the ratio between the total cross section obtained with these couplings, and the SM cross section. The SM prediction is included in every plot for reference.

If S has a non-zero VEV, $\langle S \rangle$, we can write $S = (s + \langle S \rangle)/\sqrt{2}$. Denoting by ϕ_0 the neutral SM scalar expanded around its VEV (v), we obtain the mass eigenstates h and H as mixtures of the original fields:

$$\begin{pmatrix} \mathrm{h} \\ \mathrm{H} \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \phi_0 \\ s \end{pmatrix} \tag{1.66.}$$

The field h is identified with the SM Higgs boson, and the new state H can decay into hh if $m_\mathrm{H} > 2m_\mathrm{h}$. The free parameters of this model are m_H, the mixing angle $\cos\alpha$, and the ratio between the VEVs, $\tan\beta = v/\langle S \rangle$, not to mention the measured values for m_h and v. The mixing between S and ϕ_0 leads to deviations from SM predictions for the couplings between the light Higgs boson and SM particles, and the lack of large such deviations indicates that $\alpha \approx 0$. While this suppresses the decays of H into SM particles, the branching ratio for $\mathrm{H} \to \mathrm{hh}$ can be quite large, $O(20-30\%)$, and an enhancement of more than one order of magnitude of the di-Higgs cross section compared to the SM is still possible. Note that even without visible resonant $\mathrm{H} \to \mathrm{hh}$ production, SM-Higgs pair production in singlet models can be affected by non-standard values of the hhh coupling.

- Other relatively simple extensions of the SM are the different realisations of the **2 Higgs doublet model (2HDM)**, which postulate the existence of two complex scalar $\mathrm{SU}(2)_L$ doublets, see e.g. Refs. [119,120]. Supersymmetric models such as the **minimal supersymmetric standard model (MSSM)** [121,122] also feature two such doublets. After spontaneous symmetry breaking, the physical states are two charged Higgs bosons $\mathrm{H}^\pm$, a pseudoscalar A, and two scalars h and H (with $m_\mathrm{H} > m_\mathrm{h}$). Depending on how the doublets are assumed to couple with the fermions, one obtains e.g. the type-I, type-II, "lepton-specific" or "flipped" 2HDMs. The coupling of the heavy scalar H with the light state h, assimilated with the SM Higgs boson, is obtained from the mixing angle α between the neutral states, and the ratio $\tan\beta = v_2/v_1$ between the VEVs of the neutral components of the two doublets. The decay channel $\mathrm{H} \to \mathrm{hh}$ has been extensively studied [67,123,124] but experimental constraints on 2HDMs are strong, and this channel stands in competition with other decay modes such as $\mathrm{H} \to \mathrm{ZA}$, $\mathrm{H} \to \mathrm{AA}$ or $\mathrm{H} \to \mathrm{t\bar{t}}$. Nevertheless, unexplored regions of the parameter space leave open the possibility of observable effects in Higgs pair production [124,125].
- In the **Georgi–Machacek (GM)** model [126,127], two scalar triplets of $\mathrm{SU}(2)_L$ are introduced in addition to the usual doublet ϕ. Among the several resulting physical fields, we again obtain two scalars h and H, with h identified as the SM Higgs boson. The trilinear coupling hhh is modified and the decay $\mathrm{H} \to \mathrm{hh}$ is allowed as soon as $m_\mathrm{H} > 2m_\mathrm{h}$, leading to possibly large enhancements of Higgs pair production [128].
- **Randall–Sundrum (RS)** models and their offsprings [129–135] postulate the existence of an additional, warped dimension of space (WED), and have been proposed in an attempt to solve the hierarchy problem. The extra dimension separates two flat four-dimensional boundaries, the so-called Planck/UV and TeV/IR branes. By locating the Higgs field on the TeV brane, the warping of the fifth dimension generates a large hierarchy between the Planck and weak scales. Free parameters of the theory are the warp factor k and the size of the extra dimension L; the required

hierarchy is obtained when $kL \approx 35$.
In the original RS1 model, all SM fields are restricted to the TeV brane, and only the graviton is allowed to propagate through the 5D bulk. The apparent weakness of gravity is then due to the localisation of the graviton close to the Planck brane. The main experimental prediction of RS1 is the existence of so called Kaluza–Klein (KK) excitations of the graviton. The lightest KK mode would have mass around the weak scale, and preferentially decay to pairs of fermions, gluons or photons.
In "Bulk" models, the SM fields are pulled into the bulk, with only the Higgs field localised on the TeV brane. Interestingly, this provides an explanation for the flavour puzzle: heavy fermions are simply localised closer to the TeV brane (and thus to the Higgs boson) than light fermions. In contrast to RS1, graviton decays to pairs of Higgs bosons are sizable in bulk models, whereas decays to leptons, light quarks, or massless bosons are suppressed. Bulk models also predict the existence of KK-excitations of SM particles (except H), but these do not give rise to HH final states.
Besides spin-2 tensor modes (gravitons), WED models feature scalar excitations of the metric, i.e. fluctuations of the size of the extra dimension. The corresponding particles are massive scalars called *radions*, whose couplings to SM fields are suppressed by a scale $\Lambda_R \approx O(\text{TeV})$, related to k, L, and the 5-dimensional Planck mass M_5. Radions are produced predominantly in gluon fusion, have a narrow width provided their mass is $m_R \lesssim O(\text{TeV})$, and decay mainly into pairs of W, Z and H bosons.

To the extent that the extra *scalar* state decaying to pairs of SM-like Higgs bosons has a narrow width, the production and decay kinematics of that state do not depend on the specifics of the model. Thus, experimental limits on the production cross section of such states (multiplied with their branching ratio to Higgs pairs) are model-independent.

For what concerns the production of spin-2 states, the production mode in proton collisions (gluon or quark fusion) and the specific form of the coupling with quarks or gluons has in principle some impact on the experimental sensitivity, and it is necessary to choose a benchmark model. In this work, we have considered the gluon-induced production of KK-gravitons. As it turns out, the experimental sensitivity for that particular particle does not strongly differ from what is obtained for spin-0 resonances, which indicates that while the results for these spin-2 resonances cannot be interpreted in a strict model-independent way, the searches are not blind to other possibilities.

If the new states introduced by the models listed above are too heavy to be produced directly, their indirect effects on HH production can be described and probed using the EFT approach described in the previous sections, see e.g. Ref. [110]. Another possibility, which we have not addressed, arises when the new degrees of freedom are too light to decay to H pairs, but nonetheless have an impact on HH [136–138], leading to enhanced rates. Indeed, new particles might circulate in the loops of diagrams in Fig. 1.2, or yield an additional s-channel contribution interfering with the SM amplitude, among different possibilities. In addition, some models predict the production of pairs of new particles that each decay to a Higgs boson and additional visible or invisible particles (see e.g. Ref. [139]), leading to HH + X final states where X can be SM particles or other new states that escape the detector undetected. However, all these modifications to HH

production are sensitive to the specifics of the BSM models considered, involving several unknown new parameters. Given our goal of providing experimental results that can be interpreted in a large variety of models, we have not explicitly targeted such effects in this work.

1.7. Review of current experimental results

In this section we give a short overview of the current sensitivity to Higgs pair production and the Higgs boson self-coupling. Higgs pair production has been probed at $\sqrt{s} = 8$ TeV (Run 1) and 13 TeV (Run 2) by the ATLAS and CMS collaborations. Given the Higgs boson's rich phenomenology, HH events are scattered across numerous decay channels with different experimental sensitivities. Table 1.3 shows the expected and observed limits[1] on the SM cross section obtained by ATLAS and CMS in different channels. The results obtained by CMS in the $b\bar{b}VV$ final state corresponds to those presented in this thesis and described in Chap. 3. Due to the dependence of the experimental sensitivity on the Higgs boson self-coupling, these limits cannot be used directly to constrain its value. Hence, CMS has explicitly set limits on σ_{HH} as a function of the self-coupling modifier, κ_λ. Using the most sensitive result obtained in the $b\bar{b}\gamma\gamma$ final state [140], this translates into a range of allowed values for κ_λ of about [−11,17] at 95% CL ([−8,14] expected).

Table 1.3. Expected and observed limits on $\sigma_{HH}/\sigma_{HH}^{SM}$ obtained by ATLAS and CMS in different decay channels. The integrated luminosity used for each analysis is given to ease the comparison of the figures. The ATLAS Run 1 result was obtained by combining analyses in the $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$, $b\bar{b}\tau\tau$ and $\gamma\gamma WW$ final states. The CMS Run 1 combined limit was achieved using the $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$ and $b\bar{b}\tau\tau$ channels. SM branching ratios of the Higgs boson are assumed for rescaling the results in different decay channels to the total HH cross section. All results are statistically compatible with SM expectations.

Experiment	Channel	$\int \mathcal{L} dt$ (fb^{-1})	Exp. /	obs. limit
ATLAS (Run 1)	Combined [141]	20.3	48	70
ATLAS (Run 2)	$b\bar{b}b\bar{b}$ [142]	27.4	21	13
	$b\bar{b}\gamma\gamma$ [143]	3.2	162	117
	$\gamma\gamma W(\ell\nu)W(jj)$ [144]	13.3	386	750
CMS (Run 1)	Combined [145]	≤ 19.7	47	43
CMS (Run 2)	$b\bar{b}\gamma\gamma$ [140]	35.9	19	24
	$b\bar{b}\tau\tau$ [146]	35.9	25	30
	$b\bar{b}VV(\ell\nu\ell\nu)$ [147]	35.9	89	79
	$b\bar{b}b\bar{b}$ (boosted) [148]	35.9	120	188
	$b\bar{b}b\bar{b}$ [149]	2.3	308	342

ATLAS and CMS have also probed the resonant production of Higgs boson pairs, in the same final states as those shown in Tab. 1.3 [140–147, 150–152]. Limits are placed

[1] These notions will be rigorously defined in Sec. 2.4.2.

on the product of the production cross section of spin-0 and/or spin-2 narrow-width resonances X with the branching fraction for $X \to HH$, as a function of the mass of the resonance, m_X. The limits range from about 1 pb at the kinematic threshold of 260 GeV for the production of narrow resonances decaying to Higgs boson pairs, down to O(fb) at $m_X = 4$ TeV.

The sensitivity to HH production is going to improve dramatically as more data are collected by the LHC experiments, as shown by preliminary studies on the HL-LHC reach (see Sec. 2.1.2, Chap. 4 and Refs. [153–158]). However, the weakness of the bounds on κ_λ quoted above raises two (not entirely unrelated) questions:

1. What are the largest values that κ_λ could take in any conceivable scenario?
2. Can κ_λ be bound through other means than HH production?

The first question has been addressed in Ref. [159] using two different approaches. A model-independent bound can be set through perturbativity arguments, such as the requirement that the scattering amplitude for $HH \to HH$ remains within unitarity bounds, which yields the allowed range of $|\kappa_\lambda| \lesssim 6.5$. Alternatively, specific BSM models inducing large deviations in the Higgs boson self-coupling without conflicting with measurements of single Higgs production can be studied. By scanning the allowed parameter space of these models, taking into account perturbativity bounds, vacuum stability, precision measurements or results from direct searches, it is found difficult to generate values of κ_λ larger than a few.

The second question can be answered positively by realising that the trilinear coupling contributes to NLO electroweak corrections to single Higgs boson production and decay [160–163], and to NNLO corrections to precision electroweak observables [164,165]. In both cases, the range of allowed values for κ_λ turns out to be comparable with the direct bounds from HH production, both when considering present experimental results on single and double Higgs production, or when using the expected precision attainable at the HL-LHC. However, these studies rely on the assumption that the only deviation from SM predictions in single Higgs processes is due to an anomalous Higgs self-coupling, contrary to many BSM scenarios where other Higgs boson interactions are simultaneously affected. In Ref. [101] an attempt was made at a global fit of nine effective operators in the linear EFT to single Higgs boson measurements (branching ratios and total rates), taking into account the corrections to single Higgs processes due to an anomalous Higgs trilinear coupling. It was shown that, under these hypotheses, measurements of single Higgs boson processes alone are degenerate and cannot be used to place meaningful indirect bounds on κ_λ, and that including a measurement of Higgs pair production is necessary to lift the degeneracy in the fit[1]. These observations hold also when considering differential measurements of single or double Higgs production. Note that if the analysis is repeated in the nonlinear EFT, the connection between H and HH processes is partially lost and the bounds on the self-coupling become significantly weaker. In conclusion, observing and characterising Higgs pair production remains a crucial experimental task for our understanding of electroweak-scale phenomena.

[1] Conversely, this implies that the precision on Higgs boson couplings that can be attained in global fits is limited by the lack of strong constraints on the Higgs trilinear coupling.

Chapter

2. The CMS experiment: Event reconstruction and Data analysis techniques

The previous section was dedicated to the description of the current best model of the fundamental interactions between elementary particles, and of the methods used to generate predictions that can be confronted with reality. We now consider the other side of the story and explain how to test these predictions in controlled and repeated conditions, and how to quantify the level of agreement between calculations and experimental data. We start with a brief description of the Large Hadron Collider (LHC) and of the CMS detector, our main experimental tools. The methods used to make sense of the data recorded by CMS are described next. Finally, we introduce some of the techniques required to extract and characterise the rare processes we are after.

2.1. The Large Hadron Collider

High energy collisions, a proven technique to study the Standard Model, are provided in a controlled and reproducible way by the Large Hadron Collider (LHC) [166]. The LHC, located in a 26.6 km–long tunnel under the French–Swiss border, at the CERN laboratory next to Geneva, is the largest and most powerful particle collider ever built. It is a superconducting synchrotron, able to accelerate, circulate and collide protons at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. A complex accelerator chain, depicted on Fig. 2.1, is required to produce beams consisting of more than 2000 "well-behaved" *bunches* of about 1.1×10^{11} protons with an energy of 6.5 TeV.

Protons are extracted from a plasma obtained by heating H_2 gas. They are then accelerated by the LINAC2 up to an energy of 50 MeV, and injected into the Proton Synchrotron Booster (PSB), consisting of four stacked rings each accelerating a bunch of protons to 1.4 GeV. The Proton Synchrotron (PS) is filled with 4+2 of these bunches, splits each of them into three, smaller bunches, and accelerates those to 25 GeV. Two further splits provide *trains* of 72 bunches separated by 25 ns (7.5 m), which are injected into the Super Proton Synchrotron (SPS). The SPS, which can accommodate four of these trains, brings them to an energy of 450 GeV and passes them to the LHC as two counter-circulating beams of up to 2808 bunches each.

Due to a leak in the SPS beam dump system detected in 2016, the number of bunches that could be accelerated by the SPS and injected into the LHC had to be limited to 96 and 2220, respectively. To mitigate the impact on the amount of data that could be delivered to the experiments, an alternative injection scheme has been used starting in summer 2016: the batch compression merging and splitting (BCMS) scheme [168].

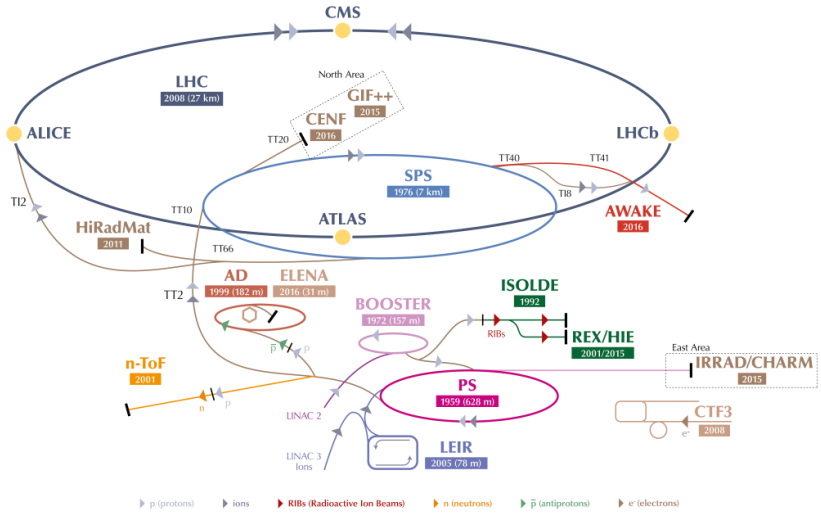

Figure 2.1. | The CERN accelerator complex and some of the related experiments, as of 2018. Figure taken from Ref. [167].

Instead of using only six bunches from the PSB to fill the PS, all eight bunches from the two PSB cycles are kept. These bunches are first compressed and merged into four; the acceleration and splitting steps then proceeds as in the nominal scheme, yielding 48 bunches with increased brightness (higher intensity, reduced transverse emittance) with respect to the nominal scheme.

The beams in the LHC each circulate in a pipe kept at an ultra-high vacuum ranging from 10^{-6} to 10^{-11} mbar, necessary to minimise beam losses. The LHC consists of 1232 superconducting NbTi dipole magnets, cooled to 1.9 K using superfluid helium, which produce the field of 0.54 to 7.7 T needed to bend the beams around a circular trajectory. The beams are kept focused by 392 quadrupole magnets providing a maximum gradient of 223 T/m, and further corrected by higher-order fields (sexta-, octo- and decapoles). The LHC features a "twin-bore" design, in which the two beam pipes are contained within a common cold mass. The magnet coils are retained by non-magnetic collars, enclosed by an iron flux return yoke, itself maintained within a vacuum vessel for thermal insulation. Eight superconducting radio-frequency (RF) cavities (per beam) increase (*ramp*) the protons' energy to 6.5 TeV using standing electromagnetic waves of about 400.79 MHz with a peak field strength of 5.5 MV/m. Their frequency has to be adjusted by less than 1 kHz during the ramp to match the slight increase in velocity of the protons as they gain momentum. Every revolution, each cavity increases the protons' energy by 60.6 keV, resulting in a ramp time of about 20 min. When the ramp is over, the RF system also compensates for small energy losses due to synchrotron radiation (7 keV/turn). The beams cross at four interaction points (IPs) but are kept separated during injection and ramp by dedicated dipole magnets. Before being brought into collision, the beams are *squeezed* at the IPs by sets of quadrupole triplets to maximise the interaction rate.

Particle detectors are installed around the IPs to detect the debris coming out of the

collisions. These experiments are:

- ATLAS [169] and CMS [170]: two large, general-purpose detectors.
- LHCb [171], optimised for the study of B mesons.
- ALICE [172], specialised in the study of the quark-gluon plasma through the analysis of collisions between lead ions.
- Three smaller, dedicated experiments are also present at the LHC: LHCf [173], MOEDAL [174], and TOTEM [175]. They share a collision point with the previous, main four detectors, and only record a fraction of the scattered particles.

The beams are kept in a stable, colliding configuration for several hours, after which they are *dumped* into massive graphite absorbers and the cycle can start again. In ideal conditions, each *fill* of the LHC lasts from 10 to 16 hours, and the time between the beams are dumped and collisions can be declared again is about two hours.

2.1.1. Luminosity and pileup

The instantaneous luminosity $\mathcal{L}$ is the proportionality factor between the rate of a scattering process and the cross section of that process. The luminosity should be as high as possible to maximise the amount of data available for analysis. At particle colliders such as the LHC, it depends on the characteristics of the beams and can be expressed as:

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

where, along with ultimate values for the LHC during 2016 data-taking:

- $f_r = 11.245\,\text{kHz}$ is the proton revolution frequency,
- $N_b = 2208$ is the number of *colliding* bunches per beam,
- $N_p = 1.15\times10^{11}$ is the number of protons per bunch,
- $\sigma_{x,y} = 11\,\mu\text{m}$ is the standard deviation of the beam density profile in the transverse plane (here assumed to be equal for both beams and to follow a Gaussian distribution), at the IP and at $\sqrt{s}$ =6.5 TeV,
- $\epsilon = 0.3\,\text{nm}$ is the (un-normalised) beam emittance, and
- $\beta_{x,y}^* = 40\,\text{cm}$ is the minimum, at the IP, of the betatron functions $\beta_{x,y}(s)$ describing the envelope of the proton trajectories, parameterised by the position s along the ring. The waist profile of the betatron function around the IP ("minibeta insertion") is created by the squeeze. By definition, we have $\sigma_{x,y} = \sqrt{\epsilon\beta_{x,y}}$.

In practice, additional factors will play a role, such as:

- The beams do not collide head-on but with a small crossing angle Φ, necessary to suppress parasitic interactions between leading and trailing bunches (collisions and long-range beam-beam electromagnetic interactions). The reduction factor can be approximated by

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

where σ_s is the bunch length. In 2016, the design crossing angle of 370 µrad was reduced to 280 µrad after the switch to BCMS, increasing the luminosity by about 15% (with $F \approx 0.7$).

- The beams might not cross at the true waist of the minibeta insertion.
- A slight misalignment of the beams would reduce the overlap between the bunches and hence the luminosity.
- The beam profile might not be strictly Gaussian, resulting in deviations from (2.1).

Plugging the numbers into (2.1) yields $\mathcal{L} \approx 1.6 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, close to the 2016 record of $1.5 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, itself significantly above the design luminosity of the LHC of $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ ($= 10\,\mathrm{nb}^{-1}\mathrm{s}^{-1}$).

The total amount of data delivered or recorded during a period of time T is measured in terms of *integrated* luminosity:

$$\mathcal{L}_{\mathrm{int}} = \int_0^T \mathcal{L}(t)\,\mathrm{d}t \tag{2.3.}$$

Naturally, we do not have $\mathcal{L}_{\mathrm{int}} = T \cdot \mathcal{L}$ since the luminosity is not constant during a fill. The beams "burn off" (reducing N_p), mostly due to the collisions in the interaction points, but also due to losses around the LHC ring (interaction with residual gas in the beam pipe, with collimators, ...), leading to an exponential decay of the luminosity: $\mathcal{L}(t) = \mathcal{L}_0 \exp(-t/\tau)$, with $\tau \sim 24\,\mathrm{h}$ and where $\mathcal{L}_0$ is the so-called *peak* luminosity at the start of the fill. In addition, the machine needs a few hours between the moment the beams are dumped, and collisions ("Stable Beams") can be declared again: typically two to six hours, or much longer in case of technical problems. The ratio $\mathcal{L}_{\mathrm{int}}/(T \cdot \mathcal{L}_0)$ is the so-called Hübner Factor, and was about 0.5 in 2016, which constitutes an exceptional achievement by the CERN beams department.

Measuring precisely the luminosity is essential for all the LHC experiments, as this quantity enters virtually every data analysis. This measurement happens in two distinct steps [176] and is based on the following expression for the luminosity:

$$\mathcal{L} = \frac{\mu f_r}{\sigma_{\mathrm{inel}}} = \frac{\epsilon \mu f_r}{\epsilon \sigma_{\mathrm{inel}}} = \frac{\mu_{\mathrm{vis}} f_r}{\sigma_{\mathrm{vis}}} = \frac{R}{\sigma_{\mathrm{vis}}} \tag{2.4.}$$

Here μ is the mean number of inelastic pp interactions per bunch crossing and σ_{inel} is the total inelastic interaction cross section, which for pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ amounts to $\sigma_{\mathrm{inel}} \approx 80\,\mathrm{mb}$. The overlap of multiple collisions in the same bunch crossing is referred to as *pileup*. The average pileup in 2016 data is $< \mu >= 27$, and this number is expected to grow together with increased luminosity in the LHC since the only terms not affecting pileup in (2.1) are f_r and N_b, which cannot be increased beyond design. Other terms in (2.4) are ϵ, the efficiency with which an interaction is recorded as an "event" by a detector, μ_{vis} and σ_{vis}, the number of interactions and interaction cross section visible by the detector, and R, the event rate recorded by the detector. The definition of

"event", and the corresponding efficiency, depend on the method chosen to measure the luminosity. Precise measurements of R over time allow a *relative* determination of the instantaneous luminosity during data taking. Several independent detectors are usually employed to that end, in order to ensure consistency and stability of the measurements. An *absolute* calibration of the luminosity is still required for determination of σ_{vis}. At the LHC, this is mostly achieved through dedicated data-taking runs: van der Meer (VdM) scans [177]. Measuring R while scanning over the vertical and horizontal separation between the beams at the IP yields a precise measurement of the size of the luminous region (beam spot), which enters (2.1) as the term $2\sigma_x\sigma_y$. Combined with the knowledge of N_p, which can be precisely measured, this fixes $\mathcal{L}$ and hence σ_{vis} through (2.4). These runs are carried out once per year under special conditions (low rate and low pileup, beam profile as close to Gaussian as possible).

2.1.2. LHC timeline and data-taking periods

The LHC is a long-term project: planning began as early as 1984, construction commenced in 1998, and operations were launched in 2008 and are due to last until 2035. The accelerator and detectors being extremely complex machines, they cannot be operated continuously but have to undergo regular maintenance. Thus, data taking is divided into "runs", lasting several years and separated by long shutdown periods. In addition, no data is delivered during winters, facilitating repairs or upgrades.

The first run ("Run 1") of the LHC was due to start in 2008, but was delayed until 2010 because of an incident during commissioning. While the design beam energy of the LHC was 7 TeV, for safety reasons it had to be operated at 3.5 TeV in 2010 and 2011, during which 6.1 fb^{-1} of data were delivered to each ATLAS and CMS. In 2012, the energy could be increased to 4 TeV, and 23.3 fb^{-1} of data were delivered, enabling the discovery of the Higgs boson. The LHC then underwent a two-year-long repair program that allowed it to be operated at 6.5 TeV during its second run ("Run 2"), which started in 2015 and is still ongoing. 4.2 fb^{-1}, 40.8 fb^{-1} and 49.3 fb^{-1} were delivered in 2015, 2016 and 2017, respectively. Run 2 will last until 2018, at which point almost 150 fb^{-1} are expected to have been delivered. The next run, from 2020 to 2023, should provide an additional 300 fb^{-1}. A further – and final – step in the LHC programme is a major upgrade of CERN's whole accelerator complex, after which the LHC will be operated at 7 TeV and with 10 times higher instantaneous luminosity. This "High-Luminosity" LHC (HL-LHC) is scheduled to deliver 3000 fb^{-1} from 2026 to 2035. Figure 2.2 shows a comparison of the integrated luminosity collected during each year of Run 1 and Run 2 so far. The results presented in this work have been obtained using data recorded during 2016.

2.2. The CMS experiment

The Compact Muon Solenoid (CMS), depicted on Fig. 2.3, is a multi-purpose particle detector built around one of the LHC's IPs, at access point nr. 5. Its role is to detect the particles produced during LHC collisions, and measure their kinematic properties (direction, momentum or energy). The resulting data is used for a wide variety of physics analyses. Its function dictates its geometry and architecture: CMS is arranged

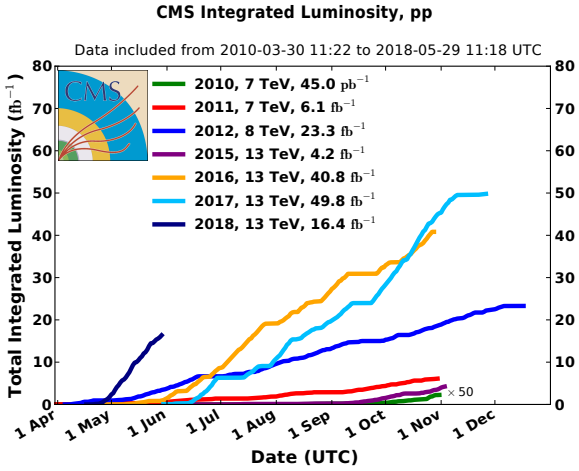

Figure 2.2. | Integrated luminosity delivered by the LHC to CMS versus time for each year of data taking during Run 1 and Run 2. Taken from Ref. [178].

as a cylinder, 21.6 m long and 14.6 m high, built symmetrically around the IP, in order to be hermetic and record as many of the outgoing particles as possible. The cylinder is divided into a barrel defining the *central* acceptance region, and two endcaps covering the *forward* regions. It consists of 10 different subdetectors, intertwined as the layers of a 12500 t onion, each responsible for the detection and characterisation of different particles, as well as a giant superconducting solenoid providing an homogeneous magnetic field of 3.8 T to bend the trajectories of charged particles. The subdetectors can be categorised into trackers and calorimeters. The former measure the direction and curvature of tracks created by charged particles, their curvature giving access to their momentum. The trackers in CMS are divided into an inner tracking system, close to the IP, consisting of the Pixel and the Strip tracker, and an outer tracking system, located outside of the solenoid and responsible for the detection of muons. The muon tracker is embedded within a steel yoke which ensures that the magnetic flux lines are closed. This arrangement provides a magnetic field of up to 2 T in the outer tracker. The electromagnetic and hadron calorimeters, both contained within the magnet volume, are responsible for measuring the energy of different types of particles, and provide a rough estimate of their direction. Finally, CMS depends on a highly efficient *trigger* system, since at a nominal collision rate of 40 MHz, about 40 TB of data are produced each second: way more than what can be stored and analysed. The trigger is responsible for selecting in real time the 0.001% of collisions, or events, that are deemed interesting enough to be stored for further analysis.

We repeat and extend here the conventions given at the beginning of Chap. 1, to clarify their relationship with the geometry of CMS. The coordinate system used by CMS is a right-handed orthonormal system with its origin at the nominal IP, its x-axis directed toward the centre of the LHC ring, its y-axis pointing vertically upwards, and its z-axis thus being tangent to the counter-clockwise (when viewed from the top) circulating beam trajectory, pointing away from Lake Geneva. These define in turn the azimuthal

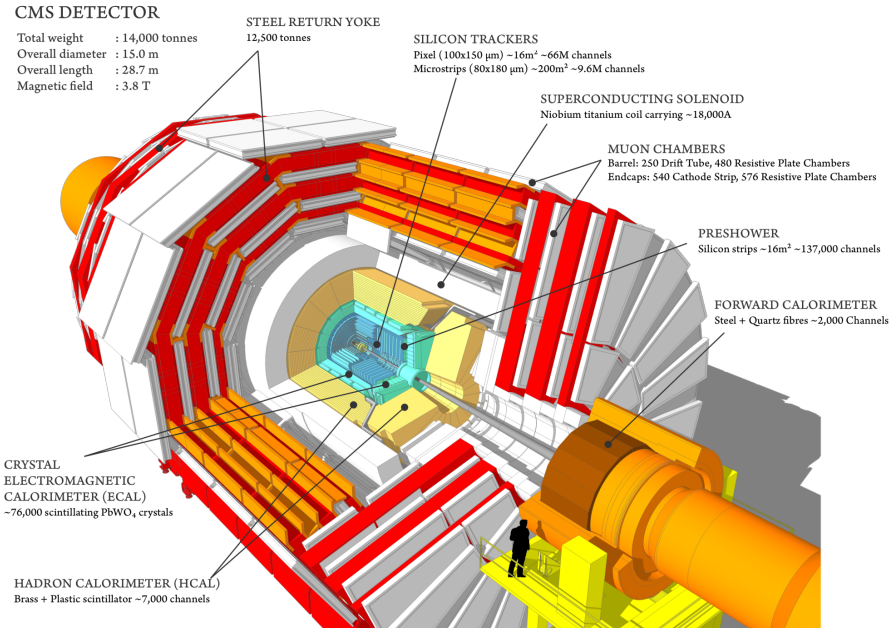

Figure 2.3. The CMS experiment and its various subdetectors. Figure taken from Ref. [179].

angle ϕ, measured from the x axis in the x-y-plane (transverse plane), and the polar angle θ, measured from the z-axis. The distance from the z-axis is defined as $r = \sqrt{x^2 + y^2}$. The transverse momentum p_T is computed as the magnitude of the projection of the measured momentum $\vec{p} = (p_x, p_y, p_z)$ in the transverse plane; the "transverse energy" of a particle with energy E is given by $E_T = \sin\theta \cdot E$. In the following, the direction of particles produced at the IP will be quoted as a function of ϕ and the pseudorapidity:

$$\eta = -\ln \tan \theta/2 \tag{2.5.}$$

For massless particles, this quantity is equal to the rapidity y. From this point on, angular distances are defined using the pseudorapidity, i.e.:

$$\Delta R = \sqrt{(\eta_1 - \eta_2)^2 + (\Delta\phi)^2}, \text{ with} \tag{2.6.}$$

$$\Delta\phi = \min\left(|\phi_1 - \phi_2|, 2\pi - |\phi_1 - \phi_2|\right). \tag{2.7.}$$

The subdetectors composing CMS, as well as the trigger and data acquisition systems, are briefly described in the following sections. More information about the CMS detector can be found in Ref. [170].

2.2.1. Inner tracker

Closest to the IP lies the inner tracking system. It is responsible for measuring the trajectories of charged particles, as well as for the reconstruction of interaction and decay vertices. The efficient tracking of several hundreds of charged particles produced

during each bunch crossing, at a rate of 40 MHz, requires a high granularity and a low response time. The resulting power density of detector electronics imposes the use of an efficient cooling system. This in turn conflicts with the need to minimise the number of nuclear interactions, multiple scatterings and bremsstrahlung the particles undergo when traversing the detector, since these processes spoil the precision and efficiency of the track reconstruction. Given the high particle flux (100 MHz/cm^2 at $r = 4$ cm), the system must also be sufficiently radiation-hard to be operated efficiently during its lifetime of more than 10 years. The choice of silicon detector technology for the CMS tracker represents a good compromise between these conflicting constraints. The CMS inner tracker has a diameter of 2.5 m and a length of 5.8 m, extending over $|\eta| < 2.5$, and it is composed of a pixel and a strip detector, as depicted on Fig. 2.4.

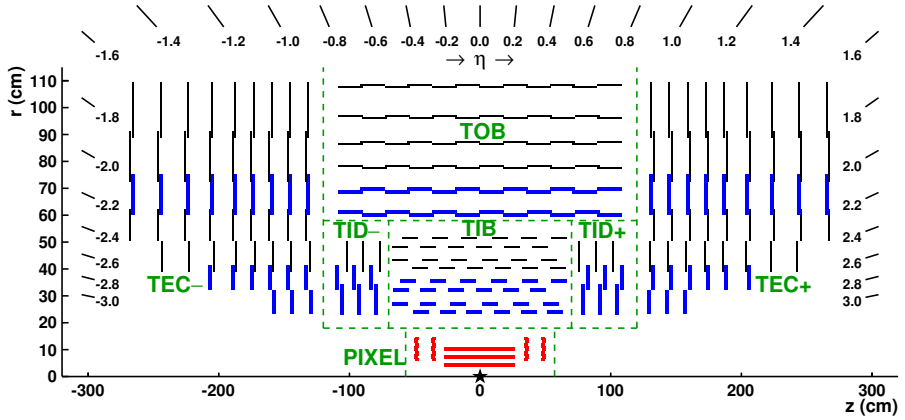

Figure 2.4. | Cross section in the $r-z$ plane of the CMS tracker layout and its partitions. The IP is depicted by a star. Thick red (pixel) and blue (strips) lines indicate modules providing 3D measurements of hit positions. Figure taken from Ref. [180].

The pixel detector features three barrel layers at $r = 4.4$ cm, 7.3 cm and 10.2 cm, and two disks in each of the forward regions, located at $|z| = 34.5$ cm and 46.6 cm and covering $6 < r < 15$ cm. Its 1440 modules contain 66 million pixel cells of 100×150 µm^2 area and 285 µm thickness, providing measurements of hit positions in two directions (z/ϕ in the barrel, r/ϕ in the endcaps) with a resolution of about 20 µm. Precise knowledge of module positions yields the missing component (r in the barrel, z in the endcaps). The occupancy in the pixel detector during 2016 data-taking, i.e. the mean number of particles hitting a cell per bunch crossing, was less than 6×10^{-4} thanks to the small pixel size.

The strip tracker consists of several subdetectors, with a total of 15148 modules of different shapes containing about 9.3 million strips. The inner strip tracker is composed of the tracker inner barrel (TIB), with four barrel layers, and the tracker inner disks (TIDs), with three disks at each end. Their strip sensors have a thickness of 320 µm and a pitch varying from 80 µm to 141 µm, depending on their distance from the IP. The number of measurements per track is extended by the outer strip tracker, itself composed of the tracker outer barrel (TOB), with six layers, and the tracker endcaps (TECs), with

two sets of nine disks. Detector thickness and strip pitch in the outer tracker vary from 320 µm to 500 µm, and from 97 µm to 184 µm, respectively. The strips are up to 20 cm long and are arranged parallelly to the beam line in the barrel, and radially in the endcap disks. Hence, the strip tracker can only measure precise hit positions in two directions once module position is taken into account: r and ϕ in the barrel, and z and ϕ in the endcaps. To supplement the pixels and provide additional measurements in the missing directions, the first two layers and rings of TIB, TID and TOB, as well as three rings of the TECs contain *stereo* modules: pairs of modules mounted back-to-back with a slight tilt of 100 mrad. Matching 2D hits from both modules provides a 3D hit. Within a same layer, modules are arranged with a slight overlap to ensure full coverage. Despite the particle flux being lower in the strip tracker compared to the pixel, occupancy is higher in the latter due to the size of the strips, and ranges from 0.3% in the outer layers to 3% closer to the IP (assuming a mean pileup of 27, as in 2016).

The pixel and strip trackers are cooled to respectively −5 °C and −15 °C to evacuate the 60 kW consumed by the electronics, minimise leakage currents and thermal noise, and slow down radiation damage. The whole inner tracker represents a material budget ranging from 0.4 to 1.8 radiation lengths (X_0), or 0.1 to 0.5 nuclear interaction lengths (λ_i), depending on the pseudorapidity.

2.2.2. Electromagnetic calorimeter

The CMS electromagnetic calorimeter (ECAL), depicted on Fig. 2.5, is wrapped around the tracker volume, its inner surface located at $r = 129$ cm. It is made of lead tungstate ($PbWO_4$) crystals and is designed to initiate and detect electromagnetic showers created by neutral and charged particles (chiefly electrons and photons), with the criteria of fast response time, fine granularity, good energy resolution and radiation hardness. The ECAL is composed of a barrel (EB), covering the pseudorapidity range $|\eta| < 1.479$, and two endcaps (EE) covering $1.479 < |\eta| < 3.0$.

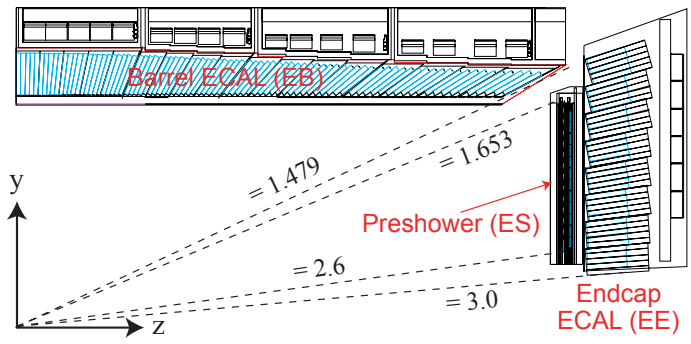

Figure 2.5. Longitudinal layout of one quadrant of the electromagnetic calorimeter, showing the barrel, endcap and preshower. Figure taken from Ref. [181].

Lead tungstate crystals feature a high density (8.23 g cm^{-3}) and short radiation length (89 mm): the resulting system is compact, but with a total thickness of about 25 X_0 it ensures that showers are fully contained within its volume, improving the energy

resolution. The material's small Molière radius (2.2 cm) yields narrow showers, which helps in determining their position and in resolving nearby particles. The crystals are trapezoidal, laid out radially around the IP, with their axes slightly tilted to ensure particle trajectories are never aligned with the inter-crystal cracks. The EB contains 61200 crystals with a front-face cross section of about $22\times22\,\mathrm{mm}^2$ and length of 23 cm, while the EEs contain each 7324 crystals of about $29\times29\times220\,\mathrm{mm}^3$. Scintillation light output amounts to about 4.5 photoelectrons per MeV of the incident particle, with a maximum yield at wavelengths of 420–430 nm. The signal decay time is such that about 80 % of the light is emitted within the LHC bunch crossing time. Since light yield varies with the temperature, the ECAL is maintained at a constant 18 °C. The produced light is amplified and collected by different types of photodetectors in the barrel—avalanche photodiodes (APDs)—and endcaps—vacuum phototriodes (VPTs)—due to the different radiation levels and magnetic field orientations.

To achieve a high level of precision, the ECAL needs to be carefully calibrated. Apart from a global calibration of the energy scale, an uniform response can only be achieved through an *intercalibration* of the individual channels due to small variations in light yield, and photodetector and electronics response. In addition, the crystals show a small but immediate loss of transparency due to irradiation, which depends on the instantaneous luminosity and recovers between fills. To control this effect, crystal transparency is continuously measured by means of laser pulses injected into the detector using optical fibres. The EB energy resolution can be parameterised as [182]:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2, \tag{2.8.}$$

where $S = 2.8\times10^{-2}\,\mathrm{GeV}^{\frac{1}{2}}$ is the stochastic term, due to fluctuations in the lateral shower size or in photo-electrons conversion, $N = 0.12\,\mathrm{GeV}$ is due to noise from electronics and pileup, and $C = 0.3\,\%$ is the constant term, due to variations in longitudinal shower development, intercalibration errors and energy leakage from the back of the crystals. Above $E \approx 100\,\mathrm{GeV}$, the constant terms dominates the resolution.

An extra detector, the Preshower, is placed in front of EE and covers $1.653 < |\eta| < 2.6$. The Preshower is a sampling calorimeter consisting of two lead layers, to initiate electromagnetic showers, behind which are placed silicon strip sensors which measure the deposited energy and transverse shower profile. Its total thickness amounts to 3 X_0.

2.2.3. Hadron calorimeter

The hadron calorimeter (HCAL) is a sampling calorimeter sitting behind ECAL. Its role is to intercept charged and neutral particles such as pions, kaons, protons and neutrons, and measure their energy as well as their position. It is segmented into a barrel (HB), radially constrained to $1.77 < r < 2.95\,\mathrm{m}$ by the solenoid and covering $|\eta| < 1.3$, two endcaps (HE) with a pseudorapidity coverage of $1.3 < |\eta| < 3$, two forward calorimeters (HF) extending to $|\eta| = 5.2$, and an outer calorimeter (HO) located on the outer surface of the solenoid. Figure 2.6 shows the general layout of HCAL.

In both HB and HE, the hadronic showers are created by brass plates acting as absorbers, supplemented by outer front and back plates of stainless steel for structural support.

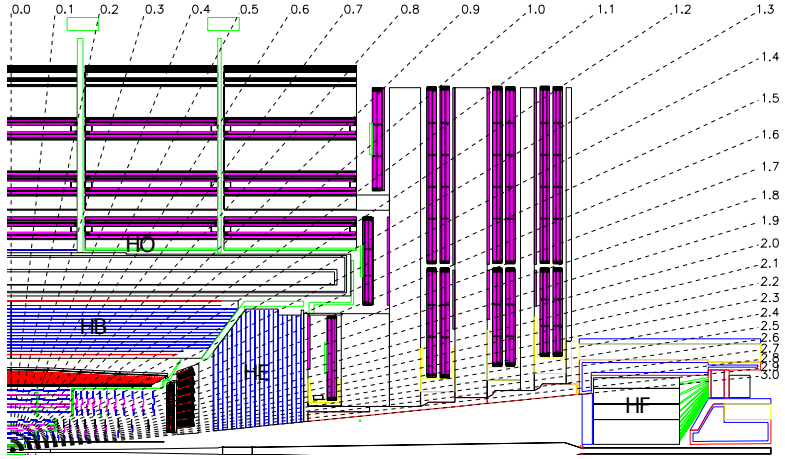

Figure 2.6. Longitudinal cross section showing one quadrant of the hadron calorimeter. Figure taken from Ref. [170].

The brass plates are 56.5 mm thick in HB and 79 mm thick in HE, with gaps respectively 3.7 mm and 9 mm wide, yielding a total absorber thickness ranging from 5.82 λ_I at $\eta = 0$ to $\approx 10\lambda_I$ at $\eta = 1.3$ and beyond. The active medium is provided by about 70000 plastic scintillator tiles inserted into the absorber gaps. The collected light is guided by wavelength shifting fibres (WLSs) to hybrid photodiodes (HPDs). The granularity of the tiles varies from $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ for $|\eta| < 1.6$ to 0.17×0.17 for $|\eta| \geq 1.6$. These $\Delta\eta \times \Delta\phi$ segments form calorimeter *towers*, most of them having a single longitudinal readout. The towers in the transition regions between barrel and endcaps, as well as the endcaps, are divided into two to three readout segments, facilitating the correction of radiation damage effects through a separate calibration of the different layers. The scintillator signal is such that about 68 % of the pulse is contained within a window of 25 ns.

In the central pseudorapidity region, the combined material of EB and HB is not sufficient to contain the hadronic showers. Since this energy leakage unacceptably degrades the resolution, a tail catcher system (HO) is installed outside of the solenoid, increasing the total calorimeter thickness to 11.8 λ_I in the barrel, with the magnet coil working as an extra absorption layer corresponding to $1.4/\sin\theta\,\lambda_I$. The HO consists of five rings of scintillator tiles, with a segmentation roughly mapping the towers in HB. An additional 19.5 cm thick iron plate and a second layer of sensitive material are placed around $\eta = 0$, since absorber depth is minimal in that region.

The forward calorimeters, located at $|z| = 11.2$ m, must endure extremely high levels of radiation during their lifetime, leading to a design consisting of steel absorber plates and quartz fibres laid out along the z-axis. The showers produce Cherenkov light in the fibres, which is guided to photomultiplier tubes (PMTs) located behind a steel-and-concrete shield. The resulting pulses are only 10 ns wide, so that HF is not subject to out-of-time pileup.

Using the parameterisation (2.8), the single-pion energy resolution of the HB is given

by $S = 1.15\,\mathrm{GeV}^{\frac{1}{2}}$, $N = 0.52\,\mathrm{GeV}$, and $C = 5.5\,\%$ [183].

2.2.4. Outer tracker

Since the final states of a wide range of essential physics processes involve muons, it is crucial for CMS to have the capability to identify muons and estimate their p_T quickly enough at the trigger stage, and to measure their momentum with excellent resolution up to very high p_T. Although muons are detected by the inner tracker, that information cannot be used by the trigger. Hence, CMS is equipped with an outer tracker located behind the calorimeters and the solenoid, since the only detectable particles not absorbed by the inner layers are essentially muons. Due to the large volume to be covered, the muon system relies on various gaseous detector technologies. As shown on Fig. 2.7, the components are inserted into the gaps of the flux-return yoke, which further absorbs any punch-through hadrons and provides the magnetic field needed to bend the trajectories of muons and measure their momentum.

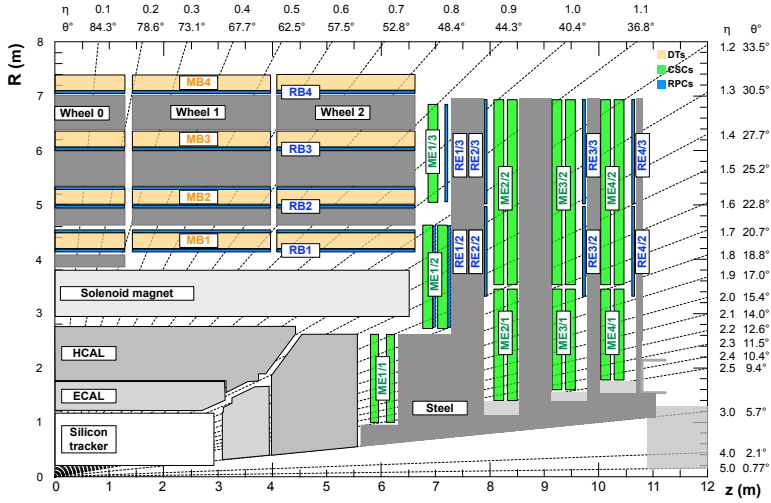

Figure 2.7. | General $r-z$ layout of the CMS muon spectrometer. The DTs, CSCs and RPCs are labelled MB, ME and RB/RE, respectively. Figure taken from Ref. [184].

The drift tubes (DTs), arranged cylindrically around the solenoid, cover the pseudorapidity range $|\eta| < 1.2$. The rectangular drift cells have a cross section of $13 \times 42\,\mathrm{mm}^2$, a maximum length of 2.4 m, and are traversed by a 50 µm-thick gold-plated steel anode wire. The tubes are filled with a mixture of argon and CO_2, yielding a gas gain of 10^5 and a maximum drift time of about 400 ns. The drift time to the wire provides a measurement, transversely to the wire, of the muon position inside of the cell. Sets of four staggered layers of cells (*superlayer*) are combined into chambers which are arranged in four layers (*stations*) around the barrel. Each chamber contains two superlayers with their wires parallel to the beam and providing $r-\phi$ hit positions, and a third one (not present in the outer stations) with its wires orthogonal to the beam, yielding a measurement of the z coordinate. In total, the barrel contains 250 chambers with about 172000 sensitive

wires. The resolution on single-cell hit positions ranges from about 200 µm for $r-\phi$ superlayers to 200–600 µm for z superlayers.

In the endcaps, due to the non-uniform magnetic field and the higher signal and background rates, multiwire chambers are used instead, covering $0.9 < |\eta| < 2.4$. The cathode strip chambers (CSCs) are trapezoidal chambers containing radial copper cathode strips and, perpendicular to those, gold-plated tungsten anode wires. The chambers are filled with a mixture of Ar, CO_2 and CF_4, and their operating voltage is 3.6 kV, providing a gas gain of about 7×10^4. The strips are 8.4 mm to 16 mm wide and provide precise ϕ measurements through interpolation of collected charges, with a single-layer resolution of 300–900 µm. The wires are spaced by about 3 mm and are ganged in groups of 12. They are also read out to obtain a measurement of the crossing time and a rough estimate of the r coordinate. For $|\eta| > 1.2$, muons cross four CSC stations, each containing six cathode panels and wire planes, for a total of 480000 readout channels.

Although DTs and CSCs are intrinsically slow, they achieve a per-station time resolution of about 5 ns, and can be used to trigger on the p_T of muons. To supplement the DTs and CSCs, a parallel set of detectors is installed in the barrel and in the endcaps (over $|\eta| < 1.9$): resistive plate chambers (RPCs). They consist of double-gap chambers operated in avalanche mode, providing a fast response with an excellent time resolution of about 3 ns, which further helps in the triggering and in the assignment of muon candidates to the correct bunch crossing, however their spatial resolution is low (O(cm)). The barrel and the endcaps contain six and four RPC stations, respectively, representing about 130000 channels.

2.2.5. Trigger and data acquisition

The trigger is responsible for selecting the events to be stored and used for further analysis. In CMS, the trigger consists of two stages: the level 1 trigger (L1), which reduces the event rate from 40 MHz to a maximum of 100 kHz, and the high-level trigger (HLT), which further lowers the rate to about 1 kHz. Readout of the detector is handled by the data acquisition (DAQ) system, in which the HLT is integrated. The L1 was upgraded between 2015 and 2016 [185], to enhance its flexibility and to be able to cope with evolving LHC conditions (higher luminosity and pileup). The DAQ and HLT hardware were almost completely replaced during the Long Shutdown 1, between LHC Run 1 and Run 2, to take advantage of advances in computing technology and handle the increase in bandwidth due to higher pileup, and hence larger event size [186].

Level 1 Trigger

The L1 is built out of custom hardware (FPGAs and ASICs) and takes decisions within a fixed delay of 3.8 µs. It receives data from ECAL, HCAL and the muon systems, and has the task of promptly identifying whether electrons or photons (which cannot be distinguished at that stage since no information from the inner tracker is available), hadronic τ lepton decays, jets, large amounts of (missing) transverse energy, or muons are present in the event, and of determining whether these objects pass some pre-defined requirements, such as minimal p_T, quality or isolation. In addition, since for HCAL, DTs and CSCs signals are spread out over several bunch crossings (BXs), it has to combine

the data generated during consecutive BXs, and tag the one in which the detected candidates have been produced. If the event is to be kept, it generates an L1 accept (L1A) signal which is propagated by the trigger control and distribution system (TCDS) to all subdetectors, *triggering* the full read-out of CMS. As shown on Fig. 2.8, the L1 is composed of several subsystems. In the following, only the parts pertaining to trigger candidates used in this work (electrons and muons) will be detailed.

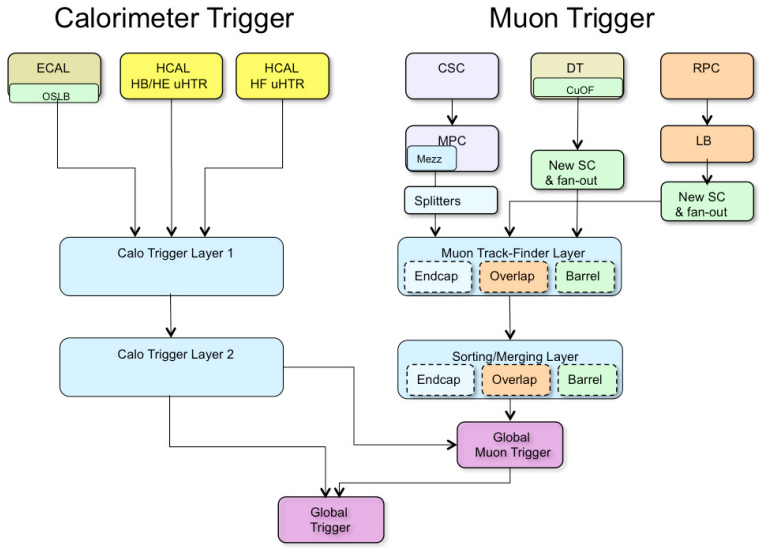

Figure 2.8. | Layout of the CMS level 1 trigger (L1) as of 2016. Figure taken from Ref. [185].

The Calorimeter Trigger processes data from ECAL and HCAL (including HF), segmented into trigger towers (TTs) corresponding to detector regions in $\Delta\eta\times\Delta\phi$ of about 0.087×0.087 (matching the HCAL towers and corresponding to 5×5 ECAL crystals), each encoding the energy deposits in the calorimeters at a specific position. Its two-layer architecture enables the trigger to exploit information from the full detector with TT granularity. Electrons (or photons) are found by searching for local maxima (above a certain threshold) of deposited energy in the ECAL towers. These *seeds* are dynamically clustered with deposits in up to eight neighbouring towers, to recover the full energy of the shower, since it is spread out along ϕ due to bremsstrahlung or photon conversions. The shape of the resulting clusters is used to reject backgrounds. In addition, the clusters are required to be *isolated* by applying a veto on the energy deposit in a region of 6×9 towers in $\eta\times\phi$ around them, and depending i.a. on the estimated level of pileup in the event. The leading 12 e/γ candidates, sorted by their cluster's pileup-corrected E_T, are then passed to the next L1 stage.

Working in parallel, the Muon Trigger receives information from the DTs, CSCs and RPCs. For the first two systems, the local front-end electronics combines hits from the different layers in each station, forming track segments giving a rough estimate of muon position, direction and bending angle. For the RPCs, adjacent hits are clustered together. These *trigger primitives* (segments and clusters) from all three muon detectors

are optimally exploited to search for muon tracks. The Muon Track Finder is partitioned to process in parallel the barrel (BMTF), the endcaps (EMTF) and the overlap region (OMTF). In a further stage, the muon candidate collections are merged and sorted, and possible duplicates are removed. The global muon trigger (GMT) then uses information from the Calorimeter Trigger to compute the pileup-corrected muon isolation, and forwards up to eight candidates to the next layer.

The final stage of the L1 trigger is the micro global trigger (μGT). The μGT combines calorimeter and muon candidates and accommodates about 300 different algorithms (trigger *paths*) which, based on the candidates' position, momentum, reconstruction quality and isolation, decide whether or not the event is to be read out. The algorithms can apply kinematic cuts, require the presence of candidates of different types, or check for topological correlations between candidates. The set of algorithms in use at any moment in time is referred to as a trigger *menu*, and the L1A signal is sent if any one of the paths' decisions is positive. The information reconstructed by the μGT is also forwarded for further analysis. Some trigger paths, in particular technical triggers used for the calibration of trigger paths used for physics analysis, are *prescaled* to limit the overall L1 rate, i.e. only one out of a fixed number of events is kept.

DAQ and High-Level Trigger

The subdetectors are read out via analog optical links by 750 front-end drivers (FEDs) located in an underground service cavern. The FEDs perform the digitisation and first steps of data reduction and local reconstruction, and send the data in fragments of 4–8 kbit to the front-end readout optical links (FEROLs) via copper (400 MB/s) and optical (4/10 Gbit/s) SLINK readout links. The FEROLs forward the data to the surface via 10/40 Gbit/s Ethernet links, where the fragments are received by 84 readout unit PCs (RUs), at a rate of up to 200 GB/s. The 72 builder unit PCs (BUs) have the task of unpacking and combining FED fragments from the different subdetectors to form a complete event, and to that end are connected to the RUs using a 56 Gbit/s Infiniband network with a total bandwidth of 6 Tbit/s (core event builder).

The HLT runs on a farm of filter unit PCs (FUs) connected to the BUs, and relies on the file-based modular software framework used by CMS for offline reconstruction and analysis, CMSSW [187]. The HLT profits from the complete detector data at full granularity to filter the events. High-level physics objects such as electrons, photons, jets, displaced vertices, . . . , are built from raw data using simplified, faster versions of the algorithms used for offline event reconstruction. The HLT paths are independent sequences of several reconstruction and filtering steps, designed to reject events as quickly as possible. On average, events are processed in about 150 ms by the $\approx$ 22000 CPU cores comprising the HLT farm. The set of $\approx$ 500 available paths constitutes the HLT menu and is carefully tuned to remain within the rate budget. Since during an LHC fill both the instantaneous luminosity and pileup decrease, the prescales applied to the technical paths are adjusted accordingly. Due to the evolving performances of the LHC throughout a year, different menus are used depending on the peak luminosity reached. Selected events are output as several *streams* for physics analysis, calibration or monitoring purposes, and the data are written to disk as one file per stream per luminosity section (LS). The LS is the finest time granularity at which the data are

certified as good for further analysis, and corresponds to about 23 s of data-taking. Finally, the files are merged by the storage manager, and sent to the CERN computing centre (Tier0) at a rate of 1 GB/s for offline, full event reconstruction.

2.3. Event reconstruction and selection

This section describes the process of reconstructing *physics objects* usable for analysis, such as electrons, muons and jets, starting from raw detector data. The fine spatial granularity of its detectors has enabled CMS to implement a particle flow (PF) approach to event reconstruction [188]. The PF algorithm is *holistic*, in the sense that it correlates information from all different subdetectors to identify each particle present in an event, and to measure their properties based on this identification. The basic elements used for PF reconstruction (tracks and calorimeter clusters) are described first. Building on those, the identification of physics objects and the measurement of their properties is explained. Next, the triggering algorithms used in this work are listed and described. Finally, we outline the event simulation as well as the methods used to calibrate the reconstruction algorithms and correct for differences between simulation and real data.

2.3.1. Track reconstruction

The first step in reconstructing data from the inner tracker (*local reconstruction*) consists of grouping zero-suppressed signals from pixels or strips into *clusters* (i.e., hits). The procedure profits from the property that a charged particle traversing the tracker deposits a signal in a few neighbouring pixels or strips. Since a measurement of the amount of deposited charge in each sensor is available, this charge-sharing enables a determination of the hit position to a precision finer than the width of the sensors. For the pixel, this is done by comparing the expected cluster charge distributions (*templates*), obtained from a detailed simulation of the sensors, with the recorded signals. This accounts for the radiation damage to the pixel detector, as well as for the Lorentz drift of the collected charge due to the magnetic field. Strip signals are clustered based on their signal-to-noise ratios; the charge-weighted average of strip positions, corrected for Lorentz drift in the barrel, defines a cluster position. The pixel is affected by a dynamic inefficiency causing the efficiency to reconstruct a hit to decrease with the instantaneous luminosity. The efficiency remains above 99 %, except for the innermost layer where it decreases to 95 % for the highest luminosities reached in 2016 [189]. In that year, the strip tracker was affected by a dynamic inefficiency due to a saturation of the front-end readout chips, causing the hit efficiency to decrease from the nominal value of 99.8 % to about 92 % at high luminosity [190]. However, the efficiency could be fully recovered in the summer of 2016 after changing the readout chip settings. Hit resolution varies from 10 µm to 50 µm depending on cluster position and the considered direction.

At first order, charged particles follow an helical trajectory through the tracker volume, with the helix axis parallel to the beam axis. The trajectory can hence be defined by five parameters: the direction in η, the position in (x, y, z) of the point of closest approach (PCA) w.r.t. a reference point (e.g. centre of the detector), and the track curvature radius R. The latter gives access to the track transverse momentum through the relation:

$$p_{\mathrm{T}}(\mathrm{GeV}) = 0.3 \cdot B(\mathrm{T}) \cdot R(\mathrm{m}) \tag{2.9.}$$

Reconstructing tracks from hits and measuring these five parameters is a challenging task due to the high multiplicity of charged particles produced in each event. In CMS, it is rendered even more difficult by the amount of material present in the tracker: an electron has typically a 85 % probability to emit a bremsstrahlung photon when interacting with the material, and a hadron has a 20 % chance to undergo a nuclear interaction before reaching the calorimeters. To maintain a good track-finding efficiency while keeping the rate of fake tracks low, track reconstruction proceeds in several iterations of the tracking sequence called the combinatorial track finder (CTF) [180]. By starting to reconstruct the easiest tracks (i.e. high-p_{T} tracks produced near the primary interaction region) and removing their associated hits, the combinatorial complexity is reduced for each subsequent iteration, simplifying the search for low-p_{T} or displaced tracks (*iterative tracking*). Every iteration consists of the following steps:

- Generation of *seeds*, i.e. track candidates consisting only of two or three hits, yielding a rough initial trajectory estimate. Hits from the pixel detector are used for the first iterations, since it provides 3D spatial measurements and its occupancy is lower than the strip tracker's. Matched stereo strip hits are still used in later iterations to recover displaced tracks such as tracks produced by the decay of long-lived hadrons. Depending on the iteration, different constraints are applied on the seeds, such as having a minimum p_{T} or originating from a region close to the beam spot (transversely and/or longitudinally).
- Track finding, based on a Kalman filter (KF) method, consists of extrapolating the seed trajectory (navigation) and finding hits in the next layers compatible with the current estimate of the track parameters. The extrapolation takes into account the added uncertainty in the trajectory due to energy losses and multiple scattering in the layer material. Once an extra hit is found, the trajectory is updated using its position. The steps of navigation, hit finding and trajectory update are repeated until the outer tracker layer is reached. Depending on the CTF iteration, different requirements on track p_{T} and number of found and missing hits are applied.
- Once the track is built, its parameters are refit using a Kalman filter and smoother to profit from the full information now available about its trajectory. This step uses a precise description of the tracker material and takes into account inhomogeneities in the magnetic field, which imply deviations from a simple helical trajectory even in between the layers.
- Due to the complexity of the pattern finding procedure, a large number of fake tracks can be present. To reject those, a selection is applied based on a multivariate analysis of track quality criteria, such as the number of missing hits, the fit χ^2 or the compatibility with the beam spot, as a function of track p_{T} and η. The requirements are tuned separately for each iteration.

Finally, the tracks found by the different iterations are merged and possible duplicates are removed. In 2016, 10 iterations were used for the track reconstruction. For hadrons produced in collisions with a mean pileup of 25, the track-finding efficiency ranges from 80 % to 95 % depending on track p_{T} and η, while the fake rate varies from 5 % to 10 % [191]. For prompt isolated muons however, the efficiency remains above 99 % [192].

Muon tracking

For muons [184], the tracking can profit from information coming from the outer tracker, where backgrounds and occupancy are much lower. In the DTs, the arrival time of electrons collected by the anode wires gives access to the distance of closest approach of the crossing muons to the wires. Using recorded hits, straight-line track *segments* are fitted separately in each chamber. Local reconstruction in the CSCs consists in building hits as the intersection points between activated wires and strips. As in the DTs, straight segments are built from the hits in each layer. Adjacent RPC strip hits are clustered, and their charge-weighted average defines the cluster's position. A KF is used to build tracks using the hits provided by all three muon subdetectors. These *standalone* muons are then combined with geometrically compatible inner tracks to define *global* muons. Two of the CTF iterations are tuned specifically for muon reconstruction, to improve the efficiency in high-pileup conditions. One uses identified global muons to trigger an outside-in track reconstruction, whereas the other re-builds tracks that match hits in the muon system (*tracker muons*), using looser quality constraints.

Electron tracking

Electrons also leave a signal in the inner tracker, yet the algorithm outlined above is not ideal for their reconstruction. Indeed, they radiate a significant fraction of their energy through bremsstrahlung, both due to their curved trajectory and when crossing tracker material. These losses are highly non-Gaussian in nature, and since the standard KF relies on uncertainties being Gaussian, it is bound to fail. Tracking of electrons thus relies on a modified KF: the Gaussian sum filter (GSF) [193, 194].

The seeds provided as input to the GSF are built by using information from the ECAL, which is done in two ways. The first method starts from clusters of energy deposits in the ECAL. To recover energy carried by bremsstrahlung photons, clusters are merged to form super clusters (SCs). Due to the curvature of electrons in the axial magnetic field, these photons generally hit the ECAL at polar angles similar to that of the initial cluster considered, but are spread out along the azimuthal direction. Tracker seeds are formed using the constraint that the SCs provide on possible electron trajectories. The second method relies on the standard track collection, selecting tracks compatible with being electrons. These tracks can be extrapolated to ECAL clusters, or have a poor fit quality (large χ^2, missing hits) due to bremsstrahlung.

In the GSF, the Bethe-Heitler formula governing the distribution of fractional energy losses is approximated as a sum of Gaussians. The trajectory is modelled as a weighted mixture of several components, and each of them is propagated independently between pairs of layers using an energy loss and uncertainty based on the Gaussian component to which it corresponds. The components are weighted by their relative importance in the Gaussian sum. At each layer, the number of components increases by the number of Gaussians considered, and to avoid this exponential growth only the most probable estimates are retained. Additional hits are added to the trajectory components as in the KF, but with looser constraints on the compatibility of the hit with the current track estimate. Ultimately, the track parameters are defined by the modes of the distributions obtained from the weighted sums of the posteriors for the parameters in each remaining GSF component.

Primary vertex finding

Reconstructed tracks can be used to identify *primary vertices*, i.e. the locations of all proton-proton interactions in the event. These include the vertex corresponding to the "hard" interaction, used for physics analysis and hereafter referred to as *the* primary vertex (PV), as well as additional parasitic interactions (pileup). The primary vertices are used to measure the position and size of the beam spot, i.e. the 3-D distribution of the luminous region, which feeds back into the track reconstruction sequence (both online and offline) as it is used as a constraint on track origin in some CTF iterations. Identifying pileup vertices is also of fundamental importance to mitigate the effect of pileup on object reconstruction performance, as will be described in Sec. 2.3.8. Lastly, knowledge of the position of the primary vertex is a crucial ingredient for b tagging (see Sec. 2.3.7).

Vertex reconstruction consists of selecting a subset of tracks using quality criteria, and clustering them based on the z-coordinate of their PCA to the centre of the beam spot. To efficiently resolve close-by interactions while avoiding to split clusters corresponding to genuine vertices, the clustering is based on a deterministic annealing algorithm, seeking the global minimum of an analogue of free energy through step-wise reductions of the "temperature" T. At infinite T, all tracks are assigned to a single vertex. As T is reduced, the vertices are allowed to split if the resulting configuration is more favoured. Once the minimum is reached, vertices containing at least two tracks are fitted to estimate their position. The resolution on vertex position varies from 10 µm to 100 µm, depending on the number and p_{T} of the clustered tracks [195]. Finally, the beam spot can be measured by fitting the distribution of reconstructed vertices, a procedure repeated for every LS in the data.

To pinpoint the vertex corresponding to the hard interaction in the event, objects are built, for each vertex candidate, using a jet clustering algorithm taking as input all the tracks associated to the vertex, as well as the vertex' missing transverse momentum. The vertex with the highest $\sum p_{\mathrm{T}}^2$, where the sum runs over the thus defined objects, is taken to be the hard interaction vertex.

2.3.2. Calorimeter clusters and particle-flow links

In the PF algorithm, the calorimeters primarily serve to identify and measure neutral hadrons and photons, help in the reconstruction of electrons, and improve the energy measurement for charged hadrons (in particular at high p_{T}). The clustering of energy deposits used in PF was specifically designed to resolve individual particles. Local cell energy maxima, above a certain threshold, define cluster *seeds*. Contiguous deposits are merged with the seeds, forming *topological clusters*. The procedure is carried out independently for each subdetector (ECAL, HCAL and HF) and partition (barrel and endcaps). Within a topological cluster, to take into account the overlap of energy deposits due to individual particles, *clusters* are identified using a Gaussian-mixture model, where the number of Gaussian energy deposits corresponds to the number of seeds, and the position and amplitude of each Gaussian are the cluster parameters. The single-particle PF response needs to be carefully calibrated, to correct i.a. for threshold effects and for the nonlinearity of the detector response. This calibration is carried out using simulated data, separately for photons using only the ECAL, and for neutral hadrons, which deposit

energy sometimes in both the ECAL and the HCAL, sometimes mostly in the HCAL.

To identify individual particles in the events, the PF approach relies on information from the different subdetectors. To that end, the *link algorithm* combines PF *elements* (tracks, clusters) to create PF *blocks*, which form the basis for the different object reconstruction algorithms. For instance, an inner track is linked with a calorimeter cluster if it can be extrapolated to a position compatible with that of the cluster. Further, if tangents to a GSF track can be extrapolated to ECAL clusters, these clusters can be linked with the track to recover bremsstrahlung photons. Pairs of tracks compatible with originating from a photon conversion into an e^+e^- pair also become linked. Similarly, groups of track are formed for reconstructed nuclear interactions within the tracker volume. Links between calorimeter clusters are only formed out of the tracker acceptance, when ECAL clusters are contained within the envelope of HCAL clusters. Lastly, as already hinted at in Sec. 2.3.1, links between inner tracks and muon tracks or segments establish global or tracker muons, respectively. With all the possible blocks in hand, object identification proceeds in steps. PF blocks corresponding to the candidates reconstructed at a given step are masked, and not considered further.

2.3.3. Muons

Muon identification (ID) [184] aims at rejecting backgrounds such as cosmic muons crossing the detector, or *punch-through* hadrons, i.e. high-p_T hadrons that are not completely absorbed by the calorimeters and magnet and create spurious signals in the muon chambers. All prompt muons considered in the analysis presented in this work are required to satisfy the *tight* ID specifications. Tight muons are global muons with additional requirements on the number of hits in the pixel, tracker and muon chambers, on the track transverse and longitudinal distance to the primary vertex, and on the global track fit χ^2. To suppress muons inside jets, originating from the leptonic decays of heavy or long-lived hadrons, the prompt muons are required to be isolated, namely that the scalar sum of all charged-particle p_T and neutral-particle E_T within a cone of radius $\Delta R = 0.4$ around the muon does not exceed 15% of the muon p_T, corresponding to the tight isolation working point (WP). The efficiency ϵ with which a prompt muon is selected for analysis can be broken up as:

$$\epsilon = \epsilon_{\text{trk}} \cdot \epsilon_{\text{ID}\,|\,\text{trk}} \cdot \epsilon_{\text{iso}\,|\,\text{ID}} \tag{2.10.}$$

The first term corresponds to the efficiency to reconstruct the muon track, as already mentioned in Sec. 2.3.1; the two other terms are the relative efficiencies for prompt muons to pass the ID and isolation requirements. For the former the efficiency is about 94 to 97%, depending on muon pseudorapidity, while the latter varies from 85% for muon $p_T \approx 20$ GeV to > 99% for $p_T > 60$ GeV [196]. The dependence of $\epsilon_{\text{ID}\,|\,\text{trk}}$ and $\epsilon_{\text{iso}\,|\,\text{ID}}$ on muon kinematics are shown on Fig. 2.9.

Although muons within jets are not considered as prompt muons, it is important to identify them since they contribute to the jet momentum, and failing to do so would negatively impact the jet resolution. In addition, they can provide information about the presence of heavy-flavoured hadrons inside the jets. These non-isolated muons are required to pass the tight selection, as well as additional criteria designed to further

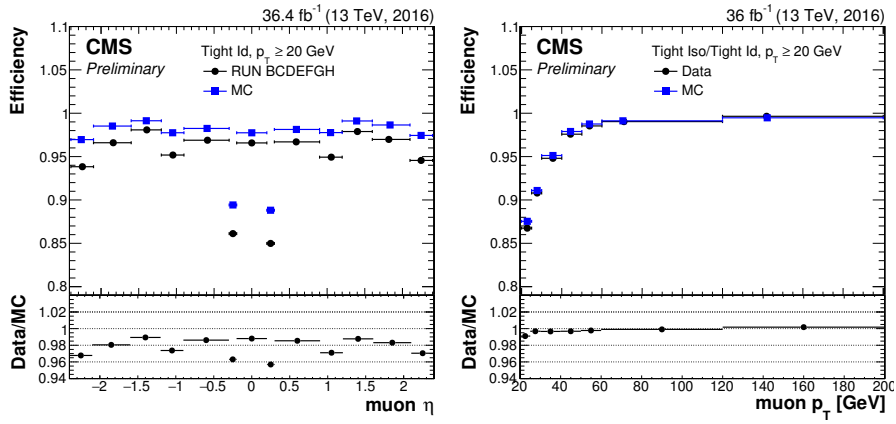

Figure 2.9. | Left: tight muon identification efficiency, as a function of muon pseudorapidity. Right: efficiency of the tight isolation requirement, for muons that pass the tight ID criteria, as a function of muon p_T. Both efficiencies are measured in data using the tag-and-probe (T&P) method, described in Sec. 2.3.11. Figures taken from Ref. [196].

remove punch-through hadrons.

The muon momentum is obtained from a combination of the inner track and different global muon fits, depending on their quality and associated uncertainties. For a muon p_T below 200 GeV the inner track dominates the resolution, while for higher p_T values the longer lever arm of the outer system significantly improves the measurement, as shown on Fig. 2.10 (left). Ultimately, the momentum resolution is not limited by the spatial resolution on hits in outer stations but by multiple scattering as muons cross the detector. The p_T resolution for muons of $20 < p_T < 100$ GeV varies from 1% in the barrel to 5% in the endcaps.

2.3.4. Electrons

Identification of prompt electrons, described in Ref. [194], proceeds on the basis of GSF tracks, provided that the PF-associated ECAL cluster is linked to at most two additional tracks. The electron direction is taken to be that of the GSF track, whereas its momentum is obtained by combining track p_T and cluster energy. The latter dominates the momentum resolution for electron $p_T \gtrsim 20$ GeV, as shown on Fig. 2.10 (right), but has to be corrected for energy missed by the clustering. The resolution in this range of p_T varies from less than 2% for non-showering, central electrons, to 4% for showering electrons in the endcaps.

A cut-based selection is employed to reject backgrounds of wrongly identified prompt electrons, such as converted photons ($\gamma \to e^+e^-$), hadrons, or electrons from weak decays of hadrons within jets. The selection is based on purely tracking- or calorimetry-related information, as well as comparisons between observables from both detectors. These variables are:

- the extension of the shower in the η direction;
- the angular distances between the energy-weighted centre of the ECAL cluster and

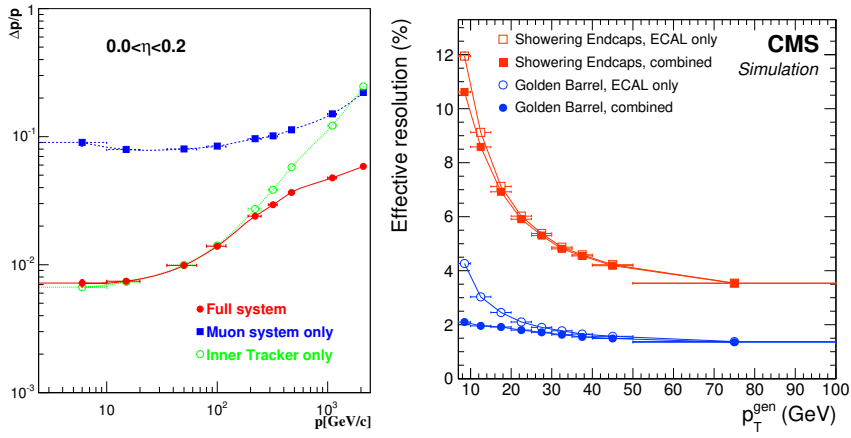

Figure 2.10. Left: resolution on reconstructed muon momentum, as a function of true muon momentum, for central muons. The resolution obtained with either the inner or outer tracker is compared with that achieved when combining the hits in both detectors. Figure taken from Ref. [181]. Right: resolution on electron p_T as a function of true electron p_T, for "golden" electrons in the barrel, i.e. electrons which do not radiate and are reconstructed as a single ECAL cluster, and for "showering" electrons in the endcaps, which emit significant amounts of radiation and yield several clusters. The open symbols denote the resolution obtained using only information from the ECAL, whereas for the solid symbols the track p_T and cluster energy are combined. Figure taken from Ref. [194].

the position of the track (extrapolated to the ECAL);

- the compatibility between the track momentum and cluster energy;
- the ratio between the energy of the HCAL and ECAL clusters linked to the track;
- the number of missing hits in the tracker;
- the compatibility of the track with originating from the primary vertex.

In addition, electrons from photon conversions are rejected using a dedicated algorithm, and electrons are required to be isolated. The isolation is computed similarly to the muons, using reconstructed PF candidates in a cone of size $\Delta R = 0.3$ around the electron direction. All the previous requirements have been optimised separately for electrons detected within EB or EE. The values used in the present analysis correspond to the *medium* ID working point, for which the efficiency varies from 60% at $p_T \approx 20$ GeV to 90% for $p_T > 45$ GeV [197].

2.3.5. Hadrons and jets

After the identification of muons, electrons and isolated photons (not described here) is completed, the left-over PF blocks can be used to reconstruct charged and neutral hadrons such as $\pi^{\pm}$, $K^{\pm}$, protons, K_L^0 or neutrons, and non-isolated photons such as those stemming from π^0 decays. The resulting charged and neutral PF candidates are fed to a clustering algorithm to reconstruct *jets*, as described in Sec. 1.1.4. The PF algorithm is here crucial, since on average two thirds of the jet energy is carried by charged hadrons, for which the momentum can be precisely measured by the tracker. Three quarters of the

remaining energy consists of photons, whose energy is measured with good precision by the ECAL. Hence, the HCAL is mainly responsible for measuring the energy of neutral hadrons, which represents less than 10% of the total jet energy. Given the much better momentum and angular resolution of the tracker and ECAL compared to the HCAL, the jet energy and angular resolution in the PF approach is significantly improved w.r.t. a purely calorimetry-based method. The difficulty resides in the need to properly identify each particle in the event, despite the fact that 20% of hadrons interact with the tracker material, to avoid double-counting the momentum of tracks and the associated energy deposits, as this would negatively impact the jet energy scale.

Hadrons are identified as follows. ECAL and HCAL clusters within the tracker acceptance which do not have associated tracks are taken to be photons and neutral hadrons, respectively. Outside of the tracker acceptance, neutral and charged particles cannot be distinguished: ECAL-only blocks are considered as photons, whereas blocks consisting of both an ECAL and HCAL cluster give rise to a single hadron candidate. Tracks linked with calorimeter clusters give rise to charged PF candidates. If the calibrated calorimetric energy is compatible with the sum of linked track p_{T}, the candidate momenta are re-fit using the calorimeter information. At very high energy, or in the case of poor track fits, this improves the resolution over a purely tracking-based measurement. If, on the other hand, the cluster energies are significantly above the track p_{T}, the charged hadron momenta are taken to be those of the tracks, and the excess energy is interpreted as photons (in ECAL) and neutral hadrons (in HCAL).

The jets used in this analysis are obtained via the anti-k_{T} algorithm using a radius parameter $R = 0.4$. Spurious jets, primarily due to detector noise in the calorimeters, are removed by a loose selection (jet ID) requiring that the jets contain more than one candidate and at least one charged candidate, and that they are composed of a mixture of charged and neutral electromagnetic and hadronic energy, as is observed for genuine jets produced through the showering and hadronisation of energetic quarks and gluons. This selection has an efficiency better than 99% and suppresses nearly 100% of the backgrounds [198].

2.3.6. Missing transverse momentum

Neutrinos can be produced through the leptonic decay of W bosons, such as those coming from top quarks or Higgs bosons, or through the decay of hadrons. They escape the detector without leaving a signal, yet their presence can be inferred from an imbalance in the total measured transverse momentum in the event. Indeed, the protons in a collision travel in opposite directions along the z axis, and hence have zero total momentum. Conservation of momentum dictates that the total momentum after the collision should still be naught. However, since proton debris after the collision travel inside the beam pipe and remain undetected, only the *transverse* total momentum of the produced neutrinos can be retrieved.

The PF missing transverse momentum, $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, is defined as the negative vector sum of the transverse momenta of all reconstructed PF candidates in the event,

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = -\sum_{i=1}^{N_{\mathrm{PF\,cands.}}} \vec{p}_{\mathrm{T},i}\,, \tag{2.11.}$$

and its magnitude is denoted as $p_{\mathrm{T}}^{\mathrm{miss}}$.

Measuring a non-zero $p_{\mathrm{T}}^{\mathrm{miss}}$ does not imply that prompt neutrinos have been produced in the hard scattering, since neutrinos in jets from weak decays of hadrons will contribute to the $p_{\mathrm{T}}^{\mathrm{miss}}$. In addition, $p_{\mathrm{T}}^{\mathrm{miss}}$ is sensitive to any mismeasurement of the momenta in (2.11) and to the fact that some particles escape detection due to detector and reconstruction inefficiencies or limited acceptance. In this regard, the extended forward coverage of HF is crucial to maintain a good $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution. Some events might feature an abnormally large $p_{\mathrm{T}}^{\mathrm{miss}}$ due to instrumental effects, such as noise in the calorimeters, or due to backgrounds such as beam halo, i.e. muons produced from parasitic collisions between protons and residual gas in the beam pipe away from the IP. Severe mismeasurements of muon momenta, or misidentified muons, can also induce a large fake $p_{\mathrm{T}}^{\mathrm{miss}}$. A dedicated post-processing or filtering of affected events successfully corrects for these effects [199, 200]. In 2016 data, some muons were observed to be duplicates of genuine muons. These fake muons would not pass the tight muon ID used in this analysis, but would still enter the computation of the $p_{\mathrm{T}}^{\mathrm{miss}}$ and significantly alter its value. Fortunately, it was possible to flag these duplicate muons and correct the measured $p_{\mathrm{T}}^{\mathrm{miss}}$.

2.3.7. Secondary vertices and b tagging

Identifying, or *tagging* jets originating from the hadronisation of b quarks (b jets) [201] is a powerful method to reduce backgrounds in analyses involving Higgs bosons or top quarks, since their decays involve b quarks in almost 60% and 100% of cases, respectively. Such an identification is possible thanks to the fact that those jets contain B mesons (B^0, $\mathrm{B}^{\pm}$), which have proper lifetimes of $c\tau \approx 500\,\mu\mathrm{m}$. Hence, given an energy of several tens of GeV, these hadrons travel typically a few mm ($c\tau\gamma$, $\gamma = E/m$ with $m \approx 5$ GeV) in the detector before decaying, opening the possibility to reconstruct displaced, *secondary* vertices (SV), as displayed on Fig. 2.11. Even when no such vertex can be reconstructed, tracks within b jets will have different properties from tracks of jets produced by gluons or light quarks (light jets), as described below.

Tracks used for b tagging are tracks clustered within jets that satisfy quality criteria, listed below, designed to reject fake tracks, tracks coming from pileup vertices, or tracks that might originate from the decay of long-lived hadrons (such as $\mathrm{K}_{\mathrm{S}}^0$, which has $c\tau \approx 2.7\,\mathrm{cm}$) or from nuclear interactions in the beam pipe or in the pixel material:

- $p_{\mathrm{T}} > 1\,\mathrm{GeV}$, fit $\chi^2/\mathrm{NDoF} < 5$, at least one hit in the pixel detector.
- The track impact parameter (IP) is defined as the distance between the primary vertex and the track at its PCA w.r.t. the vertex, as illustrated on Fig. 2.11. Requirements are applied separately on the longitudinal (< 17 cm) and transverse (< 2 mm) projections of the IP. The sign of the IP is taken to be negative if the angle between the vector joining the PV with the PCA, and the jet direction, is larger than 90°. For light jets one expects the IP distribution to be symmetric around zero, with a width reflecting the experimental resolution on this quantity. Tracks within b jets are expected to be displaced, i.e. they have mostly positive and large IP values. A related quantity, the impact parameter significance (IPS), is defined as the IP divided by its uncertainty.
- The minimal distance between the jet axis and the track is to be smaller than 700 µm.
- The track decay length, i.e. the distance between the PV and the PCA between the

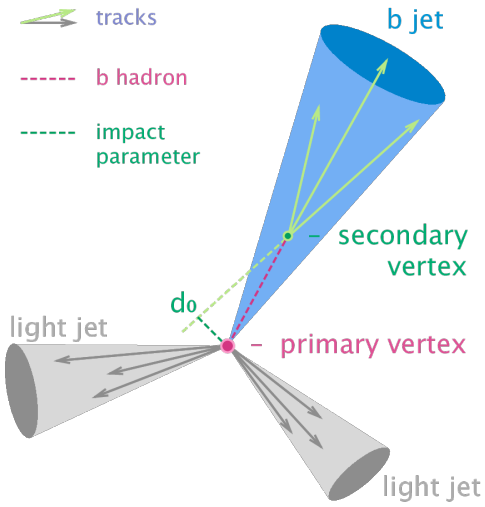

Figure 2.11. Schematic representation of the main ingredients entering b tagging. Figure taken from Ref. [202].

jet axis and the track, should be smaller than 5 cm.

Secondary vertices are reconstructed using two different algorithms. The "legacy" adaptive vertex reconstruction (AVR) algorithm is based on the adaptive vertex fitter used for primary vertex fitting and considers only tracks associated with jets. Requirements, e.g. on the number of tracks shared with the primary vertex, the vertex flight distance (distance between the PV and the SV), the angle between the vertex direction and jet axis, or the vertex mass, are applied on the vertices to reject those not compatible with the decay of a B hadron. The second algorithm is the inclusive vertex finder (IVF) and processes all the tracks present in the event, after a selection looser than the one described above. Seed tracks are clustered based on the minimum distance and angle between them, and the resulting clusters are fitted using the same adaptive fitter as for AVR. An arbitration procedure is applied in case tracks are shared among several vertices or with the PV, either removing tracks from vertices and re-fitting them, or removing vertices entirely. The IVF vertices are then cleaned using a selection similar to that used for AVR.

Several algorithms (taggers) for b jet identification have been developed, each taking advantage of different properties of such jets. In the jet probability (JP) and jet b probability (JBP) taggers, the probability for a jet to be compatible with the PV is computed using the IPS of the tracks associated with the jet. The tracks with negative IPS naturally define a resolution function p(IPS), from which the probability for a track to originate from the PV is given by $\int_{\text{IPS}}^{+\infty} p(x)\mathrm{d}x$. These probabilities for several tracks in the jets are then combined in different ways to define the J(B)P taggers. The soft electron tagger (SET) and soft muon tagger (SMT) are dedicated taggers that exploit the presence and properties of soft leptons within jets, aiming at tagging those coming from the leptonic decays of B hadrons. However, these can only be useful for the small fraction of jets actually containing leptons. The combined secondary vertex (CSVv2)

tagger is a multivariate discriminant built using information about displaced tracks and secondary vertices associated to jets. The discriminant, a multilayer perceptron (MLP) (see Sec. 2.4.1), is trained separately for jets containing either at least one SV, no SV but at least two tracks forming a so-called "pseudo" SV, or neither of those. Variables used include track IPS, decay length, angle with the jet axis; SV mass, angle with jet axis, flight distance significance, number of SVs, etc. The b tagging algorithm used in this analysis is the combined multivariate algorithm (cMVAv2). This *supertagging* algorithm relies on a boosted decision tree (BDT) discriminant (see also Sec. 2.4.1) and combines the scores (outputs) of six different taggers: JP, JBP, CSVv2 using either AVR or IVF vertices, SET and SMT. The performance of taggers entering cMVAv2, as well as that of the cMVAv2 tagger, are shown on Fig. 2.12.

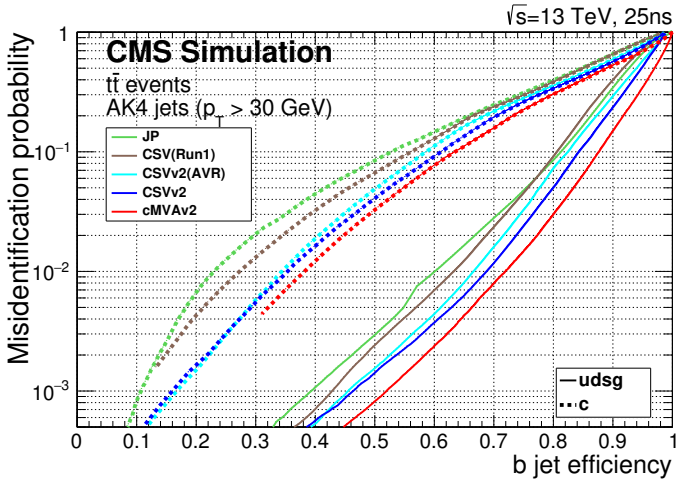

Figure 2.12. | Comparison of the average performance of several b taggers used in CMS. The x axis is the efficiency to correctly select a true b jet for varying thresholds on the algorithm score, while the y axis gives the probability to wrongly select a light or c jet with the corresponding threshold. The working point used in this work corresponds to the point on the cMVAv2 curve for which the light mistag rate is of 1%. Figure taken from Ref. [201].

The most straightforward way to employ a b tagger is to consider as b jets all jets for which the tagging score is above a chosen value (working point). The "Medium" operating point used in this analysis (cMVAv2M) yields an *average* selection efficiency for true b jets of about 70%, with a corresponding mistag rate, i.e. probability to wrongly identify light (c) jets, of less than 1% (20%). For a given working point (WP), both the b jet efficiency and the mistag rate depend on the jet p_{T} and $|\eta|$. The light-jet mistag rate is optimal for central jets and gradually increases with both p_{T} and $|\eta|$. The mistag rate for c jets is approximately constant, whereas the b-jet tagging efficiency is highest for $100 < p_{\mathrm{T}} < 200$ GeV and is somewhat degraded outside of this range.

2.3.8. Pileup mitigation

As first mentioned in Sec. 2.1.1, in 2016 on average $< \mu >= 27$ inelastic pp interactions happened during each bunch crossing. Moreover, the distribution of the number of

interactions per bunch crossing, shown on Fig. 2.13, features tails with a significant fraction of events containing more than 40 interactions. These soft interactions (pileup, or PU) create additional low-p_{T} charged and neutral particles that overlap with the products of the hard collisions, in which we are interested. In addition, the time between consecutive bunch crossings can be shorter than the time some particles need to cross the detector, or shorter than the duration of electronic signals in the detectors, creating out-of-time (OOT) pileup. Pileup particles give rise to additional deposits in the calorimeters, approximately uniformly distributed throughout the detector, that degrade the energy resolution of jets and of the $p_{\mathrm{T}}^{\mathrm{miss}}$. Extra hits from charged pileup particles in the trackers render track reconstruction more difficult due to the increased combinatorial background, resulting in a lower tracking efficiency and higher fake rate. As a consequence, but also due to the additional reconstructed tracks in the event, tasks such as b tagging become more challenging as well. Finally, the extra energy due to pileup impacts the estimate of lepton isolation, resulting in lower muon and electron selection efficiencies for a given isolation requirement. This sections describes the methods used to mitigate these effects.

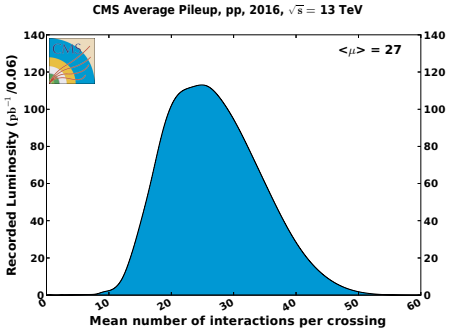

Figure 2.13. | Distribution of the mean number of interactions per bunch crossing during 2016 data taking, assuming an inelastic pp cross section of 80 mb. Figure taken from Ref. [178].

Pileup interactions which have produced charged hadrons reconstructed as tracks can be identified as additional primary vertices (see Sec. 2.3.1). These hadrons can therefore be flagged as pileup and removed from the list of candidates in the event. This procedure is dubbed charged-hadron subtraction (CHS), and it should be understood that the reconstruction algorithms described in previous sections are based on CHS-cleaned candidate collections (PF+CHS) [188].

For neutral hadrons, photons, and charged hadrons beyond tracker acceptance or not removed by CHS, no such straightforward procedure is possible. In one strategy pursued, the average p_{T} density due to pileup is estimated, relying on the uniformity of pileup energy deposits. This density, ρ, can then be multiplied by the "area" A of a candidate, and the result $\rho \cdot A$ subtracted from the candidate's p_{T}. For jets, A is taken to be their catchment area [203,204]. The *effective* area A_{eff} used to correct the electron isolation [194] is directly related to the area of the isolation cone, $(\Delta R)^2$, and their (absolute) isolation is then given by

$$\mathrm{Iso}_{\mathrm{PF}}^{\mathrm{e,\,abs.}} = \sum_{i \in \mathrm{h}^{\pm}} p_{\mathrm{T}}^{i} + \max\left[0, \sum_{i \in \mathrm{h}^{0}} p_{\mathrm{T}}^{i} + \sum_{i \in \gamma} p_{\mathrm{T}}^{i} - \rho A_{\mathrm{eff}}\right], \tag{2.12.}$$

where the sums run over all PF charged CHS ($\mathrm{h}^{\pm}$) and neutral (h^{0}, γ) candidates within a cone of $\Delta R = 0.3$ around the electron. An alternative method, used to correct muon isolation, is based on the observed ratio of $\Delta\beta \approx 0.5$ between neutral and charged pileup energy. The contribution to the isolation due to neutral pileup can therefore be estimated locally from the measured charged pileup energy in the isolation cone:

$$\mathrm{Iso}_{\mathrm{PF}}^{\mu,\,\mathrm{abs.}} = \sum_{i \in \mathrm{h}^{\pm}} p_{\mathrm{T}}^{i} + \max\left[0, \sum_{i \in \mathrm{h}^{0}} p_{\mathrm{T}}^{i} + \sum_{i \in \gamma} p_{\mathrm{T}}^{i} - \Delta\beta \sum_{i \in \mathrm{PU\,h}^{\pm}} p_{\mathrm{T}}^{i}\right] \tag{2.13.}$$

Here the sums run over the same collections as in (2.12), in a cone of $\Delta R = 0.4$ around the muon. In the last term of (2.13), only charged hadrons associated with pileup vertices are considered.

The calibration of the $p_{\mathrm{T}}^{\mathrm{miss}}$ is not expected to be affected by pileup, since pileup interactions produce little to no true $p_{\mathrm{T}}^{\mathrm{miss}}$. However, its resolution can be significantly degraded due to the extra energy in the events. In addition to computing the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ using CHS-cleaned candidates, a correction for the effect of neutral energy deposits, referred to as Type-0 correction, is applied. This correction is computed from the observed charged pileup candidates, assuming that for pileup the true charged and neutral total transverse momenta are exactly balanced, and that the charged candidates are perfectly measured.

2.3.9. Trigger paths and datasets

As will become clear in the next chapter, the signals considered in this analysis involve pairs of prompt, isolated leptons (electrons or muons). This provides us with a clear signature that the trigger system can identify with a high efficiency. The HLT paths we have used only triggered the recording of an event if two leptons, passing certain quality criteria, could be reconstructed. Depending on the flavour of the leptons, several such paths, listed in Tab. 2.1, were available during 2016 data taking. Data events entering the analysis have fired at least one of those paths, and are stored in the following, possibly overlapping primary datasets (PDs): DoubleMuon, DoubleEG, and MuonEG. Each dataset is divided into *eras*, labelled B through H. An era corresponds to a data-taking period with relatively homogeneous conditions (rate, trigger menus, . . .). Data in era H were reconstructed during data-taking itself, while eras B–G were *re-reconstructed* at the end of 2016 to profit from updated calibration and alignment measurements.

At the HLT, a simplified, quicker version of the PF event reconstruction sequence described above is carried out. Tracking is made substantially faster by performing *regional* track finding and fitting, e.g. by considering hits only in regions of interest defined by muon or electron L1 candidates. In addition, only a few iterations of the CTF sequence are ran, focusing on high-p_{T} tracks produced near the IP. The primary vertex and the beam spot are built from tracks reconstructed using only hits from the pixel detector.

Table 2.1. | Dilepton trigger paths available during 2016 data taking. The channels and p_T cuts refer to the p_T-leading and subleading leptons. Each path consists of two sets of identification ("Id") and/or isolation ("Iso") criteria applied to the reconstructed leptons. For some paths (labelled "DZ"), an additional requirement on the compatibility of the two leptons as originating from the same PV is applied.

Channel	HLT paths	p_T thresholds (GeV)
$\mu\mu$	Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ Mu17_TrkIsoVVL_Mu8_TrkIsoVVL Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL	17, 8
ee	Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	23, 12
μe	Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	23, 12
eμ	Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ	23, 8

Muons [184] are first reconstructed using information from the muon system, starting from L1 candidates. HLT muons (denoted *Mu* in Tab. 2.1) are then built by either propagating the track inwards, using hits from the inner tracker, or by starting from an inner track whose reconstruction was seeded by the muon candidate, and propagating it outwards through the muon system. In addition, HLT *tracker* muons (*TkMu* in Tab. 2.1) are reconstructed similarly to their offline counterparts (see Sec. 2.3.1), using L1 muons as seeds. Quality and p_T criteria are applied and these muons are required to be isolated, where the isolation is computed using the p_T of inner tracks within a cone $\Delta R = 0.3$ surrounding the muons (*TrkIso*).

The HLT electrons [194] are reconstructed by clustering energy deposits in ECAL, using e/γ candidates from the L1 trigger as seeds, and by searching for a compatible track. Identification requirements are applied on the track (*TrackId*) and on the cluster energy profile (*CaloId*), as well as on the compatibility between the track and the cluster. Electrons are required to be isolated, where the isolation is computed separately using energy in the tracker, ECAL and HCAL, in a cone $\Delta R = 0.3$, and corrected for pileup contributions.

The paths listed in Tab. 2.1 demand the presence of either at least two electrons, two muons, or one muon and one electron, with varying quality, isolation and p_T requirements. The criteria applied independently to each lepton are referred to as "leg". These paths were chosen for their low p_T thresholds, however due to the increase in the instantaneous luminosity in 2016 data taking, some of them had to be prescaled during era H. In that case, a non-prescaled version of the path remained available, but with an additional requirement on the longitudinal distance Δz between the points of closest approach of both leptons with the beam line: $\Delta z < 2\,\text{mm}$. These unprescaled paths are hence labelled "DZ" in Tab. 2.1.

2.3.10. Luminosity measurement

Several detectors are used by CMS to measure the luminosity delivered by the LHC: the DT, the HF, the pixel detector, the pixel luminosity telescope (PLT) and the beam conditions monitor (BCM). Most of these are primarily used for a fast and redundant determination of the instantaneous luminosity during data taking (online), crucial for diagnostics and optimisation of the LHC parameters.

The determination of the integrated luminosity used for offline data analysis is obtained chiefly using the pixel detector, following the approach sketched in Sec. 2.1.1. In the pixel cluster counting (PCC) method, the number of reconstructed pixel clusters defines the "event" rate R in (2.4). The low occupancy in the pixel detector ensures a good linearity between the number of clusters and the number of interactions μ, which is an implicit hypothesis in (2.4). Furthermore, it has been shown to provide an exceptionally stable response over time.

The stability and linearity of the PCC response is checked with the DT system, by comparing the relative rates measured in both detectors. Uncertainties of 0.6% and 1.5% are assigned to the measured luminosity to cover for a residual non-linearity and shifts in the relative rate over time, respectively.

An absolute calibration of the PCC luminosity is obtained through VdM scans (see Sec. 2.1.1). Systematic uncertainties on the absolute measurement include e.g. an uncertainty about the transverse beam profile (non-factorisability of bunch densities along x and y directions), on the effect of long-range beam-beam interactions, or on the beam separation length during scans. The total uncertainty in the integrated luminosity for 2016 data was estimated at 2.5% [205].

The number of interactions per bunch crossing can be computed by combining the measured luminosity with the total inelastic proton-proton cross section, which was measured as (71.3 ± 3.5) mb [206]. When building the distribution of the number of interactions, shown on Fig. 2.13, the luminosity is averaged for each LS.

2.3.11. Event simulation, corrections and calibrations

To bridge the gap between theoretical predictions provided in the form of samples of events generated under a fixed hypothesis, as described in Sec. 1.2, and the data recorded and reconstructed by CMS, it is essential to understand how the detector and the reconstruction procedure affect the events. This can be achieved by running, for each generated event, a *simulation* of the CMS detector. The simulation, based on the GEANT framework [207], features a detailed model of the geometry and material of the apparatus and describes the propagation of particles through the detector, including effects due to the magnetic field, energy losses, and electromagnetic and nuclear interactions with matter. It also handles the decay of long-lived particles, as well as the modelling of the response of each individual detector channel. The behaviour of the readout electronics (digitisation, ...) is reproduced using as input the simulated hits (*simhits*) given by GEANT. Finally, the very same reconstruction algorithms used for real data are employed to process the simulated events, producing collections of objects that can be compared one-to-one with data, but including additional information about the detailed simulated history of each event ("Monte-Carlo truth"). The simulation features random components

and can thus be seen as a Monte-Carlo integration over the phase space of microscopic degrees of freedom.

Modelling the effect of pileup is achieved by building a library of inelastic pp collisions (so-called minimum-bias events) with Pythia and Geant, randomly drawing events from this library following a given distribution, and overlaying their simhits with those of the hard scattering events being simulated. The remainder of the simulation chain then proceeds with these additional hits as input. Since the production of simulated event samples starts before the end of data-taking, one has to make an assumption about the ultimate shape of the pileup distribution. Hence, simulated events are reweighted to correct for the difference between assumed and actual distributions. The weight for an event with $\hat{\mu}$ interactions is given by the ratio between these two distributions, evaluated at $\mu = \hat{\mu}$. The pileup distribution in data was obtained using a minimum-bias cross section of 69.2 mb, since that value was shown to better accommodate pileup-sensitive observables. The uncertainty in that cross section, evaluated at 5%, translates into an uncertainty in the pileup reweighting. Resulting up- and down-variations of the weights are propagated through the analysis and are used to quantify the systematic uncertainty in pileup modelling.

Although the simulation of the detector is extremely detailed, it does not perfectly reproduce its true behaviour. This is due in part to the approximations and the shortcomings in the models used to simulate the interaction of particles with matter and in the material description of the detector. Moreover, as already mentioned the simulation is carried out before the end of data-taking, forcing one to make assumptions about the expected quality of the data and about one's knowledge of the relative alignment between subdetectors. Finally, data-taking conditions (alignment, active channels, ...) fluctuate during the year, whereas simulated samples assume a fixed set of conditions, aiming at an averaged-out description of the data. In particular, the dynamic inefficiency of the strip tracker mentioned in Sec. 2.3.1, affecting about 50% of the data, was not modelled in the simulation, resulting in a slightly lower efficiency of identification algorithms in data compared to the simulation. To account for all these differences, the performance of the different reconstruction algorithms is measured in data, and correction factors ("scale factors", SF) are applied on top of the simulation. In addition, the estimated uncertainties of these measurements are crucial to quantify one's confidence in the modelling of the detector. These corrections are briefly described in the following.

Lepton efficiencies and the Tag-and-Probe method

The efficiency to trigger on, reconstruct, identify and select electrons and muons can be measured in similar ways using the tag-and-probe (T&P) method, relying on the presence of the well-known $Z \to ee/\mu\mu$ resonance. In this setting, one lepton (the "tag") is required to pass tight identification criteria in order to achieve a very high purity. The other lepton (the "probe") has to pass only a basic selection $\mathcal{B}$, and the invariant mass of both leptons must lie close to the Z boson mass. The efficiency of the criterion $\mathcal{S}$, *relative* to the baseline $\mathcal{B}$, is then estimated from the number of probes passing $\mathcal{B}$ and $\mathcal{S}$, i.e. $\epsilon_{\mathcal{S}}^{\text{data}} = N_{\mathcal{S}}/N_{\mathcal{B}}$. In order to disentangle true probe leptons from backgrounds, $N_{\mathcal{S}}$ and $N_{\mathcal{B}}$ are obtained by performing parametric fits of the dilepton invariant mass distribution close to the Z resonance, before and after application of $\mathcal{S}$. To account for

the dependency of the efficiency on lepton kinematics, the measurement is repeated in windows of the probe's p_{T} and η. The uncertainty in the measurement stems from statistical uncertainties in the fits, as well as a systematic uncertainty estimated by reproducing the fits using alternate parameterisations or tag criteria. The procedure is repeated using simulated $\mathrm{Z} \to \mathrm{ee}/\mu\mu$ events to compute the Monte-Carlo efficiency $\epsilon_{\mathcal{S}}^{\mathrm{sim.}}$, for which the uncertainty is estimated i.a. by using different event generators. Since no backgrounds are present in the simulation, a simple counting approach is employed. The efficiency in the simulation can therefore be corrected by weighting events with scale factors $\mathrm{SF}_{\mathcal{S}} = \epsilon_{\mathcal{S}}^{\mathrm{data}}/\epsilon_{\mathcal{S}}^{\mathrm{sim.}}$, using as many SFs as there are leptons and selection criteria $\mathcal{S}$ to correct.

We have measured the efficiency of the trigger paths listed in Tab. 2.1 using the T&P method. With T&P, efficiencies are measured for each *leg*, i.e. for each specific subset of criteria applied independently on each lepton. For lack of proper emulation of the L1 trigger in the simulation, no attempt was made to require the firing of these paths on simulated events. Hence, no SFs were applied, but simulated events were directly weighted by the trigger efficiency as measured in data.

A condition for the latter procedure to work is that the identification and isolation criteria required on electrons and muons in the analysis be stricter than those used by the HLT paths. Since this was not the case for electrons, we required that they also passed an *HLT-safe* ID, essentially consisting in a tighter isolation requirement, on top of the medium ID described in Sec. 2.3.4. Unfortunately, these criteria were tuned for single-electron trigger paths, resulting in a significantly lower selection efficiency for electrons as compared with muons.

In part of 2016 data, the muon L1 trigger was affected by a misconfiguration of the EMTF, which would send only one muon candidate per 60°-wide azimuthal sector (instead of up to three). This created an inefficiency for double-muon triggers on events where both muons ended up in the same EMTF sector, which affected 52.6% of the recorded luminosity. Our efficiency measurement for the relevant muon legs took this issue into account, and we corrected the simulation by applying a weight of 0.526 on relevant events.

Efficiencies as a function of lepton p_{T} and η are shown on Fig. 2.14 for two of the six legs for which the measurement was carried out. The appearance of a turn-on curve, instead of a sharp drop in efficiency below the p_{T} threshold, is due to the worse resolution on the HLT lepton's p_{T} (on which the threshold is applied) compared to the fully reconstructed lepton (w.r.t. which the efficiency is plotted).

The efficiency of an HLT path given the efficiencies for each of its two legs is computed as follows. For same-flavour triggers we define ϵ_{ij} as the efficiency of lepton i to pass the leg j of the path, where leg 1 and 2 refers to the high- and low-p_{T} leg, respectively. The path's efficiency is then given by:

$$\epsilon_p = \epsilon_{12}\,\epsilon_{21} + \epsilon_{11}\,\epsilon_{22} - \epsilon_{11}\,\epsilon_{21} \qquad \text{(same flavours)}, \tag{2.14.}$$

where we have taken into account that a lepton that fired the high-p_{T} leg of a path will have automatically fired its low-p_{T} one. The reasoning behind this expression is explained schematically on Fig. 2.15. For different-flavour triggers, we denote by ϵ_{ij} the

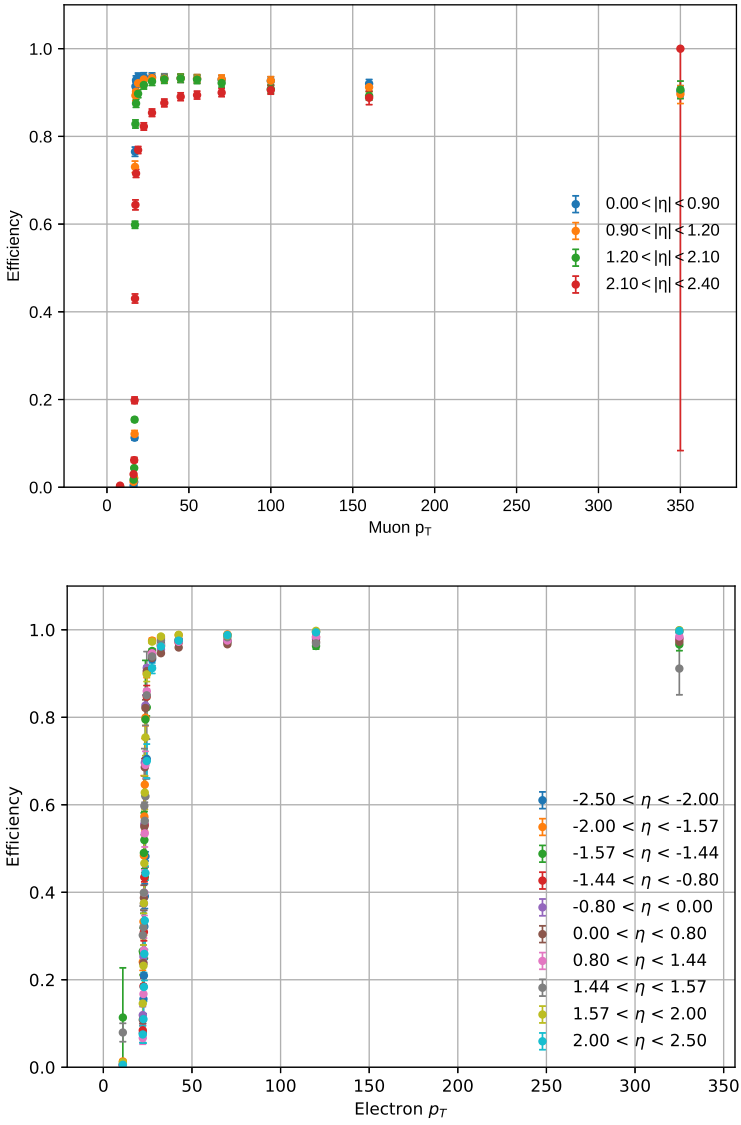

Figure 2.14. Efficiencies measured with the T&P method as a function of lepton p_T, in different η or $|\eta|$ ranges, of the muon leg `Mu17_TrkIsoVVL` used by double-muon triggers (top) and the electron leg `Ele23_CaloIdL_TrackIdL_IsoVL` used both in double-electron and muon-electron trigger paths (bottom). Error bars correspond to both statistical and systematic uncertainties.

efficiency of the lepton of flavour i to pass the leg of flavour j of the path, and we have simply:

$$\epsilon_p = \epsilon_{11}\,\epsilon_{22} \qquad \text{(different flavours)}. \tag{2.15.}$$

Measurement uncertainties in ϵ_{ij} are propagated to ϵ_p, yielding total uncertainties in the trigger efficiency of about 1%.

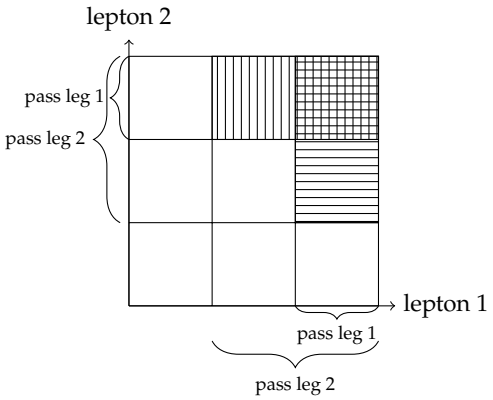

Figure 2.15. Schematic explanation of (2.14). The probabilities for lepton 1 (lepton 2) to pass the high-p_{T} or low-p_{T} trigger legs, denoted respectively by leg 1 and 2, correspond to the horizontal (vertical) segmentations. The hatched area is the probability for an event to be selected by the trigger, ϵ_p. The first term of (2.14) is represented with vertical hatching, the second with horizontal hatching, and the third term arises from the double counting in the upper right corner of the figure. Key for this representation are the facts that both leptons are processed independently, and that a lepton passing the high-p_{T} leg automatically satisfies the low-p_{T} one.

Some of the considered HLT paths were present in two versions: with or without DZ filters in addition to the lepton legs. We estimate these filters' efficiency using data in the following way:

1. Count the number of events passing the selection and the non-DZ version of the path (= D).
2. Count the number of events passing the selection and both DZ and non-DZ versions of the path (= N).
3. The filter efficiency is then $\epsilon_F = N/D$.

For the double-electron path, the DZ filter is always required, hence its efficiency can be used as is. For double-muon and muon-electron paths however, both DZ and non-DZ versions exist, and the non-DZ version is unprescaled for more than half the integrated luminosity. Since we use both versions, we see only the non-DZ version when it is unprescaled, and only the DZ one in the opposite case. Hence, the effect of the filter averaged over the whole data-taking period is $\overline{\epsilon}_F = f + (1-f)\epsilon_F$, where f is the fraction of luminosity where the non-DZ version is unprescaled[1]. The final path efficiency is thus

[1] Strictly speaking, this neglects the cases where the DZ path doesn't fire, but the prescaled non-DZ one does. The overall efficiency $\overline{\epsilon}_F$ is thus biased low due to a missing term $(1-f)(1-\epsilon_F)/\langle S\rangle$, where $\langle S\rangle$ is the average prescale factor of the non-DZ path. However, this correction of less than 1 % is expected to have a negligible effect on the analysis.

given by $\epsilon_p \cdot \overline{\epsilon}_F$. The results for the different channels are shown in Tab. 2.2. To account for a possible run-dependence of the filter efficiencies (related to pileup conditions and the strip tracker dynamic inefficiency), we compute them using only events in runs where the non-DZ versions are prescaled. The statistical uncertainty of the procedure detailed above is negligible.

The efficiency of the electron identification criteria used in this analysis has also been measured using the T&P method and is shown on Fig. 2.16. These measurements are combined with those relative to the GSF electron reconstruction [208]. On average, uncertainties in these measurements amount to about 2%.

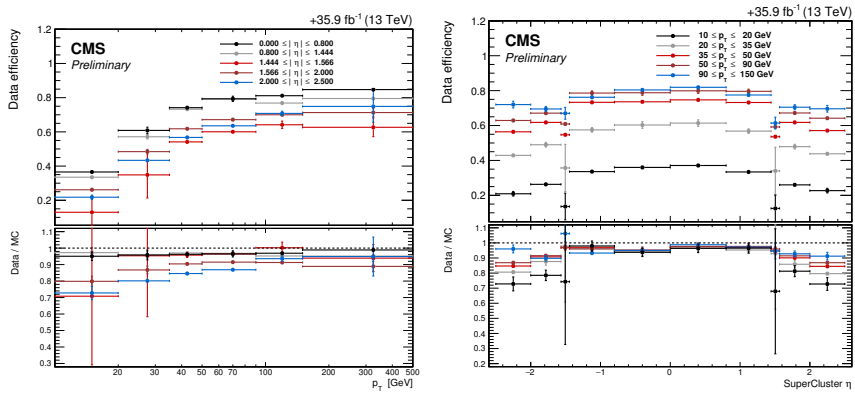

Figure 2.16. | Identification efficiency on data (top panels) and scale factors (bottom panels) as a function of electron p_{T} (left) and supercluster η (right). Error bars correspond to both statistical and systematic uncertainties.

Finally, for muons the efficiencies of tracking, identification and isolation (as introduced in (2.10)) in the simulation are each corrected using a dedicated set of SFs provided by the CMS collaboration [192, 196], shown on Fig. 2.9. The precision on each of these SFs is better than 1%.

Lepton energy scale and resolution

The electron reconstruction method presented in Sec. 2.3.4 does not fully account for effects such as energy leakage out of the superclusters, in gaps between crystals and in the

Table 2.2. | Raw and year-averaged DZ filter efficiencies for the combination of HLT paths used in the different channels, along with the fractions f of the total luminosity where the non-DZ versions of the paths are unprescaled.

Channel	DZ filter efficiency (ϵ_F)	f	Averaged efficiency ($\overline{\epsilon}_F$)
ee	98.3%	/	98.3%
$\mu\mu$	97.0%	76.7%	99.3%
μe	95.1%	75.5%	98.8%
eμ	92.4%	76.3%	98.2%

HCAL, energy losses in the tracker, and additional energy from pileup interactions [194]. To remedy these, the SC energies are corrected using a multivariate regression trained on the simulation. The regressed target is the true electron energy, whereas input variables are related to SC position and shape, and to pileup. This regression, applied both on data and simulation, reduces the bias and improves the resolution of the reconstructed electron energy.

Both electrons and muons in the data are subject to a residual offset in their momentum scale, i.e. a shift between the average reconstructed p_{T} and true p_{T}. The scale offset is due to residual errors in the alignment of the tracker and the muon system, and to imperfect corrections for the temporal evolution of ECAL crystal transparency and noise. In addition, owing to a different tracker misalignment between data and simulation and to an imperfect description of the ECAL material in simulation, both the scale offset and the momentum resolution suffer from a mismodelling. For muons, consequences are an undesired dependency on muon charge, ϕ and η of the reconstructed Z boson mass in $\mathrm{Z} \to \mu\mu$ events, and a shift and a distortion in the shape of the reconstructed Z mass distribution. These observations are used to apply a residual correction on the p_{T} of muons based on their charge and direction ("Rochester" correction), in data and simulation [184]. A similar effect is visible for electrons in $\mathrm{Z} \to \mathrm{ee}$ events, where in addition the reconstructed Z mass is observed to vary in time. This temporal dependency is removed by correcting the energy scale in data, so that the peak of the Z mass distribution matches that in the simulation. The agreement between data and simulation is further improved by altering the SC energy resolution in the simulation to match the one observed in data [194, 208].

Jet energy scale and resolution

Due to the nonlinearity of the calorimeter response and despite the calibration of individual PF clusters, there is a nontrivial relationship between a jet's true energy and its measured energy. In the simulation, "true" jets are defined by applying the same jet clustering algorithm used on PF candidates (in our case, the anti-k_{T} algorithm with parameter $R = 0.4$) on the particles given by the event generator, excluding neutrinos. These "particle-level" jets are geometrically matched to reconstructed jets, and the jet response is defined as the ratio between reconstructed and true jet energy. Several corrections are applied to bring the jet response as close as possible to unity, and to improve the agreement between data and simulation [209]. These jet energy corrections (JECs) are applied sequentially and each modify the jet 4-momenta by a multiplicative factor. All the corrections detailed below are propagated to the measured $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ to ensure a consistent global description of the events.

The first step is an offset correction to remove contributions from pileup, determined as a function of event ρ and jet p_{T}, η and area. This procedure was already mentioned in Sec. 2.3.8; it suffices to say that to account for differences between data and simulation, slightly different corrections are applied in either. In the second step, common to data and simulation, the reconstructed jets are corrected to ensure an uniform response as a function of p_{T} and η. The correction factors are derived from the simulation. The last step is needed to cure residual differences in the jet energy scale between data and simulation. These corrections are obtained from the data using di-jet, γ + jets and $\mathrm{Z}(\ell\ell)$ + jets events,

and depend on jet p_T and η. Uncertainties in the measurement of these corrections can be broken down into 27 different sources of statistical and systematic uncertainties (not detailed here). Each of these sources yields two different JEC factors, corresponding to up- and down variations of the correction within the corresponding uncertainty, that are applied on the reconstructed jets and propagated throughout the analysis. The data-to-simulation scale factors range from 1 % to 2 %, with total uncertainties of the same order [210].

In addition, the jet energy resolution (JER) is observed to be 10 % to 15 % worse in data than in the simulation [209]. To remedy this, simulated jets are smeared to worsen their resolution. Given the ratio s_{JER} between measured and simulated jet resolutions, obtained as a function of jet $|\eta|$, the jets which can be matched to a particle-level jet see their momenta rescaled by a factor:

$$c_{\text{JER}} = \max\left(0, 1 + (s_{\text{JER}} - 1)\frac{p_T - p_T^{\text{true}}}{p_T}\right). \tag{2.16.}$$

If no particle-level jet can be matched to a jet, its momentum is rescaled stochastically by a factor:

$$c_{\text{JER}} = \max\left(0, 1 + \mathcal{N}(0, \sigma_{\text{JER}})\sqrt{\max(0, s_{\text{JER}}^2 - 1)}\right), \tag{2.17.}$$

where σ_{JER} is the relative p_T resolution in the simulation, and $\mathcal{N}(0,\sigma)$ denotes a random number sampled from a normal distribution with a zero mean and standard deviation σ. Uncertainties on the scale factors s_{JER} range from 1 % to 5 %. As for the JEC, varying the scale factors up and down within one standard deviation yields two alternate jet collections that can be used to estimate the impact of these uncertainties on the analysis.

b tagging efficiency

The variables used as input to the b tagging algorithms, as well as the distributions of the tagger scores, are not entirely reproduced by the simulation [201]. This is mostly due to the sensitivity of these variables, such as track IP, to the (mis-)alignment of the tracker. An imperfect description of the tracker material budget, as well as a mismodelling of the parton shower and hadronisation processes, can also contribute to these discrepancies. Additionally, in 2016 the strip tracker dynamic inefficiency resulted in a lower b tagging performance in part of the data, namely a lower efficiency to correctly tag b jets and a higher probability to mistag b or light jets, compared to the simulation. To calibrate the performance of the taggers, the efficiency to tag a jet of flavour F is measured as a function of jet p_T in the data (ϵ_F^{data}) and in the simulation (ϵ_F^{sim}), which defines scale factors $\text{SF}_F = \epsilon_F^{\text{data}}/\epsilon_F^{\text{sim}}$ with which the simulation can be corrected. The SFs used in this analysis are determined and provided by the CMS collaboration. Using these SFs, a weight is assigned to each jet in the simulation based on its "true" flavour. The true flavour is defined from the content of particle-level jets matched to reconstructed jets: if the jet contains a B or D hadron, it is considered a b- or c-flavour jet, respectively. If the particle-level jet contains no heavy hadron, or if the jet cannot be matched with any particle-level jet, the jet is considered as a light-flavour jet. Simulated events are

reweighted by combining the jet weights for all jets in the events. Uncertainties in the measured SFs are propagated to the event weights, separately for light and b/c jets, to assess their impact on the analysis.

The measurement of tagging efficiencies in data relies on a variety of methods. The light-jet mistag rate is obtained from multijet events using the "negative-tag" method. In this method, "negative" and "positive" tagger scores are computed using only tracks with negative and positive IP, and SVs with negative and positive flight distances, respectively. Light jets dominate the distribution of negative scores, and contribute approximately symmetrically to the negative and positive scores. Modulo a correction factor obtained from the simulation, the efficiency to select jets in data using negative taggers yields the light-jet mistag rate for the corresponding tagger. Light jet SFs vary from 1.1 to 1.3, depending on jet p_{T}, and suffer from uncertainties of about 10%.

For the cMVAv2 tagger used in this analysis, b-jet scale factors are measured using samples of $\mathrm{t\bar{t}}$ events. In the "Kin" method, a multivariate discriminant is built on simulated dilepton $\mathrm{t\bar{t}}$ events to separate b jets from light jets, using only kinematic information in the events. The b-jet tagging efficiency in data is then obtained from a template fit [1] to the distribution of the discriminant, before and after applying the tagger on the jets. In addition, the T&P method is applied to semileptonic $\mathrm{t\bar{t}}$ events. First, a kinematic reconstruction of the $\mathrm{t\bar{t}}$ system is attempted, yielding an assignment between jets and b (light) quarks from top quark (W boson) decays. This procedure also delivers, for each event, a likelihood quantifying the agreement between the $\mathrm{t\bar{t}}$ hypothesis and the event content. Then, the "tag" jet is obtained by applying the CSVv2 tagger on one of the b jets, whereas the other candidate b jet defines the "probe". Finally, the b-jet tagging efficiencies are obtained from a combined template fit to the distributions of the kinematic likelihood and $p_{\mathrm{T}}^{\mathrm{miss}}$, before and after applying the tagger on the probe jet. In both these methods, the templates for different jet flavours are obtained from the simulation. The measurements are repeated in different bins of jet p_{T}, and the results from both methods are combined. The SFs vary from 0.92 to 0.96 depending on jet p_{T}, with uncertainties of 1 % to 7 %.

2.4. Analysis methods

Before describing the analysis itself, we shall lay out the fundamental techniques that will be instrumental in extracting meaningful results out of the collected data. These concern the problem of distinguishing the phenomena we are interested in from the deluge of backgrounds present in the selected events, as well as the necessity of interpreting the data in a statistical manner. The generality of the employed methods means that they may be described without reference to the particular case that concerns us, so that we can concentrate on the actual analysis in the following chapter.

2.4.1. Machine learning techniques for enhanced sensitivity

In many of the identification algorithms described in Sec. 2.3, as well as in physics analyses, there is an obvious need to *discriminate* a *signal* from one or several *backgrounds*. The power of a discriminating algorithm can be characterised e.g. in terms of true and

[1] See Sec. 2.4.2 for a description of the statistical tools being used.

false positive classification rates (in other words, efficiency and purity). Algorithms might use a single physical variable to separate signals from backgrounds, however their power can be significantly enhanced by combining several variables into an artificial score (often without direct physical interpretation) upon which the classification decision is based. Such *multivariate* discriminants can be constructed using machine learning (ML) techniques, two of which are used in this work and are briefly described below.

Given the *target* $y \in \{0,1\}$ and a set of M variables or *features*, $\boldsymbol{x} = (x^1, \ldots, x^M)$, if signal and background respectively follow the probability density functions (pdfs) $p(\boldsymbol{x}|y=1)$ and $p(\boldsymbol{x}|y=0)$, the Neyman-Pearson lemma [211] states that the most powerful discriminating variable between the signal (defined by $y=1$) and the background ($y=0$) is given from the likelihood ratio,

$$\Lambda(\boldsymbol{x}) = \frac{p(\boldsymbol{x}|y=1)}{p(\boldsymbol{x}|y=0)}. \tag{2.18.}$$

In other words, of all functions $f(\boldsymbol{x})$ such that a requirement $f(\boldsymbol{x}) > c_f$ retains a fixed fraction of the signal, the likelihood ratio is the one with which such a requirement rejects the largest proportion of background. However, in practice $p(\boldsymbol{x}, y)$ is rarely known in closed form, and we can only rely on a finite sample of independent realisations (events), $\mathcal{D} = \{(\boldsymbol{x}, y)_i \sim p(\boldsymbol{x}, y), i = 1 \ldots N\}$, obtained e.g. using a simulator (see Sec. 2.3.11). A goal of ML is to provide methods to construct, i.e. *train* a *discriminator* (or *classifier*) function $F(\boldsymbol{x})$ that approaches the likelihood ratio, or a bijection of it, using only the examples provided by the dataset $\mathcal{D}$. Ideally, this discriminator should have:

1. Low bias: the model should be flexible enough, and trained in a manner as to provide good separation between the signal and background hypotheses.
2. Low variance: The separation power should be robust when applying the trained model on a dataset that is statistically independent from the one used for training, and sampled from the same distribution $p(\boldsymbol{x}, y)$.

Two popular methods (among many others) to build such a discriminator are boosted decision trees (BDTs) and multilayer perceptrons (MLPs).

Boosted Decision Trees

A decision tree can be represented as a sequence of binary splits on the input features, as depicted on Fig. 2.17. This procedure builds rectangular regions in the input space, which can be then assigned a score of +1 or -1 if they are signal- or background-dominated[1]. In other words, a tree is a step function $h : \boldsymbol{x} \to h(\boldsymbol{x}) \in \{-1, 1\}$. Each split is defined by the variable that is used to cut on, and the position of the cut. Both choices are determined using an impurity criterion $I(p_n)$, function of the signal purity p_n in node n, that is maximal for $p_n = 0.5$ (no discrimination) and minimal for $p_n = 0$ or $p_n = 1$ (perfect discrimination). The chosen split should maximise the gain $G = I(p_m) - f_1 I(p_{c_1}) - f_2 I(p_{c_2})$, where c_1 and c_2 are the two "child" nodes obtained by splitting the "mother" node m, and f_i is the fraction of events in m falling into node c_i. A widely used criterion is the Gini coefficient $I_G(p) = p(1-p)$. This iterative tree growing procedure can be stopped

[1] In this section, we take $y \in \{-1, 1\}$ instead of $\{0, 1\}$, following common usage for BDTs in the literature.

by requiring a maximum depth (number of consecutive splits) or a minimum number of events in the final nodes.

Trees built in this way are subject to high variance (overtraining), i.e. they are strongly sensitive to statistical fluctuations in the training sample and do not generalise well. A successful method to alleviate overtraining is *boosting*, in which a large number T of shallow trees $h_t(\boldsymbol{x})$, grown with only a few splits, are combined by taking a weighted average of their output scores: $F_T(\boldsymbol{x}) = \sum_{t=1}^{T} \alpha_t h_t(\boldsymbol{x})$. Each single shallow tree has poor discrimination power but is less prone to overtraining, and the ensemble average of such *weak learners* results in a *strong learner* with high and stable performance.

The boosting algorithm used in this work is AdaBoost [212]. First, the outputs are set e.g. to $F_0(x) = 0$, and events are assigned equal weights $w_{i,0} = 1/N$. At iteration t, a weak learner $h_t(\boldsymbol{x})$ is grown according to the strategy detailed above, typically with only two or three splits. The weighted classification error $\epsilon_t = \sum_{i,h_t(\boldsymbol{x}_i)\neq y_i} w_{i,t}$ is then computed, and with $\alpha_t = \frac{1}{2}\log\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$ the new tree is added to the ensemble as $F_t(\boldsymbol{x}) = F_{t-1}(\boldsymbol{x}) + \alpha_t h_t(\boldsymbol{x})$. The event weights are then updated as $w_{i,t+1} = Z \cdot w_{i,t} E(y_i, h_t(\boldsymbol{x}_i))^{\alpha_t}$, where the error function is $E(y,h) = e^{-yh}$ and Z is a normalisation factor s.t. $\sum_i w_{i,t+1} = 1$. Thus, misclassified events at iterations $\leq t$ are given increased weights in the creation of the tree at iteration $t+1$, and trees which achieve a small classification error contribute more to the ensemble average.

Other ensemble methods, such as gradient boosting (of which AdaBoost can be seen as a particular case) [213] and bootstrap aggregation (bagging) [214], will not be detailed here. Finally, it should be clear that those methods can be applied to any form of learners, but have been particularly successful when used in conjunction with decision trees.

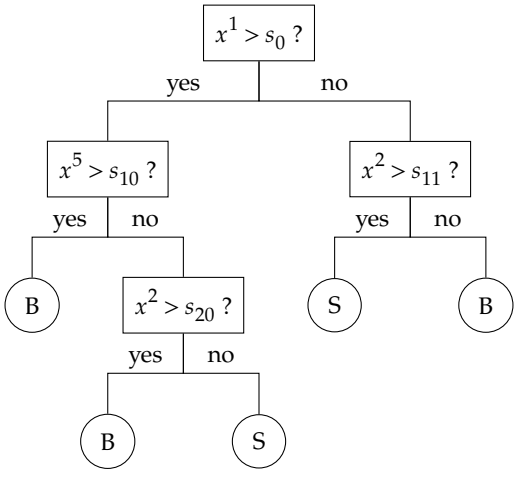

Figure 2.17. | Simple realisation of a decision tree, with four splits (on features x^1, x^2 and x^5) resulting in five final nodes or "leaves", two of which are signal-dominated ("S"), and three of which are background-dominated ("B"). When evaluating the tree, samples falling into an "S" or "B" node are assigned the score +1 or −1, respectively.

Multilayer Perceptrons

The task of building a classifier $F(\boldsymbol{x})$ that discriminates between $y = 0$ and $y = 1$ can be approached by defining a *loss function* $L(F, y)$ that quantifies the quality of the classification achieved by F, and finding F such that the average loss $\mathcal{L}$ over the previously defined dataset $\mathcal{D}$ is minimised:

$$\mathcal{L}(F, \mathcal{D}) \equiv \sum_{i=1}^{N} L(F(\boldsymbol{x}_i), y_i). \tag{2.19.}$$

If we now assume a parameterised family of functions $F_{\boldsymbol{\theta}}(\boldsymbol{x})$, and if the loss $L(F_{\boldsymbol{\theta}}, y)$ is a differentiable function of the parameters $\boldsymbol{\theta}$, then the classification problem is reframed as a well-known optimisation problem. A commonly chosen loss function for classification is the cross-entropy loss:

$$L(F, y) = -y \log(F) - (1 - y) \log(1 - F), \tag{2.20.}$$

for which the minimisation of (2.19) can be interpreted as a maximum-likelihood fit of $\boldsymbol{\theta}$, using the set of observations $\mathcal{D}$, assuming y_i follows a Bernoulli law with probability of success $p(y_i = 1|\boldsymbol{x}_i)$ estimated by $F(\boldsymbol{x})$. In practice, other choices such as the square loss $L(F, y) = (F - y)^2$ might work just as well. Indeed, given either the cross-entropy or squared loss, it can be shown that asymptotically the expected loss $\mathcal{L}$ is minimised when

$$F(\boldsymbol{x}) = p(y = 1|\boldsymbol{x}) = \left(1 + \frac{p(\boldsymbol{x}|y = 0)}{p(\boldsymbol{x}|y = 1)} \frac{p(y = 0)}{p(y = 1)}\right)^{-1}, \tag{2.21.}$$

which is one-to-one with the likelihood ratio defined in (2.18), i.e. the best possible classifier in terms of statistical power.

There are many possible choices for parameterising the functions $F_{\boldsymbol{\theta}}(\boldsymbol{x})$. A particularly successful approach is that of multilayer perceptrons (MLPs), designed by *analogy* with biological brains. The original perceptron model [215] is:

$$F_{\boldsymbol{\theta}}(\boldsymbol{x}) = \sigma\left(\theta_0 + \sum_{j=1}^{M} \theta_j x^j\right), \tag{2.22.}$$

where σ is a nonlinear *activation* function, e.g. a sigmoid $\sigma(x) = \left(1 + e^{-x}\right)^{-1}$, and $\boldsymbol{\theta}$ are weights (bias and synapse strengths) which can be *learned* by minimising the loss (2.19) using a training sample $\mathcal{D}$. In general, unless signal and background are linearly separable, functions of the form (2.22) are not sufficiently flexible to provide a high discrimination power. However, by chaining several layers of multiple perceptrons ("neurons"), one layer's outputs being used as the next layer's inputs (see Fig. 2.18), one obtains a highly versatile, nonlinear function of the input features: an MLP, also called feed-forward artificial neural network (ANN).

Using the above definitions, the loss $\mathcal{L}(F_{\boldsymbol{\theta}}, \mathcal{D})$ is a complex, non-convex, differentiable function of the weights. Fortunately, it is feasible to minimise it using gradient descent, since the gradient $\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})$ can be efficiently computed by *back-propagation* [216] (a direct

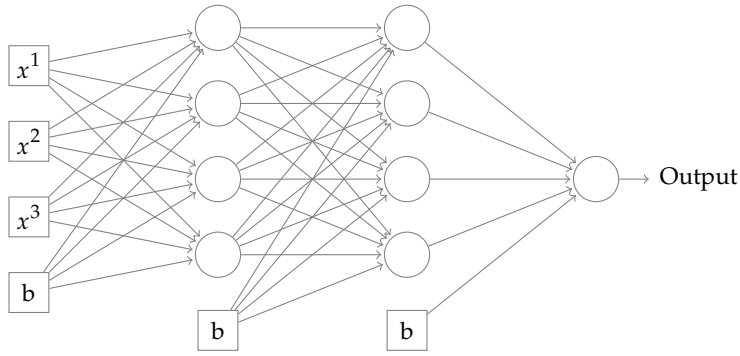

Figure 2.18. Example of an artificial neural network, with three input variables and two hidden layers of four neurons. Each circular node is a function of the form (2.22), working on the previous layer's values. Nodes labelled "b" are biases (constants). Edges between nodes are synapses; each corresponds to a weight that has to be learned by minimisation of the loss computed from the output value of the node on the far right.

consequence of the derivative chain rule). At each iteration, the weights are updated as $\boldsymbol{\theta}_t \to \boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \boldsymbol{\nu}_t$, $\boldsymbol{\nu}_t = \eta \cdot \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})$, where η is the *learning rate*.

The danger of the minimisation (learning) procedure getting stuck in a local minimum can be reduced by using stochastic gradient descent (SGD), in which the gradient is not computed using the full dataset, but on sub-samples (mini-batches) of size $B \ll N$ (strictly speaking, SGD corresponds to $B = 1$). In addition, once the number of layers and neurons is sufficiently large, local minima become rare compared to saddle points, and finding the global minimum can be undesirable since it might correspond to an overtrained model [217]. However, saddle points still represent a challenge for plain gradient descent since the gradient becomes vanishingly small around those points. Numerous strategies exist to mitigate this issue, some of which are to:

- Add "momentum" [218] by updating the weights on each batch using a combination of the gradient and the previous batch's step: $\boldsymbol{\nu}_t = \gamma \cdot \boldsymbol{\nu}_{t-1} + \eta \cdot \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})$, for some chosen γ.
- Use Nesterov accelerated gradient [219], where the gradient is computed at the position given by the momentum step: $\boldsymbol{\nu}_t = \gamma \cdot \boldsymbol{\nu}_{t-1} + \eta \cdot \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta} - \gamma \cdot \boldsymbol{\nu}_{t-1})$.
- Consider different learning rates for each parameter θ_i, as larger steps can be taken in flat directions. Each weight's step size is adapted using the magnitude of past squared gradients in that direction, while annealing the learning rate over time (Adagrad [220]).
- Compute a running, exponentially decaying averages of past squared gradients to rescale the learning rates, such as with Adadelta [221] or RMSprop, which keeps the learning rate from dropping too quickly (a deficiency of Adagrad).
- Combine these adaptive strategies with a momentum term computed as a decaying average of past gradients (Adam [222]).

Just as plain decision trees, complex ANNs are sensitive to overtraining. A straightforward way to avoid overfitting the network is to compare the value of the loss function evaluated on the training sample and on an independent test sample during training,

and to stop the training if they start diverging [223]. However, it is more efficient to also regularise the model and the training themselves, using different methods:

- Add a constraint term $+\lambda \sum_i \theta_i^n$ to the loss function, with $\lambda > 0$ and usually $n = 2$ (L_2-regularisation) [223]. This prevents the weights from growing too quickly and accommodate statistical fluctuations in the training sample.
- With dropout [224, 225], nodes in a given layer are inactivated (fixed to 0) randomly for each mini-batch update during training, with a fixed probability p. For inference, the outputs from that layer are then weighted by $1-p$. In effect, this replaces the full network with a random ensemble of smaller networks. Similarly to tree ensemble methods, the resulting model is less prone to overfitting.
- A recent suggestion [226] is to normalise each hidden node's output to have zero mean and unit variance, before feeding it to the next layer's nodes ("batch normalisation"). The normalisation is computed from the mean and variance of each mini-batch during training, and from mean and variance of the whole training set for inference. This procedure both speeds up the training and prevents overfitting.

Apart from the previous considerations, training an ANN can be hindered by a poor choice of activation function. For instance, the sigmoid mentioned above yields hard-to-train networks as it easily saturates, i.e. its gradient vanishes if its argument becomes large. Rectifiers, such as $\mathrm{ReLU}(x) = \max(0, x)$, are not subject to this deficiency and still provide sufficient nonlinearity to not impede the network's fitting capacity.

The ability to find a minimum of the loss function also depends on the initial values of the parameters $\boldsymbol{\theta}$. Choosing a good initialisation procedure depends on the activation functions used [227]. For instance, setting $\theta_i < 0$ with ReLU or $\theta_i \gg 1$ with sigmoid activations might create a vanishing-gradient problem. Usually, parameters are initialised by drawing from a random distribution, and several strategies for choosing this distribution have been suggested, see e.g. Refs. [223, 228, 229].

2.4.2. Statistical model and tools

We want to produce statements about the theory using the data at hand, and such inference is possible only once we have defined a statistical model of the data. The theory parameter we shall be interested in is the *signal strength* $\mu \geq 0$, where $\mu = 1$ corresponds to the presence of the signal according to theory predictions or some pre-defined normalisation, and $\mu = 0$ corresponds to an absence of signal. The signal strength is often only a proxy for model parameters such as the mass m of a particle, in which case a statistical exclusion $\mu(m) < 1$ translates into an excluded value of m. In other cases the signal cross section might be proportional to unconstrained couplings and thus be free to vary, hence statements on μ can be directly related to the signal cross section.

In addition to μ, other parameters $\boldsymbol{\alpha}$ need to be included in the statistical model. These *nuisance parameters* are linked with imperfect knowledge of both theory and detector. The model used in this work follows a cut-and-count approach, in which the number of data events n_c falling into mutually exclusive regions c is counted and compared with the theory predictions $\nu_c(\mu, \boldsymbol{\alpha})$. These regions can be *channels* (orthogonal event selections) and/or bins in histograms. The likelihood is obtained from the product of

the Poisson probabilities for each region [230]:

$$L(\mu,\boldsymbol{\alpha}) = L(\boldsymbol{n}|\mu,\boldsymbol{\alpha}) = \prod_c \mathrm{Pois}\left(n_c, \nu_c(\mu,\boldsymbol{\alpha})\right) = \prod_c \frac{\nu_c^{n_c}}{n_c!} e^{-\nu_c} \tag{2.23.}$$

Note that this formulation is equivalent to a marked Poisson process where the densities ("templates") are step functions given by properly normalised histograms for signal and backgrounds. The expected *yield* in each bin or region is given by $\nu_c(\mu,\boldsymbol{\alpha}) = \mu\hat{\sigma}_{s,c}(\boldsymbol{\alpha}) + \sum_b \hat{\sigma}_{b,c}(\boldsymbol{\alpha})$, where $\hat{\sigma}_s$ and $\hat{\sigma}_b$ are the yields of the signal and the different background processes contributing to that bin, respectively. For every bin and process we have:

$$\hat{\sigma}_{i,c} = \mathcal{L}_{\mathrm{int}} \cdot \sigma_i \cdot \epsilon_{i,c} = \mathcal{L}_{\mathrm{int}} \cdot \sigma_i \cdot \frac{\sum_{j\in c} c_j w_j}{\sum_j w_j}, \tag{2.24.}$$

where $\mathcal{L}_{\mathrm{int}}$ is the integrated luminosity, σ_i is the cross section of process i, and $\epsilon_{i,c}$ is the acceptance/efficiency, i.e. the probability for an event of process i to end up in bin c. It can be computed using the weights w_j of the simulated events for that process (see Sec. 1.2), possibly modified by event-dependent correction factors c_j (see Sec. 2.3.11).

In (2.23), nuisance parameters are free to float. However, we would like to incorporate some "prior" knowledge about their possible values, resulting e.g. from auxiliary measurements of these parameters, such as those detailed in Sec. 2.3.11. It is not practical to include the full likelihoods of these measurements into the statistical model (2.23), but we can idealise them using constraint terms. Thus, for each nuisance α_p we multiply (2.23) with a constraint term $f(a_p|\alpha_p)$, where a_p is some default/measured value of that parameter. This procedure ensures that a consistent frequentist treatment of the model is possible. We thus arrive at the complete statistical model:

$$L(\mu,\boldsymbol{\alpha}) = L(\boldsymbol{n},\boldsymbol{a}|\mu,\boldsymbol{\alpha}) = \prod_c \frac{\nu_c^{n_c}}{n_c!} e^{-\nu_c} \prod_p f(a_p|\alpha_p) \tag{2.25.}$$

Nuisances affecting the normalisation of a process, entering as an overall multiplicative term in (2.24) (independent of c), are usually assigned log-normal constraint terms with a width reflecting the uncertainty in their measured value a_p. However, many sources of systematic uncertainty affect in a non-trivial way both the overall normalisation and the distribution across bins of the predictions entering (2.25). We model each of those *shape* uncertainties by building alternate histograms $\hat{\sigma}^{\pm}_{p,i,c}$ using up and down variations, within one standard deviation, of the quantity associated with that uncertainty. These variations, along with the nominal prediction, are interpolated and extrapolated to yield a continuous dependence of the yields on the nuisance parameter. The "down", "nominal" and "up" histograms are associated with the values $\alpha_p = -1,0,1$. The inter-/extrapolation is done "vertically", i.e. using independently the yields of every bin of the templates. Working on *normalised* histograms, we consider a quadratic interpolation over $-1 \leq \alpha_p \leq 1$, continued linearly beyond these bounds. The overall normalisation is interpolated linearly in log-space, ensuring $\hat{\sigma} \geq 0$ for every α_p, and we add a Gaussian constraint term on α_p with unit variance and zero mean.

Since the amount of simulated events used to construct the templates entering the

likelihood is finite, we need to consider a statistical uncertainty on the predicted yields $\hat{\sigma}_{i,c}$ in (2.24). Following the suggestion of Ref. [231], for each process i and bin c we introduce an additional nuisance parameter scaling the yield according to its uncertainty. To prevent the number of nuisance parameters from becoming too large, the yield uncertainties on processes contributing to less than half of the statistical uncertainty on the total yield of a given bin are added in quadrature and assigned a single nuisance parameter. This procedure is commonly referred to as the Barlow–Beeston lite method.

The search for a new signal may be formulated as an hypothesis test, with the null H_0 being the absence of signal, and the alternative H_1 being the presence of a signal on top of the background. Without nuisance parameters, the most discriminating test statistic between these hypotheses is given by the likelihood ratio: $L(\mu=1)/L(\mu=0)$. However, since in the context of this work we do not expect to reject H_0 (i.e. observe a signal), we will instead set an *upper limit* on the signal strength. Furthermore, we need to incorporate the nuisance parameters into the construction of the confidence interval [232–235]. We thus consider the *profile log-likelihood ratio modified for upper limits*:

$$q_\mu = \begin{cases} -2\log\dfrac{L(\mu,\hat{\hat{\boldsymbol{\alpha}}}(\mu))}{L(0,\hat{\hat{\boldsymbol{\alpha}}}(0))} & \hat{\mu}<0 \\ -2\log\dfrac{L(\mu,\hat{\hat{\boldsymbol{\alpha}}}(\mu))}{L(\hat{\mu},\hat{\boldsymbol{\alpha}})} & 0\le\hat{\mu}\le\mu \\ 0 & \hat{\mu}>\mu, \end{cases} \tag{2.26.}$$

where $\hat{\mu}$ denotes the maximum-likelihood estimate (MLE) of parameter μ, based on (2.25), while $\hat{\hat{\boldsymbol{\alpha}}}(\mu)$ denotes the conditional MLE of parameter $\boldsymbol{\alpha}$ under the constraint of μ (profiled value of $\boldsymbol{\alpha}$).

If q_μ follows the pdf $f(q_\mu|\mu')$, which depends on the tested and assumed values μ and μ', and $q_{\mu,\text{obs}}$ is the observed statistic, we can compute the p-values:

$$p_\mu(q_{\mu,\text{obs}}) = \int_{q_{\mu,\text{obs}}}^{\infty} f(q_\mu|\mu)\,\mathrm{d}q_\mu \equiv \text{CL}_{\text{s+b}} \tag{2.27.}$$

$$p_0(q_{\mu,\text{obs}}) = \int_{q_{\mu,\text{obs}}}^{\infty} f(q_\mu|\mu=0)\,\mathrm{d}q_\mu \equiv \text{CL}_{\text{b}} \tag{2.28.}$$

The *upper limit* μ_{up} at 95% confidence level (CL) is the largest value of μ for which $\text{CL}_{\text{s+b}} \ge 0.05$. Given the definition in (2.26), the property $\mu_{\text{up}} \ge 0$ is then satisfied. Working in a frequentist setting we are allowed to be wrong 5% of the time, however we would like to avoid excluding a signal to which we are not sensitive just because the background happens to under-fluctuate. This can be addressed by using the CL_{s} criterion, defined in Refs. [236, 237], in which the upper limit is obtained by solving:

$$\text{CL}_{\text{s}} \equiv p'_{\mu_{\text{up}}} = \frac{p_{\mu_{\text{up}}}}{p_0} = \frac{\text{CL}_{\text{s+b}}}{\text{CL}_{\text{b}}} = 0.05, \tag{2.29.}$$

as depicted on Fig. 2.19. Note that the resulting confidence interval has coverage larger than nominal: $P(\mu_{\text{up}} \ge \mu|\mu) > 95\%$.

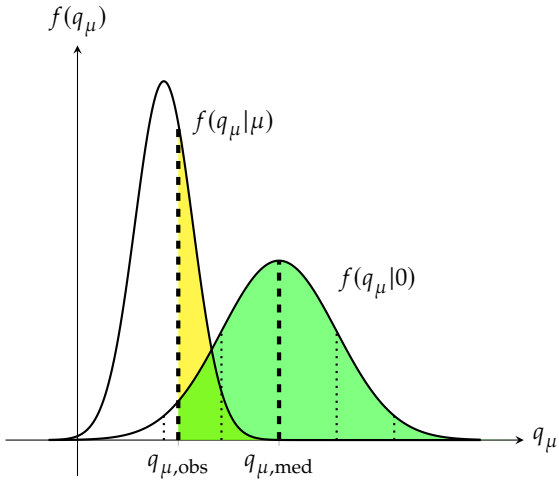

Figure 2.19. Schematic depiction of the CL_s criterion for upper limits. The 95% CL upper limit is obtained by finding μ s.t. the ratio between the yellow and green areas equals 5%. Similarly, replacing $q_{\mu,\text{obs}}$ by $q_{\mu,\text{med}}$ and repeating the procedure yields the expected upper limit.

It is also useful to quantify the *expected sensitivity* by computing the median upper limit μ_{exp} under the assumption that $\mu = 0$, obtained by solving $p_{\mu_{\text{exp}}}(q_{\mu,\text{med}}) = 0.05$, where $q_{\mu,\text{med}}$ is the median test statistic under the background-only hypothesis $\mu = 0$:

$$\int_{q_{\mu,\text{med}}}^{\infty} f(q_\mu|0)\,\mathrm{d}q_\mu = 0.5 \tag{2.30.}$$

Similarly, "1- and 2-sigma bands" around the expected limit can be defined by replacing the p-value of 50% in (2.30) by $(50 \pm 16)\%$ and $(50 \pm 22.5)\%$, respectively. If no signal is present in the data, or if we are not sensitive to it, the observed limit should be compatible with the expected one: there is a 95% probability that it lies within the "2-sigma band" as previously defined.

The procedures detailed above require the knowledge of the pdfs $f(q_\mu|\mu)$ and $f(q_\mu|0)$, for every value of μ. They can be obtained by generating ensembles of pseudo-experiments ("toy Monte Carlo"). For each toy, the number of events n_c and the auxiliary measurements a_p are generated according to the model (2.25) (using $\boldsymbol{\alpha} = \hat{\hat{\boldsymbol{\alpha}}}(\mu)$), and the test statistic is evaluated [232]. With sufficiently large samples of toys, the integrals in (2.27) and (2.28) can be evaluated and the upper limit computed.

Obtaining limits using toys can be computationally intensive; fortunately the chosen test statistic features useful asymptotic properties [238]. Wilk's [239] and Wald's [240] theorems provide expressions for the pdfs $f(q_\mu|\mu')$ in the limit of large data samples. In practice, these approximations work well even for samples of a few tens of events. Using these results, the CL_s upper limit at 95% CL is the solution to

$$\frac{1 - \Phi\left(\sqrt{q_\mu}\right)}{\Phi\left(\sqrt{q_{\mu,A}} - \sqrt{q_\mu}\right)} = 0.05, \tag{2.31.}$$

where Φ is the cumulative distribution function (cdf) of the Normal distribution, and $q_{\mu,A}$ is the test statistic evaluated on a special toy dataset, the "Asimov" toy, obtained by fixing the observed yields to the expected values in the background-only hypothesis: $n_c \equiv \nu_c(0, \boldsymbol{\alpha} = \hat{\hat{\boldsymbol{\alpha}}}(0))$. The expected N-sigma and median ($N = 0$) limits are obtained by solving the following for μ:

$$q_{\mu,A} = \left(\Phi^{-1}\left(1 - 0.05 \cdot \Phi(\pm N)\right) \pm N\right)^2 \tag{2.32.}$$

The results presented in this work have been obtained using the `HiggsCombine` software, based on the `RooFit` [241] and `RooStats` [242] frameworks and relying on the `Minuit` toolkit [243] for the minimisation of the likelihood.

Chapter

3. Search for Higgs boson pair production in the $b\bar{b}\ell\nu\ell\nu$ final state

In this chapter, we report on a search for the resonant and nonresonant production of Higgs boson pairs, decaying to a final state consisting of a pair of bottom quarks, a pair of charged leptons (muons or electrons), and neutrinos ($b\bar{b}\ell\nu\ell\nu$). This analysis was conducted using 35.9 fb^{-1} of data collected at $\sqrt{s} = 13$ TeV by CMS during 2016, and was subject to a publication in Ref. [147].

The motivations for probing resonant and nonresonant double Higgs production have been presented in Chap. 1. We search for the production of narrow-width spin-0 and spin-2 resonances X decaying to HH, with masses m_X between 260 and 900 GeV. For what regards the nonresonant production of Higgs boson pairs, in addition to the HH process in the SM we explicitly search for deviations from the SM. We follow the parameterisation introduced in Sec. 1.5, where $\kappa_\lambda = \lambda_{HHH}/\lambda_{SM}$ and $\kappa_t = y_t/y_{t,SM}$. For the most part, the methods and strategies used in the searches for resonant and nonresonant production are common. We thus describe both searches in a single body, and highlight the key differences between them where necessary.

In the first section of this chapter we describe the signal and background processes in the chosen event topology, as well as the event selection strategy. In the following, the data-driven estimation method of one of the background contaminations is detailed. Next, we describe the methods used to enhance the sensitivity to the signals. We then summarise the estimation of systematic uncertainties affecting this analysis, and finally we present the results of these searches.

3.1. Analysis setup and event selection

Since the SM Higgs boson can decay to a variety of final states, the production of Higgs boson pairs is fragmented into an even larger number of channels. The relative rate of each channel is governed by the branching ratios of the Higgs boson, which can be found in Refs. [72, 244]. The final state we are considering, $b\bar{b}\ell\nu\ell\nu$, was selected the following reasons:

1. The first and second most frequent decays of the Higgs boson are $H \to b\bar{b}$ and $H \to WW^*$. Through the $W \to \ell\nu$ decay, this renders the resulting rate in the chosen channel sizable (see discussion below).
2. The main background process populating this final state, top quark pair production, can be reasonably well modelled.
3. This channel was found to be complementary to other channels already considered by the CMS experiment in a prospective sensitivity study for the HL-LHC [153]. In

addition, the discovery of HH production in the SM is expected to be extremely challenging and will require the combination of as many channel as possible.

4. Should a deviation from SM predictions for HH production be observed in another channel, it would be crucial to provide several independent confirmations.

The branching ratio for $\mathrm{H}(\to X)\mathrm{H}(\to Y)$ is given by $\mathcal{B}(\mathrm{H}\to X)^2$ if X and Y represent the same final state, and by $2\cdot\mathcal{B}(\mathrm{H}\to X)\cdot\mathcal{B}(\mathrm{H}\to Y)$ otherwise. For $m_{\mathrm{H}} = 125.0\,\mathrm{GeV}$ we have $\mathcal{B}(\mathrm{H}\to\mathrm{b\bar{b}}) = 58.24(7)\%$ and $\mathcal{B}(\mathrm{H}\to\ell\nu\ell\nu) = 1.055(2)\%$ (for $\ell = \mu$,e and any neutrino flavour). The leptonic Higgs boson decay happens through diagrams involving both $\mathrm{H}\to\mathrm{WW}^*$ and $\mathrm{H}\to\mathrm{ZZ}^*$, whose interference has to be taken into account when computing the branching ratio to a specific final state. Nevertheless, the interference is small and the former amplitude (WW^*) dominates the latter (ZZ^*) by an order of magnitude. With these numbers, we obtain the total branching ratio to our final state as $\mathcal{B} = 1.223(5)\%$. Table 3.1 shows the branching ratios of channels that have been probed or may be probed in the future.

Table 3.1. Branching ratios to a given final state of Higgs pair production, for the main experimental channels. ℓ denotes both e and μ; ν denotes all three neutrino flavours; q stands for all quarks with exception of the top quark. The branching ratios are obtained from Refs. [72, 244]. The expected yields are given indicatively for the SM hypothesis assuming the cross section (1.44) quoted in Sec. 1.3.

Final state	Branching ratio	Expected yields (SM, $35.9\,\mathrm{fb}^{-1}$)
All	100 %	1200
$\mathrm{b\bar{b}b\bar{b}}$	33.9 %	407
$\mathrm{b\bar{b}}\tau\tau$	7.31 %	87.8
$\mathrm{b\bar{b}}\ell\nu q\bar{q}$	7.30 %	87.7
$\mathrm{b\bar{b}}\ell\nu\ell\nu$ (this work)	1.22 %	14.7
$\tau\tau\tau\tau$	0.393 %	4.72
$\mathrm{b\bar{b}}\ell\ell q\bar{q}$	0.285 %	3.42
$\mathrm{b\bar{b}}\gamma\gamma$	0.264 %	3.17

While the hadronically decaying Higgs boson in the signals can be reconstructed using the observed jets resulting from the hadronisation of the b quarks in $\mathrm{H}\to\mathrm{b\bar{b}}$, there is no way to fully reconstruct the kinematics of the leptonically decaying Higgs boson due to the presence of two unobserved neutrinos in the final state. This prevents us from accessing the di-Higgs invariant mass, m_{HH}, and justifies the absence of significant differences in the analysis strategies for resonant and nonresonant production. The salient features of the signals, common to all signal hypotheses, are on the one hand a pair of b jets with invariant mass peaking near that of the Higgs boson, standing out over a smooth background distribution, and on the other hand a pair of prompt, isolated charged leptons. The pair of leptons provides us with a clean signature to trigger the collection of events in data, as described in Sec. 2.3.9. Noteworthy properties of these leptons stem from the spin-0 nature of the Higgs boson and the $\mathrm{H}\to\mathrm{VV}^*\to\ell\nu\ell\nu$ decay

chain [245, 246]. Indeed, the charged leptons mostly have low angular separation and thus low invariant mass, with the distribution of $m_{\ell\ell}$ peaking around 30 GeV. In addition, for resonant production with $m_X \lesssim 600$ GeV as well as for nonresonant production, the Higgs bosons are produced nearly at rest. Due to the 4-body leptonic decay of the Higgs boson, this yields leptons with relatively low p_T: for the SM hypothesis, the distributions of the p_T-leading and subleading leptons only peak around 50 and 20 GeV, respectively. This represents a challenge for triggering on signal events, and justifies the choice of low-threshold dilepton trigger paths over the alternative of higher-threshold single-lepton paths.

The most frequent SM processes contributing to the considered event topology are top quark pair production ($t\bar{t}$) and $Z/\gamma^*(\to \ell^+\ell^-)$ plus jets associated production (Drell–Yan process). The total cross section of the $t\bar{t}$ process is known to NNLO in QCD and amounts to 830 pb [247]. Taking into account the branching fraction $\mathcal{B}(t \to \ell\nu b) = 10.9\%$ (for each of $\ell = e, \mu$) for both top quarks [5], this translates into a cross section of 39.4 pb, corresponding to about 1.4 million events in the signal topology. Since $t\bar{t}$ production contributes to the exact same final state as the signals, it is said to be an *irreducible* background. This implies that the only possibility to reduce its rate is to apply clever requirements on the kinematics of the reconstructed particles, so that the sensitivity to the signals can be enhanced. Furthermore, we can only rely on the simulation to compute its contribution due to the lack of clear resonant signatures in the signal. On the other hand, the Drell–Yan (DY) process is *reducible*, as it can be cut back by requiring the presence of b-tagged jets. Unfortunately, this requirement has no effect on the contamination from Z/γ^* plus b jets associated production, for which other methods will have to be employed. Further minor backgrounds include single top quark and W boson associated production (tW), single top quark production in the t and s channels, diboson production (ZZ, WW, ZW, denoted VV in the following), $t\bar{t}$ and vector boson associated production ($t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}\gamma$, denoted $t\bar{t}V$), and various single Higgs boson production processes (chiefly $t\bar{t}H$ and ZH). Experimental backgrounds due to jets misidentified as leptons, from W plus jets or QCD multijet production, have a negligible impact on the analysis thanks to the stringent requirements on the quality of the reconstructed electrons and muons.

3.1.1. Samples

The data samples used for this analysis, collected at $\sqrt{s} = 13$ TeV during 2016, were described in Sec. 2.3.9. Only the luminosity sections (LSs) (see Sec. 2.2.5) of data certified as sufficiently good to be used for analysis were considered, yielding a dataset corresponding to an integrated luminosity of 35.9 fb^{-1}.

The background simulation samples have been generated at NLO in QCD using Powheg 2 [44,45,55,248,249] and MadGraph5_aMC@NLO versions 2.2.2.0 and 2.3.2.2 [54]. MadSpin [39] was used to model the decay of heavy resonances, and the matching and merging of different parton multiplicities for samples generated by MG5_aMC@NLO was achieved in the FxFx scheme [51] (see Sec. 1.2). For all samples, Pythia version 8.212 [41, 42] with the CUETP8M1 tune [250] has been used for simulation of parton showering, hadronisation and underlying event. The modelling of pileup as well as the simulation of the CMS detector have already been described in Sec. 2.3.11.

The main background processes and the generators used to model them are listed in Tab. 3.2 along with the cross sections used to normalise their respective contributions. For each process, the most precise theoretical cross section is used. The cross section of the main $\mathrm{t\bar{t}}$ background was obtained at NNLO+NNLL precision in QCD [247]; the DY process is normalised to NNLO in QCD and NLO electroweak precision [251]. For single top quark production in the tW channel, an approximate NNLO QCD computation is used [252]. The WW samples are normalised to NNLO precision in QCD [253]; further diboson, as well as t- and s-channel single top quark, $\mathrm{t\bar{t}H}$ and $\mathrm{t\bar{t}V}$ processes are normalised to NLO precision in QCD [54, 254]. The cross sections for remaining single Higgs boson production processes are computed at NNLO in QCD and NLO in electroweak corrections [72].

Table 3.2. Parton-level generators used to model the major backgrounds entering the event selection, and cross section values used to normalise their contributions. A top quark mass of $m_{\mathrm{t}} = 172.5\,\mathrm{GeV}$ is used for both the event generation and cross section computation. For the DY and tW processes, the generation is restricted to final states containing any charged leptons (e, μ, τ), and the cross sections are rescaled using the relevant branching ratios. A restriction on the invariant mass of lepton pairs, $m_{\ell\ell} > 10\,\mathrm{GeV}$, is imposed for the DY process, and quarkonia resonances are not modelled in this sample. Both charge conjugates of the tW process have the same cross section. Uncertainties in the cross sections are quoted separately for what concerns the renormalisation and factorisation scale uncertainties (first figure), and the value of α_s and the PDFs (second figure).

Process	Generator	Cross section (pb)
$\mathrm{t\bar{t}}$ (inclusive)	Powheg 2	$831.8^{+19.8}_{-29.2} \pm 35.1$ [247]
$\mathrm{Z}/\gamma^* (\to \ell^+\ell^-)$ + jets (DY)	MG5_aMC@NLO	$24640^{+250}_{-190} \pm 1280$ [251]
$\mathrm{tW^-}$, $\mathrm{\bar{t}W^+}$ $(\to \geq 1\ell)$	Powheg 2	$39.1 \pm 1.8 \pm 3.4$ [252]

For the modelling of the signal processes in the search for resonant enhancements of Higgs pair production, we have considered a model of warped extra dimension (WED). With this model, radions and KK-gravitons are used as benchmarks for generic spin-0 and spin-2 narrow-width resonances produced in gluon fusion. As mentioned in Sec. 1.6, the results obtained in the spin-0 case (radions) can be interpreted in a model-independent way, while strictly speaking the results pertaining to the spin-2 benchmarks depend on chosen production model. These signal samples have been generated at LO in QCD using MG5_aMC@NLO version 2.2.2.0. In order to account for the dependence of the analysis acceptance and event kinematics on the hypothesised resonance mass m_X, 13 samples with different m_X have been produced, for each of the the spin-0 and spin-2 cases. The considered values for m_X are 260 GeV, 270 GeV, 300 to 700 GeV in steps of 50 GeV, and 800 and 900 GeV. The low end of this range corresponds to the kinematic threshold for the decay of narrow-width resonances to Higgs boson pairs, while above $m_X = 900\,\mathrm{GeV}$, the products of the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay can not be reconstructed efficiently as two separate jets with $R = 0.4$. Alternate ("boosted") analysis techniques have then to be employed; this possibility is left for future work.

In the case of nonresonant double Higgs production, 14 different samples have been generated with MG5_aMC@NLO versions 2.2.2.0 at LO in QCD with exact top quark mass dependence. These correspond to the 12 benchmarks resulting from the clustering procedure laid out in Sec. 1.5, as well as one sample in the SM hypothesis, and one

sample generated using only "box" diagrams (see Fig. 1.2, right), i.e. assuming $\kappa_t = 1$ and $\kappa_\lambda = 0$. While we would like to probe the effects of anomalous values of κ_λ and κ_t for any value of these couplings, the available samples can not directly be used to that end. In order to obtain predictions for arbitrary points in the $(\kappa_\lambda, \kappa_t)$ plane, we have implemented a matrix-element based event reweighting, as described in Sec. 1.2 and 1.5. The matrix element expression used to that end was generated from MG5_aMC@NLO in C++ format. The reweighting took into account the dynamic renormalisation scale used during the generation of the original samples, by accessing the event-dependent value of the strong coupling constant.

In the generation of both resonant and nonresonant signal samples, the Higgs boson decays were simulated by Pythia. However, the decay channels used in the sample generation do not strictly correspond to the final state described in Sec. 3.1, but also include the possibility in which one of the Higgs bosons decays to two τ leptons and two neutrinos. Since the τ can again decay to an electron or a muon, plus two additional neutrinos, this extra channel contributes to our final state and has to be taken into account, even if we do not specifically consider τ leptons in this analysis. The branching ratio used from now on for $HH \to b\bar{b}\ell\nu\ell\nu$, with $\ell = e, \mu, \tau$, is then $\mathcal{B} = 2.72\%$, and all results quoted for this final state should be understood to include all three lepton flavours.

3.1.2. Event selection

The event selection corresponds to a set of simple requirements applied on the content of reconstructed events in data and simulation, aimed at keeping as many signal events as possible while reducing the rate of the various backgrounds. The selection process proceeds in steps, so that the agreement between data and simulation can be assessed at multiple levels.

Data events are collected using the set of trigger paths described in Sec. 2.3.9. We recall that no emulation of the trigger is available in the simulation. We start by requiring the presence of two leptons (muons or electrons) of opposite charges, passing the identification and isolation criteria introduced in Sec. 2.3.3 and 2.3.4. The selected events are categorised based on the flavours of the leptons, into three different channels: $\mu^+\mu^-$ and e^+e^- ("same-flavour"), and $\mu^\pm e^\mp$ ("different-flavour"). If more than two leptons are present, we consider the pair with the highest scalar sum of p_T. We do not veto additional leptons, since we are not affected by large background processes yielding three or more leptons. The two leading leptons are required to have a p_T greater than 25 GeV and 15 GeV for ee events, 20 GeV and 10 GeV for $\mu\mu$ events, 25 GeV and 15 GeV for μe events, and 25 GeV and 10 GeV for $e\mu$ events, for the high- and low-p_T lepton, respectively. These thresholds are tuned to the thresholds of the HLT trigger paths in each category. For electrons (muons), the pseudo-rapidity range $|\eta| < 2.5$ ($|\eta| < 2.4$) is considered, matching the geometrical acceptance of the inner (outer) tracker. A dilepton mass requirement of $m_{\ell\ell} > 12$ GeV is applied in all categories in order to suppress backgrounds from quarkonia resonances and jets misidentified as leptons.

In data, the selected leptons are required to correspond to the leptons, reconstructed at the HLT, which triggered the recording of the events. The *matching* between a lepton reconstructed offline, O, and online (HLT) lepton H, is achieved by asking that:

$$\Delta R(O,H) < 0.1, \tag{3.1.}$$

$$\frac{\left|p_{\mathrm{T}}(O) - p_{\mathrm{T}}(H)\right|}{p_{\mathrm{T}}(O)} < 0.5. \tag{3.2.}$$

This trigger matching procedure ensures that the trigger efficiencies computed with the T&P method in Sec. 2.3.11 can be applied to the event selection.

We consider jets as defined in Sec. 2.3.5, which have a p_{T} greater than 20 GeV, lie in the pseudorapidity range $|\eta| < 2.4$, and are separated from selected leptons by $\Delta R > 0.3$. At least two such jets have to be present. If more than two jets are available, we choose the pair with the largest scores of the cMVAv2 b tagging discriminator. This pair of jets defines the hadronically decaying Higgs boson candidate. Among different possible jet pairing techniques, this method was found to be efficient in selecting the jets coming from the H decay, without creating an artificial peak around m_{H} in the distribution of the dijet invariant mass (m_{jj}) in the $\mathrm{t\bar{t}}$ background [255].

In a further stage of the selection, we require the two selected jets to pass the medium working point of the cMVAv2 b tagging algorithm. Finally, we ask that $m_{\ell\ell} < 76$ GeV, which has the effect of removing the majority of the DY background, for which $m_{\ell\ell}$ peaks around $m_{\mathrm{Z}} \approx 91$ GeV. It also efficiently suppresses the $\mathrm{t\bar{t}}$ background, at the only cost of removing the minuscule part of the signal where the H boson decays to ZZ^*, and where the on-shell Z boson decays to charged leptons. The set of all previous requirements, summarised in Tab. 3.4, define the "signal region". The predicted yields for the various groups of backgrounds, for a few representative signals (normalised to a cross section of 5 pb), and the observed yields in data in the signal region are shown on Tab. 3.3 for the three flavour channels. A good agreement is observed between predictions and observations.

The selection efficiencies for the signals, for different stages of the event selection, are shown on Fig. 3.1 as a function of the signal parameters, i.e. $\kappa_\lambda/\kappa_{\mathrm{t}}$ and m_{X} in the nonresonant and resonant case, respectively. See App. A for plots of the selection efficiencies broken down into the different channels.

The predicted and observed distributions of the dilepton invariant mass ($m_{\ell\ell}$), before and after requiring the two selected jets to be b-tagged, are shown on Fig. 3.2. On those figures, as in many others shown throughout this chapter, the predictions for the various background processes, normalised to theoretical cross sections and to the measured integrated luminosity, are shown as stacked histograms. The shaded bands indicate the systematic uncertainties on the total background yield in each bin, including the uncertainty due to the finite amount of simulated events. The various sources of systematic uncertainties are combined in quadrature. Unless specified otherwise, the uncertainties shown are *pre-fit*, i.e. as they have been estimated and before being constrained through the profile construction described in Sec. 2.4.2. The data are shown as black dots, with vertical bars indicating the corresponding confidence interval at 68% CL on the yields in each bin. The bottom panels show the ratio between the observed and expected yields in each bin. Predictions for a few representative signal samples are shown as solid lines. For visualisation purposes, the signals are normalised to an

Table 3.3. Predicted yields in the signal region with two leptons, two b-tagged jets, and $12 < m_{\ell\ell} < 76$ GeV. The uncertainties quoted for the different background processes correspond to total uncertainties. For the total background yields, uncertainties are broken down into a term due to the limited amount of simulated events, and the total systematic uncertainty. The predictions for the Drell–Yan process in the e^+e^- and $\mu^+\mu^-$ channels are obtained using the method described in Sec. 3.2; all other predictions are obtained using the simulation, normalised to the most precise available theoretical cross sections.

	e^+e^-	$\mu^\pm e^\mp$	$\mu^+\mu^-$
Signals (norm. to 5 pb)			
$(\kappa_\lambda, \kappa_t) = (-20, 0.5)$	815.6 ± 220.4	2509.6 ± 674.0	2287.0 ± 612.3
$(\kappa_\lambda, \kappa_t) = (5, 2.5)$	1109.1 ± 317.3	3257.3 ± 928.6	2874.6 ± 818.9
SM: $(\kappa_\lambda, \kappa_t) = (1, 1)$	1068.3 ± 302.0	3179.3 ± 895.3	2788.6 ± 784.8
$m_X = 400$ GeV (spin 0)	924.3 ± 63.3	2856.3 ± 176.1	2559.3 ± 152.9
$m_X = 900$ GeV (spin 0)	1444.8 ± 163.7	4032.7 ± 446.2	3908.6 ± 427.4
Backgrounds			
$t\bar{t}$	7696.3 ± 1074.1	24918.5 ± 3398.8	21829.5 ± 2959.4
Drell–Yan	565.1 ± 32.1	167.7 ± 56.5	2389.5 ± 134.9
Single top (tW, t- & s-chan.)	226.2 ± 11.8	700.6 ± 30.3	608.9 ± 24.0
$t\bar{t}V$	21.7 ± 4.2	58.2 ± 9.3	63.4 ± 9.6
$t\bar{t}H$	12.1 ± 1.3	38.0 ± 3.9	33.8 ± 3.4
VV	12.9 ± 1.7	13.9 ± 2.3	43.9 ± 4.6
Other single Higgs	4.9 ± 0.6	6.5 ± 1.3	14.6 ± 1.5
Total ± (stat.) ± (syst.)	8539.2 ± 34.4 ± 1087.7	25903.5 ± 65.0 ± 3473.5	24983.6 ± 59.7 ± 2998.4
Data	8597	26746	25880
Data / prediction	1.01 ± 0.13	1.03 ± 0.14	1.04 ± 0.12

Table 3.4. Summary of object definitions and selection requirements described in Sec. 3.1.2. The identification (ID) algorithms are detailed in Sec. 2.3.3 for muons, and Sec. 2.3.4 and 2.3.9 for electrons. Jet reconstruction and b tagging are covered in Sec. 2.3.5 and 2.3.7, respectively. "μe events" refers to events in the $\mu^\pm e^\mp$ channel where the leading lepton is a muon. The leptons are ordered by their p_T, whereas jets are ordered by their score of the cMVAv2 b tagging algorithm.

Object	Definition	Selection	
Lead. (sub-lead.) e	Medium + HLT safe ID	$p_T > 25(15)$ GeV, $\lvert\eta\rvert < 2.5$	$12 < m_{\ell\ell} < 76$ GeV
Lead. (sub-lead.) μ	Tight ID	$p_T > 20(10)$ GeV, $\lvert\eta\rvert < 2.4$	
	Rel. PF iso. < 0.15	$p_T > 25$ GeV for μe events	
Jets	PF, anti-k_T $R = 0.4$	≥ 2 jets: $p_T > 20$ GeV, $\lvert\eta\rvert < 2.4$, $\Delta R(j, \ell) > 0.3$	
b tagging	cMVAv2 medium WP	≥ 2 b-tagged jets	

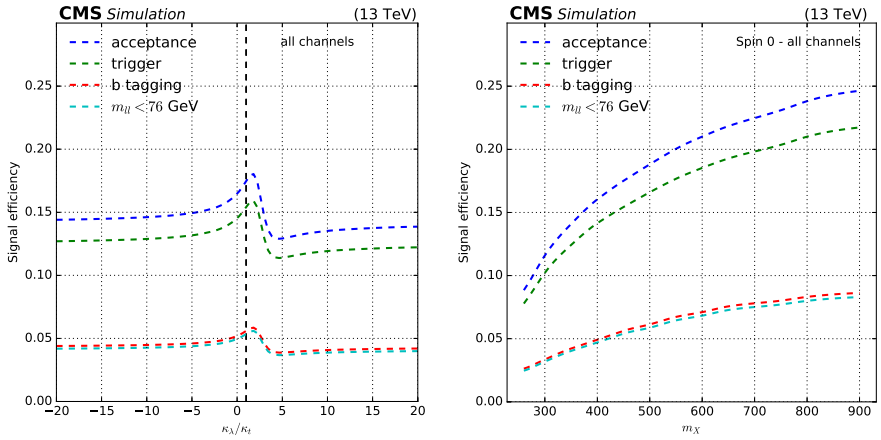

Figure 3.1. Selection efficiency of signal events, for the different steps of the selection shown on Tab. 3.4. Left: nonresonant production, shown as a function of κ_λ/κ_t. Right: resonant production (spin-0 case), shown as a function of m_X. The spin-2 case is available in App. A. All efficiencies are given with respect to the total signal samples. The "acceptance" step corresponds to the requirements on lepton identification, isolation, p_T and η, and jet p_T, η and angular separation from the leptons. The efficiency of the trigger is computed after applying the previous requirements. Note that almost 30% of signal events contain at least one hadronically decaying τ lepton in the final state, which we do not reconstruct.

arbitrary cross section times branching ratio of 5 pb. This corresponds to a total HH production cross section of 184 pb, several orders of magnitude larger than the final sensitivity.

3.2. Estimation of the Drell–Yan background

In the e^+e^- and $\mu^+\mu^-$ channels, the amount of simulated events available to model the DY background is not sufficient to provide a reliable estimate of its contribution in the signal region, and a different background estimation method has thus been developed. In the $\mu^\pm e^\mp$ channel however, the DY background's only contribution is due to leptonically decaying pairs of τ leptons from $Z/\gamma^* \to \tau^+\tau^-$. Most of these leptons have low p_T and do not pass the event selection, hence the effect of DY in that channel is almost negligible (see Tab. 3.3) and can be estimated directly using the simulation.

The idea behind this estimation method is to harness the larger amounts of DY events in data and in simulation when b tagging criteria are relaxed ("untagged" events). The goal is to reweight DY events present in untagged data to provide a *data-driven* estimate of the DY background with two b-tagged jets. Since other minor backgrounds, such as $t\bar{t}$, are present in untagged data, the untagged events for these processes undergo the same reweighting. This unwanted contribution is then subtracted using the simulated samples for these backgrounds as follows:

$$\mathrm{Data}^{2j} = \mathrm{DY}^{2j} + \mathrm{t\bar{t}}^{2j} + \ldots \tag{3.3.}$$

$$\Rightarrow \mathrm{DY}^{2b}_{\mathrm{est.}} \equiv W_{\mathrm{sim.}} \times \mathrm{DY}^{2j} = W_{\mathrm{sim.}} \times \mathrm{Data}^{2j} - \left(W_{\mathrm{sim.}} \times \mathrm{t\bar{t}}^{2j}_{\mathrm{sim.}} + \ldots \right), \tag{3.4.}$$

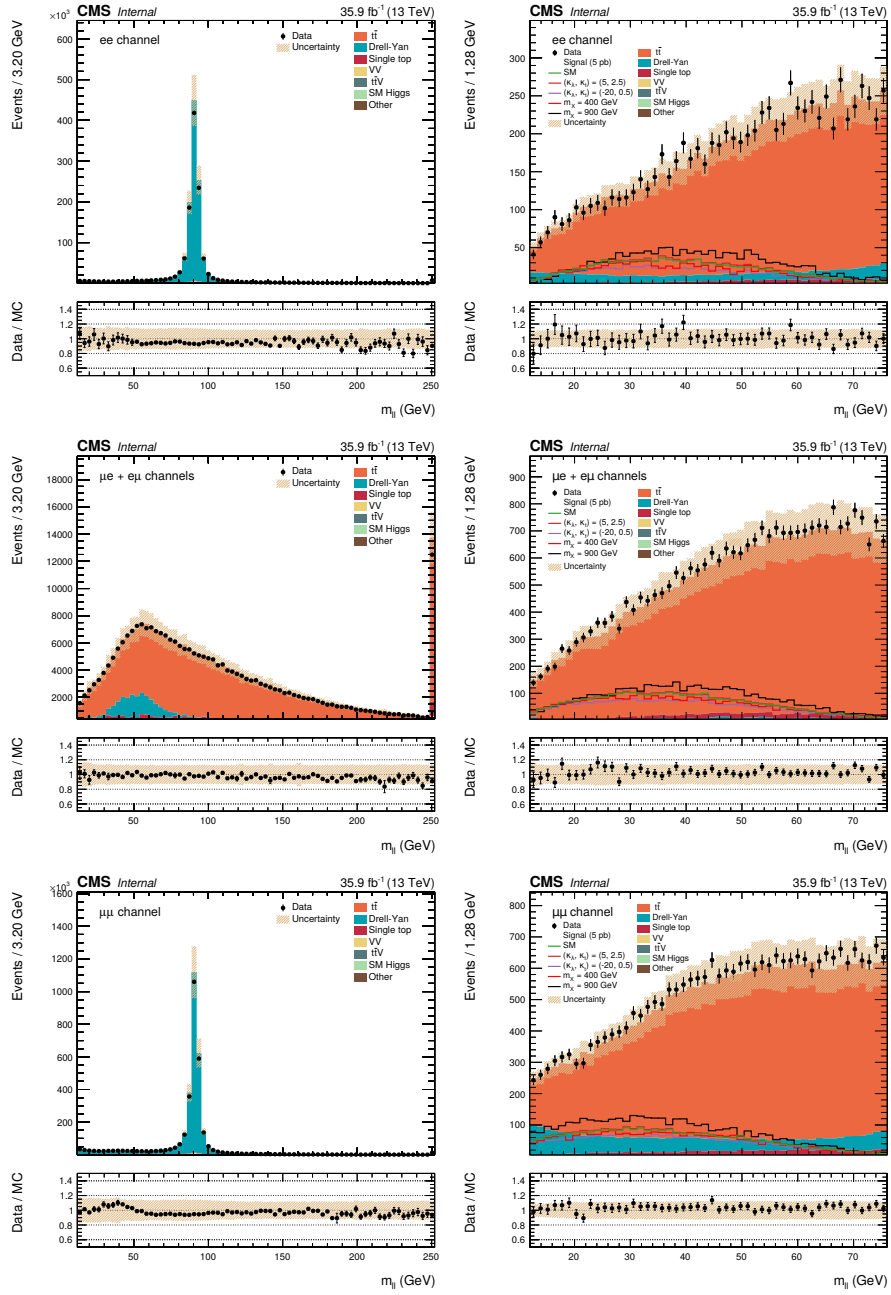

Figure 3.2. Distribution of the dilepton invariant mass, $m_{\ell\ell}$. The selection summarised in Tab. 3.4 is applied, with the exception of the b tagging requirements for the plots in the left column. The e^+e^-, $\mu^\pm e^\mp$ and $\mu^+\mu^-$ channels are shown on the top, middle and bottom, respectively. Shaded bands show pre-fit systematic uncertainties in the background predictions. The signals are normalised to an arbitrary cross section of 5 pb for visualisation purposes. All backgrounds are estimated using the simulation, except for the Drell–Yan process in the right column, in the e^+e^- and $\mu^+\mu^-$ channels, which is estimated according to the method described in Sec. 3.2.

where "2j" and "2b" denotes a sample before and after the two selected jets are required to be b-tagged, respectively, and $W_{\text{sim.}}$ denotes an event-dependent weight, modelled using the simulation as described below. Note that in the implementation of the statistical model for inference on the signal, we do not use the estimated b-tagged DY contribution directly, but consider the different terms of (3.4) (i.e. the reweighted data and non-DY background processes) separately.

Instead of actually requiring b-tagged jets in DY events, we can parameterise the effect of b tagging as a function of jet kinematics (p_{T} and $|\eta|$), separately for each jet, and weight events with the product of the b tagging efficiencies for the two selected jets in the event. However, these efficiencies chiefly depend on the true flavour of the jet, which is unknown in data. A solution is to compute a weighted average of the b tagging efficiencies, using the relative contributions F_{kl}, estimated using the simulation, of DY plus two jets of flavours k and l, where $k,l = \mathrm{b},\mathrm{c},$or light-flavour, to the DY plus two jets process. These contributions are not constant throughout the event phase space, which implies that modelling the effect of b tagging requires to parameterise these flavour fractions as a function of event kinematics. The expected fractions F_{kl} of jets with flavours k and l in the DY process are parameterised as a function of the output value of a boosted decision tree (BDT) (see Sec. 2.4.1), and estimated from the simulated DY samples. Their dependency on the BDT output value accounts for the different kinematical behaviours of heavy- or light-flavour associated DY processes, effectively reducing the dimensionality of the phase space to a single variable. The BDT is trained to discriminate DY+$\mathrm{b\bar{b}}$,$\mathrm{c\bar{c}}$ from other DY associated production processes using the following input variables: $p_{\mathrm{T}}^{\mathrm{j}_1}$, $p_{\mathrm{T}}^{\mathrm{j}_2}$, η^{j_1}, η^{j_2}, $p_{\mathrm{T}}^{\mathrm{jj}}$, $p_{\mathrm{T}}^{\ell\ell}$, $\eta^{\ell\ell}$, $\Delta\phi(\ell\ell,\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$ (defined as the $\Delta\phi$ between the dilepton system and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$), number of jets, and H_{T} defined as the scalar sum of the transverse momenta of all selected leptons and jets. To account for a residual dependence of the F_{kl} on $m_{\ell\ell}$, the fractions F_{kl} are computed separately in three $m_{\ell\ell}$ regions: $12 < m_{\ell\ell} < 76$ GeV, $76 \leq m_{\ell\ell} < 106$ GeV, and $m_{\ell\ell} \geq 106$ GeV. Figure 3.3 shows the BDT distribution in data, and the the fractions F_{kl} (for $(k,l) = (\mathrm{b},\mathrm{b})$ and (light,light)) are shown on Fig. 3.5. The event-dependent weights applied on untagged events in data are then given by:

$$W_{\text{sim}} = \sum_{k,l=\mathrm{b},\mathrm{c},\text{light-flavour}} F_{kl}(\mathrm{BDT})\,\epsilon_k(p_{\mathrm{T}}^{\mathrm{j}_1},\eta^{\mathrm{j}_1})\,\epsilon_l(p_{\mathrm{T}}^{\mathrm{j}_2},\eta^{\mathrm{j}_2}), \tag{3.5.}$$

where ϵ_k and ϵ_l are the b tagging efficiencies for k- and l-flavour jets calculated using the simulated DY samples as a function of p_{T} and η of the jets, and j_1 and j_2 denote the two selected jets, ordered as $p_{\mathrm{T}}^{\mathrm{j}_1} > p_{\mathrm{T}}^{\mathrm{j}_2}$. The indices k and l refer to the assumed flavour of j_1 and j_2, respectively. The b tagging efficiencies are corrected for differences between data and simulation using the scale factors introduced in Sec. 2.3.11. The computed values of ϵ_k (for $k = \mathrm{b}$ and light) are shown on Fig. 3.4.

Before applying b tagging requirements, data and simulation agree within systematic uncertainties. However, any residual difference between data and simulation regarding the normalisation of the various background processes subject to the subtraction in (3.4) might bias the resulting shape of the estimated DY contribution after b tagging. We thus rely on a binned likelihood fit to the distribution of $m_{\ell\ell}$ in the untagged sample, in a

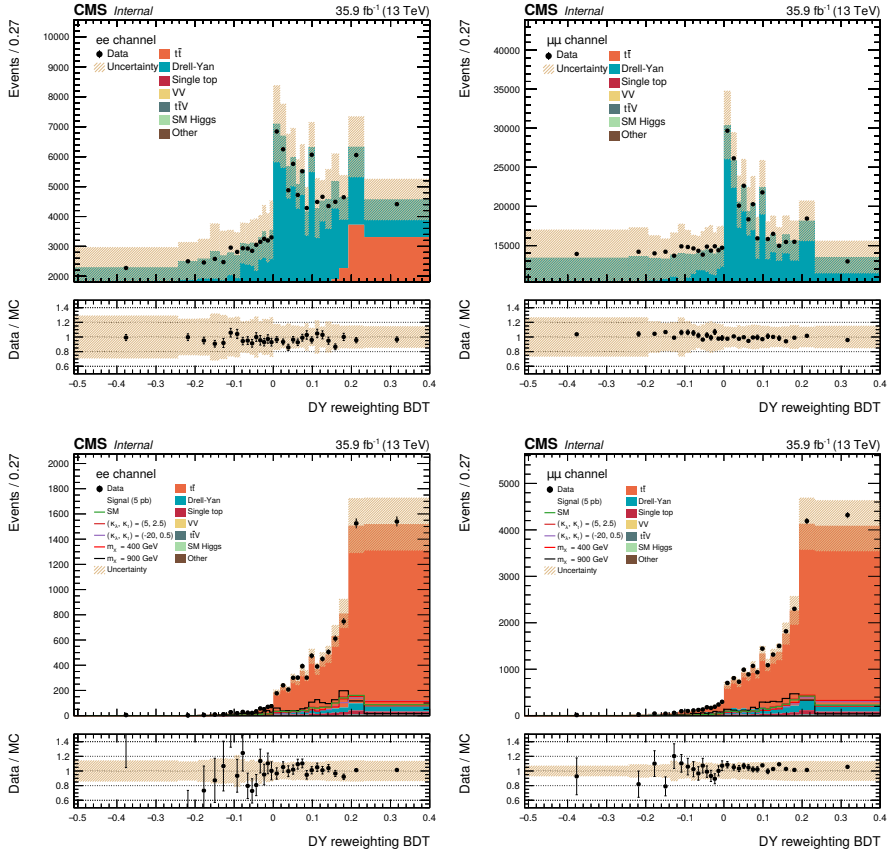

Figure 3.3. Distribution of the output score of the BDT used for the estimation of the DY background, with $12 < m_{\ell\ell} < 76$ GeV, before (top) and after (bottom) requiring the two selected jets to be b-tagged, in the e^+e^- (left) and $\mu^+\mu^-$ (right) channels. All processes are estimated using the simulation on the upper plots, whereas on the bottom, the DY process is estimated according to the method described in Sec. 3.2.

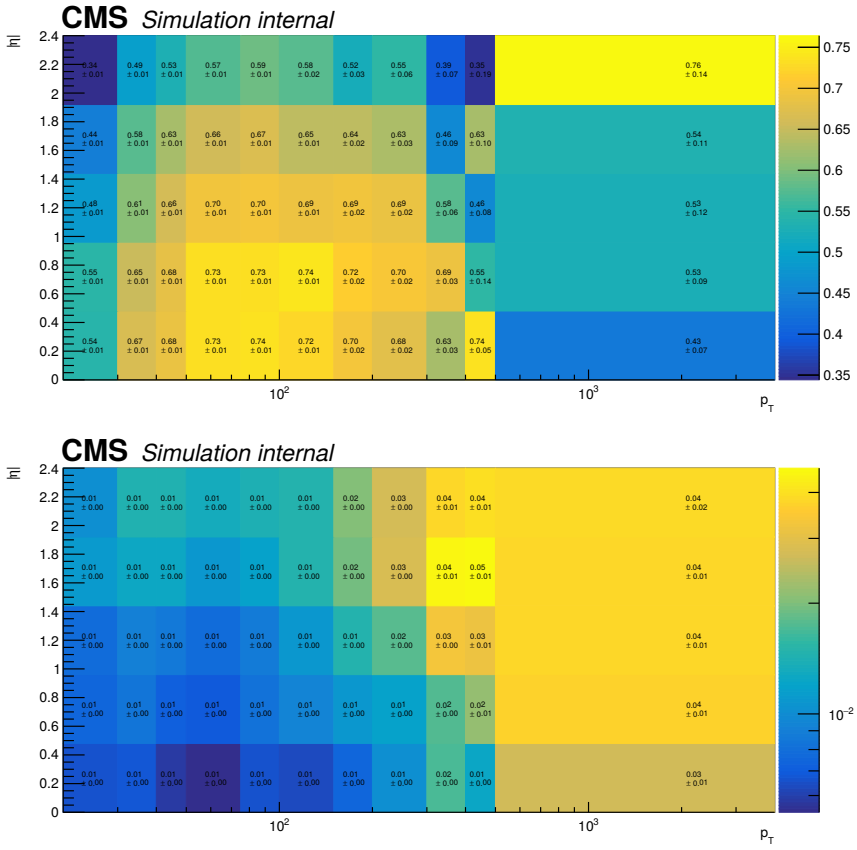

Figure 3.4. | b tagging efficiency for true b jets (top) and for light jets (bottom), parameterised as a function of reconstructed jet p_{T} and $|\eta|$, and computed using a simulated DY sample. Uncertainties shown are statistical only.

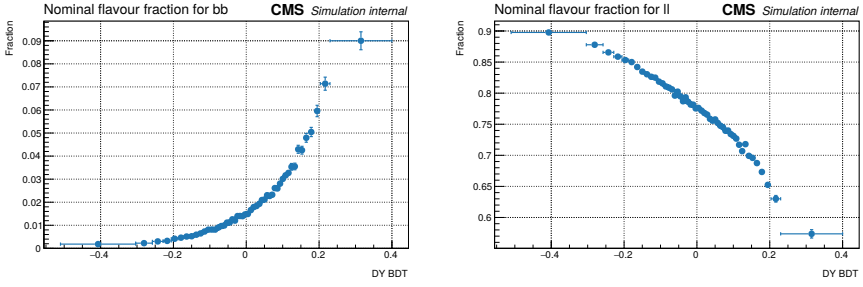

Figure 3.5. | Fraction of DY plus two b jets (left) and DY plus two light-flavour jets (right), as a function of the output score of the BDT used in the estimation of the DY background. No cut on $m_{\ell\ell}$ is applied here. Uncertainties shown are statistical only. The binning is chosen to ensure relatively uniform uncertainties across most of the range of the score.

control region defined by $m_{\ell\ell} \geq 76$ GeV, to derive a corrective factor, denoted S_1 in (3.6), for the normalisation of the total contribution of these processes. The normalisation of the DY process in the untagged sample is left free to float in the fit. After requiring b tagging, we observe a small disagreement in the overall normalisation of the estimated DY background. Hence, we again fit the distribution of $m_{\ell\ell}$, with $m_{\ell\ell} \geq 76$ GeV, to derive a second correction factor for the normalisation of the prediction for the b-tagged DY process (S_2 in (3.6)). We can then rewrite (3.4) as:

$$\mathrm{DY}^{2\mathrm{b}}_{\mathrm{est.}} \equiv S_2 \left(W_{\mathrm{sim.}} \times \mathrm{Data}^{2\mathrm{j}} - S_1 \left(W_{\mathrm{sim.}} \times \mathrm{t\bar{t}}^{2\mathrm{j}}_{\mathrm{sim.}} + \ldots \right) \right) \tag{3.6.}$$

The correction factors S_1 and S_2 are derived separately in the $\mathrm{e}^+\mathrm{e}^-$ and $\mu^+\mu^-$ channels and are given in Tab. 3.5.

Table 3.5. | Scale factors, as defined in (3.6), needed to correct the data-driven prediction of the DY background.

Factor	$\mathrm{e}^+\mathrm{e}^-$	$\mu^+\mu^-$
S_1	94 %	97 %
S_2	83 %	88 %

The estimation method is validated both in the simulation and in a DY-dominated data control region defined by requiring all selection criteria summarised in Tab. 3.4, but with $76 \leq m_{\ell\ell} < 106$ GeV. In the simulation, we compare the normalised distributions of various kinematic variables, after b tagging but inclusively in $m_{\ell\ell}$, between simulated DY events and the result of the method described above, as shown on Fig. 3.6. The agreement is overall satisfactory, given the large statistical uncertainty inherent in the simulated DY sample when applying b tagging, even for distributions showing important kinematic differences between the untagged and b-tagged samples. Figure 3.7 shows the comparison between data and predictions for the same quantities as in Fig. 3.6, in the DY-dominated data control region. While the agreement is not perfect, it was deemed sufficient for the purpose of estimating a process representing about 5% of the overall background in the signal region.

3.3. Parameterised discriminators for signal extraction

As explained in Sec. 3.1, it is impossible to fully reconstruct the Higgs boson momenta in the signal. Furthermore, there is no single variable that provides satisfying discrimination between signals and backgrounds. While the dijet invariant mass in the signal features a resonant peak close to the true mass of the Higgs boson, standing out over the smooth background, the poor energy resolution on the jets limits the sensitivity than can be attained using this variable only. We thus follow a multivariate approach to harness the information available in other quantities relative to the leptonically decaying Higgs boson. The method used here is an evolution of the strategy followed by two preliminary analyses in the same final state, Refs. [256, 257]. We consider the same variables as those that were previously identified since they showed both a good discrimination power and a good agreement between data and simulation. In addition, the multivariate classifiers

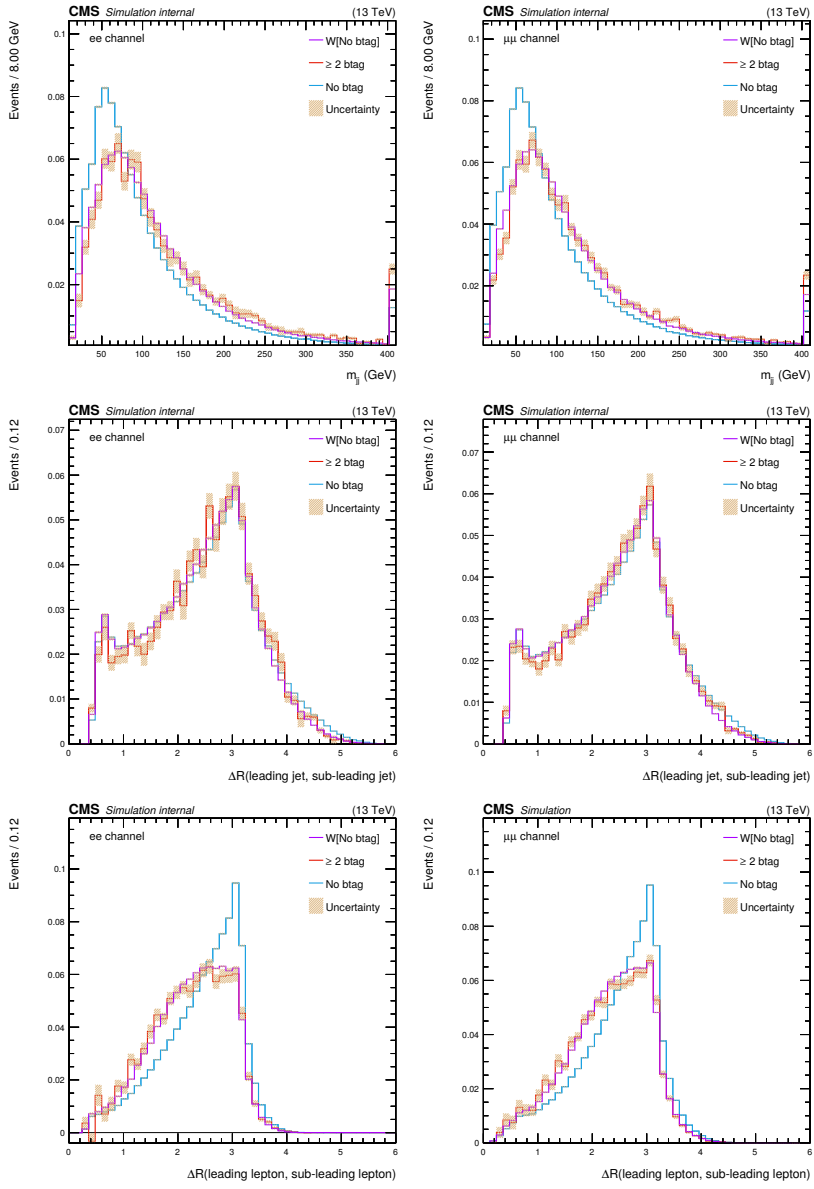

Figure 3.6. Validation of the DY estimation method using the simulation, in the $\mathrm{e^+e^-}$ (left) and $\mu^+\mu^-$ (right) channels and inclusively in $m_{\ell\ell}$. The invariant mass and ΔR separation between the jets are shown on the first and second row, respectively; the third row shows the ΔR between the leptons. The light blue curve corresponds to the untagged DY plus two jets process, i.e. without any b tagging applied. The red curve is obtained from events with two b-tagged jets, whereas the purple curve is obtained by reweighting the untagged events using the weights defined in (3.5). The two latter distributions show satisfactory agreement. The shaded bands only show the statistical uncertainties due to the limited number of events in each bin.

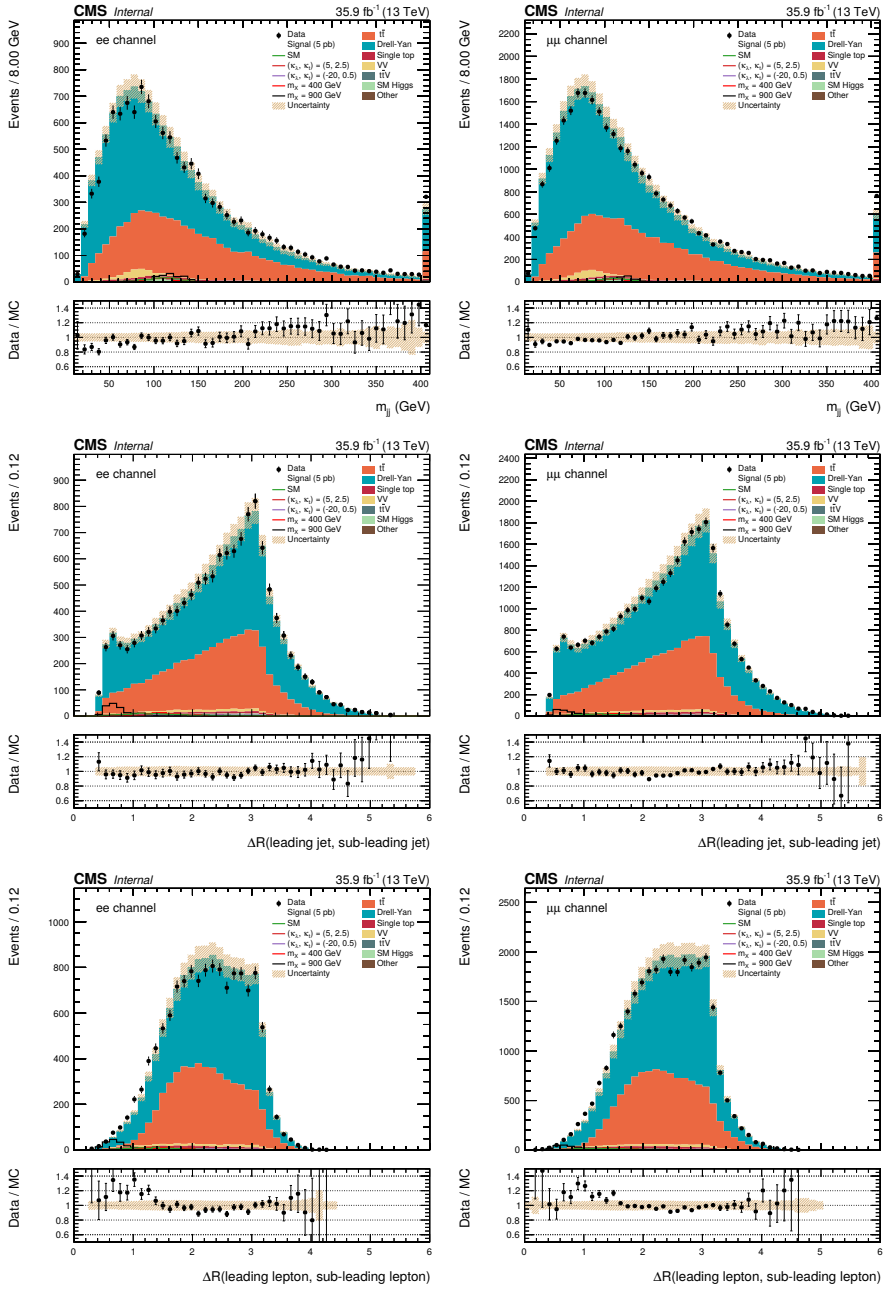

Figure 3.7. | Validation of the DY estimation method in a data control region dominated by the DY process, $76 \leq m_{\ell\ell} < 106$ GeV, in the e^+e^- (left) and $\mu^+\mu^-$ (right) channels. The variables shown are the same as in Fig. 3.6, i.e. (from top to bottom) the invariant mass and ΔR separation between the jets, and the ΔR between the leptons. All backgrounds but DY are taken from the simulation.

(BDTs) built using those variables turned out to be weakly correlated with m_{jj}. This provides a straightforward way to define signal-free control regions in data, in order to check the agreement between data and simulation in the distribution of the classifier, and possibly constrain some of the systematic uncertainties affecting the background processes.

The kinematic variables used rely on the different decay chains in the signals and in the main $t\bar{t}$ background. In the signals, the systems formed by the two jets and the two leptons each originate from a different resonance, whereas the top quarks in the $t\bar{t}$ process each decay to a lepton and a jet. In particular for high-mass resonances decaying to HH, one would therefore expect widely separated high-p_T dijet and dilepton systems, with small angular separation between the leptons and between the jets, respectively. Additionally, due to the spin-0 nature of the Higgs boson the two leptons in the signal feature low angular separation and low invariant mass. The eight variables chosen are: $m_{\ell\ell}$, $\Delta R_{\ell\ell}$, ΔR_{jj}, $\Delta\phi_{\ell\ell,jj}$ (defined as the $\Delta\phi$ between the dijet and the dilepton systems), $p_T^{\ell\ell}$, p_T^{jj}, $\min(\Delta R_{j,\ell})$, and m_T. The latter is used for the separation it provides between the signals and the lesser DY background, and is defined as:

$$m_T = \sqrt{2 p_T^{\ell\ell} p_T^{miss} \left[1 - \cos\Delta\phi(\ell\ell, \vec{p}_T^{miss})\right]}. \tag{3.7.}$$

For all of these variables, good agreement between data and simulation is found. Figures 3.2 (right), 3.8, and 3.15 (right) show four of these input variables; the remaining four are available in App. B. On top of these kinematic quantities, we add a Boolean variable indicating whether the event had same- or different-flavour leptons. This allows the classifier to adapt its response to the different background composition of the $e^+e^-/\mu^+\mu^-$ and $\mu^\pm e^\mp$ channels, without having to train a different classifier for each channel.

A major difficulty in building a multivariate classifier in a search for new physics processes is the fact that the signal hypothesis is not fixed, contrary to the discussion of Sec. 2.4.1 where a specific pdf for the signal was assumed. Crucially, the signal kinematics present a strong dependence on the hypothesised signal parameters, m_X for the resonant and κ_λ and κ_t for the nonresonant case, over the range of parameter values considered. This implies that a classifier trained to recognise a signal for a specific parameter value will not perform well when applied to a different choice of signal parameter. The obvious solution to train a different classifier for each available signal sample is clearly not practical, given the large number of samples and the difficulty inherent in building such classifiers. The alternative of training one or several classifier(s) using the complete set or subsets of considered signal hypotheses will only impose a suboptimal compromise between performance and practicality. A recent suggestion for avoiding these concerns is to build a *parameterised* classifier [258].

Parameterised classifiers differ from regular classifiers in that the parameter(s) (mass, couplings, ...) relative to the signal hypotheses are treated like other input variables. By using all the available signal samples in the training phase, the classifier is then able to infer the dependence of the signal behaviour on these parameters. When evaluating the classifier on data and simulation to derive the distributions needed for the signal search, the signal parameters are frozen to a specific value, and only the signal sample corresponding to that value is considered. This procedure is repeated for every parameter

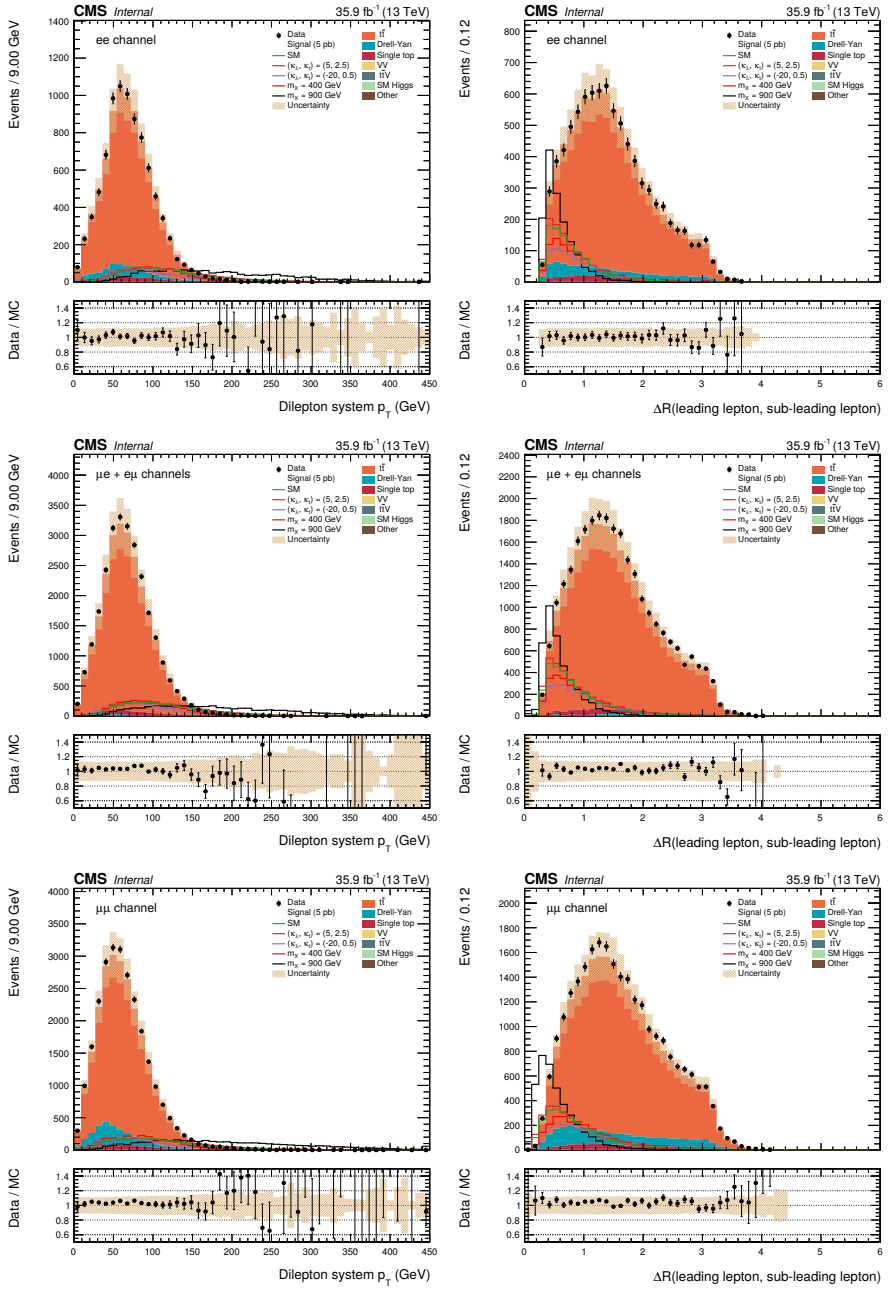

Figure 3.8. | Distributions of two of the eight kinematic input variables of the ANNs used to discriminate signal from background events, in the e^+e^- (top), $\mu^\pm e^\mp$ (middle) and $\mu^+\mu^-$ (bottom) channels, after requiring all selection criteria given in Tab. 3.4. Left: p_T of the dilepton system. Right: angular distance between the two leptons.

value for which a signal sample is available. While in principle this technique can be applied to any multivariate classifying algorithm, in practice ANNs are the tool of choice for the task. Indeed, we expect the signal kinematics to depend in a continuous way on the signal parameters, and ANNs are precisely built in a way that defines a continuous function of the input variables. Since signal parameters are not defined for background processes, background events are assigned random parameter values. The values used should be the same as those of the signal samples used in the training, and in the same proportions as the selected events for each sample. The resulting classifier

1. should perform as well for each signal as dedicated non-parameterised classifiers, and
2. should be able to interpolate the behaviour of the signal as a function of the signal parameters, and also perform well on samples not seen during the training phase.

For this analysis we have constructed two parameterised neural networks, which we stress represent the first-ever application of parameterised classifiers in an analysis of LHC data. The first network was trained to recognise the resonant signals, and was given m_X in addition to the nine variables introduced above. All 13 samples of spin-0 resonant production were used to that end. The second covered the nonresonant case, using κ_λ and κ_t as extra input variables, with 32 signal points at $\kappa_\lambda = -20, -5, 0, 1, 2.4, 3.8, 5, 20$ and $\kappa_t = 0.5, 1, 1.75, 2.5$. We considered only the main background processes in the training, namely $t\bar{t}$, Drell–Yan and single top quark production, after requiring all the selection criteria described in Tab. 3.4. For the DY process, we used simulated events without b tagging requirements, applying the per-event weights as defined by (3.5) to model the effect of b tagging. The background contributions were scaled so that each process contributed with the same weight to the loss function as what is expected in data, whereas the signal samples were scaled so that the sum of signals had equal weight as the sum of backgrounds.

The definition and training of the ANNs was carried out with the `Keras` toolkit [259], building upon the `TensorFlow` machine learning framework [260], on a computer system featuring a general-purpose graphical processing unit (GPGPU). This set of tools provided us with a high flexibility in defining the network architecture and yielded training times of under a minute, allowing us to perform extensive experimentation with the network parameters. The various architectures tried featured different number of layers (never more than 5) and of neurons (a few tens to a few hundred), a dropout layer with varying dropout fraction, or different minimisation algorithms with varying learning rates. Using batch normalisation or loss regularisation did not significantly improve the training convergence. The following network structures were found to perform satisfyingly well for the resonant (nonresonant) case (see Sec. 2.4.1 for terminology):

- 10 (11) input variables,
- 5 hidden layers with 100 neurons each, ReLU activation function,
- 1 dropout layer with $p = 0.2$ ($p = 0.35$),
- 1 output with sigmoid activation,
- cross-entropy loss,
- batch size of 5000 events.

The gradient descent algorithm used to train the networks was Adam, with the default

parameters recommended by the authors of Ref. [222]. The initial learning rate was fixed at 0.001 (0.005) and the training was stopped after 100 iterations of the minimisation algorithm over the full training dataset (epochs). In addition, it was found that decaying the learning rate by a factor 10 after 50 epochs further improved the convergence of the minimisation.

The performance of the "resonant" and "nonresonant" networks is illustrated on Fig. 3.9 with so-called ROC curves. These curves show the efficiency to select signal events as a function of the efficiency to keep background events, when applying a sliding requirement on the score of the classifier. A high signal efficiency for a fixed background rejection indicates good performance.

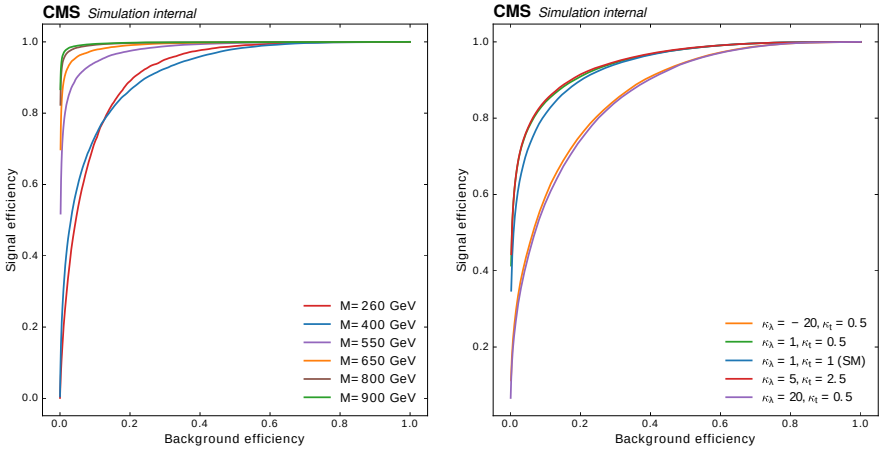

Figure 3.9. | Visualisation, using ROC curves, of the performance of the parameterised ANN classifiers used in the search for resonant (left) and nonresonant (right) HH production. The networks are evaluated using one signal sample at a time, fixing the input signal parameters (m_X on the left and κ_λ and κ_t on the right) to their corresponding value.

Distributions of the resonant and nonresonant classifier scores in data and in simulation are shown on Figs. 3.10 and 3.11. The classifiers are evaluated using a few indicative signal samples and parameter values. As expected with the parameterised ML technique, the distributions of data and background are different for each assumed value. To further illustrate the technique we consider the resonant case, since having only one signal parameter simplifies the visualisations. Figure 3.12 shows how the distribution of the score evolves as a function of m_X if for the signal, the assumed mass of the resonance in the signal samples is varied simultaneously with the value of the parameter used as input to the classifier. On the other hand, Fig. 3.13 illustrates the dependence of the score distribution w.r.t. the input parameter value when keeping the signal sample fixed at either $m_X = 350$ GeV or $m_X = 800$ GeV. It can be seen that the classifier has learned to recognise signal events corresponding specifically to $m_X = 350$ GeV, even though the amount of kinematic information provided to the ANN is not sufficient to fully reconstruct m_{HH}. This indicates that the classifier relies on different features of the signals and backgrounds at masses below, at, or above $m_X = 350$ GeV. On the other hand,

for resonances heavier than $m_X \gtrsim 600$ GeV the classifier response saturates, showing that beyond a certain threshold the same set of features can be used to distinguish signal from background events.

In order to further validate the behaviour of the parameterised ANNs, we compared their performance with that obtained with nonparameterised alternatives. For the resonant case, we trained dedicated ANNs using the signals at $m_X = 400$ GeV and $m_X = 900$ GeV, as well as one ANN with a mixture of all signals samples. As expected, the dedicated classifiers only performed well on a narrow range of m_X values, and could not recognise signal events for different m_X than those used for training them. What's more, the parameterised classifier performed significantly better than the "regular" ANNs for nearly all values of m_X, showing that providing the network with knowledge about the true m_X helped it adapting its response to the different signal hypotheses. Finally, we trained a parameterised ANN using all signal samples, except for $m_X = 650$ GeV. As it turned out, that leave-one-out model yielded comparable performance to the full model, even when evaluated on the signal at $m_X = 650$ GeV. This indicates that the classifier learned the dependence of the signal kinematics as a function of m_X, and was able to interpolate it to cases not seen during the training phase. Figure 3.14 shows the expected asymptotic upper limits (see Sec. 2.4.2) on the cross section for $X \to HH \to b\bar{b}VV \to b\bar{b}\ell\nu\ell\nu$, as a function of m_X, obtained with these different classifiers. For the nonresonant search, we compared the ROC curves obtained with the parameterised classifiers, evaluated on a few parameter points, with dedicated regular ANNs trained with only the corresponding signals samples (see Fig. 3.14, right), and checked that the performances were comparable.

For the statistical inference on the different considered signal hypotheses, we divided the selected events into three exclusive regions, defined by $m_{jj} < 75$ GeV, $75 \leq m_{jj} < 140$ GeV and $m_{jj} \geq 140$ GeV. The central region is enriched in signal due to $H \to b\bar{b}$ decays yielding pairs of jet with an invariant mass close to the true m_H. Mostly due to neutrinos present in b jets (from leptonic B and D hadron decays) escaping detection, the distribution of m_{jj} for signals does not peak at m_H but at lower values. The boundaries of the three m_{jj} bands were not optimised for signal sensitivity, but with the aim of defining two signal-free control regions which enable us to validate the agreement between data and simulation of the ANN score distribution, and constrain nuisance parameters in the background model using data. Figures 3.15 and 3.16 show the distribution of m_{jj} and p_T^{jj}, and of the resonant and nonresonant ANN score distribution in the three m_{jj} regions defined above.

While signal samples for resonant Higgs pair production are only available for 13 values of m_X, we would like to probe a larger number of values to ensure we do not miss a potential signal in data. To that end, we *interpolate* the predicted signal yields, independently in each bin of the ANN templates in the m_{jj} regions, as a function of m_X. This enables us to build signal templates for arbitrary values of m_X, even though the number of simulated samples is limited. The interpolation algorithm used to that end is the Akima sub-spline method [261], which avoids pitfalls common with polynomial or spline interpolation strategies, such as the appearance of an unwanted oscillatory behaviour. For the background samples and the data no interpolation is necessary, and we only need to reevaluate the parameterised ANN using the same set of m_X values

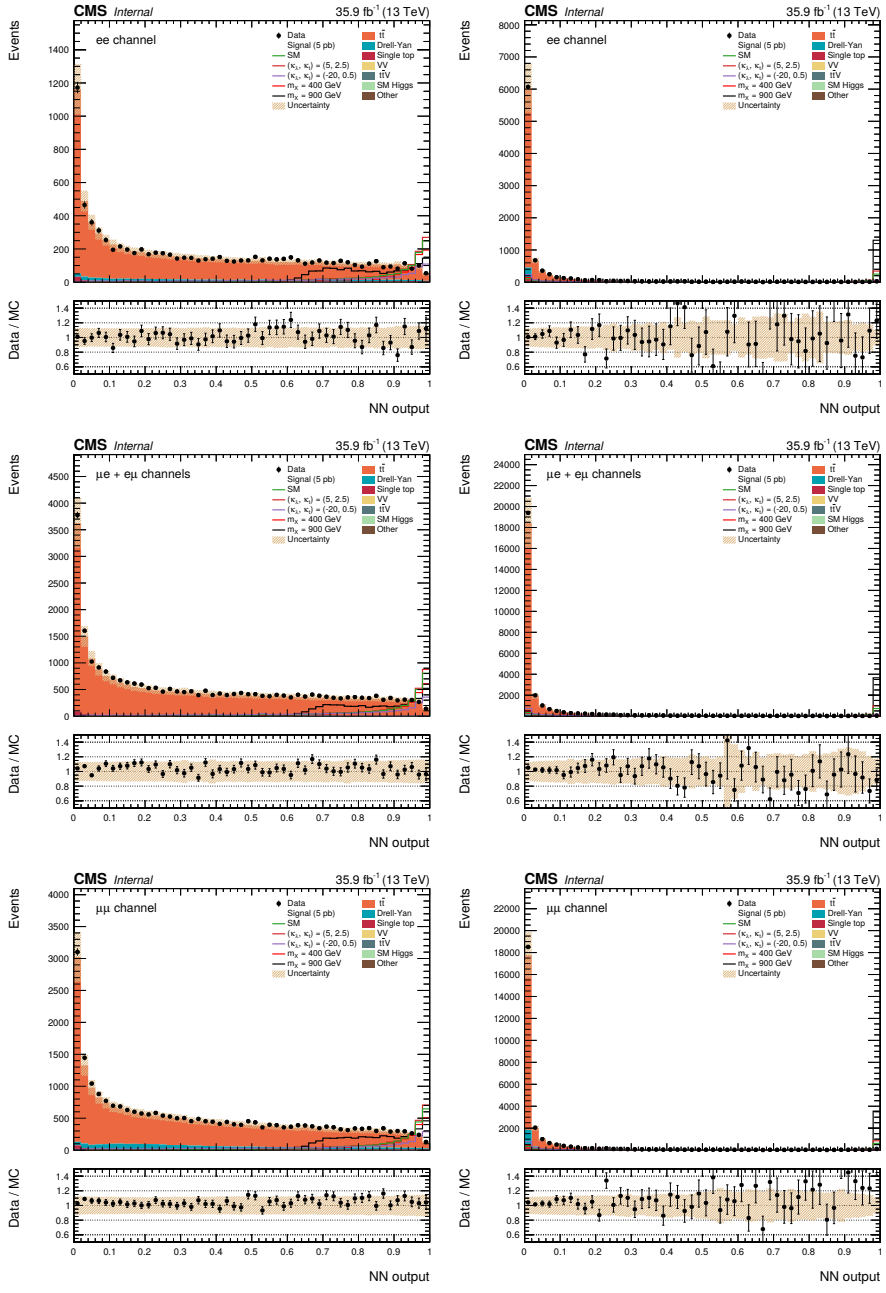

Figure 3.10. ‖ Distribution of the parameterised classifier for the resonant case, in the e^+e^- (top), $\mu^\pm e^\mp$ (middle) and $\mu^+\mu^-$ (bottom) channels, after requiring the selection criteria given in Tab. 3.4. The classifier is evaluated assuming $m_X = 400$ GeV on the left, and $m_X = 900$ GeV on the right.

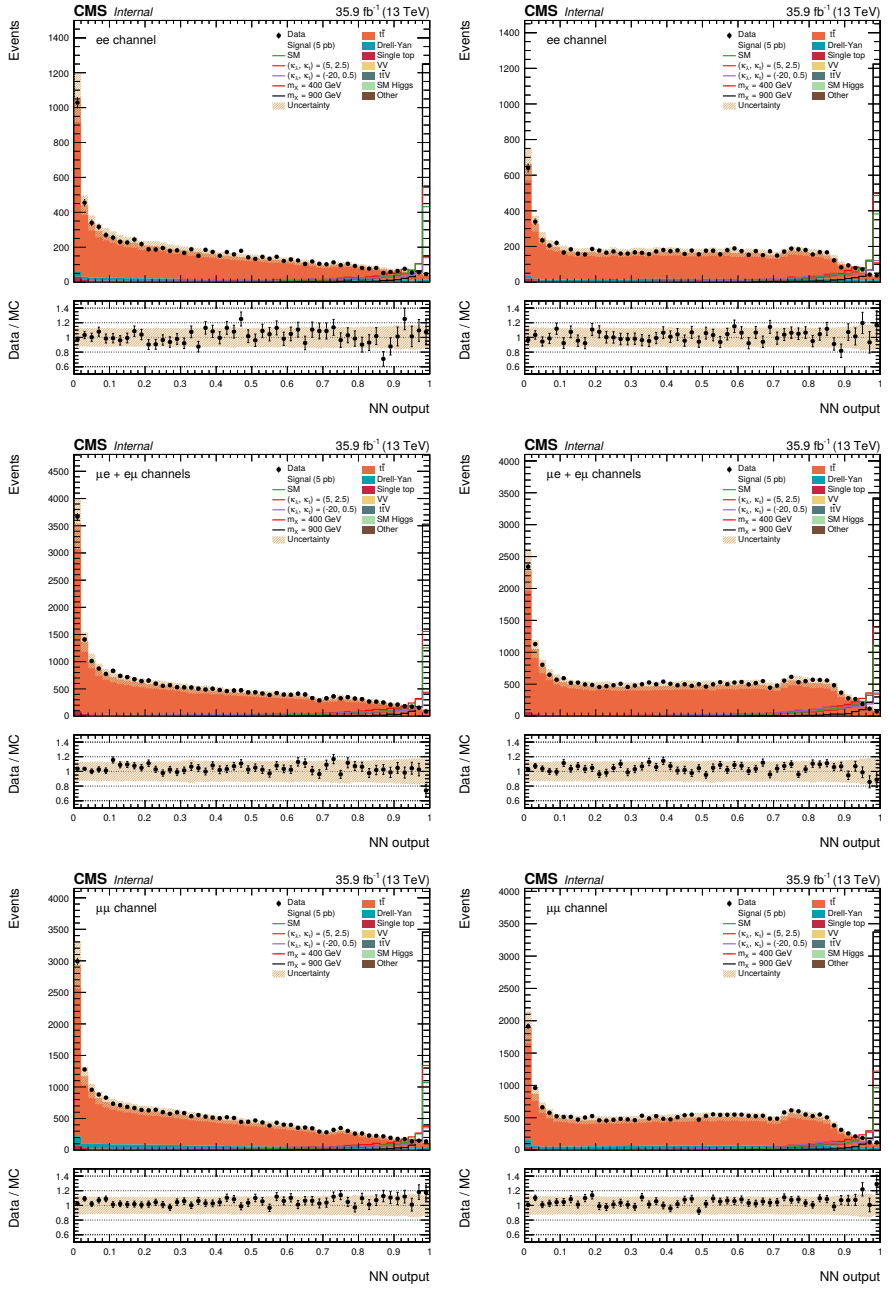

Figure 3.11. Distribution of the parameterised classifier for the nonresonant case, in the $\mathrm{e^+e^-}$ (top), $\mu^{\pm}\mathrm{e}^{\mp}$ (middle) and $\mu^+\mu^-$ (bottom) channels, after requiring the selection criteria given in Tab. 3.4. The classifier is evaluated assuming $(\kappa_\lambda, \kappa_t) = (1,1)$ on the left, and $(\kappa_\lambda, \kappa_t) = (-20, 0.5)$ on the right.

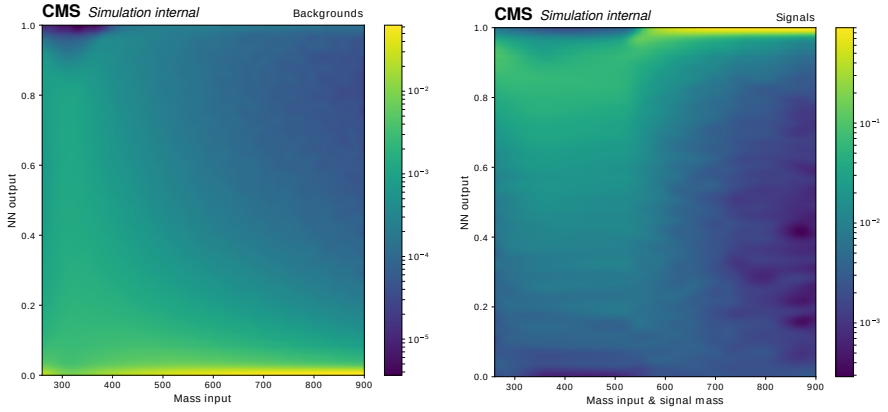

Figure 3.12. Normalised distributions of the resonant classifier score (ordinate) on the backgrounds (left) and the signals (right), conditional on the signal parameter m_X (abscissa). The classifier is evaluated on the same background events for each value of m_X, whereas for the signals only the sample corresponding to each mass hypothesis is used. The plots of Fig. 3.10 correspond to vertical slices (at $m_X = 400$ GeV and $m_X = 900$ GeV) of these two-dimensional distributions.

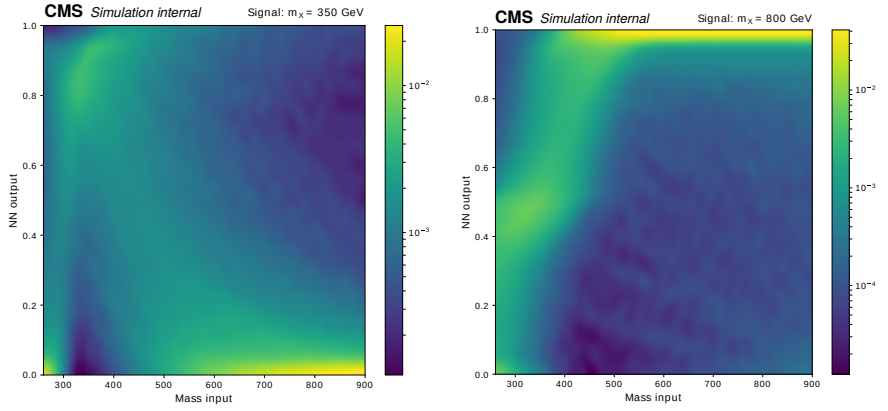

Figure 3.13. Distributions of the resonant classifier score (ordinate) on two different signals samples, $m_X = 350$ GeV (left) and $m_X = 800$ GeV (right) as a function of the signal parameter m_X given as input to the ANN (abscissa). Contrary to Fig. 3.12 (right), the same events are used for every assumed value of m_X, i.e. these signals are treated in the same way as the backgrounds on Fig. 3.12 (left).

as that used for the signals. In the case of nonresonant production no interpolation is needed either since it is possible to reweight the signal samples to arbitrary probed values of κ_λ and κ_t.

3.4. Systematic uncertainties

We investigate sources of systematic uncertainties and their impact on the statistical interpretation of the results by considering both uncertainties in the normalisation of the

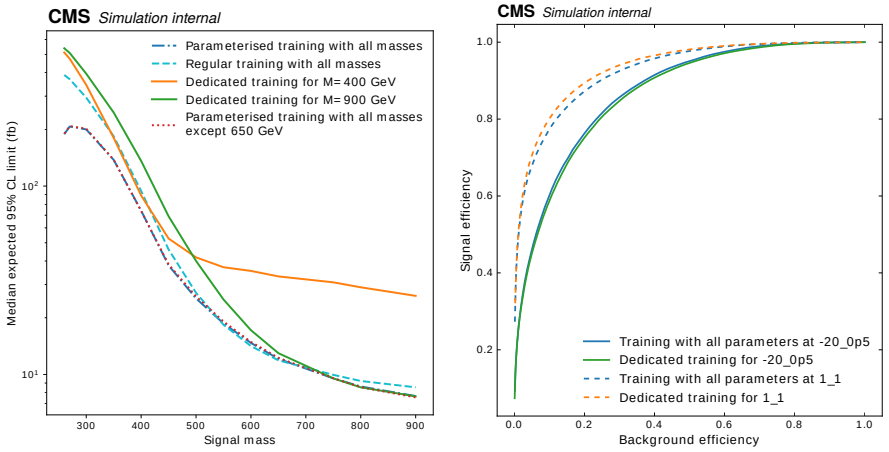

Figure 3.14. Left: median expected upper limit at 95% CL on the product of the production cross section for spin-0 X and branching fraction for $X \to HH \to b\bar{b}VV \to b\bar{b}\ell\nu\ell\nu$, as a function of their mass m_X, obtained (without systematic uncertainties) with the binned shape of the classifier scores using different parameterised and non-parameterised ANNs. The dashed-dotted line corresponds to the ANN used in the search for resonant HH production. The dashed line was obtained using a non-parameterised ANN trained using all different signals samples; the solid yellow and green lines using only the signal samples at $m_X = 400$ GeV and $m_X = 900$ GeV, respectively. For the dotted line, we trained a parameterised ANN using all signal samples, except for $m_X = 650$ GeV. Right: ROC curves showing the discrimination between signal and background obtained with the parameterised ANN used in the search for nonresonant HH production, evaluated at $(\kappa_\lambda, \kappa_t) = (1,1)$ and $(\kappa_\lambda, \kappa_t) = (-20, 0.5)$ with the corresponding signal samples, compared with that obtained with non-parameterised ANNs trained with only the signals at $(\kappa_\lambda, \kappa_t) = (1,1)$ and $(\kappa_\lambda, \kappa_t) = (-20, 0.5)$.

various processes in the analysis, as well as those affecting the shapes of the distributions. To each source of systematic uncertainty corresponds a nuisance parameter in the statistical model used for inference, as described in Sec. 2.4.2. When several processes are affected by the same source of experimental uncertainty, they are assigned a single, common nuisance parameter.

Theoretical uncertainties in the cross sections of backgrounds estimated using simulation are considered as systematic uncertainties in the yield predictions. We only consider the uncertainties on total cross sections that are not related to the renormalisation and factorisation scale and to the PDFs, since those are already taken into account for both shape and normalisation uncertainties through the simulated samples. As described in Sec. 2.3.10, the uncertainty in the total integrated luminosity is determined to be 2.5%. Since the different sources of theoretical and experimental systematic uncertainties and their evaluation have been extensively discussed in Sec. 1.2 and 2.3.11, they will only be briefly recalled here.

The following sources of systematic uncertainties that affect both the normalisation and shape of the templates used in the statistical evaluation are considered:

- **Trigger efficiency, lepton identification and isolation:** uncertainties in the T&P measurements of trigger efficiencies as well as electron and muon isolation and

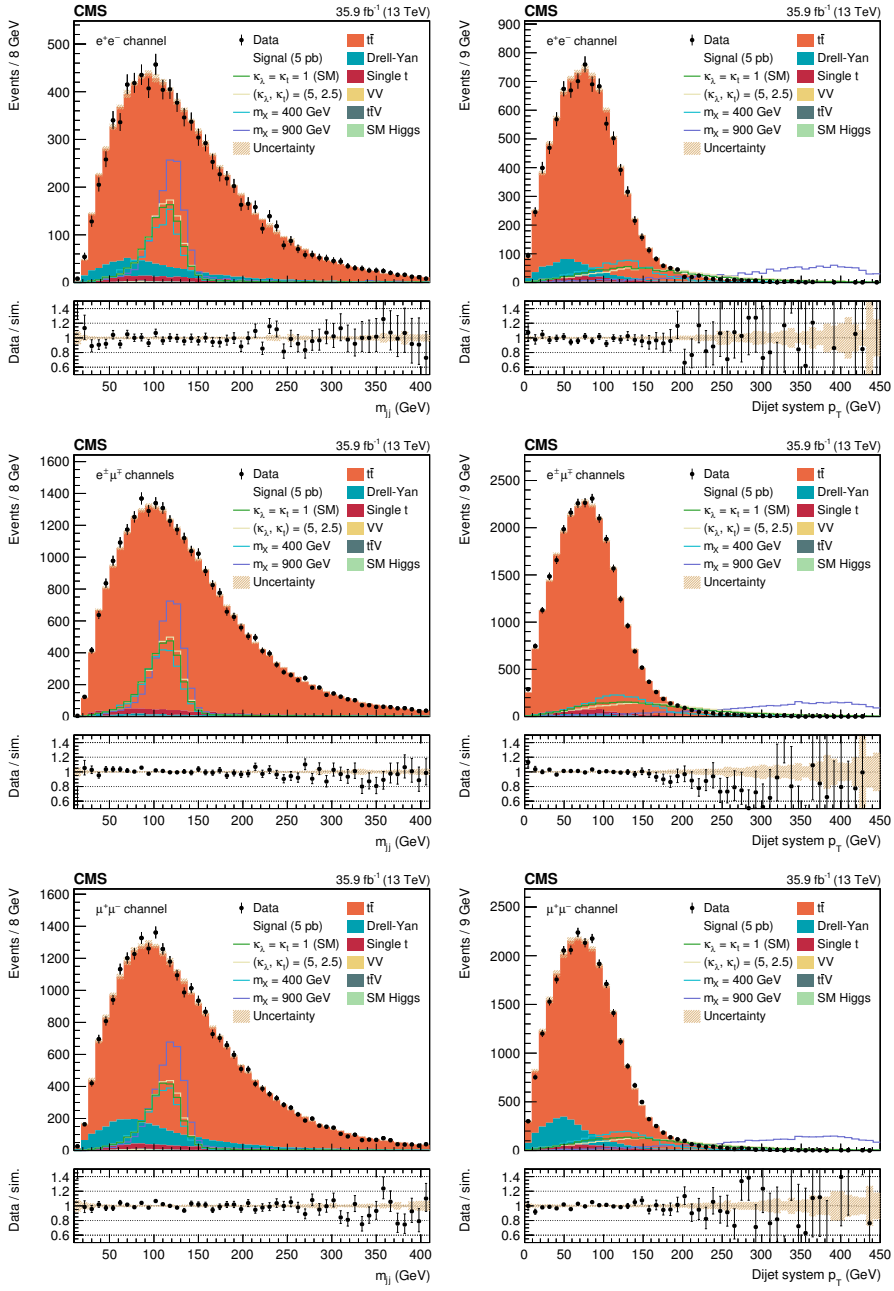

Figure 3.15. The dijet system invariant mass (left) and p_T (right) distributions in the e^+e^- (top), $\mu^\pm e^\mp$ (middle), and $\mu^+\mu^-$ (bottom) channels. Shaded bands show post-fit systematic uncertainties.

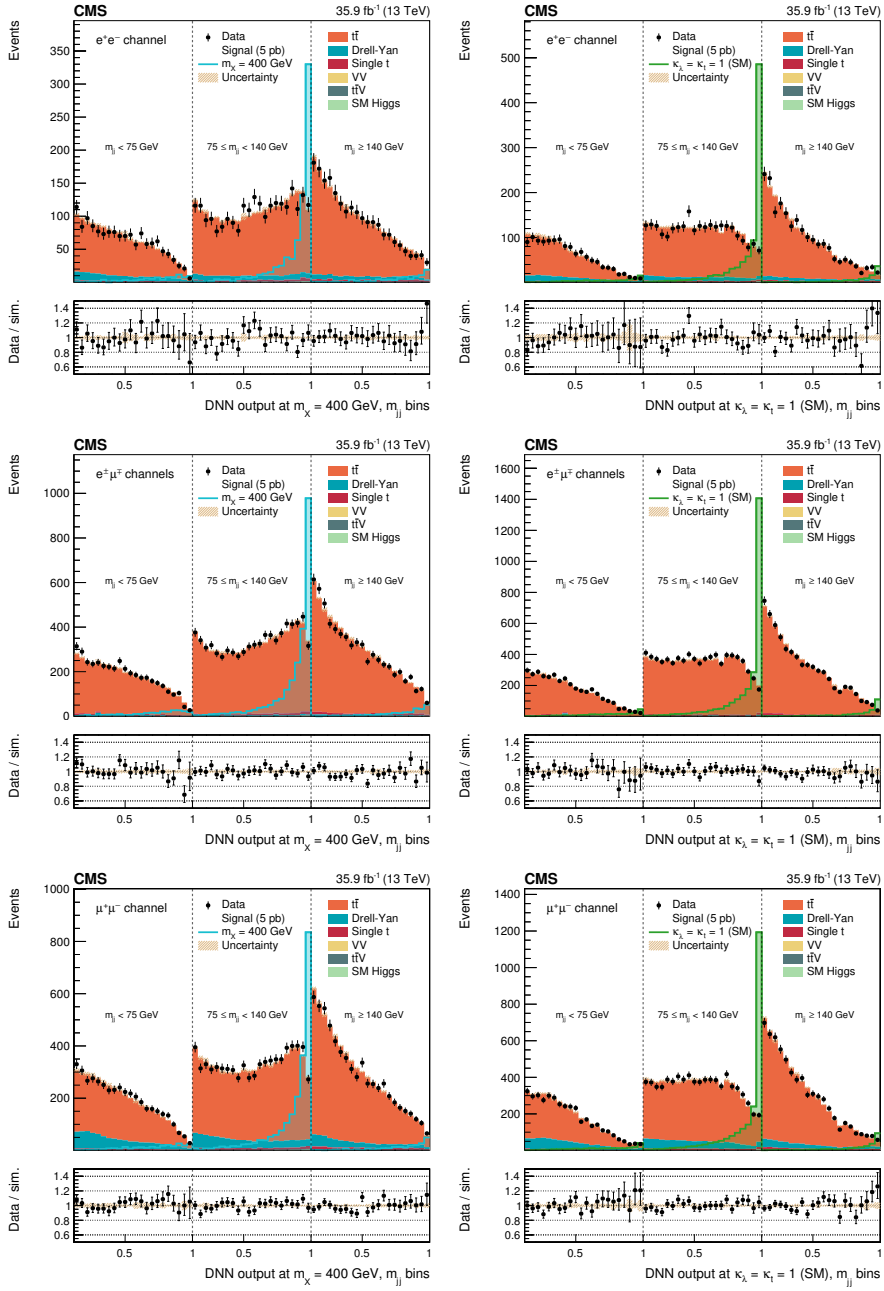

Figure 3.16. ‖ The ANN output distributions for the e^+e^- (top), $\mu^\pm e^\mp$ (middle), and $\mu^+\mu^-$ (bottom) channels, in three different m_{jj} regions: $m_{jj} < 75$ GeV, $m_{jj} \in [75,140[$ GeV, and $m_{jj} \geq 140$ GeV. The parameterised resonant ANN output (left) is evaluated at $m_X = 400$ GeV and the parameterised nonresonant ANN output (right) is evaluated at $\kappa_\lambda = 1$, $\kappa_t = 1$. Shaded bands show post-fit systematic uncertainties.

identification efficiencies are considered as sources of systematic uncertainties. These are evaluated as a function of lepton p_T and η, and their effect on the analysis is estimated by varying the corrections to the efficiencies by ±1 standard deviation.

- **Jet energy scale and resolution:** uncertainties in the jet energy scale are of the order of a few percent and are computed as a function of jet p_T and η. A difference in the jet energy resolution of about 10% between data and simulation is accounted for by worsening the jet energy resolution in simulation by η-dependent factors. The uncertainty due to these corrections is estimated by a variation of the factors applied by ±1 standard deviation. For the jet energy scale corrections, 27 different sources of uncertainty are considered. All variations of jet energies are propagated to $\vec{p}_T^{\,miss}$.
- **b tagging:** b tagging efficiency and light-flavour mistag rate corrections and associated uncertainties are determined as a function of the jet p_T. Their effect on the analysis is estimated by varying these corrections by ±1 standard deviation, separately for heavy- and light-flavour jets. Figure 3.17 (left) shows the uncertainty from the heavy-flavour jet tagging efficiency on the shape and normalisation of the $t\bar{t}$ background.
- **Pileup:** the measured total inelastic cross section is varied by ±5% to produce different expected pileup distributions.
- **Renormalisation and factorisation scale uncertainty:** this uncertainty is estimated by varying the renormalisation (μ_R) and the factorisation (μ_F) scales used during the generation of the simulated samples independently by factors of 0.5, 1, or 2. Unphysical cases, where the two scales are at opposite extremes, are not considered. An envelope is built from the six possible combinations by keeping maximum and minimum variations for each bin of the distributions, and is used as an estimate of the scale uncertainties for all the background and signal samples[1]. The impact of this uncertainty on the shape and normalisation of the $t\bar{t}$ template is shown on Fig. 3.17 (right). A deficiency of this method is that one assumes that the variations within the envelopes across all bins and channels in the likelihood are correlated. This can be optimistic, as the full possible variations in the shape of the templates are not considered. For instance, Fig. 1.3 shows that missing NLO corrections result in changes in the SM signal kinematics that are anti-correlated between low and high values of m_{HH}, whereas the scale uncertainty envelopes at LO mostly impact the overall normalisation of the process. However, the alternative of assuming complete decorrelation between the variations in different bins would be overly pessimistic. While higher-order theoretical calculations—when available—can provide some guidance on the effects missed by the use of lower-order simulated samples, it is often unclear how these effects propagate to nontrivial observables such as multivariate classifier scores. Applying differential K-factors on a specific parton-level variable to

[1] The scale uncertainty of the tW background, which dominates the single top quark contribution, has not been evaluated. However, the effect of this omission on the results is expected to be negligible, as this process only corresponds to about 2% of the total background yields. Enlarging the scale uncertainties of other single top quark processes by a factor 100, so that the relative uncertainty on the combined single top quark contribution is twice the one seen for $t\bar{t}$ production, or enlarging the scale uncertainty of the $t\bar{t}$ process to account for the missing effect only degrades the observed limit on SM HH production by less than 0.5%.

correct its shape is indeed risky, since that might impact other variables in a poorly controlled manner. Furthermore, the case can be made that for setting a limit on the signal cross section, the sensitivity is driven by the overall normalisation of the predicted signal contribution rather than by its shape, so that this evaluation procedure does not yield an optimistic result. Resolving this question, which affects many analyses in the field, goes beyond the scope of the present work, and in order to remain consistent with HH searches in other final states and thus enable the combination of their results we have stuck to the recommended procedure described above.

- **PDF uncertainty:** the magnitudes of the uncertainties related to the PDFs and the value of the strong coupling constant for each simulated background and signal process are obtained using variations of the NNPDF 3.0 set [26], following the PDF4LHC prescriptions [60,61].
- **Simulated sample size:** the finite nature of simulated samples is considered as an additional source of systematic uncertainty. For each bin of the distributions, one additional uncertainty is added, where only the considered bin is altered by ±1 standard deviation, keeping the others at their nominal value.
- **DY background estimate from data:** the systematic uncertainties listed above, which affect the simulation samples, are propagated to ϵ_k and F_{kl}, both computed from simulation. These uncertainties are then propagated to the weights W_{sim} and to the normalisation and shape of the estimated DY background contribution. The uncertainty due to the finite size of the simulation samples used for the determination of ϵ_k and F_{kl} is also taken into account. More details about the propagation of uncertainties to the estimate of the DY background are given in the next subsection. Since previous measurements [262,263] have shown that the flavour composition of DY events with associated jets in data is compatible with the simulation within scale uncertainties, which are taken into account, no extra source of theoretical uncertainty has been considered for F_{kl}. To account for residual differences between the e^+e^- and $\mu^+\mu^-$ channels not taken into account by F_{kl}, due to the different requirements on lepton p_T, a 5% uncertainty in the normalisation of the DY background estimate is added in both channels. This corresponds to the difference in the corrections to the normalisation of the data-driven DY predictions in the two channels, which are given in Tab. 3.5.

The effects of these uncertainties on the total yields in the signal region are summarised in Tab. 3.6. However, these figures give only a limited insight into the impact of the different sources of systematic uncertainty on the final analysis sensitivity. To better diagnose the behaviour of the profile likelihood fit of (2.25), two methods are often used. First, it is important to check whether the profiled values of the nuisance parameters are not too different from their default values. Indeed, the modelling of shape uncertainties relies on *interpolations* between templates corresponding to variations of ±1 standard deviations of the associated parameter, and there is no sufficient control far beyond these bounds to *extrapolate* their behaviour much. The left panel of Fig. 3.18 shows the post-fit value and uncertainty of a set of nuisance parameters, for the statistical model used for the resonant spin-0 signal at $m_X = 400$ GeV. None of the 150 nuisance parameters in the model (not shown here) are *pulled* beyond their pre-fit uncertainty. In addition,

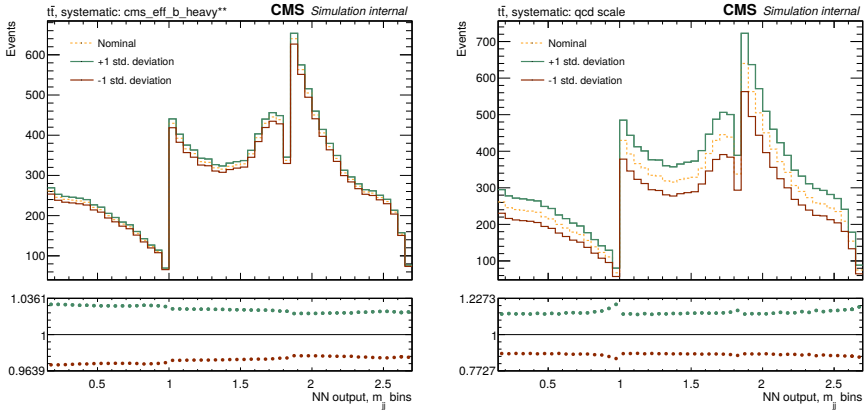

Figure 3.17. Effect of heavy-flavour jet b tagging efficiency uncertainties (left) and of the renormalisation and factorisation scale variations (right) and on the shape and normalisation of the templates for the $t\bar{t}$ process. The distribution of the resonant ANN evaluated at $m_X = 400$ GeV in the $e^{\pm}\mu^{\mp}$ channel is shown (other channels yield similar behaviours). The yield in each bin of the distribution is continuously interpolated between the down ("-1 std. deviation"), nominal and up ("+1 std. deviation") predictions, as detailed in Sec. 2.4.2.

one can ask how much the fitted signal strength is correlated with a specific nuisance parameter, which gives a measure of its impact on the uncertainty in the signal strength, i.e. the analysis sensitivity. This can be estimated by freezing that parameter to its $\pm 1\sigma$ post-fit values and repeating the fit while profiling the remaining nuisances as before. The differences between the signal strengths given by those fits and the nominal one, for the ten nuisance parameters leading to the largest differences, are shown on Fig. 3.18 (right panel). Clearly, theoretical uncertainties in the $t\bar{t}$ background modelling have the largest impact on the sensitivity to the signal, followed by experimental uncertainties such as jet energy scale, integrated luminosity, and lepton identification efficiencies.

Drell–Yan estimation

Uncertainties inherent to the estimation method used for the DY background described in Sec. 3.2 stem from two sources:

1. Uncertainties on the simulated samples (chiefly $t\bar{t}$) being subtracted from the reweighted data. These are estimated in the same way as for non-reweighted (i.e. truly b-tagged) contributions.
2. Uncertainties in the estimation of the weights W_{sim}. These propagate to *both* the reweighted data, and the reweighted simulation being subtracted from the data. Their treatment is detailed in the following sub-sections.

Since the reweighted untagged data and subtracted background simulations are implemented as separate contributions in the statistical model, we can assign to each only the relevant sources of uncertainties. Uncertainties affecting both the event reweighting through the weights W_{sim} and the subtracted background contributions are taken to be fully correlated.

Table 3.6. Summary of the systematic uncertainties and their impact on total background yields and on the SM and $m_X = 400$ GeV signal hypotheses in the signal region. Uncertainties are quoted as a percentage of the yield in each channel separately; uncertainties affecting specific processes are given relatively to the yield of these processes.

Source	Background yield variation	Signal yield variation
Electron identification and isolation	2.0–3.2%	1.9–2.9%
Jet b tagging (heavy-flavour jets)	2.5%	2.5–2.7%
Integrated luminosity	2.5%	2.5%
Trigger efficiency	0.5–1.4%	0.4–1.4%
Pileup	0.3–1.4%	0.3–1.5%
Muon identification	0.4–0.8%	0.4–0.7%
PDFs	0.6–0.7%	1.0–1.4%
Jet b tagging (light-flavour jets)	0.3%	0.3–0.4%
Muon isolation	0.2–0.3%	0.1–0.2%
Jet energy scale	<0.1–0.3%	0.7–1.0%
Jet energy resolution	0.1%	<0.1%
Affecting only $t\bar{t}$ (87.4–96.2% of the total bkg.)		
μ_R and μ_F scales	12.8–12.9%	
$t\bar{t}$ cross section	5.2%	
Simulated sample size	<0.1%	
Affecting only DY in $e^{\pm}\mu^{\mp}$ channel (0.6% of the total bkg.)		
μ_R and μ_F scales	24.6–24.7%	
Simulated sample size	7.7–11.6%	
DY cross section	4.9%	
Affecting only DY estimate from data in same-flavour events (6.6–9.6% of the total bkg.)		
Simulated sample size	18.8–19.0%	
Normalisation	5.0%	
Affecting only single top quark (2.4–2.7% of the total bkg.)		
Single t cross section	7.0%	
Simulated sample size	<0.1–1.0%	
μ_R and μ_F scales (excluding tW)	<0.1–0.2%	
Affecting only signal	SM signal	$m_X = 400$ GeV
μ_R and μ_F scales	24.2%	4.6–4.7%
Simulated sample size	<0.1%	<0.1%

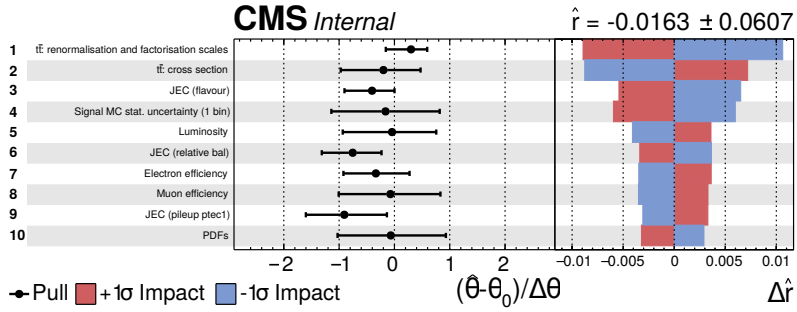

Figure 3.18. The ten nuisance parameters with the largest impacts on the fitted signal strength $\hat{r}$ for the resonant (spin 0) signal at $m_X = 400$ GeV. These nuisances amount to about 8 % of the overall variance of $\hat{r}$. The left panel shows the shifts between post- and pre-fit values $(\hat{\theta} - \theta_0)$ and uncertainties of these parameters, relative to their pre-fit uncertainty $\Delta\theta$, while the right panel shows the impact of each parameter on the signal strength, when varied within ±1 standard deviation. The fourth entry corresponds to the statistical uncertainty due to the finite number of simulated events for the signal in the bin with the largest signal yield (see Fig. 3.16, right).

Statistical uncertainties Due to the finite size of the simulated samples used to compute ϵ_k and F_{kl}, there is a statistical uncertainty attached to these numbers (for each bin they are computed in). We consider a Bayesian approach to propagate these uncertainties and obtain a 68%-level central credible interval on W_{sim}, on each single event. The low and high edges of this interval define "down" and "up" weight variations, which are used to build corresponding "down" and "up" reweighted estimates, so that this uncertainty is propagated all the way through the analysis.

For a given event to be reweighted, its jet kinematics and BDT score define what values of ϵ_k, ϵ_l and F_{kl} are to be used. Their uncertainties are computed and propagated to W_{sim} as follows:

1. We start from a flat prior on ϵ_k. The likelihood for this quantity is obtained using a binomial probability mass function (pmf), hence the posterior distribution for ϵ_k is a Beta distribution whose parameters depend on the number of jets of flavour k falling in the $p_{\text{T}}/|\eta|$ bin considered and passing/failing the b tagging requirement in the simulation.
2. Similarly, we also take a flat prior for the set of F_{kl} fractions. The likelihood for these nine quantities is obtained from a multinominal pmf, hence the posterior distribution for the F_{kl}'s is a Dirichlet distribution whose parameters depend on the number of simulated events with jet flavours (k,l) falling in the considered bin of the BDT output variable.
3. We then estimate the per-event uncertainty on the weight W_{sim} using toys. For each event we generate 10000 toys, whereby ϵ_k, ϵ_l (independently for the two jets) and F_{kl} follow their posterior distributions as described above. From the resulting distribution of values for W_{sim} we compute 16% and 84% quantiles which define the desired low and high variations of W_{sim}.

This approach has the following advantages:

- We fully take into account the non-Gaussian nature of uncertainties on ratios such

as ϵ_k and F_{kl}.
- The dependence between the different F_{kl} fractions (due to the constraint $\sum_{k,l} F_{kl} = 1$) is propagated to the final uncertainties.
- The resulting low and high error values for W_{sim} are well-defined, in particular one always has $W_{\text{sim}} \geq 0$.

The dependence of the estimated uncertainties on the choice of priors has been checked by choosing reasonable alternate priors (such as Jeffreys' priors [264]) and turns out to be negligible. Given the size of the available simulated samples, the resulting statistical uncertainty on W_{sim} turns out to be (on average) around 20% for $12 < m_{\ell\ell} < 76$ GeV. This is the same order of magnitude as the statistical uncertainty of the DY prediction when requiring b-tagged jets in the simulation, which might question the relevance of the approach followed here. However, the impact of these uncertainties is quite different: when relying solely in the simulation, every bin in the final discriminant is subject to an independent uncertainty of around 20%, i.e. a large uncertainty on the overall shape of the predicted distributions. On the other hand, the dedicated procedure for estimating the DY contribution yields smooth distributions with a much more limited shape uncertainty and thus a reduced impact on the overall sensitivity.

Systematic uncertainties The effect on ϵ_k and F_{kl} due to all the uncertainties considered in the analysis and relevant for the considered DY simulation samples has been checked. Some of them turn out to be entirely negligible, and only the following are propagated to the analysis:

- On ϵ_k, ϵ_l: jet energy scale, jet energy resolution and pileup. Since the b tagging efficiencies are corrected by scale factors, the uncertainties on these scale factors are also considered.
- On F_{kl}: jet energy scale, jet energy resolution, pileup and renormalisation and factorisation scales.

Given an event and a source of systematic uncertainty, its effect on W_{sim} is propagated as follows:

1. The errors on ϵ_k, ϵ_l (independently for both jets) and F_{kl} are computed from the nominal and corresponding $\pm 1\sigma$ ("up"/"down") variations:
$$\Delta(\epsilon_k) = \epsilon_k^{\text{up/down}} - \epsilon_k \tag{3.8.}$$
$$\Delta(F_{kl}) = F_{kl}^{\text{up/down}} - F_{kl} \tag{3.9.}$$
2. These errors are propagated to W_{sim} using a Taylor expansion:
$$\Delta^{\text{up/down}}(W_{\text{sim}}) = \sum_{k,l=\text{b,c,light}} \left(\epsilon_l^{\text{j}_2} F_{kl} \Delta(\epsilon_k^{\text{j}_1}) + \epsilon_k^{\text{j}_1} F_{kl} \Delta(\epsilon_l^{\text{j}_2}) + \epsilon_k^{\text{j}_1} \epsilon_l^{\text{j}_2} \Delta(F_{kl}) \right) \tag{3.10.}$$
3. The "up" and "down" values for W_{sim} are taken to be $W_{\text{sim}} + \Delta^{\text{up/down}}(W_{\text{sim}})$.

The systematic up and down variations on W_{sim} are then propagated just as for the statistical uncertainties described above.

3.5. Results

A binned maximum likelihood fit is performed in order to extract best fit signal cross sections. The fit is performed using templates built from the DNN output distributions in the three m_{jj} regions, as shown in Fig. 3.16, and in the three channels (e^+e^-, $\mu^+\mu^-$, and $e^\pm\mu^\mp$). The likelihood function is of the form of (2.25), i.e. it is the product of the Poisson likelihoods over all bins of the templates and over the three channels, as well as constraint terms associated with the nuisance parameters of the model.

The best-fit values for all the nuisance parameters, as well as the corresponding post-fit uncertainties, are extracted by performing another binned maximum likelihood fit, in the background-only hypothesis, of the m_{jj} vs. DNN output distributions (such as Fig. 3.16) to the data. Only nuisance parameters affecting the backgrounds are considered in that case. These best-fit values are used for the visualisation of post-fit background predictions shown in Figs. 3.15 and 3.16. The post-fit uncertainties are obtained by drawing random samples from the fit's covariance matrix and building envelope templates for each background process, thereby taking into account the correlations between fitted nuisance parameters.

3.5.1. Resonant production

The fit results in signal cross sections compatible with zero; no significant excess above background predictions is observed for X particle mass hypotheses between 260 and 900 GeV. We set upper limits at 95% confidence level (CL) on the product of the production cross section for X and branching fraction for $X \to HH \to b\bar{b}VV \to b\bar{b}\ell\nu\ell\nu$ using the asymptotic modified frequentist method with the CL_s citerion, as described in Sec. 2.4.2, as a function of the X mass hypothesis. The limits are shown on Fig. 3.19. The observed upper limits on the product of the production cross section and branching fraction for a narrow-width spin-0 resonance range from 430 to 17 fb, in agreement with expected upper limits of 340^{+140}_{-100} to 14^{+6}_{-4} fb. For narrow-width spin-2 particles produced in gluon fusion with minimal gravity-like coupling, the observed upper limits range from 450 to 14 fb, in agreement with expected upper limits of 360^{+140}_{-100} to 13^{+6}_{-4} fb.

The left plot of Fig. 3.19 shows possible cross sections for the production of a radion, for the parameters $\Lambda_R = 1$ TeV (mass scale) and $kL = 35$ (size of the extra dimension). The right plot of Fig. 3.19 shows possible cross sections for the production of a Kaluza–Klein graviton, for the parameters $k/\overline{M_{Pl}} = 0.1$ (curvature) and $kL = 35$. These cross sections are taken from Ref. [135], and assume absence of mixing with the Higgs boson.

3.5.2. Nonresonant production

Likewise for the nonresonant case, the fit results in signal cross sections compatible with zero; no significant excess above background predictions is seen. We set upper limits at 95% CL on the product of the Higgs boson pair production cross section and branching fraction for $HH \to b\bar{b}VV \to b\bar{b}\ell\nu\ell\nu$ using the asymptotic CL_s, combining the e^+e^-, $\mu^+\mu^-$ and $e^\pm\mu^\mp$ channels. The observed upper limit on the SM $HH \to b\bar{b}VV \to b\bar{b}\ell\nu\ell\nu$ cross section is found to be 72 fb, in agreement with an expected upper limit of 81^{+42}_{-25} fb. Including theoretical uncertainties in the SM signal cross section, this observed upper

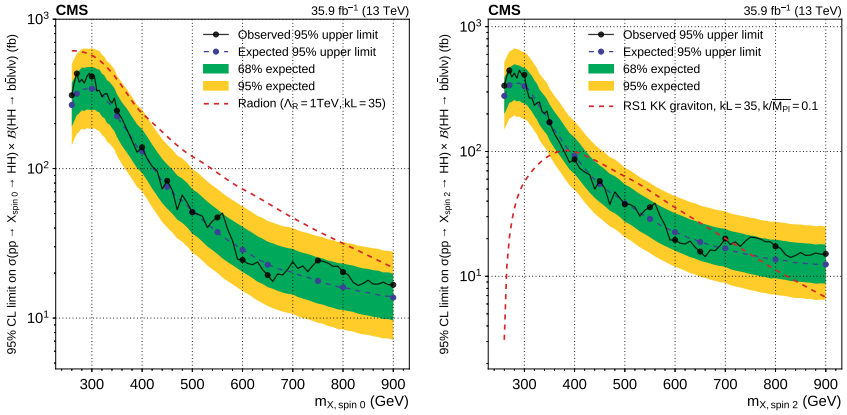

Figure 3.19. Expected (dashed) and observed (continuous) 95% CL upper limits on the product of the production cross section for X and branching fraction for $X \to HH \to b\bar{b}VV \to b\bar{b}\ell\nu\ell\nu$, as a function of m_X. The inner (green) band and the outer (yellow) band indicate the regions containing 68 and 95%, respectively, of the distribution of limits expected under the background-only hypothesis. These limits are computed using the asymptotic CL_s method, combining the e^+e^-, $\mu^+\mu^-$ and $e^\pm\mu^\mp$ channels, for spin-0 (left) and spin-2 (right) hypotheses. The solid circles represent fully-simulated mass points; the interpolation method described in Sec. 3.3 is used between those points. The dashed red lines represent possible cross sections for the production of a radion (left) or a Kaluza–Klein graviton (right), assuming absence of mixing with the Higgs boson [135]. Parameters used to compute these cross sections can be found in the legend.

limit amounts to 79 times the SM prediction, in agreement with an expected upper limit of 89^{+47}_{-28} times the SM prediction.

In the BSM hypothesis, upper limits are set as a function of κ_λ/κ_t, as shown on Fig. 3.20 (left), since the signal kinematics depend only on this ratio of couplings. Red lines show the theoretical cross sections, along with their uncertainties, for $\kappa_t = 1$ (SM) and $\kappa_t = 2$. The theoretical signal cross section is minimal for $\kappa_\lambda/\kappa_t = 2.45$, corresponding to a maximal interference between the diagrams shown on Fig. 1.2.

Excluded regions in the κ_t vs. κ_λ plane are shown on Fig. 3.20 (right). The signal cross sections and kinematics are invariant under a $(\kappa_\lambda, \kappa_t) \leftrightarrow (-\kappa_\lambda, -\kappa_t)$ transformation, hence the expected and observed limits on the production cross section, as well as the constraints on the κ_λ and κ_t parameters respect the same symmetry. The red region corresponds to parameters excluded at 95% CL with the observed data, whereas the dashed black line and the blue areas correspond to the expected exclusions and the 68 and 95% bands. Isolines of the product of the theoretical cross section and branching fraction for $HH \to b\bar{b}VV \to b\bar{b}\ell\nu\ell\nu$ are shown as dashed-dotted lines.

These results show that the sensitivity to nonresonant HH production in the considered channel, which is three times smaller than the next-best result obtained by CMS (see Tab. 1.3), is not sufficient to probe values of the Higgs boson self-coupling that might be generated by reasonable scenarios of BSM physics, *assuming* all the other Higgs boson couplings take their SM value. However, given the amount of available data, we start being sensitive to simultaneous deviations of both the self-coupling and the top

quark Yukawa coupling. Such large deviations are still conceivable when considering marginalised constraints on the $O_{t\phi}$ operator (see e.g. Ref. [94]), and while not explicitly considered in this work, other non-SM Higgs boson couplings might yet increase the cross section of Higgs boson pair production. This indicates that HH production remains a crucial experimental channel for probing nonresonant new physics effects.

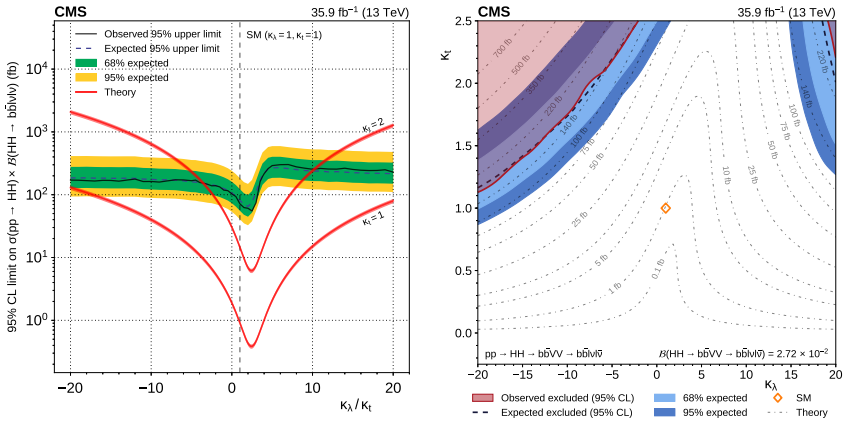

Figure 3.20. Left: expected (dashed) and observed (continuous) 95% CL upper limits on the product of the Higgs boson pair production cross section and branching fraction for $\mathrm{HH} \to \mathrm{b\bar{b}VV} \to \mathrm{b\bar{b}}\ell\nu\ell\nu$ as a function of κ_λ/κ_t. The inner (green) band and the outer (yellow) band indicate the regions containing 68 and 95%, respectively, of the distribution of limits expected under the background-only hypothesis. Red lines show the theoretical cross sections, along with their uncertainties, for $\kappa_t = 1$ (SM) and $\kappa_t = 2$. Right: exclusions in the $(\kappa_\lambda, \kappa_t)$ plane. The red region corresponds to parameters excluded at 95% CL with the observed data, whereas the dashed black line and the blue areas correspond to the expected exclusions and the 68 and 95% bands (light and dark respectively). Isolines of the product of the theoretical cross section and branching fraction for $\mathrm{HH} \to \mathrm{b\bar{b}VV} \to \mathrm{b\bar{b}}\ell\nu\ell\nu$ are shown as dashed-dotted lines. The diamond marker indicates the prediction of the SM. All theoretical predictions are extracted from Refs. [68–73].

4. Chapter
Perspectives and conclusion

Before presenting our conclusions, we will briefly discuss some possibilities for improvement that have been considered, as well as prospects for the near future. On the one hand, the results obtained in the $b\bar{b}\ell\nu\ell\nu$ final state are being combined with the analyses carried out by the CMS collaboration in the three other final states mentioned in Sec. 1.7 ($b\bar{b}\gamma\gamma$, $b\bar{b}b\bar{b}$ and $b\bar{b}\tau\tau$), as well as with the results obtained at $\sqrt{s} = 8$ TeV with LHC Run 1 data. Figure 4.1 shows the limits obtained for scalar resonances by the CMS experiment in the different final states probed so far, as a function of m_X. The combined sensitivity on Higgs boson pair production should significantly exceed that obtained in the single most sensitive final state, $b\bar{b}\gamma\gamma$. On the other hand, Run 2 data delivery by the LHC is ending this year. Run 2 will be followed by a year-long shutdown, which will provide us with the opportunity to analyse the whole dataset obtained at $\sqrt{s} = 13$ TeV, expected to amount to almost 150 fb^{-1}. For the analysis of that large dataset, a few improvements to the methodology presented in Chap. 3 that could be studied and implemented are listed in the following.

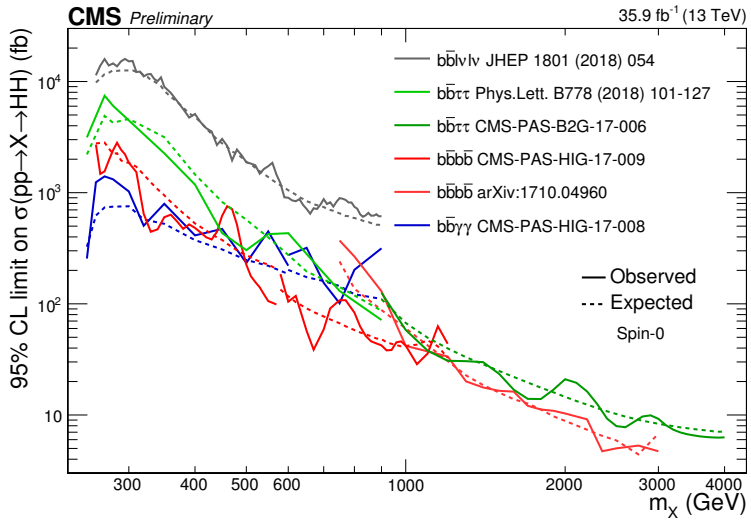

Figure 4.1. Expected (dashed) and observed (solid) limits on the product of the production cross section of spin-0 narrow-width resonances X with the branching fraction for $X \to HH$, obtained as a function of m_X by the CMS experiment in different final states. The limits are rescaled to the HH final state assuming SM Higgs boson branching ratios. The grey line corresponds to the result obtained in this work. Figure taken from Ref. [265].

Event selection

As explained in Sec. 2.3.11, the identification algorithm used for selecting electrons was unnecessarily restrictive. Using a more efficient algorithm might bring the selection efficiency for the signals in the e^+e^- and $e^\pm\mu^\mp$ channels to almost the same level as in the $\mu^+\mu^-$ channel. If possible, emulating the triggers in the simulation would also suppress the need for the tight HLT-safe criteria used for identifying electrons. More generally, the identification and isolation criteria for electrons and muons could be loosened, since this analysis is not plagued by backgrounds due to jets faking leptons. Doing so will likely require to estimate the level of backgrounds—however small—from fake leptons using proven data-driven techniques (see e.g. Ref. [266]).

The use of dilepton trigger paths in this analysis was justified by the observation that the signals feature relatively low-p_T leptons. Considering single-lepton trigger paths, although with a higher threshold on the leading lepton, would allow us to relax the criteria applied on the sub-leading lepton. However, this requires extra care to avoid any double-counting of events selected by both paths, and should only bring a small improvement ($\lesssim 10\%$) to the selection efficiency. Other possibly suitable trigger paths require the presence of two leptons and one or several jets or b-tagged jets. Their lepton p_T thresholds are impressively low (going as low as 8 GeV), but their high thresholds on the p_T of the jets ($\mathcal{O}(300\text{GeV})$) limits their relevance to our situation.

The working point used for the b tagging algorithm (see Sec. 2.3.7) was chosen to balance efficiency and purity, but has not been rigorously optimised. Using a looser requirement in the $e^\pm\mu^\mp$ channel might increase the sensitivity to the signal since that channel is dominated by the irreducible $t\bar{t}$ background, so that the signal-to-background ratio is relatively unaffected by the b tagging requirements. In addition, completely avoiding the use of b tagging in that channel would remove one of the leading sources of systematic uncertainty (see Tab. 3.6). Yet, doing so will increase the contamination from lesser backgrounds such as DY plus light jets, tW or VV, and it is hence necessary to precisely quantify the potential sensitivity gains.

The estimation method of the DY background, presented in Sec. 3.2, was designed to be general and provide us with predictions for the DY background in every possible kinematic distribution. Strictly speaking, one could dispense with this requirement and tailor the method for a few key distributions, such as the shape of the ANN classifier used for signal extraction. This could be done by computing the flavour fractions F_{kl} not in bins of the score of a BDT, but directly in the bins of the chosen distributions. The ultimate goal of designing a completely data-driven estimation method of all the backgrounds in this analysis (chiefly $t\bar{t}$) is extremely challenging, given that the main backgrounds are irreducible and that the signals do not feature any clear resonance standing out over a smooth background distribution.

Other decay channels of the Higgs boson pair could be considered, in particular the $HH \to b\bar{b}W(\ell\nu)W(jj)$ channel. While its branching ratio is higher than that of the channel chosen in this work (see Tab. 3.1), a difficulty arises from the combinatorial ambiguity in the assignment of selected jets to the decay of either Higgs bosons. In addition, triggering on the signal in this channel is challenging, given that leptons in the signal have low p_T, and that single-lepton trigger paths have thresholds of $\mathcal{O}(45\text{GeV})$.

Preliminary studies have shown that with this single-lepton channel, a sensitivity to HH comparable to that of the dilepton channel studied in this work can be attained [153,154].

Multivariate classifiers

We suspect that there is some latitude to improve the performance of the multivariate classifiers used in this analysis. To begin with, new input variables can be added on top of the eight kinematical quantities used to build the classifiers. Provided the simulated samples are large enough so that the extra information provided by these variables can be used efficiently, this should ameliorate the discrimination between signals and backgrounds. Since the main backgrounds are irreducible, all the relevant information is in principle contained in the kinematics (p_{T}, η and ϕ) of the reconstructed particles. If we remove an irrelevant overall azimuthal angle, this represents a reasonable total of 13 input variables.

A major difficulty in training multivariate classifiers resides in choosing values of parameters that have to be specified *a priori*. In the case of BDTs, one could vary the maximum depth of the individual trees, or the number of trees in the ensemble. With ANNs, both the structure of the network and the loss minimisation algorithm depend on a large number of these so-called *hyperparameters*. There is no general rule for choosing the values of all these parameters, and they are very often tuned by trial-and-error based on some *ad-hoc* objective (as was done in Sec. 3.3). Since these parameters can have a large impact on the convergence of the training procedure and the performance of the classifier, *hyperparameter optimisation* constitutes an active area of research. By quantifying the performance of the classifier in terms of a single objective (cost), the problem can be phrased as that of minimising a noisy, costly and discontinuous function over the multi-dimensional space of hyperparameters. The function is noisy because the training involves random components (e.g. the initial values of the weights in an ANN), and because it is evaluated on a sample of finite size. It is also costly, since evaluating the function means going through a complete training and evaluation of a multivariate classifier. Finally, it is discontinuous since the space of parameters can be partly categorical. The simple method of drawing random points in the parameter space has been shown to find good parameter values significantly faster than an exhaustive grid-based sampling [267]. A faster convergence can be attained by attempting to model the objective function, e.g. using Gaussian processes [268,269], so that previous evaluations of the function provide some insight on where the next point should be sampled[1].

We have attempted to implement an hyperparameter optimisation method for the parameterised classifier of Sec. 3.3 in the nonresonant case, using the `hyperopt` package [270]. The objective to be minimised was defined as the expected asymptotic limit obtained using the binned shape of the classifier score (without systematic uncertainties), averaged over the previously used set of nonresonant signals. For each set of hyperparameters proposed by the optimisation algorithm, we have trained two classifiers on independent subsamples, evaluated them on yet other independent samples, and averaged the two results. This *cross validation* procedure somewhat reduces the variance of the evaluated

[1] These optimisation algorithms often have hyperparameters of their own.

cost. The space of hyperparameters can be defined, for instance, by the following range of possible choices for the ANN structure and training procedure:

- Number of neurons per layer: 20, 50, or 100
- Number of hidden layers: 3, 4, or 5
- Dropout layer: yes or no
  - If yes, dropout rate $\in [0, 1]$ (sampled uniformly)
- Batch normalisation: yes or no
- Batch size: 512, 1024, 2048, 4096 or 8192
- Learning rate $\in [10^{-5}, 10^{-3}]$ (sampled uniformly in log space)

After 60 evaluations of the objective, a set of parameters was found that ameliorated the expected limit on the SM signal by about 10%. Figure 4.2 shows the average expected limits as a function of the learning rates and batch sizes sampled by the optimisation algorithm. The set of considered hyperparameters might be extended by including e.g. different choices of input variables, other parameters of the minimisation procedure, training stopping criteria, regularisation parameters, . . . , or refined to a lower-dimensional space by removing configurations that never lead to a competitive limit. In conclusion, for the next iteration of this analysis we can recommend the use of hyperparameter optimisation techniques based on a physics-driven objective.

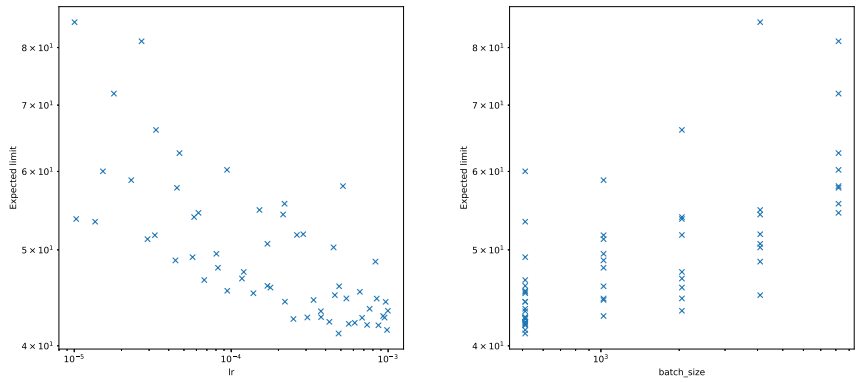

Figure 4.2. Expected limits obtained without systematic uncertainties using the binned distribution of the score of parameterised classifiers trained on the set of nonresonant signals defined in Sec. 3.3, and averaged over that set, as a function of the learning rate (left) and the batch size (right) used for the training of the classifiers. While there is a clear trend, in this case, that larger training rates and smaller batch sizes yield better-performing classifiers, that conclusion can by no means be generalised.

Two issues stand in the way of further improving the sensitivity of the analysis using a more powerful classifier. First, the sensitivity does not only depend on the level of discrimination between the signals and the backgrounds, but also on the impact of systematic uncertainties on the shape of the classifier score. Second, when including too much information as input to the classifier, it can become increasingly difficult to define

control regions in which to validate the agreement between data and simulation of the classifier score distribution. These problems are related to the machine-learning field of *domain adaptation*. In Ref. [271] a method was proposed that could address both of the quoted issues, based on the notion of domain-adversarial training first introduced in Ref. [272]. The goal is to make the distribution of the classifier score independent of a given variable, which can be e.g. a nuisance parameter used to model a systematic uncertainty, or the property of being present or not in a signal-free control region. This can be achieved by building a second ANN, the "adversary", that tries to retrieve the true value of that variable using as only information the score of the initial classifier. The loss function of the classifier can then be modified so that the performance of the adversary acts as penalty during the training, thus compelling the classifier score to provide no information on the value of the variable from which it should be independent.

We have shown that this technique could be implemented in our analysis, for the purpose of building a classifier that has the same distribution in the signal region defined by $75 \leq m_{jj} < 140$ GeV as in the complementary region. For this exploration, we have considered a non-parameterised classifier trained to recognise solely the SM HH signal from the $t\bar{t}$ background. Obviously, the condition of independence is only imposed on the background sample. Figure 4.3 shows that using adversarial training, the shape of the classifier score evaluated on the background can become almost identical in the signal and control regions. However, we have found that training the adversary classifiers is extremely delicate and that obtaining the desired behaviour requires a large amount of trial-and-error. Furthermore, it is unclear whether applying the technique to nuisance parameters would bring any significant gain in the overall sensitivity of the analysis. Hence, more work is needed to clarify the relevance of the proposed method for this analysis.

Signal definition and interpretation of the results

While in this work we have followed a conservative approach in the optimisation and interpretation of the analysis in terms of nonresonant New Physics effects, based on the discussion of Sec. 1.5 we believe that these points should be refined as the sensitivity to HH production increases. Since there is currently no consensus on which of the linear or nonlinear EFT parameterisations should be favoured, it seems reasonable to call for an agnostic approach in which both models are considered. By relying on reweighting and morphing techniques, the difficulty of probing the EFT parameter space and taking into account effects on experimental acceptance and sensitivity can be efficiently addressed. As explained in Sec. 1.5 and 1.7, ideally these measurements should be combined with the results obtained on single Higgs production modes, so that a consistent picture of our knowledge of all Higgs boson couplings can be extracted out of LHC data.

For what concerns the resonant production of Higgs boson pairs, there are numerous difficulties in going beyond the narrow-width assumption used in this analysis. Relaxing this hypothesis implies taking into account interference effects between amplitudes involving the new heavy state X and the SM amplitude. These effects depend on the considered models (see e.g. Ref. [72]), which may involve numerous parameters—some also affecting the SM HH background. Since event reweighting is largely inefficient as soon as resonances of unknown mass are involved, it is unclear on how to best probe the

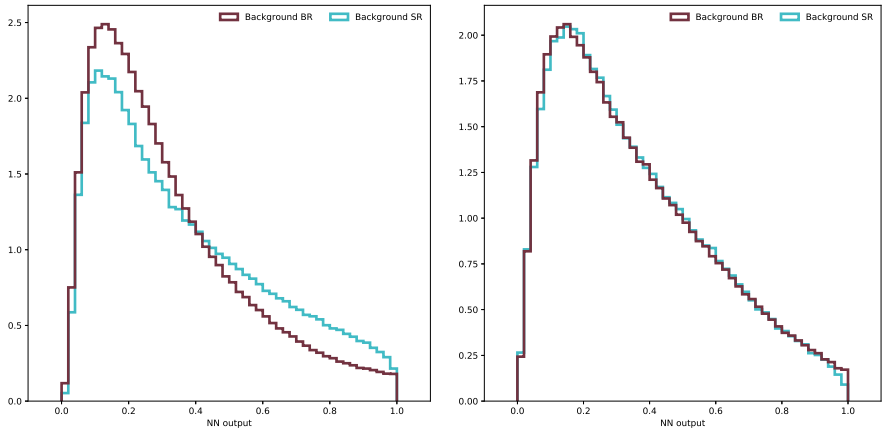

Figure 4.3. Distribution of the score of ANNs evaluated on the $t\bar{t}$ background, in the signal region (SR) defined by $75 \leq m_{jj} < 140$ GeV, and in the background region (BR) defined as the complement of the SR. The ANNs are trained to discriminate the HH signal in the SM hypothesis from the $t\bar{t}$ background. On the left, where the ANN used is a regular one, the distributions of background scores are significantly different in the SR and BR. On the right, the technique of adversarial training is successfully employed to force the shape of the classifier output to be nearly identical in the SR and BR. Although the signal distribution is not shown here, the performances of both ANNs are equivalent.

parameter space of these models, and it will be likely necessary to settle for a limited set of benchmark models and parameter points. In any case, there are two separate issues that need to be considered:

1. Making sure experimental analyses are not blind to deviations from SM predictions that are not explicitly searched for, and
2. Interpreting the results to constrain the parameter space of arbitrary models.

While the second point is certainly the most challenging given the above discussion, we would argue that the first point represents the most pressing one.

Prospects for the HL-LHC

The expected sensitivity to Higgs pair production in the SM at the end of the HL-LHC programme, in the different channels considered by CMS, has been estimated using different approaches. In Refs. [153, 155, 156] the HH and background processes were simulated using simplified detector response functions, or using the DELPHES fast parametric simulation framework [273]. Conversely,in Ref. [154] the sensitivity at the HL-LHC was projected based on preliminary results obtained at $\sqrt{s} = 13$ TeV using data collected by CMS in 2015, by scaling the expected yields to 3000 fb^{-1} and using different sets of assumptions on the size of the systematic uncertainties. While some assumptions taken in these projections are optimistic, other are pessimistic, so that their effects on the quoted sensitivity should average out. The expected uncertainties on the HH production cross section in the different final states are shown on Fig. 4.4. Clearly, the four channels probed—including the one studied in this work—are complementary,

and only by combining these and possibly further channels can we hope to reach a meaningful precision on double Higgs production.

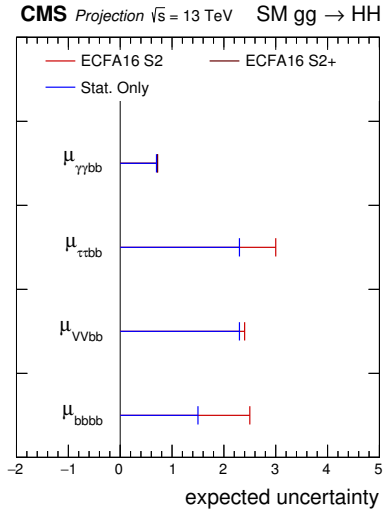

Figure 4.4. | Expected uncertainty on $\mu = \sigma_{HH}/\sigma_{HH}^{SM}$ in various final states with 3000 fb^{-1}, computed by scaling preliminary results obtained with 2.3 fb^{-1} [154]. In the "Stat. Only" scenario, all systematic uncertainties have been removed. The "ECFA16 S2(+)" correspond to different sets of assumptions about the size of the theoretical and systematic uncertainties at the end of the HL-LHC data taking, in 2035.

Conclusions

To summarise, we have analysed 35.9 fb^{-1} of data collected in 2016 by the CMS detector at the LHC at $\sqrt{s} = 13$ TeV, which (at the time) represented the largest amount of data ever collected by a hadron collider, at the highest centre-of-mass energy ever achieved by any particle collider. We have used these data to put SM predictions for the production of Higgs boson pairs (HH) to the test, in the case where the Higgs bosons decay as $H \to b\bar{b}$ on the one hand and $H \to VV(\to \ell\nu\ell\nu)$ on the other hand (with $\ell = e, \mu, \tau$). Observing and characterising Higgs boson pair production is the most direct method for measuring the trilinear Higgs boson self-coupling. The self-coupling is a crucial parameter of the mechanism of spontaneous symmetry breaking at the heart of the SM, whose value can be predicted but which has never been directly measured. Moreover, HH production is highly sensitive to possible indirect effects from New Physics present at higher energy scales. We have targeted some of these effects by searching for deviations from SM predictions for HH as a function of modifications to the Higgs boson self-coupling (κ_λ) and the coupling between the Higgs boson and the top quark (κ_t), the heaviest fermion in the SM. In addition, we have searched for new, heavy states decaying to pairs of Higgs bosons, a possibility predicted by numerous models that have been developed in attempts to extend the SM.

The sample of data used in this work consists of events with two charged leptons (electrons or muons) and two b-tagged jets. These events were divided into three

categories with different background compositions, depending on the flavour of the selected leptons. We have computed the contamination from SM background processes in part by relying on a detailed simulation of these processes and of the CMS detector, as for the main background due to top quark pair production, in part using a dedicated data-driven technique. The latter technique has been designed to estimate the contribution of Z/γ^* plus heavy flavour jet production to the selected dataset, using a sample of events where the b tagging requirements have been relaxed. In order to increase the sensitivity of our analysis, we have built multivariate classifiers designed to discriminate signal from background events. These classifiers make use of modern machine-learning technology and employ the novel approach of parameterised learning, used here for the first time in an analysis of LHC data, to ensure an optimal sensitivity over the range of considered signal hypotheses.

We have confronted the distributions of the classifier score in data with the background predictions in three different ranges of the invariant mass of the pair of selected b-tagged jets. One of these regions was designed to be enriched in signal, taking advantage of the $H \to b\bar{b}$ decay yielding pairs of b-jets with invariant mass close to the true Higgs boson mass. The remaining two regions were used to check the agreement between data and predictions and to constrain the nuisance parameters associated with systematic uncertainties in the modelling of the background processes. A good agreement between data and SM predictions has been found, hence we have set limits at 95% CL on the cross section of Higgs boson pair production in the considered final state. These limits indicate the largest possible signal cross sections that are statistically compatible with the observed data. Since the sensitivity to HH production strongly depends on the hypothesised production mechanism, the limits are computed as a function of the parameters governing the signal properties. In the hypothesis of resonant production of Higgs boson pairs, we give limits on the product of the production cross section for narrow-width spin-0 and spin-2 resonances and the branching fraction for $X \to HH \to b\bar{b}\ell\nu\ell\nu$ as a function of the mass of the resonance. For nonresonant HH production, limits have been set on the cross section for $HH \to b\bar{b}\ell\nu\ell\nu$ as a function of κ_λ and κ_t, thereby constraining the possible values of these couplings. In particular, in the hypothesis of SM HH production, we have excluded production rates larger than 79 times the SM prediction, which represents a significant improvement of a factor of five compared to previous results obtained in the same final state. Our results are being combined with those obtained by analyses targeting different decay channels, thereby increasing our overall sensitivity to the crucial HH process. Finally, we have given several indications for possible improvements, in the hope that they be useful to those who may attempt a similar analysis with an even larger amount of data.

A. Appendix

Event selection efficiencies

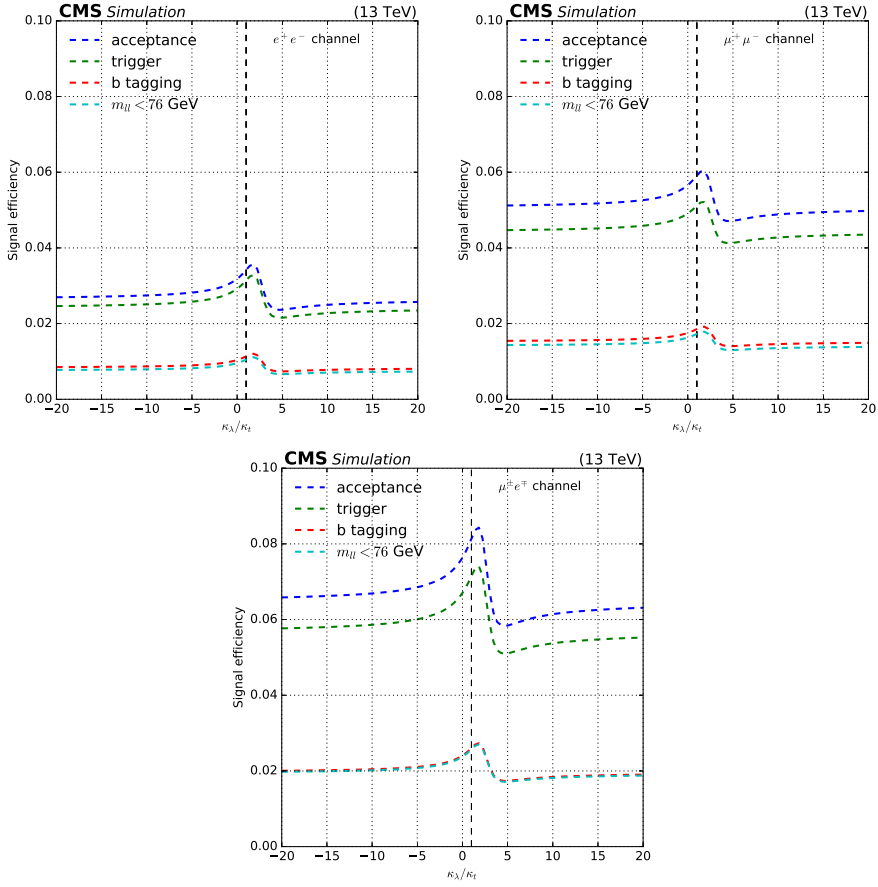

Figure 1.1. Selection efficiency of nonresonant signal events as a function of κ_λ/κ_t, for the different steps of our selection described in Sec. 3.1.2, in the e^+e^- (top left), $\mu^+\mu^-$ (top right) and $\mu^{\pm}e^{\mp}$ (bottom) channels.

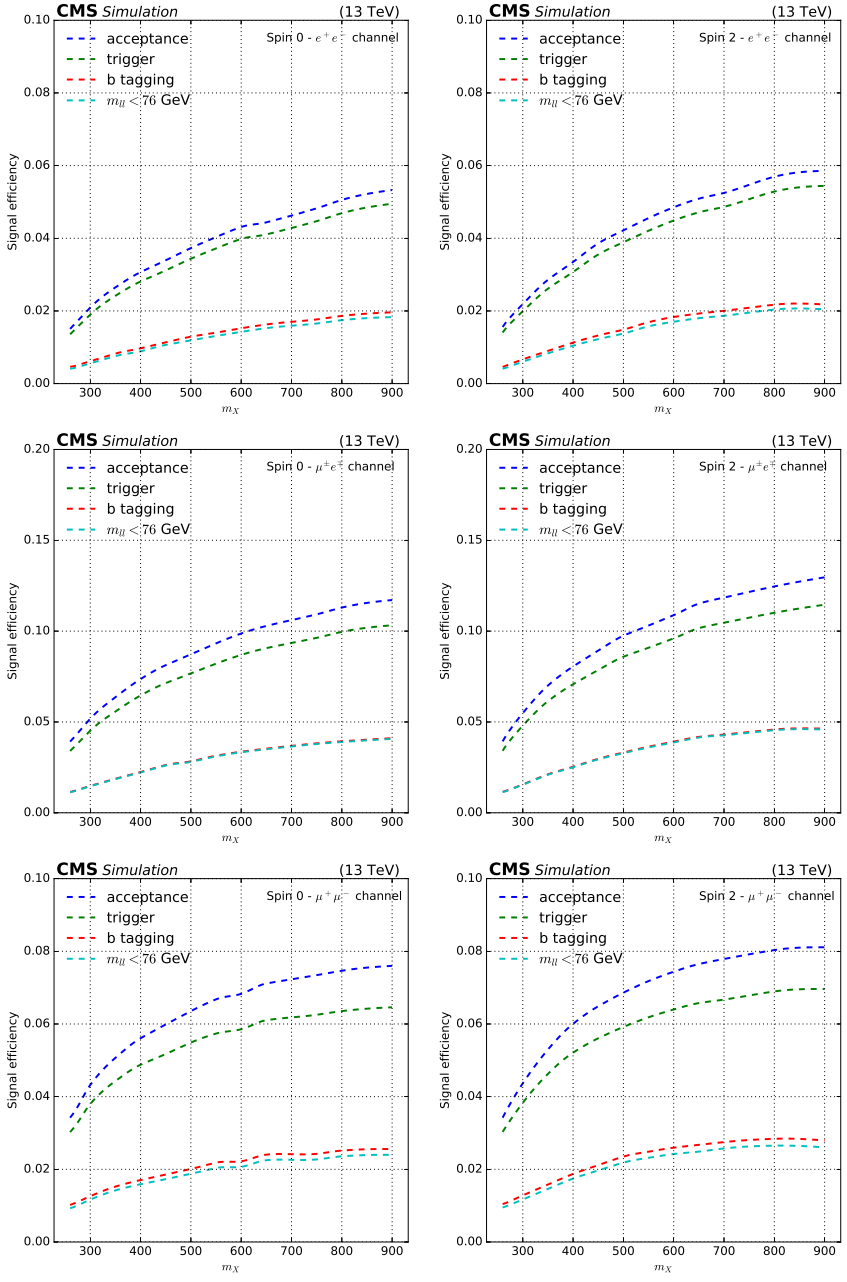

Figure 1.2. | Selection efficiency of resonant signal events as a function of m_X, for the different steps of our selection described in Sec. 3.1.2. The spin-0 and spin-2 signals are shown on the left and on the right, respectively, for the e^+e^- (top), $\mu^\pm e^\mp$ (middle) and $\mu^+\mu^-$ (bottom) channels. All efficiencies are given with respect to the total signal samples, which include the possibility of W bosons decaying to τ leptons. For the "acceptance" step we apply the listed requirements on lepton identification, isolation, p_T and η, and jet p_T, η and angular separation from the leptons. The efficiency of the trigger step is computed after having applied the previous requirements.

Appendix

B.

Classifier input variables

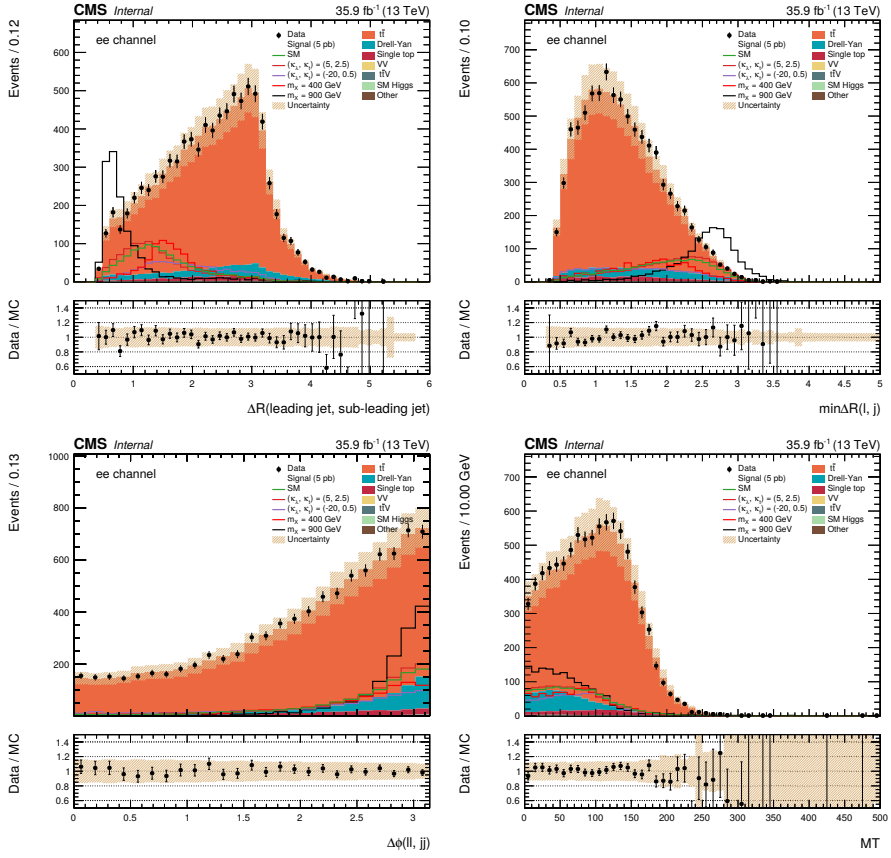

Figure 2.1. | Distributions of the remaining kinematic input variables, not shown in Chap. 3, of the ANNs used to discriminate signal from background events, in the e^+e^- channel. All selection criteria described in Sec. 3.1.2 are applied. Top left: pseudo-angular distance between the b-tagged jets. Top right: minimal pseudo-angular distance between the leptons and jets. Bottom left: $\Delta\phi$ between the dilepton and dijet systems. Bottom right: $m_T = \sqrt{2p_T^{\ell\ell} p_T^{\text{miss}}[1-\cos\Delta\phi(\ell\ell, \vec{p}_T^{\text{miss}})]}$. Shaded bands show pre-fit systematic uncertainties.

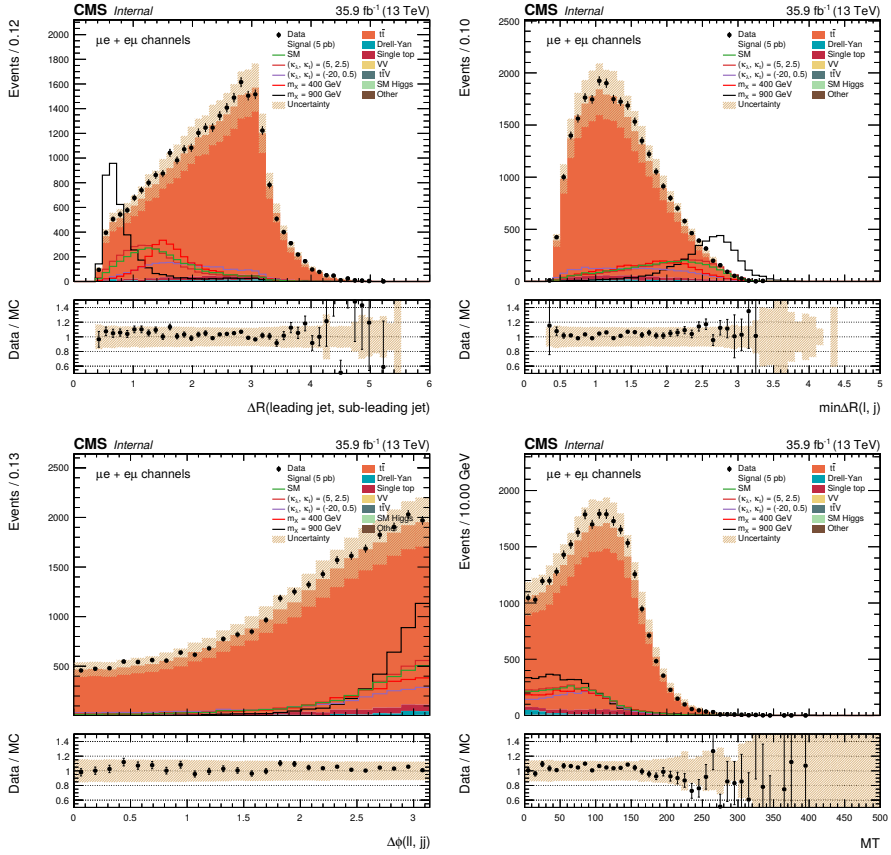

Figure 2.2. Distributions of the remaining kinematic input variables, not shown in Chap. 3, of the ANNs used to discriminate signal from background events, in the $e^{\pm}\mu^{\mp}$ channel. All selection criteria described in Sec. 3.1.2 are applied. Top left: pseudo-angular distance between the b-tagged jets. Top right: minimal pseudo-angular distance between the leptons and jets. Bottom left: $\Delta\phi$ between the dilepton and dijet systems. Bottom right: $m_{\mathrm{T}} = \sqrt{2p_{\mathrm{T}}^{\ell\ell} p_{\mathrm{T}}^{\mathrm{miss}}[1-\cos\Delta\phi(\ell\ell,\vec{p}_{\mathrm{T}}^{\mathrm{miss}})]}$. Shaded bands show pre-fit systematic uncertainties.

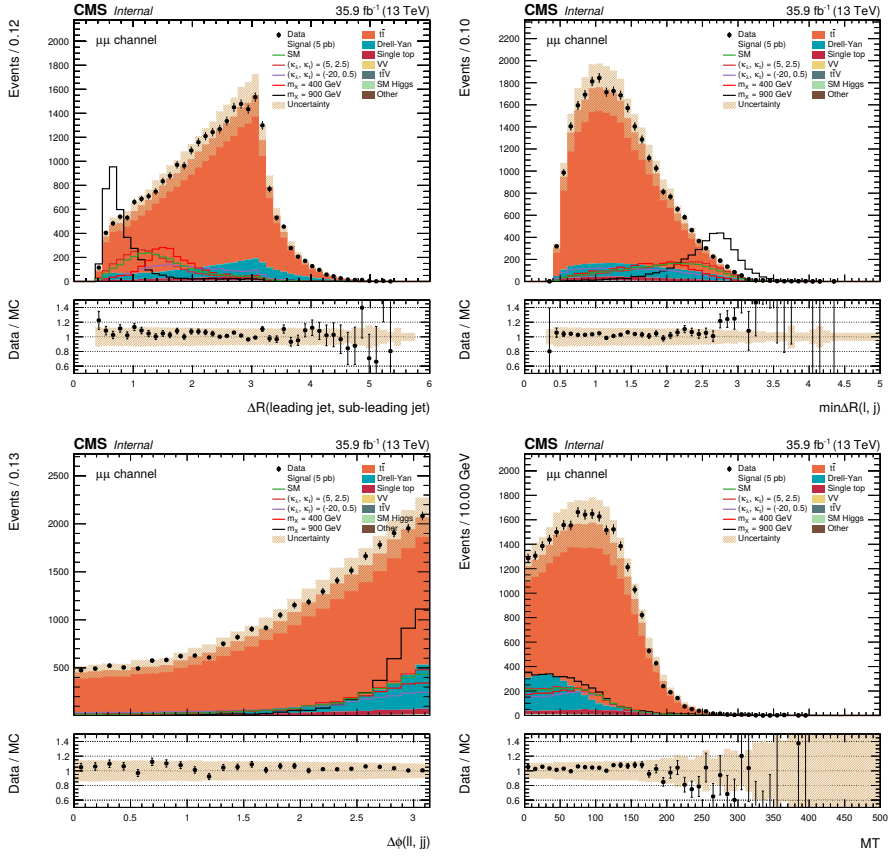

Figure 2.3. Distributions of the remaining kinematic input variables, not shown in Chap. 3, of the ANNs used to discriminate signal from background events, in the $\mu^+\mu^-$ channel. All selection criteria described in Sec. 3.1.2 are applied. Top left: pseudo-angular distance between the b-tagged jets. Top right: minimal pseudo-angular distance between the leptons and jets. Bottom left: $\Delta\phi$ between the dilepton and dijet systems. Bottom right: $m_{\mathrm{T}} = \sqrt{2p_{\mathrm{T}}^{\ell\ell} p_{\mathrm{T}}^{\mathrm{miss}}[1-\cos\Delta\phi(\ell\ell,\vec{p}_{\mathrm{T}}^{\mathrm{miss}})]}$. Shaded bands show pre-fit systematic uncertainties.

Acronyms

μGT micro global trigger **65**

2HDM 2 Higgs doublet model **47**

ANN artificial neural network **91**, 92, 93, 115–118, 121, 122, 124, 127, 136–140, 145–147

APD avalanche photodiode **60**

ASIC application-specific integrated circuit 63

AVR adaptive vertex reconstruction **75**, 76

BCM beam conditions monitor **80**

BCMS batch compression merging and splitting **51**, 54

BDT boosted decision tree 7, **76**, **89**, **108**, 109, 110, 114, 136, 137

BMTF barrel muon track finder 65

BSM beyond the SM **35**, 36, 38–41, 45, 49, 50, 132

BU builder unit PC **65**

BX bunch crossing **63**, 64

cdf cumulative distribution function **97**

CERN European Organisation for Nuclear Research 51, 52, 54, 55, 66

CHS charged-hadron subtraction **77**, 78

CKM Cabibbo–Kobayashi–Maskawa **23**, 34

CL confidence level 49, **95**, 96, 104, 122, **131**, 132, 133, 142

CMS Compact Muon Solenoid 49, 51, 53, **55**, 56–59, 62–67, 76, 80, 85, 87, 99, 101, 132, 135, 140–142

CMSSW CMS software 65

cMVAv2 combined multivariate algorithm **76**, 88, 104, 105

CSC cathode strip chamber 62, **63**, 64, 68

CSVv2 combined secondary vertex **75**, 76, 88

CTF combinatorial track finder **67**, 68, 69, 78

DAQ data acquisition **63**

DGLAP Dokshitzer–Gribov–Lipatov–Altarelli–Parisi **26**

DT drift tube **62**, 63, 64, 68, 80

DY Drell–Yan **101**, 102, 104, 106, 108–114, 116, 126–128, 130, 136

EB electromagnetic calorimeter barrel 59–61, 72

ECAL electromagnetic calorimeter **59**, 60, 63, 64, 68–73, 79, 86

EE electromagnetic calorimeter endcap 59, 60, 72

EFT effective field theory **35**, 36–46, 48, 50, 139, 150

EMTF endcap muon track finder 65, 82

FED front-end driver **65**

FEROL front-end readout optical link **65**

FPGA field-programmable gate array 63

FSR final-state radiation 24, 27

FU filter unit PC **65**

GM Georgi–Machacek **47**

GMT global muon trigger **65**

GoF goodness-of-fit **42**

GPGPU general-purpose graphical processing unit **116**

GSF Gaussian sum filter **68**, 70, 71, 85

GUT grand unified theory **35**

HB hadron calorimeter barrel 60, 61

HCAL hadron calorimeter **60**, 63, 64, 69, 70, 72, 73, 79, 86

HE hadron calorimeter endcap 60, 61

HEFT Higgs EFT 31–33, 40

HF forward hadron calorimeter 60, 61, 64, 69, 74, 80

HL-LHC High-Luminosity Large Hadron Collider 55, 99, 140, 141

HLT high-level trigger **63**, 65, 78, 79, 82, 85, 103, 105

HO outer hadron calorimeter 60, 61

HPD hybrid photodiode **61**

ID identification **70**, 71–74, 82, **105**

IP impact parameter **74**, 87, 88

IP interaction point **52**, 53, 55–60, 74, 78

IPS impact parameter significance **74**, 75, 76

ISR initial-state radiation 26, 27

IVF inclusive vertex finder **75**, 76

JBP jet b probability **75**, 76

JEC jet energy correction **86**, 87

JER jet energy resolution **87**

JP jet probability **75**, 76

KF Kalman filter **67**, 68

KK Kaluza–Klein **48**, 102

KLN Kinoshita–Lee–Nauenberg **26**

L1 level 1 trigger **63, 64**, 65, 78, 79, 82

L1A L1 accept **64**, 65

LHC Large Hadron Collider **11**, 12, 35, **51**, 52–56, 60, 63, 65, 80, 116, 139, 141, 142

LO leading order **16**, 26–28, 30–34, 38, 39, 41, 43, 44, 102, 125

LS luminosity section **65**, 69, 80, **101**

ME matrix element **15**, 28–30

ML machine learning **89**, 117

MLE maximum-likelihood estimate **95**

MLP multilayer perceptron **76**, **89**, **91**

MPI multiple parton interactions 27, 29

MSSM minimal supersymmetric standard model **47**

NLO next-to-leading order **16**, 27–30, 32, 33, 39, 50, 101, 102, 125

NNLL next-to-next-to-leading log 31–33, 102

NNLO next-to-next-to-leading order **16**, 28, 30–33, 39, 40, 50, 101, 102

NWA narrow-width approximation **28**, 34

OMTF overlap muon track finder 65

OOT out of time **77**

PCA point of closest approach **66**, 69, 74

PCC pixel cluster counting **80**

PD primary dataset **78**

PDF parton distribution function **26**, 31–33, 44, 102, 122, 126

pdf probability density function **89**, 95, 96, 114

PF particle flow **66**, 69–73, 77, 78, 86, 105

PLT pixel luminosity telescope **80**

pmf probability mass function **129**

PMNS Pontecorvo–Maki–Nakagawa–Sakata **23**

PMT photomultiplier tube **61**

PS Proton Synchrotron **51**, 52

PSB Proton Synchrotron Booster **51**, 52

PV primary vertex **69**, 74, 75, 79

QCD quantum chromodynamics **23, 24**, 25, 26, 29–32, 34, 39, 101, 102

QED quantum electrodynamics **19**, 21, 37

QFT quantum field theory **13**, 14

RF radio frequency **52**

ROC receiver operating characteristics 117, 118, 122

RPC resistive plate chamber 62, **63**, 64, 68

RS Randall–Sundrum **47**, 48

RU readout unit PC **65**

SC super cluster **68**, 86

SET soft electron tagger **75**, 76

SF scale factor 81, 82, 85, 87, 88

SGD stochastic gradient descent **92**

SI International System of units **14**

SM standard model **11**, 12, **13**, 14, 17, 18, 23, 24, 31, 33–50, 99–102, 125, 132, 133, 135, 138–142, 149

SMT soft muon tagger **75**, 76

SPS Super Proton Synchrotron **51**

SV secondary vertex 74–76, 88

T&P tag and probe **71**, **81**, 82, 83, 85, 88, 104, 122

TCDS trigger control and distribution system **64**

TEC tracker endcap **58**, 59

TIB tracker inner barrel **58**, 59

TID tracker inner disk **58**, 59

TOB tracker outer barrel **58**, 59

TT trigger tower **64**

UE underlying event 27, **29**

VdM van der Meer scan **55**, 80

VEV vacuum expectation value **21**, 22, 31, 34, 37, 47

VPT vacuum phototriode **60**

WED warped extra dimension 47, 48, **102**

WLS wavelength shifting fibre **61**

WP working point **70**, **76**, 105

References

[1] M. Peskin et al., "An Introduction to Quantum Field Theory". Advanced book classics. Avalon Publishing, 1995.

[2] F. Mandl et al., "Quantum Field Theory". Wiley, 2013.

[3] M. Kramer et al., "Large Hadron Collider Phenomenology". Scottish Graduate Series. CRC Press, 2004.

[4] C. Smith, "Introduction to the Standard Model", 2014. Lecture notes.

[5] Particle Data Group Collaboration, "Review of Particle Physics", *Chin. Phys. C* **40** (2016), no. 10, 100001. doi:10.1088/1674-1137/40/10/100001.

[6] E. Di Valentino et al., "Cosmological limits on neutrino unknowns versus low redshift priors", *Phys. Rev. D* **93** (2016), no. 8, 083527. doi:10.1103/PhysRevD.93.083527, arXiv:1511.00975.

[7] E. Fermi, "An attempt of a theory of beta radiation", *Z. Phys.* **88** (1934) 161 – 177. doi:10.1007/BF01351864.

[8] C. S. Wu et al., "Experimental Test of Parity Conservation in Beta Decay", *Phys. Rev.* **105** (1957) 1413–1415. doi:10.1103/PhysRev.105.1413.

[9] Gargamelle Collaboration, "Observation of Neutrino Like Interactions Without Muon Or Electron in the Gargamelle Neutrino Experiment", *Phys. Lett. B* **46** (1973) 138 – 140. doi:10.1016/0370-2693(73)90499-1.

[10] S. L. Glashow, "The renormalizability of vector meson interactions", *Nucl. Phys.* **10** (1959) 107 – 117. doi:https://doi.org/10.1016/0029-5582(59)90196-8.

[11] S. L. Glashow, "Partial Symmetries of Weak Interactions", *Nucl. Phys.* **22** (1961) 579 – 588. doi:10.1016/0029-5582(61)90469-2.

[12] A. Salam et al., "Weak and electromagnetic interactions", *Il Nuovo Cimento (1955-1965)* **11** (1959), no. 4, 568 – 577. doi:10.1007/BF02726525.

[13] S. Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264 – 1266. doi:10.1103/PhysRevLett.19.1264.

[14] F. Englert et al., "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (1964) 321 – 323. doi:10.1103/PhysRevLett.13.321.

[15] P. W. Higgs, "Broken symmetries, massless particles and gauge fields", *Phys. Lett.* **12** (1964) 132 – 133. doi:10.1016/0031-9163(64)91136-9.

[16] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508 – 509. doi:10.1103/PhysRevLett.13.508.

[17] G. S. Guralnik et al., "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.* **13** (1964) 585 – 587. doi:10.1103/PhysRevLett.13.585.

[18] 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", *Phys. Rev. Lett.* **114** (2015) 191803. doi:10.1103/PhysRevLett.114.191803, arXiv:1503.07589.

[19] D. J. Gross et al., "Ultraviolet Behavior of Non-Abelian Gauge Theories", *Phys. Rev. Lett.* **30** (1973) 1343–1346. doi:10.1103/PhysRevLett.30.1343.

[20] H. D. Politzer, "Reliable Perturbative Results for Strong Interactions?", *Phys. Rev. Lett.* **30** (1973) 1346–1349. doi:10.1103/PhysRevLett.30.1346.

[21] M. Cacciari et al., "The anti-k_T jet clustering algorithm", *JHEP* **04** (2008) 063. doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[22] M. Cacciari et al., "FastJet user manual", *Eur. Phys. J. C* **72** (2012) 1896. doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[23] G. Altarelli et al., "Asymptotic Freedom in Parton Language", *Nucl. Phys. B* **126** (1977) 298–318. doi:10.1016/0550-3213(77)90384-4.

[24] Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics.", *Sov. Phys. JETP* **46** (1977) 641–653. [Zh. Eksp. Teor. Fiz.73,1216(1977)].

[25] V. N. Gribov et al., "Deep inelastic ep scattering in perturbation theory", *Sov. J. Nucl. Phys.* **15** (1972) 438–450. [Yad. Fiz.15,781(1972)].

[26] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040. doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[27] A. Buckley et al., "LHAPDF6: parton density access in the LHC precision era", *Eur. Phys. J. C* **75** (2015) 132. doi:10.1140/epjc/s10052-015-3318-8, arXiv:1412.7420.

[28] T. Kinoshita, "Mass Singularities of Feynman Amplitudes", *J. Math. Phys.* **3** (1962), no. 4, 650–677. doi:10.1063/1.1724268.

[29] T. D. Lee et al., "Degenerate Systems and Mass Singularities", *Phys. Rev.* **133** (Mar, 1964) B1549–B1562. doi:10.1103/PhysRev.133.B1549.

[30] A. Buckley et al., "General-purpose event generators for LHC physics", *Phys. Rept.* **504** (2011) 145–233. doi:10.1016/j.physrep.2011.03.005, arXiv:1101.2599.

[31] S. Höche, "Introduction to parton-shower event generators", in *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders (TASI 2014): Boulder, Colorado, June 2-27, 2014*, pp. 235–295. 2015. arXiv:1411.4085.

[32] G. P. Lepage, "A New Algorithm for Adaptive Multidimensional Integration", *J. Comput. Phys.* **27** (1978) 192. doi:10.1016/0021-9991(78)90004-9.

[33] H. Murayama et al., "HELAS: HELicity amplitude subroutines for Feynman diagram evaluations", technical report, 1992.

[34] F. A. Berends et al., "Recursive Calculations for Processes with n Gluons", *Nucl. Phys. B* **306** (1988) 759–808. doi:10.1016/0550-3213(88)90442-7.

[35] S. Frixione et al., "Three jet cross-sections to next-to-leading order", *Nucl. Phys. B* **467** (1996) 399–442. doi:10.1016/0550-3213(96)00110-1, arXiv:hep-ph/9512328.

[36] S. Catani et al., "A General algorithm for calculating jet cross-sections in NLO QCD", *Nucl. Phys. B* **485** (1997) 291–419. doi:10.1016/S0550-3213(96)00589-5, 10.1016/S0550-3213(98)81022-5, arXiv:hep-ph/9605323. [Erratum: Nucl. Phys.B510,503(1998)].

[37] R. Frederix et al., "Automation of next-to-leading order computations in QCD: the FKS subtraction", *JHEP* **10** (2009) 003. doi:10.1088/1126-6708/2009/10/003, arXiv:0908.4272.

[38] T. Gleisberg et al., "Automating dipole subtraction for QCD NLO calculations", *Eur. Phys. J. C* **53** (2008) 501–523. doi:10.1140/epjc/s10052-007-0495-0, arXiv:0709.2881.

[39] P. Artoisenet et al., "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations", *JHEP* **03** (2013) 015. doi:10.1007/JHEP03(2013)015, arXiv:1212.3460.

[40] N. Davidson et al., "Universal Interface of TAUOLA Technical and Physics Documentation", *Comput. Phys. Commun.* **183** (2012) 821–843. doi:10.1016/j.cpc.2011.12.009, arXiv:1002.0543.

[41] T. Sjöstrand et al., "PYTHIA 6.4 physics and manual", *JHEP* **05** (2006) 026. doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[42] T. Sjöstrand et al., "A brief introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852. doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.

[43] M. Bahr et al., "Herwig++ Physics and Manual", *Eur. Phys. J. C* **58** (2008) 639–707. doi:10.1140/epjc/s10052-008-0798-9, arXiv:0803.0883.

[44] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **11** (2004) 040. doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.

[45] S. Frixione et al., "Matching NLO QCD computations with parton shower simulations: the POWHEG method", *JHEP* **11** (2007) 070. doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[46] S. Frixione et al., "Matching NLO QCD computations and parton shower simulations", *JHEP* **06** (2002) 029. doi:10.1088/1126-6708/2002/06/029, arXiv:hep-ph/0204244.

[47] J. Alwall et al., "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions", *Eur. Phys. J. C* **53** (2008) 473–500. doi:10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.

[48] S. Catani et al., "QCD matrix elements + parton showers", *JHEP* **11** (2001) 063. doi:10.1088/1126-6708/2001/11/063, arXiv:hep-ph/0109231.

[49] F. Krauss, "Matrix elements and parton showers in hadronic interactions", *JHEP* **08** (2002) 015. doi:10.1088/1126-6708/2002/08/015, arXiv:hep-ph/0205283.

[50] M. L. Mangano et al., "Matching matrix elements and shower evolution for top-quark production in hadronic collisions", *JHEP* **01** (2007) 013. doi:10.1088/1126-6708/2007/01/013, arXiv:hep-ph/0611129.

[51] R. Frederix et al., "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061. doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.

[52] N. D. Christensen et al., "FeynRules - Feynman rules made easy", *Comput. Phys. Commun.* **180** (2009) 1614–1641. doi:10.1016/j.cpc.2009.02.018, arXiv:0806.4194.

[53] F. Staub, "SARAH 4 : A tool for (not only SUSY) model builders", *Comput. Phys. Commun.* **185** (2014) 1773–1790. doi:10.1016/j.cpc.2014.02.018, arXiv:1309.7223.

[54] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079. doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[55] S. Alioli et al., "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043. doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[56] T. Gleisberg et al., "Event generation with SHERPA 1.1", *JHEP* **02** (2009) 007. doi:10.1088/1126-6708/2009/02/007, arXiv:0811.4622.

[57] W. Kilian et al., "WHIZARD: Simulating Multi-Particle Processes at LHC and ILC", *Eur. Phys. J. C* **71** (2011) 1742. doi:10.1140/epjc/s10052-011-1742-y, arXiv:0708.4233.

[58] O. Mattelaer, "On the maximal use of Monte Carlo samples: re-weighting events at NLO accuracy", *Eur. Phys. J. C* **76** (2016), no. 12, 674. doi:10.1140/epjc/s10052-016-4533-7, arXiv:1607.00763.

[59] A. Kalogeropoulos et al., "The SysCalc code: A tool to derive theoretical systematic uncertainties", (2018). arXiv:1801.08401.

[60] M. Botje et al., "The PDF4LHC Working Group Interim Recommendations", 2011. arXiv:1101.0538.

[61] S. Alekhin et al., "The PDF4LHC Working Group Interim Report", 2011. arXiv:1101.0536.

[62] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1–29. doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[63] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* **716** (2012) 30–61. doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[64] F. Maltoni et al., "Top-quark mass effects in double and triple Higgs production in gluon-gluon fusion at NLO", *JHEP* **11** (2014) 079. doi:10.1007/JHEP11(2014)079, arXiv:1408.6542.

[65] O. Éboli et al., "Twin Higgs-boson production", *Phys. Lett. B* **197** (1987), no. 1, 269–272. doi:https://doi.org/10.1016/0370-2693(87)90381-9.

[66] E. Glover et al., "Higgs boson pair production via gluon fusion", *Nucl. Phys. B* **309** (1988), no. 2, 282–294. doi:https://doi.org/10.1016/0550-3213(88)90083-1.

[67] T. Plehn et al., "Pair production of neutral Higgs particles in gluon-gluon collisions", *Nucl. Phys. B* **479** (1996) 46–64. doi:10.1016/0550-3213(96)00418-X, 10.1016/S0550-3213(98)00406-4, arXiv:hep-ph/9603205. [Erratum: Nucl. Phys.B531,655(1998)].

[68] D. de Florian et al., "Higgs Boson Pair Production at Next-to-Next-to-Leading Order in QCD", *Phys. Rev. Lett.* **111** (2013) 201801. doi:10.1103/PhysRevLett.111.201801, arXiv:1309.6594.

[69] D. de Florian et al., "Higgs pair production at next-to-next-to-leading logarithmic accuracy at the LHC", *JHEP* **09** (2015) 053. doi:10.1007/JHEP09(2015)053, arXiv:1505.07122.

[70] S. Borowka et al., "Full top quark mass dependence in Higgs boson pair production at NLO", *JHEP* **10** (2016) 107. doi:10.1007/JHEP10(2016)107, arXiv:1608.04798.

[71] S. Borowka et al., "Higgs Boson Pair Production in Gluon Fusion at Next-to-Leading Order with Full Top-Quark Mass Dependence", *Phys. Rev. Lett.* **117** (2016), no. 1, 012001. doi:10.1103/PhysRevLett.117.079901, 10.1103/PhysRevLett.117.012001, arXiv:1604.06447. [Erratum: Phys. Rev. Lett.117,no.7,079901(2016)].

[72] D. de Florian et al., "Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector", CERN Yellow Report CERN-2017-002-M, 2016. arXiv:1610.07922.

[73] J. Grigo et al., "Higgs boson pair production: top quark mass effects at NLO and NNLO", *Nucl. Phys. B* **900** (2015) 412–430. doi:10.1016/j.nuclphysb.2015.09.012, arXiv:1508.00909.

[74] M. Grazzini et al., "Higgs boson pair production at NNLO with top quark mass effects", (2018). arXiv:1803.02463.

[75] D. de Florian et al., "Differential Higgs Boson Pair Production at Next-to-Next-to-Leading Order in QCD", *JHEP* **09** (2016) 151. doi:10.1007/JHEP09(2016)151, arXiv:1606.09519.

[76] G. Heinrich et al., "NLO predictions for Higgs boson pair production with full top quark mass dependence matched to parton showers", *JHEP* **08** (2017) 088. doi:10.1007/JHEP08(2017)088, arXiv:1703.09252.

[77] S. Jones et al., "Parton Shower and NLO-Matching uncertainties in Higgs Boson Pair Production", *JHEP* **02** (2018) 176. doi:10.1007/JHEP02(2018)176, arXiv:1711.03319.

[78] J. C. Collins et al., "Angular distribution of dileptons in high-energy hadron collisions", *Phys. Rev. D* **16** (Oct, 1977) 2219–2225. doi:10.1103/PhysRevD.16.2219.

[79] Q.-H. Cao et al., "Double Higgs production at the 14 TeV LHC and a 100 TeV pp collider", *Phys. Rev. D* **96** (2017), no. 9, 095031. doi:10.1103/PhysRevD.96.095031, arXiv:1611.09336.

[80] M. Reichert et al., "Probing baryogenesis through the Higgs boson self-coupling", *Phys. Rev. D* **97** (2018), no. 7, 075008. doi:10.1103/PhysRevD.97.075008, arXiv:1711.00019.

[81] T. Appelquist et al., "Infrared singularities and massive fields", *Phys. Rev. D* **11** (1975) 2856–2861. doi:10.1103/PhysRevD.11.2856.

[82] S. Weinberg, "Baryon- and Lepton-Nonconserving Processes", *Phys. Rev. Lett.* **43** (1979) 1566–1570. doi:10.1103/PhysRevLett.43.1566.

[83] C. N. Leung et al., "Low-energy manifestations of a new interactions scale: Operator analysis", *Zeitschrift für Physik C* **31** (1986), no. 3, 433–437. doi:10.1007/BF01588041.

[84] W. Buchmüller et al., "Effective lagrangian analysis of new interactions and flavour conservation", *Nucl. Phys. B* **268** (1986), no. 3, 621 – 653. doi:10.1016/0550-3213(86)90262-2.

[85] R. Contino et al., "On the Validity of the Effective Field Theory Approach to SM Precision Tests", *JHEP* **07** (2016) 144. doi:10.1007/JHEP07(2016)144, arXiv:1604.06444.

[86] E. E. Jenkins et al., "Renormalization Group Evolution of the Standard Model Dimension Six Operators I: Formalism and lambda Dependence", *JHEP* **10** (2013) 087. doi:10.1007/JHEP10(2013)087, arXiv:1308.2627.

[87] E. E. Jenkins et al., "Renormalization Group Evolution of the Standard Model Dimension Six Operators II: Yukawa Dependence", *JHEP* **01** (2014) 035. doi:10.1007/JHEP01(2014)035, arXiv:1310.4838.

[88] R. Alonso et al., "Renormalization Group Evolution of the Standard Model Dimension Six Operators III: Gauge Coupling Dependence and Phenomenology", *JHEP* **04** (2014) 159. doi:10.1007/JHEP04(2014)159, arXiv:1312.2014.

[89] A. Biekötter et al., "Extending the limits of Higgs effective theory", *Phys. Rev. D* **94** (2016), no. 5, 055032. doi:10.1103/PhysRevD.94.055032, arXiv:1602.05202.

[90] B. Grzadkowski et al., "Dimension-Six Terms in the Standard Model Lagrangian", *JHEP* **10** (2010) 085. doi:10.1007/JHEP10(2010)085, arXiv:1008.4884.

[91] G. F. Giudice et al., "The Strongly-Interacting Light Higgs", *JHEP* **06** (2007) 045. doi:10.1088/1126-6708/2007/06/045, arXiv:hep-ph/0703164.

[92] R. Contino et al., "Effective Lagrangian for a light Higgs-like scalar", *JHEP* **07** (2013) 035. doi:10.1007/JHEP07(2013)035, arXiv:1303.3876.

[93] F. Goertz et al., "Higgs boson pair production in the D=6 extension of the SM", *JHEP* **04** (2015) 167. doi:10.1007/JHEP04(2015)167, arXiv:1410.3471.

[94] F. Maltoni et al., "Higgs production in association with a top-antitop pair in the Standard Model Effective Field Theory at NLO in QCD", *JHEP* **10** (2016) 123. doi:10.1007/JHEP10(2016)123, arXiv:1607.05330.

[95] A. Azatov et al., "Effective field theory analysis of double Higgs boson production via gluon fusion", *Phys. Rev. D* **92** (2015) 035001. doi:10.1103/PhysRevD.92.035001.

[96] D. Buarque Franzosi et al., "Probing the top-quark chromomagnetic dipole moment at next-to-leading order in QCD", *Phys. Rev. D* **91** (2015), no. 11, 114010. doi:10.1103/PhysRevD.91.114010, arXiv:1503.08841.

[97] A. Azatov et al., "Helicity selection rules and noninterference for BSM amplitudes", *Phys. Rev. D* **95** (2017), no. 6, 065014. doi:10.1103/PhysRevD.95.065014, arXiv:1607.05236.

[98] R. Gröber et al., "NLO QCD Corrections to Higgs Pair Production including Dimension-6 Operators", *JHEP* **09** (2015) 092. doi:10.1007/JHEP09(2015)092, arXiv:1504.06577.

[99] D. de Florian et al., "Higgs boson pair production at NNLO in QCD including dimension 6 operators", *JHEP* **10** (2017) 215. doi:10.1007/JHEP10(2017)215, arXiv:1704.05700.

[100] A. Carvalho et al., "Higgs Pair Production: Choosing Benchmarks With Cluster Analysis", *JHEP* **04** (2016) 126. doi:10.1007/JHEP04(2016)126, arXiv:1507.02245.

[101] S. Di Vita et al., "A global view on the Higgs self-coupling", *JHEP* **09** (2017) 069. doi:10.1007/JHEP09(2017)069, arXiv:1704.01953.

[102] G. Buchalla et al., "Complete Electroweak Chiral Lagrangian with a Light Higgs at NLO", *Nucl. Phys. B* **880** (2014) 552–573. doi:10.1016/j.nuclphysb.2016.09.010, 10.1016/j.nuclphysb.2014.01.018, arXiv:1307.5017. [Erratum: Nucl. Phys.B913,475(2016)].

[103] A. Carvalho et al., "Analytical parametrization and shape classification of anomalous HH production in EFT approach", Technical Report LHCHXSWG-2016-001, 2016. https://cds.cern.ch/record/2199287.

[104] A. Carvalho et al., "On the reinterpretation of non-resonant searches for Higgs boson pairs", (2017). arXiv:1710.08261.

[105] S. Brochet et al., "MoMEMta - MadGraph Matrix Element Exporter", 2016. https://github.com/MoMEMta/MoMEMta-MaGMEE/tree/standalone.

[106] M. Baak et al., "Interpolation between multi-dimensional histograms using a new non-linear moment morphing method", *Nucl. Instrum. Meth. A* **771** (2015) 39 – 48. doi:https://doi.org/10.1016/j.nima.2014.10.033.

[107] L. Brenner et al., "A morphing technique for signal modelling in a multidimensional space of coupling parameters", Technical Report LHCHXSWG-INT-2016-004, 2016. https://cds.cern.ch/record/2143180.

[108] M. J. Dolan et al., "New Physics in LHC Higgs boson pair production", *Phys. Rev. D* **87** (2013), no. 5, 055002. doi:10.1103/PhysRevD.87.055002, arXiv:1210.8166.

[109] V. Barger et al., "New physics in resonant production of Higgs boson pairs", *Phys. Rev. Lett.* **114** (2015), no. 1, 011801. doi:10.1103/PhysRevLett.114.011801, arXiv:1408.0003.

[110] S. Dawson et al., "Standard Model EFT and Extended Scalar Sectors", *Phys. Rev. D* **96** (2017), no. 1, 015041. doi:10.1103/PhysRevD.96.015041, arXiv:1704.07851.

[111] T. Binoth et al., "Influence of strongly coupled, hidden scalars on Higgs signals", *Z. Phys. C* **75** (1997) 17–25. doi:10.1007/s002880050442, arXiv:hep-ph/9608245.

[112] B. Patt et al., "Higgs-field portal into hidden sectors", (2006). arXiv:hep-ph/0605188.

[113] R. Schabinger et al., "Minimal spontaneously broken hidden sector and its impact on Higgs boson physics at the CERN Large Hadron Collider", *Phys. Rev. D* **72** (Nov, 2005) 093007. doi:10.1103/PhysRevD.72.093007.

[114] C.-Y. Chen et al., "Exploring resonant di-Higgs boson production in the Higgs singlet model", *Phys. Rev. D* **91** (2015), no. 3, 035015. doi:10.1103/PhysRevD.91.035015, arXiv:1410.5488.

[115] J. M. No et al., "Probing the Higgs Portal at the LHC Through Resonant di-Higgs Production", *Phys. Rev. D* **89** (2014), no. 9, 095031. doi:10.1103/PhysRevD.89.095031, arXiv:1310.6035.

[116] T. Robens et al., "LHC Benchmark Scenarios for the Real Higgs Singlet Extension of the Standard Model", *Eur. Phys. J. C* **76** (2016), no. 5, 268. doi:10.1140/epjc/s10052-016-4115-8, arXiv:1601.07880.

[117] T. Robens et al., "Status of the Higgs Singlet Extension of the Standard Model after LHC Run 1", *Eur. Phys. J. C* **75** (2015) 104. doi:10.1140/epjc/s10052-015-3323-y, arXiv:1501.02234.

[118] S. Ghosh et al., "Potential of a singlet scalar enhanced Standard Model", *Phys. Rev. D* **93** (2016), no. 11, 115034. doi:10.1103/PhysRevD.93.115034, arXiv:1512.05786.

[119] G. C. Branco et al., "Theory and phenomenology of two-Higgs-doublet models", *Phys. Rept.* **516** (2012) 1–102. doi:10.1016/j.physrep.2012.02.002, arXiv:1106.0034.

[120] J. Mrazek et al., "The Other Natural Two Higgs Doublet Model", *Nucl. Phys. B* **853** (2011) 1–48. doi:10.1016/j.nuclphysb.2011.07.008, arXiv:1105.5403.

[121] A. Djouadi et al., "The post- Higgs MSSM scenario: Habemus MSSM?", *Eur. Phys. J. C* **73** (2013) 2650. doi:10.1140/epjc/s10052-013-2650-0, arXiv:1307.5205.

[122] A. Djouadi et al., "Fully covering the MSSM Higgs sector at the LHC", *JHEP* **06** (2015) 168. doi:10.1007/JHEP06(2015)168, arXiv:1502.05653.

[123] A. Arhrib et al., "Double Neutral Higgs production in the Two-Higgs doublet model at the LHC", *JHEP* **08** (2009) 035. doi:10.1088/1126-6708/2009/08/035, arXiv:0906.0387.

[124] B. Hespel et al., "Higgs pair production via gluon fusion in the Two-Higgs-Doublet Model", *JHEP* **09** (2014) 124. doi:10.1007/JHEP09(2014)124, arXiv:1407.0281.

[125] H. E. Haber et al., "New LHC benchmarks for the $\mathcal{CP}$ -conserving two-Higgs-doublet model", *Eur. Phys. J. C* **75** (2015), no. 10, 491. doi:10.1140/epjc/s10052-015-3697-x, 10.1140/epjc/s10052-016-4151-4, arXiv:1507.04281. [Erratum: Eur. Phys. J.C76,no.6,312(2016)].

[126] H. Georgi et al., "Doubly charged Higgs bosons", *Nucl. Phys. B* **262** (1985) 463–477. doi:10.1016/0550-3213(85)90325-6.

[127] M. S. Chanowitz et al., "Higgs Boson Triplets With $M_W = M_Z \cos\theta_W$", *Phys. Lett. B* **165** (1985) 105–108. doi:10.1016/0370-2693(85)90700-2.

[128] J. Chang et al., "Higgs boson pair productions in the Georgi-Machacek model at the LHC", *JHEP* **03** (2017) 137. doi:10.1007/JHEP03(2017)137, arXiv:1701.06291.

[129] L. Randall et al., "A Large mass hierarchy from a small extra dimension", *Phys. Rev. Lett.* **83** (1999) 3370–3373. doi:10.1103/PhysRevLett.83.3370, arXiv:hep-ph/9905221.

[130] H. Davoudiasl et al., "Bulk gauge fields in the Randall-Sundrum model", *Phys. Lett. B* **473** (2000) 43–49. doi:10.1016/S0370-2693(99)01430-6, arXiv:hep-ph/9911262.

[131] A. L. Fitzpatrick et al., "Searching for the Kaluza-Klein Graviton in Bulk RS Models", *JHEP* **09** (2007) 013. doi:10.1088/1126-6708/2007/09/013, arXiv:hep-ph/0701150.

[132] K. Agashe et al., "Warped Gravitons at the LHC and Beyond", *Phys. Rev. D* **76** (2007) 036006. doi:10.1103/PhysRevD.76.036006, arXiv:hep-ph/0701186.

[133] H. Davoudiasl et al., "Warped 5-Dimensional Models: Phenomenological Status and Experimental Prospects", *New J. Phys.* **12** (2010) 075011. doi:10.1088/1367-2630/12/7/075011, arXiv:0908.1968.

[134] C. Csaki et al., "Radion phenomenology in realistic warped space models", *Phys. Rev. D* **76** (2007) 125015. doi:10.1103/PhysRevD.76.125015, arXiv:0705.3844.

[135] A. Oliveira, "Gravity particles from Warped Extra Dimensions, predictions for LHC", (2014). arXiv:1404.0102.

[136] A. C. A. Oliveira et al., "Hidden sector effects on double higgs production near threshold at the LHC", *Phys. Lett. B* **702** (2011) 201–204. doi:10.1016/j.physletb.2011.06.086, arXiv:1009.4497.

[137] S. Dawson et al., "What's in the loop? The anatomy of double Higgs production", *Phys. Rev. D* **91** (Jun, 2015) 115008. doi:10.1103/PhysRevD.91.115008.

[138] G. Cacciapaglia et al., "Probing vector-like quark models with Higgs-boson pair production", *JHEP* **07** (2017), no. 7, 005. doi:10.1007/JHEP07(2017)005, arXiv:1703.10614.

[139] M. van Beekveld et al., "Higgs, di-Higgs and tri-Higgs production via SUSY processes at the LHC with 14 TeV", *JHEP* **05** (2015) 044. doi:10.1007/JHEP05(2015)044, arXiv:1501.02145.

[140] CMS Collaboration, "Search for Higgs boson pair production in the $\gamma\gamma b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13$ TeV",. arXiv:1806.00408.

[141] ATLAS Collaboration, "Searches for Higgs boson pair production in the $hh \to b\bar{b}\tau\tau, \gamma\gamma WW^*, \gamma\gamma b\bar{b}, b\bar{b}b\bar{b}$ channels with the ATLAS detector", *Phys. Rev. D* **92** (2015) 092004. doi:10.1103/PhysRevD.92.092004, arXiv:1509.04670.

[142] ATLAS Collaboration, "Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", (2018). arXiv:1804.06174. Submitted to JHEP.

[143] ATLAS Collaboration, "Search for Higgs boson pair production in the $b\bar{b}\gamma\gamma$ final state using pp collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector", Technical Report ATLAS-CONF-2016-004, 2016. http://cds.cern.ch/record/2138949.

[144] ATLAS Collaboration, "Search for Higgs boson pair production in the final state of $\gamma\gamma WW^*(\to \ell\nu jj)$ using 13.3 fb^{-1} of pp collision data recorded at $\sqrt{s} = 13$ TeV with the ATLAS detector", Technical Report ATLAS-CONF-2016-071, 2016. http://cds.cern.ch/record/2206222.

[145] CMS Collaboration, "Search for Higgs boson pair production in the $b\bar{b}\tau\tau$ final state in proton-proton collisions at $\sqrt{(s)} = 8$ TeV", *Phys. Rev. D* **96** (Oct, 2017) 072004. doi:10.1103/PhysRevD.96.072004.

[146] CMS Collaboration, "Search for Higgs boson pair production in events with two bottom quarks and two tau leptons in proton–proton collisions at $\sqrt{s}$ =13TeV", *Phys. Lett. B* **778** (2018) 101–127. doi:10.1016/j.physletb.2018.01.001, arXiv:1707.02909.

[147] CMS Collaboration, "Search for resonant and nonresonant Higgs boson pair production in the $b\bar{b}\ell\nu\ell\nu$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **01** (2018) 054. doi:10.1007/JHEP01(2018)054, arXiv:1708.04188.

[148] CMS Collaboration, "Search for resonant and non-resonant production of Higgs boson pairs in the four b quark final state using boosted jets in proton-proton collisions at $\sqrt{s} = 13$ TeV", Technical Report CMS-PAS-B2G-17-019, 2018. https://cds.cern.ch/record/2621541.

[149] CMS Collaboration, "Search for non-resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state with 13 TeV CMS data", Technical Report CMS-PAS-HIG-16-026, 2016. http://cds.cern.ch/record/2209572.

[150] CMS Collaboration, "Search for a massive resonance decaying to a pair of Higgs bosons in the four b quark final state in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Lett. B* **781** (2018) 244–269. doi:10.1016/j.physletb.2018.03.084, arXiv:1710.04960.

[151] CMS Collaboration, "Search for resonant pair production of Higgs bosons decaying to bottom quark-antiquark pairs in proton-proton collisions at 13 TeV", Technical Report CMS-PAS-HIG-17-009, 2017. http://cds.cern.ch/record/2292044.

[152] CMS Collaboration, "Search for heavy resonances decaying into two Higgs bosons or into a Higgs and a vector boson in proton-proton collisions at 13 TeV", Technical Report CMS-PAS-B2G-17-006, 2017. http://cds.cern.ch/record/2296716.

[153] CMS Collaboration, "Higgs pair production at the High Luminosity LHC", Technical Report CMS-PAS-FTR-15-002, 2015. https://cds.cern.ch/record/2063038.

[154] CMS Collaboration, "Projected performance of Higgs analyses at the HL-LHC for ECFA 2016", Technical Report CMS-PAS-FTR-16-002, 2017. http://cds.cern.ch/record/2266165.

[155] CMS Collaboration, "The Phase-2 Upgrade of the CMS Barrel Calorimeters Technical Design Report", Technical Report CERN-LHCC-2017-011. CMS-TDR-015, 2017. https://cds.cern.ch/record/2283187.

[156] CMS Collaboration, "The Phase-2 Upgrade of the CMS Tracker", Technical Report CERN-LHCC-2017-009, CMS-TDR-014, 2017. http://cds.cern.ch/record/2272264.

[157] ATLAS Collaboration, "Study of the double Higgs production channel $H(\to b\bar{b})H(\to \gamma\gamma)$ with the ATLAS experiment at the HL-LHC", Technical Report ATL-PHYS-PUB-2017-001, 2017. https://cds.cern.ch/record/2243387.

[158] ATLAS Collaboration, "Studies of the ATLAS potential for Higgs self-coupling measurements at a High Luminosity LHC", Technical Report ATL-PHYS-PUB-2013-001, 2013. https://cds.cern.ch/record/1507263.

[159] L. Di Luzio et al., "Maxi-sizing the trilinear Higgs self-coupling: how large could it be?", *Eur. Phys. J. C* **77** (2017) 788. doi:10.1140/epjc/s10052-017-5361-0, arXiv:1704.02311.

[160] M. McCullough, "An Indirect Model-Dependent Probe of the Higgs Self-Coupling", *Phys. Rev. D* **90** (2014), no. 1, 015001. doi:10.1103/PhysRevD.90.015001, 10.1103/PhysRevD.92.039903, arXiv:1312.3322. [Erratum: Phys. Rev.D92,no.3,039903(2015)].

[161] M. Gorbahn et al., "Indirect probes of the trilinear Higgs coupling: $gg \to h$ and $h \to \gamma\gamma$", *JHEP* **10** (2016) 094. doi:10.1007/JHEP10(2016)094, arXiv:1607.03773.

[162] G. Degrassi et al., "Probing the Higgs self coupling via single Higgs production at the LHC", *JHEP* **12** (2016) 080. doi:10.1007/JHEP12(2016)080, arXiv:1607.04251.

[163] W. Bizon et al., "Constraints on the trilinear Higgs coupling from vector boson fusion and associated Higgs production at the LHC", *JHEP* **07** (2017) 083. doi:10.1007/JHEP07(2017)083, arXiv:1610.05771.

[164] G. Degrassi et al., "Constraints on the trilinear Higgs self coupling from precision observables", *JHEP* **04** (2017) 155. doi:10.1007/JHEP04(2017)155, arXiv:1702.01737.

[165] G. D. Kribs et al., "Electroweak oblique parameters as a probe of the trilinear Higgs boson self-interaction", *Phys. Rev. D* **95** (2017), no. 9, 093004. doi:10.1103/PhysRevD.95.093004, arXiv:1702.07678.

[166] L. Evans et al., "LHC Machine", *JINST* **3** (2008) S08001. doi:10.1088/1748-0221/3/08/S08001.

[167] E. Mobs, "The CERN accelerator complex. Complexe des accélérateurs du CERN", 2018. https://cds.cern.ch/record/2197559.

[168] H. Bartosik et al., "Performance potential of the injectors after LS1", Technical Report CERN-2012-006, 2012. https://cds.cern.ch/record/1492996.

[169] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003. doi:10.1088/1748-0221/3/08/S08003.

[170] CMS Collaboration, "The CMS Experiment at the CERN LHC", *JINST* **3** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004.

[171] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005. doi:10.1088/1748-0221/3/08/S08005.

[172] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002. doi:10.1088/1748-0221/3/08/S08002.

[173] LHCf Collaboration, "The LHCf detector at the CERN Large Hadron Collider", *JINST* **3** (2008) S08006. doi:10.1088/1748-0221/3/08/S08006.

[174] MoEDAL Collaboration, "The Physics Programme Of The MoEDAL Experiment At The LHC", *Int. J. Mod. Phys. A* **29** (2014) 1430050. doi:10.1142/S0217751X14300506, arXiv:1405.7662.

[175] TOTEM Collaboration, "The TOTEM experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08007. doi:10.1088/1748-0221/3/08/S08007.

[176] P. Grafström et al., "Luminosity determination at proton colliders", *Prog. Part. Nucl. Phys.* **81** (2015) 97–148. doi:10.1016/j.ppnp.2014.11.002.

[177] S. van der Meer, "Calibration of the effective beam height in the ISR", Technical Report CERN-ISR-PO-68-31. ISR-PO-68-31, 1968. https://cds.cern.ch/record/296752.

[178] "CMS luminosity - public results". https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. Accessed on 29/05/2018.

[179] T. Sakuma et al., "Detector and Event Visualization with SketchUp at the CMS Experiment", *Journal of Physics: Conference Series* **513** (2014), no. 2, 022032.

[180] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014), no. 10, P10009. doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[181] CMS Collaboration, "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software", Technical Report CERN-LHCC-2006-001. CMS-TDR-8-1, 2006. https://cds.cern.ch/record/922757.

[182] P. Adzic et al., "Energy resolution of the barrel of the CMS electromagnetic calorimeter", *JINST* **2** (2007) P04004. doi:10.1088/1748-0221/2/04/P04004.

[183] CMS Collaboration, "Design, Performance, and Calibration of CMS Hadron-Barrel Calorimeter Wedges", Technical Report CMS-NOTE-2006-138, 2007. https://cds.cern.ch/record/1049915.

[184] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV", (2018). arXiv:1804.04528. Submitted to *JINST*.

[185] CMS Collaboration, "CMS Technical Design Report for the Level-1 Trigger Upgrade", Technical Report CERN-LHCC-2013-011. CMS-TDR-12, 2013. https://cds.cern.ch/record/1556311.

[186] T. Bawej et al., "The New CMS DAQ System for Run-2 of the LHC", in *2014 19th IEEE-NPSS Real Time Conference*. IEEE, 2014. https://doi.org/10.1109/RTC.2014.7097437.

[187] CMS Collaboration, "CMS computing: Technical Design Report", Technical Report CERN-LHCC-2005-023. CMS-TDR-7, 2005. https://cds.cern.ch/record/838359.

[188] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017), no. 10, P10003. doi:10.1088/1748-0221/12/10/P10003.

[189] "Run II Pixel Performance plots for data and simulation". https://twiki.cern.ch/twiki/bin/view/CMSPublic/PixelOfflinePlots2016. Accessed on 24/10/2017.

[190] "CMS Silicon Strip Performance Results 2016". https://twiki.cern.ch/twiki/bin/view/CMS/StripsOfflinePlots2016. Accessed on 24/10/2017.

[191] "CMS Tracking POG Performance Plots for Vertex 2014". https://twiki.cern.ch/twiki/bin/view/CMSPublic/TrackingPOGPlotsVertex2014. Accessed on 25/10/2017.

[192] "CMS Tracking POG Performance Plots for year 2016". https://twiki.cern.ch/twiki/bin/view/CMSPublic/TrackingPOGPlots2016. Accessed on 25/10/2017.

[193] W. Adam et al., "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. doi:10.1088/0954-3899/31/9/N01.

[194] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015), no. 06, P06005. doi:10.1088/1748-0221/10/06/P06005.

[195] "CMS Tracking POG Performance Plots 2016 (targeting ICHEP conference)". https://twiki.cern.ch/twiki/bin/view/CMSPublic/TrackingPOGPlotsICHEP2016. Accessed on 15/11/2017.

[196] CMS Collaboration, "Muon Identification and Isolation efficiency on full 2016 dataset", Technical Report CMS-DP-2017-007, 2017. https://cds.cern.ch/record/2257968.

[197] CMS Collaboration, "Electron and photon performance in CMS with the full 2016 data sample.", Technical Report CMS-DP-2017-004, 2017. https://cds.cern.ch/record/2255497.

[198] CMS Collaboration, "Jet algorithms performance in 13 TeV data", Technical Report CMS-PAS-JME-16-003, 2017. https://cds.cern.ch/record/2256875.

[199] CMS Collaboration, "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s}$ = 8 TeV", *JINST* **10** (2015), no. 02, P02006. doi:10.1088/1748-0221/10/02/P02006, arXiv:1411.0511.

[200] CMS Collaboration, "Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector", Technical Report CMS-PAS-JME-16-004, 2016. https://cds.cern.ch/record/2205284.

[201] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", (2017). arXiv:1712.07158. Submitted to *JINST*.

[202] N. Bartosik, "Diagram showing the common principle of identification of jets initiated by b-hadron decays". https://commons.wikimedia.org/wiki/File:B-tagging_diagram.png. Accessed on 27/11/2017.

[203] M. Cacciari et al., "Pileup subtraction using jet areas", *Phys. Lett. B* **659** (2008) 119–126. doi:10.1016/j.physletb.2007.09.077, arXiv:0707.1378.

[204] M. Cacciari et al., "The Catchment Area of Jets", *JHEP* **04** (2008) 005. doi:10.1088/1126-6708/2008/04/005, arXiv:0802.1188.

[205] CMS Collaboration, "CMS Luminosity Measurements for the 2016 Data Taking Period", Technical Report CMS-PAS-LUM-17-001, 2017. https://cds.cern.ch/record/2257069.

[206] CMS Collaboration, "Measurement of the inelastic proton-proton cross section at $\sqrt{s}$ = 13 TeV", Technical Report CMS-PAS-FSQ-15-005, 2016. https://cds.cern.ch/record/2145896.

[207] GEANT4 Collaboration, "GEANT4—a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250. doi:10.1016/S0168-9002(03)01368-8.

[208] CMS Collaboration, "Electron and photon performance in CMS with first 12.9/fb of 2016 data", Technical Report CMS-DP-2016-049, 2016. https://cds.cern.ch/record/2203016.

[209] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017), no. 02, P02014. doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

[210] CMS Collaboration, "Jet energy scale and resolution performances with 13TeV data", Technical Report CMS-DP-2016-020, 2016. https://cds.cern.ch/record/2160347.

[211] J. Neyman et al., "On the Problem of the Most Efficient Tests of Statistical Hypotheses", *Phil. Trans. R. Soc. Lond. A* **231** (1933), no. 694-706, 289–337. doi:10.1098/rsta.1933.0009.

[212] Y. Freund et al., "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting", *Journal of Computer and System Sciences* **55** (1997), no. 1, 119–139. doi:https://doi.org/10.1006/jcss.1997.1504.

[213] L. Mason et al., "Boosting Algorithms as Gradient Descent", in *Advances in Neural Information Processing Systems 12*. MIT Press, 1999. http://papers.nips.cc/paper/1766-boosting-algorithms-as-gradient-descent.

[214] L. Breiman, "Bagging predictors", *Machine Learning* **24** (1996), no. 2, 123–140. doi:10.1007/BF00058655.

[215] F. Rosenblatt, "The Perceptron: A Probabilistic Model for Information Storage and Organization in The Brain", *Psychological Review* **65** (1958), no. 6, 386. doi:10.1037/h0042519.

[216] D. E. Rumelhart et al., "Learning Representations by Back-propagating Errors", *Nature* **323** (1986) 533–536. doi:10.1038/323533a0.

[217] G. H. Yann LeCun, Yoshua Bengio, "Deep Learning", *Nature* **521** (2015) 436–444. doi:10.1038/nature14539.

[218] N. Qian, "On the momentum term in gradient descent learning algorithms", *Neural Networks* **12** (1999), no. 1, 145 – 151. doi:https://doi.org/10.1016/S0893-6080(98)00116-6.

[219] Y. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$.", *Soviet Mathematics Doklady* **27** (1983), no. 2, 372–376.

[220] J. Duchi et al., "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization", *Journal of Machine Learning Research* (2011) 2121–2159.

[221] M. Zeiler, "ADADELTA: An Adaptive Learning Rate Method", (2012). arXiv:1212.5701.

[222] D. P. Kingma et al., "Adam: A Method for Stochastic Optimization", (2014). arXiv:1412.6980.

[223] G. Montavon et al., eds., "Neural Networks: Tricks of the Trade", volume 7700 of *Lecture Notes in Computer Science*. Springer, Berlin (Germany), 2nd edition, 2012.

[224] G. E. Hinton et al., "Improving neural networks by preventing co-adaptation of feature detectors", (2012). arXiv:1207.0580.

[225] N. Srivastava et al., "Dropout: A Simple Way to Prevent Neural Networks from Overfitting", *Journal of Machine Learning Research* **15** (2014) 1929–1958.

[226] S. Ioffe et al., "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", (2015). arXiv:1502.03167.

[227] D. Mishkin et al., "All you need is a good init", (2015). arXiv:1511.06422.

[228] X. Glorot et al., "Understanding the difficulty of training deep feedforward neural networks", in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9, pp. 249–256. PMLR, 2010. http://proceedings.mlr.press/v9/glorot10a.html.

[229] K. He et al., "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification", (2015). arXiv:1502.01852.

[230] K. Cranmer, "Practical Statistics for the LHC", in *Proceedings, 2011 European School of High-Energy Physics (ESHEP 2011): Cheile Gradistei, Romania, September 7-20, 2011*, number 247, pp. 267 – 308. 2015. arXiv:1503.07622.

[231] R. J. Barlow et al., "Fitting using finite Monte Carlo samples", *Comput. Phys. Commun.* **77** (1993) 219–228. doi:10.1016/0010-4655(93)90005-W.

[232] C.-S. Chuang et al., "Hybrid Resampling Methods For Confidence Intervals", *Statistica Sinica* **10** (2000), no. 1, 1 – 33.

[233] B. Sen et al., "On the Unified Method with Nuisance Parameters", *Statistica Sinica* **19** (2009), no. 1, 301 – 314.

[234] W. A. Rolke et al., "Limits and confidence intervals in the presence of nuisance parameters", *Nucl. Instrum. Meth. A* **551** (2005) 493 – 503. doi:10.1016/j.nima.2005.05.068, arXiv:physics/0403059.

[235] R. D. Cousins et al., "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", *Nucl. Instrum. Meth. A* **595** (2008), no. 2, 480 – 501. doi:10.1016/j.nima.2008.07.086, arXiv:physics/0702156.

[236] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth. A* **434** (1999) 435–443. doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.

[237] A. L. Read, "Presentation of search results: The CL_s technique", *J. Phys. G* **28** (2002), no. 11, 2693–2704. doi:10.1088/0954-3899/28/10/313.

[238] G. Cowan et al., "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554. doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727.

[239] S. S. Wilks, "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses", *Annals Math. Statist.* **9** (1938), no. 1, 60 – 62. doi:10.1214/aoms/1177732360.

[240] A. Wald, "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large", *Transactions of the American Mathematical Society* **54** (1943), no. 3, 426 – 482.

[241] W. Verkerke et al., "The RooFit toolkit for data modeling", in *Statistical Problems in Particle Physics, Astrophysics and Cosmology (PHYSTAT 05): Proceedings, Oxford, UK, September 12-15, 2005*, number 186, p. MOLT007. 2003. arXiv:physics/0306116.

[242] L. Moneta et al., "The RooStats Project", in *Proceedings, 13th International Workshop on Advanced computing and analysis techniques in physics research (ACAT2010): Jaipur, India, February 22-27, 2010*, p. 057. 2010. arXiv:1009.1003.

[243] F. James et al., "Minuit – a system for function minimization and analysis of the parameter errors and correlations", *Computer Physics Communications* **10** (1975), no. 6, 343 – 367. doi:https://doi.org/10.1016/0010-4655(75)90039-9.

[244] LHC Higgs Cross Section Working Group, "SM Higgs Branching Ratios and Total Decay Widths", 2016. https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR.

[245] S. Bolognesi et al., "On the spin and parity of a single-produced resonance at the LHC", *Phys. Rev. D* **86** (2012) 095031. doi:10.1103/PhysRevD.86.095031, arXiv:1208.4018.

[246] J. Ellis et al., "Does the 'Higgs' have Spin Zero?", *JHEP* **09** (2012) 071. doi:10.1007/JHEP09(2012)071, arXiv:1202.6660.

[247] M. Czakon et al., "Top++: A program for the calculation of the top-pair cross-section at hadron colliders", *Comput. Phys. Commun.* **185** (2014) 2930. doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.

[248] S. Alioli et al., "Hadronic top-quark pair-production with one jet and parton showering", *JHEP* **01** (2012) 137. doi:10.1007/JHEP01(2012)137, arXiv:1110.5251.

[249] S. Alioli et al., "NLO single-top production matched with shower in POWHEG: s- and t-channel contributions", *JHEP* **09** (2009) 111. doi:10.1088/1126-6708/2009/09/111, arXiv:0907.4076.

[250] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C* **76** (2015) 155. doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

[251] Y. Li et al., "Combining QCD and electroweak corrections to dilepton production in the framework of the FEWZ simulation code", *Phys. Rev. D* **86** (2012) 094034. doi:10.1103/PhysRevD.86.094034, arXiv:1208.5967.

[252] N. Kidonakis, "Top Quark Production", (2014). arXiv:1311.0283.

[253] T. Gehrmann et al., "W^+W^- Production at Hadron Colliders in Next to Next to Leading Order QCD", *Phys. Rev. Lett.* **113** (2014) 212001. doi:10.1103/PhysRevLett.113.212001, arXiv:1408.5243.

[254] J. M. Campbell et al., "Vector boson pair production at the LHC", *JHEP* **07** (2011) 018. doi:10.1007/JHEP07(2011)018, arXiv:1105.0020.

[255] O. Bondu et al., "Search for resonances decaying in two Higgs bosons in the $b\bar{b}l\nu l\nu$ final state", CMS Analysis Note 2015/280, 2016.

[256] CMS Collaboration, "Search for resonant Higgs boson pair production in the $b\bar{b}l\nu l\nu$ final state at $\sqrt{s} = 13$ TeV", Technical Report CMS-PAS-HIG-16-011, 2016. https://cds.cern.ch/record/2141743.

[257] CMS Collaboration, "Search for Higgs boson pair production in the $b\bar{b}l\nu l\nu$ final state at $\sqrt{s} = 13$ TeV", Technical Report CMS-PAS-HIG-16-024, 2016. https://cds.cern.ch/record/2205782.

[258] P. Baldi et al., "Parameterized neural networks for high-energy physics", *Eur. Phys. J. C* **76** (2016) 235. doi:10.1140/epjc/s10052-016-4099-4, arXiv:1601.07913.

[259] F. Chollet, "Keras", 2015. https://github.com/fchollet/keras.

[260] M. Abadi et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems", 2015. https://www.tensorflow.org/.

[261] H. Akima, "A New Method of Interpolation and Smooth Curve Fitting Based on Local Procedures", *J. ACM* **17** (October, 1970) 589–602. doi:10.1145/321607.321609.

[262] CMS Collaboration, "Measurement of the production cross sections for a Z boson and one or more b jets in pp collisions at $\sqrt{s} = 7$ TeV", *JHEP* **06** (2014) 120. doi:10.1007/JHEP06(2014)120, arXiv:1402.1521.

[263] CMS Collaboration, "Measurements of the associated production of a Z boson and b jets in pp collisions at $\sqrt{s} = 8$ TeV", *Eur. Phys. J. C* **77** (2017) 751. doi:10.1140/epjc/s10052-017-5140-y, arXiv:1611.06507.

[264] "An invariant form for the prior probability in estimation problems", *Proc. R. Soc. Lond. A* **186** (1946), no. 1007, 453–461. doi:10.1098/rspa.1946.0056, arXiv:http://rspa.royalsocietypublishing.org/content/186/1007/453.full.pdf.

[265] "Higgs PAG Summary Plots". https://twiki.cern.ch/twiki/bin/view/CMSPublic/SummaryResultsHIG. Accessed on 18/04/2018.

[266] T. P. S. Gillam et al., "Improving estimates of the number of 'fake' leptons and other mis-reconstructed objects in hadron collider events: BoB's your UNCLE", *JHEP* **11** (2014) 031. doi:10.1007/JHEP11(2014)031, arXiv:1407.5624.

[267] J. Bergstra et al., "Random Search for Hyper-parameter Optimization", *J. Mach. Learn. Res.* **13** (2012) 281–305.

[268] C. Rasmussen et al., "Gaussian Processes for Machine Learning". Adaptative computation and machine learning series. University Press Group Limited, 2006.

[269] J. Bergstra et al., "Algorithms for Hyper-parameter Optimization", in *Proceedings of the 24th International Conference on Neural Information Processing Systems*, pp. 2546–2554. 2011. http://dl.acm.org/citation.cfm?id=2986459.2986743.

[270] "Hyperopt". http://hyperopt.github.io/hyperopt/. Accessed on 26/04/2018.

[271] G. Louppe et al., "Learning to Pivot with Adversarial Networks", (2016). arXiv:1611.01046.

[272] Y. Ganin et al., "Domain-Adversarial Training of Neural Networks", (2015). arXiv:1505.07818.

[273] J. de Favereau et al., "DELPHES 3, A modular framework for fast simulation of a generic collider experiment", *JHEP* **02** (2014) 057. doi:10.1007/JHEP02(2014)057, arXiv:1307.6346.