QCD Theory meets Information Theory

Jesse Thaler

Bay Area Particle Theory Seminar — April 4, 2025

Returning to My Roots!

Event Generation with GenEvA, version of parton shower/matrix element merging/matching

Background Estimation

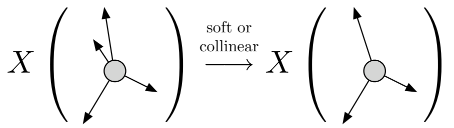
Essential for Understanding BSM signals at LHC

Partonic Strategy

Inclusive Measurement Function

$$\sigma_X = \sum_n \int d\Phi_n \, |\mathcal{M}_n|^2 \, X(\Phi_n)$$

Well-Defined if Infrared Safe

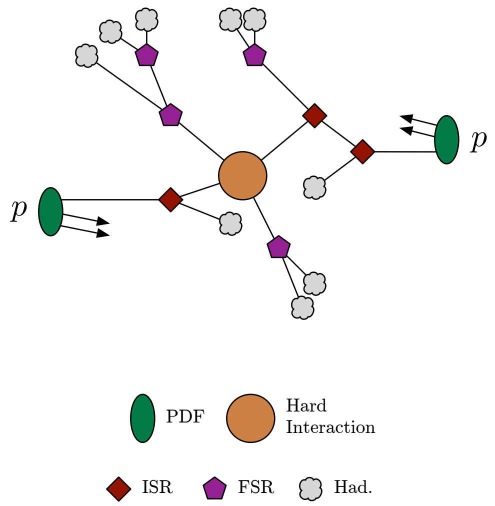


Infrared Divergences Cancel

$$|\mathcal{M}^{\text{loop}}|^2 \, X(\Phi) + \int_{+1} |\mathcal{M}^{\text{tree}}|^2 \, X(\Phi_{+1}) = \text{finite}$$

Hadronic Strategy

Exclusive Event Generation



"Ideal" Strategy

Exclusive event generator that reproduces the results of any inclusive measurement function to the calculated accuracy.

N^{i}LO : O(α_s^i) beyond Born Level
N^{j}LL : Resummation of $\alpha_s^n \log^{2n-j} r$ terms
(r is ratio of kinematic scales)

Fully hadronized events (exclusive) using best perturbative calculations (inclusive).

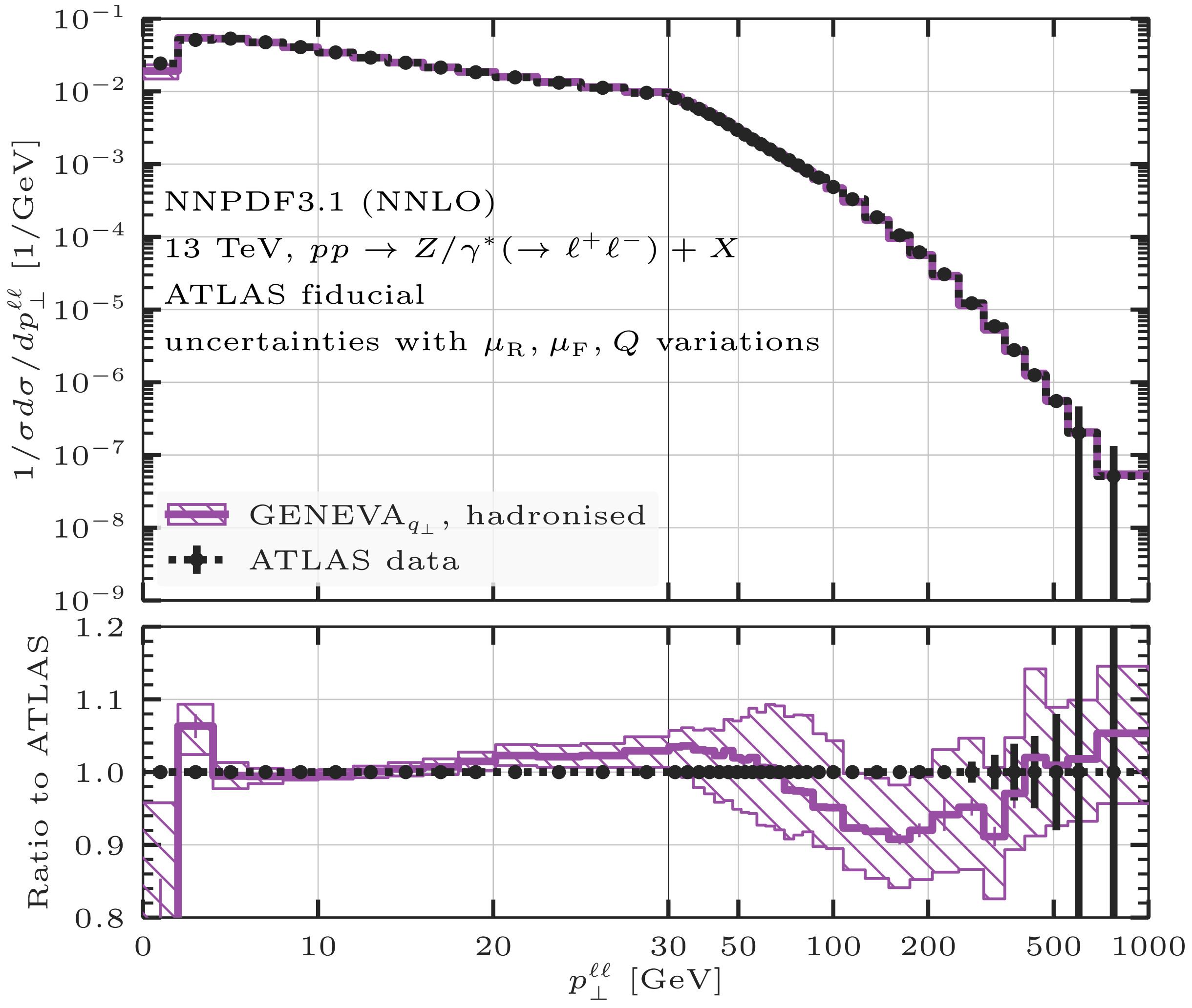
MC@NLO is a concrete example in this direction for NLO/LL calculations (see also POWHEG)

I was "dissuaded" from pursuing this further by a DOE Reviewer in 2011…

[slides from my 2008 talk; Bauer, Tackmann, JDT, JHEP 2008a, JHEP 2008b]

Returning to My Roots!

Event Generation with GenEvA, version of parton shower/matrix element merging/matching



Now NNLO+PS matching with GENEVA (and NNLOPS, MiNNLO$_{PS}$, TOMTE, …)!

Can these be merged with showers that go to NLL accuracy or beyond leading color (e.g. PanScales, Deductor)?

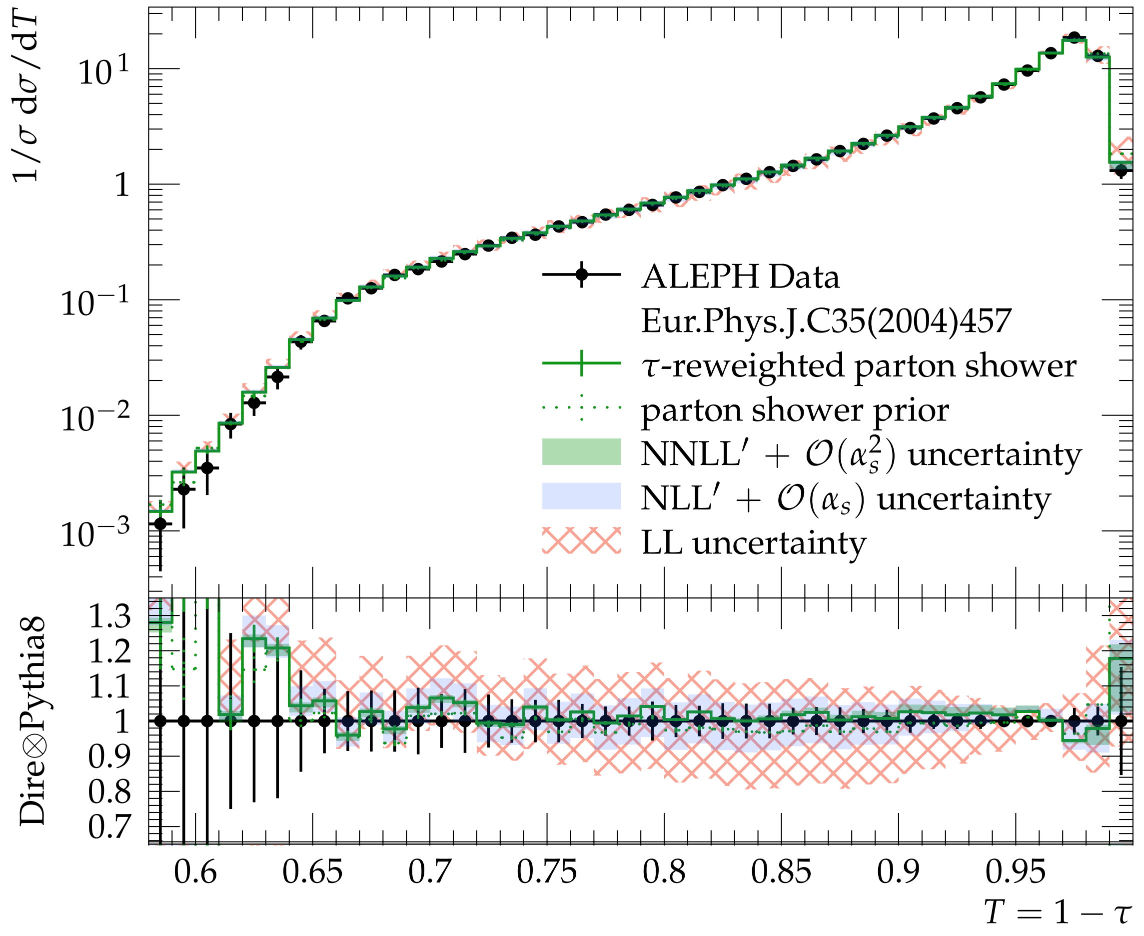
Non-perturbative corrections?
"Proper" theory systematics?

[from Campbell, Diefenthaler, Hobbs, Höche, Isaacson, Kling, Mrenna, Reuter, et al., Snowmass 2022]

[slides from my 2008 talk; Bauer, Tackmann, JDT, JHEP 2008a, JHEP 2008b]

Today: Theory Synthesis from "Thinking like a Machine"

Proof of concept: thrust at LEP



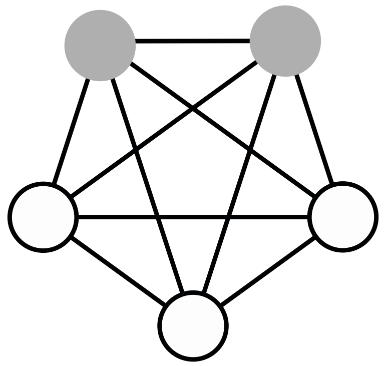
Instead of treating parton shower/matrix element merging/matching as a domain-specific problem in QCD, we can treat it as a generic problem in information theory

Remarkably, this will reveal a new understanding about the difference between fixed-order and resummed calculations, as captured by ordinary versus logarithmic moments

Possibility that this approach could dramatically simplify the experiment/theory interface for precision predictions

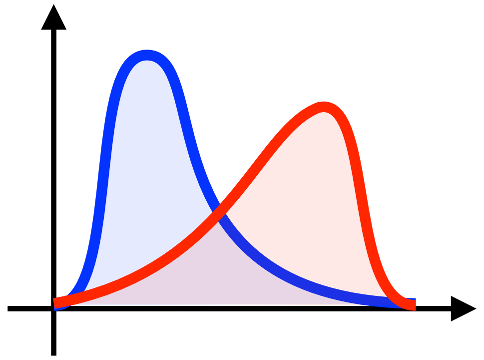
[Assi, Lee, Höche, JDT, arXiv 2025]

QCD Theory meets Information Theory



Colloquium Version of the Story

When physicists "think like a machine", we can translate thorny theoretical challenges into optimization problems

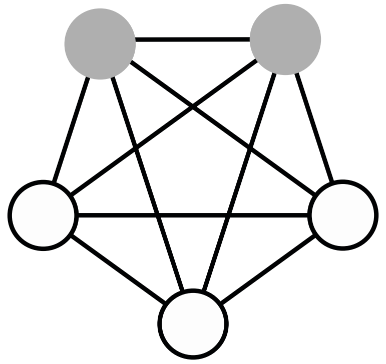


Basics of ML and Information Theory

We can leverage the flexibility of loss functions to learn and constrain physically meaningful moments of distributions

Logarithmic Moments and QCD

Logarithmic moments are particularly powerful probes of QCD that distinguish resummed from fixed-order calculations



Colloquium Version of the Story

When physicists "think like a machine", we can translate thorny theoretical challenges into optimization problems

Basics of ML and Information Theory

We can leverage the flexibility of loss functions to learn and constrain physically meaningful moments of distributions

Logarithmic Moments and QCD

Logarithmic moments are particularly powerful probes of QCD that distinguish resummed from fixed-order calculations

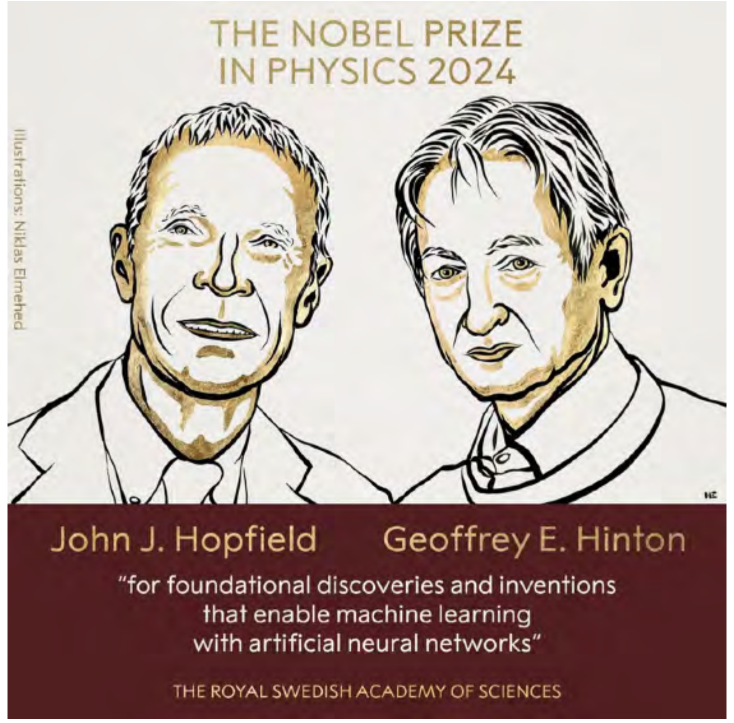
The NSF Institute for Artificial Intelligence and Fundamental Interactions (IAIFI /aɪ-faɪ/ iaifi.org)

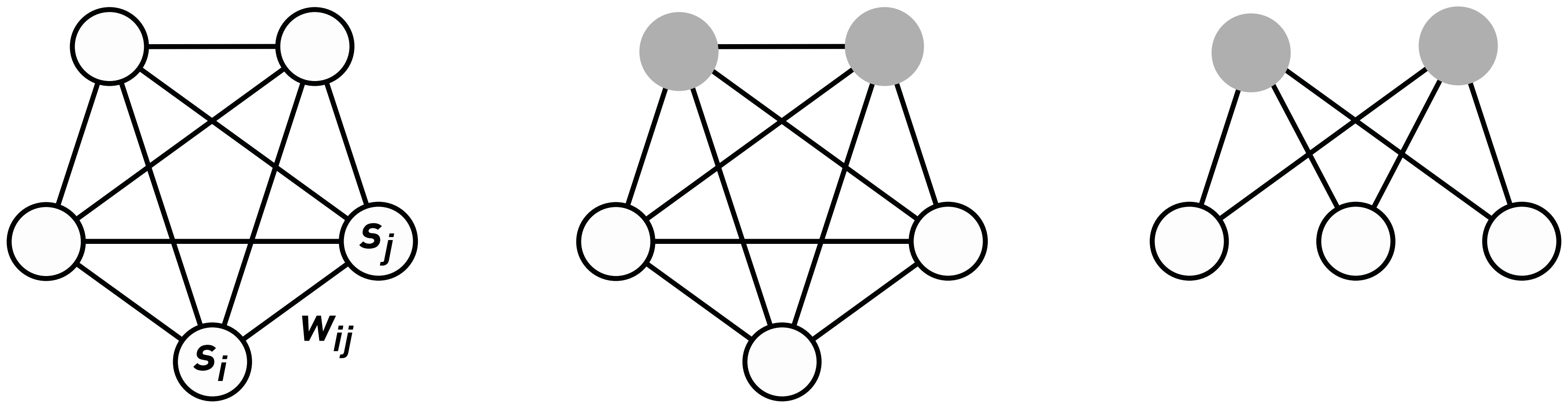
Launched August 2020

Deep Learning (AI) + Deep Thinking (Physics) = Deeper Understanding

2024: Recognition for "AI + Physics"!

Along with AI+Science more generally!

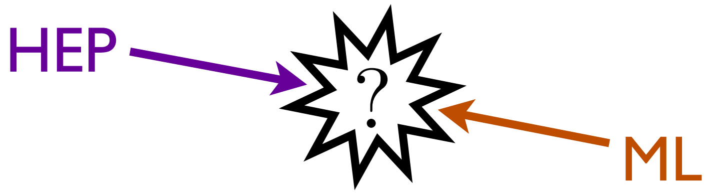




Learning through dynamical systems, power of statistical reasoning,
broad implications for scientific discovery

[discussed at IAIFI Fireside Chat]

Lessons from the HEP + ML Collision



To meet the standards of scientific rigor and performance,
we need to teach machines to "Think like a Physicist"

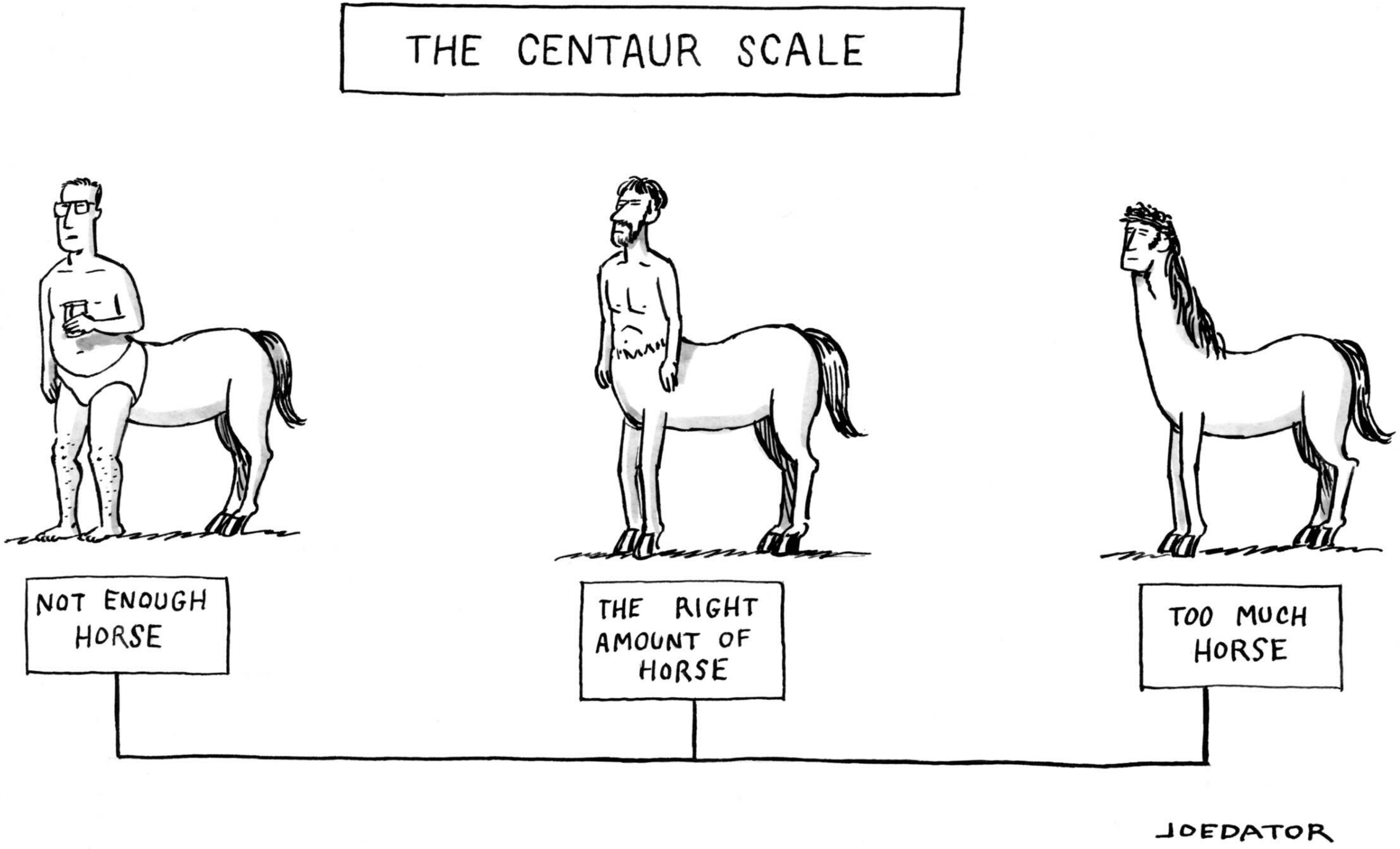
E.g.: symmetries, robustness to systematics, exactness guarantees, statistical inference, …

But to fully capitalize on AI technologies,
we also need to teach physicists to "Think like a Machine"

E.g.: computational complexity, reframing via optimization/search, algorithmic reasoning, …

The Power of "Centaur Science"

Progress in computation and information theory has long been intertwined with progress in the physical sciences (e.g. statistical mechanics, quantum computers)



[Joe Dator, The New Yorker 2025; h/t Kyle Cranmer]

Danger of "Thinking Like a Machine"?

An inspiring/infuriating paper!

***Position:* Is machine learning good or bad for the natural sciences?**

Abstract

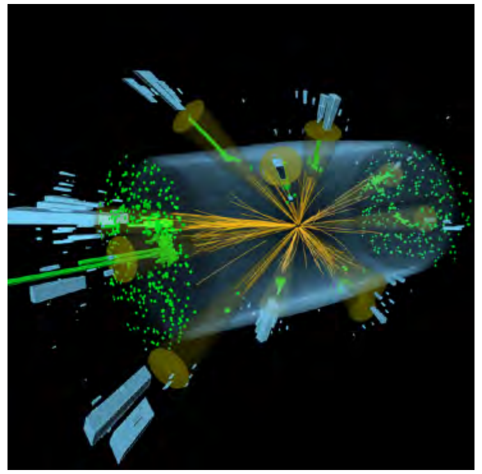
Machine learning (ML) methods are having a huge impact across all of the sciences. However, ML has a strong ontology—in which only the data exist—and a strong epistemology—in which a model is considered good if it performs well on held-out training data. These philosophies are in strong conflict with both standard practices and key philosophies in the natural sciences. . . .

My attempt at a cartoon summary:

	Machine Learning	Natural Sciences
What "exists"?	Data	Latent Structures
What is "success"?	Accurate Modeling	Understanding

[Hogg, Villar, PMLR 2024]

A Generative Model to Mimic the LHC?



What if you could generate synthetic LHC collisions that were indistinguishable from real LHC data?

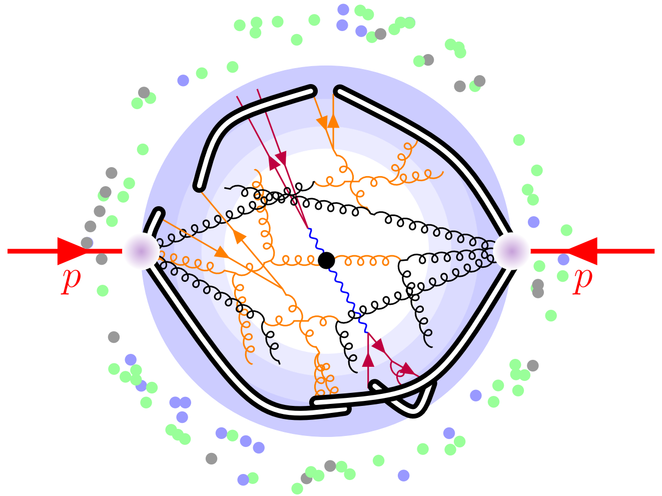
[extreme extrapolation of Butter, Diefenbacher, Kasieczka, Nachman, Plehn, Shih, Winterhalder, SciPost 2022]

According to Hogg/Villar, this conflicts with the epistemology of natural science…

What if you could generate synthetic LHC collisions based entirely on first-principles QFT calculations?

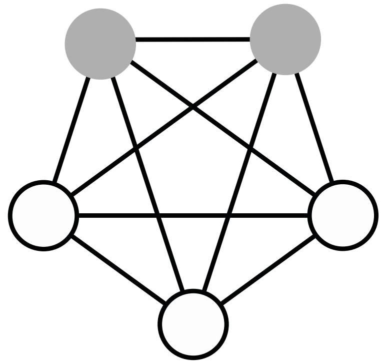
[aspirational goal in Campbell, Diefenthaler, Hobbs, Hoeche, Isaacson, Kling, Mrenna, Reuter, et al., Snowmass 2022]

Amazing if you had it! But we are nowhere near being able to do this…



Centaur Science: Combine models of LHC data for the full phase space with first-principles predictions whenever available

Taking Inspiration from Boltzmann (Machines)



Boltzmann Machine: Generative model that samples according to a statistical distribution

$P(\boldsymbol{s}) \propto e^{-E/T}$ $\qquad E = -\sum_{i<j} w_{ij} s_i s_j - \sum_i \theta_i s_i$

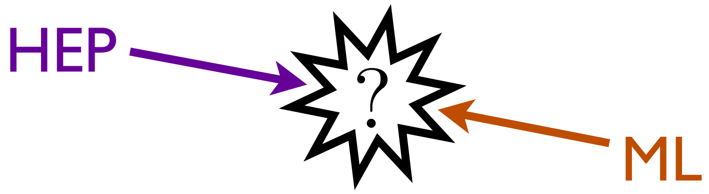
Boltzmann Factor: Maximum entropy solution subject to constraint on average energy

From spins to continuous phase space (and entropy to relative entropy)

$$p(x) = \underbrace{q(x)}_{\text{Prior}} \exp\Big[- \underbrace{\beta_0}_{\text{Sets Partition Function}} - \underbrace{\beta_1}_{\text{Lagrange Multipler}} \underbrace{f_1}_{\text{Constrained Quantity}}(x) - \beta_2 f_2(x) - \ldots \Big]$$

[Fahlman, Hinton, Sejnowski, AAAI 1983; Ackley, Hinton, Sejnowski, Cogn. Sci. 1985; Boltzmann 1877]

QCD Theory meets Information Theory



Plucking a random QCD calculation:

[e.g. Larkoski, Salam, JDT, JHEP 2013; based on Banfi, Salam, Zanderighi, JHEP 2005]

$$\Sigma(e^{-L}) = N \frac{e^{-\gamma_E R'}}{\Gamma(1+R')} e^{-R} \qquad R = \frac{\alpha_s}{\pi}\frac{C}{\beta}(L+B)^2 \qquad L \equiv \ln \frac{R_0^\beta}{C_1^{(\beta)}}$$

Sudakov Form Factor

To a QCD theorist, this is a (simplified) next-to-leading-logarithmic calculation

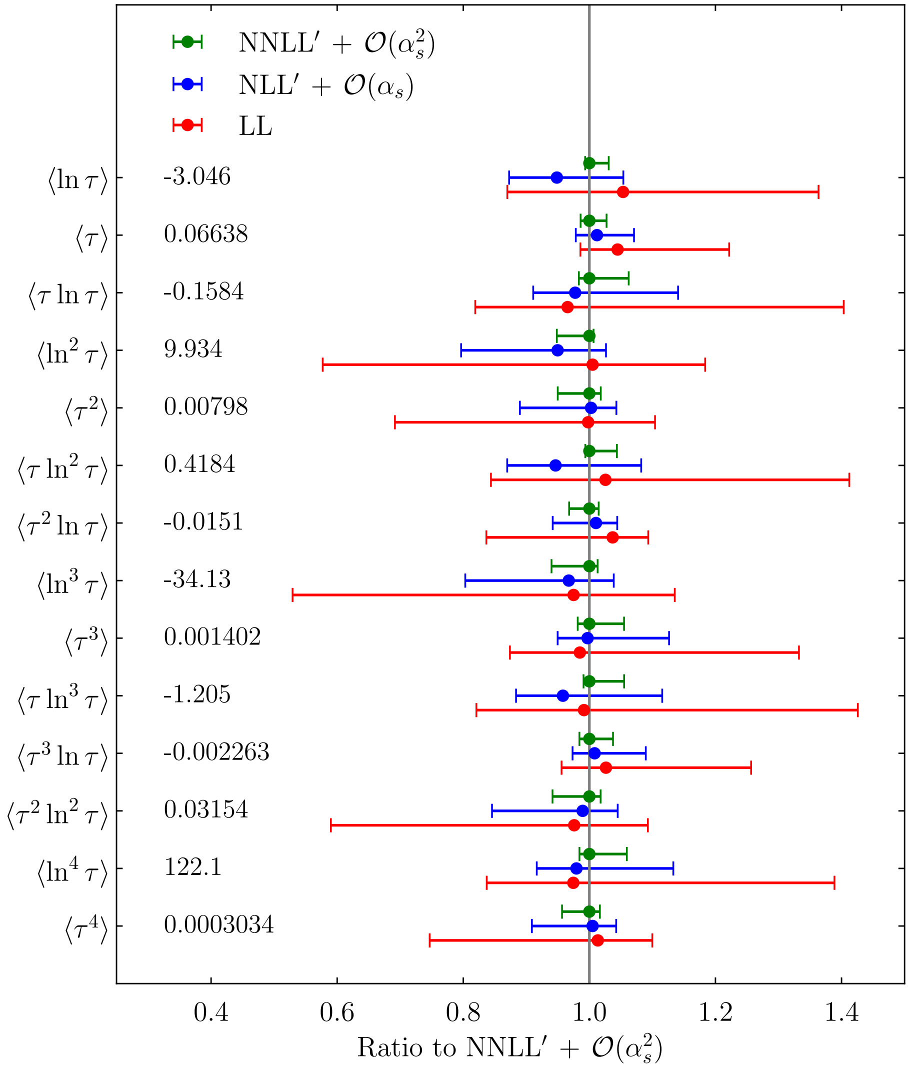
To an ML practitioner, this is an opportunity for neurosymbolic learning

To a Centaur Scientist, the Sudakov form factor is a kind of Boltzmann factor, where the cusp anomalous dimension is like a Lagrange multiplier that enforces a constraint on the second logarithmic moment of the distribution

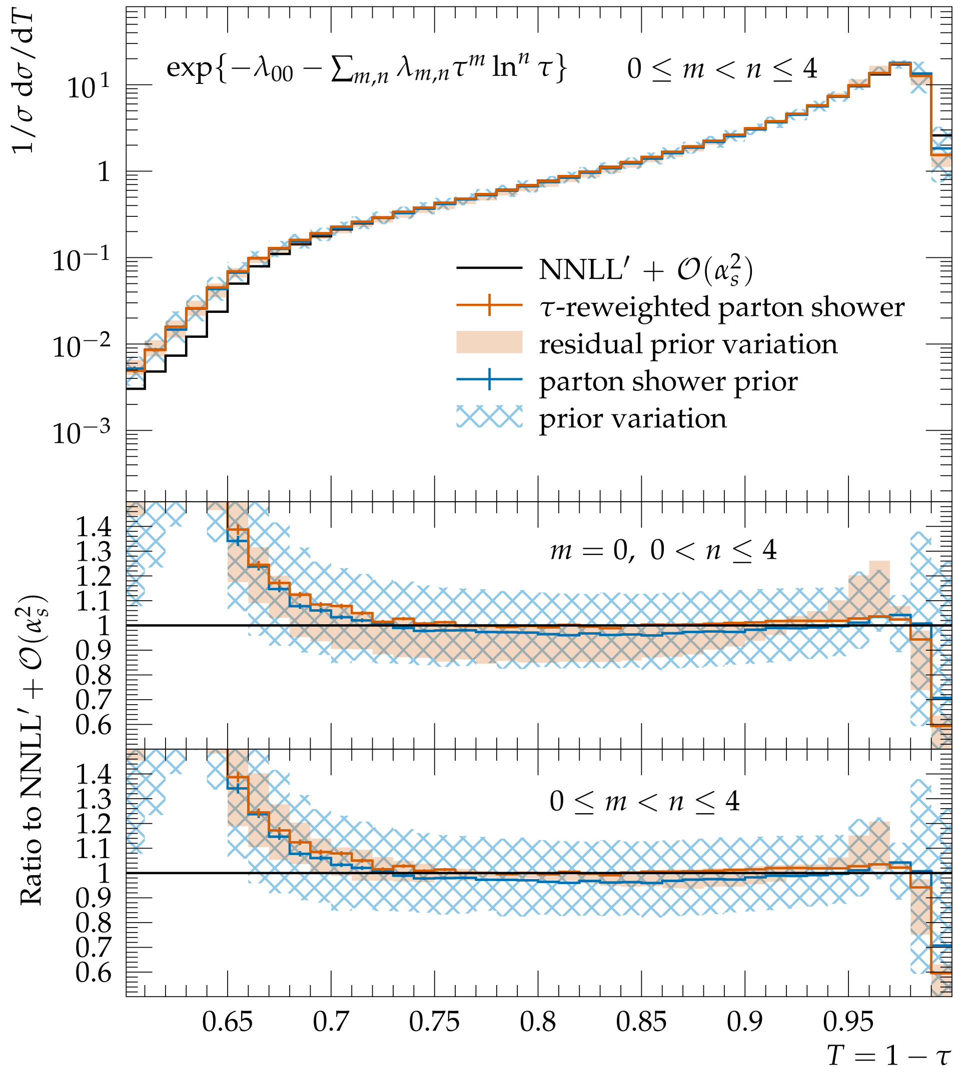
To my knowledge, logarithmic moments have never been measured or calculated before in QCD!

Proof-of-Concept Study with "Thrust"

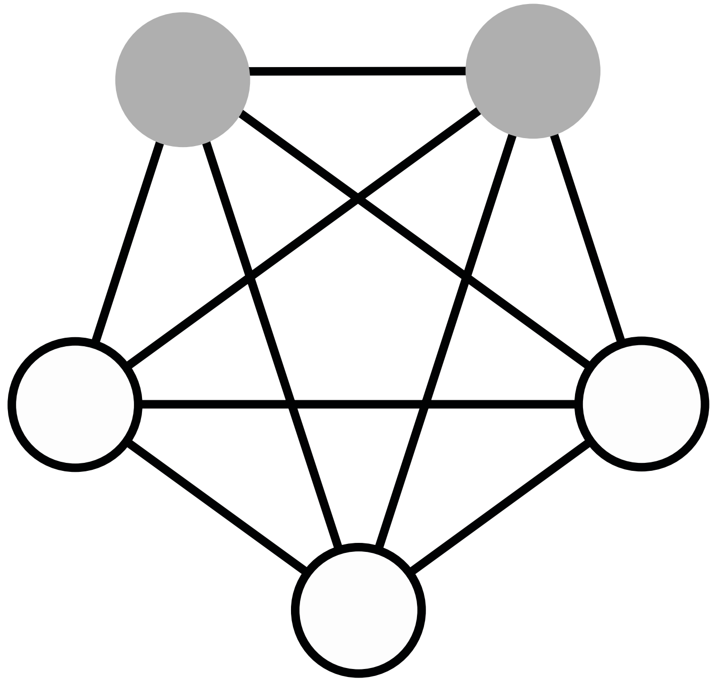
First QCD Calculation of Thrust Logarithmic Moments!



Priors from imperfect generators yield consistent results after moment reweighting!



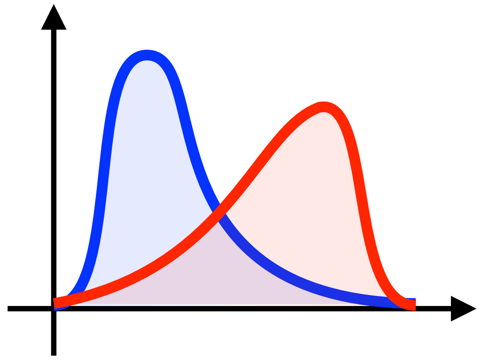
[Assi, Lee, Höche, JDT, arXiv 2025;
see Sudakov safety in Larkoski, Marzani, JDT, PRD 2015;
see related moment construction in Desai, Nachman, JDT, PRD 2024]



Centaur science continuing the tradition of relating properties of physical systems to concepts in computation/information theory

Colloquium Version of the Story

When physicists "think like a machine", we can translate thorny theoretical challenges into optimization problems

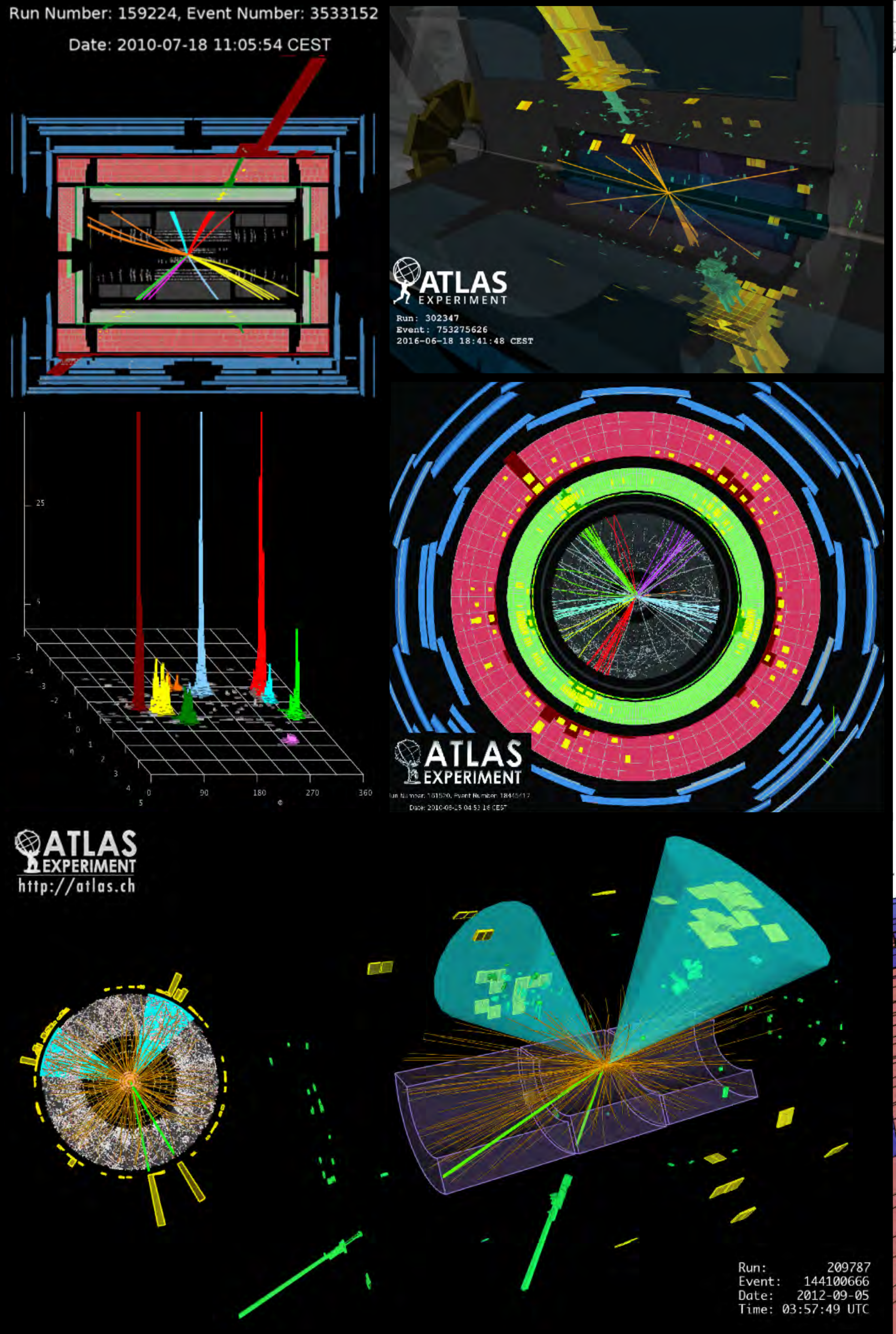


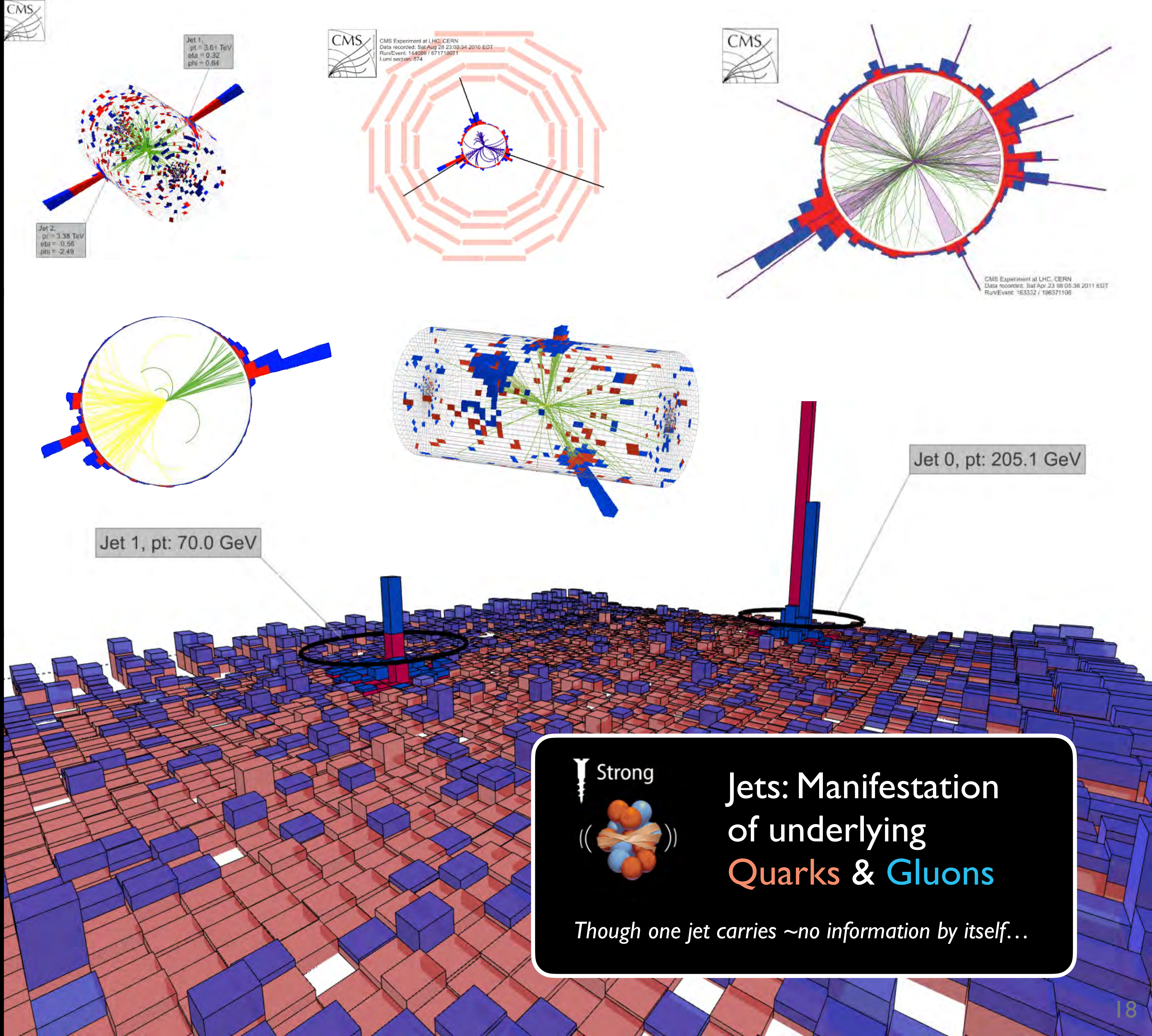
Basics of ML and Information Theory

We can leverage the flexibility of loss functions to learn and constrain physically meaningful moments of distributions

Logarithmic Moments and QCD

Logarithmic moments are particularly powerful probes of QCD that distinguish resummed from fixed-order calculations





Strong

Jets: Manifestation of underlying Quarks & Gluons

Though one jet carries ~no information by itself…

Machine Learning 101: Likelihood Ratio Trick

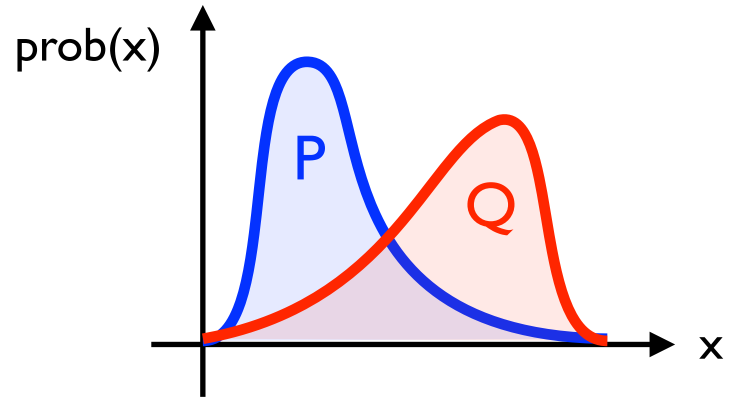
Many HEP problems can be expressed in this form!

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

[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, JDT, PRD 2021]

Machine Learning 101: Likelihood Ratio Trick

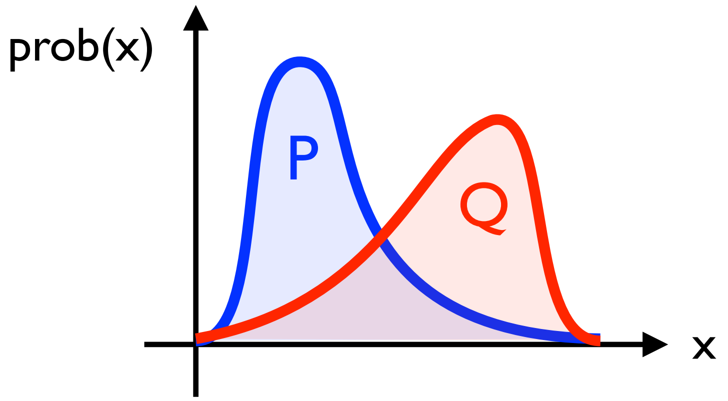
Many HEP problems can be expressed in this form!

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):

$$L = -\langle \log f(x) \rangle_P + \langle f(x) - 1 \rangle_Q$$

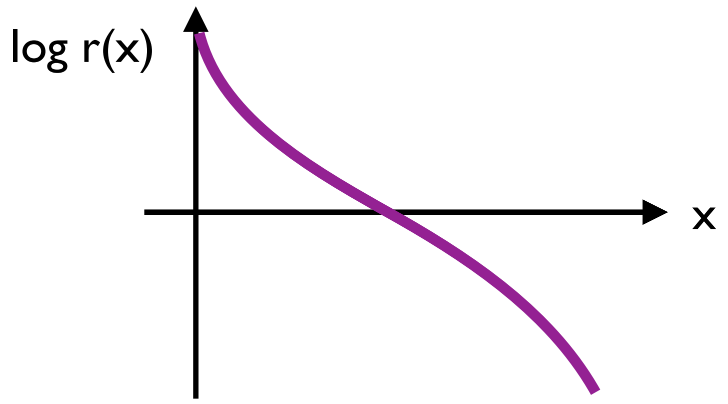
Asymptotically:

$$\arg\min_{f(x)} L = \frac{p(x)}{q(x)}$$

Likelihood ratio

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

Kullback–Leibler divergence



[see e.g. 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, JDT, PRD 2021]

Machine Learning 101: Likelihood Ratio Trick

Many HEP problems can be expressed in this form!

Key example of simulation-based inference

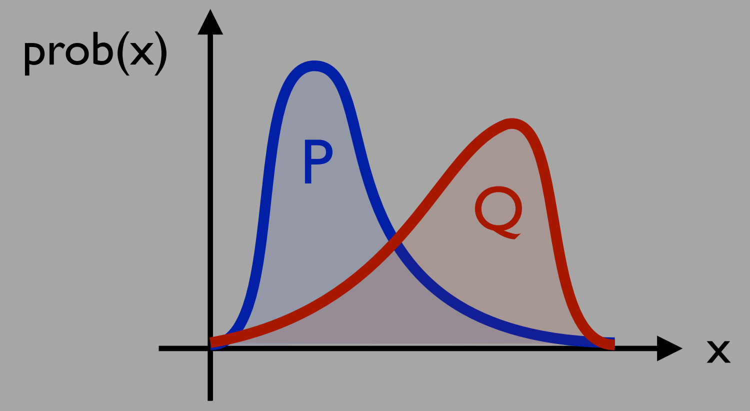
Asymptotically, same structure as Lagrangian mechanics!

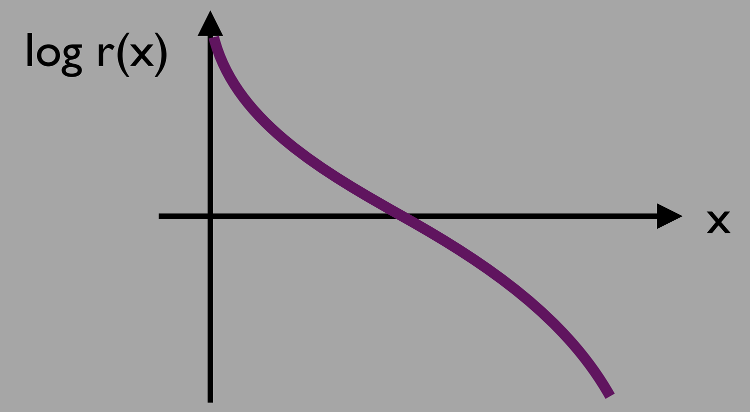
Action: $$L = \int \mathrm{d}x\, \mathcal{L}(x)$$

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 focus from solving problems to specifying problems

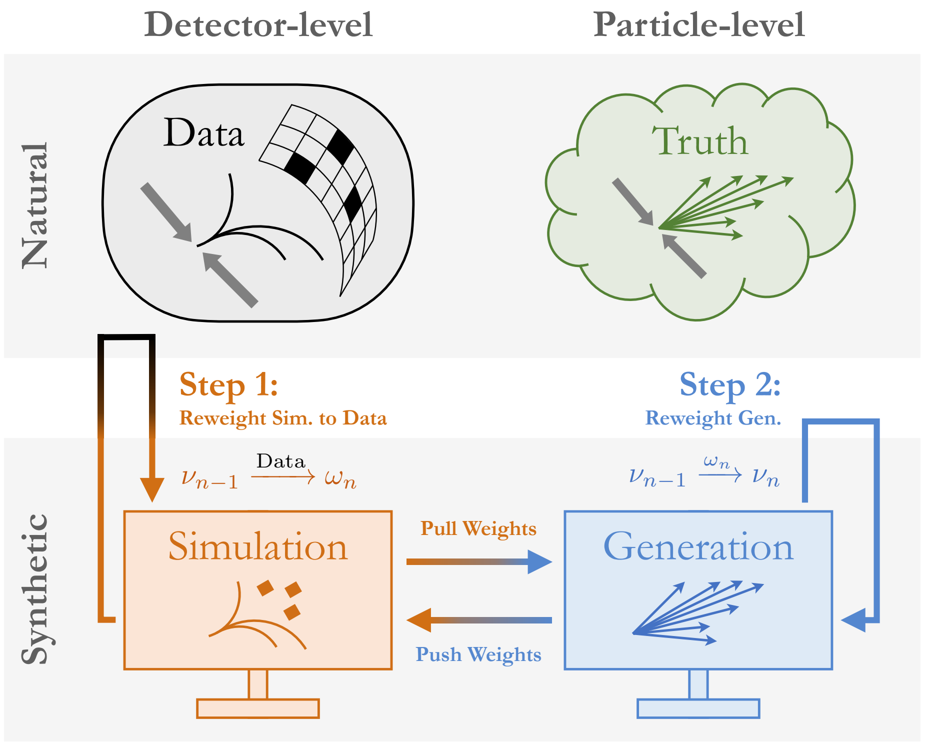




[see e.g. Cranmer, Pavez, Louppe, arXiv 2015; D'Agnolo, Wulzer, PRD 2019; simulation-based inference in Cranmer, Brehmer, Louppe, PNAS 2020; relation to f-divergences in Nguyen, Wainwright, Jordan, AoS 2009; Nachman, JDT, PRD 2021]

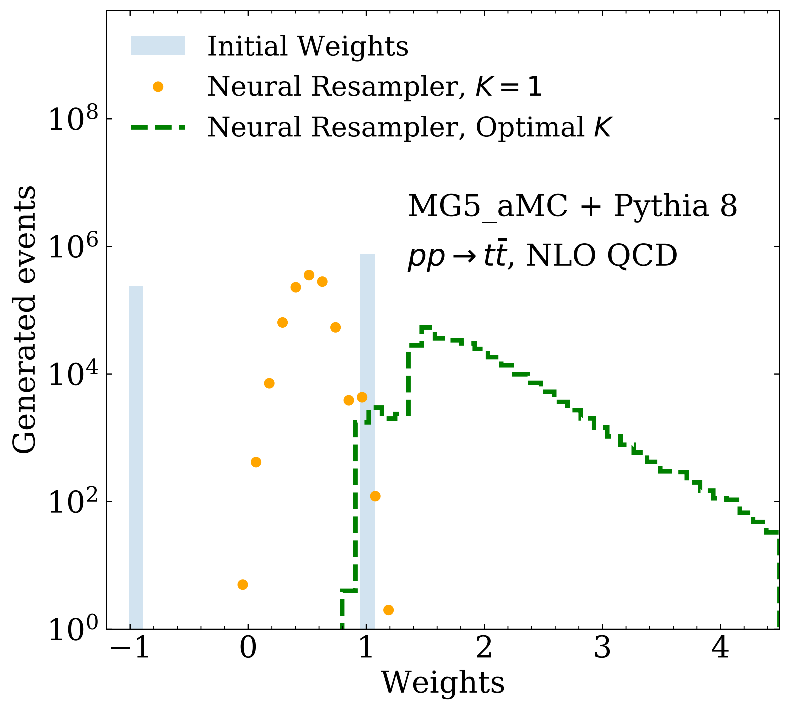
Applications of Likelihood Ratio Trick

Detector Unfolding



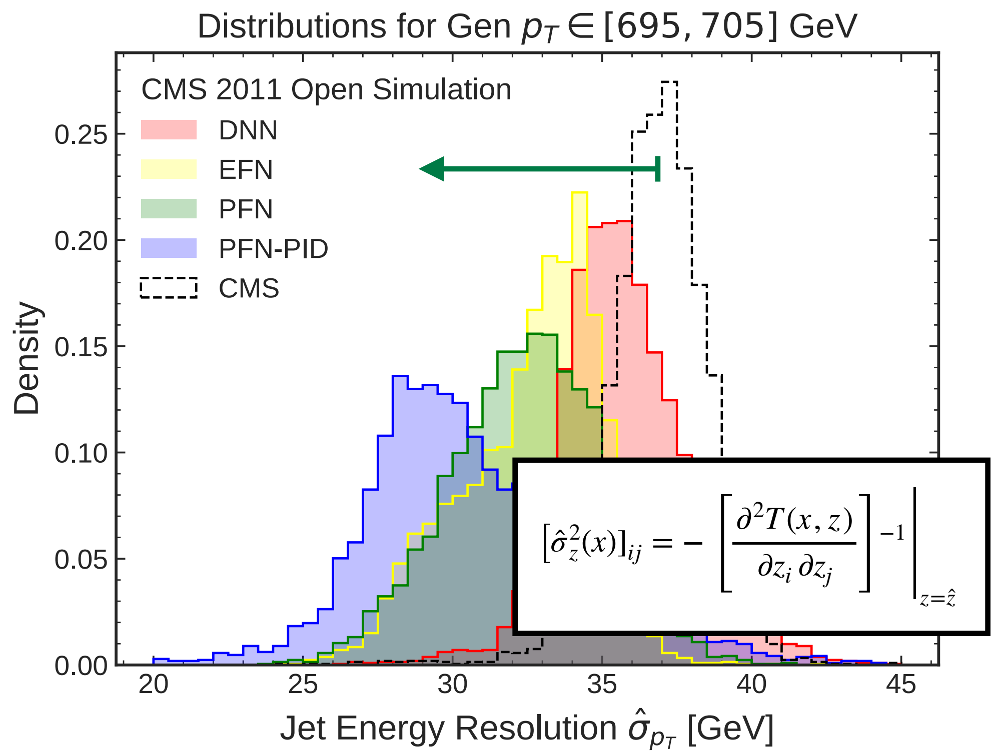
[Andreassen, Komiske, Metodiev, Nachman, JDT, PRL 2020; + Suresh, ICLR SimDL 2021]

Monte Carlo Reweighting



[Nachman, JDT, PRD 2020; inspired by Andersen, Gutschow, Maier, Prestel, EPJC 2020]

Resolution Estimation



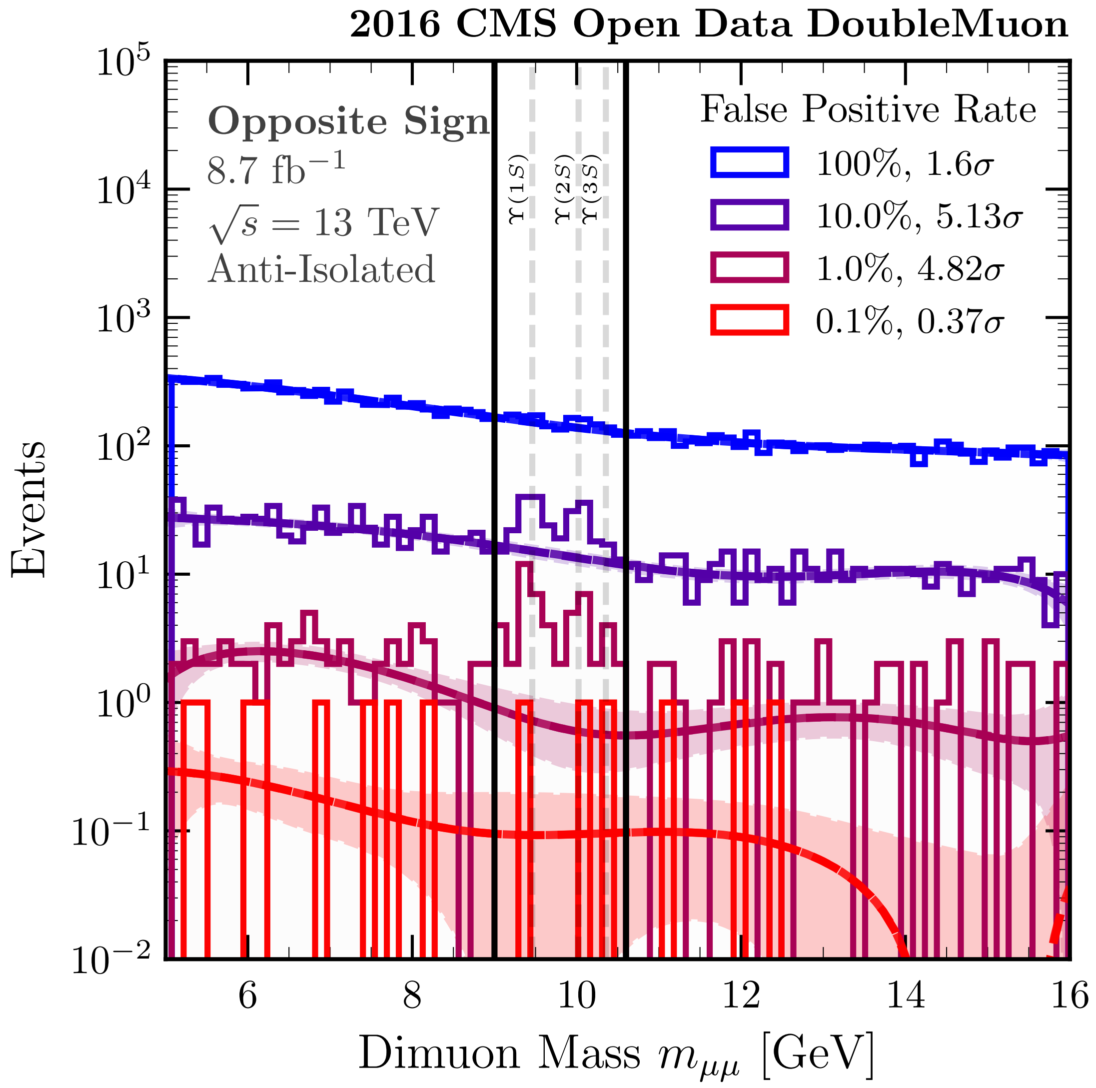
[Gambhir, Nachman, JDT, PRL 2022, PRD 2022]

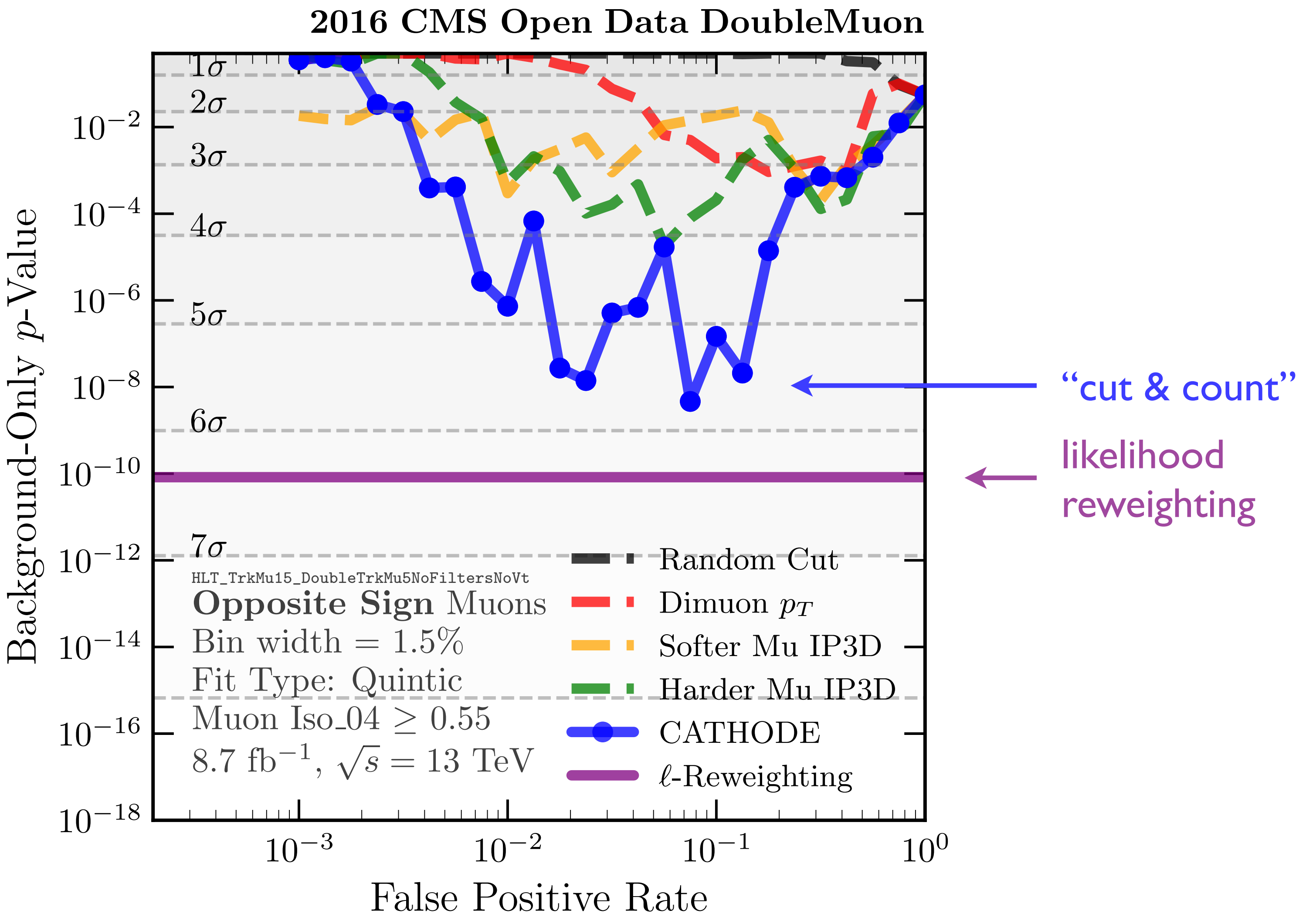
Theme: Convert sampled data into usable function approximation

Important subtlety (not for this talk): How do you quantify the accuracy/uncertainty of your approximation?

Recent Example: Rediscovering the Upsilon ($\Upsilon \to \mu^+\mu^-$)

Classification of observed data in signal region versus background interpolation from sidebands





Leverage relation to likelihood ratio to enhance sensitivity beyond simple cuts

[Gambhir, Mastandrea, Nachman, JDT, arXiv 2025;
using CATHODE in Hallin, Isaacson, Kasieczka, Krause, Nachman, Quadfasel, Schlaffer, Shih, Sommerhalder, PRD 2022]

Information Theory and Optimization

"Predicting" a distribution by maximizing entropy

Shannon Entropy: $$H(X) = \sum_{i=1}^{n} p(x_i) \log \frac{1}{p(x_i)}$$

Maximum: $$p(x_i) = \frac{1}{n}$$

Continuum Version

Relative Entropy: $$-D_{\mathrm{KL}}(P\|Q) = \int dx\, p(x) \log \frac{q(x)}{p(x)} + (\beta_0 - 1)\Big(1 - \int dx\, p(x)\Big)$$

(KL divergence we saw before)

Lagrange Multiplier to enforce normalization

Extremum: $$p(x) = e^{-\beta_0} q(x) \qquad \beta_0 = 0 \qquad \Longrightarrow \qquad p(x) = q(x)$$

Maximizing (unconstrained) entropy is equivalent to choosing a prior

Information Theory and Constrained Optimization

"Predicting" a distribution by maximizing entropy with fixed moments

Relative Entropy: $$-D_{\mathrm{KL}}(P\|Q) = \int dx\, p(x) \log \frac{q(x)}{p(x)} + (\beta_0 - 1)\left(1 - \int dx\, p(x)\right)$$

Plus Moment Constraints: $$+\sum_j \beta_j \left(c_j - \int dx\, p(x) f_j(x)\right)$$

Extremum: $$p(x) = \exp\left[-\beta_0(x) - \sum_j \beta_j(x)\, f_j(x)\right] q(x)$$

Same manipulation as Boltzmann's approach to statistical mechanics!

Lagrange multipliers set to values that satisfy constraints

Information Theory and Numerical Constrained Optimization

"Predicting" a distribution by maximizing entropy with fixed moments and auto-differentiation

Extremum: $$p(x) = \exp\Big[-\beta_0(x) - \sum_j \beta_j(x)\, f_j(x)\Big]\, q(x)$$

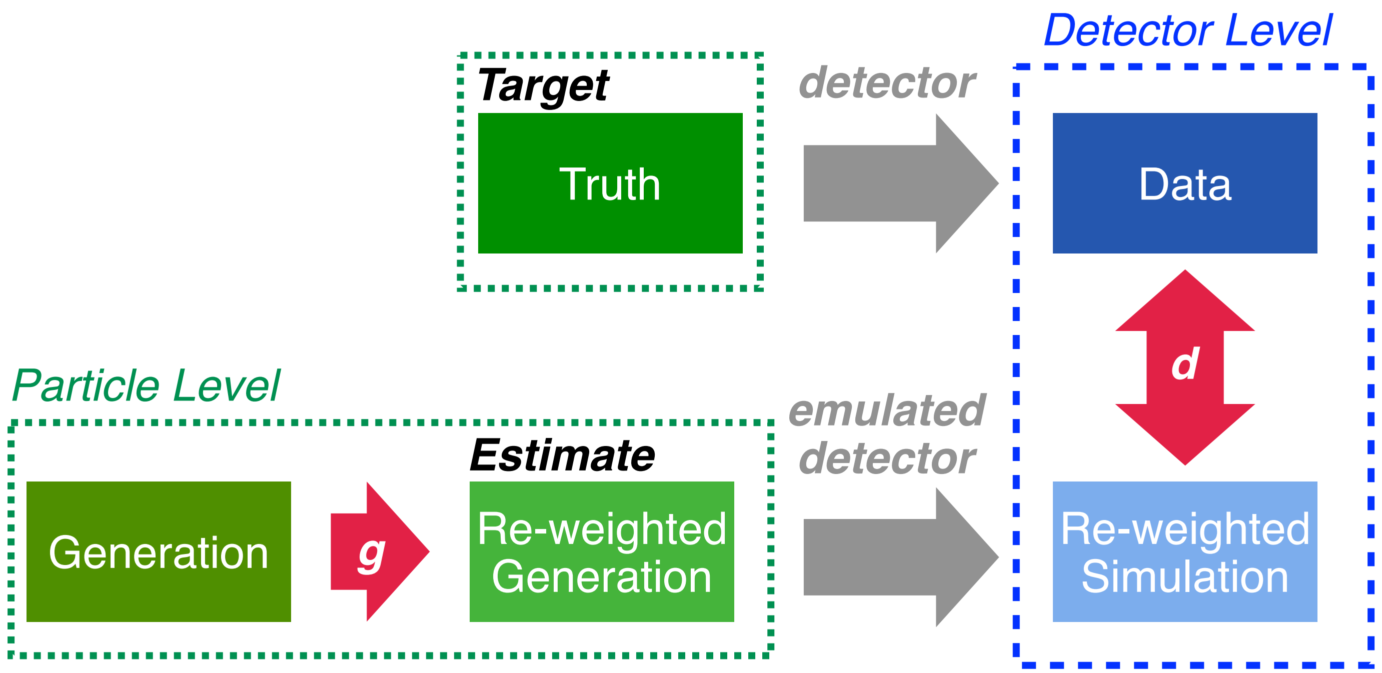
Target Moment: c_j

Current Moment: $$d_j = \int dx\, p(x)\, f_j(x)$$

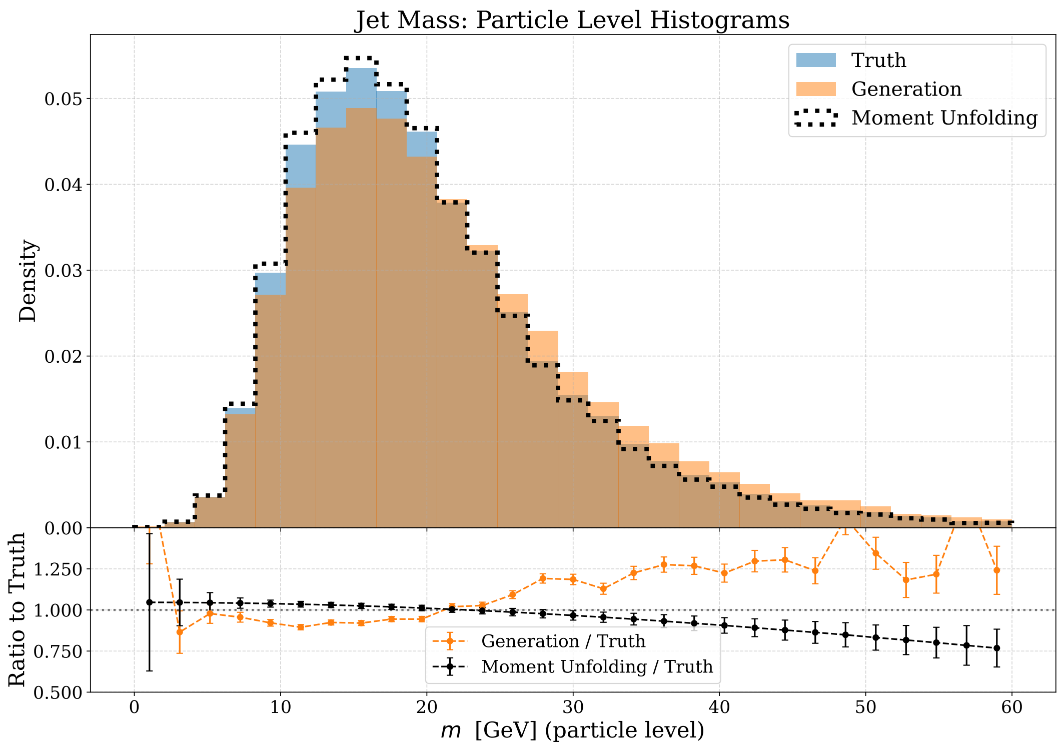
Loss Function: $$\mathcal{L} = \sum_j \left(\frac{c_j - d_j}{c_j + d_j}\right)^2$$

Numerically minimize loss to estimate Lagrange multipliers
As long as they aren't strictly incompatible, works for any set of basis functions

Alternative: Information Theory and Adversarial Networks



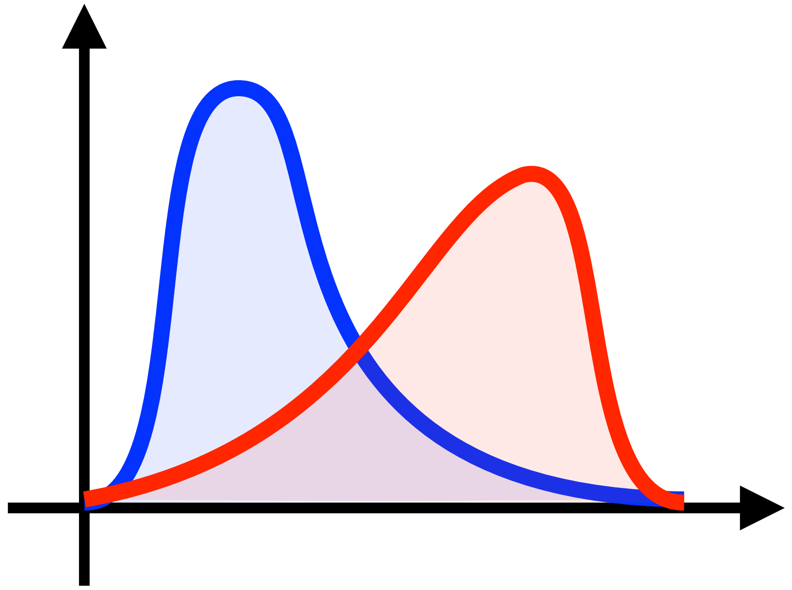
Reversing likelihood ratio trick, you can "integrate in" discriminator (d) to turn entropy minimization to adversarial optimization of weight function (g)



This allows you to unfold the detector response and learn moments from experimental data

(Prior is theory MC generator used in simulation)

[Desai, Nachman, JDT, PRD 2024]



Machine learning provides a bridge between well-established statistical methodologies and sampled data in particle physics (and beyond)

Colloquium Version of the Story

When physicists "think like a machine", we can translate thorny theoretical challenges into optimization problems

Basics of ML and Information Theory

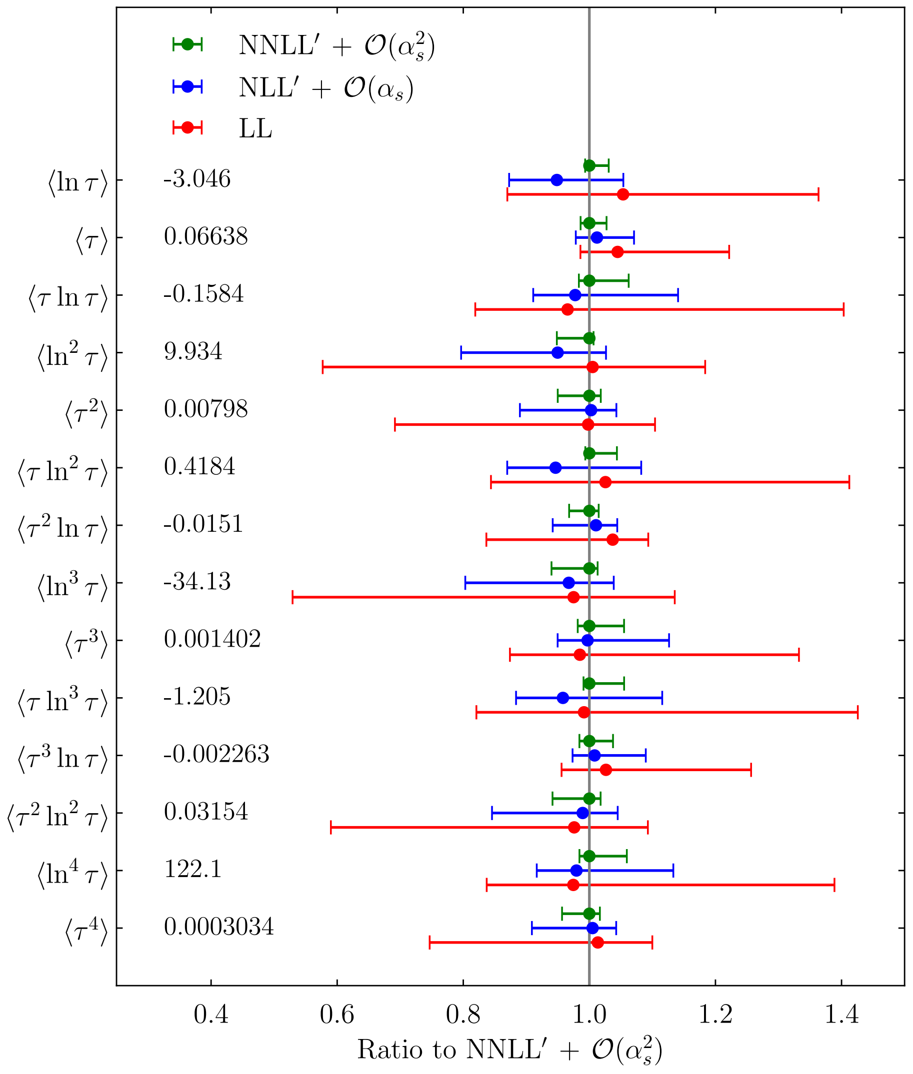
We can leverage the flexibility of loss functions to learn and constrain physically meaningful moments of distributions

Logarithmic Moments and QCD

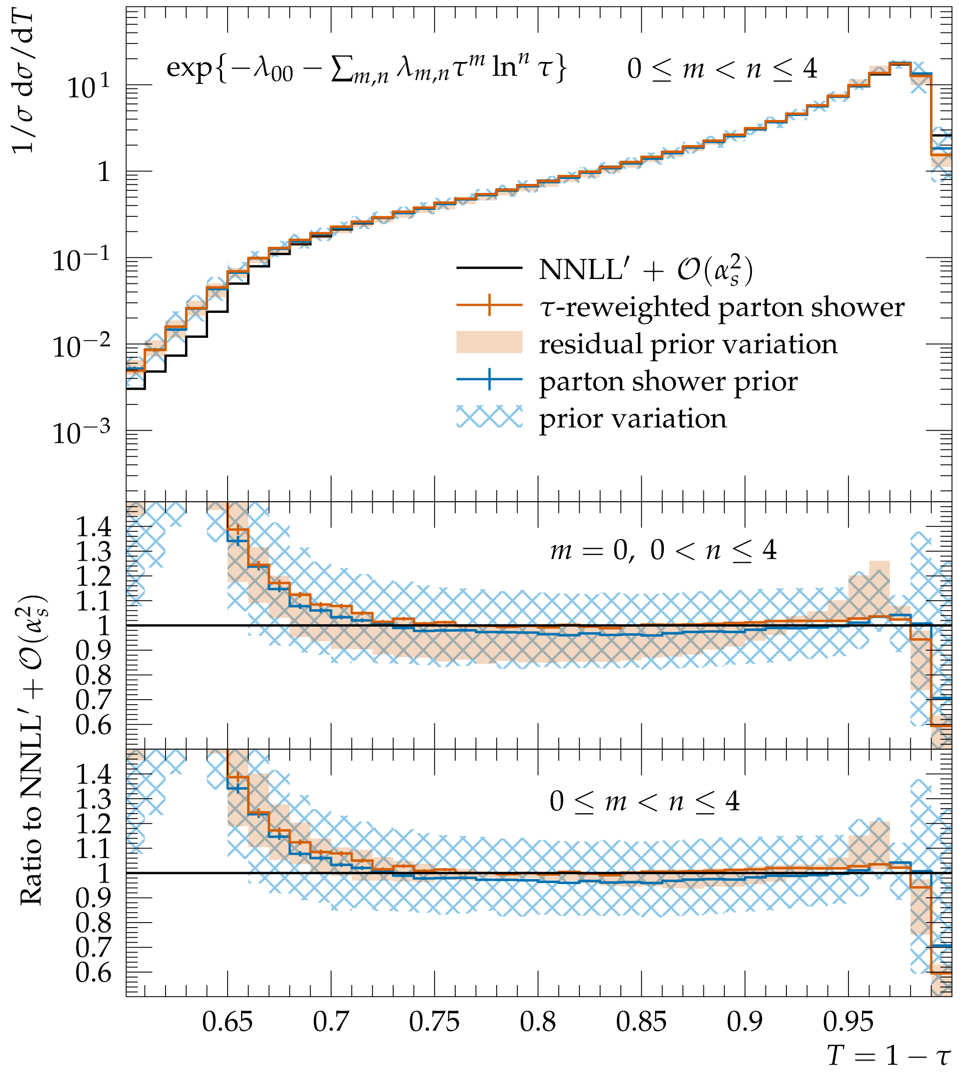
Logarithmic moments are particularly powerful probes of QCD that distinguish resummed from fixed-order calculations

Returning to Proof-of-Concept Study with “Thrust”

First QCD Calculation of Thrust Logarithmic Moments!



Priors from imperfect generators yield consistent results after moment reweighting!

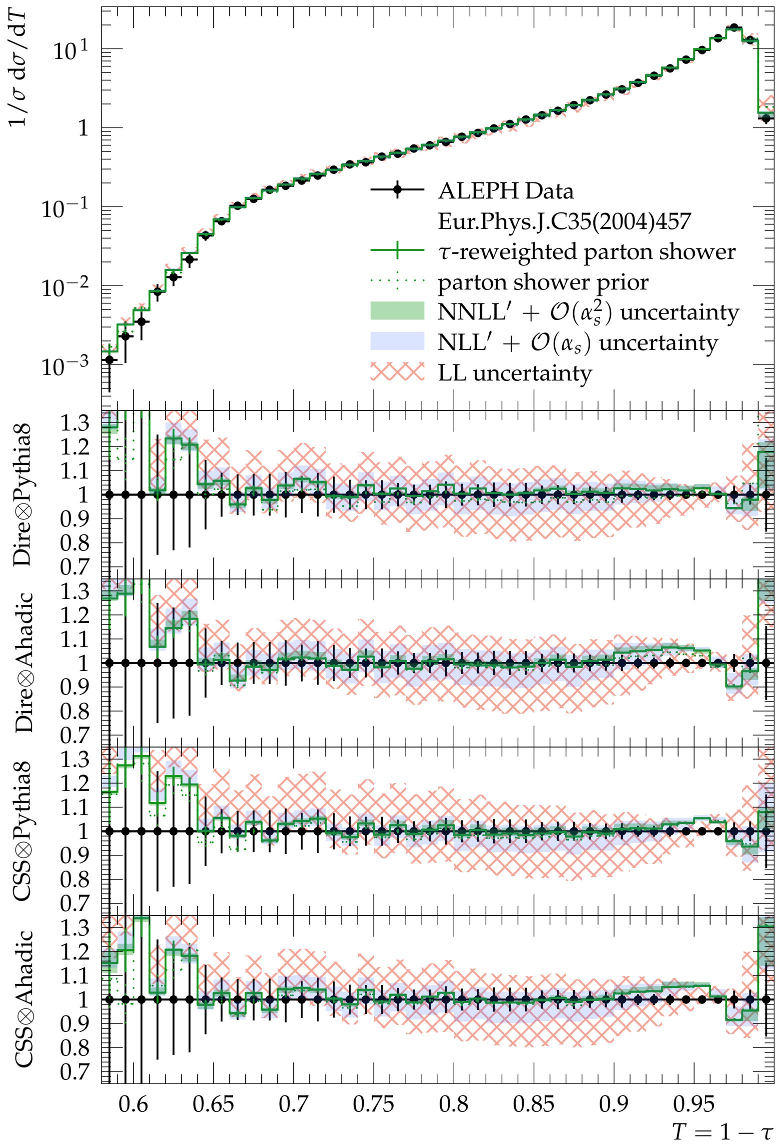


[Assi, Lee, Höche, JDT, arXiv 2025;
see Sudakov safety in Larkoski, Marzani, JDT, PRD 2015;
see related moment construction in Desai, Nachman, JDT, PRD 2024]



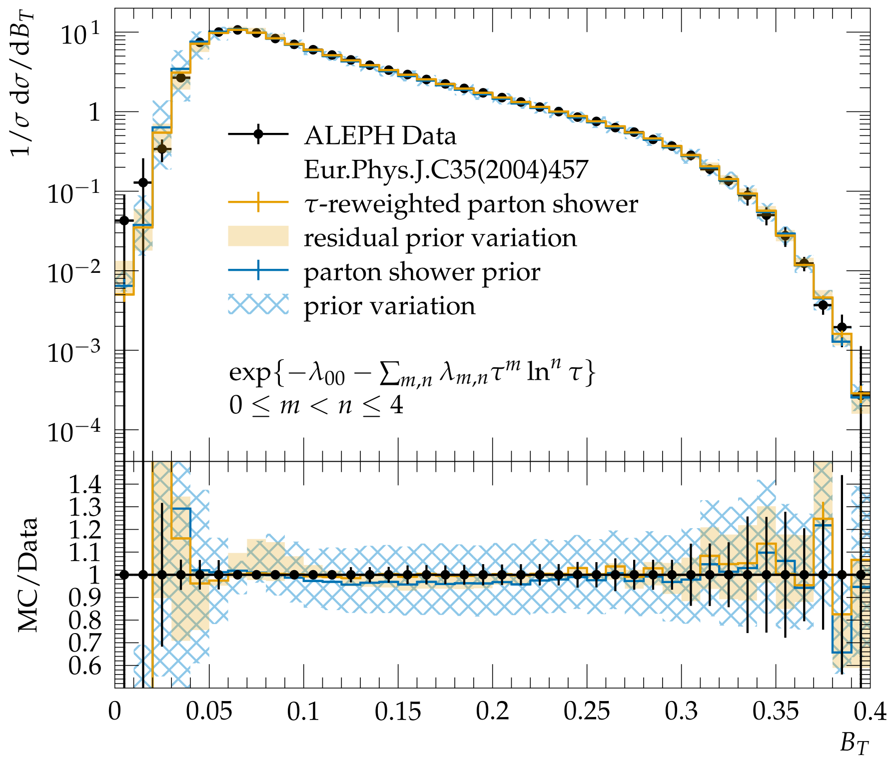
Pushing the Precision Frontier

Comparison to LEP Data

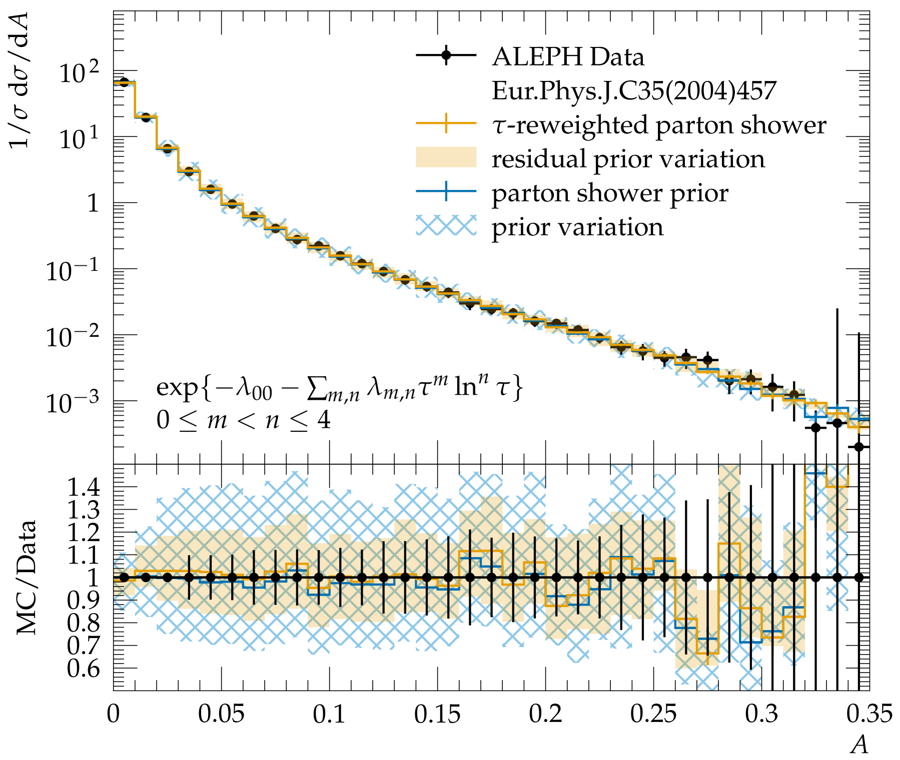


Impact on Complementary Observables

Broadening ≃ Thrust



Aplanarity ≉ Thrust



Thrust moments improve predictions for dijet kinematics

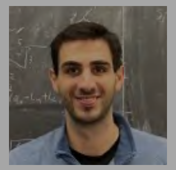
[Assi, Lee, Höche, JDT, arXiv 2025]

Pushing the Precision Frontier

By pattern matching, we drew an analogy between Sudakov form factors and Boltzmann factors, which motivated calculation of logarithmic moments

How can we understand the importance of logarithmic moments directly from QCD perspective?

[Assi, Lee, Höche, JDT, arXiv 2025]

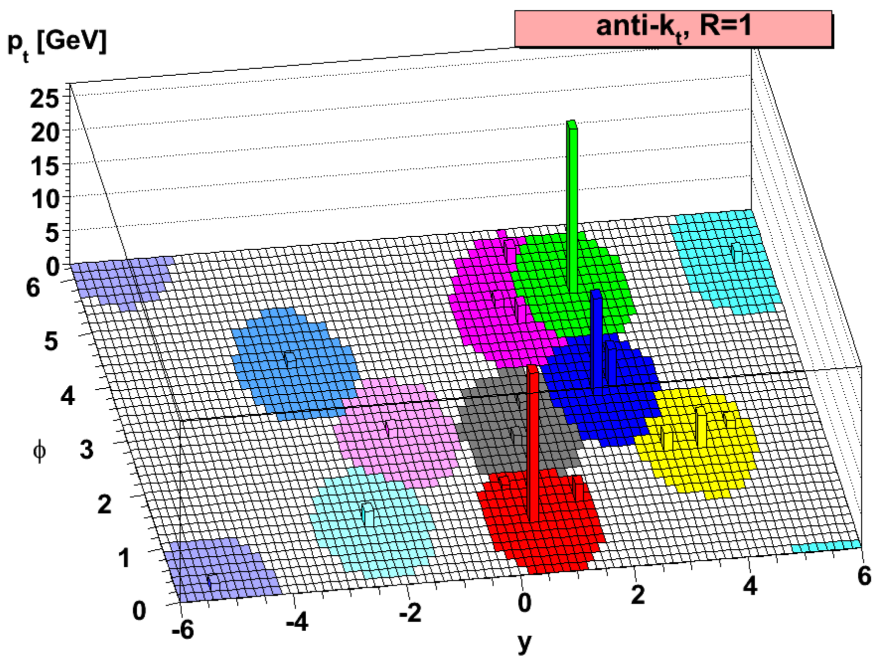


(

Aside 1: Computational Complexity as Muse

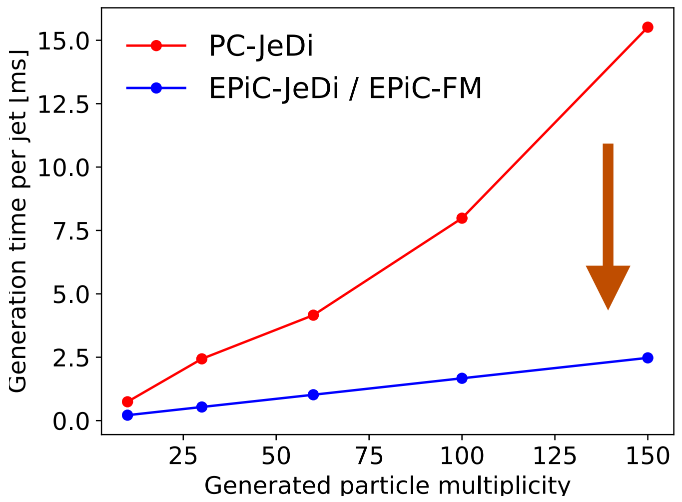
Jets with Anti-k_T:
O(M log M) vs. O(M³)
to cluster M particles

[Cacciari, Salam, Soyez, JHEP 2008]



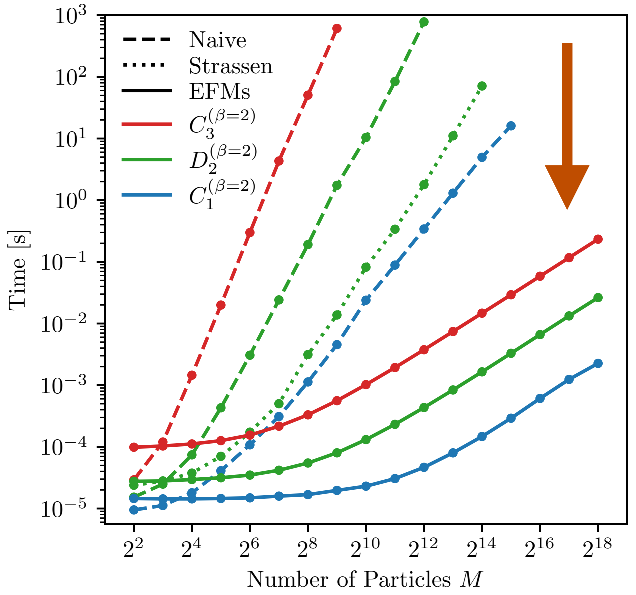
When you are trying to make sub-microsecond decisions, polynomial speedups can open up new scientific vistas!

O(M) vs. O(M²) Jet Generation



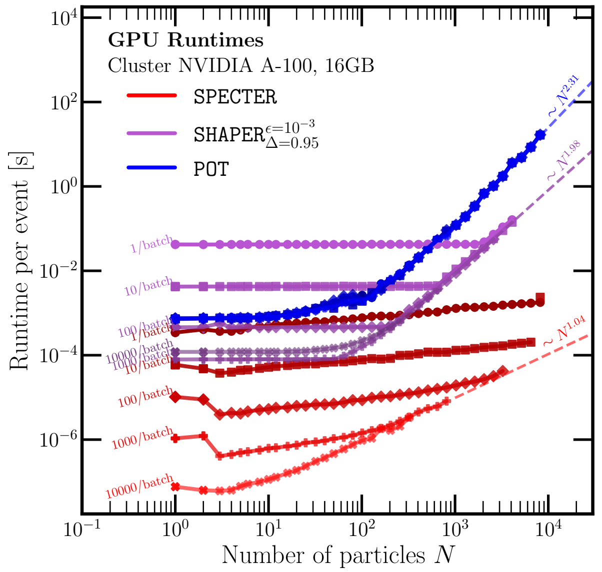
[Figure from Buhmann, et al., arXiv 2023;
based on Buhmann, Kasieczka, JDT, SciPost 2023]

O(M) vs. O(M⁴) Jet Classification



[Komiske, Metodiev, JDT, PRD 2020]

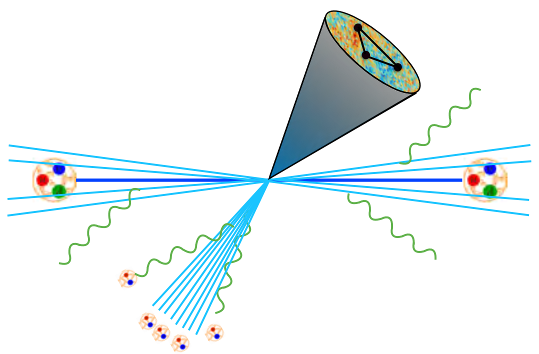
1000x Faster Optimal Transport



[Gambhir, Larkoski, JDT, JHEP 2024]

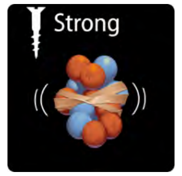
Faster Computation for Better Science

Example outside of ML: Energy correlators

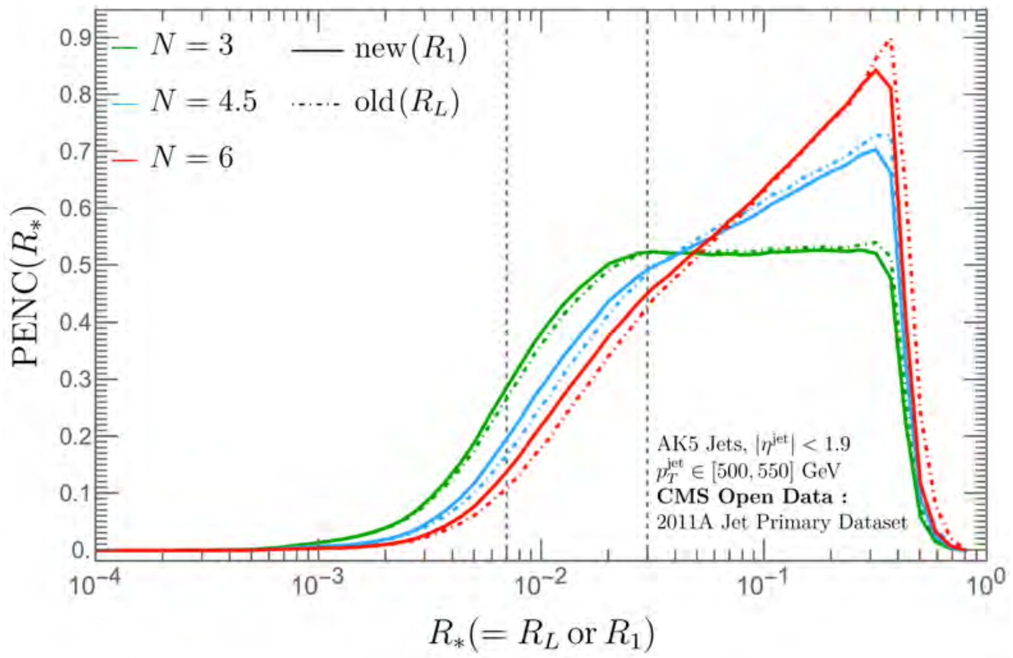


Probing the QCD Phase Transition

Hadronic Phase



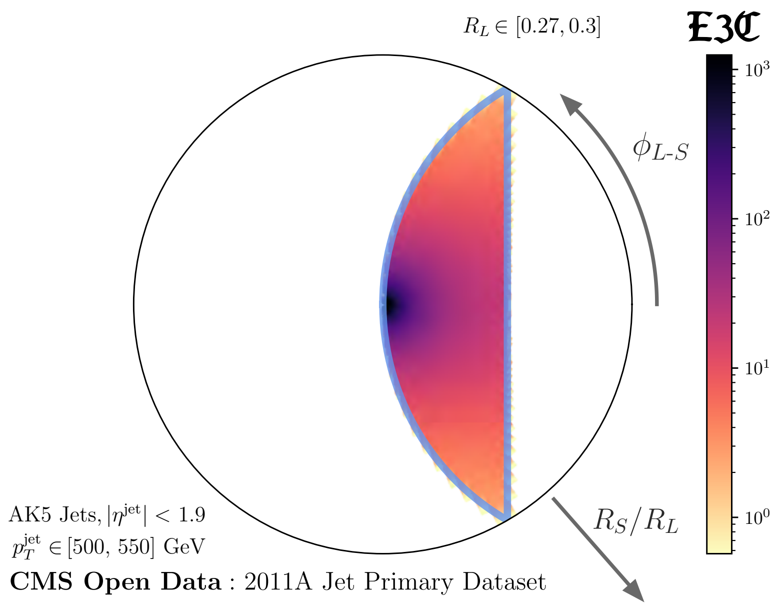
Partonic Phase



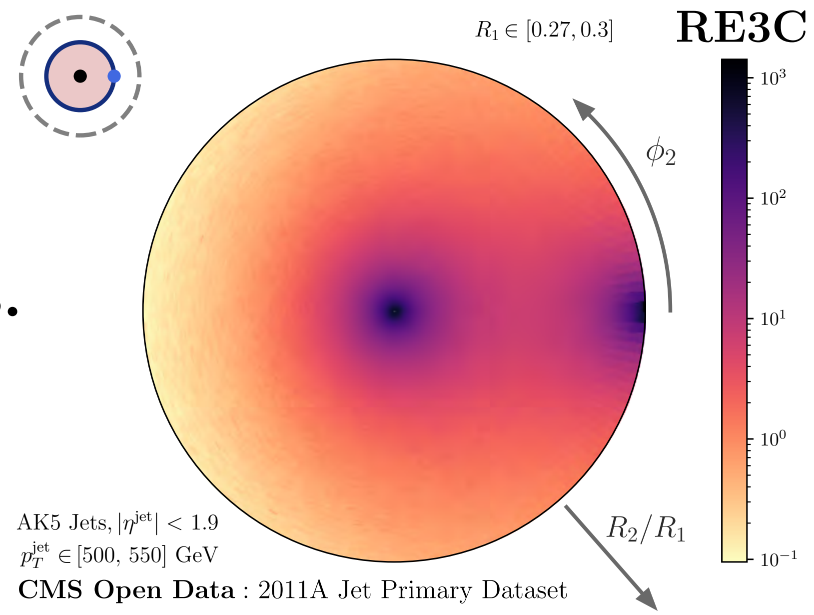
Old method (dashed): O(M^N)

New method (solid): O(M^2 log M) ∀ N

Algorithmic Advances inspire Enhanced Visualizations!



vs.

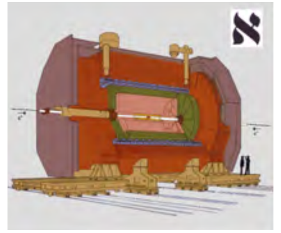


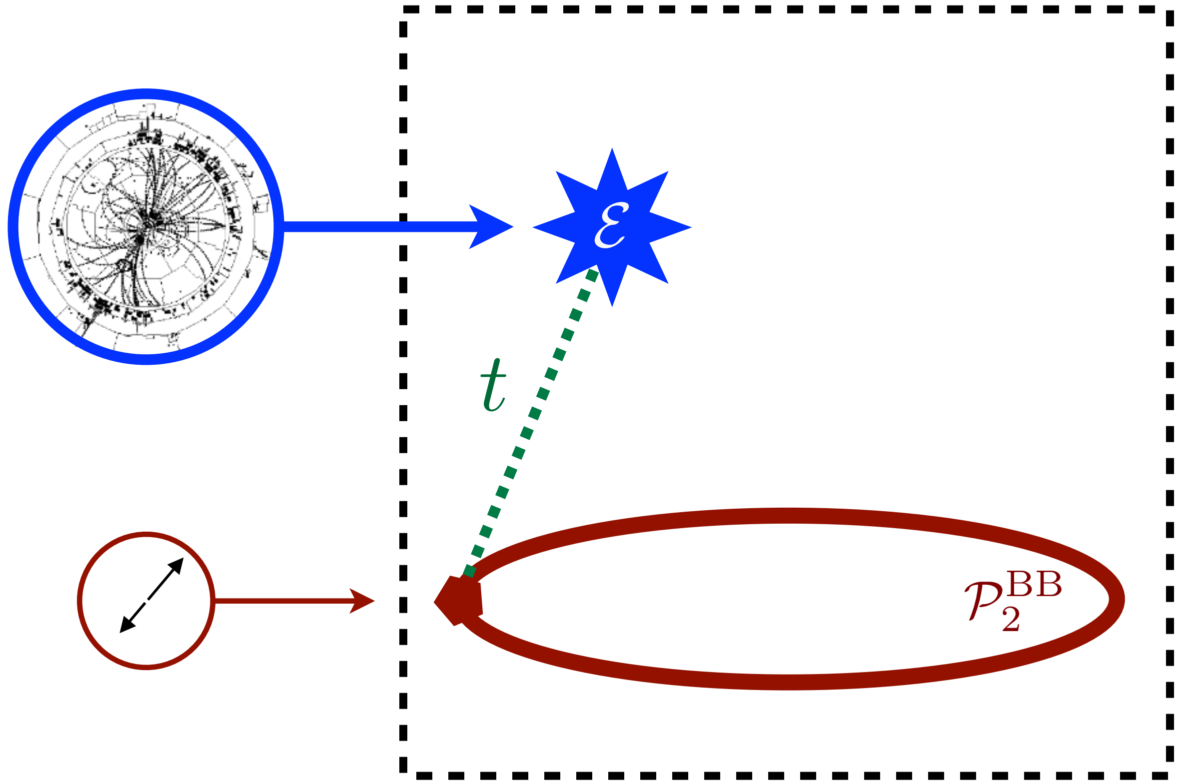
Both are O(M^3) but new method isolates distinct physics effects

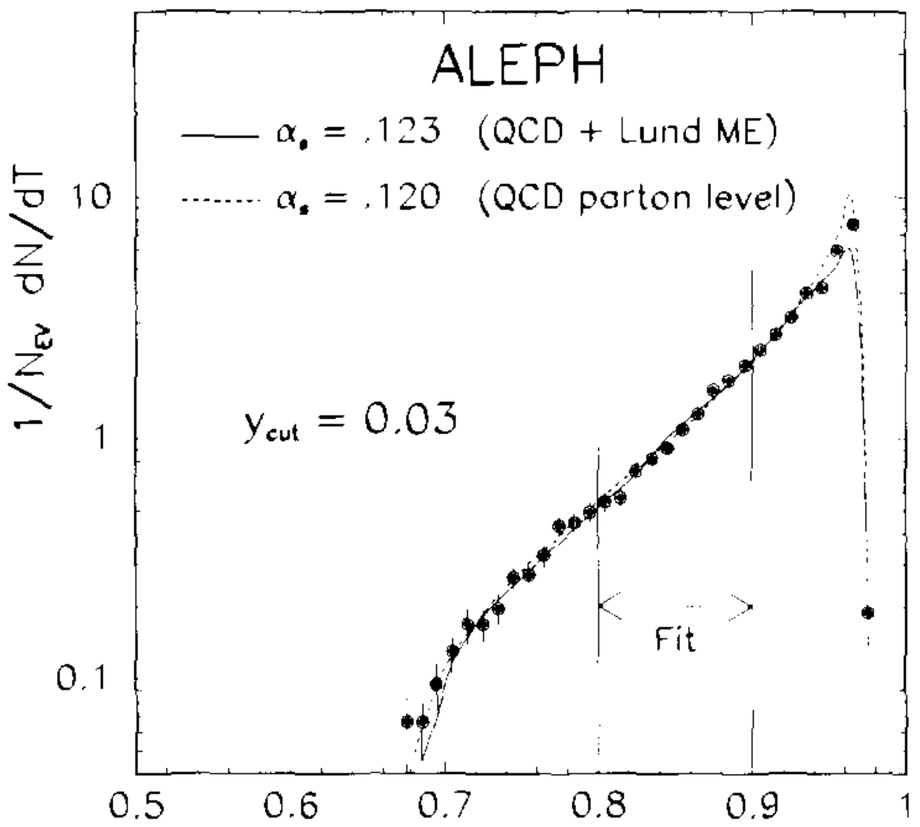
[Alipour-fard, Budhraja, JDT, Waalewijn, arXiv 2024;
[variant of Basham, Brown, Ellis, Love, PRL 1978; see also Komiske, Moult, JDT, Zhu, PRL 2022]

Aside 2: Collider Physics meets Optimal Transport

E.g. Thrust: how dijet-like is an event?







$$1 - \frac{t}{2E_{\mathrm{CM}}}$$

All Back-to-Back Two Particle Configurations

$$\mathcal{P}_2^{\mathrm{BB}} = \{ \quad \cdots \}$$

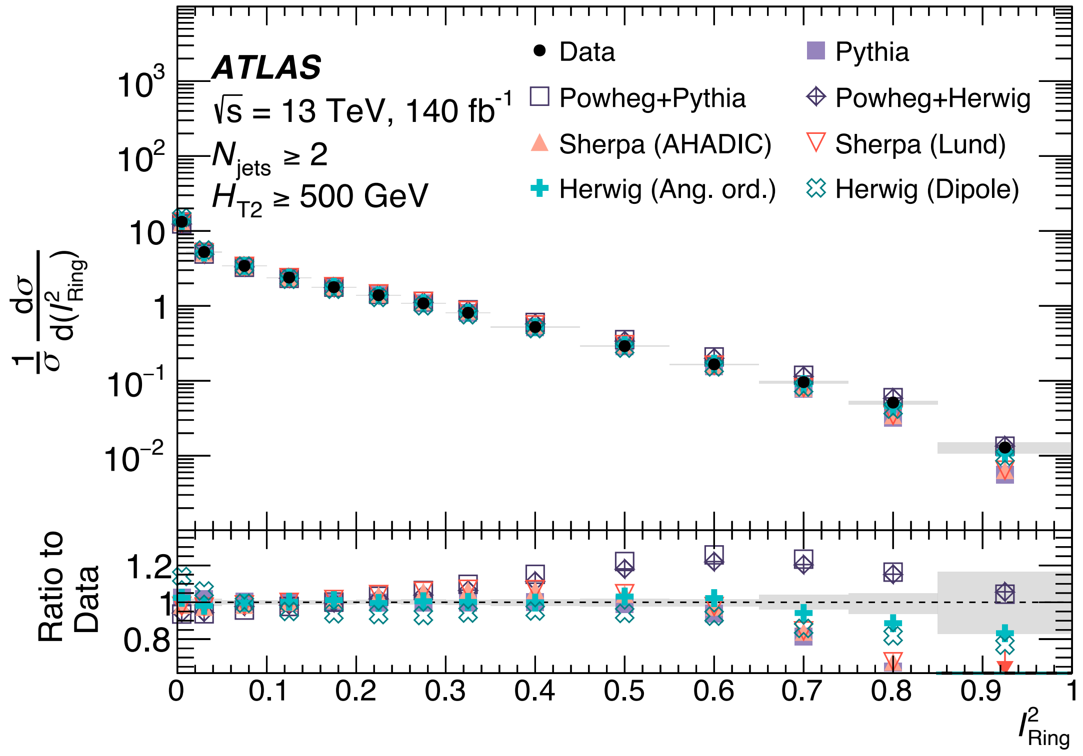
(using β=2 EMD variant)

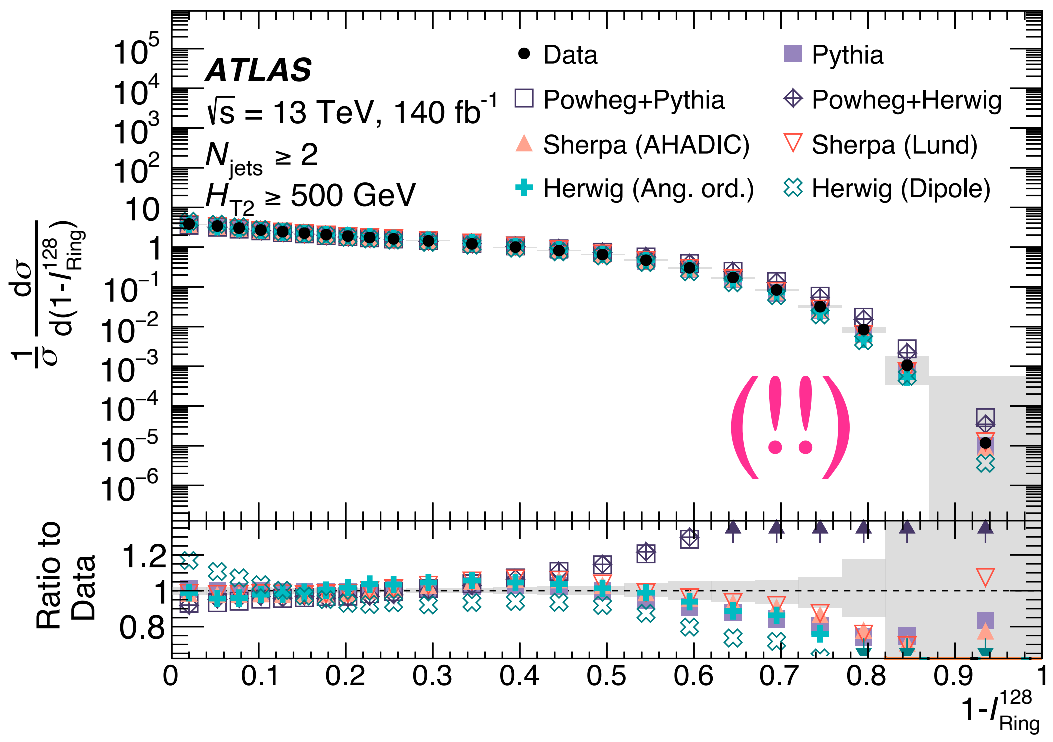
$$t(\mathcal{E}) = \min_{\mathcal{E}' \in \mathcal{P}_2^{\mathrm{BB}}} \mathrm{EMD}_2(\mathcal{E}, \mathcal{E}')$$

[Komiske, Metodiev, JDT, JHEP 2020]
[Brandt, Peyrou, Sosnowski, Wroblewski, PL 1964; Farhi, PRL 1977; ALEPH, PLB 1991]

Event Isotropy through Optimal Transport

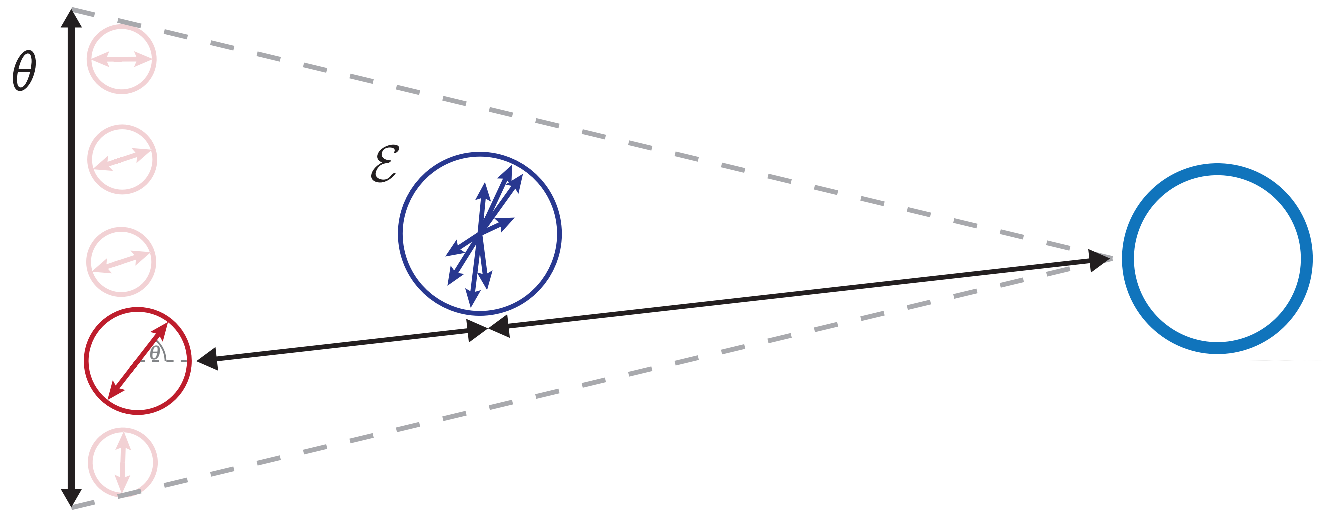
How isotropic is an event?





← Dijet-like

Isotropic →



[Cesarotti, JDT, JHEP 2020; ATLAS, JHEP 2023; see also Cesarotti, Reece, Strassler, JHEP 2021]
[generalizations in Ba, Dogra, Gambhir, Tasissa, JDT, JHEP 2023; Gambhir, Larkoski, JDT, JHEP 2024]

)

Lies, Damned Lies, and Statistics

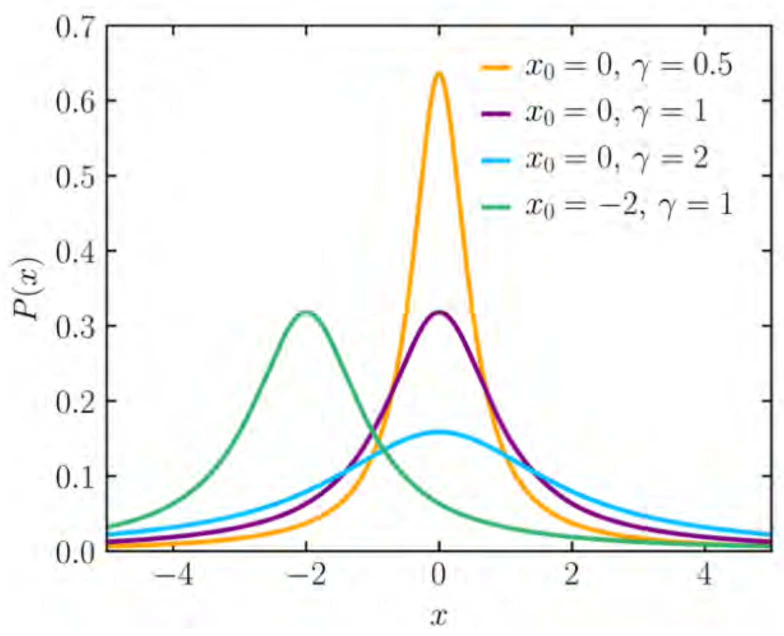
Problem of Moments (in 1-D): $d_j = \int dx\, p(x)\, x^j \qquad j \in \{1, 2, 3, \ldots\}$

Given infinite set of moments, reconstruct the underlying distribution

For physics applications, this is rarely the right question to ask

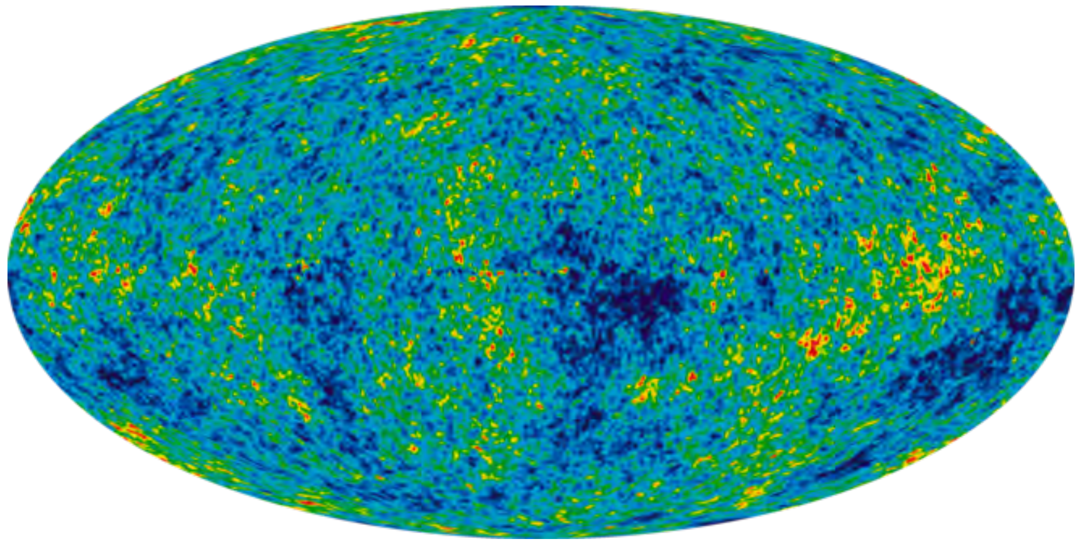
E.g. Cauchy distribution: $$p(x; x_0, \gamma) = \frac{1}{\pi} \frac{\gamma}{(x - x_0)^2 + \gamma^2}$$

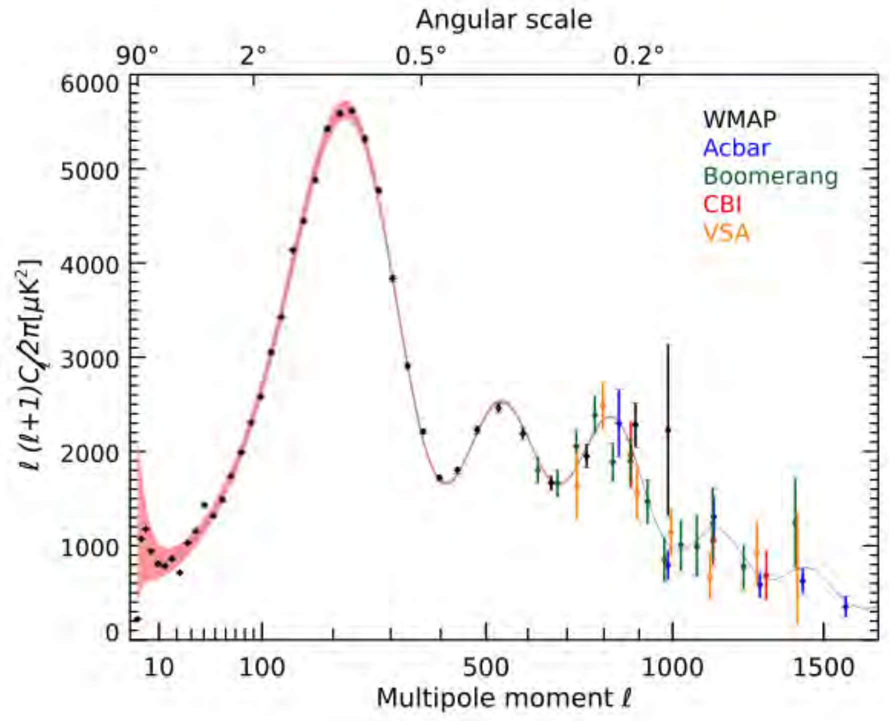
Famously, this has no mean (or higher moments), even though it is "obvious" that $\langle x \rangle$ should be x_0



Lies, Damned Lies, and Statistics

More relevant question: Given a probability distribution, what summary statistics best capture the key physical properties?



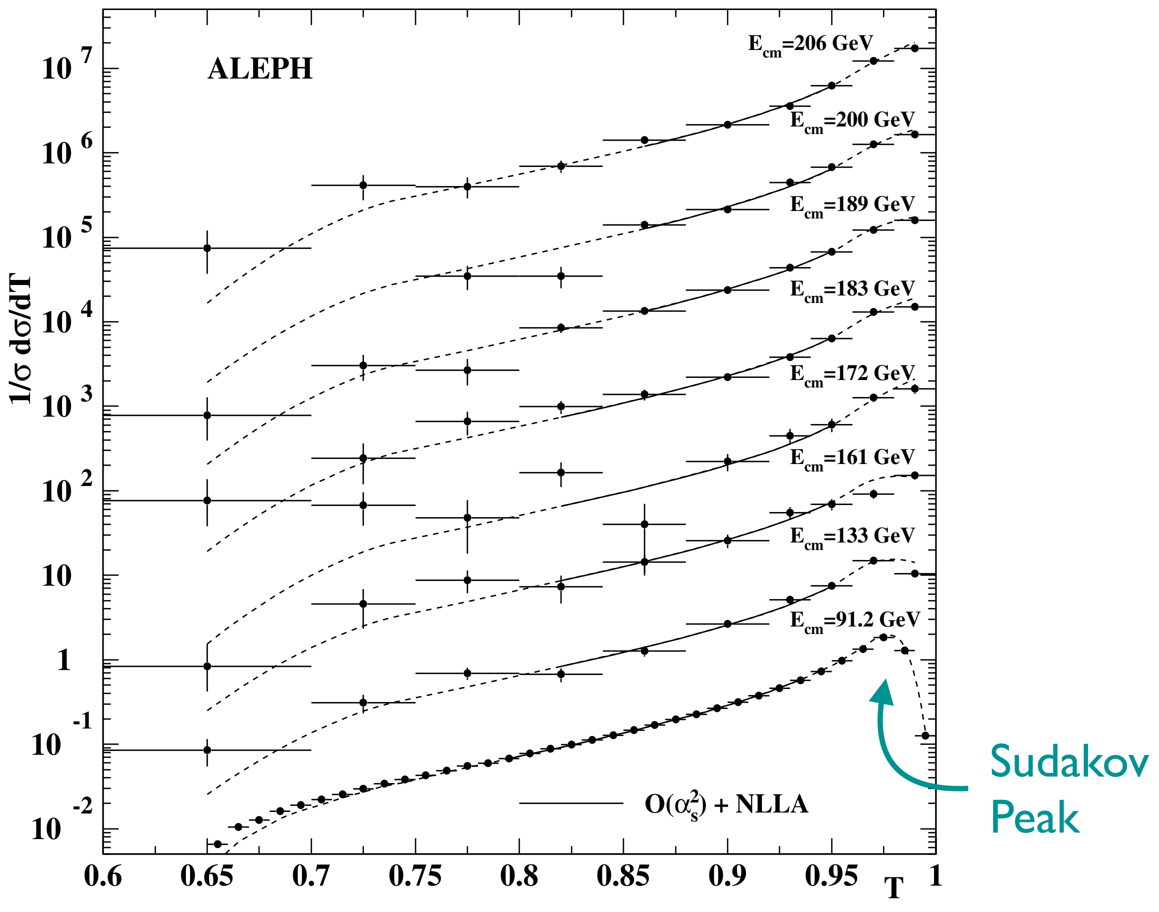


I don't know a generic way to answer this question (do you?)

Calculating Thrust

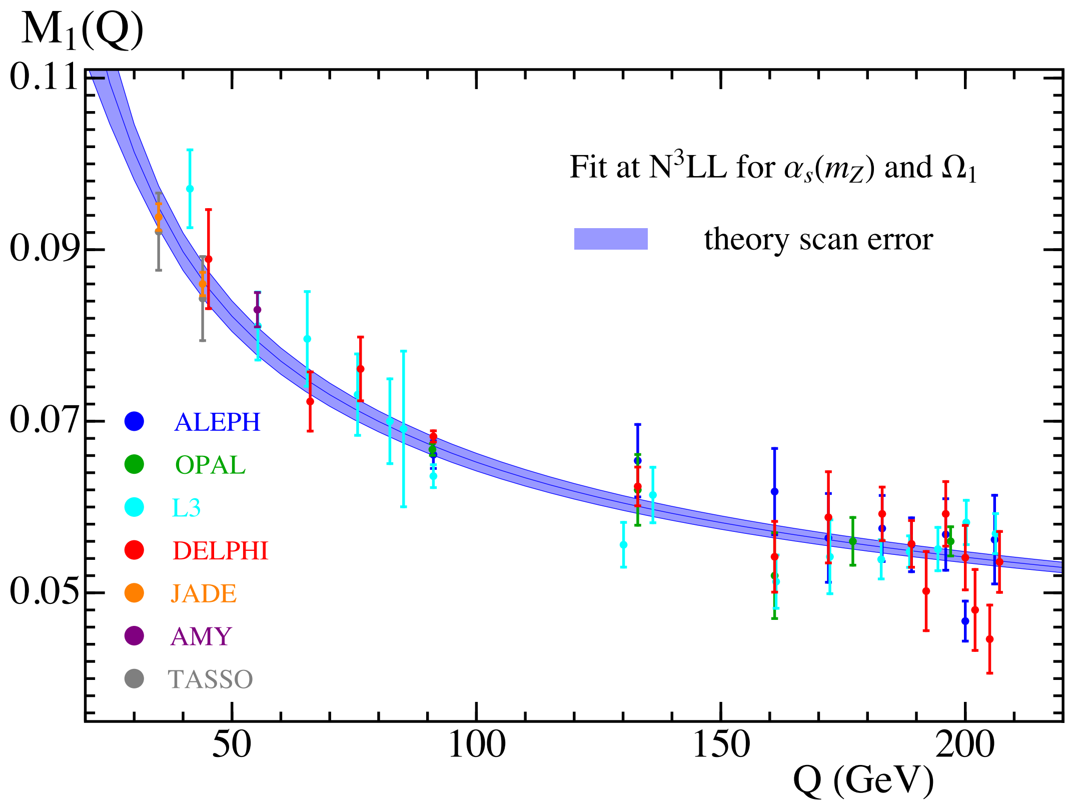
Extracting strong coupling constant from QCD event shapes

Thrust Distribution vs. Q



[ALEPH, EPJC 2024]

(One Minus) Thrust Mean vs. Q



[Abbate, Fickinger, Hoang, Mateu, Stewart, PRD 2012]

What moments best characterize… { Experimental thrust distribution? Theoretical thrust calculation?

Simplified Thrust Calculation

Leading log at fixed coupling

Sudakov Form Factor

$$p(\tau) = \frac{-2\alpha_s C_F}{\pi} \frac{\ln \tau}{\tau} \exp\left[-\frac{\alpha_s C_F}{\pi} \ln^2 \tau\right]$$

Ordinary Moments

$$\langle \tau^m \rangle = \frac{2\alpha_s C_F}{\pi} \frac{1}{m^2} + \mathcal{O}(\alpha_s^2)$$

Logarithmic Moments

$$\langle \ln^n \tau \rangle = (-1)^n \left(\frac{\pi}{\alpha_s C_F}\right)^{n/2} \Gamma\left[1 + \frac{n}{2}\right]$$

Example of "Sudakov Safe" Observable

[Assi, Lee, Höche, JDT, arXiv 2025;
Sudakov safety in Larkoski, Marzani, JDT, PRD 2015]



Simplified Thrust Calculation

Leading log at fixed coupling

Sudakov Form Factor

$$p(\tau) = \frac{-2\alpha_s C_F}{\pi} \frac{\ln \tau}{\tau} \exp\left[-\frac{\alpha_s C_F}{\pi} \ln^2 \tau\right]$$

Mean ⇒ Strong Coupling

$$\langle \tau \rangle = \frac{2\alpha_s C_F}{\pi} + \mathcal{O}(\alpha_s^2)$$

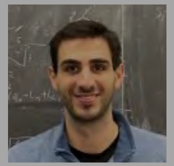
Characterizes Fixed-Order Information

Logarithmic Mean ⇒ Sudakov Peak

$$\langle \ln \tau \rangle = -\frac{\pi}{2\sqrt{\alpha_s C_F}}$$

Characterizes Resummed Information

[Assi, Lee, Höche, JDT, arXiv 2025;
Sudakov safety in Larkoski, Marzani, JDT, PRD 2015]

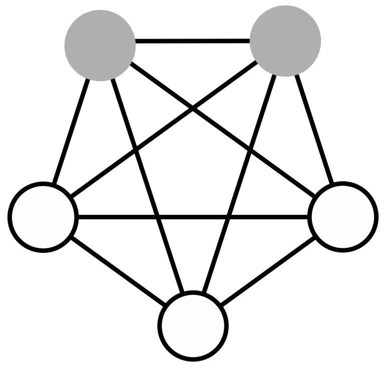


Many QCD Subtleties to Incorporate and Understand

For Thrust Specifically:	Sudakov Shoulder Resummation
	Three Jet Power Corrections
	Shape Function Parametrization
In General:	Choice of Priors
	Estimating Theory Uncertainties [see Tackmann, arXiv 2024]
	Non-Perturbative Corrections
	Treatment of Mixed Moments
	Convergence of Moment Expansion
Future Directions:	Multi-differential Distributions
	Alternative Basis Functions
	Inclusive Observables (e.g. EECs)

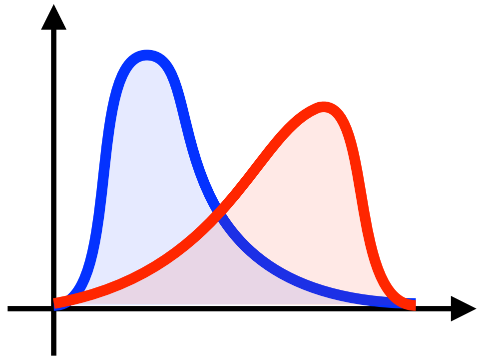
Different summary statistics capture different physical effects, and logarithmic moments offer important insights into QCD resummation

QCD Theory meets Information Theory



Colloquium Version of the Story

When physicists "think like a machine", we can translate thorny theoretical challenges into optimization problems



Basics of ML and Information Theory

We can leverage the flexibility of loss functions to learn and constrain physically meaningful moments of distributions

Logarithmic Moments and QCD

Logarithmic moments are particularly powerful probes of QCD that distinguish resummed from fixed-order calculations