

Applying machine learning techniques to intermediate-length cascade decays

Maaz Ul Haq,^{*} Can Kilic[✉], Benjamin Lawrence-Sanderson, and Ram Purandhar Reddy Sudha[✉]

Theory Group, Department of Physics, University of Texas at Austin, Austin, Texas 78712, USA



(Received 18 October 2022; accepted 13 July 2023; published 1 August 2023)

In the collider phenomenology of extensions of the Standard Model with partner particles, cascade decays occur generically, and they can be challenging to discover when the spectrum of new particles is compressed and the signal cross section is low. Achieving discovery-level significance and measuring the properties of the new particles appearing as intermediate states in the cascade decays is a long-standing problem, with analysis techniques for some decay topologies already optimized. We focus our attention on a benchmark decay topology with four final state particles where there is room for improvement and where multidimensional analysis techniques have been shown to be effective in the past. We apply machine learning techniques in order to identify effective human-level kinematic observables for discovery, spin determination, and mass measurement. We quantify the performance of these analyses as a function of the signal size. In agreement with past work, we confirm that the kinematic observable Δ_4 is highly effective.

DOI: [10.1103/PhysRevD.108.035002](https://doi.org/10.1103/PhysRevD.108.035002)

I. INTRODUCTION

While the Standard Model (SM) of particle physics is extremely successful in describing the known particles and their interactions, it is also known to be an incomplete description of fundamental physics. Some of the best-studied extensions of the Standard Model that aim to stabilize the electroweak scale or explain the observed dark matter (DM) relic abundance hint at the existence of new degrees of freedom at roughly the TeV scale. Unfortunately, collider searches for such new particles have yielded only null results until now. Therefore, as the Large Hadron Collider (LHC) is getting ready to start its third run, even if new physics is finally discovered, the signal will in all likelihood either have low statistics or be difficult to distinguish from backgrounds, or possibly both. This makes it all the more crucial that LHC searches be optimized for maximal efficiency with these challenges in mind. Similarly, postdiscovery, the measurement of the properties of the new particles, such as their masses, will be challenging for the same reasons.

Among possible final states for new physics at the LHC, supersymmetry (SUSY)-like production and decay channels of color-neutral particles, especially with a compressed spectrum for the new particles, offer good examples for the type of signatures mentioned above, as electroweak-only

charged particles have low production cross sections, and compressed spectra result in soft momenta in the final state, so backgrounds cannot be reduced by using hard cuts. Our definition of a SUSY-like channel is that new particles can be produced only in pairs due to a Z_2 symmetry and that each one decays to a lighter new particle plus SM particles, until the lightest new particle is reached, which is collider stable and a DM candidate and which cannot be detected. Note that this definition applies equally well to scenarios where the new particles have the same spin as their SM partners, such as in the case of extra dimensional models and others. Since there is an invisible particle at the end of any decay chain in such final states, no resonance can be reconstructed from any subset of visible particles.

For short decay chains, very little kinematic information is available event by event. Observables such as missing transverse energy (MET) and transverse mass variables such as m_T and m_{T2} [1–9] fully use this available information and provide the best chance for discovery and for the mass measurement of the unknown particles. As a result, all available information can be extracted from one-dimensional distributions of a small number of kinematic variables. In the other extreme, for sufficiently long decay chains, there is sufficient kinematic information available in the events for the determination of the complete spectrum by algebraic methods [10–30]. Both of these possibilities have been well studied, and there is little room for improvement. For an extensive review of kinematic variables used in collider phenomenology, we direct the reader to Ref. [31].

There are, however, final states that lie between these extremes, where algebraic methods cannot be used, but

^{*}Deceased.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

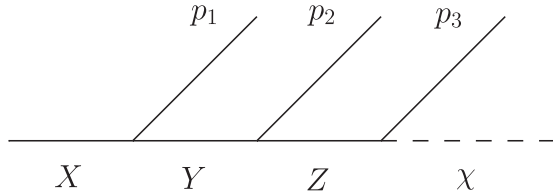


FIG. 1. Feynman diagram for our benchmark decay chain. X , Y , Z , and χ are all new particles, while $p_{1,2,3}$ are SM particles.

there are sufficiently many kinematic observables such that their correlations also contain crucial information that cannot be extracted by plotting only commonly used one-dimensional distributions such as kinematic edges and end points. It was shown in Ref. [32] that a decay chain that proceeds via three consecutive two-body decays has this property (see Fig. 1), and subsequent papers [33,34] explored how an analysis based on the full dimensionality of the Lorentz-invariant observables can be used to enhance discovery prospects as well as to improve the precision of mass measurements. In other words, in these event topologies there are features of the phase space distribution that reveal themselves only with a multidimensional study but not in the standard one-dimensional projections. In particular, it was shown that the variable Δ_4 , to be reviewed in the next section, is a highly effective observable for measuring not only the mass differences between successive particles in the cascade decay, but also the overall mass scale along the “flat direction,” where the locations of kinematic edges and end points in each step of the cascade remain fixed. In this paper, we also adopt the same decay chain as the benchmark case for our study. The decay chain that we consider in this paper has been previously studied using end points in kinematic observables [35]. The effectiveness of a number of kinematic observables for spin determination has also been studied [27,36–38].

Since we are considering SUSY-like collider signatures, the new particles need to be pair produced. However, if the decay chain in Fig. 1 appears on both sides of the event, then this event topology would, in fact, contain sufficiently many visible final state particles that algebraic methods could be used. We instead focus on scenarios where the associated production of X - χ is the dominant production channel (or possibly the next-to-dominant channel, after χ - χ production, but, for sufficiently heavy χ , this channel is challenging to observe). One can, for instance, consider a t -channel production diagram, where the mediator in the t channel couples more strongly to χ than it does to X . This is illustrated in Fig. 2 for two choices of particle spins, which we will introduce in Sec. IV. All the visible information in the event is then coming from the X side of the event, and the χ on the other side of the event affects only the total MET vector among the observables.

Recently, advances in machine learning techniques have led to significant improvements in multidimensional data

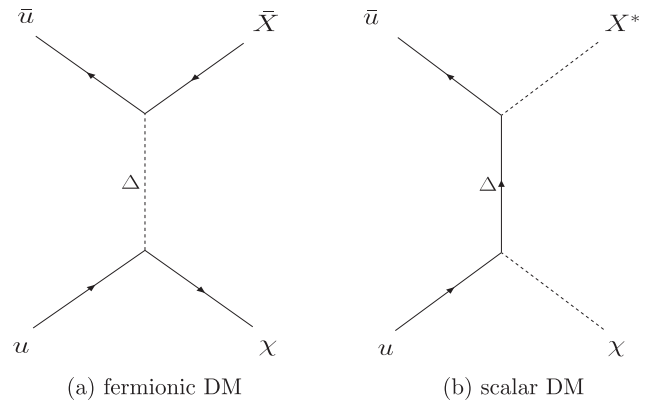


FIG. 2. The Feynman diagrams for the production of the new particles in our benchmark event topology with the two spin configurations we consider. Fermions are represented by solid lines, while scalars are represented by dotted lines.

analysis, and in this paper we evaluate the effectiveness of these techniques for the discovery, spin determination, and mass measurement in this benchmark decay chain, as a representative of SUSY-like decays that have sufficiently many kinematic observables to warrant a multidimensional analysis but not enough for a full reconstruction via algebraic techniques. In this study, we first consider how well a deep neural network (DNN) can achieve these goals without human guidance, that is, by taking the final state momenta in the events as its only input. We then proceed to interpret the DNN in terms of human-level (HL) kinematic variables, and we show that an analysis based on the HL variables results in performance comparable to that of the detector-level (DL) inputs. In particular, our results confirm that Δ_4 is highly correlated with the output of the DNN, in agreement with the conclusions of Refs. [32–34].

Supervised machine-learning-based analyses have been previously applied for the discovery of exotic collider events in Refs. [39,40]. In Ref. [41], the authors have proposed methods for discovery across a range of model parameters. Applications also include resolution of combinatorics in complex topologies [42] and parameter inference [43,44].

The visible particles $p_{1,2,3}$ can be chosen among a large set of SM final states. In order to keep our analysis as simple as possible, we adopt a benchmark scenario that is as clean as possible in terms of its collider signatures. Once the proof of concept for the methods we present has been established, additional analyses can be performed to adapt these methods to more challenging final states as well. To be specific, we choose p_1 and p_2 to be same-flavor, opposite-charge leptons, such as $\mu^\pm\text{-}\mu^\mp$, and we choose p_3 to be a photon. This final state can arise, for example, if Y is a muon partner, while X , Z , and χ are partners to gauge bosons. We emphasize that we are not promoting this to be a particularly plausible scenario of beyond-the-SM physics. It is meant only to provide a relatively clean example of our

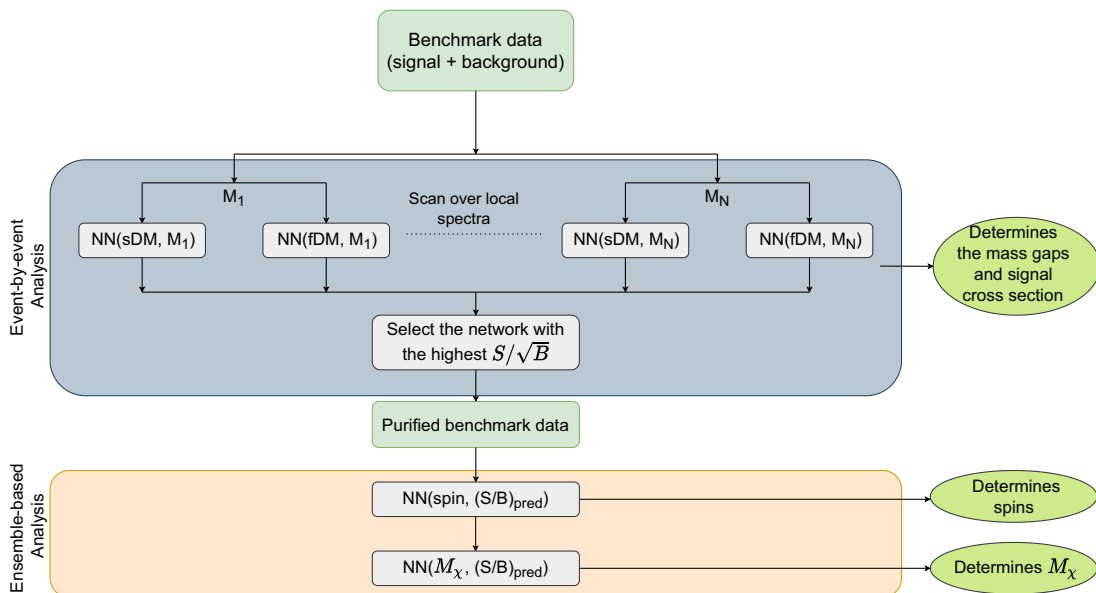


FIG. 3. Flow chart detailing the overall procedure for discovery, mass gap determination, spin determination, and overall mass scale determination.

benchmark event topology, to study the effectiveness of the methods we are proposing. X , Z , and χ can all be fermions and Y a boson, or the other way around, and we, in fact, study both assignments to address the questions not only of discovery and mass measurement, but of spin determination as well.

With these choices, the final state we focus on in this study is $\mu^+\mu^-\gamma + \text{MET}$. Since there is no reason to expect a resonant structure in the $\mu^+\mu^-$ system in the signal, we impose a Z veto in the analysis to eliminate the leading backgrounds. We also veto any significant hadronic activity. With these choices, there are two SM backgrounds contributing to this final state. The first is $Z\text{-}\gamma$ pair production, with the Z decaying to taus, and both taus decaying to muons. The second is triboson production, such as $W^+W^-\gamma$ (with both W 's decaying to muons) and $ZZ^*\gamma$ (with the on-shell Z decaying invisibly and the off-shell Z decaying to muons). These are the backgrounds we include in our analysis.

Given the signal and the background described above, a problem that needs to be dealt with early on is the following: Since information about the masses of the new particles and their spins is not known prior to their discovery, we cannot train the neural networks using Monte Carlo events generated with the correct signal spectrum and the correct spin assignments. As a possible resolution to this problem, we study the effectiveness of “scanning” over the parameter space using a performance metric that quantifies discovery (S/\sqrt{B} in our case). A full scan over the parameter space is computationally very expensive, and we will perform only a local scan in order to

demonstrate that the true spectrum is at least a local maximum.

First, we perform the scan using networks trained solely on the final state momenta, which we will refer to as the DL variables. We show that, over the region of this local scan, the discovery metric is maximized for the networks trained using mass spectra for which the mass differences or, more precisely, the kinematic edges and end points in the three m_{ij} variables have the correct values. Then, we proceed to identify a set of HL variables that match the DNN. In particular, we show that performing the scan using the HL variables rather than the final state momenta results in an improved resolution for the mass gap measurement. As we will see, both the DL and the HL variable-based scans do not perform well in measuring the spins and the overall mass scale. We show that analyses based on ensemble methods using HL kinematic variables can be used to determine the spins and the overall mass scale. Figure 3 shows a schematic representation of our procedure.

This paper is organized as follows: We start with a short review of the relevant kinematic observables in Sec. II and a review of the relevant machine learning techniques in Sec. III. Then, in Sec. IV, we study the prospects for discovery and mass measurement relying on the final state momenta of the visible particles and a deep neural network trained on individual events. We then proceed to interpret the DNN in terms of HL variables in Sec. V, and we perform the second, ensemble-based stage of the analysis in order to determine the spins of the new particles (in Sec. VI) and their masses (in Sec. VII) much more accurately. We conclude in Sec. VIII.

II. REVIEW OF PHASE SPACE AND KINEMATIC OBSERVABLES

In this section, we review important properties of kinematic observables that are used in the analysis of the next few sections. Using the notation of Fig. 1, we label the four-momenta of the new particles in a given event as p_X^μ , etc., and those of the visible final states as p_1^μ , etc. Very generally, the distribution of events in phase space is given by

$$d\Gamma = d\Pi_{\text{PS}} |\mathcal{M}|^2, \quad (2.1)$$

where $d\Pi_{\text{PS}}$ stands for the phase space volume element and $|\mathcal{M}|^2$ for the amplitude squared. While the latter contains valuable information about the spins of the particles and the angular correlations of the final state particles, the information about the spectrum of the new particles and about the boundary of the kinematically available phase space is entirely contained in the phase space factor. As mentioned in the introduction, we consider two possible spin assignments for the new particles. In our analysis of discovery and mass measurement prospects, we consider both possibilities, and we look for commonalities as well as differences in the list of the optimal HL variables in these two cases.

The simplest way to characterize the data is, of course, through the momentum vectors $\vec{p}_{1,2,3}$ of the three visible final state particles and that of the transverse missing energy (MET). As usual in collider analyses, we express these vectors in terms of p_T , the pseudorapidity η (except in the case of the MET), and the azimuthal angle ϕ of the final state particles. Considering boost invariance along the beam direction, differences $\Delta\eta$ and $\Delta\phi$ between any two particles are very useful observables, as is the combination $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. The fully Lorentz-invariant observables include the pair invariant masses $m_{ij}^2 = (p_i^\mu + p_j^\mu)^2$, as well as the total invariant mass of all three visible particles m_{123}^2 . One less well-known Lorentz-invariant observable is Δ_4 , which we define below. This completes the list of kinematic observables that we consider in our paper.

The maximum values of the m_{ij}^2 and m_{123}^2 variables for this final state topology are well known [2,8,19,35,45,46] and often used for mass (difference) measurements in this event topology. They are given by

$$(m_{23}^2)_{\text{max}} = (M_Y^2 - M_Z^2)(M_Z^2 - M_X^2)/M_Z^2, \quad (2.2)$$

$$(m_{12}^2)_{\text{max}} = (M_X^2 - M_Y^2)(M_Y^2 - M_Z^2)/M_Y^2, \quad (2.3)$$

$$(m_{13}^2)_{\text{max}} = (M_X^2 - M_Y^2)(M_Z^2 - M_X^2)/M_Z^2, \quad (2.4)$$

$$(m_{123}^2)_{\text{max}} = \begin{cases} \frac{(M_X^2 - M_Y^2)(M_Y^2 - M_Z^2)}{M_Y^2} & \frac{M_X}{M_Y} > \frac{M_Y}{M_Z} \frac{M_Z}{M_X}, \\ \frac{(M_X^2 M_Z^2 - M_Y^2 M_X^2)(M_Y^2 - M_Z^2)}{M_Y^2 M_Z^2} & \frac{M_Y}{M_Z} > \frac{M_Z}{M_X} \frac{M_X}{M_Y}, \\ \frac{(M_X^2 - M_Z^2)(M_Z^2 - M_Y^2)}{M_Z^2} & \frac{M_Z}{M_X} > \frac{M_X}{M_Y} \frac{M_Y}{M_Z}, \\ (M_X - M_X)^2 & \text{otherwise.} \end{cases} \quad (2.5)$$

It is straightforward to see that these formulas are sensitive to differences of masses; however, there exists a flat direction along which the masses can be varied such that the end points of all m_{ij}^2 variables remain constant. The kinematic variable Δ_4 we are about to describe has been shown to be useful for measurements of the mass spectrum along this challenging direction [32–34].

For each decaying X particle, the momenta of the final states $p^{1,2,3}$ and χ can be represented as a point in four-body phase space. While it is not commonly used, there is an elegant description of four-body phase space [47] that is manifestly Lorentz invariant. Consider the 4×4 matrix \mathcal{Z}_{ij} whose elements are given by $p_i \cdot p_j$ (where we are taking χ to be the fourth particle). Define the functions Δ_i of these momenta as the coefficients in the characteristic polynomial of \mathcal{Z} , namely,

$$\text{Det}[\lambda I_{4 \times 4} - \mathcal{Z}] \equiv \lambda^4 - \lambda^3 \Delta_1 - \lambda^2 \Delta_2 - \lambda \Delta_3 - \Delta_4. \quad (2.6)$$

The kinematically allowed region in phase space for X decay can be shown [47] to be defined by the conditions $\Delta_i > 0$ for all $i = 1, \dots, 4$, with the boundary of the region defined by $\Delta_4 = 0$ (with $\Delta_{1,2,3}$ still positive). As already mentioned, Δ_4 has been shown to be a powerful observable for analyzing this decay chain, both for discovery and for mass measurement purposes [32–34,48,49]. There is one subtlety that needs to be mentioned. Using the definition above, calculating Δ_4 requires knowledge of all final state momenta, including that of χ , which is invisible. However, if the masses $\{M_X, M_Y, M_Z, M_\chi\}$ in the spectrum are assumed known, then all dot products $p_i \cdot p_\chi$ (and, therefore, Δ_4) can be calculated from only the visible particle momenta, by using the on-shell conditions for the intermediate particles.

The reason that Δ_4 is such a useful variable is that the volume element of four-body phase space, expressed in the differentials dm_{ij}^2 , is given by $\Delta_4^{-1/2}$, up to a constant and an energy-momentum-conserving delta function. As a result, in the Lorentz-invariant coordinates m_{ij}^2 , the distribution of signal events is strongly clustered near $\Delta_4 = 0$. Background events, on the other hand, do not arise from X decays, and their Δ_4 distribution has no similar sharp feature at $\Delta_4 = 0$. As a result, if the spectrum were somehow known, then an excess of signal events over background can be easily discovered due to the sharp peak in the Δ_4 distribution near zero. Of course, discovery must

precede a knowledge of the spectrum; however, Δ_4 can be used to accomplish both goals simultaneously. In principle, one can scan over the possible spectra $\{M_X, M_Y, M_Z, M_\chi\}$ and look for an excess near $\Delta_4 = 0$. This feature will be most significant when the correct spectrum is used, so the presence of the excess will serve both as a discovery variable and as a way to measure the unknown particle masses. Of course, in practice, it is computationally prohibitive to perform a scan in all four mass variables. Fortunately, the well-known kinematic edges and end points in the m_{ij}^2 distributions already provide good sensitivity for the mass differences, and, therefore, they can be used first, leaving only the overall mass scale undetermined (parametrized by M_χ , say) along the flat direction. Then, one can perform a one-dimensional scan over M_χ and use Δ_4 to fix the spectrum completely. Our analysis in the rest of this paper will demonstrate that neural network techniques based on Δ_4 do indeed result in a high-precision measurement of the masses along the flat direction.

III. REVIEW OF MACHINE LEARNING TOOLS

For the results presented here, all neural networks were implemented using the Keras [50] package with the TensorFlow [51] back end. Wherever we implement neural networks in our analysis, we provide details about the number of nodes, layers, and activation functions within the respective section.

In the initial phase of our analysis focused on discovery, we implement fully connected deep networks as binary classifiers, the two target classes being the signal and the background. We use the binary cross-entropy as the loss function.

Unlike the first stage of the analysis described in Secs. IV and V, where the neural network analyzes one event at a time and assigns each event an output in the range $[0, 1]$, where 0 corresponds to background and 1 corresponds to signal, we will see in Secs. VI and VII that the spin determination and the measurement of the overall mass scale are more challenging, and an ensemble-based analysis is required. In the ensemble-based analysis, the entire dataset is considered as a single input. We also identify the HL variables whose distributions are particularly effective in separating the different hypotheses. Histograms in these variables are then used for the training and evaluation of the networks. Specific details about the histogram binning are provided in the respective sections of the paper.

In our case, postdiscovery, the determination of the mass spectrum reduces to determining a single mass value along the flat direction, which we parametrize by M_χ . We treat this as a regression problem, and we use the mean squared error loss function for training the network.

The average decision ordering (ADO) metric introduced in Ref. [52] is very useful for quantifying the correlation between two functions, one of which may be the neural network output and the other an analytic function defined in terms of the inputs. ADO is constructed to quantify the degree to which two functions f and g rank pairs of events belonging to the two classes A and B in the same order. We use the following discrete version of the ADO:

$$\text{ADO}[f, g] = \sum_{x \in A} \sum_{x' \in B} \Theta((f(x) - f(x'))(g(x) - g(x'))), \quad (3.1)$$

where Θ is the Heaviside function.

IV. ANALYSIS BASED SOLELY ON FINAL STATE MOMENTA

In this section, we start by studying how well a “black box” DNN can discriminate between signal and background based on detector-level variables alone, by which we mean the momenta of the final state particles, in other words, without using any guidance in the form of human-level variables. Since the DL variables represent all the available information at the detector, the DNN learns an approximation of the optimal discriminating function, namely, the ratio of signal and background distributions. In the next section, we will identify a small combination of HL variables that are sufficient to approximate the optimal discriminator, relying on ADO as a metric. Finally, in Secs. VI and VII, we will combine the strength of these HL variables with ensemble-based analysis to tackle the more challenging problems of spin determination and the measurement of the overall mass scale. Figure 3 provides a visual overview of the stages of the analysis.

We treat the signal cross section essentially as a free parameter varying in a range consistent with the new particles having electroweak couplings and masses of a few hundred GeV. In order to work on a specific example, we choose the following signal spectrum as a benchmark:

$$\begin{aligned} M_X &= 390 \text{ GeV}, & M_Y &= 360 \text{ GeV}, \\ M_Z &= 330 \text{ GeV}, & M_\chi &= 300 \text{ GeV}. \end{aligned} \quad (4.1)$$

This will be denoted as the *truth spectrum* for the rest of the paper. For the background cross sections, we use the leading-order values obtained from Monte Carlo simulations.

For the spin assignments of the new particles, we work with two possibilities, denoting these as the fDM and sDM models, based on whether χ is a fermion or boson (see Fig. 2). In these two models, the new particles are taken to be as follows.

- (i) *fDM*.— X , Z , and χ are neutral fermions, while Y is a charged scalar.

- (ii) *sDM*.— X , Z , and χ are neutral scalars, while Y is a charged fermion.

We begin by describing the details of Monte Carlo methods we use in our analysis.

A. Monte Carlo methods and selection cuts

As mentioned in the introduction, the production mechanism of interest to us is $pp \rightarrow X\chi$. We take the production to proceed via a heavy t -channel mediator Δ [which is a scalar (fermion) for the fDM (sDM) signal model]. We take this mediator to couple to up-type quarks. This choice is mostly arbitrary and motivated by the fact that the constraints on new physics from flavor violation are weaker compared to new particles coupling to down-type quarks. Since the analysis below is based entirely on the decay of X , we treat the signal cross section as a free parameter, with our results parametrized by this parameter. For the background, we generate $pp \rightarrow \mu^+\mu^-\gamma\nu_i\bar{\nu}_i$ and $pp \rightarrow \mu^+\mu^-\gamma\nu_i\bar{\nu}_i\nu_l\bar{\nu}_l$ events, with all possible neutrino flavor combinations. These contain the diboson and triboson processes (including off-shell W/Z bosons) discussed in the introduction. Signal and background events are all generated using MadGraph [53].

The final state can be described by nine momentum components of the three observable particles. We remind the reader that any significant hadronic activity will be vetoed—as a result, the MET does not carry additional information to the momenta of the visible particles. There is a combinatorial ambiguity in the final state, since p_1 can be a μ^+ and p_2 a μ^- or the other way around (see Fig. 1). Therefore, it is useful to denote the final state momenta by their charges, namely, p_+ , p_- , and p_γ .

In Monte Carlo generation, we use relatively loose cuts, demanding only $p_{T,\gamma} > 4$ GeV to avoid singularities in the matrix element, and we impose the acceptance cut of $|\eta| < 2.5$. In order to simulate the detector energy resolution for muons and photons, we use the parameters estimated by ATLAS for the high-luminosity run. For muons, these are given by [54]

$$\frac{\sigma(E)}{E} = 1.61 \times 10^{-2} \oplus 2.76 \times 10^{-3} \text{ GeV}^{-1/2} \sqrt{E}, \quad (4.2)$$

and for photons they are given by [55]

$$\frac{\sigma(E)}{E} = 9.84 \times 10^{-3} \oplus \frac{9.41 \times 10^{-2} \text{ GeV}^{1/2}}{\sqrt{E}} \oplus \frac{1.19 \text{ GeV}}{E}. \quad (4.3)$$

There are two (nonprescaled) ATLAS triggers that are relevant for our final state: a dimuon trigger requiring $p_T > 14$ GeV for both muons and a $\mu\mu\gamma$ trigger requiring $p_T > 10$ GeV for both muons and $p_T > 15$ GeV for the photon. We include both trigger paths in our event selection. Since our main interest in this paper is in signal spectra with small mass splittings, the invariant mass values

take on relatively small values, and a Z veto can be applied to the $\mu^+\mu^-$ system to eliminate the large backgrounds with on-shell Z 's. In order to also avoid the increased background at low invariant mass due to photon conversions, we impose the selection cut

$$15 \text{ GeV} < m_{+-} < 65 \text{ GeV}. \quad (4.4)$$

After all selection cuts, the background cross section is given by 2.88 fb.

We will work with event samples that correspond to the full HL-LHC luminosity of 3000 fb^{-1} , and we report on the performance of the analysis as a function of the signal strength, quantified by S/B , where S and B are defined as the number of signal and background events after the selection cuts listed above.

B. Measurement of mass differences

We start our analysis using a DNN trained on DL variables. As mentioned before, the questions of discovery and mass measurement are linked, since the DNN needs to be trained without prior knowledge of the spectrum. We will see below that the first phase of our analysis is efficient in measuring the mass differences between the new particles but not the overall mass scale.

As mentioned in Sec. I, it is beyond our computational resources to scan over all possible spectra. We settle for a less ambitious goal of at least demonstrating that the correct mass differences present an optimal point in a *local* scan of a “testing spectrum” (\mathbf{M}_0) given by

$$\mathbf{M}_0 = (M_X, M_Y, M_Z, M_\chi) = (691, 660, 631, 600) \text{ GeV}, \quad (4.5)$$

which has the correct mass differences but is displaced from the truth spectrum [Eq. (4.1)] along the flat direction, which corresponds approximately to the vector $(1, 1, 1, 1)$ in the (M_X, M_Y, M_Z, M_χ) space. We set up the local scan around this testing spectrum, using the following orthogonal basis:

$$\begin{aligned} \mathbf{v}_1 &= \frac{1}{2}(0, 1, 0, -1), & \mathbf{v}_2 &= \frac{1}{2}(1, 0, -1, 0), \\ \mathbf{v}_3 &= \frac{1}{2}(1, -1, 1, -1), & \mathbf{v}_4 &\approx (1, 1, 1, 1), \end{aligned}$$

where \mathbf{v}_4 is taken along the true flat direction, which deviates slightly from the vector $(1, 1, 1, 1)$. The spectra over which we perform the scans are parametrized as

$$\mathbf{M} = \mathbf{M}_0 + (\alpha\mathbf{v}_1 + \beta\mathbf{v}_2 + \gamma\mathbf{v}_3 + \delta\mathbf{v}_4). \quad (4.6)$$

For each of the spectra \mathbf{M} , we generate two sets of Monte Carlo signal events: one with the spins assigned according to the fDM hypothesis and the other according to

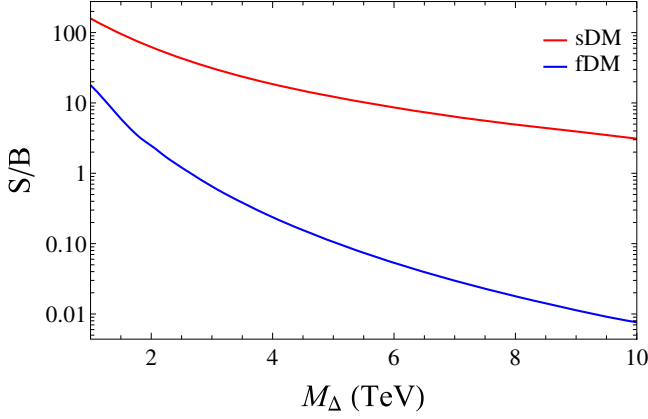


FIG. 4. Cross sections for the sDM (red line) and fDM (blue line) signal (at the truth spectrum) as a function of the mediator mass. All relevant couplings are taken to be 1.

the sDM hypothesis. Then, we train an ensemble of DNNs, two for each M , corresponding to the two possible spin assignments, to distinguish the respective signal from the background. Based on the events used to train the network, we refer to these networks as the fDM network and the sDM network, respectively.

Each DNN is made up of three hidden layers containing 200 hidden nodes. Nodes in the intermediate layers are assigned ReLU activation functions, and the output node is assigned the sigmoid activation function. The input layer contains nine nodes corresponding to the $3 \times 3 = 9$ observable momentum components. Additionally, we implement an early stopping monitor, which has a patience of 20 epochs on the validation set. Each network is trained using 1 000 000 events each for the signal and the background.

Once the DNNs are trained, we study their output on the benchmark sample (signal at the truth spectrum plus background). In order to infer the dependence of the performance on the signal cross section, we list our results below for three values of S/B , namely, 1.0, 0.1, and 0.01.

In order to provide context for these values of S/B , Fig. 4 shows how S/B depends on the mass of the mediator (M_Δ) in the sDM and fDM spin assignments, with the couplings at all vertices in Fig. 2 taken to have the numerical value 1.

We quantify the performance by $\epsilon_S/\sqrt{\epsilon_B}$, defined as the enhancement in S/\sqrt{B} after applying a cut on the output of the network, compared to the selection cuts in Sec. IV A. The choice of $\epsilon_S/\sqrt{\epsilon_B}$ (as opposed to ϵ_S/ϵ_B) as the performance metric is motivated by the expectation that statistical uncertainties will dominate over systematic ones.

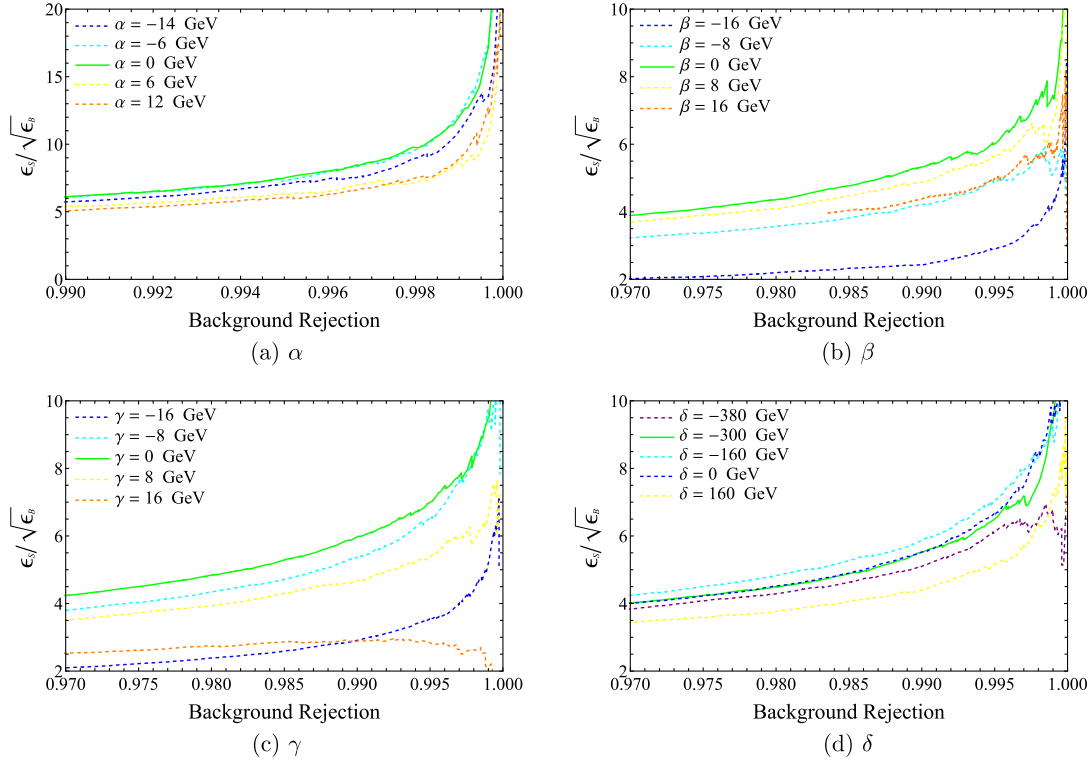


FIG. 5. For the local mass scan around the testing spectrum with $S/B = 1.0$, the parameters α , β , and γ favor the correct mass differences in the spectrum, and stronger cuts give rise to better performance. In contrast, the scan along the flat direction shows no preference for the true value of the overall mass scale. The discovery cut we impose corresponds to a background rejection of 0.995. In each panel, the solid green contour corresponds to the true value of the parameter.

In Fig. 5, we plot $\epsilon_S/\sqrt{\epsilon_B}$ for $S/B = 1.0$. It can be seen that the performance continues to improve toward stronger cuts on the network output. Here, we plot only the metric with the correct spin assignment for simplicity. However, the trend remains the same even if the incorrect spin assignment is used. In order to preserve sufficient signal statistics for the later stages of the analysis, in each network, we apply a cut on the network output that results in $\epsilon_B = 0.005$ (this tends to correspond to $\epsilon_S \approx 0.5$ for the best-performing networks in our analysis). Hereafter, this will be referred to as the “discovery cut.” The value of the discovery cut is inferred from Monte Carlo samples at each mass point. With the discovery cut, the minimum value of S/B such that the signal statistical significance is boosted above 5σ with the discovery cut is $(S/B)_{\min} \sim 6.7 \times 10^{-3}$. We will, therefore, focus on the range $S/B > 0.01$ for the rest of our analysis.

We perform one-dimensional scans using the parametrization of Eq. (4.6). The winning mass hypothesis is chosen to be the one that results in the highest $\epsilon_S/\sqrt{\epsilon_B}$ on the benchmark sample, with the discovery cut, setting ϵ_B to 0.005. To account for statistical uncertainties, we run 50 iterations of the scan. From the example scan in Fig. 5, it can be seen that this procedure results in an accurate measurement of the parameters α , β , and γ , thereby fixing the mass gaps. In contrast, we find that the $\epsilon_S/\sqrt{\epsilon_B}$ metric (at a fixed background rejection) fluctuates randomly as δ is varied, and, therefore, the value of δ maximizing the metric is uncorrelated with the true value.

To most efficiently utilize our computing resources, we limit the local mass scan to the following range around the testing spectrum: $\alpha, \beta, \gamma \in (-20, 20)$ GeV and $\delta \in (-500, 500)$ GeV, with spacings of $\Delta\alpha = \Delta\beta = \Delta\gamma = 0.5$ GeV and $\Delta\delta = 20$ GeV. Note that, with these choices, the mass hierarchy $M_X > M_Y > M_Z > M_\chi$ is automatically preserved.

Table I summarizes the results of the scans (averaged over the scan iterations). As expected, we find that the mass

TABLE I. The average values and uncertainties of the scan parameters obtained over 50 scan iterations along each direction (all numbers in GeV).

S/B	α	β	γ
(a) sDM			
1.0	-1.6 ± 1.4	5.5 ± 0.4	1.5 ± 0.3
0.1	-3.7 ± 3.3	4.7 ± 2.2	-0.1 ± 1.9
0.01	-4.0 ± 9.2	3.4 ± 5.7	-0.4 ± 4.5
S/B	α	β	γ
(b) fDM			
1.0	-12.4 ± 3.4	4.8 ± 1.1	-4.8 ± 2.2
0.1	-10.1 ± 6.0	4.1 ± 2.3	-3.2 ± 3.0
0.01	-9.0 ± 9.9	4.0 ± 8.0	-2.4 ± 5.5

differences are determined to a precision of a few GeV by this procedure, while the overall mass scale is left essentially unconstrained. Note that for high values of S/B there is a bias in the results for α , β , and γ . As we will see in the next section, when HL variables are used, these biases are eliminated.

V. IDENTIFICATION OF OPTIMIZED VARIABLES

After the DNN analysis based on the DL inputs, our next goal is to search for HL variables which match the performance of this analysis. We employ the ADO-guided method prescribed in Ref. [52] for identifying the set of HL variables that have the highest correlation with respect to the DL network. The definition of ADO was given in Sec. III. For our purposes here, the ADO is calculated from the fraction of pairs of events (one drawn from the signal generated at the testing spectrum and the other drawn from the background) which are ranked in the same order by both the deep network operating on the DL inputs and a given set of HL variables.

To summarize the method presented in Ref. [52], one picks the HL variable with the highest ADO with respect to the DL network (say, f_1) in the first iteration. In the next iteration, only the pairs of events for which f_1 and the DL network result in dissimilar orderings are considered.

TABLE II. HL variables with the leading ADOs for the sDM and the fDM networks. The ADOs have been computed over pairs of signal events at the testing spectrum and the background. We consider 20 000 events each of signal and background.

Variable	ADO
(a) sDM network	
MET	0.835
Δ_4	0.812
p_{T_γ}	0.724
p_{T_+}	0.702
p_{T_-}	0.699
$m_{-\gamma}$	0.631
$m_{+\gamma}$	0.631
$m_{+-\gamma}$	0.622
m_{+-}	0.568
Variable	ADO
(b) fDM network	
MET	0.880
Δ_4	0.827
p_{T_+}	0.739
p_{T_-}	0.721
p_{T_γ}	0.711
$m_{-\gamma}$	0.617
$m_{+\gamma}$	0.632
m_{+-}	0.558
$m_{+-\gamma}$	0.552

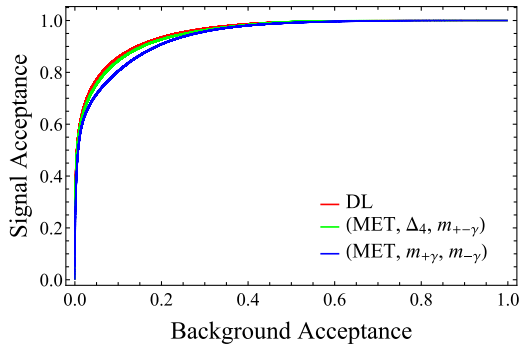
TABLE III. Triplets of HL variables with and without Δ_4 having the highest AUCs over the benchmark dataset. For simplicity, we list only the numbers for when the training dataset uses the correct spin model.

Variables	AUC
DL	0.9686
(a) sDM	
(MET, Δ_4 , $m_{+-\gamma}$)	0.9618
(MET, Δ_4 , m_{+-})	0.9572
(MET, Δ_4 , $m_{+\gamma}$)	0.9401
(MET, m_{+-} , $m_{+-\gamma}$)	0.9448
(MET, $m_{+\gamma}$, $m_{-\gamma}$)	0.9401
(MET, $m_{+-\gamma}$, $m_{+\gamma}$)	0.9361
(b) fDM	
(MET, Δ_4 , $p_{T\gamma}$)	0.9505
(MET, Δ_4 , $m_{+-\gamma}$)	0.9492
(MET, Δ_4 , m_{+-})	0.9445
(MET, $m_{+\gamma}$, $m_{-\gamma}$)	0.9285
(MET, m_{+-} , $m_{+-\gamma}$)	0.9265
(MET, $m_{+\gamma}$, $m_{+-\gamma}$)	0.9187

The second HL variable f_2 is then picked to be the one with the highest ADO over these subset of pairs and so on. We terminate the process when the additional HL variable does not lead to a significant improvement in the area under the ROC curve (AUC), where ROC stands for receiver operating characteristic.

As we described in Sec. II, we will consider the following list of HL variables:

- (i) the transverse momenta of the three visible final state particles, $(p_{T+}, p_{T-}, p_{T\gamma})$;
- (ii) the transverse missing energy, MET;
- (iii) the invariant masses of the three visible final state particle pairs, $(m_{+-}, m_{+\gamma}, m_{-\gamma})$;



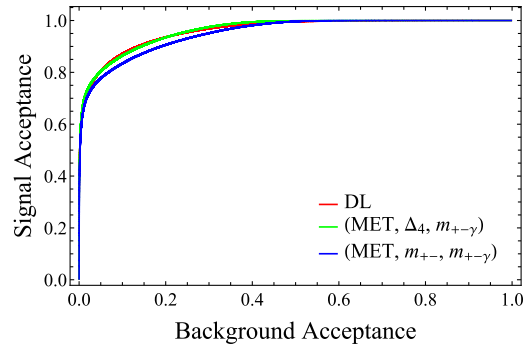
(a) sDM

- (iv) the total invariant mass of the three visible final state particles, $m_{+-\gamma}$;
- (v) Δ_4 computed assuming the testing spectrum; and
- (vi) $\Delta R_{ij} = \sqrt{\Delta\eta_{ij}^2 + \Delta\phi_{ij}^2}$ for all pairs of final state particles.

Tables II(a) and II(b) show the ADOs of individual HL variables for the two spin models generated at the testing spectrum. The AUC that is obtained by combining triplets of these variables is listed in Table III. The HL networks have three hidden layers containing 50 nodes each. The combination of the variables (MET, Δ_4 , and $m_{+-\gamma}$) is the most effective when both spin models are considered, and it effectively matches the AUC of the DNN based on DL inputs. We also illustrate this visually in Fig. 6, showing the performance of the HL variables (MET, Δ_4 , and $m_{+-\gamma}$) (in green) and the DL-based neural network (in red). For comparison, we also show the leading HL triplet that does not include Δ_4 (in blue). Our results confirm the power of Δ_4 as a kinematic variable in analyzing this decay chain, in accordance with prior studies. In Fig. 7, we show the distributions of the HL variables for the two signal models and for the background.

Next, we turn our attention to studying the performance of the HL variables for determining the spectrum, namely, fixing the parameters α , β , γ , and δ . Figure 8 is the counterpart of Fig. 5, made using the leading triplet of HL variables. Similarly, Table IV is the counterpart of Table I.

Given the relatively large dimensionality of the DL inputs, combined with finite training samples and network parameters, it is hard for the DL network to converge toward the global minimum of the loss function. Within the limitations of our analysis, we find that identification of the optimal set of HL variables leads to a better convergence toward the global minimum. This is evident from comparing the resolution for mass differences (Tables I and IV). The overall mass scale, however, still remains unresolved. In the next two sections, we will further improve the performance of the neural networks in order to determine



(b) fDM

FIG. 6. ROC curves for both spin models, for the DNN trained DL inputs (in red) and the HL variables (in green). For comparison, we also present the performance of the leading triplet of HL variables that does not contain Δ_4 (in blue).

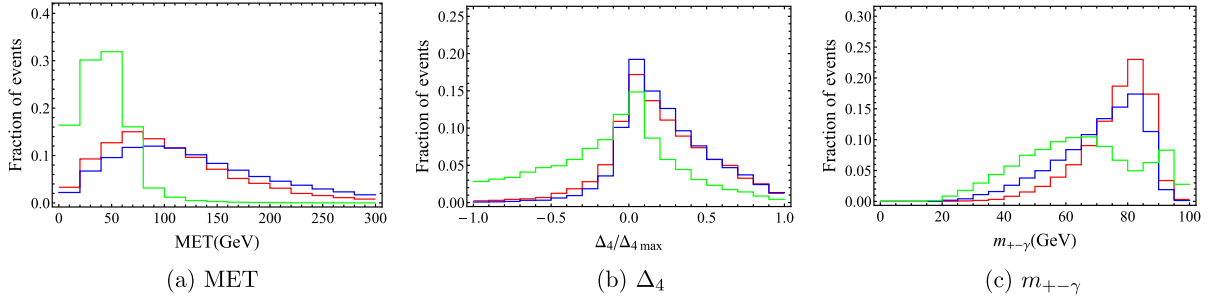


FIG. 7. Distributions of the leading triplet of HL variables for pure benchmark signal in the sDM (red line) and fDM (blue line) models and for background (green line). Δ_4 is calculated using the testing spectrum.

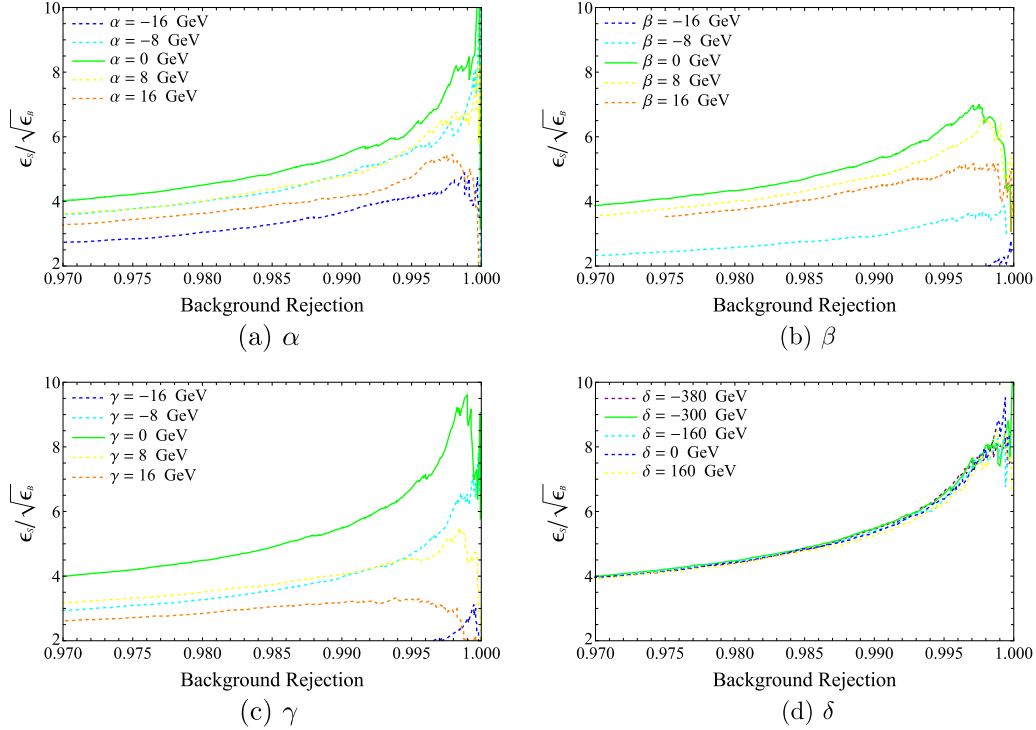


FIG. 8. Local scans around the testing spectrum as in Fig. 5 but using the leading triplet of HL variables.

TABLE IV. The average values and uncertainties of the scan parameters (all numbers in GeV), as in Table I, using the leading triplet of HL variables.

S/B	α	β	γ
(a) sDM			
1.0	-0.7 ± 0.3	1.2 ± 0.5	-0.2 ± 0.2
0.1	-0.6 ± 1.3	1.1 ± 0.9	-0.5 ± 0.8
0.01	2.3 ± 6.0	2.0 ± 2.7	-0.5 ± 2.9
S/B	α	β	γ
(b) fDM			
1.0	0.0 ± 0.8	1.3 ± 0.3	-0.6 ± 0.8
0.1	0.0 ± 1.5	0.9 ± 0.8	-0.8 ± 0.8
0.01	1.6 ± 7.1	3.0 ± 3.2	-1.3 ± 5.4

the spins of the new particles and measure the overall mass scale at a much higher precision. In order to achieve this goal, we will first purify the signal in the data by passing the data through the HL classifier network (trained at the testing spectrum) and apply the discovery cut, which eliminates 99.5% of background events in each network. Since this corresponds to roughly 50% efficiency for signal events, it results in an enhancement in S/B by a factor of ~ 100 .

We can also estimate S/B , without prior knowledge of the overall mass scale. This will be used in the second stage of the analysis, where we employ ensemble-based methods. We have

$$(S/B)_{\text{est}} = \frac{\epsilon_{S+B} - \epsilon_B}{\epsilon_S - \epsilon_{S+B}}, \quad (5.1)$$

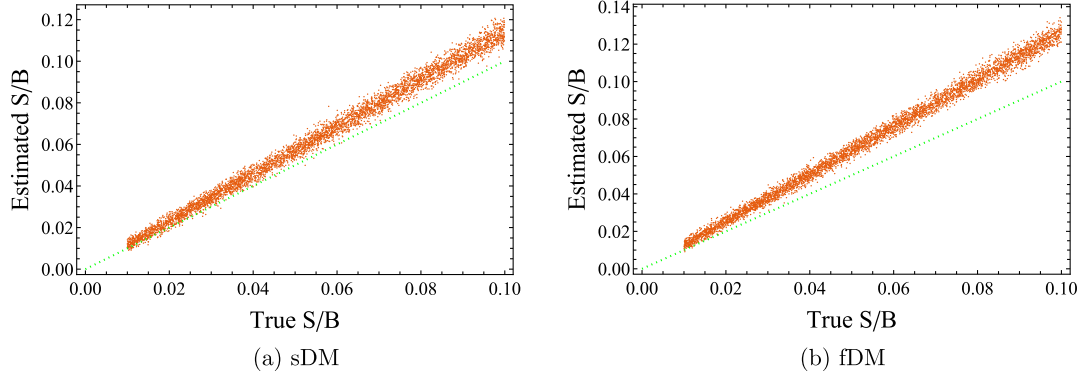


FIG. 9. The estimated value of S/B for the benchmark sDM and fDM data samples (red points), plotted against the true S/B (green curve), based on the formula in Eq. (5.1). The results were obtained over 5000 pseudoexperiments in each case, using the HL networks.

where ϵ_S is evaluated in the network with the highest $\epsilon_S/\sqrt{\epsilon_B}$ after applying the discovery cut, using signal events generated at the testing spectrum. ϵ_{S+B} is the efficiency of the discovery cut on the actual data. In other words, ϵ_B and ϵ_S are parameters obtained from simulations, while ϵ_{S+B} is measured from data. Figure 9 shows the accuracy of this procedure.

VI. SPIN DETERMINATION

We now turn our attention to determining the spins of the new particles. We treat this as a binary classification problem. We start by introducing an ensemble-based analysis method which will also allow us to determine the overall mass scale in the next section. With the ensemble-based method, we assess the performance of DL variables first, and we then identify the HL variables that result in a similar performance.

A. Ensemble-based method

Consider binary classification of ensembles containing N events each, $\{x_i\} \in \mathbb{E}^N$, where each event $x_i \in \mathbb{E} = \mathbb{R}^k$. In our case, the ensemble corresponds to the entire dataset. A neural network performing this classification task must have $k \times N$ nodes in the input layer. Also, one must take into account the permutational invariance of the input vectors (in other words, the ordering of events within an ensemble should not affect the outcome). This can be achieved either by having a special architecture for the neural network or by considering all the allowed permutations of the input vectors during training. Given the size of the event samples in our case and the dimensionality of the final state ($k = 9$ for DL inputs), we will not attempt to do this. We instead apply the method introduced in Ref. [56], to build an ensemble-based classifier based on an event-by-event classifier. We first build a simple event-by-event classifier operating on events x_i represented by the complete set of DL variables. Then, using the output of the

event-by-event classifier $y(x_i)$, the ensemble-based classification function $y_N(\{x_i\})$ is generated:

$$y_N(\{x_i\}) = \frac{\prod_i y(x_i)}{\prod_i y(x_i) + \prod_i (1 - y(x_i))}. \quad (6.1)$$

B. Performance with DL variables

The event-by-event network is made up of three hidden layers containing 100 nodes each. We use 1 000 000 events of each spin model to train the networks. Since our analysis so far has not allowed us to measure the overall mass scale, we continue generating the training samples at the testing spectrum. The network is trained using the labels $\hat{y}(x_i) = 1$ if $x_i \in \text{sDM}$ and $\hat{y}(x_i) = 0$ if $x_i \in \text{fDM}$. We use 100 000 ensembles of each spin model.

We remind the reader that we are performing the spin determination analysis after the first stage of the analysis described in the previous section has already been performed. With the amount of signal purification we gained by applying the discovery cut, even in the most conservative cases ($S/B \gtrsim 0.01$ needed for discovery), the signal fraction in the events passing the cut has been boosted to an $\mathcal{O}(1)$ number. In Fig. 10, we present the ROC curves for representative values of S/B , for an integrated luminosity of 3 ab^{-1} . As can be seen in that figure, the spin determination is very accurate and starts to degrade only near the lowest S/B values of interest.

C. Performance with HL variables

Next, we identify the HL variables that match the performance of the event-by-event DL spin determination network and then use those to perform the ensemble-based analysis in terms of these variables.

As before, we compute the ADOs of the HL variables (at the testing spectrum). The variables with the highest ADOs for spin determination are shown in Table V. Replacing the DL variables with (m_{+-}, MET) results in a similar

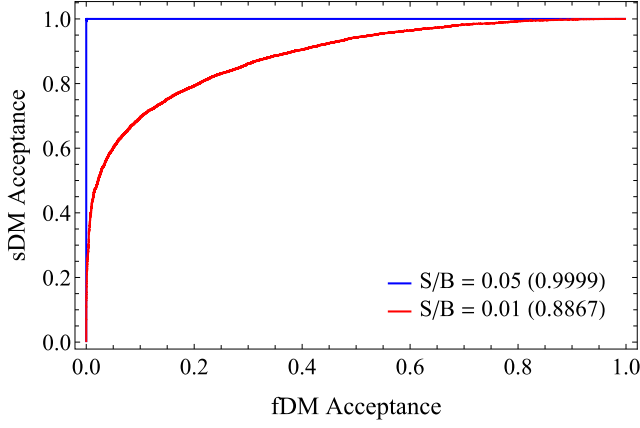


FIG. 10. ROC curves for the ensemble-based spin determination using DL variables. The given S/B numbers denote the ratio before the first stage of the analysis. The corresponding AUCs are given in brackets.

performance for event-by-event spin determination. m_{+-} is identified by the ADO method to be very effective in spin determination, matching the results presented in Ref. [36].

The HL network we use has two hidden layers containing 30 units each. Using the output of this event-by-event HL classifier network, we construct ensemble-based classifiers as we did for the DL variables. Figure 11 shows the performance of the HL ensemble-based classifiers. Comparing Figs. 10 and 11 shows that, with an ensemble-based analysis, just the two HL variables (m_{+-} and MET) are sufficient to match the performance of the DL network.

It is worth spending some time looking into how the spin information is encoded in the variables m_{+-} and MET. In Fig. 12, we show the distribution of events in these two variables within the sDM and fDM signal models after the discovery cut. Note that the discovery cut eliminates almost

TABLE V. (a) HL variables with the highest ADOs for spin determination. (b) Comparison of AUCs (for an event-by-event network) for the DL and HL inputs for spin determination. The values shown are computed for pure signal sDM and fDM events passing the discovery cut.

Variable	ADO
(a)	
m_{+-}	0.721
ΔR_{+-}	0.663
MET	0.656
$m_{+-\gamma}$	0.631
Variables	AUC
(b)	
DL	0.6727
m_{+-}	0.6487
(m_{+-} , MET)	0.6614

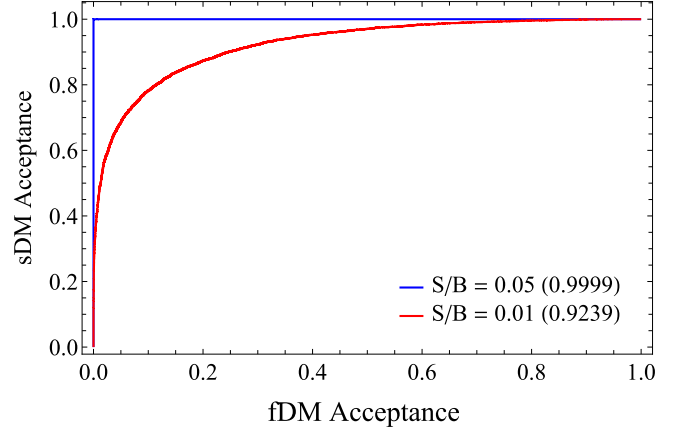


FIG. 11. The same as Fig. 10 but using the HL variables (m_{+-} and MET).

all events with $MET < 100$ GeV, as that region of phase space is background dominated. We also show in Fig. 13 the contours of the event-by-event HL network in the (m_{+-} , MET) space. The shape of these contours can be understood with the following observations. The matrix element squared of the decay process $X \rightarrow \mu^+ \mu^- \gamma \chi$ is proportional to a factor of m_{+-}^2 in the sDM model, but not in the fDM model, resulting in a significant difference in the m_{+-} distributions in the two models, which can also be seen in Fig. 12. In fact, at low m_{+-} , the network output is basically purely based on m_{+-} , as can be seen on the left side in Fig. 13. As m_{+-} approaches its maximum value, most of the kinetic energy from the decay in the X frame is taken by the muons, leaving the χ with little kinetic energy. As a result, there is a boost imbalance between this softer χ produced in the X decay and the one produced directly from the initial state and recoiling against the X (see Fig. 2). In this region, the MET and m_{+-} variables are correlated, which leads to the contours on the right side in Fig. 13 bending toward the diagonal.

D. Optimizing the cut thresholds

For the ensemble-based method described above, AUC quantifies the separation between the two spin models achieved by the output function $y_N(\{x_i\})$. However, we are ultimately interested in finding the optimal value of the cut on the output that maximizes the classification accuracy. Note that the HL variables of interest do not include Δ_4 ; it is, therefore, sufficient to use the correct mass gaps, and the overall mass scale is not needed. This being the case, we can use the testing spectrum for the determination of the optimal cut values.

For a given cut y'_N on the output function $y_N(\{x_i\})$, we define the overall accuracy rate (AR) as the sum of the accuracy rates of sDM and fDM ensembles, as a function of S/B . Then, scanning through the range of the output function, the cut y'_N that maximizes the overall AR is

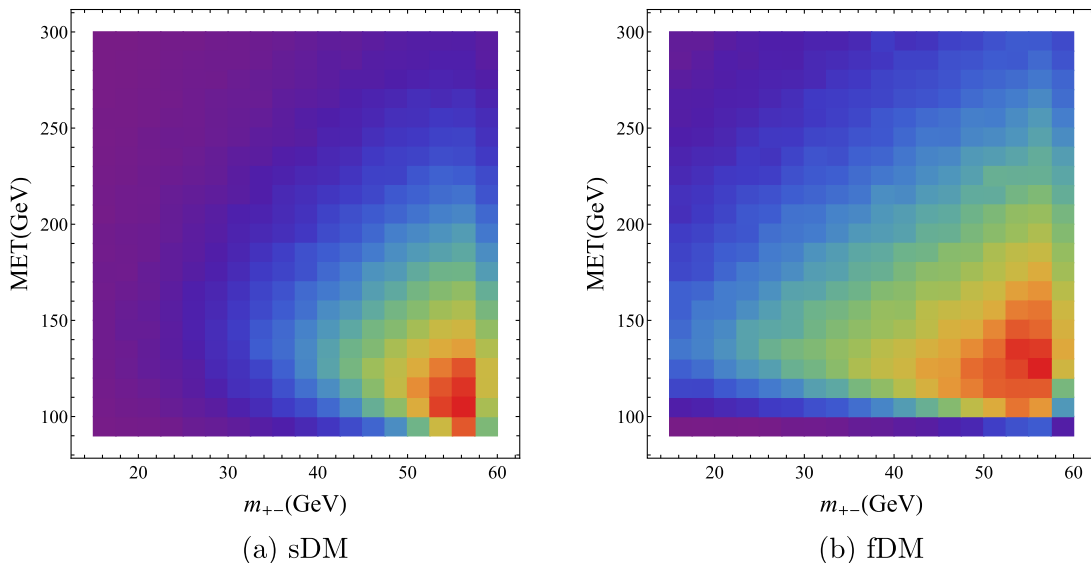


FIG. 12. Distributions of the sDM and fDM signal events after the discovery cut in the (m_{+-}, MET) plane.

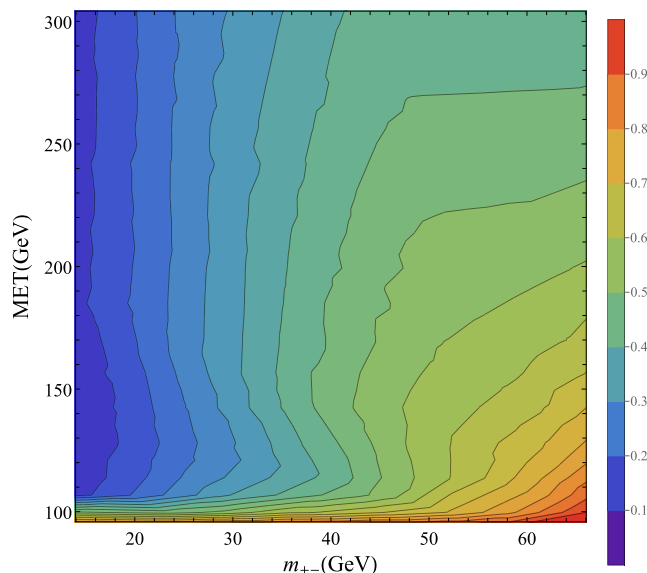


FIG. 13. Contours of the output of spin determination network based on the HL variables $(m_{+-}$ and MET). An output of 0 corresponds to fDM-like events, and an output of 1 corresponds to sDM-like events.

found. The accuracy rates for spin determination of the benchmark signal, using HL variables, are shown in Fig. 14.

VII. MASS SCALE DETERMINATION

Having fixed the mass differences in the earlier stage of the analysis (Sec. V), the spectrum hypotheses between which we are trying to distinguish at this stage are labeled in terms of a single variable, which we can take to be M_χ . It was shown in Refs. [32,34] that using Δ_4 as a kinematic

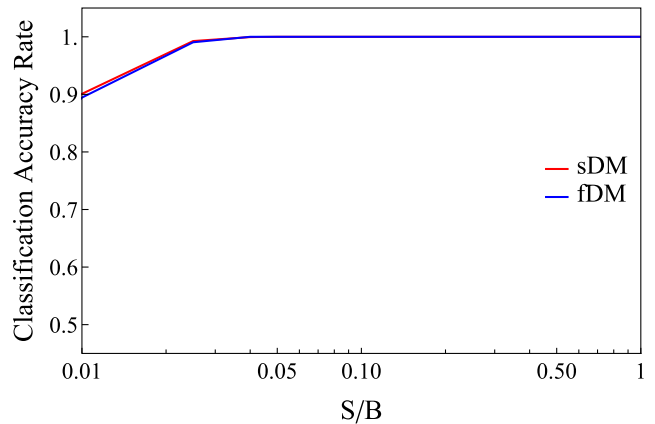


FIG. 14. Accuracy rates for the spin determination using the ensemble classifier built from the event-by-event HL network.

observable helps break the degeneracy along the flat direction and determine the overall mass scale, and we therefore expect Δ_4 to be a powerful variable when performing a neural-network-based analysis as well. Even so, we saw at the end of Sec. V that the overall mass scale still has a large uncertainty when analyzing the data event by event, even with Δ_4 as one of the HL variables. In this section, we combine the power of Δ_4 with an ensemble-based analysis to pin down the overall mass scale.

We remind the reader that a histogram of Δ_4 depends on not only the true value of M_χ in the data, but also the input value M_χ that is assumed when calculating Δ_4 . The core concept for the method we are about to present relies on training the neural network on the shape of this distribution as either the true or input value for M_χ is varied.

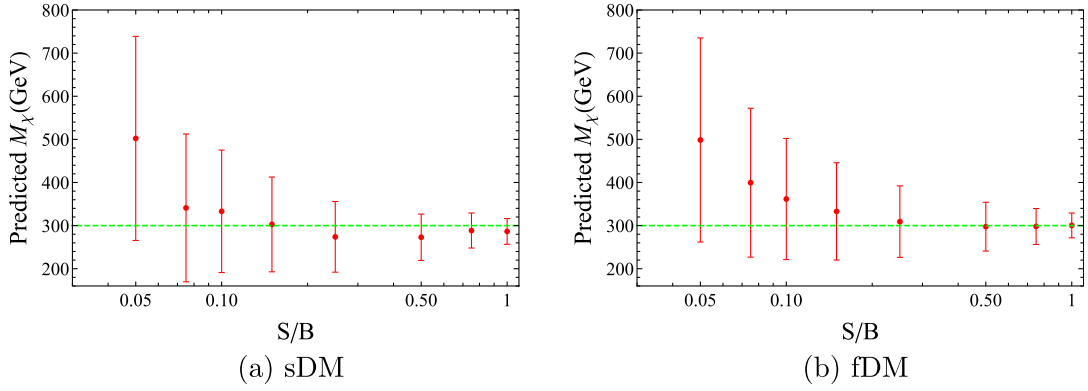


FIG. 15. The determined M_χ values and the uncertainties are plotted for $S/B \gtrsim 0.05$. Below this value, our methods cannot significantly narrow down the possible range of M_χ within our scan interval.

To determine the true value of M_χ , we specify a range $[M_1, M_2]$ that we expect it to lie in, and we transform each event x_i to a point in the following two-dimensional space, with M_1 and M_2 used as input values:

$$x_i \rightarrow (\Delta_4(x_i; M_1), \Delta_4(x_i; M_2)). \quad (7.1)$$

Any data sample with the true value M_χ is then converted into a two-dimensional scatter plot, and with an appropriate binning, into a pixellated heat map. These heat maps are used to train the neural network, as the true M_χ is varied between M_1 and M_2 , and M_χ being assigned as the output of the neural network during the training. In the testing phase, the output of the neural network is then taken as the measured value of M_χ .

To minimize bias, we choose the relatively broad mass range of (100,900) GeV for this last stage of our analysis. We use a step size of $\delta M_\chi = 20$ when varying M_χ . We have checked that a smaller step size does not result in an improvement of the accuracy of the final result.

To generate the training data, we construct the heat maps described above from the data samples for each value of m_χ and S/B of interest, after having applied the discovery cut (see Sec. IV B). We restrict the heat maps to the range $-\Delta_{4,\max} \leq \Delta_4 \leq \Delta_{4,\max}$ for each sample. Events lying beyond this range are discarded. We refine the resolution of the heat maps until a saturation in performance is reached in terms of the loss function. In our case, this is achieved for a bin size of $\Delta_{4,\max}/20$ along each of the two axes. Based on this number, we construct 41×41 pixel heat maps. The networks have 1681 nodes in the input layer and three hidden layers with 100, 300, and 100 nodes, respectively.

Having completed the training phase, we move on to the determination of M_χ from the data. In the testing phase, the trained network is run on the data, with the estimated S/B value obtained from Eq. (5.1) as input. In 1000 iterations for each value of S/B , the prediction for M_χ is obtained

from the network output in each run. In Fig. 15, we show the central value and standard deviation for M_χ obtained in this way over the S/B range of interest. The central value is found to be within 1σ of the true value of M_χ over the entire S/B range, with the precision approaching 20 GeV as S/B approaches 1.0.

VIII. CONCLUSIONS

We have applied machine learning techniques to optimize discovery sensitivity, spin determination, and mass measurement in the SUSY-like decay chain in Fig. 1. This event topology was chosen as a representative case for when signal cross sections are relatively low and the final state particles do not have high p_T due to a compressed signal spectrum. The decay chain is not long enough for algebraic reconstruction methods to be effective and yet long enough that commonly used one-dimensional distributions of Lorentz-invariant or boost-invariant kinematic observables do not capture the full amount of useful kinematic information.

In order to narrow down the parameter space, we started our analysis with a simple neural network that was effective in determining the mass gaps in the spectrum and in enhancing signal over background for the second stage of the analysis. We have identified the kinematic observables that match the performance of a “black-box” network in the first stage of the analysis, confirming the importance of the observable Δ_4 which had previously been proposed to be an effective observable for similar decay chains.

In the second, ensemble-based stage of the analysis, we were able to achieve a much higher accuracy in determining the spins of the new particles compared to an event-by-event analysis. Similarly, we were also able to measure the overall mass scale, a significant challenge for methods based on commonly used observables such as kinematic edges and end points, with an ensemble-based analysis. Once again, Δ_4 proved to play a significant role in

maximizing the precision of the measurement of the overall mass scale.

We point out to the reader that Δ_4 is an $\mathcal{O}(8)$ polynomial of the Lorentz-invariant pairs m_{ij} . To test the efficiency of Δ_4 for discovery, a scan can be performed over polynomial functions of m_{ij} . Specifically, the binary cross-entropy loss can be minimized with respect to the coefficients of the polynomials of m_{ij} . We will address such global optimization in future work.

The application of similar methods to more general event topologies and more challenging final state particles, to which traditional SUSY searches are not sensitive, is of great interest. Having provided a proof of concept with this

analysis of an idealized final state, we will take on these challenges in future work.

ACKNOWLEDGMENTS

The authors are grateful to Konstantin Matchev for helpful discussions. The research of C. K. and R. P. R. S. is supported by the National Science Foundation Grants No. PHY-1914679 and No. PHY-2210562. The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing High Performance Computing (HPC) resources that have contributed to the research results reported within this paper.

-
- [1] C. G. Lester and D. J. Summers, Measuring masses of semiinvisibly decaying particles pair produced at hadron colliders, *Phys. Lett. B* **463**, 99 (1999).
 - [2] B. C. Allanach, C. G. Lester, M. A. Parker, and B. R. Webber, Measuring sparticle masses in nonuniversal string inspired models at the LHC, *J. High Energy Phys.* **09** (2000) 004.
 - [3] W. S. Cho, K. Choi, Y. G. Kim, and C. B. Park, Measuring superparticle masses at hadron collider using the transverse mass kink, *J. High Energy Phys.* **02** (2008) 035.
 - [4] A. J. Barr, B. Gripaios, and C. G. Lester, Weighing wimps with kinks at colliders: Invisible particle mass measurements from endpoints, *J. High Energy Phys.* **02** (2008) 014.
 - [5] A. J. Barr, G. G. Ross, and M. Serna, The precision determination of invisible-particle masses at the LHC, *Phys. Rev. D* **78**, 056006 (2008).
 - [6] M. Burns, K. Kong, K. T. Matchev, and M. Park, Using Subsystem MT2 for complete mass determinations in decay chains with missing energy at hadron colliders, *J. High Energy Phys.* **03** (2009) 143.
 - [7] A. J. Barr, B. Gripaios, and C. G. Lester, Transverse masses and kinematic constraints: From the boundary to the crease, *J. High Energy Phys.* **11** (2009) 096.
 - [8] M. Burns, K. T. Matchev, and M. Park, Using kinematic boundary lines for particle mass measurements and disambiguation in SUSY-like events with missing energy, *J. High Energy Phys.* **05** (2009) 094.
 - [9] W. S. Cho, K. Choi, Y. G. Kim, and C. B. Park, Transverse mass kink, *AIP Conf. Proc.* **1078**, 274 (2009).
 - [10] I. Hinchliffe and F. E. Paige, Measurements in gauge mediated SUSY breaking models at CERN LHC, *Phys. Rev. D* **60**, 095002 (1999).
 - [11] M. M. Nojiri, G. Polesello, and D. R. Tovey, Proposal for a new reconstruction technique for SUSY processes at the LHC, in *Proceedings of the 3rd Les Houches Workshop on Physics at TeV Colliders* (2003), 12, [arXiv:hep-ph/0312317](https://arxiv.org/abs/hep-ph/0312317).
 - [12] K. Kawagoe, M. M. Nojiri, and G. Polesello, A New SUSY mass reconstruction method at the CERN LHC, *Phys. Rev. D* **71**, 035008 (2005).
 - [13] B. K. Gjelsten, D. J. Miller, P. Osland, and A. R. Raklev, Mass ambiguities in cascade decays, *Conf. Proc. C* **060726**, 1171 (2006), <https://cds.cern.ch/record/996599?ln=en>.
 - [14] H.-C. Cheng, J. F. Gunion, Z. Han, G. Marandella, and B. McElrath, Mass determination in SUSY-like events with missing energy, *J. High Energy Phys.* **12** (2007) 076.
 - [15] M. M. Nojiri and M. Takeuchi, Study of the top reconstruction in top-partner events at the LHC, *J. High Energy Phys.* **10** (2008) 025.
 - [16] H.-C. Cheng, D. Engelhardt, J. F. Gunion, Z. Han, and B. McElrath, Accurate Mass Determinations in Decay Chains with Missing Energy, *Phys. Rev. Lett.* **100**, 252001 (2008).
 - [17] B. Webber, Mass determination in sequential particle decay chains, *J. High Energy Phys.* **09** (2009) 124.
 - [18] H.-C. Cheng, J. F. Gunion, Z. Han, and B. McElrath, Accurate mass determinations in decay chains with missing energy. II, *Phys. Rev. D* **80**, 035020 (2009).
 - [19] K. T. Matchev, F. Moortgat, L. Pape, and M. Park, Precise reconstruction of sparticle masses without ambiguities, *J. High Energy Phys.* **08** (2009) 104.
 - [20] C. Autermann, B. Mura, C. Sander, H. Schettler, and P. Schleper, Determination of supersymmetric masses using kinematic fits at the LHC, [arXiv:0911.2607](https://arxiv.org/abs/0911.2607).
 - [21] I.-W. Kim, Algebraic Singularity Method for Mass Measurement with Missing Energy, *Phys. Rev. Lett.* **104**, 081601 (2010).
 - [22] T. Han, I.-W. Kim, and J. Song, Kinematic cusps: Determining the missing particle mass at colliders, *Phys. Lett. B* **693**, 575 (2010).
 - [23] Z. Kang, N. Kersting, S. Kraml, A. R. Raklev, and M. J. White, Neutralino reconstruction at the LHC from decay-frame kinematics, *Eur. Phys. J. C* **70**, 271 (2010).
 - [24] M. M. Nojiri, K. Sakurai, and B. R. Webber, Reconstructing particle masses from pairs of decay chains, *J. High Energy Phys.* **06** (2010) 069.
 - [25] Z. Kang, N. Kersting, and M. White, Mass estimation without using MET in early LHC data, [arXiv:1007.0382](https://arxiv.org/abs/1007.0382).

- [26] J. Hubisz and J. Shao, Mass measurement in boosted decay chains with missing energy, *Phys. Rev. D* **84**, 035031 (2011).
- [27] H.-C. Cheng, Z. Han, I.-W. Kim, and L.-T. Wang, Missing momentum reconstruction and spin measurements at hadron colliders, *J. High Energy Phys.* **11** (2010) 122.
- [28] B. Gripaios, K. Sakurai, and B. Webber, Polynomials, Riemann surfaces, and reconstructing missing-energy events, *J. High Energy Phys.* **09** (2011) 140.
- [29] T. Han, I.-W. Kim, and J. Song, Kinematic cusps with two missing particles I: Antler decay topology, *Phys. Rev. D* **87**, 035003 (2013).
- [30] T. Han, I.-W. Kim, and J. Song, Kinematic cusps with two missing particles II: Cascade decay topology, *Phys. Rev. D* **87**, 035004 (2013).
- [31] R. Franceschini, D. Kim, K. Kong, K. T. Matchev, M. Park, and P. Shyamsundar, Kinematic variables and feature engineering for particle phenomenology, [arXiv:2206.13431](https://arxiv.org/abs/2206.13431).
- [32] P. Agrawal, C. Kilic, C. White, and J.-H. Yu, Improved mass measurement using the boundary of many-body phase space, *Phys. Rev. D* **89**, 015021 (2014).
- [33] D. Debnath, J. S. Gainer, C. Kilic, D. Kim, K. T. Matchev, and Y.-P. Yang, Detecting kinematic boundary surfaces in phase space: Particle mass measurements in SUSY-like events, *J. High Energy Phys.* **06** (2017) 092.
- [34] D. Debnath, J. S. Gainer, C. Kilic, D. Kim, K. T. Matchev, and Y.-P. Yang, Enhancing the discovery prospects for SUSY-like decays with a forgotten kinematic variable, *J. High Energy Phys.* **05** (2019) 008.
- [35] B. K. Gjelsten, D. J. Miller, and P. Osland, Measurement of SUSY masses via cascade decays for SPS 1a, *J. High Energy Phys.* **12** (2004) 003.
- [36] L.-T. Wang and I. Yavin, Spin measurements in cascade decays at the LHC, *J. High Energy Phys.* **04** (2007) 032.
- [37] P. Meade and M. Reece, Top partners at the LHC: Spin and mass measurement, *Phys. Rev. D* **74**, 015010 (2006).
- [38] C. Athanasiou, C. G. Lester, J. M. Smillie, and B. R. Webber, Distinguishing spins in decay chains at the large hadron collider, *J. High Energy Phys.* **08** (2006) 055.
- [39] P. Baldi, P. Sadowski, and D. Whiteson, Searching for exotic particles in high-energy physics with deep learning, *Nat. Commun.* **5**, 4308 (2014).
- [40] S. Chang, T.-K. Chen, and C.-W. Chiang, Distinguishing W' signals at hadron colliders using neural networks, *Phys. Rev. D* **103**, 036016 (2021).
- [41] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski, and D. Whiteson, Parameterized neural networks for high-energy physics, *Eur. Phys. J. C* **76** (2016).
- [42] A. Badea, W. J. Fawcett, J. Huth, T. J. Khoo, R. Poggi, and L. Lee, Solving combinatorial problems at particle colliders using machine learning, *Phys. Rev. D* **106**, 016001 (2022).
- [43] A. Andreassen and B. Nachman, Neural networks for full phase-space reweighting and parameter tuning, *Phys. Rev. D* **101**, 091901(R) (2020).
- [44] J. Brehmer and K. Cranmer, Simulation-based inference methods for particle physics, [arXiv:2010.06439](https://arxiv.org/abs/2010.06439).
- [45] B. K. Gjelsten, D. J. Miller, and P. Osland, Measurement of the gluino mass via cascade decays for SPS 1a, *J. High Energy Phys.* **06** (2005) 015.
- [46] D. J. Miller, P. Osland, and A. R. Raklev, Invariant mass distributions in cascade decays, *J. High Energy Phys.* **03** (2006) 034.
- [47] N. Byers and C. N. Yang, Physical regions in invariant variables for n particles and the phase-space volume element, *Rev. Mod. Phys.* **36**, 595 (1964).
- [48] D. Debnath, J. S. Gainer, C. Kilic, D. Kim, K. T. Matchev, and Y.-P. Yang, Identifying phase space boundaries with Voronoi tessellations, *Eur. Phys. J. C* **76**, 645 (2016).
- [49] B. Altunkaynak, C. Kilic, and M. D. Klimek, Multidimensional phase space methods for mass measurements and decay topology determination, *Eur. Phys. J. C* **77**, 61 (2017).
- [50] F. Chollet *et al.*, Keras, <https://keras.io>, 2015.
- [51] M. Abadi *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems, 2015, Software available from <http://tensorflow.org>.
- [52] T. Faucett, J. Thaler, and D. Whiteson, Mapping machine-learned physics into a human-readable space, *Phys. Rev. D* **103**, 036020 (2021).
- [53] E. Boos *et al.*, Generic user process interface for event generators, in *Proceedings of the 2nd Les Houches Workshop on Physics at TeV Colliders* (2001), 9, [arXiv:hep-ph/0109068](https://arxiv.org/abs/hep-ph/0109068).
- [54] ATLAS Collaboration, Technical design report for the phase-II upgrade of the ATLAS muon spectrometer, Report No. CERN-LHCC-2017-017; ATLAS-TDR-026, CERN, Geneva, 2017.
- [55] ATLAS Collaboration, ATLAS liquid argon calorimeter phase-II upgrade: Technical design report, Report No. CERN-LHCC-2017-018; ATLAS-TDR-027, CERN, Geneva, 2017.
- [56] B. Nachman and J. Thaler, Learning from many collider events at once, *Phys. Rev. D* **103**, 116013 (2021).