

Collision Course

Particle Physics meets Machine Learning

Jesse Thaler



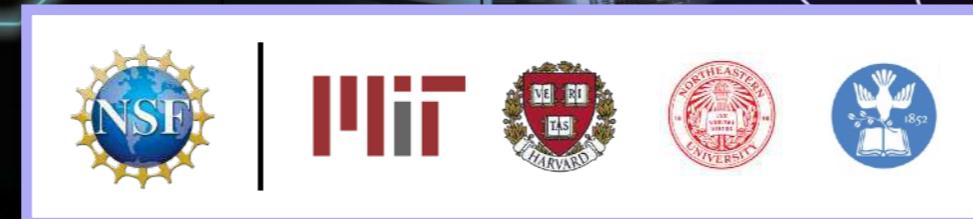
8.S50 — January 26, 2022

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

“eye- ϕ hi”



*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](#)]

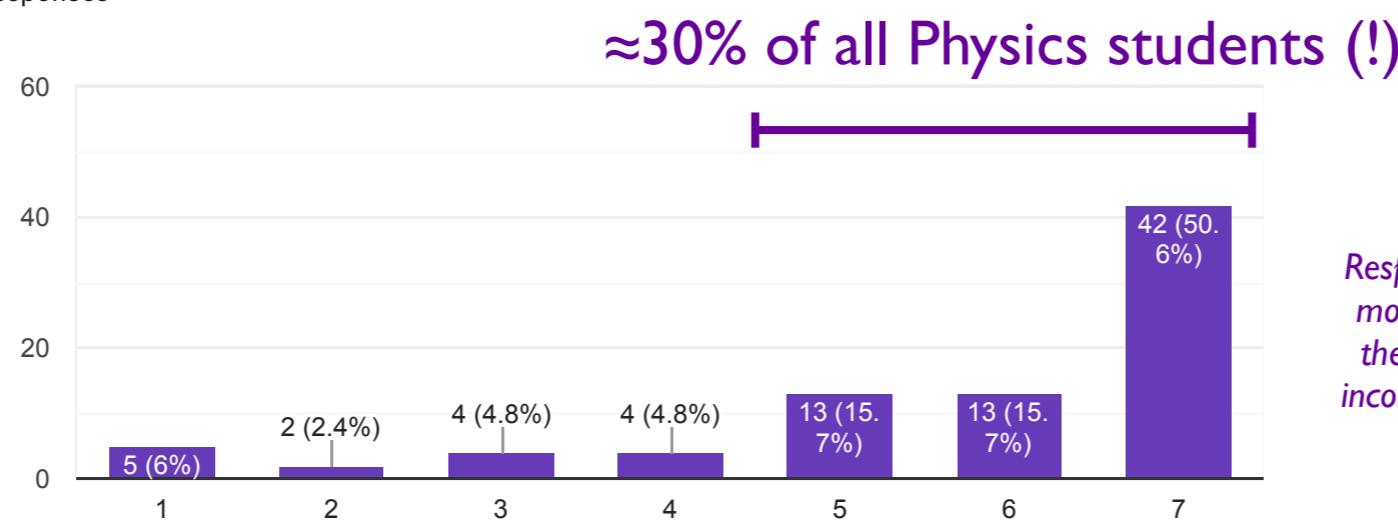
New! PhD in Physics, Statistics & Data Science

≈ Physics PhD + 4 courses (*probability, statistics, computation, data analysis*)

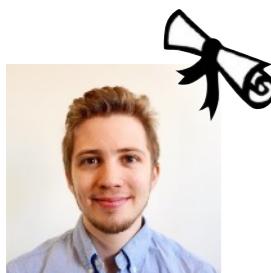


How interested would you be in submitting and defending a PhD thesis that uses statistical methods in a substantial way?

83 responses



Respondent #11: “I think ML is the most important thing happening in the world right now and should be incorporated into any STEM degree.”

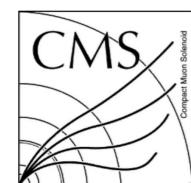
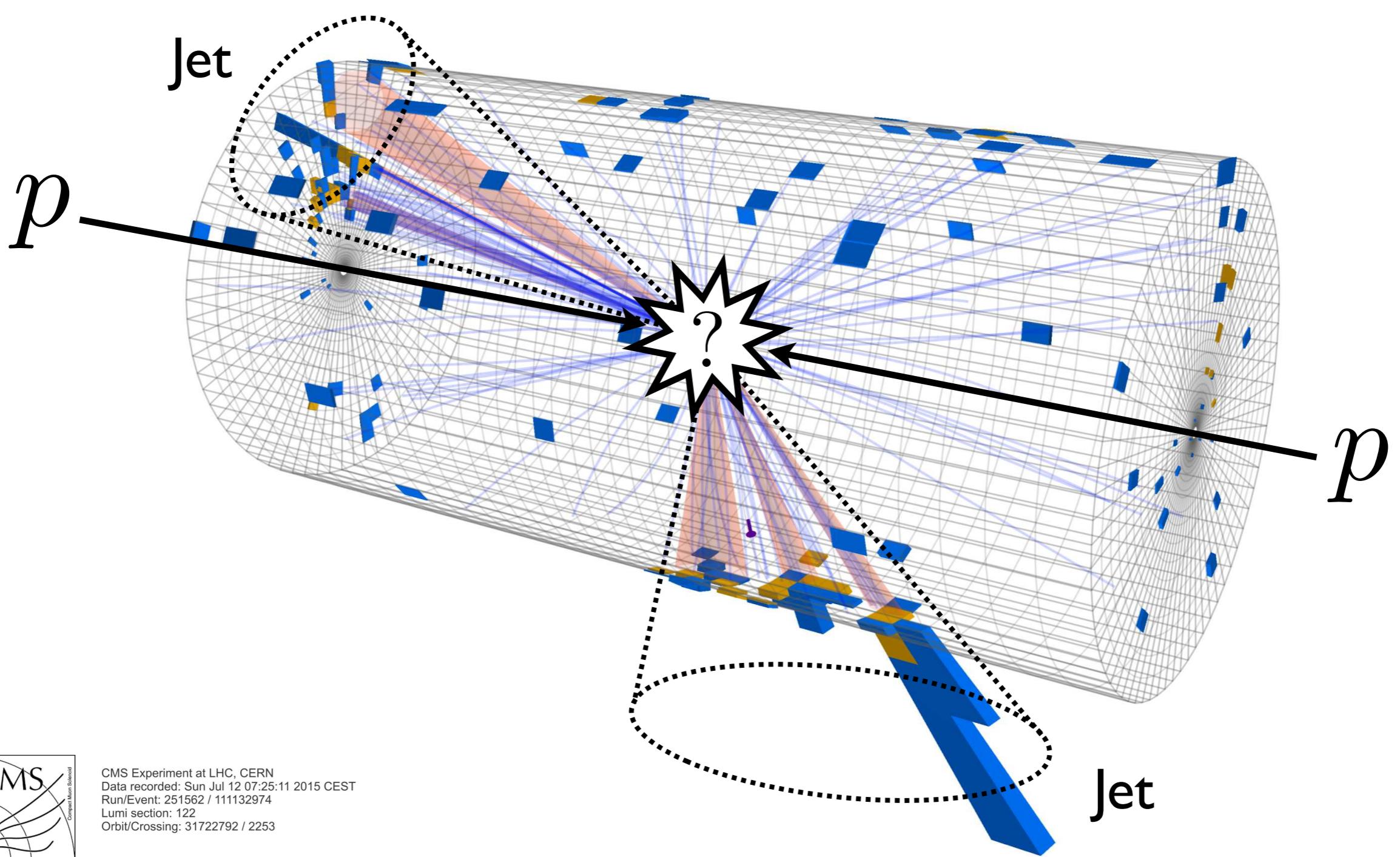


Congratulations,
Dr. Constantin Weisser!
(March 30, 2021)

[<https://physics.mit.edu/academic-programs/graduate-students/psds-phd/>]

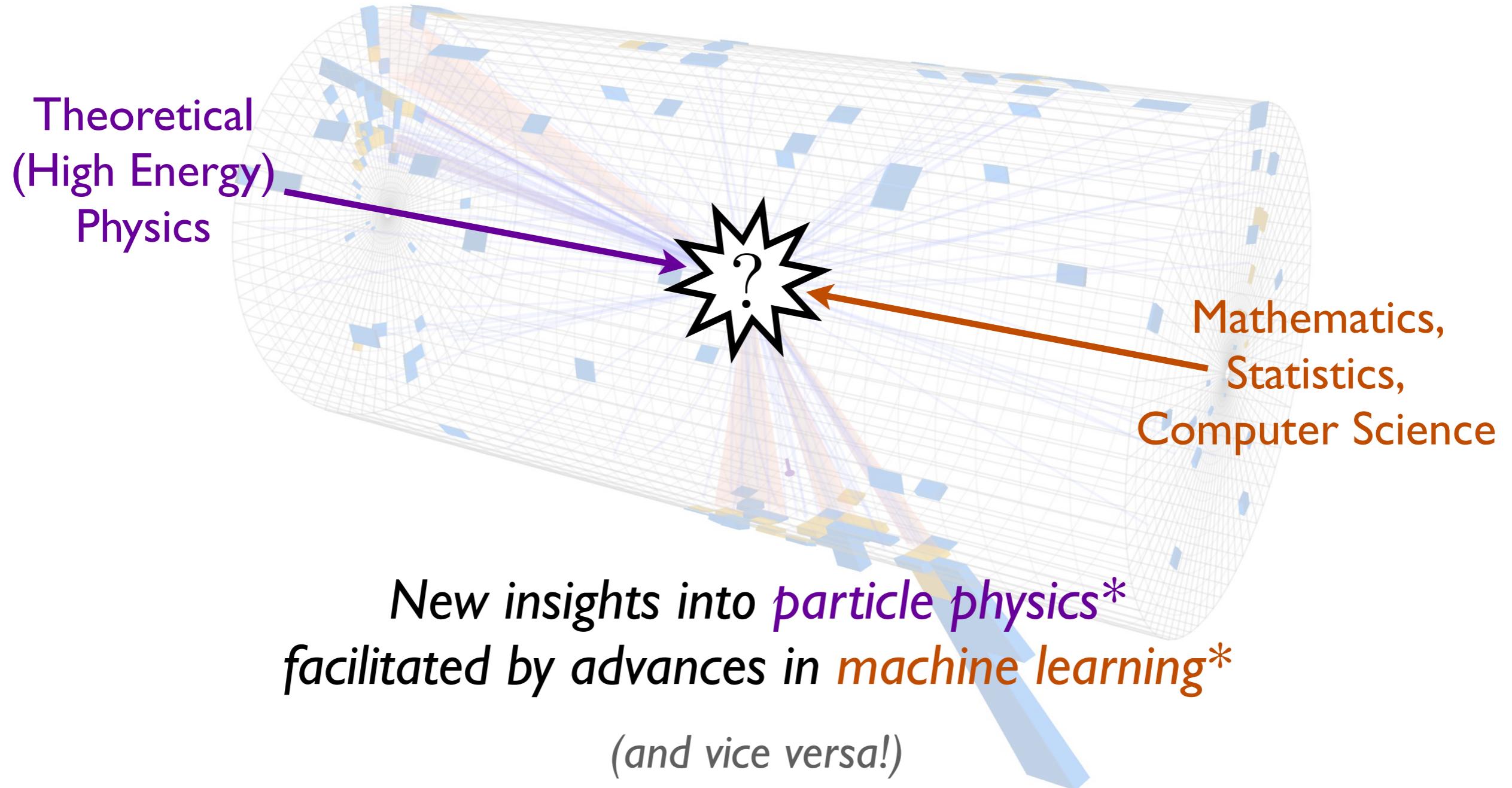
“Collision Course”

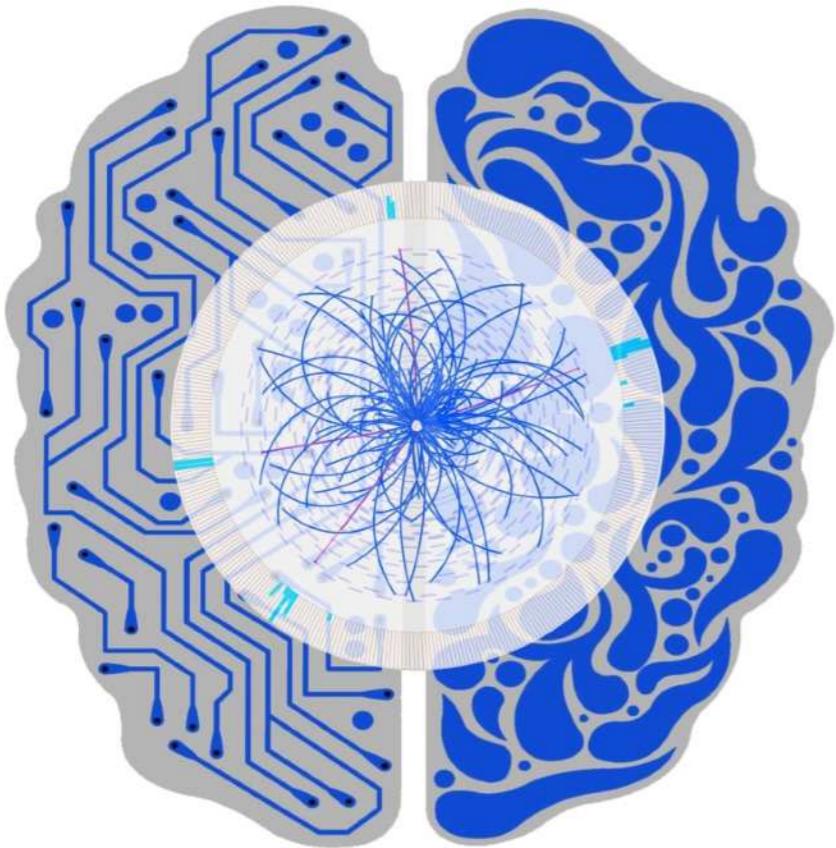
“Theoretical Physics for Machine Learning”
Aspen Center for Physics, January 2019



“Collision Course”

“Theoretical Physics for Machine Learning”
Aspen Center for Physics, January 2019





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

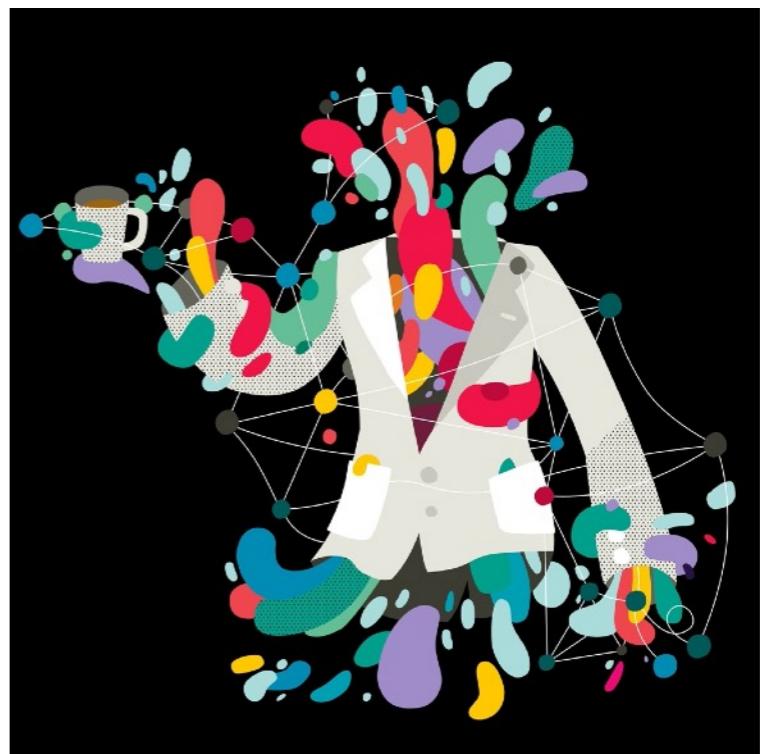
The New York Times



By Dennis Overbye

Nov. 23, 2020

Can a Computer Devise a Theory of Everything?



AI²: Ab Initio Artificial Intelligence

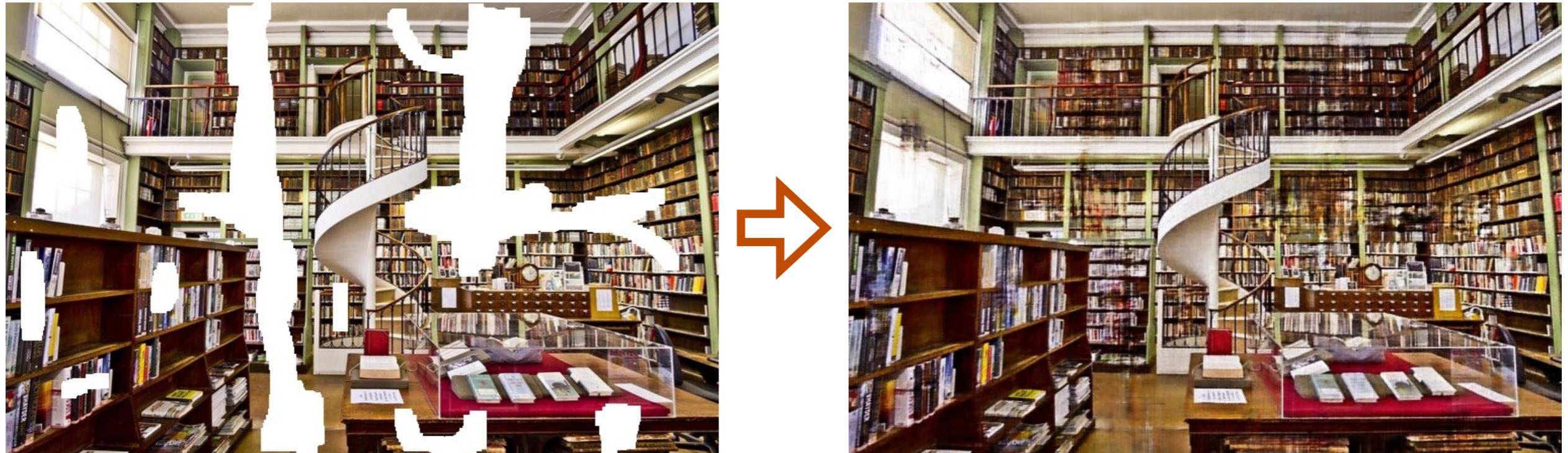


*Machine learning that incorporates
first principles, best practices, and domain knowledge
from fundamental physics*

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

Deep Learning

E.g. Inpainting



increased computational power and large data sets

[Ulyanov, Vedaldi, Lempitsky, CVPR 2018]

Deep Learning meets Deep Thinking

E.g. *Inpainting*

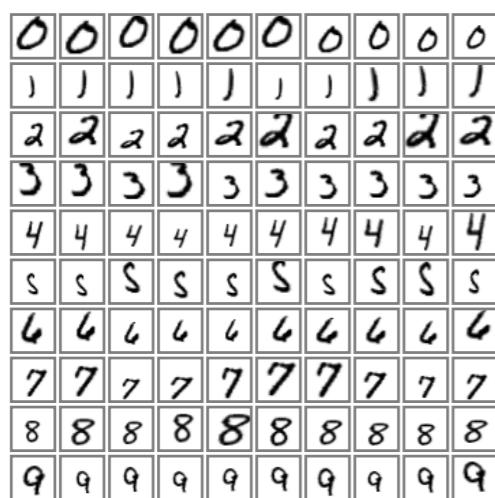
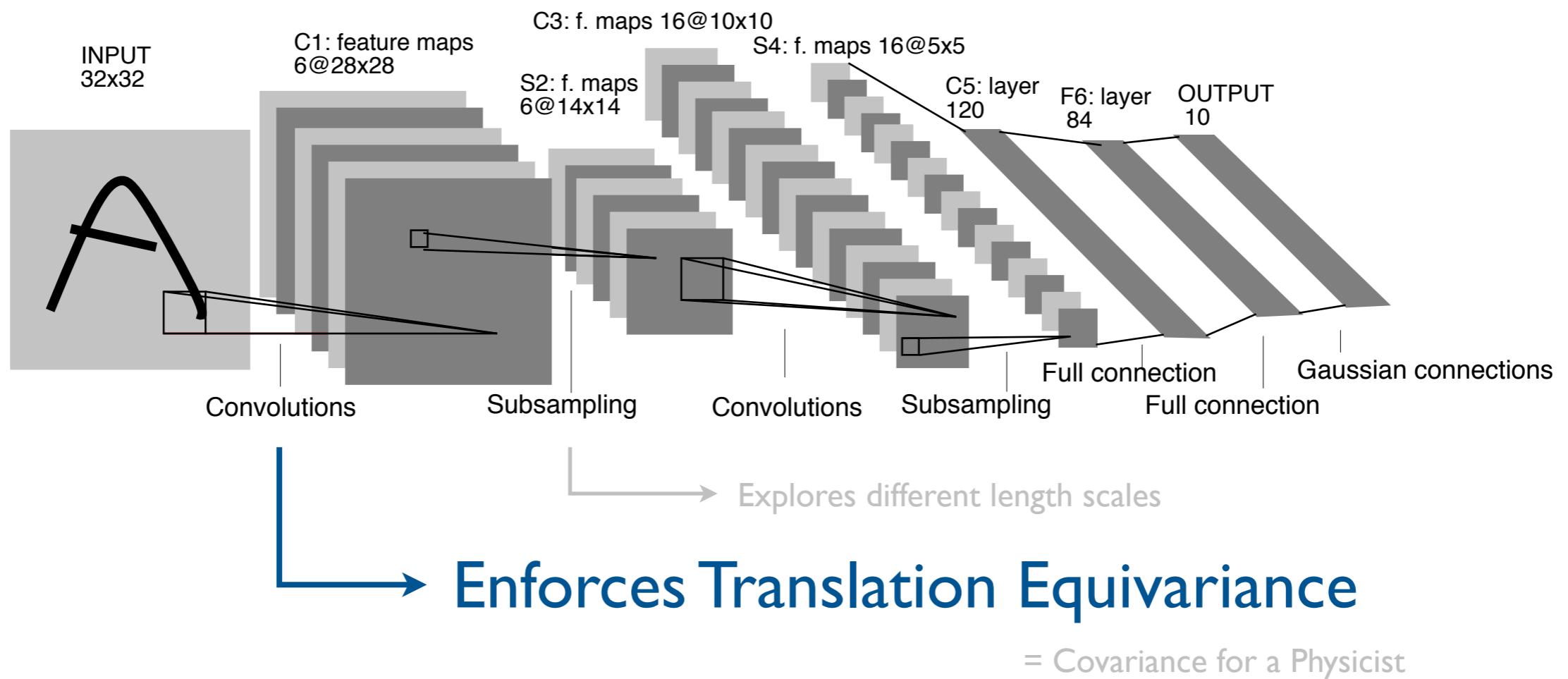


Using randomly initialized neural network (!)

Progress made by understanding the structure of problems
(not just increased computational power and large data sets)

[Ulyanov, Vedaldi, Lempitsky, CVPR 2018]

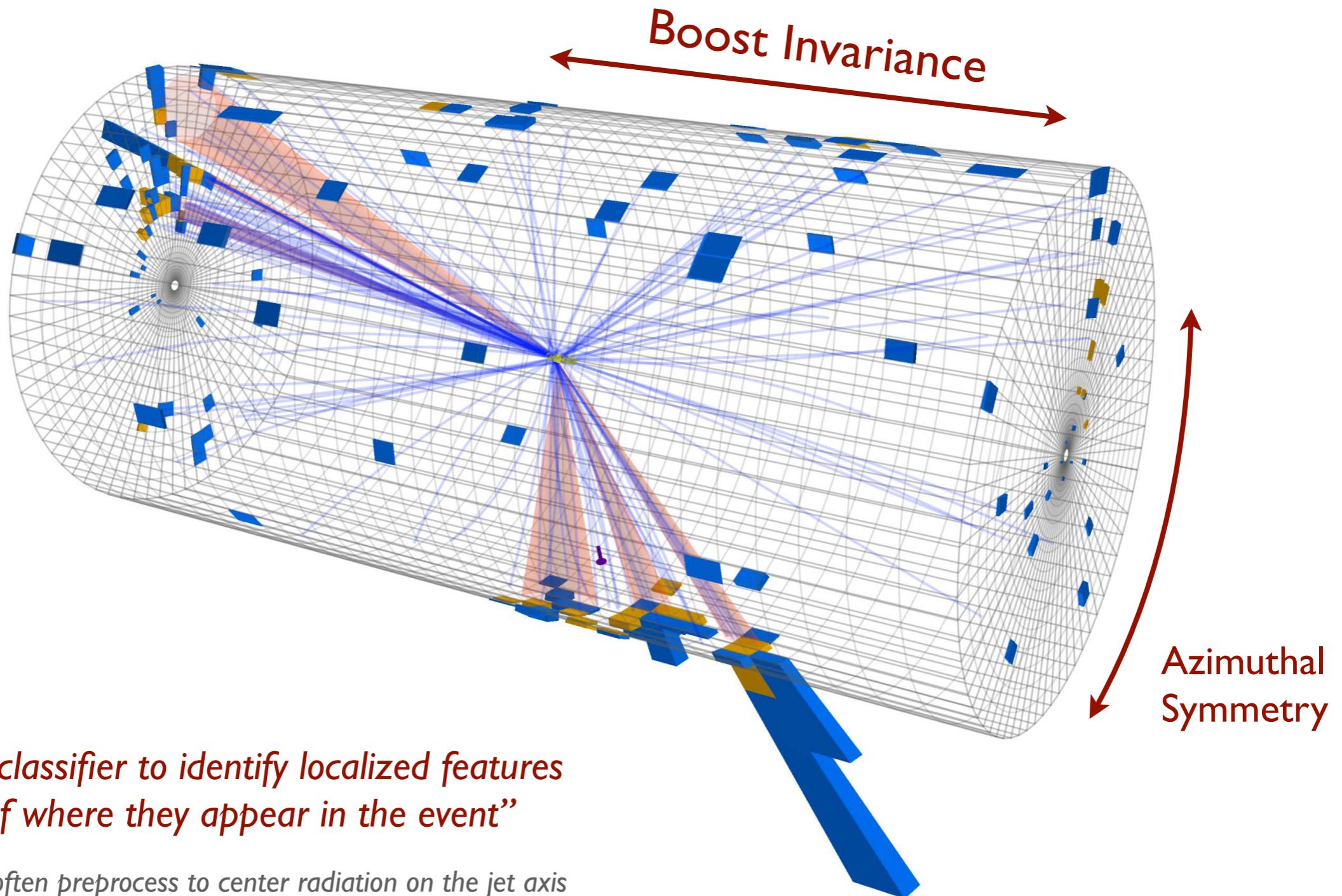
Symmetries of Convolutional NNs



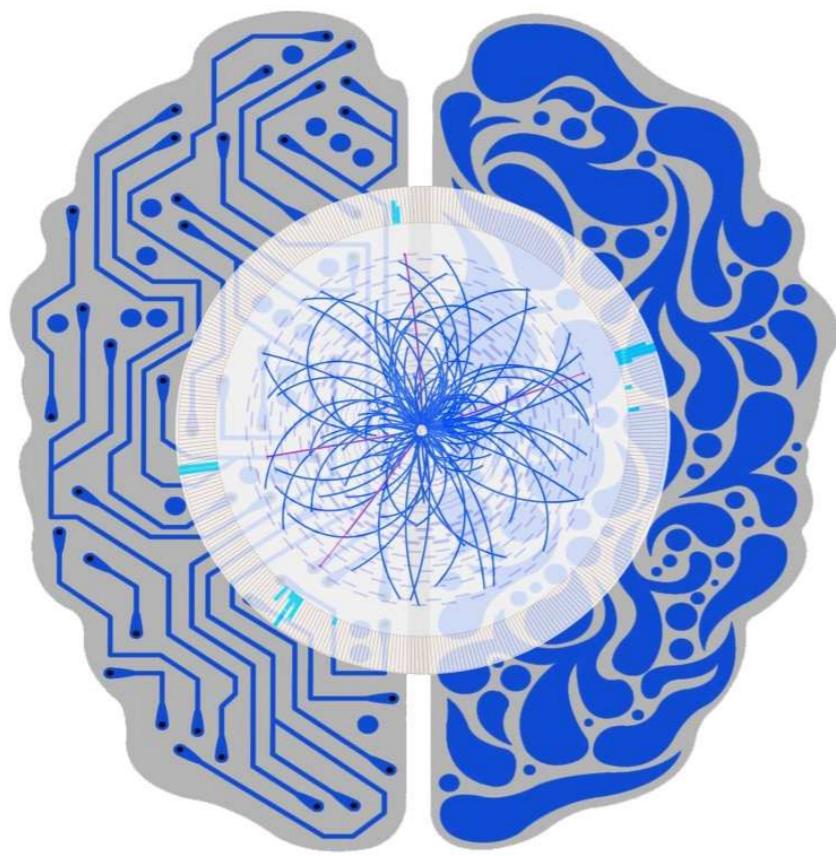
“I want my classifier to identify localized features regardless of where they appear in the image”

[image from LeCun, Bottou, Bengio, Haffner, 1998]

Symmetries of Collision Events



[image from CMS, 2015]



*Can we encode known
structures of particle
physics data into a neural
network architecture?*

AI²: Ab Initio Artificial Intelligence

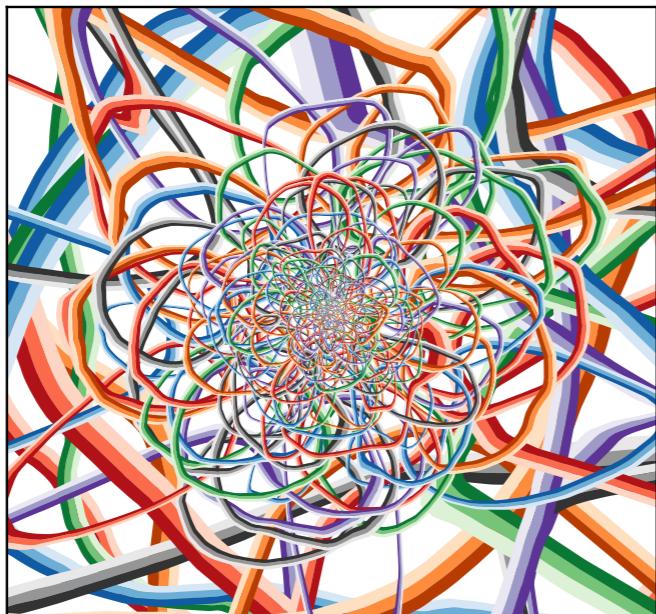


Convolutional Neural Networks \leftrightarrow Translational Equivariance

\Rightarrow Momentum Conservation

Energy Flow Networks \leftrightarrow

Identical Particles (QM)
Infrared/Collinear Safety (QFT)



[Komiske, Metodiev, JDT, JHEP 2019]

$$\begin{matrix} \text{AI} \\ \times \text{ AI} \\ = \text{AI}^2 \end{matrix}$$

Powerful strategy to
analyze LHC collisions

Efficient neural network
for point clouds

Cross-cutting research
across disciplines



Likelihood Ratio Trick

Many particle physics 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. D’Agnolo, Wulzer, [PRD 2019](#); simulation-based inference in Cranmer, Brehmer, Louppe, [PNAS 2020](#); relation to f-divergences in Nguyen, Wainwright, Jordan, [AoS 2009](#)]

Likelihood Ratio Trick

Many particle physics 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: $f(x) = \frac{p(x)}{q(x)}$

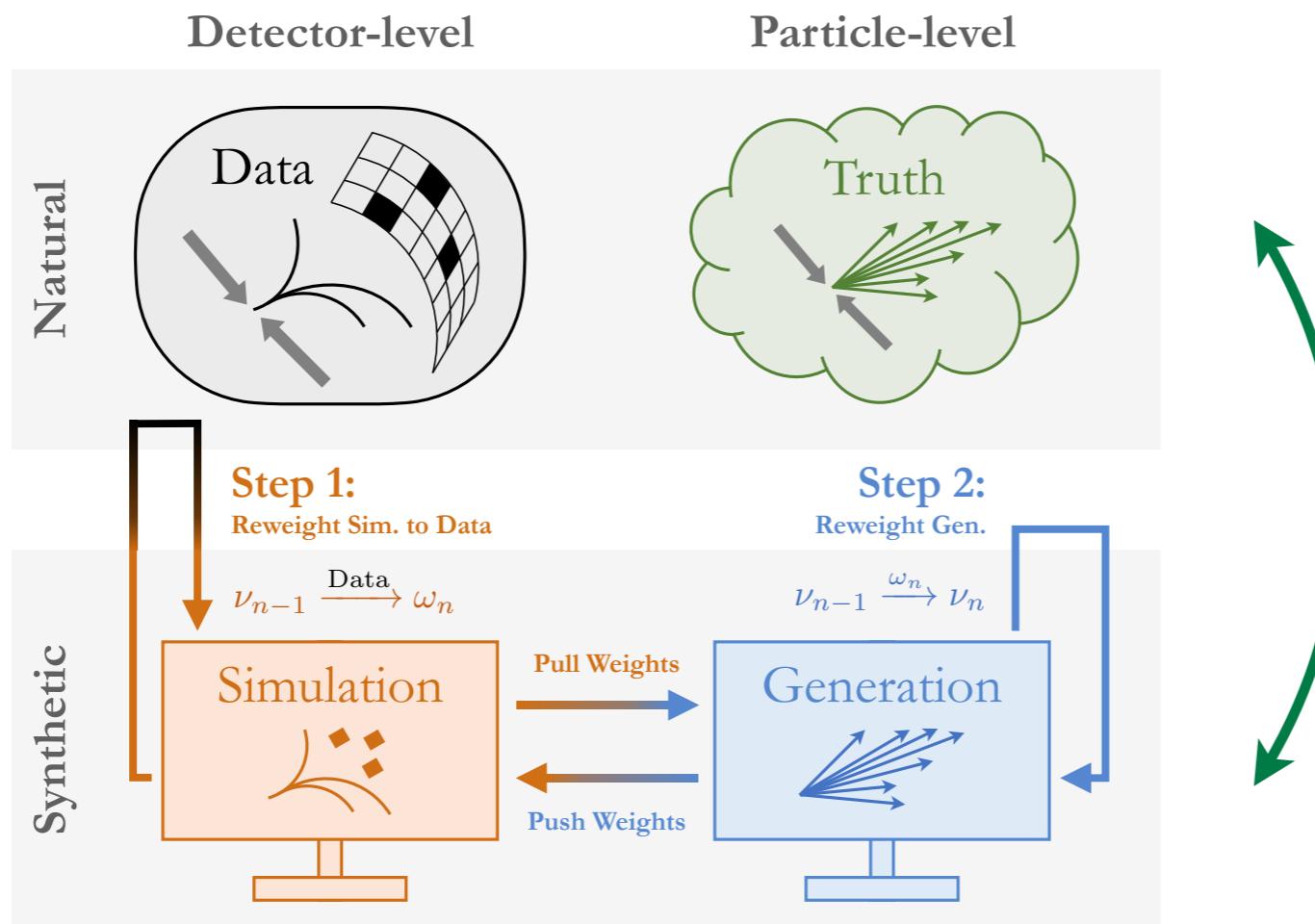
Requires shift in focus from solving problems to specifying problems

[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](#)]

Deconvolution with OmniFold



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

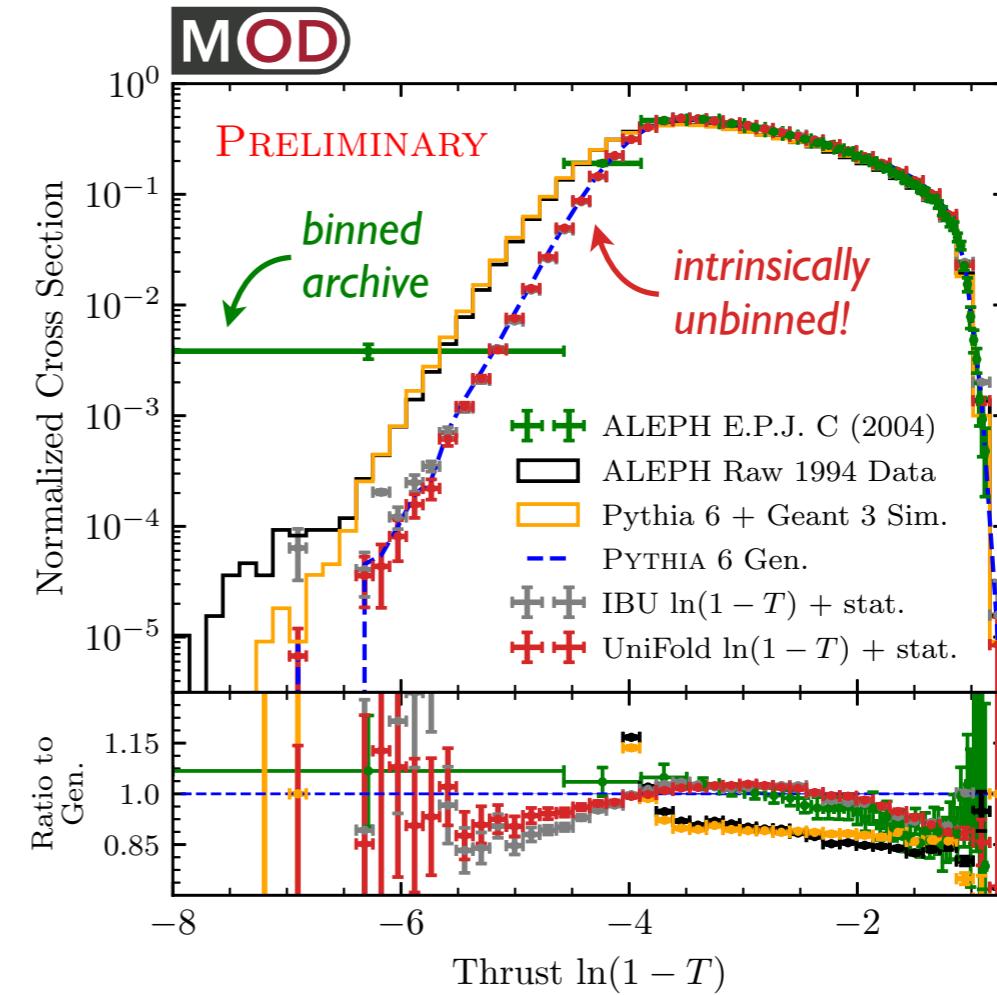
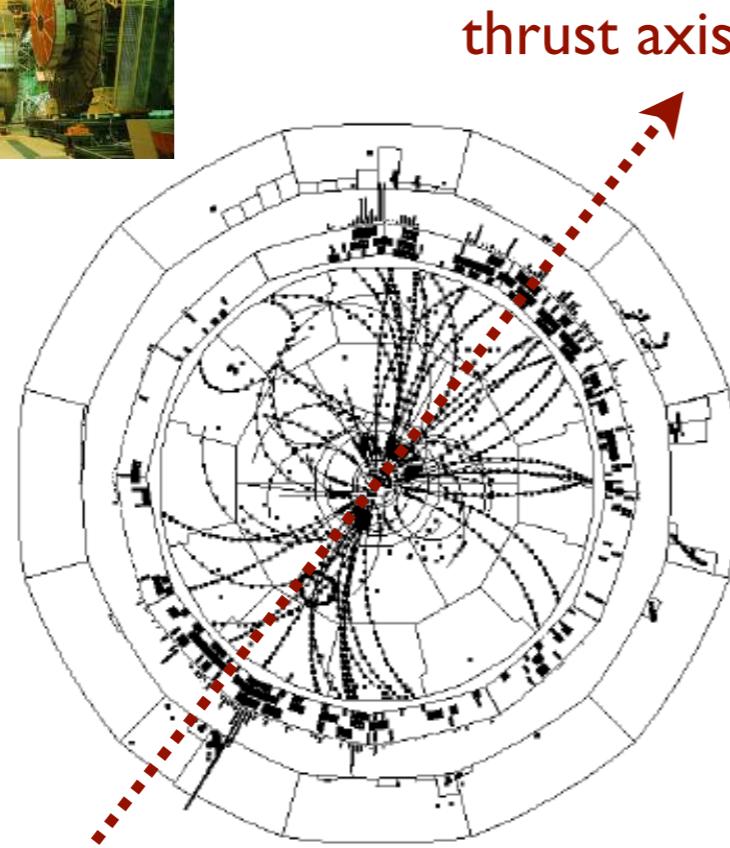


[Andreassen, Komiske, Metodiev, Nachman, JDT, [PRL 2020](#)
[see complimentary approach in Bellagente, et al., [SciPost 2020](#)]

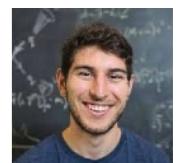


Deconvolution with OmniFold

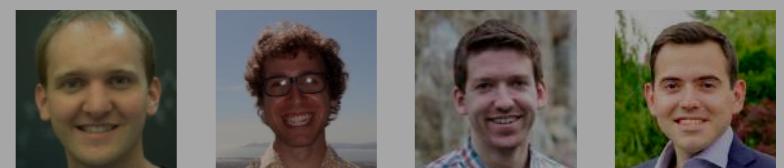
Back to the Future with ALEPH Archival Data



[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](#)]
[see complimentary approach in Bellagente, et al., [SciPost 2020](#)]



“What is the machine learning?”

For this **loss function**, an estimate of the **likelihood ratio** derived from **sampled data** and regularized by the **network architecture** and **training paradigm**

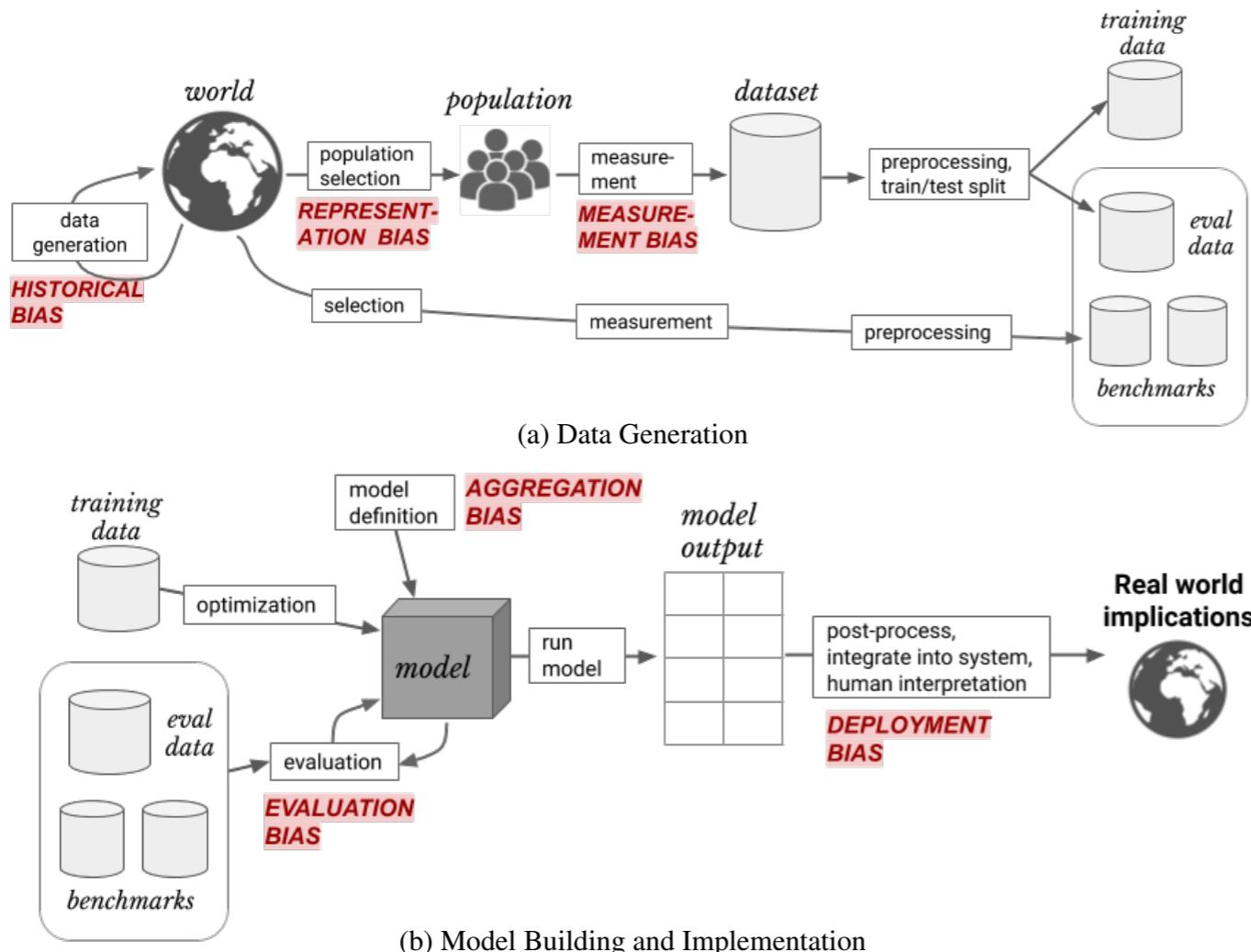
“But where’s the physics?!”

In the choice of loss function, data samples, network architecture, and training paradigm

“ ”
• • •

Many Reasons to be Wary!

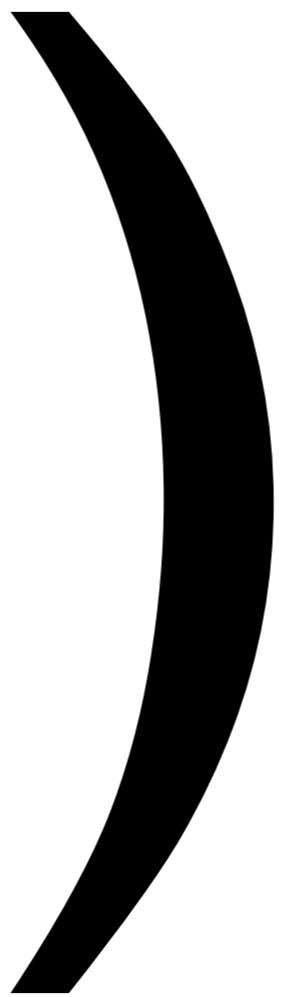
“A Framework for Understanding *Unintended Consequences* of Machine Learning”



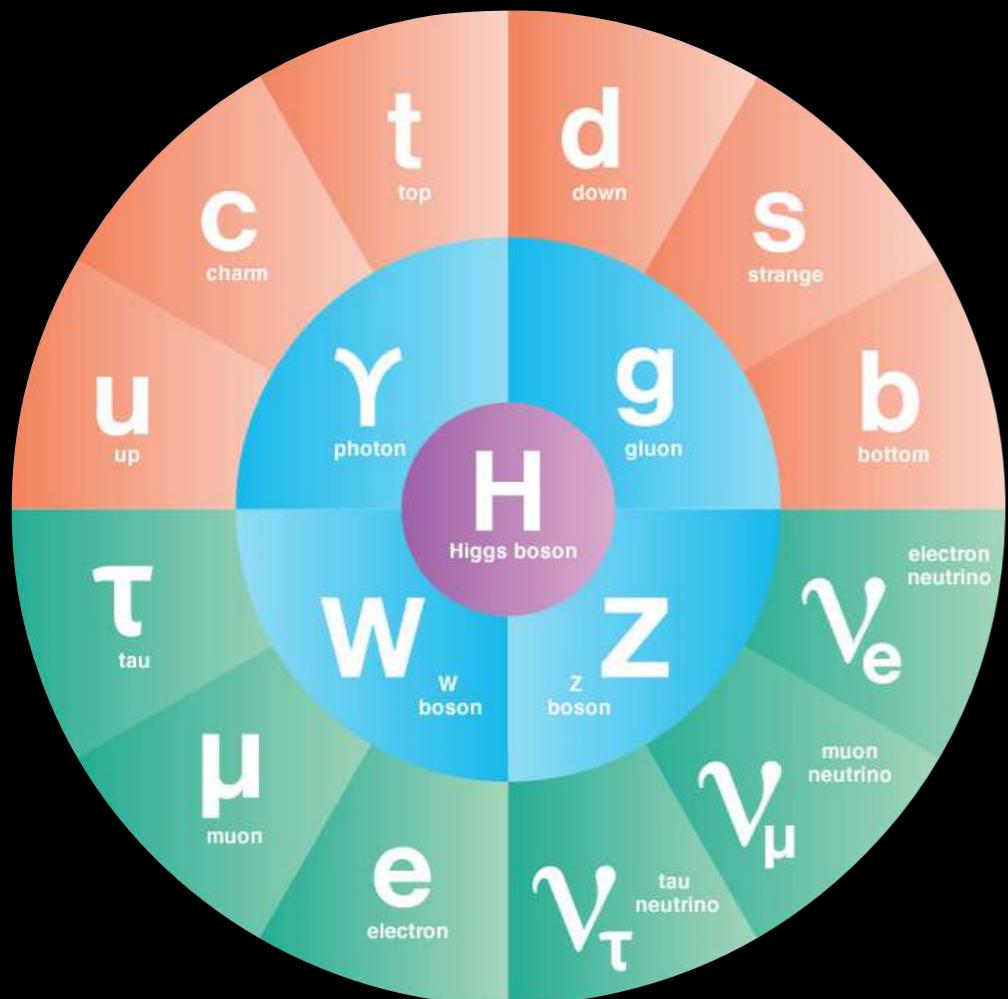
1. **Historical bias** arises when there is a misalignment between world as it is and the values or objectives to be encoded and propagated in a model. It is a normative concern with the state of the world, and exists even given perfect sampling and feature selection.
2. **Representation bias** arises while defining and sampling a development population. It occurs when the development population under-represents, and subsequently fails to generalize well, for some part of the use population.
3. **Measurement Bias** arises when choosing and measuring features and labels to use; these are often proxies for the desired quantities. The chosen set of features and labels may leave out important factors or introduce group- or input-dependent noise that leads to differential performance.
4. **Aggregation bias** arises during model construction, when distinct populations are inappropriately combined. In many applications, the population of interest is heterogeneous and a single model is unlikely to suit all subgroups.
5. **Evaluation bias** occurs during model iteration and evaluation. It can arise when the testing or external benchmark populations do not equally represent the various parts of the use population. Evaluation bias can also arise from the use of performance metrics that are not appropriate for the way in which the model will be used.
6. **Deployment Bias** occurs after model deployment, when a system is used or interpreted in inappropriate ways.

For collider physics, “bias” \approx “systematic uncertainty”

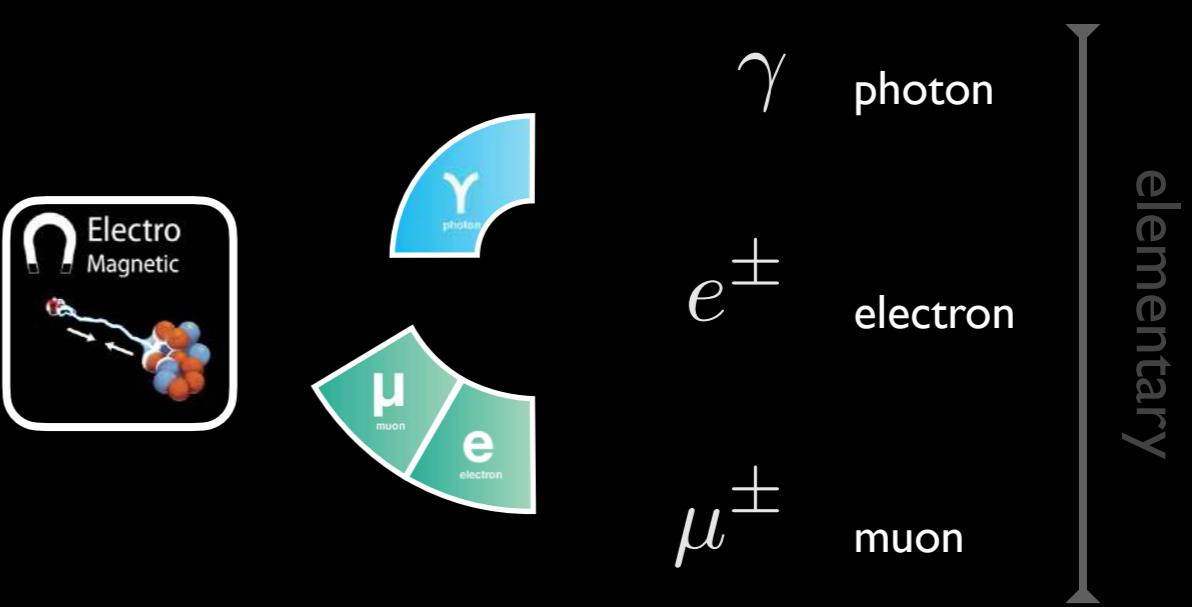
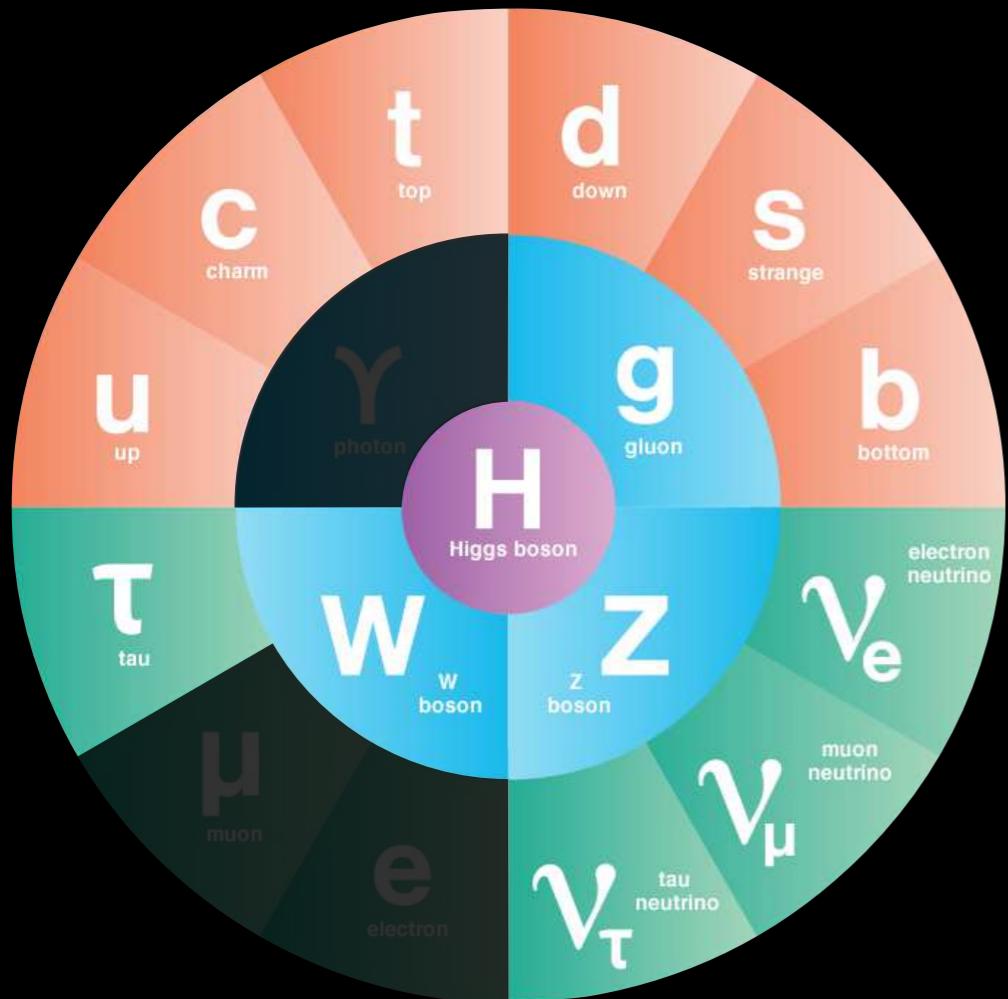
[h/t David Kaiser, [MIT SERC](#); Suresh, Guttag, [arXiv 2019](#)]



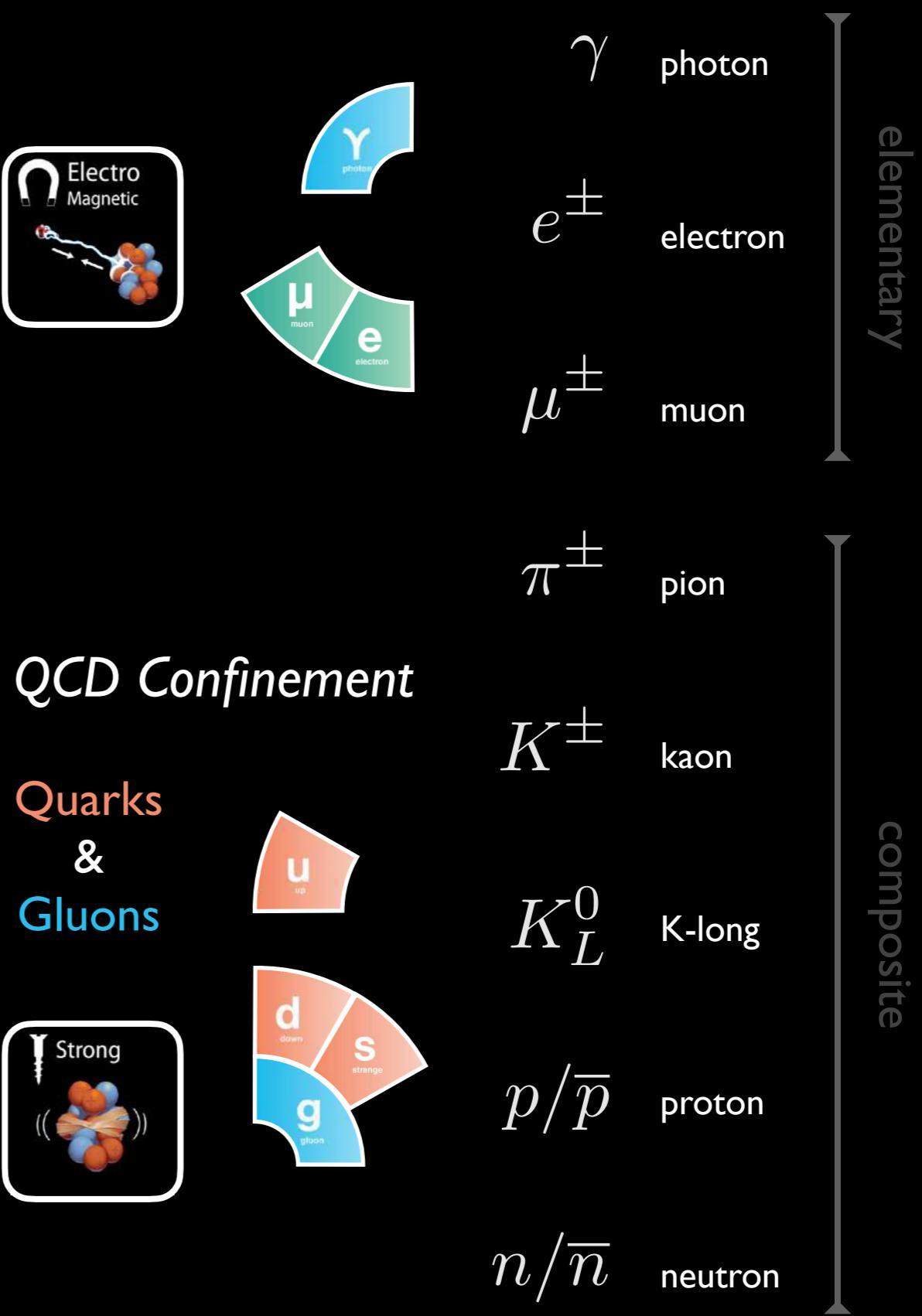
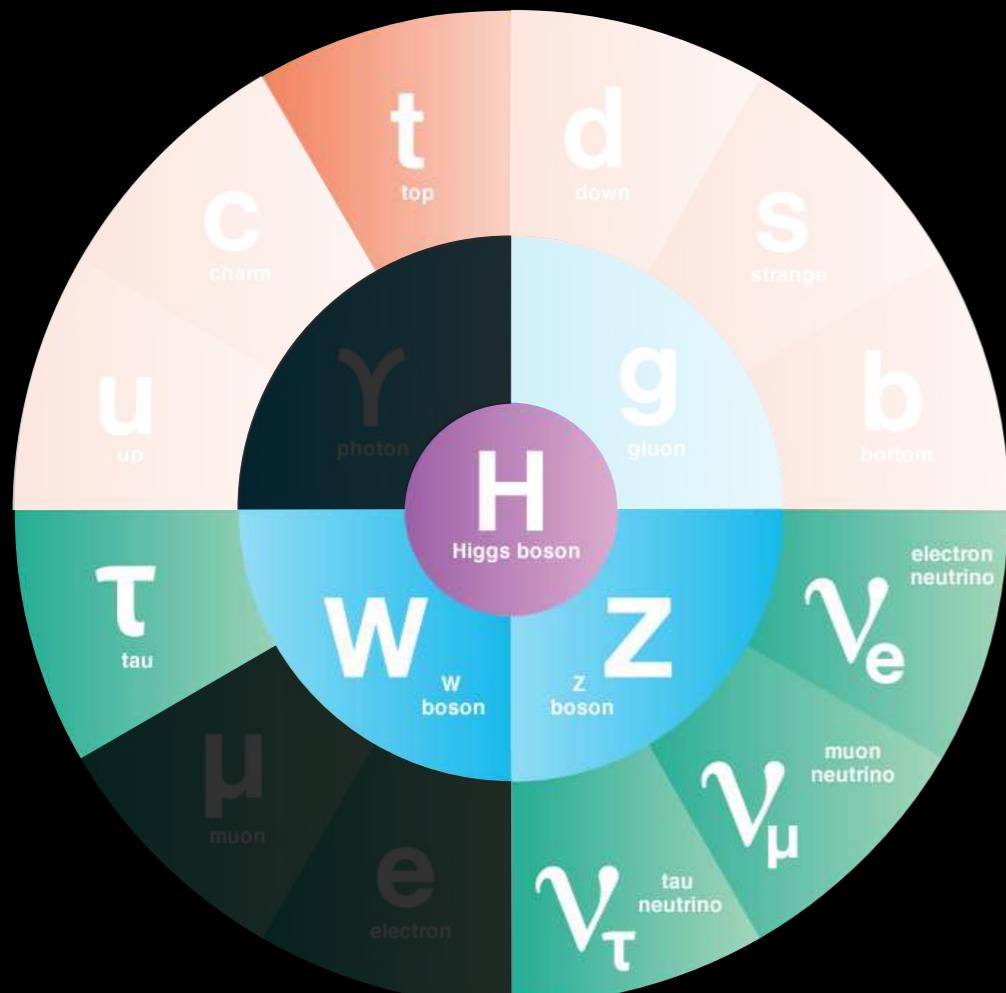
Particle Physics 101



Particle Physics 101

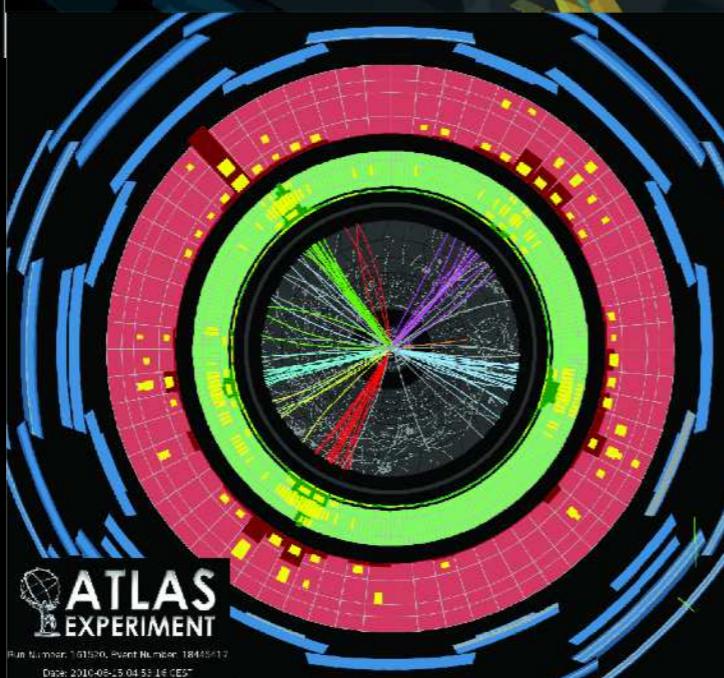
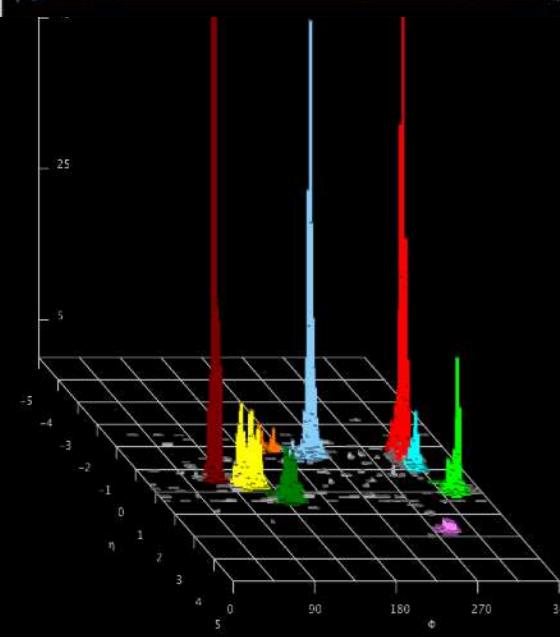
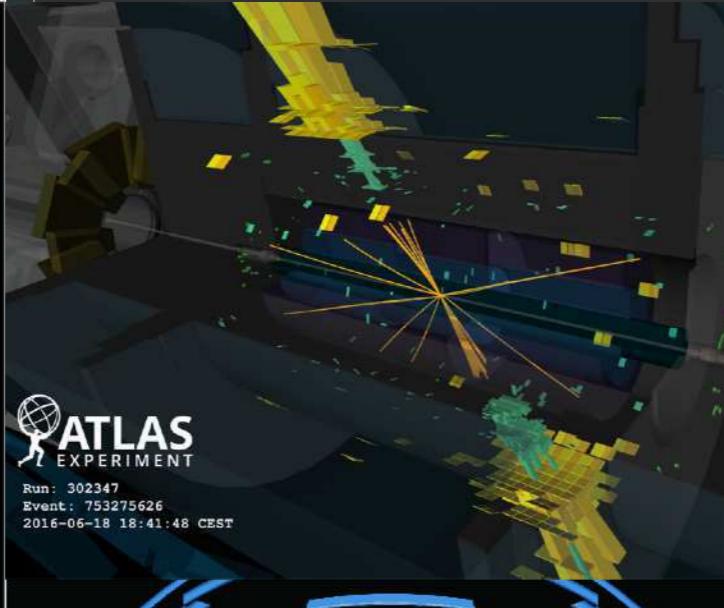
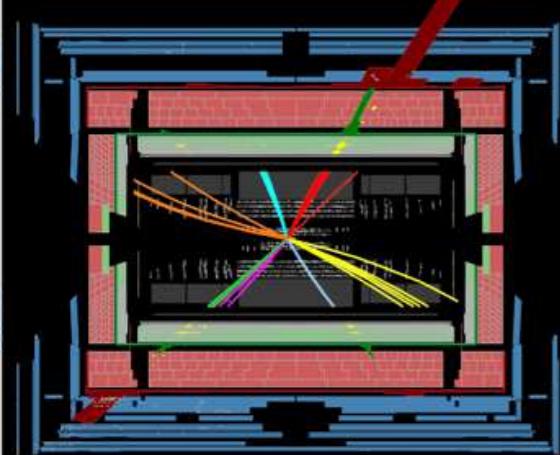


Particle Physics 101

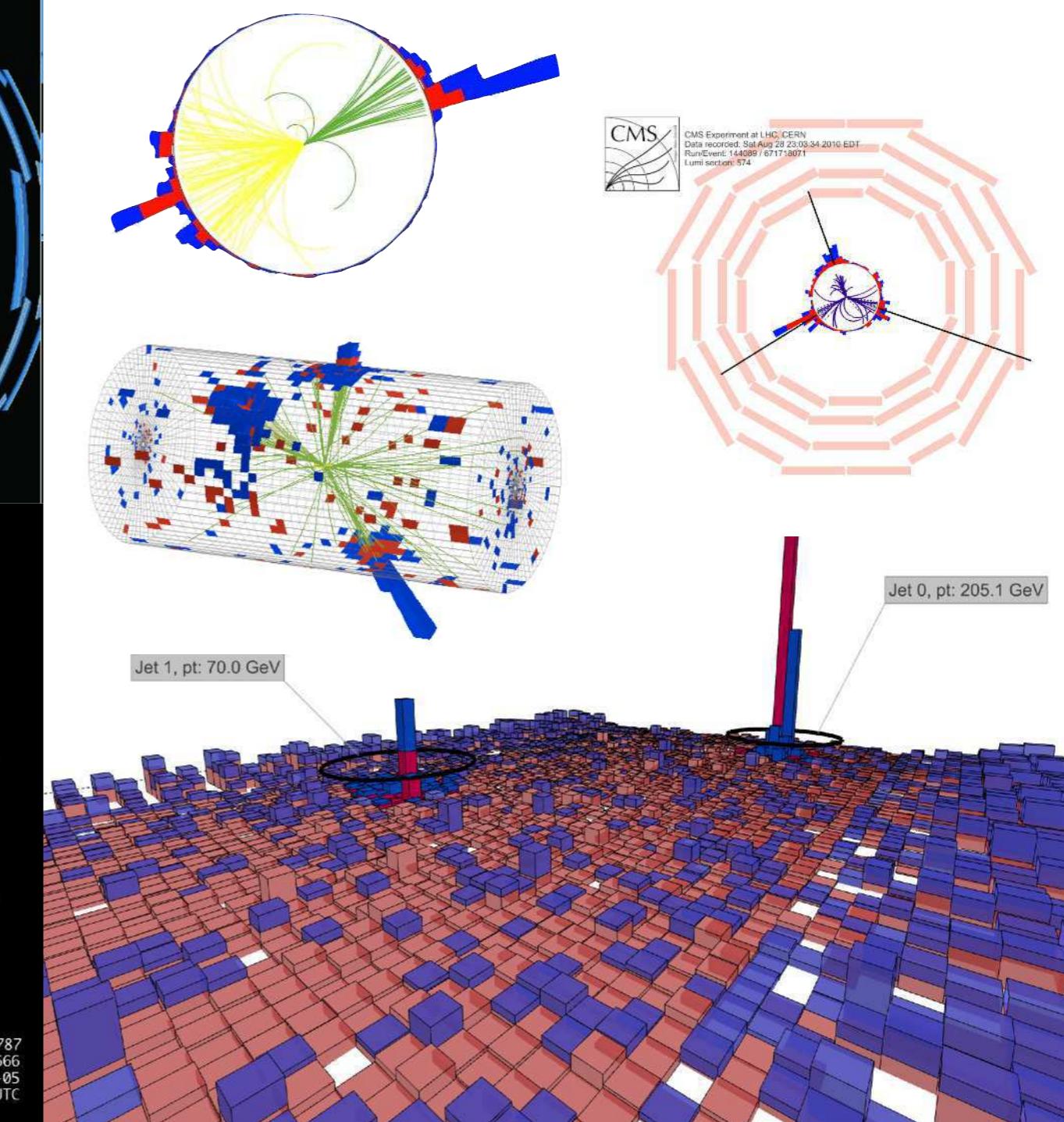
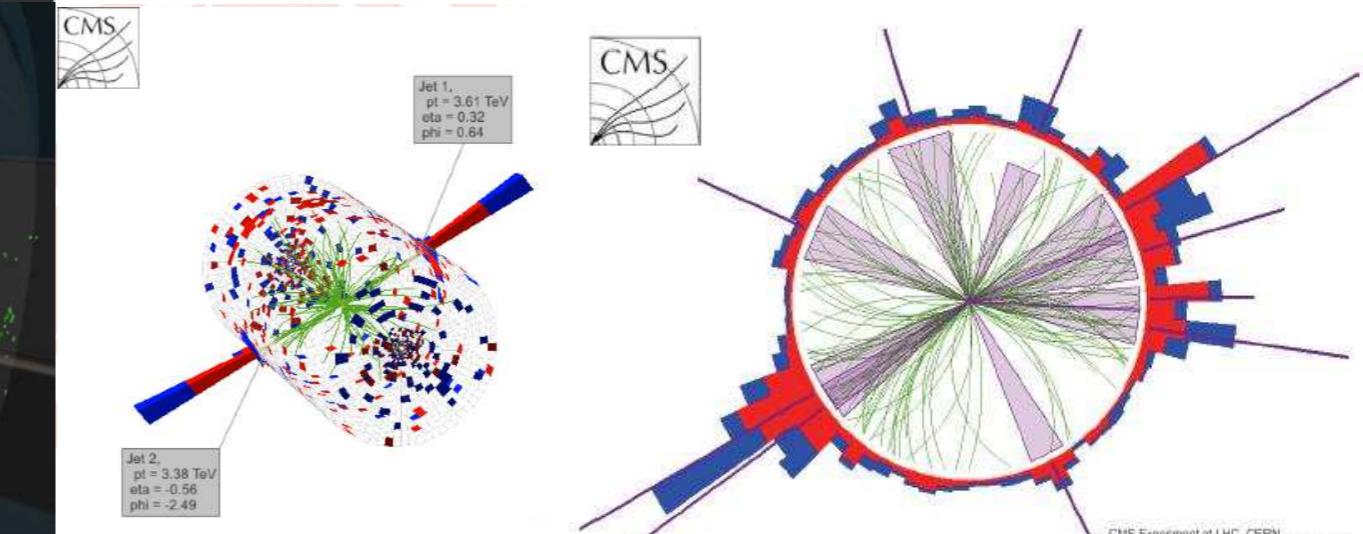
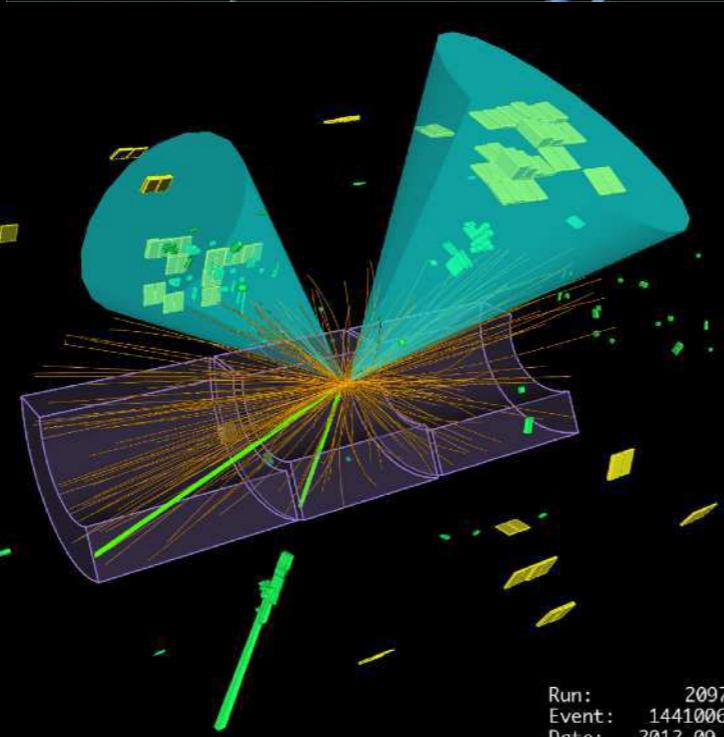


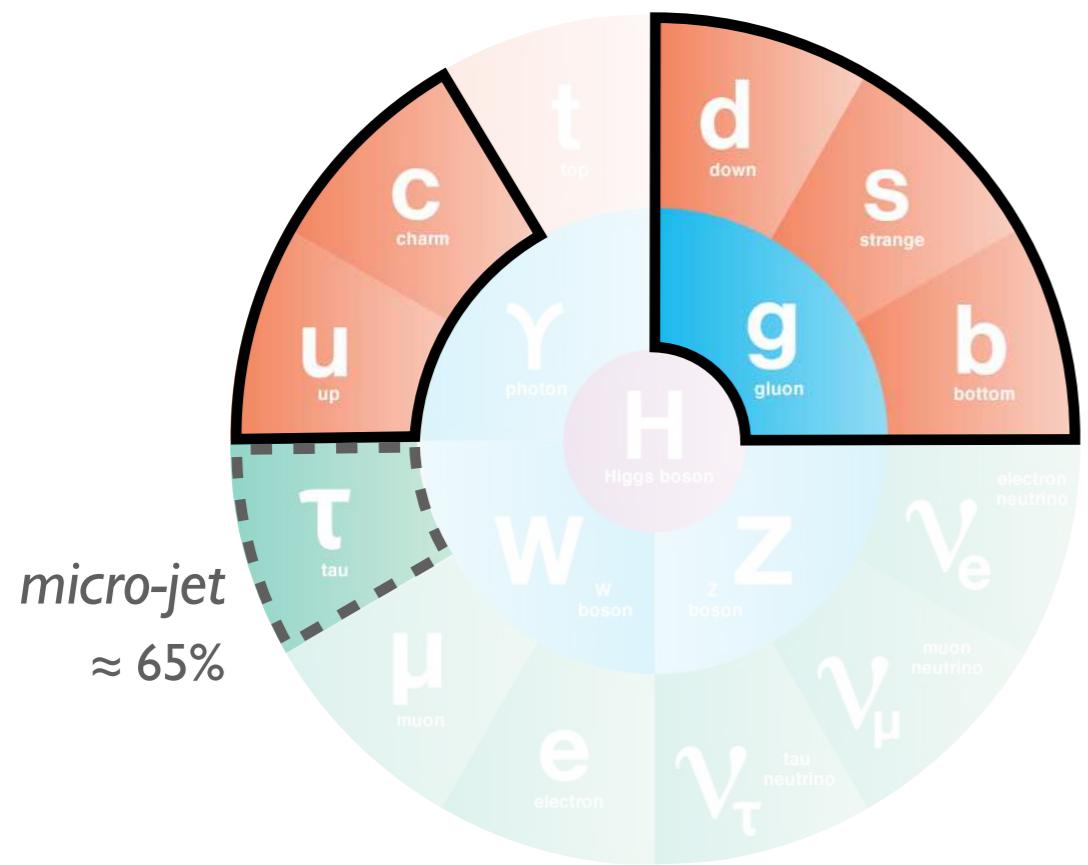
Run Number: 159224, Event Number: 3533152

Date: 2010-07-18 11:05:54 CEST



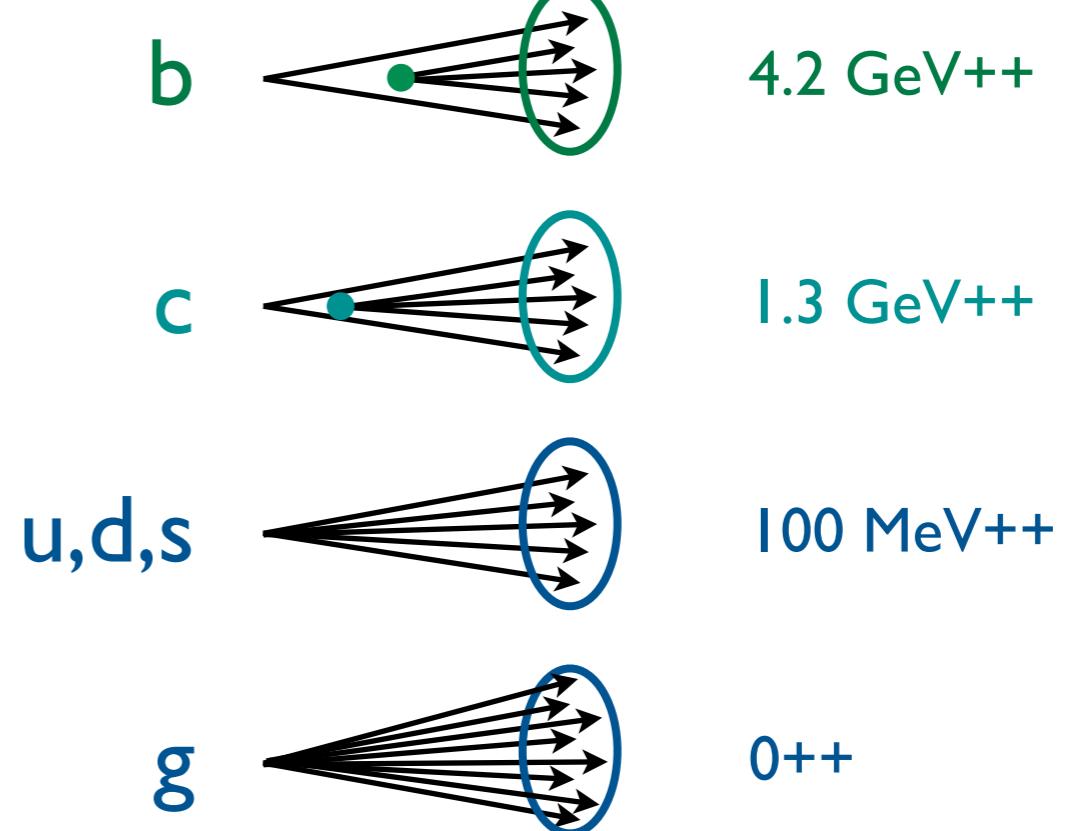
ATLAS
EXPERIMENT
<http://atlas.ch>

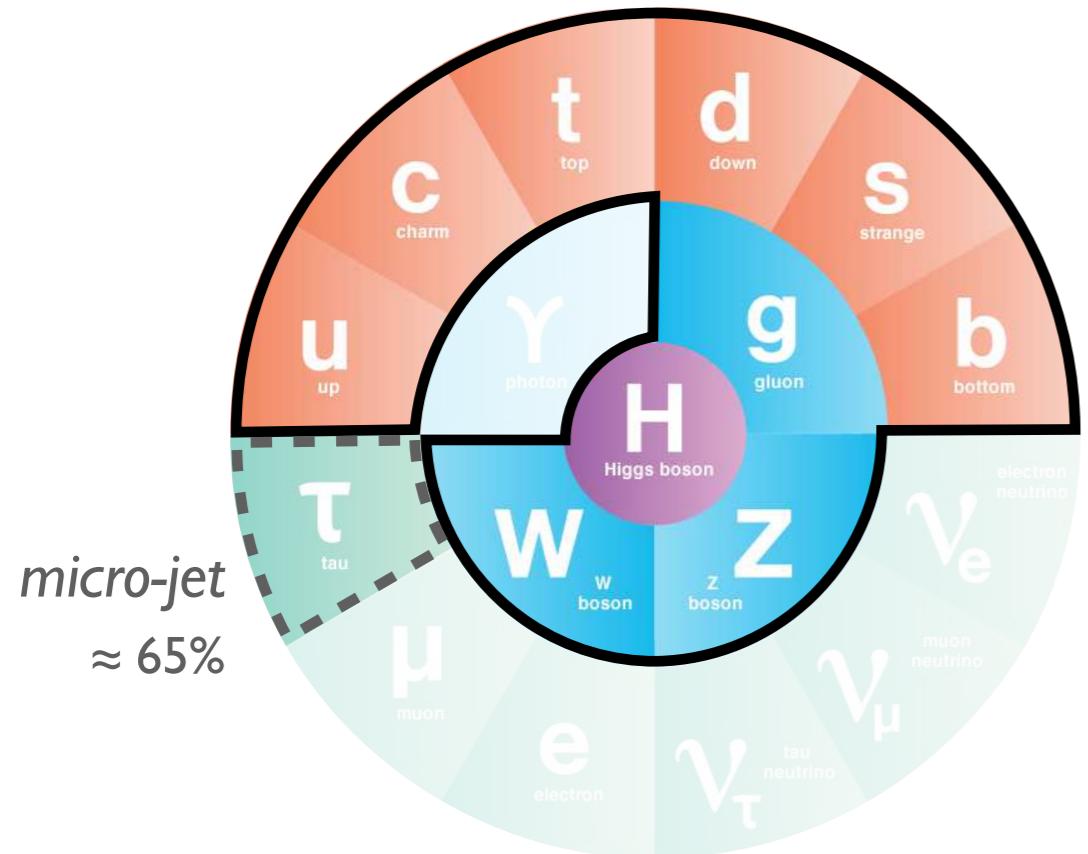




Jets from the Standard Model

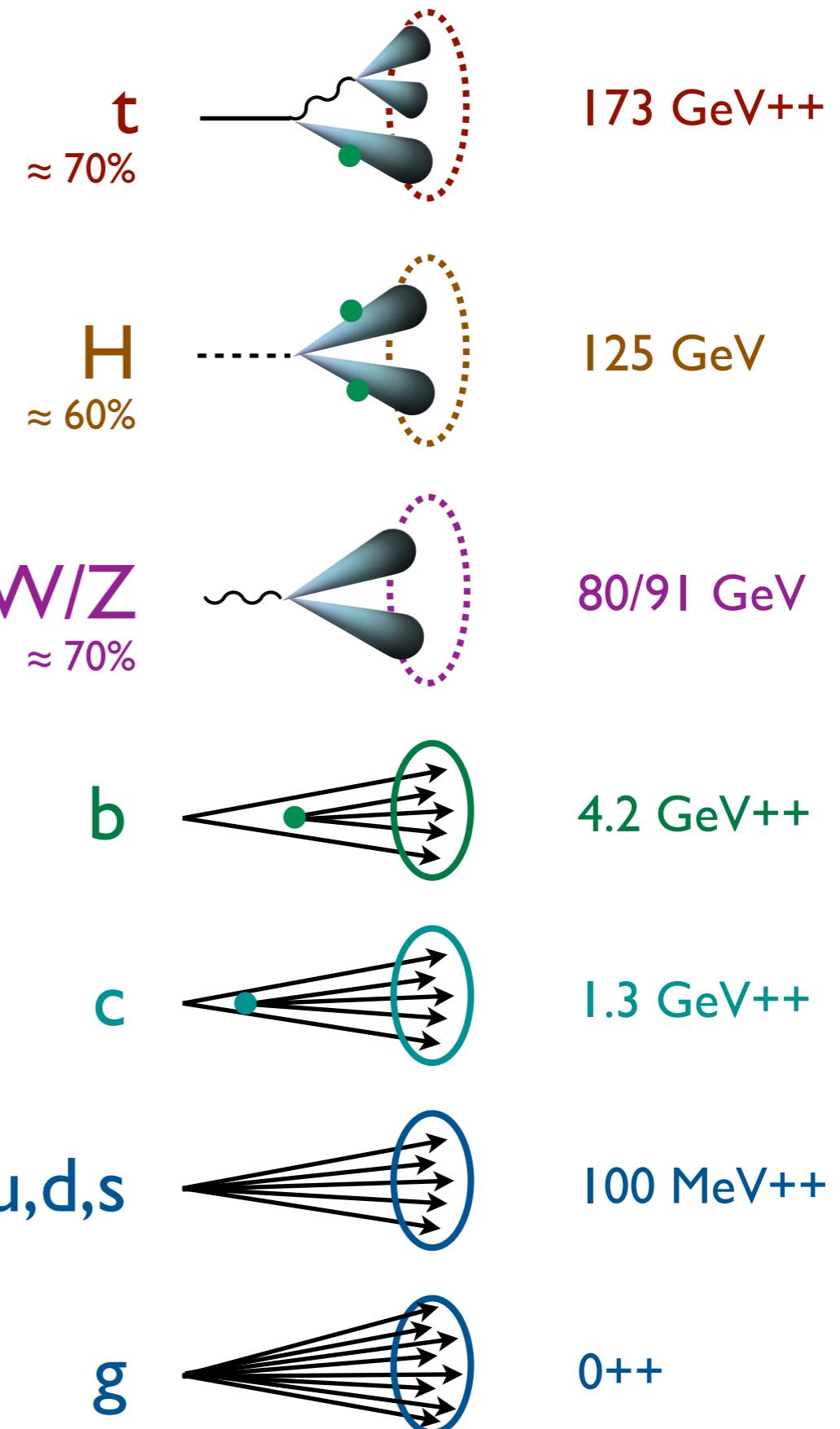
++ = Mass from QCD Radiation





Jets from the Standard Model

++ = Mass from QCD Radiation



T E H M



γ

photon



e^+

electron



μ^+

muon



π^+

pion



K^+

kaon



K_L^0

K-long



p/\bar{p}

proton



n/\bar{n}

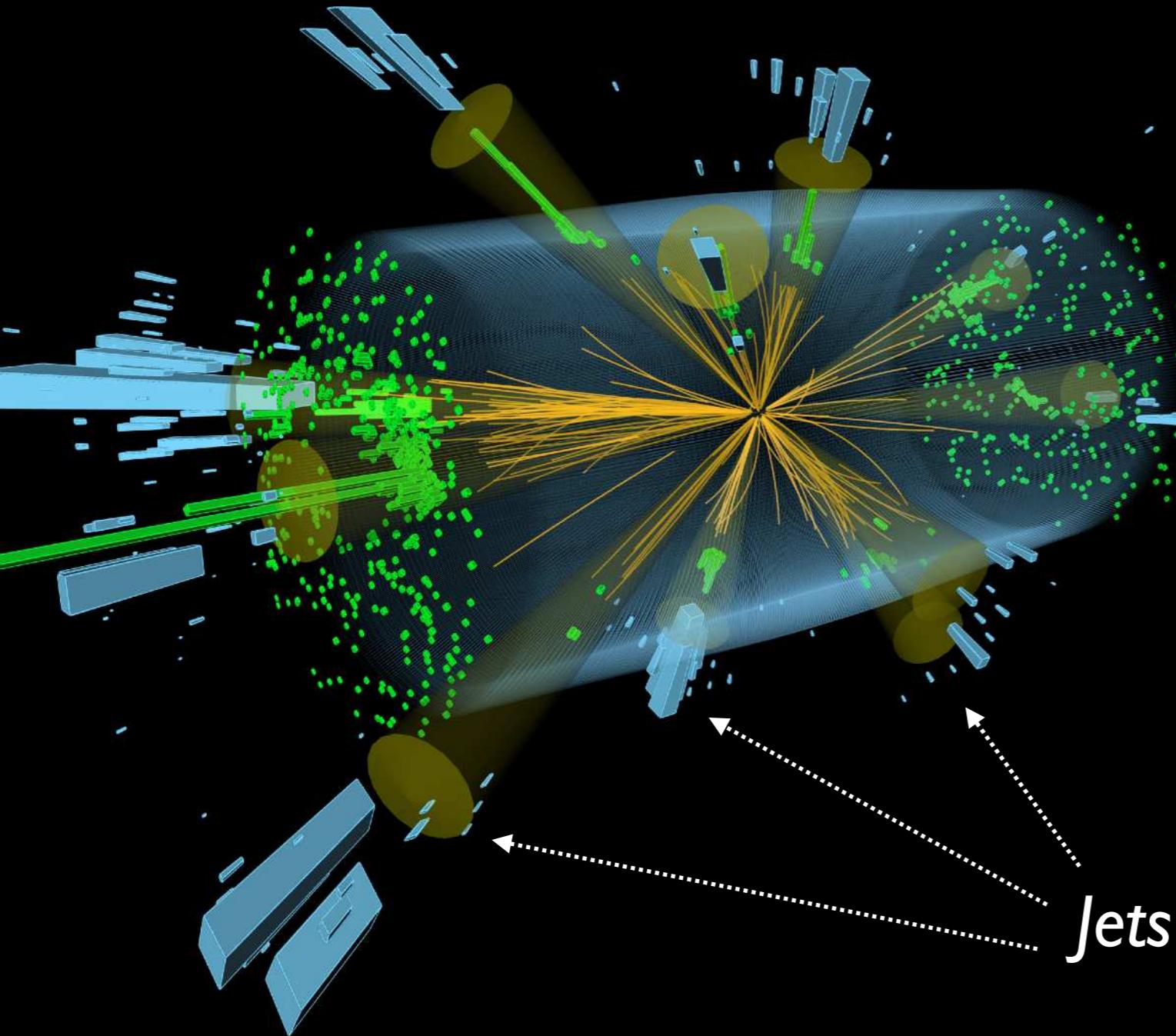
neutron

elementary

composite

Collider Event

Collection of points in (momentum) space



T E H M

 γ

photon

 e^+

electron

 μ^+

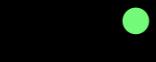
muon

 π^+

pion

 K^+

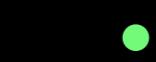
kaon

 K_L^0

K-long

 p/\bar{p}

proton

 n/\bar{n}

neutron

elementary

composite

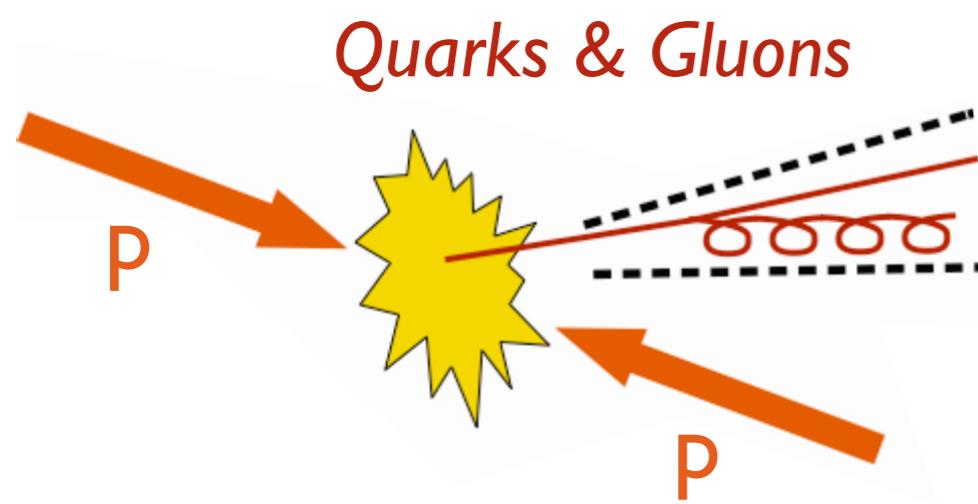
Point Cloud

Collection of points in (position) space



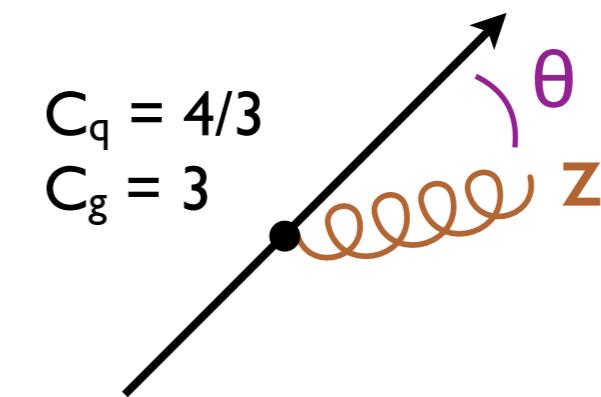
[Popular Science, 2013]

Dynamics of Jet Formation



Altarelli-Parisi Splitting

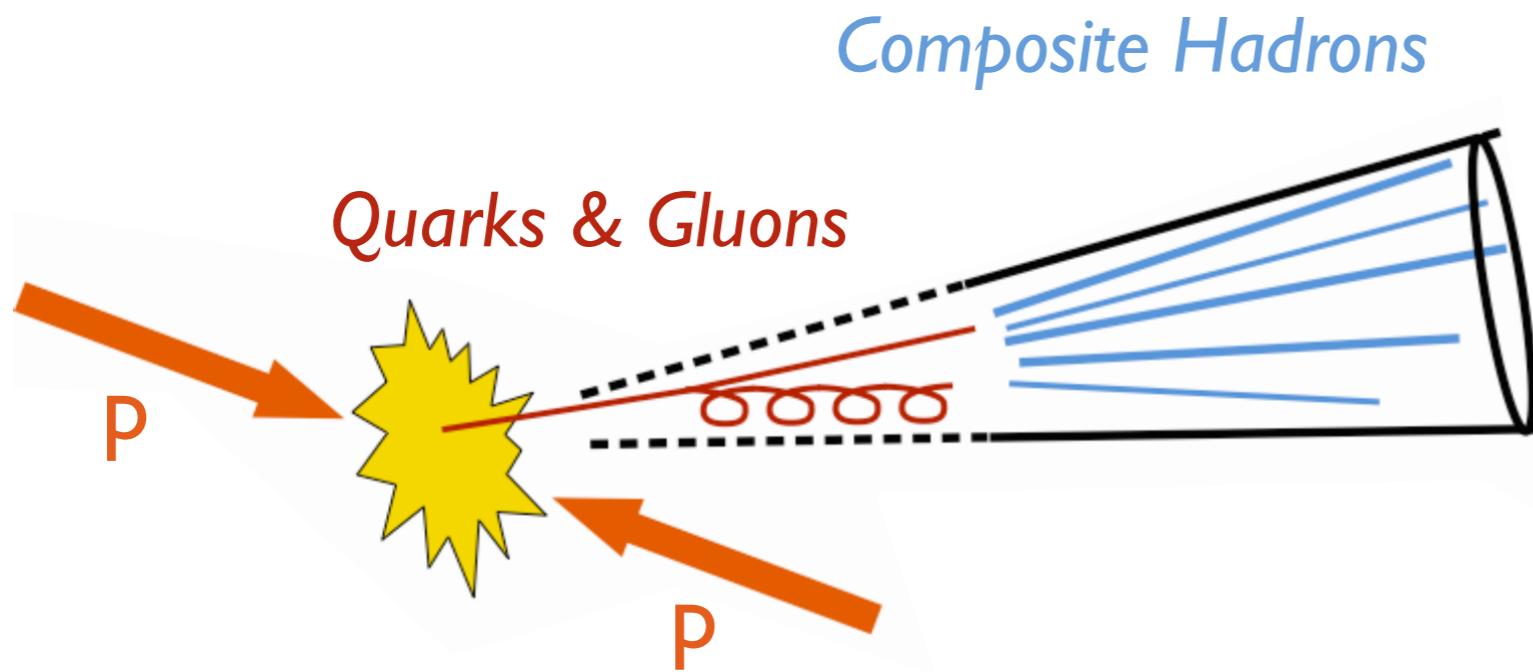
Core prediction of QCD



$$dP_{i \rightarrow ig} \simeq \frac{2\alpha_s}{\pi} C_i \frac{d\theta}{\theta} \frac{dz}{z}$$

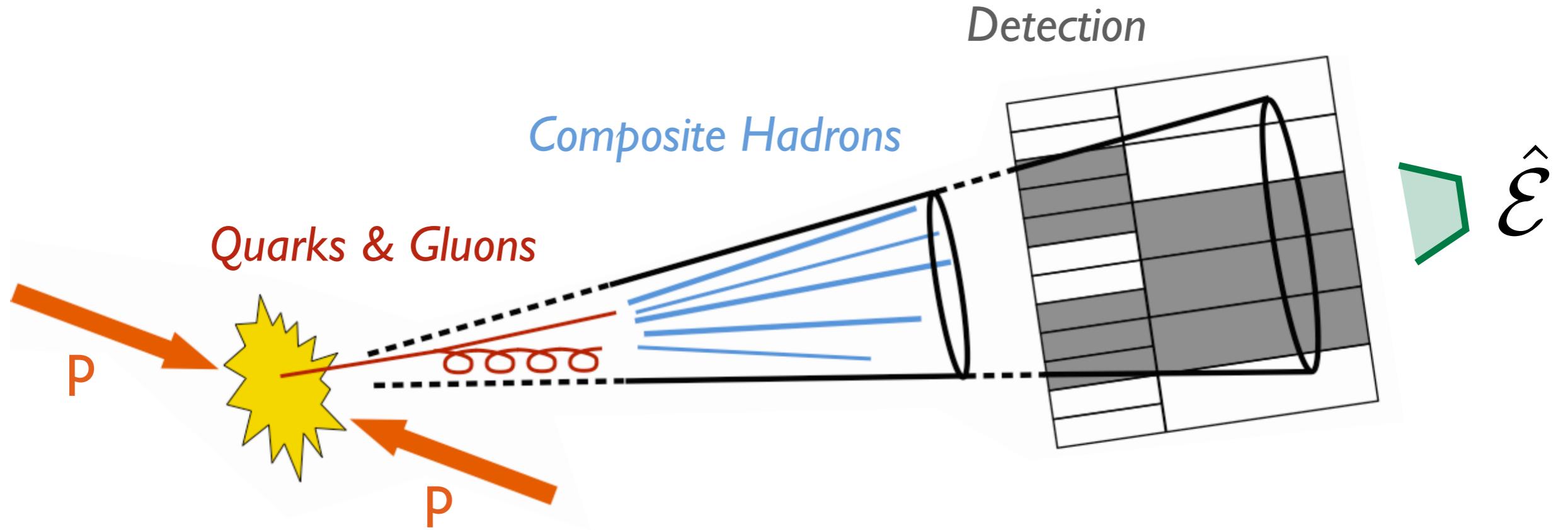
Collinear Soft

Dynamics of Jet Formation



Dynamics of Jet Formation

Theory



Energy Flow:

Robust to hadronization and detector effects

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

[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](#)]

Principles of Fundamental Physics

Robustness of Energy Flow

[Komiske, Metodiev, JDT, JHEP 2018]



Patrick Komiske



Eric Metodiev



SF



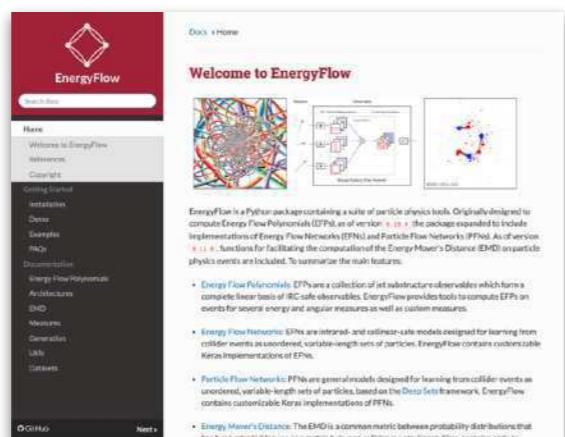
Power of Artificial Intelligence

Point Cloud Learning

Zaheer, Kottur, Ravanbakhsh, Poczos, Salakhutdinov, Smola, NIPS 2017

Energy Flow Networks

<https://energyflow.network/>
[Komiske, Metodiev, JDT, JHEP 2019]



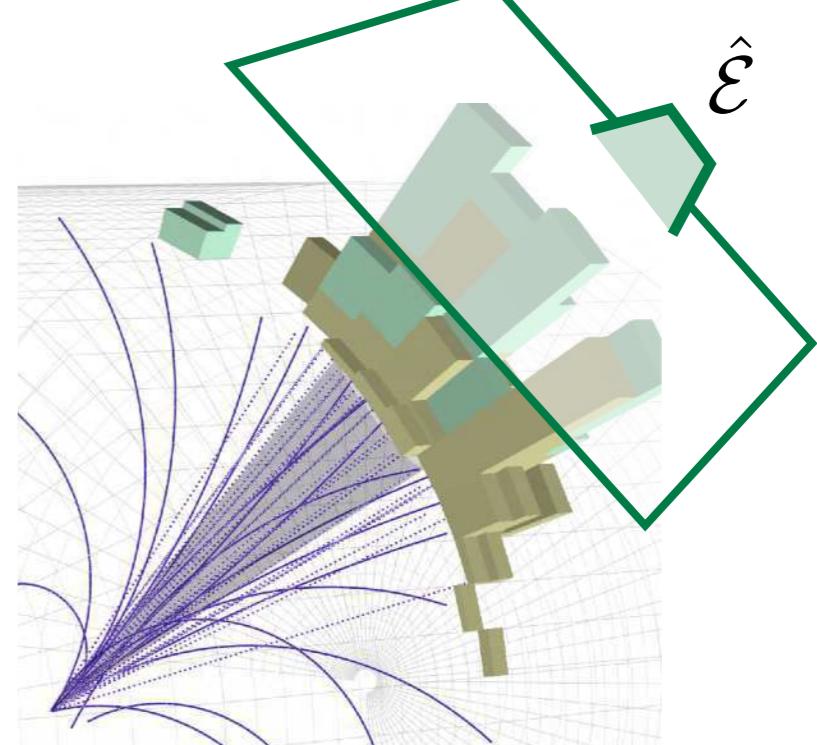
Jets as Weighted Point Clouds

- Energy-Weighted Directions

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

↑ |
Energy Direction

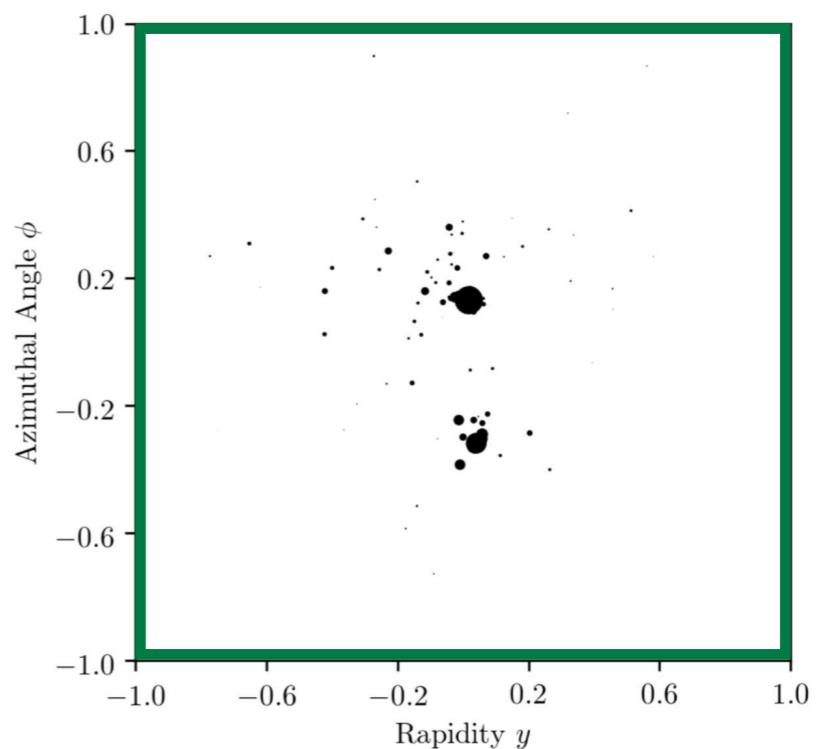
(suppressing “unsafe” charge/flavor information)



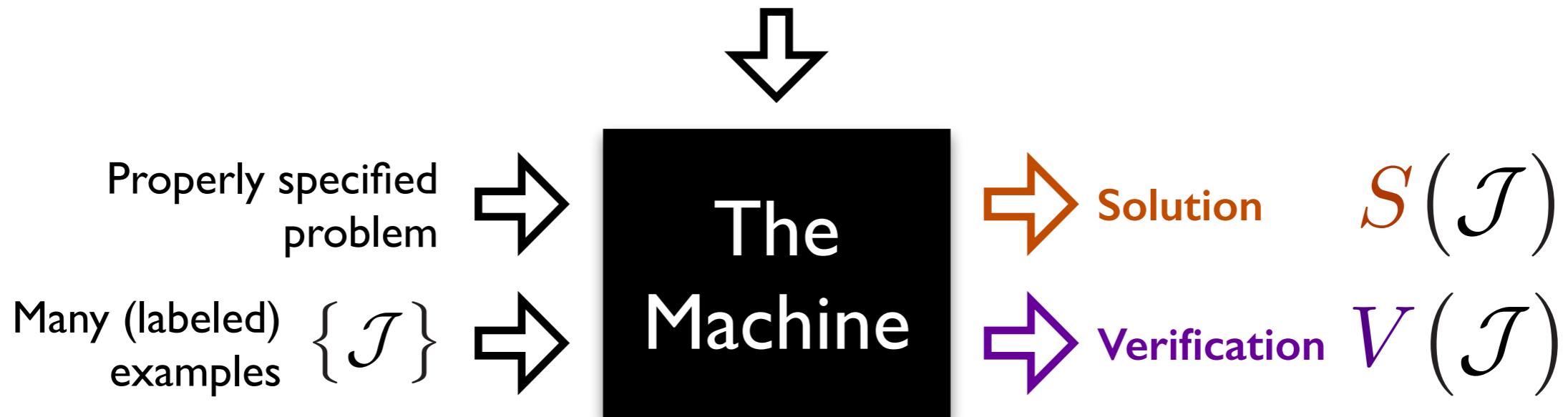
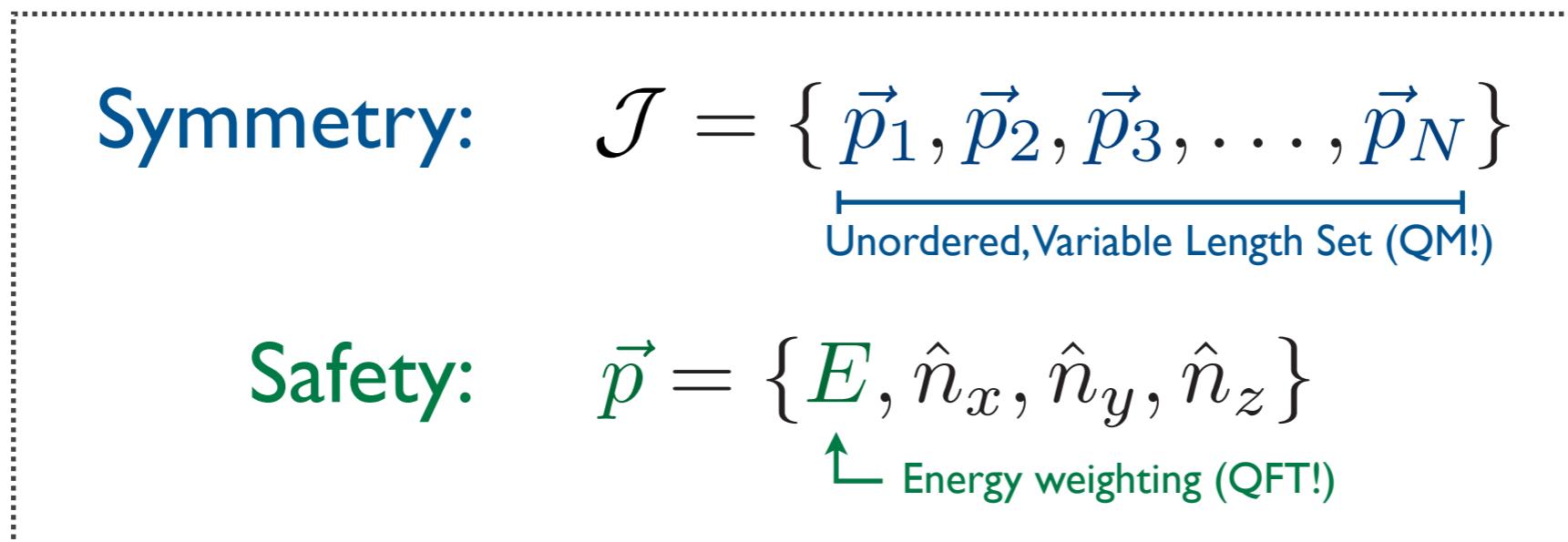
- Equivalently: Energy Density

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

↑ ↑
Energy Direction



“Thinking” Like a Physicist



*Check that answer
is physically sensible*

Energy Flow Networks

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



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

Permutation invariant \downarrow Linear weights (i.e. safe) \downarrow

Parametrized with Neural Networks

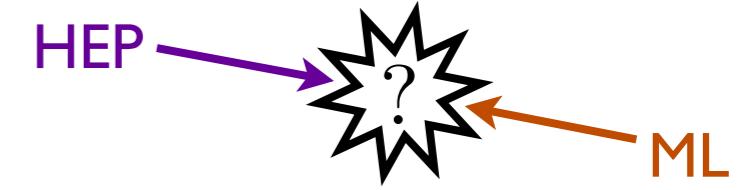
Provably describes any **safe** observable (!)*
Excellent jet classification performance

[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](#)]



Energy Flow Networks

Architecture designed around symmetries and *interpretability*

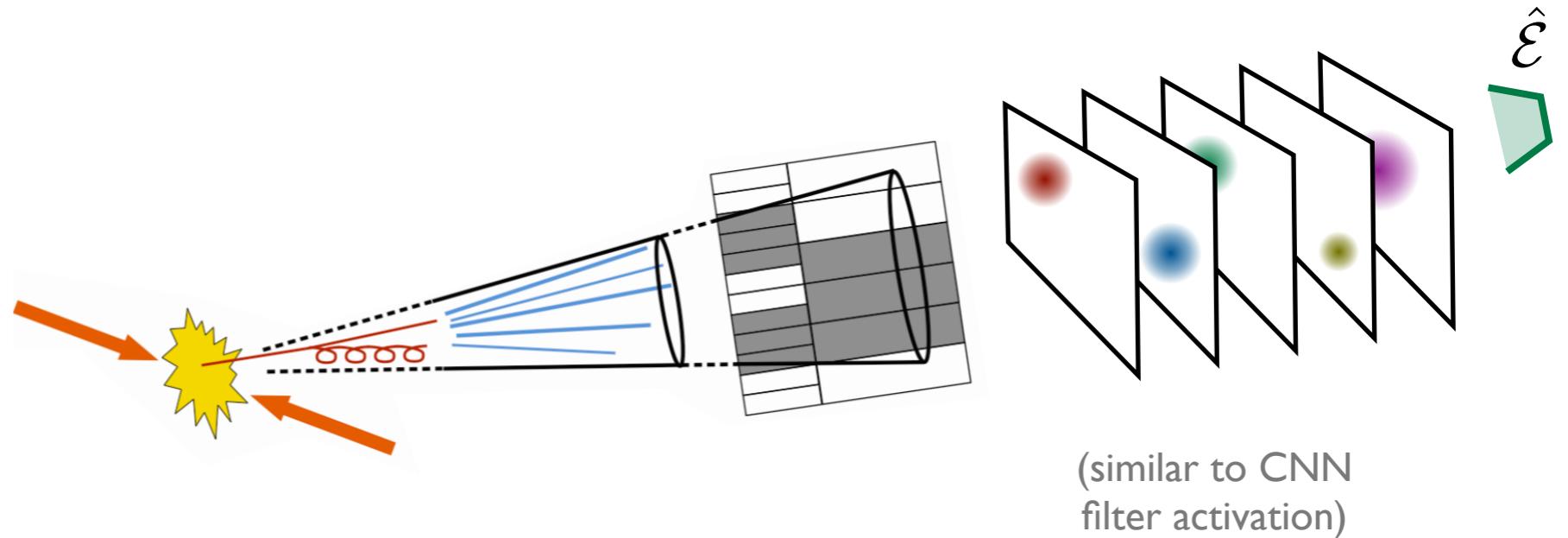


$$S(\mathcal{J}) = F(V_1, V_2, \dots, V_\ell)$$

Latent space of dim ℓ

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

Easy to plot!

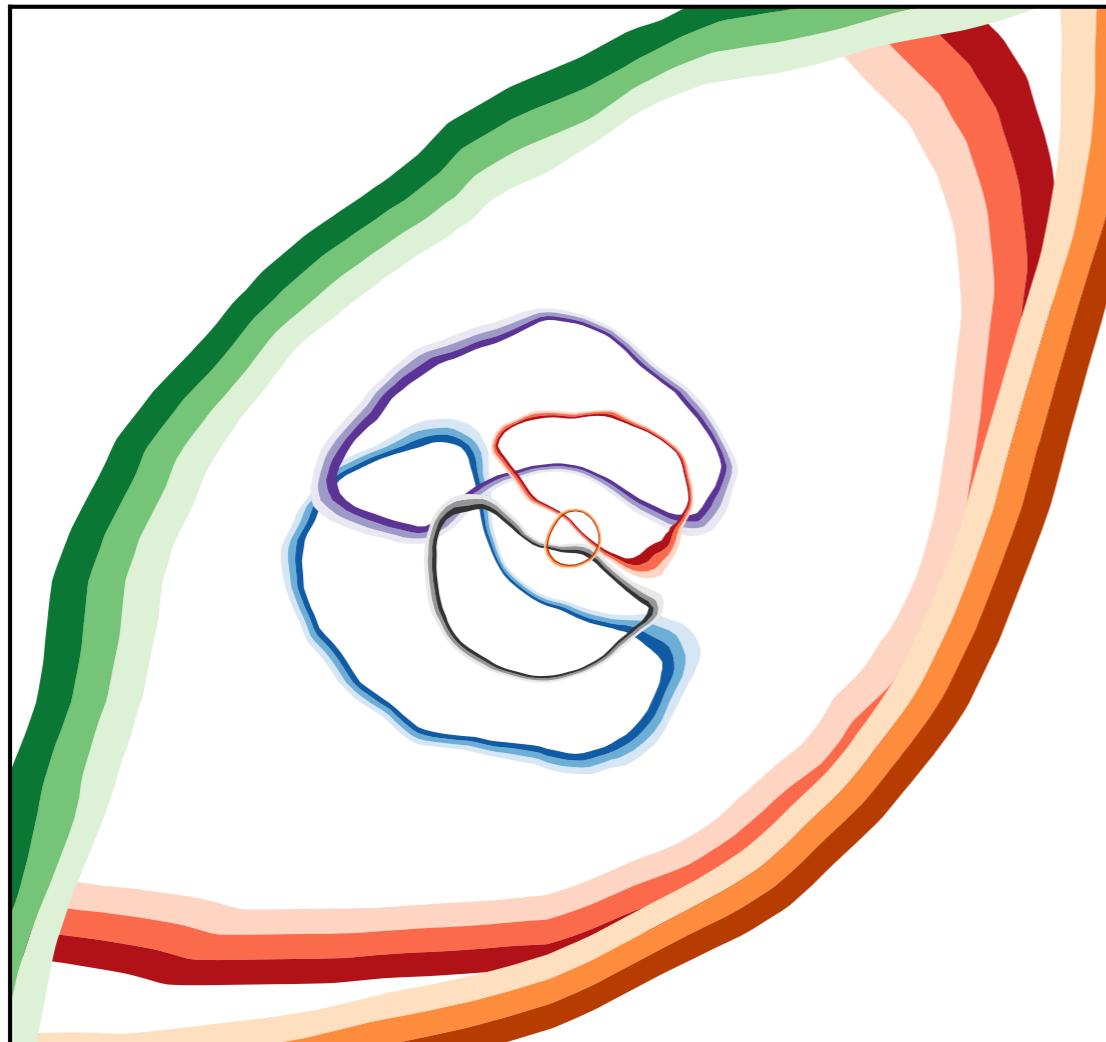


[Komiske, Metodiev, JDT, [JHEP 2019](#); see also Komiske, Metodiev, JDT, [JHEP 2018](#); code at [energyflow.network](#); special case of Zaheer, Kottur, Ravanbakhsh, Poczos, Salakhutdinov, Smola, [NIPS 2017](#)]

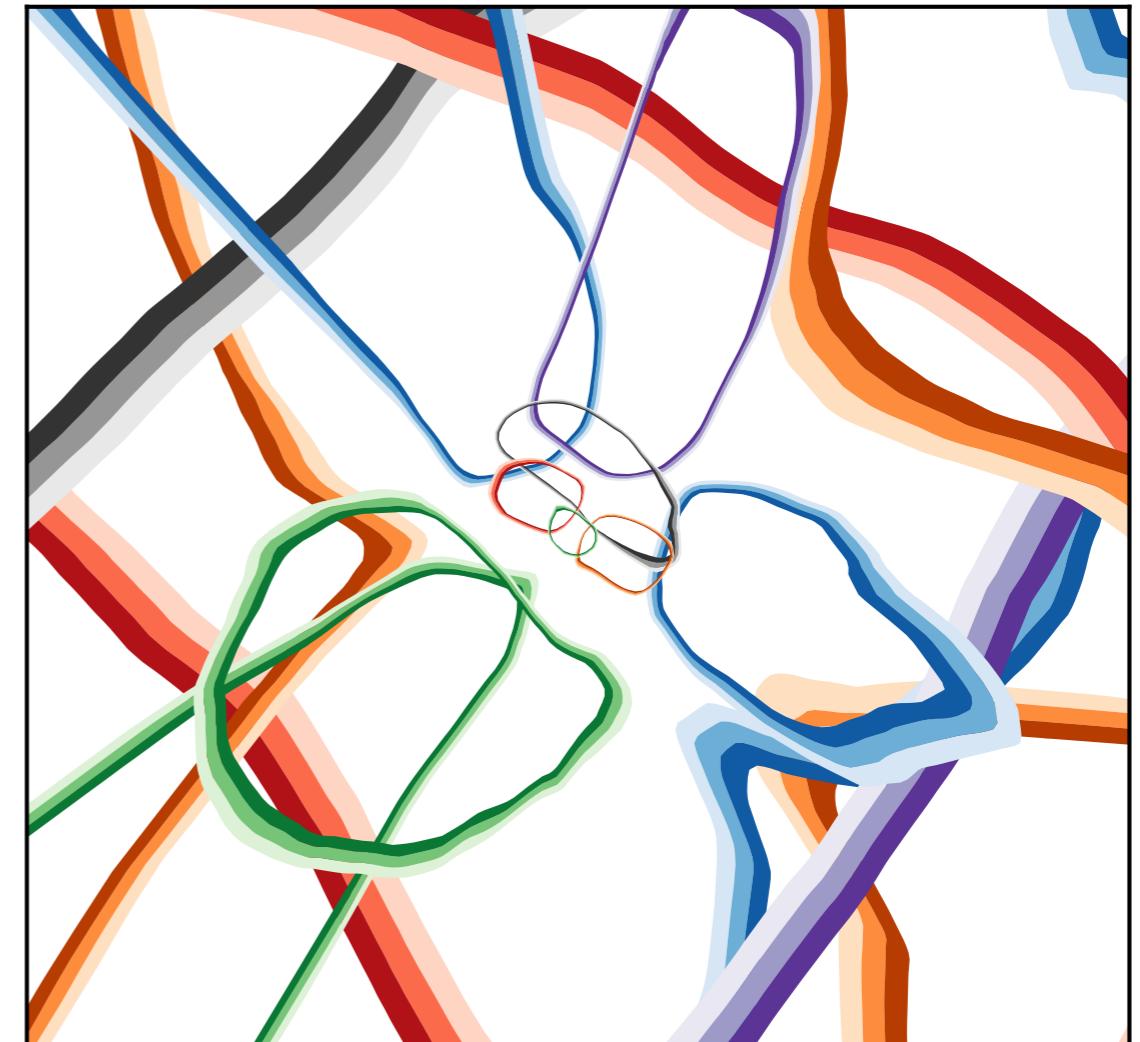


Psychedelic Network Visualization

Latent Dimension 8



Latent Dimension 16



“Hello, World!” of Jets:

Quark



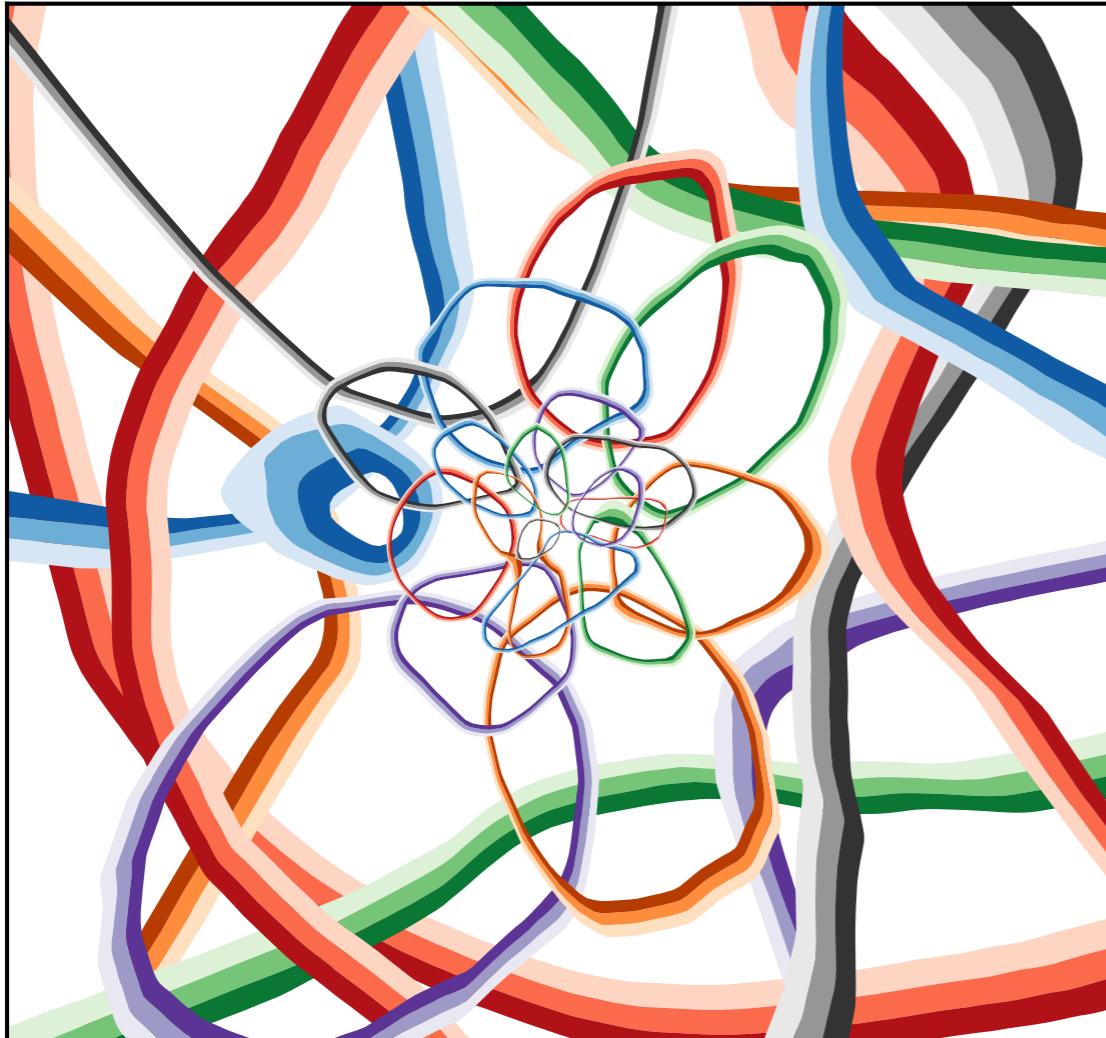
vs.

Gluon



Psychedelic Network Visualization

Latent Dimension 32

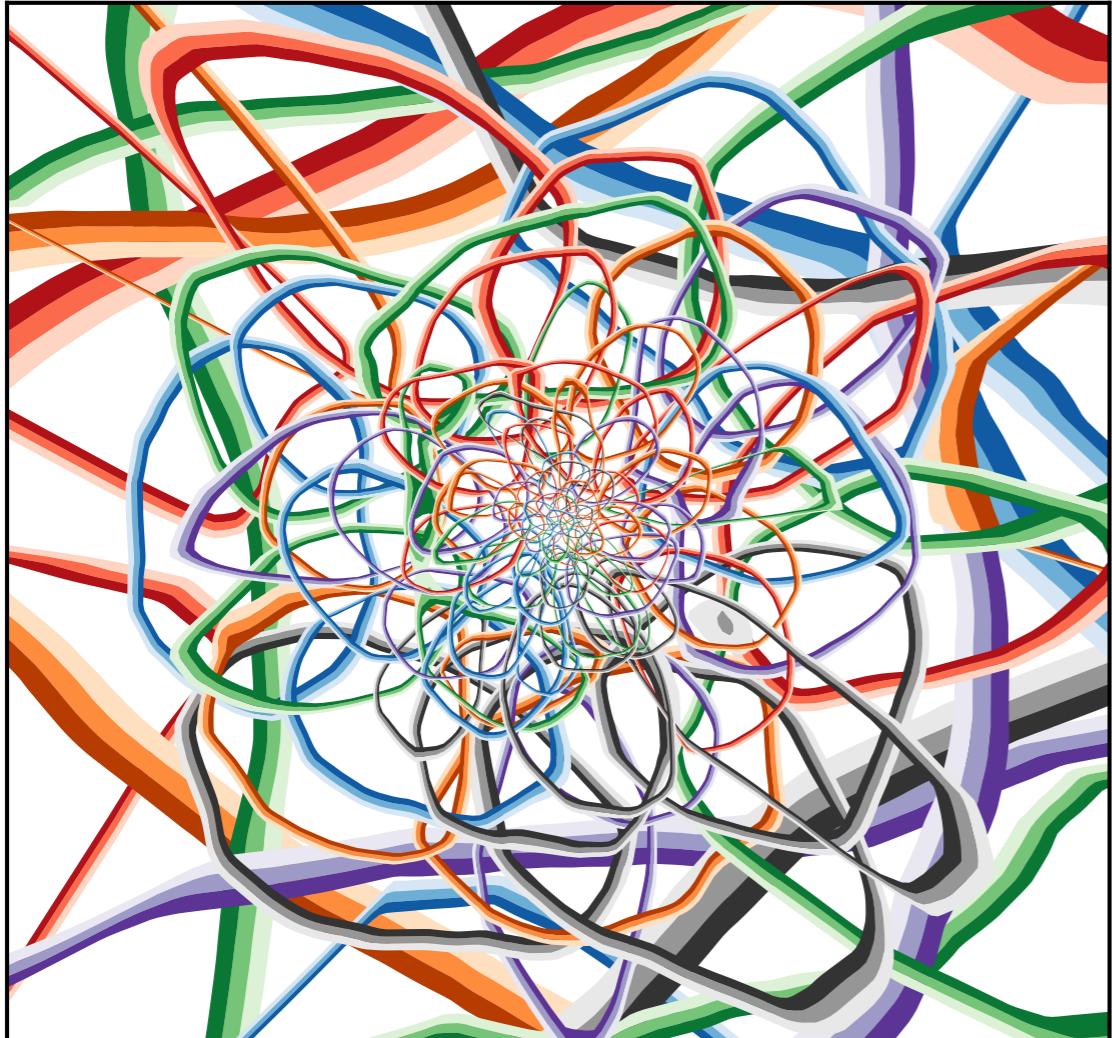


Latent Dimension 64

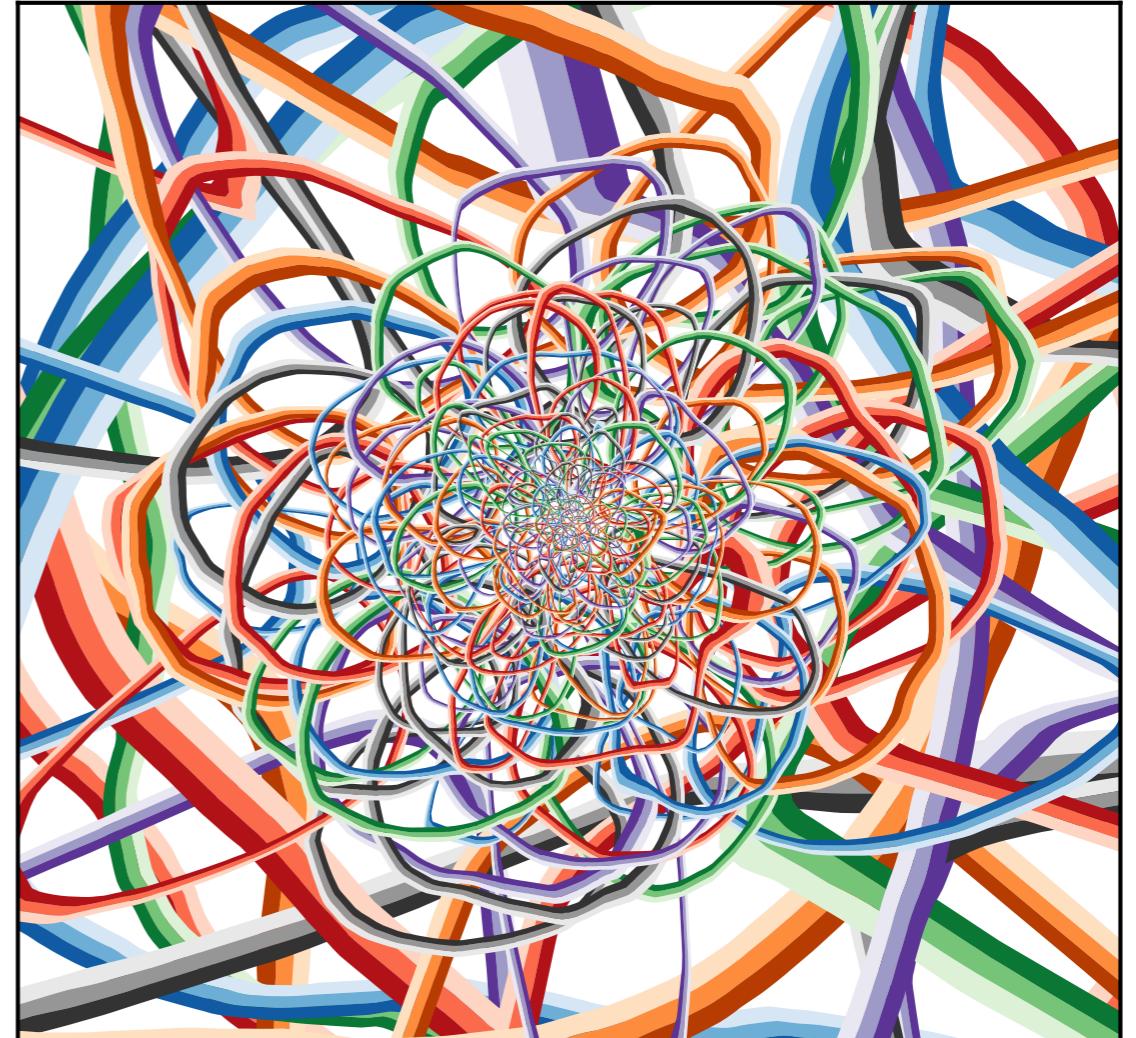


Psychedelic Network Visualization

Latent Dimension 128

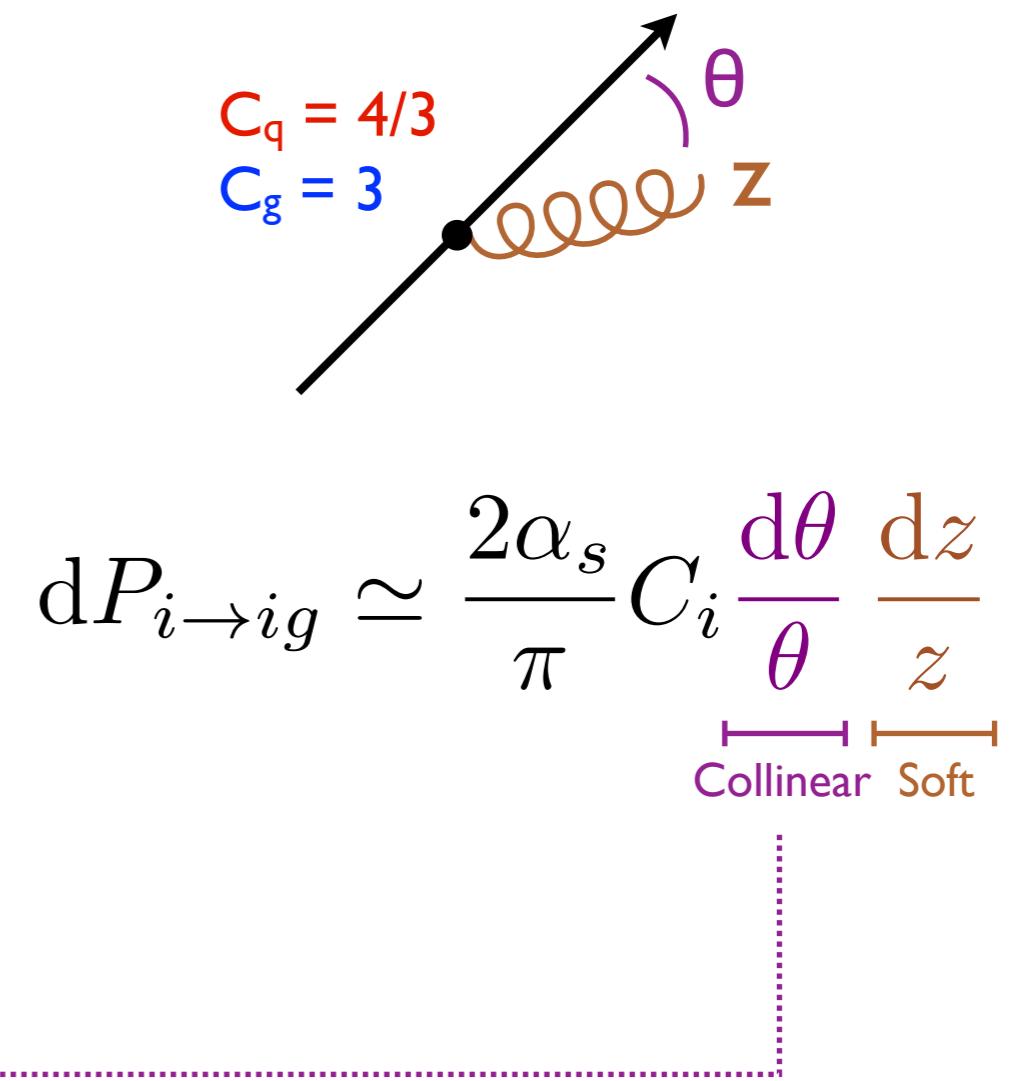
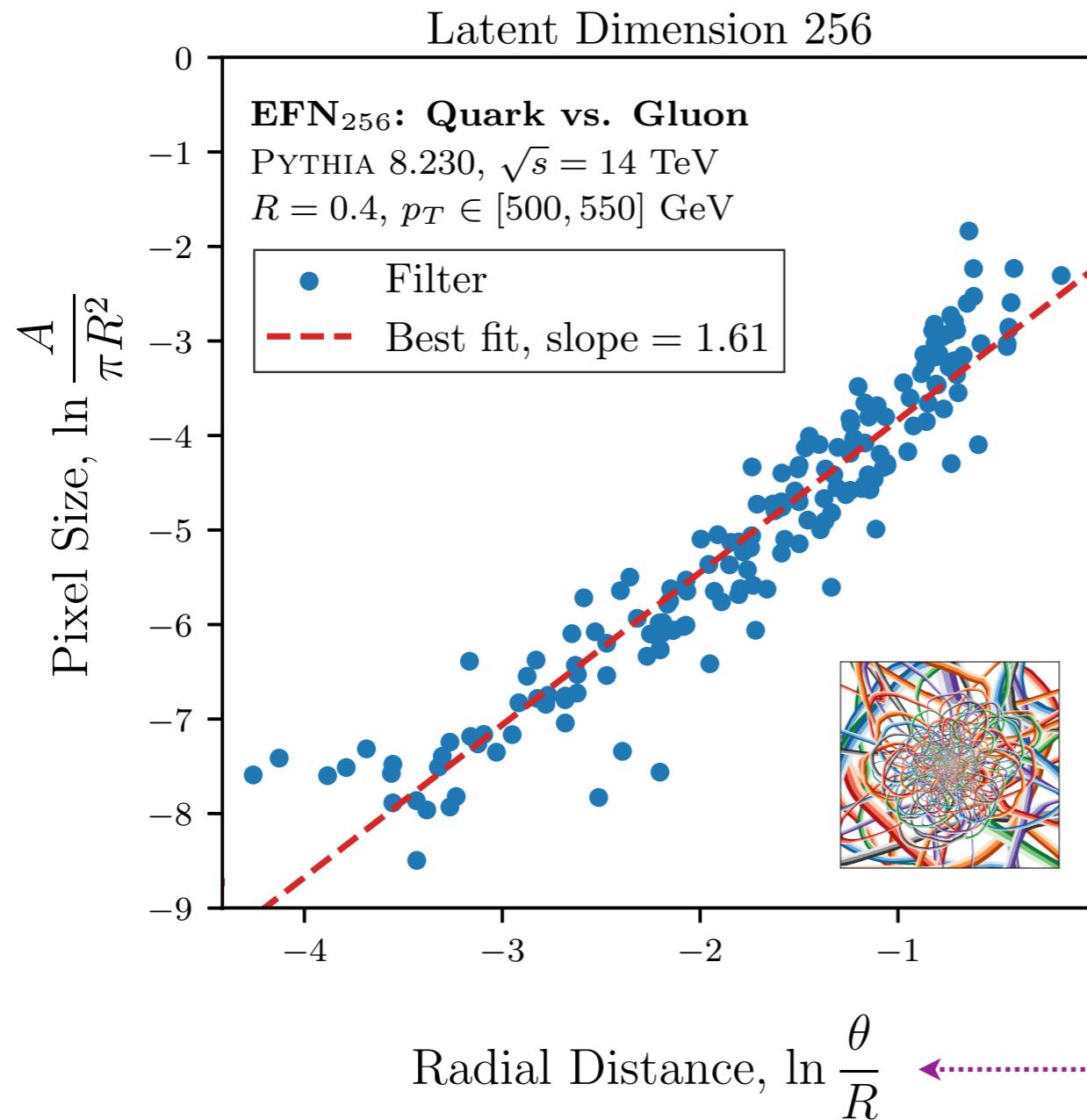
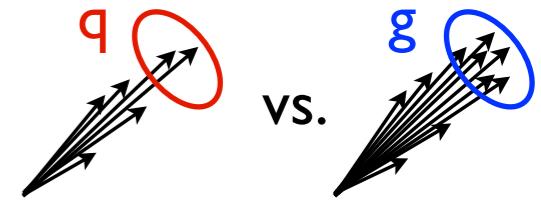


Latent Dimension 256



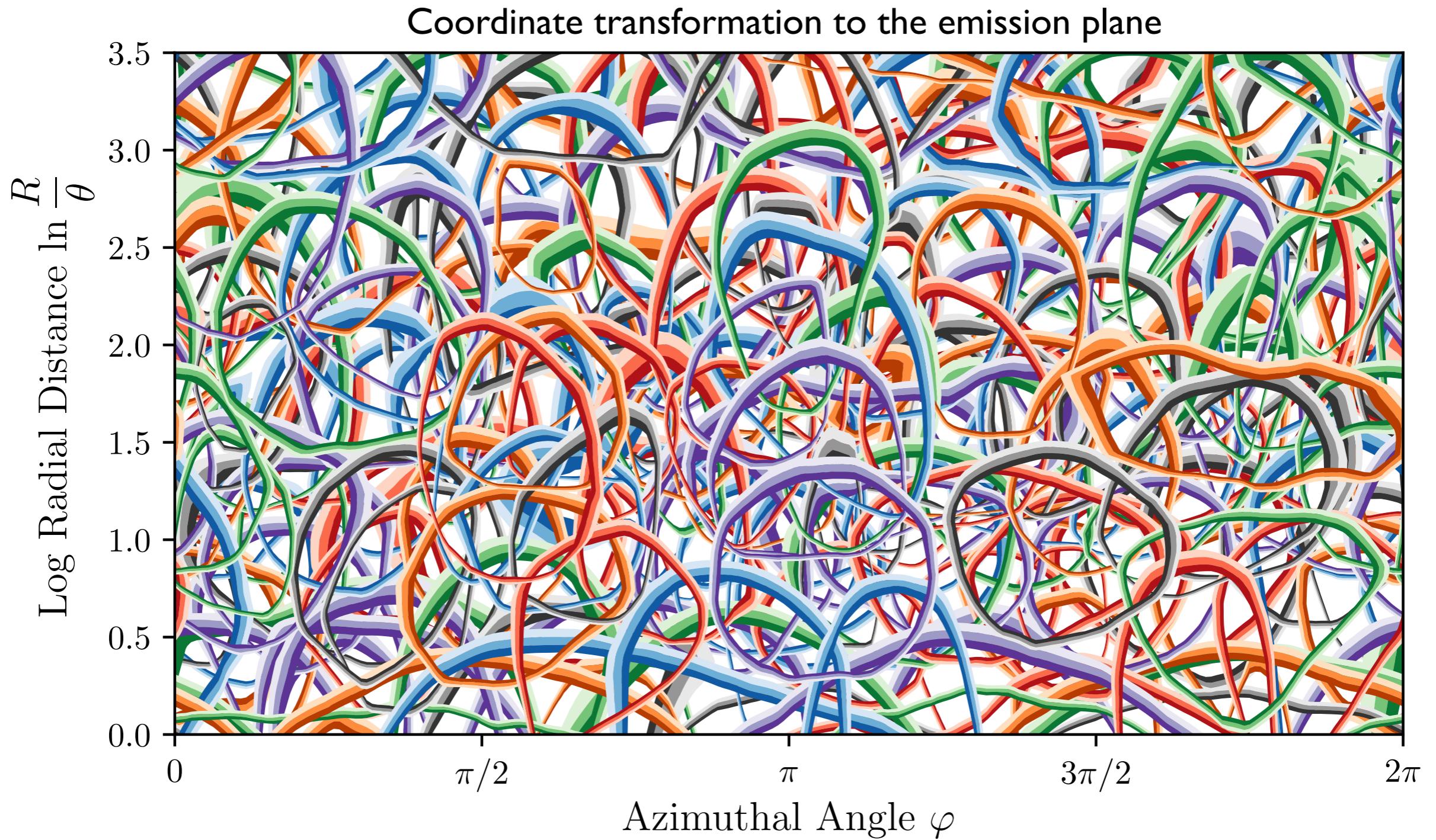
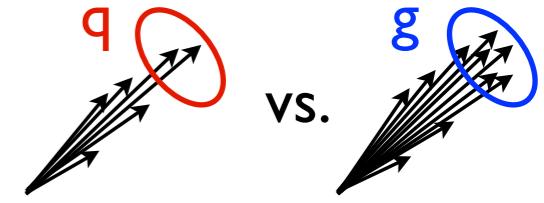
Fractal structure of the strong force!

Scaling of Strong Interactions

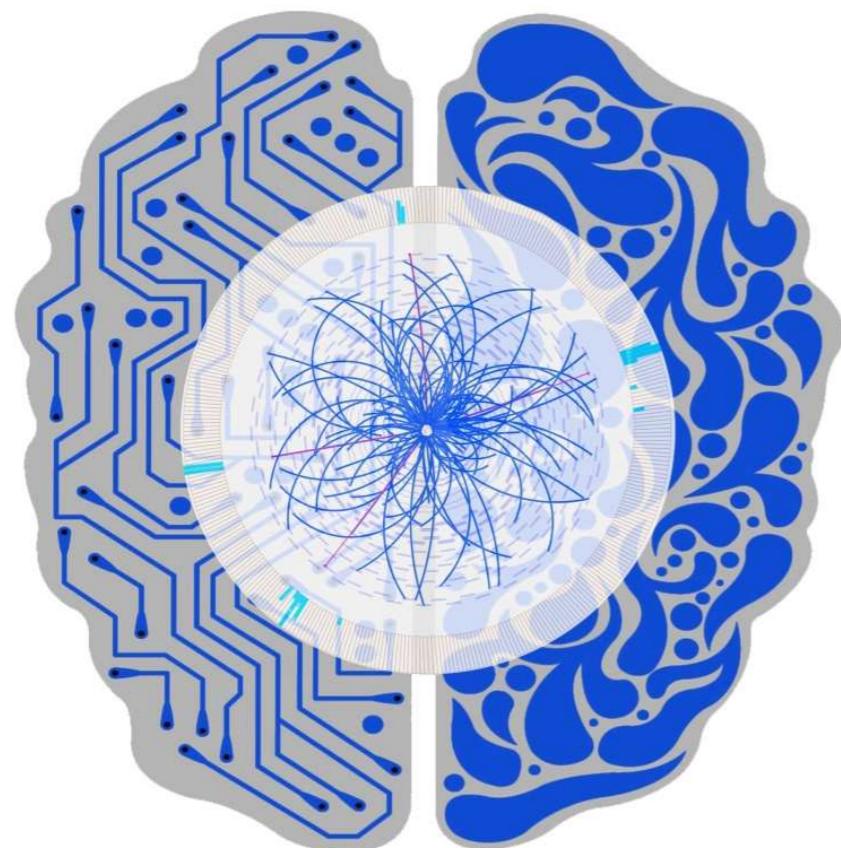


[Komiske, Metodiev, JDT, JHEP 2019]

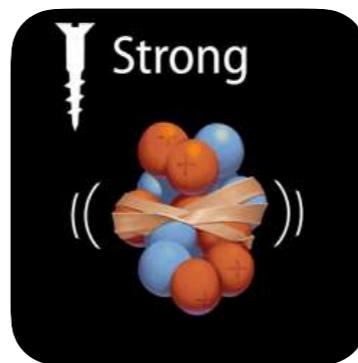
Ready for the ICA?



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



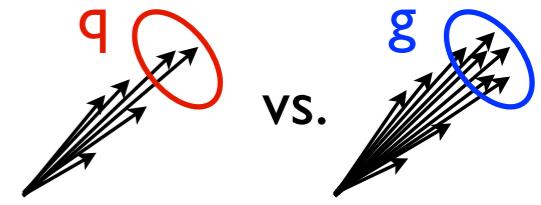
*We taught a machine to
“think” like a physicist...*



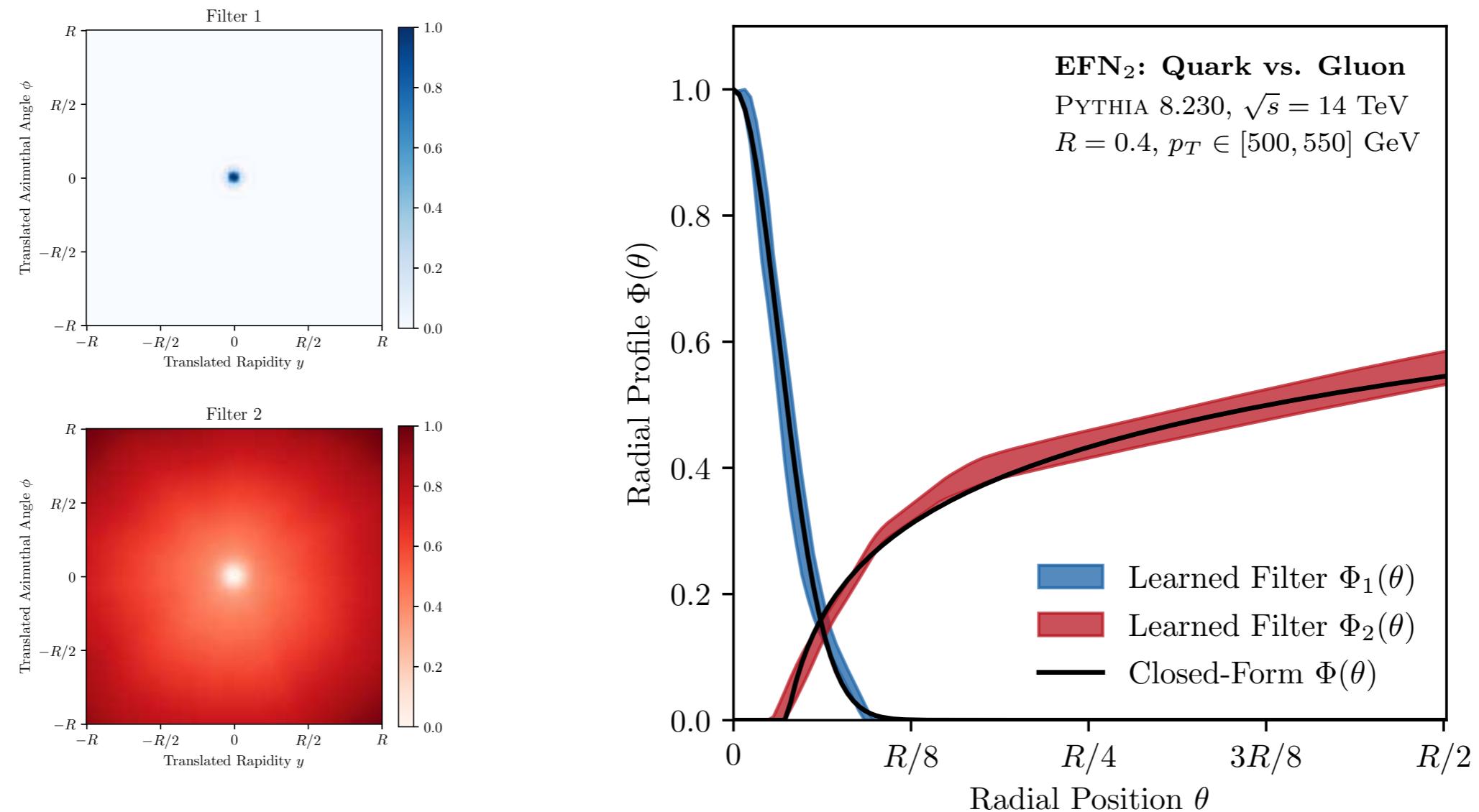
*...and it learned fractal
structure of strong force!*

*“Ok, but did you really learn something
you didn’t already know?”*

Learning from the Machine



For $\ell = 2$ EFN, radial moments: $\sum_{i \in \text{jet}} z_i f(\theta_i)$ cf. Angularities:
 $f(\theta) = \theta^\beta$



[Komiske, Metodiev, JDT, JHEP 2019;
cf. Larkoski, JDT, Waalewijn, JHEP 2014; using Berger, Kucs, Sterman, PRD 2003; Ellis, Vermilion, Walsh, Hornig, Lee, JHEP 2010]

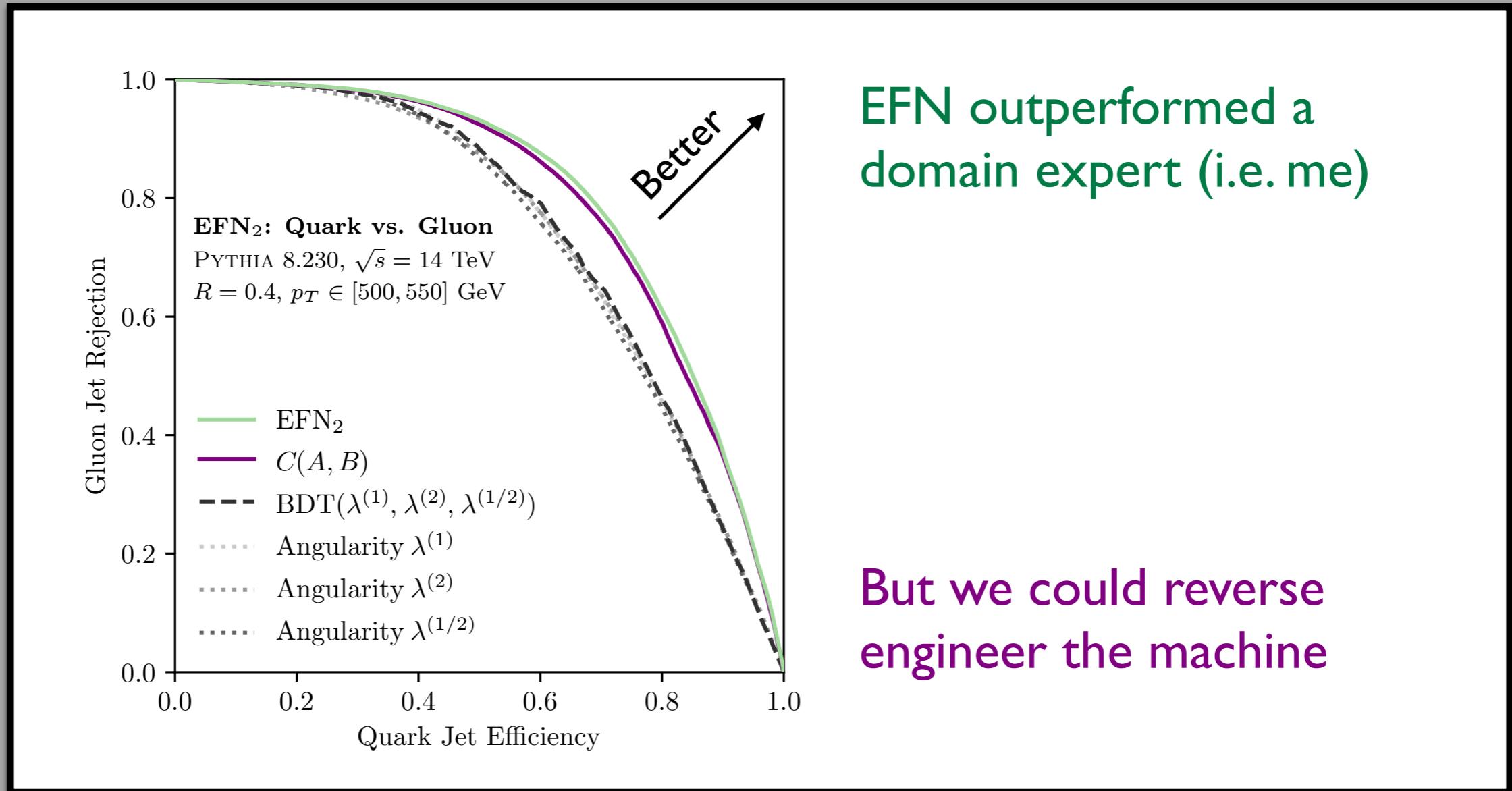
Learning from the Machine



For $\ell = 2$ EFN, radial moments:

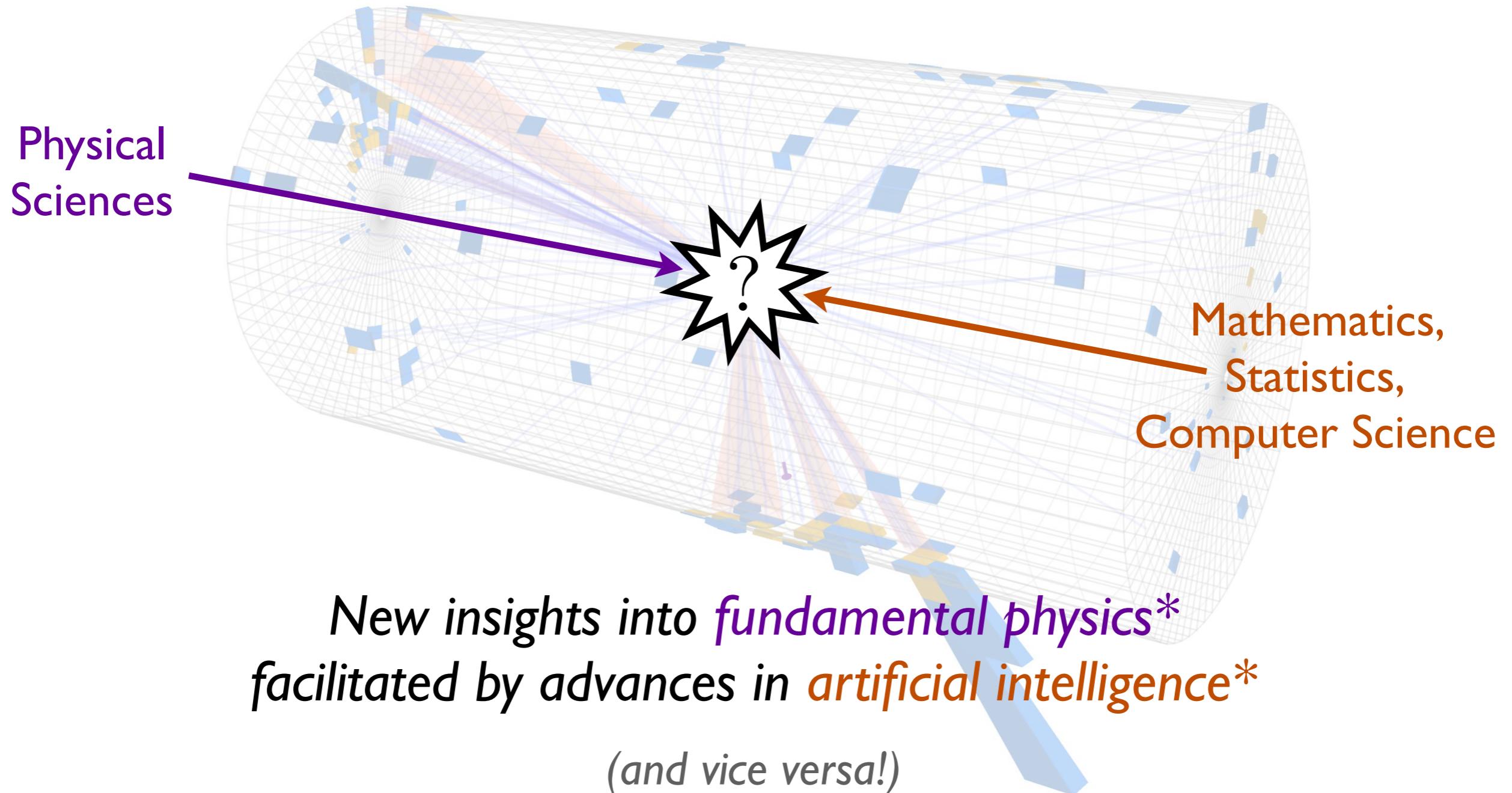
$$\sum_{i \in \text{jet}} z_i f(\theta_i)$$

cf. Angularities:
 $f(\theta) = \theta^\beta$

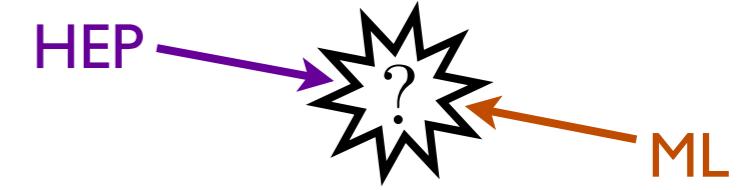


[Komiske, Metodiev, JDT, [JHEP 2019](#);
cf. Larkoski, JDT, Waalewijn, [JHEP 2014](#); using Berger, Kucs, Sterman, [PRD 2003](#); Ellis, Vermilion, Walsh, Hornig, Lee, [JHEP 2010](#)]

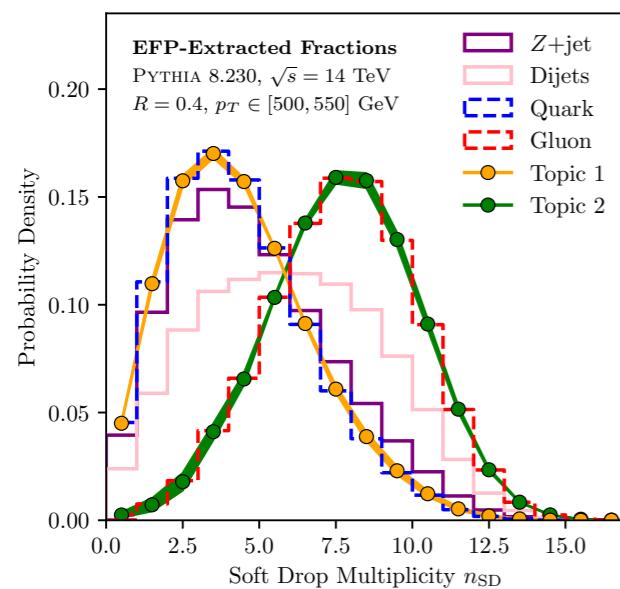
“Collision Course”



Recent Collisions

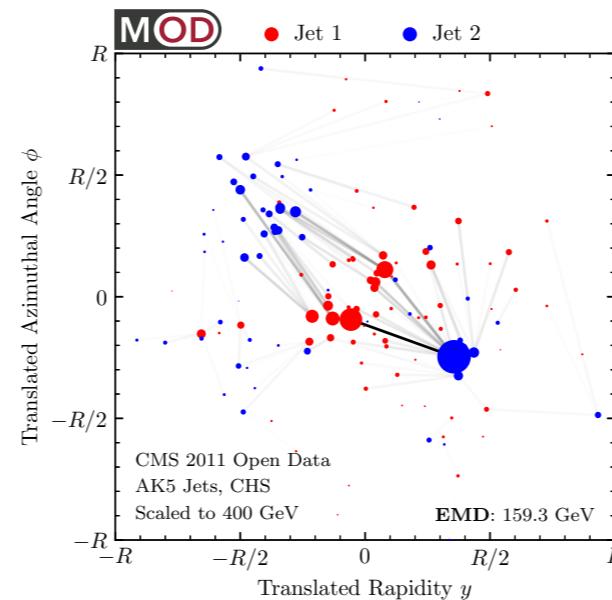


Quark/Gluon Definitions via Blind Source Separation



[Metodiev, JDT, PRL 2018;
Komiske, Metodiev, JDT, JHEP 2018]

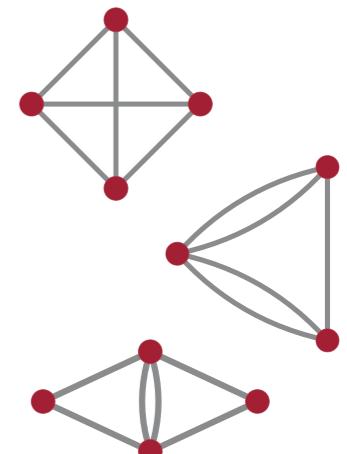
Half-Century of Collider Physics via Optimal Transport (!)



[Komiske, Metodiev, JDT, PRL 2019, JHEP 2020;
Komiske, Mastandrea, Metodiev, Naik, JDT, PRD 2020]

Kinematic Decomposition via Graph Theory

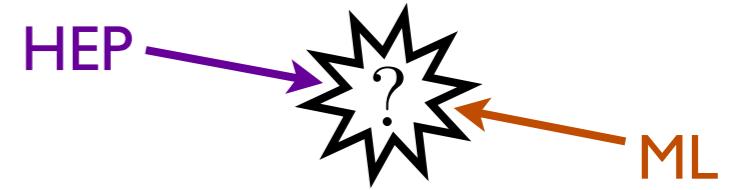
Edges d	Leafless Multigraphs			A307317
	Connected	All	A307316	
1	0	0		
2	1	1		
3	2	2		
4	4	5		
5	9	11		
6	26	34		
7	68	87		
8	217	279		
9	718	897		
10	2 553	3 129		
11	9 574	11 458		
12	38 005	44 576		
13	157 306	181 071		
14	679 682	770 237		
15	3 047 699	3 407 332		
16	14 150 278	15 641 159		



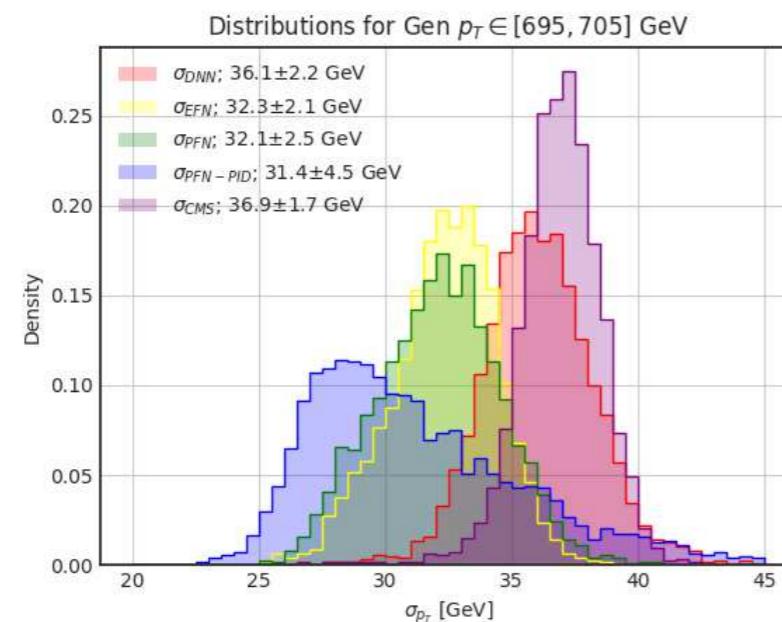
[Komiske, Metodiev, JDT, JHEP 2018, PRD 2020]

High Energy Physics \leftrightarrow Mathematics, Statistics & Computer Science

Ongoing Collisions

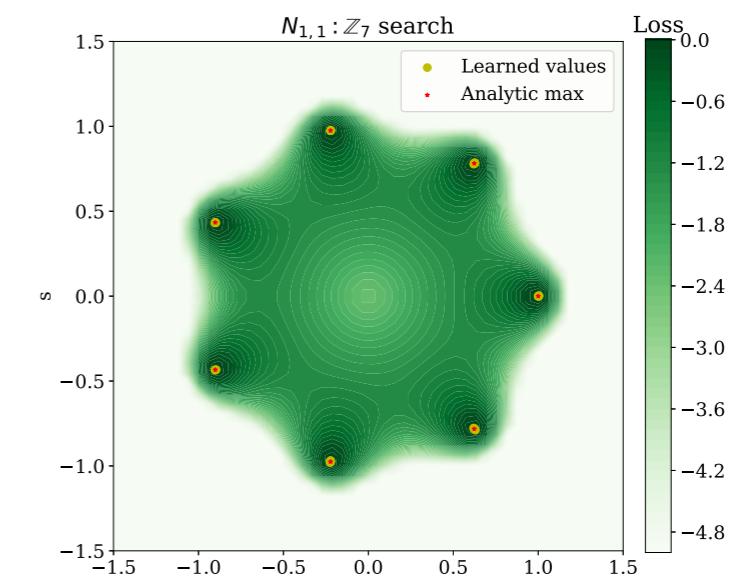


Frequentist Jet Calibration via Gaussian Ansatz



[Nachman, Gambhir, JDT, in progress]

Symmetry Discovery via Adversarial Learning



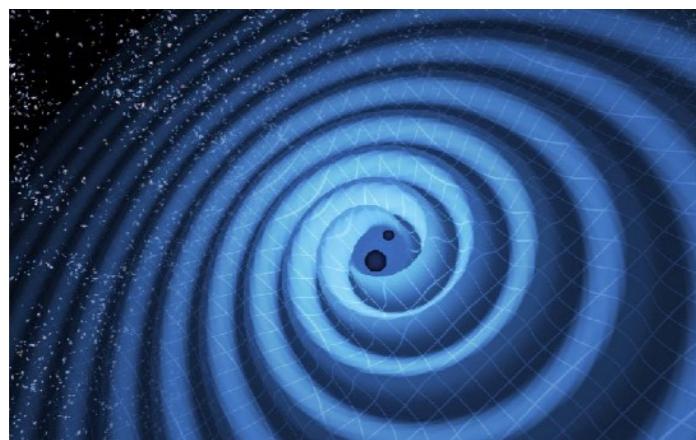
[Desai, Nachman, JDT, NeurIPS 2021 ML4PS]

High Energy Physics \leftrightarrow Mathematics, Statistics & Computer Science

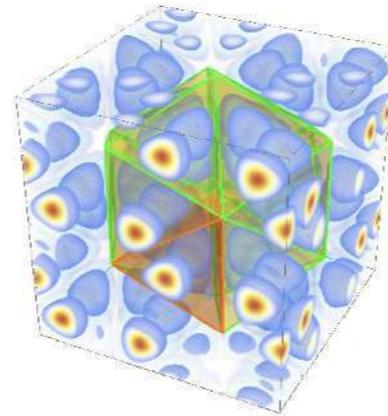
Artificial Intelligence \leftrightarrow Fundamental Interactions



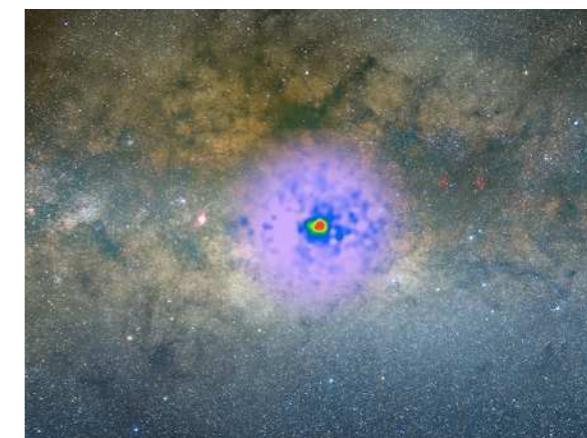
Gravitational Waves



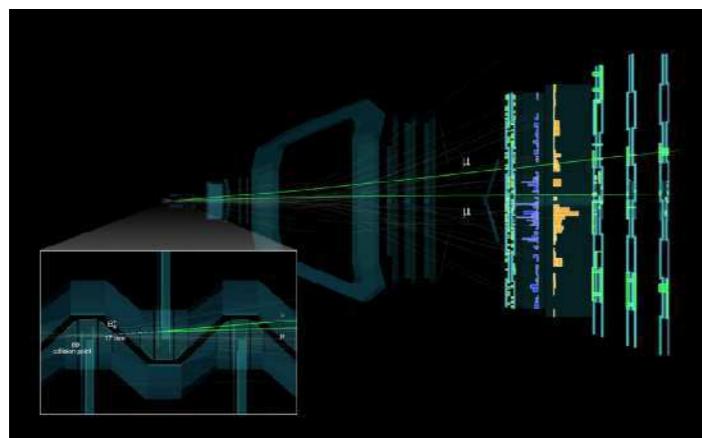
Nuclear Physics



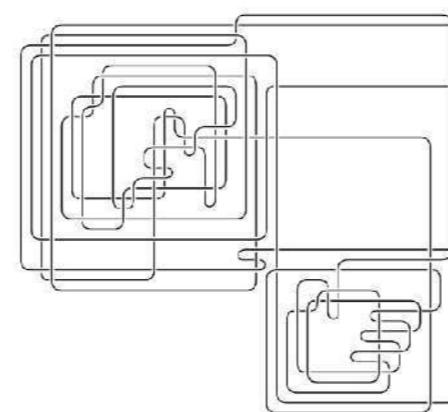
Astrophysics



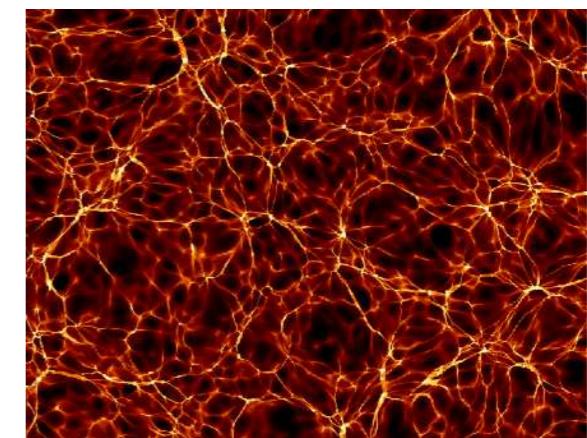
Particle Colliders



Mathematical Physics



Dark Matter



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/>

Physics
Theory



AI Foundations

Physics
Experiment

