QCD and Jets through the Lens of Machine Learning

Jesse Thaler

Presentation with Frédéric Dreyer

New Physics from Precision at High Energies, KITP, Santa Barbara — March 30, 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]

AI^2: Ab Initio Artificial Intelligence

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

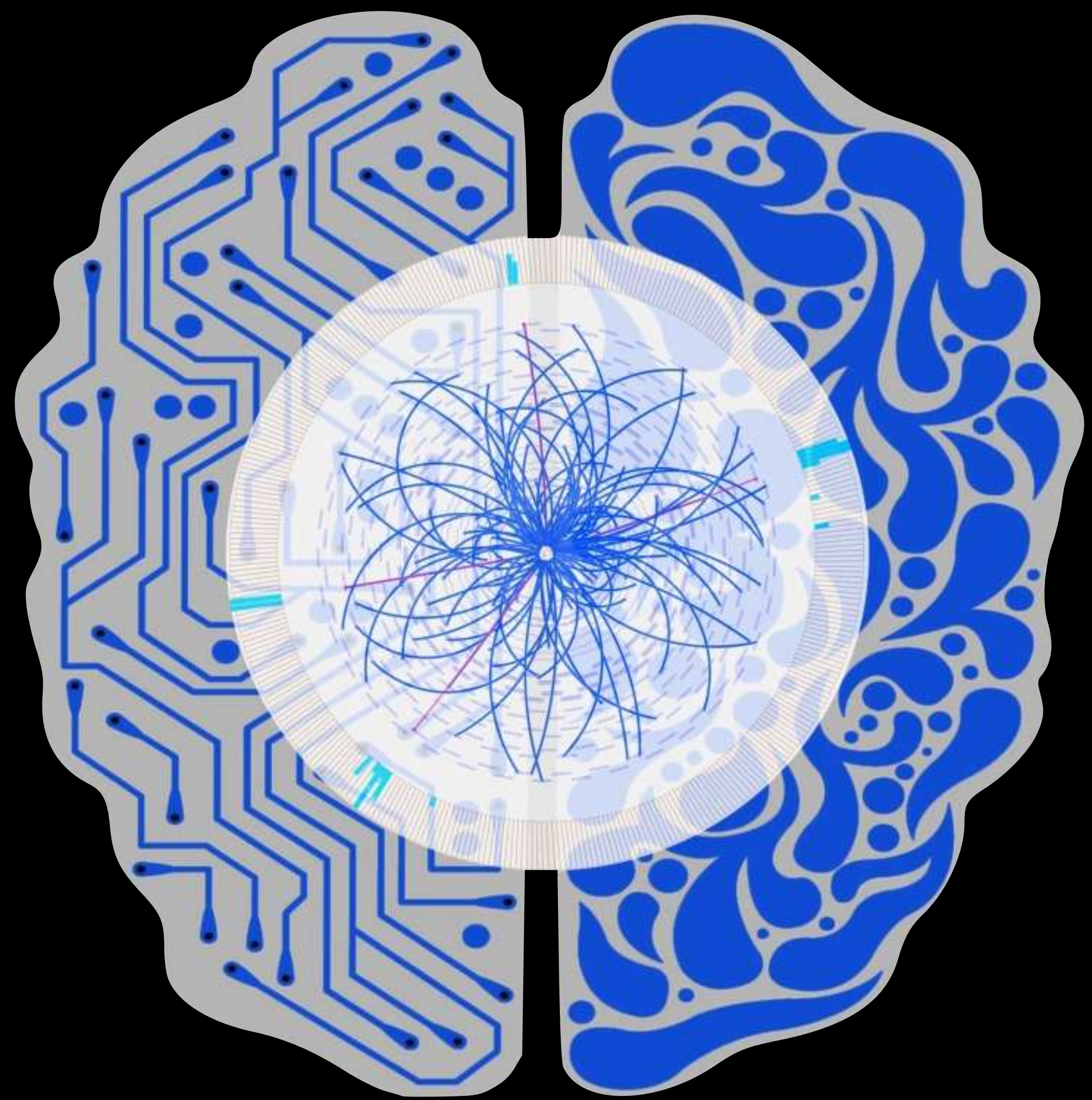
Symmetries, conservation laws, scaling relations, limiting behaviors, locality, causality, unitarity, gauge invariance, entropy, least action, factorization, unit tests, exactness, systematic uncertainties, reproducibility, verifiability, …

What aspects of QCD and Jets can be phrased as well-defined optimization problems?

Do we have deep enough physics principles and/or rich enough data sets (real or synthetic) such that machine learning will yield trustable answers?

Apologies that citations/examples in this talk are not (even close to) complete!

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

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

[see e.g. 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]

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: $L = \int \mathrm{d}x \, \mathcal{L}(x)$

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

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

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

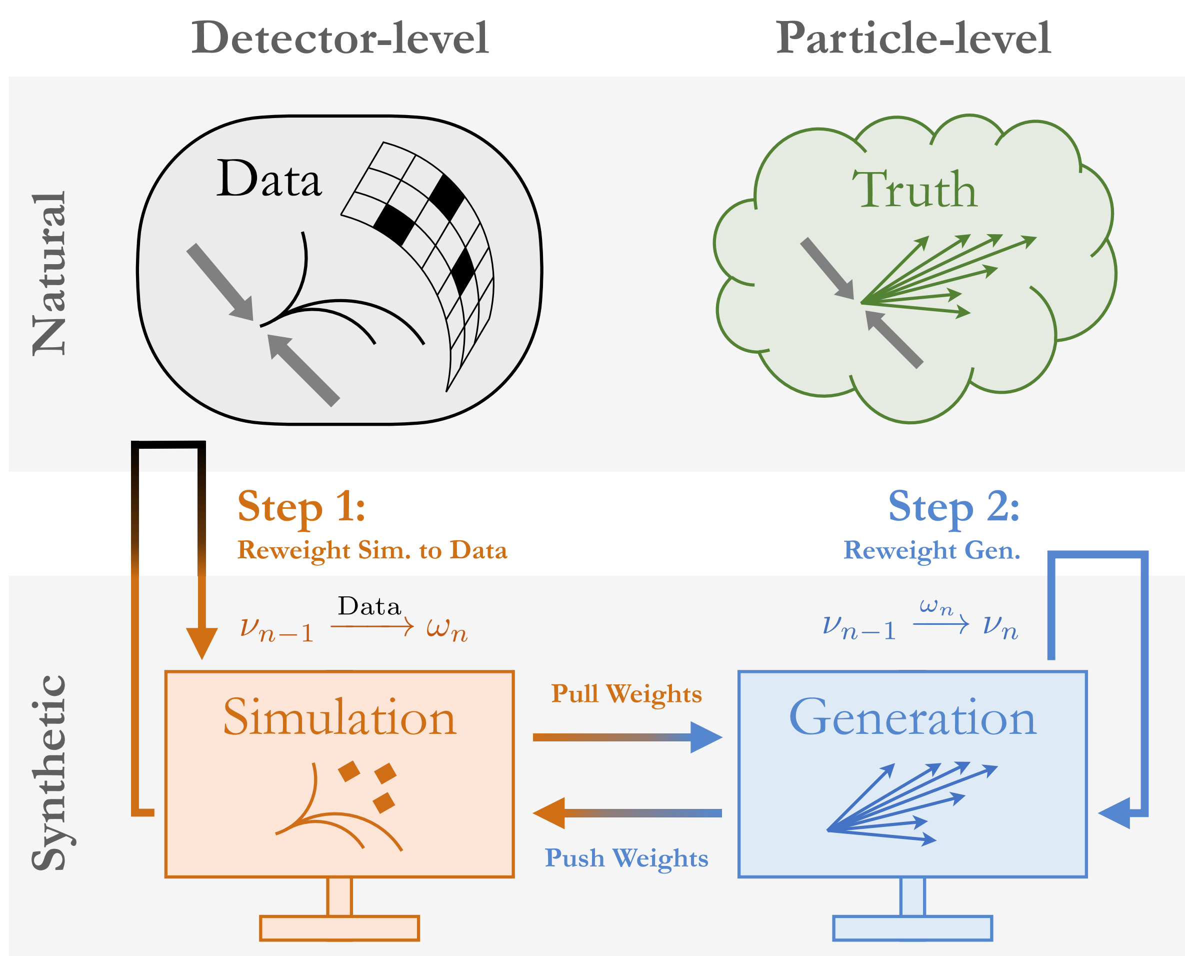
Requires shift in theoretical focus from solving problems to specifying problems

[see e.g. 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]

E.g. Detector Unfolding

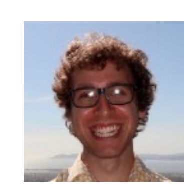
OmniFold

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



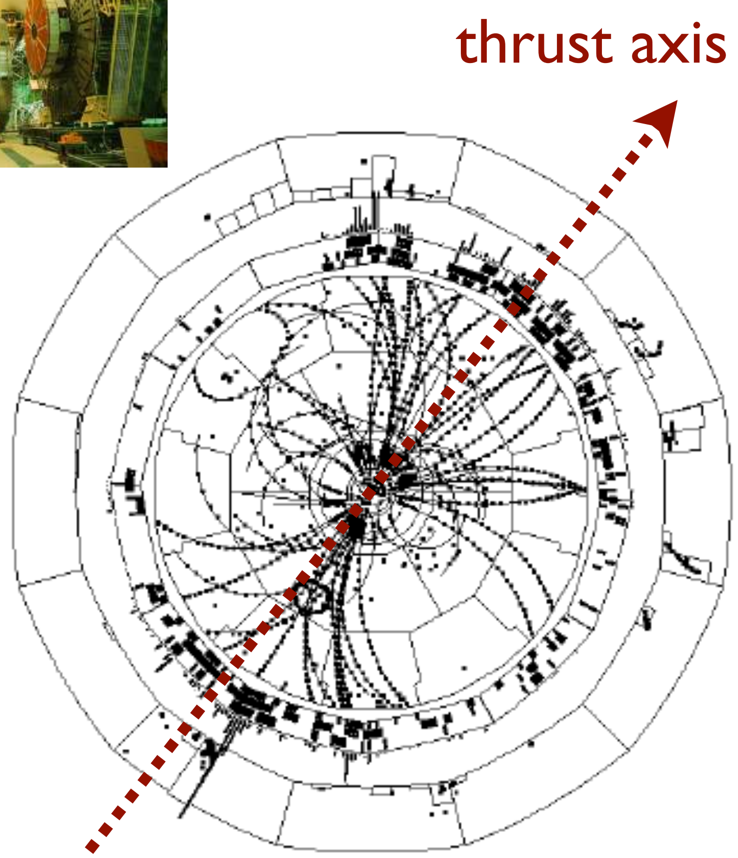
Use ML to compute reweighting factors

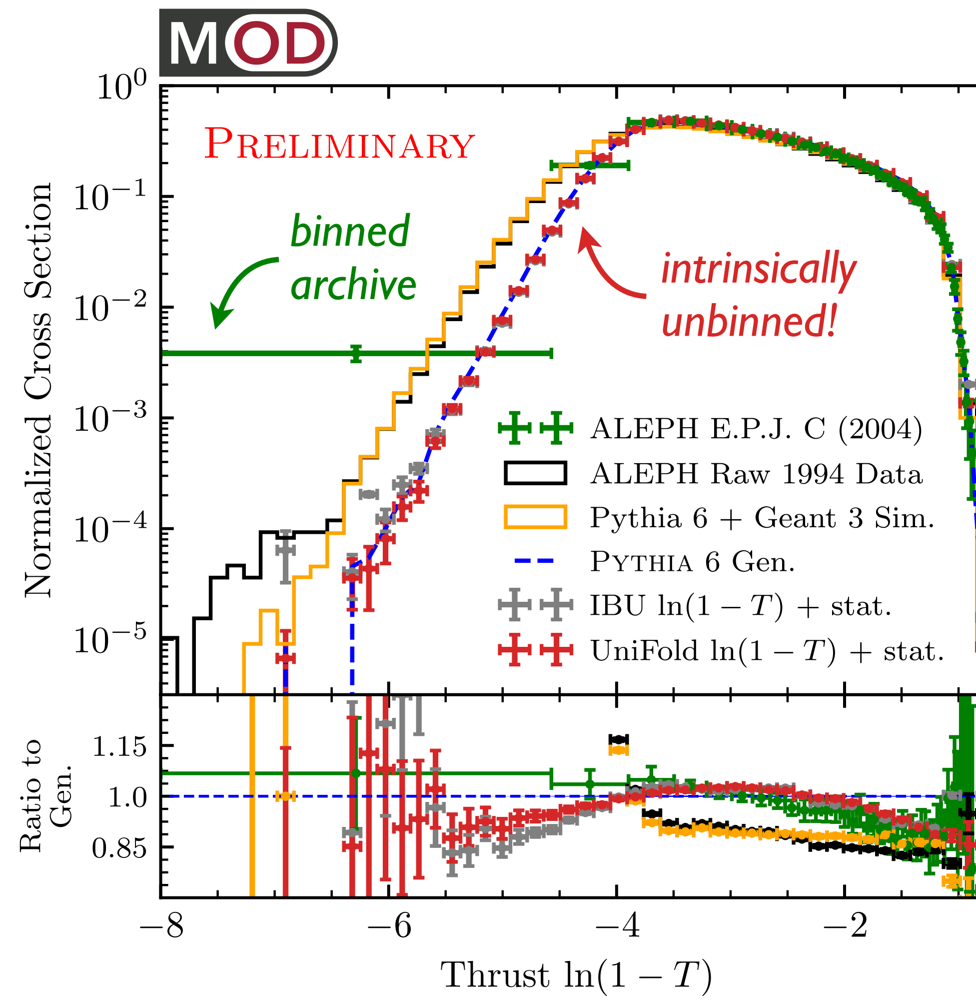
[Andreassen, Komiske, Metodiev, Nachman, JDT, PRL 2020]

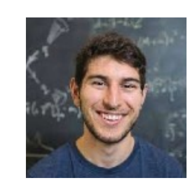



E.g. Detector Unfolding

Back to the Future with ALEPH Archival Data





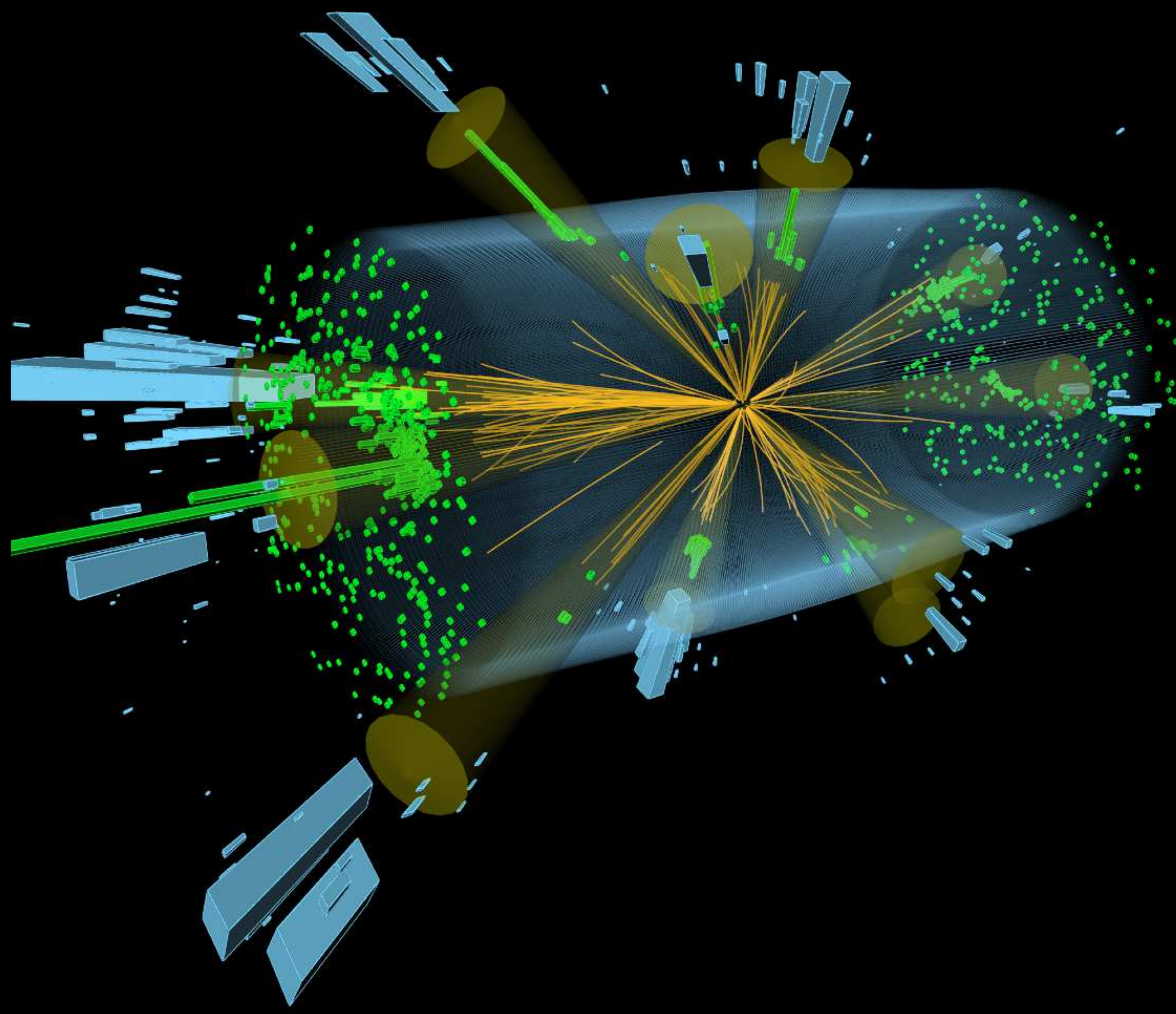


[talk by Badea, ICHEP 2020; cf. ALEPH, EPJC 2004]
[see also Badea, Baty, Chang, Innocenti, Maggi, McGinn, Peters, Sheng, JDT, Lee, PRL 2019]




[Andreassen, Komiske, Metodiev, Nachman, JDT, PRL 2020]

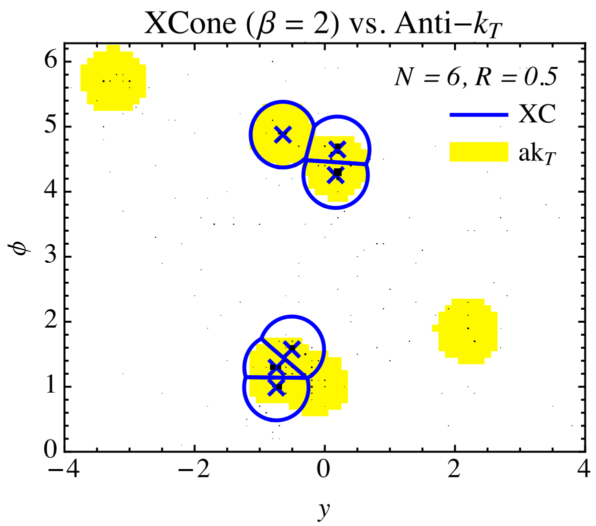
Machine Learning for QCD and Jets



What collider tasks can be phrased as optimization problems?

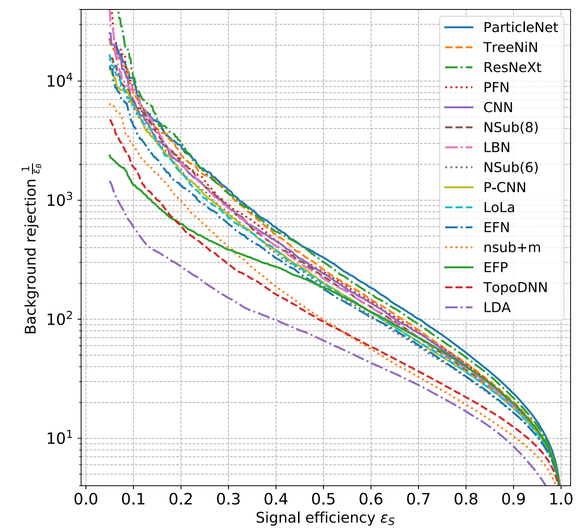
Optimization for QCD and Jets

Jet Algorithms



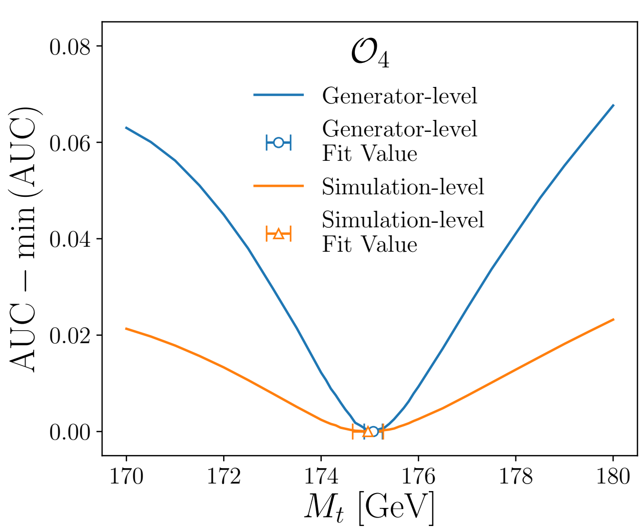
[e.g. Stewart, Tackmann, JDT, Vermilion, Wilkason, JHEP 2015]

Jet Classification



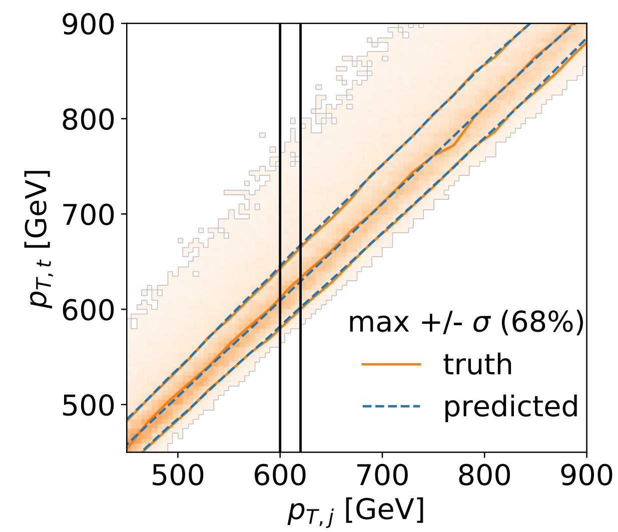
[e.g. Kasieczka, Plehn, et al., SciPost 2019]

Parameter Estimation



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

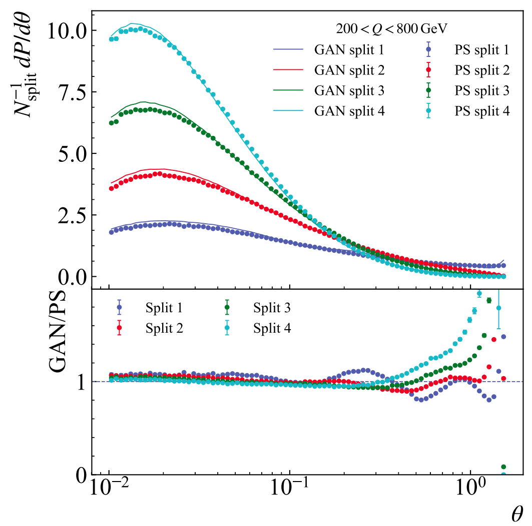
Uncertainty Quantification



[e.g. Kasieczka, Luchmann, Otterpohl, Plehn, SciPost 2020]

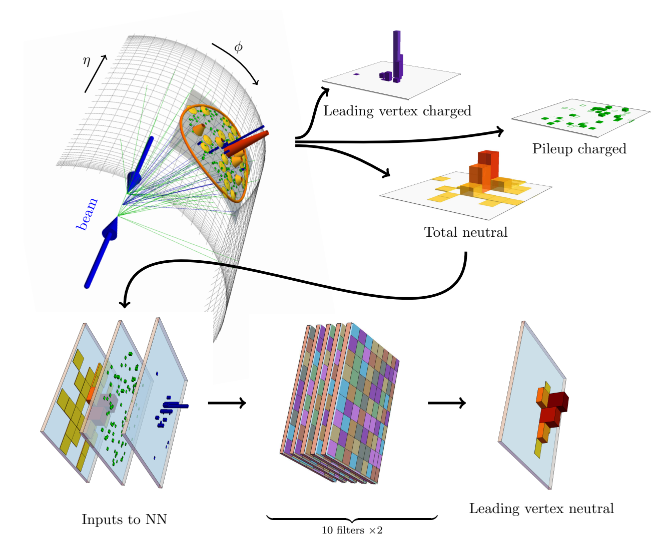
More Optimization for QCD and Jets

Parton Shower Modeling/Tuning



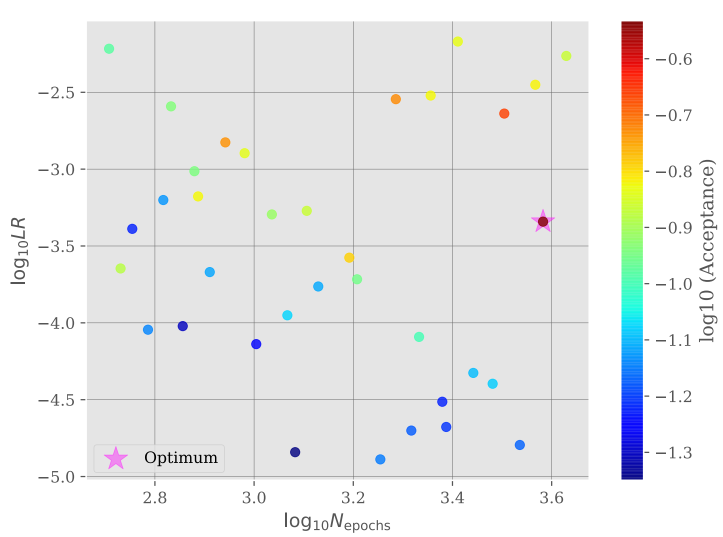
[e.g. Lai, Neill, Płoskoń, Ringer, arXiv 2020]

Pileup Mitigation



[e.g. Komiske, Metodiev, Nachman, Schwartz, JHEP 2017]

Phase Space Integration



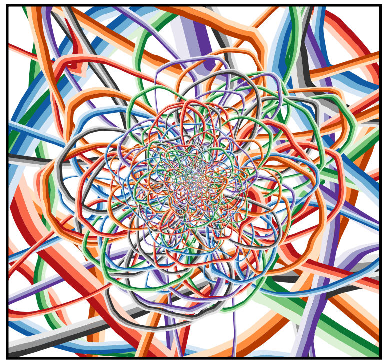
[e.g. Gao, Höche, Isaacson, Krause, Schulz, PRD 2020]

Amplitude Calculations

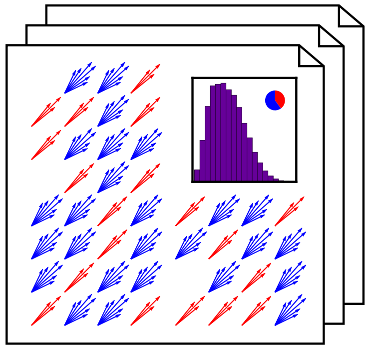
?

From Curmudgeon to Evangelist

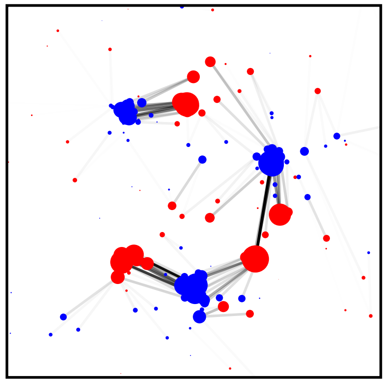
What have been helpful guides in pursing ML ⇔ QCD?



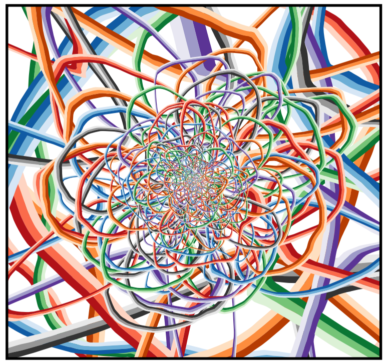
Can theoretical structures be encoded directly?



Can strategy be defined on physical final states?



Can we leverage unsupervised machine learning?



Can theoretical structures be encoded directly?

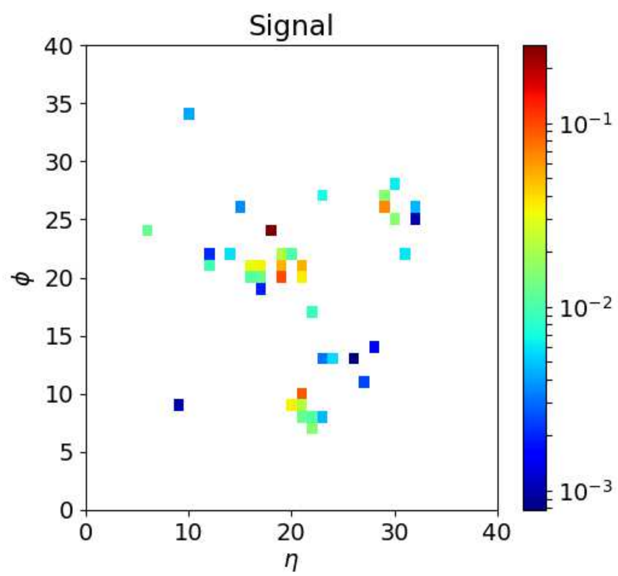
Can strategy be defined on physical final states?

Can we leverage unsupervised machine learning?

Jet Representations

Pixelized Image

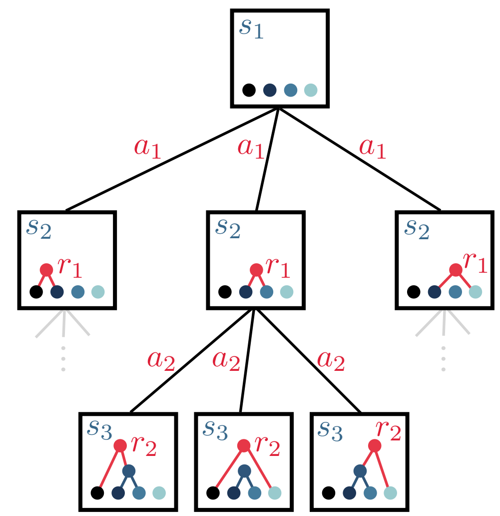
Calorimetry



[review in Kagan, arXiv 2020]

Hierarchical Tree

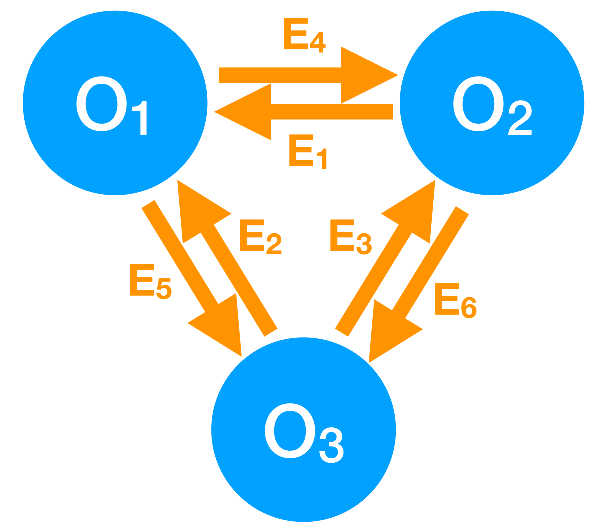
Binary Splittings



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

Graphs

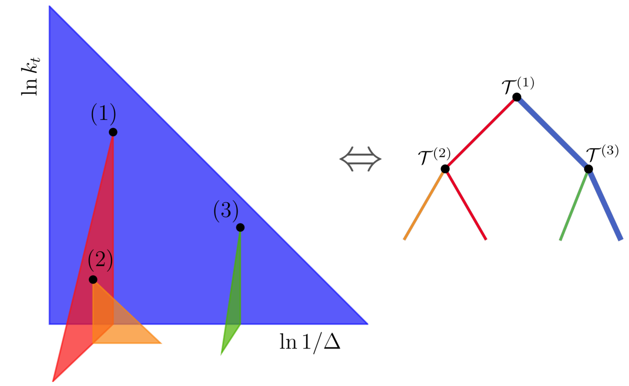
Pairwise Interactions



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

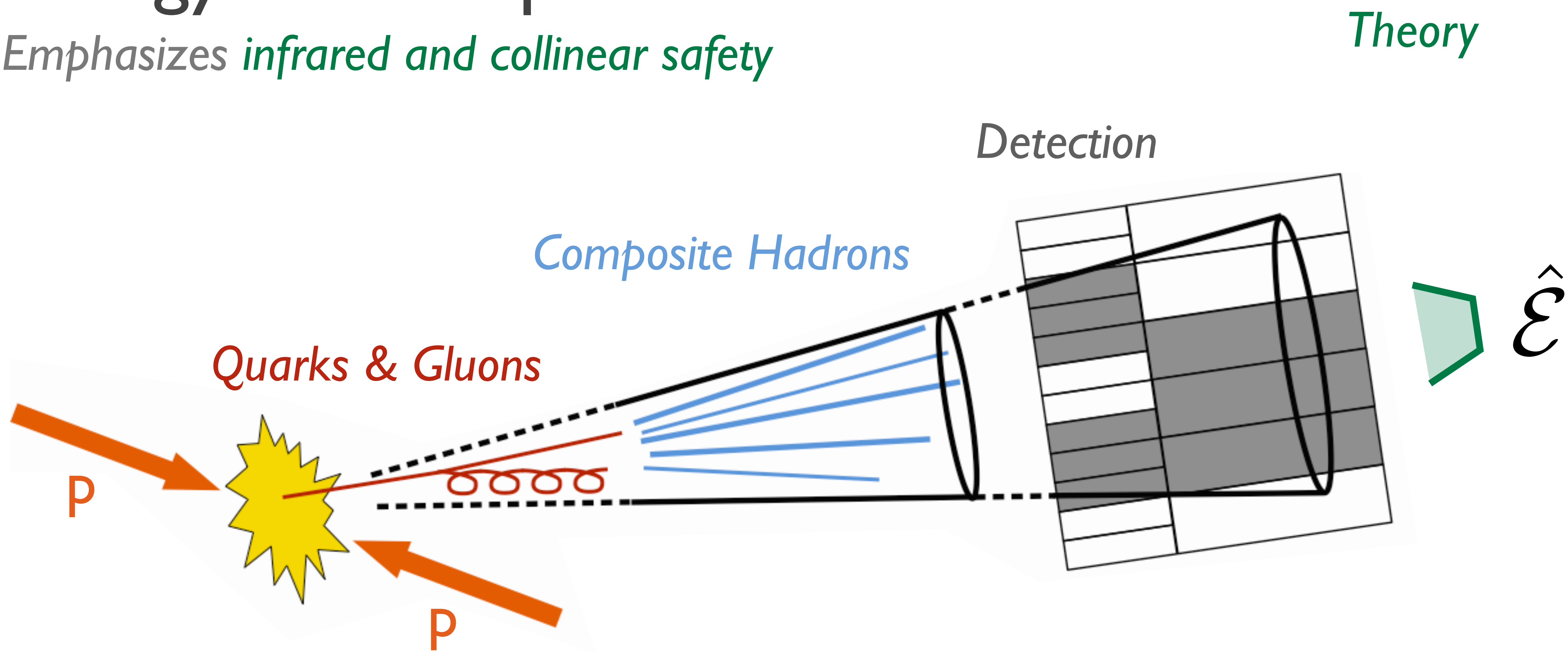
Imposes implicit theoretical prior; affects choice of network architecture

See Frédéric's talk for
Lund Plane + Graph Networks



Energy Flow Representation

Emphasizes infrared and collinear safety



Energy Flow:

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

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

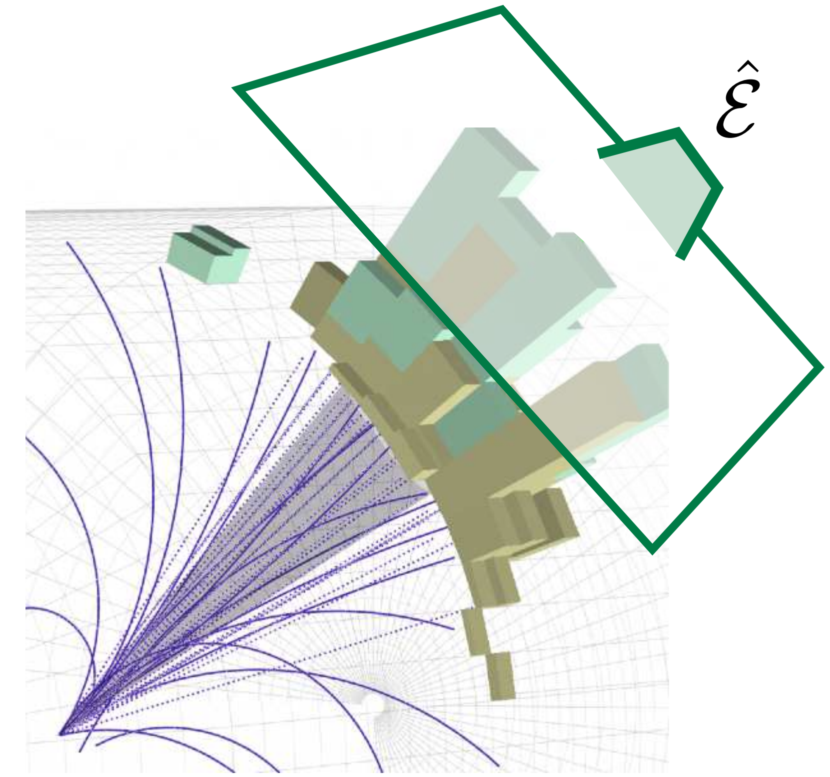
[see e.g. Sveshnikov, Tkachov, PLB 1996; Hofman, Maldacena, JHEP 2008; Mateu, Stewart, JDT, PRD 2013; Belitsky, Hohenegger, Korchemsky, Sokatchev, Zhiboedov, PRL 2014; Chen, Moult, Zhang, Zhu, PRD 2020]
[complementary perspective on IRC unsafe information in Chakraborty, Lim, Nojiri, Takeuchi, JHEP 2020]

Jets as Weighted Point Clouds

- Energy-Weighted Directions

$$\vec{p} = \{\underset{\uparrow \atop \text{Energy}}{E}, \underbrace{\hat{n}_x, \hat{n}_y, \hat{n}_z}_{\text{Direction}}\}$$

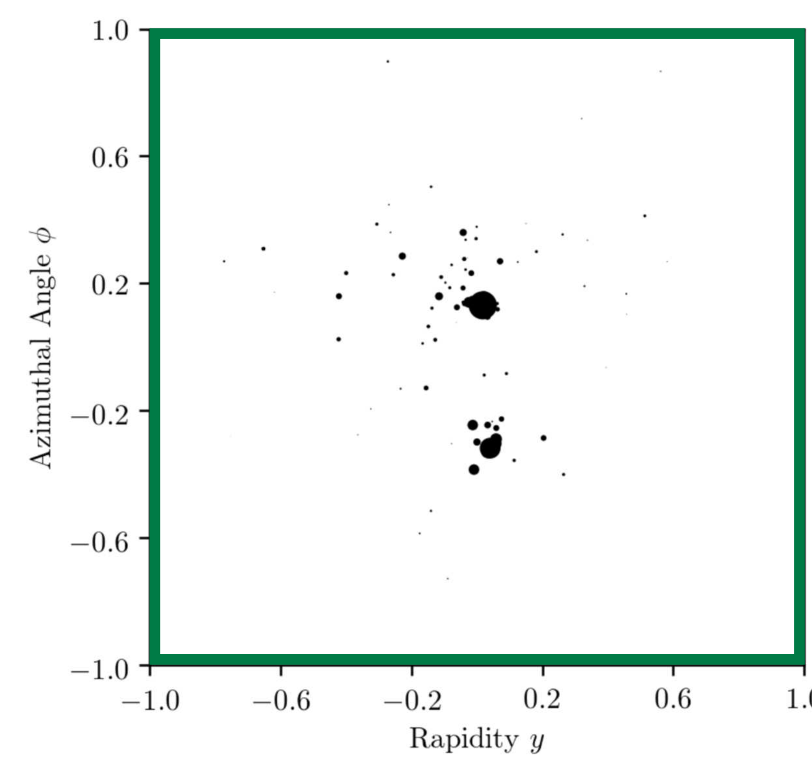
(suppressing "unsafe" charge/flavor information)



- 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

Permutation invariant

Linear weights (i.e. safe)

$$S(\mathcal{J}) = F(V_1, V_2, \ldots, V_\ell) \qquad V_a(\mathcal{J}) = \sum_{i \in \mathcal{J}} E_i \, \Phi_a(\hat{n}_i)$$

Parametrized with Neural Networks

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]

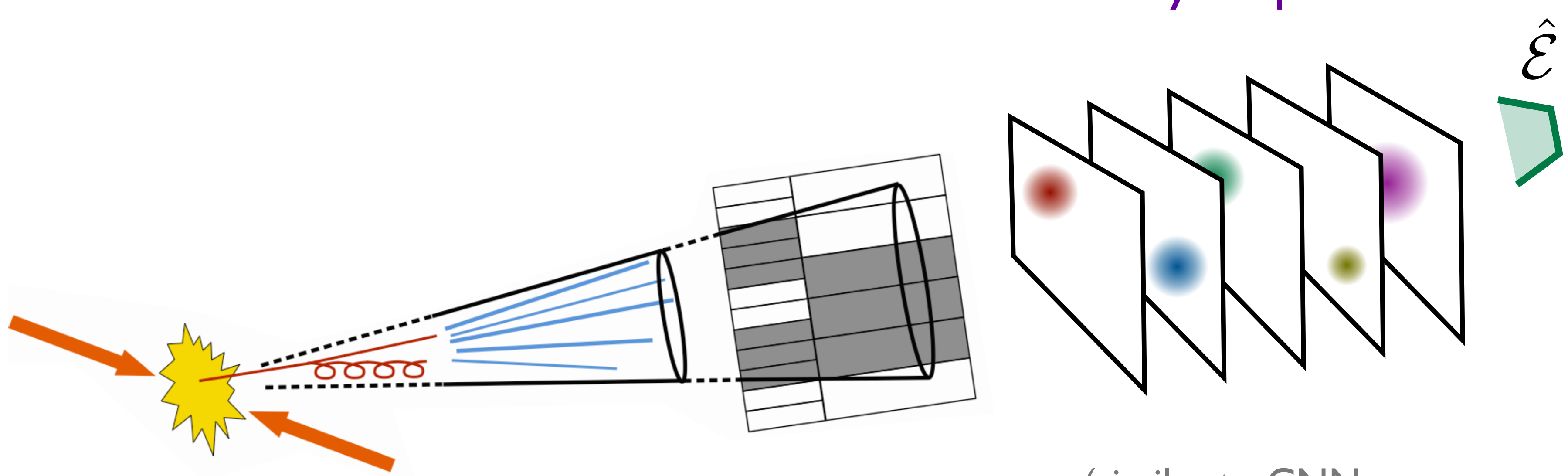
Energy Flow Networks

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

Latent space of dim ℓ

$$S(\mathcal{J}) = F\left(V_1, V_2, \ldots, V_\ell\right) \qquad V_a(\mathcal{J}) = \sum_{i \in \mathcal{J}} E_i \, \Phi_a(\hat{n}_i)$$

Easy to plot!

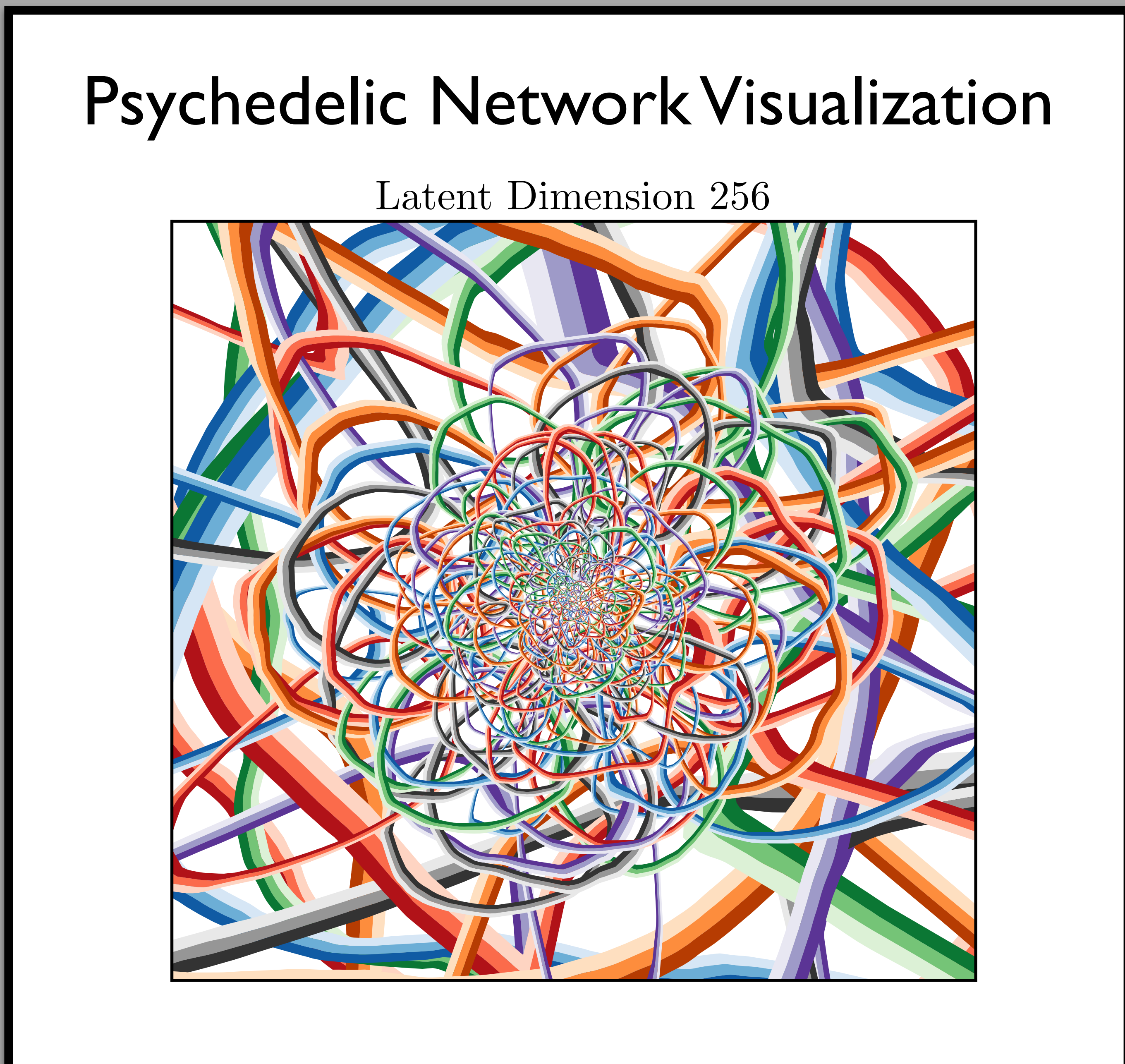


(similar to CNN filter activation)

[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]

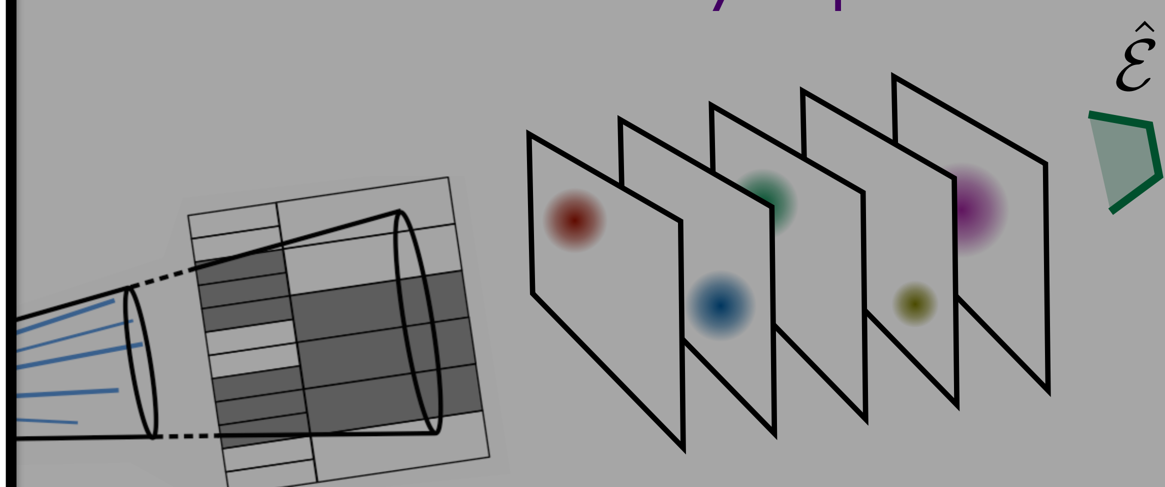
Energy Flow Networks

Architecture designed around symmetries and interpretability



$$V_a(\mathcal{J}) = \sum_{i \in \mathcal{J}} E_i \, \Phi_a(\hat{n}_i)$$

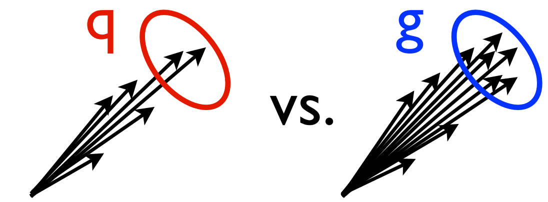
Easy to plot!

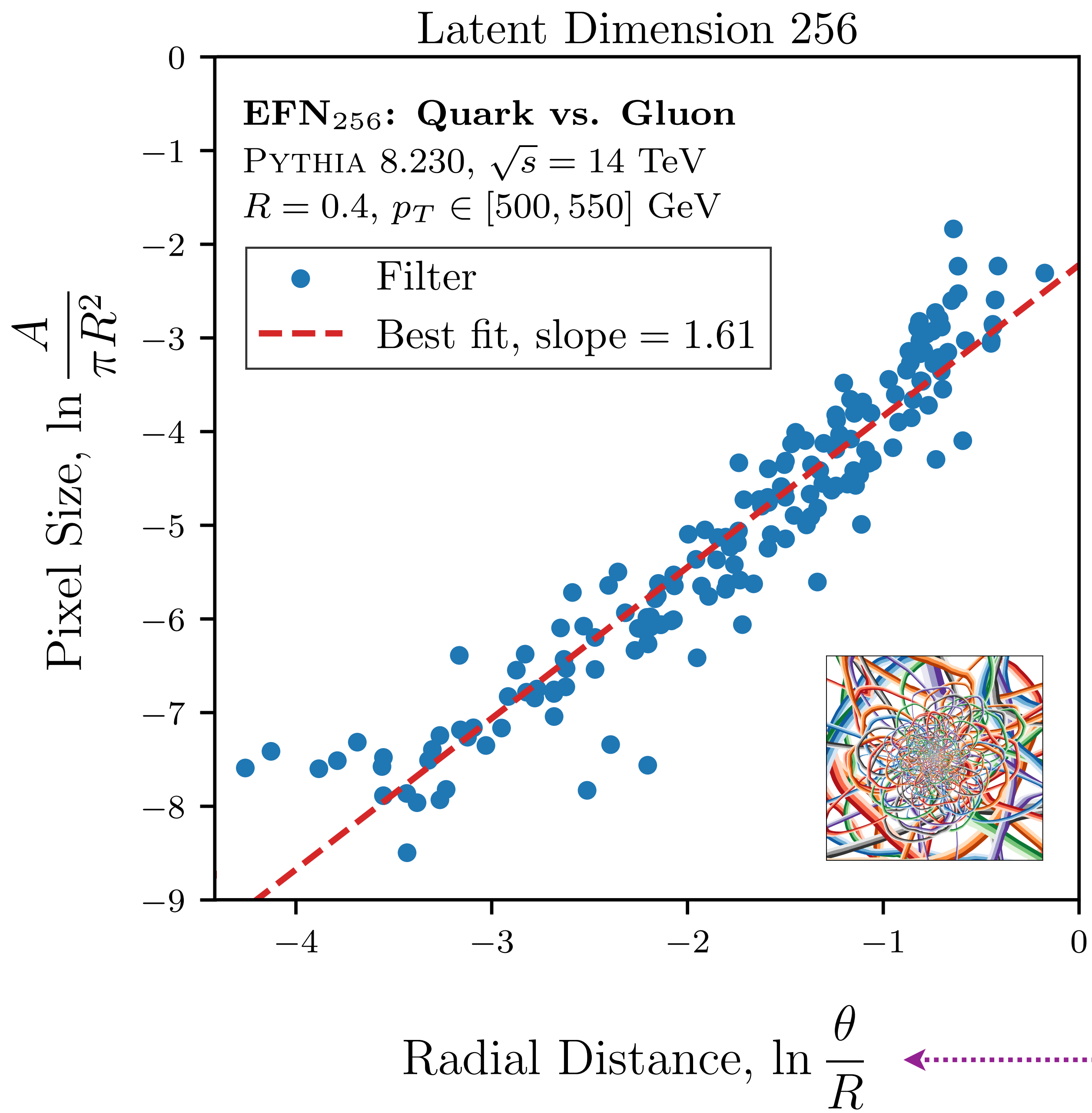


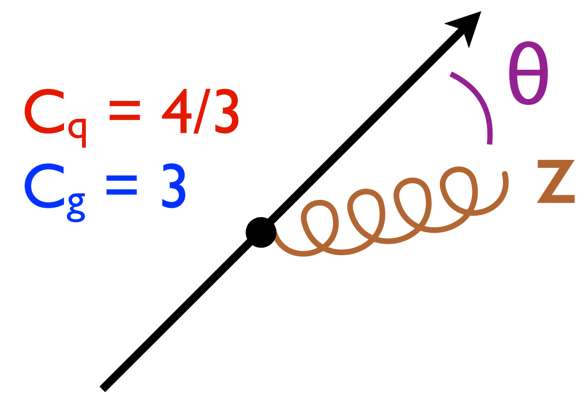
(similar to CNN filter activation)

[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



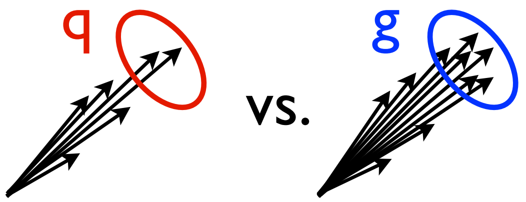


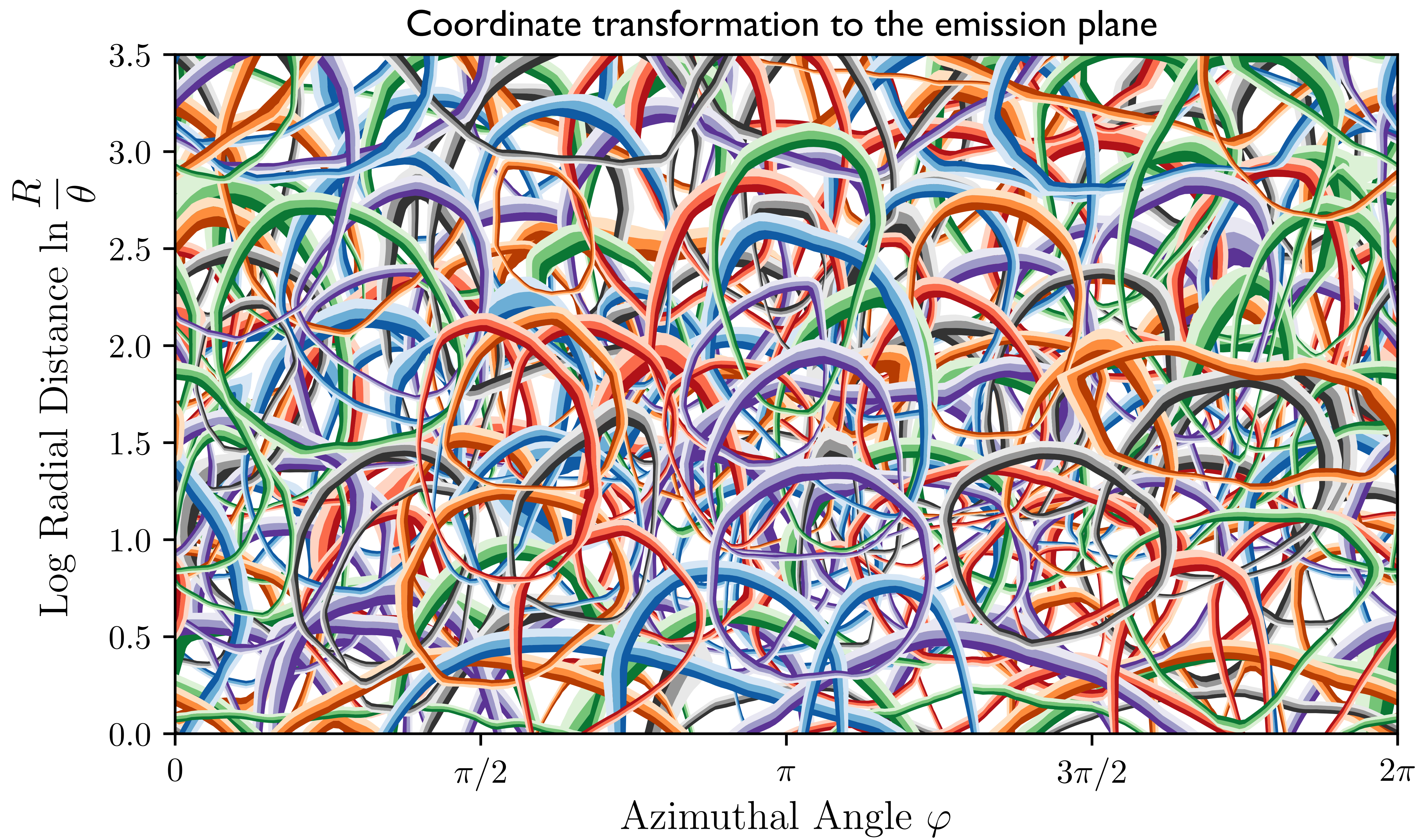


$$\mathrm{d}P_{i\to ig} \simeq \frac{2\alpha_s}{\pi} C_i \underbrace{\frac{\mathrm{d}\theta}{\theta}}_{\text{Collinear}} \underbrace{\frac{\mathrm{d}z}{z}}_{\text{Soft}}$$

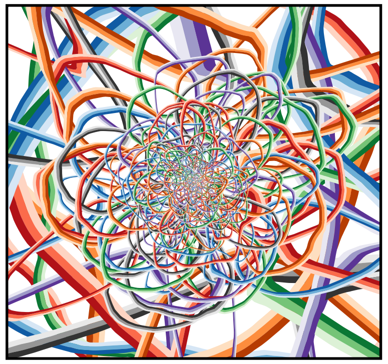
[Komiske, Metodiev, JDT, JHEP 2019]

En Route to the Lund Plane





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



Can theoretical structures be encoded directly?

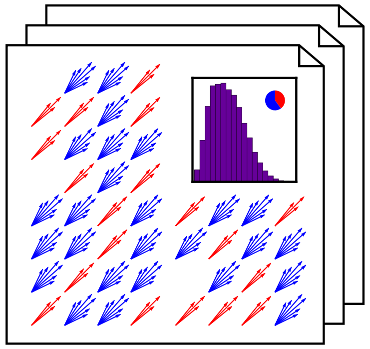
Energy Flow Networks ⇔ IRC Safety + Permutations

Can strategy be defined on physical final states?

Can we leverage unsupervised machine learning?

Can theoretical structures be encoded directly?

Energy Flow Networks ⇔ IRC Safety + Permutations

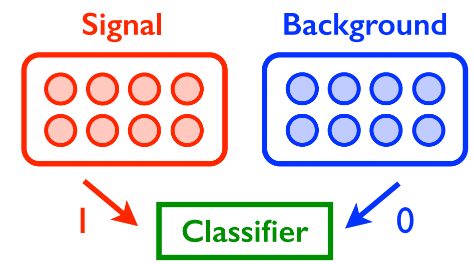


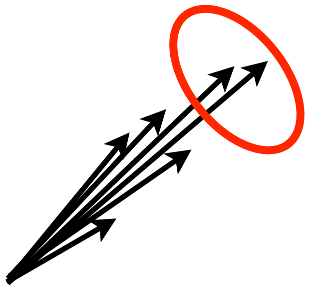
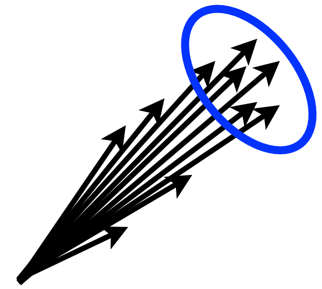
Can strategy be defined on physical final states?

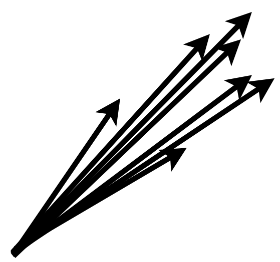
Can we leverage unsupervised machine learning?

Quark/Gluon Classification

"Hello, World!" of Jet Physics



e.g.  **Quark** $C_q = 4/3$ vs.  **Gluon** $C_g = 3 = 9/3$

Find $h\Big($  $\Big)$ such that

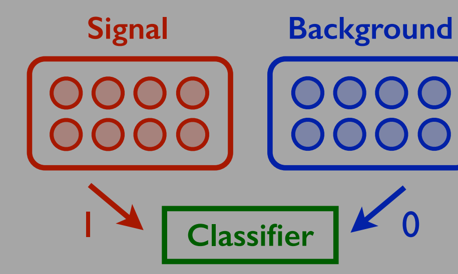
$$h(\text{Quark}) = 1$$

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

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

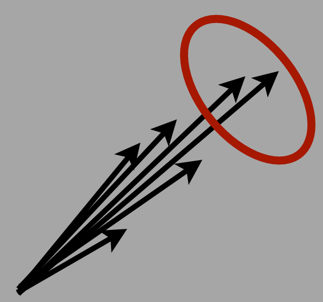
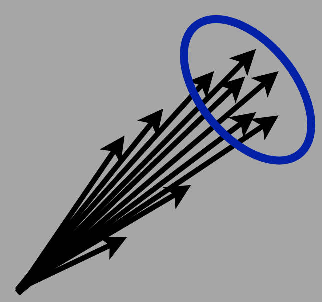
(Neyman-Pearson lemma)

[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

e.g.  **Quark** $C_q = 4/3$ vs.  **Gluon** $C_g = 3 = 9/3$

What do you mean by "quark" and "gluon"?

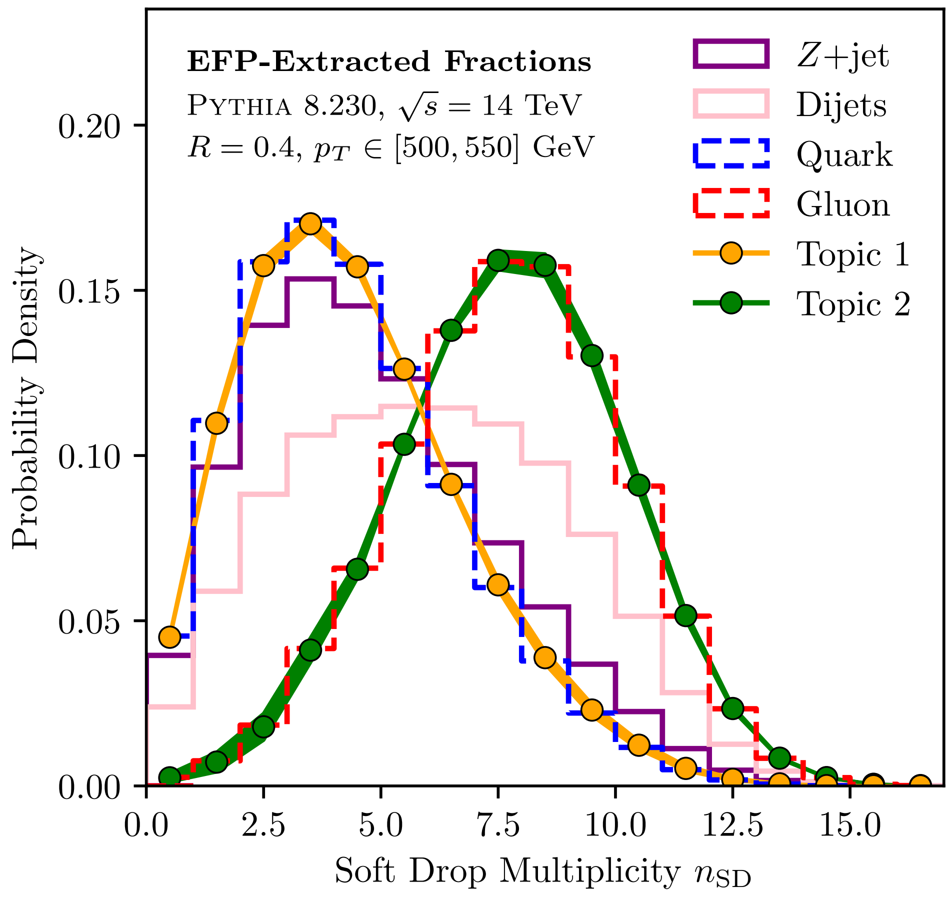
Jets are clusters of colorless hadrons!

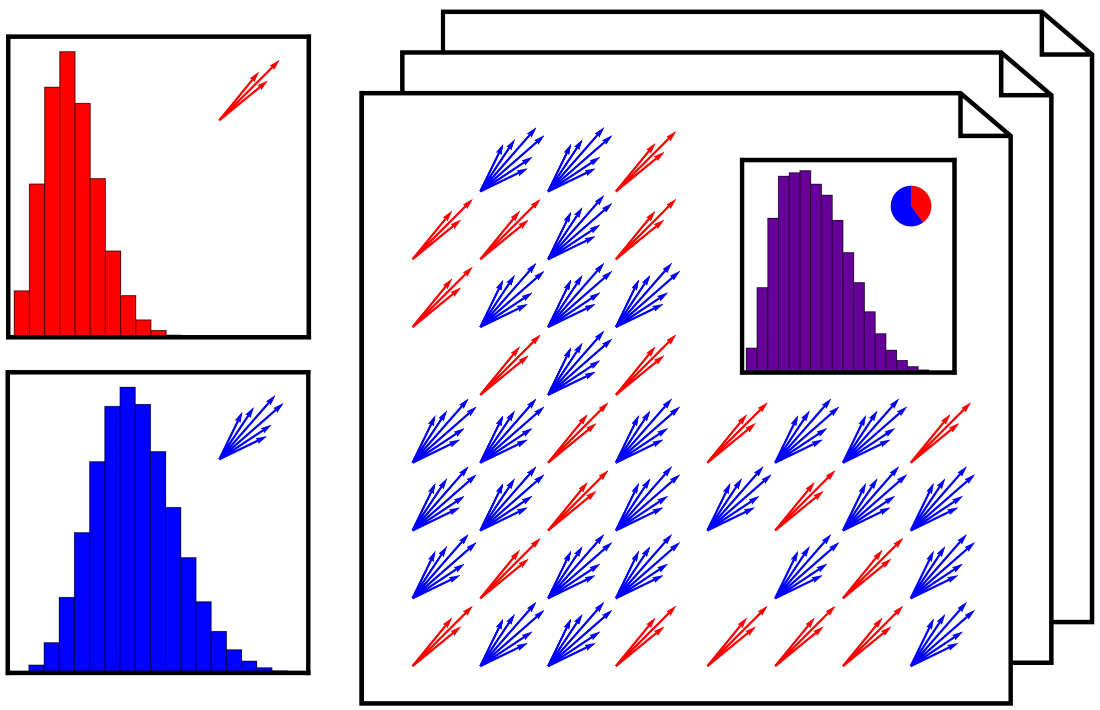
Parton shower "truth" is but a (useful) fiction!

[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



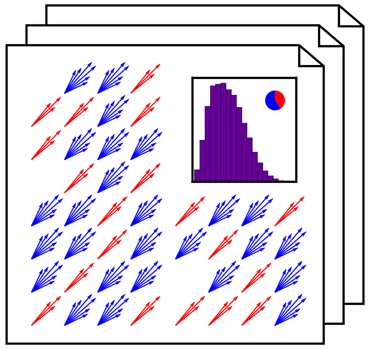


Key concept from natural language processing: "anchor words"

[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 ⇔ IRC Safety + Permutations



Can strategy be defined on physical final states?

Jet Topics ⇔ Hadron-Level Approach to QCD Partons

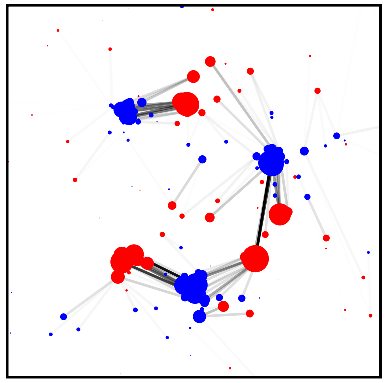
Can we leverage unsupervised machine learning?

Can theoretical structures be encoded directly?

Energy Flow Networks ⇔ IRC Safety + Permutations

Can strategy be defined on physical final states?

Jet Topics ⇔ Hadron-Level Approach to QCD Partons



Can we leverage unsupervised machine learning?

The Earth Mover's Distance

Optimal Transport: Minimum "work" (stuff x distance) to make one distribution look like another distribution

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

Déblai

Remblai

[h/t Niles-Weed, ML4Jets 2020; Monge, 1781; Vaseršteĭn, 1969; Wikipedia]

The Earth Mover's Distance

Optimal Transport: Minimum "work" (stuff x distance) to make one distribution look like another distribution

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

Additive Weight:

Ground Distance:

$\Rightarrow$ Distance Between Distributions

[h/t Niles-Weed, ML4Jets 2020; Monge, 1781; Vaseršteĭn, 1969; Wikipedia]

The Earth Mover's Distance

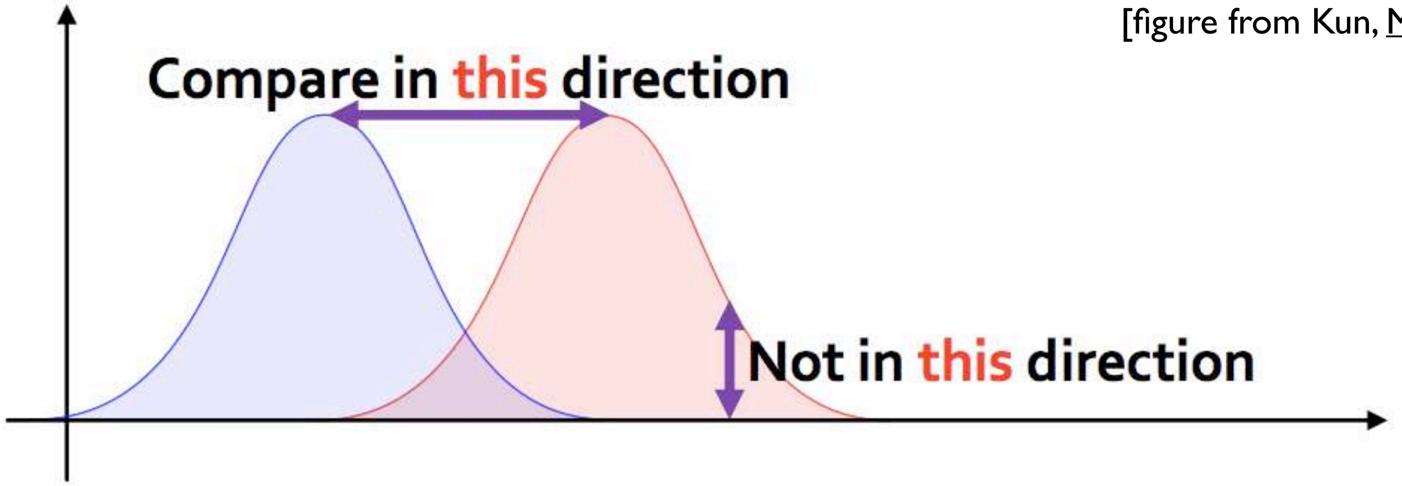
Optimal Transport: Minimum "work" (stuff x distance) to make one distribution look like another distribution

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

Additive Weight:

Ground Distance:

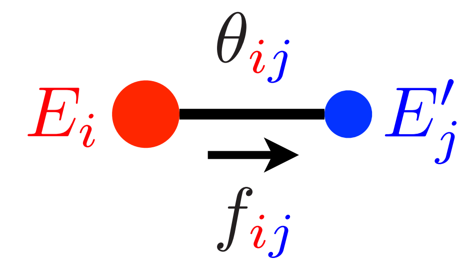
⇒ Distance Between Distributions

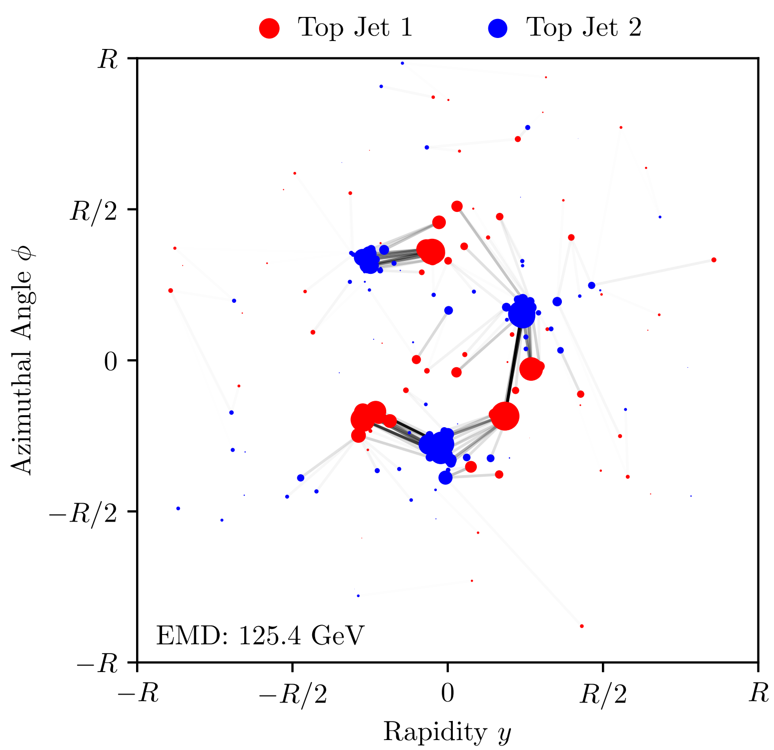


[figure from Kun, Math ∩ Programming]

[h/t Niles-Weed, ML4Jets 2020; Monge, 1781; Vaseršteĭn, 1969; Wikipedia]

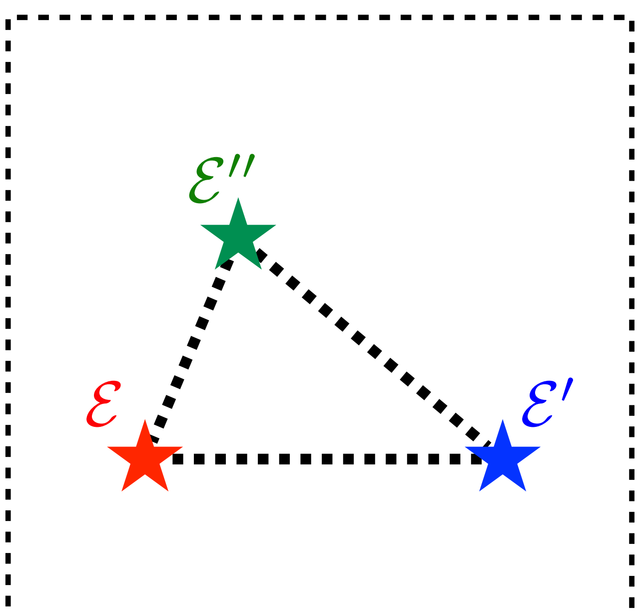
The Energy Mover's Distance





Optimal transport between energy flows…

$$\underset{\text{in GeV}}{\mathrm{EMD}(\mathcal{E},\mathcal{E}')} = \underbrace{\min_{\{f\}} \sum_i \sum_j f_{ij} \frac{\theta_{ij}}{R}}_{\text{Cost to move energy}} + \underbrace{\left| \sum_i E_i - \sum_j E'_j \right|}_{\text{Cost to create energy}}$$



…defines a metric on the space of events

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

(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

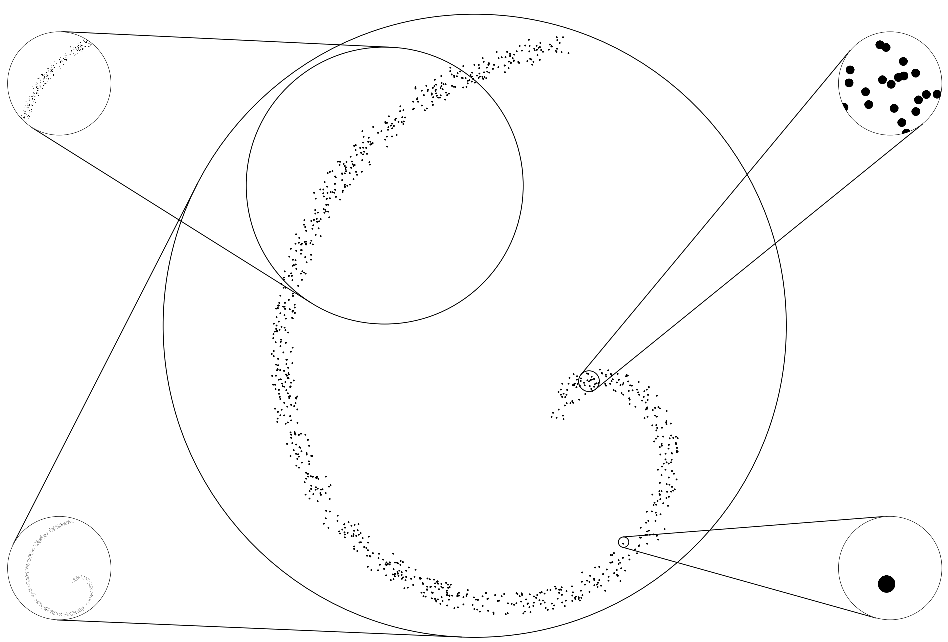
$$N_{\text{neighbors}}(r) \sim r^{\text{dim}}$$

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

[Grassberger, Procaccia, PRL 1983; Kégl, NIPS 2002]

dim≃1

dim≃2



dim≃2

(eventually 0)

dim≃0

Dimensionality of Space of Jets

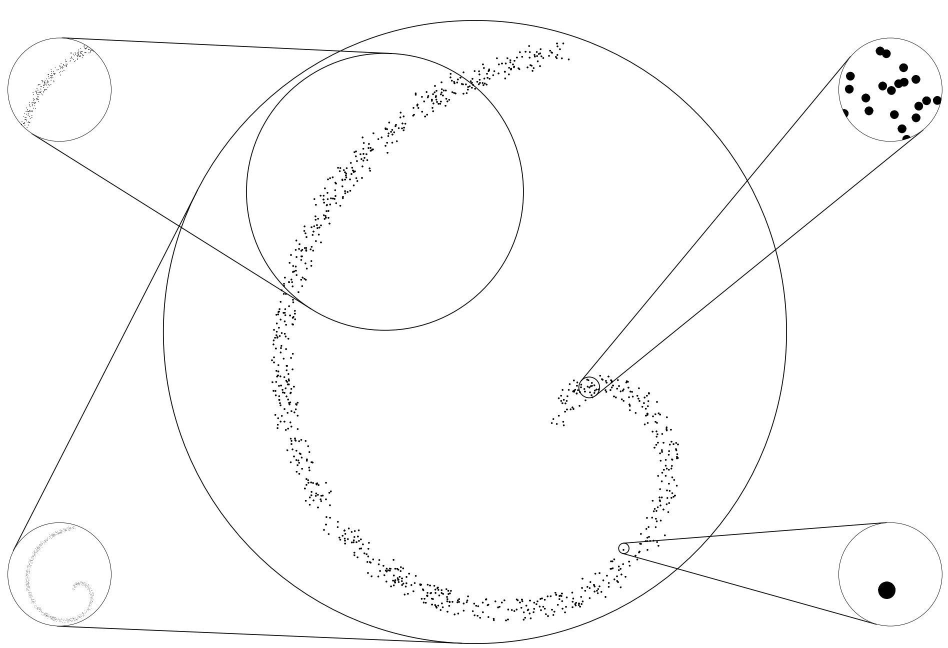
$$N_{\text{neighbors}}(r) \sim r^{\text{dim}}$$

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

[Grassberger, Procaccia, PRL 1983; Kégl, NIPS 2002]

dim≃1

dim≃2

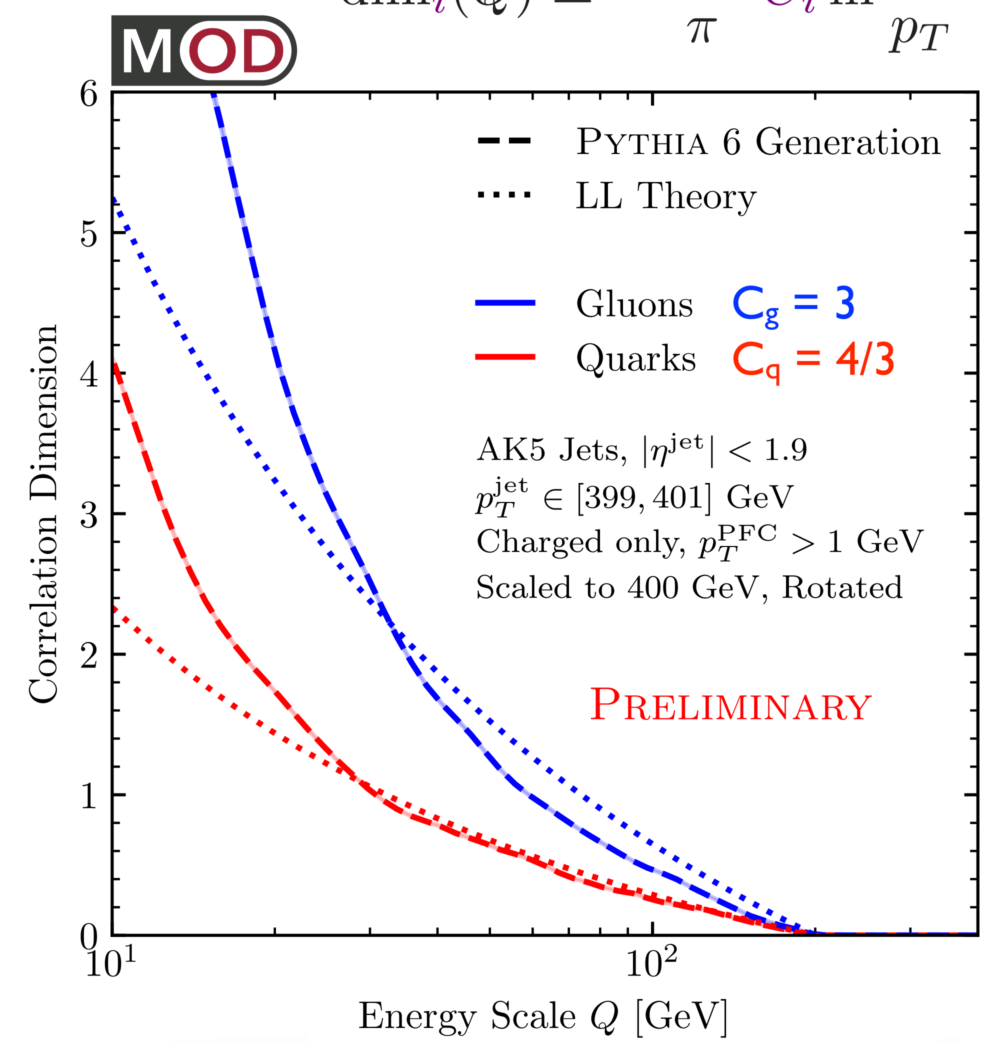


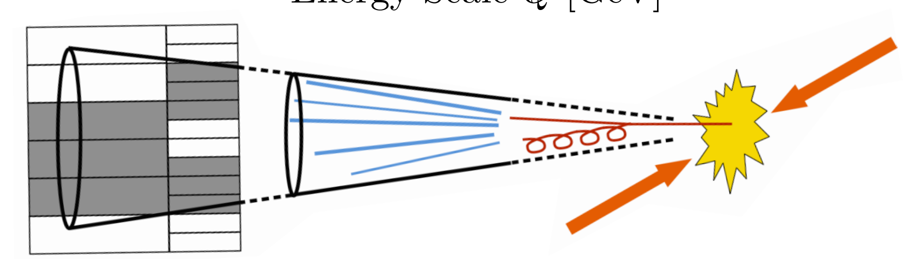
dim≃2

(eventually 0)

dim≃0

$$\text{dim}_i(Q) \simeq -\frac{8\alpha_s}{\pi} C_i \ln \frac{Q}{p_T}$$



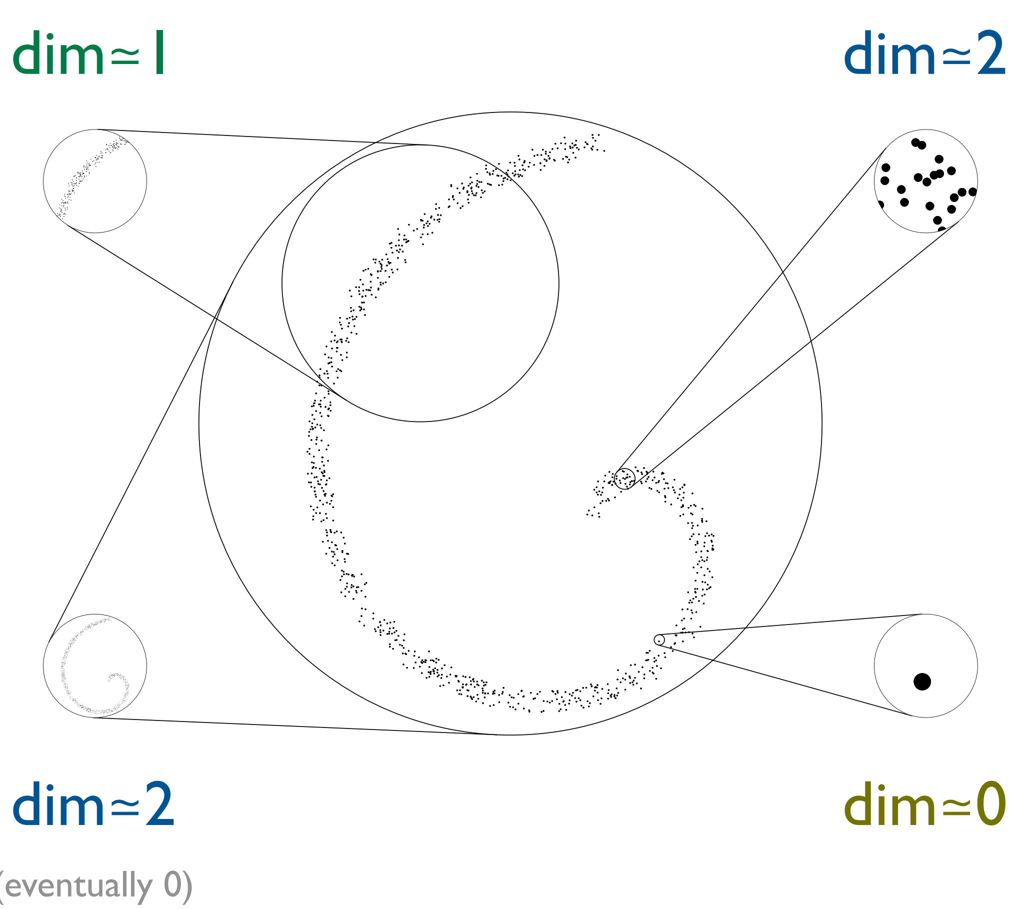


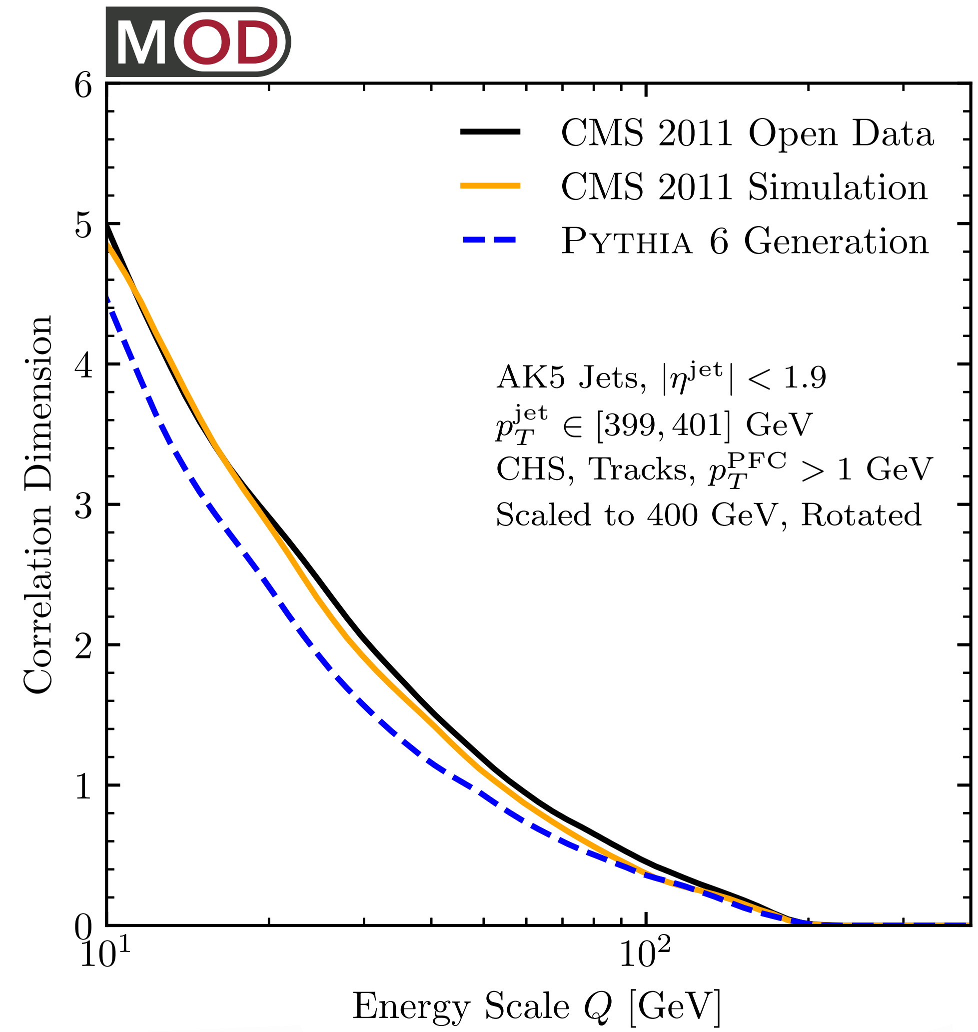
Dimensionality of Space of Jets

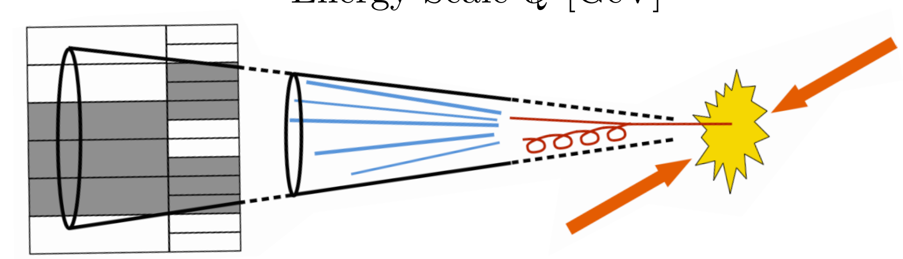
$$N_{\text{neighbors}}(r) \sim r^{\text{dim}}$$

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

[Grassberger, Procaccia, PRL 1983; Kégl, NIPS 2002]





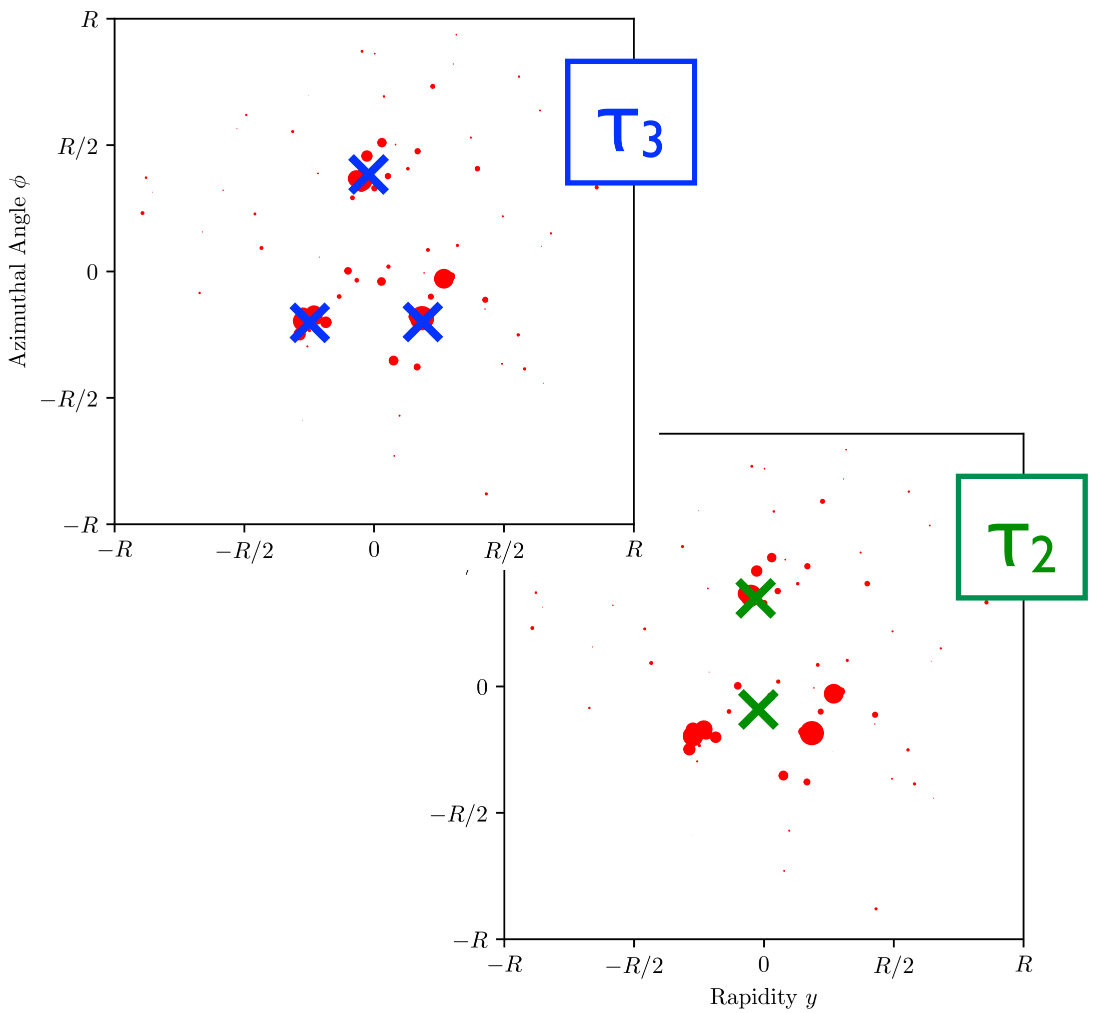


[Komiske, Mastandrea, Metodiev, Naik, JDT, PRD 2020; using CMS Open Data]

N-subjettiness

Ubiquitous jet substructure observable used for almost a decade…

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

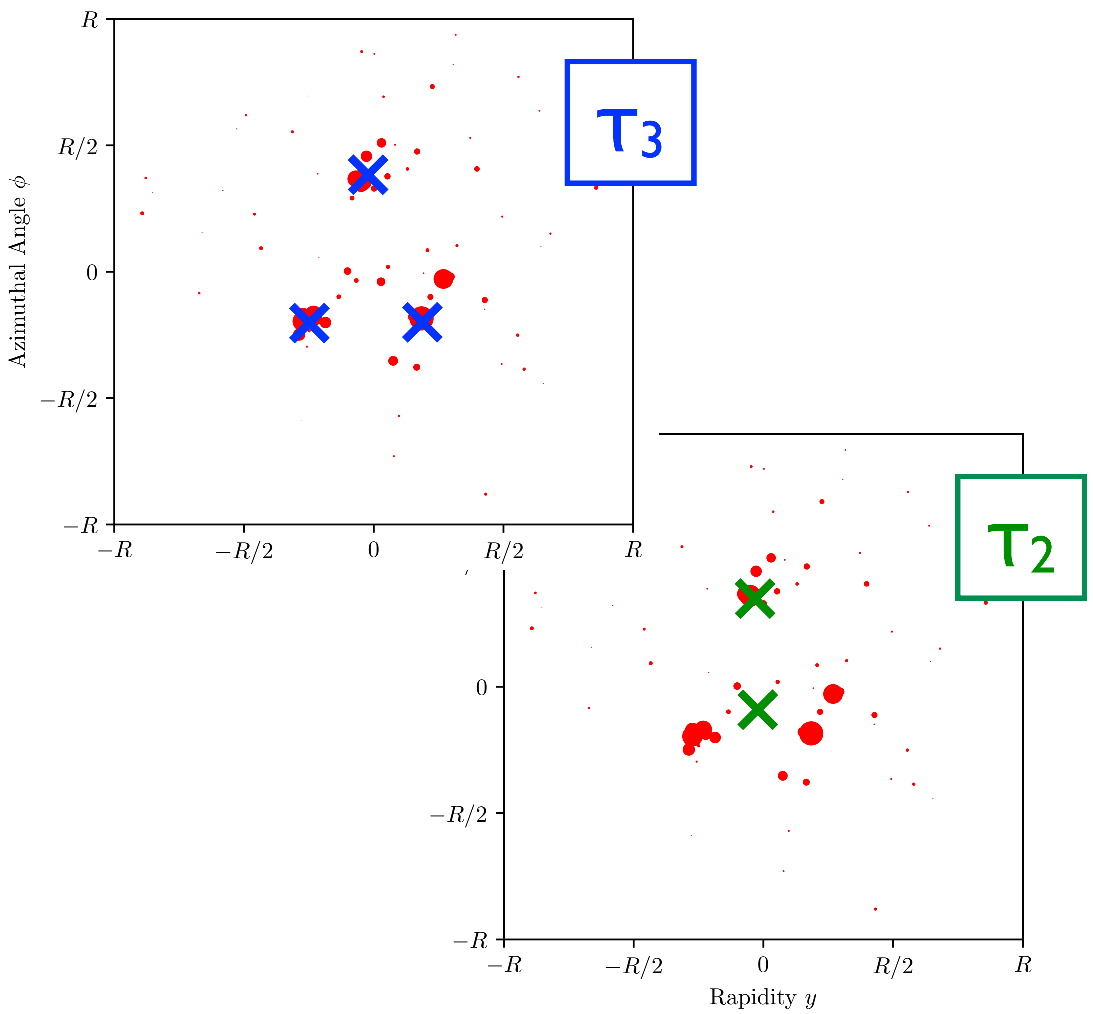


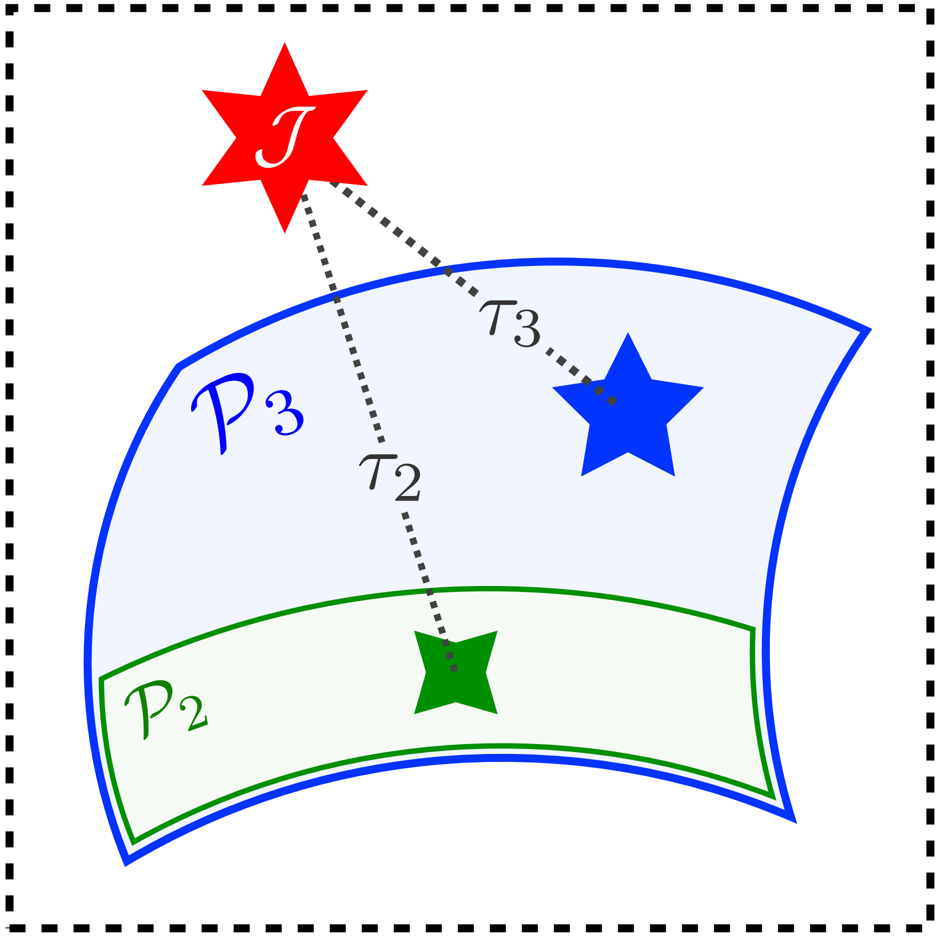
[JDT, Van Tilburg, JHEP 2011, JHEP 2012;
based on Brandt, Dahmen, ZPC 1979; Stewart, Tackmann, Waalewijn, PRL 2010]

N-subjettiness = Point to Manifold EMD

…is secretly an optimal transport problem

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

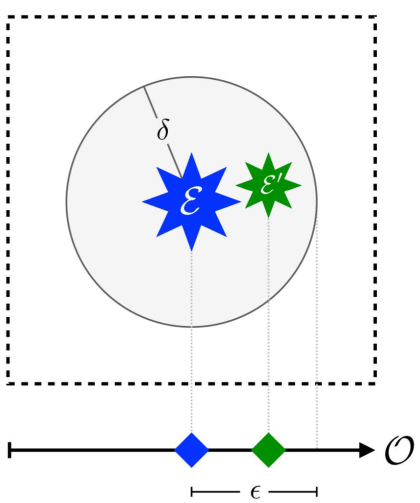




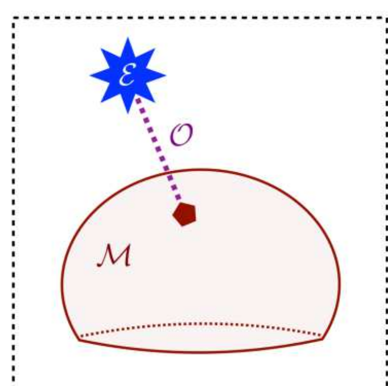
[JDT, Van Tilburg, JHEP 2011, JHEP 2012;
rephrased via Komiske, Metodiev, JDT, JHEP 2020; see opposite limit in Cesarotti, JDT, JHEP 2020]

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

IRC Safety is smoothness in the space of events

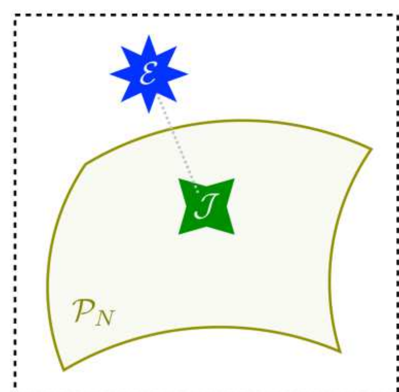


Event shapes are distances from events to manifolds.



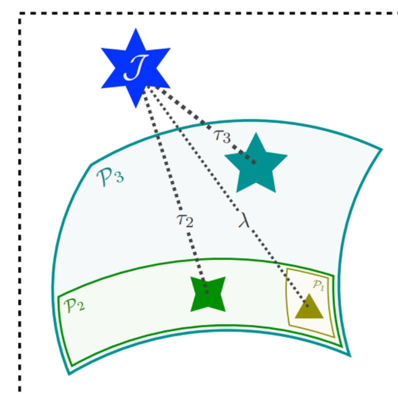
$\mathcal{O}(\mathcal{E}) = \min_{\mathcal{E}'\in\mathcal{M}} \mathrm{EMD}_{\beta,R}(\mathcal{E},\mathcal{E}')$

Jets are projections to few-particle manifolds.



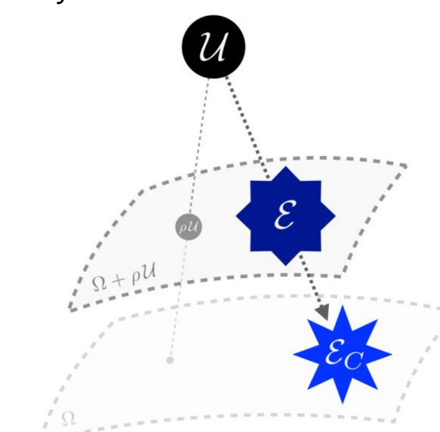
$\mathcal{J} = \mathrm{argmin}_{\mathcal{E}'\in\mathcal{P}_N} \mathrm{EMD}_{\beta,R}(\mathcal{E},\mathcal{E}')$

Substructure resolves emissions within the jet.



$\tau(\mathcal{J}) = \min_{\mathcal{E}'\in\mathcal{P}_N} \mathrm{EMD}_{\beta}(\mathcal{J},\mathcal{E}')$.

Pileup mitigation moves away from uniform radiation.



$\mathcal{E}_C = \arg\min_{\mathcal{E}'} \mathrm{EMD}(\mathcal{E},\mathcal{E}' + \rho\,\mathcal{U})$.

Taming infinities — Event Shapes — Jet Algorithms — Jet Substructure — Pileup

1960 — 2020

1962-1964
Infrared Safety
[Kinoshita, JMP 1962]
[Lee, Nauenberg, PR 1964]

1977
Thrust, Sphericity
[Farhi, PRL 1977]
[Georgi, Machacek, PRL 1977]

1993
k_T jet clustering
[Ellis, Soper, PRD 1993]
[Catani, Dokshitzer, Seymour, Webber, NPB 1993]

1997-1998
C/A jet clustering
[Wobisch, Wengler, 1998]
[Doskhitzer, Leder, Moretti, Webber, JHEP 1997]

2010-2015
N-(sub)jettiness, XCone
[Stewart, Tackmann, Waalewijn, PRL 2010]
[Thaler, Van Tilburg, JHEP 2011]
[Stewart, Tackmann, Thaler, Vermilion, Wilkason, JHEP 2015]

2014-2019
Constituent Subtraction
[Berta, Spousta, Miller, Leitner, JHEP 2014]
[Berta, Masetti, Miller, Spousta, JHEP 2019]

And many more!

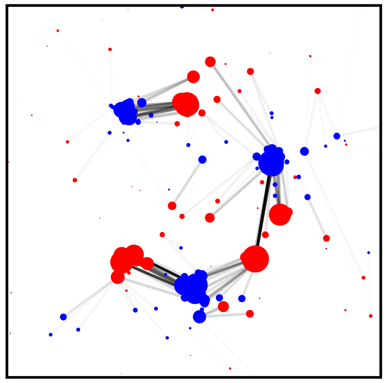
[Komiske, Metodiev, JDT, JHEP 2020; timeline by Metodiev]

Can theoretical structures be encoded directly?

Energy Flow Networks ⇔ IRC Safety + Permutations

Can strategy be defined on physical final states?

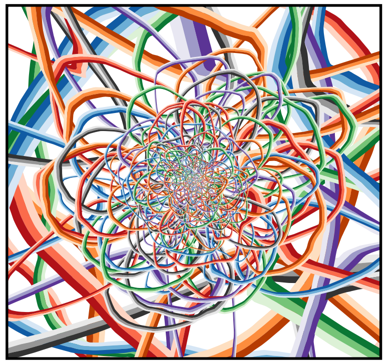
Jet Topics ⇔ Hadron-Level Approach to QCD Partons



Can we leverage unsupervised machine learning?

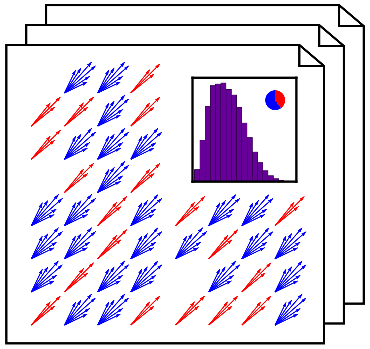
Energy Mover's Distance ⇔ Geometric Strategies for Collider Physics

QCD and Jets through the Lens of ML



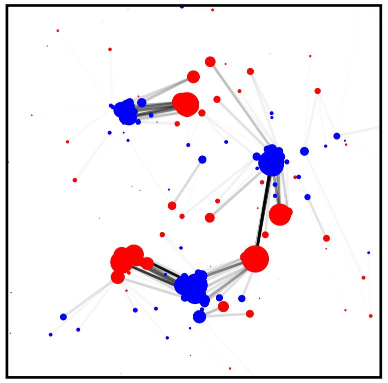
Can theoretical structures be encoded directly?

Energy Flow Networks ⇔ IRC Safety + Permutations



Can strategy be defined on physical final states?

Jet Topics ⇔ Hadron-Level Approach to QCD Partons



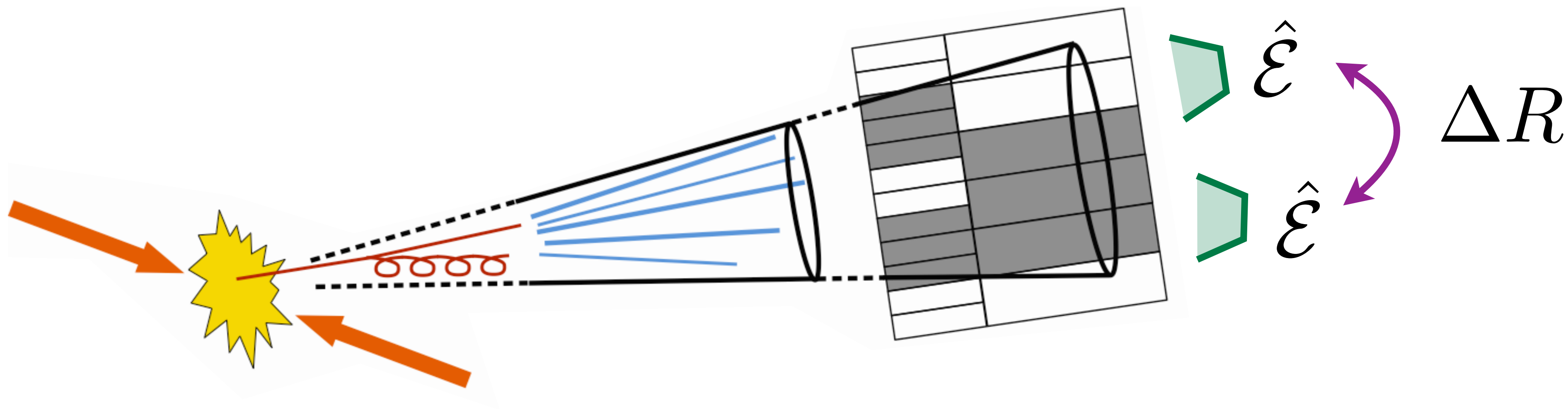
Can we leverage unsupervised machine learning?

Energy Mover's Distance ⇔ Geometric Strategies for Collider Physics

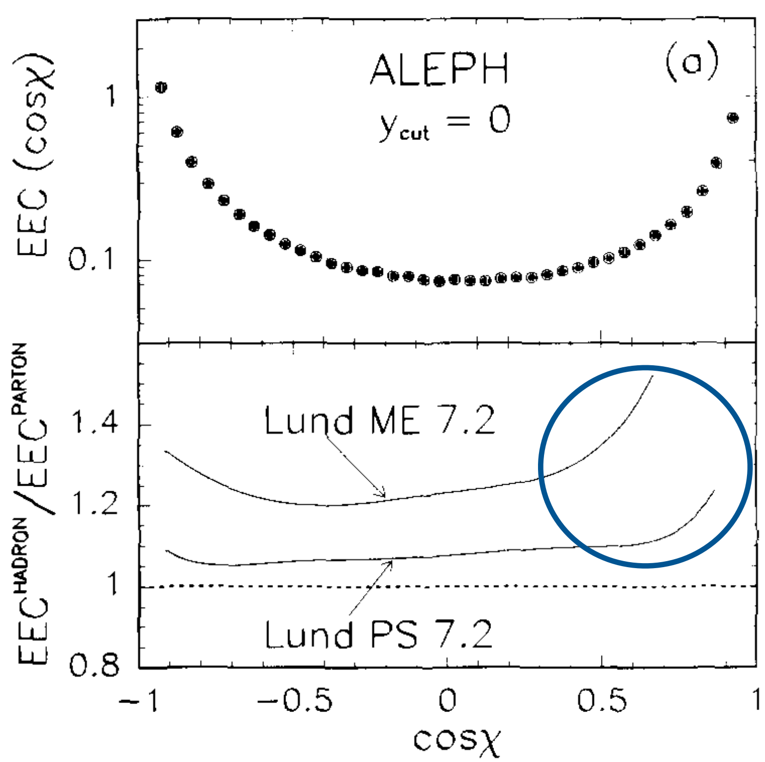
And now… Frédéric's perspective on ML ⇔ QCD!

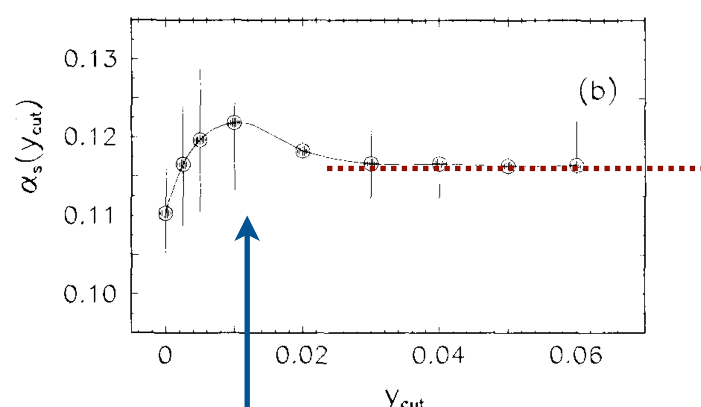
Backup Slides

Energy-Energy Correlators



A long history in probing collinear dynamics of QCD



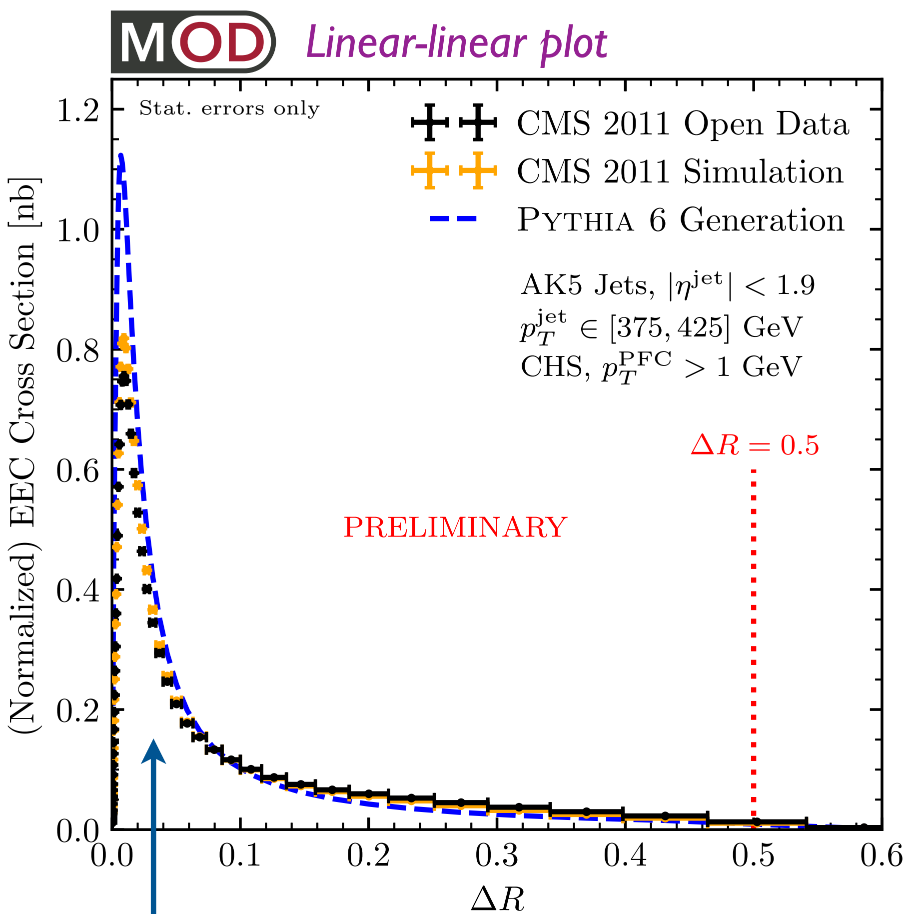


Extracting the strong coupling constant

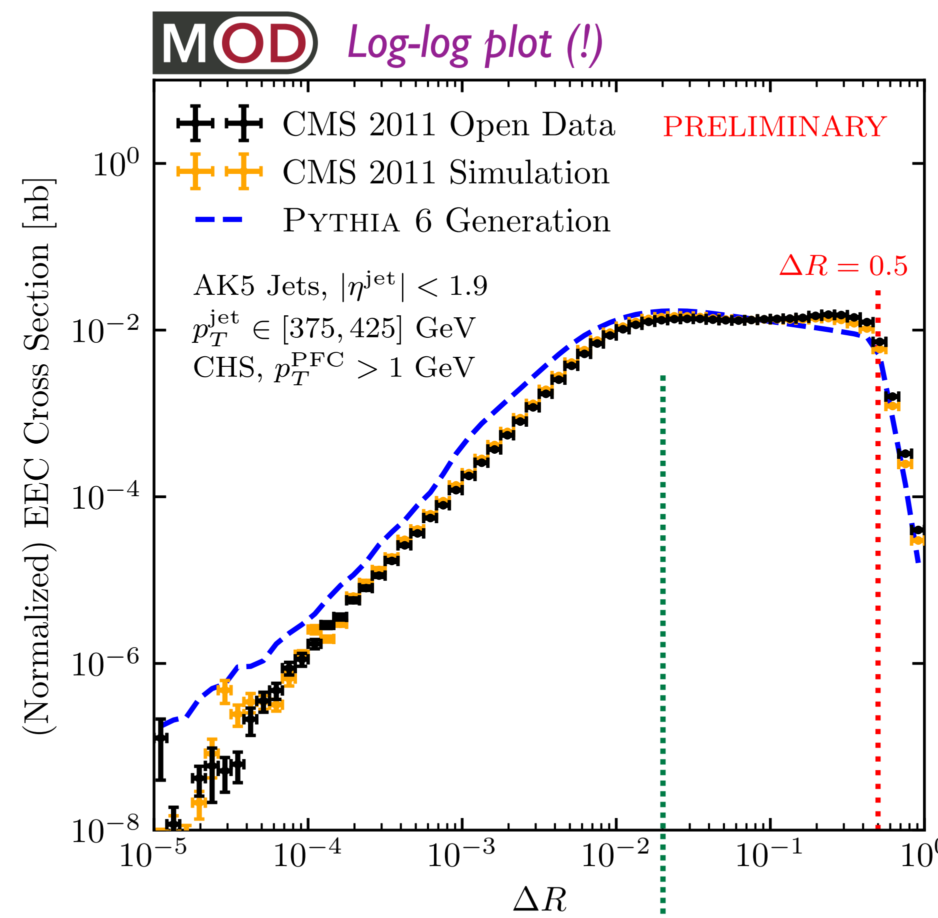
Theoretical challenges with small angle (collinear) limit

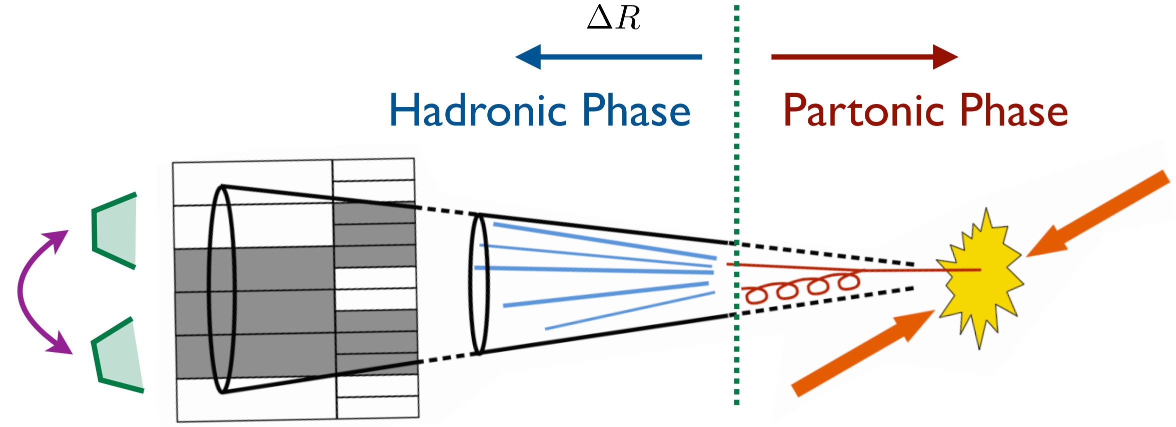
[Basham, Brown, Ellis, Love, PRL 1978; ALEPH, PLB 1991; see Chen, Moult, Zhang, Zhu, PRD 2020]

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]

QCD and Jets through the Lens of Machine Learning II

KITP, New Physics from Precision at High Energies, 30 March 2021

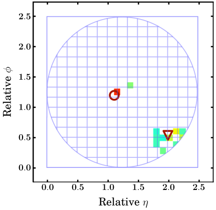
Frédéric Dreyer

UNIVERSITY OF OXFORD

with Jesse Thaler

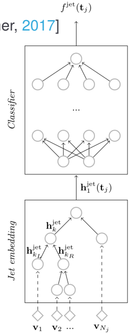
Representation of jets

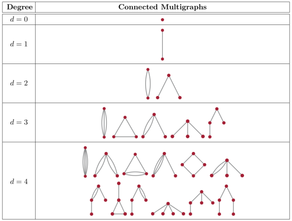
Jets are high-dimensional objects, can be represented in numerous ways



[Cogan, Kagan, Strauss, Schwartzman, 2014]

[Louppe, Cho, Becot, Cranmer, 2017]





[Komiske, Metodiev, Thaler, 2017]

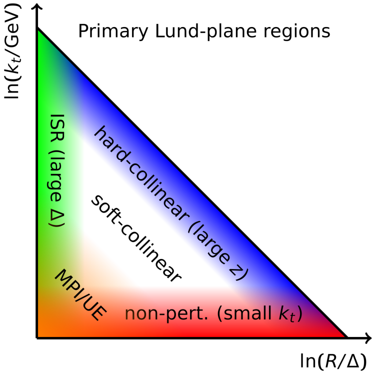
Choice of basis is a balance between

- Theory-motivated encoding of radiation patterns
- Constraints on neural network architecture

Lund diagrams

- Lund diagrams in the $(\ln z\theta, \ln\theta)$ plane are a very useful way of representing emissions.
- Different kinematic regimes are clearly separated, used to illustrate branching phase space in parton shower Monte Carlo simulations and in perturbative QCD resummations.
- Soft-collinear emissions are emitted uniformly in the Lund plane

$$dw^2 \propto \alpha_s \frac{dz}{z}\frac{d\theta}{\theta}$$

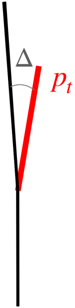


[Andersson et al, Z.Phys. C43 (1989) 625]
[FD, Salam, Soyez, JHEP 1812 (2018) 064]

Lund diagrams

- Lund diagrams in the $(\ln z\theta, \ln\theta)$ plane are a very useful way of representing emissions.
- Different kinematic regimes are clearly separated, used to illustrate branching phase space in parton shower Monte Carlo simulations and in perturbative QCD resummations.
- Soft-collinear emissions are emitted uniformly in the Lund plane

$$dw^2 \propto \alpha_s \frac{dz}{z}\frac{d\theta}{\theta}$$



Δ opening angle of a splitting

$$k_t = p_t \Delta$$

p_t (or $p_\perp$) is transverse momentum wrt beam

k_t is $\sim$ transverse momentum wrt jet axis

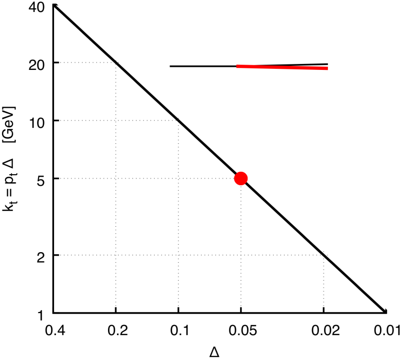
[Andersson et al, Z.Phys. C43 (1989) 625]

[FD, Salam, Soyez, JHEP 1812 (2018) 064]

Lund diagrams

- Lund diagrams in the $(\ln z\theta, \ln\theta)$ plane are a very useful way of representing emissions.
- Different kinematic regimes are clearly separated, used to illustrate branching phase space in parton shower Monte Carlo simulations and in perturbative QCD resummations.
- Soft-collinear emissions are emitted uniformly in the Lund plane

$$dw^2 \propto \alpha_s \frac{dz}{z}\frac{d\theta}{\theta}$$



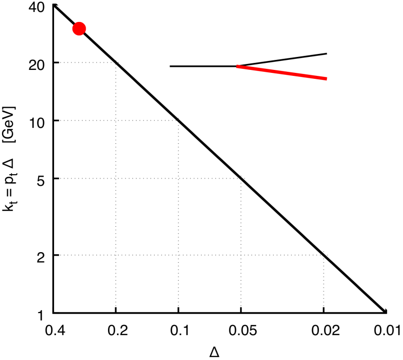
[Andersson et al, Z.Phys. C43 (1989) 625]

[FD, Salam, Soyez, JHEP 1812 (2018) 064]

Lund diagrams

- Lund diagrams in the $(\ln z\theta, \ln\theta)$ plane are a very useful way of representing emissions.
- Different kinematic regimes are clearly separated, used to illustrate branching phase space in parton shower Monte Carlo simulations and in perturbative QCD resummations.
- Soft-collinear emissions are emitted uniformly in the Lund plane

$$dw^2 \propto \alpha_s \frac{dz}{z}\frac{d\theta}{\theta}$$



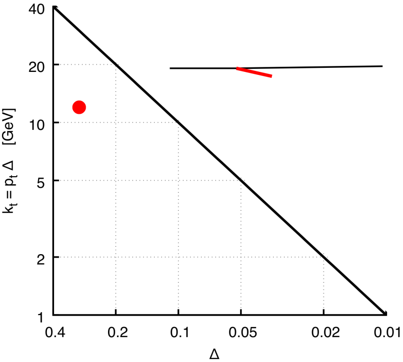
[Andersson et al, Z.Phys. C43 (1989) 625]

[FD, Salam, Soyez, JHEP 1812 (2018) 064]

Lund diagrams

- Lund diagrams in the $(\ln z\theta, \ln\theta)$ plane are a very useful way of representing emissions.
- Different kinematic regimes are clearly separated, used to illustrate branching phase space in parton shower Monte Carlo simulations and in perturbative QCD resummations.
- Soft-collinear emissions are emitted uniformly in the Lund plane

$$dw^2 \propto \alpha_s \frac{dz}{z}\frac{d\theta}{\theta}$$



[Andersson et al, Z.Phys. C43 (1989) 625]

[FD, Salam, Soyez, JHEP 1812 (2018) 064]

Lund plane representation

To create a Lund plane representation of a jet, use the (Cambridge/Aachen) clustering sequence of the jet to associate a unique Lund tree to each jet.

1. Undo the last clustering step, defining two subjets j_1, j_2 ordered in transverse momentum.
2. Save the kinematics of the current declustering step i as a tuple $\mathcal{T}^{(i)} = \{k_t, \Delta, z, m, \psi\}$

$$\Delta \equiv (y_1 - y_2)^2 + (\phi_1 - \phi_2)^2, \quad k_t \equiv p_{t2}\Delta,$$
$$m^2 \equiv (p_1 + p_2)^2, \quad z \equiv \frac{p_{t2}}{p_{t1}+p_{t2}}, \quad \psi \equiv \tan^{-1}\frac{y_2-y_1}{\phi_2-\phi_1}.$$

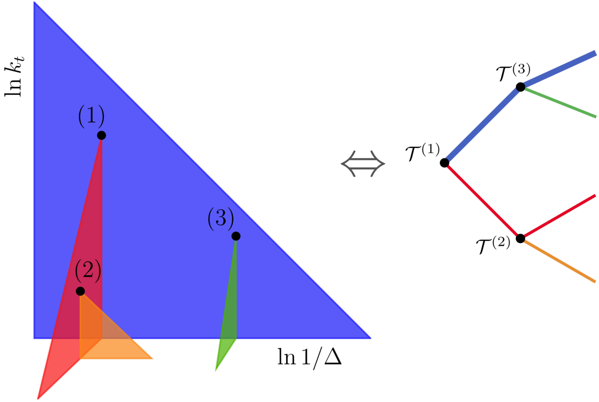
3. Repeat this procedure on both j_1 and j_2 until they are single particles.

Cambridge/Aachen clustering: pairwise recombination of particles with smallest Δ separation.

[FD, Salam, Soyez, JHEP 1812 (2018) 064]

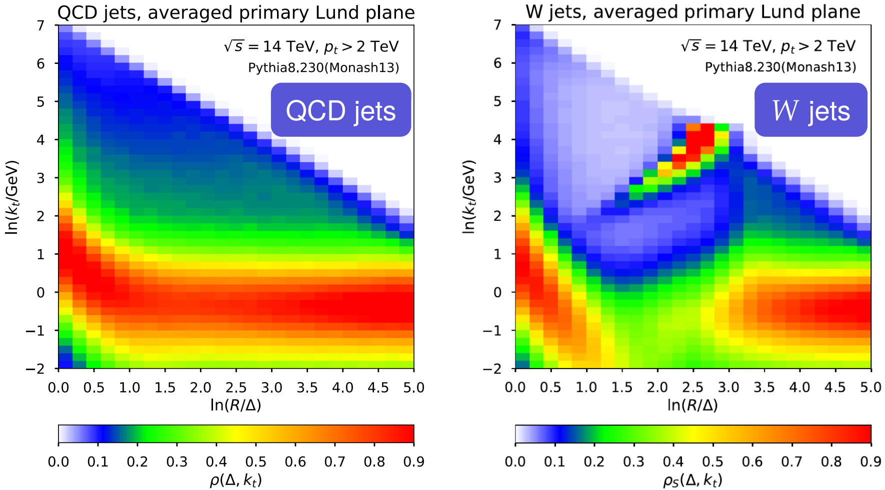
Lund plane representation

- Each jet is thus mapped onto a tree of Lund declusterings from its clustering sequence.
- Primary sequence of hardest transverse momentum branch is of particular interest for measurements and visualisation.



Jets as Lund images

Average over declusterings of hardest branch for 2 TeV jets.

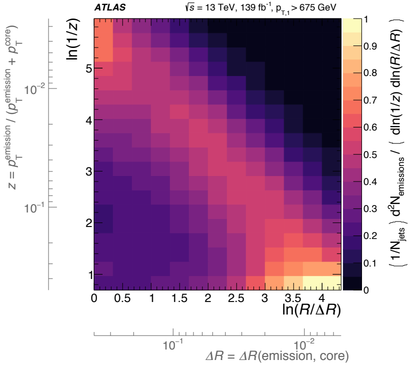


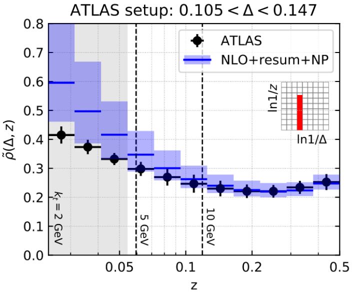
- Hard splittings visible along the diagonal line with jet mass $m = m_W$.

Lund images provide an opportunity for experimental measurements and comparisons with theory

- Lund plane can be predicted analytically, and the calculation is systematically improvable.
- Can be compared to data and used e.g. for α_s extractions.

[ATLAS Collaboration, PRL 124 (2020) 22, 222002]





[Lifson, Salam, Soyez, JHEP 10 (2020) 170]

Log-likelihood use of Lund Plane

Log-likelihood approach takes two inputs:

- First one obtained from the “leading” emission, defined as first emision satisfying $z > 0.025$ ($\sim$ mMDT tagger).

$$\mathcal{L}_\ell(m,z) = \ln\left(\frac{1}{N_S}\frac{dN_S}{dm\,dz}\bigg/\frac{1}{N_B}\frac{dN_B}{dm\,dz}\right)$$

- The second one which brings sensitivity to non-leading emissions.

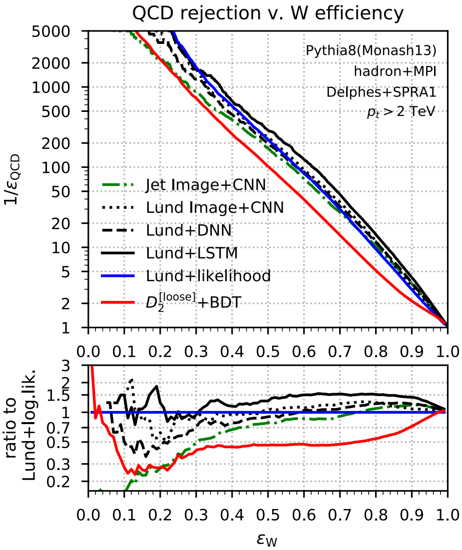
$$\mathcal{L}_{n\ell}(\Delta, k_t; \Delta^{(\ell)}) = \ln\left(\rho_S^{(n\ell)}/\rho_B^{(n\ell)}\right)$$

Overall log-likelihood signal-background discriminator for a given jet is then given by

$$\mathcal{L}_{\text{tot}} = \mathcal{L}_\ell(m^{(\ell)}, z^{(\ell)}) + \sum_{i\neq\ell}\mathcal{L}_{n\ell}(\Delta^{(i)}, k_t^{(i)}; \Delta^{(\ell)}) + \mathcal{N}(\Delta^{(\ell)})$$

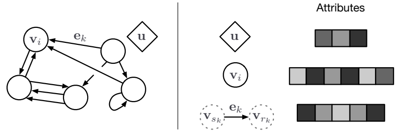
Boosted W tagging with the primary Lund plane

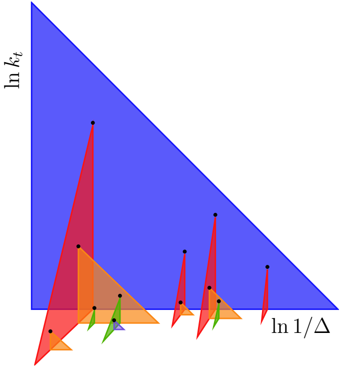
- LL approach already provides substantial improvement over best-performing substructure observable.
- LSTM network substantially improves on results obtained with other methods.
- Large gain in performance, particularly at higher efficiencies.



Mapping the full Lund plane to a graph

- Performance can be improved further by taking secondary/tertiary Lund planes into account, particularly relevant for top tagging.
- Treat each declustering of the Lund tree as a node on a graph.





Many promising applications of graphs, e.g.

[Henrion et al. DLPS NIPS '17]

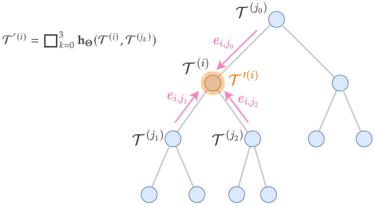
[Martinez et al. EPJP 134 (2019) 7, 333]

[Moreno et al. EPJC 80, 58 (2020)]

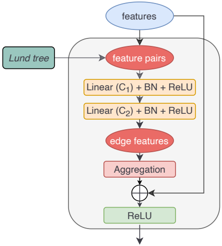
[Qu, Gouskos, PRD 101, 056019 (2020)]

LundNet models

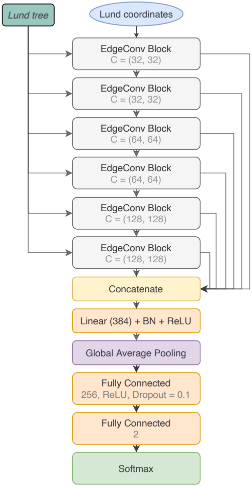
Tuple of kinematic variables as input for each node $\begin{cases} \text{LundNet-5} : (\ln k_t, \ln\Delta, \ln z, \ln m, \psi) \\ \text{LundNet-3} : (\ln k_t, \ln\Delta, \ln z) \end{cases}$



(a)



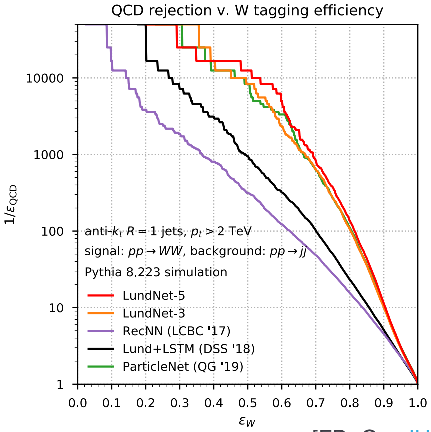
(b)

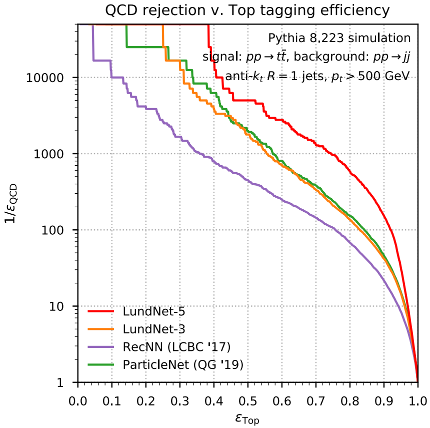


(c)

Boosted object tagging with graph networks

- Graph-based methods outperform our previous benchmarks significantly.
- LundNet model provides substantial improvement over ParticleNet and is an order of magnitude faster to train/deploy.

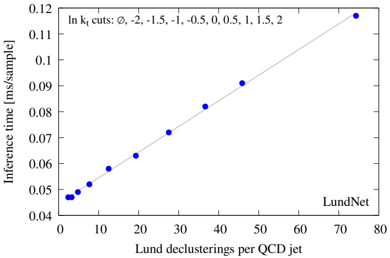




[FD, Qu, JHEP 03 (2021) 052]

Complexity of models

- Direct use of the Lund tree as the graph structure removes the need for a costly nearest-neighbour search.
- LundNet reduces training and inference time by order of magnitude compared to previous graph methods.
- Due to their higher-level kinematic inputs, LundNet takes significantly less epochs to converge to a good solution.
- Training and inference time of the model are reduced as transverse momentum cut is increased and more nodes are removed from input.



	Number of parameters	Training time [ms/sample/epoch]	Inference time [ms/sample]
LundNet	395k	0.472	0.117
ParticleNet	369k	3.488	1.036
Lund+LSTM	67k	0.424	0.131

Understanding what the network is learning

Can we determine what is driving performance of a neural network?

- Consider their application on a simple task where we have first principle understanding.
- Build analytic likelihood-ratio discriminatant for this configuration and compare them with ML models.

We will consider quark/gluon discrimination.

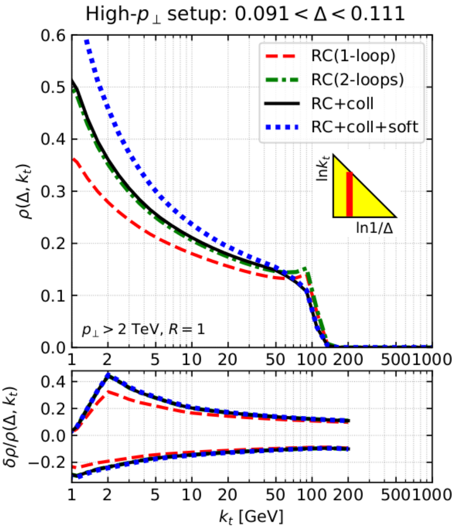
Calculating Lund plane variables

Primary Lund-plane density can be computed to single-logarithmic accuracy for both quarks and gluons.

[Lifson, Salam, Soyez, JHEP 10 (2020) 170]

For given jet with Lund declusterings $\{\Delta_i, k_{t,i}, \ldots\}$ define likelihood ratio

$$\mathbb{L}_{\text{density}} = \prod_i \frac{\rho_g(\Delta_i, k_{t,i})}{\rho_q(\Delta_i, k_{t,i})}$$



Building an analytic q/g discriminant

For a jet with primary declusterings $\{\Delta_i, k_{t,i}, z_i, \dots\}$ compute the likelihood ratio

$$\mathbb{L}_{\text{primary}} = \frac{p_g(\{\Delta_i, k_{t,i}, z_i, \dots\})}{p_q(\{\Delta_i, k_{t,i}, z_i, \dots\})}$$

where $p_{q,g}(\{\Delta_i, k_{t,i}, z_i, \dots)\}$ is the probability to observe the given set of declusterings if the jet were a quark or a gluon.

$$p_q(\{\Delta_i, k_{t,i}, z_i, \dots\}) = p^{\text{(final)}}(q|q_0) + p^{\text{(final)}}(g|q_0)$$
$$p_g(\{\Delta_i, k_{t,i}, z_i, \dots\}) = p^{\text{(final)}}(q|g_0) + p^{\text{(final)}}(g|g_0)$$

We can compute all single-logarithms from running coupling and collinear effects.

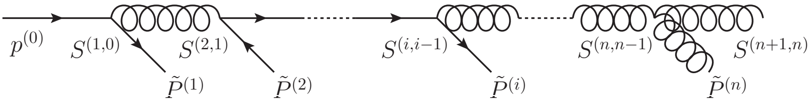
Optimal discriminant at single-logarithmic accuracy

- Computation in the collinear limit where Lund declusterings are strongly ordered in angle $\Delta_1 \gg \Delta_2 \gg \cdots \gg \Delta_n$.
- Construct the quark & gluon probability distribution iteratively from first splitting.

Probabilities after including all Lund declusterings expressed as

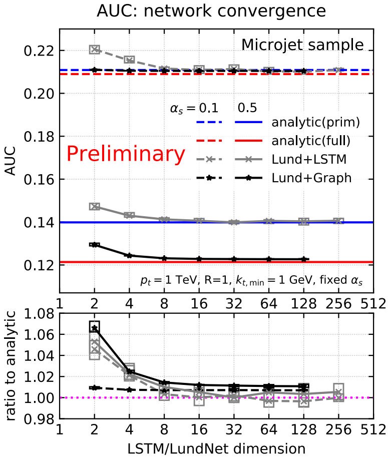
$$p^{(\text{final})} = S^{n+1,n}\tilde{P}^{(n)}S^{n,n-1}\ldots\tilde{P}^{(i)}S^{i,i-1}\ldots\tilde{P}^{(1)}S^{1,0}p^{(0)}$$

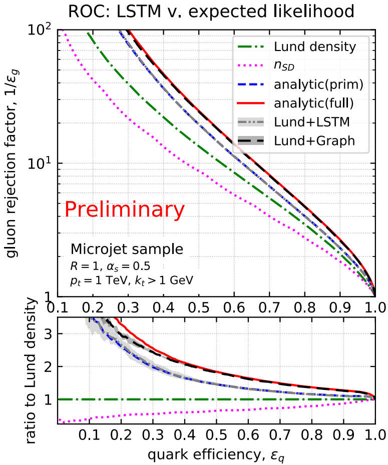
where S is a NLL Sudakov matrix and P a matrix of splitting kernels.



Comparison with pure-collinear parton shower

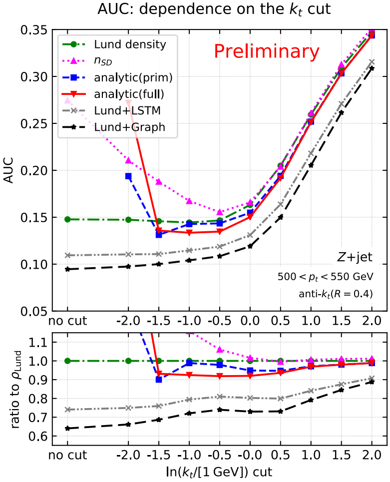
- Compare analytic and deep learning approaches in events generated in the strong-angular-ordered limit.
- In this limit analytic approach is exact and becomes optimal discriminant.

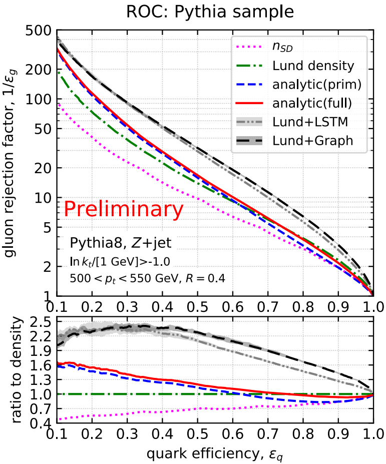




Application to full Monte Carlo

- Applying to Z+jet events generated with Pythia 8: difference in performance, but same qualitative behaviour.

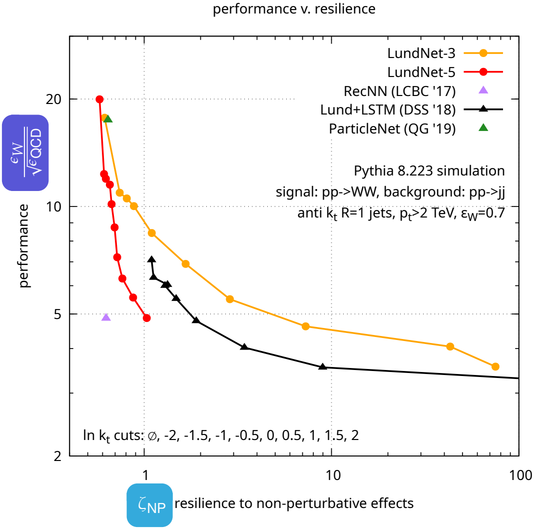




[FD, Soyez, Takacs, in progress]

Robustness to model-dependent effects

- Performance compared to resilience to MPI and hadronisation corrections.
- Vary Lund plane cut on k_t, which reduces sensitivity to the non-pert. region.



$\Delta\epsilon = \epsilon - \epsilon'$

$$\zeta_{\mathrm{NP}} = \left(\frac{\Delta\epsilon_S^2}{\langle\epsilon\rangle_S^2} + \frac{\Delta\epsilon_B^2}{\langle\epsilon\rangle_B^2} \right)^{-\frac{1}{2}}$$

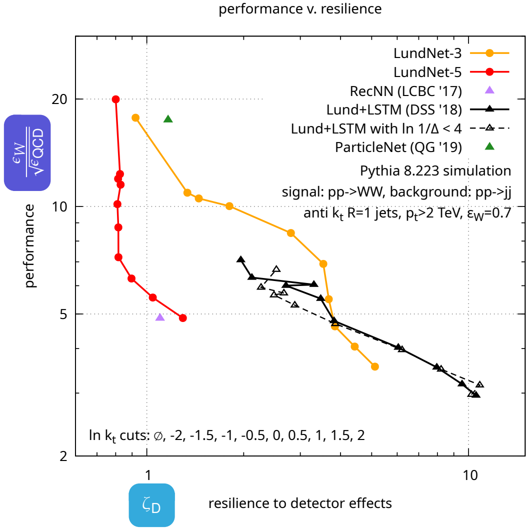
(c.f. arXiv:1803.07977)

$\langle\epsilon\rangle = \frac{1}{2}(\epsilon + \epsilon')$

- LundNet-3 performs well even at high resilience.
- Most ML models can reach very good performance but are not particularly resilient to non-perturbative effects.

Robustness to model-dependent effects

- Performance compared to resilience to detector smearing effects.
- Vary Lund plane cut on k_t, which partly reduces sensitivity to detector effects.



$\Delta\epsilon = \epsilon - \epsilon'$

$$\zeta_D = \left(\frac{\Delta\epsilon_S^2}{\langle\epsilon\rangle_S^2} + \frac{\Delta\epsilon_B^2}{\langle\epsilon\rangle_B^2} \right)^{-\frac{1}{2}}$$

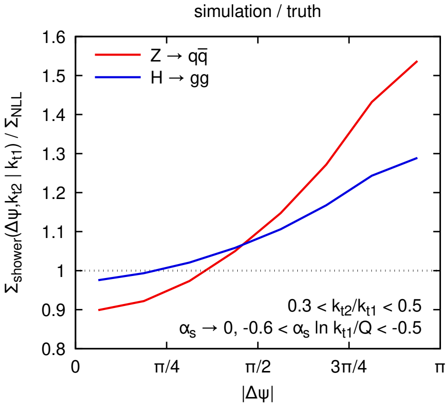
(c.f. arXiv:1803.07977)

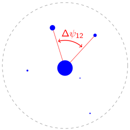
$\langle\epsilon\rangle = \frac{1}{2}(\epsilon + \epsilon')$

- LundNet-3 performs well even at high resilience.
- Most ML models can reach very good performance but are not particularly resilient to detector effects.

But what does the machine learn?

- Important limitation stems from the fact that labelled training data is usually obtained from Monte Carlo event generators.
- But parton shower simulations are not perfect tools!



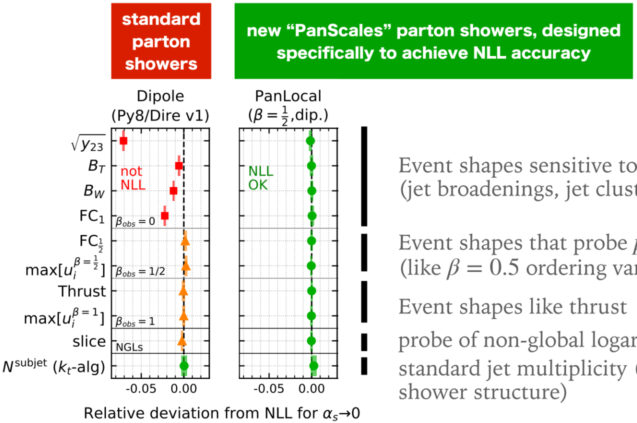


Common dipole showers display quark/gluon differences that should not be there.

- How to be sure ML models are not overfitting unphysical features?

[Dasgupta, FD, Hamilton, Monni, Salam, Soyez, Phys.Rev.Lett. 125 (2020) 5, 052002]

Designing new showers for precision physics



Event shapes sensitive to transverse momentum (jet broadenings, jet clustering transitions)

Event shapes that probe $p_t e^{-0.5|\eta|}$ (like $\beta = 0.5$ ordering variable)

Event shapes like thrust

probe of non-global logarithms

standard jet multiplicity (probe of full recursive shower structure)

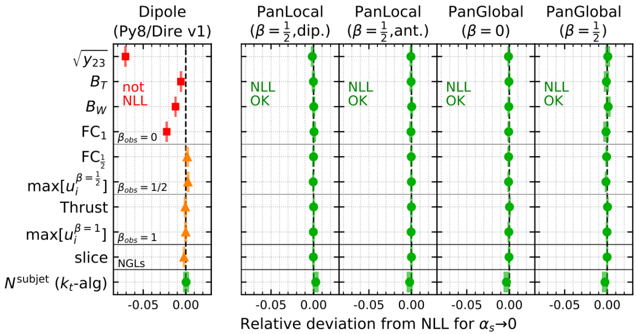
[Dasgupta, FD, Hamilton, Monni, Salam, Soyez, Phys.Rev.Lett. 125 (2020) 5, 052002]

Paves the way for improved simulations with more accurate physical description of perturbative radiation.

Designing new showers for precision physics

standard parton showers

new "PanScales" parton showers, designed specifically to achieve NLL accuracy



All PanScales shower that are expected to agree with NLL pass these tests

(Standard dipole showers don't)

[Dasgupta, FD, Hamilton, Monni, Salam, Soyez, Phys.Rev.Lett. 125 (2020) 5, 052002]

Paves the way for improved simulations with more accurate physical description of perturbative radiation.

Conclusions

- Jet substructure provides a unique practical playground for recent developments in machine learning.
- Discussed new ways to study and exploit radiation patterns in a jet using the Lund plane.
- Through appropriate benchmarks, it is possible to gain some first-principles understanding of neural network performance.
- Combination of physical insight and machine learning needed to design models that combine performance and robustness.