

## Project description

## Contents

<b>1</b>	<b>Science</b>	<b>2</b>
<b>2</b>	<b>Research hypothesis</b>	<b>3</b>
2.1	Effective field theory couplings in $V (\rightarrow \ell \ell) H (\rightarrow b \bar{b})$ production . . . . .	3
2.2	Direct measurements of SMEFT couplings in $V (\rightarrow \ell \ell) H (\rightarrow b \bar{b})$ . . . . .	3
<b>3</b>	<b>Objectives</b>	<b>4</b>
<b>4</b>	<b>Scientific originality</b>	<b>5</b>
<b>5</b>	<b>Relevance to the research field</b>	<b>6</b>
<b>6</b>	<b>Method</b>	<b>6</b>
6.1	Final states and observables . . . . .	7
6.2	Event selection and categorization . . . . .	8
6.3	Neutrino reconstruction . . . . .	9
6.4	Multivariate analysis . . . . .	10
6.5	Sensitivity results . . . . .	11
6.6	Possible extensions of ideas . . . . .	13
<b>7</b>	<b>Planned cooperation arrangement</b>	<b>13</b>
<b>8</b>	<b>Work and plan</b>	<b>14</b>
<b>9</b>	<b>Ethical, safety-related or regulatory aspects</b>	<b>17</b>
<b>10</b>	<b>Sex-specific and gender related issues</b>	<b>17</b>
<b>11</b>	<b>Human resources</b>	<b>17</b>
<b>Annex 1</b>	<b>References</b>	<b>20</b>
<b>Annex 2</b>	<b>Research institution and required funding</b>	<b>24</b>
<b>Annex 3</b>	<b>CV of the principal investigator</b>	<b>26</b>

# 1 Science

The large hadron collider (LHC) at CERN has provided a large volume of data till 2018 and is expected to deliver a larger amount of data in the near future to be compared with theoretical predictions. During the Run 2 (2015–2018) era of LHC, CMS experiment [1] has collected proton-proton collision data corresponding to an integrated luminosity of  $\sim 140 \text{ fb}^{-1}$  corresponding to  $\sim 10^{16}$  collisions and has been performing an extensive set of analyses to extract the maximum information from those. A few key milestones LHC reached in last ten years are the following:

- Discovery of Higgs boson (H) in 2012 [2, 3] resulting in the completion of the standard model (SM) of particle physics. This is followed by precision measurements of Higgs boson properties, such as mass [4, 5], CP nature [6, 7], coupling to Gauge bosons, third-generation fermions, and second-generation leptons [8], among others.
- Precision measurements of processes involving electroweak Gauge bosons (W, Z,  $\gamma$ ), light- and heavy-flavor quarks (u, d, s, c, b, t), gluons (g).
- Discovery of processes not observed previously in colliders, such as light-by-light scattering [9], single-production of a top quark in association with a Z boson [10], vector boson scattering [11], among others.
- Constraining a large region of parameter space in models beyond the SM (BSM) constructed with extended symmetries constructed to explain several observed phenomena and to provide an explanation of theoretical quests.

The lack of a smoking gun evidence of BSM physics predicting new particles within the LHC energy range has motivated the particle physicists to look for subtle deviations in measured distributions of observables from the SM predictions and by systematic field-theoretic extension of the SM Lagrangian adding higher-dimensional operators [12–14]. Higher-dimensional operators, in general, modify the couplings between different particles and result to a smooth change in the shape of particular distributions as compared to SM predictions in precision measurements. Precision measurement of Higgs boson couplings, particularly to the vector bosons (V), is extremely important in pinning down the structure of the extended SM Lagrangian. For example, the existence of a heavy Gauge boson of mass beyond the LHC reach can manifest itself modifying H-V couplings [15]. The second most dominant mechanism of Higgs boson production at LHC is in association with a vector boson, referred to as VH production, which involves a  $s$ -channel vector boson. The VH production with V decaying to leptons is an ideal place to probe H in its dominant decay mode to a pair of b quarks as the leptons can be used to select the targeted events and also to reduce the multijet background from quantum chromodynamics (QCD) interaction. The production of vector boson from quark-antiquark annihilation and the decay of H to a pair of b quarks are sensitive to the footprint of new physics at respective vertices, namely vector couplings and Gauge couplings at VH production and Yukawa coupling in  $H \rightarrow b\bar{b}$  decay. A number of such couplings involving the Higgs boson can be directly probed for

the first time at LHC [16]. The rest of the couplings, which could have left signatures in earlier experiments like LEP, can also be probed with better accuracy at LHC because of the increase of their impact with energy [17, 18].

The coupling measurements can already be performed with Run 2 LHC data. With the addition of  $\sim 160 \text{ fb}^{-1}$  data to be delivered by LHC in Run 3 (2022–2024), these measurements are expected to have a significant leap in precision. Therefore, I propose performing an extensive measurement of the couplings relevant in VH production, where V decays leptonically and H decays to a pair of b quarks, using differential distributions of a number of observables, some of which have not been measured before.

## 2 Research hypothesis

### 2.1 Effective field theory couplings in $V (\rightarrow \ell \ell) H (\rightarrow b \bar{b})$ production

Standard model effective field theory (SMEFT), a generalized extension of the SM, consists of all the possible operators of dimensions greater than four, satisfying the symmetries in SM [19–23].

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{c_i^{(5)}}{\Lambda} \mathcal{L}_5^i + \sum_i \frac{c_i^{(6)}}{\Lambda^2} \mathcal{L}_6^i + \dots \quad (2.1)$$

In Eq. (2.1),  $\Lambda$  denotes the cut-off scale which sets an upper limit of the validity of the theory,  $c$ 's represent dimensionless numbers, referred to as Wilson coefficients, capturing short-range physics at an energy scale above  $\Lambda$ , and  $\mathcal{L}_{5,6}$  are the operators at dimension-5 and -6, respectively. The only possible operator at dimension-5, known as Weinberg operator [24], violates the lepton number conservation and generates Majorana mass to neutrinos [25]; the smallness of neutrino mass points to the relevance of this operator only at energy much higher than that relevant for LHC physics. The number of independent dimension-6 operators that do not violate baryon and lepton number is known to be 2499, which can be brought down to 76 assuming the maximum flavor symmetry,  $U(3)^5$ , thereby ignoring differences between different fermion generations [20]. The complete set of non-redundant operators at dimension-6, except for the Hermitian conjugates, written for the first time in 2010 forms the so-called Warsaw basis [26]. SMEFT operators at dimension-6 modify the production and decay kinematics of the Higgs boson in VH production compared with those predicted by the SM, which results in differences in shapes of numerous observables constructed using the particles measured in the experiment [27–30].

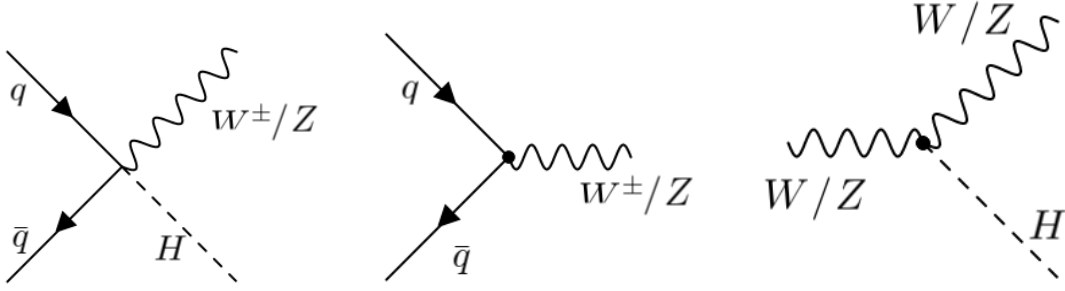
List of dimension-6 operators in the Warsaw basis that affects the VH production are listed in Table 1. First four operators in the left column of Table 1 introduce 4-point interactions, shown in Fig. 1 (left). The same operators along with  $\mathcal{O}_{HD}$  and  $\mathcal{O}_{HWB}$  affect the coupling of V to fermions, shown in Fig. 1 (middle). The remaining operators modifies H–V coupling as shown in Fig. 1 (right).

### 2.2 Direct measurements of SMEFT couplings in $V (\rightarrow \ell \ell) H (\rightarrow b \bar{b})$

SMEFT parameters can, in general, be constrained using data from the previous low-energy experiments by replacing H by its vacuum expectation value ( $v$ ) in operators in

**Table 1.** Dimension-6 operators in the Warsaw basis affecting the VH production.

$\mathcal{O}_{Hq}^{(1)} = iH^\dagger \overleftrightarrow{D}_\mu H \bar{q} \gamma^\mu q$	$\mathcal{O}_{HWB} = H^\dagger \sigma^a H W_{\mu\nu}^a B^{\mu\nu}$
$\mathcal{O}_{Hq}^{(3)} = iH^\dagger \sigma^a \overleftrightarrow{D}_\mu H \bar{q} \sigma^a \gamma^\mu q$	$\mathcal{O}_{H\tilde{W}B} = H^\dagger \sigma^a H W_{\mu\nu}^a \tilde{B}^{\mu\nu}$
$\mathcal{O}_{Hu} = iH^\dagger \overleftrightarrow{D}_\mu H \bar{u}_R \gamma^\mu u_R$	$\mathcal{O}_{HW} =  H ^2 W_{\mu\nu} W^{\mu\nu}$
$\mathcal{O}_{Hd} = iH^\dagger \overleftrightarrow{D}_\mu H \bar{d}_R \gamma^\mu d_R$	$\mathcal{O}_{H\tilde{W}} =  H ^2 W_{\mu\nu}^a \tilde{W}^{a\mu\nu}$
$\mathcal{O}_{HD} = (H^\dagger D_\mu H)^* (H^\dagger D_\mu H)$	$\mathcal{O}_{HB} =  H ^2 B_{\mu\nu} B^{\mu\nu}$
$\mathcal{O}_{H\Box} = (H^\dagger H) \Box (H^\dagger H)$	$\mathcal{O}_{H\tilde{B}} =  H ^2 B_{\mu\nu} \tilde{B}^{\mu\nu}$



**Figure 1.** Fragments of the Feynman diagram for VH production sensitive to different dimension-6 operators.

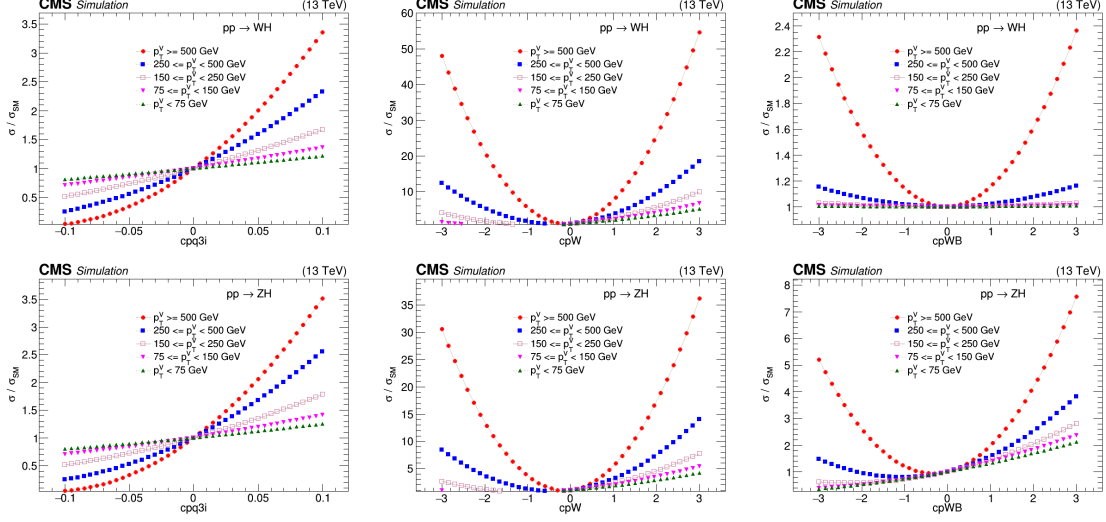
Table 1. For a majority of the operators at work here, which are of the form  $H^2 \mathcal{L}_{\text{SM}}$ , this just corresponds to a shift of the parameters in SM Lagrangian. Therefore, those can be probed directly for the first time at LHC.

Now, the operators at Table 1 can also be constrained using the existing measurements in Higgs sector at LHC [8, 31, 32]. However, this methodology assumes that all the object and event selection conditions have the same efficiency and acceptance for the SM and SMEFT operators, which is known to be not the case always. Also, as will be shown in Sec. 6, some important effects can be lost in traditional measurements of kinematic variables. Therefore, it is necessary to construct a set of observables that can extract telltale signatures of EFT operators without any loss of information. Also, the impact of all the conditions applied at the analysis level needs to be checked for the operators involved. These goals can only be achieved in a dedicated measurement as proposed here.

### 3 Objectives

The goal of this proposal is to probe the operators listed in Table 1 in VH production as precisely as possible. Dependence of the production cross section of VH process on the Wilson coefficients corresponding to  $\mathcal{O}_{Hq}^{(3)}$ ,  $\mathcal{O}_{HW}$ , and  $\mathcal{O}_{HWB}$ , denoted as cpq3i and cpW,

respectively, in different regions in vector boson transverse momentum ( $p_T$ ) are shown in Fig. 2 for WH and ZH productions. These three operators, along with their corresponding



**Figure 2.** Variation of cross section of WH (top) and ZH (bottom) productions relative to the SM prediction as a function of Wilson coefficients corresponding to  $\mathcal{O}_{Hq}^{(3)}$  (left),  $\mathcal{O}_{HW}$  (middle), and  $\mathcal{O}_{HWB}$  (right) in five regions of W and Z boson  $p_T$ , respectively.

CP-odd counterparts, and  $\mathcal{O}_{HD}$ ,  $\mathcal{O}_{H\Box}$  affect both WH and ZH productions, whereas  $\mathcal{O}_{Hq}^{(1)}$ ,  $\mathcal{O}_{Hu}$ ,  $\mathcal{O}_{Hd}$ ,  $\mathcal{O}_{HB}$ , and  $\mathcal{O}_{H\tilde{B}}$  affect only ZH production because of the helicity structure involved in  $q\bar{q} \rightarrow Z, W$  interaction. The SM cross section for WH production, where W decays to either a muon or an electron is roughly 20 times larger than ZH production, where Z decays to a pair of electrons and muons, so the former results in better statistical precision in the measurements. Therefore, measurement of couplings in both WH and ZH processes is necessary to probe the complete set of operators relevant to VH production as also shown in Ref. [33]. A preliminary study to test the sensitivity to certain operators is presented in Sec. 6.

#### 4 Scientific originality

Although the SMEFT operators can be indirectly probed using results from the previous experiments, LHC gives a unique opportunity to directly measure a number of SMEFT operators, particularly those involving the Higgs field [16, 34]. So far, there exist very few measurements where all the efficiency and acceptance effects of object and event selection conditions are taken into account, and complete information contained in an event is used. The proposed measurement will extend the scope of probing new physics scenarios using previously unexplored angular observables. A list of new things aimed to be developed and applied during the proposed measurements are the following:

- The complete set of event information, parameterized by kinematic and angular variables, will be used for the first time in VH production followed by  $V \rightarrow \ell \ell$  and

$H \rightarrow b\bar{b}$ . The newly proposed methods which can be applied in this context are ‘method of moments’ [33, 35] and boosted information tree [36]. The latter one is developed by the group where the proposer belongs to. Development of the methods and applying those in a CMS measurement by the same group will greatly accelerate the preparation of the results.

- Deep neural network based algorithms are used to identify  $H$  decaying to a pair of  $b$  quarks, and efficiency of the tagger is conventionally measured in data and simulation in  $g \rightarrow b\bar{b}$  process. However,  $g$  and  $H$  are very different in mass, spin, and color properties. During the proposed work, an effort will be pursued to improve the calibration of the tagger, particularly using  $Z \rightarrow b\bar{b}$  decay.

## 5 Relevance to the research field

The idea of EFT has prevailed in the context of particle physics since as late as the 1930s with the Fermi theory of beta decay [37] and Yukawa theory of meson exchange [38]. Those guided the particle physics community through a long and adventurous journey, which has led to the development of the SM in its present form. The lack of convincing hints of new physics at LHC has automatically shifted the focus of the community’s interests towards EFT as the future step [39–41] and measurements of higher-dimensional operators has started to take the center stage of LHC physics program [42–44]. The interests in both theory and experiment communities are growing at an increasing rate, which can be gauged by the formation of the LHC EFT working group [45] and the conduction of a series of lectures under the ‘All things EFT’ forum [46]. So, it is a call of time to perform dedicated EFT measurements at LHC. The proposed measurement not only probes a set of unique EFT operators in the Higgs sector but also complements the ongoing measurements of EFT operators in the area of electroweak bosons and top quark and enhances the physics potential of LHC experiments on a broad scale. For example, measurements of the vector couplings involving light quarks nicely complement the measurement of the similar couplings involving top quarks and they together can probe the flavor universality in quark sector. The results of these measurements will constitute the output from LHC to be used in a global EFT fit, combining results of many different experiments. Therefore, it is high time for the group to perform the proposed measurement and strengthen its position in the international community.

## 6 Method

Achieving goals in time requires setting up a concrete plan in advance. The target is to utilize the data available for physics analysis from LHC Run 2, corresponding to an integrated luminosity of  $139 \text{ fb}^{-1}$  at  $\sqrt{s} = 13 \text{ TeV}$ , and also the data to be recorded during LHC Run 3, which is expected to be  $\sim 160 \text{ fb}^{-1}$  at around the same  $\sqrt{s}$ . With  $\sim 45\%$  of the total dataset already in the disk, the strategy is to finalize the analysis procedures with Run 2 data and publish the first set of results, which will be followed by a final paper using data to be collected by 2024.

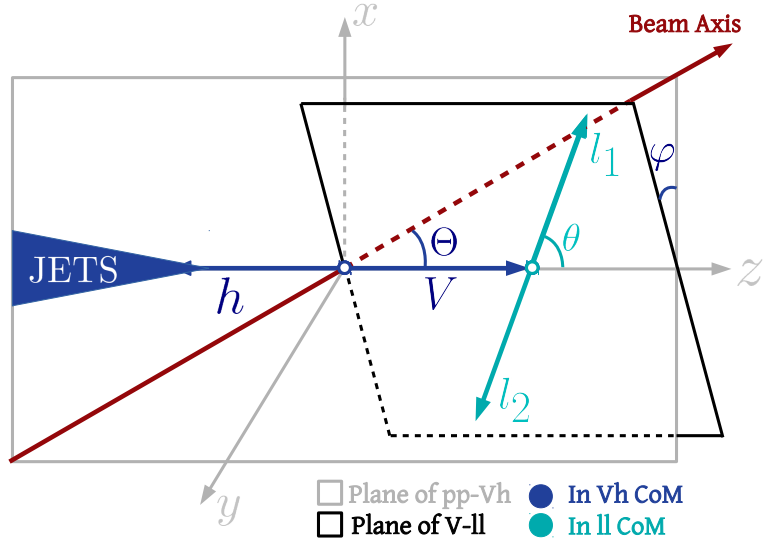
In the following, a brief discussion of the analysis strategy to achieve goals mentioned in Sec. 3 is made. The feasibility study reported here is performed focusing primarily on WH production, nevertheless, also some additional features available in ZH are mentioned.

### 6.1 Final states and observables

The following final states are considered.

- $W (\rightarrow \ell \nu) H (\rightarrow b \bar{b})$
- $Z (\rightarrow \ell^+ \ell^-) H (\rightarrow b \bar{b})$

As mentioned in Sec. 2 and 3, operators listed in Table. 1 introduce an energy growth, which will be visible in distributions of traditional kinematic observables, like  $p_T$  of vector boson or Higgs boson. However, only a few kinematic observables are not enough to disentangle the effects of a relatively large number of SMEFT operators. A systematic analysis of VH production process starting from helicity amplitudes, as performed in Ref. [33], enables to learn about angular variables, shown in Fig. 3, which are complementary to the kinematic variables exhibiting energy growth, and they together fully characterize the event structure. The helicity amplitudes corresponding to longitudinal



**Figure 3.** Pictorial depiction of the angles describing the final state in  $V (\rightarrow \ell\ell H (\rightarrow PbPAb)$  production.

and transverse polarizations of the vector boson contributing to the full amplitude of VH production are affected differently by the operators involved. Thus, rich information about the event structure can be obtained by a joint analysis of kinematic and angular variables. In a traditional cross section measurement integrated over angular variables, effects of the interference between helicity amplitudes are lost, and those can be resurrected using the angular information [47].

## 6.2 Event selection and categorization

Lepton(s) from W or Z boson is (are) used to trigger the events to be analyzed. CMS trigger system, developed centrally within the Collaboration, is used to select events with at least one lepton ( $\ell$ ) satisfying a threshold in  $p_T$ . At very high  $p_T$  of the vector boson, backup triggers requiring a considerable hadronic activity can be used particularly for the final state with electron(s). Measurement of efficiency of the triggers both data and simulation is an essential step of the analysis to understand the modeling of the selected events in data by simulation. Tag-and-probe method in  $Z \rightarrow \ell^+ \ell^-$  events will be used to measure the trigger efficiencies. Corrections will be derived in simulation to match the trigger efficiency in data. For the feasibility study, it is assumed that trigger efficiencies are the same in data and simulation.

An event to be selected as a WH candidate event is required to have an electron or a muon with  $p_T > 32$  and  $> 25$  GeV, respectively passing the identification conditions corresponding to 90 and 95% efficiency for genuine electron and muon, respectively. The four-momentum of neutrino is reconstructed using the missing transverse momentum vector ( $\vec{p}_T^{\text{miss}}$ ) and assuming that the invariant mass of neutrino and lepton system is always the value of W boson mass provided by the particle data group. This will be discussed further in Sec. 6.3. The W boson reconstructed from lepton and neutrino momenta is required to satisfy  $p_T > 150$  GeV, which brings down the contamination from multijet production to a negligible extent. The lepton and  $\vec{p}_T^{\text{miss}}$  are required to satisfy  $\Delta\phi > 2$  radian, where  $\Delta\phi$  is the angular separation in the plane transverse to the collision axis. The event is also required not to contain any additional electrons or muons with  $p_T > 25$  GeV satisfying the identification requirement specified earlier.

The selected events are further categorized into regions depending on the jet system. If H has a large  $p_T$ , its decay products are merged into a large-sized jet. In this case, referred to as boosted topology, jets reconstructed using anti- $k_T$  algorithm with distance parameter 0.8, referred to as AK8 jets are used to identify H. Otherwise, H is reconstructed using AK4 jets initiated by its daughter b quarks; this is referred to as resolved topology. In CMS, DNN-based taggers, henceforth referred to as DeepJet and DeepAK8 taggers, are used to identify AK4 jets initiated by b quarks [48] and AK8 jets initiated by H decaying to a pair of b quarks [49], respectively. DeepAK8 tagger actually provides multiple outputs that correspond to the probability of AK8 originating from top quark, W boson, light quark or gluon, etc.; this information is also used later.

In resolved topology, the system of two AK4 jets with highest b tagging scores, referred to as b1 and b2, is taken as the H candidate, whereas the AK8 jet with the highest H tagging score is taken as the H candidate in boosted topology. Signal region (SR), enriched by  $W (\rightarrow \ell \nu) H (\rightarrow b\bar{b})$  events, is constructed using the conditions listed in Table 2. In Table 2,  $\text{btag}_{\text{max}}^{\text{cut}}$  and  $\text{btag}_{\text{min}}^{\text{cut}}$  represent thresholds on DeepJet score corresponding to 1% and 10% mistag rate for light quark or gluon jets, respectively, and  $\text{Htag}^{\text{cut}}$  corresponds to threshold on H tagging score corresponding to 1% mistag rate for the same background. Changing the conditions on variables in Table 2, one also gets control regions (CRs) enriched by different background processes. For example, requiring no. of additional jets (no. of b-tagged jets



**Table 2.** Selection conditions targeted for WH events.

Resolved category	Boosted category
$\geq 2$ b-tagged AK4 jets ( $p_T > 30$ GeV, $ \eta  < 2.5$ )	$\geq 1$ AK8 jets ( $p_T > 250$ GeV, $ \eta  < 2.5$ )
B tagging score of b1 $> \text{btag}_{\text{max}}^{\text{cut}}$	H tagging score of H $> \text{Htag}^{\text{cut}}$
B tagging score of b2 $> \text{btag}_{\text{min}}^{\text{cut}}$	
$< 2$ additional AK4 jets	No. of b-tagged AK4 jets outside H = 0
$M(b1 + b2) \in [90, 150]$ GeV	Soft-drop mass of H $\in [90, 150]$ GeV

outside the H candidate)  $\geq 2$  in the resolved (boosted) category enables to select a phase space enriched by the events with pair production of top quarks ( $t\bar{t}$ ). The phase space one gets inverting the mass condition on the H candidate in the SR is dominated by vector boson in association with jets initiated by b and c quarks ( $W + b/c$ ). Finally, events failing the condition of the presence of two b-tagged AK4 jets (one Higgs-tagged AK8 jet) are dominated by the production of vector boson in association with light quark jets ( $W+j$ ). These CRs can be used to measure the corresponding backgrounds and validate those using data.

Distribution of W boson  $p_T$  in resolved and boosted categories of the SR are shown in Fig. 4, where the change in WH signal with the operators  $\mathcal{O}_{Hq}^{(3)}$  and  $\mathcal{O}_{HW}$  turned on with the corresponding Wilson coefficients set to 1 are also overlaid. Fig. 4 shows that the main backgrounds to WH signal are events from  $t\bar{t}$  and  $W + b/c$  productions.

The shape of the distribution of angular variables is also sensitive to some of the dimension-6 operators. As shown in Fig. 5, the SM production dominated by longitudinal polarization of W boson has some differences w.r.t the case when  $\mathcal{O}_{HW}$  operator is turned on since it causes interference between amplitudes with longitudinal and transverse polarizations of W boson. Variable  $\phi$ , on the other hand, is very sensitive to the CP-nature of the operators, and the difference in its modulation between CP-even and CP-odd operators is clearly visible in ZH production. CP-sensitivity in  $\phi$  is reduced in WH production due to the ambiguity in neutrino reconstruction as will be shown in Sec. 6.3 and the impact of other particles in the process on  $p_T^{\text{miss}}$ ; this needs a dedicated study in future.

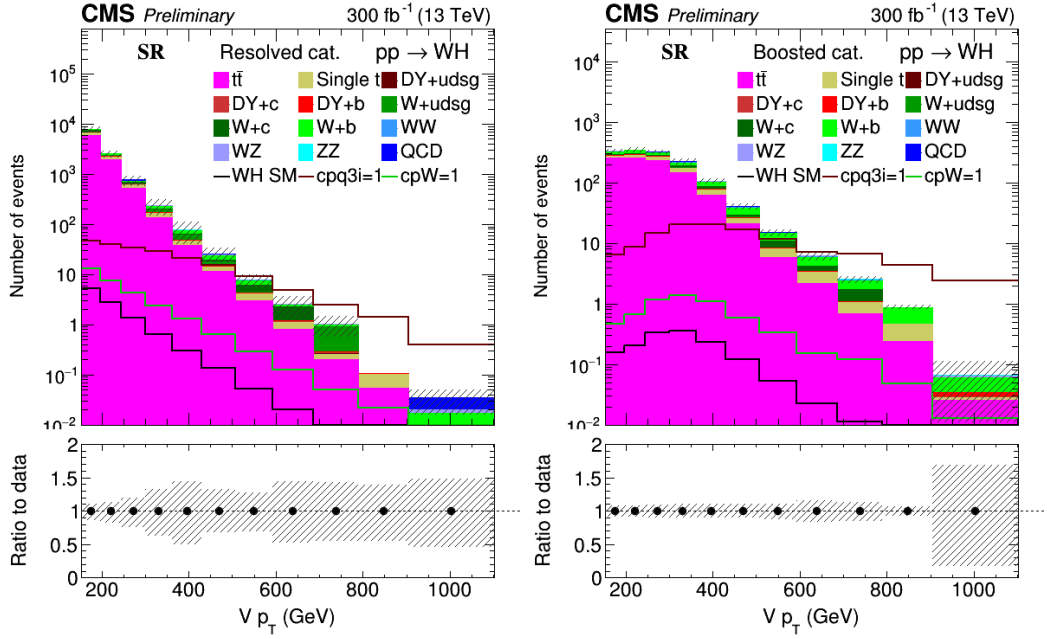
### 6.3 Neutrino reconstruction

In order to reconstruct the momentum of invisible neutrino in  $W \rightarrow \ell\nu$ , we assume that the  $\nu$  is the only source of  $p_T^{\text{miss}}$  in the event and take its  $p_T$  and  $\phi$  as the same of  $\nu$ . Finally, we use W-mass constraint to calculate  $\eta$  of  $\nu$ , denoted as  $\eta^\nu$ . However, the quadratic nature of mass constraint gives rise to two possible solution for  $\eta^\nu$  as follows.

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

where  $\Delta^2 = \frac{m_W^2 - m_T^2(p^l, p^{\text{miss}})}{2p_T p_T^{\text{miss}}}$ . When W boson has large  $p_T$ , which results to  $\Delta \ll 1$ , these two solutions become

$$\eta^\nu \simeq \eta^l \pm \sqrt{2}\Delta + \mathcal{O}(\Delta^3), \quad (6.2)$$



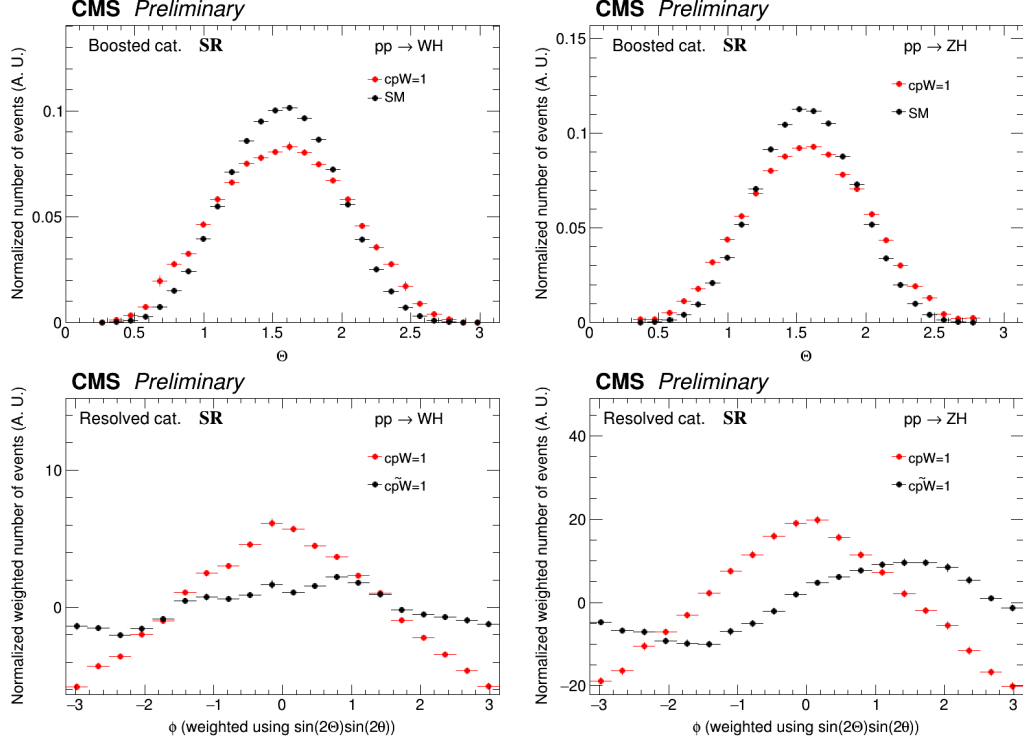
**Figure 4.** Distribution of W boson  $p_T$  in resolved (left) and boosted (left) categories of the SR. Impact of systematic uncertainties on jet energy scale and resolution, correction factor used to match the b-tagging and Higgs-tagging efficiencies in data and simulation are shown in the bottom panel.

In this limit, the angular variable  $\theta$  and  $\Theta$  computed with these two solutions converge to the same values, whereas the variable  $\phi$  does not. It can be shown that  $\phi$  computed with two solutions are related by the  $\phi_+ \simeq \pi - \phi_-$ . This ambiguity is visible in Fig. 6.

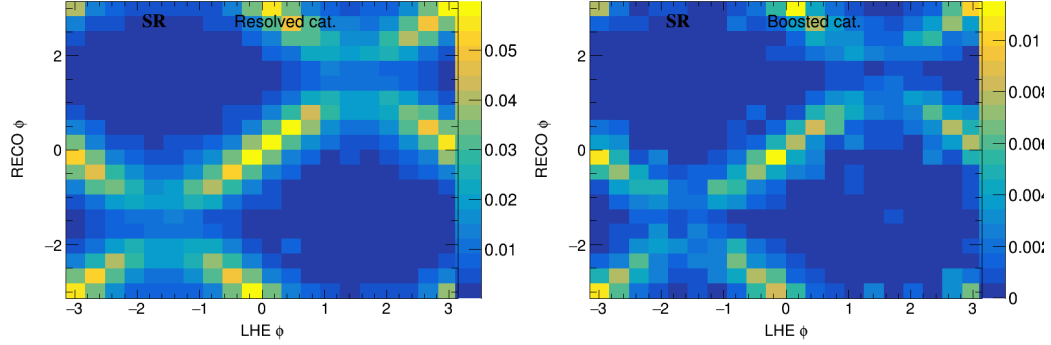
The ambiguity in  $\phi$  reduces the CP-sensitivity in WH production. A dedicated study needs to be performed to overcome this limitation.

#### 6.4 Multivariate analysis

As seen from Fig. 4, the region SR has a large number of background events. To separate the WH signal even more from the background events, particularly due to  $t\bar{t}$  production, a multivariate (MVA) discriminator is constructed using a DNN with two fully connected dense layers. Input variables to the DNN in the resolved category are listed in Table 3. A similar discriminator is built in the boosted category using the variables in the left column of Table 3 along with the following variables corresponding to the H candidate AK8 jet: the output of DeepAK8 taggers quantifying the probability of the jet originated due to top quark, Higgs boson, W boson relative to the same originated from a light quark or gluon. Normalized distributions of MVA discriminators in SM WH and  $t\bar{t}$  events are shown in Fig. 7. Resistive operator characteristics (ROC) curves are obtained by computing the fraction of signal and background events varying thresholds on the MVA discriminators.



**Figure 5.** Normalized distribution of polar angle  $\Theta$  in WH (top left) and ZH (top right) productions for  $cpW = 1$  and SM scenarios in boosted category. Normalized distribution of azimuthal angle  $\phi$  in WH (bottom left) and ZH (bottom right) productions for  $cpW = 1$  and  $cp\bar{W} = 1$ , respectively; SM component is subtracted from in these distributions.



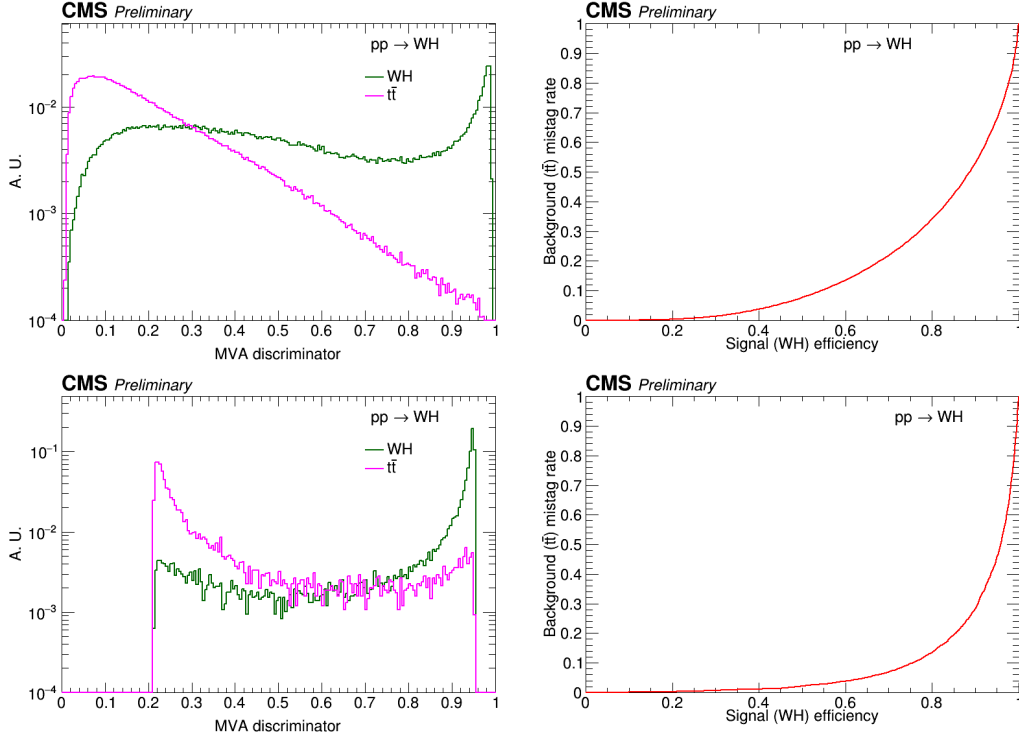
**Figure 6.** Correlation between  $\phi$  calculated at parton level and detector level in resolved (left) and boosted (right) categories in WH production.

## 6.5 Sensitivity results

Although the final condition on the MVA discriminator needs to be optimized, we go on checking the sensitivity by choosing a threshold on the MVA discriminator, which corresponds to 80% signal efficiency; this corresponds  $\sim 30\%$  ( $\sim 15\%$ ) background efficiency

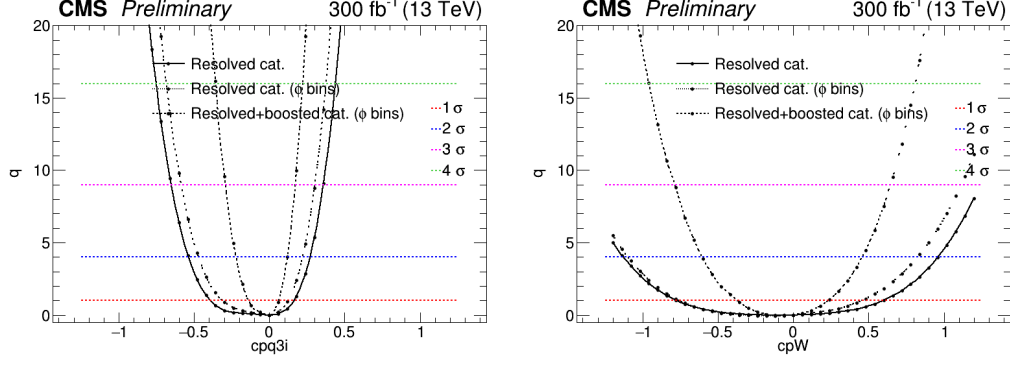
**Table 3.** Input variables to the DNN designed to separate events from WH and  $t\bar{t}$  productions.

$p_T$ component of $\vec{p}_T^{\text{miss}}$	No. of AK4 jets ( $p_T > 30 \text{ GeV}$ , $ \eta  < 2.5$ )
$p_T$ and $\eta$ of the lepton	B tagging scores of b1 and b2
$\Delta\phi(\ell, \vec{p}_T^{\text{miss}})$	Ratio of the $p_T$ of b1 and b2
No. of additional jets	Minimum of the $p_T$ of b1 and b2
Event thrust (see definition in [50])	Maximum of the $p_T$ of b1 and b2
Rapidity and mass of H candidate	Maximum of relative $p_T$ of the lepton w.r.t. b1 and b2
$\Delta\phi(W, H)$	$\Delta\phi(b1, b2)$



**Figure 7.** Normalized distributions of the MVA discriminator in signal (WH) and background ( $t\bar{t}$ ) events (left) and ROC curve characterizing the performance of MVA discriminator (right) in the resolved (top) and boosted (bottom) categories.

in the resolved (boosted) category. For simplicity, we use the efficiency numbers for different event samples as the first step and perform one-dimensional likelihood scans for two Wilson coefficients corresponding to operators  $\mathcal{O}_{Hq}^{(3)}$  and  $\mathcal{O}_{HW}$ , respectively, while setting others to 0. The results are shown in Fig. 8 for the integrated luminosity of  $300 \text{ fb}^{-1}$ . Fig. 8 shows that a large gain in sensitivity is obtained by adding the boosted category due to the energy growth induced by the operators. The preliminary results show that the Wilson coefficient corresponding to  $\mathcal{O}_{Hq}^{(3)}$  can be constrained to percent level and the one corresponding to  $\mathcal{O}_{HW}$  within  $\sim [-0.4, 0.25]$  at 95% confidence level. An improvement due to a binning in



**Figure 8.** Variation of the profile likelihood ratio  $q$  as a function of Wilson coefficients corresponding operators  $\mathcal{O}_{Hq}^{(3)}$  (left) and  $\mathcal{O}_{HW}$  (right) for an integrated luminosity of  $300 \text{ fb}^{-1}$ .

$\phi$  is also observed in the resolved category. The use of information about all three angles  $\Theta, \theta, \phi$  is expected to result in a large increase in sensitivity.

## 6.6 Possible extensions of ideas

The research plan proposed here can also be extended in multiple directions, such as the following.

- The constraints obtained from the proposed measurements can also be reinterpreted as bounds on new physics models of interest, for example, on models predicting the existence of a  $Z'$  and  $W'$  bosons in  $Z' \rightarrow ZH$  and  $W' \rightarrow WH$  decays, respectively. Particularly, new physics models with  $Z'$  is in highlights in recent times to explain the anomalies in the flavor sector.
- Here, the proposal is to probe SMEFT operators in  $VH$  production, where  $V$  decays to leptons. But, the methodology developed also holds for hadronically decaying  $V$ ; the only bottleneck 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.

## 7 Planned cooperation arrangement

The work will be conducted within the CMS Collaboration. The reconstructed and simulated data, trigger information, object calibration factors provided centrally by the Collaboration will be used. Ideas and updates will be presented at regular intervals in internal meetings. In general, no cooperative arrangement is planned for this proposal.

However, there are synergies with ongoing activities within the CMS data analysis group at HEPHY. The group recently submitted a paper on measurement of  $t\bar{t}\gamma$  process and constraining  $t-\gamma$  coupling in SMEFT; this work was supported by FWF grant P31578. There is also ongoing work on constraining SMEFT operators in  $tWZ$  process supported by

FWF grant P33771. Thus the proposed measurement will add a complementary direction in terms of the group’s activities.

The CMS data analysis group at HEPHY consists of five staff scientists, three post-doctoral researchers, and five Ph.D. students working on supersymmetry, long-lived signatures, search for BSM Higgs boson decaying to a pair of  $\tau$  leptons, top quark physics, among others. Diverse experience in the group helps to develop new ideas and implement those in the CMS Collaboration.

## 8 Work and plan

The work will be carried out by a new Ph.D. student along with the applicant. Since the proposed work is on Higgs physics, which is one of the main centers of activities at LHC, the Ph.D. student will get high visibility within the CMS Collaboration.

Tasks in chronological order are described in the following and summarized in Table 4.

### 1. Signal simulation for Run 2:

The proposed study requires the simulation of VH events, including the effects of SMEFT operators both at the levels of production and decay using recently developed software packages, e.g., SMEFTsim [51] at leading-order or using SMEFT@NLO [52] at next-to-leading order in perturbative QCD. Particular care needs to be taken for the matching of generation of the hard process using MADGRAPH5\_aMC@NLO [53] and the parton shower with PYTHIA8 [54] since the parton shower does not involve higher-dimensional operators.

### 2. SMEFT parameterization:

Information about SMEFT operators will be simulated using a second-order polynomial in the Wilson coefficients involved. A number of weights will be stored for each event. It needs to be checked if a second-order polynomial is sufficient to parameterize the effects of SMEFT operators.

### 3. Optimization of lepton selection:

The conditions used for lepton identification needs to be optimized in such a way that the variation of acceptance for different operators and the total systematic uncertainty are minimized while not compromising on the statistical power of data.

### 4. Optimization of b tagging condition:

The CMS Collaboration provides the correction to be applied to simulation to match the efficiency of DNN-based b in data and simulation in two variants. The first one, the working point-based method, requires to use certain pre-defined thresholds on b tagging score of a jet to qualified a 'b-tagged' and the corresponding correction has smaller uncertainties. In the second approach, the shape-based method, one can put the threshold at any value in b tagging discriminator and the corrections have larger uncertainties. Since the b tagging information is used for signal-to-background

discrimination, it needs to be checked which of the two options provides better overall sensitivity.

5. Optimization of H tagging condition and calculation of correction:

The threshold on the DNN-based H tagging condition needs to be optimized since this is required for signal-to-background separation. Correction factors need to be appropriately calculated to match the tagging efficiency in data and simulation.

6. Trigger efficiency measurement:

The efficiency of lepton-based triggers needs to be measured in both data and simulation. Corrections to match the trigger simulation and data will be derived and applied to the simulation. The same needs to be performed for the backup jet triggers.

7. Neutrino reconstruction:

A careful analysis is required to assign the correct  $\eta$  to the neutrino, which also benefits the CP-sensitivity of the measurements in case of WH production. A powerful prescription developed for neutrino reconstruction will also help similar analyses ongoing in CMS.

8. Background estimation:

In the proposed measurement, the backgrounds are primarily planned to be taken from simulation with a thorough validation by comparing those with data in dedicated control regions. Corrections need to be derived wherever needed, and the corresponding systematic uncertainties need to be quantified.

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 of the MVA discriminator need to be optimized. The modeling of the MVA discriminator for different background processes will be checked, and data-to-simulation corrections will be derived if needed.

10. SMEFT to SM separation:

In the feasibility study, the constraints on Wilson coefficients are derived just by separating events in different regions. A more sophisticated study with larger number of variables and taking into account their correlation needs to be used. This will be a test-bed of a method proposed by the group at HEPHY using boosted decision trees [36].

11. Experimental systematic uncertainties:

Systematic uncertainties related to different objects, e.g. leptons, 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.

12. Modeling uncertainties:

Modeling uncertainties in simulation from the choice of renormalization and factorization scales, parton distribution function, parton shower need to be derived and applied to all the predictions taken from simulation. The neutrino reconstruction is also expected to be affected by the modeling uncertainties.

13. Quantification of sensitivity:

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 correlation between the operators. The correlation between systematic uncertainty sources in different years needs to be properly defined. It will be carefully checked if some of the systematic uncertainties are unexpectedly constrained.

14. SMEFT effects in the background:

So far, all the discussions regarding the SMEFT 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 boson equivalence theorem 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 SMEFT 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 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 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 datasets will be combined. Final sensitivity on Wilson coefficients will be reported based on the combined dataset.

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 sole aim of the project is to improve the human understanding of nature. Thus the topic itself does not have any gender-related aspects. The job advertisement will be formulated gender neutrally. In order to improve the gender balance at HEPHY ( $\sim 20\%$ ), a suitable female candidate will be given preference.

## 11 Human resources

A Ph.D. student will be recruited solely for this project to be conducted at HEPHY, Vienna. The Ph.D. student will carry out the project under the supervision of the project applicant, PI Dr. S. Chatterjee. According to the plan, the PI will spend 50% of his time for 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 professional theorists and published papers in international journals.

Tasks will be carried out in consultation with the leader of CMS data analysis group of the institute: Dr. Robert Schöfbeck (RA), who is also performing SMEFT measurements

in top quark physics in association with a post-doctoral researcher Dr. Dennis Schwartz (DS). Cross-talk with RA and DS will be extremely useful for this project. Particularly, the simulation of SMEFT operators (task 1), SMEFT parameterization (task 2), and SMEFT to SM separation (task 10) will be performed together with RA and DS. RA is also the former convener of the 'top quark mass and properties' physics analysis group and the 'jet and MET' physics object group of the CMS Collaboration. His feedback on documentation and presentation of the results will be constructive.

Task / Year	2022 ( $\int \mathcal{L} dt \sim 30 \text{ fb}^{-1}$ )				2023 ( $\int \mathcal{L} dt \sim 120 \text{ fb}^{-1}$ )				2024 ( $\int \mathcal{L} dt \sim 160 \text{ fb}^{-1}$ )			
	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec	Jan-Mar	Apr-Jun	Jul-Dec	
1. Signal simulation for Run 2	✓											
2. SMEFT 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. SMEFT to SM separation					✓							
11. Exp. systematic uncertainties					✓							
12. Modeling uncertainties												
13. Quantification of sensitivity						✓						
14. SMEFT 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 Ph.D. student, ✓ by the project applicant and Ph.D. student, ✓ by the project applicant and Ph.D. student in collaboration with other members in HEPHY CMS data analysis group.

## Annex 1 References

- [1] **CMS** Collaboration, S. Chatrchyan et al., *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) S08004.
- [2] **ATLAS** Collaboration, G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1, [[arXiv:1207.7214](#)].
- [3] **CMS** Collaboration, S. Chatrchyan et al., *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, [[arXiv:1207.7235](#)].
- [4] **CMS** Collaboration, A. M. Sirunyan et al., *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, [[arXiv:1706.09936](#)].
- [5] **CMS** Collaboration, A. M. Sirunyan et al., *A measurement of the Higgs boson mass in the diphoton decay channel*, *Phys. Lett. B* **805** (2020) 135425, [[arXiv:2002.06398](#)].
- [6] **CMS** Collaboration, A. M. Sirunyan et al., *Constraints on anomalous HVV couplings from the production of Higgs bosons decaying to  $\tau$  lepton pairs*, *Phys. Rev. D* **100** (2019), no. 11 112002, [[arXiv:1903.06973](#)].
- [7] **CMS** Collaboration, A. M. Sirunyan et al., *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, [[arXiv:2003.10866](#)].
- [8] **CMS Collaboration** Collaboration, *Combined Higgs boson production and decay measurements with up to 137 fb<sup>-1</sup> of proton-proton collision data at  $\sqrt{s} = 13$  TeV*, CMS Physics Analysis Summary CMS-PAS-HIG-19-005, 2020.
- [9] **CMS** Collaboration, A. M. Sirunyan et al., *Evidence for light-by-light scattering and searches for axion-like particles in ultraperipheral PbPb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV*, *Phys. Lett. B* **797** (2019) 134826, [[arXiv:1810.04602](#)].
- [10] **CMS** Collaboration, A. M. Sirunyan et al., *Observation of Single Top Quark Production in Association with a Z Boson in Proton-Proton Collisions at  $\sqrt{s} = 13$  TeV*, *Phys. Rev. Lett.* **122** (2019), no. 13 132003, [[arXiv:1812.05900](#)].
- [11] **CMS** Collaboration, A. M. Sirunyan et al., *Observation of electroweak production of same-sign W boson pairs in the two jet and two same-sign lepton final state in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *Phys. Rev. Lett.* **120** (2018), no. 8 081801, [[arXiv:1709.05822](#)].
- [12] B. Grinstein and M. B. Wise, *Operator analysis for precision electroweak physics*, *Phys. Lett. B* **265** (1991) 326.
- [13] 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, [[arXiv:0712.0396](#)].
- [14] G. Passarino and M. Trott, *The Standard Model Effective Field Theory and Next to Leading Order*, [[arXiv:1610.08356](#)].
- [15] T. Appelquist and J. Carazzone, *Infrared Singularities and Massive Fields*, *Phys. Rev. D* **11** (1975) 2856.

- [16] R. S. Gupta, A. Pomarol, and F. Riva, *BSM Primary Effects*, *Phys. Rev. D* **91** (2015), no. 3 035001, [[arXiv:1405.0181](#)].
- [17] J. Ellis, V. Sanz, and T. You, *The Effective Standard Model after LHC Run I*, *JHEP* **03** (2015) 157, [[arXiv:1410.7703](#)].
- [18] C. Grojean, M. Montull, and M. Riembau, *Diboson at the LHC vs LEP*, *JHEP* **03** (2019) 020, [[arXiv:1810.05149](#)].
- [19] 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, [[arXiv:1308.2627](#)].
- [20] 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, [[arXiv:1312.2014](#)].
- [21] 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, [[arXiv:1310.4838](#)].
- [22] C. Englert and M. Spannowsky, *Effective Theories and Measurements at Colliders*, *Phys. Lett. B* **740** (2015) 8–15, [[arXiv:1408.5147](#)].
- [23] I. Brivio and M. Trott, *The Standard Model as an Effective Field Theory*, *Phys. Rept.* **793** (2019) 1, [[arXiv:1706.08945](#)].
- [24] S. Weinberg, *Baryon- and lepton-nonconserving processes*, *Phys. Rev. Lett.* **43** (Nov, 1979) 1566.
- [25] F. Bonnet, D. Hernandez, T. Ota, and W. Winter, *Neutrino masses from higher than  $d=5$  effective operators*, *JHEP* **10** (2009) 076, [[arXiv:0907.3143](#)].
- [26] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, *Dimension-Six Terms in the Standard Model Lagrangian*, *JHEP* **10** (2010) 085, [[arXiv:1008.4884](#)].
- [27] K. Hagiwara, R. Szalapski, and D. Zeppenfeld, *Anomalous Higgs boson production and decay*, *Phys. Lett. B* **318** (1993) 155–162, [[hep-ph/9308347](#)].
- [28] J. Ellis, V. Sanz, and T. You, *Complete Higgs Sector Constraints on Dimension-6 Operators*, *JHEP* **07** (2014) 036, [[arXiv:1404.3667](#)].
- [29] C. W. Murphy, *Statistical approach to Higgs boson couplings in the standard model effective field theory*, *Phys. Rev. D* **97** (2018), no. 1 015007, [[arXiv:1710.02008](#)].
- [30] J. Baglio, S. Dawson, S. Homiller, S. D. Lane, and I. M. Lewis, *Validity of standard model EFT studies of  $VH$  and  $VV$  production at NLO*, *Phys. Rev. D* **101** (2020) 115004, [[arXiv:2003.07862](#)].
- [31] **ATLAS** Collaboration, G. Aad et al., *Measurements of  $WH$  and  $ZH$  production in the  $H \rightarrow b\bar{b}$  decay channel in  $pp$  collisions at 13 TeV with the ATLAS detector*, *Eur. Phys. J. C* **81** (2021) 178, [[arXiv:2007.02873](#)].
- [32] **ATLAS** Collaboration, G. Aad et al., *Measurement of the associated production of a Higgs boson decaying into  $b$ -quarks with a vector boson at high transverse momentum in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *Phys. Lett. B* **816** (2021) 136204, [[arXiv:2008.02508](#)].

- [33] S. Banerjee, R. S. Gupta, J. Y. Reiness, S. Seth, and M. Spannowsky, *Towards the ultimate differential SMEFT analysis*, *JHEP* **09** (2020) 170, [[arXiv:1912.07628](#)].
- [34] 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, [[arXiv:1308.1879](#)].
- [35] S. Banerjee, R. S. Gupta, O. Ochoa-Valeriano, M. Spannowsky, and E. Venturini, *A fully differential SMEFT analysis of the golden channel using the method of moments*, *JHEP* **06** (2021) 031, [[arXiv:2012.11631](#)].
- [36] S. Chatterjee, N. Frohner, L. Lechner, R. Schöfbeck, and D. Schwarz, *Tree boosting for learning EFT parameters*, [arXiv:2107.10859](#).
- [37] E. Fermi, *An attempt of a theory of beta radiation. 1.*, *Z. Phys.* **88** (1934) 161.
- [38] H. Yukawa, *On the Interaction of Elementary Particles I*, *Proc. Phys. Math. Soc. Jap.* **17** (1935) 48–57.
- [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, [[arXiv:1803.03252](#)].
- [40] J. Ellis, M. Madigan, K. Mimasu, V. Sanz, and T. You, *Top, Higgs, Diboson and Electroweak Fit to the Standard Model Effective Field Theory*, *JHEP* **04** (2021) 279, [[arXiv:2012.02779](#)].
- [41] J. J. Ethier, G. Magni, F. Maltoni, L. Mantani, E. R. Nocera, J. Rojo, E. Slade, E. Vryonidou, and C. Zhang, *Combined SMEFT interpretation of Higgs, diboson, and top quark data from the LHC*, [arXiv:2105.00006](#).
- [42] CMS Collaboration, A. M. Sirunyan et al., *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, A. Tumasyan et al., *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, A. Tumasyan et al., *Measurement of the electroweak production of  $Z\gamma$  and two jets in proton-proton collisions at  $\sqrt{s} = 13$  TeV and constraints on anomalous quartic gauge couplings*, [arXiv:2106.11082](#).
- [45] L. P. C. at CERN, “Lhc effective field theory wg.”
- [46] *All things eft*, 2021.
- [47] G. Panico, F. Riva, and A. Wulzer, *Diboson interference resurrection*, *Phys. Lett. B* **776** (2018) 473, [[arXiv:1708.07823](#)].
- [48] E. Bols, J. Kieseler, M. Verzett, M. Stoye, and A. Stakia, *Jet flavour classification using DeepJet*, *JINST* **15** (2020) P12012, [[arXiv:2008.10519](#)].
- [49] CMS Collaboration, A. M. Sirunyan et al., *Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques*, *JINST* **15** (2020) P06005, [[arXiv:2004.08262](#)].
- [50] CMS Collaboration, V. Khachatryan et al., *Study of Hadronic Event-Shape Variables in Multijet Final States in  $pp$  Collisions at  $\sqrt{s} = 7$  TeV*, *JHEP* **10** (2014) 087, [[arXiv:1407.2856](#)].
- [51] I. Brivio, *SMEFTsim 3.0 — a practical guide*, *JHEP* **04** (2021) 073, [[arXiv:2012.11343](#)].

- [52] C. Degrande, G. Durieux, F. Maltoni, K. Mimasu, E. Vryonidou, and C. Zhang, *Automated one-loop computations in the standard model effective field theory*, *Phys. Rev. D* **103** (2021) 096024, [[arXiv:2008.11743](#)].
- [53] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *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, [[arXiv:1405.0301](#)].
- [54] T. Sjstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *An Introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159, [[arXiv:1410.3012](#)].

## 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 exploit 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. Schöffbeck, 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 Ph.D. students and two master students doing data analysis in CMS, as well as with the scientist of the New Physics theory group. Additionally, four members are based at CERN. Among them is Dr. W. Adam, the former physics coordinator of the CMS experiment and the leader of 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 for participation in meetings in the CMS collaboration.

This project application requests funding for one Ph.D. 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 Ph.D. 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 ( $\sim 20\%$ ), 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 Ph.D. student will travel four times a year (approximately every two months) to CERN for, on average, one week. The purpose of 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 Ph.D. student will get an 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. 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 travel to CERN is provided in Table 5.

The institute needs to pay an amount (MoE) 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



**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.** Total amount of requested funding.

Item	Cost
Ph.D. student's salary	EUR 119340
Ph.D. student's travel costs of MoE of the PI (for 3 years)	EUR 15465
	EUR 30000
General costs (5%)	EUR 8240
Total	EUR 173045

EUR 164805 is requested to cover other unexpected costs. Total requested funding sums up to EUR 173045. A summary of the total costs of the project is provided in [Table 6](#).

### **Annex 3 CV of the principal investigator**

#### **Personal Data**

**Age:** 27 years (as of February 1, 2021)

**Nationality:** Indian

**Sex:** Male

**Affiliation:** Institute of High Energy Physics (HEPHY), Vienna

**Designation:** Post-doctoral researcher

#### **Contact Details**

**Email:** s7384705218@gmail.com, suman.chatterjee@cern.ch

**Phone:** +91-7710835260, +43-6769394076

**Website:** <https://sumanchatterjeetifr.wordpress.com>

#### **Academic Records**

**2016-2020:** Ph.D. in Experimental Particle Physics

Department of High Energy Physics

Tata Institute of Fundamental Research, Mumbai 400005

Degree awarded in August 2020

**2014-2016:** Master of Science in Physics

Tata Institute of Fundamental Research, Mumbai 400005

Rank: 1st

**2011-2014:** Bachelor of Science in Physics

Department of Physics

Jadavpur University, Kolkata 700032

Remark: 1st class

## Achievements

- 2020** Received **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** Awarded **Professor Sukumar Biswas Ph.D. Student Award for Excellence in Physics** for scoring the highest grades in graduate courses in the Physics Integrated Ph.D. program in Tata Institute of Fundamental Research
- 2015** Qualified CSIR-UGC National Eligibility Test in Physics with **All India Rank 43**
- 2014** Secured **All India Rank 3** with 99.63 percentile in Joint Entrance Screening Test 2014 conducted by the research institutes in India
- 2011-2014** Awarded **INSPIRE SCHOLARSHIP** by Department of Science and Technology, India for being within top 1% students appeared from West Bengal Board of Higher Secondary Education in Higher Secondary (10+2) Examination
- 2011** Awarded by the Chief Minister of the state of West Bengal for securing **the 4<sup>th</sup> 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

SMEFT results in Higgs sector from the CMS experiment

**July 2021** European Conference on High Energy Physics (EPS-HEP 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 (QEIC), IIT Bombay, Mumbai, India

Plenary talk (on behalf of ATLAS, CMS, and LHCb Collaborations)

Understanding parton distribution function from LHC measurements

**Sep 2019** Workshop on Top Quark Physics (TOP2019), IHEP, Beijing, China

Invited talk (on behalf of ATLAS and CMS Collaborations)

Heavy resonance searches in ATLAS and CMS

**July 2019 European Conference on High Energy Physics (EPS-HEP 2019), Ghent, Belgium**

Parallel talk (on behalf of CMS Collaboration)

Inclusive jet results from CMS experiment

**July 2019 European Conference on High Energy Physics (EPS-HEP 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 (on behalf of CMS Collaboration)

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 (on behalf of ATLAS, CMS, and LHCb Collaborations)

Top quark and jet measurements sensitive to PDFs

**Aug 2018 QCD at LHC 2018, Technische Universität Dresden, Dresden, Germany**

Parallel talk (on behalf of CMS Collaboration)

Differential jet cross sections at the CMS experiment

**Dec 2017 Workshop on High Energy Physics Phenomenology XV, IISER Bhopal, Bhopal, India**

Parallel talk (on behalf of CMS Collaboration)

Radius Scan for Inclusive Jets in CMS Experiment at  $\sqrt{s} = 13$  TeV

**Dec 2016 XXII DAE-BRNS High Energy Physics Symposium, Delhi, India**

Parallel talk (on behalf of CMS Collaboration)

Perspectives for a Radius Scan in Inclusive Jets at 13 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  $\sqrt{s} = 13$  TeV

**Sep 2018 Asia-Europe-Pacific School on High Energy Physics, Quy Nhon, Vietnam**

Radius scan for inclusive jets at  $\sqrt{s} = 13$  TeV

**Dec 2016 XXII DAE-BRNS High Energy Physics Symposium, Delhi, India**  
HO weight factor in particle flow algorithm in CMS

### **Teaching Experience**

**Autumn 2019: Teaching assistant**

Course: Statistical Methods in Physics

Instructor: Prof. G. Majumder, TIFR

**January 2019: Co-coordinator**

Software session (lecture and hands-on exercise on algorithms used in high energy 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

**Autumn 2016: Teaching assistant**

Course: Electrodynamics II

Instructor: Prof. G. Ravindra Kumar, TIFR

### **Supervision Experience**

**Anirban Bala** (Ph.D. student at TIFR, Mumbai, Advisor: Prof. Monoranjan Guchait):  
During his departmental project I (July-September, 2020) under Prof. Gobinda Majumder at TIFR

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  $\sqrt{s} = 13$  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  $\sqrt{s} = 13$  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  $\sqrt{s} = 13$  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.*