Project description

Contents

1 Science

Since the start of operations in 2008, the Large Hadron Collider (LHC) at CERN has provided a data set amounting to about 10^{16} proton-proton collisions that are expected to double, once more, until 2024. During the LHC Run 2 (2015–2018), the CMS experiment [1] collected $137\,\text{fb}^{-1}$ of proton-proton collision data at $\sqrt{s} = 13\,\text{TeV}$ and is now performing extensive analyses to extract the maximum knowledge on fundamental particle physics processes at the TeV scale. The most prominent key milestone in the LHC experiments' history is the discovery of the Higgs boson (H), jointly announced by the ATLAS and CMS Collaborations in 2012 [2, 3]. It completed the particle content of the standard model (SM) of particle physics. Since then, precision measurements of the Higgs boson's mass [4, 5], its charge-parity (CP) nature [6, 7], its couplings to electroweak gauge bosons ($\text{W}^{\pm}$, Z, γ), and other couplings have been at the center of the LHC research program with many already entering the precision regime. At the same time, the lack of smoking-gun evidence for resonant phenomena beyond the SM (BSM) has restricted many sophisticated models, like supersymmetry or models with extra dimensions, that were conceived to cure the SM's shortcomings: the absence of a suitable dark matter candidate, the unknown origin of nonzero neutrino masses, the baryon asymmetry of the universe, and the lack of a theoretical explanation of the fine tuning required for the insensitivity of the Higgs boson mass to quantum corrections.

Is the case, therefore, closed? In this proposal, I argue on the contrary. In fact, the recent application of effective field theories (EFT) [8–10] demonstrates impressively that subtle deviations, hiding in the observables' distributions, can probe BSM phenomena at energy scales often exceeding the LHC's reach in the direct searches. The underlying theoretical framework considered in this proposal is the SM effective field theory (SM-EFT) [11–15] that systematically extends the SM with operators of higher mass dimension that smoothly change the kinematic spectra. Specifically, SM-EFT modifications to the SM Higgs sector result in deviations that grow with the momenta of the final state particles. The joint production of a H boson and a vector boson (VH with $\text{V} = \text{W}^{\pm}, \text{Z}$) are particularly strongly affected, turning the highly energetic VH final states into extremely important discovery tools for BSM phenomena. In addition to the changes in one-dimensional kinematic distributions, BSM sources of CP violation change subtle triple-correlations of angular observables that are easily lost if not targeted explicitly by means of "interference resurrection" [16]. The latter techniques have been demonstrated in the simpler $\text{W}\gamma$ final states [17], but the order-of-magnitude sensitivity gain for precision measurements, although proposed recently [18], has not yet been exploited experimentally in the Higgs sector.

To close these gaps, I propose an extensive measurement of the H–V couplings in VH final states. Resurrecting the process' interference pattern and simultaneously leveraging the signature of energy growth in the kinematic tails of the H, $\text{W}^{\pm}$, and Z bosons promise unique sensitivity and a wide reach to SM-EFT effects, including the CP structure of the coupling. The proposed analysis includes differential cross section measurements of a number of observables never considered before. Combining H boson decays to a pair of b quarks with the requirement of leptonic V decays (electrons or muons) leads to a high signal acceptance while suppressing the complex hadronic backgrounds.

Sensitivity to the SM-EFT effects mentioned previously can already be obtained using the existing data set collected during the LHC Run 2. With the additional $160\,\text{fb}^{-1}$ of data to be delivered by the LHC during Run 3 (2022–2024), these measurements are expected to have a significant leap in precision. In the remainder of the proposal, I attempt to demonstrate these claims.

2 Research hypothesis

I assume that there exist no unknown resonances within the kinematic reach of LHC and the TeV-scale phenomena are described by the standard model effective theory (SM-EFT). The SM-EFT progressively includes operators of mass dimension greater than four that respect the SM symmetries [11–15] and is defined by the Lagrangian

$$\mathcal{L}_{\text{SM-EFT}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{c_i^{(5)}}{\Lambda} \mathcal{O}_{5,i} + \sum_i \frac{c_i^{(6)}}{\Lambda^2} \mathcal{O}_{6,i} + \ldots . \tag{2.1}$$

It captures all non-resonant BSM phenomena below an arbitrarily chosen energy scale Λ. The dimensionless Wilson coefficients $c^{(n)}$ are used to parameterize the effects on observables, while the terms $\mathcal{O}_5$ and $\mathcal{O}_6$ are the operators at mass dimensions 5 and 6, respectively. The only possible dimension-5 candidate is the Weinberg operator [19], not relevant for the phenomenology of this proposal [20]. The number of baryon and lepton number conserving dimension-6 operators is 2499 when all flavor degrees of freedom are counted. There are 76 operators respecting a $U(3)^5$ flavor symmetry [12] among which the 12 listed in Table 1 affect the VH process. The operators are written in the so-called Warsaw basis [21], and it is the central aim of this proposal to constrain their Wilson coefficients via their effects on the Higgs boson production properties and decay kinematics [22–25].

Table 1. The dimension-6 operators in the Warsaw basis [21] affecting VH processes at leading order.

$\mathcal{O}_{Hq}^{(1)}$	$iH^\dagger \overleftrightarrow{D}_\mu H \bar{q}\gamma^\mu q$	$\mathcal{O}_{HWB}$	$H^\dagger \sigma^a H W_{\mu\nu}^a B^{\mu\nu}$
$\mathcal{O}_{Hq}^{(3)}$	$iH^\dagger \sigma^a \overleftrightarrow{D}_\mu H \bar{q}\sigma^a\gamma^\mu q$	$\mathcal{O}_{H\tilde{W}B}$	$H^\dagger \sigma^a H W_{\mu\nu}^a \tilde{B}^{\mu\nu}$
$\mathcal{O}_{Hu}$	$iH^\dagger \overleftrightarrow{D}_\mu H \bar{u}_R\gamma^\mu u_R$	$\mathcal{O}_{HW}$	$(H^\dagger H) W_{\mu\nu} W^{\mu\nu}$
$\mathcal{O}_{Hd}$	$iH^\dagger \overleftrightarrow{D}_\mu H \bar{d}_R\gamma^\mu d_R$	$\mathcal{O}_{H\tilde{W}}$	$(H^\dagger H) W_{\mu\nu}^a \tilde{W}^{a\mu\nu}$
$\mathcal{O}_{HD}$	$(H^\dagger D_\mu H)^*(H^\dagger D_\mu H)$	$\mathcal{O}_{HB}$	$(H^\dagger H) B_{\mu\nu} B^{\mu\nu}$
$\mathcal{O}_{H\square}$	$(H^\dagger H)\square(H^\dagger H)$	$\mathcal{O}_{H\tilde{B}}$	$(H^\dagger H) B_{\mu\nu} \tilde{B}^{\mu\nu}$

The operators $\mathcal{O}_{Hq}^{(1)}$, $\mathcal{O}_{Hq}^{(3)}$, $\mathcal{O}_{Hu}$, and $\mathcal{O}_{Hd}$ introduce the 4-point interactions depicted in Fig. 1 (left). Together with $\mathcal{O}_{HD}$ and $\mathcal{O}_{HWB}$, these six operators describe all relevant V-fermion coupling modifications (Fig. 1, middle). The remaining operators modify the H–V coupling, shown in Fig. 1 (right). There is also the Yukawa-type operator $H^\dagger H \bar{Q} H b$, which only changes the H $\to$ b$\bar{\text{b}}$ branching ratio and, thus, can be probed in an inclusive cross section measurement.

The SU(2) doublet H is expanded as $\begin{pmatrix} 0 \\ v + h \end{pmatrix}$, where v is the vacuum expectation value and h corresponds to the physical Higgs boson. At the weak scale, H is replaced by v in operators shown in Table 1, and some of them can be constrained using the existing collider results from the LEP and the Tevatron experiments. Nevertheless, the bounds from the LHC are expected to be more stringent [26, 27]. Moreover, the operators of the form $H^2\mathcal{L}_{\text{SM}}$, can be directly probed for the first time at the LHC [28].
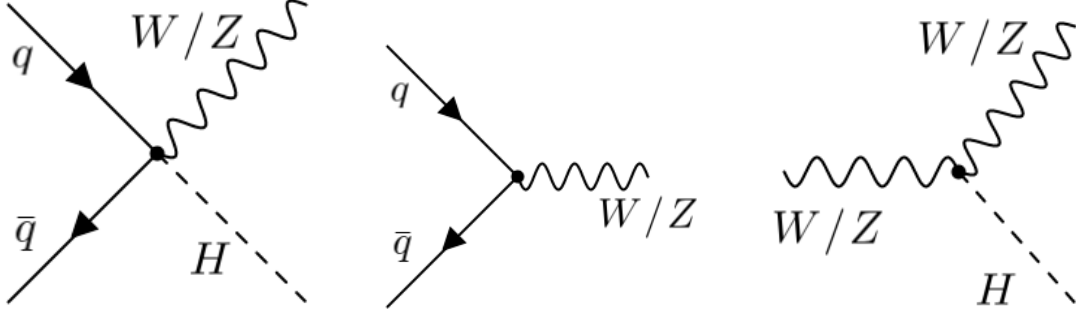
Figure 1. Feynman diagrams for VH production sensitive to different dimension-6 operators.

The energy dependence of the VH production cross section is critical for the sensitivity of the proposed analyses. The Feynman diagram with four-point interaction (Fig. 1, left) involves one scalar and one vector field and two fermion fields. Power counting with naive dimensional analysis [29] implies a scaling of the cross section with E^2. The diagram with the modification of the fermion–V coupling (Fig. 1, middle) corresponds to a Lagrangian term of the form $\frac{v^2}{\Lambda^2} V_\mu \bar{\psi} \gamma^\mu \psi$, which does not induce a cross section enhancement with energy. The operators modifying H–V coupling (Fig. 1, right) give rise to terms $vh \partial^\mu \partial^\nu V_\mu V_\nu$. When interfering with the SM term $(H^\dagger H) V^\mu V_\mu$, this gives rise to a linear increase of cross section with energy. The proposed measurement will probe the operators in Table 1 using kinematic variables exhibiting energy growth and, for the first time, the angular structure that is sensitive to subtle interference effects. At high energy, the H boson has large $p_{\rm T}$, and its decay products are collimated in a single large-radius jet, thus providing a testbed for the application of the cutting-edge jet substructure techniques [30, 31]. For the angular information, we exploit the techniques of Ref. [18] and define the angles and decay planes of the VH process as shown in Fig. 2.

The squared amplitude of V $(\to \ell\ell)$H $(\to b\bar{b})$ production, when summing over the lepton helicities, has the form

$$|\mathcal{M}(\hat{s}, \Theta, \theta, \varphi)|^2 = \sum_i a_i(\hat{s}) f_i(\Theta, \theta, \varphi), \tag{2.2}$$

where $a_i(\hat{s})$ are functions of Wilson coefficients and the energy transfer $\hat{s}$ involved in the process. The functions f_i depend on the three angles in Fig. 2

$$
\begin{aligned}
&f_1 = f_{LL} = \sin^2\Theta \sin^2\theta, &&f_6 = \tilde{f}_{LT}^1 = \sin\varphi \sin\Theta \sin\theta, \\
&f_2 = f_{TT}^1 = \cos\Theta \cos\theta, &&f_7 = \tilde{f}_{LT}^2 = \sin\varphi \sin\Theta \sin\theta \cos\Theta \cos\theta, \\
&f_3 = f_{TT}^2 = (1 + \cos^2\Theta)(1 + \cos^2\theta), &&f_8 = f_{TT'} = \cos^2\varphi \sin^2\Theta \sin^2\theta, \\
&f_4 = f_{LT}^1 = \cos\varphi \sin\Theta \sin\theta, &&f_9 = \tilde{f}_{TT'} = \sin^2\varphi \sin^2\Theta \sin^2\theta, \\
&f_5 = f_{LT}^2 = \cos\varphi \sin\Theta \sin\theta \cos\Theta \cos\theta
\end{aligned}
\tag{2.3}
$$

where indices refer to the polarizations of the intermediate vector boson.

A traditional inclusive measurement implicitly integrates over Θ, θ, φ, thus removing all the terms in Eq. (2.3), except f_{LL} and f_{TT}^2. This is the cause of loss of important information, which can be recovered for f_{LT}^2, $\tilde{f}_{LT}^2$, $f_{TT'}$, and $\tilde{f}_{TT'}$ by using a triple differential analysis with respect to
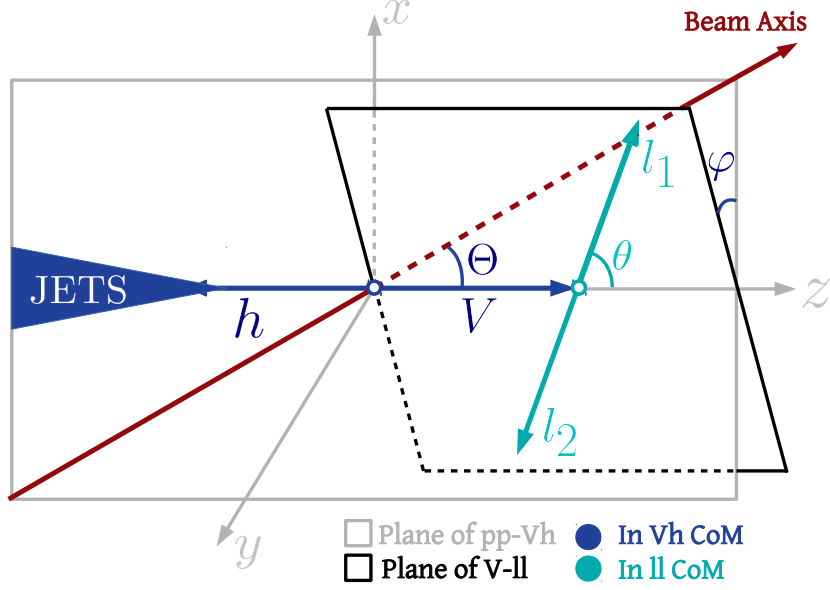
Figure 2. Decay planes and angles in the V $(\rightarrow \ell\ell)$H $(\rightarrow b\bar{b})$ production. Note that Θ is defined in the VH rest frame, while θ is defined in the V rest frame.

all three angles. Because the incoming quark and anti-quark direction is not known at the LHC, the terms proportional to f_{TT}^1, f_{LT}^1, and $\tilde{f}_{LT}^1$ can not be extracted.

3 Objectives

The goal of the proposal is to probe the SM-EFT operators in Table 1 as precisely as possible using events from the VH processes in the data collected by the CMS experiment during the LHC Run 2 and 3. Two experimental objectives separately tackle the energy dependence of the differential cross section and the CP nature of H–V coupling due to SM-EFT effects. At the final stage of the work, those will be combined to achieve optimal sensitivity.

The first objective is the exploitation of energy growth induced by SM-EFT operators to boost the sensitivity of differential cross section measurements in VH production with leptonic V decays. Latest jet substructure techniques for boosted H bosons will be applied for this purpose. The three operators $\mathcal{O}_{Hq}^{(3)}$, $\mathcal{O}_{HW}$, $\mathcal{O}_{HWB}$, their CP-odd counterparts, and the operators $\mathcal{O}_{HD}$, $\mathcal{O}_{H\square}$ affect both WH and ZH productions. The dependence of the production cross section on the first three Wilson coefficients for different selections in vector boson transverse momentum, $p_{\mathrm{T}}(V)$ is shown in Fig. 3. The energy growth is visible as an increased dependence on the Wilson coefficient for selections with higher $p_{\mathrm{T}}(V)$. The operators $\mathcal{O}_{Hq}^{(1)}$, $\mathcal{O}_{Hu}$, $\mathcal{O}_{Hd}$, $\mathcal{O}_{HB}$, and $\mathcal{O}_{H\tilde{B}}$ affect only the ZH production, because both left- and right-handed initial-state quarks of same flavor take part in ZH production, while only the left-handed quarks of different flavors are involved in WH production [32, 33]. Therefore, simultaneous measurements of the couplings in both the WH and the ZH processes are necessary to confidently probe the complete set of relevant operators [18]. The SM cross section for WH production with H $\rightarrow$ b$\bar{b}$ and W $\rightarrow \ell\nu$ is 0.28 pb, whereas it is only 0.06 pb for ZH production, where Z $\rightarrow \ell\ell$. Thus, the WH process provides better statistical precision in the
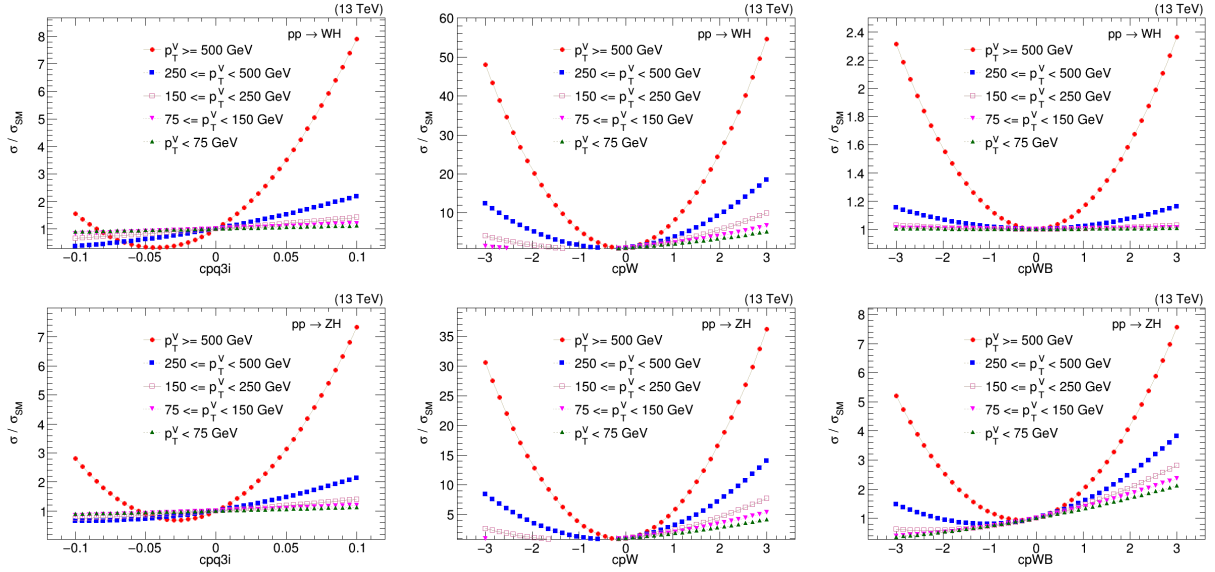
Figure 3. The ratio of the cross section of the WH (top) and ZH (bottom) processes relative to the SM prediction, $\sigma/\sigma_{\mathrm{SM}}$, as a function of Wilson coefficients corresponding to operators $\mathcal{O}_{Hq}^{(3)}$ (left), $\mathcal{O}_{HW}$ (middle), and $\mathcal{O}_{HWB}$ (right), respectively, in five regions of W and Z boson p_{T}.

measurements. The Z final state is less populated, but contributions from background processes are smaller, particularly at high energy, providing an independent sample to measure the sensitivity to SM-EFT operators and two processes together can lead to evidence of the BSM signal if any of the Wilson coefficients are found to be significantly different from zero.

The second objective is the use of angular observables to further boost the sensitivity and tackle the CP-structure of the H–V interaction. As an example, distribution of one of the angular variables, φ, is shown in Fig. 4. Here, the leading contribution for the CP-even and -odd operators come from the terms corresponding to f_{LT}^2 and $\tilde{f}_{LT}^2$ in Eq. (2.3), respectively, which have $\cos\varphi$ and $\sin\varphi$ dependence. In the simpler Wγ final state, the technique increased the sensitivity to interference between the SM and the SM-EFT operator of interest by about a factor of 10 [17]. Section 6 presents a preliminary sensitivity study for the three most important operators.

4 Scientific originality

In the V $\rightarrow \ell\ell$ and H $\rightarrow$ b$\bar{\text{b}}$ final state, I extend traditional search strategies, based on the simplified template cross section binning scheme [34], by angular observables sensitive to subtle SM-EFT effects and the CP-structure of the H–V interaction. For the first time,

- the complete multi-dimensional event information is used in VH production to directly probe dimension-6 SM-EFT operators inducing energy growth. Besides kinematic event properties, novel angular variables (Fig. 6) crucially refine the analysis strategy. They disentangle the various SM-EFT effects and can largely boost the sensitivity [17]. Furthermore,

- the angular information, not explored previously in VH production, is used to probe the CP structure of SM-EFT modifications of the H–V coupling. We can, therefore, measure new sources of CP violation beyond the SM.
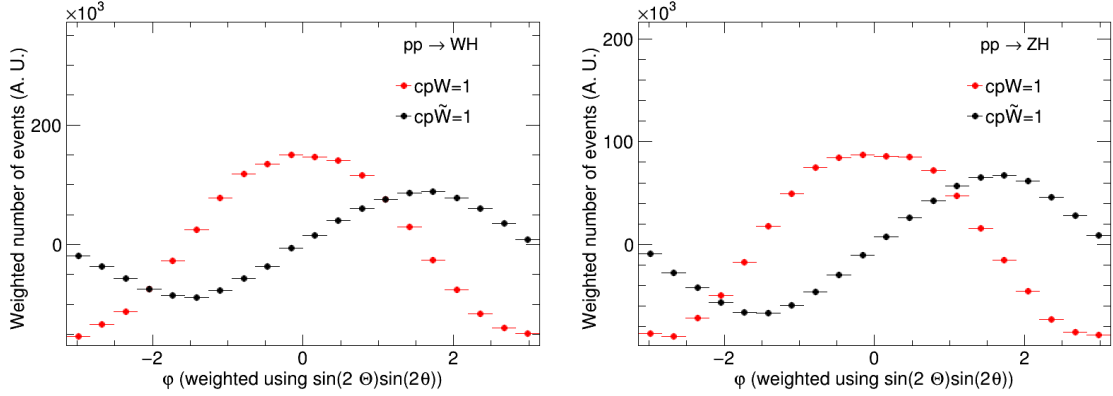
6

Figure 4. Distribution of angle φ at the particle level for Wilson coefficients corresponding to CP-even operator $\mathcal{O}_{HW}$ (red points) and CP-odd operator $\mathcal{O}_{H\tilde{W}}$ (black points) set to 1 in case of WH (left) and ZH (right) productions. For each entry in the distribution, a weight of $+1$ or -1 is applied depending on the sign of $\sin(2\Theta)\sin(2\theta)$.

A recent differential cross section measurement with an EFT interpretation of the Run 2 data set by the ATLAS collaboration [35] includes measurement regions with boosted H candidates but ignores the angular information. The most recent EFT interpretation of VH measurement is restricted to inclusive measurements [36]. The scope of this proposal significantly extends beyond both cases.

Smaller novel items of this proposal pertain to the multivariate analysis (MVA) strategy employed to achieve these goals. For example, the proposer's group has recently developed a new technique to train tree-based classifiers to optimally separate effects of SM-EFT operators from the SM background [37]. The combined data sets from the LHC Run 2 and the future LHC Run 3 provide an opportunity to measure SM-EFT operators involving the Higgs field [28, 38] with unprecedented precision.

5 Relevance to the research field

On a broad scale, the measurements provide a new piece of the TeV-scale SM-EFT puzzle. The findings will be interpreted in the context of SM-EFT, which guarantees maximal independence from model assumptions and a smooth interface for theory interpretations.

The absence of hints for resonant BSM phenomena has shifted the focus from ultraviolet-complete theories to EFTs, with first attempts at complete TeV-scale characterizations emerging relatively recently [39–41]. At the experimental collaborations, the measurements of SM-EFT operator coefficients have just arrived at the center stage of the LHC physics program [42–44]. The interest from the theoretical and experimental communities is still growing, reflected in various new working groups (e.g., the LHC EFT working group [45]) and lecture series (e.g., "All things EFT" [46]).

In particular, the proposed research contributes a unique probe of SM-EFT operator coefficients in the Higgs sector. Besides the sensitivity boost from novel experimental techniques, it also complements the existing results by extending to CP-sensitive angular observables, tightly constraining the H–V couplings. A deviation from the SM hinting at new sources of CP-violation will

have wide-ranging significance in the explanation of the origin of the matter-antimatter asymmetry of the universe [47–49].

The relevance of the proposed research extends to adjacent fields when the interplay of LHC measurements in global SM-EFT analysis is considered [39, 41]. For example, specific assumptions of the BSM flavor structure provide a tight connection of the proposed measurement with the sector of top-quark physics. Moreover, the SM electroweak symmetry breaking relates the 4-point interactions in Table 1 to complementary measurements of the fermion-gauge boson couplings. An example of this link can be obtained from Fig. 1 (left) when replacing the H boson line with v.

The results obtained from the proposed measurements can also be reinterpreted as constraints on ultraviolet-complete new physics models describing the underlying dynamics at high energy. Examples of models the proposed analysis will be capable of probing are the following:

- Singlet scalar models, where a scalar field, singlet under the SM gauge group, is introduced [50, 51]. This class of models gives rise to operators of the form of $\mathcal{O}_{H\square}$, $\mathcal{O}_{HD}$, $\mathcal{O}_{HW}$, $\mathcal{O}_{HB}$, $\mathcal{O}_{HWB}$, $\mathcal{O}_{Hu}$, and $\mathcal{O}_{Hd}$.

- Models predicting the existence of a new vector boson [52], for example, Z$'$ and W$'$. The models with extra spatial dimensions [53], and little Higgs models [54] fall in this category. If the new bosons coupling to both fermions and gauge bosons have large masses ($>$ few TeV), they give rise to the contact interaction shown in Fig. 1 (left) and thus contributes to the operators $\mathcal{O}_{Hq}^{(1)}$, $\mathcal{O}_{Hq}^{(3)}$, $\mathcal{O}_{Hu}$, and $\mathcal{O}_{Hd}$.

- Colored extensions of the SM with vector-like quarks (VLQs), triplet under SU(3) and singlet under SU(2)$_L$ [55, 56]. When VLQs are allowed to couple to light quarks, they can give rise to $\mathcal{O}_{Hq}^{(1)}$ and $\mathcal{O}_{Hq}^{(3)}$ operators. New physics models with Z$'$ or VLQs are in highlights in recent times to explain the anomalies in the flavor sector.

The phenomenological implications of the H–V coupling measurement thus radiate into neighboring research areas, while the proposed experimental program is self-contained. In the following section, I aim to establish that.

6   Method

We target to use the combined data set collected by the CMS experiment during the LHC Run 2, corresponding to an integrated luminosity of $139\,\mathrm{fb}^{-1}$ at $\sqrt{s} = 13\,\mathrm{TeV}$, and the LHC Run 3, currently estimated at $160\,\mathrm{fb}^{-1}$. In the following, we report a feasibility study of the proposed measurements and exemplify the key methodology. Important sources of systematic uncertainty modeling the differences of the simulation to data are considered. We focus on the WH final states with challenging background rates. The results are also valid for the ZH process, which provides independent measurement channels with orthogonal SM-EFT sensitivity, higher signal purity, but lower signal yield.

6.1   Final states and observables

In the SM, the H $\rightarrow$ b$\overline{\mathrm{b}}$ decay channel has the largest branching ratio, motivating to choose it for improving the statistical precision of the measurement. Leptonic decay modes of the vector bosons help in reducing the contribution from multijet production in quantum chromodynamics (QCD).

Therefore, we consider the WH and ZH production processes with the $W \to \ell\nu$, $Z \to \ell^+\ell^-$, and $H \to b\bar{b}$ decay channels, where $\ell = (e, \mu)$.

As shown in Sec. 3 and Fig. 3, the 4-point interactions provided by the operators in Table 1 induce energy growth, modifying the high-energy tail of distributions of kinematic observables such as the transverse momenta, $p_T(W)$, $p_T(Z)$, or $p_T(H)$. However, these kinematic observables alone are not sufficient to disentangle the various effects from a large number of possible SM-EFT operators. A systematic analysis of helicity amplitudes in VH production followed by vector boson decay is provided in Ref. [18]. Amplitudes involving the longitudinal and transverse vector boson polarizations contribute differently to VH production in SM and SM-EFT. The interference among helicity amplitudes, including CP effects, is preserved at the event level in subtle angular triple-correlations. For example, the cosinusoidal and sinusoidal modulations in φ distributions shown in Fig. 4 are obtained after using the sign of $\sin(2\Theta)\sin(2\theta)$ as seen from $\tilde{f}_{LT}^2$ and f_{LT}^2 terms, respectively, in Eq. 2.3. This triple angular correlation allows extracting rich information about the event structure from a joint analysis of kinematic and angular variables. It is important to note that traditional strategies can not disentangle the subtle angular triple-correlation: integrating over any of the angles in the final observables removes the interference effects [16].

While SM-EFT effects in background processes, e.g. from diboson production, are ignored for this simple feasibility study, they will be considered in the final measurement.

6.2 Event preselection and categorization

All results are obtained from events generated with `Madgraph5_aMC@NLO` [57] and simulated up to the CMS detector level. The SM-EFT effects in the VH signal are incorporated with `MadWeight` [58] and using the `SMEFTsim` model [59].

The CMS trigger system selects the VH events based on the presence of leptons (electrons or muons) originating from the W or Z bosons. After the trigger decision, we select WH candidate events by requiring an electron or a muon with $p_T > 32$ and $> 25\,\text{GeV}$, respectively. As will be discussed in Sec. 6.3, the four-momentum of the neutrino from the W decay is reconstructed using the missing transverse momentum vector ($\vec{p}_T^{\,\text{miss}}$), assuming that the invariant mass of the neutrino-lepton system equals the W boson mass m_W.

A sample enriched in boosted W bosons is selected by the requirement $p_T(W) > 150\,\text{GeV}$. After this selection, the level of QCD multijet production is negligible. Requiring a minimum difference between the azimuthal lepton- and p_T^{miss} angles, $\Delta\phi(\vec{\ell}, \vec{p}_T^{\,\text{miss}}) > 2$, removes spurious events from detector noise and reconstruction failure while retaining good signal efficiency. Finally, events must not contain additional leptons with $p_T > 25\,\text{GeV}$.

The selected events are further categorized into measurement regions depending on quantities from the jet system. If the H boson has large p_T, its decay products are merged into a jet with a large cone size. In this "boosted topology", the H boson is reconstructed and identified in the set of jets provided by the anti-k_T algorithm with a distance parameter of $R = 0.8$ (AK8 jets). Otherwise, the H boson is in the "resolved topology", where it can be reconstructed using the standard anti-k_T jets with distance parameter of $R = 0.4$ (AK4 jets) originating from the b quark pair. Deep neural network (DNN) based algorithms DeepJet [60] and DeepAK8 [31] are used to identify AK4 jets initated by b quarks and the AK8 jet from $H \to b\bar{b}$ decay, respectively. The DeepAK8 tagger also provides probability-normalized classifier values for the top quark, W boson, light quark or gluon hypothesis for each AK8 jet. We use this information later for signal-to-background discrimination in the boosted topology.

Table 2. Selection conditions targeted for WH events.

Resolved category	Boosted category
At least 2 b-tagged AK4 jets (satisfying $p_\mathrm{T} > 30\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.5$)	At least 1 AK8 jets (satisfying $p_\mathrm{T} > 250\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.5$)
The b tagging score of $b_1 > \mathrm{btag}_\mathrm{max}^\mathrm{cut}$	The H tagging score of H $> \mathrm{Htag}^\mathrm{cut}$
The b tagging score of $b_2 > \mathrm{btag}_\mathrm{min}^\mathrm{cut}$	
Up to 1 additional AK4 jets	Veto on extra b-tagged AK4 jets
$M(b_1 + b_2) \in [90, 150]$ GeV	Soft-drop mass of J $\in [90, 150]$ GeV

In the resolved topology, the two AK4 jets with highest b tagging scores, referred to as $b_{1,2}$, form the H candidate, while the AK8 jet (J) with the highest DeepAK8 score for H tagging provides the H candidate in the boosted topology. For AK8 jets, soft drop grooming [61] is used to calculate the jet mass by removing the uncorrelated radiation clustered into the jet. In boosted category, a condition on the number of b-tagged jets outside the H candidate is applied, which reduces its overlap with the resolved category. The inclusive signal region (SR), enriched by $W(\to \ell\nu)\,H(\to b\overline{b})$ events, is constructed using the conditions listed in Table 2. In the following, the threshold values $\mathrm{btag}_\mathrm{max}^\mathrm{cut}$ and $\mathrm{btag}_\mathrm{min}^\mathrm{cut}$ are chosen to ensure a DeepJet b tagging mistag rate of 1% and 10% for light quark or gluon jets, respectively. The threshold value $\mathrm{Htag}^\mathrm{cut}$ corresponds to a H mistag rate of 1%. These numerical values are preliminary and will be optimized in the context of the final analysis.

By inverting specific requirements in Table 2, control regions (CRs) can be defined. These enrich the various background processes such as top quark pair production ($t\overline{t}$), vector boson production in association with b- and c-quark jets ($W + b/c$), and with light quark or gluon jets (W+udsg). These CRs will be used to measure and validate the corresponding backgrounds in data.

The distribution of reconstructed $p_\mathrm{T}(W)$ in the WH signal and all relevant background processes in the resolved and boosted SR are shown in Fig. 5, where alternate WH signal hypothesis for the operators $\mathcal{O}_{Hq}^{(3)}$ and $\mathcal{O}_{HW}$ are overlaid. The figure shows the importance of the $t\overline{t}$, $W + b/c$, and $W + $ udsg backgrounds in the WH SRs. Fig. 5 also shows that a large gain in sensitivity to the SM-EFT operators is expected from the high-p_T region, where the signal-to-background ratio improves due to their characteristic energy growth.

Next, we turn to the sensitivity of the angular observables after using the selection conditions summarized in Table 2. Distributions of two angular variables Θ and φ in the WH signal after applying the preselection conditions listed in Table 2 are shown in Fig. 6 for different SM-EFT hypotheses. For completeness, similar distributions in the case of the ZH signal sample are also shown. Fig. 6 shows that the SM production, dominated by the longitudinal polarization of W bosons, exhibits differences in Θ for nonzero $\mathcal{O}_{HW}$ coefficients because this operator enhances the helicity amplitude of WH production with transversely polarized W bosons. The observable φ, on the other hand, is very sensitive to the CP-nature of the H–V coupling, and the difference in the modulation of CP-even and CP-odd operator coefficients is clearly visible. The CP-sensitivity in the φ observable is slightly affected by the ambiguities introduced by the unknown z-component of the neutrino momentum. This necessitates the development of an algorithm for the neutrino reconstruction, possibly similar to the MVA-based techniques designed in top quark analyses for
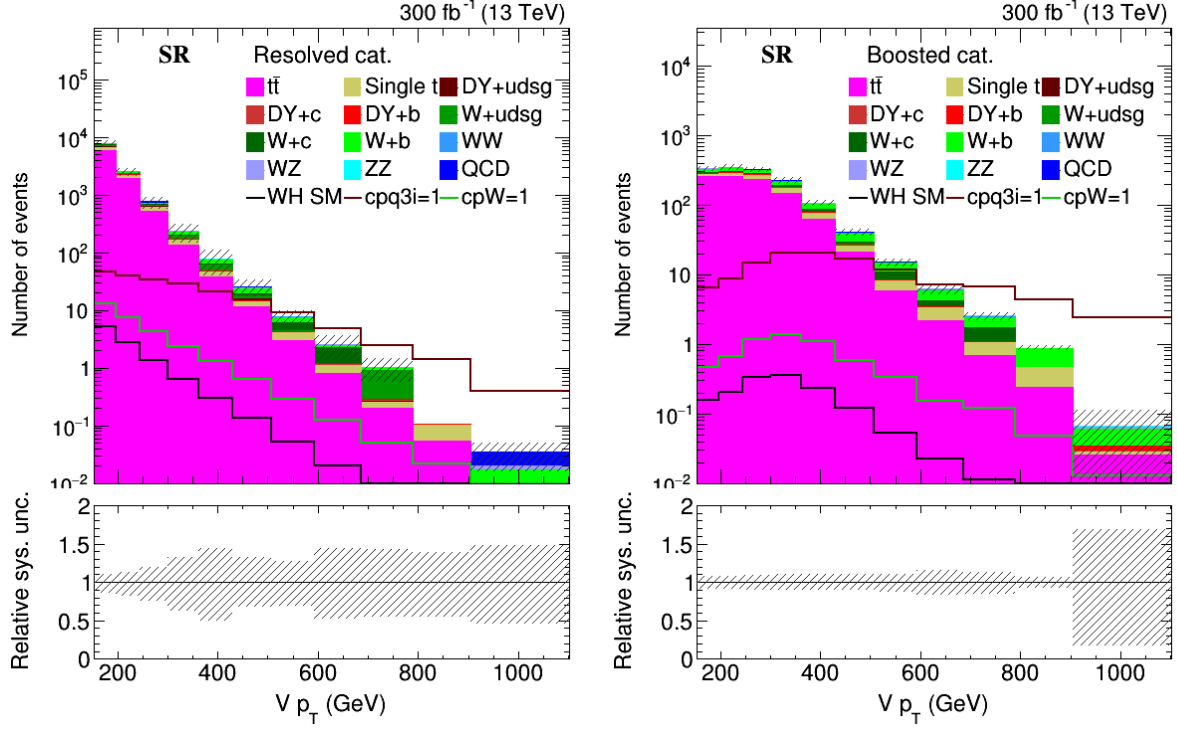
Figure 5. Distribution of W boson p_T in the resolved (left) and boosted (left) categories of the SR. Impact of systematic uncertainties on jet energy scale and resolution as well as correction factors used to match the b tagging and Higgs tagging efficiencies in data and simulation are shown in the bottom panel.

assigning reconstructed objects to particles produced in the hard interaction [62]. The φ observable is weighted by $\sin 2\Theta \sin 2\theta$ to lift the cancellation otherwise enforced by the triple angular correlation. For the first time, we thus resurrect the interference of the helicity amplitudes in H final states at the detector level.

6.3 Neutrino reconstruction

In order to reconstruct the momentum of the invisible neutrino in $W \to \ell\nu$, we assume that the ν is the only source of $\vec{p}_T^{\text{miss}}$ in the event and, hence, $\vec{p}_T^{\text{miss}}$ to be equal to the transverse mometum vector of ν. In a simple approach, we use the quadratic equation from the W-mass constraint to calculate $\eta(\nu)$. The two possible solutions are

$$\eta(\nu) = \eta(l) \pm \cosh^{-1}(1 + \Delta^2), \text{ where} \tag{6.1}$$

$$\Delta^2 = \frac{m_W^2 - m_T^2(\ell, p_T^{\text{miss}})}{2p_T^l \, p_T^{\text{miss}}}. \tag{6.2}$$

Large $p_T(W)$ leads to $\Delta \ll 1$, and the two solutions become

$$\eta(\nu) \simeq \eta(\ell) \pm \sqrt{2}\Delta + \mathcal{O}(\Delta^3) \ . \tag{6.3}$$

In this limit, the angular variable θ and Θ converge to the same values, whereas a two-fold ambiguity $\varphi_+ \simeq \pi - \varphi_-$ remains for φ. This ambiguity is visible in Fig. 7 for the WH signal sample and currently reduces the CP-sensitivity in WH production. No attempt was made to reduce the impact
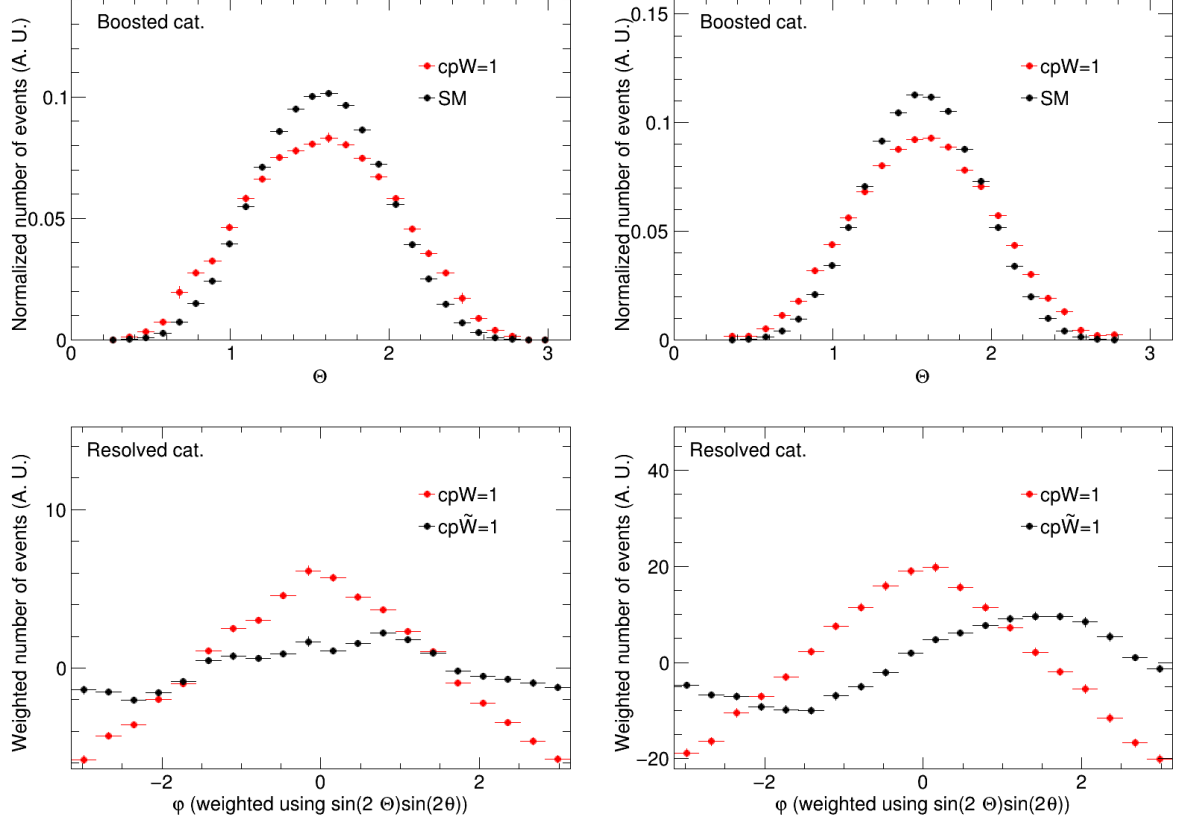
11

Figure 6. Normalized distribution of polar angle Θ in WH (top left) and ZH (top right) signal samples for $cpW = 1$ and SM scenarios in the boosted category. Normalized distribution of azimuthal angle φ in WH (bottom left) and ZH (bottom right) productions for $cpW = 1$ and $cp\tilde{W} = 1$, respectively; SM component is subtracted from these distributions.
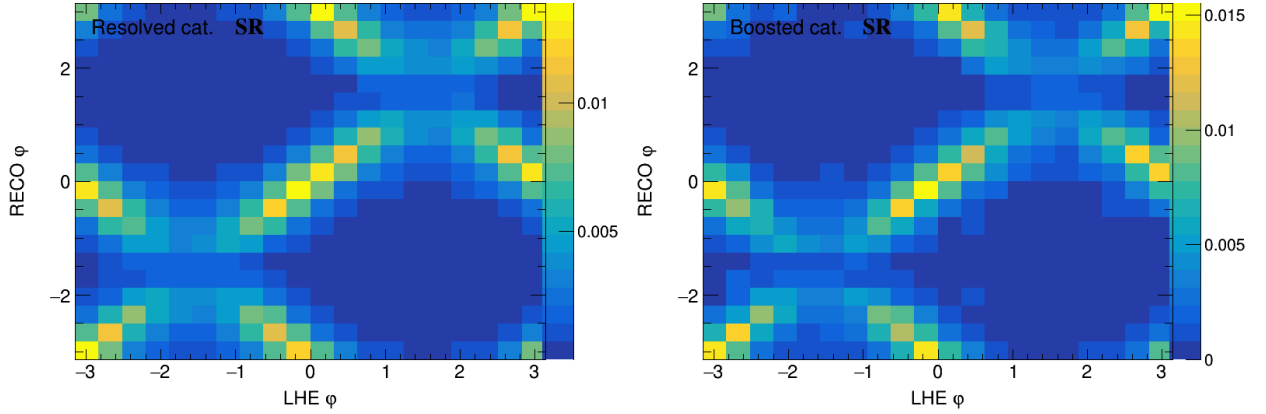

Figure 7. Relation between reconstructed- and parton-level φ in the resolved (left) and boosted (right) categories in WH production illustrating the effect of the ambiguity in the reconstruction of the neutrino momentum.

Table 3. Input variables to the NN discriminator designed to separate events from VH and $t\bar{t}$ productions.

p_T component of $\vec{p}_T^{\,\text{miss}}$	b tagging scores of b_1 and b_2	DeepAK8 top tagging score
p_T and η of the lepton	$\text{Min}(p_T(b_1), p_T(b_2))$	DeepAK8 W tagging score
$\Delta\phi(\ell, \vec{p}_T^{\,\text{miss}})$	$\text{Max}(p_T(b_1), p_T(b_2))$	DeepAK8 H tagging score
$\Delta\phi(W, H)$	$\text{Max}(p_T(b_1)/p_T(\ell), p_T(b_2)/p_T(\ell))$	
Rapidity and mass of H	$\Delta\phi(b_1, b_2)$	
Event thrust	AK4 jet multiplicity	
(see definition in [63])	(satisfying $p_T > 30\,\text{GeV}$, $|\eta| < 2.5$)	

of the ambiguity for the purpose of this proposal. As part of the analysis development, we will develop a multi-variate discriminator to select the correct solution based on the kinematic event properties.

6.4   Multivariate analysis

As shown in Fig. 5, the SRs are subject to significant backgrounds from $t\bar{t}$, $W+b/c$, and $W+udsg$. To separate the WH signals optimally from the backgrounds, neural network (NN) discriminators are constructed. In this study, it comprises two fully connected dense layers, while the discriminator will be promoted to a DNN as part of the proposed work. A very preliminary set of input variables is listed in Table 3. The variables in the left column of Table 3 are common to both the resolved and the boosted categories, while the ones in the middle and right columns characterize the jet system in the resolved and the boosted categories, respectively, for the separation of the WH signal from the $t\bar{t}$ background.   Normalized distributions of NN discriminators in SM WH and $t\bar{t}$ events are shown in Fig. 8 (left). The performance of the discriminator, quantified with resistive operator characteristics (ROC) curves [64], are shown in Fig. 8 (middle). Benefiting from the less combinatorial background, the discriminator constructed in the boosted category performs slightly better. In both cases, the signal distribution is sharply peaked at high discriminator values and no overtraining is observed, indicating a successfully trained classifier. In the next section, we quantify the sensitivity gain from this strategy. In the final work, a machine learning based algorithm developed by the applicant's group to separate the SM-EFT effects from the SM [37] will be used.

6.5   Sensitivity results

Although the final condition on the NN discriminator is subject to development during the project, we can obtain a simple assessment of the sensitivity with a threshold on the NN discriminator corresponding to a WH signal efficiency of 80%. This choice corresponds to a $t\bar{t}$ background efficiency of $\sim 30\%$ ($\sim 15\%$) in the resolved (boosted) category. Distributions of the W boson p_T after applying this threshold on the NN discriminator are shown in Fig. 8 (right). Next, we perform one-dimensional likelihood scans for the Wilson coefficients corresponding to three operators $\mathcal{O}_{Hq}^{(3)}$, $\mathcal{O}_{HW}$, and $\mathcal{O}_{H\tilde{W}}$, respectively, while setting all other Wilson coefficients to 0. In the proposed work, sensitivity for a particular Wilson coefficient will also be measured by allowing others to vary. Normalizing the predicted yields to an integrated luminosity of $300\,\text{fb}^{-1}$, we show the negative log-likelihood results in Fig. 9. For completeness, the results for the ZH signal, obtained by following
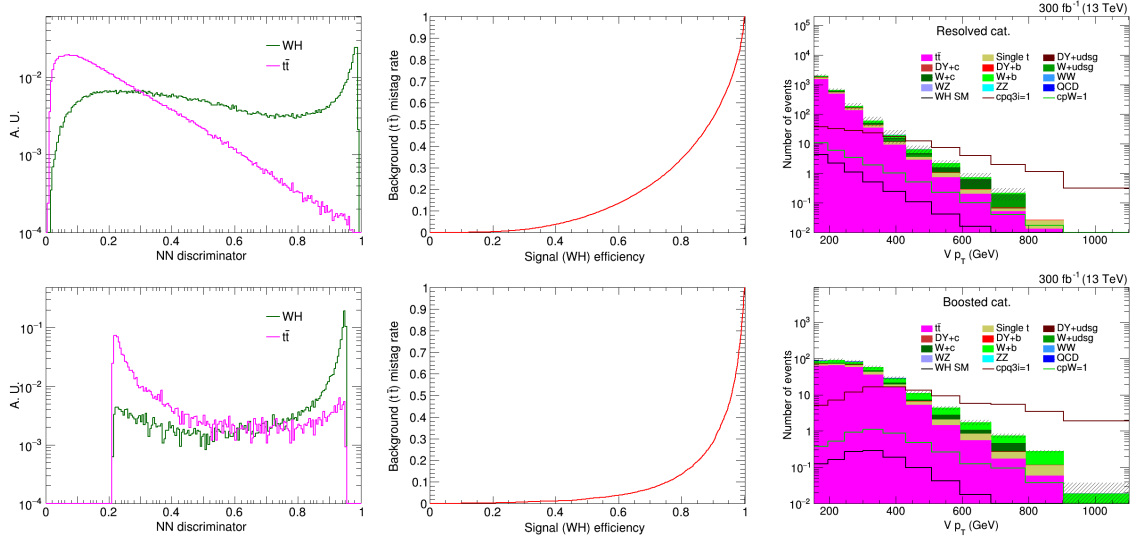
Figure 8. Normalized distributions of the NN discriminator in signal (WH) and background ($t\bar{t}$) events (left) and ROC curves characterizing the performance of NN discriminator (middle) and distributions of the W boson p_T after applying a condition on the NN discriminator (right) in the resolved (top) and boosted (bottom) categories.

a similar analysis strategy as mentioned for the WH process, are also combined. Fig. 9 shows that a large gain in sensitivity is obtained by adding the boosted category due to the energy growth induced by the operators: The coefficient corresponding to $\mathcal{O}_{Hq}^{(3)}$ can be constrained to percent level and the ones corresponding to $\mathcal{O}_{HW}$ and $\mathcal{O}_{H\tilde{W}}$ within tens of percent at 95% confidence level with luminosity amounting to $300\,\mathrm{fb}^{-1}$. An improvement due to a binning in φ is achieved in the resolved category. Addition of the results for the ZH process improves the sensitivity, particularly for the $\mathcal{O}_{HW}$ operator coefficient. The results of this preliminary study are at a similar level as the recent results from the ATLAS Collaboration [35], however, many improvements are expected in the proposed work during its execution. Also, simultaneous use of information about all three angles Θ, θ, φ is expected to result in a significant increase in sensitivity. The proposed work is also capable of disentangling the effects of CP-even and -odd SM-EFT operators, which is missing in Ref. [35].

6.6 Possible extensions

The proposal centers around SM-EFT operators in VH production, where the gauge boson V decays to leptons. The methodology developed is also suitable for hadronically decaying V; the only limitation in the case of V = W in sensitivity is the absence of knowledge on V charge. Latest developments in DNN-based methods for vector boson tagging can be used to increase the sensitivity.

Technologies developed for the proposed measurement will also help searches with $V \to \ell^+\ell^-/\ell\nu$ and $H \to b\bar{b}$ in the final state, particularly in the context of a heavy resonance decaying to V and H.
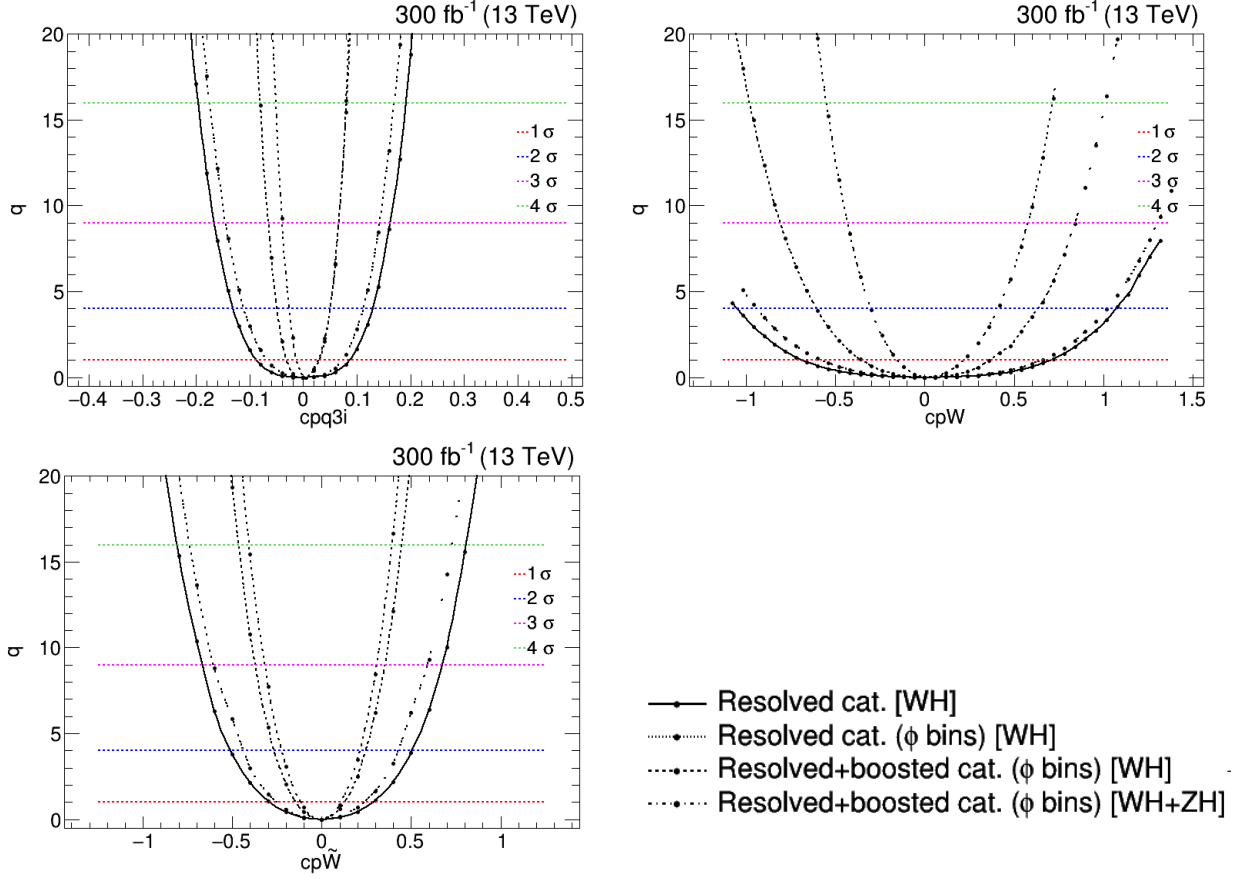
Figure 9. The profiled likelihood ratio test statistic q as a function of the Wilson coefficient $\mathcal{O}_{Hq}^{(3)}$ (top left), $\mathcal{O}_{HW}$ (top right), and $\mathcal{O}_{H\tilde{W}}$ (bottom left) for an integrated luminosity of $300\,\mathrm{fb}^{-1}$.

7 Planned cooperation arrangement

The work will be conducted by a PhD student (N.N.) under the supervision of the principal investigator (PI) Dr. Suman Chatterjee and within the CMS data analysis group at the Institute of High Energy Physics (HEPHY) of the Austrian Academy of Sciences. The PI and the PhD student are members of the CMS Collaboration and will work full time on the project.

The reconstructed data, the simulated background samples, and the object calibration are provided centrally by the CMS Collaboration. Ideas and updates will be presented at regular intervals in internal meetings at HEPHY and within the CMS working groups.

There are synergies with ongoing activities within the HEPHY CMS data analysis group. The group recently submitted a paper on measuring the $t\bar{t}\gamma$ process and constraining the $t-\gamma$ coupling in SM-EFT; this work was supported by FWF grant P31578. There is also ongoing work on constraining SM-EFT operators in the tWZ process supported by FWF grant P33771. The existing expertise in the group will help in the smooth execution of the proposed measurement, which will also broaden the scope of the group's activities.

The CMS data analysis group at HEPHY consists of five staff scientists, three post-doctoral researchers, and five PhD students working on supersymmetry, long-lived signatures, searches for

BSM Higgs boson decaying to a pair of τ leptons, and top quark physics. The diverse expertise of the group members will help to develop new ideas and implement those in the CMS Collaboration.

8 Work and plan

The work will be carried out by a PhD student (N.N.) and the applicant and PI Dr. Suman Chatterjee, both working full time on the project. In the foliing, the tasks are described in approximate chronological order and are summarized in Table 4.

1. **Signal simulation for the Run 2 analysis.** Simulate VH event samples including the effects of SM-EFT operators both at in production and decay using with, e.g., `SMEFTsim` [65] at leading order or `SMEFT@NLO` [66] at next-to-leading order in perturbative QCD. Particular care needs to be taken for the matching of the hard process generated using MAD-GRAPH5_aMC@NLO [57] and the parton shower with PYTHIA8 [67] since the parton shower does not involve higher-dimensional operators. Event reconstruction is done within CMS reconstruction software (CMSSW).

2. **SM-EFT parameterization.** Obtain polynomial parametrization of signal yields for arbitrary selections as a function of the Wilson coefficients. Validate the accuracy of the procedure with dedicated simulation of event samples with non-zero Wilson coefficients.

3. **Optimization of the lepton selection.** Optimize the lepton identification criteria minimizing the total systematic uncertainty. The optimization is done separately for the resolved and the boosted signal regions, as well as for the analysis of the angular observables. Assess the systematic uncertainty in the angular observables associated with the lepton identification.

4. **Optimization of b tagging criteria.** The CMS Collaboration provides the correction to be applied to simulation to match the efficiency of DNN-based b tagging in data and simulation in two variants. The first method corrects the simulated efficiency of pre-defined b tagging working points. The second approach is based on reweighting the simulated b-tagging discriminator shape to match the shape observed in data. Since the b tagging information is used for signal-to-background discrimination, in this task we study the performance and resulting uncertainties for the two options in terms of better overall sensitivity.

5. **Optimization of H tagging criterion and scale-factor measurement.** Optimize the DNN-based H tagging criterion for signal-to-background separation. Measure correction factors need to be measured in appropriate control samples, e.g. in $Z \to \bar{b}b$ event samples, to match the tagging efficiency in data.

6. **Trigger efficiency measurement.** The efficiencies of lepton-based triggers are measured in both data and simulation. Corrections to match the trigger efficiency in simulation and data are derived and applied to the simulation.

7. **MVA based neutrino momentum reconstruction.** A finely tuned momentum reconstruction is required to resolve the ambiguities in neutrino momentum reconstruction. Using global event observables and kinematic properties of the reconstructed objects, we develop an MVA algorithm based on neural nets to select the correct solution. This task particularly benefits the CP-sensitivity of the H–W coupling measurements. A powerful prescription developed for neutrino reconstruction will be a valuable asset beyond the scope of this proposal.

8. **Background estimation.** Scaling simulation to measured yields in side band regions is an important method of estimating backgrounds. The procedure is validated in statistically independent selections defined by inverting appropriate selection requirements. Corrections

accounting for are derived wherever needed, and the corresponding systematic uncertainties are quantified. Different techniques for the background estimation, e.g., using the mass sidebands, are be checked.

9. **Multivariate analysis.** Modeling of the variables used for signal-to-background separation in DNN-based multivariate analysis and their correlations need to be properly checked. The architecture of the DNN and the threshold on the NN discriminator need to be optimized. The modeling of the NN discriminator for different background processes will be checked, and data-to-simulation corrections will be derived if needed.

10. **SM-EFT to SM separation.** In the feasibility study, the constraints on Wilson coefficients are derived just by using a few kinematic variables. A more sophisticated study with a larger number of observables, taking into account their correlation, needs to be performed. This will be a test-bed of a method developed by the proposer's group using decision trees [37].

11. **Experimental systematic uncertainties.** Systematic uncertainties related to different objects, e.g., leptons and jets, and corrections at event-level need to be applied. The systematic uncertainties need to be properly derived for the corrections developed during this particular analysis. The correlation between systematic uncertainty sources in different years needs to be appropriately defined. It will be carefully checked if some of the systematic uncertainties are unexpectedly constrained.

12. **Modeling uncertainties.** Systematic uncertainties in prediction from the choice of renormalization and factorization scales, parton distribution function, and the parton shower need to be derived. The neutrino reconstruction is also expected to be affected by the modeling uncertainties.

13. **Sensitivity report.** Results from different years of Run 2 will be combined and the sensitivity of the analysis will be quantified using likelihood scans, taking into account correlations between the operators.

14. **SM-EFT effects in the background.** So far, all the discussions regarding the SM-EFT operators are made for VH production, which is the signal targeted. However, some of the operators in Table 1 also affect some of the background processes. For example, $\mathcal{O}_{Hq}^{(1)}$, $\mathcal{O}_{Hq}^{(3)}$ affect the diboson production, which is related to VH production by Goldstone equivalence theorem [68] at high energy. These are also expected to affect V+jets production. A significant amount of effort needs to be dedicated to the simulation of SM-EFT effects for the backgrounds, followed by a joint analysis of signal and background to report the final sensitivity to the corresponding Wilson coefficients.

15. **Publication based on Run 2 data.** A first paper will be written reporting the results based on Run 2 data. This involves a series of reviews within the CMS Collaboration and finally by the journal reviewers.

16. **Signal simulation for Run 3.** The setup developed for the signal simulation for Run 2 will be used in Run 3. However, options will be kept open to incorporate the developments in the theory community by this period.

17. **Trigger efficiency measurement and optimization of object selection for Run 3.** The triggers which were operational during Run 2 are expected to be available for Run 3 also, but the pileup condition is likely to change. So, the trigger efficiency and corresponding corrections need to be also derived for Run 3. For the same reason, object selection criteria need to be reevaluated for Run 3.

18. **Choice of H tagger for Run 3.** With the fast progress in heavy particle tagging, more powerful taggers are expected to be available during the Run 3 schedule. An evaluation will be performed based on the available taggers, and the best performant one will be used.

19. **Background estimation for Run 3.** Improvements are expected in the Monte Carlo simulation of different processes in the next few years. Thus, the background estimation strategy to be developed for Run 2 analysis will need a validation for Run 3.

20. **Experimental and modeling systematic uncertainties for Run 3.** Systematic uncertainties due to experimental sources will be added. Modeling uncertainties are expected to be very similar to Run 2.

21. **Combination of Run 2 and 3 results.** Results from the analyses based on Run 2 and Run 3 data samples will be combined. Final sensitivity on Wilson coefficients will be reported using the combined data set.

22. **Publication of final results.** The final results will be reported in a paper, which will go through the internal review process by the CMS Collaboration and finally be submitted to the journal for publication.

9   Ethical, safety-related or regulatory aspects

The sole aim of the project is to improve the human understanding of nature. The topic itself has no ethical, safety-related, or regulatory aspects that need to be considered.

10   Sex-specific and gender-related issues

The project aims to advance our fundamental understanding of nature. Thus the topic itself has no gender-relevant aspects. However, the PI is aware of the gender and diversity issues in sciences in general and actively engages in creating equal opportunities for all groups. The hiring process at HEPHY is accompanied by a gender-and-diversity officer (Barbara Weber), employed by the institute, who ensures equal opportunities in all communications with candidates. W. Adam from HEPHY is a member of the CMS committee for diversity and inclusion and will oversee the hiring process. The job advertisement will be formulated gender neutrally. In order to improve the gender balance at HEPHY (23%), a suitable female candidate will be given preference.

11   Human resources

A PhD student will be recruited solely for this project to be conducted at HEPHY, Vienna. The PhD student will carry out the project under the supervision of the project applicant, PI Dr. S. Chatterjee. The PI will spend 100% of his time on this project. The PI has five years of experience in CMS data analysis in the context of precision jet measurements, searches for new physics, detector calibration, and implementation of algorithms in the CMS software framework. The PI has also worked on high energy physics phenomenology in collaboration with theorists and published papers in international journals.

The tasks will be carried out in consultation with the leader of CMS data analysis group of the institute: Dr. Robert Schöfbeck (RS), who is also performing SM-EFT measurements in top quark physics in association with a post-doctoral researcher Dr. Dennis Schwartz (DS). Particularly, SM-EFT parameterization (task 2) and SM-EFT to SM separation (task 10) will be performed in collaboration with RS and DS, but they do not take responsibility for the tasks in the work and

time plan. RS is also the former convener of the "top quark mass and properties" physics analysis group and the "Jets and MET" physics object group of the CMS Collaboration. His feedback on documentation and presentation of the results will be constructive.

Task / Year	2022 ($\int \mathcal{L}dt \sim 30\,\mathrm{fb}^{-1}$)			2023 ($\int \mathcal{L}dt \sim 120\,\mathrm{fb}^{-1}$)				2024 ($\int \mathcal{L}dt \sim 160\,\mathrm{fb}^{-1}$)			2025
	Apr-Jun	Jul-Sep	Oct-Dec	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec	Jan-Mar	Apr-Jun	Jul-Dec	Jan-Mar
1. Signal simulation for Run 2	✓										
2. SM-EFT parameterization	✓										
3. Optimization of lepton selection		✓									
4. Optimization of b tagging condition		✓									
5. Optimization of H tagging condition		✓									
6. Trigger efficiency measurement			✓								
7. Neutrino reconstruction			✓								
8. Background estimation				✓							
9. Multivariate analysis				✓							
10. SM-EFT to SM separation					✓						
11. Exp. systematic uncertainties					✓						
12. Modeling uncertainties					✓						
13. Quantification of sensitivity						✓					
14. SM-EFT effects in background						✓					
15. Publication based on Run 2 data							✓				
16. Signal simulation for Run 3							✓				
17. Selections for Run 3								✓			
18. Choice of H tagger for Run 3								✓			
19. Background estimation for Run 3									✓		
20. Systematic uncertainties for Run 3										✓	
21. Combination of Run 2 and 3 results										✓	
22. Publication of final results											✓

Table 4. Tasks shown using symbol ✓ will be covered by the PhD student, ✓ by the project applicant and PhD student, ✓ by the project applicant and PhD student in collaboration with other members in HEPHY CMS data analysis group.

Annex 1   References

[1] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, `doi:10.1088/1748-0221/3/08/S08004`.

[2] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett.* **B716** (2012) 1, `doi:10.1016/j.physletb.2012.08.020`, `arXiv:1207.7214`.

[3] CMS Collaboration, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC", *Phys. Lett. B* **716** (2012) 30–61, `doi:10.1016/j.physletb.2012.08.021`, `arXiv:1207.7235`.

[4] CMS Collaboration, "Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} = 13$ TeV", *JHEP* **11** (2017) 047, `doi:10.1007/JHEP11(2017)047`, `arXiv:1706.09936`.

[5] CMS Collaboration, "A measurement of the Higgs boson mass in the diphoton decay channel", *Phys. Lett. B* **805** (2020) 135425, `doi:10.1016/j.physletb.2020.135425`, `arXiv:2002.06398`.

[6] CMS Collaboration, "Constraints on anomalous HVV couplings from the production of Higgs bosons decaying to τ lepton pairs", *Phys. Rev. D* **100** (2019), no. 11, 112002, `doi:10.1103/PhysRevD.100.112002`, `arXiv:1903.06973`.

[7] CMS Collaboration, "Measurements of $t\bar{t}H$ Production and the CP Structure of the Yukawa Interaction between the Higgs Boson and Top Quark in the Diphoton Decay Channel", *Phys. Rev. Lett.* **125** (2020) 061801, `doi:10.1103/PhysRevLett.125.061801`, `arXiv:2003.10866`.

[8] B. Grinstein and M. B. Wise, "Operator analysis for precision electroweak physics", *Phys. Lett. B* **265** (1991) 326, `doi:10.1016/0370-2693(91)90061-T`.

[9] J.-y. Chiu, F. Golf, R. Kelley, and A. V. Manohar, "Electroweak Corrections in High Energy Processes using Effective Field Theory", *Phys. Rev. D* **77** (2008) 053004, `doi:10.1103/PhysRevD.77.053004`, `arXiv:0712.0396`.

[10] G. Passarino and M. Trott, "The Standard Model Effective Field Theory and Next to Leading Order", `arXiv:1610.08356`.

[11] E. E. Jenkins, A. V. Manohar, and M. Trott, "Renormalization Group Evolution of the Standard Model Dimension Six Operators I: Formalism and lambda Dependence", *JHEP* **10** (2013) 087, `doi:10.1007/JHEP10(2013)087`, `arXiv:1308.2627`.

[12] R. Alonso, E. E. Jenkins, A. V. Manohar, and M. Trott, "Renormalization Group Evolution of the Standard Model Dimension Six Operators III: Gauge Coupling Dependence and Phenomenology", *JHEP* **04** (2014) 159, `doi:10.1007/JHEP04(2014)159`, `arXiv:1312.2014`.

[13] E. E. Jenkins, A. V. Manohar, and M. Trott, "Renormalization Group Evolution of the Standard Model Dimension Six Operators II: Yukawa Dependence", *JHEP* **01** (2014) 035, `doi:10.1007/JHEP01(2014)035`, `arXiv:1310.4838`.

[14] C. Englert and M. Spannowsky, "Effective Theories and Measurements at Colliders", *Phys. Lett. B* **740** (2015) 8–15, `doi:10.1016/j.physletb.2014.11.035`, `arXiv:1408.5147`.

[15] I. Brivio and M. Trott, "The Standard Model as an Effective Field Theory", *Phys. Rept.* **793** (2019) 1, `doi:10.1016/j.physrep.2018.11.002`, `arXiv:1706.08945`.

[16] G. Panico, F. Riva, and A. Wulzer, "Diboson interference resurrection", *Phys. Lett. B* **776** (2018) 473, `doi:10.1016/j.physletb.2017.11.068`, `arXiv:1708.07823`.

[17] CMS Collaboration Collaboration, "W$^{\pm}\gamma$ differential cross sections and effective field theory constraints at $\sqrt{s} = 13\,\text{TeV}$", CMS Physics Analysis Summary CMS-PAS-SMP-20-005, 2021.

[18] S. Banerjee et al., "Towards the ultimate differential SMEFT analysis", *JHEP* **09** (2020) 170, `doi:10.1007/JHEP09(2020)170`, `arXiv:1912.07628`.

[19] S. Weinberg, "Baryon- and lepton-nonconserving processes", *Phys. Rev. Lett.* **43** (Nov, 1979) 1566, `doi:10.1103/PhysRevLett.43.1566`.

[20] F. Bonnet, D. Hernandez, T. Ota, and W. Winter, "Neutrino masses from higher than d=5 effective operators", *JHEP* **10** (2009) 076, `doi:10.1088/1126-6708/2009/10/076`, `arXiv:0907.3143`.

[21] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, "Dimension-Six Terms in the Standard Model Lagrangian", *JHEP* **10** (2010) 085, `doi:10.1007/JHEP10(2010)085`, `arXiv:1008.4884`.

[22] K. Hagiwara, R. Szalapski, and D. Zeppenfeld, "Anomalous Higgs boson production and decay", *Phys. Lett. B* **318** (1993) 155–162, `doi:10.1016/0370-2693(93)91799-S`, `arXiv:hep-ph/9308347`.

[23] J. Ellis, V. Sanz, and T. You, "Complete Higgs Sector Constraints on Dimension-6 Operators", *JHEP* **07** (2014) 036, `doi:10.1007/JHEP07(2014)036`, `arXiv:1404.3667`.

[24] C. W. Murphy, "Statistical approach to Higgs boson couplings in the standard model effective field theory", *Phys. Rev. D* **97** (2018), no. 1, 015007, `doi:10.1103/PhysRevD.97.015007`, `arXiv:1710.02008`.

[25] J. Baglio et al., "Validity of standard model EFT studies of VH and VV production at NLO", *Phys. Rev. D* **101** (2020) 115004, `doi:10.1103/PhysRevD.101.115004`, `arXiv:2003.07862`.

[26] J. Ellis, V. Sanz, and T. You, "The Effective Standard Model after LHC Run I", *JHEP* **03** (2015) 157, `doi:10.1007/JHEP03(2015)157`, `arXiv:1410.7703`.

[27] C. Grojean, M. Montull, and M. Riembau, "Diboson at the LHC vs LEP", *JHEP* **03** (2019) 020, `doi:10.1007/JHEP03(2019)020`, `arXiv:1810.05149`.

[28] R. S. Gupta, A. Pomarol, and F. Riva, "BSM Primary Effects", *Phys. Rev. D* **91** (2015), no. 3, 035001, `doi:10.1103/PhysRevD.91.035001`, `arXiv:1405.0181`.

[29] A. Manohar and H. Georgi, "Chiral Quarks and the Nonrelativistic Quark Model", *Nucl. Phys. B* **234** (1984) 189–212, `doi:10.1016/0550-3213(84)90231-1`.

[30] H. Qu and L. Gouskos, "ParticleNet: Jet Tagging via Particle Clouds", `arXiv:1902.08570`.

[31] CMS Collaboration, "Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques", *JINST* **15** (2020) P06005, `doi:10.1088/1748-0221/15/06/P06005`, `arXiv:2004.08262`.

[32] A. Falkowski and F. Riva, "Model-independent precision constraints on dimension-6 operators", *JHEP* **02** (2015) 039, `doi:10.1007/JHEP02(2015)039`, `arXiv:1411.0669`.

[33] S. Banerjee, C. Englert, R. S. Gupta, and M. Spannowsky, "Probing Electroweak Precision Physics via boosted Higgs-strahlung at the LHC", *Phys. Rev. D* **98** (2018), no. 9, 095012, `doi:10.1103/PhysRevD.98.095012`, `arXiv:1807.01796`.

[34] N. Berger et al., "Simplified Template Cross Sections - Stage 1.1", `arXiv:1906.02754`.

[35] ATLAS Collaboration, "Combination of measurements of Higgs boson production in association with a W or Z boson in the $b\bar{b}$ decay channel with the ATLAS experiment at $\sqrt{s} = 13$ TeV", Technical Report ATLAS-CONF-2021-051, CERN, Geneva, 2021.

[36] CMS Collaboration Collaboration, "Combined Higgs boson production and decay measurements with up to 137 fb-1 of proton-proton collision data at sqrts = 13 TeV", CMS Physics Analysis Summary CMS-PAS-HIG-19-005, 2020.

[37] S. Chatterjee et al., "Tree boosting for learning EFT parameters", arXiv:2107.10859.

[38] J. Elias-Miro, J. R. Espinosa, E. Masso, and A. Pomarol, "Higgs windows to new physics through d=6 operators: constraints and one-loop anomalous dimensions", *JHEP* **11** (2013) 066, doi:10.1007/JHEP11(2013)066, arXiv:1308.1879.

[39] J. Ellis, C. W. Murphy, V. Sanz, and T. You, "Updated Global SMEFT Fit to Higgs, Diboson and Electroweak Data", *JHEP* **06** (2018) 146, doi:10.1007/JHEP06(2018)146, arXiv:1803.03252.

[40] J. Ellis et al., "Top, Higgs, Diboson and Electroweak Fit to the Standard Model Effective Field Theory", *JHEP* **04** (2021) 279, doi:10.1007/JHEP04(2021)279, arXiv:2012.02779.

[41] J. J. Ethier et al., "Combined SMEFT interpretation of Higgs, diboson, and top quark data from the LHC", arXiv:2105.00006.

[42] CMS Collaboration, "Constraints on anomalous Higgs boson couplings to vector bosons and fermions in its production and decay using the four-lepton final state", arXiv:2104.12152.

[43] CMS Collaboration, "Probing effective field theory operators in the associated production of top quarks with a Z boson in multilepton final states at $\sqrt{s} = 13$ TeV", arXiv:2107.13896.

[44] CMS Collaboration, "Measurement of the electroweak production of Zγ and two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV and constraints on anomalous quartic gauge couplings", arXiv:2106.11082.

[45] "LHC Effective Field Theory WG". https://lpcc.web.cern.ch/lhc-eft-wg. https://lpcc.web.cern.ch/lhc-eft-wg.

[46] "All Things EFT". https://sites.google.com/view/all-things-eft. doi:https://sites.google.com/view/all-things-eft, https://sites.google.com/view/all-things-eft.

[47] A. G. Cohen, A. De Rujula, and S. Glashow, "A matter - antimatter universe?", *Astrophys. J.* **495** (1998) 539, doi:10.1086/305328, arXiv:astro-ph/9707087.

[48] P. H. Damgaard, A. Haarr, D. O'Connell, and A. Tranberg, "Effective Field Theory and Electroweak Baryogenesis in the Singlet-Extended Standard Model", *JHEP* **02** (2016) 107, doi:10.1007/JHEP02(2016)107, arXiv:1512.01963.

[49] B. Grzadkowski and D. Huang, "Spontaneous *CP*-Violating Electroweak Baryogenesis and Dark Matter from a Complex Singlet Scalar", *JHEP* **08** (2018) 135, doi:10.1007/JHEP08(2018)135, arXiv:1807.06987.

[50] J. de Blas, M. Chala, M. Perez-Victoria, and J. Santiago, "Observable Effects of General New Scalar Particles", *JHEP* **04** (2015) 078, doi:10.1007/JHEP04(2015)078, arXiv:1412.8480.

[51] S. Profumo, M. J. Ramsey-Musolf, C. L. Wainwright, and P. Winslow, "Singlet-catalyzed electroweak phase transitions and precision Higgs boson studies", *Phys. Rev. D* **91** (2015), no. 3, 035018, doi:10.1103/PhysRevD.91.035018, arXiv:1407.5342.

[52] F. del Aguila, J. de Blas, and M. Perez-Victoria, "Electroweak Limits on General New Vector Bosons", *JHEP* **09** (2010) 033, doi:10.1007/JHEP09(2010)033, arXiv:1005.3998.

[53] G. Burdman, B. A. Dobrescu, and E. Ponton, "Resonances from two universal extra dimensions", *Phys. Rev. D* **74** (2006) 075008, doi:10.1103/PhysRevD.74.075008, arXiv:hep-ph/0601186.

[54] J. C. Pati and A. Salam, "Lepton number as the fourth color", *Phys. Rev. D* **10** (1974) 275, doi:10.1103/PhysRevD.10.275.

[55] F. del Aguila, M. Perez-Victoria, and J. Santiago, "Effective description of quark mixing", *Phys. Lett. B* **492** (2000) 98, doi:10.1016/S0370-2693(00)01071-6, arXiv:hep-ph/0007160.

[56] S. Dawson and E. Furlan, "A Higgs Conundrum with Vector Fermions", *Phys. Rev. D* **86** (2012) 015021, doi:10.1103/PhysRevD.86.015021, arXiv:1205.4733.

[57] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, `doi:10.1007/JHEP07(2014)079`, `arXiv:1405.0301`.

[58] P. Artoisenet and O. Mattelaer, "MadWeight: Automatic event reweighting with matrix elements", *PoS* **CHARGED2008** (2008) 025, `doi:10.22323/1.073.0025`.

[59] I. Brivio, Y. Jiang, and M. Trott, "The SMEFTsim package, theory and tools", *JHEP* **12** (2017) 070, `doi:10.1007/JHEP12(2017)070`, `arXiv:1709.06492`.

[60] E. Bols et al., "Jet flavour classification using DeepJet", *JINST* **15** (2020) P12012, `doi:10.1088/1748-0221/15/12/P12012`, `arXiv:2008.10519`.

[61] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, "Soft Drop", *JHEP* **05** (2014) 146, `doi:10.1007/JHEP05(2014)146`, `arXiv:1402.2657`.

[62] CMS Collaboration, "Measurement of $t\bar{t}$ normalised multi-differential cross sections in pp collisions at $\sqrt{s} = 13$ TeV, and simultaneous determination of the strong coupling strength, top quark pole mass, and parton distribution functions", *Eur. Phys. J. C* **80** (2020), no. 7, 658, `doi:10.1140/epjc/s10052-020-7917-7`, `arXiv:1904.05237`.

[63] CMS Collaboration, "Study of Hadronic Event-Shape Variables in Multijet Final States in pp Collisions at $\sqrt{s} = 7$ TeV", *JHEP* **10** (2014) 087, `doi:10.1007/JHEP10(2014)087`, `arXiv:1407.2856`.

[64] T. Fawcett, "An introduction to roc analysis", *Pattern Recognition Letters* **27** (2006), no. 8, 861–874, `doi:https://doi.org/10.1016/j.patrec.2005.10.010`. ROC Analysis in Pattern Recognition.

[65] I. Brivio, "SMEFTsim 3.0 — a practical guide", *JHEP* **04** (2021) 073, `doi:10.1007/JHEP04(2021)073`, `arXiv:2012.11343`.

[66] C. Degrande et al., "Automated one-loop computations in the standard model effective field theory", *Phys. Rev. D* **103** (2021) 096024, `doi:10.1103/PhysRevD.103.096024`, `arXiv:2008.11743`.

[67] T. Sjöstrand et al., "An Introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159, `doi:10.1016/j.cpc.2015.01.024`, `arXiv:1410.3012`.

[68] J. M. Cornwall, D. N. Levin, and G. Tiktopoulos, "Derivation of gauge invariance from high-energy unitarity bounds on the s matrix", *Phys. Rev. D* **10** (Aug, 1974) 1145, `doi:10.1103/PhysRevD.10.1145`.

Annex 2   Research institution and required funding

The Institute of High Energy Physics (HEPHY) of the Austrian Academy of Sciences was founded in 1966. Its main purpose is the research in high energy physics and to highlight Austria's membership at CERN. On the hardware side, HEPHY has made significant contributions to the CMS inner tracker and to the trigger system. The CMS data analysis group, led by Dr. R. Schöfbeck, consists of 14 members based at offices in the Apostelgasse 23 in Vienna and at CERN. The successful candidate will be based in Vienna. This choice will enable a direct collaboration with the other PhD students and two master students doing data analysis in CMS, as well as with the scientist of the New Physics theory group. Four members are based at CERN. Among them is W. Adam, the former physics coordinator of the CMS experiment and the leader of the HEPHY team at CERN.

The infrastructure available to the HEPHY CMS analysis group comprises:

- Office space with personal workstations at Apostelgasse 23.

- Access to an LHC-Grid (LCG) Tier-2 cluster with 1000 CPU cores and 500 TByte storage. Access to the CLIP computing cluster at the Gregor Mendel Institute of the Austrian Academy of sciences with approx. 10k CPU cores and 50 TB of storage for analysis work.

- Video-conferencing equipment for daily communication with other CMS physicists and participation in CMS Collaboration meetings.

This project application requests funding for one PhD student for 36 months (EUR 39.780 per year in 2021). The PI S. Chatterjee is a post-doctoral researcher at HEPHY. Upon a positive funding decision by the FWF, a call for the PhD student will be opened, advertised inside and outside Austria, and made equally accessible to physicists regardless of their race, color, creed, national origin, religion, family status, sexual orientation, age, and political beliefs. In order to improve the gender balance at HEPHY ($\sim 23\%$), a suitable female candidate will be given preference.

Furthermore, a total of EUR 5.155 per year for travel expenses between Vienna and CERN is requested. These costs are justified as follows: The PhD student will travel four times a year (approximately every two months) to CERN for, on average, one week. The purpose of the four trips is the collaboration with the other experimentalists in the CMS HIG group and to present in person to the other members of the CMS Collaboration. Also, the PhD student will get the opportunity to participate in the detector operation during the data-taking periods. If possible, the CERN trips will overlap with a CMS Collaboration week (internal CMS working meeting occurring four times a year at CERN). Furthermore, one trip to the annual Higgs conference is foreseen, where recent developments in experimental and theoretical results on the Higgs boson are presented. The cost for this trip is assumed to be the same as for a trip to CERN. A breakdown of the expected costs of a working trip to CERN is provided in Table 5.

The institute needs to pay an amount (memorandum of operation) for the PI, which amounts to EUR 10.000 per year, this is also requested in the proposal. An extra 5% of the total that amounts to EUR 164.805 is requested to cover other unexpected costs. Total requested funding sums up to EUR 173.045. A summary of the total costs of the project is provided in Table 6.

Table 5. Estimation of travel costs for trips to CERN.

Item	Cost		Comment
Flight VIE-GVA-VIE	EUR	300	
7 per diem at EUR 28,10	EUR	196,70	
7 per noctem at EUR 24,90	EUR	174,30	
6 hotel nights at EUR 60	EUR	360	
Total per year	EUR	5.155	5 trips per year
Total	EUR	15.465	For 3 years in total

Table 6. The total amount of requested funding along with divisions for different items.

Item	Cost	
PhD student's salary	EUR	119.340
PhD student's travel costs of	EUR	15.465
MoE of the PI (for 3 years)	EUR	30.000
General costs (5%)	EUR	8.240
Total	EUR	173.045

Annex 3   CV of the PI

Personal Data

Age: 27 years (as of February 1, 2021)
Nationality: Indian
Sex: Male
Affiliation: Institute of High Energy Physics of the Austrian Academy of Sciences, 1050 Vienna
Designation: Post-doctoral researcher

Contact Details

Email: suman.chatterjee@oeaw.ac.at, suman.chatterjee@cern.ch
Phone: +91-7710835260, +43-6769390476
Personal website: https://sumanchatterjeetifr.wordpress.com

Academic Records

2016-2020: PhD in Experimental Particle Physics
Department of High Energy Physics
Tata Institute of Fundamental Research, Mumbai 400005, India
PhD Supervisor: Prof. Gobinda Majumder

2014-2016: Master of Science in Physics (Rank: 1st)
Tata Institute of Fundamental Research, Mumbai 400005, India

2011-2014: Bachelor of Science in Physics (Remark: 1st class)
Department of Physics
Jadavpur University, Kolkata 700032, India

<u>Achievements</u>

2020	**Honorable mention** for the **IPA Rahul Basu Memorial Award for Best Thesis in High Energy Physics** for the period 2018–2020 in XXIV DAE-BRNS HEP Symposium 2020, India
2015	**Professor Sukumar Biswas PhD Student Award for Excellence in Physics** for scoring the highest grades in graduate courses in the Physics Integrated PhD program in Tata Institute of Fundamental Research
2014	**All India Rank 3** in Joint Entrance Screening Test 2014 conducted by the research institutes in India
2011	Awarded by the Chief Minister of the state of West Bengal for securing **the 4th rank in the state** in Higher Secondary (10+2) Examination (conducted by West Bengal Council for Higher Secondary Education)

Oral Presentations at Conferences, Workshops

September 2021 Joint annual meeting of Austrian and Swiss physical societies, Innsbruck, Austria
Parallel talk   SM-EFT results in Higgs sector from the CMS experiment

July 2021 European Conference on High Energy Physics 2021, Hamburg, Germany (online)
Parallel talk   Jet substructure measurements in CMS

Dec 2020 XXIV DAE-BRNS HIGH ENERGY PHYSICS SYMPOSIUM 2020
Rahul Basu Memorial Thesis Award Seminar   Jets as probes for precision measurement and candles for physics beyond standard model

Jan 2020 QCD with Electron-Ion Collider, IIT Bombay, Mumbai, India
Plenary talk   Understanding parton distribution function from LHC measurements

Sep 2019 Workshop on Top Quark Physics 2019, IHEP, Beijing, China
Invited talk   Heavy resonance searches in ATLAS and CMS

July 2019 European Conference on High Energy Physics 2019, Ghent, Belgium
Parallel talk   Inclusive jet results from CMS experiment

July 2019 European Conference on High Energy Physics 2019, Ghent, Belgium
Parallel talk   Tagging top in leptonic decay

Feb 2019 International Workshop on Forward and Jet Physics at LHC, Bose Institute, Kolkata, India
Invited talk   Measurement of inclusive jet cross section and event properties in CMS experiment

Aug 2018 QCD at LHC 2018, Technische Universität Dresden, Dresden, Germany
Parallel talk   LHC top quark and jet measurements sensitive to PDFs

Aug 2018 QCD at LHC 2018, Technische Universität Dresden, Dresden, Germany
Parallel talk   Differential jet cross sections at the CMS experiment

Dec 2017 WHEPP XV, IISER Bhopal, Bhopal, India
Parallel talk   Radius Scan for Inclusive Jets in CMS Experiment at $\sqrt{s} = 13\,\mathrm{TeV}$

Dec 2016 XXII DAE-BRNS High Energy Physics Symposium, Delhi, India
Parallel talk   Perspectives for a Radius Scan in Inclusive Jets at $13\,\mathrm{TeV}$ with CMS

Poster Presentations

June 2021 Large Hadron Collider Physics (LHCP 2021) conference, Paris, France (online)
Search for W' bosons decaying to a top and a bottom quark at $\sqrt{s} = 13\,\mathrm{TeV}$

Sep 2018 Asia-Europe-Pacific School on High Energy Physics, Quy Nohn, Vietnam
Radius scan for inclusive jets at $\sqrt{s} = 13\,\mathrm{TeV}$

Teaching Experience

Autumn 2019: Teaching assistant
Course: Statistical Methods in Physics at TIFR, Mumbai

January 2019: Co-coordinator
Lecture and hands-on exercise on algorithms used in particle physics for event generation, object reconstruction, and analysis at the XII School on Experimental High Energy Physics conducted by the Science and Engineering Research Board in India at TIFR, Mumbai

Autumn 2016: Teaching assistant
Course: Advanced Electrodynamics at TIFR, Mumbai

Supervision Experience

Anirban Bala (PhD student at TIFR, Mumbai, Advisor: Prof. M. Guchait):
During his departmental project I (July-September, 2020) under Prof. G. Majumder
Designed the project and mentored thoroughly during its implementation

Organizational Responsibilities

XII School on Experimental High Energy Physics at TIFR, Mumbai
Helped the organization at the institute level

SUSY 2017 conference at TIFR, Mumbai
Led the team of volunteers

Selected Publications

1: Search for a W$'$ boson decaying to a top and a bottom quark at $\sqrt{s} = 13\,\text{TeV}$
CMS Collaboration
Phys. Lett. B 820 (2021) 136535
Most stringent limits published to date on right- and left-handed W$'$ bosons decaying to a top and a bottom quark.

2: Dependence of inclusive jet production on the anti-k_T distance parameter in pp collisions at $\sqrt{s} = 13\,\text{TeV}$
CMS Collaboration
JHEP 12 (2020) 082
First public results from colliders on dependence of inclusive jet cross section on jet size.

3: Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s} = 13\,\text{TeV}$
CMS Collaboration
JINST 15 (2020) P05002
Absolute calibration of outer hadron calorimeter in particle-flow algorithm in CMS.

4: Jets with electrons from boosted top quarks
Suman Chatterjee, Rohini Godbole, and Tuhin S. Roy
JHEP 01 (2020) 170
An algorithm to identify boosted top quarks decaying where W boson from top quark decays to an electron using jet substructure techniques.

5: Mixed WIMP-axion dark matter
Suman Chatterjee, Anirban Das, Tousik Samui, and Manibrata Sen
Phys. Rev. D 100, 115050
A detailed study of the phenomenology of a multi-component dark matter scenario with a scalar WIMP and the QCD axion in the light of the latest data from dark matter detection experiments and LHC results.

6: Dose rate effects in the radiation damage of the plastic scintillators of the CMS hadron endcap calorimeter
CMS Collaboration
2016 JINST 11 T10004
Measurements of change in light output by plastic scintillators in the CMS hadron endcap due to irradiation during the collisions at LHC at $\sqrt{s} = 8\,TeV$.