

A Framework for Finding Anomalous Objects at the LHC

Amit Chakraborty^a, Abhishek M. Iyer^{b,c}, Tuhin S. Roy^{c,d}

^a*Theory Center, Institute of Particle and Nuclear Studies,
KEK, 1-1 Oho, Tsukuba, Ibaraki 305-0801, Japan*

^b*INFN-Sezione di Napoli, Via Cintia, 80126 Napoli, Italia*

^c*Department of Theoretical Physics, Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Mumbai 400 005, India*

^d*Theory Division T-2, Los Alamos National laboratory, Los Alamos, NM 87545, USA*

Abstract

Search for new physics events at the LHC mostly rely on the assumption that the events are characterized in terms of standard-reconstructed objects such as isolated photons, leptons, and jets initiated by QCD-partons. While such strategy works for a vast majority of physics beyond the standard model scenarios, there are examples aplenty where new physics give rise to anomalous objects (such as collimated and equally energetic particles, decays due to long lived particles etc.) in the detectors, which can not be classified as any of the standard-objects. Varied methods and search strategies have been proposed, each of which is trained and optimized for specific models, topologies, and model parameters. Further, as LHC keeps excluding all expected candidates for new physics, the need for a generic method/tool that is capable of finding the unexpected can not be understated. In this paper, we propose one such method that relies on the philosophy that all anomalous objects are *not* standard-objects. The anomaly finder, we suggest, simply is a collection of vetoes that eliminate all standard-objects up to a pre-determined acceptance rate. Any event containing at least one anomalous object (that passes all these vetoes), can be identified as a candidate for new physics. Subsequent offline analyses can determine the nature of the anomalous object as well as of the event, paving a robust way to search for these new physics scenarios in a model-independent fashion. Further, since the method relies on learning only the standard-objects, for which control samples are readily available from data, one can build the analysis in an entirely data-driven way.

Keywords: Standard Model, BSM physics, LHC, Jets, Jet substructure, QCD

Email addresses: amit@post.kek.jp (Amit Chakraborty), iyera@na.infn.it (Abhishek M. Iyer),
tuhin@theory.tifr.res.in (Tuhin S. Roy)

1. Introduction

The discovery of the Higgs boson of the Standard Model (SM) of particle physics in the Large Hadron Collider (LHC) [1, 2], was believed to be a precursor towards the realization of non-standard physics at around the TeV scale. However, the analysis of all data from Run-I and Run-II so far have failed to yield any statistically significant excess over the SM expectations in any of the channels being looked at [3–8]. While it is highly likely that new physics (NP) is just around the corner and is going to show up as LHC keeps accumulating data, it is worthwhile to think through whether there remains gaps in aspects of our search strategies where events due to NP might show up and yet elude our grasp.

However, before proceeding further, let us deconstruct the general search strategy being employed at the LHC. Broadly speaking, at the detector level events due to collisions are recorded in terms of the charged tracks observed at the trackers and the muon spectrometers, energies deposited at different cells of the electromagnetic calorimeters (namely, ECAL), and the hadronic calorimeters (namely, HCAL). The CMS collaboration of the LHC employs a sophisticated **particle-flow algorithm** [9] which combines all this information and generates outputs as a set of 4-vectors, which are then classified into objects such as electrons, muons, photons, charged and neutral hadrons. Note that, these particle-flow objects, even though carry names of the particles, should still be treated as detector objects since further processing is required before one can start the process of identifying the physics of short-distance that might have given rise to the event.

The *detector-objects* (either tracks and calorimeter cells or even particle-flow objects) are the inputs to a series of algorithms and techniques that are used to obtain the *reconstructed objects* such as isolated photons, electrons, muons, taus and jets¹. An event is now described in terms of these ‘standard’ reconstructed objects along with some variables that carry the global detector information such as, missing energy, H_T etc. Standard phenomenological studies to search for NP as well as SM physics at the LHC use this information.

The above-mentioned strategy works fairly well for the SM and a large fraction of NP physics processes. However, the fundamental assumption that *all* NP events can be described in terms of these reconstructed objects is not true. Take, for example, reconstructed photons – these are outputs of an algorithm which identifies a cluster of ECAL energy depositions to be a photon if the pattern of energy deposits is consistent with the shower of a photon in the calorimeter [10, 11]. However, it is not implausible to imagine a NP scenario which gives rise to *only* collimated photons (known as photon-jets [12–24]) instead of single photons, where the degree of collimation is less than the size of a reconstructed photon. In this case, the photon-finder algorithm, trained on the samples of showers from single photons may not find any photon in the event. As a result, either we completely miss the event or, at best, the event gets classified as an event consisting of QCD-jets. Photon jets are not the only example – one can again find such examples where NP gives rise to events consisting of ‘anomalous’ or ‘non-standard’ objects, such as collimated electrons (or, electron-jets [25, 26], or, say lepton jets in general [27–35]), collimated taus (or tau-jets [36, 37]), particle with large life-times (e.g., long lived particles [38–48]), etc. to name a few.

Several methods and search strategies have been proposed, trained, and optimized to find many of these scenarios by identifying these anomalous objects. An essential problem remaining is that these strategies are powerful when it comes to finding specific NP models and topologies for which the searches have been optimized, but lose sensitivity fast, when models/topologies/parameters are varied. In other words, no general framework exists to probe and trigger these events with anomalous objects at the LHC. In this paper we attempt to provide one such general framework that can be used to select (and store) these events containing anomalous objects (equivalently signatures of NP) for further physics analysis.

The framework proposed here relies on the broad definition of anomalous objects, as objects that are *not* standard such as photons, electrons, taus, or QCD-jets. The philosophy is, therefore, straightforward –

¹These jets, often understood to be initiated by hard partons from short distance physics, which undergoes further showering and hadronization are usually thought to be synonymous with ‘QCD-jets’.

understand the standard-objects enough to be able to veto these at a desired level of efficiency. The objects that pass through these series of vetoes are, therefore, anomalous. The working principle can be briefly summarized as follows:

1. First, we find reconstructed-objects by clustering all the calorimeter information, using a *single* algorithm and a *single* set of clustering parameters (this conforms with the philosophy first proposed in Ref. [16, 17]). The output then becomes the superset of all standard as well as anomalous objects. Additionally, we demand that these outputs satisfy certain **hardness criteria**, which ensures that these objects can not be resultants from noise only.
2. Using a set of **judiciously chosen variables**, we find representations of these reconstructed-objects in a multi-dimensional space. By training **MultiVariateAnalyses** (MVAs) we identify patches in this multi-dimensional space occupied by the standard objects (namely, single photons, single electrons, single tau (hadronic), and QCD-jets).
3. Finally, we construct vetoes that simply block these patches rich in standard objects. In quantitative terms, these vetoes require ‘target-rates’, defined as the rates at which standard-objects will be acceptable. For example, if one sets the target-rate for QCD-jets to be 1%, this in turn determines the veto-boundary such that only 1% of QCD-jets can pass it.
4. Objects that pass through these vetoes are then identified as anomalous objects. Events containing at least one anomalous object become candidates for events due to NP and need to be recorded. One can look at the multidimensional representation of an anomalous object (offline) to learn about the object itself (such as whether it contains collimated photons, or it corresponds to long-lived objects, etc.). Coupled with the event information (such as the number, the nature, and the kinematic features of the accompanying objects in the event), one can then identify whether the event arises from NP or from SM.

The crucial feature of this strategy is that the whole exercise relies on **knowing standard objects**, such as single photons, single electrons, single taus, QCD-parton initiated jets etc., for which we have ample data that can work as control samples. Therefore, the entire formalism can be easily turned into a data-driven exercise, even though, in this paper we rely on Monte Carlo in order to demonstrate its working principle. Furthermore, this framework has plenty of rooms to improve, since it offers flexibility in terms of easily including new variables. We also emphasize that, even though, standard objects such as isolated photons, electrons, etc. can be subsets of outputs in the first step, we are not proposing any new method/changes in the way these standard objects are identified currently. Rather, we propose that this procedure be implemented in parallel to current strategies, and be used only to identify the presence of anomalous objects in the event.

The paper is organized as follows: in Sec.2 we outline the working principle and the philosophy of the proposed framework; in Sec.3 we discuss an ensemble of jet-variables that we employ in order to construct the veto; in Sec.4 we demonstrate the construction of vetoes, using responses of carefully constructed MVAs; in Sec.5 we give examples, where anomalous objects manage to pass these vetoes at acceptable rates (in particular, we give examples of collimated photons, electrons, and taus) even though vetoes did not use any information pertaining to these anomalous objects; and finally in Sec.6 we conclude.

2. The philosophy and the Framework

As mentioned in the introduction, the aim of this paper is to construct a *tool* or a *methodology* that can identify an “anomalous” or “non-standard” object, where the adjective anomalous or non-standard refers to the fact that the chance for the chosen object to be a standard object (such as e, γ, τ or QCD-jet) is highly unlikely (statistically speaking). The fundamental feature of the tool that we attempt to build is that it can be designed/optimized in an entirely data-driven procedure, even though in this work we use Monte Carlo in

order to construct a complete proposal as well as to demonstrate its efficacy. This constraint is non-trivial, since we can not expect to have controlled samples of anomalous objects available at the LHC.

The aim of this section is to discuss the philosophy of this paper along with its blueprint. This lays the groundwork before we move on and describe the procedures in detail in the following sections.

2.1. A universal framework for analyzing all objects

A difficulty arises while implementing such an analysis is the fact that the “standard-objects” are reconstructed objects. Even though the experimental analysis reconstructs these using the same detector elements such as calorimeter cells and tracks, or more refined objects such as particle flow elements, however varied reconstruction algorithms and/or parameters are used to find different objects. This makes a direct comparison among different reconstructed objects somewhat ambiguous. A robust analysis needs a universal construct for all objects “standard” or “non-standard”, built from calorimeters and trackers. In this work we implement a formalism as proposed in Ref. [16, 17]. The key ingredient is that one uses ‘jets’, defined as the output of a standard Infrared (IR) safe jet algorithm, to be the common construct for all physics objects that deposit energy in the calorimeters.

Note that the formalism adopted here maintains a clear distinction between the terminology of ‘jets’ and ‘QCD-jets’. We define ‘jets’ as the output of IR safe jet algorithms such as anti- k_T [49], k_T [50, 51], or C/A [52], which, in some instances, may have nothing to do with partons in QCD. A jet, therefore, becomes a generic concept that is defined in terms of the energy deposits in calorimeter cells and is identified by a jet algorithm. With this definition a QCD-jet is simply a special kind of jet (or rather, a standard-jet). The set of jets, therefore, also includes clustered energetic cells due to a single photon, or an electron, or a tau.

Our next strategy would be to devise a set of chosen variables in order to identify/classify the jets into categories. The working principle behind this is simple: the variables pave a way to map a jet to a point in a multi-dimensional space; a potent set of variables can ensure that jets of different kinds cluster in different corners in the space; as a consequence, by identifying these corners one can tag photons/electrons/taus/QCD-jets at the same time while minimizing the mistag rates due to jets of other kinds. It turns out that jet substructure techniques [53–55], developed to distinguish QCD-jets from jets containing boosted heavy particle decays by probing in detail the energy distribution within the jet, are ideal for this job. In fact, this treatment has been demonstrated to yield higher tagging efficiency for photons for the same mistag rate due to QCD-jets. Additionally, this method imparts the advantage of using grooming techniques [56–58] in photon tagging, making the tagging performance to be more pile-up robust.

Refs.[16, 17] also show that the same treatment can be used to find jets consisting of energetic and collimated photons (also known as photon-jets). Since kinematic features of the underlying physics (*e.g.*, the masses and spins of intermediary particles, whose decay give rise to these objects) are responsible for these distributions, the existence of structures within photon-jets is guaranteed. Substructure variables, therefore, should be efficient at finding and discriminating photon-jets from QCD-jets and even from single photons.

In this paper we use a slightly altered philosophy. In Refs.[16, 17], the authors rely on understanding photon-jets in order to separate these from single photons and QCD-jets. The analysis was more focused to obtain the best signal acceptance rate through performing a signal-background optimization procedure using several jet observables in a MVA framework. The analysis in Ref [17], like any other supervised learning, is extremely powerful in discriminating the photon-jets from QCD-jets. However, this technique quickly loses its discrimination power if, for example, photon-jets are replaced by ditau jets, or even use collimated photons but produced with different kinematics. Thus, the analysis of [16, 17], though extremely useful, uses knowledge on the type of NP and thereby limited to the specific new physics scenario under consideration. In this paper, however, we use a slightly altered philosophy; we follow the ‘unsupervised learning’ technique. We start with various standard objects (electron, photon, tau and QCD samples), while being completely agnostic of the type of NP, and go on understanding various properties of each of these standard objects. We then systematically construct vetoes to identify regions of phase space where the standard jets have small acceptance rate. As a

result, jets that escape these vetos, will, to a high probability, be considered as ‘non-standard’ objects, and corresponding events will be triggered as potential candidates for new physics events.

It is to be stressed that while constructing the vetoes only the known properties of the standard jets are used, no new physics input has been considered here. The proposed framework is thus less powerful compared to the one obtained after supervised learning that discriminates a specific kind of non-standard object from the standard objects, for example [16, 17], however, is more powerful in terms of its applicability in finding wide varieties of non-standard objects, and, therefore, can be used as a universal trigger for probing new physics signatures at the LHC.

2.2. Standard-jets

The second step towards constructing vetoes is to learn about the standard-jets. In this work we focus on four kinds of standard-jets, namely photons, electrons, taus (hadronic), and QCD-jets. The purpose of this subsection is to outline the operational definitions of these objects. The details of event generation, object reconstruction and the involved pile-up analysis are discussed in [Appendix A](#).

- **Photons:** We cluster the calorimeter responses for the events $pp \rightarrow h \rightarrow \gamma\gamma$ using anti- k_T jet algorithm for $R = 0.4$ and $p_T > 50$ GeV. From each event only the hardest jet, obtained after performing a pile-up subtraction, is selected. In order to create a pure sample of jets initiated by photons, we impose an additional consistency criterion using Monte Carlo (MC) truth. We check that the selected jet indeed contains at least one energetic photon inside. To be specific, there should be at least one MC photon within $\Delta R < 0.4$ from the jet axis, where the angular separation between two four-vectors is defined via $(\Delta R)^2 \equiv (\Delta\eta)^2 + (\Delta\phi)^2$. The quantities $\Delta\eta$ and $\Delta\phi$, refer to the differences in pseudo-rapidity and azimuthal angle of the two four-vectors respectively. From now on each jet of this sample will be known as a jet of type photon, or a jet initiated by a photon (or, often simply a photon or γ).
- **Electrons:** The simplest and the most practical choice is to use jets initiated by electrons from the decay $Z \rightarrow ee$. In this work, however, we use electrons from Monte Carlo sample where Higgs is being used as the intermediate particle in order to generate samples. We have explicitly checked that the distributions of the substructure variables we employ here remain identical irrespective of whether we use Z or h as the intermediate particle. To be specific, we cluster the calorimetric responses for the events $pp \rightarrow h \rightarrow ee$ using the anti- k_T jet algorithm with $R = 0.4$ and $p_T > 50$ GeV. We then select the leading jet, obtained after performing a pile-up subtraction, from each event as long as it also contains at least one MC electron within $\Delta R < 0.4$. We call jets from this sample to be a jet of type electron or a jet initiated by an electron (or, often simply an electron or e).
- **Taus:** Similar to the case of electrons, the most practical choice is to have jets initiated by taus from decays $Z \rightarrow \tau\tau$. However, we simulate the events $pp \rightarrow h \rightarrow \tau^+\tau^-$ with the τ decaying hadronically. The jets are then constructed using anti- k_T jet algorithm with jet radius $R = 0.4$ and $p_T > 50$ GeV. Similar to the earlier cases, the leading jet from each event, obtained after performing a pile-up subtraction, is selected as long as there is at least one MC tau within $\Delta R < 0.4$. We denote each jet from this sample as a jet of type tau or a jet initiated by tau (or, often simply as a tau or τ).
- **QCD-jets:** Hard QCD processes are simulated with a minimum p_T threshold of 50 GeV. Jets are then constructed from the calorimetric four-vectors using anti- k_T jet algorithm with $R = 0.4$ and $p_T > 50$ GeV. For each event, the leading (p_T ordered) jet obtained after performing a pile-up subtraction, is selected for further analysis. We require no further purity criteria for these jets. We denote the jets from this sample as jets of type QCD-jets or jets initiated by QCD-partons or simply QCD-jets or simply as j .

Before concluding this subsection, let us discuss two important issues: first, the choice of jet radius $R = 0.4$ and second, the use of Higgs boson as the intermediate particle. In a typical search for boosted massive

resonances, the jet radius R is chosen such that the resultant jet contains (almost) all the decay products of the resonance. The search strategy then needs to customize R by optimizing the discovery potential of the target resonance. The problem we are solving here is unconventional; we do not have any particular target resonance mass in mind. By aiming at those cases where the angular separation among the decay products is such that the standard techniques fail, we get a target R – namely R needs to be smaller than (or, at most equal to) the size of the standard reconstructed objects (~ 0.4). In fact, given the choice of the new physics model under consideration (see [Appendix A](#)), we find the choice of $R = 0.4$ includes all the decay products of the collimated objects, and therefore it’s already a very robust choice. Therefore, increasing R will not improve signal acceptance, however, it will necessarily increase hadronic contaminations of the underlying events and pile-ups. In such a case, QCD-jets need to be controlled separately to improve the sensitivity of the new physics. We use the Higgs scalar as the intermediate particle to generate standard-jets only for convenience. During implementation, we rather recommend the use of Z for generating electrons and taus. For example, leading jets in events with di-boson (namely, $ZZ \rightarrow 4e$) can be used to populate the electron sample.

2.3. From substructure variables to a veto

The next agenda on the list is to construct a veto for all standard objects by looking at these objects only. We do this in multiple stages:

1. Using a **carefully chosen set of variables**, a jet is mapped to a point in a multidimensional space. To elaborate, one can translate the statement such as “mass of a jet (say, J) is m_J ” to the statement that the variable mass maps $J \mapsto m_J$. Following the same logic, we use a set of variables $\{V_1, V_2, \dots, V_D\}$, to map each jet J to a set of numbers $\{v_1, v_2, \dots, v_D\}$. Assigning the jet J a vector of numbers $\vec{v} \equiv \{v_1, v_2, \dots, v_D\}$, one finds a representation of the jet J in the D -dimensional space.
2. We use Greek indices to denote the type of jets. In particular, if a standard-jet is designated as J_α , then α represents one of γ, e, τ or j . A set of variables, therefore, maps the i -th jet of kind α (namely, $J_{\alpha,i}$) to a representation $\vec{v}_{\alpha,i}$.
3. As we noted before, the variables are chosen in such a manner that one can simply find corners (or close regions) in the D -dimensional space where the standard-jets occupy and use D -dimensional boxes to isolate these samples. However, as D increases the analysis simply becomes tedious and less and less manageable.
4. In order to overcome the difficulty mentioned above, we incur a mechanism that maps the D -dimensional vector of numbers \vec{v} to a vector of fewer numbers, while still keeping jets of different types separated from each other. To be specific we use **MultiVariateAnalyses**, in particular, **Boosted Decision Tree** or BDT in the ROOT framework [59] (see [Appendix A](#) for BDT specific parameter details.). The process can be described as follows:
 - i. The input to a BDT is jets of two kinds with a set of variables that are, ideally, efficient in discriminating these two jets. As explained before, this set of variables give jets their representations. The job for the BDT is, therefore, to separate a list of jets of type α (or, the set of vectors $\{\vec{v}_\alpha\}$) from another list of jets of type β (or, the set of vectors $\{\vec{v}_\beta\}$).
 - ii. Broadly speaking, the **BDT optimizes the separation of jets**, by dividing the multidimensional space in many hyper-boxes, which are dominantly populated by jets of one kind in an algorithmic way. Now, given any point in this multi-dimensional space, a BDT can associate with it a response that is calculated based on the hyper-boxes that the point belongs to, as well as the purity contents of each box. Once a BDT is successfully trained to separate signals from backgrounds, it assigns large responses for signal-like jets whereas small responses to background-like jets. We denote a BDT treating jets of type α to be signal like, and jets of type β to be background like, by $\mathcal{B}^{\alpha/\beta}$ and

its responses by $r^{\alpha/\beta}$. We rescale the responses such that, the distribution for responses for jets of type- α (namely, $r_{\alpha}^{\alpha/\beta}$) peaks at large values (close to 1), whereas the same for jets of type- β (namely, $r_{\beta}^{\alpha/\beta}$) peaks at smaller values (close to 0).

- iii. Summarizing, a BDT optimized to separate jets of type α from type β (represented by $\mathcal{B}^{\alpha/\beta}$), maps any jet J (represented by a vector \vec{v}_J), to a response (a number) $r_J^{\alpha/\beta}$.

$$J \xrightarrow{\{V_1, V_2, \dots, V_D\}} \vec{v}_J \xrightarrow{\mathcal{B}^{\alpha/\beta}} r_J^{\alpha/\beta}. \quad (1)$$

As explained before, we expect $r_{\alpha}^{\alpha/\beta}$ close to 1, whereas $r_{\beta}^{\alpha/\beta}$ close to 0. There is no definite prediction for any other kind of jets (except that we expect it to be somewhere between 0 and 1).

The advantage of the above procedure is straightforward. Even if more and more variables are added to the existing set $\{V_1, V_2, \dots, V_D\}$, the jet still gets mapped to a single number for a BDT.

5. In this work, we end up using three BDTs ($\mathcal{B}^{j/\tau}$, $\mathcal{B}^{\gamma/\tau}$, and $\mathcal{B}^{\gamma/j}$), and therefore map all jets to a point in the $(r^{j/\tau}, r^{\gamma/\tau}, r^{\gamma/j})$ space. The entire procedure reduces a D -dimensional representation to a 3-dimensional representation without sacrificing information pertaining to pair-wise differences between the standard-jets.
6. As we show later, by construction, standard-jets occupy rather small corners in this space. Finally, after identifying bins in these three dimensions rich in standard-jets, we can veto most of standard-jets.

2.4. Summary

- We attempt to devise a tool which identifies anomalous objects, defined as the objects that are not the standard-objects such as electrons, photons, taus, and QCD-jets. The procedure therefore is synonymous to the construction of vetoes that block these objects.
- The fundamental problem in comparing all of these standard or anomalous objects is that we need a universal construct. For this purpose, we employ IR safe jet algorithms whose output (namely, jets) become the common construct. Electrons, photons, taus, and QCD-jets are therefore jets of specific types, so as all anomalous objects.
- We represent jets by points in a D -dimensional space spanned by outputs of D -number of jet variables. A judicious choice of variables is needed that emphasizes the differences among the jets of different types.
- For the vetoes to be effective, we need D to be large which makes the construction of vetoes hard. Increasing D , even by 1, only increases the difficulty associated with the procedure exponentially. We use MVAs (in particular, BDTs) that collapses D -dimensional representations to 3-dimensional representations of the responses. By construction, this reduction of dimensionality preserves information pertaining to pair-wise differences between the standard-jets.
- As a result, standard-jets get maximally separated from each other in this space. We block these corners rich in standard-jets to construct vetoes.

3. The Variables

In this section we describe the list of variables which can be useful in characterizing a jet of a given type. The variables are based on the tracker and calorimeter information, and also take into account the information associated to the constituents of the jets.

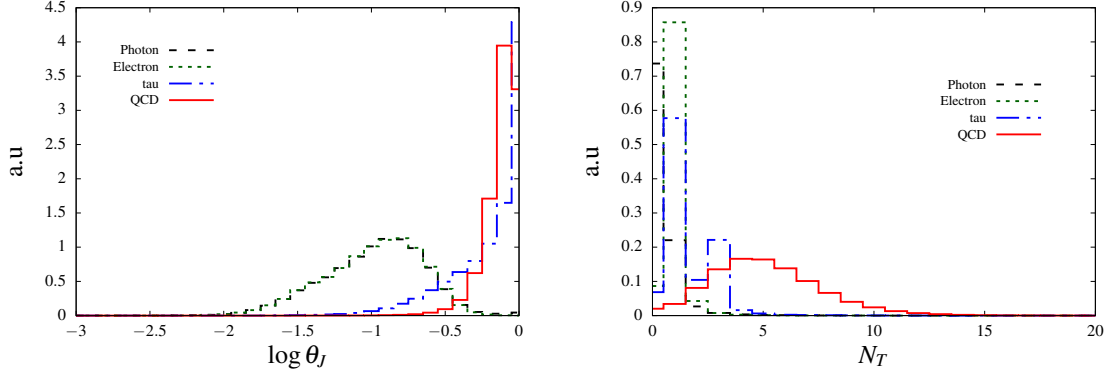


Figure 1: The distributions of Hadronic energy fraction (left) and the number of tracks (right) in the leading jet for the standard objects.

3.1. Hadronic energy fraction (namely, θ_J)

Since we construct jets from the calorimeter towers, calculating the hadronic energy fraction is particularly easy. Given a jet, we define its hadronic energy function from its constituents, which are calorimeter cells by definition.

$$\theta_J \equiv \frac{1}{E_J} \sum_{i \in \{J, \text{HCAL}\}} E_i \quad \text{where} \quad E_J \equiv \sum_{i \in J} E_i. \quad (2)$$

In the above definitions, the sum runs over all constituents of the jet. The total energy of the jet is therefore given by E_J .

The $\log \theta_J$ distributions for various kinds of background jets (or say ‘standard objects’) are shown in the left panel of Fig.1. As expected, θ_J peaks at 1 for τ -jets, since it dominantly decay to charged pions which deposit almost entire energy in the hadronic calorimeter. On the other hand, QCD-jets contain a significant number of neutral pions (1/3 on average because of isospin symmetry) which decay to pair of photons, and thus θ_J peaks at a smaller value. However, the electron and photon initiated jets deposit almost all their energy in the electromagnetic calorimeter leading to much smaller values of $\log \theta_J$. Not surprisingly, θ_J is widely used for providing pure samples of electrons and photons. Precise prediction of these distributions for standard objects helps us to understand and probe the presence, if any, of non-standard objects in an event. We are thus going to use this variable extensively in our analysis.

3.2. Tracks (namely, N_T)

The number of tracks associated to a jet is a measure of charged particle multiplicity inside a jet. Since the multiplicity of particles (charged or not) inside a jet is IR-unsafe, we set a lower p_T threshold and accept tracks which satisfy $p_T > 2$ GeV. The number of tracks in the leading jet is counted by calculating ΔR between the leading jet and each pile-up subtracted track. We then accept those tracks which satisfy $\Delta R < 0.4$, where $(\Delta R)^2 \equiv (\Delta \eta)^2 + (\Delta \phi)^2$ with $\Delta \eta$ and $\Delta \phi$ being the differences in pseudo-rapidity and azimuthal angle of the jet and the given track respectively.

The N_T distributions for each kind of background jets are shown in the right panel of Fig.1. A QCD jet or a jet initiated by colored partons (quarks or gluons) is mostly characterized by a large number of charged particles (*i.e.*, a large N_T). These charge particles are mostly hadrons, generated in the hadronization of partons after the initiating parton showers and split into multiple partons. In Fig.1 the N_T distribution is peaked around 5. Note that this value of peak is a function of the size of the jet (*i.e.*, the R parameter in jet clustering), and the minimum value of p_T of the tracks. The distribution moves to the right if R is increased

or if the cut on track p_T is lowered. Also note that the N_T distribution depends on the flavor of the parton initiating jets, and often are used for discriminating quark/gluon initiating jets [60–66].

Among the rest of the background jets, photons peaks at zero, while electrons and τ -jets dominantly peak around unity. The τ -jet samples also have a fair amount of three track events due to three charged pions. Because of conversion of photons into charged particles inside the tracker, some of the photons appear in $N_T = 1$ bin. We outline the details of photon conversions as implemented in our simulation in [Appendix A](#).

3.3. Energy-momentum distribution in subjets

In order to quantify the energy-momentum distributions among the subjets of a given jet, we recluster its constituents using k_T algorithm [50, 51] such that all constituent 4-vectors are combined and reproduces the original jet 4-vector. Even though, the final jet 4-vector remain the same, this procedure assigns the jet a new clustering history. Using this procedure of reclustering the constituents, one can assign a k_T ordered clustering history to any jet irrespective of the jet-algorithm used to find the jet. After reclustering, we obtain exclusive k_t -subjets. Of course, the number of exclusive subjets n_t is a free parameter. We then order these subjets according to their transverse momenta such that the subjet momenta follow the relation $p_{T_i} > p_{T_j}$ for $j > i$, with the 0-th subjet being the hardest. We primarily concentrate on two variables: the first one quantifies the fraction of the jet energy (or rather the p_T) carried by the leading subjet (namely, λ_J), while the second variable contains additional information of the next-to-leading as well as next-to-next-to-leading jets (namely, energy-energy correlation or ϵ_J).

$$\begin{aligned}\lambda_J &\equiv \log \left(1 - \frac{p_{T_0}}{p_{T_J}} \right) \\ \epsilon_J &\equiv \frac{1}{E_J^2} \sum_{n_f > i > j} E_i E_j\end{aligned}\tag{3}$$

where, as explained before, p_{T_i}, E_i is the transverse momentum and energy of the i -th subjet (ordered in p_T , such the 0-th subjet is the hardest); p_{T_J}, E_J is transverse momentum and energy of the given jet; and n_f is less or equal to the total number of exclusive subjets (n_t) of the given jet. In this work, following Ref. [16, 17], we ask for $n_t = 5$ and $n_f = 3$.

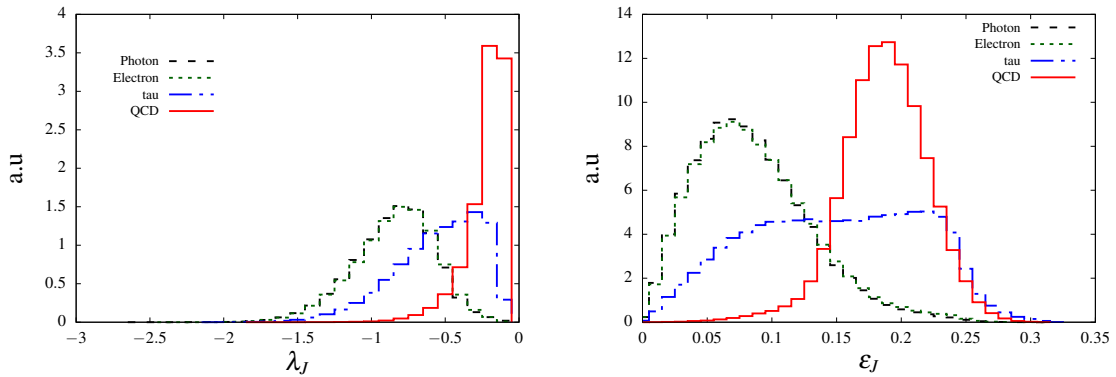


Figure 2: The distributions of λ_J (left) and ϵ_J (right) for the standard objects.

For a narrow pencil like (i.e., single prong) jet, the leading subjet carries most of the energy. For these jets one typically gets $p_{T_L} \lesssim p_{T_J}$, and consequently small λ_J and ϵ_J . To be specific, consider a jet consisting of n -number of energetic subjets, then by definition we have the following inequalities:

$$p_{T_0} \geq p_{T_J}/n \quad \implies \quad \lambda \leq \log \left(1 - \frac{1}{n} \right) \quad \implies \quad \lambda \rightarrow \begin{cases} -\infty & \text{as } n \rightarrow 1 \\ 0 & \text{as } n \rightarrow \infty \end{cases}, \tag{4}$$

where we have used the notations as used previously in Eq. (3). Note that for $n = 2, 3, 4, \dots$, one obtains $\lambda = -0.30, -0.18, -0.12, \dots$ respectively. As a result, a cut on λ is straightforward to understand and interpret. For example, for a jet with $\lambda > -0.30$, the leading subjet contains less than 50% of the total p_T . One can intuit from the above fact that the jet most likely contains at least two energetic subjets. Similarly, a jet with $\lambda > -0.18$, most likely will be characterized by three energetic subjets. Therefore, a cut $\lambda > -0.18$, for example, typically allows jets with three or more prongs.

Similar qualitative understanding can be obtained for ϵ_J . For example, if we assume the leading jet carries 90% of the jet energy, then the remaining 10% will be distributed among other subjets. In that case, the ϵ_J is expected to be around 0.08-0.09. However, if we assume that the energy distribution among the leading and two sub-leading jets are 50%, 30% and 20% of the total jet energy respectively, then we expect ϵ_J to be around 0.3. As the number of subjets increases with equal share of energies, ϵ_J increases. For e or γ initiated jets we expect the distributions of λ_J to be peaked at lower values than the QCD jets. Such intuitions are validated in Fig. 2, where we plot λ_J (left) and ϵ_J (right) for all the standard objects. From Fig. 2 and the discussion above, it is evident that λ_J and ϵ_J are qualitatively similar in describing the substructure of a given jet. A cut on λ_J can be mapped to a corresponding cut in ϵ_J , thereby exhibiting a strong correlation between the two.

3.4. N -subjettiness (namely, τ_N)

N -subjettiness [53] is a measure of the number of energetic subjets (or energy lobes) *inside a jet* as opposed to N -jettiness [67] which is an example of an event shape. We compute τ_N of the given jet using the definition in Ref. [53]. Given a set of N -axes, one defines

$$\tau_N \equiv \frac{\sum_k p_{T_k} \times \min(\Delta R_{1k}, \Delta R_{2k} \dots \Delta R_{Nk})}{\sum p_{T_k} \times R}, \quad \text{and} \quad \tau_{ab} \equiv \tau_a / \tau_b, \quad (5)$$

where k runs over the constituents of the jet, p_{T_k} is the transverse momentum for the k -th constituent, ΔR_{ak} is the angular distance between the k -th constituent and the a -th axis. Further, in order to calculate τ_N , one needs N -axes. In this work, we use axes collinear to the N exclusive k_t -subjets of the jets. Finally, Eq. (5) also gives the notation for the ratio of two N -subjettiness.

In order to understand the physics of N -subjettiness, consider for example a jet with l number of distinct lobes of energy. If one calculates τ_N as a function of N starting with $N = 1$, one finds that τ_N keeps decreasing with increasing N , with the rate of decrease maximized around $N = l$. The jet with l prongs, is then characterized by a large drop $\tau_{l-1} \gg \tau_l$. We can therefore use the ratio variable $\tau_{N(N-1)}$ to identify the energy distribution inside jet. In an ideal scenario, a jet with l prongs, will be given by a small $\tau_{N(N-1)}$ for $N = l$. We also find that it is often useful to consider the product of ratios $\tau_{a(a-1)} \times \tau_{b(b-1)}$, in order to isolate mixed samples containing primarily jets with a or b number of distinct prongs.

Out of various possible τ_N and the ratios τ_{ab} , we find τ_1 and τ_{31} particularly to be interesting. In the left panel of Fig. 3, we display the $\log(\tau_1)$ estimated for various background jets. Jets with energy distributed in a single and narrow prong (such an e or a γ initiated jet), is characterized by a small τ_1 , whereas jets with broader distributions of energy (such as jets due to QCD) will give rise to sizable τ_1 s. From the left panel one can also, see that τ initiated jets lie in-between the parton-initiated jets and the e/γ -initiated jets, since these are still “cleaner” than the qcd-jets. In fact some of the τ -initiated jets are characterized by a single pencil like distribution of energy as one sees with e/γ -jets. The τ - jets lie in between the e, γ and QCD jets as they either exhibit a 1- or 3- pronged structure. For the latter case, $\tau_1 \gg 0$ and thus has a reasonable overlap with the QCD jets. In the right plot of Fig. 3, we also show the distribution of $\tau_{31} = \frac{\tau_3}{\tau_1}$, which complements the $\log(\tau_1)$ distribution. Since a QCD jet exhibits a broader distribution of energy, it is likely to have multiple prongs inside the jet. As a result, τ_3 may not be significantly smaller to τ_1 . For the e, γ jets however, τ_3 is much smaller in comparison and is reflected in the plot. The τ - jets are characterized by $\tau_1(\tau_3) \rightarrow 0$ corresponding to a 1-(3-) pronged structure. Thus the ratio behaves similar to the pencil like jets of e, γ .

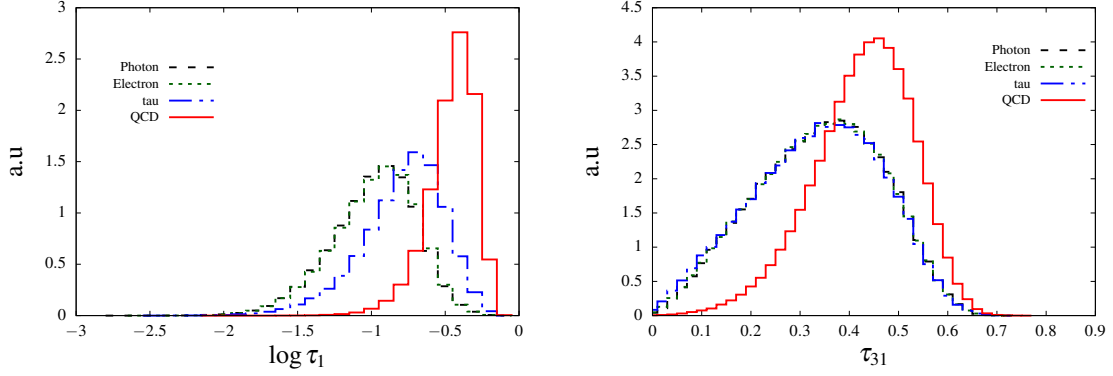


Figure 3: *Distribution of N -subjettiness variables $\log(\tau_1)$ (left) and τ_{31} (right) for all the standard objects.*

3.5. Energy Correlation functions and their ratios

Similar to N -subjettiness, energy correlation functions (namely e_N) also quantify the distribution of energy inside a jet. The key difference is that the N -subjettiness is constructed using the p_T of the constituents weighted by their angular distances from a set of axes, whereas in the definitions of e_N , the weighing parameters are the angles between the constituents themselves. In particular, we use the following [55, 68],

$$e_N = \sum_{i_1 < i_2 < \dots < i_N \in J} z_{i_1} z_{i_2} \dots z_{i_N} \left(\prod_{b=1}^{N-1} \prod_{c=b+1}^N \Delta R_{i_b i_c} \right)^\beta, \quad \text{where } z_i \equiv \frac{p_{T_i}}{p_{T_J}}. \quad (6)$$

In the equation above the sum runs over all constituents of the jet, and we assume the angular exponent (β) to be equal to unity. Note that in order to construct e_N we use dimensionless quantity z_i , which describes the fraction of the jet's transverse momentum carried by its i -th constituent. Consequently e_N is dimensionless. Also note that e_0 is taken to be equal to be 1. Additionally, we also use correlation ratios and double ratios (ratios of ratios):

$$r_N = \frac{e_{N+1}}{e_N} \\ C_N = \frac{r_N}{r_{N-1}} = \frac{e_{N+1} e_{N-1}}{e_N^2}. \quad (7)$$

Understanding the correlation functions and their ratios are straightforward. For a jet with n distinct pencil-like structures, it is clear that there can at maximum be n -number of subjets, where all are separated from each other by large angles. Therefore, e_{n+1} is suppressed w.r.t e_n . Both ratios and double ratios are sensitive to this fact. The double ratio, in fact, can be employed to measure the higher-order radiation from leading order substructure.

In Fig.4, we show the distribution of the two ECFs, namely e_2 and e_3 , while Fig.5 displays the distribution of the variables involving the ratios of the ECFs. As we have already discussed, for single prong objects like single electron, single photon and signal tau we expect both e_2 and e_3 to be sufficiently small. However, for QCD-jets, being multi-prong structure, both e_2 and e_3 can be large enough. Thus, we expect the distributions of $r_2 \equiv \frac{e_3}{e_2}$ will be shifted towards left for the e, γ, τ and right shifted for the QCD-jets. Similar behavior can be seen in C_2 , however the separation is not so significant as it involves a ratio $\frac{e_1}{e_2}$ which is comparable for all of these standard objects.

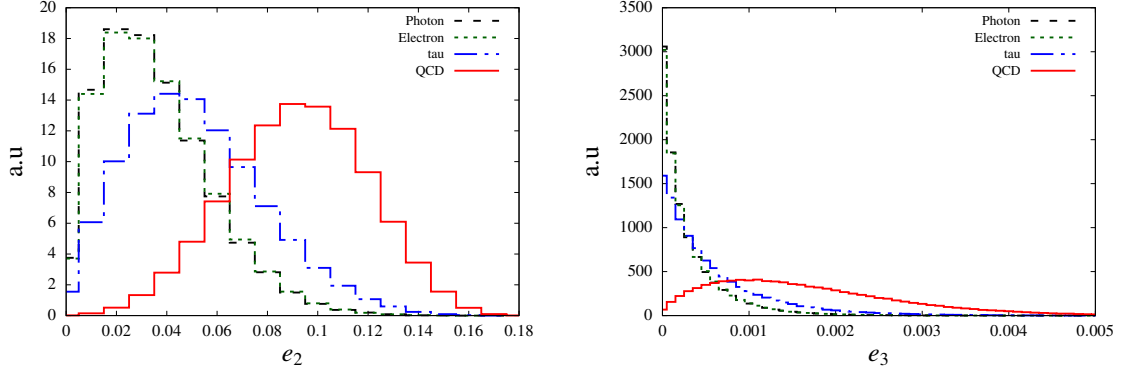


Figure 4: *Distribution of the Energy Correlation Function (ECF) variables e_2 (left) and e_3 (right) for all the standard objects.*

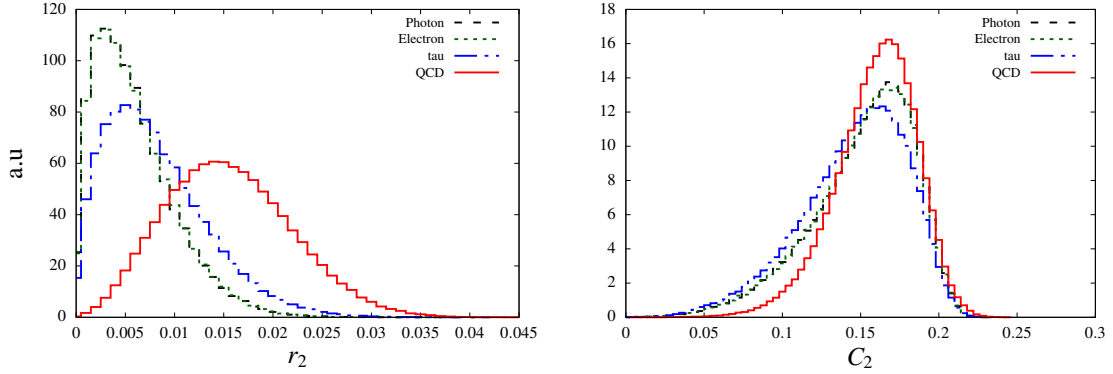


Figure 5: *Distribution of one of the ratios and double ratios of Energy Correlation Function (ECF) r_2 (left) and C_2 (right) respectively for all the standard objects.*

4. From substructure variables to a veto: a demonstration

The purpose of this paper is to provide a simple example where we design a relatively simple veto to discard all standard-jets. In the previous section we have summarized a set of variables and for each of these we have examined the distributions of jets of various kinds. As explained before, after examining the distributions $\{\vec{v}_\alpha\}$, we can identify the patches in the multidimensional space which predominantly get occupied by jets of kind α . We can simply block these patches in order to veto standard-jets. Even though the procedure seems simple, difficulties arise because of the large number of variables – one needs to be clever.

Note that the variables discussed in the last section are all efficient in highlighting differences among jets of different types. However, two among these, namely θ_J and N_T are special. These are the easiest to comprehend and at the same time, no other variables separate different jets as efficiently as these two. In our analysis, we will first employ these two variables to separate the phase space into many segments (see Subsec. 4.1). In Subsec. 4.2, we proceed to analyze those different segments by constructing a realistic veto using multivariate analysis.

4.1. Segmentation of phase space

Schematically, we segment jets first binning according to their electromagnetic characters and then further binning using the number of associated tracks. The arguments are simple: jets with $\theta < \theta_0$ is rich with electromagnetic radiation (mostly neutral pions), and is less likely to be initiated from partons. The count of

tracks is also a fairly good indicators of the origin of the jet. Small track multiplicities (small charged hadron multiplicities) indicate small particle multiplicities overall in the jets, which makes them unlikely to be due to QCD partons. It is then clear that even the use of simple variables such as θ_J , and N_T can already generate these patches where these are primarily occupied by standard-jets of distinct types. In Table 1 we display the

	ϵ_γ (in %)		ϵ_e (in %)		ϵ_τ (in %)		ϵ_j (in %)	
	$\theta < \theta_0$	$\theta \geq \theta_0$	$\theta < \theta_0$	$\theta \geq \theta_0$	$\theta < \theta_0$	$\theta \geq \theta_0$	$\theta < \theta_0$	$\theta \geq \theta_0$
	93.1	6.9	93.9	6.1	10.8	89.2	0.9	99.1
$N_T = 0$	69.7	3.9	8.0	0.64	1.6	5.3	0.15	1.9
$N_T = 1$	20.6	1.4	81.3	4.5	8.5	49.2	0.21	3.2
$N_T = 2$	2.2	0.57	3.8	0.45	0.44	9.9	0.23	7.9
$N_T \geq 3$	0.6	0.99	0.8	0.55	0.22	24.8	0.34	86.1

Table 1: Segmentation of the entire phase-space based on the information of number of tracks associated to the jets and the HCAL information. The regions with $\theta < \theta_0$ are dominated by the ECAL energy deposition (less HCAL deposition) while $\theta \geq \theta_0$ is the same but with HCAL information only where we choose $\theta_0 = 0.25$. Each sample has been divided into two regions, one ECAL-rich while other HCAL-rich. We then further separate them in terms of number of charged tracks associated to the leading jet. For details, see the text.

θ_J	$N_T = 0$	$N_T = 1$	$N_T = 2$	$N_T \geq 3$
$\theta_J < \theta_0$	EC0	EC1	EC2	EC3+
$\theta_J > \theta_0$	HC0	HC1	HC2	HC3+

Table 2: The nomenclature of the regions based on the charged track multiplicity and calorimetry information.

result of segmenting the entire phase-space based on θ_J and N_T . For a jet of kind α , we define the efficiency in a patch/bin as

$$\epsilon_\alpha(\text{bin}) = \frac{\text{Number of jets of type } \alpha \text{ in the bin}}{\text{Total number of jets of type } \alpha} \quad (8)$$

Additionally, in Table 2, we denote how we refer to these regions in this work. For example, the segment EC1, represents the region occupied by the jets with $\theta < \theta_0$ and $N_T = 1$, whereas the segment HC2 represents the region occupied by jets with $\theta \geq \theta_0$ and $N_T = 2$. As seen from the left plot in Fig.1, one expects regions with $\theta \geq \theta_0$ and a large number of tracks are rich in parton initiated jets and further binning these jets in N_T does not really help in finding regions relatively free of QCD-jets. We simply group these regions occupied by HCAL rich jets with large tracks under the designation HC3+.

4.2. A realistic veto using multivariate analyses

Once we segment the entire phase-space in terms of number of tracks and energy profile associated to a jet of standard objects, next goal is to find regions of the phase-space where the contribution coming from these

standard objects are at the sub-percent level. We incorporate all the variables discussed in Sec.3, important in terms of its discrimination power, and then perform a multivariate analysis in order to achieve the maximum sensitivity.

As explained in the guideline discussed in Sec.2.3, we begin with constructing three BDTs, namely

1. $\mathcal{B}^{\gamma/j}$: A BDT to separate photons (signal) from QCD-jets (background).
2. $\mathcal{B}^{j/\tau}$: A BDT to separate QCD-jets (signal) from taus (background).
3. $\mathcal{B}^{\gamma/\tau}$: A BDT to separate photons (signal) from taus (background).

The working principle in a BDT is straightforward. It is a collection of decision trees whose main purposes are to pairwise discriminate two samples. For the sake of notation we refer to one sample as ‘signal’ and the other as ‘background’. Each tree is characterized by different levels of hard cuts on the variables, which selects regions rich in signals. Since a single tree can be sensitive to the choice of the cuts on the variables, multiple trees are constructed, which is followed by a weighing procedure. As mentioned before, the final outcome of the BDT is a single real number (namely, the ‘response’) for each object in the sample. We reweigh responses such that it lies in the range 0 to +1. For a good discriminator, the background and signal events are characterized by $r \sim 0$ and $r \sim +1$ respectively.

In our case, the samples consist of jets. In $\mathcal{B}^{\gamma/j}$, for example, we call the set of photons (or $\{J_\gamma\}$) as signals and the set of QCD-jets (or $\{J_j\}$) as backgrounds. Corresponding to a decision, each jet in the sample (mixed signal and background) is assigned a response of the given analysis. In this example, we expect responses for photons to lie at around 1, whereas QCD-jets to accumulate around 0. Further, as explained in Sec.2.3, we use a naming convention for the responses, similar to the BDTs. For example, the responses for $\mathcal{B}^{\gamma/j}$ will be denoted by $r^{\gamma/j}$.

A crucial part for the construction of BDTs is to find a set of variables. Even though one can use the full set of variables described in Sec.3.3 for all the BDTs, we rather make judicious choices for each of the BDTs. For example, for $\mathcal{B}^{\gamma/j}$, we select variables which exhibit good discriminatory power between photons and QCD-jets. In Table 3, we provide the list of the variables we consider for the three BDTs.

BDT	Variables
$\mathcal{B}^{\gamma/j}$	λ_J, C_1, r_1
$\mathcal{B}^{j/\tau}$	$\lambda_J, r_2, \tau_{31}$
$\mathcal{B}^{\gamma/\tau}$	$\epsilon_J, C_1, \lambda_J, r_1, e_2$

Table 3: List of variables for the discrimination of a given pair of standard-jets. The first variable represent the one best suited (highest weighted) for this discrimination when θ_J and N_T are excluded.

We summarize the results of the BDT analyses (responses) in Fig.6. Each of the plots in Fig.6 shows two dimensional probability distributions corresponding to various standard-jets. The left column corresponds to responses for photons: the top plot shows 2D-histogram in $r^{j/\tau}$ – $r^{\gamma/j}$ plane, whereas the bottom plot shows that in $r^{j/\tau}$ – $r^{\gamma/\tau}$ plane. The color coding associated with each bin reflects the *probability* (not probability-density) of a photon to occupy the bin. The physics understanding of these plots are simple. Note that the y axes in both the plots represent responses for the BDTs $\mathcal{B}^{\gamma/j}$ and $\mathcal{B}^{\gamma/\tau}$, which treat photons as signals and therefore assign large responses correctly. As far as the x -axis is concerned, the BDT $\mathcal{B}^{j/\tau}$ considers photons more τ -like (background) than qcd-jets (signal). Therefore, photons show up mostly in top left corner in both the figures.

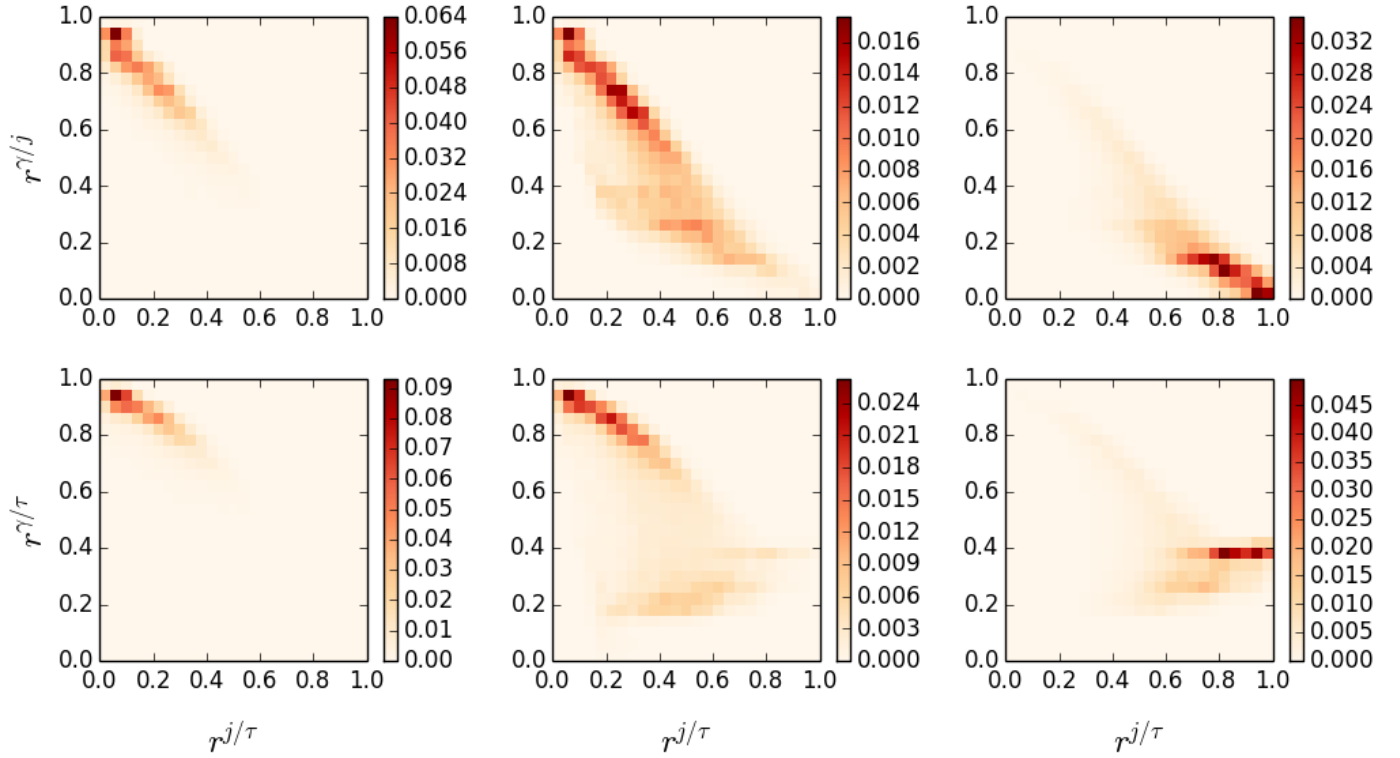


Figure 6: The 2-dimensional distributions of the BDT response variables $r^{\gamma/j}$, $r^{j/\tau}$, and $r^{\gamma/\tau}$ for the standard-jets. The columns from the left to right represent the distributions for the photon, τ and QCD-jets respectively. In these plots we have used $2D$ -bins of size (0.04×0.04) in units of responses.

The central column of plots in Fig.6 show the same distributions, but for τ s. These follow patterns quite similar to that of the photons, and occupy mostly in the top left corner of both the plots. A striking feature in both the plots is that there is quite a few of these jets get characterized by large responses under BDT $\mathcal{B}^{j/\tau}$ even though τ s are treated as background jets. This suggests that the characteristics identified by $\mathcal{B}^{j/\tau}$ to separate j from τ , does not perform as well for a small fraction of tau jets. We think that $\mathcal{B}^{j/\tau}$ becomes efficient in separating taus with single prongs (the largest fraction of tau samples) from QCD-jets. In fact support for this argument can be found in the $\mathcal{B}^{j/\tau}$ responses for photons, which assign all photons (single pronged) small responses. Taus with multi-prong structures show us with large responses. The response of $\mathcal{B}^{\gamma/\tau}$, on the other hand, is quite disappointing. It simply shows that the variables we select here, which mostly analyzes the transverse features of energy depositions in the calorimeter cells are not very efficient in discriminating photons from the most of the tau samples (mostly single pronged). The substructure variables only manage to find taus with multi-prong structures to be substantially different from the photon samples.

Finally, the rightmost column in Fig.6 we show the same probability distributions for QCD-jets. The top plot does not require any subtle explanation. The BDTs $\mathcal{B}^{j/\tau}$ and $\mathcal{B}^{\gamma/j}$ treat qcd-jets as signals and backgrounds respectively, giving these preferred positions in the bottom right corner. The bottom plot is quite interesting. The BDT, $\mathcal{B}^{\gamma/\tau}$ is trained on discriminating photons from the taus. Even though it does not turn out to be very good at separating taus from photons, it nevertheless assigns most of qcd-jets responses within a narrow zone. As we show later, it will end up being highly useful in constructing a veto for QCD-jets.

One can use the phase space distributions to construct vetoes for these standard objects. For example, the

region rich in QCD-jets can be roughly parameterized as:

$$C_1 \leq r^{\gamma/\tau} \leq C_2 \quad \text{AND} \quad \left(\left| r^{\gamma/j} + r^{j/\tau} - 1.0 \right| \leq C_3 \quad \text{OR} \quad \left| r^{\gamma/j} - r^{j/\tau} \right| \geq C_4 \right). \quad (9)$$

In the above equation, C_i s are parameters that can be adjusted to contain most of QCD-jets. A QCD-veto will then reject all jets in the phase-space described in Eq. (9).

In this work, instead of finding a region rich in QCD-jets by eye, we rather take a different approach in order to construct a QCD-veto. We discretize the 3D space of responses $\{r^{j/\tau}, r^{\gamma/j}, r^{\gamma/\tau}\}$ into bins; we calculate the probability of finding QCD-jets in each of the bins; we sort bins in decreasing probability; and finally keep vetoing sorted-bins until only a small (desired) fraction of QCD-jets remain.

Let us elaborate on the procedure described above with a concrete example. Consider the region HC2. As reflected in Table 1, in HC2 $\epsilon_j = 0.079$. This implies that 7.9% of all QCD-jets occupy this section of the phase space. The goal of the following exercise will be to reduce QCD rate below an acceptable level, say R_j . In short we want $\epsilon_j \leq R_j$ in the region HC2.

- We begin with binning the full phase space into cubes of sizes $(0.04 \times 0.04 \times 0.04)$ in units of responses. We can represent each bin either in 3D (for example, the bin (i, j, k) represents the i -th in $r^{j/\tau}$ direction, j -th in $r^{\gamma/j}$ direction, and k -th in $r^{\gamma/\tau}$ direction), or in 1D (for example, the (i, j, k) -th bin gets represented as the b -th bin, where $b = i + n_b \times j + n_b^2 \times k$ with n_b being the number of bins, here $n_b = 25$).
- Each bin is characterized by the probability of QCD-jets occupying the bin. In particular we define bin probabilities to be

$$P_b = \frac{1}{N} \sum_j \begin{cases} 1 & \text{if } j \in b \\ 0 & \text{if } j \notin b \end{cases}, \quad (10)$$

where N is the total number of QCD-jets studied and the index j runs over all QCD-jets. Also, clearly by construction $\sum_b P_b = 1$. The cumulative probability of each bin (namely, C_b) is defined as

$$C_b = \sum_{b'} \begin{cases} P_{b'} & \text{if } P_{b'} \geq P_b \\ 0 & \text{else} \end{cases}, \quad (11)$$

where we sum over all bins b' . A better pictorial representation can be obtained if bins are sorted in decreasing probabilities as shown in Fig.7. In the left-most plot we have shown the distributions P_b and C_b for QCD jets by solid and dashed lines respectively. Note that C_b asymptotes towards 1 as per expectations.

- We also determine bin probabilities in each segment. For example the bin probabilities in HC2 will be given by

$$P_b^{\text{HC2}} = \frac{1}{N} \sum_j \begin{cases} 1 & \text{if } j \in b \quad \& \quad \theta \geq \theta_0 \quad \& \quad N_T = 2 \\ 0 & \text{else} \end{cases}. \quad (12)$$

$$C_b^{\text{HC2}} = \sum_{b'} \begin{cases} P_{b'}^{\text{HC2}} & \text{if } P_{b'}^{\text{HC2}} \geq P_b^{\text{HC2}} \\ 0 & \text{else} \end{cases}. \quad (13)$$

Note that the denominator in Eq. (12) is still given by the total number of QCD-jets. Therefore, one gets $\sum_b P_b^{\text{HC2}} = \epsilon_j^{\text{HC2}}$. In the central plot of Fig.7 we show P_b^{HC2} and C_b^{HC2} again by solid and dashed lines respectively. The distribution C_b^{HC2} now asymptotes to ϵ_j^{HC2} .

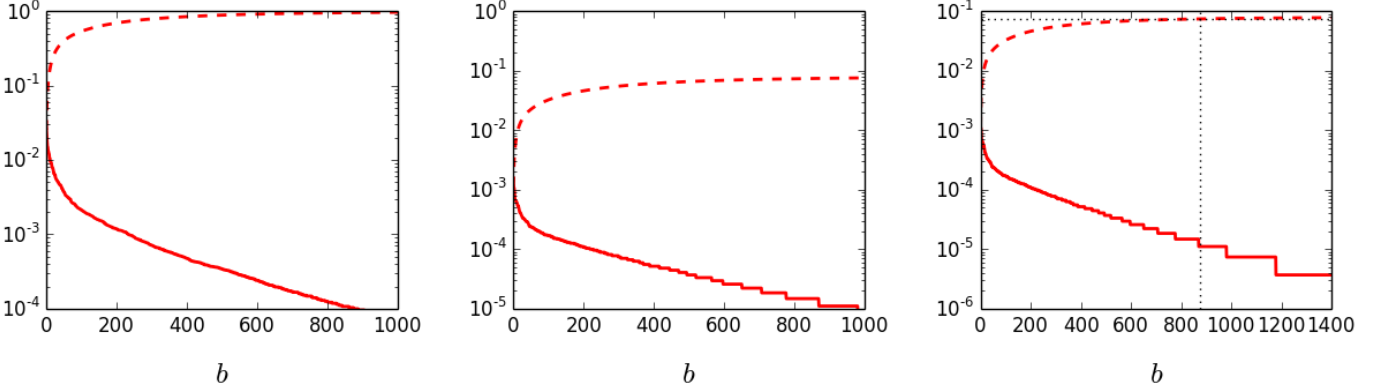


Figure 7: Left: The distributions P_b (solid) and C_b (dashed) for QCD-jets. The bins are sorted in decreasing P_b and we only show the first 1000 bins. Center: The distributions P_b^{HC2} (solid) and C_b^{HC2} (dashed). Right: The same plot as the center one, but for the first 1500 bins. The dotted horizontal line represents $y = 0.074$ (see text for explanation). The QCD-veto for HC2 as described in Eq. (14) and in Eq. (15) blocks 875 bins left of the right dotted line.

- The *QCD-veto* is simply about blocking a collection of bins rich in QCD-jets so that only a small fraction of QCD-jets are allowed. Given a tolerance rate R_j (defined as the rate at which QCD-jets can be allowed), one can then determine the QCD veto function (for HC2) using

$$f_b^j(\text{HC2}) = \begin{cases} 1 & \text{if } C_b^{\text{HC2}} \geq (\epsilon_j^{\text{HC2}} - R_j) \\ 0 & \text{else} \end{cases} \quad (14)$$

where 0 represents bins vetoed and 1 the bins accepted. The logic behind the equation above can be explained in the rightmost plot in Fig.7. The plot is identical to the middle plot except that we only plot first 1500 bins. The dotted horizontal line represents at $y = \epsilon_j^{\text{HC2}} - R_j = 0.074$ (here we have taken $R_j = 0.005$ and ϵ_j^{HC2} is given as 0.079 from Table 1). The vertical dotted line represents the bin for which $C_b^{\text{HC2}} = (\epsilon_j^{\text{HC2}} - R_j) = 0.074$. The veto function in Eq. (14) simply vetoes bins on the left of the line and accepts the bins on the right. The veto function in Eq. (14) can be rewritten in terms of P_b as well. Naming the point where the vertical line intersects P_b^{HC2} to be $P_{R_j}(\text{HC2})$, we can restate

$$f_b^j(\text{HC2}) = \begin{cases} 1 & \text{if } P_b^{\text{HC2}} \leq P_{R_j}(\text{HC2}) \\ 0 & \text{else.} \end{cases} \quad (15)$$

Note that vetoes as stated in Eq. (14) and in Eq. (15) are slightly different, may yield slightly different values of ϵ_j after vetoes are enforced. Differences arise since we did not impose strict inequalities (rather we use \geq and \leq), which get magnified especially in case there are multiple bins corresponding to the $P_b^{\text{HC2}} = P_{R_j}(\text{HC2})$.

We impose QCD-veto as described above in all HC segments. Similar constructions are used to construct photon-veto (for EC0), electron-veto (for EC1 and EC2), and tau-veto (for HC1). The procedure is identical except that we can allow for a larger rate for other vetoes. To be specific, we mainly use two different target rates

$$R_j = 0.005, \quad \text{and} \quad R_\gamma = R_e = R_\tau = 0.05. \quad (16)$$

This implies that we target blocking order 199 in 200 (or target allowing only 1 in every 200) QCD-jets. For jets of other types, we could be less restrictive and allow more jets to pass through (since the production rate for

Regions	Vetoed used	ϵ_γ (in %)	ϵ_e (in %)	ϵ_τ (in %)	ϵ_j (in %)
EC0	Photon Veto with $R_\gamma = 0.05$	5.0	0.70	0.59	0.07
EC1	Electron Veto with $R_e = 0.05$	0.93	5.0	3.8	0.13
EC2	No Veto	2.1	3.8	0.44	0.22
EC3+	No Veto	0.61	0.81	0.22	0.34
HC0	QCD Veto with $R_j = 0.005$	2.4	0.34	3.1	0.58
HC1	QCD Veto with $R_j = 0.005$ Tau Veto with $R_\tau = 0.05$	0.21	0.66	4.9	0.25
HC2	QCD Veto with $R_j = 0.005$	0.14	0.08	2.3	0.55
HC3	QCD Veto with $R_j = 0.005$	0.06	0.03	5.2	0.54
HC4+	QCD Veto with $R_j = 0.005$	0.04	0.02	0.21	0.56

Table 4: Efficiencies in various segments after vetoes are imposed. The vetoes are applied in a way so that QCD-jets are allowed only at the level of $\lesssim 0.5\%$, whereas for other standard-jets we allow efficiencies of order $\lesssim 5\%$.

these jets are small compared to the QCD-jets). In particular, we try blocking roughly 19 out of 20 photons, for example. Note that these number are in sync with what we typically target as tolerable mis-tagging efficiency when designing a tagger. For example, in standard jet-flavor-tagging procedure the working point typically involves 1% or higher mistag efficiency from light-flavor QCD-jets. Similarly, for photon tagging, we tolerate around 5-6% of mistag from electrons.

In Table 4 we show the results as we impose vetoes judiciously on different segments. It turns out that single vetoes are efficient enough to bring down the rate of standard-jets below the acceptable range in all but one segment. In HC2, we need a tau-veto along with a QCD-veto. Note that, given our target, we do not need any veto for EC2, and EC3+, since these segments are already pure.

5. Example Non-standard objects after vetoes

The generality of our analysis enables its application across a wide range of models which includes various non-standard objects, e.g., highly collimated particles, long lived particles etc. In this section, we discuss, as an example, the sensitivity of this analysis to capture some of these non-standard objects, especially collimated di-photon, di-electron and di-tau samples.

Let us emphasize that the purpose of this section is not to categorize, describe or even to tabulate all possible anomalous objects – simply because such tasks are more or less rendered less important due to the nature of our proposal. The vetoes are constructed around the standard objects only, and thus we can always be agnostic of the exact form of new physics while attempting to find traces of new physics.

In order to demonstrate the efficacy of our method, we take three examples of anomalous objects:

- i. Jets initiated by a pair of collimated photons.

- ii. Jets initiated by a pair of collimated electrons.
- iii. Jets initiated by a pair of collimated taus (hadronic).

Note again that vetoes we use (as tabulated in Table 4), have no information regarding the exact nature of any of these anomalous objects. Of course, if analyses use this information, they would perform better – the job here is to demonstrate that even without using any information of anomalous objects we can capture decent amount of these objects.

In order to evaluate the rate at which these objects pass the vetoes, we first need to generate samples, which requires a toy Lagrangian. Once again the details of Lagrangian does not matter. Following the example shown in Ref. [16, 17], we consider a handful of toy models here. The simplistic model by extending the SM Higgs sector with a new scalar field (say, n_1) can be written as:

$$\mathcal{L}_{\text{toy1}} = \frac{1}{2} (\partial^2 - m_1^2) n_1^2 + \frac{1}{2} \mu_1 h n_1^2 + \eta_a \frac{1}{\Lambda} n_1 F^{\mu\nu} \tilde{F}_{\mu\nu} + \eta_e n_1 e e^c + \eta_\tau n_1 \tau \tau^c, \quad (17)$$

where h represents the SM Higgs scalar (of mass $m_h \sim 125$ GeV); m_1, μ_1 are masses much smaller than the cut-off Λ ; and finally all η_i are dimensionless constants. Now, the limit $\eta_e, \eta_\tau \rightarrow 0$, gives rise to Higgs decay to four photons via $p p \rightarrow h \rightarrow n_1(\gamma\gamma)n_1(\gamma\gamma)$. In the limit, $m_1 \ll m_h$, one actually finds each n_1 giving rise to a collimated pair of photons (say, the diphoton-jets). Similarly in the same limit, one finds dielectron-jets or ditau-jets for $\eta_a, \eta_\tau \rightarrow 0$ or $\eta_a, \eta_e \rightarrow 0$ respectively.

We further emphasize that we only use this to generate sample of anomalous objects that tests our proposed anomaly finder. While the Lagrangian is Eq. (17) is easy to understand as well as to implement in a Monte Carlo, the use of Higgs scalar always raises the question whether we can search of it indirectly just using some variations of current search strategies. Such questions are irrelevant. If Higgs is replaced by a new particle of mass say, 1 TeV, which decays only to di-tau-jets, of course, no current strategy will work satisfactorily unless one devises a method to look for di-tau-jets in particular.

Note that the toy model in Eq. (17), can be easily UV-completed in a electroweak symmetric model, where n_1 arise from a electroweak singlet. The mixing term with the Higgs scalar can arise from mixed quartic $|H|^2 n_1^2$, where H is the electroweak doublet. This term also give rise to a quadratic piece in n_1 , that gets absorbed in m_1 . The term with electromagnetic gauge fields easily goes through with the replacement of $F_{\mu\nu} \rightarrow B_{\mu\nu}$, the field strength for hypercharge. Finally, terms with fermions break electroweak symmetry, and therefore must be proportional to the Higgs vacuum expectation value (namely, v). These terms, therefore, can arise from Higher dimensional terms (for example, $\frac{1}{\Lambda} n_1 H l_1 e^c$), where l_1 is the lepton electroweak doublet of the first generation. The coupling η_e is v/Λ suppressed.

The toy model can be extended easily to find non-standard jets with varied particle contents and topologies. A simple modification by adding a new scalar particle n_2 ,

$$\mathcal{L}_{\text{toy2}} = \mathcal{L}_{\text{toy1}} + \frac{1}{2} (\partial^2 - m_2^2) n_2^2 + \frac{1}{2} \mu_2 h n_2^2 + \frac{1}{2} \mu_{12} n_1 n_2^2. \quad (18)$$

Now, setting μ_1 to be zero in $\mathcal{L}_{\text{toy2}}$, opens up Higgs width to eight particles. Of course, in our preferred limit (i.e., $m_2 \ll m_h$), Higgs decays to two non-standard jets, with each of these standard-jets containing various combinations of four collimated particles. Exploring all sorts of topologies for a varied range of parameters is beyond the scope of this paper. As an example, we consider the Lagrangian in Eq. (17), i.e., only study non-standard-jets consisting of pairs of photons, electrons, and tau particles. For the generation of the non-standard topologies, the parameters μ_1 the decay of Higgs into two scalars (n_1) and is chosen to be 0.5. The light scalar n_1 (of mass $m_{n_1} \sim 10$ GeV) couples to a pair of photons, electrons and taus. To generate a collimated process, we assume the decay mode of n_1 into a given final state to be 100%. For instance, for the collimated photon topology, we assume $\eta_\gamma = 1$ and set $\eta_e = \eta_\tau = 0$. It is imperative to note that the decay of the Higgs (h) to

a pair of n_1 with mass around 10 GeV provides the sufficient boost to n_1 (and thus to its decay products) so that it get clustered inside a single jet.

Before proceeding, we outline the behavior of the selected anomalous objects under the variables discussed in Sec. 3.

- $\log(\theta_J)$: The left plot of Fig.8 displays the distribution of the hadronic energy fraction in the leading jet for the non-standard jets. The di-photon (purple-dashed) and di-electron (blue-dotted) exhibit a behavior similar to the single photon and single electron jets as majority of both of the di-samples get deposited at the ECAL with no (or small) energy deposition at the HCAL. The di-tau jets, on the other hand, with both the taus decaying hadronically deposit a significant fraction of their energy in the HCAL, and thereby display a behavior similar to single τ and QCD jets. Thus, as expected, θ_J can be used efficiently to separate the ECAL-rich and HCAL-rich non-standard objects.

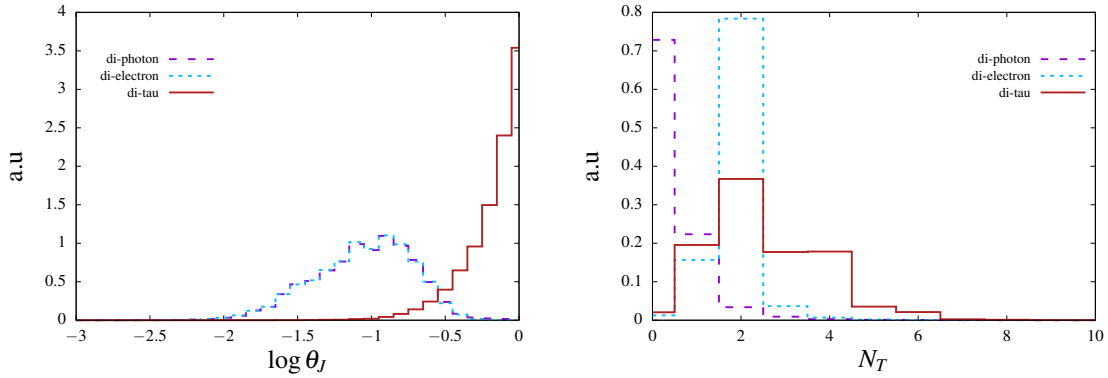


Figure 8: The distribution of hadronic energy fraction (left) and the number of tracks (right) in the leading jet for the non-standard objects.

- N_T : In the right plot of Fig.8 we provide the distribution of the number of tracks inside the leading jet. The track multiplicity for the di-photon and the di-electron are expected to peak at 0 and 2 respectively, while for the di-tau it is a bit more involved owing the single or three pronged nature of a single tau (see Fig.1). As we observe the single-tau being dominantly single pronged, the corresponding track distribution for di-tau peaks at 2. However events with higher track multiplicities can be attributed to different combinations of the single and three pronged nature of the two taus inside the jet. Comparing Fig.1 and Fig.8 one can observe the track multiplicity distribution for the di-tau sample lies somewhat in between the single-tau (and other two di-samples) and QCD-jets, and thus N_T (along with $\log \theta_J$) plays an important role while segmenting the phase space.
- λ_J and ϵ_J : We have already discussed in Sec.3, λ_J quantifies the fraction of the p_T of the jet carried by the leading subjet. For single prong jets (with pencil like structure) λ_J is expected to be small. For example, a jet with $\lambda_J > -0.3$ confirms the presence of two or more subjets. By construction the non-standard samples under consideration are of two prongs structure, as a result the distribution of λ_J expectedly peaks at smaller negative values as opposed to the single electron, photon or tau jets, see Fig.9. From the distributions of λ_J , it is evident that the QCD-jets have a significant overlap with these non-standard objects. The behavior of ϵ_J , which is also a measure of the energy distribution inside a jet, exhibits a pattern similar to λ_J , see right plot of Fig.9. In this case also the non-standard jets have a pattern very much similar to the QCD-jets.
- Energy-Correlation functions (ECFs) and ratios: The key feature of this variable is that it quantifies the distribution of the energy inside a jet utilizing the information of the jet constituents. It is thus a direct

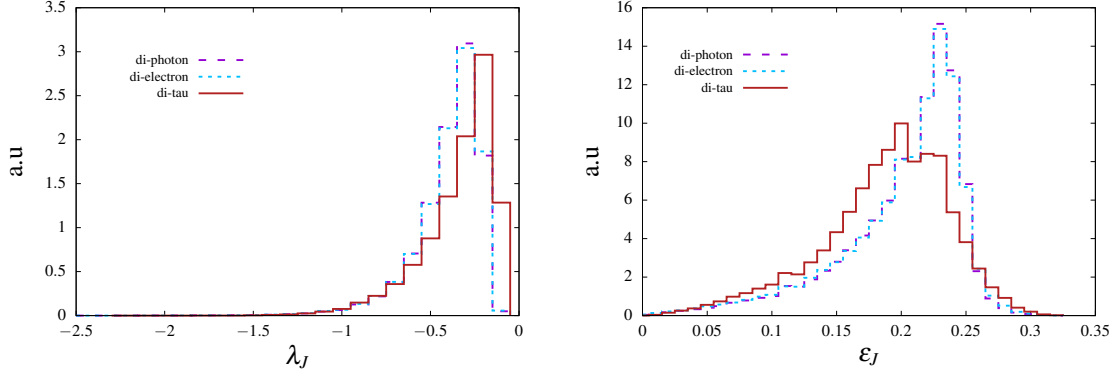


Figure 9: The distribution of λ_J (left) and ϵ_J (right) for the non-standard objects.

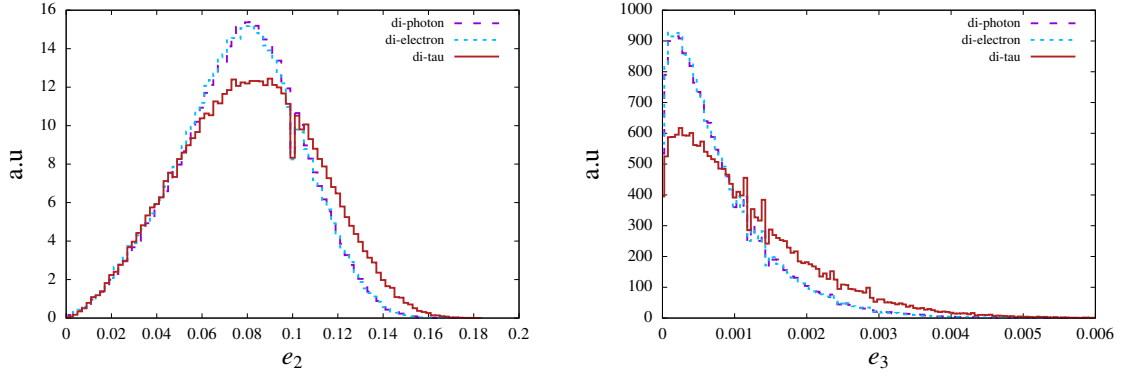


Figure 10: Distribution of the Energy Correlation Function (ECF) variables e_2 (left) and e_3 (right) for all the non-standard objects.

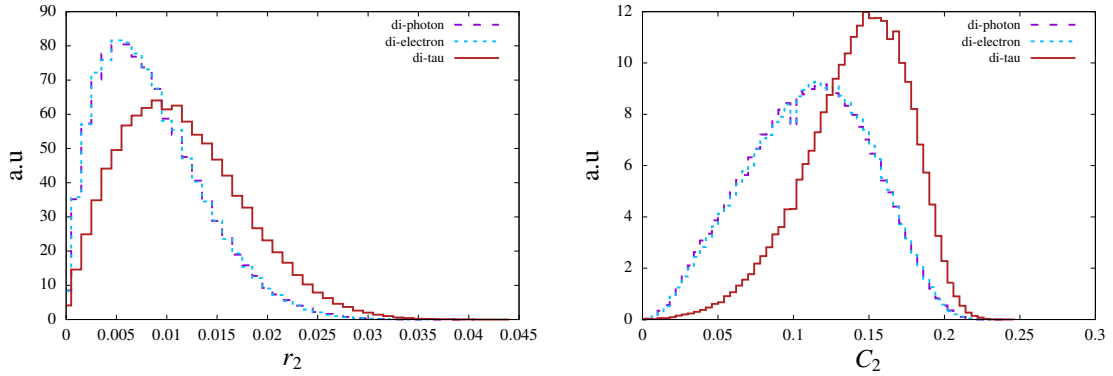


Figure 11: Distribution of one of the ratios and double ratios of Energy Correlation Function (ECF) r_2 (left) and C_2 (right) respectively for all the non-standard objects.

probe of the pronginess of the jet. In Fig.10 we show the distribution for the two ECF variables, namely e_2 (left) and e_3 (right) (see Eq. (6) for definition) for the non-standard objects. As already discussed in Sec.3, the e_{n+1} computed for a jet with n -energetic prongs is always suppressed w.r.t. e_n . Now, as all the non-standard objects are primarily two-pronged, e_3 is expected to peak at much lower values compared

to e_2 . We validate our expectation in Fig.10 where one can indeed see the distribution of e_3 is left-shifted (towards lower values) compared to e_2 . A similar feature can be observed in Fig.11 where we plot the ECF ratios $r_2 = e_3/e_2$ (left) and $C_2 = e_3 e_1/e_2^2$ (right). Since e_2 is always greater than e_3 , r_2 peaks at values ≤ 1 for all the non-standard objects. The larger values of r_2 can be understood from the long tail in the e_3 and e_2 distributions. It is interesting to note that C_2 has some discrimination power for the di-tau jets from the other two di-samples. This can be attributed to the slight difference observed in the peak (and tail) of e_2 and e_3 distributions for di-tau jets. These minor differences get accentuated and, thereby the peak for the di-tau samples get shifted towards slightly higher values.

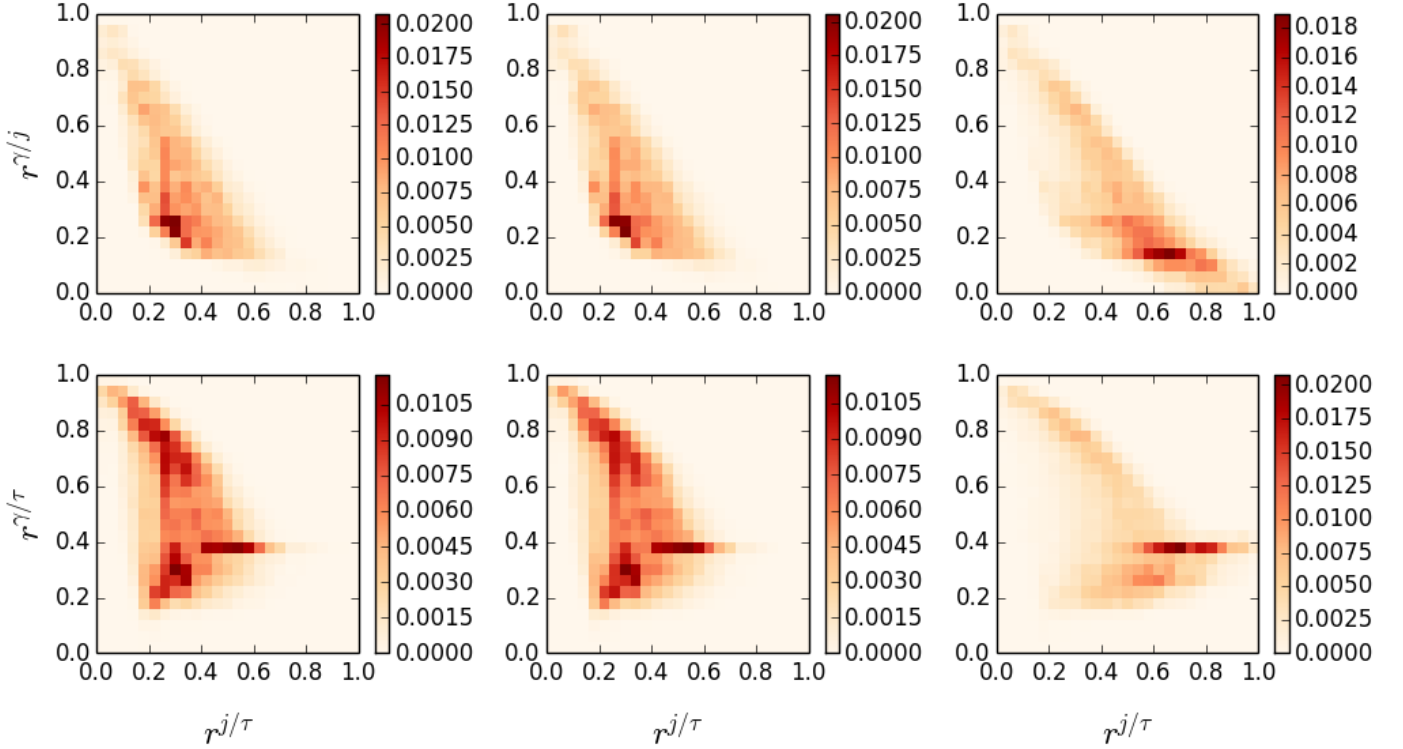


Figure 12: The 2-dimensional distributions of the BDT response variables $r^{\gamma/j}$, $r^{j/\tau}$, and $r^{\gamma/\tau}$ for the anomalous jets. The columns from the left to right represent the distributions for the di-photon, di-electron, and di- τ jets respectively. In these plots we have used 2D-bins of size (0.04×0.04) in units of responses.

It is clear that the segmentation of phase space already separates these three kinds from each other and identify their potential backgrounds. The di-photon jets mostly occupy EC0, di-electrons occupy mostly EC2, whereas di-taus can be found from HC1 to HC3. Following the guideline discussed in Sec.2.3, we find representations of the anomalous objects in three dimensions (given by the three BDTs as discussed in Sec. 4.2). We show the 2-dimensional distributions of the BDT response variables $r^{\gamma/j}$, $r^{j/\tau}$, and $r^{\gamma/\tau}$ for these anomalous objects in Fig.12 where the three columns represent di-photons (left), di-electrons (middle), and di-tau (right) jets.

In Table 5, we summarize the effect the SM vetoes (as tabulated in Table 4) on the non-standard di-samples. The numbers denote the fraction of the anomalous objects remain after the vetoes have been imposed in both the ECAL and HCAL regions segmented with different track multiplicities. The di-photon and di-electron jets are conspicuous by their presence in the EC0 and EC2 regions respectively. The photon veto with $R_\gamma = 0.05$

Regions	Vetoed used	$\epsilon_{\gamma\gamma}$ (in %)	ϵ_{ee} (in %)	$\epsilon_{\tau\tau}$ (in %)
EC0	Photon Veto with $R_\gamma = 0.05$	59.6	0.96	0.21
EC1	Electron Veto with $R_e = 0.05$	17.1	11.8	1.4
EC2	No Veto	2.8	76.3	2.5
EC3+	No Veto	0.83	4.27	0.48
HC0	QCD Veto with $R_j = 0.005$	1.3	0.04	0.99
HC1	QCD Veto with $R_j = 0.005$ Tau Veto with $R_\tau = 0.05$	0.21	0.1	2.7
HC2	QCD Veto with $R_j = 0.005$	0.19	0.42	9.3
HC3	QCD Veto with $R_j = 0.005$	0.06	0.02	2.86
HC4+	QCD Veto with $R_j = 0.005$	0.01	0	1.79

Table 5: The numbers represent the fraction of non-standard jets remain after the vetoes have been imposed.

selects events with the leading jet having two or more photons. Thus, we observe relatively low yield ($\sim 5\%$) for the single photon samples in the EC0, however a large yield of 60% for the di-photon samples. Similar arguments hold for the di-electron samples in the EC2 region, where single electron and τ jets yields are 3.8% and 0.44% respectively in comparison to 76% yield for the di-electron. In EC1 we observe a higher efficiency for the di-photon which can be attributed to the fact that one of the photons can get converted to an electron-positron pair with one of them showing up in the tracker. The di-tau sample with both the taus decaying hadronically is expected to have relatively lower yields in the ECAL regions with varying track multiplicities, and thus mild sensitivity is observed for the photon and/or electron vetoes. It is worth mentioning that EC3+ being supposedly free from the standard objects has a significantly larger efficiency for the di-electron and somewhat milder ($\sim 1\%$) efficiency for the di-photon and di-tau samples.

The single tau and QCD-jets constitute the major background in the hadronic calorimeter region. A QCD veto with $R_j = 0.005$ is imposed for all the segments irrespective of the track multiplicities. The di-tau jets, which is predominantly composed of two tracks has the maximum acceptance in the HC2 region with an efficiency of 9.3% with an acceptance rate of 2.3% and 0.55% for the single tau and QCD-jets respectively. Segments with one and three tracks (HC1 and HC3) also provide an appreciable amount of sensitivity for the di-tau samples when for HC1 a tau veto is additionally imposed. It is interesting to note that the cumulative percentage from the HC1 to HC3 for the di-tau sample is characterized by an acceptance of 16.7%, while for the single tau and QCD are 12.6% and 2% respectively. One may be worried about the yield of di-tau to be comparable to that of a single tau jet, however note that the vetoes were developed by adapting an approach of being agnostic of any non-standard physics. Furthermore, production rate for these single tau events are also much smaller compared to the QCD-jets. Thus, once events with a hint of di-tau signals are triggered, one may repeat the analysis by optimizing the separation of the di-tau jets from single tau and QCD jets as demonstrated for example in [37].

To summarize, in this subsection we demonstrate examples of anomalous objects (collinear particles) passing vetoes that restrict all standard objects (below a pre-determined acceptance rate). Even though the vetoes are constructed without using any information about the anomalous objects, we manage to find anomalous objects at a reasonable rate.

6. Conclusions

The hunt for new physics constitutes an essential ingredient for the current and future run of the LHC. A fundamental assumption employed in these searches is that any new physics is characterized in terms of the standard reconstructed objects, such as isolated photons, electrons, taus, QCD-jets etc. This strategy fails when new physics, instead, give rise to anomalous objects, such as collimated and equally energetic particles, or particles with long lifetime, to name a few. These objects either are missed or are mis-identified as standard-objects. In case these are missed, we lose events unless associated particles trigger. In case, these are mis-identified, we mischaracterize the full event information. Specifically, if we mis-identify these objects as QCD-jets the event gets lost in the sea of SM events due to QCD. Various studies have been proposed towards the discovery of these anomalous objects. However, proposals, typically, rely heavily on specifics of the anomalous objects themselves, which implies that these methods may lose sensitivity fast even for slightly altered NP scenarios.

In this work we propose a framework where we identify these anomalous objects entirely by constructing vetoes around the standard objects. The occurrence of an object passing all vetoes signify the detection of anomalous objects, which, in turn, gives hint of NP. The framework for constructing vetoes as proposed here rely on, (i) the use of jet-clustering algorithms as a universal construct for all objects (standard or non-standard), (ii) an ensemble of conventional and jet-substructure variables to find representations of jets in a multi-dimensional space, (iii) the combination of phase-space segmentation and MVAs to reduce the dimensionality of the space without sacrificing information pertaining to pairwise differences among standard-objects, and finally (iv) an algorithm (loosely based on the greedy algorithm) to identify regions rich in standard-jets. The procedure proposed here is completely agnostic of the form of new physics and therefore can be widely applied across different new physics scenarios which may give rise to such anomalous objects.

Notice that the current set up of the proposed “Anomaly Finder” does not include the Muons and b-jets. The identification and reconstruction of Muons and b-jets at the LHC involve specialized techniques. In the existing set up, the b-jets would fall into the category of identified QCD-jets. However, note that b-jet reconstruction strategy at the LHC includes the combined information of the calorimeter energy deposits as well as information of displaced tracks and properties of secondary and tertiary decay vertices reconstructed within the jet [69, 70]. These additional information will thus introduce a collection of new kinematic variables, especially in terms of vertex and life-time information of the B-hadrons. The inclusion of this information in the proposed framework is indeed interesting and a straightforward extension of the proposed framework. The Muons, at the LHC, are reconstructed from the tracks in the inner detector and muon spectrometer information, which are then combined to improve the reconstruction efficiency and background rejection rates [71, 72]. Moreover, the Muon candidates are also required to satisfy stringent lepton isolation cuts. In this work, we reconstruct jets using calorimeter information only, and so we don’t have the full information for the Muons. However, we can still define a region of parameter space which should be Muon-rich. For example, we can look for events with exactly one track associated to the jet with negligible energy depositions both in ECAL and HCAL. This segment of phase-space is very unique, and has almost no overlap with τ - or QCD-rich jets. Additionally, we can also extend the existing set up incorporating variables based on Muon spectrometer information.

Before we end, a practical guideline on the implementation of this proposal is worth mentioning. Here we propose two strategies to categorize the data samples to be analyzed at the LHC. First one is an *offline* analysis, while the other an *online* implementation. The offline mode assumes that the event has already

been triggered at the HLT level through the existing trigger menu by reconstructing, for each event, objects like electrons, muons, and jets and then selected based on several identification criteria and physics related goals. Once the events are triggered and selected, the proposed analysis, namely the ‘anomaly finder’, can be performed independently to look for new physics signatures. Of course, here we assume that the anomaly finder has already been optimized using the control sample, and thus one needs to simply pass the registered events through the anomaly finder. Further, one can also use additional information from the processes like the associated production of Higgs boson with a Z-boson (with Z decaying to muons or invisibly), or say pair produced Z-bosons with both Z decaying leptonically etc., to model the standard objects in the Higgs or, Z channels. Here we stress that all of these analyses can be performed offline, and thus, this proposal provides a unique framework to probe a wide range of new physics scenarios by directly identifying events containing anomalous objects. Note that, one can always perform supervised analysis later to probe the origin and nature of those anomalous objects.

The second approach, a bit more aggressive, is to combine the proposed ‘anomaly finder’ with the existing HLTs, which will provide a unified framework to look for direct imprints of new physics in the LHC data. It is interesting to note that both the ATLAS and CMS collaborations at the LHC have modified and redesigned the trigger menu significantly to cope with the higher event rates at run-2 as well as high luminosity runs of LHC [73, 74]. The HLT softwares are now upgraded to enhance the acceptance rates by making the algorithms and selections criteria similar to the offline reconstruction techniques for objects like electrons, muons and jets. Interestingly, anti- k_T jets with varying values of jet radius are reconstructed at the HLT with the calorimeter topo-clusters constructed from the calorimeter cells. These jets are then calibrated for the nonlinearity of the calorimeter response and pileup effects using a combination of studies based on simulation and collision data. Identification and tagging the flavour of these reconstructed jets, e.g. b-jet tagging, tau-tagging etc., are now an integral part of the HLT system. Moreover, these updated online flavour tagging templates now include advanced multivariate analysis (MVA) incorporating various discriminating variables mimicking their offline templates [75–78]. Search for exotic new physics signatures at the LHC, for example, long-lived particles, displaced jets, displaced leptons etc., also utilize sophisticated MVA-based techniques and algorithms especially designed to trigger these rare events, for example [79]. Thus, we understand that the existing HLT set up is already efficient enough to handle sophisticated algorithms similar to their offline counterparts, and provide impressive results. The proposed ‘anomaly finder’ require to construct several variables utilizing the tracker and calorimeter information, and perform a MVA to obtain a collection of vetoes that eliminate all standard-objects upto a pre-determined acceptance rate. In this work we assume the acceptance rate for the QCD-jets to be 0.5%, while the existing HLT photon trigger menu accepts isolated photons ($p_T > 20$ GeV, loose selection) with an efficiency of 97% with a rejection factor for the QCD-jets around 1000 [80, 81]. A crucial aspect of the proposed anomaly finder is that it includes a free/input parameter that directly controls the rate at which QCD jets get accepted. Our choice was essentially aimed to provide a concrete example, however one can always tune the parameter associated to the QCD rejection rate to a desired value while probing a wide class of new physics models.

Therefore, this proposal can be used either as a stand-alone framework (offline mode) once we select events after the HLT with acceptable event rates, or we combine it with the existing HLT menu (online mode) with moderate thresholds for the SM event rates. Both the strategies are expected to work reasonably well with the real data.

Acknowledgements

We thank Adam Martin and Michael Graesser for careful reading of an earlier version of the draft and sending critical remarks. A significant part of the computations was completed in the Gaggles cluster at TIFR. Some preliminary simulations were also carried out at the Mapache cluster in the HPC facility at LANL. TSR was supported by the Early Career Research Award by Science and Engineering Research Board, Dept. of Science and Technology, Govt. of India (grant no: ECR/2015/000196). We also thank Sreerup Raychaudhuri for helping us with computational resources.

Appendix A. Simulation Details

1. Event simulation: As outlined in Sec.2.2, standard-jets of various kinds (i.e., single photon, single electron, single tau, QCD-parton initiated jets) are constructed from the leading jet of an event. For example, photon (or jet of type photon) is the leading jet in events with $pp \rightarrow h \rightarrow \gamma\gamma$ where h represents the 125 GeV SM Higgs boson. The event generation as well as parton showering and hadronization have been performed using **Pythia** 8.2 [82] with parton distribution function NNPDF2.3 [83]. For the non-standard jets, we implement the toy Lagrangian described in Eq. (17) in **FeynRules** 2.0 [84]. The generated model files are then used to generate the events using **MadGraph** 2.3.3 [85]. The events are then passed to **Pythia** for showering and hadronization.
2. Detector simulation: In order to perform a fast detector simulation, we use **Delphes** 3.3.2 [86, 87] with the CMS card. The default charged and neutral particle identification efficiencies as implemented in the card have been used. We simulate low- Q^2 soft QCD pile-up events using **Pythia** and then pass it through **Delphes**. The default parametrization as implemented in the CMS card has been used to distribute the minimum-bias pile-up events and hard scattering events in time and z positions. The mean number of soft events merged with each hard scattering, denoted by $\langle N_{PU} \rangle$, is considered to be 40. Note that, after adding these low- Q^2 soft QCD events, one has to identify the primary vertex and then remove those collisions which are not associated to the primary vertex; one can achieve this by performing a pile-up subtraction technique.

A combination of vertex and tracker information helps to identify (and then remove) the contamination of the charged particles originating from the pile-ups. On the other hand, contribution of neutral particles to the pile-up events can be estimated, and then physical observables can be accordingly corrected, by using the jet area method [88, 89]. In this work, we follow the default set up of **Delphes** CMS card to perform the pile-up subtraction. A spatial vertex resolution parameter $|z|$ is used to perform the charged pile-up subtraction; every charged particle originating from a reconstructed vertex with $|z| > 0.01$ cm are considered as coming from pile-ups. We consider those tracks which are passed through the **TrackPileUpSubtractor** module in **Delphes**. Jets are constructed with the calorimeter tower elements using **Fastjet** 3.1.3 [90] with anti- k_T jet algorithm [49], jet radius $R = 0.4$ with $p_T > 50$ GeV. Similar to the tracks, we require to correct the reconstructed jets from low- Q^2 pile-up events containing neutral particles. Note that, charged particles that have failed to be reconstructed as tracks or, are outside the tracker volume can also contribute here. In **Delphes**, the residual pile-up subtraction is achieved by using an algorithm based on the jet area. This technique helps to correct the jet momenta by calculating pile-up density (ρ) and jet area. Here we use the jets constructed using the calorimetric information and allow the default estimation of ρ with the **EFlow** elements. Finally, we recluster the constituents of the pile-up subtracted leading jet (p_T ordered), obtained from the **JetPileUpSubtractor** module, to find an exclusive C/A jet [52]. This pile-up subtracted C/A jet is considered in rest of our analysis. The last step of jet clustering is performed just to have a C/A-based clustering history of the jet. The number of tracks associated to the leading jet is counted by calculating ΔR between the jet and each pile-up subtracted

track, and then accept those tracks with $p_T \geq 2$ GeV and $\Delta R < 0.4$, where $(\Delta R)^2 \equiv (\Delta\eta)^2 + (\Delta\phi)^2$ with $\Delta\eta$ and $\Delta\phi$ being the differences in pseudo-rapidity and azimuthal angle between them respectively.

3. Photon conversion: In order to implement conversion of photons in the tracker portion of the detector we simply follow the prescription as described in Ref. [16, 17]. We register a track for photons after drawing a random number from 0 to 1 in a flat grid. The probability of conversion is η -dependent, since the amount of material a photon passes through (*i.e.*, the number of radiation lengths) varies with directions. For simplicity, in this analysis we assign a flat conversion probability of 20%.
4. BDT parameters: The parameters associated to BDT analyses are chosen as follows: the number of trees in the forest `NTrees` = 800, the maximum depth of the decision tree `MaxDepth` = 3, and finally, the minimum percentage of training events required in a leaf node `MinNodeSize` = 2.5%. All other necessary variables are kept at their default values. Furthermore, we consider the `AdaBoost` method [91] for boosting the decision trees with the boost parameter `AdaBoostBeta` = 0.5.

References

- [1] 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. B716 (2012) 1–29, CERN-PH-EP-2012-218. [arXiv:1207.7214](#), [doi:10.1016/j.physletb.2012.08.020](#).
- [2] S. Chatrchyan, et al., Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, Phys. Lett. B716 (2012) 30–61, CMS-HIG-12-028, CERN-PH-EP-2012-220. [arXiv:1207.7235](#), [doi:10.1016/j.physletb.2012.08.021](#).
- [3] A. M. Sirunyan, et al., Search for supersymmetry in pp collisions at $\sqrt{s} = 13$ TeV in the single-lepton final state using the sum of masses of large-radius jets CMS-SUS-16-037, CERN-EP-2017-088. [arXiv:1705.04673](#).
- [4] C. Collaboration, Search for direct stop pair production in the dilepton final state at $\sqrt{s} = 13$ TeV CMS-PAS-SUS-17-001.
- [5] C. Collaboration, Combined search for electroweak production of charginos and neutralinos in pp collisions at $\sqrt{s} = 13$ TeV CMS-PAS-SUS-17-004.
- [6] T. A. collaboration, Search for direct top squark pair production in final states with two leptons in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector ATLAS-CONF-2017-034.
- [7] M. Aaboud, et al., Search for supersymmetry in final states with two same-sign or three leptons and jets using 36 fb^{-1} of $\sqrt{s} = 13$ TeV pp collision data with the ATLAS detector CERN-EP-2017-108. [arXiv:1706.03731](#).
- [8] T. A. collaboration, Search for electroweak production of supersymmetric particles in the two and three lepton final state at $\sqrt{s} = 13$ TeV with the ATLAS detector ATLAS-CONF-2017-039.
- [9] F. Beaudette, [The CMS Particle Flow Algorithm](#), in: Proceedings, International Conference on Calorimetry for the High Energy Frontier (CHEF 2013): Paris, France, April 22-25, 2013, 2013, pp. 295–304. [arXiv:1401.8155](#).
URL <https://inspirehep.net/record/1279774/files/arXiv:1401.8155.pdf>

- [10] J. Mitrevski, [Electron and photon reconstruction with the atlas detector](#), Nuclear and Particle Physics Proceedings 273 (2016) 2539 – 2541, 37th International Conference on High Energy Physics (ICHEP). [doi:http://dx.doi.org/10.1016/j.nuclphysbps.2015.09.452](http://dx.doi.org/10.1016/j.nuclphysbps.2015.09.452).
URL <http://www.sciencedirect.com/science/article/pii/S2405601415009414>
- [11] V. Khachatryan, et al., Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV, JINST 10 (08) (2015) P08010, cMS-EGM-14-001, CERN-PH-EP-2015-006. [arXiv:1502.02702](#), [doi:10.1088/1748-0221/10/08/P08010](https://doi.org/10.1088/1748-0221/10/08/P08010).
- [12] B. A. Dobrescu, G. L. Landsberg, K. T. Matchev, Higgs boson decays to CP odd scalars at the Tevatron and beyond, Phys. Rev. D63 (2001) 075003. [arXiv:hep-ph/0005308](#), [doi:10.1103/PhysRevD.63.075003](https://doi.org/10.1103/PhysRevD.63.075003).
- [13] S. Chang, P. J. Fox, N. Weiner, Visible Cascade Higgs Decays to Four Photons at Hadron Colliders, Phys. Rev. Lett. 98 (2007) 111802. [arXiv:hep-ph/0608310](#), [doi:10.1103/PhysRevLett.98.111802](https://doi.org/10.1103/PhysRevLett.98.111802).
- [14] N. Toro, I. Yavin, Multiphotons and photon jets from new heavy vector bosons, Phys. Rev. D86 (2012) 055005. [arXiv:1202.6377](#), [doi:10.1103/PhysRevD.86.055005](https://doi.org/10.1103/PhysRevD.86.055005).
- [15] P. Draper, D. McKeen, Diphotons from Tetrphotons in the Decay of a 125 GeV Higgs at the LHC, Phys. Rev. D85 (2012) 115023. [arXiv:1204.1061](#), [doi:10.1103/PhysRevD.85.115023](https://doi.org/10.1103/PhysRevD.85.115023).
- [16] S. D. Ellis, T. S. Roy, J. Scholtz, Jets and Photons, Phys. Rev. Lett. 110 (12) (2013) 122003. [arXiv:1210.1855](#), [doi:10.1103/PhysRevLett.110.122003](https://doi.org/10.1103/PhysRevLett.110.122003).
- [17] S. D. Ellis, T. S. Roy, J. Scholtz, Phenomenology of Photon-Jets, Phys. Rev. D87 (1) (2013) 014015. [arXiv:1210.3657](#), [doi:10.1103/PhysRevD.87.014015](https://doi.org/10.1103/PhysRevD.87.014015).
- [18] P. Agrawal, J. Fan, B. Heidenreich, M. Reece, M. Strassler, Experimental Considerations Motivated by the Diphoton Excess at the LHC, JHEP 06 (2016) 082. [arXiv:1512.05775](#), [doi:10.1007/JHEP06\(2016\)082](https://doi.org/10.1007/JHEP06(2016)082).
- [19] H. Fukuda, M. Ibe, O. Jinnouchi, M. Nojiri, Cracking down on fake photons: Cases of diphoton resonance imposters, PTEP 2017 (3) (2017) 033B05. [arXiv:1607.01936](#), [doi:10.1093/ptep/ptx019](https://doi.org/10.1093/ptep/ptx019).
- [20] S. Knapen, T. Melia, M. Papucci, K. Zurek, Rays of light from the LHC, Phys. Rev. D93 (7) (2016) 075020. [arXiv:1512.04928](#), [doi:10.1103/PhysRevD.93.075020](https://doi.org/10.1103/PhysRevD.93.075020).
- [21] J. Chang, K. Cheung, C.-T. Lu, Interpreting the 750 GeV diphoton resonance using photon jets in hidden-valley-like models, Phys. Rev. D93 (7) (2016) 075013. [arXiv:1512.06671](#), [doi:10.1103/PhysRevD.93.075013](https://doi.org/10.1103/PhysRevD.93.075013).
- [22] D. Curtin, et al., Exotic decays of the 125 GeV Higgs boson, Phys. Rev. D90 (7) (2014) 075004. [arXiv:1312.4992](#), [doi:10.1103/PhysRevD.90.075004](https://doi.org/10.1103/PhysRevD.90.075004).
- [23] B. Dasgupta, J. Kopp, P. Schwaller, Photons, photon jets, and dark photons at 750 GeV and beyond, Eur. Phys. J. C76 (5) (2016) 277. [arXiv:1602.04692](#), [doi:10.1140/epjc/s10052-016-4127-4](https://doi.org/10.1140/epjc/s10052-016-4127-4).
- [24] B. C. Allanach, D. Bhatia, A. M. Iyer, Dissecting Multi-Photon Resonances at the Large Hadron Collider [arXiv:1706.09039](#).
- [25] G. Aad, et al., A search for prompt lepton-jets in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector, Phys. Lett. B719 (2013) 299–317, cERN-PH-EP-2012-319. [arXiv:1212.5409](#), [doi:10.1016/j.physletb.2013.01.034](https://doi.org/10.1016/j.physletb.2013.01.034).

- [26] G. Aad, et al., Search for long-lived neutral particles decaying into lepton jets in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector, JHEP 11 (2014) 088, cERN-PH-EP-2014-209. [arXiv:1409.0746](#), [doi:10.1007/JHEP11\(2014\)088](#).
- [27] A. Falkowski, J. T. Ruderman, T. Volansky, J. Zupan, Discovering Higgs Decays to Lepton Jets at Hadron Colliders, Phys. Rev. Lett. 105 (2010) 241801. [arXiv:1007.3496](#), [doi:10.1103/PhysRevLett.105.241801](#).
- [28] A. Falkowski, J. T. Ruderman, T. Volansky, J. Zupan, Hidden Higgs Decaying to Lepton Jets, JHEP 05 (2010) 077. [arXiv:1002.2952](#), [doi:10.1007/JHEP05\(2010\)077](#).
- [29] J. T. Ruderman, T. Volansky, Decaying into the Hidden Sector, JHEP 02 (2010) 024. [arXiv:0908.1570](#), [doi:10.1007/JHEP02\(2010\)024](#).
- [30] A. M. Iyer, U. Maitra, Dissecting new physics models through kinematic edges, Phys. Rev. D95 (3) (2017) 035039. [arXiv:1609.06502](#), [doi:10.1103/PhysRevD.95.035039](#).
- [31] G. Barelo, S. Chang, C. A. Newby, B. Ostdiek, Don't be left in the dark: Improving LHC searches for dark photons using lepton-jet substructure, Phys. Rev. D95 (5) (2017) 055007. [arXiv:1612.00026](#), [doi:10.1103/PhysRevD.95.055007](#).
- [32] S. Dube, D. Gadkari, A. M. Thalapillil, Lepton-Jets and Low-Mass Sterile Neutrinos at Hadron Colliders [arXiv:1707.00008](#).
- [33] C. Cheung, J. T. Ruderman, L.-T. Wang, I. Yavin, Lepton Jets in (Supersymmetric) Electroweak Processes, JHEP 04 (2010) 116. [arXiv:0909.0290](#), [doi:10.1007/JHEP04\(2010\)116](#).
- [34] J. Chang, K. Cheung, S.-C. Hsu, C.-T. Lu, Detecting multimueon jets from the Higgs boson exotic decays in the Higgs portal framework, Phys. Rev. D95 (3) (2017) 035012. [arXiv:1607.07550](#), [doi:10.1103/PhysRevD.95.035012](#).
- [35] M. Buschmann, J. Kopp, J. Liu, P. A. N. Machado, Lepton Jets from Radiating Dark Matter, JHEP 07 (2015) 045. [arXiv:1505.07459](#), [doi:10.1007/JHEP07\(2015\)045](#).
- [36] A. Katz, M. Son, B. Tweedie, Dilepton-Jet Tagging and Boosted Higgses from a Multi-TeV Resonance, Phys. Rev. D83 (2011) 114033. [arXiv:1011.4523](#), [doi:10.1103/PhysRevD.83.114033](#).
- [37] C. Englert, T. S. Roy, M. Spannowsky, Dilepton jets in Higgs searches, Phys. Rev. D84 (2011) 075026. [arXiv:1106.4545](#), [doi:10.1103/PhysRevD.84.075026](#).
- [38] J. A. Evans, J. Shelton, Long-Lived Staus and Displaced Leptons at the LHC, JHEP 04 (2016) 056. [arXiv:1601.01326](#), [doi:10.1007/JHEP04\(2016\)056](#).
- [39] B. C. Allanach, M. Badziak, G. Cottin, N. Desai, C. Hugonie, R. Ziegler, Prompt Signals and Displaced Vertices in Sparticle Searches for Next-to-Minimal Gauge Mediated Supersymmetric Models, Eur. Phys. J. C76 (9) (2016) 482. [arXiv:1606.03099](#), [doi:10.1140/epjc/s10052-016-4330-3](#).
- [40] G. Aad, et al., Search for massive, long-lived particles using multitrack displaced vertices or displaced lepton pairs in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector, Phys. Rev. D92 (7) (2015) 072004. [arXiv:1504.05162](#), [doi:10.1103/PhysRevD.92.072004](#).
- [41] T. A. collaboration, Search for long-lived neutral particles decaying into displaced lepton jets in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector ATLAS-CONF-2016-042.

- [42] C. Collaboration, Search for Displaced SUSY in Dilepton Final States CMS-PAS-B2G-12-024.
- [43] G. Aad, et al., Search for pair-produced long-lived neutral particles decaying in the ATLAS hadronic calorimeter in pp collisions at $\sqrt{s} = 8$ TeV, Phys. Lett. B743 (2015) 15–34. [arXiv:1501.04020](#), [doi:10.1016/j.physletb.2015.02.015](#).
- [44] R. Aaij, et al., Search for long-lived particles decaying to jet pairs, Eur. Phys. J. C75 (4) (2015) 152. [arXiv:1412.3021](#), [doi:10.1140/epjc/s10052-015-3344-6](#).
- [45] V. Khachatryan, et al., Search for Long-Lived Neutral Particles Decaying to Quark-Antiquark Pairs in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV, Phys. Rev. D91 (1) (2015) 012007. [arXiv:1411.6530](#), [doi:10.1103/PhysRevD.91.012007](#).
- [46] P. Bhupal Dev, R. N. Mohapatra, Y. Zhang, Displaced photon signal from a possible light scalar in minimal left-right seesaw model, Phys. Rev. D95 (11) (2017) 115001. [arXiv:1612.09587](#), [doi:10.1103/PhysRevD.95.115001](#).
- [47] H. Ito, O. Jinnouchi, T. Moroi, N. Nagata, H. Otono, Extending the LHC Reach for New Physics with Sub-Millimeter Displaced Vertices, Phys. Lett. B771 (2017) 568–575. [arXiv:1702.08613](#), [doi:10.1016/j.physletb.2017.06.003](#).
- [48] S. Banerjee, G. Blanger, B. Bhattacharjee, F. Boudjema, R. M. Godbole, S. Mukherjee, Novel signatures for long-lived particles at the LHC [arXiv:1706.07407](#).
- [49] M. Cacciari, G. P. Salam, G. Soyez, The Anti- $k(t)$ jet clustering algorithm, JHEP 04 (2008) 063. [arXiv:0802.1189](#), [doi:10.1088/1126-6708/2008/04/063](#).
- [50] S. D. Ellis, D. E. Soper, Successive combination jet algorithm for hadron collisions, Phys. Rev. D48 (1993) 3160–3166. [arXiv:hep-ph/9305266](#), [doi:10.1103/PhysRevD.48.3160](#).
- [51] S. Catani, Y. L. Dokshitzer, M. H. Seymour, B. R. Webber, Longitudinally invariant K_t clustering algorithms for hadron hadron collisions, Nucl. Phys. B406 (1993) 187–224. [doi:10.1016/0550-3213\(93\)90166-M](#).
- [52] Y. L. Dokshitzer, G. D. Leder, S. Moretti, B. R. Webber, Better jet clustering algorithms, JHEP 08 (1997) 001. [arXiv:hep-ph/9707323](#), [doi:10.1088/1126-6708/1997/08/001](#).
- [53] J. Thaler, K. Van Tilburg, Identifying Boosted Objects with N-subjettiness, JHEP 03 (2011) 015. [arXiv:1011.2268](#), [doi:10.1007/JHEP03\(2011\)015](#).
- [54] J. Thaler, K. Van Tilburg, Maximizing Boosted Top Identification by Minimizing N-subjettiness, JHEP 02 (2012) 093. [arXiv:1108.2701](#), [doi:10.1007/JHEP02\(2012\)093](#).
- [55] A. J. Larkoski, G. P. Salam, J. Thaler, Energy Correlation Functions for Jet Substructure, JHEP 06 (2013) 108. [arXiv:1305.0007](#), [doi:10.1007/JHEP06\(2013\)108](#).
- [56] J. M. Butterworth, A. R. Davison, M. Rubin, G. P. Salam, Jet substructure as a new Higgs search channel at the LHC, Phys. Rev. Lett. 100 (2008) 242001. [arXiv:0802.2470](#), [doi:10.1103/PhysRevLett.100.242001](#).
- [57] D. Krohn, J. Thaler, L.-T. Wang, Jet Trimming, JHEP 02 (2010) 084. [arXiv:0912.1342](#), [doi:10.1007/JHEP02\(2010\)084](#).

- [58] S. D. Ellis, C. K. Vermilion, J. R. Walsh, Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches, *Phys. Rev. D* 81 (2010) 094023. [arXiv:0912.0033](#), [doi:10.1103/PhysRevD.81.094023](#).
- [59] P. Speckmayer, A. Hocker, J. Stelzer, H. Voss, The toolkit for multivariate data analysis, TMVA 4, *J. Phys. Conf. Ser.* 219 (2010) 032057. [doi:10.1088/1742-6596/219/3/032057](#).
- [60] S. J. Brodsky, J. F. Gunion, Hadron Multiplicity in Color Gauge Theory Models, *Phys. Rev. Lett.* 37 (1976) 402–405. [doi:10.1103/PhysRevLett.37.402](#).
- [61] J. Gallicchio, M. D. Schwartz, Seeing in Color: Jet Superstructure, *Phys. Rev. Lett.* 105 (2010) 022001. [arXiv:1001.5027](#), [doi:10.1103/PhysRevLett.105.022001](#).
- [62] J. Gallicchio, M. D. Schwartz, Pure Samples of Quark and Gluon Jets at the LHC, *JHEP* 10 (2011) 103. [arXiv:1104.1175](#), [doi:10.1007/JHEP10\(2011\)103](#).
- [63] J. Gallicchio, M. D. Schwartz, Quark and Gluon Tagging at the LHC, *Phys. Rev. Lett.* 107 (2011) 172001. [arXiv:1106.3076](#), [doi:10.1103/PhysRevLett.107.172001](#).
- [64] J. Gallicchio, M. D. Schwartz, Quark and Gluon Jet Substructure, *JHEP* 04 (2013) 090. [arXiv:1211.7038](#), [doi:10.1007/JHEP04\(2013\)090](#).
- [65] B. Bhattacherjee, S. Mukhopadhyay, M. M. Nojiri, Y. Sakaki, B. R. Webber, Associated jet and subjet rates in light-quark and gluon jet discrimination, *JHEP* 04 (2015) 131. [arXiv:1501.04794](#), [doi:10.1007/JHEP04\(2015\)131](#).
- [66] B. Bhattacherjee, S. Mukhopadhyay, M. M. Nojiri, Y. Sakaki, B. R. Webber, Quark-gluon discrimination in the search for gluino pair production at the LHC, *JHEP* 01 (2017) 044. [arXiv:1609.08781](#), [doi:10.1007/JHEP01\(2017\)044](#).
- [67] I. W. Stewart, F. J. Tackmann, W. J. Waalewijn, N-Jettiness: An Inclusive Event Shape to Veto Jets, *Phys. Rev. Lett.* 105 (2010) 092002. [arXiv:1004.2489](#), [doi:10.1103/PhysRevLett.105.092002](#).
- [68] I. Moult, L. Necib, J. Thaler, New Angles on Energy Correlation Functions, *JHEP* 12 (2016) 153. [arXiv:1609.07483](#), [doi:10.1007/JHEP12\(2016\)153](#).
- [69] A. M. Sirunyan, et al., Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV, *JINST* 13 (05) (2018) P05011. [arXiv:1712.07158](#), [doi:10.1088/1748-0221/13/05/P05011](#).
- [70] G. Aad, et al., Performance of b -Jet Identification in the ATLAS Experiment, *JINST* 11 (04) (2016) P04008. [arXiv:1512.01094](#), [doi:10.1088/1748-0221/11/04/P04008](#).
- [71] A. M. Sirunyan, et al., Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV [arXiv:1804.04528](#).
- [72] G. Aad, et al., Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s} = 13$ TeV, *Eur. Phys. J. C* 76 (5) (2016) 292. [arXiv:1603.05598](#), [doi:10.1140/epjc/s10052-016-4120-y](#).
- [73] M. Aaboud, et al., Performance of the ATLAS Trigger System in 2015, *Eur. Phys. J. C* 77 (5) (2017) 317. [arXiv:1611.09661](#), [doi:10.1140/epjc/s10052-017-4852-3](#).
- [74] V. Khachatryan, et al., The CMS trigger system, *JINST* 12 (01) (2017) P01020. [arXiv:1609.02366](#), [doi:10.1088/1748-0221/12/01/P01020](#).

- [75] [Optimisation and performance studies of the ATLAS \$b\$ -tagging algorithms for the 2017-18 LHC run](#), Tech. Rep. ATL-PHYS-PUB-2017-013, CERN, Geneva (Jul 2017).
URL <http://cds.cern.ch/record/2273281>
- [76] S. Chatrchyan, et al., Identification of b -quark jets with the CMS experiment, JINST 8 (2013) P04013. [arXiv:1211.4462](#), [doi:10.1088/1748-0221/8/04/P04013](#).
- [77] T. A. collaboration, Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s} = 13$ TeV.
- [78] A. M. Sirunyan, et al., Observation of the Higgs boson decay to a pair of tau leptons [arXiv:1708.00373](#).
- [79] C. Collaboration, The potential to study exotic physics signatures at HL-LHC using the phase 2 upgraded CMS detector.
- [80] Expected photon performance in the ATLAS experiment.
- [81] M. Aaboud, et al., Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run-1 data, Eur. Phys. J. C76 (12) (2016) 666. [arXiv:1606.01813](#), [doi:10.1140/epjc/s10052-016-4507-9](#).
- [82] T. Sjostrand, S. Mrenna, P. Z. Skands, A Brief Introduction to PYTHIA 8.1, Comput. Phys. Commun. 178 (2008) 852–867. [arXiv:0710.3820](#), [doi:10.1016/j.cpc.2008.01.036](#).
- [83] R. D. Ball, V. Bertone, S. Carrazza, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. Rojo, Parton distributions with QED corrections, Nucl. Phys. B877 (2013) 290–320. [arXiv:1308.0598](#), [doi:10.1016/j.nuclphysb.2013.10.010](#).
- [84] N. D. Christensen, C. Duhr, FeynRules - Feynman rules made easy, Comput.Phys.Comm. 180 (2009) 1614–1641. [arXiv:0806.4194](#), [doi:10.1016/j.cpc.2009.02.018](#).
- [85] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, 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](#), [doi:10.1007/JHEP07\(2014\)079](#).
- [86] S. Ovin, X. Rouby, V. Lemaitre, DELPHES, a framework for fast simulation of a generic collider experiment [arXiv:0903.2225](#).
- [87] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lematre, A. Mertens, M. Selvaggi, DELPHES 3, A modular framework for fast simulation of a generic collider experiment, JHEP 02 (2014) 057. [arXiv:1307.6346](#), [doi:10.1007/JHEP02\(2014\)057](#).
- [88] M. Cacciari, G. P. Salam, G. Soyez, The Catchment Area of Jets, JHEP 04 (2008) 005. [arXiv:0802.1188](#), [doi:10.1088/1126-6708/2008/04/005](#).
- [89] M. Cacciari, G. P. Salam, Pileup subtraction using jet areas, Phys. Lett. B659 (2008) 119–126. [arXiv:0707.1378](#), [doi:10.1016/j.physletb.2007.09.077](#).
- [90] M. Cacciari, G. P. Salam, G. Soyez, FastJet User Manual, Eur. Phys. J. C72 (2012) 1896, cERN-PH-TH-2011-297. [arXiv:1111.6097](#), [doi:10.1140/epjc/s10052-012-1896-2](#).
- [91] Y. Freund, R. E. Schapire, A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting, J. Comput. Syst. Sci. 55 (1) (1997) 119–139. [doi:10.1006/jcss.1997.1504](#).