SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Dipartimento di / Department of

Fisica “Giuseppe Occhialini”

Dottorato di Ricerca in / PhD program Fisica e Astronomia Ciclo / Cycle 38

Curriculum in (se presente / if it is) Fisica Subnucleare

TITOLO TESI / THESIS TITLE

Measurement of the Electroweak Production of Z Bosons in Proton Collisions and the Data Acquisition System for the CMS MIP Timing Detector

Cognome / Surname Pizzati Nome / Name Giorgio

Matricola / Registration number 826389

Tutore / Tutor: Prof. Marco Paganoni

Cotutore / Co-tutor: Dott. Andrea Massironi
(se presente / if there is one)

Supervisor: Prof. Pietro Govoni
(se presente / if there is one)

Coordinatore / Coordinator: Prof. Stefano Ragazzi

ANNO ACCADEMICO / ACADEMIC YEAR 2024/2025

Abstract

This thesis presents the outcome of a three-year research activity conducted within the Compact Muon Solenoid (CMS) collaboration at the CERN Large Hadron Collider (LHC), focusing on two complementary domains: precision electroweak physics and detector instrumentation for the Phase-2 upgrade.
The first part of this work is devoted to the measurement of the Vector Boson Fusion (VBF) production of the Z bosons and its interpretation within the framework of the Standard Model Effective Field Theory (SMEFT). I am the main analyzer for the full Run 2 dataset collected by CMS at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 138 fb^{-1}. The results include the inclusive, fiducial, and differential cross sections of the electroweak Z + 2 jets process and the unfolding to particle level for key observables that characterize the VBF topology. A Deep Neural Network (DNN) was trained to enhance the separation between the electroweak signal and the Drell-Yan background, maximazing the precision of the measurement.
The analysis is currently in the CMS approval process both for the Standard Model (SM) measurement and for its SMEFT interpretation. The results of my thesis are also included in a EFT CMS combination, where the VBF-Z channel provides leading sensitivity to two bosonic dimension-six operators in the Warsaw basis, c_{HWB} and c_{HDD}, playing a central role in constraining new physics effects in the electroweak sector.
The second part of the thesis focuses on the development of the Data Acquisition (DAQ) system for the MIP Timing Detector (MTD), a key upgrade of the CMS detector designed to mitigate the effects of high pileup during the High-Luminosity LHC (HL-LHC) era exploiting 4D tracking (space and time). I was the main developer of the MTD DAQ software, which provides full control, monitoring, and data readout capabilities for the Barrel Timing Layer (BTL). The software architecture was designed to be modular and scalable, enabling seamless communication with front-end and control electronics.
A stable and production-ready version of the DAQ was achieved and is now routinely used at the BTL Assembly Centers (BACs) for large-scale tray validation and quality-control operations. The same framework was successfully employed during BTL test beam campaigns to perform timing and calibration studies under realistic conditions, demonstrating its robustness and readiness for detector integration.
Together, these two projects illustrate the dual nature of experimental high-energy physics: the pursuit of precision measurements that probe the Standard Model at unprecedented accuracy, and the development of advanced instrumentation that enables those measurements. The results presented in this work contribute both to the understanding of electroweak interactions through the VBF-Z process and to the technological foundation of the CMS detector in preparation for the HL-LHC program.

Contents

Chapter 1

Introduction

The production of a Z boson in association with two jets via vector boson fusion (VBF-Z) provides a clean probe of the electroweak sector at the LHC. Beyond enabling a precise measurement of an electroweak production mode, this topology is directly sensitive to the non-Abelian structure of the Standard Model through the presence of triple gauge couplings (TGCs), which enter the VBF amplitudes and govern the interactions among electroweak gauge bosons. Given the central role of the TGC vertex in tests of gauge invariance and its enhanced sensitivity to potential new-physics contributions at high momentum transfer, this measurement is complemented by an interpretation in the Standard Model Effective Field Theory (SMEFT) framework. In this approach, possible deviations from the Standard Model predictions are parameterized by higher-dimensional operators, allowing for a systematic and model-independent search for effects beyond the Standard Model in VBF-Z production. The analysis is performed using data collected by the CMS experiment during the full Run 2 period (2016–2018) in proton–proton collisions at $\sqrt{s} = 13$ TeV.

1.1 Foundations of the Standard Model and Effective Field Theory

The Standard Model (SM) of particle physics is a quantum field theory founded on the principle of local gauge symmetry. It provides a unified description of the fundamental interactions among elementary particles and has been confirmed with remarkable precision by experimental results across the highest energy scales probed to date.

The SM represents the culmination of decades of theoretical development and experimental discovery that began in the early twentieth century. From a mathematical standpoint, it is formulated as a relativistic Quantum Field Theory (QFT), where quantum fields—defined at every point in spacetime—constitute the fundamental entities of nature. Particles emerge as excitations of these fields, while their interactions arise naturally from the symmetries governing the theory.

Two concepts lie at the core of this framework: *gauge invariance* and *Noether's theorem*. Gauge invariance requires that the physical content of a theory remain unchanged under specific local transformations of the fields. This condition is not merely a mathematical constraint; it determines the form of the interactions and necessitates the introduction of gauge bosons—the mediators of the fundamental forces. Noether's theorem establishes the profound link between symmetries and conservation laws, asserting that each con-

tinuous symmetry corresponds to a conserved quantity. This correspondence connects conserved charges, such as electric charge, color charge, and weak isospin, directly to the symmetries of the underlying fields.
Together, these principles provide a systematic way to classify particles according to their quantum numbers and to identify the allowed interactions consistent with the symmetries of the theory. The overarching goal of the Standard Model is to describe all known forms of matter and their interactions within a single, coherent theoretical framework built upon a small number of fundamental postulates.
This chapter introduces the theoretical foundations of the Standard Model, beginning with the concept of symmetry and its mathematical formulation through group theory. It then discusses how local symmetries give rise to gauge fields, using Quantum Electrodynamics (QED) as a guiding example, and finally extends these concepts to the complete structure of the SM, encompassing its three main sectors: the Color, Electroweak, and Scalar components.

1.1.1 The Standard Model Formalism

Quantum Field Theory's historical development has been intrinsically tied to the canonical formalism [1, 2, 3, 4]. By utilizing the Lagrangian formalism, achieving Lorentz invariance becomes more straightforward.
By denoting x the spacetime coordinates, by $\phi(x)$ a generic field and by $\partial_\mu\phi$ its derivatives, we can define a Lagrangian density $\mathcal{L}(\phi, \partial_\mu)$. The Lagrangian density is deeply connected to the action S, which is a scalar quantity describing how a physical system has changed over time

$$S = \int L dt = \int \mathcal{L}(\phi, \partial_\mu\phi) d^4x \tag{1.1}$$

1.1.2 Noether's theorem

Many of the quantities used by physicists to describe natural phenomena—such as energy, momentum, or angular momentum—are deeply connected to underlying symmetries. This connection is formalized by Noether's theorem [5], which states that every continuous symmetry of the action S of a physical system with conservative forces corresponds to a conserved quantity.
In classical field theory, one can distinguish between two classes of continuous symmetries: (i) *spacetime symmetries*, which act on the spacetime coordinates (e.g. translations $x_\mu \to x_\mu + a_\mu$), and (ii) *internal symmetries*, which act on the fields themselves (e.g. $\phi(x) \to \phi(x) + \alpha\,\Delta\phi(x)$).
In the case of spacetime translations, the field transformation can be written as $\phi(x) \to \phi(x) + a_\mu\partial^\mu\phi(x)$. Since the Lagrangian density is a scalar quantity, it transforms accordingly as $\mathcal{L} \to \mathcal{L} + a_\mu\partial^\mu\mathcal{L}$. A change of the Lagrangian by a total derivative (a surface term) leaves the equations of motion unchanged, as such terms vanish under the derivatives in the Euler-Lagrange equation. Comparing the expected variation of $\mathcal{L}$ with the result obtained by varying the fields gives

$$a_\mu\partial^\mu\mathcal{L} = a_\nu\partial^\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\,\partial_\nu\phi\right) + a_\nu\left[\frac{\partial\mathcal{L}}{\partial\phi} - \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\right)\right]\partial_\nu\phi = 0. \tag{1.2}$$

The term in square brackets vanishes due to the Euler–Lagrange equations, leaving

$$a_\nu \partial^\mu \left(\delta^\nu_\mu \mathcal{L} - \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \partial_\nu \phi \right) = 0, \qquad \partial^\mu T^\nu_\mu = 0, \tag{1.3}$$

where T^ν_μ is identified as the *energy–momentum tensor* of the field ϕ.
Equation 1.3 expresses the conservation of four independent currents—one for each space-time component ν. For time translations ($\nu = 0$), the conserved quantity corresponds to the Hamiltonian or total energy of the system ($\partial^0 T^0_0 = -\partial^i T^0_i$). For spatial translations ($\nu = i$, $i = 1, 2, 3$), the conserved quantities correspond to the components of the total momentum of the field configuration. More generally, the conserved charges associated with a given continuous symmetry (e.g. translations, rotations, or internal gauge transformations) also serve as the generators of that symmetry [6].
In the most general case, a conservation law for a quantity Q_a (often referred to as a *charge*) can be expressed using the divergence theorem as

$$Q_a = \int d^{n-1}x \, j^0_a, \tag{1.4}$$

where j^μ_a is the conserved current (a tensor whose rank depends on the specific transformation) and n denotes the dimensionality of the spacetime manifold.

1.1.3 Gauge transformations

From Maxwell's equations we know that electromagnetic fields are defined in terms of derivatives of the scalar and vector potentials, implying that infinitely many potential configurations can lead to the same physical observables. For instance, the electrostatic potential is defined only up to an arbitrary additive constant. In quantum field theory, a similar principle holds: the fundamental fields themselves are not directly observable, and different field configurations may produce identical measurable quantities. Transformations between such equivalent configurations are known as *gauge transformations*. Because these transformations leave all physical observables unchanged, they correspond to a symmetry of the system, known as a *gauge symmetry*, and are associated with conserved quantities through Noether's theorem.
Consider a theory invariant under a global $U(1)$ phase transformation, $\psi \to e^{i\alpha}\psi$. To describe local interactions, we extend this symmetry to depend on space-time, requiring that the transformation parameter varies as $\alpha \to \alpha(x)$, so that $\psi(x) \to e^{i\alpha(x)}\psi(x) \equiv C(x)\psi(x)$. In this case, the action of the ordinary derivative on the transformed field, $n^\mu \partial_\mu \psi(x) = \lim_{\varepsilon \to 0} \varepsilon^{-1}[\psi(x+\varepsilon n) - \psi(x)]$, no longer factors out as in the global case, since $\psi(x)$ and $\psi(x + \varepsilon n)$ transform differently under $C(x)$. The derivative term thus breaks local invariance.
To compare field values at neighbouring points consistently, we introduce the quantity $U(y, x)$, known as a *Wilson line* [7], which transforms under $U(1)$ as $U(y, x) \to e^{i\alpha(y)} U(y, x) e^{-i\alpha(x)}$ and satisfies $U(x, x) = \mathbb{1}$. Using this, we can define a *covariant derivative* D_μ that transforms in the same way as the field itself:

$$\eta^\mu D_\mu \psi(x) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \big[\psi(x + \varepsilon\eta) - U(x + \varepsilon\eta, x)\psi(x) \big]. \tag{1.5}$$

If $U(y, x)$ is continuous in both arguments, we may expand it for small ε. Since $U(1)$ is a Lie group, infinitesimal transformations near the identity can be written as

$$U(x + \varepsilon n, x) = \mathbb{1} - ie\varepsilon n^\mu A_\mu(x) + \mathcal{O}(\varepsilon^2),$$

where e is an arbitrary constant and the coefficient $A_\mu(x)$ defines a new vector field, known as an *Ehresmann connection* [8]. In the infinitesimal limit, the covariant derivative becomes

$$D_\mu\psi(x) = \partial_\mu\psi(x) + ieA_\mu(x)\psi(x), \tag{1.6}$$

and the connection transforms under a local gauge transformation as

$$A_\mu(x) \to A_\mu(x) - \frac{1}{e}\,\partial_\mu\alpha(x). \tag{1.7}$$

We can now construct Lagrangians that remain invariant under local gauge transformations. Any combination of ψ and its covariant derivatives is locally invariant, but we must also include a kinetic term for the gauge field A_μ that depends only on A_μ and its derivatives. The simplest such term is built from the commutator of covariant derivatives,

$$[D_\mu, D_\nu] = ie(\partial_\mu A_\nu - \partial_\nu A_\mu) = ieF_{\mu\nu}, \tag{1.8}$$

where $F_{\mu\nu}$ is the electromagnetic field-strength tensor.
This simple example forms the basis of quantum electrodynamics (QED), where ψ represents the electron field, A_μ the electromagnetic four-potential, and e the electric charge. Gauge fields thus acquire a clear geometric interpretation as connections ensuring local symmetry of the underlying field theory.

1.1.4 The Standard Model of Particle Physics

This section introduces the Standard Model (SM) of particle physics, the quantum field theory that describes the fundamental constituents of matter and their interactions. Requiring local gauge invariance of the SM implies that its Lagrangian remains unchanged under space–time–dependent transformations belonging to specific Lie groups. These local symmetry transformations dictate the form of the interactions and the structure of the underlying theory.
The Standard Model is invariant under the local gauge group

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_{Y_L}, \tag{1.9}$$

which unifies three of the four known fundamental forces: the strong, weak, and electromagnetic interactions. Incorporating the remaining fundamental force—gravity—into this same quantum field–theoretical framework remains one of the foremost challenges in modern theoretical physics.

Quantum Electrodynamics

The development of Quantum Electrodynamics (QED) represents the first major step toward the formulation of the modern Standard Model. We begin by considering the Lagrangian density of a free electron field, described by the Dirac Lagrangian [9, 10]. The spinor ψ, which transforms under the $(1/2,0)\oplus(0,1/2)$ representation of the Lorentz group, is governed by $\mathcal{L} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi$, where $\bar{\psi} = \psi^\dagger\gamma^0$ and m denotes the Dirac mass. The γ^μ matrices in Minkowski space are at least 4×4 and satisfy the Dirac algebra $\{\gamma^\mu,\gamma^\nu\} = 2\eta^{\mu\nu}\mathbb{1}_{n\times n}$. A representation of the Lorentz algebra can be written in terms of these matrices as $\mathcal{S}^{\mu\nu} = i/4\,[\gamma^\mu,\gamma^\nu]$, where $\mathcal{S}^{0i}$ generate spatial rotations and $\mathcal{S}^{ij}$ generate

Lorentz boosts. A four-component field transforming according to $\mathcal{S}^{\mu\nu}$ is referred to as a Dirac spinor.
In the chiral basis, the diagonal structure of the γ matrices reveals that this representation is reducible. The spinor ψ can therefore be decomposed into two two-component objects, $\psi = (\psi_L, \psi_R)$, corresponding to eigenstates of the chirality operator γ_5:

$$\psi_L = \frac{1-\gamma_5}{2}\psi, \qquad \psi_R = \frac{1+\gamma_5}{2}\psi. \tag{1.10}$$

The free Lagrangian is invariant under a global $U(1)$ transformation. As discussed in Chapter 1.1.3, promoting this symmetry to a local one requires the replacement of ordinary derivatives ∂_μ with covariant derivatives D_μ that transform in the same way as ψ. This introduces a new interaction term involving the gauge field A_μ: $\mathcal{L} = i\bar{\psi}(x)(\gamma^\mu\partial_\mu - m)\psi(x) - e\bar{\psi}(x)\gamma^\mu A_\mu(x)\psi(x)$. The requirement of local gauge invariance also dictates the form of the kinetic term for the gauge field, which is constructed from the field-strength tensor $F^{\mu\nu}$ introduced in Chapter 1.1.3.
To recover classical electrodynamics in the appropriate limit, a physical spin-1 description of the gauge field can be written in terms of the Proca Lagrangian [11, 12], together with its coupling to fermions via the covariant derivative. However, local gauge invariance forces the mass term for the gauge field to vanish, as expected for the photon. The resulting expression is the familiar QED Lagrangian:

$$\mathcal{L}_{QED} = \bar{\psi}(x)(i\gamma^\mu D_\mu - m)\psi(x) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{1.11}$$

The coupling constant e is identified with the electric charge, whose conservation follows from Noether's theorem (see Chapter 1.1.2). Applying the Euler-Lagrange equations to this Lagrangian yields the Dirac equation in the presence of an electromagnetic field, $i\gamma^\mu\partial_\mu\psi - m\psi = e\gamma^\mu A_\mu\psi$.

Quantum Chromodynamics

The remaining sectors of the Standard Model are constructed following the same guiding principles used in QED, but with increasingly rich symmetry structures. Yang-Mills theories [13] extend QED by promoting the gauge symmetry to non-Abelian groups, while retaining renormalizability [14, 15].
Quantum Chromodynamics (QCD) describes the strong interaction and is based on the non-Abelian symmetry group $SU(3)_C$. This group has eight independent generators, corresponding to eight massless spin-1 gauge bosons. Because the generators do not commute, the conserved charges implied by Noether's theorem are gauge dependent, and the gauge bosons themselves carry color charge. The associated fundamental charge is called color and appears in $N = 3$ types.
The matter fields interacting via the strong force are the quarks, denoted $q_f(x)$, which are spin-$\frac{1}{2}$ fermions carrying both fractional electric charge and color charge. They obey the Dirac equation and come in different flavors f. The covariant derivative associated with a local $SU(3)$ symmetry takes the form $D_\mu = \partial_\mu - ig_S G^a_\mu \lambda^a/2$, where G^a_μ are the eight gluon gauge fields. The corresponding field-strength tensor, $G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_S f^{abc} G^b_\mu G^c_\nu$, contains a self-interaction term arising from the structure constants f^{abc} of the $SU(3)$ Lie algebra. As in QED, the gauge-invariant kinetic term for these fields is $-G^a_{\mu\nu}G^{a\mu\nu}/4$.

The full QCD Lagrangian density [16] is therefore constructed by combining this kinetic term with the Dirac Lagrangian for quarks:

$$\mathcal{L}_{QCD} = \sum_f \bar{q}_f(x)(i\gamma^\mu D_\mu - m_f)\, q_f(x) - \frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu}. \tag{1.12}$$

Color-charged particles cannot be observed in isolation. When the strong coupling becomes large, QCD exhibits confinement [7], implying that observable asymptotic states must form color-singlet combinations [17].
The QCD Lagrangian possesses additional global $U(1)$ and $SU(N_f)$ symmetries. The $U(1)$ transformation rotates left- and right-handed quarks by a common phase, and its associated conserved current corresponds to baryon number B, i.e. the net number of quarks is conserved in QCD interactions. The $SU(N_f)$ symmetry acts in flavor space, reflecting the fact that strong interactions are flavor blind. It can be decomposed into isospin symmetry, with charges (I, I_3), and a chiral symmetry in which the left- and right-handed components of the quark fields transform independently, because QCD preserves parity.
However, the $SU(N_f)$ chiral symmetry is spontaneously broken [18], as the QCD vacuum is not invariant under the same symmetries as the Lagrangian. The vacuum expectation value of the quark condensate, $\langle 0|\bar{Q}Q|0\rangle = \langle 0|\bar{Q}_L Q_R + \bar{Q}_R Q_L|0\rangle \neq 0$, explains why only about 1% of the proton mass originates from its constituent quarks. For $N_f = 2$, the near degeneracy of the up and down quark masses ($m_u \simeq m_d$) implies that $SU(2)$ is an approximate symmetry of QCD. For $N_f = 3$, one can extend this symmetry to include the strange quark, but the significantly larger strange-quark mass means that $SU(3)$ flavor symmetry is only approximate.

Weak Interactions and Electroweak unification

The existence of a third fundamental interaction was first inferred in the early 1900s through studies of radioactive decays. Experimental observations puzzled the physics community for decades, until Fermi proposed an interpretation of neutron β-decay as a four-fermion contact interaction [19]. Subsequent experiments by C. S. Wu et al. [20] demonstrated that the left- and right-handed components defined in Eq. 1.10 behave differently under the weak interaction, establishing that weak processes violate spatial parity, $P : x \to -x$. Later, Christenson et al. [21] showed that the combined operation of charge conjugation and parity (CP) is also violated by observing the decays of the long-lived neutral kaon K^0_L.
The conserved Noether current associated with weak interactions must therefore be maximally parity violating and can be written in a vector minus axial-vector ($V - A$) form. For leptons l, this current takes the form

$$J_{\mu,l} = \bar{\psi}_{\nu_l}\gamma_\mu(1-\gamma_5)\psi_l = J^V_\mu - J^A_\mu = 2\,\bar{\psi}^L_{\nu_l}\gamma_\mu\psi^L_l,$$

making it explicit that right-handed fermions do not participate in charged-current weak interactions.
Formulating a theory of charged weak currents within a gauge framework requires at least two generators. No such group exists, but the simplest suitable choice is $SU(2)$, which has three generators. An $SU(2)$ gauge theory naturally predicts two charged gauge bosons, $W^\pm$, with purely left-handed couplings, as well as a neutral boson, Z, coupling to

both chiralities. Since the weak interaction is short ranged, its mediators must be massive spin-1 bosons to reproduce Fermi's four-fermion theory at low energies. In contrast, QED is mediated by the massless photon, resulting in an infinite-range force. Additionally, fermion mass terms of the form $-m_l(\bar{l}_L l_R + \bar{l}_R l_L)$ are forbidden by gauge invariance, as l_L and l_R transform differently under $SU(2)$ and carry different $U(1)$ charges.
These difficulties were resolved through the electroweak unification formulated by Glashow [22], Salam [23], and Weinberg [24], which incorporates spontaneous symmetry breaking via the Higgs mechanism [25, 26, 27]. The electroweak theory is chiral: left- and right-handed fermions transform differently under the gauge group. The left-handed components form $SU(2)_L$ doublets, while the right-handed components are singlets. We can therefore define the left-handed quark and lepton doublets as

$$q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix} \qquad l_L = \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}, \begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix}, \begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix}. \tag{1.13}$$

We consider for the electroweak Yang-Mills theory a gauge symmetry group $SU(2)_L \times U(1)_{Y_L}$, whose generators are denoted weak isospin T^a and weak hypercharge Y_L, respectively. The associated gauge fields are W_μ^a for $SU(2)_L$ and B_μ for $U(1)_{Y_L}$. Before electroweak symmetry breaking, the Lagrangian density consists of four contributions, $\mathcal{L}_{EW} = \mathcal{L}_g + \mathcal{L}_f + \mathcal{L}_h + \mathcal{L}_y$, corresponding to the gauge kinetic terms, the kinetic terms for Standard Model fermions, the Higgs-sector Lagrangian, and Yukawa interactions. Focusing on the first two pieces, we obtain

$$\mathcal{L}_{g+f} = -\frac{1}{4} W_a^{\mu\nu} W_{\mu\nu}^a - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} + \bar{\psi}_L \, i\gamma^\mu D_\mu \psi_L + \bar{\psi}_R \, i\gamma^\mu D_\mu \psi_R, \tag{1.14}$$

where $B_{\mu\nu}$ and $W_{\mu\nu}^a$ are the field-strength tensors introduced in Chapters 1.1.4 and 1.1.4. For a fermion in a generic $SU(2)$ representation, the covariant derivative is given by

$$D_\mu = \partial_\mu - ig' Y_L B_\mu - igT^a W_\mu^a. \tag{1.15}$$

The generator of electric charge is defined by $Q = T^3 + Y_L$.
We now derive the mass eigenstates of the theory. For the charged mediators, only $SU(2)$ gauge fields mix, leading to $W^\pm = (W^1 \pm iW^2)/\sqrt{2}$. The remaining component W_μ^3 mixes with the $U(1)_{Y_L}$ field B_μ to form the neutral mass eigenstates Z_μ and A_μ via an $SO(2)$ rotation parameterized by the Weinberg angle [28] θ_ω: $(Z_\mu, A_\mu) = R(\theta_\omega)(W_\mu^3, B_\mu)$. Expressed in this basis, the covariant derivative becomes

$$D_\mu = \partial_\mu - i\frac{g}{\sqrt{2}} \left(W_\mu^+ T^+ + W_\mu^- T^- \right) - \frac{ig}{\cos\theta_\omega} Z_\mu \left(T^3 - \sin^2\theta_\omega \, Q \right) - ieA_\mu Q, \tag{1.16}$$

where the charged generators are defined as $T^\pm = T^1 \pm iT^2$. The Weinberg angle is determined by the requirement that QED couples universally to left- and right-handed fermions with charge e, which implies $g \sin\theta_\omega = g' \cos\theta_\omega = e$ and therefore $\tan\theta_\omega = g'/g$. The charged current interaction Lagrangian then takes the form [29]

$$\mathcal{L}_{CC} = \frac{g}{\sqrt{2}} \left[\bar{\psi}_L T^+ W_\mu^+ \psi_L + \bar{\psi}_L T^- W_\mu^- \psi_L \right], \tag{1.17}$$

while the neutral sector contains both the familiar electromagnetic interaction mediated by A_μ and an additional weak neutral current mediated by Z_μ:

$$\mathcal{L}_{NC} = \frac{g}{2\cos\theta_\omega} \left[\bar{\psi}_L T^3 \gamma_\mu \psi_L - 2\,\bar{\psi} Q \sin^2\theta_\omega \gamma_\mu \psi \right] Z^\mu. \tag{1.18}$$

Several properties of weak interactions follow immediately. Charged weak interactions are the only known interactions capable of changing particle flavor. Weak neutral currents can couple to both left- and right-handed fermions, although flavor-changing processes mediated in this way are highly suppressed.

Spontaneous symmetry breaking and the Higgs mechanism

The problem of generating masses for the electroweak gauge bosons and fermions while preserving gauge invariance is resolved by the Brout–Englert–Higgs (BEH) mechanism, described by the Lagrangian term $\mathcal{L}_h$. Consider a system of complex scalar fields $\phi \equiv (\phi_u, \phi_d)$ transforming as a doublet under $SU(2)$. Under an infinitesimal $SU(2)$ transformation, the change in the field is given by $\delta S(k) \equiv \phi' - \phi \simeq i\,\varepsilon_k(x)[\tau_k/2, \phi]$, where τ_k $(k = 1, 2, 3)$ are the Pauli matrices, generators of the $SU(2)$ group, and $\varepsilon_k(x)$ are infinitesimal parameters. This relation remains valid when taking the vacuum expectation value (v.e.v.) of the field, $\langle 0|\phi|0\rangle = \langle\phi\rangle_0$. If at least one component of ϕ acquires a non-zero v.e.v., the vacuum is no longer invariant under the action of the generators, i.e. $\tau_k|0\rangle \neq 0$. The Lagrangian remains $SU(2)$-invariant, but the vacuum does not — a phenomenon known as spontaneous symmetry breaking.
An $SU(2)$-invariant Lagrangian exhibiting this behaviour can be constructed by introducing a potential of the Coleman–Weinberg form [30]:

$$\mathcal{L}_h = (D_\mu\phi)^\dagger(D^\mu\phi) + \underbrace{\mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2}_{-V(\phi^\dagger\phi)}. \tag{1.19}$$

The conditions $\lambda > 0$ (stability) and $\mu^2 > 0$ (non-zero minimum) ensure that the potential has a degenerate set of minima, located at $|\phi| = e^{i\rho}\sqrt{\mu^2/2\lambda}$, as illustrated in Figure 1.1. Choosing the vacuum alignment such that $\langle\phi_d\rangle_0 = v/\sqrt{2}$ and $\langle\phi_u\rangle_0 = 0$, the field can be expanded around the new minimum as

$$\phi(x) = \frac{e^{i\theta_k\tau_k/2}}{\sqrt{2}}\begin{pmatrix}0\\ v+h(x)\end{pmatrix} \approx \frac{1}{\sqrt{2}}\begin{pmatrix}\eta_1(x)+i\eta_2(x)\\ v+h(x)+i\eta_3(x)\end{pmatrix} \xrightarrow{SU(2)} \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v+h(x)\end{pmatrix}, \tag{1.20}$$

where $\eta_i(x)$ are the massless spin-0 Nambu–Goldstone bosons [31, 18] associated with the spontaneously broken generators. Since no such massless scalar particles are observed, they must be absorbed by the gauge fields. This occurs because the $SU(2)$ symmetry is local: the would-be Goldstone bosons become the longitudinal components of the now-massive gauge bosons. This absorption does not occur for a purely global symmetry. In the unitary gauge, the exponential term in Eq. 1.20 is removed through a gauge rotation, leaving only the physical scalar degree of freedom $h(x)$, the Higgs field. The corresponding potential takes the form $V(\phi^\dagger\phi) = -\mu^2h^2 - \lambda v h^3 - \frac{\lambda}{4}h^4$, revealing a scalar particle with mass $m_h = \sqrt{2\lambda}\,v$.
The origin of the electroweak gauge boson masses can be traced to the kinetic term $(D_\mu\phi)^\dagger(D^\mu\phi)$ under the $SU(2)_L \times U(1)_{Y_L}$ symmetry. Evaluating this term at the vacuum expectation value yields

$$\begin{aligned}\left\langle (D_\mu\phi)^\dagger(D^\mu\phi)\right\rangle_0 &= \frac{1}{2}\begin{pmatrix}0 & v\end{pmatrix}\left(gW_\mu^aT^a + g'Y_LB_\mu\right)^2\begin{pmatrix}0\\ v\end{pmatrix}\\ &= \frac{v^2}{8}\left[g^2(W_\mu^1)^2 + g^2(W_\mu^2)^2 + (-gW_\mu^3 + g'B_\mu)^2\right].\end{aligned} \tag{1.21}$$

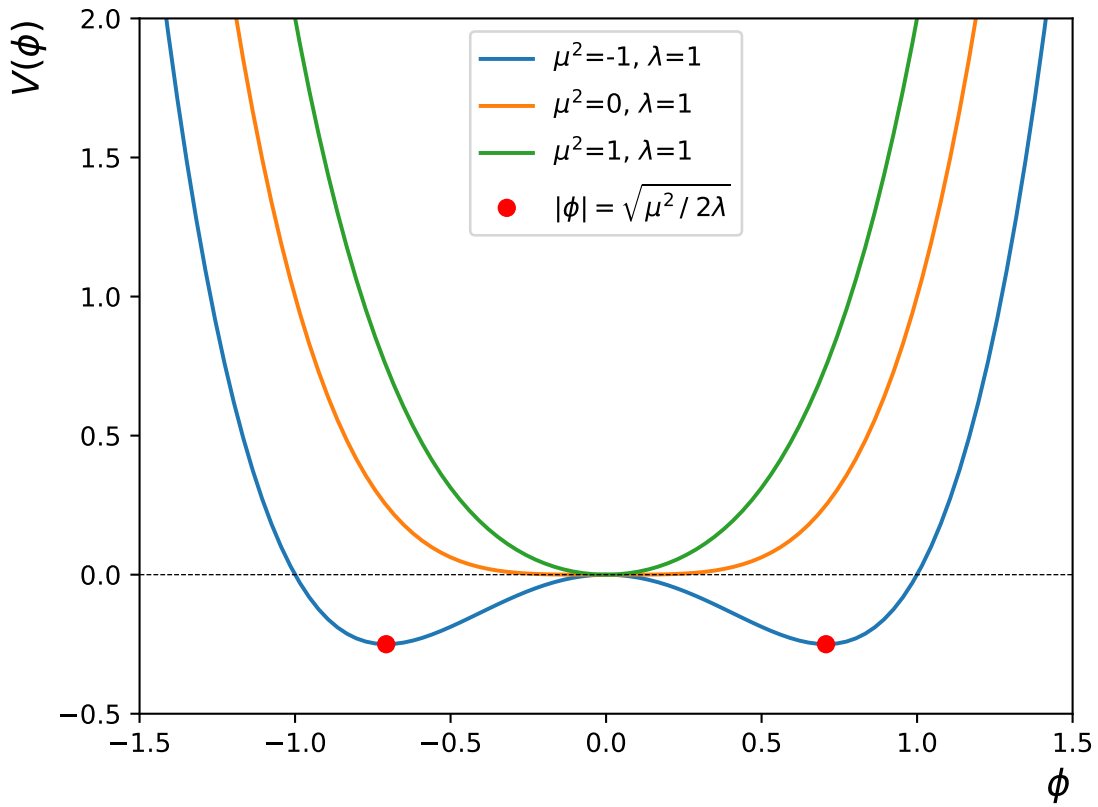

Figure 1.1: Projection of the Coleman-Weinberg potential for a complex scalar field ϕ on the real axis. While the λ parameter is fixed to 1, the potential is drawn for $\mu^2 = -1, 0, 1$ respectively in blue, orange and green showing how only $\mu^2 > 0$ induces a degeneracy of the minima.

This expression reveals three massive gauge bosons and one massless boson:

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \qquad m_W = \frac{gv}{2}, \\ Z_\mu &= \frac{1}{\sqrt{g^2+g'^2}}(gW_\mu^3 - g'B_\mu), \qquad m_Z = \frac{v}{2}\sqrt{g^2+g'^2}, \\ A_\mu &= \frac{1}{\sqrt{g^2+g'^2}}(g'W_\mu^3 + gB_\mu), \qquad m_A = 0. \end{aligned} \tag{1.22}$$

Finally, fermion masses arise from Yukawa interactions between the Higgs field and fermion fields. Gauge-invariant terms coupling left-handed doublets and right-handed singlets can be written as $\bar{\psi}_L \Gamma \psi_R \phi$, where Γ includes the appropriate coupling constants ensuring Lorentz and gauge invariance. After symmetry breaking, the Yukawa Lagrangian becomes

$$\mathcal{L}_y = -\frac{v+h}{\sqrt{2}} \sum_f y_f \bar{\psi}_L^f \psi_R^f, \tag{1.23}$$

where y_f are the Yukawa couplings, proportional to the fermion masses as $m_f = y_f v/\sqrt{2}$.

	ψ	Mass	Q	Color	$\mathbf{T}_3$	$\mathbf{Y_L}$
$\mathbf{1}^{st}$	$\boldsymbol{\nu_e}$	< 225 eV	0	0	L:+$\frac{1}{2}$; R:0	L:-1; R:0
	$\boldsymbol{e}$	511 keV	-1	0	L:-$\frac{1}{2}$; R:0	L:-1; R:-2
	$\boldsymbol{u}$	$2.16^{+0.49}_{-0.26}$ MeV	1/3	r, g, b	L:+$\frac{1}{2}$; R:0	L:+$\frac{1}{3}$; R:+$\frac{4}{3}$
	$\boldsymbol{d}$	$4.67^{+0.48}_{-0.17}$ MeV	-2/3	r, g, b	L:-$\frac{1}{2}$; R:0	L:+$\frac{1}{3}$; R:-$\frac{2}{3}$
$\mathbf{2}^{nd}$	$\boldsymbol{\nu_\mu}$	< 0.19 MeV	0	0	L:+$\frac{1}{2}$; R:0	L:-1; R:0
	$\boldsymbol{\mu}$	105.7 MeV	-1	0	L:-$\frac{1}{2}$; R:0	L:-1; R:-2
	$\boldsymbol{c}$	1.27 ± 0.02 MeV	1/3	r, g, b	L:+$\frac{1}{2}$; R:0	L:+$\frac{1}{3}$; R:+$\frac{4}{3}$
	$\boldsymbol{s}$	$93.4^{+8.6}_{-3.4}$ MeV	-2/3	r, g, b	L:-$\frac{1}{2}$; R:0	L:+$\frac{1}{3}$; R:-$\frac{2}{3}$
$\mathbf{3}^{rd}$	$\boldsymbol{\nu_\tau}$	< 18.2 MeV	0	0	L:+$\frac{1}{2}$; R:0	L:-1; R:0
	$\boldsymbol{\tau}$	1.777 GeV	-1	0	L:-$\frac{1}{2}$; R:0	L:-1; R:-2
	$\boldsymbol{t}$	172.69 ± 0.30 GeV	1/3	r, g, b	L:+$\frac{1}{2}$; R:0	L:+$\frac{1}{3}$; R:+$\frac{4}{3}$
	$\boldsymbol{b}$	$4.18^{+0.03}_{-0.02}$ GeV	-2/3	r, g, b	L:-$\frac{1}{2}$;R:0	L:+$\frac{1}{3}$; R:-$\frac{2}{3}$

Table 1.1: Leptons (e, μ, τ) and quarks (u, d, c, s, t, b) masses [32] and charges as spin 1/2 fermions

Thus, the Higgs mechanism provides a unified explanation for the origin of both gauge boson and fermion masses while preserving the gauge symmetry of the Standard Model.

Gauge and matter fields of the Standard Model

The field content of the Standard Model can be grouped into two main categories: *matter fields* and *gauge fields*. The matter fields comprise all spin-$\frac{1}{2}$ fermions, which obey Fermi-Dirac statistics, and are organized into leptons and quarks. The gauge fields, instead, are spin-1 bosons obeying Bose-Einstein statistics and mediate the fundamental interactions associated with the $SU(3)_C \times SU(2)_L \times U(1)_{Y_L}$ gauge symmetry. In addition, the model includes a scalar field, the Higgs field, responsible for electroweak symmetry breaking and for generating particle masses through the Yukawa mechanism.
The fermionic sector is arranged in three generations of quarks and leptons. Each generation exhibits identical gauge quantum numbers but differs in mass, a structure that underlies the observed pattern of weak interactions and flavor mixing. Table 1.1 summarizes the spin-$\frac{1}{2}$ matter fields and their quantum numbers, while Table 1.2 lists the gauge bosons and the Higgs field with their main properties.

1.1.5 Vector Boson Scattering as a probe of electroweak symmetry breaking

Following the discussion of spontaneous electroweak symmetry breaking and the construction of the gauge and matter sectors of the Standard Model (SM), it is natural to ask how the dynamics of the electroweak sector can be tested at high energies. Since the discovery of a new scalar boson in 2012 by the ATLAS and CMS collaborations [33, 34], its measured properties—including its couplings to SM particles—have so far been found compatible with those of the Higgs boson predicted by the SM [35, 36]. An independent

	Spin	Mass	Q	Color	$\mathbf{T_3}$	$\mathbf{Y_L}$
Photon (γ)	1	$< 10^{-18}$ eV	0	0	1,0	0
Gluon (g)	1	~ 0	0	r, g, b	0	0
$\mathbf{Z^0}$	1	91.1876 ± 0.0021 GeV	0	0	1,0	0
$\mathbf{W^{\pm}}$	1	80.377 ± 0.012 GeV	± 1	0	1	0
h	0	125.25 ± 0.17 GeV	0	0	0	0

Table 1.2: Gauge bosons spin, masses [32] and charges

and particularly incisive handle on the electroweak symmetry breaking (EWSB) mechanism is provided by the high-energy behaviour of electroweak vector bosons, especially in scattering processes involving longitudinally polarised states.
Vector Boson Scattering (VBS) denotes the scattering of two electroweak gauge bosons, schematically $VV \to VV$, where $V = W^{\pm}, Z$. In hadronic collisions such processes are accessed through diagrams in which the initial-state quarks radiate electroweak bosons, leading to a characteristic "vector-boson-fusion" topology with two energetic jets at large rapidity separation and a central diboson system. From a theoretical perspective, the interest of VBS stems from the fact that the scattering amplitudes are directly sensitive to the non-Abelian structure of the electroweak theory and to the scalar sector responsible for EWSB.
A central role is played by the longitudinal polarisations. In the absence of the Higgs boson exchange, the amplitude for longitudinal vector boson scattering, $V_L V_L \to V_L V_L$, grows with the center-of-mass energy and violates perturbative unitarity at scales of order the TeV [37, 38]. In the SM this potentially problematic energy growth is cancelled by the contribution of diagrams containing Higgs exchange, restoring a unitary high-energy behaviour [39, 40]. This cancellation is a direct consequence of the gauge structure together with the Higgs mechanism, and therefore provides a conceptually clean test of EWSB that is complementary to on-shell Higgs measurements. Any deviation of the Higgs–gauge couplings from their SM values would spoil this delicate compensation, reintroducing an energy-dependent enhancement in VBS amplitudes and distorting kinematic spectra at high invariant masses [40], as illustrated in Figure 1.2.
Beyond its role as a consistency test of EWSB, VBS is also an excellent laboratory to probe the self-interactions of the electroweak gauge bosons. These interactions are encoded in the SM by the triple gauge couplings (TGC) and quartic gauge couplings (QGC), which arise from the non-Abelian structure of the $SU(2)_L \times U(1)_Y$ gauge symmetry. In VBS, both TGC and QGC enter already at tree level through diagrams with three- and four-gauge-boson vertices, and their interplay with Higgs-mediated contributions governs the high-energy behaviour of the amplitudes.
The same TGC structures that appear in VBS also play an important role in vector-boson-fusion (VBF) production modes, where a single electroweak boson or a Higgs boson is produced in association with two forward jets. In these processes, the electroweak currents radiated from the incoming quarks can interact through intermediate gauge bosons, and the corresponding amplitudes depend on the same underlying gauge vertices. As a result, precision measurements of both VBS and VBF topologies provide complementary sensitivity to possible deviations from the SM gauge structure.

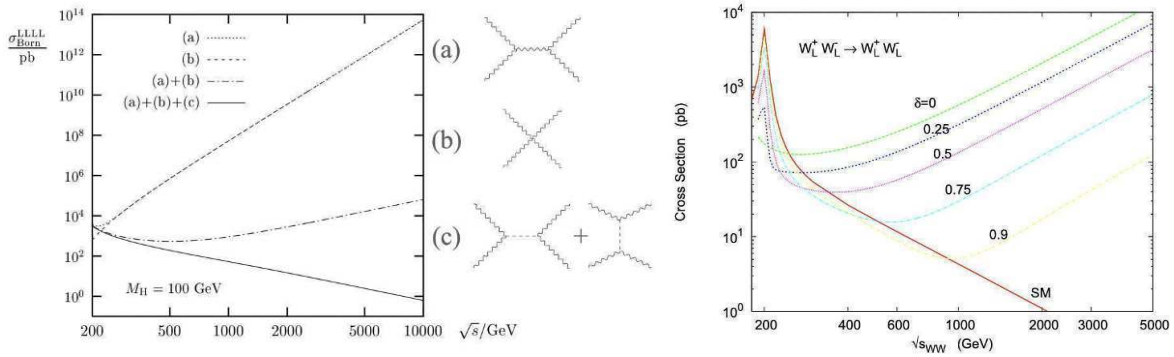

Figure 1.2: Left figure from [39] shows how the cross section for $V_LV_L \to V_LV_L$ diverges in the absence of Higgs-mediated diagrams (a)+(b) while is it unitarized by s and t-channel Higgs diagrams (a)+(b)+(c). In the right figure the authors of [40] assume that the hW^+W^- coupling is a fraction $\sqrt{\delta}$ of its SM value thus the cancellation of the energy growth of amplitudes is only partial and distorts the spectrum of the diboson invariant mass $\sqrt{s}_{WW}$ in the process $W_L^+W_L^- \to W_L^+W_L^-$

This motivates the effective-field-theory approach adopted in the next chapter. In particular, within the Standard Model Effective Field Theory (SMEFT), higher-dimensional operators can induce correlated modifications of Higgs–gauge interactions as well as of TGC and QGC, leading to measurable effects in the tails of kinematic distributions and in the overall rates of electroweak processes. The following chapter introduces the SMEFT framework and provides the tools to parameterize such deviations in a systematic expansion, enabling a consistent interpretation of VBS and VBF measurements in terms of Wilson coefficients.

1.2 Effective Field Theory

In the absence of direct evidence for new particles via resonant peaks in invariant mass distributions, indirect searches play a crucial role in exploring potential physics beyond the Standard Model. Such searches aim to detect deviations from Standard Model predictions in kinematic distributions induced by anomalous couplings between new particles and the Standard Model fields. However, these deviations are often not uniquely attributable to a single underlying theory, since multiple high-energy models can produce similar low-energy effects. This ambiguity arises from the decoupling theorem [41], which states that if the characteristic energy scale of new physics, Λ, is much larger than the experimentally accessible energy E, then processes involving particles of mass $\mathcal{O}(\Lambda)$ contribute negligibly to low-energy observables.
The same theorem implies that low-energy phenomena can be described without explicitly including heavy degrees of freedom, as their effects are suppressed by powers of E/Λ. This observation forms the foundation of the *Effective Field Theory* (EFT) approach, which provides a general, model-independent framework to parameterize possible deviations from the Standard Model. An EFT is a quantum field theory built on the same principles of Lorentz invariance and quantum mechanics, but valid only up to a finite energy scale Λ. In this formulation, all interactions consistent with the symmetries of the theory are included, organized as an expansion in inverse powers of Λ. A familiar historical example is the Fermi four-fermion interaction, which represents the low-energy limit of the electroweak theory.
EFTs allow one to compute measurable quantities with a well-defined and controllable theoretical uncertainty. Formally, this is achieved by extending the renormalizable Standard Model Lagrangian with higher-dimensional operators, each suppressed by increasing powers of $1/\Lambda$. This systematic expansion—known as the *power counting*—provides a consistent prescription for computing higher-order corrections and progressively improving the accuracy of theoretical predictions.

1.2.1 Counting dimensions and EFT renormalization

The Standard Model (SM) Lagrangian has a canonical mass dimension of four, $[\mathcal{L}] = 4$, with the covariant derivative carrying dimension $[D_\mu] = 1$. Consequently, all terms in the SM Lagrangian must also have total dimension four, for instance:

$$\begin{aligned} &[D_\mu \phi^\dagger D^\mu \phi] = 4 \quad \Rightarrow \quad [\phi] = 1, \\ &[X_{\mu\nu} X^{\mu\nu}] = 4 \quad \Rightarrow \quad [X_\mu] = 1, \\ &[\bar{\psi} D_\mu \psi] = 4 \quad \Rightarrow \quad [\psi] = \tfrac{3}{2}, \\ &[\bar{\psi} y_L \phi \psi] = 4 \quad \Rightarrow \quad [y_L] = 0. \end{aligned} \tag{1.24}$$

Extending the theory beyond renormalizable interactions introduces higher-dimensional operators that encode potential effects of new physics. These appear in the Lagrangian as additional terms of the form $\mathcal{L} \supset c\,\mathcal{O}^{(d)}$, where $\mathcal{O}^{(d)}$ is an operator of mass dimension d, and c is the corresponding Wilson coefficient quantifying its strength. While the SM Lagrangian ($d = 4$) is renormalizable, terms with $d > 4$ are not. Nevertheless, an Effective Field Theory (EFT) remains renormalizable *order by order* in its expansion: if truncated at a given operator dimension d, the theory provides predictions that are finite and well-defined up to corrections of higher order in $1/\Lambda$.

Consider a Feynman diagram containing a generic loop integral $\int F(p,\dots)\,d^4p$. The ultraviolet (UV) behaviour of this integral can be characterized by the *superficial degree of divergence* [16], D, defined such that the asymptotic form is $\int dp\, p^{D-1}$. Integrals with $D<0$ are convergent, those with $D=0$ are logarithmically divergent, and those with $D>0$ are power divergent. For a diagram involving a set of external lines E_f (for field species f with spin s_f), interaction vertices N_k, and operators of dimension d_k, one finds:

$$D = 4 - \sum_f E_f(1+s_f) + \sum_k N_k d_k. \tag{1.25}$$

If the loop is approximated by an EFT vertex in the limit of large internal mass, the insertion of one or two dimension-six operators ($N_k = 1, 2$ and $d_k = 6$) leads to divergent integrals with $D = 6$ and $D = 8$, respectively. To preserve renormalizability at a given order, counterterms of matching dimension must be introduced to absorb these divergences. The factor $-\sum_k N_k(d_k-4)$ precisely indicates the power of the mass dimension that regularizes the theory. Using the cutoff scale Λ as the expansion parameter, this scaling is expressed as $c\,\mathcal{O}^{(d)} \to c\,\mathcal{O}^{(d)}/\Lambda^{d-4}$.

1.2.2 The Standard Model Effective Field Theory

The *Standard Model Effective Field Theory* (SMEFT) [42] provides a systematic and gauge-invariant extension of the SM, constructed from all possible higher-dimensional operators built solely from SM fields. These operators are invariant under the full $SU(3)_C \times SU(2)_L \times U(1)_{Y_L}$ gauge symmetry and include the Higgs field h embedded in an $SU(2)$ doublet [43]. The SMEFT Lagrangian is organized as an expansion in canonical mass dimension:

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \mathcal{L}^{(5)} + \mathcal{L}^{(6)} + \mathcal{L}^{(7)} + \dots, \qquad \mathcal{L}^{(d)} = \sum_{i=1}^{n_d} \frac{c_i^{(d)}}{\Lambda^{d-4}}\,\mathcal{O}_i^{(d)}. \tag{1.26}$$

Here, Λ denotes the energy scale of new physics, assumed to be much larger than the masses of SM particles and the kinematic scales of the studied processes. The condition $(E/\Lambda) \ll 1$ ensures the decoupling of heavy states and the validity of the EFT expansion. At dimension five, $\mathcal{L}^{(5)}$ contains a single lepton-number-violating operator (and its Hermitian conjugate), known as the *Weinberg operator* [44]:

$$\mathcal{L}_5 = \frac{c_{ll'}}{\Lambda}\left[\phi\cdot\bar{l}_L^c\right]\left[l'_L\cdot\phi\right] + \text{h.c.},$$

which generates dimension-three Majorana mass terms $m_{ll'} = c_{ll'}v^2/\Lambda$ for the left-handed neutrinos l_L after spontaneous symmetry breaking [45].
At dimension six, and assuming three generations of fermions, there exist 2499 independent operators [46] that conserve both baryon and lepton numbers ($\Delta B = \Delta L = 0$). These provide the leading SMEFT corrections under the assumption that $\Lambda \ll \Lambda_{\not L}$, where $\Lambda_{\not L}$ denotes the scale of lepton-number violation. An additional 273 operators with $\Delta B = \Delta L = 1$ can also be constructed, allowing for baryon-number–violating processes such as proton decay. Higher-dimensional operators ($d = 7, 8, \dots$) further suppressed by powers of $1/\Lambda$ have also been studied extensively [47, 48, 49, 50, 51], and the number of operators at each mass dimension is summarized in Figure 1.3.

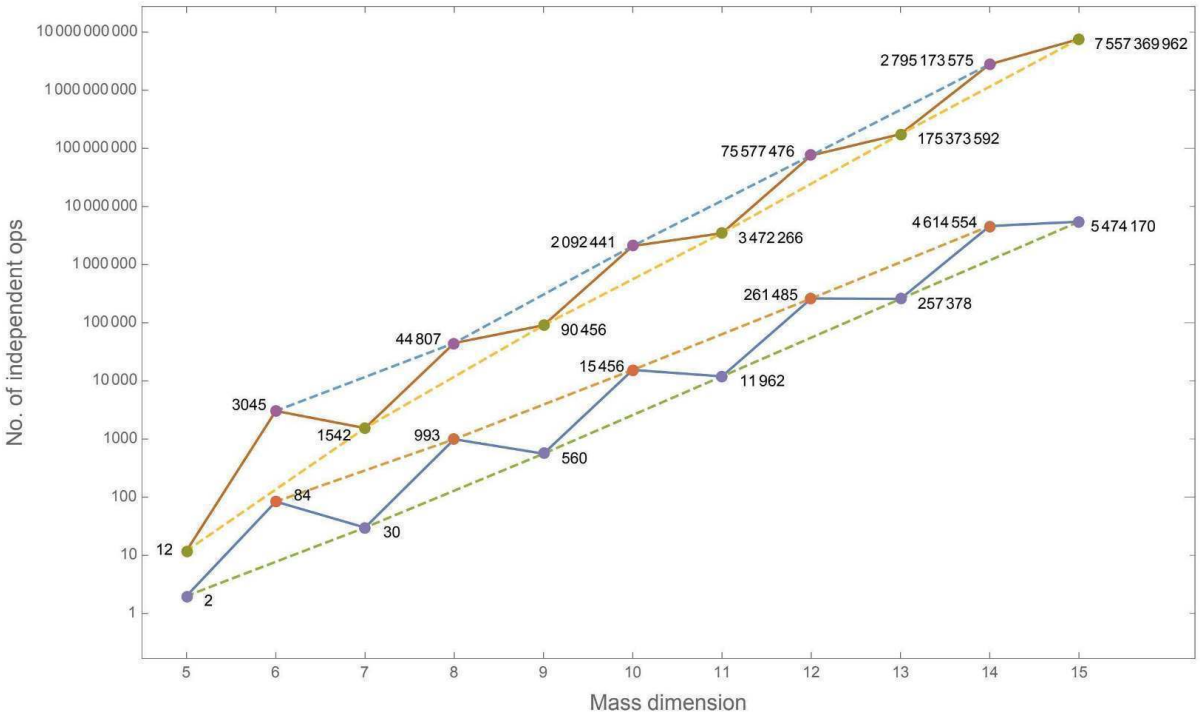

Figure 1.3: Evolution of the number of independent operators in the SMEFT as a function of the operator dimension [51]. For each dimension, the lower markers correspond to the case of a single fermion generation, while the upper markers represent the result for three generations. Dashed lines connect the values obtained for even and odd operator dimensions.

1.2.3 SMEFT at dimension six

The leading baryon- and lepton-number–conserving contributions in the SMEFT arise from dimension-six operators. To identify a minimal, non-redundant set of such operators, all possible combinations of Standard Model fields consistent with the gauge symmetry group $SU(3)_C \times SU(2)_L \times U(1)_{Y_L}$ must be constructed, and redundancies removed. These redundancies can be eliminated using integration by parts, momentum conservation, the equations of motion, and Fierz rearrangements.

Several incomplete operator bases have appeared in the literature, including the HISZ basis [52], the SILH basis [53], and the Buchmüller–Wyler basis [42]. However, the first complete and non-redundant formulation of the dimension-six Lagrangian, known as the *Warsaw basis* [46], identified 59 independent operators. Subsequent work has refined other bases, such as the SILH basis, to achieve full equivalence with the Warsaw basis [54]. The Warsaw basis is expressed in terms of SM fields transforming under $SU(2)_L$, as summarized in Table 1.3. The Higgs doublet is denoted by ϕ, while its conjugate $\tilde{\phi}$ is defined through the two-index Levi–Civita tensor ε_{jk} as $\tilde{\phi}^j = \varepsilon^{jk}(\phi_k)^*$. The Hermitian derivatives acting on ϕ are defined as

$$\phi^\dagger i \overleftrightarrow{D}_\mu \phi = \phi^\dagger i D_\mu \phi - i(D_\mu \phi)^\dagger \phi, \qquad \phi^\dagger i \overleftrightarrow{D}^i_\mu \phi = \phi^\dagger i \sigma^i D_\mu \phi - i(D_\mu \phi)^\dagger \sigma^i \phi,$$

where σ^i are the Pauli matrices. The gauge field-strength tensors, collectively denoted as $X_{\mu\nu}$, are defined as in the SM: $G_{\mu\nu}$ for $SU(3)_C$, $W^a_{\mu\nu}$ for $SU(2)_L$, and $B_{\mu\nu}$ for $U(1)_{Y_L}$, as described in Section 1.1.4.

	Fermions					Scalars
Field	$l^j_{L,p}$	$e_{R,p}$	$q^{\alpha,j}_{L,p}$	$u^\alpha_{R,p}$	$d^\alpha_{R,p}$	ϕ^j
$\boldsymbol{Y_L}$	$-\frac{1}{2}$	-1	$\frac{1}{6}$	$\frac{2}{3}$	$-\frac{1}{3}$	$\frac{1}{2}$

Table 1.3: Standard Model matter content in the $SU(2)$ representation. Here $j = 1, 2$ denotes weak isospin, $\alpha = r, g, b$ the color index, and $p = 1, 2, 3$ the fermion generation index.

Following the dimensional analysis introduced in Section 1.2.1, dimension-six operators can be systematically constructed from SM fields and categorized according to their field content. The bosonic operator classes include X^3, ϕ^6, $\phi^4 D^2$, and $X^2\phi^2$, while the fermionic classes comprise $\psi^2 X\phi$ and $\psi^2\phi^2 D$. Four-fermion contact interactions are further divided according to fermion chirality as $(\bar{L}L)(\bar{L}L)$, $(\bar{R}R)(\bar{R}R)$, $(\bar{L}L)(\bar{R}R)$, and the mixed structures $(\bar{L}R)(\bar{R}L)$ or $(\bar{L}R)(\bar{L}R)$.
The complete list of dimension-six operators in the Warsaw basis is reported in Table 1.4.

1.2.4 Electroweak input parameters

The Standard Model (SM) can provide quantitative predictions only once a set of experimentally measured quantities is used to fix the parameters of the theory. Assuming massless neutrinos, the SM requires 19 independent input parameters. Allowing for massive neutrinos introduces seven additional parameters, bringing the total to 26. These are summarized in Table 1.5, where θ_{ij} denote the mixing angles of the unitary Cabibbo–Kobayashi–Maskawa (CKM) matrix [55, 56], δ its CP-violating phase, and U_{ij} the elements of the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix [57, 58], with associated CP-violating phase δ', describing neutrino flavour mixing.
The Lagrangian parameters $g = \{g_1, \ldots, g_N\}$ are fixed by imposing a set of defining relations between the parameters and a collection of experimentally measured observables $\mathcal{O} = \{\mathcal{O}_1, \ldots, \mathcal{O}_M\}$, with $M \geq N$. At a chosen perturbative order, these relations can be expressed as $\mathcal{O}_n = F_n^{(0)}(g)$ for $n = 1, \ldots, M$, where $F_n^{(0)}$ denotes the functional dependence of the observable on the parameters. If the number of observables equals the number of parameters ($M = N$), the system can be inverted, yielding

$$g_i = K_i^{(0)}(\mathcal{O}), \qquad i = 1, \ldots, N.$$

Extending the SM to the SMEFT introduces new parameters: the cutoff scale Λ and the Wilson coefficients $c_i^{(d)}$. In this framework, the same observables $\mathcal{O}$ used to fix the SM parameters now receive higher-order EFT corrections:

$$\begin{aligned} \mathcal{O}_n &= F_n^{(0)}(g) + \frac{1}{\Lambda^2} F_n^{(2)}(g, c^{(6)}) + \frac{1}{\Lambda^4} F_n^{(4)}(g, c^{(8)}) + \ldots, \\ g_i &= K_i^{(0)}(\mathcal{O}) + \frac{1}{\Lambda^2} K_i^{(2)}(\mathcal{O}, c^{(6)}) + \frac{1}{\Lambda^4} K_i^{(4)}(\mathcal{O}, c^{(8)}) + \ldots \end{aligned} \tag{1.27}$$

A generic predicted observable $\mathcal{P}$ can then be expressed as

$$\mathcal{P} = P^{(0)}(g) + \sum_{d=2,4,8,\ldots} \frac{P^{(d)}(g, c^{(4+d)})}{\Lambda^d},$$

X^3		ϕ^6 and $\phi^4 D^2$		$\psi^2\phi^3$	
Q_G	$f^{abc}G^{a\nu}_{\mu}G^{b\rho}_{\nu}G^{c\mu}_{\rho}$	Q_{ϕ}	$(\phi^\dagger\phi)^3$	$Q_{e\phi}$	$(\phi^\dagger\phi)(\bar{l}_p e_r \phi)$
$Q_{\tilde{G}}$	$f^{abc}\tilde{G}^{a\nu}_{\mu}G^{b\rho}_{\nu}G^{c\mu}_{\rho}$	$Q_{\phi\Box}$	$(\phi^\dagger\phi)\Box(\phi^\dagger\phi)$	$Q_{u\phi}$	$(\phi^\dagger\phi)(\bar{q}_p u_r \tilde{\phi})$
Q_W	$\varepsilon^{ijk}W^{i\nu}_{\mu}W^{j\rho}_{\nu}W^{k\mu}_{\rho}$	$Q_{\phi D}$	$(D^\mu\phi^\dagger\phi)(\phi^\dagger D_\mu\phi)$	$Q_{d\phi}$	$(\phi^\dagger\phi)(\bar{q}_p d_r \phi)$
$Q_{\tilde{W}}$	$\varepsilon^{ijk}\tilde{W}^{i\nu}_{\mu}W^{j\rho}_{\nu}W^{k\mu}_{\rho}$				
$X^2\phi^2$		$\psi^2 X\phi$		$\psi^2\phi^2 D$	
$Q_{\phi G}$	$\phi^\dagger\phi G^a_{\mu\nu}G^{a\mu\nu}$	Q_{eW}	$(\bar{l}_p\sigma^{\mu\nu}e_r)\sigma^i\phi W^i_{\mu\nu}$	$Q^{(1)}_{\phi l}$	$(\phi^\dagger i\overleftrightarrow{D}_\mu\phi)(\bar{l}_p\gamma^\mu l_r)$
$Q_{\phi\tilde{G}}$	$\phi^\dagger\phi\tilde{G}^a_{\mu\nu}G^{a\mu\nu}$	Q_{eB}	$(\bar{l}_p\sigma^{\mu\nu}e_r)\phi B_{\mu\nu}$	$Q^{(3)}_{\phi l}$	$(\phi^\dagger i\overleftrightarrow{D}^i_\mu\phi)(\bar{l}_p\sigma^i\gamma^\mu l_r)$
$Q_{\phi W}$	$\phi^\dagger\phi W^a_{\mu\nu}W^{a\mu\nu}$	Q_{uG}	$(\bar{q}_p\sigma^{\mu\nu}T^a u_r)\tilde{\phi}G^a_{\mu\nu}$	$Q_{\phi e}$	$(\phi^\dagger i\overleftrightarrow{D}_\mu\phi)(\bar{e}_p\gamma^\mu e_r)$
$Q_{\phi\tilde{W}}$	$\phi^\dagger\phi\tilde{W}^a_{\mu\nu}W^{a\mu\nu}$	Q_{uW}	$(\bar{q}_p\sigma^{\mu\nu}u_r)\sigma^i\tilde{\phi}W^i_{\mu\nu}$	$Q^{(1)}_{\phi q}$	$(\phi^\dagger i\overleftrightarrow{D}_\mu\phi)(\bar{q}_p\gamma^\mu q_r)$
$Q_{\phi B}$	$\phi^\dagger\phi B_{\mu\nu}B^{\mu\nu}$	Q_{uB}	$(\bar{q}_p\sigma^{\mu\nu}u_r)\tilde{\phi}B_{\mu\nu}$	$Q^{(3)}_{\phi q}$	$(\phi^\dagger i\overleftrightarrow{D}^i_\mu\phi)(\bar{q}_p\sigma^i\gamma^\mu q_r)$
$Q_{\phi\tilde{B}}$	$\phi^\dagger\phi\tilde{B}_{\mu\nu}B^{\mu\nu}$	Q_{dG}	$(\bar{q}_p\sigma^{\mu\nu}T^a d_r)\phi G^a_{\mu\nu}$	$Q_{\phi u}$	$(\phi^\dagger i\overleftrightarrow{D}_\mu\phi)(\bar{u}_p\gamma^\mu u_r)$
$Q_{\phi WB}$	$\phi^\dagger\sigma^i\phi W^i_{\mu\nu}B^{\mu\nu}$	Q_{dW}	$(\bar{q}_p\sigma^{\mu\nu}d_r)\sigma^i\phi W^i_{\mu\nu}$	$Q_{\phi d}$	$(\phi^\dagger i\overleftrightarrow{D}_\mu\phi)(\bar{d}_p\gamma^\mu d_r)$
$Q_{\phi\tilde{W}B}$	$\phi^\dagger\sigma^i\phi\tilde{W}^i_{\mu\nu}B^{\mu\nu}$	Q_{dB}	$(\bar{q}_p\sigma^{\mu\nu}d_r)\phi B_{\mu\nu}$	$Q_{\phi ud}$	$i(\tilde{\phi}^\dagger D_\mu\phi)(\bar{u}_p\gamma^\mu d_r)$
$(\bar{L}L)(\bar{L}L)$		$(\bar{R}R)(\bar{R}R)$		$(\bar{L}R)(\bar{L}R)+(\bar{L}R)(\bar{R}L)$	
Q_{ll}	$(\bar{l}_p\gamma_\mu l_r)(\bar{l}_s\gamma^\mu l_t)$	Q_{ee}	$(\bar{e}_p\gamma_\mu e_r)(\bar{e}_s\gamma^\mu e_t)$	Q_{le}	$(\bar{l}_p\gamma_\mu r)(\bar{e}_s\gamma^\mu e_t)$
$Q^{(1)}_{qq}$	$(\bar{q}_p\gamma_\mu q_r)(\bar{q}_s\gamma_\mu q_t)$	Q_{uu}	$(\bar{u}_p\gamma_\mu u_r)(\bar{u}_s\gamma_\mu u_t)$	Q_{lu}	$(\bar{l}_p\gamma_\mu l_r)(\bar{u}_s\gamma_\mu u_t)$
$Q^{(3)}_{qq}$	$(\bar{q}_p\gamma_\mu\sigma^i q_r)(\bar{q}_s\gamma_\mu\sigma^i q_t)$	Q_{dd}	$(\bar{d}_p\gamma_\mu d_r)(\bar{d}_s\gamma_\mu d_t)$	Q_{ld}	$(\bar{l}_p\gamma_\mu l_r)(\bar{d}_s\gamma_\mu d_t)$
$Q^{(1)}_{lq}$	$(\bar{l}_p\gamma_\mu l_r)(\bar{q}_s\gamma^\mu q_t)$	Q_{eu}	$(\bar{e}_p\gamma_\mu e_r)(\bar{u}_s\gamma_\mu u_t)$	Q_{qe}	$(\bar{q}_p\gamma_\mu q_r)(\bar{e}_s\gamma_\mu e_t)$
$Q^{(3)}_{lq}$	$(\bar{l}_p\gamma_\mu\sigma^i l_r)(\bar{q}_s\gamma^\mu\sigma^i q_t)$	Q_{ed}	$(\bar{e}_p\gamma_\mu e_r)(\bar{d}_s\gamma_\mu d_t)$	$Q^{(1)}_{qu}$	$(\bar{q}_p\gamma_\mu q_r)(\bar{u}_s\gamma^\mu u_t)$
		$Q^{(1)}_{ud}$	$(\bar{u}_p\gamma_\mu u_r)(\bar{d}_s\gamma^\mu d_t)$	$Q^{(8)}_{qu}$	$(\bar{q}_p\gamma_\mu T^a q_r)(\bar{u}_s\gamma^\mu T^a u_t)$
		$Q^{(8)}_{ud}$	$(\bar{u}_p\gamma_\mu T^a u_r)(\bar{d}_s\gamma^\mu T^a d_t)$	$Q^{(1)}_{qd}$	$(\bar{q}_p\gamma_\mu q_r)(\bar{d}_s\gamma^\mu d_t)$
				$Q^{(8)}_{qd}$	$(\bar{q}_p\gamma_\mu T^a q_r)(\bar{d}_s\gamma^\mu T^a d_t)$
$(\bar{L}R)(\bar{L}R)+(\bar{L}R)(\bar{R}R)$		Barion number violating			
Q_{leqd}	$(\bar{l}^j_p e_r)(\bar{d}_s q^j_t)$	Q_{duq}	$\varepsilon^{\alpha\beta\gamma}\varepsilon_{jk}\left[(d^\alpha_p)^T C u^\beta_r\right]\left[(q^{\gamma j}_s)^T C l^k_t\right]$		
$Q^{(1)}_{quqd}$	$(\bar{q}^j_p u_r)\varepsilon_{jk}(\bar{q}^k_s d_t)$	Q_{qqu}	$\varepsilon^{\alpha\beta\gamma}\varepsilon_{jk}\left[(q^{\alpha j}_p)^T C q^{\beta k}_r\right]\left[(u^\gamma_s)^T C e_t\right]$		
$Q^{(8)}_{quqd}$	$(\bar{q}^j_p T^a u_r)\varepsilon_{jk}(\bar{q}^k_s T^a d_t)$	$Q^{(1)}_{qqq}$	$\varepsilon^{\alpha\beta\gamma}\varepsilon_{jk}\varepsilon_{mn}\left[(q^{\alpha j}_p)^T C q^{\beta k}_r\right]\left[(q^{\gamma m}_s)^T C l^n_t\right]$		
$Q^{(1)}_{lequ}$	$(\bar{l}^j_p e_r)\varepsilon_{jk}(\bar{q}^k_s u_t)$	$Q^{(3)}_{qqq}$	$\varepsilon^{\alpha\beta\gamma}(\sigma^i\varepsilon)_{jk}(\sigma^i\varepsilon)_{mn}\left[(q^{\alpha j}_p)^T C q^{\beta k}_r\right]\left[(q^{\gamma m}_s)^T C l^n_t\right]$		
$Q^{(3)}_{lequ}$	$(\bar{l}^j_p\sigma_{\mu\nu}e_r)\varepsilon_{jk}(\bar{q}^k_s\sigma^{\mu\nu}u_t)$	Q_{duu}	$\varepsilon^{\alpha\beta\gamma}\left[(d^\alpha_p)^T C u^\beta_r\right]\left[(u^\gamma_s)^T C e_t\right]$		

Table 1.4: Dimension-six operators as defined in the Warsaw basis [46].

receiving both *direct* SMEFT contributions from higher-dimensional operators ($P^{d\geq 2}$) and *indirect* ones through modified input parameters ($F_n^{d\geq 2}$). The net effect is that EFT corrections to input observables are absorbed into redefinitions of the predicted quantities. If $\mathcal{P}$ coincides with one of the input observables $\mathcal{O}_p$ used to fix g, all SMEFT corrections

$\boldsymbol{m_\nu = 0}$	
α_S, θ_{QCD}	QCD
$y_e, y_\mu, y_\tau, y_u, y_d, y_c, y_s, y_t, y_b$	Yukawas
g, g', v, λ	Electroweak and Higgs sector
$\theta_{12}, \theta_{13}, \theta_{23}, \delta$	CKM
$\boldsymbol{m_\nu \neq 0}$	
$y_{\nu_e}, y_{\nu_\mu}, y_{\nu_\tau}$	Yukawas
$U_{12}, U_{13}, U_{23}, \delta'$	PMNS

Table 1.5: Input parameters of the Standard Model for the cases of massless and massive neutrinos.

cancel order by order. Otherwise, the prediction for $\mathcal{P}$ depends on the chosen set of input observables $\mathcal{O}$. The difference between two predictions based on distinct input sets $\mathcal{O}$ and $\mathcal{O}'$ is given by

$$\mathcal{P}(\mathcal{O}) - \mathcal{P}(\mathcal{O}') = \frac{1}{\Lambda^2}\left[-\frac{\partial P^{(0)}}{\partial F_n^{(0)}}F_n^{(2)} + \frac{\partial P^{(0)}}{\partial F_n'^{(0)}}F_n'^{(2)}\right] + \mathcal{O}\left(\frac{1}{\Lambda^4}\right). \tag{1.28}$$

Hence, the choice of the input observable set $\mathcal{O}$ must be made carefully, as predictions depend on this choice [59].
In the Higgs and electroweak sectors, the SM involves four independent parameters, $g = \{g, g', v, \lambda\}$, which can be fixed using the observable set $\mathcal{O} = \{\alpha_{em}, G_F, m_W, m_Z, m_h\}$. The Higgs boson mass m_h is always required to determine λ, but the selection of the remaining three observables is not unique. Three common electroweak input schemes are used:

- $\{\alpha_{em}, G_F, m_Z\}$**:** Historically adopted at LEP, this scheme employs the most precisely measured parameters. Its drawback is that m_W then receives SMEFT corrections, introducing nonlinear dependencies on Wilson coefficients in many observables and in the W-boson propagator.

- $\{G_F, m_Z, m_W\}$**:** This choice fixes m_W to its measured value, preventing SMEFT corrections to the W mass but requiring corrections to the Fermi constant G_F, even for processes unrelated to leptonic observables.

- $\{\alpha_{em}, m_Z, m_W\}$**:** This alternative avoids SMEFT-induced shifts in G_F and the associated electroweak couplings, but can slightly degrade perturbative convergence in both SM and SMEFT predictions [60].

Currently, most SMEFT computational frameworks adopt the $\{G_F, m_Z, m_W\}$ input scheme as the default choice [61].

1.2.5 Flavour assumptions at dimension six

While the electroweak sector contains only 59 independent dimension-six operators [46], this number grows dramatically once flavour indices are included. If no flavour symmetry is imposed, the SMEFT Lagrangian in Eq. 1.26 depends on 1350 CP-even and 1149

CP-odd independent Wilson coefficients [62]. Given the current experimental precision, it is unfeasible to constrain such a large parameter space directly, making dimensionality reduction strategies essential. This is achieved by imposing flavour symmetries—possibly with controlled symmetry breaking—at the cost of introducing some degree of model dependence. These symmetries need not be fundamental properties of the ultraviolet (UV) theory but can instead arise accidentally if, for instance, the UV completion distinguishes between fermion generations [63]. The most common flavour-symmetry scenarios adopted in SMEFT analyses are summarized below.

Maximal $U(3)^5$ symmetry

The largest flavour symmetry compatible with the gauge symmetries of the SM Lagrangian is

$$U(3)^5 = U(3)_l \times U(3)_q \times U(3)_e \times U(3)_u \times U(3)_d = SU(3)^5 \times U(1)^5. \tag{1.29}$$

This symmetry implies that fermions sharing the same electroweak quantum numbers are indistinguishable under gauge interactions [64]. The most restrictive flavour assumption corresponds to treating $U(3)^5$ as an exact symmetry of the beyond-the-Standard-Model (BSM) sector, which effectively removes all flavour dependence in the Wilson coefficients. Such an assumption, however, is valid only if Yukawa couplings are neglected.

Linear Minimal Flavour Violation

In the Standard Model, the Yukawa matrices $Y_{e,u,d}$ and the CKM phase δ explicitly break the $U(3)^5$ flavour symmetry. The framework of *Minimal Flavour Violation* (MFV) [65, 66] restores flavour symmetry by promoting the Yukawa couplings to spurion fields that transform under $U(3)^5$. This prescription allows one to construct all possible $U(3)^5$-invariant combinations of SM fields and Yukawa insertions. Flavour violation then arises only through controlled insertions of these spurions, ensuring that any new physics respects the observed SM flavour structure. Spurion insertions are organized as a power expansion in the Yukawa couplings—for example, single insertions $\mathcal{O}(Y_{u,d,e})$ or higher-order combinations such as $\mathcal{O}(Y_u^2)$, $\mathcal{O}(Y_u Y_d)$, and $\mathcal{O}(Y_u^2 Y_d)$ [63].

$U(2)^3$ symmetry in the quark sector

Although the CKM mechanism provides an excellent description of flavour mixing and CP violation in the quark sector, the origin of the hierarchical structure of quark masses and mixing angles remains unknown. Within the MFV framework based on a broken $U(3)^3$ symmetry [67], this hierarchy is not naturally explained. An alternative approach introduces a minimally broken $U(2)^3$ symmetry acting only on the first two generations of quarks [68, 69]. This construction naturally suppresses flavour-changing effects while distinguishing the light generations from the third one.
Denoting by (q_p, u_p, d_p) the left- and right-handed quark fields of the first two generations (with generation index $p = 1, 2$), by (Q, t, b) those of the third generation, and by Ω a generic $U(2)$ transformation, the symmetry can be expressed as

$$\begin{aligned} &U(2)^3 = U(2)_q \times U(2)_u \times U(2)_d, \\ &q \to \Omega_q q, \quad u \to \Omega_u u, \quad d \to \Omega_d d, \quad Q \to Q, \quad t \to t, \quad b \to b. \end{aligned} \tag{1.30}$$

In this scheme, the first two generations transform as $U(2)$ doublets, while the third generation remains a singlet. The lepton sector is independent of this construction, allowing different levels of flavour symmetry to be imposed. The most restrictive choice is $U(3)^2 = U(3)_l \times U(3)_e$, corresponding to the maximal or MFV assumptions. More relaxed options include $U(2)^2 = U(2)_l \times U(2)_e$, which combined with the quark sector leads to $U(2)^5$ [68], or a completely flavour-diagonal symmetry $U(1)^3_{l+e} = U(1)_e \times U(1)_\mu \times U(1)_\tau$ [70, 66].
The number of independent SMEFT parameters under each flavour assumption is summarized in Table 1.6.

	General		$\boldsymbol{U(3)^5}$		**MFV**		$\boldsymbol{U(2)^3 \times U(1)^3}$		$\boldsymbol{U(2)^3 \times U(3)^2}$	
	All	CP	All	CP	All	CP	All	CP	All	CP
X^3	4	2	4	2	2	-	4	2	4	2
ϕ^6 and $\phi^4 D^2$	3	-	3	-	3	-	3	-	3	-
$\psi^2 X \phi$	8	4	8	4	4	-	8	4	8	4
$\psi^2 \phi^3$	54	27	6	3	7	-	14	7	10	5
$\psi^2 X \phi$	144	72	16	8	20	-	36	18	28	14
$\psi^2 \phi^2 D$	81	30	9	1	14	-	21	2	15	2
$(\bar{L}L)(\bar{L}L)$	297	126	8	-	10	-	31	-	16	-
$(\bar{R}R)(\bar{R}R)$	450	195	9	-	19	-	40	2	27	2
$(\bar{L}L)(\bar{R}R)$	648	288	8	-	28	-	54	4	31	4
$(\bar{L}R)(\bar{L}R) + (\bar{L}R)(\bar{R}L)$	810	405	14	7	13	-	64	32	40	20
total	2499	1149	85	25	120	-	275	71	182	53

Table 1.6: Number of independent parameters for each class of dimension-six operators as defined in the Warsaw basis under different flavour assumptions [66].

1.3 The Large Hadron Collider and the Compact Muon Solenoid

1.3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a double-ring superconducting hadron accelerator and collider located at CERN. It occupies the 27 km circular tunnel that previously hosted the Large Electron–Positron Collider (LEP) [71]. The main scientific motivation behind its design and construction is the exploration of the scalar sector of the Standard Model (SM) and the investigation of the mechanism responsible for electroweak symmetry breaking, presumed to be explained by the Higgs mechanism. Beyond the discovery of the Higgs boson, the LHC physics program enables precision tests of the SM at the TeV scale, detailed studies of Quantum Chromodynamics (QCD), electroweak and flavour physics, and extensive searches for phenomena predicted by ultraviolet (UV)-complete theories, such as supersymmetry or extra dimensions.

The LHC operates at unprecedented parameters, reaching a centre-of-mass energy of 13.6 TeV and instantaneous luminosities in excess $\mathcal{O}(10^{34})$ $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. Achieving such performance presents major technological challenges. The required beam intensity and stringent collimation conditions rule out the possibility of a particle–antiparticle collider sharing a common vacuum and magnetic system. Consequently, the LHC was designed as a proton–proton collider, employing two separate magnetic fields and vacuum chambers arranged in a twin-bore magnet architecture—an innovative solution dictated by the 3.7 m diameter of the LEP tunnel. Out of the eight possible interaction regions, only four host detectors, reducing unnecessary beam crossings and ensuring beam stability.

Reaching the design beam energy of 7 TeV per proton requires dipole magnetic fields of 8.33 T, achievable only with superconducting technology. The LHC therefore employs 1232 superconducting dipole magnets and 858 superconducting quadrupoles, fabricated from a niobium–titanium alloy and cooled with superfluid helium to an operating temperature of 1.9 K.

Acceleration of the proton beams is achieved using 16 superconducting radio-frequency (RF) cavities—eight per beam—operating at 400 MHz. These cavities alternately accelerate slower protons and decelerate faster ones, ensuring longitudinal bunch stability in both position and momentum through the principle of phase stability. The RF cavities are housed within cryogenic units maintained at 4.5 K. The LHC injects one proton bunch every ten RF cycles. Given the 27 km circumference, the ring could, in principle, host up to 3600 bunches, but operational constraints related to injection and extraction reduce this to 2808 bunches per beam. This configuration corresponds to a bunch spacing of 25 ns, resulting in a nominal collision rate of approximately 40 MHz.

Protons are delivered to the LHC through a complex accelerator chain (Figure 1.4). They are first accelerated to 160 MeV by the Linac4, then injected into the Proton Synchrotron Booster (PSB), and subsequently into the Proton Synchrotron (PS), which raises the energy to 26 GeV. The Super Proton Synchrotron (SPS) provides the final acceleration stage, boosting the beam energy to 450 GeV before injection into the LHC for the last stage of acceleration to the nominal energy.

Four major experiments are located at the LHC's beam-crossing points. Two of them—ATLAS (A Toroidal LHC ApparatuS) at Point 1 and CMS (Compact Muon Solenoid) at Point 5—are general-purpose detectors operating at the highest-luminosity interaction regions, with peak luminosities designed to be 10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$, but during the operations of Run 3 a

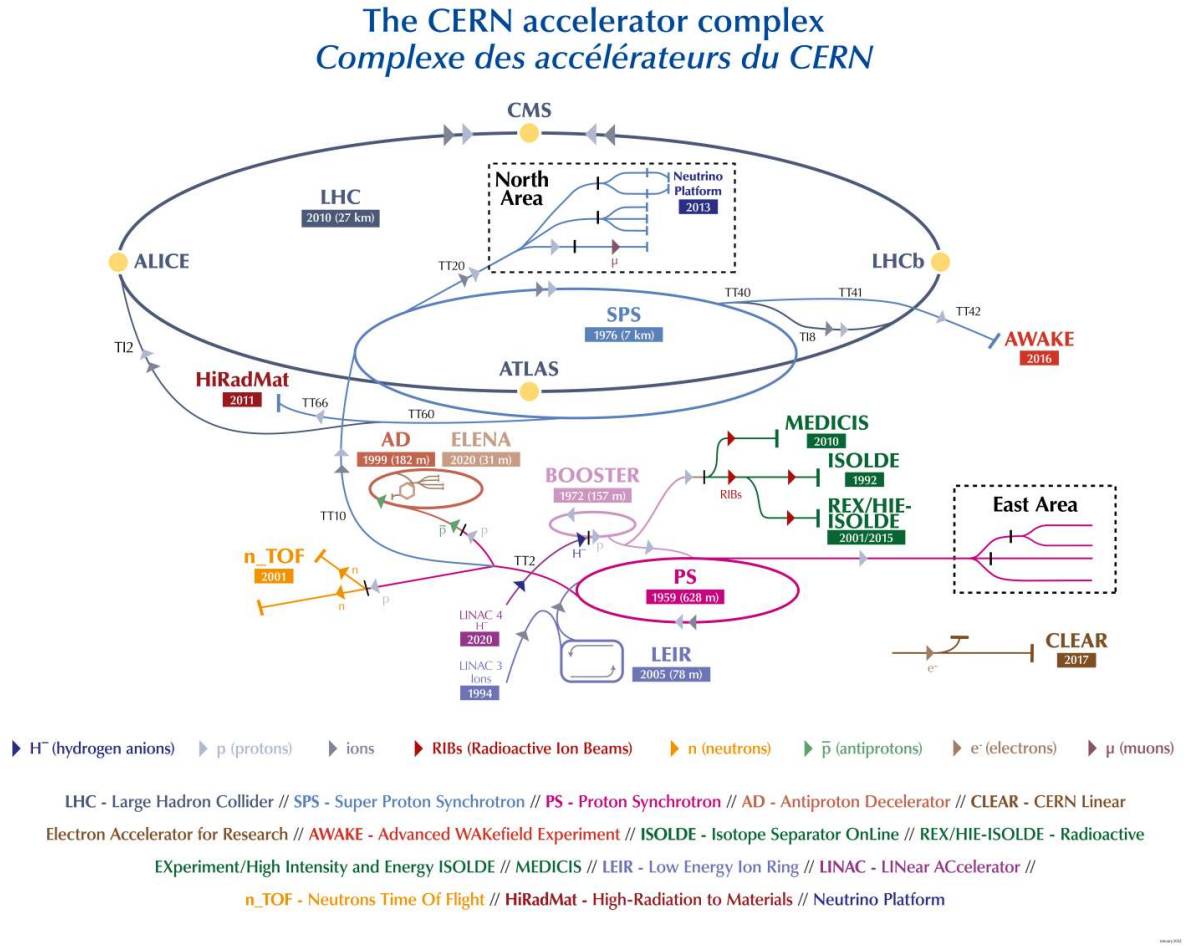

Figure 1.4: The CERN accelerator complex, experiments, and injection chain.

peak luminosity two times higher was reached. The other two experiments, ALICE (A Large Ion Collider Experiment) and LHCb (LHC beauty), are dedicated to the study of heavy-ion collisions and flavour physics, respectively, and operate at lower luminosities of 10^{27} cm^{-2}s^{-1} and 2×10^{29} cm^{-2}s^{-1}.
The stochastic nature of high-energy interactions implies that the observation of rare processes—such as those predicted by beyond-the-Standard-Model (BSM) theories—requires large event samples. The expected number of events N for a given process can be expressed as

$$N = L\,\sigma\,\varepsilon\,A, \tag{1.31}$$

where L denotes the integrated luminosity, σ the production cross section, ε the detector efficiency, and A the geometrical acceptance. This relationship highlights the importance of both accelerator performance and detector design in enabling precision measurements and discovery potential at the LHC.

Luminosity and operational performance

The luminosity of a collider quantifies the rate of particle interactions and depends on the beam parameters. For Gaussian beam profiles, the instantaneous luminosity can be

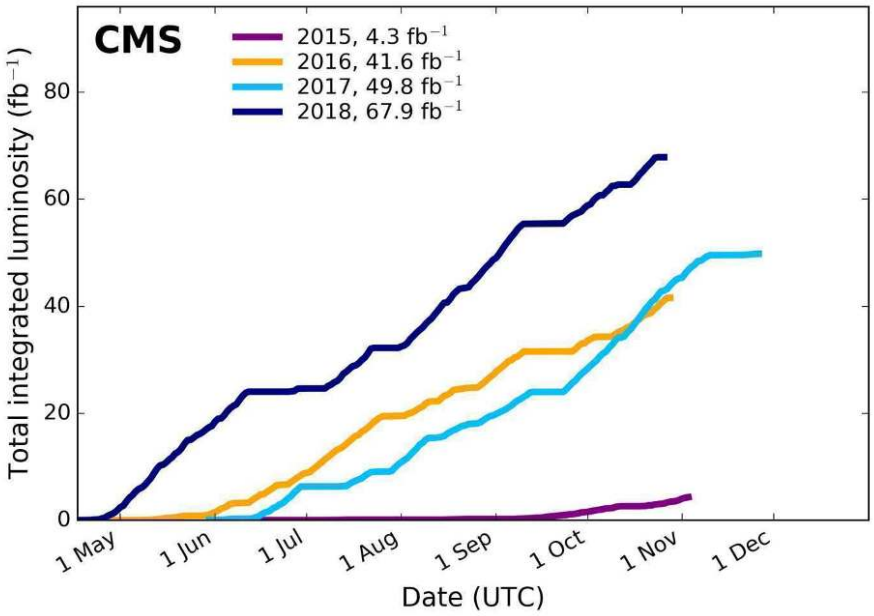

Figure 1.5: Total integrated luminosity delivered to CMS during Run 2 (2015–2018) as a function of time.

expressed as

$$L = \frac{N_b^2 \, n_b \, f_{\text{rev}} \, \gamma_r}{4\pi \, \varepsilon_n \, \beta^*} \, F, \tag{1.32}$$

where N_b and n_b denote, respectively, the number of protons per bunch and the number of bunches per beam, f_{rev} is the revolution frequency, γ_r the relativistic Lorentz factor, ε_n the normalized transverse emittance, and β^* the value of the betatron function at the interaction point (IP), which determines the beam size. The factor F is the geometric luminosity reduction factor, accounting for the finite crossing angle, transverse beam size, and bunch length at the IP. The total amount of data delivered to the experiments is quantified by the integrated luminosity, obtained as $\int_0^t L \, dt$.

Symbol	Description	2016	2017	2018
$\sqrt{s}$	Centre-of-mass energy (TeV)	13	13	13
Δt	Bunch spacing (ns)	25	25	25
L	Peak inst. luminosity (10^{34} cm^{-2}s^{-1})	1.4	2.1	2.1
μ	Average pileup	25	38	37
n_b	Bunches per beam	2220	2556/1868	2556
N_b	Protons per bunch (10^{11})	1.0–1.25	1.0–1.25	1.0–1.25
ε_n	Transverse emittance (μm)	2.2	2.2	1.9
f_{rev}	Revolution frequency (Hz)	11245	11245	11245
β^*	Beta function (cm)	40	40–30	30–25

Table 1.7: LHC Run 2 operating parameters for 2016–2018.

The LHC entered operation in 2009, delivering its first proton–proton collisions at a centre-of-mass energy of 0.9 TeV. In 2010, the energy was increased to 7 TeV, and an

integrated luminosity of 45 pb^{-1} was collected. The following year, the machine reached a peak luminosity of 4×10^{33} $cm^{-2}s^{-1}$ and delivered about 6 fb^{-1} of data. Run 1 concluded in 2012 with stable operation at 8 TeV and an instantaneous luminosity of 6×10^{33} $cm^{-2}s^{-1}$ maintained for most of the year.
A two-year technical shutdown followed to consolidate the high-current splices between the superconducting magnets. In 2015, the LHC resumed operation at a centre-of-mass energy of 13 TeV, marking the beginning of Run 2. Between 2015 and 2018, the collider consistently exceeded its design luminosity, reaching peaks of 2.1×10^{34} $cm^{-2}s^{-1}$. This increase resulted in higher levels of simultaneous interactions per bunch crossing (pileup), with averages of around 40 and peaks of up to 70 in 2018 (see Table 1.7). Over this period, each of the main experiments recorded approximately 140 fb^{-1} of data, as shown in Figure 1.5.
A second long shutdown (2019–2021) prepared the LHC for its next operational phase, the High-Luminosity LHC (HL-LHC), scheduled to begin in 2030. In 2022, the machine restarted with a centre-of-mass energy of 13.6 TeV, close to its design value, and an average pile-up of 60. It has been able to deliver already 300 fb^{-1} of data and the data-taking will continue in 2026 before the shutdown for the HL-LHC upgrades. The HL-LHC will ultimately reach a peak instantaneous luminosity of up to seven times the nominal value, $L = 7 \times 10^{34}$ $cm^{-2}s^{-1}$, enabling the collection of over 3 ab^{-1} of data during its lifetime.

1.3.2 The Compact Muon Solenoid Detector

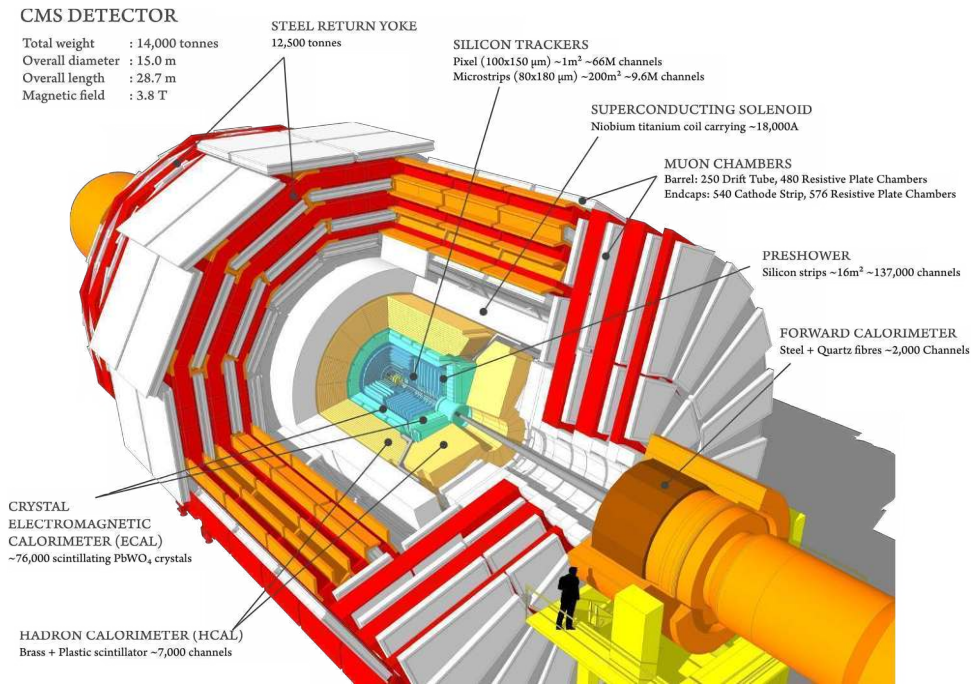

Figure 1.6: CMS detector layout

The Compact Muon Solenoid (CMS) detector, one of the two general-purpose experiments at the LHC, was designed to investigate high-energy proton-proton interactions and provides outstanding performance in the reconstruction of electrons, photons, and muons. At its core lies a niobium–titanium superconducting solenoid capable of generat-

ing a 3.8 T magnetic field over a diameter of 6 m, making it one of the largest and most powerful solenoids ever constructed. The solenoid and its return yoke together weigh nearly 2000 tonnes. Its compact geometry enables the entire detector to fit within the experimental cavern located about 100 m underground at LHC Point 5.
The broad physics program of CMS required a design that optimizes particle identification and momentum measurement over a wide energy range. This led to exceptional performance from the electromagnetic calorimeter (ECAL), the hadron calorimeter (HCAL), the inner tracking system, and the dedicated muon detectors. Many potential Beyond the Standard Model (BSM) signatures involve weakly interacting or invisible particles, which manifest experimentally as an imbalance in the total visible transverse energy. To detect such signatures, the ECAL and HCAL were designed to be nearly hermetic, providing almost complete solid-angle coverage around the interaction point and allowing accurate measurements of missing transverse momentum.
The tracking system forms the innermost component of CMS, located within the solenoid and extending to within about 3 cm of the beam line. Together with the ECAL and HCAL, it benefits from the solenoid's magnetic field to measure the curvature and hence the momentum of charged particles. Outside the solenoid, within the return yoke where the magnetic field is about 2 T, reside the muon detectors—the outermost layer of CMS [72]. These consist of three technologies optimized for different regions: drift tubes, cathode strip chambers, and resistive plate chambers. Due to the variation in particle flux across the detector, the CMS subdetectors are divided into a central barrel region and two forward endcaps, each adopting distinct design solutions optimized for their geometric and radiation environments. A schematic view of the CMS detector is shown in Figure 1.6.
A right-handed Cartesian coordinate system is used to describe the CMS geometry and the reconstructed particle kinematics. The origin is located at the nominal interaction point, with the x-axis pointing towards the centre of the LHC ring, the y-axis directed vertically upward (perpendicular to the LHC plane), and the z-axis aligned with the anticlockwise beam direction. Spherical coordinates are also employed, with θ denoting the polar angle relative to the z-axis and ϕ the azimuthal angle in the x–y plane. The pseudorapidity, defined as $\eta = -\ln[\tan(\theta/2)]$, is often used in place of θ since particle production rates in hadron collisions are approximately uniform as a function of η, and differences in rapidity are Lorentz invariant. The spatial separation between two objects is expressed as the distance in the η–ϕ plane, $(\Delta R)^2 = (\Delta\eta)^2+(\Delta\phi)^2$. Transverse quantities, defined with respect to the beam axis, are frequently used because they are invariant under Lorentz boosts along the z direction. For instance, the transverse momentum of a particle is given by $(p_T)^2 = (p_x)^2 + (p_y)^2$.

Tracking system

The CMS tracking system [74] is the innermost detector surrounding the interaction point. It extends 5.6 m in length and up to 2.4 m in radius, providing a total active area of about 200 m^2 with nearly 80 million readout channels. Given the intense radiation environment near the beam line, the tracker was designed with a lightweight, two-tiered structure entirely based on silicon technology. This approach provides large surfaces of finely segmented active detectors with minimal material budget. The uniform 3.8 T magnetic field inside the solenoid allows the precise measurement of charged-particle trajectories. The individual position measurements (hits) recorded in the silicon sensors

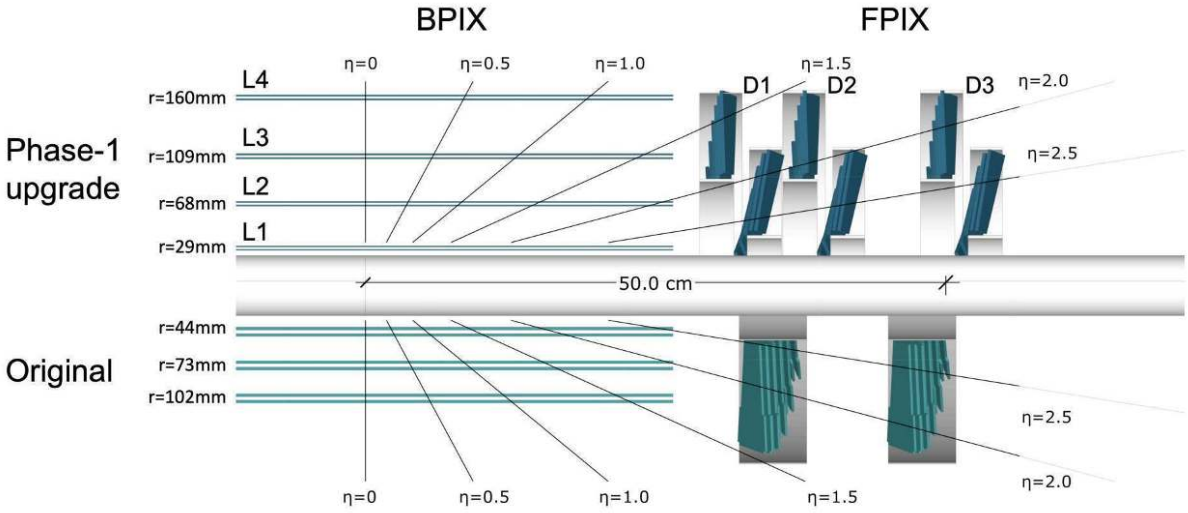

Figure 1.7: Schematic view of the CMS tracking system in the r–z plane before and after the upgrades introduced during the 2016–2017 year-end technical stop [73].

are combined into tracks, from which the particle momentum and charge are determined. By correlating multiple reconstructed tracks, the tracker identifies the hard-scatter interaction point (primary vertex) and distinguishes it from other simultaneous proton–proton collisions occurring in the same bunch crossing, commonly referred to as pileup. It also enables the reconstruction of displaced vertices from in-flight decays of long-lived particles such as B- and D-hadrons or τ leptons, providing key information for heavy-flavour physics analyses.

Because the particle flux decreases rapidly with radial distance r from the beam axis, the tracker employs a hybrid layout of pixel and strip detectors. The innermost region ($r < 20$ cm) is instrumented with high-granularity silicon pixel sensors, while the outer region ($20 < r < 250$ cm) is covered by silicon microstrip modules with a pitch increasing proportionally to r, ensuring optimal resolution and cost efficiency.

The pixel detector consists of three barrel layers and two endcap disks, covering pseudorapidities up to $|\eta| < 2.5$. Following a major upgrade in 2017, an additional layer was added both in the barrel and endcap regions to enhance pattern recognition and radiation tolerance. The four barrel layers are located at radii of 2.9, 6.8, 10.2, and 16.0 cm. Each pixel cell measures approximately 100×150 μm^2, resulting in about 66 million channels—roughly 90% of all tracker readout channels—and an active area of about 1 m^2. The typical spatial resolution is 10 μm in the (r, ϕ) plane and 20 μm along the z direction. In the outer tracker, larger silicon-strip modules are used, with a pitch varying from 80 to 180 μm depending on the radial position. This coarser granularity is acceptable due to the lower hit density and is essential for covering the large surface area economically. The strip system in the barrel is divided into two regions: the Tracker Inner Barrel (TIB), extending up to $|z| < 65$ cm with four active layers, and the Tracker Outer Barrel (TOB), covering $|z| < 110$ cm with six layers. The 45 cm gap between the TIB and TOB is instrumented with four Tracker Inner Disks (TID) on each side. To achieve full pseudorapidity coverage up to $|\eta| < 2.5$, two Tracker Endcaps (TEC) are installed, each consisting of nine disks and spanning $120 < |z| < 280$ cm.

The single-hit spatial resolution varies with radius, ranging from 20–50 μm in the trans-

verse direction and 200–500 μm longitudinally. To mitigate radiation damage and dissipate the heat generated by the on-detector electronics, both pixel and strip subsystems are actively cooled, operating at approximately -20°C and -15°C, respectively. A schematic representation of the CMS tracker is shown in Figure 1.7.

Electromagnetic Calorimeter

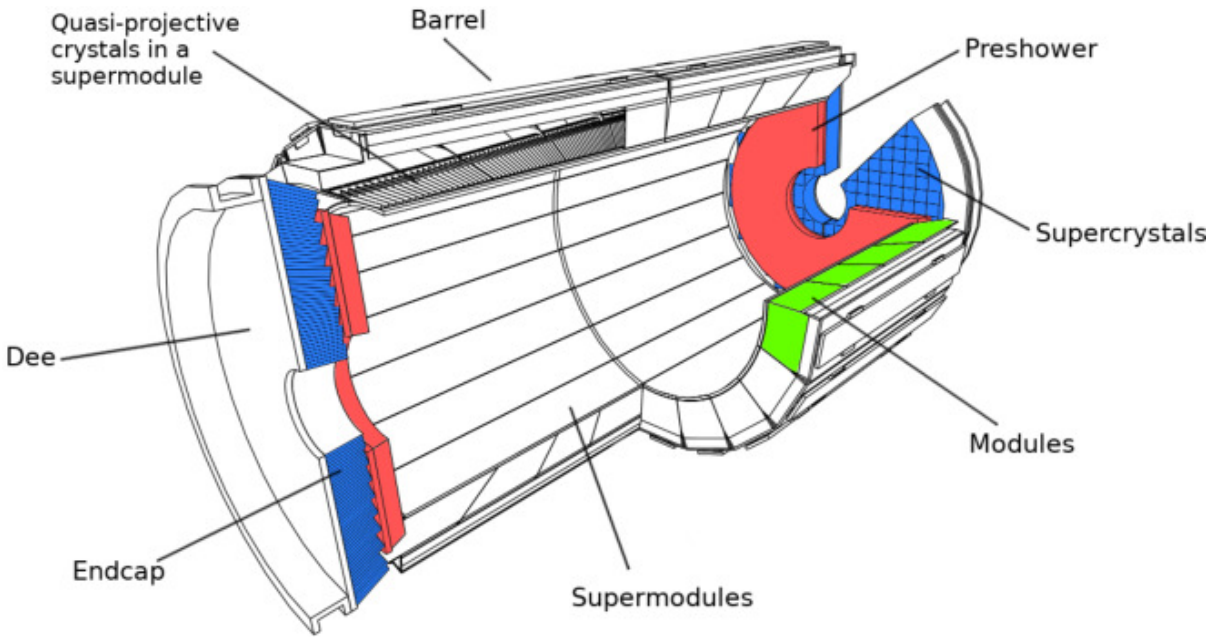

Figure 1.8: Layout and geometry of the CMS Electromagnetic Calorimeter (ECAL).

The Electromagnetic Calorimeter (ECAL) [75] is a homogeneous calorimeter composed of 75 848 lead tungstate ($PbWO_4$) scintillating crystals, designed to measure the energy of electrons and photons with excellent precision. A schematic view of its geometry is shown in Figure 1.8. The crystals serve simultaneously as dense absorber and active scintillating medium, enabling high energy resolution within a compact volume.
Lead tungstate was chosen for its combination of density (8.28 g/cm^3), short radiation length (0.89 cm), and small Molière radius (2.19 cm), ensuring efficient containment of electromagnetic showers. Its fast scintillation response—about 80% of light emitted within 25 ns—makes it well suited to the 40 MHz bunch-crossing frequency of the LHC. $PbWO_4$ has also descrete radiation tolerance properties, though its relatively low light yield ($\sim$30 photons/MeV, depending on temperature) requires photodetectors with internal amplification capable of operating within the 3.8 T axial magnetic field of CMS. Avalanche photodiodes are used in the barrel region, while vacuum phototriodes, more resistant to radiation, are employed in the endcaps. The front-end electronics amplify and digitize the signals using 12-bit ADCs sampling at 40 MHz.
The ECAL barrel consists of 61 200 crystals arranged in 36 supermodules (18 per half-barrel), each spanning 20° in ϕ. The barrel crystals have a truncated-pyramid shape, with a front face of 22×22 mm^2, a length of 23 cm ($\sim$25.8 X_0), and cover the pseudorapidity range $|\eta| < 1.479$. Two endcap disks extend the coverage to $|\eta| < 3.0$, each containing 7 324 crystals with a front face of 2.86×2.86 cm^2, located about 314 cm from

the interaction point. The endcap crystals are arranged in supercrystals (arrays of 5×5 crystals), grouped into two semi-circular halves called "Dees," each comprising 3 662 crystals organized into 138 standard and 18 partial supercrystals. In the forward region, a silicon–lead sampling detector (the preshower) is installed in front of the endcaps to help distinguish isolated photons from $\pi^0 \to \gamma\gamma$ decays.

The crystals are mounted in a quasi-projective geometry, tilted by about 3° in both η and ϕ with respect to the interaction point [75], minimizing energy loss in inter-crystal gaps. The $PbWO_4$ light yield and photodetector response depend strongly on temperature (-2%/°C and -2.3%/°C respectively). A precision cooling system maintains the temperature at 18°C, with stability better than ±0.05°C in the barrel and ±0.1°C in the endcaps, while removing the heat dissipated by the readout electronics. This operating temperature also helps mitigate radiation-induced transmission losses. To monitor and correct for variations in crystal transparency, a 440 nm laser system injects light into each crystal to derive time-dependent correction factors.

The calorimetric energy resolution can be parameterized as the quadratic sum of three contributions:

$$\left(\frac{\sigma}{E}\right) = \frac{S}{\sqrt{E}} \oplus \frac{N}{E} \oplus C, \tag{1.33}$$

where S, N, and C are the stochastic, noise, and constant terms, respectively. The stochastic term models the statistical fluctuations in the number of scintillation photons, scaling as $\sqrt{E}$; the noise term represents the contribution from electronic and pileup noise, independent of E; and the constant term accounts for non-uniformities in crystal response, light collection, and longitudinal leakage.

Beam tests with electrons and cosmic-ray studies have demonstrated excellent performance. Measurements with a 3×3 crystal matrix [76] (without magnetic field or upstream material) yielded $S = 2.8\%$, $N = 12\%$, and $C = 0.3\%$. Absolute and inter-channel calibrations are essential to correct for variations in crystal light yield (∼15%) and phototriode gain dispersion in the endcaps (∼25%). Initial calibration, derived from laboratory and cosmic-ray data, is refined using collision events, achieving an energy resolution of about 2% for 45 GeV electrons in the barrel and 2–5% in other regions. For central electrons, where bremsstrahlung losses are minimal, the resolution improves to approximately 1.5%. A comprehensive discussion of ECAL calibration and resolution in pp collisions at $\sqrt{s} = 7$ TeV is provided in Ref. [77].

Hadron Calorimeter

The CMS hadron calorimeter (HCAL) [78] is a sampling calorimeter composed of alternating layers of brass absorber and plastic scintillator. It is designed to measure the energy of hadrons that traverse the electromagnetic calorimeter (ECAL) without being fully absorbed. In contrast to electromagnetic showers, hadronic cascades are more complex: they involve both nuclear and hadronic interactions that generate non-Poissonian fluctuations and a significant fraction of undetectable energy. Additionally, neutral pions produced within the shower decay promptly into photons, giving rise to an electromagnetic component that the detector responds to differently than to purely hadronic energy deposits. These effects intrinsically limit the calorimetric resolution for hadrons, which can only be improved through advanced offline techniques such as the particle-flow reconstruction. Despite these challenges, the HCAL plays a central role in reconstructing hadronic jets and in measuring missing transverse momentum arising from undetected or weakly interacting particles such as neutrinos.

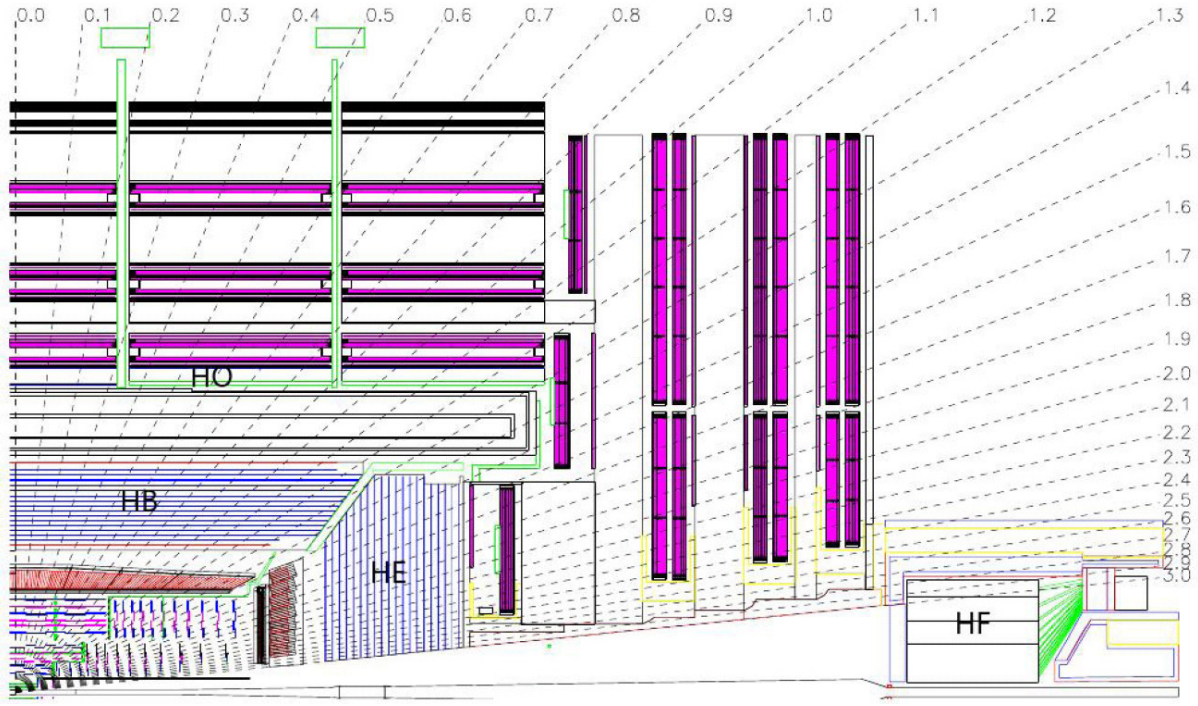

Figure 1.9: Layout and geometry of the CMS Hadron Calorimeter (HCAL). The longitudinal cross section illustrates the barrel region (HB) positioned between the ECAL and the solenoid, the endcaps (HE), the outer barrel (HO) located beyond the solenoid, and the forward calorimeter (HF).

The design of the HCAL is strongly constrained by the CMS magnet geometry and the need for nearly hermetic coverage up to $|\eta| < 5$. Because of the limited radial space between ECAL and the superconducting solenoid, a sampling design was adopted. Brass was chosen as absorber material owing to its short nuclear interaction length (16.24 cm) and non-magnetic properties, while the interleaved plastic scintillator tiles provide the active medium. When hadrons enter the calorimeter, they undergo inelastic interactions with the absorber nuclei, producing a cascade of secondary particles. These secondaries traverse the scintillator layers, where the deposited energy is measured through the scintillation light they emit. In total, the HCAL contains about 70 000 scintillator tiles distributed across four major subsystems: the barrel (HB), the outer (HO), the endcaps (HE), and the forward calorimeter (HF).

The HCAL barrel (HB) is located between the ECAL and the solenoid, covering the pseudorapidity range $|\eta| < 1.3$. Because the limited thickness of the HB does not allow full containment of high-energy showers, an additional outer calorimeter (HO) is installed between the solenoid and the muon system for $|\eta| < 1.26$. The HO collects energy leaking from the HB, extending the total depth to about 11 interaction lengths (λ_0) and thereby improving energy resolution. The endcap calorimeters (HE) cover $1.3 < |\eta| < 3.0$, while the forward calorimeters (HF), positioned 11.2 m from the interaction point, extend the detector coverage to $|\eta| = 5.2$. A schematic view of the HCAL layout is shown in Figure 1.9.

While the HB, HO, and HE employ brass absorbers and plastic scintillators, the extreme radiation environment in the forward region required a different approach. The HF uses 1.65-m-thick steel absorbers interleaved with 0.6-mm radiation-hard quartz fibers as the active medium, where light is produced via the Cherenkov effect rather than scintillation. In HB and HE, the light from the scintillators is collected by wavelength-shifting fibers embedded in the tiles and read out by hybrid photodiodes (HPDs). Each readout channel corresponds to a "tower" of stacked scintillator layers within a fixed (η, ϕ) region. The tower segmentation is $\Delta\eta \times \Delta\phi \approx 0.087 \times 0.087$ in HB and approximately 0.17×0.17 in HE. In the HF, Cherenkov light is detected by photomultiplier tubes (PMTs); two fiber lengths are used per tower to separate the electromagnetic and hadronic components of the shower.

The HCAL energy resolution is largely driven by the incomplete containment of hadronic showers. Test-beam measurements with pions in the 4–300 GeV range indicate a stochastic term of about $S = 110\%$ and a constant term of $C = 9\%$ for the raw energy resolution:

$$\left(\frac{\sigma}{E}\right) = \frac{S}{\sqrt{E}} \oplus C. \tag{1.34}$$

This performance, combined with the information from ECAL and the tracker in the particle-flow algorithm, enables accurate reconstruction of jets, missing transverse momentum, and other hadronic observables critical for CMS physics analyses.

Muon Chambers

Located outside the solenoid, the outermost region of CMS houses the muon detection system, which covers an area of approximately 25 000 m^2 [79]. Muons, being much more massive than electrons, experience minimal energy loss through bremsstrahlung and can traverse the tracker, calorimeters, and solenoid without being absorbed. The muon system thus serves two essential purposes: it identifies muons and provides an independent

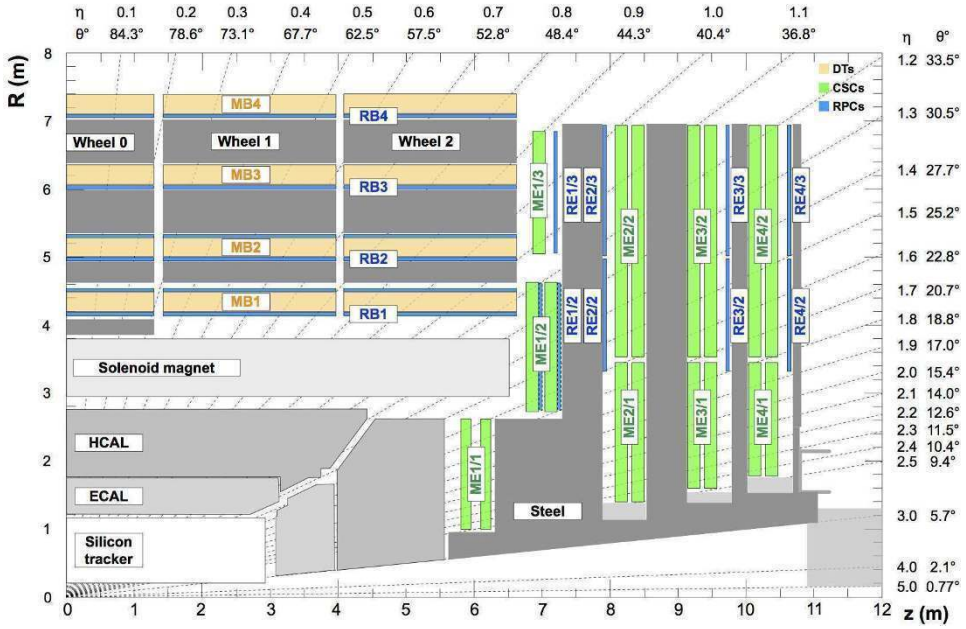

Figure 1.10: r–z cross section of one quarter of the CMS detector in the longitudinal plane. The orange, green, and blue regions correspond to the DT, CSC, and RPC subsystems, respectively, while the dark grey areas represent the steel return yoke disks.

measurement of their momentum using the 2 T magnetic return field generated in the iron yoke surrounding the solenoid. Because isolated muons are clean signatures of many key physics processes, their trajectories are measured twice in CMS—first in the inner tracker and then in the outer muon chambers—allowing for precise momentum reconstruction and efficient triggering.

The muon detector is composed of a cylindrical barrel section and two endcap regions, each instrumented with gaseous detectors optimized for the expected background rates and magnetic field conditions. Three detector technologies are employed: drift tubes (DTs), cathode strip chambers (CSCs), and resistive plate chambers (RPCs). This combination ensures uniform coverage and robust performance across the full pseudorapidity range, exploiting the cost-effectiveness and large-area capabilities of gaseous detectors. An overview of the muon system layout is shown in Figure 1.10.

In the barrel region ($|\eta| < 1.2$), where the magnetic field is relatively uniform (< 0.4 T) and the background rates are low, CMS employs 250 DT chambers [80]. These are arranged in five cylindrical wheels along the z axis—W0 at the detector center, W$\pm$1, and W$\pm$2 in the forward and backward directions. Each wheel is divided into 12 sectors, each spanning 30° in azimuth. Along the radial direction, the DTs are organized into four stations located at radii of approximately 4.0, 4.9, 5.9, and 7.0 m from the beam axis, interleaved with the steel plates of the return yoke. Each drift cell has transverse dimensions of 42×13 mm^2 with a 50 μm anode wire at the center and is filled with an Ar/CO_2 (85%/15%) gas mixture, which provides efficient quenching, a saturated drift velocity of about 55 μm/ns, and a maximum drift time of 400 ns [81]. The effective electric field is shaped by four electrodes—two cathode strips on the sides and two additional electrodes located on the ground planes between cell layers.

Each DT chamber consists of three superlayers (SLs), each made of four staggered layers of parallel drift cells. Two SLs measure the r–ϕ coordinate (with wires parallel to the

beam line), while the third measures the r–z coordinate. The spatial resolution of a single cell is about 200 μm, yielding a combined chamber resolution of 80–120 μm in position and approximately 1 mrad in angle. In addition to precise tracking, the DTs also provide a fast trigger signal with a timing resolution of about 5 ns, correlated with the muon transverse momentum.

In the endcap regions ($0.9 < |\eta| < 2.4$), where the particle rates and magnetic field gradients are significantly higher, cathode strip chambers (CSCs) are used. CSCs are multiwire proportional chambers with cathode strip readout that can tolerate high radiation levels, non-uniform magnetic fields, and high occupancy [82]. Each chamber has a trapezoidal shape and consists of six anode wire layers interleaved with seven orthogonal cathode strip planes, allowing simultaneous measurements in the bending (r–ϕ) and longitudinal (z) directions. The chambers are organized in four stations per endcap, mounted perpendicular to the beam line. They operate with a gas mixture of 50% CO_2, 40% Ar, and 10% CF_4. When a muon passes through the chamber, the gas is ionized, and the intersection of hit strips and wire groups defines the reconstructed hit position, with a typical resolution of 100 μm in r and 10 mrad in ϕ. The timing is calibrated such that hits from promptly produced muons are centered at zero. Together, the DTs and CSCs provide continuous coverage up to $|\eta| < 2.4$ without acceptance gaps, ensuring efficient muon identification throughout this region.

To complement these systems with precise timing capabilities, CMS employs resistive plate chambers (RPCs), which provide an excellent time resolution of about 1 ns and are used in both triggering and bunch-crossing identification. RPCs cover the range $|\eta| < 1.6$ and are installed in both the barrel and endcap regions. Each RPC consists of two parallel bakelite plates, 2 mm thick, separated by a narrow gas gap filled with a mixture of 95.2% Freon ($C_2H_2F_4$), 4.5% isobutane (i-C_4H_{10}), and 0.3% SF_6, with humidity regulated between 40–50%. Operated in avalanche mode, a charged particle traversing the gas volume initiates a localized avalanche in the strong electric field. The induced image charge is collected on external readout strips aligned along the η direction, with a typical pitch of up to 2 cm. This provides a spatial resolution between 0.8 and 1.2 cm. Because the ionization charge can spread over multiple adjacent strips, clusters are combined, and the hit position is reconstructed as the centroid of the cluster. The excellent time resolution of RPCs enables precise bunch-crossing identification, which is crucial at high LHC luminosities.

Trigger System

At the LHC design luminosity of 10^{34} cm^{-2}s^{-1}, the proton–proton interaction rate exceeds 1 GHz. Most of these interactions are inelastic, producing low-energy multi-jet final states with a cross section of approximately $\sigma_{\text{inel}} \sim 70$ mb—several orders of magnitude larger than the cross sections of the rare processes targeted by the CMS physics program ($\sigma < 10^5$ pb). Given that bunch crossings occur every 25 ns, the detector records an immense volume of data, corresponding to an aggregate throughput of roughly 70 Tb/s. It is neither technically feasible nor scientifically useful to read out, store, and process all these events.

The CMS trigger system is therefore designed to efficiently identify potentially interesting collisions in real time, while rejecting the overwhelming majority of uninformative events. Its purpose is to reduce the event rate from the 1 GHz collision frequency to a manageable level of about 10^5 events/s for full detector readout, and further down to approximately

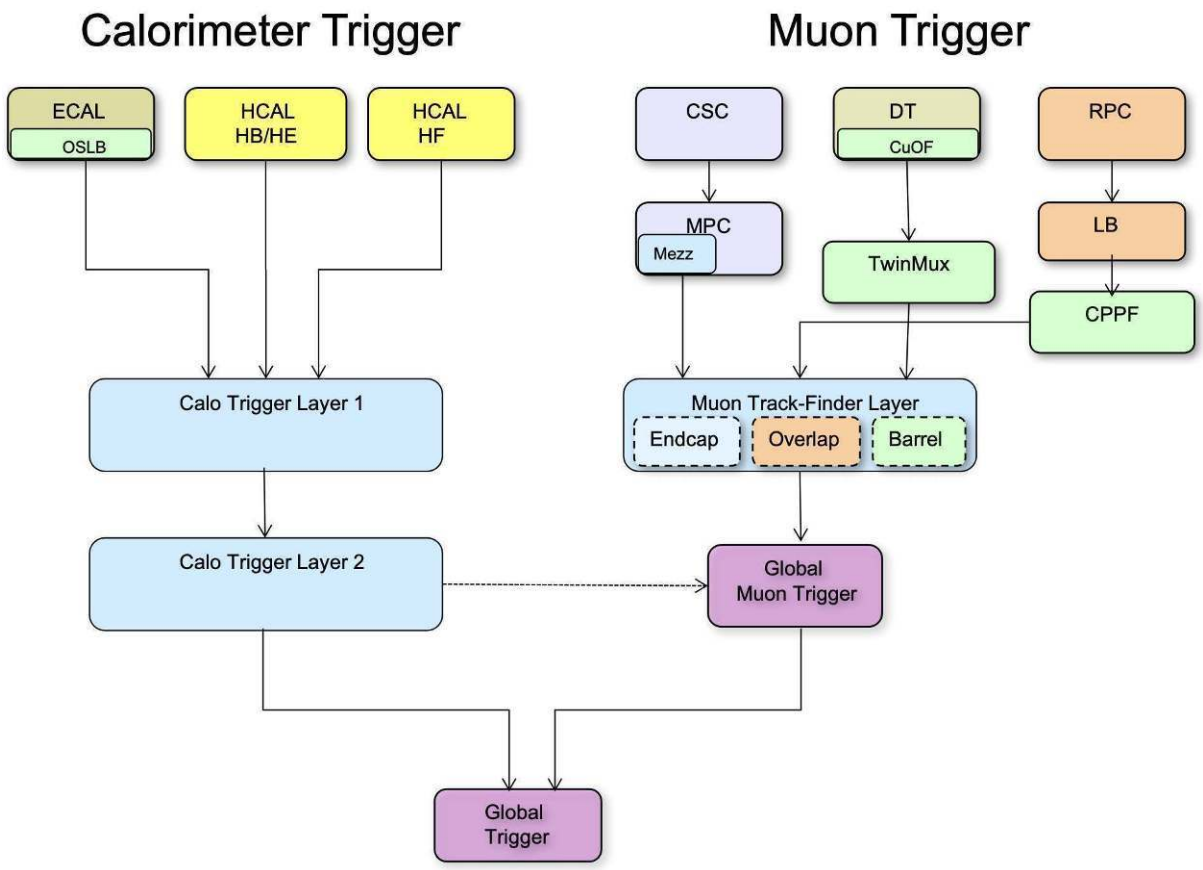

Figure 1.11: Schematic representation of the CMS Level-1 (L1) trigger system [83].

10^3 events/s for permanent offline storage and analysis.
To accomplish this task, CMS employs a two-level trigger architecture composed of the *Level-1 Trigger* (L1) and the *High-Level Trigger* (HLT). The L1 trigger is implemented entirely in custom hardware and firmware, designed for sub-microsecond decision times using coarse but fast information from the calorimeters and muon systems. The HLT, on the other hand, operates in software on a large computing farm and performs a refined event reconstruction using the full detector information available after the L1 selection. Together, these two systems enable CMS to retain the most interesting events for physics analyses while operating efficiently under the extreme data-taking conditions of the LHC.

Level-1 Trigger

The Level-1 (L1) trigger is implemented entirely in custom hardware and processes detector information in real time with fixed latency. It uses coarse calorimeter and muon data, available at reduced spatial granularity compared to the full detector readout, to make the initial event selection. Operating with a constant latency of 3.8 μs, the L1 system decides whether to retain or reject an event before data from the following bunch crossings overwrite the buffers. At this stage, the event rate is reduced from the LHC bunch crossing frequency of 40 MHz to about 100 kHz—compatible with the CMS readout and data acquisition system.
Currently, tracker information is not available to the L1 trigger, as reading out the entire tracker at 40 MHz is not technically feasible. Dedicated hardware boards are under

development for the High-Luminosity LHC (HL-LHC) upgrade [84], which will enable the outer tracker to contribute to L1 decisions. The absence of tracking information limits the ability of L1 to separate overlapping interactions from multiple collisions within the same bunch crossing and prevents distinguishing electrons from photons, which appear identical at this stage.

The L1 event selection relies on a set of configurable algorithms known as *seeds*. Each seed applies requirements on one or more candidate objects reconstructed from energy deposits or ionization signals corresponding to muons, electrons, photons, jets, τ leptons, or missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$). These L1 objects are built from low-level inputs called *trigger primitives*, produced independently by each subdetector. Most seeds require a minimum number of candidate objects, typically one or two, satisfying thresholds on transverse energy (E_{T}) and pseudorapidity. Single- and double-object seeds occupy about 75% of the available bandwidth, and as of 2018, the L1 system can host up to 512 algorithms in total.

A schematic of the L1 architecture is shown in Figure 1.11 [83]. Two parallel processing chains handle the calorimetric information (from ECAL, HCAL, and HF) and the muon subsystems (DT, CSC, and RPC). Their outputs are combined in the micro Global Trigger (μGT), which makes the final L1 decision to accept or reject an event.

Experience from Run I led to a major upgrade of the calorimeter trigger [85], increasing the effective readout granularity from 4×4 trigger towers to full trigger-tower resolution. The typical L1 calorimeter granularity is $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ radians across most of the acceptance (slightly coarser at high $|\eta|$). To process this fine-grained information within latency constraints, CMS adopted a time-multiplexed architecture composed of two layers. The first layer (Layer-1) performs preprocessing—such as energy calibration and synchronization—on advanced mezzanine cards equipped with Xilinx Virtex-7 FPGAs (CTP7 boards) [86]. Eighteen Layer-1 boards operate in parallel, forwarding calibrated trigger primitives to the second layer (Layer-2), which reconstructs higher-level physics objects such as jets, τ leptons, electrons, and photons using dedicated shape and isolation algorithms. Layer-2 nodes employ MP7 processor cards, also based on Virtex-7 FPGAs [83], designed for flexible firmware implementation. A demultiplexer (DeMux) board then reorders and formats the processed information for the μGT.

The muon L1 trigger reconstructs muon candidates by combining inputs from the DT, CSC, and RPC detectors. It is divided into three pseudorapidity regions, each handled by a dedicated track-finding subsystem: the Barrel Muon Track Finder (BMTF), the Overlap Muon Track Finder (OMTF), and the Endcap Muon Track Finder (EMTF). The BMTF ($|\eta| < 0.83$) combines DT and RPC primitives into "super-primitives" using the TwinMUX system, which enhances hit-position precision and incorporates bending-angle information for track finding via road-search algorithms. In the overlap region ($0.83 < |\eta| < 1.24$), the OMTF merges information from all three muon subsystems, while the EMTF ($|\eta| > 1.24$) reconstructs endcap muons from CSC and RPC data. Each subsystem operates with a latency of about 750 ns, using advanced pattern-recognition algorithms to assign muon transverse momentum and quality criteria. Each track finder can transmit up to 36 muon candidates to the Global Muon Trigger (GMT), which resolves duplicates across systems and forwards up to eight highest-quality candidates, ordered by p_{T}, to the μGT.

Different hardware solutions are employed for the three muon track finders, optimized for their algorithmic complexity and resource needs. The BMTF uses MP7 cards for intensive computations, while the OMTF and EMTF rely on modular track-finder boards equipped

with large memory resources for pattern-recognition operations [87].
The final L1 trigger decision is taken by the μGT, which combines calorimetric and muon information to evaluate complex multi-object correlations and global event quantities. It computes, for example, invariant masses of object pairs, spatial separations, and vectorial momentum sums. The μGT executes all L1 algorithms in parallel on MP7 processor boards, offering substantial computational power. During Run II, the CMS μGT configuration included approximately 350-400 active algorithms.

High-Level Trigger

Events accepted by the Level-1 (L1) trigger are passed to the High-Level Trigger (HLT) for further selection. While the L1 system reduces the event rate to a level sustainable for detector readout, the purpose of the HLT is to identify events of physics interest at a rate manageable for permanent offline storage. The HLT operates under less stringent latency constraints than L1, but must still satisfy real-time requirements while ensuring that trigger efficiencies can be accurately measured from data.
Unlike the fixed-latency, hardware-based L1 system, the HLT is fully software-driven and runs on a dedicated processor farm known as the *Event Filter Farm* (EVF) [88]. This farm consists of commercial multi-core CPUs running Scientific Linux, and is divided into two main components: *builder units* and *filter units*. The builder units retrieve event fragments from the detector front-end corresponding to accepted L1 triggers and assemble them into complete events. The filter units then unpack the raw data, perform fast reconstruction, and apply the trigger selection algorithms. Both functions are implemented on the same physical machines, which in total comprise roughly 36 000 CPU cores. The HLT processes the L1 input rate of about 100 kHz and reduces it to a final output rate below 1 kHz, limited by event size and the downstream storage and processing capacity. The average decision latency per event is approximately 260 ms, with a dependence on event complexity; the most demanding events can require processing times up to $\mathcal{O}(1\text{ s})$.
Events accepted by the HLT are first handled by the *Storage Manager*, which writes them to local disks before transferring them to the CMS Tier-0 facility for prompt reconstruction and permanent archiving. Roughly half of these events are reconstructed promptly (within 48 hours) using the full offline algorithms, while the remaining half—typically lower-priority datasets—are *parked* for later processing.
The HLT is organized around the concept of *paths*, which are predefined sequences of algorithmic modules executed in a specific order. Each HLT path is seeded by one or more L1 triggers and is designed to progressively reconstruct and select physics objects with increasing precision. Early steps in each path apply coarse selection criteria based on calorimetric or muon information, significantly reducing the event rate before more computationally intensive tracking algorithms are invoked. At this stage, coarse reconstruction of jets, τ leptons, electrons, photons, and muons typically consumes about 50 ms per event. Subsequent inclusion of tracking information enables refined object identification: distinguishing electrons from photons, identifying heavy-flavour jets and hadronically decaying τ leptons, evaluating lepton isolation, and precisely measuring lepton momenta. To reduce computational cost, tracking is usually performed only in regions of interest (regional tracking) and for a subset of events.
The final stage of the reconstruction sequence employs the *particle-flow* (PF) algorithm [89], which provides the most accurate event description achievable within the HLT framework. The PF sequence combines information from all CMS subdetectors to

reconstruct individual particles, using a simplified but highly optimized version of the offline algorithm. Although it is the most time-consuming step in the HLT processing chain, the PF reconstruction significantly enhances the physics performance of the online selections, bringing them close to offline quality.

Chapter 2

VBF-Z Analysis

2.1 Introduction

The study of the vector boson fusion (VBF) production of Z bosons is an important probe of the electroweak sector of the SM and offers sensitivity to potential new physics beyond it. The VBF process, characterized by the presence of two forward jets and a centrally produced Z boson, provides a clean environment for precision measurements of electroweak interactions. Such measurements are crucial for testing the SM predictions and for constraining EFT operators that could describe deviations arising from new physics.

Previous analyses [90, 91] of Z boson production through VBF have demonstrated the capability of the LHC experiments to measure this process with high precision. The CMS collaboration has previously measured the inclusive cross section for Z boson production via VBF using the 2016 dataset corresponding to 35.9 fb^{-1} of integrated luminosity at a center-of-mass energy of 13 TeV. This analysis provided an inclusive measurement but did not explore the differential cross sections as a function of kinematic variables, which are essential for a more detailed understanding of the underlying dynamics and for enhancing sensitivity to new physics effects.

In contrast, the ATLAS collaboration has recently published a comprehensive analysis using the full Run 2 dataset, which corresponds to an integrated luminosity of 139 fb^{-1} at 13 TeV. Their study included differential cross section measurements as a function of various observables, such as the transverse momentum (p_T) and rapidity (y) of the Z boson, as well as the properties of the associated jets. The differential approach offers a more sensitive probe of the SM and provides an opportunity to test for deviations through an EFT interpretation. This allows for constraints on the coefficients of dimension-6 operators, thereby exploring possible new physics scenarios.

Building upon these previous studies, we present an analysis using the full Run 2 dataset collected by the CMS experiment, corresponding to an integrated luminosity of 138 fb^{-1} at 13 TeV. Our analysis measures both inclusive and differential cross sections for VBF Z boson production, exploring various kinematic distributions. Additionally, we interpret these results using the EFT framework to set constraints on dimension-6 operators, offering a comprehensive test of the SM and potential deviations arising from new physics. This study represents the most detailed exploration of VBF Z production at CMS to date and extends the understanding of the electroweak sector through precision measurements.

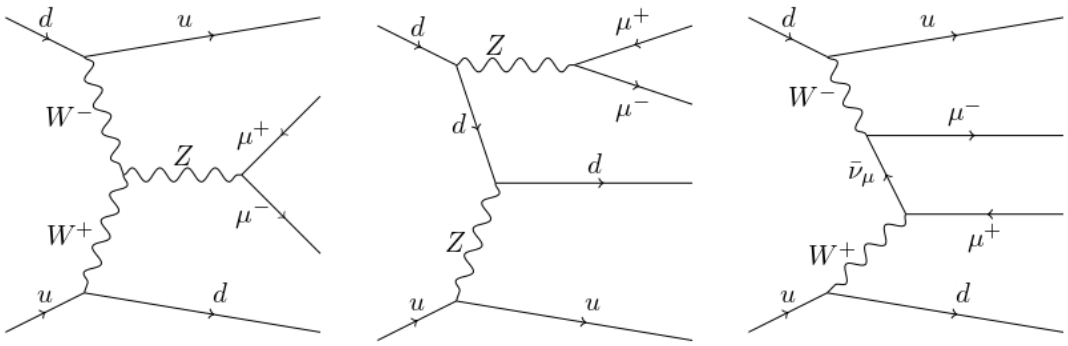

Figure 2.1: EW signal Feynman diagrams: VBF on the left, non-resonant on the center and right

2.2 Vector Boson Fusion signal

With vector boson fusion, we refer to a process where two vector bosons interact producing a new boson, be it the Z as in this thesis, as well as W or the Higgs. It is an electroweak (EW) process, meaning no QCD vertices are involved in the Feynman diagrams and the cross-section is small compared to other processes.
Weak vector bosons, $W^{\pm}$ and Z^0, are massive (respectively $80.39 \pm 0.02\ GeV/c^2$ and $91.188 \pm 0.002\ GeV/c^2$) and they gain their mass thanks to their interactions with the Higgs field. They differ from the other two vector bosons, the gluon and photon, because of a lacking vertex of Feynman diagrams, where the latter bosons are directly interacting with the Higgs. This is due to the Electroweak symmetry breaking of SU(2) $\times$ U(1) Yang-Mills theory, and the Higgs mechanism that converts three Goldstone bosons in additional degrees of freedom for the $W^{\pm}$ and the Z^0.
VBF gives access to one of the core concepts of the EW sector of the SM: the triple gauge coupling (TGC), where three bosons interact. The TGC of main interest for the SM analysis is depicted at the center of the first diagram in Figure 2.1, where a W^- and a W^+ fuse into a Z. The latter vertex may be conceived as an effective vertex, in which new physics could exist. Therefore a precise measurement of the latter might lead to deviations from the TGC theorized by the SM. This is why, in this thesis, the EFT study is conducted with VBF.
The only channel considered in this thesis is the fully leptonic decay of the Z into charged leptons. This was chosen because of its better reconstruction and signal-to-background ratio.
The process is thus:

$$p\,p \longrightarrow Z\,j\,j \longrightarrow l^+\,l^-\,j\,j \tag{2.1}$$

and the requirement of not having any QCD vertices at Leading Order.
In order not to break gauge invariance, the inclusive signal is the whole:

$$p\,p \longrightarrow l^+\,l^-\,j\,j \tag{2.2}$$

which includes also two kinds of non resonant diagrams which are reported in figure 2.1.

2.3 Backgrounds

The Drell Yan process, where a quark and an antiquark annihilate into a colorless neutral boson, is the dominant background as it presents similar topology and the same final state as the signal, thus forming an irreducible background with a cross section many times larger than the EW-Z process.
Two QCD induced Feynman diagrams are reported in Figure 2.2, they are of the order of α_S^2 and belong to the DY process. They constitute the main background for this measurement. Dedicated selections are exploited in order to reduce the DY process while retaining a high number of signal events. However its cross-section is larger than the EW Zjj by a factor of $O(100)$, so even within a narrow phase space where the signal yield is enhanced, the DY will be dominant in the number of events by a factor $O(10)$.

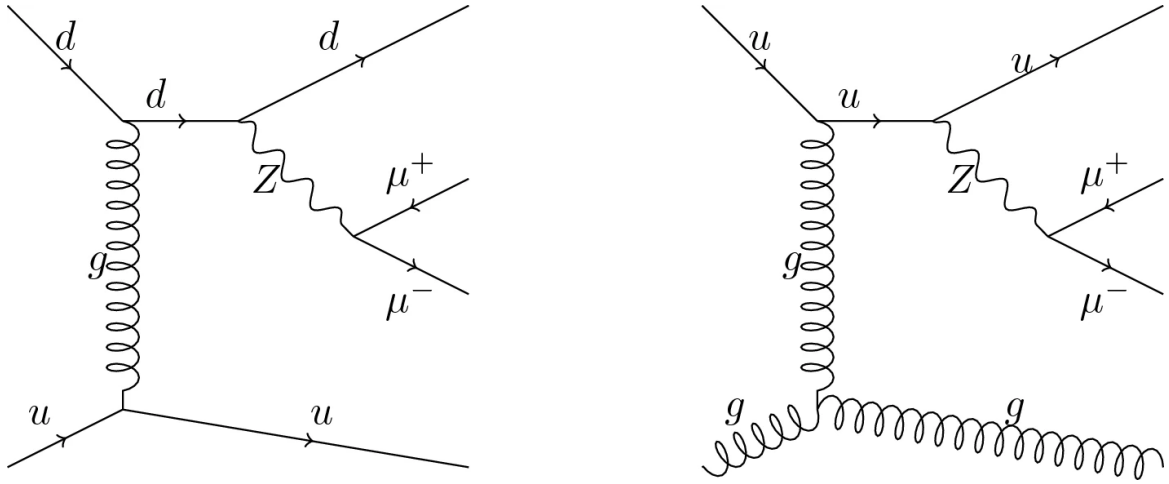

Figure 2.2: Examples of Feynman diagrams of Drell-Yan process with the addition of Initial State Radiation (ISR).

Second for importance is the background class of the multi-boson processes where two or more vector bosons are produced. In the case where one of the bosons decay leptonically and the other decays hadronically, the event will have the same signature as the signal.
Finally, the processes where a top-quark or a pair of top-quarks are produced are considered, since e.g. two top-quarks will decay into two b-jets and two W bosons which in the fully leptonic decay would lead to a final state with two leptons, two neutrinos but most importantly two b-jets.
The presence of b-jets in the final state is a characteristic signature that can be identified by means of b-tagger exploiting properties of long-lived b-hadrons. A pure region enriched in top events can be designed by requiring b-jets in the event. On the other hand top events can be suppressed from other regions by imposing a veto on b-jets.

2.4 Data and Monte Carlo Samples

2.4.1 Data samples

The analysis is based on the full dataset of proton-proton collisions collected by the CMS experiment during Run 2 of the LHC, corresponding to an integrated luminosity of approximately 138 fb^{-1} at a center-of-mass energy of 13 TeV. The dataset includes data

from the 2016, 2017, and 2018 data-taking periods, each recorded under slightly different detector and accelerator conditions.
The 2016 data, corresponding to about 36 fb^{-1}, are divided into two distinct periods known as *pre-VFP* and *post-VFP*. This split accounts for changes in the readout conditions of the silicon tracker electronics introduced during the year to improve charge collection and timing performance. Treating the two periods separately ensures that the associated simulation accurately reflects the detector behavior before and after this adjustment. The 2017 and 2018 datasets, corresponding to integrated luminosities of approximately 41 fb^{-1} and 60 fb^{-1} respectively, are each treated as single, coherent data-taking periods.
Events are selected using a combination of single- and double-lepton triggers to ensure high efficiency for the dilepton final state characteristic of $Z \to \ell^+\ell^-$ decays. For the muon channel, the single-muon trigger `HLT_IsoMu24` requires at least one muon with transverse momentum $p_{\text{T}} > 24$ GeV passing an isolation requirement at HLT level (`Iso`). In addition, a dimuon trigger `HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8` is used, requiring two muons with asymmetric thresholds ($p_{\text{T}} > 17$ GeV for the leading leg and $p_{\text{T}} > 8$ GeV for the subleading leg). Both muon legs must satisfy a very-very-loose track-based isolation working point (`TrkIsoVVL`); a longitudinal impact-parameter compatibility requirement between the two muons (`DZ`) is applied to favor a common primary vertex; and a minimum dimuon invariant mass requirement (`Mass3p8`, i.e. $m_{\mu\mu} > 3.8$ GeV) suppresses very low-mass backgrounds.
For the electron channel, the single-electron trigger `HLT_Ele32_WPTight_Gsf` requires one electron with $p_{\text{T}} > 32$ GeV reconstructed with the Gaussian-sum filter tracking (`Gsf`) and passing a tight HLT identification and isolation working point (`WPTight`). The di-electron trigger `HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL` requires two electrons with $p_{\text{T}} > 23$ GeV and $p_{\text{T}} > 12$ GeV for the leading and subleading legs, respectively, along with loose calorimeter- and track-based identification criteria (`CaloIdL`, `TrackIdL`) and a very-loose isolation requirement (`IsoVL`). Offline lepton kinematic and identification selections are chosen such that the trigger efficiencies are on the plateau of these HLT requirements.
Only periods in which all detector subsystems were fully operational and certified for physics analysis are included, ensuring data of consistent quality across all years.
The corresponding Monte Carlo simulations are produced using the same year-dependent detector configurations and are normalized to the integrated luminosity of each data-taking period. This consistent treatment of data and simulation guarantees that differences in detector conditions and reconstruction algorithms are properly accounted for in the final measurement.

2.4.2 Simulated samples

The samples used for the EW Zjj signal and the EW QCD interference are reported in Table 2.1 together with their cross sections. The EW Zjj, EW Zjj EFT and EW QCD interference are generated with Madgraph at Leading Order and showered with Pythia. The difference between the EW Zjj and the EW Zjj EFT samples is that the latter includes the set of SMEFT operators of the analysis, c_W, c_{HWB}, c_{HDD}, implemented via the SMEFTsim package [59] and using 1.0 as Wilson coefficient values for all the operators.
Two additional samples are used for the comparison with the unfolded results and both

give NLO predictions: POWHEG and MadGraph@NLO. The cross-sections for the EW Zjj signal differ due to different generator level selections. After applying the analysis selections, the effective cross-sections are closer to each other but a remaining 10-20% difference is observed between LO and NLO.

Sample name	σ **(pb)**
EW Zjj MG@LO	1.719 (LO)
EW Zjj EFT MG@LO	2.217 (Full BSM LO)
Zjj interference MG@LO	0.2102 (LO)
EW Zjj Powheg	2.255 (NLO)
EW Zjj MG@NLO	6.761 (NLO)

Table 2.1: Signal and interference cross-sections. EW Zjj samples have different generator level selections.

The main background in this analysis is Drell-Yan (DY) production, whose simulated samples are summarized in Table 2.2. All DY samples are generated with the program MadGraph5_aMC@NLO at NLO order accuracy in QCD and interfaced with Pythia8 for parton showering and hadronization. The FxFx merging scheme [92] is employed to consistently combine matrix-element calculations with up to two additional partons at NLO with the parton shower, avoiding double counting of radiation and improving the modeling of jet-related observables. All samples have a parton level cut on the dilepton invariant mass: $m_{ll} > 50$ GeV.
Three complementary sets of DY samples are used in the analysis:

- **Inclusive samples:** cover the full DY phase space with NLO accuracy but limited statistical precision due to their smaller size.
- **Jet-binned samples:** generated in exclusive jet multiplicity bins, providing a better description of events with additional jets and moderate statistics.
- p_T^Z**-binned samples:** generated in bins of the transverse momentum of the Z boson, ensuring high statistical precision in the kinematic tails and in the regions most sensitive to the electroweak Zjj signal, see Figure 2.3.

Cross sections for the inclusive and jet-binned samples are taken from the FEWZ prediction providing NLO EW and NNLO QCD accurate results, Ref. [93]. For the p_T^Z-binned instead, the cross sections are computed at NLO accuracy directly from MadGraph5_aMC@NLO. Given the superior statistical power and coverage of the p_T^Z-binned samples, they are used as the nominal DY background prediction in the main analysis, while the jet-binned samples are employed for the DNN training to ensure a consistent description of the event kinematics across different jet multiplicities.

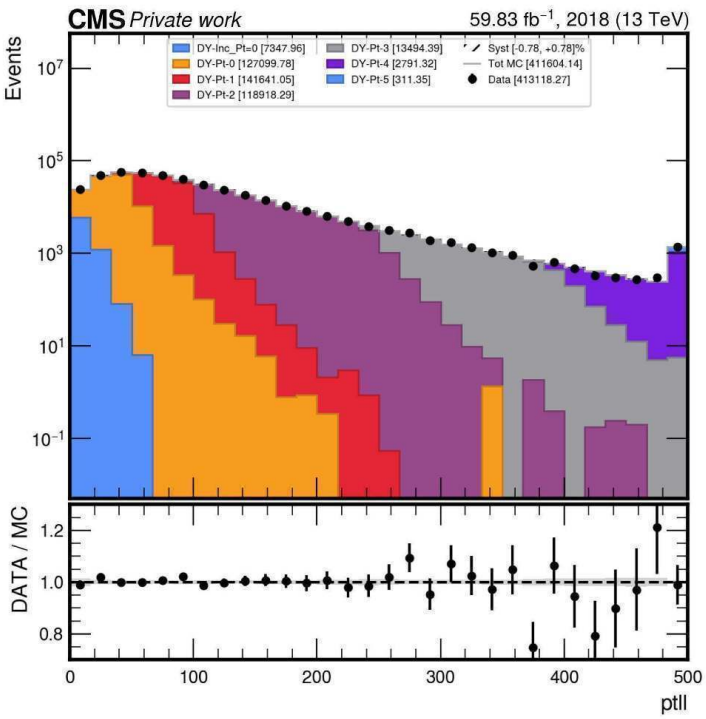

Figure 2.3: Comparison of the dilepton transverse momentum ($p_T^{\ell\ell}$) distribution in the signal region between the inclusive DY sample (Data in the plot) and the stitched DY p_T^Z-binned samples, demonstrating good agreement after stitching and most importantly a reduced statistical uncertainty.

Sample	**Description**	**σ (pb)**
DY inclusive	Inclusive sample	6077.22 (NNLO FEWZ)
DY 0J	0-jet bin	4620.52 (NNLO)
DY 1J	1-jet bin	859.59 (NNLO)
DY 2J	2-jet bin	338.26 (NNLO)
DY p_T^Z 0-50	p_T^Z bin: 0-50 GeV	1414.50 (NLO)
DY p_T^Z 50-100	p_T^Z bin: 50-100 GeV	354.30 (NLO)
DY p_T^Z 100-250	p_T^Z bin: 100-250 GeV	83.12 (NLO)
DY p_T^Z 250-400	p_T^Z bin: 250-400 GeV	3.047 (NLO)
DY p_T^Z 400-650	p_T^Z bin: 400-650 GeV	0.3921 (NLO)
DY $p_T^Z > 650$	$p_T^Z > 650$ GeV	0.03636 (NLO)

Table 2.2: Cross sections of the Drell-Yan samples used in the analysis. All samples are generated with MG5_aMC@NLO + Pythia8 using FxFx merging.

The diboson background processes (WW, WZ, and ZZ) are simulated using Pythia8 with the CUETP8M1 tune. These samples account for contributions where two electroweak gauge bosons are produced, either both decaying leptonically or involving additional jets in the final state. Although their overall contribution to the selected event yield is small compared to the Drell-Yan process, they are included for completeness in the background estimation. The cross sections used for normalization are summarized in Table 2.3.

Process	σ (pb)
WW	75.8
WZ	27.6
ZZ	12.1

Table 2.3: Cross sections of diboson background samples generated with PYTHIA8 (CUETP8M1 tune).

Top-quark related backgrounds include both the pair-production ($t\bar{t}$) process and the single-top production channels (tW, t-channel, and s-channel). All samples are generated using POWHEG or MG5_AMC@NLO interfaced with PYTHIA8 for parton showering and hadronization, employing the CP5 tune. The tW and single-top processes are simulated in the five-flavor scheme and include both top and anti-top contributions. These backgrounds are particularly relevant in the top control region and in regions with high jet activity or b-tagged jets. The cross sections used for normalization are summarized in Table 2.4.

Process	σ (pb)
$t\bar{t} \to 2\ell 2\nu$	87.3
tW (top)	35.9
tW (anti-top)	35.9
Single top t-ch. (top)	44.1
Single top t-ch. (anti-top)	26.2
Single top s-ch.	3.34

Table 2.4: Cross sections of top-quark background samples generated with POWHEG or MG5_AMC@NLO + PYTHIA8 (CP5 tune).

2.5 Physics reconstruction and online selections

A consistent and precise definition of reconstructed physics objects is essential for any measurement at the LHC. In this analysis, electrons, muons, jets, and missing transverse momentum are reconstructed and calibrated following the standard procedures developed within the CMS collaboration. These common definitions ensure compatibility with other measurements targeting similar final states and rely on well-validated selections and calibration strategies.

Each object type is subject to specific reconstruction and identification requirements designed to balance efficiency and purity. Corrections are applied to the simulated samples to reproduce the performance observed in data, including scale factors for identification, isolation, and trigger efficiencies, as well as energy scale and resolution calibrations. The corresponding uncertainties are propagated as systematic variations in the final statistical interpretation.

These calibrations and efficiency measurements are obtained using control samples of well-understood processes, such as $Z \to \ell^+\ell^-$ and photon+jet events, through methods like tag-and-probe or in-situ energy balancing. Applying these corrections ensures that the Monte Carlo simulation accurately models detector performance and maintains a consistent physics interpretation across all data-taking years.

The following sections summarize the reconstruction, selection, and calibration criteria for each physics object used in the analysis and describe how trigger efficiencies are measured and applied to achieve a uniform, data-driven treatment of detector effects.

2.5.1 Muons

Muon reconstruction and identification are essential for this analysis, as the signal topology involves final states with two high-quality muons originating from a $Z \to \mu^+\mu^-$ decay. The selection is based on the standard CMS tight identification (98% efficiency) and particle flow based isolation criteria, optimized to ensure high purity while maintaining good efficiency over the full kinematic range of the analysis.
Muons are required to have pseudo-rapidity within $|\eta| < 2.4$ and either a transverse momentum above 20 GeV and be quite displaced from the vertex ($|d_{xy}| < 0.2$ and $|d_z| < 0.1$) or $p_T < 20$ GeV and be close to the vertex ($|d_{xy}| < 0.01$ and $|d_z| < 0.1$). To suppress background contributions from hadrons misidentified as muons or from secondary decays, a particle-flow-based relative isolation variable is used, which accounts for energy deposits in a cone around the muon direction and corrects for residual pileup effects.

Muon Energy Scale Corrections

An accurate calibration of the muon momentum is crucial, as it directly affects the reconstructed invariant mass of the dilepton system, $m_{\mu\mu}$, which is used to identify Z-boson candidates. To ensure a precise and uniform momentum scale across the detector, data-driven energy scale corrections, known as Rochester corrections [94], are applied.
These corrections compensate for detector-related effects such as alignment, magnetic field variations, and residual biases in the reconstruction. The calibration constants are derived from well-known resonances, notably the $Z \to \mu^+\mu^-$ process, by comparing the reconstructed and true invariant mass distributions in both data and simulation.
Applying these corrections leads to an improved description of the Z-boson peak and reduces systematic uncertainties associated with the muon momentum measurement. As a result, the reconstructed dilepton invariant mass is unbiased and stable across different detector regions, providing the accuracy required for precision measurements in vector boson fusion Z-boson production.

2.5.2 Electrons

Electrons play a key role in this analysis, as they are used to reconstruct $Z \to e^+e^-$ decays that define the signal final state. The selection aims to identify prompt, isolated electrons produced in the hard interaction, while suppressing backgrounds from jets misidentified as electrons, photon conversions, and electrons from heavy-flavour decays.
Electron candidates are reconstructed by combining information from the electromagnetic calorimeter (ECAL) and the inner tracking detector. The resulting tracks and energy clusters are matched to ensure an accurate measurement of the electron energy and direction. Selected electrons are required to have transverse momentum above 13 GeV and pseudo-rapidity within $|\eta| < 2.5$, excluding the barrel-endcap transition region where the detector response is less reliable.
To ensure a high level of purity, electrons must satisfy tight identification (with 90% efficiency) and isolation requirements. The identification relies on a multivariate algorithm that combines several observables related to shower shape, track-cluster matching, and

track quality. This approach efficiently separates genuine prompt electrons from misidentified hadrons or secondary sources. Electron isolation is computed using the sum of nearby particle-flow candidates within a cone around the electron direction, corrected for contributions from additional proton-proton interactions (pileup).

Energy Scale and Efficiency Corrections

Accurate energy calibration is essential for a precise reconstruction of the dilepton invariant mass, m_{ee}, which is used to identify Z-boson candidates. The electron energy scale is corrected using data-driven methods based on well-known resonances, such as $Z \to e^+e^-$, to align the reconstructed mass peak in data and simulation. Additional corrections are applied to account for differences in the detector response between the barrel and endcap regions.
Residual discrepancies between data and simulation in the identification and isolation efficiencies are corrected through scale factors derived using the "tag-and-probe" technique applied to $Z \to e^+e^-$ events. These factors ensure that the efficiency of the electron selection is accurately reproduced in simulated samples, and their uncertainties are propagated as systematic effects in the analysis.
The combination of these reconstruction, identification, and calibration procedures ensures a uniform and well-understood electron performance across the detector, providing the precision required for differential and effective field theory measurements of vector boson fusion Z-boson production.

2.5.3 Lepton trigger efficiencies

The identification of events containing high-quality leptons requires an efficient and well-calibrated online trigger system. Throughout Run 2, a set of single- and double-lepton triggers was employed to select final states with electrons or muons, maintaining high acceptance over a broad range of transverse momenta and detector pseudorapidities. These triggers are particularly important for the collection of $Z \to \ell^+\ell^-$ decays, which provide the main event sample used in this analysis.
Small differences in trigger performance between data and simulation are corrected by applying scale factors (SFs) derived from dedicated efficiency measurements. These efficiencies are obtained with the standard *tag-and-probe* method using $Z \to \ell^+\ell^-$ events, where one lepton (the "tag") satisfies tight offline and trigger requirements, while the second lepton (the "probe") is used to assess the trigger response. To suppress backgrounds and ensure a clean control sample, the invariant mass of the lepton pair is required to lie near the nominal Z-boson mass.
The trigger efficiency ε is defined as

$$\varepsilon = \frac{N_{\text{passing probes}}}{N_{\text{all probes}}},$$

and is measured in both data and simulation as a function of lepton transverse momentum and pseudorapidity. The corresponding SF applied to simulated events is given by

$$\text{SF} = \frac{\varepsilon_{\text{data}}}{\varepsilon_{\text{MC}}},$$

thereby correcting residual discrepancies in the trigger modeling.

Separate calibrations are derived for electrons and muons in each data-taking year (2016, 2017, and 2018) to account for evolving detector conditions and trigger menus. For the 2016 dataset, an additional distinction is made between the *pre-VFP* and *post-VFP* periods to reflect differences in the detector readout configuration.
Applying these corrections ensures that the simulation accurately reflects the trigger performance observed in data. The associated uncertainties, typically at the level of a few per mille, are propagated through the analysis as correlated systematic effects and are included in the final statistical interpretation.

2.5.4 Jets

Jets are key observables in this analysis, as the vector boson fusion (VBF) topology is characterized by the presence of two energetic jets with large separation in pseudorapidity and high invariant mass. Their accurate reconstruction and calibration are essential to identify the VBF process and to measure kinematic distributions sensitive to potential new physics effects.

Jet Reconstruction

Jets are reconstructed using the anti-k_T clustering algorithm [95] with a distance parameter of $R = 0.4$ (AK4 jets). The reconstruction is based on particle-flow (PF) candidates, which combine information from all CMS subdetectors to provide a complete list of reconstructed particles. To reduce the impact of additional proton-proton interactions occurring in the same bunch crossing (pileup), the Charged Hadron Subtraction (CHS) technique is employed, which removes charged particles not associated with the primary interaction vertex. This improves the accuracy of jet momentum reconstruction and ensures a more stable jet response across varying pileup conditions.

Jet Energy Corrections (JECs)

The energy measured for a reconstructed jet does not directly correspond to the true particle-level jet energy, due to effects such as non-uniform calorimeter response, residual pileup contamination, and detector noise. To correct for these effects, a set of multiplicative calibration factors, referred to as Jet Energy Corrections (JECs), are applied. These corrections are derived from a combination of simulation and in-situ measurements using well-known reference processes, such as dijet and photon+jet events. After the corrections are applied, the reconstructed jet energy scale is uniform across the detector and consistent between data and simulation.
Among the various components of the JECs, one of the key corrections is the absolute p_T-dependent residual correction, which ensures that the jet energy response is properly modeled as a function of the jet transverse momentum. This correction is obtained from a detailed data-to-simulation comparison of the jet response in several reference channels, including $Z(\to \mu\mu)$+ jet, $Z(\to ee)$+ jet, γ+ jet, multijet, and hadronic $W \to q\bar{q}'$ decays in $t\bar{t}$ events. The combination of these processes allows the calibration to cover a wide p_T range and to control different sources of systematic uncertainty.
In these studies, both the Missing p_T Projection Fraction (MPF) and Direct Balance (DB) methods are used to measure the jet response, with the MPF method from Z+ jet events providing the dominant precision. The calibration procedure is performed separately for each data-taking period, and for the 2016 and 2018 datasets, additional

constraints from the particle-flow jet composition are included to reduce degeneracies among fit parameters. Post-fit values of the nuisance parameters, which encode the scales of the reference objects, demonstrate the internal consistency between the different datasets and reference channels.
Figure 2.4 shows the absolute p_T-dependent residual correction derived from these measurements. The yellow band indicates the systematic uncertainty associated with this correction, which remains below 1% across a large portion of the phase space. This high level of precision ensures a robust and consistent jet energy scale across the full kinematic range used in this analysis.

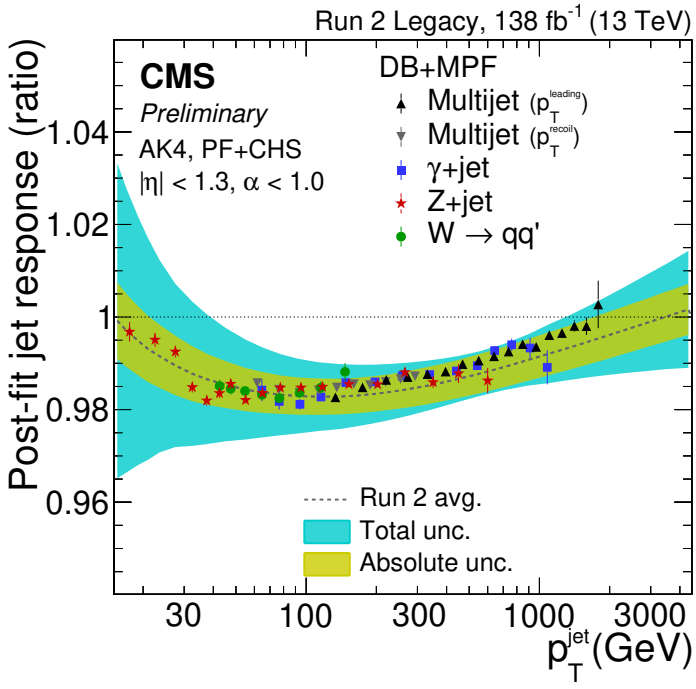

Figure 2.4: Absolute p_T-dependent residual Jet Energy Correction (JEC) derived from data-to-simulation comparisons in $Z+$ jet, $\gamma+$ jet, multijet, and $t\bar{t}$ events. The yellow band represents the systematic uncertainty associated with this correction.

Jet Energy Resolution (JER) Corrections

Even after calibration, the finite detector resolution causes a spread in the measured jet energies relative to their true values. To account for this effect, a Jet Energy Resolution (JER) correction is applied, ensuring that the resolution in simulated events matches that observed in data. This step improves the modeling of jet kinematics, particularly in analyses sensitive to the precise shape of distributions such as the dijet invariant mass or pseudorapidity separation.
The JER is measured in data and simulated samples using dijet events with the p_T-balance method. This approach exploits the momentum balance between two back-to-back jets in an ideal two-body topology, which is then extrapolated to realistic conditions including additional jet activity arising from initial- and final-state radiation (ISR/FSR). The method is applied using jets that have already been corrected with the full set of Jet Energy Scale (JES) calibrations to ensure that residual discrepancies are due solely to resolution effects.

From these measurements, data-to-simulation scale factors (SFs) are derived to bring the simulated resolution in agreement with the one observed in data. For the Run 2 Legacy reprocessing, these JER scale factors are determined as a function of the jet pseudorapidity (η_{jet}) for jets with transverse momentum $p_T \geq 100$ GeV. The resulting SFs are typically in the range 1.0-1.2, with slightly larger values in the ECAL-HCAL transition region ($|\eta_{\text{jet}}| \in [2.5, 3.0]$), where the detector response is less uniform.
Figure 2.5 shows the JER scale factors derived for Run 2 using the dijet p_T-balance method. The colored bands represent the associated statistical uncertainty on the measurement. These corrections are applied on a per-jet basis in simulated samples to smear their energies, ensuring that the reconstructed resolution matches that observed in data.

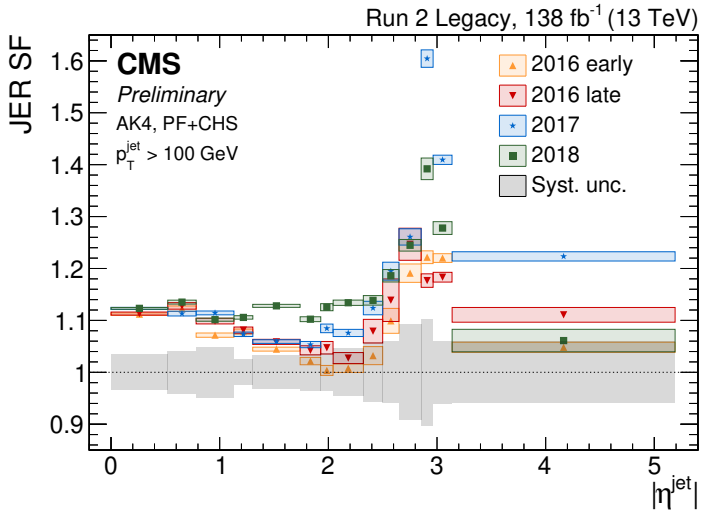

Figure 2.5: Jet Energy Resolution (JER) data-to-simulation scale factors for the Run 2 Legacy dataset, derived using the dijet p_T-balance method. The measurement is performed as a function of jet pseudorapidity for jets with $p_T \geq 100$ GeV. The colored bands represent the associated statistical uncertainty.

Jet Pileup Identification

Under LHC running conditions, additional soft proton-proton interactions often give rise to low-momentum jets unrelated to the primary hard-scatter process. These so-called pileup jets are suppressed using a dedicated multivariate discriminant, referred to as the jet pileup identification (PU ID). The method combines information on the jet's internal structure, track multiplicity, and vertex association to separate genuine hard-scatter jets from those originating from pileup activity.
For this analysis, the *loose* PU ID working point is employed. It maintains a high efficiency for signal jets while efficiently rejecting pileup contributions. At transverse momenta above 50 GeV, pileup effects become negligible, and no additional selection is applied. Scale factors extracted from data are used to correct the simulation, ensuring that the PU ID efficiency is consistently modeled between data and Monte Carlo samples. Together with the reconstruction and calibration steps described above, these procedures guarantee an accurate description of the jet energy scale and resolution, and a faithful modeling of the VBF dijet system that constitutes the core of the measurement.

2.5.5 B-Tagging

The identification of jets originating from bottom quarks, commonly known as b-tagging, is a key element of modern collider analyses. In this study, b-tagging is employed primarily to suppress backgrounds containing top quarks, whose decays almost always produce one or more b jets. It also serves to enhance the purity of the phase-space enriched by the VBF-Z production, where genuine b jets are expected less frequently in the final state.

Algorithm and Working Point B-tagging is performed using a deep neural network-based algorithm, which combines information from tracks, secondary vertices, and jet kinematic properties to distinguish jets initiated by b quarks from those originating from lighter partons or gluons. The algorithm outputs a continuous discriminator value, with higher scores corresponding to a higher probability that a jet contains a b hadron.
For this analysis, the medium working point of the discriminator is used. It provides a good balance between efficiency and purity, typically identifying about 70% of genuine b jets while misidentifying light-flavor jets at the per-mille level. Only jets with transverse momentum above 30 GeV and pseudorapidity within $|\eta| < 2.5$ are considered for b-tagging, corresponding to the region where the algorithm is well calibrated.

Calibration and Data-Simulation Corrections As with other reconstructed objects, the performance of the b-tagging algorithm can differ slightly between simulation and data due to imperfect modeling of detector response and track reconstruction. To correct for these effects, data-driven calibration factors are derived using control samples enriched in b jets from top-quark and multijet events. These scale factors are applied to simulated jets to reproduce the b-tagging efficiency observed in data.
Uncertainties associated with the calibration procedure—such as the limited size of control samples, modeling of fragmentation and hadronization, or differences in detector conditions—are propagated as systematic variations in the analysis. Separate uncertainties are assigned for b, c, and light-flavor jets to account for their distinct tagging efficiencies and modeling accuracy.

Impact on the Analysis The use of b-tagging serves two purposes in this measurement. First, events containing one or more b-tagged jets are vetoed to reduce backgrounds from top-quark production. Second, the events with b-tagged jets are used to estimate the normalization of the top-quark samples.

2.6 Statistical methods

The statistical interpretation of the results is performed using the Combine framework [96], the standard tool adopted within CMS for parameter estimation, hypothesis testing, and uncertainty propagation. The framework follows a likelihood-based approach in which the observed data are compared to the signal and background expectations, allowing a consistent treatment of both statistical and systematic uncertainties.

Likelihood modeling

The analysis employs a binned maximum-likelihood fit constructed from templates of the discriminating observables. The likelihood function $\mathcal{L}(\text{data}|\mu, \theta)$ depends on the signal

strength parameter μ, defined as the ratio of the measured to the Standard Model (SM) predicted cross section, and on a set of nuisance parameters θ representing systematic effects. The expected yield in each bin is modeled as the sum of signal and background contributions, with systematic variations implemented as shape or normalization changes depending on
theta. The best-fit values of μ and θ are obtained by maximizing the likelihood function, and the corresponding uncertainties are derived from the profile likelihood ratio using the asymptotic approximation of Ref. [97].

Treatment of nuisance parameters

Each nuisance parameter represents a specific source of uncertainty, encompassing detector effects, theoretical modeling, and background normalization. These parameters are constrained in the fit through prior probability distributions that encode their estimated uncertainties. Experimental effects are generally modeled with Gaussian or log-normal priors, while Poisson or gamma terms are used to describe the statistical uncertainties originating from the limited size of the simulated samples (see Section 2.6). Background normalization factors that are left free in the fit are implemented as unconstrained nuisance parameters. The fit procedure automatically propagates correlations among all nuisance parameters, ensuring that their combined impact is correctly reflected in the final uncertainty on the signal strength μ.

Barlow-Beeston treatment of Monte Carlo statistical uncertainties

The statistical uncertainty arising from the limited number of simulated events in each bin of the Monte Carlo (MC) templates is incorporated in the likelihood following the Barlow-Beeston method [98]. This approach introduces an additional Poisson term for every bin of each template, treating the MC prediction in that bin as a quantity with its own statistical uncertainty rather than as an exact value. In practice, each bin yield n_i is modeled as:

$$\mathcal{L}_i = \text{Poisson}(N_i^{\text{data}} \,|\, \mu S_i + \sum_b B_{ib}) \times \prod_b \text{Poisson}(B_{ib} \,|\, \hat{B}_{ib}),$$

where N_i^{data} is the observed number of events in bin i, μS_i is the signal contribution, and B_{ib} represents the background components estimated from simulation with nominal expectations $\hat{B}_{ib}$. This formulation allows the fit to account for statistical fluctuations in the MC templates, ensuring that bins with lower MC statistics contribute appropriately larger uncertainties.
In the COMBINE implementation, the so-called Barlow-Beeston lite option is used, in which a single effective nuisance parameter per bin encapsulates the MC statistical uncertainty of that bin. This approach provides an efficient and numerically stable approximation while preserving the essential treatment of finite MC statistics. It is particularly relevant for processes with limited simulated event counts, such as in the high-mass or large-$\Delta\eta_{jj}$ regions of the VBF phase space, where the MC statistical precision can have a non-negligible impact on the overall uncertainty of the measurement.

Asimov dataset

Before unblinding the analysis, all statistical studies are performed using a pseudo-dataset known as the Asimov dataset [97]. This dataset is constructed directly from the simulated Monte Carlo (MC) predictions and represents an idealized observation in which the number of events in each bin is exactly equal to the expected value from the model under consideration, typically corresponding to the Standard Model (SM) signal and background expectations.
The use of the Asimov dataset allows the full statistical machinery of the analysis to be validated—such as the fit model, nuisance-parameter treatment, and uncertainty propagation—without accessing the real collision data. It also provides a consistent way to compute the expected sensitivities, confidence intervals, and likelihood scans that would be obtained if the data exactly matched the model prediction.
In practical terms, the Asimov dataset is used as input to the combined likelihood fit implemented in the COMBINE framework. The fitted signal strength μ, the corresponding uncertainties, and the shapes of the likelihood scans shown in this thesis all refer to expected results, not measurements on observed data. This procedure ensures full statistical blindness in compliance with the CMS analysis policy: the real data are only examined after the final fit strategy, systematic treatment, and validation steps have been completely defined and frozen.
As a result, all the results and plots presented in this chapter—including likelihood scans, impact plots, and EFT parameter constraints—are based on the Asimov dataset and therefore represent the expected performance and sensitivity of the analysis prior to unblinding. No data is shown in this chapter where a high sensitivity is expected.

Likelihood scans

Likelihood scans are a central tool in the statistical interpretation of this analysis. They provide a direct visualization of how well different parameter values describe the observed data, and how uncertainties affect the precision of the measurement.

General procedure

In the profile-likelihood approach adopted by the CMS analysis framework, the estimator of a parameter of interest (such as the signal strength μ or a Wilson coefficient c_i) is obtained by maximizing the likelihood function $\mathcal{L}(\text{data} \mid \mu, \theta)$, where θ denotes the set of nuisance parameters representing systematic uncertainties. The likelihood ratio

$$-2\Delta \ln \mathcal{L} = -2 \ln \frac{\mathcal{L}(\text{data} \mid \text{parameter value})}{\mathcal{L}(\text{data} \mid \text{best fit})}$$

is used as the test statistic to quantify deviations from the best-fit point. Under the asymptotic approximation, a change of $+1$ in $-2\Delta \ln \mathcal{L}$ corresponds to a 68% confidence interval for one degree of freedom, and a change of $+2.3$ defines the 68% confidence contour in two dimensions.

One- and two-dimensional scans

One-dimensional scans are performed to extract the best-fit value of the signal strength μ and its uncertainty. In the Effective Field Theory (EFT) interpretation, two-dimensional

likelihood scans are produced for pairs of Wilson coefficients, allowing the study of correlations between operators and the determination of joint confidence regions. These scans often reveal parameter degeneracies and the extent to which different operators interfere in the VBF process.

Decomposition of uncertainties

To better understand the impact of different sources of uncertainty, multiple versions of the likelihood scans are typically produced:

- **Full scan:** All nuisance parameters, including statistical, experimental, and theoretical components, are profiled. This represents the total uncertainty in the measurement.
- **Freeze theory scan:** Theoretical uncertainties, such as those from renormalization and factorization scale variations or parton distribution functions, are fixed to their nominal values. The resulting curve isolates the impact of experimental and statistical uncertainties.
- **Statistical-only scan:** All nuisance parameters corresponding to systematic uncertainties are fixed to their best-fit values, leaving only statistical fluctuations. This scan highlights the intrinsic precision of the dataset independent of systematic effects.

The comparison among these scans quantifies the relative contribution of statistical, experimental, and theoretical sources to the total uncertainty. In the case of the signal strength modifier, for example, the statistical component often dominates, while in EFT fits the theoretical modeling uncertainties may play a more significant role due to acceptance and shape effects.

Example

Figure 2.6 shows an example of a likelihood scan for the fiducial signal strength modifier μ_{fid}. The best-fit value corresponds to the minimum of the curve, while the intersections with the horizontal line at $-2\Delta \ln \mathcal{L} = 1$ define the 1σ confidence interval. The decomposition into total, systematic, and statistical components illustrates the relative impact of different uncertainty sources on the measurement precision.

Impact plots

Impact plots provide a concise summary of how individual sources of uncertainty affect the main measurement outcome, typically the signal strength modifier μ. They are derived from the profile-likelihood fit and illustrate both the sensitivity of the result to each nuisance parameter and the post-fit behavior of these parameters. An example is shown in Figure 2.7.

Definition and interpretation

For each nuisance parameter in the fit, two quantities are shown:

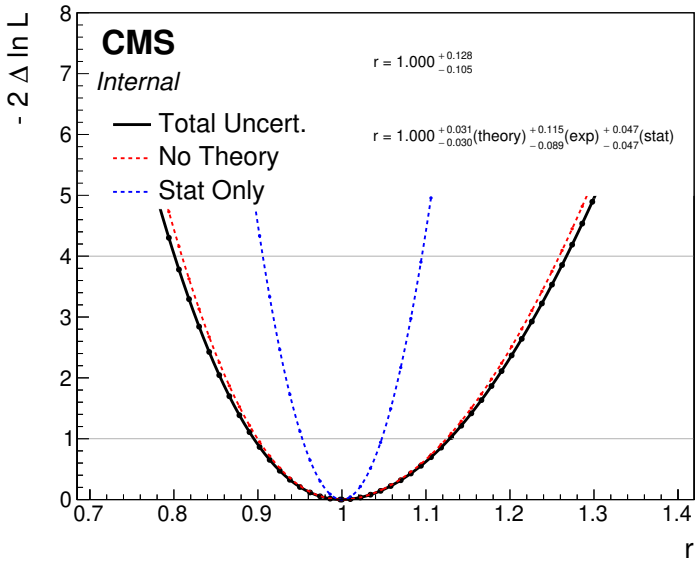

Figure 2.6: Likelihood scan for the fiducial signal strength modifier μ_{fid}. The total uncertainty is decomposed into statistical, experimental systematics, and theoretical components. The dashed lines illustrate scans where selected nuisance groups (e.g. theoretical or all systematics) are frozen to their best-fit values, isolating the corresponding uncertainty contributions.

- The **pull**, defined as the post-fit deviation of the nuisance parameter from its nominal (pre-fit) value in units of its pre-fit uncertainty. This indicates how much the fit has adjusted a given parameter to best describe the data and how much the prefit uncertainty are constrained by the data.
- The **impact**, defined as the change in the fitted value of μ when the nuisance parameter is shifted by $\pm 1\sigma$ around its best-fit value, with all other parameters profiled. This quantifies how sensitive the measurement is to that particular uncertainty source.

Each entry in the plot thus shows both the relative constraint imposed by the data (via the pull) and the effect of the uncertainty on the parameter of interest (via the impact). A well-behaved fit typically features pulls distributed around zero, with post-fit uncertainties narrower than the pre-fit ones for the most constrained parameters, reflecting information gained from the data.

Ranking and uncertainty decomposition

Nuisance parameters are ranked in decreasing order of their impact on the signal strength. The parameters with the largest impacts identify the dominant sources of systematic uncertainty in the measurement. In the context of this analysis, the leading contributions often originate from:

- **Jet-related uncertainties**, such as the jet energy scale and resolution, which directly affect the reconstruction of the VBF topology.
- **Background normalization and modeling**, particularly for the Drell-Yan and electroweak Zjj components.

- **Theoretical uncertainties**, including variations of renormalization and factorization scales or parton distribution functions, which influence both normalization and acceptance.

By identifying these key uncertainty sources, the impact plot guides future improvements in both the experimental calibration and theoretical modeling, showing where reductions in uncertainty would most effectively enhance the measurement precision.

Relation to the likelihood scans

While likelihood scans provide a global view of the uncertainty decomposition (statistical, systematic, and theoretical), impact plots offer a more granular perspective by quantifying the contribution of each individual nuisance parameter. Together, they form a complementary pair: likelihood scans describe the overall uncertainty envelope, whereas impact plots reveal its detailed composition.

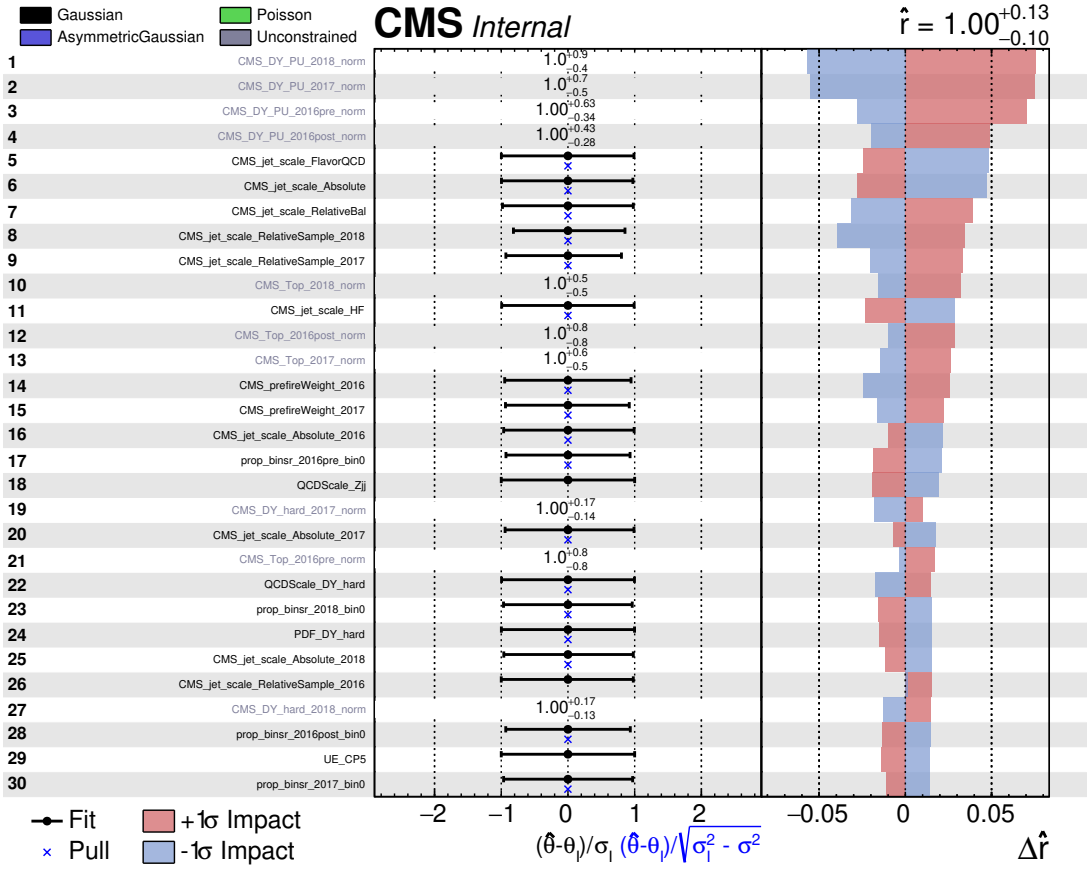

Figure 2.7: Pre-fit impact plot showing the ranking of nuisance parameters by their effect on the fitted signal strength modifier μ_{fid}. In the center panel, the horizontal bars represent the best-fit values and the corresponding uncertainties relative to the prior uncertainty. From this one can understand which systematics are constrained by the data. If no prior is available, e.g. simple normalizing parameters, only best-fit value and uncertainty are reported. In the right panel, the horizontal bars represent the variation in μ when each parameter is shifted by $\pm 1\sigma$. The most significant impacts correspond to DY PU normalization, and the jet energy scales

2.7 Analysis Selections

In this chapter, we describe the selection criteria applied in the analysis, as summarized in Table 2.5. The primary objective is to isolate vector boson fusion (VBF) events with Z boson production in a clean and efficient manner. The selection targets events with two oppositely charged leptons of the same flavor (electron or muon) and two energetic jets consistent with the VBF topology.
The baseline event selection includes the following requirements:

- **Lepton Pair**: Events must contain exactly two leptons (muons or electrons) with opposite charge and the same flavor, consistent with a $Z \to \ell^+\ell^-$ decay.
- **Leading Lepton** p_T: The leading lepton must have $p_T > 25$ GeV to ensure high trigger efficiency.
- **Subleading Lepton** p_T: The subleading lepton must have $p_T > 13$ GeV to maintain good acceptance while reducing background contamination.
- **Jet Selection**: Events must contain at least two jets with $p_T > 30$ GeV. These jets are used to reconstruct the VBF topology.
- **Dijet Invariant Mass** (m_{jj}): The invariant mass of the two leading jets must satisfy $m_{jj} > 200$ GeV, enhancing the contribution from genuine VBF processes.

A multivariate classifier (DNN) is used to further enhance the signal purity and is further discussed in Section 2.9. For all analysis regions—signal and control—the DNN output is required to be greater than 0.9, with the exception of the DY Hard control region. In this region, the same baseline selections as in the signal region are applied, but the DNN selection is inverted to define a sideband: $0.8 < \text{DNN} \leq 0.9$. This allows for a background-enriched region, with a purity of 70% with minimal signal contamination, useful for validating the modeling and constraining systematic uncertainties.
These selection criteria are optimized to maximize sensitivity to the VBF Z boson production process while suppressing backgrounds from Drell-Yan and other electroweak or QCD-induced processes.

Selection	**Description**
Lepton Pair	$e^{\pm}\,e^{\mp}$ or $\mu^{\pm}\,\mu^{\mp}$
High dilepton mass	$m_{ll} > 50$ GeV
Jet Selection	At least two jets with $p_T > 30$ GeV
Leading Lepton p_T	$p_T > 25$ GeV
Subleading Lepton p_T	$p_T > 13$ GeV
Dijet Invariant Mass	$m_{jj} > 200$ GeV

Table 2.5: Preselections common to all analysis regions

Selection	**Description**
Z mass peak	$\lvert m_{ll} - m_Z \rvert < 15$ GeV
Jet Selection	Two jets with $p_T > 50$ GeV
B veto	no b-tagged jet
High DNN	DNN > 0.9

Table 2.6: Signal region selections

2.7.1 Drell-Yan Control Regions: Hard and PU Components

The Drell-Yan (DY) process is one of the dominant backgrounds in this analysis and represents a significant source of contamination in the signal region. Due to its high production rate and broad kinematic coverage, a precise modeling of the DY background is essential for isolating the electroweak (EWK) Zjj signal and extracting reliable cross section measurements.
To improve the modeling of the DY background and enhance sensitivity to the VBF topology, the DY process is separated into two distinct components based on generator-level information:

- **DY Hard**: Events in which both of the two leading jets are matched to generator-level jets. These events are representative of genuine hard-scattering processes and are most relevant for the kinematic regions targeted by the VBF selection.
- **DY Pileup**: Events where at least one of the two leading jets fails to match a gen jet, indicating a likely origin from pileup or soft radiation. These events typically populate lower-purity regions and can distort the event topology.

To constrain the modeling of these two DY components, two dedicated control regions are defined in the analysis:

DY Hard Control Region: This region is designed to closely mimic the signal region in terms of kinematic selections but targets events in a lower-purity range of the DNN discriminant. Specifically, the same baseline selection as the signal region is applied, but the DNN requirement is inverted to select events with $0.8 < \text{DNN} \leq 0.9$. This region is enriched in DY Hard events while maintaining negligible signal contamination. It is used to validate the normalization and shape modeling of the DY Hard component and ensure sufficient flexibility in the fit.

DY PU Control Region: To isolate events dominated by pileup-affected jets, the DY PU control region is defined by requiring that the transverse momentum of either the leading or subleading jet is below 50 GeV. This selection enhances the fraction of DY PU events, which tend to populate the lower jet-p_T regime. This region plays a key role in constraining the normalization and modeling of the DY PU contribution.
By explicitly separating and controlling the two DY components, the analysis ensures a more accurate background estimation. The DY Hard and DY PU control regions provide orthogonal constraints that improve the stability of the fit and reduce the associated systematic uncertainties. This strategy is crucial for achieving high precision in the measurement of the electroweak Zjj signal.

Selection	**Description**
Z mass peak	$\lvert m_{ll} - m_Z \rvert < 15$ GeV
Jet Selection	Two jets with $p_T > 50$ GeV
B veto	no b-tagged jet
High DNN	$0.8 < \text{DNN} \leq 0.9$

Table 2.7: DY Hard control region selections

Selection	Description
Z mass peak	$\lvert m_{ll} - m_Z \rvert < 15$ GeV
Jet Selection	At least one jet with $p_T < 50$ GeV
B veto	no b-tagged jet
High DNN	DNN > 0.9

Table 2.8: DY PU control region selections

2.7.2 Top Control Region

The Top control region (Top CR) is designed to isolate a sample enriched in top quark backgrounds, primarily from $t\bar{t}$ and single top processes. This region is used both to validate the modeling of these backgrounds and to constrain their normalization in the statistical analysis.
To minimize extrapolation uncertainties between the Top CR and the signal region (SR), the selection is defined to be as close as possible to the SR. Specifically, all selections applied in the SR are retained, with the sole exception that events are required to have at least one of the final state jets is b-tagged and that the dilepton mass should be outside the Z mass peak in order to minimize the DY contamination. This requirement significantly suppresses the VBF Z boson signal, while enhancing top-enriched processes through the presence of real b-jets from top decays.
This strategy ensures that systematic uncertainties, particularly those affecting jet kinematics and modeling of top production, are less affected by extrapolation in the SR. As a result, corrections and systematic variations derived in this region are more reliably extrapolated to the SR.
The signal contamination in this region is found to be negligible.

Selection	Description
Outside Z mass peak	$\lvert m_{ll} - m_Z \rvert \geq 15$ GeV
Jet Selection	Two jets with $p_T > 50$ GeV
B tag	at least one b-tagged jet
High DNN	DNN > 0.9

Table 2.9: Top control region selections

2.7.3 Data-Monte Carlo Agreement in Control Regions

A crucial validation step in the analysis consists of assessing the agreement between data and Monte Carlo (MC) simulation for the most important input variables of the Deep Neural Network (DNN). This procedure is performed in the three analysis regions introduced previously: the Drell-Yan pileup control region (DY PU CR), the top-quark control region (Top CR), and the signal region (SR). To reduce potential sculpting effects from the DNN and to remain sensitive to phase space populated predominantly by background processes, this comparison is restricted to events with a DNN score below 0.9.
The nine variables shown in Figures 2.8, 2.9, and 2.10 correspond to the highest-ranked DNN inputs determined from the SHAP-based feature importance study (Section 2.9). These variables include the dijet pseudorapidity separation $\Delta\eta_{jj}$, the dijet invariant mass m_{jj}, the Z-boson Zeppenfeld variable $Z_{\ell\ell}$, the transverse momenta of the leading and subleading jets (p_T^{j1} and p_T^{j2}), the dijet azimuthal separation $\Delta\phi_{jj}$, the dilepton transverse

momentum $p_T^{\ell\ell}$, the dijet system transverse momentum p_T^{jj}, and the pseudorapidity of the leading lepton $\eta_{\ell 1}$.
Overall, good agreement is observed between data and simulation across all three regions, within the associated statistical and systematic uncertainties. Although a few ratio points are outside the uncertainty band, the data-driven normalization of the main backgrounds is not included and is able to recover the agreement. The shapes of the key kinematic variables, those with the highest discriminating power between electroweak Zjj production and the dominant Drell-Yan background, are well modelled in both control regions, indicating that the MC description of the background processes is reliable. In the signal region, where the electroweak contribution becomes more prominent, the level of agreement remains consistent, providing confidence that the simulation accurately describes the relevant phase space in which the fit is performed.

2.8 Optimization of the Analysis Strategy

The selection defining the signal region is based on the output of a deep neural network (DNN) classifier, trained to distinguish signal events from background. To maximize the sensitivity of the analysis, an optimization procedure was carried out to determine the optimal cut on the DNN score that defines the regions entering the statistical inference. The goal of the optimization was to identify the DNN selection that yields the best expected precision on the signal strength modifier, μ. To this end, we employed a simplified binned likelihood model, in which the expected number of signal and background events in each region is treated as fixed, and only three sources of uncertainty are considered:

- the statistical uncertainty on the observed data,
- the statistical uncertainty on the simulated background predictions (MC stats, propbins),
- the signal strength modifier r, treated as the only parameter of interest.

No additional nuisance parameters were included in this optimization step, allowing for a clean evaluation of the intrinsic statistical power of different selections.
The fitted variable had 6 bins in DNN starting from the point under study up to 1.0.
The likelihood was constructed for various DNN threshold values, and for each configuration, the expected uncertainty on r was computed using the Asimov dataset. The optimal working point was defined as the DNN cut that minimized this uncertainty, corresponding to the best expected precision in the measurement of the signal strength.
This optimization strategy ensures that the final selection maximizes the sensitivity of the analysis, while keeping the region definition straightforward and robust for the final statistical interpretation.
Given the results of the scan 2.11, the final point to cut on was set to 0.8. We use the region from 0.8 to 0.9 as DY hard CR and from 0.9 to 1.0 as SR.

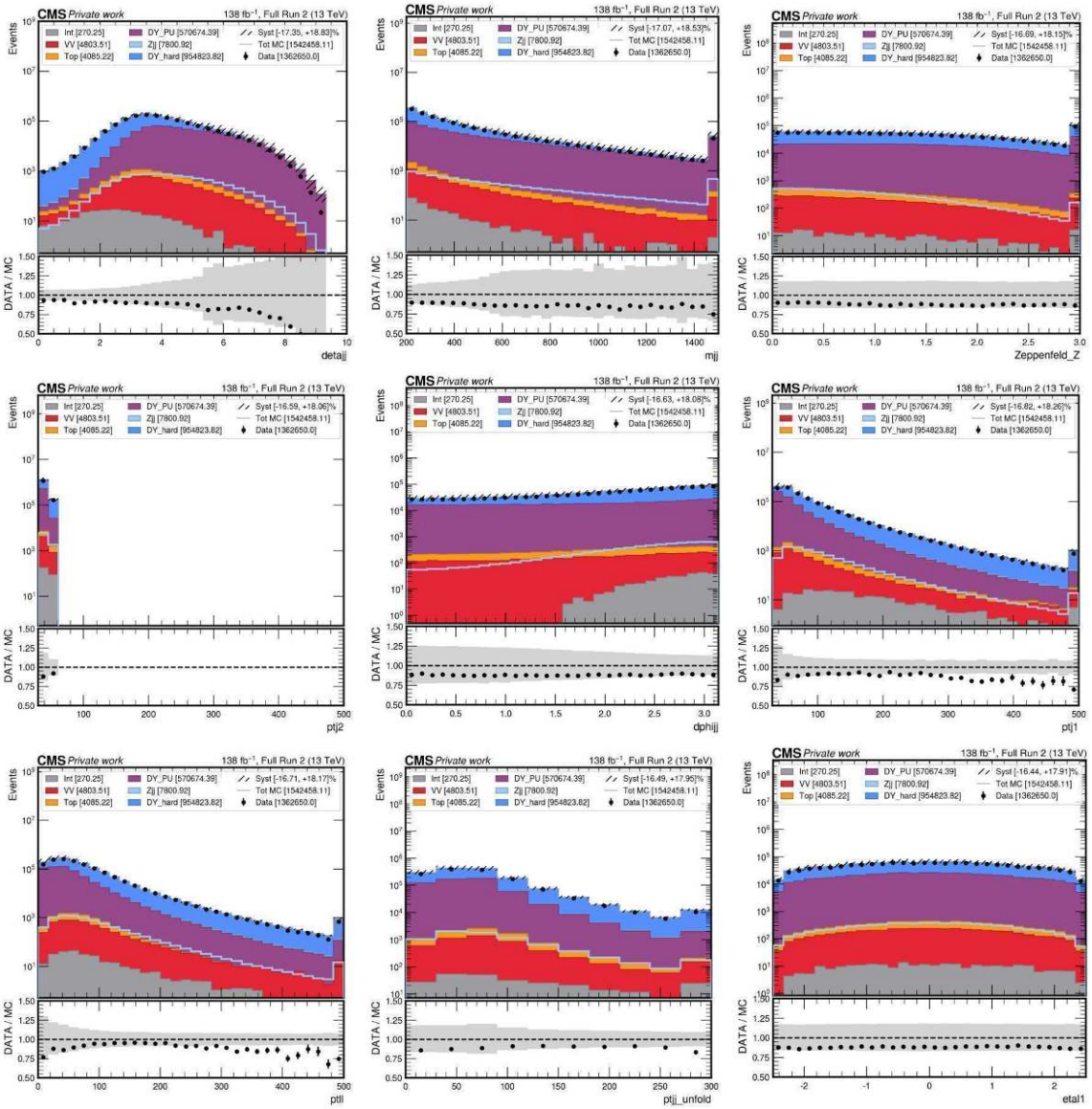

Figure 2.8: Input variables to the DNN in the DY PU CR with DNN < 0.9. From left to right, from top to bottom, ranked by importance in the DNN: $\Delta\eta_{jj}$, m_{jj}, Z_{ll}, p^T_{j2}, $\Delta\phi_{jj}$, p^T_{j1}, p^T_{ll}, p^T_{jj}, η_{l1}

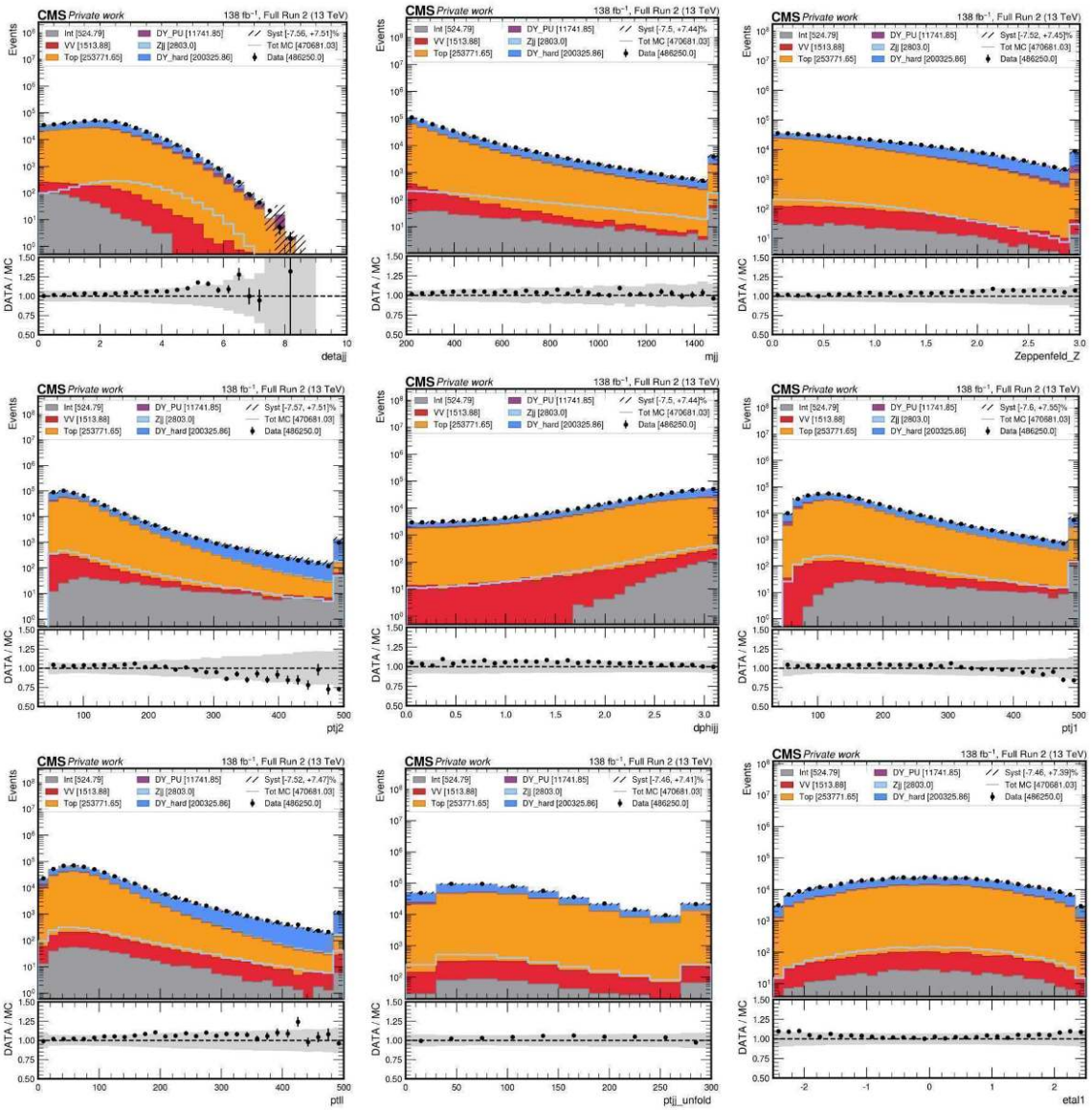

Figure 2.9: Input variables to the DNN in the Top CR with DNN < 0.9. From left to right, from top to bottom, ranked by importance in the DNN: $\Delta\eta_{jj}$, m_{jj}, Z_{ll}, p^T_{j2}, $\Delta\phi_{jj}$, p^T_{j1}, p^T_{ll}, p^T_{jj}, η_{l1}

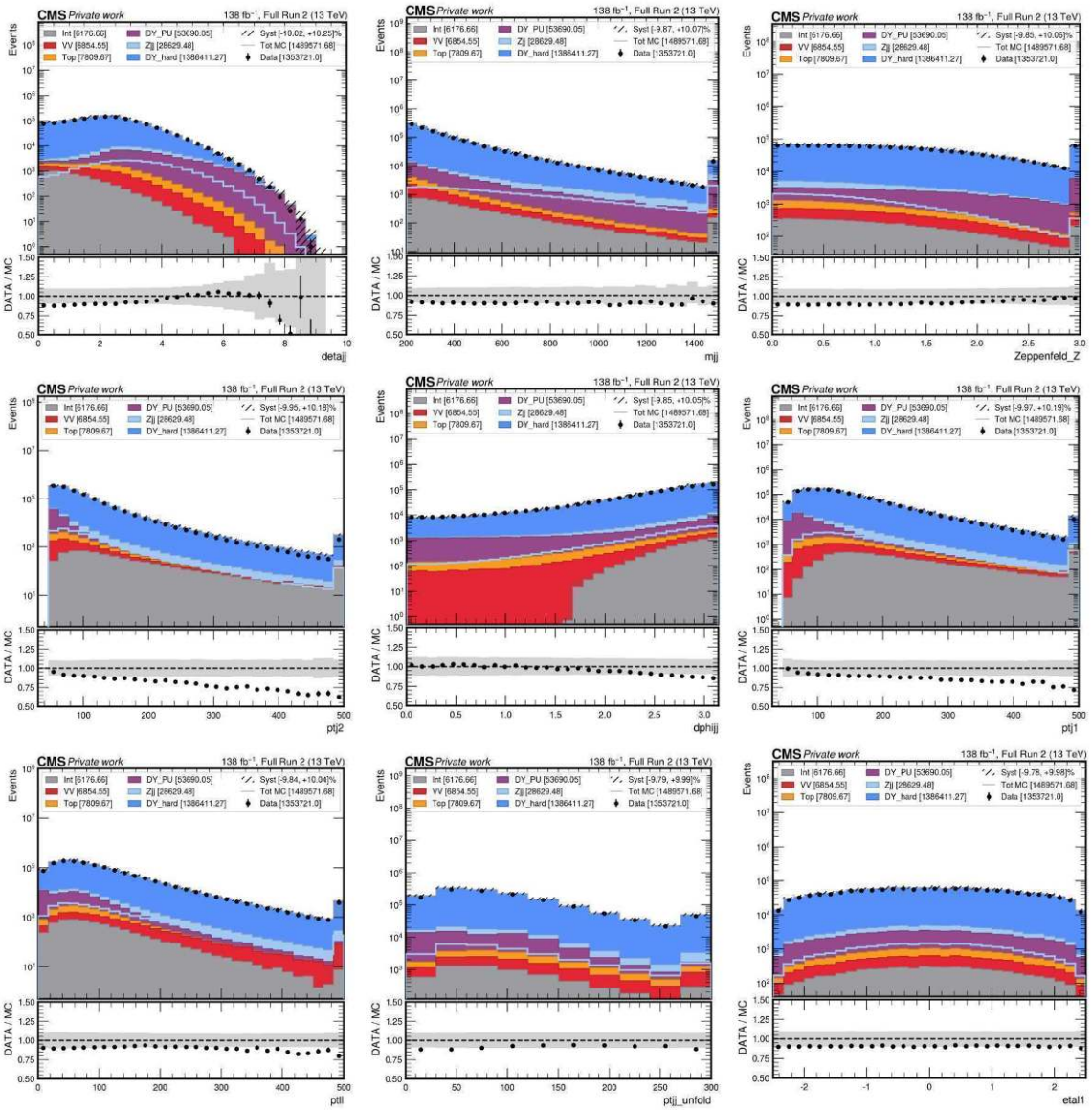

Figure 2.10: Input variables to the DNN in the SR with DNN < 0.9. From left to right, from top to bottom, ranked by importance in the DNN: $\Delta\eta_{jj}$, m_{jj}, Z_{ll}, p_{j2}^T, $\Delta\phi_{jj}$, p_{j1}^T, p_{ll}^T, p_{jj}^T, η_{l1}

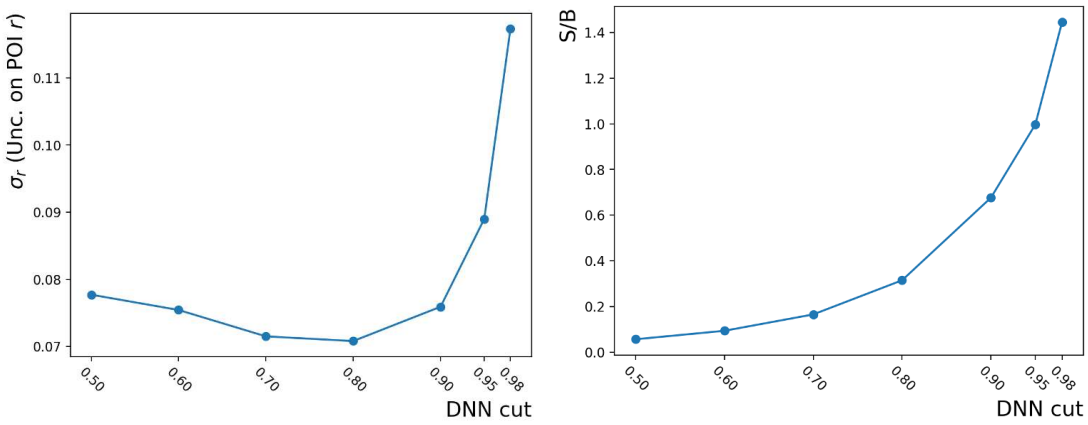

Figure 2.11: DNN cut optimization for the expected uncertainty on the signal strength modifier (left) and the signal over background ratio (right). The chosen cut is 0.8.

2.9 Signal vs background discrimination

To enhance the discrimination between electroweak Zjj (EW Zjj) signal events and the Drell-Yan (DY) background, a Deep Neural Network (DNN) is trained. The DNN leverages a set of kinematic and event-level variables to differentiate the EW Zjj process from the dominant background processes, improving the overall sensitivity of the analysis. The event classifier is implemented as a fully connected feed-forward neural network using the `Keras` for a binary classification task. The network consists of four hidden dense layers with ReLU activation functions and an output layer with a sigmoid activation. The hidden-layer topology is $N_{\text{in}} \to 32 \to 64 \to 128 \to 32$, followed by a single-node output layer. To reduce overfitting, an ℓ_2 regularization term is applied to the kernel weights of all hidden layers with coefficient $\lambda = 0.05$ (`kernel_regularizer=l2(0.05)`). In addition, dropout is used after the 128- and 32-neuron layers with a rate of 0.02. Batch normalization layers are enabled and applied after the first two dense layers and after the `Dense(128)+Dropout`, for a total of three batch-normalization layers. The model is trained by minimizing the binary cross-entropy loss (`BinaryCrossentropy`) with the Adam optimizer and a fixed learning rate of 10^{-3}. The training procedure includes early stopping based on the validation loss (`val_loss`) with a patience of 2 epochs; the best-performing weights are restored at the end of training, and an improvement threshold of `min_delta`=10^{-4} is required to continue training.

Class Imbalance Compensation The dataset used for training the DNN exhibits a class imbalance, as EW Zjj events (signal) are much less frequent than DY background events. To address this imbalance, event weights are applied during training, increasing the weight of signal events relative to background events. This approach ensures that the DNN learns to recognize the features of the EW Zjj signal more effectively, despite its lower impact in the dataset, since MC events are weighted with their cross-section including SFs.

Input Variables The DNN is trained using a set of discriminative variables, including:

- Single lepton and single jet variables:

  - Transverse momentum (p_T)
  - pseudorapidity (η)
  - azimuthal angle (ϕ).
- Dilepton and dijet system variables:
  - Transverse momentum ($p_T^{ll/jj}$)
  - pseudo-rapidity ($\Delta\eta$)
  - azimuthal angle ($\Delta\phi$).
  - Invariant mass ($m_{ll/jj}$)
- MET
- Zeppenfeld variable defined as:

$$Z_{ll} = \frac{1}{2}|(\eta^{l1} + \eta^{l2}) - (\eta^{j1} + \eta^{j2})| \tag{2.3}$$

The full list of variables is reported in table 2.10.

Variable	Description
MET	Missing Transverse Energy
Zeppenfeld$_Z$	Zeppenfeld Variable for the Z boson
$\Delta\eta_{jj}$	Pseudorapidity difference between the two leading jets
$\Delta\eta_{\ell\ell}$	Pseudorapidity difference between the two leptons
$\Delta\phi_{jj}$	Azimuthal angle difference between the two leading jets
$\Delta\phi_{\ell\ell}$	Azimuthal angle difference between the two leptons
η_{j_1}	Pseudorapidity of the leading jet
η_{j_2}	Pseudorapidity of the subleading jet
η_{ℓ_1}	Pseudorapidity of the leading lepton
η_{ℓ_2}	Pseudorapidity of the subleading lepton
m_{jj}	Invariant mass of the two leading jets
$m_{\ell\ell}$	Invariant mass of the two leptons
ϕ_{j_1}	Azimuthal angle of the leading jet
ϕ_{j_2}	Azimuthal angle of the subleading jet
ϕ_{ℓ_1}	Azimuthal angle of the leading lepton
ϕ_{ℓ_2}	Azimuthal angle of the subleading lepton
p_{T,j_1}	Transverse momentum of the leading jet
p_{T,j_2}	Transverse momentum of the subleading jet
$p_{T,jj}$	Transverse momentum of the dijet system
p_{T,ℓ_1}	Transverse momentum of the leading lepton
p_{T,ℓ_2}	Transverse momentum of the subleading lepton
$p_{T,\ell\ell}$	Transverse momentum of the dilepton system

Table 2.10: Input variables used for the DNN training.

The correlation of the above variables is reported in Figure 2.12 for a dataset including both signal and background. Two variables are added in the correlation matrix: the flavor ("ee" boolean variable) and the year. The correlation of the two variables with

the whole set is found negligible. The training is performed with simulations from both flavors (e^+e^- and $\mu^+\mu^-$) and from all the data-taking years. Tests were made training the model with the flavor and year variables but no effective improvement is seen. As cross-check the ROC is validated for the different flavors and years and no major differences are found. These two variables are not take into account in the final trained model.

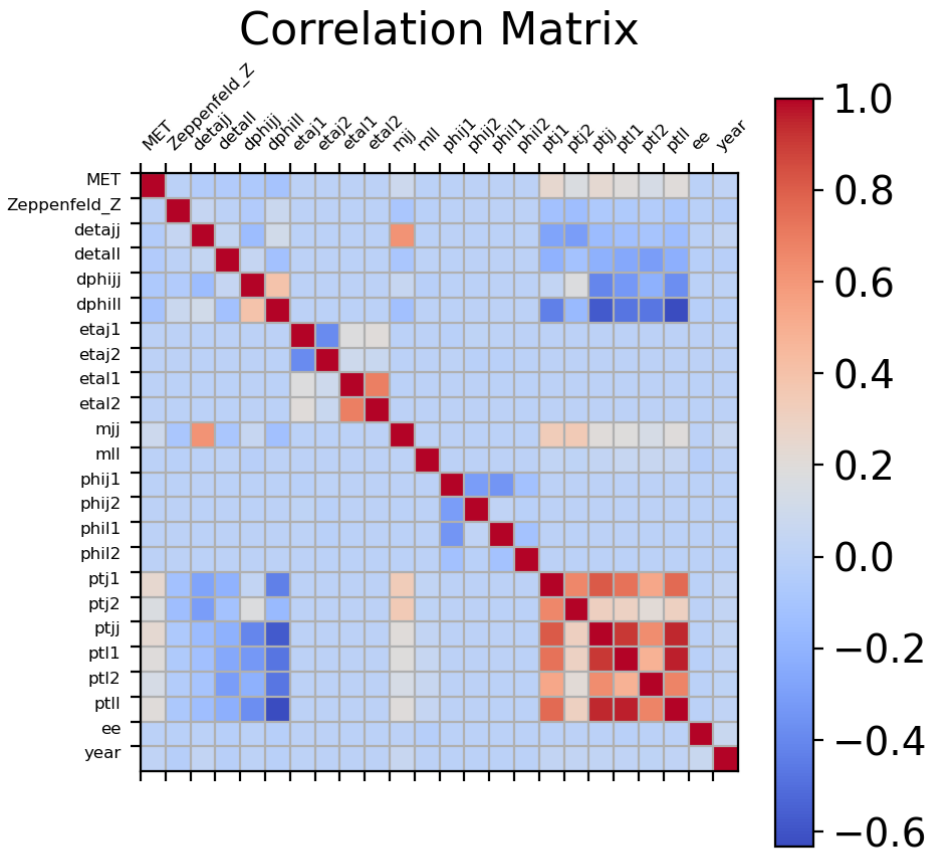

Figure 2.12: Correlation of input variables of the DNN for a dataset including both signal and background. Two variables are added in the correlation matrix: the flavor ("ee" boolean variable) and the year

The loss evaluated at the different epochs during the training, together with the final DNN output is ported in Figure 2.13.

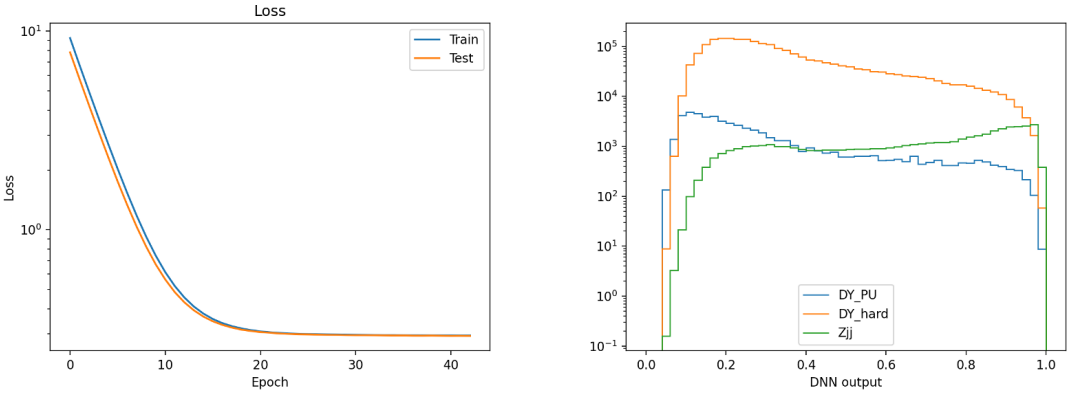

Figure 2.13: On the left: Training and testing loss at the different epochs. On the right: the final DNN output

Performance: ROC Curve The performance of the DNN is evaluated using the Receiver Operating Characteristic (ROC) curve, which plots the true positive rate (signal efficiency) against the false positive rate (background rejection). The ROC curve, shown in Figure 2.14, demonstrates the DNN's capability to distinguish between the EW Zjj signal and the DY background, with the area under the curve (AUC) indicating a high level of discriminative power.

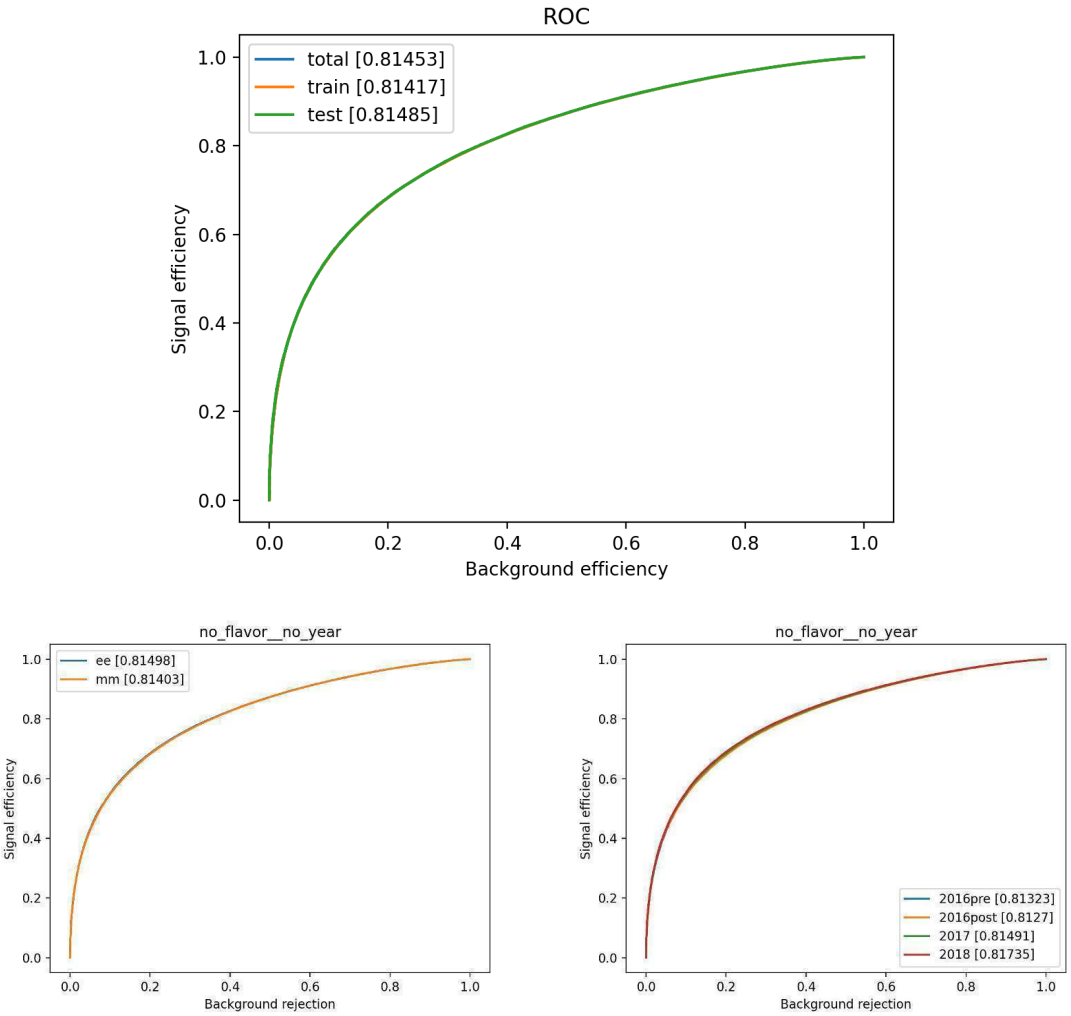

Figure 2.14: ROC curve for the DNN trained to separate EW Zjj signal events from the DY background. The area under the curve (AUC) is reported in square brackets for each input sample. On the bottom left the ROC is split for the different flavors and on the bottom right for the different data-taking years.

SHAP Analysis: Variable Ranking To understand the impact of each input variable on the DNN's decision-making process, SHapley Additive exPlanations (SHAP) [99] values are computed. SHAP is a game-theoretic approach that quantifies the contribution of each input feature to the model output by assigning to every variable a *Shapley value*, analogous to the concept of fair value distribution among players in a cooperative game. For each event, SHAP evaluates how the prediction changes when a variable is included or excluded from the model, averaged over all possible feature combinations. This results in an additive decomposition of the DNN output into feature contributions, providing a consistent and interpretable measure of variable importance.
The absolute mean of the SHAP values across the dataset is used to rank the input variables according to their overall influence on the model output. The ranking obtained from this analysis is shown in Figure 2.15. The most influential variables for distinguishing the EW Zjj signal from the DY background are found to be the dijet invariant mass (m_{jj}), the angular separation between the jets ($\Delta\eta_{jj}$), and the Zeppenfeld variable. These

results are consistent with the physical intuition of the VBF topology, where large m_{jj} and $\Delta\eta_{jj}$ values, as well as the centrality of the leptons encoded in the Zeppenfeld variable, characterize the electroweak signal relative to the Drell-Yan background. This is also confirmed by the comparison of the normalized shapes for the most important variables of the DNN reported in Figure 2.16 and the comparison of the correlation matrices shown in Figure 2.17.

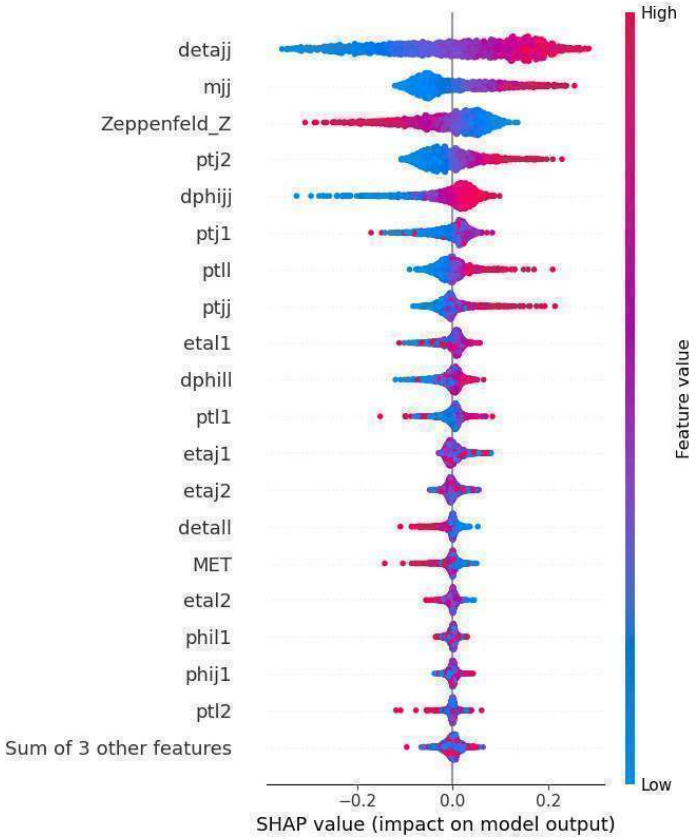

Figure 2.15: Ranking of input variables based on SHAP values, illustrating their impact on the DNN's output. Variables such as m_{jj}, $\Delta\eta_{jj}$, and Zeppenfeld Z boson are identified as the most significant for signal-background discrimination.

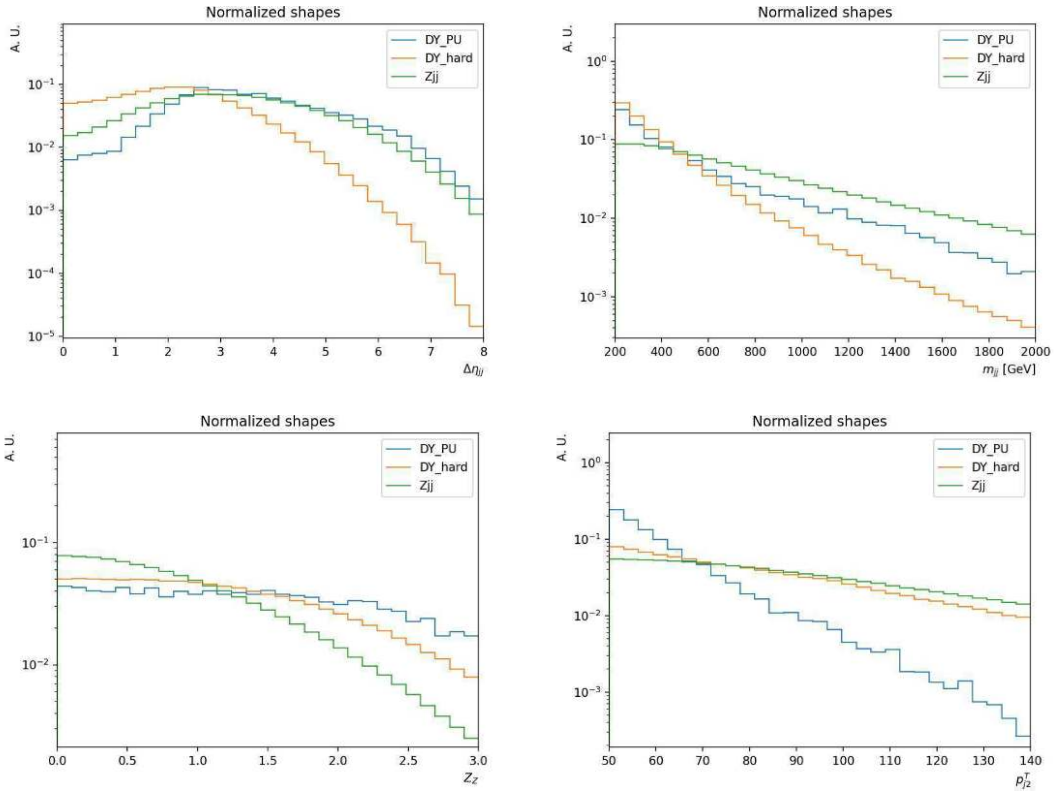

Figure 2.16: Normalized shapes comparison for the four most important variables of the DNN

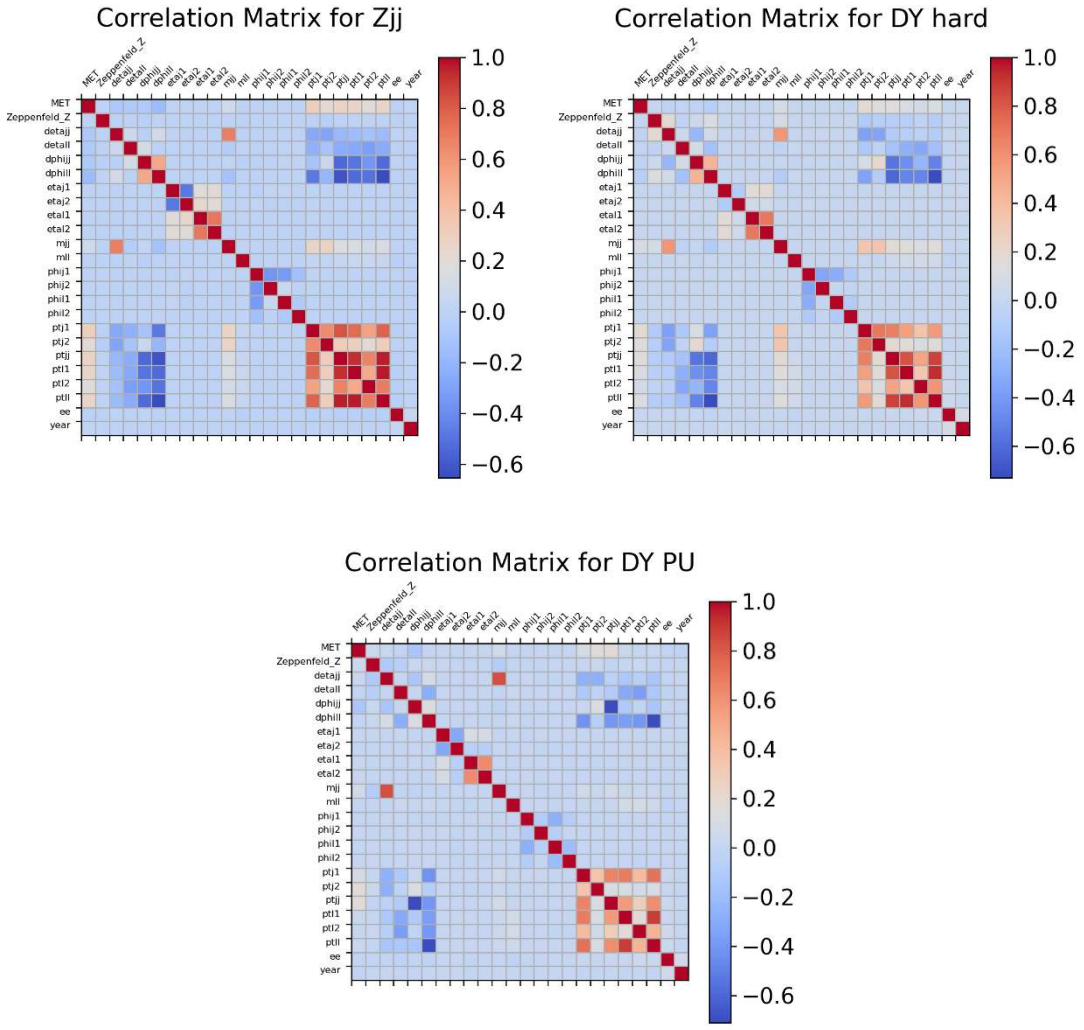

Figure 2.17: Correlation matrices comparison for the three samples: EW Zjj signal (top left), DY hard (top right) and the DY PU (bottom)

2.10 Systematic Uncertainties

Systematic uncertainties are crucial for understanding the robustness of the results and their impact on the measurement of the cross sections and the effective field theory (EFT) interpretations. The uncertainties are categorized into several groups, including theoretical uncertainties, detector-related uncertainties, and uncertainties related to event selection and calibration. A detailed description of each category is provided below.

Theoretical Uncertainties

The theoretical uncertainties arise from the modeling of the signal and background processes and are an important component of the overall uncertainty budget. The following theoretical uncertainties are considered for the signal and the DY:

- **QCD Scale Uncertainty**: This uncertainty is associated with the choice of renormalization and factorization scales in the quantum chromodynamics (QCD) calcu-

lations taking the combinations of $\mu_R = 0.5,\ 1,\ 2$ and $\mu_F = 0.5,\ 1,\ 2$. Variations in these scales can affect the predicted cross sections and event kinematics. We consider the envelope of the 6 variations.

- **PDF Uncertainty**: Uncertainties in the parton distribution functions (PDFs) are evaluated by using different variations of the same PDF set (NNPDF, with $\alpha_S = 0.118$ and LHAPDF id 325300). These uncertainties impact the modeling of the initial state partons and, consequently, the predicted cross sections. We take the envelope of the 100 variations.

- **Parton Shower Uncertainty**: The parton shower (PS) uncertainty accounts for differences in the modeling of the hadronization and parton shower evolution in the simulation. The PS reweights are provided by the parton shower used by each samples and are used to compute the Initial State Radiation (ISR) and Final State Radiation (FSR) components.

For the theoretical uncertainties a different treatment is adopted for signal and background based on this AN [100].

- Signal: only an acceptance effect is considered, renormalizing the variations to the effect before the selections of the analysis, and this depends on the type of measurement under study:

  1. For the inclusive measurement, the variations are normalized to the nominal case for all the events, i.e. no selections applied
  2. For the fiducial measurement, the variations are normalized to the nominal case for the events falling into the fiducial region
  3. For the differential measurement, the variations are normalized to the nominal case for each fiducial bin

- Background (DY): only a shape effect is considered, normalizing the variations to the nominal passing the reco selections of the analysis

Background normalization

To minimize the dependence on theoretical cross-section predictions, the overall normalization of the DY hard, DY PU and Top backgrounds is determined directly from the data. The normalization factors for the DY and Top processes are treated as free parameters in the fit, allowing the data in the signal and mainly in control regions to constrain their values.
The Drell-Yan background is separated into two components according to the jet multiplicity in the particle level events: a hard component, corresponding to $Z + 2$ jet events, and a pileup-enriched component, dominated by $Z + 1$ jet configurations. Independent normalization parameters are assigned to these two components, reflecting their distinct theoretical modeling and kinematic behavior. The top-quark background, which mainly arises from $t\bar{t}$ and single-top production, is treated with a single floating normalization parameter.
This approach effectively factorizes the uncertainty related to the overall cross-section normalization from the theoretical uncertainties that affect the shape of key observables.

Once the normalizations are allowed to float freely, the residual theoretical uncertainties—such as those from renormalization and factorization scale variations, parton shower modeling, or parton distribution functions—are applied only as *shape* uncertainties with no normalizing effect.

Charged Lepton Systematics

The reconstruction, identification, and efficiency of leptons (electrons and muons) are subject to several systematic uncertainties and are all de-correlated between years. These include:

- **Lepton Reconstruction , Isolation and Identification**: The efficiency of reconstructing, isolating and identifying leptons is varied to account for differences between data and simulation. Working point systematic uncertainties are also included to reflect variations in the criteria used for lepton Id. and Iso.

Jet Energy Scale and Resolution

Jets are corrected using jet energy scales (JES) and jet energy resolution (JER) corrections, both of which have associated uncertainties:

- **Jet Energy Scale (JES) Uncertainty**: Variations in the jet energy scale are considered, reflecting uncertainties in the calibration of jet energy measurements. We use the regrouped version of JES uncertainties as described in [101].
- **Jet Energy Resolution (JER) Uncertainty**: The JER uncertainty is evaluated by varying the resolution applied to jets in simulation to match the resolution observed in data. This uncertainty is particularly relevant for jets with pseudorapidity (η) less than 2.5, where corrections are applied. The JER are not correlated between years and are split in 6 different variations as suggested by the JET/MET group in bins of pt and eta.

Trigger Scale Factor Uncertainty

Trigger scale factors are used to account for the efficiency of the triggers applied in the analysis. The uncertainties associated with these scale factors are propagated as systematic uncertainties, affecting the normalization of the signal and background events. Trigger uncertainties are not correlated between years.

b-Tagging Scale Factor Uncertainty

The identification of jets originating from heavy-flavor quarks is an important handle for background suppression, particularly for processes involving top quarks. To account for small differences in the b-tagging performance between data and simulation, correction factors are applied to simulated events. The uncertainties associated with these corrections are propagated as systematic variations in the fit.
Separate uncertainty components are defined for jets originating from light-flavor quarks or gluons, from heavy-flavor (b) quarks, and from charm (c) quarks. The latter are represented by dedicated variations that capture the limited knowledge of the charm-tagging efficiency. All heavy-flavor-related tagging uncertainties are treated as correlated across data-taking years, reflecting their common experimental and modeling origin.

Pileup (PU) Variations

The modeling of pileup (PU) events, which are additional interactions occurring in the same bunch crossing, introduces systematic uncertainties. PU variations are included by changing the number of pileup interactions within the range allowed by data. The PU variations are not correlated between years.

PU Scale Factor for jet Identification

Uncertainties of the scale factor for PU Identification of jets are only applied to jets that pass the loose PU jet ID, that have a $15 < p_T < 50$ GeV. These uncertainties are uncorrelated between years.

Limited events in Monte Carlo samples

The statistical uncertainty associated with the finite size of the simulated Monte Carlo (MC) samples is explicitly taken into account in the fit. This uncertainty arises from the limited number of generated events in each process and bin of the analysis templates, which can lead to fluctuations in the predicted yields. To correctly propagate these effects to the final result, the so-called Barlow-Beeston approach [98] is employed (see Section 2.6). In this method, each bin of the MC template is assigned an independent uncertainty corresponding to the finite number of effective events contributing to it, allowing the likelihood fit to account for the statistical precision of the simulation itself. This treatment ensures a consistent propagation of MC statistical uncertainties to the fitted signal strength and nuisance parameters, without artificially inflating systematic effects or introducing biases from low-statistic bins.

Effect of uncertainties on key variables

The effect of the different uncertainties on the variables that will be unfolded is checked and reported for the DY Hard and for the signal in Figures 2.18-2.19. As a general note the signal uncertainties are comparable to the data statistical uncertainties. For the DY Hard the JES, the theory and limited MC statistical uncertainties have a larger impact compared to data statistics uncertainty.
The binning chosen for each variable is the same as the one chosen for the final unfolding fit discussed in section 2.13.

2.11 Inclusive Signal Strength Measurement

The goal of this measurement is to quantify the overall normalization of the VBF-Z signal, independently of its differential behavior. This signal strength modifier μ scales the expected signal yield and incorporates both experimental and theoretical uncertainties.
The fit is performed on a single-bin observable in each of the three control regions (Top CR, DY Hard CR, and DY PU CR) and also in the signal region. This approach simplifies the likelihood model and allows a robust constraint of the dominant backgrounds while focusing on the overall signal normalization.
Theoretical uncertainties associated with the signal are treated as nuisance parameters in the fit, allowing their effect to be included in the total uncertainty on μ. For this reason the theoretical uncertainties are normalized to the inclusive cross-sections.

Figure 2.20 shows the pre-fit distributions in the four analysis regions, before any fit to data is performed.

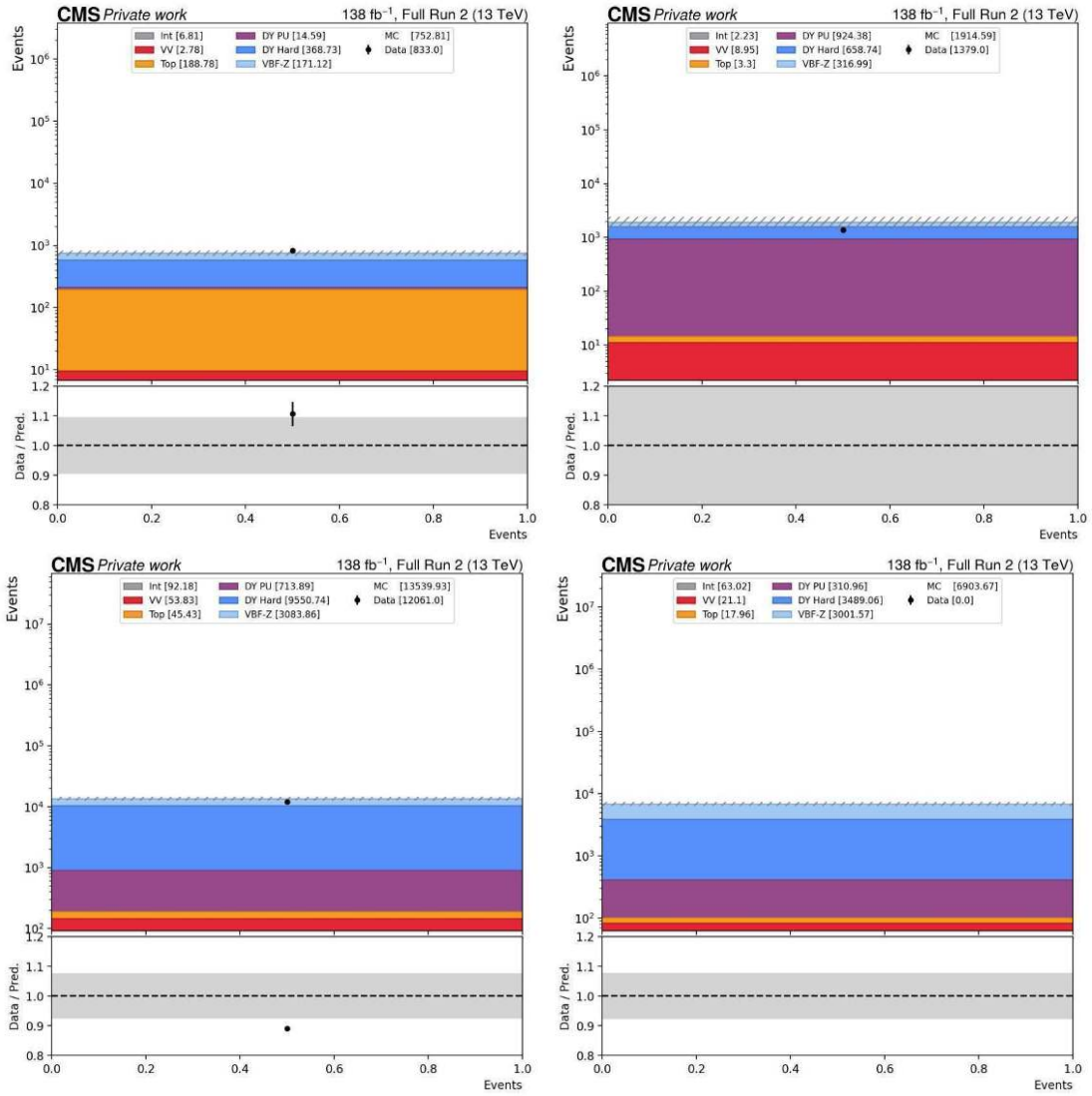

Figure 2.20: Pre-fit yields in the four analysis regions used in the inclusive fit: Top CR, DY PU CR, DY Hard CR, and SR (blind).

To better understand what's the impact of the different types of systematics on the measurement of μ, we check the likelihood scan and the impact plots (see Section 2.6). The likelihood scan for μ is shown in Figure 2.21. The scan also reports the decomposition of the total uncertainty into its statistical, systematic, and theoretical components, obtained using the standard CMS breakdown procedure, where in the "Stat Only" component all the non-constrained nuisances enter (i.e. the freely floating normalizations).

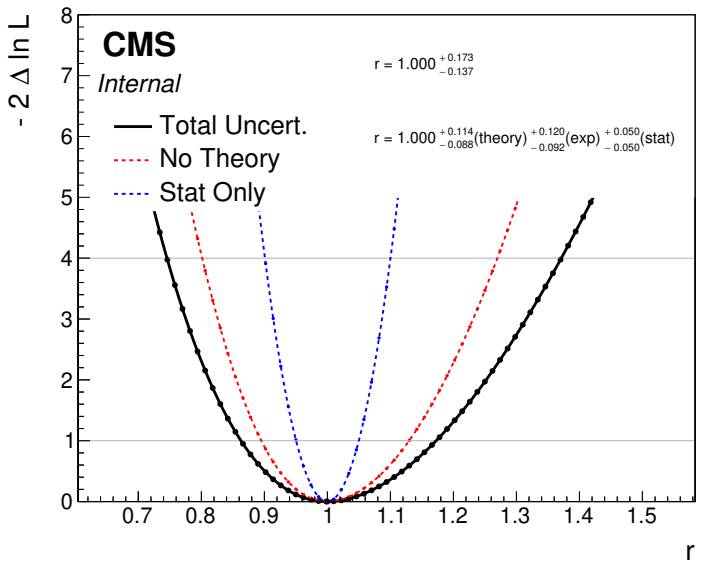

Figure 2.21: Likelihood scan for the inclusive signal strength μ. The total uncertainty is decomposed into statistical, experimental systematics, and theoretical components.

The ranking of nuisance parameters by their impact on the signal strength modifier μ is summarized in Figure 2.22. This ranking provides insight into the dominant sources of systematic uncertainty affecting the inclusive measurement. The leading uncertainties are primarily associated with the signal theoretical uncertanties, the jet energy scale (JES) and normalization factors of major background processes. In particular, variations related to the jet energy scale — such as the absolute calibration, flavor dependence, and year-specific relative corrections — have the largest effect on the fitted signal strength, reflecting the strong dependence of the VBF-like topology on the dijet kinematics. Other significant contributors include the normalization uncertainties for the DY PU and top-quark backgrounds, especially in the 2016 and 2018 datasets, as well as theoretical modeling uncertainties like the QCD scale variations and PDF uncertainties for the DY process. Experimental effects such as the pileup reweighting and the prefire correction also show non-negligible impacts, though their influence remains subdominant.
Overall, the fit is well constrained, with most nuisance parameters exhibiting pulls consistent with their pre-fit expectations and limited post-fit variations, indicating that the data statistical uncertainty is actually the dominant one.
The total uncertainty in the inclusive signal strength is dominated by experimental and theoretical systematic effects, with Monte Carlo statistical contributions (see Section 2.10) playing a smaller role.

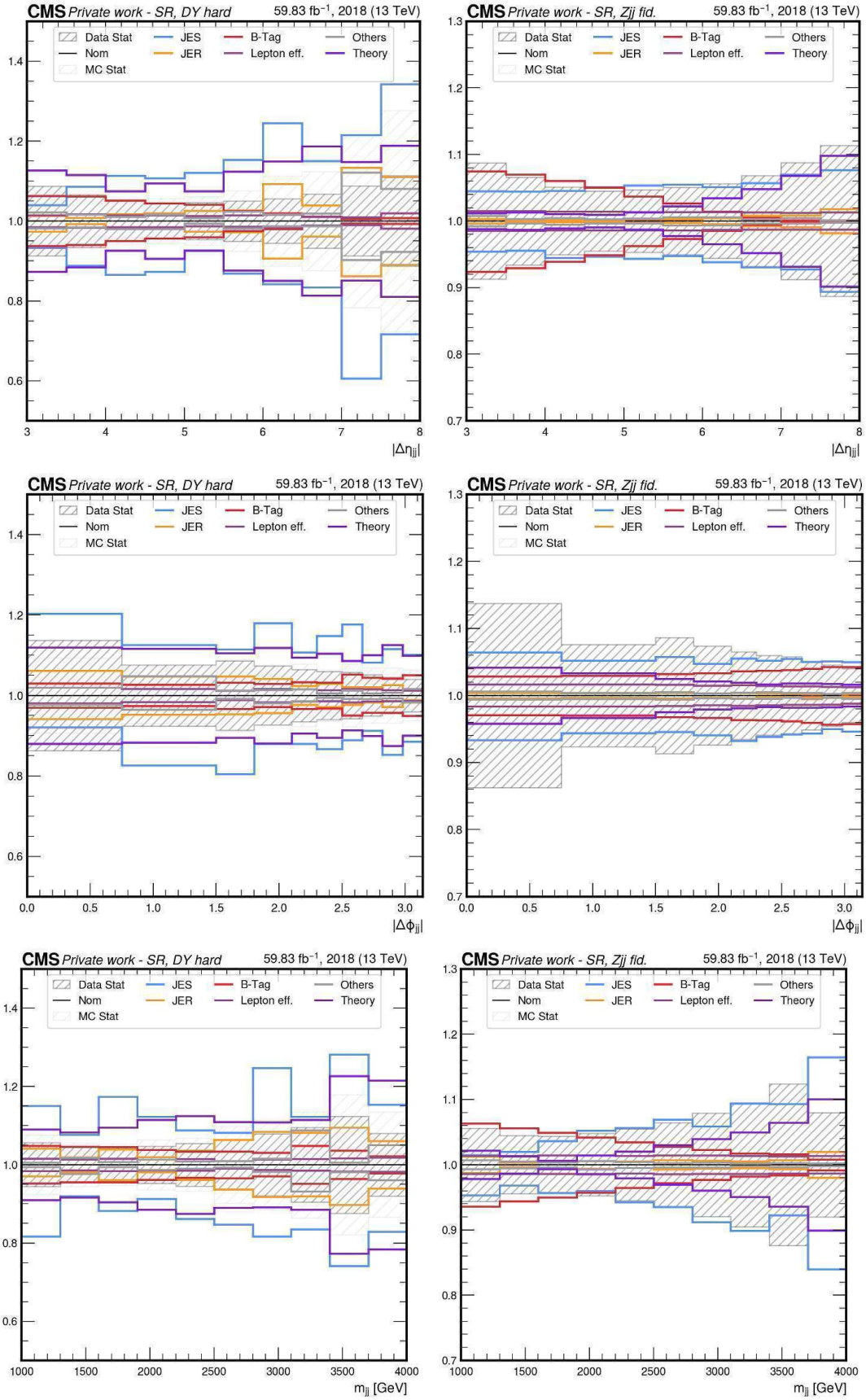

Figure 2.18: Main systematics, SR $\Delta\phi_{jj}$. Left DY hard, right Zjj EW.

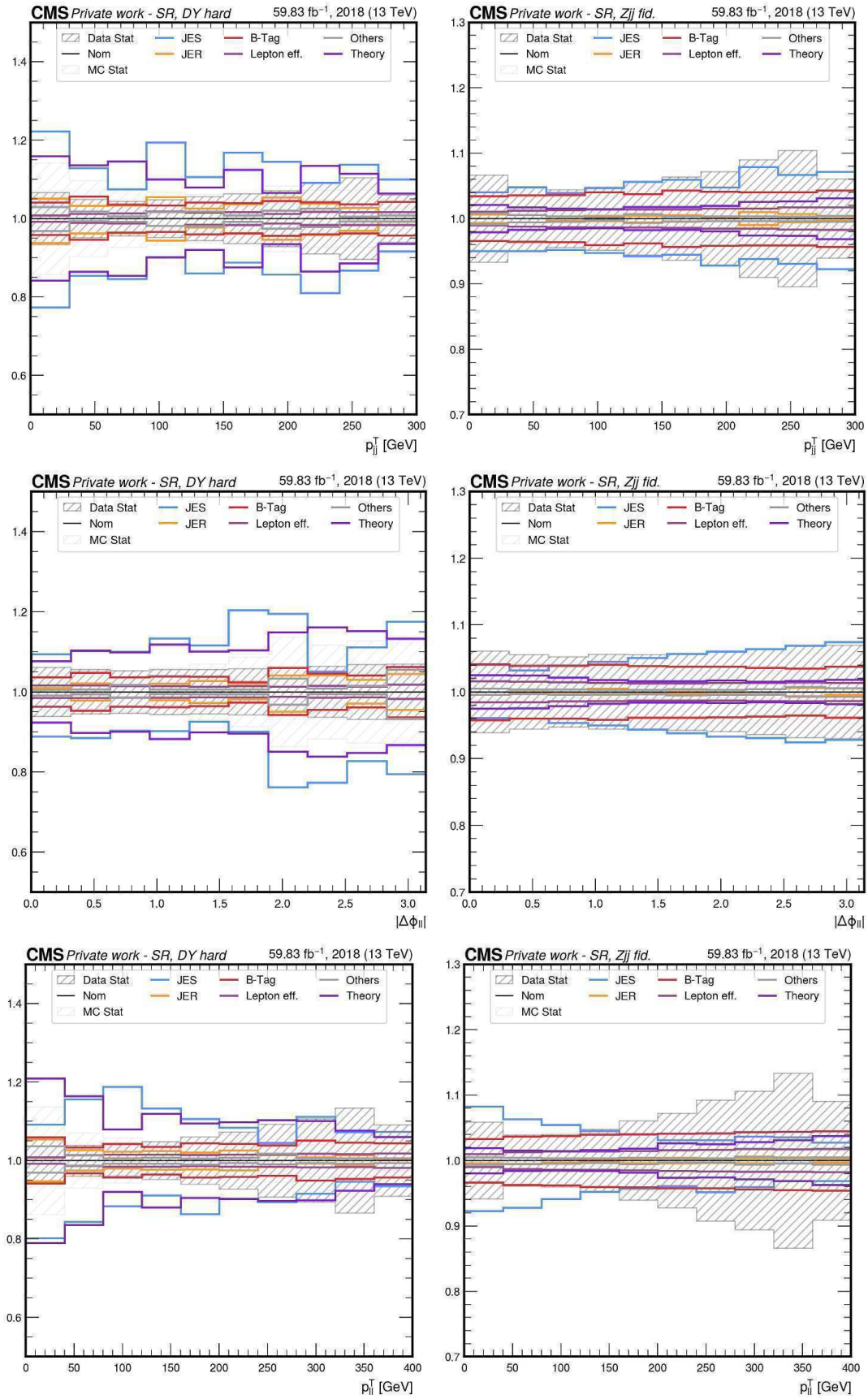

Figure 2.19: Main systematics, SR p_{jj}^{T}. Left DY hard, right Zjj EW.

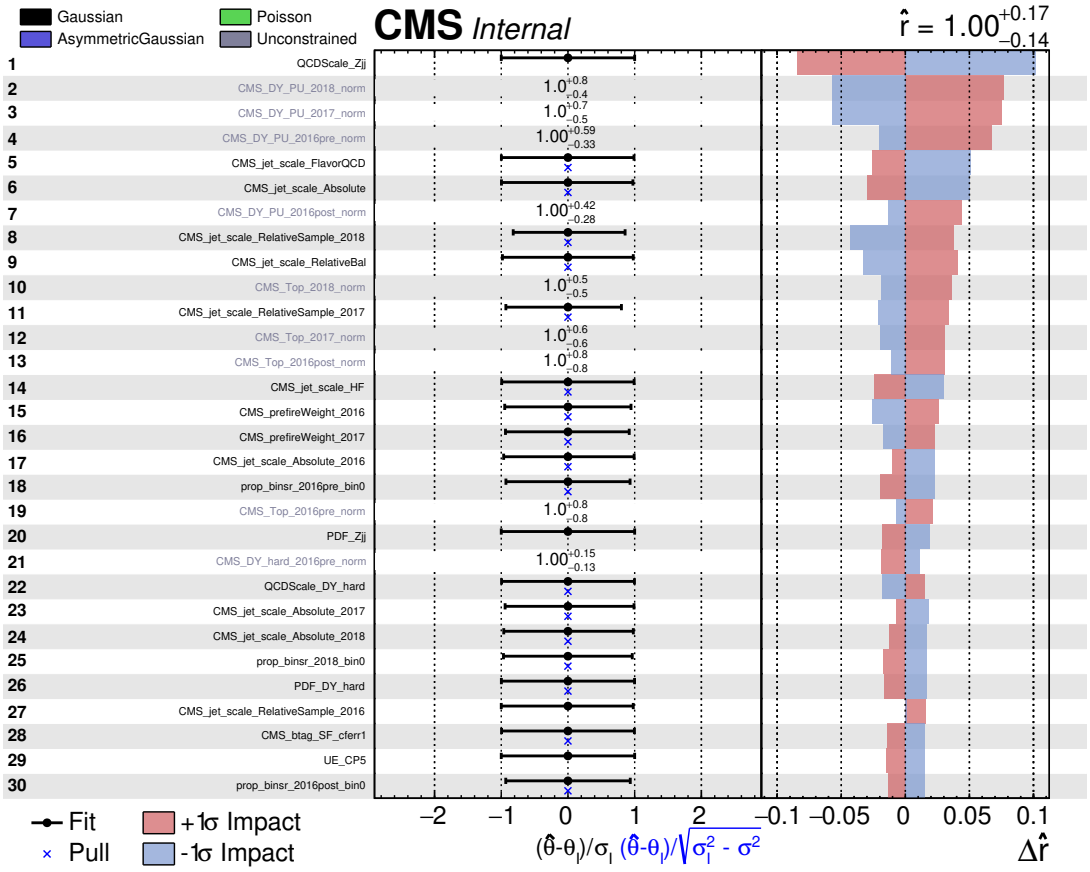

Figure 2.22: Ranking of nuisance parameters according to their impact on the inclusive signal strength μ.

2.12 Fiducial Cross Section Measurement

The fiducial cross-section modifier, μ_{fid}, is extracted via a simultaneous fit to the three control regions and the signal region, using a single-bin observable in each. In contrast to the inclusive measurement, μ_{fid} is defined within a fiducial phase space at generator level that matches as closely as possible the reconstructed selections applied in the signal region. This strategy allows a more direct comparison with theoretical predictions limited to the experimentally accessible phase space.
Theoretical uncertainties on the signal are included only as acceptance uncertainties. In particular, variations due to PDFs and QCD scales are normalized to the fiducial nominal cross-section ensuring that only migrations in and out of the fiducial volume contribute to the uncertainty.
Table 2.11 summarizes the generator-level selections used to define the fiducial region. These requirements mirror the ones used at reconstruction level in the signal region, allowing for a consistent unfolding and interpretation. The overlap between the fiducial region and the analysis reconstructed phase spaces is approximately 45%. Extrapolation effects due to detector acceptance and resolution are corrected for via the signal acceptance, defined as the ratio of events passing both the fiducial and reconstructed selections to those passing only the fiducial requirements. The events that fail the fiducial selection but pass the reconstructed-level criteria are treated as background in the fit. The DNN selection (DNN > 0.9) is mimicked at generator level with the selections on the dijet mass ($m_{jj} > 1000\,\text{GeV}$) and on the dijet pseudo-rapidity separation ($\Delta\eta_{jj} > 3$).
The fiducial LO cross-section is:

$$\sigma_{\text{fid}} = 102.31 \pm 7.38\,\text{fb} = 102.31 \pm 1.6\,(\text{PDF}) \pm 7.2\,(\text{QCD}) \pm 0.56\,(\text{PS})\,\text{fb} \qquad (2.4)$$

Quantity	Requirement
Leptons	e^+e^-or $\mu^+\mu^-$
Lepton p_T	$p_T^{l1} > 25$ GeV, $p_T^{l2} > 10$ GeV
Lepton $\lvert\eta\rvert$	$\lvert\eta\rvert < 2.4$ for both leptons
Dilepton mass	$\lvert m_{ll} - m_Z\rvert < 15$ GeV
Jets	At least two jets with $p_T > 50$ GeV and $\lvert\eta\rvert < 4.7$
Dijet invariant mass	$m_{jj} > 1000$ GeV
Dijet pseudo-rapidity separation	$\Delta\eta_{jj} > 3$

Table 2.11: Definition of the fiducial phase space at generator level. These selections are designed to reproduce as closely as possible the reconstructed-level requirements in the signal region.

Figure 2.23 shows the pre-fit distributions in the control and signal regions.
The likelihood scan for the fiducial signal strength modifier μ_{fid} is shown in Figure 2.24, including the breakdown of the total uncertainty into statistical, systematic, and theoretical (acceptance) components according to Section 2.6.
Finally, the impact of individual nuisance parameters on μ_{fid} is reported in Figure 2.25, highlighting the dominant sources of uncertainty in the fiducial cross-section extraction. Compared to the inclusive measurement, the fiducial fit shows a relatively smaller impact from theoretical acceptance uncertainties as expected given the treatment of these uncertainties for the fiducial measurement.

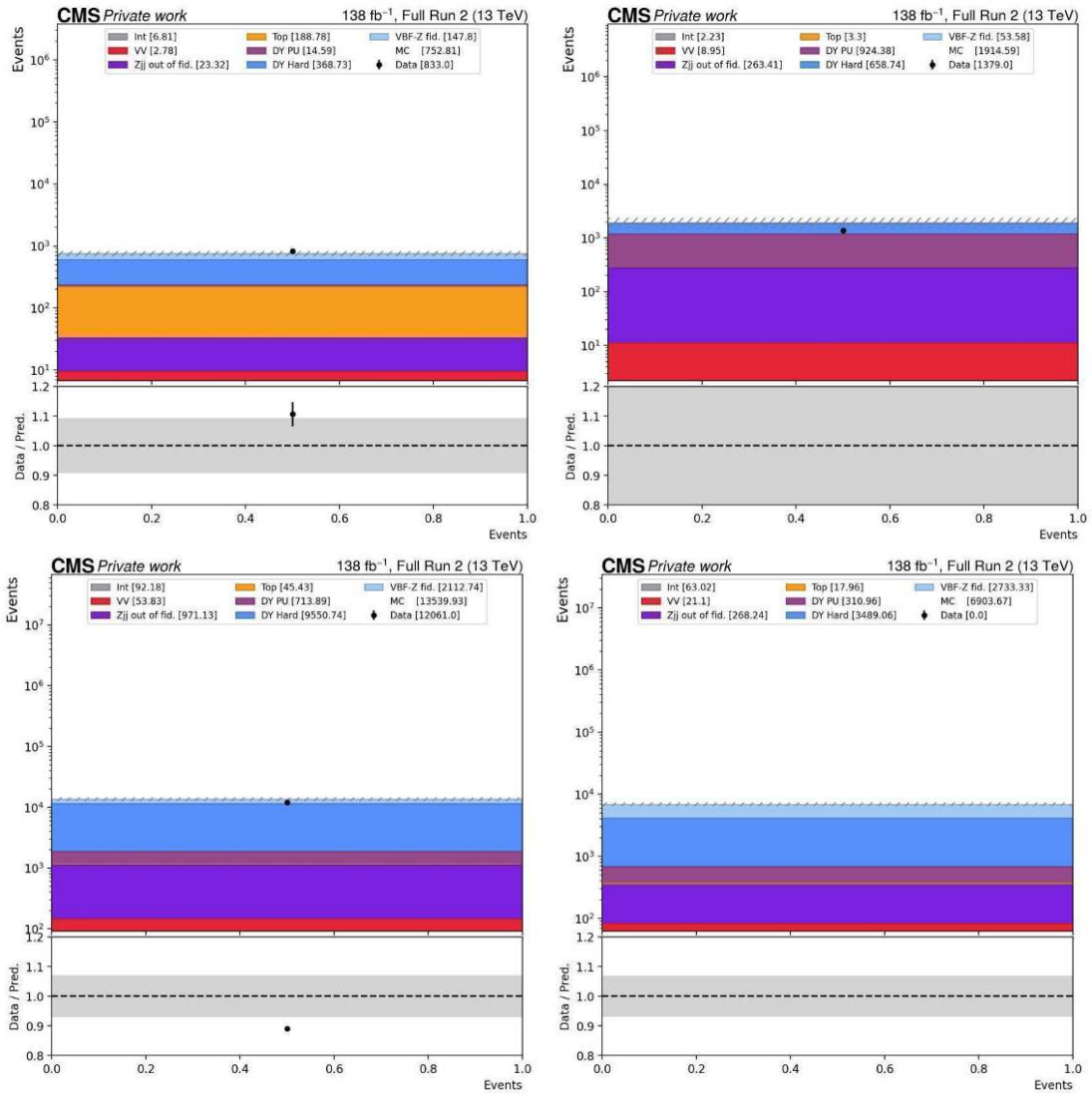

Figure 2.23: Pre-fit yields in the four analysis regions used in the fiducial cross-section fit: Top CR, DY PU CR, DY Hard CR, and SR (blind).

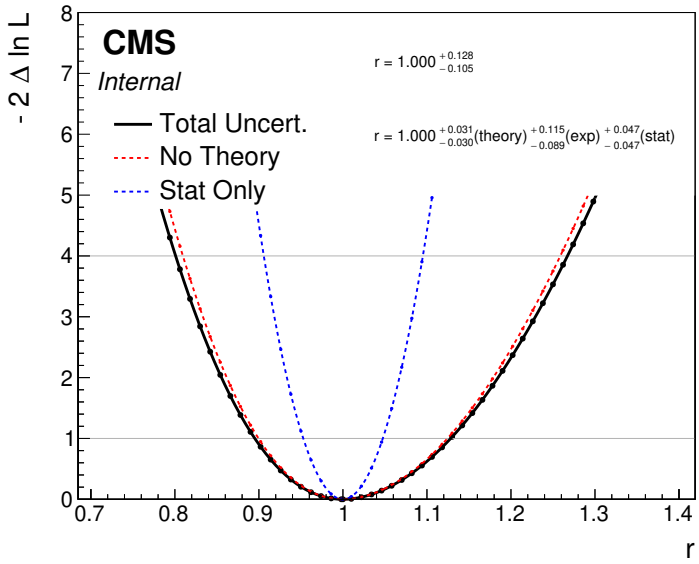

Figure 2.24: Likelihood scan for the fiducial signal strength modifier μ_{fid}. The total uncertainty is decomposed into statistical, experimental systematic, and theoretical (acceptance) components.

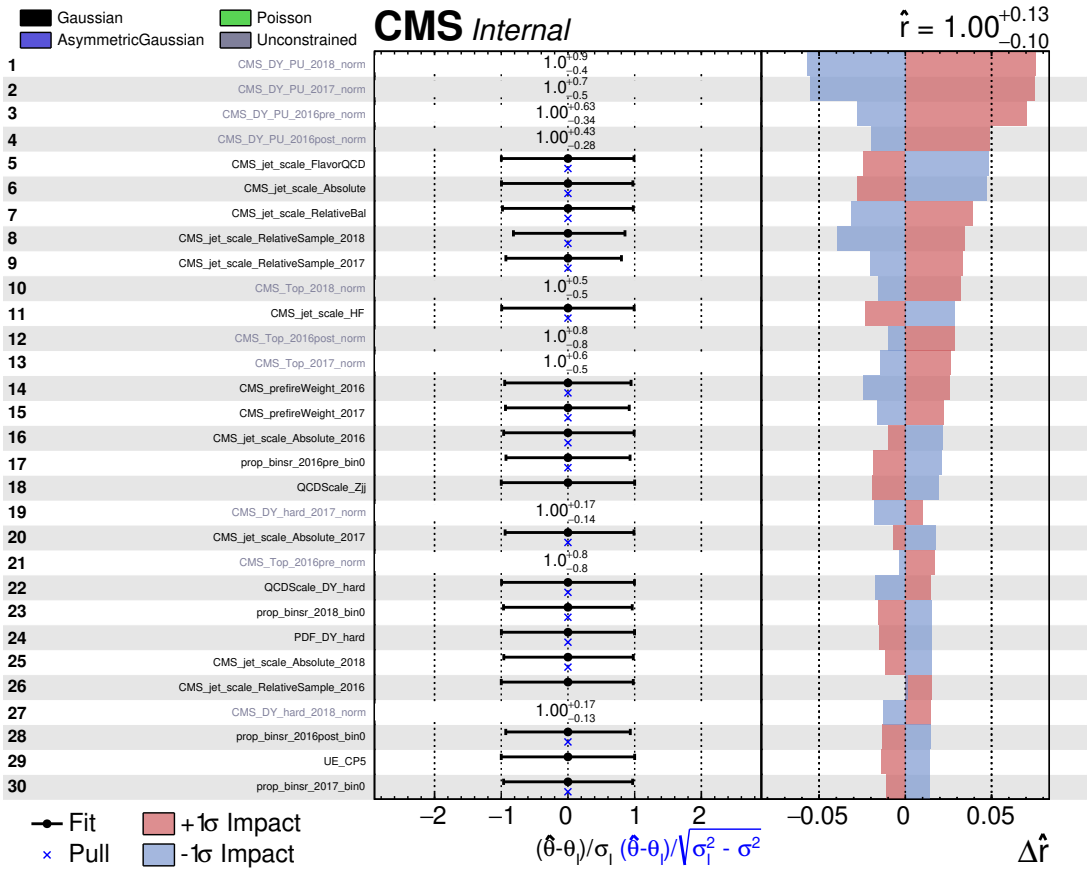

Figure 2.25: Ranking of nuisance parameters according to their impact on the fiducial signal strength μ_{fid}.

2.13 Unfolded measurement

2.13.1 Fit strategy

To perform the differential cross-section measurement we perform the unfolding of the kinematic distributions with with a maximum likelihood method. The signal is split into different generator-level (GEN) bins for each observable to be unfolded. The maximum likelihood unfolding is performed with the `combine` tool, similarly to what is done for the inclusive and fiducial measurements described in Sections 2.11 and 2.12. The signal is divided into different GEN-level bins for each observable to be unfolded. A signal strength modifier is assigned to each GEN bin, allowing the fit to extract the differential cross-section in each bin. No regularization is applied in the unfolding procedure, as the condition numbers, reported in Table 2.13, indicate a well-conditioned problem.
The variable for the fit in the signal region is the kinematic observable under study, with a binning that matches the GEN-level binning, but with twice as many bins. This choice provides additional constraints on the different signal strength modifiers associated with each bin, improving the sensitivity of the measurement.
The fiducial region definition used to extract the unfolded results is the same as that described in Table 2.11, ensuring consistency between the measurement phase space and the theory predictions. Theoretical uncertainties are treated as acceptance uncertainties and are normalized to the nominal fiducial yield in each GEN bin. That is, for each bin, the variations from PDFs, QCD scales and PS are normalized with respect to the total nominal cross-section in the corresponding fiducial bin.
The list of generator-level observables considered for the unfolding, together with their binning, is reported in Table 2.12. When the range of the binning does not match the selection, we take into account the overflow. The binning choice is optimized to balance statistical precision and resolution effects, ensuring that each bin has sufficient events while minimizing migration effects due to detector resolution.

Observable	**Binning**
$\lvert\Delta\eta_{jj}\rvert$	[3, 4, 5, 6, 7, 8]
m_{jj} [GeV]	[1000, 1600, 2200, 2800, 3400, 4000]
p_{jj}^{T} [GeV]	[0, 60, 120, 180, 240, 300]
$\lvert\Delta\phi_{jj}\rvert$	[0, 1.5, 2.1, 2.5, 2.82, π]
$\lvert\Delta\phi_{ll}\rvert$	[0, $\frac{\pi}{5}$, $\frac{2\pi}{5}$, $\frac{3\pi}{5}$, $\frac{4\pi}{5}$, π]
p_{ll}^{T} [GeV]	[0, 80, 160, 240, 320, 400]

Table 2.12: Generator-level observables considered in the unfolding and their binning.

2.13.2 Unfolding variables

For each kinematic variable we present in Figures 2.26 - 2.27, the detector level distribution in the SR for a year of example, 2018, and the migration matrix.
The migration matrix of GEN bins at detector level and in the ideal case it's completely diagonal. The matrix is normalized such that the sum of the expected events in each gen bin is 1.0.
For m_{jj} and p_T^{jj}, the migration matrices exhibit a degree of non-diagonality, as expected due to the jet energy resolution and the detector response. As a result, we anticipate

Observable	Condition Number
$\lvert\Delta\eta_{jj}\rvert$	5.86
m_{jj} [GeV]	4.65
p_{jj}^{T} [GeV]	4.04
$\lvert\Delta\phi_{jj}\rvert$	4.80
$\lvert\Delta\phi_{ll}\rvert$	1.67
p_{ll}^{T} [GeV]	5.25

Table 2.13: Condition numbers for the unfolding migration matrices of each observable. The condition numbers indicate a well-conditioned unfolding problem, with values significantly lower than typical thresholds for ill-conditioning.

larger uncertainties in the unfolding of these observables. Nonetheless, they remain key variables for characterizing the VBF process.

2.13.3 Results

We perform the unfolding using the official CMS electroweak Zjj sample generated with MadGraph5 LO, and compare the resulting differential distributions with the two alternative predictions at NLO: one generated with MadGraph5_aMC@NLO and another produced with POHWEG. All samples are interfaced with Pythia8 using the dipole recoil showering scheme. The comparison allows us to assess the modeling uncertainties related to higher-order corrections and matrix-element implementations.
A likelihood scan is performed for each parameter of interest (POI), profiling all other parameters to assess the expected uncertainties. We report together with the scans, in Figures 2.28-2.29, the expected differential distribution of the unfolded variables. The data points are the r_i scaled by MG@LO fiducial cross-sections, while the gray boxes include all the uncertainties on r_i. The theoretical prediction uncertainties depicted with colored boxes include theoretical uncertainties and the MC stats.

2.13.4 Control of the DY Hard Shape

The accuracy of the unfolded measurement of electroweak (EWK) Zjj production strongly depends on the precise modeling of the DY background, in particular the so-called DY Hard component. This process constitutes roughly 50% of the expected yield in the signal region, making its description one of the dominant factors in the stability and robustness of the unfolding strategy. Any mismodeling of the DY Hard shape could propagate into biases in the signal extraction and, consequently, in the unfolded differential cross sections.

Motivation The DY Hard background corresponds to events with a hard-scattering topology that closely mimics the kinematics of the electroweak signal. In this phase space, the accuracy of the parton shower, higher-order QCD corrections, and matching to the matrix element play a crucial role. Although dedicated NLO+PS simulations are employed, residual mismodeling cannot be excluded. It is therefore essential to validate that the fit procedure has enough flexibility to absorb such discrepancies through nuisance parameters and template morphing, ensuring that the unfolding is not biased by limitations of the MC description.

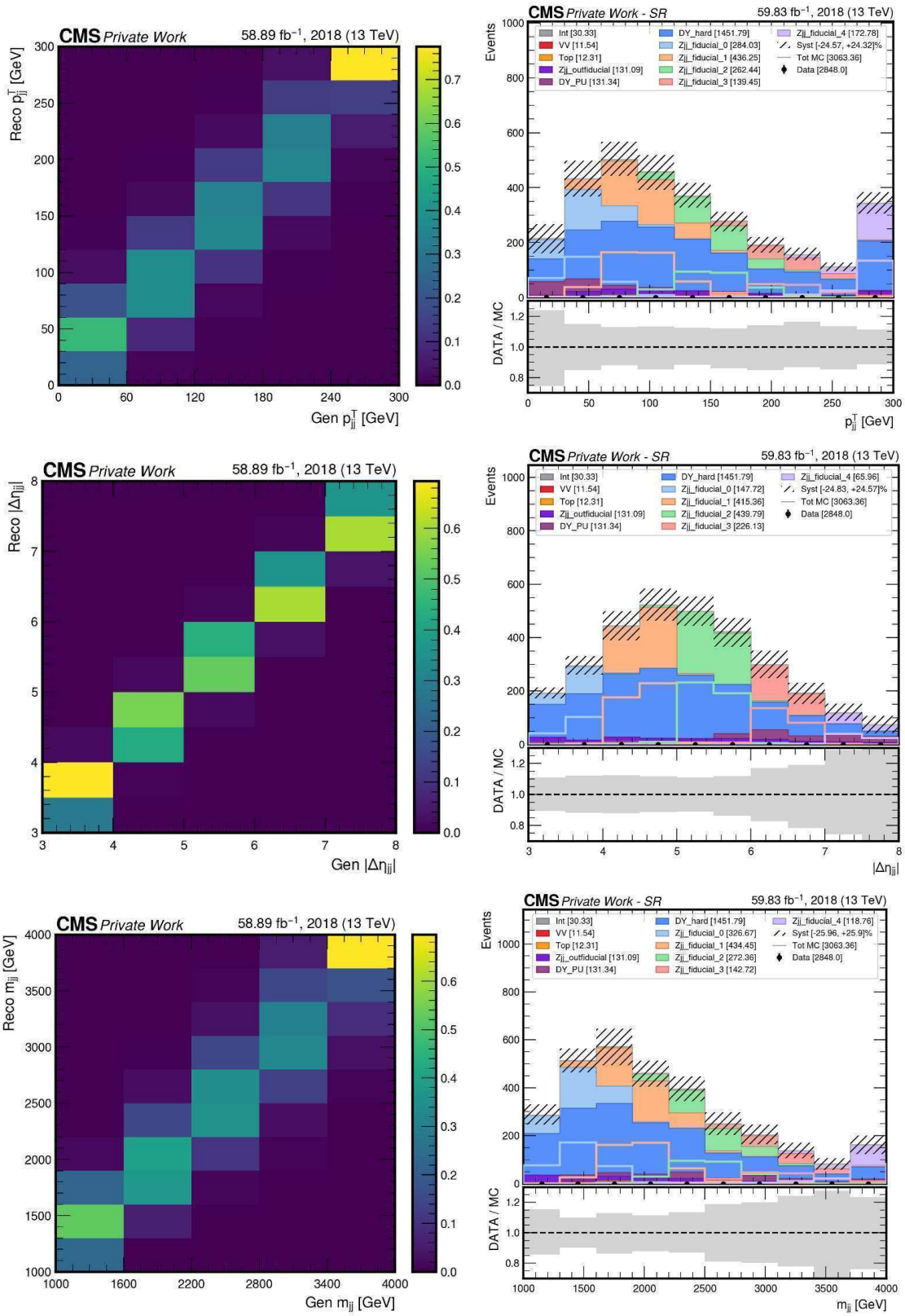

Figure 2.26: Migration matrices (left) and reconstructed-level distributions (right) for the main observables used in the unfolding procedure: dijet transverse momentum (p_{jj}^T), pseudo-rapidity separation ($\Delta\eta_{jj}$), and dijet invariant mass (m_{jj}). The migration matrices illustrate the detector response and the purity/stability of each Gen bin, normalizing for each column, while the right panels show the corresponding reconstructed-level distributions in the SR blinded.

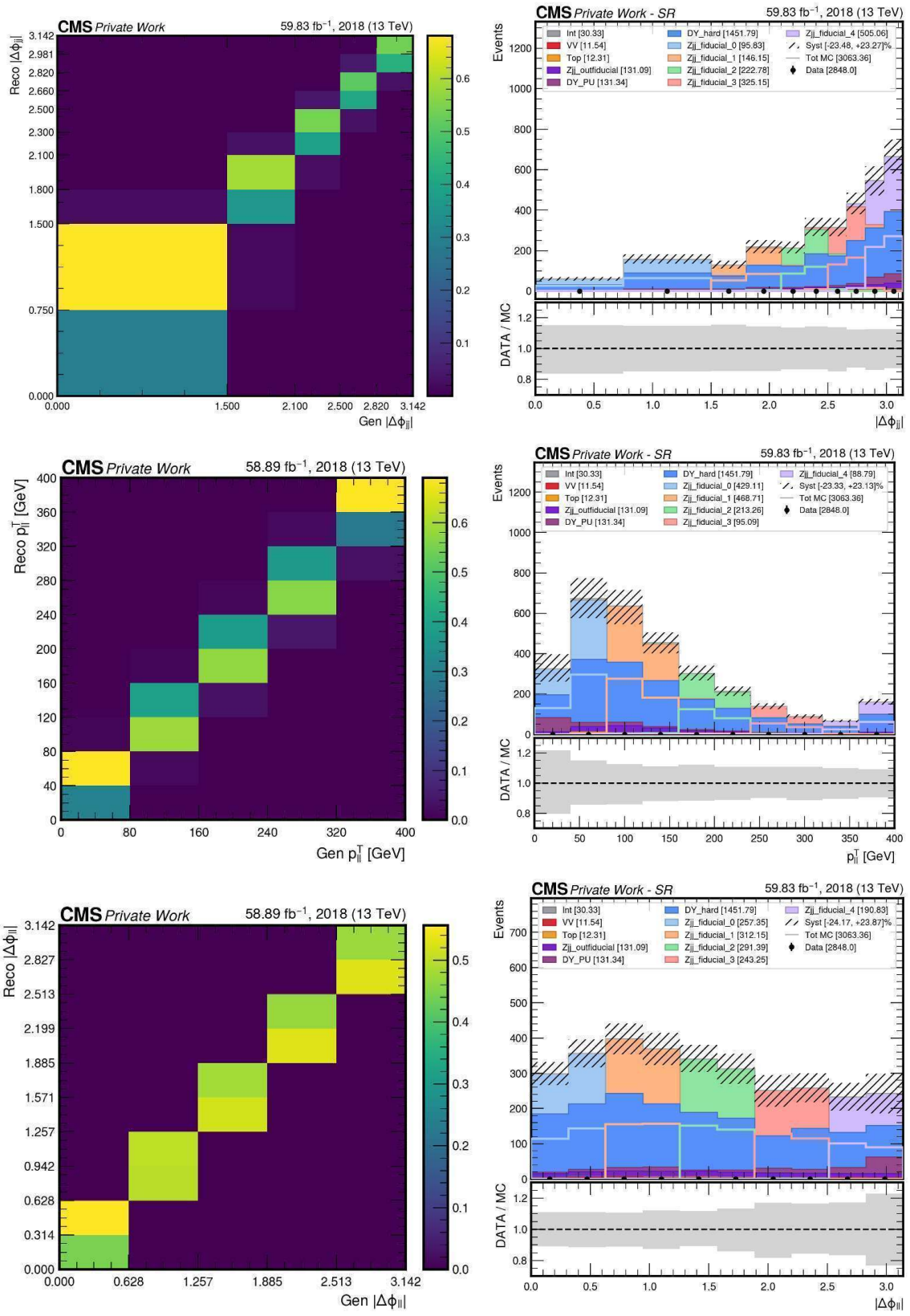

Figure 2.27: Migration matrices (left) and reconstructed-level distributions (right) for the main observables used in the unfolding procedure: dijet azimuthal separation ($\Delta\phi_{jj}$), dilepton transverse momentum (p_{ll}^T), and dilepton azimuthal separation ($\Delta\phi_{ll}$). The migration matrices illustrate the detector response and the purity/stability of each Gen bin, normlizing for each column, while the right panels show the corresponding reconstructed-level distributions in the SR blinded.

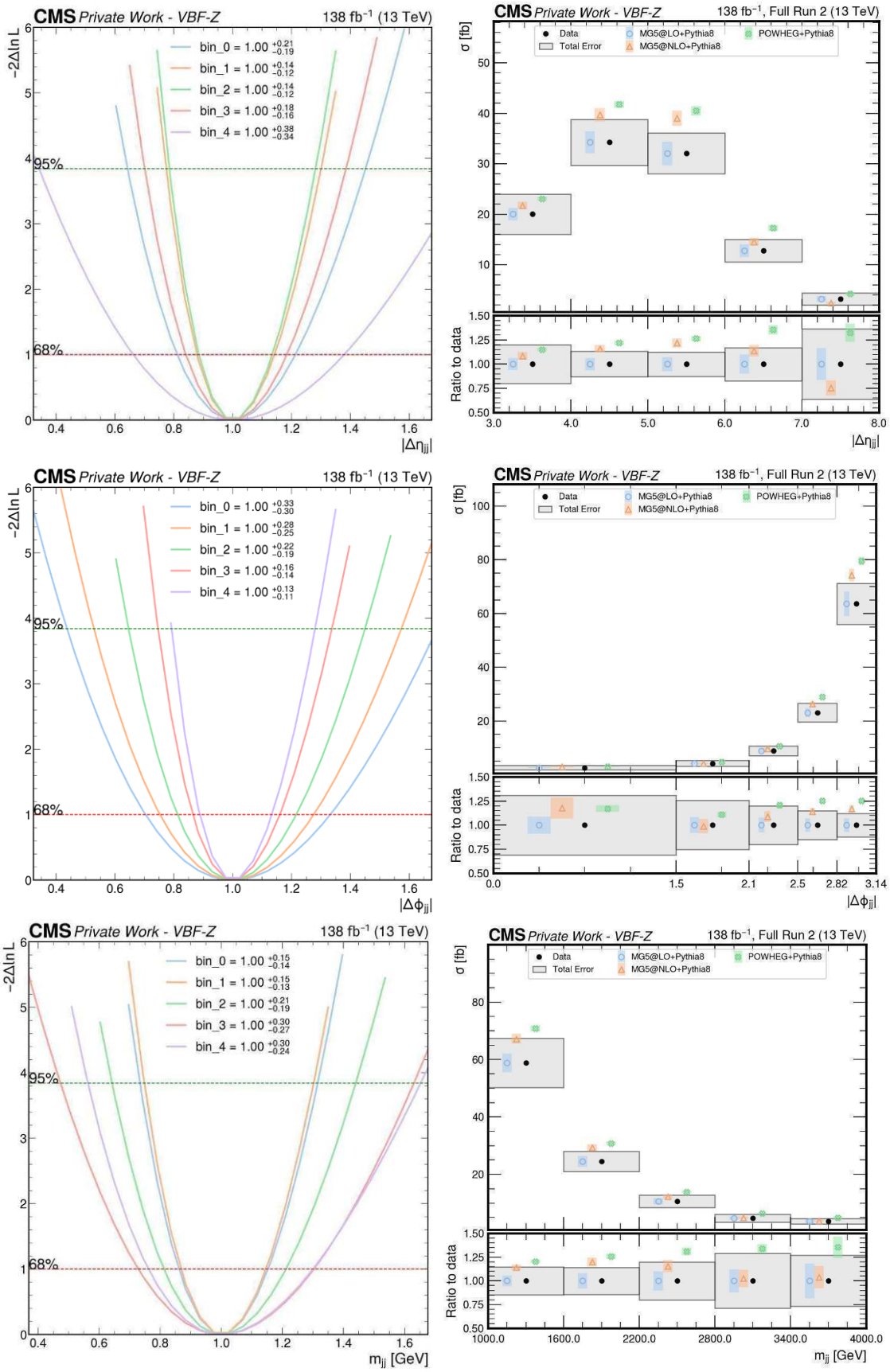

Figure 2.28: Expected likelihood scans (left) and unfolded differential cross sections (right) for the observables $\Delta\eta_{jj}$ (top), $\Delta\phi_{jj}$ (middle), and m_{jj} (bottom). The data points represent the unfolded measurements scaled by the MG@LO fiducial cross-sections, with gray boxes indicating the total uncertainties. The theoretical predictions from MG@NLO and POWHEG are shown with colored boxes representing their respective uncertainties.

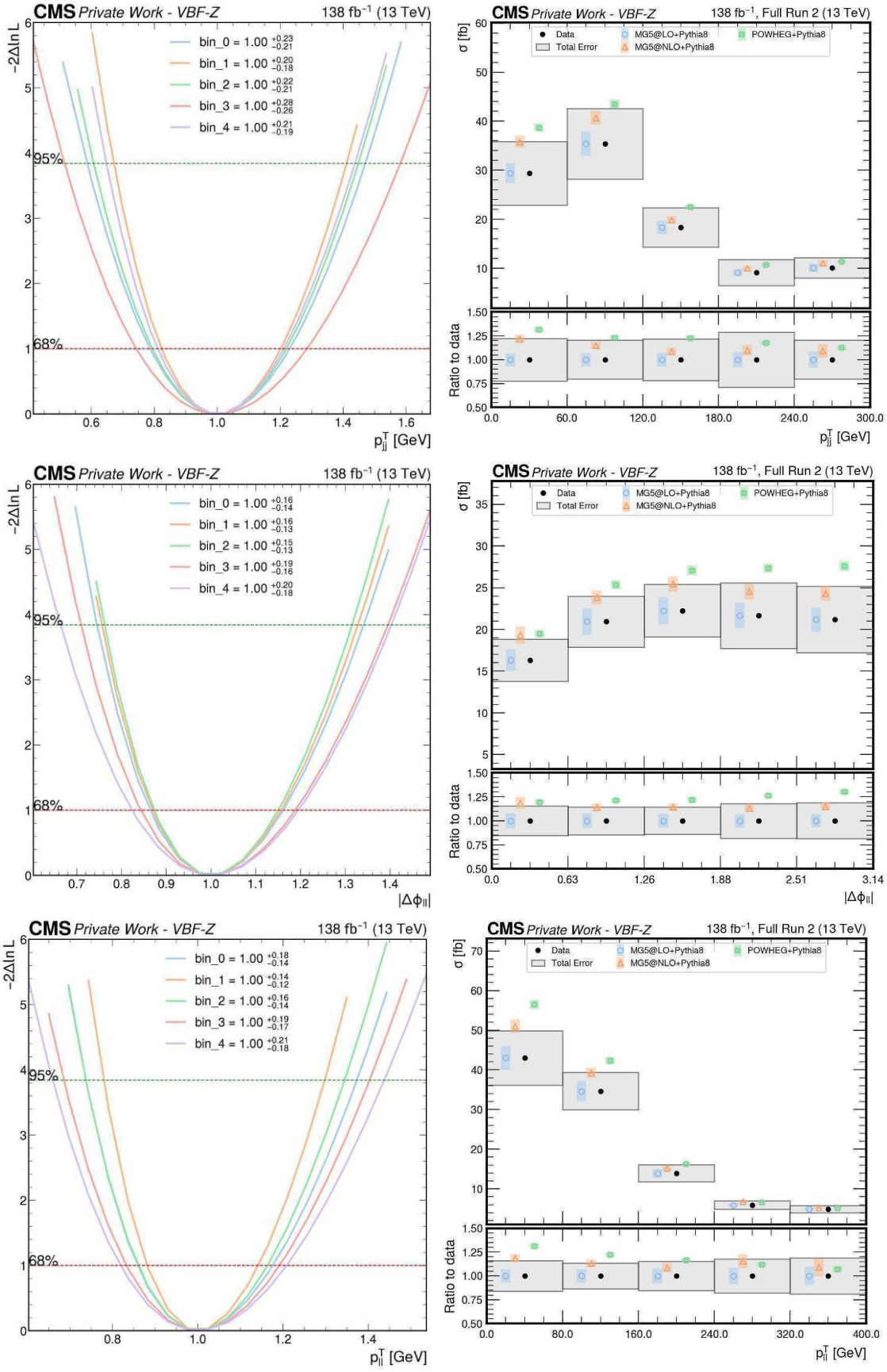

Figure 2.29: Expected likelihood scans (left) and unfolded differential cross sections (right) for the observables p_T^{jj} (top), $\Delta\phi_{\ell\ell}$ (middle), and $p_T^{\ell\ell}$ (bottom). The data points represent the unfolded measurements scaled by the MG@LO fiducial cross-sections, with gray boxes indicating the total uncertainties. The theoretical predictions from MG@NLO and POWHEG are shown with colored boxes representing their respective uncertainties.

To validate the DY Hard modeling we check that it well describes data in the DY CR, when rescaling each bin according to the SR fit results.
Figures 2.30 to 2.31, where the distributions for the analyzed variables, are shown before and after the fit, clearly display a very good level of agreement as quantified by the reported χ^2 values.

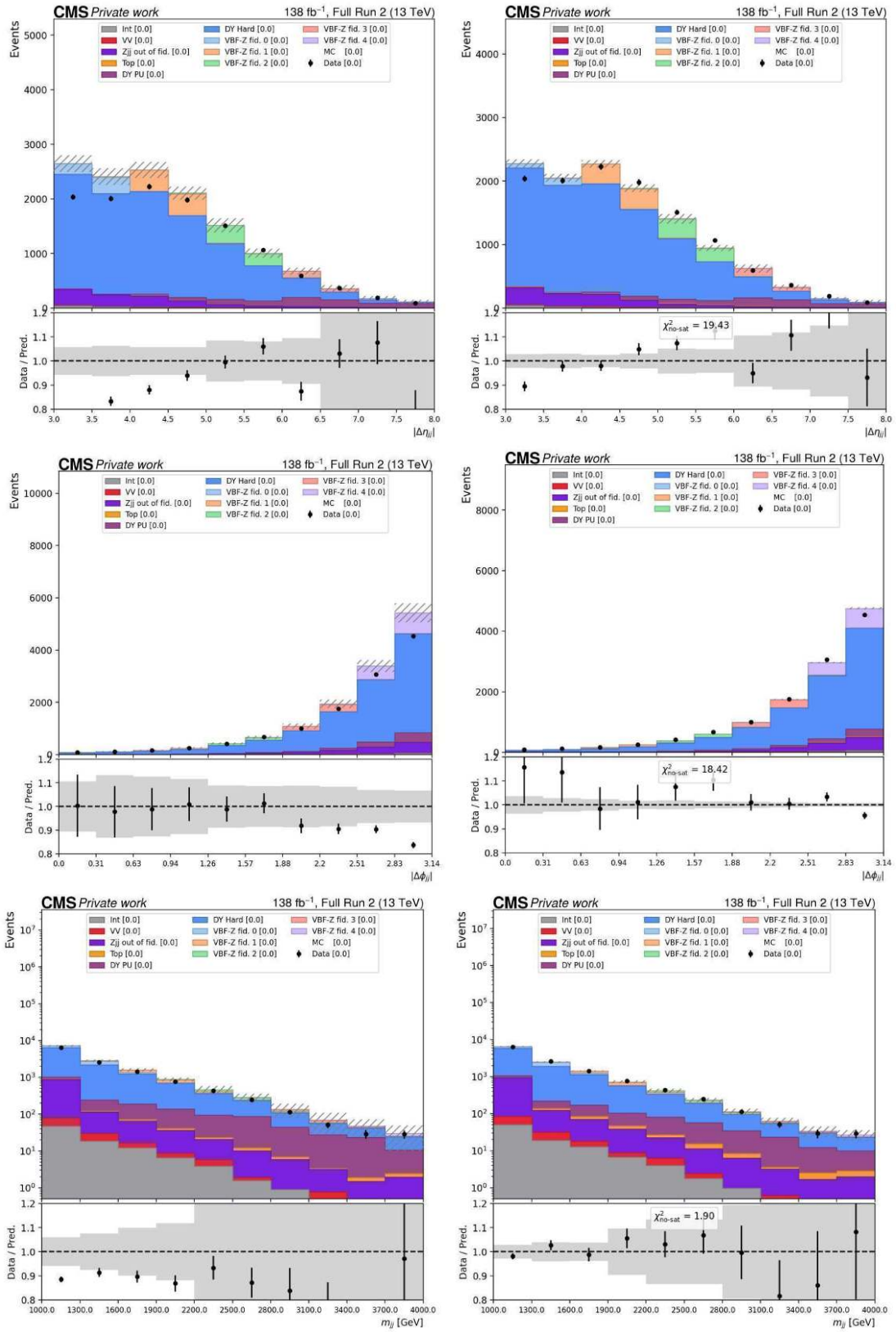

Figure 2.30: Pre-fit (left) and post-fit (right) distributions of $|\Delta\eta_{jj}|$ (top), $\Delta\phi_{jj}$ (middle), and m_{jj} (bottom) in the DY Hard control region. In the post-fit plots, the χ^2 accounts only for the statistical uncertainties of data and is not divided by the number of degrees of freedom.

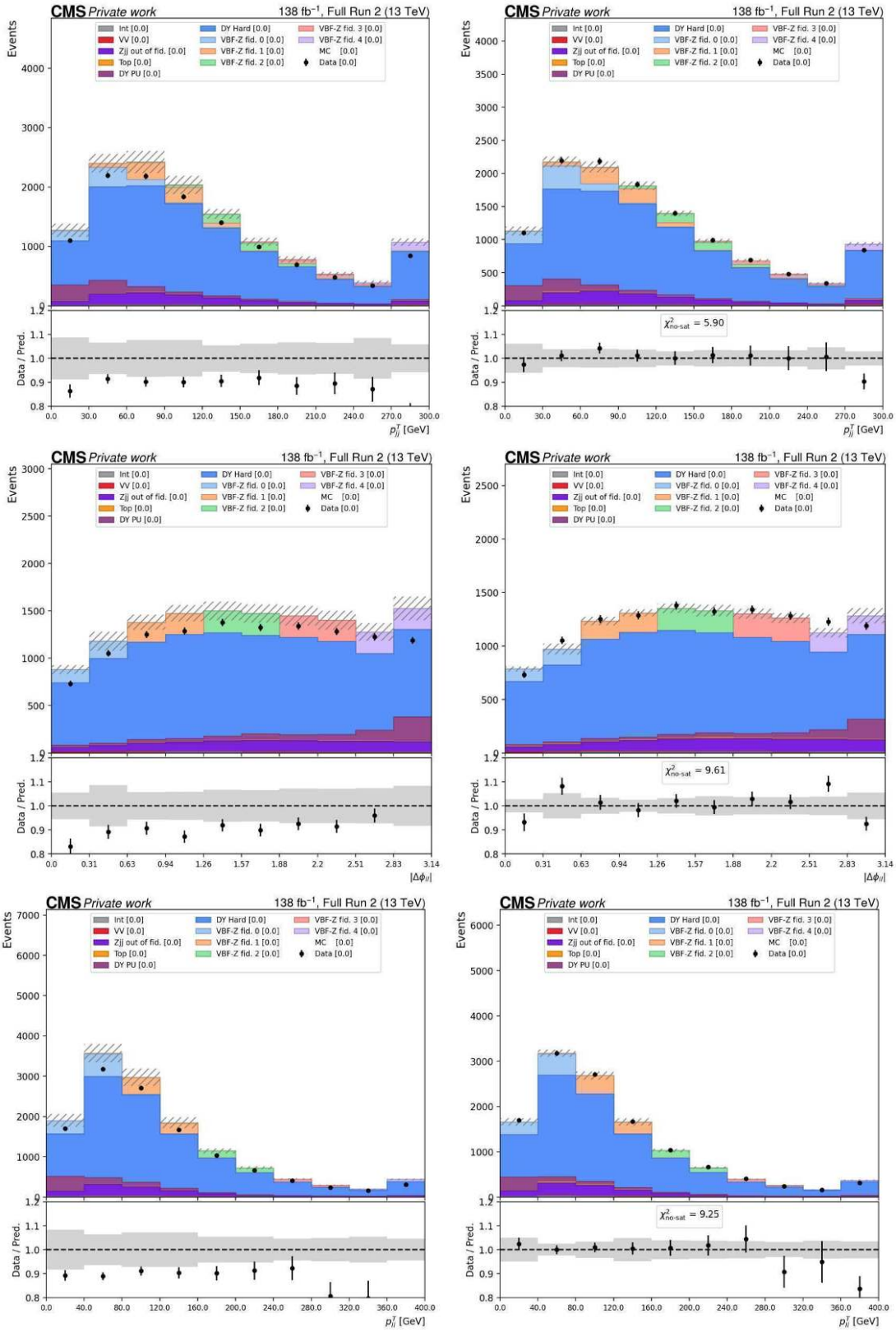

Figure 2.31: Pre-fit (left) and post-fit (right) distributions of p_T^{jj} (top), $\Delta\phi_{\ell\ell}$ (middle), and $p_T^{\ell\ell}$ (bottom) in the DY Hard control region. In the post-fit plots, the χ^2 accounts only for the statistical uncertainties of data and is not divided by the number of degrees of freedom.

2.14 Effective Field Theory Interpretations

The search for deviations from the predictions of the Standard Model (SM) can be systematically pursued within the framework of the Standard Model Effective Field Theory (SMEFT). In this approach, potential new physics at energy scales higher than those directly accessible at the LHC is parametrized through higher-dimensional operators constructed with only SM fields and respecting SM symmetries, which excludes odd dimension operators. The leading contributions correspond to operators of dimension six, whose effects can modify both the total rate and the kinematic distributions of SM processes. These deviations often become more pronounced in the high-energy tails of observables, where the sensitivity to new interactions is enhanced.
The VBF production of Z bosons provides an especially attractive laboratory for SMEFT studies. Its characteristic event topology grants direct access to the electroweak sector in a phase-space region with reduced QCD backgrounds. Several dimension-six operators can alter the electroweak couplings relevant to this process, leading to distortions in the shapes of key observables such as m_{jj}, $\Delta\eta_{jj}$, and $\Delta\phi_{jj}$. The theoretical cleanliness of the VBF topology, combined with the full Run 2 integrated luminosity, results in a competitive and complementary probe of possible beyond-the-SM effects.
Effective Field Theory studies for VBF-Z have been previously conducted using 2016 only dataset by CMS [91] and by ATLAS using the full Run 2 dataset [90]. Building upon these foundations, the current analysis extends the EFT interpretation to the complete Run 2 dataset collected by CMS. This expanded dataset enhances the statistical power of the study and allows for more stringent constraints on the Wilson coefficients of dimension-six operators.
In the SMEFT interpretation, possible EFT-induced shape and normalization modifications are considered for the electroweak Zjj signal process, which is the primary target of the measurement. Conversely, EFT effects are not explicitly modeled in the dominant Drell–Yan (DY) Z+jets background. This choice is motivated by the analysis strategy adopted for the DY contribution, which is constrained using a data-driven procedure through dedicated control regions. Since the DY normalization (and, to a large extent, residual mismodeling in the relevant phase space) is already absorbed by this data-driven background determination, introducing SMEFT variations in the DY simulation would be largely degenerate with the fitted background constraints and would not provide additional, robust sensitivity. The EFT interpretation is therefore focused on deviations affecting the electroweak production amplitude, where genuine sensitivity to anomalous gauge interactions is expected.
The following section presents the SMEFT interpretation performed in the context of a broader SM EFT combination within CMS. This global effort combines multiple electroweak processes—covering VBF, VBS, diboson, and triboson final states—to simultaneously constrain a set of six bosonic Wilson coefficients. Among these six operators, only three yield leading-order, physically meaningful contributions to this process. As a result, the sensitivity of this channel is particularly valuable in improving the overall precision of the combination, and in reducing flat directions when interpreted jointly with other measurements.

EFT Reweighting

To incorporate the effects of dimension-six SMEFT operators into the VBF-Z signal model, this analysis employs Monte Carlo samples generated with MadGraph5. The

SMEFT framework is implemented through a dedicated UFO model with SMEFTSim [66], which allows the modification of electroweak interactions at leading order by introducing new Lorentz structures and coupling dependencies. Directly generating independent samples at every point in the Wilson coefficient parameter space would be computationally prohibitive, given the large number of contributing diagrams in this topology. Instead, a reweighting strategy is adopted, in which a single underlying event sample is enriched with additional weights that encode the dependence of the cross-section on the chosen EFT parameters. This approach efficiently captures shape and rate variations induced by the operators of interest while keeping the computational cost manageable.
The process command used in MadGraph5 for the generation of the EFT samples is the following:

- `p p > l+ l- j j QCD=0 NP=1 SMHLOOP=0`

The EFT operators considered in this studies are c_W, c_{HWB} and c_{HDD}. The latters are the subset of Bosonic operators studied in the EFT SMP VBS/VBF combination that are sensitive to EW-Zjj.
The effects of this operators is only included in vertices and not propagators, see Section 2.14 for more details.
To accelerate the generation of Standard Model (SM) and Effective Field Theory (EFT) samples, a reweighting strategy is employed. In this approach, only a limited phase space—typically corresponding to the SM or the SM supplemented by a small subset of dimension-six operators—is explicitly integrated. New event weights corresponding to arbitrary points in the EFT parameter space are then computed using the reweighting functionality provided by MadGraph5. For each generated event, a new weight is assigned according to [102]:

$$\begin{aligned} \omega_{\text{new}} &= \omega_{\text{orig}} \cdot |M_{\text{new}}|^2/|M_{\text{orig}}|^2 \\ \omega_{\text{orig}} &= f_1(x_1,\mu_F) \cdot f_2(x_2,\mu_F) \cdot |M_{\text{orig}}|^2 \cdot \Omega_{\text{PS}} \end{aligned} \tag{2.5}$$

with f denoting parton distribution functions and Ω_{PS} is the phase space measure associated to the event. The usage of reweighting for SMEFT studies addresses two problems:

- events are generated only once drastically reducing the computation needed for independent generations of the $n(n+3)/2+1$ terms of the EFT expansion for N number of operators;
- the phase space factor Ω_{PS} is the same for all the new hypothesis. This ensures that the handling of the reweighting weights, needed to extract the terms in the EFT parametrization, will mathematically return quantities with well-defined properties. The most important one is the positive-definiteness of the full EFT prediction, inclusive or differential for each bin of a distribution, for arbitrary values of a Wilson Coefficient.

By denoting k a general Wilson coefficient, the event weight ω will depend on the Wilson coefficients as:

$$\begin{aligned} \omega(k) &= \omega(\text{SM} + k\text{Lin} + k^2\text{Quad}) \\ \omega(k_m,k_n) &= \omega(\text{SM} + \sum_{i\in[m,n]} k_i\text{Lin}_i + k_i^2\text{Quad}_i + 2\text{Mix}_{ij}) \end{aligned} \tag{2.6}$$

The weights corresponding to the components of the EFT parametrisation can be retrieved via simple algebra from the generation with $\omega(k=0)$, $\omega(k=1)$ and $\omega(k=-1)$ for the one-dimensional case and with $\omega(k_i=k_j=1)$, $\omega(k_i=1,k_j=0)$, $\omega(k_i=0,k_j=1)$ and $\omega(k_i=k_j=0)$ for the two-dimensional case. The algebraic relations are reported in Equation 2.7.

$$\begin{aligned}
\omega_{\mathrm{SM}} &= \omega(k=0) \\
\omega_{\mathrm{Lin}} &= \frac{1}{2}[\omega(k=1)-\omega(k=-1)] \\
\omega_{\mathrm{Quad}} &= \frac{1}{2}[\omega(k=1)+\omega(k=-1)-2\omega(k=0)] \\
\omega_{\mathrm{Mix}_{ij}} &= \omega(k_i=k_j=1)+\omega(k_i=k_j=0)-\omega(k_i=1,k_j=0)-\omega(k_i=0,k_j=0)
\end{aligned} \tag{2.7}$$

EFT propagators

The choice of not including propagators effects is motivated by the fact that including EFT in propagators would require a complex generation where each operator is generated alone. This would also potentially break the EFT fitting algebra where the algebric sum of all the EFT components for a given combination of Wilson coefficients should always return a positive-definite result.
Nontheless the effect of propagators on c_{HWB} and c_{HDD} has been generated for the standalone components and studied finding non-negligible differences in the yield and shapes for $\Delta\phi_{jj}$:DNN. These effects are reported in Figure 2.32. This choice is also made to be consistent with the work of a parton level EFT study of VBS and diboson [103].

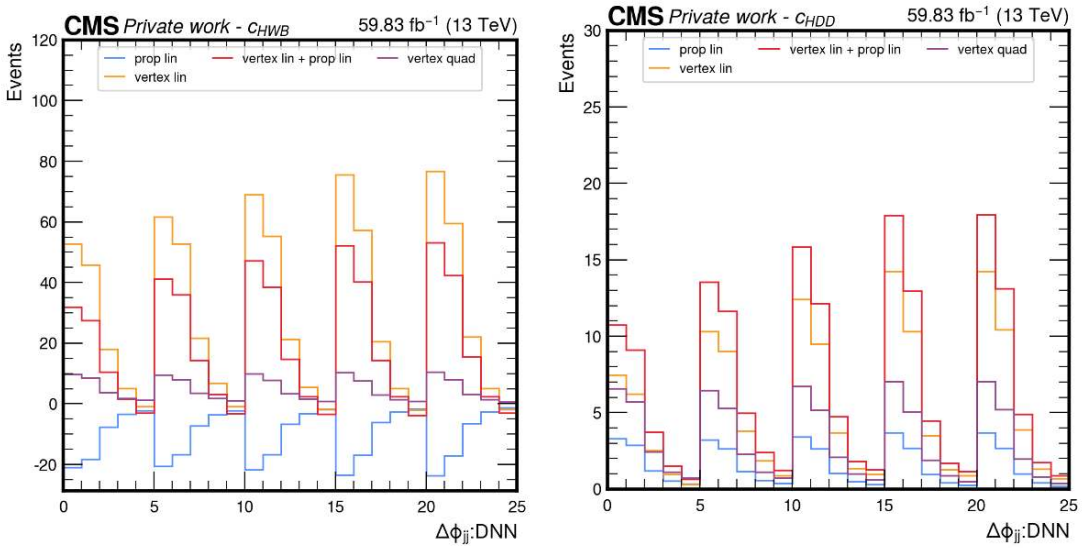

Figure 2.32: Comparison of the EFT shapes with and without propagator effects for c_{HWB} (left) and c_{HDD} (right) in $\Delta\phi_{jj}$:DNN.

2.14.1 EFT effects on kinematics

The effects of the selected EFT operators, c_W, c_{HDD}, and c_{HWB}, were studied in order to assess their impact on the kinematic distributions relevant to the VBF-Z channel.

Each operator induces characteristic modifications to specific observables, which provide differential sensitivity in the EFT interpretation. These operator-dependent distortions can be exploited to optimize the analysis strategy and to maximize sensitivity to potential deviations from the SM.
In particular, the variable found to be most sensitive to c_{HDD} is the DNN discriminant, while for c_W and c_{HWB} the $\Delta\phi_{jj}$ distribution provides the strongest discriminating power. The analysis strategy therefore exploits both multivariate and one-dimensional discriminants to maximize sensitivity. The DNN achieves improved separation for operators such as c_{HDD}, where correlations among leptonic and global event variables are particularly relevant. Conversely, the $\Delta\phi_{jj}$ observable is retained as an explicit differential measurement due to its strong response to modifications induced by c_W and c_{HWB}.
This combined approach avoids relying on a single discriminant and instead leverages the complementary information provided by different observables. The resulting two-dimensional distribution in the plane defined by the DNN output and $\Delta\phi_{jj}$ is shown in Figure 2.34, highlighting the characteristic distortions introduced by each operator.

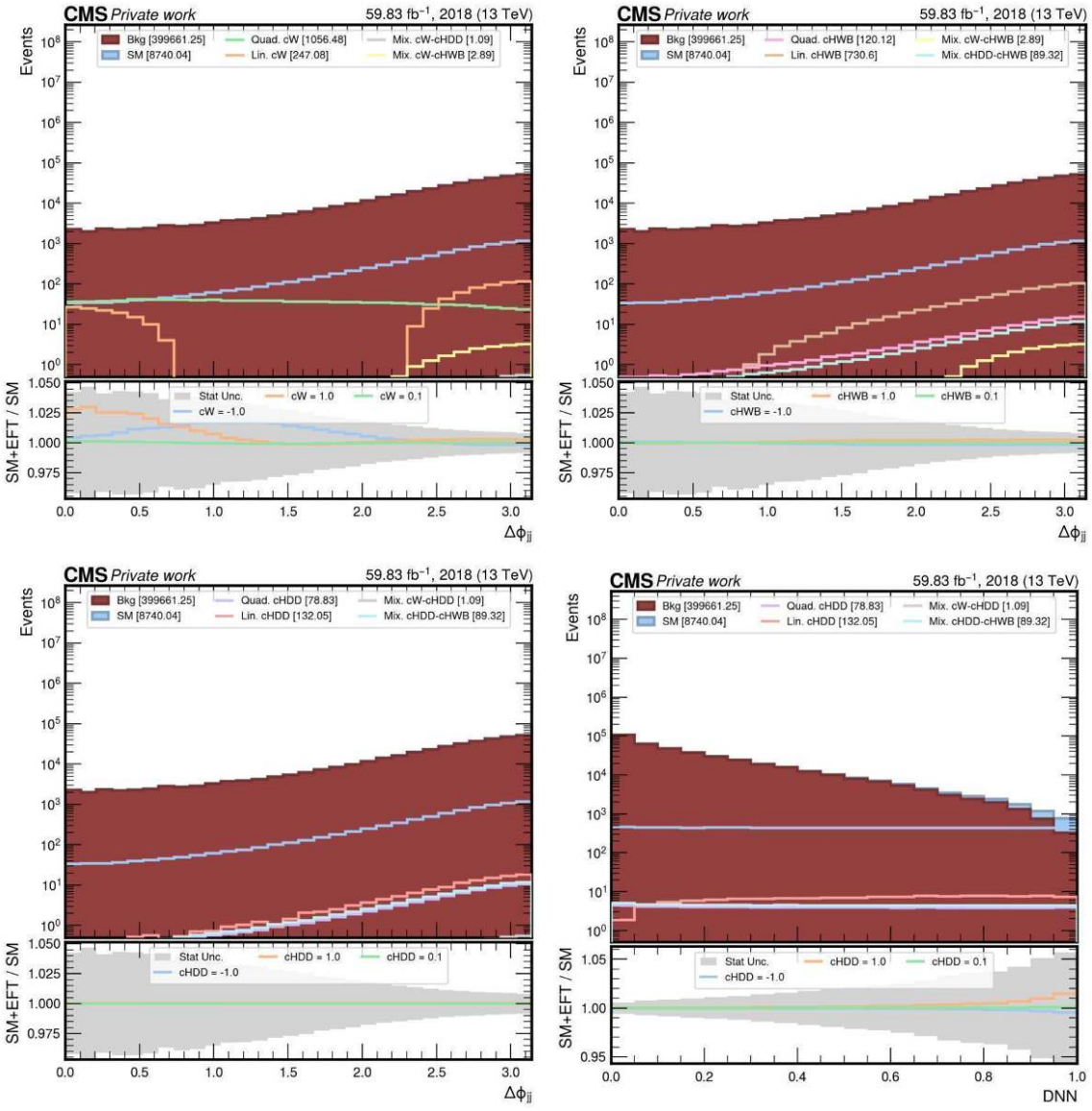

Figure 2.33: EFT distributions in the SR for c_W, c_{HWB} and c_{HDD} in $\Delta\phi_{jj}$

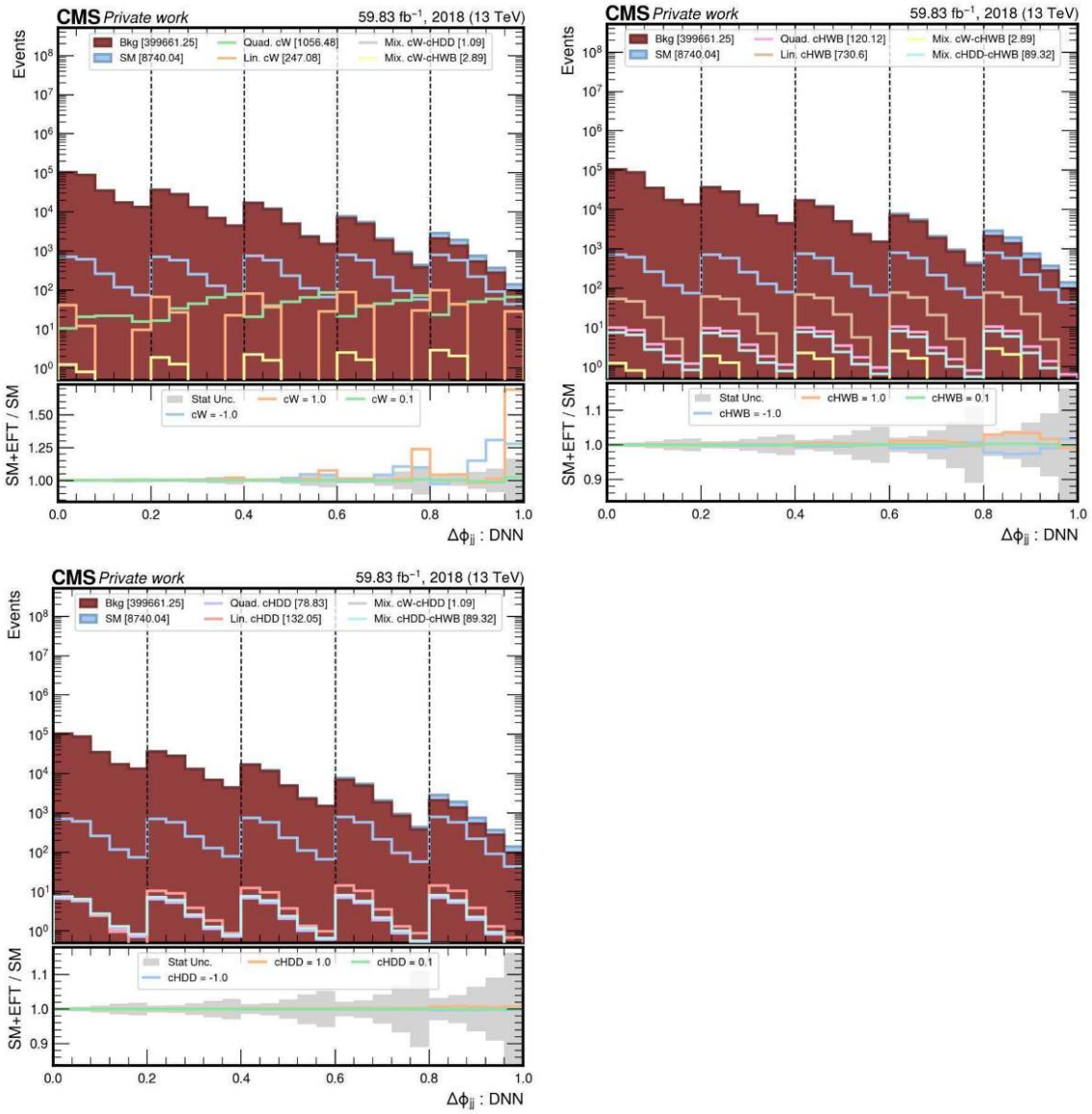

Figure 2.34: EFT distributions in the SR for c_W, c_{HWB} and c_{HDD} in $\Delta\phi_{jj}$: DNN

2.14.2 EFT results

Analysis strategy

The strategy adopted for the SMEFT interpretation closely follows the structure of the Standard Model (SM) analysis described in Section 2.7, with a few notable modifications motivated by the need to preserve sensitivity to potential EFT effects. In contrast to the SM measurement, no selection on the output of the Deep Neural Network (DNN) discriminant is applied. While such a requirement is effective in enhancing the purity of the electroweak Zjj signal, it may sculpt the kinematic distributions and bias the phase space where EFT contributions manifest, particularly in the lower range of the discriminant where interference effects can be more relevant. For this reason, the full spectrum of the DNN output is exploited in the likelihood fit.
The set of analysis regions is also adapted to reflect this strategy. The SR in the SMEFT fit incorporates the phase space previously associated with the DY Hard CR, since the

absence of a DNN threshold reduces the need for a separate high-purity selection. Two control regions are retained to constrain the dominant backgrounds: the DY PU CR and the Top CR. This simplified region structure ensures that the global fit remains sensitive to potential SMEFT-induced distortions while maintaining adequate control of leading systematic uncertainties.
Overall the regions are:

- **Signal Region (SR):** enriched in electroweak Zjj events and sensitive to potential EFT contributions;
- **Drell-Yan Pileup Control Region (DY PU CR):** dominated by Drell-Yan events with additional jets from pileup or lower-order QCD radiation, used to constrain the corresponding background normalization;
- **Top Control Region (Top CR):** enriched in events containing top-quark production, constraining the top background normalization.

A particular focus is given to the modeling of the Drell-Yan background, which dominates the event yield in the inclusive phase space. To account for residual mismodeling observed in the $\Delta\phi_{jj}$ distribution, see Section 2.14.2, which is the most sensitive observable to EFT effects, the Drell-Yan sample is divided into bins of this variable.
Moreover, the two DY components used in the simulation, referred to as *DY hard* (matrix element $Z + 2$ jets) and *DY PU* (matrix element $Z + 1$ jet), are assigned independent normalization parameters.
These normalizations are left free to float in the fit, allowing the data to constrain them directly and effectively making the background normalization fully data-driven for the two components of the DY and the different bins of $\Delta\phi_{jj}$.
The observables entering the fit are chosen to maximize sensitivity to EFT effects. In both the SR and the DY PU CR, a two-dimensional template built from the ($\Delta\phi_{jj}$: DNN) distribution is used, providing complementary information from the dijet topology and the DNN discriminant output.
The DY PU CR is important for constraining the normalization of both the DY hard and the DY PU components across the different $\Delta\phi_{jj}$ bins, primarily through the low-DNN region where the background dominates.
For the Top CR, where the purpose is solely to constrain the normalization of the top background, a single-bin variable is employed.
In summary, the EFT interpretation of the VBF-Z process follows the same conceptual structure as the SM analysis but with fully data-driven control regions.

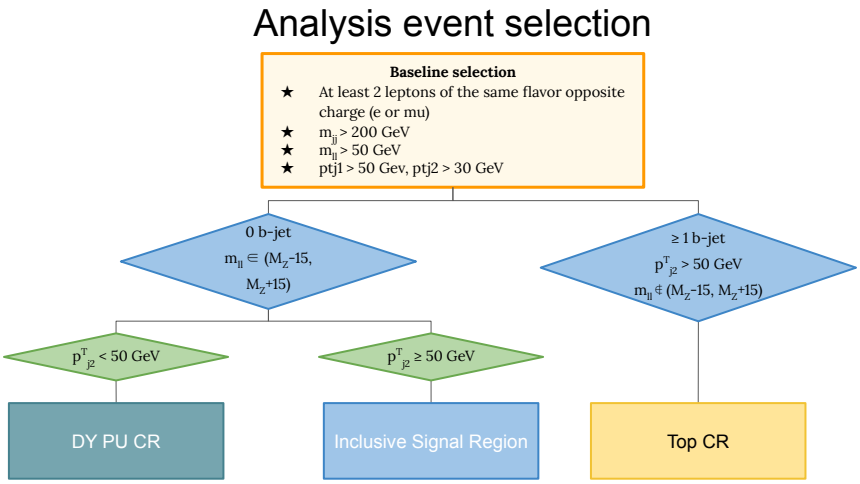

Figure 2.35: Control flow diagram summarizing the analysis strategy.

Prefit Distributions

As a sanity check we looked at the $\Delta\phi_{jj}$ variable in the DY PU CR and in the low significance region of the SR. A trend of disagreement, is observed that cannot be absorbed by the flat normalizations of the DY-hard and DY-PU. This trend is present in all the years and both lepton flavours. Only 2018 is reported in Figures 2.36-2.37, while the rest of years are in the Appendix B.0.1.
In the following sections a blinding based on the significance is applied in the SR to avoid biases in the studies.

- For the inclusive $\Delta\phi_{jj}$ the last two bins are blinded.
- For the 2D $\Delta\phi_{jj}$:DNN the last 5 bins are blinded.

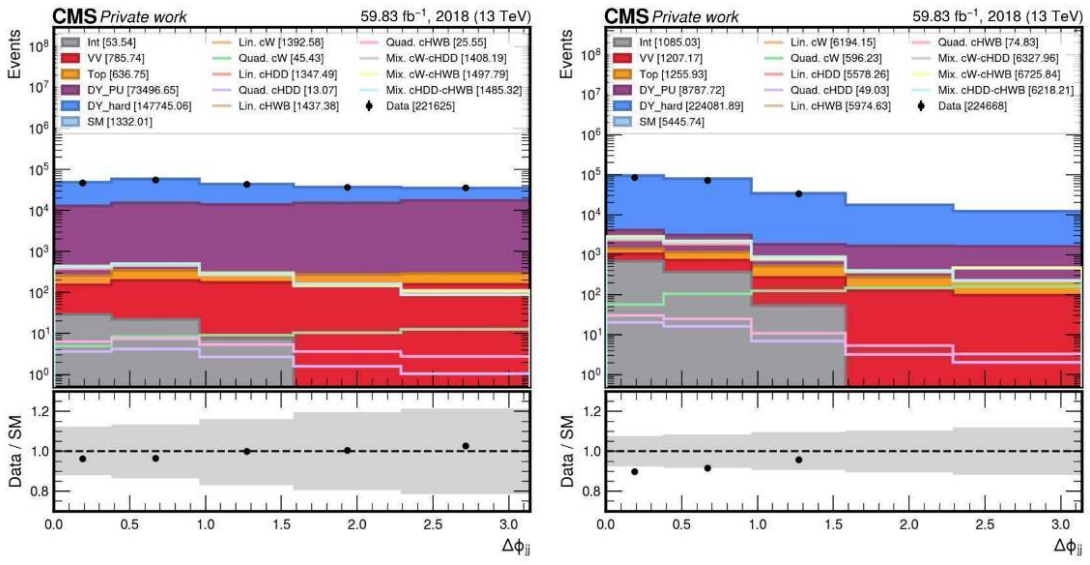

Figure 2.36: Bad modelling of $\Delta\phi_{jj}$ in 2018, left DY PU CR, right SR (last 2 bins blinded), e^+e^-channel

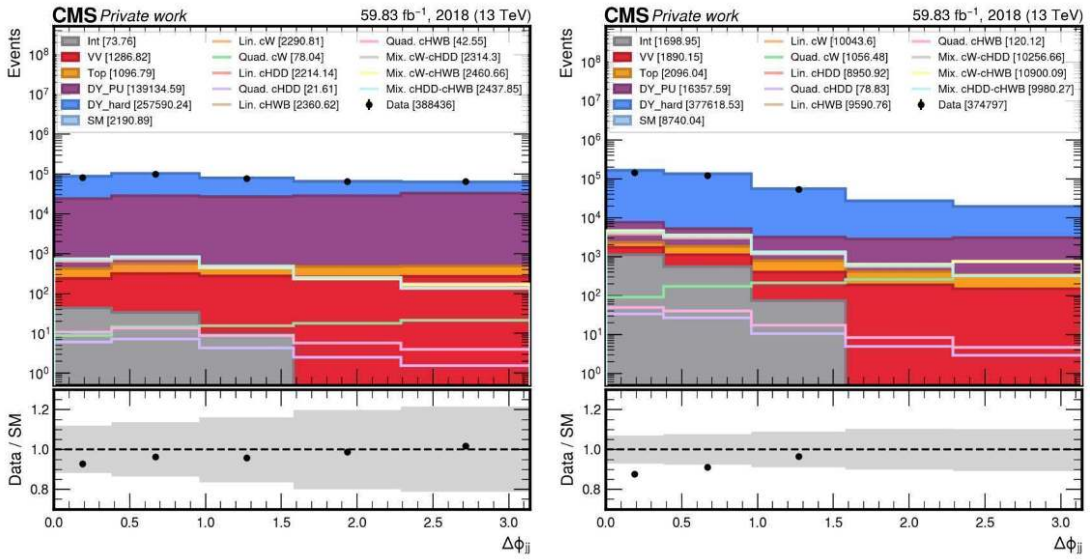

Figure 2.37: Bad modelling of $\Delta\phi_{jj}$ in 2018, left DY PU CR, right SR (last 2 bins blinded), $\mu^+\mu^-$channel

Overall the whole set of variables in the different regions are reported in Figures 2.38-2.40 for 2018 as example of the analysis strategy. The rest of years are in the Appendix B.0.2.

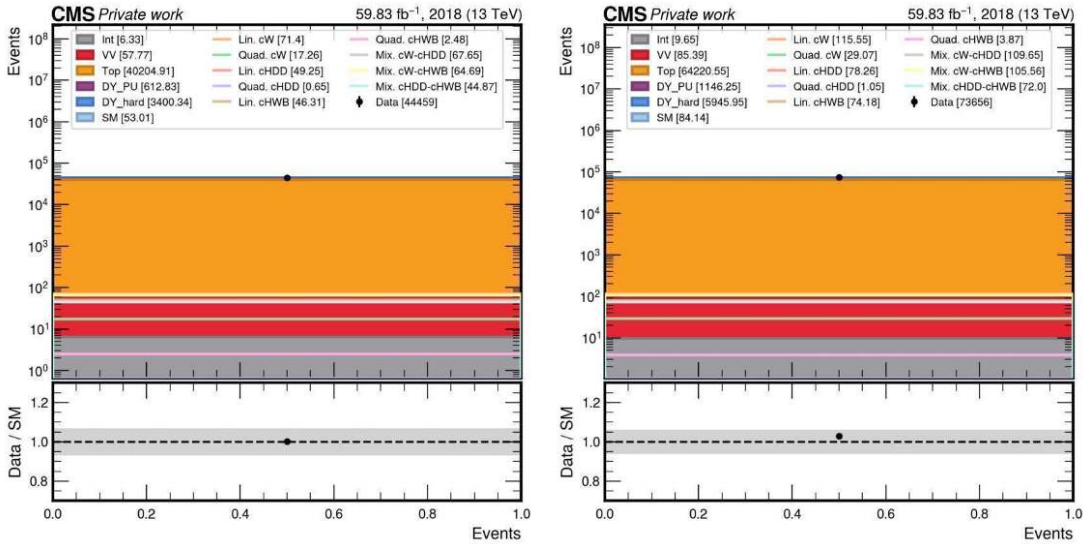

Figure 2.38: Analysis strategy in 2018 for the Top CR. Left is the e^+e^-channel, right $\mu^+\mu^-$

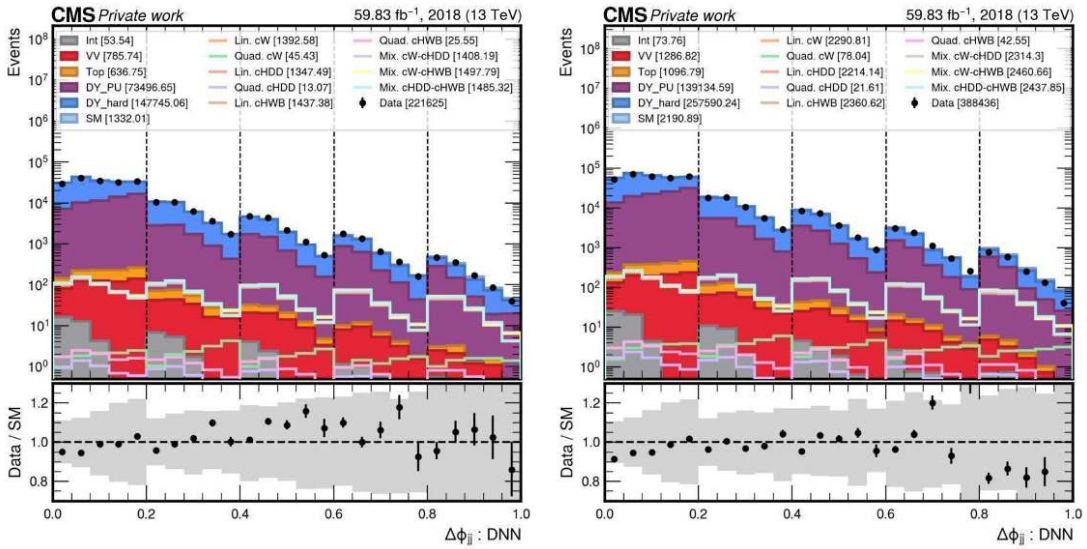

Figure 2.39: Analysis strategy in 2018 for the DY PU CR. Left is the e^+e^-channel, right $\mu^+\mu^-$

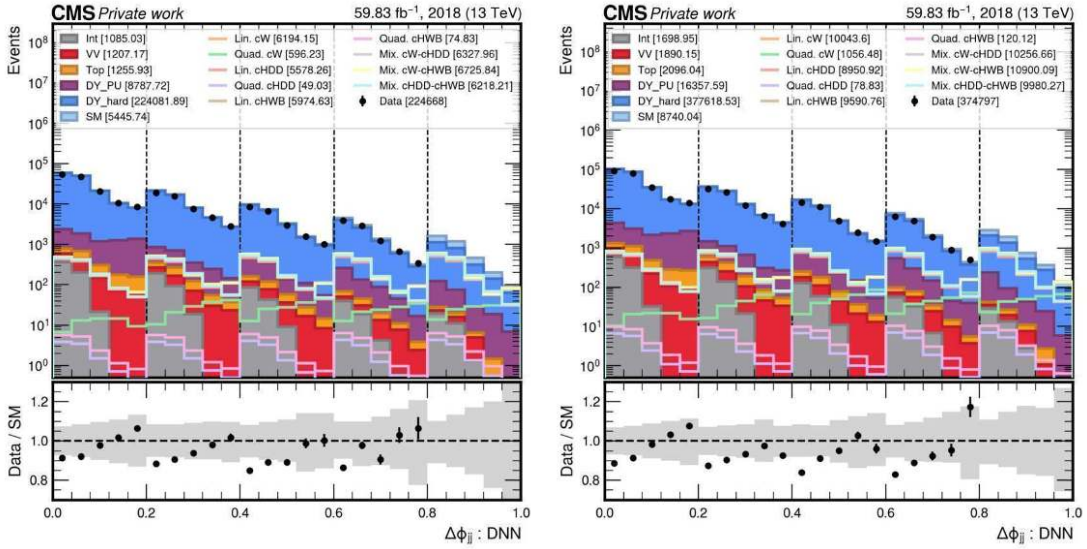

Figure 2.40: Analysis strategy in 2018 for the SR (last 5 bins blinded). Left is the e^+e^-channel, right $\mu^+\mu^-$

Postfit Distributions

Figures 2.41-2.43 show the post-fit agreement in the different regions when fitting c_W for 2018. The same plots for the other years are in the Appendix B.0.3.
In the post-fit the same blinding strategy as the previous section is used.
In all post-fit the χ^2_0 is checked to ensure the goodness of the fit. No significant discrepancies are observed.

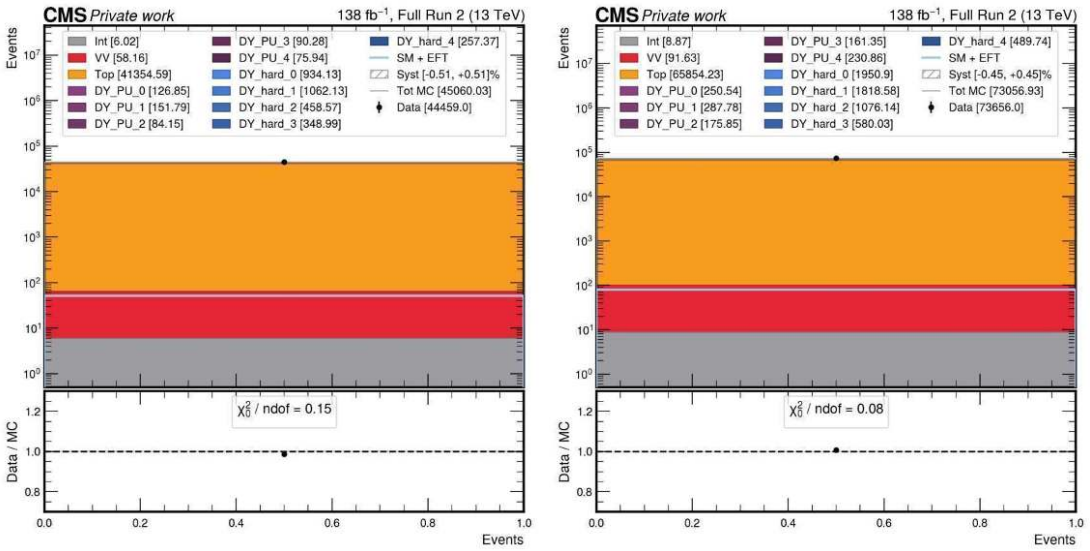

Figure 2.41: Post fit distributions in 2018 for Top CR. Left is the e^+e^-channel, right $\mu^+\mu^-$. EFT includes the postfit of cW.

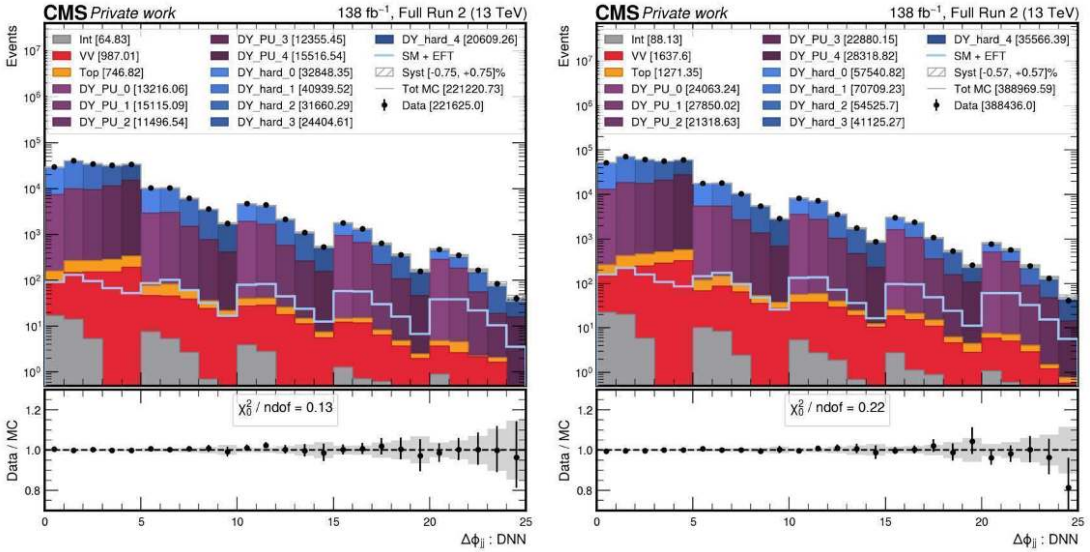

Figure 2.42: Post fit distributions in 2018 for DY PU CR. Left is the e^+e^-channel, right $\mu^+\mu^-$. EFT includes the postfit of cW.

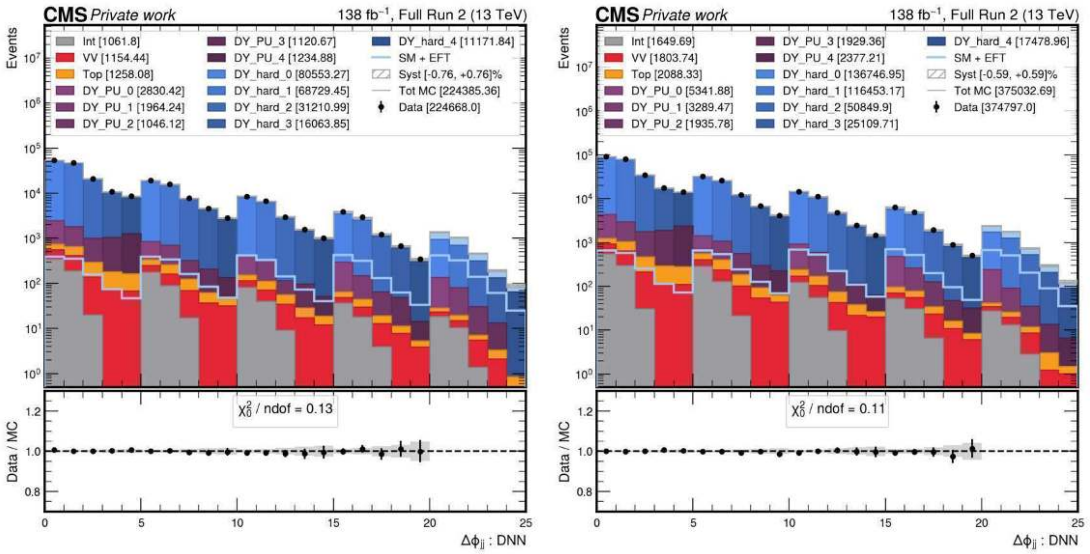

Figure 2.43: Post fit distributions in 2018 for SR (last 5 bins blinded). Left is the e^+e^-channel, right $\mu^+\mu^-$. EFT includes the postfit of cW.

Limits on Wilson Coefficients and Impacts of Systematic Uncertainties

Using the Full Run 2 dataset (138 fb^{-1}), we derived constraints on the Wilson coefficients associated with the selected EFT operators. The fits were performed by simultaneously incorporating the signal regions and control regions, allowing the nuisance parameters corresponding to theoretical and experimental systematics to be profiled consistently. The results include:

1. **One-dimensional (1D) limits**, where each Wilson coefficient is constrained individually while all others are set to zero. These results highlight the sensitivity of the analysis to individual operators and are particularly useful for comparison across channels and with global SMEFT combinations.

2. **Two-dimensional (2D) limits**, where pairs of Wilson coefficients are floated simultaneously. These contours provide insight into correlations between operators and the degree to which VBF-Z production can disentangle their contributions.

The constraints derived here represent the most stringent limits obtained so far in the VBF-Z topology, underlining the unique role of this channel in probing dimension-6 bosonic operators. In particular, the sensitivity to c_W, c_{HWB}, and c_{HDD} complements that of other diboson and VBS channels, contributing significantly to the global EFT picture.

Breakdown of Uncertainties The impact of systematic uncertainties was carefully quantified by profiling the likelihood with all nuisance parameters included. The uncertainties are grouped into three categories:

- **Experimental lnN/shape systematics:** these include lepton identification and trigger efficiencies, jet energy scale and resolution, luminosity determination, and pileup reweighting.

- **Rate parameters:** used to control normalization uncertainties of subdominant backgrounds such as $t\bar{t}$ and diboson production.
- **Monte Carlo statistical uncertainties:** arising from the limited size of simulated samples, treated as described in Section 2.6 with bin-by-bin uncertainties.

The results for all three operators classify the MC statistical uncertainties as a leading source of uncertainty. The dominant contribution originates from the Drell-Yan (DY) background. Despite the use of the most precise DY sample currently available in CMS, simulated at NLO accuracy in the second jet and produced in a dedicated jet-binned configuration to maximize statistics in the high-p_{T} and high-m_{jj} regime, the limited size of this sample prevents a further reduction of the corresponding uncertainty in the fit.
To illustrate the origin of this limitation, one can compare the expected number of DY events with at least two jets at matrix-element level in data to the number of generated events in the DY-2J sample used in 2018. The generated sample contains only about a factor of two more events than the number observed in data. After applying the full selection of the analysis and accounting for the fraction of events with negative weights, the resulting MC statistical uncertainty becomes comparable to the data statistical uncertainty itself.
This effect is more pronounced in the SMEFT interpretation than in the Standard Model cross-section measurement. The SMEFT analysis relies on the jet-binned DY production, which, while optimized to populate the VBF-sensitive phase space, yields lower statistics overall compared to the $p_{\mathrm{T}}(Z)$-binned DY sample used in the SM analysis. As a consequence, the sensitivity is limited by the available MC statistics. An update of the SMEFT analysis is foreseen to replace the jet-binned production with the higher-statistics $p_{\mathrm{T}}(Z)$-binned sample, thereby reducing the dominant MC statistical uncertainty and improving the expected constraints on the Wilson coefficients.

Theoretical Systematic Uncertainties Theoretical nuisances affecting the DY background, in particular the DY QCD scale and parton shower final state radiation (PS FSR), are found to be significantly constrained by the fit thanks to the high statistical power of the SR and the DY PU CR at low DNN score.
As an illustrative example, in the 2018 dataset the post-fit impact of the PS FSR on the DY is larger than the statistical uncertainty of the expected data in both the SR $e^+e^-$and DY-PU CR $\mu^+\mu^-$regions (see Table 2.14). Similarly, the DY QCD scale uncertainty is constrained in multiple high-statistics regions, including SR e^+e^-, SR $\mu^+\mu^-$, and DY-PU CR $\mu^+\mu^-$, as shown in Table 2.15.
This behavior confirms that, while MC statistics remain a limiting factor, the VBF-Z analysis has sufficient statistical power in the control and signal regions to constrain major theoretical uncertainties associated with the DY background.

Summary of Results The one-dimensional expected limits on the Wilson coefficients c_W, c_{HWB}, and c_{HDD} obtained from the Full Run 2 fit are summarized in Table 2.16 using both the linear and quadratic component of the EFT expansion. These intervals represent the most stringent constraints on these operators in the VBF-Z channel to date. A comparison with other public analyses targeting the same operators is provided in Table 2.17, highlighting the complementary sensitivity of the VBF-Z process within the broader SMEFT program. As a direct comparison with the VBF-Z ATLAS analysis [90],

Region	Exp. num. of events	$1/\sqrt{N}$ (%)	Syst. Unc. Cumulative Effect (%)
sr e^+e^-	232687.17	0.207	0.305
sr $\mu^+\mu^-$	393629.77	0.159	0.055
DY-PU CR e^+e^-	219968.39	0.213	0.104
DY-PU CR $\mu^+\mu^-$	394241.86	0.159	0.265

Table 2.14: Statistical uncertainty and PS FSR effect on the DY in the different regions. The systematic uncertainty cumulative effect is the ratio of the integral of the variation with the integral of the nominal sample in a specified region.

Region	Exp. num. of events	$1/\sqrt{N}$ (%)	Syst. Unc. Cumulative Effect (%)
sr e^+e^-	232687.17	0.207	0.302
sr $\mu^+\mu^-$	393629.77	0.159	0.193
DY-PU CR e^+e^-	219968.39	0.213	0.103
DY-PU CR $\mu^+\mu^-$	394241.86	0.159	0.342

Table 2.15: Statistical uncertainty and QCD scale effect on the DY in the different regions. The systematic uncertainty cumulative effect is the ratio of the integral of the variation with the integral of the nominal sample in a specified region.

this work provides more stringent expected limits on c_{HWB} while loosing some sensitivity to c_W. The ATLAS collaboration EFT limits on EW Zjj are also obtained using only the linear component of the EFT expansion. The comparison of the expected limits using only the linear component is reported in Table 2.18.

Detailed Results for c_W The results for the Wilson coefficient c_W are presented in terms of both the likelihood profile and the corresponding nuisance parameter impacts. Figure 2.44 shows the one-dimensional likelihood scan for c_W obtained from the Full Run 2 fit using the Asimov dataset. The scan is symmetric and smooth around the best-fit value, indicating a well-behaved likelihood and stable profiling of the associated nuisance parameters. The fitted value of c_W is compatible with the Standard Model expectation within uncertainties, and the curvature of the likelihood function defines the one- and two-standard-deviation confidence intervals reported in the global summary.
To better understand the dominant sources of uncertainty affecting this coefficient, Table 2.19 lists the most relevant contributions to the total uncertainty of c_W.

Wilson Coefficient	68% Confidence Interval Expected [TeV^{-2}]	95% Confidence Interval Expected [TeV^{-2}]
c_W	[-0.17, 0.17]	[-0.38, 0.32]
c_{HWB}	[-0.90, 0.70]	[-1.90, 1.50]
c_{HDD}	[-4.60, 2.20]	[-6.20, 3.80]

Table 2.16: Summary of expected limits with Full Run 2 data

Analysis	c_W 95% C. I. [TeV^{-2}]	c_{HWB} 95% C. I. [TeV^{-2}]	c_{HDD} 95% C. I. [TeV^{-2}]
This Work: EW Zjj Run II	-0.381,0.323	-1.901,1.504	-6.202,3.803
CMS VBS WV $\rightarrow$ lνqq	-0.349,0.345	-4.347,4.349	-
SSWW 2018	-0.396,0.382	-	-
CMS VBF+VH+H$\rightarrow$4l	-	-0.310,0.420	-4.600,1.060
CMS VBS SSWW+τ_h	-0.842,0.818	-49.600,110.000	-31.400,45.500
CMS VBS VV$\rightarrow$ 4q	-0.258,0.256	-	-
CMS $W^{\pm}\gamma$	-0.062,0.052	-	-
ATLAS H$\rightarrow$ WW	-	-1.200,1.100	-
ATLAS EW Zjj	-0.310,0.290	-3.110,2.100	-
ATLAS pp$\rightarrow$ 4l	-	-0.210,0.200	-0.460,0.450
CMS WZ$\rightarrow$ 3lν	-0.090,0.090	-	-
CMS WW$\rightarrow$ 2l2ν	-0.188,0.188	-	-
CMS WV$\rightarrow$ lν qq	-0.110,0.111	-	-
CMS EW Zjj: 2016 Data	-0.257,0.250	-	-
CMS EW Wjj: 2016 Data	-0.174,0.174	-	-

Table 2.17: Summary and comparison of expected limits with other public analyses.

Analysis	c_W 95% C. I. [TeV^{-2}]	c_{HWB} 95% C. I. [TeV^{-2}]	c_{HDD} 95% C. I. [TeV^{-2}]
This Work: EW Zjj Run II	-0.392,0.335	-1.520,1.890	-9.100,12.305
ATLAS EW Zjj	-0.300,0.300	-2.450,2.450	-

Table 2.18: Summary and comparison of expected limits with ATLAS using the linear component only.

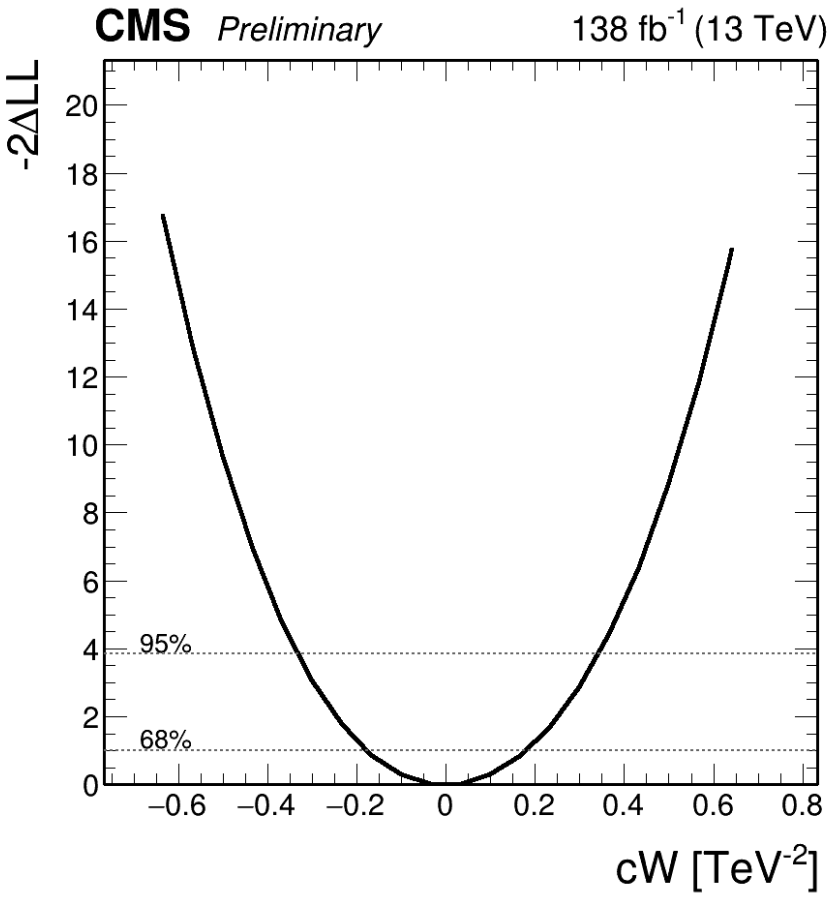

Figure 2.44: Expected likelihood scan of the POI c_W for Full Run 2 fit

c_W bestfit:	$0.00^{+0.18}_{-0.17}$
Unc. type	**Rel. (abs.) unc. to c_W**
Overall statistical unc.	63% $(^{+0.10}_{-0.10})$
Overall systematic unc.	77% $(^{+0.11}_{-0.12})$
Unc. in the data-driven background prediction	
Bin-by-bin unc.	93% $(^{+0.11}_{-0.11})$
Shape unc.	38% $(^{+0.04}_{-0.04})$
Unc. in the signal prediction	
Theory	79% $(^{+0.04}_{-0.03})$
Jet energy scale/resolution	46% $(^{+0.02}_{-0.02})$
B-tagging unc.	24% $(^{+0.01}_{-0.01})$
Other sources	30% $(^{+0.02}_{-0.01})$

Table 2.19: Major source of uncertainty in the measurement of c_W. Both relative contribution and absolute uncertainty are shown. The post-fit uncertainty is first separated into statistical and systematic components. The systematic component is further divided into uncertainties affecting the data-driven background prediction, primarily driven by the limited sample size, and those associated with the simulation.

Detailed Results for c_{HWB} The results for the Wilson coefficient c_{HWB} are presented through its one-dimensional likelihood profile and the corresponding impact tables. Figure 2.45 shows the likelihood scan for c_{HWB}, obtained from the Full Run 2 fit with the Asimov dataset. The likelihood is well-behaved and displays a symmetric profile around the minimum, with no visible secondary structures or instabilities. The best-fit value of c_{HWB} is consistent with the Standard Model prediction, and the extracted confidence intervals confirm the sensitivity of the VBF Z channel to this operator, which primarily modifies the electroweak gauge-boson mixing.
To assess the main sources of uncertainty impacting this result, Table 2.20 lists the major contributions to the total uncertainty of c_{HWB}. The theory uncertainty constitutes the largest single contribution, reflecting the sensitivity of this operator to higher-order corrections. This is due to the fact that c_{HWB} exhibits milder shape effects compared to c_W, leading to a greater reliance on the overall signal normalization and thus amplifying the impact of theoretical modeling uncertainties.

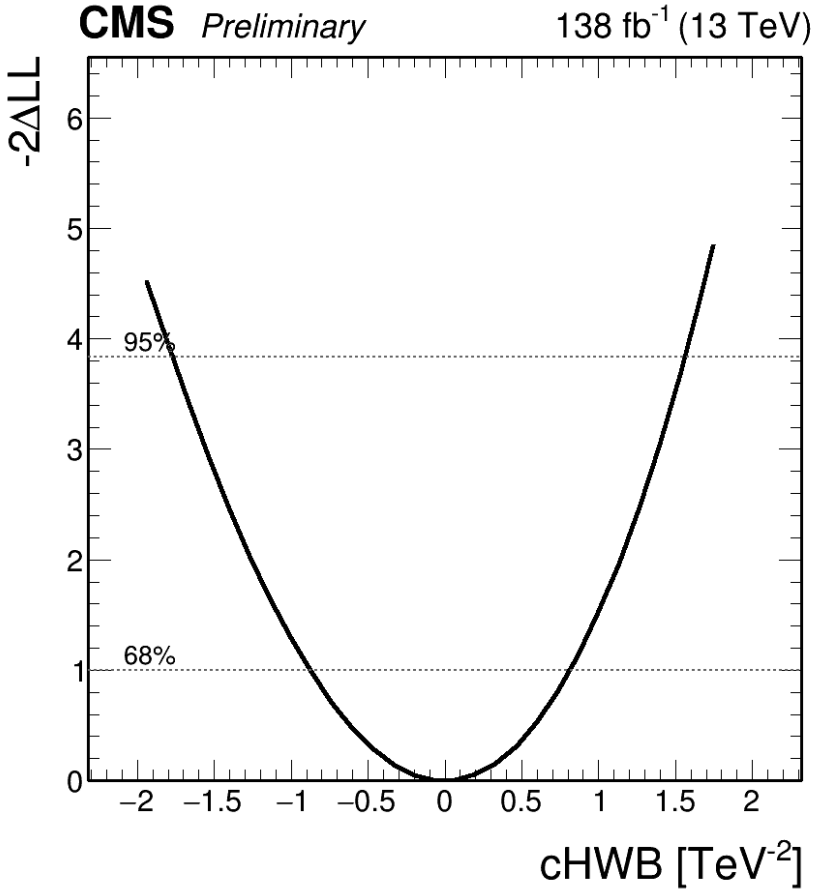

Figure 2.45: Expected likelihood scan of the POI c_{HWB} for Full Run 2 fit

c_{HWB} bestfit:	$0.00^{+0.81}_{-0.87}$
Unc. type	**Rel. (abs.) unc. to c_{HWB}**
Overall statistical unc.	21% $(^{+0.19}_{-0.21})$
Overall systematic unc.	98% $(^{+0.91}_{-0.91})$
Unc. in the data-driven background prediction	
Bin-by-bin unc.	60% $(^{+0.52}_{-0.58})$
Shape unc.	80% $(^{+0.75}_{-0.71})$
Unc. in the signal prediction	
Theory	91% $(^{+0.69}_{-0.64})$
Jet energy scale/resolution	33% $(^{+0.23}_{-0.24})$
B-tagging unc.	13% $(^{+0.09}_{-0.10})$
Other sources	20% $(^{+0.14}_{-0.15})$

Table 2.20: Major source of uncertainty in the measurement of c_{HWB}. Both relative contribution and absolute uncertainty are shown. The post-fit uncertainty is first separated into statistical and systematic components. The systematic component is further divided into uncertainties affecting the data-driven background prediction, primarily driven by the limited sample size, and those associated with the simulation.

Detailed Results for c_{HDD} The likelihood profile and impacts of uncertanties for the Wilson coefficient c_{HDD} are shown in Figure 2.46 and Table 2.21, respectively. The likelihood scan, obtained from the Full Run 2 Asimov fit, exhibits a smooth and asymmetric behavior around the minimum, confirming the numerical stability of the fit and the absence of degeneracies with other operators. The asymmetric nature of the likelihood is due to the quadratic term in the EFT expansion becoming dominant at large absolute values of c_{HDD}. The best-fit value is compatible with the Standard Model expectation within uncertainties. Given that c_{HDD} modifies the derivative structure of the Higgs kinetic term and thus affects the electroweak boson propagators, this result provides a direct constraint on new physics effects influencing the Z-boson kinematics in VBF Z production.
To identify the dominant sources of uncertainty, Table 2.21 reports the biggest uncertainties affecting c_{HDD}.

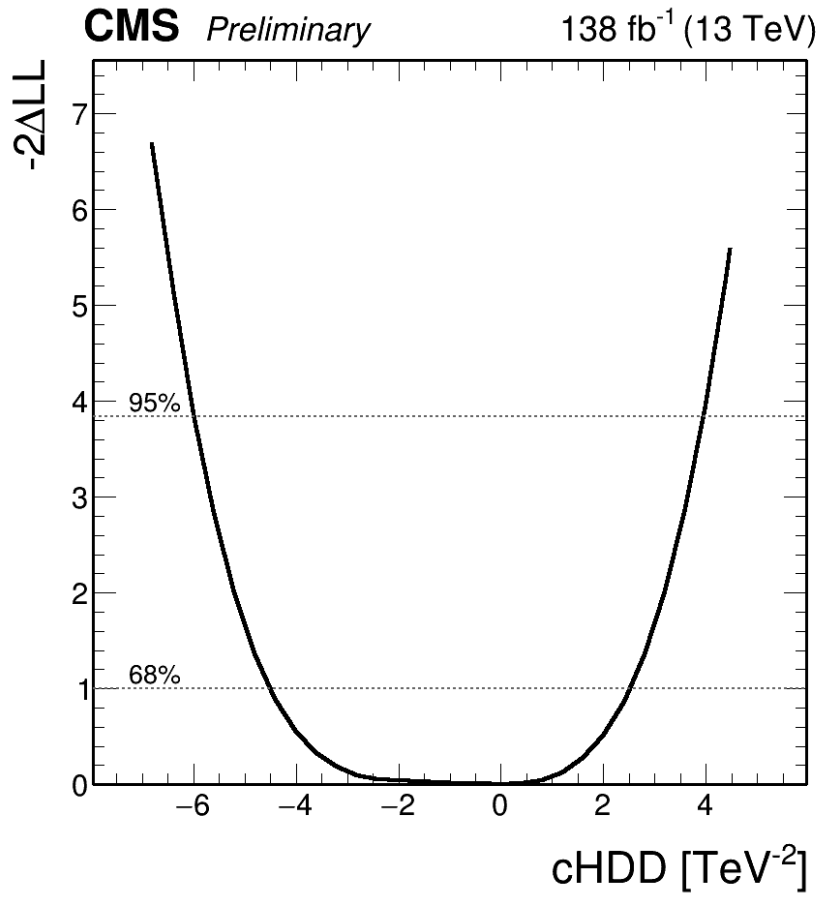

Figure 2.46: Expected likelihood scan of the POI c_{HDD} for Full Run 2 fit

c_{HDD} bestfit:	$0.00^{+2.52}_{-4.49}$
Unc. type	**Rel. (abs.) unc. to c_{HDD}**
Overall statistical unc.	37% $(^{+0.78}_{-2.58})$
Overall systematic unc.	92% $(^{+3.12}_{-4.38})$
Unc. in the data-driven background prediction	
Bin-by-bin unc.	48% $(^{+1.53}_{-2.09})$
Shape unc.	88% $(^{+2.72}_{-3.85})$
Unc. in the signal prediction	
Theory	87% $(^{+2.49}_{-3.21})$
Jet energy scale/resolution	38% $(^{+0.91}_{-1.61})$
B-tagging unc.	12% $(^{+0.33}_{-0.47})$
Other sources	26% $(^{+0.51}_{-1.30})$

Table 2.21: Major source of uncertainty in the measurement of c_{HDD}. Both relative contribution and absolute uncertainty are shown. The post-fit uncertainty is first separated into statistical and systematic components. The systematic component is further divided into uncertainties affecting the data-driven background prediction, primarily driven by the limited sample size, and those associated with the simulation.

2D Limits

The two-dimensional likelihood scans provide information not only on the individual sensitivity to each operator but also on the correlations among them. Figure 2.47 shows the expected contours at 68% and 95% confidence level for the pairs (c_W, c_{HDD}), (c_W, c_{HWB}), and (c_{HWB}, c_{HDD}).
The contours for the (c_W, c_{HDD}) and (c_W, c_{HWB}) planes appear respectibly rectangular, and circular, indicating that these pairs of operators are largely uncorrelated in the VBF-Z channel.
In contrast, the (c_{HWB}, c_{HDD}) scan exhibits a tilted, elongated contour, reflecting a stronger correlation between these two operators, as expected from the c_{HWB} and c_{HDD} mixture term that had an impact as large as the quadrature term of c_{HDD} in Figure 2.34. The reason of such a high correlation lies in the nature of the two operators that affects the normalization rather then the shape of the distributions.
Overall, the 2D scans confirm that the VBF-Z analysis is capable of disentangling the impact of different EFT operators, with only moderate correlations observed. This ensures that the limits obtained are robust and that the sensitivity of the measurement is not dominated by degeneracies among the chosen operators.

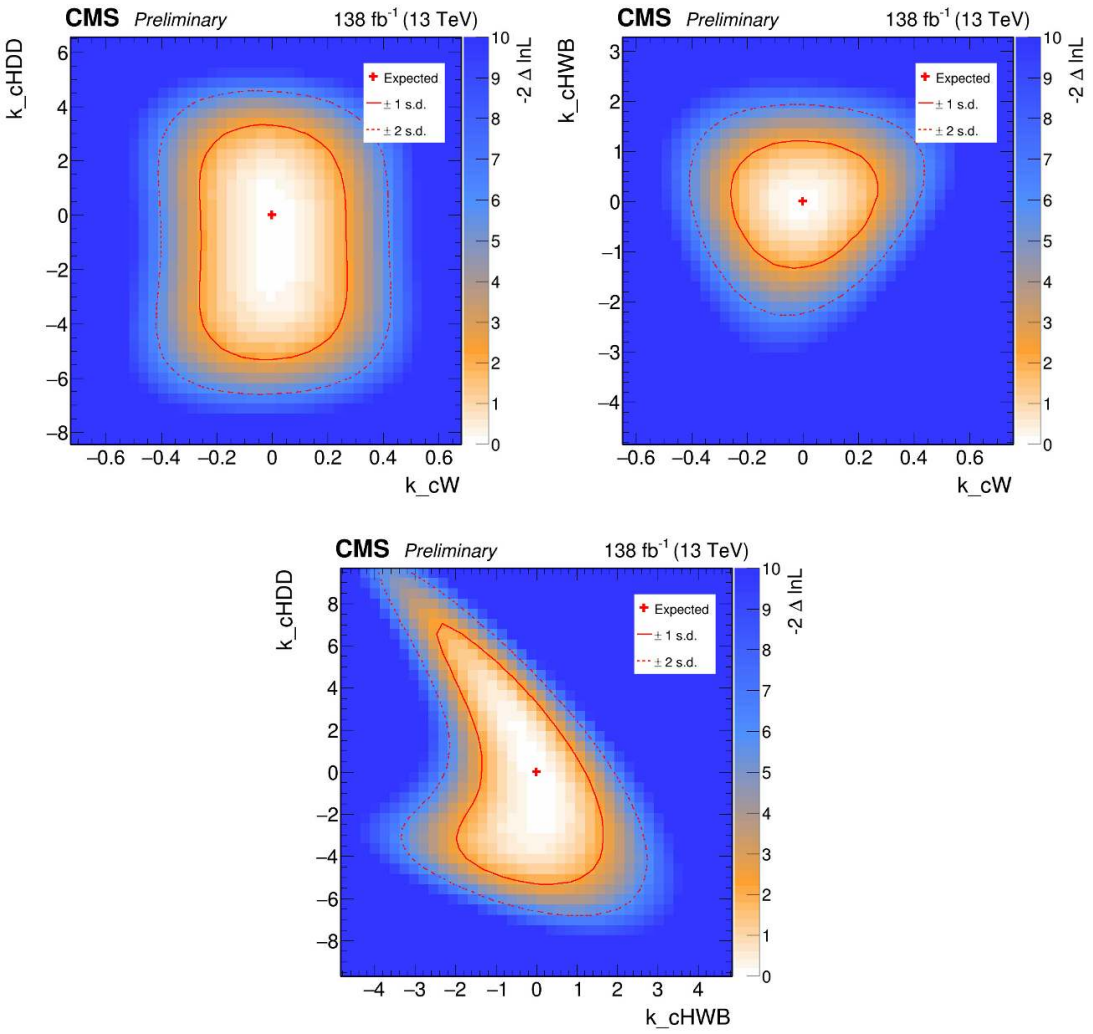

Figure 2.47: Expected likelihood scan for Full Run 2 fit of the POIs: c_W, c_{HDD} (top left), c_W, c_{HWB} (top right), and c_{HWB}, c_{HDD} (bottom). The 68% and 95% CL contours are shown.

2.15 Summary

In this analysis, we measured the inclusive cross section for vector boson fusion (VBF) production of the Z boson using the full Run 2 dataset collected by the CMS experiment at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 138 fb^{-1}. The inclusive signal strength was extracted with an expected precision of 15.5% ($\mu = 1.0^{+0.17}_{-0.14}$), where theoretical uncertainties—particularly those affecting the overall signal normalization—constitute the dominant source of uncertainty.

In addition to the inclusive cross-section, we measured fiducial and differential ones for a set of key observables that characterize both the VBF topology and Z boson kinematics. The signal strength in the fiducial phase space was extracted with an expected uncertainty of 11.5% ($\mu = 1.0^{+0.13}_{-0.10}$), with experimental uncertainties dominating the total uncertainty in this case. The differential distributions were unfolded to particle level in a fiducial region closely matching the reconstructed selection, and validated with detailed

robustness tests of the background modeling.
The full list of observables unfolded in this analysis includes:

- dijet invariant mass (m_{jj}),
- pseudorapidity separation between the two leading jets ($\Delta\eta_{jj}$),
- azimuthal angle separation between the two leading jets ($\Delta\phi_{jj}$),
- transverse momentum of the dijet system (p_T^{jj}),
- azimuthal angle separation between the two leptons ($\Delta\phi_{\ell\ell}$),
- transverse momentum of the dilepton system ($p_T^{\ell\ell}$).

Particular care was taken to ensure the accurate modeling of the dominant background process, Drell-Yan + jets, especially in the high-purity regions relevant to the VBF selection. Control regions and masked bin validation studies demonstrated that the fit strategy provides sufficient flexibility to absorb shape mismodeling and constrain this background effectively.
Beyond the Standard Model interpretation, the same analysis framework was extended to perform an Effective Field Theory study, probing deviations in the electroweak Zjj production through dimension-6 operators in the Warsaw basis. The resulting constraints on the most sensitive operators—particularly c_{HWB} and c_{HDD}, will be included in the CMS EFT combination of VBF, VBS, diboson and triboson measurements, where VBF-Z plays a key role in improving the precision on the bosonic sector of the theory.
Overall, the results enhance our understanding of electroweak Z+jets production in the VBF topology, provide high-precision tests of the Standard Model, and place stringent constraints on potential deviations from it in a model-independent EFT framework.

Chapter 3

Data acquisition system for the CMS MIP Timing Detector

While the results presented in the previous chapter focus on exploiting the full Run 2 (138 fb^{-1}) dataset for precision measurements and EFT interpretations in the VBF-Z channel, the High-Luminosity LHC (HL-LHC) will open a new era of physics opportunities, given the much higher integrated luminosity of 3000 fb^{-1}, and experimental challenges. The unprecedented instantaneous luminosity will lead to an average of 140-200 simultaneous proton-proton interactions per bunch crossing, drastically increasing the level of pileup and complicating event reconstruction in all CMS analyses, including those based on vector-boson fusion and scattering topologies.
To mitigate these effects, CMS will be equipped with the MIP Timing Detector (MTD), designed to provide precise time-of-arrival measurements for charged particles with a resolution of about 30-40 ps. Covering the pseudorapidity range up to $|\eta| < 3$, the MTD will add a fourth dimension to event reconstruction, enabling the assignment of tracks and vertices to individual interactions in time. This capability will substantially reduce pileup contamination in jets, missing transverse momentum, and vertex association.
For processes such as VBF and VBS, characterized by forward jets and electroweak topologies particularly sensitive to pileup, the MTD will play a crucial role in improving jet and vertex reconstruction efficiency, enhancing signal purity, and ultimately increasing the sensitivity of these measurements in the HL-LHC phase. The following sections describe the design, development, and validation of the MTD data acquisition (DAQ) system, with a focus on the Barrel Timing Layer (BTL).

3.1 The LHC High Luminosity phase and its challenges

To fully exploit the physics potential of the Large Hadron Collider (LHC), CERN launched the High-Luminosity LHC (HL-LHC) project in late 2010. The upgrade has two main objectives: to achieve a peak instantaneous luminosity of 5-7.5 $\times 10^{34}, \text{cm}^{-2}\text{s}^{-1}$ and to deliver an annual integrated luminosity of about 250 fb^{-1}, ultimately accumulating $\sim$3000 fb^{-1} over 12 years of operation [104], as shown in Figure 3.1. During this period, the accelerator will also operate at its nominal design energy of 7 TeV per beam. A dataset of such unprecedented size will open a new era of precision studies across all sectors of the Standard Model (SM).

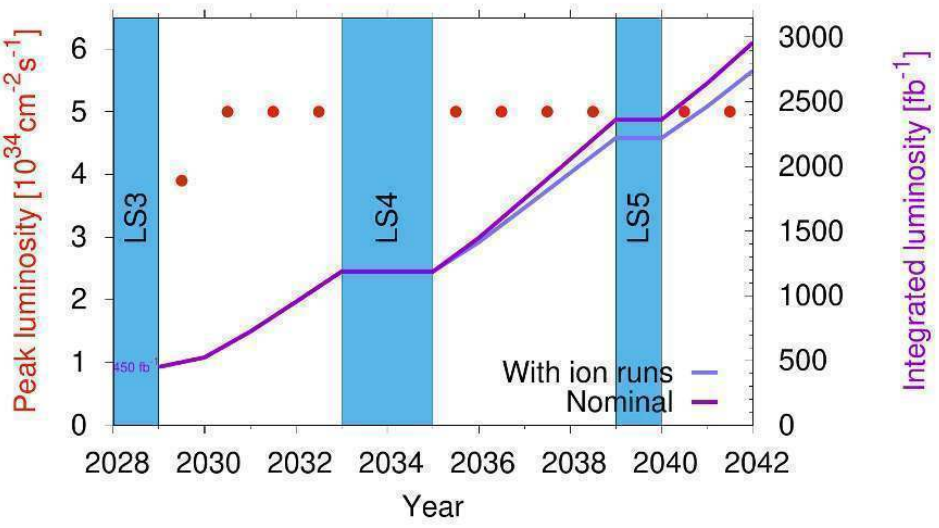

Figure 3.1: Projected integrated luminosity delivery by the LHC and HL-LHC through the year 2042, including the Long Shutdown 3 (LS3) period for the HL-LHC upgrade.

At the forefront of this program is the Higgs sector. Precise measurements of Higgs boson properties and their relation to the electroweak symmetry breaking (EWSB) mechanism form a central goal of the HL-LHC physics case [105]. The key Higgs couplings, together with the boson width Γ_h, are expected to be determined with percent-level accuracy [106]. Higgs boson pair production, which provides direct sensitivity to the trilinear self-coupling λ, will be probed with a projected combined significance of about 4 standard deviations under the SM hypothesis by the ATLAS and CMS collaborations.
Beyond the Higgs, the HL-LHC will enable unprecedented scrutiny of the EWSB mechanism at energy scales above m_h. Processes involving pairs or triplets of electroweak gauge bosons will be measured with high precision [107]. In particular, the combined ATLAS and CMS analysis of same-sign WW scattering is expected to establish the longitudinal component (WLWL) with a significance exceeding 3 standard deviations (estimated with 2000 fb^{-1}), thereby probing the subtle interplay between Higgs-mediated and purely gauge-mediated contributions. The production of three massive gauge bosons should be accessible with $\sim$20% precision and with more than 3 standard deviations significance in most final states.
Several fundamental SM parameters will also benefit from the HL-LHC dataset. The effective weak mixing angle $\sin^2\theta_{\text{eff}}$, the W boson mass m_{W}, and the top-quark mass m_{top} are expected to be measured with unprecedented accuracy, at the per-mille level. The statistical power of the dataset will further strengthen searches for physics beyond the Standard Model (BSM), including dark matter and supersymmetric candidates [108]. Finally, the precision of parton distribution functions (PDFs) will be significantly enhanced. Studies of processes involving jets, top quarks, photons, and electroweak bosons at the HL-LHC are projected to reduce PDF uncertainties by a factor of 2-4 [107].
To deliver the full physics potential of the HL-LHC programme, the upgraded detectors must maintain efficient performance in extreme conditions: they will need to operate reliably under high pile-up, withstand intense radiation over 12 years of running, and handle significantly larger data throughput from the detector to long-term storage. At the HL-LHC design luminosity, each bunch crossing will contain on average about 140 interactions, reaching up to 200 in the ultimate scenario of $7.5 \times 10^{34}, \text{cm}^{-2}\text{s}^{-1}$. The

additional tracks and energy deposits from these simultaneous collisions complicate the reconstruction of the hard-scattering vertex and increase the probability of false triggers. To meet these challenges, all LHC experiments will undergo major upgrades, incorporating advanced detector components and state-of-the-art technologies to mitigate radiation damage and cope with unprecedented pile-up levels. On the computing side, trigger rates and data volumes are expected to rise by more than an order of magnitude. These demands will be met through improvements in software scalability, efficiency, and performance, as well as by exploiting progress in heterogeneous computing architectures, storage systems, and network infrastructure [109].

3.2 CMS upgrades for the HL-LHC

The Phase-II upgrade programme of CMS is designed to ensure that the detector continues to deliver the same level of performance achieved during the first years of LHC running, while operating under the much harsher conditions of the HL-LHC. The central goals are to maintain high efficiencies, precise measurements, and robust background rejection for all physics objects employed in data analyses [110]. The main driving factors behind the upgrades are the cumulative effects of radiation damage and the challenges posed by extreme pileup, both of which could otherwise compromise the ability of CMS to operate effectively as the integrated luminosity approaches 3 ab^{-1}.
Radiation effects manifest differently across the various sub-detectors. In silicon-based tracking detectors, radiation induces lattice defects that reduce charge collection efficiency and significantly increase the depletion voltage required to operate the sensors. Eventually, the voltages become unmanageable, forcing operation under partial depletion and thus leading to a reduced signal yield. In the calorimeter systems, radiation alters the optical properties of the active materials: for $PbWO_4$ scintillating crystals and for plastic scintillators with wavelength-shifting fibers, the transmission of scintillation light deteriorates, with signal losses reaching up to 90%.
The second major challenge is the high number of simultaneous proton-proton interactions. At design conditions, up to 200 pileup events per bunch crossing are expected, comprising both in-time pileup and out-of-time (OOT) pileup. These additional collisions degrade physics performance in several ways. In the tracker, extra hits can result in mis-reconstructed or mis-identified tracks, complicating the reconstruction of isolated leptons that are crucial for electroweak measurements. In the calorimeters, pileup produces spurious energy deposits that are difficult to distinguish from those of the primary hard scatter. To address this, the HL-LHC trackers will rely on sensors with much finer granularity, enabling the precise association of charged particles to their production vertices even in the presence of hundreds of interactions. Although calorimeters cannot themselves link energy deposits to interaction vertices, advanced reconstruction algorithms such as Particle Flow [111] can correlate calorimetric information with charged tracks, allowing pileup contributions to be statistically subtracted [112, 113].
The impact of OOT pileup is strongly dependent on the signal shaping time of the readout electronics. Sub-detectors with fast response times, such as the tracking system, are less sensitive because the shaping time is short compared to the 25 ns bunch spacing. In contrast, calorimeters, which require longer shaping times, suffer more from OOT contamination: signals from earlier or later bunches can overlap with the triggered pulse, as illustrated in the right panel of Figure 3.2. For these systems, techniques based on precise

timing and pulse-shape analysis are essential, and will play an increasingly important role in the HL-LHC era.

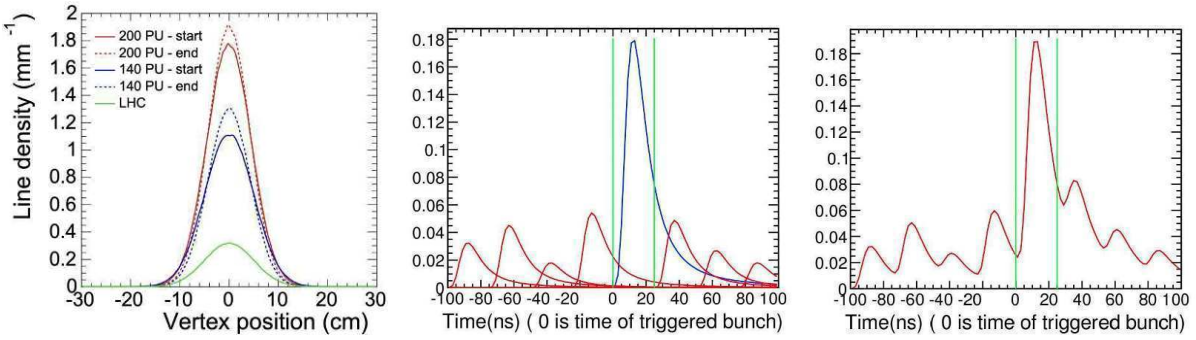

Figure 3.2: The left panel, adapted from [114], illustrates the distribution of reconstructed vertices and their z-coordinate along the beam axis at the LHC (green), corresponding to an average pileup of ~ 30, and at the HL-LHC for scenarios with 140 (blue) and 200 (red) simultaneous interactions. Solid curves denote the expected distributions at the start of a fill, while dashed curves refer to the end of a fill. The central and right panels, taken from [110], highlight the impact of out-of-time pileup. The central plot shows individual detector pulses as a function of time: the blue pulse corresponds to the triggered event (bounded by the green vertical lines), while the red pulses represent contributions from earlier and later bunch crossings. The right plot demonstrates that, when these pulses are summed, the triggered signal is contaminated by the additional energy of overlapping pileup events, requiring dedicated corrections.

Tracker

Both the strip and pixel tracking detectors must be entirely replaced prior to HL-LHC operation. For the pixel system, the accumulated radiation will progressively degrade the hit efficiency and spatial resolution [115]. The strip tracker will also suffer from rising depletion voltages and leakage currents. While cooling can partially mitigate the latter, the increasing bias requirements cannot be sustainably managed. Dedicated studies indicate that the current strip tracker becomes inoperable after collecting an integrated luminosity of about 1000 fb^{-1}. Radiation levels in the innermost regions are expected to reach fluences of $2.3 \times 10^{16}\,\mathrm{n_{eq}/cm^2}$ (1 MeV equivalent neutrons), making radiation damage a central design challenge for the Phase-II tracker.

To guarantee efficient performance in the presence of 140-200 pileup interactions per bunch crossing, the HL-LHC tracker will feature a fourfold increase in granularity compared to the current design. The acceptance is also extended in pseudorapidity, reaching up to $|\eta| = 4$, thereby matching the forward coverage of the calorimeter system, as illustrated in Figure 3.3.

Additional improvements include a reduction of the material budget and upgraded reconstruction algorithms to cope with higher track densities. Importantly, tracking information will be integrated into the Level-1 (L1) trigger at 40 MHz for tracks with $p_T > 2\,\mathrm{GeV}$, providing strong background suppression already at the hardware trigger level.

The upgraded tracker consists of two subsystems: the Inner Tracker (IT) and the Outer Tracker (OT). Located closest to the beamline, the IT must withstand the highest radiation doses. Two sensor technologies are used, both with an active area of 2500 $\mu\mathrm{m}^2$,

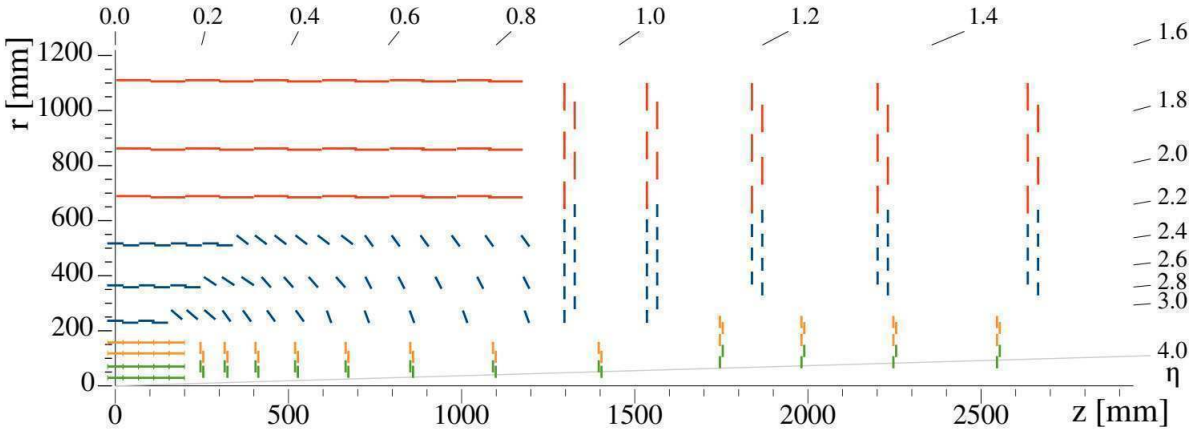

Figure 3.3: Schematic view of one quarter of the Phase-II tracker in the r-z plane. Inner tracker modules in green (orange) correspond to pixel sensors read out by two (four) chips, extending coverage up to $|\eta| = 4$. The outer tracker modules are shown as blue (PS) and red (2S) lines.

approximately one-sixth of the Phase-I pixels. One technology employs planar silicon sensors with cell sizes of $25 \times 100\,\mu\text{m}^2$ or $50 \times 50\,\mu\text{m}^2$. Alternatively, 3D silicon sensors, which offer superior radiation tolerance albeit at higher production cost, will be deployed in the most irradiated regions. Readout is performed with 65 nm CMOS ASICs (CMS Read Out Chips) developed within the RD53 collaboration [116], bump-bonded to single pixel sensors. This technology minimizes the number of service cables, contributing to a reduced material budget. The baseline IT layout is derived from the Phase-I geometry of four barrel layers and three endcap disks, extended with additional layers to achieve full coverage up to $|\eta| = 4$.

The OT begins at a radius of about 200 mm, where particle fluxes are an order of magnitude lower than in the IT, with fluences up to $1.5 \times 10^{15}\,\text{n}_{\text{eq}}/\text{cm}^2$. Covering up to $|\eta| = 2.4$, the OT employs two sensor types [110]. The so-called PS modules, used in the inner barrel layers and forward regions, combine pixel (p) and strip (s) sensors. PS-s modules consist of two strips, each 5 cm long with a pitch of 100 μm. PS-p modules are macro-pixels, 1.5 mm long with a 100 μm pitch, arranged into sensors of 10×5 cm^2. The remaining OT volume is instrumented with 2S modules, each comprising two strip sensors of 5 cm length and 90 μm pitch, covering an area of 10×10 cm^2.

Both PS and 2S modules are arranged in parallel pairs, known as p_T-modules, with a separation of 1.6-4 mm. Correlating hits across the paired sensors enables on-detector momentum discrimination, reducing the fraction of tracks transmitted to the trigger system to roughly 2.4% when applying a threshold of $p_T > 2\,\text{GeV}$.

Custom ASICs provide the readout: 2S modules are served by the CMS Binary Chip (CBC), a 130 nm CMOS ASIC that processes 254 strips, performs clustering, correlates hits, and delivers L1 data at 40 MHz. A secondary 65 nm CMOS chip aggregates outputs from eight CBCs with 160 Mb/s bandwidth. For PS modules, two types of front-end ASICs are used. The Short-Strip ASIC (SSA) reads PS-s modules, transmits unsparsified data to the Macro-Pixel ASIC (MPA) at 40 MHz, and performs functions analogous to the CBC. Each PS-p module is bump-bonded to 16 MPAs, each handling about 2000 connections, where the hit correlation between pixel and strip sensors is performed.

The OT will operate at $-20°$C at the start of HL-LHC running, with the capability of being cooled to $-30°$C as the integrated luminosity increases.

Muon Detectors

The present CMS muon system was originally designed to operate at instantaneous luminosities up to $10^{34}, \mathrm{cm}^{-2}\mathrm{s}^{-1}$. With the advent of the HL-LHC, several major upgrades are required to ensure its long-term performance. These include measures to limit radiation damage, new electronics to cope with the higher L1 trigger rates and increased latency, and an extension of the pseudorapidity coverage beyond $|\eta| = 2.4$. In the forward region $1.6 < |\eta| < 2.4$, where backgrounds are most severe and the bending power of the magnetic field is reduced, novel detector technologies will be deployed [110]. A key requirement is to preserve the efficiency of the L1 muon trigger while maintaining low p_T thresholds, thereby enabling sensitivity to a broad spectrum of physics channels. With the planned upgrades, the excellent performance demonstrated in Phase-I is expected to be sustained throughout the HL-LHC programme.

Longevity studies indicate that most existing chambers can continue efficient operation through Phase-II. The drift tube (DT) chambers in MB1 and MB4 are most susceptible to aging, as deposits forming on the anode wires could reduce their hit efficiency below 50% if the anticipated radiation dose is scaled by a safety factor of three. Nonetheless, because of the large redundancy in muon hit information, even such a reduction would only modestly affect the global efficiency: the barrel efficiency is projected to remain above 90%, and the overlap between barrel and endcap above 95%. To meet the demands of the HL-LHC trigger, the DT and cathode strip chamber (CSC) subsystems will see their front-end electronics replaced, while the RPC system will undergo an upgrade of its link electronics. Although the existing RPC front-end boards are sufficiently radiation-tolerant, the legacy link system is based on outdated and fragile components, necessitating replacement with modern electronics capable of fully exploiting the intrinsic timing resolution of the RPCs. To reinforce the system in the high-background forward region, Gas Electron Multiplier (GEM) detectors will be introduced. This technology has proven to be highly radiation resistant and capable of operating under HL-LHC conditions. Two new GEM stations, GE1/1 and GE2/1, each consisting of two layers of triple-GEMs, will provide additional redundancy in $1.6 < |\eta| < 2.4$. These detectors complement the CSCs, which alone are insufficient to reject low-p_T muons and punch-through segments. The GEMs increase the effective path length for track segments within each station to 28-55 cm, a factor of 2-5 improvement over the current layout [117]. In the third and fourth stations, CSCs will be paired with new RPC layers (RE3/1 and RE4/1), enhancing rejection of neutron-induced backgrounds. The upgraded RPCs feature excellent timing resolution, around 100 ps, making them valuable also for pileup mitigation.

Further extending the detector acceptance is justified by the improved forward pixel tracker coverage ($|\eta| < 4$) and by the replacement of the endcap calorimeters with more compact devices, which frees $\sim$30 cm along the z direction. In the forward pseudorapidity interval $2.1 < |\eta| < 2.8$, an additional six-layer GEM station (ME0) will be installed ahead of the ME1/1 chambers. This detector is designed to withstand extreme background hit rates, in the range of 10-100 kHz/cm^2.

The combined effect of the new RPC and GEM detectors will be a substantial increase in trigger efficiency. The fine spatial resolution of the GEM system will also enable precise determination of the transverse momentum at the L1 trigger level. This improvement

allows the use of a muon trigger threshold as low as 20 GeV while keeping the output trigger rate manageable at approximately 1 kHz. In the $2.1 < |\eta| < 2.4$ region, the redundant GEM layers reduce the trigger rate by an order of magnitude while boosting efficiency by about 10

Electromagnetic and hadronic calorimeters

The CMS electromagnetic calorimeter (ECAL) was originally designed for LHC Phase-I operation, with optimal performance expected up to an integrated luminosity of about 500 fb^{-1}. The scintillation mechanism of $PbWO_4$ crystals is intrinsically stable under irradiation, but high doses at the HL-LHC lead to the formation of crystal defects that reduce transparency and consequently the light yield. During Phase-I, radiation-induced damage from electromagnetic showers could self-recover at room temperature. However, at HL-LHC conditions, hadron-induced damage becomes significant, and the annealing temperature required to restore transparency ($\sim$350°C) cannot be reached in situ. These radiation effects deteriorate the ECAL energy resolution by impacting the stochastic, noise, and constant terms.

For the ECAL barrel (EB), the degradation is mitigated by lowering the operating temperature from 18°C to 9°C. After an integrated luminosity of 3000 fb^{-1}, the expected loss in light output compared to 2010 conditions ranges from 50% to 65%, depending on $|\eta|$ [118]. The EB photodetectors, avalanche photodiodes (APDs) operated with a gain of 50, are sensitive to γ-rays and hadrons, leading to increased leakage currents and electronic noise. Cooling to 9°C restores performance sufficiently for APDs to remain operational throughout the HL-LHC period.

Upgrades to the EB are primarily motivated by the new L1 trigger requirements: a latency of 12.5 μs and a maximum rate of 750 kHz. The front-end electronics will be replaced with faster components, enabling a 160 MHz sampling frequency with 12-bit resolution, albeit with a corresponding increase in data volume. This enhancement allows precision timing for spike rejection, pileup mitigation, and vertex association, and makes it possible to provide single-crystal information to the trigger rather than the current 5×5 trigger towers. After the upgrade, the EB is expected to achieve 99% spike rejection while maintaining nearly full electromagnetic signal acceptance at 3000 fb^{-1} and an average pileup of 140. In addition, reducing the electronic shaping time by a factor of two helps suppress radiation-induced noise, restoring Phase-I energy resolution.

The hadronic barrel calorimeter (HB) will also suffer from light loss at high integrated luminosity, affecting both electromagnetic object identification and neutral hadron reconstruction. To counter this, new silicon photomultipliers (SiPMs) with a photo-detection efficiency 2.5 times higher than the original design will be installed, ensuring functionality over the HL-LHC lifetime. Some scintillator tiles near the beam pipe will be replaced with doubly-doped plastic scintillator, which provides improved radiation stability albeit with lower light yield. Since the HB data contribute only about 10% to the total bandwidth, its off-detector electronics will be merged with those of the EB [110].

Radiation effects at high $|\eta|$ in the EB after 10 years of HL-LHC running are comparable to those already observed in the endcap calorimeters (EE) after only 30 fb^{-1}. The expected performance degradation in the EE is unacceptable, especially in forward regions where high QCD multijet rates under pileup conditions complicate electromagnetic object identification, e.g. in VBF and VBS topologies. For this reason, CMS will replace both the EE and the hadronic endcaps (HE) with a single high-granularity calorimeter

(HGC), capable of withstanding fluences up to $1.5 \times 10^{16}\,\mathrm{n_{eq}/cm^2}$ [119].
The HGC concept is based on developments from ILC/CALICE, featuring 3D shower imaging and precision timing. It consists of three parts: an electromagnetic section (EE) with tungsten/copper absorbers and silicon sensors of depth 1.5 interaction lengths (λ), followed by a hadronic section (FH) based on brass absorbers and silicon sensors with 3.5λ depth, and finally a brass-scintillator hadronic section (BH), located in a lower-radiation region and extending another 5λ. The detector provides coverage over $1.5 < |\eta| < 3.0$ and will be operated at $-30°$C to limit radiation-induced leakage currents in the silicon. The active HGC elements are 320 μm thick silicon sensors capable of minimum ionizing particle detection with high efficiency. In total, the calorimeter will comprise about six million readout channels. Monte Carlo studies of the HGC energy resolution predict a stochastic term between 20% and 24%, with a constant term below 1% for electrons in the energy range 5-500 GeV.

3.2.1 MIP Timing Detector

The CMS MIP Timing Detector (MTD) is a new subsystem that will be introduced for HL-LHC operations. Its role is to deliver precision timing for minimum ionizing particles (MIPs), thereby reducing the confusion from the $\mathcal{O}(200)$ pileup interactions per bunch crossing that obscure the reconstruction of the hard scatter. The expected MTD resolution is 30-40 ps at the start of Phase-II, degrading to about 50-60 ps after the collection of 3000 fb^{-1}. This precision is roughly a factor of five better than the intrinsic time spread of the collisions in a single bunch crossing, which originates from the longitudinal bunch size and corresponds to an rms of 180-200 ps. By exploiting this additional timing coordinate, tracks that originate near the same interaction point in space but differ in time can be disentangled. The MTD will provide such measurements in both barrel and endcap regions, with coverage up to $|\eta| = 3$.
The upgraded CMS tracker and pixel detectors, thanks to their finer granularity, will be able to resolve primary and pileup vertices for average pileup rates of around 140 interactions per crossing. However, at 200 simultaneous collisions the density of reconstructed vertices along the beamline becomes too high, and algorithms based solely on spatial separation, such as those used in Particle Flow (PF), begin to fail. Simulation studies show that performance degrades significantly when the vertex line density exceeds $\sim 1\,\mathrm{mm}^{-1}$: isolation for leptons is spoiled by nearby pileup tracks, while jet reconstruction and missing transverse momentum are also affected.
The addition of MTD information allows PF to incorporate a fourth dimension, time, effectively slicing the luminous region into windows of 30-40 ps. Each window contains on the order of 40-60 interactions, reducing the effective pileup per reconstruction step to levels comparable to LHC Run II. This effect is illustrated in Figure 3.4, which shows how vertex separation improves once timing is introduced. With this capability, CMS will retain Phase-I-like pileup mitigation performance even at HL-LHC conditions.
The physics impact of MTD is directly tied to improvements in time-aware reconstruction. For jets, the rejection of pileup-induced objects is expected to improve by 25-40%, depending on the achieved timing resolution, with significant benefits for VBF and VBS analyses. For missing transverse momentum, where a $\sim$15% resolution degradation is expected at 200 pileup, timing allows a recovery of 10-15%, reducing the tails above 130 GeV by roughly 40%. This directly translates into restored sensitivity for searches that rely on p_T^{miss}, such as supersymmetry.

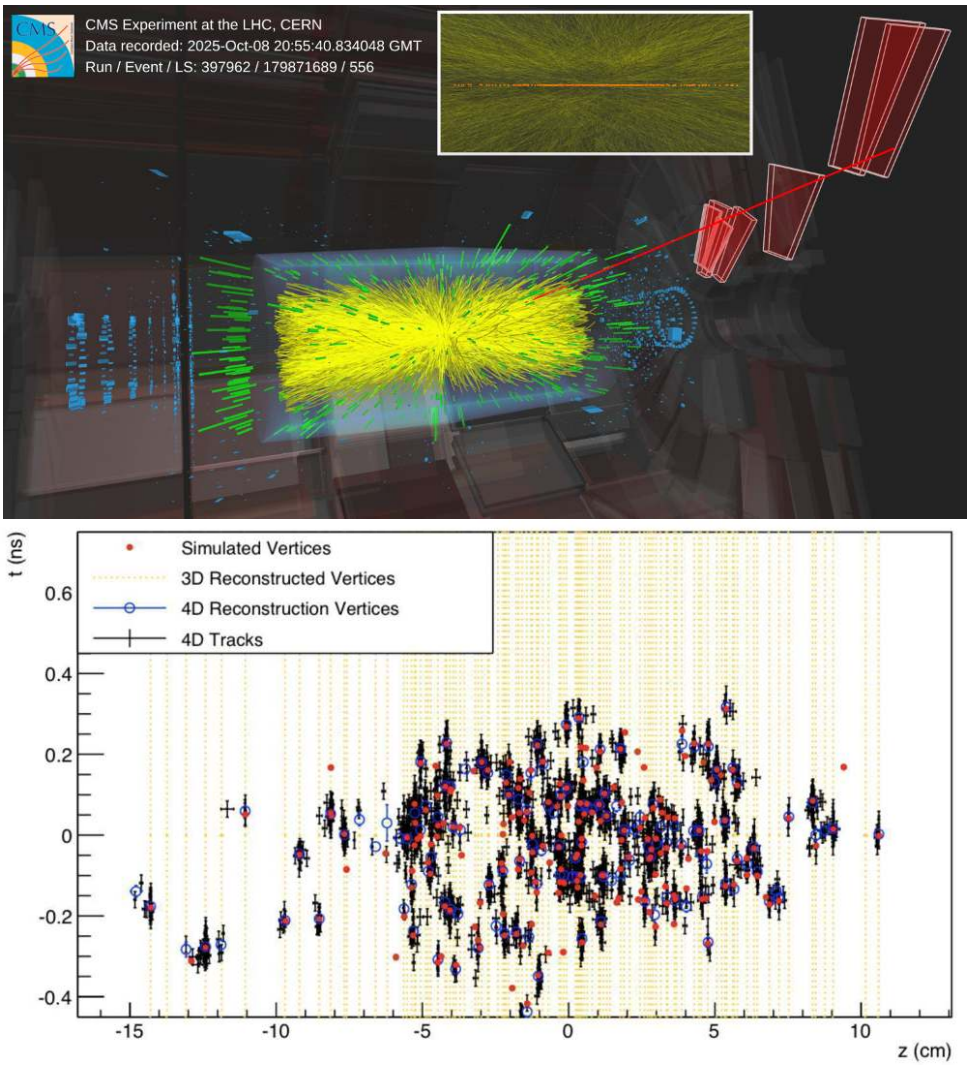

Figure 3.4: Top: LHC high PU test event ($<$ PU $>\sim$ 152) during Run 3, 206 reconstructed vertices [120]. Bottom: simulated and reconstructed vertices at 200 pileup. Without timing (yellow dashed lines), vertices overlap spatially if aligned along z. With MTD timing (blue points for tracks, black crosses for vertices), the spatial degeneracy is removed.

Heavy-flavour tagging, which relies on displaced secondary vertices, is also compromised by pileup tracks. Timing information reduces by about 30% the rate of falsely reconstructed secondary vertices. In di-Higgs searches, which exploit the $H \to b\bar{b}$ channel, improvements in b-tagging translate to a ∼20% gain in signal yield at 200 pileup, for constant backgrounds. Isolated lepton reconstruction also benefits: at 200 pileup, efficiency is increased by 3-4% in the barrel and 6-7% in the endcaps. Such improvements are crucial for precision electroweak studies, including channels like $H \to 4\ell$ and $H \to \tau\tau$.
The design of the MIP Timing Detector (MTD) is guided by both scientific goals and engineering constraints. From an engineering standpoint, the system must withstand the strong CMS magnetic field, ensure mechanical stability over the entire HL-LHC program, and integrate seamlessly into the existing detector layout within the upgrade schedule. On the scientific side, the central performance target is the delivered time resolution. Simulation studies indicate that a precision of 30-40 ps, attainable at the start of HL-LHC operation, is sufficient to mitigate the impact of pileup. Even after radiation-induced degradation, leading to a resolution of 50-60 ps at the end of Phase-II with 3000 fb^{-1} collected, the MTD will continue to provide the necessary performance. Since service channels for cooling, power distribution, and data transmission are already foreseen in the HL-LHC upgrade plan, the integration of the MTD into the CMS detector will ensure nearly hermetic coverage without major changes to the planned geometry.
The detector will be located in the radial gap between the outer tracker (OT) and the electromagnetic calorimeter barrel (EB), see Figure 3.5. To provide coverage up to $|\eta| = 3$, the system is divided into two regions: a cylindrical barrel timing layer (BTL) covering $|\eta| < 1.5$ and two endcap timing layers (ETL), placed as disks, covering $1.6 < |\eta| < 3.0$. The BTL will be mounted on the OT support tube, extending radially inward by about 40 mm, with its services (cables, cooling, fibers) integrated alongside those of the OT. The ETL will be positioned between the high-granularity calorimeter (HGC) and the tracker, with up to 45 mm of space available for installation. Due to the HGC geometry, ETL coverage will be limited to $|\eta| = 3$ rather than extending to the tracker's full $|\eta| = 4$. The MTD geometry is illustrated in Figure 3.6.

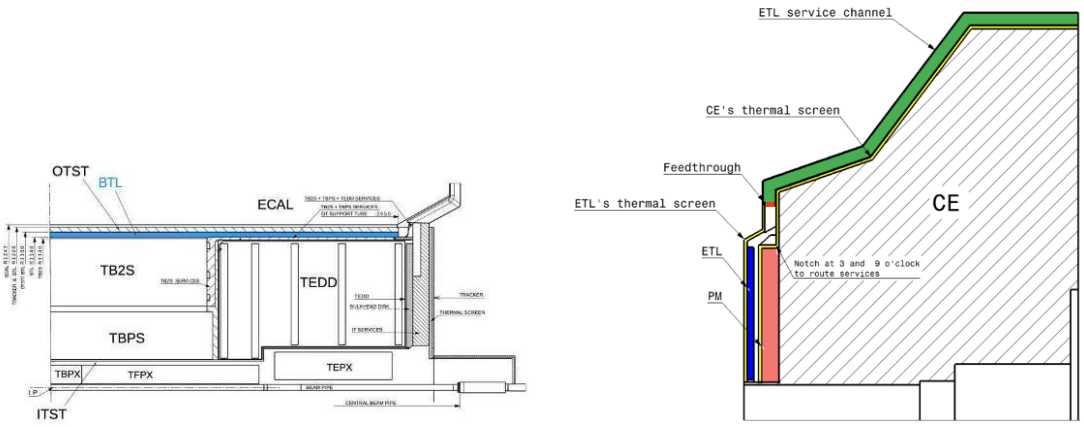

Figure 3.5: Schematic positioning of the CMS MIP Timing Detector (MTD). On the left BTL position in the barrel. On the right, ETL position in the endcap.

The technology choices for BTL and ETL are driven by radiation levels close to the interaction point. For an integrated luminosity of 3000 fb^{-1}, the expected fluence amounts to $1.9 \times 10^{14}\,\text{n}_{\text{eq}}/\text{cm}^2$ for the BTL region and $1.6 \times 10^{15}\,\text{n}_{\text{eq}}/\text{cm}^2$ for the ETL, i.e. about

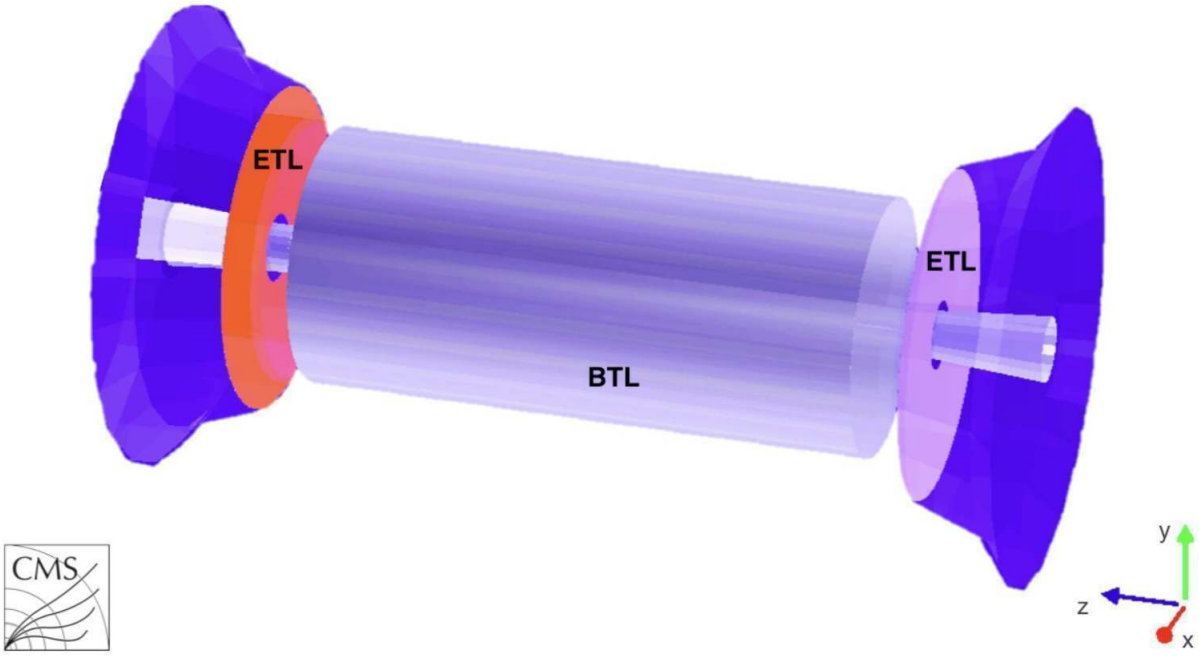

Figure 3.6: Schematic layout of the CMS MIP Timing Detector (MTD).

one order of magnitude higher. The ETL disks will be accessible during long shutdowns and can therefore be serviced if necessary, whereas the BTL will not be replaceable and must maintain functionality throughout the HL-LHC lifetime. For this reason, different sensor technologies are adopted: the ETL will use silicon-based Low Gain Avalanche Detectors (LGADs), while the BTL will be based on Cerium-doped Lutetium-Yttrium Oxyorthosilicate (LYSO:Ce) scintillating crystals read out by silicon photomultipliers (SiPMs). Both systems will operate at $-35°C$ with a dedicated cooling infrastructure to limit leakage currents and electronic noise. To counteract the degradation of BTL timing performance caused by radiation-induced increases in dark current, the sensors will undergo periodic annealing cycles during HL-LHC shutdowns, reaching temperatures up to $+40°C$ to $+60°C$.

3.3 The Barrell Timing Layer

Barrel Timing Layer

The Barrel Timing Layer (BTL) constitutes the central, hollow cylindrical section of the MTD, providing precise timing measurements up to $|\eta| = 1.48$. It will be mounted on the tracker support structure between the Outer Tracker (OT) and the Electromagnetic Calorimeter Barrel (EB). The BTL cylinder has a radial thickness of 40 mm, extending from 1144 mm to 1188 mm from the beam axis, and spans 5000 mm along the z direction. The BTL active medium consists of cerium-doped lutetium–yttrium orthosilicate (LYSO:Ce) scintillating crystals. Each crystal bar has a length of 57 mm, a width of 3.12 mm, and an average thickness of 3 mm, with a constant longitudinal thickness of 3.75 mm along the detector z-axis. When a minimum-ionizing particle (MIP) crosses a bar, scintillation light is produced and collected by silicon photomultipliers (SiPMs) optically coupled to

both ends of the crystal. Each module contains 32 SiPMs—two per crystal bar—operated in Geiger mode with a typical gain of 10^5.
Since the SiPM dark current doubles for every 10°C temperature increase, all modules are thermally anchored to an aluminum cooling plate maintained at −35°C by a CO_2 evaporative cooling system during HL-LHC operation. To further suppress the dark-count rate (DCR) and allow for controlled annealing cycles, thermo-electric coolers (TECs) are mounted on the rear side of the SiPM packages (see Section 3.3.2). The TECs can cool the sensors down to −45°C during operation and heat them up to +60°C during shutdown periods to mitigate radiation-induced damage. Because the SiPM overvoltage (OV) directly affects both the photon detection efficiency (PDE) and the DCR, the operating voltage will be gradually reduced throughout the detector lifetime to optimize the balance between signal amplitude and noise rate. The SiPM active area is specifically designed to match the crystal end face geometry to maximize photon collection efficiency, as detailed in Section 3.3.3.
A BTL detector module comprises two arrays of 16 LYSO:Ce bars and their corresponding SiPMs, all enclosed within a copper housing and interfaced with a custom front-end ASIC named TOFHIR. This ASIC digitizes the SiPM signals and provides precise timing measurements for all 32 channels. The timing is extracted by discriminating the leading edges of the analog pulses and measuring them with an integrated time-to-digital converter (TDC). Each BTL readout unit groups 3×8 detector modules, and six readout units form a *tray* covering half the length of the BTL cylinder. The full detector comprises 72 trays, totaling 165,888 LYSO:Ce bars and 331,776 SiPM channels.
The time resolution for a MIP traversing a BTL crystal depends primarily on the number of detected photons and on several independent contributions to the timing uncertainty:

$$\sigma_t^{\mathrm{BTL}} = \sigma_t^{\mathrm{clock}} \oplus \sigma_t^{\mathrm{digi}} \oplus \sigma_t^{\mathrm{ele}} \oplus \sigma_t^{\mathrm{photo}} \oplus \sigma_t^{\mathrm{DCR}}, \tag{3.1}$$

where $\sigma_t^{\mathrm{clock}}$ represents the CMS clock distribution jitter, σ_t^{digi} arises from the digitization process, σ_t^{ele} reflects the electronic noise, $\sigma_t^{\mathrm{photo}}$ accounts for stochastic fluctuations in photon arrival times, and σ_t^{DCR} denotes the noise induced by SiPM dark counts. As the photostatistics and noise terms are expected to dominate the overall timing performance, extensive optimization has been devoted to their minimization.
The contribution from photostatistics scales as $\sigma_t^{\mathrm{photo}} \sim \sqrt{\tau_r \tau_d / N_{\mathrm{phe}}}$, where τ_r and τ_d are the rise and decay times of the scintillation pulse, and N_{phe} is the number of detected photoelectrons. LYSO:Ce provides fast scintillation characteristics, with typical rise and decay times of approximately 100 ps and 40 ps, respectively. The number of detected photons per MIP depends on the deposited energy, the crystal light yield (LY), the light collection efficiency (LCE), and the SiPM PDE. In LYSO:Ce, the energy deposited by a MIP follows a Landau distribution with a most probable value (MPV) of 0.86 MeV/mm [114], corresponding to an average energy deposition of about 4.2 MeV per crystal. Assuming an LCE of 15% and a PDE of 20%, roughly 5,000 scintillation photons are detected per MIP. This yields an intrinsic time jitter from photostatistics of approximately 25–30 ps, consistent with the expected single-bar performance.

3.3.1 BTL performance at the beginning of HL-LHC

A comprehensive characterization campaign of LYSO:Ce scintillating crystals began in 2019, focusing on 57 mm-long bars produced by 12 different vendors, with the goal of assessing the key parameters governing timing performance [121]. The optical and

scintillation properties of the crystals were studied in detail, including measurements of light yield, decay time, and transmission, using unwrapped samples to avoid systematic effects from reflective coatings. Tests were performed before and after irradiation to emulate end-of-life conditions corresponding to an integrated luminosity of 3000 fb^{-1}, with the crystals operated at temperatures down to $-30°C$.
The study demonstrated that all LYSO:Ce samples from the tested vendors meet the performance requirements for precise timing measurements at the HL-LHC, exhibiting a variation in key characteristics within approximately 10%. After irradiation, a light-output reduction of about 9% was observed, which was fully recovered after five days of dark annealing at room temperature [122]. These results confirm the excellent radiation tolerance and operational stability of LYSO:Ce crystals under HL-LHC conditions.
Further investigations were performed at the Fermilab Test Beam Facility using a 120 GeV proton beam to evaluate the combined response of LYSO:Ce crystals and SiPMs in realistic conditions [123]. Non-irradiated BTL modules equipped with 57 mm-long LYSO:Ce bars of varying thicknesses (2, 3, and 4 mm) were tested. Light signals were read out by two SiPM technologies, developed by Hamamatsu Photonics (HPK) and Fondazione Bruno Kessler (FBK). The measurements revealed the dependence of the time resolution on the SiPM photon detection efficiency (PDE) and applied overvoltage (OV). For FBK sensors, the time resolution was primarily limited by photostatistical fluctuations, whereas for HPK devices a larger contribution from noise was observed. The measured resolution scales approximately with the inverse square root of the crystal thickness, yielding about 36 ps for 2 mm-thick bars and 26 ps for 4 mm-thick bars at an operating overvoltage of 6 V.
Overall, the Fermilab test-beam campaign validated the BTL sensor concept, demonstrating a uniform time response across the detector and confirming that the design can achieve a time resolution better than 30 ps for MIPs at the start of HL-LHC operation. These results are in excellent agreement with the original performance goals defined for the CMS MIP Timing Detector.

3.3.2 Strategies for performance recovery at end of life

Although the BTL meets the required timing performance at the beginning of HL-LHC operations, studies with irradiated modules—representative of end-of-life conditions—have shown a significant degradation of the timing resolution. After an integrated luminosity of 3000 fb^{-1}, the achievable resolution with the nominal design exceeds 100 ps, compared to the target value of approximately 60 ps. To recover the expected performance, several mitigation strategies have been developed, focusing on the optimization of the active component geometries, the operational parameters, and the front-end electronics. With the implementation of these improvements, detailed in the following subsections, the BTL is expected to retain its design specifications throughout the entire HL-LHC running period.

Thermo-electric coolers and thermal annealing

During HL-LHC operation, the BTL crystals and SiPMs will be subjected to a cumulative ionizing dose of approximately 50 kGy and a neutron fluence of the order of $10^{14}\,n_{eq}/cm^2$. The radiation response of SiPMs from the two manufacturers, Hamamatsu Photonics (HPK) and Fondazione Bruno Kessler (FBK), has been extensively studied and docu-

mented in Refs. [124, 125]. Both devices exhibit an increase in dark count rate (DCR) that scales linearly with neutron fluence up to about $5 \times 10^{13}\,\mathrm{n_{eq}/cm^2}$. Beyond this level—corresponding to mid-to-late HL-LHC conditions—additional degradation mechanisms become significant. At a fluence of $2 \times 10^{14}\,\mathrm{n_{eq}/cm^2}$, the breakdown voltage of both SiPM types increases, resulting in a reduction of the signal amplitude by roughly 25% at an overvoltage (OV) of 1 V. This observation suggests that the SiPM internal gain and photon detection efficiency (PDE) are reduced after heavy irradiation. Moreover, the large DCR contributes to increased power dissipation, which can lead to self-heating effects and alter the effective OV, further impacting performance.

The baseline BTL cooling system was originally designed to maintain a stable temperature of -35°C, since the DCR approximately doubles for every 10°C temperature increase, largely independent of the integrated radiation dose. By the end of HL-LHC operation, the expected DCR ranges between 10 and 100 GHz, depending on the operating temperature, OV, and annealing history—representing the dominant source of timing degradation.

Radiation-induced defects in silicon can, however, be partially mitigated through thermal annealing, which allows for the recovery of the lattice structure and a reduction of the DCR. During shutdown or technical stop periods, annealing at elevated temperatures can significantly reduce the DCR and restore timing performance. Extensive tests on irradiated SiPM prototypes have been performed to quantify the effects of annealing time and temperature [126]. In these studies, one SiPM array was annealed at +40°C for varying durations, while another was subjected to a 40-minute annealing cycle at +70°C. The results show that annealing at +40°C for about four days reduces the DCR by approximately 50%, as shown in the left panel of Figure 3.7. Furthermore, lowering the operating temperature by 10°C below the design value yields an additional DCR reduction by roughly another factor of two, as shown in the right panel.

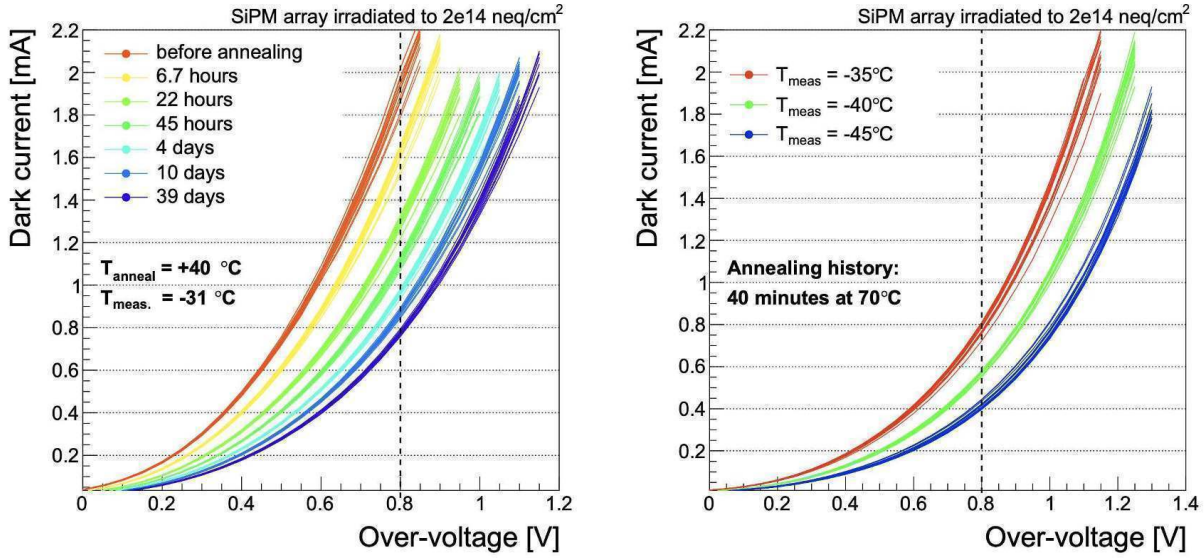

Figure 3.7: Left: dependence of the SiPM dark count rate (DCR) on the operating overvoltage for various annealing durations, from 0 h to 39 days. Right: DCR behaviour as a function of overvoltage for different operating temperatures at or below the nominal design temperature of -35°C, following a 40-minute annealing cycle at +70°C [126].

During HL-LHC technical stops, the BTL cold plate—housing the CO_2 cooling pipes—can only be warmed from $-35°C$ up to approximately $+10°C$, limiting the achievable annealing temperature. As the BTL will not be accessible for maintenance throughout the 12-year HL-LHC program, mitigation strategies must be implemented in the detector design itself. A twofold approach has been proposed to address this challenge. First, by lowering the operational temperature by an additional 10°C (down to $-45°C$) during data-taking, and second, by enabling localized heating of the SiPMs up to about $+60°C$ during technical stops to promote accelerated annealing. To achieve this, small thermo-electric coolers (TECs) are mounted directly on the back of the SiPM packages, serving as controllable thermal interfaces between the sensors and the copper housing.
The expected evolution of the DCR with luminosity for an SiPM operated at 1 V OV is shown in Figure 3.8, for various combinations of operating and annealing temperatures. This dual-temperature strategy is predicted to reduce the end-of-life DCR by more than a factor of five, achieving a value around 10 GHz at 3000 fb^{-1} for operating (annealing) temperatures of $-45°C$ ($+60°C$). As a result, the proposed TEC-based system is expected to preserve—and even enhance—the BTL timing performance throughout the full HL-LHC lifetime.

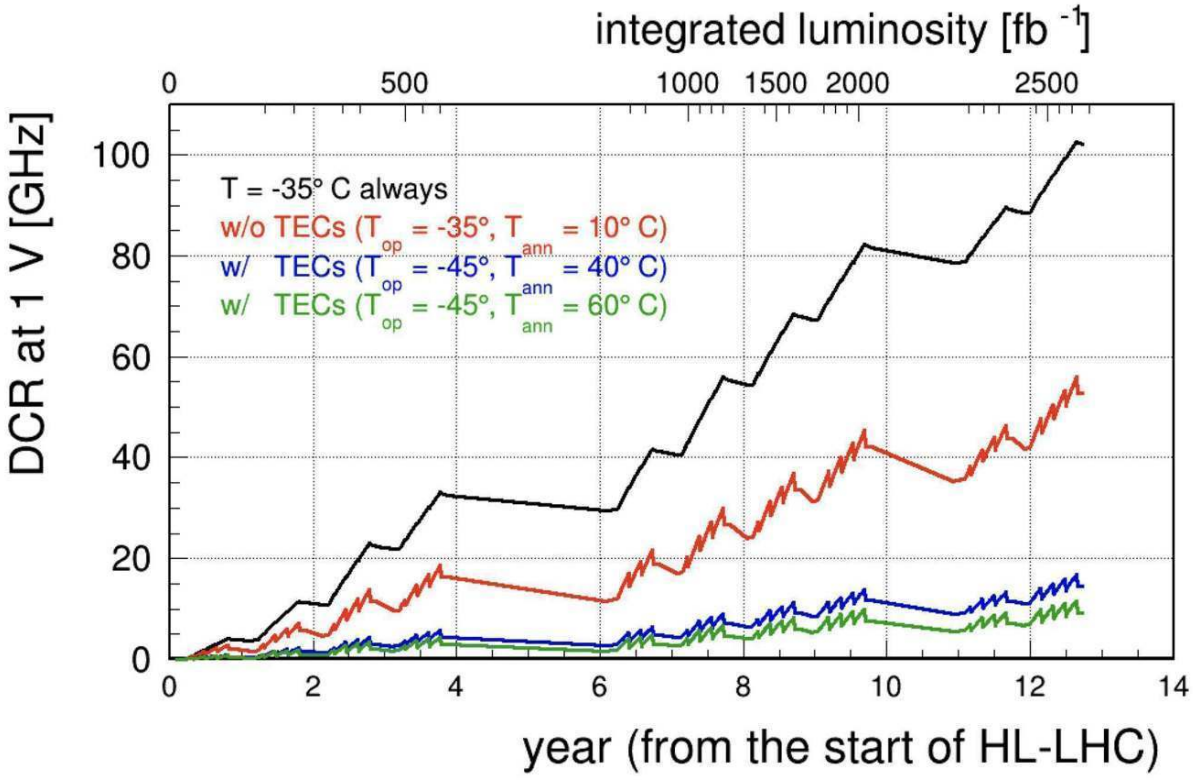

Figure 3.8: Expected SiPM dark count rate as a function of the accumulated integrated luminosity, or equivalently the HL-LHC lifetime, under different operational scenarios. The baseline configuration corresponding to the original BTL design, with an operating temperature of $-35°C$, is shown in black. The effect of annealing at $+10°C$ using the existing cooling system is represented by the red curve. The improvement achieved with the addition of thermo-electric coolers (TECs) is illustrated by the blue and green curves, corresponding to operating temperatures of $-45°C$ and annealing temperatures of $+40°C$ and $+60°C$, respectively. The overvoltage (OV) is fixed at 1 V in all scenarios.

3.3.3 State-of-the-art BTL performance

At present, the BTL module design is nearly finalized, incorporating several refinements with respect to the original proposal to ensure the required time resolution is achieved both at the start and at the end of the detector's operational lifetime. The BTL will be equipped with LYSO:Ce scintillating tiles measuring $3.75 \times 3.0 \times 54.7$ mm^3, featuring a uniform thickness across the full pseudorapidity range up to $|\eta| = 1.48$. Adjacent tiles are optically isolated using layers of Enhanced Specular Reflector (ESR) film, while each detector module will consist of an array of 16 bars. The top and bottom surfaces of the module will also be wrapped with ESR to maximize light collection efficiency. Scintillation light will be read out at both ends of each bar by SiPMs with an active area precisely matching the crystal front face and a microcell size of 25 μm^2.
Thermo-electric coolers (TECs) will maintain the operating temperature at -45°C and enable local heating to +40–+60°C during LHC shutdown periods to perform thermal annealing, as described in Section 3.3.2. The SiPM signals will be processed by a dedicated front-end ASIC, the TOFHIR chip. The latest version, TOFHIR v2, is implemented in a 130 nm CMOS technology with improved radiation tolerance [127]. This version also integrates a DCR noise-cancellation circuit to mitigate end-of-life performance degradation. Additionally, the TOFHIR analog gain will be progressively increased throughout the HL-LHC period to compensate for the reduction in SiPM overvoltage and the corresponding loss in photon detection efficiency (PDE).
Recent test-beam campaigns have demonstrated that the updated BTL design achieves a timing resolution of approximately 20 ps at the beginning of HL-LHC operation and about 60 ps after an integrated luminosity of 3000 fb^{-1}. Figure 3.9 presents the expected timing performance as a function of integrated luminosity, based on test-beam data for the final crystal thickness (3.75 mm) and SiPM cell size (25 μm^2). Assuming an operating temperature of -45°C and an annealing temperature of +60°C, the chosen combination of thicker crystals and larger SiPM microcells meets the required BTL performance targets, validating the final detector configuration.

3.4 MTD Data Acquisition System

The Data Acquisition (DAQ) system of the MIP Timing Detector (MTD) combines radiation-tolerant front-end electronics, located inside the experimental cavern, with FPGA-based back-end electronics situated in radiation-safe areas outside the cavern [114]. The MTD DAQ is fully integrated into the central CMS DAQ system [128], shared by all CMS subdetectors. Upon receiving a Level-1 (L1) trigger signal from the global CMS trigger, the MTD front-end electronics transmit the corresponding detector data to the back-end system via high-speed optical links. Based on the current layout and expected data rates, the Barrel Timing Layer (BTL) will utilize 864 optical links, while the Endcap Timing Layer (ETL) will require 1688. A unified DAQ architecture serves both subsystems, providing bidirectional communication for fast and slow control, configuration, and monitoring of the on-detector electronics.
The DAQ system also manages the distribution of a low-jitter sampling clock, synchronized with the 40 MHz LHC bunch-crossing frequency, to the front-end electronics. Achieving the target MTD timing resolution requires that the link-to-link clock jitter remains below 15 ps RMS across all distribution channels. For example, a 35 ps intrinsic timing resolution from the MTD sensors and electronics would be degraded by only 3 ps

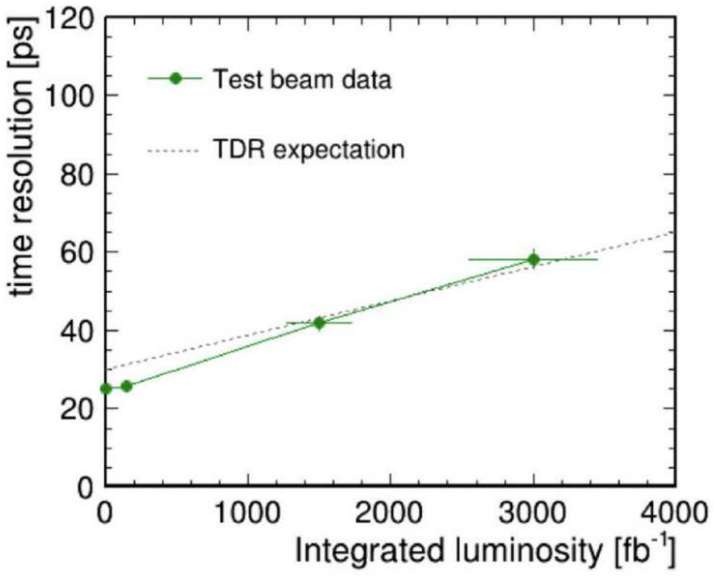

Figure 3.9: BTL time resolution as a function of the accumulated integrated luminosity, assuming an operating temperature of −45°C and an annealing temperature of +60°C. The projections are derived from test-beam measurements performed with both irradiated and non-irradiated SiPMs. Grey markers correspond to LYSO:Ce crystals with a thickness of 3.75 mm coupled to SiPMs featuring a microcell size of 25 μm^2.

under a 15 ps clock jitter contribution. The LHC master clock, generated by the RF system, exhibits an intrinsic jitter of approximately 9 ps RMS. This reference clock is distributed by the CMS central systems to both on-detector and off-detector electronics through multi-mode optical fibers.
Two possible clock-distribution schemes are foreseen. In the baseline configuration, the encoded clock signal is transmitted over the bidirectional DAQ links to the front-end readout electronics. If this approach fails to meet the required precision, a direct, unencoded clock distribution will be adopted. Although this alternative requires additional optical fibers and transceivers, it ensures optimal performance. A dedicated fan-out ASIC, called *RAFAEL*, has been developed to serve as a low-jitter clock multiplexer for this purpose. Test measurements have demonstrated that this solution achieves a link-to-link jitter of 3.2 ps RMS, comfortably within MTD design specifications [129].
This chapter is organized as follows. Section 3.4.1 provides an overview of the hardware components of the MTD DAQ. Section 3.4.2 describes the unified software and firmware framework developed for the system, with special emphasis on its architecture — a major original contribution of this work. Finally, Section 3.5.2 presents the validation procedures developed for BTL tray testing at the assembly centers, performed using the dedicated MTD DAQ software.

3.4.1 Overview of the DAQ Hardware

This section provides an overview of the key electronic components that constitute the BTL Data Acquisition (DAQ) system, including both on-detector and off-detector elements. While some front-end components differ between the BTL and ETL due to their distinct design and geometric constraints, the off-detector infrastructure is common to both subsystems and ensures a unified DAQ architecture across the MTD.

The BTL DAQ enables fully bidirectional communication between the on-detector electronics and the back-end system. Through these links, the system distributes the sampling clock, configuration parameters for the front-end chips, synchronization and reset commands, the Level-1 (L1) trigger signal, and power. In return, it receives both the high-speed data stream corresponding to triggered events and the slow-control monitoring information, such as SiPM temperature, bias voltage, and other diagnostic quantities essential for detector operation and stability.

TOFHIR Chip

The TOFHIR2C ASIC acquires and digitizes the signals from the SiPMs. The chip is radiation tolerant, and its design is optimized to meet the MTD timing requirements for the digitization of both time and energy of minimum ionizing particles (MIPs). Additionally, TOFHIR is able to reject low-energy signals from particles below the MIP threshold. The front-end chip measures the time of arrival of MIP signals with a time jitter of approximately 14 ps. The timing threshold discriminator has a 6-bit precision and can be configured within the 0-100 photoelectron range.
At the input stage, the preamplifier acts as a current buffer, isolating the large SiPM load from the analog front-end (AFE) core, and mirroring the output current to separate timing and energy branches. Owing to the large dark count rate (DCR) of the SiPMs and the high event rate expected at the HL-LHC, the preamplifier output is affected by baseline drifts and pulse pile-up. Since the comparators rely on thresholds referenced to this baseline, such drifts could introduce timing inaccuracies. Moreover, DarkCount and signal pulses exhibit similar shapes and frequency content, differing primarily in amplitude. To overcome these limitations and maintain sub-60 ps timing accuracy at the end of operation, TOFHIR employs a signal processing technique known as the Differential Leading Edge Discriminator (DLED).
The DLED converts the unipolar SiPM pulse into a bipolar waveform by subtracting a delayed replica of the signal from the original, see Figure 3.10. With a proper delay setting, the DLED preserves the rising edge of the pulse while cancelling its trailing tail, see Figure 3.11, thus mitigating baseline fluctuations and reducing pile-up effects. In TOFHIR2, the delay line is realized using a chain of RC taps acting as a low-pass filter, providing a programmable delay between 200 and 1400 ps. Adjusting the delay not only allows fine-tuning of the pulse amplitude but also provides flexibility to optimize timing performance. To compensate for process variations between the PMOS and NMOS devices in the delay circuit, a 5-bit DAC is implemented to trim the pulse shape. This calibration can be performed by measuring the width of the positive lobe above the baseline using the T1 comparator and the two TDCs. A baseline holder feedback loop stabilizes the DC operating point without affecting the pulse dynamics, as its bandwidth is limited to the mHz range.
The differential DLED output current is mirrored into two comparators, T1 and T2, used for precise time-tagging and for low-energy hit rejection, respectively. The T1 comparator operates at a low threshold to maximize timing resolution, while T2 is set at a higher threshold to suppress noise and low-energy events. The T1 signal is delayed to ensure that the timing measurement is triggered only for pulses exceeding the T2 threshold, effectively reducing spurious triggers from dark counts. Both comparators employ a fast, low-impedance architecture with positive feedback to ensure sharp switching. Their threshold DACs feature configurable least significant bits (LSBs) of [0.156, 0.313, 0.47,

0.63] µA for T1 and [0.313, 0.63, 0.94, 1.25] µA for T2, referenced to a baseline current of 45 µA.
The preamplifier current is mirrored into two DLED paths: one dedicated to the timing branch and one to the energy branch. The energy branch prioritizes linearity over speed and includes a programmable gain stage with eight selectable settings (2/8-9/8) to optimally match the integrator's dynamic range. An additional comparator (E) is included for energy discrimination, sharing the same threshold DAC as T2. This dual-branch architecture ensures stable timing and energy measurements even under high DCR and pile-up conditions, crucial for achieving the MTD's target time resolution.
Upon the passage of a MIP, two timestamps are digitized by the TOFHIR chip and stored in the ASIC memory, corresponding to the times when the signal crosses the timing and MIP thresholds. The conversion is performed using a Time-to-Digital Converter (TDC) circuit triggered by these two discriminators. The first timing measurement corresponds to the rising edge of the signal, while the second can be taken either on the falling or rising edge, allowing the reconstruction of the time-over-threshold (ToT), which provides information on the signal width. The TDC achieves a quantization step of 20 ps and consists of a time-to-amplitude converter (TAC) followed by a 40 MHz, 10-bit successive-approximation (SAR) analog-to-digital converter (ADC) [130]. A coarse 16-bit counter counts the number of cycles of the external 160.32 MHz reference clock, while a fine counter provides sub-cycle resolution [114]. TOFHIR also features a charge-to-digital converter (QDC), composed of a charge-to-amplitude converter (QAC) followed by the same ADC used for timing measurements. The integration window is 25 ns, starting from the rising-edge trigger.

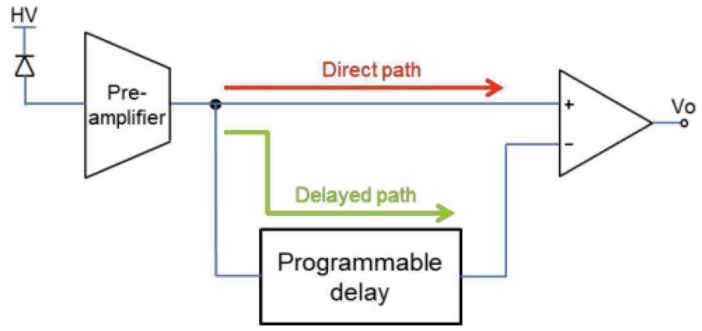

Figure 3.10: Block diagram of the TOFHIR2 ASIC DLED.

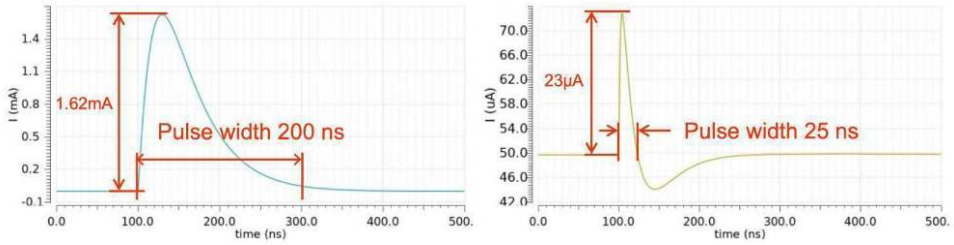

Figure 3.11: DLED input (left) and output (right) waveforms.

The TOFHIR will store the information corresponding to a MIP hit with two digitised times, the signal amplitude and the associated clock cycle number. This data words

will be shipped to other electronic components for further processing as twelve 8/10 bit encoded symbols. The initial symbol is K28.1, while the subsequent 11 symbols collectively represent 88 bits of data their content is described in Table 3.1.

Bits	Stored information
0-4	channel ID
5-7	time-to-amplitude converter ID
8-11	previous trigger flags
12-21	idle time since previous trigger
22-31	fine counter of the 1st time measurement
32-41	fine counter of the 2nd time measurement
42-51	integrated charge measurement
52-61	coarse counter of the end of charge integration
62-71	coarse counter of the 2nd time measurement
72-86	coarse counter of the 1st time measurement
87	0 packet type identifier

Table 3.1: The TOFHIR frame from the DAQ perspective, where each bit/s content is described

The data words can be sent upon the reception of an external trigger signal or via a self-triggering mechanism. TOFHIR is instrumented with two output ports dedicated to data transmission through E-links [131] at 320 Mb/s. An additional E-link input port, operated at 80 Mb/s, allows to communicate with the ASIC from the outside for e.g. configuring the chip. Two additional input ports are dedicated for the handling of fast control command. The first of these ports handles L0 and L1 signals. The second port can receive RESET and RESYNC commands that reset the timing coarse and fine counters and, in case of RESET, also the content of all FIFOs registers and state machines. The TOFHIR identification number can be configured with four dedicated pins.

Front-End Board

The BTL front-end module represents the fundamental building block of the detector's Data Acquisition (DAQ) system. Since the initial MTD Technical Design Proposal, the layout of the front-end boards has undergone several iterations to optimize performance, integration, and radiation tolerance.
In the current design, each BTL module consists of two arrays of sixteen LYSO:Ce crystals. Each crystal is optically coupled to a Silicon Photomultiplier (SiPM) on either end, forming a total of 32 readout channels per module. To ensure the required operational temperature of $-45°C$ for the SiPMs, four thermo-electric coolers (TECs) are mounted on each side of the module.
Signal processing is performed by two TOFHIR2 ASICs, each reading out one crystal array and interfacing with the 32 SiPMs through flexible printed circuits. Power regulation and bias stability are provided by two ALDO2 chips [132], one per TOFHIR2. Each ALDO2 supplies a regulated voltage and current up to 500 mA, sufficient to power the associated SiPM array.
The front-end boards are connected to a concentrator card via low-power GigaBit Transceiver (lpGBT) links, which manage both uplink and downlink communication for data, control, and synchronization. Photographs of the crystal assemblies and the front-end boards are shown in Figure 3.12.

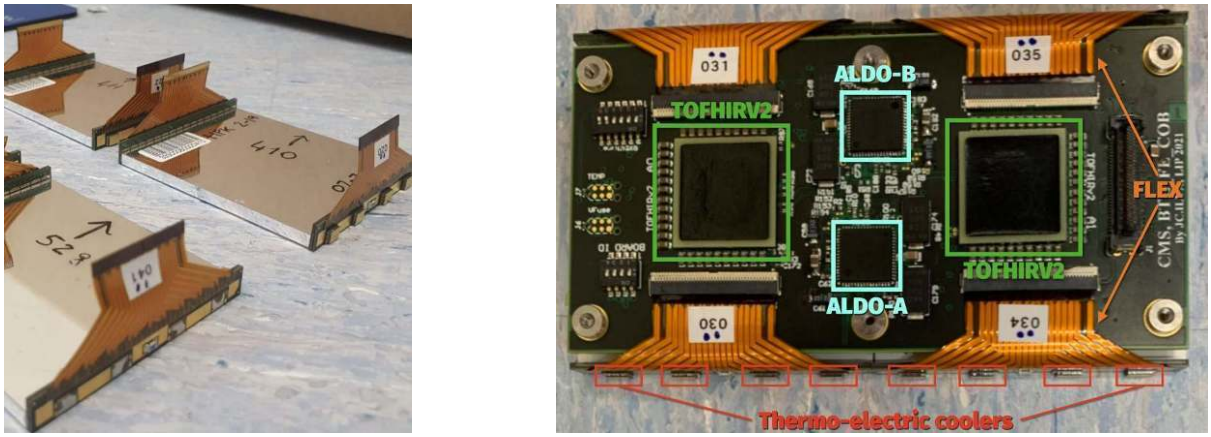

Figure 3.12: Left: BTL module comprising an array of sixteen LYSO:Ce bars, each wrapped in Enhanced Specular Reflector (ESR) foil and coupled on both ends to SiPMs. The SiPMs are connected via flexible circuits (orange) to the TOFHIR readout electronics. Right: BTL front-end board featuring two TOFHIR2 chips (green), each interfaced with 32 SiPM channels through the flex connections (orange). Two ALDO2 power regulation chips (cyan) supply the TOFHIR2s, while four thermo-electric coolers (TECs, red) are mounted back-to-back with the SiPM arrays, each TEC serving four SiPMs.

Low-Power GigaBit Transceivers (lpGBT)

Optical communication between the on-detector and off-detector DAQ systems is provided by the Low-Power GigaBit-Versatile Link (lpGBT) system [133]. This bidirectional architecture connects the front-end electronics to the DAQ back-end through radiation-tolerant, high-speed optical links.
The data produced by the TOFHIR2 ASICs are transmitted via E-link interfaces to two lpGBTs, each operating at a bandwidth of 320 Mb/s. Conversely, configuration commands for the front-end chips—together with fast control signals such as RESYNC and trigger commands—are sent from the lpGBT to the TOFHIR2 and ALDO2 chips through an E-link operating at 80 Mb/s. In addition, the lpGBT distributes the 160 MHz sampling clock to the TOFHIR ASICs, ensuring synchronization with the LHC bunch crossing frequency.
A schematic representation of the lpGBT architecture is shown in Figure 3.13, while its detailed integration with the front-end and back-end electronics is discussed in the following sections.
The lpGBT is a radiation-hard ASIC specifically designed to implement flexible, high-speed, bidirectional optical links in high-radiation environments typical of high-energy physics experiments. It supports downlink communication from the counting room to the detector at 2.56 Gb/s, and uplink data transmission at either 5.12 or 10.24 Gb/s, depending on the selected operating mode. The lpGBT provides the very front-end electronics with three logical data paths dedicated to timing and trigger distribution, DAQ data transfer, and slow control communication.
The device is highly configurable, featuring 11 hardware configuration pins and 494 programmable 8-bit registers that allow fine-grained tuning of its functionality. It can operate in three distinct modes—as a bidirectional transceiver, a unidirectional transmitter, or a unidirectional receiver—making it a highly adaptable component for a wide range of applications within the MTD DAQ system.

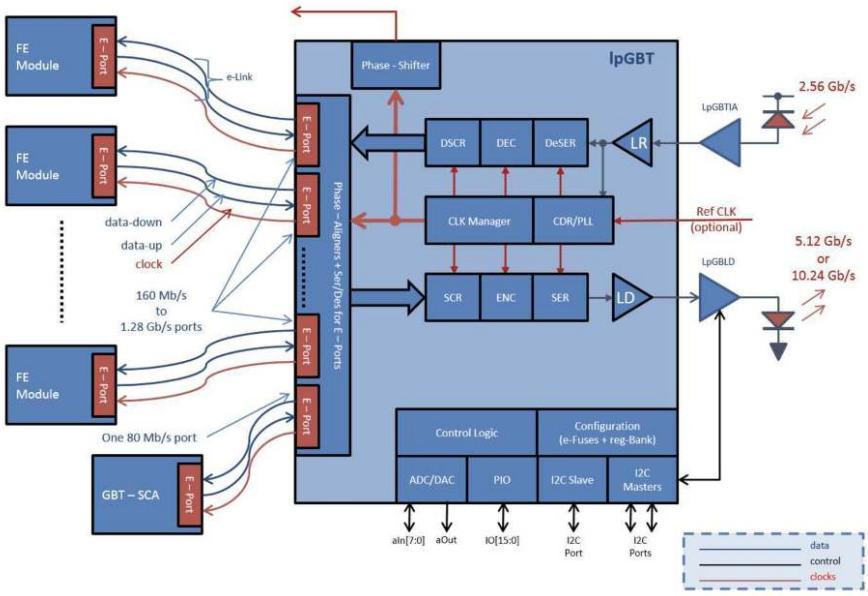

Figure 3.13: Architecture of the Low-Power GigaBit Transceiver (lpGBT). Figure adapted from [133].

The communication between the lpGBT and the detector modules is established through E-links, which are driven by a set of output ports (ePortTx) matched to the corresponding front-end devices. These ePorts are organized into groups, referred to as eGroups, each comprising four ePorts. Every ePort is associated with a differential eClock line responsible for transmitting the clock signal to the front-end electronics. Both data and clock signals are transmitted and received through the ePorts using dedicated Line Drivers (eTx) and Line Receivers (eRx) that implement the CERN Low-Power Signaling (CLPS) standard. On the downlink, transmission ePorts serialize the parallel data before sending it to the front-end devices. Conversely, on the uplink, data received from the front-end modules by the lpGBT passes through the receiving ePorts (ePortRx), where it is deserialized, scrambled, and encoded with Forward Error Correction (FEC) before transmission to the off-detector electronics. A laser driver, configured via an I^2C interface from the lpGBT, handles the optical uplink communication between the lpGBT and the counting room.
Although the primary function of the lpGBT is to act as a bidirectional transceiver, it also integrates extensive control and monitoring capabilities. These include three I^2C master interfaces for configuring external devices (e.g., the laser driver), a 16-bit programmable I/O port, a 10-bit ADC, on-chip temperature monitoring, configurable current sources for external temperature sensors, and an 8-bit voltage DAC.
Downlink data from the counting room to the lpGBT is transmitted in 64-bit frames at a rate of 2.56 Gb/s. Each frame is composed of three distinct fields:

- **Internal Control (IC)** — used for ASIC control and configuration, with a bandwidth of 80 Mb/s;
- **External Control (EC)** — employing a dedicated ePort with 80 Mb/s bandwidth, designed for compatibility with the GBT-SCA chip [134] used for slow control and monitoring;

- **Data Channel (DC)** — reserved for user data transmitted to the front-end ASICs, with a bandwidth of 1.28 Gb/s.

The uplink frame from the lpGBT to the counting room mirrors this structure, operating at either 5.12 Gb/s or 10.24 Gb/s with frame sizes of 128 or 256 bits, respectively. Two FEC modes are available, capable of correcting up to 5 consecutive bits (FEC5) or 12 bits (FEC12). For an uplink running at 5.12 Gb/s, the IC and EC fields maintain the same bandwidths as in the downlink, while the DC field reaches 3.84 Gb/s (with FEC12) or 4.48 Gb/s (with FEC5). When operating at 10.24 Gb/s, the IC and EC fields remain unchanged, whereas the DC bandwidth doubles accordingly.
A schematic view of the uplink frame structure, as seen by the lpGBT user for a 5.12 Gb/s data rate with FEC5 encoding, is shown in Figure 3.14.

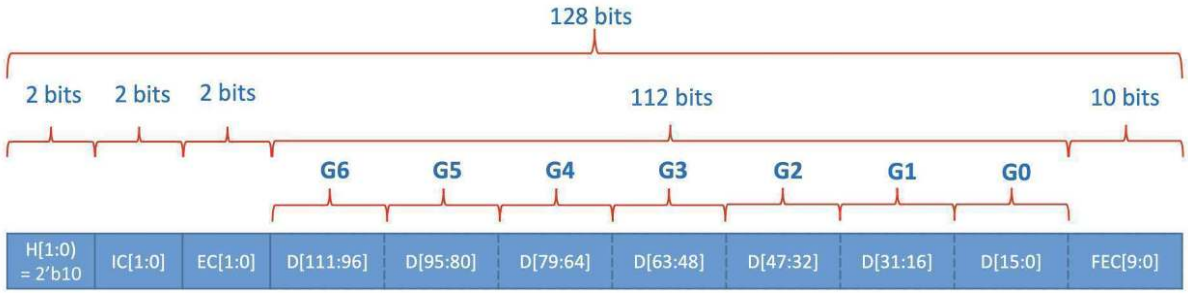

Figure 3.14: lpGBT uplink frame structure for a data rate of 5.12 Gb/s with FEC5 encoding. This frame corresponds to the format received by the user after processing by the lpGBT-FPGA interface. Bits G0-G6 denote the fields allocated to the ePort groups. For the downlink direction, the number of eGroups is reduced to four due to the asymmetric data rate, and the frame length is 64 bits.

The number of eLink connections to the front-end ASICs depends on several factors, including the uplink data rate, the selected FEC encoding scheme, and the bandwidth configuration of the uplink and downlink channels, as summarized in Table 3.2.

	Downlink e-links			**Uplink e-links**											
Data rate	2.56 Gb/s			5.12 Gb/s						10.24 Gb/s					
FEC coding	-			FEC5			FEC12			FEC5			FEC12		
eLink (Mb/s)	80	160	320	160	320	640	160	320	640	320	640	1280	320	640	1280
Number	16	8	4	28	14	7	24	12	6	28	14	7	24	12	6

Table 3.2: Number of available uplink and downlink lpGBT electrical eLinks

Due to the asymmetric bandwidths of the uplink and downlink channels, the number of input and output eLinks available on the lpGBT differs. The total number of available eClocks, however, remains constant at 29, independent of the programmed clock frequency. The timing reference for these eClocks can be either externally supplied (synchronized to the LHC bunch-crossing frequency) or derived from the downlink data stream. For proper operation, the clock provided to the lpGBT must maintain a well-defined phase relationship with respect to the LHC bunch-crossing clock. When this condition is met, both the eClock and eLink data outputs exhibit a fixed phase relative to the bunch-crossing signal. This phase is hardcoded within the lpGBT firmware and cannot be user-configured. Additionally, the lpGBT provides four programmable clock

outputs with a phase adjustment granularity of 50 ps. To ensure uniform phase alignment across all eLink data channels, the lpGBT integrates a dedicated internal phase-alignment mechanism.

GBT-SCA

While the lpGBT includes basic slow-control and monitoring functionalities, these are insufficient to meet the full operational requirements of the BTL. To extend these capabilities, two GBT-SCA (Slow Control Adapter) ASICs are employed, providing additional interfaces for configuration, monitoring, and environmental readout. The GBT-SCA inherits the radiation-tolerant design and optical link architecture of the GBT family [133], and is optimized for the distribution of slow-control and monitoring signals to the on-detector front-end electronics. It communicates with the lpGBT via the External Control (EC) channel using a dedicated ePort operating at 80 Mb/s and synchronized to a 40 MHz clock.
The ASIC offers a rich set of user interfaces to interact with the detector electronics, including 16 I^2C master controllers, a JTAG master controller, and 32 general-purpose input/output (GPIO) pins. Furthermore, it features 31 analog input channels multiplexed to a 12-bit ADC with offset calibration and gain correction capabilities, enabling precise measurements of temperatures, voltages, and currents from the SiPM sensors. Communication through the SCA's eLink port follows the HDLC transport protocol, with the data-packet structure illustrated in Figure 3.15.

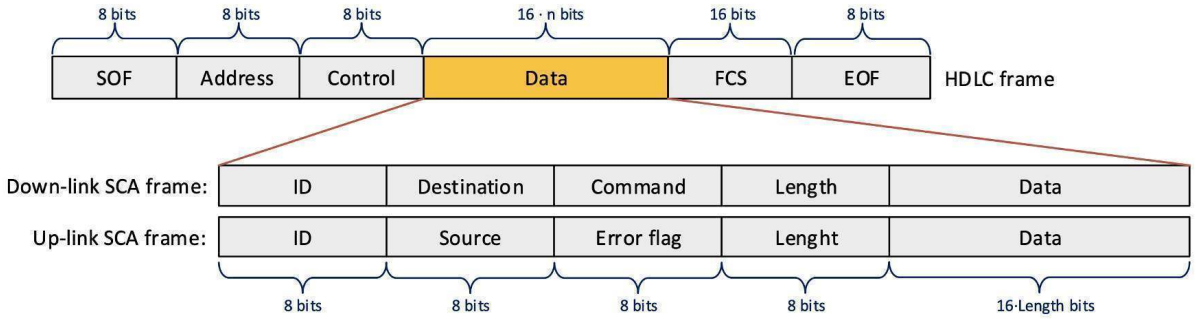

Figure 3.15: Data structure of the GBT-SCA uplink and downlink communication

The HDLC frame structure consists of several distinct fields. It begins with an 8-bit start-of-frame (SOF) delimiter composed of six consecutive logic "1" bits, followed by an 8-bit address field and an 8-bit control field. These are succeeded by a variable-length payload field, a 16-bit Frame Check Sequence (FCS) used for error detection, and finally, an end-of-frame (EOF) delimiter. The payload content follows the SCA channel command protocol, which specifies communication with the internal interface channels of the ASIC. It includes four 8-bit fields: a transaction identification number, a destination/source address (depending on the data direction), a command or error flag field, and a data-length register, followed by a variable-length data segment. The Control field can additionally instruct the SCA to perform one of three special operations: `CONNECT`, which activates the SCA ePort to receive data; `RESET`, which clears the ePort, FIFOs, and internal state machines; and `TEST`, which enables an internal loopback mode for diagnostic purposes.

Readout Unit

A BTL Readout Unit (RU) is implemented as a Concentrator Card (CC), a custom PCB that integrates two lpGBT transceivers responsible for aggregating data from 12 front-end boards, along with two GBT-SCA chips, two RAFAEL clock fan-out ASICs, and two Power Converter Cards (PCC-bPOL12) that supply low-voltage power to all components described above. Each RU services 768 SiPM channels by interfacing the two lpGBTs with a total of 24 TOFHIR ASICs. The lpGBTs are operated in full transceiver mode, each configured for an uplink bandwidth of 5.12 Gb/s, yielding a combined data throughput of 10.24 Gb/s. This configuration is determined by the expected average data rate of approximately 230 Mb/s per eLink, estimated assuming an 8% channel occupancy, 120 bits per hit, and a 0.75 MHz Level-1 trigger rate [114]. This corresponds to a total uplink bandwidth requirement of about 5.5 Gb/s, which can be comfortably sustained by two lpGBTs operating at their lower-speed configuration.
Each PCC hosts three radiation- and magnetic-field-tolerant DC/DC converters based on the FEASTMP_CLP ASIC [135]. The two GBT-SCA devices on the CC handle slow control and monitoring tasks. Specifically, they regulate the six DC/DC converter output voltages, monitor the temperatures of the SiPM arrays, and supervise thermal sensors distributed across the front-end boards, PCCs, and the CC itself. Additionally, the SCAs track the SiPM bias currents—one per group of 16 SiPMs, for a total of 48 current measurements per RU.
The final layout and routing of the CC define the interconnections among the various ASICs, establishing a tight coupling between hardware implementation and firmware/software development. Accurate knowledge of the signal mapping and interface pinout is essential for reliable read and write operations at the register level. A photograph of the final RU assembly, showing the complete CC design and all major components, is displayed in Figure 3.16.

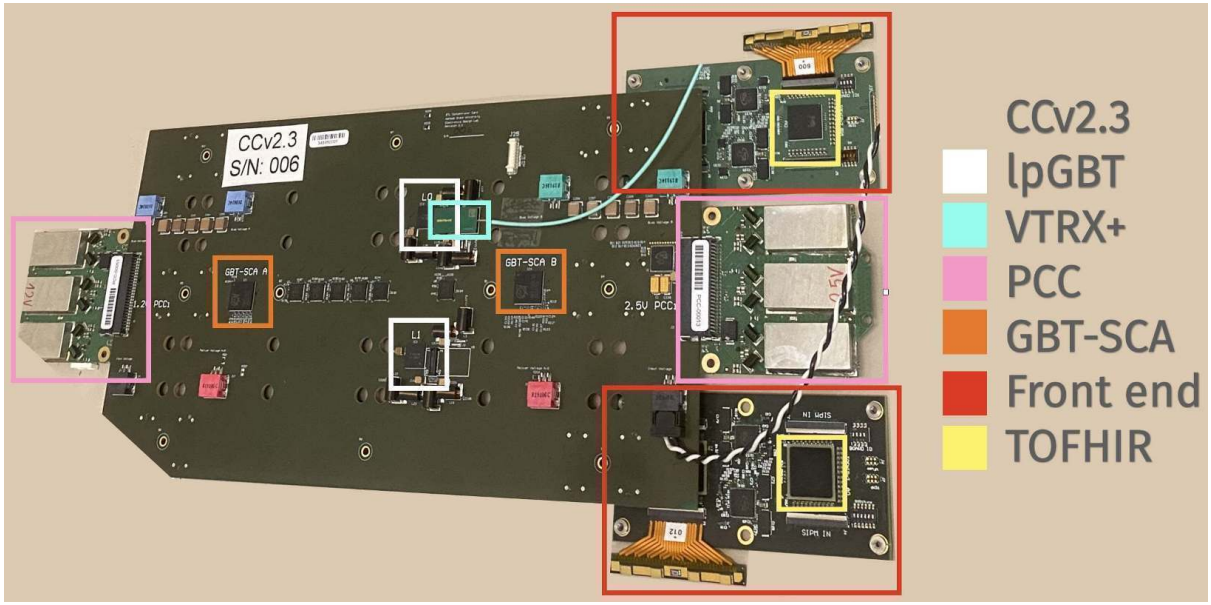

Figure 3.16: Layout of the BTL readout unit with CCv2.3 final design with two front end boards connected.

Back-End Electronics

The off-detector electronics are shared between the BTL and ETL subsystems. The back-end system is responsible for receiving, decoding, and aggregating data from approximately 864 lpGBTs in the BTL and 1600 in the ETL. Data are transmitted from the lpGBTs via radiation-tolerant optical transceivers (VTRx+) through 70 m multimode optical fibers to the back-end electronics. At the receiving end, the optical signals are captured by Firefly transceivers mounted on the MTD DAQ nodes, where the data are processed, formatted, and subsequently forwarded to the Timing and Control Distribution System (TCDS). This communication path is implemented through the DAQ and Timing Hub (DTH400) board [136], with optional additional bandwidth provided by the DAQ800 board. The DTH400 delivers a total DAQ throughput of 400 Gb/s, while the DAQ800 doubles this capacity to 800 Gb/s. Each MTD DAQ node connects to the DTH board through 25 Gb/s front-panel optical links. The aggregated data from all MTD DAQ boards are transmitted to the CMS event builder via the 100 Gb/s data-to-surface (D2S) protocol [114].
The MTD back-end architecture for the CMS Phase-II upgrade adopts the Advanced Telecommunications Computing Architecture (ATCA) standard. All electronic boards are installed in ATCA crates, with the total number of boards determined by the data bandwidth requirements of each subsystem. The DAQ system must sustain the expected data throughput, assuming an average channel occupancy of 8% for the BTL and 6% for the ETL, an event size of 120 bits, and a Level-1 trigger rate of 750 kHz. These parameters correspond to an average data rate of approximately 5.5 Gb/s per readout unit for the BTL and 1.5 Gb/s for the ETL. Considering the full detector coverage—with 432 and 1136 service hybrids instrumented for the BTL and ETL, respectively—the total expected data throughput amounts to about 2300 Gb/s for BTL and 1900 Gb/s for ETL. These bandwidth requirements can be accommodated using four MTD DAQ boards for the BTL and eight for the ETL, installed within an ATCA crate together with one DTH400 and two DAQ800 boards. Both the BTL and ETL back-end systems will each occupy two ATCA crates, as summarized in Table 3.3.

Component	**BTL**	**ETL**	**Component Bandwidth [Gb/s]**
lpGBT links	864	1688	5.12 (BTL) - 5.12 (ETL)
Serenity boards	8	16	25 (BTL) - 16 (ETL)
DTH400	2	2	400
DAQ800	4	4	800
Crates	2	2	-

Table 3.3: BTL and ETL back-end components and their respective bandwidths.

The core element of the MTD-specific DAQ system is the MTD DAQ node board, based on the Serenity platform [137, 114], which is shared with other CMS subsystems such as the Tracker, HGC, and L1 Trigger. The Serenity is an ATCA-compliant platform comprising three main components: an ATCA carrier card providing power, clock distribution, optical and electrical interfaces; an IPMC mezzanine card for board management; and an onboard CPU for control and monitoring. Data received from the lpGBTs are processed by FPGA-based daughterboards mounted on the carrier card. In addition to the hardware components, a flexible suite of generic firmware and software is integrated into the system to manage data handling and communication.

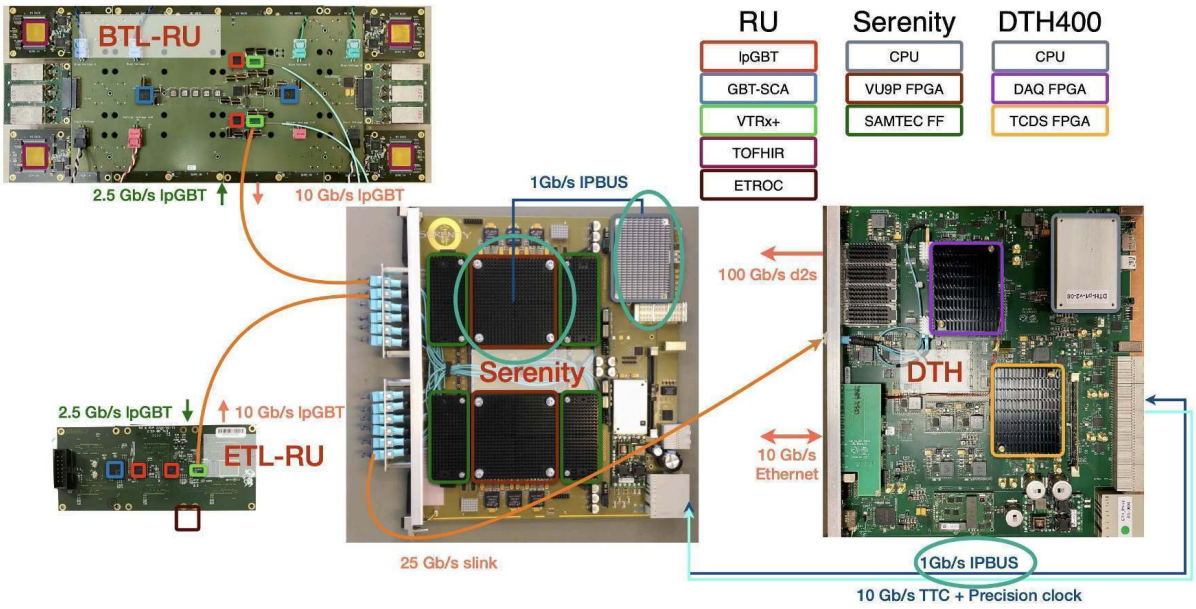

Figure 3.17: Schematic representation of the MTD DAQ chain, from the on-detector readout units (RUs) to the off-detector electronics (Serenity and DTH400). Image courtesy of Ö. Sahin.

Each Serenity board hosts two FPGAs connected to twelve SAMTEC Firefly optical transceivers, providing a total of 72 bidirectional optical links per FPGA, or 144 per board. These links interface either with lpGBTs on the front-end electronics or with the DTH400/DAQ800 boards on the back end. The Serenity platform and firmware are designed to accommodate multiple FPGA architectures; for the final MTD configuration, two symmetric Xilinx Virtex UltraScale+ VU13P FPGAs will be used.
The complete MTD DAQ chain, spanning from the on-detector readout electronics to the off-detector DTH400 board, is illustrated in Figure 3.17.

3.4.2 Development of the DAQ Software Framework

The Serenity board is accompanied by a suite of firmware and software tools. Its highly flexible FPGA architecture, while powerful, introduces challenges for the development of a common firmware framework. Since the Serenity back end is shared among multiple CMS subsystems—HGC, Tracker, L1-Trigger, and MTD—the adoption of a unified and homogeneous firmware environment is particularly advantageous. The Serenity design leverages a general-purpose firmware framework known as EMP (Extensible Modular Platform), which serves as the foundation for firmware development and integration across different subdetectors.
The EMP framework provides a modular design strategy that separates the core FPGA infrastructure from user-defined algorithmic "payloads." It implements top-level designs for various FPGA models supported by Serenity, connecting their input/output ports to the clocking infrastructure, control bus, and I/O buffers. The main goal of EMP is to allow seamless integration of arbitrary user-developed firmware payloads across heterogeneous FPGA architectures. The framework is written in the Very High Speed Integrated Circuit Hardware Description Language (VHDL), one of the standard languages for dig-

ital logic design. EMP also supports High-Level Synthesis (HLS) through Vivado HLS, which allows firmware algorithms written in high-level languages such as C or C++ to be compiled into register-transfer-level (RTL) code.
Because modern FPGA projects involve a large number of components, packages, and dependencies, EMP employs the `ipbb` tool to simplify project creation, building, and reproducibility. Additional utilities such as `SMASH` and `IPMC` are provided for slow-control monitoring, power management, and diagnostic functions. Communication between the Serenity board and the control room is implemented via the IPBus protocol, a standard mechanism for reading and modifying memory-mapped resources within FPGA-based, IP-aware hardware devices [138]. The IPBus suite includes both firmware and software components: the firmware integrates the protocol into the FPGA design, while the software layer—known as μHal—provides a C++ and Python interface enabling users to perform read, write, and read-modify-write operations. The EMP framework is actively maintained and continuously updated with common developments from all Serenity-based subsystems.
On the software side, each subdetector retains a high degree of freedom in defining its architecture, reflecting the diversity of ASIC configurations and operational needs. For the MTD, a dedicated software package has been developed from scratch, optimized for BTL but extendable to ETL. The architecture combines low-level C++ functions—responsible for direct hardware communication—with high-level Python interfaces, which provide a readable, modular, and efficient abstraction layer. This modular approach is particularly suited to both production and testing workflows, establishing a one-to-one correspondence between the physical hardware components described in Section 3.4.1 and their high-level software representations.
In this framework, the user interacts primarily through Python handles, while the underlying C++ code manages optimized communication with the hardware. This design ensures flexibility in defining ASIC initialization routines, data readout sequences, and slow-control monitoring. Extensive testing has confirmed that the framework is reliable, scalable, and capable of supporting the full granularity of the MTD. The detailed procedures for initializing and communicating with the BTL ASICs are described in the following sections.
Each ASIC is configured through a dedicated, human-readable configuration file that retains flexibility and synthesis efficiency. Custom parsers translate these files from user-level definitions into machine-level register settings using predefined keywords. This approach minimizes redundancy in configuration management while preserving all necessary information for bookkeeping, as discussed in Section 3.4.2. Although the software provides a high level of abstraction, its use may still be nontrivial for inexperienced operators. In practice, the MTD assembly centers—where the DAQ software is essential for quality assurance and validation of detector components—will involve users with varying levels of experience who must quickly retrieve key diagnostic information during detector assembly and testing.
While the firmware focuses on the implementation of hardware communication protocols and interconnections, the MTD DAQ software manages user-level interaction, chip configuration, and sensor monitoring across all ASICs in the system. After an initial prototype developed for TOFHIR data readout using a provisional back end, the MTD DAQ firmware and software were redesigned from scratch to be fully compliant with the Serenity back-end standard. The new software framework consists of a modular, scalable Python interface coupled to an optimized C++ hardware layer. Its design ensures

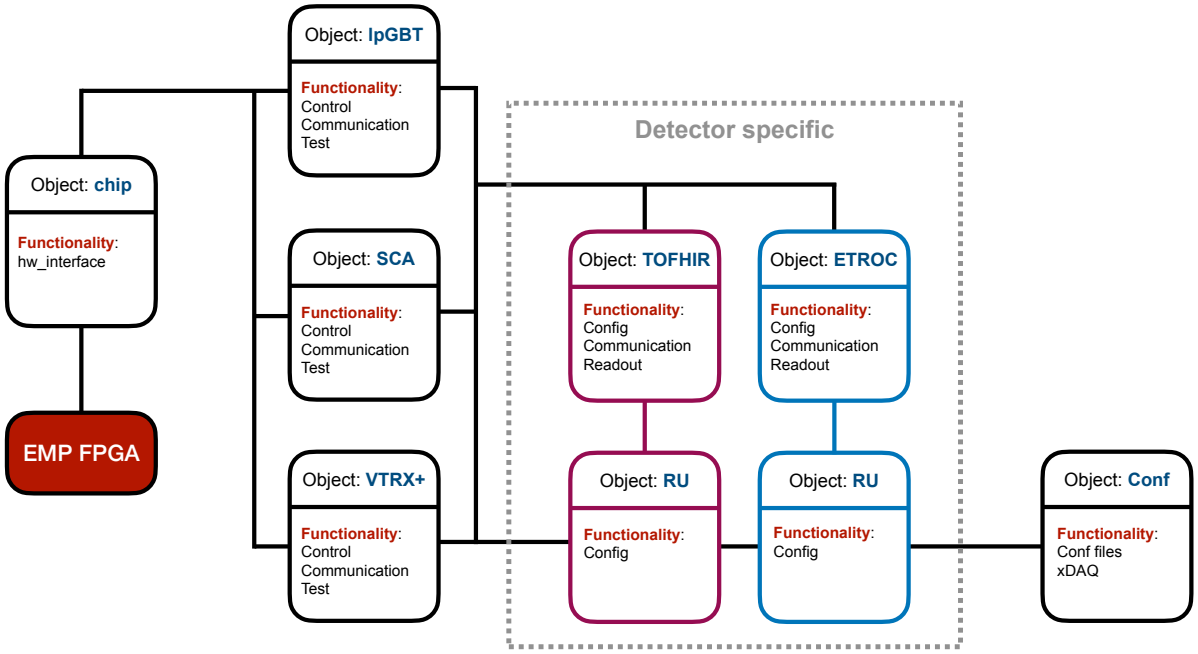

Figure 3.18: Software architecture for the Phase-II MTD DAQ

flexibility to accommodate both BTL and ETL requirements by adapting the software abstractions to the hardware topology.

As the Serenity board will serve as the final back end of the MTD during the Phase-II operation of the LHC, the DAQ firmware and software suite developed here are intended for online use during the commissioning and integration phases, and will later evolve into the "local DAQ" system used for calibration and testing after detector installation within CMS.

The core concept of the software architecture is to represent each MTD ASIC as an independent object, with higher-level abstractions naturally emerging from the hardware hierarchy. A simplified schematic representation of the MTD DAQ software architecture is shown in Figure 3.18.

All ASIC software classes are structured into three layers, each providing a progressively higher level of abstraction. The top layer offers high-level user functionalities, implementing predefined routines for chip initialization, configuration, and operation. Beneath it, an intermediate layer serves as an abstraction of the communication interfaces, providing general-purpose `read` and `write` methods independent of the specific hardware protocol. The lowest layer handles the physical communication through I2C, IC, or EC data paths, interfacing directly with the C++ functions of the μHal hardware access library via `pybind11`. This layered design conceals the details of the hardware communication from the user, ensuring that only high-level interfaces are exposed while maintaining full control over low-level operations.

Each ASIC is configured through a structured system of configuration files that store all essential operational parameters. These configurations are defined in human-readable text files parsed by a custom configuration parser. The parser efficiently manages large sets of register operations by allowing users to define keyword groups that simultaneously set multiple registers. This design keeps configuration files concise and easily interpretable, even for non-expert users. Each ASIC class is associated with a dedicated parser tailored

to its specific register map and functional requirements. Hardware connection maps and firmware link definitions are likewise stored in dedicated configuration files.
Since most detector ASICs operate under common settings, the MTD DAQ software framework was designed to minimize unnecessary duplication of configuration files while preserving flexibility for module-specific adjustments. This fine-grained configuration system ensures accurate bookkeeping of ASIC operational conditions during data taking. Moreover, the modular ASIC software structure supports multi-threaded and multi-process initialization of the chips, maximizing the computational efficiency of the CPU embedded on the Serenity board.

Hardware Communication

The hardware communication layer common to all ASICs is implemented through the `chip` class. This class provides the interface to μHal by specifying the configuration files that define the IP address and protocol for the IPBus communication endpoints, which correspond to the FPGAs on the Serenity board. An additional file specifies the address layout of the IPBus endpoints, mapping all ASIC connections. Both configuration files are written in XML.
The first XML file defines the communication endpoints and the associated address tables, referencing the two FPGAs (`x0` and `x1`), each sharing an identical address structure:

```
<?xml version="1.0" encoding="UTF-8"?>
<connections>
  <connection
   id="artix"
   uri="ipbuspcie-2.0:///dev/xdma/card0/h2c0,/dev/xdma/card0/c2h0"
   address_table="file://serenity_services.xml"
  />
  <connection
   id="x0"
   uri="ipbuspcie-2.0:///dev/serenity_pcie/x0/h2c,/dev/serenity_pcie/x0/c2h"
   address_table="file://addrtab/top_emp.xml"
  />
  <connection
   id="x1"
   uri="ipbuspcie-2.0:///dev/serenity_pcie/x1/h2c,/dev/serenity_pcie/x1/c2h"
   address_table="file://addrtab/top_emp.xml"
  />
</connections>
```

The `id` field defines a unique identifier for each endpoint, while the `uri` field specifies both the protocol and the access path to the target device. For the MTD DAQ, which employs IPBus 2.0 hardware, the protocol `ipbuspcie-2.0` is used. The `address_table` field provides the relative path to the corresponding XML file that describes the firmware's address space. Each entry in this table corresponds to a memory-mapped register, uniquely identified by name, allowing high-level software access to the underlying hardware. Since IPBus operates as an A32/D32 protocol, it supports 32-bit addressing and 32-bit data width, independent of the firmware or ASIC-specific implementation details.
The register layout defined in the address tables is directly interfaced with the firmware through the EMP module. For example, the Slow Control Commands implemented in the IC and EC fields are defined in XML as follows:

```
<node description="slow command control registers" fwinfo="endpoint;width=6" class="SCCNode">
   <node id="ic" address="0x10" class="SCCICNode">
        ...
        <node id="reply_data_count" address="0x2"/>
        <node id="gbtx_addr"        address="0x3"/>
        <node id="txdata_fifo"      address="0x4" size="0x200" mode="port"/>
        <node id="rxdata_fifo"      address="0x5" size="0x200" mode="port"/>

    <node id="ec" address="0x20" class="SCCECNode">
        <node id="txram">
                <node id="data"       address="0x0" size="0x400" mode="port"/>
                <node id="input_addr" address="0x1"/>
```

```
        <node id="frame_addr" address="0x2"/>
</node>
<node id="rxram" address="0x4">
        <node id="data" address="0x0" size="0x400" mode="port"/>
        <node id="output_addr" address="0x1"/>
        <node id="frame_addr"  address="0x2"/>
</node>
...
```

Through this structure, transmitted and received data can be easily accessed from the software by navigating the XML hierarchy. For instance, to send data to the lpGBT via the EC interface (from the Serenity board through the high-speed link to the lpGBT and subsequently to the SCA), one can simply write to `ec.txram.data`. The response from the ASIC can then be read, after a short latency, by querying `ec.rxram.data`.

Readout Unit and Configuration Parser

The fundamental block of the BTL data acquisition (DAQ) system is the Readout Unit (RU), which serves as the interface between the front-end electronics and the DAQ back-end. Each RU connects to the front-end via lpGBTs and provides access to the slow-control infrastructure through two GBT-SCAs. From a hardware perspective, the RU is an integrated assembly of ASICs and ancillary components working in concert to enable stable and synchronized data acquisition. The main ASICs hosted on a single RU are listed in Table 3.4.

Component	TOFHIR	ALDO	GBT-SCA	lpGBT	VTRX+	PCC	RAFAEL
Quantity	24	24	2	2	2	2	6

Table 3.4: Number of ASICs integrated in a single BTL Readout Unit (RU). The full detector will employ 432 RUs in total.

In the software, this hardware structure is abstracted by a container class `RU`, designed to encapsulate a collection of sub-classes, each corresponding to one ASIC on the Central Controller (CC). The main responsibility of the `RU` class is to initialize all ASICs present on the physical CC, serving as the primary interface between the user and the detector hardware. Its initialization relies on configuration files that describe the hardware layout and operating parameters. A dedicated configuration parser decodes these files and distributes the relevant information to the individual ASIC classes, ensuring their correct initialization and communication. An additional configuration file specifies the Serenity board parameters, including its connection XML and the mapping between optical links, lpGBTs, and GBT-SCAs.
The configuration framework supports hierarchical register initialization and flexible keyword-based expansion to handle redundancy and grouped settings efficiently. For instance, the configuration of all eLink clocks at 160 MHz can be achieved with a single command using the suffix "_all" appended to the desired field name. Similarly, the suffixes "_group" and "_channel" allow configuration of subsets of eLinks or individual channels. For fine-grained debugging or expert adjustments, users may directly address single registers within the configuration file.
To preserve the modularity of the software framework, each ASIC is associated with its own configuration file and custom parser, which convert user-defined parameters into software-level register settings. The RU itself is assigned a higher-level configuration file that enumerates the connected ASICs and links them to their respective configuration

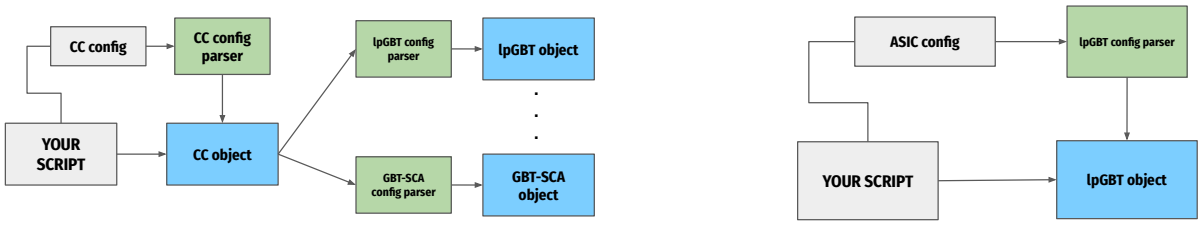

Figure 3.19: Left: workflow for the initialization of a complete Readout Unit (RU) starting from configuration files. Right: modular software structure enabling independent initialization of individual ASICs for debugging.

files. It also includes a mapping table defining the Serenity connections specific to that RU. This design enables flexible initialization: the entire RU can be configured in a single operation, or individual ASICs can be initialized independently—an approach particularly useful for debugging. The overall workflow is illustrated in Figure 3.19. The initialization of a single ASIC, such as an lpGBT or GBT-SCA, typically requires about 15 ms when performed from its configuration file.

3.5 Tray Validation

The tray validation procedure performed at the BTL Assembly Centers (BACs) was developed using the MTD DAQ software framework operating with the Serenity back-end platform. This setup provides full control and readout of the tray electronics, enabling both low-level communication checks and calibration tasks. The validation campaign represents the final quality-control stage before certifying a tray for integration into the detector.
The procedure begins with basic functionality checks of the front-end ASICs. The lpGBT and GBT-SCA chips are first tested by reading temperatures and currents through their slow-control interfaces, ensuring proper communication and power distribution. The TOFHIR ASICs and SiPMs are then validated using current-voltage (IV) scans to confirm correct connectivity and detector response.
Once these basic checks are completed, a series of calibration runs is performed. These include the discriminator (DISC) or threshold calibration for the T1 timing branch, followed by the time-to-digital converter (TDC) calibration to align fine timing measurements, and finally the energy calibration to intercalibrate the channel response. A complete set of reference calibrations is provided by the Front-End (FE) QA/QC center at the Lisbon Laboratory of Instrumentation and Experimental Particle Physics (LIP), against which the locally measured calibrations are compared to verify their consistency and stability. Together, these steps ensure that all front-end components of the tray are functioning within specifications before the tray is approved for integration.

3.5.1 Validation of CC Communication

To confirm the correct operation of the CCs, three complementary checks are performed:

- **Low-level communication:** we establish contact with the two lpGBTs and the two GBT-SCAs present in each CC. This ensures that all control and monitoring

links are functional.

- **Slow-control readout:** the internal ADCs of the lpGBTs and SCAs are operated to measure bias currents and temperatures of the FE. These values are compared to reference conditions to detect possible anomalies.

- **Long-term stability:** extended data-taking runs (typically of ∼20 minutes) with LYSO signals are performed, i.e. where we use the Lutetium in the crystals as a radioactivity source. During these runs, both lpGBTs, SCAs and all the Tofhirs are continuously operated, and their stability is monitored over time. Any loss of synchronization or drift in the readout would flag a potential communication issue.

3.5.2 Validation of Detector Modules Operation

The correct functioning of the FE electronics and, more generally, of the Detector Modules (DMs) within a tray is verified through a series of dedicated runs. These tests provide complementary information on the timing, energy, and stability of the system, and constitute the core of the tray validation workflow.

IV Scans

The current-voltage (IV) characteristics of the SiPM arrays are measured to extract the breakdown voltage V_{bd}, which defines the onset of avalanche multiplication in the microcells. The V_{bd} can be measured from the IV curve as a peak in the derivative of the logatithmic of the current, with respect to the voltage, i.e.:

$$V_{\text{bd}} = V \text{ such that } \max \frac{d \log I}{dV} \tag{3.2}$$

Knowledge of V_{bd} is essential to bias the SiPMs at a controlled overvoltage (OV), defined as the difference between the applied bias voltage and V_{bd}, ensuring uniform gain and timing performance across all channels.
This procedure relies on the bias currents read out by the SCAs, while the applied voltages are reconstructed from the ALDO DAC calibration provided by LIP. The latter maps the programmed DAC values, combined with the gain selection (0 or 1), into an effective voltage delivered to the SiPMs. The measured breakdown voltages are then correlated with those obtained during the SiPM QA/QC campaign. A systematic shift between the two is typically observed, consistent with temperature differences between the tray validation environment and the original QA/QC laboratory. Figure 3.21 shows an example IV curve, while Figure 3.21 the distribution of differences with respect to the QA/QC values, and the overall correlation map.

Threshold Calibration

Each TOFHIR channel features three discriminator branches, referred to as T1, T2, and Energy. To ensure accurate time and energy measurements, the baseline position of the discriminator thresholds must be calibrated for each branch. This procedure, commonly known as the discriminator or threshold calibration, is one of the essential validation steps in the tray testing campaign.

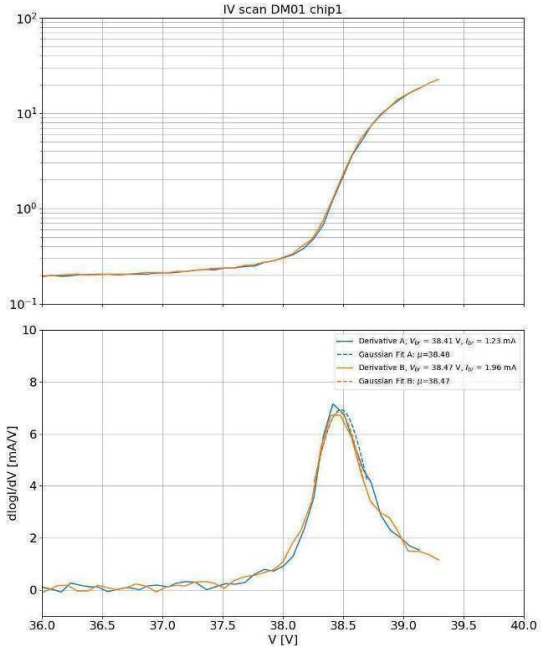

Figure 3.20: Tray validation through IV scan. An example of the IV curve for a single SM

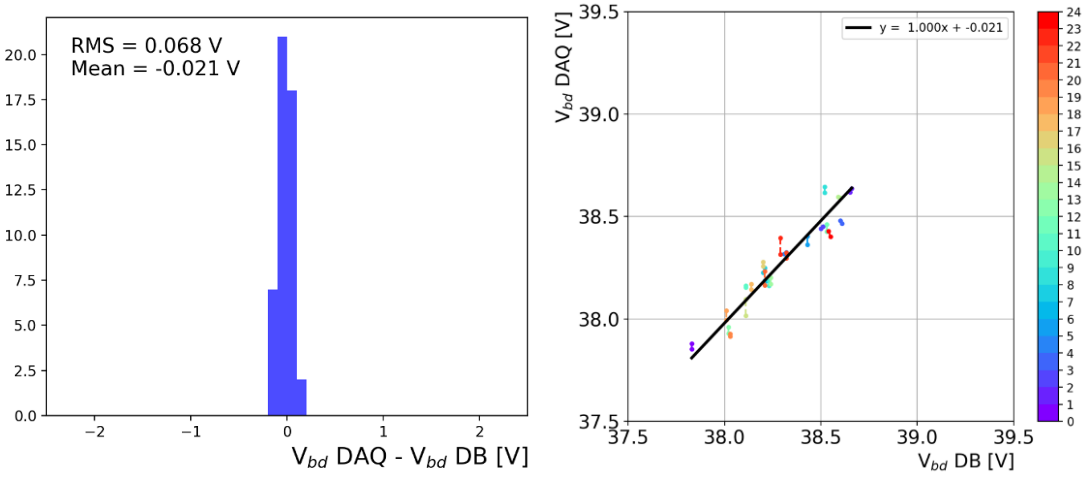

Figure 3.21: Tray validation through IV scan. On the left the difference between the measured V_{bd} and the SiPM QA/QC values. On the right the correlation of the values.

The calibration is performed by triggering on the noise current and scanning different discriminator threshold values. As the threshold increases, the probability for the discriminator to fire decreases from nearly 100% (at very low thresholds, when the noise is always above threshold) to 0% (at very high thresholds). The transition point between these two regimes, or better the point for which the rate is around 50%, defines the baseline position of the discriminator.
To accelerate the process, the internal counter functionality of the TOFHIR ASIC is exploited. By configuring the counter time parameter (`c_counter_t`), corresponding to the number of clock cycles during which triggered events are counted, the chip automatically returns the number of discriminator hits recorded in each time window. The resulting discriminator occupancy as a function of threshold value forms a characteristic S-curve, which is fitted with an error function. From this fit, the baseline threshold position (μ) and noise level (σ) are extracted. An example of such an S-curve is shown in Figure 3.22.

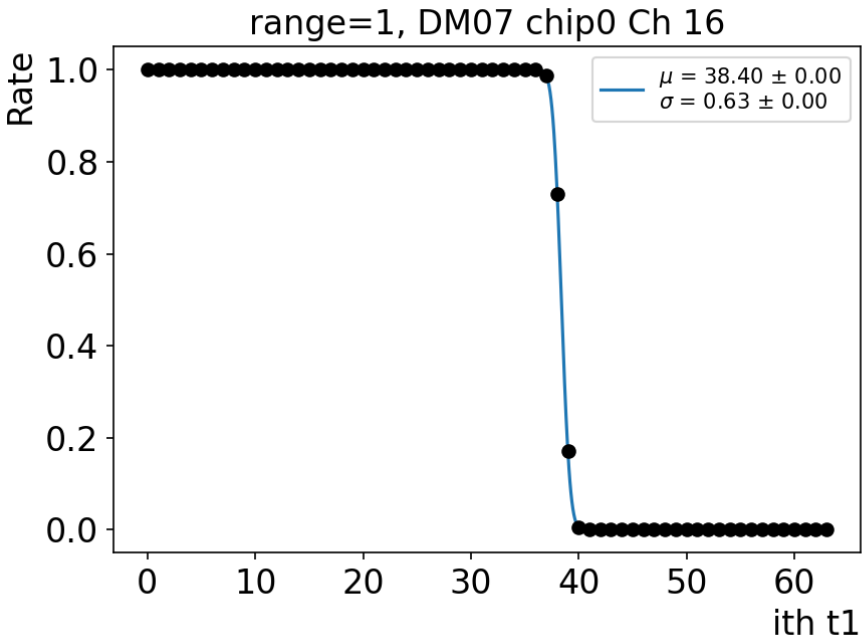

Figure 3.22: Threshold scan for the T1 branch and corresponding error function fit.

The extracted calibration constants are stored and later used during the configuration of the TOFHIR front end when setting the discriminator thresholds. The effective threshold applied to a given branch is thus defined as:

$$\mathrm{th}_{\mathrm{effective}} = \mathrm{th}_{\mathrm{user}} + \mathrm{baseline}.$$

where the $\mathrm{th}_{\mathrm{user}}$ is the user provided value, the baseline is the one found with the threshold calibrations, and the assumption of a linear scaling of the threshold is made, so that we can just sum the baseline to the user specified value.
Following the calibration, threshold scans are compared to reference measurements obtained during single-module characterization at LIP. A strong correlation between the baseline positions and noise values is expected, with noise typically observed to be slightly

lower in the tray setup due to the reduced electrical noise in the test environment. Figure 3.23 shows an example of the noise and baseline correlations between the tray validation and LIP measurements.

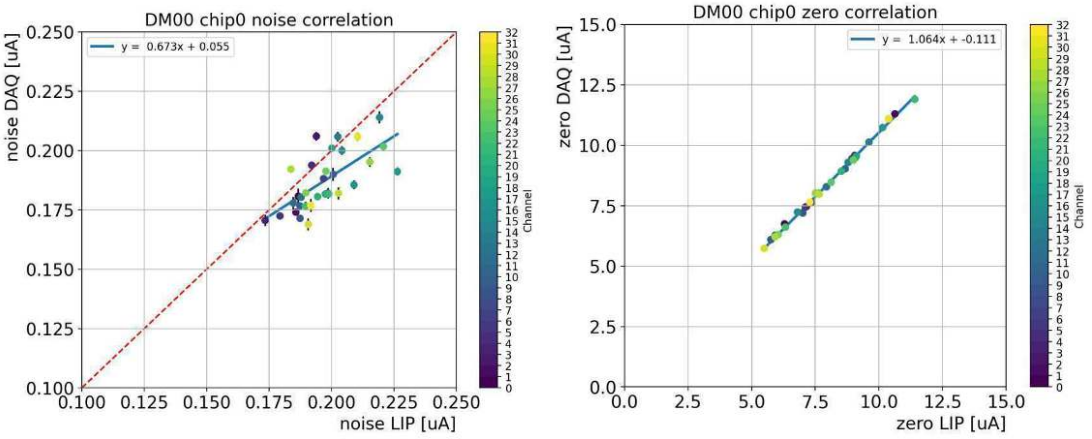

Figure 3.23: Tray validation through threshold scans. Left: correlation of the noise values with those measured at LIP. Right: comparison of the baseline threshold positions.

Time Calibration and TDC Validation

Precise time measurements in the TOFHIR ASIC rely on a Time-to-Digital Converter (TDC) architecture. Each discriminator branch (T1 and T2) provides a time measurement composed of two parts: a coarse counter, which counts full clock cycles, and a fine counter (`t1fine` or `t2fine`), which measures the fraction of a clock cycle by integrating a current.

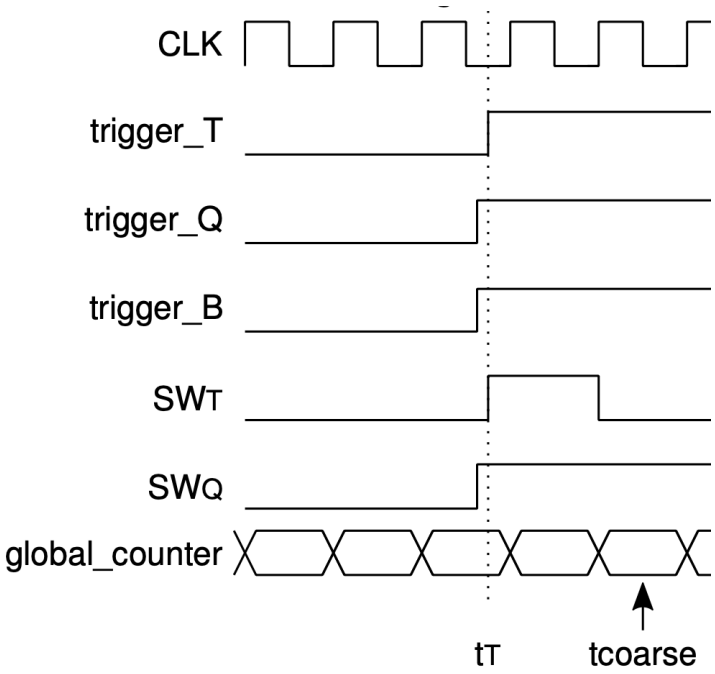

Figure 3.24: Time measurement operation with respect to the clock signal.

As shown in Figure 3.24, when a trigger signal (t_T) is received, the corresponding switch (SW_T) closes, initiating the integration of a current on the Time-to-Amplitude Converter

(TAC). The integration continues until the next falling and rising clock edges are encountered, at which point the switch opens, the accumulated charge is digitized, and the coarse time is registered. The minimum `tfine` value is obtained when the trigger occurs just before the clock falling edge, corresponding to an integration time of 0.5 clock cycles. Conversely, the maximum `tfine` value occurs when the trigger follows immediately after the falling edge, yielding a total integration of 1.5 clock cycles.
A dedicated calibration procedure is required to map the raw `tfine` values into precise time delays, expressed in fractions of the TOFHIR clock period. This is accomplished using external digital test pulses generated by the lpGBT. These pulses are synchronous with the TOFHIR clock and feature a programmable delay, enabling precise control of the input timing. The calibration procedure proceeds as follows:

- External test pulses are injected from the lpGBT into the TOFHIR input channels.
- The programmable delay of the lpGBT is varied in steps of 97.6 ps, from 0 up to 64 steps (6.24 ns), covering the entire 6.25 ns clock cycle.
- For each delay setting, the corresponding `tfine` value is recorded.
- The distribution of `tfine` as a function of the injected delay is analyzed to extract the calibration curve.

Since the test pulses are synchronous with the TOFHIR clock, the integration time is related to the applied delay by:

$$\text{int. time} = 1 - \text{delay}, \quad \text{with int. time} \in (0.5, 1.5)\,\text{clock cycles}.$$

The relationship between the fine time and the integration time is modeled with a second-order polynomial:

$$\text{tfine} = a_2 \cdot (\text{int. time} - t_0)^2 + a_1 \cdot (\text{int. time} - t_0) + a_0.$$

This mapping is then inverted to compute the calibrated fine-time component:

$$f(\text{tfine}) = \frac{-a_1 + \sqrt{a_1^2 - 4a_2(a_0 - \text{tfine})}}{2a_2} + t_0,$$

allowing the absolute time of the event to be reconstructed as:

$$t = t_{\text{coarse}} - f(\text{tfine}),$$

with a periodicity of one clock cycle and a synchronization offset fixed at 0.5.
The calibration results are generally consistent with those obtained during the single-module characterization at LIP, with only a small offset observed in the t_0 parameter. Figure 3.25 shows an example of the TDC calibration for all TACs and both timing branches, where the red dashed curve corresponds to the original LIP calibration, the green dashed line to the new DAQ calibration, and the red dotted line to the LIP calibration shifted by the DAQ t_0 offset.

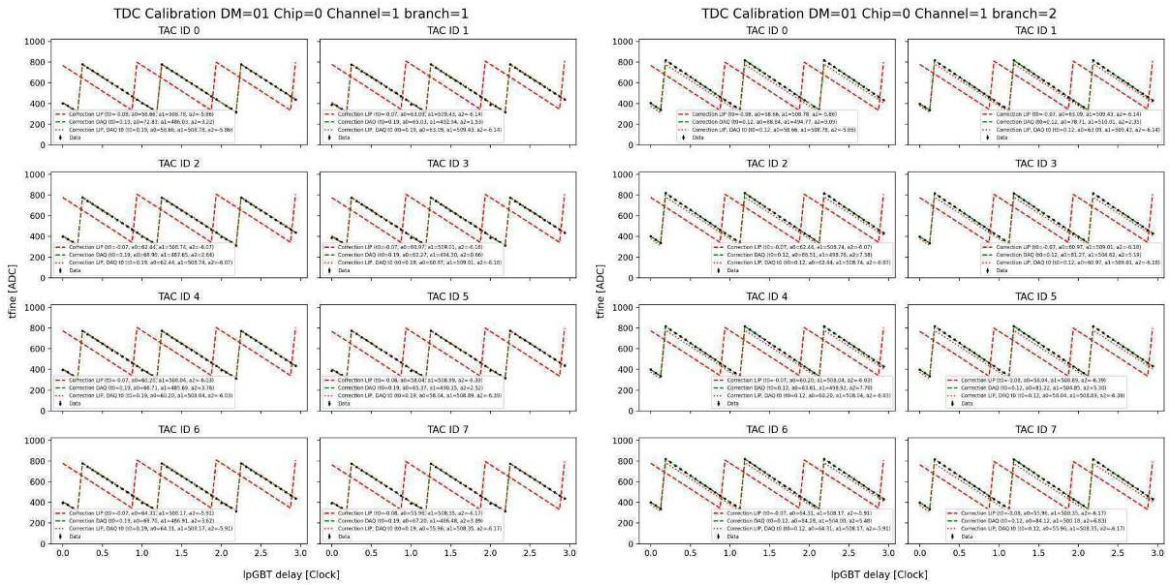

Figure 3.25: lpGBT test pulse delay scan for TDC calibration for a single channel, showing all TACs and the two timing branches.

Once the calibration constants are determined, the TDC performance is validated through dedicated digital test-pulse runs. In this phase, synchronous digital pulses are distributed to all channels, and the reconstructed times of arrival (ToA) are compared between pairs of channels. For each channel, the ToA is measured with respect to a fixed reference (the `RESYNC` command), and the distribution of time differences is fitted to extract the per-channel resolution. The expected performance requirement is a channel-level timing resolution below 25 ps, consistent with the intrinsic precision of the TOFHIR ASIC. It may happen that a channel exhibits a time resolution worse than 25 ps because of an improper calibration. The ToA is a sum of gaussians, one from each TAC, where each of gaussians might have shifted mean compared to the others, but a proper time resolution. This is the case for two channels of the map below.
An example of the validation results is shown in Figure 3.26, where the top panel displays examples of the ToA distribution for a single channel relative to a reference, the bottom panel shows the map of timing resolutions across the RU.

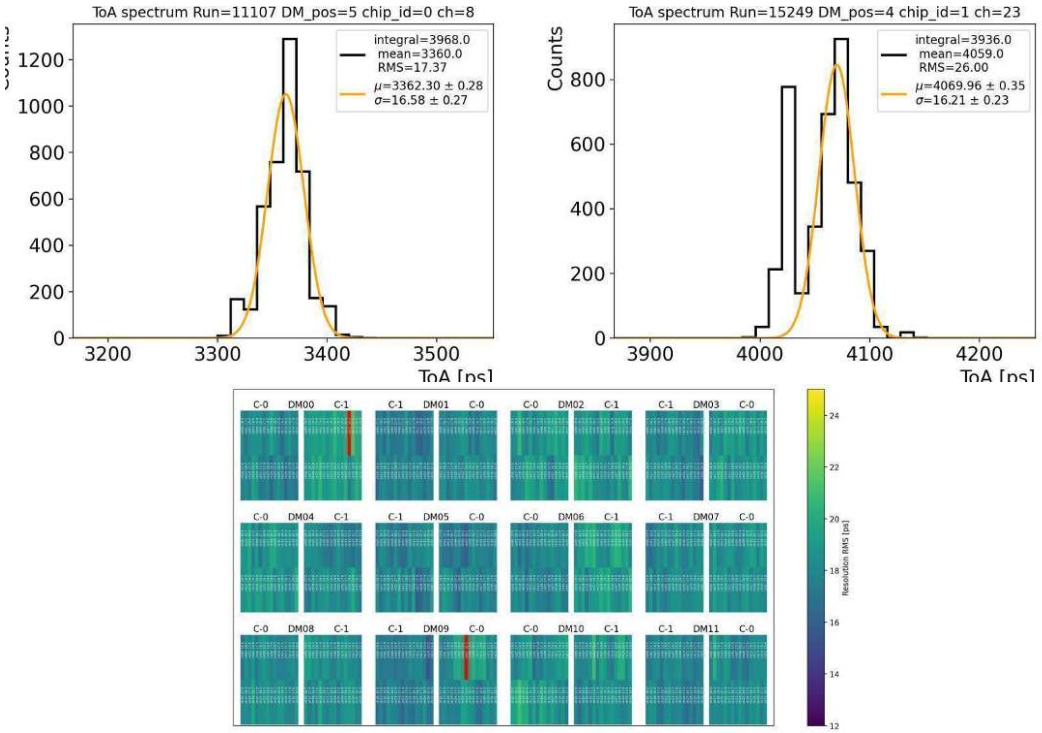

Figure 3.26: Tray validation through TDC resolution. Top left: time of arrival for a single channel relative to a reference; top right: an example of a miscalibrated TAC with a proper RMS per TAC but wrong offset; bottom: resolution map across all channels in one RU.

Energy Calibration and Validation

The energy calibration of the TOFHIR front-end ASIC is a multi-step procedure designed to ensure a uniform and stable energy response across all channels of a tray. It includes the calibration of the DC current subtraction (I_{trim}), the determination of pedestals, the inter-calibration of the energy scale using the Front-End Test Pulse (FETP), and the validation of the overall energy response through long LYSO runs. All these steps are performed using the `mtd_daq` software in conjunction with the Serenity back-end platform.

I_{trim} and Pedestal Calibration In the TOFHIR ASIC, the energy of a signal is obtained by integrating the input current using a charge-to-amplitude converter (QAC) with a configurable integration window. The measured amplitude is affected by two main contributions:

- a DC current component that is integrated together with the signal, and
- an offset (the pedestal) that must be subtracted.

Each channel employs eight QACs operating in a round-robin fashion to minimize dead time. While the DC current component is common to all QACs of a given channel, the pedestal may vary from one QAC to another. Therefore, pedestal calibrations are performed for all QACs individually, whereas the I_{trim} calibration is common to the channel.

The TOFHIR provides a tunable DC bias current subtraction through the I_{trim} parameter. The optimal value of I_{trim} for each channel is determined by injecting digital test pulses—generated internally by the front-end board—while biasing the SiPMs so that they do not produce any analog signal. In this configuration, the only current contribution at the TOFHIR input is the DC component. The integration window is scanned while varying the I_{trim} setting, and the resulting integrated charge is measured. For small I_{trim} values, the integrated charge increases with the integration window (positive slope), while for larger settings the slope becomes negative. The optimal calibration point corresponds to the smallest positive slope, indicating an accurate DC current subtraction without affecting the signal.
Once the I_{trim} is determined, the integrated charge as a function of the integration window is fitted with a third-degree polynomial to extract the pedestal value. Figure 3.27 shows an example of a successful I_{trim} calibration compared with the reference from the LIP QA/QC center, along with the resulting QDC (pedestal subtraction) fit.

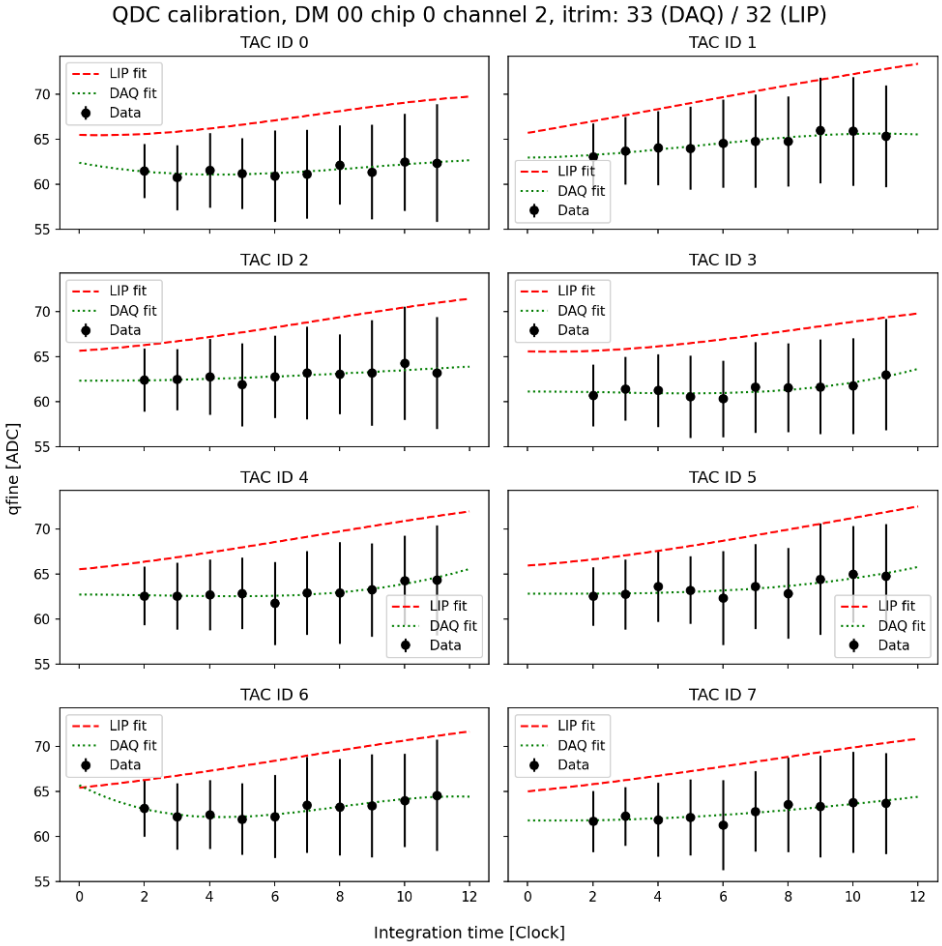

Figure 3.27: Result of I_{trim} and QDC calibration. The DAQ determines a slightly higher value (33) than the LIP reference (32) for the DC trim, yielding a smaller positive slope as expected. The green line shows the fitted QDC response.

Energy Inter-Calibration via Front-End Test Pulse (FETP) After pedestal subtraction, an inter-calibration procedure equalizes the energy response across all TOFHIR channels. This is achieved using the Front-End Test Pulse (FETP), which injects analog

signals of varying amplitude directly into the channels, allowing a controlled measurement of the energy response.
For each channel, the calibrated energy is computed from the `qfine` observable, the integral of the charge over time with the pedestal removed. By scanning the FETP amplitude, the relation between the measured and expected calibrated energies is established, where the expected energy is defined as the average response across all channels for a given pulse amplitude. A linear fit of the expected energy versus the measured energy provides a correction function that maps the measured calibrated energy to the inter-calibrated scale. This procedure ensures uniform energy response across channels, minimizing channel-to-channel variations and improving the global energy resolution.
An example of the inter-calibration across a single RU is shown in Figure 3.28. The left panel displays the measured versus expected calibrated energy before correction, while the right panel shows the improved uniformity after applying the calibration constants.

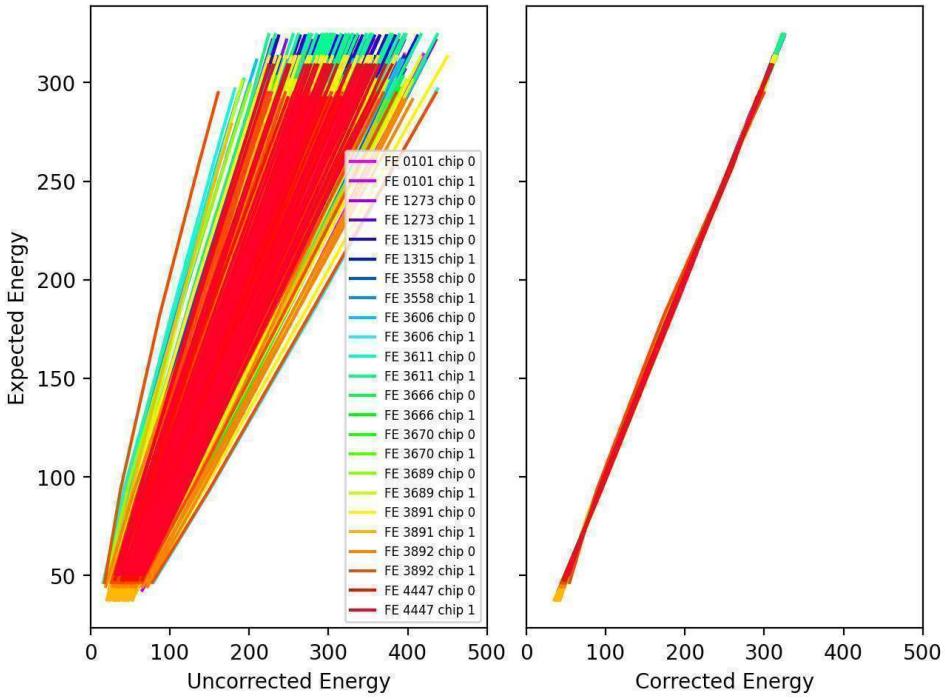

Figure 3.28: Front-End Test Pulse (FETP) inter-calibration. Left: calibrated energy vs. expected energy. Right: corrected and inter-calibrated energy vs. expected energy. Each line corresponds to a single channel within one RU.

LYSO Validation Runs The energy calibration procedures described above are validated through long data-taking runs using LYSO scintillation signals. Natural radioactivity from lutetium in LYSO produces two characteristic peaks at 202 keV and 305 keV [139], which serve as in-situ calibration markers. During these runs, the positions of the LYSO peaks are monitored across channels within each SM and across the full RU. The data can be acquired using the reference LIP calibrations or with the custom calibrations derived from the MTD DAQ. The latter typically yield a smoother and more uniform distribution of peak positions, demonstrating improved channel equalization.
Figure 3.29 illustrates the outcome of these LYSO runs. The panels on the left correspond to runs acquired with the LIP calibration constants, while those on the right show the improved uniformity achieved using the DAQ-derived calibrations. Bars or channels with

peak positions deviating by more than 20% from the expected value are flagged for further inspection.

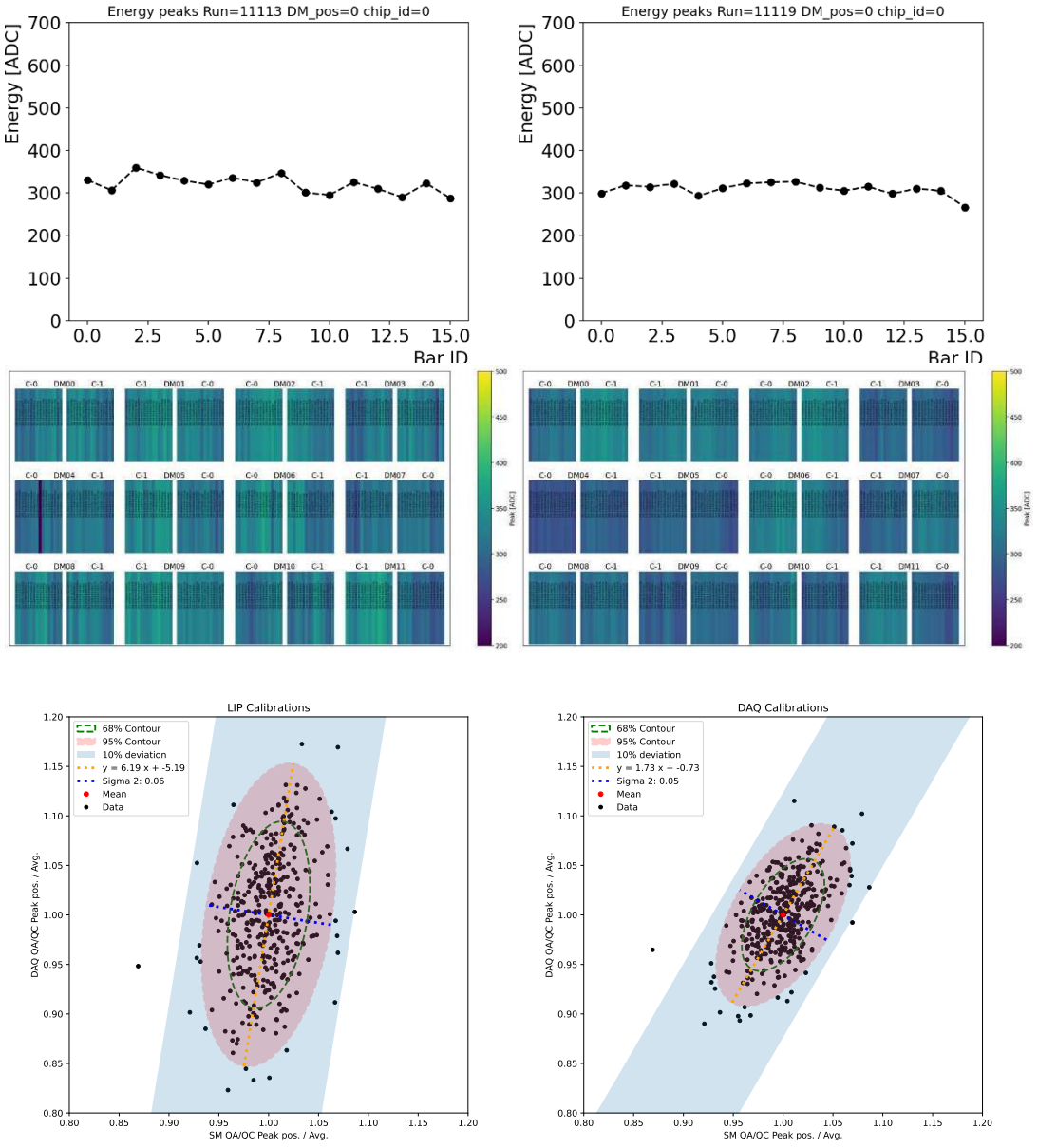

Figure 3.29: Tray validation through long LYSO runs for each single bar. Left: results with LIP calibrations. Right: improved uniformity with custom DAQ calibrations. The smoothing achieved using the DAQ-derived constants is evident. On top: peak position distributions for a single SM; middle: peak position maps across the RU; bottom: correlation of peak positions across the RU with the SM QA/QC results, assuming a 2D Gaussian we report 68% and 95% contour and check for bars outside the 10% range (light blue).

Together, the I_{trim}, pedestal, and FETP calibrations, validated through LYSO runs provide a complete characterization of the energy response for each tray. Only trays that meet the performance and uniformity requirements are certified for integration into the detector.

3.6 Testbeam Data Acquisition

In September 2025, a dedicated test beam campaign was carried out at the CERN SPS to validate the performance of the `mtd_daq` system in realistic beam conditions. This represented the first large-scale test of a good part of the full data acquisition chain, from the Serenity board to the CCs, Front-End (FE) modules, and TOFHIR ASICs, operating with particle signals rather than calibration pulses. The primary objectives were to demonstrate stable operation of the DAQ during beam spills, to verify the synchronization of the readout with external triggers, and most importantly to perform the ultimate time resolution measurement with production like components.
The test was carried out on the H8 beamline, where a 180 GeV pion beam with a spill structure of roughly 10 s every 30 s was available. Two BTL RU with only one DM attached to each were installed and read through the entire DAQ chain. A scintillator detector upstream of the RUs provided the external trigger reference, while two downstream MCPs provided the absolute time reference.

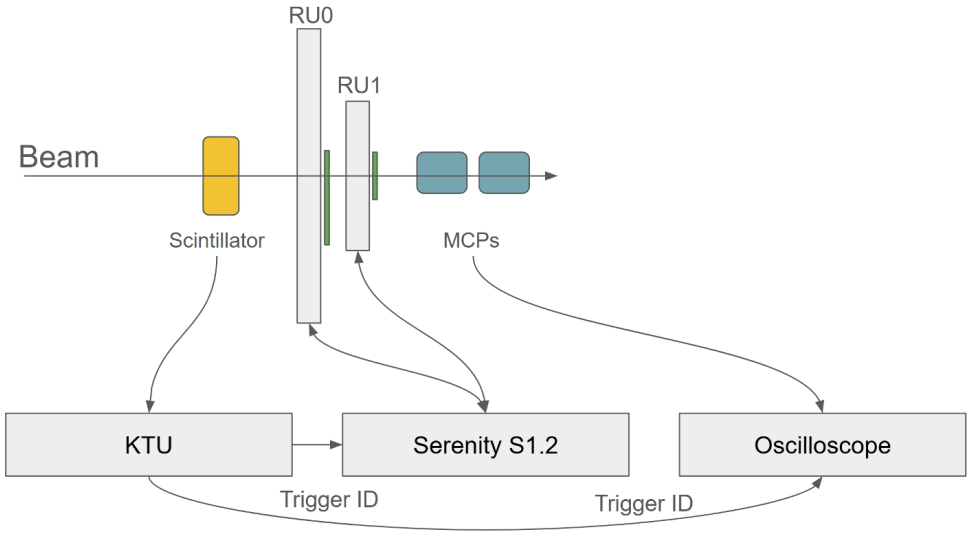

Figure 3.30: Testbeam setup with KCU105 Trigger Unit, Serenity S1.2 and Oscilloscope backend.

The DAQ software was employed to manage the typical run structure (run initialization, spill start/end, run termination) and to read data continuously. During the two week long campaign, the system demonstrated sustained operation at a throughput of up to 6.1 Mbit/s. The low throughput (lower than the production rate by a factor 1000) was due to a missing key component of the DAQ chain: the DTH. The latter is supposed to convey data from 6 different backend boards (Serenity), each with a rate of 100 Gbit/s. Previous test-beam studies [140] have shown that the expected single-bar time resolution at normal incidence ($\theta = 0°$) is approximately 30 ps. This value serves as a reference benchmark for the present measurement. In particular, the time resolution is known to depend both on the applied SiPM overvoltage and on the angle of incidence of the traversing particle. 3.31 illustrates the dependence of the resolution on the over-voltage, as well as its variation as a function of the incident angle. These results highlight that given the testbeam operating condition, over-voltage set to 3V and incident angle of 0°($|\eta| = 0$), 30ps should be achieved.

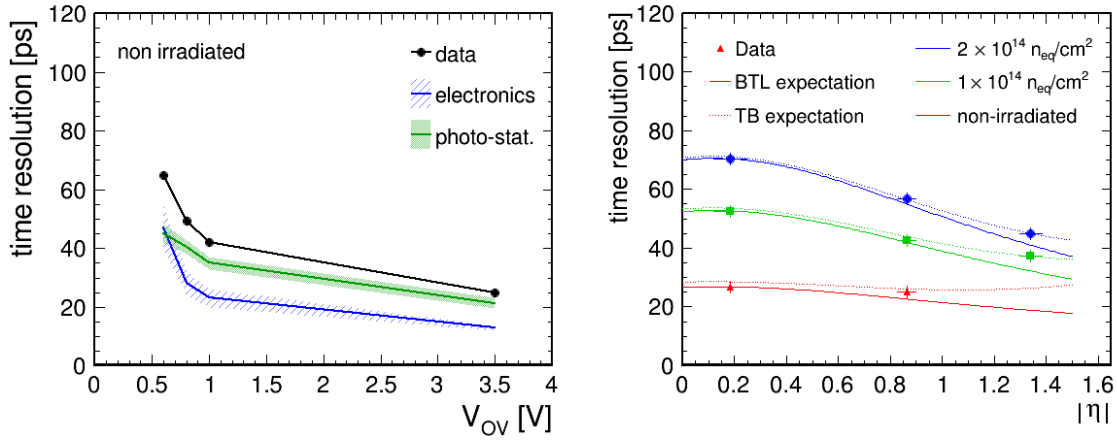

Figure 3.31: Left: BTL time resolution for different incident angles at different overvoltages. Right: expected time resolution at different $|\eta|$ in BTL.

3.6.1 System Synchronization

A critical aspect of the September test beam campaign was the synchronization of the readout chain across multiple Readout Units (RUs). In the test setup, two TOFHIR chips were operated in parallel, each mounted on a different RU exposed to the beam. Since the final MTD system will involve thousands of channels distributed over many RUs, establishing and validating a reliable synchronization procedure was a central goal of this test.

Resynchronization Signal

Synchronization of the two TOFHIRs was achieved using a dedicated short resynchronization (*resync*) signal. This signal was distributed simultaneously to both RUs via the optical fibers connecting the Serenity board to the CCs. When received, resync enforced the alignment of the internal counters and ensured that subsequent data frames from the two TOFHIRs could be consistently matched.

Event Structure and Timing Information

Each TOFHIR readout frame is encapsulated in a 128-bit word. In addition to the hit information (energy and timing fine counters), these words contain two global identifiers:

- **Timetag:** a coarse timestamp counter that increases at the LHC bunch 0 crossing frequency (89.1 μs) and serves as a global time reference for the readout.
- **Trigger ID:** an event identifier associated with the external trigger.

For synchronized operation, the `trigger ID`, `timetag`, and `t1coarse` counters must evolve consistently between the two RUs. This condition ensures that events recorded by the two TOFHIRs can be unambiguously aligned during offline reconstruction.
The data frames read out from the TOFHIRs are stored in separate buffers within the FPGA memory, each capable of holding up to 16 383 words of 128 bits. To maintain synchronization at the buffer level, the FPGA implements a dedicated function that resets the buffer address pointer after each download. Reading out the buffers over many

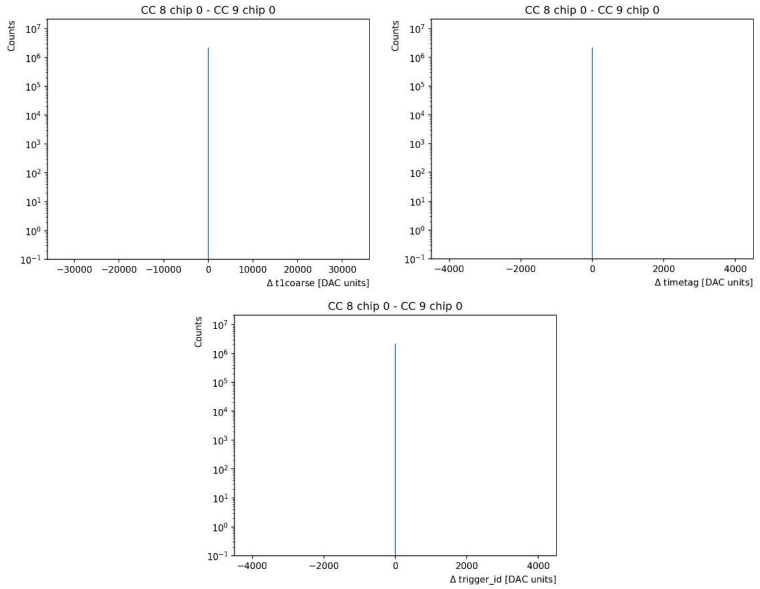

Figure 3.32: Synchronization validation during testbeam. Top left: t1coarse alignment between the two TOFHIRs. Top right: timetag alignment between the two TOFHIRs. Bottom: trigger ID alignment between the two TOFHIRs.

cycles, events are written starting from the beginning of the buffer at the end of every cycle, ensuring consistent alignment between the two data streams.
The number of cycles to read events in a spill (5.5 s) was fine-tuned to fill the entire spill duration and also to ensure alignment of the events. The final number of cycles was 8 corresponding to 131 064 events for each Tofhir.

Validation During the Testbeam

During data taking, the synchronization was validated by comparing t1coarse, the timetag and trigger ID fields in the output streams of the two RUs. As illustrated in Figure. 3.32, the timetag counters remained aligned for the duration of the 10 spills. Similarly, t1coarse and the trigger IDs showed one-to-one correspondence between the two devices, confirming that external trigger signals were distributed consistently.
This successful validation demonstrated that the MTD DAQ framework can operate two independent RUs in a fully synchronized mode, a crucial milestone for scaling to the full BTL system. It also provided the first confirmation that the combined use of resync signals and embedded timing metadata (timetag and trigger ID) is sufficient to guarantee event alignment in realistic beam conditions.

3.6.2 Data Analysis

The analysis of the testbeam data began with the decoding of the raw output words produced by the TOFHIR ASICs. Each readout frame, described in Table 3.1, is encapsulated in a 128-bit word containing the hit information together with timing and trigger metadata. From this raw information, reconstructed hits were obtained through a series of calibration and grouping steps.

Hit Calibration

Each hit was calibrated using the dedicated timing and charge calibration constants derived from the TOFHIR test-pulse scans:

- **TDC calibration:** the `t1fine` and `t2fine` fields were mapped to precise time offsets using the calibration curves described in Section 3.5.2. This procedure converts the raw fine-time counters into physical time units, allowing the reconstruction of hit times with sub-nanosecond precision.
- **QDC calibration:** the charge measurement (`qfine`) was corrected using the energy inter-calibration described in Section 3.5.2, mapping the raw integral of the waveform to an energy scale consistent across all channels.

After this step, each entry in the dataset corresponded to a calibrated hit, characterized by an absolute time, an energy estimate, and the associated channel identifier.

Event Building Across TOFHIRs

Since the testbeam setup involved two TOFHIRs operating in parallel on different RUs, hits from both devices needed to be combined into common events. This was achieved by grouping hits according to three synchronization keys:

- the `trigger ID`, distributed by the external scintillator trigger system;
- the `timetag`, providing the coarse timestamp reference;
- the `cycle number`, corresponding to the internal DAQ cycle.

Within each group, hits were further clustered in time using a sliding window of 20 TOFHIR clock cycles, corresponding to 125 ns. Hits falling within this window were considered to belong to the same physical event. This procedure ensured that signals from the two TOFHIRs produced by the same beam particle were consistently merged, even in the presence of intrinsic timing fluctuations.

Bar Hit Reconstruction

Within each reconstructed event, the signals on the left and right channels of a LYSO bar were combined in order to form a *bar hit*. A bar hit was defined whenever a pair of hits, one from the left channel and one from the right channel of the same bar, were found to be time-compatible within one TOFHIR clock cycle (6.25 ns).
For each bar hit, the calibrated energy was obtained as the sum of the energies measured on the two channels, while the hit time was defined as the average of the two arrival times:

$$E_{\text{bar}} = E_{\text{left}} + E_{\text{right}}, \tag{3.3}$$

$$t_{\text{bar}} = \frac{1}{2}\left(t_{\text{left}} + t_{\text{right}}\right). \tag{3.4}$$

This procedure suppressed spurious single-channel hits and provided a more robust estimate of both the deposited energy and the interaction time of the traversing particle.

Energy Spectrum of Bar Hits

The distribution of E_{bar} for all reconstructed bar hits is shown in Figure 3.33. As expected, the spectrum exhibits a Landau-like peak corresponding to the energy deposited by minimum-ionizing particles (MIPs). In addition to the MIP peak, two further features can be identified:

- A low-energy peak originating from optical cross-talk between adjacent channels, producing spurious coincidences below the MIP scale.
- A high-energy tail associated with showering events in the material upstream of the detector, leading to multiple-particle hits in the same bar. This tail if strongly suppressed by the requirement of having less than 3 bar hits in the same event.

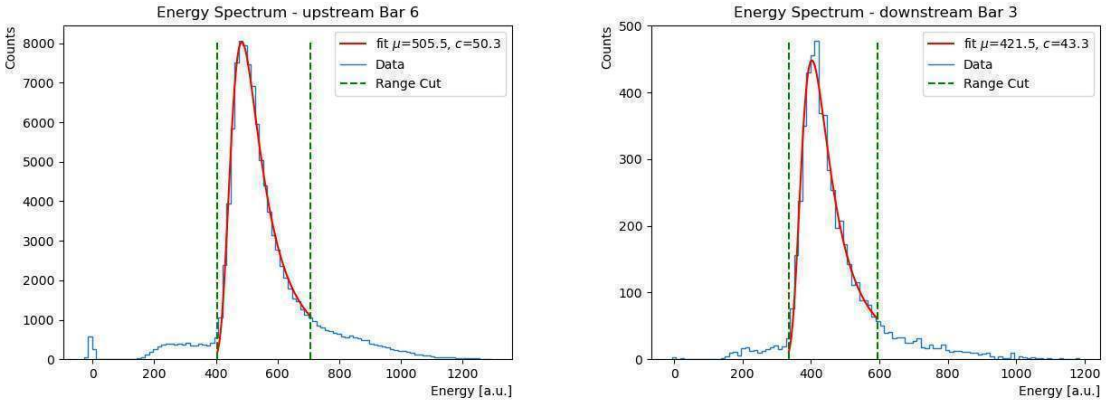

Figure 3.33: Bar spectra upstream (left) and downstream (right)

3.6.3 Time Resolution Measurement

The ultimate performance metric of the system is the achievable time resolution. In the testbeam, two independent methods were employed to evaluate it:

- the time difference between the left and right channels of the same scintillator bar;
- the time difference between two bars aligned along the beam axis (upstream-downstream configuration).

In both cases, the intrinsic resolution of a single measurement can be extracted from the width of the time-difference distribution, after applying dedicated corrections to remove systematic dependencies.
The time difference is computed only on bar hits that fall under the MIP peak to simplify the analysis.

Energy-Ratio Correction

For the left-right configuration, the raw time difference $\Delta t = t_{\text{left}} - t_{\text{right}}$ exhibits a dependence on the relative signal amplitudes of the two channels. This effect originates from residual differences in the photon collection and SiPM response. The same argument applies for the upstream-downstream configuration, where differences in light collection,

SiPM response and position of the hit inside the bar between the two crystals can introduce a systematic bias in the time difference. To correct for this, the time difference was studied as a function of the ratio of the measured energies $E_{\text{left}}/E_{\text{right}}$. For the time difference over an upstream and the downstream bar, the energy correction is calculated as a function of $E_{\text{upstream}}/E_{\text{downstream}}$. The distribution was profiled in bins of the energy ratio over the time difference, and the observed dependence was parameterized. The correction was then applied event-by-event, removing this systematic bias. Figure 3.34 illustrates the effect of this correction on the time-difference distributions. This correction has a mild effect on the time resolution in the same bar case, while it is more significant in the upstream-downstream configuration, where the light collection and SiPM response can differ more substantially between the two bars.

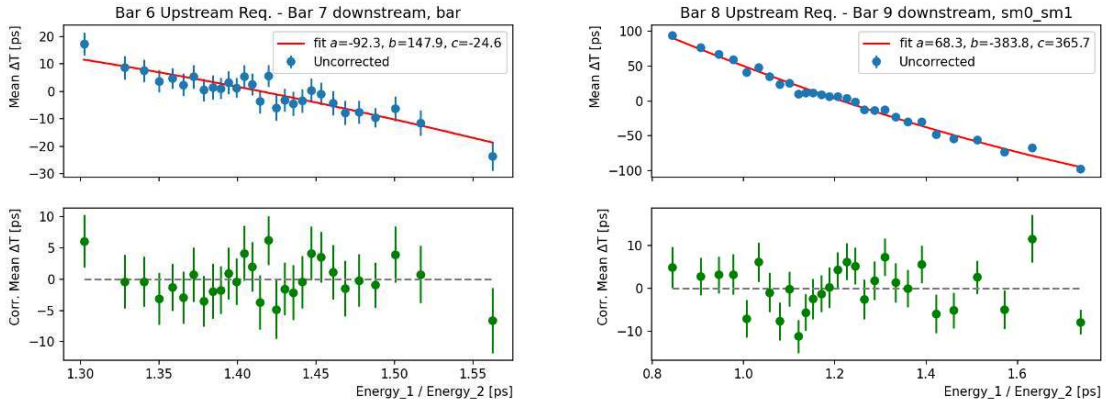

Figure 3.34: Energy Ratio correction for time difference in the same bar (left) and an upstream bar with respect to a downstream bar

Phase Correction

A second correction is applied to account for the dependence of the reconstructed time on the signal phase with respect to the sampling clock. The TOFHIR provides fine-time information through the `t1fine` counter, which is affected by intrinsic phase variations that modulate the effective threshold crossing. This introduces a residual time bias that appears as a periodic structure in the measured time difference Δt.
To correct for this effect, the average of the left and right `t1fine` values is used as a phase estimator. The dependence of Δt on this average phase is profiled, and the resulting modulation is removed through a fitted spline. Since the amplitude dependence of this effect can vary, the correction is derived separately for each over-voltage and threshold configuration. In the Upstream-Downstream configuration, the correction is computed independently as a function of the average `t1fine` upstream and then downstream.
This phase dependence is attributed to periodic activity of the sampling clock inside the chip, which induces an oscillation on the post-amplifier baseline. When the signal pulse is superimposed on this oscillating baseline, the effective threshold shifts periodically, inducing an additional time-walk-like contribution. As shown in Fig. 3.35, applying this correction significantly improves the time-difference resolution in the Upstream-Downstream configuration, where the two ends are read out by different chips on separate RUs. Within the same bar, the effect is smaller since both SiPMs are processed by the same TOFHIR and RU.

It is worth noting that this periodic phase effect is not expected during standard detector operation, as the BTL readout will run at a fixed phase determined by the bunch-crossing timing.

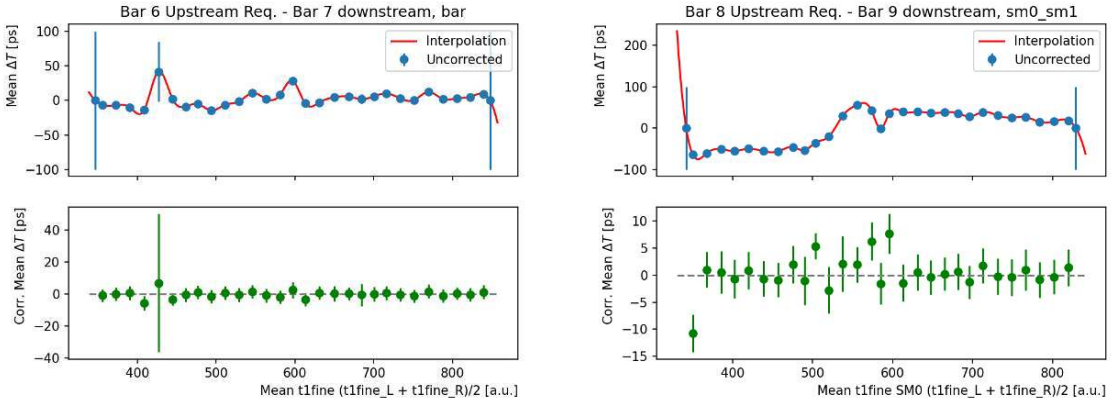

Figure 3.35: Phase correction for time difference in the same bar (left) and upstream bar - downstream bar

Extraction of the Resolution

After applying the energy-ratio and phase corrections, the distribution of the time differences was fitted with a Gaussian function. The single bar time resolution was then obtained as:

$$\sigma_t = \frac{\sigma_{\Delta t, L-R}}{2}, \tag{3.5}$$

with the assumption thant the left and right channel resolutions are independent, and

$$\sigma_t = \frac{\sigma_{\Delta t, \text{Upstream}-\text{Downstream}}}{\sqrt{2}}, \tag{3.6}$$

where $\sigma_{\Delta t}$ is the standard deviation of the corrected time-difference distribution.
Figure 3.36 shows the time-difference distributions after each correction, along with the Gaussian fits used to extract the resolutions.

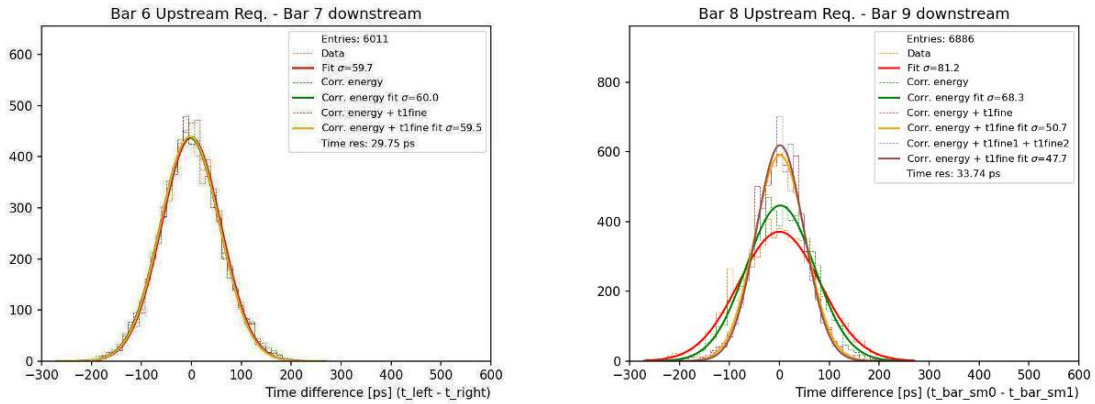

Figure 3.36: Time difference with different level of corrections, fitted with a Gaussian model. Left is Left-Right, right is upstream bar - downstream bar

Results

In the left-right configuration, a time resolution of $\sigma_t = 31.1$ ps was achieved, consistent with the design expectations for the BTL channels. The upstream-downstream configuration provided an independent cross-check, yielding a resolution of $\sigma_t = 33.0$ ps (Figure 3.37). The time resolution between two bars is expected to be higher than the single bar measurement due to the clock propagation to different RUs.
The difference between the time resolution in the same bar and across different bars is measured to be 11.03 ps, while the expected difference due to clock distribution was 5 ps. The unexpected difference will be further studied by both the BTL and DAQ teams with laser measurements and scope.
These results demonstrate that the full DAQ chain, including synchronization, and event reconstruction, is capable of delivering the timing performance required for the MTD at the HL-LHC.

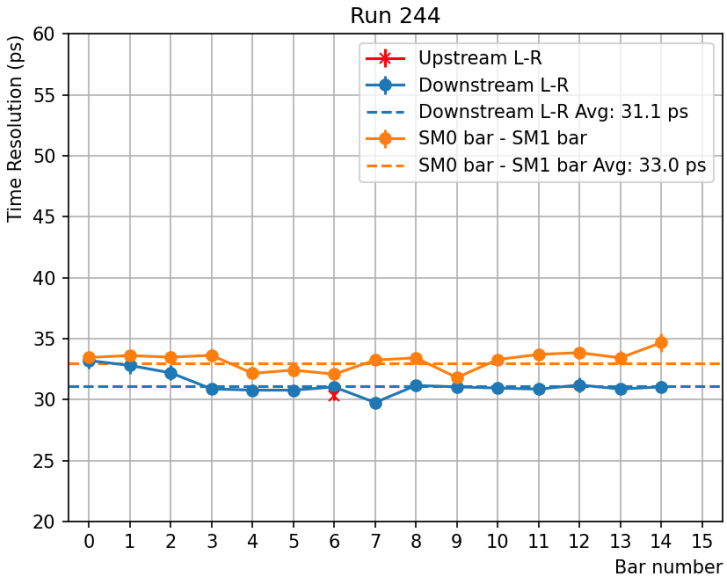

Figure 3.37: Time resolution over the different bars

3.7 Summary

The MTD DAQ system has evolved into a complete and reliable framework for the operation, calibration, and validation of the Barrel Timing Layer (BTL) detector. The system combines firmware and software components developed around the Serenity platform, which serves as the back-end electronics for both the BTL and the Endcap Timing Layer (ETL). Within this unified architecture, the MTD DAQ provides a consistent and modular interface to control all front-end ASICs — including the lpGBT, GBT-SCA, and TOFHIR — ensuring full compatibility with the final Phase-2 CMS back-end infrastructure.
A major milestone for the MTD DAQ has been its successful deployment at the BTL Assembly Centers (BACs), where it is routinely used to validate fully assembled trays. In this context, the DAQ framework handles the communication with the complete readout chain, from the Concentrator Cards down to the individual Sensor Modules. Through

automated routines, it performs essential functionality checks of the front-end components, such as reading temperatures and currents via the lpGBT and SCA interfaces, and measuring the I-V characteristics of the SiPM arrays to determine the breakdown voltage and operating over-voltage. Once these preliminary tests are completed, the system executes a full suite of calibration procedures — including the discriminator (DISC) or threshold calibration for the T1 timing branch, the time-to-digital converter (TDC) calibration for fine-time alignment, and the energy calibration to inter-calibrate the channel response. This validation chain ensures that every tray satisfies the performance and stability criteria before being certified for integration into the detector.

In parallel, the MTD DAQ was successfully employed during the September 2025 test-beam campaign, marking the first large-scale operation of the system outside the laboratory environment. During these tests, the Serenity board was used to read out two TOFHIR chips connected on separate optical fibers, demonstrating the capability to acquire synchronized data from multiple front-end paths. The runs confirmed that timing alignment between the two TOFHIRs could be achieved with new synchronization procedures implemented in the DAQ firmware and software. Although the absence of the DTH limited the maximum throughput during the campaign, several dedicated runs collected sufficient statistics to study the intrinsic time resolution of individual BTL bars. These measurements validated the correct functioning of the DAQ chain and provided valuable feedback for future firmware and DAQ optimizations.

Overall, the MTD DAQ system has proven to be a robust and scalable solution, fully compatible with the CMS Phase-2 architecture and capable of supporting both production-level quality control and detector performance studies. Its modular design ensures smooth integration with the final DTH-based back-end during commissioning, while its demonstrated reliability at the Assembly Centers and in test beam confirms its readiness for large-scale operation during the forthcoming installation and commissioning phases of the CMS MTD.

Chapter 4

Conclusions

This thesis presented two complementary research activities carried out with the CMS experiment at the CERN Large Hadron Collider: the measurement and interpretation of Vector Boson Fusion production of the Z boson, and the development of the Data Acquisition system for the Barrel Timing Layer of the MIP Timing Detector, one of the key upgrades for the Phase-2 CMS detector for High Lumi LHC.

VBF-Z Analysis and EFT Interpretation I served as the main analyzer for the measurement of the VBF production of the Z boson using the full Run 2 dataset collected by CMS at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 138 fb^{-1}. The analysis measured the inclusive, fiducial, and differential cross sections of the electroweak Zjj process, unfolding the results to particle level for a set of observables that characterize the VBF topology. A dedicated DNN-based discriminant was trained to enhance the separation between the electroweak signal and the Drell-Yan background, allowing a precise extraction of the signal strength.
The analysis has reached an advanced stage within the CMS review process, currently undergoing the official approval for both the Standard Model measurement and its Effective Field Theory interpretation. In the latter, the VBF-Z channel plays a key role in the global CMS combination of processes sensitive to dimension-six operators in the Warsaw basis, providing the leading sensitivity to the bosonic coefficients c_{HWB} and c_{HDD}. The final results will contribute to the CMS public documentation and to the forthcoming SMEFT combination of electroweak measurements.

MTD DAQ Development and Commissioning In parallel, I am the main developer of the MTD DAQ software, the official Data Acquisition framework for the BTL. The software was designed to control and monitor the full readout chain, including the TOFHIR front-end electronics, lpGBTs, and GBT-SCAs, and to automate calibration and validation procedures across multiple hardware configurations. Over the course of this work, a mature and stable version of the DAQ was achieved, which is now routinely used at the BTL Assembly Centers for large-scale tray validation and quality-control operations.
The same DAQ system was successfully deployed during the BTL test beam campaigns, where it was used to perform timing, energy, and calibration studies under realistic running conditions. These results demonstrate the robustness and versatility of the framework, which is expected to remain the reference software for BTL integration and com-

missioning activities in the coming years.

Outlook The VBF-Z study contributes to the understanding of electroweak interactions and provides a sensitive probe for new physics through the EFT framework, while the MTD DAQ development ensures that the next-generation CMS detector will achieve the timing precision required to fully exploit the physics potential of the High-Luminosity LHC. Together, these contributions represent an important step toward the future of precision measurements and detector performance within the CMS experiment.
Looking ahead, both the analysis and detector-development components of this work naturally open the way to future improvements and extensions. On the physics side, the precision of the differential measurement in the VBF-Z channel is currently limited by the statistical uncertainty of the Run 2 dataset. The inclusion of Run 3 data, corresponding to an integrated luminosity of approximately 300 fb^{-1}, will substantially increase the available statistics, enabling more precise unfolded measurements and improving the sensitivity to potential SMEFT effects through refined differential fits.
On the detector side, the MTD DAQ activities will continue toward the integration of the Data and Timing Hub in the full readout chain, providing a complete demonstration of the end-to-end data path from the front-end electronics to the back-end system. This would allow tests of fully assembled BTL trays with cosmic rays within the coming year, representing a key validation step prior to CMS installation. These forthcoming efforts will consolidate both the performance validation and operational readiness of the BTL system for the HL-LHC era.

Bibliography

[1] Paul A. M. Dirac. "The Lagrangian in quantum mechanics". In: *Phys. Z. Sowjetunion* 3 (1933), pp. 64–72.

[2] Richard Phillips Feynman and Albert Roach Hibbs. *Quantum mechanics and path integrals*. International series in pure and applied physics. New York, NY: McGraw-Hill, 1965. URL: https://cds.cern.ch/record/100771.

[3] Julian Seymour Schwinger. "Quantum Kinematics And Dynamics". In: W.A. Benjamin Inc. Publisher, 1970. URL: https://api.semanticscholar.org/CorpusID:117850892.

[4] Richard Phillips Feynman. "The Principle of Least Action in Quantum Mechanics." PhD thesis. Princeton University, New Jersey, May 1942.

[5] E. Noether. "Invariante Variationsprobleme". ger. In: *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* 1918 (1918), pp. 235–257. URL: http://eudml.org/doc/59024.

[6] Steven Weinberg. *The Quantum Theory of Fields*. Vol. 1. Cambridge University Press, 1995. DOI: 10.1017/CBO9781139644167.

[7] Kenneth G. Wilson. "Confinement of quarks". In: *Phys. Rev. D* 10 (8 Oct. 1974), pp. 2445–2459. DOI: 10.1103/PhysRevD.10.2445. URL: https://link.aps.org/doi/10.1103/PhysRevD.10.2445.

[8] Collectif. "Les connexions infinitésimales dans un espace fibré différentiable". fr. In: *Séminaire Bourbaki : années 1948/49 - 1949/50 - 1950/51, exposés 1-49*. Séminaire Bourbaki 1. talk:24. Société mathématique de France, 1952. URL: http://www.numdam.org/item/SB_1948-1951__1__153_0/.

[9] Paul A. M. Dirac. "The quantum theory of the electron". In: *Proc. Roy. Soc. Lond. A* 117 (1928), pp. 610–624. DOI: 10.1098/rspa.1928.0023.

[10] P. A. M. Dirac. "A Theory of Electrons and Protons". In: *Proc. Roy. Soc. Lond. A* 126.801 (1930), pp. 360–365. DOI: 10.1098/rspa.1930.0013.

[11] Al. Proca. "Particules libres photons et particules " charge pure "". In: *J. Phys. Radium* 8.1 (1937), pp. 23–28. DOI: 10.1051/jphysrad:019370080102300. URL: https://hal.science/jpa-00233468.

[12] Al. Proca. "Sur la théorie ondulatoire des électrons positifs et négatifs". In: *J. Phys. Radium* 7.8 (1936), pp. 347–353. DOI: 10.1051/jphysrad:0193600708034700. URL: https://hal.science/jpa-00233436.

[13] C. N. Yang and R. L. Mills. "Conservation of Isotopic Spin and Isotopic Gauge Invariance". In: *Phys. Rev.* 96 (1 Oct. 1954), pp. 191–195. DOI: 10.1103/PhysRev.96.191. URL: https://link.aps.org/doi/10.1103/PhysRev.96.191.

[14] Gerard 't Hooft and M. J. G. Veltman. "Regularization and Renormalization of Gauge Fields". In: *Nucl. Phys. B* 44 (1972), pp. 189–213. DOI: 10.1016/0550-3213(72)90279-9.

[15] Gerard 't Hooft. "Renormalization of Massless Yang-Mills Fields". In: *Nucl. Phys. B* 33 (1971), pp. 173–199. DOI: 10.1016/0550-3213(71)90395-6.

[16] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Reading, USA: Addison-Wesley, 1995. ISBN: 978-0-201-50397-5.

[17] Murray Gell-Mann. "A Schematic Model of Baryons and Mesons". In: *Phys. Lett.* 8 (1964), pp. 214–215. DOI: 10.1016/S0031-9163(64)92001-3.

[18] Yoichiro Nambu and G. Jona-Lasinio. "Dynamical model of elementary particles based on an analogy with superconductivity. II." In: *Phys. Rev.* 124 (1961). Ed. by T. Eguchi, pp. 246–254. DOI: 10.1103/PhysRev.124.246.

[19] Enrico Fermi. "Tentativo di una teoria dell'emissione dei raggi beta". In: *Ric. Sci.* 4 (1933), pp. 491–495.

[20] C. S. Wu et al. "Experimental Test of Parity Conservation in Beta Decay". In: *Phys. Rev.* 105 (4 Feb. 1957), pp. 1413–1415. DOI: 10.1103/PhysRev.105.1413. URL: https://link.aps.org/doi/10.1103/PhysRev.105.1413.

[21] J. H. Christenson et al. "Evidence for the 2π Decay of the K_2^0 Meson". In: *Phys. Rev. Lett.* 13 (4 July 1964), pp. 138–140. DOI: 10.1103/PhysRevLett.13.138. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.138.

[22] Sheldon Lee Glashow. "The renormalizability of vector meson interactions". In: *Nuclear Physics* 10 (1959), pp. 107–117. URL: https://api.semanticscholar.org/CorpusID:123163322.

[23] Abdus Salam. "Weak and Electromagnetic Interactions". In: *Conf. Proc. C* 680519 (1968), pp. 367–377. DOI: 10.1142/9789812795915_0034.

[24] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264. URL: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[25] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[26] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[27] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles". In: *Phys. Rev. Lett.* 13 (20 Nov. 1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.585.

[28] Sheldon L. Glashow. "Partial-symmetries of weak interactions". In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: https://doi.org/10.1016/0029-5582(61)90469-2. URL: https://www.sciencedirect.com/science/article/pii/0029558261904692.

[29] G. Altarelli. *The Standard Electroweak Theory and Beyond*. 2000. arXiv: hep-ph/0011078 [hep-ph].

[30] Sidney Coleman and Erick Weinberg. "Radiative Corrections as the Origin of Spontaneous Symmetry Breaking". In: *Phys. Rev. D* 7 (6 Mar. 1973), pp. 1888–1910. DOI: 10.1103/PhysRevD.7.1888. URL: https://link.aps.org/doi/10.1103/PhysRevD.7.1888.

[31] J. Goldstone. "Field Theories with Superconductor Solutions". In: *Nuovo Cim.* 19 (1961), pp. 154–164. DOI: 10.1007/BF02812722.

[32] R. L. Workman et al. "Review of Particle Physics". In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.

[33] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[34] Serguei Chatrchyan et al. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[35] Armen Tumasyan et al. "A portrait of the Higgs boson by the CMS experiment ten years after the discovery". In: *Nature* 607.7917 (2022), pp. 60–68. DOI: 10.1038/s41586-022-04892-x. arXiv: 2207.00043 [hep-ex].

[36] "A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery". In: *Nature* 607.7917 (2022). [Erratum: Nature 612, E24 (2022)], pp. 52–59. DOI: 10.1038/s41586-022-04893-w. arXiv: 2207.00092 [hep-ex].

[37] Benjamin W. Lee, C. Quigg, and H. B. Thacker. "Weak interactions at very high energies: The role of the Higgs-boson mass". In: *Phys. Rev. D* 16 (5 Sept. 1977), pp. 1519–1531. DOI: 10.1103/PhysRevD.16.1519. URL: https://link.aps.org/doi/10.1103/PhysRevD.16.1519.

[38] Benjamin W. Lee, C. Quigg, and H. B. Thacker. "Strength of Weak Interactions at Very High Energies and the Higgs Boson Mass". In: *Phys. Rev. Lett.* 38 (16 Apr. 1977), pp. 883–885. DOI: 10.1103/PhysRevLett.38.883. URL: https://link.aps.org/doi/10.1103/PhysRevLett.38.883.

[39] A. Denner and T. Hahn. "Radiative corrections to W+W- → W+W- in the electroweak standard model". In: *Nuclear Physics B* 525.1-2 (Aug. 1998), pp. 27–50. DOI: 10.1016/s0550-3213(98)00287-9. URL: https://doi.org/10.1016%2Fs0550-3213%2898%2900287-9.

[40] Kingman Cheung, Cheng-Wei Chiang, and Tzu-Chiang Yuan. "Partially strong WW scattering". In: *Phys. Rev. D* 78 (5 Sept. 2008), p. 051701. DOI: 10.1103/PhysRevD.78.051701. URL: https://link.aps.org/doi/10.1103/PhysRevD.78.051701.

[41] Thomas Appelquist and J. Carazzone. "Infrared singularities and massive fields". In: *Phys. Rev. D* 11 (10 May 1975), pp. 2856–2861. DOI: 10.1103/PhysRevD.11.2856. URL: https://link.aps.org/doi/10.1103/PhysRevD.11.2856.

[42] W. Buchmuller and D. Wyler. "Effective Lagrangian Analysis of New Interactions and Flavor Conservation". In: *Nucl. Phys. B* 268 (1986), pp. 621–653. DOI: 10.1016/0550-3213(86)90262-2.

[43] Ilaria Brivio and Michael Trott. "The standard model as an effective field theory". In: *Physics Reports* 793 (Feb. 2019), pp. 1–98. DOI: 10.1016/j.physrep.2018.11.002. URL: https://doi.org/10.1016%2Fj.physrep.2018.11.002.

[44] Steven Weinberg. "Baryon- and Lepton-Nonconserving Processes". In: *Phys. Rev. Lett.* 43 (21 Nov. 1979), pp. 1566–1570. DOI: 10.1103/PhysRevLett.43.1566. URL: https://link.aps.org/doi/10.1103/PhysRevLett.43.1566.

[45] Elizabeth E. Jenkins, Aneesh V. Manohar, and Peter Stoffer. "Low-energy effective field theory below the electroweak scale: operators and matching". In: *Journal of High Energy Physics* 2018.3 (Mar. 2018). DOI: 10.1007/jhep03(2018)016. URL: https://doi.org/10.1007%2Fjhep03%282018%29016.

[46] B. Grzadkowski et al. "Dimension-six terms in the Standard Model Lagrangian". In: *Journal of High Energy Physics* 2010.10 (Oct. 2010). DOI: 10.1007/jhep10(2010)085. URL: https://doi.org/10.1007%2Fjhep10%282010%29085.

[47] Gitte Elgaard-Clausen and Michael Trott. "On expansions in neutrino effective field theory". In: *Journal of High Energy Physics* 2017.11 (Nov. 2017). DOI: 10.1007/jhep11(2017)088. URL: https://doi.org/10.1007%2Fjhep11%282017%29088.

[48] L. F. Abbott and Mark B. Wise. "Effective Hamiltonian for nucleon decay". In: *Phys. Rev. D* 22 (9 Nov. 1980), pp. 2208–2212. DOI: 10.1103/PhysRevD.22.2208. URL: https://link.aps.org/doi/10.1103/PhysRevD.22.2208.

[49] Landon Lehman. "Extending the standard model effective field theory with the complete set of dimension-7 operators". In: *Physical Review D* 90.12 (Dec. 2014). DOI: 10.1103/physrevd.90.125023. URL: https://doi.org/10.1103%2Fphysrevd.90.125023.

[50] Landon Lehman and Adam Martin. "Low-derivative operators of the Standard Model effective field theory via Hilbert series methods". In: *Journal of High Energy Physics* 2016.2 (Feb. 2016). DOI: 10.1007/jhep02(2016)081. URL: https://doi.org/10.1007%2Fjhep02%282016%29081.

[51] Brian Henning et al. *2, 84, 30, 993, 560, 15456, 11962, 261485, ...: Higher dimension operators in the SM EFT*. 2019. arXiv: 1512.03433 [hep-ph].

[52] K. Hagiwara et al. "Low energy effects of new interactions in the electroweak boson sector". In: *Phys. Rev. D* 48 (5 Sept. 1993), pp. 2182–2203. DOI: 10.1103/PhysRevD.48.2182. URL: https://link.aps.org/doi/10.1103/PhysRevD.48.2182.

[53] Gian Francesco Giudice et al. "The strongly-interacting light Higgs". In: *Journal of High Energy Physics* 2007.06 (June 2007), pp. 045–045. DOI: 10.1088/1126-6708/2007/06/045. URL: https://doi.org/10.1088%2F1126-6708%2F2007%2F06%2F045.

[54] Joan Elias-Miró et al. "Scaling and tuning of EW and Higgs observables". In: *Journal of High Energy Physics* 2014.5 (May 2014). DOI: 10.1007/jhep05(2014)019. URL: https://doi.org/10.1007%2Fjhep05%282014%29019.

[55] Nicola Cabibbo. "Unitary Symmetry and Leptonic Decays". In: *Phys. Rev. Lett.* 10 (12 June 1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531. URL: https://link.aps.org/doi/10.1103/PhysRevLett.10.531.

[56] Makoto Kobayashi and Toshihide Maskawa. "CP Violation in the Renormalizable Theory of Weak Interaction". In: *Prog. Theor. Phys.* 49 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652.

[57] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. "Remarks on the unified model of elementary particles". In: *Prog. Theor. Phys.* 28 (1962), pp. 870–880. DOI: 10.1143/PTP.28.870.

[58] B. Pontecorvo. "Inverse beta processes and nonconservation of lepton charge". In: *Zh. Eksp. Teor. Fiz.* 34 (1957), p. 247.

[59] Ilaria Brivio and Michael Trott. "Scheming in the SMEFT. . . and a reparameterization invariance!" In: *Journal of High Energy Physics* 2017.7 (July 2017). DOI: 10.1007/jhep07(2017)148. URL: https://doi.org/10.1007%2Fjhep07%282017%29148.

[60] Ansgar Denner and Stefan Dittmaier. "Electroweak radiative corrections for collider physics". In: *Physics Reports* 864 (June 2020), pp. 1–163. DOI: 10.1016/j.physrep.2020.04.001. URL: https://doi.org/10.1016%2Fj.physrep.2020.04.001.

[61] Ilaria Brivio et al. "Electroweak input parameters". In: (Nov. 2021). arXiv: 2111.12515 [hep-ph].

[62] Rodrigo Alonso et al. "Renormalization Group Evolution of the Standard Model Dimension Six Operators III: Gauge Coupling Dependence and Phenomenology". In: *JHEP* 04 (2014), p. 159. DOI: 10.1007/JHEP04(2014)159. arXiv: 1312.2014 [hep-ph].

[63] Darius A. Faroughy et al. "Flavour symmetries in the SMEFT". In: *JHEP* 08 (2020), p. 166. DOI: 10.1007/JHEP08(2020)166. arXiv: 2005.05366 [hep-ph].

[64] J. M. Gerard. "FERMION MASS SPECTRUM IN SU(2)-L x U(1)". In: *Z. Phys. C* 18 (1983), p. 145. DOI: 10.1007/BF01572477.

[65] Alexander L. Kagan et al. "General Minimal Flavor Violation". In: *Phys. Rev. D* 80 (2009), p. 076002. DOI: 10.1103/PhysRevD.80.076002. arXiv: 0903.1794 [hep-ph].

[66] Ilaria Brivio. "SMEFTsim 3.0 — a practical guide". In: *Journal of High Energy Physics* 2021.4 (Apr. 2021). DOI: 10.1007/jhep04(2021)073. URL: https://doi.org/10.1007%2Fjhep04%282021%29073.

[67] G. D'Ambrosio et al. "Minimal flavour violation: an effective field theory approach". In: *Nuclear Physics B* 645.1-2 (Nov. 2002), pp. 155–187. DOI: 10.1016/s0550-3213(02)00836-2. URL: https://doi.org/10.1016%2Fs0550-3213%2802%2900836-2.

[68] Gianluca Blankenburg, Gino Isidori, and Joel Jones-Pérez. "Neutrino masses and LFV from minimal breaking of U(3)5 and U(2)5 flavor symmetries". In: *The European Physical Journal C* 72.8 (Aug. 2012). DOI: 10.1140/epjc/s10052-012-2126-7. URL: https://doi.org/10.1140%2Fepjc%2Fs10052-012-2126-7.

[69] Riccardo Barbieri et al. "Flavour physics from an approximate U(2)3 symmetry". In: *Journal of High Energy Physics* 2012.7 (July 2012). DOI: 10.1007/jhep07(2012)181. URL: https://doi.org/10.1007%2Fjhep07%282012%29181.

[70] D. Barducci et al. "Interpreting top-quark LHC measurements in the standard-model effective field theory". In: (Feb. 2018). Ed. by Juan Antonio Aguilar-Saavedra et al. arXiv: 1802.07237 [hep-ph].

[71] "LHC Machine". In: *JINST* 3 (2008). Ed. by Lyndon Evans and Philip Bryant, S08001. DOI: 10.1088/1748-0221/3/08/S08001.

[72] S. Chatrchyan et al. "The CMS Experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[73] W. Adam et al. "The CMS Phase-1 Pixel Detector Upgrade". In: *JINST* 16.02 (2021), P02027. DOI: 10.1088/1748-0221/16/02/P02027. arXiv: 2012.14304 [physics.ins-det].

[74] "CMS Technical Design Report for the Pixel Detector Upgrade". In: (Sept. 2012). Ed. by David Aaron Matzner Dominguez et al. DOI: 10.2172/1151650.

[75] "The CMS electromagnetic calorimeter project: Technical Design Report". In: (1997).

[76] P. Adzic et al. "Energy resolution of the barrel of the CMS electromagnetic calorimeter". In: *JINST* 2 (2007), P04004. DOI: 10.1088/1748-0221/2/04/P04004.

[77] Serguei Chatrchyan et al. "Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s} = 7$ TeV". In: *JINST* 8 (2013), P09009. DOI: 10.1088/1748-0221/8/09/P09009. arXiv: 1306.2016 [hep-ex].

[78] S. Abdullin et al. "The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c". In: *Eur. Phys. J. C* 60 (2009). [Erratum: Eur.Phys.J.C 61, 353–356 (2009)], pp. 359–373. DOI: 10.1140/epjc/s10052-009-0959-5.

[79] "The CMS muon project: Technical Design Report". In: (1997).

[80] G. L. Bayatian et al. "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software". In: (2006).

[81] Serguei Chatrchyan et al. "The Performance of the CMS Muon Detector in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV at the LHC". In: *JINST* 8 (2013), P11002. DOI: 10.1088/1748-0221/8/11/P11002. arXiv: 1306.6905 [physics.ins-det].

[82] Georges Charpak et al. "High Accuracy Localization of Minimum Ionizing Particles Using the Cathode Induced Charge Center of Gravity Readout". In: *Nucl. Instrum. Meth.* 167 (1979). Ed. by Georges Charpak, p. 455. DOI: 10.1016/0029-554X(79)90227-1.

[83] "CMS Technical Design Report for the Level-1 Trigger Upgrade". In: (June 2013). Ed. by A. Tapper and Darin Acosta.

[84] Alexandre Zabi et al. "The Phase-2 Upgrade of the CMS Level-1 Trigger". In: (2020).

[85] A. Zabi et al. "The CMS Level-1 Calorimeter Trigger for the LHC Run II". In: *JINST* 12.01 (2017), p. C01065. DOI: 10.1088/1748-0221/12/01/C01065.

[86] A. Svetek et al. "The Calorimeter Trigger Processor Card: the next generation of high speed algorithmic data processing at CMS". In: *Journal of Instrumentation* 11.02 (Feb. 2016), p. C02011. DOI: 10.1088/1748-0221/11/02/C02011. URL: https://dx.doi.org/10.1088/1748-0221/11/02/C02011.

[87] D. Acosta et al. "The CMS Modular Track Finder boards, MTF6 and MTF7". In: *JINST* 8 (2013), p. C12034. DOI: 10.1088/1748-0221/8/12/C12034.

[88] Vardan Khachatryan et al. "The CMS trigger system". In: *JINST* 12.01 (2017), P01020. DOI: 10.1088/1748-0221/12/01/P01020. arXiv: 1609.02366 [physics.ins-det].

[89] "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET". In: (Apr. 2009).

[90] G. Aad. "Differential cross-section measurements for the electroweak production of dijets in association with a Z boson in proton–proton collisions at ATLAS". In: *The European Physical Journal C* 81.2 (Feb. 2021). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-020-08734-w. URL: http://dx.doi.org/10.1140/epjc/s10052-020-08734-w.

[91] Sirunyan. "Electroweak production of two jets in association with a Z boson in proton–proton collisions at $\sqrt{s} = 13$TeV". In: *The European Physical Journal C* 78.7 (July 2018). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-018-6049-9. URL: http://dx.doi.org/10.1140/epjc/s10052-018-6049-9.

[92] Rikkert Frederix and Stefano Frixione. "Merging meets matching in MC@NLO". In: *Journal of High Energy Physics* 2012.12 (Dec. 2012). ISSN: 1029-8479. DOI: 10.1007/jhep12(2012)061. URL: http://dx.doi.org/10.1007/JHEP12(2012)061.

[93] Ryan Gavin et al. "FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order". In: *Computer Physics Communications* 182.11 (Nov. 2011), pp. 2388–2403. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2011.06.008. URL: http://dx.doi.org/10.1016/j.cpc.2011.06.008.

[94] CMS. *Rochester Corrections*. https://twiki.cern.ch/twiki/bin/viewauth/CMS/RochcorMuon.

[95] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. "The anti-ktjet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/04/063. URL: http://dx.doi.org/10.1088/1126-6708/2008/04/063.

[96] A. Hayrapetyan, A. Tumasyan, and Adam. "The CMS Statistical Analysis and Combination Tool: Combine". In: *Computing and Software for Big Science* 8.1 (Nov. 2024). ISSN: 2510-2044. DOI: 10.1007/s41781-024-00121-4. URL: http://dx.doi.org/10.1007/s41781-024-00121-4.

[97] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J. C* 71 (2011). [Erratum: Eur.Phys.J.C 73, 2501 (2013)], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].

[98] Roger J. Barlow and Christine Beeston. "Fitting using finite Monte Carlo samples". In: *Comput. Phys. Commun.* 77 (1993), pp. 219–228. DOI: 10.1016/0010-4655(93)90005-W.

[99] Scott M Lundberg and Su-In Lee. "A Unified Approach to Interpreting Model Predictions". In: *Advances in Neural Information Processing Systems 30*. Ed. by I. Guyon et al. Curran Associates, Inc., 2017, pp. 4765–4774. URL: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

[100] Andrea Massironi. "Unfolding, fiducial cross section, and more ". In: *CMS Note* 2024/240 (2024).

[101] *Jet energy scale uncertainty sources*. https://twiki.cern.ch/twiki/bin/viewauth/CMS/JECUncertaintySources.

[102] Olivier Mattelaer. "On the maximal use of Monte Carlo samples: re-weighting events at NLO accuracy". In: *Eur. Phys. J. C* 76.12 (2016), p. 674. DOI: 10.1140/epjc/s10052-016-4533-7. arXiv: 1607.00763 [hep-ph].

[103] Riccardo Bellan et al. "A sensitivity study of VBS and diboson WW to dimension-6 EFT operators at the LHC". In: *JHEP* 05 (2022), p. 039. DOI: 10.1007/JHEP05(2022)039. arXiv: 2108.03199 [hep-ph].

[104] I. Zurbano Fernandez et al. "High-Luminosity Large Hadron Collider (HL-LHC): Technical design report". In: 10/2020 (Dec. 2020). Ed. by I. Béjar Alonso et al. DOI: 10.23731/CYRM-2020-0010.

[105] Andrea Dainese et al., eds. *Report on the Physics at the HL-LHC,and Perspectives for the HE-LHC*. Vol. 7/2019. CERN Yellow Reports: Monographs. Geneva, Switzerland: CERN, 2019. ISBN: 978-92-9083-549-3. DOI: 10.23731/CYRM-2019-007.

[106] M. Cepeda et al. "Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC". In: *CERN Yellow Rep. Monogr.* 7 (2019). Ed. by Andrea Dainese et al., pp. 221–584. DOI: 10.23731/CYRM-2019-007.221. arXiv: 1902.00134 [hep-ph].

[107] P. Azzi et al. "Report from Working Group 1: Standard Model Physics at the HL-LHC and HE-LHC". In: *CERN Yellow Rep. Monogr.* 7 (2019). Ed. by Andrea Dainese et al., pp. 1–220. DOI: 10.23731/CYRM-2019-007.1. arXiv: 1902.04070 [hep-ph].

[108] Xabier Cid Vidal et al. "Report from Working Group 3: Beyond the Standard Model physics at the HL-LHC and HE-LHC". In: *CERN Yellow Rep. Monogr.* 7 (2019). Ed. by Andrea Dainese et al., pp. 585–865. DOI: 10.23731/CYRM-2019-007.585. arXiv: 1812.07831 [hep-ph].

[109] Johannes Albrecht et al. "A Roadmap for HEP Software and Computing R&D for the 2020s". In: *Comput. Softw. Big Sci.* 3.1 (2019), p. 7. DOI: 10.1007/s41781-018-0018-8. arXiv: 1712.06982 [physics.comp-ph].

[110] "Technical Proposal for the Phase-II Upgrade of the CMS Detector". In: (June 2015). Ed. by D. Contardo et al. DOI: 10.17181/CERN.VU8I.D59J.

[111] A. M. Sirunyan et al. "Particle-flow reconstruction and global event description with the CMS detector". In: *JINST* 12.10 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].

[112] "Pileup Removal Algorithms". In: (2014).

[113] Daniele Bertolini et al. "Pileup Per Particle Identification". In: *JHEP* 10 (2014), p. 059. DOI: 10.1007/JHEP10(2014)059. arXiv: 1407.6013 [hep-ph].

[114] Joel N. Butler and Tommaso Tabarelli de Fatis. "A MIP Timing Detector for the CMS Phase-2 Upgrade". In: (2019).

[115] Armen Tumasyan et al. "The Phase-2 Upgrade of the CMS Tracker". In: (June 2017). Ed. by K. Klein. DOI: 10.17181/CERN.QZ28.FLHW.

[116] S. Bonacini et al. "Characterization of a commercial 65 nm CMOS technology for SLHC applications". In: *JINST* 7 (2012), P01015. DOI: 10.1088/1748-0221/7/01/P01015.

[117] Thomas Hebbeker and Andrey Korytov. "The Phase-2 Upgrade of the CMS Muon Detectors". In: (Sept. 2017).

[118] *The Phase-2 Upgrade of the CMS Barrel Calorimeters*. Tech. rep. This is the final version, approved by the LHCC. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2283187.

[119] "The Phase-2 Upgrade of the CMS Endcap Calorimeter". In: (2017).

[120] CMS Collaboration. "An event seen in the CMS detector during the high-pileup fill of 8 October 2025". CMS Collection. 2025. URL: https://cds.cern.ch/record/2945596.

[121] F.M. Addesa et al. "Comparative characterization study of LYSO:Ce crystals for timing applications". In: *JINST* 17.08 (2022). 38 pages, 22 figures, P08028. DOI: 10.1088/1748-0221/17/08/P08028. arXiv: 2205.14890. URL: http://cds.cern.ch/record/2812139.

[122] Francesca Maria Addesa et al. "OPTICAL SPECTROSCOPIC CHARACTERIZATION OF LYSO CRYSTALS AT THE CALLIOPE FACILITY (ENEA CASACCIA R.C.)" In: 2020. URL: https://api.semanticscholar.org/CorpusID:233322027.

[123] R. Abbott et al. "Test beam characterization of sensor prototypes for the CMS Barrel MIP Timing Detector". In: *JINST* 16.07 (2021), P07023. DOI: 10.1088/1748-0221/16/07/P07023. arXiv: 2104.07786 [physics.ins-det].

[124] A. Heering et al. "Effects of very high radiation on SiPMs". In: *Nucl. Instrum. Meth. A* 824 (2016). Ed. by Maria Giuseppina Bisogni et al., pp. 111–114. DOI: 10.1016/j.nima.2015.11.037.

[125] E. Garutti and Yu. Musienko. "Radiation damage of SiPMs". In: *Nucl. Instrum. Meth. A* 926 (2019), pp. 69–84. DOI: 10.1016/j.nima.2018.10.191. arXiv: 1809.06361 [physics.ins-det].

[126] A. Bornheim et al. "Integration of thermo-electric coolers into the CMS MTD SiPM arrays for operation under high neutron fluence". In: *JINST* 18.08 (2023), P08020. DOI: 10.1088/1748-0221/18/08/P08020. arXiv: 2306.00818 [physics.ins-det].

[127] E. Albuquerque et al. "TOFHIR2: The readout ASIC of the CMS Barrel MIP Timing Detector". In: *2020 IEEE Nuclear Science Symposium (NSS) and Medical Imaging Conference (MIC)*. Oct. 2020. DOI: 10.1109/NSS/MIC42677.2020.9507749.

[128] Gilbert Badaro et al. "The Phase-2 Upgrade of the CMS Data Acquisition". In: *EPJ Web Conf.* 251 (2021), p. 04023. DOI: 10.1051/epjconf/202125104023.

[129] Polina Simkina. *The DAQ and clock distribution system of CMS MIP Timing Detector*. Tech. rep. Geneva: CERN, 2023. DOI: 10.1016/j.nima.2022.167802. URL: https://cds.cern.ch/record/2824604.

[130] M. Firlej et al. "A fast, ultra-low and frequency-scalable power consumption, 10-bit SAR ADC for particle physics detectors". In: *JINST* 10.11 (2015), P11012. DOI: 10.1088/1748-0221/10/11/P11012.

[131] S Bonacini, K Kloukinas, and P Moreira. "E-link: A Radiation-Hard Low-Power Electrical Link for Chip-to-Chip Communication". In: (2009). DOI: 10.5170/CERN-2009-006.422. URL: https://cds.cern.ch/record/1235849.

[132] Paolo Carniti, Claudio Gotti, and Gianluigi Pessina. "ALDO2, a multi-function rad-hard linear regulator for SiPM-based HEP detectors". In: *Nucl. Instrum. Meth. A* 1039 (2022), p. 167028. DOI: 10.1016/j.nima.2022.167028. arXiv: 2203.16098 [physics.ins-det].

[133] Paulo Moreira et al. *lpGBT documentation: release*. 2022. URL: https://cds.cern.ch/record/2809058.

[134] A Caratelli et al. "The GBT-SCA, a radiation tolerant ASIC for detector control and monitoring applications in HEP experiments". In: *JINST* 10.03 (2015), p. C03034. DOI: 10.1088/1748-0221/10/03/C03034. URL: https://cds.cern.ch/record/2158969.

[135] T. Gadek et al. "DC-DC converters for the CMS MTD BTL and ECAL for HL-LHC". In: *JINST* 18.02 (2023), p. C02038. DOI: 10.1088/1748-0221/18/02/C02038.

[136] Jean-Marc André et al. "The CMS Data Acquisition System for the Phase-2 Upgrade". In: (June 2018). arXiv: 1806.08975 [physics.ins-det].

[137] Andrew Rose. *Serenity - An ATCA prototyping platform for CMS Phase-2*. Tech. rep. Geneva: CERN, 2019. DOI: 10.22323/1.343.0115. URL: https://cds.cern.ch/record/2646388.

[138] Thomas Stephen Williams. *IPbus A flexible Ethernet-based control system for xTCA hardware*. Tech. rep. Geneva: CERN, 2014. URL: https://cds.cern.ch/record/2020872.

[139] H Alva-Sánchez et al. "Understanding the intrinsic radioactivity energy spectrum from 176Lu in LYSO/LSO scintillation crystals". en. In: *Sci. Rep.* 8.1 (Nov. 2018).

[140] F. Addesa et al. "The CMS barrel timing layer: test beam confirmation of module timing performance". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1081 (Jan. 2026), p. 170823. ISSN: 0168-9002. DOI: 10.1016/j.nima.2025.170823. URL: http://dx.doi.org/10.1016/j.nima.2025.170823.

Appendix A

VBF-Z SM Analysis

A.0.1 Data/MC Comparison for DNN Input Variables

In Section 2.9 we described the training procedure of the deep neural network (DNN) and presented the ranking of the most important input variables contributing to the classification. Since the DNN relies on a large set of kinematic and event-level features, it is essential to validate that these inputs are accurately modeled in simulation. Any significant mismodeling could bias the classifier training or propagate into the extraction of the signal, hence a systematic validation step is required.

To this end, we perform a detailed Data/MC comparison of the top 18 variables identified by the importance ranking discussed earlier. The comparisons are presented in Figures A.1-A.0.1, and cover several regions of phase space:

In addition to the control regions, we present the same comparisons in a more inclusive phase space defined by requiring DNN < 0.9. This selection retains a broad mixture of DY PU, Top CR, and SR events, and is less affected by potential biases from the highly discriminant signal region (DNN > 0.9). The inclusive sample thus provides an additional global validation of the modeling of input variables before applying the final selection aimed at enhancing the electroweak Zjj signal.

Overall, these comparisons confirm a satisfactory level of agreement between data and simulation across all considered regions and observables. The residual discrepancies are within the assigned systematic uncertainties and do not exhibit trends that could bias the DNN training or its application in the analysis. This validation step supports the reliability of the DNN inputs and the robustness of the downstream event classification strategy.

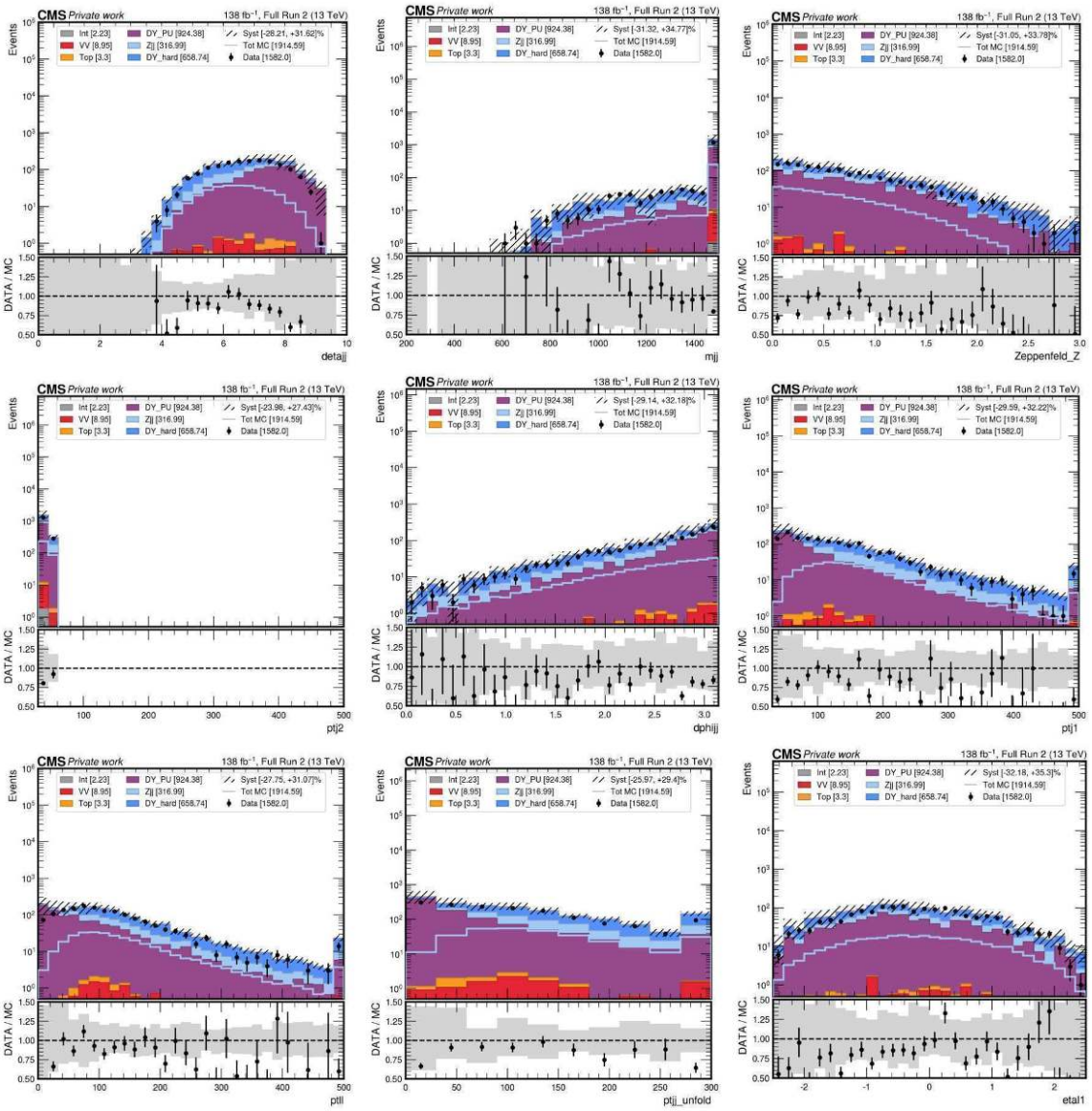

Figure A.1: Input variables to the DNN in the DY PU CR with $0.9 < \text{DNN} < 1.0$. From left to right, from top to bottom, ranked by importance in the DNN: $\Delta\eta_{jj}$, m_{jj}, Z_{ll}, p^T_{j2}, $\Delta\phi_{jj}$, p^T_{j1}, p^T_{ll}, p^T_{jj}, η_{l1}

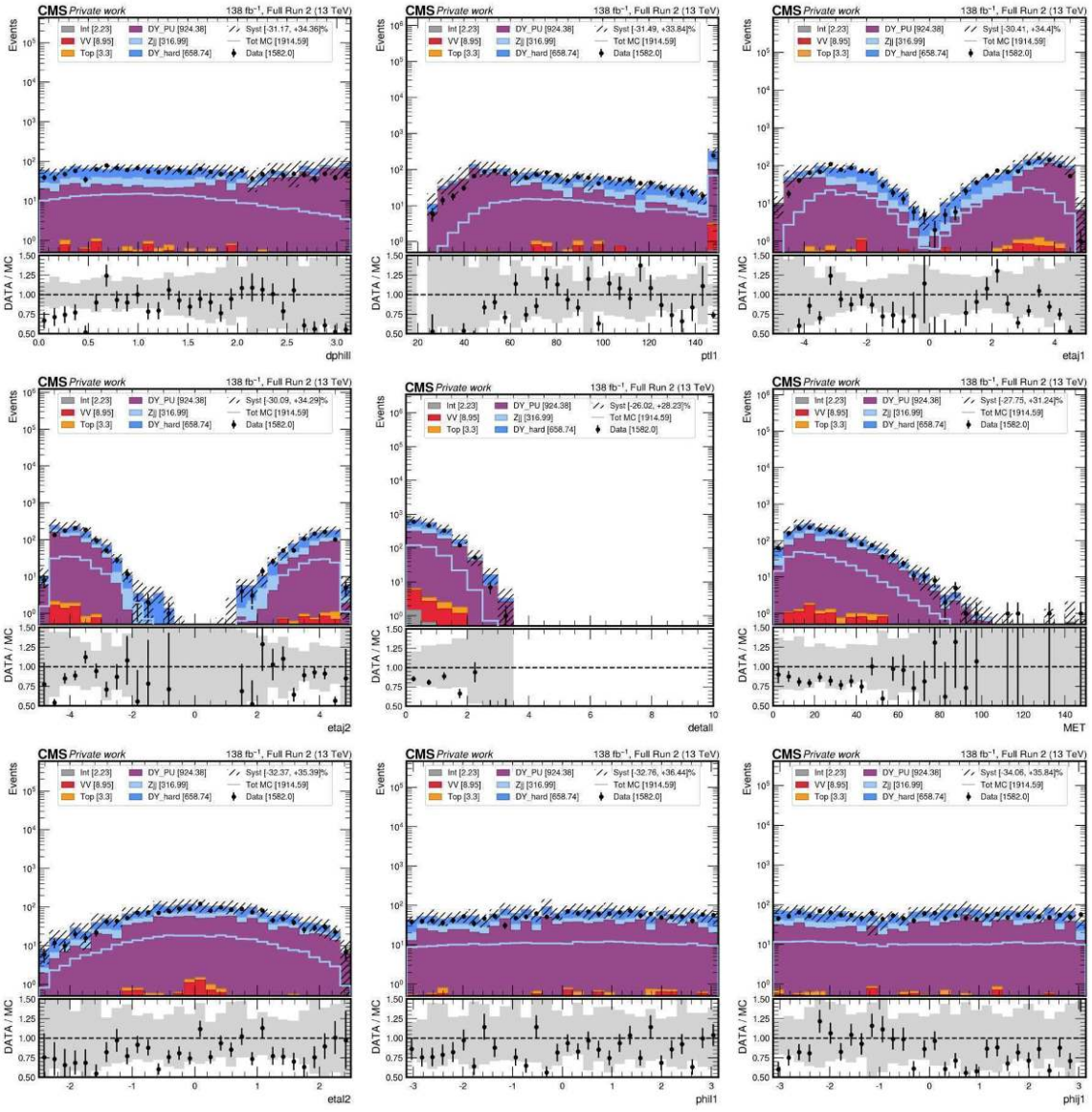

Figure A.2: Input variables to the DNN in the DY PU CR with $0.9 < \text{DNN} < 1.0$. From left to right, from top to bottom, ranked by importance in the DNN: $\Delta\phi_{ll}$, p_{l1}^{T}, η_{j1}, η_{j2}, $\Delta\eta_{ll}$, MET, η_{l2}, ϕ_{l1}, ϕ_{j1}

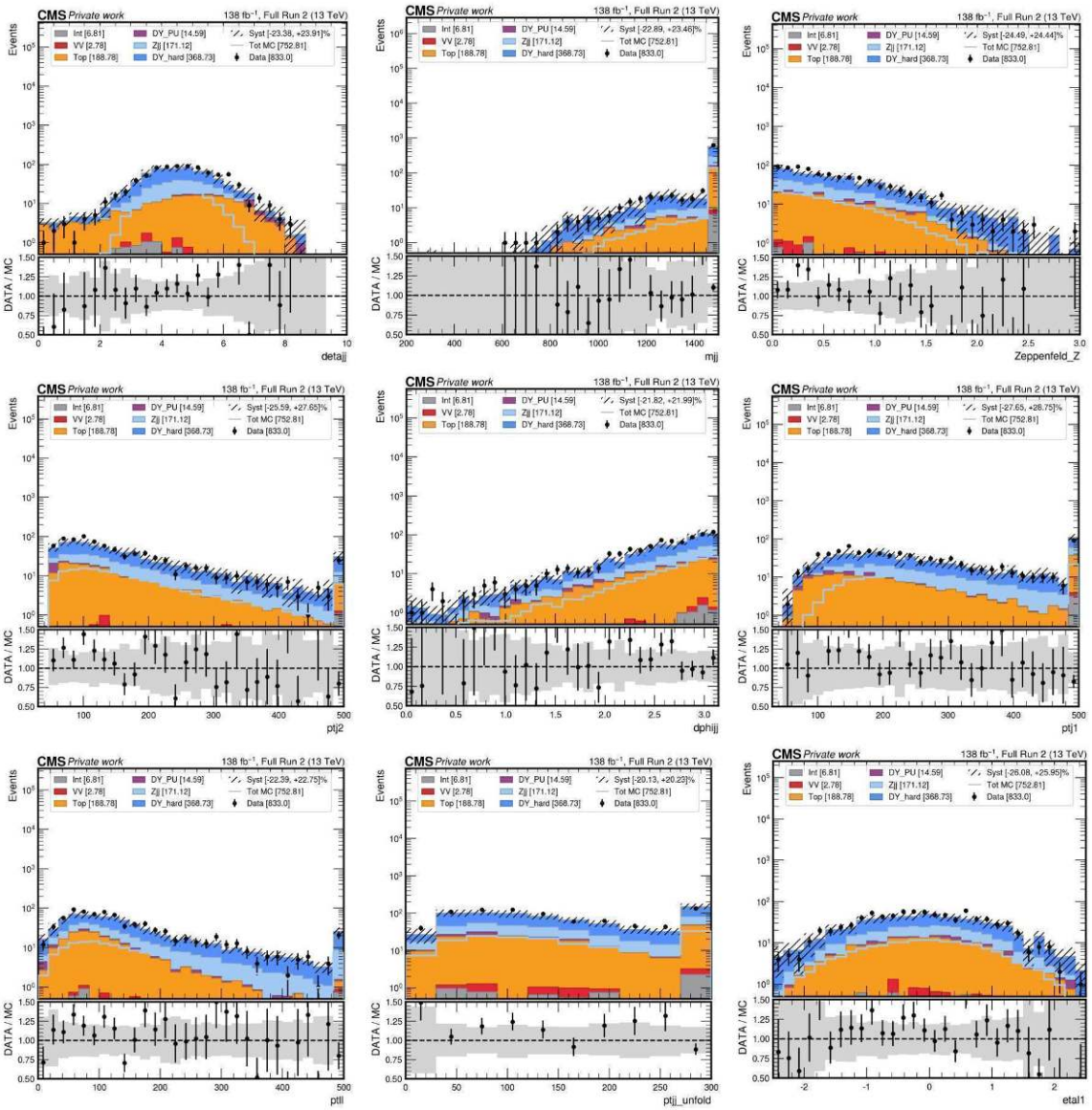

Figure A.3: Input variables to the DNN in the Top CR with $0.9 < \text{DNN} < 1.0$. From left to right, from top to bottom, ranked by importance in the DNN: $\Delta\eta_{jj}$, m_{jj}, Z_{ll}, p^T_{j2}, $\Delta\phi_{jj}$, p^T_{j1}, p^T_{ll}, p^T_{jj}, η_{l1}

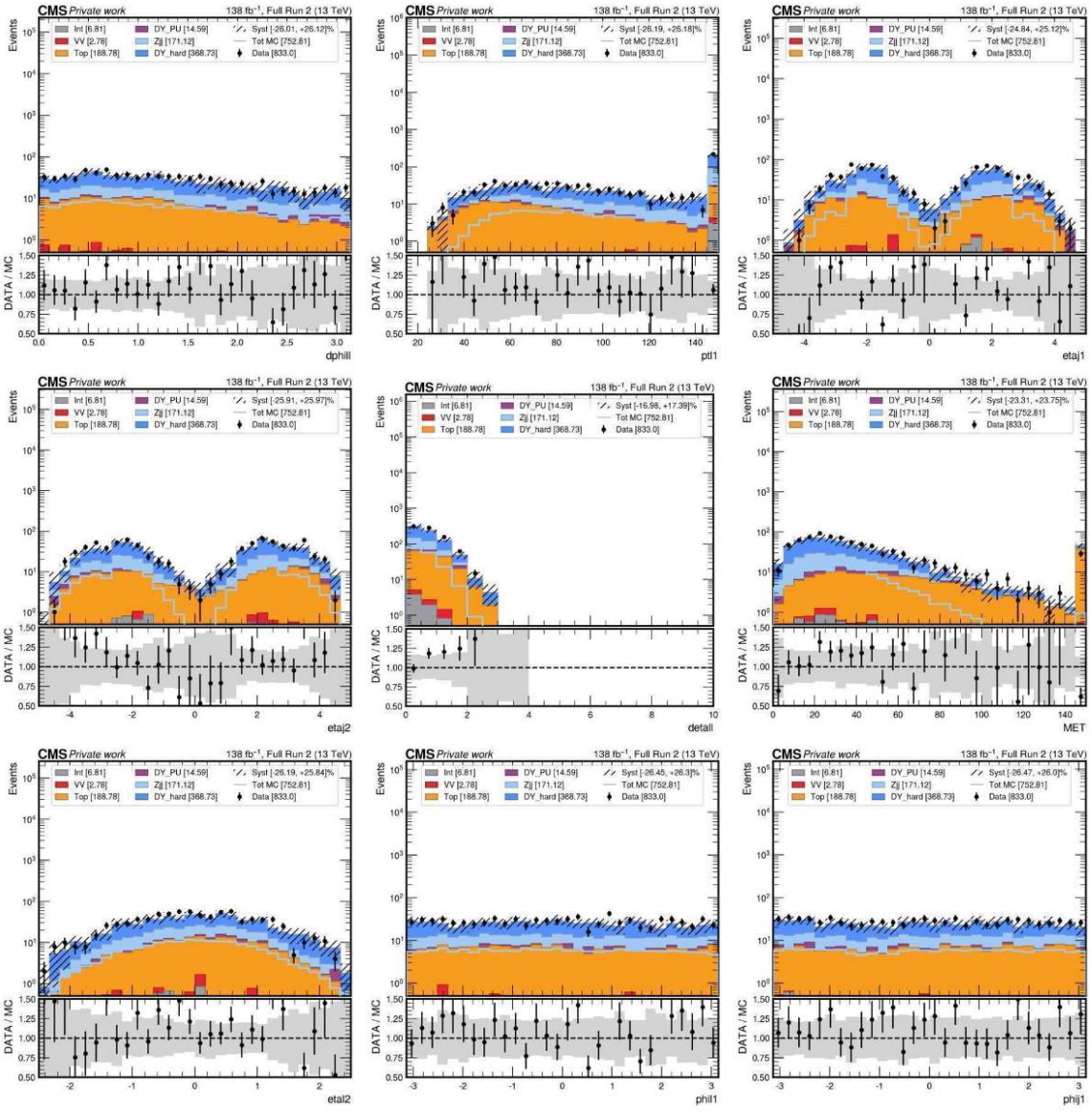

Figure A.4: Input variables to the DNN in the Top CR with $0.9 < \text{DNN} < 1.0$. From left to right, from top to bottom, ranked by importance in the DNN: $\Delta\phi_{ll}$, p_{l1}^{T}, η_{j1}, η_{j2}, $\Delta\eta_{ll}$, MET, η_{l2}, ϕ_{l1}, ϕ_{j1}

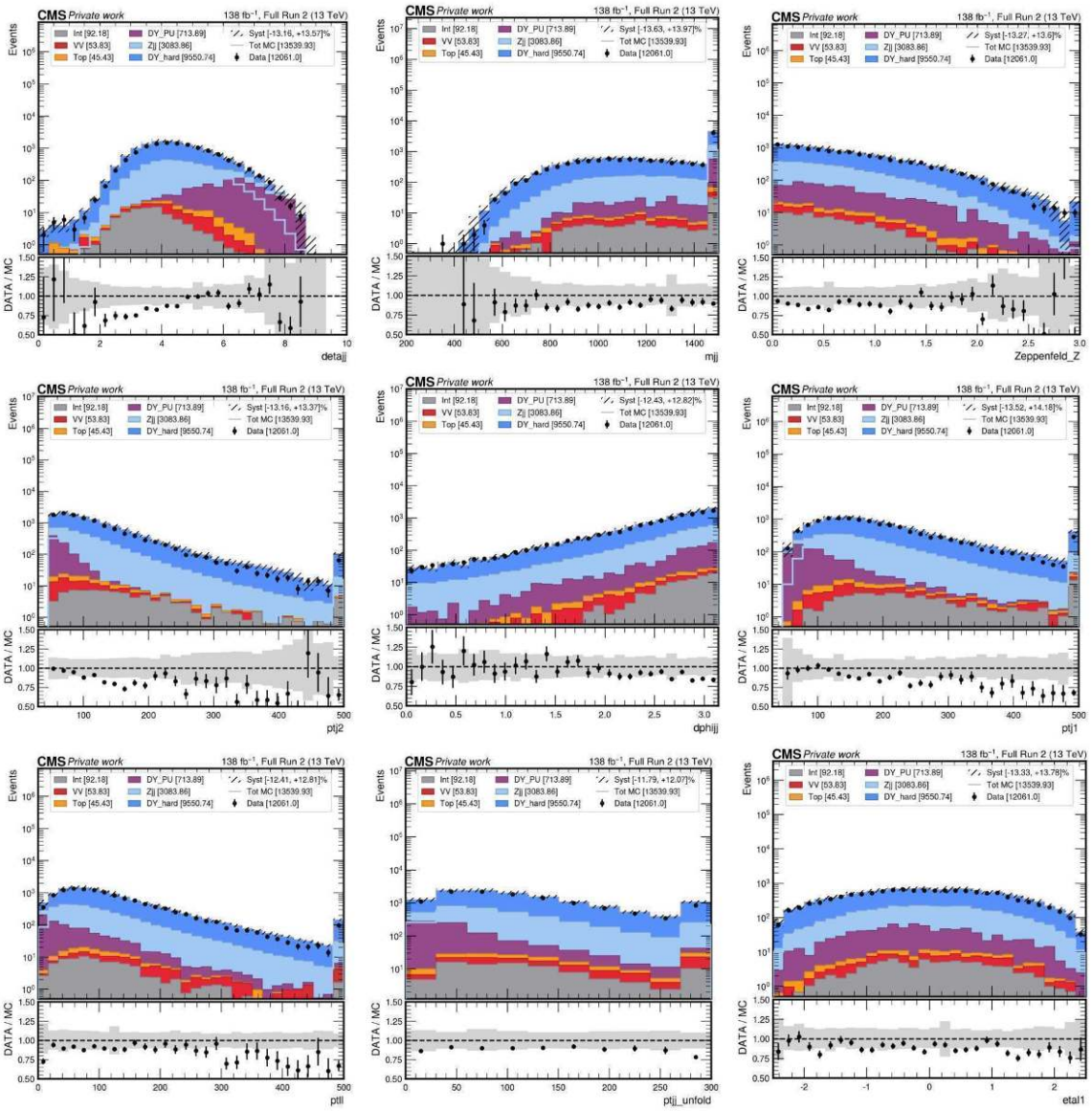

Figure A.5: Input variables to the DNN in the SR with $0.8 < \text{DNN} < 0.9$. From left to right, from top to bottom, ranked by importance in the DNN: $\Delta\eta_{jj}$, m_{jj}, Z_{ll}, p_{j2}^{T}, $\Delta\phi_{jj}$, p_{j1}^{T}, p_{ll}^{T}, p_{jj}^{T}, η_{l1}

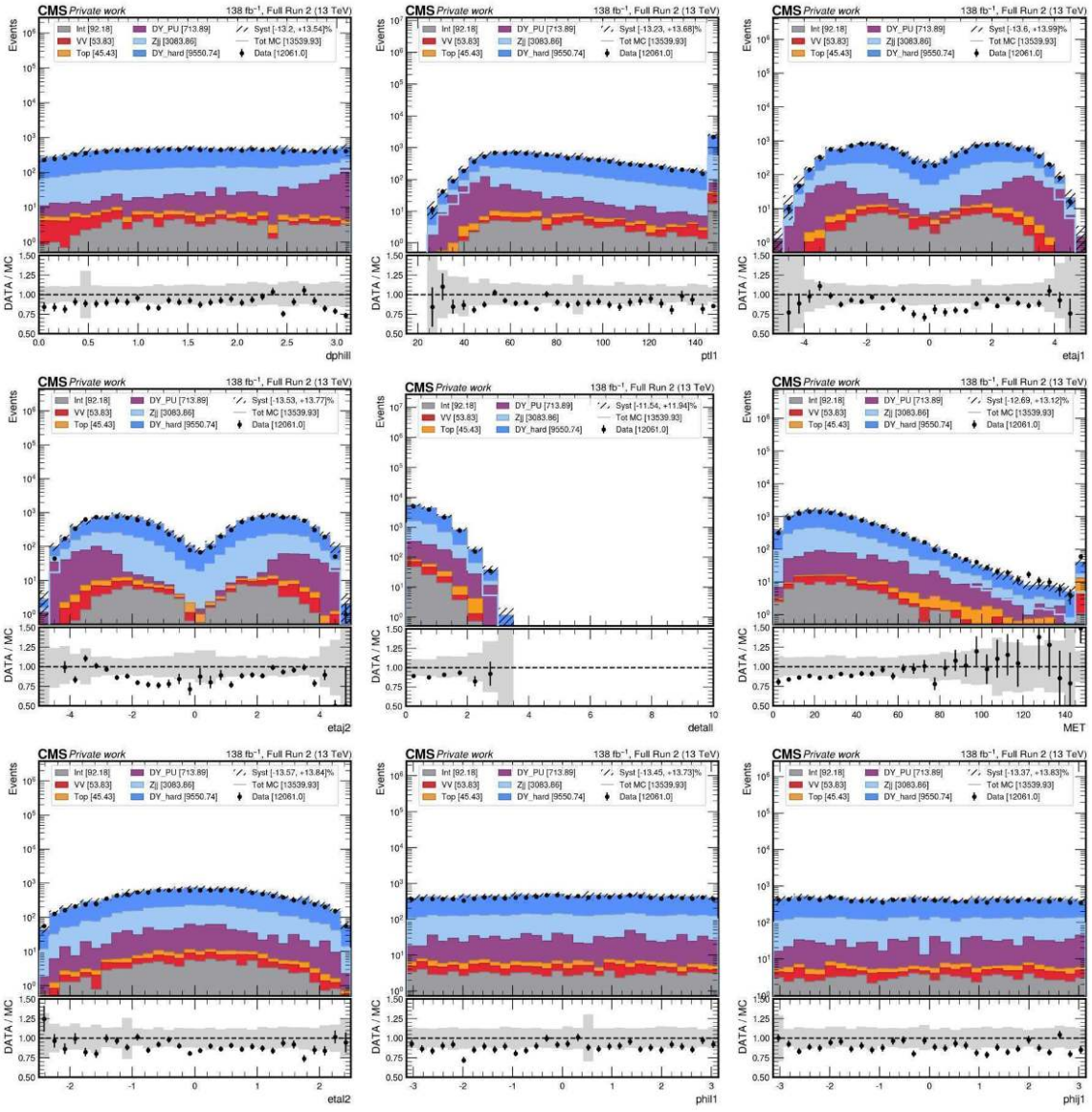

Figure A.6: Input variables to the DNN in the SR with $0.8 < \text{DNN} < 0.9$. From left to right, from top to bottom, ranked by importance in the DNN: $\Delta\phi_{ll}$, p_{l1}^{T}, η_{j1}, η_{j2}, $\Delta\eta_{ll}$, MET, η_{l2}, ϕ_{l1}, ϕ_{j1}

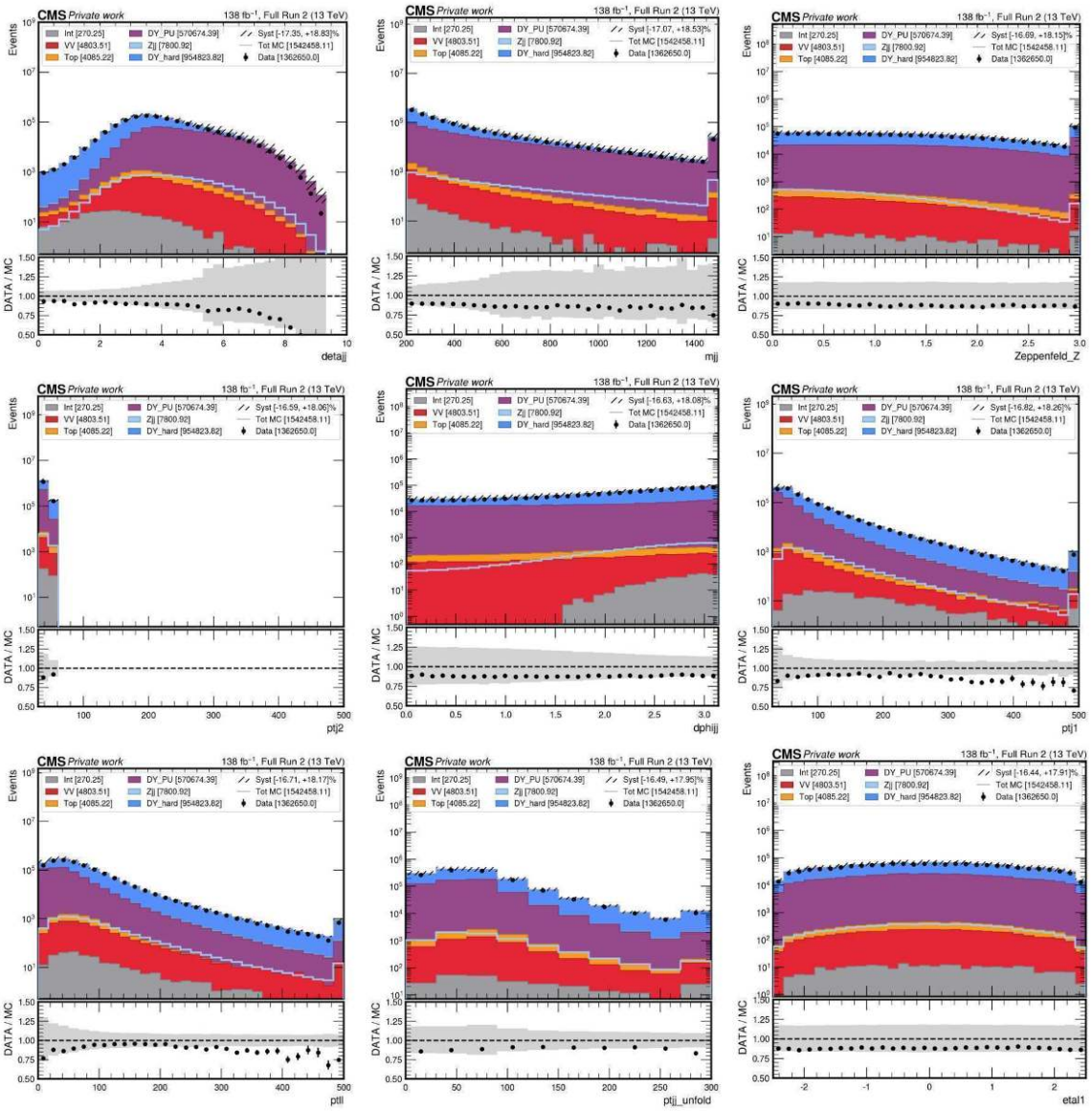

Figure A.7: Input variables to the DNN in the DY PU CR with DNN < 0.9. From left to right, from top to bottom, ranked by importance in the DNN: $\Delta\eta_{jj}$, m_{jj}, Z_{ll}, p^T_{j2}, $\Delta\phi_{jj}$, p^T_{j1}, p^T_{ll}, p^T_{jj}, η_{l1}

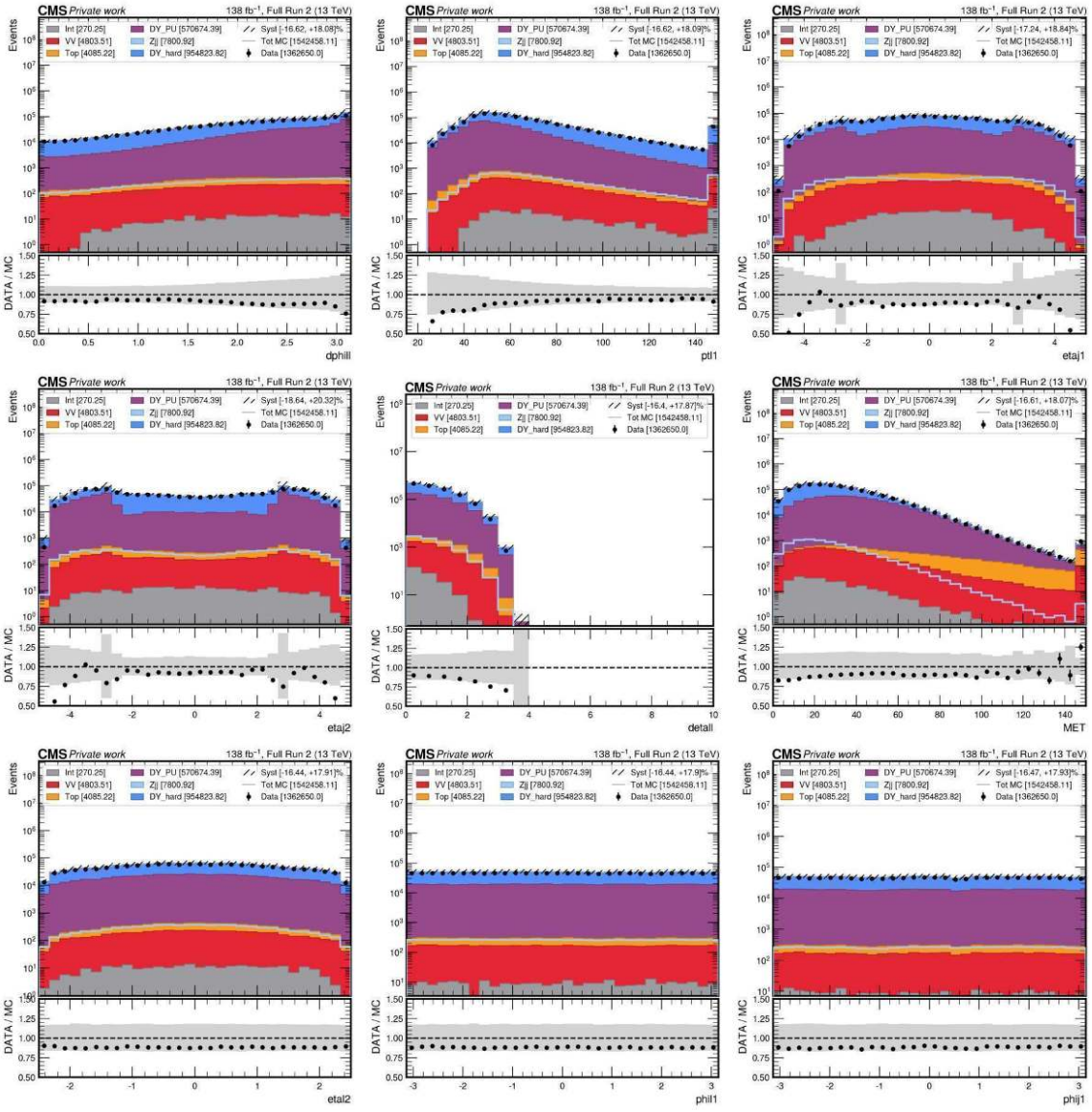

Figure A.8: Input variables to the DNN in the DY PU CR with DNN < 0.9. From left to right, from top to bottom, ranked by importance in the DNN: $\Delta\phi_{ll}$, p_{l1}^{T}, η_{j1}, η_{j2}, $\Delta\eta_{ll}$, MET, η_{l2}, ϕ_{l1}, ϕ_{j1}

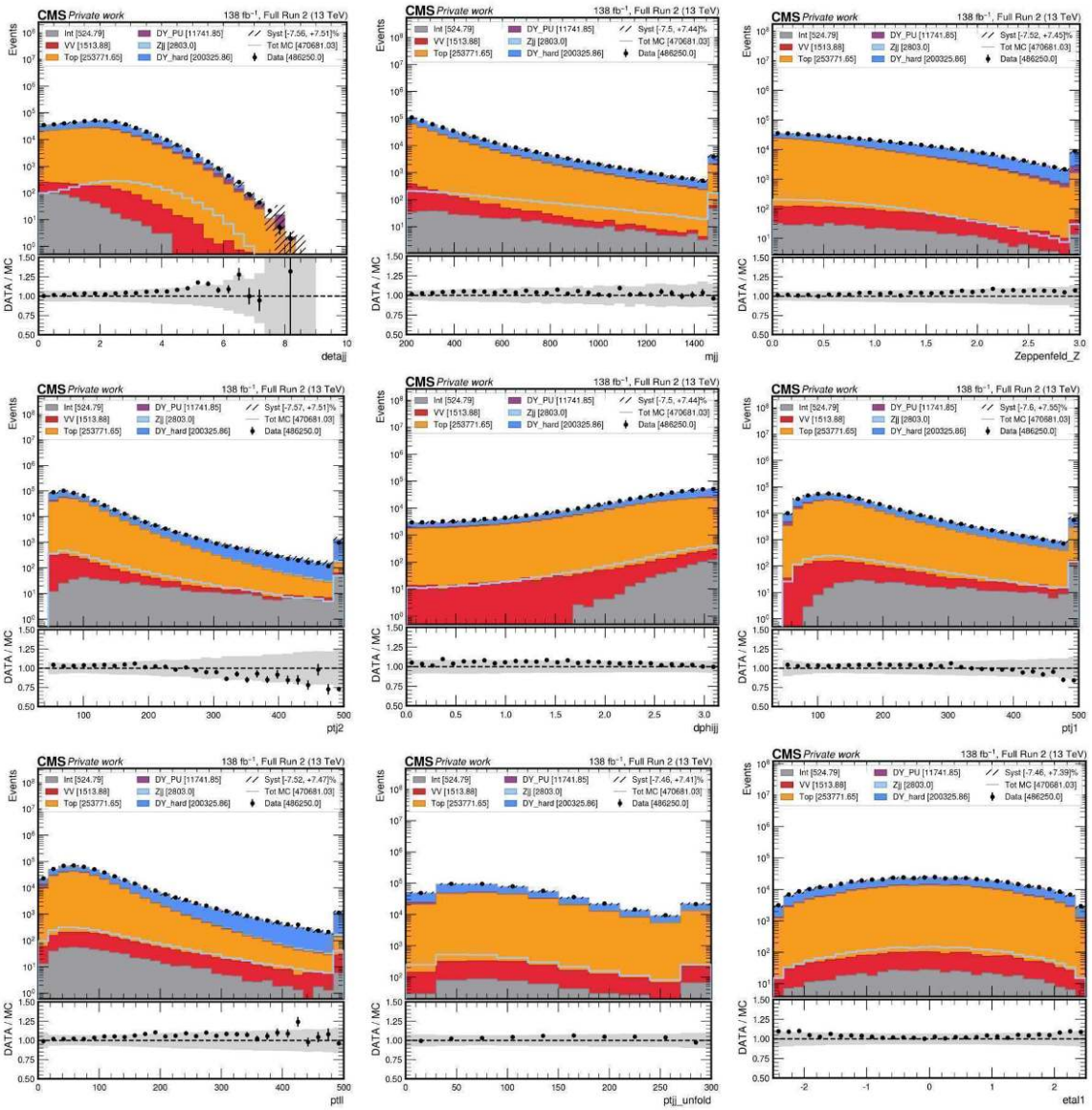

Figure A.9: Input variables to the DNN in the Top CR with DNN < 0.9. From left to right, from top to bottom, ranked by importance in the DNN: $\Delta\eta_{jj}$, m_{jj}, Z_{ll}, p^T_{j2}, $\Delta\phi_{jj}$, p^T_{j1}, p^T_{ll}, p^T_{jj}, η_{l1}

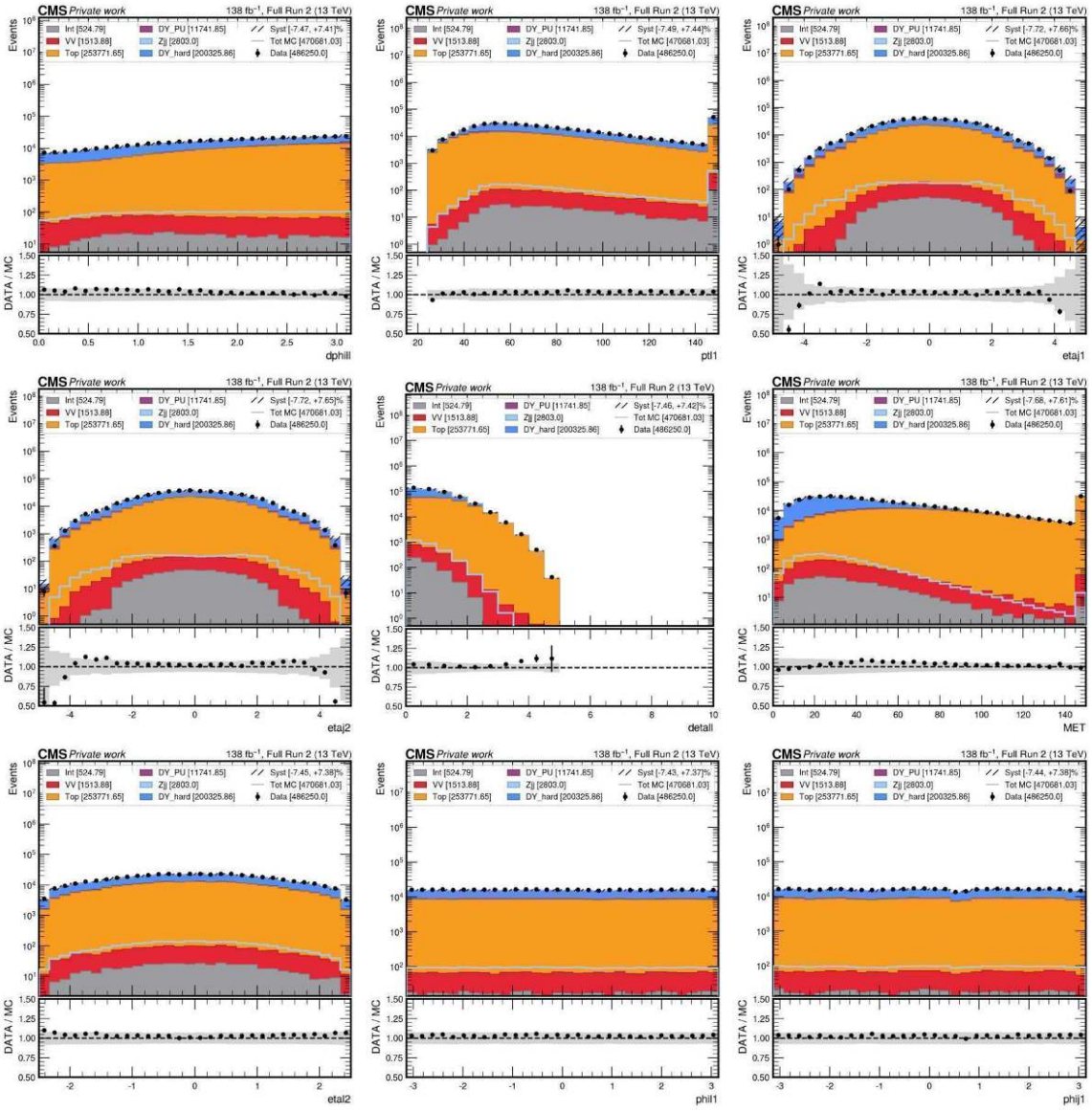

Figure A.10: Input variables to the DNN in the Top CR with DNN < 0.9. From left to right, from top to bottom, ranked by importance in the DNN: $\Delta\phi_{ll}$, p_{l1}^{T}, η_{j1}, η_{j2}, $\Delta\eta_{ll}$, MET, η_{l2}, ϕ_{l1}, ϕ_{j1}

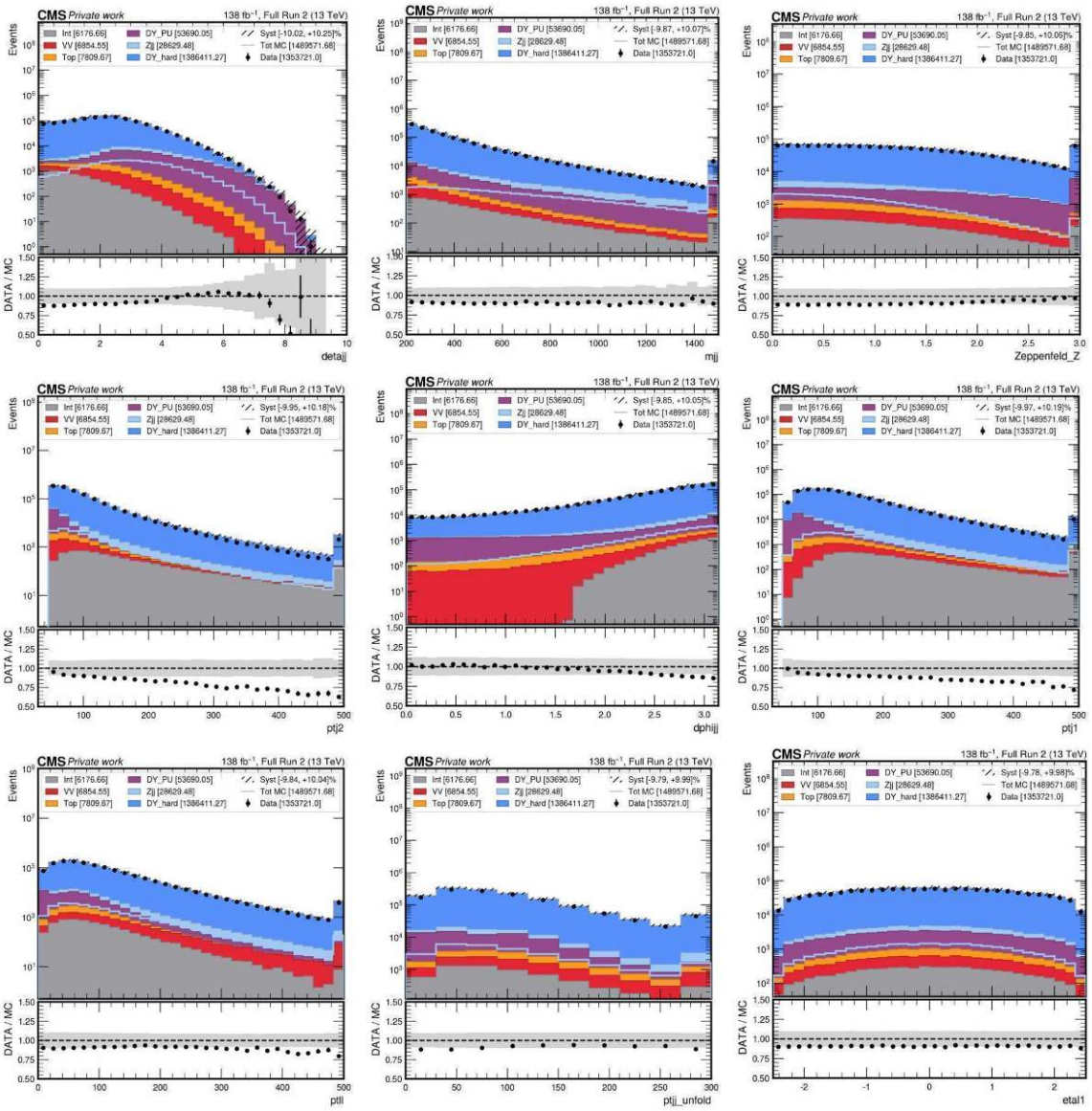

Figure A.11: Input variables to the DNN in the SR with DNN < 0.9. From left to right, from top to bottom, ranked by importance in the DNN: $\Delta\eta_{jj}$, m_{jj}, Z_{ll}, p_{j2}^T, $\Delta\phi_{jj}$, p_{j1}^T, p_{ll}^T, p_{jj}^T, η_{l1}

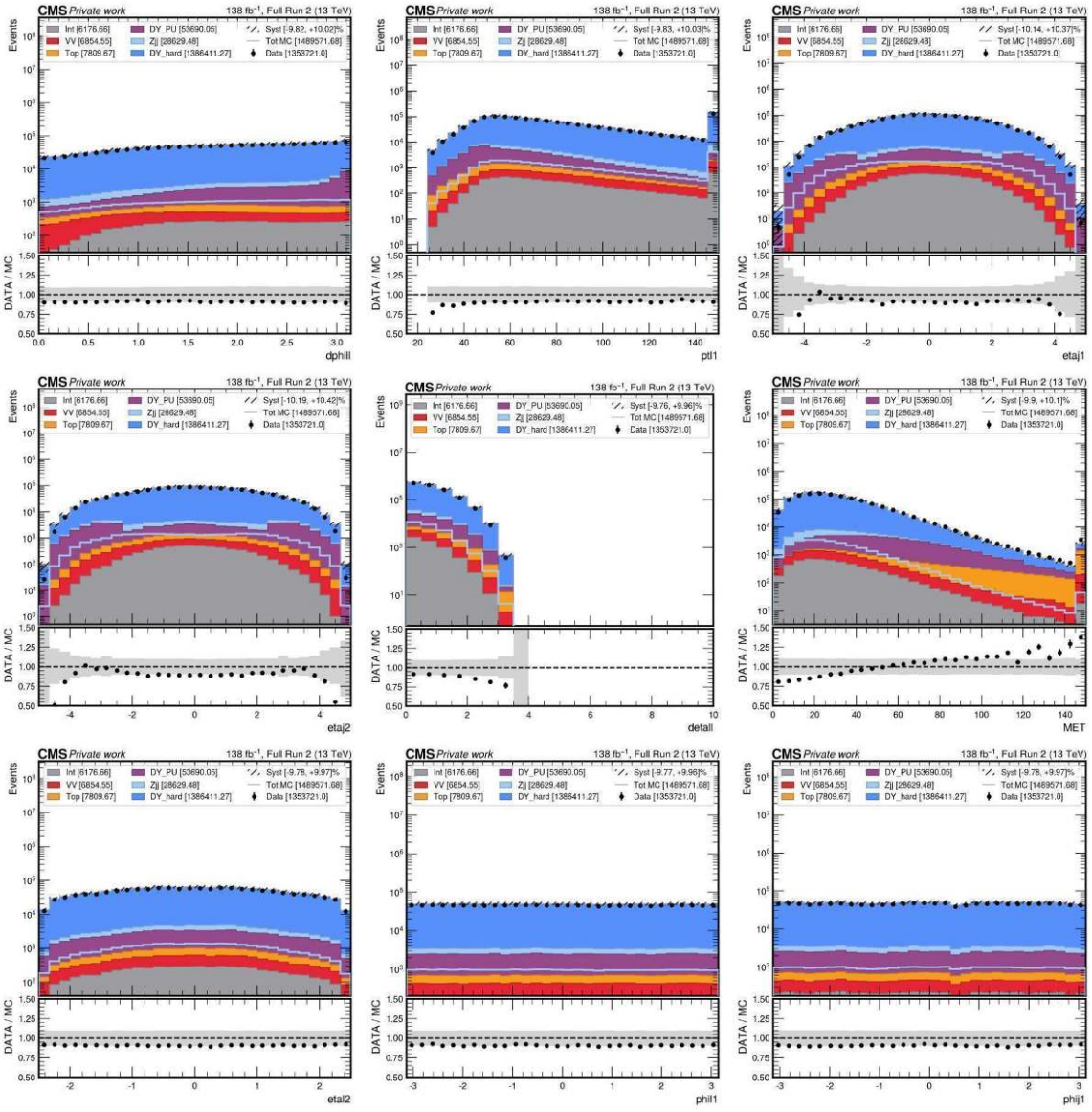

Figure A.12: Input variables to the DNN in the SR with DNN < 0.9. From left to right, from top to bottom, ranked by importance in the DNN: $\Delta\phi_{ll}$, p_{l1}^{T}, η_{j1}, η_{j2}, $\Delta\eta_{ll}$, MET, η_{l2}, ϕ_{l1}, ϕ_{j1}

Appendix B

VBF-Z EFT Analysis

B.0.1 Disagreement in $\Delta\phi_{jj}$

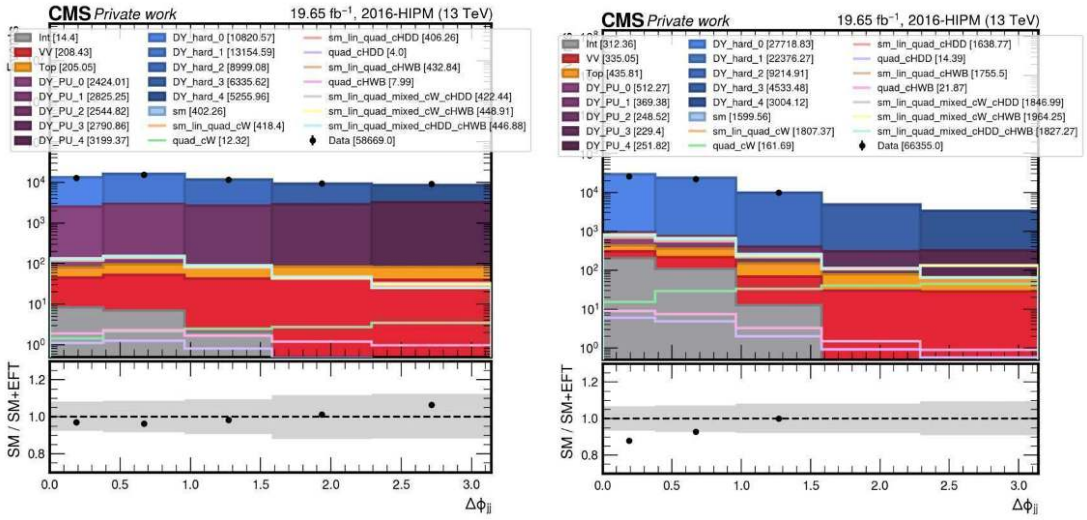

Figure B.1: Bad modelling of $\Delta\phi_{jj}$ in 2016 pre VFP, left DY PU CR, right SR, e^+e^-channel

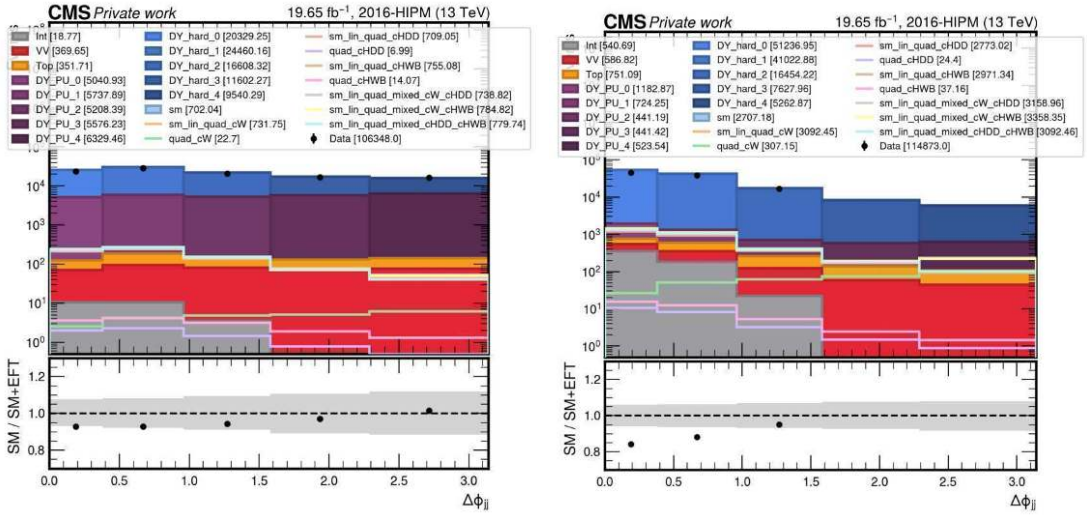

Figure B.2: Bad modelling of $\Delta\phi_{jj}$ in 2016 pre VFP, left DY PU CR, right SR, mm channel

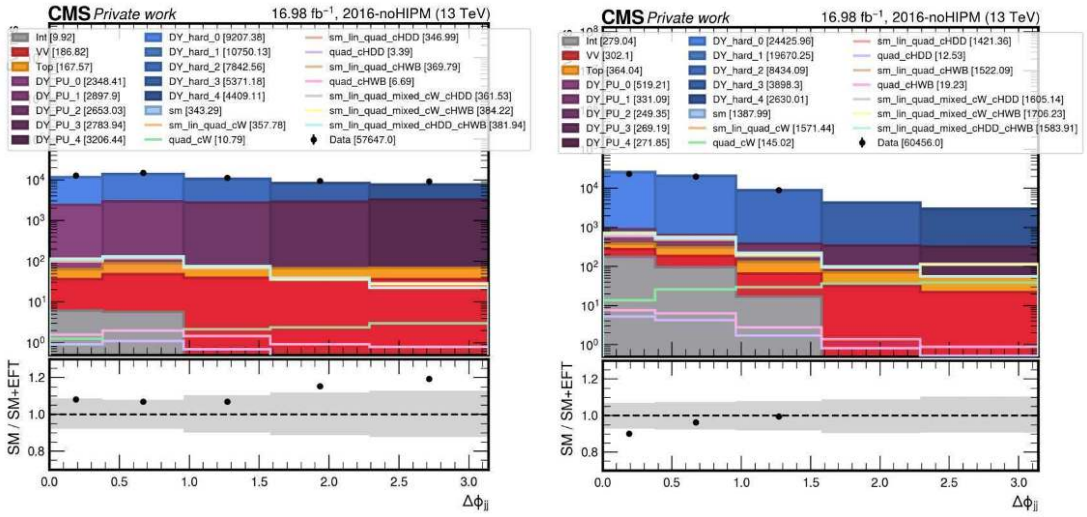

Figure B.3: Bad modelling of $\Delta\phi_{jj}$ in 2016 post VFP, left DY PU CR, right SR, e^+e^-channel

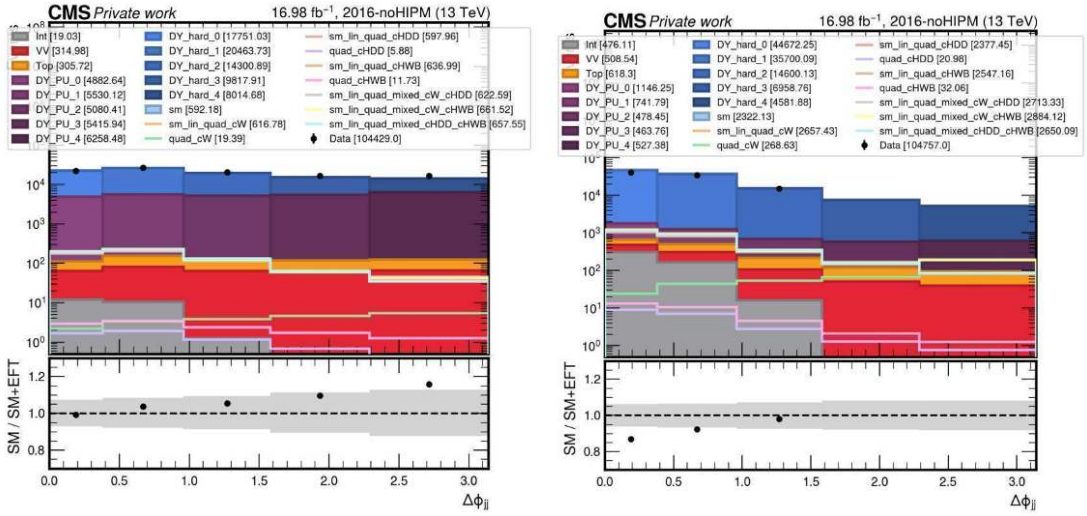

Figure B.4: Bad modelling of $\Delta\phi_{jj}$ in 2016 post VFP, left DY PU CR, right SR, mm channel

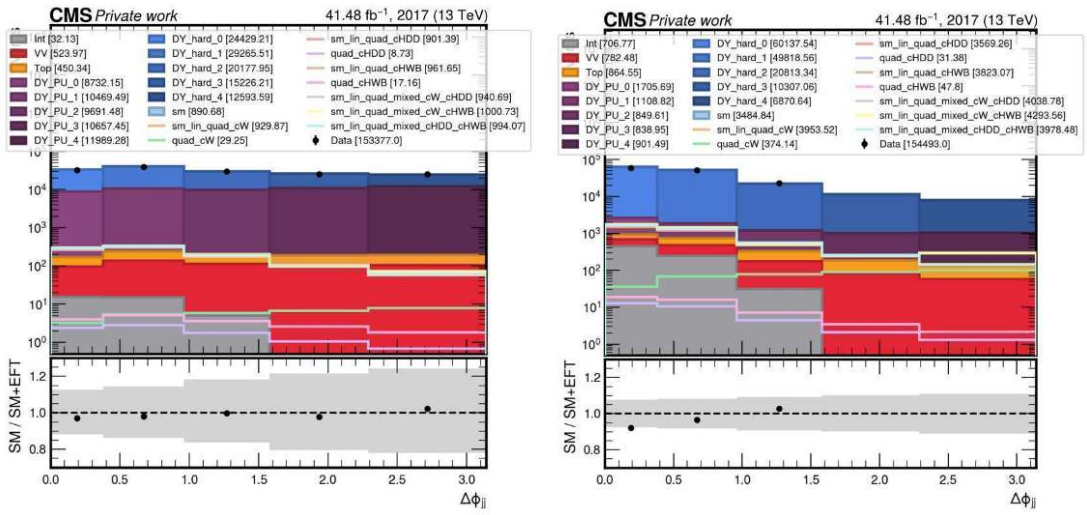

Figure B.5: Bad modelling of $\Delta\phi_{jj}$ in 2017, left DY PU CR, right SR, e^+e^-channel

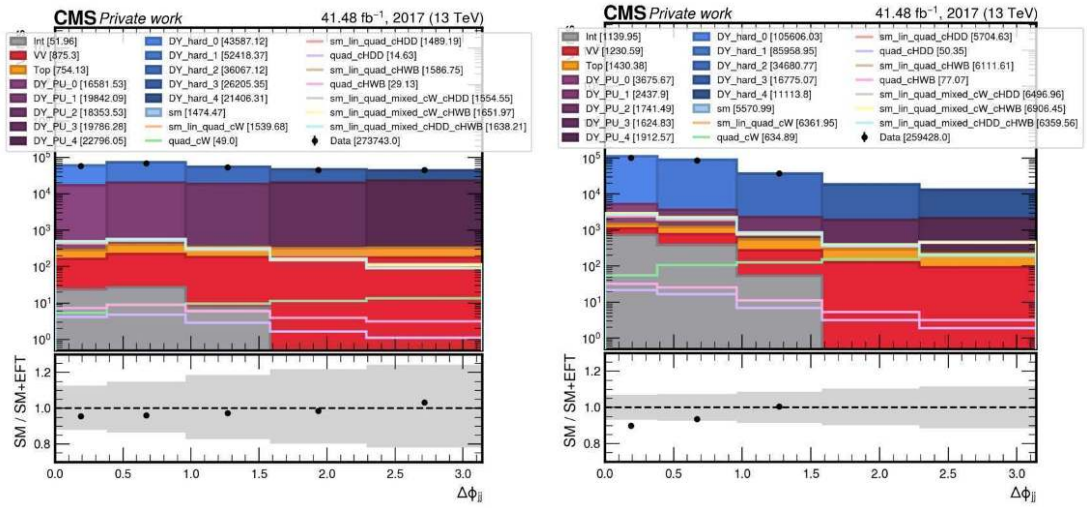

Figure B.6: Bad modelling of $\Delta\phi_{jj}$ in 2017, left DY PU CR, right SR, mm channel

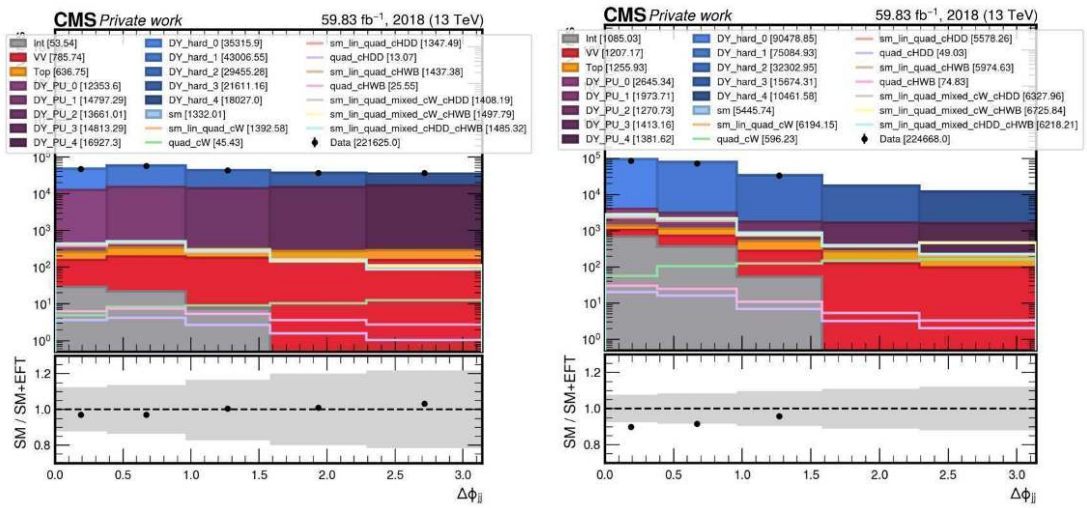

Figure B.7: Bad modelling of $\Delta\phi_{jj}$ in 2018, left DY PU CR, right SR, e^+e^-channel

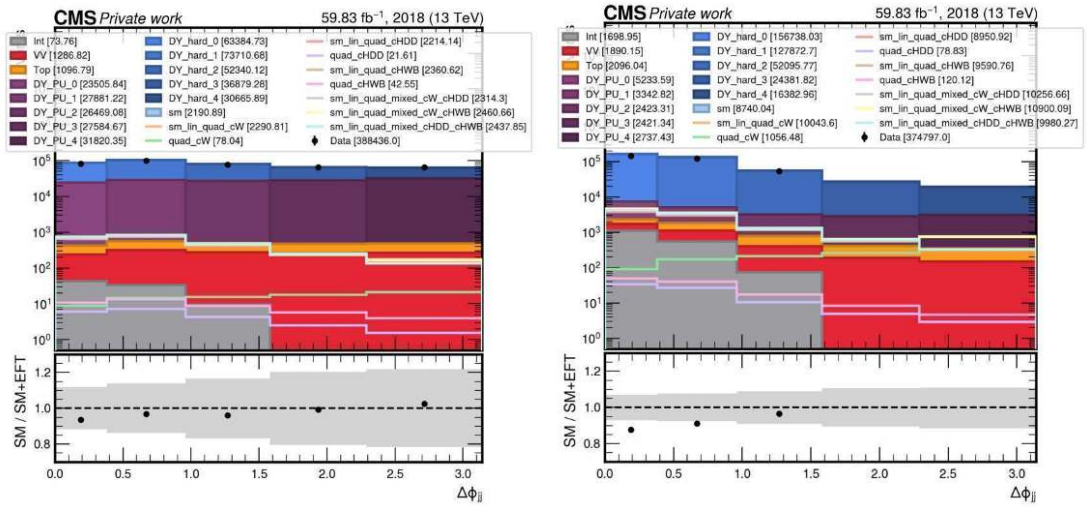

Figure B.8: Bad modelling of $\Delta\phi_{jj}$ in 2018, left DY PU CR, right SR, mm channel

B.0.2 Analysis strategy plots

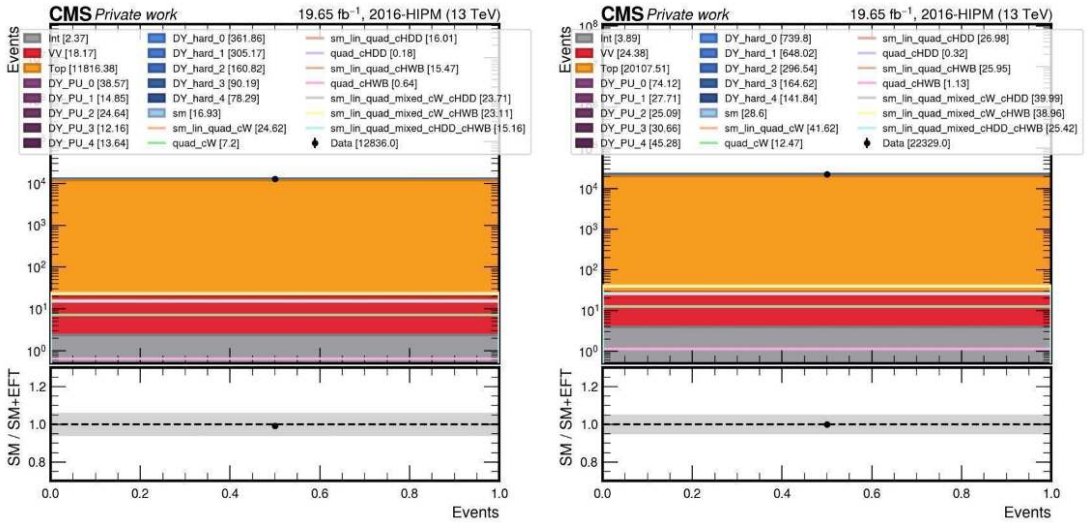

Figure B.9: Analysis strategy in 2016 pre VFP in the Top CR, left ee channel, right mm

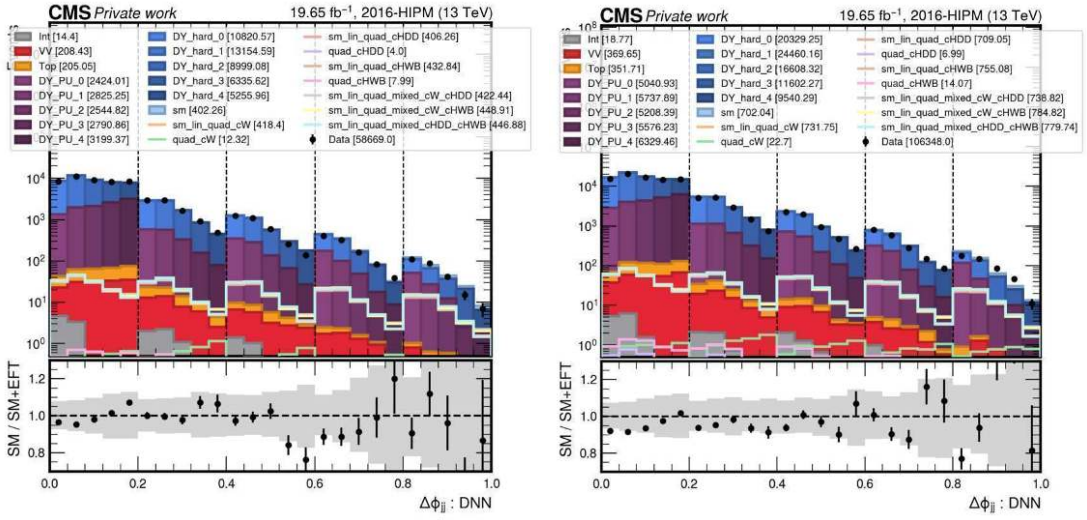

Figure B.10: Analysis strategy in 2016 pre VFP in the DY PU CR, left ee channel, right mm

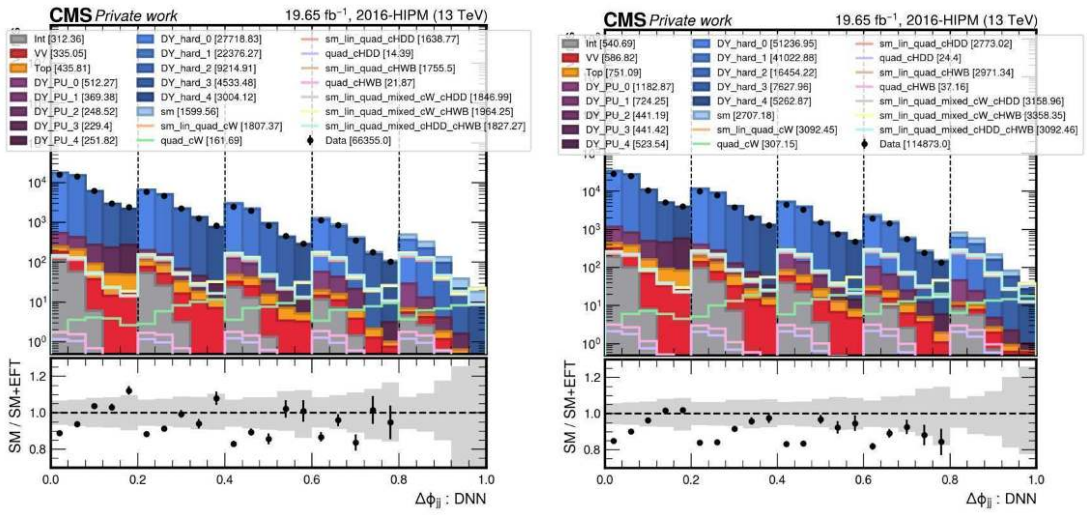

Figure B.11: Analysis strategy in 2016 pre VFP in the SR, left ee channel, right mm

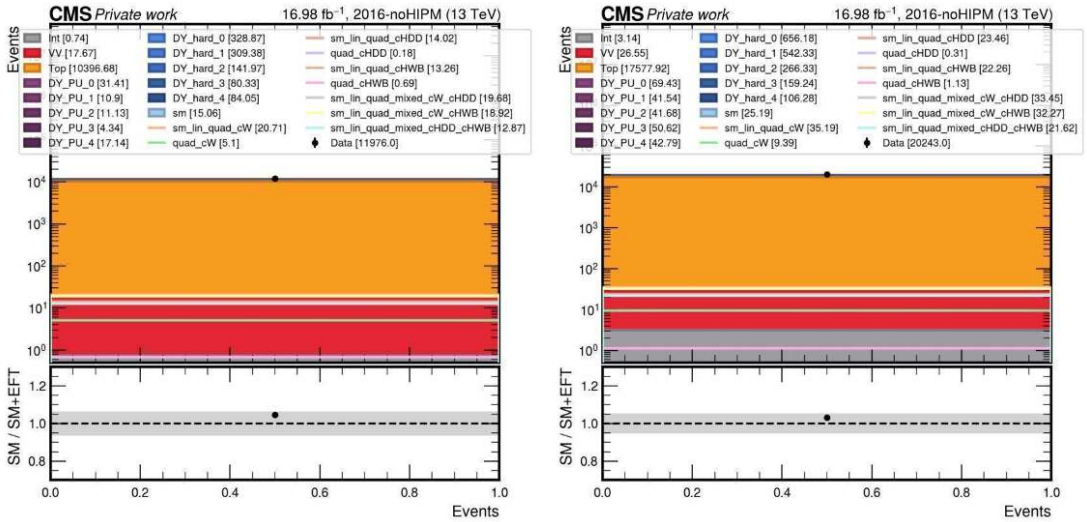

Figure B.12: Analysis strategy in 2016 post VFP in the Top CR, left ee channel, right mm

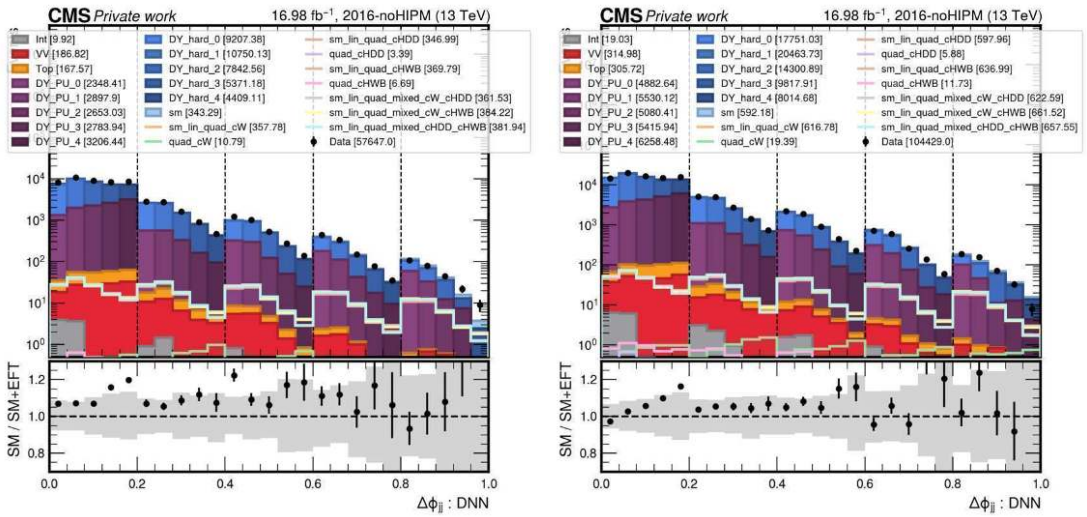

Figure B.13: Analysis strategy in 2016 post VFP in the DY PU CR, left ee channel, right mm

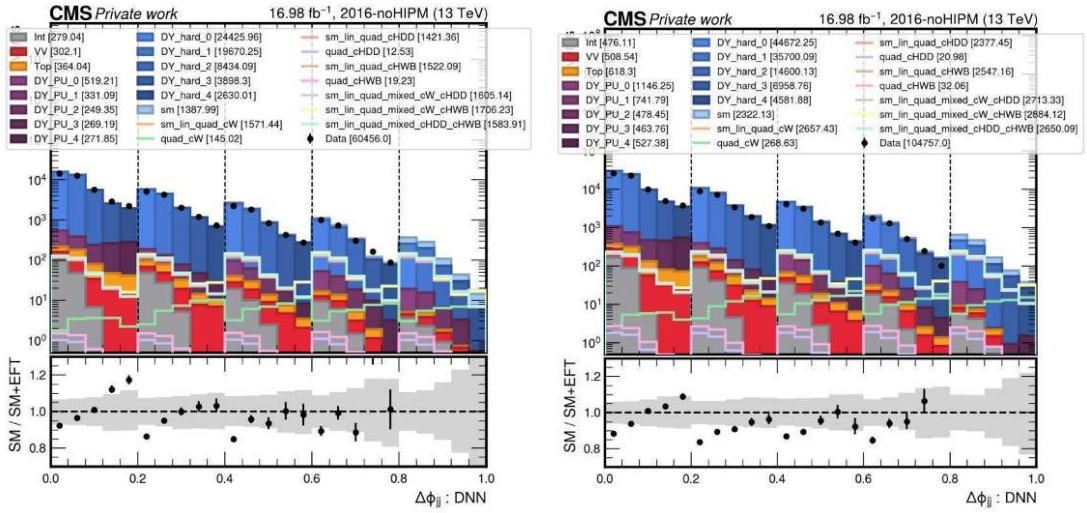

Figure B.14: Analysis strategy in 2016 post VFP in the SR, left ee channel, right mm

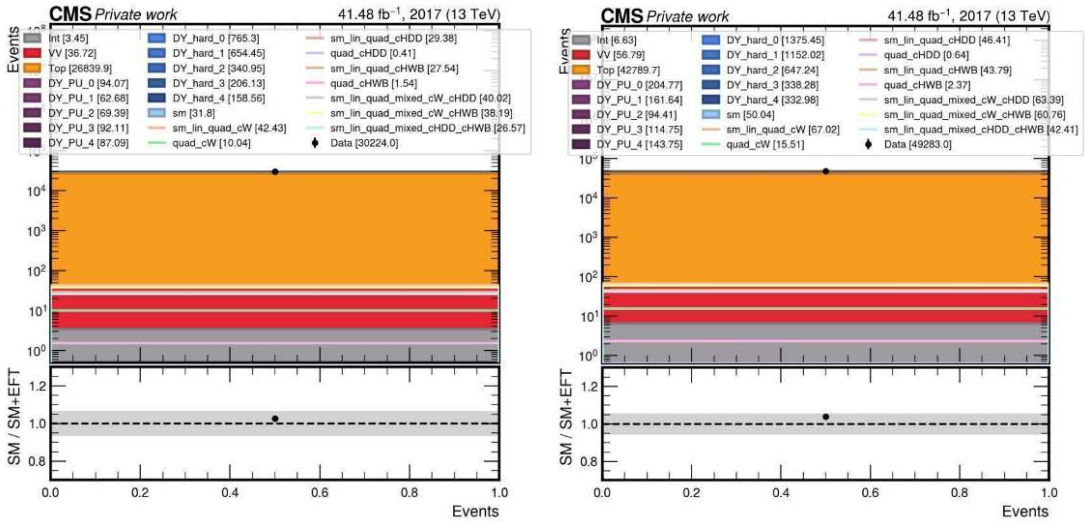

Figure B.15: Analysis strategy in 2017 in the Top CR, left ee channel, right mm

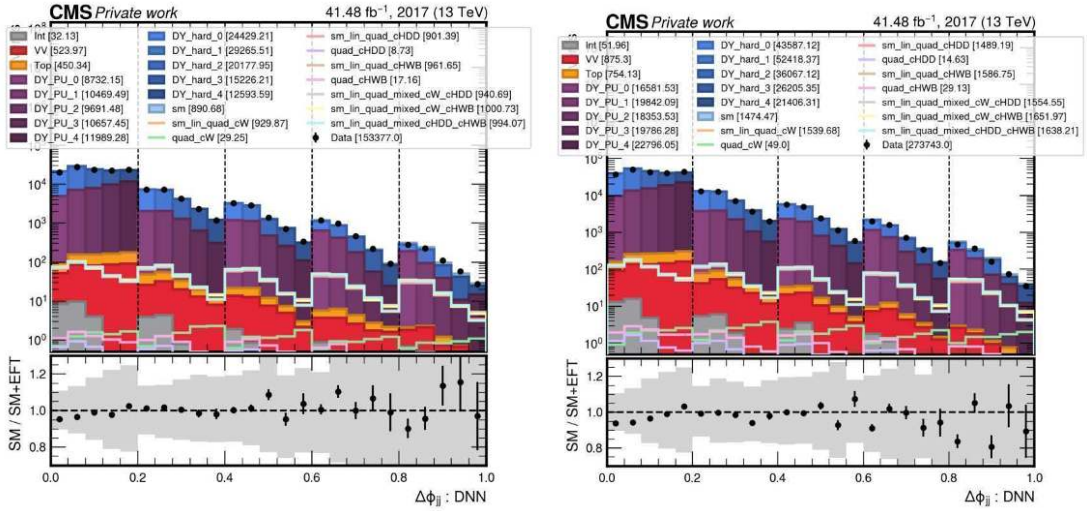

Figure B.16: Analysis strategy in 2017 in the DY PU CR, left ee channel, right mm

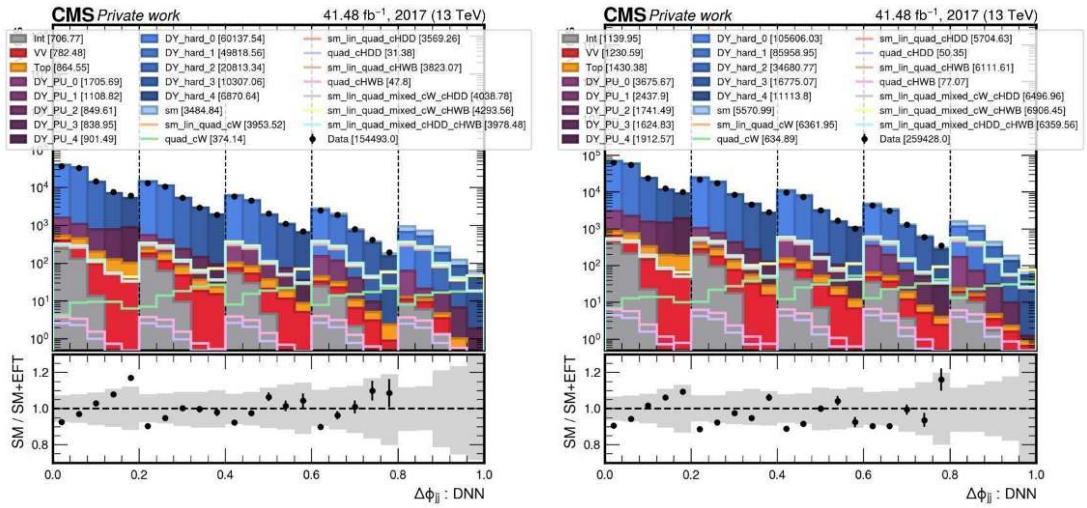

Figure B.17: Analysis strategy in 2017 in the SR, left ee channel, right mm

B.0.3 Post-fit plots

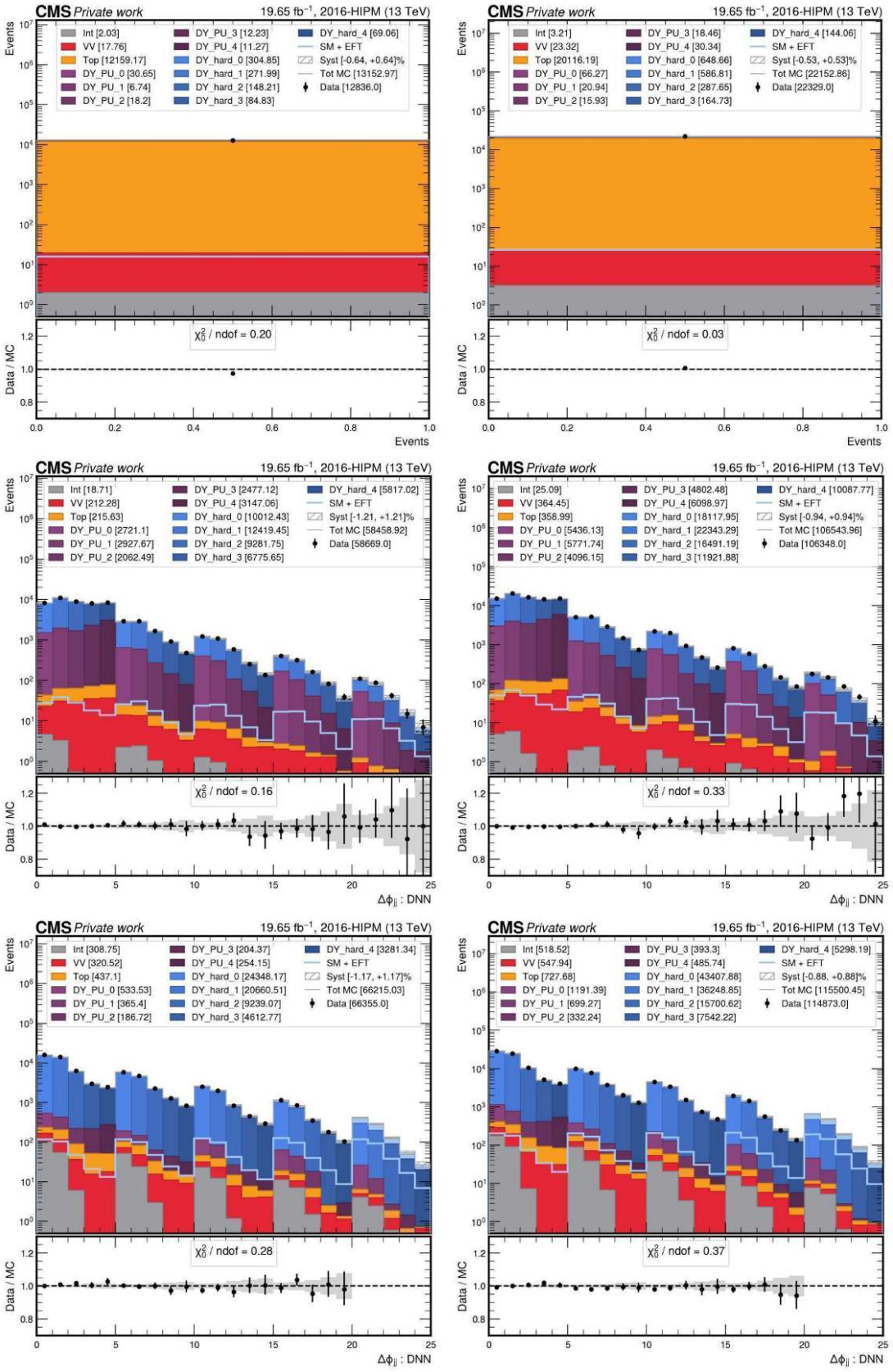

Figure B.18: Post fit distributions in 2016 pre VFP. On the top, the Top CR, middle the DY PU CR, bottom the SR. Left is the e^+e^-channel, right $\mu^+\mu^-$

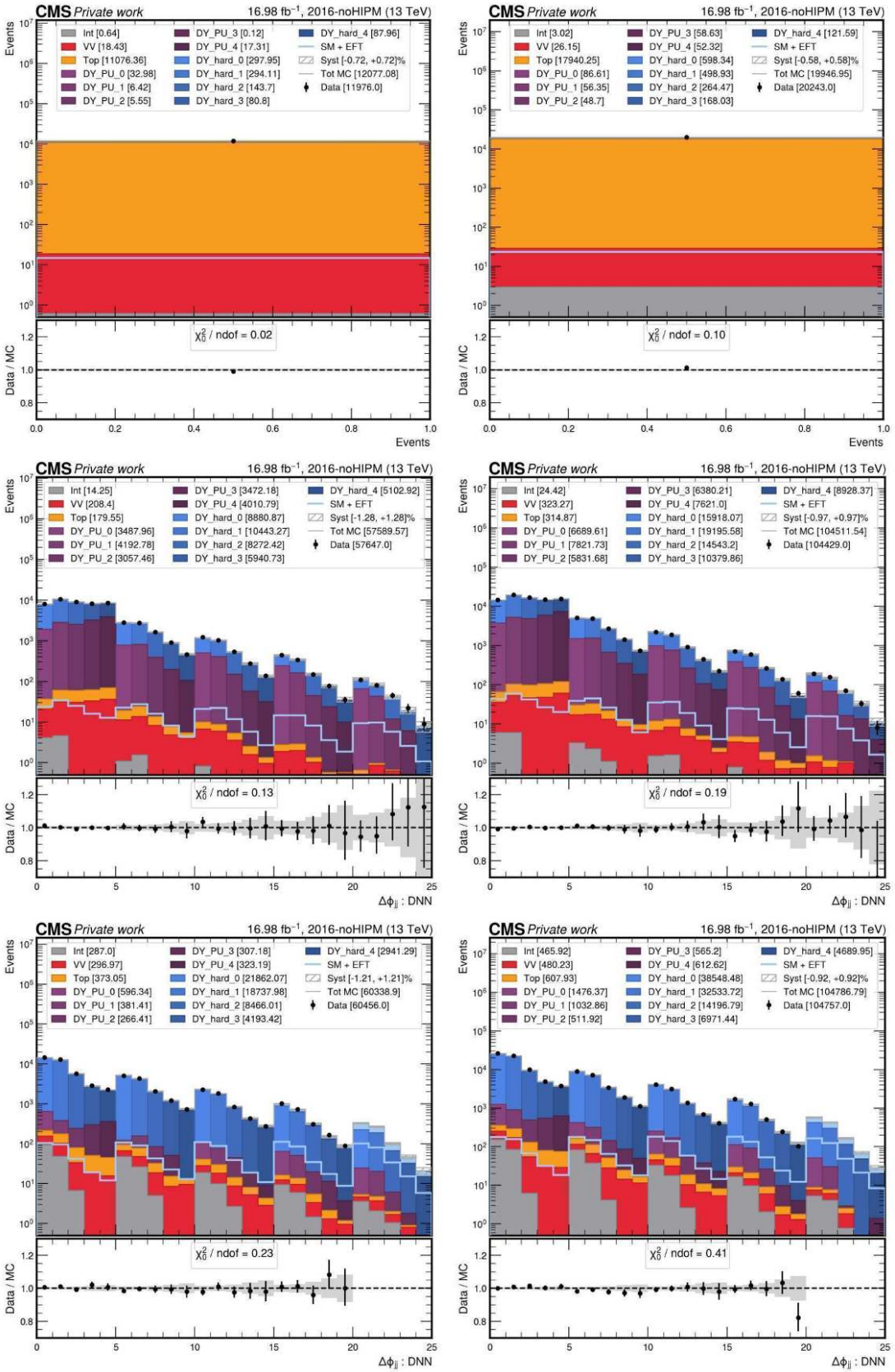

Figure B.19: Post fit distributions in 2016 post VFP. On the top, the Top CR, middle the DY PU CR, bottom the SR. Left is the e^+e^-channel, right $\mu^+\mu^-$

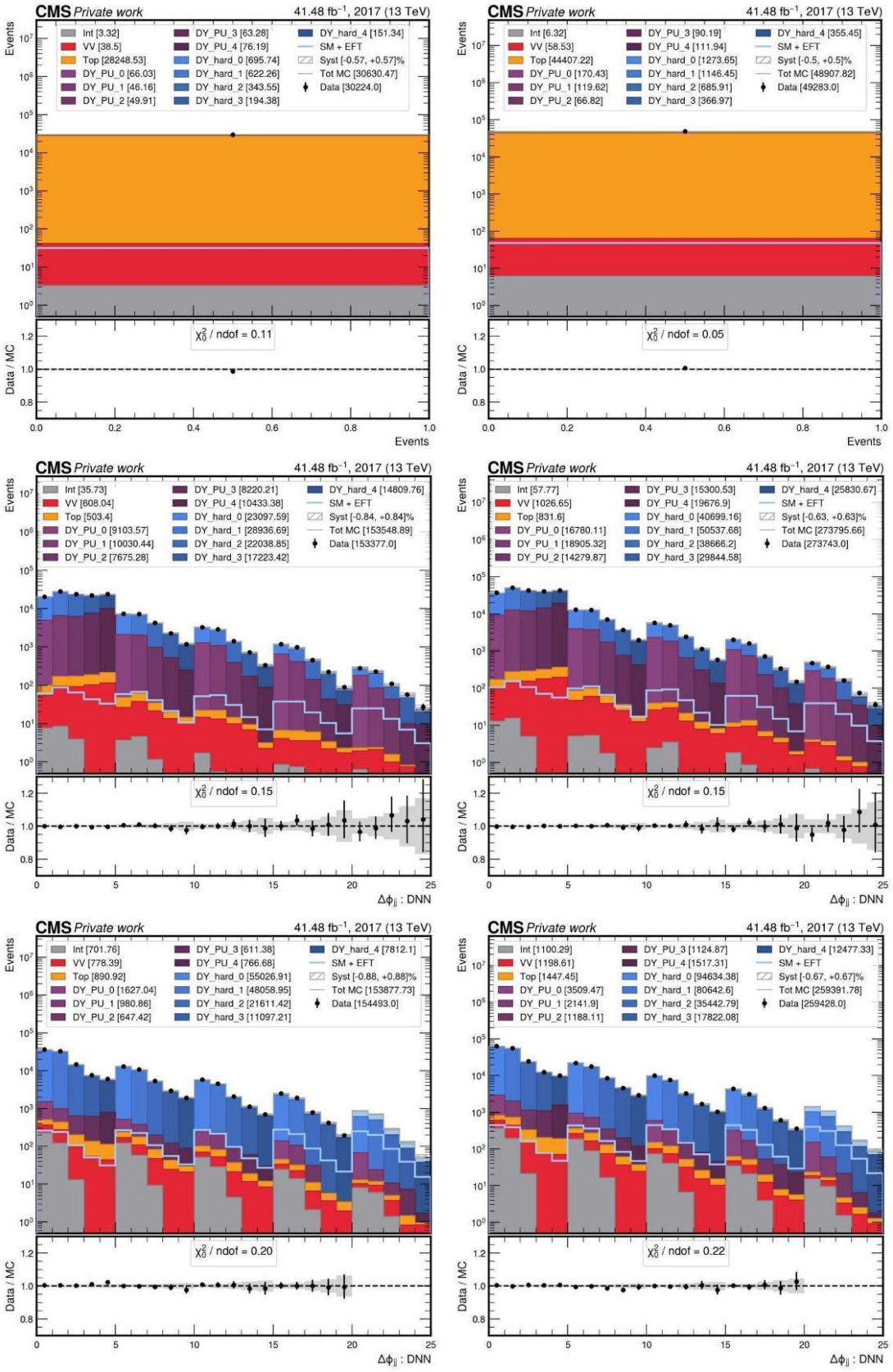

Figure B.20: Post fit distributions in 2017. On the top, the Top CR, middle the DY PU CR, bottom the SR. Left is the e^+e^-channel, right $\mu^+\mu^-$

B.0.4 Impact breakdown for the EFT parameters

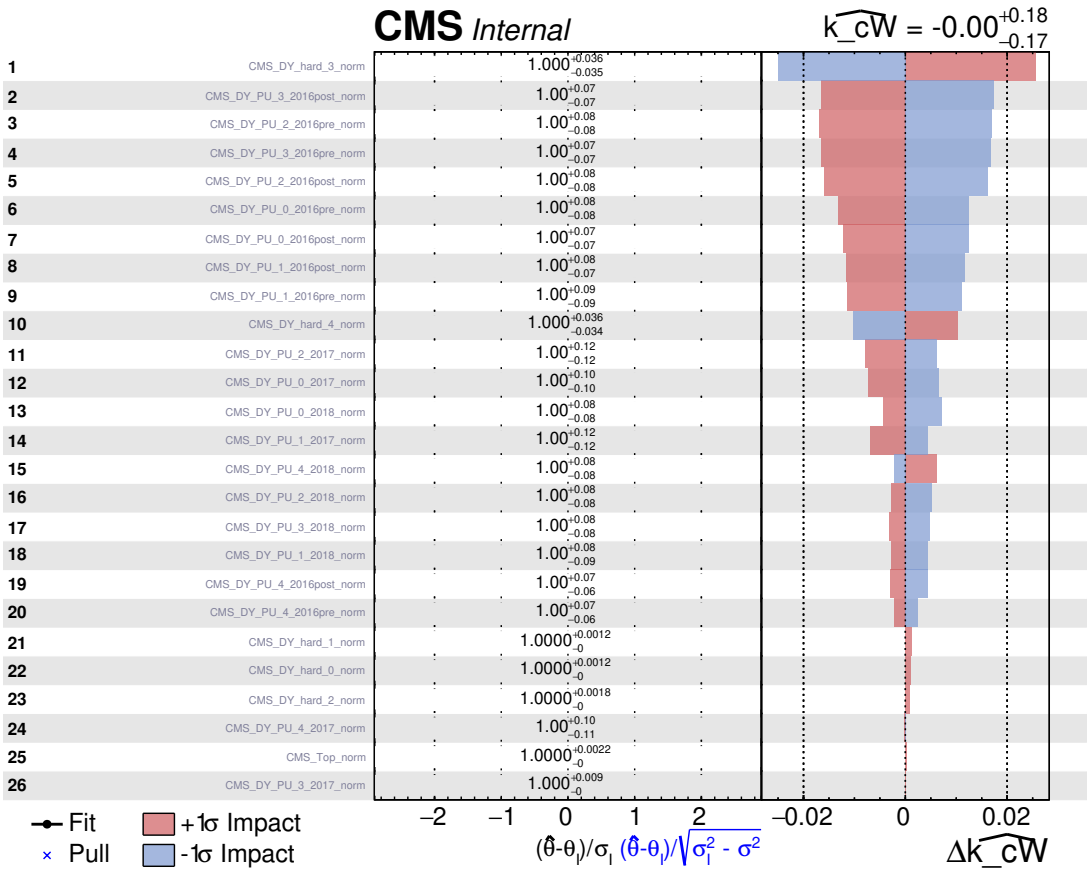

Figure B.21: Ranking of nuisance parameters, of normalization type, according to their impact on the c_W.

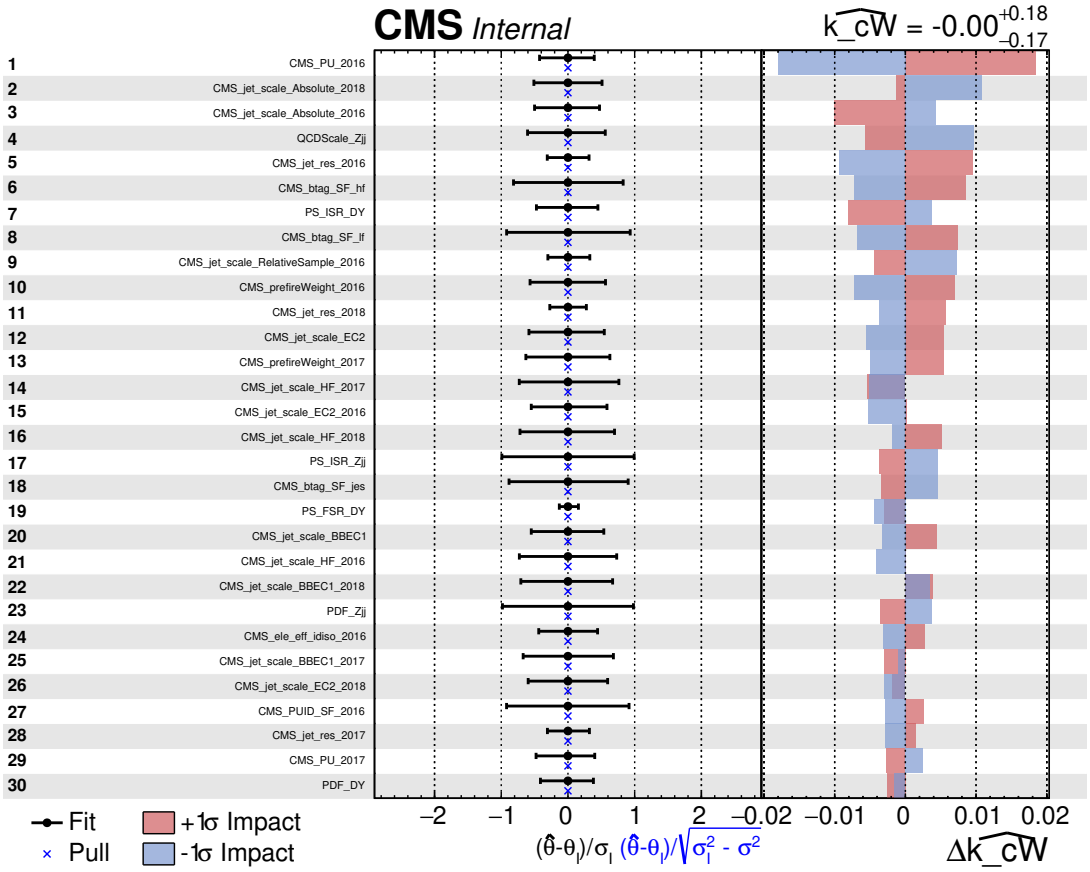

Figure B.22: Ranking of nuisance parameters, of shape systematic type, according to their impact on the c_W.

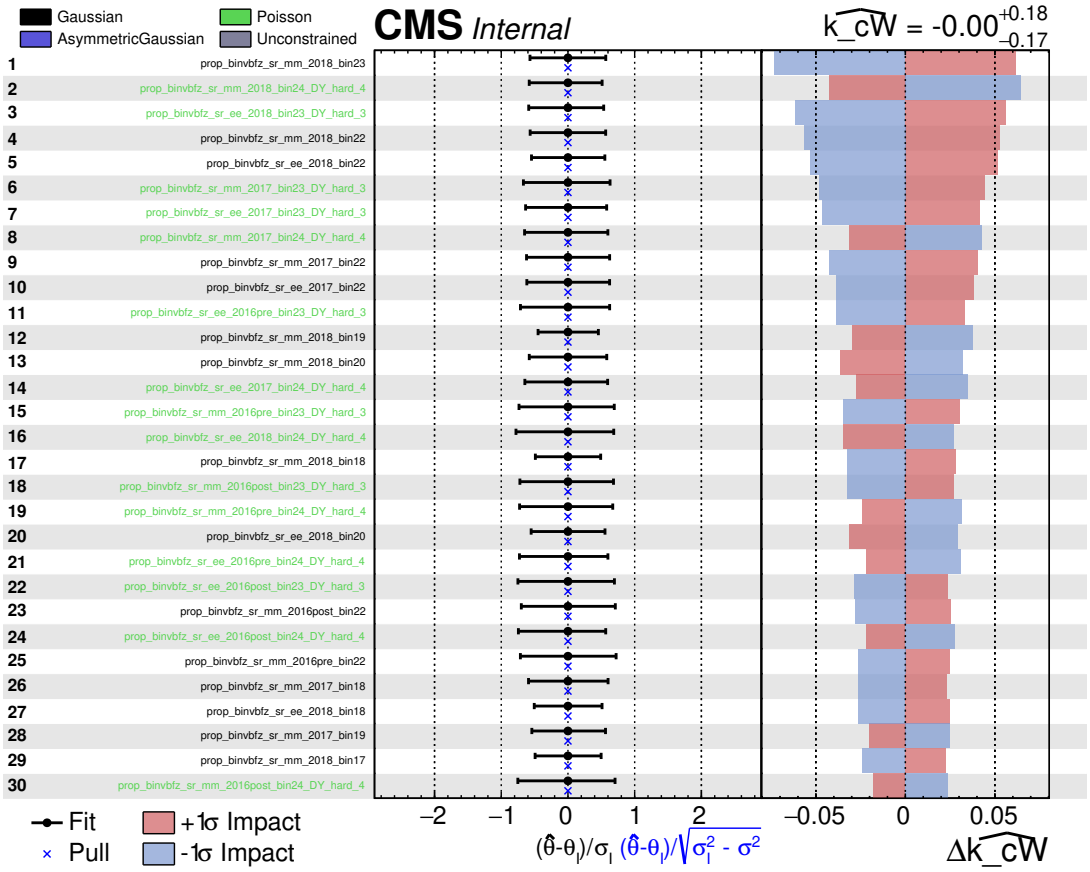

Figure B.23: Ranking of nuisance parameters, of MC statistical type, according to their impact on the c_W.

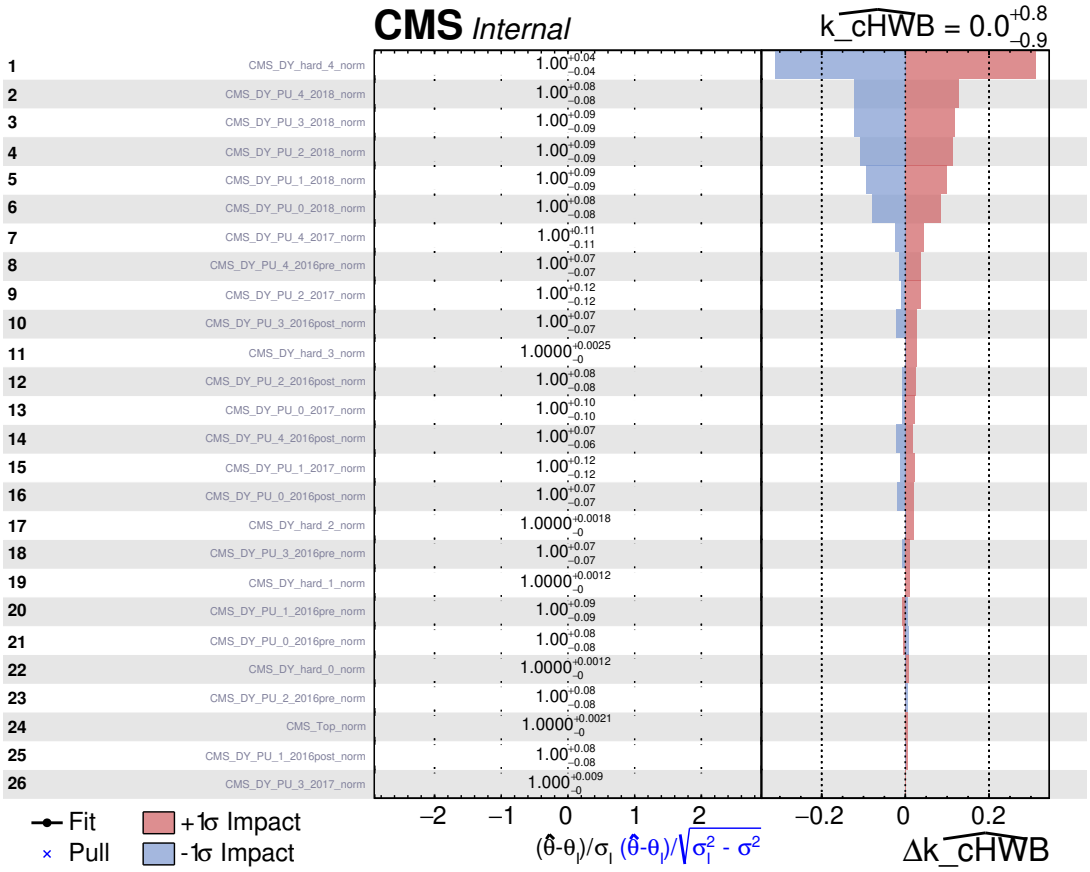

Figure B.24: Ranking of nuisance parameters, of normalization type, according to their impact on the c_{HWB}.

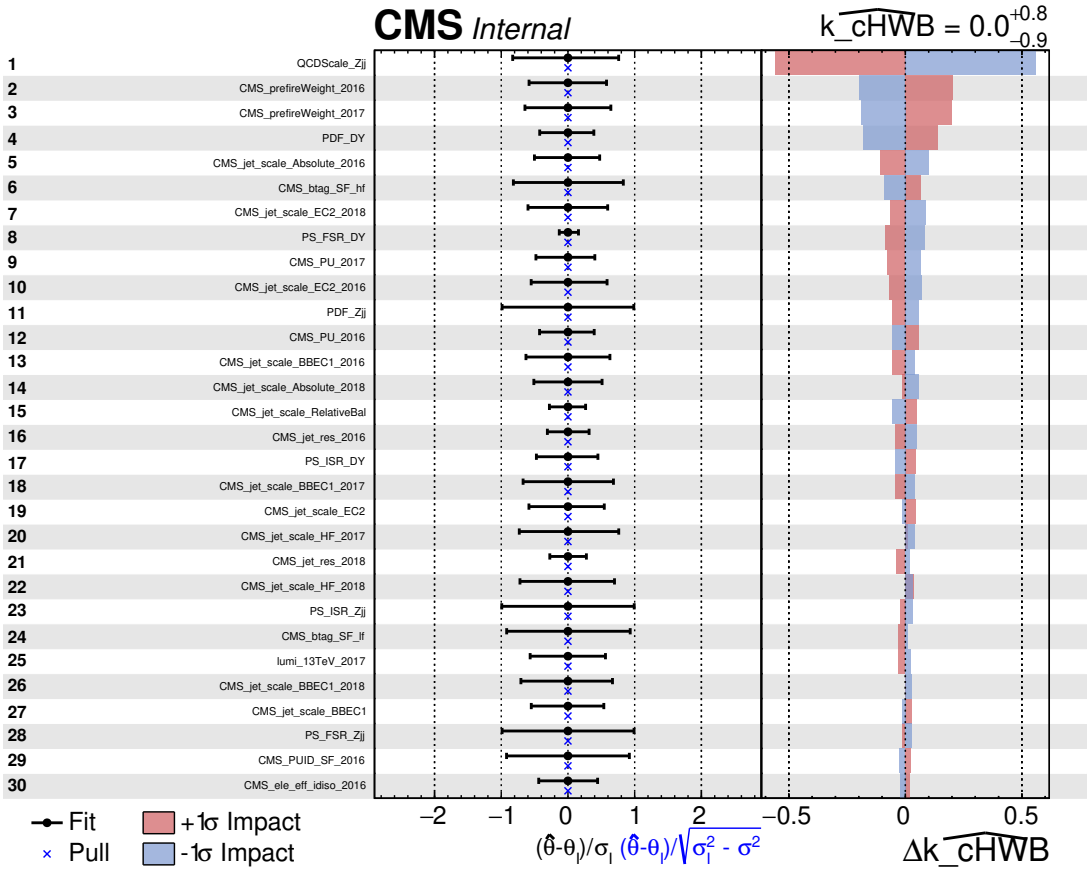

Figure B.25: Ranking of nuisance parameters, of shape systematic type, according to their impact on the c_{HWB}.

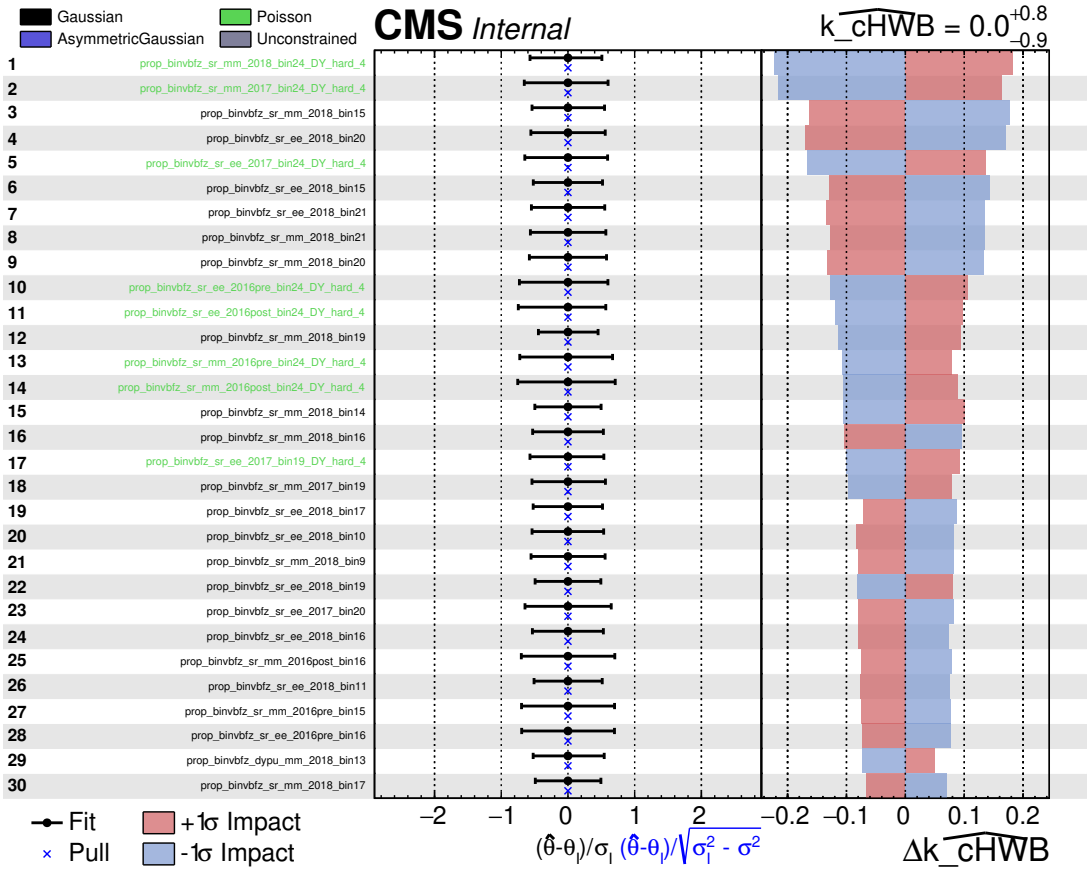

Figure B.26: Ranking of nuisance parameters, of MC statistical type, according to their impact on the c_{HWB}.

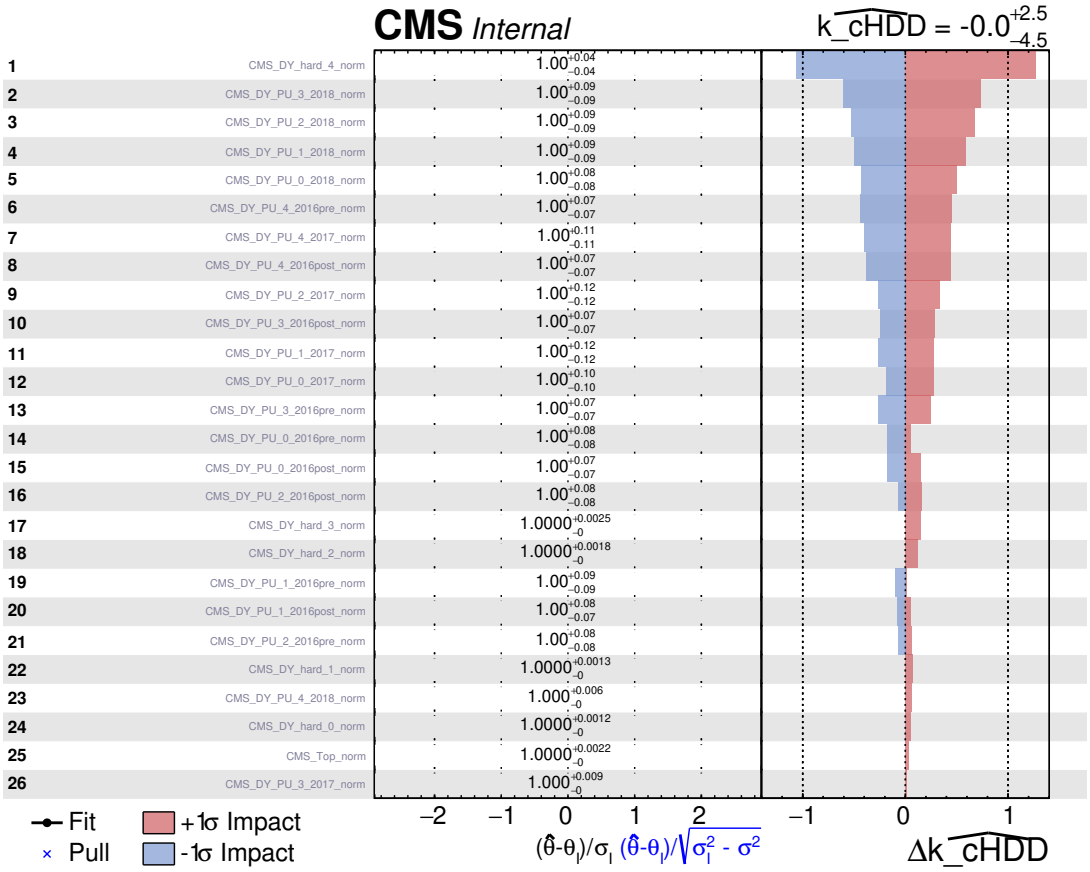

Figure B.27: Ranking of nuisance parameters, of normalization type, according to their impact on the c_{HDD}.

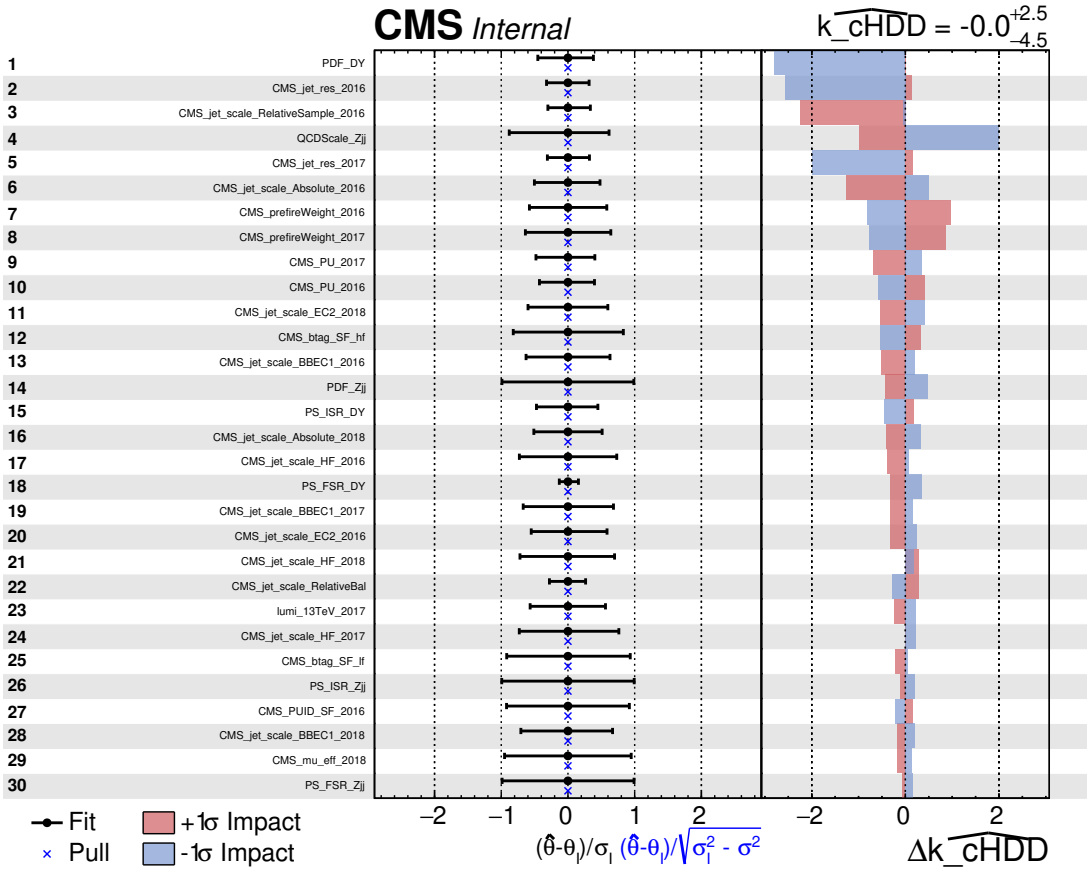

Figure B.28: Ranking of nuisance parameters, of shape systematic type, according to their impact on the c_{HDD}.

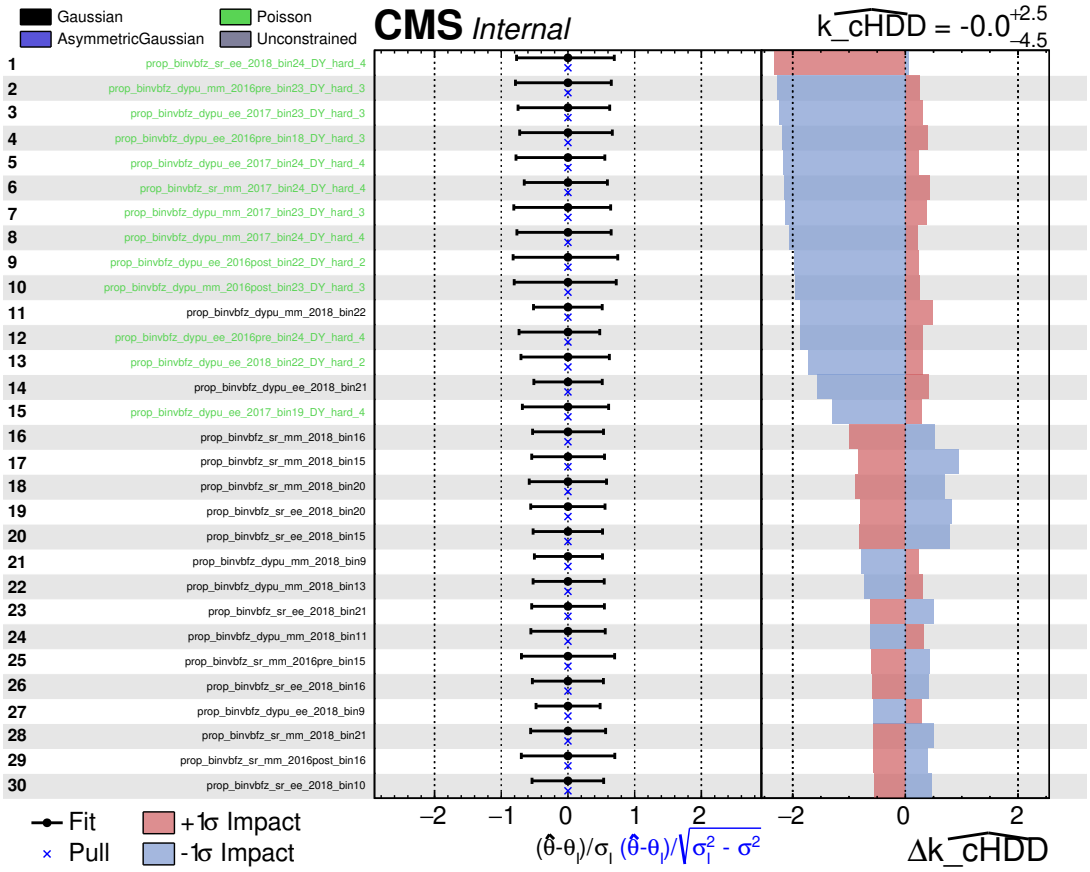

Figure B.29: Ranking of nuisance parameters, of MC statistical type, according to their impact on the c_{HDD}.

Appendix C

MTD lpGBT and GBT-SCA Details

C.0.1 lpGBT

lpGBT Communication and Initialization

The lpGBT software suite is organized into three classes with increasing levels of abstraction. Two of these handle low-level communication, which is fully transparent to the user, while the main class manages standard initialization routines. This top-level class inherits from the central `lpgbt_control_lib`, ensuring compatibility and support from the original lpGBT software framework developed by CERN.
At the lowest level, the `lpgbt_com` class interfaces directly with the μHal hardware abstraction layer, inheriting from the chip class. It provides the core methods for integrated circuit (IC) read and write operations. A dedicated function is available to optimize write performance through a "multi-write" mode, which aggregates multiple data frames into a single transfer, achieving approximately a factor-of-two improvement in communication speed. In standard read/write mode, data are sent to or retrieved from the ASICs upon each query. In contrast, the multi-write mode temporarily stores the addresses and data in vectors and transmits them in bulk, reducing I/O overhead. The `lpgbt_com` interface also supports I^2C communication through the GBT-SCA. A lightweight wrapper class, `lpgbt_cont`, provides a simplified interface to `lpgbt_com`, exposing only `read` and `write` functions. This class adapts the input and output data formats according to the selected communication protocol and datapath configuration.
The high-level `lpgbt_chip` class is designed for general user operations. Upon instantiation, it automatically initializes the communication between the FPGA and the lpGBT, abstracting all low-level control details.
The lpGBT supports three main configuration flows: via serial control, I^2C, and e-fuses. In the current implementation, initialization is performed through the serial control channel, which provides the minimal configuration required to establish communication with the counting room. This includes setting up the reference clock, locking the Phase-Locked Loop (PLL) and Delay-Locked Loop (DLL), and configuring the Clock and Data Recovery (CDR) circuit from a Read-Only Memory (ROM). Additional parameters must also be defined for the downlink high-speed equalizer, the line drivers connecting the lpGBT to the laser driver, and the EC/IC channel settings. These configurations must be applied carefully, as modifying them can compromise the integrity of the serial link. Once the minimal setup is complete, the lpGBT can be powered up, specifying whether the PLL and DLL configurations have been written. During power-up, the chip executes an automatic initialization sequence managed by the Power-Up State Machine (PUSM), a

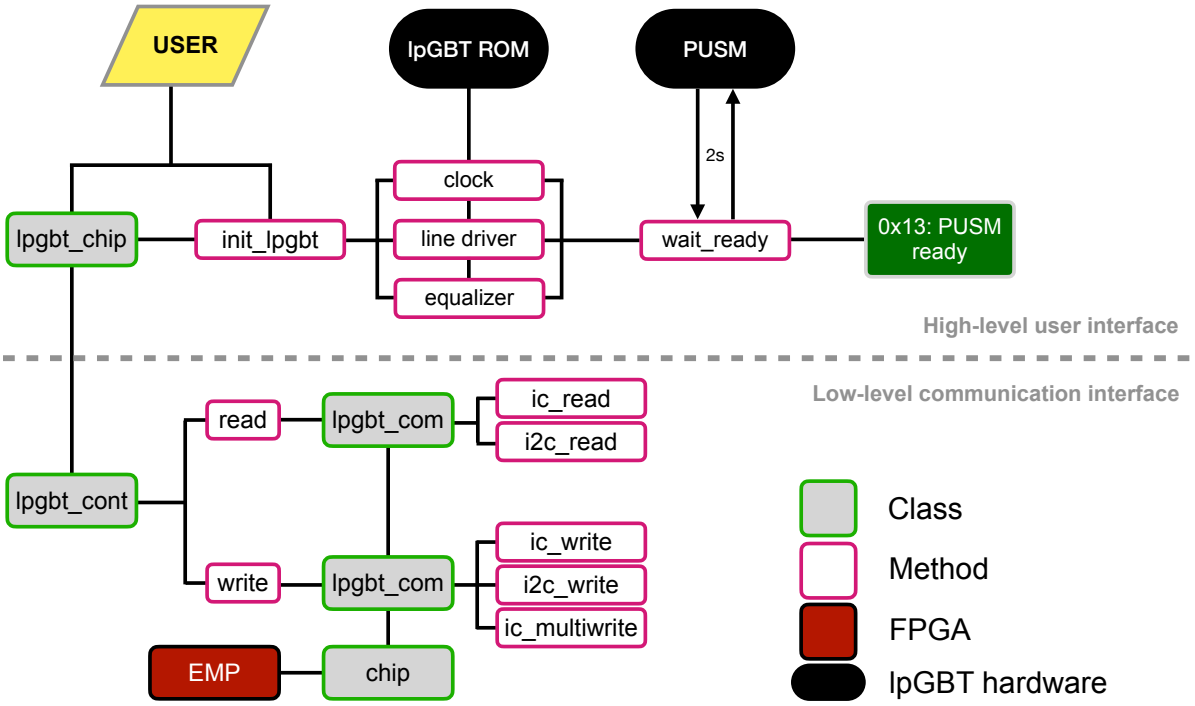

Figure C.1: The lpGBT software architecture and the initialization procedure

finite-state machine that issues resets to various subsystems and monitors their status until the device reaches full operational readiness. The software waits until the PUSM reaches the "ready" state (hexadecimal 0x13) before proceeding.
The complete lpGBT initialization flow, together with the corresponding software architecture, is illustrated in Figure C.1.
Once the lpGBT is powered on and reaches the ready state, the chip configuration sequence can begin. Additional settings for the EC connections are required, primarily to define the delay-line phase. When receiving data from the ASICs via the EC channel, the ePorts can operate either in a fixed-phase mode or in a continuous phase-tracking mode. For data transmission through the EC ePorts, the main configurable parameters include the output driver strength, pre-emphasis level, and signal polarity inversion.

lpGBT eLinks Configuration

To ensure proper communication with external peripherals, the lpGBT eLinks and their associated eClocks must be configured according to the hardware specifications. The 28 eClocks are set with a driver strength of 2.5 mA, pre-emphasis and signal inversion disabled, and full power supplied to the filter resistance driver. All eClocks operate at 160 MHz, except for the 28th, which is dedicated to the GBT-SCA connection. This link runs at a fixed data rate of 80 Mb/s, and therefore its corresponding ePort clock is configured to 40 MHz.
The eLink configuration is divided into two stages: reception (RX) and transmission (TX). For the reception stage, the delay-locked loop (DLL) of the uplink phase aligner—shared among all ePort RX groups—is configured with an 8 μA charge pump. The phase detector must report the "late" condition for 31 consecutive clock cycles. Both clock and

data gating of the DLL initialization state machine are disabled, as is the coarse lock detection. Reinitialization is enabled for each ePort RX group when the phase selection drifts out of range.
Each ePort RX group is then configured individually. In both the BTL and ETL systems, all seven ePort RX groups (each consisting of four eLinks, for a total of 28) are activated with continuous phase tracking at a data rate of 320 Mb/s. For every eLink, the signal polarity remains non-inverted, 100 Ω termination and AC biasing are enabled, and the fixed phase value is set to zero. The continuous phase tracking dynamically adjusts the phase to its optimal value during operation. The only exception is the ePort RX group 3 (for lpGBT0) and group 2 (for lpGBT1) in the BTL CC v2.3 design, which host the GBT-SCA eLink connection. In this configuration, only channel 0 of group 3 (for lpGBT0) and channel 3 of group 2 (for lpGBT1) are activated.
The transmission stage consists of four ePort TX groups, each comprising four eLinks. The TX groups do not implement DLLs. Group 3 is hardwired to the GBT-SCA and requires specific configuration. All ePort TX groups share a common setup: all channels are enabled with a data rate of 80 Mb/s, mirroring and pre-emphasis are disabled, and the driver strength is set to 2.0 mA. Only a subset of transmission eLinks is used for downlink communication, carrying trigger and RESYNC signals, as well as data from the GBT-SCA.
A summary of the RX and TX ePort group configurations for both lpGBTs implemented on the latest Central Controller version (v2.3) is reported in Table C.1.

	Receiving ePorts (RX)				**Transmission ePorts (TX)**			
	lpGBT0		**lpGBT1**		**lpGBT0**		**lpGBT1**	
Group.Channel	**Function**	**[Mb/s]**	**Function**	**[Mb/s]**	**Function**	**[Mb/s]**	**Function**	**[Mb/s]**
0.0	FE10.1	320	FE9.1	320	Trig	80	-	-
0.1	FE10.0	320	FE8.1	320	-	-	-	-
0.2	FE8.1	320	FE10.0	320	Rx	80	Trigger	80
0.3	FE8.0	320	FE10.1	320	-	-	-	-
1.0	FE7.1	320	FE8.0	320	RESYNC	80	Rx	80
1.1	FE7.0	320	FE9.0	320	-	-	-	-
1.2	FE1.0	320	FE7.1	320	Trigger Alt.	-	RESYNC	80
1.3	FE1.1	320	FE7.0	320	-	-	-	-
2.0	FE2.1	320	-	-	Rx Alt.	80	-	-
2.1	FE2.0	320	-	-	-	-	-	-
2.2	FE4.0	320	-	-	RESYNC Alt.	80	-	-
2.3	FE4.1	320	GBTB-IN	320	-	-	-	-
3.0	GBTA-IN	320	FE2.0	320	GBTA Out	80	GBTB Out	80
3.1	-	-	FE2.1	320	-	-	-	-
3.2	-	-	FE1.0	320	-	-	-	-
3.3	-	-	FE1.1	320	-	-	-	-
4.0	FE3.1	320	FE4.1	320				
4.1	FE3.0	320	FE4.0	320				
4.2	FE5.0	320	FE3.0	320				
4.3	FE6.1	320	FE3.1	320				
5.0	FE6.0	320	FE6.1	320				
5.1	FE5.1	320	FE6.0	320				
5.2	FE11.0	320	FE5.0	320				
5.3	FE11.1	320	FE5.1	320				
6.0	FE12.1	320	FE11.0	320				
6.1	FE12.0	320	FE11.1	320				
6.2	FE9.1	320	FE12.0	320				
6.3	FE9.0	320	FE12.1	320				

Table C.1: Active lpGBT ePort connections from CC v2.3. board featuring two lpGBT ASICs (0 and 1)

lpGBT Analog-to-Digital and Digital-to-Analog Converters (ADC and DAC)

The lpGBT integrates a 10-bit analog-to-digital converter (ADC) with eight multiplexed external input channels, an internal temperature sensor, and monitoring of the internal power supplies. In addition, it includes a 12-bit voltage DAC and a programmable current DAC, either of which can be attached to any of the ADC input channels. Each ASIC pin associated with the ADC is equipped with an internal current generator, allowing it to operate as a source for analog current measurements. The lpGBT also provides an internal 1.0 V reference voltage generator. A `PT1000` device can be connected to one of the ADC input channels to perform on-chip temperature measurements. The analog input ports can also be used to measure the internal or external reference voltages of the lpGBT. Although the lpGBT in the MTD is not primarily used for slow-control monitoring, the capability to read temperature values remains important for assessing the system reliability and validating hardware connections.

The accuracy of the ADC, DACs, reference voltage, internal temperature sensors, and supply monitors is improved through a two-temperature calibration procedure performed during production testing. These calibrations compensate for offset and gain variations, which can be affected by irradiation and temperature drift. Calibration data correct two dominant sources of systematic uncertainty: chip-to-chip variations and temperature-dependent effects. Other contributions, such as slope and offset errors intrinsic to the ADC and DACs, must also be characterized to achieve precise analog performance.

A critical quantity for all calibration procedures is the lpGBT junction temperature (t_J). Laboratory measurements have shown that the temperature difference between the printed circuit board (PCB) and the lpGBT junction is approximately 20 K, with a spread of about 10 K. The junction temperature can also be estimated using the uncalibrated lpGBT ADC, although this approach typically introduces larger errors.

To calibrate the reference voltage, a two-point linear calibration is applied to optimize the output around 1.0 V. The optimal tuning parameter, T_VREF, is defined as:

$$T_\mathrm{VREF} = t_\mathrm{J} \times S_\mathrm{VREF} + O_\mathrm{VREF}, \tag{C.1}$$

where S_VREF and O_VREF are the slope and offset constants provided by the manufacturer for each chip. The standard uncertainty on the reference voltage is approximately 2 mV, depending on the precision of the t_J estimate. In the current BTL setup, the optimal tuning value was determined to be 32.

Because a single ADC converter is shared among all analog inputs, no per-channel calibration is required. The ADC readings can be converted to calibrated voltages using:

$$V_\mathrm{ADC} = \mathrm{ADC[9{:}0]} \times \left(O^\mathrm{ADC}_\mathrm{slope}(G) + t_\mathrm{J} S^\mathrm{ADC}_\mathrm{slope}(G)\right) + \left(O^\mathrm{ADC}_\mathrm{off}(G) + t_\mathrm{J} S^\mathrm{ADC}_\mathrm{off}(G)\right), \tag{C.2}$$

where ADC[9:0] is the 10-bit digital output of the ADC, and O and S are the offset and slope calibration constants corresponding to a given gain configuration G.

The voltage and current DACs follow the same analytical dependence on the junction temperature as the ADC. For the current DAC, accuracy is limited by uncertainties in both the junction temperature and supply voltage, which can modify the DAC slope. Additionally, channel-dependent output resistance introduces further variation. Neglecting resistance effects, the calibrated output current $\mathrm{C_A}$ for one of the eight channels n is given by:

$$\mathrm{CDAC}^n = \left(O^{\mathrm{DAC}-n}_\mathrm{slope} + t_\mathrm{J} S^{\mathrm{DAC}-n}_\mathrm{slope}\right) \mathrm{C_A} + O^{\mathrm{DAC}-n}_\mathrm{off} + t_\mathrm{J} S^{\mathrm{DAC}-n}_\mathrm{off}. \tag{C.3}$$

Calibration coefficients O and S are provided by the manufacturer for each DAC channel. The channel output resistance is also characterized and corrected with a linear dependence on temperature:

$$R_{\text{out}} = \frac{O_{\text{R}}^{\text{DAC}-n} + t_{\text{J}} S_{\text{R}}^{\text{DAC}-n}}{\text{CDAC}^{n}}. \tag{C.4}$$

The on-chip temperature measurements are likewise corrected using linear slope and offset parameters.
In the BTL configuration employing the latest CC v2.3, which hosts two lpGBTs, the accessible temperature and voltage measurements are limited by the physical connections available on the board. ADC input channels that are not enabled cannot be read out. The available connections for the analog monitoring channels are summarized in Table C.2.

	lpGBT0			**lpGBT1**		
ADC Port	**Function**	**Enabled**	**Range**	**Function**	**Enabled**	**Range**
0	SiPM FE6 temp1	Yes	Pt1000	SiPM FE12 temp1	Yes	Pt1000
1	SiPM FE6 temp2	Yes	Pt1000	SiPM FE12 temp2	Yes	Pt1000
2	PCC A temp1	Yes	Pt1000	PCC B temp1	Yes	Pt1000
3	PCC A temp2	Yes	Pt1000	PCC B temp2	Yes	Pt1000
4	VDDA A voltage	No	0-2.5 V	VDDA C voltage	No	0-2.5 V
5	VDDA B voltage	No	0-2.5 V	VDDA D voltage	No	0-2.5 V
6	Vin voltage	No	0-15 V	CC temp2	Yes	Pt1000
7	CC temp1	Yes	Pt1000	CC temp3	Yes	Pt1000

Table C.2: Active lpGBT analog connections from CC v2.3. board featuring two lpGBT ASICs (0 and 1)

Temperature of voltage readings are made querying the ADC for a specific port and converting its value to volts. For voltage measurements this value is sent back to the uplink to the counting room while for temperature reading the voltages are further converted to equivalent resistances. Calibrations to the ADC value are applied according to the calculated correction factors as

$$\text{R}_{\text{eq}} = 0.000962441 \cdot \text{ADC}_{\text{port}} - 0.0253751 \tag{C.5}$$

The temperature is then extracted from a linear interpolation of R_{eq} with an array of 200 tabulated resistance value with a 1°C temperature difference between two points covering a temperature range between -50 and 150°C.

lpGBT General Purpose I/O (GPIO)

The lpGBT provides 16 general-purpose input/output (GPIO) pins, logically divided into two 8-pin ports: the low (L) and high (H) ports. Each pin can be independently configured as either an input or an output, with adjustable driver strength and pull-up/pull-down settings. All GPIO operations are synchronous with the internal 40 MHz system clock. In the BTL front-end, these GPIOs are used for monitoring and control of the ASICs through simple logical signals.
Each pin is managed through five dedicated registers that define:

- the pin direction (input or output),
- the output state (low or high),

- the configuration of pull-up or pull-down resistors,
- the driver strength, and
- the output slew-rate control to limit electromagnetic emissions.

While input pins can take two logical states (0 or 1), output pins support three states: logical 0, logical 1, and a third state depending on whether the output is driven low or high. Multiple GPIOs can be configured simultaneously, although all parameters are also individually programmable per pin.
It is essential to assign a defined logic level to all unused pins to prevent undefined or floating states. The simplest approach is to enable the internal pull-up or pull-down resistors, which stabilize the pin voltage when the pin is set as an input but left unconnected. Alternatively, unused pins can be configured as outputs set to a fixed logic level. The complete mapping and function of the lpGBT GPIO pins implemented in the final version of the BTL Central Controller (CC v2.3) are summarized in Table C.3.

	lpGBT0			**lpGBT1**		
Port	**Function**	**Dir.**	**Driv.**	**Function**	**Dir.**	**Driv.**
0	SCA_A Reset	Out	High	Not connected	-	-
1	Not connected	-	-	eLink Raphael pre-emphasis duration	In	-
2	CLK Raphael pre-emphasis configuration	Out	High	CLK Raphael pre-emphasis configuration	In	-
3	VDDA B Power Good	In	-	Calib Raphael pre-emphasis duration	In	-
4	1.2V Power Good	In	-	Not connected	-	-
5	VDDA A Power Good	In	-	CLK Raphael pre-emphasis duration	In	-
6	Calib Raphael pre-emphasis configuration	Out	High	VDDA B Power Good	In	-
7	Calib Raphael pre-emphasis duration	Out	High	Not connected	-	-
8	eLink Raphael pre-emphasis configuration	Out	High	eLink Raphael pre-emphasis configuration	In	-
9	eLink Raphael pre-emphasis duration	Out	High	Not connected	-	-
10	CLK Raphael pre-emphasis duration	Out	High	SCA_A Reset	Out	High
11	Not connected	-	-	Calib Raphael pre-emphasis configuration	In	-
12	Not connected	-	-	VDDA C Power Good	In	-
13	SCA_B Reset	Out	High	2.5V Power Good	In	-
14	Not connected	-	-	SCA_B Reset	Out	High
15	ALDO temp. shut-down disable	Out	High	ALDO temp. shut-down disable	Out	High

Table C.3: GPIO functions as of v2.3 of the concentrator card for BTL featuring two lpGBT ASICs (0 and 1). Dir. stands for the GPIO pin direction (input or output pin) while Driv. indicates whether the pin is driven low or high.

C.0.2 GBT-SCA

The software abstraction of the GBT-SCA ASIC closely follows that developed for the lpGBT. The primary objective in operating the GBT-SCA is to access and read the slow-control data from the front-end electronics. This functionality is achieved through 32

general-purpose digital I/O pins, 31 analog input ports multiplexed to a 12-bit ADC, and four analog output ports controlled by independent 8-bit DACs. As previously described, the GBT-SCA interfaces with the lpGBT via a dedicated ePort, operating at 40 MHz with an effective data rate of 80 Mb/s. The chip includes two identical eLink ports; in the current Central Controller (CC) v2.3 design, each GBT-SCA is connected to both lpGBTs to enable redundancy in case of optical link failures. In this configuration, only one eLink port is active at a time, while the second remains inactive.
The GBT-SCA provides 16 independent I^2C master channels with programmable data rates ranging from 100 kHz to 1 MHz and supporting both 7- and 10-bit addressing. These channels allow single- or multi-byte transactions to perform read, write, and read/modify/write operations. On the CC v2.3, GBT-SCA A uses a single I^2C master connected through the VTRX+ of lpGBT0, while GBT-SCA B employs two masters: one linked through the VTRX+ of lpGBT1 (master 0) and the other through the lpGBT slave I^2C (master 15). All I^2C communications have been successfully validated. Read and write operations to the GBT-SCA are performed via the EC primary link; redundant eLink communication has not yet been tested.
Communication with the GBT-SCA uses fixed-length packets of 120 bits, structured according to the SCA and eLink transfer protocols. Separate payload formats are used for read, write, and reset operations, differing only in the command field. Upon receiving any valid command—read, write, or reset—the GBT-SCA always replies with a payload. For control register operations, the response includes a command field of `0x00`; otherwise, an error flag is returned.
In software, each payload is constructed from four components:

- **Channel:** identifies the functional block being addressed;
- **Length:** fixed to 4 bytes for write operations and 1 byte for read operations;
- **Command:** a unique identifier in the range `0x00-0xFF`, determining the action to perform;
- **Data:** up to 32 bits of payload data for write operations (empty for read queries).

The header is built from the first three fields (channel, length, and command) and always occupies 32 bits, while the data section is organized in 8-bit chunks. Padding to the required packet size is handled internally by the C++ implementation and remains transparent to the user. Start-of-frame (SOF) and end-of-frame (EOF) delimiters are implemented in firmware on the EC interface.
To perform a register read, the software transmits a payload with an empty data field; the GBT-SCA responds with the requested data or an error code. In write mode, the payload includes the data to be written, and the GBT-SCA returns an acknowledgment payload or an error report. A command/error field value of `0x00` indicates successful execution. Otherwise, the returned 8-bit error field specifies the failure mode:

- bit 0 - generic error;
- bit 1 - invalid channel;
- bit 2 - invalid command;
- bit 3 - invalid transaction number;

	Description				Payload							
	Mode	Bit	Name	Function		TYPE	CH	CMD	D[31:24]	D[23:16]	D[15:8]	D[7:0]
ID	r	23-0	Chip ID	SCA unique ID number	R	TX	0x14	0xD1	-	-	-	1
						RX	0x14	0x00	-	ID	ID	ID
CR-B	r/w	0	-	-	R	TX	0	0x03	-	-	-	-
	r/w	1	ENSPI	SPI enable flag								
	r/w	2	ENGPIO	GPIO enable flag		RX	0	0x00	VAL	-	-	-
	r/w	3	ENI2C0	I2C-0 enable flag								
	r/w	4	ENI2C1	I2C-1 enable flag	W	TX	0	0x02	VAL	-	-	-
	r/w	5	ENI2C2	I2C-2 enable flag								
	r/w	6	ENI2C3	I2C-3 enable flag		RX	0	0x00	-	-	-	-
	r/w	7	ENI2C4	I2C-4 enable flag								
CR-C	r/w	0	ENI2C5	I2C-5 enable flag	R	TX	0	0x05	-	-	-	-
	r/w	1	ENI2C6	I2C-6 enable flag								
	r/w	2	ENI2C7	I2C-7 enable flag		RX	0	0x00	VAL	-	-	-
	r/w	3	ENI2C8	I2C-8 enable flag								
	r/w	4	ENI2C9	I2C-9 enable flag	W	TX	0	0x04	VAL	-	-	-
	r/w	5	ENI2CA	I2C-10 enable flag								
	r/w	6	ENI2CB	I2C-11 enable flag		RX	0	0x00	-	-	-	-
	r/w	7	ENI2CC	I2C-12 enable flag								
CR-D	r/w	0	ENI2CD	I2C-13 enable flag	R	TX	0	0x07	-	-	-	-
	r/w	1	ENI2CE	I2C-14 enable flag								
	r/w	2	ENI2CF	I2C-15 enable flag		RX	0	0x00	VAL	-	-	-
	r/w	3	ENJTAG	JTAG enable flag								
	r/w	4	ENADC	ADC enable flag	W	TX	0	0x06	VAL	-	-	-
	r/w	5	-	-								
	r/w	6	ENDAC	DAC enable flag		RX	0	0x00	-	-	-	-
	r/w	7	-	-								
SEU	r	31-0	SEU	SEU counter	R	TX	0x13	0xF1	-	-	-	-
						RX	0x13	0x00	-	-	-	VAL
					Rs	TX	0x13	0xF0	-	-	-	0
						RX	0x13	0x00	-	-	-	-

Table C.4: GBT-SCA internal control registers. Rs in the SEU field stands for Reset of the SEU counter. Type refers to a transmitted (TX) or a received (RX) payload. Where data fields are empty they should be assumed to be all 0 in the sent payload.

- bit 4 - invalid payload length;
- bit 5 - channel not enabled;
- bit 6 - channel busy;
- bit 7 - command under processing.

GBT-SCA Internal Registers

A dedicated logic block within the GBT-SCA is responsible for supervising the internal and network channels. This block can be accessed by sending a message with the channel field set to `0x0`, `0x13`, or `0x14`. It contains three 8-bit generic control registers, a 24-bit read-only chip identification (ID) register, and a 32-bit single-event upset (SEU) counter. The control registers, summarized in Table C.4 together with their corresponding payload values, control the enabling of the GBT-SCA channel interfaces, such as I2C, JTAG, ADC, DAC, and SPI.
As a best practice, unused registers should be disabled to minimize power consumption. A notable detail is that reading the SCA ID requires enabling the ADC interface, since the ID is stored in the internal e-fuse bank located within the ADC block (`0x14`). The e-fuses are programmed during the production and testing stages by the manufacturer, providing a unique identifier for each chip.

GBT-SCA Analog-to-Digital Converter (ADC)

The GBT-SCA includes a 12-bit analog-to-digital converter (ADC) featuring 31 analog input channels with a reference voltage of 1.0 V. The ADC performs an internal auto-calibration procedure for offset cancellation and applies an internal gain correction factor, provided by the manufacturer and stored in a dedicated register that is automatically

Command	Function	TYPE	CH	CMD/ERR	D[31:24]	D[23:16]	D[15:8]	D[7:0]
ENADC	Enable ADC	TX	0	0x06	0x4	-	-	-
		RX	0	0x0	-	-	-	-
ADC_W_CURR	100 μA current	TX	0x14	0x60	D[31:24]	D[32:16]	D[15:8]	D[7:0]
		RX	0x14	Flag	-	-	-	-
ADC_W_MUX	Select active line	TX	0x14	0x50	-	-	-	D[4:0]
		RX	0x14	Flag	-	-	-	-
ADC_R_DATA	Read corrected ADC conversion	TX	0x14	0x21	-	-	-	1
		RX	0x14	Flag	-	-	D[11:8]	D[7:0]

Table C.5: Sequence of commands issued to setup the GBT-SCA ADC and to read the latest gain and offset corrected conversion.

loaded at power-up. These analog channels are primarily used to monitor temperatures and currents on the board. Each input is connected to a `PT1000` resistor, allowing temperature measurements in the range of -50 to $+100$°C, similar to the implementation used for the lpGBT.
In the current setup, the ADC is used to read temperatures and also to monitor the SiPM bias voltages. The ADC readout sequence begins by enabling the converter through the corresponding control register (`Control Register D`). Next, the 100 μA bias current for the selected temperature-sensing channel is activated via an n-bit mask. The ADC is then configured to read from the desired analog port by writing to the active input-line register (port 31 corresponds to the internal temperature sensor). A write command initiates the conversion, and the resulting data are retrieved from the ADC data register payload.
The raw ADC counts are converted to voltages using a linear calibration relationship, $\mathrm{V} = \mathrm{ADC} \times a + b$, where the a and b are calibrations provided by the CCs QA/QC center. The corresponding current values used for conversion are also provided by the QA/QC center for both SCAs. The measured voltage is divided by the corresponding current to obtain the resistance, which is then converted to temperature by linearly interpolating from standard `PT1000` calibration tables, following the same approach adopted for the lpGBT.
The software framework also includes functionality for slope and offset calibration of the ADC voltage, although this feature has not yet been validated. Table C.5 summarizes the GBT-SCA command payloads used for temperature and voltage readout, while Table C.6 lists the analog input port connections of the GBT-SCA to the front-end components for the CC v2.3 board.

GBT-SCA General Purpose I/O (GPIO)

The GBT-SCA features 32 general-purpose digital input/output (GPIO) lines that, similar to those of the lpGBT, can be individually configured as inputs or outputs (tri-state mode). Each pin must be assigned a direction bit: “1” for output mode and “0” for input mode. The GPIO interface of the GBT-SCA has been validated in a basic configuration supporting read-only operations. On the Central Controller (CC) version 2.3, the GPIOs are used to enable the ALDO chips that supply the bias voltage to the TOFHIR ASICs, to configure the clock of the RAFAEL chip, and to activate the drain-to-drain voltage provided by the PCC components.
As the communication with the TOFHIR ASICs is still under development, the GPIO functionality of the GBT-SCA is expected to evolve in parallel with the ongoing firmware updates. In the current implementation, all pins are configured as inputs and only the communication layer is tested by reading back the status of each pin.
From the software standpoint, interaction with the GPIO involves first enabling the

	GBT-SCA A			GBT-SCA B		
ADC Port	**Function**	**Range (gain)**	**Enabled 100 μA**	**Function**	**Range**	**Enabled 100 μA**
0	FE2 Bias Curr 3	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	FE9 Bias Curr 1	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
1	FE2 Bias Curr 4	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	FE9 Bias Curr 3	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
2	FE6 Bias Curr 4	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	FE9 Bias Curr 2	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
3	FE3 Bias Curr 4	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	FE9 Bias Curr 4	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
4	FE2 Bias Curr 2	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	FE8 Bias Curr 4	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
5	FE6 Bias Curr 3	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	FE8 Bias Curr 3	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
6	FE3 Bias Curr 3	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	FE12 Bias Curr 3	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
7	FE2 Bias Curr 1	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	FE8 Bias Curr 2	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
8	FE6 Bias Curr 1	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	FE10 Bias Curr 4	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
9	FE3 Bias Curr 1	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	FE12 Bias Curr 4	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
10	FE4 Bias Curr 4	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	FE8 Bias Curr 1	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
11	FE6 Bias Curr 2	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	FE10 Bias Curr 3	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
12	FE5 Bias Curr 4	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	FE12 Bias Curr 1	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
13	FE1 Bias Curr 4	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	FE7 Bias Curr 4	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
14	FE1 Temp 2	-50,100 °C	Yes	FE10 Bias Curr 2	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
15	FE1 Temp 1	-50,100 °C	Yes	FE12 Bias Curr 2	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
16	FE4 Bias Curr 3	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	FE7 Bias Curr 3	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
17	FE1 Bias Curr 3	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	FE7 Bias Curr 2	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
18	FE3 Bias Curr 1	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	FE10 Bias Curr 1	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
19	FE1 Bias Curr 2	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	FE7 Temp 1	-50,100 °C	No
20	FE5 Bias Curr 2	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	FE11 Bias Curr 4	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
21	FE4 Bias Curr 1	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	V Bias B	0-61 V	No
22	FE1 Bias Curr 1	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	FE7 Temp 2	-50,100 °C	No
23	FE4 Bias Curr 2	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	FE11 Bias Curr 3	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
24	FE5 Temp 2	-50,100 °C	Yes	FE11 Bias Curr 2	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
25	FE2 Temp 2	-50,100 °C	Yes	FE7 Bias Curr 1	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
26	FE5 Temp 1	-50,100 °C	Yes	FE11 Bias Curr 1	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No
27	FE5 Bias Curr 3	0-2.247 mA (1/40) 0-44.94 mA (1/800)	No	FE11 Temp 2	-50,100 °C	Yes
28	FE2 Temp 1	-50,100 °C	Yes	FE8 Temp 2	-50,100 °C	Yes
29	FE5 Bias Curr 1	0-1.02 mA (1/40) 0-20.04 mA (1/800)	No	FE8 Temp 1	-50,100 °C	Yes
30	V Bias A	0-61 V	No	FE11 Temp 1	-50,100 °C	Yes

Table C.6: Active GBT-SCA analog connections from CC v2.3 board featuring two SCA ASICs (A and B)

relevant bits in the control register, then setting the pin direction according to the CC v2.3 hardware specifications. Once configured, the software can either write to output pins or read from input/output pins as required. Table C.7 summarizes the sequence of software operations needed for the GBT-SCA GPIO configuration, while Table C.8 lists the GPIO hardware connections implemented on the CC v2.3 and their corresponding functions.

Command	Function	TYPE	CH	CMD/ERR	D[31:24]	D[23:16]	D[15:8]	D[7:0]
ENGPIO	Enable GPIOs	TX	0	0x02	0x2	-	-	-
		RX	0	0x0	-	-	-	-
GPIO_W_CLKSEL	Enable 40 MHz clock	TX	0x02	0x80	0[31:24]	0[23:16]	0[15:8]	0[7:0]
		RX	0	0x0	-	-	-	-
GPIO_W_DIRECTION	Set the GPIO direction	TX	0x02	0x20	D[31:24]	D[32:16]	D[15:8]	D[7:0]
		RX	0x14	Flag	-	-	-	-
GPIO_R_DATAOUT	Read output pins	TX	0x02	0x11	-	-	-	-
		RX	0x02	Flag	D[31:24]	D[32:16]	D[15:8]	D[7:0]
GPIO_R_DATAIN	Read input pins	TX	0x02	0x01	-	-	-	-
		RX	0x02	Flag	D[31:24]	D[32:16]	D[15:8]	D[7:0]

Table C.7: Sequence of commands issued to setup the GBT-SCA GPIOs and to read the input and output pins.

	GBT-SCA A		GBT-SCA B	
Port	**Function**	**Dir.**	**Function**	**Dir.**
0	FE2 ALDO ENABLE 2	In	FE9 ALDO Enable 2	-
1	FE5 ALDO Enable 2	-	FE9 ALDO Enable 1	In
2	FE8 ALDO Enable 2	In	PCC B VDDA C Enable	In
3	FE8 ALDO Enable 1	In	PCC B VDDA D Enable	In
4	FE11 ALDO Enable 2	In	Select clock RAFAEL 0	-
5	FE1 ALDO Enable 2	In	FE10 ALDO Enable 1	In
6	FE1 ALDO Enable 1	In	FE10 ALDO Enable 2	In
7	FE11 ALDO Enable 1	In	Not connected	-
8	FE2 ALDO Enable 1	In	FE7 ALDO Enable 1	In
9	FE4 ALDO Enable 1	In	FE7 ALDO Enable 2	-
10	FE12 ALDO Enable 1	In	RAFAEL MUX Calib	In
11	PCC A VDDA B ENABLE	-	FE3 ALDO Enable 1	In
12	PCC A VDDA A ENABLE	-	FE5 ALDO Enable 1	In
13	FE12 ALDO Enable 2	In	Enable clock RAFAEL 1	In
14	FE3 ALDO Enable 2	-	FE6 ALDO Enable 1	In
15	FE4 ALDO Enable 2	In	FE6 ALDO Enable 2	In
16	Not connected 2	In	FE12 ALDO Enable 2	-
17	FE6 ALDO Enable 1	-	FE3 ALDO Enable 2	In
18	FE6 ALDO Enable 2	In	FE4 ALDO Enable 2	In
19	RAFAEL MUX Calib	In	FE12 ALDO Enable 1	In
20	FE3 ALDO Enable 1	In	PCC A VDDA B ENABLE	-
21	FE5 ALDO Enable 1	In	PCC A VDDA A ENABLE	In
22	Enable clock RAFAEL 1	In	FE11 ALDO Enable 1	In
23	FE7 ALDO Enable 1	In	FE2 ALDO Enable 1	-
24	FE7 ALDO Enable 2	In	FE4 ALDO Enable 1	In
25	Select clock RAFAEL	In	FE11 ALDO Enable 2	-
26	FE9 ALDO Enable 2	In	FE2 ALDO Enable 2	In
27	FE9 ALDO Enable 1	-	FE1 ALDO Enable 1	In
28	PCC B VDDA D Enable	-	FE5 ALDO Enable 2	In
29	FE10 ALDO Enable 1	In	FE1 ALDO Enable 2	In
30	FE10 ALDO Enable 2	-	FE8 ALDO Enable 1	In
31	PCC B VDDA C Enable	In	FE8 ALDO Enable 2	In

Table C.8: GPIO functions as of v2.3 of the concentrator card for BTL featuring two GBT-SCA ASICs (A and B). All pins are configured as input pins for simplicity as no front end board is available. This will change in future developments of the BTL software.