

Topological reconstruction of particle physics processes using graph neural networks

Lukas Ehrke^{1,*}, John Andrew Raine^{1,†}, Knut Zoch^{1,2}, Manuel Guth¹ and Tobias Golling¹

¹*Département de physique nucléaire et corpusculaire, University of Geneva, 1211 Geneva, Switzerland*

²*Laboratory for Particle Physics and Cosmology, Harvard University, Cambridge, Massachusetts 02138, USA*



(Received 27 March 2023; accepted 22 May 2023; published 23 June 2023)

We present a new approach, the Topograph, which reconstructs underlying physics processes, including the intermediary particles, by leveraging underlying priors from the nature of particle physics decays and the flexibility of message passing graph neural networks. The Topograph not only solves the combinatoric assignment of observed final state objects, associating them to their original mother particles, but directly predicts the properties of intermediate particles in hard scatter processes and their subsequent decays. In comparison to standard combinatoric approaches or modern approaches using graph neural networks, which scale exponentially or quadratically, the complexity of Topographs scales linearly with the number of reconstructed objects. We apply Topographs to top quark pair production in the all hadronic decay channel, where we outperform the standard approach and match the performance of the state-of-the-art machine learning technique.

DOI: [10.1103/PhysRevD.107.116019](https://doi.org/10.1103/PhysRevD.107.116019)

I. INTRODUCTION

At the Large Hadron Collider (LHC) at CERN, beams of protons are collided together at incredibly high energies to probe the underlying nature of the universe. From the objects recorded by the detectors on the LHC, the underlying physics processes in the collisions are attempted to be reconstructed.

In processes with a high multiplicity of objects, reconstructing intermediate particles, such as the two top quarks and W bosons in top quark pair production, is a crucial and challenging component in analyses [1–17]. This is of particular interest when measuring the mass of an intermediary particle, or its kinematic properties as part of a cross section measurement.

In this work we introduce Topographs, a new machine learning approach for reconstructing the full hypothesized decay chain from observed objects. It leverages state-of-the-art machine learning (ML) with underlying priors from particle physics, through the nature of particle decays, to reconstruct underlying physics processes including the intermediary particles. The computational complexity of Topographs scales

linearly with object multiplicity, whereas alternative methods scale quadratically [18,19] or exponentially [20,21].

We apply Topographs to reconstruct the underlying processes in the production of top quark pairs in the all hadronic decay channel. However, the architecture can be applied to any particle physics process and is not limited to event level reconstruction.

II. MOTIVATION

In the case of top quark pair production, with each top quark decaying hadronically $t \rightarrow Wb \rightarrow qq'b$, each quark is expected to initiate a shower in the detector which is reconstructed as a jet, resulting in six jets. The number of combinations to match the reconstructed jets to quarks from the top quark pair system is computationally intractable and grows exponentially with additional jets reconstructed in the final state, even when taking underlying symmetries into account. Solving the combinatorics of this system is a key challenge in the measurement of the top quark and $t\bar{t}$ pair, and one which is often computationally limited. The combinatorics can be restricted by looking at event topologies where the high momentum of the top quarks results in all three quarks being reconstructed in a single large radius jet, however this restricts the phase space to such topologies which represent only a small fraction of all events.

III. CURRENT APPROACHES

In top quark physics kinematic event reconstruction forms a key part of many measurements. The χ^2 method [4] and

*lukas.ehrke@unige.ch

†john.raine@unige.ch

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

the kinematic likelihood fitter (KLFitter) [20] have been employed in a large number of analyses [1–14]. In both approaches, all combinatorics of jet matching to final state quarks and gluons (partons) in the $t\bar{t}$ final state are tested with kinematic constraints based on the masses of the reconstructed W bosons and top quarks as minimization criteria. In the case of KLFitter these are used in conjunction with transfer functions and the particle decay widths.

Although good performance can be achieved with such an approach, as the number of jets in an event increases, as well as the multiplicity of final state objects to be reconstructed, the number of combinations increases exponentially. For example, in an all hadronic top quark pair event with six jets, there are 720 potential combinations. This is reduced to 90 by exploiting underlying symmetries. However, for events with seven jets the combinations increase to 630, and for eight jets they increase further to 2520. Furthermore, as the exact values of the mass of the top quark and W boson are used to test the likelihood of a combination, this leads to a biased estimator which focuses on assigning jets which together are closest to the hypothesized particles mass, rather than exploiting all the information about the pairs or triplets of objects. It also assumes that in all cases the top quarks and W bosons are on shell.

Building on the previous combinatoric approaches, simple approaches using machine learning (ML) have been developed. Instead of finding the most probable assignment using just the masses of intermediary particles, machine learning discriminants are used to identify correct assignments, exploiting more information from the event [21]. Nonetheless, these approaches still suffer from the same problems as the KLFitter and χ^2 methods, with each combination needing to be tested to identify the most likely.

Another approach which uses more information from the event is the matrix element method (MEM) [15,21–23]. The MEM not only attempts to match objects to the final state objects in an event, but directly assesses the likelihood of observing an event given the matrix element for a process. This can be evaluated for each potential combination with the highest resulting probability chosen as the correct assignment. However, it is extremely slow and computationally intensive. To calculate the likelihood of an event, an integral over the whole phase space of possible final state particle momenta must be performed. It is also reliant on a transfer function, which is used to convert the jets, charged leptons and missing transverse momentum recorded by the detector to the partons, charged leptons and neutrinos before any hadronization and detector effects. As there is no accurate function to model this, it is at best an approximation optimized by hand. Normalizing flows present a solution to the computational challenge and approximate functions [24], however do not yet address the combinatoric solving.

A. State of the art

The state of the art machine learning approach uses attention transformers [19,25,26] to identify the indices of final state objects coming from intermediate particles. In this approach no graph structure is used and only the permutation invariant collection of objects is considered. The complexity of the approach can be reduced by taking into account the symmetries, as performed in Refs. [19,26] (SPA-Net), corresponding to removing potential solutions in the combinatoric approaches, which leads to an overall complexity of $\mathcal{O}(N^2)$.

Graph neural networks (GNNs) [27,28] are also employed in High energy physics to associate objects to a common origin, for example in secondary vertex reconstruction [18] and could similarly be applied to combinatoric solving at the event level. These approaches have fully connected graphs with $N(N-1)$ edges.

In addition to their reduced computational complexity in comparison to traditional approaches, both attention and GNN approaches also demonstrate reduction in biases towards particle masses, as often seen in the combinatoric approaches. However, in both GNNs and SPA-Net the target is to identify the two triplets of objects which correspond to the decay of each top quark, neglecting the structure of the decay, and the properties of the intermediary particles.

Other approaches employ physics inspired layers in order to assign parton labels [29] or try to predict the properties of intermediate particles directly [30].

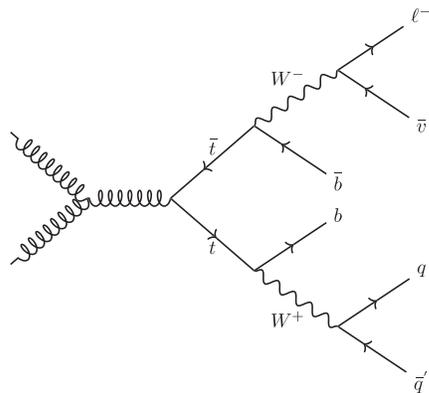
IV. THE TOPOGRAPH

The use of GNNs in high energy physics applications [31–44] is a recent development which is gaining in popularity. However, graphs have also long been used to describe underlying processes occurring in particle physics in the form of Feynman diagrams.

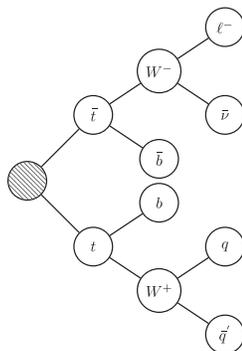
In a Feynman diagram, the vertices (nodes) represent the interactions between particles and the edges represent the particles themselves. This representation can equally be converted into a node-and-edge graph by representing the particles as nodes, and defining edges based on the vertices. An example of the Feynman diagram and one such graph for top quark pair production is shown in Fig. 1. In current graph based approaches this physical inspired graph representation is not exploited, and instead fully connected graphs are constructed from all objects recorded by the detector.

We introduce a new method, called the Topograph, which builds upon the structure seen in Fig. 1(b) to define neural networks which can be used to predict the correct edges from final state particles and the properties of intermediate particles.

First, the intermediate particles in a chosen physics process are injected into the graph. Then, instead of connecting all objects to one another, they are instead



(a) Feynman diagram



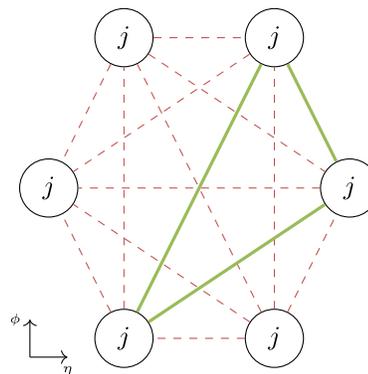
(b) Node and edge graph

FIG. 1. Two graph representations of top pair production, with one top decaying semileptonically and the other hadronically. The production mechanism in (a) is shown to be gluon fusion however in (b) it is represented by the hashed circle. Time flows from left to right in both graphs.

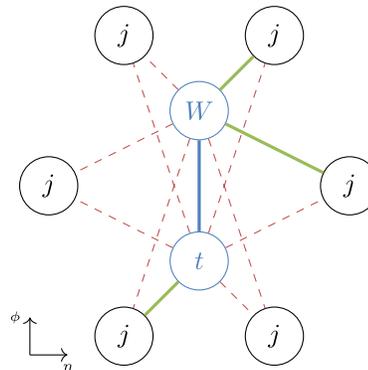
connected to all of their potential mother particles. From this new graph, instead of identifying the edges based on a shared origin, true edges are identified as being between a mother and daughter particle, and the result is the reconstruction of the Feynman diagram as depicted in Fig. 1.

Second, as the intermediate particles are now represented by their own nodes, during training the kinematic properties of the injected nodes are predicted as auxiliary tasks with dedicated regression networks for each particle type. Edges in GNNs are not only used to identify connections but also to propagate information. By utilizing the edges of the Topograph as message passing layers to update the properties of the injected nodes, properties of the intermediary particles are extracted from the graph and regression networks can predict their kinematic properties. This is advantageous both through improvements in the edge classification, but also from the additional extracted information.

One leading advantage of Topographs over fully connected GNNs can be seen in Fig. 2, showing the simple case of identifying the jets from the decay of a single top quark. In comparison to the fully connected GNN, which



(a) Fully connected graph



(b) Topograph

FIG. 2. Comparison of the edges in (a) fully connected GNNs and (b) Topographs, when identifying the three jets j which originate from a top quark decay. The blue nodes in (b) represent the injected nodes, a key component of the Topograph, for both the top quark (t) and W boson (W). The true edges, which identify jets which (a) originate from the same top quark or (b) reconstruct the decay chain, are green. All false edges in the graphs are dashed red lines. The predefined connection between t and the W boson is blue.

has $N(N - 1)$ edges, the Topograph only requires $\mathcal{O}(N)$, which scales linearly with the number of intermediary particles; in this case there are $2N$ edges. For cases where $N > M + 1$, where M is the number of intermediary particles, there are always fewer edges associated to reconstructed objects in a Topograph than a fully connected GNN.

Furthermore, having a handle on the underlying kinematics of the intermediary particles with auxiliary tasks should improve the resolution and accuracy of current differential measurements, which are often a function of intermediate particle kinematics. These properties could also be used in a likelihood test of the process, as is done in the MEM, instead of only using objects recorded by the detector.

With Topographs, complex underlying physics processes can be injected as priors by changing the injected particles and their potential connections. This enables additional

information to be included when designing and training the networks over standard approaches.

A. Building blocks

Although the Topograph could be visualized as one large graph with injected nodes and edges, we break them down into a simple set of building blocks. The core building block of the Topograph is the particle block. It includes the edge definitions between an input set of particles, or nodes, and the target mother particle. In addition, it contains the regression network used to predict the kinematic properties of the injected mother particle. If a Topograph is visualized as a Feynman diagram of a process, a particle block is the subcomponent which determines the correct connections to recreate a single vertex alongside the properties of the incoming particle. The basic representation of a particle block for a mother particle M is depicted in Fig. 3 alongside a representative Feynman diagram vertex. The injected mother particle M can be initialized with random values, or using information extracted from all potential daughter particles. Its properties are learned from message passing layers between itself and the input particles.

Any hypothesized process can be described by combining multiple particle blocks into a single network, connecting them to the input particles and to one another. Edges between objects and particle blocks can also be

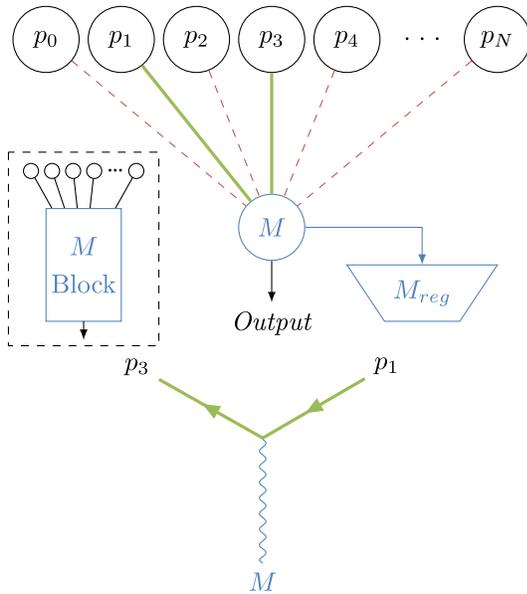


FIG. 3. The particle block for a given mother particle M . All input particles are connected to the mother particle by an edge; for illustrative purposes the true edges are shown in green and the false edges are represented by dashed red lines. The trapezoid M_{reg} is the regression network which predicts the kinematic properties of M . The inset is the notation used to represent the whole particle block. This is shown alongside the Feynman diagram representing the process, with time running from bottom to top.

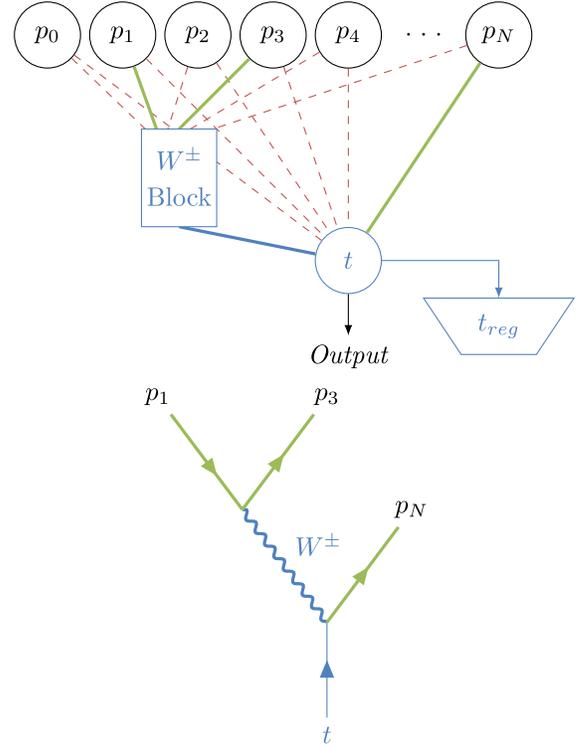


FIG. 4. Particle block of a top quark t . A particle block for the W is nested within the t block. A connection between the W boson and the top quark is predefined (shown in blue). This is shown alongside the corresponding Feynman diagram.

predefined, for example between the particle blocks for a W boson and a top quark, as shown in Fig. 4.

B. Assembling a neural network with Topographs

For many processes, there will be more than one reconstructed object type or final state particle in the chosen process. To address this a set of neural networks ϕ_p , one for each particle type p , can be incorporated into the Topograph model as a series of particle embedding networks. Furthermore, in order to maximize initial information exchange before the Topograph it may be beneficial to include a normal message passing layer before the Topograph. Options for this layer include attention transformers and standard message passing GNN layers. Furthermore, in complex processes it is possible to define which edges are to be predicted and which are fixed. For example, two leptons in the production of a Z boson in association with two top quarks could be set to originate from the Z boson, or one from each W boson.

V. SOLVING COMBINATORICS IN $\bar{t}t$ EVENTS

For an initial application and for direct comparison with other state-of-the-art methods we apply Topographs to top quark pair production with both tops decaying hadronically. We compare the performance of our method to a

benchmark non-ML approach used in many top quark analyses, the χ^2 method, and the state-of-the-art ML approach, SPA-Net [26]. All models are trained and evaluated on the same dataset.

20 million $t\bar{t}$ events with a center of mass energy $\sqrt{s} = 13$ TeV are simulated using MadGraph5_aMC@NLO [45] (v3.1.0), with decays of top quarks and W bosons modeled with MadSpin [46], with both W bosons decaying to two quarks.¹ The parton shower and hadronization is performed with PYTHIA [48] (v8.243). The detector response is simulated using DELPHES [49] (v3.4.2) with a parametrization similar to the response of the ATLAS detector. Jet clustering is performed using the anti- k_r algorithm [50] with a radius parameter $R = 0.4$ using the FastJet [51] package. Jets originating from b quarks (b jets) are identified with a simple binary discriminant corresponding to an inclusive 70% signal efficiency.

For training, events with at least six jets are selected, keeping up to 16 jets per event as ordered by their transverse momentum. Truth matching of jets to the partons in the hard scatter is performed using a cone of $\Delta R < 0.4$. Events with partons matched to multiple jets, or jets matched to multiple partons, are discarded. Events are further required to have zero reconstructed leptons, though no requirement on the number of b jets. Finally, events are required to be fully reconstructable, where jets are matched to all six partons of the $t\bar{t}$ decay. After these requirements there are 1,340,000 training events and 71,000 validation events.

An additional 298,000 events are reserved for evaluating the final performance. These also contain events where not all partons have a jet matched to them. In 76,000 of these events it is possible to associate a jet to all six partons. After requiring at least two b jets in the event, there are 147,000 events of which 44,000 are fully reconstructable.

As inputs for both ML models the four momentum (p_x, p_y, p_z, E) together with a Boolean flag showing whether the jet was b tagged or not are used. The three momenta are normalized subtracting the mean and dividing by the standard deviation. The energy is normalized in the same way after applying the logarithm. No normalization or transformation is applied to the b -tagging flag. As both models represent the single jets as nodes, additional information per jet can easily be added for both models as additional node features.

A. Topograph implementation

Before the particle blocks for the two top quarks, two fully connected message-passing graph layers are used to provide an information exchange between the jets in the event and update the jet features. From these updated jets, the W nodes are initialized using attention weighted

pooling. Two different networks are used to obtain two sets of attention weights, one for each of the W nodes. The top nodes are initialized in the same way from the jets but with the corresponding W node concatenated to the pooled jet information. The regression targets of the W and top nodes are the truth level properties of the particles. In both cases the three momentum (p_x, p_y, p_z) is chosen as the regression target. In our dataset the truth mass is fixed to the Monte Carlo mass values. The network structure is shown in Fig. 5.

Within both the information exchange and Topograph blocks, message passing is bidirectional with a separate edge for each direction. Shared weights are used for calculating the attention pooling weights for each category of edge, defined by the sending and receiving particle type. Edge features are formed by concatenating the features of the sending node and the receiving node. Edge features are persistent between updates, and after the first iteration the current values are concatenated to the new features.

Four message passing steps in the Topograph update the jets, edges and injected W and top nodes. After the final message passing step, the properties of the W bosons and top quarks and particle matching scores are extracted. There are four matching scores for each jet, from which six jets are assigned to the six partons. The scores are calculated by passing the edge properties of the jet to the W and top node through a classification network to determine if it is a true edge.

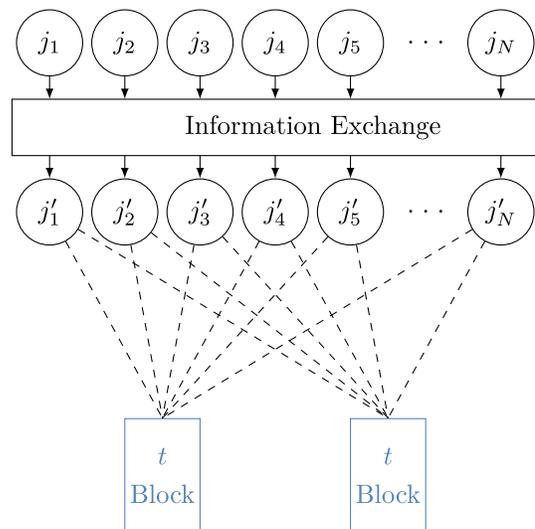


FIG. 5. Topograph network for the $t\bar{t}$ process comprising two t blocks. The jets are passed through an information exchange layer, using a fully connected graph message passible layer. All jets are connected to all possible mother particles, as shown by the dashed edges. The information exchange comprises multiple message passing graph layers and the node updates with the Topograph are performed N times before the edge values are used for parton assignment and the properties for the top quarks and W bosons are extracted.

¹The dataset is available at <https://zenodo.org/record/7737248> [47].

1. Loss function

The loss function is calculated from both the edge classification task and the regression tasks. Edges are classified as true if a jet originated from the b quark, in the case of the top nodes, or one of the two quarks from the decay of the W boson, for the W nodes. Here, binary cross entropy is used for the loss function, and it is weighted using the number of true and false edges across the dataset in order to improve convergence during training. For each regression task, the mean absolute error function is used between the predicted values of the node and the true values of the top quark or W boson.

Since it is not possible for the network to determine which top node corresponds to the top quark or the antitop quark, a symmetrized version of the losses is calculated in which the loss is calculated for both possible cases. The loss is then defined as the minimum of the two. Since the W boson nodes are directly connected to the top nodes, the loss terms for the W bosons are included in the loss calculation without need for additional symmetrization. The full loss is given by

$$\mathcal{L}_{\text{symm}} = \min(\mathcal{L}(t_1, t) + \mathcal{L}(W_1, W^+) + \mathcal{L}(t_2, \bar{t}) + \mathcal{L}(W_2, W^-), \\ \mathcal{L}(t_1, \bar{t}) + \mathcal{L}(W_1, W^-) + \mathcal{L}(t_2, t) + \mathcal{L}(W_2, W^+)),$$

where $\mathcal{L}(p_i, p)$ corresponds to the combined edge classification loss and regression loss for each of the injected particles $p_i \in t_1, t_2, W_1$, and W_2 , with respect to the truth particle $p \in t, \bar{t}, W^+, W^-$.

2. Parton assignment

Several options could be tested for assigning jets to the partons for the edge score. For this initial study a simple iterative approach is chosen. First, the edge with the highest score is labeled as a true edge, with the jet assigned to this parton. Next all edges connected to the corresponding jet and parton are removed, and the next highest edge is chosen. In the all hadronic $t\bar{t}$ case, there is one parton per top quark, corresponding to the b quarks in the decay, and two per W boson. This assignment procedure is repeated until all six partons are assigned, and results in a solution where neither a jet or parton can be assigned to multiple targets.

B. Reference methods

1. χ^2 implementation

The χ^2 score for $t\bar{t}$ decays used in this work is given by

$$\chi^2 = \frac{(m_{b_1 q_1 q_2} - m_t)^2}{\sigma_t^2} + \frac{(m_{b_2 q_3 q_4} - m_t)^2}{\sigma_t^2} \\ + \frac{(m_{q_1 q_2} - m_W)^2}{\sigma_W^2} + \frac{(m_{q_3 q_4} - m_W)^2}{\sigma_W^2},$$

where $m_{b_i q_j q_k}$ is the invariant mass of the jets in the permutation, m_t and m_W are the masses of the top quark and W boson, and σ_t and σ_W are the widths of the top quark and W boson in the dataset. The values for m_t and m_W are obtained by taking the mean of the reconstructed invariant masses in our dataset, while σ_t and σ_W are set to the standard deviation. The calculated values are $m_t = 169.8$ GeV, $m_W = 81.0$ GeV, $\sigma_t = 29.0$ GeV, and $\sigma_W = 18.5$ GeV. Jets associated to the b quark of the top decays are required to be b tagged, while no requirement is placed on the jets associated to the W boson decays.

2. SPA-Net implementation

The implementation is taken from the GitHub repository.² The hyperparameters of the model are optimized for the dataset using fully matched events, with the values in Ref. [19] using *v1.0* resulting in the highest overall efficiencies. All SPA-Net models are trained for 100 epochs, taking the weights after the epoch with the lowest loss on the validation set.

C. Partial event trainings

It is also possible to train Topographs on events in which not all partons from the $t\bar{t}$ decay are matched to jets. For these events, the same network is used but the edge classification and regression loss terms are not considered for the W boson or top quarks which have partons not able to be matched to jets. In the case of the b quark from the top quark decay not being matched to a jet, the $\mathcal{L}(W_i, W^+)$ term is still considered. Where a quark from the W boson decay is missing, both the top quark and W boson loss terms are not considered. At least one W boson is required to be fully reconstructable, with both quarks matched to jets in the event.

As introduced in Ref. [26], SPA-Net can also be trained on nonfully reconstructable events, however in comparison to Topographs this is only at the level of each top quark. When a b quark from the top quark is missing, the corresponding W boson is also not considered. For a fairer comparison with both models trained on the same number of events, we compare models trained only on fully reconstructable events. Results for the Topograph and SPA-Net models trained with partial events are presented in Appendix A 3.

VI. RESULTS

For the evaluation of the performance only events with at least two b -tagged jets are considered. This is a common requirement in physics analysis to reduce the contribution from the multijet background. This requirement is not imposed during training as it is found to reduce the

²<https://github.com/Alexanders101/SPANet/tree/v1.0>.

TABLE I. Event reconstruction efficiencies (%) for the χ^2 method, the SPA-Net model and our Topograph model in different jet and b -jet multiplicities. The mean and standard deviation are calculated using five trainings, and the best model corresponds to the training with the highest efficiency for events with at least six jets and exactly two b jets. The highest efficiency is highlighted in bold.

N_{jets}	$N_{b\text{-jets}}$	Mean		Best		
		SPA-Net	Topograph	SPA-Net	Topograph	χ^2
6	2	81.43 \pm 0.03	81.54 \pm 0.12	81.47 \pm 0.31	81.70 \pm 0.31	72.73 \pm 0.36
6	≥ 2	79.45 \pm 0.07	79.76 \pm 0.14	79.56 \pm 0.30	79.95 \pm 0.29	70.94 \pm 0.33
7	2	64.47 \pm 0.08	65.46 \pm 0.20	64.69 \pm 0.44	65.90 \pm 0.44	54.28 \pm 0.46
7	≥ 2	62.38 \pm 0.12	63.51 \pm 0.25	62.86 \pm 0.40	63.94 \pm 0.40	52.11 \pm 0.41
≥ 6	2	68.48 \pm 0.06	69.12 \pm 0.19	68.65 \pm 0.25	69.54 \pm 0.25	58.57 \pm 0.26
≥ 6	≥ 2	65.76 \pm 0.08	66.59 \pm 0.22	66.05 \pm 0.23	67.03 \pm 0.22	55.90 \pm 0.24

performance as a result of the reduction in training statistics.

A. Jet parton assignment

Although Topographs predict the kinematics of injected particles, the most important measure of performance is the association of jets to partons. Efficiencies of reconstructing the whole event correctly are shown in Table I for varying jet and b -jet multiplicities. Only events which are fully reconstructable are considered. Topographs and SPA-Net show very similar performance across all jet multiplicities, with subpercent differences in the reconstruction efficiencies. Both clearly outperform the χ^2 method by around 10% with larger differences at higher jet multiplicities.

Figures 6 and 7 show the reconstructed W boson and top quark invariant masses. The distributions are shown separately for events with the correct and incorrect jet assignments, as well as events which are not possible to be fully reconstructed due to not all partons being matched to reconstructed jets (impossible). Events where the correct triplet is assigned to the top quark but with the incorrect

matching account for 5% of the incorrect events for Topographs and SPA-Net, and 3% for χ^2 . All methods are able to reconstruct candidates closest to the particle mass, with a broader distribution for events without the corresponding partons matched to jets in the event. Normalized and stacked distributions can be found in Appendix A 6. SPA-Net and Topographs have very similar distributions showing no difference in potential sculpting of the mass distributions. The χ^2 method has distributions which are shifted towards slightly higher values compared to SPA-Net and Topographs.

For events which do not contain all partons in the final state, it is of interest to see how many of the partons are correctly matched to jets. In Table II the matching efficiencies of only the partons which are present in the event are compared for the three approaches. The Topograph performs slightly better than SPA-Net, with higher efficiencies for correctly matching all available partons or only one incorrectly matched parton. Both Topographs and SPA-Net are substantially better than the χ^2 at correctly identifying all available partons. It should be noted that the

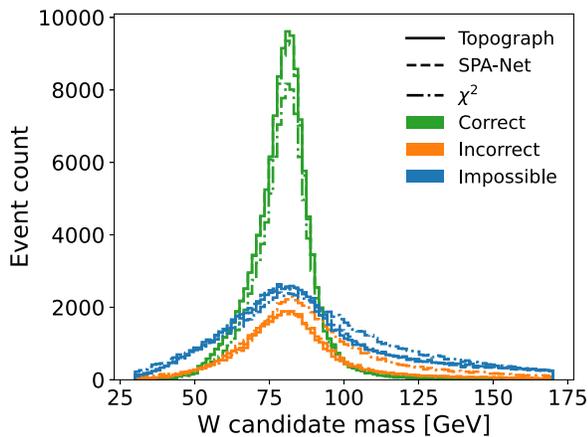


FIG. 6. Reconstructed W boson invariant mass m_W using jets assigned by the Topograph (solid), SPA-Net (dashed), and the χ^2 method (dash-dotted). Events are categorized into correct (green) and incorrect assignments (orange), with events missing a parton at reconstruction level labeled as impossible (blue).

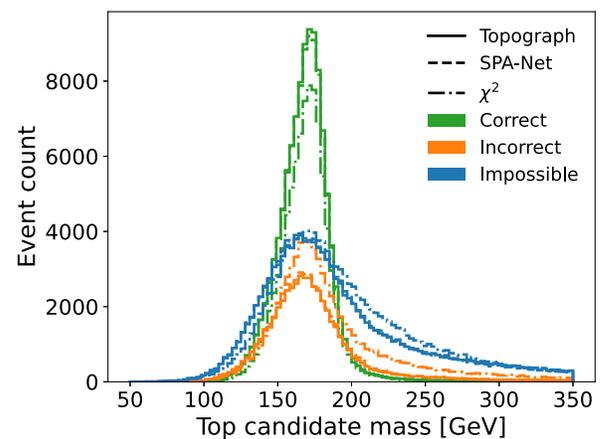


FIG. 7. Reconstructed top quark invariant mass m_{top} using jets assigned by the Topograph (solid), SPA-Net (dashed), and the χ^2 method (dash-dotted). Events are categorized into correct (green) and incorrect assignments (orange), with events missing a parton at reconstruction level labeled as impossible (blue).

TABLE II. Percentage of events with at most N incorrectly matched partons using the χ^2 method, SPA-Net and Topographs. Events are categorized based on the number of partons matched to jets at truth level. The highest efficiency is highlighted in bold.

N_{partons} matched	Model	Incorrectly matched partons [%]					
		0	≤ 1	≤ 2	≤ 3	≤ 4	≤ 5
3	SPA-Net	34.34 ± 0.65	74.88 ± 0.59	98.03 ± 0.19	100.00 ± 0.00
	Topograph	36.57 ± 0.66	77.80 ± 0.57	98.70 ± 0.15	100.00 ± 0.00
	χ^2	34.30 ± 0.65	80.12 ± 0.54	99.05 ± 0.13	100.00 ± 0.00
4	SPA-Net	35.69 ± 0.28	64.53 ± 0.28	90.99 ± 0.17	99.41 ± 0.04	100.00 ± 0.00	...
	Topograph	37.74 ± 0.28	67.06 ± 0.27	92.21 ± 0.16	99.53 ± 0.04	100.00 ± 0.00	...
	χ^2	31.01 ± 0.27	63.84 ± 0.28	92.19 ± 0.16	99.73 ± 0.03	100.00 ± 0.00	...
5	SPA-Net	44.74 ± 0.19	64.87 ± 0.18	86.92 ± 0.13	97.91 ± 0.05	99.94 ± 0.01	100.00 ± 0.00
	Topograph	46.57 ± 0.19	66.88 ± 0.18	88.58 ± 0.12	98.45 ± 0.05	99.96 ± 0.01	100.00 ± 0.00
	χ^2	32.36 ± 0.18	58.27 ± 0.19	84.70 ± 0.14	98.21 ± 0.05	99.94 ± 0.01	100.00 ± 0.00
6	SPA-Net	66.05 ± 0.23	74.43 ± 0.21	90.06 ± 0.14	96.22 ± 0.09	99.52 ± 0.03	99.98 ± 0.01
	Topograph	67.03 ± 0.22	75.71 ± 0.20	91.33 ± 0.13	97.03 ± 0.08	99.70 ± 0.03	100.00 ± 0.00
	χ^2	55.90 ± 0.24	64.93 ± 0.23	84.14 ± 0.17	93.97 ± 0.11	99.43 ± 0.04	99.98 ± 0.01

perfect reconstruction in this case is substantially lower than the fully reconstructable events in Table I.

In Table III we break down the matching efficiencies of jets in incorrect events by the truth flavor of the jet. The jet truth flavor is assigned using ΔR matching to B and C hadrons. We observe that Topographs and SPA-Net have similar jet matching efficiencies for all jet flavors. Topographs have a slightly higher efficiency for b jets, which correspond to the jets coming from the b quarks in the top quark decays as well as in matching light jets to the quarks in the W boson decay. The χ^2 method has the largest efficiency drop for truth flavor c jets. This results from c jets being misidentified by the b tagging, which in the χ^2 method allows them to be matched to the b quarks from the top decays.

The matching efficiencies of partons in incorrect events are shown in Table IV. Here we see that the χ^2 method mostly performs worse for matching jets to the quarks in the W boson decay. Topographs and SPA-Net show similar behavior, with the slightly better performance in Topographs for incorrect events evenly distributed across all partons.

TABLE III. Efficiencies (%) of correctly associating a jet to its corresponding parton for the χ^2 method, SPA-Net and Topographs. Jets are categorized into their truth flavor using parton matching. Only jets coming from the $t\bar{t}$ decays in events with at least one incorrect assignment are considered. The highest efficiency is highlighted in bold.

Correct	SPA-Net	Topograph	χ^2
b jets	58.94 ± 0.29	59.99 ± 0.29	57.68 ± 0.25
c jets	55.08 ± 0.39	55.81 ± 0.39	50.38 ± 0.34
Light jets	70.42 ± 0.22	72.01 ± 0.22	68.51 ± 0.20

B. Interpreting edge scores

Because of the individual edge scores, the confidence of the combinatoric assignment from Topographs can be obtained by aggregating the edge scores. This could be used to filter events as likely to be incorrect, as well as identify events for which it is not possible to match jets to all partons.

In Fig. 8 the product of assigned jet edge scores is used as an event matching score. There is reasonable separation between events with a correct assignment and those with the incorrect and impossible assignments. However, a shoulder towards higher values is visible for the impossible events.

In Fig. 9 the impossible events are categorized into the number of jets which are correctly assigned. It can be seen that the score is still high for events where there is a single parton missing, but all remaining partons are correctly matched to jets. Whether using this score to remove events will benefit other downstream applications, such as top quark mass measurements, would need to be tested.

C. Regression performance

Instead of reconstructing the top quark and W boson kinematics from the matched jets, Topographs are trained to predict their properties directly. Figure 10 compares the resolution of the W boson p_T and Fig. 11 the top quark p_T from the Topograph regression networks and the invariant system of the assigned jets. The predictions are shown for both correct and incorrect jet assignments.

For events with all jets correctly assigned, using the p_T values from the Topograph regression networks leads to a narrower peak than the reconstructed quantities, demonstrating a more accurate prediction. They also show a reduced bias compared to the values reconstructed solely using the jets. For events with incorrectly assigned jets, no

TABLE IV. Efficiencies (%) of correctly associating a jet to the corresponding parton for events with at least one incorrect assignment for the χ^2 method, SPA-Net and Topographs. Top quarks are classified by the momentum of the corresponding b quark. The quarks from the W boson decays are ordered by their p_T . The highest efficiency is highlighted in bold.

	Correctly identified	SPA-Net	Topograph	χ^2
Top quark (leading b)	b -quark	61.83 ± 0.40	62.94 ± 0.40	59.17 ± 0.35
	Leading quark (W)	67.79 ± 0.38	68.92 ± 0.39	61.80 ± 0.35
	Subleading quark (W)	63.22 ± 0.40	64.70 ± 0.40	65.09 ± 0.34
Top quark (subleading b)	b -quark	56.05 ± 0.41	57.03 ± 0.41	56.20 ± 0.36
	Leading quark (W)	67.49 ± 0.38	68.22 ± 0.39	61.30 ± 0.35
	Subleading quark (W)	66.41 ± 0.39	68.31 ± 0.39	65.94 ± 0.34

difference is observed between the two predictions. This shows that the Topograph is not learning to guess the p_T from the training data, but instead is using the associated jets. Accurate reconstruction of the intermediate particles has a strong dependence on whether the model has correctly assigned jets to partons. There are no biases

observed using either method for the top quark and W boson, and there is very little difference for reconstruction of the top quark η and ϕ coordinates. However, reconstructing the W boson η and ϕ coordinates using the assigned jets performs better than the prediction from the network. In all cases, the Topograph is trained to predict the momentum components p_x, p_y, p_z rather than p_T, η, ϕ .

VII. CONCLUSION

In this work we have introduced Topographs, a novel approach for solving the combinatorics and reconstructing the topology of a particle physics process from final state objects reconstructed by a detector. The performance matches the current state-of-the-art technique using symmetry preserving attention transformers, and surpasses the standard approach commonly used in analyses, with a computational complexity which scales only linearly with increasing final state object multiplicity.

The edge scores from Topographs can be combined into discriminants to assign a confidence to the jet-parton assignments, which could be useful in downstream applications. Furthermore, the additional regression tasks included in Topographs demonstrate good predictive power with similar accuracy but reduced bias compared to using only the jets assigned to intermediate particles. However, in both cases there remains room for improvement.

There are several other areas open for further optimization. Because of the message passing layers used to define the Topograph, it was found that fully connected graph layers between all jets for information exchange lead to faster convergence whilst training, and also resulted in requiring fewer learnable parameters in the model. This causes the complexity of the network in this work to scale quadratically with the number of jets, and not linearly as can be achieved by only using the particle blocks in Topographs. By moving to other architectures such as transformer encoders with cross attention, the need for information exchange layers could be mitigated. Alternative approaches for assigning jets to partons based on the edge scores could also improve the overall performance.

Applications of Topographs are not limited to the case study presented in this paper, and due to their modular

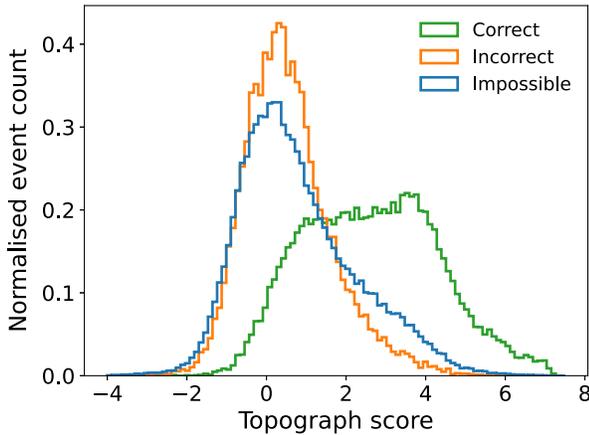


FIG. 8. Event matching score for events with the correct assignment (green), incorrect assignment (orange), and non-fully-reconstructable events (impossible, blue).

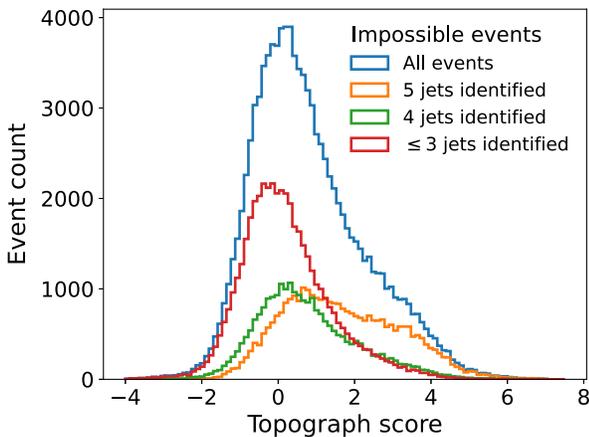


FIG. 9. Event matching score for non-fully-reconstructable events, categorized by the number of correctly assigned jets.

nature Topographs can be generalized to almost all particle physics processes. Their applications are also not limited to matching final state objects to an underlying physics process, but could also find use in jet identification, reconstructing displaced vertices from heavy flavor hadron decays or using the constituents in large radius jets in boosted topologies.

ACKNOWLEDGMENTS

The authors would like to acknowledge funding through the SNSF Sinergia grant called ‘‘Robust Deep Density Models for High-Energy Particle Physics and Solar Flare Analysis (RODEM)’’ with funding No. CRSII5_193716, the SNSF Project Grant No. 200020_212127 called ‘‘At the two upgrade frontiers: Machine learning and the ITk Pixel detector,’’ and the Alexander von Humboldt foundation Feodor Lynen fellowship program.

L. E. and J. A. R. contributed equally to this work.

APPENDIX

1. Hyperparameters

Table V shows the hyperparameters used for the training of the Topograph models presented in this paper. The models were trained using TensorFlow v2.10 [52].

2. Example Topograph

A complete Topograph network is shown in Fig. 12 for reconstruction of $t\bar{t}W$ in events containing exactly one lepton, one neutrino and multiple jets. For the production of a top quark pair in association with a W boson ($t\bar{t}W$), two top quark blocks and one W block are connected to the input particles, but not to one another. The Topograph network is trained to identify the edges of the true daughter particles of each particle in the process, and predicts the kinematics of the two top quarks and the three W bosons. As Topographs are defined to represent an underlying physics process, it is also a choice whether to predefine whether the lepton and neutrino originate from a W boson coming from a top decay, or the additional W boson. Figure 13 shows a Topograph where the lepton is required to come from a top decay.

3. Comparisons with partial event trainings

Instead of training using only complete events, both the Topograph and SPA-Net models can be trained including partial events, that is, events where some partons are not able to be matched to reconstructed jets. In the training events, for Topographs at least one W boson and for SPA-Net at least one top quark are required to be fully reconstructable. For these models the efficiencies of fully reconstructable events change slightly, with Topographs having a slight reduction in efficiencies and SPA-Net a slight increase, as shown in Table VI. However, for most

TABLE V. Hyperparameters used for the training of Topographs.

Hyperparameter	Value
Optimizer	<i>AdamW</i>
Epochs	100
n original message passing	2
n topograph updates	4
lr schedule	Cosine annealing
Initial learning rate	0.001
Decay steps	Two epochs
Batch size	256
Pooling	Attention
W initialization	Attention pooling
Top initialization	Attention pooling
Attention units	[32, 32, 1]
Activation	Gelu
Normalization	Layer norm
Regression units	[64, 64, 3]
Edge classification units	[128, 128, 128, 1]
Graph processing units	[256, 256, 64]
Persistent edges	True
Classification loss	Weighted binary cross entropy
Regression loss	Mean absolute error

selections, the efficiencies are still within the uncertainties of models trained only on fully reconstructable events. Table VII shows the parton matching efficiencies for events categorized based on how many partons can be matched to reconstructed jets. Here the benefit of training on partial events can be seen, with both Topographs and SPA-Net having higher efficiencies of correctly matching jets to the all available partons.

4. Impact of systematic variations

For applications in high energy physics, it is crucial that any new approach is not sensitive to changes under

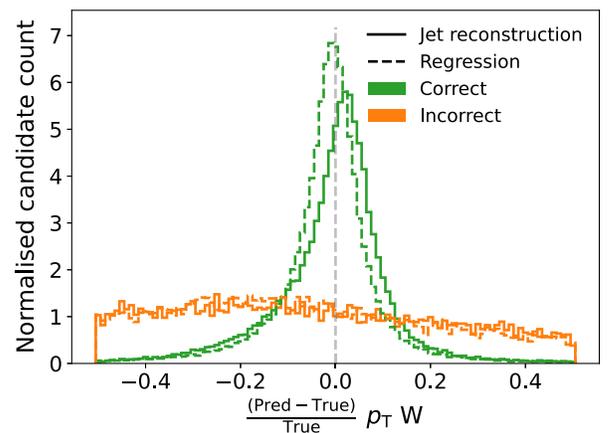


FIG. 10. Resolution of the reconstructed p_T of the W boson from the Topograph, comparing the prediction from the invariant system of the assigned jets (solid line) and the Topograph regression network (dashed line) for correct assigned events (green) and incorrect assigned events (orange).

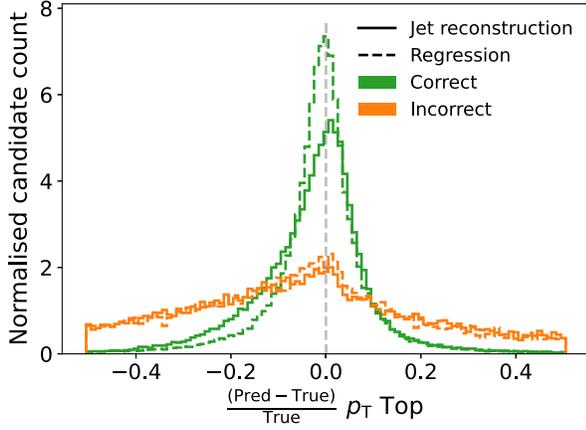


FIG. 11. Resolution of the reconstructed p_T of the top quark from the Topograph, comparing the prediction from the invariant system of the assigned jets (solid line) and the Topograph regression network (dashed line) for correct assigned events (green) and incorrect assigned events (orange).

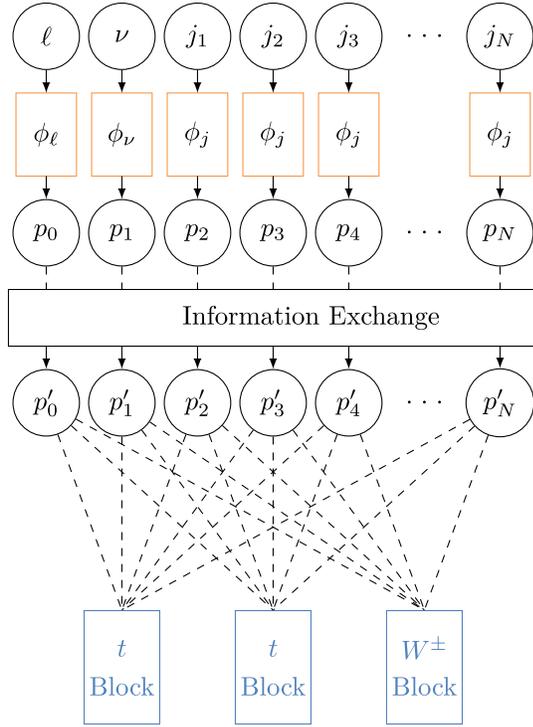


FIG. 12. Topograph network for the $t\bar{t}W$ process comprising two t blocks and a W block. The input particles are first passed through an embedding network ϕ unique to each type of particle; either a jet j , lepton ℓ , or neutrino ν . Embedded particles are then passed through an (optional) information exchange layer. All particles are connected to all possible mother particles, as shown by the dashed edges.

systematic variations. In particular with machine learning approaches, it would be problematic if methods were sensitive to underlying and nonphysical effects arising from the simulated samples on which they were trained.

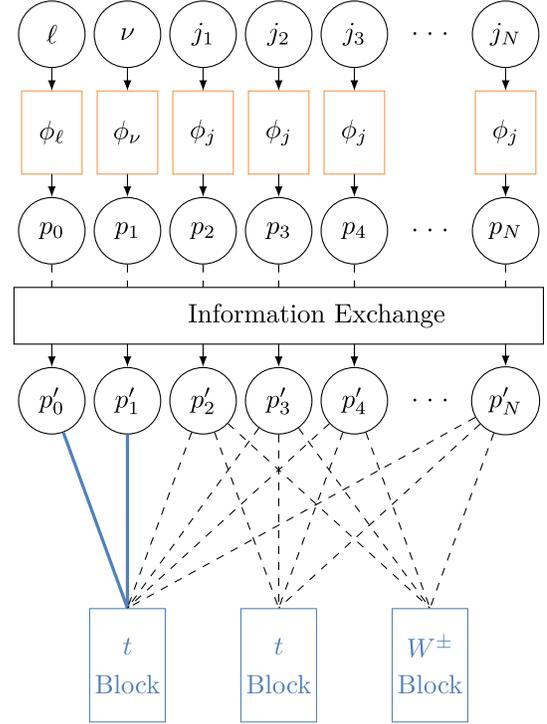


FIG. 13. Topograph network for the $t\bar{t}W$ process comprising two t blocks and a W block where the ℓ and ν are predefined to connect to the W in one top block.

Other sources of variation come from differences in the calibration or reconstruction of physics objects between simulation and data.

To test the dependence on the simulated samples used to train the Topograph, we evaluate the best performing model trained on the nominal MadGraph data on an alternative independent dataset. This alternative sample consisting of all-hadronic $t\bar{t}$ events simulates both the hard interactions and parton shower are with PYTHIA8 (v8.307), using the Monash tuned set of parameters [53] at leading order accuracy.

To test the dependence on reconstruction effects, we apply a shift or Gaussian smearing to the energy of reconstructed jets in the events.

The absolute change in performance arising from the systematic variations is summarized in Table VIII.

TABLE VI. Event reconstruction efficiencies (%) for the χ^2 method, the SPA-Net model and our Topograph model in different jet and b -jet multiplicities. Both models were trained on complete and partial events. The highest efficiency is highlighted in bold.

N_{jets}	$N_{b\text{-jets}}$	SPA-Net	Topograph	χ^2
6	2	81.03 ± 0.32	80.77 ± 0.32	72.73 ± 0.36
6	≥2	79.11 ± 0.30	78.93 ± 0.30	70.94 ± 0.33
7	2	65.35 ± 0.44	65.18 ± 0.44	54.28 ± 0.46
7	≥2	63.40 ± 0.40	63.13 ± 0.40	52.11 ± 0.41
≥6	2	68.93 ± 0.25	68.75 ± 0.25	58.57 ± 0.26
≥6	≥2	66.33 ± 0.23	66.21 ± 0.23	55.90 ± 0.24

The impact is compared for Topographs, SPA-Net, and χ^2 . Evaluating on the alternative sample results in a slightly reduced overall efficiency for all three approaches. This affects Topographs and SPA-Net slightly more than χ^2 , however, the overall gain in performance remains similar. Both Topographs and SPA-Net are robust under systematic shifts or reduced jet energy resolution, whereas the χ^2 method suffers from a substantial drop in efficiency, especially at higher jet multiplicities.

5. Studying edge scores

Figure 14 shows the distribution of edge scores of the jets to the W helper node which is decided to be the W^+ based on the loss. Figure 14(a) includes all jets in the event. The distribution for the jets which originate from the W^+ peak at one, whereas the distributions of all other jet types peak at zero. A small peak at one can be seen for the b jets originating from the top quark. To investigate this peak, the scores of the b jets from both

the top and the antitop quark are shown in Fig. 14(b). They are further split based on their b -tagging score. The small peak at one originates from b jets from the top quark which are not b tagged. Furthermore, the impact of the b tagging on the score can be seen. The b jets from the antitop quark which are not b tagged have on average higher scores than the b jets from the top quark which are b tagged. This could point to the b -tagging result being more important for the association to the W nodes than kinematics.

Figure 15 shows the same distributions but to the top helper node which is decided to be the top quark. Again, good separation of the true edges from the false edges can be seen. Considering only b jets and splitting them based on their b -tagging score, it can be seen that the b -tagged jets originating from the antitop quark have on average lower scores than the non- b -tagged jets originating from the top quark. So, the b -tagging decision is not as important for the association to the top helper nodes.

TABLE VII. Percentage of events with at most N incorrectly matched partons using the χ^2 method, SPA-Net and Topographs. The SPA-Net model was trained on complete and partial events. Events are categorized based on the number of partons matched to jets at truth level. The highest efficiency is highlighted in bold.

$N_{\text{partons matched}}$	Model	Incorrectly matched partons [%]					
		0	≤ 1	≤ 2	≤ 3	≤ 4	≤ 5
3	SPA-Net	36.68 ± 0.66	76.94 ± 0.57	98.55 ± 0.16	100.00 ± 0.00
	Topograph	37.27 ± 0.66	77.89 ± 0.57	98.77 ± 0.15	100.00 ± 0.00
	χ^2	34.30 ± 0.65	80.12 ± 0.54	99.05 ± 0.13	100.00 ± 0.00
4	SPA-Net	37.78 ± 0.28	66.29 ± 0.28	91.78 ± 0.16	99.46 ± 0.04	100.00 ± 0.00	...
	Topograph	38.08 ± 0.28	67.53 ± 0.27	92.57 ± 0.15	99.62 ± 0.04	100.00 ± 0.00	...
	χ^2	31.01 ± 0.27	63.84 ± 0.28	92.19 ± 0.16	99.73 ± 0.03	100.00 ± 0.00	...
5	SPA-Net	48.14 ± 0.19	67.24 ± 0.18	88.18 ± 0.12	98.14 ± 0.05	99.95 ± 0.01	100.00 ± 0.00
	Topograph	48.22 ± 0.19	69.27 ± 0.18	89.52 ± 0.12	98.68 ± 0.04	99.96 ± 0.01	100.00 ± 0.00
	χ^2	32.36 ± 0.18	58.27 ± 0.19	84.70 ± 0.14	98.21 ± 0.05	99.94 ± 0.01	100.00 ± 0.00
6	SPA-Net	66.33 ± 0.23	74.38 ± 0.21	89.86 ± 0.14	96.18 ± 0.09	99.54 ± 0.03	100.00 ± 0.00
	Topograph	66.20 ± 0.23	74.87 ± 0.21	91.11 ± 0.14	97.04 ± 0.08	99.73 ± 0.02	100.00 ± 0.00
	χ^2	55.90 ± 0.24	64.93 ± 0.23	84.14 ± 0.17	93.97 ± 0.11	99.43 ± 0.04	99.98 ± 0.01

TABLE VIII. Difference in reconstruction efficiencies when evaluating the different methods on systematic variations. The difference is taken with respect to Table I. The systematic variations include evaluating on a dataset produced with a Pythia for the matrix element generation and the parton shower, evaluating on the nominal dataset scaling all jet energies by 2.5%, and evaluating on the nominal dataset smearing all jet energies by 5%.

N_{jets}	$N_{b\text{-jets}}$	PYTHIA			Jet scale			Jet resolution		
		SPA-Net	Topograph	χ^2	SPA-Net	Topograph	χ^2	SPA-Net	Topograph	χ^2
6	2	-2.13	-2.00	-0.94	+0.04	± 0.00	-2.82	+0.08	± 0.00	-12.79
6	≥ 2	-2.28	-2.13	-1.33	+0.07	+0.03	-3.63	+0.04	+0.01	-14.30
7	2	-2.04	-2.14	-1.66	± 0.00	± 0.00	-12.30	-0.05	-0.06	-21.52
7	≥ 2	-2.36	-2.37	-1.62	-0.02	+0.04	-12.69	-0.10	-0.07	-21.63
≥ 6	2	-1.64	-1.53	-0.72	+0.04	+0.01	-8.23	+0.02	+0.02	-17.18
≥ 6	≥ 2	-1.86	-1.89	-0.94	+0.06	+0.05	-9.02	-0.01	+0.03	-17.96

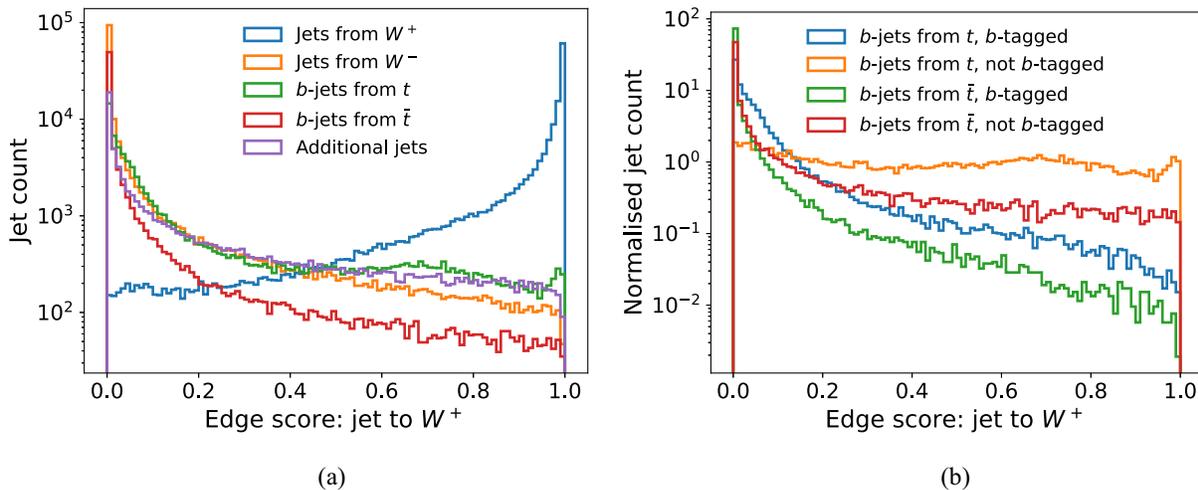


FIG. 14. Edge scores of jets to the helper node representing the W^+ . The decision which W node is the W^+ is taken by choosing the minimum of the loss under both hypotheses. Part (a) shows all jets, while (b) only shows the b jets from the two tops. They are further split into whether the jet was tagged as a b jet or not. No requirement is placed on the number of b tags.

This reliance on the b -tagging result is not unexpected. With a b -tagging efficiency of 70%, around 30% of the b jets will not be tagged as a b jet, leaving a large contribution of true edges to the top helper nodes which are not b tagged. The observed mis-tag efficiency is around 5%. Therefore, only a small fraction of the true edges to the W helper nodes is b tagged.

Figures 16 and 17 show the same plots for the W^- and antitop. No qualitative differences can be observed to the plots for the W^+ and the top.

6. Additional figures

Figures 18 and 19 show the regression results for the η and ϕ coordinate. The difference between the value of the predicted η or ϕ and the true parton property is

shown. For the predicted properties, the parton is either reconstructed from the jets that the model associates to the parton or the regression result is taken. ‘‘Correct’’ and ‘‘incorrect’’ events are shown in separate distributions. For incorrect events an additional prediction is shown by taking the true jets to reconstruct the parton properties. For correct events, the regression has for both quantities a wider distribution than the model association. For incorrect events, the regression and the model association perform worse, however using the true jets also has a worse resolution for these events. Especially for the W boson, the regression and the model association have a worse performance. However, for the top quark incorrect labeled events can still contain the correct three jets but with a wrong

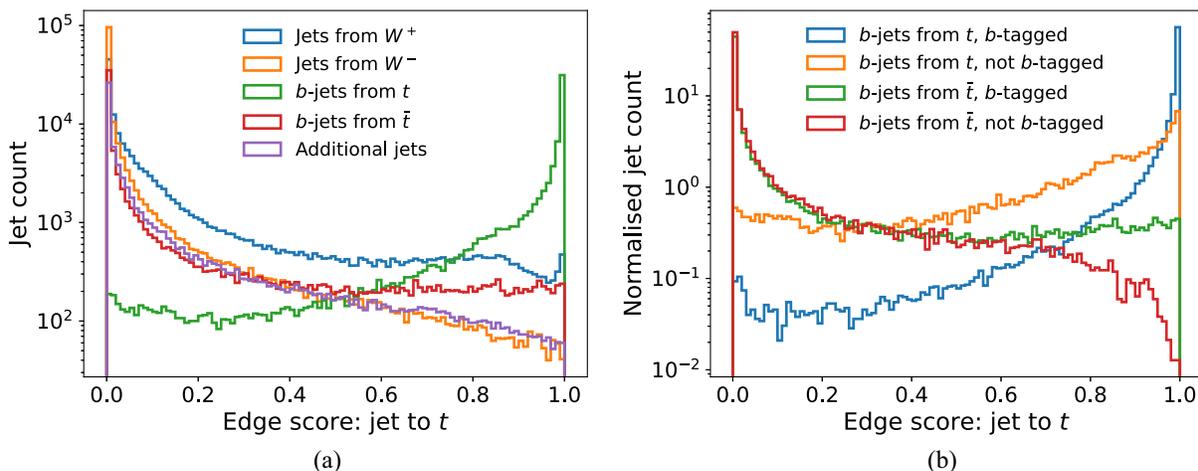


FIG. 15. Edge scores of jets to the helper node representing the t . The decision which t node is the t is taken by choosing the minimum of the loss under both hypotheses. Part (a) shows all jets, while (b) only shows the b jets from the two tops. They are further split into whether the jet was tagged as a b jet or not. No requirement is placed on the number of b tags.

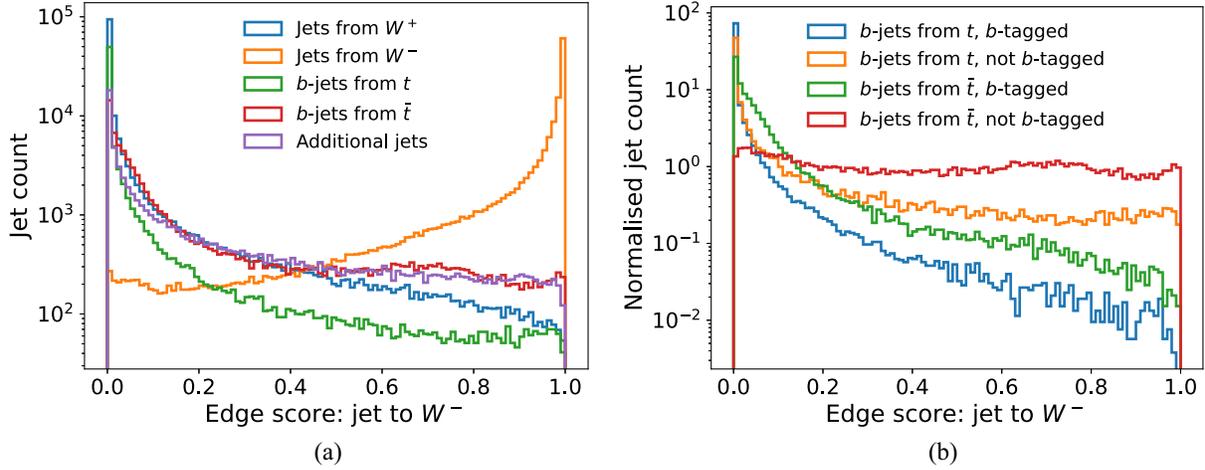


FIG. 16. Edge scores of jets to the helper node representing the W^- . The decision which W node is the W^- is taken by choosing the minimum of the loss under both hypotheses. Part (a) shows all jets, while (b) only shows the b jets from the two tops. They are further split into whether the jet was tagged as a b jet or not. No requirement is placed on the number of b tags.

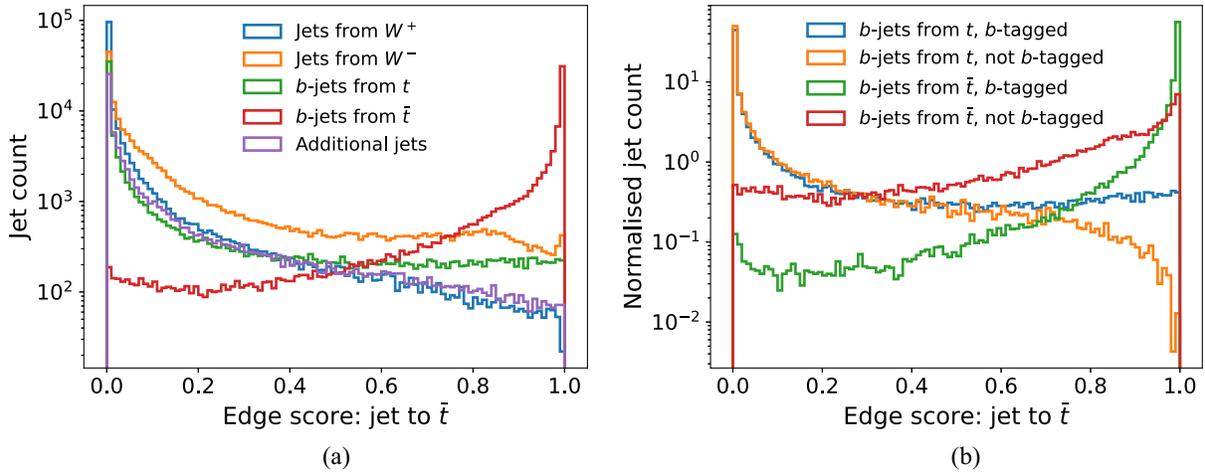


FIG. 17. Edge scores of jets to the helper node representing the \bar{t} . The decision which t node is the \bar{t} is taken by choosing the minimum of the loss under both hypotheses. Part (a) shows all jets, while (b) only shows the b jets from the two tops. They are further split into whether the jet was tagged as a b jet or not. No requirement is placed on the number of b tags.

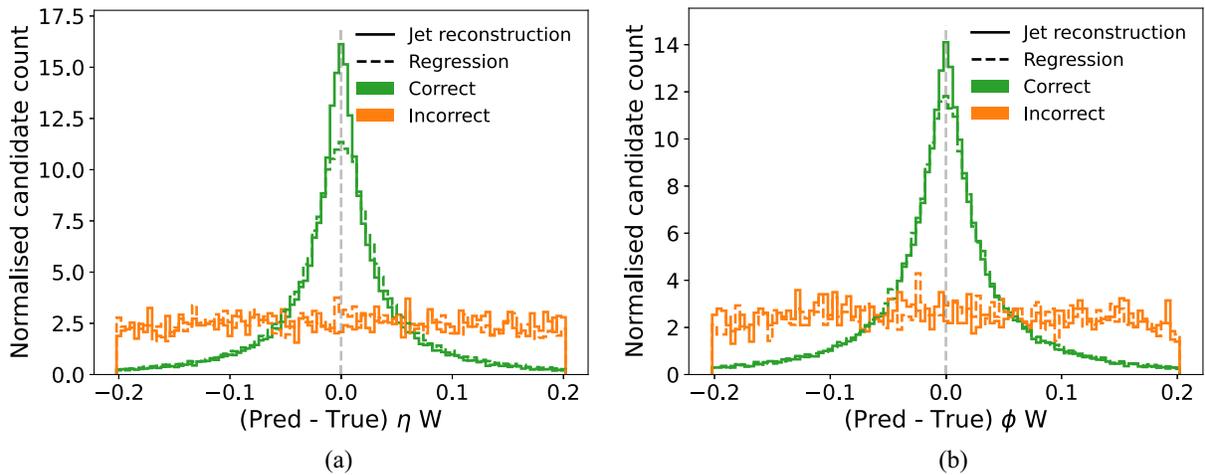


FIG. 18. Resolution of the reconstructed W boson η and ϕ coordinates from the Topograph, comparing the prediction from the invariant system of the assigned jets (solid line) and the Topograph regression network (dashed line) for correct assigned events (green) and incorrect assigned events (orange).

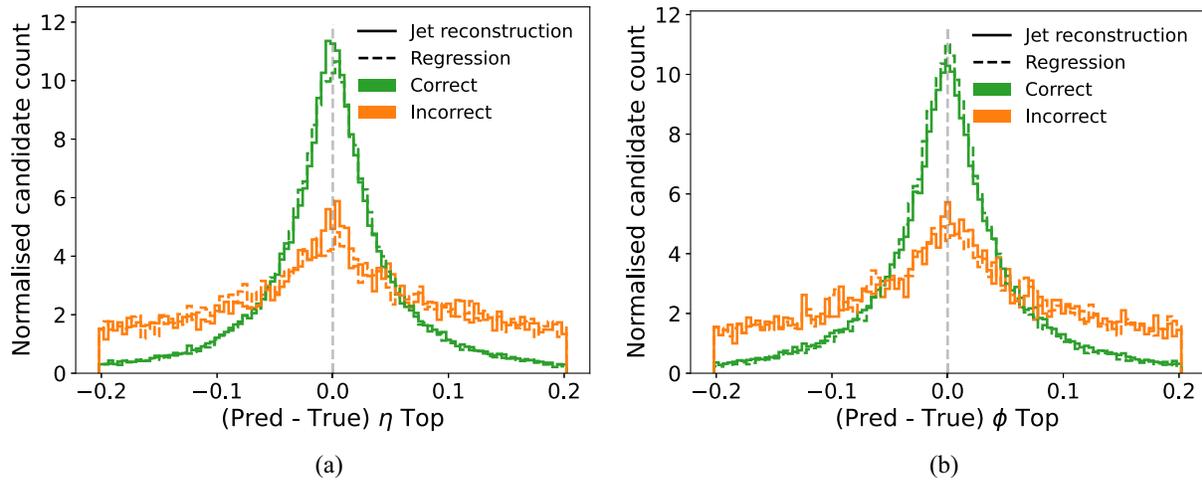


FIG. 19. Resolution of the reconstructed top quark η and ϕ coordinates from the Topograph, comparing the prediction from the invariant system of the assigned jets (solid line) and the Topograph regression network (dashed line) for correct assigned events (green) and incorrect assigned events (orange).

assignment by switching one of the W jets with the b jet.

Figure 20 shows the distribution of the reconstructed W mass split into correct, incorrect, and “impossible”

events as stacked histograms as an alternative representation compared to Fig. 6. Similarly, Fig. 21 is an alternative representation of Fig. 7 showing the reconstructed top quark mass.

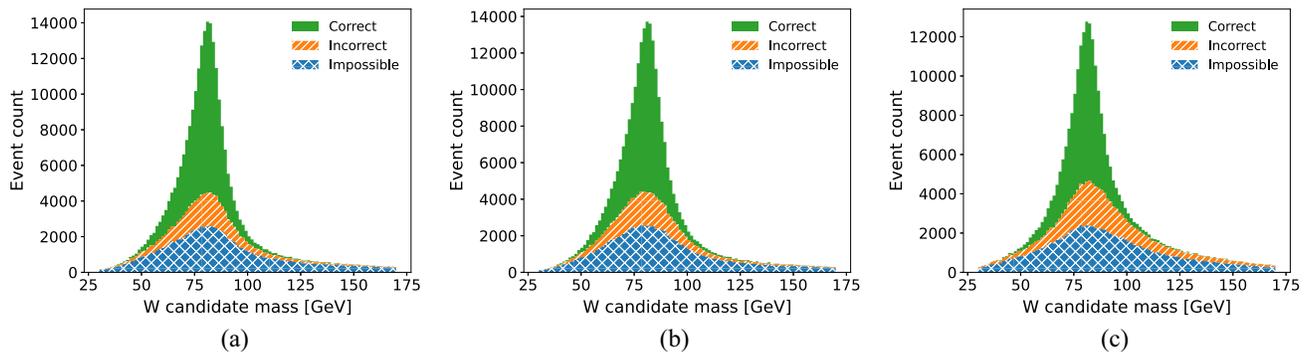


FIG. 20. Stacked distributions of the reconstructed m_W using (a) Topographs, (b) SPA-Net, and (c) χ^2 . The single histograms split the candidates into correct, incorrect, and impossible candidates.

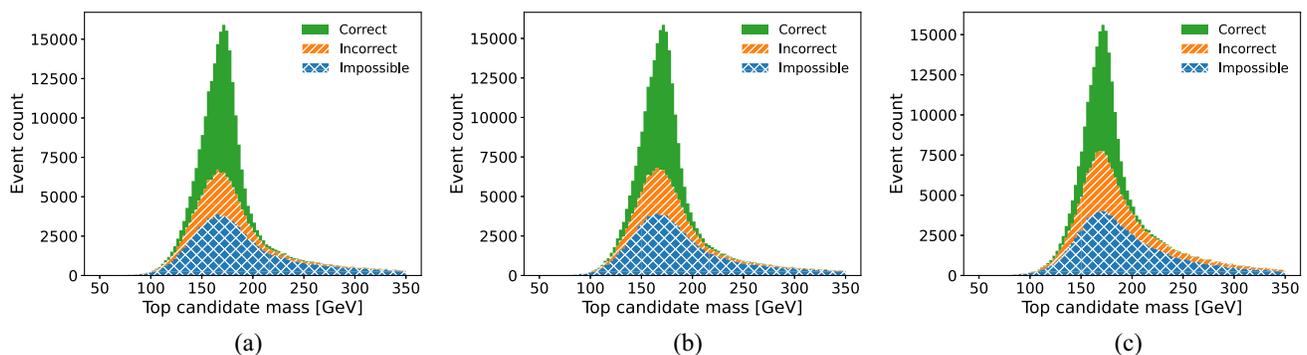


FIG. 21. Stacked distributions of the reconstructed m_{top} using (a) Topographs, (b) SPA-Net, and (c) χ^2 . The single histograms split the candidates into correct, incorrect, and impossible candidates.

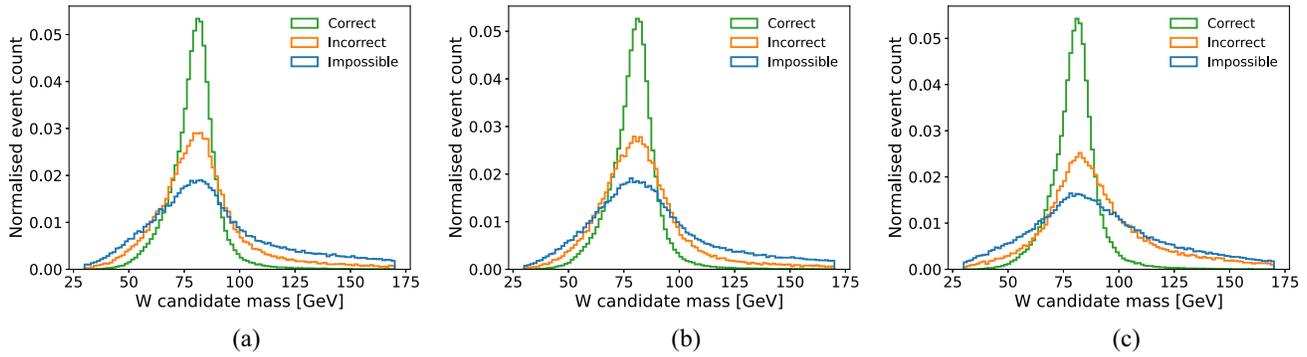


FIG. 22. Normalized distributions of the reconstructed m_W using (a) Topographs, (b) SPA-Net, and (c) χ^2 . The single histograms split the candidates into correct, incorrect, and impossible candidates.

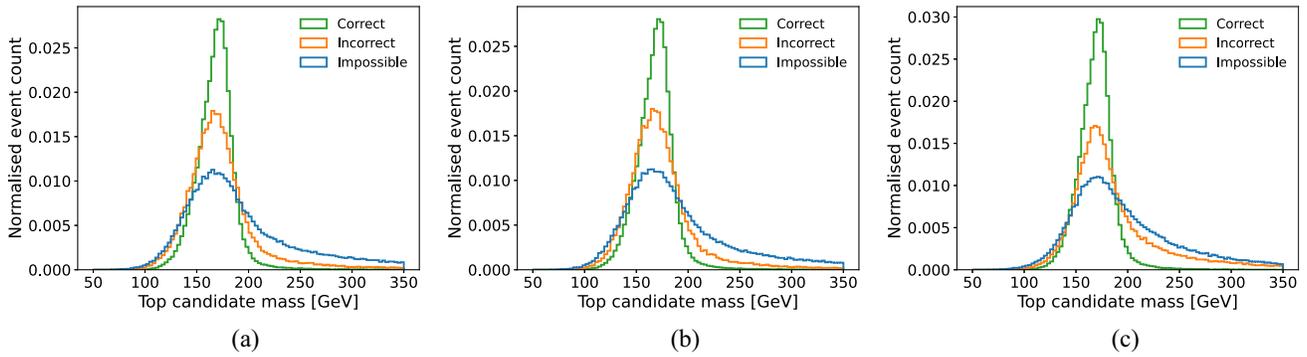


FIG. 23. Normalized distributions of the reconstructed m_{top} using (a) Topographs, (b) SPA-Net, and (c) χ^2 . The single histograms split the candidates into correct, incorrect, and impossible candidates.

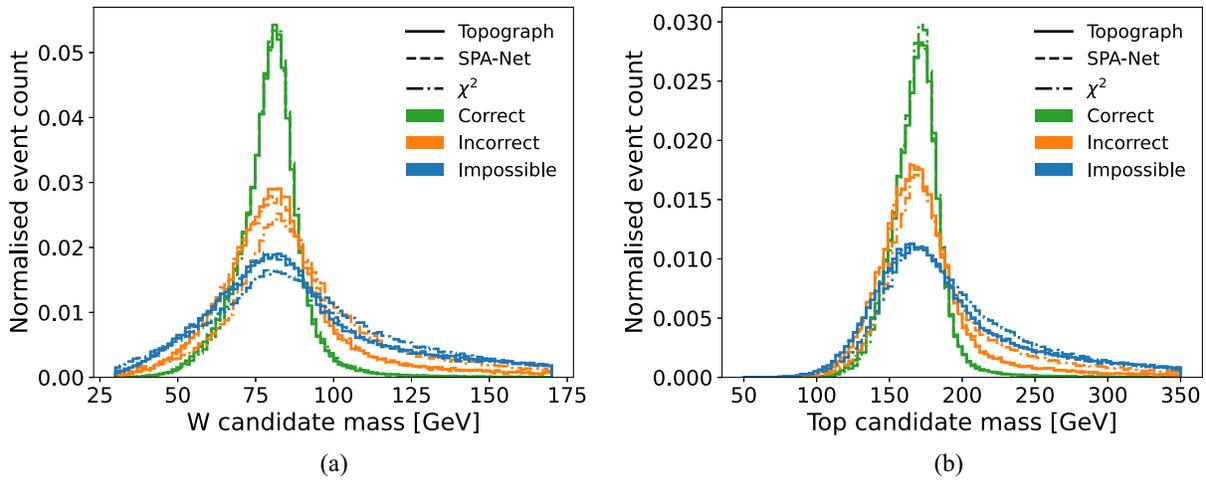


FIG. 24. Distributions of the reconstructed (a) m_W and (b) m_{top} . Each histogram is normalized to an area of one. The solid lines show the distributions for the Topograph, the dashed lines show the distributions for the SPA-Net, and the dash-dotted lines show the distributions for the χ^2 method. The different colors show the different types of events based on the assignment of the model: correct, incorrect, impossible.

Figures 22 and 23 show the same distributions for the reconstructed W and top quark mass, respectively, but every distribution is normalized to unity.

Figure 24 shows the same distributions but with the different models combined in single plots.

- [1] ATLAS Collaboration, Measurements of normalized differential cross-sections for $t\bar{t}$ production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector, *Phys. Rev. D* **90**, 072004 (2014).
- [2] CMS Collaboration, Measurement of the top quark mass using proton-proton data at $\sqrt{s} = 7$ and 8 TeV, *Phys. Rev. D* **93**, 072004 (2016).
- [3] ATLAS Collaboration, Determination of the top-quark pole mass using $t\bar{t} + 1$ -jet events collected with the ATLAS experiment in 7 TeV pp collisions, *J. High Energy Phys.* **10** (2015) 121.
- [4] ATLAS Collaboration, Top-quark mass measurement in the all-hadronic $t\bar{t}$ decay channel at $\sqrt{s} = 8$ TeV with the ATLAS detector, *J. High Energy Phys.* **09** (2017) 118.
- [5] ATLAS Collaboration, Measurements of top-quark pair differential cross-sections in the lepton + jets channel in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector, *J. High Energy Phys.* **11** (2017) 191.
- [6] CMS Collaboration, Measurement of differential cross sections for top quark pair production using the lepton + jets final state in proton-proton collisions at 13 TeV, *Phys. Rev. D* **95**, 092001 (2017).
- [7] CMS Collaboration, Measurement of the top quark mass in the all-jets final state at $\sqrt{s} = 13$ TeV and combination with the lepton + jets channel, *Eur. Phys. J. C* **79**, 313 (2019).
- [8] CMS Collaboration, Measurement of differential cross sections for the production of top quark pairs and of additional jets in lepton + jets events from pp collisions at $\sqrt{s} = 13$ TeV, *Phys. Rev. D* **97**, 112003 (2018).
- [9] CMS Collaboration, Measurement of the top quark mass with lepton + jets final states using pp collisions at $\sqrt{s} = 13$ TeV, *Eur. Phys. J. C* **78**, 891 (2018); **82**, 323 (E) (2022).
- [10] ATLAS Collaboration, Measurement of the top quark mass in the $t\bar{t} \rightarrow$ lepton + jets channel from $\sqrt{s} = 8$ TeV ATLAS data and combination with previous results, *Eur. Phys. J. C* **79**, 290 (2019).
- [11] ATLAS Collaboration, Measurement of the top-quark mass in $t\bar{t} + 1$ -jet events collected with the ATLAS detector in pp collisions at $\sqrt{s} = 8$ TeV, *J. High Energy Phys.* **11** (2019) 150.
- [12] ATLAS Collaboration, Measurements of top-quark pair differential and double-differential cross-sections in the $\ell +$ jets channel with pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector, *Eur. Phys. J. C* **79**, 1028 (2019); **80**, 1092(E) (2020).
- [13] CMS Collaboration, Measurement of differential $t\bar{t}$ production cross sections in the full kinematic range using lepton + jets events from proton-proton collisions at $\sqrt{s} = 13$ TeV, *Phys. Rev. D* **104**, 092013 (2021).
- [14] ATLAS Collaboration, Evidence for the charge asymmetry in $pp \rightarrow t\bar{t}$ production at $\sqrt{s} = 13$ TeV with the ATLAS detector, [arXiv:2208.12095](https://arxiv.org/abs/2208.12095).
- [15] ATLAS Collaboration, Evidence for single top-quark production in the s -channel in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector using the matrix element method, *Phys. Lett. B* **756**, 228 (2016).
- [16] CMS Collaboration, Measurement of spin correlations in $t\bar{t}$ production using the matrix element method in the muon + jets final state in pp collisions at $\sqrt{s} = 8$ TeV, *Phys. Lett. B* **758**, 321 (2016).
- [17] ATLAS Collaboration, Measurements of top-quark pair single- and double-differential cross sections in the all-hadronic channel in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector, *J. High Energy Phys.* **01** (2021) 033.
- [18] J. Shlomi, S. Ganguly, E. Gross, K. Cranmer, Y. Lipman, H. Serviansky, H. Maron, and N. Segol, Secondary vertex finding in jets with neural networks, *Eur. Phys. J. C* **81**, 540 (2021).
- [19] M.J. Fenton, A. Shmakov, T.-W. Ho, S.-C. Hsu, D. Whiteson, and P. Baldi, Permutationless many-jet event reconstruction with symmetry preserving attention networks, *Phys. Rev. D* **105**, 112008 (2022).
- [20] J. Erdmann, S. Guindon, K. Kröninger, B. Lemmer, O. Nackenhorst, A. Quadt, and P. Stolte, A likelihood-based reconstruction algorithm for top-quark pairs and the KLFilter framework, *Nucl. Instrum. Methods Phys. Res., Sect. A* **748**, 18 (2014).
- [21] ATLAS Collaboration, Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, *Phys. Rev. D* **97**, 072016 (2018).
- [22] F. Fiedler, A. Grohsjean, P. Haefner, and P. Schieferdecker, The matrix element method and its application to measurements of the top quark mass, *Nucl. Instrum. Methods Phys. Res., Sect. A* **624**, 203 (2010).
- [23] CMS Collaboration, Measurement of the production cross sections for a Z boson and one or more b jets in pp collisions at $\sqrt{s} = 7$ TeV, *J. High Energy Phys.* **06** (2014) 120.
- [24] A. Butter, T. Heimes, T. Martini, S. Peitzsch, and T. Plehn, Two invertible networks for the matrix element method, [arXiv:2210.00019](https://arxiv.org/abs/2210.00019).
- [25] J. S. H. Lee *et al.*, Zero-permutation jet-parton assignment using a self-attention network, [arXiv:2012.03542](https://arxiv.org/abs/2012.03542).
- [26] A. Shmakov, M.J. Fenton, T.-W. Ho, S.-C. Hsu, D. Whiteson, and P. Baldi, SPANet: Generalized permutationless set assignment for particle physics using symmetry preserving attention, *SciPost Phys.* **12**, 178 (2022).
- [27] P. W. Battaglia *et al.*, Relational inductive biases, deep learning, and graph networks, [arXiv:1806.01261](https://arxiv.org/abs/1806.01261).
- [28] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, A comprehensive survey on graph neural networks, *IEEE Trans. Neural Networks Learn. Syst.* **32**, 4 (2020).
- [29] A. Badea, W. J. Fawcett, J. Huth, T. J. Khoo, R. Poggi, and L. Lee, Solving combinatorial problems at particle colliders using machine learning, *Phys. Rev. D* **106**, 016001 (2022).
- [30] S. Qiu, S. Han, X. Ju, B. Nachman, and H. Wang, A holistic approach to predicting top quark kinematic properties with the covariant particle transformer, [arXiv:2203.05687](https://arxiv.org/abs/2203.05687).
- [31] H. Qu and L. Gouskos, ParticleNet: Jet tagging via particle clouds, *Phys. Rev. D* **101**, 056019 (2020).
- [32] E. A. Moreno, O. Cerri, J. M. Duarte, H. B. Newman, T. Q. Nguyen, A. Perival, M. Pierini, A. Serikova, M. Spiropulu, and J.-R. Vlimant, JEDI-net: A jet identification algorithm based on interaction networks, *Eur. Phys. J. C* **80**, 58 (2020).

- [33] J. Shlomi, P. Battaglia, and J.-R. Vlimant, Graph neural networks in particle physics, *Mach. Learn. Sci. Technol.* **2**, 021001 (2020).
- [34] E. Bernreuther, T. Finke, F. Kahlhoefer, M. Krämer, and A. Mück, Casting a graph net to catch dark showers, *SciPost Phys.* **10**, 046 (2021).
- [35] X. Ju and B. Nachman, Supervised jet clustering with graph neural networks for Lorentz boosted bosons, *Phys. Rev. D* **102**, 075014 (2020).
- [36] J. Guo, J. Li, T. Li, and R. Zhang, Boosted Higgs boson jet reconstruction via a graph neural network, *Phys. Rev. D* **103**, 116025 (2021).
- [37] F. A. Dreyer and H. Qu, Jet tagging in the Lund plane with graph networks, *J. High Energy Phys.* **03** (2021) 052.
- [38] A. Hariri, D. Dyachkova, and S. Gleyzer, Graph generative models for fast detector simulations in high energy physics, [arXiv:2104.01725](https://arxiv.org/abs/2104.01725).
- [39] G. DeZoort, S. Thais, J. Duarte, V. Razavimaleki, M. Atkinson, I. Ojalvo, M. Neubauer, and P. Elmer, Charged particle tracking via edge-classifying interaction networks, *Comput. Software Big Sci.* **5**, 26 (2021).
- [40] O. Atkinson, A. Bhardwaj, C. Englert, V. S. Ngairangbam, and M. Spannowsky, Anomaly detection with convolutional graph neural networks, *J. High Energy Phys.* **08** (2021) 080.
- [41] S. Thais, P. Calafiura, G. Chachamis, G. DeZoort, J. Duarte, S. Ganguly, M. Kagan, D. Murnane, M. S. Neubauer, and K. Terao, Graph neural networks in particle physics: Implementations, innovations, and challenges, in *2022 Snowmass Summer Study* (2022), [arXiv:2203.12852](https://arxiv.org/abs/2203.12852).
- [42] S. Gong, Q. Meng, J. Zhang, H. Qu, C. Li, S. Qian, W. Du, Z.-M. Ma, and T.-Y. Liu, An efficient Lorentz equivariant graph neural network for jet tagging, *J. High Energy Phys.* **07** (2022) 030.
- [43] CMS Collaboration, Identification of highly Lorentz-boosted heavy particles using graph neural networks and new mass decorrelation techniques, Report No. CMS-DP-2020-002, 2020.
- [44] ATLAS Collaboration, Graph neural network jet flavour tagging with the ATLAS detector, Report No. ATL-PHYS-PUB-2022-027, 2022.
- [45] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, *J. High Energy Phys.* **07** (2014) 079.
- [46] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk, Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations, *J. High Energy Phys.* **03** (2013) 15.
- [47] K. Zoch, J. A. Raine, L. Ehrke, D. Sengupta, M. Guth, and T. Golling, Top quark pair production at the LHC, with all hadronic resolved decays for solving event combinatorics, [10.5281/zenodo.7737248](https://doi.org/10.5281/zenodo.7737248) (2023).
- [48] T. Sjöstrand, S. Mrenna, and P. Skands, A brief introduction to PYTHIA 8.1, *Comput. Phys. Commun.* **178**, 852.
- [49] J. de Favereau, C. Delaere, C. Delaere, A. Giammanco, V. Lemaître, V. Lemaître, and M. Selvaggi, DELPHES 3, A modular framework for fast simulation of a generic collider experiment, *J. High Energy Phys.* **02** (2014) 057.
- [50] M. Cacciari, G. P. Salam, and G. Soyez, The anti-kt jet clustering algorithm, *J. High Energy Phys.* **04** (2008) 063.
- [51] M. Cacciari, G. P. Salam, and G. Soyez, FastJet user manual, *Eur. Phys. J. C* **72**, 1 (2012).
- [52] M. Abadi *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems (2015), software available from <https://www.tensorflow.org/>.
- [53] P. Skands, S. Carrazza, and J. Rojo, Tuning PYTHIA 8.1: The Monash 2013 tune, *Eur. Phys. J. C* **74**, 3024 (2014).