Deep learning jet substructure from two-particle correlations

Kai-Feng Chen[1,*] and Yang-Ting Chien[2,3,†]

[1]*Department of Physics, National Taiwan University, Taipei 10617, Taiwan*
[2]*C. N. Yang Institute for Theoretical Physics, Stony Brook University, Stony Brook, New York 11794, USA*
[3]*Center for Frontiers in Nuclear Science, Stony Brook University, Stony Brook, New York 11794, USA*

Deciphering the complex information contained in jets produced in collider events requires a physical organization of the jet data. We introduce two-particle correlations (2PCs) by pairing individual particles as the initial jet representation from which a probabilistic model can be built. Particle momenta as well as particle types and vertex information are included in the correlation. A novel, two-particle correlation neural network (2PCNN) architecture is constructed by combining neural-network-based filters on 2PCs and a deep neural network for capturing jet kinematic information. The 2PCNN is applied to boosted boson and heavy flavor tagging, and it achieves excellent performance by comparing to image-based convolutional neural network and telescoping deconstruction. Major correlation pairs exploited in the trained models are also identified, which shed light on the physical significance of certain jet substructure.

In high-energy collider events, hundreds or even thousands of particles are produced, and the understanding of their high-dimensional probability distributions can be a formidable task. An emergent structure consisting of collimated particles, referred to as jets, is typically observed. The kinematic distribution and many aspects of the internal structure of jets have been the testing ground of quantum chromodynamics (QCD) in perturbative calculations and nonperturbative modeling, with remarkable success witnessed in reasonably accurate descriptions of collider data via Monte Carlo (MC) simulations and analytic calculations. However, the dynamical hadronization process which turns partonic degrees of freedom to hadronic degrees of freedom has not been fully understood, and new ways of quantifying hadron distributions may shed light on the description of hadronization.

In this paper, two-particle correlations (2PCs) are explored as a novel representation of jet information to illuminate the physics underlying jet formation from parton evolution to hadronization. This is one step beyond processing individual particle information by constructing particle pairs as basic information elements, from which probabilistic models can be built and physical analysis can be performed. The model includes particle momenta for energy flow information, and it can also seamlessly include electric charge and vertex information which are sensitive to hadronization as well as bottom quark decays. Significant features can build up, because relevant jet information is already contained in 2PC pairs.

Remarkably, modern computation power has made possible the use of machine learning techniques on complex data, and many methods have been applied successfully on classification and regression problems in particle and nuclear physics, such as jet classification [1–20], correlation of particles [21,22], anomaly detection [23–26], event generation [27–29], and other tasks [30,31]. We will tackle classic classification problems such as boosted boson and heavy flavor jet tagging, as a way to discover and highlight certain jet properties which are relevant in these tasks. Specifically, the discrimination of two-prong jets (W jets and Higgs jets from the $H \to b\bar{b}$ decay channel) and three-prong jets (fully hadronic top jets) against light-quark q ($q = u, d, c, s$ quark) jets, as well as W^+ vs W^- [20], and quark vs gluon jet discrimination, are studied. Excellent performance of a 2PC-based neural network and comparisons with an image-based convolutional neural network (CNN) [20] and telescoping deconstruction [9] will be presented in all of the tasks. In particular, the network optimized for W tagging successfully identifies the two-prong structure and color-singlet nature of W jets by weighing strongly on these two features. The model behavior will be analyzed by examining collinear and soft contributions from soft-drop [32,33] and collinear-drop [34] constituents and their correlations. We demonstrate that a combination of machine learning and physics analysis methods benefits significantly from the use of a physically organized and unbiased jet representation so that one can extract the physics features the model identifies.

[*]kfjack@phys.ntu.edu.tw
[†]yang-ting.chien@stonybrook.edu

The analysis is performed with samples generated from MC simulations using MADGRAPH [35] for hard scattering processes and PYTHIA8 [36] for parton shower and hadronization. Jets are defined using the anti-k_T algorithm [37] implemented in FastJet 3 [38], with $R = 0.8$ for the studies of tagging high-p_T two- or three-prong jets and with $R = 0.4$ for the studies of quark gluon discrimination. The high-p_T $R = 0.8$ jets are generated using decays of hypothetical heavy Z' bosons ($Z' \rightarrow W^+W^-, ZH, t\bar{t}, q\bar{q}$) with the invariant mass fixed at 2 TeV, while the jets used in quark gluon discrimination are generated with the standard model QCD processes. For the samples generated using Z' decays, jets are produced and reconstructed in the same kinematic region; therefore, the classification is not affected by the hard process kinematics. The truth particle information is passed through a DELPHES [39] fast detector simulation and converted into particle flow candidates, with track, electromagnetic calorimeter, and hadronic calorimeter information. A parametric model based on the CMS detector [40] at the Large Hadron Collider is introduced in the simulation.

Two different sets of 2PC inputs are included. A basic set contains only the energy flow information,[1] including the transverse momentum fraction $z = p_T^i/p_T(\text{jet})$, relative pseudorapidity $\Delta\eta = \eta^i - \eta(\text{jet})$, and relative azimuthal angle $\Delta\phi = \phi^i - \phi(\text{jet})$ of the jet constituents labeled by the index i. Here p_T^i, η^i, and ϕ^i are the transverse momentum, pseudorapidity, and azimuthal angle of particle i, respectively, and $p_T(\text{jet})$, $\eta(\text{jet})$, and $\phi(\text{jet})$ are the corresponding quantities of the jet. The other set of inputs contains the 2PCs of charged tracks, including the vertex position and the charge of each particle.

Based on the 2PC inputs, we design a two-particle correlation neural network (2PCNN)[2] to model the probability distribution of jet particles (see Fig. 1), which is implemented using Keras [42] with TensorFlow back end [43]. Since the number of jet particles can vary, the 2PCNN layer is designed to handle inputs with variable sizes. Inspired by one of the key ideas from the CNN, the 2PCNN model implements a collection of filters[3] with shared weights to process the input of all 2PC data, while CNN filters process only local patches of the jet image. In the prototype model, the number of filters is set to 64 to extract features from the energy flow information. The vertex and charge information is processed with a parallel 2PCNN layer containing 32 filters. Each filter processes and gives outputs to all input 2PCs. The filter outputs are then ranked according to their

---

[1]The energy flow input here includes infrared and collinear unsafe information.

[2]The prototype 2PCNN example code and test samples are available from Ref. [41].

[3]The filter consists of a fully connected dense network with 2PCs as the input, processed with a hidden layer, and then a layer of single nodes as the output. We use the ReLU [44] activation function at each layer; therefore, the output can be only nonnegative floating-point numbers.
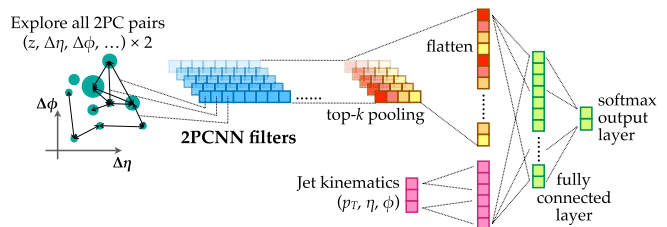

FIG. 1. The schematic view of the 2PCNN model. It processes two-particle correlations as inputs and uses filters with shared weights to benchmark the importance of each 2PC pair. The top-k-ranked filter outputs, together with jet kinematic information, are fed into a fully connected network for decision making.

numerical values, and only the top-k-ranked 2PCs of each filter are kept as the inputs for the subsequent decision-making, fully connected network. In order to balance between performance and complexity, $k = 4$ has been set; therefore, the total number of output nodes is $256 = 64 \times 4$, which is equal to the number of filters times k.

Besides the 2PCNN layers, we use a dense network to include the jet kinematic information $p_T(\text{jet})$, $\eta(\text{jet})$, and $\phi(\text{jet})$ which is the baseline input for standard analysis. The outputs of the dense network and the 2PCNN layer are sent to another fully connected layer of 128 nodes (or 256 nodes if two 2PCNN layers are used), followed by two output nodes with the softmax activation function for the final decision. The model is optimized by minimizing a categorical cross-entropy loss function with the Adam optimizer [45]. Input samples for each task are split into three subsets: one set consisting of 80 thousand jets is used to optimize the weights in the model, and another set of 40 thousand jets is used to validate if the model reaches its optimal performance. The other set of 40 thousand jets is used for an independent measure of the model performance.

In order to benchmark the 2PCNN performance, we compare with a deep neural network model based on telescoping deconstruction of energy flow information (referred to as the T-jet model) [9,46–48] and CNN models based on jet image information. The telescoping deconstruction method systematically decomposes jet information into a fast-converging subjet series expansion $\sum_N T_N$ which is ordered by the number of subjets N. These subjets are defined as the sets of particles along dominant energy flow directions within a variable subjet radius. Such organization is motivated by the infrared structure of QCD. Energetic, collinear particles are captured at lower orders, and the series gradually reaches out to soft, wide-angle particles. In this paper, the T-jet model includes jet information up to the T_3 order and scans energy flows with four values of subjet radius. The energy flow directions and subjet kinematics consist of 60 input variables. Together with the jet kinematic information, these inputs are processed by a fully connected network layer of 128 nodes followed by two output nodes. The same activation function, loss function, and optimizer are adopted as in the 2PCNN model.
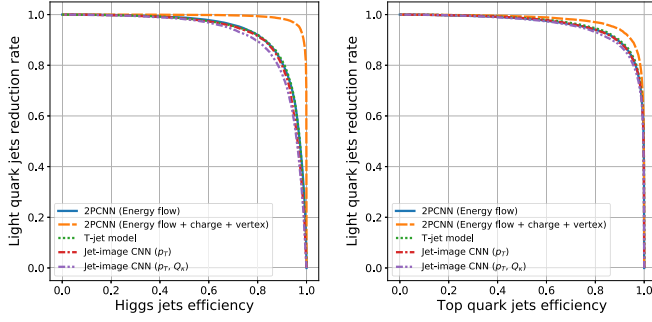
FIG. 2. The receiver operating characteristic curves for classification of Higgs jets vs light-quark jets (left) and top jets vs light-quark jets (right). The solid curves show the performance of the 2PCNN model based on energy flow information. The dashed curves correspond to the 2PCNN model with additional electric charges and vertex inputs. The dotted curves give the result from the T-jet model.

On the other hand, the reference CNN models are constructed based on the implementation described in Ref. [20]. The particle data have been converted to jet images with a dimension of 75×75 pixels, in the range of $|\Delta\eta| < 0.8$ and $|\Delta\phi| < 0.8$. A rotation in $\Delta\eta - \Delta\phi$ coordinate system is performed to align the principle axis of the jet constituents horizontally. Two channels are considered in the CNN models: a p_T channel and a jet charge (Q_κ) channel. In the p_T channel, the value of each jet image pixel is the sum of the p_T's for the particles within the pixel. In the Q_κ channel, the pixel value is equal to the sum of $Q_i \times (p_T^i)^\kappa/(p_T(\text{jet}))^\kappa$ for the particles within the pixel, where Q_i is the charge of the particle and κ is set to 0.15. The input jet images are connected to a combination of three two-dimensional convolutional layers (32 filters of 6×6 pixels, 128 filters of 4×4 pixels, and 256 filters of 6×6 pixels) and max-pooling layers of 2×2 pixels. The network is ending with two fully connected layers of 512 nodes, followed by two output nodes with the softmax activation function for final classification. We compared with the

CNN model with only the p_T channel, which provides a direct comparison with models using only energy flow information. We also tested the model including both p_T and Q^κ channels.

Figure 2 shows the receiver operating characteristic (ROC) curves, plotting the background rejection rate as a function of the signal efficiency, for two discrimination tasks as representative examples: high-p_T Higgs jet vs light-quark jet, as well as top jet vs light-quark jet. The model performances are quantified by the area under the ROC curve (AUC) and the average accuracy (ACC), which is the fraction of correctly predicted jet samples. As summarized in Table I, the 2PCNN, the T-jet model, and the CNN model with the p_T channel based on energy flow information show nearly the same performance. This confirms the baseline capability of the 2PCNN model, which is comparable to the state-of-the-art methods that are all capable of modeling the energy flow probability distributions very well. With the additional vertex and charge information, the 2PCNN model achieves excellent performance in all the classification tasks. The vertex information has a strong impact on tagging jets which contain one or more secondary vertices such as the high-p_T Higgs and top jets. The electric charges of particles are also essential for separating jets from W^+ and W^- bosons, although the CNN model with both p_T and Q_κ channels still provides a better discriminant power. The power-law factor with preassigned $\kappa = 0.15$ is found to be optimal in the task of W^+ and W^- separation. With $\kappa = 1$, the performance of CNN model is comparable to the performance of 2PCNN.

We also compare the actual numbers of trainable parameters in these models. The 2PCNN model with only energy flow information includes 6.2 thousand parameters for feature extraction and 34 thousand parameters for the decision-making fully connected network. The 2PCNN model with full information contains 15 thousand parameters for feature extraction and 101 thousand parameters for the subsequent fully connected network. The T-jet model is implemented with a fully connected network of

TABLE I. The performance of the 2PCNN, T-jet, and CNN models, as quantified by the average accuracy (ACC) and the area under the receiver operating characteristic curve (AUC), for W, Higgs, and top tagging as well as W^+ vs W^- and quark vs gluon discrimination. The energy flow 2PCNN model has comparable performance with the T-jet model and the CNN model with the p_T channel. The 2PCNN model with additional information of electric charges and vertex of charged tracks outperforms significantly the other models in most of the tasks, except the W^+ and W^- separation. The uncertainty due to the finite sample size in ACC is smaller than 0.003.

|	2PCNN(E-flow)		2PCNN(full)		T-jet model		CNN(p_T)		CNN(p_T, Q_κ)	
Task	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC
W vs quark	0.879	0.942	0.877	0.943	0.880	0.945	0.881	0.944	0.869	0.934
Higgs vs quark	0.865	0.935	0.961	0.993	0.866	0.934	0.864	0.927	0.850	0.918
Top vs quark	0.897	0.960	0.924	0.976	0.900	0.963	0.894	0.960	0.888	0.955
Top vs W	0.940	0.983	0.952	0.987	0.941	0.982	0.936	0.981	0.934	0.979
W^+ vs W^-	0.502	0.504	0.742	0.823	0.499	0.499	0.499	0.500	0.815	0.894
Quark vs gluon	0.739	0.811	0.749	0.823	0.732	0.803	0.731	0.800	0.730	0.797
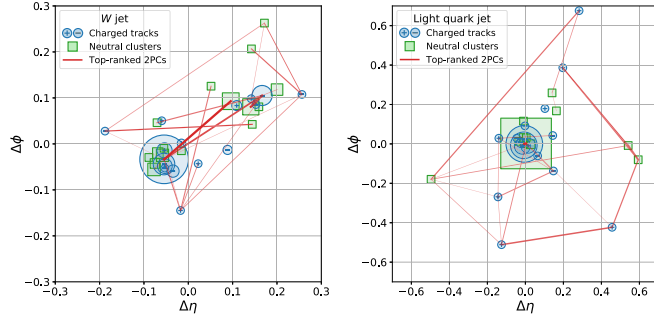
FIG. 3.   Displays of a typical W jet (left) and a typical light-quark jet (right) in the $\Delta\eta - \Delta\phi$ plane. The charged tracks of jet particles are shown as circles with charge signs, while the neutral clusters are shown as squares. The sizes of the circles or the squares are proportional to the p_T's of jet constituents. The solid lines indicate the top-one-ranked 2PCs of the filters in the energy flow 2PCNN model. The strength of filter outputs are represented by the line thickness.

26 thousand free parameters only, since the feature extraction procedures have been carried out in the telescoping deconstruction step. The CNN models include 1.2 million parameters in the three convolutional layers, following by 3.5 million parameters for the decision-making network. The 2PCNN models in our studies are still rather light weighted, which already shows a good efficiency for describing the jet data. A scaled-up 2PCNN configuration, such as a model with more filters, with deeper or wider filters, or with multiple layers, may improve the yet-optimal performance in certain tasks such as W^+ and W^- tagging.

We now discuss the physics properties of the 2PCs and focus on the task of W jet and light-quark jet separation using the energy flow 2PCNN model, aiming to identify the key features which are useful for distinguishing the two jet samples. Many other detailed studies will be presented in a forthcoming paper. Thanks to the internal ranking of 2PC pairs, the importance of the top-k-ranked 2PC pairs

within a filter can potentially be quantified by their filter output values. These sets of outputs represent the weights on 2PCs which the 2PCNN has learned from separating the two samples and are task dependent. Therefore, intrinsic features of each jet sample can be illuminated by contrasting with different jet samples potentially having distinct features.

Figure 3 shows the display of a typical two-prong W jet and a typical one-prong light-quark jet. The jet constituents are shown as scattered circles and squares, with their sizes proportional to the particle transverse momenta. The top-one-ranked 2PC pair of each active 2PCNN filter is indicated by a solid line, with the thickness representing the strength of the filter output. Two distinct signatures of the high-ranked 2PCs are identified: (i) strong internal correlations within and between the prongs and (ii) strong correlations between high-p_T constituents within the prongs and low-p_T constituents scattered at wide angle.

Such behaviors of the high-ranked 2PC pairs are further examined by the spatial distance $\Delta R = \sqrt{(\eta^i - \eta^j)^2 + (\phi^i - \phi^j)^2}$ between the ith and jth particles forming the 2PC, and their p_T asymmetry $\mathcal{A}(p_T) = |p_T^i - p_T^j|/(p_T^i + p_T^j)$. Figure 4 shows the comparisons of a variety of ΔR and $\mathcal{A}(p_T)$ distributions of W jets and light-quark jets. In order to maximize the sensitivity to the features extracted by the 2PCNN, the distributions corresponding to the top-ranked 2PCs weighed by the output values of 2PCNN filters, as an indication of their importance, are presented in the lower panels. For W jets, strong features are identified at $\Delta R \approx 0$ and $\Delta R \approx 0.2 \sim 2m_W/p_T(\text{jet})$, whereas for light-quark jets the $\Delta R \approx 0$ feature is strong and the $\Delta R \approx 0.2$ feature is absent. This indicates the intrinsic jet property of particle collimation for both samples and the two-prong structure of W jets. The filters tend to either select the 2PCs within the same prong (therefore, with small ΔR values) or emphasize the correlations between the two prongs for W jets and build up the
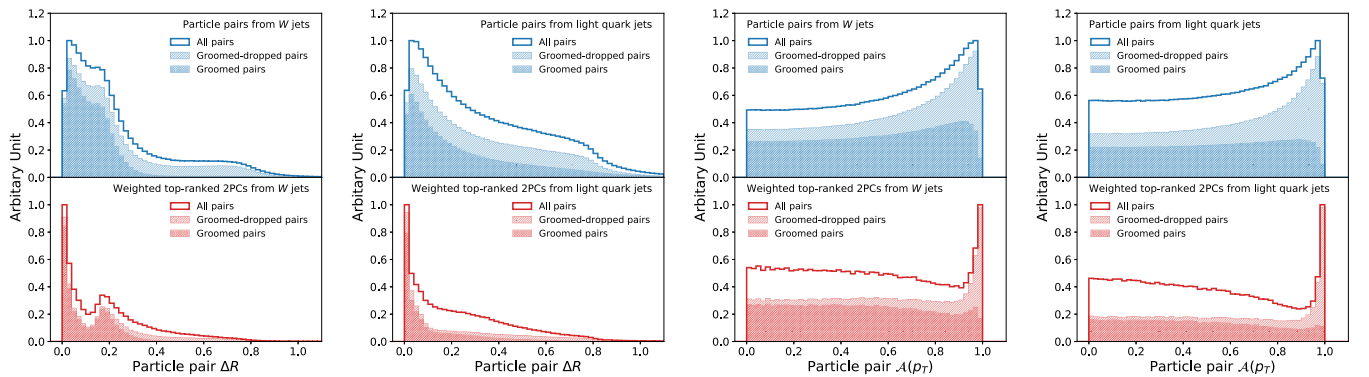

FIG. 4.   The distributions of spatial distance ΔR (left two panels) and p_T asymmetry $\mathcal{A}(p_T)$ (right two panels) between particles in 2PCs for W jets and light-quark jets. The lower panels show the distributions for top-ranked 2PCs weighed by the 2PCNN filter outputs, while the upper panels have equal weights for all the 2PCs. In each subfigure, the solid lines represent the histograms from all 2PC pairs, which are decomposed into three stacked, hatched, or unfilled components corresponding to groomed-groomed, groomed-dropped, and dropped-dropped 2PC pairs, respectively, as categorized by soft-drop and collinear-drop methods.

$\Delta R \approx 0.2$ feature. On the other hand, a clear feature at $\mathcal{A}(p_T) \approx 1$ shows up in the filter-output weighed $\mathcal{A}(p_T)$ distributions for both samples. Such a signature corresponds to highly unbalanced p_T's in the 2PCs; therefore, one of the particles has to be soft. This shows the importance of low-p_T constituents which are often neglected or suppressed in many other jet tagging methods.

In order to further examine the properties of 2PCs which are responsible for the learned jet features, soft drop and collinear drop with parameters $z_{\mathrm{cut}} = 0.2$ and $\beta = 0$ are used to classify jet constituents into two categories. The jet constituents surviving soft drop are referred to as "groomed," while those surviving collinear drop belong to the "dropped" category. Therefore, the 2PCs form three distinct sets: groomed-groomed, groomed-dropped, and dropped-dropped. We can see that the one- and two-prong structures are dominantly determined by the groomed-groomed 2PC pairs from the two soft-drop branches. Also, there is a significant dropped-dropped contribution at medium and large ΔR values for light-quark jets. On the other hand, the feature at $\mathcal{A}(p_T) \approx 1$ dominantly comes from the groomed-dropped 2PC pairs which correlate hard, collinear particles to soft, wide-angle particles, while most other 2PC pairs form a fairly flat $\mathcal{A}(p_T)$ distribution.

To highlight the power and sensitivity of 2PCNN in feature extraction, we contrast with the ΔR and $\mathcal{A}(p_T)$ distributions of W and light-quark jets formed with equal weight for all the 2PC pairs (upper panels in Fig. 4). Evidence of one- or two-prong structure from the falling ΔR distribution with a "shoulder" around $\Delta R \approx 0.2$, as well as the significant soft particle contributions in the $\mathcal{A}(p_T) \approx 1$ region, is observed. Similar conclusions can be reached by decomposing the distributions into groomed-groomed, groomed-dropped, and dropped-dropped components; however, all the features are much more convincingly identified by the 2PCNN model as a very useful guide for physics analysis.

We discuss the similarity and difference among 2PCNN and other related methods. While 2PCNN and graph-based networks both have particle inputs as sets and use correlations [21,22], 2PCNN does not perform convolution, whereas ParticleNet [22] does edge convolution. Also, 2PCNN treats exclusively particle correlations, while energy flow network [21] and spectral analysis [17–19] use concretely defined jet observables.

The number of 2PCs scales quadratically with the number of particles; therefore, it creates a redundancy in the jet representation and increases computational complexity in the analysis of 2PC data. However, since the processing of 2PC data is carried out only once for each filter, the actual computational cost is similar only to a process of generic two-body decay reconstruction. The 2PC pairs can be reduced by introducing preselections such as a minimum particle p_T threshold or a maximum ΔR angular separation. On the other hand, the 2PCNN filters concern only the top-ranked 2PCs; therefore, a dynamical way of reducing 2PCs can be achieved if those low-ranked 2PC pairs are removed during or after the training process. A further optimization of the algorithm in terms of computational complexity is foreseeable.

In conclusion, we have constructed a new neural network architecture which utilizes 2PCs as a fundamental description of jets. The input for 2PCNN is dynamically determined by the number of jet constituents with no artificial reduction of input information and no particular biased ordering of jet particles. The structure of the 2PC neural network is driven by the physics needs rather than a direct application of existing deep learning methods developed for solving problems in other subjects. We demonstrate that the 2PCNN model based on energy flow information has comparable performance with the model using variables from telescoping deconstruction as well as the image-based convolutional neural network model, which are the most effective methods for factorizing jet information. By including additional information from charged tracks, such as electric charges and vertex, the 2PCNN model achieves an unprecedentedly promising power for a variety of jet tagging tasks. Besides the excellent tagging performance, an important benefit of the 2PCNN model is the ranking of 2PCs which can be directly extracted from the filter outputs. Since two-particle correlations are fundamental descriptions of particles relations, this physical machine-learning method can be potentially useful in subsequent physics studies such as the hadronization process and collective behaviors of quark-gluon plasma remnants in high-energy collisions. The 2PCNN will shed light on physics signatures which are difficult to identify with conventional methods.

[1] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman, J. High Energy Phys. 07 (2016) 069.

[2] P. Baldi, K. Bauer, C. Eng, P. Sadowski, and D. Whiteson, Phys. Rev. D **93,** 094034 (2016).

[3] G. Kasieczka, T. Plehn, M. Russell, and T. Schell, J. High Energy Phys. 05 (2017) 006.

[4] S. Macaluso and D. Shih, J. High Energy Phys. 10 (2018) 121.

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

[6] G. Louppe, K. Cho, C. Becot, and K. Cranmer, J. High Energy Phys. 01 (2019) 057.

[7] T. Cheng, Comput. Softw. Big Sci. **2,** 3 (2018).

[8] S. Egan, W. Fedorko, A. Lister, J. Pearkes, and C. Gay, arXiv:1711.09059.

[9] Y.-T. Chien and R. Kunnawalkam Elayavalli, arXiv: 1803.03589.

[10] K. Fraser and M. D. Schwartz, J. High Energy Phys. 10 (2018) 093.

[11] J. Lin, M. Freytsis, I. Moult, and B. Nachman, J. High Energy Phys. 10 (2018) 101.

[12] A. Andreassen, I. Feige, C. Frye, and M. D. Schwartz, Eur. Phys. J. C **79,** 102 (2019).

[13] A. Andreassen, I. Feige, C. Frye, and M. D. Schwartz, Phys. Rev. Lett. **123,** 182001 (2019).

[14] A. Butter, T. Plehn, and R. Winterhalder, SciPost Phys. **7,** 075 (2019).

[15] L.-G. Pang, K. Zhou, N. Su, H. Petersen, H. Stöcker, and X.-N. Wang, Nat. Commun. **9,** 210 (2018).

[16] L.-G. Pang, K. Zhou, and X.-N. Wang, arXiv:1906.06429.

[17] S. H. Lim and M. M. Nojiri, J. High Energy Phys. 10 (2018) 181.

[18] A. Chakraborty, S. H. Lim, and M. M. Nojiri, J. High Energy Phys. 07 (2019) 135.

[19] A. Chakraborty, S. H. Lim, M. M. Nojiri, and M. Takeuchi, arXiv:2003.11787.

[20] Y.-C. J. Chen, C.-W. Chiang, G. Cottin, and D. Shih, Phys. Rev. D **101,** 053001 (2020).

[21] P. T. Komiske, E. M. Metodiev, and J. Thaler, J. High Energy Phys. 01 (2019) 121.

[22] H. Qu and L. Gouskos, Phys. Rev. D **101,** 056019 (2020).

[23] J. H. Collins, K. Howe, and B. Nachman, Phys. Rev. Lett. **121,** 241803 (2018).

[24] M. Farina, Y. Nakai, and D. Shih, Phys. Rev. D **101,** 075021 (2020).

[25] A. Blance, M. Spannowsky, and P. Waite, J. High Energy Phys. 10 (2019) 047.

[26] T. S. Roy and A. H. Vijay, arXiv:1903.02032.

[27] M. Paganini, L. de Oliveira, and B. Nachman, Phys. Rev. D **97,** 014021 (2018).

[28] L. de Oliveira, M. Paganini, and B. Nachman, Comput. Softw. Big Sci. **1,** 4 (2017).

[29] M. Paganini, L. de Oliveira, and B. Nachman, Phys. Rev. Lett. **120,** 042003 (2018).

[30] P. T. Komiske, E. M. Metodiev, B. Nachman, and M. D. Schwartz, J. High Energy Phys. 12 (2017) 051.

[31] A. Andreassen, P. T. Komiske, E. M. Metodiev, B. Nachman, and J. Thaler, Phys. Rev. Lett. **124,** 182001 (2020).

[32] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, J. High Energy Phys. 09 (**2013**) 029.

[33] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, J. High Energy Phys. 05 (2014) 146.

[34] Y.-T. Chien and I. W. Stewart, J. High Energy Phys. 06 (2020) 064.

[35] 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.

[36] T. Sjostrand, S. Mrenna, and P. Z. Skands, Comput. Phys. Commun. **178,** 852 (2008).

[37] M. Cacciari, G. P. Salam, and G. Soyez, J. High Energy Phys. 04 (2008) 063.

[38] M. Cacciari, G. P. Salam, and G. Soyez, Eur. Phys. J. C **72,** 1896 (2012).

[39] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi (DELPHES 3 Collaboration), J. High Energy Phys. 02 (2014) 057.

[40] S. Chatrchyan et al. (CMS Collaboration), J. Instrum. **3,** S08004 (2008).

[41] https://github.com/kfjack/2PCNN.

[42] F. Chollet et al., Keras, https://keras.io (2015).

[43] M. Abadi et al., TensorFlow: Large-scale machine learning on heterogeneous systems (2015), software available from tensorflow.org.

[44] V. Nair and G. E. Hinton, in Proceedings of the 27th International Conference on Machine Learning (ICML-10) (Omnipress, Madison, WI, 2010), pp. 807–814.

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

[46] Y.-T. Chien, Phys. Rev. D **90,** 054008 (2014).

[47] Y.-T. Chien, D. Farhi, D. Krohn, A. Marantan, D. Lopez Mateos, and M. Schwartz, J. High Energy Phys. 12 (2014) 140.

[48] Y.-T. Chien, A. Emerman, S.-C. Hsu, S. Meehan, and Z. Montague, arXiv:1711.11041.