

Artificial Intelligence and High-Energy Physics

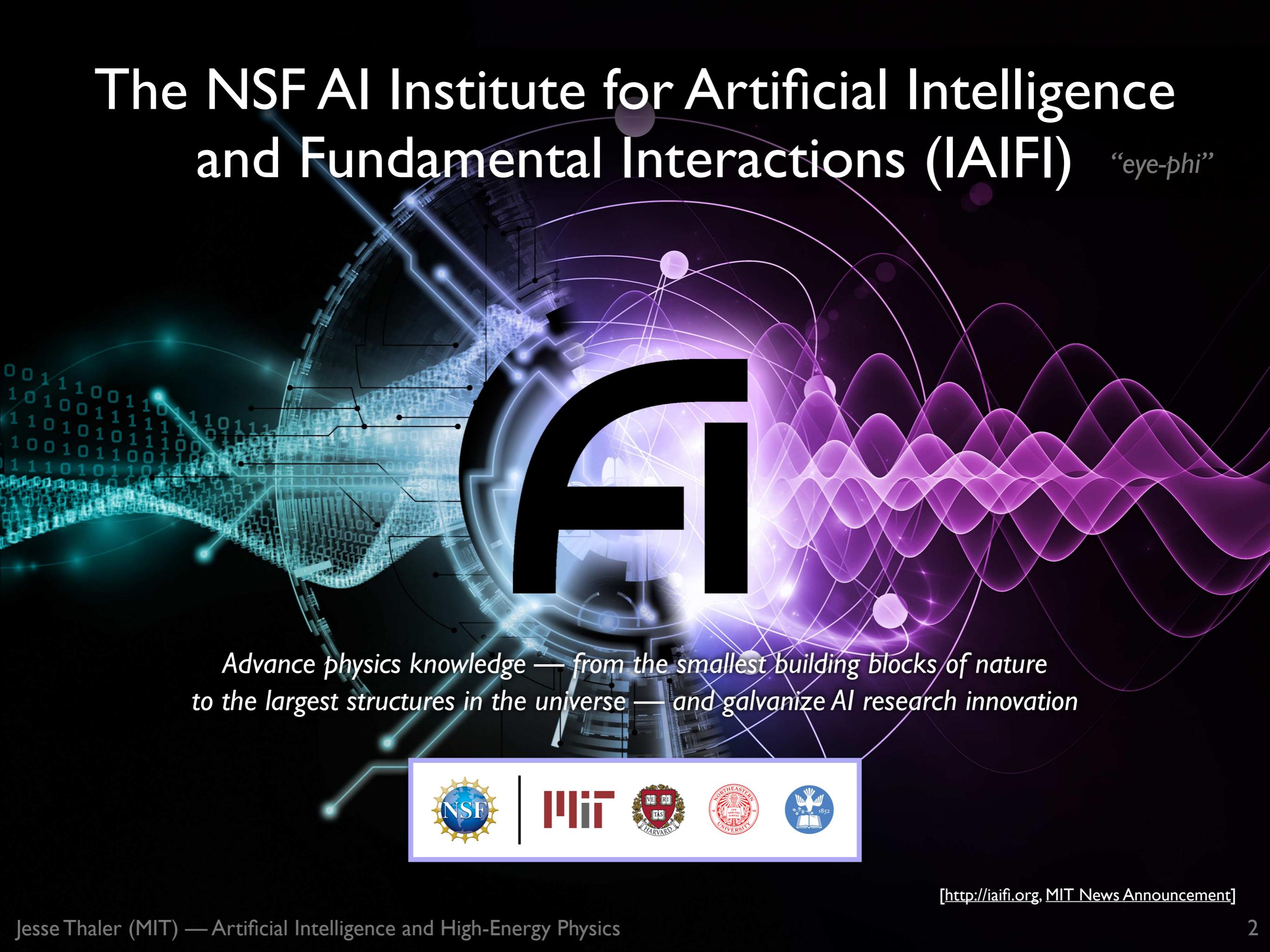
Jesse Thaler



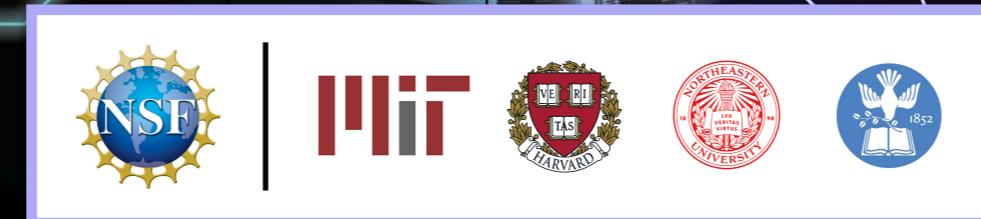
APS April Meeting, H02: Recent Advances in Theoretical Physics — April 18, 2021

The NSF AI Institute for Artificial Intelligence and Fundamental Interactions (IAIFI)

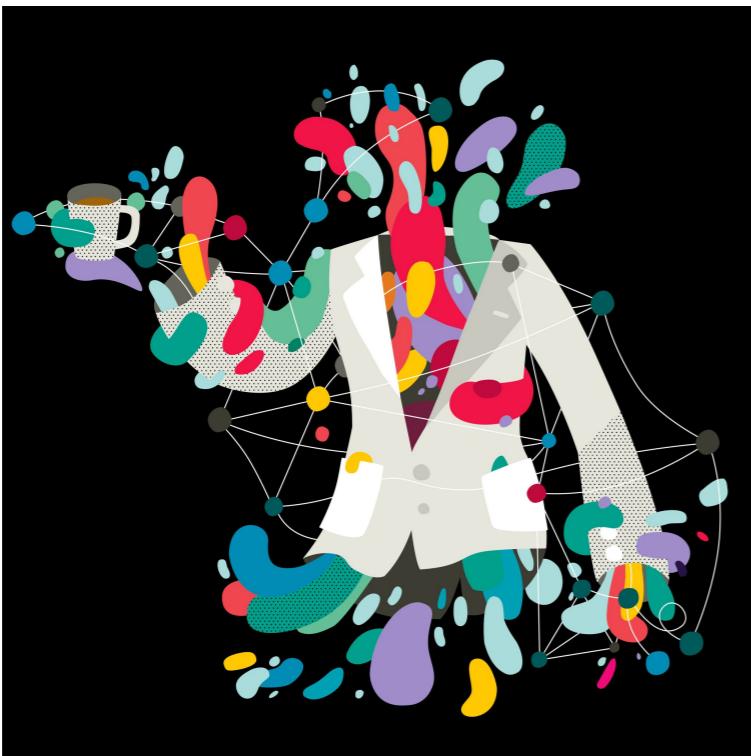
“eye-phi”



*Advance physics knowledge — from the smallest building blocks of nature
to the largest structures in the universe — and galvanize AI research innovation*



[<http://iaifi.org>, MIT News Announcement]



*Can we teach a machine
to “think” like a physicist?*

The New York Times



By Dennis Overbye

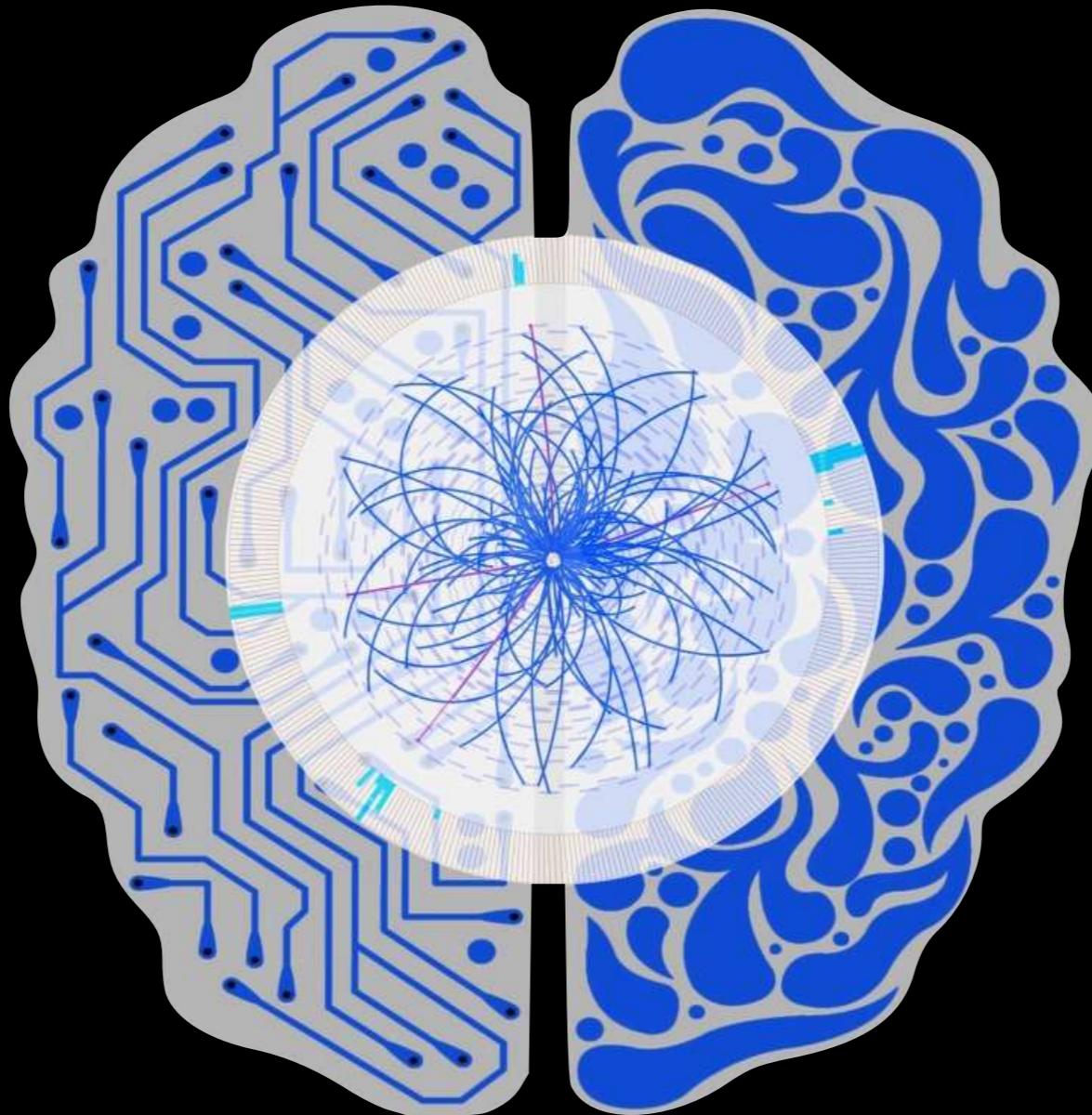
Nov. 23, 2020

Can a Computer Devise a Theory of Everything?

Deep Learning meets “Deep Thinking”

*Machine learning that incorporates
first principles, best practices, and domain knowledge
from fundamental physics*

The Lens of Machine Learning



What formalisms are needed to leverage ML for HEP?

Likelihood Ratio Trick

Many HEP problems can be expressed in this form!

Key example of *simulation-based inference*

Goal: Estimate $p(x)$ / $q(x)$

Training Data: Finite samples P and Q

Learnable Function: $f(x)$ parametrized by, e.g., neural networks

Loss Function(al): $L = -\langle \log f(x) \rangle_P + \langle f(x) - 1 \rangle_Q$

Asymptotically: $\arg \min L = \frac{p(x)}{q(x)}$ *Likelihood ratio*

$-\min_{f(x)} L = \int dx p(x) \log \frac{p(x)}{q(x)}$ *Kullback–Leibler divergence*

[see e.g. Cranmer, Pavez, Loupe, [arXiv 2015](#); D’Agnolo, Wulzer, [PRD 2019](#); simulation-based inference in Cranmer, Brehmer, Loupe, [PNAS 2020](#); relation to f-divergences in Nguyen, Wainwright, Jordan, [AoS 2009](#); Nachman, Thaler, [arXiv 2021](#)]

Likelihood Ratio Trick

Many HEP problems can be expressed in this form!

Key example of *simulation-based inference*

Asymptotically, same structure as *Lagrangian mechanics!*

Action: $L = \int dx \mathcal{L}(x)$

Lagrangian: $\mathcal{L}(x) = -p(x) \log f(x) + q(x)(f(x) - 1)$

Euler-Lagrange: $\frac{\partial \mathcal{L}}{\partial f} = 0$

Solution: $f(x) = \frac{p(x)}{q(x)}$

Requires shift in theoretical focus from solving problems to *specifying problems*

[see e.g. Cranmer, Pavez, Louppe, [arXiv 2015](#); D'Agnolo, Wulzer, [PRD 2019](#); simulation-based inference in Cranmer, Brehmer, Louppe, [PNAS 2020](#); relation to f-divergences in Nguyen, Wainwright, Jordan, [AoS 2009](#); Nachman, Thaler, [arXiv 2021](#)]

Machine Learning Requirements

If you have in hand...

Well-specified loss
Reliable training data
Learnable function

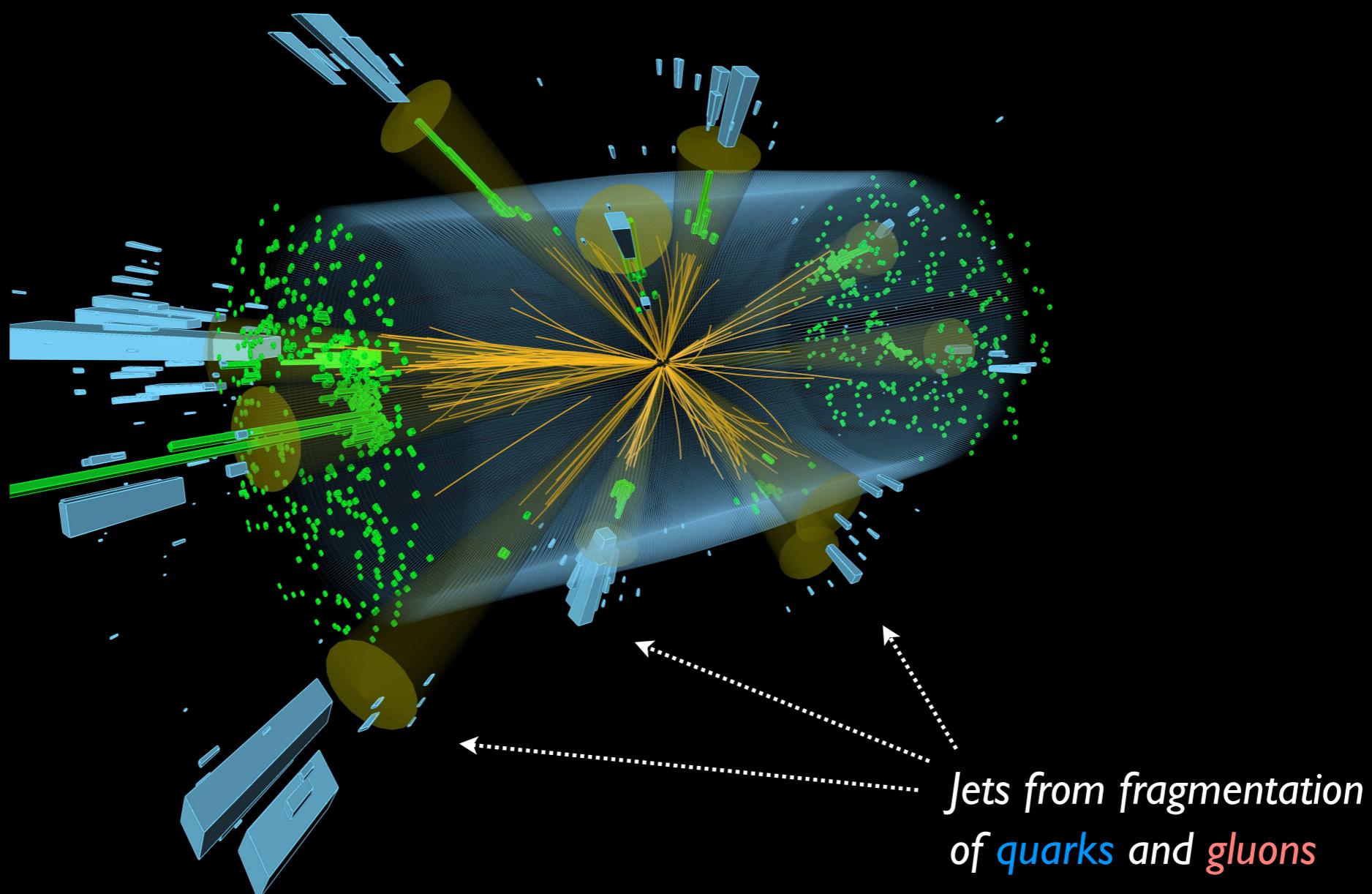
...then you can leverage ML!

Many HEP tasks can be phrased in this language

Theory input essential for robust usage of these tools

[see [HEPML-LivingReview](#) for extensive bibliography]

Machine Learning for High-Energy Physics

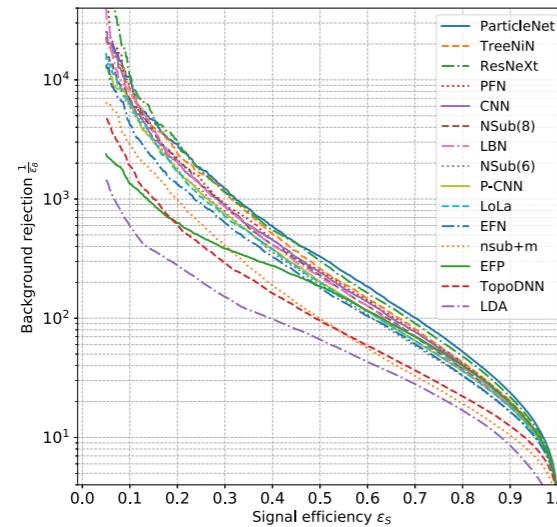


What tasks are amendable to a machine learned approach?

Optimization in Collider Physics

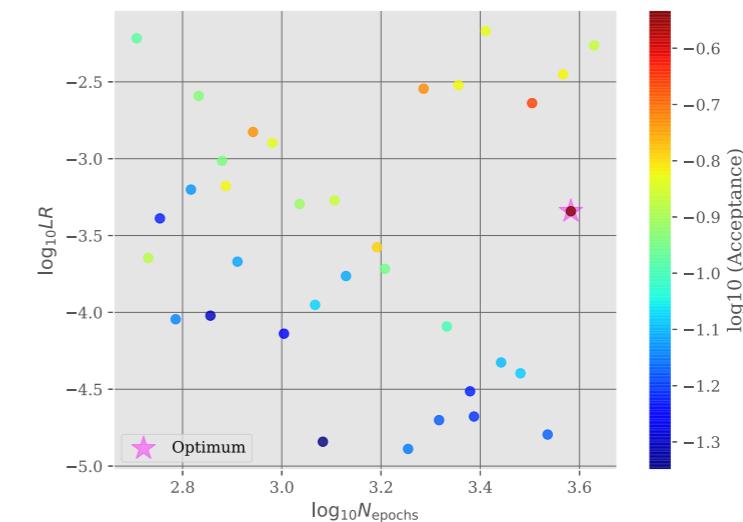
This slide is far from exhaustive

Jet Classification



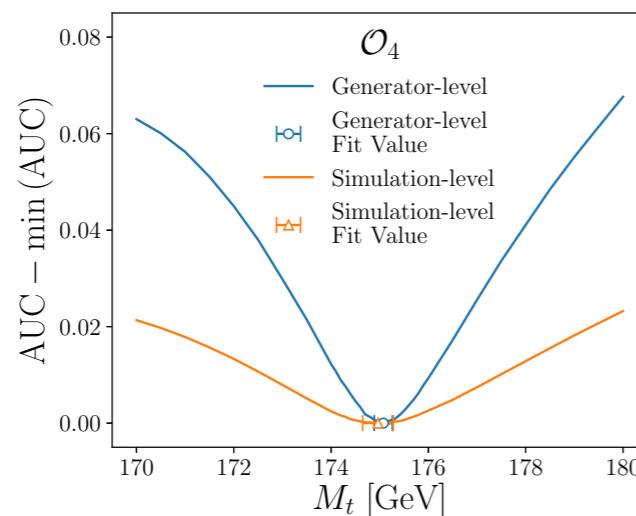
[e.g. Kasieczka, Plehn, et al., [SciPost 2019](#)]

Phase Space Integration



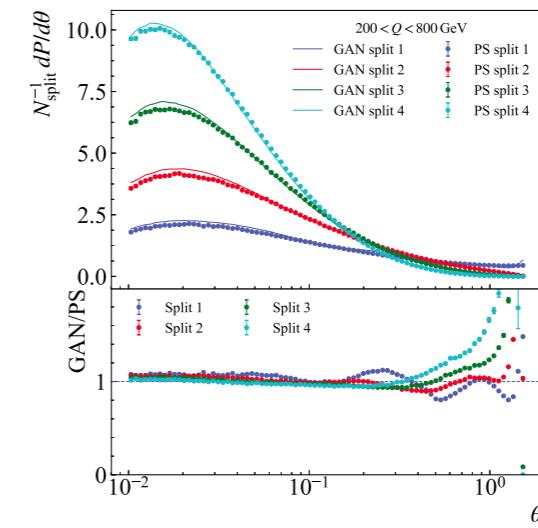
[e.g. Gao, Höche, Isaacson, Krause, Schulz, [PRD 2020](#)]

Parameter Estimation



[e.g. Andreassen, Hsu, Nachman, Suaysom, Suresh, [PRD 2021](#)]

Parton Shower Modeling

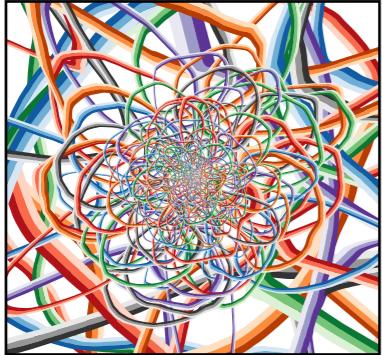


[e.g. Lai, Neill, Płoskoń, Ringer, [arXiv 2020](#)]

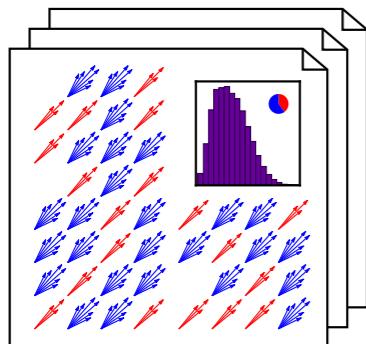
From Curmudgeon to Evangelist



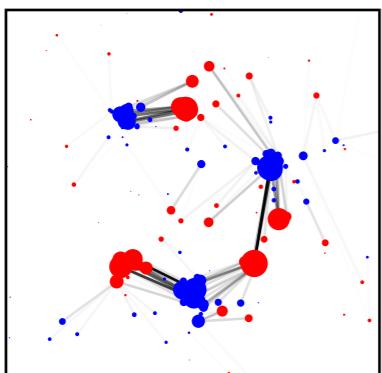
What have been helpful guides in pursuing ML \leftrightarrow HEP?



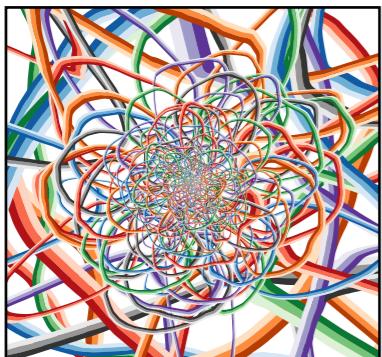
Can theoretical structures be encoded directly?



Can strategy be defined on physical quantities?



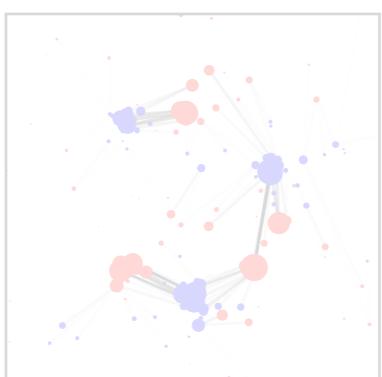
Can we leverage unsupervised machine learning?



Can theoretical structures be encoded directly?



Can strategy be defined on physical quantities?

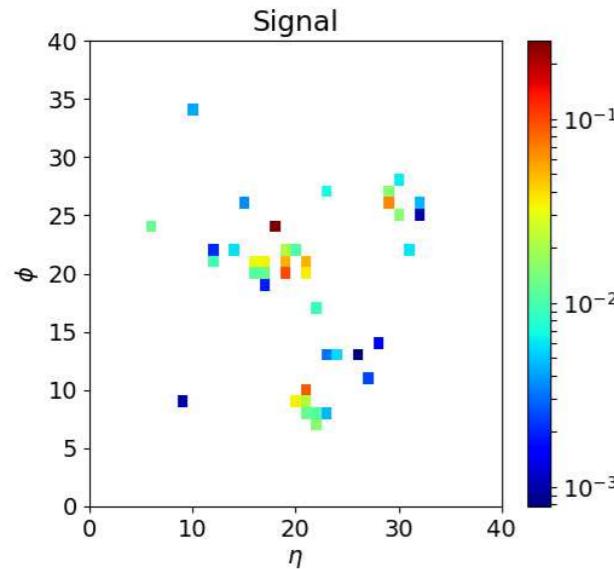


Can we leverage unsupervised machine learning?

Jet Representations

Pixelized Image

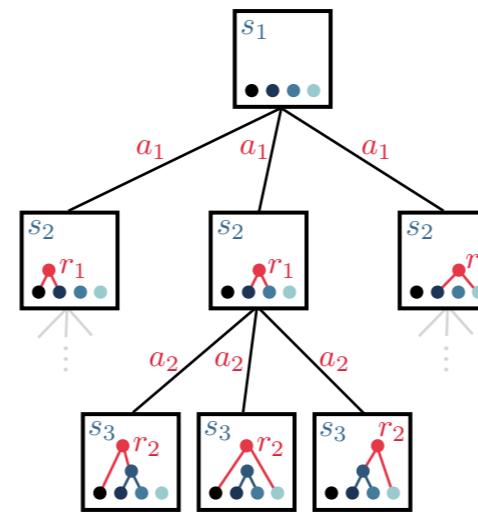
Calorimetry



[review in Kagan, [arXiv 2020](#)]

Hierarchical Tree

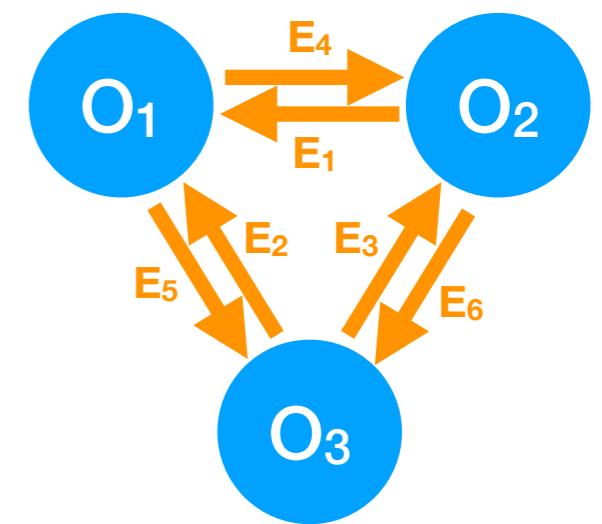
Binary Splittings



[e.g. Brehmer, Macaluso, Pappadopulo, Cranmer, [NeurIPS 2020](#)]

Graphs

Pairwise Interactions

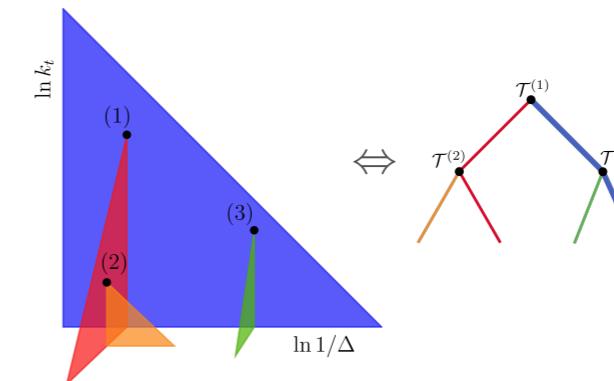


[e.g. Moreno, Cerri, Duarte, Newman, Nguyen, Periwal, Pierini, Serikova, Spiropulu, Vlimant, [EPJC 2020](#)]

Imposes implicit theoretical prior; affects choice of network architecture

E.g. recent progress with Lund Plane + Graph Networks

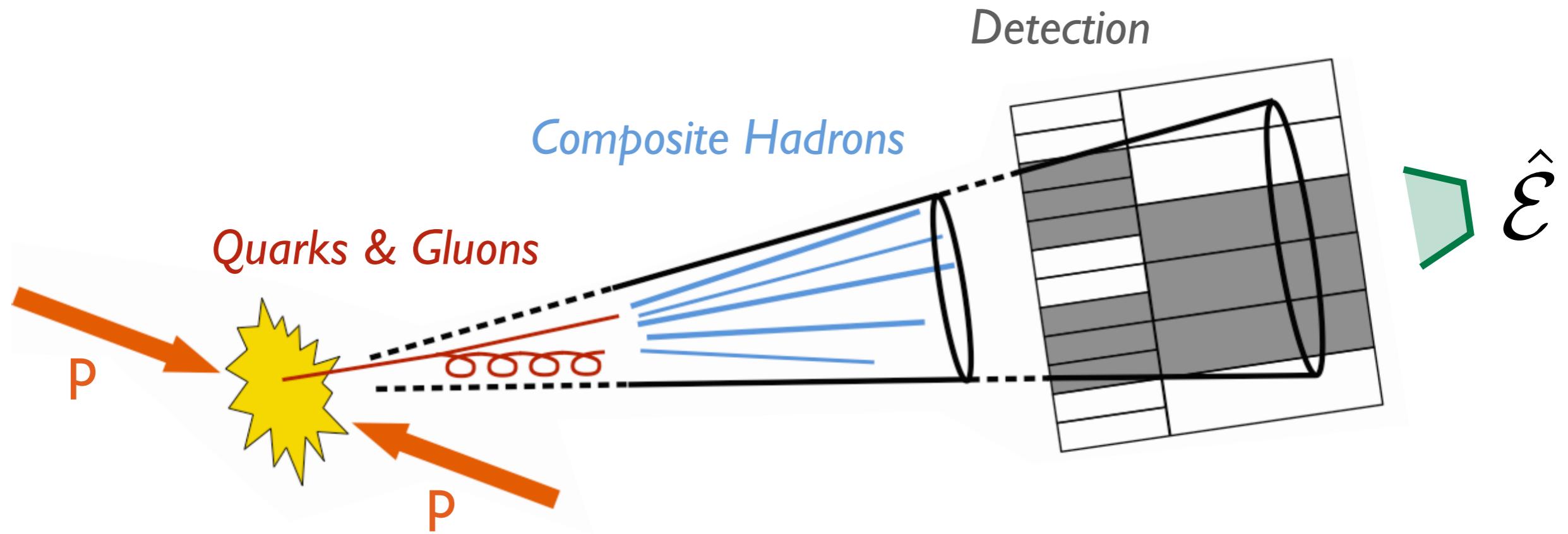
[Dreyer, Qu, [JHEP 2021](#)]



Energy Flow Representation

Emphasizes *infrared and collinear safety*

Theory



Energy Flow:

Robust to hadronization and detector effects
Well-defined for massless gauge theories

$$\hat{\mathcal{E}} \simeq \lim_{t \rightarrow \infty} \hat{n}_i T^{0i}(t, vt\hat{n})$$

[see e.g. Sveshnikov, Tkachov, [PLB 1996](#); Hofman, Maldacena, [JHEP 2008](#); Mateu, Stewart, [JDT, PRD 2013](#); Belitsky, Hohenegger, Korchemsky, Sokatchev, Zhiboedov, [PRL 2014](#); Chen, Moult, Zhang, Zhu, [PRD 2020](#)]
[complementary perspective on IRC unsafe information in Chakraborty, Lim, Nojiri, Takeuchi, [JHEP 2020](#)]

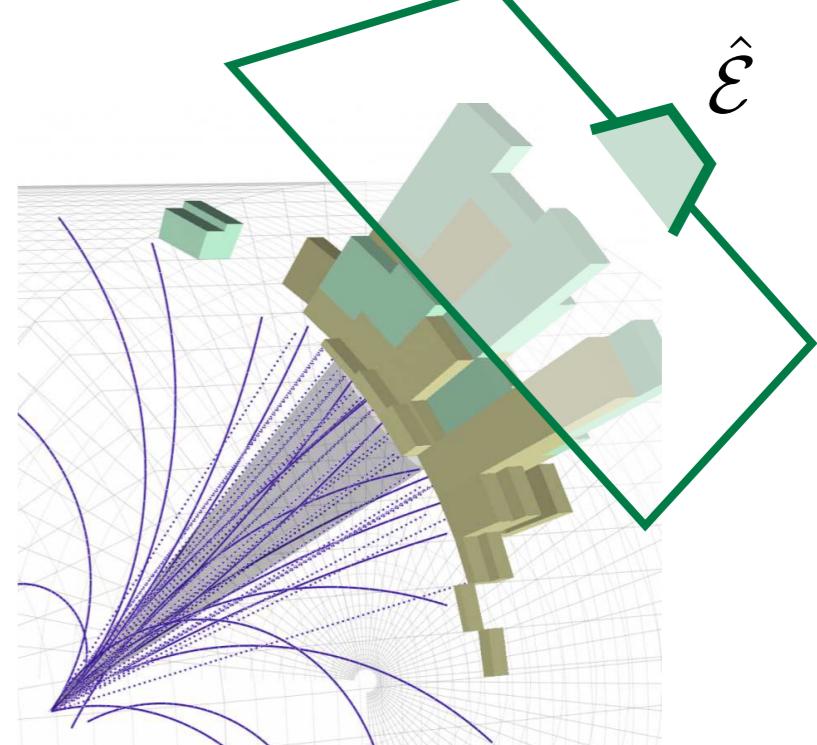
Jets as Weighted Point Clouds

- Energy-Weighted Directions

$$\vec{p} = \{E, \hat{n}_x, \hat{n}_y, \hat{n}_z\}$$

↑ |
Energy Direction

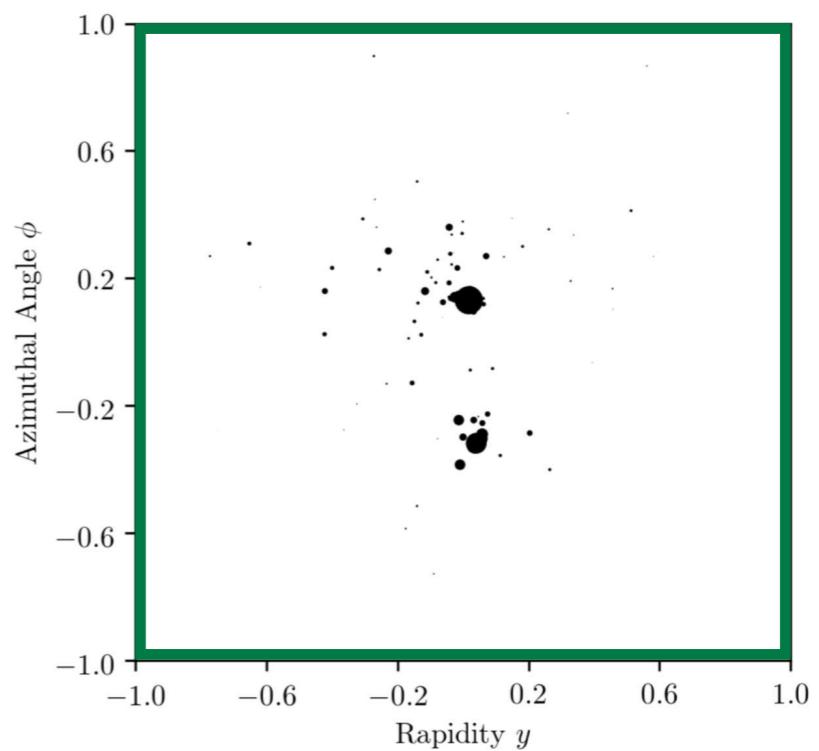
(suppressing “unsafe” charge/flavor information)



- Equivalently: Energy Density

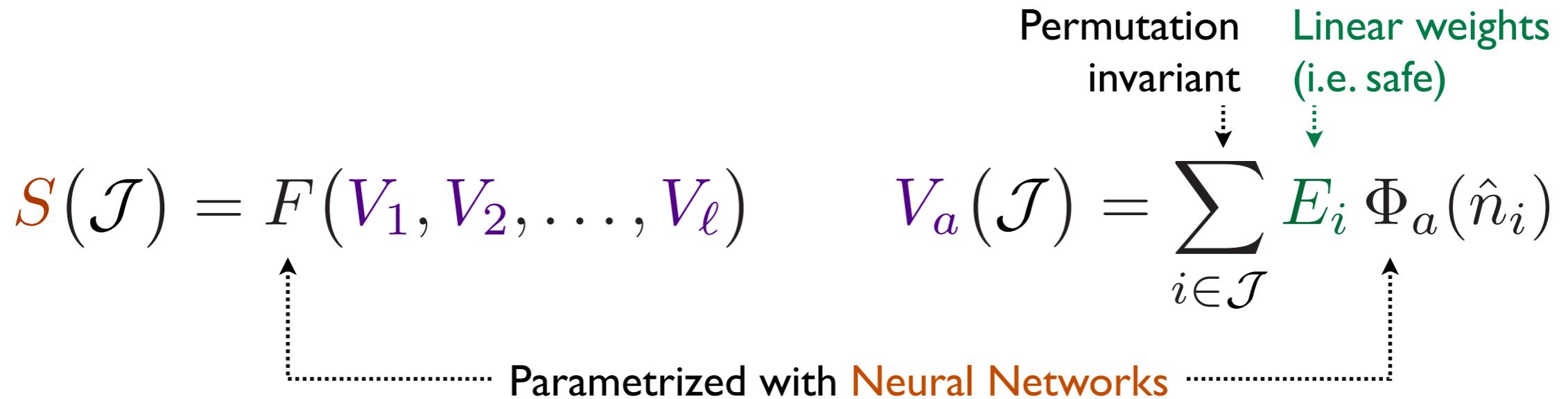
$$\rho(\hat{n}) = \sum_{i \in \mathcal{J}} E_i \delta^{(2)}(\hat{n} - \hat{n}_i)$$

↑ ↑
Energy Direction



Energy Flow Networks

Architecture designed around **symmetries** and **interpretability**



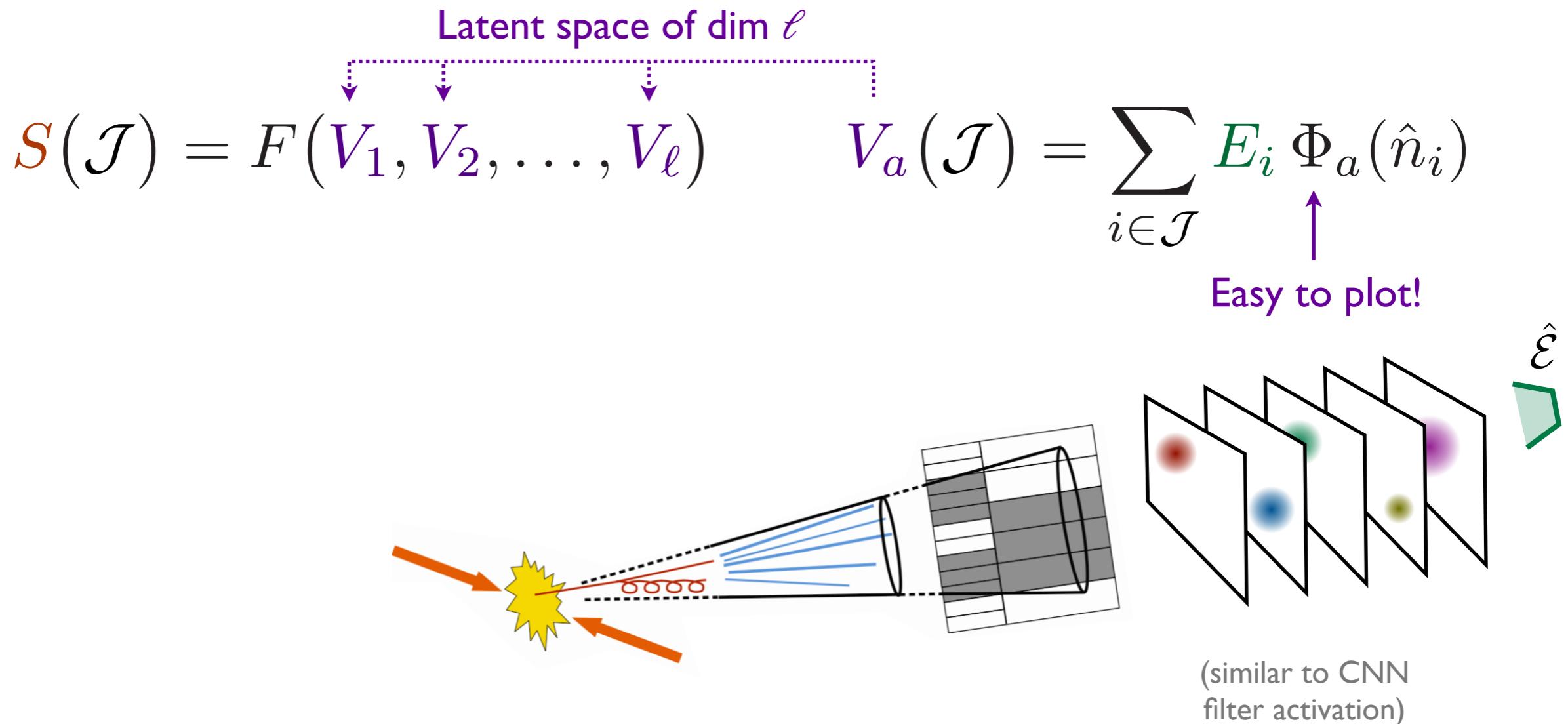
Provably describes any **safe** observable (!)
Excellent jet classification performance*

[Komiske, Metodiev, JDT, [JHEP 2019](#); see also Komiske, Metodiev, JDT, [JHEP 2018](#); code at [energyflow.network](#); special case of Zaheer, Kottur, Ravanbakhsh, Poczos, Salakhutdinov, Smola, [NIPS 2017](#); other set-based architecture in Qu, Gouskos, [PRD 2020](#); Mikuni, Canelli, [EPJP 2020](#); Dolan, Ore, [arXiv 2020](#); Lorentz-equivariant approach in Bogatskiy, Anderson, Offermann, Roussi, Miller, Kondor, [arXiv 2020](#)]



Energy Flow Networks

Architecture designed around symmetries and *interpretability*

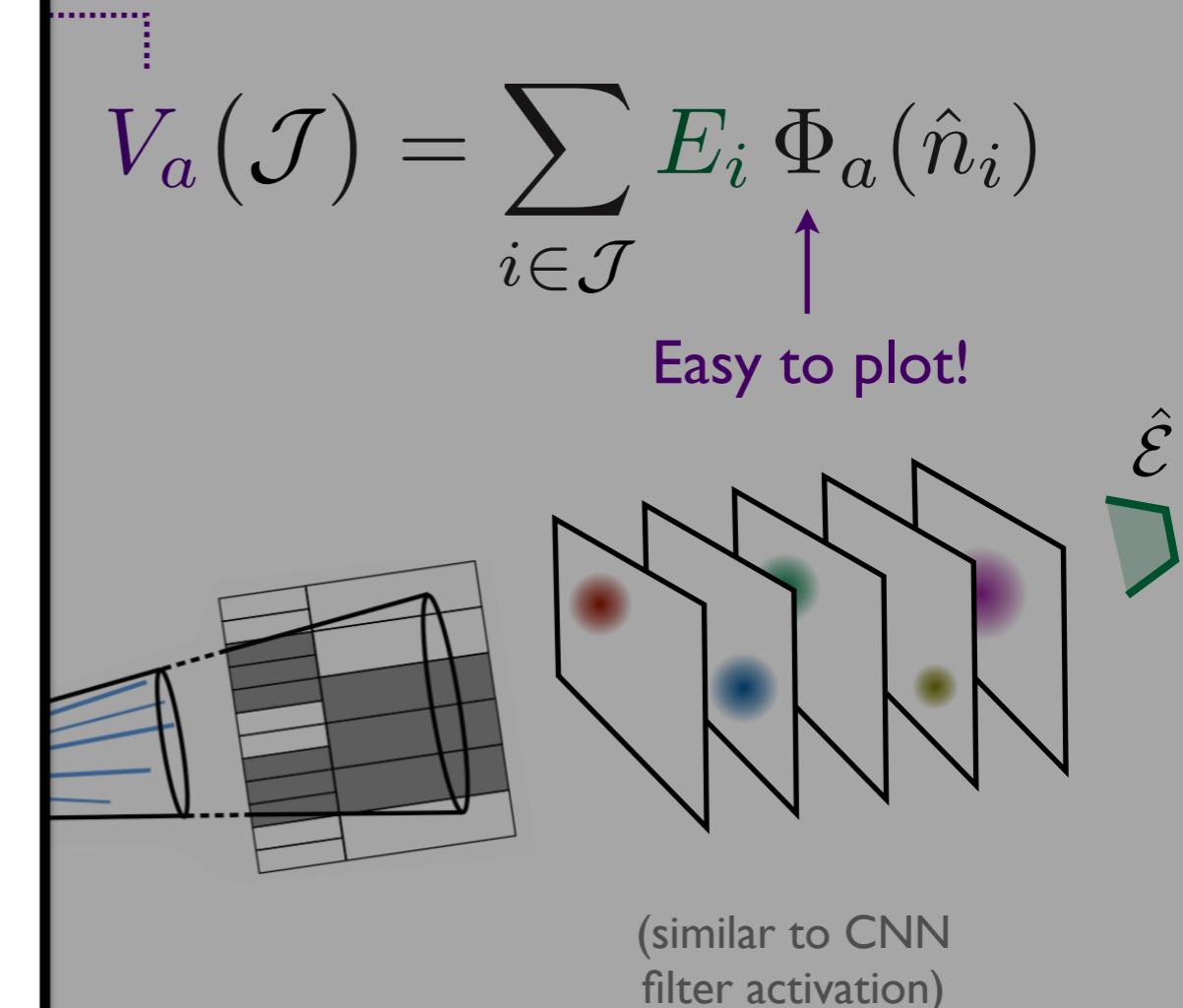
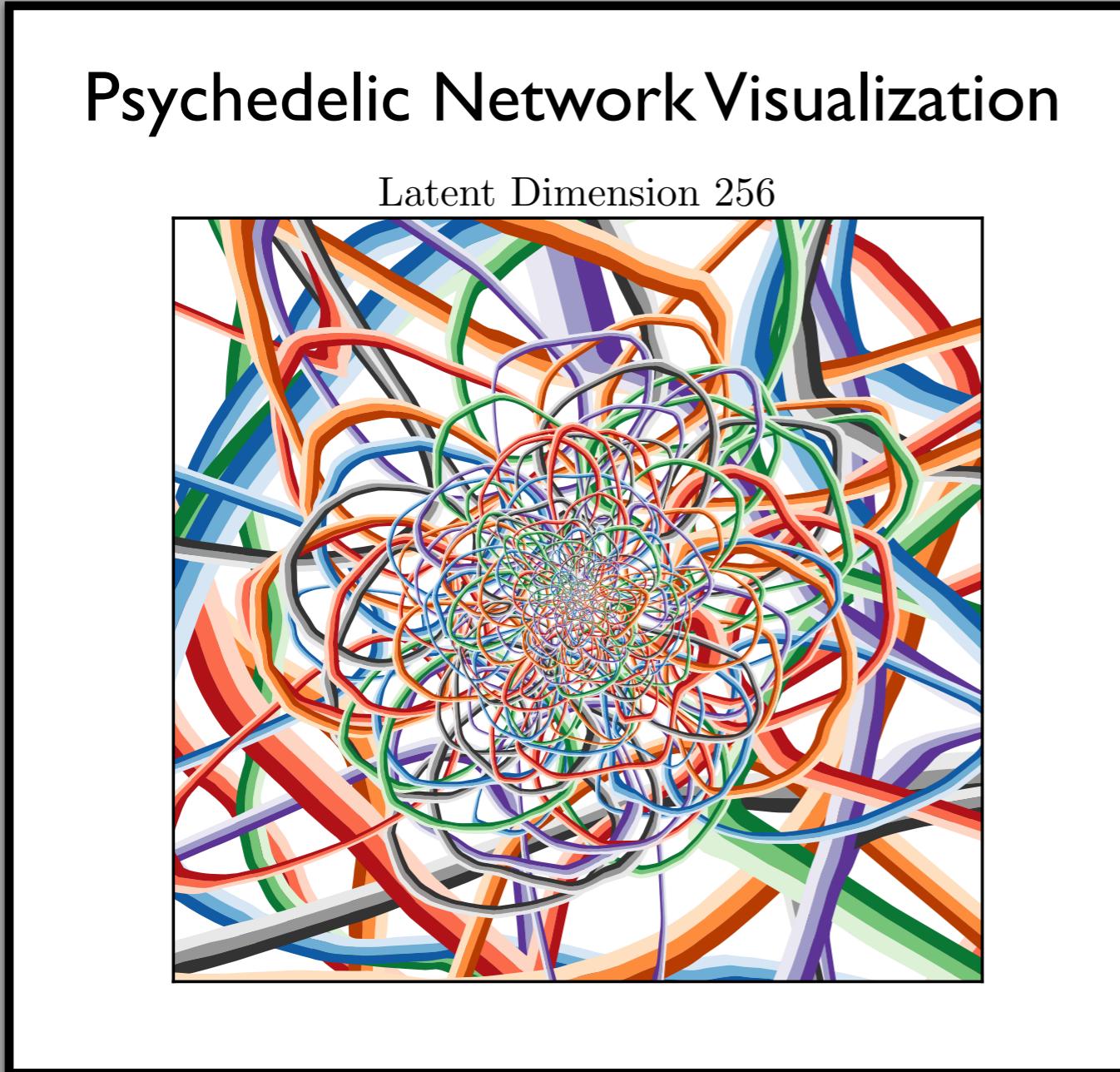


[Komiske, Metodiev, JDT, *JHEP* 2019; see also Komiske, Metodiev, JDT, *JHEP* 2018; code at energyflow.network; special case of Zaheer, Kottur, Ravanbakhsh, Poczos, Salakhutdinov, Smola, *NIPS* 2017; other set-based architecture in Qu, Gouskos, *PRD* 2020; Mikuni, Canelli, *EPJP* 2020; Dolan, Ore, *arXiv* 2020; Lorentz-equivariant approach in Bogatskiy, Anderson, Offermann, Roussi, Miller, Kondor, *arXiv* 2020]



Energy Flow Networks

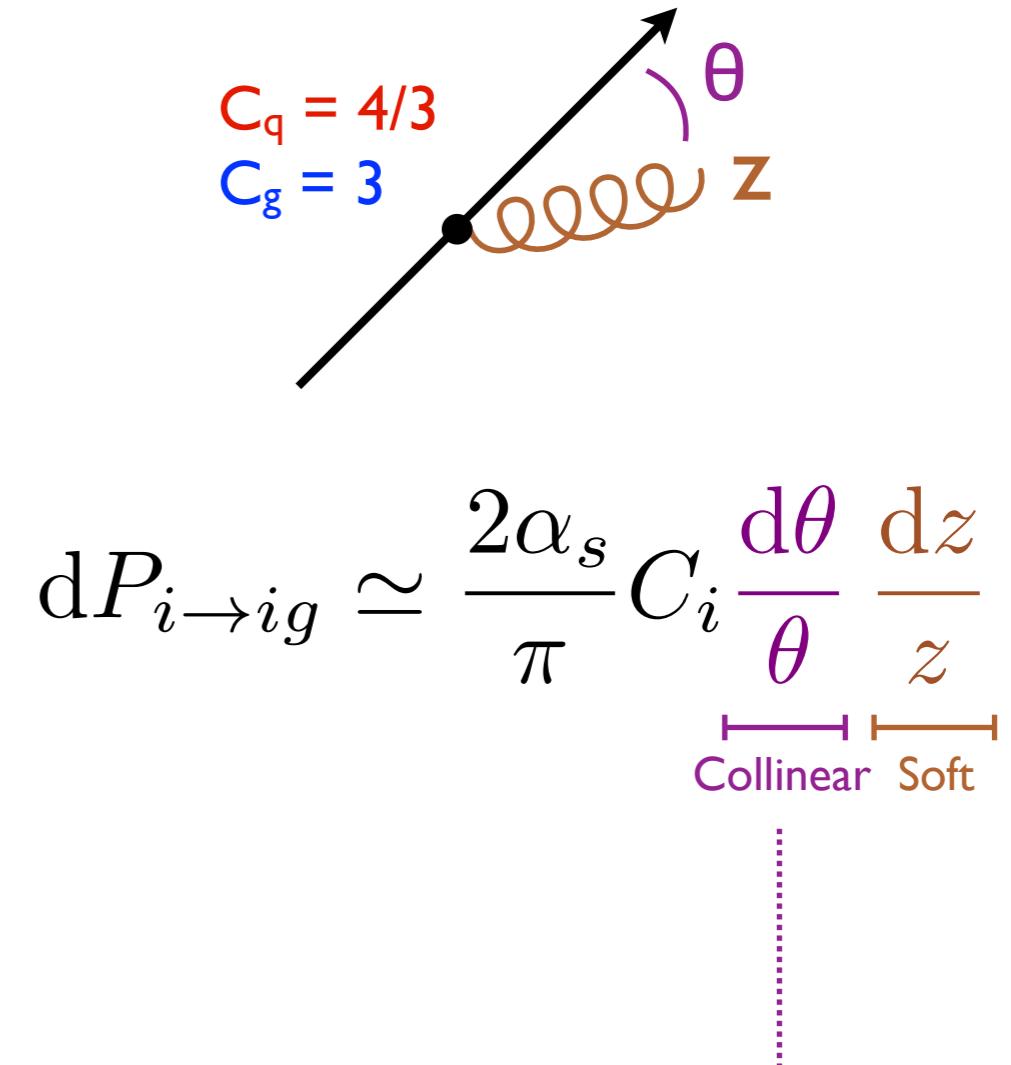
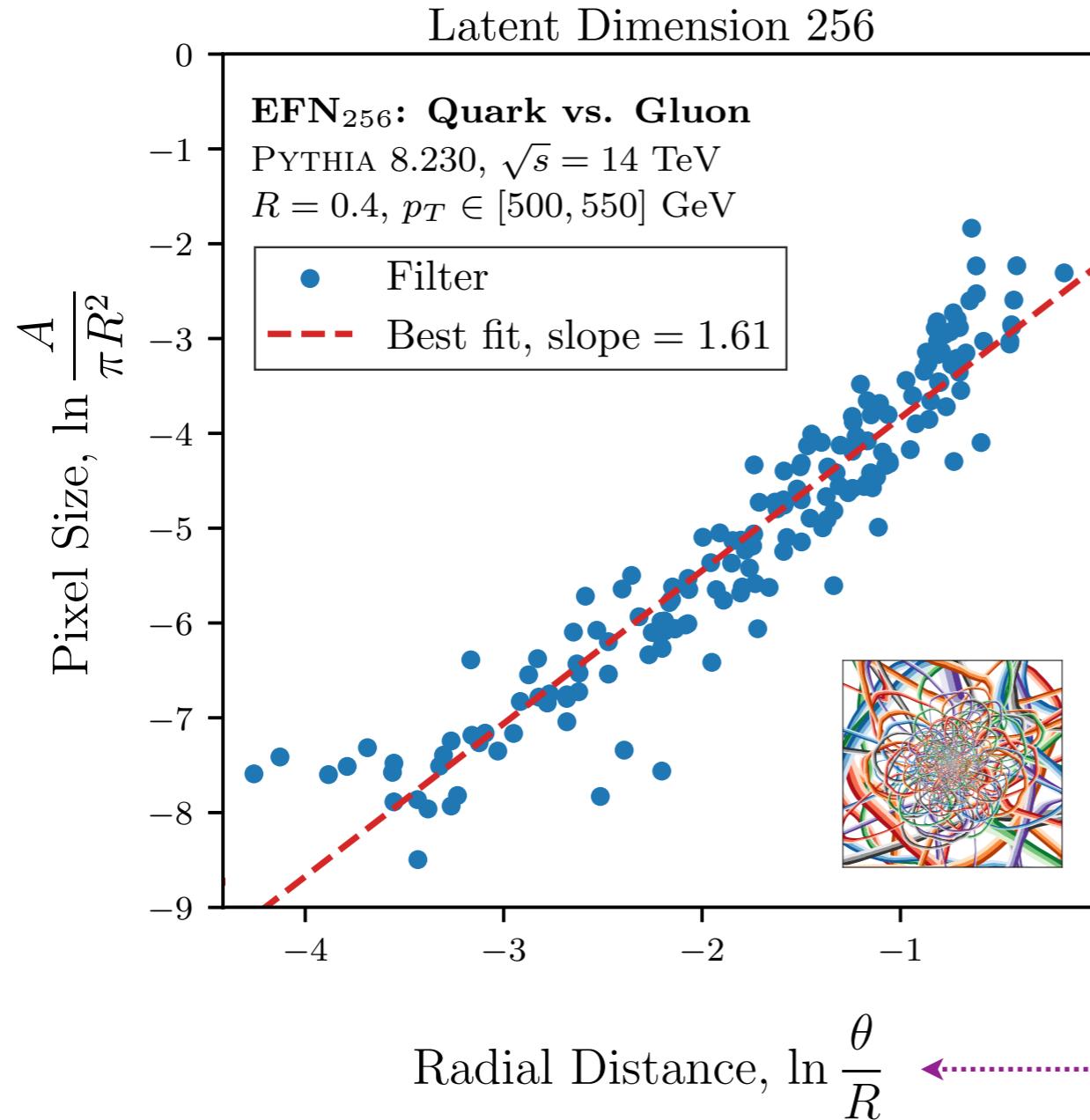
Architecture designed around symmetries and *interpretability*



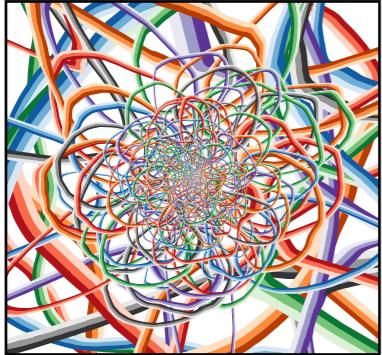
[Komiske, Metodiev, JDT, *JHEP* 2019; see also Komiske, Metodiev, JDT, *JHEP* 2018; code at energyflow.network; special case of Zaheer, Kottur, Ravanbakhsh, Poczos, Salakhutdinov, Smola, *NIPS* 2017; other set-based architecture in Qu, Gouskos, *PRD* 2020; Mikuni, Canelli, *EPJP* 2020; Dolan, Ore, *arXiv* 2020; Lorentz-equivariant approach in Bogatskiy, Anderson, Offermann, Roussi, Miller, Kondor, *arXiv* 2020]



Machine Learning Collinear QCD

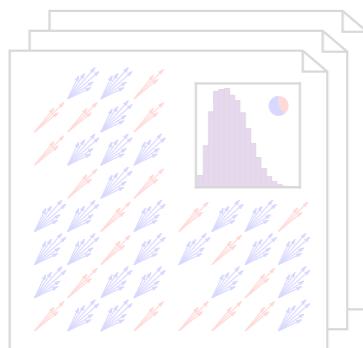


[Komiske, Metodiev, JDT, JHEP 2019]

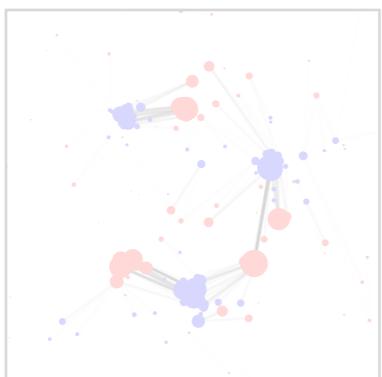


Can theoretical structures be encoded directly?

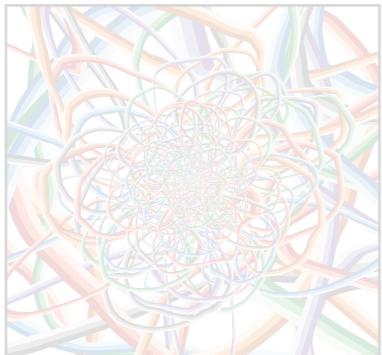
Energy Flow Networks \Leftrightarrow IRC Safety + Permutations



Can strategy be defined on physical quantities?

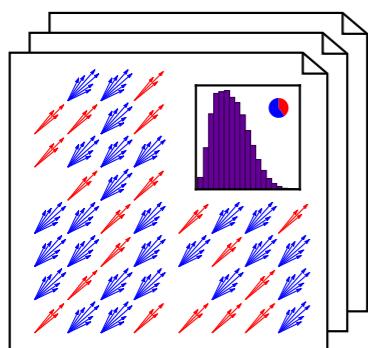


Can we leverage unsupervised machine learning?

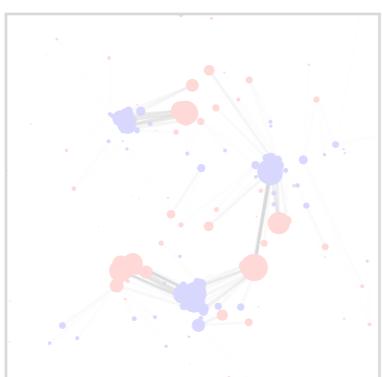


Can theoretical structures be encoded directly?

Energy Flow Networks \leftrightarrow IRC Safety + Permutations



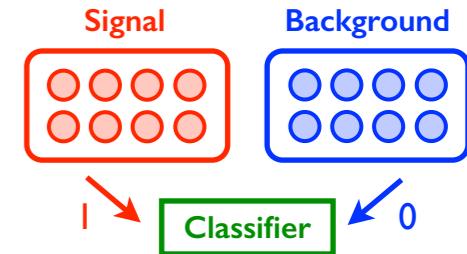
Can strategy be defined on physical quantities?



Can we leverage unsupervised machine learning?

Quark/Gluon Classification

“Hello, World!” of Jet Physics



Find $h\left(\begin{array}{c} \nearrow \\ \nearrow \\ \nearrow \\ \nearrow \end{array}\right)$ such that

$$h(\text{Quark}) = 1$$
$$h(\text{Gluon}) = 0$$

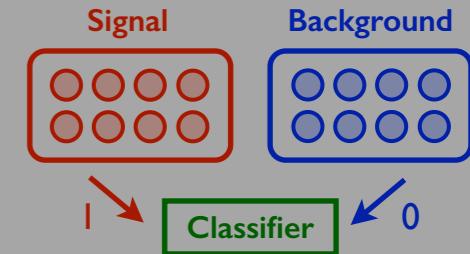
Best you can do: $h(\mathcal{J}) = \frac{p(\mathcal{J}|Q)}{p(\mathcal{J}|Q) + p(\mathcal{J}|G)}$

(Neyman-Pearson lemma)

[see e.g. Gras, Höche, Kar, Larkoski, Lönnblad, Plätzer, Siódmok, Skands, Soyez, JDT, [JHEP 2017](#); Komiske, Metodiev, Schwartz, [JHEP 2017](#); Komiske, Metodiev, JDT, [JHEP 2018](#)]

Quark/Gluon Classification

“Hello, World!” of Jet Physics



What do you mean by “quark” and “gluon”?

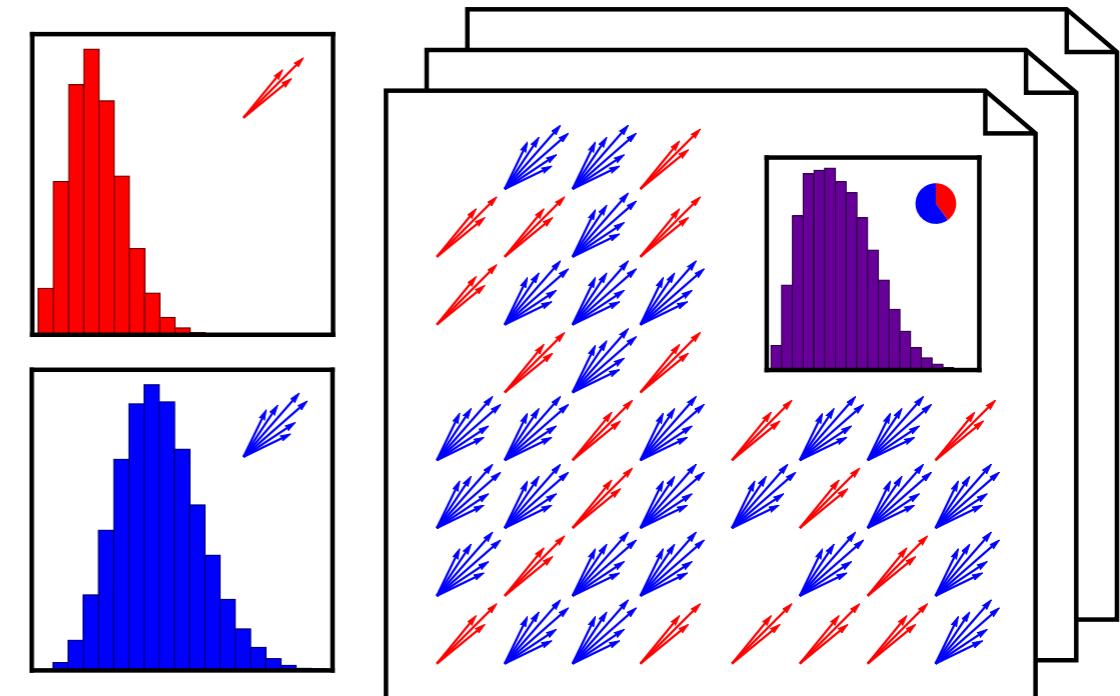
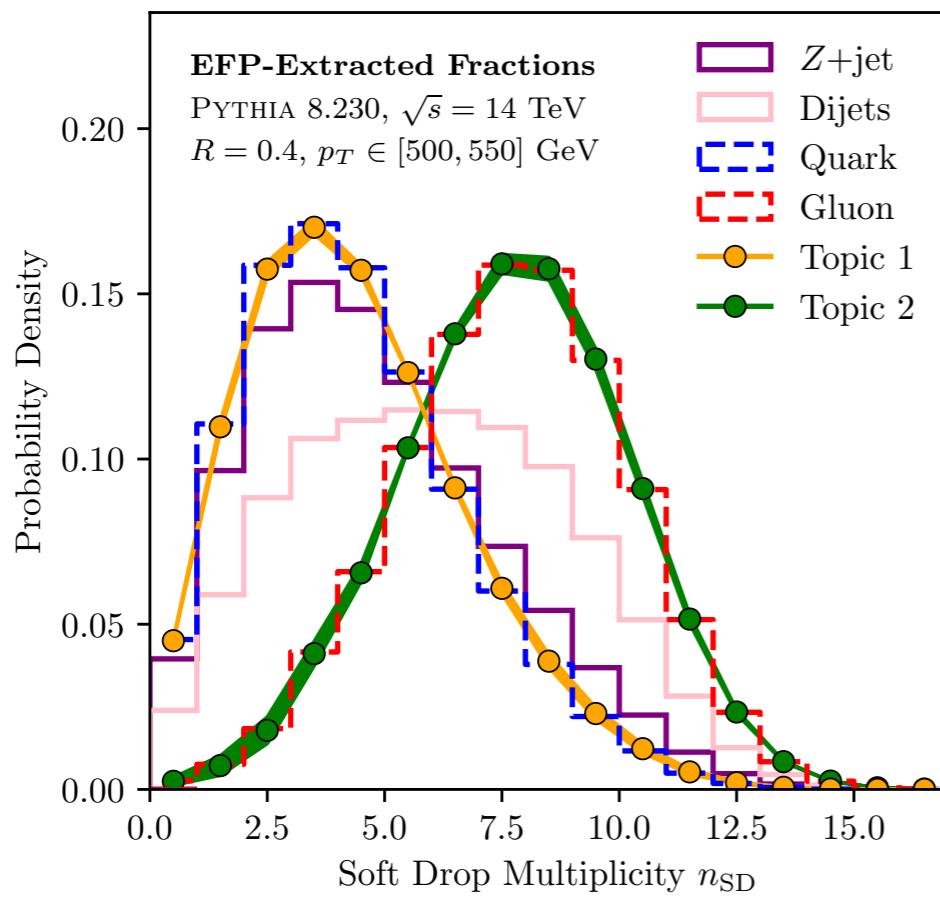
Jets are clusters of colorless hadrons!

Parton shower “truth” is but a (useful) fiction!

[see e.g. Gras, Höche, Kar, Larkoski, Lönnblad, Plätzer, Siódmok, Skands, Soyez, JDT, [JHEP 2017](#); Komiske, Metodiev, Schwartz, [JHEP 2017](#); Komiske, Metodiev, JDT, [JHEP 2018](#)]

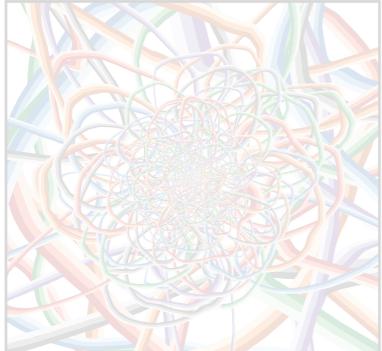
Topic Modeling to Disentangle Jet Categories

While you can't unambiguously label individual jets, you can extract **quark** and **gluon** distributions from **hadron-level measurements**



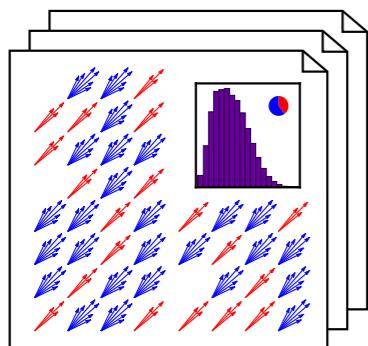
Key concept from natural language processing: “**anchor words**”

[Komiske, Metodiev, JDT, [JHEP 2018](#); using Metodiev, Nachman, JDT, [JHEP 2017](#); Metodiev, JDT, [PRL 2018](#)]
see also Blanchard, Flaska, Handy, Pozzi, Scott, [PLMR 2013](#); Katz-Samuels, Blanchard, Scott, [JMLR 2016](#)]



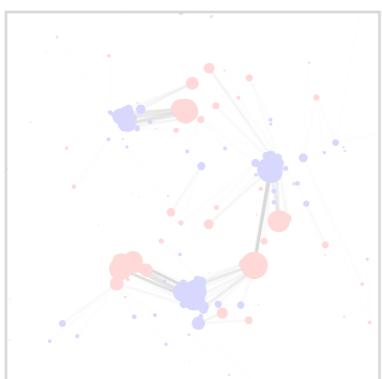
Can theoretical structures be encoded directly?

Energy Flow Networks \Leftrightarrow IRC Safety + Permutations

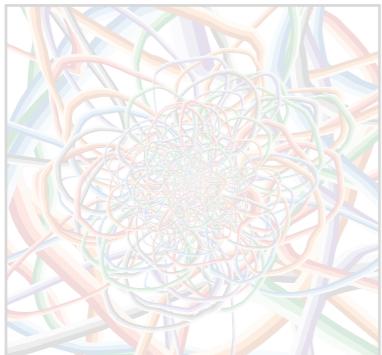


Can strategy be defined on physical quantities?

Jet Topics \Leftrightarrow Hadron-Level Approach to QCD Partons

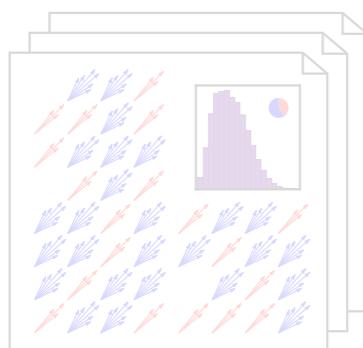


Can we leverage unsupervised machine learning?



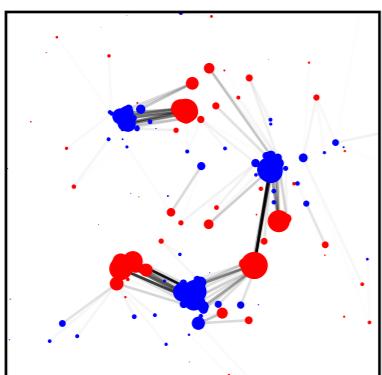
Can theoretical structures be encoded directly?

Energy Flow Networks \leftrightarrow IRC Safety + Permutations



Can strategy be defined on physical quantities?

Jet Topics \leftrightarrow Hadron-Level Approach to QCD Partons



Can we leverage unsupervised machine learning?

The Earth Mover's Distance

Optimal Transport:

[Peleg, Werman, Rom, [IEEE 1989](#);
Rubner, Tomasi, Guibas, [ICCV 1998](#), [ICCV 2000](#);
Pele, Werman, [ECCV 2008](#); Pele Taskar, [GSI 2013](#)]

Minimum “work” (stuff \times distance) to make one distribution look like another distribution



Déblai

Remblai

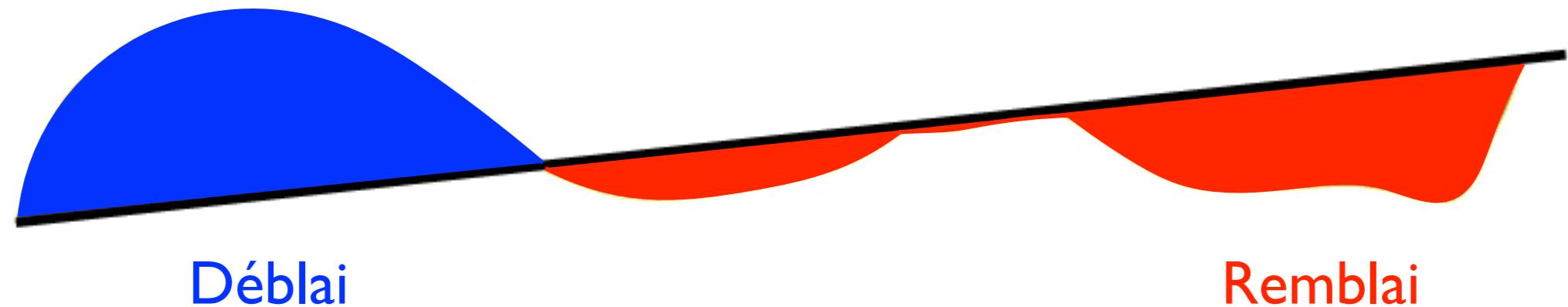
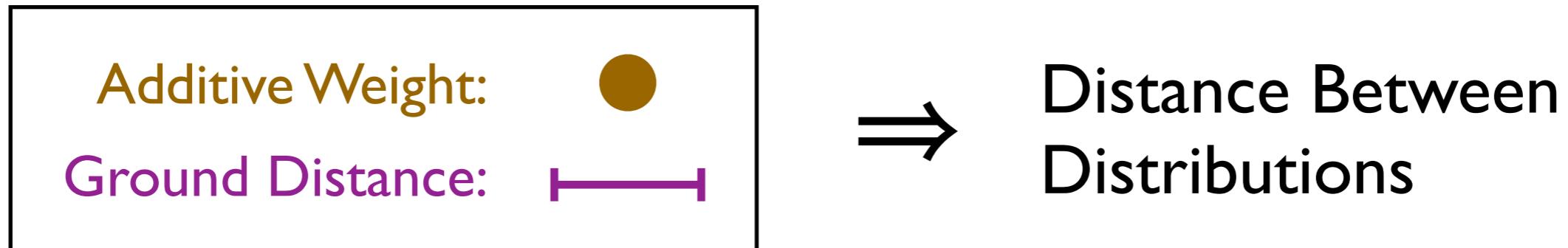
[h/t Niles-Weed, [ML4Jets 2020](#); Monge, 1781; Vaserštejn, 1969; [Wikipedia](#)]

The Earth Mover's Distance

Optimal Transport:

[Peleg, Werman, Rom, [IEEE 1989](#);
Rubner, Tomasi, Guibas, [ICCV 1998](#), [ICCV 2000](#);
Pele, Werman, [ECCV 2008](#); Pele Taskar, [GSI 2013](#)]

Minimum “work” (**stuff** × **distance**) to make
one distribution look like **another distribution**



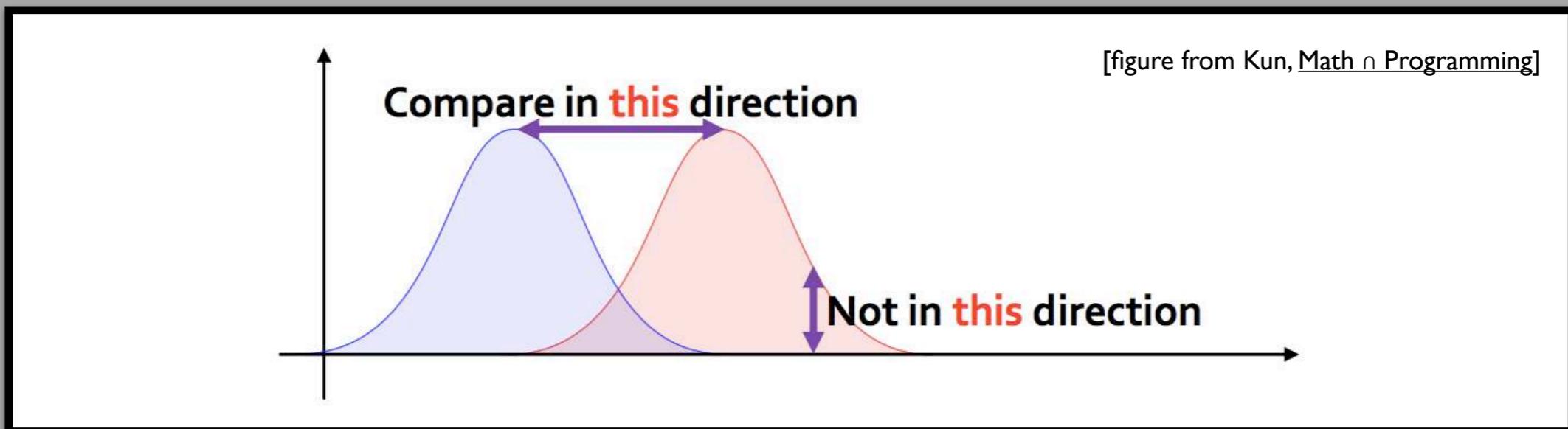
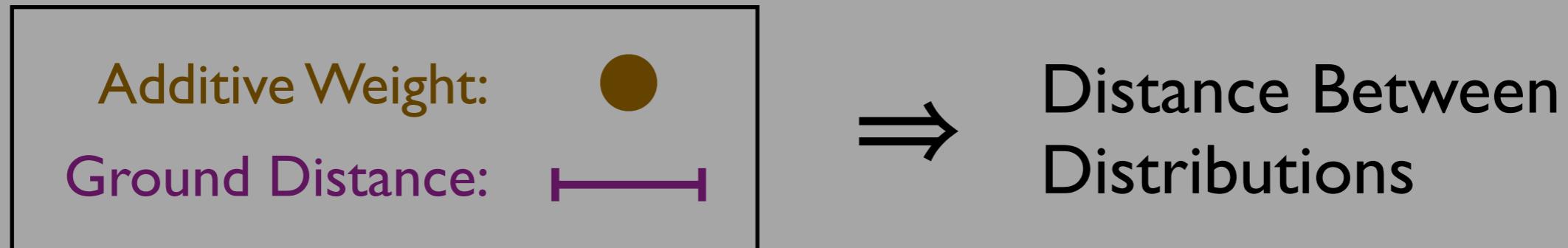
[h/t Niles-Weed, [ML4Jets 2020](#); Monge, 1781; Vaserštejn, 1969; [Wikipedia](#)]

The Earth Mover's Distance

Optimal Transport:

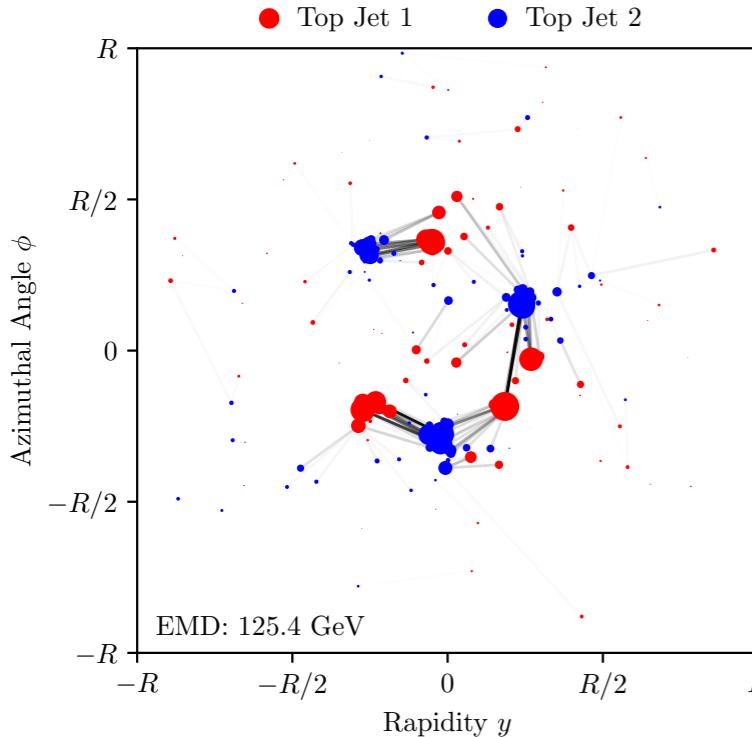
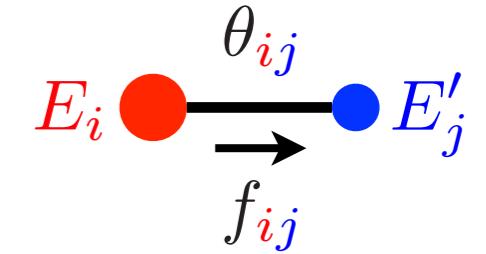
[Peleg, Werman, Rom, [IEEE 1989](#);
Rubner, Tomasi, Guibas, [ICCV 1998](#), [ICCV 2000](#);
Pele, Werman, [ECCV 2008](#); Pele Taskar, [GSI 2013](#)]

Minimum “work” (**stuff \times distance**) to make
one distribution look like **another distribution**



[h/t Niles-Weed, [ML4Jets 2020](#); Monge, 1781; Vaserštejn, 1969; [Wikipedia](#)]

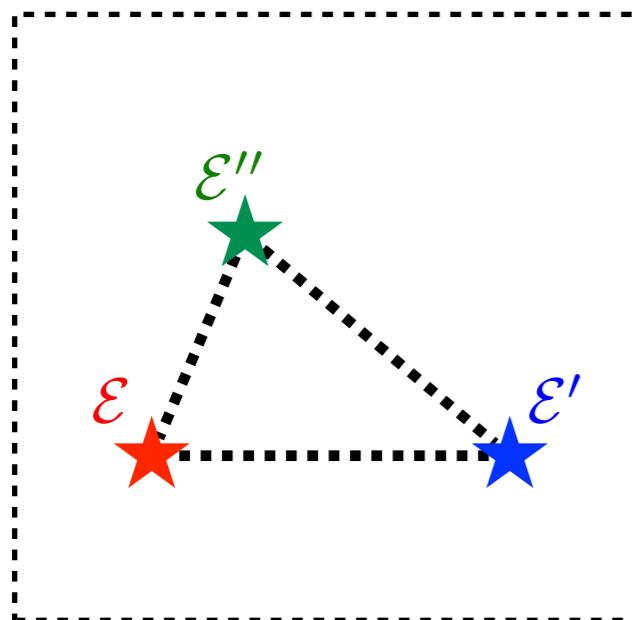
The Energy Mover's Distance



Optimal transport between energy flows...

$$\text{EMD}(\mathcal{E}, \mathcal{E}') = \min_{\{f\}} \sum_i \sum_j f_{ij} \frac{\theta_{ij}}{R} + \left| \sum_i E_i - \sum_j E'_j \right|$$

↑
in GeV
Cost to move energy
Cost to create energy



...defines a metric on the space of events

$$0 \leq \text{EMD}(\mathcal{E}, \mathcal{E}') \leq \text{EMD}(\mathcal{E}, \mathcal{E}'') + \text{EMD}(\mathcal{E}', \mathcal{E}'')$$

(assuming $R \geq \theta_{\max}/2$, i.e. $R \geq$ jet radius for conical jets)

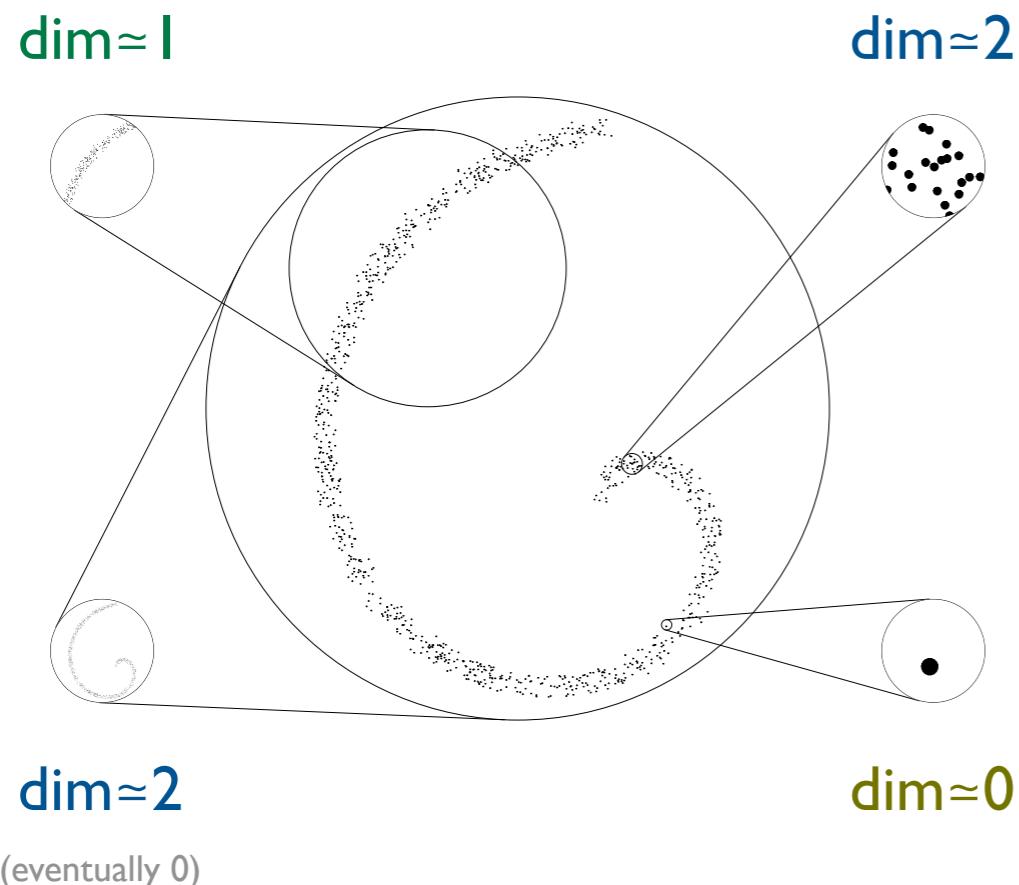
[Komiske, Metodiev, JDT, [PRL 2019](#); see also Pele, Werman, [ECCV 2008](#); Pele, Taskar, [GSI 2013](#)
 [see flavored variant in Crispim Romão, Castro, Milhano, Pedro, Vale, [EPJC 2021](#)]
 [see computational speed up in Cai, Cheng, Craig, Craig, [PRD 2020](#)]
 [see graph network approach in Mullin, Pacey, Parker, White, Williams, [arXiv 2019](#)]

Dimensionality of Space of Jets

$$N_{\text{neighbors}}(r) \sim r^{\dim}$$

$$\Rightarrow \dim(r) \sim r \frac{\partial}{\partial r} \ln N_{\text{neighbors}}(r)$$

[Grassberger, Procaccia, [PRL 1983](#); Kégl, [NIPS 2002](#)]



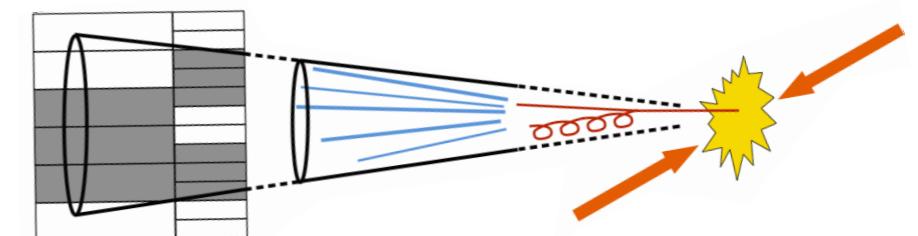
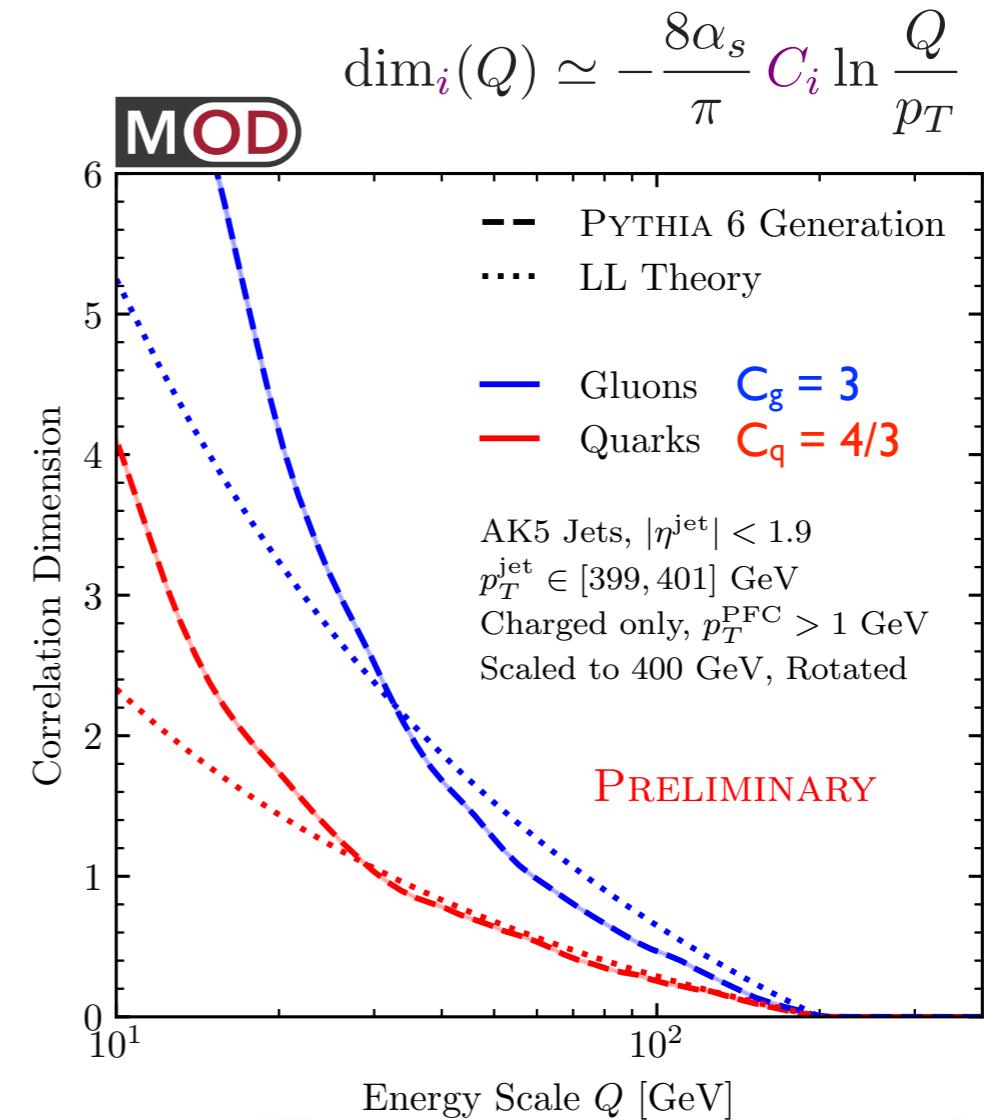
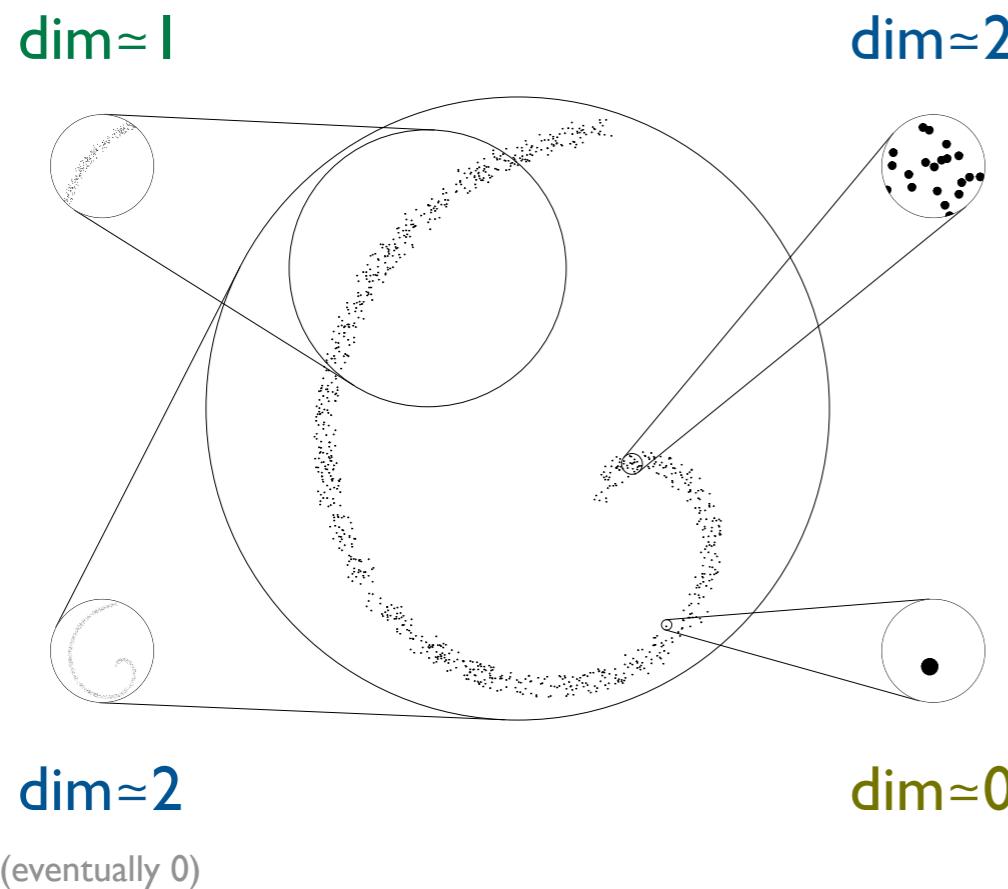
Dimensionality of Space of Jets



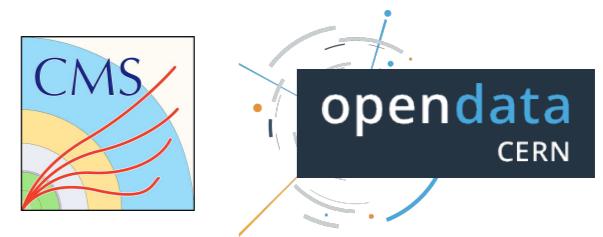
$$N_{\text{neighbors}}(r) \sim r^{\dim}$$

$$\Rightarrow \dim(r) \sim r \frac{\partial}{\partial r} \ln N_{\text{neighbors}}(r)$$

[Grassberger, Procaccia, [PRL 1983](#); Kégl, [NIPS 2002](#)]



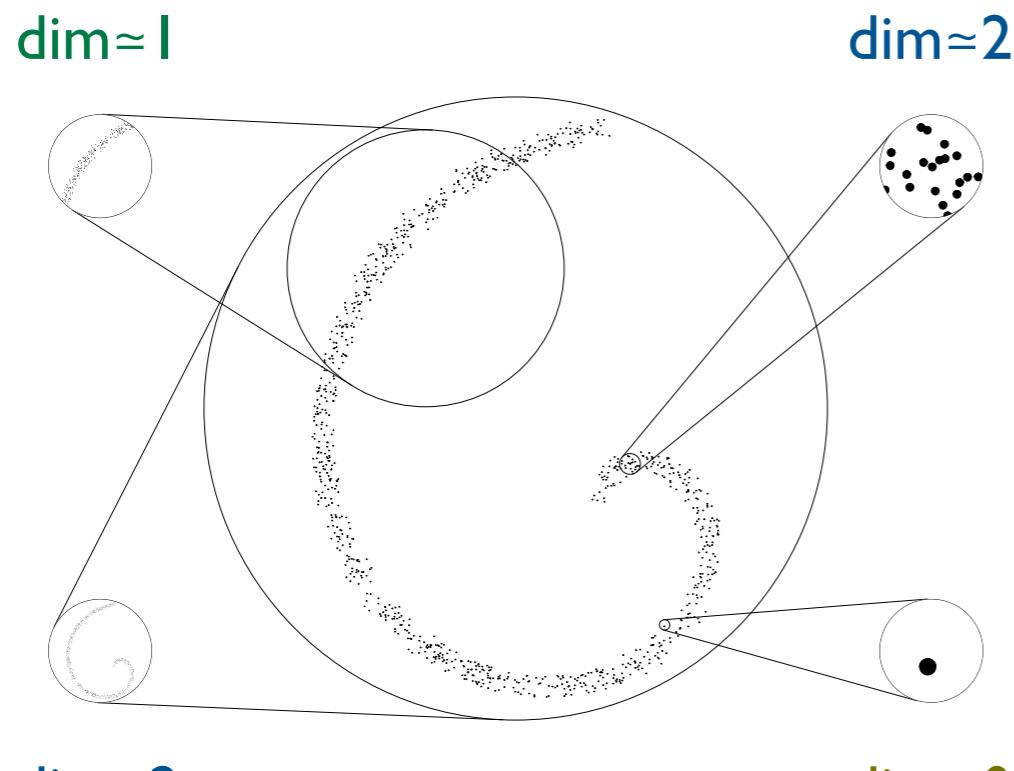
Dimensionality of Space of Jets



$$N_{\text{neighbors}}(r) \sim r^{\dim}$$

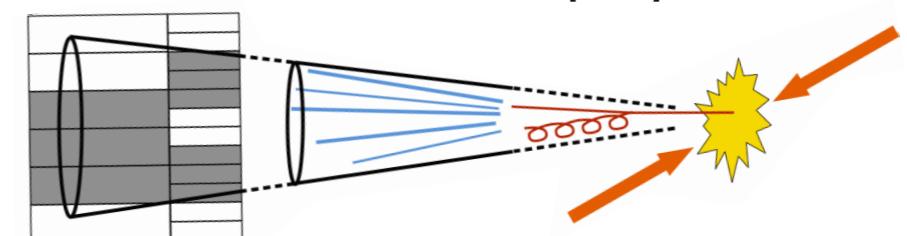
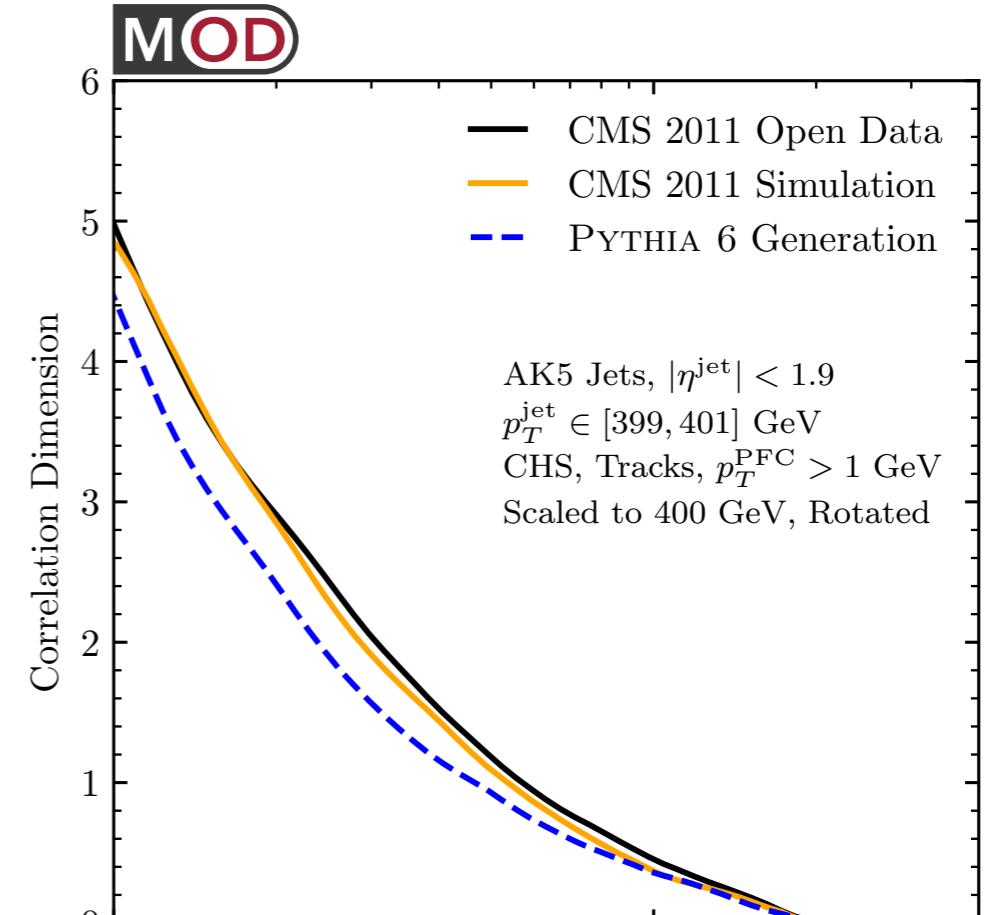
$$\Rightarrow \dim(r) \sim r \frac{\partial}{\partial r} \ln N_{\text{neighbors}}(r)$$

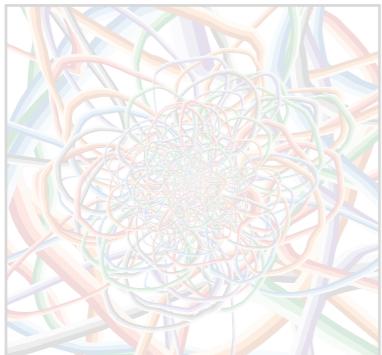
[Grassberger, Procaccia, [PRL 1983](#); Kégl, [NIPS 2002](#)]



(eventually 0)

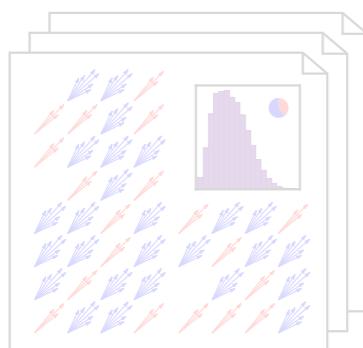
[Komiske, Mastandrea, Metodiev, Naik, [JDT, PRD 2020](#);
using [CMS Open Data](#)]





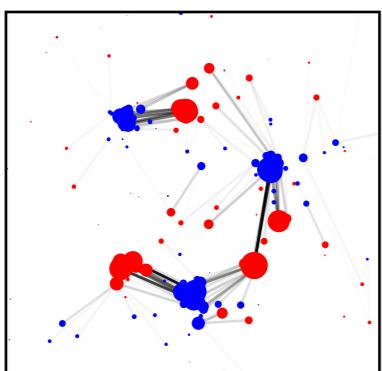
Can theoretical structures be encoded directly?

Energy Flow Networks \Leftrightarrow IRC Safety + Permutations



Can strategy be defined on physical quantities?

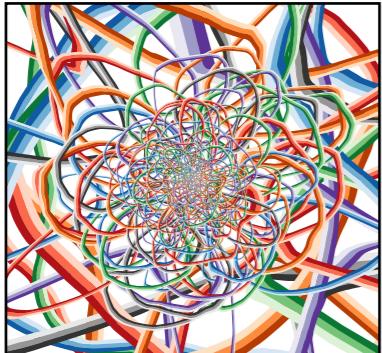
Jet Topics \Leftrightarrow Hadron-Level Approach to QCD Partons



Can we leverage unsupervised machine learning?

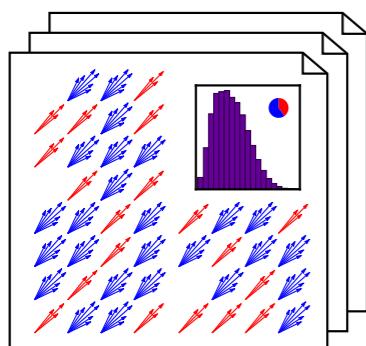
Energy Mover's Distance \Leftrightarrow Geometric Strategies for Collider Physics

Artificial Intelligence and High-Energy Physics



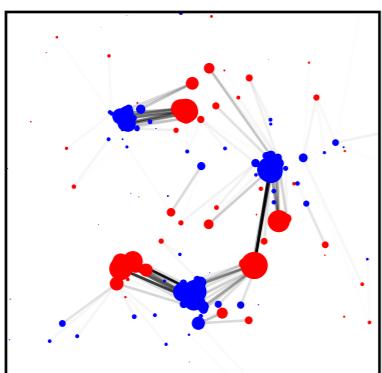
Can theoretical structures be encoded directly?

Energy Flow Networks \Leftrightarrow IRC Safety + Permutations



Can strategy be defined on physical quantities?

Jet Topics \Leftrightarrow Hadron-Level Approach to QCD Partons



Can we leverage unsupervised machine learning?

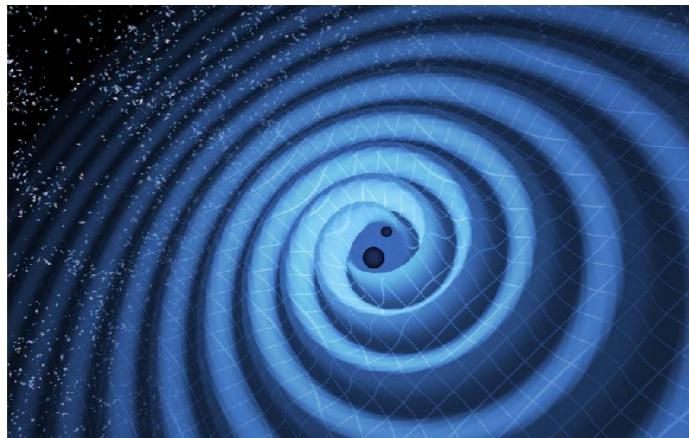
Energy Mover's Distance \Leftrightarrow Geometric Strategies for Collider Physics

Theory insights essential for developing these tools

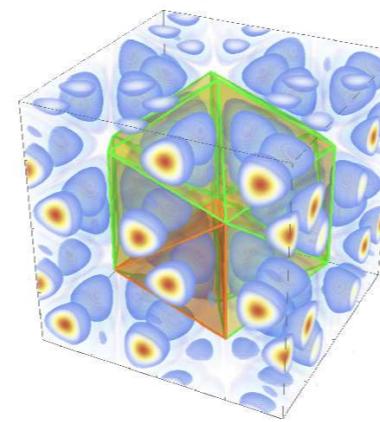
Artificial Intelligence \leftrightarrow Fundamental Interactions



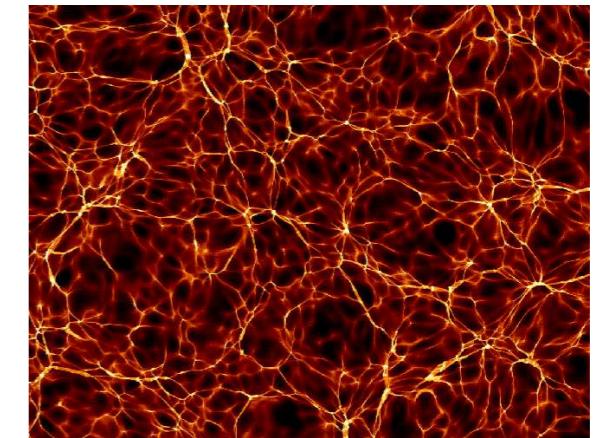
Gravitational Waves



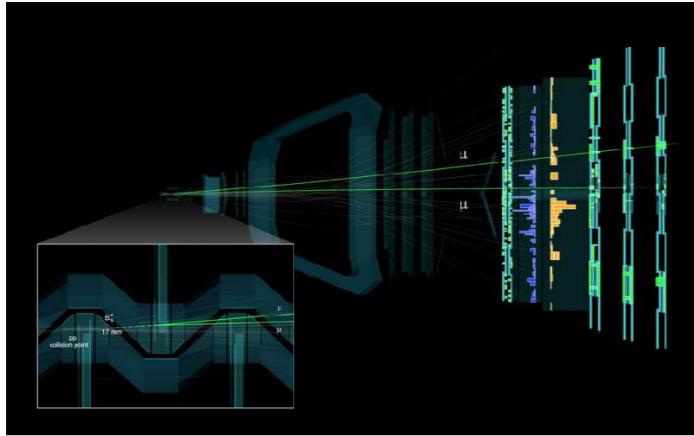
Nuclear Physics



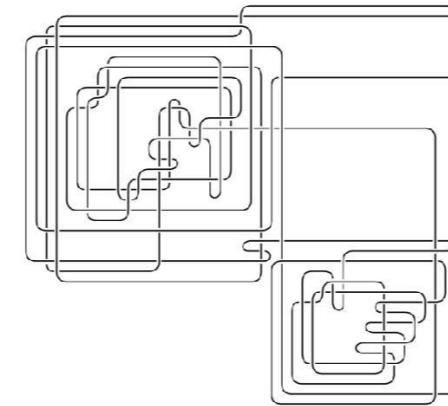
Dark Matter



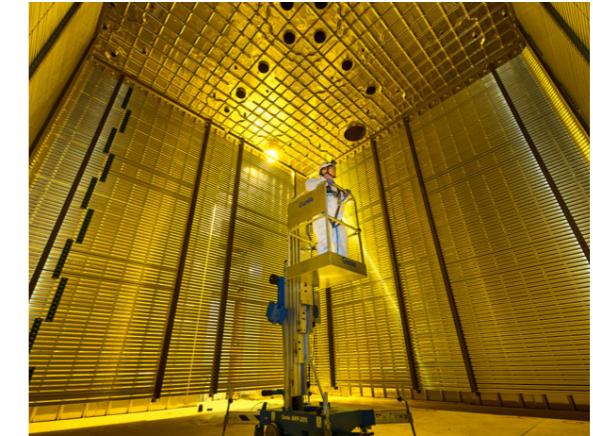
Particle Colliders



Mathematical Physics



Neutrino Detection



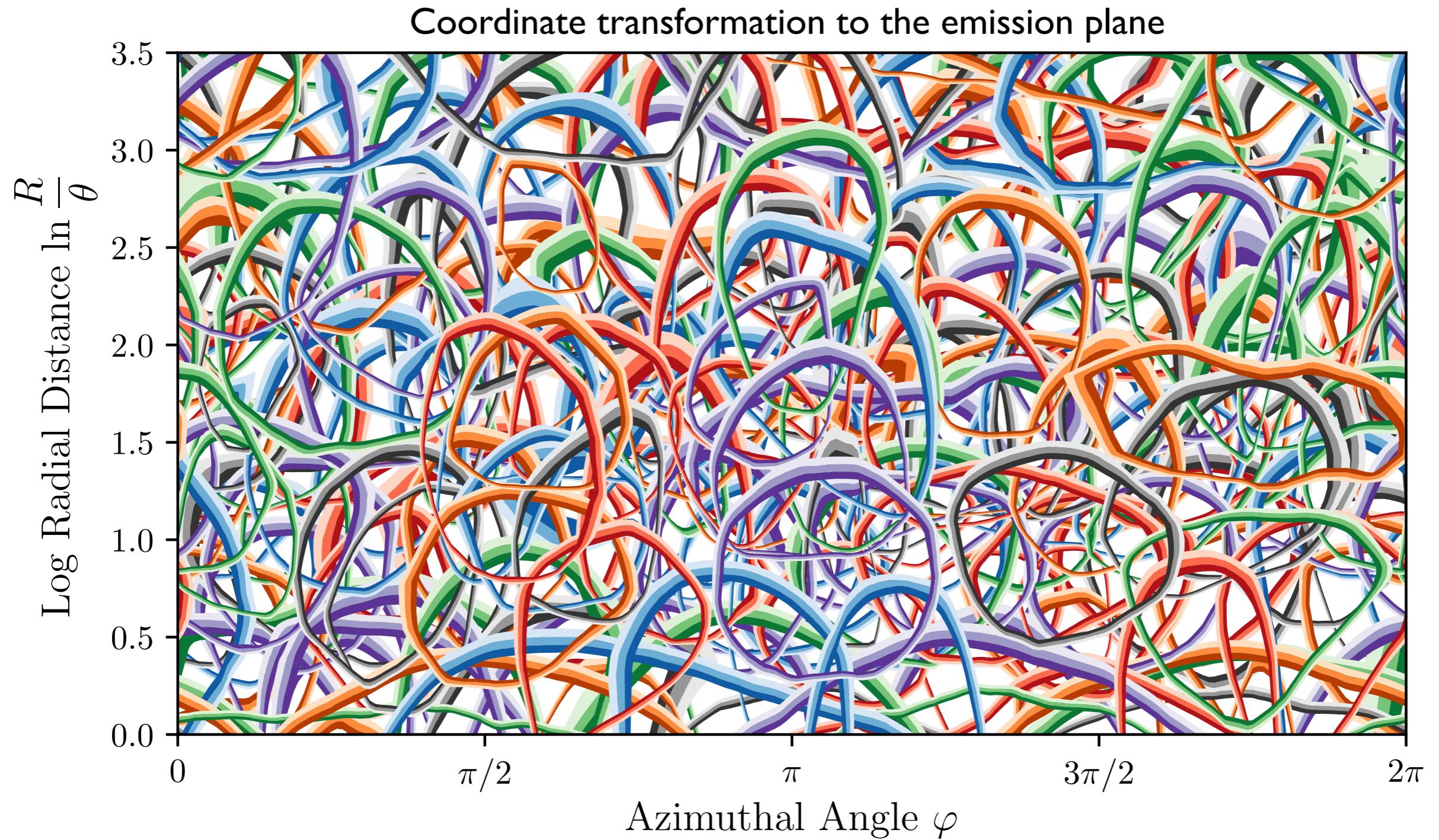
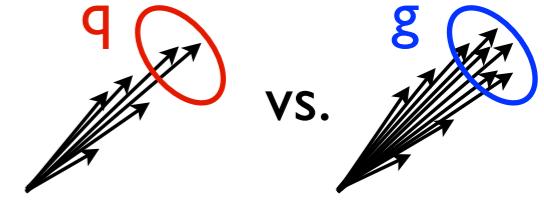
...

Machine learning that incorporates first principles, best practices, and domain knowledge from fundamental physics

[<http://iaifi.org>]

Backup Slides

En Route to the Lund Plane



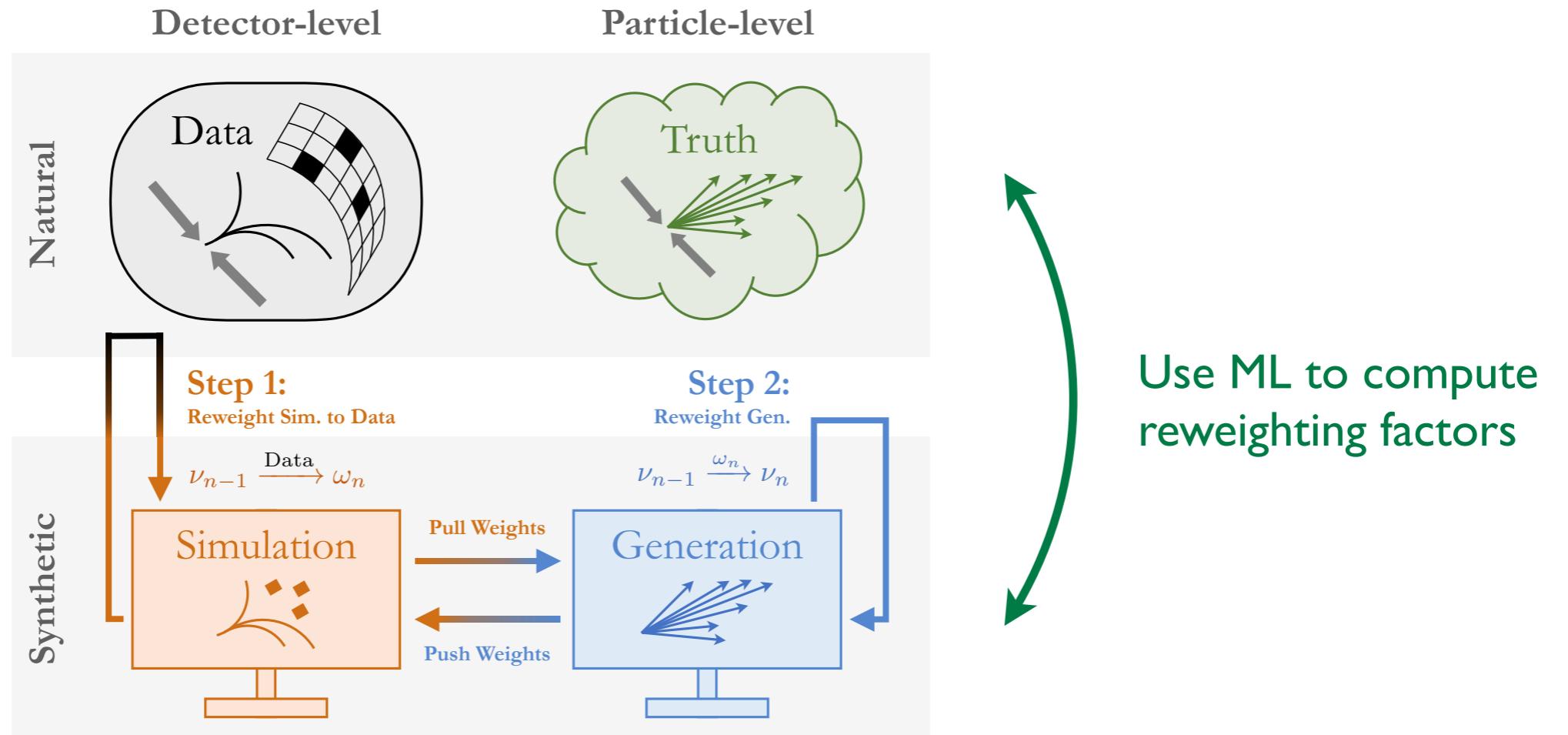
[Komiske, Metodiev, JDT, [JHEP 2019](#); see also Dreyer, Salam, Soyez, [JHEP 2018](#)]

E.g. Detector Unfolding

OmniFold



*Multi-dimensional unbinned detector corrections
via iterated application of likelihood ratio trick*

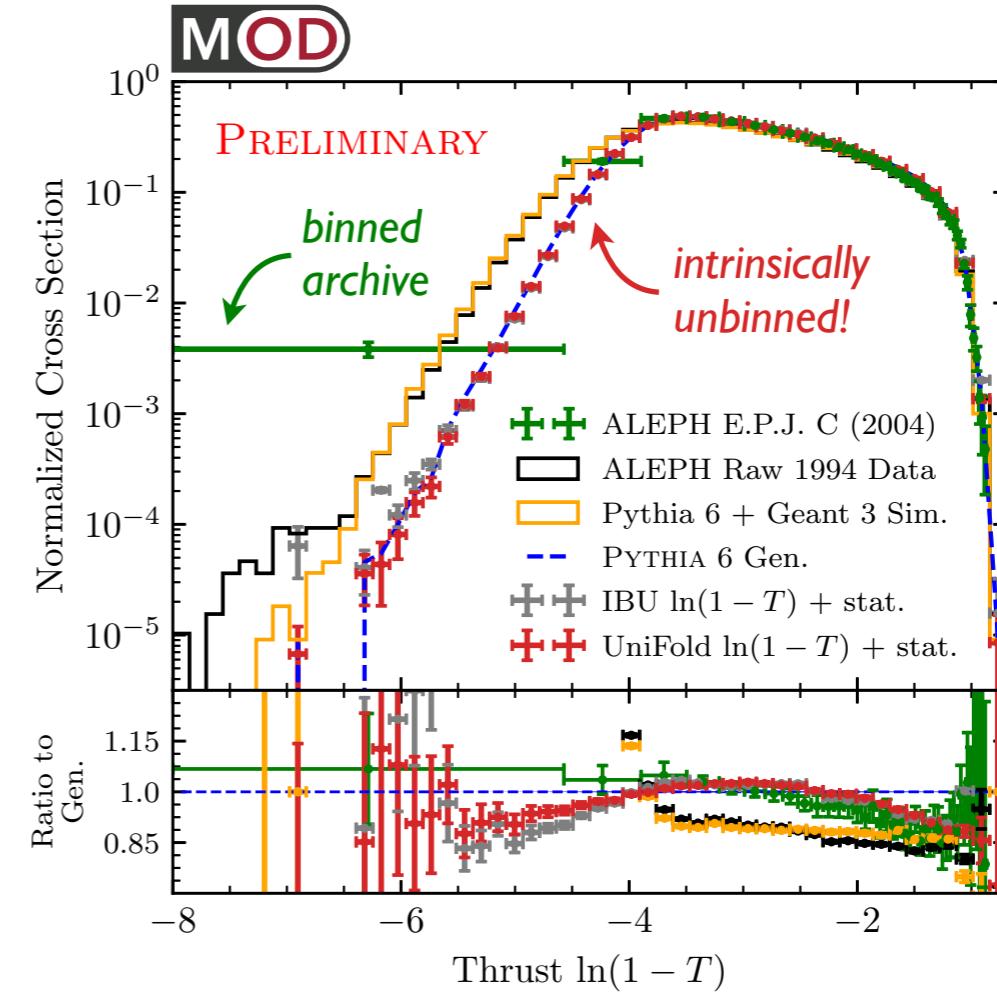
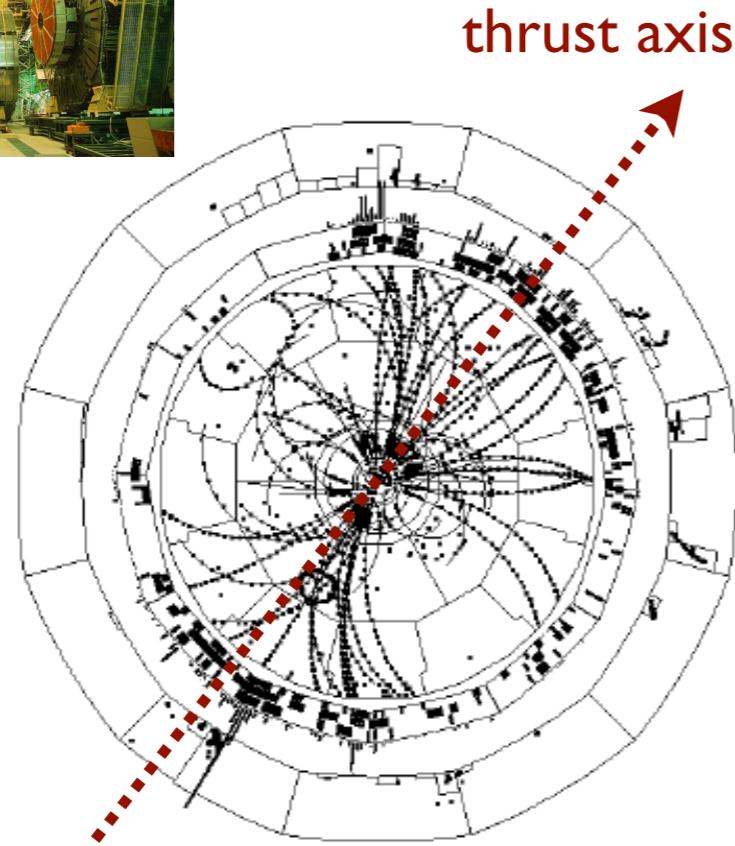


[Andreassen, Komiske, Metodiev, Nachman, JDT, [PRL 2020](#)]

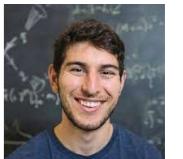


E.g. Detector Unfolding

Back to the Future with ALEPH Archival Data



[see also Badea, Baty, Chang, Innocenti, Maggi, McGinn, Peters, Sheng, JDT, Lee, [PRL 2019](#)]



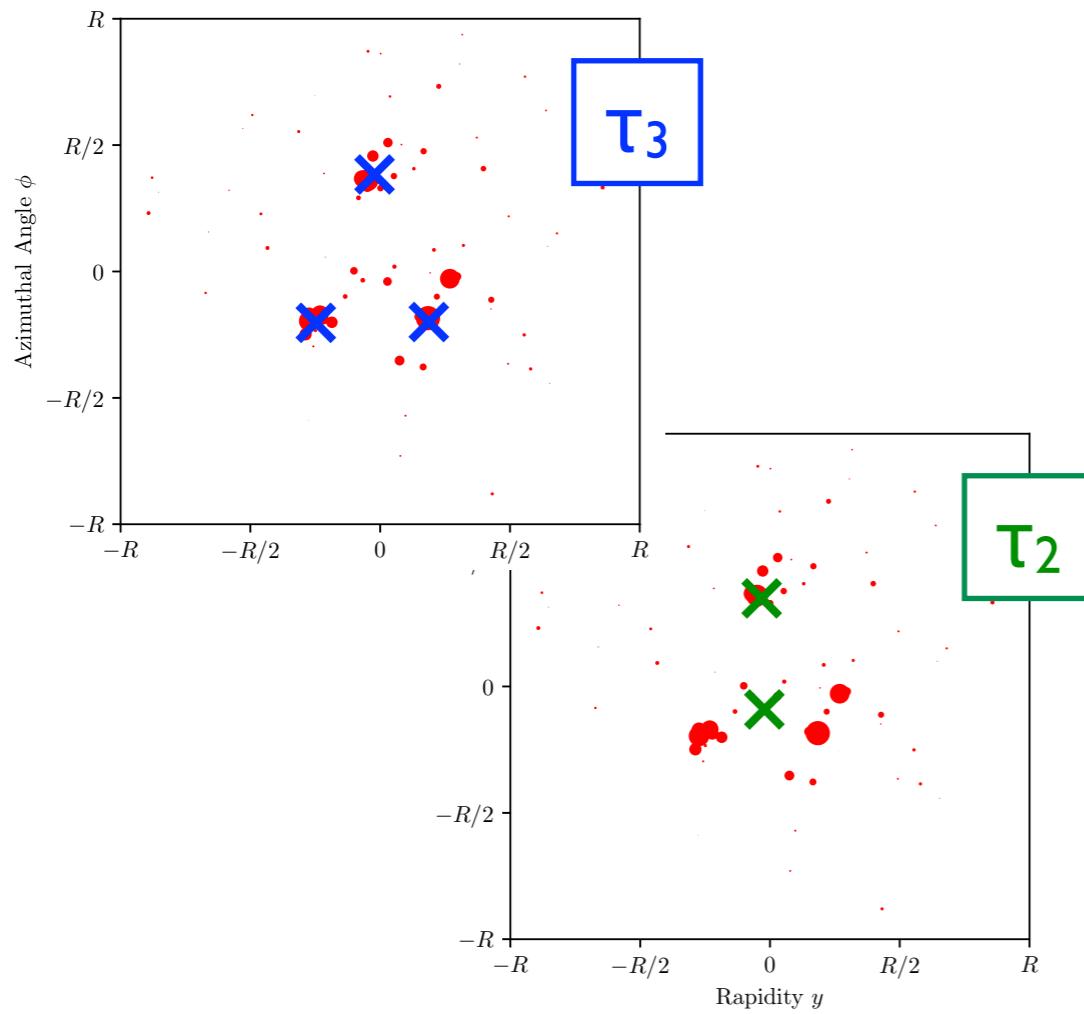
[Andreassen, Komiske, Metodiev, Nachman, JDT, [PRL 2020](#)]



N-subjettiness

Ubiquitous jet substructure observable used for almost a decade...

$$\tau_N(\mathcal{J}) = \min_{N \text{ axes}} \sum_i E_i \min \{\theta_{1,i}, \theta_{2,i}, \dots, \theta_{N,i}\}$$

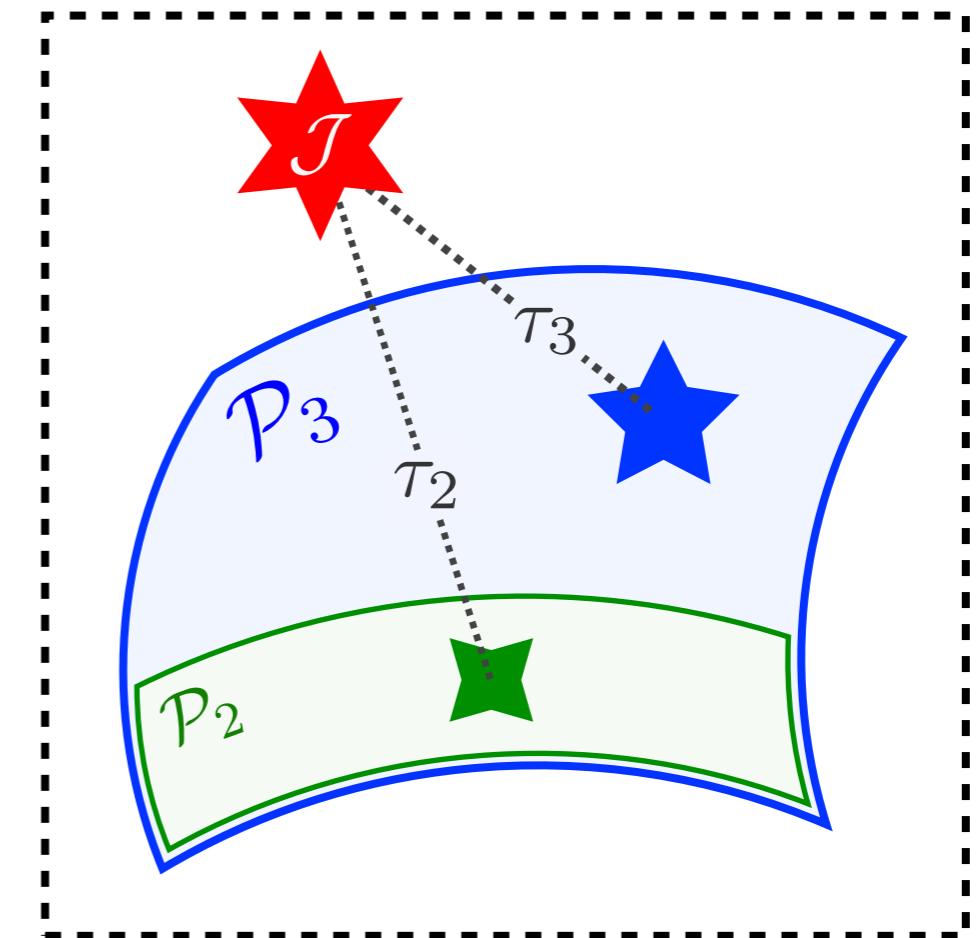
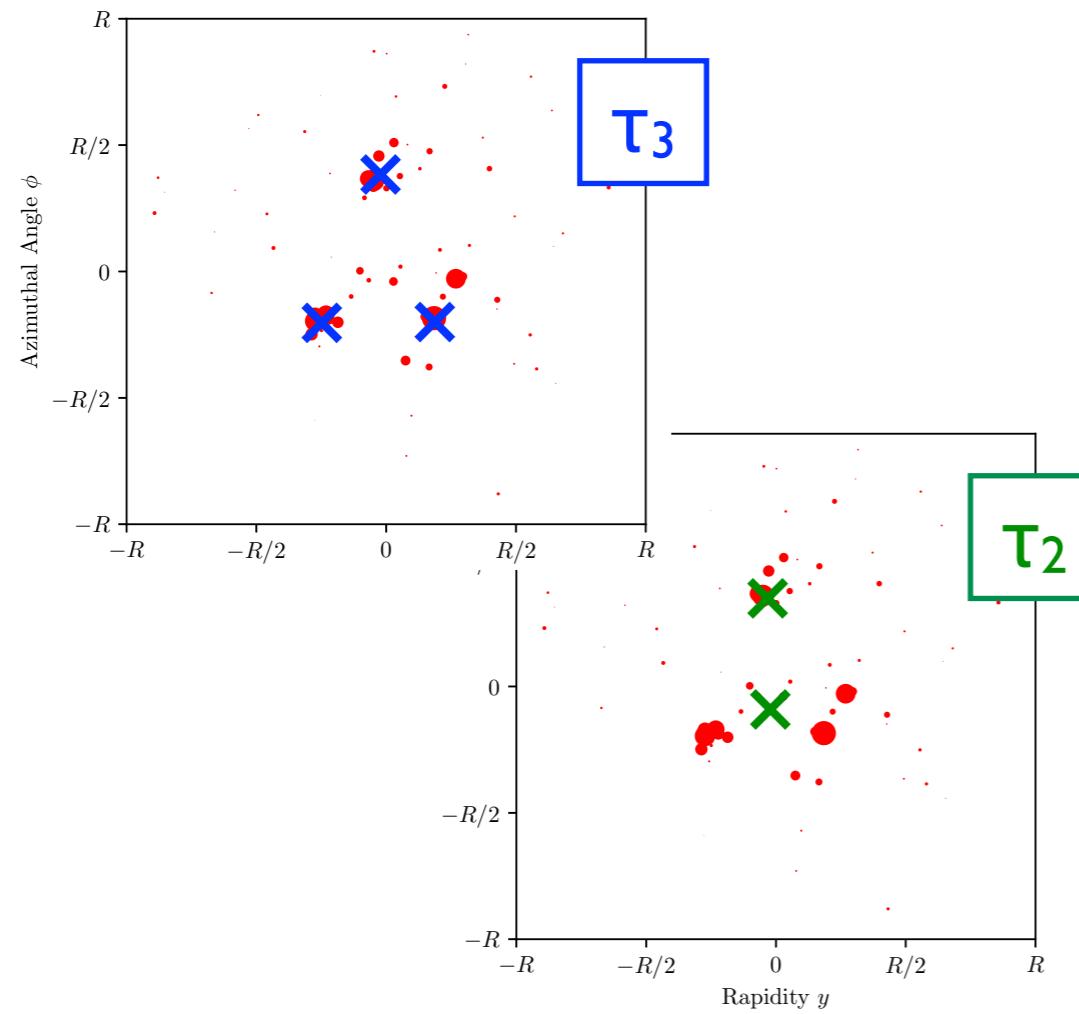


JDT, Van Tilburg, [JHEP 2011](#), [JHEP 2012](#);
based on Brandt, Dahmen, [ZPC 1979](#); Stewart, Tackmann, Waalewijn, [PRL 2010](#)

N-subjettiness = Point to Manifold EMD

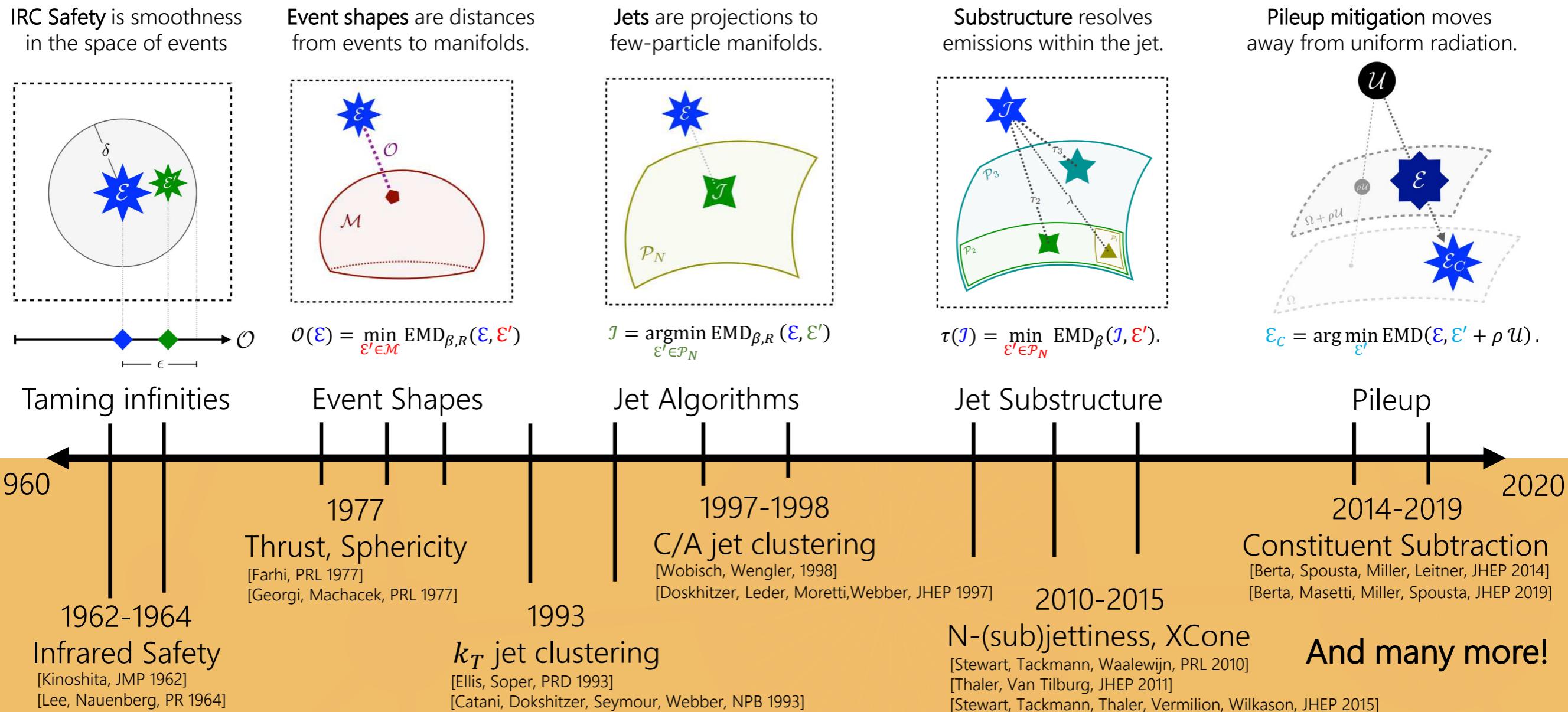
...is secretly an optimal transport problem

$$\tau_N(\mathcal{J}) = \min_{\mathcal{J}' \in \mathcal{P}_N} \text{EMD}(\mathcal{J}, \mathcal{J}')$$



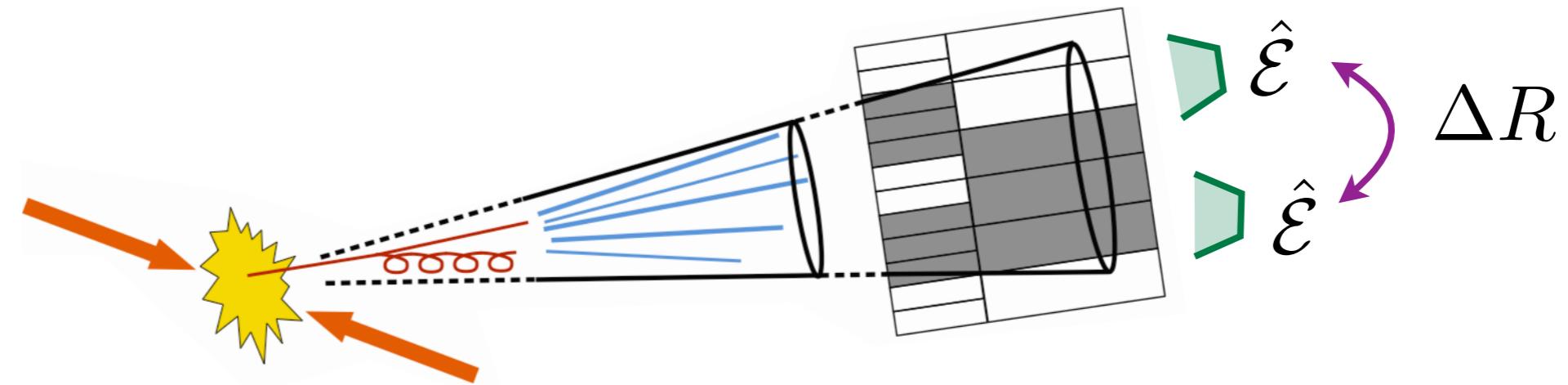
[JDT, Van Tilburg, [JHEP 2011](#), [JHEP 2012](#);
rephrased via Komiske, Metodiev, JDT, [JHEP 2020](#); see opposite limit in Cesarotti, JDT, [JHEP 2020](#)]

Six Decades of Collider Physics Translated into a New Geometric Language!

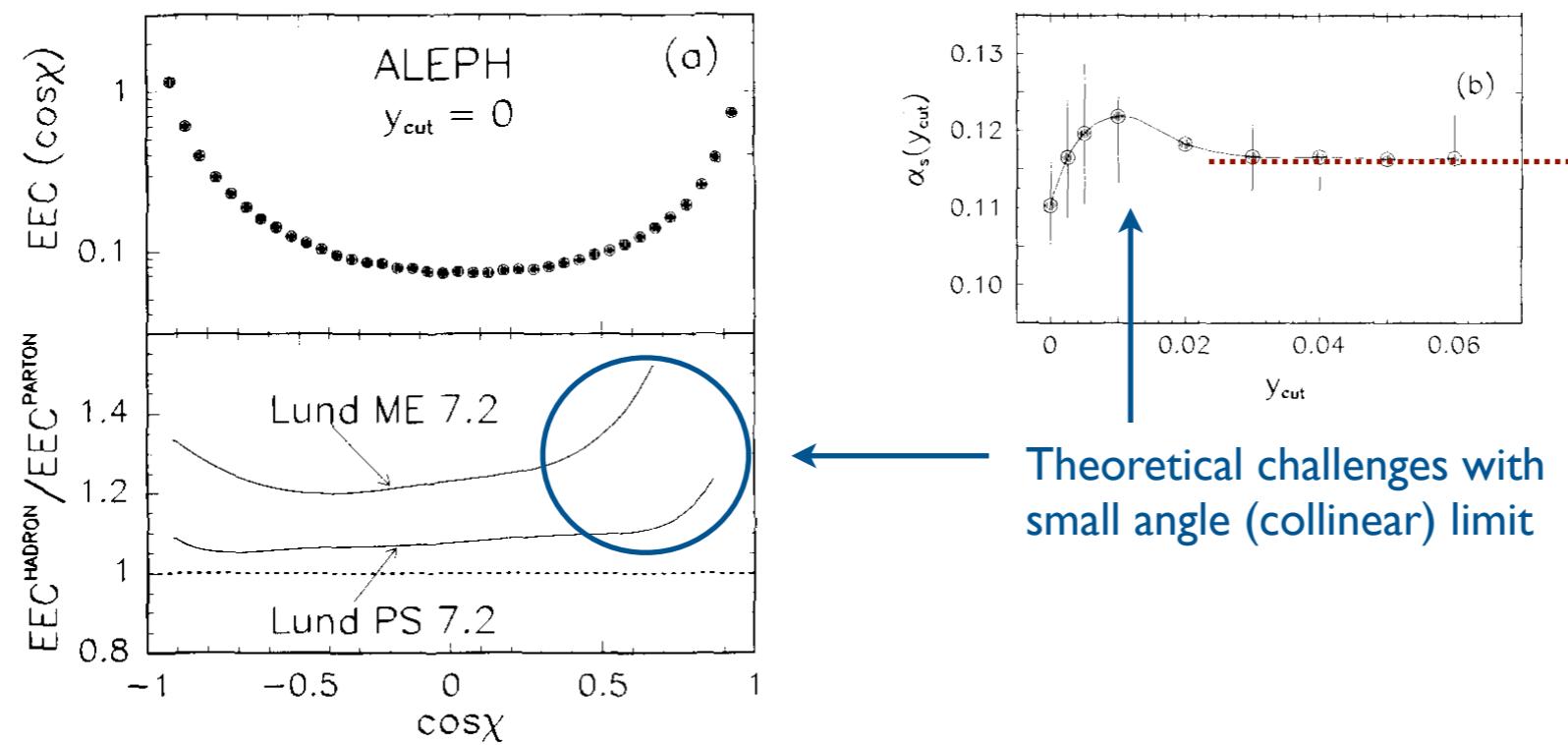


[Komiske, Metodiev, JDT, JHEP 2020; timeline by Metodiev]

Energy-Energy Correlators

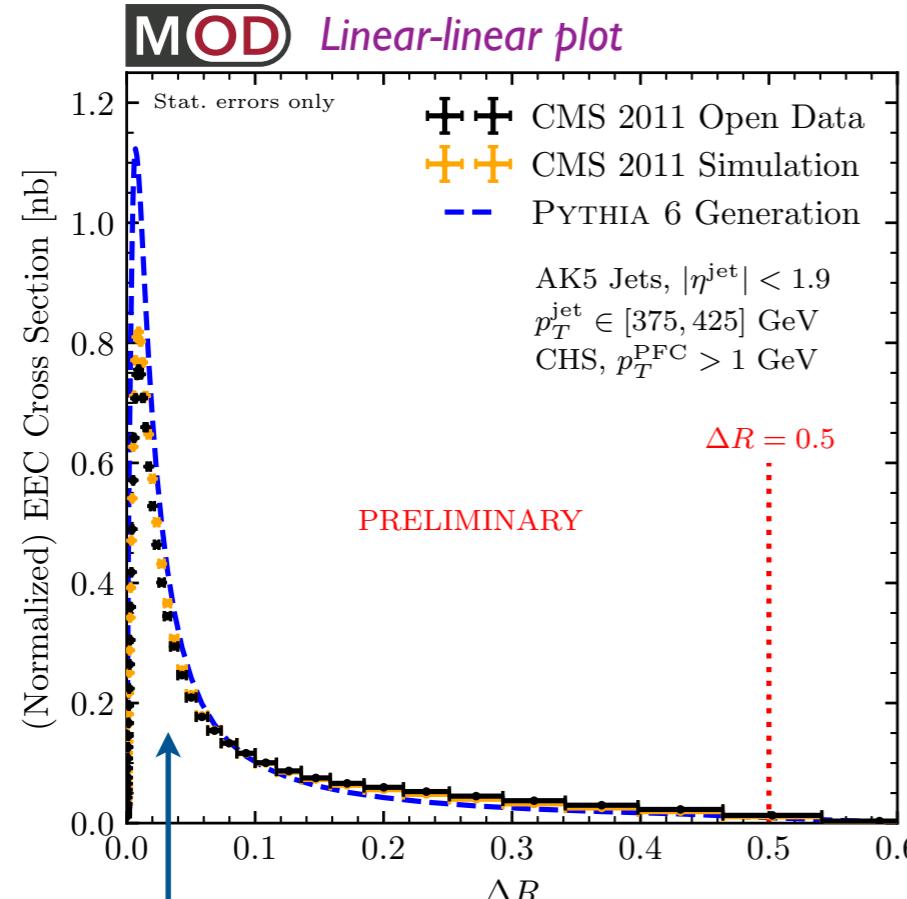


A long history in probing collinear dynamics of QCD



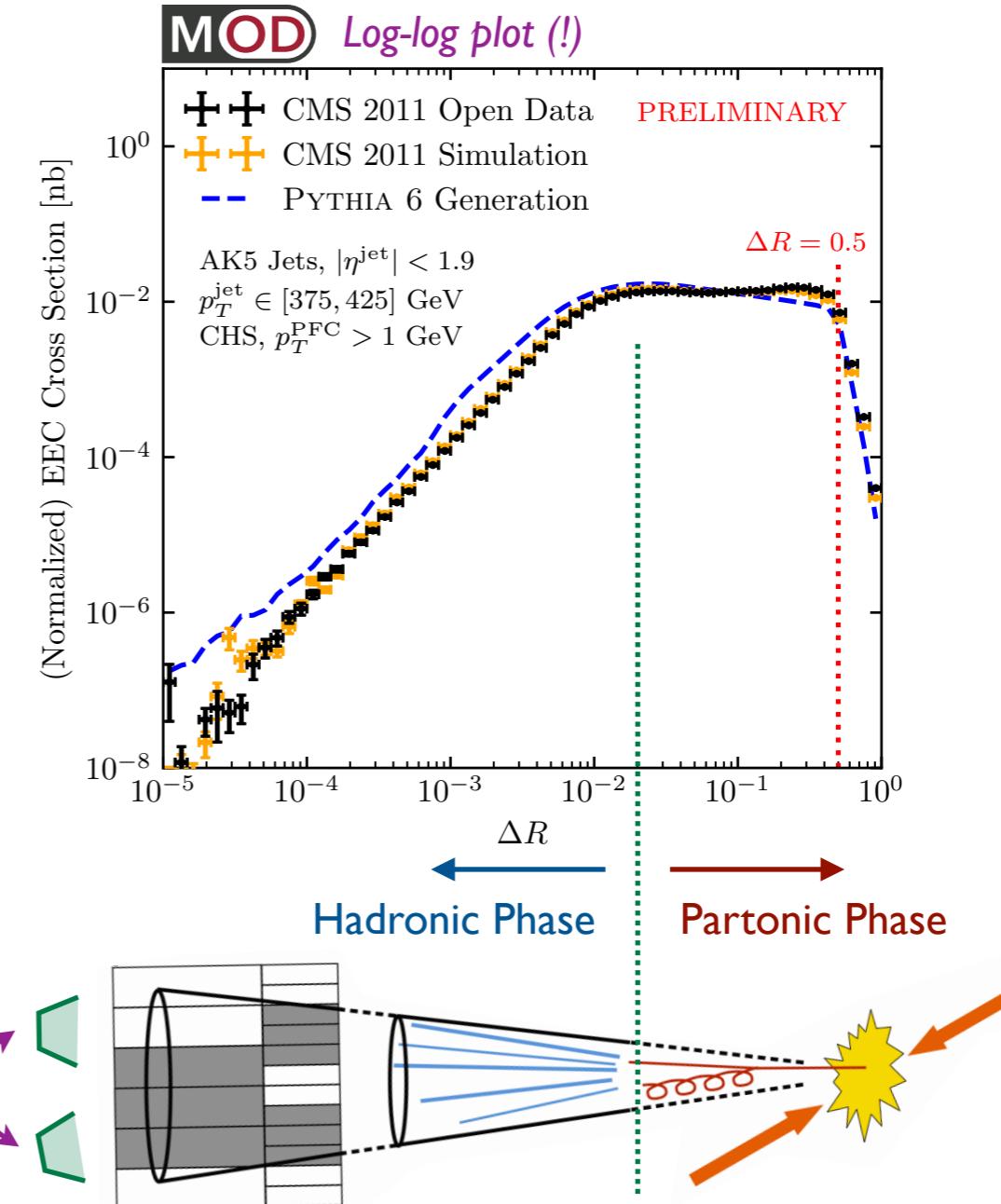
[Basham, Brown, Ellis, Love, [PRL 1978](#); ALEPH, [PLB 1991](#); see Chen, Moult, Zhang, Zhu, [PRD 2020](#)]

QCD Phase Transition in Jets?



Are we learning something about small angle limit of QCD?

First Jet EEC Plot from the LHC (!)



[Komiske, Moult, JDT, Zhu, in progress; see talks by Moult, [BOOST 2019](#), [BOOST 2020](#)]

