# Effects of Jet Clustering on Top Tagging with a Hybrid Tree-Sequence Neural Net

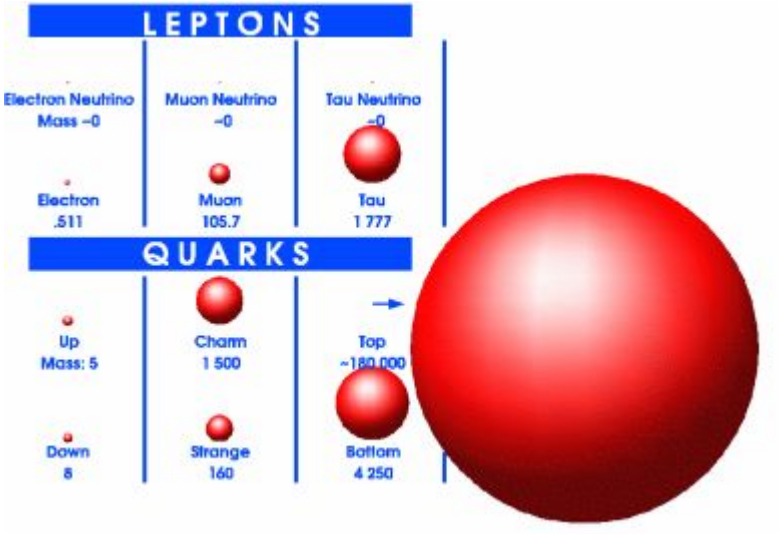
Alex Wen
ATLAS Canada Summer Student Presentations
22 August 2019

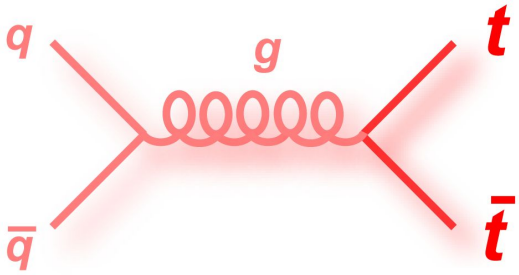
Supervisors: Colin Gay, Alison Lister
ATLAS Group @ UBC

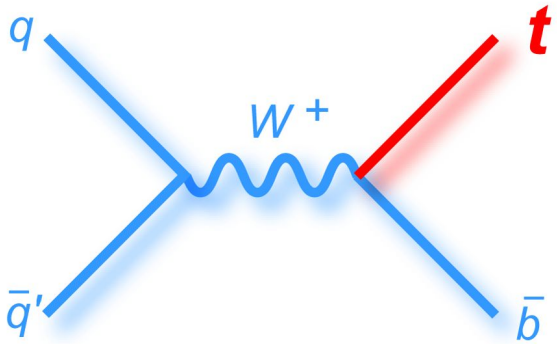
ATLAS EXPERIMENT

THE UNIVERSITY OF BRITISH COLUMBIA

# Top Quarks - Motivation and Production

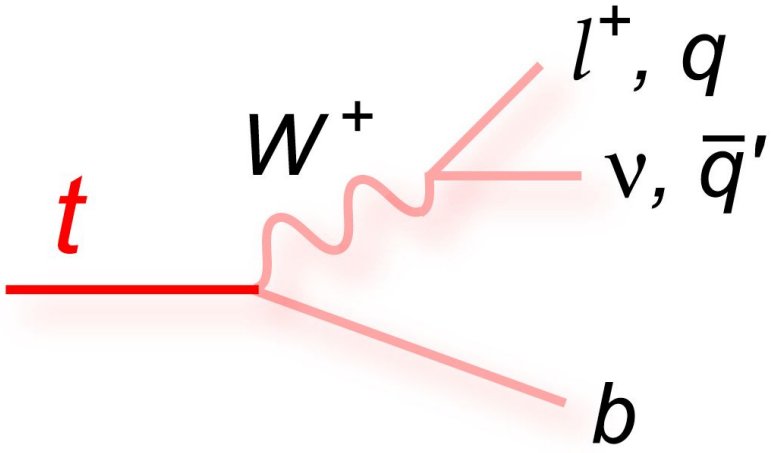


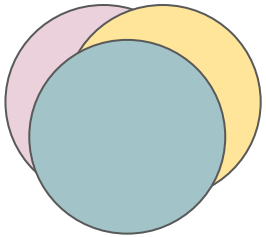




Images: Fermilab

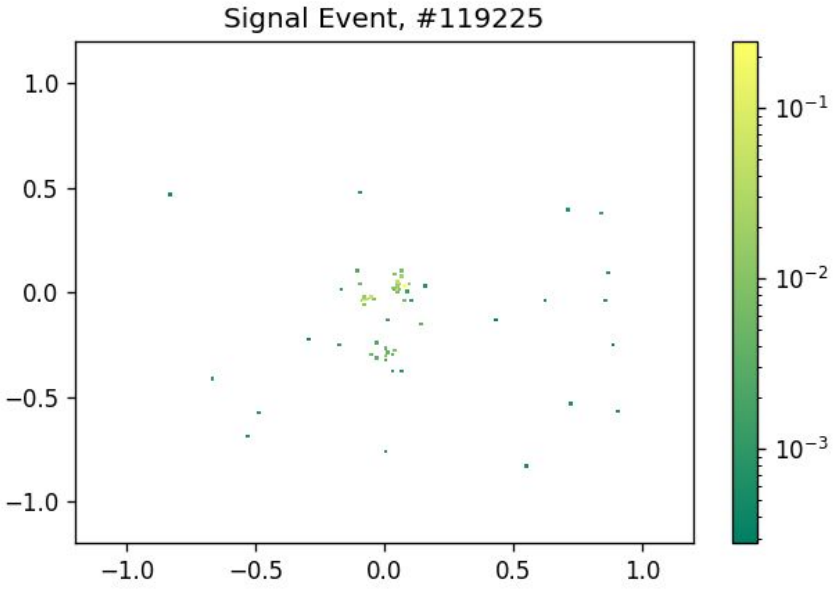
# Top Quarks - Decay

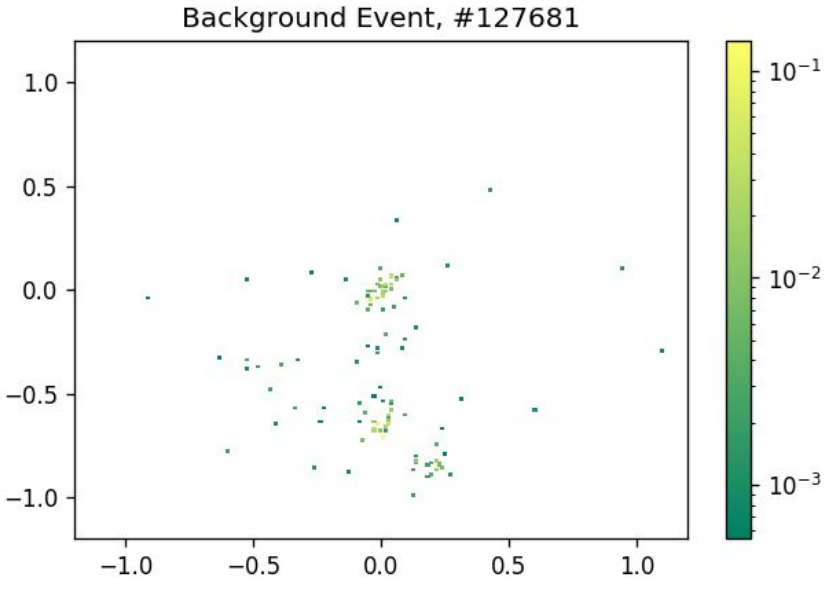



Images: Fermilab



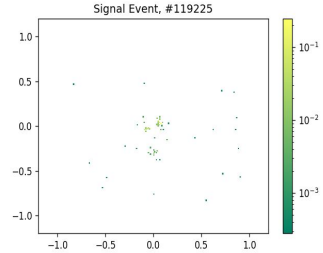
Boosted


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


Background Event, #127681
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
$10^{-1}$
$10^{-2}$
$10^{-3}$

# The Neural Net

Intuitive thing to do:



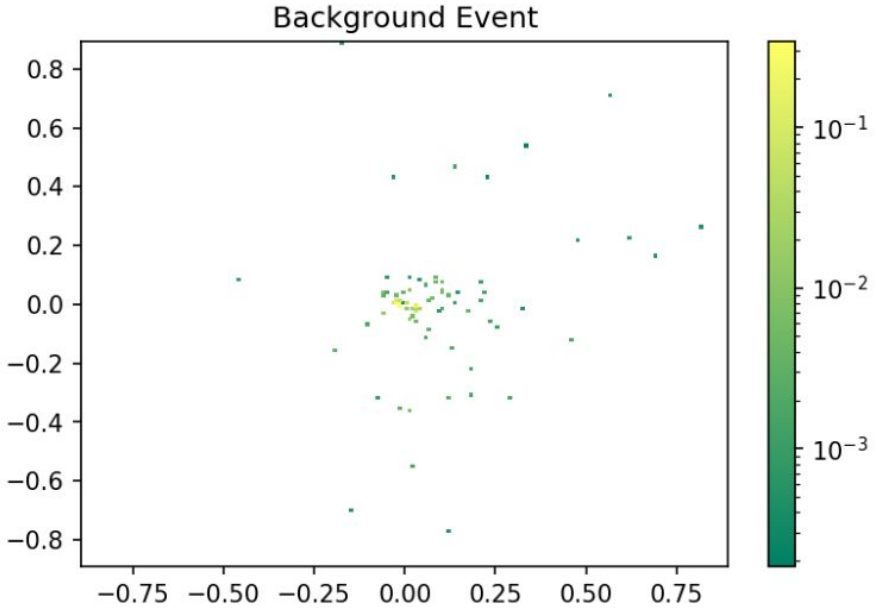

Neural Net of some kind

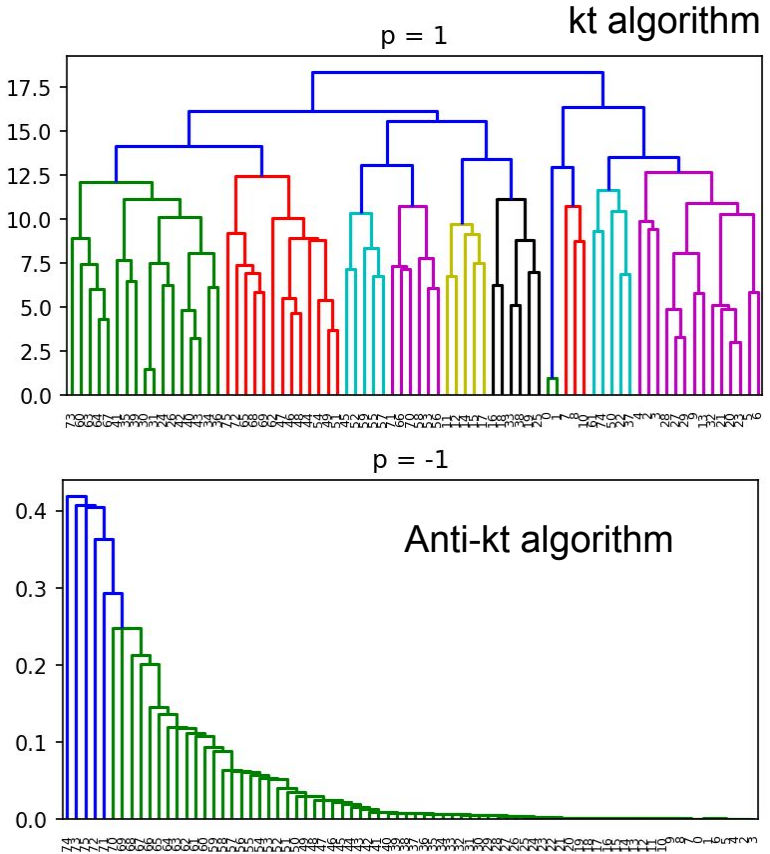
- DNN
- LSTM

Top/not top

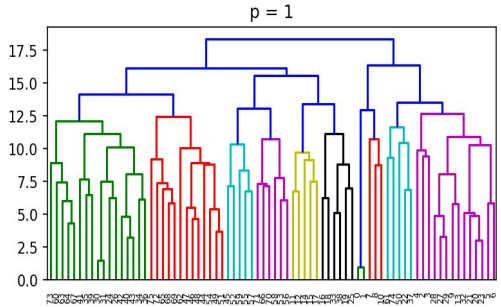
# But … we know physics

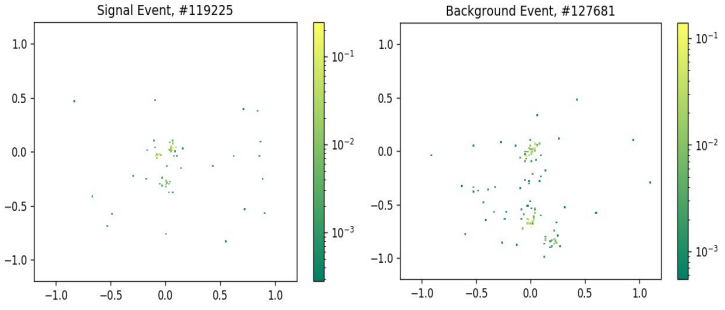
# Clustering Algorithms

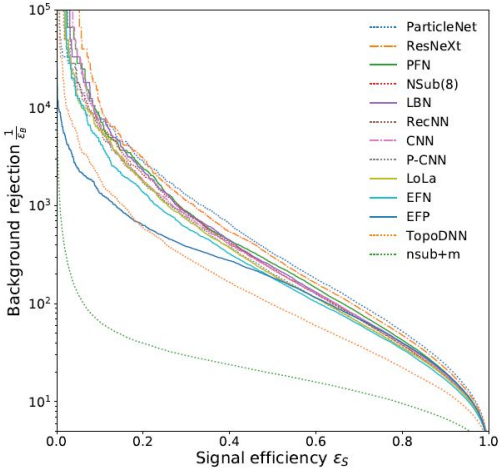


# What kind of neural network?



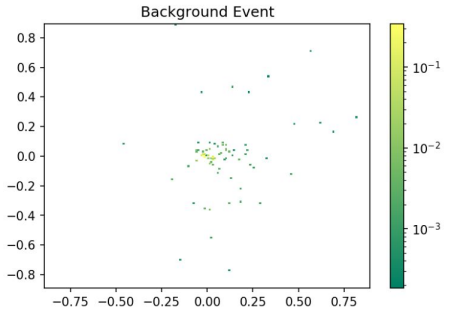

Tree structure & combination of constituents

Really complicated data

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

Image from [2]

# Stack-augmented Parser Interpreter (SPINN)



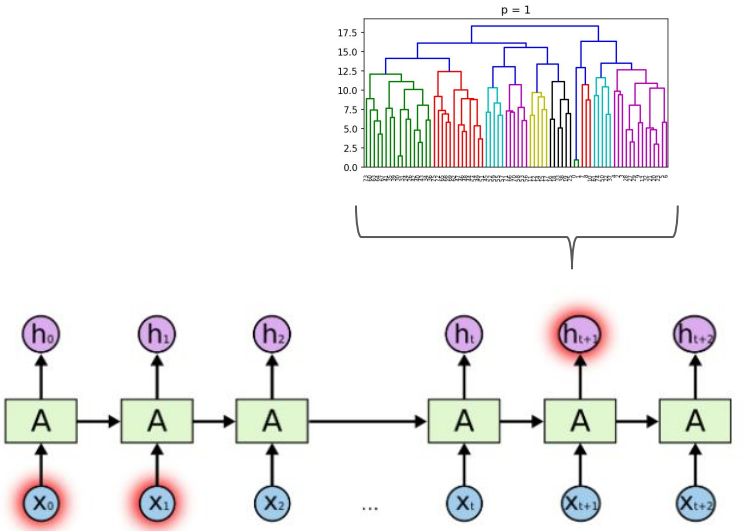



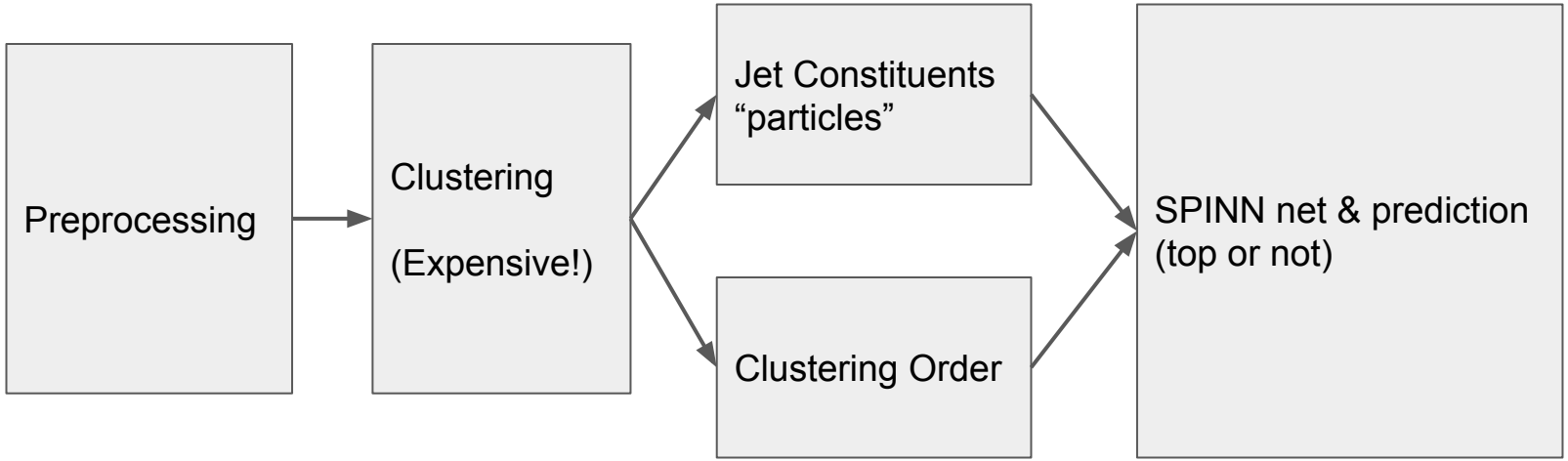

Image: C. Olah

00010010001010101001010101001011111

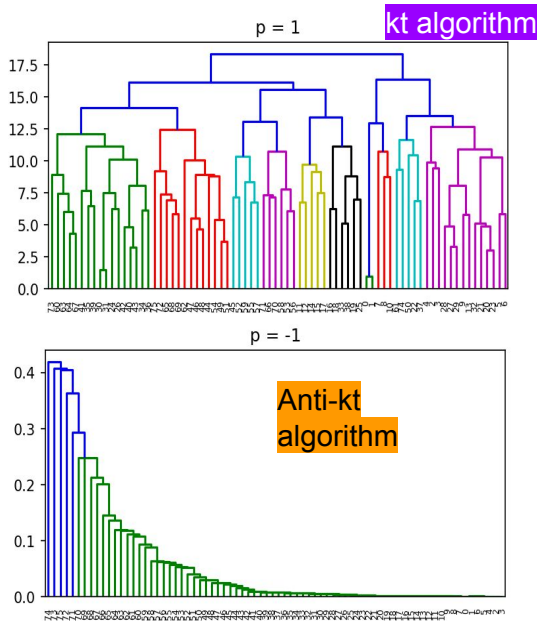
Samuel R. Bowman, Jon Gauthier, Abhinav Rastogi, Raghav Gupta, Christopher D. Manning, and Christopher Potts. A fast unified model for parsing and sentence understanding. *CoRR*, abs/1603.06021, 2016.

# Training

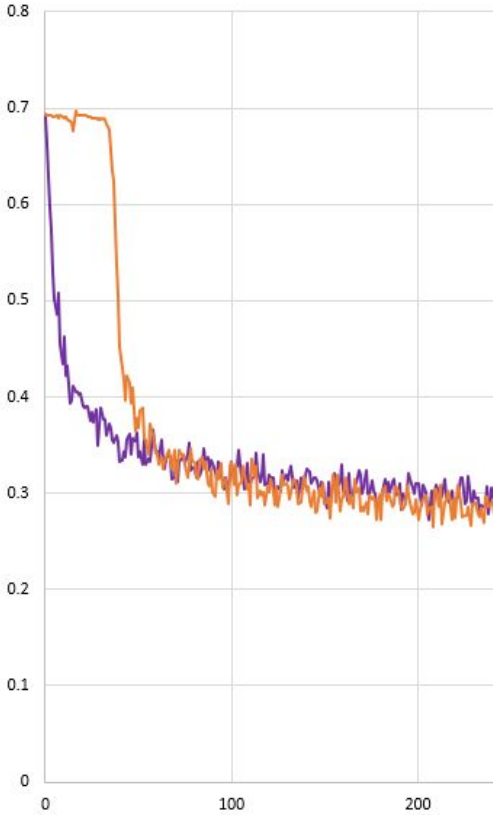
Total data: ~ 2 M events



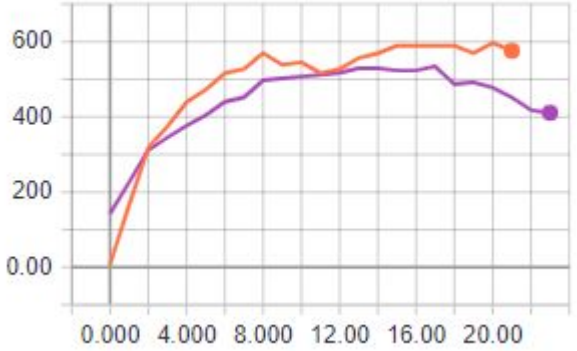
# Results



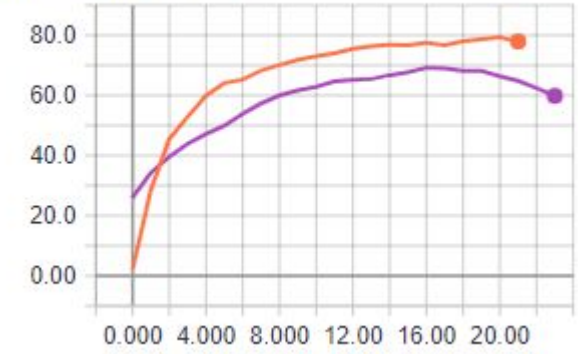

Loss in Time




Rejection at Eff 20%




Rejection at Eff 50%




ATLAS EXPERIMENT

THE UNIVERSITY OF BRITISH COLUMBIA

# Takeaways

- Currently the anti-kt algorithm is most competitive, despite having unrealistic tree structure
- The SPINN network is a logical fit

- Top quarks are interesting
- Identifying them is **hard**
- We can inject **physics** into machine learning
- But this doesn’t always improve performance

# What now?

- More data
- Different clustering algorithms
  - But should we be clustering at all?
  - How much should we care about how “realistic” the clustering is?
- Different neural networks
- Using a less diverse momentum data set
- Injecting more/different physics

# Thanks!

# Backup

[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.

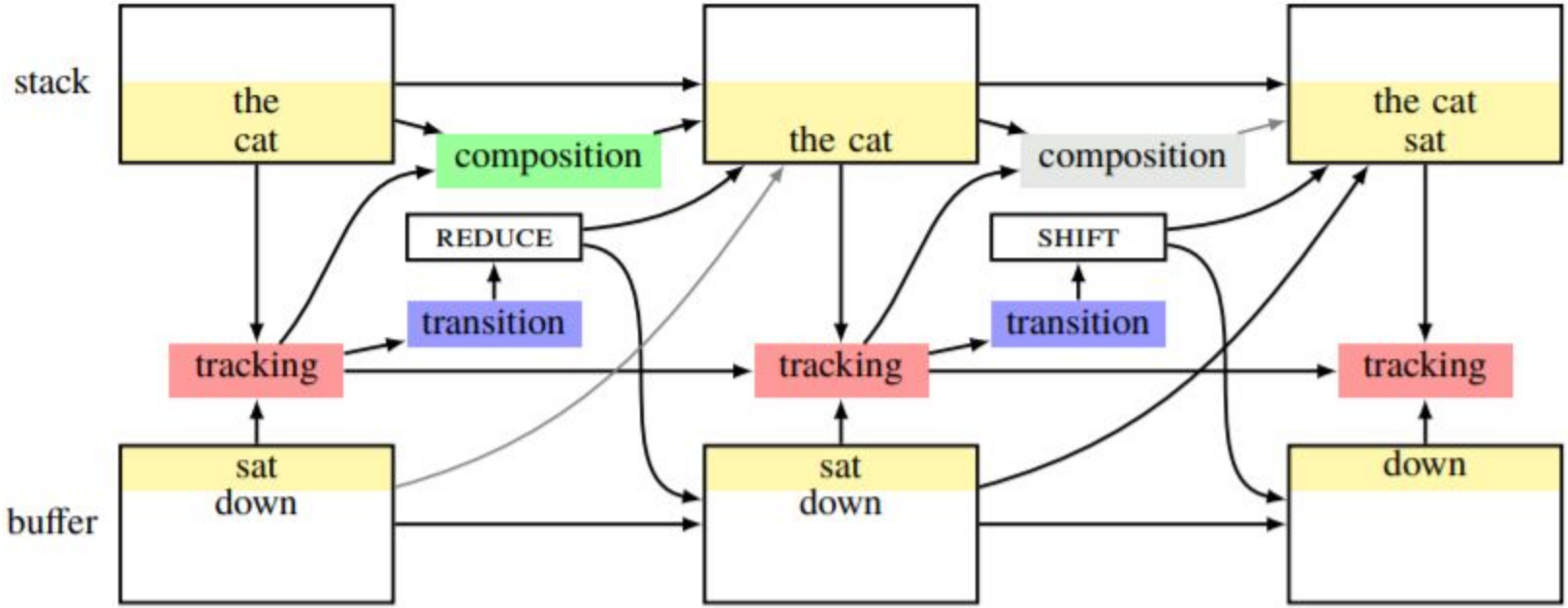
[4] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-$k_t$ jet clustering algorithm. *JHEP*, 04:063, 2008.

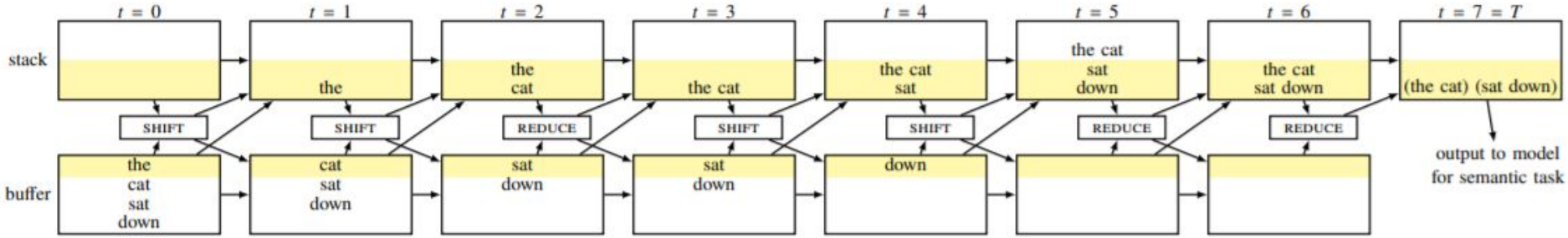
[5] Jannicke Pearkes, Wojciech Fedorko, Alison Lister, and Colin Gay. Jet Constituents for Deep Neural Network Based Top Quark Tagging. 2017.

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

# Paper for reference

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


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




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