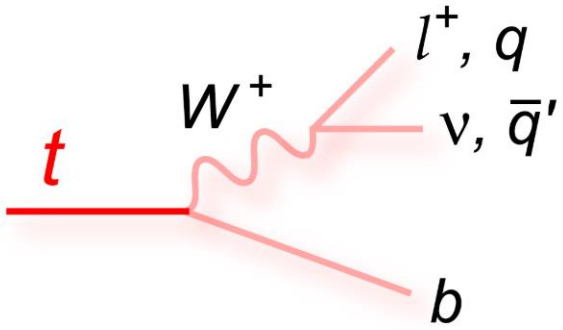
Machine Learning for Top Tagging at ATLAS



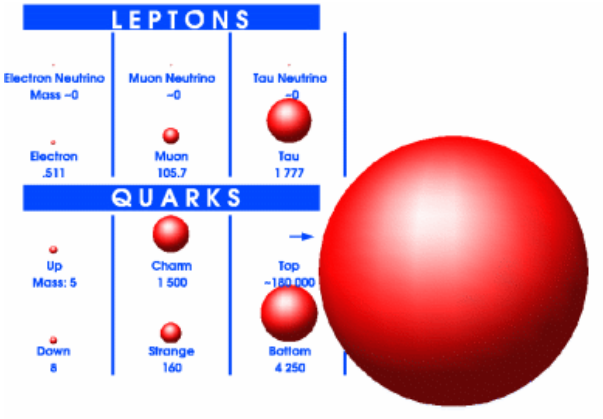
Alex Wen
ATLAS Group
University of British Columbia

CUPC 2019 – Montreal, QC
7-10 November 2019

ATLAS EXPERIMENT

THE UNIVERSITY OF BRITISH COLUMBIA

Top Quarks - Production

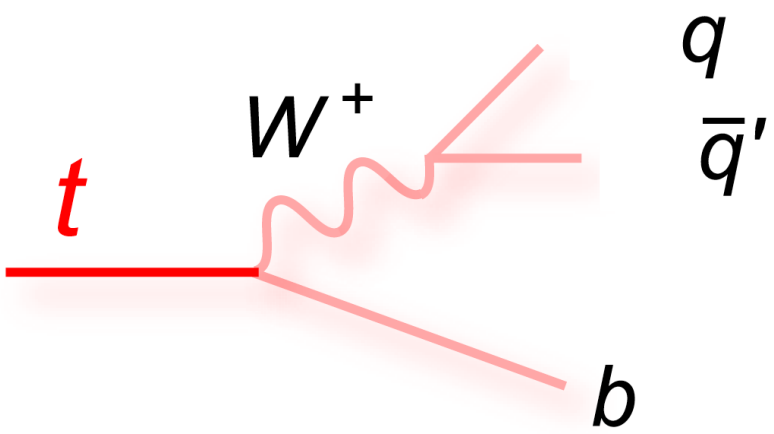


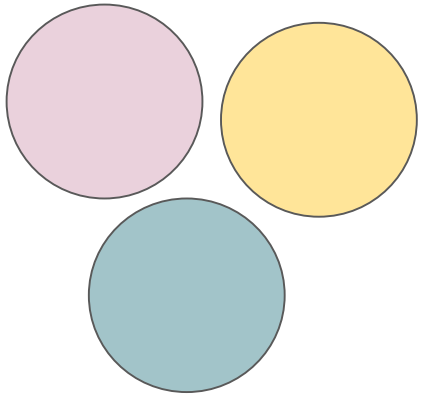
173 GeV!

Re-187 atom

Images: Fermilab

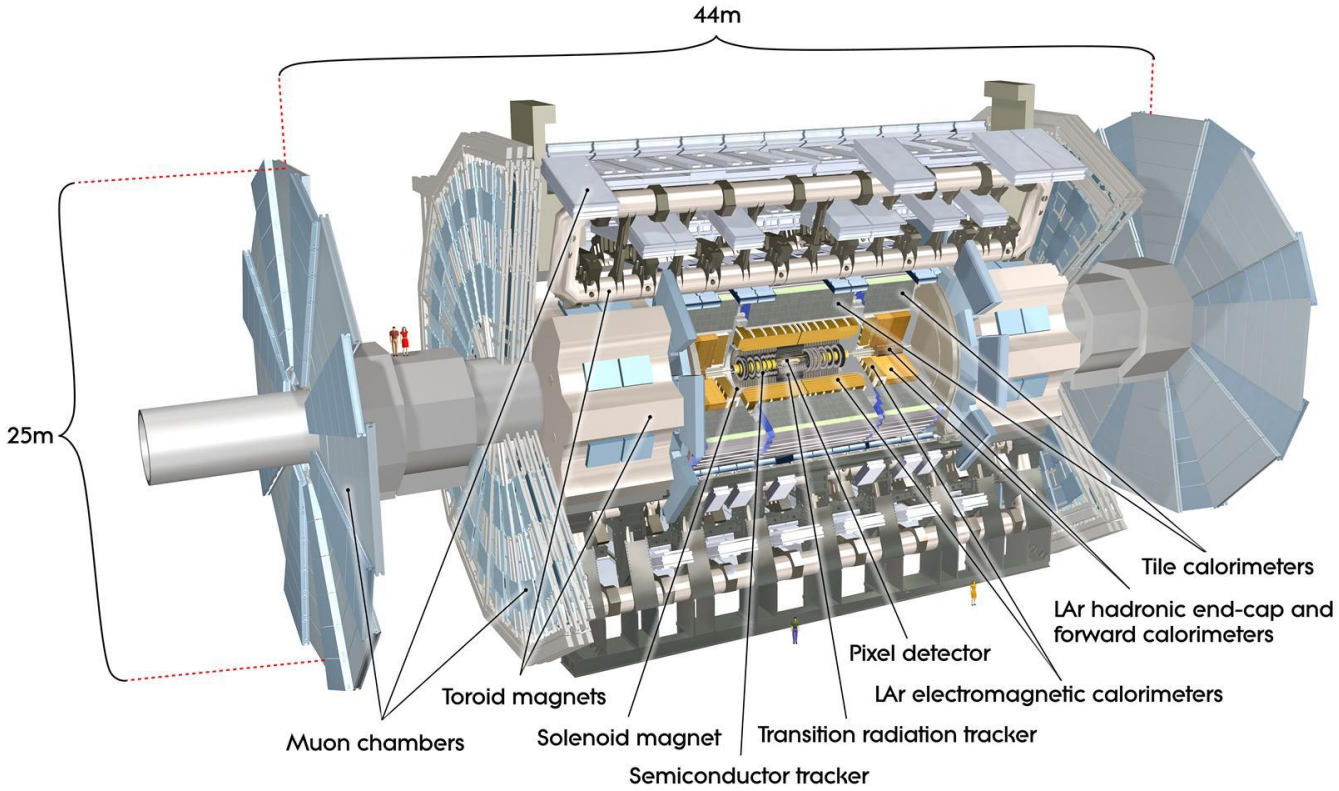
Top Quarks - Decay





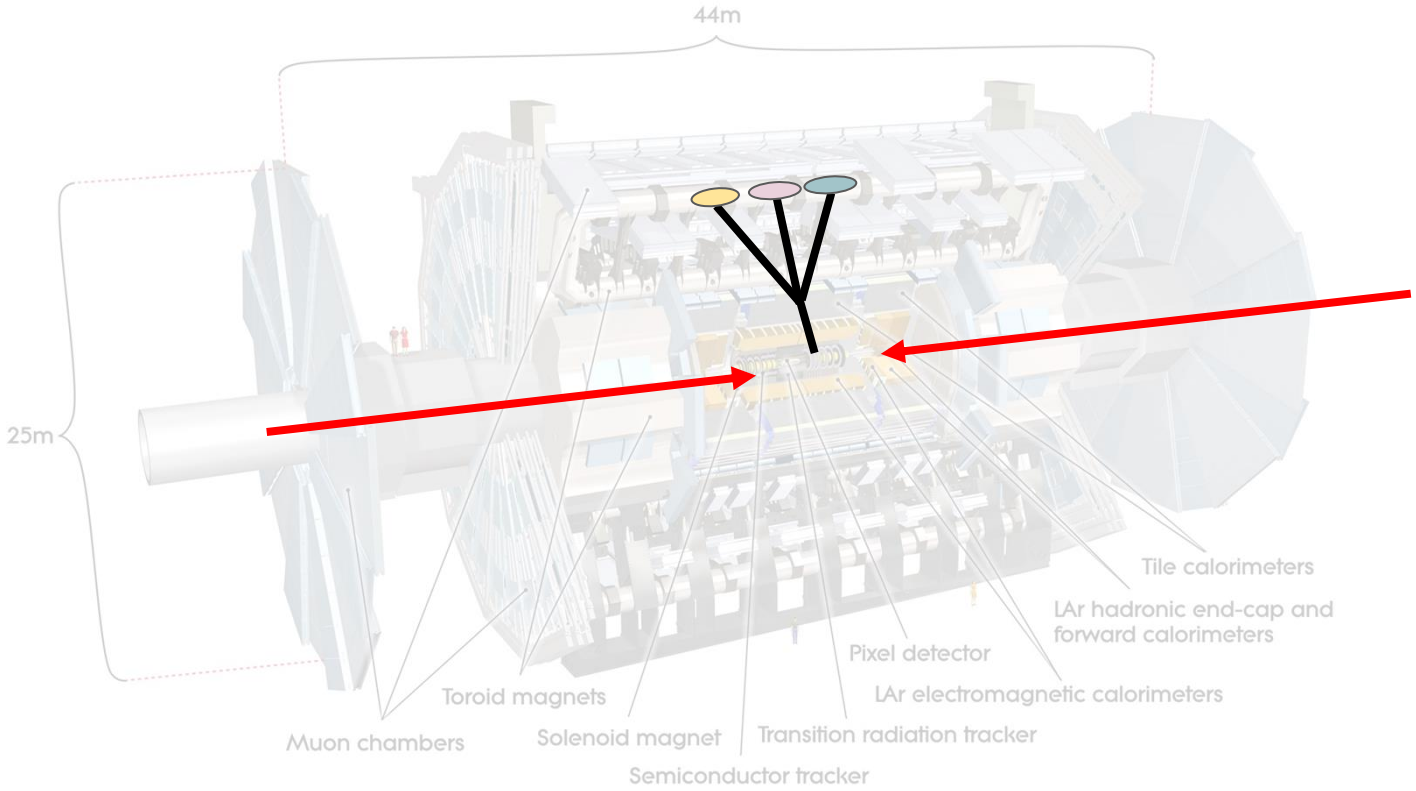
Images: Fermilab

ATLAS Experiment

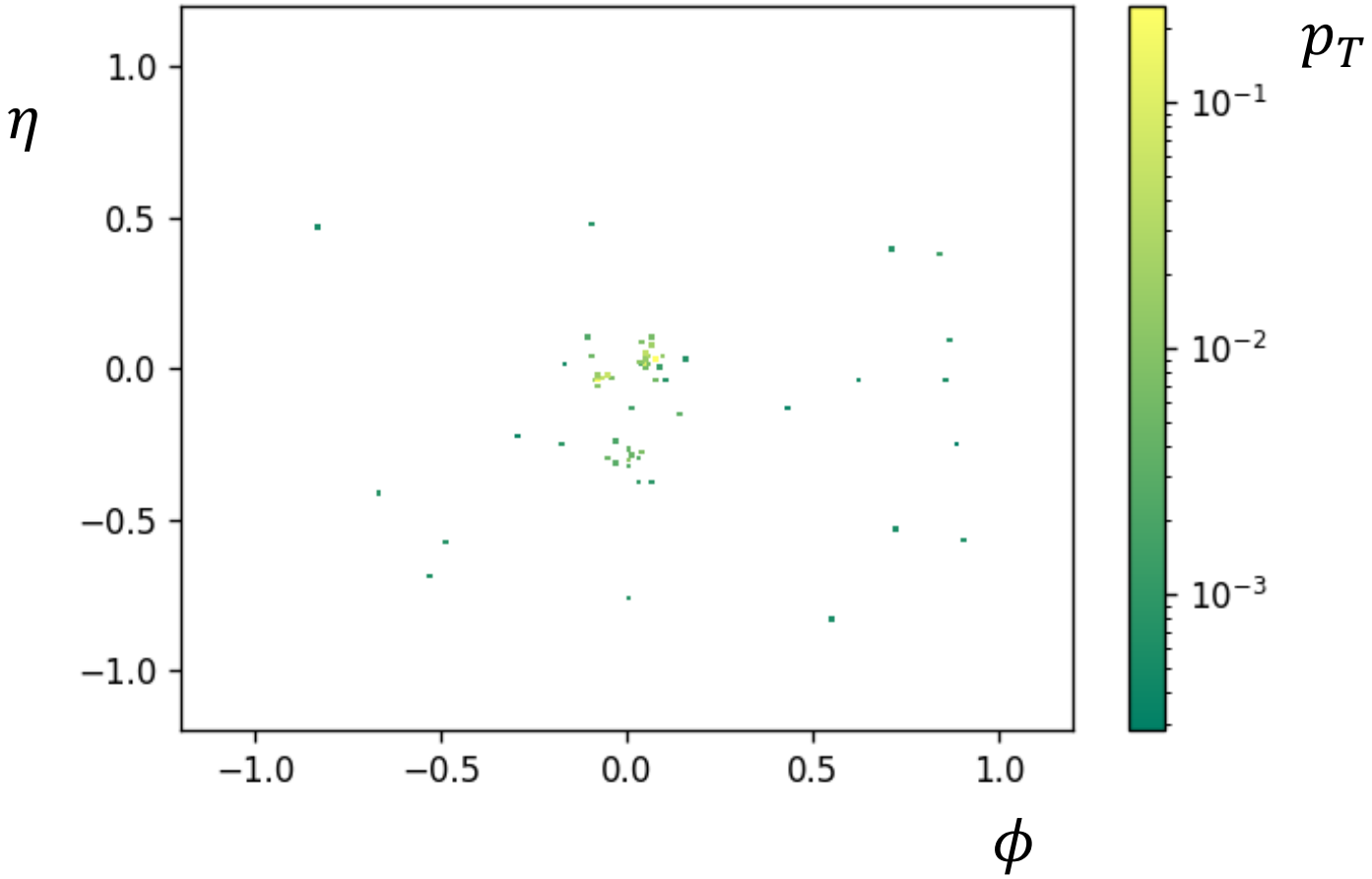


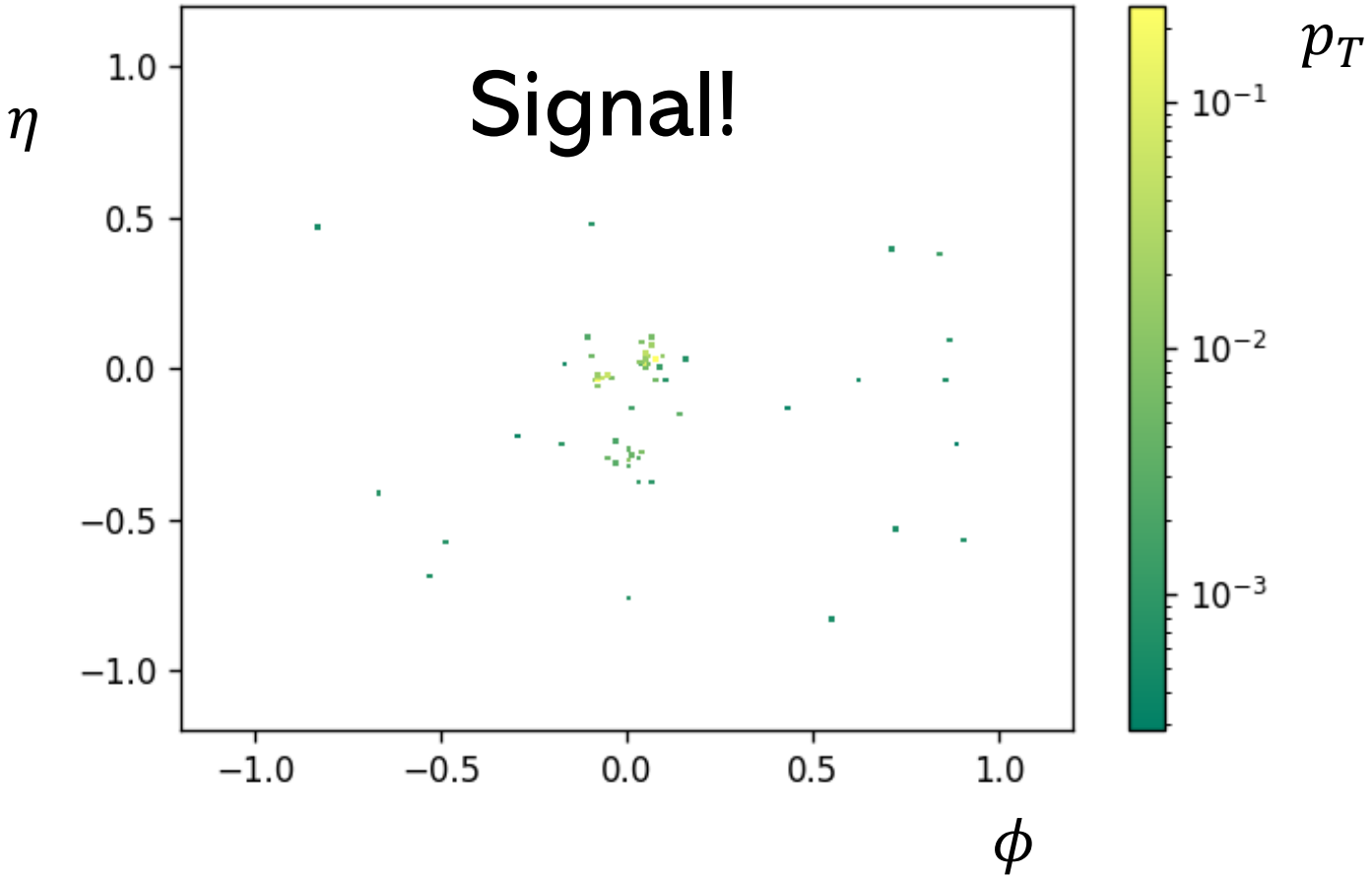
Images: CERN

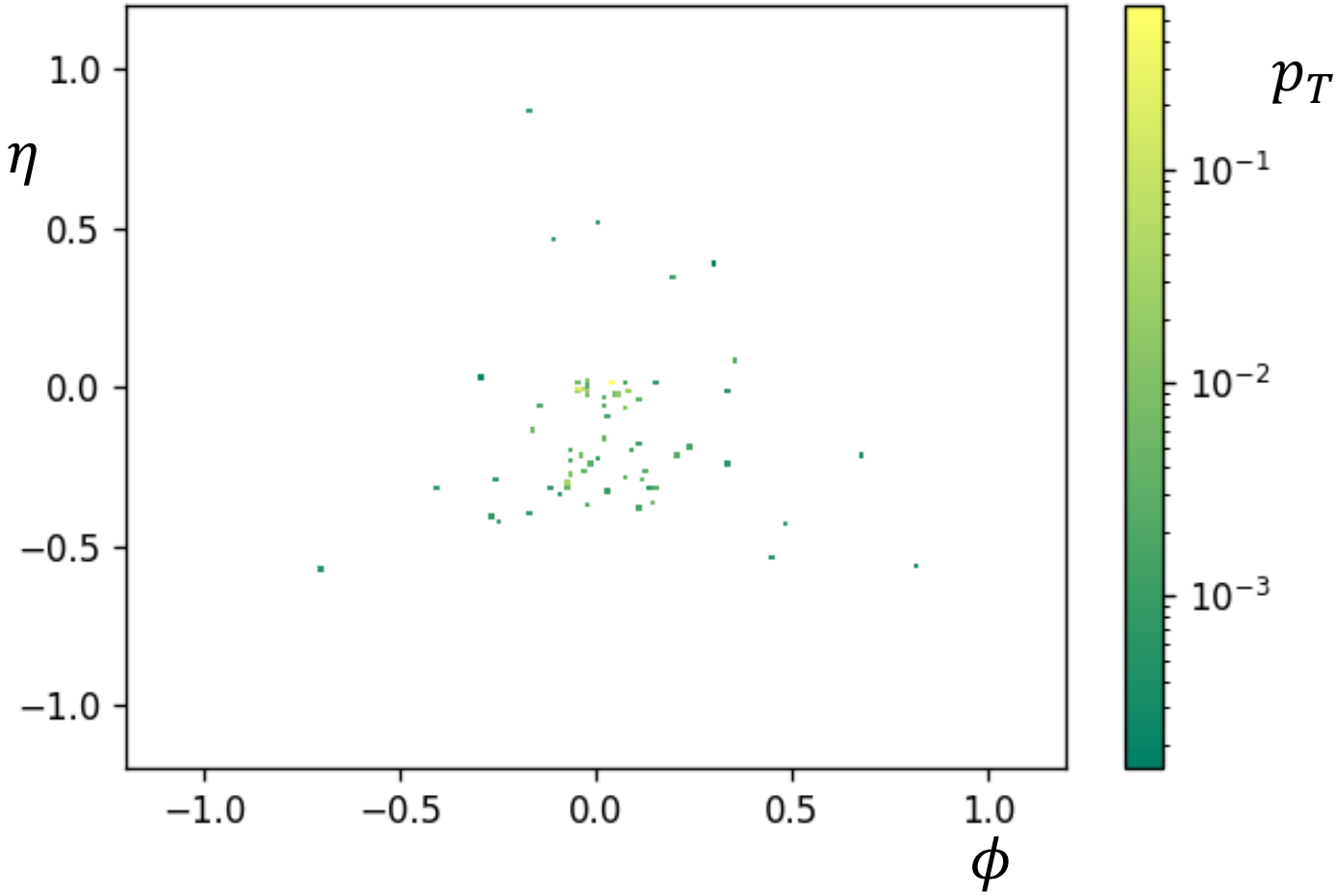
ATLAS Experiment

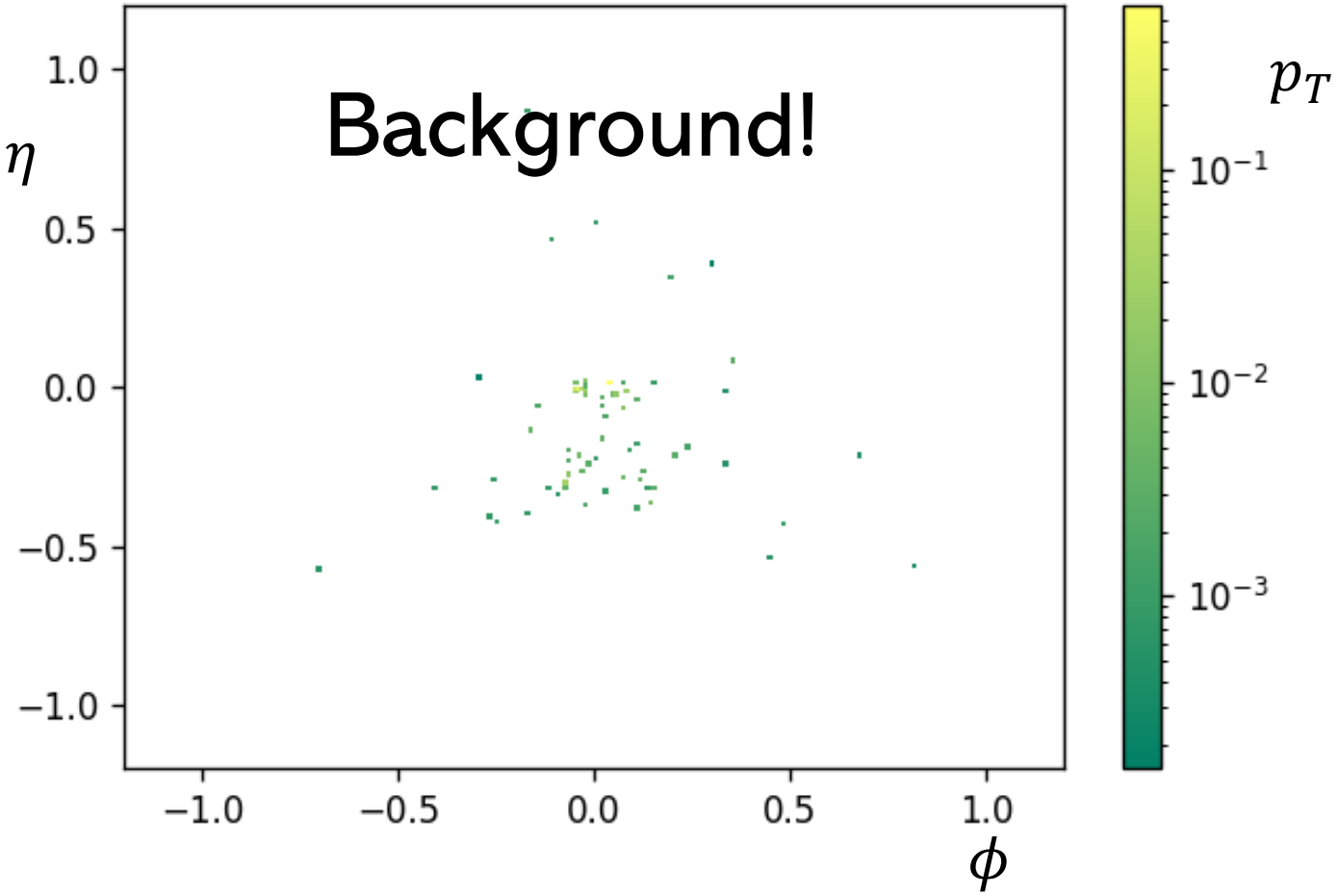


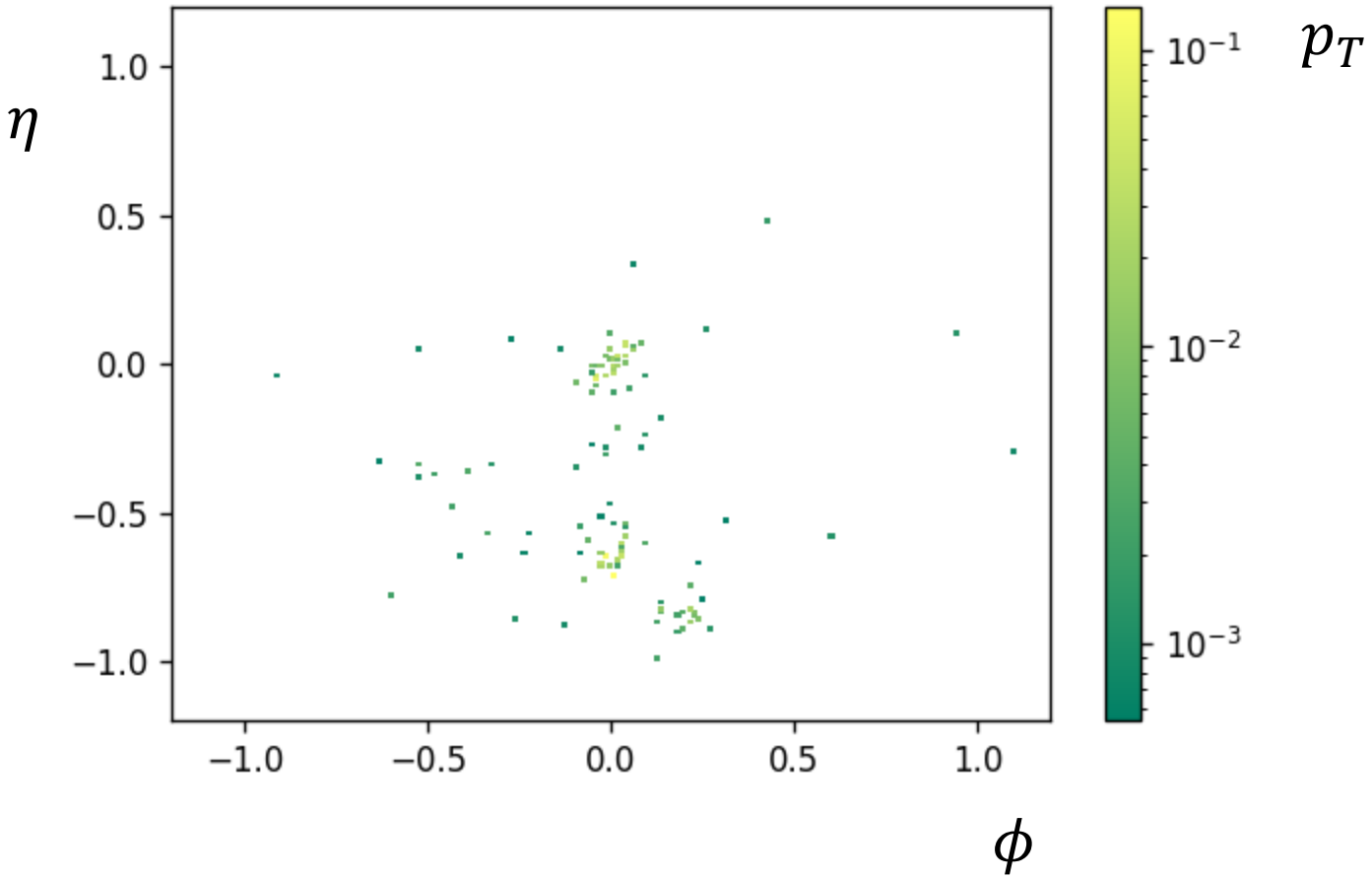
Images: CERN

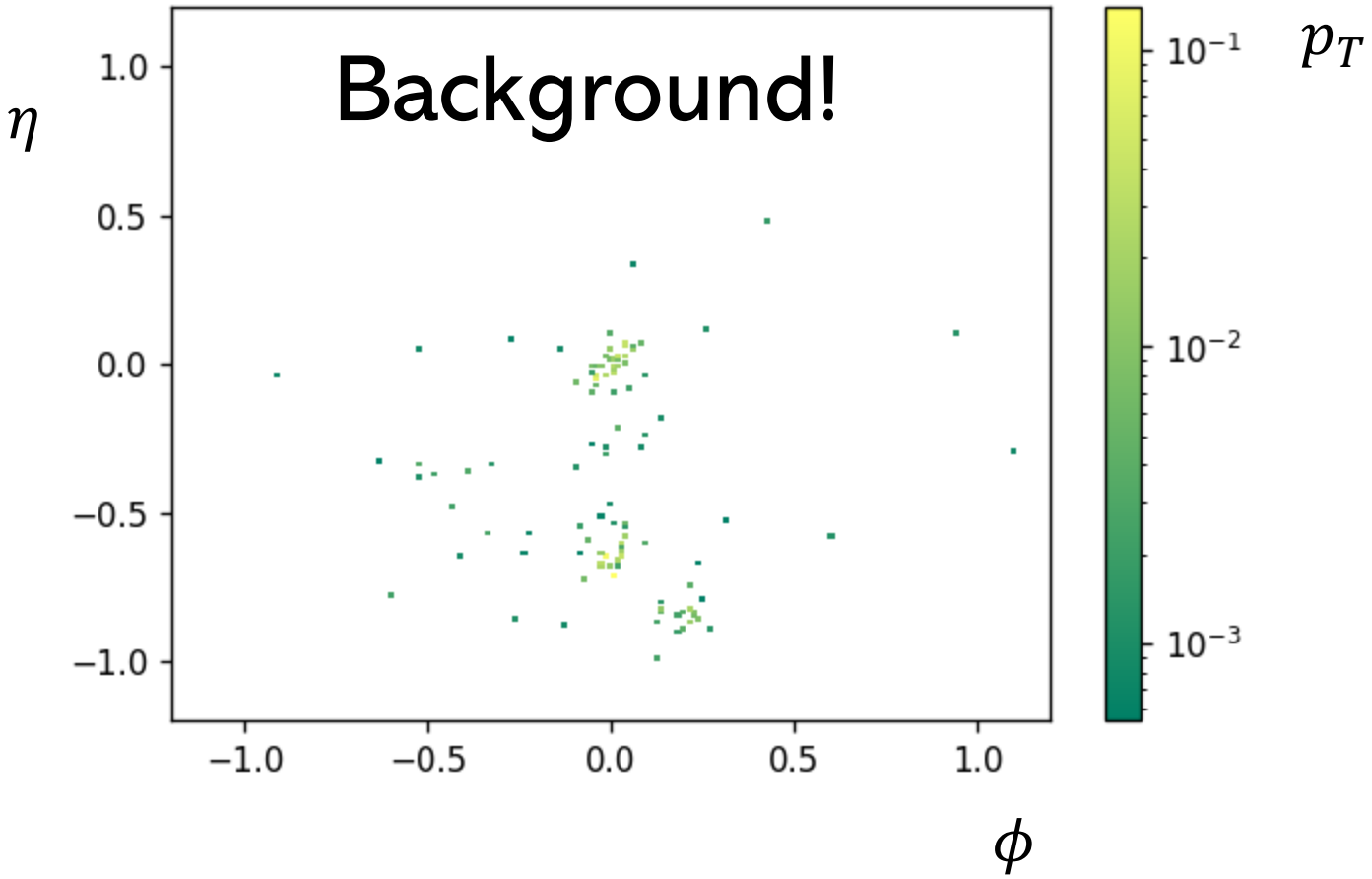

η
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
ϕ
p_T
10^{-1}
10^{-2}
10^{-3}


Signal!
η
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
ϕ
p_T
10^{-1}
10^{-2}
10^{-3}

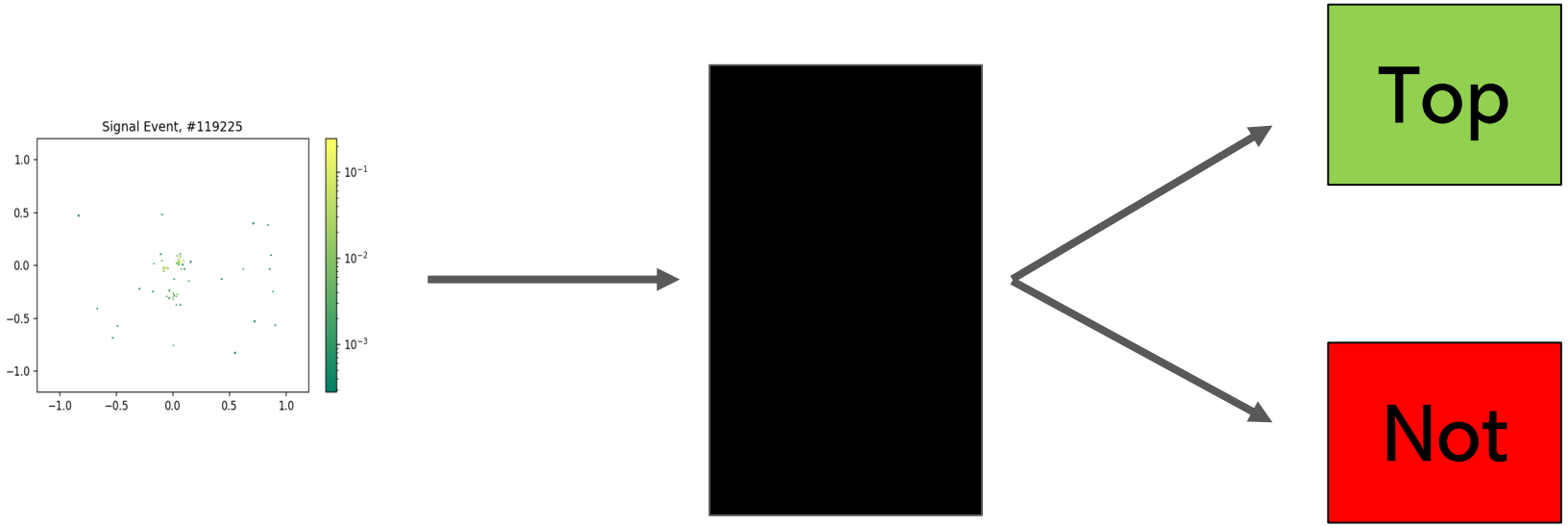

1.0
0.5
0.0
−0.5
−1.0
η
−1.0
−0.5
0.0
0.5
1.0
ϕ
p_T
10^{-1}
10^{-2}
10^{-3}

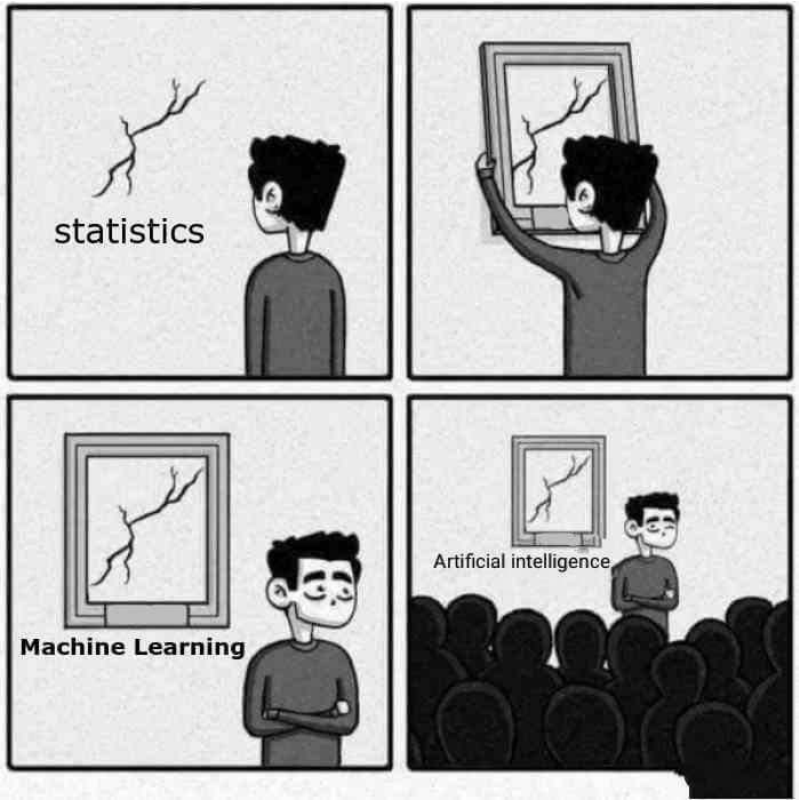

Background!
η
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
ϕ
p_T
10^{-1}
10^{-2}
10^{-3}

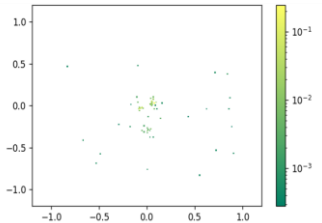

η
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
ϕ
10^{-1}
10^{-2}
10^{-3}
p_T


Background!
η
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
ϕ
10^{-1}
10^{-2}
10^{-3}
p_T

Want:


Signal Event, #119225
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
10^{-1}
10^{-2}
10^{-3}
Top
Not

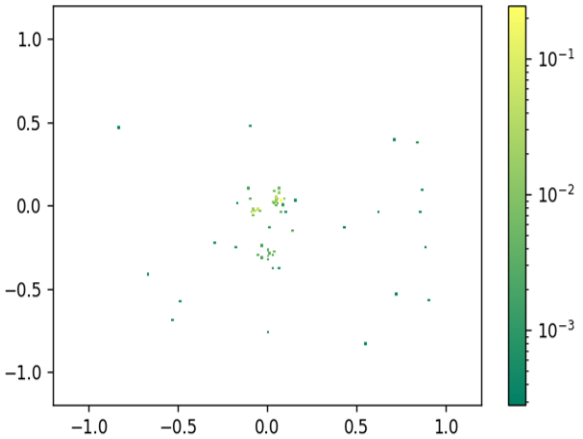

statistics
Machine Learning
Artificial intelligence

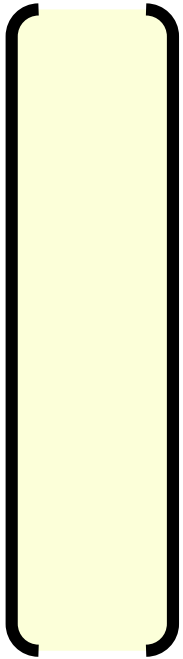
$$f($$  $$) =$$

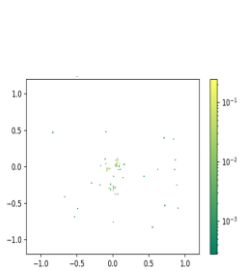
1

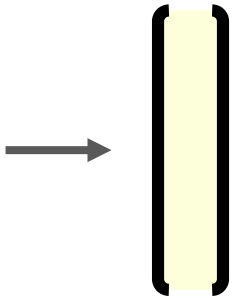
0

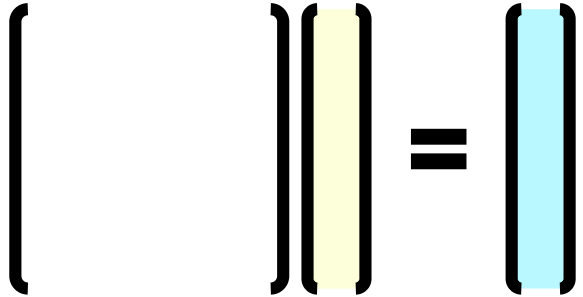
Idea:

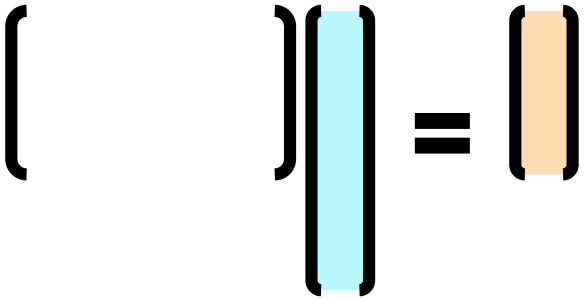


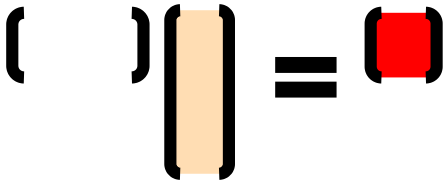


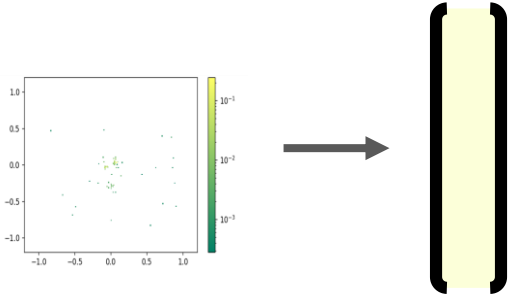












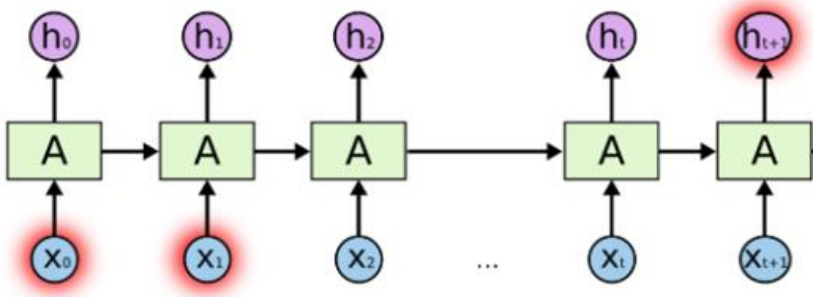
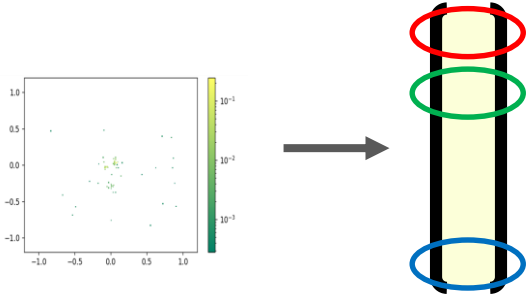


Image: C. Olah



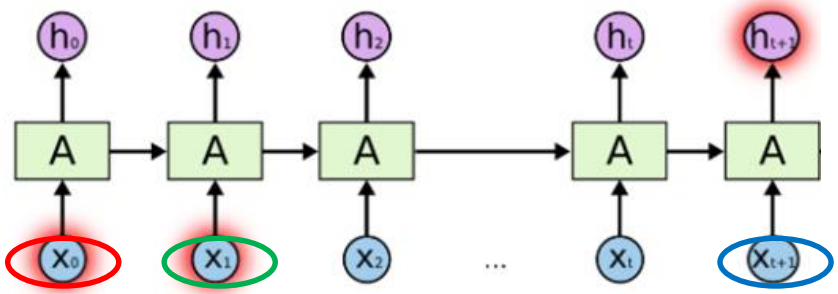
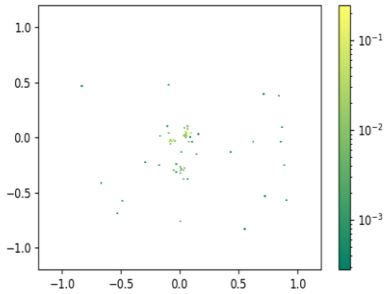
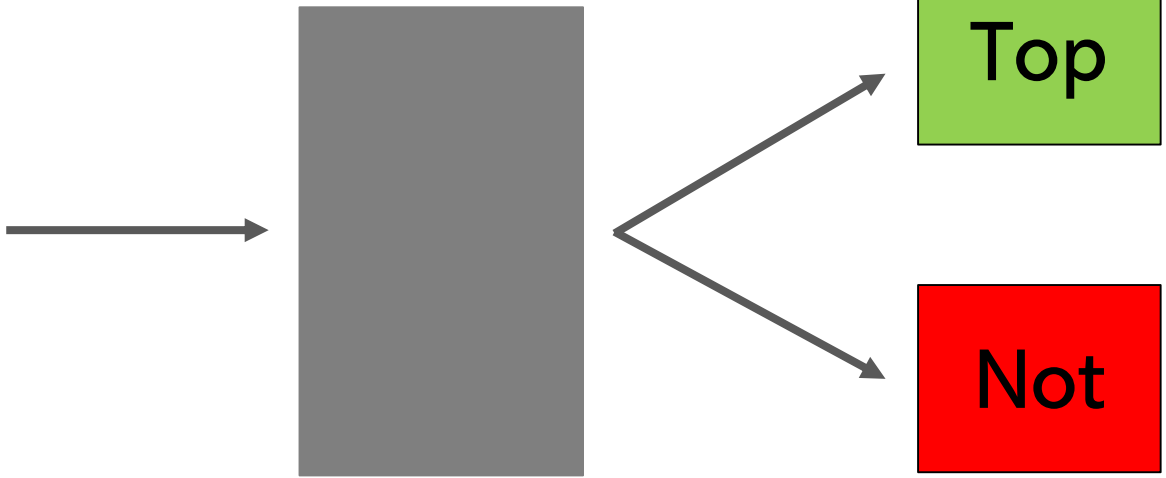
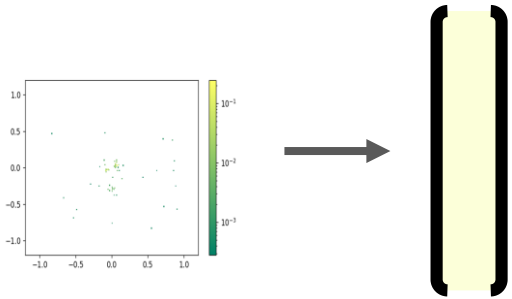


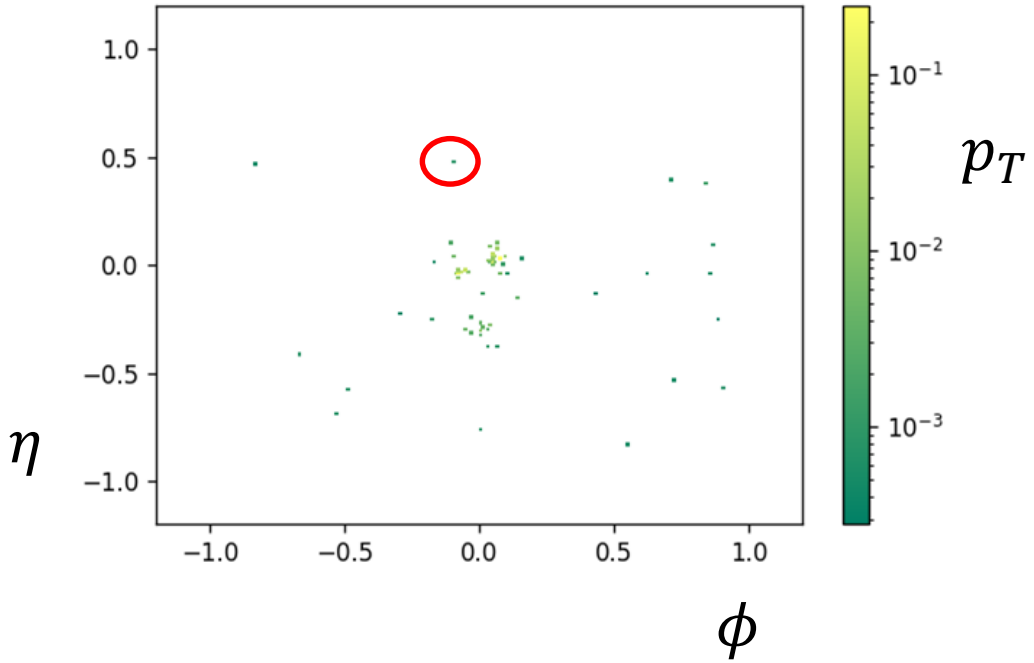
Image: C. Olah

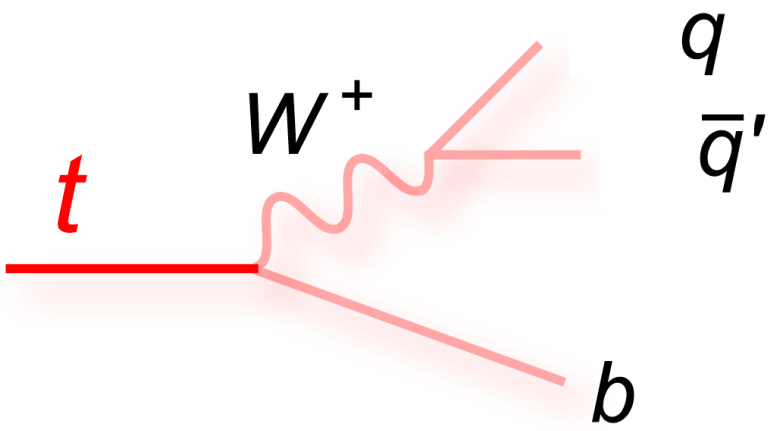

1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
10^{-1}
10^{-2}
10^{-3}


Top
Not


1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
10^{-1}
10^{-2}
10^{-3}

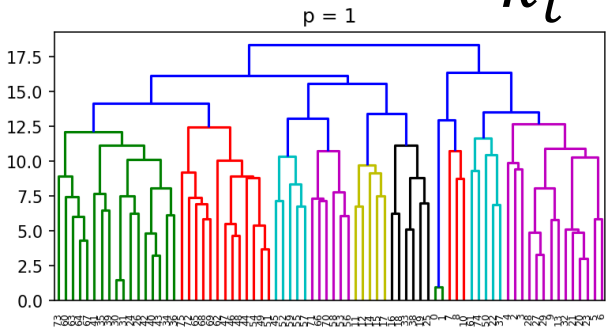
p_T, η, ϕ

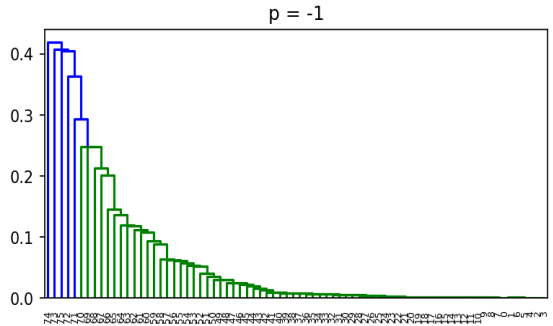




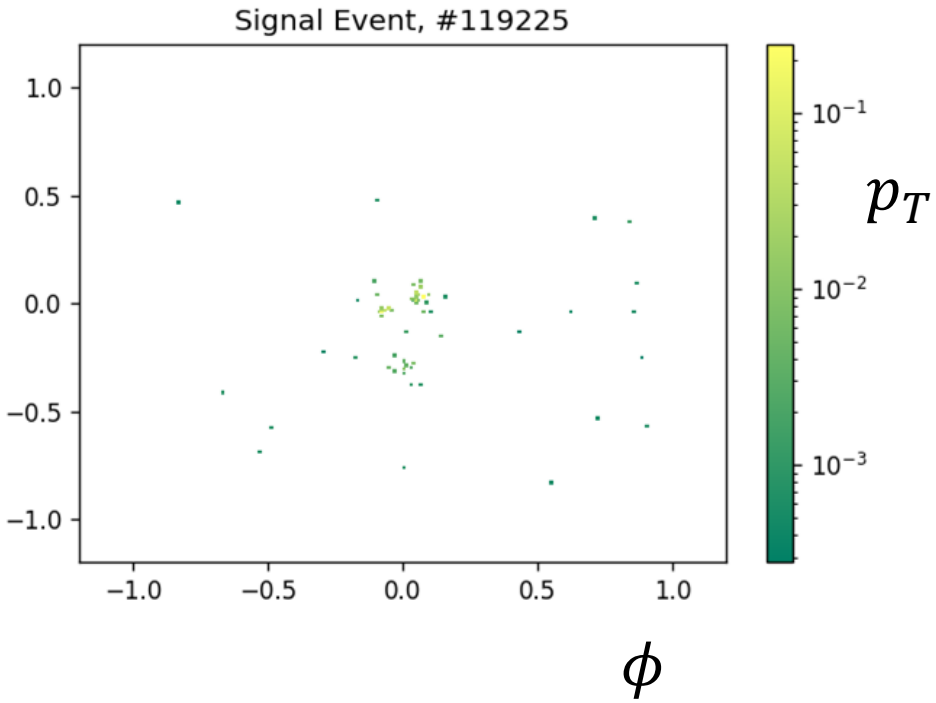
Images: Fermilab

k_t

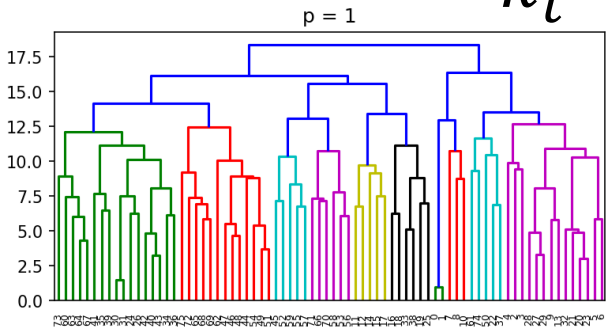

p = 1
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0

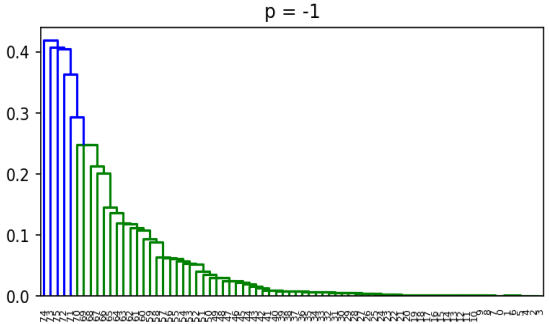

p = -1
0.4
0.3
0.2
0.1
0.0

$anti - k_t$

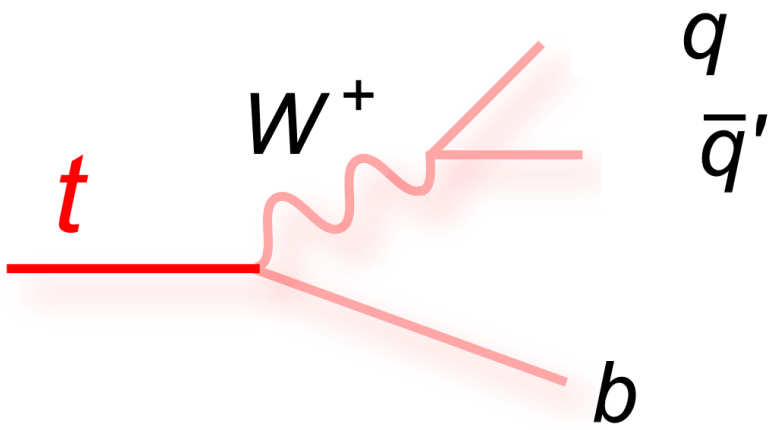

Signal Event, #119225
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
10^{-1}
10^{-2}
10^{-3}
p_T
η
ϕ

k_t

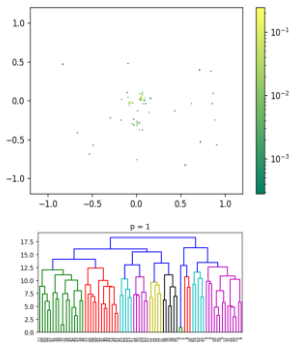




$anti - k_t$

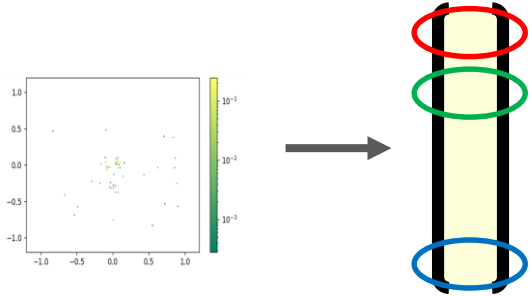


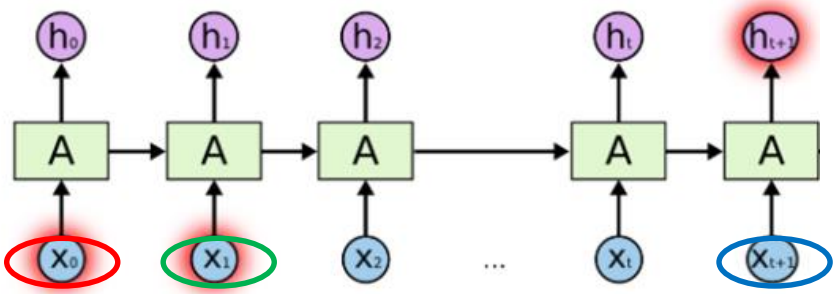
Images: Fermilab

$$f($$  $$) =$$

1

0





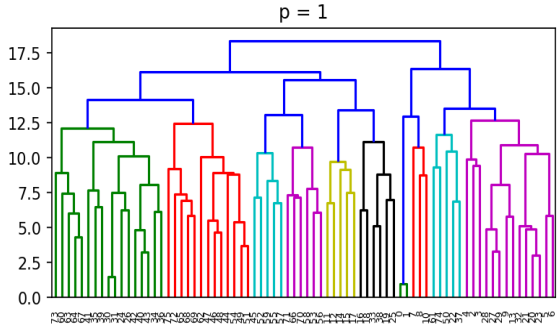
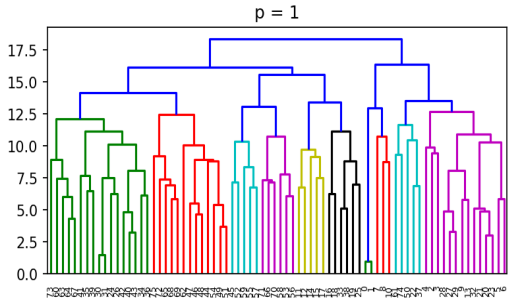
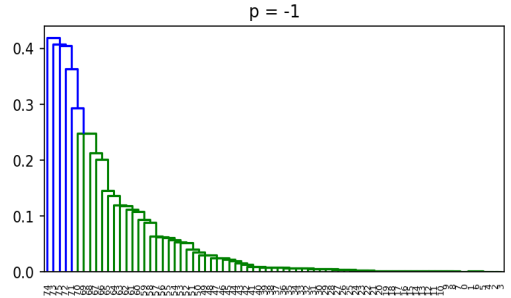


Image: C. Olah

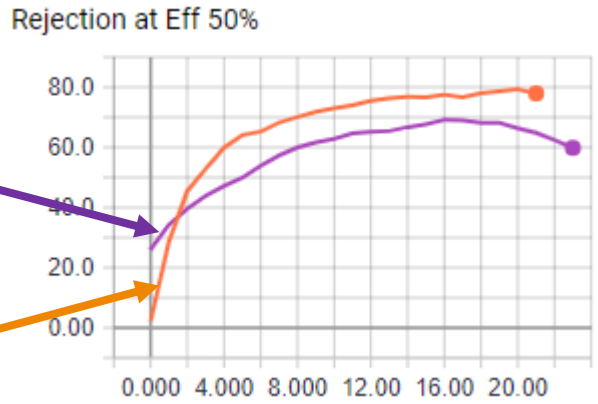
Effects of Clustering





k_t

$anti - k_t$



Takeaways:

- ML is critical

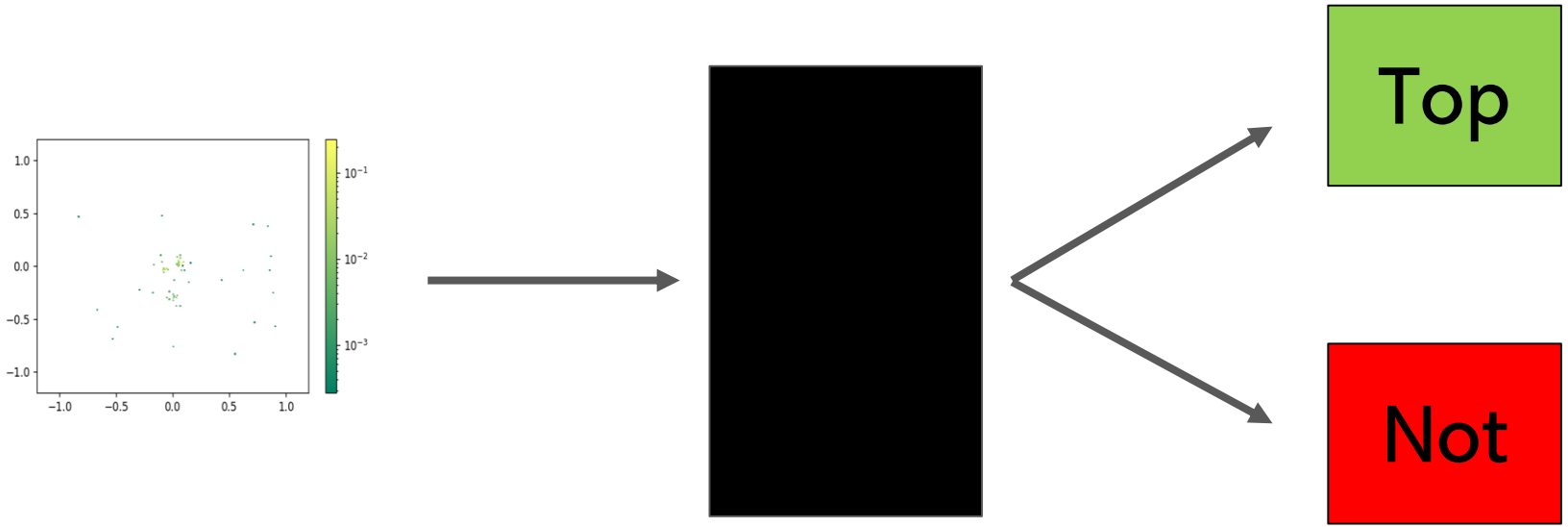
Takeaways:

- ML is critical
- We can inject physics into analysis

Takeaways:

- ML is critical
- We can inject physics into analysis
- But it doesn’t necessarily help!

Want:



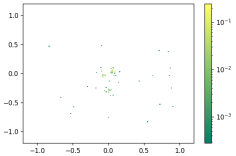
Thanks!

- Alison Lister and Colin Gay of UBC ATLAS.
- Past students Kevin Zhang, Daniel Hortela, Shannon Egan, Jannicke Pearkes.

Backup Slides

Overview

~ 2 M events

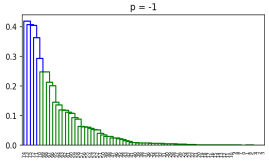


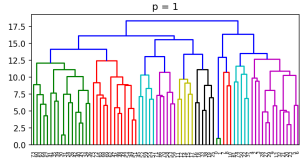
Data

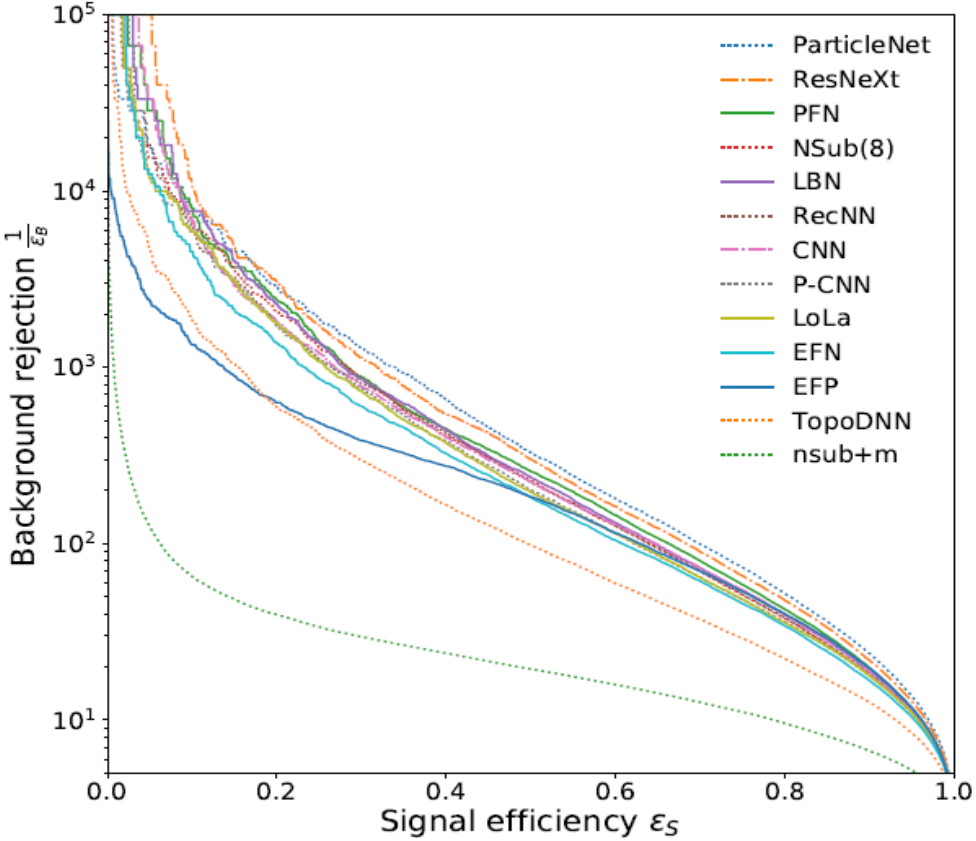
Jet Constituents
“Particles”

Clustering Order
“Tree”

Neural Network &
prediction (top or not)

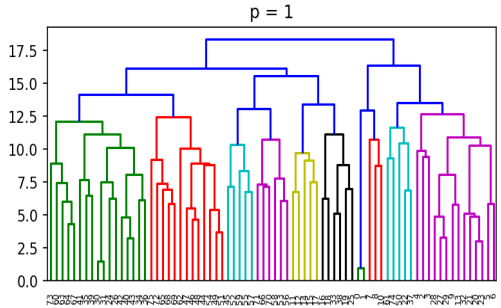


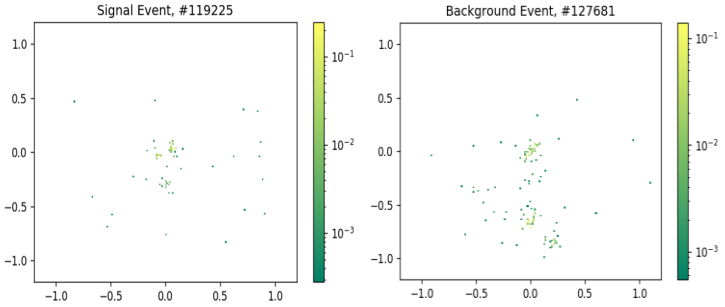


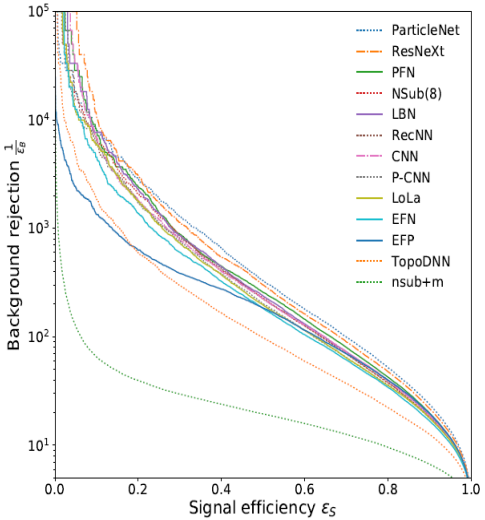


A. Butter et al. The Machine Learning Landscape of Top Taggers. 2019.

What kind of neural network?







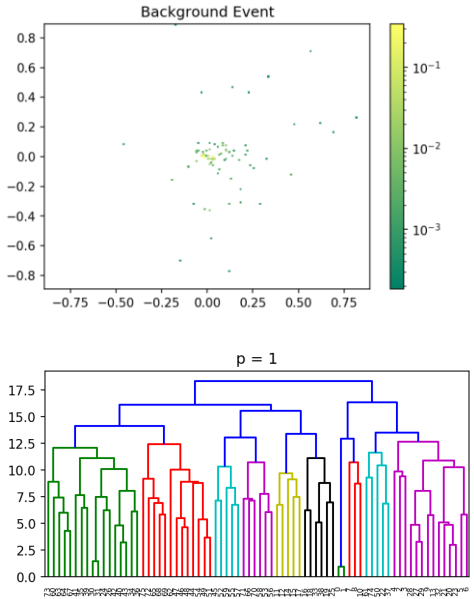
Tree structure & combination of constituents

Really complicated data

Competition - some models have
>1000 rejection @ 30% signal eff.

Image from [2]

Stack-augmented Parser Interpreter (SPINN)



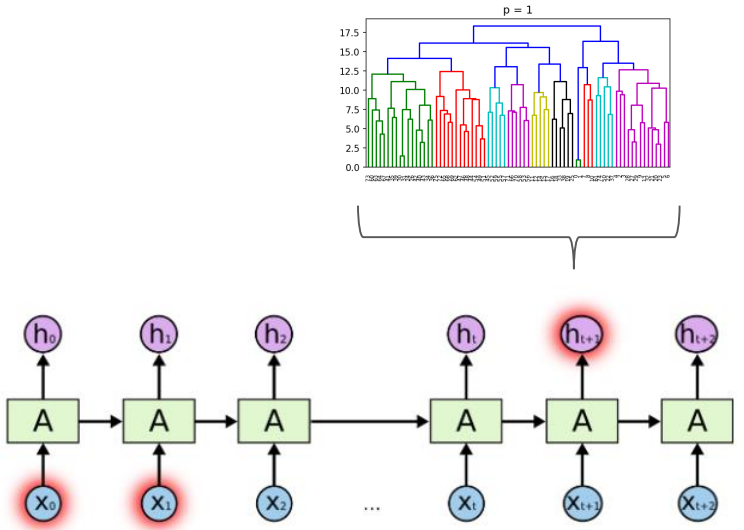
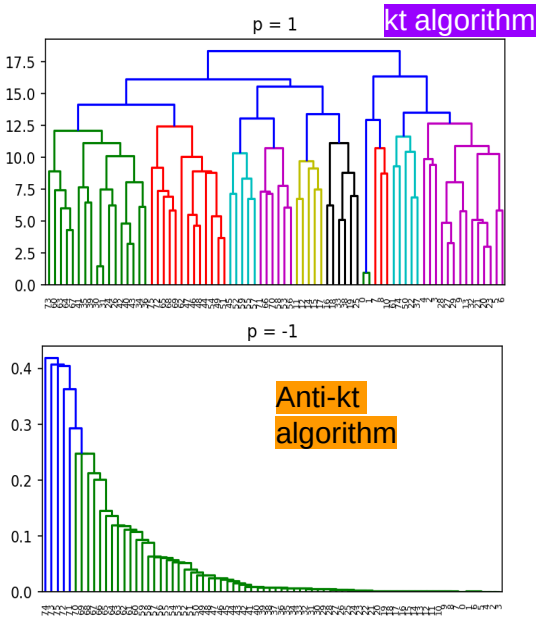


Image: C. Olah

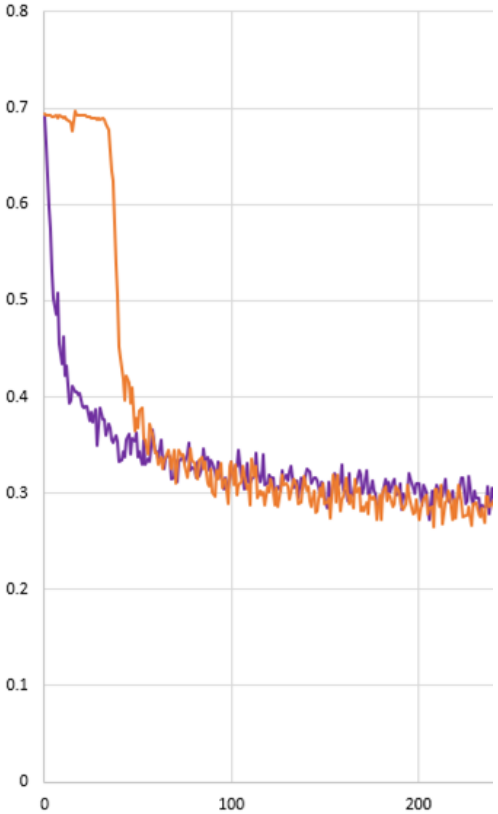
0001001000101010100101010100101111

Samuel R. Bowman, Jon Gauthier, Abhinav Rastogi, Raghav Gupta, Christopher D. Manning, and Christopher Potts. A fast unified model for parsing and sentence understanding. *CoRR*, abs/1603.06021, 2016.

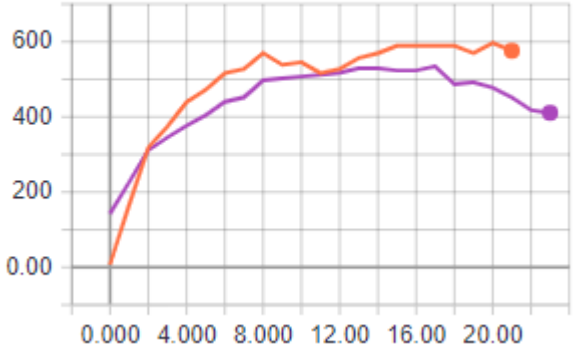
Results



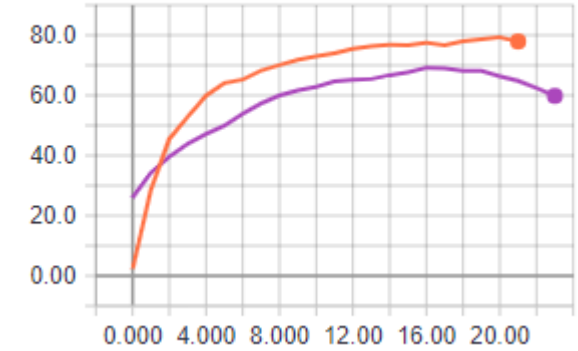
Loss in Time



Rejection at Eff 20%



Rejection at Eff 50%



[2] Samuel R. Bowman, Jon Gauthier, Abhinav Rastogi, Raghav Gupta, Christopher D. Manning, and Christopher Potts. A fast unified model for parsing and sentence understanding. *CoRR*, abs/1603.06021, 2016.

[3] A. Butter et al. The Machine Learning Landscape of Top Taggers. 2019.

[4] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

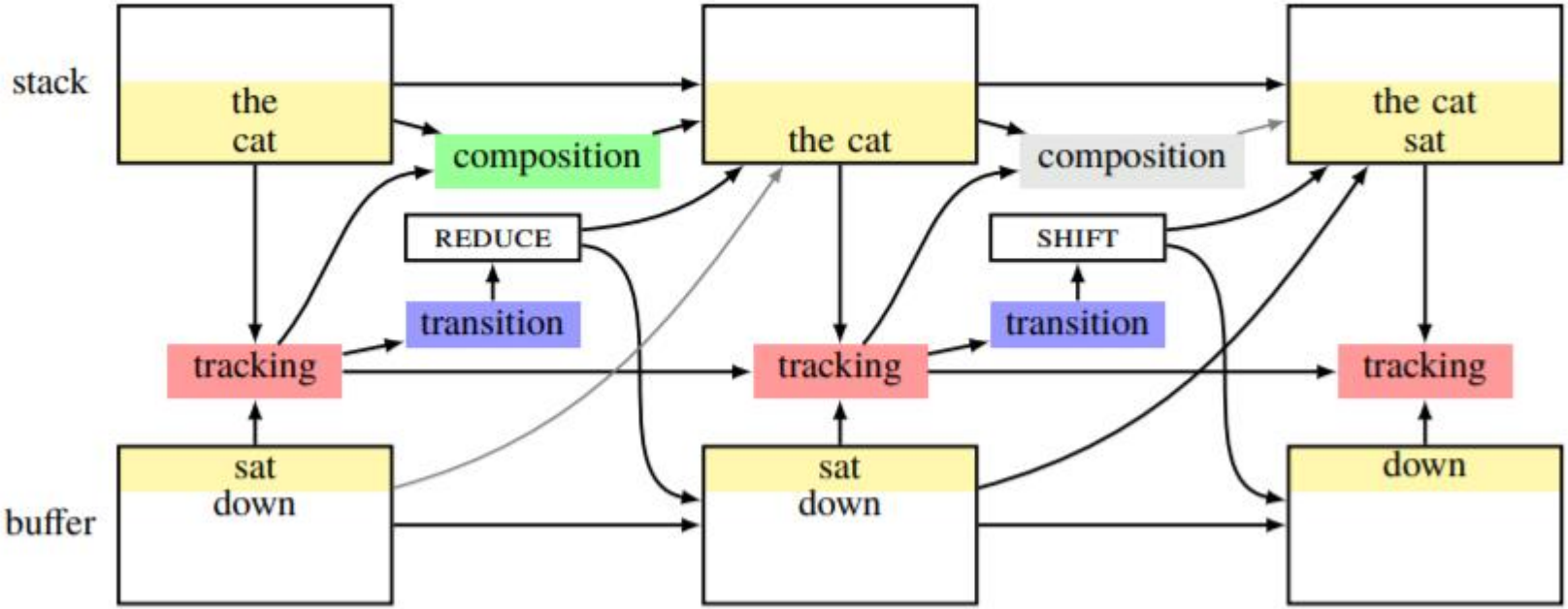
[5] Jannicke Pearkes, Wojciech Fedorko, Alison Lister, and Colin Gay. Jet Constituents for Deep Neural Network Based Top Quark Tagging. 2017.

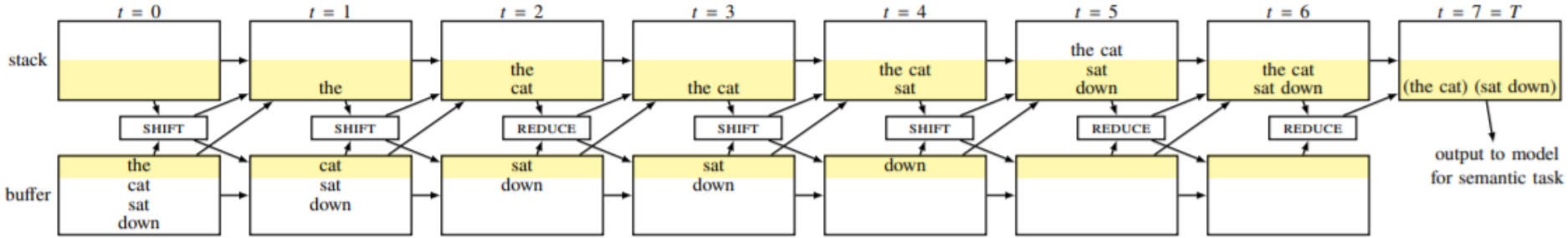
[6] M. Tanabashi et al. Review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018.

$$ak_t\text{:}\ d_{ij} = \min\left(p_i^{-2}, p_j^{-2}\right)\frac{\Delta^2}{R^2}$$

$$k_t\text{:}\ d_{ij} = \min\left(p_i^{2}, p_j^{2}\right)\frac{\Delta^2}{R^2}$$

$$\Delta^2 = \left(\eta_i - \eta_j\right)^2 + \left(\phi_i - \phi_j\right)^2$$

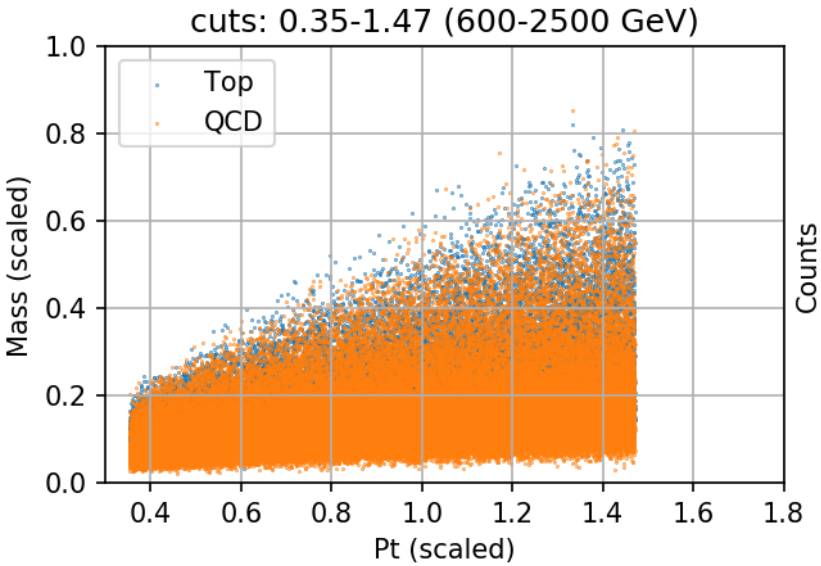

stack
the
cat
composition
the cat
composition
the cat
sat
REDUCE
transition
SHIFT
transition
tracking
tracking
tracking
buffer
sat
down
sat
down
down

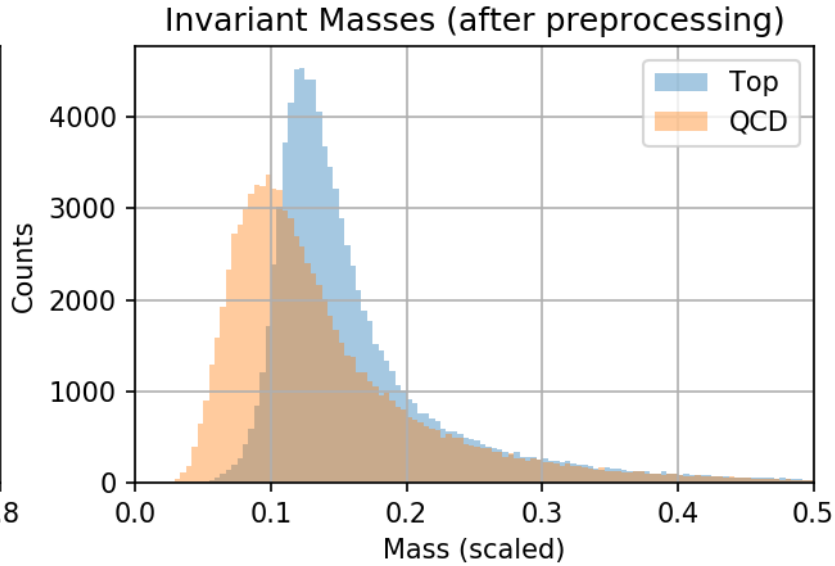


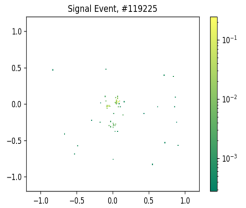
(b) The fully unrolled SPINN for *the cat sat down*, with neural network layers omitted for clarity.

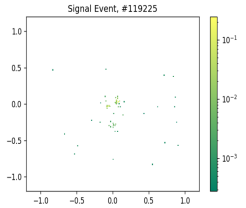
Paper for reference

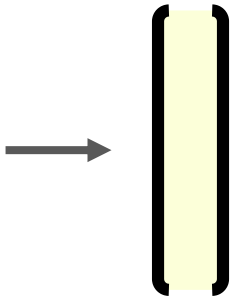
- Bowman et al.: A Fast Unified Model for Parsing and Sentence Understanding https://nlp.stanford.edu/pubs/bowman2016spinn.pdf

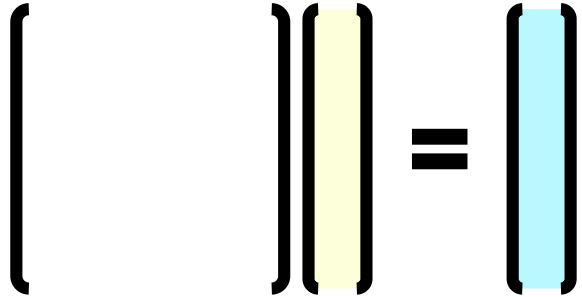

cuts: 0.35-1.47 (600-2500 GeV)
Top
QCD
Mass (scaled)
Pt (scaled)
0.0
0.2
0.4
0.6
0.8
1.0
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8

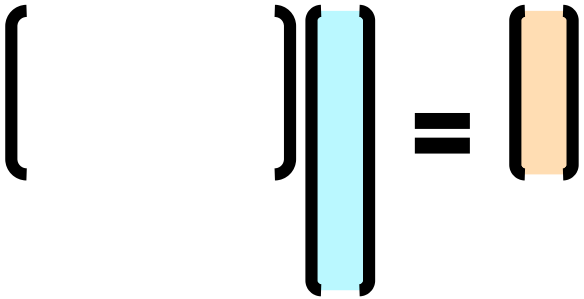

Invariant Masses (after preprocessing)
Top
QCD
Counts
Mass (scaled)
0
1000
2000
3000
4000
0.0
0.1
0.2
0.3
0.4
0.5

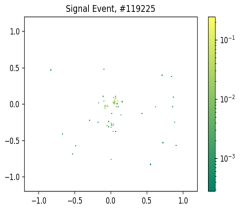

Signal Event, #119225
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
10^{-1}
10^{-2}
10^{-3}

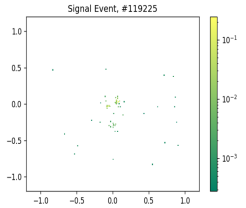

Signal Event, #119225
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
10^{-1}
10^{-2}
10^{-3}

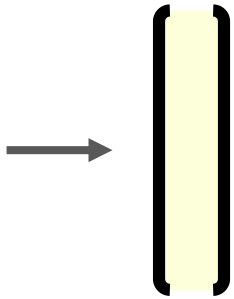


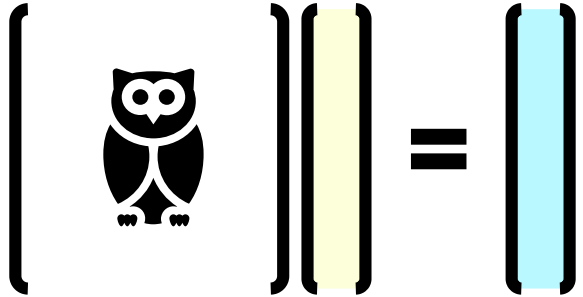


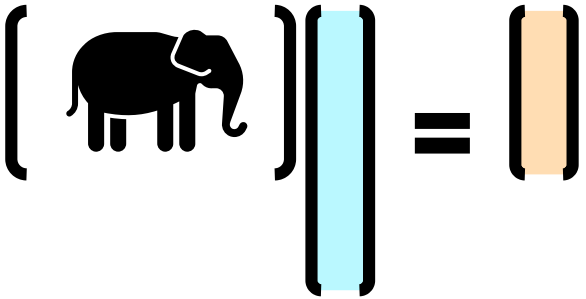


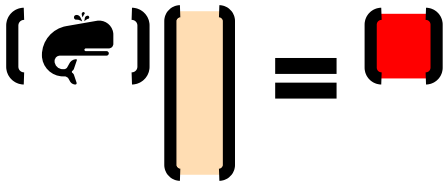

Signal Event, #119225
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
10^{-1}
10^{-2}
10^{-3}

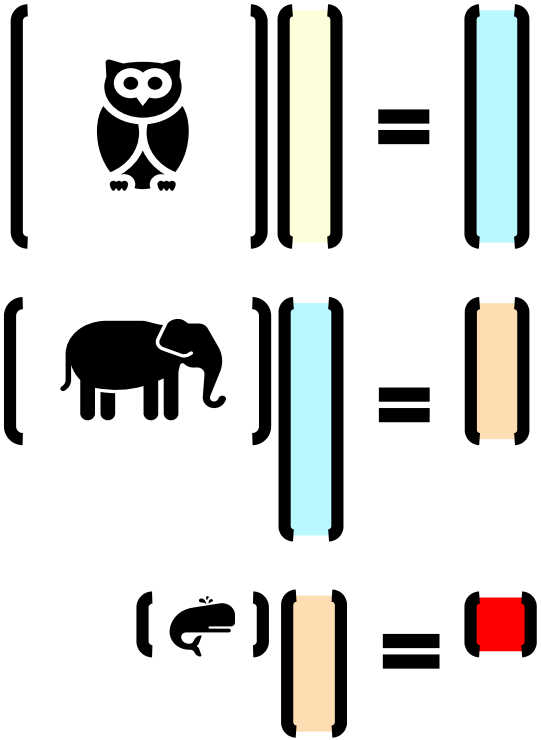

Signal Event, #119225
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
10^{-1}
10^{-2}
10^{-3}

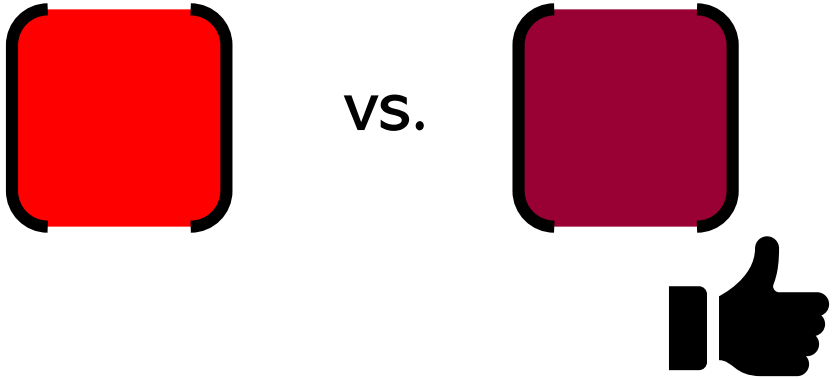











VS.