



PAPER • OPEN ACCESS

Wasserstein normalized autoencoder for anomaly detection

To cite this article: The CMS Collaboration 2026 *Mach. Learn.: Sci. Technol.* **7** 035030

View the [article online](#) for updates and enhancements.

You may also like

- [Flow measurements and modelling in helical RFX-mod equilibria](#)
F. Bonomo, D. Bonfiglio, P. Piovesan et al.
- [Forthcoming events](#)
- [Diary of events](#)



PAPER

OPEN ACCESS

RECEIVED

2 October 2025

REVISED

9 April 2026

ACCEPTED FOR PUBLICATION

17 April 2026

PUBLISHED

26 May 2026

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Wasserstein normalized autoencoder for anomaly detection

The CMS Collaboration*

CERN, Geneva, Switzerland

* Author to whom any correspondence should be addressed.

E-mail: cms-publication-committee-chair@cern.ch**Keywords:** CMS, anomaly detection, autoencoder, dark matter, semivisible jetsSupplementary material for this article is available [online](#)

Abstract

A novel anomaly detection algorithm is presented. The Wasserstein normalized autoencoder (WNAE) is a normalized probabilistic model that minimizes the Wasserstein distance between the learned probability distribution—a Boltzmann distribution where the energy is the reconstruction error of the autoencoder (AE)—and the distribution of the training data. This algorithm has been developed and applied to the identification of semivisible jets—conical sprays of visible standard model (SM) particles and invisible dark matter states—with the CMS experiment at the CERN LHC. Trained on jets of particles from simulated SM processes, the WNAE is shown to learn the probability distribution of the input data in a fully unsupervised fashion, such that it effectively identifies new physics jets as anomalies. The model exhibits stable, convergent training and recovers strong classification performance for a wide range of signals against the selected background process, for which a standard AE fails because of outlier reconstruction. In addition, the model improves upon standard normalized autoencoders while remaining fully agnostic to the signal. The WNAE directly tackles the problem of outlier reconstruction, a common failure mode of autoencoders in anomaly detection tasks.

1. Introduction

Unsupervised machine-learning algorithms have proven to be powerful tools in the search for new physics at the LHC [1, 2]. They can effectively separate events from known standard model (SM) processes (background) from potential signal events arising from interactions beyond the SM (BSM), without relying on the simulation of specific signal hypotheses. Autoencoders (AEs) are frequently used for unsupervised learning tasks. One of the main challenges when using AE-based unsupervised machine-learning algorithms for anomaly detection is preventing the reconstruction of outliers—examples that are not part of the training data. A well-known example of outlier reconstruction called complexity bias [3] occurs when outliers are less complex than the training data. In this paper, we introduce the Wasserstein normalized AE (WNAE), an evolution of the normalized AE (NAE) algorithm [4], to solve these challenges. The identification of semivisible jets (SVJs) [5] in the CMS experiment [6] is used as a case study to demonstrate the effectiveness of the WNAE in separating signal from background events.

The CMS apparatus [6, 7] is a multipurpose, nearly hermetic detector, designed to trigger on [8–10] and identify electrons, muons, photons, and (charged and neutral) hadrons [11–13]. A global ‘particle-flow’ algorithm [14] aims to reconstruct all individual particles in an event, combining information provided by the all-silicon inner tracker and by the crystal electromagnetic and brass-scintillator hadron calorimeters, operating inside a 3.8 T superconducting solenoid, with data from the gas-ionization muon detectors embedded in the flux-return yoke outside the solenoid. The reconstructed particles are used to build jets [15]—conical sprays of particles—and missing transverse momentum [16]—the imbalance of visible momentum in the transverse plane. Events of interest are selected using a two-tiered trigger system. The first level, composed of custom hardware processors, uses information from the calorimeters and muon detectors to select events at a rate of around 100 kHz within a fixed latency of 4 μ s [8]. The second level, known as the high-level trigger, consists of a farm of processors running a version of the

full event reconstruction software optimized for fast processing, and reduces the event rate to a few kHz before data storage [9, 10].

The paper is organized as follows. Section 2 discusses the use and limitations of (AEs) as anomaly detection tools, focusing on the outlier reconstruction problem, with a brief review of the NAE algorithm. In section 3, the AE and NAE algorithms are applied to the search for SVJs, and we introduce a metric based on the Wasserstein distance [17–19] to quantify the degree of outlier reconstruction without relying on the signal in any form during training. Section 4 presents the major advance in this paper: the use of this metric as the direct objective for the AE training to build a more robust anomaly detection algorithm, the WNAE. The advantages of the WNAE over the NAE are discussed. Finally, section 5 discusses the implications of the findings and possible future directions.

2. AE-based anomaly detection

2.1. Autoencoders

Autoencoders [20] are a class of neural networks that are intended to learn a compressed representation of the input data. This is usually achieved by mapping the input feature space $\mathcal{X} \subset \mathbb{R}^d$ to a lower-dimensional latent space $\mathcal{Z} \subset \mathbb{R}^m$ via an encoder network $f_e: \mathbb{R}^d \mapsto \mathbb{R}^m$ ($m < d$) and mapping this latent space to an output space with the same dimension as the input space via a decoder network $f_d: \mathbb{R}^m \mapsto \mathbb{R}^d$. autoencoders are typically trained by minimizing a loss function that penalizes the difference between the input and output, referred to as the reconstruction error of the AE. The latter is denoted l_θ in the following, where θ represents the parameters of the AE network.

Crucially, when (AEs) are used for anomaly detection, the network is only exposed to non-anomalous (background) examples during training. If the network is able to learn a compressed representation of these examples, it can reconstruct them with low error. However, since the anomalous (signal) examples are not part of the training set, the AE is expected not to reconstruct them well, resulting in a larger average reconstruction error.

While this reasoning holds for many practical applications of (AEs) for anomaly detection, it does not follow in general that achieving low reconstruction error on the background examples is sufficient for an AE to effectively identify anomalies, as discussed below. This shortcoming is addressed by the WNAE algorithm introduced in this paper.

2.2. The problem of outlier reconstruction

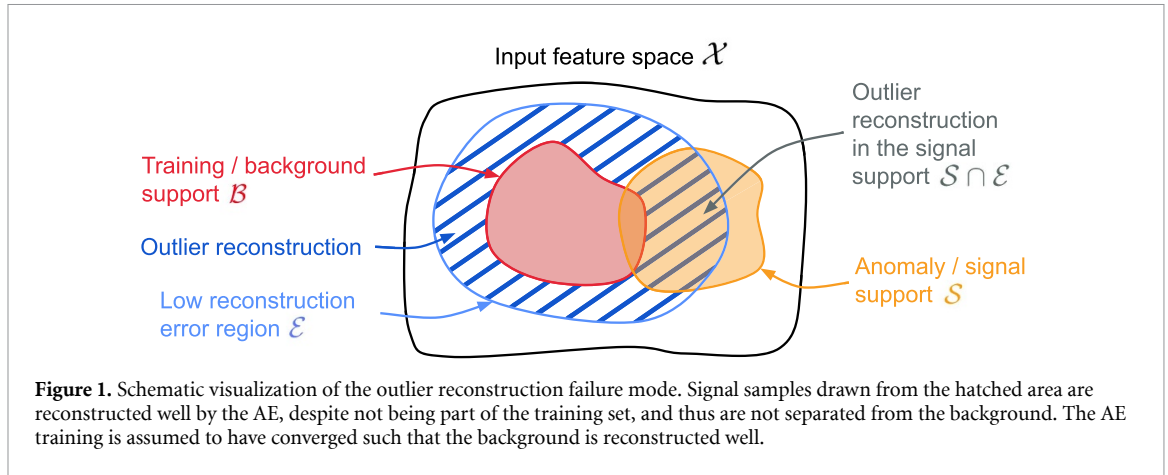
Given a data distribution p_{data} , an example datum $x \in \mathcal{X}$ is an outlier if it lies in a region with probability density lower than a given threshold τ : $p_{\text{data}}(x) \leq \tau$. In the extreme case where $\tau = 0$, the outlier lies outside of the support of the data distribution.

Outlier reconstruction occurs when an AE, while learning to reconstruct the training examples, also learns to reconstruct outlier examples. More formally: let $\mathcal{B}$ and $\mathcal{S}$ be the supports of the background and signal distributions in the input feature space $\mathcal{X}$, respectively. The AE is trained to achieve minimum reconstruction error on background samples $x_b \in \mathcal{B}$. Let $\mathcal{E}$ be the region where the AE achieves reconstruction error below some threshold. Since there is no constraint on the reconstruction error of signal samples $x_s \in \mathcal{S}$, there can be a sizable overlap $\mathcal{S} \cap \mathcal{E}$. This overlap can in general be at least partly disjoint from the training data: $(\mathcal{S} \cap \mathcal{E}) - \mathcal{B} \neq \emptyset$. Examples drawn from this region of phase space are assigned low reconstruction error by the AE, even though they are outside the training data. Thus, these examples and the background both have low reconstruction error, decreasing the performance of the reconstruction error as a discriminator. This is illustrated in figure 1.

Ideally, the reconstruction error l_θ should be a function of the probability density of the training data: the lower the probability density of the training data, the higher the reconstruction error. Thus, examples x with $p_{\text{data}}(x) \leq \tau$ must have reconstruction error above a threshold τ_r : $l_\theta(x) \geq \tau_r$. The reconstruction error itself can accordingly be used as a summary statistic to find anomalies. The objective of the AE is therefore to learn the distribution of the training data (the ‘inlier’ distribution), rather than simply to minimize the reconstruction error.

2.3. The NAE paradigm

The NAE [4] is based on a proposed mechanism to learn the probability distribution of the training data. In analogy with energy-based models [21–23], often used for generative modeling, the NAE is assigned a normalized probability distribution p_θ , parametrized by the AE network weights θ . Following



the maximum entropy principle [24], the probability distribution is chosen as the Boltzmann distribution from statistical mechanics:

$$p_{\theta}(x) = \frac{1}{\Omega_{\theta}} \exp(-E_{\theta}(x)), \quad (1)$$

where E_{θ} is the energy of the system and Ω_{θ} is the partition function:

$$\Omega_{\theta} = \int dx \exp(-E_{\theta}(x)). \quad (2)$$

The energy of the system is defined as the reconstruction error of the AE:

$$E_{\theta}(x) = l_{\theta}(x). \quad (3)$$

The problem is then recast as a maximum likelihood estimation problem, where the goal is to find the weights θ that maximize the likelihood of the training data, with probability distribution denoted p_{data} . As is common, the negative log likelihood is minimized instead:

$$\mathbb{E}_{x \sim p_{\text{data}}} [-\log p_{\theta}(x)] = \mathbb{E}_{x \sim p_{\text{data}}} [E_{\theta}(x)] + \log \Omega_{\theta}. \quad (4)$$

Calculating the partition function Ω_{θ} is in general intractable, but the problem can be circumvented when using gradient descent to find the optimum, since the gradient of the partition function can be calculated:

$$\nabla_{\theta} \log \Omega_{\theta} = \frac{1}{\Omega_{\theta}} \nabla_{\theta} \Omega_{\theta} = \frac{1}{\Omega_{\theta}} \int_B dx \nabla_{\theta} \exp(-E_{\theta}(x)) \quad (5)$$

$$= \int dx \frac{1}{\Omega_{\theta}} \exp(-E_{\theta}(x)) \nabla_{\theta} (-E_{\theta}(x)) \quad (6)$$

$$= \int dx p_{\theta}(x) \nabla_{\theta} (-E_{\theta}(x)) \quad (7)$$

$$= -\mathbb{E}_{x \sim p_{\theta}} [\nabla_{\theta} E_{\theta}(x)], \quad (8)$$

which is the expectation value of the gradient of the AE reconstruction error over the probability distribution p_{θ} . The gradient of the loss can thus be calculated by sampling from the probability distribution p_{θ} . This can be done via a Markov chain Monte Carlo (MCMC) algorithm. The application of the MCMC in the NAE training procedure is discussed in more detail in section 3.4 and appendix D. The loss function to be minimized is then:

$$\mathcal{L}_{\text{NAE}} = \mathbb{E}_{x \sim p_{\text{data}}} [E_{\theta}(x)] - \mathbb{E}_{x' \sim p_{\theta}} [E_{\theta}(x')]. \quad (9)$$

Both terms in equation (9) are expectation values of the reconstruction error of the AE. The first term is the expectation value over the training data x (the usual AE loss), referred to as the positive energy. The second term is the expectation value over samples x' drawn from the probability distribution p_{θ} ,

referred to as the negative energy. The NAE loss is thus the difference between the positive and negative energy, denoted E_+ and E_- , respectively:

$$\mathcal{L}_{\text{NAE}} \equiv E_+ - E_- . \quad (10)$$

In analogy, the training examples are referred to as positive examples, and examples drawn from p_θ are referred to as negative examples. In this way, the NAE is not trained just to reconstruct background examples well, but rather to learn the probability distribution of the training data. The NAE is therefore penalized for having low reconstruction error in regions with low probability density of the training data, hence suppressing the reconstruction of examples outside the background support $\mathcal{B}$. It is worth emphasizing that the whole procedure is still fully unsupervised. The NAE is trained only on background examples, and the signal is not used in any form during training.

2.4. MCMC

The key step in the NAE paradigm is to estimate the AE probability p_θ . This is achieved by running an MCMC algorithm to obtain a set of examples following the p_θ distribution. The Markov chain is only needed to train the network and is not used for inference. The Langevin Monte Carlo algorithm is employed in this work. First, a starting set of points $\{x\}_0$ is drawn from an initial distribution, following the procedure laid out in section 3.4. The negative samples are obtained by repeating a fixed number of steps, each step proceeding as follows:

$$x_{i+1} = x_i + \lambda \nabla_x \log p_\theta(x_i) + \sigma \epsilon_i, \quad (11)$$

where λ is the step size, σ is the noise coefficient, ϵ_i is an independent draw from a normal distribution on $\mathbb{R}^d$ with mean 0 and covariance matrix equal to the $d \times d$ identity matrix, and d is the number of dimensions of the input feature space. The Markov chain has therefore the following dependence on the reconstruction error of the AE:

$$x_{i+1} = x_i - \lambda \nabla_x E_\theta(x_i) + \sigma \epsilon_i. \quad (12)$$

With ρ_i denoting the probability distribution of the sample $\{x\}_i$, choosing:

$$\sigma = \sqrt{2\lambda}, \quad (13)$$

and assuming an infinite number of steps, the probability ρ_i converges towards the target probability p_θ : $\lim_{i \rightarrow \infty} \rho_i = p_\theta$. In practice, p_θ needs to be estimated at every gradient descent step, once per batch, while training an NAE. For that reason, the number of Markov chain steps cannot be too large, and the probability p_θ can only be approximated. In order to enhance the gradient term in equation (11), the temperature T is introduced:

$$x_{i+1} = x_i + \frac{\lambda}{T} \nabla_x \log p_\theta(x_i) + \sigma \epsilon_i. \quad (14)$$

Temperatures $T < 1$ result in an effective increase of the gradient term.

2.5. Related work

There has long been interest in applying autoencoders to model a data distribution, mostly for generative tasks. Variational autoencoders (VAE) [25] make use of the Kullback–Leibler divergence to impose a Gaussian prior in the latent space $\mathcal{Z}$ and the decoder f_d learns to model the data distribution from this Gaussian prior. Adversarial autoencoders (AAE) [26] minimize the reconstruction error and enforce a Gaussian prior in the latent space by training the network in an adversarial way, similarly to generative adversarial networks (GANs) [27]. The encoder f_e plays the role of the generator network and a discriminator network estimates the probability that a latent sample arises from the prior distribution rather than being the output of the generator $f_e(x)$, where x is a training datum. In Wasserstein autoencoders (WAE) [28], the loss function is composed of two terms: a divergence between the latent space $\mathcal{Z}$ and a (Gaussian) prior, and the Wasserstein distance between the training data distribution and that learned by the AE (the output of the decoder f_d). The Wasserstein distance between two probability distributions p_1 and p_2 is defined as:

$$W(p_1, p_2) = \inf_{\gamma \in \Pi(p_1, p_2)} \mathbb{E}_{(x, x') \sim \gamma} [\|x - x'\|], \quad (15)$$

where $\Pi(p_1, p_2)$ denotes the set of all joint distributions whose marginal distributions are p_1 and p_2 and $\inf$ indicates the infimum over this set. This distance measures the cost of morphing p_1 into p_2 . Using the Kantorovich–Rubinstein duality [29], this expression can be written in the dual form as:

$$W(p_1, p_2) = \sup_{f \in \mathcal{F}_{\mathcal{L}}} \mathbb{E}_{x \sim p_1} [f(x)] - \mathbb{E}_{x \sim p_2} [f(x)], \quad (16)$$

where the supremum is over all 1-Lipschitz functions $\mathcal{F}_{\mathcal{L}}$. The Wasserstein distance in equations (15) and (16) is typically approximated by relaxing the constraint on the coupling distribution γ or the function f and replacing it with a simpler constraint or regularizing the loss function. In Wasserstein GANs [30], the 1-Lipschitz constraint is approximated by implementing f as a neural network and clipping its weights. In WAEs, the constraint on the coupling distribution γ is relaxed and the loss function is regularized by adding this term to the loss function: $\lambda[D_f(\gamma_1, p_1) + D_f(\gamma_2, p_2)]$, where λ is a positive constant, γ_1 and γ_2 are the marginal distributions of γ , and D_f denotes an f -divergence [28]. An exact solution to the optimal transport problem in equation (15) is implemented in [31], but does not scale well to large data sets. Nevertheless, computing the Wasserstein distance on minibatches of the entire data set is efficient and was found to provide an unbiased estimator of the gradients [32].

As explained in section 2.2, an ideal outlier detection algorithm must learn the training data distribution. For that reason, and as further highlighted in section 3.3, the vanilla AE defined in section 2.1 is not a suitable algorithm for anomaly detection. While the reconstruction error of VAEs can directly be used as an anomaly score, VAEs have been shown to be prone to outlier reconstruction [33, 34].

Normalizing flows present a similar shortcoming. This class of networks learns to model complex probability distributions by mapping a simple distribution to a chosen complex distribution via a series of invertible and differentiable transformations. This procedure can be used to compute the likelihood of the data, which can serve as an anomaly detection score. However, normalizing flows can assign high likelihoods to outliers, hindering their ability to be used as anomaly detection algorithms [33, 35]. Though recent work has tried mitigating this issue [36, 37], a general solution is not known.

Algorithms based on GANs, often constructed similarly to AAEs, need to be modified to provide an anomaly score. Generative probabilistic novelty detection [38], GANomaly [39], and AnoGAN [40] follow different approaches, but all are challenging to train because of the iterative adversarial procedure [27].

The NAE, described in section 2.3, is based on a different principle: it relies on an MCMC algorithm to sample data points from the probability distribution learned by the network. Though it is a normalized probabilistic model, its loss function takes the form of a minimax rule, which leads to numerical instability and therefore requires ad-hoc regularization [4, 41]. In addition, as detailed in section 3.4, the NAE has a failure mode that can cause outlier reconstruction.

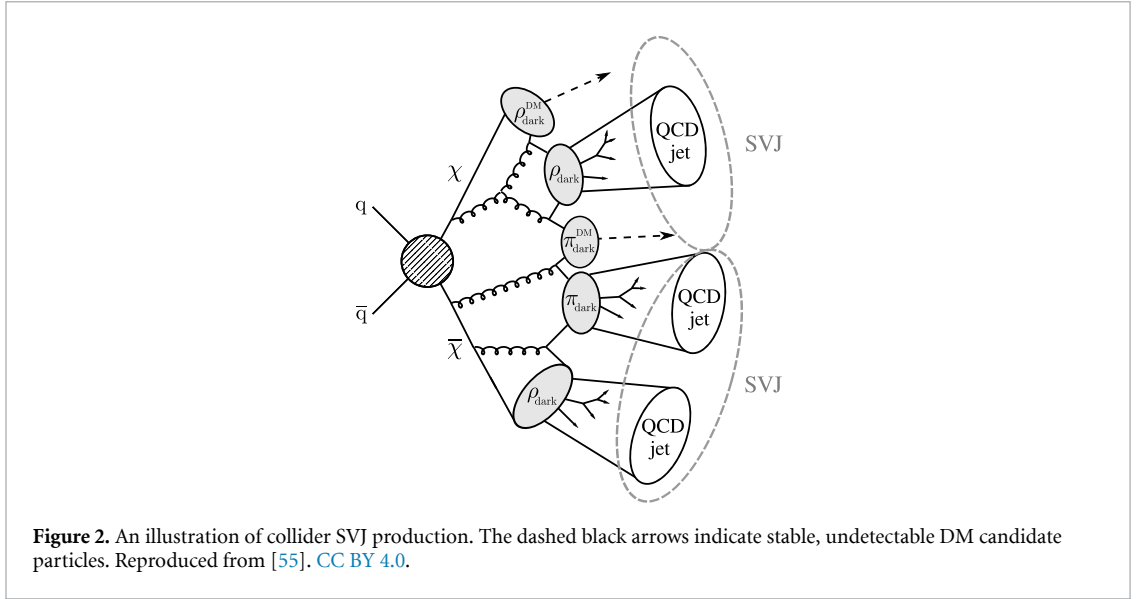
Since the first results using AEs to search for new physics at the LHC [1, 2], there has been active research on this topic. The CMS experiment has implemented two realtime anomaly detection algorithms in the level-1 trigger to select new physics signatures: one based on a VAE using global trigger information, and another based on a convolutional AE using calorimeter trigger information [42]. The CMS experiment has recently conducted a search for new physics in dijet resonances with anomalous jet substructure using five different anomaly detection algorithms, two of which are based on AEs [43]. The ATLAS experiment looked for new physics in the invariant mass distribution of pairs of SM objects consisting of one light-flavor or b jet, and a lepton, a photon, or a second light-flavor or b jet, using an AE to select anomalous data [44]. NAEs were first used in high-energy physics in the context of jet tagging for new physics searches [45].

3. Case studies

3.1. Data set generation

Astronomical observations [46–50] provide compelling evidence for dark matter (DM) in the Universe. However, particles that could constitute the DM are not found in the SM of particle physics. If DM particles can be produced in high-energy collisions at the LHC by new physics processes, their existence can be inferred from the signature they leave in detectors. This work considers a specific class of models of new physics, described below.

Hidden valley models [51] are new physics scenarios in which the SM is expanded with a dark sector that has its own particles and forces. The dark sector particles are not charged under any SM gauge group and hence are not directly detectable. Similarly, SM particles are not charged under the dark



gauge group. Hidden valley theories may, under certain conditions, result in SVJs as an experimental signature [5], where some of the particles in the final state are not detectable. If the dark sector has a new non-Abelian $SU(N)$ confining interaction, it would exhibit showering and hadronization qualitatively similar to the SM quantum chromodynamics (QCD). In analogy with the SM sector, the new confining force and the new particles charged under it are referred to as dark QCD and dark quarks, respectively. More specifically, we consider a dark QCD interaction $SU(N_c^{\text{dark}})$, where N_c^{dark} denotes the number of dark colors, with a confinement scale Λ_{dark} and N_f^{dark} flavors of dark quark χ with a mass m_χ , which communicates with the SM via a mediator Φ with mass m_Φ . In such a model, QCD-like showers occur when $m_\Phi \gg \Lambda_{\text{dark}}$ and $m_\chi \sim \Lambda_{\text{dark}}$. The dark quarks therefore form bound states called dark hadrons. The unstable dark hadrons, π_{dark} and ρ_{dark} , decay back to SM particles to produce a visible signature, depending on the symmetries of the theory. The stable dark hadrons, $\pi_{\text{dark}}^{\text{DM}}$ and $\rho_{\text{dark}}^{\text{DM}}$, are invisible and escape detection, giving rise to missing transverse momentum. The ensemble of stable dark hadrons and the decay products of unstable dark hadrons from the dark QCD shower constitutes an SVJ. The fraction of stable dark hadrons in the dark QCD shower can be parametrized by the invisible fraction r_{inv} :

$$r_{\text{inv}} = \left\langle \frac{\text{Number of stable dark hadrons}}{\text{Total number of dark hadrons}} \right\rangle. \quad (17)$$

Figure 2 illustrates the production of SVJs. The CMS and ATLAS Collaborations have published searches for resonant production of pairs of SVJs [52, 53]. The production of SVJs via t -channel interactions through a bifundamental scalar mediator, charged under both SM QCD and dark QCD, is also possible [54], with one published search from ATLAS [53]. The latter production mode is considered in this paper.

The SVJs may have different radiation patterns than SM jets, which can be exploited to identify them. The details depend on the exact spectrum of the dark hadrons, which in turn depends on N_c^{dark} , N_f^{dark} , Λ_{dark} , the masses m_χ , and the nonperturbative physics in the dark sector, all of which are unknown. Therefore, anomaly detection, which does not require a detailed simulation of every target signal hypothesis in this large parameter space, is a powerful complement to traditional, supervised searches for SVJs, such as [52].

Simulated events are used to train and evaluate the performance of the machine-learning models discussed in this paper. The background considered in this study consists of top-antitop quark production ($t\bar{t}$) with additional jets, where each top quark decays either hadronically or leptonically. Background and SVJ signal events are generated at the parton level with MadGraph5_aMC@nlo 2.6.5 [56], final-state partons are hadronized with PYTHIA 8.240 [57], and the interactions of the resulting particles with the CMS detector [6] are simulated using GEANT4 [58]. The NNPDF3.1 parton distribution functions are used for all processes [59]. The hadronization in the dark sector during the signal generation follows the procedure described in reference [52], and all dark hadrons have a fixed mass, 20 GeV. Because of the double hadronization, first in the dark sector and then in the SM sector, the substructure of SVJs

is expected to be different from that of SM jets. In particular, SVJs are expected to be wide, and therefore jets are reconstructed using the anti- k_T algorithm [60, 61] with radius parameter $R = 0.8$. Each jet is represented by the following features, describing its substructure:

- the major and minor axes [62], which characterize the elliptical shape of the jet in the η - ϕ plane (where η is the pseudorapidity and ϕ is the azimuthal angle);
- the first energy flow polynomial (EFP1) [63], a multiparticle energy correlator that directly results from infrared and collinear safety;
- the $C_2^{(0.5)}$ energy correlation function [64], a ratio of two ratios of energy correlation functions that can be used to determine if a jet has two hard subjects;
- the transverse momentum dispersion p_T^D [62], which describes the spread of transverse momenta of the jet constituents;
- the softdrop mass [65], the mass of the jet after soft wide-angle radiation has been removed; and
- the 2- and 3-subjettiness τ_2 and τ_3 [66], describing the compatibility of the jet with a two- or three-prong structure, respectively;

for a total of $d = 8$ inputs to the AE. These input features are chosen to exploit potential differences between SM jets and BSM jets in shape (major and minor axes), energy distribution (p_T^D and EFP1), prong structure ($C_2^{(0.5)}$, τ_2 , τ_3), and mass of the high-momentum core of the jet (softdrop mass). The distributions of these features in the physical space, before any preprocessing, can be found in appendix A. These features represent a space of physical variables that facilitates the classification task by providing some separation between signal and background, even though a single boundary in this space, unbiased across all possible signal models, may be difficult to define. For each event, the two jets with highest transverse momenta are used for training and evaluation. In addition, among the two selected jets, only jets identified via generator-level information as originating from dark-sector particles are considered as anomalous signal jets when evaluating the network performance. The inputs are preprocessed to follow a normal distribution using quantile normalization, as implemented in the SCIKIT-LEARN package [67].

3.2. Training setup

The architecture is a fully connected network with five hidden layers. All layers have ten nodes, except for the third one, which, as the bottleneck, has six. The performance of the different models discussed in this paper was found to have little dependence on either the number of layers or the number of nodes per layer; the listed choices are found to produce a sufficiently expressive network for this anomaly detection task. The reconstruction of the AE is parametrized by the network weights θ . The reconstruction error l_θ for an example x is defined as the mean squared error between the features x_i of x and the features $\hat{x}_{\theta,i}$ of the reconstructed output:

$$l_\theta(x) = \frac{1}{d} \sum_{i=1}^d (x_i - \hat{x}_{\theta,i})^2. \quad (18)$$

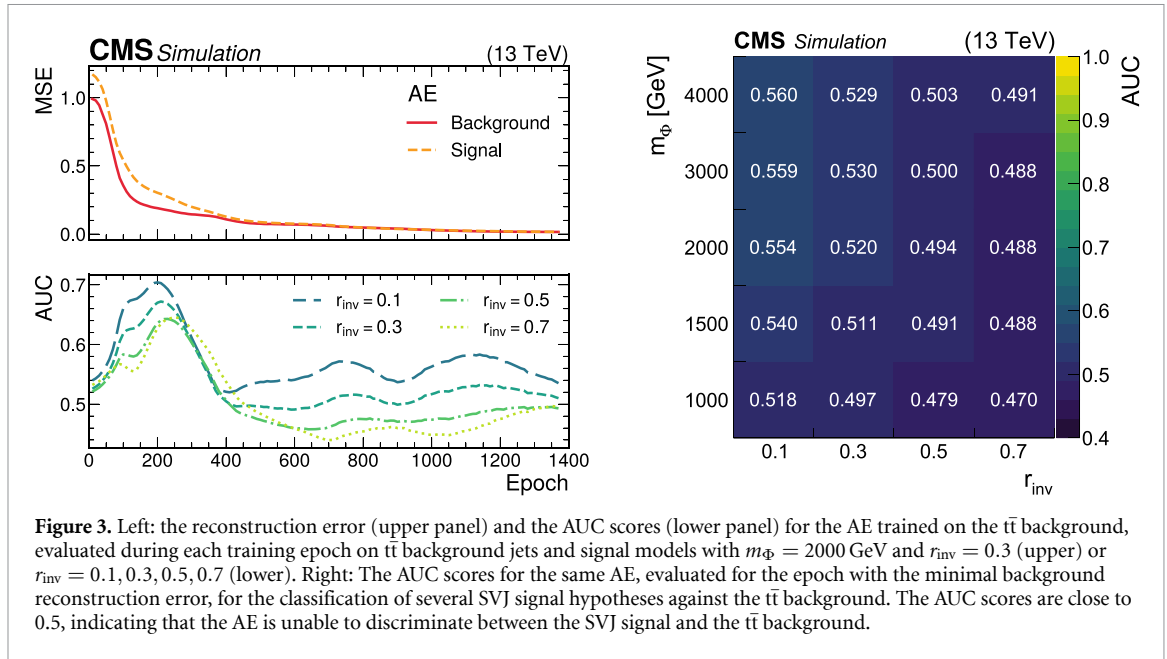
The loss function of the AE is the reconstruction error averaged over the batch:

$$\mathcal{L}_{\text{AE}} = \frac{1}{n} \sum_{j=1}^n l_\theta(x^{(j)}), \quad (19)$$

where $x^{(j)}$ denotes the example in position j in the batch and n is the number of training examples in the batch. The background data set is split into three independent parts: a training set on which the loss is minimized (70%), a validation set used to monitor the training (15%), and a test set used to evaluate the performance of the network (15%). The background data set has a total of 47 000 jets and the number of trainable parameters is 534. The AE is trained until the loss function evaluated on the validation set ceases to improve. The AE is then tested using multiple SVJ signal models, as well as the background test sample.

3.3. AE training

The issue of outlier reconstruction is found to be especially impactful in the case of the SVJ search. While previous studies have shown that AEs are an effective tool for the task of discriminating between



the signal and the SM background from light-flavor jets [68], the performance is found to degrade significantly when jets originating from t quarks in the $t\bar{t}$ process are included. This is shown by training an AE solely on the $t\bar{t}$ background. The cause is understood to be outlier reconstruction, based on the nature of $t\bar{t}$ events. Boosted jets from high-momentum top quarks, which contain multiple prongs and genuine missing transverse momentum from neutrinos, are the closest SM analog to SVJs. These events span multiple topologies, as they can also include boosted W boson jets, light-flavor quark jets, and gluon jets from initial- or final-state radiation.

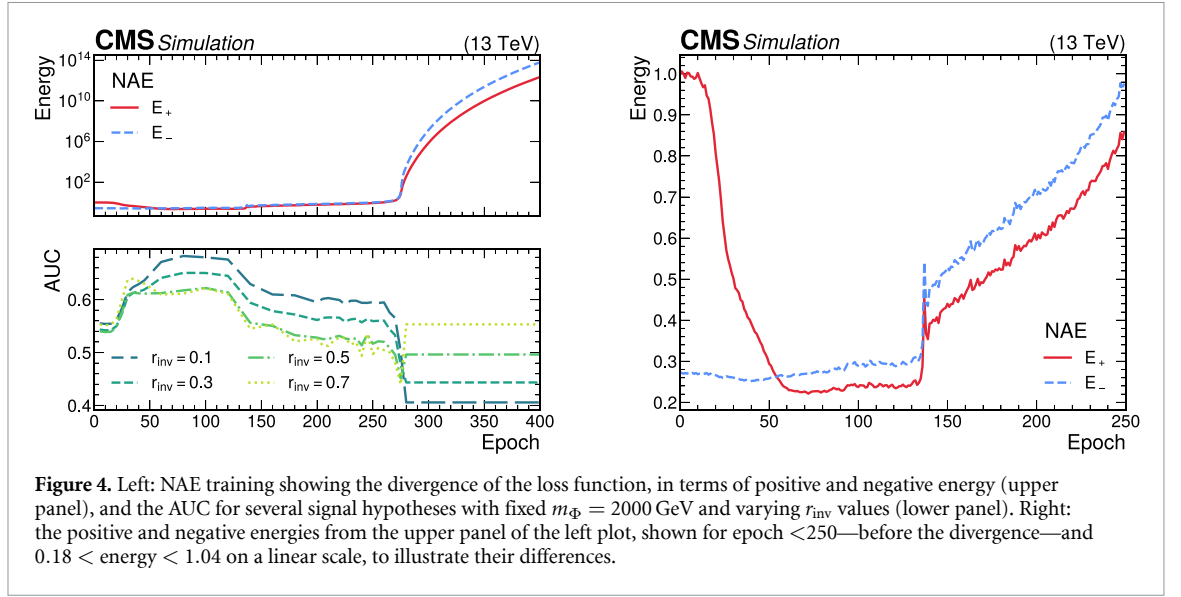
The receiver operating characteristic curve compares the true- and false-positive rates for the AE reconstruction error used as a discriminator, and the area under the curve (AUC) is used to quantify the performance of this discriminator. As shown in figure 3, when achieving minimum reconstruction error on the $t\bar{t}$ background, the AE is unable to discriminate between the SVJ signal and the $t\bar{t}$ background, and the AUC score is close to 0.5. In order to check that this behavior does not occur because the bottleneck was too large, the same experiment was repeated with a bottleneck of only five, four, three, or two nodes. The results were found to be consistent with a bottleneck size of six, indicating that the issue does not arise from insufficient compression of the data.

The AUC scores evolve in nontrivial ways throughout the training. As shown in figure 3, the AUC score improves during the first stages of the training, indicating that the AE is performing as desired. However, after further training, the AUC score starts to degrade, and eventually converges to a value close to 0.5. This is a clear indication that the requirement to minimize the reconstruction error on the $t\bar{t}$ background is too weak to achieve good discrimination; it does not prevent the AE from also reconstructing signal jets.

3.4. NAE training

The most delicate part of the NAE training is the tuning of the MCMC. A suboptimal MCMC would not accurately sample the distribution p_θ , and in turn, the NAE would not be able to learn the training data probability distribution p_{data} . In theory, the convergence of an MCMC algorithm does not depend on the random initial set of points from which the Markov chain starts. However, there are two motivations for using a nonrandom initial set of points. First, the Markov chain is not infinitely long, and so there is no guarantee of convergence starting from any initial distribution. Second, from one batch to another, the neural network weights and p_θ both change only slightly, so a good initialization for batch k usually also works for batch $k + 1$. In addition, tuning the hyperparameters of the Markov chain is a complex task since the p_θ distribution is unknown. The procedure to initialize the MCMC and tune its associated hyperparameters, which accounts for these considerations, is detailed in appendix D.

After tuning the MCMC algorithm, the evolution of the positive and negative energies of an NAE with the loss function defined in equation (10) is presented in figure 4. It shows two failure modes: negative loss values, indicating the negative energy is higher than positive energy; and divergence of both energies.



Negative values of the energy difference occur when the network did not learn the background probability distribution, such that p_{data} is higher than p_θ in regions with low reconstruction error. This is a severe failure mode, illustrating that energy difference minimization is not sufficient to guarantee the intended behavior.

The divergence of the negative energy was also reported in previous work [4, 45]. As a consequence of the form of the loss function, after the positive energy has reached a minimal value, further minimization of the loss is achieved by increasing the negative energy, effectively leading to a divergence of both energy terms and the loss itself, as shown in figure 4. The distributions of a selected input feature τ_3 for the positive, negative, and signal samples before and after the start of the divergence are provided in figure 5. The large reconstruction error after the divergence leads to very large gradient values and thus large steps in the MCMC. This results in the negative samples reaching the boundaries of the sampling phase space and no longer being representative of p_θ .

A simple modification of the NAE loss can mitigate these issues:

$$\mathcal{L} = \log \cosh(E_+ - E_-). \quad (20)$$

Because the logcosh function is positive definite, this loss removes the incentive to favor negative values of the energy difference and leads to a more stable training procedure. The evolution of the energies, loss value, and AUC score for a training with this loss function are shown in figure 6. Their behavior in this figure can be qualitatively understood as follows. There is a first stage in which the energy difference is minimized, giving the largest improvement in AUC. Once the difference between positive and negative energies is approximately zero, there is a second stage in which the network fine-tunes its knowledge of p_θ , further increasing the AUC score until it reaches a plateau around epoch 4000. There is subsequently a third stage in which a new form of mode collapse is observed: the energy difference is still approximately zero and the positive energy is further minimized, but the AUC score begins to decrease. Crucially, it is not possible to avoid this degradation by relying solely on the positive and negative energies. On the other hand, making use of the AUC score as a stopping condition for the training would be in contradiction with the unsupervised nature of the NAE. The evolution of the Wasserstein distance, also shown in this figure, is discussed in section 4.

This collapse arises as a form of overtraining of the network, leading to outlier reconstruction. While keeping the reconstruction error low, the network eventually achieves low reconstruction in a region $\mathcal{E}$ that contains, but is larger than, the support of the training data $\mathcal{B}$, while keeping $\mathbb{E}_{x \sim \mathcal{E}}[E(x)]$ comparable to $\mathbb{E}_{x \sim \mathcal{B}}[E(x)]$. As the overlap of $\mathcal{E}$ with the support of the signal $\mathcal{S}$ becomes larger, the AUC score decreases. This behavior is shown schematically in figure 7. Before the mode collapse, $\mathcal{E}$ and $\mathcal{B}$ overlap, while $\mathcal{E}$ and $\mathcal{S}$ do not. After the mode collapse, $\mathcal{E}$ expands and can partially include $\mathcal{S}$, reducing the difference in reconstruction error between background and signal and correspondingly lowering the AUC score. In both cases, the difference between the positive and the negative energies is zero.

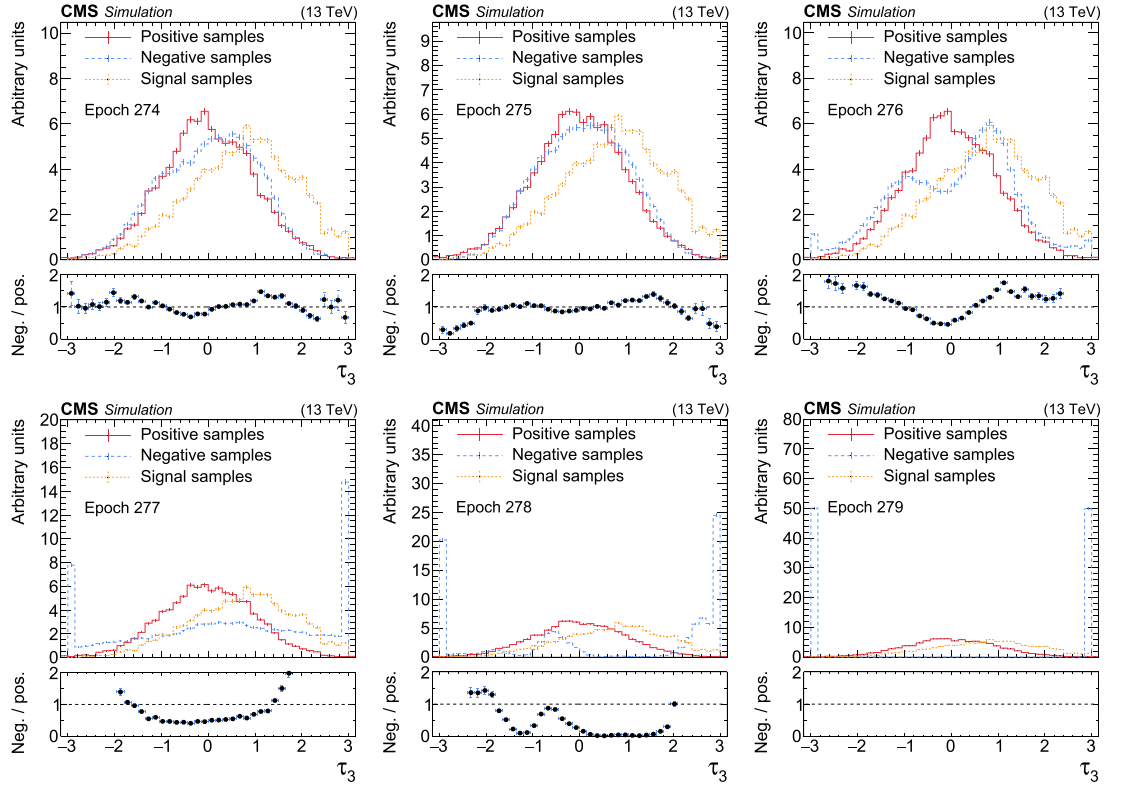


Figure 5. Distributions of the input feature τ_3 for positive, negative, and signal samples, before (epochs 274) and after (epochs 275–279) the start of the divergence of the NAE loss. The signal distributions are overlaid for illustration; signal samples are not used during the training. All distributions are normalized such that their integral is 100.

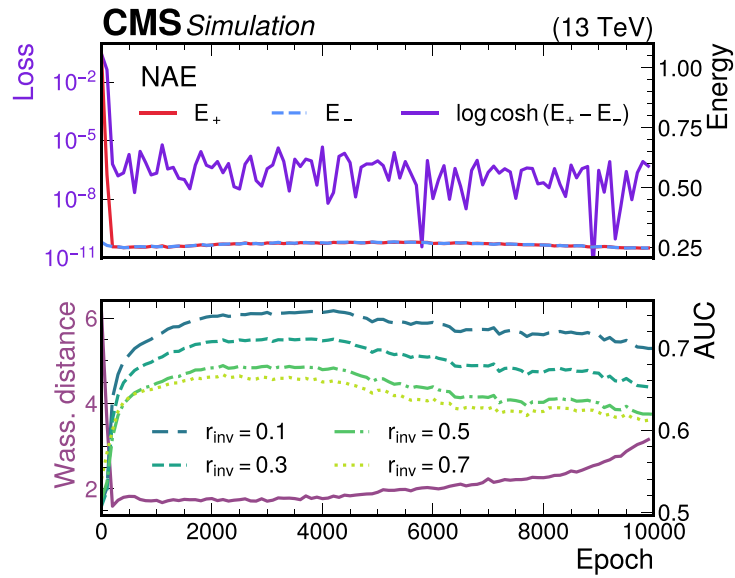
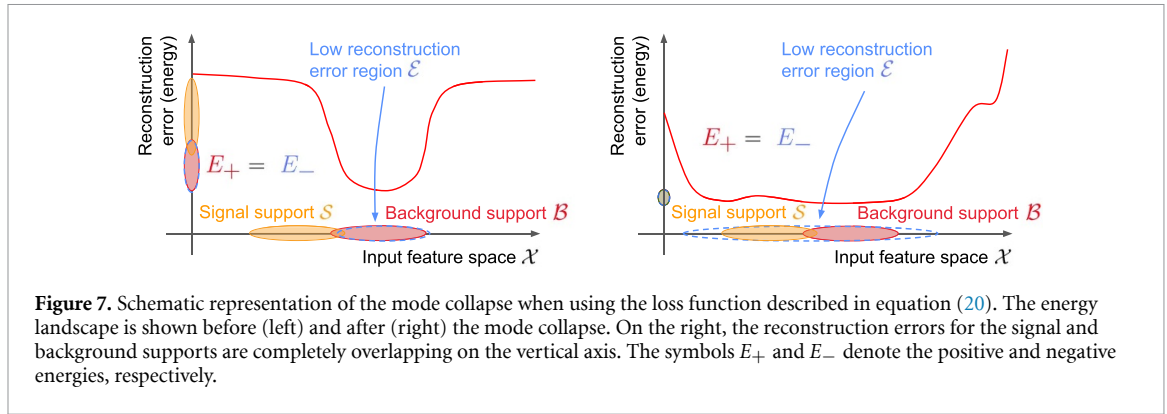


Figure 6. As a function of epoch during the training of an NAE with the loss function from equation (20): the positive and negative energies and the value of the loss function (upper panel); the Wasserstein distance between negative and positive samples and the AUC for several signal hypotheses with fixed $m_\Phi = 2000$ GeV and varying r_{inv} (lower panel).

4. The WNAE

4.1. The Wasserstein distance as a measure of outlier reconstruction

The failure mode described at the end of section 3.4 is a consequence of the fact that the energy difference is unable to distinguish the two cases shown in figure 7 (left). In fact, no stopping condition based solely on combinations of positive and negative energies was found to be a good indicator of the start



of this mode collapse. In other words, without explicitly monitoring the AUC during training, there is no clear stopping condition; choosing a stopping condition based on the AUC, however, violates the principle of unsupervised learning. In order to overcome this problem, a different metric is used: the 1-Wasserstein distance W (also known as the earth mover's distance) from equation (15) between the probability distribution of the training data p_{data} and the probability distribution learned by the network p_{θ} , in analogy to what was applied successfully to generative adversarial networks [30]:

$$W(p_{\text{data}}, p_{\theta}) = \inf_{\gamma \in \Pi(p_{\text{data}}, p_{\theta})} \mathbb{E}_{(x, x') \sim \gamma} [\|x - x'\|]. \quad (21)$$

The 1-Wasserstein distance is chosen for consistency with other machine learning applications, which use this distance rather than any other p -Wasserstein distance because the Kantorovich–Rubinstein duality is only known to hold for $p = 1$. Experiments using the 2-Wasserstein distance demonstrated similar performance. This metric is expected to solve the problem of outlier reconstruction by measuring the distance between the p_{θ} and p_{data} distributions directly in the input feature space. By minimizing $W(p_{\text{data}}, p_{\theta})$, these two probability distributions are forced to match, which prevents feature space regions distinct from p_{data} from having low reconstruction error, achieving the desired behavior illustrated in figure 7 (left). The Wasserstein distance is, indeed, found to be a good indicator of the start of the mode collapse, as seen in figure 6 around epoch 4000: the Wasserstein distance increases significantly when this collapse occurs and the AUC decreases. The point where the Wasserstein distance increases, which can be quantified as the point where its numerical derivative exceeds an empirical threshold for the first time during the training, can then be identified as a stopping condition for the training of the NAE. It is worth stressing that the Wasserstein distance, calculated from the positive and negative samples, is completely agnostic to the signal, and yet it is found to be closely correlated with the performance of the NAE as an outlier detector. With this prescription, the NAE is able to successfully learn the $t\bar{t}$ distribution and effectively detect the signal, as shown in figure 8.

4.2. The Wasserstein distance as a loss function

Having shown that the Wasserstein distance is a good indicator of the performance of the NAE as an outlier detection algorithm, it is natural to use it directly as a loss function. Compared to other common distances or divergences between probability distributions, such as the Kullback–Leibler or Jensen–Shannon divergences, the Wasserstein distance has better analytic properties. In particular, if g_{θ} is a feed-forward neural network parametrized by weights θ with probability distribution p_{θ} , and p is a fixed distribution, then the Wasserstein distance $W(p_{\theta}, p)$ is continuous everywhere and differentiable almost everywhere with respect to θ . This is not the case for the Kullback–Leibler and Jensen–Shannon divergences between p_{θ} and p . A review and proof of these analytic properties can be found in [30].

Correctly incorporating the Wasserstein distance in the neural network training requires two computational graphs: one for differentiation with respect to the input feature space during each step of the Markov chain, another for differentiation with respect to the network weights θ during gradient descent. The procedure to implement this is detailed in appendix B. The resulting model is called the WNAE, with source code publicly available online [69]. The flowchart for training the WNAE is shown in figure 9.

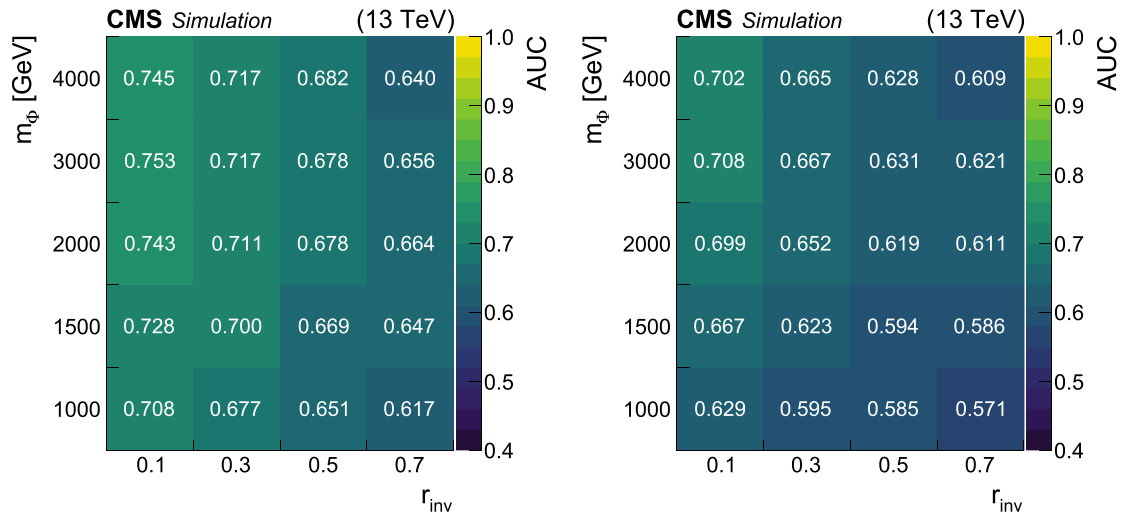


Figure 8. Left: the AUC scores for an NAE trained on the $t\bar{t}$ background and tested against a grid of possible SVJ signals, before the increase of the Wasserstein distance (at epoch 3000). Right: the AUC scores for the same NAE after the increase in Wasserstein distance (at epoch 10000).

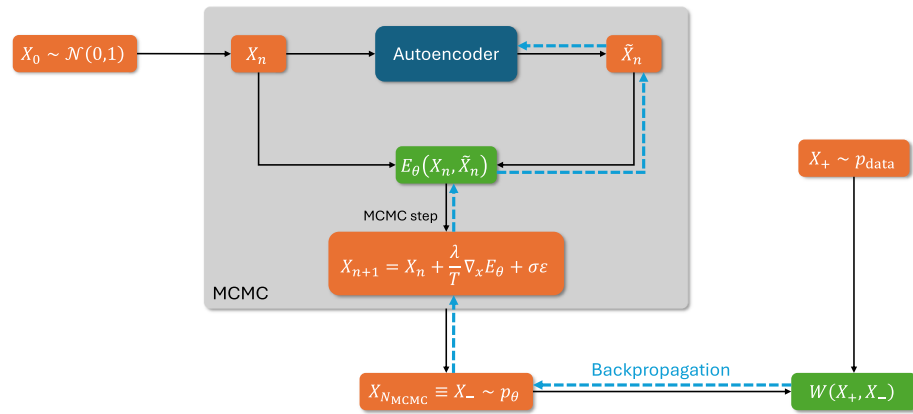


Figure 9. Flowchart of the Wasserstein normalized autoencoder training. The negative examples are generated via MCMC using the energy function of the model, which is the Boltzmann-distributed reconstruction error. The energy function is computed from random input feature values X_n and the corresponding reconstructed feature values $\tilde{X}_n$, obtained by passing the inputs through the autoencoder. The positive examples are compared to the negative examples through the Wasserstein distance. The gradients are backpropagated through the entire MCMC chain.

4.3. Training the WNAE

The training is found to qualitatively proceed in two phases: first, a coarse adjustment in which the Wasserstein distance decreases sharply while the model adapts p_θ to produce negative samples that resemble physical examples, and second, a fine-tuning phase in which the model learns the specifics of the training sample. The use of a learning rate scheduler was found to decrease the training time while preserving the stability of the training. In the case of a constant learning rate, if the learning rate is too small, the time needed for the first phase increases significantly, while if it is too large, the training becomes unstable in the second phase. Technical details about the learning rate scheduler can be found in appendix C.

The computational bottleneck for training the WNAE is the calculation of the Wasserstein distance between the positive and negative samples, which takes approximately 90% of the computation time for each training batch. The complexity of the Wasserstein distance scales linearly with the dimensionality of the data d and at worst cubically with the number of data points n [70]. In practice, for the data set used in this paper, the complexity is measured to scale quadratically with n , similar to reference [70]. The second most impactful operation is the backpropagation through the MCMC chain, which takes

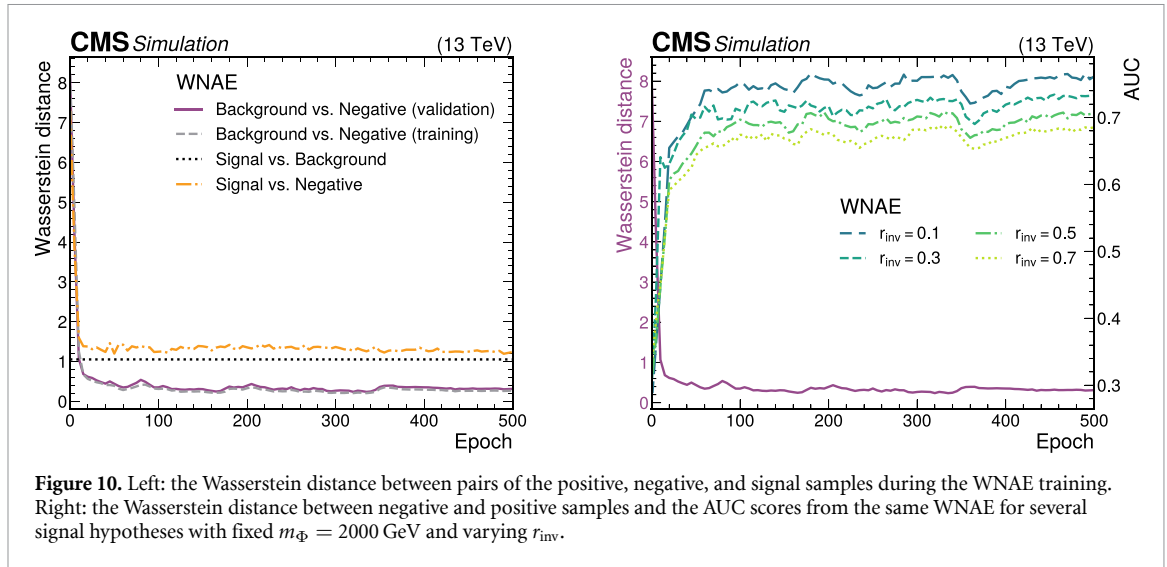


Figure 10. Left: the Wasserstein distance between pairs of the positive, negative, and signal samples during the WNAE training. Right: the Wasserstein distance between negative and positive samples and the AUC scores from the same WNAE for several signal hypotheses with fixed $m_{\Phi} = 2000$ GeV and varying r_{inv} .

approximately 7% of the computation time. The time needed for the calculation of the Wasserstein distance heavily depends on the batch size, which was set to 4096 to allow for a robust estimation of the distance between positive and negative samples. Lowering the batch size reduces the computation time at the expense of decreasing the stability of the loss function. Depending on the batch size, the training of the WNAE can thus be significantly slower than that of the NAE, by up to an order of magnitude for large enough batch sizes. However, the Wasserstein distance is found to provide a much more reliable and stable stopping condition. All of these computational considerations only impact training; at inference time, the WNAE behaves like a standard AE, with no additional computational overhead.

4.4. Performance of the WNAE

The WNAE is found to be significantly more stable in training, not showing any of the forms of divergence described previously in section 3.4. It is able to achieve strong anomaly detection performance while using the minimum of the validation loss, the Wasserstein distance evaluated on the validation set, as a clear stopping condition. Figure 10 shows the evolution of the Wasserstein distance and the AUC score for several SVJ signal hypotheses. The minimum of the Wasserstein distance corresponds to the maximum discrimination power. Figure 11 shows the AUC score evaluated for a grid of SVJ signals, showing good discrimination power across the parameter space, on par with or better than any of the NAE models trained above. The WNAE is thus found to be a more reliable, stable, and effective tool for anomaly detection in the context of the SVJ search. In addition, this algorithm can achieve the same performance with all hidden layers having higher dimension than the input feature space. The requirement to minimize the Wasserstein distance controls the behavior of the network without separately requiring an architecture with a lower-dimensional bottleneck.

The fact that the WNAE behaves as desired can be seen in figures 12 and 13, where the distributions of the input features for the positive, negative, and signal samples are shown at the start and end of the training. Starting from an essentially multivariate distribution, induced by the random weight initialization of the network, the WNAE is able to learn the distribution of the training data, as evidenced by the distribution of negative samples eventually converging toward that of the positive samples.

4.5. Complexity bias in the WNAE

A common weakness of plain AEs is the so-called complexity bias, in which the AE tends to be able to discriminate effectively only against anomalous examples whose distribution in input space is more complex than that of the training data. This can be demonstrated by inverting the background and signal samples and observing that, in this case, an AE trained on SM jets can tag SVJs as anomalous, but the reverse is not true. Because the WNAE learns the probability distribution of the training data in input space, this issue is mitigated, as shown in figure 14. The high AUC at the start of the training is caused by the random uniform initialization of the network weights and the different correlations between the input features of the signal and background jets. The AUC then decreases as the network learns a distribution closer to the physical input data, before increasing again as the network starts to more precisely learn the distribution of SVJ examples and thus becomes able to tag SM jets as anomalous.

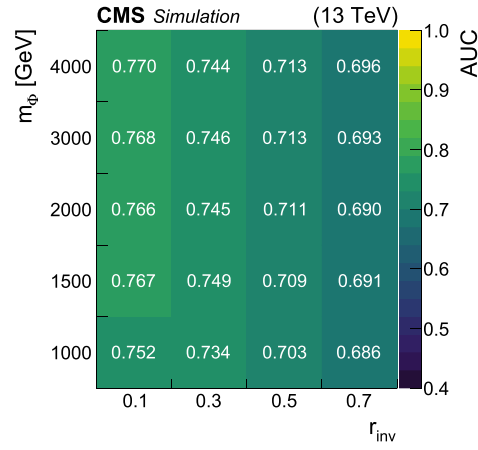


Figure 11. The AUC scores for a WNAE trained on the $t\bar{t}$ background and tested on a grid of possible SVJ signal models. The statistical uncertainty in the AUC scores, evaluated by retraining the WNAE with different random initializations, is found to be less than 0.02.

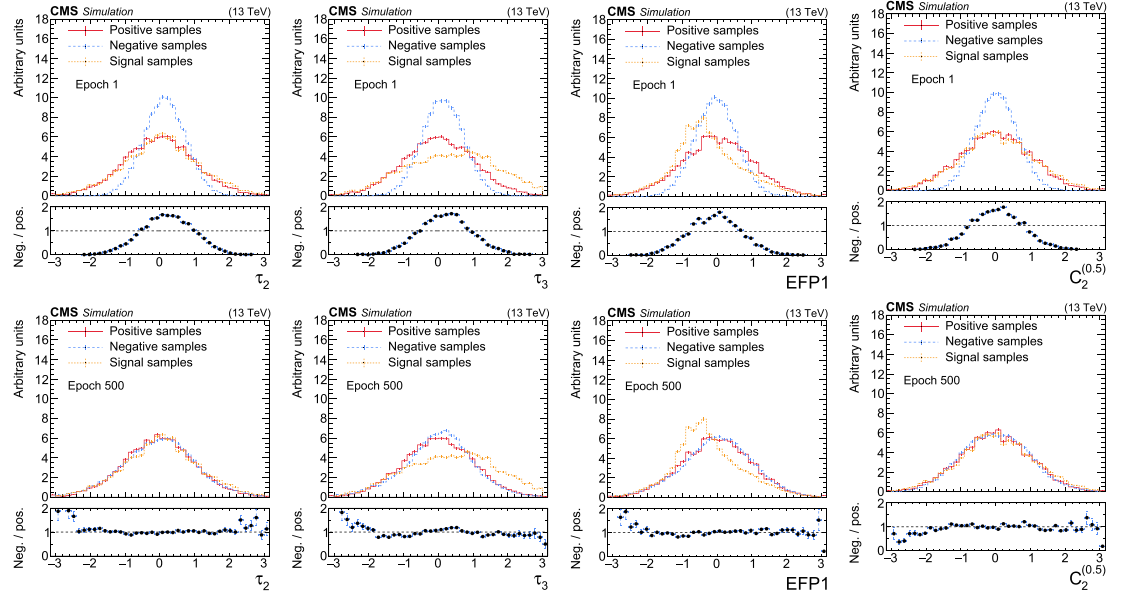


Figure 12. The distributions of half of the input variables, τ_2 , τ_3 , EFP1, and $C_2^{(0.5)}$, for the positive, negative, and signal samples, at the start (upper) and at the end (lower) of the WNAE training. The signal distributions are overlaid for illustration; signal samples are not used during the training. All distributions are normalized such that their integral is 100.

5. Conclusions

Anomaly detection using AEs relies on learning a reconstruction function that gives high reconstruction error to phase space regions with low probability density, such that they can be identified as anomalous. However, standard AEs are prone to learn to reconstruct outliers because they are free to minimize the reconstruction error outside the training phase space. In addition, they may exhibit complexity bias, learning to identify examples as anomalous only if their feature distribution is more complex than the training data. The NAE paradigm promotes the AE reconstruction error to an energy function in the framework of energy-based models, in order to define a normalized probabilistic model. This is achieved by minimizing the negative log-likelihood of the training data from the energy-based model probability. In practice, this method presents a number of failure modes, such as divergence of the loss function and phase space degeneracy, leading to low reconstruction error for phase space regions distinct from the training data.

The WNAE, an improvement over the NAE, is introduced to solve these failure modes. This is achieved by directly minimizing the Wasserstein distance between the probability distribution of the training data and the Boltzmann distribution of the energy function of the model. This Wasserstein

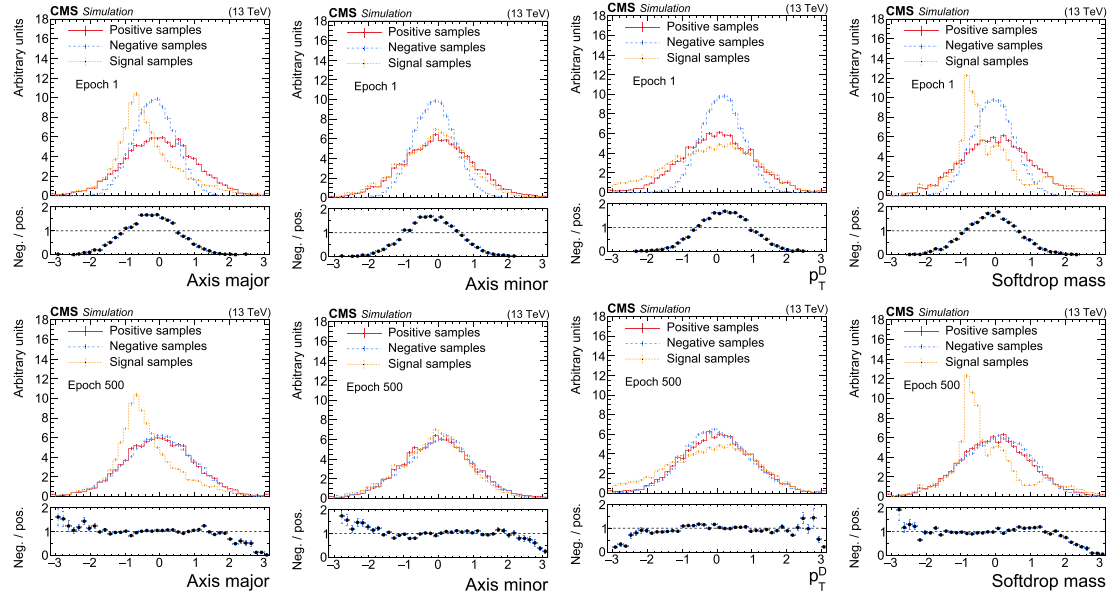


Figure 13. The distributions of the other half of the input variables, axis major, axis minor, p_T^D , and softdrop mass, for the positive, negative, and signal samples, at the start (upper) and at the end (lower) of the WNAE training. The signal distributions are overlaid for illustration; signal samples are not used during the training. All distributions are normalized such that their integral is 100.

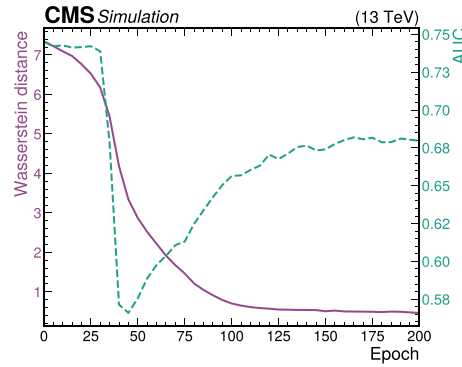


Figure 14. The Wasserstein distance between the positive and negative samples and the AUC score during the training of a WNAE on an SVJ signal ($m_\Phi = 2000$ GeV, $r_{\text{inv}} = 0.3$), with the $t\bar{t}$ background used for testing.

distance is found to be highly correlated with signal identification performance while still being fully signal-agnostic, preserving the unsupervised nature of the approach. The performance is studied in the context of a search for new physics with the CMS experiment, using top-antitop quark production as the SM background and nonresonant SVJ production from a strongly coupled dark sector as the proposed signal model. The classification of the signal jets as outliers by the WNAE is shown to be on par with or better than that of the NAE. Further, the WNAE approach is found to mitigate complexity bias, as it can effectively identify top quark jets as anomalous when trained on SVJs from signal events. While the WNAE is found to be stable and effective for the anomaly detection task studied here, it may suffer from the generic failure modes inherent to anomaly detection models and energy-based models, including: overlap of the signal and background distributions in input space; training data contaminated by anomalous examples; or failure to effectively sample the model probability distribution.

Though simulated samples were used to develop the WNAE, in practice it may be preferable to use observed data directly for training, in order to limit biases arising from differences between simulation and observation. In this case, the training data may contain anomalies, which would reduce the anomaly detection performance of the WNAE. The WNAE can be straightforwardly trained using observed data from a control region with no anomalous examples, if such a region can be defined and follows the same probability distribution as the observed data where the WNAE will be applied. When no assumption at all can be made about the nature of the anomalies, such as the case of triggering at a high-energy

physics experiment, alternative solutions may exist. The WNAE associates low probability density regions with high reconstruction error; because anomalies necessarily have low probability density, they still tend to have relatively high reconstruction error even when included in the training data set. Therefore, the training data set can be iteratively refined by selecting a given fraction of examples with the lowest reconstruction error, in order to reduce the proportion of anomalous data. This would result in a self-supervised training for the WNAE. We leave the development of such a procedure for future work.

When applying this method to learn higher-dimensional probability distributions, the slower convergence of the Wasserstein distance with increasing dimensionality may pose practical challenges. For an arbitrary d -dimensional distribution μ and an independent and identically distributed sample from μ with n data points μ_n , the convergence rate of the 1-Wasserstein distance between μ and μ_n is $n^{-1/d}$: $W(\mu, \mu_n) \gtrsim n^{-1/d}$ [71, 72]. To circumvent this behavior, approximations of the Wasserstein distance that do not suffer from this curse of dimensionality can be employed. The sliced Wasserstein distance [73, 74] has been increasingly favored because of its interesting statistical properties and low computational cost. It approximates the Wasserstein distance by computing the average of one-dimensional Wasserstein distances—for which an analytical and easily computable solution is known—between projections of the two probability distributions uniformly distributed on the unit sphere. Its convergence is equivalent to the weak convergence of the exact Wasserstein distance [75] and its convergence rate does not depend on the number of dimensions [76]. The sliced Wasserstein distance therefore converges faster than the exact Wasserstein distance when the number of dimensions increases. Differentiable implementations are available, so it can straightforwardly be used in place of the exact Wasserstein distance to perform anomaly detection in very high-dimensional feature spaces.

Acknowledgments

We congratulate our colleagues in the CERN accelerator departments for the excellent performance of the LHC and thank the technical and administrative staffs at CERN and at other CMS institutes for their contributions to the success of the CMS effort. In addition, we gratefully acknowledge the computing centers and personnel of the Worldwide LHC Computing Grid and other centers for delivering so effectively the computing infrastructure essential to our analyses. Finally, we acknowledge the enduring support for the construction and operation of the LHC, the CMS detector, and the supporting computing infrastructure provided by the following funding agencies: SC (Armenia), BMBWF and FWF (Austria); FNRS and FWO (Belgium); CNPq, CAPES, FAPERJ, FAPERGS, and FAPESP (Brazil); MES and BNSF (Bulgaria); CERN; CAS, MoST, and NSFC (China); MINCIENCIAS (Colombia); MSES and CSF (Croatia); RIF (Cyprus); SENESCYT (Ecuador); ERC PRG, TARISTU24-TK10 and MoER TK202 (Estonia); Academy of Finland, MEC, and HIP (Finland); CEA and CNRS/IN2P3 (France); SRNSF (Georgia); BMFTR, DFG, and HGF (Germany); GSRI (Greece); NKFIH (Hungary); DAE and DST (India); IPM (Iran); SFI (Ireland); INFN (Italy); MSIT and NRF (Republic of Korea); MES (Latvia); LMTLT (Lithuania); MOE and UM (Malaysia); BUAP, CINVESTAV, CONACYT, LNS, SEP, and UASLP-FAI (Mexico); MOS (Montenegro); MBIE (New Zealand); PAEC (Pakistan); MES, NSC, and NAWA (Poland); FCT (Portugal); MESTD (Serbia); MICIU/AEI and PCTI (Spain); MOSTR (Sri Lanka); Swiss Funding Agencies (Switzerland); MST (Taipei); MHESI (Thailand); TUBITAK and TENMAK (Türkiye); NASU (Ukraine); STFC (United Kingdom); DOE and NSF (USA).

Individuals have received support from the Marie-Curie program and the European Research Council and Horizon 2020 Grant, Contract Nos. 675440, 724704, 752730, 758316, 765710, 824093, 101115353, 101002207, 101001205, and COST Action CA16108 (European Union); the Leventis Foundation; the Alfred P. Sloan Foundation; the Alexander von Humboldt Foundation; the Science Committee, Project No. 22rl-037 (Armenia); the Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture (FRIA) and Fonds voor Wetenschappelijk Onderzoek Contract No. 1228724N (Belgium); the Beijing Municipal Science & Technology Commission, No. Z191100007219010, the Fundamental Research Funds for the Central Universities, the Ministry of Science and Technology of China under Grant No. 2023YFA1605804, and the Natural Science Foundation of China under Grant No. 12061141002, 12535004 (China); the Ministry of Education, Youth and Sports (MEYS) of the Czech Republic; the Shota Rustaveli National Science Foundation, Grant FR-22-985 (Georgia); the Deutsche Forschungsgemeinschaft (DFG), among others, under Germany's Excellence Strategy—EXC 2121 'Quantum Universe'—390833306, and under Project Number 400140256—GRK2497; the Hellenic Foundation for Research and Innovation (HFRI), Project Number 2288 (Greece); the Hungarian Academy of Sciences, the New National Excellence Program—ÚNKP, the NKFIH research Grants

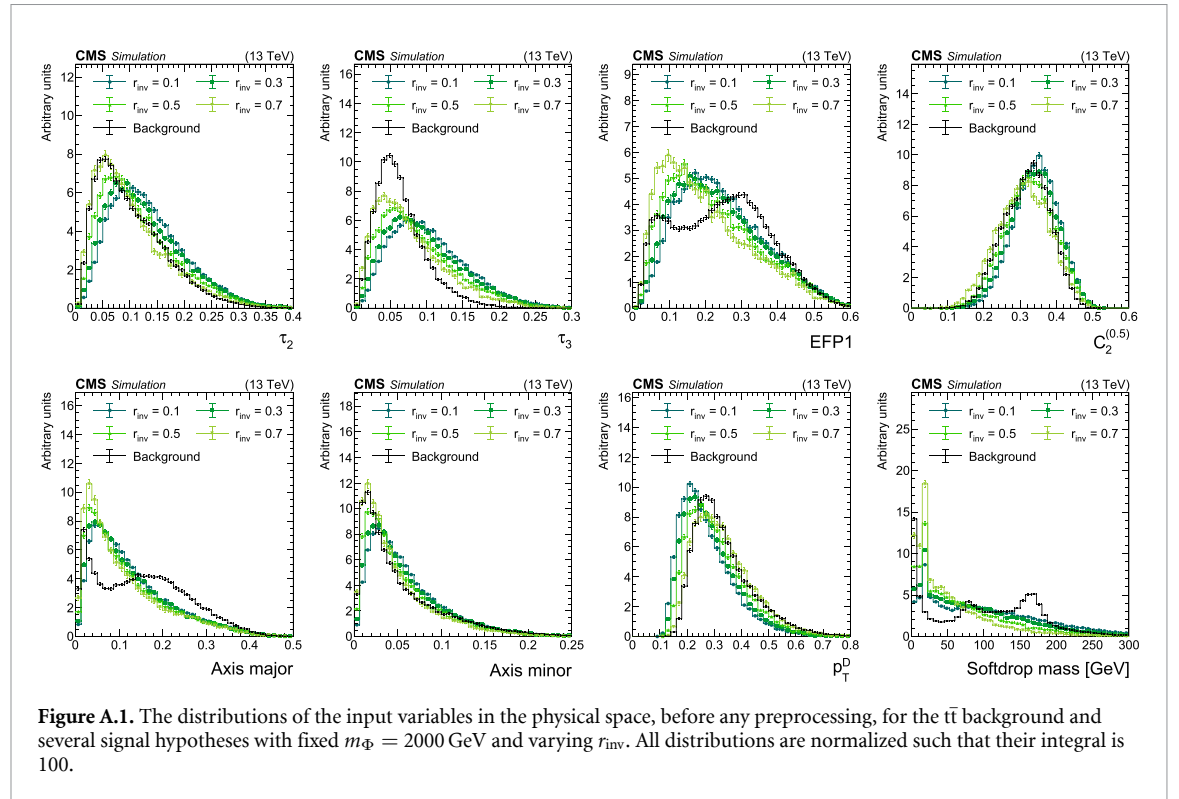
K 131991, K 133046, K 138136, K 143460, K 143477, K 146913, K 146914, K 147048, 2020-2.2.1-ED-2021-00181, TKP2021-NKTA-64, and 2021-4.1.2-NEMZ_KI-2024-00036 (Hungary); the Council of Science and Industrial Research, India; ICSC—National Research Center for High Performance Computing, Big Data and Quantum Computing, FAIR—Future Artificial Intelligence Research, and CUP I53D23001070006 (Mission 4 Component 1), funded by the NextGenerationEU program (Italy); the Latvian Council of Science; the Ministry of Education and Science, Project No. 2022/WK/14, and the National Science Center, Contracts Opus 2021/41/B/ST2/01369, 2021/43/B/ST2/01552, 2023/49/B/ST2/03273, and the NAWA Contract BPN/PPO/2021/1/00011 (Poland); the Fundação para a Ciência e a Tecnologia, Grant CEECIND/01334/2018 (Portugal); the National Priorities Research Program by Qatar National Research Fund; MICIU/AEI/10.13039/501100011033, ERDF/EU, ‘European Union NextGenerationEU/PRTR’, and Programa Severo Ochoa del Principado de Asturias (Spain); the Chulalongkorn Academic into Its 2nd Century Project Advancement Project, the National Science, Research and Innovation Fund program IND_FF_68_369_2300_097, and the Program Management Unit for Human Resources & Institutional Development, Research and Innovation, Grant B39G680009 (Thailand); the Kavli Foundation; the Nvidia Corporation; the SuperMicro Corporation; the Welch Foundation, Contract C-1845; and the Weston Havens Foundation (USA).

Data availability statement

Release and preservation of data used by the CMS Collaboration as the basis for publications is guided by the CMS data preservation, re-use, and open access policy at <https://doi.org/10.7483/OPENDATA.CMS.1BNU.8V1W>.

Appendix A. Input features in physical space

The distributions of the 8 input features used to train the models presented in this paper are presented in figure A.1, in the physical space before any preprocessing is applied. Semivisible jets tend to resemble background $t\bar{t}$ jets more when r_{inv} increases.



Appendix B. Gradient dependence between MCMC and AE

During the MCMC process, the gradient with respect to the feature space needs to be calculated to compute each step. At the same time, the dependency on the weights of the neural network must be retained in order to compute the gradient with respect to the weights for backpropagation.

For generative applications, a model g_θ , parametrized by weights θ , may be trained to minimize the Wasserstein distance W between the training examples x that it is trying to replicate and the samples $g_\theta(z)$ that it generates from random noise z with prior distribution p_z :

$$\mathcal{L}_{\text{generative}} = \inf_{\gamma \in \Pi(p_{\text{data}}, p_z)} \mathbb{E}_{(x,z) \sim \gamma} [\|x - g_\theta(z)\|]. \quad (\text{B.1})$$

In this case, the Wasserstein distance depends explicitly on the network weights θ via $g_\theta(z)$.

However, this is not the case for an AE with the loss function defined in equation (21), since neither the positive nor the negative examples pass through the network explicitly. The dependency is in the generation of the negative examples, which is done via the MCMC described in section 2.3. Equation (14) shows that the dependency on the network weights is in the gradient of the energy function with respect to the input features. As such, in order to use the Wasserstein distance as a loss function, the dependency of each step of the MCMC on the weights of the network must be retained:

$$\nabla_\theta x_{i+1, \theta} = \nabla_\theta x_{i, \theta} - \lambda \nabla_\theta \nabla_x E_\theta(x_{i, \theta}). \quad (\text{B.2})$$

The differentiable implementation of the Wasserstein distance metric is provided in the POT package [31]. In the PYTORCH package [77], the dependency retention in equation (B.2) is accomplished by explicitly enabling the `create_graph` flag when calling the `autograd` method on the AE energy in each MCMC step. This ensures that the computational graph, which PYTORCH uses to perform backpropagation, includes the dependency of the negative examples on the network weights throughout the MCMC.

Appendix C. Learning rate scheduler

The `torch.optim.lr_scheduler.ReduceLROnPlateau` method [77] is used to reduce the learning rate by a factor of 0.8 when the Wasserstein distance does not decrease for 20 epochs. The values for all hyperparameters of the learning rate scheduler can be found in table C.1. The evolution of the learning rate during the training of the WNAE is shown in figure C.1.

Table C.1. The hyperparameters of the learning rate scheduler.

Hyperparameter	Value
Initial learning rate	2×10^{-4}
Mode	'min'
Factor	0.8
Patience	20
Cooldown	10
Threshold	0.001
Threshold mode	'rel'
Minimum learning rate	0

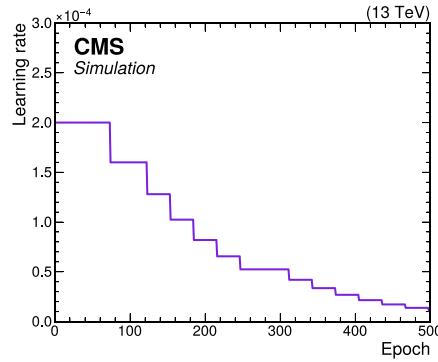


Figure C.1. The learning rate during the training of the WNAE from section 4.3.

Appendix D. Tuning the MCMC

Markov chains initialized with persistent contrastive divergence (PCD) [78] are found to accurately sample from the p_θ distribution. Therefore, PCD is used to initialize the Markov chain for each batch. In PCD, the Markov chain in batch k starts not from a new random set of points, but rather from the final set of points obtained from the Markov chain in batch $k - 1$. This procedure can be tuned by adding a restart fraction r , meaning that at every batch, a fraction r of the points sampled for the previous batch are discarded and replaced by new, random points, while the remaining points from the previous batch are reused [41]. The restart fraction r is a hyperparameter of the MCMC algorithm and is set to 5%.

The tuning of the Markov chain hyperparameters—the coefficients of equation (11), the step size λ and the noise σ —is performed as follows. The number of MCMC steps N_{MCMC} is fixed to 10 as a compromise between training time and sampling accuracy. The MCMC hyperparameters λ and σ are chosen to satisfy equation (13), leaving one degree of freedom, and the temperature T was constrained to the interval $(0, 1]$. These parameters are optimized to maximize the AUC for discrimination between SVJs and top quark jets, using equation (20) as the loss function, with two constraints to mitigate bias toward the signal hypothesis. First, at the maximum AUC, the input feature distributions of the negative and positive samples should match, as discussed in section 3.4. Second, the Wasserstein distance between the positive and negative samples should be minimal, as discussed in section 4.1. If the MCMC parameters are correctly set and the network is sufficiently trained, then the network learns $p_\theta = p_{\text{data}}$, so the AUC is greater than 0.5, assuming the distribution of the signal is not also equal to p_{data} , in which case no discrimination would be possible. In this case, the distribution of the negative samples is representative of the p_{data} distribution. The values of the Markov chain hyperparameters obtained from this optimization are provided in table D.1.

Trainings with a considerably larger gradient coefficient λ compared to the optimum were found to reach larger AUCs; however, these failed to achieve a good match between the distributions of negative and positive samples. Such MCMC configurations, though appealing if only the AUC is considered, result in an inadequate NAE that has not learned $p_\theta = p_{\text{data}}$: the AUC is large only because of the specific location of the signal hypotheses. This choice would therefore bias the performance of the NAE towards the signal hypotheses used for tuning, spoiling the unsupervised nature of the approach.

Table D.1. The hyperparameters of the MCMC.

Hyperparameter	Value
Number of steps N_{MCMC}	10
Restart fraction r	0.05
Temperature T	0.063
Noise σ	0.22
Step size λ	0.0242

The CMS Collaboration

A Hayrapetyan, **V Makarenko**, **A Tumasyan**¹

Yerevan Physics Institute, Yerevan, Armenia

W Adam, **J W Andrejkovic**, **L Benato**, **T Bergauer**, **M Dragicevic**, **C Giordano**, **P S Hussain**, **M Jeitler**², **N Krammer**, **A Li**, **D Liko**, **M Matthewman**, **I Mikulec**, **J Schieck**², **R Schöffbeck**², **D Schwarz**, **M Shooshtari**, **M Sonawane**, **W Waltenberger**, **C-E Wulz**²

Institut für Hochenergiephysik, Vienna, Austria

T Janssen, **H Kwon**, **D Ocampo Henao**, **T Van Laer**, **P Van Mechelen**

Universiteit Antwerpen, Antwerpen, Belgium

J Bierkens, **N Breugelmans**, **J D'Hondt**, **S Dansana**, **A De Moor**, **M Delcourt**, **F Heyen**, **Y Hong**, **P Kashko**, **S Lowette**, **I Makarenko**, **D Müller**, **J Song**, **S Tavernier**, **M Tytgat**³, **G P Van Onsem**, **S Van Putte**, **D Vannerom**

Vrije Universiteit Brussel, Brussel, Belgium

B Bilin, **B Clerbaux**, **A K Das**, **I De Bruyn**, **G De Lentdecker**, **H Evard**, **L Favart**, **P Gianneios**, **A Khalilzadeh**, **F A Khan**, **A Malara**, **M A Shahzad**, **L Thomas**, **M Vanden Bemden**, **C Vander Velde**, **P Vanlaer**, **F Zhang**

Université Libre de Bruxelles, Bruxelles, Belgium

M De Coen, **D Dobur**, **G Gokbulut**, **J Knolle**, **D Marckx**, **K Skovpen**, **N Van Den Bossche**, **J van der Linden**, **J Vandenbroeck**, **L Wezenbeek**

Ghent University, Ghent, Belgium

S Bein, **A Benecke**, **A Bethani**, **G Bruno**, **A Cappati**, **J De Favereau De Jeneret**, **C Delaere**, **A Giammanco**, **A O Guzel**, **V Lemaitre**, **J Lidrych**, **P Malek**, **P Mastrapasqua**, **S Turcpar**

Université Catholique de Louvain, Louvain-la-Neuve, Belgium

G A Alves, **M Barroso Ferreira Filho**, **E Coelho**, **C Hensel**, **T Menezes De Oliveira**, **C Mora Herrera**⁴, **P Rebello Teles**, **M Soeiro**, **E J Tonelli Manganote**⁵, **A Vilela Pereira**⁴

Centro Brasileiro de Pesquisas Fisicas, Rio de Janeiro, Brazil

W L Aldá Júnior, **H Brandao Malbouisson**, **W Carvalho**, **J Chinellato**⁶, **M Costa Reis**, **E M Da Costa**, **G G Da Silva**⁷, **D De Jesus Damiao**, **S Fonseca De Souza**, **R Gomes De Souza**, **S S Jesus**, **T Laux Kuhn**⁷, **M Macedo**, **K Mota Amarilo**, **L Mundim**, **H Nogima**, **J P Pinheiro**, **A Santoro**, **A Sznajder**, **M Thiel**, **F Torres Da Silva De Araujo**⁸

Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil

C A Bernardes⁷, **F Damas**, **T R Fernandez Perez Tomei**, **E M Gregores**, **B Lopes Da Costa**, **I Maietto Silverio**, **P G Mercadante**, **S F Novaes**, **B Orzari**, **Sandra S Padula**, **V Scheurer**

Universidade Estadual Paulista, Universidade Federal do ABC, São Paulo, Brazil

A Aleksandrov, **G Antchev**, **P Danev**, **R Hadjiiska**, **P Iaydjiev**, **M Shopova**, **G Sultanov**

Institute for Nuclear Research and Nuclear Energy, Bulgarian Academy of Sciences, Sofia, Bulgaria

A Dimitrov, **L Litov**, **B Pavlov**, **P Petkov**, **A Petrov**

University of Sofia, Sofia, Bulgaria

S Keshri, **D Laroze**, **S Thakur**

Instituto De Alta Investigación, Universidad de Tarapacá, Casilla 7 D, Arica, Chile

W Brooks

Universidad Tecnica Federico Santa Maria, Valparaiso, Chile

T Cheng, **T Javaid**, **L Wang**, **L Yuan**

Beihang University, Beijing, People's Republic of China

Z Hu, **Z Liang**, **J Liu**, **X Wang**

Department of Physics, Tsinghua University, Beijing, People's Republic of China

G M Chen⁹ , **H S Chen⁹** , **M Chen⁹** , **Y Chen⁹** , **Q Hou⁹** , **X Hou⁹**, **F Iemmi⁹** , **C H Jiang⁹**, **A Kapoor¹⁰** , **H Liao⁹** , **G Liu⁹** , **Z-A Liu¹¹** , **J N Song¹¹**, **S Song⁹**, **J Tao⁹** , **C Wang⁹**, **J Wang⁹** , **H Zhang⁹** , **J Zhao⁹** 

Institute of High Energy Physics, Beijing, People's Republic of China

A Agapitos⁹ , **Y Ban⁹** , **A Carvalho Antunes De Oliveira⁹** , **S Deng⁹** , **B Guo⁹**, **Q Guo⁹**, **C Jiang⁹** , **A Levin⁹** , **C Li⁹** , **Q Li⁹** , **Y Mao⁹**, **S Qian⁹**, **S J Qian⁹** , **X Qin⁹**, **X Sun⁹** , **D Wang⁹** , **J Wang⁹**, **H Yang⁹**, **M Zhang⁹**, **Y Zhao⁹**, **C Zhou⁹** 

State Key Laboratory of Nuclear Physics and Technology, Peking University, Beijing, People's Republic of China

S Yang⁹ 

State Key Laboratory of Nuclear Physics and Technology, Institute of Quantum Matter, South China Normal University, Guangzhou, People's Republic of China

Z You⁹ 

Sun Yat-Sen University, Guangzhou, People's Republic of China

K Jaffel⁹ , **N Lu⁹** 

University of Science and Technology of China, Hefei, People's Republic of China

G Bauer^{12,13}, **B Li¹⁴**, **H Wang⁹** , **K Yi¹⁵** , **J Zhang⁹** 

Nanjing Normal University, Nanjing, People's Republic of China

Y Li⁹

Institute of Modern Physics and Key Laboratory of Nuclear Physics and Ion-beam Application (MOE)—Fudan University, Shanghai, People's Republic of China

Z Lin⁹ , **C Lu⁹** , **M Xiao¹⁶** 

Zhejiang University, Hangzhou, Zhejiang, People's Republic of China

C Avila⁹ , **D A Barbosa Trujillo⁹** , **A Cabrera⁹** , **C Florez⁹** , **J Fraga⁹** , **J A Reyes Vega⁹**

Universidad de Los Andes, Bogota, Colombia

C Rendón⁹ , **M Rodriguez⁹** , **A A Ruales Barbosa⁹** , **J D Ruiz Alvarez⁹** 

Universidad de Antioquia, Medellin, Colombia

N Godinovic⁹ , **D Lelas⁹** , **A Sculac⁹** 

University of Split, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, Split, Croatia

M Kovac⁹ , **A Petkovic⁹** , **T Sculac⁹** 

University of Split, Faculty of Science, Split, Croatia

P Bargassa⁹ , **V Brigljevic⁹** , **B K Chitroda⁹** , **D Ferencek⁹** , **K Jakovcic⁹**, **A Starodumov⁹** , **T Susa⁹** 

Institute Rudjer Boskovic, Zagreb, Croatia

A Attikis⁹ , **K Christoforou⁹** , **A Hadjiagapiou⁹**, **C Leonidou⁹** , **C Nicolaou⁹**, **L Paizanos⁹** , **F Ptochos⁹** , **P A Razis⁹** , **H Rykaczewski⁹**, **H Saka⁹** , **A Steppenov⁹** 

University of Cyprus, Nicosia, Cyprus

M Finger[†] , **M Finger Jr⁹** 

Charles University, Prague, Czechia

E Ayala⁹ 

Escuela Politecnica Nacional, Quito, Ecuador

E Carrera Jarrin⁹ 

Universidad San Francisco de Quito, Quito, Ecuador

S Khalil¹⁷ , **E Salama^{18,19}** 

Academy of Scientific Research and Technology of the Arab Republic of Egypt, Egyptian Network of High Energy Physics, Cairo, Egypt

M Abdullah Al-Mashad , **A Hussein, H Mohammed** 

Center for High Energy Physics (CHEP-FU), Fayoum University, El-Fayoum, Egypt

K Ehataht , **M Kadastik** , **T Lange** , **C Nielsen** , **J Pata** , **M Raidal** , **N Seeba** , **L Tani** 

National Institute of Chemical Physics and Biophysics, Tallinn, Estonia

A Milieva , **K Osterberg** , **M Voutilainen** 

Department of Physics, University of Helsinki, Helsinki, Finland

N Bin Norjoharuddeen , **E Brücken** , **F Garcia** , **P Inkaew** , **K T S Kallonen** , **R Kumar Verma** ,
T Lampén , **K Lassila-Perini** , **B Lehtela** , **S Lehti** , **T Lindén** , **N R Mancilla Xinto** , **M**
Myllymäki , **M M Rantanen** , **S Saariokari** , **N T Toikka** , **J Tuominiemi** 

Helsinki Institute of Physics, Helsinki, Finland

H Kirschenmann , **P Luukka** , **H Petrow** 

Lappeenranta-Lahti University of Technology, Lappeenranta, Finland

M Besancon , **F Couderc** , **M Dejardin** , **D Denegri** , **P Devoue** , **J L Faure** , **F Ferri** , **P Gaigne** , **S**
Ganjour , **P Gras** , **G Hamel de Monchenault** , **M Kumar** , **V Lohezic** , **J Malcles** , **F Orlandi** , **L**
Portales , **S Ronchi** , **M Ö Sahin** , **A Savoy-Navarro** ²⁰ , **P Simkina** , **M Titov** , **M Tornago** 

IRFU, CEA, Université Paris-Saclay, Gif-sur-Yvette, France

F Beaudette , **G Boldrini** , **P Busson** , **C Charlot** , **M Chiusi** , **T D Cuisset** , **O Davignon** , **A**
De Wit , **T Debnath** , **I T Ehle** , **B A Fontana Santos Alves** , **S Ghosh** , **A Gilbert** , **R Granier de**
Cassagnac , **L Kalipoliti** , **M Manoni** , **M Nguyen** , **S Obraztsov** , **C Ochando** , **R Salerno** , **J B**
Sauvan , **Y Sirois** , **G Sokmen** , **L Urda Gómez** , **A Zabi** , **A Zghiche** 

Laboratoire Leprince-Ringuet, CNRS/IN2P3, Ecole Polytechnique, Institut Polytechnique de Paris,
Palaiseau, France

J-L Agram ²¹ , **J Andrea** , **D Bloch** , **J-M Brom** , **E C Chabert** , **C Collard** , **G Coulon** , **S Falke** , **U**
Goerlach , **R Haeberle** , **A-C Le Bihan** , **M Meena** , **O Poncet** , **G Saha** , **P Vaucelle** 

Université de Strasbourg, CNRS, IPHC UMR 7178, Strasbourg, France

A Di Florio 

Centre de Calcul de l'Institut National de Physique Nucleaire et de Physique des Particules,
CNRS/IN2P3, Villeurbanne, France

D Amram , **S Beauceron** , **B Blancon** , **G Boudoul** , **N Chanon** , **D Contardo** , **P Depasse** , **H El**
Mamouni , **J Fay** , **S Gascon** , **M Gouzevitch** , **C Greenberg** , **G Grenier** , **B Ille** , **E Jourdain** , **I B**
Laktineh , **M Lethuillier** , **B Massoteau** , **L Mirabito** , **A Purohit** , **M Vander Donckt** , **J Xiao** 

Institut de Physique des 2 Infinis de Lyon (IP2I), Villeurbanne, France

G Adamov , **I Lomidze** , **Z Tsamalaidze** ²² 

Georgian Technical University, Tbilisi, Georgia

V Botta , **S Consuegra Rodríguez** , **L Feld** , **K Klein** , **M Lipinski** , **D Meuser** , **P Nattland** , **V**
Oppenländer , **A Pauls** , **D Pérez Adán** , **N Röwert** , **M Teroerde** 

RWTH Aachen University, I Physikalisches Institut, Aachen, Germany

C Daumann , **S Diekmann** , **A Dodonova** , **N Eich** , **D Eliseev** , **F Engelke** , **J Erdmann** , **M**
Erdmann , **B Fischer** , **T Hebbeker** , **K Hoepfner** , **F Ivone** , **A Jung** , **N Kumar** , **M Y Lee** ,
F Mausolf , **M Merschmeyer** , **A Meyer** , **F Nowotny** , **A Pozdnyakov** , **W Redjeb** , **H Reithler** , **U**
Sarkar , **V Sarkisovi** , **A Schmidt** , **C Seth** , **A Sharma** , **J L Spah** , **V Vaulin** , **S Zaleski** 

RWTH Aachen University, III. Physikalisches Institut A, Aachen, Germany

M R Beckers , **C Dziwok** , **G Flügge** , **N Hoeflich** , **T Kress** , **A Nowack** , **O Pooth** , **A Stahl** , **A**
Zotz 

RWTH Aachen University, III. Physikalisches Institut B, Aachen, Germany

H Aarup Petersen , **A Abel** , **M Aldaya Martin** , **J Alimena** , **S Amoroso** , **Y An** , **I Andreev** ,
J Bach , **S Baxter** , **M Bayatmakou** , **H Becerril Gonzalez** , **O Behnke** , **A Belvedere** , **F**
Blekman ²³ , **K Borrás** ²⁴ , **A Campbell** , **S Chatterjee** , **L X Coll Saravia** , **G Eckerlin** , **D Eckstein** 

E Gallo²³ , A Geiser , V Guglielmi , M Guthoff , A Hinzmann , L Jeppe , M Kasemann , C Kleinwort , R Kogler , M Komm , D Krücker , W Lange, D Leyva Pernia , K-Y Lin , K Lipka²⁵ , W Lohmann²⁶ , J Malvaso , R Mankel , I-A Melzer-Pellmann , M Mendizabal Morentin , A B Meyer , G Milella , K Moral Figueroa , A Mussgiller , L P Nair , J Niedziela , A Nürnberg , J Park , E Ranken , A Raspereza , D Rastorguev , L Rygaard, M Scham^{27,24} , S Schnake²⁴ , P Schütze , C Schwanenberger²³ , D Selivanova , K Sharko , M Shchedrolosiev , D Stafford , M Torkian, F Vazzoler , A Ventura Barroso , R Walsh , D Wang , Q Wang , K Wichmann, L Wiens²⁴ , C Wissing , Y Yang , S Zakharov, A Zimmermann Castro Santos 
Deutsches Elektronen-Synchrotron, Hamburg, Germany

A R Alves Andrade , M Antonello , S Bollweg, M Bonanomi , K El Morabit , Y Fischer , M Frahm, E Garutti , A Grohsjean , A A Guvenli , J Haller , D Hundhausen, G Kasieczka , P Keicher , R Klanner , W Korcari , T Kramer , C C Kuo, F Labe , J Lange , A Lobanov , L Moureaux , A Nigamova , K Nikolopoulos , A Paasch , K J Pena Rodriguez , N Prouvost, T Quadfasel , B Raciti , M Rieger , D Savoie , P Schleper , M Schröder , J Schwandt , M Sommerhalder , H Stadie , G Steinbrück , R Ward , B Wiederspan, M Wolf 
University of Hamburg, Hamburg, Germany

S Brommer , E Butz , Y M Chen , T Chwalek , A Dierlamm , G G Dincer , U Elicabuk, N Faltermann , M Giffels , A Gottmann , F Hartmann²⁸ , R Hofsaess , M Horzela , F Hummer , U Husemann , J Kieseler , M Klute , R Kunnilan Muhammed Rafeek, O Lavoryk , J M Lawhorn , A Lintuluoto , S Maier , M Mormile , Th Müller , E Pfeffer , M Presilla , G Quast , K Rabbertz , B Regnery , R Schmieder, N Shadskiy , I Shvetsov , H J Simonis , L Sowa , L Stockmeier, K Tauqeer, M Toms , B Topko , N Trevisani , C Verstege , T Voigtländer , R F Von Cube , J Von Den Driesch, M Wassmer , R Wolf , W D Zeuner , X Zuo 
Karlsruher Institut fuer Technologie, Karlsruhe, Germany

G Anagnostou , G Daskalakis , A Kyriakis 
Institute of Nuclear and Particle Physics (INPP), NCSR Demokritos, Aghia Paraskevi, Greece

G Melachroinos, Z Painesis , I Paraskevas , N Saoulidou , K Theofilatos , E Tziaferi , E Tzovara , K Vellidis , I Zisopoulos 
National and Kapodistrian University of Athens, Athens, Greece

T Chatzistavrou , G Karapostoli , K Kousouris , E Siamarkou, G Tsipolitis 
National Technical University of Athens, Athens, Greece

I Bestintzanos, I Evangelou , C Foudas, P Katsoulis, P Kokkas , P G Kosmoglou Kioseoglou , N Manthos , I Papadopoulos , J Strologas 
University of Ioánnina, Ioánnina, Greece

D Druzhkin , C Hajdu , D Horvath^{29,30} , K Márton, A J Rádl³¹ , F Sikler , V Veszpremi 
HUN-REN Wigner Research Centre for Physics, Budapest, Hungary

M Csanád , K Farkas , A Fehérkúti³² , M M A Gadallah³³ , Á Kadlecsek , M León Coello , G Pásztor , G I Veres 
MTA-ELTE Lendület CMS Particle and Nuclear Physics Group, Eötvös Loránd University, Budapest, Hungary

B Ujvari , G Zilizi 
Faculty of Informatics, University of Debrecen, Debrecen, Hungary

G Bencze, S Czellar, J Molnar, Z Szillasi
HUN-REN ATOMKI—Institute of Nuclear Research, Debrecen, Hungary

T Csorgo³² , F Nemes³² , T Novak , I Szanyi³⁴ 
Karoly Robert Campus, MATE Institute of Technology, Gyongyos, Hungary

S Bansal , S B Beri, V Bhatnagar , G Chaudhary , S Chauhan , N Dhingra³⁵ , A Kaur , A Kaur , H Kaur , M Kaur , S Kumar , T Sheokand, J B Singh , A Singla 
Panjab University, Chandigarh, India

A Bhardwaj, **A Chhetri**, **B C Choudhary**, **A Kumar**, **A Kumar**, **M Naimuddin**, **S Phor**, **K Ranjan**, **M K Saini**

University of Delhi, Delhi, India

S Acharya³⁶, **B Gomber**³⁶, **B Sahu**³⁶

University of Hyderabad, Hyderabad, India

S Mukherjee

Indian Institute of Technology Kanpur, Kanpur, India

S Baradia, **S Bhattacharya**, **S Das Gupta**, **S Dutta**, **S Dutta**, **S Sarkar**

Saha Institute of Nuclear Physics, HBNI, Kolkata, India

M M Ameen, **P K Behera**, **S Chatterjee**, **G Dash**, **A Dattamunsi**, **P Jana**, **P Kalbhor**, **S Kamble**, **J R Komaragiri**³⁷, **T Mishra**, **P R Pujahari**, **A K Sikdar**, **R K Singh**, **P Verma**, **S Verma**, **A Vijay**

Indian Institute of Technology Madras, Madras, India

B K Sirasva

IISER Mohali, India, Mohali, India

L Bhatt, **S Dugad**, **G B Mohanty**, **M Shelake**, **P Suryadevara**

Tata Institute of Fundamental Research-A, Mumbai, India

A Bala, **S Banerjee**, **S Barman**³⁸, **R M Chatterjee**, **M Guchait**, **Sh Jain**, **A Jaiswal**, **B M Joshi**, **S Kumar**, **M Maity**³⁸, **G Majumder**, **K Mazumdar**, **S Parolia**, **R Saxena**, **A Thachayath**

Tata Institute of Fundamental Research-B, Mumbai, India

S Bahinipati³⁹, **D Maity**⁴⁰, **P Mal**, **K Naskar**⁴⁰, **A Nayak**⁴⁰, **S Nayak**, **K Pal**, **R Raturi**, **P Sadangi**, **S K Swain**, **S Varghese**⁴⁰, **D Vats**⁴⁰

National Institute of Science Education and Research, An OCC of Homi Bhabha National Institute, Bhubaneswar, Odisha, India

A Alpina, **S Dube**, **P Hazarika**, **B Kansal**, **A Laha**, **R Sharma**, **S Sharma**, **K Y Vaish**

Indian Institute of Science Education and Research (IISER), Pune, India

S Ghosh

Indian Institute of Technology Hyderabad, Telangana, India

H Bakhshiansohi⁴¹, **A Jafari**⁴², **V Sedighzadeh Dalavi**, **M Zeinali**⁴³

Isfahan University of Technology, Isfahan, Iran

S Bashiri, **S Chenarani**⁴⁴, **S M Etesami**, **Y Hosseini**, **M Khakzad**, **E Khazaie**, **M Mohammadi Najafabadi**, **S Tizchang**⁴⁵

Institute for Research in Fundamental Sciences (IPM), Tehran, Iran

M Felcini, **M Grunewald**

University College Dublin, Dublin, Ireland

M Abbrescia^{a,b}, **M Barbieri**^{a,b}, **M Buonsante**^{a,b}, **A Colaleo**^{a,b}, **D Creanza**^{a,c}, **N De Filippis**^{a,c}, **M De Palma**^{a,b}, **W Elmetenawee**^{a,b,46}, **N Ferrara**^{a,c}, **L Fiore**^a, **L Longo**^a, **M Louka**^{a,b}, **G Maggi**^{a,c}, **M Maggi**^a, **I Margjeka**^a, **V Mastrapasqua**^{a,b}, **S My**^{a,b}, **F Nenna**^{a,b}, **S Nuzzo**^{a,b}, **A Pellecchia**^{a,b}, **A Pompili**^{a,b}, **G Pugliese**^{a,c}, **R Radogna**^{a,b}, **D Ramos**^a, **A Ranieri**^a, **L Silvestris**^a, **F M Simone**^{a,c}, **Ü Sözbilir**^a, **A Stamerra**^{a,b}, **D Troiano**^{a,b}, **R Venditti**^{a,b}, **P Verwilligen**^a, **A Zaza**^{a,b}

INFN Sezione di Bari^a, Università di Bari^b, Politecnico di Bari^c, Bari, Italy

G Abbiendi^a, **D Bonacorsi**^{a,b}, **P Capiluppi**^{a,b}, **F R Cavallo**^a, **M Cuffiani**^{a,b}, **G M Dallavalle**^a, **T Diotallevi**^{a,b}, **F Fabbri**^a, **A Fanfani**^{a,b}, **D Fasanella**^a, **P Giacomelli**^a, **L Gionmmi**^{a,b}, **L Guiducci**^{a,b}, **S Lo Meo**^{a,47}, **M Lorusso**^{a,b}, **L Lunerti**^a, **S Marcellini**^a, **G Masetti**^a, **F L Navarria**^{a,b}, **G Paggi**^{a,b}, **A Perrotta**^a, **F Primavera**^{a,b}, **A M Rossi**^{a,b}, **S Rossi Tisbeni**^{a,b}, **T Rovelli**^{a,b}, **G P Siroli**^{a,b}

INFN Sezione di Bologna^a, Università di Bologna^b, Bologna, Italy

S Costa^{a,b,48} , A Di Mattia^a , A Lapertosa^a , R Potenza^{a,b} , A Tricomi^{a,b,48} 

INFN Sezione di Catania^a, Università di Catania^b, Catania, Italy

J Altork^{a,b} , P Assiouras^a , G Barbagli^a , G Bardelli^a , M Bartolini^{a,b} , A Calandri^{a,b} , B Camaiani^{a,b} , A Cassese^a , R Ceccarelli^a , V Ciulli^{a,b} , C Civinini^a , R D'Alessandro^{a,b} , L Damenti^{a,b} , E Focardi^{a,b} , T Kello^a , G Latino^{a,b} , P Lenzi^{a,b} , M Lizzo^a , M Meschini^a , S Paoletti^a , A Papanastassiou^{a,b} , G Sguazzoni^a , L Viliani^a

INFN Sezione di Firenze^a, Università di Firenze^b, Firenze, Italy

L Benussi , S Bianco , S Meola⁴⁹ , D Piccolo 

INFN Laboratori Nazionali di Frascati, Frascati, Italy

M Alves Gallo Pereira^a , F Ferro^a , E Robutti^a , S Tosi^{a,b} 

INFN Sezione di Genova^a, Università di Genova^b, Genova, Italy

A Benaglia^a , F Brivio^a , V Camagni^{a,b} , F Cetorelli^{a,b} , F De Guio^{a,b} , M E Dinardo^{a,b} , P Dini^a , S Gennai^a , R Gerosa^{a,b} , A Ghezzi^{a,b} , P Govoni^{a,b} , L Guzzi^a , M R Kim^a , G Lavizzari^{a,b} , M T Lucchini^{a,b} , M Malberti^a , S Malvezzi^a , A Massironi^a , D Menasce^a , L Moroni^a , M Paganoni^{a,b} , S Palluotto^{a,b} , D Pedrini^a , A Perego^{a,b} , G Pizzati^{a,b} , S Ragazzi^{a,b} , T Tabarelli de Fatis^{a,b}

INFN Sezione di Milano-Bicocca^a, Università di Milano-Bicocca^b, Milano, Italy

S Buontempo^a , C Di Fraia^{a,b} , F Fabozzi^{a,c} , L Favilla^{a,d} , A O M Iorio^{a,b} , L Lista^{a,b,50} , P Paolucci^{a,28} , B Rossi^a

INFN Sezione di Napoli^a, Università di Napoli 'Federico II'^b, Napoli, Italy; Università della Basilicata^c, Potenza, Italy; Scuola Superiore Meridionale (SSM)^d, Napoli, Italy

P Azzi^a , N Bacchetta^{a,51} , D Bisello^{a,b} , P Bortignon^{a,c} , G Bortolato^{a,b} , A C M Bulla^{a,c} , R Carlin^{a,b} , P Checchia^a , T Dorigo^{a,52} , U Gasparini^{a,b} , S Giorgetti^a , E Lusiani^a , M Margoni^{a,b} , A T Meneguzzo^{a,b} , M Michelotto^a , J Pazzini^{a,b} , P Ronchese^{a,b} , R Rossin^{a,b} , F Simonetto^{a,b} , M Tosi^{a,b} , A Triossi^{a,b} , M Zanetti^{a,b} , P Zotto^{a,b} , A Zucchetta^{a,b} , G Zumerle^{a,b}

INFN Sezione di Padova^a, Università di Padova^b, Padova, Italy; Università degli Studi di Cagliari^c, Cagliari, Italy

A Braghieri^a , S Calzaferri^a , P Montagna^{a,b} , M Pelliccioni^a , V Re^a , C Riccardi^{a,b} , P Salvini^a , I Vai^{a,b} , P Vitulo^{a,b}

INFN Sezione di Pavia^a, Università di Pavia^b, Pavia, Italy

S Ajmal^{a,b} , M E Ascioti^{a,b} , G M Bilei^a , C Carrivale^{a,b} , D Ciangottini^{a,b} , L Della Penna^{a,b} , L Fanò^{a,b} , V Mariani^{a,b} , M Menichelli^a , F Moscatelli^{a,53} , A Rossi^{a,b} , A Santocchia^{a,b} , D Spiga^a , T Tedeschi^{a,b}

INFN Sezione di Perugia^a, Università di Perugia^b, Perugia, Italy

C Aimè^{a,b} , C A Alexe^{a,c} , P Asenov^{a,b} , P Azzurri^a , G Bagliesi^a , R Bhattacharya^a , L Bianchini^{a,b} , T Boccali^a , E Bossini^a , D Bruschini^{a,c} , L Calligaris^{a,b} , R Castaldi^a , F Cattafesta^{a,c} , M A Ciocci^{a,d} , M Cipriani^{a,b} , R Dell'Orso^a , S Donato^{a,b} , R Forti^{a,b} , A Giassi^a , F Ligabue^{a,c} , A C Marini^{a,b} , D Matos Figueiredo^a , A Messineo^{a,b} , S Mishra^a , V K Muraleedharan Nair Bindhu^{a,b} , S Nandan^a , F Palla^a , M Riggirello^{a,c} , A Rizzi^{a,b} , G Rolandi^{a,c} , S Roy Chowdhury^{a,54} , T Sarkar^a , A Scribano^a , P Solanki^{a,b} , P Spagnolo^a , F Tenchini^{a,b} , R Tenchini^a , G Tonelli^{a,b} , N Turini^{a,d} , F Vaselli^{a,c} , A Venturi^a , P G Verdini^a

INFN Sezione di Pisa^a, Università di Pisa^b, Scuola Normale Superiore di Pisa^c, Pisa, Italy; Università di Siena^d, Siena, Italy

P Akrap^{a,b} , C Basile^{a,b} , S C Behera^a , F Cavallari^a , L Cunqueiro Mendez^{a,b} , F De Ruggi^{a,b} , D Del Re^{a,b} , E Di Marco^a , M Diemoz^a , F Errico^a , L Frosina^{a,b} , R Gargiulo^{a,b} , B Harikrishnan^{a,b} , F Lombardi^{a,b} , E Longo^{a,b} , L Martikainen^{a,b} , J Mijuskovic^{a,b} , G Organtini^{a,b} , N Palmeri^{a,b} , R Paramatti^{a,b} , C Quaranta^{a,b} , S Rahatlou^{a,b} , C Rovelli^a , F Santanastasio^{a,b} , L Soffi^a , V Vladimirov^{a,b}

INFN Sezione di Roma^a, Sapienza Università di Roma^b, Roma, Italy

N Amapane^{a,b} , **R Arcidiacono^{a,c}** , **S Argiro^{a,b}** , **M Arneodo^{a,c}** , **N Bartosik^{a,c}** , **R Bellan^{a,b}** , **A Bellora^{a,b}** , **C Biino^a** , **C Borca^{a,b}** , **N Cartiglia^a** , **M Costa^{a,b}** , **R Covarelli^{a,b}** , **N Demaria^a** , **L Finco^a** , **M Grippo^{a,b}** , **B Kiani^{a,b}** , **L Lanteri^{a,b}** , **F Legger^a** , **F Luongo^{a,b}** , **C Mariotti^a** , **S Maselli^a** , **A Mecca^{a,b}** , **L Menzio^{a,b}** , **P Meridiani^a** , **E Migliore^{a,b}** , **M Monteno^a** , **M M Obertino^{a,b}** , **G Ortona^a** , **L Pacher^{a,b}** , **N Pastrone^a** , **M Ruspa^{a,c}** , **F Siviero^{a,b}** , **V Sola^{a,b}** , **A Solano^{a,b}** , **A Staiano^a** , **C Tarricone^{a,b}** , **D Trocino^a** , **G Umoret^{a,b}** , **E Vlasov^{a,b}** , **R White^{a,b}** 
INFN Sezione di Torino^a, Università di Torino^b, Torino, Italy; Università del Piemonte Orientale^c, Novara, Italy

J Babbar^{a,b} , **S Belforte^a** , **V Candelise^{a,b}** , **M Casarsa^a** , **F Cossutti^a** , **K De Leo^a** , **G Della Ricca^{a,b}** , **R Delli Gatti^{a,b}** 
INFN Sezione di Trieste^a, Università di Trieste^b, Trieste, Italy

S Dogra , **J Hong** , **J Kim** , **T Kim** , **D Lee** , **H Lee** , **J Lee** , **S W Lee** , **C S Moon** , **Y D Oh** , **S Sekmen** , **B Tae** , **Y C Yang** 
Kyungpook National University, Daegu, Republic of Korea

M S Kim 
Department of Mathematics and Physics—GWNu, Gangneung, Republic of Korea

G Bak , **P Gwak** , **H Kim** , **D H Moon** , **J Seo** 
Chonnam National University, Institute for Universe and Elementary Particles, Kwangju, Republic of Korea

E Asilar , **F Carnevali** , **J Choi⁵⁵** , **T J Kim** , **Y Ryou** 
Hanyang University, Seoul, Republic of Korea

S Ha , **S Han** , **B Hong** , **J Kim** , **K Lee** , **K S Lee** , **S Lee** , **J Yoo** 
Korea University, Seoul, Republic of Korea

J Goh , **J Shin** , **S Yang** 
Kyung Hee University, Department of Physics, Seoul, Republic of Korea

Y Kang , **H S Kim** , **Y Kim** , **S Lee**
Sejong University, Seoul, Republic of Korea

J Almond , **J H Bhyun** , **J Choi** , **J Choi** , **W Jun** , **H Kim** , **J Kim** , **T Kim** , **Y Kim** , **Y W Kim** , **S Ko** , **H Lee** , **J Lee** , **J Lee** , **B H Oh** , **S B Oh** , **J Shin** , **U K Yang** , **I Yoon** 
Seoul National University, Seoul, Republic of Korea

W Jang , **D Y Kang** , **D Kim** , **S Kim** , **B Ko** , **J S H Lee** , **Y Lee** , **I C Park** , **Y Roh** , **I J Watson** 
University of Seoul, Seoul, Republic of Korea

G Cho , **K Hwang** , **B Kim** , **S Kim** , **K Lee** , **H D Yoo** 
Yonsei University, Department of Physics, Seoul, Republic of Korea

M Choi , **Y Lee** , **I Yu** 
Sungkyunkwan University, Suwon, Republic of Korea

T Beyrouthy , **Y Gharbia** 
College of Engineering and Technology, American University of the Middle East (AUM), Dasman, Kuwait

F Alazemi 
Kuwait University—College of Science—Department of Physics, Safat, Kuwait

K Dreimanis , **O M Eberlins** , **A Gaile** , **C Munoz Diaz** , **D Osite** , **G Pikurs** , **R Plese** , **A Potrebko** , **M Seidel** , **D Sidiropoulos Kontos** 
Riga Technical University, Riga, Latvia

N R Strautnieks 
University of Latvia (LU), Riga, Latvia

M Ambroz^{[1](#)}, **A Juodagalvis**^{[1](#)}, **S Nargelas**^{[1](#)}, **A Rinkevicius**^{[1](#)}, **G Tamulaitis**^{[1](#)}
Vilnius University, Vilnius, Lithuania

I Yusuff^{[56](#)}, **Z Zolkapli**
National Centre for Particle Physics, Universiti Malaya, Kuala Lumpur, Malaysia

J F Benitez^{[1](#)}, **A Castaneda Hernandez**^{[1](#)}, **A Cota Rodriguez**^{[1](#)}, **L E Cuevas Picos**, **H A Encinas Acosta**, **L G Gallegos Maríñez**, **J A Murillo Quijada**^{[1](#)}, **L Valencia Palomo**^{[1](#)}
Universidad de Sonora (UNISON), Hermosillo, Mexico

G Ayala^{[1](#)}, **H Castilla-Valdez**^{[1](#)}, **H Crotte Ledesma**^{[1](#)}, **R Lopez-Fernandez**^{[1](#)}, **J Mejia Guisao**^{[1](#)}, **R Reyes-Almanza**^{[1](#)}, **A Sánchez Hernández**^{[1](#)}
Centro de Investigacion y de Estudios Avanzados del IPN, Mexico City, Mexico

C Oropeza Barrera^{[1](#)}, **D L Ramirez Guadarrama**, **M Ramírez García**^{[1](#)}
Universidad Iberoamericana, Mexico City, Mexico

I Bautista^{[1](#)}, **F E Neri Huerta**^{[1](#)}, **I Pedraza**^{[1](#)}, **H A Salazar Ibarguen**^{[1](#)}, **C Uribe Estrada**^{[1](#)}
Benemerita Universidad Autonoma de Puebla, Puebla, Mexico

I Bubanja^{[1](#)}, **N Raicevic**^{[1](#)}
University of Montenegro, Podgorica, Montenegro

P H Butler^{[1](#)}
University of Canterbury, Christchurch, New Zealand

A Ahmad^{[1](#)}, **M I Asghar**^{[1](#)}, **A Awais**^{[1](#)}, **M I M Awan**, **W A Khan**^{[1](#)}
National Centre for Physics, Quaid-I-Azam University, Islamabad, Pakistan

V Avati, **L Forthomme**^{[1](#)}, **L Grzanka**^{[1](#)}, **M Malawski**^{[1](#)}, **K Piotrkowski**^{[1](#)}
AGH University of Krakow, Krakow, Poland

M Bluj^{[1](#)}, **M Górski**^{[1](#)}, **M Kazana**^{[1](#)}, **M Szeleper**^{[1](#)}, **P Zalewski**^{[1](#)}
National Centre for Nuclear Research, Swierk, Poland

K Bunkowski^{[1](#)}, **K Doroba**^{[1](#)}, **A Kalinowski**^{[1](#)}, **M Konecki**^{[1](#)}, **J Krolikowski**^{[1](#)}, **A Muhammad**^{[1](#)}
Institute of Experimental Physics, Faculty of Physics, University of Warsaw, Warsaw, Poland

P Fokow^{[1](#)}, **K Pozniak**^{[1](#)}, **W Zabolotny**^{[1](#)}
Warsaw University of Technology, Warsaw, Poland

M Araujo^{[1](#)}, **D Bastos**^{[1](#)}, **C Beirão Da Cruz E Silva**^{[1](#)}, **A Boletti**^{[1](#)}, **M Bozzo**^{[1](#)}, **T Camporesi**^{[1](#)}, **G Da Molin**^{[1](#)}, **M Gallinaro**^{[1](#)}, **J Hollar**^{[1](#)}, **N Leonardo**^{[1](#)}, **G B Marozzo**^{[1](#)}, **A Petrilli**^{[1](#)}, **M Pisano**^{[1](#)}, **J Seixas**^{[1](#)}, **J Varela**^{[1](#)}, **J W Wulff**^{[1](#)}
Laboratório de Instrumentação e Física Experimental de Partículas, Lisboa, Portugal

P Adzic^{[1](#)}, **L Markovic**^{[1](#)}, **P Milenovic**^{[1](#)}, **V Milosevic**^{[1](#)}
Faculty of Physics, University of Belgrade, Belgrade, Serbia

D Devetak^{[1](#)}, **M Dordevic**^{[1](#)}, **J Milosevic**^{[1](#)}, **L Nadderd**^{[1](#)}, **V Rekovic**, **M Stojanovic**^{[1](#)}
VINCA Institute of Nuclear Sciences, University of Belgrade, Belgrade, Serbia

M Alcalde Martinez^{[1](#)}, **J Alcaraz Maestre**^{[1](#)}, **Cristina F Bedoya**^{[1](#)}, **J A Brochero Cifuentes**^{[1](#)}, **Oliver M Carretero**^{[1](#)}, **M Cepeda**^{[1](#)}, **M Cerrada**^{[1](#)}, **N Colino**^{[1](#)}, **J Cuchillo Ortega**, **B De La Cruz**^{[1](#)}, **A Delgado Peris**^{[1](#)}, **A Escalante Del Valle**^{[1](#)}, **D Fernández Del Val**^{[1](#)}, **J P Fernández Ramos**^{[1](#)}, **J Flix**^{[1](#)}, **M C Fouz**^{[1](#)}, **M Gonzalez Hernandez**^{[1](#)}, **O Gonzalez Lopez**^{[1](#)}, **S Goy Lopez**^{[1](#)}, **J M Hernandez**^{[1](#)}, **M I Josa**^{[1](#)}, **J Llorente Merino**^{[1](#)}, **C Martin Perez**^{[1](#)}, **E Martin Viscasillas**^{[1](#)}, **D Moran**^{[1](#)}, **C M Morcillo Perez**^{[1](#)}, **R Paz Herrera**^{[1](#)}, **C Perez Dengra**^{[1](#)}, **A Pérez-Calero Yzquierdo**^{[1](#)}, **J Puerta Pelayo**^{[1](#)}, **I Redondo**^{[1](#)}, **J Vazquez Escobar**^{[1](#)}
Centro de Investigaciones Energéticas Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain

J F de Trocóniz^{[1](#)}
Universidad Autónoma de Madrid, Madrid, Spain

B Alvarez Gonzalez , **J Ayllon Torresano** , **A Cardini** , **J Cuevas** , **J Del Riego Badas** , **D Estrada Acevedo** , **J Fernandez Menendez** , **S Folgueras** , **I Gonzalez Caballero** , **P Leguina** , **M Obeso Menendez** , **E Palencia Cortezon** , **J Prado Pico** , **A Soto Rodríguez** , **C Vico Villalba** , **P Vischia** 
 Universidad de Oviedo, Instituto Universitario de Ciencias y Tecnologías Espaciales de Asturias (ICTEA), Oviedo, Spain

S Blanco Fernández , **I J Cabrillo** , **A Calderon** , **J Duarte Campderros** , **M Fernandez** , **G Gomez** , **C Lasasa García** , **R Lopez Ruiz** , **C Martinez Rivero** , **P Martinez Ruiz del Arbol** , **F Matorras** , **P Matorras Cuevas** , **E Navarrete Ramos** , **J Piedra Gomez** , **C Quintana San Emeterio** , **L Scodellaro** , **I Vila** , **R Vilar Cortabitarte** , **J M Vizan Garcia** 
 Instituto de Física de Cantabria (IFCA), CSIC-Universidad de Cantabria, Santander, Spain

B Kailasapathy⁵⁷ , **D D C Wickramarathna** 
 University of Colombo, Colombo, Sri Lanka

W G D Dharmaratna⁵⁸ , **K Liyanage** , **N Perera** 
 University of Ruhuna, Department of Physics, Matara, Sri Lanka

D Abbaneo , **C Amendola** , **R Ardino** , **E Auffray** , **J Baechler** , **D Barney** , **J Bendavid** , **M Bianco** , **A Bocci** , **L Borbonovi** , **C Botta** , **A Bragagnolo** , **C E Brown** , **C Caillol** , **G Cerminara** , **P Connor** , **D d'Enterria** , **A Dabrowski** , **A David** , **A De Roeck** , **M M Defranchis** , **M Deile** , **M Dobson** , **P J Fernández Manteca** , **W Funk** , **A Gaddi** , **S Giani** , **D Gigi** , **K Gill** , **F Glege** , **M Glowacki** , **A Gruber** , **J Hegeman** , **J K Heikkilä** , **B Huber** , **V Innocente** , **T James** , **P Janot** , **O Kaluzinska** , **O Karacheban**²⁶ , **G Karathanasis** , **S Laurila** , **P Lecoq** , **C Lourenço** , **A-M Lyon** , **M Magherini** , **L Malgeri** , **M Mannelli** , **A Mehta** , **F Meijers** , **J A Merlin** , **S Mersi** , **E Meschi** , **M Migliorini** , **F Monti** , **F Moortgat** , **M Mulders** , **M Musich** , **I Neutelings** , **S Orfanelli** , **F Pantaleo** , **M Pari** , **G Petrucciani** , **A Pfeiffer** , **M Pierini** , **M Pitt** , **H Qu** , **D Rabady** , **A Reimers** , **B Ribeiro Lopes** , **F Riti** , **P Rosado** , **M Rovere** , **H Sakulin** , **R Salvatico** , **S Sanchez Cruz** , **S Scarfi** , **M Selvaggi** , **A Sharma** , **K Shchelina** , **P Silva** , **P Sphicas**⁵⁹ , **A G Stahl Leiton** , **A Steen** , **S Summers** , **D Treille** , **P Tropea** , **E Vernazza** , **J Wanczyk**⁶⁰ , **J Wang** , **S Wuchterl** , **M Zarucki** , **P Zehetner** , **P Zejdl** , **G Zevi Della Porta** 
 CERN, European Organization for Nuclear Research, Geneva, Switzerland

T Bevilacqua⁶¹ , **L Caminada**⁶¹ , **W Erdmann** , **R Horisberger** , **Q Ingram** , **H C Kaestli** , **D Kotlinski** , **C Lange** , **U Langenegger** , **L Nohte**⁶¹ , **T Rohe** , **A Samalan** 
 PSI Center for Neutron and Muon Sciences, Villigen, Switzerland

T K Aarrestad , **M Backhaus** , **G Bonomelli** , **C Cazzaniga** , **K Datta** , **P De Bryas Dexmiers** , **D'archiacchiac**⁶⁰ , **A De Cosa** , **G Dissertori** , **M Dittmar** , **M Donegà** , **F Eble** , **K Gedia** , **F Glessgen** , **C Grab** , **N Härringer** , **T G Harte** , **W Lustermann** , **M Malucchi** , **R A Manzoni** , **M Marchegiani** , **L Marchese** , **A Mascellani**⁶⁰ , **F Nessi-Tedaldi** , **F Pauss** , **V Perovic** , **B Ristic** , **R Seidita** , **J Steggemann**⁶⁰ , **A Tarabini** , **D Valsecchi** , **R Wallny** 
 ETH Zurich—Institute for Particle Physics and Astrophysics (IPA), Zurich, Switzerland

C Amsler⁶² , **P Bärtschi** , **F Bilandzija** , **M F Canelli** , **G Celotto** , **K Cormier** , **M Huwiler** , **W Jin** , **A Jofrehei** , **B Kilminster** , **T H Kwok** , **S Leontsinis** , **V Lukashenko** , **A Macchiolo** , **F Meng** , **M Missiroli** , **J Motta** , **P Robmann** , **M Senger** , **E Shokr** , **F Stäger** , **R Tramontano** , **P Viscone** 
 Universität Zürich, Zurich, Switzerland

D Bhowmik , **C M Kuo** , **P K Rout** , **S Taj** , **P C Tiwari**³⁷ 
 National Central University, Chung-Li, Taiwan

L Ceard , **K F Chen** , **Z G Chen** , **A De Iorio** , **W-S Hou** , **T H Hsu** , **Y W Kao** , **S Karmakar** , **G Kole** , **Y Y Li** , **R-S Lu** , **E Paganis** , **X F Su** , **J Thomas-Wilsker** , **L S Tsai** , **D Tsionou** , **H Y Wu** , **E Yazgan** 
 National Taiwan University (NTU), Taipei, Taiwan

C Asawatangtrakuldee , **N Srimanobhas** 
 High Energy Physics Research Unit, Department of Physics, Faculty of Science, Chulalongkorn University, Bangkok, Thailand

Y Maghrbi 

Tunis El Manar University, Tunis, Tunisia

D Agyel , **F Dolek** , **I Dumanoglu**⁶³ , **Y Guler**⁶⁴ , **E Gurpinar Guler**⁶⁴ , **C Isik** , **O Kara** , **A Kayis Topaksu** , **Y Komurcu** , **G Onengut** , **K Ozdemir**⁶⁵ , **B Tali**⁶⁶ , **U G Tok** , **E Uslan** , **I S Zorbakir** 

Çukurova University, Physics Department, Science and Art Faculty, Adana, Türkiye

M Yalvac⁶⁷ 

Middle East Technical University, Physics Department, Ankara, Türkiye

B Akgun , **I O Atakisi**⁶⁸ , **E Gülmez** , **M Kaya**⁶⁹ , **O Kaya**⁷⁰ , **M A Sarkisla**⁷¹ , **S Tekten**⁷² 

Bogazici University, Istanbul, Türkiye

A Cakir , **K Cankocak**^{63,73} , **S Sen**⁷⁴ 

Istanbul Technical University, Istanbul, Türkiye

O Aydilek⁷⁵ , **B Haciasahinoglu** , **I Hos**⁷⁶ , **B Kaynak** , **S Ozkorucuklu** , **O Potok** , **H Sert** , **C Simsek** , **C Zorbilmez** 

Istanbul University, Istanbul, Türkiye

S Cerci , **B Isildak**⁷⁷ , **E Simsek** , **D Sunar Cerci** , **T Yetkin**⁷⁸ 

Yildiz Technical University, Istanbul, Türkiye

A Boyaryntsev , **O Dadazhanova**, **B Grynyov** 

Institute for Scintillation Materials of National Academy of Science of Ukraine, Kharkiv, Ukraine

L Levchuk 

National Science Centre, Kharkiv Institute of Physics and Technology, Kharkiv, Ukraine

J J Brooke , **A Bundock** , **F Bury** , **E Clement** , **D Cussans** , **D Dharmender**, **H Flacher** , **J Goldstein** , **H F Heath** , **M-L Holmberg** , **L Kreczko** , **S Paramesvaran** , **L Robertshaw**, **M S Sanjrani**⁴¹, **J Segal**, **V J Smith** 

University of Bristol, Bristol, United Kingdom

A H Ball, **K W Bell** , **A Belyaev**⁷⁹ , **C Brew** , **R M Brown** , **D J A Cockerill** , **A Elliot** , **K V Ellis**, **J Gajownik** , **K Harder** , **S Harper** , **J Linacre** , **K Manolopoulos**, **M Moallemi** , **D M Newbold** , **E Olaiya** , **D Petyt** , **T Reis** , **A R Sahasransu** , **G Salvi** , **T Schuh**, **C H Shepherd-Themistocleous** , **I R Tomalin** , **K C Whalen** , **T Williams** 

Rutherford Appleton Laboratory, Didcot, United Kingdom

I Andreou , **R Bainbridge** , **P Bloch** , **O Buchmuller**, **C A Carrillo Montoya** , **D Colling** , **J S Dancu**, **I Das** , **P Dauncey** , **G Davies** , **M Della Negra** , **S Fayer**, **G Fedi** , **G Hall** , **H R Hoorani** , **A Howard**, **G Iles** , **C R Knight** , **P Krueper** , **J Langford** , **K H Law** , **J León Holgado** , **E Leutgeb** , **L Lyons** , **A-M Magnan** , **B Maier** , **S Mallios**, **A Mastronikolis** , **M Mieskolainen** , **J Nash**⁸⁰ , **M Pesaresi** , **P B Pradeep** , **B C Radburn-Smith** , **A Richards**, **A Rose** , **L Russell** , **K Savva** , **C Seez** , **R Shukla** , **A Tapper** , **K Uchida** , **G P Uttley** , **T Virdee**²⁸ , **M Vojinovic** , **N Wardle** , **D Winterbottom** 

Imperial College, London, United Kingdom

J E Cole , **A Khan**, **P Kyberd** , **I D Reid** 

Brunel University, Uxbridge, United Kingdom

S Abdullin , **A Brinkerhoff** , **E Collins** , **M R Darwish** , **J Dittmann** , **K Hatakeyama** , **V Hegde** , **J Hiltbrand** , **B McMaster** , **J Samudio** , **S Sawant** , **C Sutantawibul** , **J Wilson** 

Baylor University, Waco, TX, United States of America

J M Hogan⁸¹ 

Bethel University, St. Paul, MN, United States of America

R Bartek , **A Dominguez** , **S Raj** , **A E Simsek** , **S S Yu** 

Catholic University of America, Washington, DC, United States of America

B Bam , **A Buchot Perraguin** , **S Campbell**, **R Chudasama** , **S I Cooper** , **C Crovella** , **G Fidalgo** , **S V Gleyzer** , **A Khukhunaishvili** , **K Matchev** , **E Pearson**, **C U Perez** , **P Rumerio**⁸² , **E Usai** , **R Yi** 

The University of Alabama, Tuscaloosa, AL, United States of America

S Cholak , **G De Castro**, **Z Demiragli** , **C Erice** , **C Fangmeier** , **C Fernandez Madrazo** , **E Fontanesi** , **J Fulcher** , **F Golf** , **S Jeon** , **J O’Cain**, **I Reed** , **J Rohlf** , **K Salyer** , **D Sperka** , **D Spitzbart** , **I Suarez** , **A Tsatsos** , **E Wurtz**, **A G Zecchinelli** 

Boston University, Boston, MA, United States of America

G Barone , **G Benelli** , **D Cutts** , **S Ellis** , **L Gouskos** , **M Hadley** , **U Heintz** , **K W Ho** , **T Kwon** , **L Lambrecht** , **G Landsberg** , **K T Lau** , **J Luo** , **S Mondal** , **J Roloff**, **T Russell** , **S Sagir**⁸³ , **X Shen** , **M Stamenkovic** , **N Venkatasubramanian** 

Brown University, Providence, RI, United States of America

S Abbott , **B Barton** , **R Breedon** , **H Cai** , **M Calderon De La Barca Sanchez** , **E Cannaert**, **M Chertok** , **M Citron** , **J Conway** , **P T Cox** , **R Erbacher** , **O Kukral** , **G Mocellin** , **S Ostrom** , **I Salazar Segovia**, **J S Tafoya Vargas** , **W Wei** , **S Yoo** 

University of California, Davis, Davis, CA, United States of America

K Adamidis, **M Bachtis** , **D Campos**, **R Cousins** , **A Datta** , **G Flores Avila** , **J Hauser** , **M Ignatenko** , **M A Iqbal** , **T Lam** , **Y F Lo** , **E Manca** , **A Nunez Del Prado** , **D Saltzberg** , **V Valuev** 

University of California, Los Angeles, CA, United States of America

R Clare , **J W Gary** , **G Hanson** 

University of California, Riverside, Riverside, CA, United States of America

A Aportela , **A Arora** , **J G Branson** , **S Cittolin** , **S Cooperstein** , **B D’Anzi** , **D Diaz** , **J Duarte** , **L Giannini** , **Y Gu**, **J Guiang** , **V Krutelyov** , **R Lee** , **J Letts** , **H Li**, **M Masciovecchio** , **F Mokhtar** , **S Mukherjee** , **M Pieri** , **D Primosch**, **M Quinnan** , **V Sharma** , **M Tadel** , **E Vourliotis** , **F Würthwein** , **A Yagil** , **Z Zhao**

University of California, San Diego, La Jolla, CA, United States of America

A Barzdukas , **L Brennan** , **C Campagnari** , **S Carron Montero**⁸⁴ , **K Downham** , **C Grieco** , **M M Hussain**, **J Incandela** , **M W K Lai**, **A J Li** , **P Masterson** , **J Richman** , **S N Santpur** , **U Sarica** , **R Schmitz** , **F Setti** , **J Sheplack** , **D Stuart** , **T Á Vami** , **X Yan** , **D Zhang** 

University of California, Santa Barbara—Department of Physics, Santa Barbara, CA, United States of America

A Albert , **S Bhattacharya** , **A Bornheim** , **O Cerri**, **R Kansal** , **J Mao** , **H B Newman** , **G Reales Gutiérrez**, **T Sievert**, **M Spiropulu** , **J R Vlimant** , **R A Wynne** , **S Xie** 

California Institute of Technology, Pasadena, CA, United States of America

J Alison , **S An** , **M Cremonesi**, **V Dutta** , **E Y Ertorer** , **T Ferguson** , **T A Gómez Espinosa** , **A Harilal** , **A Kallil Tharayil**, **M Kanemura**, **C Liu** , **P Meiring** , **T Mudholkar** , **S Murthy** , **P Palit** , **K Park** , **M Paulini** , **A Roberts** , **A Sanchez** , **W Terrill** 

Carnegie Mellon University, Pittsburgh, PA, United States of America

J P Cumalat , **W T Ford** , **A Hart** , **A Hassani** , **S Kwan** , **J Pearkes** , **C Savard** , **N Schonbeck** , **K Stenson** , **K A Ulmer** , **S R Wagner** , **N Zipper** , **D Zuolo** 

University of Colorado Boulder, Boulder, CO, United States of America

J Alexander , **X Chen** , **J Dickinson** , **A Duquette**, **J Fan** , **X Fan** , **J Grassi** , **S Hogan** , **P Kotamnives** , **J Monroy** , **G Niendorf** , **M Oshiro** , **J R Patterson** , **A Ryd** , **J Thom** , **P Wittich** , **R Zou** , **L Zygała** 

Cornell University, Ithaca, NY, United States of America

M Albrow , **M Alyari** , **O Amram** , **G Apollinari** , **A Apresyan** , **L A T Bauerdick** , **D Berry** , **J Berryhill** , **P C Bhat** , **K Burkett** , **J N Butler** , **A Canepa** , **G B Cerati** , **H W K Cheung** , **F Chlebana** , **C Cosby** , **G Cummings** , **I Dutta** , **V D Elvira** , **J Freeman** , **A Gandrakota** , **Z Gecse** , **L Gray** , **D Green**, **A Grummer** , **S Grünendahl** , **D Guerrero** , **O Gutsche** , **R M**

Harris , T C Herwig , J Hirschauer , B Jayatilaka , S Jindariani , M Johnson , U Joshi , T Klijnsma , B Klima , K H M Kwok , S Lammel , C Lee , D Lincoln , R Lipton , T Liu , K Maeshima , D Mason , P McBride , P Merkel , S Mrenna , S Nahn , J Ngadiuba , D Noonan , S Norberg , V Papadimitriou , N Pastika , K Pedro , C Pena⁸⁵ , C E Perez Lara , F Ravera , A Reinsvold Hall⁸⁶ , L Ristori , M Safdari , E Sexton-Kennedy , N Smith , A Soha , L Spiegel , S Stoynev , J Strait , L Taylor , S Tkaczuk , N V Tran , L Uplegger , E W Vaandering , C Wang , I Zoi 

Fermi National Accelerator Laboratory, Batavia, IL, United States of America

C Aruta , P Avery , D Bourilkov , P Chang , V Cherepanov , R D Field , C Huh , E Koenig , M Kolosova , J Konigsberg , A Korytov , N Menendez , G Mitselmakher , K Mohrman , A Muthirakalayil Madhu , N Rawal , S Rosenzweig , V Sulimov , Y Takahashi , J Wang 

University of Florida, Gainesville, FL, United States of America

T Adams , A Al Kadhimi , A Askew , S Bower , R Goff , R Hashmi , R S Kim , T Kolberg , G Martinez , M Mazza , H Prosper , P R Prova , R Yohay 

Florida State University, Tallahassee, FL, United States of America

B Alsufyani , S Butalla , S Das , M Hohlmann , M Lavinsky , E Yanes 

Florida Institute of Technology, Melbourne, FL, United States of America

M R Adams , N Barnett , A Baty , C Bennett , R Cavanaugh , R Escobar Franco , O Evdokimov , C E Gerber , H Gupta , M Hawksworth , A Hingrajiya , D J Hofman , J H Lee , C Mills , S Nanda , G Nigmatkulov , B Ozek , T Phan , D Pilipovic , R Pradhan , E Prifti , P Roy , T Roy , N Singh , M B Tonjes , N Varelas , M A Wadud , J Yoo 

University of Illinois Chicago, Chicago, IL, United States of America

M Alhusseini , D Blend , K Dilsiz⁸⁷ , O K Köseyan , A Mestvirishvili⁸⁸ , O Neogi , H Ogul⁸⁹ , Y Onel , A Penzo , C Snyder , E Tiras⁹⁰ 

The University of Iowa, Iowa City, IA, United States of America

B Blumenfeld , J Davis , A V Gritsan , L Kang , S Kyriacou , P Maksimovic , M Roguljic , S Sekhar , M V Srivastav , M Swartz 

Johns Hopkins University, Baltimore, MD, United States of America

A Abreu , L F Alcerro Alcerro , J Anguiano , S Arteaga Escatel , P Baringer , A Bean , Z Flowers , D Grove , J King , G Krintiras , M Lazarovits , C Le Mahieu , J Marquez , M Murray , M Nickel , S Popescu⁹¹ , C Rogan , C Royon , S Rudrabhatla , S Sanders , C Smith , G Wilson 

The University of Kansas, Lawrence, KS, United States of America

B Allmond , N Islam , A Ivanov , K Kaadze , Y Maravin , J Natoli , G G Reddy , D Roy , G Sorrentino 

Kansas State University, Manhattan, KS, United States of America

A Baden , A Belloni , J Bistany-riebman , S C Eno , N J Hadley , S Jabeen , R G Kellogg , T Koeth , B Kronheim , S Lascio , P Major , A C Mignerey , C Palmer , C Papageorgakis , M M Paranjpe , E Popova⁹² , A Shevelev , L Zhang 

University of Maryland, College Park, MD, United States of America

C Baldenegro Barrera , H Bossi , S Bright-Thonney , I A Cali , Y C Chen , P C Chou , M D'Alfonso , J Eysermans , C Freer , G Gomez-Ceballos , M Goncharov , G Grosso , P Harris , D Hoang , G M Innocenti , D Kovalskyi , J Krupa , L Lavezzo , Y-J Lee , K Long , C McGinn , A Novak , M I Park , C Paus , C Reissel , C Roland , G Roland , S Rothman , T A Sheng , G S F Stephans , D Walter , Z Wang , B Wyslouch , T J Yang 

Massachusetts Institute of Technology, Cambridge, MA, United States of America

B Crossman , W J Jackson , C Kapsiak , M Krohn , D Mahon , J Mans , B Marzocchi , R Rusack , O Sancar , R Saradhy , N Strobbe 

University of Minnesota, Minneapolis, MN, United States of America

K Bloom , **D R Claes** , **G Haza** , **J Hossain** , **C Joo** , **I Kravchenko** , **A Rohilla** , **J E Siado** , **W Tabb** , **A Vagnerini** , **A Wightman** , **F Yan** 

University of Nebraska-Lincoln, Lincoln, NE, United States of America

H Bandyopadhyay , **L Hay** , **H W Hsia** , **I Iashvili** , **A Kalogeropoulos** , **A Kharchilava** , **A Mandal** , **M Morris** , **D Nguyen** , **S Rappoccio** , **H Rejeb Sfar** , **A Williams** , **P Young** , **D Yu** 

State University of New York at Buffalo, Buffalo, NY, United States of America

G Alverson , **E Barberis** , **J Bonilla** , **B Bylsma** , **M Campana** , **J Dervan** , **Y Haddad** , **Y Han** , **I Israr** , **A Krishna** , **M Lu** , **N Manganelli** , **R Mccarthy** , **D M Morse** , **T Orimoto** , **A Parker** , **L Skinnari** , **C S Thoreson** , **E Tsai** , **D Wood** 

Northeastern University, Boston, MA, United States of America

S Dittmer , **K A Hahn** , **Y Liu** , **M McGinnis** , **Y Miao** , **D G Monk** , **M H Schmitt** , **A Taliervo** , **M Velasco** , **J Wang** 

Northwestern University, Evanston, IL, United States of America

G Agarwal , **R Band** , **R Bucci** , **S Castells** , **A Das** , **A Ehnis** , **R Goldouzian** , **M Hildreth** , **K Hurtado Anampa** , **T Ivanov** , **C Jessop** , **A Karneyeu** , **K Lannon** , **J Lawrence** , **N Loukas** , **L Lutton** , **J Mariano** , **N Marinelli** , **I Mcalister** , **T McCauley** , **C Mcgrady** , **C Moore** , **Y Musienko** ²² , **H Nelson** , **M Osherson** , **A Piccinelli** , **R Ruchti** , **A Townsend** , **Y Wan** , **M Wayne** , **H Yockey** 

University of Notre Dame, Notre Dame, IN, United States of America

A Basnet , **M Carrigan** , **R De Los Santos** , **L S Durkin** , **C Hill** , **M Joyce** , **M Nunez Ornelas** , **D A Wenzl** , **B L Winer** , **B R Yates** 

The Ohio State University, Columbus, OH, United States of America

H Bouchamaoui , **P Das** , **G Dezoort** , **P Elmer** , **A Frankenthal** , **M Galli** , **B Greenberg** , **N Haubrich** , **K Kennedy** , **G Kopp** , **Y Lai** , **D Lange** , **A Loeliger** , **D Marlow** , **I Ojalvo** , **J Olsen** , **F Simpson** , **D Stickland** , **C Tully** 

Princeton University, Princeton, NJ, United States of America

S Malik , **R Sharma** 

University of Puerto Rico, Mayaguez, PR, United States of America

S Chandra , **R Chawla** , **A Gu** , **L Gutay** , **M Jones** , **A W Jung** , **D Kondratyev** , **M Liu** , **G Negro** , **N Neumeister** , **G Paspalaki** , **S Piperov** , **N R Saha** , **J F Schulte** , **F Wang** , **A Wildridge** , **W Xie** , **Y Yao** , **Y Zhong** 

Purdue University, West Lafayette, IN, United States of America

N Parashar , **A Pathak** , **E Shumka** 

Purdue University Northwest, Hammond, IN, United States of America

D Acosta , **A Agrawal** , **C Arbour** , **T Carnahan** , **K M Ecklund** , **S Freed** , **P Gardner** , **F J M Geurts** , **T Huang** , **I Krommydas** , **N Lewis** , **W Li** , **J Lin** , **O Miguel Colin** , **B P Padley** , **R Redjimi** , **J Rotter** , **M Wulansatiti** , **E Yigitbasi** , **Y Zhang** 

Rice University, Houston, TX, United States of America

O Bessidskaia Bylund , **A Bodek** , **P de Barbaro** [†] , **R Demina** , **A Garcia-Bellido** , **H S Hare** , **O Hindrichs** , **N Parmar** , **P Parygin** ⁹² , **H Seo** , **R Taus** 

University of Rochester, Rochester, NY, United States of America

B Chiarito , **J P Chou** , **S V Clark** , **S Donnelly** , **D Gadkari** , **Y Gershtein** , **E Halkiadakis** , **C Houghton** , **D Jaroslawski** , **A Kobert** , **S Konstantinou** , **I Laflotte** , **A Lath** , **J Martins** , **M Perez Prada** , **B Rand** , **J Reichert** , **P Saha** , **S Salur** , **S Schnetzer** , **S Somalwar** , **R Stone** , **S A Thayil** , **S Thomas** , **J Vora** 

Rutgers, The State University of New Jersey, Piscataway, NJ, United States of America

D Ally , **A G Delannoy** , **S Fiorendi** , **J Harris** , **T Holmes** , **A R Kanuganti** , **N Karunarathna** , **J Lawless** , **L Lee** , **E Nibigira** , **B Skipworth** , **S Spanier** 

University of Tennessee, Knoxville, TN, United States of America

D Aebi¹, M Ahmad², T Akhter³, K Androsov⁴, A Bolshov, O Bouhali⁹³, A Cagnotta⁵, V D'Amante⁶, R Eusebi⁷, P Flanagan⁸, J Gilmore⁹, Y Guo, T Kamon¹⁰, S Luo¹¹, R Mueller¹², A Safonov¹³

Texas A&M University, College Station, TX, United States of America

N Akchurin¹⁴, J Damgov¹⁵, Y Feng¹⁶, N Gogate¹⁷, Y Kazhykarim, K Lamichhane¹⁸, S W Lee¹⁹, C Madrid²⁰, A Mankel²¹, T Peltola²², I Volobouev²³

Texas Tech University, Lubbock, TX, United States of America

E Appelt²⁴, Y Chen²⁵, S Greene, A Gurrola²⁶, W Johns²⁷, R Kunnawalkam Elayavalli²⁸, A Melo²⁹, D Rathjens³⁰, F Romeo³¹, P Sheldon³², S Tuo³³, J Velkovska³⁴, J Viinikainen³⁵, J Zhang

Vanderbilt University, Nashville, TN, United States of America

B Cardwell³⁶, H Chung³⁷, B Cox³⁸, J Hakala³⁹, G Hamilton Ilha Machado, R Hirosky⁴⁰, M Jose, A Ledovsky⁴¹, C Mantilla⁴², C Neu⁴³, C Ramón Álvarez⁴⁴

University of Virginia, Charlottesville, VA, United States of America

S Bhattacharya⁴⁵, P E Karchin⁴⁶

Wayne State University, Detroit, MI, United States of America

A Aravind⁴⁷, S Banerjee⁴⁸, K Black⁴⁹, T Bose⁵⁰, E Chavez⁵¹, S Dasu⁵², P Everaerts⁵³, C Galloni, H He⁵⁴, M Herndon⁵⁵, A Herve⁵⁶, C K Koraka⁵⁷, S Lomte⁵⁸, R Loveless⁵⁹, A Mallampalli⁶⁰, A Mohammadi⁶¹, S Mondal, T Nelson, G Parida⁶², L Pétré⁶³, D Pinna⁶⁴, A Savin, V Shang⁶⁵, V Sharma⁶⁶, W H Smith⁶⁷, D Teague, H F Tsoi⁶⁸, W Vetens⁶⁹, A Warden⁷⁰

University of Wisconsin—Madison, Madison, WI, United States of America

S Afanasiev⁷¹, V Alexakhin⁷², Yu Andreev⁷³, T Aushev⁷⁴, D Budkouski⁷⁵, R Chistov⁹⁴, M Danilov⁹⁴, T Dimova⁹⁴, A Ershov⁹⁴, S Gninenko⁷⁶, I Gorbunov⁷⁷, A Gribushin⁹⁴, A Kamenev⁷⁸, V Karjavine⁷⁹, M Kirsanov⁸⁰, V Klyukhin⁹⁴, O Kodolova⁹⁵, V Korenkov⁸¹, I Korsakov, A Kozyrev⁹⁴, N Krasnikov⁸², A Lanev⁸³, A Malakhov⁸⁴, V Matveev⁹⁴, A Nikitenko^{96,95}, V Palichik⁸⁵, V Perelygin⁸⁶, S Petrushanko⁹⁴, S Polikarpov⁹⁴, O Radchenko⁹⁴, M Savina⁸⁷, V Shalaev⁸⁸, S Shmatov⁸⁹, S Shulha⁹⁰, Y Skovpen⁹⁴, K Slizhevskiy, V Smirnov⁹¹, O Teryaev⁹², I Tlisova⁹⁴, A Toropin⁹³, N Voytishin⁹⁴, B S Yuldashev^{†,97}, A Zarubin⁹⁸, I Zhizhin⁹⁹

Authors affiliated with an international laboratory covered by a cooperation agreement with CERN

E Boos¹⁰⁰, V Bunichev¹⁰¹, M Dubinin⁸⁵, V Savrin¹⁰², A Snigirev¹⁰³, L Dudko¹⁰⁴, K Ivanov¹⁰⁵, V Kim²², V Murzin¹⁰⁶, V Oreshkin¹⁰⁷, D Sosnov¹⁰⁸

Authors affiliated with an institute formerly covered by a cooperation agreement with CERN

[†]Deceased

¹Also at Yerevan State University, Yerevan, Armenia

²Also at TU Wien, Vienna, Austria

³Also at Ghent University, Ghent, Belgium

⁴Also at Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil

⁵Also at FACAMP—Faculdades de Campinas, Sao Paulo, Brazil

⁶Also at Universidade Estadual de Campinas, Campinas, Brazil

⁷Also at Federal University of Rio Grande do Sul, Porto Alegre, Brazil

⁸Also at The University of the State of Amazonas, Manaus, Brazil

⁹Also at University of Chinese Academy of Sciences, Beijing, People's Republic of China

¹⁰Also at China Center of Advanced Science and Technology, Beijing, People's Republic of China

¹¹Also at University of Chinese Academy of Sciences, Beijing, People's Republic of China

¹²Also at School of Physics, Zhengzhou University, Zhengzhou, People's Republic of China

¹³Now at Henan Normal University, Xinxiang, People's Republic of China

¹⁴Also at University of Shanghai for Science and Technology, Shanghai, People's Republic of China

¹⁵Now at The University of Iowa, Iowa City, IA, United States of America

¹⁶Also at Center for High Energy Physics, Peking University, Beijing, People's Republic of China

¹⁷Also at Zewail City of Science and Technology, Zewail, Egypt

¹⁸Also at British University in Egypt, Cairo, Egypt

¹⁹Now at Ain Shams University, Cairo, Egypt

²⁰Also at Purdue University, West Lafayette, IN, United States of America

²¹Also at Université de Haute Alsace, Mulhouse, France

- ²²Also at an institute formerly covered by a cooperation agreement with CERN
- ²³Also at University of Hamburg, Hamburg, Germany
- ²⁴Also at RWTH Aachen University, III. Physikalisches Institut A, Aachen, Germany
- ²⁵Also at Bergische University Wuppertal (BUW), Wuppertal, Germany
- ²⁶Also at Brandenburg University of Technology, Cottbus, Germany
- ²⁷Also at Forschungszentrum Jülich, Juelich, Germany
- ²⁸Also at CERN, European Organization for Nuclear Research, Geneva, Switzerland
- ²⁹Also at HUN-REN ATOMKI—Institute of Nuclear Research, Debrecen, Hungary
- ³⁰Now at Universitatea Babes-Bolyai—Facultatea de Fizica, Cluj-Napoca, Romania
- ³¹Also at MTA-ELTE Lendület CMS Particle and Nuclear Physics Group, Eötvös Loránd University, Budapest, Hungary
- ³²Also at HUN-REN Wigner Research Centre for Physics, Budapest, Hungary
- ³³Also at Physics Department, Faculty of Science, Assiut University, Assiut, Egypt
- ³⁴Also at The University of Kansas, Lawrence, KS, United States of America
- ³⁵Also at Punjab Agricultural University, Ludhiana, India
- ³⁶Also at University of Hyderabad, Hyderabad, India
- ³⁷Also at Indian Institute of Science (IISc), Bangalore, India
- ³⁸Also at University of Visva-Bharati, Santiniketan, India
- ³⁹Also at IIT Bhubaneswar, Bhubaneswar, India
- ⁴⁰Also at Institute of Physics, Bhubaneswar, India
- ⁴¹Also at Deutsches Elektronen-Synchrotron, Hamburg, Germany
- ⁴²Also at Isfahan University of Technology, Isfahan, Iran
- ⁴³Also at Sharif University of Technology, Tehran, Iran
- ⁴⁴Also at Department of Physics, University of Science and Technology of Mazandaran, Behshahr, Iran
- ⁴⁵Also at Department of Physics, Faculty of Science, Arak University, ARAK, Iran
- ⁴⁶Also at Helwan University, Cairo, Egypt
- ⁴⁷Also at Italian National Agency for New Technologies, Energy and Sustainable Economic Development, Bologna, Italy
- ⁴⁸Also at Centro Siciliano di Fisica Nucleare e di Struttura Della Materia, Catania, Italy
- ⁴⁹Also at Università degli Studi Guglielmo Marconi, Roma, Italy
- ⁵⁰Also at Scuola Superiore Meridionale, Università di Napoli ‘Federico II’, Napoli, Italy
- ⁵¹Also at Fermi National Accelerator Laboratory, Batavia, IL, United States of America
- ⁵²Also at Lulea University of Technology, Lulea, Sweden
- ⁵³Also at Consiglio Nazionale delle Ricerche—Istituto Officina dei Materiali, Perugia, Italy
- ⁵⁴Also at UPES—University of Petroleum and Energy Studies, Dehradun, India
- ⁵⁵Also at Institut de Physique des 2 Infinis de Lyon (IP2I), Villeurbanne, France
- ⁵⁶Also at Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia
- ⁵⁷Also at Trincomalee Campus, Eastern University, Sri Lanka, Nilaveli, Sri Lanka
- ⁵⁸Also at Saegis Campus, Nugegoda, Sri Lanka
- ⁵⁹Also at National and Kapodistrian University of Athens, Athens, Greece
- ⁶⁰Also at Ecole Polytechnique Fédérale Lausanne, Lausanne, Switzerland
- ⁶¹Also at Universität Zürich, Zurich, Switzerland
- ⁶²Also at Stefan Meyer Institute for Subatomic Physics, Vienna, Austria
- ⁶³Also at Near East University, Research Center of Experimental Health Science, Mersin, Türkiye
- ⁶⁴Also at Konya Technical University, Konya, Türkiye
- ⁶⁵Also at Izmir Bakircay University, Izmir, Türkiye
- ⁶⁶Also at Adiyaman University, Adiyaman, Türkiye
- ⁶⁷Also at Bozok Universitetesi Rektörlüğü, Yozgat, Türkiye
- ⁶⁸Also at Istanbul Sabahattin Zaim University, Istanbul, Türkiye
- ⁶⁹Also at Marmara University, Istanbul, Türkiye
- ⁷⁰Also at Milli Savunma University, Istanbul, Türkiye
- ⁷¹Also at Informatics and Information Security Research Center, Gebze/Kocaeli, Türkiye
- ⁷²Also at Kafkas University, Kars, Türkiye
- ⁷³Now at Istanbul Okan University, Istanbul, Türkiye
- ⁷⁴Also at Hacettepe University, Ankara, Türkiye
- ⁷⁵Also at Erzincan Binali Yildirim University, Erzincan, Türkiye
- ⁷⁶Also at Istanbul University—Cerrahpasa, Faculty of Engineering, Istanbul, Türkiye
- ⁷⁷Also at Yildiz Technical University, Istanbul, Türkiye

- ⁷⁸Also at Istinye University, Istanbul, Türkiye
- ⁷⁹Also at School of Physics and Astronomy, University of Southampton, Southampton, United Kingdom
- ⁸⁰Also at Monash University, Faculty of Science, Clayton, Australia
- ⁸¹Also at Bethel University, St. Paul, MN, United States of America
- ⁸²Also at Università di Torino, Torino, Italy
- ⁸³Also at Karamanoğlu Mehmetbey University, Karaman, Türkiye
- ⁸⁴Also at California Lutheran University, Thousand Oaks, CA, United States of America
- ⁸⁵Also at California Institute of Technology, Pasadena, CA, United States of America
- ⁸⁶Also at United States Naval Academy, Annapolis, MD, United States of America
- ⁸⁷Also at Bingöl University, Bingöl, Türkiye
- ⁸⁸Also at Georgian Technical University, Tbilisi, Georgia
- ⁸⁹Also at Sinop University, Sinop, Türkiye
- ⁹⁰Also at Erciyes University, Kayseri, Türkiye
- ⁹¹Also at Horia Hulubei National Institute of Physics and Nuclear Engineering (IFIN-HH), Bucharest, Romania
- ⁹²Now at another institute formerly covered by a cooperation agreement with CERN
- ⁹³Also at Hamad Bin Khalifa University (HBKU), Doha, Qatar
- ⁹⁴Also at another institute formerly covered by a cooperation agreement with CERN
- ⁹⁵Also at Yerevan Physics Institute, Yerevan, Armenia
- ⁹⁶Also at Imperial College, London, United Kingdom
- ⁹⁷Also at Institute of Nuclear Physics of the Uzbekistan Academy of Sciences, Tashkent, Uzbekistan

References

- [1] Heimel T, Kasieczka G, Plehn T and Thompson J M 2019 QCD or what? *SciPost Phys.* **6** 030
- [2] Farina M, Nakai Y and Shih D 2020 Searching for new physics with deep autoencoders *Phys. Rev. D* **101** 075021
- [3] Finke T *et al* 2021 Autoencoders for unsupervised anomaly detection in high energy physics *J. High Energy Phys.* **JHEP06(2021)161**
- [4] Yoon S, Noh Y-K and Park F 2021 Autoencoding under normalization constraints *Proc. 38th Int. Conf. on Machine Learning* p 12087
- [5] Cohen T, Lisanti M and Lou H K 2015 Semivisible jets: dark matter undercover at the LHC *Phys. Rev. Lett.* **115** 171804
- [6] CMS Collaboration 2008 The CMS experiment at the CERN LHC *J. Instrum.* **3** S08004
- [7] CMS Collaboration 2024 Development of the CMS detector for the CERN LHC run 3 *J. Instrum.* **19** 05064
- [8] CMS Collaboration 2020 Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV *J. Instrum.* **15** 10017
- [9] CMS Collaboration 2017 The CMS trigger system *J. Instrum.* **12** 01020
- [10] CMS Collaboration 2024 Performance of the CMS high-level trigger during LHC run 2 *J. Instrum.* **19** 11021
- [11] CMS Collaboration 2021 Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC *J. Instrum.* **16** 05014
- [12] CMS Collaboration 2018 Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV *J. Instrum.* **13** 06015
- [13] CMS Collaboration 2014 Description and performance of track and primary-vertex reconstruction with the CMS tracker *J. Instrum.* **9** 10009
- [14] CMS Collaboration 2017 Particle-flow reconstruction and global event description with the CMS detector *J. Instrum.* **12** 10003
- [15] CMS Collaboration 2017 Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV *J. Instrum.* **12** 02014
- [16] CMS Collaboration 2019 Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector *J. Instrum.* **14** 07004
- [17] Kantorovich L V 1939 Mathematical methods of organizing and planning production *Manage. Sci.* **6** 366
- [18] Vaserstein L N 1969 Markov processes over denumerable products of spaces describing large systems of automata *Problems Inform. Transm.* **5** 47 (available at: <http://mi.mathnet.ru/ppi1811>)
- [19] Villani C 2009 *Optimal Transport: Old and New* (Grundlehren der Mathematischen Wissenschaften vol 338) (Springer) (<https://doi.org/10.1007/978-3-540-71050-9>)
- [20] Kramer M A 1992 Autoassociative neural networks *Comput. Chem. Eng.* **16** 313
- [21] Smolensky P 1986 Information processing in dynamical systems: foundations of harmony theory *Parallel Distributed Processing, Volume 1: Explorations in the Microstructure of Cognition: Foundations* (The MIT Press) p 194
- [22] Hinton G 2002 Training products of experts by minimizing contrastive divergence *Neural Comput.* **14** 1771
- [23] Teh Y W, Welling M, Osindero S and Hinton G E 2003 Energy-based models for sparse overcomplete representations *J. Mach. Learn. Res.* **4** 1235
- [24] Jaynes E T 1957 Information theory and statistical mechanics *Phys. Rev.* **106** 620
- [25] Kingma D P and Welling M 2014 Auto-encoding variational Bayes *2nd Int. Conf. on Learning Representations* (arXiv:1312.6114)
- [26] Makhzani A, Shlens J, Jaitly N and Goodfellow I 2016 Adversarial autoencoders *Int. Conf. on Learning Representations* (arXiv:1511.05644)
- [27] Goodfellow I J 2014 Generative adversarial nets *Proc. Int. Conf. on Advances in Neural Information Processing Systems* vol 27 (Curran Associates, Inc.) p 2672 (arXiv:1406.2661)
- [28] Tolstikhin I, Bousquet O, Gelly S and Schoelkopf B 2018 Wasserstein auto-encoders *Int. Conf. on Learning Representations* (arXiv:1711.01558)
- [29] Kantorovich L V and Rubinshtein S 1958 On a space of totally additive functions *Vestnik St. Petersburg Univ. Math.* **13** 52

- [30] Arjovsky M, Chintala S and Bottou L 2017 Wasserstein GAN *Proc. 34th Int. Conf. on Machine Learning* vol 70 p 214 (arXiv:1701.07875)
- [31] Flamary R 2021 POT: python optimal transport *J. Mach. Learn. Res.* **22** 1
- [32] Fatras K 2019 Learning with minibatch Wasserstein: asymptotic and gradient properties *Proc. 23rd Int. Conf. on Artificial Intelligence and Statistics* vol 108 p 2131 (arXiv:1910.04091)
- [33] Nalisnick E *et al* 2019 Do deep generative models know what they don't know? *Int. Conf. on Learning Representations* (arXiv:1810.09136)
- [34] Xiao Z, Yan Q and Amit Y 2020 Likelihood regret: an out-of-distribution detection score for variational auto-encoder *Proc. Int. Conf. on Advances in Neural Information Processing Systems* vol 33 p 20685 (arXiv:2003.02977)
- [35] Kirichenko P, Izmailov P and Wilson A G 2020 Why normalizing flows fail to detect out-of-distribution data *Proc. Int. Conf. on Advances in Neural Information Processing Systems* vol 33 p 20578 (arXiv:2006.08545)
- [36] Kamkari H *et al* 2024 A geometric explanation of the likelihood OOD detection paradox *Proc. 41st Int. Conf. on Machine Learning (PMLR)* (arXiv:2403.18910)
- [37] Cook E D, Lavoie M-A and Waslander S L 2024 Feature density estimation for out-of-distribution detection via normalizing flows *Proc. Conf. on Robots and Vision* (arXiv:2402.06537)
- [38] Pidhorskyi S, Almohsen R and Doretto G 2018 Generative probabilistic novelty detection with adversarial autoencoders *Proc. 32nd Int. Conf. on Neural Information Processing Systems* p 6823 (arXiv:1807.02588)
- [39] Akcay S, Atapour-Abarghouei A and Breckon T P 2018 GANomaly: semi-supervised anomaly detection via adversarial training *Asian Conf. on Computer Vision* (Springer) p 622 (arXiv:1805.06725)
- [40] Schlegl T *et al* 2019 f-AnoGAN: fast unsupervised anomaly detection with generative adversarial networks *Med. Image Anal.* **54** 30
- [41] Du Y and Mordatch I 2019 Implicit generation and modeling with energy-based models *Proc. 33rd Int. Conf. on Neural Information Processing Systems* p 324 (arXiv:1903.08689)
- [42] Gandrakota A 2025 Realtime anomaly detection at the L1 trigger of CMS experiment *Proc. Sci.* **476** 1025
- [43] CMS Collaboration 2025 Model-agnostic search for dijet resonances with anomalous jet substructure in proton-proton collisions at $\sqrt{s} = 13$ TeV *Rep. Prog. Phys.* **88** 067802
- [44] ATLAS Collaboration 2024 Search for new phenomena in two-body invariant mass distributions using unsupervised machine learning for anomaly detection at $\sqrt{s} = 13$ TeV with the ATLAS detector *Phys. Rev. Lett.* **132** 081801
- [45] Dillon B M *et al* 2023 A normalized autoencoder for LHC triggers *SciPost Phys. Core* **6** 074
- [46] Rubin V C, Thonnard N and Ford W K 1980 Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 ($R = 4$ kpc) to UGC 2885 ($R = 122$ kpc) *Astrophys. J.* **238** 471
- [47] Persic M, Salucci P and Stel F 1996 The universal rotation curve of spiral galaxies: I. The dark matter connection *Mon. Not. R. Astron. Soc.* **281** 27
- [48] Clowe D *et al* 2006 A direct empirical proof of the existence of dark matter *Astrophys. J.* **648** L109
- [49] DES Collaboration 2018 Dark energy survey year 1 results: curved-sky weak lensing mass map *Mon. Not. R. Astron. Soc.* **475** 3165
- [50] Collaboration P 2020 Planck 2018 results. VI. Cosmological parameters *Astron. Astrophys.* **641** A6
- [51] Strassler M J and Zurek K M 2007 Echoes of a hidden valley at hadron colliders *Phys. Lett. B* **651** 374
- [52] CMS Collaboration 2022 Search for resonant production of strongly coupled dark matter in proton-proton collisions at 13 TeV *J. High Energy Phys.* **JHEP06(2022)156**
- [53] ATLAS Collaboration 2025 Search for new physics in final states with semivisible jets or anomalous signatures using the ATLAS detector *Phys. Rev. D* **112** 012021
- [54] Cohen T, Lisanti M, Lou H K and Mishra-Sharma S 2017 LHC searches for dark sector showers *J. High Energy Phys.* **JHEP11(2017)196**
- [55] Bernreuther E, Kahlhoefer F, Krämer M and Tunney P 2020 Strongly interacting dark sectors in the early Universe and at the LHC through a simplified portal *J. High Energy Phys.* **JHEP01(2020)162**
- [56] Alwall J *et al* 2014 The automated computation of tree-level and next-to-leading order differential cross sections and their matching to parton shower simulations *J. High Energy Phys.* **JHEP07(2014)079**
- [57] Sjöstrand T *et al* 2015 An introduction to PYTHIA 8.2 *Comput. Phys. Commun.* **191** 159
- [58] GEANT4 Collaboration 2003 GEANT4—a simulation toolkit *Nucl. Instrum. Methods A* **506** 250
- [59] NNPDF Collaboration 2017 Parton distributions from high-precision collider data *Eur. Phys. J. C* **77** 663
- [60] Cacciari M, Salam G P and Soyez G 2008 The anti-t jet clustering algorithm *J. High Energy Phys.* **JHEP04(2008)063**
- [61] Cacciari M, Salam G P and Soyez G 2012 FastJet user manual *Eur. Phys. J. C* **72** 1896
- [62] CMS Collaboration 2013 Performance of quark/gluon discrimination in 8 TeV pp data *CMS Physics Analysis Summary CMS-PAS-JME-13-002* (available at: <http://cds.cern.ch/record/1599732>)
- [63] Komiske P T, Metodiev E M and Thaler J 2018 Energy flow polynomials: a complete linear basis for jet substructure *J. High Energy Phys.* **JHEP04(2018)013**
- [64] Larkoski A J, Salam G P and Thaler J 2013 Energy correlation functions for jet substructure *J. High Energy Phys.* **JHEP06(2013)108**
- [65] Larkoski A J, Marzani S, Soyez G and Thaler J 2014 Soft drop *J. High Energy Phys.* **JHEP05(2014)146**
- [66] Thaler J and Van Tilburg K 2011 Identifying boosted objects with N-subjettiness *J. High Energy Phys.* **JHEP03(2011)015**
- [67] Pedregosa F *et al* 2011 Scikit-learn: machine learning in Python *J. Mach. Learn. Res.* **12** 2825 (arXiv:1201.0490)
- [68] Canelli F *et al* 2022 Autoencoders for semivisible jet detection *J. High Energy Phys.* **JHEP02(2022)074**
- [69] 2025 Source code repository for this analysis (available at: <https://gitlab.cern.ch/cms-analysis/mlg/mlg-24-002/wnae>)
- [70] Bonneel N, De Panne M V, Paris S and Heidrich W 2011 Displacement interpolation using lagrangian mass transport *ACM Trans. Graph.* **30** 1
- [71] Dudley R M 1969 The speed of mean Glivenko-Cantelli convergence *Ann. Math. Stat.* **40** 40
- [72] Weed J and Bach F 2019 Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance *Bernoulli* **25** 2620
- [73] Rabin J, Peyré G, Delon J and Bernot M 2012 Wasserstein barycenter and its application to texture mixing *Scale Space and Variational Methods in Computer Vision* (Springer) p 435
- [74] Bonneel N, Rabin J, Peyré G and Pfister H 2015 Sliced and Radon Wasserstein barycenters of measures *J. Math. Imaging Vis.* **51** 22
- [75] Nadjahi K, Durmus A, Şimşekli U and Badeau R 2019 Asymptotic guarantees for learning generative models with the sliced-Wasserstein distance *Proc. Int. Conf. on Advances in Neural Information Processing Systems* (arXiv:1906.04516)

- [76] Nadjahi K *et al* 2020 Statistical and topological properties of sliced probability divergences *Proc. Int. Conf. on Advances in Neural Information Processing Systems* (arXiv:[2003.05783](#))
- [77] Paszke A *et al* 2019 PyTorch: an imperative style, high-performance deep learning library *Proc. 33rd Int. Conf. on Neural Information Processing Systems* vol 32 p 721 (arXiv:[1912.01703](#))
- [78] Tieleman T 2008 Training restricted Boltzmann machines using approximations to the likelihood gradient *Proc. 25th Int. Conf. on Machine Learning* p [1064](#)