Solving combinatorial problems at particle colliders using machine learning

Anthony Badea [1,*] William James Fawcett,[2] John Huth,[1] Teng Jian Khoo,[3]
Riccardo Poggi,[4] and Lawrence Lee[1,5,†]

[1]*Harvard University, Cambridge, Massachusetts 02138, USA*
[2]*Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, United Kingdom*
[3]*Humboldt-Universität zu Berlin, Berlin 12489, Germany*
[4]*Université de Genève, Geneva 1211, Switzerland*
[5]*University of Tennessee, Knoxville, Tennessee 37996, USA*

High-multiplicity signatures at particle colliders can arise in Standard Model processes and beyond. With such signatures, difficulties often arise from the large dimensionality of the kinematic space. For final states containing a single type of particle signature, this results in a combinatorial problem that hides underlying kinematic information. We explore using a neural network that includes a Lorentz Layer to extract high-dimensional correlations. We use the case of squark decays in *R*-Parity-violating Supersymmetry as a benchmark, comparing the performance to that of classical methods. With this approach, we demonstrate significant improvement over traditional methods.

At particle colliders such as the Large Hadron Collider (LHC) [1], there is a need to probe high-multiplicity signatures to study the decays of Standard Model (SM) particles and to search for new physics. There, the visible decay products from proton-proton interactions are measured by detectors, such as CMS and ATLAS, as a collection of N four-momentum vectors. Since each momentum contains four components, the measured kinematic feature space per reconstructed collision event is $\mathbb{R}^{4N}$. When the decay chains are complex and N is large, this dimensionality can be large enough that combinatorics prevent effective interpretation of the decay history. The aim of this letter is to highlight the capacity of machine learning (ML) to handle combinatorial problems more accurately and robustly than existing methods.

We are interested in event *interpretation*: the problem of matching parent and child particles. One often has many candidate interpretations of the decay history and must decide which is correct or accept a background from the incorrect interpretations. In the limit that all final-state particles have the same kind of experimental signature, even events with low multiplicity are swamped with combinatorial background making explicit mass resonance

reconstruction or crafting useful high-level observables difficult.

Since this problem has existed for decades, classical techniques (i.e., without ML) have been developed to overcome it. Minimizations over the set of combinations in invariant mass or angle are particularly popular for simple signatures [2]. For more complex scenarios, other analyses have used event-level χ^2 minimization [3], the Dalitz variables for three-body decays [4,5], jet substructure [6,7], and accidental substructure for avoiding combinatorics [8–10].

In this work, we demonstrate how a simple ML model can outperform the two most prominent existing methods for an example benchmark. We focus on a case inspired by *R*-parity-violating (RPV) Supersymmetry (SUSY) as a proof of concept [11], with a simplified model where pair-produced heavy stop squarks $\tilde{t}$ decay to two SM quarks each, giving a four-jet final state at leading order. This signature is ideal for our study because the all-hadronic (2×2) jet topology is one of the simplest cases facing this combinatorial background. Its relative simplicity allows traditional methods to retain some utility, unlike in more complex cases. Despite this simplicity, however, even this signature remains underprobed at the LHC [12–14].

For the (2×2) topology, at leading order, there are $\binom{4}{2}/2 = 3$ possible unique pairings. As only one of these is correct, a simple brute-force approach faces a 200% combinatorial background, even in a pure sample of signal events. The problem becomes even more difficult in practice due to the possible existence of additional high-momentum jets from initial state radiation (ISR) and

*[*]abadea@g.harvard.edu
[†]lawrence.lee.jr@cern.ch

overlapping particle interactions (pileup). This background is on top of fully SM background processes with multiple jets (primarily from pure QCD processes).

Figure 1 shows the level of combinatorial background as a function of the complexity of symmetrically decaying pair-produced multijet resonances. For larger multiplicities, brute-force solutions are significantly less effective. These high multiplicities may also arise from multistep cascade decays via on-shell intermediate resonances. In the RPV model, (2×4) and (2×5) signatures can easily arise when the RPV coupling is smaller than the gauge couplings and a cascade of decays to lighter beyond the standard model (BSM) particles can occur. These more complex topologies are particularly interesting because they can contain nested resonance structure. For example, simple RPV models may yield gluino $\tilde{g}$ pair production with each $\tilde{g}$ decaying to $qq\tilde{\chi}$, and the $\tilde{\chi}$ further decaying to qqq. To correctly interpret this topology, one must overcome the $\binom{10}{5}/2 = 126$ unique combinations to reconstruct each 5-jet resonance $\tilde{g}$, and then the additional $\binom{5}{3} = 10$ possibilities to resolve *each* 3-jet resonance $\tilde{\chi}$ correctly. The fully correct interpretation of the event is swamped in a combinatorial background of more than a million percent. Past searches for these signatures have proven rather weak for these reasons, as indicated in terms of some past search signal selection efficiencies shown in blue in Fig. 1 [5,6,10,13–17].
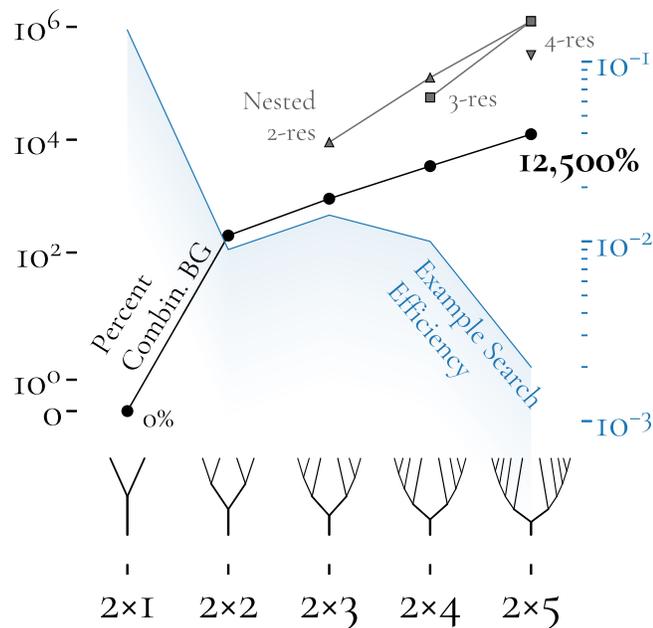
FIG. 1. Percent combinatorial background given a brute-force analysis (black) is shown as a function of final-state multiplicity. Increased complexities from nested resonances are shown for various intermediate decays (gray). Examples of past search signal selection efficiencies are also shown (blue) for 1 TeV resonances from recent LHC searches [5,6,10,13–16].

Similar challenges exist in SM measurements [18,19]. However, in searches for new particles, neither the location nor the width of a mass peak is known *a priori*. Therefore, a method must perform well across a broad range of potential signal masses and lineshapes to be useful in BSM searches. Combinatorial problems in SM event reconstruction are at a significant advantage of being able to use expected resonance masses as strong additional constraints. In the new particle searches discussed here, only aggregate peaking qualities and event-level symmetries can be exploited to reduce the complexity.

Classically, the problem can be constrained significantly beyond the brute-force approach by optimizing over some observable quantity. One prominent example for pair production is the asymmetry variable

$$\mathcal{A} = \frac{|m_1 - m_2|}{m_1 + m_2}, \quad (1)$$

where m_1 and m_2 represent the invariant masses of some exclusive subsets of objects. In this method, the combinatorial solution is taken as the particular configuration of objects into systems 1 and 2 that minimizes $\mathcal{A}$. This technique generally works well since at leading order the invariant mass of half of the objects should be equal to that of the other half. However, the performance drops with realistic detector resolutions since the width of the invariant mass distribution can allow for correct configurations with large $\mathcal{A}$.

Another approach is to minimize the opening angle between grouped objects using

$$\Delta R^\Sigma = \sum_{i \in \{1,2\}} |\Delta R_i - C|, \quad (2)$$

where $\Delta R_i \equiv \sqrt{\Delta\eta_i^2 + \Delta\phi_i^2}$ is the opening angle between momenta in group i with pseudorapidity η and azimuthal angle ϕ, and C is a customizable offset to optimize for nonzero opening angles that may be favored by the target signal. This method is successful when the resonances are produced well above threshold, with large back-to-back transverse boost. Using the limiting relation that the scale of the opening angle between the decay products of each resonance is roughly $(2m/p_T)$ for a parent particle of mass m and transverse momentum p_T, highly boosted resonances tend to have collimated decay products such that ΔR^Σ minimization performs well. This has been the primary method (with $C \approx 1$) used by ATLAS and CMS for all-hadronic RPV $\tilde{t}$ searches [13,14].

While these techniques have been useful, there is a significant amount of kinematic information not utilized, motivating the use of ML for this problem. Convolutional Neural Networks have been used on event images [20,21], but are not ideal because using images requires a choice of pixel segmentation. This results in sparse and higher

dimensional event representations while neglecting much of the complex jet physics that produces well-understood four-vectors. For example, an image of size 50×50 with three channels such as the tracker, ECAL, and HCAL has 7,500 dimensions, a far cry from the 16 dimensions of the original (2×2) problem. More recently, there is ongoing work on architecture properties tailored to kinematics problems, as in attention-based networks, for example [22–25]. Recent studies also explore the use of quantum annealers to solve similar event topology problems [26].

We deploy a relatively simple, physics-inspired architecture shown in Fig. 2 called CANNONBALL (*Combinatorial Artificial Neural Network ON a BAckronym Lorentz Layer*) based on an idea proposed for top-tagging [27–29] that a so-called *Lorentz Layer* can help efficiently utilize the relationships between four-vectors. Events are input as matrices with each row representing a separate four-momentum $(E, \vec{p})$. A *Combination Layer* enables the network to add four-momenta by taking weighted linear combinations of the input using a single learned matrix and softmax activation.[1] The fact that this matrix is learned may provide robustness against merging of nearby objects, although this has not been studied. In this implementation, a *Lorentz Layer* computes the properties (m, p_T, η, ϕ) of those combinations.[2] Lastly, a *Head* uses a batch normalization and feed-forward network to convert these features into an event label, described below.[3]

Simulated signal and background events were used to study performance. These samples were produced using MadGraph5_aMC@NLO 2.7.3 interfaced to MadEvent and the RPVMSSM UFO [30,31]. Production of $\tilde{t}$ pairs from 13 TeV pp collisions was simulated with up to one additional parton in the matrix element from ISR. Each $\tilde{t}$ is forced to decay to two light-flavor quarks. One million events at $\tilde{t}$ mass from 0.3 to 2 every 0.1 TeV were generated and split 75/25% for training and testing, resulting in a total training set size of 13.5M events. For simplicity, these events are analyzed at parton level as a function of the $\tilde{t}$ mass and a transverse momentum smearing that transforms $p_T \rightarrow \mathcal{N}(p_T, \epsilon p_T)$ for a smearing parameter ϵ to account for loss of information from fragmentation, hadronization, and detector effects.[4] Using ϵ rather than looking at reconstruction-level jets enables a quantitative understanding of how the algorithms perform as information is removed. Throughout this paper, only parton-level results

---

[1]The softmax ensures that no negative weights are applied while forming the combinations, as those could result in objects with $m^2 < 0$.

[2]This differs from Ref. [27] as there are no learned weights nor exotic observables computed. Through testing, this switch resulted in faster training and slightly better performance.

[3]The batch normalization was trained for the first epoch to fix the normalization for the remainder of the training; alternate schemes were tested and resulted in lower performance.

[4]In a dedicated study, ϵ values of roughly 20% captured the spread due to a parton shower model from PYTHIA8.2 [32,33].
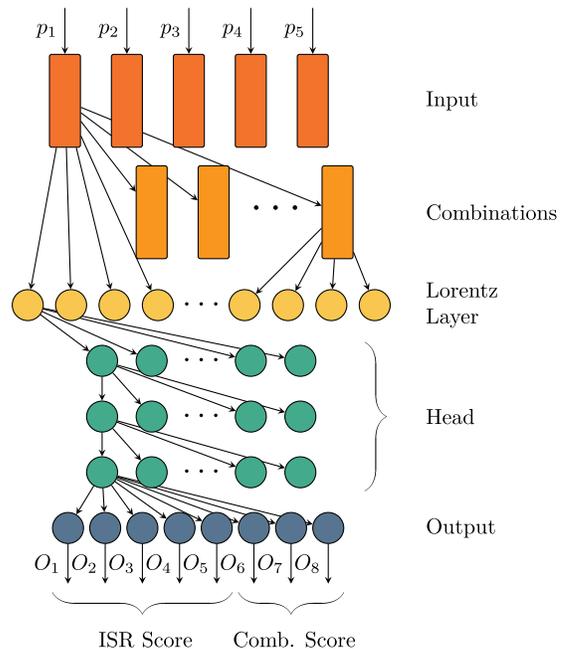
FIG. 2. Network structure of CANNONBALL including combination and Lorentz Layers. Only some connections are shown for clarity. All neurons from the Lorentz Layer and after are fully connected. Neurons represented by a circle correspond to scalar values, while rectangles denote four-vector-valued neurons.

are presented where loss of information is represented in a controlled way via ϵ. The angular extent of the hadronic shower, and therefore the potential for object merging, is ignored in this study. A background sample of QCD events containing exclusively $pp \rightarrow 4/5j$ was also generated at parton level, where j represents a parton.

Since the benchmark data is at parton level with up to one additional parton from ISR, CANNONBALL's input is a 5×4 event matrix, representing five four-vectors. There are no events with more than five partons, and those with four have the fifth row zero-padded. Explicit masking within the architecture for this padding was tested and did not improve results. The target event label is an 8-bit sequence; the first five bits identify which 4 partons are produced in the $\tilde{t}$ decays, while the last three detail how to combine them. The network produces a likelihood score for each bit.

In an attempt to achieve signal mass invariance, a single network was trained on all mass points at a given ϵ. During preprocessing, events were shuffled so that the mass points were mixed with roughly even statistics per batch, and the input four-vectors were randomly ordered per event. We also tried p_T ordering the inputs and saw better performance for fewer-epoch trainings but comparable performance for longer trainings.

The networks were implemented using PyTorch [34] and trained on an NVIDIA Quadro RTX using CUDA version 11.5 [35]. Binary Cross Entropy was minimized using the Adam optimizer over 30 epochs with a learning rate of 10^{-3}

and a batch size of 10 k [36]. The large batch size was important so that each mass sample was well-represented in the training. The *Combination Layer* and *Head* were chosen to have 30 combinations and three 200-node hidden layers, respectively. The number of learned parameters was 110 k, corresponding to $\sim$0.8% of the training set size. Networks were also trained for 20 times longer with the same architecture and for the same duration with 10$\times$-wider layers in the Combination Layer and Head, leading to only a $\sim$1% improvement, so the shorter length and smaller network was favored for reproducibility.

Figure 3 shows the fraction of correctly reconstructed events versus stop mass for CANNONBALL, $\mathcal{A}$ minimization, and ΔR^{Σ} minimization on the testing set. The bands indicate the degree of p_{T} smearing from $\epsilon = 0.1$ to 0.3. The classical methods are both significantly outperformed by the neural network. The ΔR^{Σ} minimization performance does not change with ϵ since the smearing does not modify the $\eta - \phi$ distributions and performs the worst because most of the stops are not highly boosted. The accuracy decreases with stop mass because the ratio $(2m/p_{\mathrm{T}})_{\tilde{t}}$ increases with mass, resulting in less collimated decay products. CANNONBALL performs significantly better than either of the classical methods across the full range of masses but with more mass dependence than the $\mathcal{A}$ minimization. This is expected since at lower masses the kinematics of the signal partons are more similar to those of ISR, but the general shape is heavily influenced by the training procedure, architecture choices, and the loss of information from nonzero ϵ values. At $\epsilon = 0.3$ and $m_{\tilde{t}} = 2$ TeV, CANNONBALL gives twice the performance

of the $\mathcal{A}$ minimization, and thirty-five times that of the ΔR^{Σ} minimization.

While Fig. 3 captures the performance as a function of stop mass, it is useful to understand in which regions of phase space the methods perform well. To quantify this, we calculate the Kullback-Leibler (KL) divergence of the four-dimensional stop momentum probability density functions (PDFs) [37]. The KL divergence is an entropic distance measure between multidimensional PDFs defined here as

$$\mathcal{D}_{\mathrm{KL}}(T||P) = \int T \log\left(\frac{T}{P}\right) dp^{\mu} \tag{3}$$

$$= \sum_{p_{\mathrm{T}} \text{ bins}} \sum_{\eta,\phi,m \text{ bins}} T \log\left(\frac{T}{P}\right) \tag{4}$$

$$= \sum_{p_{\mathrm{T}} \text{ bins}} \mathcal{D}_{\mathrm{KL}}^{\eta,\phi,m}(T||P, p_{\mathrm{T}}), \tag{5}$$

where $p^{\mu} = (p_{\mathrm{T}}, \eta, \phi, m)$ is a stop four-momentum and the second line converts T and P from continuous PDFs to 4D histograms of truth and prediction, respectively. A representative example of $\mathcal{D}_{\mathrm{KL}}^{\eta,\phi,m}(T||P, p_{\mathrm{T}})$ is plotted in Fig. 4 as a function of stop p_{T} for $m_{\tilde{t}} = 1$ TeV and $\epsilon = 0.2$. Smaller values represent more accurate interpretations of the events. The distribution between truth and itself is shown in black as a reference for zero divergence. There is a strong p_{T} dependence for the classical methods that is not present for the neural network, indicating that the largest performance gains are in the low stop p_{T} regime. This provides strong

FIG. 3. The fraction of correctly reconstructed events for CANNONBALL (blue), $\mathcal{A}$ minimization (orange), and ΔR^{Σ} minimization (green) as a function of stop mass for p_{T} smearing between $\epsilon = 0.1$ and 0.3.
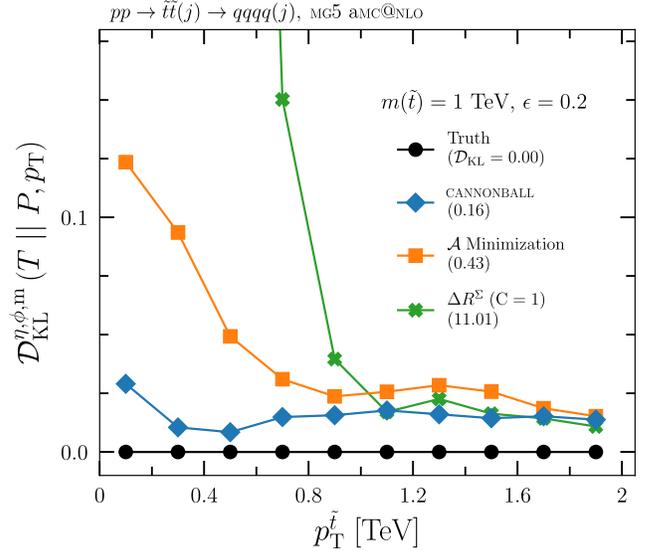
FIG. 4. KL divergence over the four-momentum phase space $p^{\mu} = (p_{\mathrm{T}}, \eta, \phi, m)$ space in bins of p_{T} of CANNONBALL (blue), $\mathcal{A}$ minimization (orange), ΔR^{Σ} minimization (green) with respect to truth for $m_{\tilde{t}} = 1$ TeV with $\epsilon = 0.2$. Zero divergence (black) is shown for visual aid. The total divergences $\mathcal{D}_{\mathrm{KL}}$ are shown in parentheses.

evidence that CANNONBALL is best at predicting the full kinematics of the parent stops.

When using these methods in a bump-hunt search, shape discrimination between the QCD multijet background and the hypothesis signal is crucial. Failing to correctly solve the combinatorial problem leads to decreased discrimination power. In particular, analyses often search in the average mass space $m_{avg} = \frac{1}{2}(m_1 + m_2)$, where m_1 and m_2 are the predicted candidate stop masses. CANNONBALL results in a more sharply peaked distribution than either of the other methods across all mass points considered. Figure 5 shows a shape comparison between the resulting m_{avg} distributions for the same signal model shown in Fig. 4. CANNONBALL most faithfully reproduces the true signal lineshape.

The blue filled histogram in Fig. 5 shows the shape of the QCD background contribution after applying the interpretation given by CANNONBALL, giving a smoothly falling distribution after an initial kinematic turn-on. None of these three methods significantly sculpt the QCD background distribution away from a steeply falling distribution in the mass range of interest. Figure 5 also shows the resulting QCD distributions from the classical methods, showing no significant sculpting. Final discovery sensitivity of each method will depend on the comparison between the shapes from QCD and signal for a given combinatorial solution. Determination of the properties of such a new particle would be best done using CANNONBALL.

In addition to its ability to improve event reconstruction, this approach would also easily fit into current analysis paradigms. The background estimation techniques, systematic uncertainty evaluation, observable construction, and statistical treatment will still apply from the classical
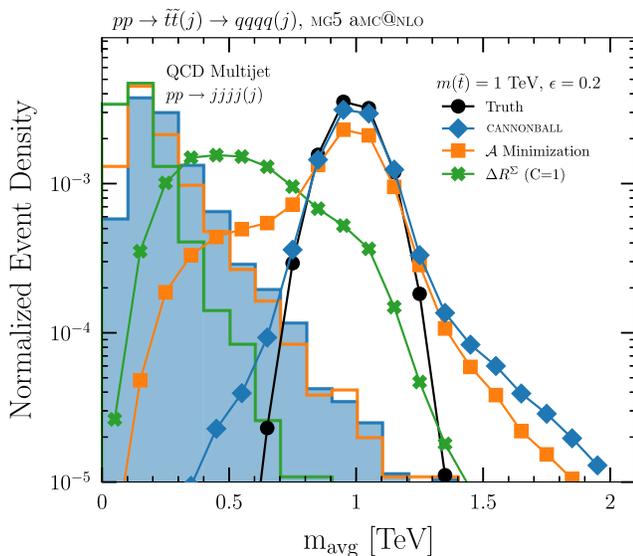
iterations of this analysis. In its simplest form, its just that the separation power of existing variables will be improved by decreasing the contributions from combinatorial backgrounds.[5]

The ability to use classical treatments for the evaluation of systematic uncertainties is a major advantage of this method. In previous work, no dedicated systematic was ever placed on the combinatorial solution method. Instead, experimental uncertainties were propagated to the details of the background estimate and on the final observables [13,14]. Replacing these classical combinatorial solutions with that from CANNONBALL gives no additional source of systematic uncertainty. Existing techniques fully cover the analysis uncertainty, without the complications seen when full-event classification networks are used [20,21].

There are several outstanding questions left for future work. One drawback of this architecture is that it operates via matrix multiplications on full events and therefore requires fixed-size inputs. In signatures with many jets, users would have to increase the architecture size considerably or exclude jets from consideration. A more robust method could natively handle variable size inputs. Theoretical analysis of the architecture, its functional representation and Jacobian, and generalizations are needed to better understand what is learned. Lastly, we aim to explore how to effectively use parton-level information to constrain learned representations of smeared or reconstruction-level events.

In summary, we have identified that the combinatorial background often becomes the primary challenge at high multiplicities and demonstrated a considerable performance increase on a simple (2×2) multijet signature as a function of resonance mass assuming realistic resolutions. Future work will extend the complexity of the signature and therefore the network, making those problems from well-motivated signatures tractable. More broadly, we advocate for an increased program of ML focusing on kinematic problems where there is an overabundance, rather than a lack, of information.

FIG. 5. A shape comparison of the per-event average candidate mass from CANNONBALL (blue), $\mathcal{A}$ minimization (orange), ΔR^{Σ} minimization (green) for $m_{\tilde{t}} = 1$ TeV with $\epsilon = 0.2$ and simulated QCD background.

---

[5]For this exact signature, a ML approach to constructing an optimal ABCD observable plane is suggested in Ref. [38], although no attempt is made at improving the combinatorial treatment.

[1] L. Evans and P. Bryant, J. Instrum. **3,** S08001 (2008).

[2] M. Aaboud *et al.* (ATLAS Collaboration), Phys. Rev. D **99,** 092004 (2019).

[3] S. Abachi *et al.* (D0 Collaboration), Phys. Rev. Lett. **74,** 2632 (1995).

[4] R. H. Dalitz, Phys. Rev. **94,** 1046 (1954).

[5] A. M. Sirunyan *et al.* (CMS Collaboration), Phys. Rev. D **99,** 012010 (2019).

[6] A. M. Sirunyan *et al.* (CMS Collaboration), Phys. Rev. Lett. **121,** 141802 (2018).

[7] G. Aad *et al.* (ATLAS Collaboration), J. High Energy Phys. 12 (2012) 086.

[8] T. Cohen, E. Izaguirre, M. Lisanti, and H. Lou, J. High Energy Phys. 03 (2013) 161.

[9] S. Chatrchyan *et al.* (CMS Collaboration), Phys. Lett. B **718,** 329 (2012).

[10] M. Aaboud *et al.* (ATLAS Collaboration), Phys. Lett. B **785,** 136 (2018).

[11] R. Barbier, C. Bérat, M. Besançon, M. Chemtob, A. Deandrea, E. Dudas, P. Fayet, S. Lavignac, G. Moreau, E. Perez *et al.*, Phys. Rep. **420,** 1 (2005).

[12] J. A. Evans and Y. Kats, J. High Energy Phys. 04 (2013) 028.

[13] M. Aaboud *et al.* (ATLAS Collaboration), Eur. Phys. J. C **78,** 250 (2018).

[14] A. M. Sirunyan *et al.* (CMS Collaboration), Phys. Rev. D **98,** 112014 (2018).

[15] G. Aad *et al.* (ATLAS Collaboration), J. High Energy Phys. 03 (2020) 145.

[16] G. Aad *et al.* (ATLAS Collaboration), Phys. Rev. D **91,** 112016 (2015); **93,** 039901(E) (2016).

[17] J. A. Evans, J. High Energy Phys. 08 (2014) 073.

[18] H. Alhazmi, Z. Dong, L. Huang, J. H. Kim, K. Kong, and D. Shih, Phys. Rev. D **105,** 115011 (2022).

[19] S. Qiu, S. Han, X. Ju, B. Nachman, and H. Wang, arXiv:2203.05687.

[20] W. Bhimji, S. Farrell, T. Kurth, M. Paganini, Prabhat, and E. Racah, J. Phys. Conf. Ser. **1085,** 042034 (2018).

[21] J. Guo, J. Li, T. Li, F. Xu, and W. Zhang, Phys. Rev. D **98,** 076017 (2018).

[22] A. Shmakov, M. J. Fenton, T.-W. Ho, S.-C. Hsu, D. Whiteson, and P. Baldi, SciPost Phys. **12,** 178 (2022).

[23] M. J. Fenton, A. Shmakov, T.-W. Ho, S.-C. Hsu, D. Whiteson, and P. Baldi, Phys. Rev. D **105,** 112008 (2022).

[24] J. S. H. Lee, I. Park, I. J. Watson, and S. Yang, arXiv:2012.03542.

[25] A. M. Sirunyan *et al.* (CMS Collaboration), Phys. Rev. D **104,** 052001 (2021).

[26] M. Kim, P. Ko, J.-h. Park, and M. Park, arXiv:2111.07806.

[27] A. Butter, G. Kasieczka, T. Plehn, and M. Russell, SciPost Phys. **5,** 028 (2018).

[28] M. Erdmann, E. Geiser, Y. Rath, and M. Rieger, J. Instrum. **14,** P06006 (2019).

[29] R. Poggi, Ph.D. thesis, Université de Genève, 2021.

[30] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, J. High Energy Phys. 07 (**2014**) 079.

[31] B. Fuks, Int. J. Mod. Phys. A **27,** 1230007 (2012).

[32] T. Sjostrand, S. Mrenna, and P. Z. Skands, J. High Energy Phys. 05 (2006) 026.

[33] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, Comput. Phys. Commun. **191,** 159 (2015).

[34] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, in *Advances in Neural Information Processing Systems 32*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., Red Hook, NY, 2019), pp. 8024–8035.

[35] NVIDIA, P. Vingelmann, and F. H. Fitzek, CUDA, release: 11.5 (2021).

[36] D. P. Kingma and J. Ba, arXiv:1412.6980.

[37] S. Kullback and R. A. Leibler, Ann. Math. Stat. **22,** 79 (1951).

[38] G. Kasieczka, B. Nachman, M. D. Schwartz, and D. Shih, Phys. Rev. D **103,** 035021 (2021).