Revisiting combinatorial problem in the dileptonic $t\bar{t}$ production with neural networks

Cosmos (Zhongtian) Dong

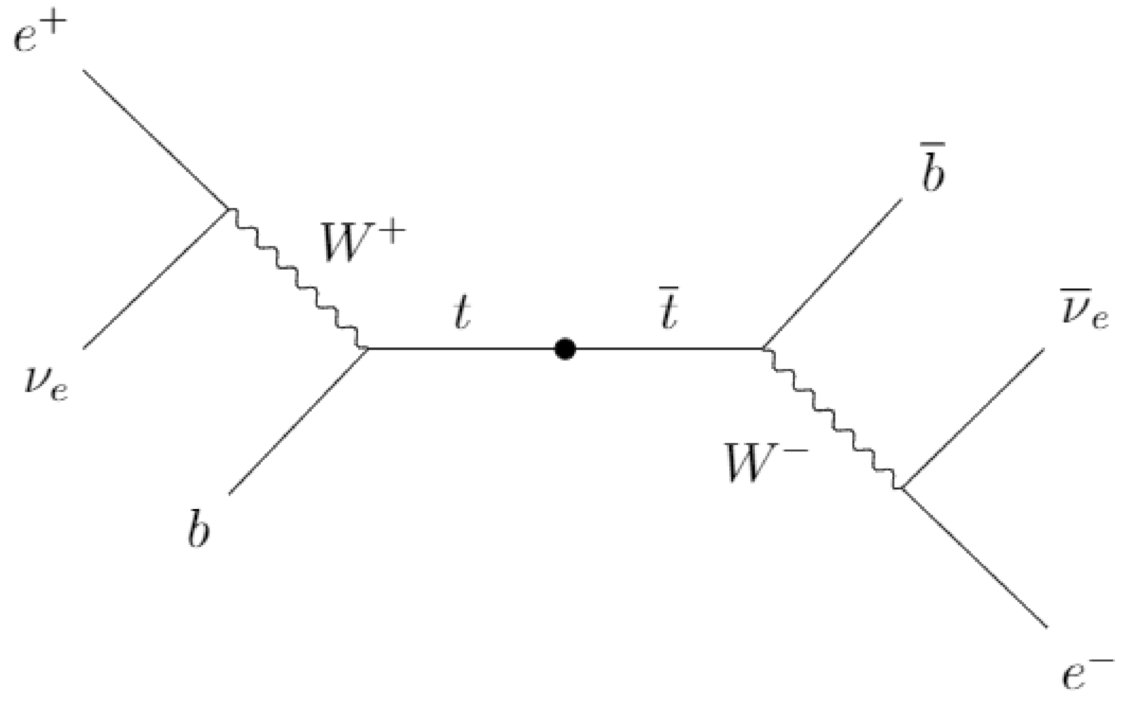
University of Kansas

Based on work done in collaboration with

H. Alhazmi, L. Huang, J. Kim, K.C. Kong, D.Shih

Dileptonic ttbar-like events topology

- Imperfect quark charge measurement and missing kinematic information
- Two-fold ambiguity in reconstructing top quark momentum
- We would like to develop an algorithm that finds the correct pairing
- We want to compare existing methods with machine learning approaches
- Event generation using MG5+ Gaussian smearing (following detector resolution given by ATLAS Collaboration)



Method 1 : endpoints

- Check if mass variables given by specific pairing violates endpoints
- Event is considered unresolved if both pairings does not violate the endpoints

$$max\{m_{b\ell^+}, m_{b\ell^-}\} \leqslant m_{b\ell}^{max} = \sqrt{\frac{(m_{top}^2 - m_W^2)(m_W^2 - m_\nu^2)}{m_W^2}}$$

$$M_{T2}(\tilde{m}) \equiv \min_{\vec{q}_{1T}, \vec{q}_{2T}} \{\max [M_{TP_1}(\vec{q}_{1T}, \tilde{m}),\ M_{TP_2}(\vec{q}_{2T}, \tilde{m})]\}$$

$$\vec{q}_{1T} + \vec{q}_{2T} = \vec{\not{P}}_T \,,$$

9906349, Lester, Summers

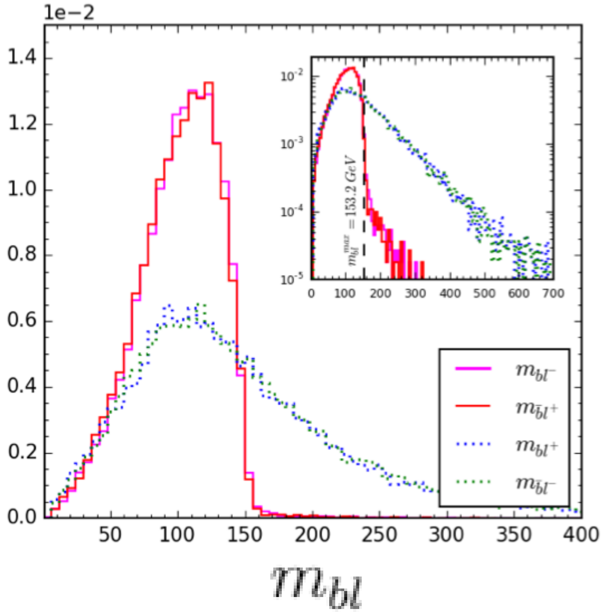
- We can use other similarly defined mass variable with different constraints such as $M_{2CC},\ M_{2CW},\ M_{2CT}$

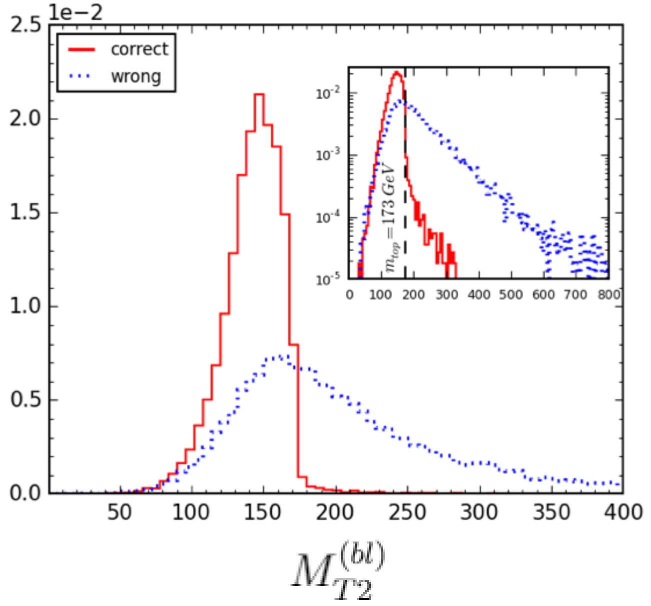
1401.1449, Cho, Gainer, Kim, Matchev, Moortgat, Pape, Park

1703.06887, Kim, Matchev, Moortgat, Pape

Parton level: efficiency 77% purity 96%
Smeared level: efficiency 74% purity 87%

Efficiency is defined by percentage of resolved events of all events. Purity is the percentage of correctly resolved events of the resolved events.





1706.04995 Debnath, Kim, Kim, Kong, Matchev

Method 2 : Topness

- Chi square minimization over MET constraint
- Smaller Chi square values are considered more likely to be the correct pairing

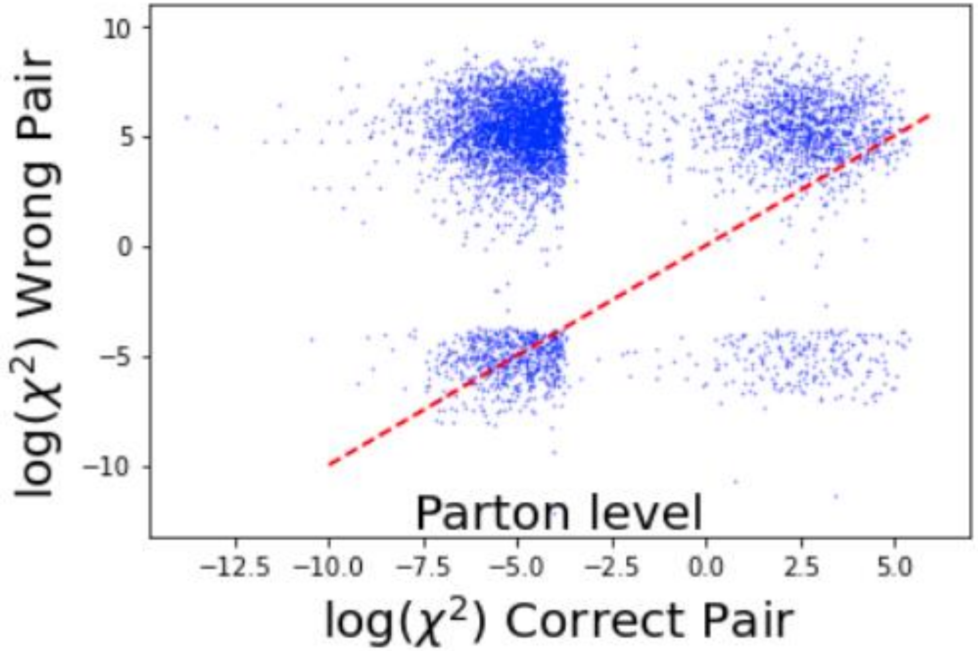
1212.4495 Graesser, Shelton

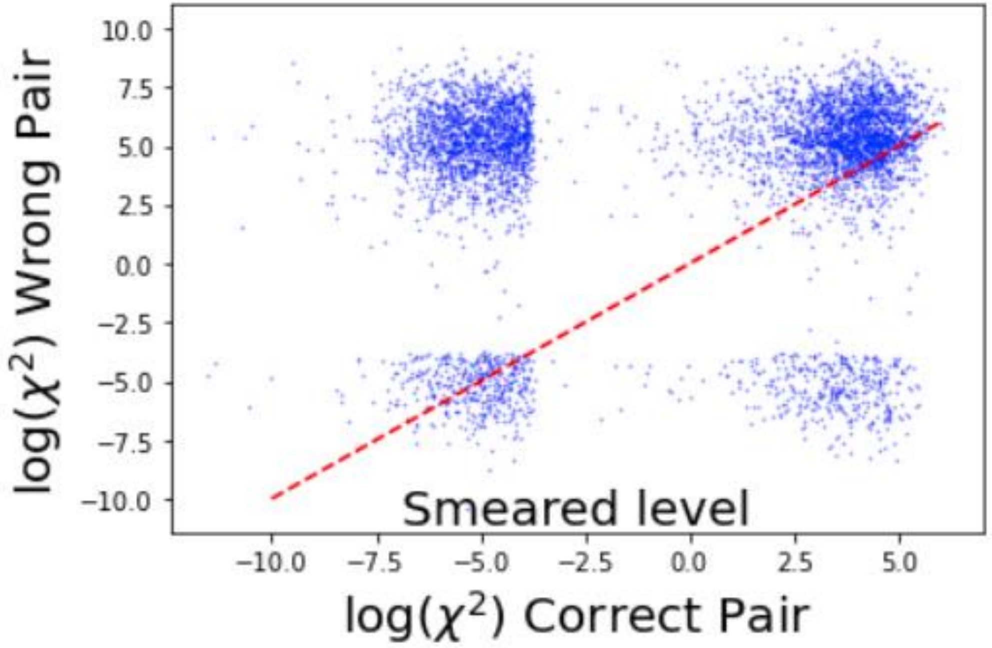
1807.11498 Kim, Kong, Matchev, Park

$$\chi^2_{ij} \equiv \min_{\not{p}_T = p_{\nu T} + p_{\bar{\nu} T}} \left[\frac{\left(m^2_{b_i l^+ \nu} - m_t^2\right)^2}{\sigma_t^4} + \frac{\left(m^2_{l^+ \nu} - m_W^2\right)^2}{\sigma_W^4} + \frac{\left(m^2_{b_j l^- \bar{\nu}} - m_t^2\right)^2}{\sigma_t^4} + \frac{\left(m^2_{l^- \bar{\nu}} - m_W^2\right)^2}{\sigma_W^4} \right]$$

Parton level: efficiency 100% purity 87%

Smeared level: efficiency 100% purity 81%





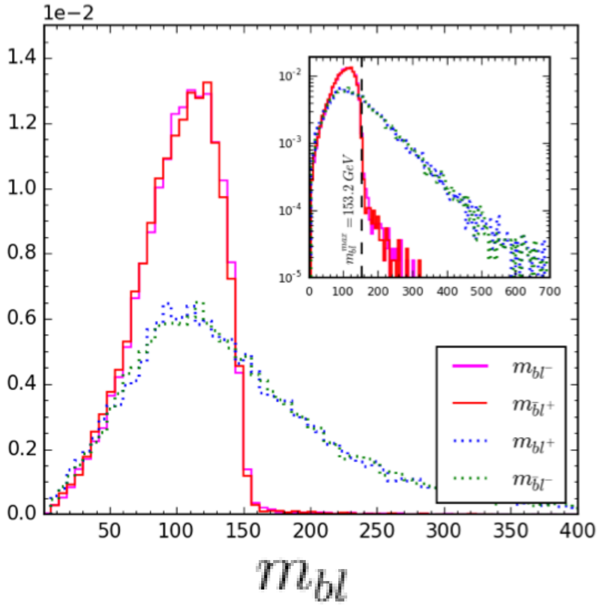
Method 3 : KLfitter

- Uses transfer functions and probability densities to estimate likelihood of measured objects given model signature
- Depend on angular correlation and invariant mass

$$\mathcal{L} = \prod_{x,y} \mathcal{G}\left(E_i^{\text{miss}} | p_i^{\nu_1}, p_i^{\nu_2}, \sigma_i^{\text{miss}}\left(m_t, m_W, \eta_{\nu_1}, \eta_{\nu_2}\right)\right) \cdot$$

$$\prod_{i=1}^{2} \mathcal{G}\left(\eta_{\nu_i} | m_t\right) \cdot \left(m_{\ell_1,q_1} + m_{\ell_2,q_2}\right)^{\alpha} \cdot$$

$$\prod_{i=1}^{2} W_{\text{jet}}\left(E_{\text{jet},i}^{\text{meas}} | E_{\text{jet},i}\right) \cdot \prod_{i=1}^{2} W_{\ell}\left(E_{\ell,i}^{\text{meas}} | E_{\ell,i}\right)$$



Parton level: efficiency 100% purity 85%

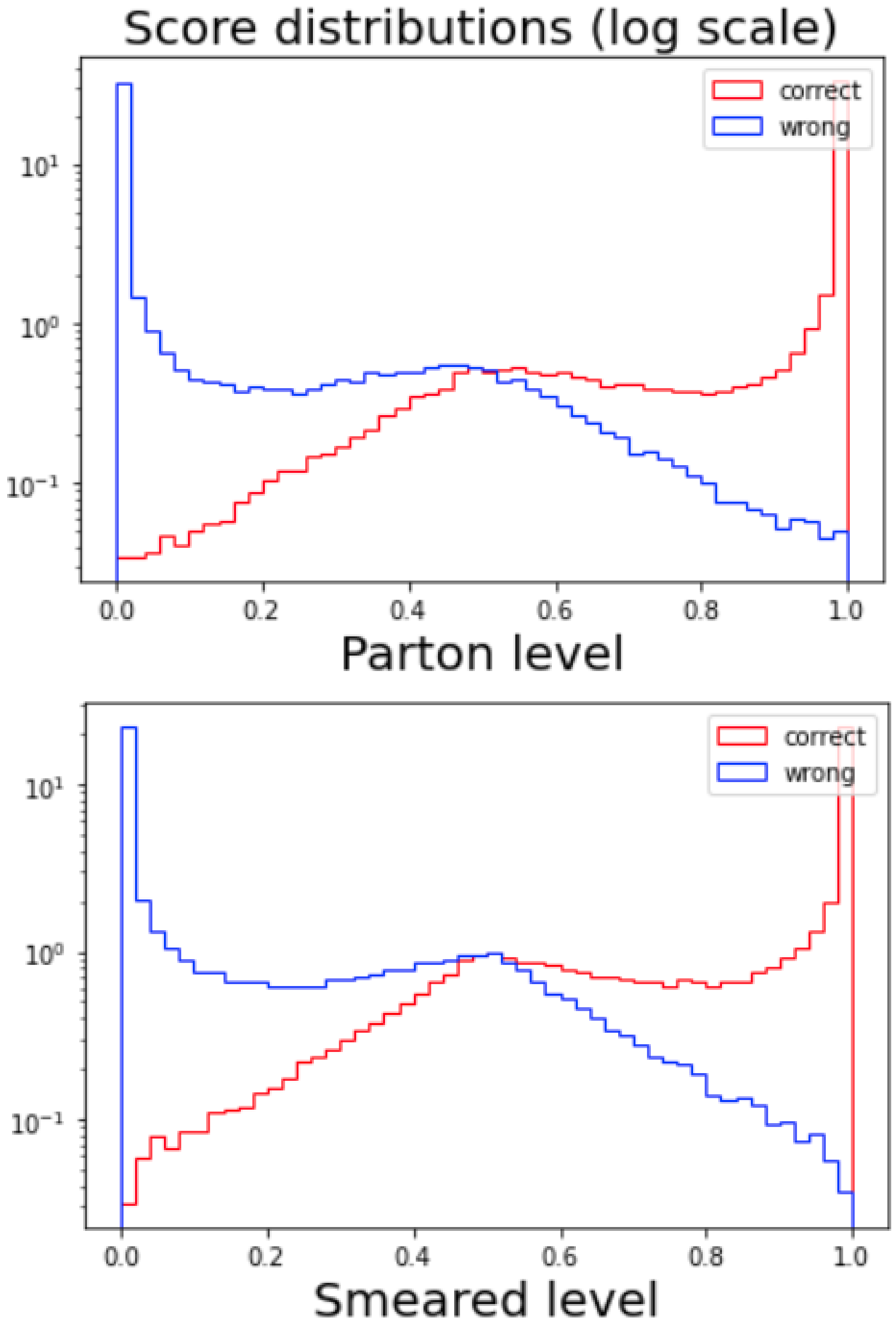
1312.5595 Erdmann, Guindon, Kroninger, Lemmer Nackenhorst, Quadt, Stolte

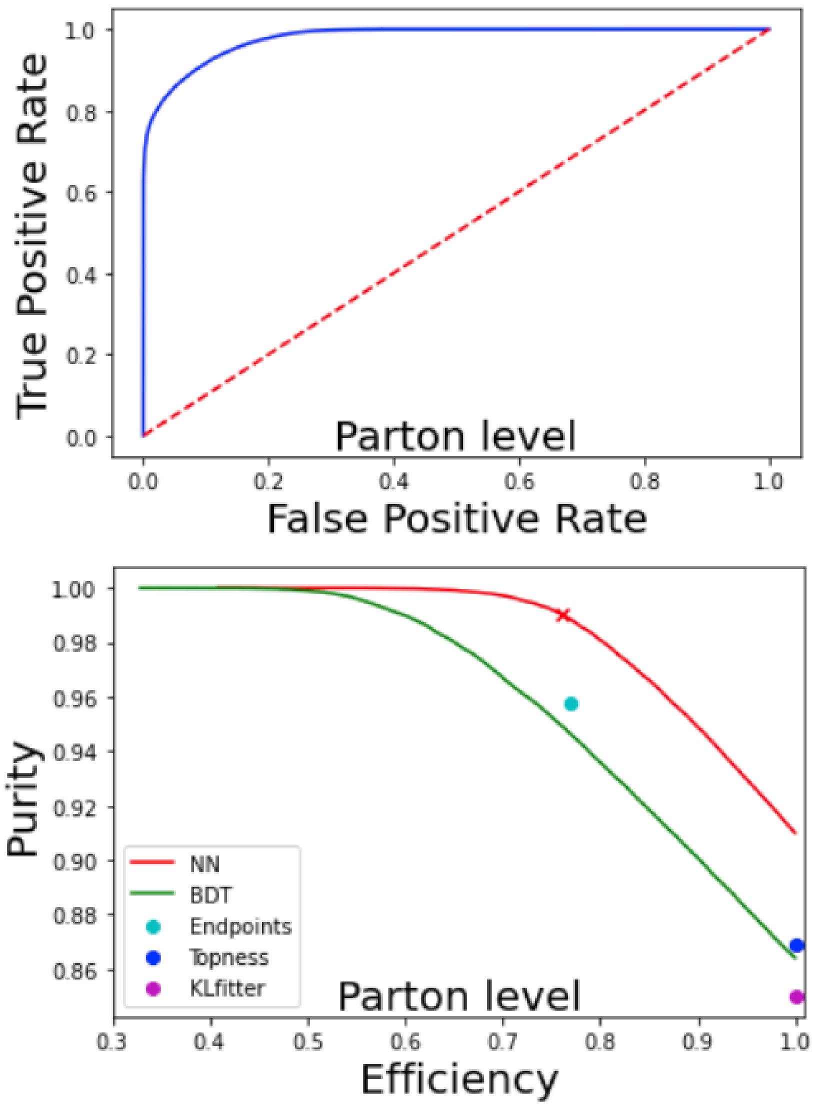
github.com/KLFitter/KLFitter

Neural network approach

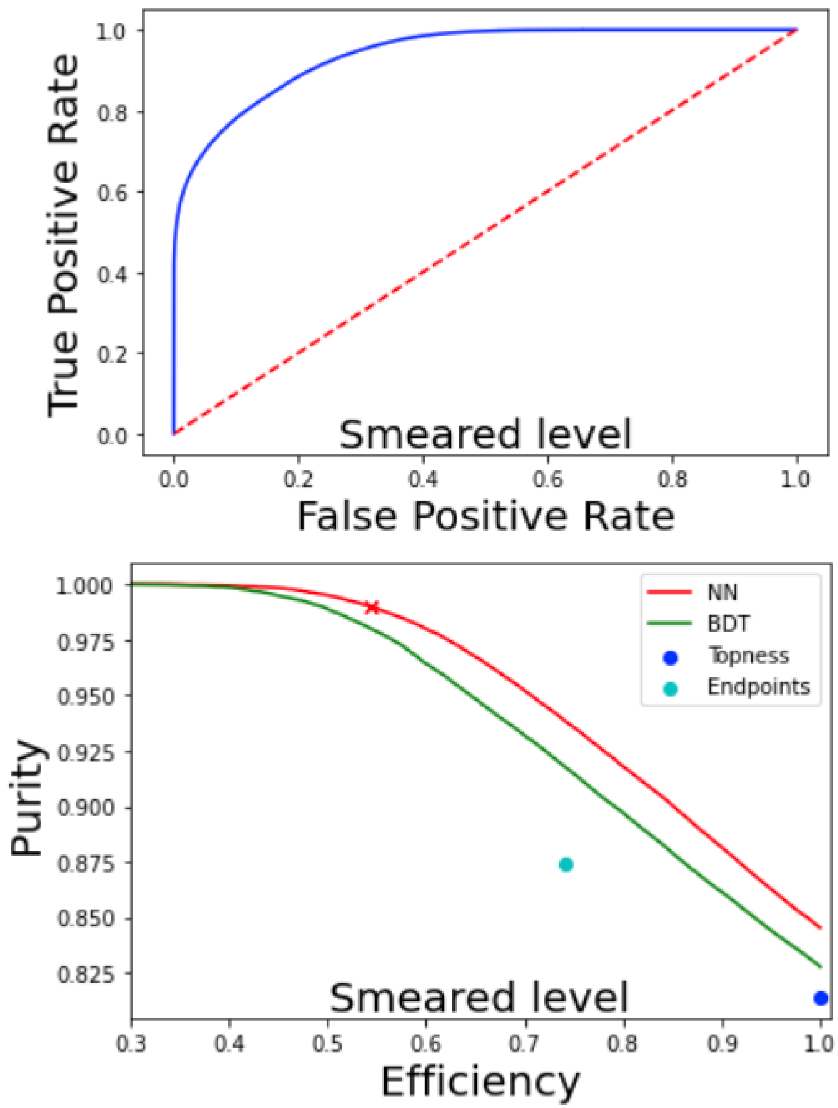
- Uses four-momentum information as inputs
- Formulated as a binary classification problem
- Three hidden layer fully connected neural network
- Using Keras library and trained with Adam optimizer

Parton level: efficiency 100% purity 90%
Smeared level: efficiency 100% purity 85%





Parton level: efficiency 76% purity 99% (marked by X)



Smeared level: efficiency 54% purity 99% (marked by X)

Attention mechanism

- Recent studies shows success in multi-jets combinatorial assignment in the top quark pair production

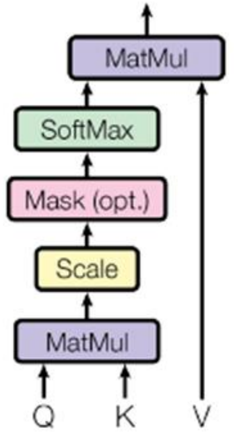
  2010.09206 Fenton, Shmakov, Ho, Hsu, Whiteson, Baldi
  2012.03542 Lee, Park, Watson, Yang

- Permutation invariant architecture
- Uses dot product between every object in the sequence

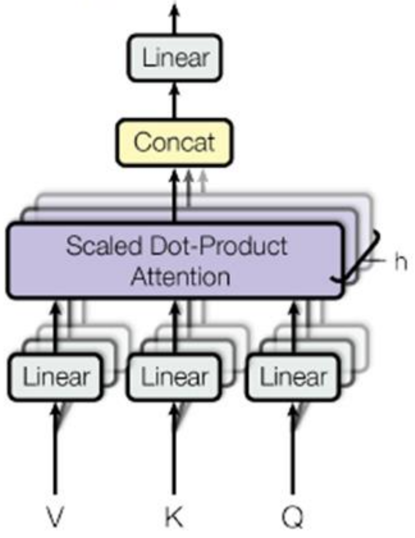
Parton level: efficiency 100% purity 90%

Smeared level: efficiency 100% purity 85%

Scaled Dot-Product Attention



Multi-Head Attention



1706.03762 Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, Polosukhin

Summary

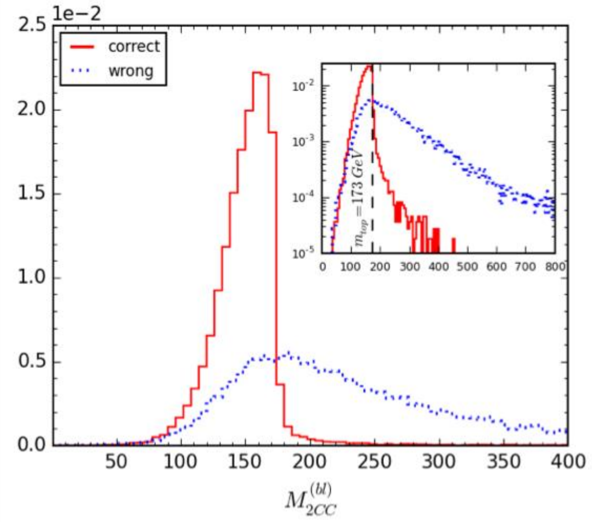
- Combinatorial ambiguity is everywhere
- Existing methods (Endpoint, Topness, KLfitter etc.) resolve two-fold ambiguity using kinematic features
- Machine learning based methods outperform existing kinematic methods
- The methods can be used for precision measurement i.e. CP phase measurement in the tth production
- The methods can be generalized for arbitrary mass spectrum. (work in progress)

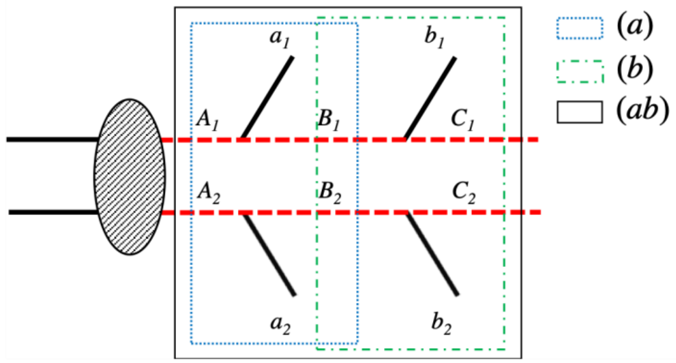
Algorithm	Parton		Smeared	
	purity	efficiency	purity	efficiency
Endpoints method	0.957	0.769	0.874	0.742
Topness method	0.869	1	0.814	1
KLfitter	0.85	1		
Neural Network	0.895	1	0.845	1
Neural Network	0.990	0.762	0.990	0.543
Attention Network	0.898	1	0.844	1
Boosted tree	0.861	1	0.824	1
BDT with masses input	0.894	1	0.840	1

Method 1 : endpoints

$$M_{2CC} \equiv \min_{\vec{q}_1,\vec{q}_2} \{\max [M_{P_1}(\vec{q}_1,\tilde{m}),\ M_{P_2}(\vec{q}_2,\tilde{m})]\}$$
$$\vec{q}_{1T} + \vec{q}_{2T} = \not{\vec{P}}_T$$
$$M_{P_1} = M_{P_2}$$
$$M_{R_1}^2 = M_{R_2}^2$$

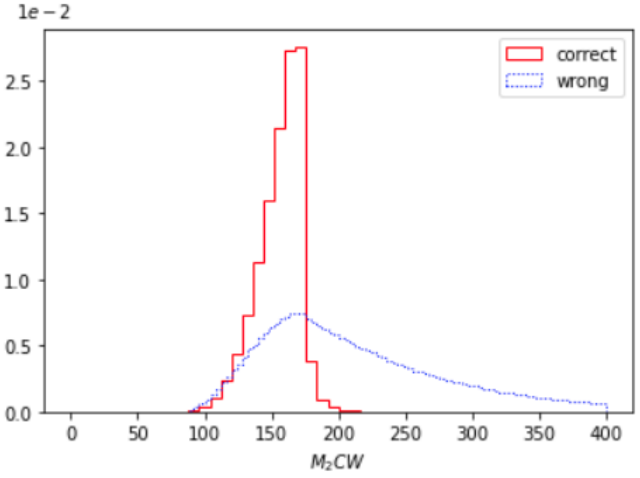
1401.1449, Cho, Gainer, Kim, Matchev, Moortgat, Pape, Park

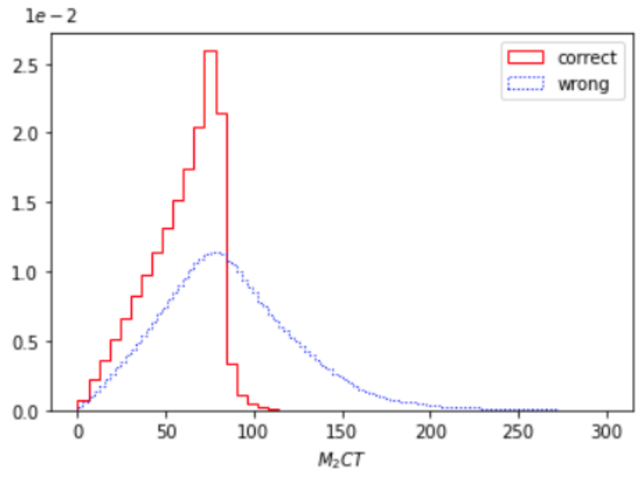




$$M_{2CW}^{(b\ell)} \equiv \min_{\vec{q}_1,\vec{q}_2} \{\max [M_{P_1}(\vec{q}_1,\tilde{m}),\ M_{P_2}(\vec{q}_2,\tilde{m})]\}$$
$$\vec{q}_{1T} + \vec{q}_{2T} = \not{\vec{P}}_T$$
$$M_{P_1} = M_{P_2}$$
$$M_{R_1}^2 = M_{R_2}^2 = m_W^2$$

$$M_{2Ct}^{(\ell)} \equiv \min_{\vec{q}_1,\vec{q}_2} \{\max [M_{P_1}(\vec{q}_1,\tilde{m}),\ M_{P_2}(\vec{q}_2,\tilde{m})]\}$$
$$\vec{q}_{1T} + \vec{q}_{2T} = \not{\vec{P}}_T$$
$$M_{P_1} = M_{P_2}$$
$$M_{R_1}^2 = M_{R_2}^2 = m_t^2$$





1703.06887, Kim, Matchev, Moortgat, Pape

Algorithm	Parton		Smeared	
	purity	efficiency	purity	efficiency
Endpoints method	0.957	0.769	0.874	0.742
Topness method	0.869	1	0.814	1
KLfitter	0.85	1		
Neural Network	0.895	1	0.845	1
Neural Network	0.990	0.762	0.990	0.543
Attention Network	0.898	1	0.844	1
Boosted tree	0.861	1	0.824	1
BDT with masses input	0.894	1	0.840	1

Data generation

- 1 million simulated events at parton level
- 14 TeV center of mass energy

- Use Gaussian smearing to simulate detector effects

$$\frac{\sigma_{jets}}{E} = \sqrt{(\frac{5.3}{E})^2 + (\frac{0.74}{\sqrt{E}})^2 + 0.05^2}$$

$$\frac{\sigma_e}{E} = \sqrt{(\frac{0.3}{E})^2 + (\frac{0.1}{\sqrt{E}})^2 + 0.01^2}$$

Matrix element method

- Uses true neutrino momentum
- 93% find the correct pairing