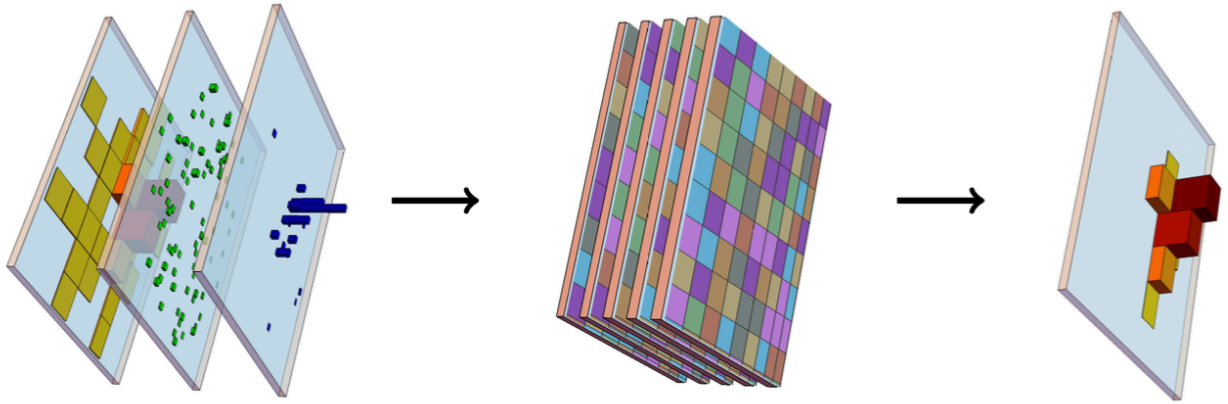
Jet Physics with Machine Learning

Hammers and Nails

August 1, 2019

Matthew Schwartz

Harvard University

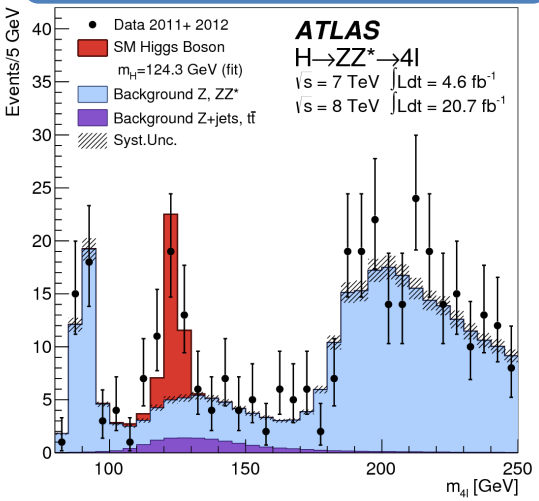


Outline

- Introduction to jet physics
- Traditional approach vs ML approach
- ML highlights from BOOST 2019

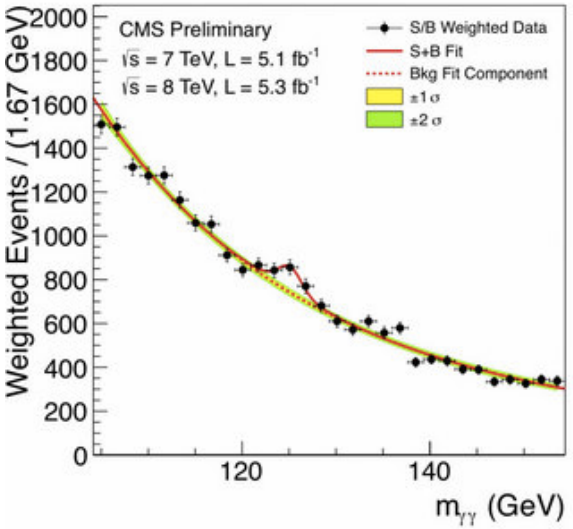
Higgs discovered 2012

$h \to e^- e^+ \mu^- \mu^+$, 0.01%



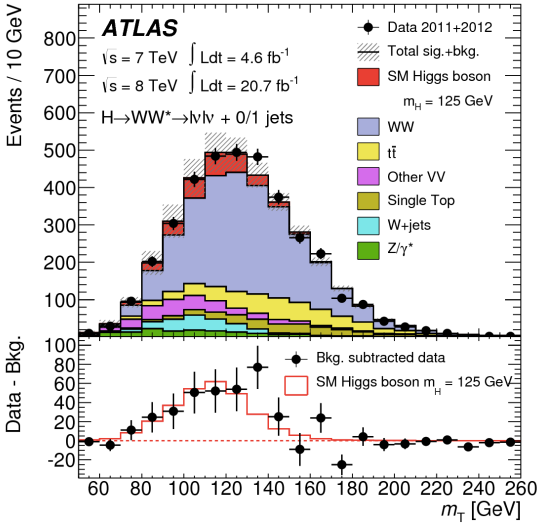
4 leptons = Golden Channel
Clear bump
Large signal/background
0.01% of Higgs decays

$h \to \gamma\gamma$, 0.1%



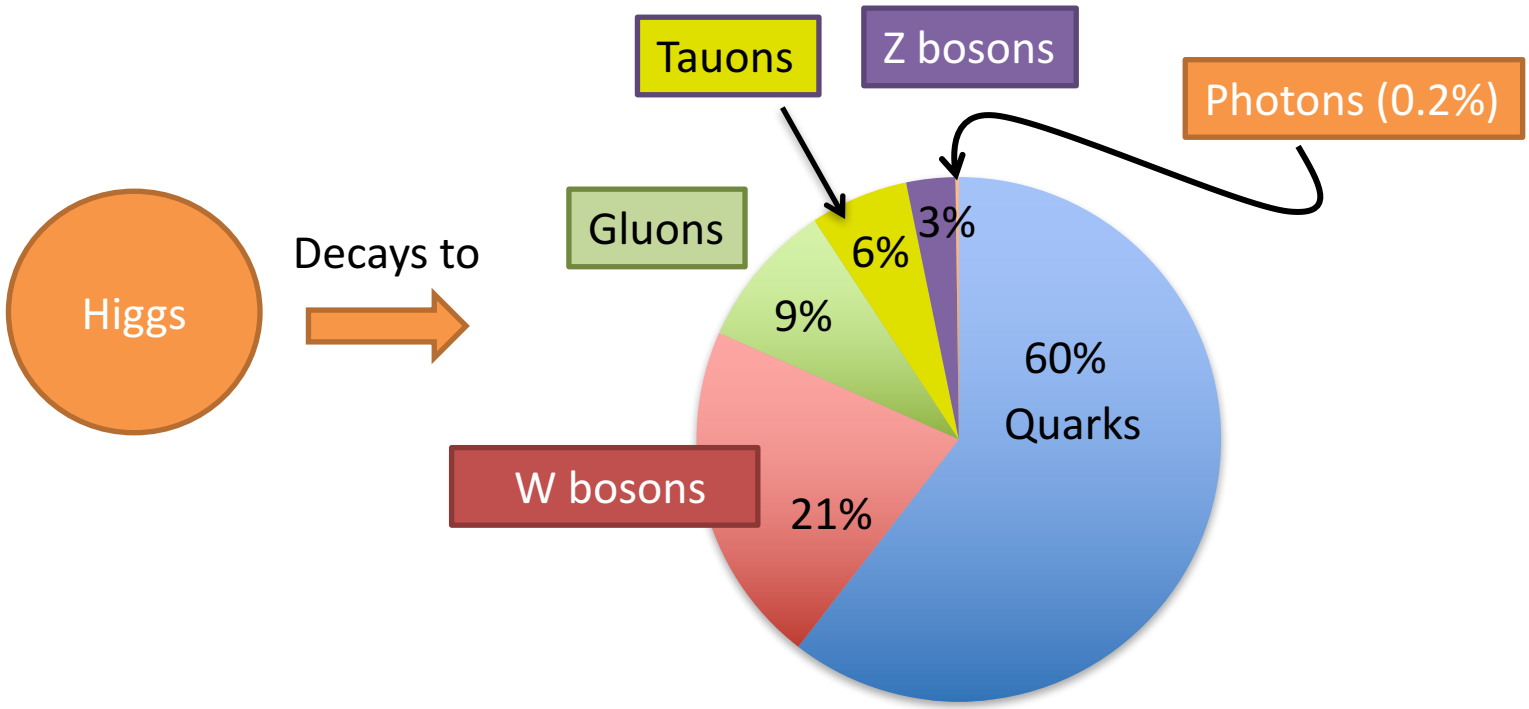
Photons
Not as clear bump
Small signal/background
0.2% of Higgs decays

$h \to e^- \mu^+ \bar{\nu}\nu$, 1%

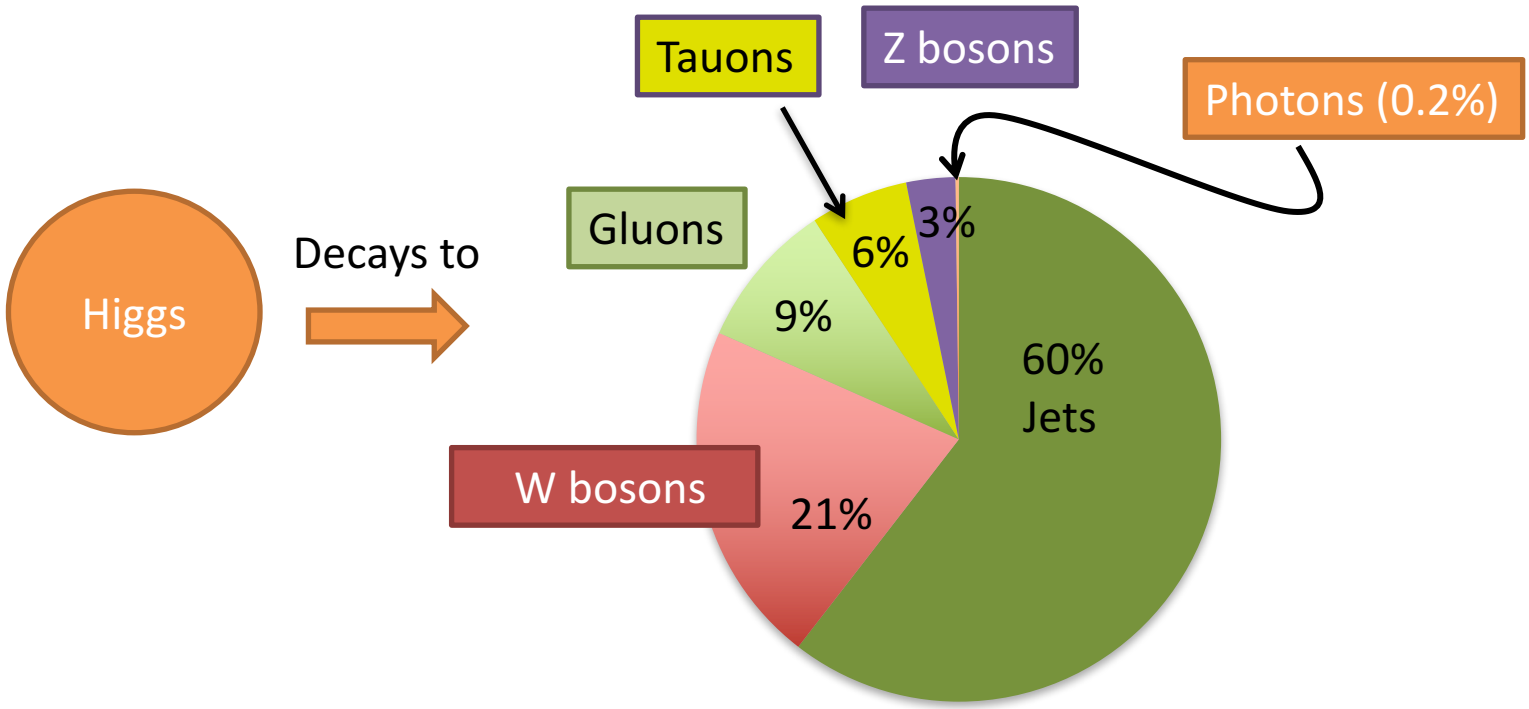


WW
No bump
Need backgrounds
0.8% of Higgs decays

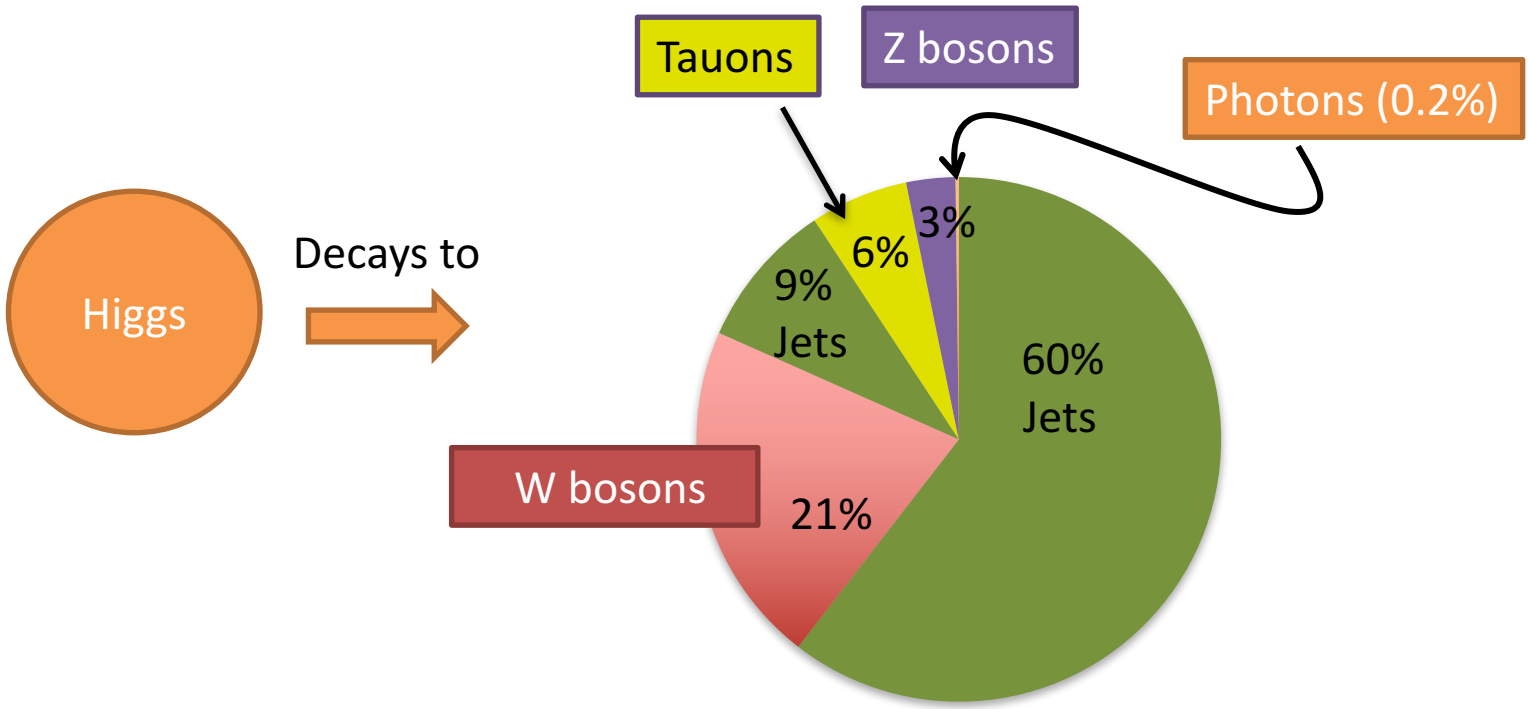
Higgs boson decay modes



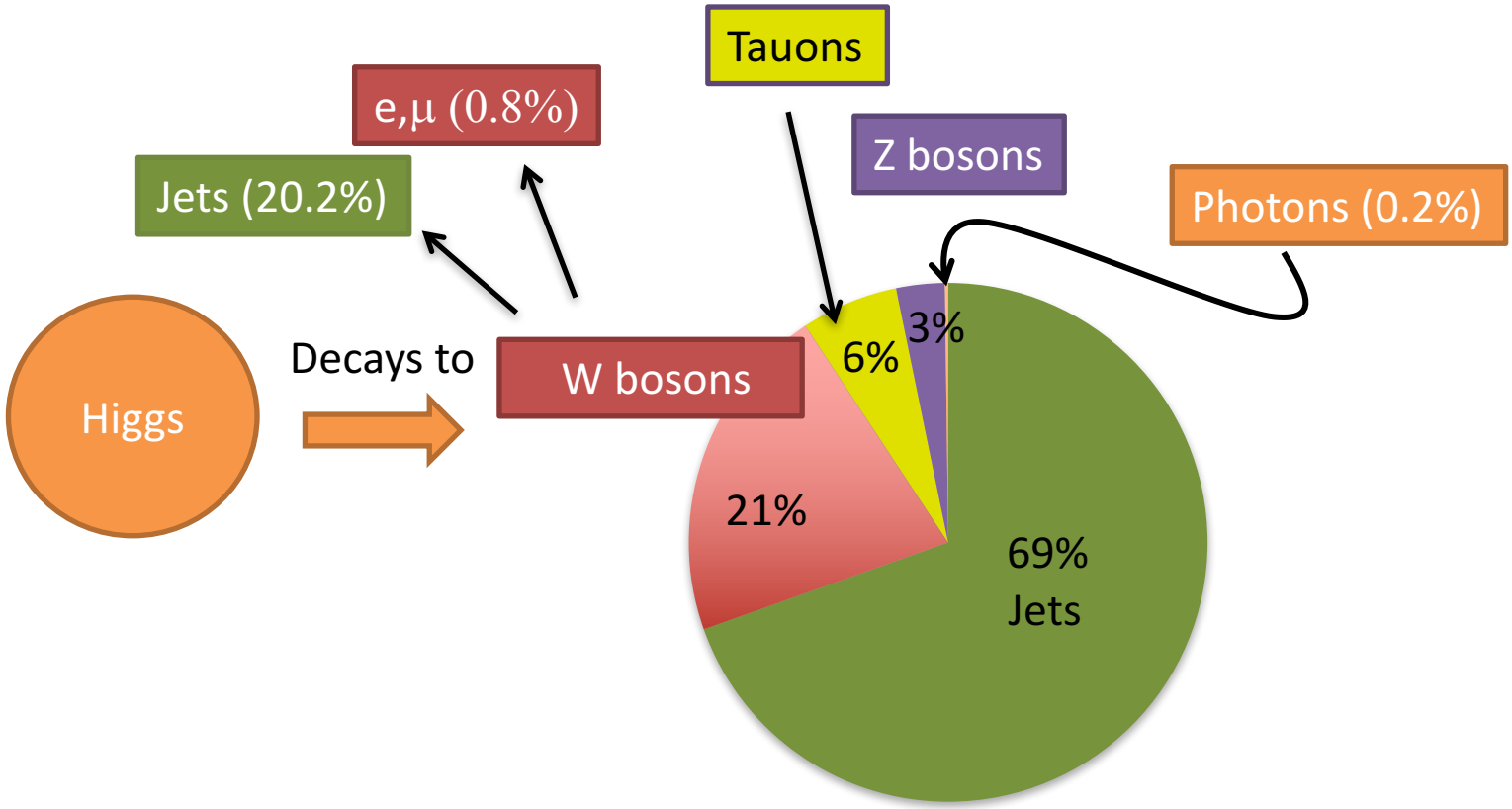
Higgs boson decay modes



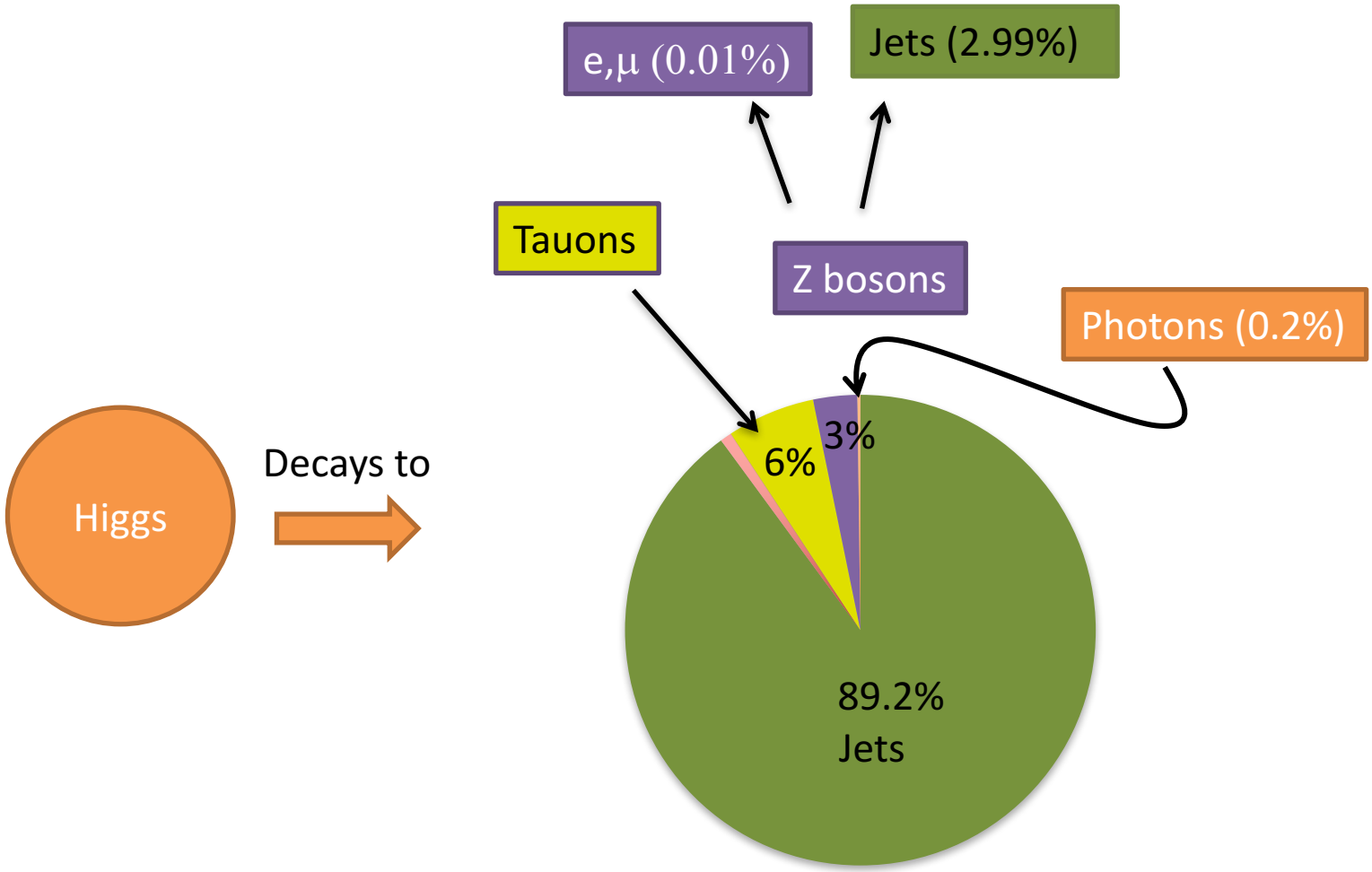
Higgs boson decay modes



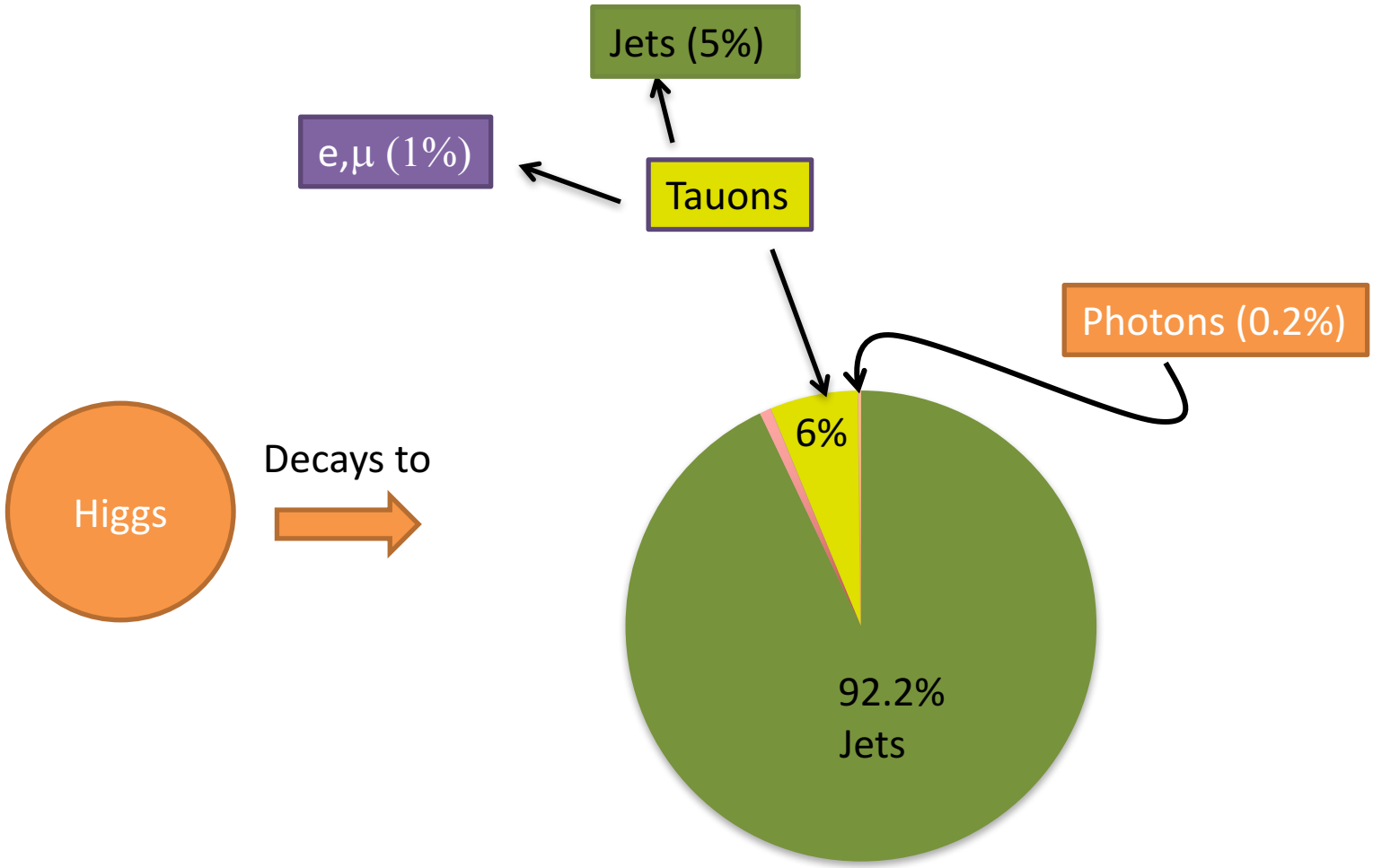
Higgs boson decay modes



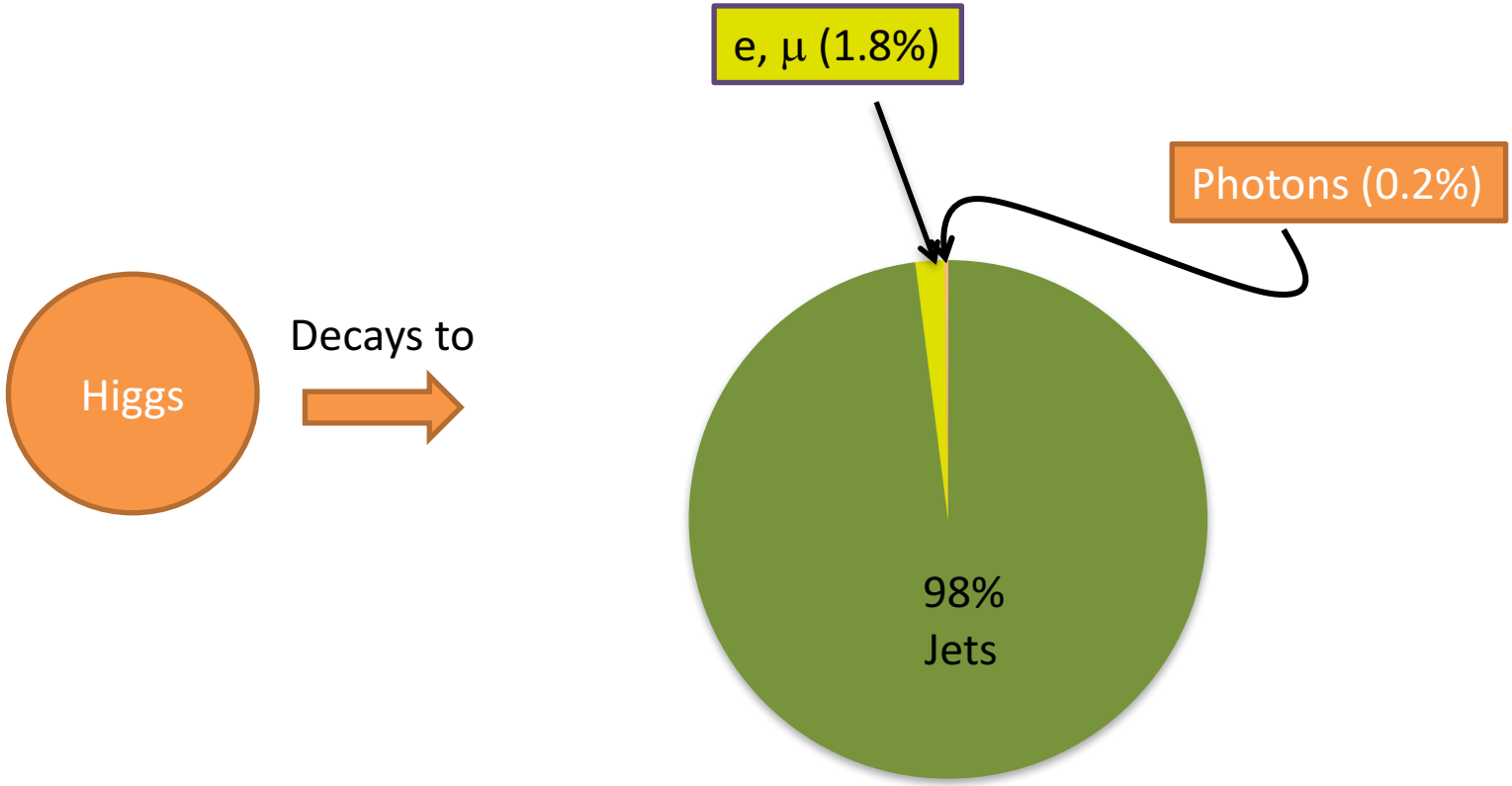
Higgs boson decay modes



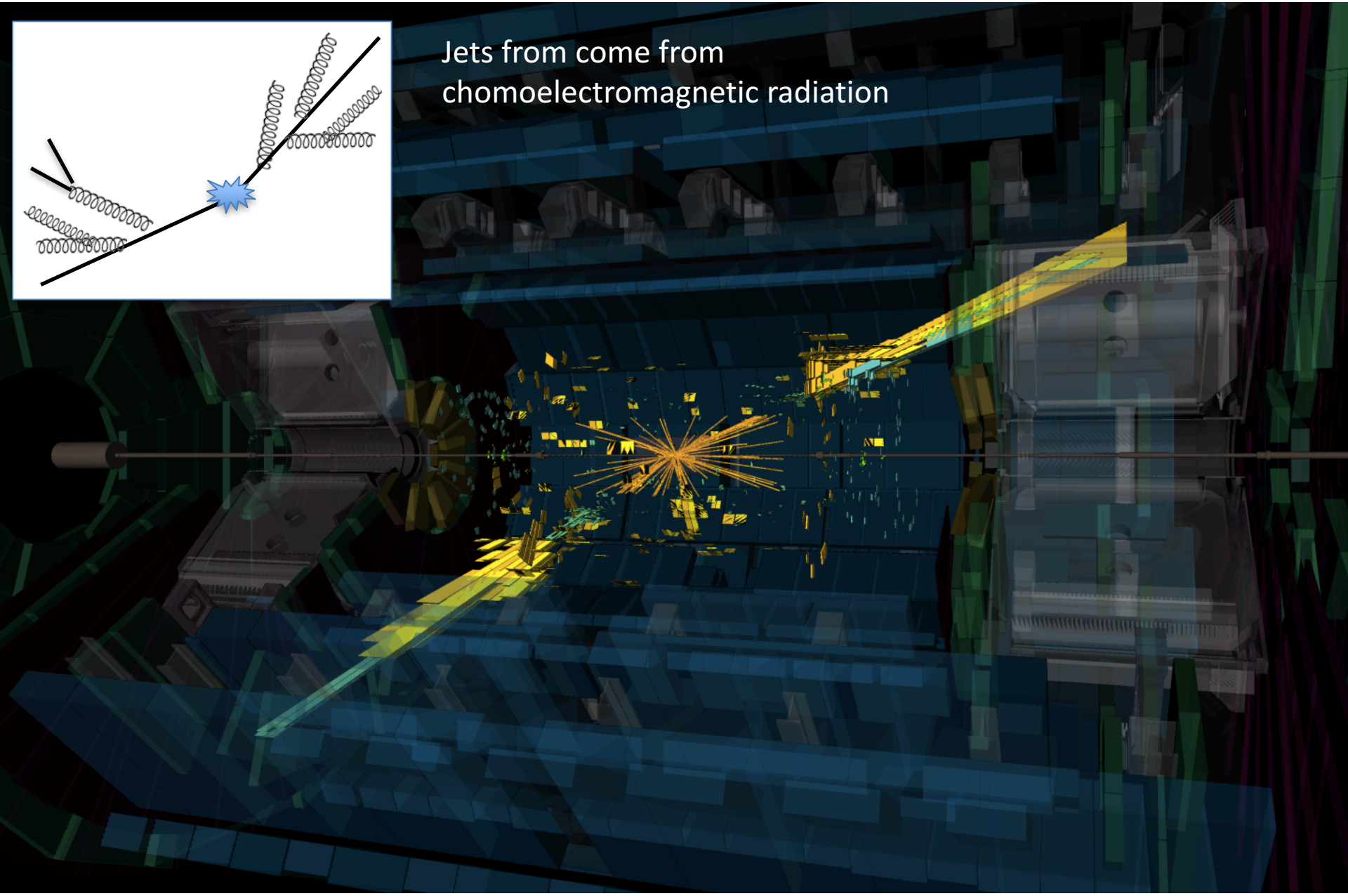
Higgs boson decay modes



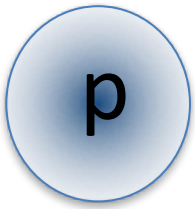
Higgs boson decay modes

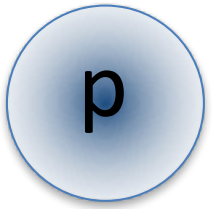


- Higgs discovery involved just these special decays

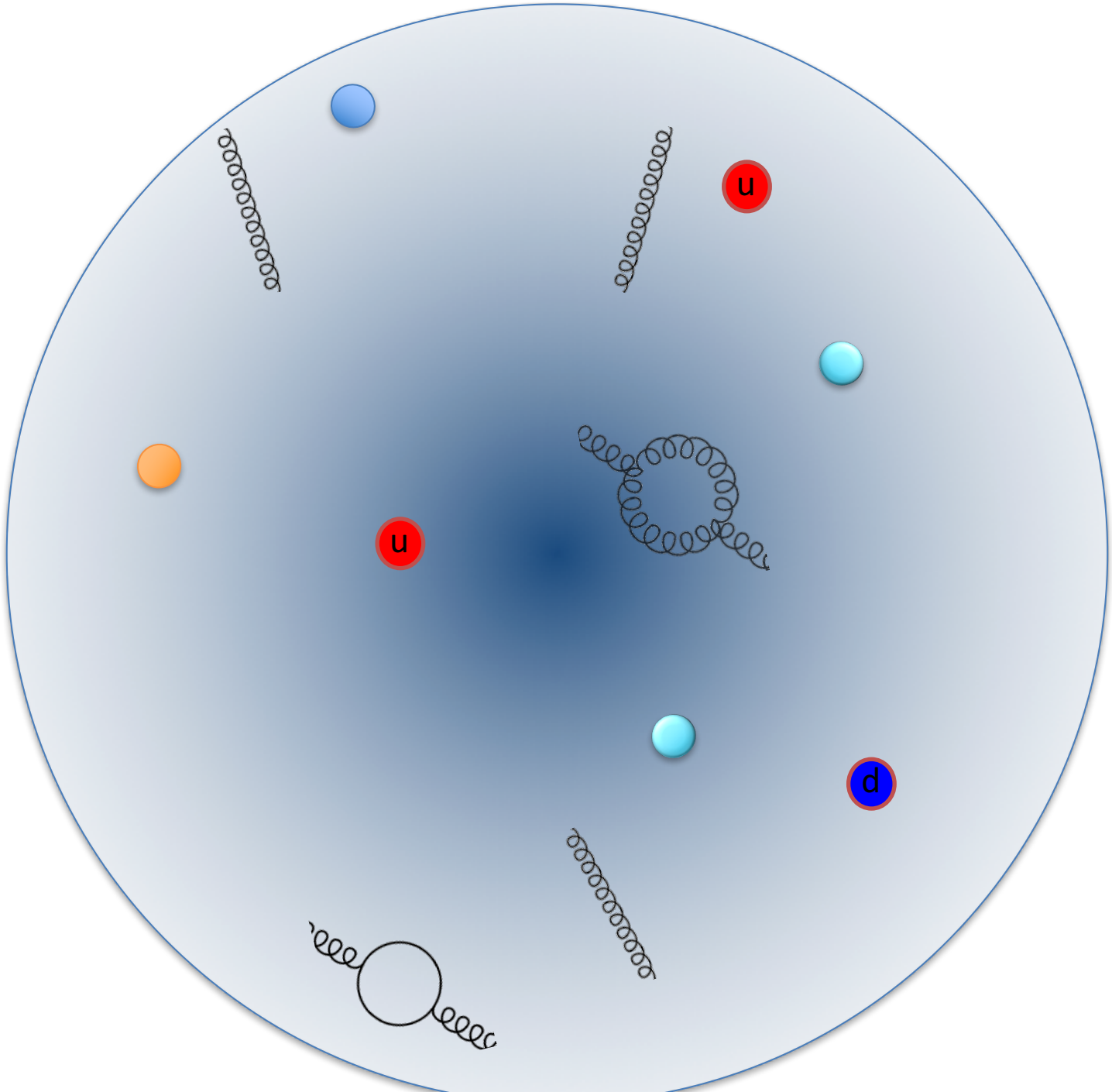

Jets from come from
chomoelectromagnetic radiation

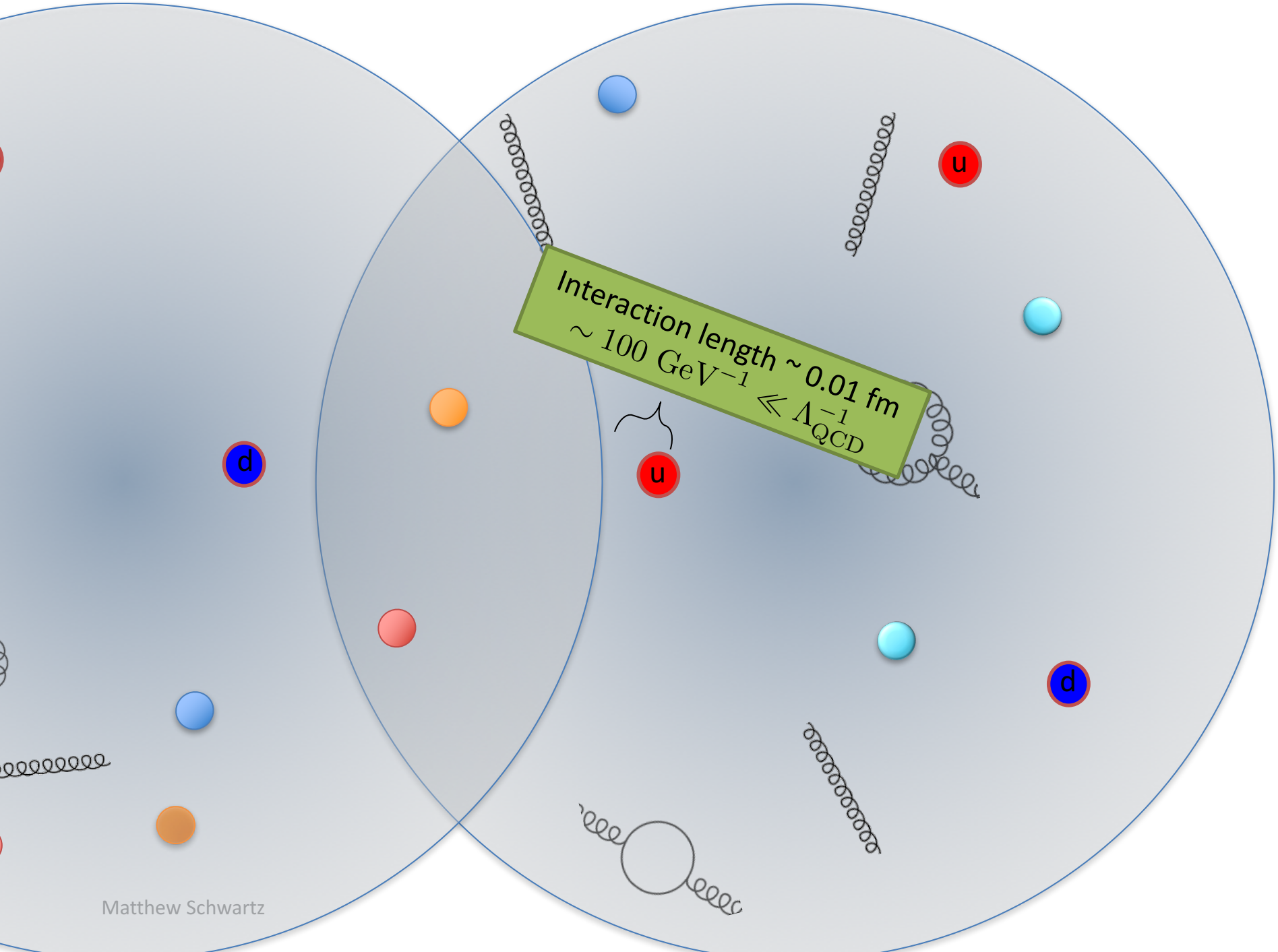
Colliding protons

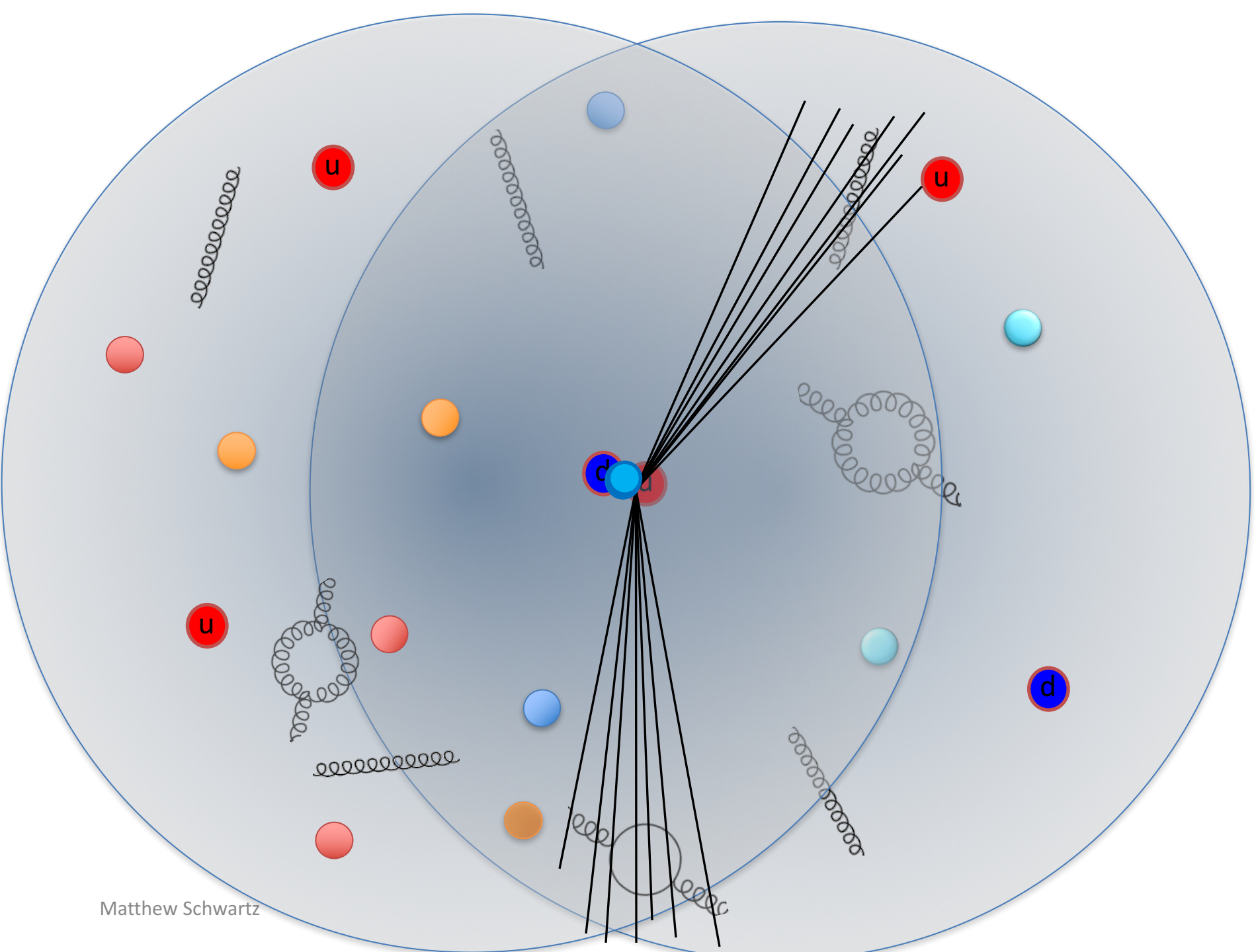




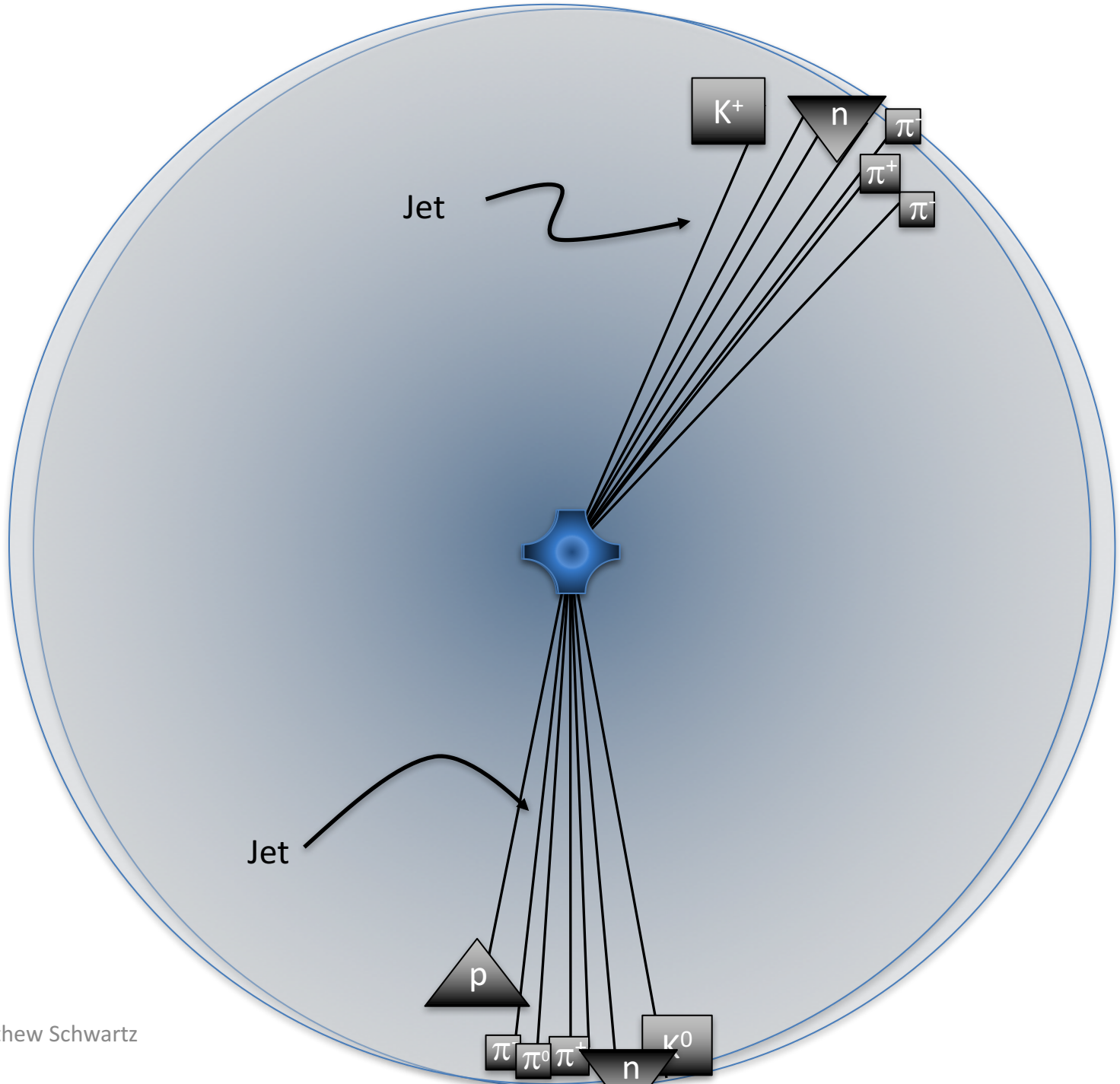
Size is
1 fm ~ Λ_{QCD}^{-1}

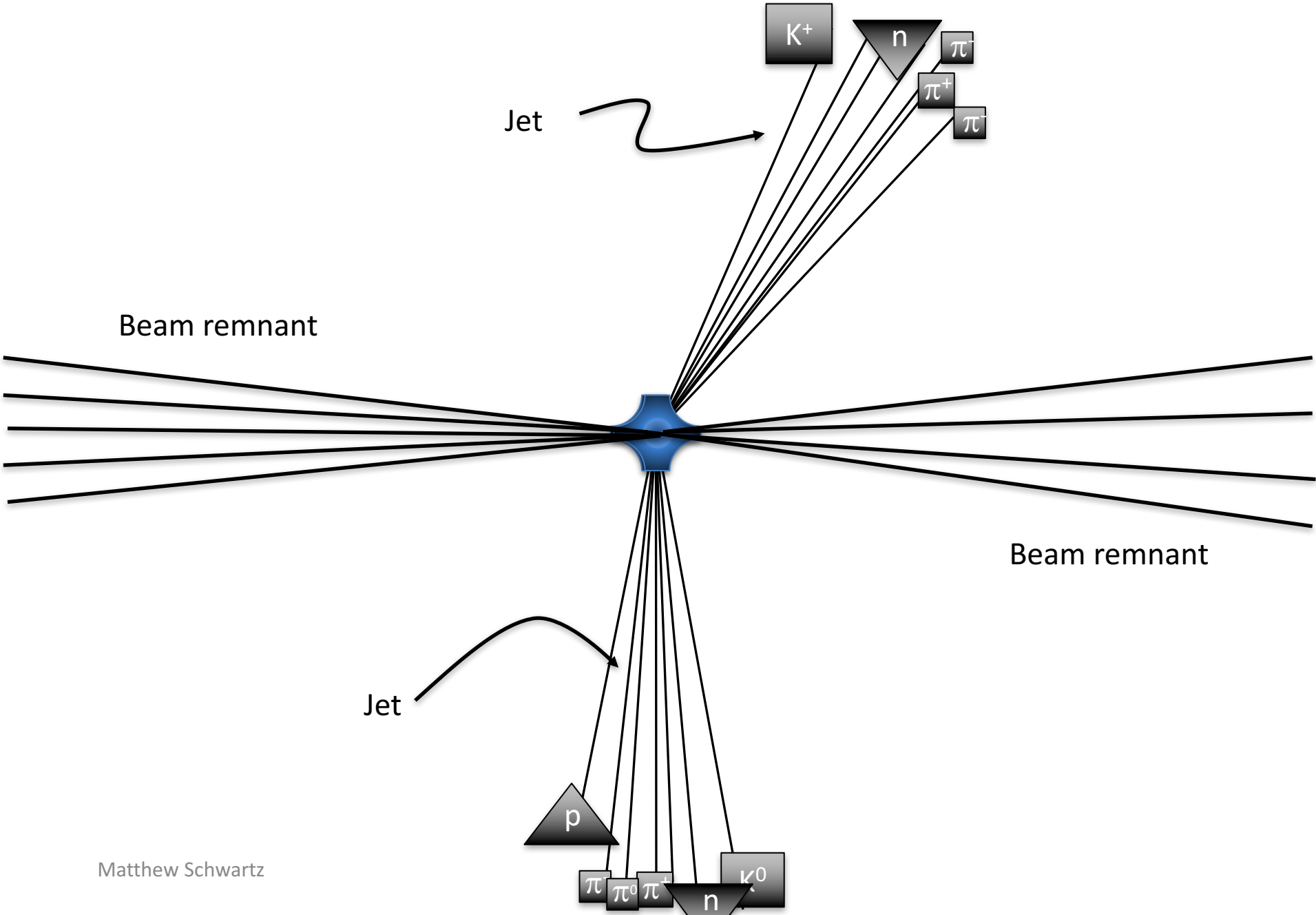


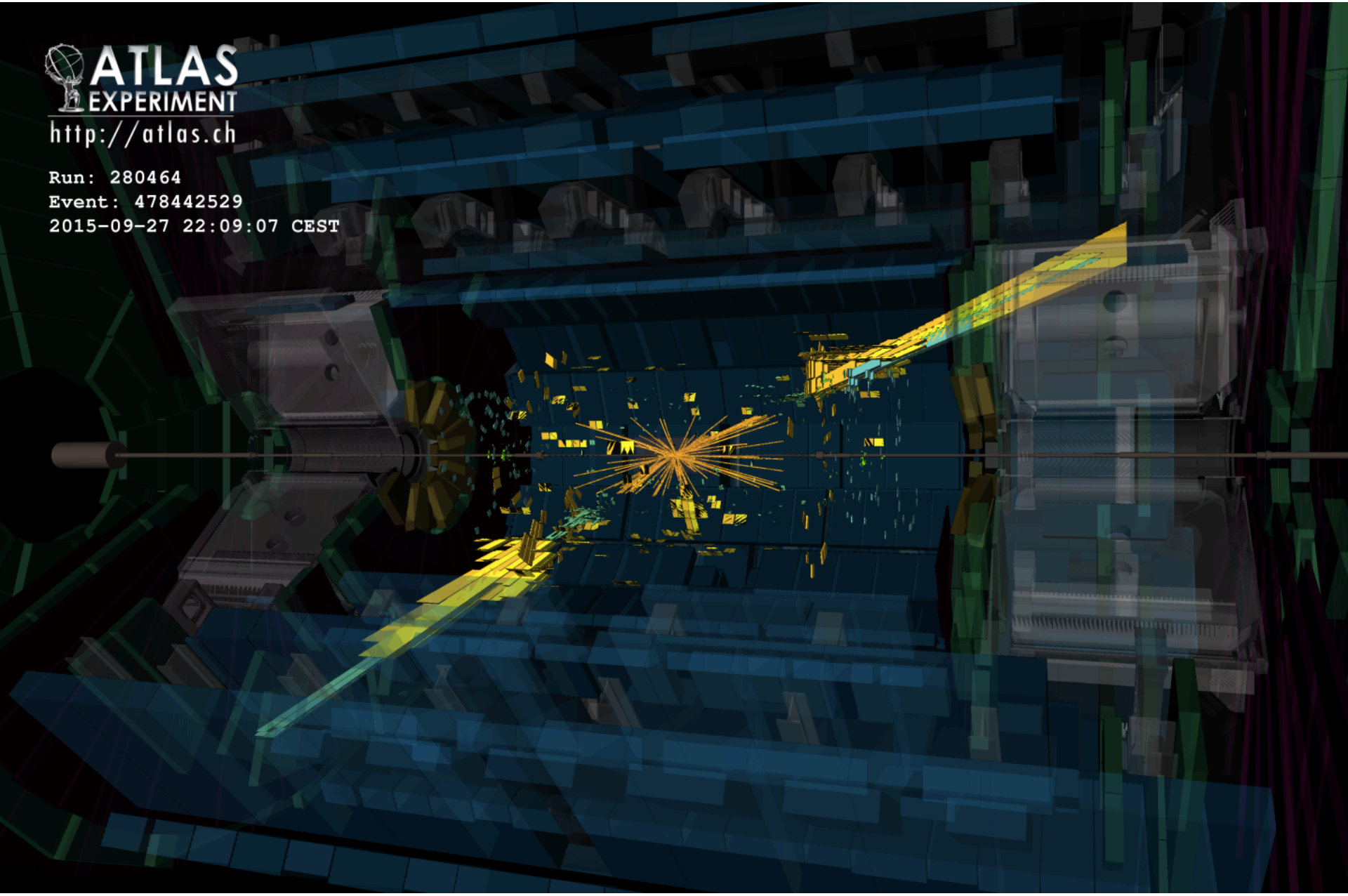

u
d
u
d
Interaction length ~ 0.01 fm
$\sim 100\ \mathrm{GeV}^{-1} \ll \Lambda_{\mathrm{QCD}}^{-1}$
Matthew Schwartz



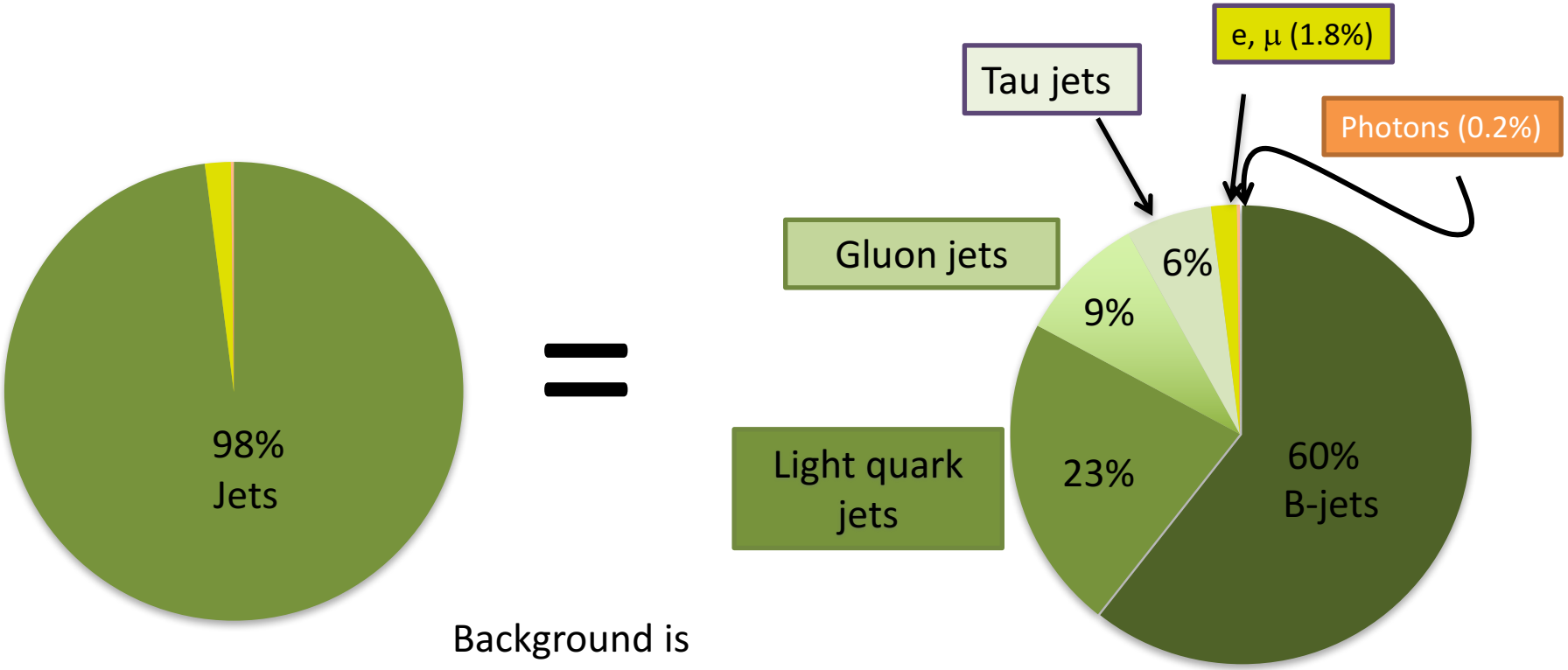
Matthew Schwartz


K+
n
π-
π+
π-
Jet
Jet
p
π
π0
π+
n
K0


K+
n
π-
π+
π-
Jet
Beam remnant
Beam remnant
Jet
p
π
π0
π+
n
K0


ATLAS
EXPERIMENT
http://atlas.ch
Run: 280464
Event: 478442529
2015-09-27 22:09:07 CEST

Not all jets are created equal

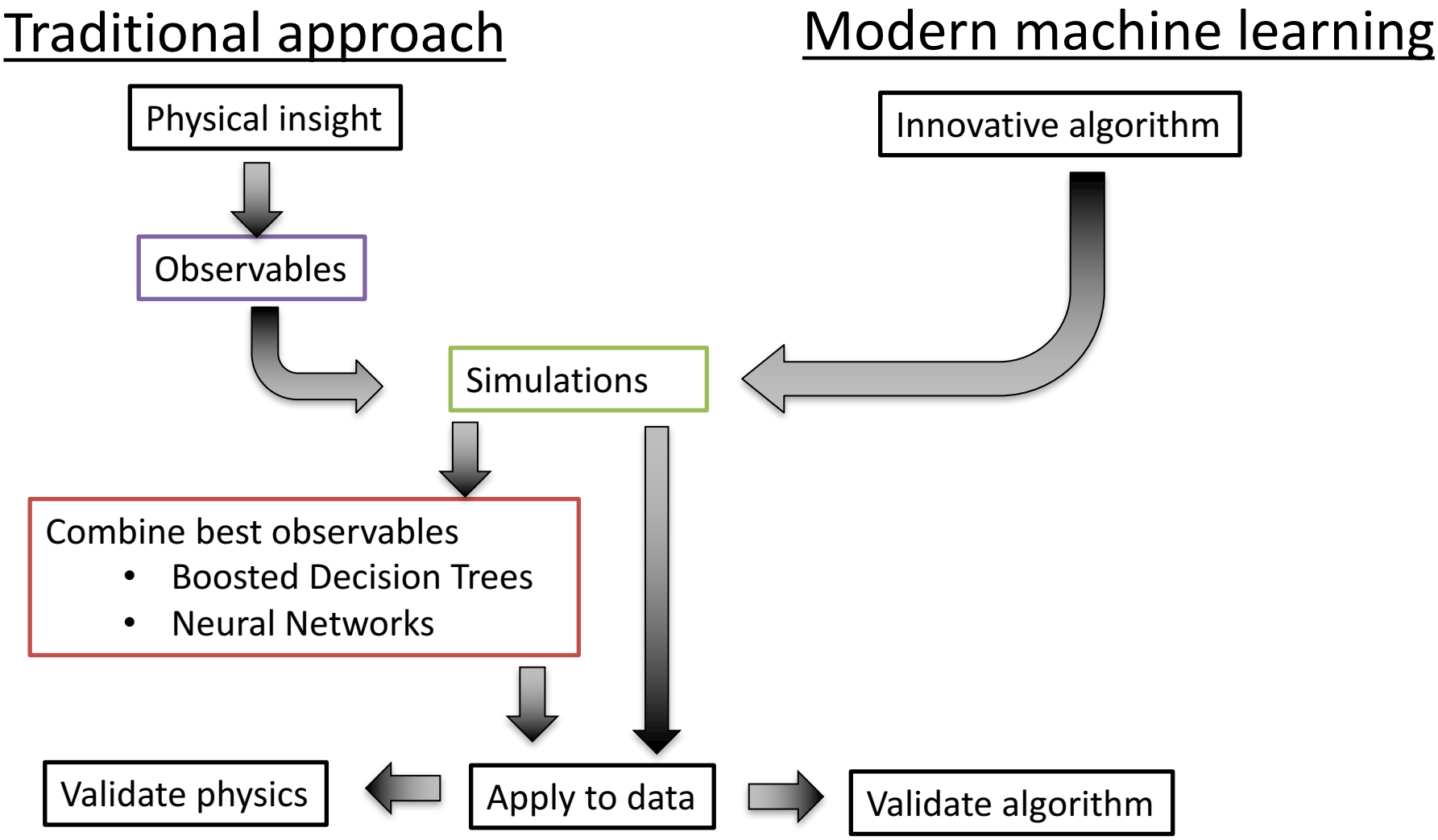


Background is

- 80 billion gluon jets
- 10 billion light quark jets

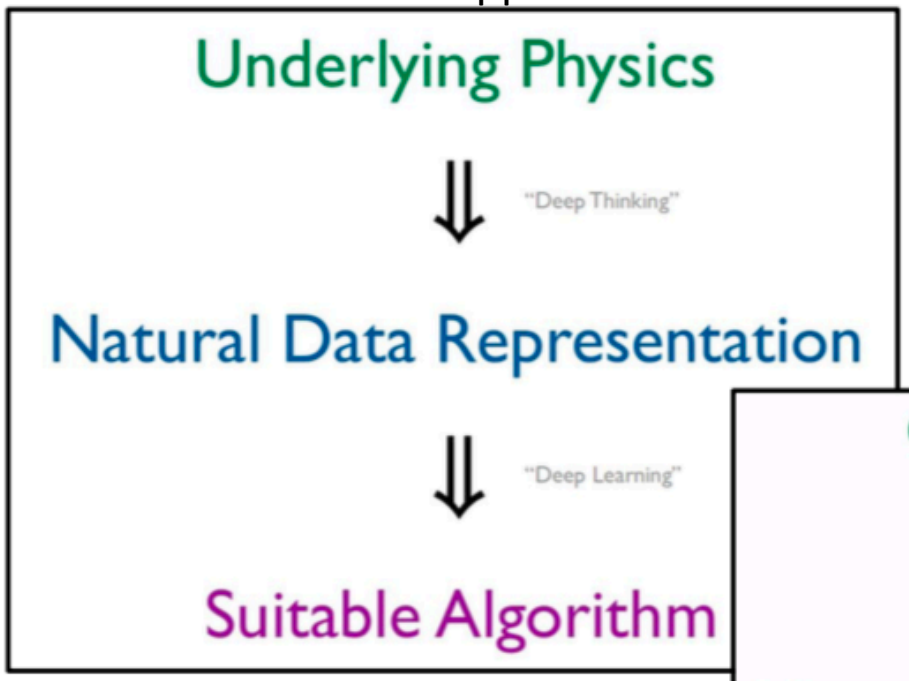
How can we tell all these different jet types apart?

Modern Machine Learning for Particle Physics

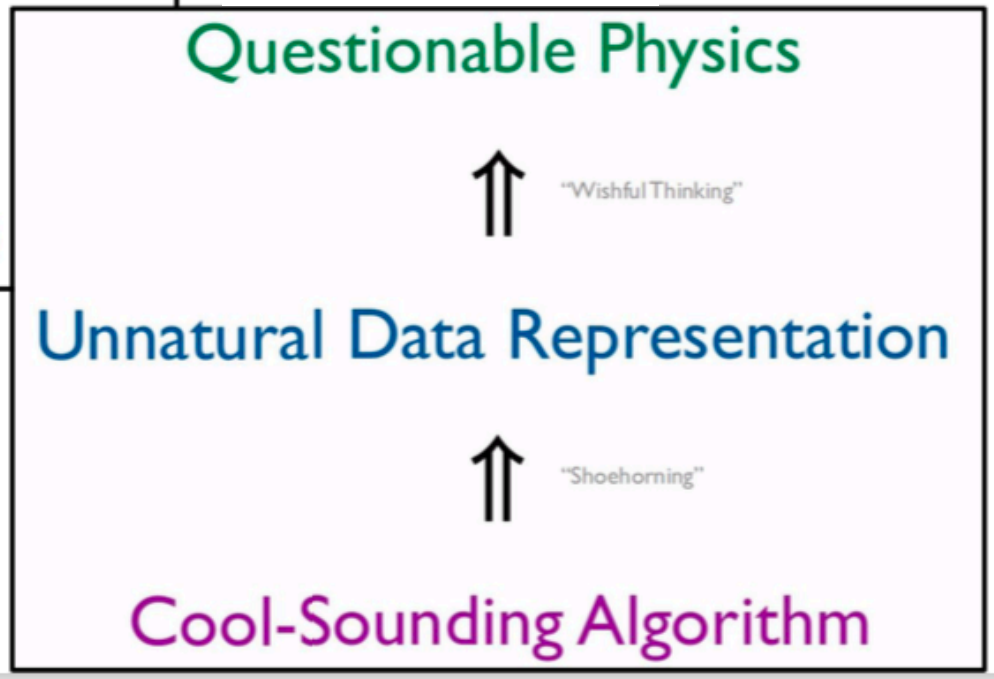


From Jesse Thaler's Boost summary 2018

traditional approach



machine-learning approach

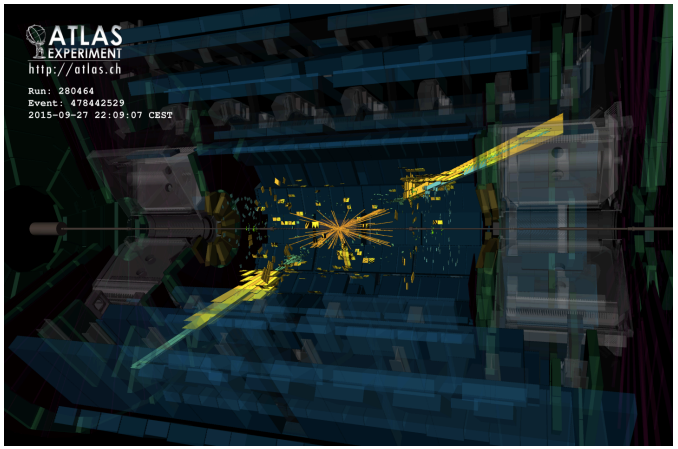


Hammers and Nails for jets

- Convolutional Neural Networks
- Recurrent Neural Networks
- Energy Flow Polynomials
- Latent Dirichlet Allocation
- Reinforcement learning
- Lorentz Boost Network
- Lorentz Layer
- Point cloud networks
- Cluster networks

- Top tagging
- W tagging
- Quark/gluon discrimination
- Pileup removal
- b/c/s-tagging
- Jet-energy scale calibration
- Missing energy measurement
- Jets in heavy ion collisions

Physics domain is *distributions*



Is this event
- two quark jets
- two gluon jets
- two Higgs bosons?

- Any individual event has no "truth" identity
- All that exists are the probability distributions for different truths

$$dP_q(x) = \frac{d^n\sigma_q}{dp_1\cdots dp_n} \qquad dP_g(x) = \frac{d^n\sigma_g}{dp_1\cdots dp_n}$$

1000-dimensional phase space
(from 10^8 dimensional measurements!)

goal is to test/measure

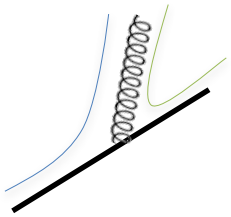
- data is a combination

$$dP_{\rm data}(x) = \alpha_q dP_q + \alpha_g dP_g + \cdots$$

Quark jets vs gluon jets: theory

Probability of quark radiating:

Color charge = 3

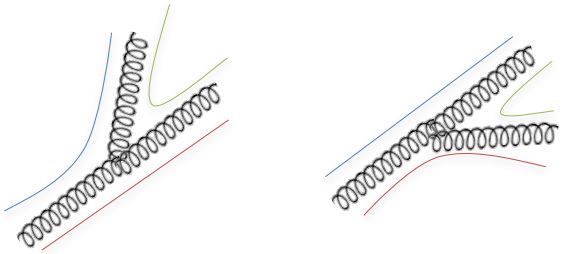


$$dP_q \sim 3 \times \frac{\alpha_s}{4\pi}(\cdots)$$

3 colors of quark

Probability of gluon radiating:

Color charge = 8

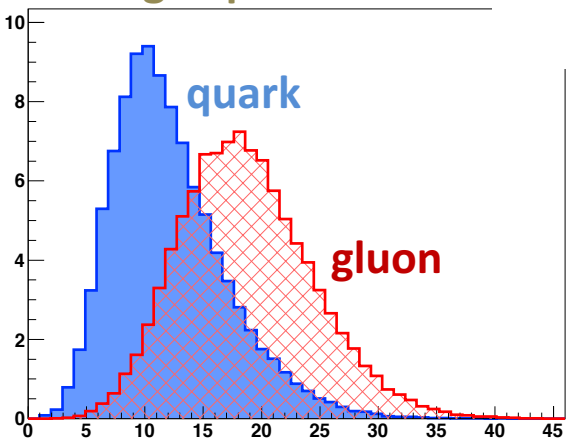


$$dP_g \sim 8 \times \frac{\alpha_s}{4\pi}(\cdots)$$

8 colors of gluon

- Gluons around twice as likely to radiate than
  - Gluon jets are fatter, on average
  - Gluon jets are more massive, on average
  - Gluon jets have more particles, on avera
  - …

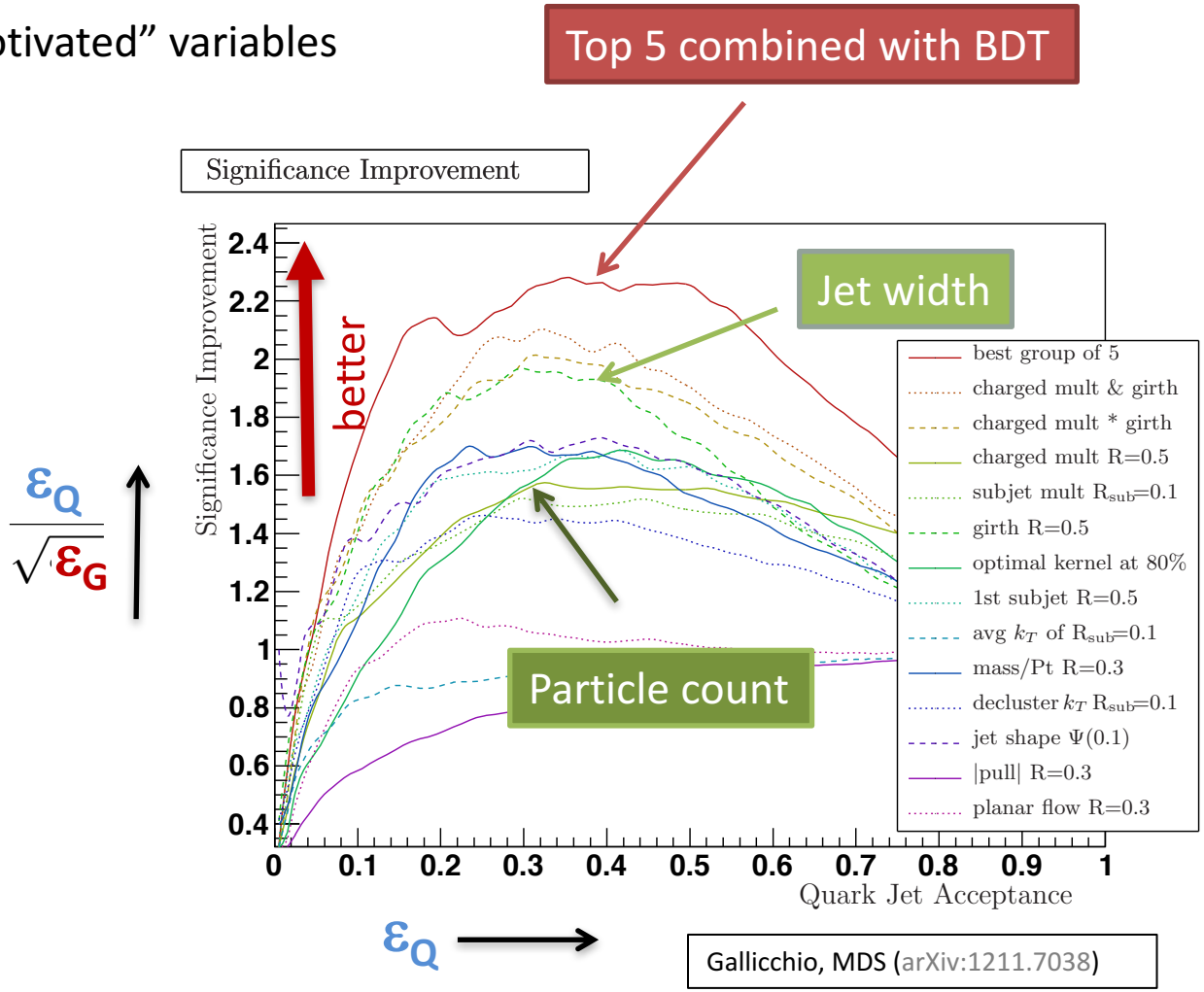
Charged particle count



Traditional approach

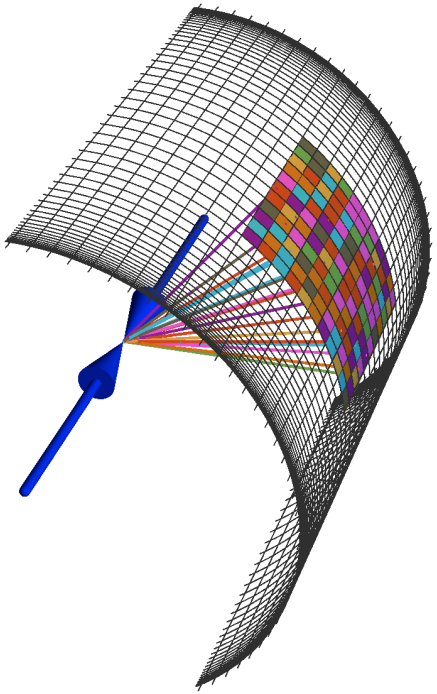
Consider lots of "motivated" variables

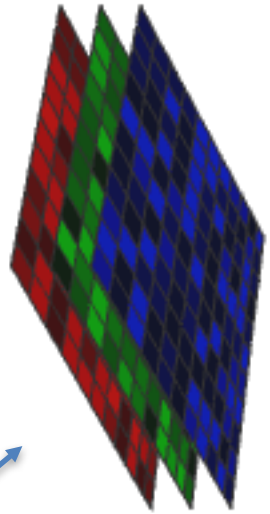
- Jet width
- # of particles
- # of subjets
- Jet "shape"
- jet mass
- …



Convolutional Neural Networks for quark/gluon jet discrimination

Komiske, Metodiev, MDS (arXiv:1612.01551)



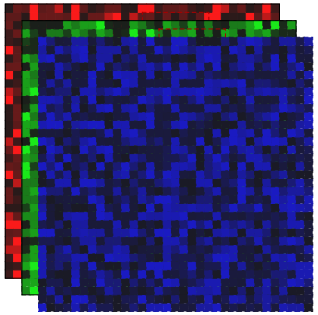


Three input layers

preprocess

- Center
- Crop
- Normalize
- Zero
- Standardize

NN inputs

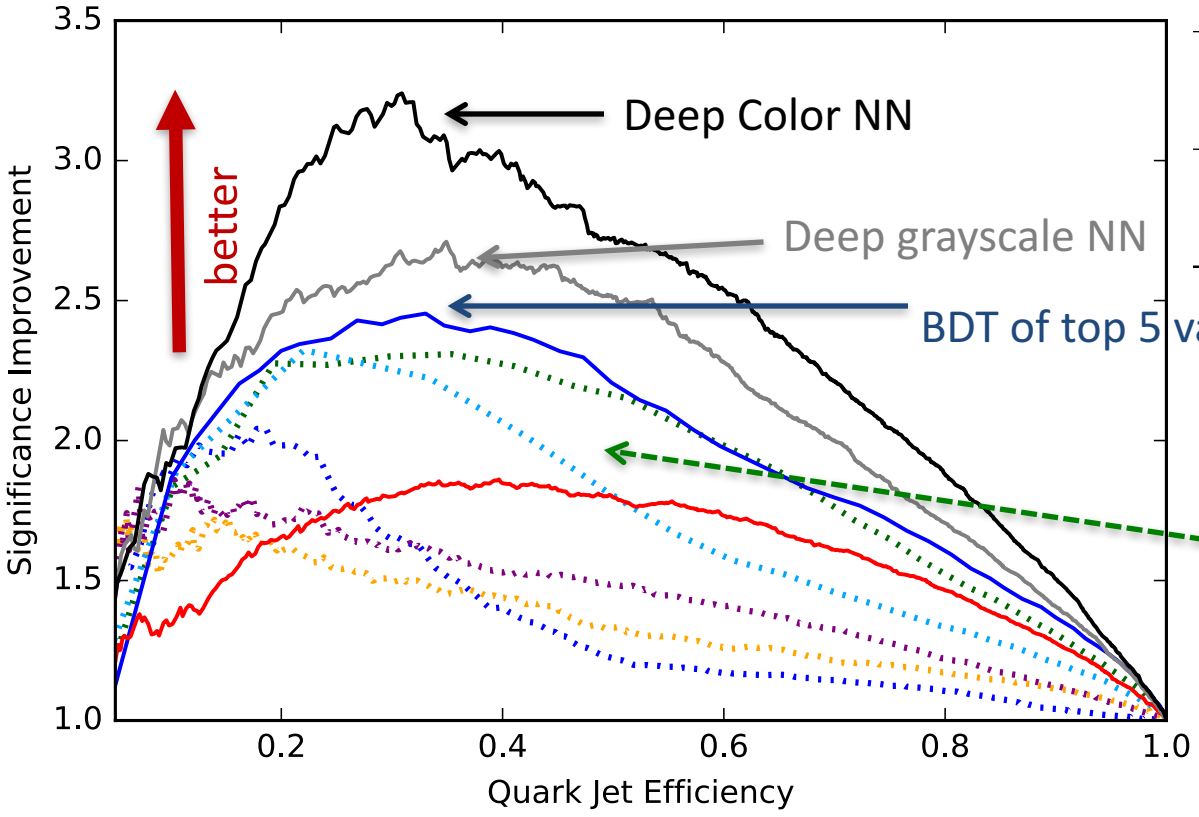


- Red = energy of charged particles
- Green = energy of neutral particles
- Blue = number of charged particles

Quark/Gluon CNN results

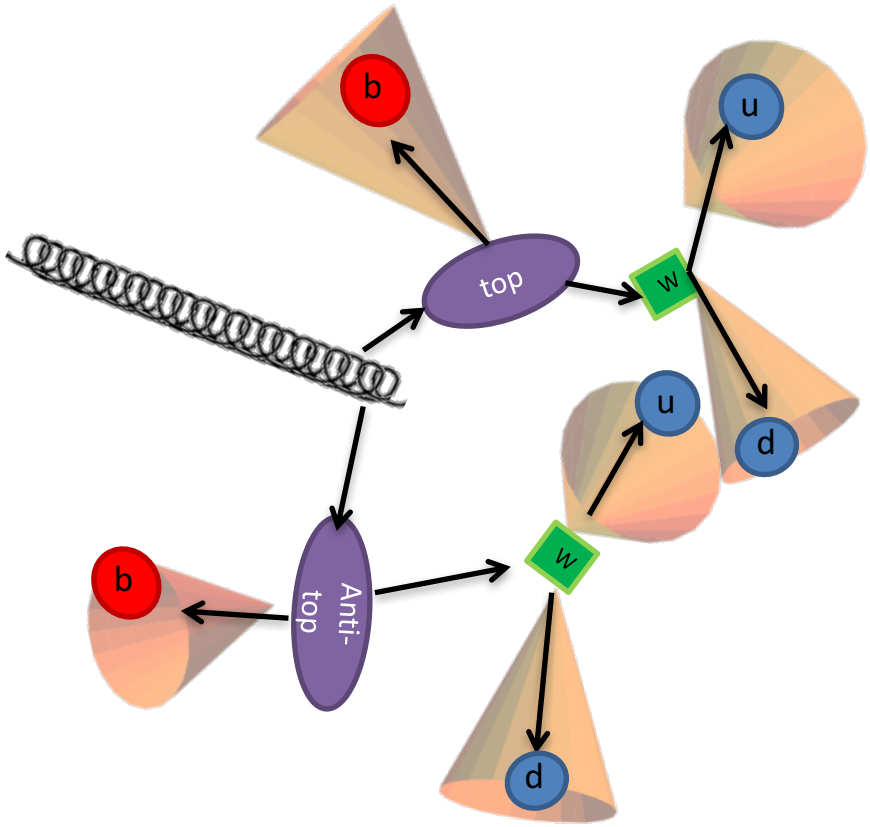
Komiske, Metodiev, MDS (arXiv:1612.01551)

$$\frac{\epsilon_Q}{\sqrt{\epsilon_G}}$$

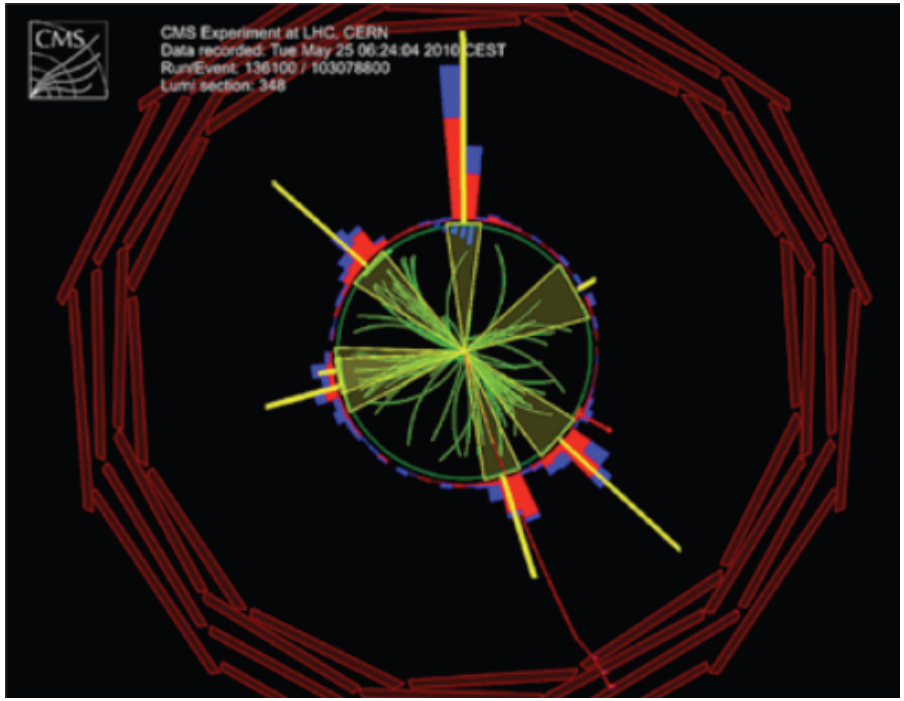


Top-tagging

Hypothetical new heavy particles often decay to top quarks:

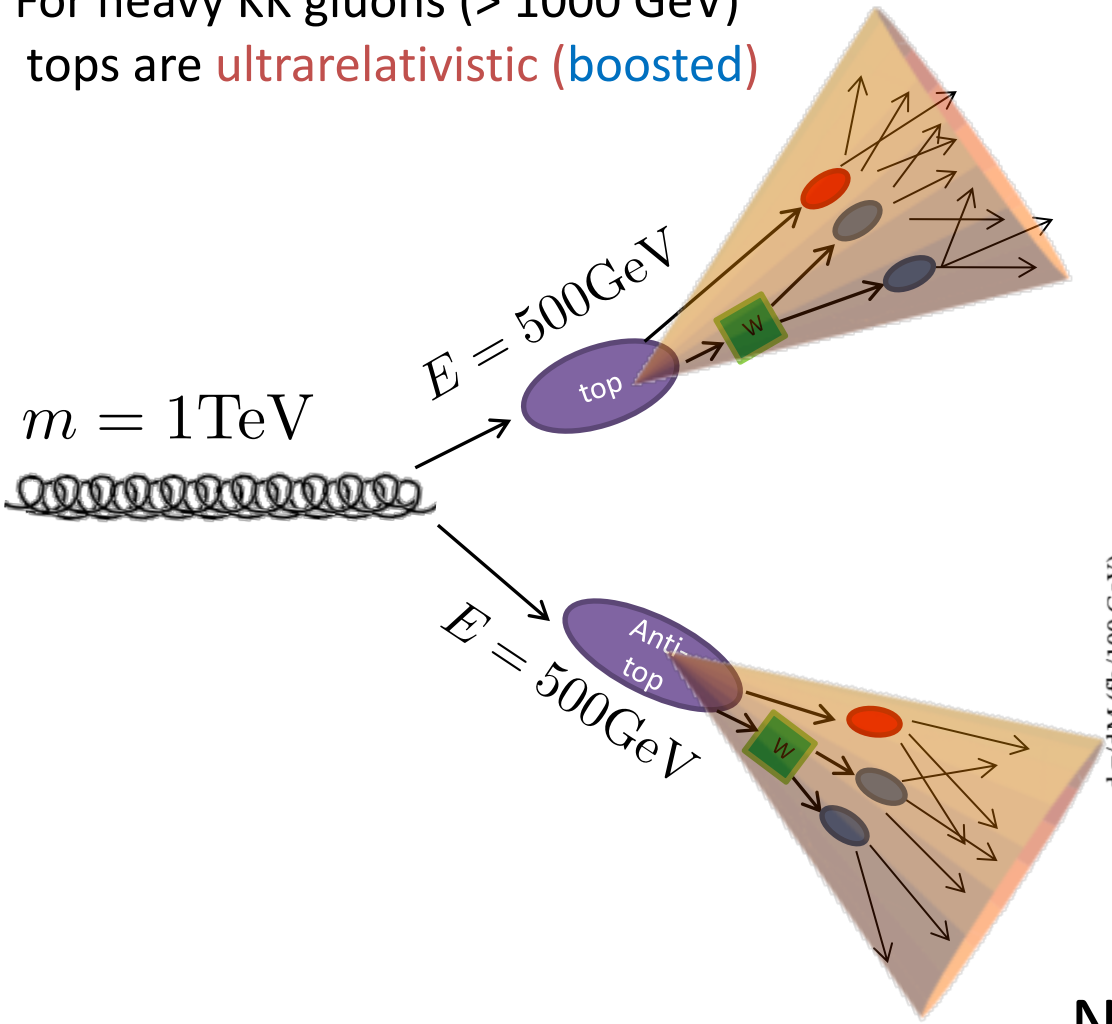


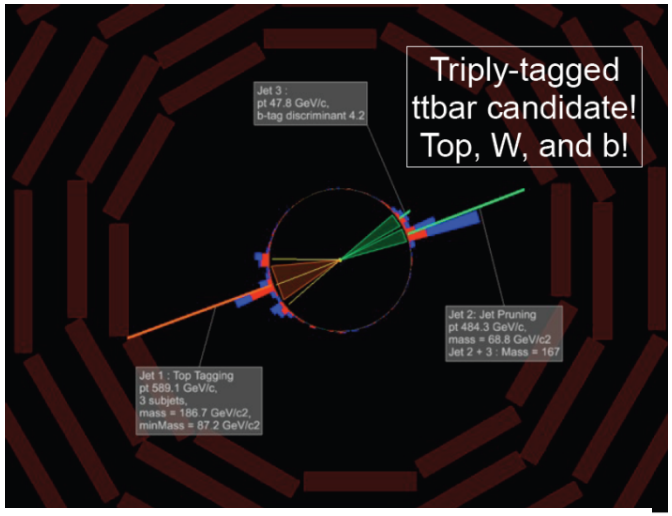
Looks like 6 Jets

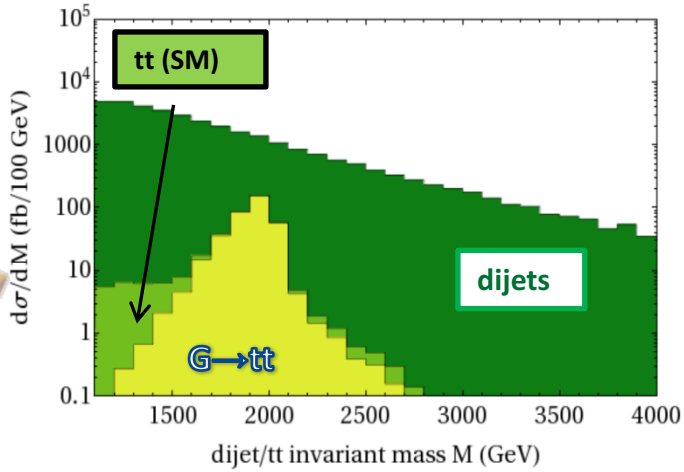


Tops are often *boosted*

For heavy KK gluons (> 1000 GeV)
tops are ultrarelativistic (boosted)



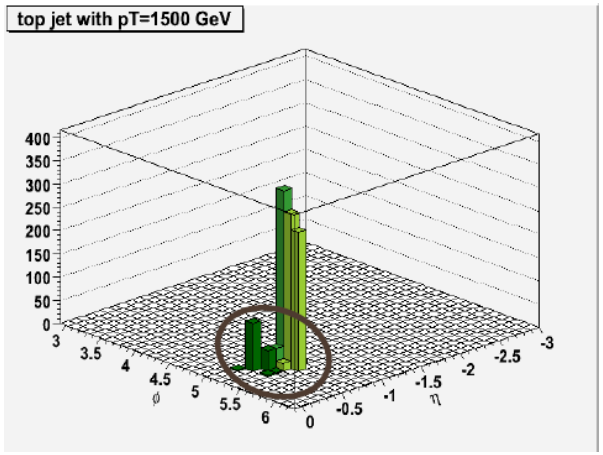


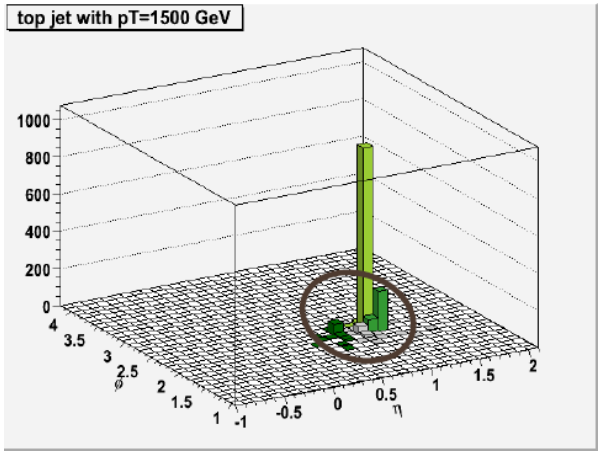


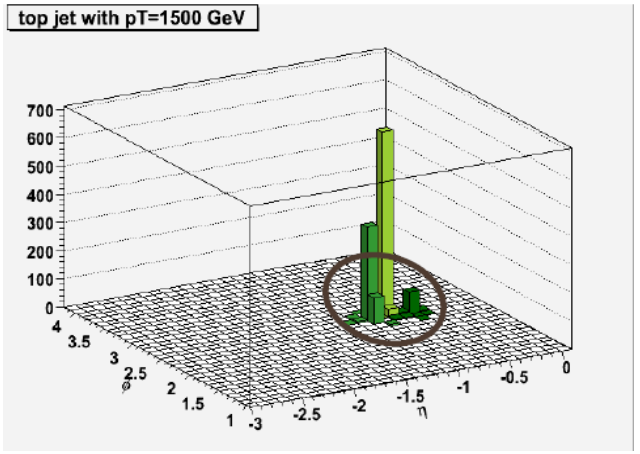
Now it looks like 2 jets!

Typical top jets

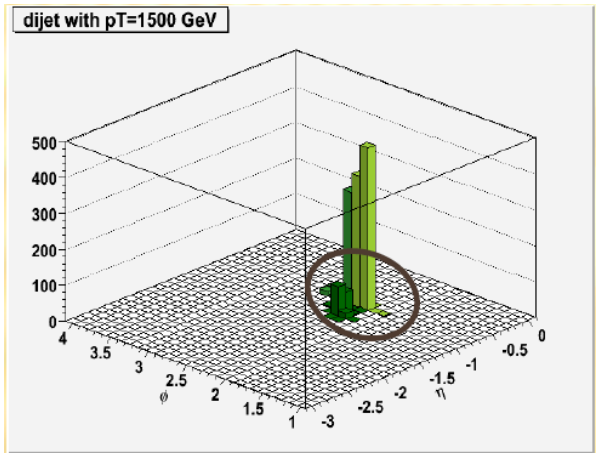
Large boost (P_T = 1500 GeV)

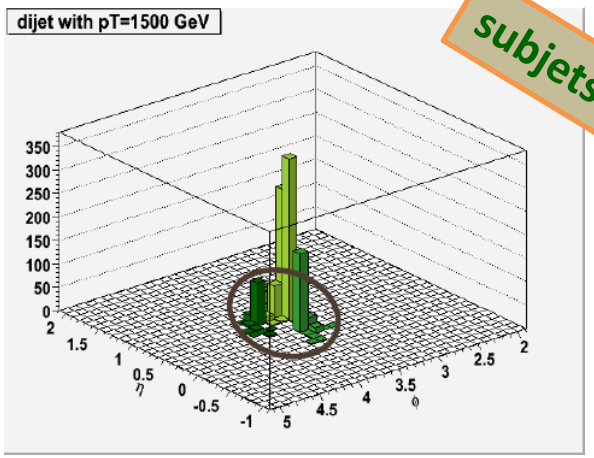




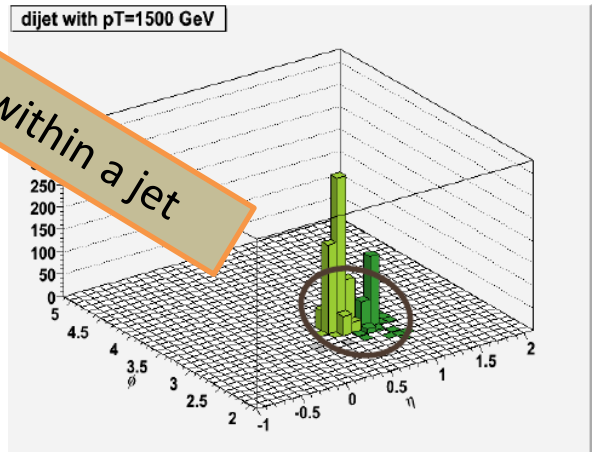


Typical background jets



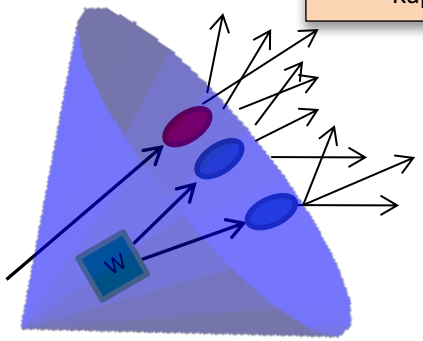


subjets within a jet

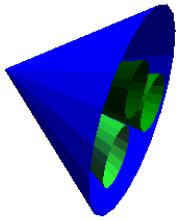


Top-tagging

Hopkins top-tagger
Kaplan et al. arXiv:0806.0848

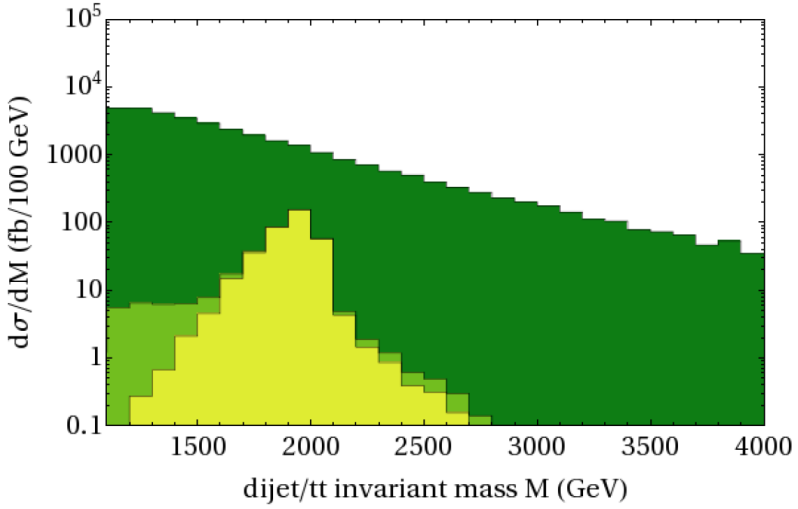


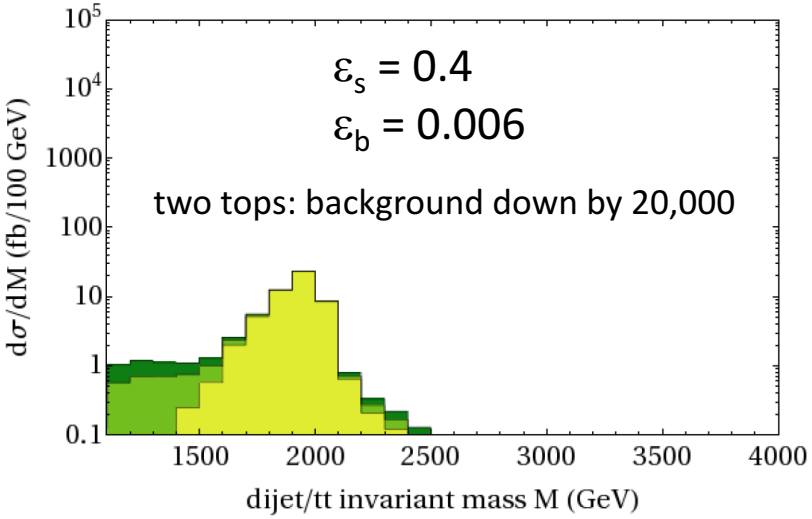
1. Look for big jets (R = 1.2)



2. with subjets within the jet

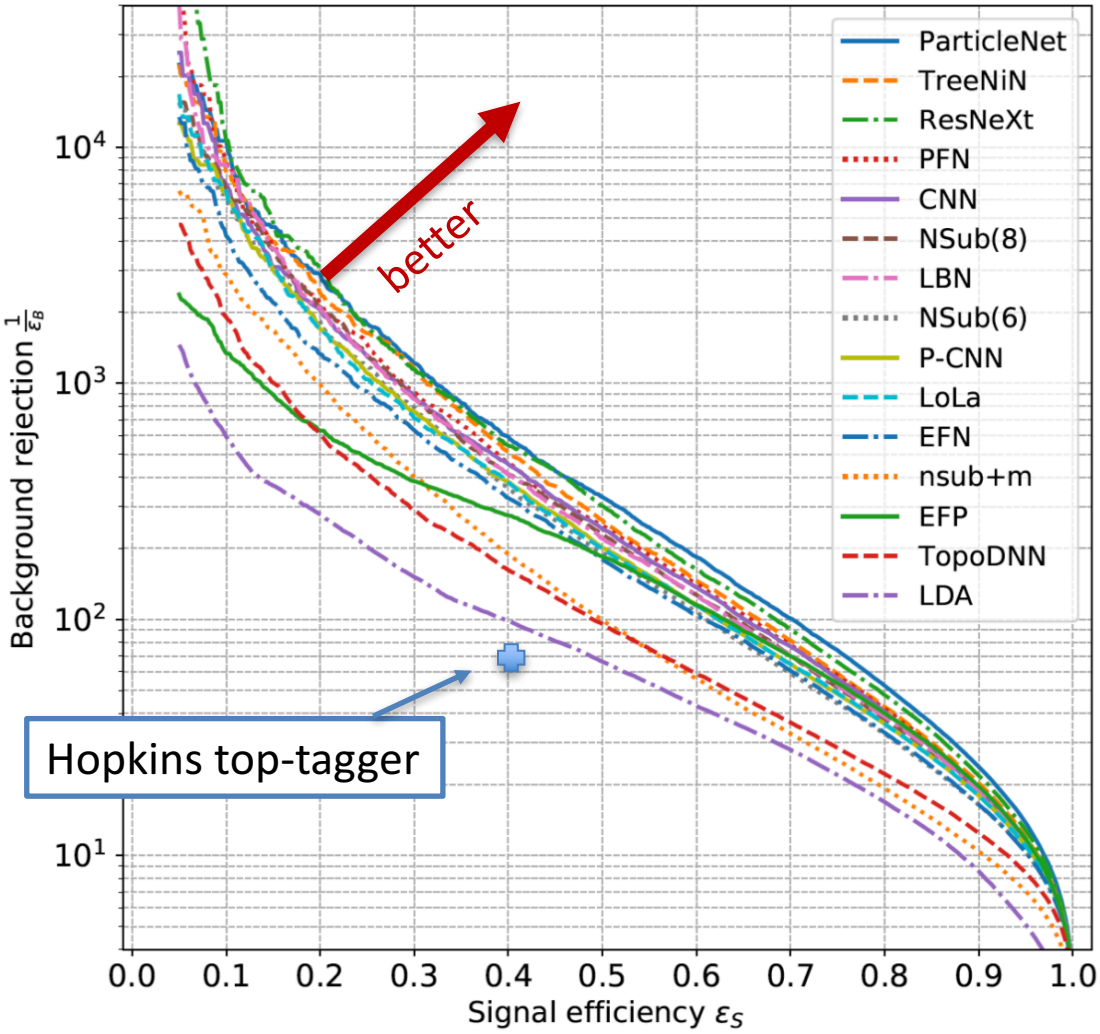
3. Analyze the subjets
   - W mass peak, top mass peak, and helicity angle





Many ML hammers applied to the top-tagging nail

arXiv:1902.09914



Apples-to-apples top-tagging comparison

arXiv:1902.09914

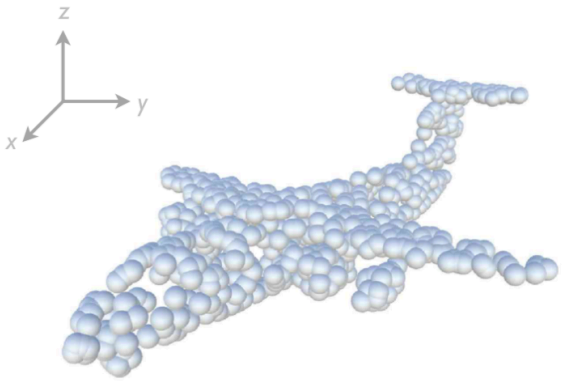
	AUC	Acc	$1/\epsilon_B$ ($\epsilon_S = 0.3$) single	mean	median	#Param
image based						
CNN [16]	0.981	0.930	914±14	995±15	975±18	610k
ResNeXt [30]	0.984	0.936	1122±47	1270±28	1286±31	1.46M
4-vector based						
TopoDNN [18]	0.972	0.916	295±5	382± 5	378 ± 8	59k
Multi-body N-subjettiness 6 [24]	0.979	0.922	792±18	798±12	808±13	57k
Multi-body N-subjettiness 8 [24]	0.981	0.929	867±15	918±20	926±18	58k
TreeNiN [43]	0.982	0.933	1025±11	1202±23	1188±24	34k
P-CNN	0.980	0.930	732±24	845±13	834±14	348k
ParticleNet [47]	0.985	0.938	1298±46	1412±45	1393±41	498k
theory inspired						
LBN [19]	0.981	0.931	836±17	859±67	966±20	705k
LoLa [22]	0.980	0.929	722±17	768±11	765±11	127k
Energy Flow Polynomials [21]	0.980	0.932	384			1k
Energy Flow Network [23]	0.979	0.927	633±31	729±13	726±11	82k
Particle Flow Network [23]	0.982	0.932	891±18	1063±21	1052±29	82k
GoaT	0.985	0.939	1368±140		1549±208	35k
particle cloud						
ParticleNet-Lite	0.984	0.937	1262±49			26k
ParticleNet	0.986	0.940	1615±93			366k

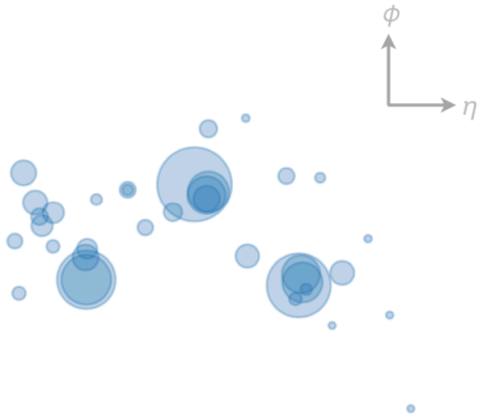
- Uses same samples
- 800k training, 200k test
  - Goat: 200k/1368 = 146 bg events survive
  - Particlenet: 200k/1615 = 124 bg events survive

Particle net uses point cloud approach

Gouskos and Qu
arXiv:1902.08570

- Respects permutation symmetry
- See also energy flow polynomials





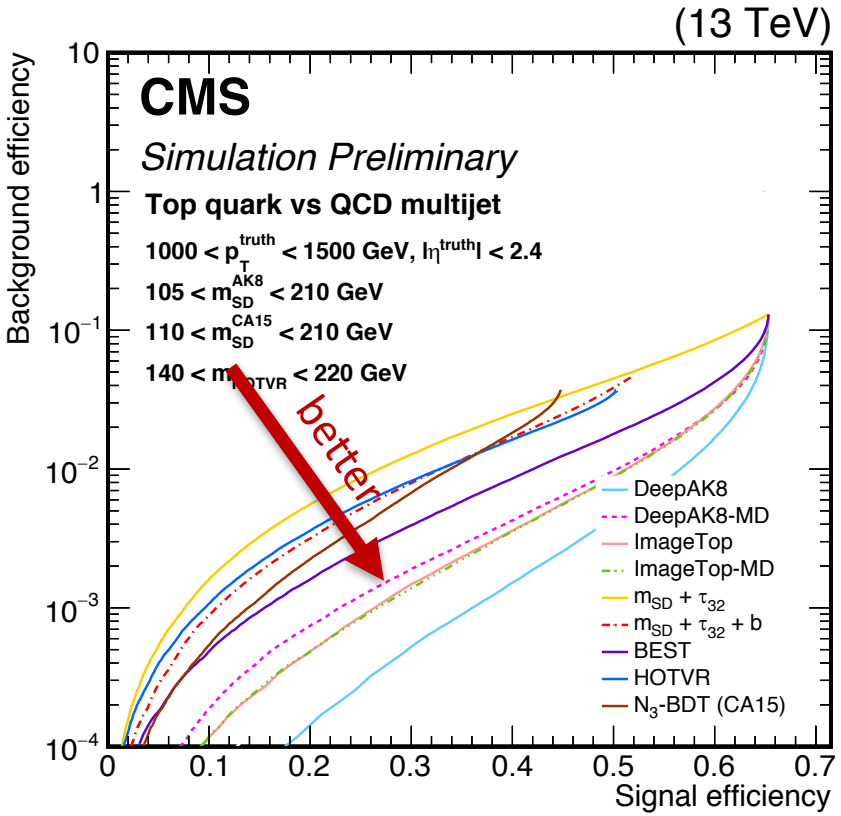
- Point cloud
  - points are intrinsically *unordered*
  - primary information:
    - 3D coordinates in the xyz space

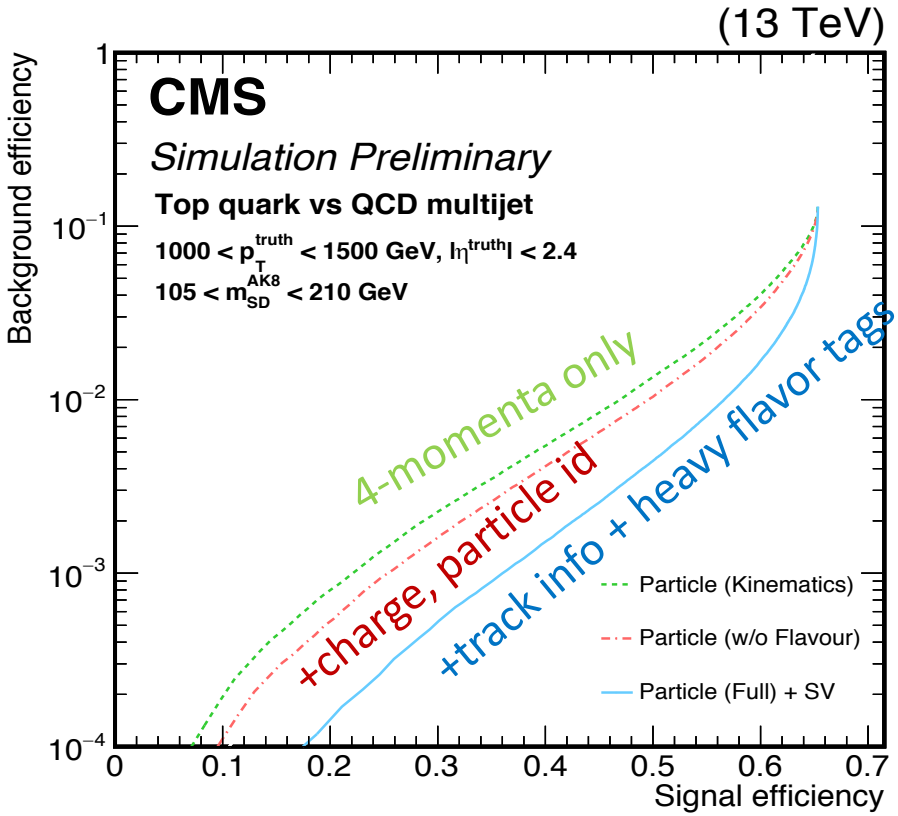
- Particle cloud
  - particles are intrinsically *unordered*
  - primary information:
    - 2D coordinates in the η-ϕ space
  - *but also many additional features!*
    - energy/momenta
    - charge/particle ID
    - track quality/impact parameters/etc.

- uses EdgeConv
- angular distance metric
- k-nearest neighbors

CMS top-tagging: DeepAK8 ML algorithm

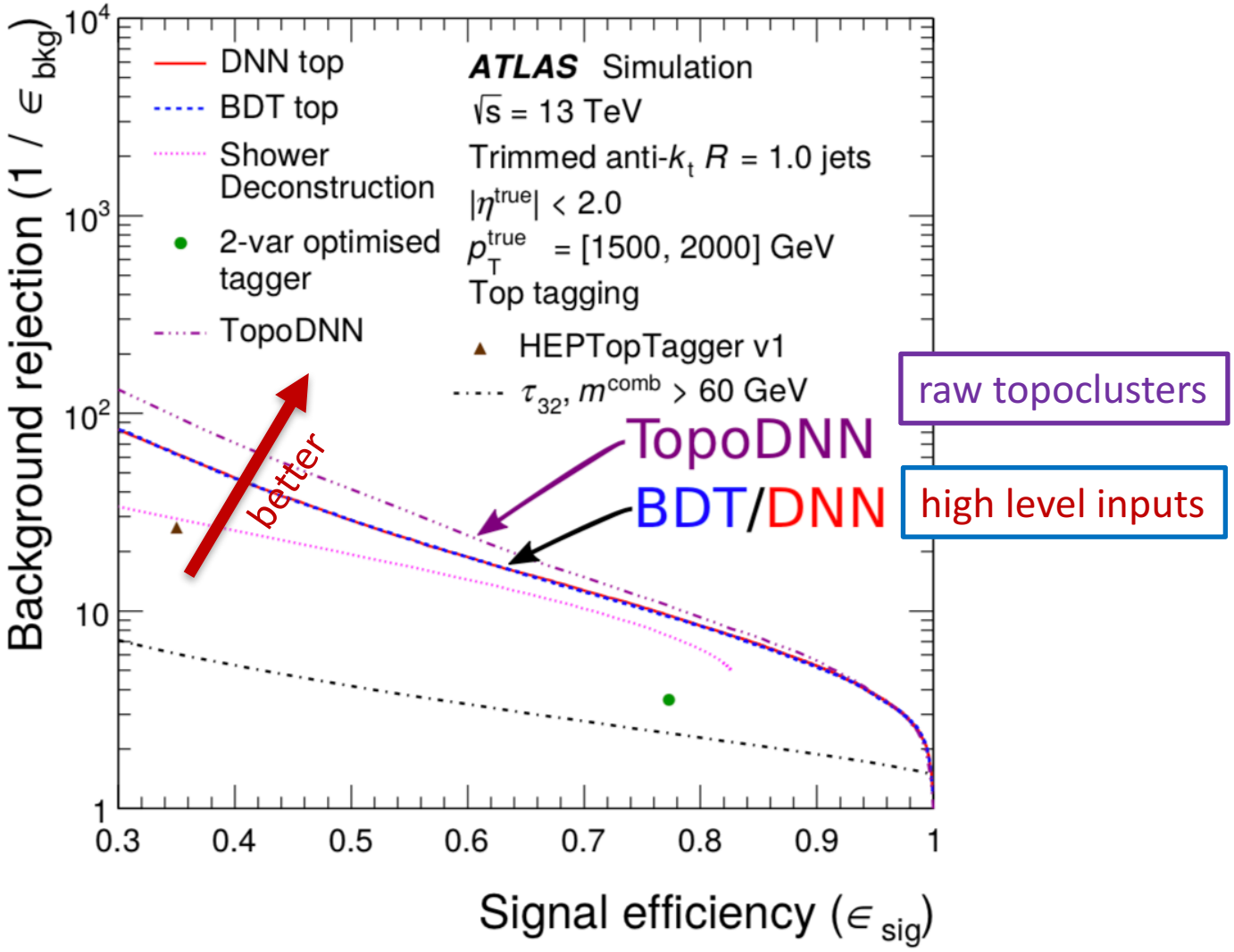
- Inputs particles/tracks to ResNet archictecture





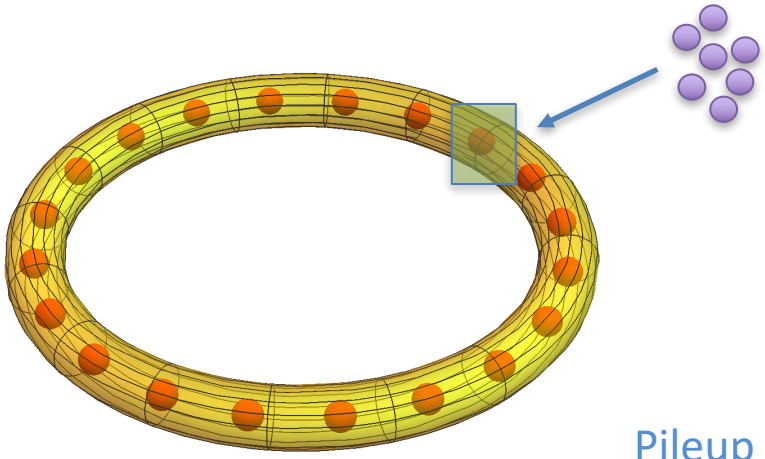
ATLAS top-tagging : topoDNN

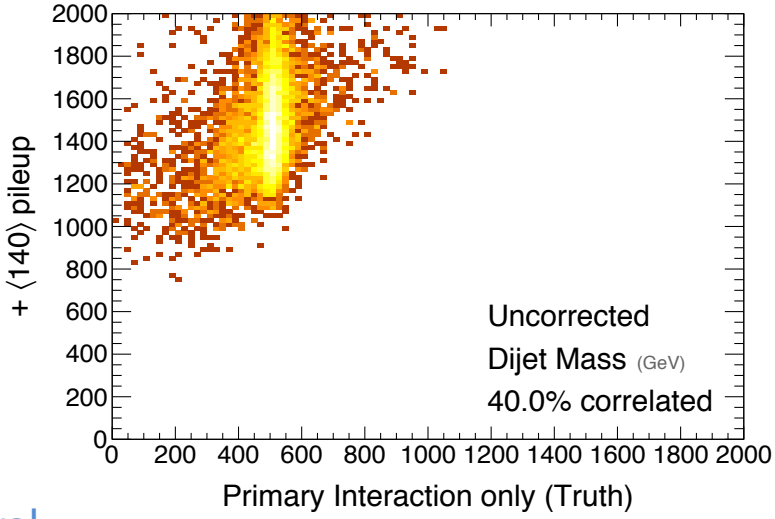
- inputs topoclusters to deep neural network
- also finds better performance with modern machine learning



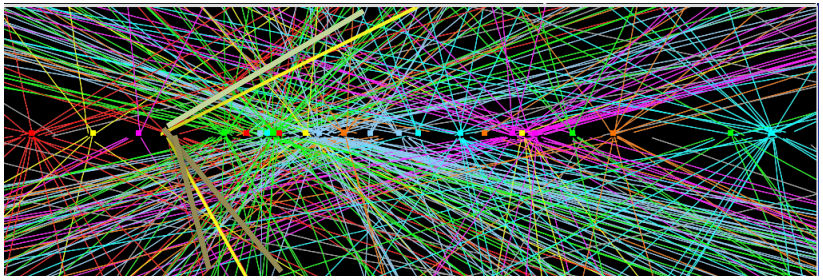
Pileup

- LHC collides protons in bunches
  - 10^{11} protons/bunch
  - Up to 200 collisions per bunch crossing





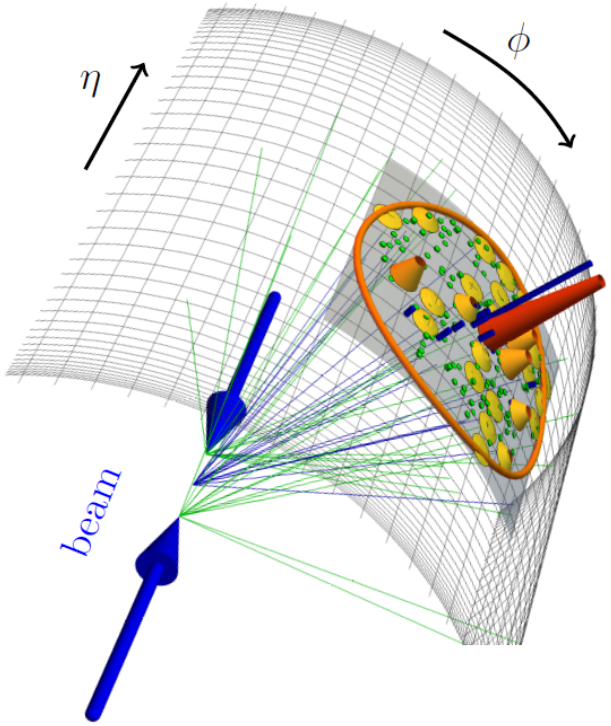
Pileup removal algorithm

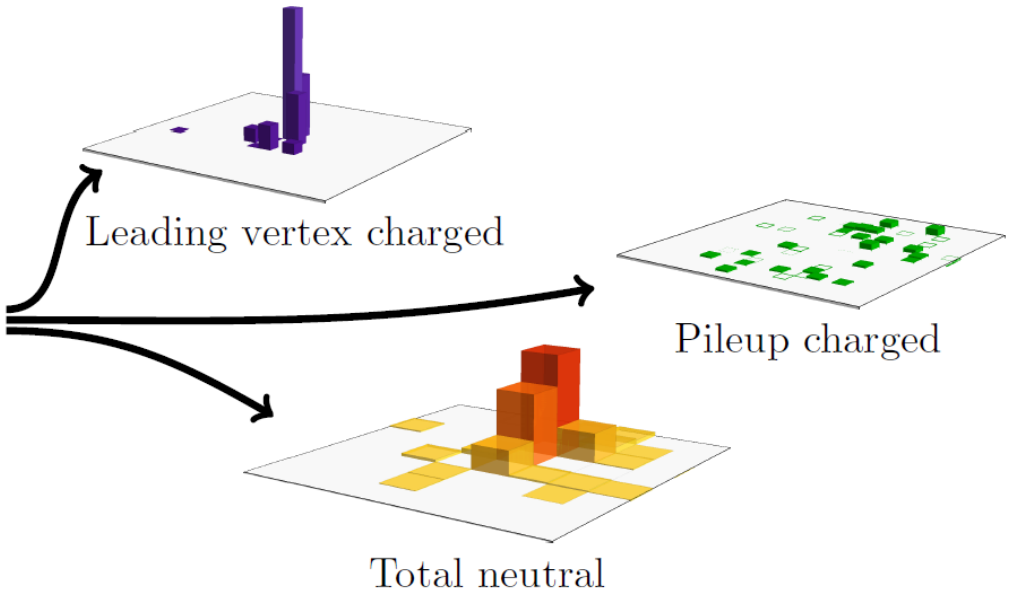


- Tracking system can resolve primary collision from secondary ”pileup” collisions
- Only charged particles can be seen this way

- Can we use machine learning to remove the pileup radiation?

Pileup removal as regression problem





Can measure

1. Leading vertex charged particles
2. Pileup charged particles
3. Total neutral particles

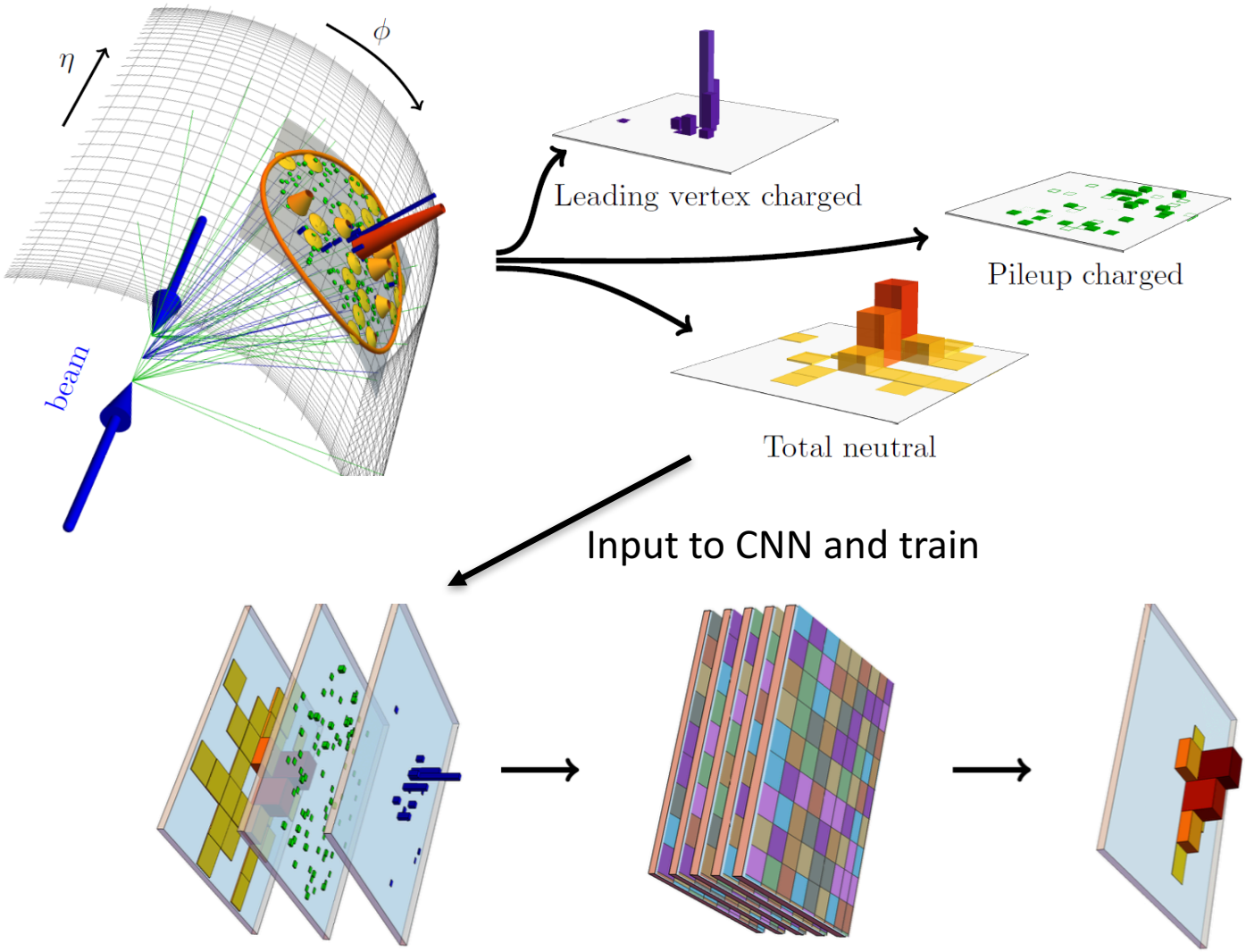
?

Leading vertex
neutral particles

CNNs for Pileup Removal

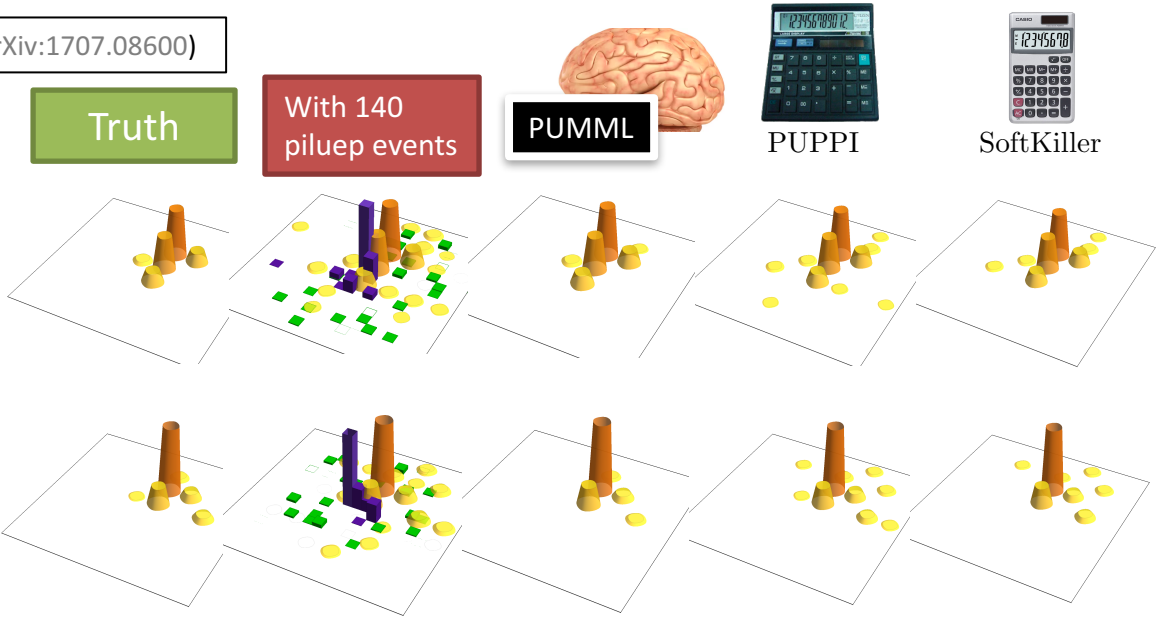
Komiske, Metodiev, Nachman, MDS (arXiv:1707.08600)

- Separate observable energy deposits into 3 images



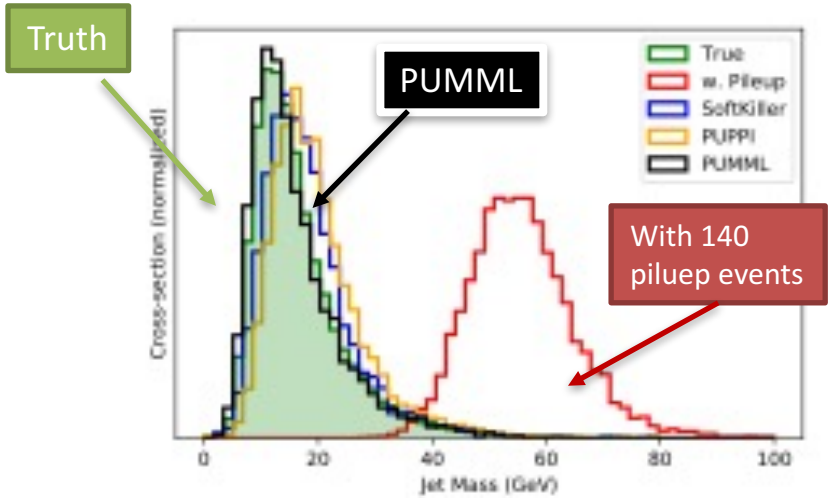
PileUp Mitigation with Machine Learning (PUMML)

Komiske, Metodiev, Nachman, MDS (arXiv:1707.08600)

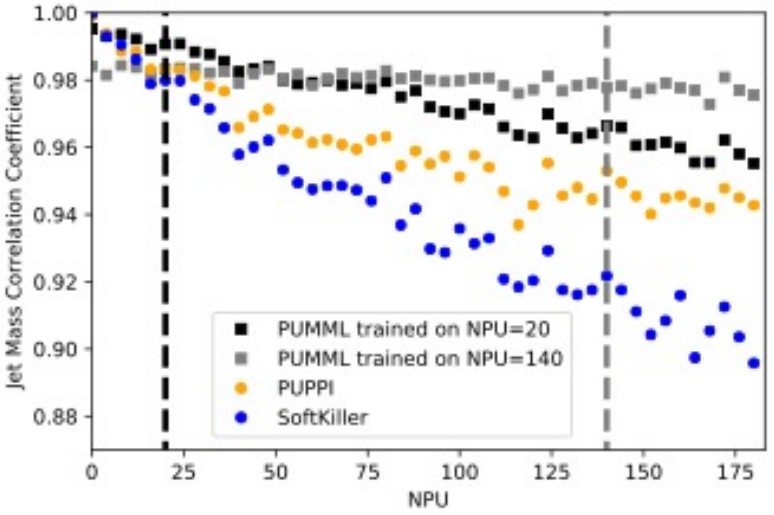


- Excellent leading vertex (truth) reconstruction

- Excellent observable reconstruction



- Excellent stability for variable piluep #



ATLAS uses CNN to measure MET (in simulation) ATL-PHYS-PUB-2019-028

$$E_{\mathrm{T}}^{\mathrm{miss}} = -\left(\sum_{i \in \mathrm{muons}} p_{\mathrm{T}}{}^{i} + \sum_{i \in \mathrm{electrons}} p_{\mathrm{T}}{}^{i} + \sum_{i \in \mathrm{photons}} p_{\mathrm{T}}{}^{i} + \sum_{i \in \mathrm{hadronic}\ \tau} p_{\mathrm{T}}{}^{i} + \sum_{i \in \mathrm{jets}} p_{\mathrm{T}}{}^{i} + \sum_{i \in \mathrm{Soft\ Term}} p_{\mathrm{T}}{}^{i} \right)$$

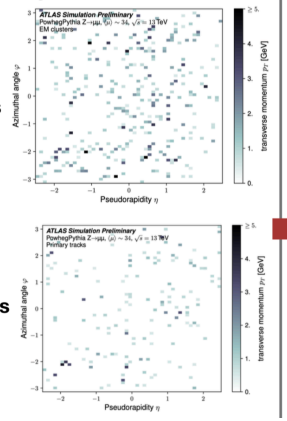
Input (*Z+jets events*)

Clusters

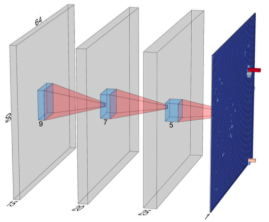
+

PV-Tracks

+

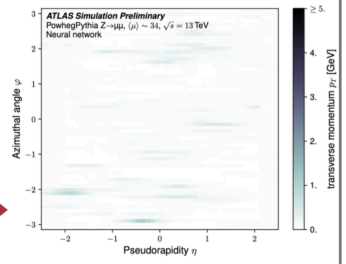


CNN, Network Architecture

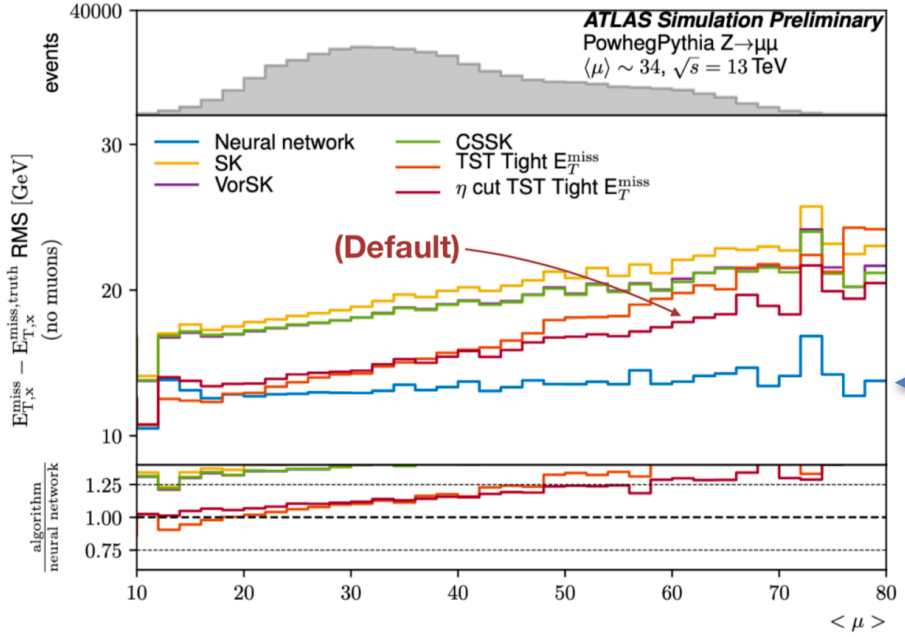


Layer	Kernelsize	Filters	Activation	Parameters
2D convolution	9	15	ReLU	7305
dropout				
2D convolution	7	20	ReLU	14720
dropout				
2D convolution	5	25	ReLU	12525
dropout				
2D convolution	1	1	ReLU	26
Total				34576

Output



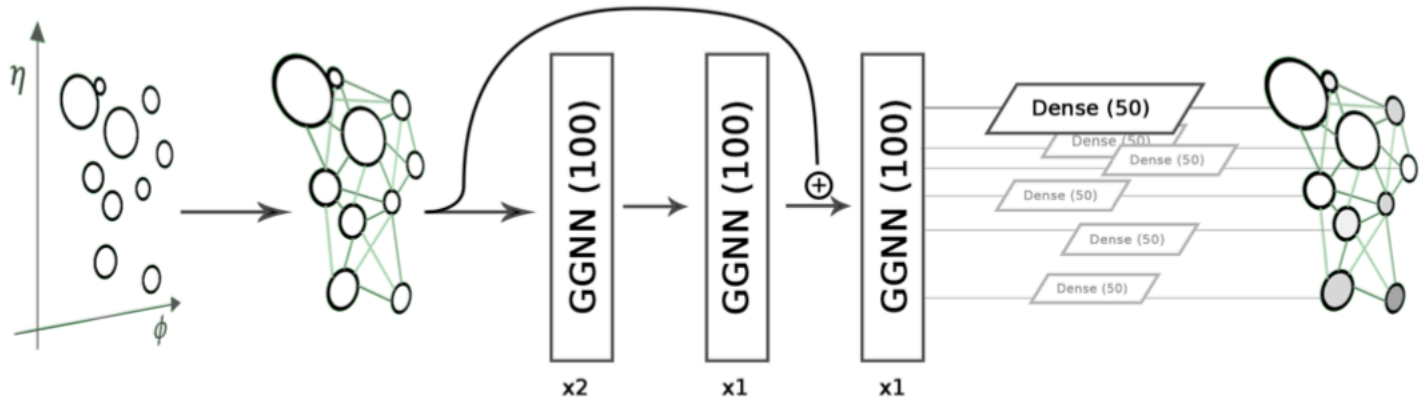
Output: **an event image with MET as-close-as-possible to the truth-level expectation.**

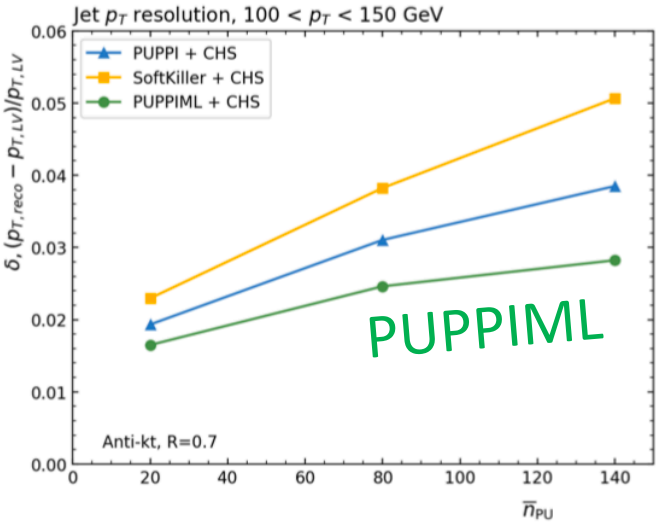


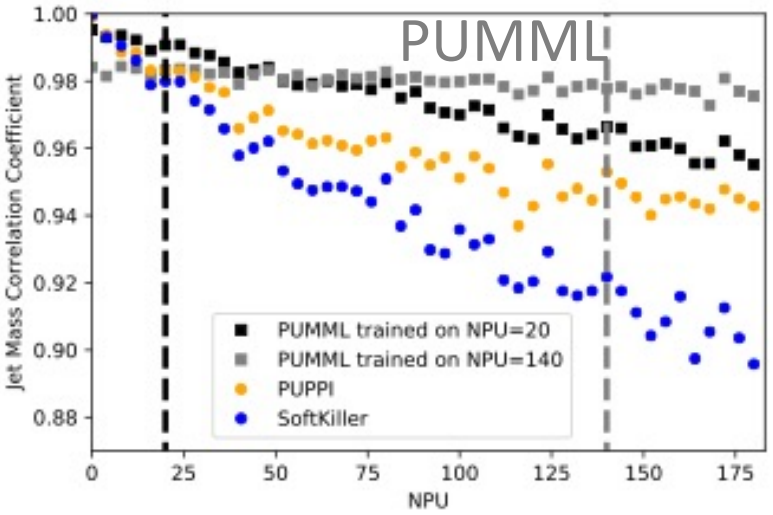
CNN-based image regression

Martinez et al. arXiv:1810.07988

PUPPIML: Graphnet approach







- good stability
- comparable to PUMML

JUNIPR

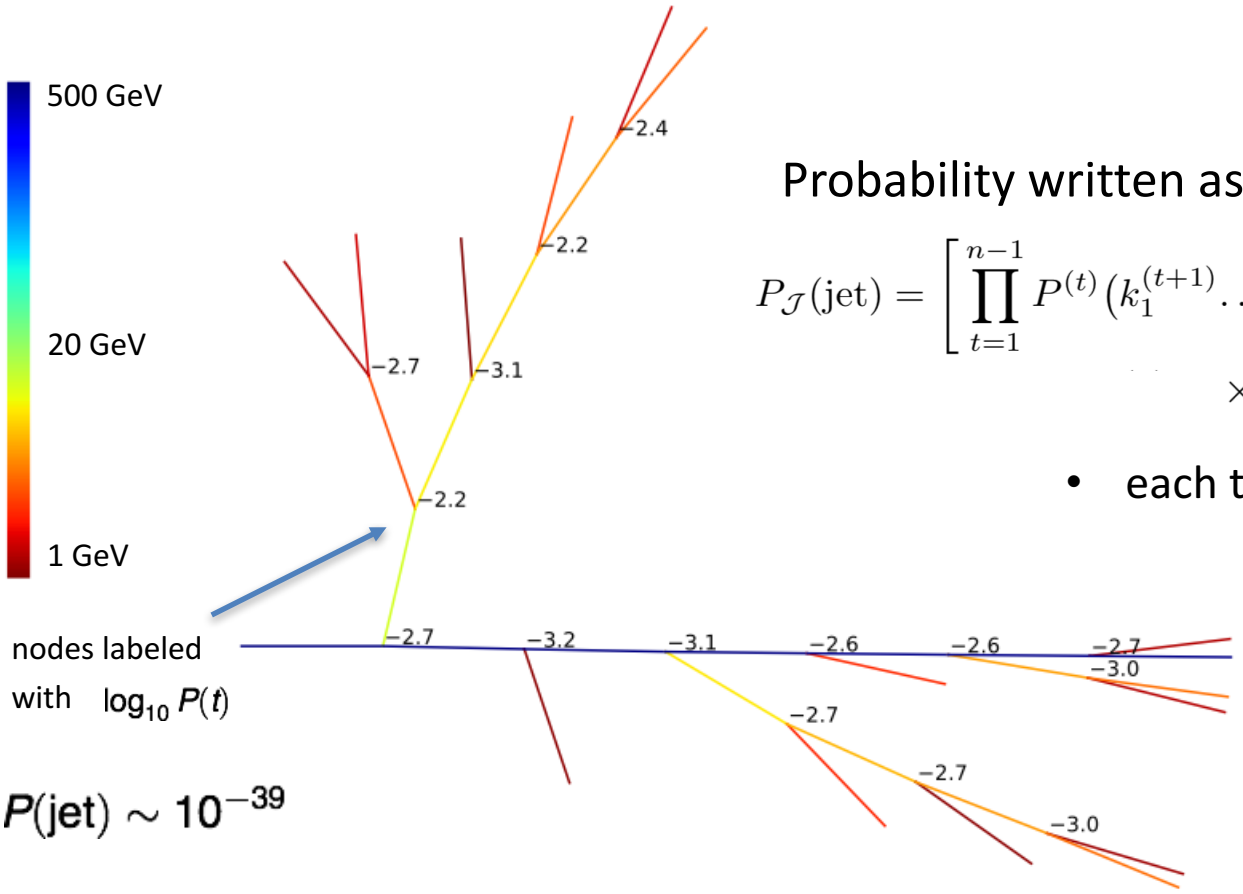
Andreassen, Feige, Frye , MDS

arXiv:1804.09720

arXiv:1906.10137

- unsupervised approach: learn probability distribution for each sample
- represent data as clustering tree
- can be used to classify or generate

$$dP_q(x) = \frac{d^n\sigma_q}{dp_1 \cdots dp_n}$$



nodes labeled with $\log_{10} P(t)$

$P(\text{jet}) \sim 10^{-39}$

Probability written as product

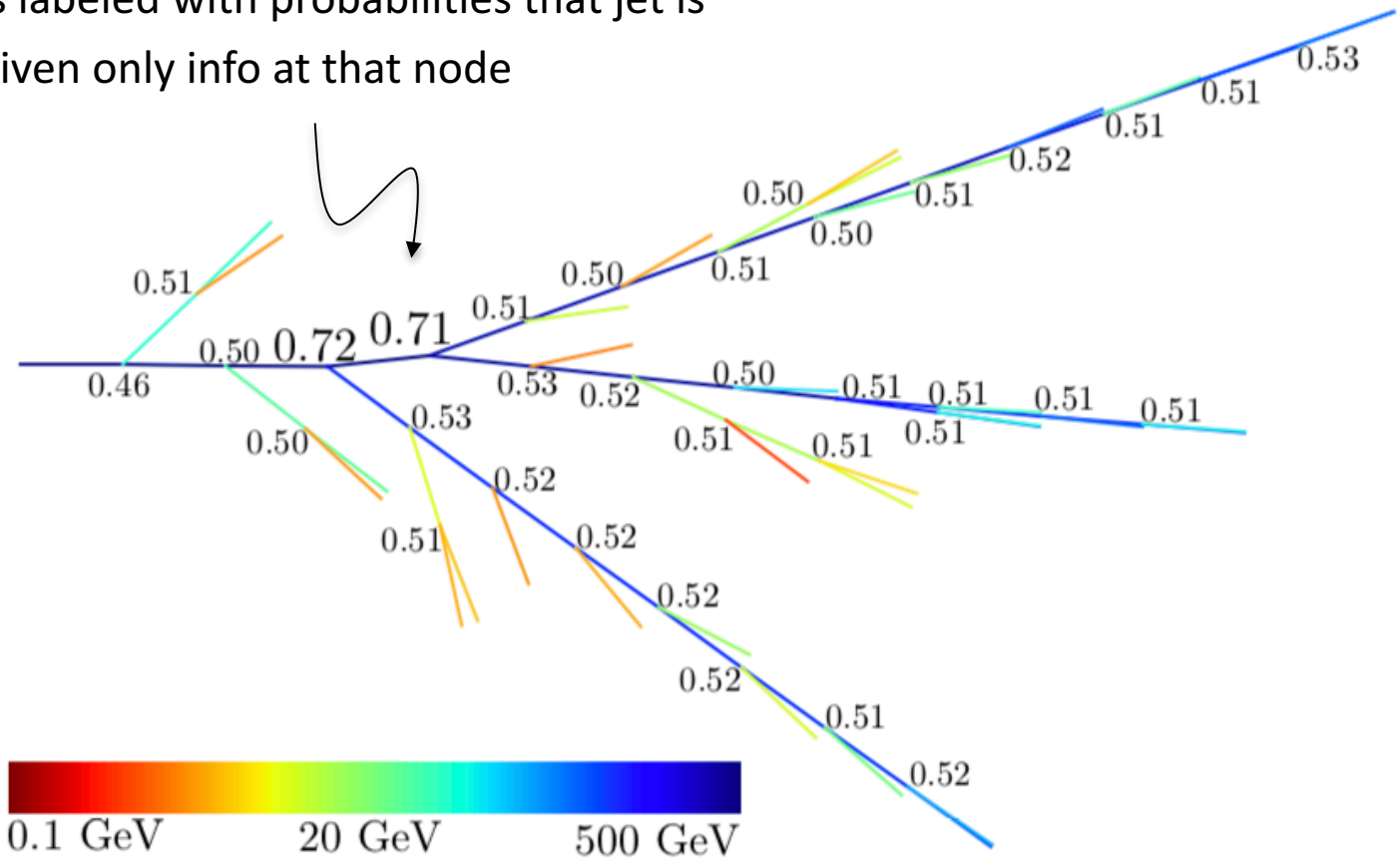
$$P_{\mathcal{J}}(\text{jet}) = \left[\prod_{t=1}^{n-1} P^{(t)}\left(k_1^{(t+1)} \dots k_{t+1}^{(t+1)} \middle| k_1^{(t)} \dots k_t^{(t)}\right)\right] \times P^{(n)}\left(\text{end} \middle| k_1^{(n)} \dots k_n^{(n)}\right)$$

- each term interpretable

JUNIPR

Visualizing discrimination power: boosted top or QCD jet?

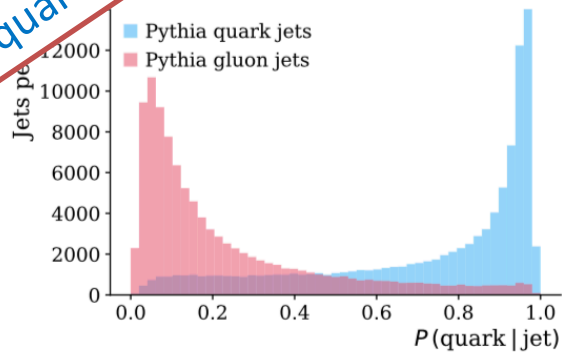
- nodes labeled with probabilities that jet is top, given only info at that node

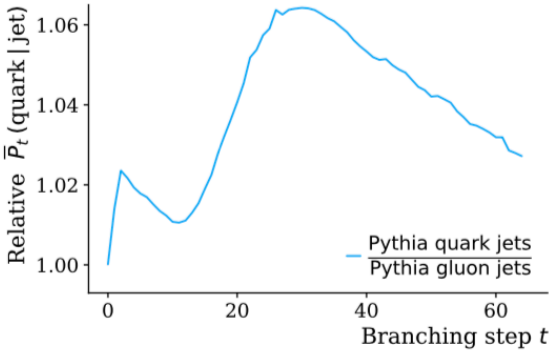


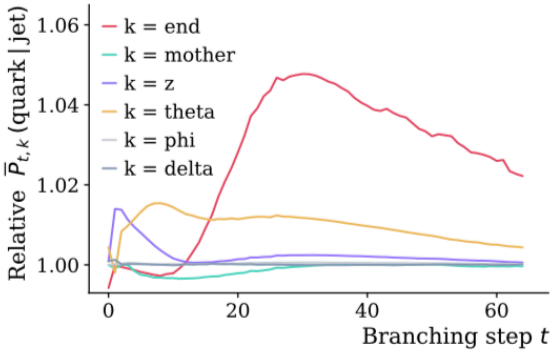
JUNIPR

What is different?

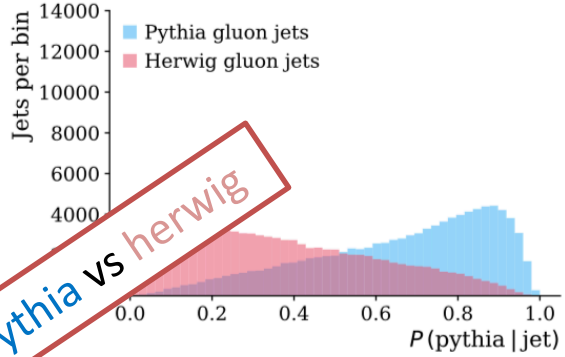
quark vs glue

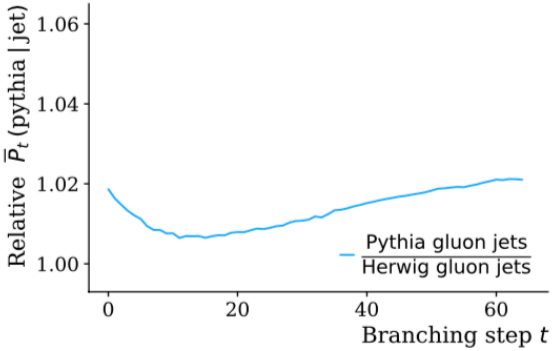


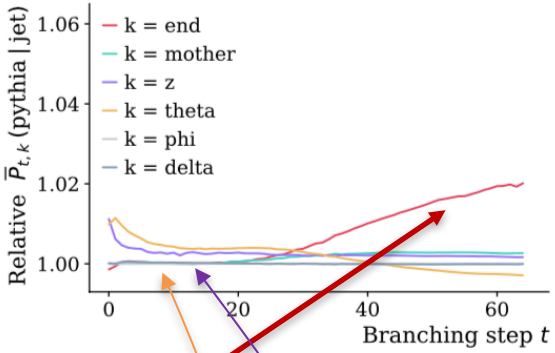




pythia vs herwig





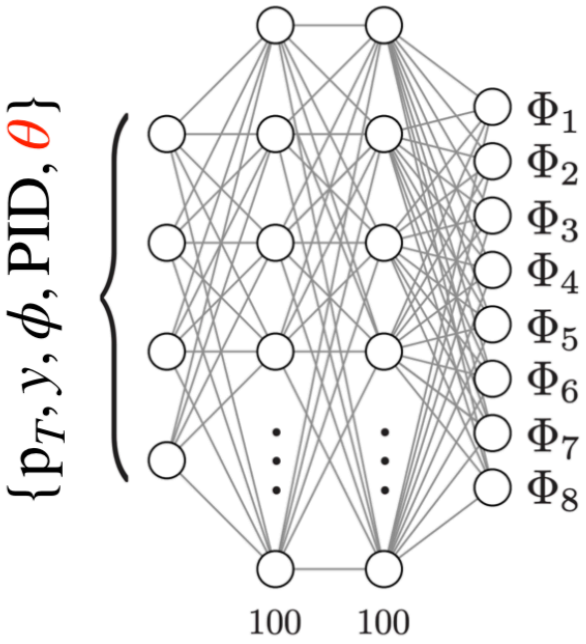


- Length of shower important
- angular distribution early on
- energy sharing different early on

Andreassen, Nachman arXiv:1907.08209

DCTR: Use relative weights for tuning

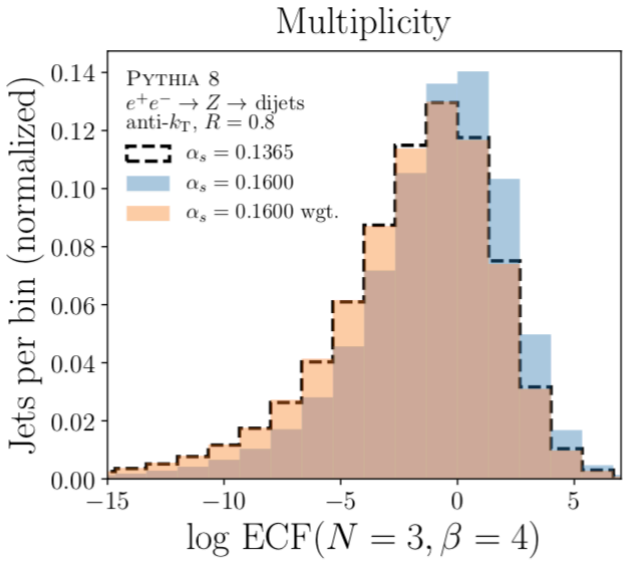
- includes simulation parameters θ in truth data



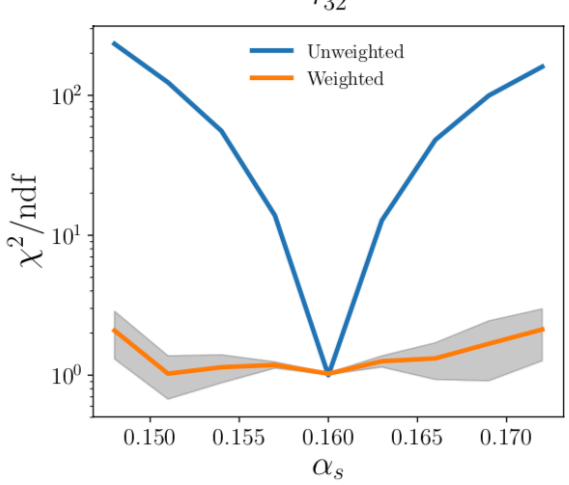
- learns relative weight P(x, θ)/P(x,θ_0)

- Could be a very efficient way to tune simulations
- or to reweight simulations to data

Reweights α_s=0.1365 distribution back to α_s=0.1600



χ^2 is minimal for α_s=0.1600



Reinforcement learning

Dreyer and Carrazza

arXiv:1903.09644

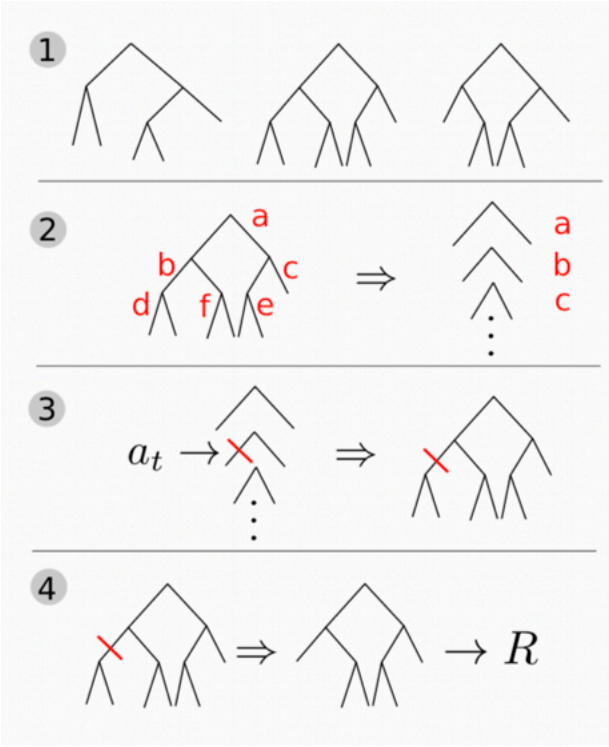
Tries to learn jet grooming and mass resolution

reward function: $R(m, a_t, \Delta, z) = R_M(m) + \frac{1}{N_{\rm SD}} R_{\rm SD}(a_t, \Delta, z)$

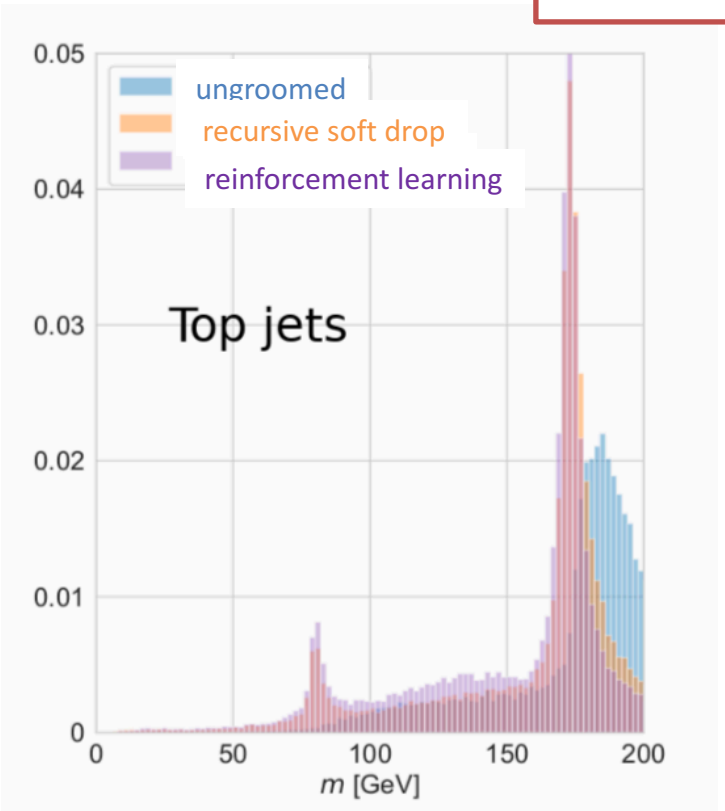
close to soft drop

close to right mass

Game is remove legs in clustering tree

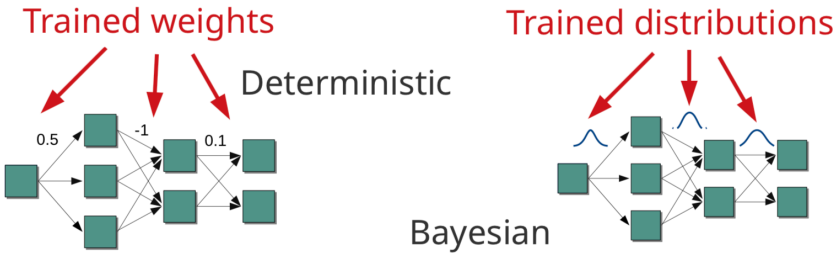


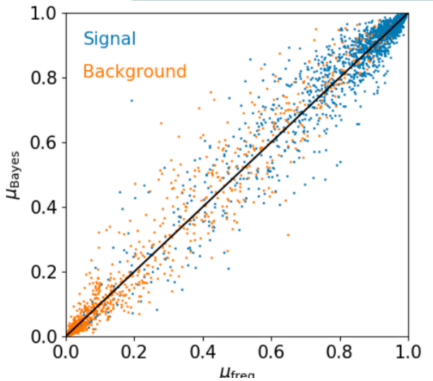
works ok



Baysian networks

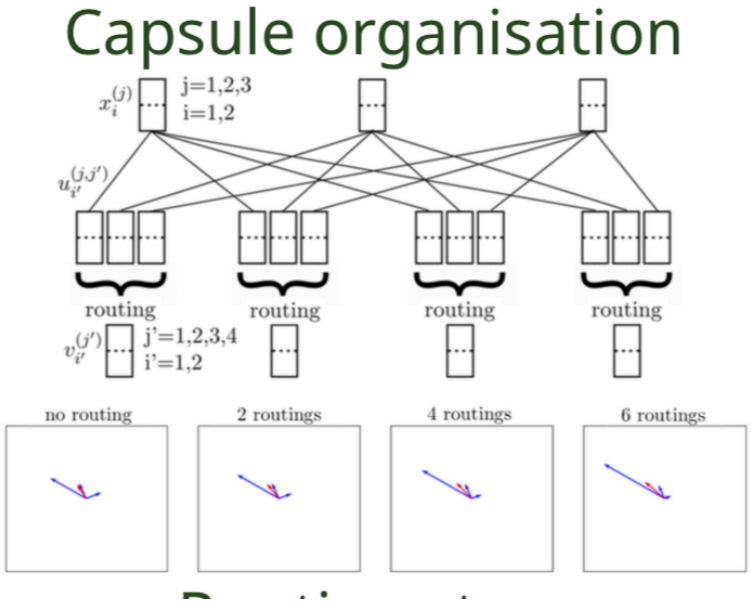
arXiv:1904.10004





Capsule networks for whole-event tagging

arXiv:1906.11265



See Gregor Kasieczka's talk

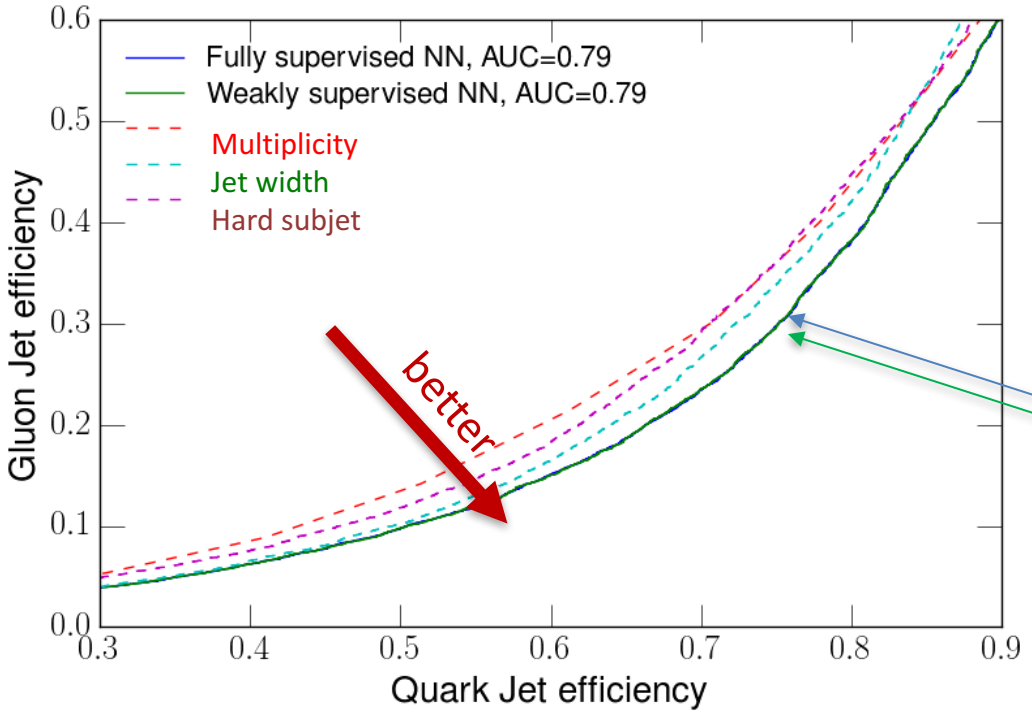
Weak supervision

Dery et al. (arXiv:1702.00414)

Supervised learning: pure samples of quark and gluons used

Weak learning: mixed samples of quarks or gluons used

Unsupervised learning: no labels at all, just find patterns

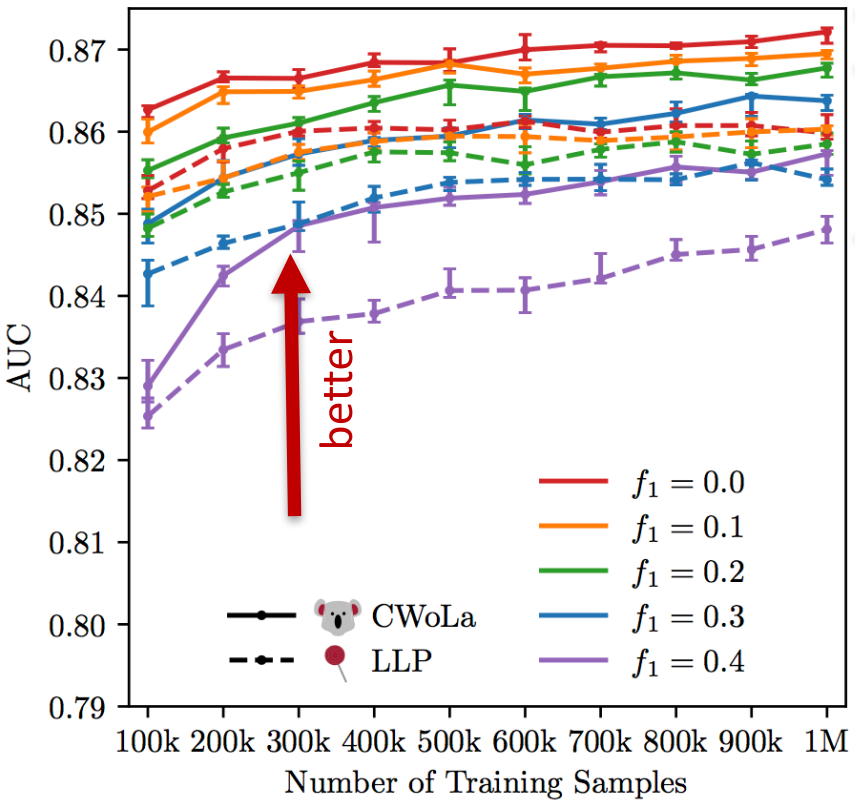


- **Weak learning** works with three Q/G discriminants

Weakly supervised works as well as fully supervised

Jet images + weak supervision

MDS, Komiske, Metodiev, Nachman (arXiv:1801.10158)



Fully supervised

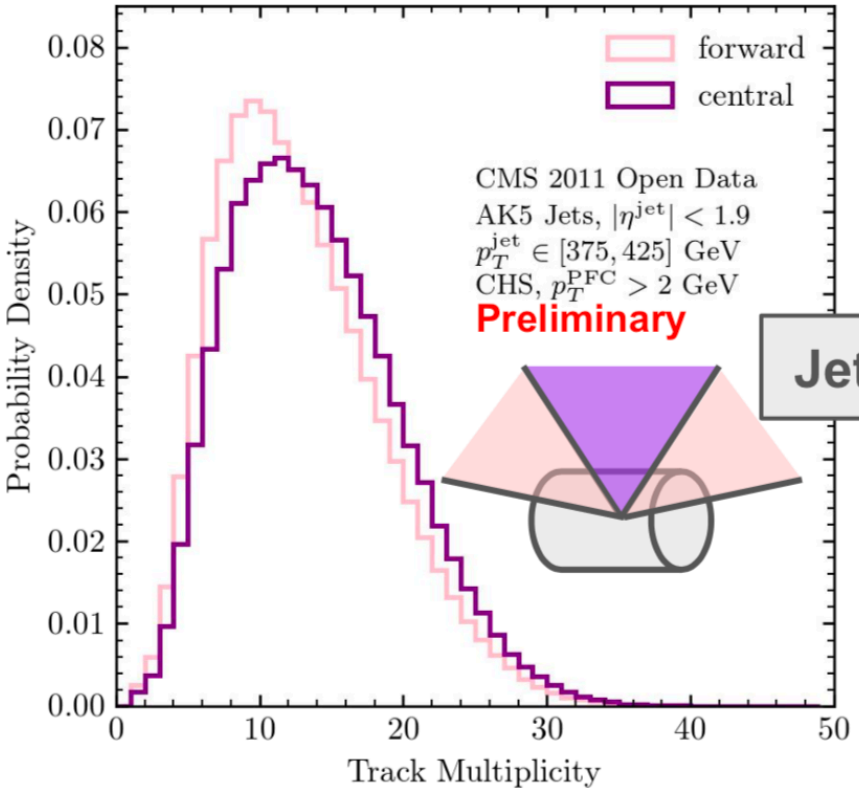
Weak supervision
(Mixed samples)

- Works as well as with full supervision
- Labels not needed even for complex inputs

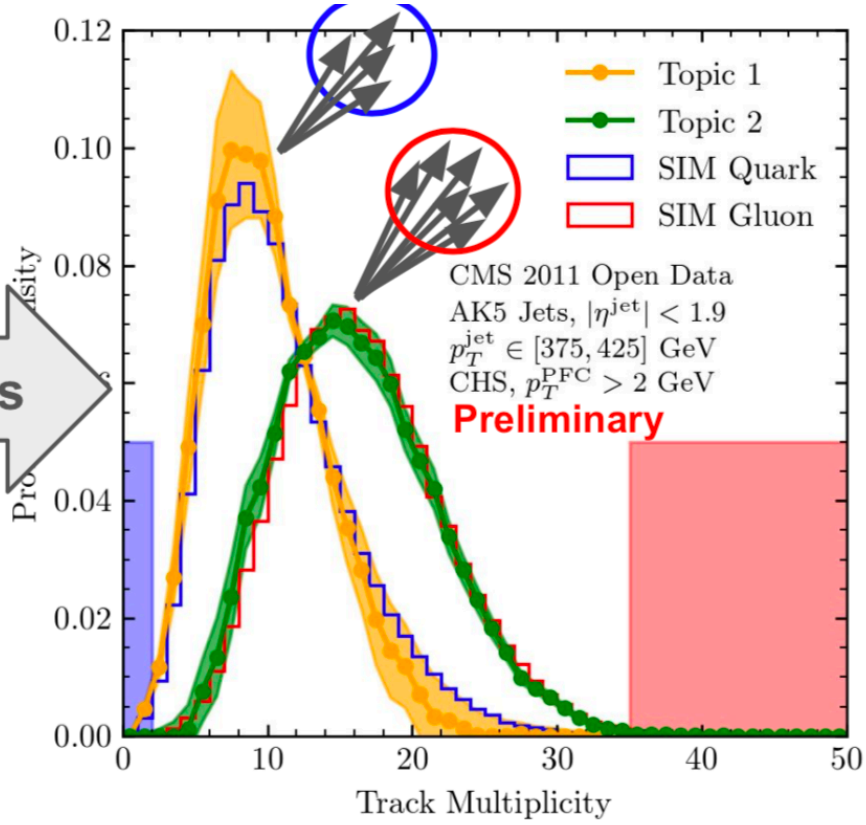
Weak supervision is a **breakthrough** for particle physics:
Can learn complex discrimination directly from data

Topic modeling: find two categories

uses CMS open data (Thaler et al., unpublished)



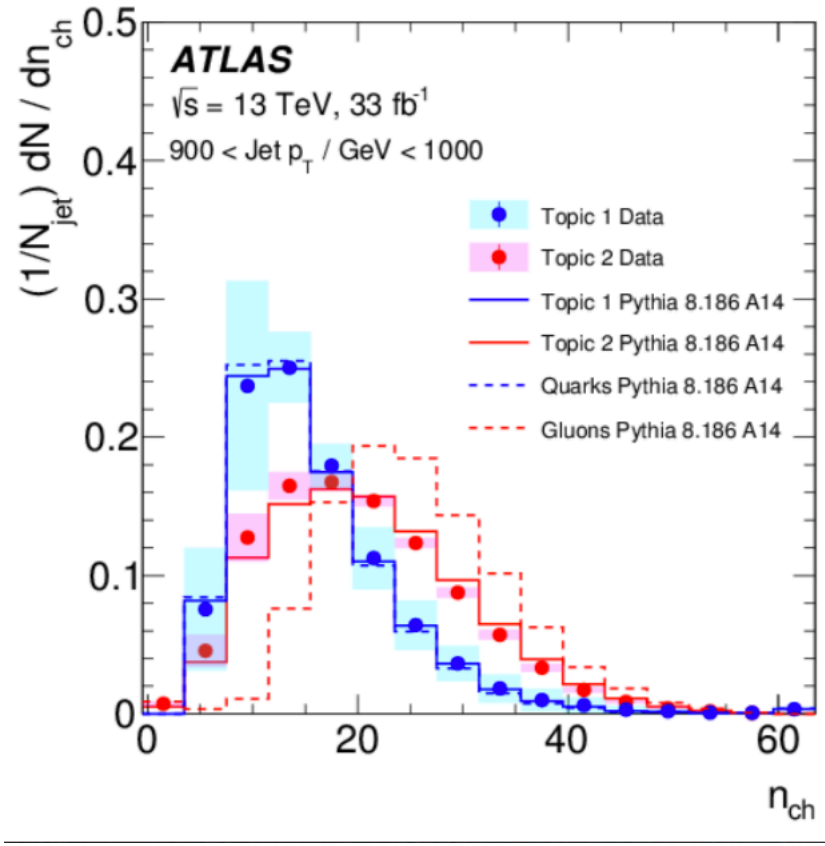
Jet Topics



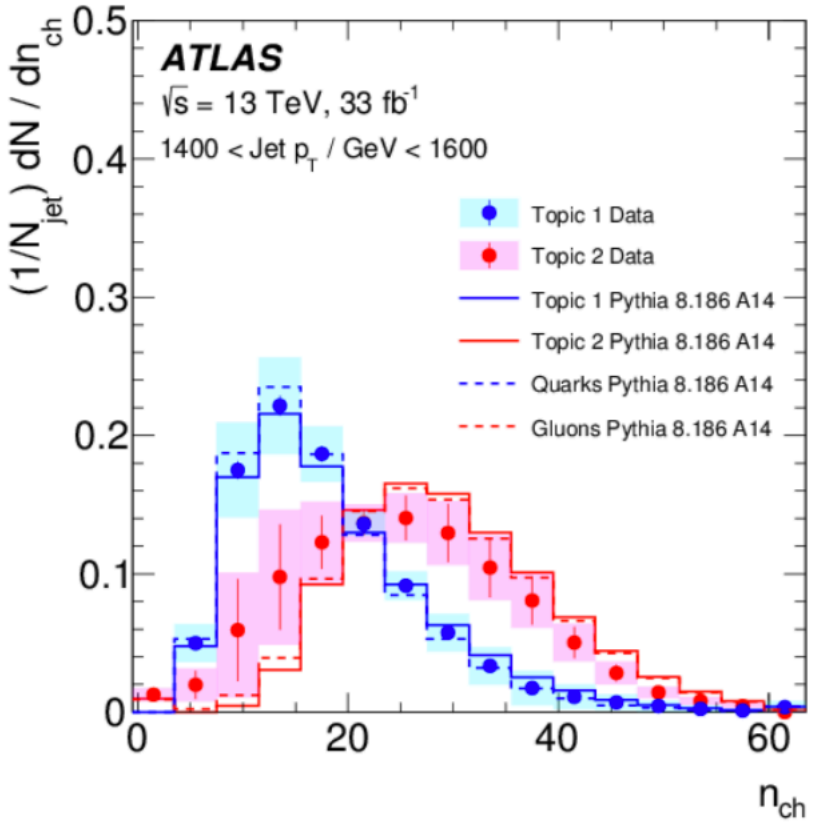
- Requires anchor bins that are pure topic A and pure topic B
- Requies sample independence

Topic modeling in ATLAS data: arXiv:1906.09254

1 TeV



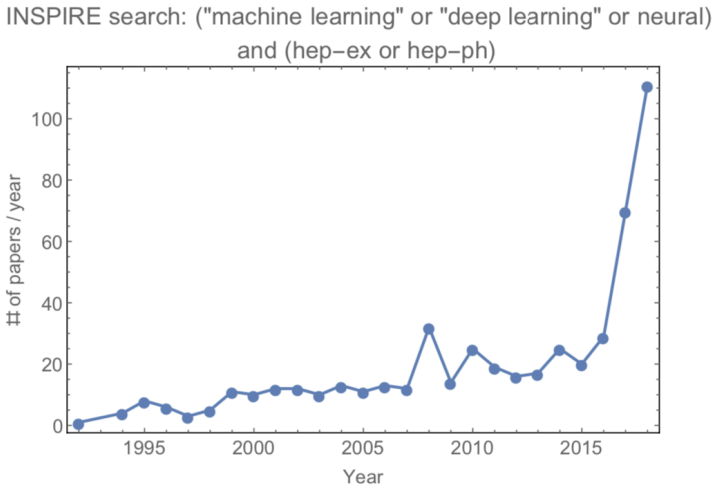
1.5 TeV



Are Topic 1 and topic 2 the same as quark and gluon?

Conclusions

- Modern machine learning is growing rapidly
- "Traditional" collider physics is dead



- Much progress on standard nails
  - top-tagging
  - quark/gluon
  - anomaly detection

- State-of-the-art methodology
  - Interpretable learning (JUNIPR)
  - Unordered representations (point clouds …)
  - ML directly on low-level data

- Past: apply hammers to nails

Matthew Sc

- Future: learn some new physics