

Topology-agnostic, hierarchical graph network for event reconstruction

Nathalie Soybelman^{✉,*}, Nilotpal Kakati[✉], and Eilam Gross[✉]

Weizmann Institute of Science, Rehovot, Israel

Francesco A. Di Bello^{✉,†}

University of Genova, Genoa, Italy



(Received 28 October 2025; accepted 1 December 2025; published 26 January 2026)

Event reconstruction at the LHC, the task of assigning observed physics objects to their true origins, is a central challenge for precision measurements and searches. Many existing machine learning approaches address this problem but rely on a single event topology, restricting their applicability to realistic analyses where multiple signal and background processes with different structures are present. To overcome this, we present TIGER, a novel hierarchical graph network that is fundamentally topology agnostic. By incorporating only the common underlying structure of sequential two-body decays, our model can reconstruct complex events without process-specific assumptions. This flexible architecture supports multitask learning, enabling simultaneous event reconstruction and classification. TIGER thus provides a powerful and generalizable tool for physics analysis at the LHC.

DOI: [10.1103/z57h-y3rx](https://doi.org/10.1103/z57h-y3rx)

I. INTRODUCTION

At the Large Hadron Collider (LHC) [1], the reconstruction of collision events is a highly important and complex task. Since unstable particles decay in cascades before reaching the detector, only their resulting decay products are captured. The raw detector signals are first reconstructed into stable final-state particles by a subsequent chain of reconstruction algorithms, which are then clustered into high-level objects like jets [2–4]. Our work addresses the final combinatorial challenge: accurately assigning these objects to their original “mother” particles. This is fundamental to nearly all particle physics analyses. However, in events involving multiple unstable particles, the large combinatorics become intractable for traditional methods such as χ^2 and likelihood approaches [5–10].

To overcome these limitations, various machine learning approaches emerged in recent years, designed to tackle the combinatorial problem by embedding physical knowledge directly into their network architectures. Notable examples include Topograph [5], a graph neural network that exploits the hierarchical decay structure of top quarks ($t \rightarrow bW \rightarrow bq\bar{q}$) to reconstruct the W -boson before the

top quark. Similarly, HyPER [11], another graph neural network, utilizes both edges and hyperedges to infer two- and three-particle decay assignments. Furthermore, SPANet [6,7,12,13] employs a transformer-based architecture with symmetry-preserving attention mechanisms to account for inherent symmetries among decay products. In addition, a theory-informed framework [14] utilizes reinforcement learning to maximize the matrix element, enabling the reconstruction of the decay chain at the parton level.

While the integration of physics knowledge is crucial for enhancing model performance and interpretability, it can also introduce significant limitations regarding generalizability. For instance, while HyPER and Topograph show strong performance in reconstructing $t\bar{t}$ events, their architectures will require modifications to address other decay topologies. In a realistic analysis, background events would be forced into the same predefined topology as the signal, limiting the performance of such specialized models. Although the SPANet architecture offers greater generality and has been evaluated across various decay processes, it still necessitates explicit input of the event topology.

To address these shortcomings, we introduce a novel topology-agnostic approach that incorporates fundamental physical constraints without sacrificing generality or introducing bias. Our method is built on the observation that, in most cases, particles decay in at most two stages, each producing two daughter particles (e.g., $t \rightarrow Wb \rightarrow qq'b$, $H \rightarrow ZZ \rightarrow 4l$). These assumptions hold for nearly all processes studied at the LHC. However, certain cases such as trilinear Higgs decays, effective theories, including Fermi theory for low energy hadron decays, or possible new physics scenarios can involve three or more prong

*Contact author: nathalie.soybelman@weizmann.ac.il

†Present address: University of Pisa, Italy.

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

decays or additional decay stages. In such situations, our method can be naturally extended, though exploring these generalizations lies beyond the scope of this work.

Importantly, our approach requires no prior knowledge of the event topology or the particles it comprises, making this a highly flexible method capable of simultaneously reconstructing and classifying both signal and background events across diverse event structures. In this work, we first demonstrate the model’s reconstruction performance on fully hadronic $t\bar{t}$ and semileptonic $t\bar{t}H$ events. Subsequently, we evaluate its effectiveness in signal-versus-background discrimination by classifying between $t\bar{t}H$ and $t\bar{t} + bb$ events.

II. DATASETS

This study utilizes two distinct datasets, taken from the studies presented in [11,12]. These datasets describe both fully hadronic and semileptonic event topologies.

The first dataset [15] comprises fully hadronic $t\bar{t}$ events. These events were generated with the following simulation chain: MadGraph5_aMC@NLO (v2.9.16) [16] for matrix element generation, Pythia8 (v8.306) [17] for parton showering and hadronization, and Delphes (v3.5.0) [18] for detector simulation, employing the ATLAS detector card.

For jet reconstruction, the anti- k_t [19] algorithm was applied with a radius parameter of $R = 0.4$. A p_T -dependent tagging efficiency was used for b tagging. Reconstructed jets were required to satisfy $p_T > 25$ GeV and $|\eta| < 2.4$. Events were selected if they contained a minimum of six jets, with at least two of these being b tagged. The dataset is partitioned into approximately 7.9 million events for training, and 440,000 events each for validation and testing. Further details regarding this simulation can be found in [11].

The second dataset [20] focuses on semileptonic $t\bar{t}H$ events, where one top quark decays hadronically and the other semileptonically, and the Higgs boson decays into two b jets. The simulation chain for these events involved MadGraph5_aMC@NLO (v3.2.0) [16], Pythia8 (v8.2) [17], and Delphes (v3.4.2) [18], configured with the CMS detector card.

Jets were clustered using the anti- k_t algorithm with a radius parameter of $R = 0.5$. Selected jets were required to have $p_T > 25$ GeV and $|\eta| < 2.5$. The same b -tagging procedure used for the fully hadronic $t\bar{t}$ dataset was applied. Electrons and muons were selected with the same p_T and η cuts. No specific requirements were imposed on the missing transverse energy (E_T^{miss}). Event selection required the presence of one lepton and a minimum of six jets. For the training and validation sets, a loose requirement of at least two b -tagged jets is applied, whereas the testing set follows a stricter criterion of at least four b -tagged jets. The difference in event selection matches how the dataset was originally defined and used in the referenced study [12], and we therefore keep the same selection to remain consistent with that benchmark.

This dataset is allocated with approximately 9.9 million events for training, 520,000 for validation, and 240,000 for testing.

For the signal and background classification study, this $t\bar{t}H$ dataset was combined with a $t\bar{t} + bb$ sample, which represents the most dominant background for the described process. The $t\bar{t} + bb$ dataset adds five million events for training, 270,000 for validation, and 150,000 for testing.

III. TIGER

We introduce a novel method, topology-independent graph-based event reconstruction (TIGER), which employs a hierarchical, physics-inspired graph learning approach.

A. Network architecture

The main building blocks of our model are graph neural networks and transformers; both have demonstrated broad applicability within particle physics, including tasks such as jet tagging [21–23], event classification [24], anomaly detection [25], reconstructions of particles [2–4] or tracks [26–30], as well as in fast simulations of detector responses [31].

The overall network architecture consists of three primary components: encoding, hierarchical graph learning, and an optional event classification block. Each of these parts is described in detail below. An illustration of the complete architecture is provided in Fig. 1.

Encoding. The encoding stage begins by taking the input features of jets and leptons (p_T , η , ϕ , m) and missing transverse energy (MET) (E_T , ϕ). These features are then concatenated with their respective one-hot-encoded b -tag or lepton-tag information. The resulting feature vectors are then embedded into a hidden dimension using a separate multi-layer perceptron (MLP) for each particle class. Following this embedding, all input objects are treated uniformly as nodes in the graph. This approach treats the MET as a particle, implicitly assuming it corresponds to a neutrino. While this simplification may not be entirely accurate from a physics perspective, as MET can arise from multiple neutrinos, new physics, or detector issues, it preserves the model’s simplicity and generality for the studies presented here. An alternative approach used in SPANet [12], where MET is treated as global event information, could be implemented, but was not found to be necessary for this work.

We then employ a diffusion transformer (DiT), an architecture originally proposed in [32], to update these hidden representations. To provide contextual information, the mean of the hidden representations of all nodes, as well as the total number of objects in the event, are passed to the transformer. The updated hidden representations are subsequently passed to the hierarchical graph learning module.

Optionally, an auxiliary task can be performed at this stage to classify the origin of the particles. This task can be designed based on the expected signal and background

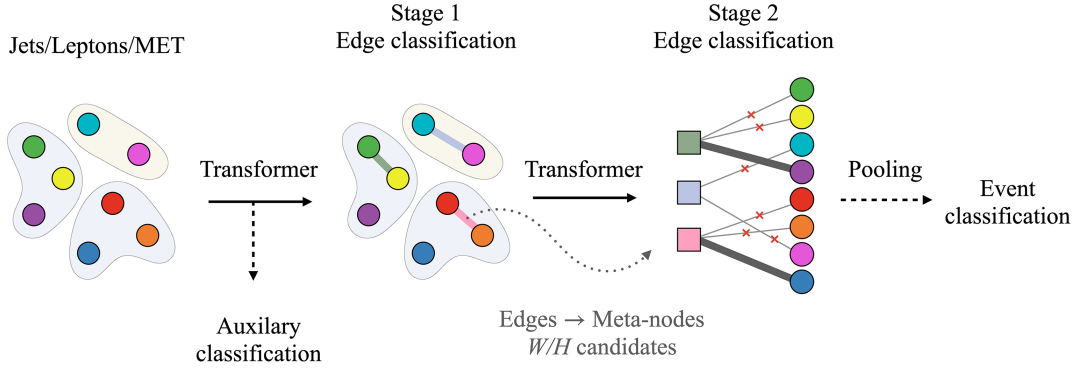


FIG. 1. Overview of the TIGER architecture. After an initial transformer-based encoding, a two-stage hierarchical learning module reconstructs the event. The first stage identifies intermediate particles (W , H) as metanodes, while the second stage combines them with other nodes to form top quarks. An optional final block pools features from the graph for event classification.

processes, by considering all possible particle origins and defining corresponding classification labels. For the specific case of the $t\bar{t}$ dataset, the classes are as follows: not associated, associated with a reconstructable W , associated with a nonreconstructable W , b jet from a reconstructable top, and b jet from a nonreconstructable top. For the semileptonic $t\bar{t}H$ dataset, these classes are naturally extended through the addition of the Higgs and the leptonic decay mode of the W . While the predictions from this auxiliary classification task are not directly used in the final reconstruction, we observe that it improves the overall network performance and interpretability.

Hierarchical graph learning. This component of the architecture is loosely inspired by previous work [33,34], but has been substantially modified to incorporate physical principles. The hierarchical graph learning proceeds in two stages, corresponding to the two levels in the decay chain.

In the first stage, a fully connected graph is constructed from the encoded input objects. Edge representations are created by summing the representations of connected nodes, ensuring invariance to the order of nodes. These are then processed through a multilayer perceptron (MLP) for edge classification. The goal of this step is to determine whether two reconstructed objects originate from the same parent particle, and if so, to identify the type of that parent. Depending on the process under study, this can be formulated as either a binary or multiclass classification task. For example, in $t\bar{t}$ events, the task involves identifying whether two objects form a W boson, while in $t\bar{t}H$ events, the classification extends to distinguishing between no connection, a W , or a Higgs boson.

In the second stage, the goal is to combine the intermediate particles identified in the first stage with other particles to form the final-state particles. To achieve this, a new graph is created with two types of nodes: the original nodes from the first graph and *metanodes*, which represent the edges classified as the intermediate parent particles. Depending on the expected processes, one can fully connect all initial nodes with the metanodes (e.g., for

top decays), the metanodes among themselves (e.g., diboson Higgs decays like $H \rightarrow WW$), or both. The following edge classification is conceptionally identical to the one in the first stage.

The metanodes are selected based on the reconstruction probability predicted in the first stage. A maximum of k metanodes are chosen if their probability exceeds a certain threshold p_{th} , which are tunable hyperparameters. Notably, nodes that are incorporated into metanodes are not discarded during the training process. This is to prevent errors from the first stage from propagating and negatively impacting the performance of the second stage. For instance, if a b jet is incorrectly included in a fake W , it would be unavailable for correct matching in the second stage. The model may also produce ambiguous assignments, such as two W candidate predictions sharing a jet, or unphysical pairings where a metanode is matched with one of its own constituents. To ensure training stability, these configurations are permitted to form but are discouraged by the loss. At the evaluation stage, a dedicated algorithm resolves these conflicts to guarantee a physically consistent output and prevent any double counting of particles.

In cases where rare event topologies involving multiple decay stages or three-prong structures are expected, the approach can be generalized through an additional hierarchical level and a transition from edge to hyperedge representations.

Event classification. The final, optional component of the architecture is an event classification block for signal versus background discrimination. The representation of the hierarchical graph is pooled by concatenating the mean of all nodes, the mean of all metanodes, the number of nodes, and the number of metanodes. This pooled representation is then passed through an MLP to perform a binary classification of the event.

This event-level classification can also serve as an auxiliary task for signal-only datasets. For the $t\bar{t}$ dataset, we found that classifying the number of reconstructable tops (0, 1, or 2) improved performance. However, a similar auxiliary task for the $t\bar{t}H$ dataset classifying the number of

reconstructable tops and Higgs ($0t0H, 1t0H, 2t0H, 0t1H, 1t1H, 2t1H$) did not yield improvements, indicating that the utility of such a task should be evaluated on a case-by-case basis.

The total loss is the weighted sum of the (binary) cross-entropy losses from the two stages of the hierarchical graph learning $\mathcal{L}_{\text{stage}\{1,2\}}$, the auxiliary task following the encoding step $\mathcal{L}_{\text{aux}}$, and the event-level classification $\mathcal{L}_{\text{event}}$ with weights $\lambda_{\{1,2\}}$:

$$\mathcal{L} = (1 - 0.5\alpha)\mathcal{L}_{\text{stage } 1} + 0.5\alpha\mathcal{L}_{\text{stage } 2} + \lambda_1\mathcal{L}_{\text{aux}} + \lambda_2\mathcal{L}_{\text{event}}. \quad (1)$$

The weights were adjusted empirically for each training to ensure that all loss components contribute on a comparable scale to the total loss. For the $t\bar{t}$ dataset we chose $\lambda_1 = 1/7$ and $\lambda_2 = 1/6$. For the pure $t\bar{t}H$ trainings, where both the auxiliary task and the event-level classification involve a larger number of classes (nine and six, respectively), the corresponding losses are naturally higher; therefore, we used lower scaling factors, $\lambda_1 = \lambda_2 = 1/50$. In the combined $t\bar{t}H + t\bar{t}b\bar{b}$ dataset, the auxiliary task remains unchanged, but the event-level classification becomes a simple binary decision, producing a significantly smaller raw loss. Consequently, we set the event-level scaling to $\lambda_2 = 1$.

The weight α is an epoch-dependent function ranging from 0 to 1: at the beginning of training, the network emphasizes the first stage of graph learning for stability, and the contribution of the two stages is gradually balanced such that $\alpha = 1$ is reached at epoch 40. The network has a total of 1.1 million parameters, and all training was performed on a single NVIDIA RTX A6000 GPU. Each network was trained for 500 epochs, taking around 30 hours.

B. Interpretation of network outputs

To translate the network's probabilistic outputs into a definite set of reconstructed particles, we employ a post-processing algorithm. This algorithm interprets the edge probabilities from both hierarchical stages to build physical objects. The primary constraint of the algorithm is the prevention of double counting, ensuring each input particle is assigned to at most one final-state object. While the algorithm can be defined for any collection of processes, we illustrate it here using fully hadronic $t\bar{t}$ events for clarity, followed by its extension to the more complex semileptonic $t\bar{t}H$ topology.

Fully hadronic $t\bar{t}$ events. For the $t\bar{t}$ topology, the reconstruction is determined by two probability thresholds p_W and p_t , for W -boson and top quark identification, respectively. The algorithm is a greedy, iterative process:

- (1) All potential W candidates (jet pairs) from the first hierarchical stage are sorted in descending order of their predicted probability.
- (2) The algorithm iterates through this sorted list. For each candidate, it checks if its probability exceeds

p_W and if its constituent jets have not already been assigned to another object.

- (3) If the W candidate is valid, the algorithm consults the second-stage predictions to see if this W metanode forms a top quark with a third, unassigned jet, with a probability exceeding p_t .
- (4) If a top quark is successfully identified, all three constituent jets are marked as “used.” If the W candidate was valid but did not form a top quark, only its two jets are marked as “used.”
- (5) This process repeats until no remaining W candidates in the list satisfy the p_W threshold.

This procedure inherently prevents double counting. The thresholds can be optimized to tune efficiency and purity. It is required that p_W is not set lower than the metanode selection threshold p_{th} used during training, as lower-probability W candidates would not have a corresponding metanode in the second stage. The algorithm does not enforce a fixed number of reconstructed objects, preserving the model's generality.

Semileptonic $t\bar{t}H$ events. The extension to $t\bar{t}H$ events introduces the Higgs boson and requires an additional probability threshold, p_H . We empirically set a stricter cut for the Higgs boson ($p_H > p_W$). Since the first-stage classification is now multiclass (distinguishing between W , H , etc.), the algorithm is adapted as follows.

The algorithm considers a single list of all potential W and H candidates, sorted by their maximum class probability. During the iteration, a candidate is accepted only if its probability exceeds the corresponding threshold and its constituents are available. If a candidate is identified, its constituents are marked as used, and it proceeds to the second stage for top quark reconstruction as before. The algorithm treats both hadronic and leptonic W boson reconstructions uniformly. The leptonic W , formed by the lepton and MET, is a trivial combination for the network to identify in this event topology.

The generalization to multiprocesses is straightforward by defining the appropriate thresholds for each particle type without modifying the architecture.

IV. RESULTS

We evaluate the performance of TIGER by comparing its reconstruction efficiency and purity against established baseline models. A direct comparison presents a challenge due to a fundamental difference in design philosophy. Baseline models like HyPER and SPANet are specialized for a known event topology and typically predict a fixed number of objects. In contrast, TIGER is intentionally topology agnostic, making no *a priori* assumptions about the event structure or the number of reconstructable particles. This agnostic approach is designed to be more applicable to realistic experimental conditions, where the event topology is not known in advance. While specialized models may achieve high efficiency by constraining the

TABLE I. Performance metrics for different jet multiplicities, comparing the HyPER and TIGER methods. The efficiencies of HyPER are taken from [11], and purities are calculated from the dataset using the efficiencies. Underlined entries indicate efficiencies at which TIGER was aligned with HyPER for a fair purity comparison. Higher purity performance is indicated in bold.

	HyPER efficiency [%]			TIGER efficiency [%]			HyPER purity [%]			TIGER purity [%]		
	$t\bar{t}$	t	W	$t\bar{t}$	t	W	$t\bar{t}$	t	W	$t\bar{t}$	t	W
6 j	80	71	75	78	70	75	16	40	49	34	54	55
7 j	66	68	71	65	67	72	19	41	49	28	49	51
≥ 8 j	51	61	65	50	61	67	18	40	47	23	44	42
All	66	<u>67</u>	<u>71</u>	65	<u>66</u>	<u>72</u>	18	40	49	29	49	50

problem space, this can come at the cost of purity when applied to datasets containing a mix of fully and partially reconstructable events. Our objective is to demonstrate that TIGER's generalized framework achieves efficiencies comparable to specialized models while yielding significantly higher purity.

A. Fully hadronic $t\bar{t}$ events

For this analysis, we set the probability thresholds to $p_W = p_T = 0.3$. These values were chosen to match the

overall top quark efficiency of the HyPER baseline, allowing for a direct comparison of purity. Efficiency for a given particle type is defined as the ratio of correctly reconstructed particles to the total number of reconstructable particles at the truth level ($N_{\text{correct}}^p / N_{\text{reconstructable}}^p$). Purity is the ratio of correctly reconstructed particles to the total number of predicted particles ($N_{\text{correct}}^p / N_{\text{all pred}}^p$).

As summarized in Table I, TIGER achieves efficiencies comparable to the baseline while demonstrating significantly higher purities. For single top quarks, our model's

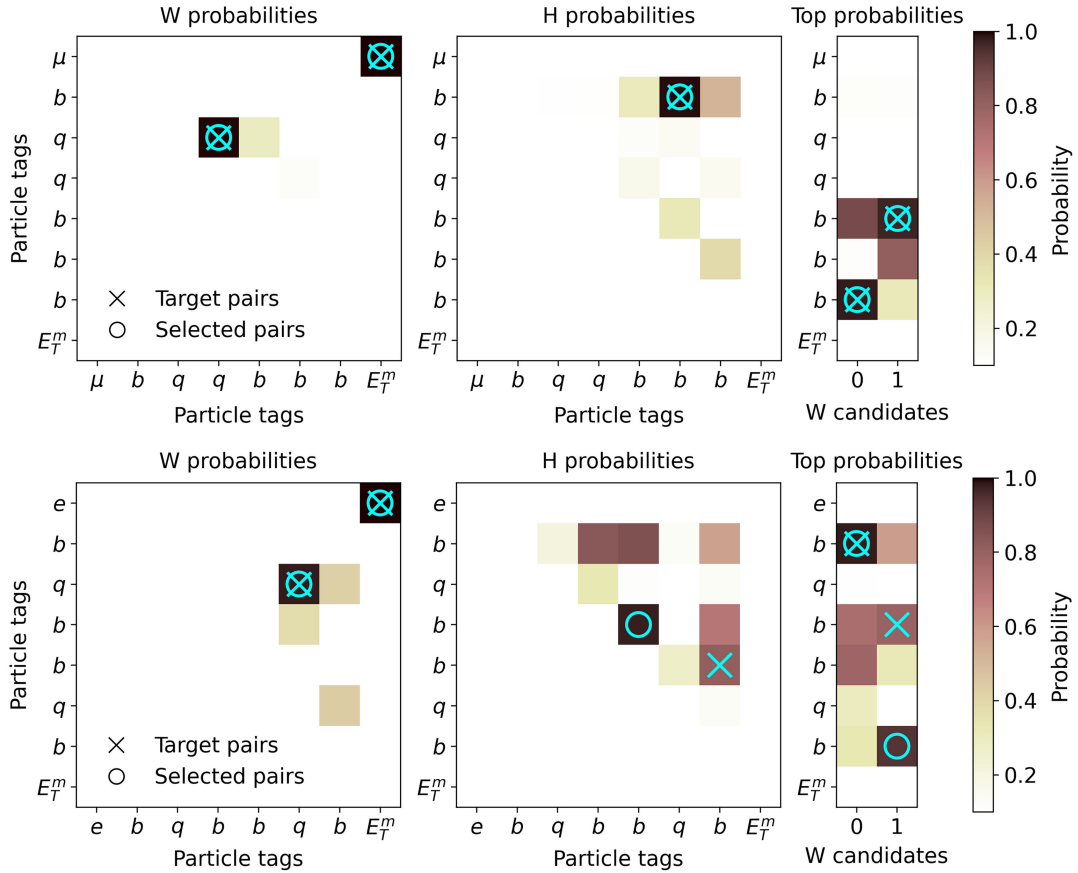


FIG. 2. Two event displays from the $t\bar{t}H$ test dataset. The left and middle matrices show the classification probabilities from the first stage of the hierarchical graph, while the right matrix depicts the second-stage incidence matrix, showing only the W candidates selected by the first stage. The selected pairs are determined by the algorithm described in Sec. III B. The top panel shows a correctly reconstructed event, whereas the bottom panel illustrates a misassignment between a b jet from the Higgs and the top.

TABLE II. Efficiency and purity comparison for $t\bar{t}H$ events. TIGER and the SPANet baseline are evaluated across varying jet multiplicities. Reconstruction targets include the full event ($t\bar{t}H$), hadronic top (t_h), leptonic top (t_l), and Higgs boson (H). SPANet values were taken from [12]. Underlined entries indicate efficiencies at which TIGER was aligned with SPANet for a fair purity comparison. Higher purity performance is indicated in bold.

	SPANet efficiency [%]				TIGER efficiency [%]				SPANet purity [%]				TIGER purity [%]			
	$t\bar{t}H$	t_h	t_l	H	$t\bar{t}H$	t_h	t_l	H	$t\bar{t}H$	t_h	t_l	H	$t\bar{t}H$	t_h	t_l	H
6 j	54	54	69	50	50	51	70	51	15	24	63	38	32	35	64	50
7 j	42	48	68	49	39	46	69	48	16	26	63	38	27	34	64	47
≥ 8 j	33	42	68	47	30	40	67	45	14	26	63	37	22	31	62	42
All	45	49	69	<u>49</u>	40	46	69	<u>49</u>	15	25	63	38	27	34	64	47

purity is nearly 10% higher on average. The most striking difference is observed in the $t\bar{t}$ event-level purity, where our performance is more than double that of the baseline at low jet multiplicity. This discrepancy arises because only approximately 27% of events in the dataset contain two fully reconstructable top quarks. Specialized models designed to always predict two top quarks are therefore inherently prone to incorrect reconstructions in the remaining 73% of events.

For single W bosons, TIGER shows a 6% purity improvement for events with six jets. At high multiplicity, HyPER achieves better purity, as it inherently limits predictions to the two highest-probability candidates, which artificially reduces the rate of fake reconstructions. While the authors of HyPER note that probability cuts could be implemented, the impact of such a change on their reported performance is not shown.

B. Semileptonic $t\bar{t}H$ events

To test our model on a more complex event topology, we use the semileptonic $t\bar{t}H$ dataset and compare our results against SPANet. We set the probability thresholds to $p_W = p_T = 0.3$ and $p_H = 0.9$, with the value of p_H chosen to align our Higgs boson efficiency with that of SPANet for a fair comparison. We note that the Higgs and hadronic top efficiencies are highly sensitive to the p_H threshold, as the primary source of misreconstruction is the confusion between b jets originating from the Higgs and those from the hadronic top decay. Two exemplary reconstructed events are displayed in Fig. 2. While the top-panel event is correctly reconstructed, the bottom panel illustrates the misassignment described above.

The performance metrics are presented in Table II. While efficiencies for the semileptonic top and Higgs boson are comparable, our model slightly underperforms in reconstructing the hadronic top and the full event. However, TIGER again demonstrates a substantial improvement in purity, outperforming the baseline by approximately 10% on average for hadronic tops, Higgs bosons, and full-event reconstructions.

This performance is achieved despite dataset imbalances between the training and testing sets due to the different

b -tagging requirements. These differences likely pose a greater challenge for our agnostic approach, which does not presuppose a specific event structure. While later work has explored applying probability cuts to SPANet outputs [13], its architecture remains fundamentally constrained by the predefined event topology. For example, it could never predict two Higgs bosons in an event, which is explicitly allowed in our model, further enhancing its generality.

C. Classification of $t\bar{t}H$ and $t\bar{t} + bb$

Finally, we evaluate the model's capability for signal-versus-background discrimination. Using the combined $t\bar{t}H$ and $t\bar{t} + bb$, we train TIGER with the event classification head simultaneously with the reconstruction components. This end-to-end approach is compared to SPANet, which also uses a global event representation for classification. The authors of [12] found that fine-tuning a pretrained SPANet backbone for classification yielded the best performance. As shown by the ROC curves in Fig. 3, TIGER outperforms even the optimized fine-tuning

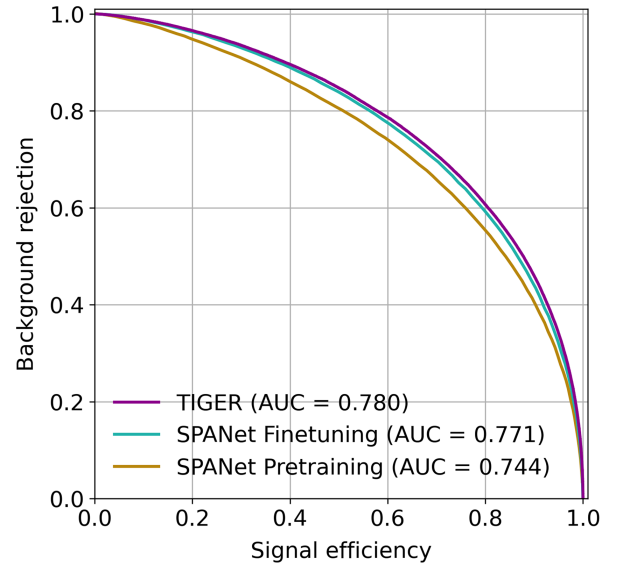


FIG. 3. Receiver operating characteristic (ROC) curves for the $t\bar{t}H$ vs $t\bar{t} + bb$ classification, comparing TIGER with SPANet baselines (results taken from [12]).

strategy, demonstrating its effectiveness as a multitask learning framework.

V. CONCLUSIONS

In this work, we introduced TIGER, a novel, topology-agnostic method for event reconstruction. Its hierarchical graph architecture is inspired by the fundamental physics principle of sequential two-body particle decays. Unlike specialized models such as HyPER and SPANet, which require a predefined event topology as an architectural prior or direct input, TIGER makes no such assumptions during either training or evaluation. This makes the learning task more challenging but yields a far more flexible and realistic framework for real-world physics analysis.

On top of being topology agnostic, we demonstrated on fully hadronic $t\bar{t}$ and semileptonic $t\bar{t}H$ events that this approach also achieves better performance compared to specialized baselines. This is reflected in the fact that, for similar reconstruction efficiencies, we deliver significantly higher purity. The improvement in purity is a direct consequence of the model's agnosticism; it avoids the inherent bias of methods that are optimized to find a complete event structure, even when many events are not fully reconstructable. Furthermore, we showed that TIGER's architecture seamlessly integrates event classification, outperforming a fine-tuned baseline in a signal-versus-background task without requiring a multistage training process. Collectively, these results establish TIGER as a powerful and flexible tool for data-driven event reconstruction.

The performance demonstrated in this study is primarily limited by the provided datasets. Future work will focus on deploying the model with more realistic detector simulations to better assess its performance under experimental conditions. The model's flexibility allows for the natural inclusion of additional information, such as jet substructure variables or outputs from advanced flavor-tagging algorithms, which could further boost performance.

Ultimately, this approach could be extended to complex analyses involving multiple overlapping signal and background topologies, such as searches for rare $H \rightarrow c\bar{c}$ decays [35,36] or di-Higgs events [37], paving the way toward a unified model capable of addressing diverse physics analyses at the LHC.

ACKNOWLEDGMENTS

We thank Alberto Ibanez, Zihan Zhang, Alexander Shmakov, and Hideki Okawa for insightful discussions and clarifications concerning the baseline models used in our comparisons. N. S., N. K., and E. G. are supported by the Minerva Stiftung under Grant No. 715027, and the Weizmann Institute for Artificial Intelligence Grant Program Reference No. 151676.

DATA AVAILABILITY

The data that support the findings of this article are openly available [15,20] and our code is available in [38].

-
- [1] L. Evans and P. Bryant, LHC machine, *J. Instrum.* **3**, S08001 (2008).
 - [2] F. A. Di Bello, E. Dreyer, S. Ganguly, E. Gross, L. Heinrich, A. Ivina, M. Kado, N. Kakati, L. Santi, J. Shlomi, and M. Tusoni, Reconstructing particles in jets using set transformer and hypergraph prediction networks, *Eur. Phys. J. C* **83**, 596 (2023).
 - [3] N. Kakati, E. Dreyer, A. Ivina, F. A. Di Bello, L. Heinrich, M. Kado, and E. Gross, HGPflow: Extending hypergraph particle flow to collider event reconstruction, *Eur. Phys. J. C* **85**, 847 (2025).
 - [4] J. Pata, J. Duarte, F. Mokhtar, E. Wulff, J. Yoo, J.-R. Vlimant, M. Pierini, and M. Girone, Machine learning for particle flow reconstruction at cms, *J. Phys. Conf. Ser.* **2438**, 012100 (2023).
 - [5] L. Ehrke, J. A. Raine, K. Zoch, M. Guth, and T. Golling, Topological reconstruction of particle physics processes using graph neural networks, *Phys. Rev. D* **107**, 116019 (2023).
 - [6] 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).
 - [7] 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).
 - [8] 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.
 - [9] 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.
 - [10] J. Erdmann, S. Guindon, K. Kroeninger, 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).
 - [11] C. Birch-Sykes, B. Le, Y. Peters, E. Simpson, and Z. Zhang, Reconstructing short-lived particles using hypergraph representation learning, *Phys. Rev. D* **111**, 032004 (2025).

- [12] M. J. Fenton, A. Shmakov, H. Okawa, Y. Li, K.-Y. Hsiao, S.-C. Hsu, D. Whiteson, and P. Baldi, Reconstruction of unstable heavy particles using deep symmetry-preserving attention networks, *Commun. Phys.* **7**, 139 (2024).
- [13] H. Li, M. Stamenkovic, A. Shmakov, M. Fenton, D. S.-C. Chao, K. M. White, C. Mikkelsen, J. Mitic, C. M. Suarez, M. Quinnan, G. Landsberg, H. Newman, P. Baldi, D. Whiteson, and J. Duarte, Reconstruction of boosted and resolved multi-higgs-boson events with symmetry-preserving attention networks, *J. High Energy Phys.* **11** (2025) 119.
- [14] B. M. Dillon and M. Spannowsky, Theory-informed neural networks for particle physics, [arXiv:2507.13447](https://arxiv.org/abs/2507.13447).
- [15] Z. Zhang, C. Birch-Sykes, B. Le, Y. Peters, and E. Simpson, Top quark pair production with fully hadronic final states at the LHC, [10.5281/zenodo.10653837](https://doi.org/10.5281/zenodo.10653837) (2024).
- [16] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, MadGraph 5: Going beyond, *J. High Energy Phys.* **06** (2011) 128.
- [17] T. Sjöstrand, S. Mrenna, and P. Skands, A brief introduction to PYTHIA 8.1, *Comput. Phys. Commun.* **178**, 852 (2008).
- [18] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi (DELPHES 3 Collaboration), DELPHES 3, A modular framework for fast simulation of a generic collider experiment, *J. High Energy Phys.* **02** (2014) 057.
- [19] M. Cacciari, G. P. Salam, and G. Soyez, The anti-ktjet clustering algorithm, *J. High Energy Phys.* **04** (2008) 063.
- [20] https://mlphysics.ics.uci.edu/data/2023_spanet/
- [21] S. Mondal and L. Mastrolorenzo, Machine learning in high energy physics: A review of heavy-flavor jet tagging at the LHC, *Eur. Phys. J. Special Topics* **233**, 2657 (2024).
- [22] G. Aad *et al.* (ATLAS Collaboration), Transforming jet flavour tagging at ATLAS, [arXiv:2505.19689](https://arxiv.org/abs/2505.19689).
- [23] J. Brehmer, V. Bresó, P. de Haan, T. Plehn, H. Qu, J. Spinner, and J. Thaler, A lorentz-equivariant transformer for all of the LHC, *SciPost Phys.* **19**, 108 (2025).
- [24] J. Ho, B. R. Roberts, S. Han, and H. Wang, Pretrained event classification model for high energy physics analysis, [arXiv:2412.10665](https://arxiv.org/abs/2412.10665).
- [25] J. Y. Araz, D. Athanasakos, M. Ploskon, and F. Ringer, Graph theory inspired anomaly detection at the LHC, [arXiv:2506.19920](https://arxiv.org/abs/2506.19920).
- [26] P. Calafiura, J. Chan, L. Delabrouille, and B. Wang, EggNet: An evolving graph-based graph attention network for particle track reconstruction, [arXiv:2407.13925](https://arxiv.org/abs/2407.13925).
- [27] D. Murnane, S. Thais, and A. Thete, Equivariant graph neural networks for charged particle tracking, in *21th International Workshop on Advanced Computing and Analysis Techniques in Physics Research: AI meets Reality* (2023), [arXiv:2304.05293](https://arxiv.org/abs/2304.05293).
- [28] R. Liu, P. Calafiura, S. Farrell, X. Ju, D. T. Murnane, and T. M. Pham, Hierarchical graph neural networks for particle track reconstruction, [arXiv:2303.01640](https://arxiv.org/abs/2303.01640).
- [29] R. Miller, A. Shmakov, K. Oh, J. Lee, P. Baldi, L. Condren, M. Vessella, and D. Whiteson, Fast and precise track fitting with machine learning, [arXiv:2505.02937](https://arxiv.org/abs/2505.02937).
- [30] L. Heinrich, B. Huth, A. Salzburger, and T. Wettig, Combined track finding with GNN & CKF, in *8th International Connecting the Dots Workshop (CTD 2023)* (2024), [arXiv:2401.16016](https://arxiv.org/abs/2401.16016).
- [31] D. Kobylanskii, N. Soybelman, E. Dreyer, and E. Gross, Graph-based diffusion model for fast shower generation in calorimeters with irregular geometry, *Phys. Rev. D* **110**, 072003 (2024).
- [32] W. S. Peebles and S. Xie, Scalable diffusion models with transformers, in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, (IEEE International Conference on Computer Vision, 2022) p. 4172.
- [33] Z. Zhang, J. Bu, M. Ester, J. Zhang, C. Yao, Z. Yu, and C. Wang, Hierarchical graph pooling with structure learning, [arXiv:1911.05954](https://arxiv.org/abs/1911.05954).
- [34] Z. Ying, J. You, C. Morris, X. Ren, W. Hamilton, and J. Leskovec, Hierarchical graph representation learning with differentiable pooling, in *Advances in Neural Information Processing Systems*, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Curran Associates, Inc., 2018), Vol. 31.
- [35] Search for Higgs boson decay to a charm quark-antiquark pair via ttH production, Technical Report No. CMS-PAS-HIG-24-018, CERN, Geneva, 2025, <https://cds.cern.ch/record/2929444>.
- [36] ATLAS Collaboration, Measurements of WH and ZH production with higgs boson decays into bottom quarks and direct constraints on the charm Yukawa coupling in 13 TeV pp collisions with the ATLAS detector, *J. High Energy Phys.* **04** (2025) 075.
- [37] ATLAS Collaboration, Combination of searches for Higgs boson pair production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, *Phys. Rev. Lett.* **133**, 101801 (2024).
- [38] <https://github.com/nathaliesoy/tiger>.