# Missing information search with deep learning for mass estimation

Kayoung Ban [1], Dong Woo Kang [2,3] Tae-Geun Kim [1], Seong Chan Park [1], and Yeji Park [1]

[1]*Department of Physics and IPAP, Yonsei University, Seoul 03722, Republic of Korea*
[2]*School of Physics, Korea Institute for Advanced Study, Seoul 02455, Republic of Korea*
[3]*Theoretical Physics Department, CERN, CH-1211 Geneva, Switzerland*

We introduce DeeLeMa, a deep learning-based network for the analysis of energy and momentum in high-energy particle collisions. This novel approach is specifically designed to address the challenge of analyzing collision events with multiple invisible particles, which are prevalent in many high-energy physics experiments. DeeLeMa is constructed based on the kinematic constraints and symmetry of the event topologies. We show that DeeLeMa can robustly estimate mass distribution even in the presence of combinatorial uncertainties and detector smearing effects. The approach is flexible and can be applied to various event topologies by leveraging the relevant kinematic symmetries. This work opens up exciting opportunities for the analysis of high-energy particle collision data, and we believe that DeeLeMa has the potential to become a valuable tool for the high-energy physics community.

I. INTRODUCTION

Despite the numerous neutrinos generated during particle collisions, the detectors at the Large Hadron Collider (LHC) are unable to observe them directly [1,2]. In addition to neutrinos, other elusive particles such as dark matter candidates, including weakly interacting massive particles (WIMP) [3,4], axions [5,6], are also challenging to detect as they pass through the detector without leaving discernible signals [7,8]. Such entities are termed as "invisible particles" in the realm of collider physics. Their existence isn't directly observed but is inferred by leveraging the principles of energy and momentum conservation, which highlight discrepancies in momentum or energy within an event.

The LHC, like other hadronic collider experiments, measures the scattering processes involving the partonic constituents of hadrons. Within this context, the reconstruction of the longitudinal component of missing momentum along the beam axis (referred to as the longitudinal direction) poses a substantial challenge. Furthermore, the formidable nature of this endeavor becomes particularly pronounced when multiple invisible particles are simultaneously generated within the same event. This challenging issue is conventionally called the "missing information problem" of invisible particles.

Researchers commonly employ "transverse" quantities to address the challenge from the longitudinal information. These transverse quantities are defined along directions perpendicular to the beam axis and include the transverse momentum $p_T = \sqrt{\vec{p}_\perp^2}$ and transverse energy $E_T \equiv \sqrt{m^2 - p_T^2}$ as observable parameters. Over the past decade or more, many kinematic variables have been devised and proposed, primarily tailored for the experiments at the LHC, such as the stransverse mass or the Cambridge M_{T2} [9–11], M_2 [12–14], and their extensions [11,15–18]. However, it is worth noting that introducing more complex kinematic variables while aiding in obtaining missing information can also introduce additional complexities in data analysis. The precision of these variables may not always meet the desired level due to inherent complexities and uncertainties, including combinatorial errors and detector effects. For a comprehensive overview, see e.g., Ref. [19].

This paper introduces an innovative approach to address the challenges posed by missing information problems in collider physics [20–26]. Instead of relying on intricate kinematic variables, our proposed method leverages the power of deep neural networks (DNNs), capitalizing on the recent rapid advancements in machine learning techniques [27–36].

DNNs have emerged as a versatile tool capable of handling vast datasets and capturing intricate correlations among diverse features. This capability renders them exceptionally well suited to tackle the complexities associated with missing information. Our newly developed kinematics-solving machine, which integrates the physical conditions and symmetries inherent in event shapes, is named "DeeLeMa ." This acronym, derived from "Deep Learning for Mass Estimation," encapsulates the essence of our machine's function. DeeLeMa represents a cutting-edge approach to the problem of kinematics estimation in collider physics, promising more robust and accurate results compared to traditional methods reliant on complex kinematic variables. The detail of the architecture is presented in the GitHub page [37] where one can download DeeLeMa code with examples.
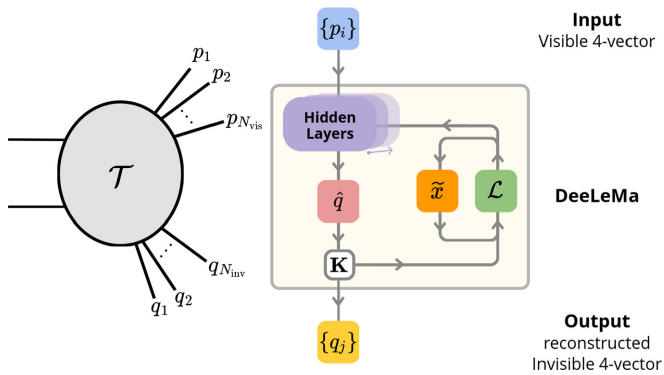
FIG. 1. General event shape with visible ($\{p_i\}_{i=1,2,\ldots,N_{\text{vis}}}$) and invisible ($\{q_j\}_{j=1,2,\ldots,N_{\text{inv}}}$) momenta in the final state.

II. DEELEMA FRAMEWORK

Our study is dedicated to unveiling concealed information within the complex landscape of high-energy collider events. We aim to achieve this objective by harnessing observable data, specifically the four-momenta of detected particles. Event topology, symmetry principles, and the steadfast application of conservation laws furnish constraints on the kinematic variables governing these events.

To illustrate, we examine a cascade event configuration consisting of N_{vis} visible particles and N_{inv} invisible particles in the final state, which can be succinctly represented as

$$pp \to \overbrace{v_1 v_2 \cdots v_{N_{\text{vis}}}}^{N_{\text{vis}}} \overbrace{i_1 i_2 \cdots i_{N_{\text{inv}}}}^{N_{\text{inv}}}.$$

Our primary goal is to utilize the input information encapsulated in the four-momenta of visible particles, denoted as p_i for $i = 1, 2, \ldots, N_{\text{vis}}$, to precisely determine the momenta of each invisible particle in the final state, which we designate as q_j for $j = 1, 2, \ldots, N_{\text{inv}}$. Nonetheless, it is crucial to notice that this kinematic problem becomes mathematically underdetermined when the count of unknown variables N_{inv} surpasses the constraining relationships governing each event's momenta.

Utilizing a physics-informed machine learning approach, we build a model that decodes concealed information in collider events under a given event topology. Central to this approach are two functions: $\mathcal{L}$, our loss function for neural optimization, and the function **K** serves as a mechanism that encapsulates kinematic relationships crucial for reconstructing the momenta of invisible particles. These functions are based on physical relations such as the on-shell mass conditions for the intermediate particles and the constraints on the transverse momentum. The structure of our DNN machine is schematically depicted in Fig. 1:

(i) The event topology of the specific event is $\mathcal{T}$, and the kinematic relations among momenta are encapsulated in **K**.

(ii) The input for DeeLeMa is the visible information from the measured momenta $\{p_i\}$, $i = 1, 2, \ldots, N_{\text{vis}}$.

(iii) The expected output from DeeLeMa is the reconstructed momenta of the invisible particles $\{q_j\}$, $j = 1, 2, \ldots, N_{\text{inv}}$.
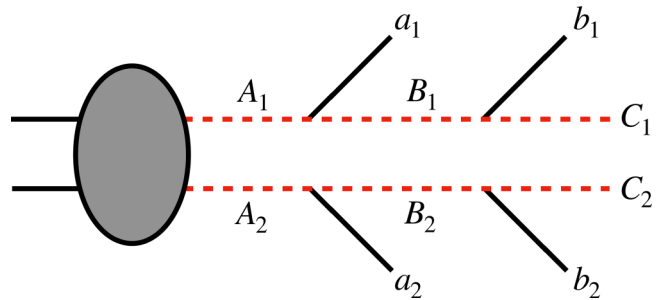

FIG. 2. Symmetric event topology for $N_{\text{vis}} = 4$, $N_{\text{inv}} = 2$.

(iv) The loss function $\mathcal{L}$ enforces the machine to learn to reconstruct the invisible information under the given event topology $\mathcal{T}$ and the kinematic relations **K**.

Additionally, we introduce the auxiliary parameters $\tilde{x}$ which act to force target physical variables x (i.e., invariant mass) to converge into a single value for all training events. The corresponding auxiliary parameters $\tilde{x}$ appear globally in all events, allowing the neural network to learn that the events come from the same physical process. Thus, they are introduced as global, trainable parameters based on prior knowledge from $\mathcal{T}$. Consequently, DeeLeMa works to optimize the reconstruction of invisible momenta by minimizing the loss function $\mathcal{L}$, which is defined in terms of the reconstructed kinematic quantities $\hat{q}$ and the auxiliary parameters $\tilde{x}$, subject to the kinematic relations **K**.

III. DEELEMA FOR PAIR PRODUCTION PROCESS

In this section, our primary focus lies on the pair production of mother particles during particle collisions, where each of these particles subsequently decays, following identical decay chains. Under such circumstances, the scenario involves an even number of both visible and invisible particles, denoted as $(N_{\text{vis}}, N_{\text{inv}}) = (2n, 2m)$. Here, the terms n and m correspond to the visible and invisible particles in each respective branch. Exploiting the inherent symmetry of this situation, we find that there are precisely $8m$ unknown components originating from the $2m$ invisible four-momenta, along with $(n + m + 2)$ constraints stemming from kinematic relations.

Mathematically speaking, the system becomes solvable when the condition $(n + m + 2) \geqslant 8m$ or equivalently $n \geqslant 7m - 2$ is satisfied. A pertinent illustration is the case of $m = 1$, wherein a single invisible particle emerges in each of the decay chain branches. In this scenario, the system can be effectively solved when $n \geqslant 5$. It is noteworthy to mention that earlier analyses on systems involving $n = 3, m = 1$ have been documented in previous works (see Refs. [23–25]), particularly when multiple events of the identical process were considered.

We now delve into a challenging "unsolvable" problem characterized by the parameters $n = 2$ and $m = 1$, visually represented in Fig. 2. This specific configuration corresponds to an event topology of $(N_{\text{vis}}, N_{\text{inv}}) = (4, 2)$. A prominent example of this event topology is found in the dilepton process of $t\bar{t}$ events, where both top quarks undergo leptonic decay, leading to $t \to bW \to b(\ell \nu_\ell)$. In a more general context, we contemplate the pair production of mother particles, denoted

as A_1 and A_2, with subscripts 1 and 2 signifying the respective branches of decay. Each A_i subsequently decays into a visible particle a_i and an intermediate heavy state B_i. Ultimately, B_i undergoes a semi-invisible decay into a visible particle b_i and an invisible particle C_i in branches $i = 1$ and $i = 2$. The event can be succinctly expressed as

$$pp \rightarrow A_1 A_2 \rightarrow (a_1(p_{a_1})B_1)(a_2(p_{a_2})B_2),$$
$$B_1 \rightarrow b_1(p_{b_1})C_1(q_1), \ \ B_2 \rightarrow b_2(p_{b_2})C_2(q_2).$$

Here, p_{a_i} and p_{b_i} symbolize the momenta of visible particles, while q_i represents the momentum of the corresponding invisible particles C_i. Despite the apparent simplicity of this event topology, it is fundamentally underdetermined from a kinematic perspective, rendering the separate measurement of each invisible particle's momentum unattainable.

To define the loss function, we first select a set of "target variables" $\{x\}$, such as the invariant masses of the intermediate states and invisible out-coming particles. For our specific example

$$\mathcal{B}\begin{cases} x^{\#1} = (m_{A_1}, m_{B_1})^{\#1} \oplus (m_{A_2}, m_{B_2})^{\#1} \\ \qquad\qquad\qquad \vdots \\ x^{\#N} = (m_{A_1}, m_{B_1})^{\#N} \oplus (m_{A_2}, m_{B_2})^{\#N}. \end{cases}$$

Consider a batch of dataset consisting of N training events. The event-wise information, denoted as $\mathcal{E}$, is derived from the symmetric event topology. This implies that identical particle masses are consistent, making $\mathcal{E}$ an independent piece of information for each event.

On the other hand, the batch-wise information, represented by $\mathcal{B}$, signifies that all training events are associated with the same physical event. We introduce auxiliary parameters, like $\tilde{m}_A$ and $\tilde{m}_B$, to ensure that the masses across all events in a batch remain consistent (e.g., m_{A_1} of all events are the same, and so on). This $\mathcal{B}$ information is dependent on the entire batch of events. Finally, our loss function is defined as

$$\mathcal{L}_{tot} \equiv \frac{1}{|\mathcal{B}|} \sum_{i=1}^{|\mathcal{B}|} \left[\sum_{f \in \{A,B\}} \mathcal{L}_f^{\#i} \right],$$

$$\mathcal{L}_f^{\#i} \equiv d_{\mathcal{E}}\left(m_{f_1}^{\#i}, m_{f_2}^{\#i}\right) + \left[d_{\mathcal{B}}\left(\tilde{m}_f, m_{f_1}^{\#i}\right) + d_{\mathcal{B}}\left(\tilde{m}_f, m_{f_2}^{\#i}\right)\right], \quad (1)$$

where $|\mathcal{B}|$ represents the batch size, indicating the number of events in a batch $\mathcal{B}$, $\#i$ denotes the event index, and f is the target variable, either A or B. The functions $d_{\mathcal{E}}(x_1, x_2)$ and $d_{\mathcal{B}}(x_1, x_2)$ are distance functions for event-wise and batch-wise information, respectively. They satisfy mathematical conditions: (d1) $d(x_1, x_2) > 0$ if $x_1 \neq x_2$, $d(x_1, x_1) = 0$, (d2) $d(x_1, x_2) = d(x_2, x_1)$, (d3) $d(x_1, x_2) \leqslant d(x_1, y) + d(y, x_2)$ for any y in the sample. Various distance functions can be used, such as $d(x_1, x_2) = |x_1 - x_2|$, $|x_1 - x_2|^2$, or $|x_1^2 - x_2^2|$. The appropriate choice depends on the specific physical process under study.

We illustrate the training procedure of DeeLeMa in Fig. 3. The target variable points $x_1^{\#i}$, $x_2^{\#i}$ are represented within spaces X_1 and X_2, accompanied by the scalar value of the auxiliary parameter $\tilde{x}$. By minimizing the loss function in Eq. (1) from
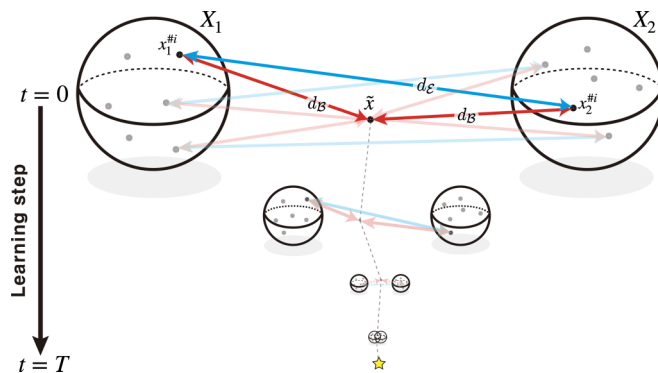

FIG. 3. Schematic representation of the role of the loss function in simultaneously bringing $d_{\mathcal{E}}$ (blue double-headed arrow) and $d_{\mathcal{B}}$ (red double-headed arrow) closer throughout the learning process $t = 0$ to $t = T$.

the initial learning step at $t = 0$ to the end of training at $t = T$, we ensure that spaces X_1 and X_2 come closer together. Additionally, the overall distribution of points within these spaces becomes more compact, leading to a reduction in their spread or dispersion. This compactness and reduction in dispersion are facilitated by the inclusion of the auxiliary parameter $\tilde{x}$. For a comprehensive model implementation of DeeLeMa, refer to Appendix.

IV. TEST OF DEELEMA PERFORMANCE

In pair production, practical experiments often encounter issues with the misidentification of branches. Termed the combinatorics problem, this complication can result in erroneous kinematic relations, leading to substantial uncertainties in the derived solutions. To quantify the extent of this contamination, we introduce the parameter $\mathcal{E}_C$, defined as the fraction of incorrectly assigned events relative to the overall number of events, expressed as

$$\mathcal{E}_C \equiv \frac{\text{wrong}}{\text{wrong} + \text{correct}}. \quad (2)$$

We assess the efficacy of DeeLeMa through three distinctive test runs:

(i) Test run (A) is conducted using a toy model featuring fixed values of $m_A = 1000\,\text{GeV}$, $m_B = 800\,\text{GeV}$, and $m_C = 700\,\text{GeV}$, with no combinatorial errors ($\mathcal{E}_C = 0$).

(ii) Test run (B) mirrors (A) but incorporates varying rates of combinatorial errors, specifically $\mathcal{E}_C = 0$, 10%, 20%, 50%. This test aims to investigate the influence of combinatorial errors on the performance of DeeLeMa.

(iii) Test run (C) is executed on the standard model $t\bar{t}$ and $t \rightarrow Wb \rightarrow (\ell\nu)b$ processes, encompassing $\mathcal{E}_C = 20\%$ and accounting for detector smearing effects. We consider this test run to closely simulate a realistic scenario.

We compare the results with those obtained using other existing methods: the transverse mass variable M_{T2} and the on-shell constrained invariant mass variables M_{2CC}, which use similar constraints as DeeLeMa. We use the YAM2 package [38] to calculate M_2 optimally.
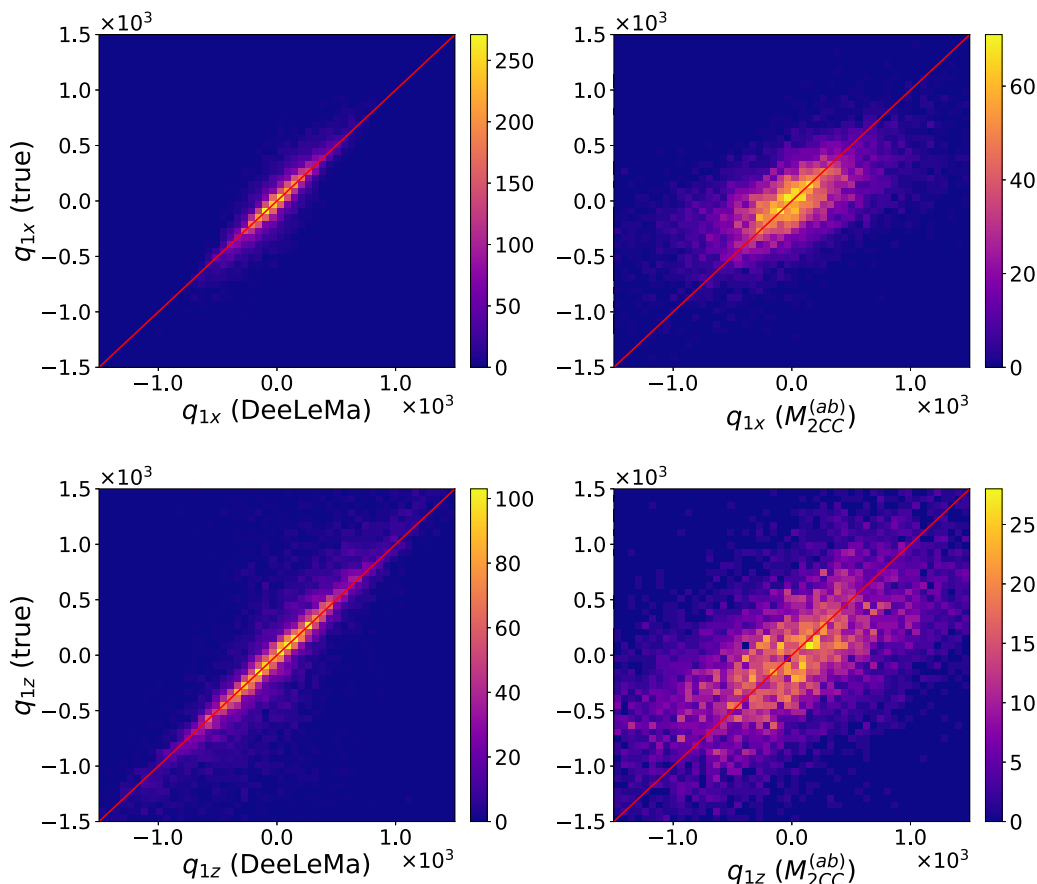
FIG. 4. [Toy] The correlation heatmap of the reconstructed momenta and the true momenta from the DeeLeMa (left) and $M_{2CC}^{(ab)}$ (right) for the toy example with $\mathcal{E}_C = 0$.

A. Toy model

1. Toy model test with no contamination ($\mathcal{E}_C = 0$)

We selected narrow width values for m_A, m_B, and m_C at 1000 GeV, 800 GeV, and 700 GeV, respectively. The correlation heatmap in Fig. 4 displays the relationship between the reconstructed momenta (horizontal axis) and the true momenta (vertical axis) for the DeeLeMa method (left) and the $M_{2CC}^{(ab)}$ method (right) applied to the toy example with $\mathcal{E}_C = 0$. Ideally, the diagonal line (red, solid line) should represent perfect efficiency with $p_{\text{recon.}} = p_{\text{true}}$. As shown in the figure, the DeeLeMa method (left) exhibits a strong diagonal correlation pattern, indicating high accuracy in reconstructing the momenta. In contrast, the $M_{2CC}^{(ab)}$ method (right) shows a weaker and more scattered correlation pattern, implying a lower accuracy in momentum reconstruction. This demonstrates the superior performance of DeeLeMa over traditional methods.

The upper panel of Fig. 5 shows the reconstructed mass distributions of B and A obtained with DeeLeMa for the toy example with $\mathcal{E}_C = 0$. The blue dashed lines indicate the reconstructed masses of $m_{B_{1,2}}$ and $m_{A_{1,2}}$, respectively. The red vertical lines indicate the true masses, and the black dot-dashed line shows the auxiliary mass after training. In the bottom panel, we compare the results with two existing methods based on M_{T2} (gray) and M_{2CC} with suitable subsystems (b) and (ab) (orange and green), respectively [15].

We can see that the reconstructed mass distributions with DeeLeMa are well centered around the true values, while the M_{T2} distribution shows the physical mass at the endpoint of
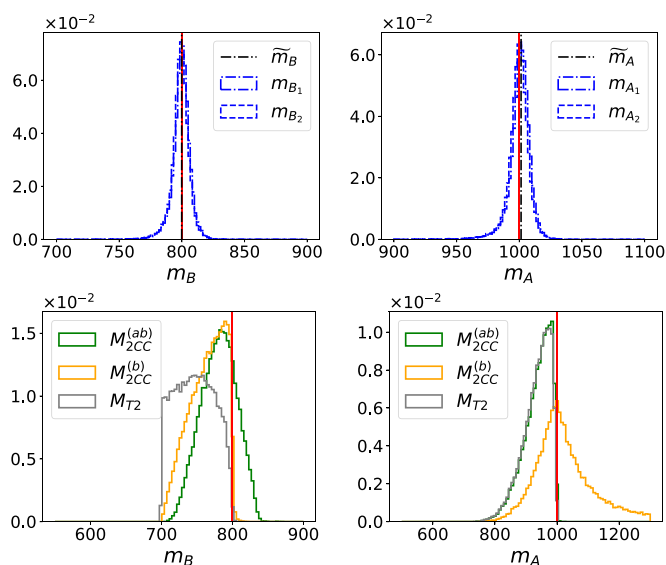

FIG. 5. [Toy] The reconstructed mass distributions of B and A using DeeLeMa (upper), and M_{T2}, $M_{2CC}^{(b)}$, and $M_{2CC}^{(ab)}$ (bottom) for the toy example with $\mathcal{E}_C = 0$.

TABLE I. The summary table for combinatorial efficiency $\mathcal{E}_C$.

$\mathcal{E}_C$	$\widetilde{m}_X$ [GeV]		$m_X \pm \sigma$ [GeV]			
[%]	$\widetilde{m}_A$	$\widetilde{m}_B$	m_{A_1}	m_{A_2}	m_{B_1}	m_{B_2}
0	1001.34	799.95	1000.45 ± 13.31	999.93 ± 13.59	799.59 ± 8.95	799.42 ± 9.05
10	1001.46	800.47	1007.41 ± 32.21	1007.18 ± 31.90	802.26 ± 16.79	802.11 ± 16.55
20	1005.16	802.25	1013.56 ± 43.24	1013.14 ± 42.24	804.59 ± 21.82	804.41 ± 21.75
50	1010.73	807.61	1028.94 ± 62.56	1029.27 ± 61.39	810.87 ± 31.80	810.97 ± 31.59

the distribution, which often causes errors. The $M_{2CC}^{(b)}$ for m_B and the $M_{2CC}^{(ab)}$ for m_A show slightly improved performances, but still **DeeLeMa** provides the best results.

The disparity arises from the manner in which global information is assimilated during the machine learning training phase, primarily facilitated through the auxiliary parameter $\tilde{x}$. Conversely, in the context of the M_{T2} or M_2 method, global information is solely derived from statistical outcomes, primarily centered around kinematic endpoints. While numerous events are typically clustered around these endpoints, leading to a reconstruction of momenta close to the actual values, there is a lack of subsequent optimization within the M_{T2}- or M_2-based reconstruction process.

Consequently, the precision is notably diminished, with the kinematic endpoints becoming less distinct, particularly when grappling with combinatorial ambiguities and accounting for the effects of detector smearing. Subsequently, this degradation in accuracy will be demonstrated in the subsequent sections.

2. Toy model test with contamination ($\mathcal{E}_C > 0$)

To explicitly see the effect of combinatorics contamination, we conducted comprehensive test runs incorporating the possibility of combinatorial errors, with a concise summary of **DeeLeMa**'s performance presented in Table I. In these instances, the peak positions have displayed a slight shift towards larger values, owing to the influence of inaccurately assigned data implying a relatively higher mass. Despite accommodating up to 20% in combinatorial errors, **DeeLeMa** exhibits sustained resilience and commendable performance, accurately reconstructing masses within the $5 - 10\%$ range of the true values.

Notably, for cases where $\mathcal{E}_C \leqslant 20\%$, the reconstructed masses are within the vicinity of $\mathcal{O}(1)\%$ of the true values, attesting to **DeeLeMa**'s remarkable ability to mitigate the impacts of combinatorial challenges effectively. Collectively, our findings underscore **DeeLeMa**'s reliability and robustness as a method proficient in the precise reconstruction of masses, even in the face of demanding conditions prevalent in collider environments.

B. Realistic test with standard model $t\bar{t}$

We finally present the results of our investigation on a more realistic case, the top quark pair production at the LHC, where top quarks decay semileptonically as $t\bar{t} \rightarrow (b\,W^+)\,(\bar{b}\,W^-) \rightarrow (b\,\ell^+\,\nu)\,(\bar{b}\,\ell^-\,\bar{\nu})$. In this case, we consider finite width effects with $\sigma_t = 1.4915$ GeV, $\sigma_W = 2.0476$ GeV, and $m_t = 173.0$ GeV, $m_W = 80.4190$ GeV for the top quark and W boson,

respectively. Moreover, we also account for the uncertainties related to the detector resolution. We simulated detector effects by applying Gaussian smearing to the momenta. However, to achieve more accurate results, we encourage the use of a more realistic detector simulation. For the two b jets, we applied Gaussian smearing with jet p_T values of $\{10, 20, 30, 50, 100, 400, 1000\}$ GeV and energy resolutions of $\{40, 28, 19, 13, 10, 6, 5\}\%$, respectively [26,39]. We took the combinatorial ambiguity at $\mathcal{E}_C = 20\%$ for our simulation.

We present the results obtained using **DeeLeMa** in Fig. 6 (upper). The distributions for the reconstructed masses (m_t, m_W) show robust peaks near the true values (red vertical line), albeit slightly widened. To compare the performance of **DeeLeMa** with conventional methods, we also present the results obtained using $M_{2CC}^{(ab)}$ and M_{T2} variables (lower). **DeeLeMa** provides more accurate results compared to conventional methods. In conventional methods, we need to read the endpoints in the lower distributions, which can be challenging in practice due to realistic effects from finite widths, detector smearing, and combinatorial mismatches.

V. CONCLUSION

We introduce **DeeLeMa**, a deep learning-based approach to analyze high-energy particle collisions with multiple invisible particles. **DeeLeMa** can reconstruct the event's invisible
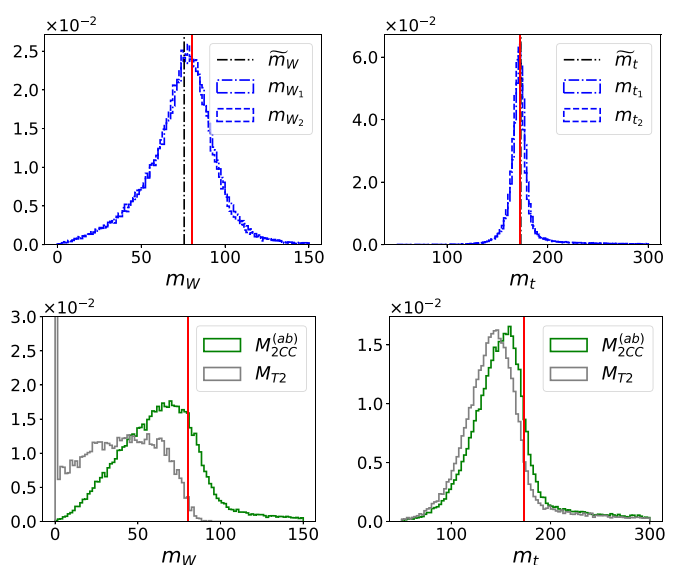
FIG. 6. [Realistic $t\bar{t}$] The reconstructed mass distributions of B and A using **DeeLeMa** (upper), and M_{T2}, $M_{2CC}^{(b)}$ and $M_{2CC}^{(ab)}$ (bottom) for the $t\bar{t}$ example.

TABLE II. Hyperparameters used for the result plots in IV. N_{node} denotes the number of nodes in each hidden layer, N_{layer} represents the number of hidden layers, η refers to the learning rate, $|\mathcal{B}|$ is the batch size, and epoch signifies the number of epochs.

Model	N_{node}	N_{layer}	η	$|\mathcal{B}|$	epoch	m_C	Δm_B^{init}	Δm_A^{init}	d
Toy	256	5	10^{-2}	2048	100	700	0.3	0.3	L_1
$t\bar{t}$	256	5	5×10^{-4}	2048	100	0	0.3	0.3	L_1

momenta and masses, even when multiple invisible particles are involved. Focusing on a challenging problem with $(N_{\text{vis}}, N_{\text{inv}}) = (4, 2)$, we demonstrate the efficiency of DeeLeMa : compared to conventional methods that rely on kinematic variables such as M_{T2} or M_2, DeeLeMa delivers a significant improvement in accuracy. The reconstructed masses show sharp peaks in the distribution, and the results are robust against the combinatorial problem of misidentification of final state particles and detector-smearing effects. In conclusion, DeeLeMa has the potential to contribute to advances in the field as a new solid tool.

APPENDIX: THE DETAIL OF MODEL

The DeeLeMa is constructed using the PyTorch package [40] and the Lightning library [41] as the front end, with the Adam optimizer [42] for training. The model is trained on GPUs with a specified batch size and number of epochs as summarized in Table II. Additionally, we employ the GELU (Gaussian Error Linear Unit) activation function [43] with a tanh approximation and apply batch normalization. The detailed architecture and hyperparameters are available on the associated GitHub page [37].

[1] G. Aad *et al.* (ATLAS Collaboration), Inclusive and differential cross-sections for dilepton $t\bar{t}$ production measured in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector, J. High Energy Phys. 07 (2023) 141.

[2] A. Tumasyan *et al.* (CMS Collaboration), First measurement of the top quark pair production cross section in proton-proton collisions at $\sqrt{s} = 13.6$ TeV, J. High Energy Phys. 08 (2023) 204.

[3] G. Jungman, M. Kamionkowski, and K. Griest, Supersymmetric dark matter, Phys. Rep. **267**, 195 (1996).

[4] G. Bertone, D. Hooper, and J. Silk, Particle dark matter: Evidence, candidates and constraints, Phys. Rep. **405**, 279 (2005).

[5] J. E. Kim, Weak interaction singlet and strong CP invariance, Phys. Rev. Lett. **43**, 103 (1979).

[6] M. Dine, W. Fischler, and M. Srednicki, Supersymmetric technicolor, Nucl. Phys. B **189**, 575 (1981).

[7] A. Tumasyan *et al.* (CMS Collaboration), Search for long-lived particles decaying in the CMS end cap muon detectors in proton-proton collisions at $\sqrt{s} = 13$ TeV, Phys. Rev. Lett. **127**, 261804 (2021).

[8] J. Gonski (ATLAS Collaboration), Highlights from long-lived particle searches at ATLAS, in 56th Rencontres de Moriond on QCD and high energy interactions, arXiv:2205.07669 [hep-ex].

[9] A. Barr, C. Lester, and P. Stephens, A variable for measuring masses at hadron colliders when missing energy is expected; m_{T2}: The truth behind the glamour, J. Phys. G: Nucl. Part. Phys. **29**, 2343 (2003).

[10] C. G. Lester and D. J. Summers, Measuring masses of semi-invisibly decaying particles pair produced at hadron colliders, Phys. Lett. B **463**, 99 (1999).

[11] W. S. Cho, K. Choi, Y. G. Kim, and C. B. Park, Measuring the top quark mass with m_{T2} at the LHC, Phys. Rev. D **78**, 034019 (2008).

[12] A. J. Barr, T. J. Khoo, P. Konar, K. Kong, C. G. Lester, K. T. Matchev, and M. Park, Guide to transverse projections and mass-constraining variables, Phys. Rev. D **84**, 095031 (2011).

[13] R. Mahbubani, K. T. Matchev, and M. Park, Re-interpreting the Oxbridge stransverse mass variable m_{T2} in general cases, J. High Energy Phys. 03 (2013) 134.

[14] W. S. Cho, J. S. Gainer, D. Kim, K. T. Matchev, F. Moortgat, L. Pape, and M. Park, On-shell constrained M_2 variables with applications to mass measurements and topology disambiguation, J. High Energy Phys. 08 (2014) 070.

[15] M. Burns, K. Kong, K. T. Matchev, and M. Park, Using subsystem M_{T2} for complete mass determinations in decay chains with missing energy at hadron colliders, J. High Energy Phys. 03 (2009) 143.

[16] A. J. Barr, B. Gripaios, and C. G. Lester, Transverse masses and kinematic constraints: from the boundary to the crease, J. High Energy Phys. 11 (2009) 096.

[17] P. Konar, K. Kong, K. T. Matchev, and M. Park, Dark matter particle spectroscopy at the LHC: Generalizing m_{T2} to asymmetric event topologies, J. High Energy Phys. 04 (2010) 086.

[18] P. Konar, K. Kong, K. T. Matchev, and M. Park, Superpartner mass measurement technique using 1d orthogonal decompositions of the cambridge transverse mass variable M_{T2}, Phys. Rev. Lett. **105**, 051802 (2010).

[19] R. Franceschini, D. Kim, K. Kong, K. T. Matchev, M. Park, and P. Shyamsundar, Kinematic variables and feature engineering for particle phenomenology, arXiv:2206.13431 [hep-ph].

[20] M. M. Nojiri, G. Polesello, and D. R. Tovey, Proposal for a new reconstruction technique for SUSY processes at the LHC, in *3rd Les Houches Workshop on Physics at TeV Colliders*, arXiv:hep-ph/0312317.

[21] K. Kawagoe, M. M. Nojiri, and G. Polesello, A new SUSY mass reconstruction method at the CERN LHC, Phys. Rev. D **71**, 035008 (2005).

[22] H.-C. Cheng, J. F. Gunion, Z. Han, G. Marandella, and B. McElrath, Mass determination in SUSY-like events with missing energy, J. High Energy Phys. **12** (2007) 076.

[23] H.-C. Cheng, D. Engelhardt, J. F. Gunion, Z. Han, and B. McElrath, Accurate mass determinations in decay chains with missing energy, Phys. Rev. Lett. **100**, 252001 (2008).

[24] H.-C. Cheng, J. F. Gunion, Z. Han, and B. McElrath, Accurate Mass determinations in decay chains with missing energy. II, Phys. Rev. D **80**, 035020 (2009).

[25] B. Webber, Mass determination in sequential particle decay chains, J. High Energy Phys. 09 (2009) 124.

[26] D. Kim, K. T. Matchev, and P. Shyamsundar, Kinematic focus point method for particle mass measurements in missing energy events, J. High Energy Phys. 10 (2019) 154,.

[27] M. Feickert and B. Nachman, A living review of machine learning for particle physics, arXiv:2102.02770 [hep-ph].

[28] A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel, A. Aurisano, K. Terao, and T. Wongjirad, Machine learning at the energy and intensity frontiers of particle physics, Nature (London) **560**, 41 (2018).

[29] P. Shanahan *et al.*, Snowmass 2021 computational frontier CompF03 topical group report: Machine learning, arXiv:2209.07559 [physics.comp-ph].

[30] G. Karagiorgi, G. Kasieczka, S. Kravitz, B. Nachman, and D. Shih, Machine learning in the search for new fundamental physics, Nat. Rev. Phys. **4**, 399 (2022).

[31] Z. Dong, K. Kong, K. T. Matchev, and K. Matcheva, Is the machine smarter than the theorist: Deriving formulas for particle kinematics with symbolic regression, Phys. Rev. D **107**, 055018 (2023).

[32] D. Kim, K. Kong, K. T. Matchev, M. Park, and P. Shyamsundar, Deep-learned event variables for collider phenomenology, Phys. Rev. D **107**, L031904 (2023).

[33] M. U. Haq, C. Kilic, B. Lawrence-Sanderson, and R. P. R. Sudha, Applying machine learning techniques to intermediate-length cascade decays, Phys. Rev. D **108**, 035002 (2023).

[34] A. Alves and C. H. Yamaguchi, Reconstruction of missing resonances combining nearest neighbors regressors and neural network classifiers, Eur. Phys. J. C **82**, 746 (2022).

[35] S. H. Lim and M. M. Nojiri, Morphology for jet classification, Phys. Rev. D **105**, 014004 (2022).

[36] A. Chakraborty, S. H. Lim, M. M. Nojiri, and M. Takeuchi, Neural network-based top tagger with two-point energy correlations and geometry of soft emissions, J. High Energy Phys. 07 (2020) 111.

[37] https://github.com/Yonsei-HEP-COSMO/DeeLeMa.

[38] C. B. Park, YAM2: Yet another library for the M_2 variables using sequential quadratic programming, Comput. Phys. Commun. **264**, 107967 (2021).

[39] V. Khachatryan *et al.* (CMS Collaboration), Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV, JINST **12**, P02014 (2017).

[40] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai *et al.*, Pytorch: An imperative style, high-performance deep learning library, in *Advances in Neural Information Processing Systems 32* (Curran Associates, Inc., 2019), pp. 8024–8035.

[41] W. Falcon, J. Borovec, A. Wälchli, N. Eggert, J. Schock, J. Jordan, N. Skafte, Ir1dXD, V. Bereznyuk, E. Harris, T. Murrell, P. Yu, S. Præsius, T. Addair, J. Zhong, D. Lipin, S. Uchida, S. Bapat, H. Schröter, B. Dayma *et al.*, PyTorchLightning/pytorch-lightning: 0.7.6 release (2019), doi: 10.5281/zenodo.3828935.

[42] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980 [cs.LG].

[43] D. Hendrycks and K. Gimpel, Gaussian error linear units (gelus), arXiv:1606.08415 [cs.LG].