Available on the CERN CDS information server **CMS PAS EXO-19-011**

CMS Physics Analysis Summary

Contact: cms-pag-conveners-exotica@cern.ch

2019/11/01

A deep neural network-based tagger to search for new long-lived particle states decaying to jets

The CMS Collaboration

Abstract

The development of a tagging algorithm to identify jets that are significantly displaced from the luminous regions of LHC proton-proton (pp) collisions is presented. Displaced jets can arise from the decay of a long-lived particle (LLP), which are predicted by several theoretical extensions to the standard model. The tagger is a multiclass classifier based on a deep neural network, which is parameterized according to the proper decay length $c\tau_0$ of the LLP. A novel scheme is defined to reliably label jets from LLP decays for supervised learning. Samples of both simulated events and pp collision data are used to train the neural network. Domain adaptation by backward propagation is performed to improve the simulation modelling of the jet class probability distributions observed in pp collision data. The tagger is applied in a search for long-lived gluinos, a manifestation of split supersymmetric models. The tagger provides a rejection factor of 10 000 for jets from standard model processes while maintaining an LLP jet tagging efficiency of 30−80% for split supersymmetric models with $1\,\mathrm{mm} \leq c\tau_0 \leq 10\,\mathrm{m}$. The expected coverage of the split supersymmetric model parameter space is presented.

1 Introduction

Machine-learned algorithms are routinely deployed to perform event reconstruction, particle identification, event classification, and other tasks [1] when analysing data samples of proton-proton (pp) collisions recorded by experiments at the CERN LHC. Machine learning techniques have been widely adopted to classify a jet, a collimated spray of particles that originate from the hadronization of a parton, according to the underlying flavour of the initial parton [2]. For example, jets that originate from the hadronization of b quarks (b jets) exhibit characteristic experimental signatures that can be exploited by dedicated algorithms in a procedure known as b tagging. The b hadrons, with lifetimes of $\mathcal{O}(10^{-12}\,\mathrm{s})$, typically travel distances of approximately 1–10 mm, depending on their momenta, before decaying. As a result, charged particle tracks in jets can originate from one or more common vertices that may be displaced with respect to the luminous region of pp collisions. Further, the impact parameter of each track, defined as the spatial distance between the originating pp collision and the track at its point of closest approach, can have a significant nonzero value. The ATLAS [3] and CMS [4] Collaborations have developed numerous algorithms based on boosted decision trees or neural networks to identify b jets [5, 6] using the aforementioned and other high-level engineered features. The latest b tagging algorithm developed by the CMS Collaboration is the DEEPJET tagger [7, 8], which is a multiclass classifier that discriminates b jets from c and light-flavour jets with unprecedented performance. The algorithm is based on a deep neural network (DNN) that exploits low-level information, such as the kinematical properties of the jet constituents, as well as engineered features used in preceding b tagging algorithms [5, 6].

Various theoretical extensions to the standard model (SM) [9–14] predict the existence of long-lived particles (LLPs) with a proper lifetime τ_0 that can be very different from those of known SM particle states. Consequently, the production and decay of LLPs at the LHC could give rise to atypical experimental signatures. These possibilities have led to the development of a broad signature-based search programme at the LHC, based around LLP simplified models [15, 16] and novel reconstruction techniques. A comprehensive review of LLP searches at the LHC can be found in Ref. [17].

In this note, we present the novel application of a DNN to identify (i.e. tag) a jet originating from the decay of an LLP (LLP jet). The charged particle tracks of an LLP jet may be significantly displaced with respect to the luminous region of pp collisions within the CMS detector. The DNN is trained and evaluated using a range of signal hypotheses comprising simplified models of split supersymmetry (SUSY) [9, 10] with R-parity [18] conservation. These models assume the production of gluino ($\widetilde{\mathrm{g}}$) pairs. The gluino is a long-lived state that decays to a quark-antiquark ($\mathrm{q}\overline{\mathrm{q}}$) pair and a weakly interacting and massive neutralino ($\widetilde{\chi}^0$), which is the lightest SUSY particle and a dark matter candidate. Simplified models of split SUSY are widely used to benchmark LLP searches in final states containing missing transverse momentum $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ and jets [19–29].

The LLP jet tagger is inspired by the DEEPJET approach, albeit with significant modifications. The DNN extends the multiclass classification scheme of the DEEPJET algorithm to accommodate the LLP jet class. A procedure to reliably label LLP jets using truth information from Monte Carlo (MC) event generators is defined. The experimental signature for an LLP jet depends strongly on $c\tau_0$. Hence, a parameterized approach [30] is adopted by using $c\tau_0$ as an external parameter to the DNN, which permits hypothesis testing using a single network for models with values of $c\tau_0$ that span several orders of magnitude. Further, the jet momenta depend strongly on the model parameters $m_{\widetilde{\mathrm{g}}}$ and $m_{\widetilde{\chi}^0}$, particularly the mass difference $m_{\widetilde{\mathrm{g}}} - m_{\widetilde{\chi}^0}$. The DNN training is performed with simulated event samples drawn from the full mass parameter space of interest to ensure a broad generalization and optimal performance over a range of jet momenta. The out-of-sample performance of the tagger is quantified for simplified models of SUSY with gauge-mediated SUSY breaking (GMSB) [12] and weak R-parity violation (RPV) [31]. Domain adaptation (DA) by backward propagation (of errors) [32] is incorporated into the network architecture to achieve similar classification performance in simulation and pp collision data, thus mitigating differences between the two domains, simulation and data, which can arise from e.g. limited simulation performance or anomalous instrumental effects.

This note is organized as follows. Section 2 describes the CMS detector and event reconstruction algorithms. Sections 3 and 4 describe, respectively, the event samples and simulation software packages used in this study. Section 5 describes the LLP jet tagger. Sections 6 and 7 demonstrate, respectively, the validation of the tagger using control samples of pp collision data and its performance based on simulated samples. Section 8 presents the expected performance of the tagger in a search for long-lived gluinos. Section 9 presents a novel, in situ determination of a correction to the LLP jet tagger efficiency. Section 10 provides a summary of this work.

2 The CMS detector and event reconstruction

The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. Forward calorimeters extend the pseudorapidity coverage provided by the barrel and endcap detectors. Muons are detected in gas-ionization chambers embedded in the steel flux-return yoke outside the solenoid. At $\eta = 0$, the outer radial dimension of the barrel section of the tracker, ECAL, HCAL, and muon subdetectors is 1.3, 1.8, 3.0, and 7.4 m, respectively. A more detailed description of the CMS detector, together with a definition of the coordinate system used and the relevant kinematic variables, can be found in Ref. [4].

The candidate vertex with the largest value of summed physics-object p_{T}^2 is taken to be the primary pp interaction vertex. The physics objects are the jets, clustered using the jet finding algorithm [33, 34] with the tracks assigned to candidate vertices as inputs, and the associated missing transverse momentum, taken as the negative vector sum of the p_{T} of those jets.

The particle-flow (PF) algorithm [35] aims to reconstruct and identify each particle in an event, with an optimized combination of all subdetector information. In this process, the identification of the particle type (photon, electron, muon, charged hadron, neutral hadron) plays an important role in the determination of the particle direction and energy. Photons [36] are identified as ECAL energy clusters not linked to the extrapolation of any charged particle trajectory to the ECAL. Electrons [37] are identified as a primary charged particle track and potentially many ECAL energy clusters corresponding to the extrapolation of this track to the ECAL and to possible bremsstrahlung photons emitted along the way through the tracker material. Muons [38] are identified as tracks in the central tracker consistent with either a track or several hits in the muon system, and associated with calorimeter deposits compatible with the muon hypothesis. Charged hadrons are identified as charged particle tracks neither identified as electrons, nor as muons. Finally, neutral hadrons are identified as HCAL energy clusters not linked to any charged hadron trajectory, or as a combined ECAL and HCAL energy excess with respect to the expected charged hadron energy deposit.

Jets are clustered from PF candidates using the anti-k_{T} algorithm [33, 34] with a distance parameter of 0.4. Additional pp interactions within the same or nearby bunch crossings (pileup) can contribute additional tracks and calorimetric energy depositions to the jet momentum. To mitigate this effect, charged particles identified to be originating from pileup vertices are discarded [39] and an offset correction is applied to correct for remaining contributions. In this study, jets are required to satisfy $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$, and are subject to a set of loose identification criteria [40] to reject anomalous activity from instrumental sources, such as detector noise. These criteria ensure that each jet contains at least two particle candidates and at least one charged particle track, the energy fraction attributed to charged- and neutral-hadron particle candidates is nonzero, and the fraction of energy deposited in the ECAL attributed to charged- and neutral-particle candidates is less than unity.

The most accurate estimator of $\vec{p}_{\mathrm{T}}^{\text{miss}}$ is computed as the negative vector sum of the transverse momenta of all the PF candidates in an event, and its magnitude is denoted as $p_{\mathrm{T}}^{\text{miss}}$ [41]. The $\vec{p}_{\mathrm{T}}^{\text{miss}}$ is modified to account for corrections to the energy scale [42] of the reconstructed jets in the event. Anomalous high-$p_{\mathrm{T}}^{\text{miss}}$ events can be due to a variety of reconstruction failures, detector malfunctions or noncollision backgrounds. Such events are rejected by event filters that are designed to identify more than 85–90% of the spurious high-$p_{\mathrm{T}}^{\text{miss}}$ events with a rejection (mistagging) rate of less than 0.1% for genuine events [41].

Events of interest are selected using a two-tiered trigger system [43]. The first level, composed of custom hardware processors, uses information from the calorimeters and muon detectors whereas a version of the full event reconstruction software optimized for fast processing is performed at the second level, which runs on a farm of processors.

3 Event selection and sample composition

The simplified models of split SUSY considered in this note would reveal themselves in events characterized by the absence of photons and leptons, the presence of jets, and significant $\vec{p}_{\mathrm{T}}^{\text{miss}}$ from the undetected neutralinos. Candidate signal events are required to satisfy a set of selection requirements that define a signal-enriched region (SR). Conversely, event samples that are enriched in the same background processes that populate the SR, while being depleted in contributions from SUSY processes, are identified as control regions (CRs).

Two control regions are used to assess differences in the performance of the tagger when using simulated events or pp collision data. The single muon (μ+jets) CR is required to contain exactly one muon satisfying $p_{\mathrm{T}} > 26\,\mathrm{GeV}$ and $|\eta| < 2.4$. The dimuon ($\mu\mu$+jets) CR is required to contain a second muon that satisfies $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ and $|\eta| < 2.4$. The muons are required to be isolated from other activity in the event and satisfy tight identification criteria [38]. Events containing additional loose muons (electrons) with $p_{\mathrm{T}} > 10(15)\,\mathrm{GeV}$ and $|\eta| < 2.4$ are vetoed. Both CRs must contain at least two jets satisfying $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$. Events in the μ+jets and $\mu\mu$+jets CRs are, respectively, required to satisfy $p_{\mathrm{T}}^{\text{miss}} > 150\,\mathrm{GeV}$ and $p_{\mathrm{T}}(\mu\mu) > 100\,\mathrm{GeV}$, where the latter variable is the vectorial p_{T} sum of the dimuon system. The μ+jets CR is dominated by the associated production of jets and a W boson, single top quark production, and the production and semileptonic decay of top quark-antiquark pairs ($\mathrm{t\bar{t}}$). The $\mu\mu$+jets CR contains events from Drell–Yan ($\mathrm{q\bar{q}} \to \mathrm{Z}^0/\gamma^* \to \mu^{\pm}\mu^{\mp}$), $\mathrm{t\bar{t}}$, and Wt-channel single top quark production. Events in both CRs are efficiently recorded with a trigger condition that requires the presence of a single isolated muon that satisfies $p_{\mathrm{T}} > 24\,\mathrm{GeV}$ and $|\eta| < 2.4$. The data samples of pp collisions at a centre-of-mass energy of 13 TeV correspond to an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$.

Candidate signal events in the SR are required to satisfy the following set of selection requirements. Events are required to contain at least three jets, as defined in Section 2. Events are vetoed if they contain at least one muon (electron), isolated from other activity in the event, that satisfies $p_{\mathrm{T}} > 10\,(15)\,\mathrm{GeV}$ and $|\eta| < 2.4$. The mass scale of each event is estimated from the scalar p_{T} sum of the jets, $H_{\mathrm{T}} = \sum_i^{\text{jets}} p_{\mathrm{T}}^i$, which is required to be larger than 300 GeV. An estimator for $p_{\mathrm{T}}^{\text{miss}}$ is given by the magnitude of the vector $\vec{p}_{\mathrm{T}}$ sum of the jets, $H_{\mathrm{T}}^{\text{miss}} = |\sum_i^{\text{jets}} \vec{p}_{\mathrm{T}}^{\,i}|$, which is also required to be larger than 300 GeV. Following these selections, the dominant contribution to the SM background is multijet events produced via the strong interaction, a manifestation of quantum chromodynamics (QCD). This contribution is reduced to a negligible level using the following criteria. Events with at least one jet that satisfies $p_{\mathrm{T}} > 50\,\mathrm{GeV}$ and $2.4 < |\eta| < 5$ are vetoed. Events are required to satisfy $H_{\mathrm{T}}^{\text{miss}}/p_{\mathrm{T}}^{\text{miss}} < 1.25$, which mitigates the rare circumstance in which several jets with p_{T} below the aforementioned 30 GeV threshold and collinear in ϕ lead to large values of $H_{\mathrm{T}}^{\text{miss}}$ relative to $p_{\mathrm{T}}^{\text{miss}}$, the latter of which is less sensitive to jet thresholds. Events are also required to satisfy $\Delta\phi^{\star}_{\min} > 0.2$, which is defined by the minimum azimuthal separation between each jet and the vector p_{T} sum of all other jets in the event [44]. Events in the SR can be efficiently recorded with a trigger condition that requires the presence of a single jet, $H_{\mathrm{T}}^{\text{miss}} > 120\,\mathrm{GeV}$, and $p_{\mathrm{T}}^{\text{miss}} > 120\,\mathrm{GeV}$.

The remaining background events in the SR are dominated by contributions from processes that involve the production of high-p_{T} neutrinos in the final state, such as the associated production of jets and a Z^0 boson that decays to $\nu\overline{\nu}$. A further significant background contribution arises from events that contain a W boson that undergoes a leptonic decay, $\mathrm{W}(\to\ell\nu)$+jets, where the charged lepton ($\ell \equiv e$, μ or τ) is outside the experimental acceptance or is not identified or is not isolated. The associated production of jets and a W boson is most relevant for $c\tau_0$ values of order $\mathcal{O}(10\,\mu\mathrm{m})$. Single top quark and $\mathrm{t\bar{t}}$ production, leading to a final state containing one leptonically decaying W boson and at least one b jet, also provides a substantial contribution. Residual contributions from rare SM processes, such as diboson production or the associated production of $\mathrm{t\bar{t}}$ and a vector or scalar boson, are not considered in this study.

4 Monte Carlo simulation

The DNN is trained to predict the jet class using supervised learning, which relies on truth information from MC generator programs. Various simulated event samples are also used during the evaluation of the DNN to benchmark the performance of the tagger.

Split SUSY predicts the unification of the gauge couplings at high energy [45–47] and a candidate dark matter particle, the neutralino. In addition to a (finely tuned) scalar Higgs boson, only the fermionic superpartners may be kinematically accessible at the LHC. All other superpartners are assumed to be ultraheavy. The gluino is only able to decay through the highly virtual squark states. Hence, the gluino hadronizes and forms a bound state with SM particles called an R-hadron. The R-hadron can travel a significant distance before the gluino undergoes a three-body decay to $\mathrm{q\overline{q}}\widetilde{\chi}^0$ according to its proper decay length $c\tau_0$.

The split SUSY simplified models are defined by three parameters: $c\tau_0$ and the masses of the gluino $m_{\widetilde{\mathrm{g}}}$ and the neutralino $m_{\widetilde{\chi}^0}$. The following model parameter space is considered by the search: $600 < m_{\widetilde{\mathrm{g}}} < 2400\,\mathrm{GeV}$, $m_{\widetilde{\mathrm{g}}} - m_{\widetilde{\chi}^0} > 100\,\mathrm{GeV}$, and $10\,\mu\mathrm{m} < c\tau_0 < 10\,\mathrm{m}$. The lower and upper bounds of the $c\tau_0$ range are motivated by the $\mathcal{O}(10\,\mu\mathrm{m})$ position resolution of the tracker subdetector [48] and the physical dimensions of the CMS detector. The tagger performance is also assessed using two benchmark split SUSY models that feature: an "uncompressed" mass spectrum, with a large mass difference between the gluino and $\widetilde{\chi}^0$, $(m_{\widetilde{\mathrm{g}}}, m_{\widetilde{\chi}^0}) = (2000, 0)\,\mathrm{GeV}$; or a "compressed" spectrum that is nearly degenerate in mass, $(1600, 1400)\,\mathrm{GeV}$. The value of

$c\tau_0$ for both models is defined in the text on a case-by-case basis.

Benchmark GMSB and RPV SUSY models are also considered for demonstrating the generalization of the DNN. A GMSB-inspired model assumes a long-lived gluino, with a mass of 2500 GeV and $c\tau_0 = 1\,\text{mm}$ or 1 m, that decays to a gluon and a light gravitino of mass 1 keV. Again, all other SUSY particles are assumed to be ultraheavy and decoupled from the interaction. An RPV-inspired model assumes the production of top squark-antisquark pairs. The decay of the long-lived top squark, with a mass of 1200 GeV and $c\tau_0 = 1\,\text{mm}$ or 1 m, to a b quark and a charged lepton is suppressed through a small R-parity violating Yukawa coupling.

Samples of simulated events are produced with several MC generator programs. For the split SUSY simplified models, the associated production of gluino pairs and up to two additional partons are generated at leading-order (LO) accuracy with the MADGRAPH5_aMC@NLO 2.2.2 [49] program. The decay of the gluino is performed with the PYTHIA 8.205 [50] program. The RHADRONS package within the PYTHIA 8.205 program, steered according to the default parameter settings, is used to describe the formation of R-hadrons through the hadronization of gluinos [51–53]. A similar treatment is performed for the GMSB- and RPV-inspired benchmark models.

The MADGRAPH5_aMC@NLO 2.2.2 event generator is used to produce samples of $\mathrm{W}(\to\ell\nu)$+jets, $\mathrm{Z}^0(\to\nu\overline{\nu})$+jets, and $\mathrm{Z}^0/\gamma^*(\to\mu\mu)$+jets events at next-to-leading-order (NLO) accuracy in QCD. The samples of $\mathrm{W}(\to\ell\nu)$+jets events are generated with up to two additional partons at the matrix-element level and are merged using the FxFx scheme [54]. The POWHEG v2 [55–57] event generator is used to simulate the production of top quark-antiquark pairs [58] and the t-channel [59] and Wt-channel [60] production of single top quarks at NLO accuracy. Multijet events are simulated at LO accuracy using PYTHIA 8.205.

Simulated samples are normalized to theoretical production cross sections that are calculated with NLO and next-to-NLO precision [49, 59–63] for SM processes and with NLO plus next-to-leading-logarithm (NLL) precision [64] for the production of gluino and top squark-antisquark pairs.

The PYTHIA 8.205 program with the CUETP8M1 tune [65] is used to describe parton showering and hadronization for all simulated samples except top quark-antiquark production, which used the CUETP8M2T4 tune [66]. The NNPDF3.0 LO and NLO parton distribution functions (PDFs) [67] are used with the LO and NLO event generators, as described above. Minimum bias events are overlaid with the hard scattering event to simulate pileup interactions. The resulting events undergo a full detector simulation using the GEANT [68] package. The model-dependent interactions of R-hadrons with the detector material are not considered by this study.

5 The LLP jet algorithm

In this section, the DNN architecture and technical implementation of the LLP jet tagger is described, along with various novel applications of techniques related to machine learning.

5.1 Jet labelling

Truth information from MC generator programs is often used to label a jet according to its initiating parton for supervised learning. A standard procedure known as "ghost" labelling [39] determines the jet flavour by clustering not only the reconstructed final-state particles into jets, but also the generator-level b and c hadrons. Only the directional information of the four-momentum of the generator-level (ghost) hadron is retained to prevent any modification to the four-momentum of the corresponding reconstructed jet. Jets containing at least one ghost hadron are assigned the corresponding flavour label, with b hadrons preferentially selected over c hadrons. Similarly, labels are defined for jets originating from gluons (g) or light-flavour (uds) quarks.

The LLP jet tagger also adopts the ghost labelling approach for jets originating from SM background processes. However, a complication arises when applying ghost labelling to the jets originating from $\widetilde{g}$ decays. The two quarks produced in the $\widetilde{g}$ decay can interact with each other, potentially leading to one or more jets that do not point in the same direction as the quarks. Two examples of $\widetilde{g} \to q\overline{q}\widetilde{\chi}_1^0$ decays are shown in Fig. 1. For each example, the final-state particles resulting from the hadronization of one of the quarks are sufficiently diffuse such that they are clustered into multiple distinct jets. By definition, ghost tagging cannot account for multiple jets originating from a single ghost particle, and it may even fail to associate the ghost particle to any of the jets if the jets are sufficiently distanced from one other in η–ϕ space.

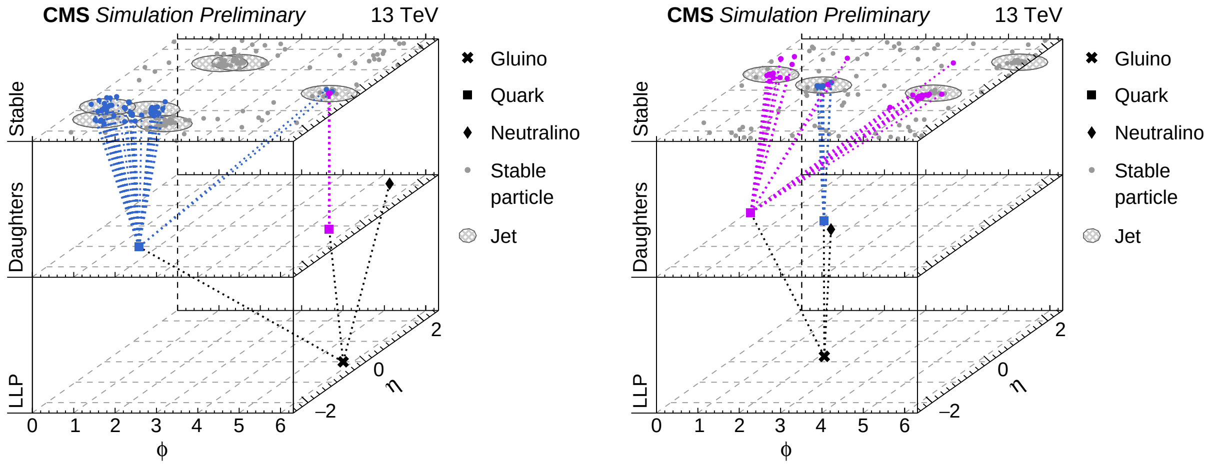

Figure 1: Two examples of a $\widetilde{g} \to q\overline{q}\widetilde{\chi}_1^0$ decay using truth information from the MADGRAPH5_aMC@NLO 2.2.2 [49] and PYTHIA 8.205 [50] programs. The positions of various particles in the η–ϕ plane are shown: the LLP ($\widetilde{g}$) and its daughter particles ($q\overline{q}\widetilde{\chi}_1^0$) are shown in the lower and middle planes, respectively; the upper plane depicts the location of the stable final-state particles after hadronization, with shaded ellipses overlaid to indicate the reconstructed jets. Each quark and its decay is assigned a unique colour. The dotted lines indicate the links between parent and daughter particles.

An alternative labelling scheme is defined for jets originating from gluino decays, which can be extended to other LLP decays. All stable SM particles are grouped according to their simulated vertex position, which can then be linked to one of the (quark) daughters from the LLP decay. All stable SM particles, except neutrinos, are clustered into generator-level jets using the anti-k_T algorithm [33, 34] with a distance parameter of 0.4. Given that constituent particles in a jet may originate from different vertices, the momentum of a given jet is shared between vertices according to the vectorial momentum sum $\sum_i \vec{p}_i$ of the constituent particles i in a jet that share the same vertex v. Per vertex, the jet-vertex shared momentum fraction f_v with respect to vertex v is then defined as

$$f_v = \frac{(\sum_i \vec{p}_i | i \in v) \cdot \vec{p}}{\vec{p}^2}, \quad f_v \in [0;1], \quad \sum_v^{\text{vertices}} f_v = 1. \tag{1}$$

Each jet is then associated to the vertex v from which the majority of its momentum originates, i.e. $\hat{v} = \text{argmax}(f_v)$. This criterion prevents the coincidental association of jets containing very few or very soft particles from a gluino decay to a vertex for which the majority of the constituent particles stem from initial-state radiation or the underlying event. A reconstructed jet is associated with a generator-level jet if their axes are aligned within a cone $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.4$, and it adopts the same label. The LLP label is prioritized over the other jet labels to prevent ambiguities. Jets from the split SUSY samples that are not labelled as LLP by this scheme can still comprise a nonnegligible fraction of displaced particles and are thus discarded to prevent class contamination. Non-LLP jets for the training of the DNN are instead taken from simulated samples of SM backgrounds. An (artificial) 20% contamination of the LLP jet class from pileup jets leads to no discernible effect on the tagger performance given the statistical precision of the study.

5.2 Deep neural network architecture to predict the jet class

The architecture used to predict the jet class, inspired by the DEEPJET algorithm, is shown in Fig. 2. The DNN considers approximately 600 input features, which can be grouped into four categories: up to 25 charged (neutral) PF particle candidates, each described by 17 (6) features; up to four secondary vertices (SVs), each described by 12 features; and 14 global features associated with the jet.

Four sequential layers of one-dimensional convolutions with a kernel size of one are used, with each layer comprising 64, 32, 16, 8 or 4 filters depending on the group of input features. The convolutional layers provide feature compression for each particle or vertex. After each layer, a leaky rectified linear (LeakyReLU) [69] activation function is used. Dropout [70] layers are interleaved throughout the network with a 10% dropout rate to mitigate overfitting. After the final convolutional layer, the compressed feature vectors are flattened and concatenated along with the global jet features. The input parameter $c\tau_0$ shown in Fig. 2 is described in Section 5.3.

The resulting feature vector is fed into a multilayer perceptron, a series of dense (fully connected) layers comprising 200, 100, or 50 neurons. The softmax activation function is used in the last layer. Categorical cross entropy is used for the loss function to predict the jet class probability. The DNN provides the probabilities for the following jet classes: LLP jet, b or c jet, uds (quark) jets, and gluon jets. The latter two classes are frequently combined when evaluating the network performance to give a light-flavour (udsg) jet class.

5.3 Network parameterization according to $c\tau_0$

The experimental signature for a jet produced in a gluino decay depends strongly on $c\tau_0$. Information from all CMS detector systems is available if the gluino decay occurs promptly, in the vicinity of the luminous region of the proton beams, while information can be limited if the decay occurs in the outermost detector systems. Hence, $c\tau_0$ is introduced as an input parameter to the dense network, as indicated in Fig. 2. This parameterized approach [30] allows for hypothesis testing with a single network for values of $c\tau_0$ that span six orders of magnitude: $10\,\mu\text{m} < c\tau_0 < 10\,\text{m}$. During the network evaluation, the $c\tau_0$ value is given by the model hypothesis under test.

5.4 Domain adaptation by backward propagation

The simulated event samples are of limited accuracy and do not exhaustively reproduce all features observed in the pp collision data. Hence, a classification algorithm may produce different identification efficiencies when evaluated on simulated samples and pp collision data

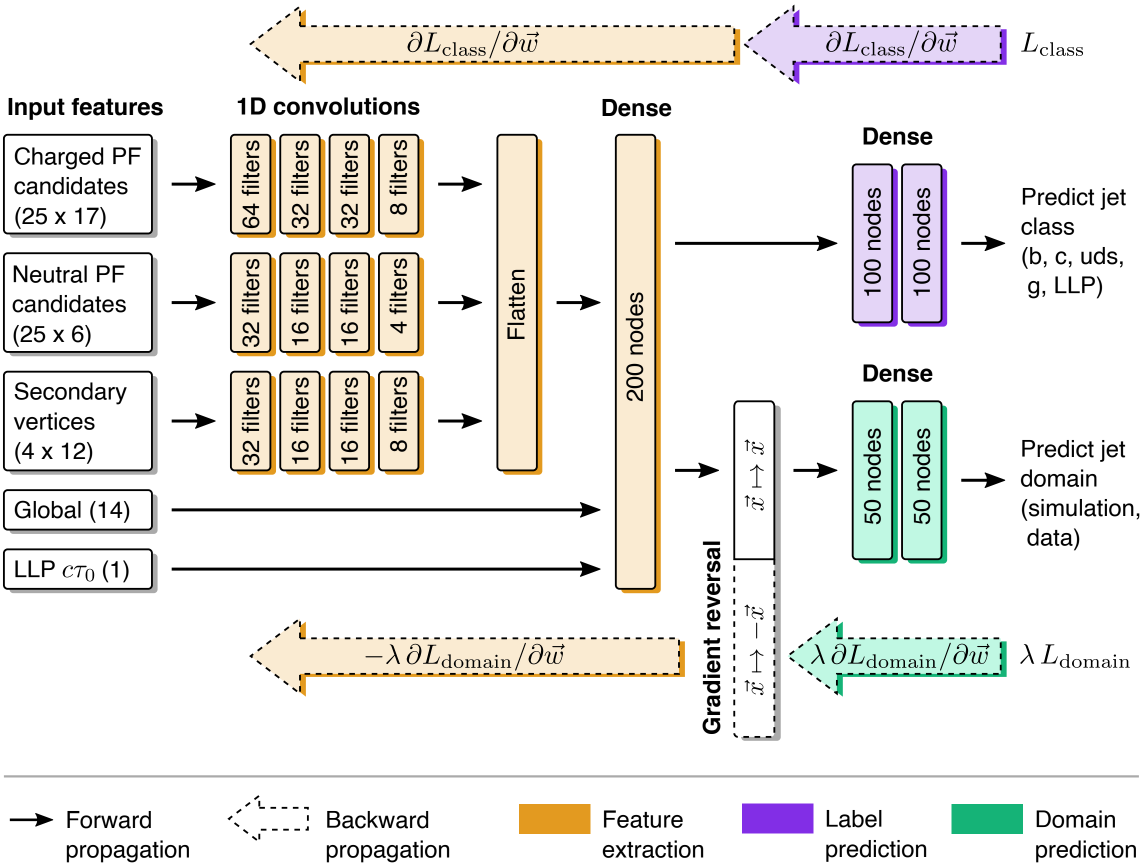

Figure 2: An overview of the DNN architecture, which comprises convolutional and dense layers; the number of filters and nodes, respectively, is indicated. Dropout layers and activation functions are not shown. The input features are grouped by object type and $(m \times n)$ indicates the maximum number of objects (m) and the number of features per object (n). The gradients of the class (L_{class}) and domain (L_{domain}) losses with respect to the weights $\vec{w}$, used during backward propagation, are shown.

if the DNN training relies solely on simulation. In this study, unsupervised DA by backward propagation [32] is employed to obtain a similar jet classification performance when applied to jets in samples of simulated events or pp collision data. This approach reduces the importance of individual or combinations of input features that exhibit differences in the two domains of pp collision data and simulation.

The network is extended to predict also the jet domain, as indicated in Fig. 2. A network branch is added after the first dense layer, the feature layer. A gradient reversal layer is inserted in the domain branch directly after the feature layer. This special layer is only active during backward propagation and reverses the gradients of the domain loss L_{domain} with respect to the network weights $\vec{w}$ in its preceding layers. At the end of the domain prediction branch, the sigmoid activation function is used. Binary cross entropy is used for the loss function to predict the jet domain probability. During DNN training, the combined loss $L_{\text{class}} + \lambda L_{\text{domain}}$ is minimized, where λ is a hyperparameter that controls the magnitude of any penalty from the jet domain loss.

5.5 Deep neural network training

Supervised training of the DNN is performed to predict the jet class. The ADAM optimizer [71] is used to minimize the loss function with respect to the network parameters. The DNN training relies on simulated events of gluino production from split SUSY models, and multijet and $\mathrm{t\bar{t}}$ production. Jets in these samples that have an uncorrected p_T greater than 20 GeV and satisfy the loose requirements outlined in Section 3 are considered for the DNN training. Approximately 20 million jets are used in the DNN training. Event samples of split SUSY are utilised over a wide range of $(m_{\tilde{\mathrm{g}}}, m_{\tilde{\chi}^0})$ hypotheses to ensure adequate generalization of the DNN over the model parameter space of interest. Models considered by the DNN training have $c\tau_0$ values that span the range $10\,\mu\mathrm{m} < c\tau_0 < 10\,\mathrm{m}$. For the jet domain prediction, a sample of 1.2 million jets are drawn from a sample of μ+jets events and simulated $\mathrm{W}(\rightarrow\ell\nu)$+jets and $\mathrm{t\bar{t}}$ events, weighted according to their respective SM cross sections, that satisfy the μ+jets CR requirements.

The training of the DNN comprises a few tens of epochs, each of which is batched into subsamples of 10 000 jets, containing approximately 2000 jets of each class: LLP, b, c, usd, and g. For each batch, the jets from SM processes are drawn randomly (from a larger sample) to obtain the same p_T and η distributions as the LLP jet class. This resampling technique is done to ensure an adequate generalization that is (largely) independent of kinematical features related to the physics process. For each batch, the $c\tau_0$ values assigned to jets from SM background processes are generated randomly according to the $c\tau_0$ distribution obtained from LLP jets.

The training of the domain branch uses batches of 10 000 jets, drawn from samples of μ+jets data and simulated $\mathrm{W}(\rightarrow\ell\nu)$+jets and $\mathrm{t\bar{t}}$ events in the relative quantities of 55%, 35%, and 10%, respectively. All parts of the network are trained simultaneously by merging the class and domain batches. Events in the domain batch are assigned the same $c\tau_0$ values as used by the class batch. For the domain batch, only the six highest p_T jets are used, and jets may be reused multiple times per epoch.

The DNN is initially trained to predict only the jet class to determine the optimal scheduling of the learning rate (α), which decays from an initial value of 0.01 according to $\alpha = 0.01/(1+\kappa n)$ where n is the epoch number and κ is the decay constant. The classifier performance is optimal for $\kappa = 0.1$ and only weakly dependent on κ. The DNN is then trained to predict both the jet class and domain, and the λ hyperparameter is increased according to $\lambda = \lambda_0[2/(1+e^{-0.2n}) - 1]$ with $\lambda_0 = 30$, such that λ increases from 0 to $0.9\lambda_0$ after 15 epochs.

5.6 Workflow

The TENSORFLOW v1.6 [72] queue system is used to read and preprocess files for the DNN training. The KERAS v2.1.5 [73] software package is used to implement the DNN architecture. A schematic overview of the pipeline used to preprocess batches of jets for the class prediction is given in Fig. 3, while a similar queue is also used for the domain prediction. At the beginning of each epoch, a queue holding a randomized list of the input file names is initialized. File names are dequeued asynchronously in multiple threads. For each thread, ROOT v6.18.00 [74] trees contained in the files are read from disk to memory in batches using a TENSORFLOW operation kernel, developed in the context of this note. The resulting batches are resampled to achieve the same distributions in p_T and η for all jet classes and are enqueued asynchronously into a second queue, which caches a list of tensors. The DNN training commences by dequeuing a randomized batch of tensors and generating $c\tau_0$ values for all SM jets within the batch. The advantages of this system lay in its flexibility to adapt to new input features or samples on-the-fly. The (p_T, η) resampling and the generation of $c\tau_0$ values for the SM jets proceeds asynchronously in multiple threads, managed by TENSORFLOW, on the CPU while the network is being trained. A demonstration of the workflow is currently under preparation and will be released in the future.

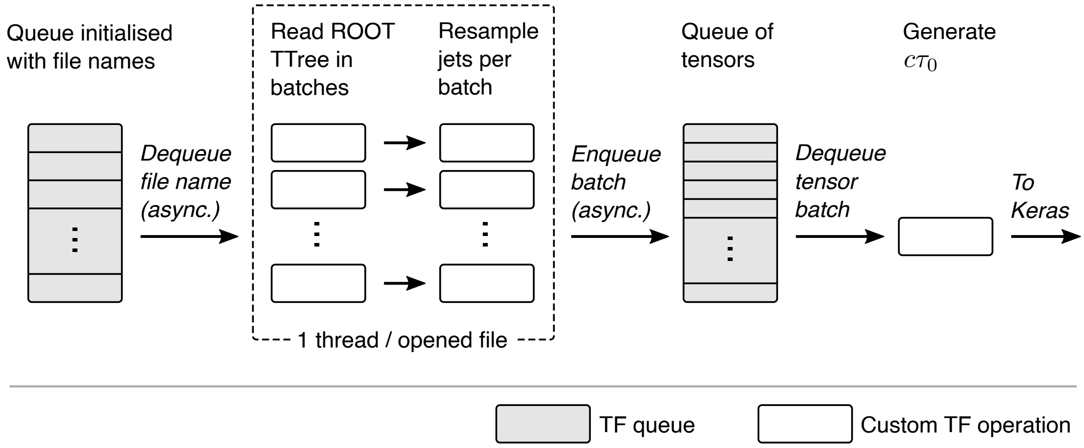

Figure 3: A schematic of the input pipeline for training the DNN, which uses the TENSORFLOW queue system with custom operation kernels for reading ROOT trees from disk, (p_T, η) resampling for SM jets, and generating random $c\tau_0$ values for jets from SM backgrounds and data.

6 Validation with pp collision data

In the absence of DA, the LLP jet probability $P(\mathrm{LLP}|c\tau_0)$ obtained from the simulation of the relevant SM backgrounds and the CR data can differ significantly, with deviations of up to 50% in distributions of $P(\mathrm{LLP}|c\tau_0)$, while a good agreement is observed for the distributions of the other jet class probabilities. A similar event-level variable is $P_\mathrm{max}(\mathrm{LLP}|c\tau_0)$, which is defined as the maximum value of $P(\mathrm{LLP}|c\tau_0)$ obtained from all selected jets in a given event. A comparison of the $P_\mathrm{max}(\mathrm{LLP}|c\tau_0)$ distributions obtained from pp collision data and simulated events in both the μ+jets and $\mu\mu$+jets CRs, using a DNN trained with and without DA, is shown in Fig. 4. The use of DA in the network leads to a significant improvement in the level of agreement in the binned counts of $P_\mathrm{max}(\mathrm{LLP}|c\tau_0)$ for the two domains of data and simulation, with only small residual differences remaining. This improvement is expected for the μ+jets CR, as the same events are used to train, evaluate, and optimize the domain branch of the DNN. The $\mu\mu$+jets CR, comprising a statistically independent event sample, validates the method.

An estimate of the uncertainty in $P_\mathrm{max}(\mathrm{LLP}|c\tau_0)$ due to simulation mismodelling is determined from jets in a statistically independent sample of $\mu\mu$+jets events that satisfy $p_\mathrm{T}(\mu\mu) < 100\,\mathrm{GeV}$. The magnitude of the uncertainty is assessed by weighting up or down the simulated events by the factor $w^\pm = \prod_i^\mathrm{jets}\Big(1 \pm \big(\xi_i(\mathrm{LLP}) - 1\big)\Big)$, where $\xi_i(\mathrm{LLP})$ is the ratio of counts from data and simulation in bin i of the $P(\mathrm{LLP}|c\tau_0)$ distribution. The ratios of event counts binned according to $P_\mathrm{max}(\mathrm{LLP}|c\tau_0)$ from pp collision data and simulation, as well as the corresponding uncertainty estimates, are shown in the lower panels of Fig. 4. The ratios are closer to unity, with reduced uncertainties, following the application of DA. The level of agreement between data and simulation is further quantified by the Jensen–Shannon divergence (JSD) [75], which is bound to $[0, 1]$. The JSD is reduced by an order of magnitude following the application of DA. The quoted uncertainties in JSD reflect the finite size of the data and simulated samples.

As for the CRs, the application of DA leads to significantly reduced biases and uncertainties in

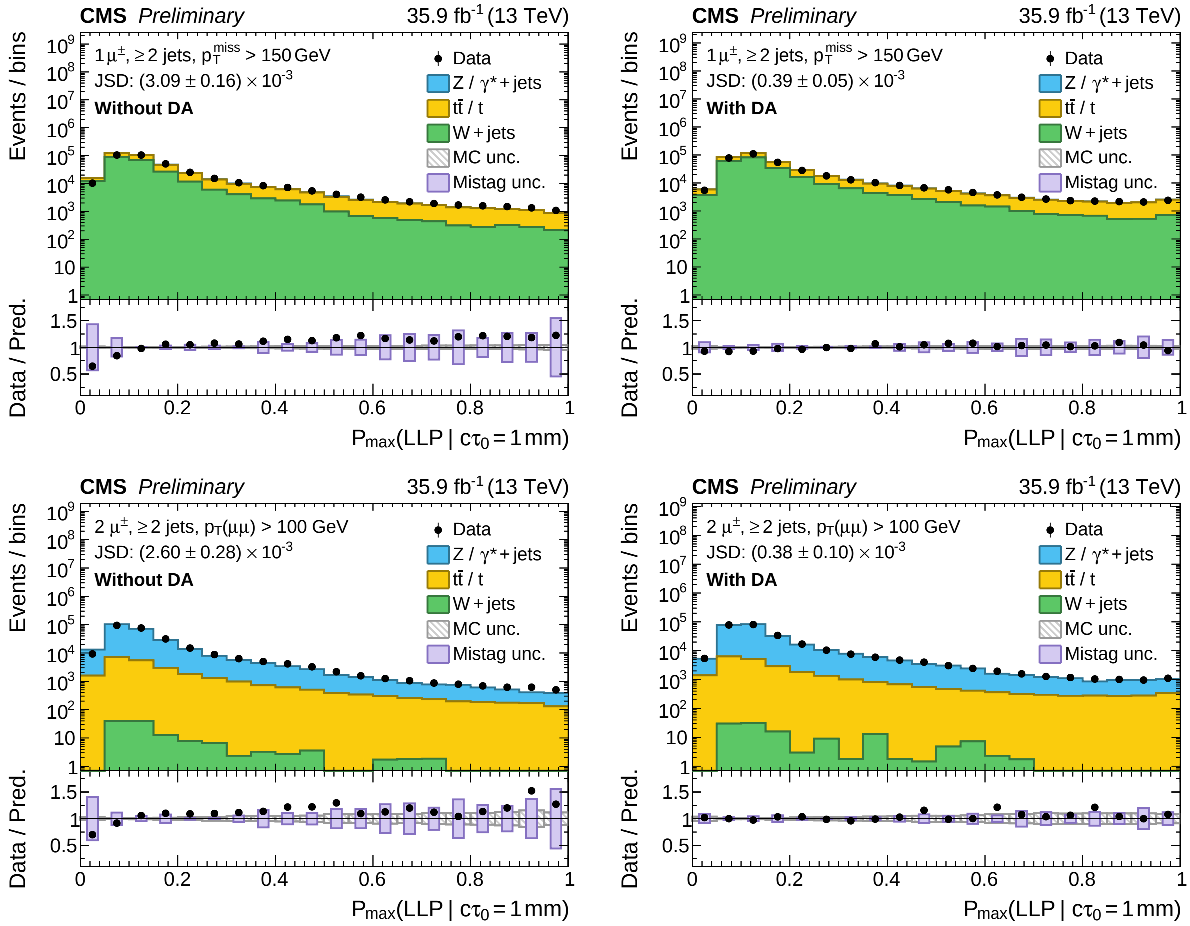

Figure 4: Distributions of the maximum probability for the LLP jet class obtained from all selected jets in a given event, $P_{\mathrm{max}}(\mathrm{LLP}|c\tau_0)$. The distributions from pp collision data (circular marker) and simulated events (histograms) are compared in the μ+jets (upper row) and $\mu\mu$+jets (lower row) CRs, using a DNN trained without (left column) and with (right column) DA. All probabilities are evaluated with $c\tau_0 = 1$ mm. The JSD is introduced in the text.

the modelling of $P(\mathrm{LLP}|c\tau_0)$ and related variables in the signal-depleted CRs. In a search, this behaviour would translate into an improved treatment of the experimental systematic uncertainties related to the estimation of SM backgrounds. However, only modest gains in sensitivity to new high-mass particle states may be expected, as the dominant uncertainties arise from the finite-size samples of pp collision data and simulated events.

7 Performance

The performance of the tagger is demonstrated using simulated event samples for split SUSY benchmark models with an uncompressed mass spectrum, as defined in Section 4, and $c\tau_0$ values of 1 mm and 1 m. The two values of $c\tau_0$ highlight the roles of the tracker and calorimeter systems, respectively. Negligible SM background contributions are expected for the 1 m scenario. An inclusive sample of $\mathrm{t\bar{t}}$ events is used to provide both light-flavour (udsg) jets, through initial state radiation and hadronic decays of the W boson, as well as b jets with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$.

The efficiency of the tagger to identify correctly the LLP jet class depends on the chosen working point, defined by a threshold requirement on the jet class probability. The mistag rates for the remaining jet classes also depend on the same working point. The receiver operating characteristic (ROC) curves that provide the LLP jet tagging efficiency and the mistag rate for the udsg jet class as a function of the working point are shown in Fig. 5. The uncertainties indicated by shaded bands are determined from a threefold cross validation. Given a mistag rate of 0.01%, equivalent to a background rejection factor of 10 000, efficiencies of 40 and 70% are obtained for $c\tau_0$ values of 1 mm and 1 m, respectively.

Figure 5 also shows ROC curves when evaluating the DNN using simulated events from the GMSB and RPV SUSY benchmark models, as defined in Section 4. The jets originate from uds quarks (gluons) from the gluino decay in the case of split (GMSB) SUSY models, and b hadrons from the top squark decay for RPV SUSY. A similar level of performance is observed for these SUSY models, in which the LLP jets have a different underlying flavour. Further, the ROC curves indicated by the thick and thin solid curves illustrate the tagger performance when the DNN is trained with or without DA, respectively, for the split SUSY benchmark models. Studies demonstrate a comparable performance for split SUSY models with and without DA for the region $c\tau_0 \leq 10$ mm; for larger values of $c\tau_0$, the performance is overestimated without DA, as indicated by the $c\tau_0 = 1$ m scenario shown in Fig. 5, because the model is able to exploit feature distributions obtained from simulation that are not representative of data.

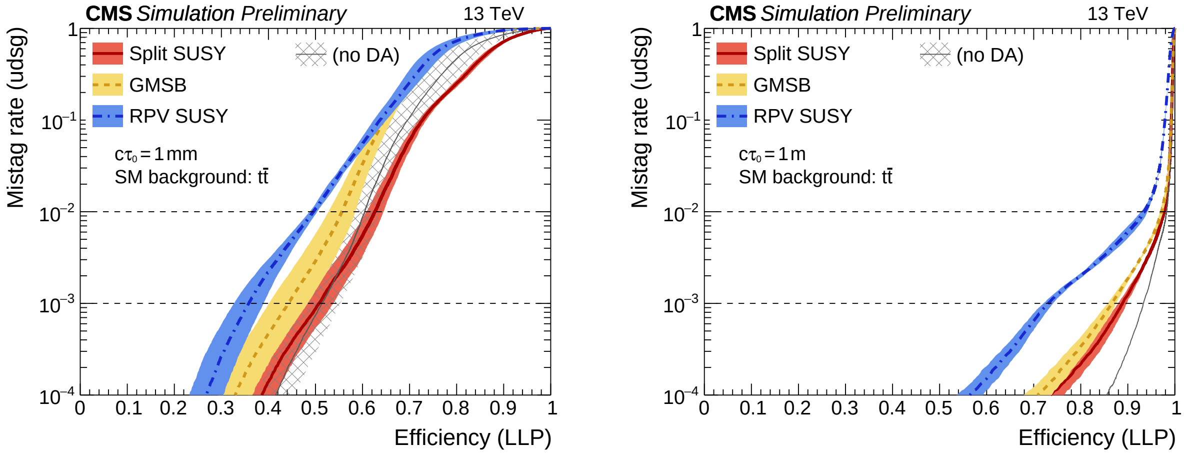

Figure 5: ROC curves illustrating the tagger performance for the split (solid line), GMSB (dashed), and RPV (dot-dashed) SUSY benchmark models, assuming $c\tau_0$ values of 1 mm (left) and 1 m (right). The thick and thin solid curves indicate the performance using the DNN trained with and without DA, respectively. The jet sample is defined in the text.

The LLP jet tagging efficiency is shown in Fig. 6 as a function of the jet p_T, η, and the number of reconstructed secondary vertices within the jet (N_{SV}) for a working point that yields a mistag rate of 0.01% for the udsg jet class obtained from simulated $t\bar{t}$ events. Efficiencies are highest for high-p_T, centrally produced jets with $N_{SV} = 0$. The latter observation demonstrates the complementary performance of the tagger with respect to a more standard approach of relying on reconstructed secondary vertices to identify displaced jets. The RPV model has lower efficiencies for $p_T \gtrsim 100$ GeV or $N_{SV} = 0$ because of reduced discrimination between the (SM) b jet class and the LLP jet class, when the jet is of b quark flavour as produced in the decay of the long-lived top squark.

The performance of the DNN parameterization according to $c\tau_0$ is shown in Figure 7. The left subfigure shows the LLP jet tagging efficiency, using a working point that yields a mistag rate of 0.01% for the udsg jet class, as a function of the generated $c\tau_0$ value for both an uncompressed

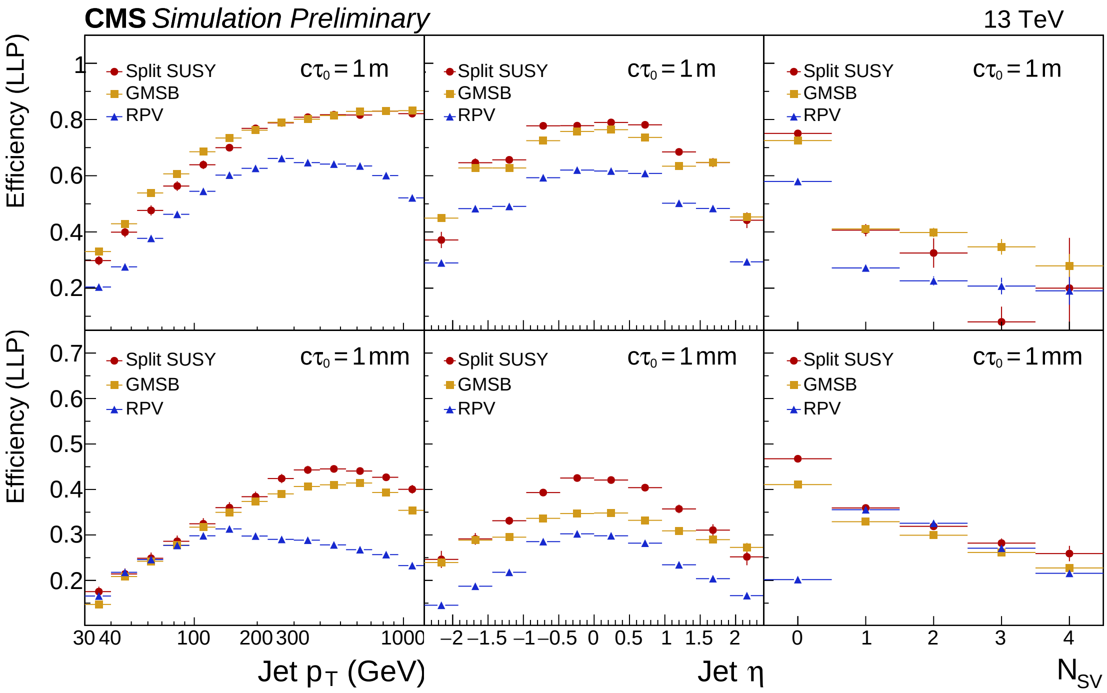

Figure 6: The LLP jet tagging efficiency as a function of the jet p_{T}, η, and N_{SV} using a working point that yields a mistag rate of 0.01% for the udsg jet class, as obtained from an inclusive sample of simulated $\mathrm{t\bar{t}}$ events. The efficiency curves are shown separately for the split (circular marker), GMSB (square), and RPV (triangle) SUSY benchmark models, assuming $c\tau_0$ values of 1 m (upper row) and 1 mm (lower row).

and a compressed split SUSY model. Efficiencies in the range 40–80% (30–70%) are achieved for uncompressed (compressed) scenarios with 1 mm $\leq c\tau_0 \leq$ 10 m. The compressed scenarios are characterized by low-p_{T} jets, resulting in lower efficiencies. The performance is further tested using two uncompressed split SUSY models with $c\tau_0 =$ 1 mm and 1 m. The LLP jet tagging efficiency is obtained by evaluating the DNN for each value of $c\tau_0$ in the range 10 μm $\leq c\tau_0 \leq$ 10 m. Again, the efficiency is determined using a working point that is tuned for each evaluated $c\tau_0$ value to yield a mistag rate of 0.01% for the udsg jet class. The efficiency as a function of the evaluated $c\tau_0$ value is shown in Fig. 7 (right). The maximum efficiency is obtained when the evaluated value of $c\tau_0$ approximately matches the parameter value used for the split SUSY model hypothesis. This behaviour may be used to characterize a potential signal contribution in terms of the model parameter $c\tau_0$. Finally, studies demonstrate that the parameterised approach does not significantly impact performance with respect to the training of multiple DNNs, one per $c\tau_0$ value.

8 Application to a search for LLPs

An application of the LLP jet tagger is demonstrated in a showcase search for long-lived gluinos with 10 μm $\leq c\tau_0 \leq$ 10 m. Expected limits on the theoretical production cross section for gluino pairs are determined. Candidate events that satisfy the SR requirements defined in Section 3 are categorised according to: the number of jets in the events, N_{jet}; the number of jets for which $P(\mathrm{LLP}|c\tau_0)$ is above a predefined threshold, N_{tag}; and H_{T}. Events are further categorized according to H_{T}. Models with an uncompressed mass spectrum, $m_{\tilde{\mathrm{g}}} - m_{\tilde{\chi}^0} \gtrsim$ 200 GeV, are characterized by high values of H_{T}. Models with a compressed mass spectrum,

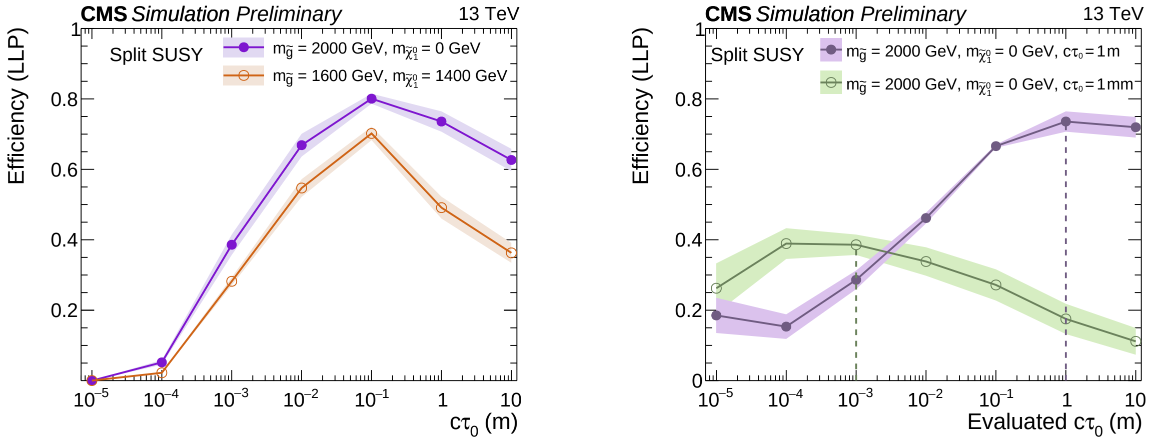

Figure 7: The LLP jet tagging efficiency, using a working point that yields a mistag rate of 0.01% for the udsg jet class obtained from an inclusive sample of simulated $t\bar{t}$ events, when (left) the DNN is evaluated as a function of the model parameter value $c\tau_0$ for an uncompressed and a compressed split SUSY model, and (right) the DNN is evaluated over a range of $c\tau_0$ values for uncompressed split SUSY models generated with $c\tau_0 = 1\,\text{mm}$ and 1 m; the dashed vertical lines indicate equality for the evaluated and generated values of $c\tau_0$ for each model. The fixed model parameters are defined in the legends.

$m_{\tilde{g}} - m_{\tilde{\chi}^0} \lesssim 200\,\text{GeV}$, are characterized by lower values of H_T because of the limited kinematic phase space available to the jets from the gluino decay and an increased reliance on associated jet production from initial-state radiation. Events that satisfy $N_{\text{tag}} < 2$ are grouped into a single category, which is used to constraint the normalization of simulated background events during the statistical evaluation.

The tagger is evaluated using simulated samples of all relevant SM backgrounds, described in Section 4. The negligible background contribution from multijet events in the SR is not considered in this exploratory study. The predefined threshold on $P(\text{LLP}|c\tau_0)$ is determined per $c\tau_0$ value such that the most sensitive event categories are nearly free of SM background contributions while maintaining control over uncertainties due to the finite-size simulated samples. The $P(\text{LLP}|c\tau_0)$ thresholds fall in the range 30–50% and yield an LLP jet tagging efficiency of about 30–90% for $c\tau_0 \geq 1\,\text{mm}$. Table 1 summarizes the a priori expected counts and uncertainties for the contributions from SM backgrounds in the various event categories for $c\tau_0 = 1\,\text{mm}$. The statistical uncertainty arising from the finite size of the simulated samples is the dominant contribution to the quoted uncertainties. Additional systematic contributions are described below. The expected event counts from an uncompressed and compressed split SUSY model are also provided.

Several sources of systematic uncertainty in the SM background expectations are considered. An uncertainty of ±20% is assumed in the normalization of each dominant background process ($\text{W}(\to\ell\nu)$+jets, $\text{Z}^0(\to\nu\bar{\nu})$+jets, and single top squark and $t\bar{t}$ production), which is motivated by theoretical uncertainties in the production cross sections and the experimental acceptance for the final state leptons [76]. The jet energy is varied within its uncertainty and resolution [42] and the resulting shifts are propagated to $\vec{p}_T^{\text{miss}}$. The unclustered component of $\vec{p}_T^{\text{miss}}$ is varied within its uncertainties. An uncertainty in the number of pileup interactions, by varying the inelastic pp cross section within its uncertainty of ±5% [77], is considered. The renormalization and factorization scales of the aforementioned four dominant SM backgrounds are varied independently by factors of 0.5 and 2 per process [78]. An uncertainty of ±2.5% in the integrated

Table 1: A priori expected counts and uncertainties for SM backgrounds and split SUSY models, as determined from simulation, in categories defined by H_T and (N_{jet}, N_{tag}). The uncompressed and compressed split SUSY models are defined in Section 4. The value of $c\tau_0$ is assumed to be 1 mm. The uncertainties include both statistical and systematic contributions. Expected counts for events that satisfy $N_{tag} < 2$ are not shown.

H_T (GeV)	300–800	300–800	300–800	>800	>800	>800
(N_{jet}, N_{tag})	(3–4, ≥2)	(5, ≥2)	(≥6, ≥3)	(3–4, ≥2)	(5, ≥2)	(≥6, ≥3)
$Z^0(\to\nu\bar{\nu})$+jets	40.7 ± 39.2	6.5 ± 5.8	0.6 ± 0.4	3.3 ± 2.8	1.6 ± 1.2	0.1 ± 0.1
$W(\to\ell\nu)$+jets	56.3 ± 44.1	11.6 ± 5.1	1.5 ± 0.5	3.6 ± 2.5	1.2 ± 3.0	< 0.1
$t\bar{t}$	39.6 ± 36.1	17.9 ± 15.7	1.9 ± 1.1	2.1 ± 1.3	3.2 ± 2.4	3.0 ± 2.1
Single top	5.7 ± 5.2	2.6 ± 2.2	0.3 ± 0.2	0.6 ± 0.4	0.5 ± 0.3	0.4 ± 0.3
Total SM	142 ± 69	38.5 ± 17.6	4.3 ± 1.3	9.7 ± 4.0	6.5 ± 4.1	3.5 ± 2.5
Uncompressed	< 0.1	< 0.1	< 0.1	3.0 ± 2.9	3.8 ± 3.7	5.7 ± 5.5
Compressed	5.4 ± 5.0	4.2 ± 3.8	2.8 ± 2.5	1.1 ± 0.9	2.5 ± 2.2	4.5 ± 4.1

luminosity is assumed [79].

A likelihood model is used to test for the presence of new-physics signals in the SR. The observed event count in each event category is modelled as a Poisson-distributed variable around the SM expectation and a potential signal contribution. The expected event counts from SM backgrounds is given by simulated event samples. The uncertainties resulting from the finite simulated samples are modelled using the Barlow-Beeston method [80]. The systematic uncertainties in the SM background estimates are accommodated in the likelihood model as nuisance parameters.

Hypothesis testing is performed using an Asimov data set to provide expected constraints on the parameter space of simplified models of split SUSY. A modified frequentist approach is used to determine the expected upper limits (ULs) at 95% confidence level (CL) on the theoretical gluino-pair production cross section as a function of $m_{\tilde{g}}$, $m_{\tilde{\chi}^0}$, and $c\tau_0$. The signal strength parameter, r_{UL}, expresses the UL relative to the production cross section. Alternatively, expected lower limits (95% CL) on $m_{\tilde{g}}$ can be determined as a function of $c\tau_0$. The approach is based on the profile likelihood ratio as the test statistic [81], the CL_s criterion [82, 83], and the asymptotic formulae [84] to approximate the distributions of the test statistic under the background-only and signal-plus-background hypotheses.

Figure 8 summarizes the expected lower limit on $m_{\tilde{g}}$ (95% CL) as a function of $c\tau_0$ for simplified models of split SUSY that assume the production of gluino pairs. The model assumptions are indicated by the legend in each panel. The left (right) subfigure presents the expected mass exclusions for a model with an uncompressed (compressed) mass spectrum. A lower limit on the gluino mass in excess of 2 TeV is obtained. The mass exclusions are compared to those obtained from an inclusive search for SUSY [28] in final states containing jets and p_T^{miss}, over the same range in $c\tau_0$ values. Significant gains in excluded values of $m_{\tilde{g}}$, of up to approximately 500 GeV, are expected for $c\tau_0 \gtrsim 1$ mm. For the region $c\tau_0 < 1$ mm, the tagger performance degrades while the reference search is able to exploit the distinguishing kinematical features of split SUSY events through a finer categorization of candidate signal events. For the region $c\tau_0 > 1$ m, the LLPs frequently decay outside the experimental acceptance, which leads to an increased reliance on the presence of initial-state radiation.

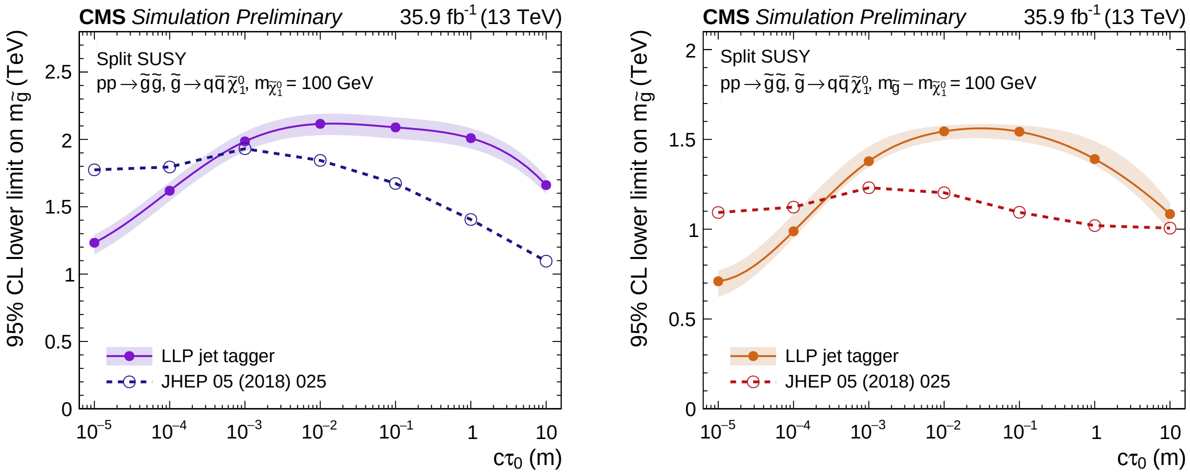

Figure 8: Expected lower limits (95% CL) on $m_{\tilde{g}}$ as a function of $c\tau_0$ for split SUSY models with an uncompressed (left) and a very compressed (right) mass spectrum. The shaded bands indicate the total uncertainty from both statistical and systematic sources. The model assumptions are indicated by the legends. The results are compared to the expected limits obtained in Ref. [28], indicated by a dashed line.

9 In situ corrections to the LLP jet tagging efficiency

Domain adaptation is used to minimize differences in the distributions of the jet class probability obtained from simulated events and pp collision data in the CRs. The LLP jet tagging efficiency determined from simulated events of split SUSY models, ϵ_{MC}, may require the application of corrections to account for sources of potential mismodelling. The corrections are applied by reweighting simulated events as follows:

$$w = \left(\frac{1-\mathrm{SF}\,\epsilon_{\mathrm{MC}}}{1-\epsilon_{\mathrm{MC}}}\right)^{(N_{\mathrm{jet}}-N_{\mathrm{tag}})} \times \mathrm{SF}^{N_{\mathrm{tag}}}, \tag{2}$$

where SF denotes the unknown scale factor correction and $\mathrm{SF}\,\epsilon_{\mathrm{MC}} \in [0,1]$. The $(N_{\mathrm{jet}}, N_{\mathrm{tag}})$ categorization scheme allows the SF to be constrained *in situ* during the derivation of upper limits on the gluino-pair production cross section. The SF is introduced as a free parameter to the likelihood model.

Figure 9 shows the negative log-likelihood from a maximum likelihood fit to the Asimov data as a function of both the SF and r/r_{UL}, where the signal strength parameter r expresses the obtained UL relative to the gluino-pair production cross section, for two benchmark split SUSY scenarios with an uncompressed and compressed mass spectrum, and $c\tau_0 = 10\,\mathrm{mm}$. The model assumptions are indicated in the figure legends. The SF can be constrained to $<$40% ($<$70%) at 68% CL for $r/r_{\mathrm{UL}} = 1$ for the uncompressed (compressed) model.

10 Summary

Many models of new physics beyond the standard model predict the production of long-lived particles (LLPs) in proton-proton (pp) collisions at the LHC. Jets arising from the decay of LLPs (LLP jets) can be appreciably displaced from the pp collisions. The development of a novel tagger to identify LLP jets is presented. The tagger employs a deep neutral network (DNN) using an architecture inspired by the DEEPJET algorithm. Simplified models of split

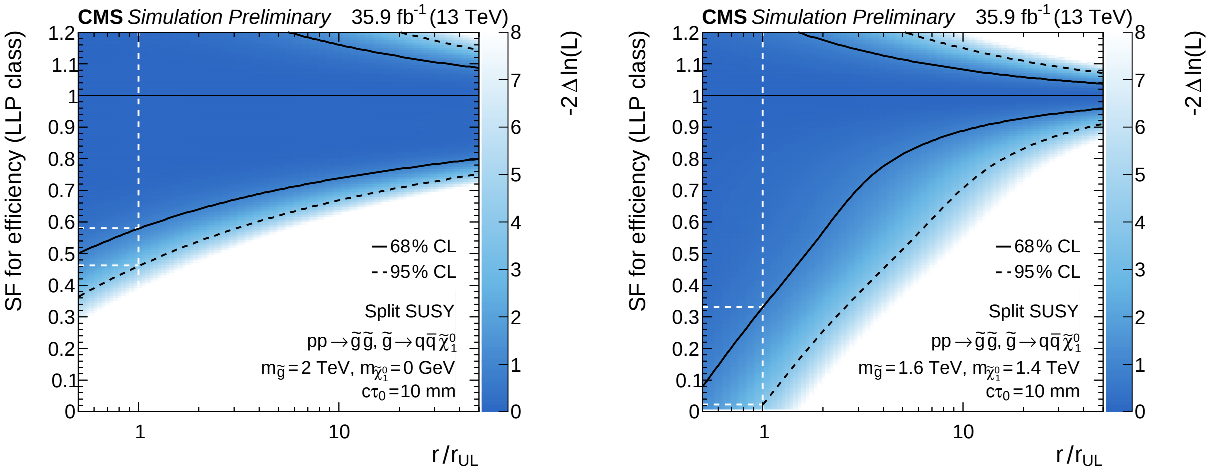

Figure 9: The negative log-likelihood of a maximum likelihood fit to the Asimov data set as a function of the expected gluino production cross section for a (left) uncompressed and (right) compressed scenario. The black solid (dashed) line indicate the 68% (95%) CL, while the white dashed lines indicate the SF constraints (68% and 95% CL) at $r = r_{\mathrm{UL}}$. The product of the LLP jet tagger efficiency and the SF is bound to $[0, 1]$.

supersymmetry (SUSY), which yield neutralinos and LLP jets from the decay of long-lived gluinos, are used to train the DNN and demonstrate its performance.

The application of various techniques related to the tagger are reported. A custom labelling scheme for LLP jets based on truth information from Monte Carlo generator programs is defined. The proper decay length $c\tau_0$ of the gluino is used as an external parameter to the DNN, which allows hypothesis testing over several orders of magnitude in $c\tau_0$ with a single DNN. The DNN was trained using both simulated and pp collision data using domain adaptation by backward propagation. This approach significantly improves the agreement between simulation and data, by an order of magnitude according to the Jensen–Shannon divergence, when compared to training the DNN with simulation only. The method is validated using signal-depleted control samples of pp collisions at a centre-of-mass energy of 13 TeV. The samples were recorded by the CMS experiment and correspond to an integrated luminosity of 35.9 fb^{-1}. Training the DNN with pp collision data does not significantly degrade the tagger performance. The tagger rejects 99.99% of light-flavour jets from SM processes, as measured in an inclusive $\mathrm{t\bar{t}}$ sample, while retaining approximately 30–80% of LLP jets for split SUSY models with 1 mm $\leq c\tau_0 \leq$ 10 m and a gluino-neutralino mass difference of at least 200 GeV.

Finally, the potential performance of the tagger is demonstrated through its use in a search for split SUSY in final states containing jets and significant transverse missing momentum. Simulated events samples provide the expected contributions from SM background processes. Candidate signal events were categorized according to the scalar sum of jet momenta, the number of jets, and the number of tagged LLP jets. Expected lower limits on the gluino mass (95% CL) are determined with a binned likelihood fit as a function of $c\tau_0$ in the range from 10 μm to 10 m. A procedure to constrain a correction to the LLP jet tagger efficiency in the likelihood fit is introduced. Competitive limits are demonstrated: long-lived gluinos of mass $\gtrsim$ 2 TeV and proper decay length 1 mm $\leq c\tau_0 \leq$ 1 m are expected to be excluded by this search.

References

[1] K. Albertsson et al., "Machine learning in high energy physics community white paper", *J. Phys. Conf. Ser.* **1085** (2018) 022008, `doi:10.1088/1742-6596/1085/2/022008`, `arXiv:1807.02876`.

[2] A. J. Larkoski, I. Moult, and B. Nachman, "Jet substructure at the Large Hadron Collider: a review of recent advances in theory and machine learning", (2017). `arXiv:1709.04464`. Submitted to *Phys. Rept.*

[3] ATLAS Collaboration, "The ATLAS experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003, `doi:10.1088/1748-0221/3/08/S08003`.

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

[5] ATLAS Collaboration, "ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13\,\text{TeV}$", `arXiv:1907.05120`. Submitted to *Eur. Phys. J. C.*

[6] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *JINST* **13** (2018) P05011, `doi:10.1088/1748-0221/13/05/P05011`, `arXiv:1712.07158`.

[7] M. Stoye, "Deep learning in jet reconstruction at CMS", *J. Phys. Conf. Ser.* **1085** (2018) 042029, `doi:10.1088/1742-6596/1085/4/042029`.

[8] CMS Collaboration, "Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13 TeV with Phase 1 CMS detector", CMS Detector Performance Note CMS-DP-2018-058, 2018.

[9] N. Arkani-Hamed and S. Dimopoulos, "Supersymmetric unification without low energy supersymmetry and signatures for fine-tuning at the LHC", *JHEP* **06** (2005) 073, `doi:10.1088/1126-6708/2005/06/073`, `arXiv:hep-th/0405159`.

[10] G. F. Giudice and A. Romanino, "Split supersymmetry", *Nucl. Phys. B* **699** (2004) 65, `doi:10.1016/j.nuclphysb.2004.08.001`, `arXiv:hep-ph/0406088`. [Erratum: *Nucl. Phys. B* **706** (2005) 487, `doi:10.1016/j.nuclphysb.2004.11.048`].

[11] J. L. Hewett, B. Lillie, M. Masip, and T. G. Rizzo, "Signatures of long-lived gluinos in split supersymmetry", *JHEP* **09** (2004) 070, `doi:10.1088/1126-6708/2004/09/070`, `arXiv:hep-ph/0408248`.

[12] G. F. Giudice and R. Rattazzi, "Theories with gauge mediated supersymmetry breaking", *Phys. Rept.* **322** (1999) 419, `doi:10.1016/S0370-1573(99)00042-3`, `arXiv:hep-ph/9801271`.

[13] M. J. Strassler and K. M. Zurek, "Discovering the Higgs through highly-displaced vertices", *Phys. Lett. B* **661** (2008) 263, `doi:10.1016/j.physletb.2008.02.008`, `arXiv:hep-ph/0605193`.

[14] T. Han, Z. Si, K. M. Zurek, and M. J. Strassler, "Phenomenology of hidden valleys at hadron colliders", *JHEP* **07** (2008) 008, `doi:10.1088/1126-6708/2008/07/008`, `arXiv:0712.2041`.

[15] D. Alves et al., "Simplified models for LHC new physics searches", *J. Phys. G* **39** (2012) 105005, `doi:10.1088/0954-3899/39/10/105005`, `arXiv:1105.2838`.

[16] O. Buchmueller et al., "Simplified models for displaced dark matter signatures", *JHEP* **09** (2017) 076, `doi:10.1007/JHEP09(2017)076`, `arXiv:1704.06515`.

[17] J. Alimena et al., "Searching for long-lived particles beyond the standard model at the Large Hadron Collider", (2019). `arXiv:1903.04497`.

[18] G. R. Farrar and P. Fayet, "Phenomenology of the production, decay, and detection of new hadronic states associated with supersymmetry", *Phys. Lett. B* **76** (1978) 575, `doi:10.1016/0370-2693(78)90858-4`.

[19] CMS Collaboration, "Search for long-lived charged particles in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Phys. Rev. D* **94** (2016) 112004, `doi:10.1103/PhysRevD.94.112004`, `arXiv:1609.08382`.

[20] LHCb Collaboration, "Updated search for long-lived particles decaying to jet pairs", *Eur. Phys. J. C* **77** (2017) 812, `doi:10.1140/epjc/s10052-017-5178-x`, `arXiv:1705.07332`.

[21] ATLAS Collaboration, "Search for long-lived neutral particles in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ that decay into displaced hadronic jets in the ATLAS calorimeter", *Eur. Phys. J. C* **79** (2019) 481, `doi:10.1140/epjc/s10052-019-6962-6`, `arXiv:1902.03094`.

[22] ATLAS Collaboration, "Search for heavy charged long-lived particles in the ATLAS detector in 36.1 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13\,\text{TeV}$", *Phys. Rev. D* **99** (2019) 092007, `doi:10.1103/PhysRevD.99.092007`, `arXiv:1902.01636`.

[23] ATLAS Collaboration, "Search for long-lived particles produced in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ that decay into displaced hadronic jets in the ATLAS muon spectrometer", *Phys. Rev. D* **99** (2019) 052005, `doi:10.1103/PhysRevD.99.052005`, `arXiv:1811.07370`.

[24] CMS Collaboration, "Search for long-lived particles using nonprompt jets and missing transverse momentum with proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Phys. Lett. B* **797** (2019) 134876, `doi:10.1016/j.physletb.2019.134876`, `arXiv:1906.06441`.

[25] CMS Collaboration, "Search for long-lived particles decaying into displaced jets in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Phys. Rev. D* **99** (2019) 032011, `doi:10.1103/PhysRevD.99.032011`, `arXiv:1811.07991`.

[26] CMS Collaboration, "Search for long-lived particles with displaced vertices in multijet events in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Phys. Rev. D* **98** (2018) 092011, `doi:10.1103/PhysRevD.98.092011`, `arXiv:1808.03078`.

[27] CMS Collaboration, "Search for decays of stopped exotic long-lived particles produced in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **05** (2018) 127, `doi:10.1007/JHEP05(2018)127`, `arXiv:1801.00359`.

[28] CMS Collaboration, "Search for natural and split supersymmetry in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ in final states with jets and missing transverse momentum", *JHEP* **05** (2018) 025, `doi:10.1007/JHEP05(2018)025`, `arXiv:1802.02110`.

[29] CMS Collaboration, "Search for new long-lived particles at $\sqrt{s} = 13\,\text{TeV}$", *Phys. Lett. B* **780** (2018) 432, `doi:10.1016/j.physletb.2018.03.019`, `arXiv:1711.09120`.

[30] P. Baldi et al., "Parameterized neural networks for high-energy physics", *Eur. Phys. J. C* **76** (2016) 235, `doi:10.1140/epjc/s10052-016-4099-4`, `arXiv:1601.07913`.

[31] R. Barbier et al., "R-parity violating supersymmetry", *Phys. Rept.* **420** (2005) 1, `doi:10.1016/j.physrep.2005.08.006`, `arXiv:hep-ph/0406039`.

[32] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation", (2014). `arXiv:1409.7495`.

[33] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_{T} jet clustering algorithm", *JHEP* **04** (2008) 063, `doi:10.1088/1126-6708/2008/04/063`, `arXiv:0802.1189`.

[34] M. Cacciari, G. P. Salam, and G. Soyez, "FASTJET user manual", *Eur. Phys. J. C* **72** (2012) 1896, `doi:10.1140/epjc/s10052-012-1896-2`, `arXiv:1111.6097`.

[35] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017) P10003, `doi:10.1088/1748-0221/12/10/P10003`, `arXiv:1706.04965`.

[36] CMS Collaboration, "Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$", *JINST* **10** (2015) P08010, `doi:10.1088/1748-0221/10/08/P08010`, `arXiv:1502.02702`.

[37] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$", *JINST* **10** (2015) P06005, `doi:10.1088/1748-0221/10/06/P06005`, `arXiv:1502.02701`.

[38] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7\,\text{TeV}$", *JINST* **7** (2012) P10002, `doi:10.1088/1748-0221/7/10/P10002`, `arXiv:1206.4071`.

[39] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Phys. Lett. B* **659** (2008) 119, `doi:10.1016/j.physletb.2007.09.077`, `arXiv:0707.1378`.

[40] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *JINST* **6** (2011) 11002, `doi:10.1088/1748-0221/6/11/P11002`, `arXiv:1107.4277`.

[41] CMS Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ using the CMS detector", *JINST* **14** (2019) P07004, `doi:10.1088/1748-0221/14/07/P07004`, `arXiv:1903.06078`.

[42] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017) P02014, `doi:10.1088/1748-0221/12/02/P02014`, `arXiv:1607.03663`.

[43] CMS Collaboration, "The CMS trigger system", *JINST* **12** (2017) P01020, `doi:10.1088/1748-0221/12/01/P01020`, `arXiv:1609.02366`.

[44] CMS Collaboration, "Search for supersymmetry in pp collisions at 7 TeV in events with jets and missing transverse energy", *Phys. Lett. B* **698** (2011) 196, `doi:10.1016/j.physletb.2011.03.021`, `arXiv:1101.1628`.

[45] S. Dimopoulos, S. Raby, and F. Wilczek, “Supersymmetry and the scale of unification”, *Phys. Rev. D* **24** (1981) 1681, `doi:10.1103/PhysRevD.24.1681`.

[46] L. E. Ibanez and G. G. Ross, “Low-energy predictions in supersymmetric grand unified theories”, *Phys. Lett. B* **105** (1981) 439, `doi:10.1016/0370-2693(81)91200-4`.

[47] W. J. Marciano and G. Senjanović, “Predictions of supersymmetric grand unified theories”, *Phys. Rev. D* **25** (1982) 3092, `doi:10.1103/PhysRevD.25.3092`.

[48] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST* **9** (2014) P10009, `doi:10.1088/1748-0221/9/10/P10009`, `arXiv:1405.6569`.

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

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

[51] M. Fairbairn et al., “Stable massive particles at colliders”, *Phys. Rept.* **438** (2007) 1, `doi:10.1016/j.physrep.2006.10.002`, `arXiv:hep-ph/0611040`.

[52] A. C. Kraan, “Interactions of heavy stable hadronizing particles”, *Eur. Phys. J. C* **37** (2004) 91, `doi:10.1140/epjc/s2004-01946-6`, `arXiv:hep-ex/0404001`.

[53] R. Mackeprang and A. Rizzi, “Interactions of coloured heavy stable particles in matter”, *Eur. Phys. J. C* **50** (2007) 353, `doi:10.1140/epjc/s10052-007-0252-4`, `arXiv:hep-ph/0612161`.

[54] R. Frederix and S. Frixione, “Merging meets matching in MC@NLO”, *JHEP* **12** (2012) 061, `doi:10.1007/JHEP12(2012)061`, `arXiv:1209.6215`.

[55] P. Nason, “A new method for combining NLO QCD with shower Monte Carlo algorithms”, *JHEP* **11** (2004) 040, `doi:10.1088/1126-6708/2004/11/040`, `arXiv:hep-ph/0409146`.

[56] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with parton shower simulations: the POWHEG method”, *JHEP* **11** (2007) 070, `doi:10.1088/1126-6708/2007/11/070`, `arXiv:0709.2092`.

[57] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG box”, *JHEP* **06** (2010) 043, `doi:10.1007/JHEP06(2010)043`, `arXiv:1002.2581`.

[58] S. Frixione, P. Nason, and G. Ridolfi, “A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction”, *JHEP* **09** (2007) 126, `doi:10.1088/1126-6708/2007/09/126`, `arXiv:0707.3088`.

[59] S. Alioli, P. Nason, C. Oleari, and E. Re, “NLO single-top production matched with shower in POWHEG: s- and t-channel contributions”, *JHEP* **09** (2009) 111, `doi:10.1088/1126-6708/2009/09/111`, `arXiv:0907.4076`. [Erratum: *JHEP* **02** (2010) 011, `doi:10.1007/JHEP02(2010)011`].

[60] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method", *Eur. Phys. J. C* **71** (2011) 1547, `doi:10.1140/epjc/s10052-011-1547-z`, `arXiv:1009.2450`.

[61] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, "W physics at the LHC with FEWZ 2.1", *Comput. Phys. Commun.* **184** (2013) 208, `doi:10.1016/j.cpc.2012.09.005`, `arXiv:1201.5896`.

[62] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, "FEWZ 2.0: a code for hadronic Z production at next-to-next-to-leading order", *Comput. Phys. Commun.* **182** (2011) 2388, `doi:10.1016/j.cpc.2011.06.008`, `arXiv:1011.3540`.

[63] M. Czakon and A. Mitov, "Top++: a program for the calculation of the top-pair cross-section at hadron colliders", *Comput. Phys. Commun.* **185** (2014) 2930, `doi:10.1016/j.cpc.2014.06.021`, `arXiv:1112.5675`.

[64] C. Borschensky et al., "Squark and gluino production cross sections in pp collisions at $\sqrt{s} = 13, 14, 33$ and 100 TeV", *Eur. Phys. J. C* **74** (2014) 3174, `doi:10.1140/epjc/s10052-014-3174-y`, `arXiv:1407.5066`.

[65] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C* **76** (2016) 155, `doi:10.1140/epjc/s10052-016-3988-x`, `arXiv:1512.00815`.

[66] CMS Collaboration, "Investigations of the impact of the parton shower tuning in PYTHIA 8 in the modelling of $t\bar{t}$ at $\sqrt{s} = 8$ and 13 TeV", CMS Physics Analysis Summary CMS-PAS-TOP-16-021, 2016.

[67] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, `doi:10.1007/JHEP04(2015)040`, `arXiv:1410.8849`.

[68] GEANT4 Collaboration, "GEANT 4—a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, `doi:10.1016/S0168-9002(03)01368-8`.

[69] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks", in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15, p. 315. 2011.

[70] N. Srivastava et al., "Dropout: a simple way to prevent neural networks from overfitting", *JMLR* **15** (2014) 1929.

[71] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization", (2014). `arXiv:1412.6980`.

[72] M. Abadi et al., "TensorFlow: large-scale machine learning on heterogeneous systems", (2015). `arXiv:1603.04467`.

[73] F. Chollet, "Keras". `https://github.com/fchollet/keras`, 2015.

[74] R. Brun and F. Rademakers, "ROOT: an object oriented data analysis framework", *Nucl. Instrum. Meth. A* **389** (1997) 81, `doi:10.1016/S0168-9002(97)00048-X`.

[75] J. Lin, "Divergence measures based on the Shannon entropy", *IEEE Trans. Inf. Theor.* **37** (2006) 145, `doi:10.1109/18.61115`.

[76] S. Kallweit et al., "NLO QCD+EW predictions for V+jets including off-shell vector-boson decays and multijet merging", *JHEP* **04** (2016) 021, `doi:10.1007/JHEP04(2016)021`, `arXiv:1511.08692`.

[77] CMS Collaboration, "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **07** (2018) 161, `doi:10.1007/JHEP07(2018)161`, `arXiv:1802.02613`.

[78] A. Kalogeropoulos and J. Alwall, "The SysCalc code: a tool to derive theoretical systematic uncertainties", (2018). `arXiv:1801.08401`.

[79] CMS Collaboration, "CMS luminosity measurements for the 2016 data taking period", CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017.

[80] R. J. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples", *Comput. Phys. Commun.* **77** (1993) 219, `doi:10.1016/0010-4655(93)90005-W`.

[81] ATLAS and CMS Collaborations, and LHC Higgs Combination Group, "Procedure for the LHC Higgs boson search combination in summer 2011", Technical Report CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11, 2011.

[82] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instr. Meth. A* **434** (1999) 435, `doi:10.1016/S0168-9002(99)00498-2`, `arXiv:hep-ex/9902006`.

[83] A. L. Read, "Presentation of search results: the $\mathrm{CL_s}$ technique", *J. Phys. G* **28** (2002) 2693, `doi:10.1088/0954-3899/28/10/313`.

[84] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554, `doi:10.1140/epjc/s10052-011-1554-0`, `arXiv:1007.1727`. [Erratum: *Eur. Phys. J. C* **73** (2013) 2501, `doi:10.1140/epjc/s10052-013-2501-z`].