DeeLeMa: Missing information search with **Dee**p **Le**arning for **Ma**ss estimation

Kayoung Ban (Yonsei University)

ArXiv: 2212.12836 [hep-ph]

In collaboration with

Dong Woo Kang (KIAS), Tae Geun Kim, Yeji Park, and Seong Chan Park (Yonsei)

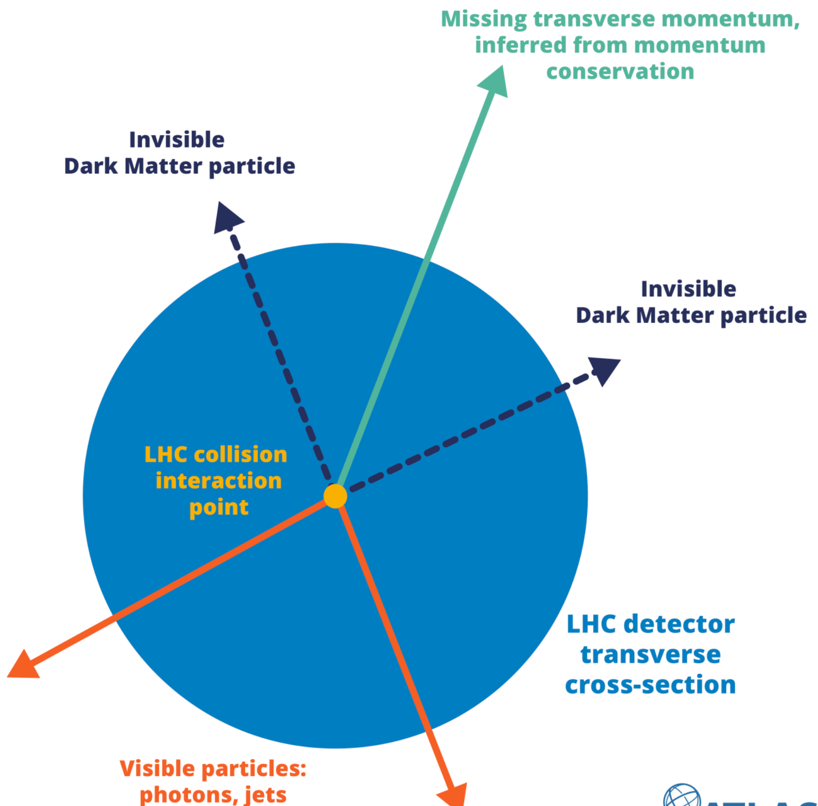
AI and Quantum Information Applications in Fundamental Physics
2023.02.12-18 Konjiam resort

Introduction

- Events with invisible particles are challenging to handle at the Large Hadron Collider (LHC), even though invisible particles often appear in particle physics models.

- **Neutrinos** and **dark matter** particles are excellent examples and particle detectors cannot track those invisible particles directly.

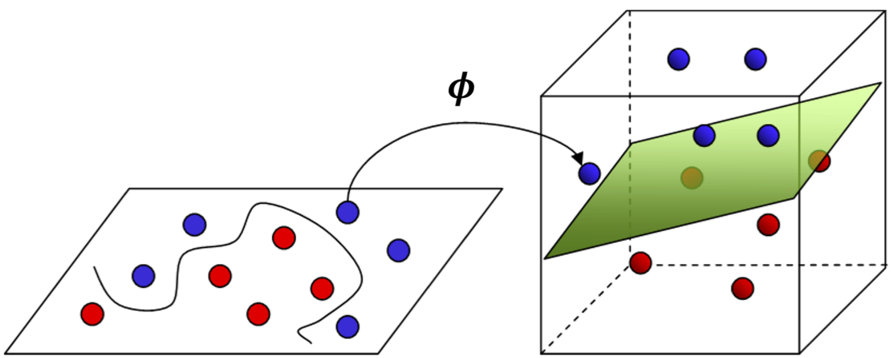
- Their existence appears as the missing momentum or the missing energy by momentum and energy conservation, especially along the transverse directions.



ATLAS EXPERIMENT

Introduction

- In particle physics, machine learning is a topic **to solve complex problems** that have been actively developed for the past few years and until now.
- Deep Neural Network (DNN) can utilize all the information contained in the data.
- In these senses, we use DNN as it exhibits strength in processing extensive data and convoluted correlations among features, critical characteristics of missing information problems.

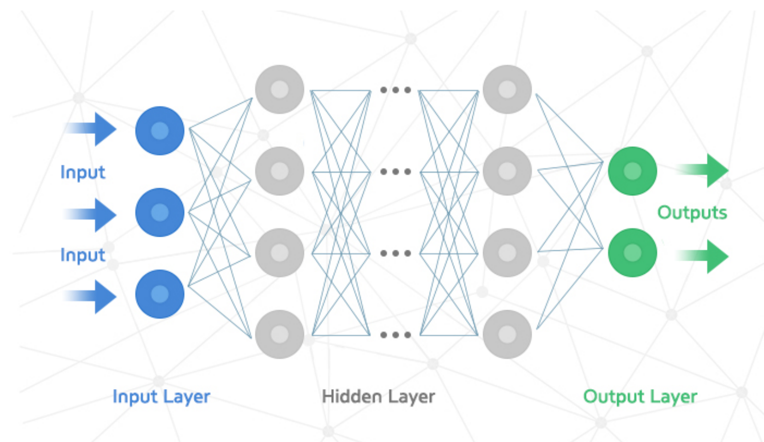


input space

(Visible space)

feature space

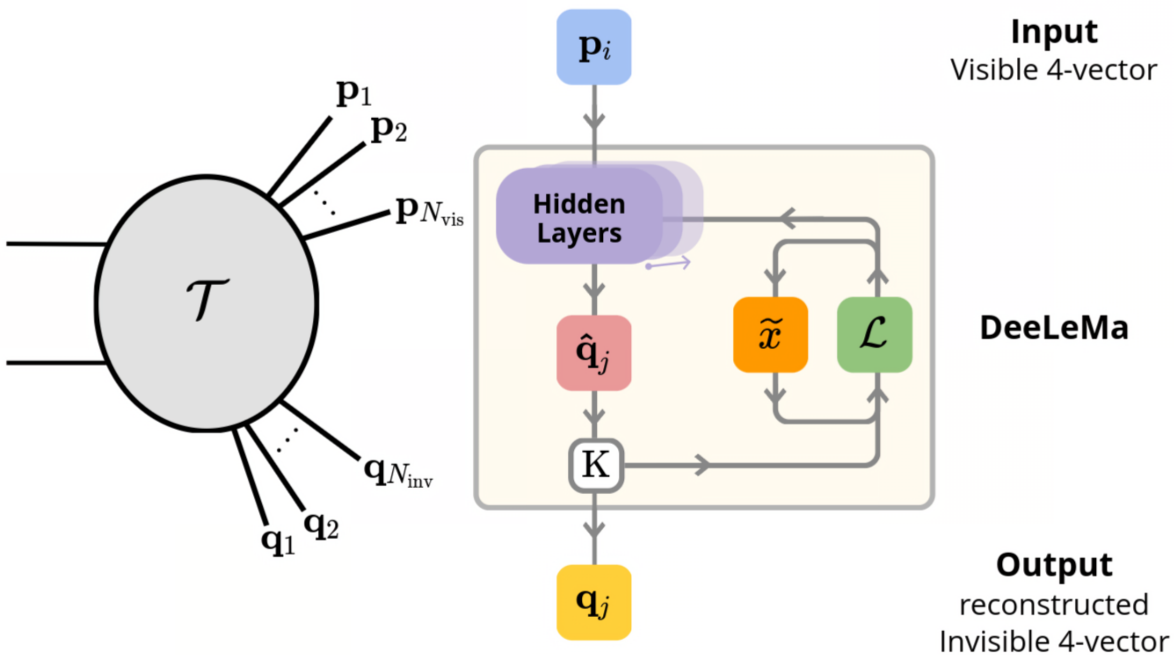
(Full space)



- Our goal is to **design a kinematics-solving machine that reconstructs the invisible information of particles** based on kinematic properties differently from kinematic methods relying on MT2 or its variants.

DeeLeMa (Deep Learning for Mass estimation)

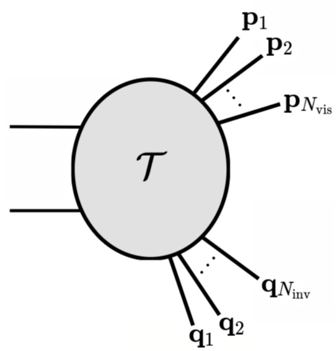
- **DeeLeMa** is a DNN-based machine to reconstruct kinematic unknowns q_j, based on knowns p_i, symmetries, and conservation laws for a given event topology ($\mathscr{T}$).



✓ We construct $\mathscr{L}$ as a loss function and **K as kinematics based on physical laws**, i.e., the on-shell condition on topology and the missing transverse momentum constraint.

✓ Additionally, we introduce an auxiliary parameter $\tilde{x}$, which is designed as global trainable parameters based on prior knowledge **that all events are of the same physics**, leading the values reconstructed from the output of DNN (denoted without tilde as x) to converge into a single value.

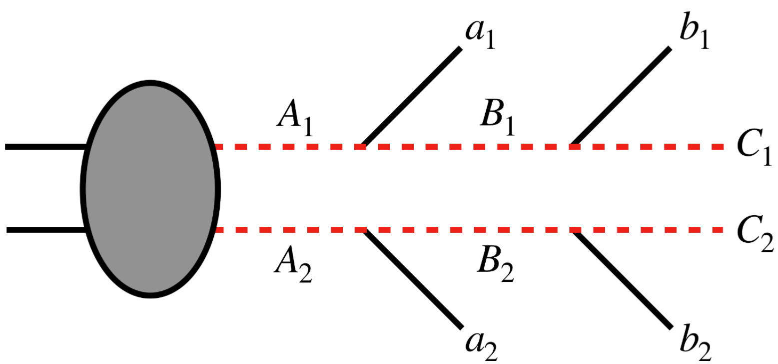
DeeLeMa (Deep Learning for Mass estimation)



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

- Mathematically, this kinematic problem is not solvable if the number of unknowns, (i.e. N(inv)), is larger than the constraining relations among momenta in each event.

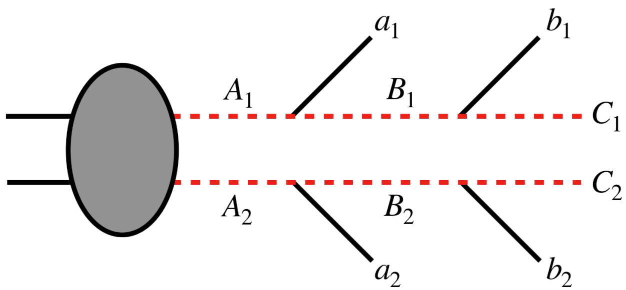
- For definiteness, we consider the event topology where N(vis) = 4 and N(inv) = 2 in the final state.



$$\vec{q}_{1T} + \vec{q}_{2T} = -\not{\vec{p}}_T$$
$$m_{A_1} = m_{A_2} \Rightarrow (p_{a_1} + p_{b_1} + q_1)^2 = (p_{a_2} + p_{b_2} + q_2)^2,$$
$$m_{B_1} = m_{B_2} \Rightarrow (p_{b_1} + q_1)^2 = (p_{b_2} + q_2)^2,$$
$$m_{C_1} = m_{C_2} \Rightarrow q_1^2 = q_2^2.$$

- The number of unknowns is 8 (two four-momenta of invisible particles) and the number of available constraining relations is (2+2+2): 2 from mother particles' masses, 2 from invisible masses if the mass is given, and 2 from transverse momenta.

DeeLeMa – Loss Function



$$\vec{q}_{1T} + \vec{q}_{2T} = -\not{\vec{p}}_T$$

$$m_{A_1} = m_{A_2} \Rightarrow (p_{a_1} + p_{b_1} + q_1)^2 = (p_{a_2} + p_{b_2} + q_2)^2,$$

$$m_{B_1} = m_{B_2} \Rightarrow (p_{b_1} + q_1)^2 = (p_{b_2} + q_2)^2,$$

$$m_{C_1} = m_{C_2} \Rightarrow q_1^2 = q_2^2.$$

$$\mathcal{L} \equiv \sum_{X=A,B} d(m_{X_1}, m_{X_2}) + \sum_{X=A,B} \left[d(\widetilde{m}_X, m_{X_1}) + d(\widetilde{m}_X, m_{X_2}) \right].$$

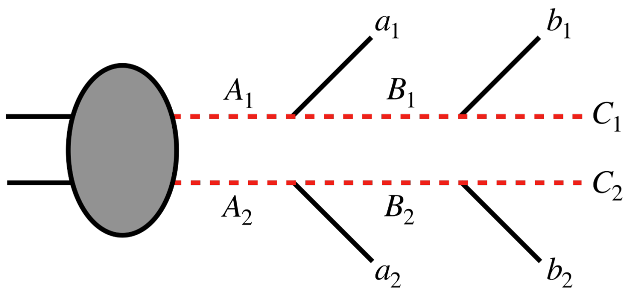
- Here $d(x_1, x_2)$ denotes the general distance function calculated with respect to two parameters x_1 and x_2.

- **The first line** reflects the symmetrical topology.

$$m_{A_1} = m_{A_2} \Rightarrow (p_{a_1} + p_{b_1} + q_1)^2 = (p_{a_2} + p_{b_2} + q_2)^2,$$

$$m_{B_1} = m_{B_2} \Rightarrow (p_{b_1} + q_1)^2 = (p_{b_2} + q_2)^2,$$

DeeLeMa – Loss Function



$$\vec{q}_{1T} + \vec{q}_{2T} = -\not{p}_T$$

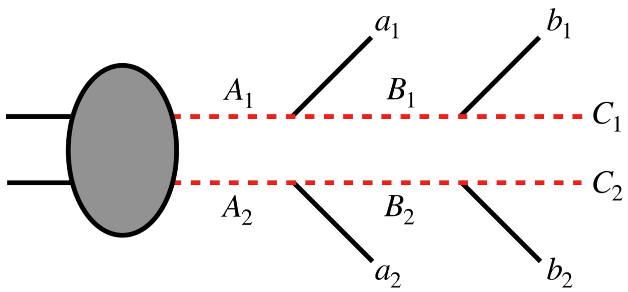
$$m_{A_1} = m_{A_2} \Rightarrow (p_{a_1} + p_{b_1} + q_1)^2 = (p_{a_2} + p_{b_2} + q_2)^2,$$
$$m_{B_1} = m_{B_2} \Rightarrow (p_{b_1} + q_1)^2 = (p_{b_2} + q_2)^2,$$
$$m_{C_1} = m_{C_2} \Rightarrow q_1^2 = q_2^2.$$

$$\mathcal{L} \equiv \boxed{\sum_{X=A,B} d(m_{X_1}, m_{X_2})} + \sum_{X=A,B} \left[d(\widetilde{m}_X, m_{X_1}) + d(\widetilde{m}_X, m_{X_2}) \right].$$

- Here $d(x_1, x_2)$ denotes the <u>general distance function</u> calculated with respect to two parameters x_1 and x_2.
- **The first line** reflects the <u>symmetrical topology.</u>

$$m_{A_1} = m_{A_2} \Rightarrow (p_{a_1} + p_{b_1} + q_1)^2 = (p_{a_2} + p_{b_2} + q_2)^2,$$
$$m_{B_1} = m_{B_2} \Rightarrow (p_{b_1} + q_1)^2 = (p_{b_2} + q_2)^2,$$

DeeLeMa – Loss Function



$$\vec{q}_{1T} + \vec{q}_{2T} = -\not{p}_T$$

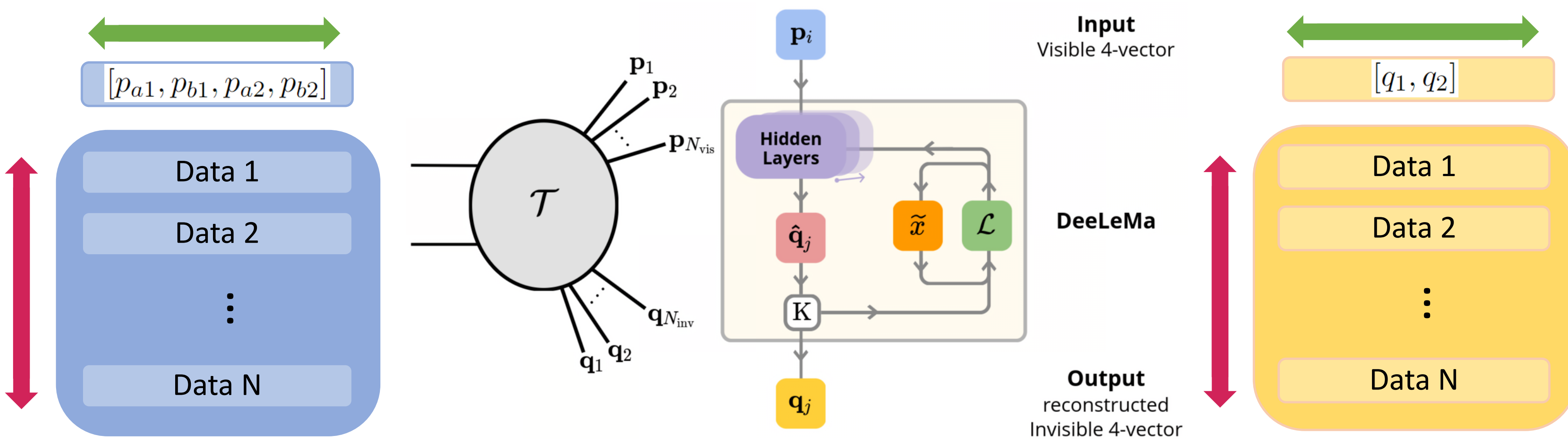
$$m_{A_1} = m_{A_2} \Rightarrow (p_{a_1} + p_{b_1} + q_1)^2 = (p_{a_2} + p_{b_2} + q_2)^2,$$
$$m_{B_1} = m_{B_2} \Rightarrow (p_{b_1} + q_1)^2 = (p_{b_2} + q_2)^2,$$
$$m_{C_1} = m_{C_2} \Rightarrow q_1^2 = q_2^2.$$

$$\mathcal{L} \equiv \sum_{X=A,B} d(m_{X_1}, m_{X_2}) + \boxed{\sum_{X=A,B} \left[d(\widetilde{m}_X, m_{X_1}) + d(\widetilde{m}_X, m_{X_2})\right]}.$$

- Here $d(x_1, x_2)$ denotes the general distance function calculated with respect to two parameters x_1 and x_2.
- **The second line** is to force each event to be generated from the same physics as we mentioned earlier.
- In this case, we use the auxiliary masses $\tilde{m}_X$ of the particle X, making the masses of all events the same.

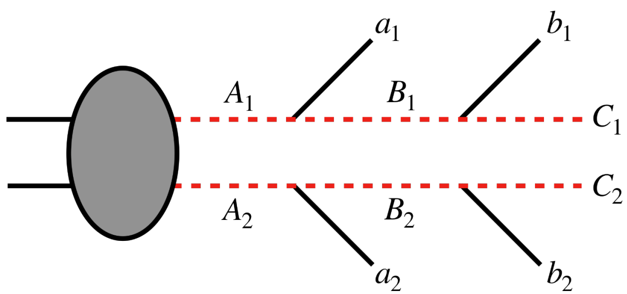
DeeLeMa – Loss Function

$$\mathcal{L} \equiv \sum_{X=A,B} d(m_{X_1}, m_{X_2}) + \sum_{X=A,B} \left[d(\tilde{m}_X, m_{X_1}) + d(\tilde{m}_X, m_{X_2}) \right].$$



- Based on **prior knowledge that all events are of the same physics**, then the reconstructed masses are imposed constraints to converge into a single value.

DeeLeMa – Loss Function and K(inematics)



$$\vec{q}_{1T} + \vec{q}_{2T} = -\not{\vec{p}}_T$$

$$\begin{aligned} m_{A_1} = m_{A_2} &\Rightarrow (p_{a_1} + p_{b_1} + q_1)^2 = (p_{a_2} + p_{b_2} + q_2)^2, \\ m_{B_1} = m_{B_2} &\Rightarrow (p_{b_1} + q_1)^2 = (p_{b_2} + q_2)^2, \end{aligned} \tag{2}$$

$$m_{C_1} = m_{C_2} \Rightarrow q_1^2 = q_2^2.$$

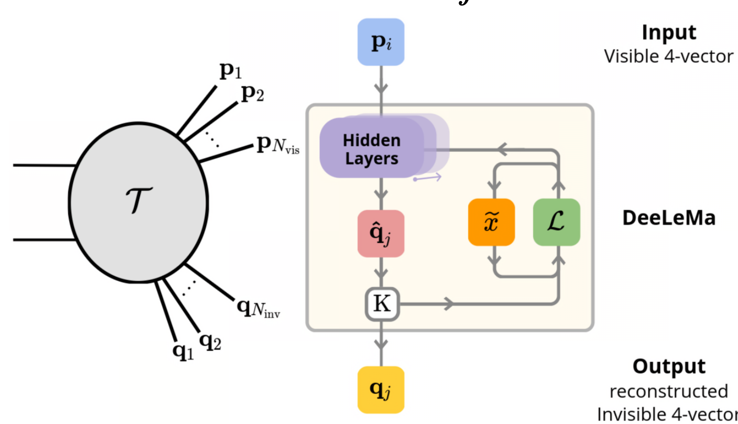
$$\begin{aligned} \mathcal{L} \equiv & \sum_{X=A,B} d(m_{X_1}, m_{X_2}) \\ & + \sum_{X=A,B} \left[d(\widetilde{m}_X, m_{X_1}) + d(\widetilde{m}_X, m_{X_2}) \right]. \end{aligned}$$

K ➡

- E_{q_1} and E_{q_2}: determined by given mass of C ($E_{q_1} = \sqrt{m_C^2 + q_{1x}^2 + q_{1y}^2 + q_{1z}^2}$).
- $q_{1,2x}, q_{1,2y}$: determined by DeeLeMa output with the transverse momentum condition $\sum \vec{p}_T = -\sum \vec{q}_T$. For example, when DeeLeMa outputs are the q_{1x} and q_{2y}, the q_{2x} and q_{1y} are automatically determined.
- q_{1z} and q_{2z}: fully determined by DeeLeMa output.

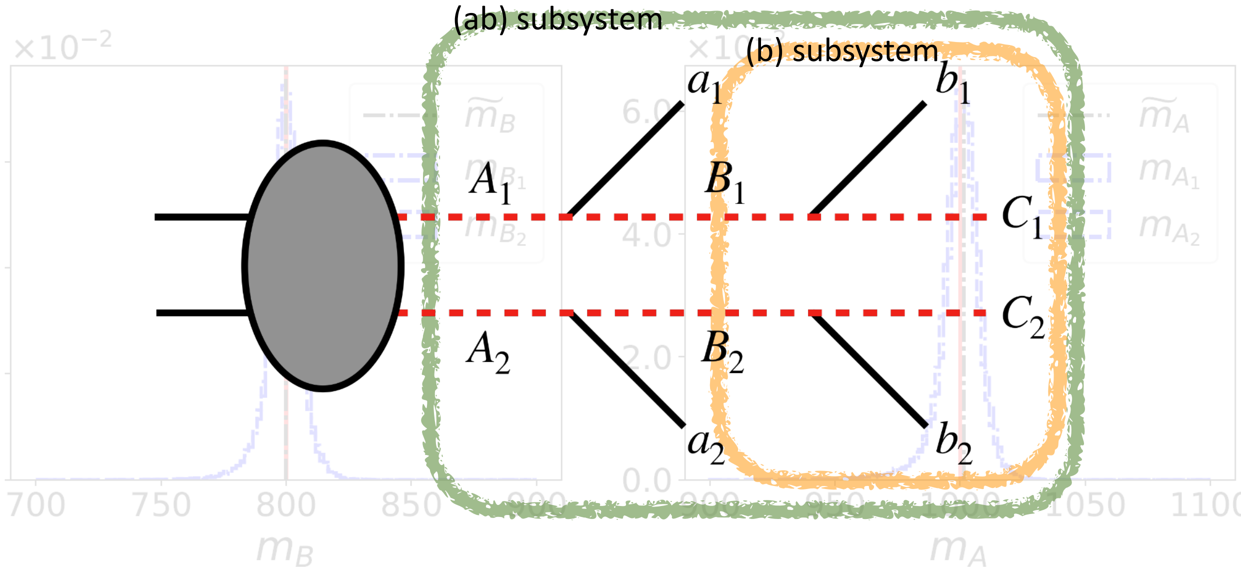
- Finally, we introduce the conjectured mass of mC, we impose these conditions and transverse momentum conditions as **K** which act on reconstructing q_j from the output $\hat{q}$ of **DeeLeMa**.

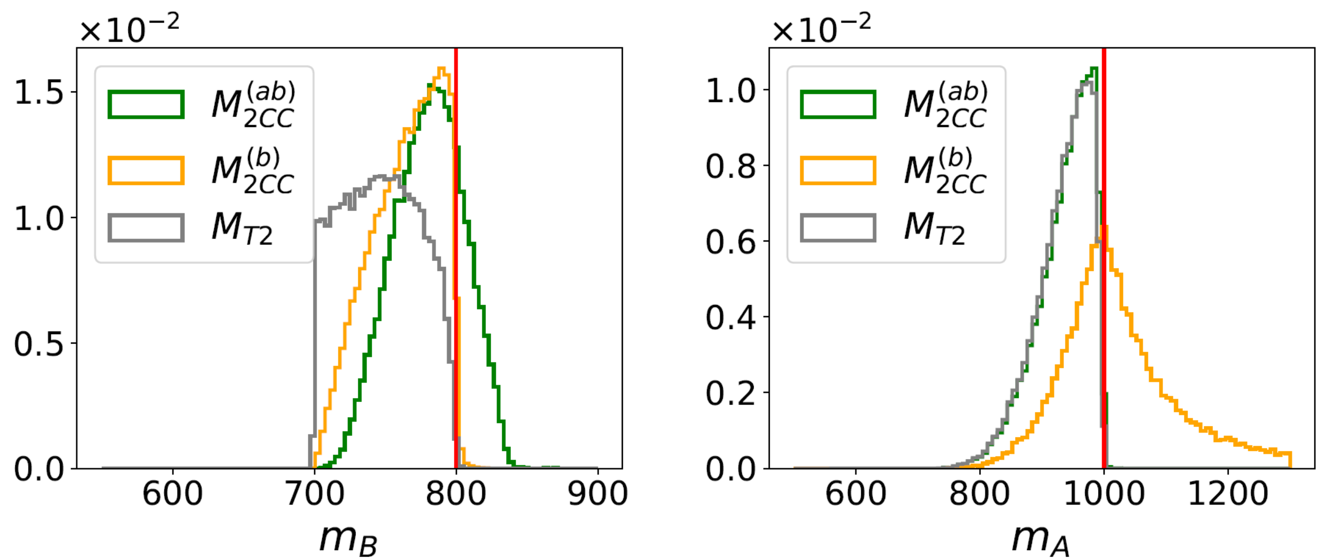
$$\text{DeeLeMa} \equiv \min_{\hat{q}, \widetilde{x}} \mathcal{L} \quad \text{subject to } \mathbf{K}.$$



Results Toy example

$m_A = 1000$ GeV, $m_B = 800$ GeV and $m_C = 700$ GeV

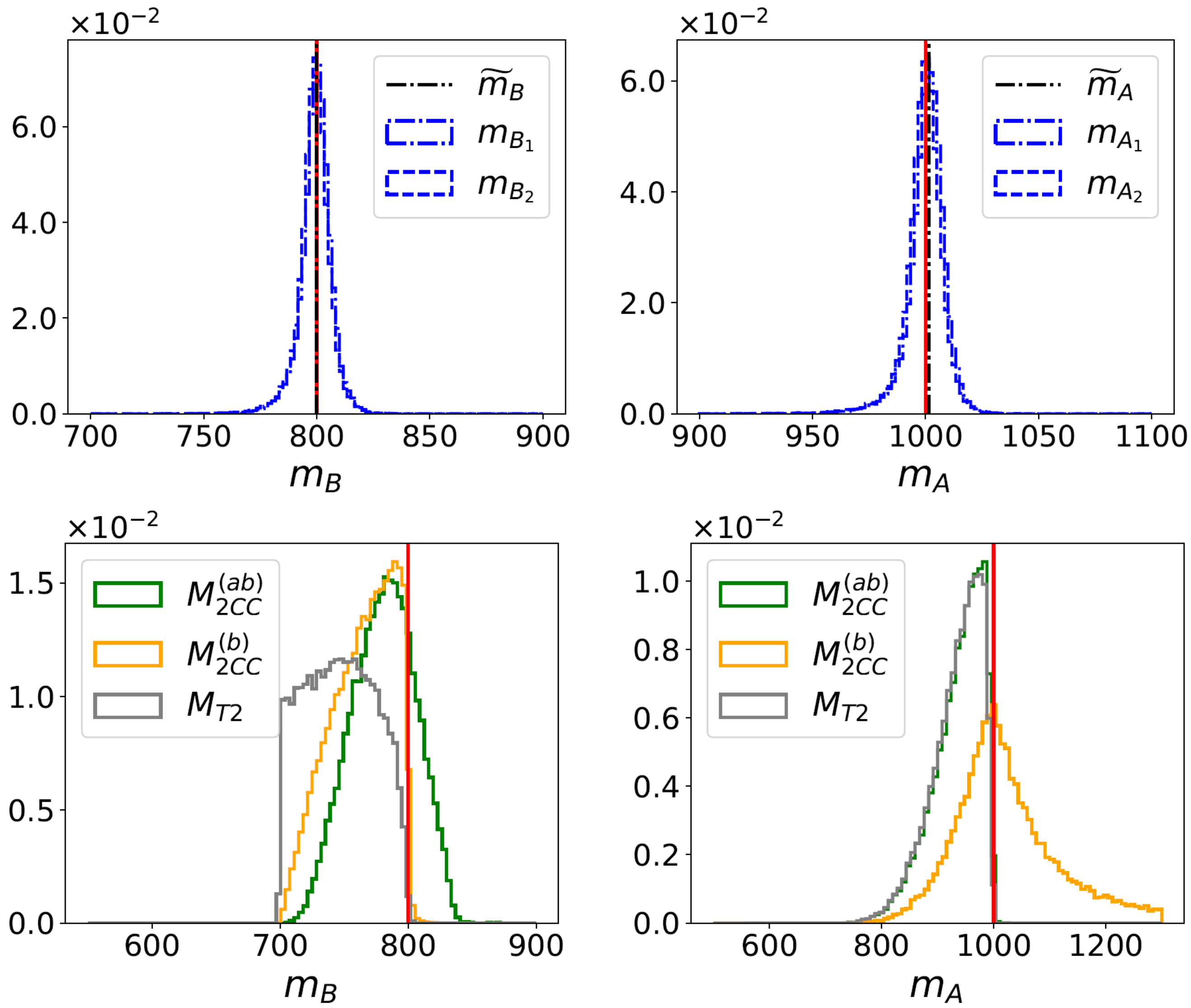




- A. Barr, C. Lester, and P. Stephens, “m(T2): The Truth behind the glamour,” J. Phys. G, vol. 29, pp. 2343–2363, 2003.
- C. G. Lester and D. J. Summers, “Measuring masses of semiinvisibly decaying particles pair produced at hadron colliders,” Phys. Lett. B, vol. 463, pp. 99–103, 1999
- W. S. Cho, J. S. Gainer, D. Kim, K. T. Matchev,F. Moortgat, L. Pape, and M. Park, “On-shell con-strained M2 variables with applications to mass measurements and topology disambiguation,” JHEP, vol. 08,p. 070, 2014

Results Toy example

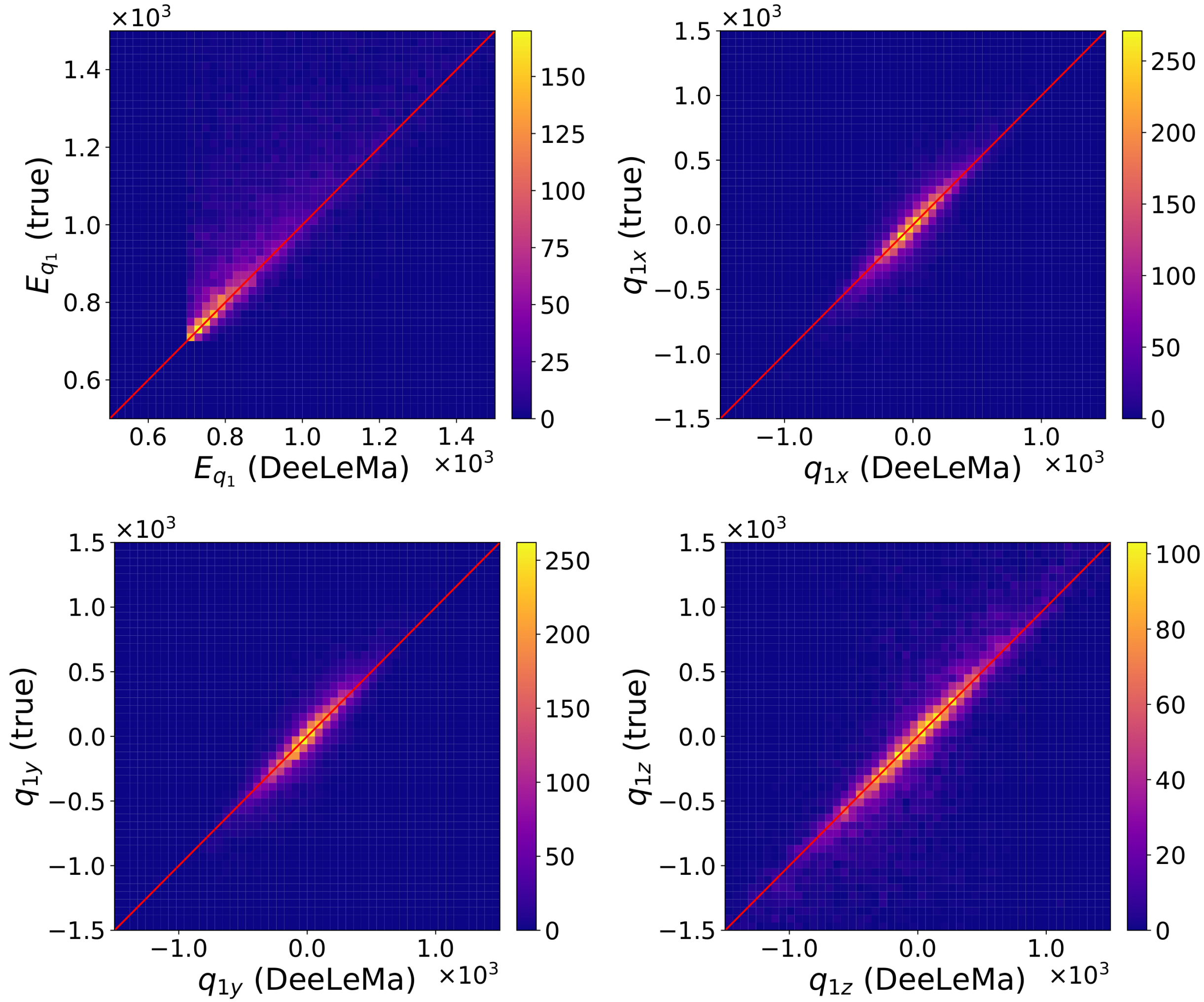
$m_A = 1000$ GeV, $m_B = 800$ GeV and $m_C = 700$ GeV



- A. Barr, C. Lester, and P. Stephens, “m(T2): The Truth behind the glamour,” J. Phys. G, vol. 29, pp. 2343–2363, 2003.
- C. G. Lester and D. J. Summers, “Measuring masses of semiinvisibly decaying particles pair produced at hadron colliders,” Phys. Lett. B, vol. 463, pp. 99–103, 1999
- W. S. Cho, J. S. Gainer, D. Kim, K. T. Matchev,F. Moortgat, L. Pape, and M. Park, “On-shell con-strainedM2variables with applications to mass measurements and topology disambiguation,” JHEP, vol. 08,p. 070, 2014

Results Toy example $(E_{q1}, q_{1x}, q_{1y}, q_{1z})$ and $(E_{q2}, q_{2x}, q_{2y}, q_{2z})$

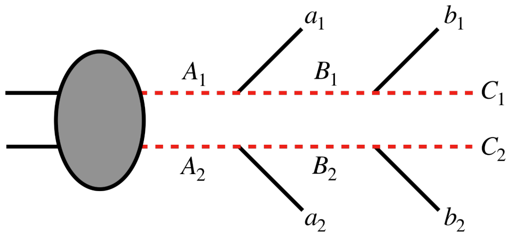
$m_A = 1000$ GeV, $m_B = 800$ GeV and $m_C = 700$ GeV



The red lines are the y=x axis and it indicates how accurate it is to the true value.

Combinatorics

$m_A = 1000$ GeV, $m_B = 800$ GeV and $m_C = 700$ GeV



- We tried {0, 10, 20, 50} [%] with the toy model.

$$\mathcal{E}_C \equiv \frac{\text{wrong}}{\text{wrong} + \text{correct}}$$

- Here, 0% means a perfect assignment, and 50% means a random assignment.

$\mathcal{E}_C$	$\widetilde{m}_X$ [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

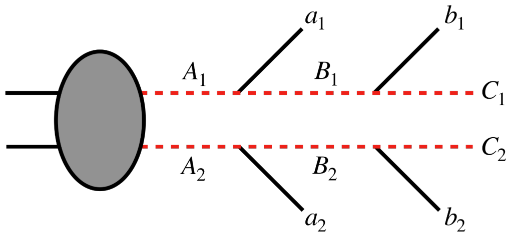
TABLE I: The summary table for combinatorial efficiency $\mathcal{E}_C$.

- Thanks to the characteristics of machine learning that solves tasks by identifying features and distributions from large amounts of data, it performs well even on data where some parts of the major are contaminated.

- <u>We conclude **DeeLeMa** is robust and practically free of combinatoric issues.</u>

Combinatorics

$m_A = 1000$ GeV, $m_B = 800$ GeV and $m_C = 700$ GeV



- We tried {0, 10, 20, 50} [%] with the toy model.

$$\mathcal{E}_C \equiv \frac{\text{wrong}}{\text{wrong} + \text{correct}}$$

- Here, 0% means a perfect assignment, and 50% means a random assignment.

$\mathcal{E}_C$	$\widetilde{m}_X$ [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

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

- Thanks to the characteristics of machine learning that solves tasks by identifying features and distributions from large amounts of data, it performs well even on data where some parts of the major are contaminated.

- <u>We conclude **DeeLeMa** is robust and practically free of combinatoric issues.</u>

Standard Model ttbar

$$\mathcal{E}_C \equiv \frac{\text{wrong}}{\text{wrong} + \text{correct}}$$

- The off-shell produced particles can cross the end-point, which is tiny for the narrow-width but inevitable and the other is the detector smearing effect.
- For the two b jets, we apply Gaussian smearing with jet pT up to {10, 20, 30, 50, 100, 400, 1000} GeV with the energy resolution {40, 28, 19, 13, 10, 6, 5}% and with $\mathcal{E}_C = 20\%$

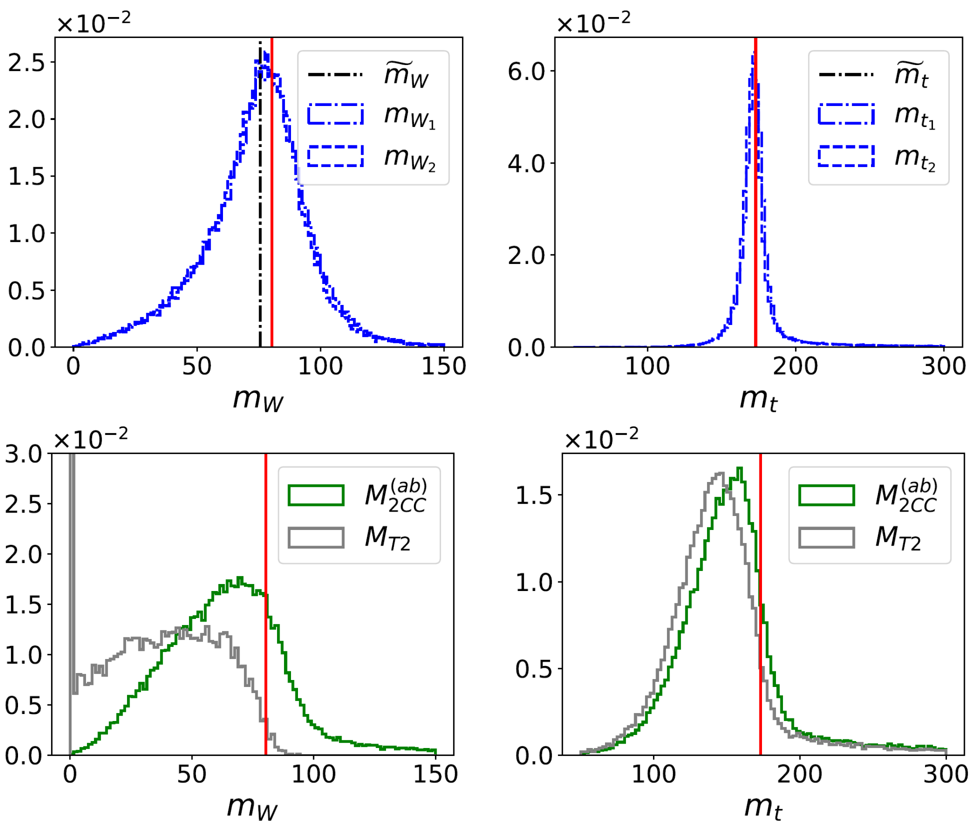
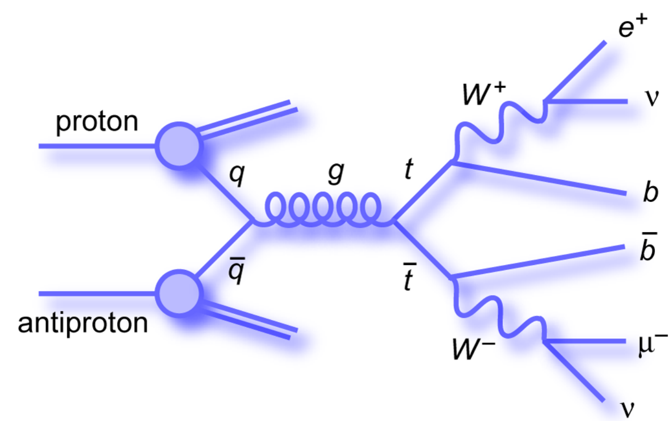


FIG. 5: Same as FIG. 4 but for the $t\bar{t}$ example.



- The sharpness of the endpoint collapses, and the endpoint is very vague after the detector smearing.
- However, in the point that **DeeLeMa** is the method to read the peak position, it is robust under these problems.

Conclusion

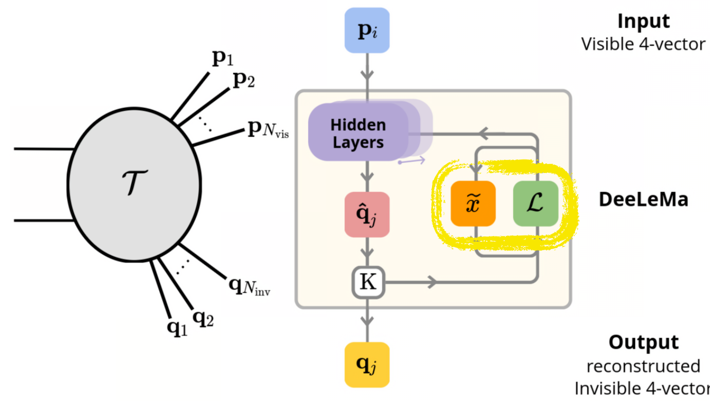
- We present a deep learning network named **DeeLeMa** to build an optimal search for events with missing kinematic information at the colliders.

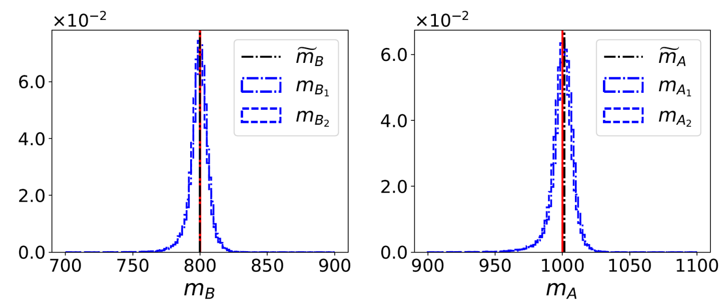
- To be specific, we present the baseline structure that can reconstruct the invisible momenta and the particle mass spectrum of the event as ttbar like antler event topology.

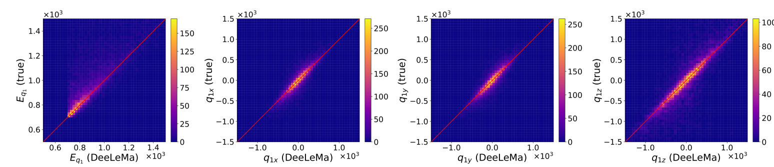
- We show that **DeeLeMa** can powerfully improve the reconstruction of not only the mass but also the momenta of invisible particles.

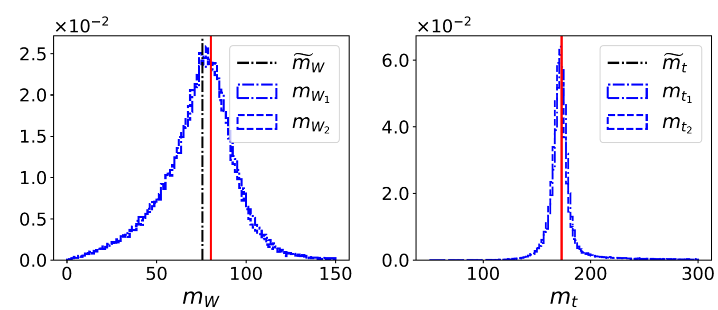
- Finally, since **DeeLeMa** reconstructs masses as a sharp peak, **it is robust under the combinatorial problem as well as detector smearing and off-shell effects**, which is a very good advantage to apply to real experiments.

- *(Ongoing work) More complicated problem*
  >> Unknown - Constraints > 2









Thank you
for your attention