LU-TP 20-43
August 2020

Searching for Dark Photons with Jet Substructure

Ashar Kamal

Department of Astronomy and Theoretical Physics, Lund University

Bachelor thesis supervised by Stefan Prestel

LUND
UNIVERSITY

Acknowledgments

A special thank you to my supervisor Stefan Prestel, who has provided me with a fun and interesting project, and whom I have learnt a lot from during this time. I would also like to thank Torbjörn Sjöstrand for spending time to review my project and offering very constructive criticism.

Abstract

Dark matter is the greatest mystery of modern particle physics and cosmology. Although an overwhelming amount of observational cosmological evidence has pointed towards a dark sector, only its gravitational effect has been experimentally confirmed. Currently, no other properties of dark matter are known, including its mass, interactions else than gravitational and potential mediators between dark matter and the particles of the Standard Model. In recent time, the idea of probing dark matter in the sub-GeV mass region with an electromagnetic coupling has become increasingly popular. In this paper, a simplified light dark matter model is adapted as a toy model. We investigate lepton jets as a promising signature for the $U(1)_D$ dark photon mediator A', using an off-the-shelf search strategy widely used by experimentalists at the Large Hadron Collider. At this collider, the $\mathcal{O}(\text{GeV})$ dark boson decays into a pair of leptons, which further go on to creating a leptonic jet signature. The thesis is begun by a study of the dark photon decay kinematics using truth distributions. The kinematical study looks for effective methods to distinguish signal from background by observing the characteristic energy, spatial and momentum distributions of signal leptons. This is followed by a jet substructure analysis including jet filtering, mass drop tagging and sub-jet correlations, where we conclude that the applied off-the-shelf analysis is not as efficient in isolating the dark photon signal as originally considered.

Populärvetenskaplig beskrivning

Standardmodellen är teorin som beskriver tre utav de fyra fundamentala naturkrafterna, och Standardmodellens komplettering har hittills varit partikelfysikens största prestation. Modellen integrerar fysikens mest elementära partiklar och deras interaktioner till en och samma teori. Den kan exempelvis beskriva universums tillstånd en kort stund efter den stora smällen, eller varför protonen inte kan sönderfalla. Däremot kan den inte förklara varför partiklarna har de massorna some de har, eller vad mörk materia är. Därav kan Standardmodellen inte betraktas som en komplett teori, och därmed har sökandet börjat för fysik bortom Standardmodellen.

Trots den stora mängden bevis från kosmologiska observationer, har forskare inget konkret svar på vad mörk materia är. De vanligaste förekommande modeller tyder på svagt interagerande massiva partiklar. Däremot har sökandet för dessa hittills varit utan framgång, vilket har fått forskarna att vända sig till andra modeller. En sådan modell är lättare mörk materia med en svag koppling till Standardmodellens foton.

I denna studie antar vi en väldigt förenklad bild av lätt mörk materia och introducerar en hypotetisk lätt mörk partikel, kallad för en ”mörk foton”. På grund av kopplingen mellan den mörka fotonen och Standardmodellen, kan den mörka fotonen sönderfalla till de kända elementarpartiklarna, och kan på så sätt detekteras experimentellt. Däremot är signalen väldigt otydlig på grund av den stora mängden bakgrund orsakad av växelverkan mellan andra partiklar. I detta projekt studerar vi dynamiken av sönderfallet av den mörka fotonen, samt utvecklar bakgrunds-reducerande metoder för att extrahera den gömda signalen. De bakgrunds-reducerande metoderna baseras på en strategi som använts i stor utsträckning av forskare på den experimentella sidan, och målet med studien är att undersöka strategins effektivitet i att extrahera signalen från mörk materia.

Contents

1 Introduction

The Standard Model (SM) of particle physics offers the most accurate description of our present knowledge of the universe at microscopic distances. It describes three of the four fundamental forces, while incorporating all the known elementary particles. However, it still falls short of being a complete theory, in part due to the incompatibility to observational cosmology. As a result, physicists are looking at physics Beyond the Standard Model (BSM) for potential explanations. Although no direct evidence within the SM suggests that such physics necessarily exists, astrophysical observations have shown lower than expected rotational velocities of stars in galaxies. These observations can be explained by a number of alternatives, most successfully by the existence of some invisible-to-the-eye, excess matter. Due to its invisible nature, the phenomenon is referred to as dark matter (DM) [1] and has in recent times become the central topic of Particle Physics. In this project, production of DM is simulated in scattering of particles at particle colliders, such that its properties can be searched for with new analysis in the future. In order to find the DM, an off-the-shelf search strategy will be applied. The objective of this thesis is to assess the efficiency of this commonly used method.

Due to the hidden nature of DM, direct identification becomes a cumbersome task. There exist a number of candidate models for DM, however all remain unproven so far. One such candidate is the Weakly Interacting Massive Particle (WIMP). WIMPs are hypothetical heavy elementary particles considered to interact via gravity and any other weak force(s), which is as weak or weaker than the weak force, but not necessarily in the SM. Hence the name "weakly interacting" [2]. In recent years, the idea of DM coupling feebly to the electromagnetic (EM) current in the sub-GeV mass range has become increasingly popular. This model has been given the suiting name light dark matter (LDM), the focus of this project.

In order to observe LDM, proton-proton (pp) collisions will be simulated in an environment similar to the Large Hadron Collider (LHC). The LHC is a 27 kilometer circular particle collider. In its latest run, it has collided in the order of a billion particles per second, at center of mass energies of 13 TeV, the highest energy ever achieved by a collider [3]. A sketched event from the LHC is depicted in Fig. 1. Once a quark is emitted from the pp vertex, it will radiate off (mostly collinear) gluons and photons. The interaction with DM occurs when the quark instead radiates off a dark photon A'; the dark photon is a proposed gauge boson of the dark sector, a sector of hypothetical, unobserved particles. The dark photon mediator is an $\mathcal{O}$(GeV) dark particle, coupled to the EM photon γ. Due to this mixing with QED, the dark photon is able to decay back into SM particles via the $A' \to \ell^+\ell^-$ decay channel, where $\ell^\pm$ is a charged SM lepton. Additionally, due to the small dark photon mass compared to the partonic center of mass energy at the LHC, its decay products are enhanced in the collinear direction, creating a narrow jet of particles. The jet

will be enriched in electrons and muons, resulting in a leptonic jet signature [4].

However, as a result of high background activity, the probed DM signal discussed above will be heavily suppressed. Firstly, due to the high interactivity of quarks and gluons, coloured partons will quickly radiate off other coloured partons. The partons are "bound" into colour neutral hadrons and form decay chains by decaying into lighter hadrons, photons and leptons. This high activity of QCD particles causes a significant amount of background noise, refereed to as QCD background [5]. A similar effect occurs for QED background activity, but at a lower rate, where SM photons pair produce electron-positron pairs and the electrons and positrons ionize and annihilate. Additionally, at the LHC, beams consisting of bunches of hadrons are accelerated and collided in order to increase production of the desired heavy particles (dark photons are, disregarding hadrons and heavy quarks, among the heaviest particles in the simulation). As a result, multiple particle collisions occur, an effect called pile-up. Furthermore, since protons are composite particles consisting of quarks and gluons, several parton interactions occur within the same hadron pair collision. Such excess interactions are called underlying events. Back-of-the-envelope calculations [6] show that pile-up and underlying events alone can increase the signal jet mass up to 20% at high luminosity, where the luminosity is a quantity that describes the number of events detected per unit time weighted by the cross section. A high luminosity is preferred when searching for heavy particles with low interaction strength, and hence, jet (sub)structure techniques can help in disentangling the signal from the background.

In section 2, some theory on the LDM model will be covered. This includes the Lagrangian and the previously mentioned mixing between the dark and EM photons. Some relevant substructure techniques will also be discussed. In section 3, we study the kinematics of the dark photon decay. This includes the distribution of the signature lepton pairs around jets and the four-momentum of the associated leptons. This is followed by the application of substructure techniques, with interpretation and discussion on the resulting plots (section 4). Finally, the paper ends with an outlook and a few ideas for future possibilities to extend the work in section 5.

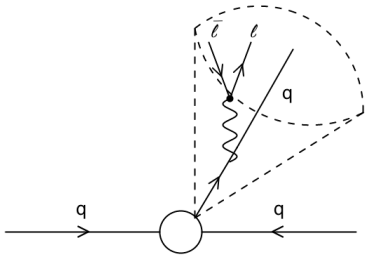

Figure 1: Sketched production of a dark photon via a pp collision at the LHC. An emitted quark from the main vertex radiates off a dark photon, which in turn decays to a charged lepton pair.

2 Theory

2.1 The Light Dark Matter Model

In our model of LDM, two assumptions are made:

- **A light dark gauge boson A'.** A small mass increases the production rate of the dark photons.
- **Dark particle neutrality.** All dark particles must be neutral, or singlets, under the full SM gauge group and are charged under the $U(1)_D$ "dark QED" group. However, the dark photon is kinetically mixed with the SM photon.

Defining the LDM particle χ' and the $U(1)_D$ gauge boson A', the Lagrangian of the dark sector is

$$L \supset -\frac{1}{4}F'^{\mu\nu}F'_{\mu\nu} + \frac{m^2_{A'}}{2}A'^{\mu}A'_{\mu} - A'_{\mu}\big(\epsilon e J^{\mu}_{EM} + g_D J^{\mu}_D\big), \tag{2.1}$$

where ϵ is the kinetic mixing parameter between the groups $U(1)$ and $U(1)_D$, $J^{\mu}_{EM} = \sum_f Q_f \bar{f}\gamma^{\mu} f$ is the SM electromagnetic current, f is a SM fermion with charge Q_f, $g_D = \sqrt{4\pi\alpha_D}$ is the $U(1)_D$ coupling constant and J^{μ}_D is the DM current [7]. The details of the final model are described in [8].

The mixing parameter ϵ, responsible for the degree of mixing between the SM and dark photons, creates the leptonic decay channel $A' \to l^+l^-$ (provided that $\epsilon \neq 0$) (Fig. 1). The dark photon can also decay into other $U(1)_D$ particles via $A' \to \chi\bar{\chi}$ (assuming $m_{A'} > 2m_{\chi}$). This possibility is discussed e.g in [7], but it is not the focus of this paper.

2.2 Geometrical intuition of a scattering event

Events of interest are studied by their final states. Typically, heavy particles cannot be observed directly, since they quickly decay into lighter particles. Hence, the final states are characterized by the (detectable) stable SM particles; muons, electrons, photons and a small number of different hadrons [5]. As discussed in the introduction, the final states are mostly visible as sets of collimated particles. The sets of collimated particles, otherwise known as jets, are detected as regions of focused energy deposits in the detector [6]. The jet size and shape (circular or irregular) are ill-defined properties. Thus, jet algorithms are required to "define" the jets, with each algorithm creating jets of different desired properties. Some of these properties will be covered below. Additionally, due to the high center of mass energies involved at the LHC, the jets often end up overlapping each other, merging into larger jets, also called fat jets. By exploiting the internal substructures of the fat jets, the original sub-jets containing the signal can be recovered. This is done by jet

substructure techniques [9]. Jet algorithms and substructure techniques are covered in the sections below.

The QCD jet description can be extended by looking at the properties of Quantum Field Theory. Beginning with the probability P for a quark to emit a gluon with an energy fraction $z = \frac{E_g}{E_q+E_g}$ at an angle θ:

$$P(z, \cos\theta)dzd\cos\theta = \frac{\alpha_s C_F}{\pi}\frac{dz}{z}\frac{d\cos\theta}{1-\cos\theta}. \tag{2.2}$$

The factor $\alpha_s C_F/\pi$ is the coupling strength between the quark and gluon. A detailed proof of the above expression can be found in [6].

It is immediately clear that the probability in Eq. 2.2 diverges in the soft ($z \to 0$) and/or collinear ($\theta \to 0$) limits. Since detectors measure particles by their energy deposits, a system with a single quark cannot be distinguished from a system with a quark radiating off a gluon with infinitely low energy fraction z (disregarding hadronization. In reality, one cannot detect individual quarks or gluons). The specific gluon energy fraction z required to distinguish the separation between the two systems will depend on the resolution of the detector. Drawing a similar argument for the collinear limit, the probability for a quark to decay into products with small relative angles and very asymmetric energies becomes very large, resulting in the previously discussed sets of collimated particles around high-energy cores, or jets. Overall, singularities are guaranteed to cancel in experimental measurements.

Considering the case where a quark radiates off a SM photon, the emission can again be described by the quark propagator $\frac{1}{z}\frac{1}{1-\cos\theta}$ in expression 2.2, where z is the SM photon energy fraction this time. Similarly as above, the $z \to 0$ divergency demonstrates that (real) SM photons will on average have low energies.

Carrying on to when the SM photon decays into a pair of leptons, its decay rate consists of propagator factors, and an additional dependence on the SM photon energy through the "splitting kernel". The SM photon propagator, in terms of the outgoing lepton momenta, can be expressed as $\sim \frac{1}{(p_{\ell+}+p_{\ell-})^2} = \frac{1}{m^2_{\ell+\ell-}}$, and the splitting kernel is $\sim z^2 + (1-z)^2$, where z is the lepton energy fraction. Due to the propagator structure, SM photons will (on average) decay into lepton pairs with very small $m^2_{\ell+\ell-}$, and due to the splitting kernel, one of the leptons will have a slight preference for either carrying all of the energy, or none of it. However, this enhancement of asymmetric splits is very small, especially when considering the highly asymmetric $\mathrm{q} \to \mathrm{q}\ \mathrm{g}/\gamma/A'$ split. Looking at the dark photon decay, the propagator now includes the rest mass $\frac{1}{(p_{\ell+}+p_{\ell-})^2-m^2_{A',\mathrm{rest}}}$, indicating of a $m^2_{\ell+\ell-}$ close to $m^2_{A',\mathrm{rest}}$ this time. The dark photon splitting kernel remains unchanged from the SM photon one, however, due to Jacobian effects, in particular for the case of final-initial

kinematic mappings, there is a small suppression of the z $\rightarrow$ 0 and z $\rightarrow$ 1 regions for decays of particles with non-zero rest mass. For these Jacobian effects, see [10].

As a result, the energy sharing between leptons in pairs will (on average) be marginally more symmetric in the case of a dark photon decay, compared to the decay of a virtual SM photon with similar mass to the dark photon. However, in distinguishing dark photons from SM photons in the jet substructure, we will mostly rely on the difference of mass distributions instead. Due to its propagator, majority of encountered SM photons will have low virtual mass, causing an asymmetrical energy sharing between its leptonic decay products in the laboratory frame. Dark photons on the other hand, have their mass distributed around 1 GeV, causing a more symmetric energy sharing between its leptonic decay products in the laboratory frame. This technique only filters out virtual SM photons with mass lower than that of dark photon. Once a virtual SM photon is produced with mass similar to that of the dark photon, there will be no difference between the two, except for the very small suppression of the z $\rightarrow$ 0 and z $\rightarrow$ 1 regions for decays of the massive dark photons, as discussed above. Although the size of the overall difference from the z $\rightarrow$ 0 and z $\rightarrow$ 1 suppression is expected to be very small by theoretical considerations, we remain agnostic to the size, and attempt to estimate the concrete smallness of the effect through numerical simulations.

2.3 Jet Algorithms

Although the intuitive definition of a jet as a large energy deposit in a small angular region in a detector is sufficient in some cases, this definition quickly becomes troublesome when jets begin to overlap and "jet borders" become unclear. A well defined analysis requires that there exists a precise jet definition, or in other words, a jet algorithm. Any jet algorithm should preferably have three properties in order to return appropriate jets. These are [11]:

- Collinear and infrared-safe - Completely collinear splittings and infinitely soft emissions should not have an affect on the jets.
- Minimal sensitivity to hadronization, underlying events and pile-up.
- Applicable at detector-level - Good computational performance.

There exist two classes of jet algorithms; cone-type algorithms and clustering algorithms. Cone-type algorithms typically lead to regular jets with simple cone-like geometric structure. Although cone-jets are simpler to apply, they often overlap and are in most cases not infrared and collinear-safe. Clustering algorithms however, can be infrared and collinear-safe, and are highly favoured by theorists as they allow for well-defined calculations. The

algorithms assign individual particles iteratively to jets. By employing a specific jet separation measure to define which particles to cluster together, the same particle is not assigned to multiple jets, ensuring that there is no overlap [12]. The choice of separation parameter is specific to the algorithm.

The Cambridge/Aachen (C/A) clustering jet algorithm has proved to be a top candidate for resolving substructure of jets, once the appropriate jet substructure is applied (section 2.4). Due to the larger size of fat jets, they become prone to carrying uncorrelated soft-energy deposits, mostly from pile-up, underlying events and soft wide-angle radiation [9]. The Cambridge/Aachen jet algorithm is especially well suited for the identification of such wide-angle emissions in jets. The algorithm starts by clustering together particles with smallest angular distances. The separation between two particles j_1 and j_2 is found by

$$\Delta R_{j_1 j_2} = \Delta y^2_{j_1 j_2} + \Delta \phi^2_{j_1 j_2}, \tag{2.3}$$

$$\Delta y_{j_1 j_2} = y_{j_1} - y_{j_2}, \quad \Delta \phi_{j_1 j_2} = \phi_{j_1} - \phi_{j_2}, \tag{2.4}$$

where y is the rapidity and ϕ is the azimuthal angle [6][13].

The pair's new collective four-vector momentum becomes the sum of the original particle four-momenta, $p_j = p_{j_1} + p_{j_2}$. The clustering process is repeated until the constituent separation becomes too large, characterized by $\Delta R_{j_1 j_2} < R$, where R is a parameter of the algorithm. After the clustering procedure terminates, each jet is governed by one four-momentum, and the separation between any pair of jets is large [6][13]. Undoing the clustering steps will yield the jets sub-jets.

2.4 Jet Substructure

Jet substructure techniques, the studies of internal jet structures, are regularly used in quest to disentangle the signal from the background. There exist three classes of substructure techniques relevant to this study; grooming, tagging and sub-jet correlations.

2.4.1 Jet Grooming

Jet groomers, together with the C/A jet algorithm, are aimed towards fat jets. They exploit the fact that contamination from pile-up, underlying events and soft wide-angle radiation, is on average uniformly distributed over the jet area, and on average also low in energy [6]. A commonly used groomer is the jet filterer, an easily applicable technique that returns a jet of size R_{filt}, smaller than the original jet size. It picks the most energetic, or hardest, sub-jets inside the original jet, and creates a smaller sized jet around those

sub-jets, avoiding much of wide-angle radiation. The filtering radius R_{filt} is a parameter of the method, and should be chosen accordingly to the problem [14].

2.4.2 Tagging & sub-jet correlations

The central part of the analysis is the tagging of leptonic jets from decaying dark photons and the study of sub-jet correlations. Tagging is the identification of jets originating from the probed signal, which starts with the mass drop tagger in our substructure. The lepton pairs from decaying dark photons mainly compete against lepton pairs from EM and hadronic showers (recall that we expect a symmetry-asymmetry differentiation in energy between leptons in pairs from SM and dark photons respectively, due to the difference in mass distribution, see section 2.2) and other soft jet emissions from Eq. 2.2. The mass drop tagger utilizes the asymmetric nature of such background jet emissions and (the expected increase of asymmetry of) background lepton pairs (compared to signal lepton pairs). It isolates the signal lepton pair by searching for relatively symmetric sub-jets (leptonic jets) that have masses smaller than that of the original jet (dark photons). The technique is applied in the following steps [14]:

1. Undo the last step of the C/A clustering by breaking the jet j into the sub-jets j_1 and j_2, such that $m_{j_1} > m_{j_2}$.
2. If there is a significant mass-drop $\mu \equiv m_{j_1}/m_j > \mu_{cut}$ and the splitting is not too asymmetric $y = \min(p^2_{tj_1}, p^2_{tj_2})\Delta R^2_{j_1 j_2}/m^2_j > y_{cut}$, then j is a tagged heavy particle.
3. If 2. is not fulfilled, redefine j_1 as the new j and repeat.

Again, $\Delta R_{j_1 j_2}$ is the angular separation between j_1 and j_2, μ_{cut} and y_{cut} are parameters of the method, describing the mass drop and asymmetry between the sub-jets.

The mass drop tagger can be understood by a simple illustration. To begin with, assume a fat jet, clustered as shown in Fig. 2 (left). Here, P_1 and P_2 represent the momenta of a signal lepton and antilepton respectively, and $P_{12} = P_1 + P_2$ is the momentum of the clustered signal lepton pair. Since the invariant mass $m^2_{ik} = (P_i + P_k)^2$ is non-zero if there is a non-vanishing angle between P_i and P_k, we know that $m_{12} > 0$. P_3 represents a single light parton inside the clustered jet and P_j represents the full clustered jet.

Applying the mass drop tagger, we begin with the branching of P_j into $P_{12}+P_3$, i.e undoing the last clustering step of the jet algorithm, see Fig. 2 (center). Considering the assumption that P_3 is a light parton, we have $m(P_{12}) > m(P_3)$ $(m(P_3) \approx 0)$ and asymmetric energy sharing. No mass drop is found between $m(P_{12})$ and $m(P_j)$, and so the mass drop tagger continues to the branch with higher invariant mass, the $P_{12} \to P_1 + P_2$ branching (Fig. 2 (right)) in this case.

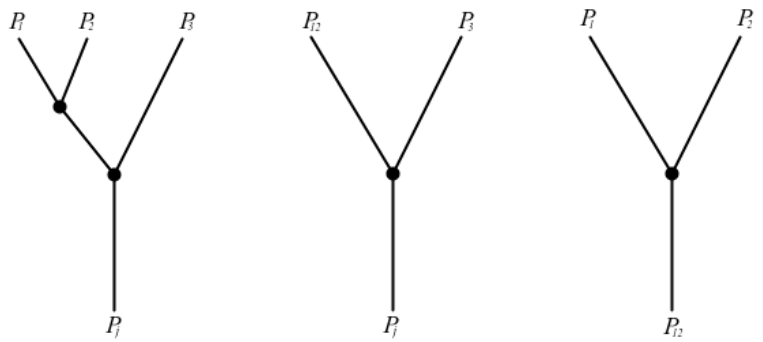

Figure 2: An illustration of a clustered jet (left) and the mass drop tagger (center and right).

In this clustering, which leads to P_{12}, the leptons with momenta P_1 and P_2 have negligable masses, $m_1 = m_2 \approx 0$, causing a significant mass drop with $m_1 << m_{12}$. Additionally, the energy is (on average) shared much more more symmetrically, triggering the mass drop tagger.

These results are obtained using three assumptions. Firstly, that the mass drop tagger moves along the correct branch when checking P_{12} vs. P_3 in Fig. 2 (center). This condition is fulfilled as long as $m_{12}^2 = (P_1 + P_2)^2 > P_3^2 = m_3^2$. A larger invariant mass means that the leptons have larger angular separation at commensurate energies, or larger overall energy than the other particles. A numerical study helps to determine the validity of this assumption. If P_3 is a composite momentum (e.g. $P_3 = P_4 + P_5$+...), then the assumption does not hold either. If the clustering $P_1 + P_2 \to P_{12}$ happens early in the jet reconstruction (i.e. before particles are clustered to P_3), the assumption holds. Again, it is useful to investigate these competing effects numerically.

Secondly, it is assumed that the drop in mass between P_{12} and P_1 is significant. This condition can be regulated by the mass drop parameter μ_{cut}. Finally, we assumed that $m_1 = m_2 \approx 0$. If e.g P_2 is no longer massless, $m_2 > 0$, then the mass drop will not be observed in this particular branching. This occurs if energetic particles or particles with large angular separation relative to P_2 are clustered to P_2. An example of this would be high-energy photons that had been radiated from the leptons. This clustering is not very likely, in particular if the clustering happens at the early stages of the jet algorithm (where average angles and energies are still moderate).

Note that the above argument relies on the invariant mass $m_{12}^2 = P_{12}^2$. Thus, an "average" SM photon-to-lepton branching will not necessarily fulfill $m_{12} > m_3$, whereas an "average" dark-photon-to-lepton branching would. However, for SM photons or dark photons with the same virtuality (i.e. filling the same $m_{\ell\ell}$ mass bins), the result of the mass drop tagger

will be identical. Thus, it became clear, during the literature review in the process of this thesis, that a mass-drop tagger will most likely not improve the signal-to-background discrimination when only looking at the mass bin ≈ 1 GeV. The numerical assessment below (section 3.5) was useful in coming to this conclusion, and confirming the argument.

The remaining background of coloured particles could be removed by dilepton tagging. At the LHC, the detectors are equipped with trackers and EM calorimeters. When an electron enters the detector region, it leaves a track in the tracker due to its electric charge, and forms an EM shower in the calorimeter, where it produces photons and other electrons/positrons due to ionization. Due to the muon's low interactivity, it is the only particle that reaches the outermost part of the detector, the muon chamber, leaving a track in all the other parts [15]. These characteristics are used for lepton tagging. Due to the large mass of the tau lepton, the $\sigma(A' \to \tau^-\tau^+)$ cross section is so small that its contributions to the signal are negligible, and so it is omitted from the tagging. The dilepton tagging is combined with the mass drop tagger in section 4.

Sub-jet correlations are studied by looking at the appropriate observables, distributions by which the signal can be distinguished from the background. Considering that the mass was one of the defining parameters of the dark photon in this proxy model, the lepton pair invariant mass is likely the most appropriate observable. Additionally, due to the high energy of the signal lepton pair, applying an appropriate lower energy constraint E_{cut} should remove some of the low energy background.

2.5 Simulation Environment

All (10 million) events are generated by Pythia 8 [16] [17] with the Dire shower plugin [18] and analyzed using Rivet [19]. We employ an LHC setup of scattering protons at center-of-mass energy $E_{\text{CM}} = 13$ TeV, focusing on [proton+proton] $\to$ [parton+parton] scatterings with a generation cut of $p_{\perp,\text{min}} = 200$ GeV on the outgoing parton transverse momentum. Only tree-level diagrams are considered. The strong coupling constant is set to $\alpha(M_Z) = 0.120$ with the QED and dark photon showers cut-off at $p_\perp = 0.001$ GeV, using the NNPDF2.3 parton distribution function set. Since the lepton masses m_e and m_μ are comparably small to other masses and energies in the simulation, they are set to zero, and a dark photon with mass $m_{A'} = 1$ GeV has been introduced into the event generator environment. Further, hadronization has been disabled in the simulation environment, meaning that individual quarks are observed in the final states instead of colour neutral hadrons. This simplifies the analysis in a number of ways. Firstly, QCD background is much more easily filtered by simply avoiding coloured particles, since now there do not exist any colour neutral QCD particles in our simulation. Secondly, with the exclusion of heavy baryons, the dark photon can be regarded as one of the heaviest particles from the pp collisions, which can be used in the analysis later. Finally, disregarding hadronization, when plotting the invariant mass distribution we do not have to be concerned with hadrons

that have mass close to that of the dark photon and also decay into lepton pairs, e.g the ρ and ϕ vector mesons. This can be seen as a "best-case scenario". If dark photons cannot be found in this simplified environment, then extracting a signal in a more realistic setup will pose a significant challenge.

Two input parameters are required in our LDM adaption; the dark photon mass $m_{A'}$ and the mixing parameter ϵ from the Lagrangian in Eq. 2.1. The dark photon mass is motivated on the one hand by the large QCD activity at low energy scales, favouring a heavy dark photon for analysis prototyping. On the other hand, a light dark photon is desired for an increased production rate. Thus, the $m_{A'} = 1$ GeV dark photon avoids much of the QCD background, while still having a large production rate. Other than this, the choice of the dark photon mass is relatively free. The dark photon properties remain the same, although the properties of its decay products, such as their energies, depend on the initial mass.

The mixing parameter is set to $\epsilon = 0.5$, such that the dark QED fine structure constant becomes $\alpha_D \sim \epsilon^2\alpha = 0.25\alpha$, where α is the QED fine structure constant. The possibility of such a high ϵ in reality is excluded in most simple models, see e.g [20] for details, but the kinematics remain the same regardless of the mixing parameter. A high ϵ simplifies testing and development, given a finite sample size. Once the signal has been isolated, ϵ can be adjusted to more realistic values in attempt to isolate the signal in an environment with less signal to background ratio. In [7] for instance, they use several ϵ values within an interval of $[10^{-4} - 10^{-1}]$. Once the signal is found for a realistic signal to background ratio, the analysis can be generalized to more complicated LDM models.

3 Kinematics and Distributions at truth level

In the previous section, a number of free parameters of the analysis were introduced; the filtering radius R_{filt} (section 2.4.1), the mass drop and p_T symmetry variables μ_{cut} and y_{cut} (section 2.4.2) and the lepton energy constraint E_{cut} (Section 2.4.2). The parameter choices must be deduced before the application of the substructure techniques, and are usually done by theoretical calculations on the particle decay kinematics.

Since this is not the focus of this project, the kinematical study in this paper is mostly done by observing "truth" variables and looking at their distributions (truth distributions). A truth variable extracts information directly from the Rivet output, without using substructure methods. As a result, truth variables carry no uncertainties and represent the "true" distributions of an observable. These are commonly used tools since they offer a deeper insight to the system properties, and allow for a "presearch" for the probed signal. However, truth variables are only a feature in computer generated events, where all the

information is already known by the software tool.

In the following section, we define a "lepton truth" variable, which extracts all the leptons in the final state from the pp collisions, to study the dark photon decay kinematics. Since the dark photon signal is leptonic, the lepton truth variable is appropriate to obtain a rough estimation of the optimal parameter values for R_{filt} and y_{cut}. The remaining parameters μ_{cut} and E_{cut} are estimated by an intuitive approximation and a quick back-of-the-envelope calculation respectively. In section 4, a substructure analysis is applied with the deduced parameter values, to find the DM signal. The truth variable will not be used in the analysis itself, but knowing the truth is useful for the development of the analysis.

3.1 The dark photon resonance at truth level

Since the lepton truth variable tags all final state leptons, in addition to the lepton pair signal, it also includes the leptonic background, e.g from QCD and QED showers, as discussed in the earlier section. Considering the large mass of the dark photon (compared to the most commonly appearing particles in pp collisions when hadronization is off, e.g electrons, photons and up quarks), it can be hypothesized that the leptons from decaying dark photons will generally be constituents of the hardest jets from the pp collisions. However, exactly which jet, or which lepton pair in the jet is not known with certainty.

Disregarding the jets momentarily, the lepton signal can be identified by plotting the invariant mass of all leptons in the system, i.e the lepton truth variable. Since the dark photon signal is a pair of hard leptons, the invariant mass of different combinations of hard lepton pairs in the entire system were checked. Recall the difference in the lepton pair mass distributions between SM and dark photons, caused by the different propagators in section 2.2. The heavy lepton pair is much more likely to be caused by an on-shell dark photon than an off-shell SM photon. Currently, lepton pairs are not required to fulfill any charge or flavour constraints, and are paired based on their energy. The lepton pair with the hardest and second hardest leptons showed the clearest peak, plotted in Fig. 3. Although there is a large background activity due to the non-discrimination against background leptons by the truth variable, a slight "bump", or resonance, can be seen in the graph at the expected dark photon mass. The resonance is greatly amplified by the current high mixing parameter ϵ. Doing a background estimation, i.e applying the same analysis to events where the hypothetical dark photon has not been included, shows that the resonance only appears in the events including the signal. This supports that the resonance in fact is from the dark photon. Note that the graphs have been normalized to unit area, explaining why the background is greater than the background+signal at small mass.

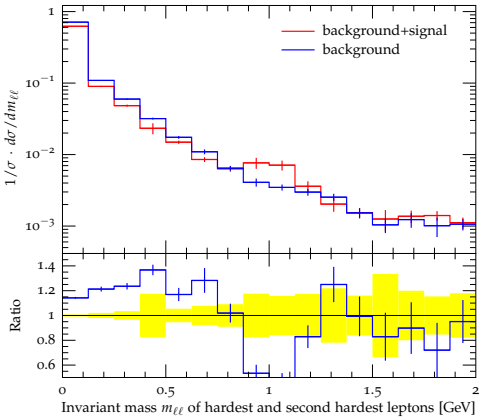

Figure 3: Truth distribution of lepton pair events from pp collisions at LHC. Distributions include background (blue) and background+signal (red) events.

3.2 Energies of signal leptons

Assuming an $m_{A'} = 1$ GeV dark photon, by conservation of energy

$$E_{A'} = E_{\ell^-\ell^+} = E_{\ell^-} + E_{\ell^+}, \tag{3.5}$$

where $E_{A'}$ and $E_{\ell^-\ell^+}$ are the dark photon and lepton pair total energies, signal leptons must have $E_\ell = 0.5$ GeV each in the dark photon reference frame. However, the signal is observed in the laboratory reference frame, so the lepton energies are expected to be shifted. This agrees very well the leptonic jet signature picture discussed earlier. The initial boost of the dark photon will cause a jet formation, since the boost makes the two leptons come closer in angle.

3.3 Constructing the jets

Since hadronization is off, a set of final states consisting of leptons, quarks, photons and the dark photons are detected, which are clustered together into jets using the C/A jet algorithm (see section 2.3). The result is a set of jets separated by large distances. Somewhere inside the jets, dark photon candidates are hidden among background jet particles. The decay products of these dark photons will isolated in the analysis in section 4.

3.4 Spatial distribution of LDM inside jets

To avoid cutting away any dark photon candidates in the jet grooming part of the analysis, one needs an appropriate filtering radius R_{filt}. The radius can be found by observing the

spatial distribution of the DM inside the constructed jets. Thus, the distribution of the lepton signal is studied, once again using the lepton truth variable.

In section 3.1, it was deduced that the lepton signal consists of the hardest lepton pair (hardest and second hardest leptons) in the jet. In Fig. 4, separations between the hardest lepton pair (found by the lepton truth variable in section 3.1) and the two hardest jets (clustered by the C/A algorithm in section 3.3) have been plotted. The LDM distribution inside the two hardest jets is studied since the high-energy signal lepton pair is most likely part of the most high-energy jets. In Fig. 4 (left), ΔR is plotted between the two hardest jets, showing them to be back-to-back, which intuitively can be agreed on being correct. In Fig. 4 (right), ΔR is plotted between the hardest lepton pair and the hardest jet. The distribution dominates at $\Delta R = 0, \pi$. Considering the back-to-back orientation between the two jets, it can be concluded that the dark photon candidate is distributed around the jet axes with a separation of $\Delta R < 1.0$ to each jet, where the concentration of lepton pairs drops by a factor 10.

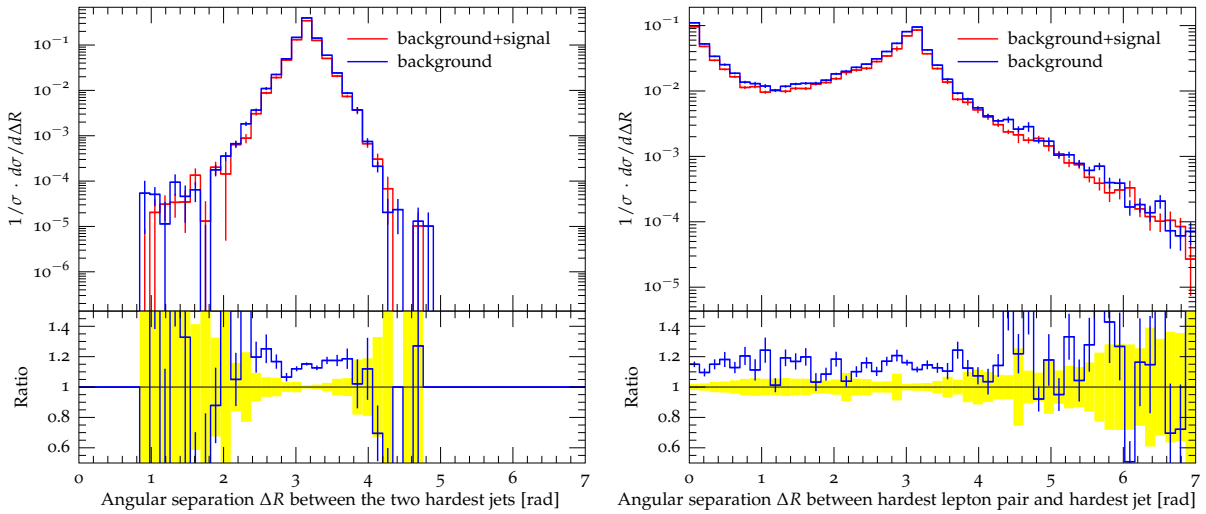

Figure 4: Truth distributions of angular separations ΔR in pp collisions at LHC. Distributions include background (blue) and background+signal (red) events.

3.5 Lepton Asymmetry

The final part of the kinematical study is the momentum asymmetry study between leptons in lepton pairs. In section 2.2 it was explained that if we do not restrict ourselves to a specific (SM/dark) photon mass, we on on average expect a marginally higher momentum asymmetry between leptons in lepton pairs from SM photons compared to leptons from dark photons. The mass drop tagger will exploit any asymmetry of background lepton pairs, in addition to the (on average) large(r) mass drop seen when dark photons decay

into two leptons, described by the y_{cut} and μ_{cut} parameters respectively (see sections 2.2, 2.4.2). In order to find the optimal momentum asymmetry separation between signal and background, the p_T asymmetry is studied by the weighted lepton energy difference $y' = \frac{\text{lep[0]-lep[1]}}{\text{lep[0]+lep[1]}}$ and the ratio $R = \frac{\text{lep[0]}}{\text{lep[1]}}$, where lep[0] and lep[1] are the hardest and second hardest leptons' energy in the jet, found by the lepton truth variable. In Fig. 5, the asymmetry is plotted for background and background+signal leptons. The two graphs show the asymmetry in two different pictures, where in Fig. 5 (left) the actual lepton energies are considered, unlike 5 (left) where only the ratio is studied. For symmetric lepton pairs, the graphs should be distributed around $y' = 0$ and $R = 1$. Both background and background+signal leptons follow very similar (within the margin of error) asymmetric distributions, even though it was originally hypothesized that the signal lepton pairs should be relatively symmetric in momentum. Hence, the asymmetry difference between background and signal lepton pairs is expected to be very small.

By a quick study of the lepton energies, it is found that the energies involved are much greater than the dark photon mass, meaning that the entire system is highly boosted and the asymmetric-symmetric differentiation between background and signal is "blurred out", supported by Fig. 5. This is explained by the large cut-off $p_{\perp,\text{min}} = 200$ GeV, set due to the large center of mass energies involved in the pp collisions. Both graphs in Fig. 5 agree very well on the large overlap between background and background+signal, indicating that if there does exist an optimal y_{cut} parameter that distinguishes lepton pairs on their symmetry, it will isolate the signal at the expense of decreasing the signal rate. By some manual optimization by scanning several parameters showed that there did exist optimal value, found to be $y_{\text{cut}} = 0.70$. Regarding the mass drop parameter, the signal $A' \to \ell^-\ell^+$ mainly competes against $\gamma \to \ell^-\ell^-$. A reasonably high mass drop condition that discriminates against the average SM photon decay is sufficient. The parameter value $\mu_{\text{cut}} = 0.65$ was found to be an appropriate fit. Although we expect (off-shell) SM photons and dark photons to behave similarly, we rely on an increased signal-to-background rate in the natural dark photon mass range (1 GeV in this case).

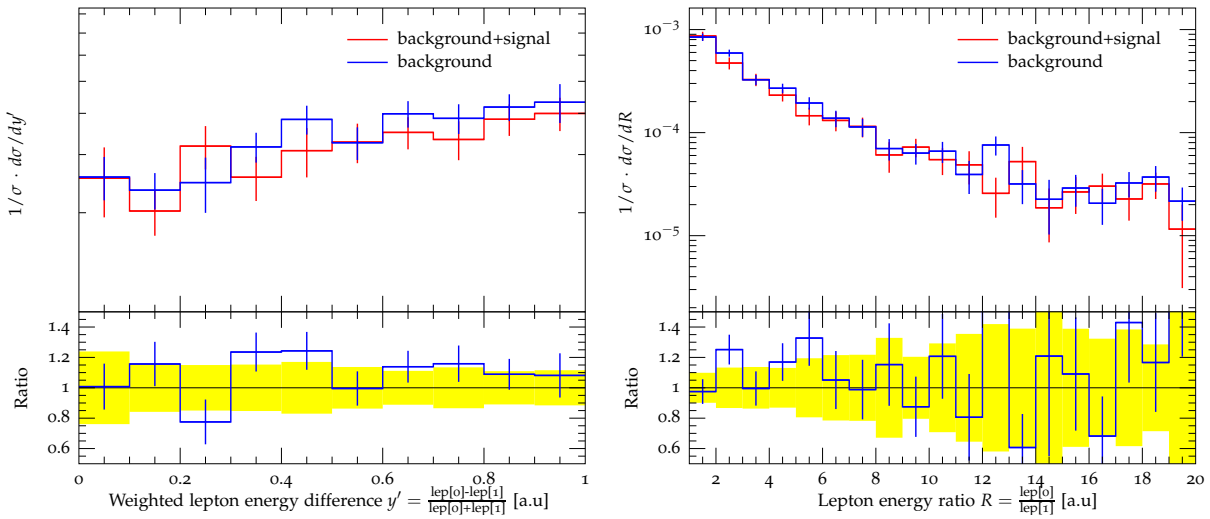

Figure 5: Truth distributions of lepton asymmetry in pp collisions at LHC. Distributions include background (blue) and background+signal (red) events.

4 Simulations and Implementation of analysis

With the parameters of the analysis deduced and some knowledge on the dark photon decay kinematics, a well defined substructure analysis can be developed. Unlike the lepton truth distribution in Fig 3, the substructure will be applicable in collider experiments, and will focus on being less susceptible to background. The analysis is summarized into the following steps:

1. **The hardest jet is extracted from the collection of C/A clustered jets** The boosted lepton signal from a dark photon is most likely found in a hard jet (Fig. 4 (left)).

2. **Jet is groomed using $R_{filt} = \min(1.0, R)$, where R is the original jet radius**. Dark photon candidates are distributed around the jet axis with $R < 1.0$ (Fig. 4 (right)).

3. **Mass drop tagger is applied on the jet, with parameters $\mu_{\text{cut}} = 0.65$ and $y_{\text{cut}} = 0.70$** . A large p_T asymmetry was observed for the signal, yielding a high y_{cut} parameter. Due to the large dark photon mass, its decay to leptons should be distinguished from the average SM photon decay by observing the mass drop μ (section 3.5).

4. **Jets with less than three "constituents", or sub-jets, will be vetoed**. Including the jet core itself, a jet including a signal lepton pair should consist of at least three constituents.

5. **Lepton pairs are tagged.** Rivet supplies an ID tool, which allows for efficient tagging of particles. The sub-jets are tagged for e^-e^+ and $\mu^-\mu^+$ pairs.

6. **Leptons in pairs with energies less than 0.5 GeV are vetoed.** Due to the large mass difference between the dark photon and leptons, the signal is high in energy (section 3.2).

7. **The sub-jets' (containing the lepton pair) invariant mass is filled into histograms.**

In Fig. 6, the jets before and after jet substructure have been plotted. In these graphs, QCD interactions have been omitted for development purposes, drastically decreasing background (no QCD background). In Fig. 6 (left), the invariant mass of the C/A clustered jets from section 3.3 are filled into the histogram. This illustrates the signal without substructure. In Fig. 6 (right), substructure without lepton tagging is applied, showing a promising signal of the dark photon. By the background estimation, it can be concluded that the signal indeed is caused by the dark photon decay. Note that the position of the peak is shifted to smaller masses. This effect might be due to filtering removing particles, most likely photons radiated from decayed leptons, such that on average, the jet masses drop slightly.

Upon applying lepton tagging, it is observed that the events from the signal region disappear. A quick study on the particle type, or ID, in the signal region shows that there are no leptons. Upon further investigation, the issue seems to be originating from the jet algorithm not transmitting the ID information correctly. We hypothesize that when this specific algorithm clusters two different particle types, the clustered particles do not have a well-defined common particle type, and so the algorithm does not assign a particle ID to the pair (designated by the particle ID zero). This is supported by Fig. 6 (bottom), where the particles are tagged for the zero particle ID instead of leptons. The resonance does not disappear this time. The missing ID is not a question of a computational error, but a fundamental question of the jet algorithm method itself. Currently, when the algorithm encounters two different particle types, there does not exist a procedure that tells the algorithm what particle ID to assign the pair. Hence, the "error" encountered here would also show up in real-life experiments, especially if particles are not well isolated from each other.

In Fig. 7 (left), the C/A clustered jet invariant mass is plotted once again, but this time for events including QCD interactions, i.e background from (perturbative) QCD parton showers is included (but still no hadronization). On the right, substructure without lepton tagging is applied, and immediately it is clear that the signal is hidden behind the large QCD activity. Applying the zero ID tagging, a small peak emerges in the Fig. 7 (bottom) graph. The peak could indicate the existence of a dark photon, but no conclusion can be drawn with certainty. Due to the large margin of error, in a sample of 10 million events, the peak cannot be excluded from being a statistical fluctuation.

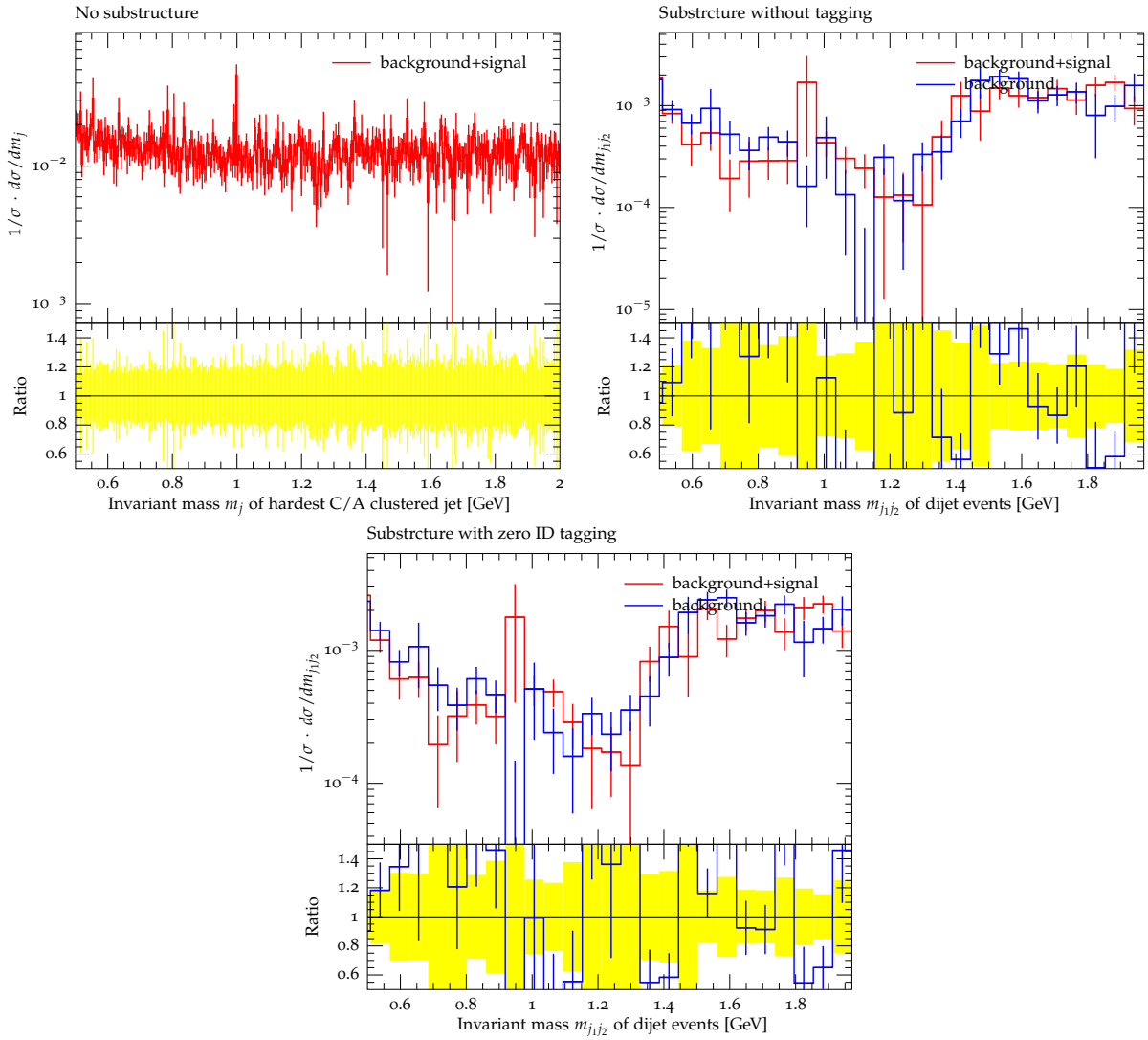

Figure 6: Invariant mass $m_{j_1 j_2}$ of dijet events from pp collisions at LHC, excluding QCD interactions. Distributions include background and background+signal, at different stages of substructure.

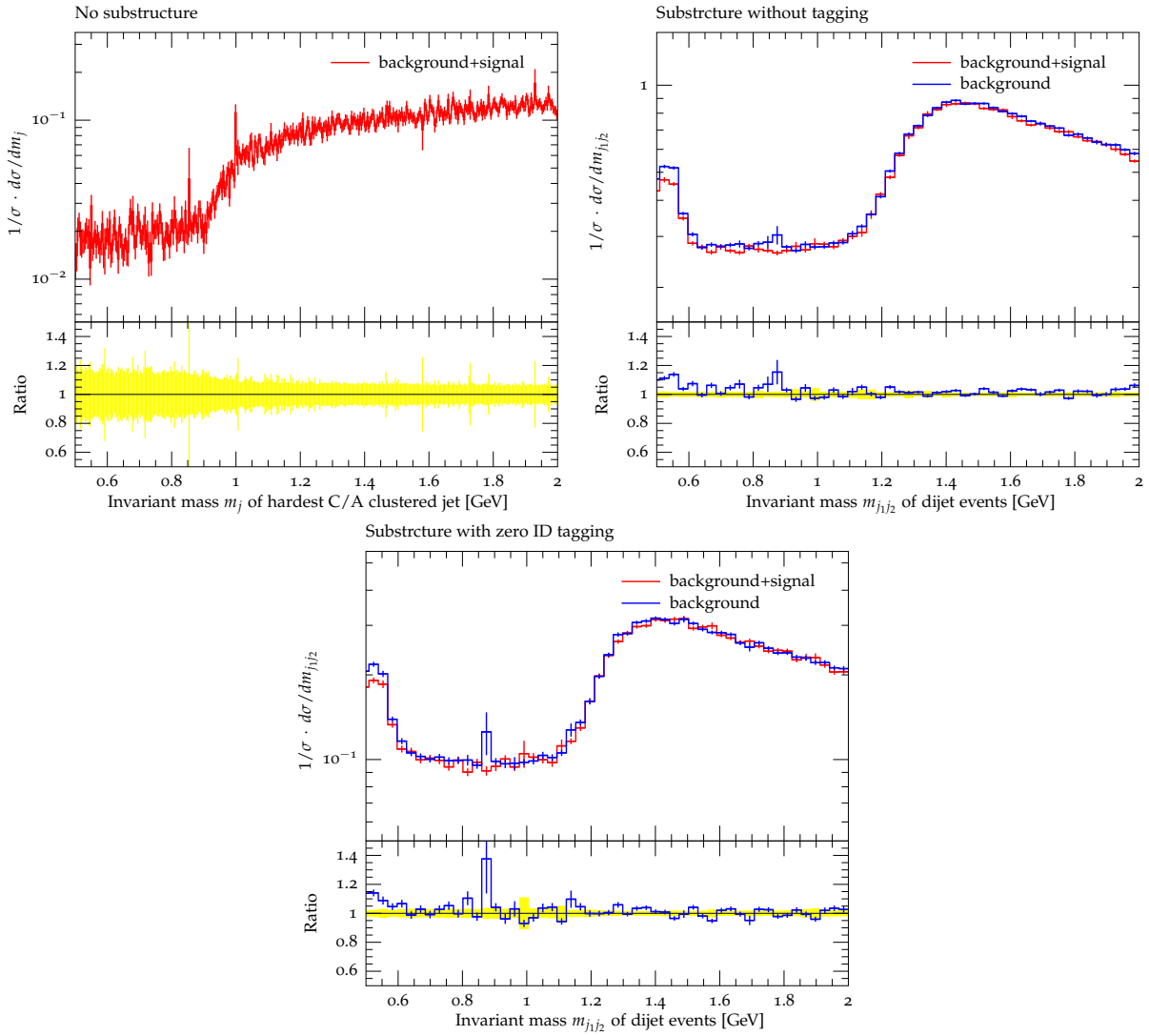

Figure 7: Invariant mass $m_{j_1 j_2}$ of dijet events from pp collisions at LHC, including QCD interactions. Distributions include background and background+signal, at different stages of substructure.

5 Outlook

In summary, we have looked at a simplified LDM model, in search of the $U(1)_D$ gauge boson. Due to the mixing between $U(1)$ and $U(1)_D$, the signal is characterized by a pair of boosted leptons. In the first part of the thesis, Rivet's powerful tools are used to assess properties on the leptonic jet signature. Firstly, the lepton pairs' invariant masses were studied, showing that the leptonic jet signature consists of the most energetic lepton pair in the jet. It was also deduced that the signal lepton pairs are distributed relatively close to the jet axes, and the leptons inside the pair are unexpectedly highly asymmetric in momentum due to the initial boost of dark photons from high-energy pp collisions. In the second part, an analysis is developed according to the observed kinematics of the dark photon decay. A resonance is successfully seen in the QED events with jet grooming and mass drop tagging. After a background estimation on the QED events, the resonance is proven to be the dark photon. In the events including QCD, a clear signal cannot be reconstructed, mainly due to the lepton tagging failing and the inefficient mass drop tagging. The lepton tagging issue was traced back to the C/A jet algorithm, where it is believed that particles' ID information is discarded. Furthermore, due to the overall boost of the system caused by the high center of mass energy, the symmetry-asymmetry differentiation between signal and background exploited by the mass drop tagger decreases. The difficulty in extracting the signal from the QCD background with the C/A jet algorithm and mass drop tagging, considering that the tagger must be applied with the C/A algorithm (since the tagger depends on the clustering history), suggests that new substructure methods might be necessary to search for dark particles at the LHC.

Regarding future possibilities, simulations of this kind are a valuable part of search for new particles. In addition to returning quicker results and more control over the substructure, simulations with toy models can be used to isolate certain aspects of complicated models, as done in this paper. The LDM model adapted in this thesis is a very simplified scenario. For instance, it does not include the remaining $U(1)_D$ particles, or explain where the dark photon mass originates from. However, this model does serve as a good proxy, since dark photons appear readily in BSM physics, with SUSY being one. Another such model is discussed in [21], where in addition to the kinetic mixing between the hidden sector and the SM, a Higgs mixing to the hidden sector is introduced as well. The paper discusses the possibility of detecting dark photons in exotic higgs boson decays, including $h \to A'Z \to 4\ell$ and $h \to A'A' \to 4\ell$.

References

[1] Cern, "Dark Matter [internet]," 2020. https://home.cern/science/physics/dark-matter.

[2] E. Britannica, "Weakly interacting massive particle [internet]," 2018. https://www.britannica.com/science/weakly-interacting-massive-particle.

[3] Cern, "Facts and figures about the LHC [internet]," 2020. https://home.cern/resources/faqs/facts-and-figures-about-lhc.

[4] M. Buschmann, J. Kopp, J. Liu, and P. A. N. Machado, "Lepton Jets from Radiating Dark Matter," *JHEP*, vol. 7, p. 45, 2015.

[5] S. D. Ellis, T. S. Roy, and J. Scholtz, "Jets and Photons," *Phys. Rev. Lett.*, vol. 110, no. 12, p. 122003, 2013.

[6] A. J. Larkoski, "An Unorthodox Introduction to QCD," 2017.

[7] T. Åkesson *et al.*, "Light Dark Matter eXperiment (LDMX)," 2018.

[8] J. D. Clarke, "Constraining portals with displaced Higgs decay searches at the LHC," *JHEP*, vol. 10, p. 61, 2015.

[9] D. Kar, "Jet substructure: a discovery tool," 2015. https://s3.cern.ch/inspire-prod-files-4/40bc7db1050f2a1b8bdb8f5664a326ad.

[10] S. Catani, S. Dittmaier, M. H. Seymour, and Z. Trocsanyi, "The Dipole formalism for next-to-leading order QCD calculations with massive partons," *Nucl. Phys. B*, vol. 627, pp. 189–265, 2002.

[11] P. Schieferdecker, "Jet Algorithms [internet]," 2009. https://twiki.cern.ch/twiki/bin/viewfile/Sandbox/Lecture?rev=1;filename=Philipp.Schieferdeckers_Lecture.pdf.

[12] S. Catani, Y. L. Dokshitzer, M. Seymour, and B. Webber, "Longitudinally invariant K_t clustering algorithms for hadron hadron collisions," *Nucl. Phys. B*, vol. 406, pp. 187–224, 1993.

[13] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, "Jet substructure as a new Higgs search channel at the LHC," *Phys. Rev. Lett.*, vol. 100, p. 242001, 2008.

[14] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual," *Eur. Phys. J. C*, vol. 72, p. 1896, 2012.

[15] M. R. Vidal, X.c., "Taking a closer look at LHC [internet]," n.d. https://www.lhc-closer.es/taking_a_closer_look_at_lhc/1.detectors.

[16] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual," *JHEP*, vol. 05, p. 026, 2006.

[17] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An introduction to PYTHIA 8.2," *Comput. Phys. Commun.*, vol. 191, pp. 159–177, 2015.

[18] S. Höche and S. Prestel, "The midpoint between dipole and parton showers," *Eur. Phys. J. C*, vol. 75, no. 9, p. 461, 2015.

[19] A. Buckley *et al.*, "Rivet user manual [internet]," n.d. `https://rivet.hepforge.org/rivet-manual.pdf`.

[20] M. Battaglieri *et al.*, "US Cosmic Visions: New Ideas in Dark Matter 2017: Community Report," 7 2017.

[21] D. Curtin, R. Essig, S. Gori, and J. Shelton, "Illuminating Dark Photons with High-Energy Colliders," *JHEP*, vol. 02, p. 157, 2015.