Community Summer Study
SNOWMASS
July 17-26 2022, Seattle

Kinematic Variables and Feature Engineering for Particle Phenomenology

TEXAS A&M UNIVERSITY
Physics & Astronomy

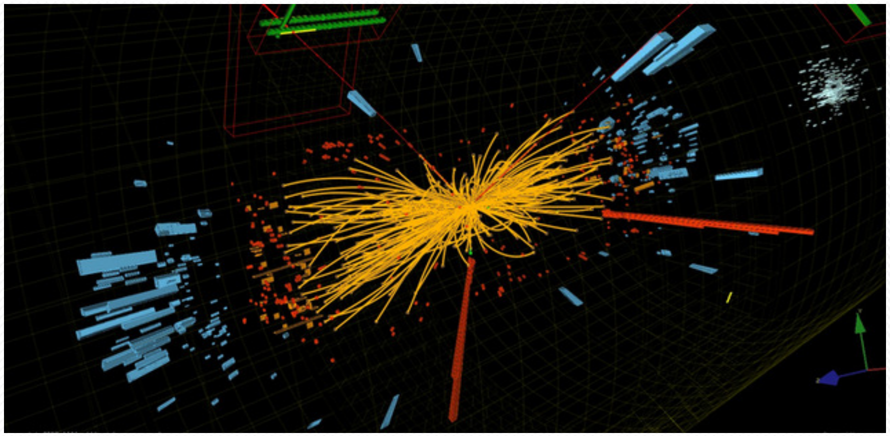
Doojin Kim

(doojin.kim@tamu.edu)

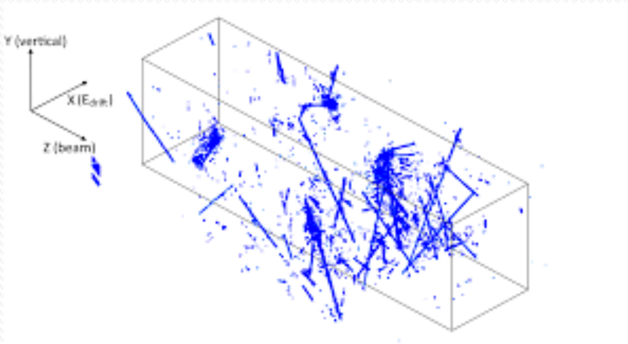
Seattle Snowmass Summer Meeting 2022

July 19th, 2022

The “Curse” of Dimensionality



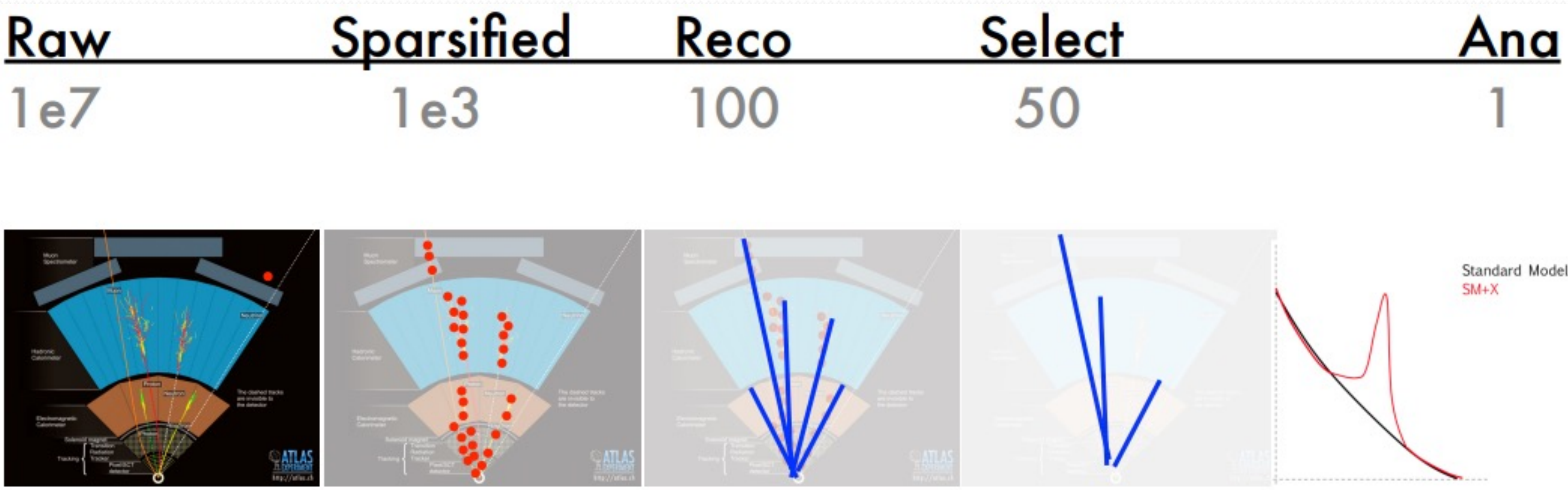
At colliders (e.g., LHC)



At neutrino experiments (e.g., μBooNE)

Dimensional Reduction

(Image from Daniel Whiteson's talk in Seattle Snowmass Summer Meeting)



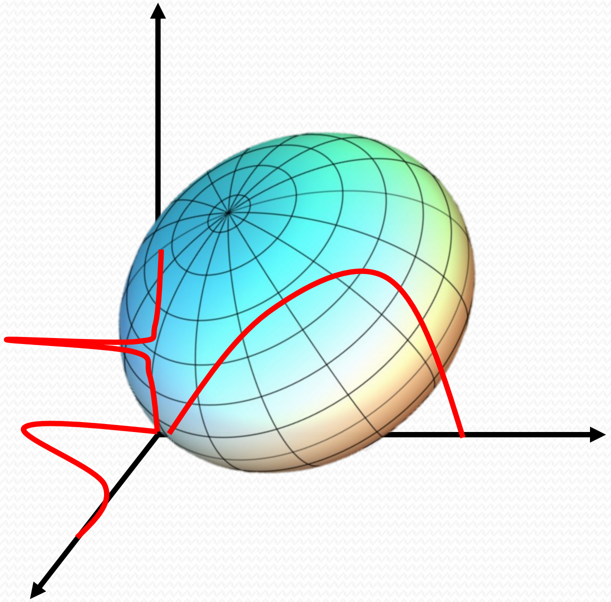
Less intuitive

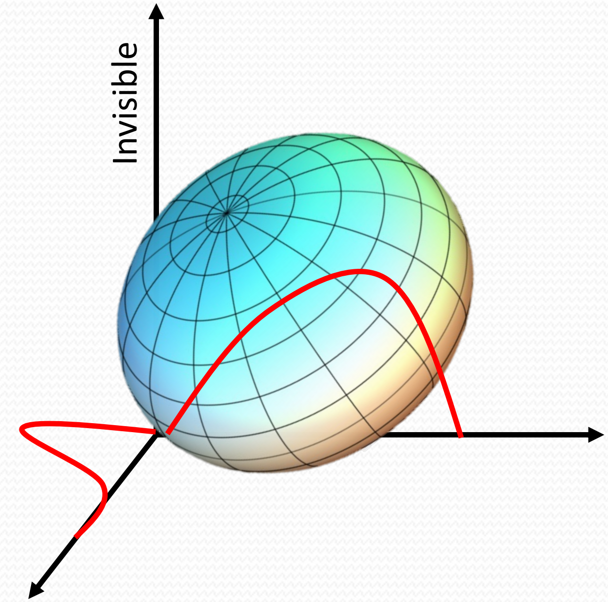
More intuitive

Keeping information

Losing information

Kinematic Variables and Challenges

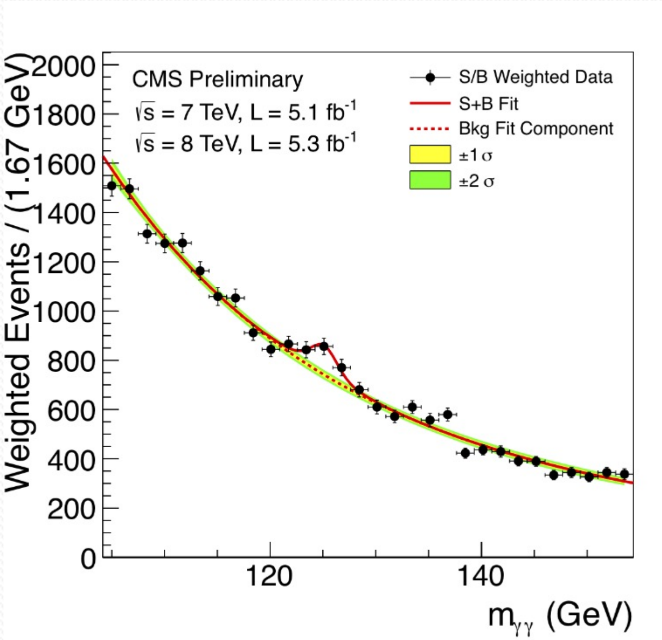


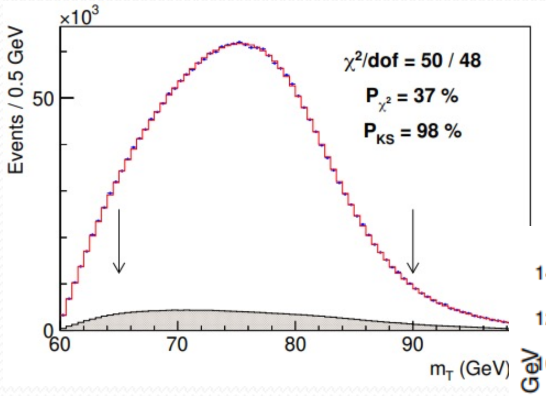


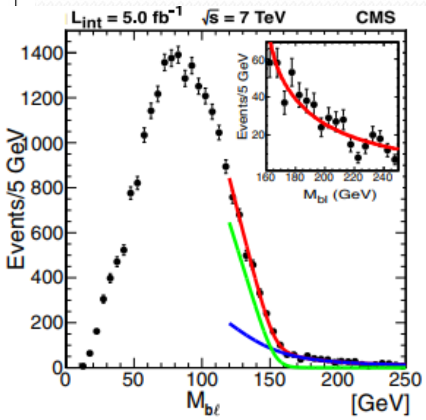
Which axis (or variable) in the visible space is best sensitive to the signal of interest?

The existence of invisible particles (e.g., neutrinos, dark matter candidates)

Discovery and Precision Measurement







Kinematic variables

- Expedite the discovery of new particles,
- Help interpreting physics data,
- Allow for particle property measurements (e.g., mass, spin, coupling)

Kinematic Variables and Feature Engineering

Kinematic Variables and Feature Engineering for Particle Phenomenology

Roberto Franceschini,[1,*] Doojin Kim,[2,†] Kyoungchul Kong,[3,‡] Konstantin T. Matchev,[4,§] Myeonghun Park,[5,6,¶] and Prasanth Shyamsundar[7,**]

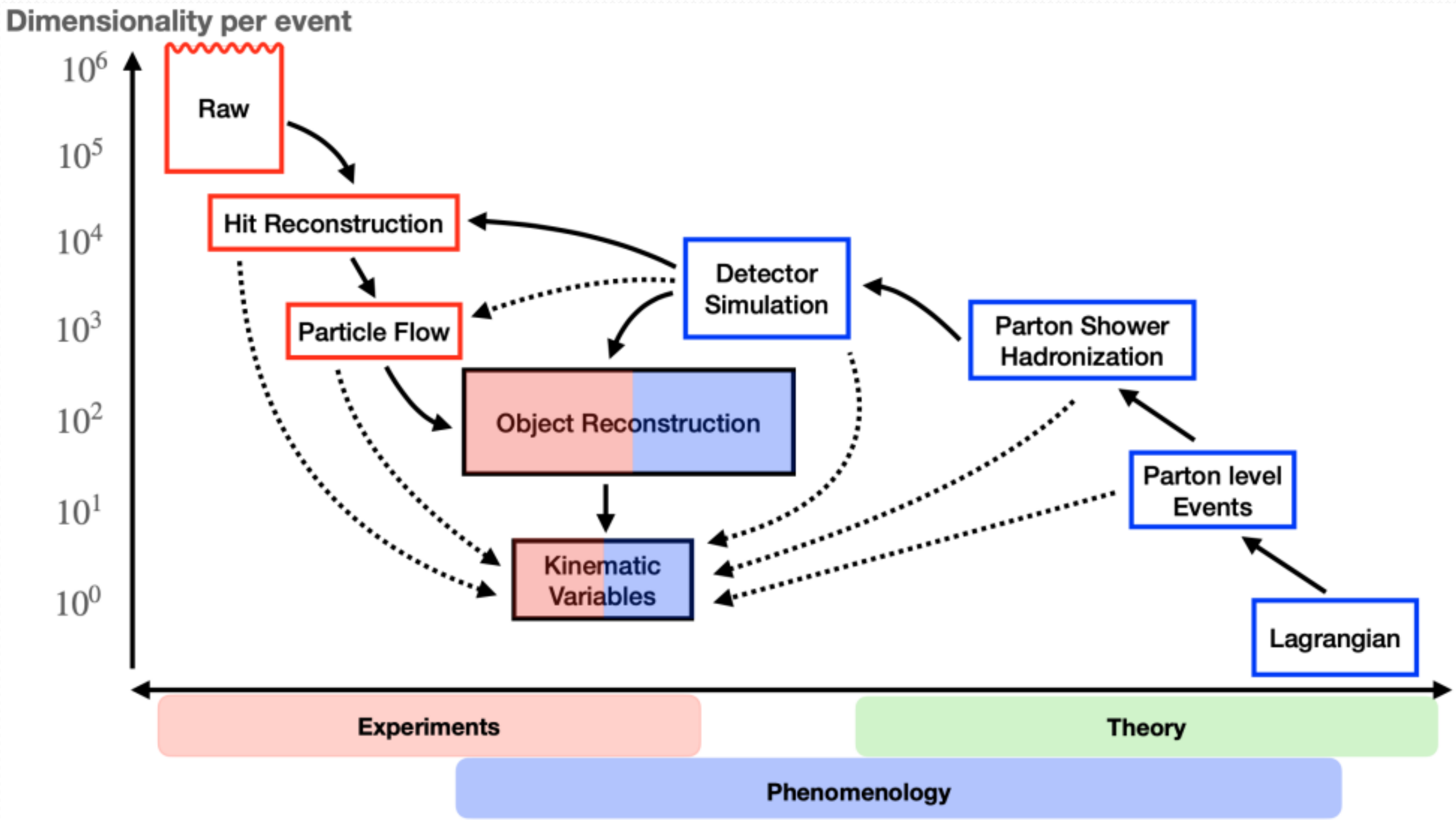
[1] *Università degli Studi Roma Tre and INFN Roma Tre, Via della Vasca Navale 84, I-00146 Roma, Italy*
[2] *Mitchell Institute for Fundamental Physics and Astronomy, Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA*
[3] *Department of Physics and Astronomy, University of Kansas, Lawrence, KS 66045, USA*
[4] *Institute for Fundamental Theory, Physics Department, University of Florida, Gainesville, FL 32611, USA*
[5] *Faculty of Natural Sciences, Seoultech, 232 Gongneung-ro, Nowon-gu, Seoul, 01811, Korea*
[6] *School of Physics, KIAS, Seoul 02455, Korea*
[7] *Fermilab Quantum Institute, Fermi National Accelerator Laboratory, Batavia, IL 60510, USA*

Kinematic variables have been playing an important role in collider phenomenology, as they expedite discoveries of new particles by separating signal events from unwanted background events and allow for measurements of particle properties such as masses, couplings, spins, etc. For the past 10 years, an enormous number of kinematic variables have been designed and proposed, primarily for the experiments at the Large Hadron Collider, allowing for a drastic reduction of high-dimensional experimental data to lower-dimensional observables, from which one can readily extract underlying features of phase space and develop better-optimized data-analysis strategies. We review these recent developments in the area of phase space kinematics, summarizing the new kinematic variables with important phenomenological implications and physics applications. We also review recently proposed analysis methods and techniques specifically designed to leverage the new kinematic variables. As machine learning is nowadays percolating through many fields of particle physics including collider phenomenology, we discuss the interconnection and mutual complementarity of kinematic variables and machine learning techniques. We finally discuss how the utilization of kinematic variables originally developed for colliders can be extended to other high-energy physics experiments including neutrino experiments.

[Franceschini, **DK**, Kong, Matchev, Park, Shyamsundar, arXiv:2206.13431]

Any feedback will be more than welcome.

Flow of Dimensionality Reduction and Kinematic Variables

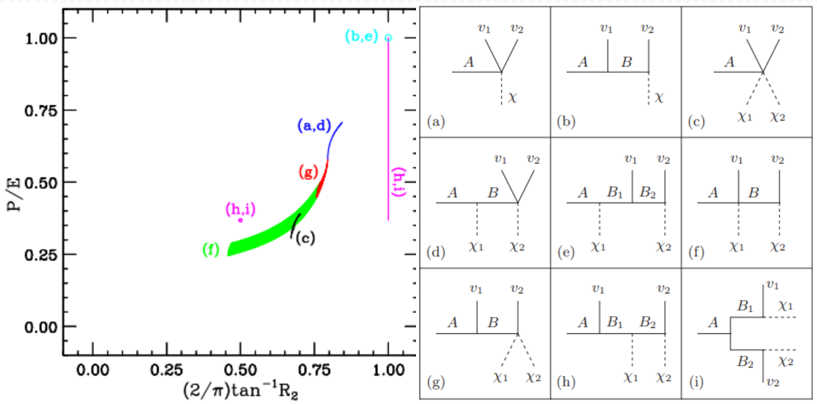


Inclusive (Shape) Variables

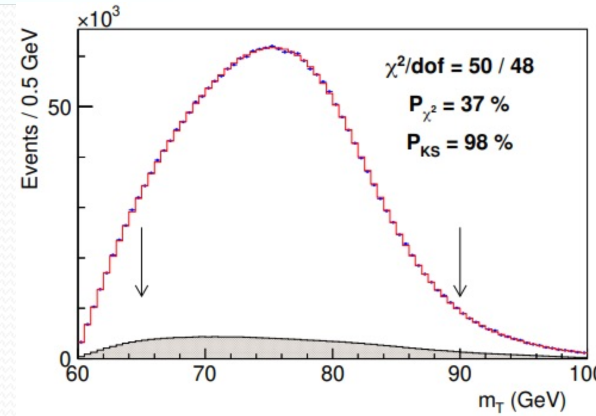
Observable	Definition
Sphericity	$S = \frac{3}{2}(\lambda_2 + \lambda_3)$, $\lambda_i (\lambda_1 \geq \lambda_2 \geq \lambda_3)$, eigenvalues of $M_{ij} = \frac{\sum_{a=1}^{n_j} p_{a,i} p_{a,j}}{\sum_{a=1}^{n_j} \lvert\vec{p}_a\rvert^2}$ with $i, j \in \{x, y, z\}$
Transverse sphericity	$S_T = \frac{2\lambda_2}{\lambda_1 + \lambda_2}$
Aplanarity	$A = \frac{3}{2}\lambda_3$
Planarity	$P = \lambda_2 - \lambda_3$
(Transverse) spherocity	$S_0 = \frac{\pi^2}{4} \min_{\hat{n}} \left(\frac{\sum_a \lvert\vec{p}_{a,T} \times \hat{n}\rvert}{\sum_a p_{a,T}} \right)^2$
Thrust	$T = \max_{\hat{n}} \left(\frac{\sum_a \lvert\vec{p}_a \cdot \hat{n}\rvert}{\sum_a \lvert\vec{p}_a\rvert} \right)$
Thrust major	$T_{\text{major}} = \max_{\hat{n}_{\text{ma}} \perp \hat{n}_T} \left(\frac{\sum_a \lvert\vec{p}_a \cdot \hat{n}_{\text{ma}}\rvert}{\sum_a \lvert\vec{p}_a\rvert} \right)$
Thrust minor	$T_{\text{minor}} = \frac{\sum_a \lvert\vec{p}_a \cdot \hat{n}_{\text{mi}}\rvert}{\sum_a \lvert\vec{p}_a\rvert}$ with $\hat{n}_{\text{mi}} = \hat{n}_T \times \hat{n}_{\text{ma}}$
Oblateness	$\mathcal{O} = T_{\text{major}} - T_{\text{minor}}$
Normalized hemisphere mass	$M_{1(2)}^2 = \frac{1}{E_{\text{CM}}^2} \left(\sum_{a \in H_{1(2)}} p_a \right)^2$ with $H_{1(2)}$ being hemispheres divided by the plane normal to $\hat{n}_T$
Heavy jet mass	$M_H^2 = \max(M_1^2, M_2^2)$
Light jet mass	$M_L^2 = \min(M_1^2, M_2^2)$
Jet mass difference	$M_D^2 = \lvert M_1^2 - M_2^2 \rvert$
Jet broadening	$B_{1(2)} = \frac{\sum_{a \in H_{1(2)}} \lvert\vec{p}_a \times \hat{n}_T\rvert}{2 \sum_b \lvert\vec{p}_b\rvert^2}$
Wide/narrow, total broadening	$B_{W/N} = \max / \min(B_1, B_2)$, $B_T = B_W + B_N$
Fox-Wolfram moments	$H_\ell = \sum_{i,j} \frac{\lvert\vec{p}_i\rvert \lvert\vec{p}_j\rvert}{E^2} P_\ell(\cos\theta_{ij})$
N-jettiness	$\tau_N = \frac{2}{Q^2} \sum_k \min\{q_a \cdot p_k, q_b \cdot p_k, q_1 \cdot p_k, \cdots, q_N \cdot p_k\}$
N-subjettiness	$\tau_N = \frac{1}{\sum_k p_{T,k} R_0} \sum_k p_{T,k} \min\{\Delta R_{1k}, \Delta R_{2k}, \cdots, \Delta R_{Nk}\}$
Energy-energy correlation	$EEC(\chi) = \frac{1}{\sigma} \frac{d\Sigma}{d\cos\chi} = \sum_{i,j} \int \frac{E_i E_j}{Q^2} \delta(\hat{p}_i \cdot \hat{p}_j - \cos\chi) d\sigma$

[Inspired by Fabio Maltoni's lecture slide at the 2013 CERN - Latin-American School of High Energy Physics]

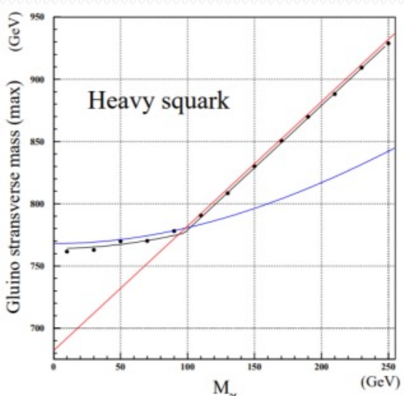
Exclusive Mass Variables



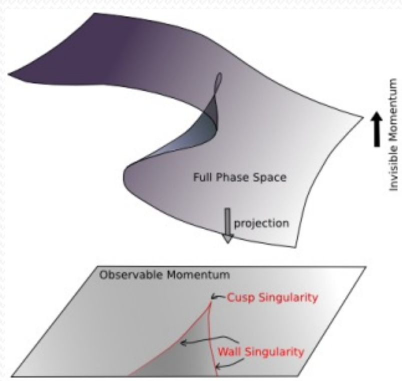
Topology disambiguation with invariant masses [Cho, **DK**, Matchev, Park, arXiv:1206.1546]



Transverse mass distribution of leptonic W [CDF Collaboration, Science 376 (2022) 6589]



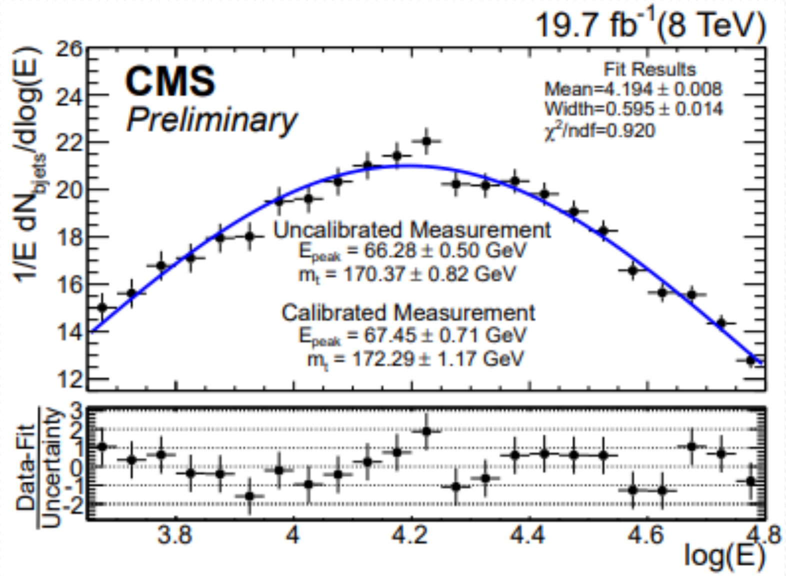
M_{T2} kink [Cho, Choi, Kim, Park, arXiv:0709.0288]



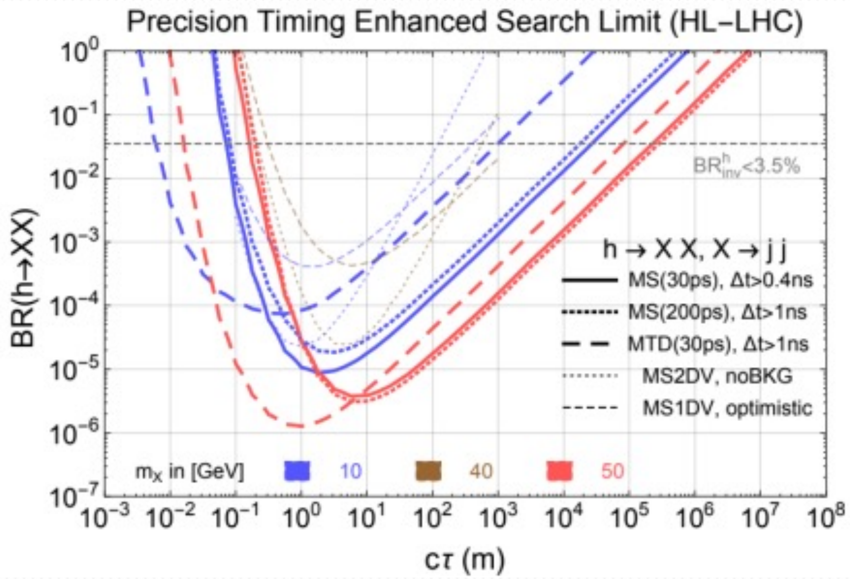
Algebraic singularity [I.-W. Kim, arXiv:0910.1149]

- Theoretical/phenomenological developments mostly motivated by missing energy events

Exclusive Variables: Energy, Time, Distance



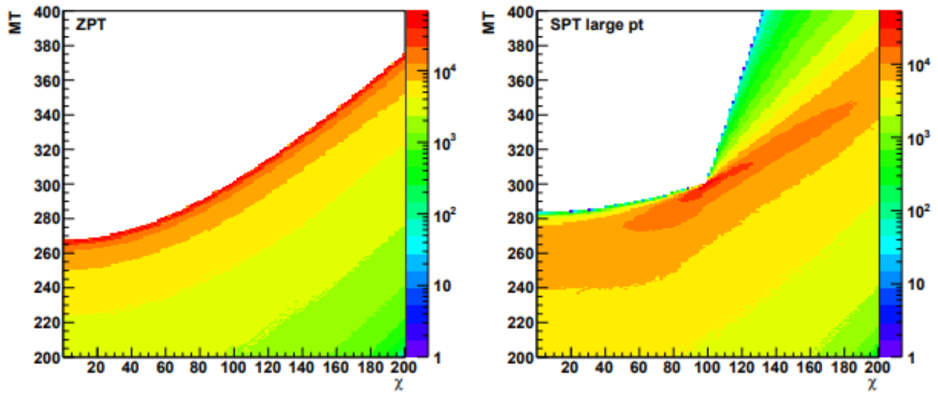
Top quark mass measurement using the energy-peak method [CMS Collaboration, CMS-PAS-TOP-15-002]



Long-lived particle searches with timing information [Liu, Lui, Wang, arXiv:1805.05957]

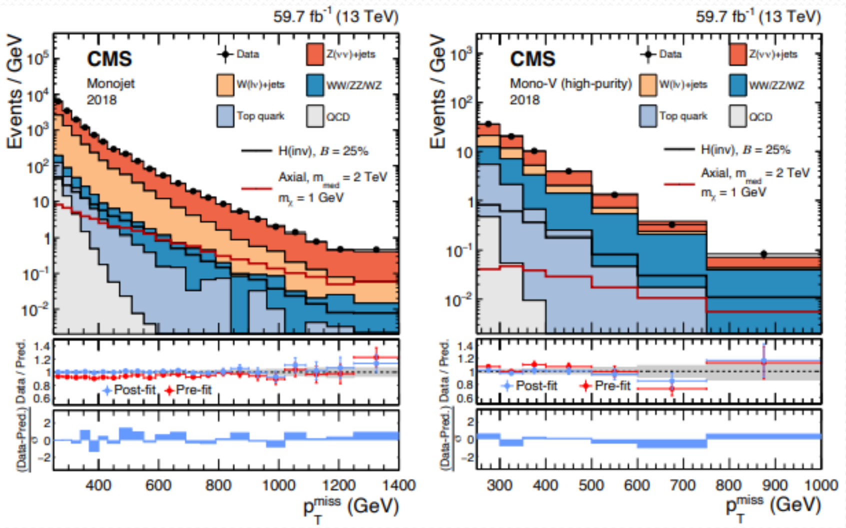
- Re-discovering the implication of the energy peak at hardon colliders
- Long-lived particle searches with timing and distance variables

Other Exclusive Variables



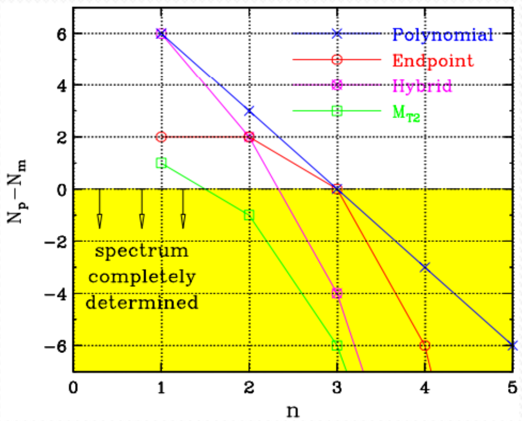
Kink structure in the transverse variable of single resonance events with ISR [Barr, Gripaios Lester, arXiv:0711.4008]

- ISR (sometimes) helps both the discovery of new phenomena and particle mass measurements.



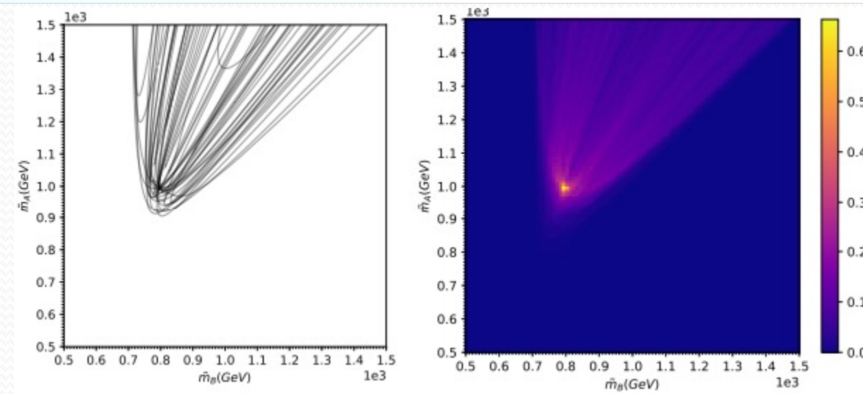
Dark matter search in the mono-X + MET channels [CMS Collaboration, arXiv:2107.13021]

Variables/Methods Using Ensembles of Events



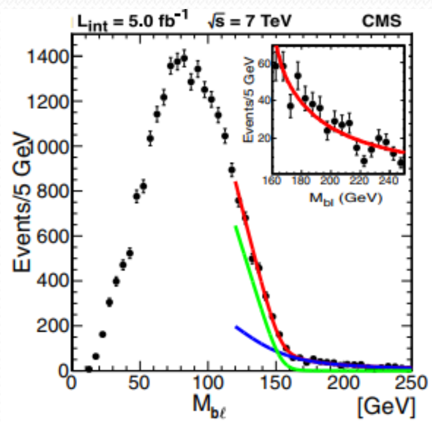
Dependence of # of undetermined parameters $N_p - N_m$ in # of intermediate resonances n

[Burns, Kong, Matchev, Park, arXiv:0810.5576]



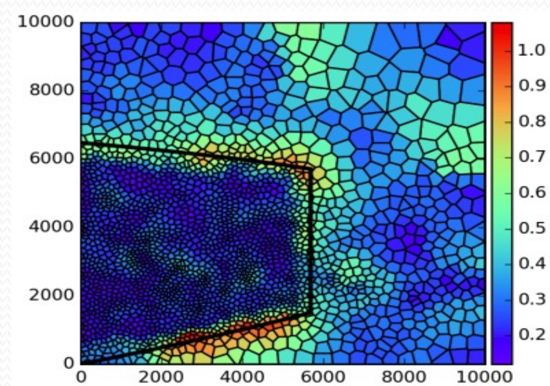
Particle mass measurement with the focus point method

[**DK**, Matchev, Shyamsundar, arXiv:1906.02821]



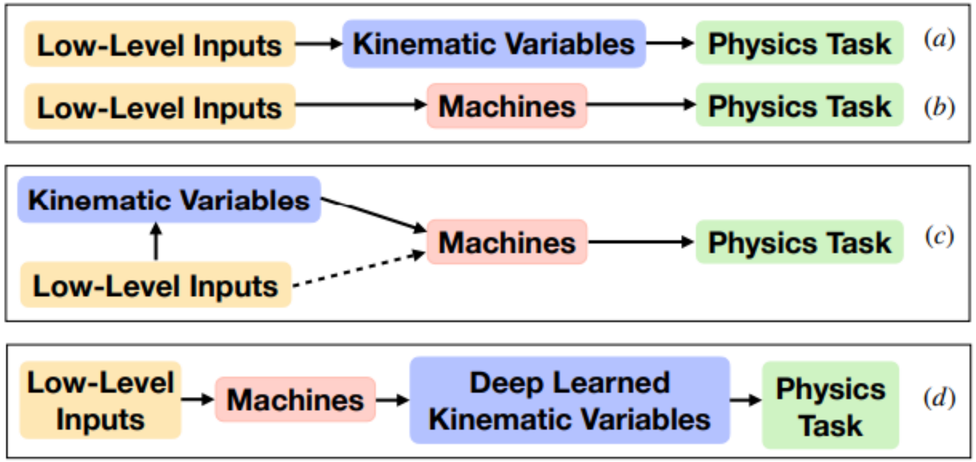
Top quark mass measurement with kinematic endpoints

[CMS Collaboration, arXiv:1304.5783]



Kinematic edge detection [Debnath, Gainer, **DK**, Matchev, arXiv:1506.04141]

Kinematic Variables (KVs) in the Machine Learning (ML) Era



Traditional approach only with KVs

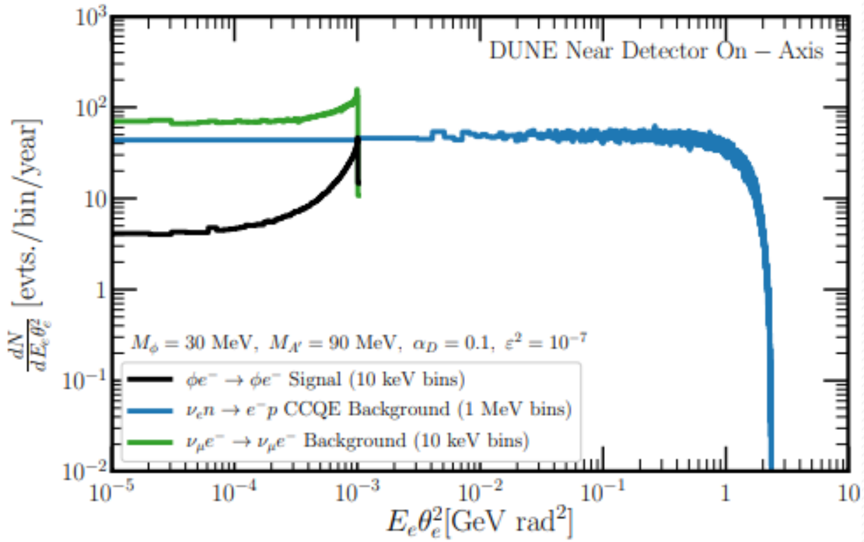
ML-based approaches with/without the aid of KVs

Construction of artificial KVs

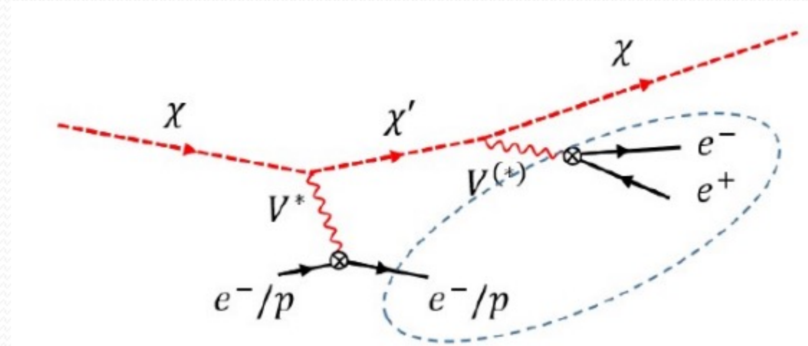
"Blackboxness" of ML-based techniques and kinematic variables

- To interpret/explain the ML blackbox ⇒ KVs for interpreting, explaining, and understanding the decisions made by ML [Chang, Cohen, Ostdiek, arXiv:1709.10106; Faucett, Thaler, Whiteson, arXiv:2010.11998; Agarwal et al, arXiv:2011.13466; Grojean, Paul, Qian, arXiv: 2011.13945]
- To try the machine less of a blackbox ⇒ Constructing KV-inspired neural architectures [Komiske, Metodiev, Thaler, arXiv:1810.05165; Erdmann, Geiser, Rath, Rieger, arXiv:1812.09722]
- To design robust ML-based analysis techniques despite blackboxness ⇒ Constructing ML-based event variables [**DK**, Kong, Matchev, Park, Shaymsundar, arXiv:2105.10126]

Kinematic Variables at Non-Collider Experiments



Elastic low-mass dark matter signal separation from CCQE-induced backgrounds at DUNE [de Romeri, Kelly, Machado, arXiv:1903.10505]



Cosmogenic inelastic boosted dark matter at large-volume neutrino detectors [DK, Park, Shin, arXiv:1612.06867]

Kinematic variables designed for collider phenomenology can be readily applied to non-collider experiments (e.g., neutrino experiments)

- Experiment-wise ⇒ high-capability detectors
- Theory-wise ⇒ Models predicting multi-particle states (e.g., non-minimal dark-sector scenarios)

Conclusions

① Kinematic variables have been playing an important role in collider phenomenology, as they expedite discoveries of new particles by separating signal events from unwanted background events and allow for measurements of particle properties such as masses, couplings, spins, et ② For the past 10 years, an enormous number of kinematic variables have been designed and proposed, primarily for the experiments at the Large Hadron Collider, allowing for a drastic reduction of high-dimensional experimental data to lower-dimensional observables, from which one can readily extract underlying features of phase space and develop better-optimized data-analysis strategie ③ We review these recent developments in the area of phase space kinematics, summarizing the new kinematic variables with important phenomenological implications and physics applications. We also review recently proposed analysis methods and techniques specifically designed to leverage the new kinematic variables. As machine learning is nowadays percolating through many fields of particle physics including collider phenomenolog ④ we discuss the interconnection and mutual complementarity of kinematic variables and machine learning techniques ⑤ Ve finally discuss how the utilization of kinematic variables originally developed for colliders can be extended to other high-energy physics experiments including neutrino experiments.

Discussion

Future directions of kinematic variables

1) Applicability of kinematic variables for the non-collider experiments, especially fixed target experiments and neutrino experiments
2) Investigation of interpretability and reliability of ML-based techniques using kinematic variables