Machine Learning for Particle Physics: What Can Humans Learn?

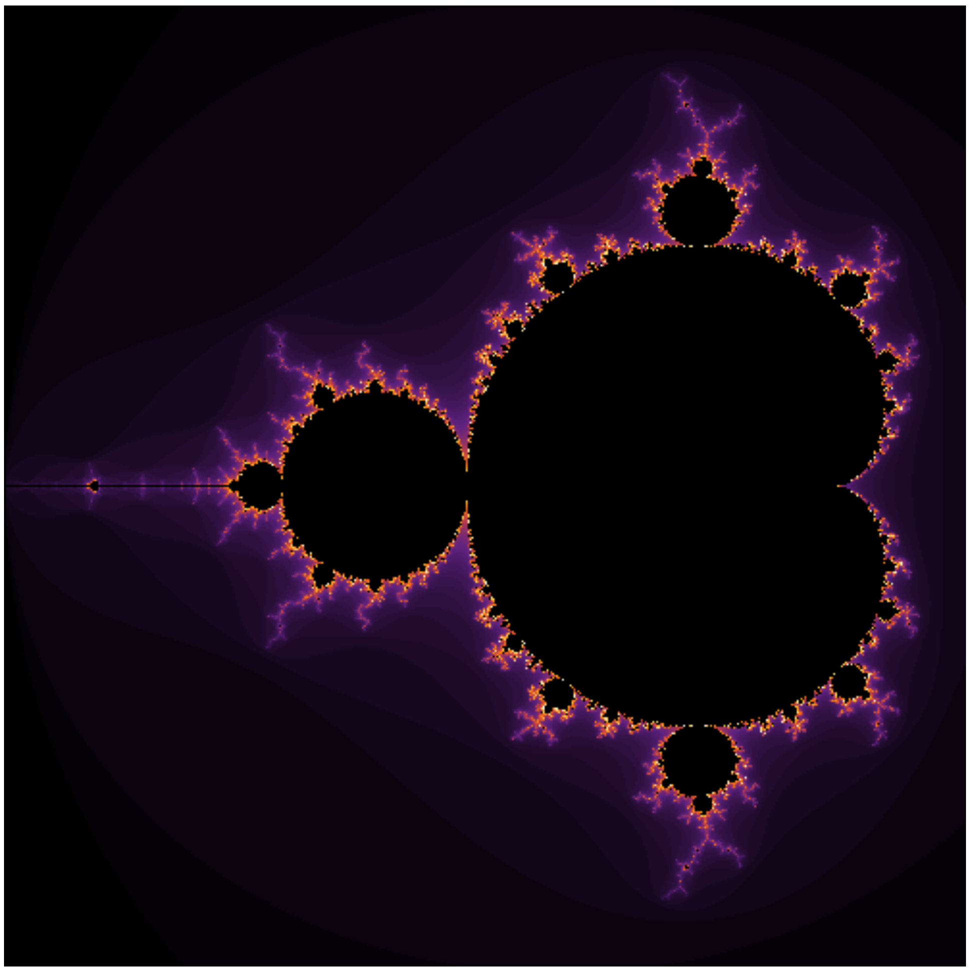
Andrew Larkoski
Reed College

Based on related work in collaboration with:
Kaustuv Datta, Ian Moult, Eric Metodiev, Duff Neill, Gavin Salam, Jesse Thaler, Wouter Waalewijn
…

IPMU, February 17, 2020

Complexity vs. Information

Here is a cool image:



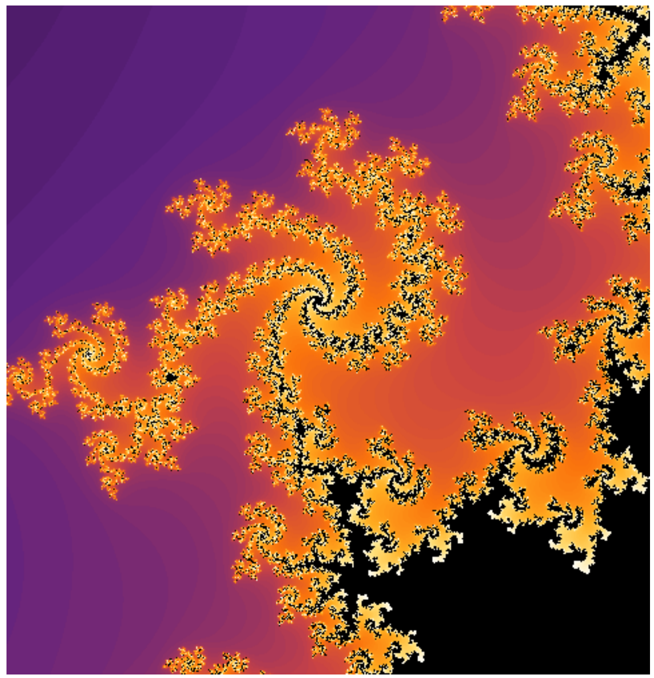
Likely, you know what it is, but I won't ruin the suspense yet.

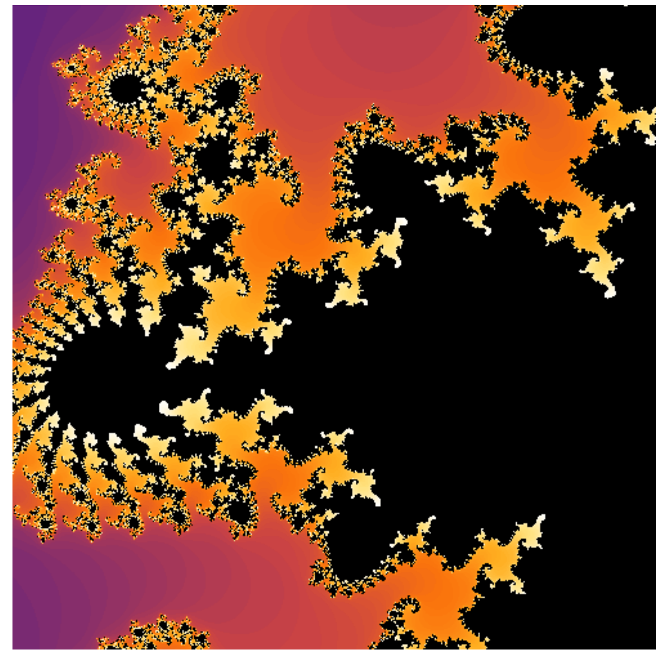
You might think there there is a huge amount of information.

The more you zoom into the image, the more you see.

Complexity vs. Information

To understand this image, you might zoom in and find:





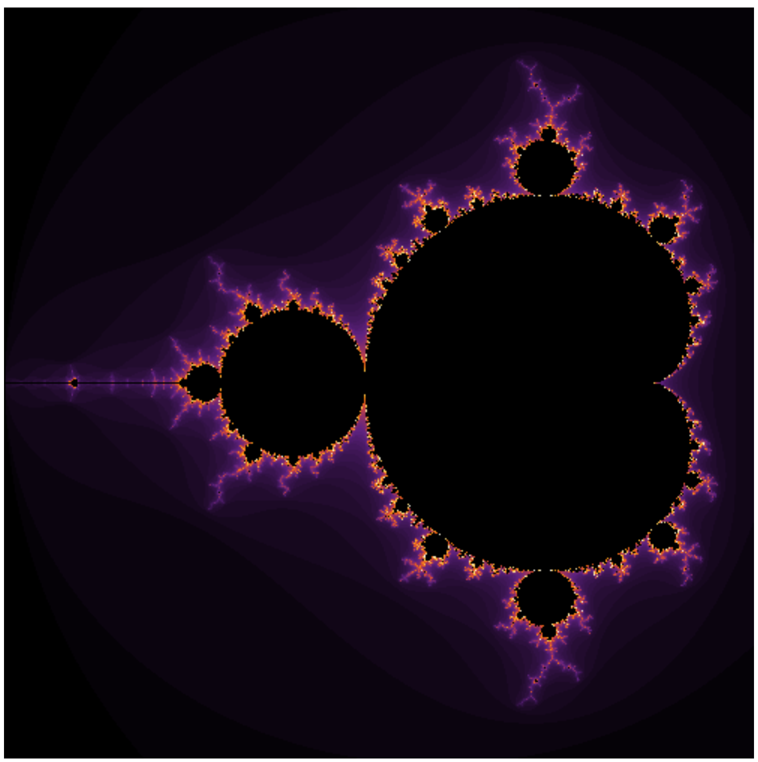
These look very different, and if you focus too much on small sections, you might not see the larger structure.

Complexity vs. Information

Then, if I tell you this is the Mandelbrot set, defined by the region of convergence from recursively applying:

$$f(z) = z^2 + c$$

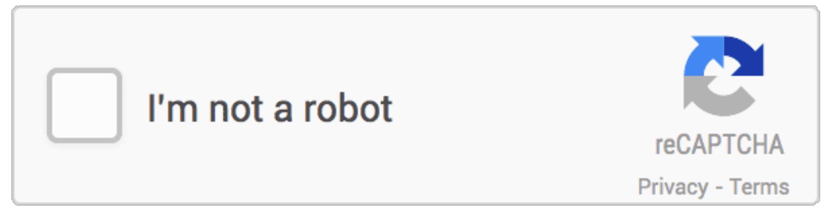
You will likely be very surprised!

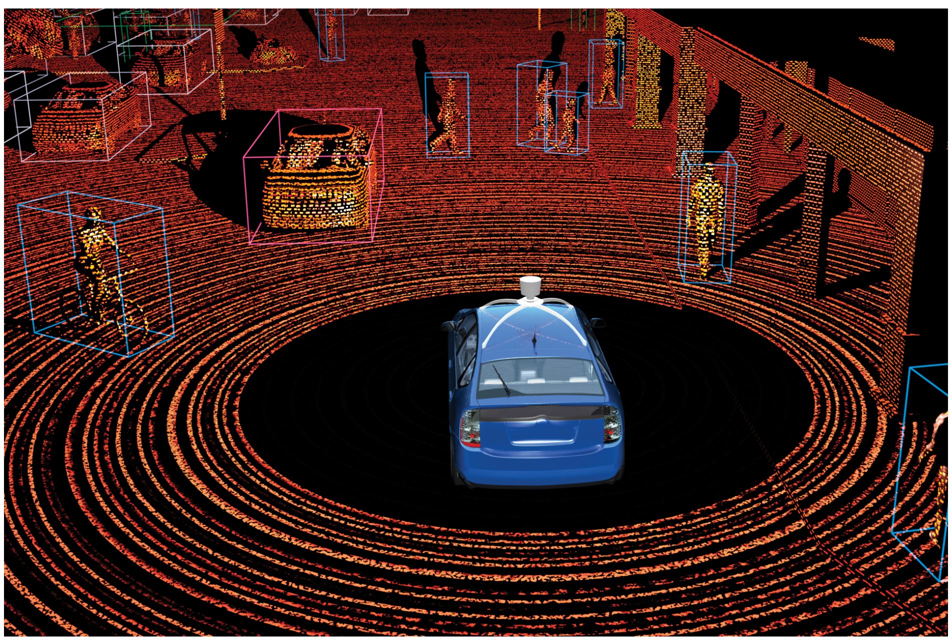


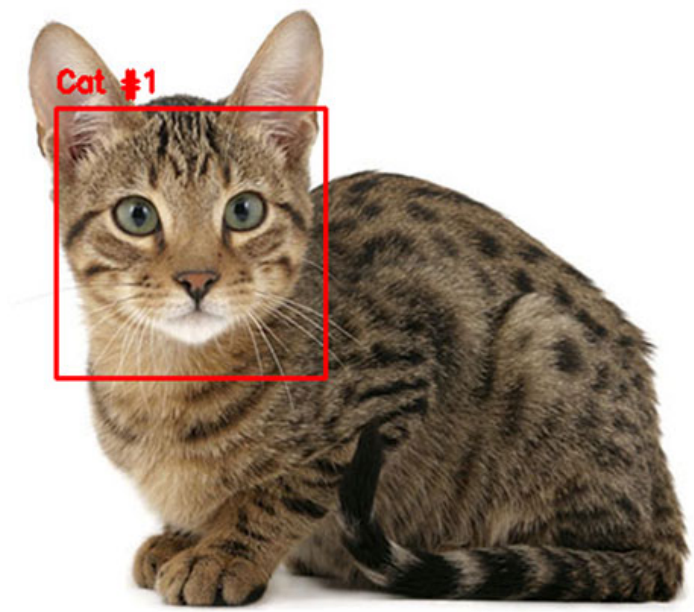
Complexity does not equal explosion of information

Fractals, like the Mandelbrot set, can have arbitrary complexity from simple rules

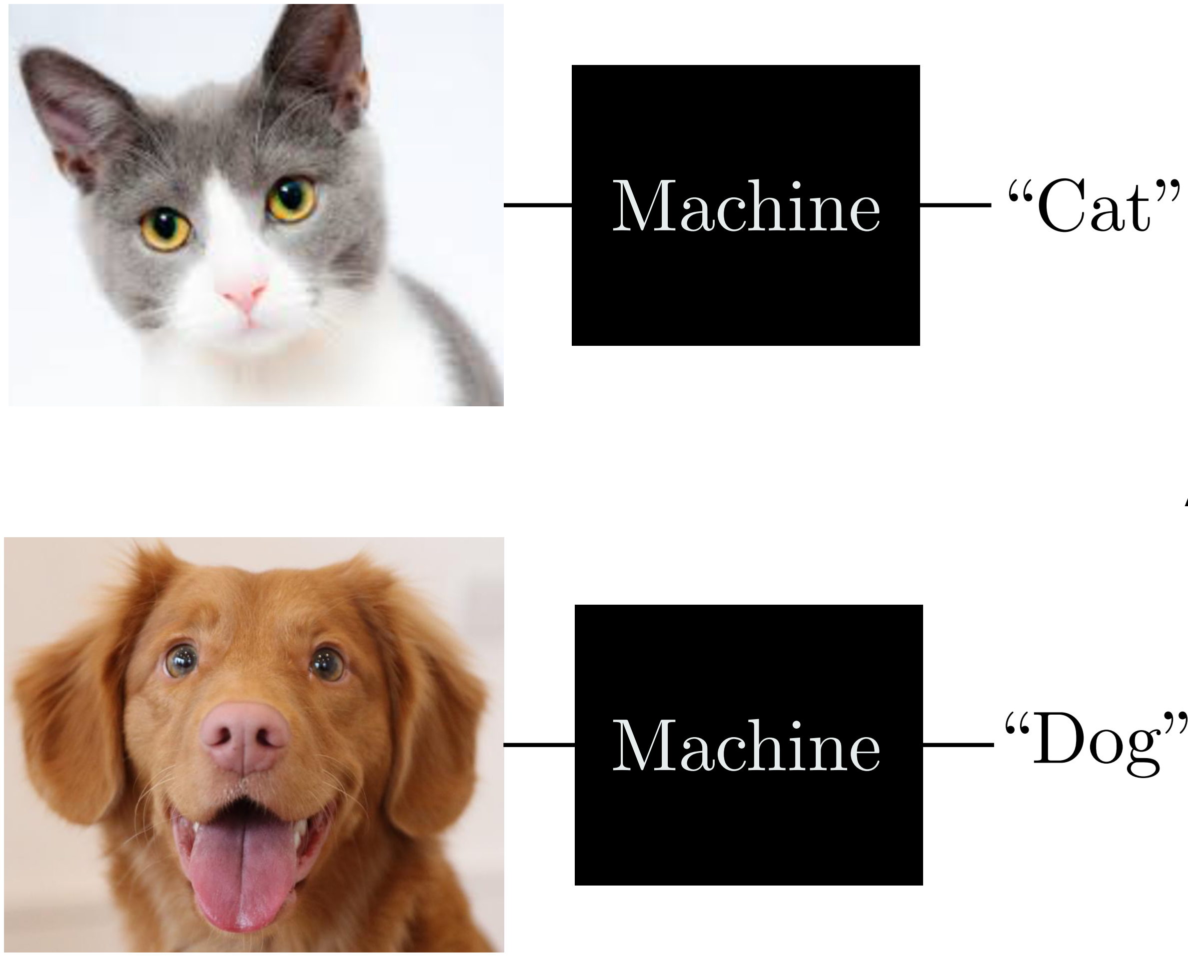
Machine Learning is Everywhere!







Canonical Problem: Binary Discrimination



The machine learns distinguishing features

As a physicist, “machine” is just a black box

Goal: Determine the output of a perfect machine

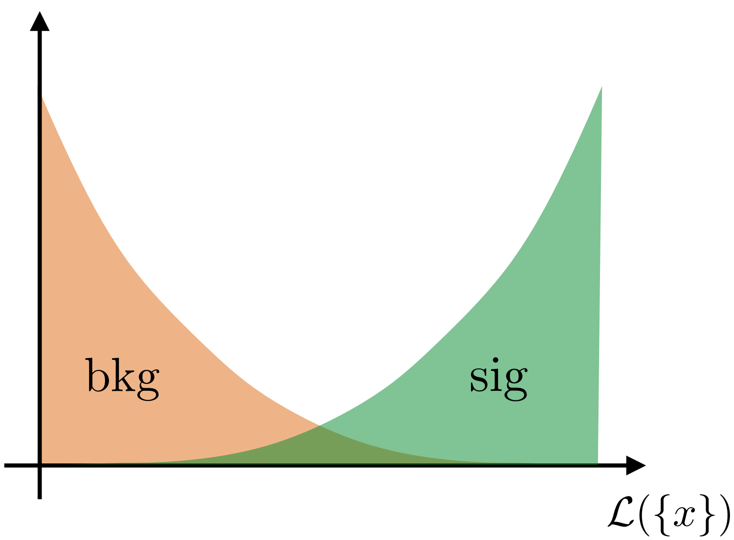
Canonical Problem: Binary Discrimination

Guiding Principles:

<u>Neyman-Pearson Lemma</u>

The optimal binary discriminant
is monotonic in the likelihood

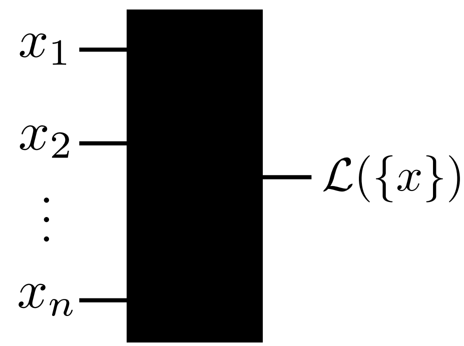
$$\mathcal{L}(\{x\}) = \frac{p_S(\{x\})}{p_B(\{x\})}$$

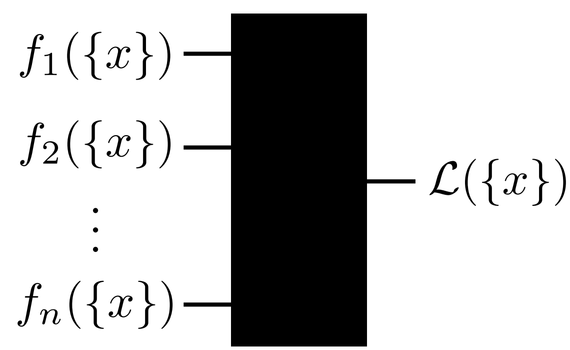


Neyman, Pearson 1933

<u>Universal Approximation Theorem</u>

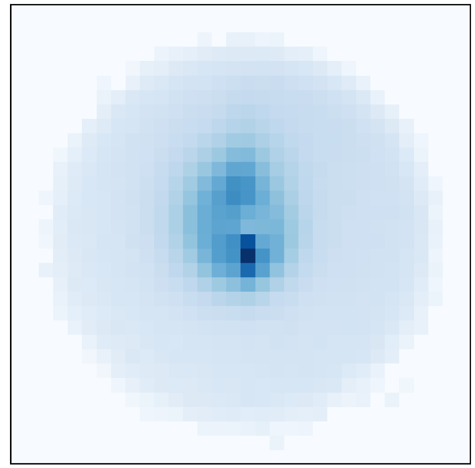
A “good” machine can
output any function of the input

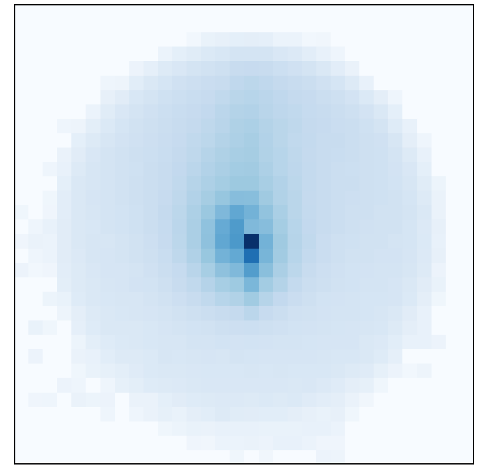




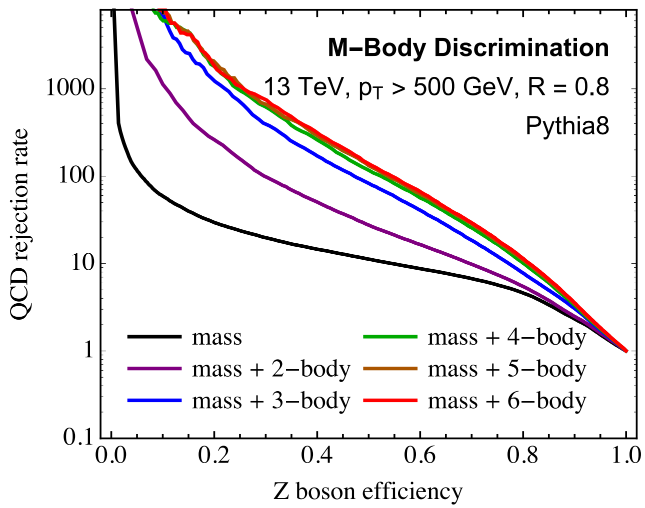
Cybenko 1989; *et al.*

Outline



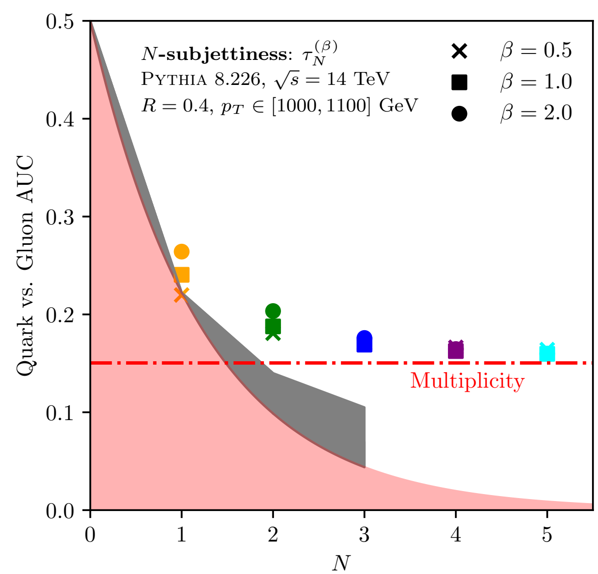


Machine Learning on Jets at the LHC



Simplifying the Discrimination Space

Insights into Quark vs. Gluon Discrimination



For More Information

Jet Substructure at the Large Hadron Collider:
A Review of Recent Advances in Theory and Machine Learning

Andrew J. Larkoski*
Physics Department, Reed College, Portland, OR 97202, USA

Ian Moult†
Berkeley Center for Theoretical Physics, University of California, Berkeley, CA 94720, USA and
Theoretical Physics Group, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

Benjamin Nachman‡
Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

arXiv:1709.04464

Annual Review of Nuclear and Particle Science

Deep Learning and Its Application to LHC Physics

Dan Guest,[1] Kyle Cranmer,[2] and Daniel Whiteson[1]

[1]Department of Physics and Astronomy, University of California, Irvine, California 92697, USA
[2]Physics Department, New York University, New York, NY 10003, USA

arXiv:1806.11484

nature
International journal of science

Review Article | Published: 01 August 2018

Machine learning at the energy and intensity frontiers of particle physics

Alexander Radovic, Mike Williams, David Rousseau, Michael Kagan, Daniele Bonacorsi, Alexander Himmel, Adam Aurisano, Kazuhiro Terao & Taritree Wongjirad

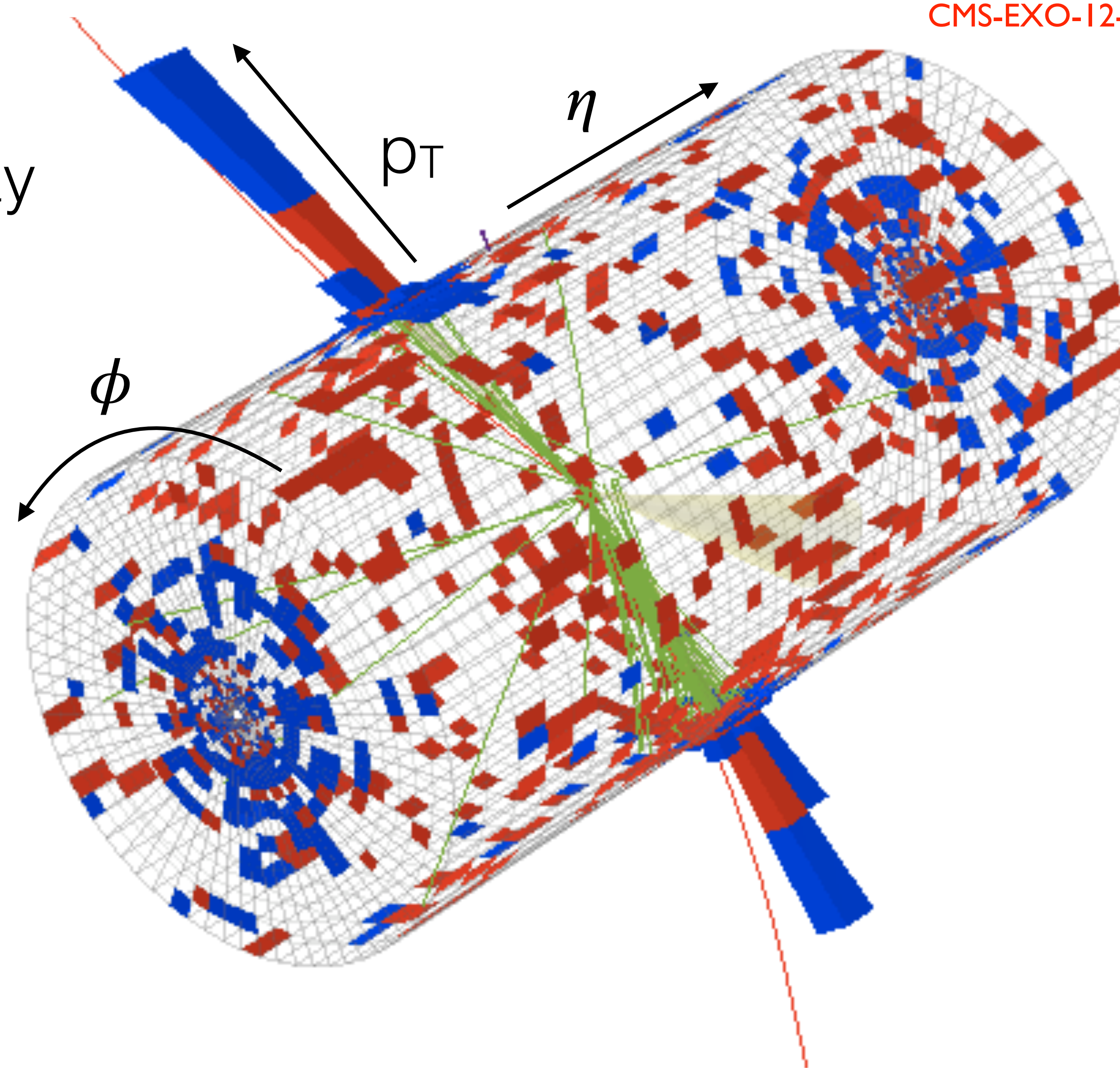
Nature 560 (2018) no.7716, 41-48

Machine Learning on Jets at the LHC

Collision Events at the LHC

Example Event Display
from CMS

CMS-EXO-12-059

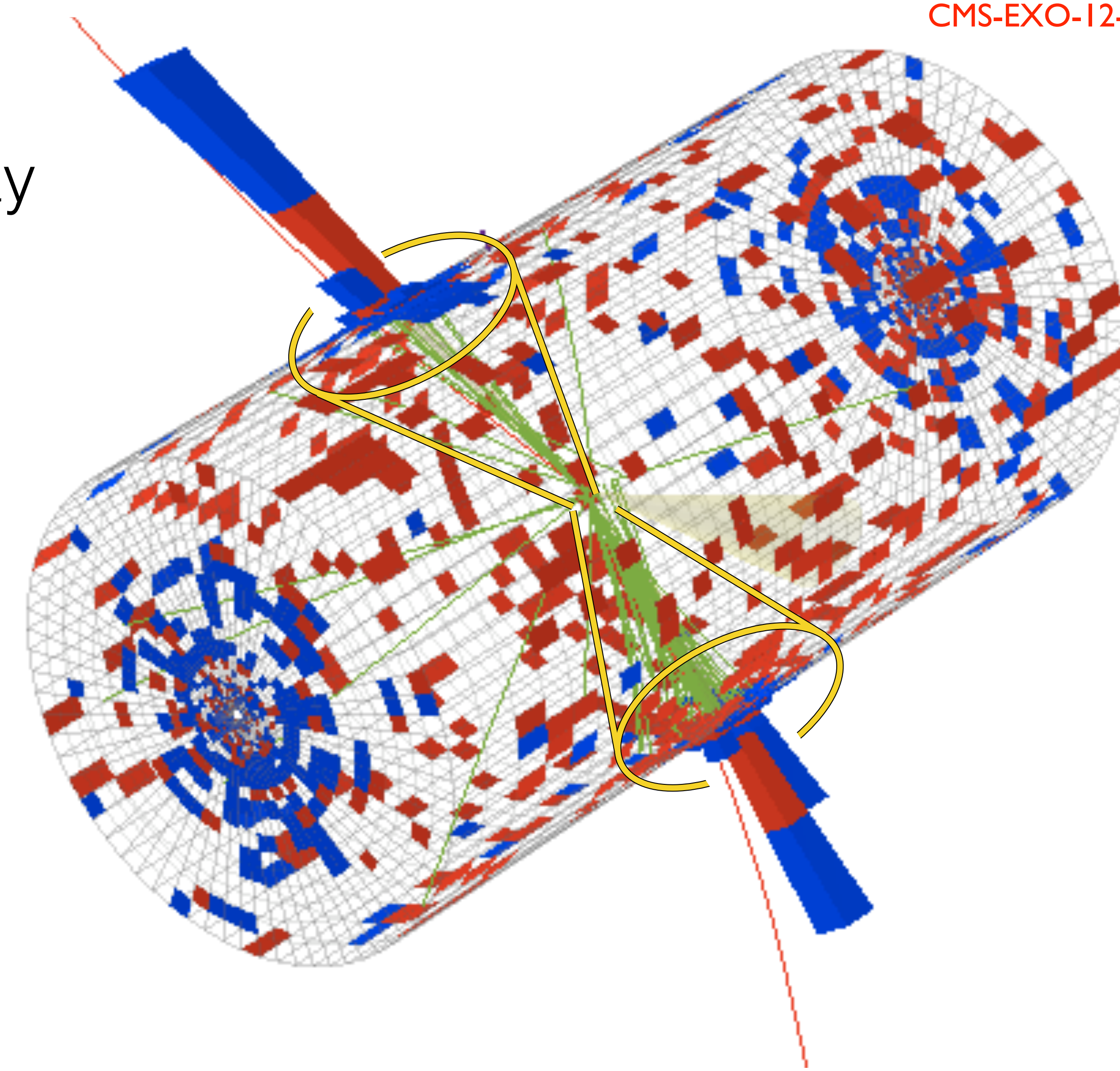


Collision Events at the LHC

Example Event Display
from CMS

Dijet Event

CMS-EXO-12-059



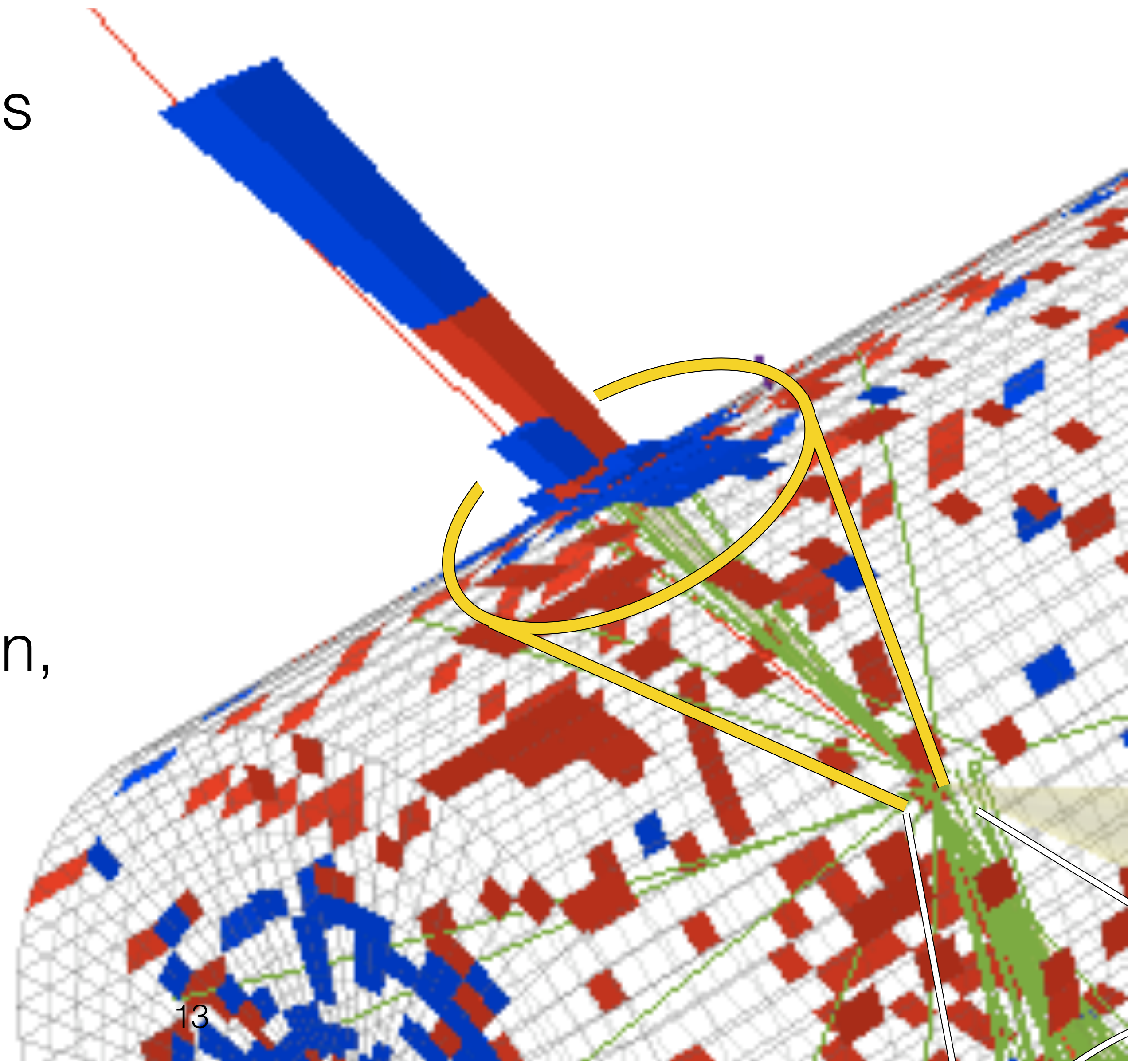
Collision Events at the LHC

CMS-EXO-12-059

Focus on one of the jets

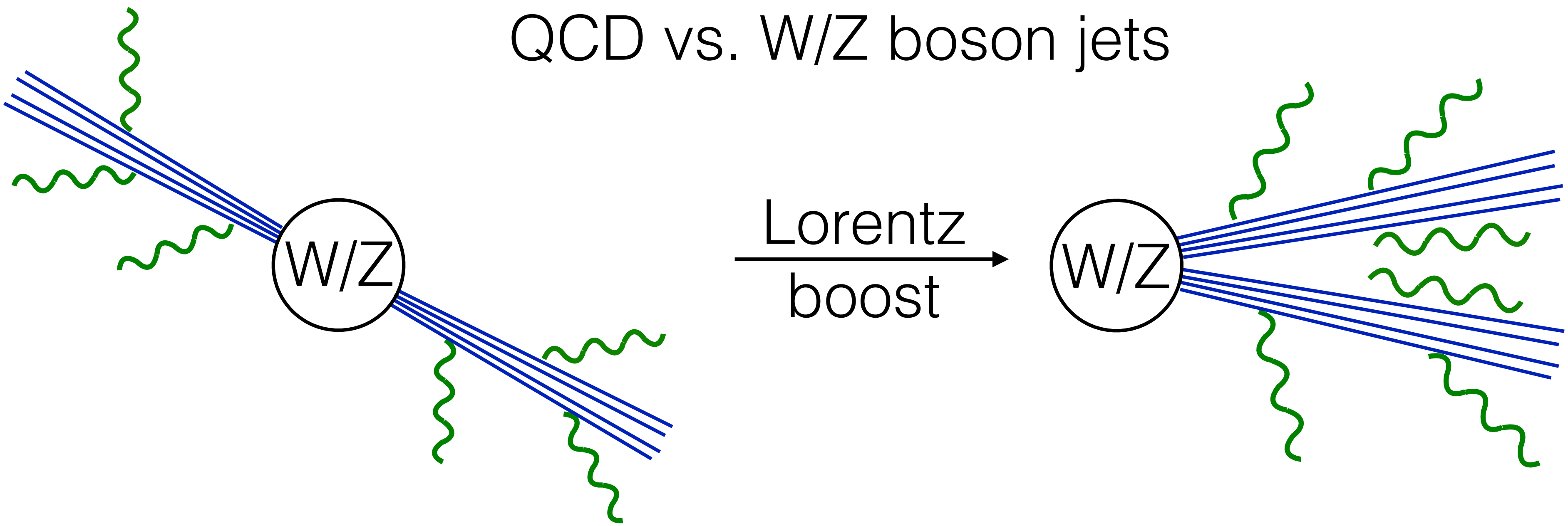
What particle initiated this jet?

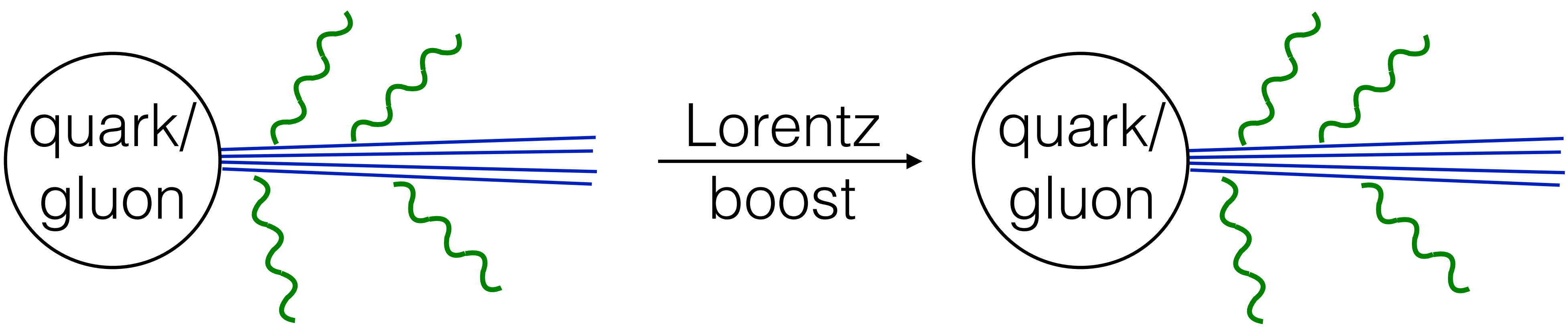
Is it just a quark or gluon, or something more interesting?



Jet Identification as Image Recognition

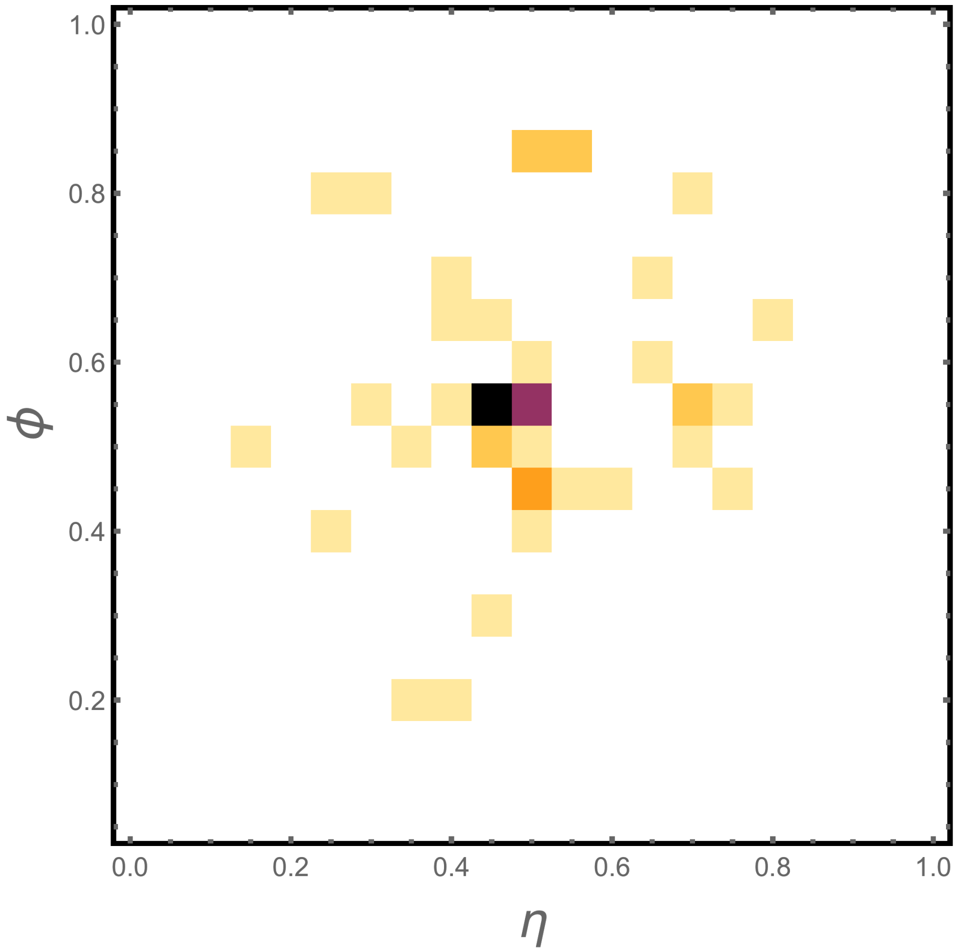
Canonical Discrimination Problem:
QCD vs. W/Z boson jets





Jet Identification as Image Recognition

Sample Jet



Pixels = Location in (η,ϕ)

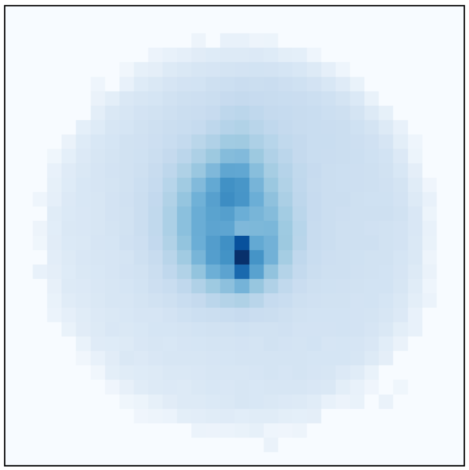
Color = Magnitude of p_T

Think of the jet as imaged by the detector

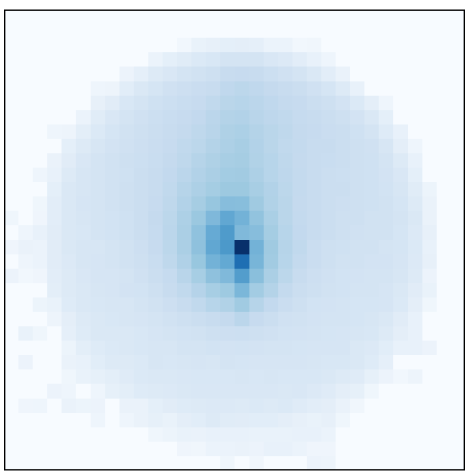
Jet Identification as Image Recognition

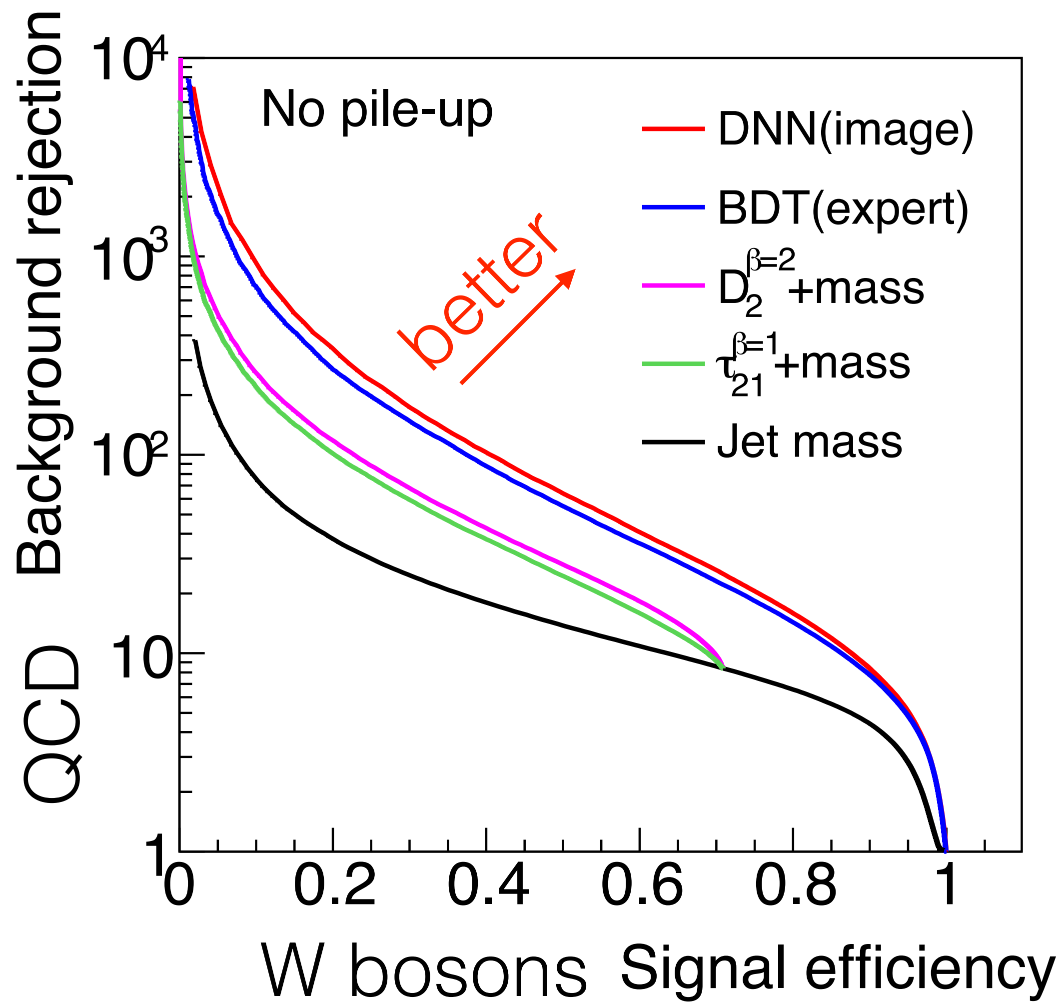
After pre-processing

Boosted W



QCD





Baldi, Bauer, Eng, Sadowski, Whiteson 2016

other work:

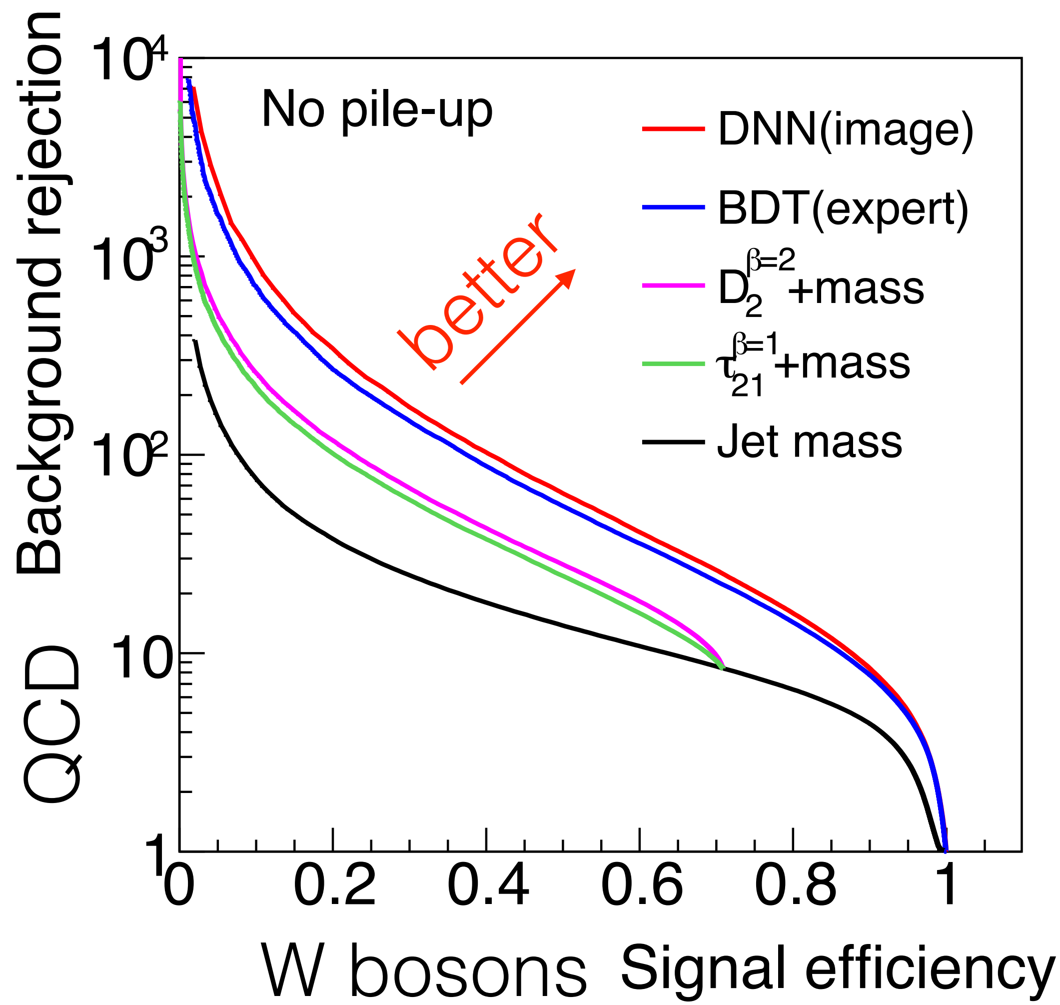
- de Oliveira, Kagan, Mackey, Nachman, Schwartzman 2015
- Louppe, Cho, Becot, Cranmer 2017

Jet Identification as Image Recognition

Image:
Large number of inputs
(32x32 grid)

Expert BDT:
Very small number of inputs
(6 variables)

Why is the image preferable
to the expert BDT?



Baldi, Bauer, Eng, Sadowski, Whiteson 2016
other work:
de Oliveira, Kagan, Mackey, Nachman, Schwartzman 2015
Louppe, Cho, Becot, Cranmer 2017

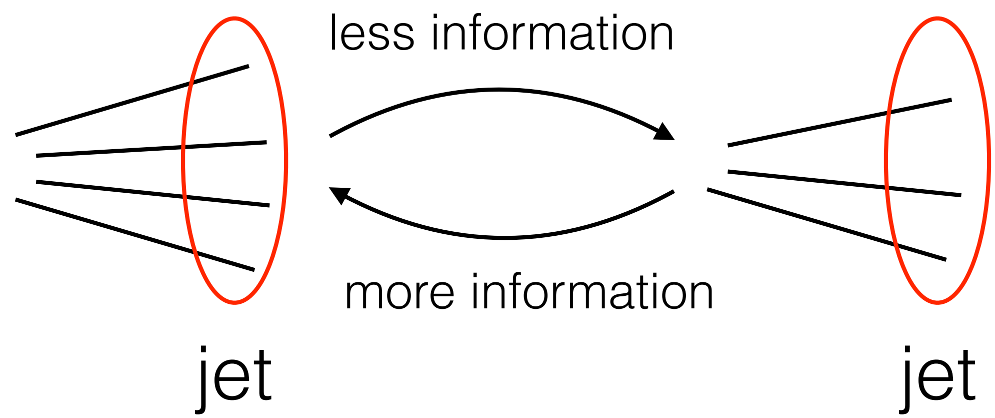
Simplifying the Discrimination Space

Human Learning on Jets

To make progress, use the guiding principles:

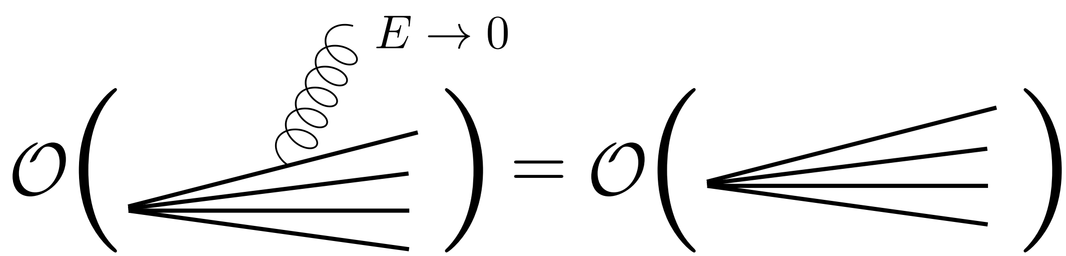
Systematic Improvability

Including more or less information in jet description is well-defined



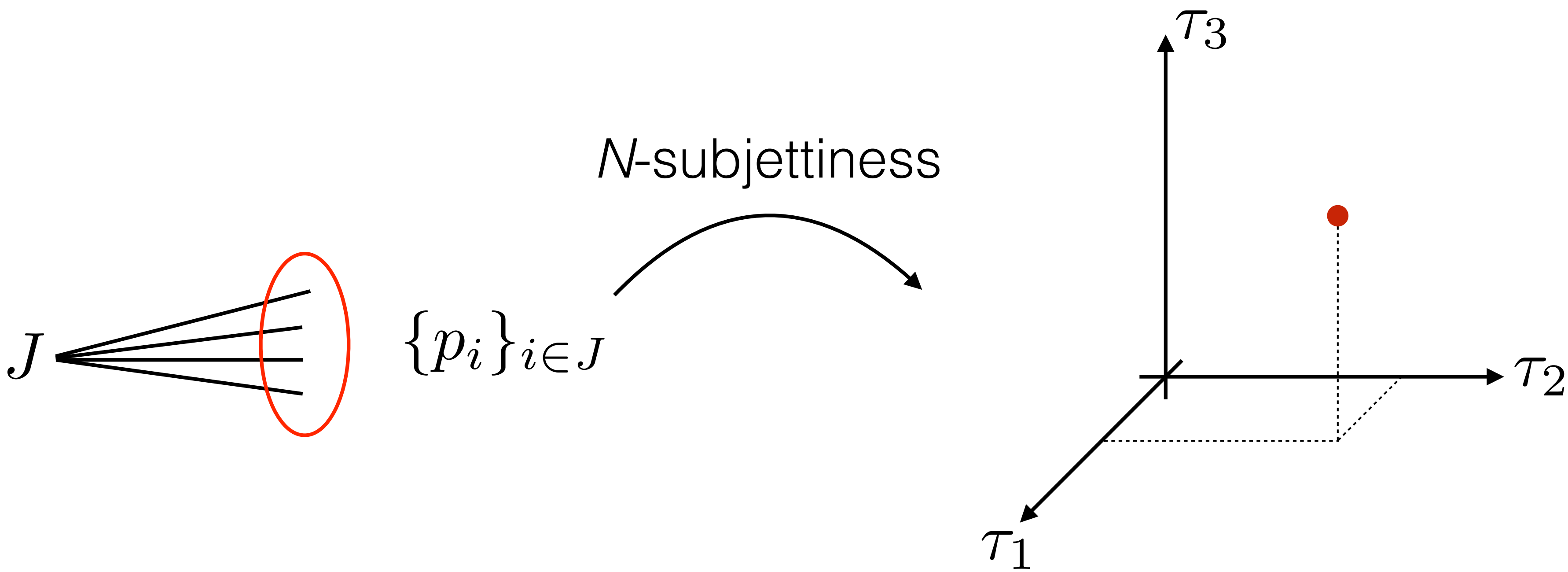
Infrared and Collinear (IRC) Safety

Ensures calculability in perturbation theory



Human Learning on Jets

N-subjettinesses and related observables accomplish this



$$\tau_N^{(\beta)} = \frac{1}{p_{TJ}} \sum_{i\in J} p_{Ti} \min\left\{R_{1i}^{\beta}, R_{2i}^{\beta}, \dots, R_{Ni}^{\beta}\right\}$$

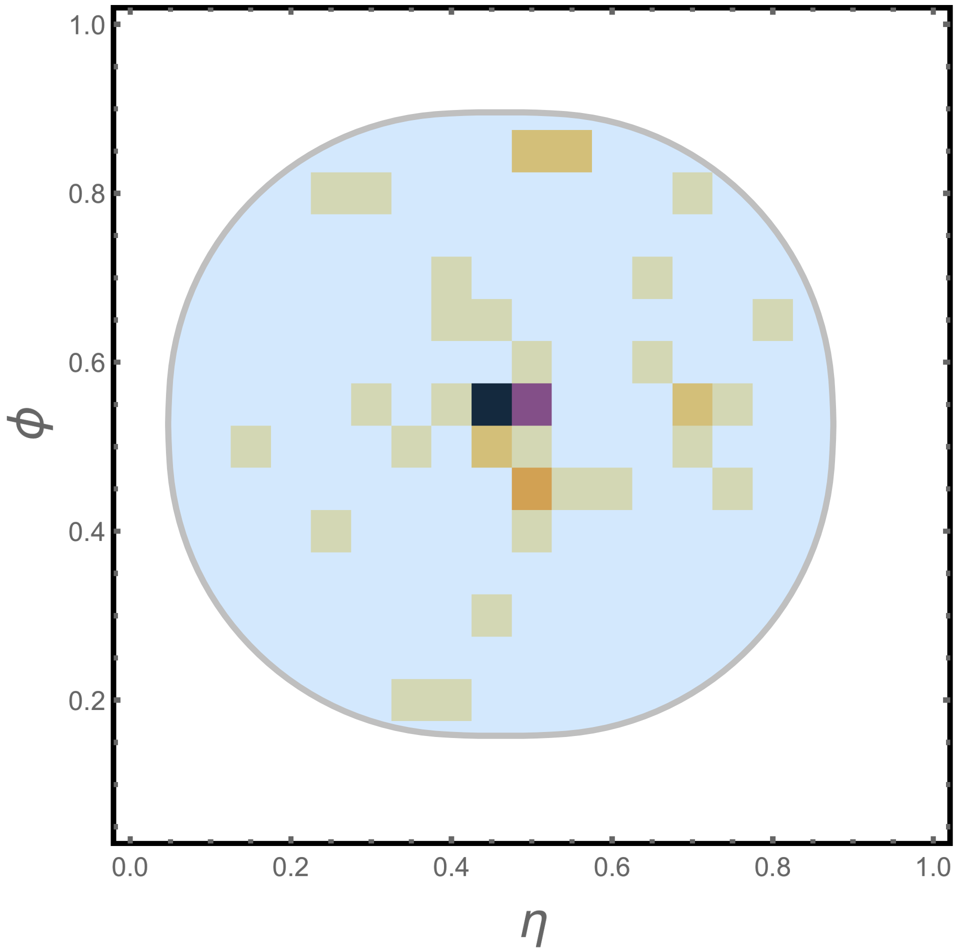
Sensitive to radiation of off *N* axes in the jet

Datta, AJL 2017

history:
Thaler, van Tilburg, 2010, 2011
Stewart, Tackmann, Waalewijn 2010
Brandt, Dahmen 1979
Wu, Zobernig 1979
Nachtmann, Reiter 1982

Human Learning on Jets

Systematically resolve more structure in the jet



Full Jet

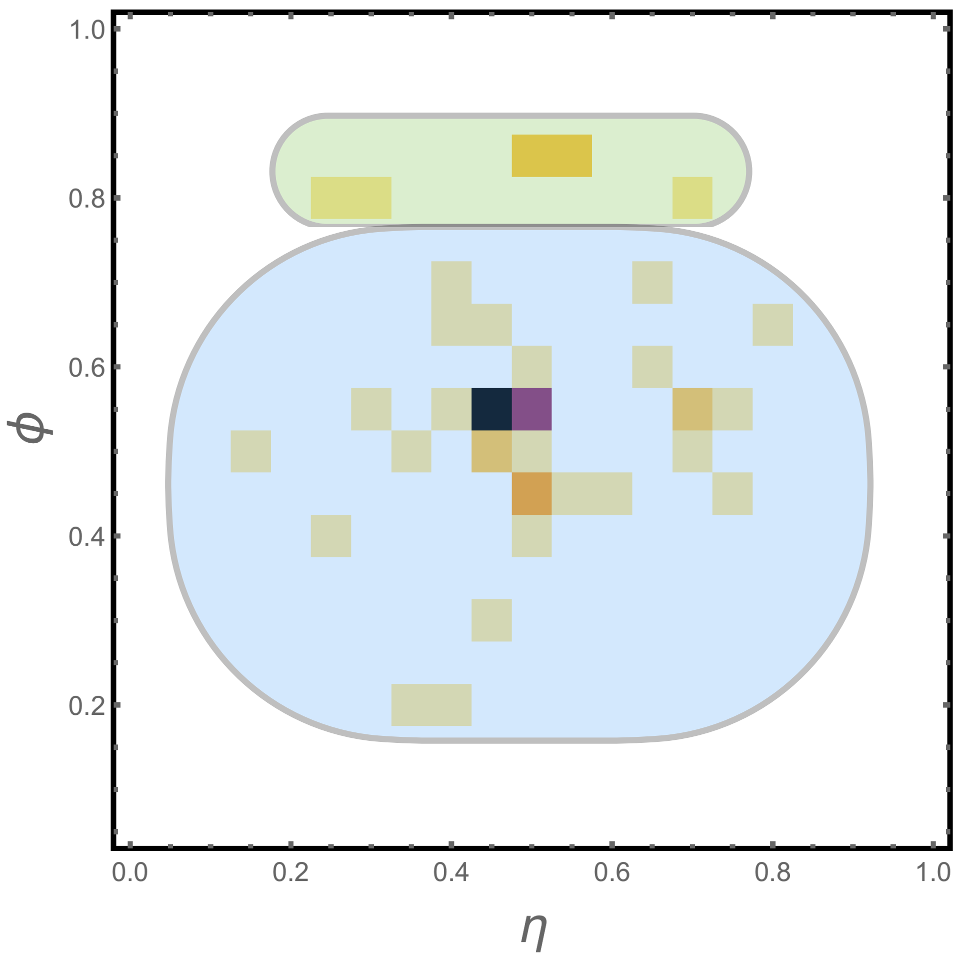
Net p_T, η, ϕ selected for

1 useful quantity:
jet invariant mass

Restrict m_J in a range about the mass of interest

Human Learning on Jets

Systematically resolve more structure in the jet



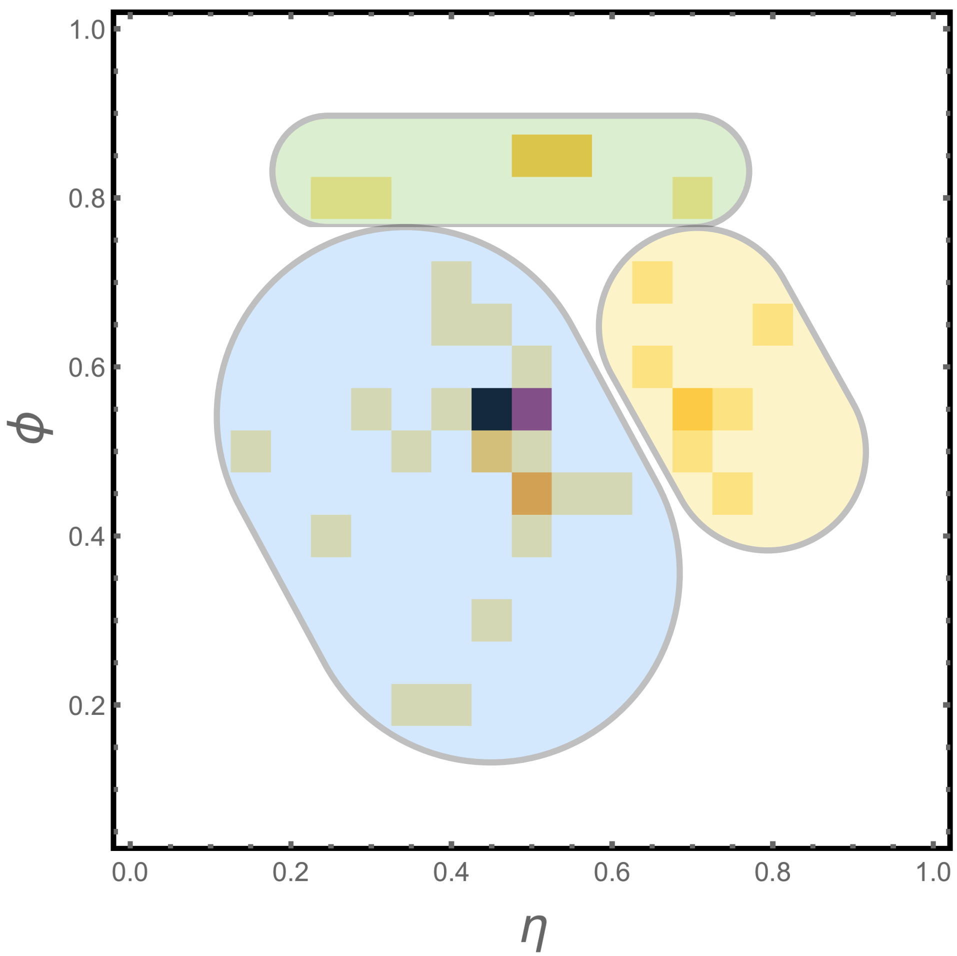
Two Subjets

Net p_T, η, ϕ, m_J selected for

2 useful quantities:
relative p_T fraction
relative angle

Human Learning on Jets

Systematically resolve more structure in the jet



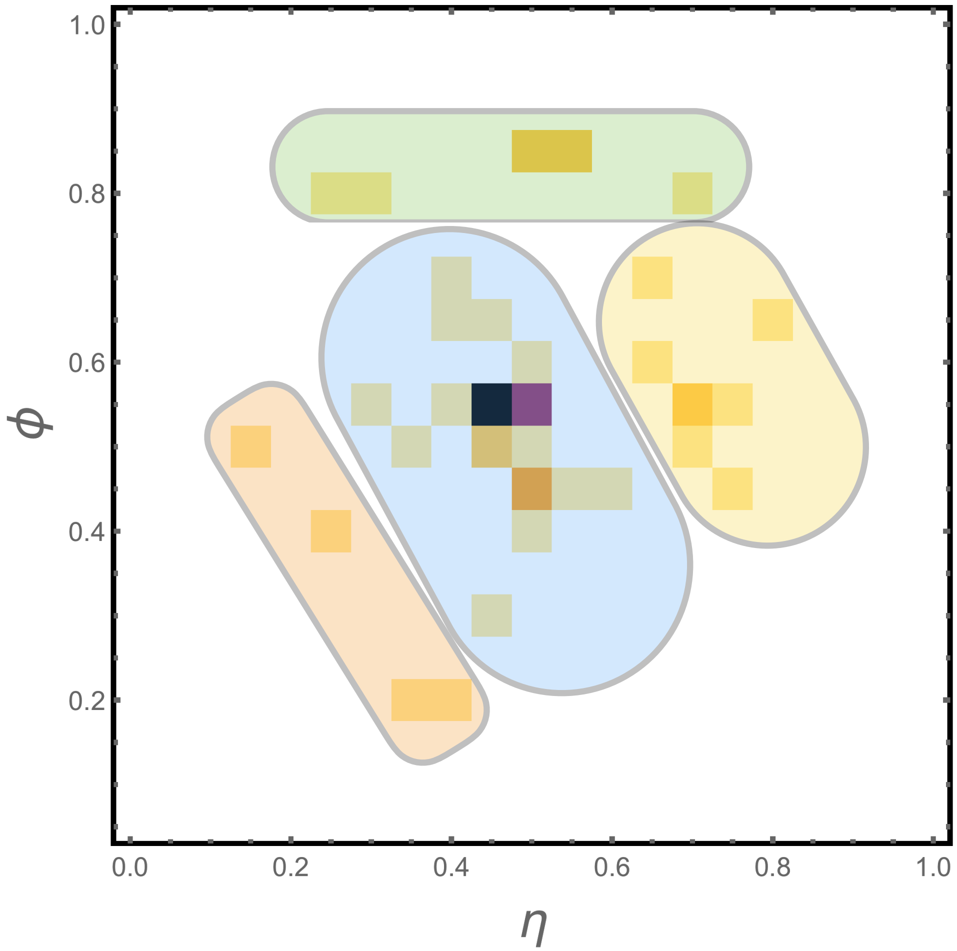
Three Subjets

Net p_T, η, ϕ, m_J selected for

5 useful quantities:
2 relative p_T fractions
3 relative angles

Human Learning on Jets

Systematically resolve more structure in the jet



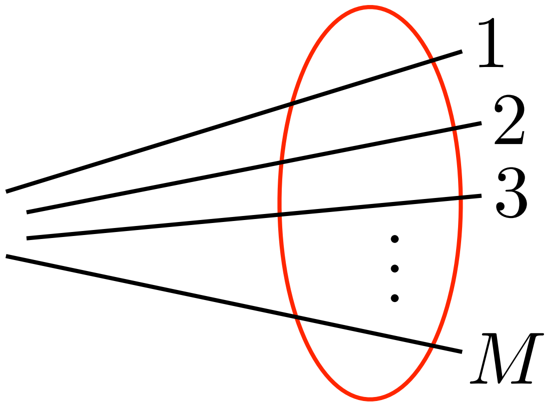
Four Subjets

Net p_T, η, ϕ, m_J selected for

8 useful quantities:
3 relative p_T fractions
5 relative angles

Can continue to resolve arbitrary structure

Human Learning on Jets

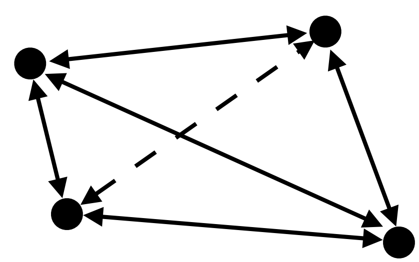


jet

Measure observables to resolve M-body phase space

$$\sigma \sim \int \underbrace{\prod_{i=1}^{M} \left[\frac{d^4 p_i}{(2\pi)^4} 2\pi \delta(p_i^2 - m_i^2) \right] \delta^{(4)} \left(Q - \sum_{i=1}^{M} p_i \right)}_{3M - 4 \text{ dimensional phase space}} |\mathcal{M}|^2$$

In general:
M - 1 relative p_T fractions
$2M$ - 3 relative angles



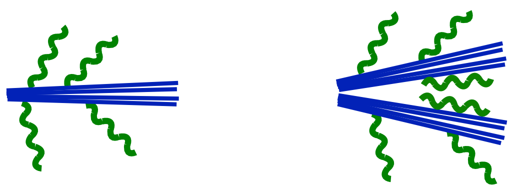
4 particle example

M-body Phase Space Machine Learning

Measure observables sensitive to
2-, 3-, 4-, 5-, and 6-body phase space + jet mass

Analyzed with a deep neural network on GPU

Calculated ROC curves for QCD vs. Z boson



If information is finite, should see saturation

Datta, AJL 2017

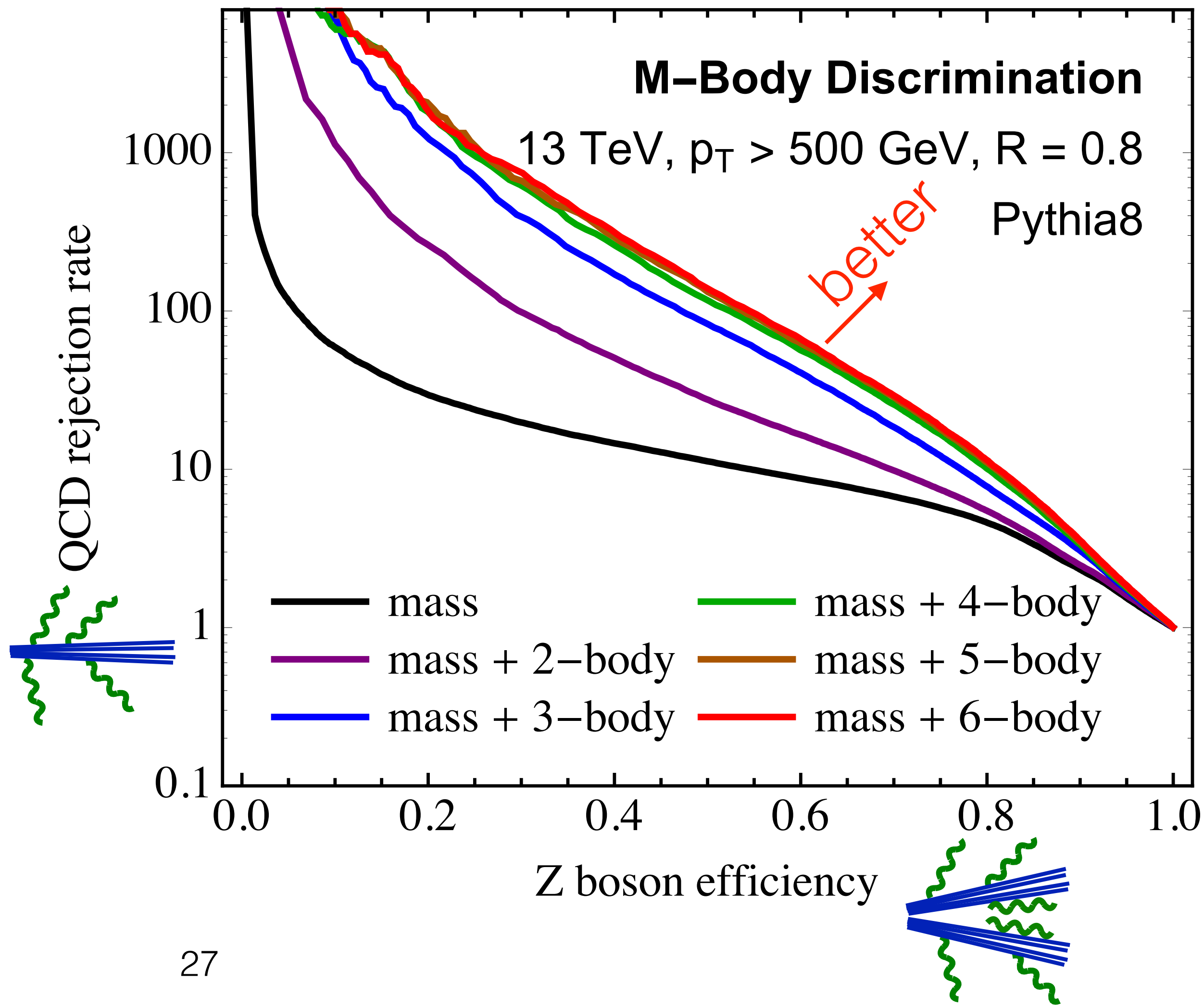
M-body Phase Space Machine Learning

Measure observables sensitive to
2-, 3-, 4-, 5-, and 6-body phase space + jet mass

Results:

Saturation observed at
4-body phase space!

4-body phase space
= 8 dimensional



Datta, AJL 2017

Insights into Quark vs. Gluon Discrimination

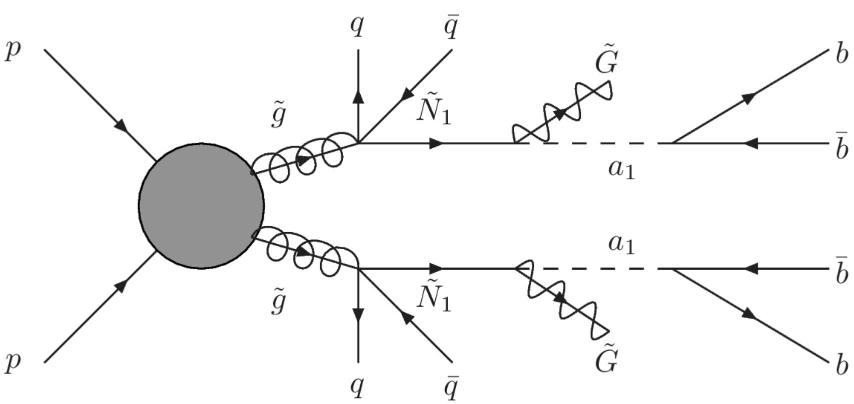
“The White Whale of Jet Physics”

-Jesse Thaler

Quark/Gluon jet discrimination impacts entire LHC program!

New Physics Searches

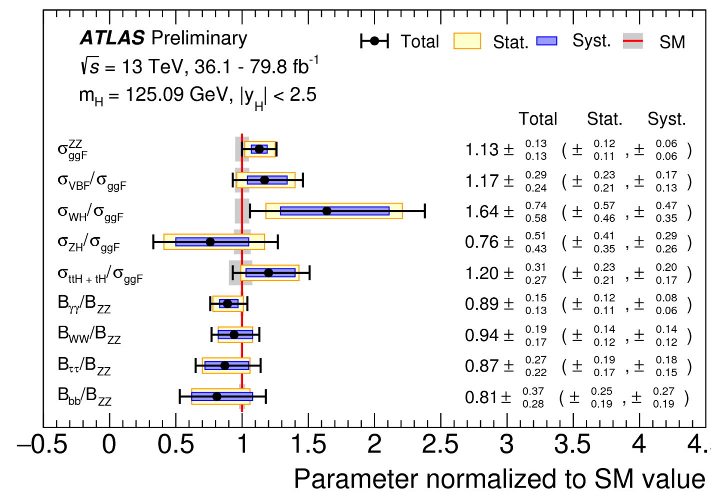
8 quark-jet event!



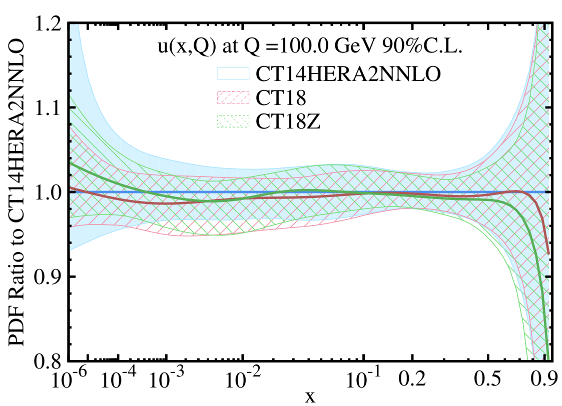
Allanach, Badziak, Cottin, Desai, Hugonie, Ziegler 2016

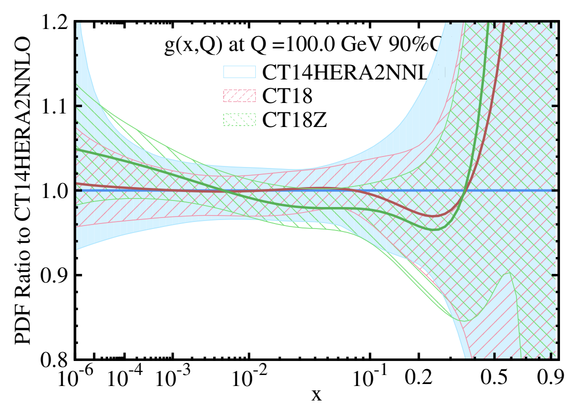
Higgs Physics

H → bb and H → gg are ~70% of total width



ATLAS-CONF-2018-031





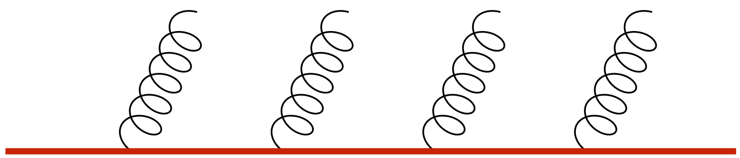
CTEQ-TEA Collaboration 2019

Gluon PDFs

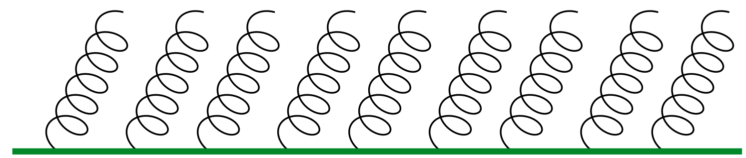
Large uncertainties at large x

Simple Picture of Quark/Gluon Jets

~Scale invariance of QCD =
Particle Production is Poisson Process



quark jet



gluon jet

Rate ∝ $C_F = 4/3$

Rate ∝ $C_A = 3$

Measure collection of *N*-subjettiness observables

Poisson Process implies

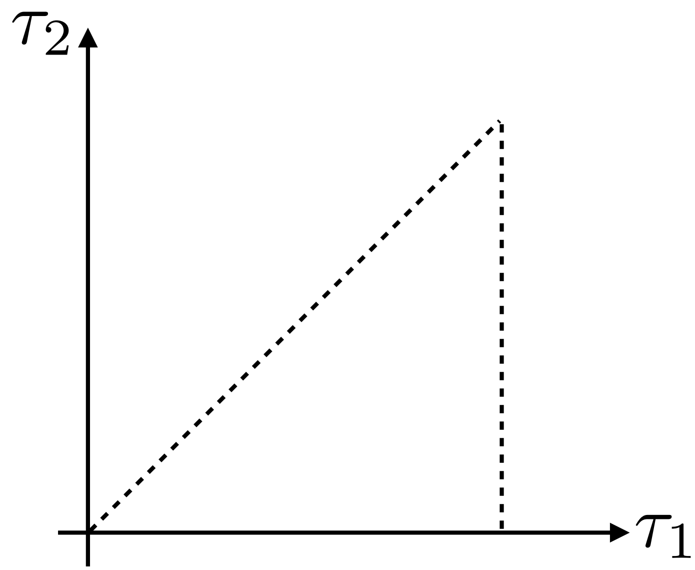
$$p_q(\{\tau_N\}) \sim e^{-C_F r(\{\tau_N\})} \qquad\qquad p_g(\{\tau_N\}) \sim e^{-C_A r(\{\tau_N\})}$$

IRC safety of *N*-subjettinesses implies

$r(\{\tau_N\}) \to \infty$ as any $\tau_N \to 0$

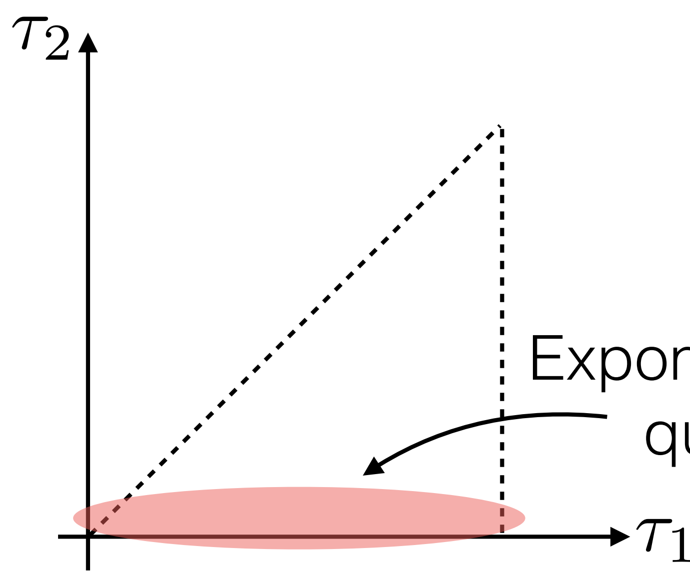
Where do jets live?

For visualization simplicity, just consider (τ_1,τ_2)

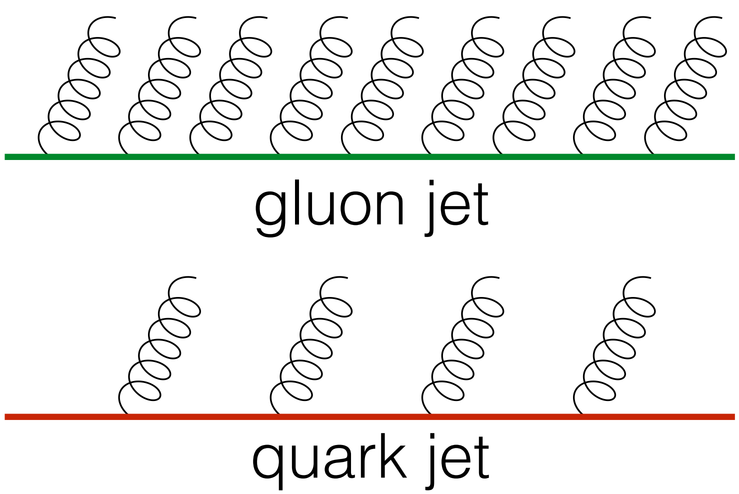


Where do jets live?

For visualization simplicity, just consider (τ_1,τ_2)



Exponentially more likely to be quark than gluon here



$\propto C_A$

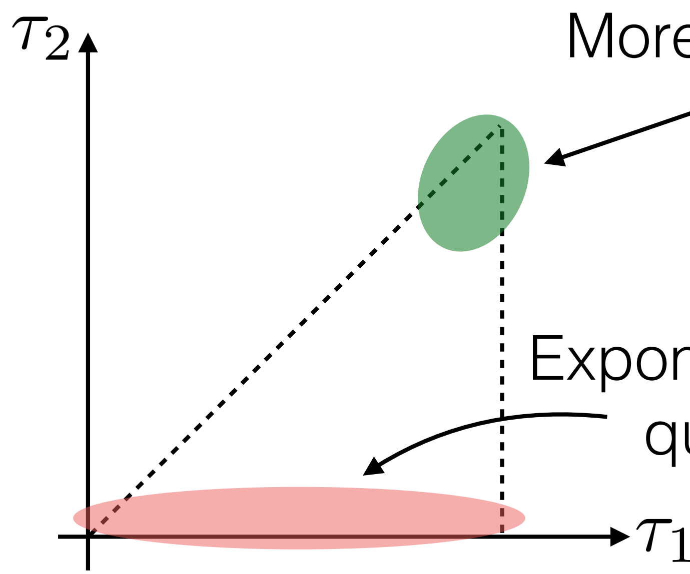
$\propto C_F < C_A$

Additivity then implies

$$\tau_N^{\text{gluon}} > \tau_N^{\text{quark}}$$

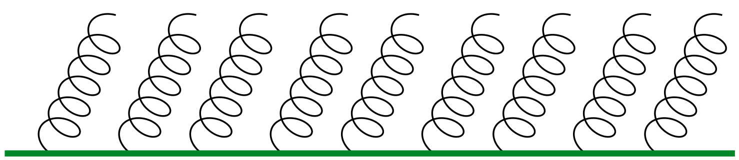
Where do jets live?

For visualization simplicity, just consider (τ_1, τ_2)



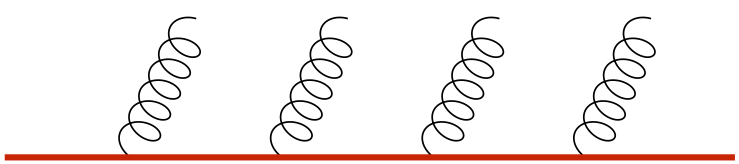
More likely to be gluon than quark here

Exponentially more likely to be quark than gluon here



gluon jet

$\propto C_A$



quark jet

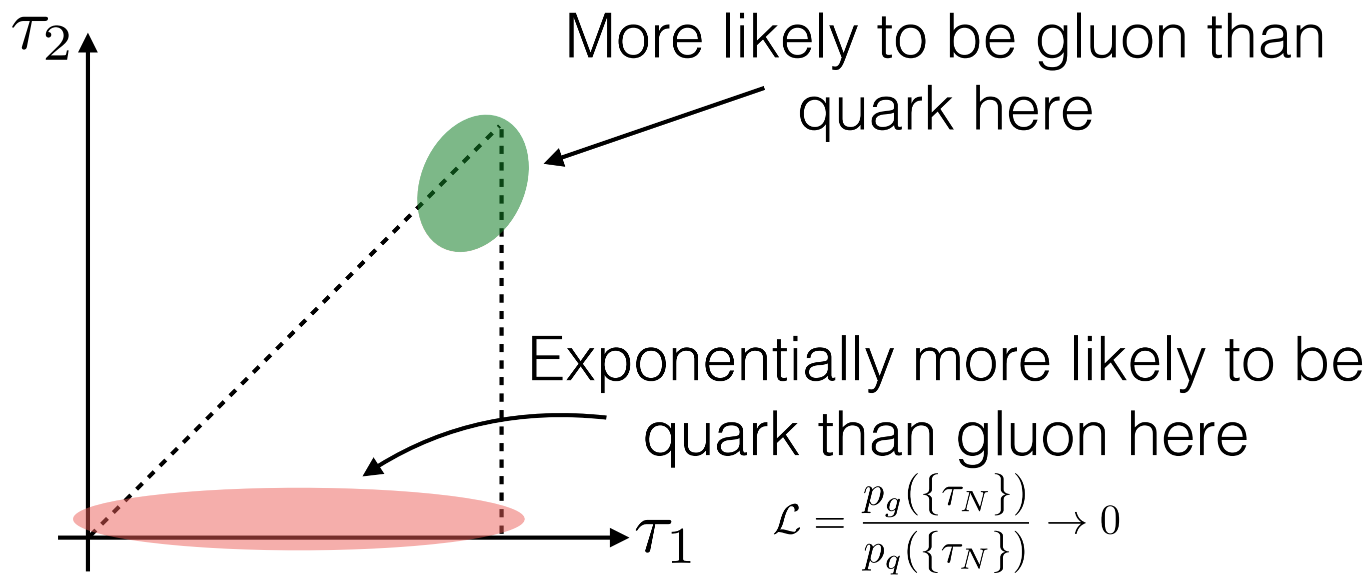
$\propto C_F < C_A$

Additivity then implies

$$\tau_N^{\text{gluon}} > \tau_N^{\text{quark}}$$

Quark vs. Gluon Likelihood

What are the properties of the likelihood ratio?



Ans: As any $\tau_N \to 0$, likelihood vanishes

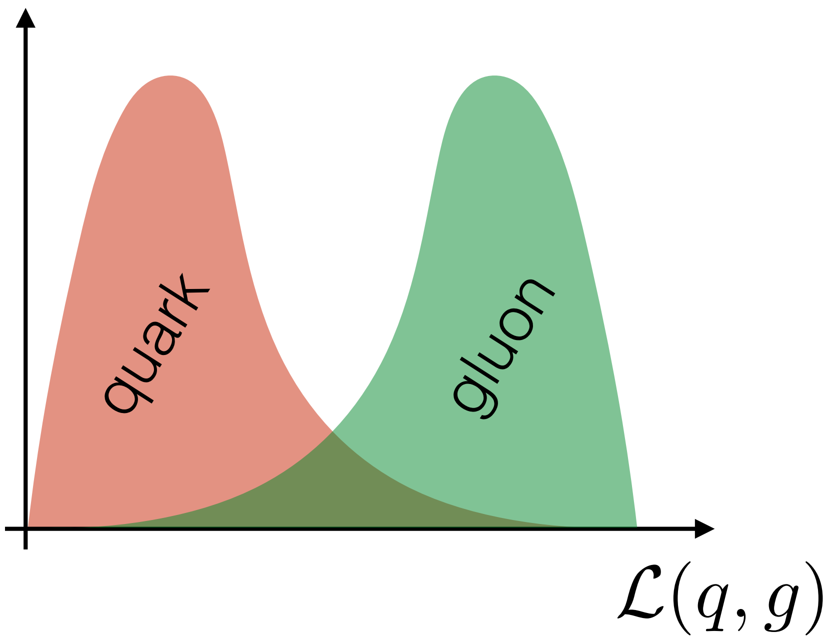
$\tau_N \to 0$ is the fixed-order divergent limit

Quark vs. Gluon likelihood ratio is IRC safe!

AJL, Metodiev 2019

IRC Safety of the Likelihood

Consequences



General

Independent of number of resolved emissions *N*

Non-vacuous

Pronginess discriminators (D_2, $\tau_{2,1}$, $\tau_{3,2}$, $\ldots$) all IRC unsafe

Soyez, Salam, Kim, Dutta, Cacciari 2012

AJL, Moult, Neill 2014

Practical

IRC safe observables are good q/g discriminants out of the box

good τ_N discrimination well-known

Gallicchio, Schwartz 2012

Caveat Emptor

Does *not* mean that likelihood can be calculated at fixed-order

Conclusions

Huge LHC datasets motivate machine learning

One approach: throw everything into a black box

Systematic organization dramatically simplifies description

Don't force the machine to work so hard!
Theory insight can get closer to Universal Approx. Thm.

Thanks!

https://xkcd.com/1838/

Don't forget that *you* are a domain expert!