

RECEIVED: February 11, 2019

REVISED: July 20, 2019

ACCEPTED: July 30, 2019

PUBLISHED: August 9, 2019

# Probing stop pair production at the LHC with graph neural networks

Murat Abdughani,<sup>a,b,c</sup> Jie Ren,<sup>b,c</sup> Lei Wu<sup>a</sup> and Jin Min Yang<sup>b,c,d</sup>

<sup>a</sup>*Department of Physics and Institute of Theoretical Physics, Nanjing Normal University, Nanjing, 210023, China*

<sup>b</sup>*CAS Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China*

<sup>c</sup>*School of Physics, University of Chinese Academy of Sciences, Beijing 100049, China*

<sup>d</sup>*Department of Physics, Tohoku University, Sendai 980-8578, Japan*

*E-mail:* [mulati@itp.ac.cn](mailto:mulati@itp.ac.cn), [renjie@itp.ac.cn](mailto:renjie@itp.ac.cn), [leiwu@itp.ac.cn](mailto:leiwu@itp.ac.cn), [jmyang@itp.ac.cn](mailto:jmyang@itp.ac.cn)

**ABSTRACT:** Top-squarks (stops) play a crucial role for the naturalness of supersymmetry (SUSY). However, searching for the stops is a tough task at the LHC. To dig the stops out of the huge LHC data, various expert-constructed kinematic variables or cutting-edge analysis techniques have been invented. In this paper, we propose to represent collision events as event graphs and use the message passing neural network (MPNN) to analyze the events. As a proof-of-concept, we use our method in the search of the stop pair production at the LHC, and find that our MPNN can efficiently discriminate the signal and background events. In comparison with other machine learning methods (e.g. DNN), MPNN can enhance the mass reach of stop mass by several tens of GeV to over a hundred GeV.

**KEYWORDS:** Supersymmetry Phenomenology

**ARXIV EPRINT:** [1807.09088](https://arxiv.org/abs/1807.09088)

---

**Contents**

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Methodology</b>	<b>2</b>
<b>3</b>	<b>Numerical results and discussions</b>	<b>5</b>
<b>4</b>	<b>Conclusions</b>	<b>8</b>

---

**1 Introduction**

After the discovery of the Higgs boson, the pursuit of new physics beyond the Standard Model (SM) is a primary goal of the LHC experiment. A major guideline in this endeavor is the naturalness principle which implies that the new physics for stabilizing the Higgs mass should appear at TeV scale. Among all the proposed scenarios, the weak scale SUSY remains as one of the most popular models, in which the quadratically divergent contribution to the Higgs mass from the top quark is canceled by the top-squarks (stops). Thus, the search for the stops is crucial for testing the naturalness of SUSY.

However, searching for the stops at the LHC is a challenging task due to the complicated nature of super-particles (sparticles). For examples, (i) when  $m_{\tilde{t}_1} \gg m_t + m_{\tilde{\chi}_1^0}$ , the stop can decay to  $t\tilde{\chi}_1^0$  and produce an energetic top quark. Using endpoint observables, like  $M_T$  or  $M_{T_2}$ , the  $t\bar{t}$  background can be efficiently reduced [1–5]. (ii) In the compressed region  $m_{\tilde{t}_1} - m_{\tilde{\chi}_1^0} \approx m_t$ , the kinematics of stop pair events closely resemble the  $t\bar{t}$  background events, rendering the searches rather difficult. Thanks to the ISR jet, the stop events in such a compressed region will have a peak-like feature around the ratio of missing transverse momentum vector to the transverse momentum vector of  $t\bar{t}$  system [6–8], while the  $t\bar{t}$  background does not show such a peak. If the LSP in the above compressed region becomes almost massless, the precision measurements of  $t\bar{t}$  cross section [9] or spin-correlation [10] can also be used to probe the light stop. (iii) When the two body decays  $\tilde{t}_1 \rightarrow t\tilde{\chi}_1^0$  and  $\tilde{t}_1 \rightarrow b\tilde{\chi}_1^+$  are kinematically forbidden, the three-body decay  $\tilde{t}_1 \rightarrow W^+b\tilde{\chi}_1^0$  [11], the two-body flavor-changing decay  $\tilde{t}_1 \rightarrow c\tilde{\chi}_1^0$  [12–14] or even the four-body decay  $\tilde{t}_1 \rightarrow bf'\tilde{f}\tilde{\chi}_1^0$  [15] would happen. But due to the small mass splitting, the decay products of the stop are usually too soft to be observed. Thus the ISR/FSR jet (plus the heavy quark tagging) is needed to trigger these stop events [16–18]. In addition, other miscellaneous cut-flow based studies of stop searches in different parameter space have been performed at the LHC [19–37]. A vast number of experimental searches for stop pair productions have been also devoted at the LHC in the past few years [38–43]. Besides the traditional cut-flow based analysis technique, the machine learning (ML) method provides an alternative way for signal/background discrimination and new particle searches, which have been used for

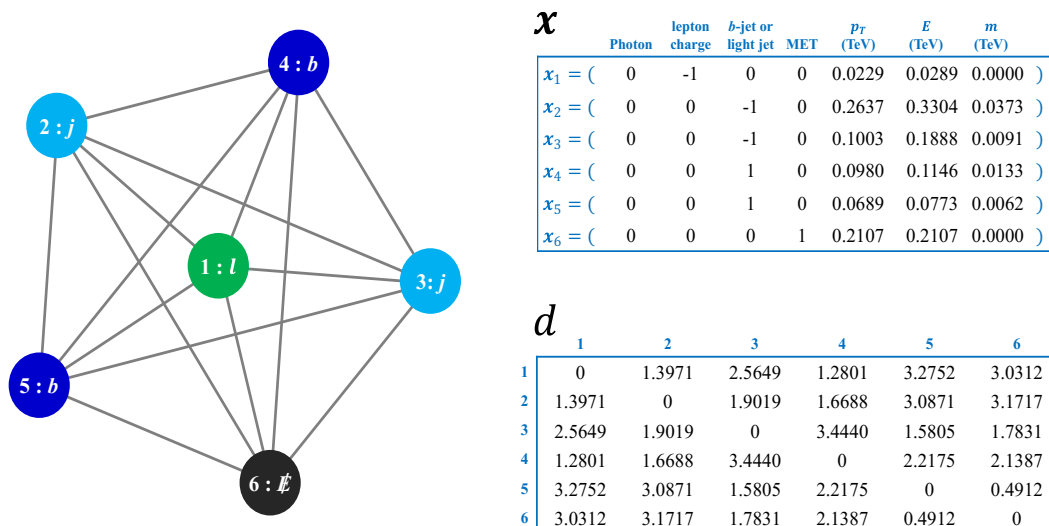
nearly three decades [44]. A great successful example in this direction is the use of Boosted Decision Trees [45] in the LHC experiment that led to the Higgs discovery. Recently, the ML techniques have been developed and applied for the studies of BSM phenomenology [46–52] increasingly.

When using ML to deal with collision events, one has to first construct a representation of event and then choose a ML model to analyze that representation. An event is usually described by a set of particles with certain kinematic features. The geometrical relationship between these particles in the event is a sensitive probe to distinguish the signal and background events. In mathematics, such a geometrical pattern of a number of entities can be represented by a graph, which can be numerically analyzed by ML algorithms. Among them, the Message Passing Neural Networks (MPNNs) [53] provide a general framework for supervised learning on graphs and are particularly suited for the problems where geometrical representation of patterns is to be learned. The MPNNs inherit the generality and powerfulness of the original Graph Neural Network (GNN) [54, 55] and improve the training efficiency. They are non-linear models with a bunch of parameters that relates the output to the input graphs, where the supervised learning finds optimized parameters. The MPNNs have been applied to jet physics [53] and other fields [56].

In this work, we first apply MPNN for the classification of signal and background events. Each HEP event is represented as an event graph, in which the nodes describe the final state particles and the edges reflect the geometric relations between each two nodes. In contrast with Deep Neural Network (DNN), MPNN is a dynamic neural network and intrinsically does not depend on the number and ordering of final state particles. Therefore, the MPNN is especially fit for processing the graph representation of collision event. As a proof-of-concept, we implement our method in the search of the stops through the process  $pp \rightarrow \tilde{t}_1 \tilde{t}_1^* \rightarrow t \bar{t} \tilde{\chi}_1^0 \tilde{\chi}_1^0$  at the LHC.

## 2 Methodology

A collision event usually produces a number of final state objects which are reconstructed as photons, leptons and jets with four-momentum information. To construct an event graph, we represent each final state object as a node and connect each pair of nodes by an edge, which constitutes an undirected weighted complete graph. Each node  $i$  has its feature vector  $\mathbf{x}_i$  and each edge between node  $i$  and  $j$  has its weight matrix  $d_{ij}$ . As an illustration, we show an event graph of a Monte Carlo simulated event of the process  $pp \rightarrow \tilde{t}_1 \tilde{t}_1^* \rightarrow t \bar{t} \tilde{\chi}_1^0 \tilde{\chi}_1^0 \rightarrow \ell + 2b + 2j + \cancel{E}_T$  in figure 1, where the detailed node feature vector  $\mathbf{x}$  and edge weights matrix  $d$  are also given. To be specific, the properties of the  $i$ -th final state are partly encoded into the 7-dimensional node feature vector  $\mathbf{x}_i$ . The first feature indicates that the final state is a photon (1) or not (0). If the final state is a lepton, its charge acts as the second feature; otherwise 0. The third feature indicates that the final state is a b-jet (1), light jet (-1) or not a jet (0). The fourth feature is used as an indicator of missing energy (MET) (1) or not (0). The rest three features are transverse momentum ( $p_T$ ), energy ( $E$ ) and mass ( $m$ ) of the object. By construction, such a feature vector is compact. The edges are weighted by pair-distances  $d_{ij} = \sqrt{\Delta y_{ij}^2 + \Delta \phi_{ij}^2}$  between



**Figure 1.** An event graph with detailed node features and distance matrix, built from a Monte Carlo simulated event of the process  $pp \rightarrow \tilde{t}_1 \tilde{t}_1^* \rightarrow t \bar{t} \tilde{\chi}_1^0 \tilde{\chi}_1^0 \rightarrow \ell + 2b + 2j + \cancel{E}_T$ .

$i$ -th and  $j$ -th node in the graph, where  $\Delta y$  and  $\Delta\phi$  are the rapidity difference and azimuthal angle difference, respectively. It should be mentioned that the azimuthal angle  $\phi$  is not encoded in the node features so that our graph representation is invariant under rotation in  $\phi$ , which is helpful for the stability of event classification.

In this paper, we design a variant of MPNN to carry out the graph classification, whose architecture is presented in figure 2. Firstly, we embed each node features  $\mathbf{x}_i$  into a higher dimensional state vector  $\mathbf{s}_i^{(0)}$ :

$$\mathbf{s}_i^{(0)} = f_e(\mathbf{x}_i), \tag{2.1}$$

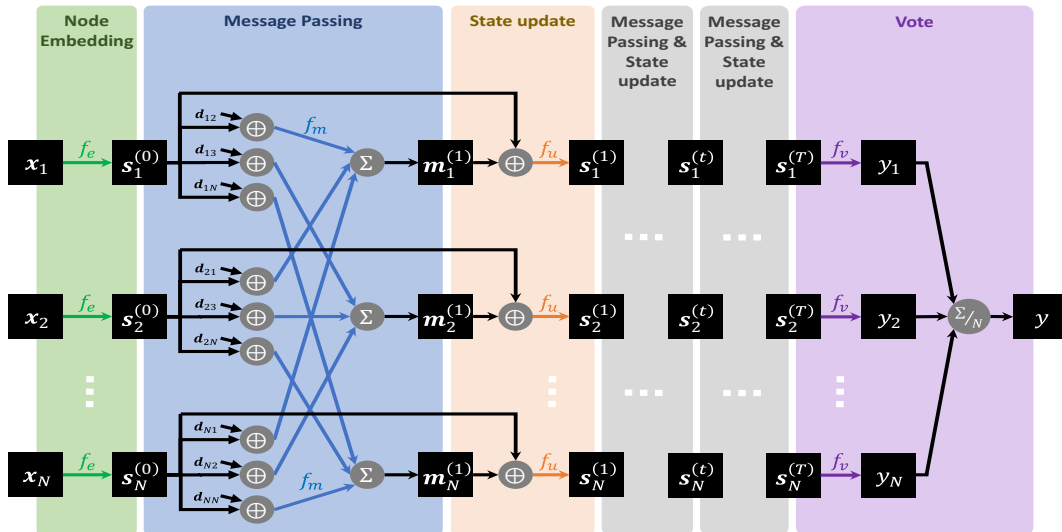
where  $f_e$  is called the node embedding function. The state vector  $\mathbf{s}_i^{(0)}$  only encodes the  $i$ -th node features  $\mathbf{x}_i$  without any information about the geometrical pattern of the graph. Then, the message passing techniques are utilized to perform event graph embedding, which will encode the whole event graph into each node state vector. At iteration  $t$ , each node  $i$  collects the messages sent from other nodes  $j$ :

$$\mathbf{m}_i^{(t)} = \sum_{j \neq i} \mathbf{m}_{i \leftarrow j}^{(t)} = \sum_{j \neq i} f_m^{(t)}(\mathbf{s}_j^{(t-1)}, d_{ij}), \tag{2.2}$$

and update its state vector

$$\mathbf{s}_i^{(t)} = f_u^{(t)}(\mathbf{s}_i^{(t-1)}, \mathbf{m}_i^{(t)}), \tag{2.3}$$

where  $f_m^{(t)}$  are the message functions and  $f_u^{(t)}$  are the update functions. By repeating this procedure, the information in node states and the distances between nodes are disseminated with the sent messages, and each node updates its knowledge of other nodes and the relationships between all nodes. Therefore, after  $T$  iterations, each resulting node state is an encoding of the whole graph, which is a compact representation of the information of both the kinematic features of all final states and the geometrical relationship between them.



**Figure 2.** The architecture of our message passing neural network (MPNN) designed for event graph classification. The network is a stack of functional layers shown as shadowed blocks, which has  $T$  pair of message passing and state update layers for automatic event feature extraction. State vectors  $\mathbf{s}$ , message vectors  $\mathbf{m}$ , votes  $y_i$  and discrimination score  $y$  are shown as black boxes. Colored arrows denote applying node embedding function  $f_e$ , message passing function  $f_m$ , state update function  $f_u$  and vote function  $f_v$ , respectively. The operators are given in gray circles, where  $\oplus$  performs vector concatenation,  $\Sigma$  and  $\Sigma/N$  are sum and average, respectively.

They are the event features that are automatically extracted from the input event graph. Next, each node votes a number as the likeness of the event to be signal-like, based on its own state vector,

$$y_i = f_v \left( \mathbf{s}_i^{(T)} \right), \quad (2.4)$$

where  $f_v$  is the vote function. Finally, to make the prediction stable, we average the votes from each node

$$y = \frac{1}{|\mathcal{V}|} \sum_i y_i \quad (2.5)$$

as the final discrimination score of the event, where  $|\mathcal{V}|$  is the number of nodes. The above operations form an end-to-end ML model, which maps event graphs directly to discrimination scores. The event selection can then be carried out by applying a specific cut  $\theta_y$  on the score  $y$ ; only events with  $y > \theta_y$  will be selected out.

In our following calculations, we use 30-dimensional state and message vectors, and choose single layer perceptrons as the node embedding, message passing, update and vote functions,

$$f_e(\mathbf{x}) = \text{relu}(W_e \mathbf{x} + \mathbf{b}_e), \quad (2.6)$$

$$f_m^{(t)}(\mathbf{s}, d) = \text{relu}\left(W_m^{(t)}(\mathbf{s} \oplus [d]) + \mathbf{b}_m^{(t)}\right), \quad (2.7)$$

$$f_u^{(t)}(\mathbf{s}, \mathbf{m}) = \text{relu}\left(W_u^{(t)}(\mathbf{s} \oplus \mathbf{m}) + \mathbf{b}_u^{(t)}\right), \quad (2.8)$$

$$f_v(\mathbf{s}) = \sigma(W_v \mathbf{s} + \mathbf{b}_v), \quad (2.9)$$

where  $\oplus$  denotes vector concatenation,  $\text{relu}$  is the rectified linear unit,  $\sigma$  is the sigmoid function,  $W$ s and  $\mathbf{b}$ s are learnable parameters. Independent message and update functions are used for each iteration  $t$  up to  $T = 2$ . Note that the pair distance  $d_{ij}$  is only a real number. To ease the learning of the message functions, we represent the distance  $d_{ij}$  in high dimensional vector space. Inspired from that radio basis function (RBF) networks can solve non-linear problems by mapping input into high dimensions using RBFs, we therefore choose the unnormalized Gaussian basis  $\exp\{(d - \mu_i)^2/2\delta^2\}$  as the RBFs to map the number  $d_{ij}$  into a vector  $[d]$ . By design, the mean values  $\mu_i$  of the RBFs are linearly distributed in the range of  $[0, 5]$  as  $\mu_1 = 0, \mu_2 = 0.25, \mu_3 = 0.5, \dots, \mu_{20} = 4.75, \mu_{21} = 5$ , and RBFs have the same width of  $\delta = 0.25$ . With such a method, we find that our result will become better, for example, the significance of benchmark point B can increase from 4.5 to 5.4 when  $d_{ij}$  is expanded from 0 to 21. Based on our practice, above choices are a good trade-off between model complexity and prediction accuracy. Our MPNN model totally have 7051 learnable parameters.

The MPNN can be efficiently trained using supervised learning techniques. We adopt binary-cross-entropy as the loss function. The Adam [57] optimizer with a learning rate of 0.001 is used to optimize the model parameters based on the gradients calculated on mini-batch of 500 training examples. A separate set of validation examples is used to measure the generalization performance while training to prevent over-fitting using the early-stopping technique. All these are implemented with the open-source deep learning framework PyTorch [58] with strong GPU acceleration.

Next, we compare the performance of MPNN with DNN in stop search at the LHC. Since the input size of DNN is fixed, we sort the nodes in each event graph by their identities and  $p_T$  and then arrange the node features and edge weights into a fixed-size feature vector of the form

$$\left[ \mathbf{x}_{\cancel{E}_T}^T, \mathbf{x}_\ell^T, \mathbf{x}_{b_1}^T, \mathbf{x}_{b_2}^T, \mathbf{x}_{j_1}^T, \dots, \mathbf{x}_{j_N}^T; d_{\cancel{E}_T \cancel{E}_T}, d_{\cancel{E}_T \ell}, \dots, d_{j_N j_N} \right], \quad (2.10)$$

where  $N = 17$  is the maximum number of light jets in our event graphs. Zero-padding is adopted to fill the missing values in the feature vector, namely for an event graph with  $n$  light jets,  $\mathbf{x}_{j_k} = \mathbf{0}$  ( $k > n$ ) and  $d_{j_k j_l} = 0$  ( $k$  or  $l > n$ ).

### 3 Numerical results and discussions

As a proof-of-concept, we apply MPNN to investigate the observability of the stop through the process  $pp \rightarrow \tilde{t}_1 \tilde{t}_1^* \rightarrow t \bar{t} \tilde{\chi}_1^0 \tilde{\chi}_1^0 \rightarrow \ell + 2b + 2j + \cancel{E}_T$  at 13 TeV LHC with the luminosity  $\mathcal{L} = 36.1 \text{ fb}^{-1}$ . We assume the LSP  $\tilde{\chi}_1^0$  is pure bino and focus on the kinematic region of  $m_{\tilde{t}_1} \geq m_t + m_{\tilde{\chi}_1^0}$ . The dominant background events in this analysis arise from  $t\bar{t}$ ,  $W + \text{jets}$  and  $tW$ . The  $t\bar{t}Z(\rightarrow \nu\bar{\nu})$  background is non-negligible for a heavy stop and is included in our calculations as well. The multi-jet background can be estimated from data using a fake-factor method, which is found to be negligible in all regions [59].

We use the event generator MadGraph5\_aMC@NLO [60] to simulate the signal and background events at the parton-level. Then we carry out the parton shower and hadronization with the Pythia8.2 [61]. Delphes-3.4.1 [62] is used for fast detector simulation. The anti- $k_t$  algorithm [63] with the distance parameter  $R = 0.4$  is chosen to cluster jets, and

Benchmark	A	B
$m_{\tilde{t}_1}$ (GeV)	525	900
$m_{\tilde{\chi}_1^0}$ (GeV)	352	330
$Z(\text{MPNN})$	$4.6\sigma$	$5.4\sigma$
$Z(\text{DNN})$	$3.0\sigma$	$4.4\sigma$
$Z(\text{MPNN6})$	$3.5\sigma$	$4.3\sigma$
$Z(\text{DNN6})$	$3.2\sigma$	$4.3\sigma$

**Table 1.** The comparison of MPNN with DNN for significance ( $Z$ ) of two benchmark points at 13 TeV LHC with the luminosity of  $\mathcal{L} = 36.1 \text{ fb}^{-1}$ . MPNN6 and DNN6 denote the results that use six objects (one lepton, two b-jets, two leading light-jets and MET) as inputs, respectively.

the  $b$ -tagging efficiency is assumed as 80%. In the end, the event preselections are performed by CheckMATE-2.0.14 [64] using the following pre-selection cuts.

- We require exactly one lepton with  $p_T(\ell) > 10 \text{ GeV}$  and  $|\eta(\ell)| < 2.5$ , and at least four jets with  $p_T(j) > 25 \text{ GeV}$  and  $|\eta| < 2.5$ .
- We also require exactly two  $b$ -jets in the events.
- The transverse missing energy should satisfy  $\cancel{E}_T > 150 \text{ GeV}$ .

The NLO QCD corrected cross section of stop pair production is calculated with the Prospino [65]. The  $t\bar{t}$  and  $W$ +jets events are further normalized with their NNLO cross-sections, respectively [66, 67]. We evaluate the statistical significance with the formula,

$$Z = S/\sqrt{B}. \tag{3.1}$$

To guarantee the statistics, we require at least 10 events after imposing cuts.

In table 1, we choose two Benchmark Points(BPs) with distinctive kinematic features to compare the performance of MPNN and DNN. The BP-A lies in the compressed region with  $m_{\tilde{t}_1} \approx m_t + m_{\tilde{\chi}_1^0}$ , while the BP-B lies in the uncompressed region with  $m_{\tilde{t}_1} \gg m_t + m_{\tilde{\chi}_1^0}$ . For BP-A/B, we generated 14/4 million signal events and 100 million background events. After the pre-selection, 300,000 signal events and 300,000 background events are collected as training examples. We also collect a separate set of 100,000 signal events and 100,000 background events as validation examples. In addition, we also show the results of MPNN6 and DNN6, which use six objects (one lepton, two b-jets, two leading light-jets and MET) used as inputs, respectively. The reduced event graphs for MPNN6 have 6 nodes and 36 edges. The input feature vectors for DNN6 are of the form of eq. (2.10) with  $N = 2$  so that it has 78 input features.

In order to optimize the DNN results, we perform a grid scan of hyperparameters for BP-A and B. In our scan, we find that the significance generally increases with more hidden layers and neurons per hidden layer. But the increasing complexity of the DNN model limits the improvement. The best DNN model for BP-A/B has 7/4 hidden layers, 700/300



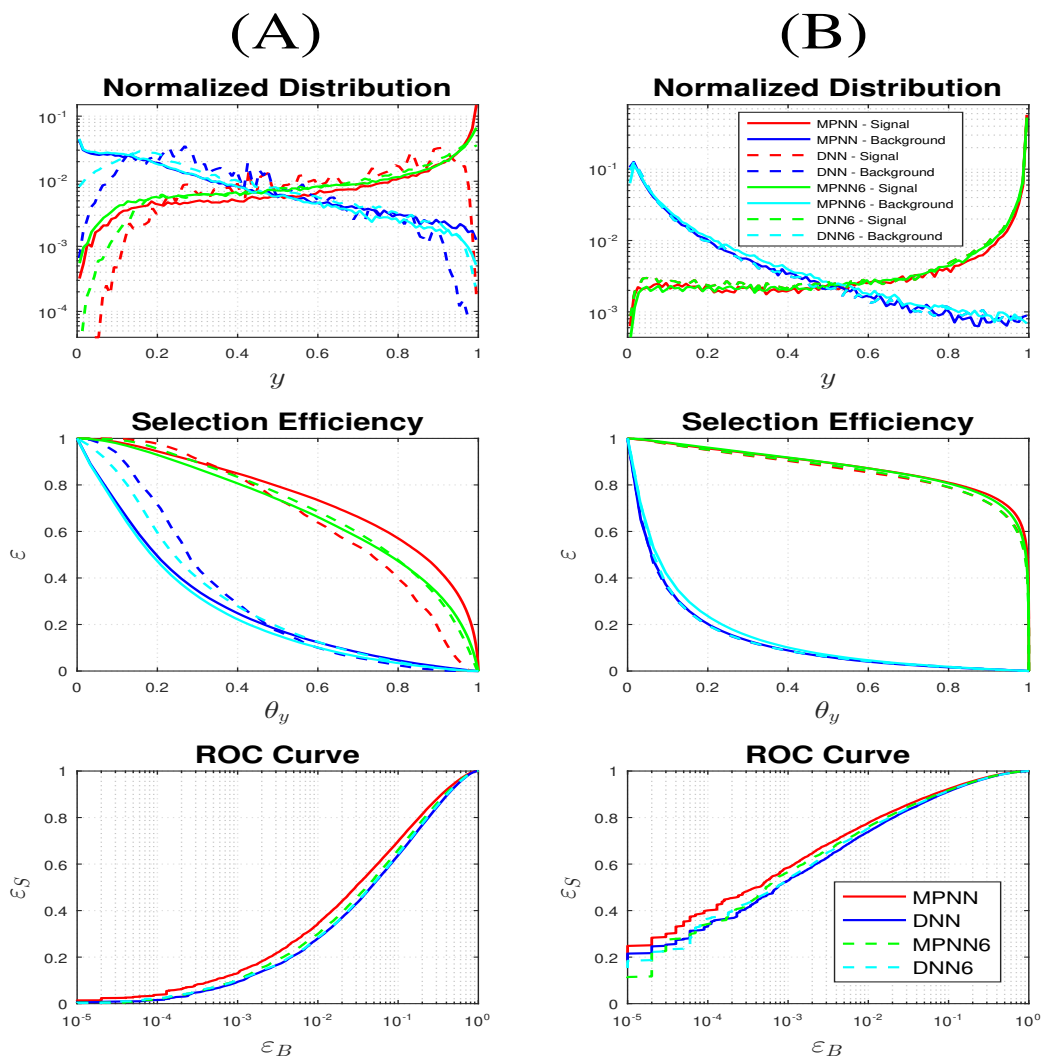
neurons per hidden layer and a dropout ratio of 0.5/0.4, which owns 3357201/447901 learning parameters and gives the best significance of  $3.0/4.4\sigma$ . The same optimization procedure is adopted for DNN6 model in table 1 and also for each sample in figure 4.

From table 1, we can see that the MPNN method have better significance than DNN for both BPs, in particular for the BP-A with  $m_{\tilde{t}_1} \approx m_{\tilde{\chi}_1^0} + m_t$ . The significance of BP-A increases from  $3.2\sigma$  (DNN6) to  $3.5\sigma$  (MPNN6), and  $3.0\sigma$  (DNN) to  $4.6\sigma$  (MPNN). With more input features, MPNN outperforms over MPNN6, while DNN has no such a feature, since more learnable parameters usually leads to more serious over-fitting for DNN.

In figure 3, we further present the discriminating power of MPNNs and DNNs on signal and background events for benchmark points A and B. Since the over-fitting is very small, we only show the results on validation set. From the top panel, we can see that the signal and background events are well separated in the distributions of the discrimination score for both of MPNNs and DNNs. But for MPNNs, the score of signals are inclined to have larger value than DNNs, while the score of backgrounds have smaller values than DNNs. From the middle panel, we can find that MPNNs has higher signal selection efficiency  $\varepsilon_S$  and lower background selection efficiency  $\varepsilon_B$  than DNNs. This leads to MPNNs having sharper receiver operating characteristic (ROC) curves than DNNs, as shown in the bottom panel of figure 3.

In figure 4, we perform a grid scan on the plane of  $m_{\tilde{t}_1}$  versus  $m_{\tilde{\chi}_1^0}$  by using MPNN(6) and DNN(6) at 13 TeV LHC with the luminosity  $\mathcal{L} = 36.1 \text{ fb}^{-1}$ . We take the right-handed  $\tilde{t}_1$  as the example and simulate 40 parameter points in our grid scan. The pre-selections for signal and background events are the same as those for the benchmark points in table 1. We optimize the DNN(6) and MPNN(6) models for each parameter point as the approaches used for benchmarks points. As a theoretical work, we define the exclusion limit when the significance  $Z = 2\sigma$ . Then, the exclusion line is drawn by using bilinear interpolation. We can see that the MPNN method can produce a stronger exclusion limit than the DNN method. For examples, in the compressed region with  $m_{\tilde{t}_1} \approx m_{\tilde{\chi}_1^0} + m_t$ , the stop mass reach from MPNN can be 670 GeV, which is about 90 GeV larger than that from the DNN. In other regions with  $m_{\tilde{t}_1} > m_{\tilde{\chi}_1^0} + m_t$ , the exclusion limit on stop mass for given neutralino mass from the MPNN can be greater than DNN by about several tens of GeV to more than one hundred GeV. Unlike the MPNN, the results of the MPNN6 are not much better than that of the DNN6. This is because that the man-made truncation on the number of final states in the MPNN6 will badly worsen its performance. In the MPNN6, the kinematical information coded in the event graphs for some parameter points having large multiplicity are incomplete. Thus the full correlations among final states, which is the most important factor for discriminating signal and backgrounds in the MPNN method, are broken. On the other hand, the performance of DNN seems worse than that of DNN6 in most of region in figure 4. This is because the DNN involves much more training parameters so that it has poorer generalization capability. Therefore, we expect that the MPNN approach will have better performance than the DNN approach when the physical processes have high-multiplicity final states at the LHC.

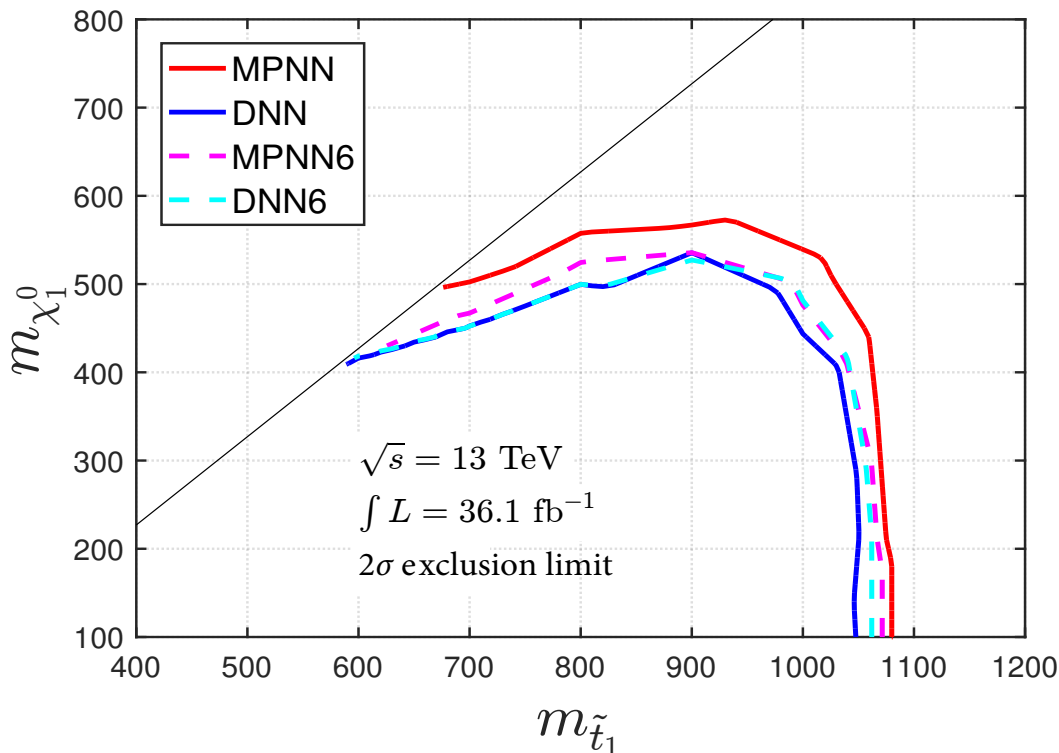




**Figure 3.** The discriminating power of MPNN(6) and DNN(6) on signal and background for benchmark points A (left panel) and B (right panel) in table 1. The top panel is the distributions of discrimination score  $y$  for signal and background events. The middle panel is the selection efficiency  $\epsilon$  versus the cut threshold  $\theta_y$ . The bottom panel shows ROC curves of the MPNN(6) and DNN(6).

## 4 Conclusions

In this paper, we proposed to represent a collision event in high energy physics as an event graph with a set of nodes and edges, and use the Message Passing Neural Networks to deal with the problems of discrimination of signal and background events at colliders. As a proof of concept, we applied our approach to the search for the stop pair production at the LHC. We found that the geometrical pattern of the event that depends on correlations among final states can be a sensitive feature in the MPNN to discriminate signal and background events. Furthermore, we compared the performance of our MPNN with the DNN, and found that our MPNN can enhance the mass reach of stop mass by several tens of GeV to over a hundred GeV.



**Figure 4.** Exclusion limits on the plane of  $m_{\tilde{t}_1}$  versus  $m_{\tilde{\chi}_1^0}$  by using MPNN(6) and DNN(6) at 13 TeV LHC with the luminosity of  $\mathcal{L} = 36.1 \text{ fb}^{-1}$ .

### Acknowledgments

Part of this work was done while M.A. was visiting Nanjing Normal University. This work was supported by the National Natural Science Foundation of China (NNSFC) under grant Nos. 11705093, 11675242 and 11851303, by Peng-Huan-Wu Theoretical Physics Innovation Center (11747601), by the CAS Center for Excellence in Particle Physics (CCEPP), by the CAS Key Research Program of Frontier Sciences and by a Key R&D Program of Ministry of Science and Technology under number 2017YFA0402200-04.

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

### References

- [1] C.G. Lester and D.J. Summers, *Measuring masses of semiinvisibly decaying particles pair produced at hadron colliders*, *Phys. Lett. B* **463** (1999) 99 [[hep-ph/9906349](#)] [[INSPIRE](#)].
- [2] A. Barr, C. Lester and P. Stephens,  *$m(T2)$ : The truth behind the glamour*, *J. Phys. G* **29** (2003) 2343 [[hep-ph/0304226](#)] [[INSPIRE](#)].
- [3] Y. Bai, H.-C. Cheng, J. Gallicchio and J. Gu, *Stop the Top Background of the Stop Search*, *JHEP* **07** (2012) 110 [[arXiv:1203.4813](#)] [[INSPIRE](#)].

- [4] J. Cao, C. Han, L. Wu, J.M. Yang and Y. Zhang, *Probing Natural SUSY from Stop Pair Production at the LHC*, *JHEP* **11** (2012) 039 [[arXiv:1206.3865](#)] [[INSPIRE](#)].
- [5] C. Kilic and B. Tweedie, *Cornering Light Stops with Dileptonic  $mT_2$* , *JHEP* **04** (2013) 110 [[arXiv:1211.6106](#)] [[INSPIRE](#)].
- [6] K. Hagiwara and T. Yamada, *Equal-velocity scenario for hiding dark matter at the LHC*, *Phys. Rev. D* **91** (2015) 094007 [[arXiv:1307.1553](#)] [[INSPIRE](#)].
- [7] H. An and L.-T. Wang, *Opening up the compressed region of top squark searches at 13 TeV LHC*, *Phys. Rev. Lett.* **115** (2015) 181602 [[arXiv:1506.00653](#)] [[INSPIRE](#)].
- [8] S. Macaluso, M. Park, D. Shih and B. Tweedie, *Revealing Compressed Stops Using High-Momentum Recoils*, *JHEP* **03** (2016) 151 [[arXiv:1506.07885](#)] [[INSPIRE](#)].
- [9] M. Czakon, A. Mitov, M. Papucci, J.T. Ruderman and A. Weiler, *Closing the stop gap*, *Phys. Rev. Lett.* **113** (2014) 201803 [[arXiv:1407.1043](#)] [[INSPIRE](#)].
- [10] Z. Han, A. Katz, D. Krohn and M. Reece, *(Light) Stop Signs*, *JHEP* **08** (2012) 083 [[arXiv:1205.5808](#)] [[INSPIRE](#)].
- [11] A. Djouadi and Y. Mambrini, *Three body decays of top and bottom squarks*, *Phys. Rev. D* **63** (2001) 115005 [[hep-ph/0011364](#)] [[INSPIRE](#)].
- [12] T. Han, K.-i. Hikasa, J.M. Yang and X.-m. Zhang, *The FCNC top squark decay as a probe of squark mixing*, *Phys. Rev. D* **70** (2004) 055001 [[hep-ph/0312129](#)] [[INSPIRE](#)].
- [13] M. Muhlleitner and E. Popeno, *Light Stop Decay in the MSSM with Minimal Flavour Violation*, *JHEP* **04** (2011) 095 [[arXiv:1102.5712](#)] [[INSPIRE](#)].
- [14] J. Aebischer, A. Crivellin and C. Greub, *One-loop SQCD corrections to the decay of top squarks to charm and neutralino in the generic MSSM*, *Phys. Rev. D* **91** (2015) 035010 [[arXiv:1410.8459](#)] [[INSPIRE](#)].
- [15] C. Boehm, A. Djouadi and Y. Mambrini, *Decays of the lightest top squark*, *Phys. Rev. D* **61** (2000) 095006 [[hep-ph/9907428](#)] [[INSPIRE](#)].
- [16] M.A. Ajaib, T. Li and Q. Shafi, *Stop-Neutralino Coannihilation in the Light of LHC*, *Phys. Rev. D* **85** (2012) 055021 [[arXiv:1111.4467](#)] [[INSPIRE](#)].
- [17] M. Drees, M. Hanussek and J.S. Kim, *Light Stop Searches at the LHC with Monojet Events*, *Phys. Rev. D* **86** (2012) 035024 [[arXiv:1201.5714](#)] [[INSPIRE](#)].
- [18] Z.-H. Yu, X.-J. Bi, Q.-S. Yan and P.-F. Yin, *Detecting light stop pairs in coannihilation scenarios at the LHC*, *Phys. Rev. D* **87** (2013) 055007 [[arXiv:1211.2997](#)] [[INSPIRE](#)].
- [19] M. Perelstein and A. Weiler, *Polarized Tops from Stop Decays at the LHC*, *JHEP* **03** (2009) 141 [[arXiv:0811.1024](#)] [[INSPIRE](#)].
- [20] T. Plehn, M. Spannowsky and M. Takeuchi, *Stop searches in 2012*, *JHEP* **08** (2012) 091 [[arXiv:1205.2696](#)] [[INSPIRE](#)].
- [21] C. Han, K.-i. Hikasa, L. Wu, J.M. Yang and Y. Zhang, *Current experimental bounds on stop mass in natural SUSY*, *JHEP* **10** (2013) 216 [[arXiv:1308.5307](#)] [[INSPIRE](#)].
- [22] M.R. Buckley, T. Plehn and M.J. Ramsey-Musolf, *Top squark with mass close to the top quark*, *Phys. Rev. D* **90** (2014) 014046 [[arXiv:1403.2726](#)] [[INSPIRE](#)].
- [23] D. Goncalves, D. Lopez-Val, K. Mawatari and T. Plehn, *Automated third generation squark production to next-to-leading order*, *Phys. Rev. D* **90** (2014) 075007 [[arXiv:1407.4302](#)] [[INSPIRE](#)].

- [24] B. Fuks, P. Richardson and A. Wilcock, *Studying the sensitivity of monotop probes to compressed supersymmetric scenarios at the LHC*, *Eur. Phys. J. C* **75** (2015) 308 [[arXiv:1408.3634](#)] [[INSPIRE](#)].
- [25] T. Eifert and B. Nachman, *Sneaky light stop*, *Phys. Lett. B* **743** (2015) 218 [[arXiv:1410.7025](#)] [[INSPIRE](#)].
- [26] A. Kobakhidze, N. Liu, L. Wu and J.M. Yang, *ATLAS Z-peaked excess in the MSSM with a light sbottom or stop*, *Phys. Rev. D* **92** (2015) 075008 [[arXiv:1504.04390](#)] [[INSPIRE](#)].
- [27] K.-i. Hikasa, J. Li, L. Wu and J.M. Yang, *Single top squark production as a probe of natural supersymmetry at the LHC*, *Phys. Rev. D* **93** (2016) 035003 [[arXiv:1505.06006](#)] [[INSPIRE](#)].
- [28] A. Kobakhidze, N. Liu, L. Wu, J.M. Yang and M. Zhang, *Closing up a light stop window in natural SUSY at LHC*, *Phys. Lett. B* **755** (2016) 76 [[arXiv:1511.02371](#)] [[INSPIRE](#)].
- [29] H.-C. Cheng, C. Gao, L. Li and N.A. Neill, *Stop Search in the Compressed Region via Semileptonic Decays*, *JHEP* **05** (2016) 036 [[arXiv:1604.00007](#)] [[INSPIRE](#)].
- [30] C. Han, J. Ren, L. Wu, J.M. Yang and M. Zhang, *Top-squark in natural SUSY under current LHC run-2 data*, *Eur. Phys. J. C* **77** (2017) 93 [[arXiv:1609.02361](#)] [[INSPIRE](#)].
- [31] G.H. Duan, K.-i. Hikasa, L. Wu, J.M. Yang and M. Zhang, *Leptonic mono-top from single stop production at the LHC*, *JHEP* **03** (2017) 091 [[arXiv:1611.05211](#)] [[INSPIRE](#)].
- [32] P. Jackson, C. Rogan and M. Santoni, *Sparticles in motion: Analyzing compressed SUSY scenarios with a new method of event reconstruction*, *Phys. Rev. D* **95** (2017) 035031 [[arXiv:1607.08307](#)] [[INSPIRE](#)].
- [33] D. Goncalves, K. Sakurai and M. Takeuchi, *Tagging a monotop signature in natural SUSY*, *Phys. Rev. D* **95** (2017) 015030 [[arXiv:1610.06179](#)] [[INSPIRE](#)].
- [34] A. Butter, G. Kasieczka, T. Plehn and M. Russell, *Deep-learned Top Tagging with a Lorentz Layer*, *SciPost Phys.* **5** (2018) 028 [[arXiv:1707.08966](#)] [[INSPIRE](#)].
- [35] Z. Kang, J. Li and M. Zhang, *Uncover Compressed Supersymmetry via Boosted Bosons from the Heavier Stop/Sbottom*, *Eur. Phys. J. C* **77** (2017) 371 [[arXiv:1703.08911](#)] [[INSPIRE](#)].
- [36] G.H. Duan, L. Wu and R. Zheng, *Resonant Higgs pair production as a probe of stop at the LHC*, *JHEP* **09** (2017) 037 [[arXiv:1706.07562](#)] [[INSPIRE](#)].
- [37] H. Baer, V. Barger, J.S. Gainer, H. Serce and X. Tata, *Reach of the high-energy LHC for gluinos and top squarks in SUSY models with light Higgsinos*, *Phys. Rev. D* **96** (2017) 115008 [[arXiv:1708.09054](#)] [[INSPIRE](#)].
- [38] ATLAS collaboration, *Search for supersymmetry in final states with charm jets and missing transverse momentum in 13 TeV pp collisions with the ATLAS detector*, *JHEP* **09** (2018) 050 [[arXiv:1805.01649](#)] [[INSPIRE](#)].
- [39] ATLAS collaboration, *Search for direct top squark pair production in final states with two leptons in  $\sqrt{s} = 13$  TeV pp collisions with the ATLAS detector*, *Eur. Phys. J. C* **77** (2017) 898 [[arXiv:1708.03247](#)] [[INSPIRE](#)].
- [40] ATLAS collaboration, *Search for a scalar partner of the top quark in the jets plus missing transverse momentum final state at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *JHEP* **12** (2017) 085 [[arXiv:1709.04183](#)] [[INSPIRE](#)].
- [41] CMS collaboration, *Search for new phenomena with the  $M_{T2}$  variable in the all-hadronic final state produced in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *Eur. Phys. J. C* **77** (2017) 710 [[arXiv:1705.04650](#)] [[INSPIRE](#)].

- [42] CMS collaboration, *Search for top squark pair production in pp collisions at  $\sqrt{s} = 13$  TeV using single lepton events*, *JHEP* **10** (2017) 019 [[arXiv:1706.04402](#)] [[INSPIRE](#)].
- [43] CMS collaboration, *Search for top squarks decaying via four-body or chargino-mediated modes in single-lepton final states in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *JHEP* **09** (2018) 065 [[arXiv:1805.05784](#)] [[INSPIRE](#)].
- [44] P.C. Bhat, *Multivariate Analysis Methods in Particle Physics*, *Ann. Rev. Nucl. Part. Sci.* **61** (2011) 281 [[INSPIRE](#)].
- [45] B.P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu and G. McGregor, *Boosted decision trees, an alternative to artificial neural networks*, *Nucl. Instrum. Meth. A* **543** (2005) 577 [[physics/0408124](#)] [[INSPIRE](#)].
- [46] P. Baldi, P. Sadowski and D. Whiteson, *Searching for Exotic Particles in High-Energy Physics with Deep Learning*, *Nature Commun.* **5** (2014) 4308 [[arXiv:1402.4735](#)] [[INSPIRE](#)].
- [47] P. Baldi, P. Sadowski and D. Whiteson, *Enhanced Higgs Boson to  $\tau^+\tau^-$  Search with Deep Learning*, *Phys. Rev. Lett.* **114** (2015) 111801 [[arXiv:1410.3469](#)] [[INSPIRE](#)].
- [48] M. Bridges, K. Cranmer, F. Feroz, M. Hobson, R. Ruiz de Austri and R. Trotta, *A Coverage Study of the CMSSM Based on ATLAS Sensitivity Using Fast Neural Networks Techniques*, *JHEP* **03** (2011) 012 [[arXiv:1011.4306](#)] [[INSPIRE](#)].
- [49] A. Buckley, A. Shilton and M.J. White, *Fast supersymmetry phenomenology at the Large Hadron Collider using machine learning techniques*, *Comput. Phys. Commun.* **183** (2012) 960 [[arXiv:1106.4613](#)] [[INSPIRE](#)].
- [50] N. Bornhauser and M. Drees, *Determination of the CMSSM Parameters using Neural Networks*, *Phys. Rev. D* **88** (2013) 075016 [[arXiv:1307.3383](#)] [[INSPIRE](#)].
- [51] S. Caron, J.S. Kim, K. Rolbiecki, R. Ruiz de Austri and B. Stienen, *The BSM-AI project: SUSY-AI-generalizing LHC limits on supersymmetry with machine learning*, *Eur. Phys. J. C* **77** (2017) 257 [[arXiv:1605.02797](#)] [[INSPIRE](#)].
- [52] G. Bertone, M.P. Deisenroth, J.S. Kim, S. Liem, R. Ruiz de Austri and M. Welling, *Accelerating the BSM interpretation of LHC data with machine learning*, *Phys. Dark Univ.* **24** (2019) 100293 [[arXiv:1611.02704](#)] [[INSPIRE](#)].
- [53] J. Gilmer, S.S. Schoenholz, P.F. Riley, O. Vinyals and G.E. Dahl, *Neural Message Passing for Quantum Chemistry*, [arXiv:1704.01212](#).
- [54] M. Gori, G. Monfardini and F. Scarselli, *A new model for learning in graph domains*, in proceedings of the *2005 IEEE International Joint Conference on Neural Networks*, Montreal, Que., Canada, 31 July – 4 August 2005, [<https://doi.org/10.1109/IJCNN.2005.1555942>].
- [55] F. Scarselli, M. Gori, A.C. Tsoi, M. Hagenbuchner and G. Monfardini, *The Graph Neural Network Model*, *IEEE Trans. Neural Networks* **20** (2009) 61.
- [56] S. Farrell et al., *Particle Track Reconstruction with Deep Learning*, proceedings of the *Deep Learning for Physical Sciences Workshop at NIPS 17*, Long Beach, CA, U.S.A., 2017.
- [57] D.P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, [arXiv:1412.6980](#) [[INSPIRE](#)].
- [58] <http://pytorch.org/>.
- [59] ATLAS collaboration, *Search for top-squark pair production in final states with one lepton, jets and missing transverse momentum using  $36\text{ fb}^{-1}$  of  $\sqrt{s} = 13$  TeV pp collision data with the ATLAS detector*, *JHEP* **06** (2018) 108 [[arXiv:1711.11520](#)] [[INSPIRE](#)].

- [60] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [[arXiv:1405.0301](#)] [[INSPIRE](#)].
- [61] T. Sjöstrand et al., *An Introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [[arXiv:1410.3012](#)] [[INSPIRE](#)].
- [62] DELPHES 3 collaboration, *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057 [[arXiv:1307.6346](#)] [[INSPIRE](#)].
- [63] M. Cacciari, G.P. Salam and G. Soyez, *The anti- $k_t$  jet clustering algorithm*, *JHEP* **04** (2008) 063 [[arXiv:0802.1189](#)] [[INSPIRE](#)].
- [64] M. Drees, H. Dreiner, D. Schmeier, J. Tattersall and J.S. Kim, *CheckMATE: Confronting your Favourite New Physics Model with LHC Data*, *Comput. Phys. Commun.* **187** (2015) 227 [[arXiv:1312.2591](#)] [[INSPIRE](#)].
- [65] W. Beenakker, M. Klasen, M. Krämer, T. Plehn, M. Spira and P.M. Zerwas, *The production of charginos/neutralinos and sleptons at hadron colliders*, *Phys. Rev. Lett.* **83** (1999) 3780 [*Erratum ibid.* **100** (2008) 029901] [[hep-ph/9906298](#)] [[INSPIRE](#)].
- [66] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, *Comput. Phys. Commun.* **185** (2014) 2930 [[arXiv:1112.5675](#)] [[INSPIRE](#)].
- [67] R. Boughezal, C. Focke, X. Liu and F. Petriello, *W-boson production in association with a jet at next-to-next-to-leading order in perturbative QCD*, *Phys. Rev. Lett.* **115** (2015) 062002 [[arXiv:1504.02131](#)] [[INSPIRE](#)].