Jet Tagging in the Lund Plane with Deep Sets

Hadronic Final State Forum 2018

M. LeBlanc, G. Stark, M. Feickert, S. Weber
(w/ thanks to B. Nachman for ML guidance)

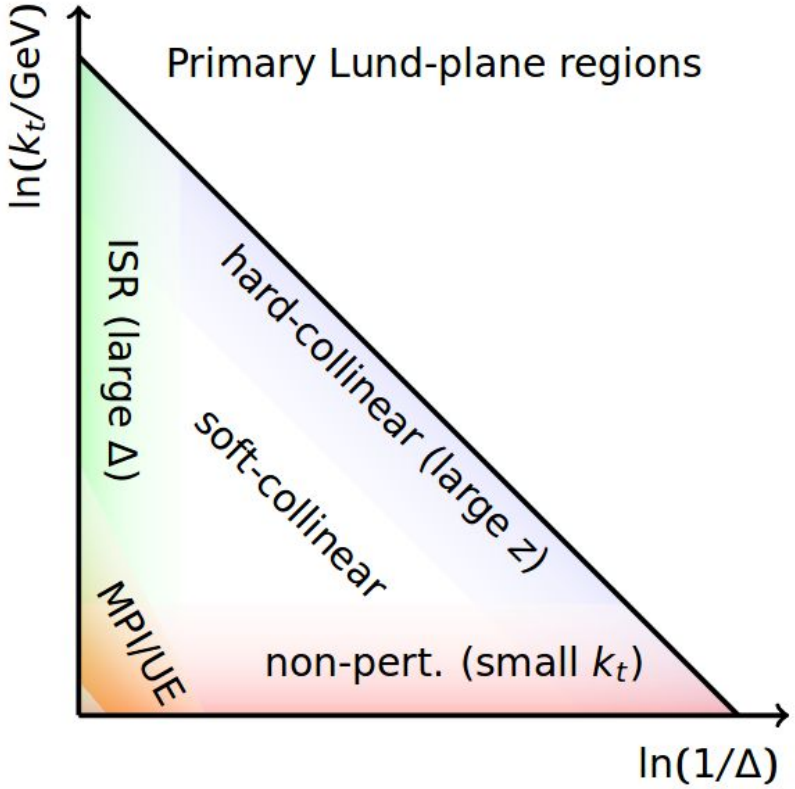
Lund Jet Plane

NB: Covered by M. LeBlanc already in a previous talk

Lund Jet Plane [arXiv:1807.04758]

- Lund Diagrams
  - Theoretical representation of the phase-space within jets
- Visual representation of jets
  - Can be created for individual jets by repeated application of Cambridge/Aachen (C/A) declustering algorithm
  - Built from tracks ghost-associated to large-R=1.0 anti-kt ungroomed jets

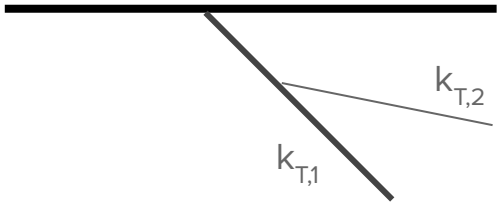
Key: allows us to represent substructure in a quantified way



NB: Covered by M. LeBlanc already in a previous talk

Lund plane physics

- Z_i is a measure of the amount of energy that is taken by the branching relative to how much energy is left in that branch
  - So if there is a branch of a jet that branches again, such that $k_{T,1} > k_{T,2}$, then z would be $k_{T,2}/k_{T,1}$



- Given only knowledge of the emissions from the jets, can we learn anything about tops, W/Zs -- compared to QCD (dijets)?
  - **Today**: include Z, DeltaR
  - **Future**: can inclusion of jet mass, jet pT, emission pT give us more improvements? Any other variables? Bin in jet pT, jet mass?

Deep Sets and Particle Flow Networks

Theory of Deep Sets [arXiv:1703.06114]

Deep Sets

Slide from P. Komiske [ML4Jets 2018]

Namespace for symmetric function parametrization

A general permutation-symmetric function is *additive* in a latent space

Deep Sets [1703.06114]

Manzil Zaheer[1,2], Satwik Kottur[1], Siamak Ravanbhakhsh[1], Barnabás Póczos[1], Ruslan Salakhutdinov[1], Alexander J Smola[1,2]
[1] Carnegie Mellon University [2] Amazon Web Services

Feature space

Variable length

Deep Sets Theorem [63]. *Let $\mathfrak{X} \subset \mathbb{R}^d$ be compact, $X \subset 2^{\mathfrak{X}}$ be the space of sets with bounded cardinality of elements in $\mathfrak{X}$, and $Y \subset \mathbb{R}$ be a bounded interval. Consider a continuous function $f : X \to Y$ that is invariant under permutations of its inputs, i.e. $f(x_1, \ldots, x_M) = f(x_{\pi(1)}, \ldots, x_{\pi(M)})$ for all $x_i \in \mathfrak{X}$ and $\pi \in S_M$. Then there exists a sufficiently large integer ℓ and continuous functions $\Phi : \mathfrak{X} \to \mathbb{R}^\ell$, $F : \mathbb{R}^\ell \to Y$ such that the following holds to an arbitrarily good approximation:*[1]

Permutation invariance

Latent space

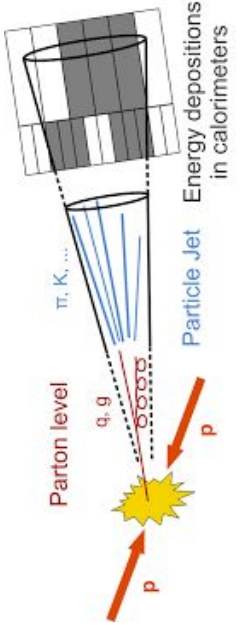
$$f(\{x_1, \ldots, x_M\}) = F\left(\sum_{i=1}^{M} \Phi(x_i)\right)$$

General parametrization for a function of sets

...Observables for jets

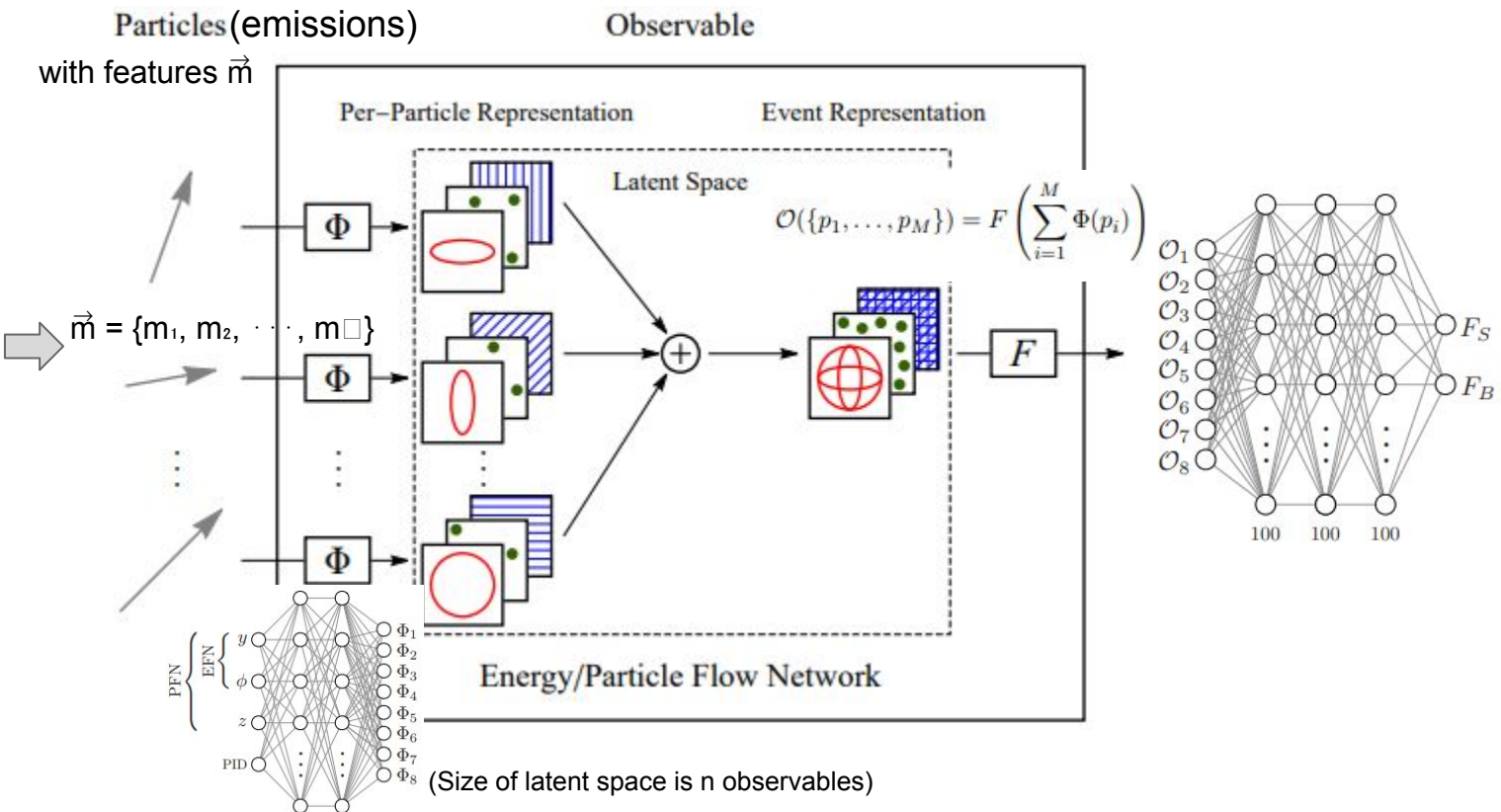
$$\mathcal{O}(\{p_1, \ldots, p_M\}) = F\left(\sum_{i=1}^{M} \Phi(p_i)\right)$$

Particle Flow Networks [arXiv:1810.05165]



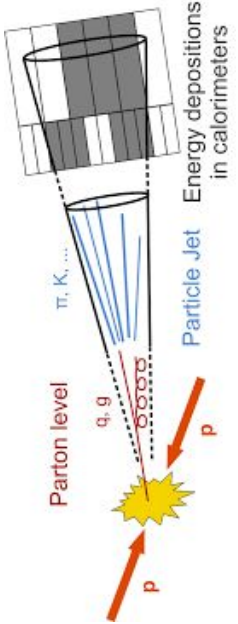
Particles (emissions) with features $\vec{m}$

$\vec{m} = \{m_1, m_2, \cdots, m_{\square}\}$



(Size of latent space is n observables)

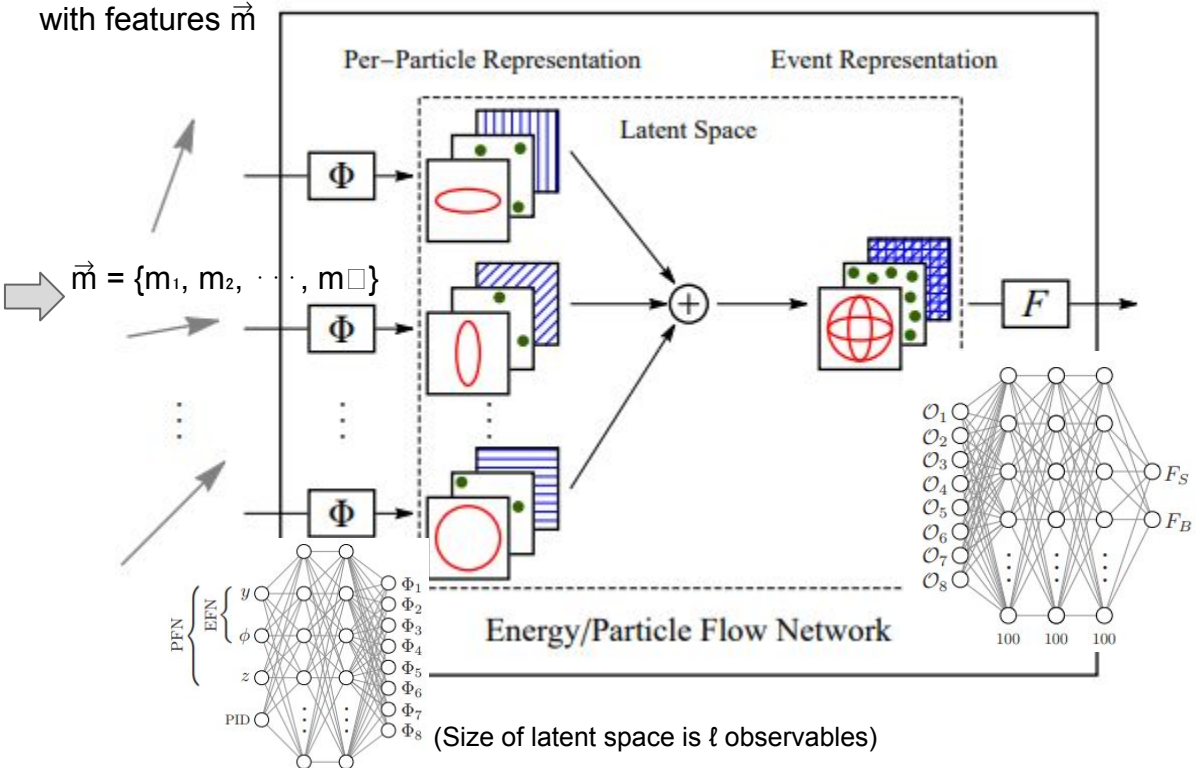
Jet



Particles (emissions) with features $\vec{m}$

$\vec{m} = \{m_1, m_2, \cdots, m_\square\}$

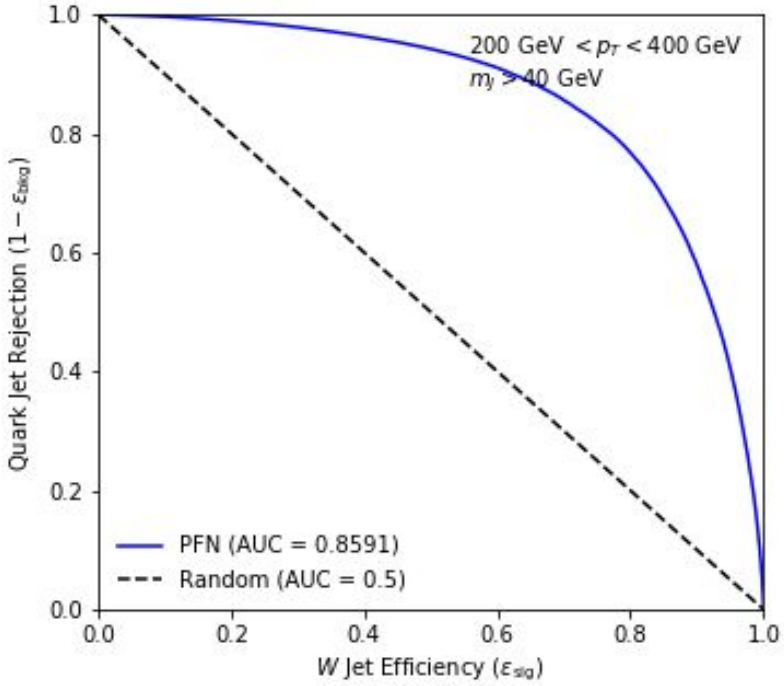
Observable

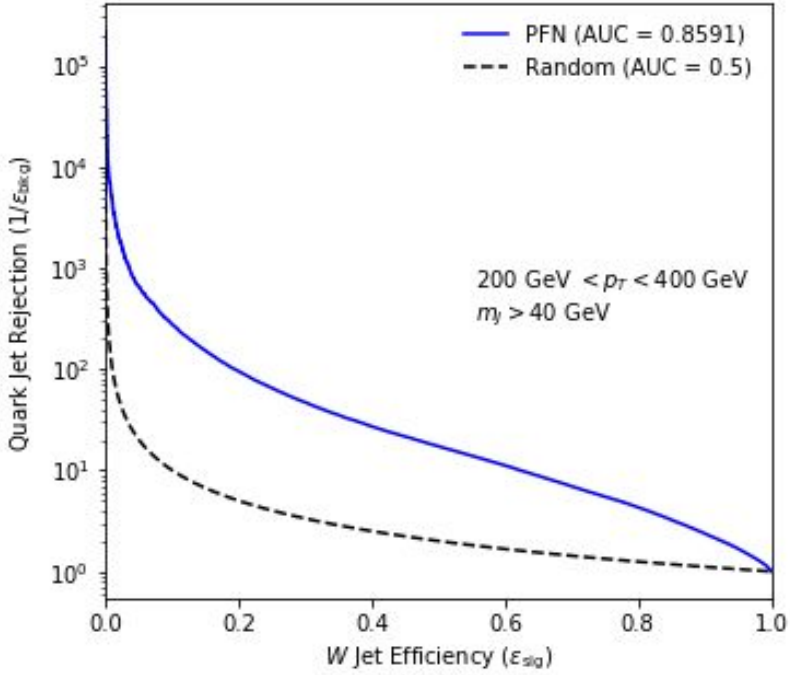


(Size of latent space is ℓ observables)

PFN Performance with 2^7 latent space

- Features used: log(1/ΔR), log(1/z)
- Bkg: JZ3W (pT: [200, 400] GeV)
- Signal: W/Z's

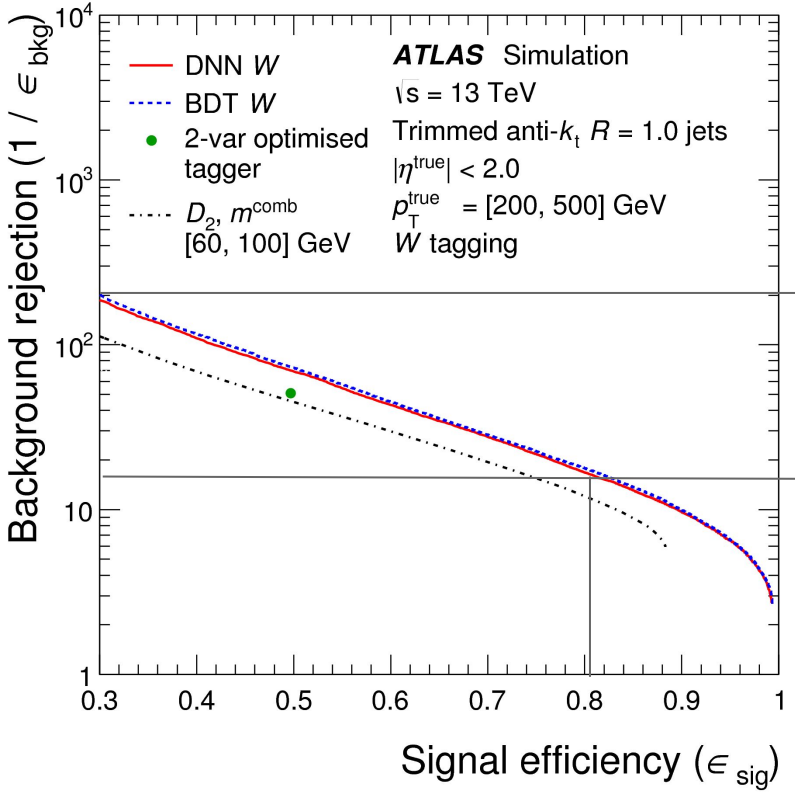


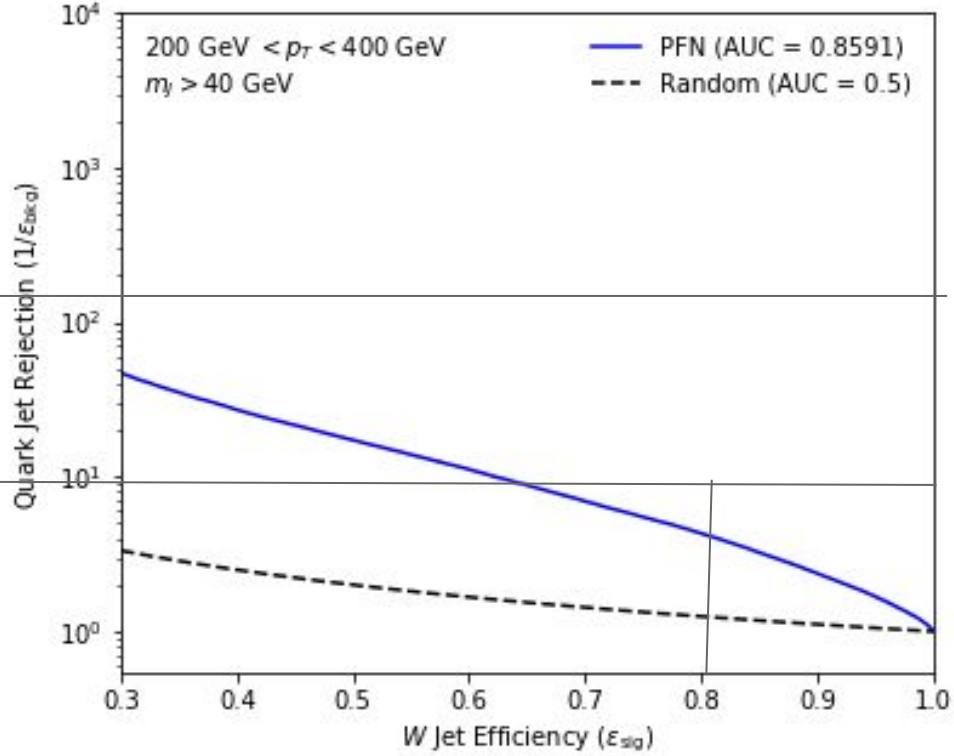


PFN vs. ATLAS Results [CERN-EP-2018-192]

PFN underperforming everything on ATLAS plot, but totally unoptimized at this point

Lots of room for improvement!

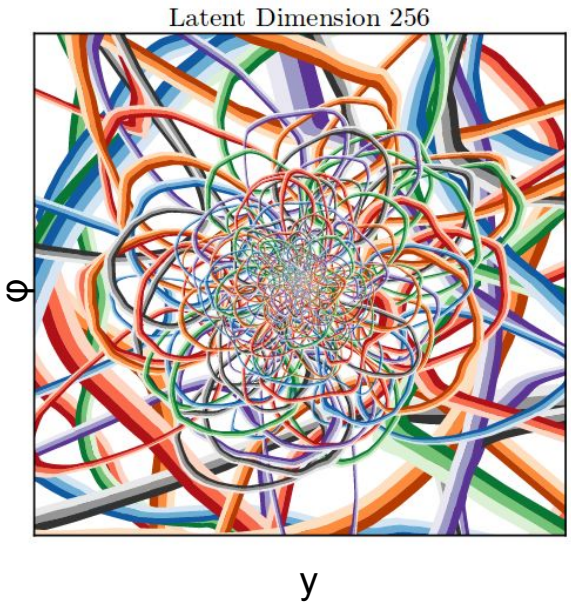




See last few code blocks here:
https://gitlab.cern.ch/USATLAS-HFSF2018/lund-tagging-deep-sets/blob/feature/add-example-NN-stephen/notebooks/lund-tagging-deep-sets.ipynb

JZ6 and Zprime->tt

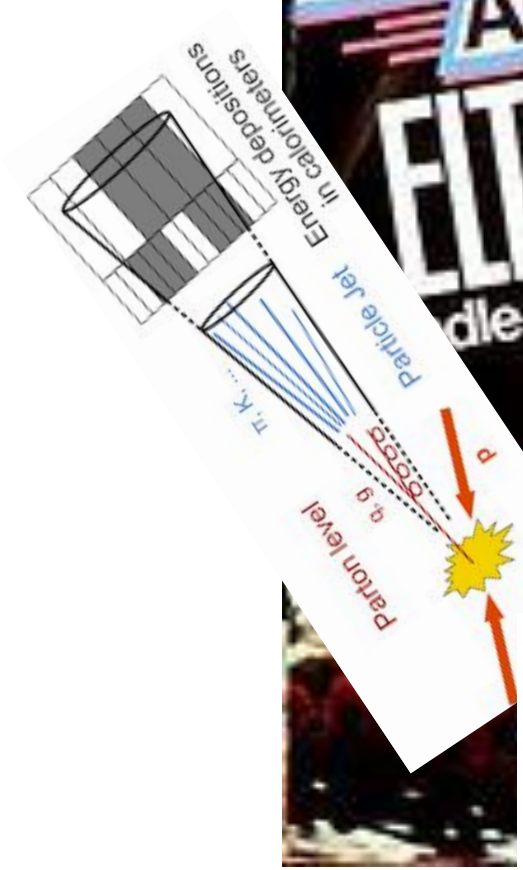
An attempt at recreating the filter contours from [1810.05165]

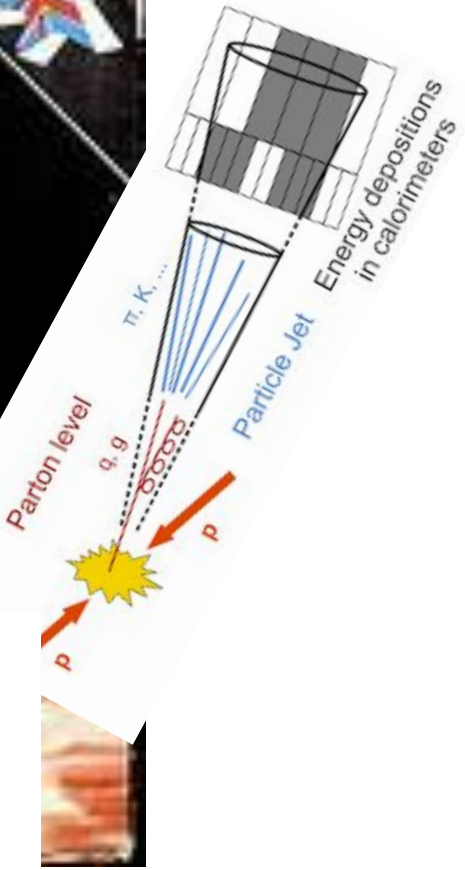


Notes and Refs

Papers and Refs

- GitLab Repo: https://gitlab.cern.ch/USATLAS-HFSF2018/lund-tagging-deep-sets
- Lund Jet tagging [arXiv:1807.04758]
- Energy Flow Networks: Deep Sets for Particle Jets [arXiv:1810.05165]
- EnergyFlow code: https://energyflow.network/
- Deep Sets: [arXiv:1703.06114]





Matt LeBlanc
@TopPhysicist

Following

Replying to @kratsg @HighEnergyDavid and 2 others

b-b-b-b-b-ben nachman and the jets @BPNachman

Translate Tweet

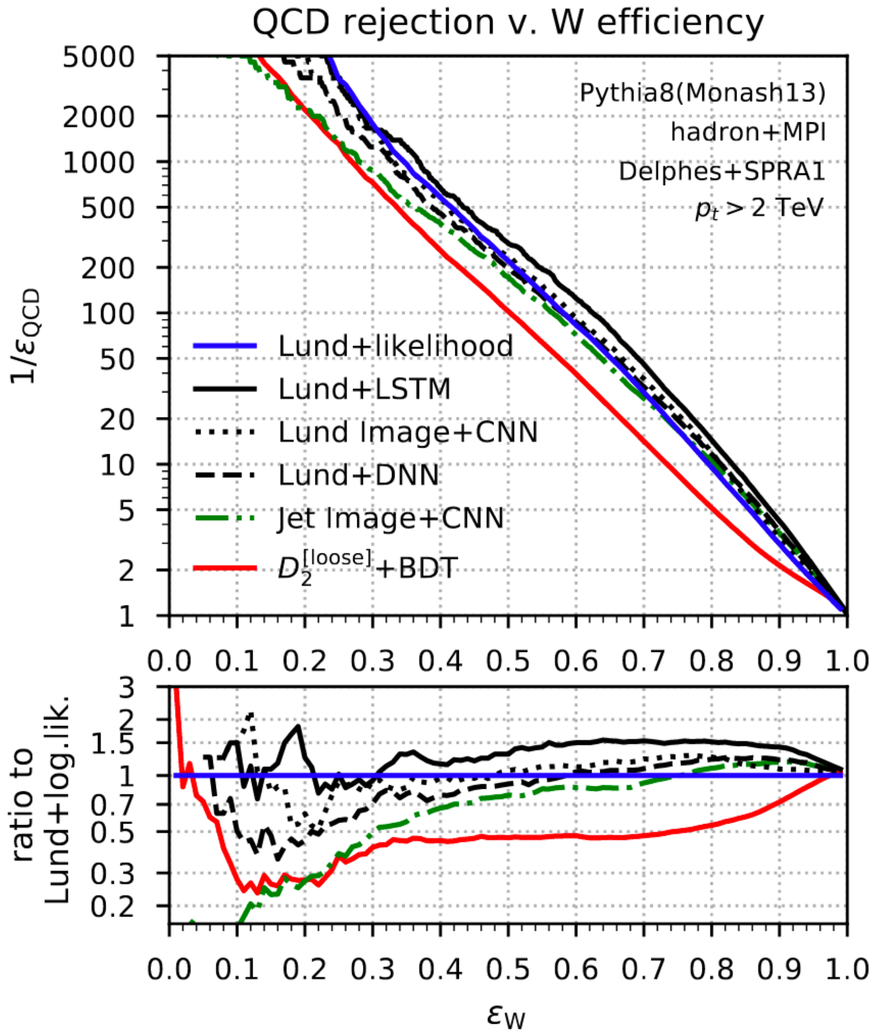
9:35 AM - 7 Dec 2018 from Meyrin, Suisse

Lund Plane Tagging @ #HFSF18

Matt LeBlanc (Arizona) o.b.o. the Lund Jet Team

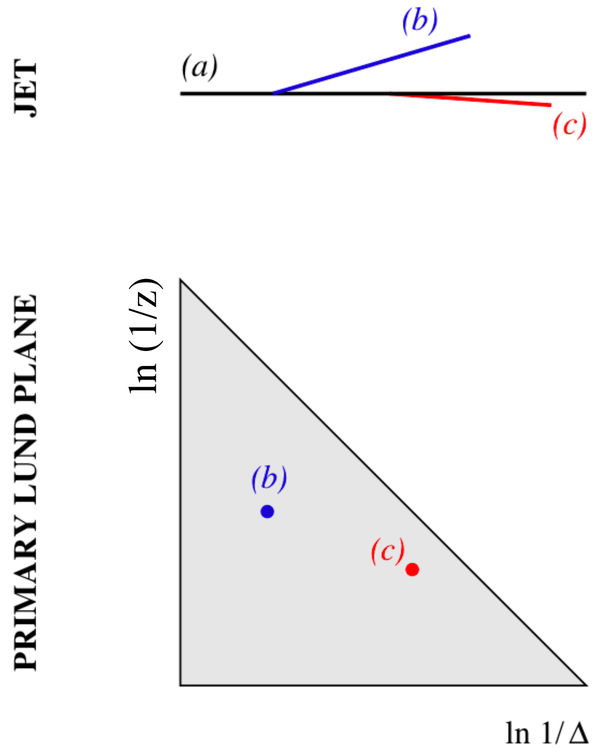
Overview

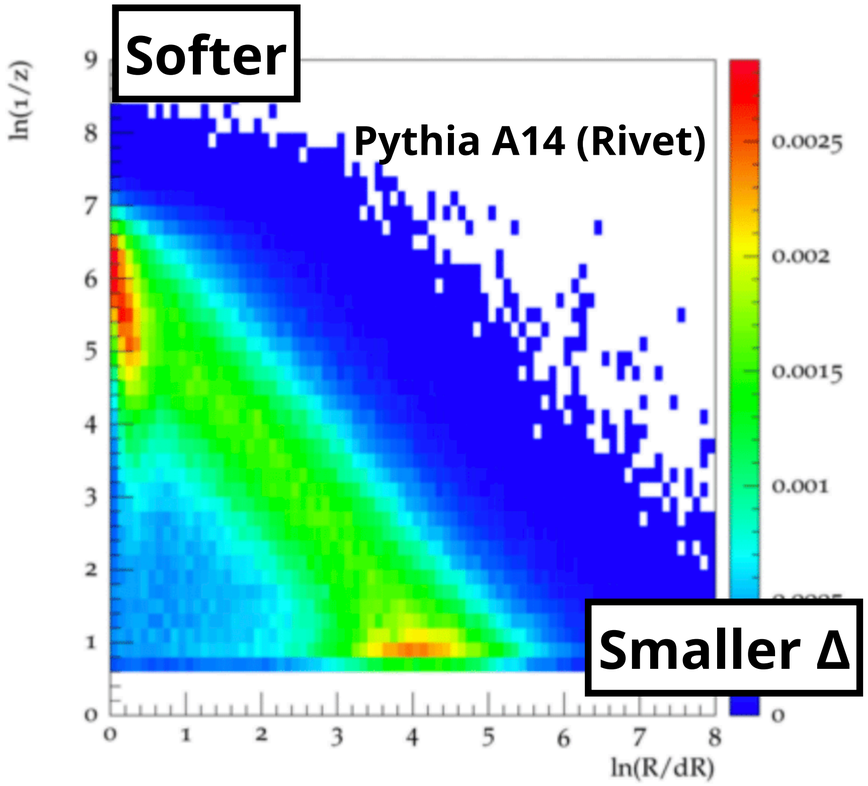
- **Representing the substructure of jets using the Lund Jet Plane** was recently proposed by Dreyer/Soyez/Salam (1807.04758).
  - The authors propose that this approach may provide new insights which will be useful for …
    - MC tuning (**seems promising**, but not the focus of this workshop)
    - **Jet-by-jet tagging (this week!)**

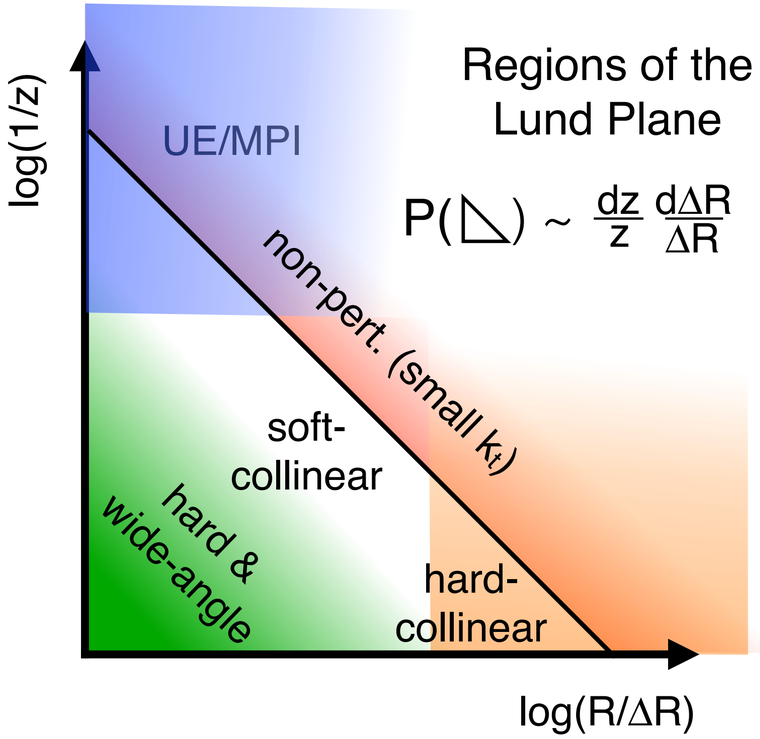


The Lund Jet Plane

- Recluster tracks matched to jet with C/A algorithm (angular ordering), then plot.







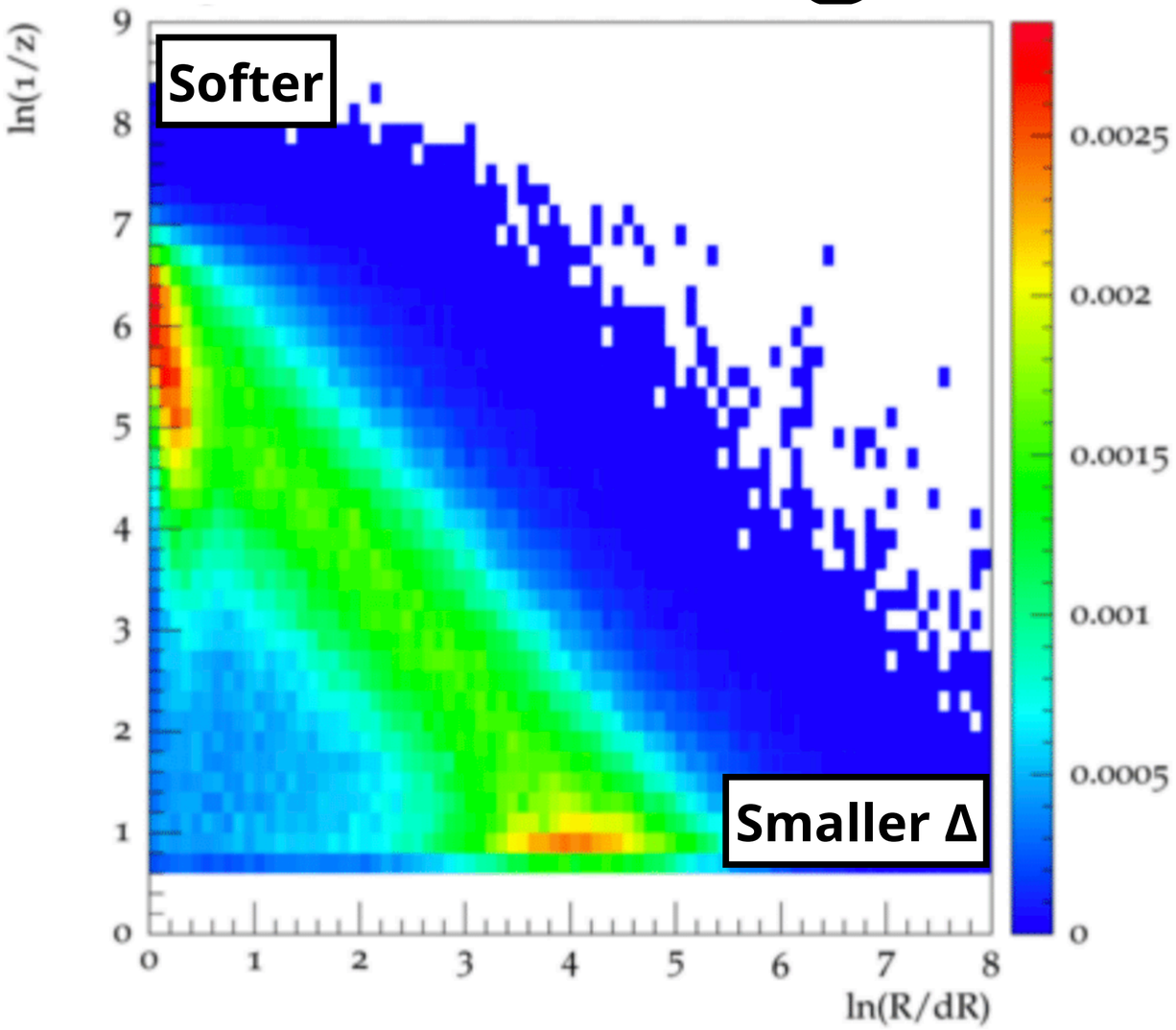
$$P(\text{◺}) \sim \frac{dz}{z}\frac{d\Delta R}{\Delta R}$$

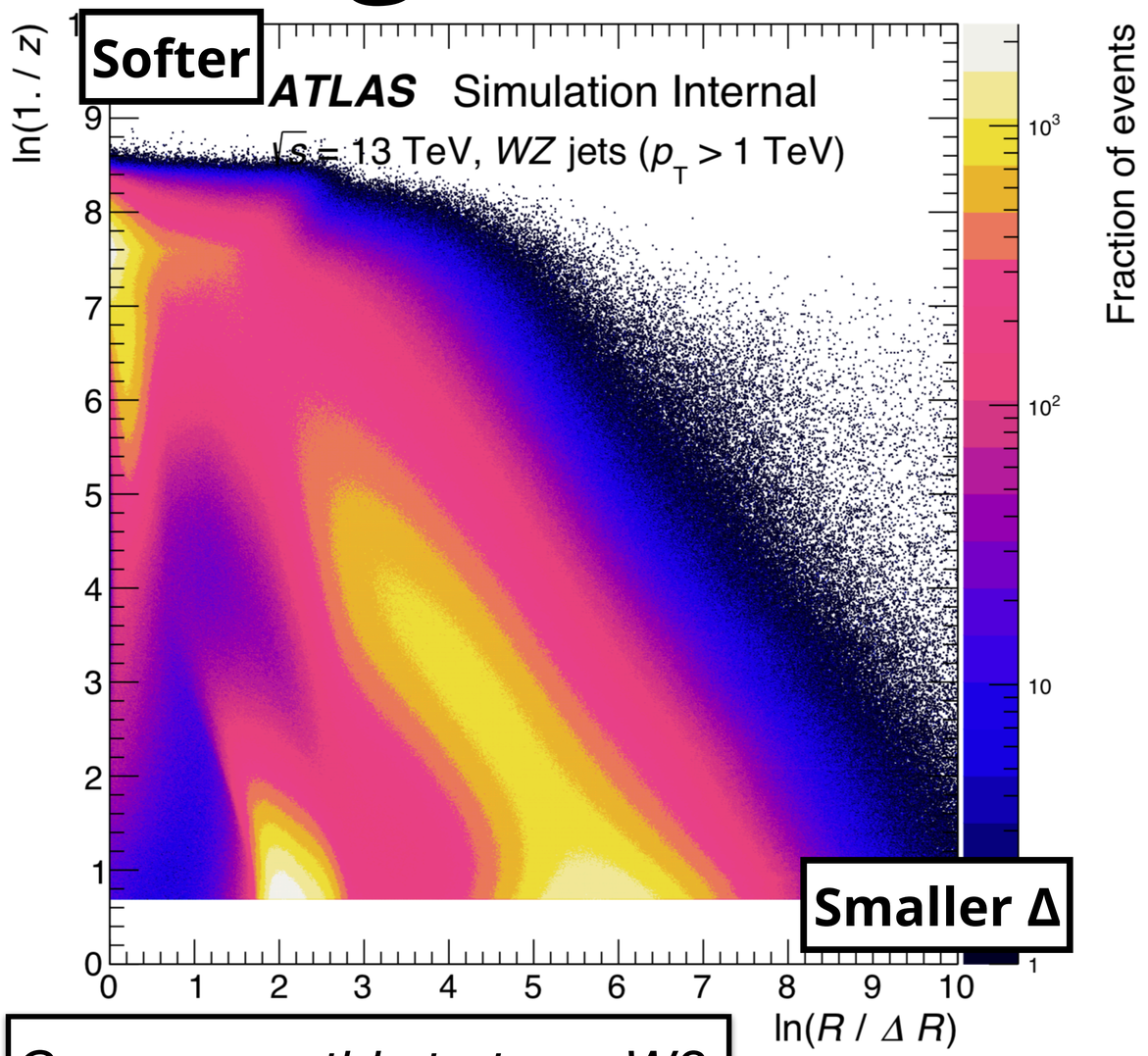
Can be done jet-by-jet …

… or for an ensemble …

… different sources of radiation are neatly compartmentalised.

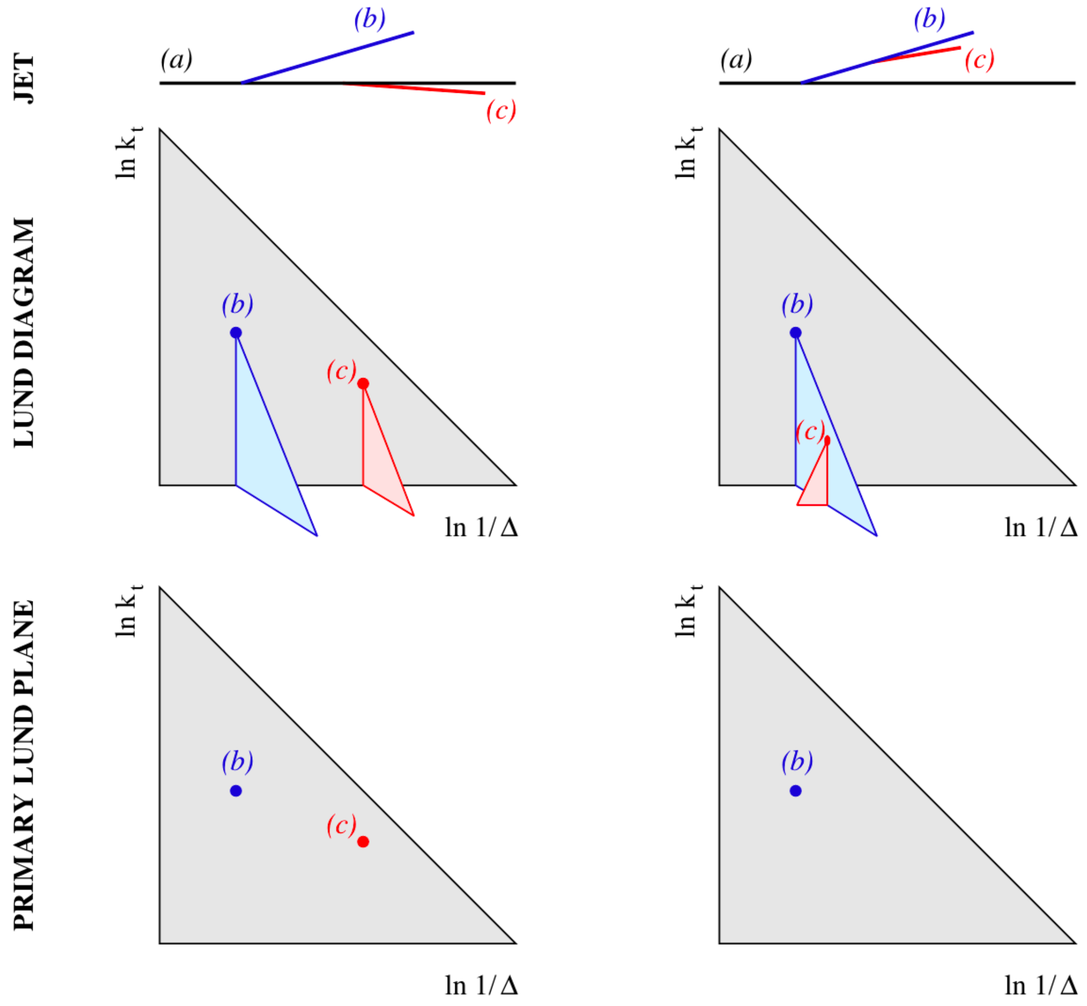
Background vs. Signal





Can we use this to tag a W?

Next time?



- Perhaps including the secondary planes would help discrimination?
- I will avoid spoiling the upcoming talks much more at this point. Speaking of which, there are quite a few!

Jet Tagging in the Lund Plane with Deep Sets — *Ben Nachman et al.*
Building 59, Room 3101, LBNL — 10:45 - 11:00

Lund Planes and RNNs — *Jakub Filipek et al.*
Building 59, Room 3101, LBNL — 11:00 - 11:15

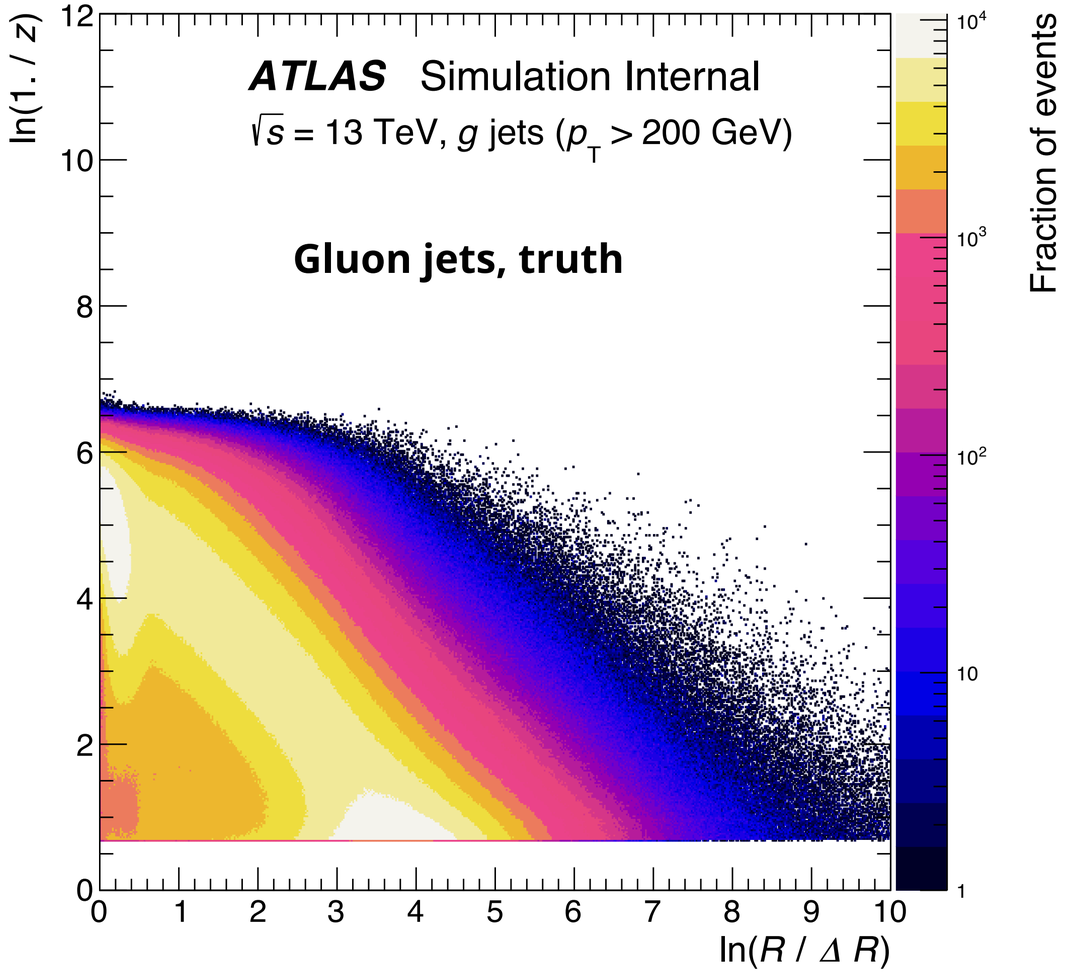
ML on the Lund Plane — *Congqiao Li et al.*
Building 59, Room 3101, LBNL — 11:15 - 11:30

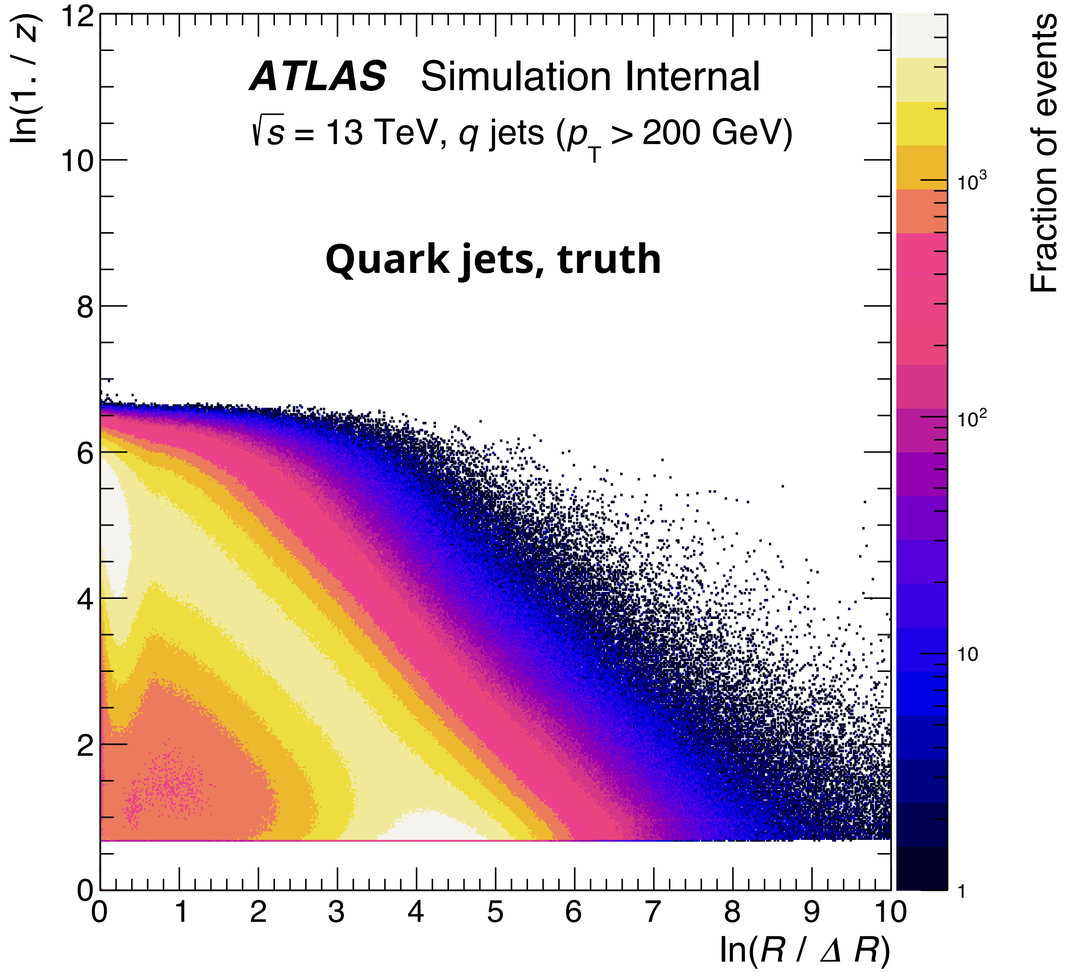
M. Feikart (SMU)
MLB (AZ)
G. Stark (UCSC)
B. Nachman (LBNL)
S. Weber (Carleton)

J. Filipek (UW)
J. Kruper (UW)

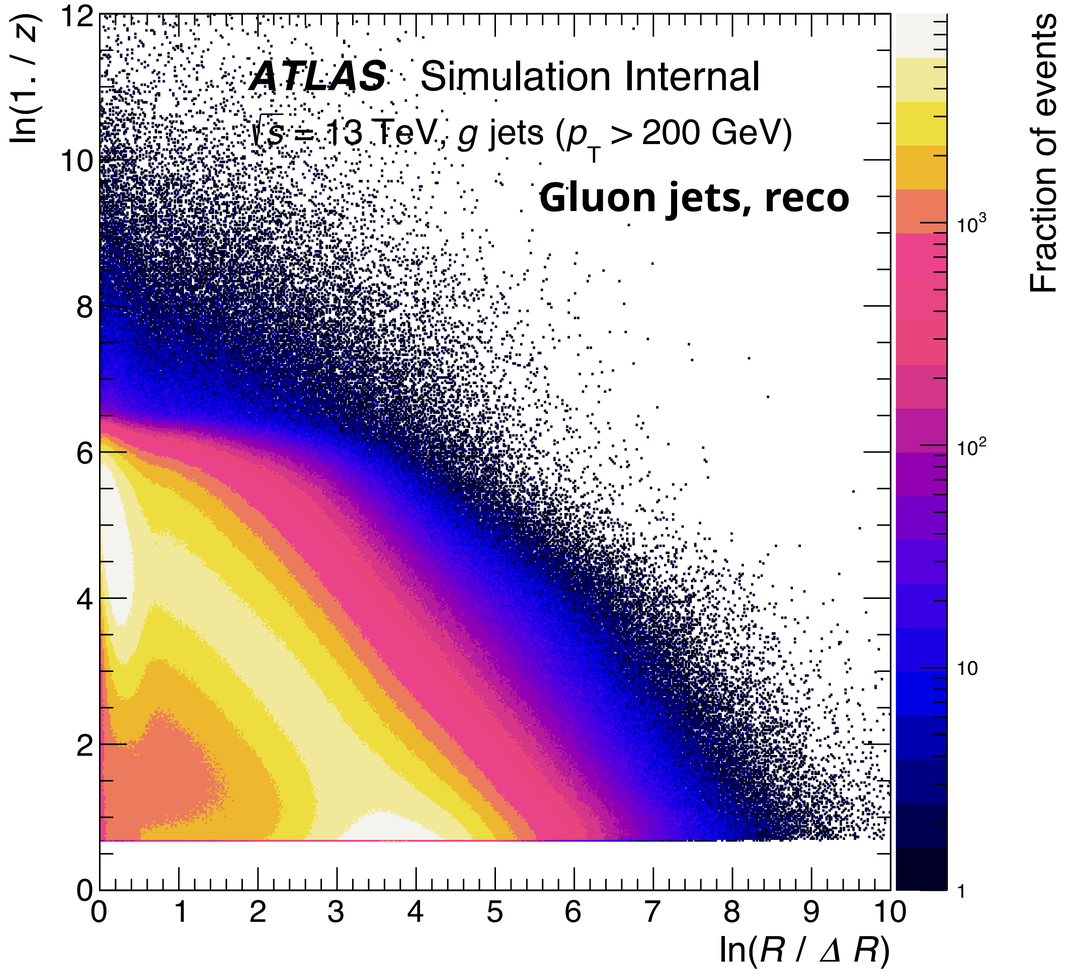
C. Li (UW)
G. Zhang (UW)
T. Ou (UW)

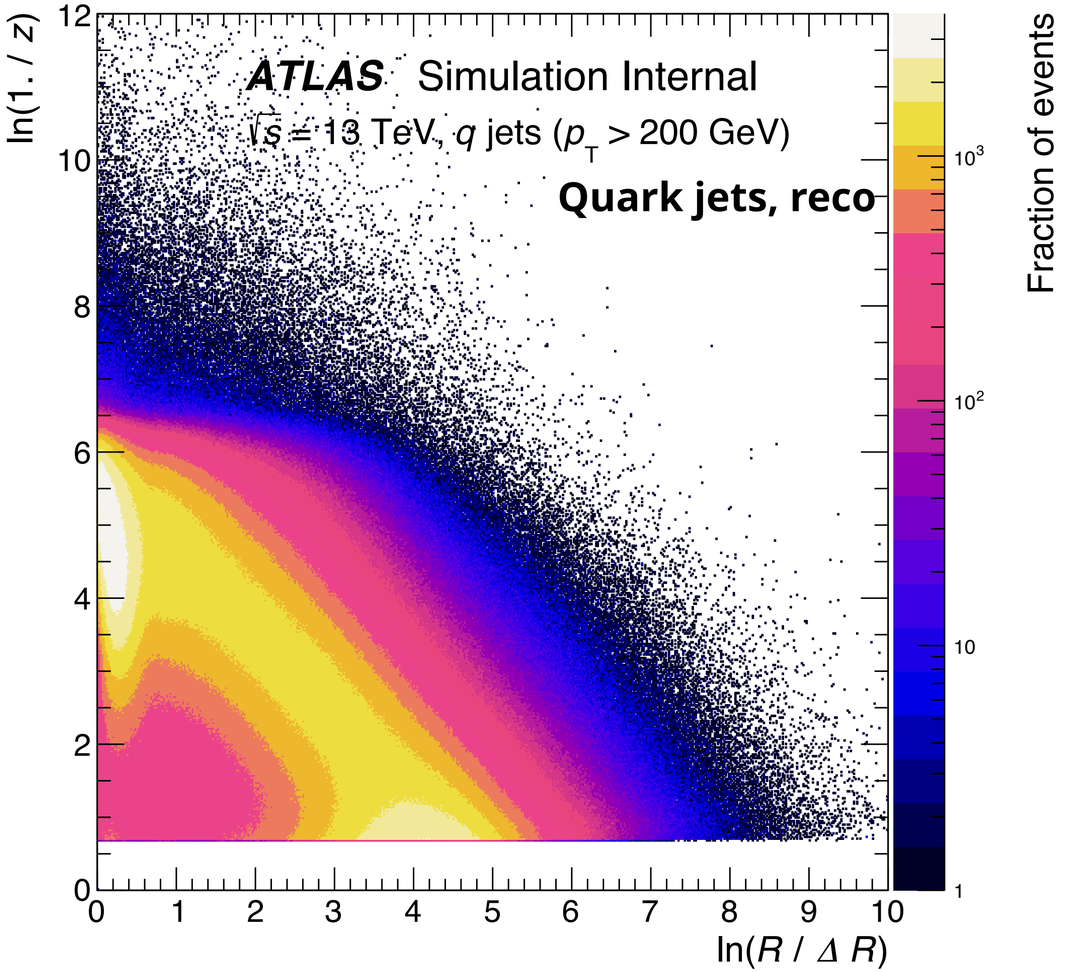
q/g : Spot the difference?





q/g : Spot the difference?

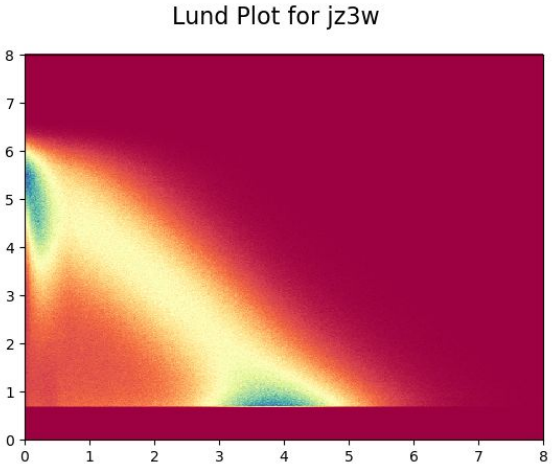


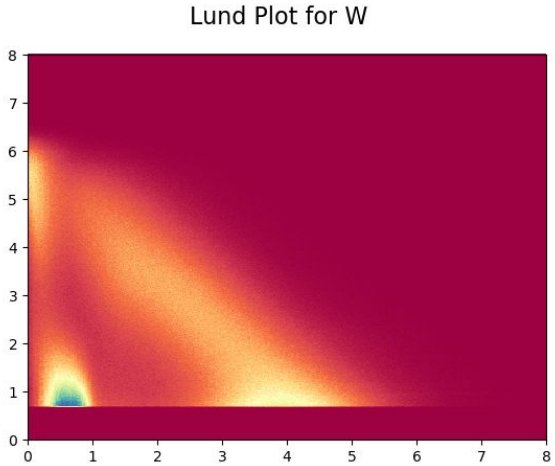


Thanks, everyone! #HFSF2018

Lund Planes & RNNs

And an aside on XGBoost on Gregor Kasieczka's Top Tagging Reference Dataset

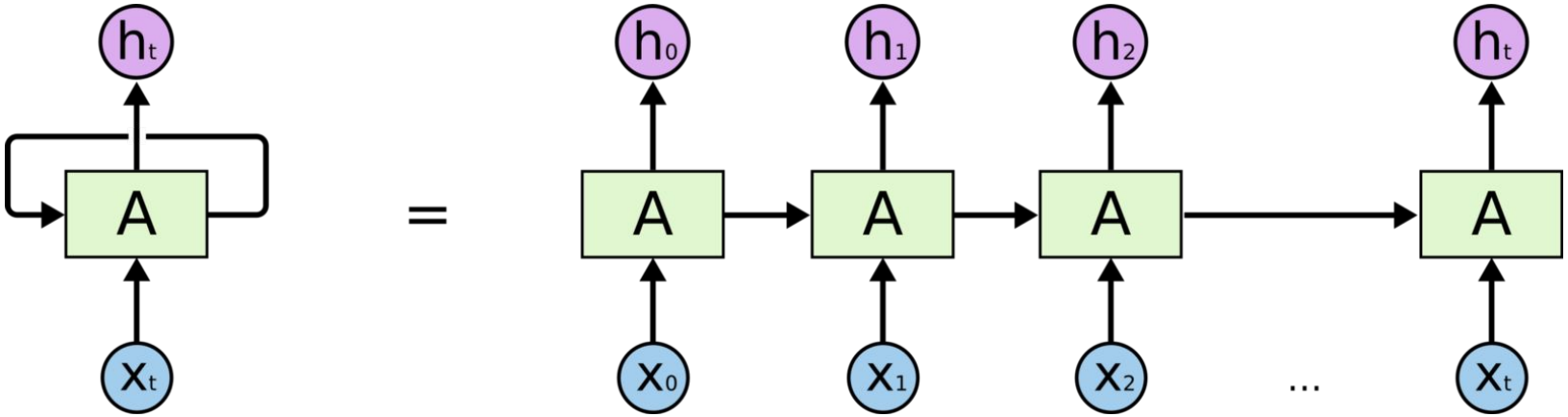




Jakub Filippek & John Kruper

2018 US ATLAS Hadronic Final State Forum

What are RNNs?

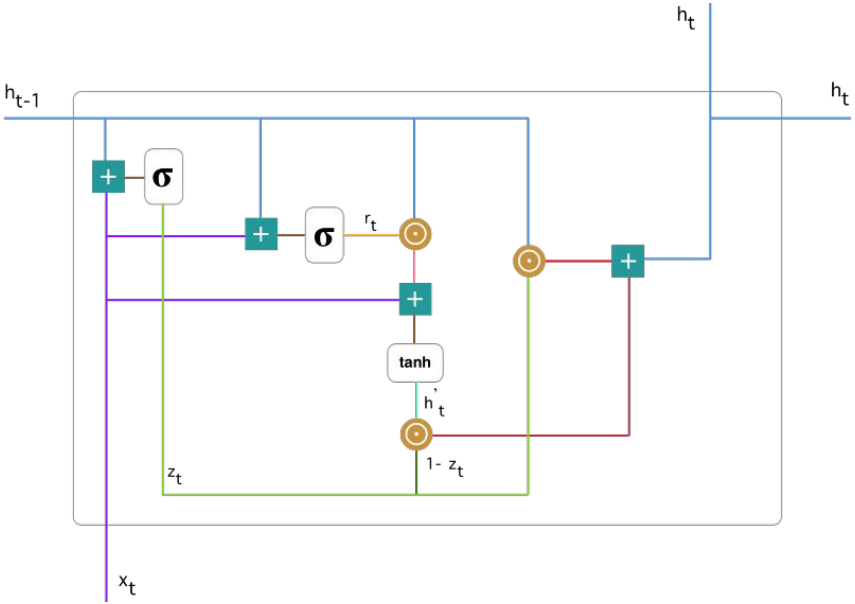


- Normal (dense) network applies a transformation (A) on input (x) to give the output (h).
- Recurrent NN, takes the sequence to inputs. For a given input in a sequence, it's output is based on that input and the output of a term previously in a sequence.
- This can be visualized on diagram on a right, where vertical arrows mean standard input/output transformation, while horizontal arrows mean passing the output of previous element of sequence.

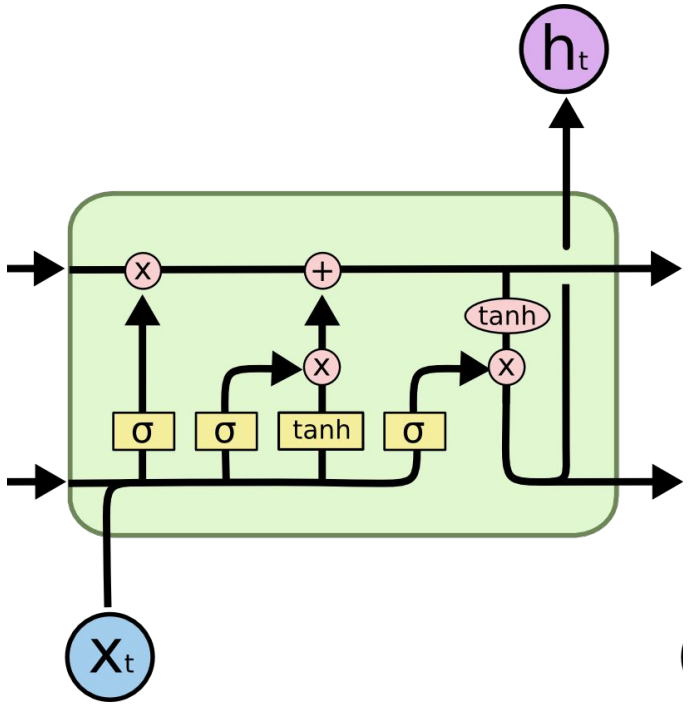
Image source

What are GRUs, and LSTMs?

GRU is a Gated Recurrent Unit.



LSTM is a Long Short Term Memory



GRU image source

LSTM image source

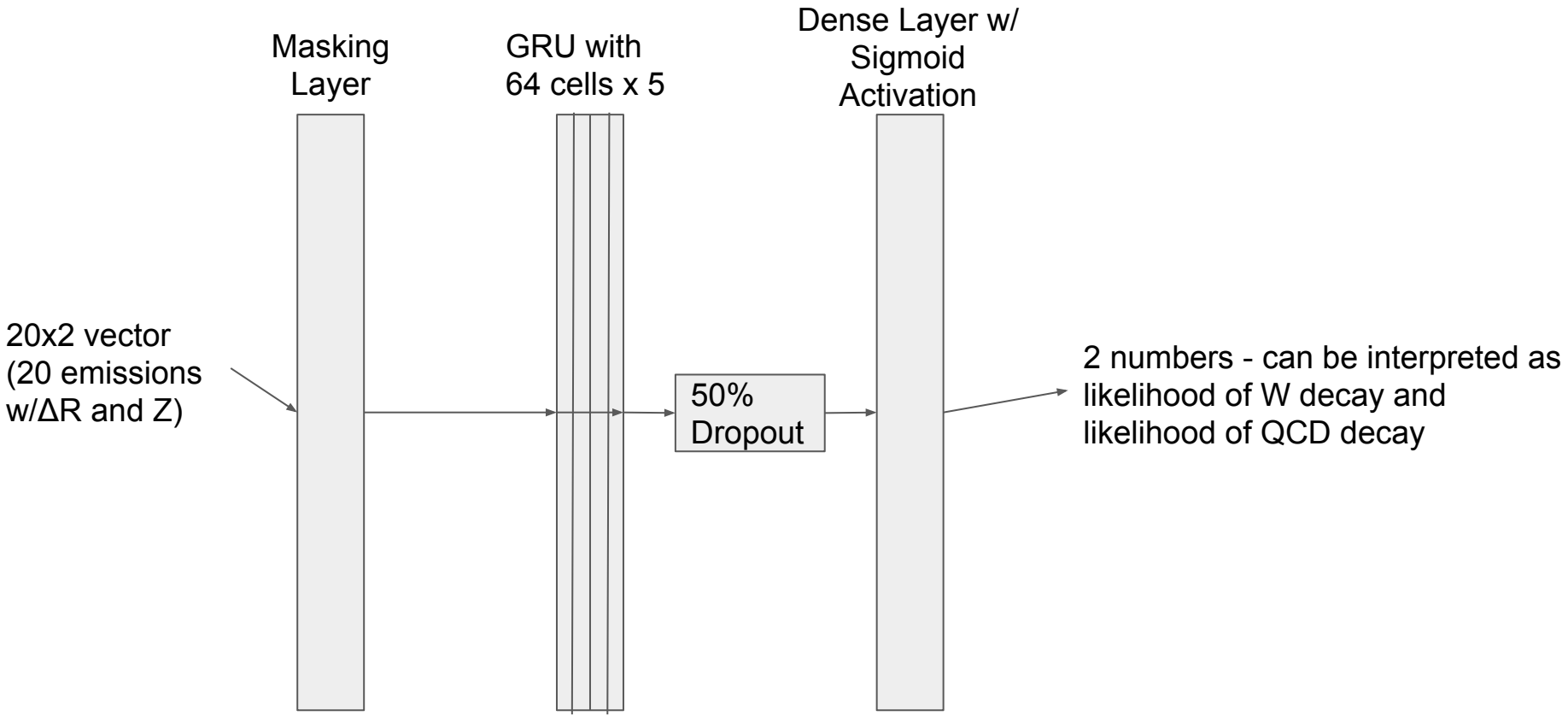
Some Scope

- W against QCD, cut to 200-400 pt
- Trained using ΔR and Z from up to 20 emissions
- ~1.6M training events

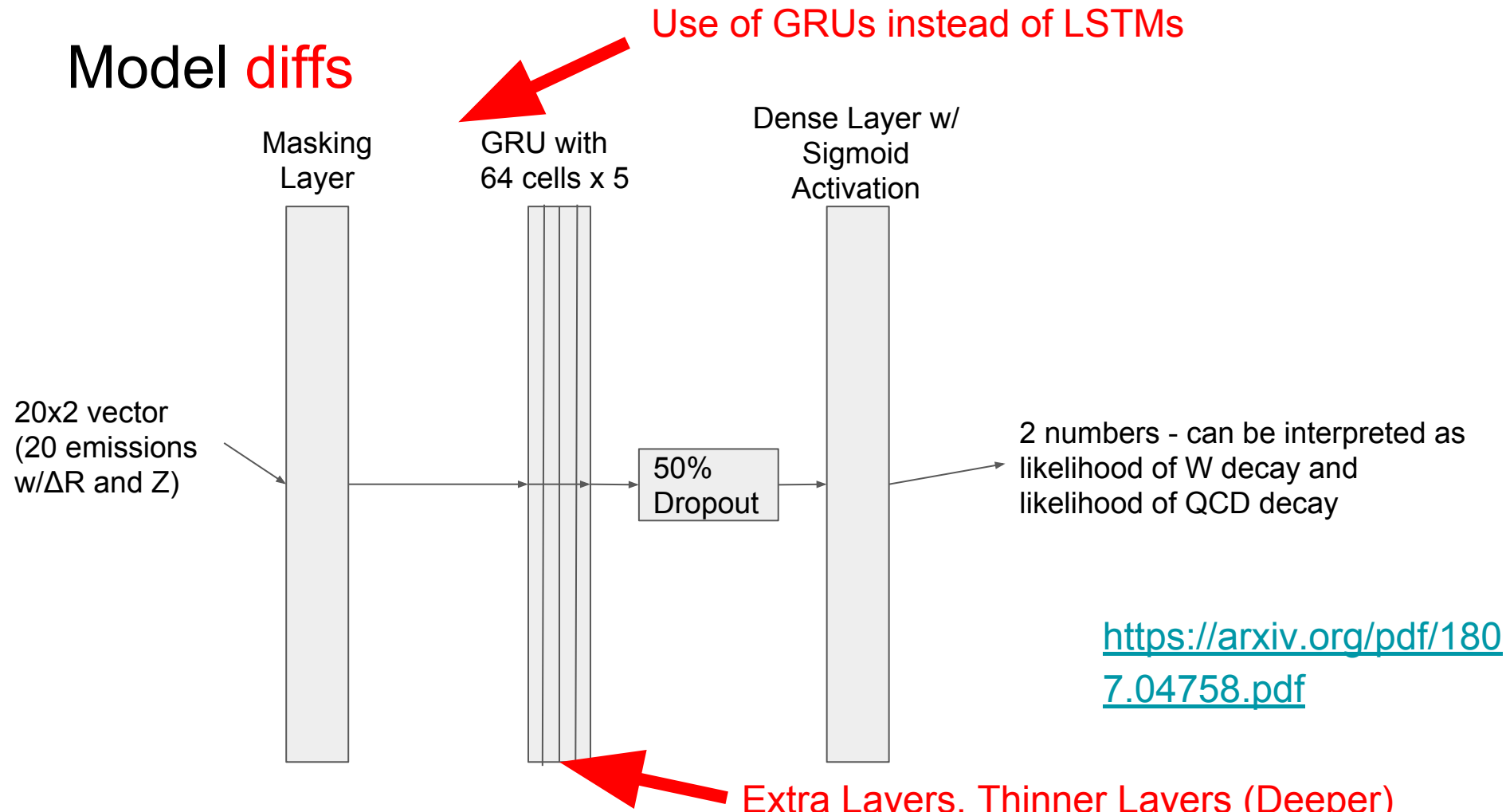
Model Choice

- Combination of trial & error, use of SMAC3
- SMAC3 is a simple Bayesian Optimizer that is easy to setup for python.
- Tried simple RNNs, LSTMs, and GRUs
- Tried various depths & widths
- Optimizer choice and activation function same as in paper
- Based off best performing model in The Lund Plane Paper.

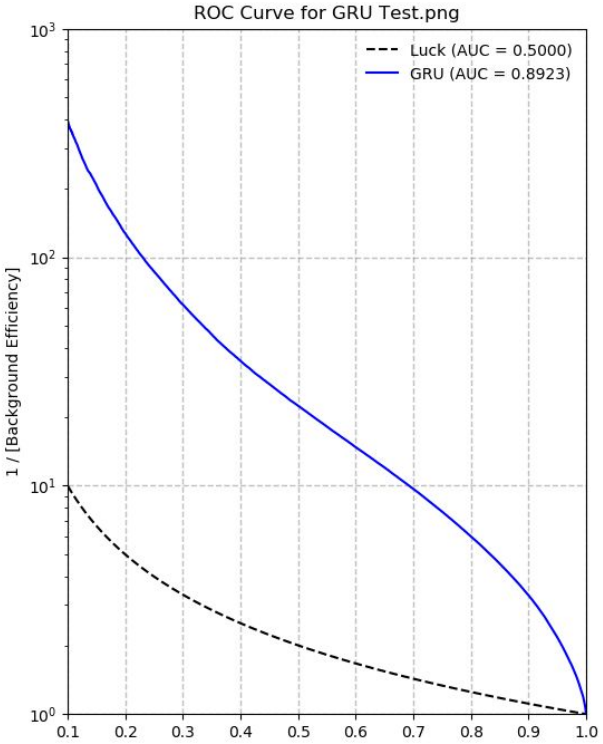
Model

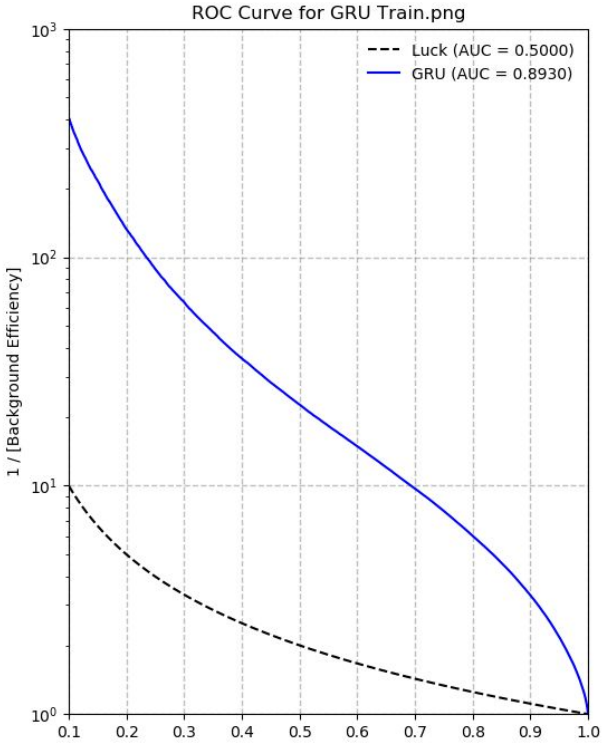

Masking Layer
GRU with 64 cells x 5
Dense Layer w/ Sigmoid Activation
20x2 vector (20 emissions w/ΔR and Z)
50% Dropout
2 numbers - can be interpreted as likelihood of W decay and likelihood of QCD decay

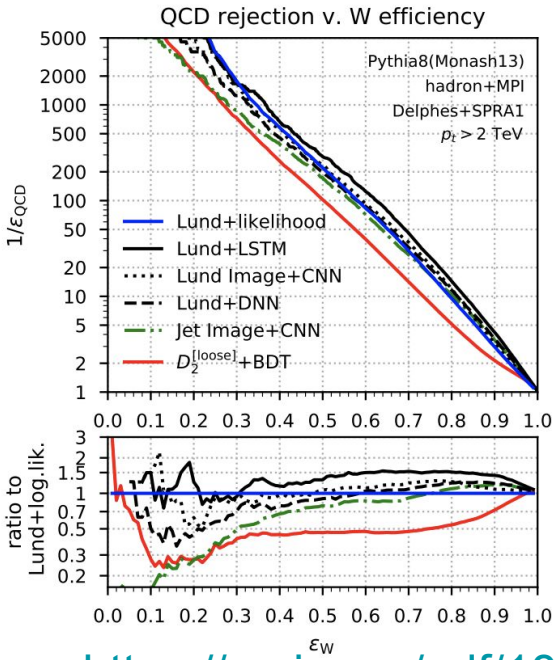
Model diffs


Use of GRUs instead of LSTMs
Masking Layer
GRU with 64 cells x 5
Dense Layer w/ Sigmoid Activation
20x2 vector (20 emissions w/ΔR and Z)
50% Dropout
2 numbers - can be interpreted as likelihood of W decay and likelihood of QCD decay
https://arxiv.org/pdf/1807.04758.pdf
Extra Layers, Thinner Layers (Deeper)

ROCs

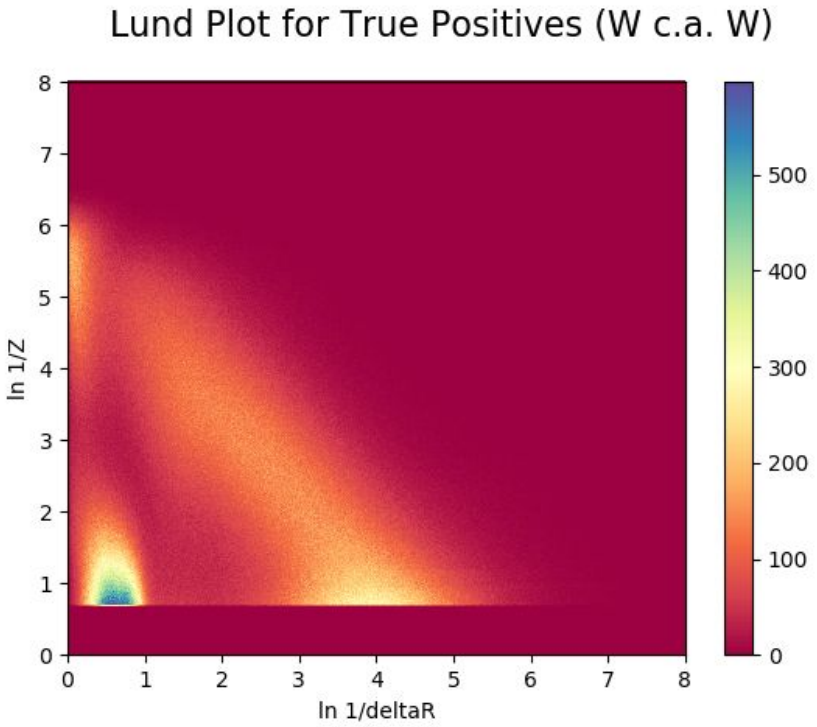


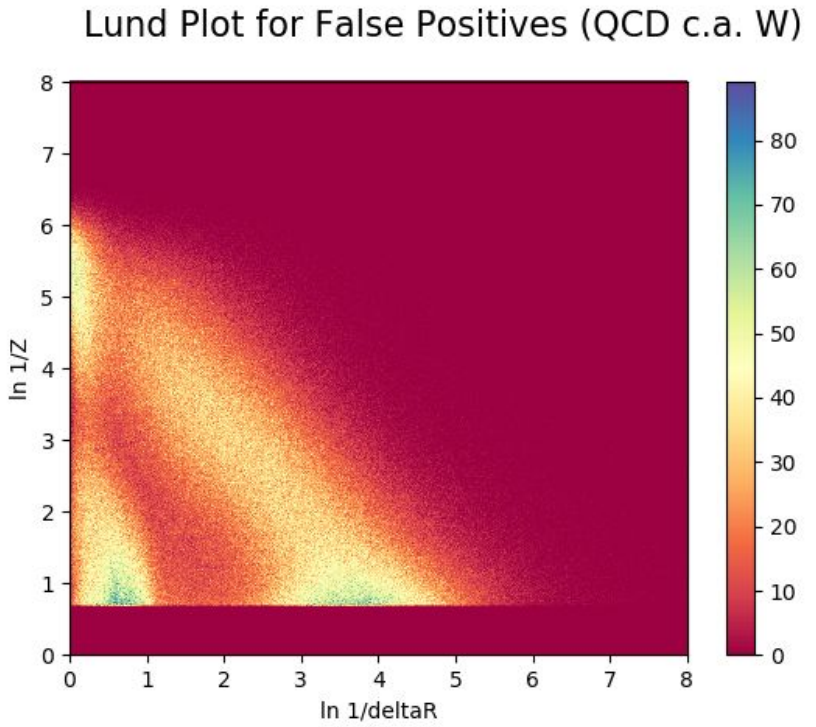




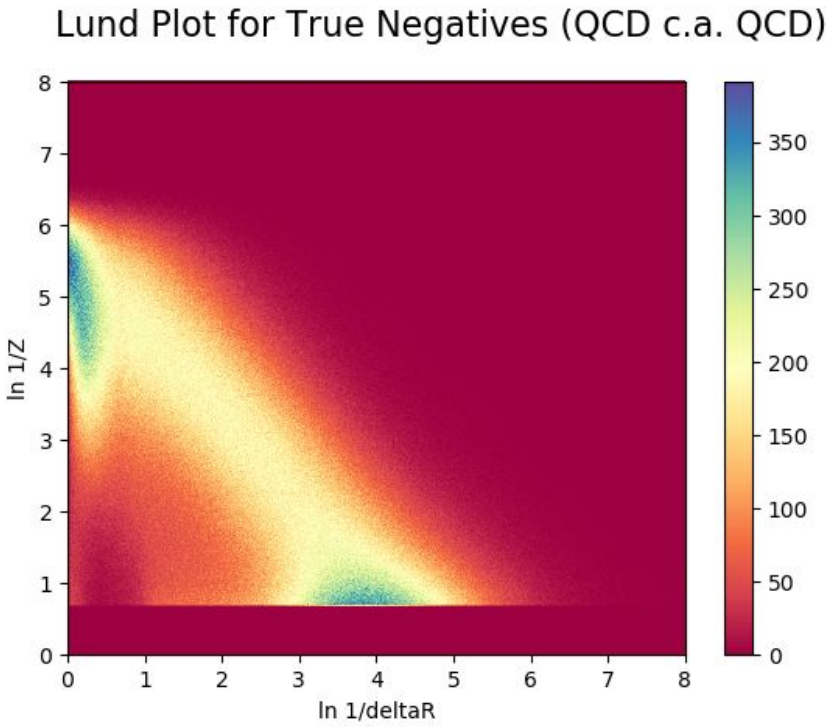
https://arxiv.org/pdf/1807.04758.pdf

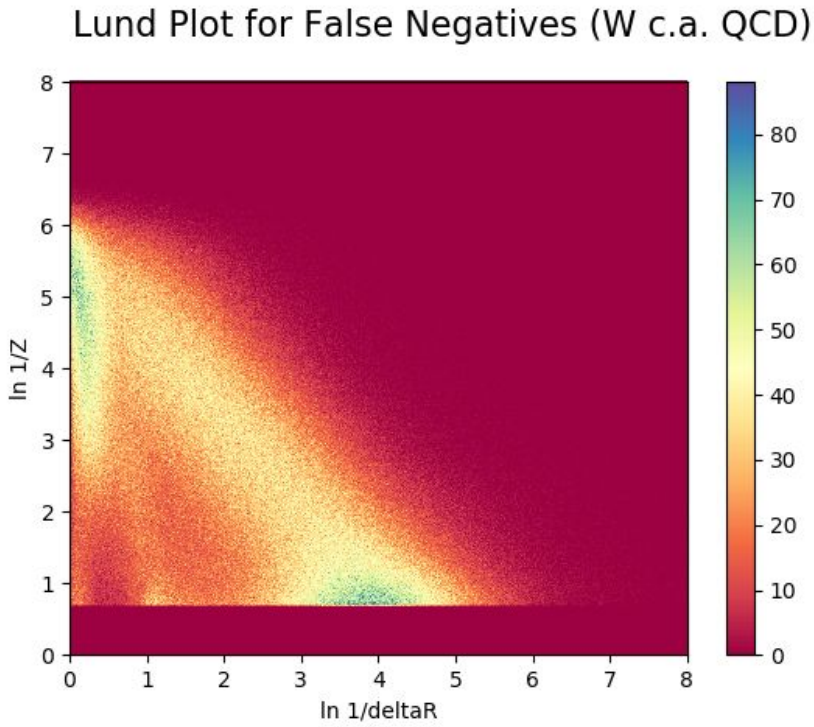
General Confusion Plots



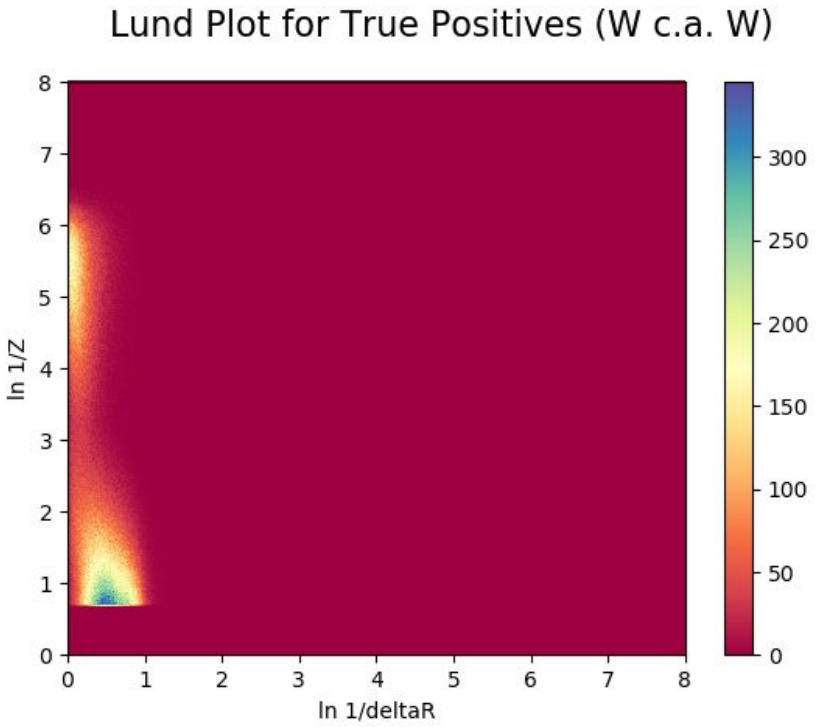


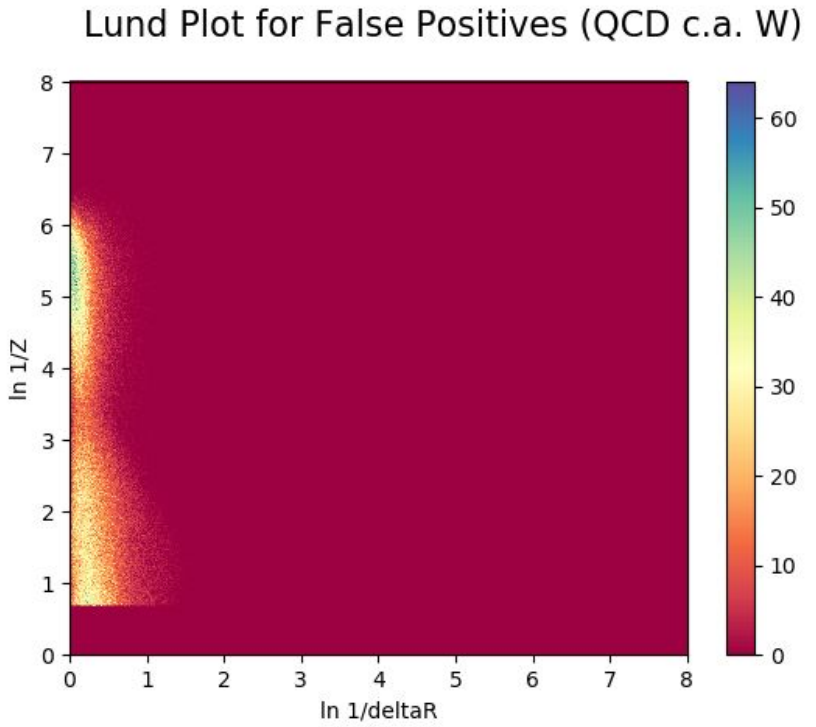
General Confusion Plots



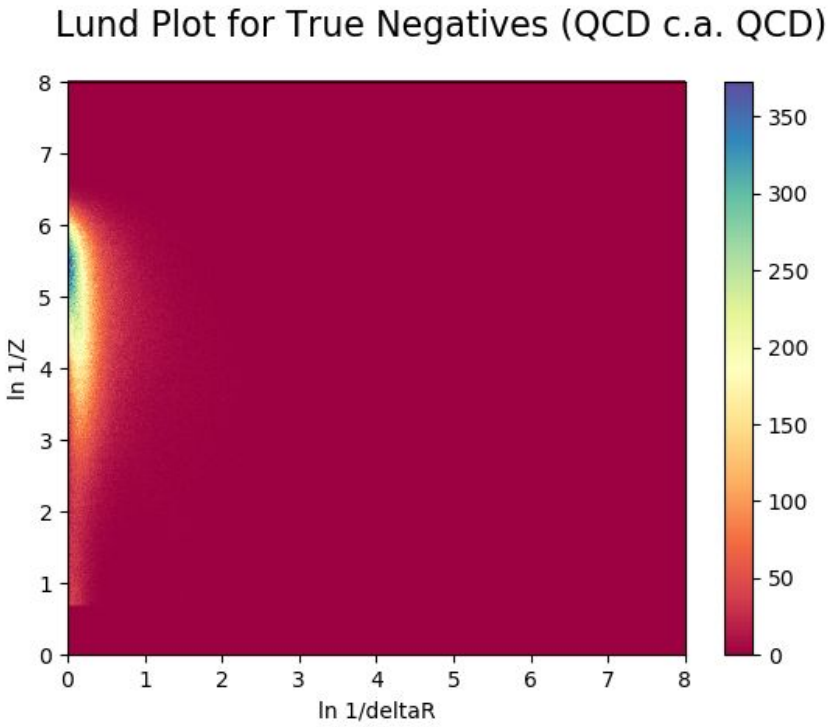


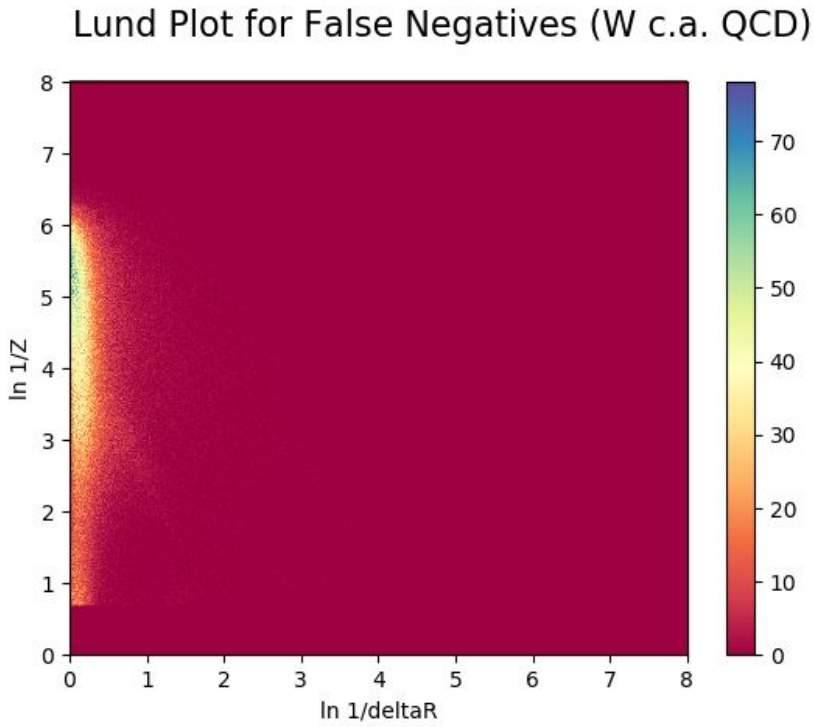
Highest ΔR Confusion Plots



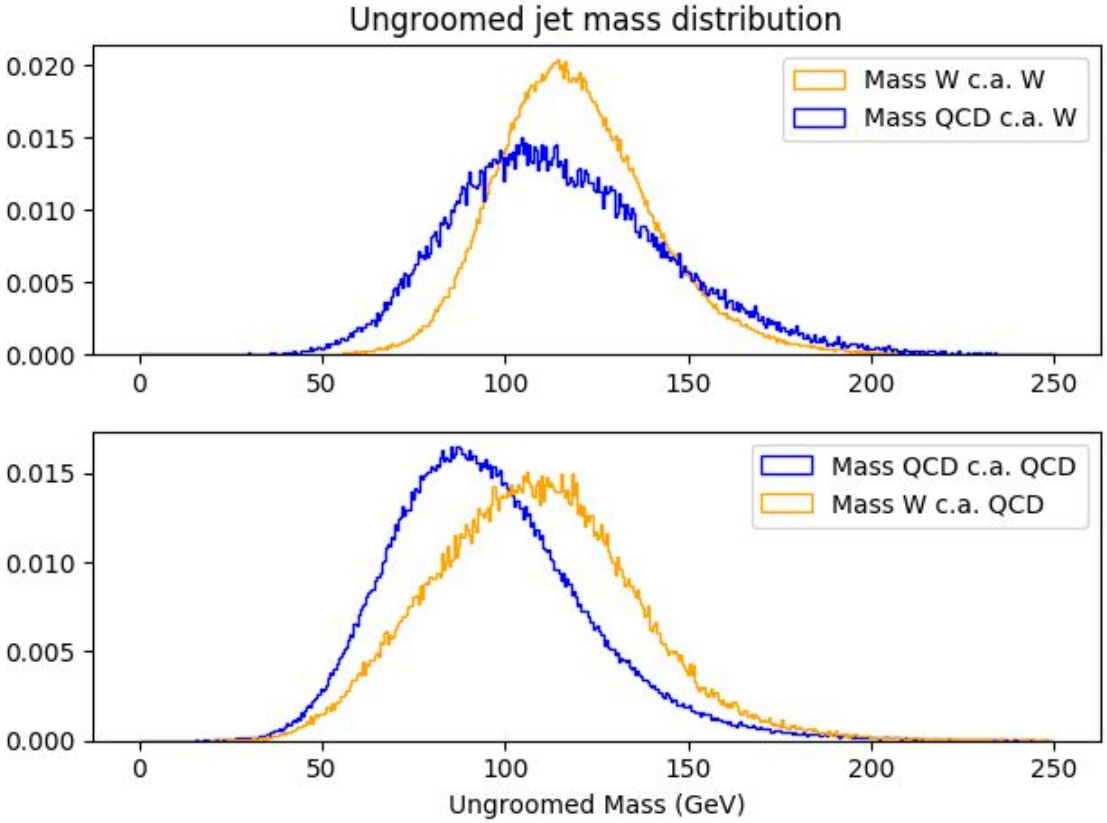


Highest ΔR Confusion Plots





Confusion Mass Plots



Possible improvements

- Better constituent sorting (possibly by Z, or some function based on ΔR and Z)
- More data or better weight normalization would allow deeper networks
- Use an ensemble of models, specialized for specific cases of jets (ΔR bins, pT bins, likelihood output of current model, etc.)

Aside to XGBoost

metric\depth	6	8	10
Test AUC	.7384	.7499	.7739
Train AUC	.7432	.7621	.8049

What went wrong?

- XGBoost requires a one dimensional input, which does not enable us to contain spatial information between jets well
  - A possible solution is to pass multidimensional input to autoencoder first, and then pass the encoding to the XGBoost
  - We did not optimize XGBoost well. In particular width of layers can provide a big gains in accuracy.

Code

- Github: https://github.com/36000/hfsf
- Repository mirrored on HFSF gitlab

Machine Learning in Lund Jet Plane

Tong Ou[+], Congqiao Li[‡], Gang Zhang[§], Samuel Meehan[*], Shih-Chieh Hsu[*]

Nanjing University[+], Peking University[‡], Tsinghua University[§],
University of Washington[*]

US ATLAS Hadronic Final State Forum 2018

14 December, 2018

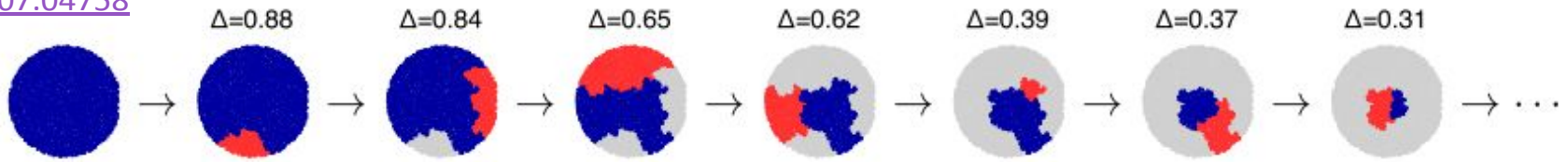
Can RNN with Lund plane help with W tagging?

Signal: W'->WZ
Bkg: QCD dijet

Many thanks to Giordon and Matt for the samples

Lund jet plane

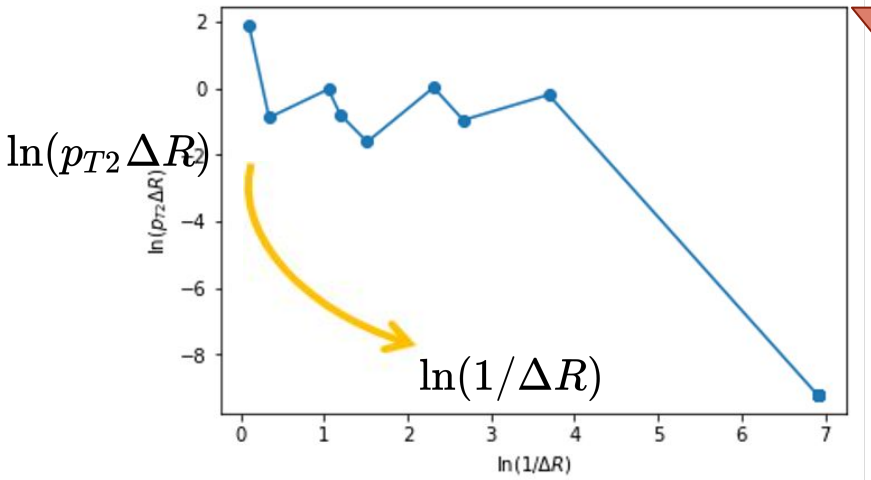
arXiv:1807.04758

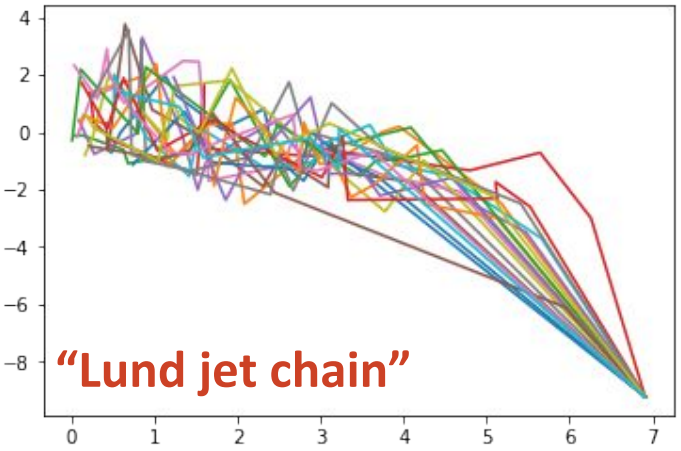


Decluster a jet using C/A algorithm

Calculate

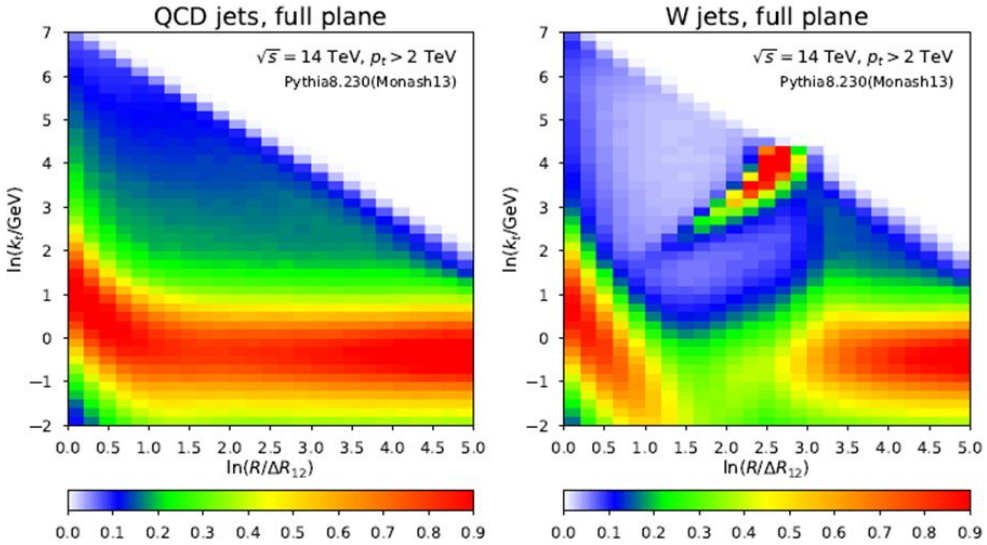
$$\Delta \equiv (y_a - y_b)^2 + (\phi_a - \phi_b)^2, \quad k_t \equiv p_{tb}\Delta_{ab},$$
$$m^2 \equiv (p_a + p_b)^2, \quad z \equiv \frac{p_{tb}}{p_{ta}+p_{tb}}, \quad \psi \equiv \tan^{-1}\frac{y_b - y_a}{\phi_b - \phi_a}.$$





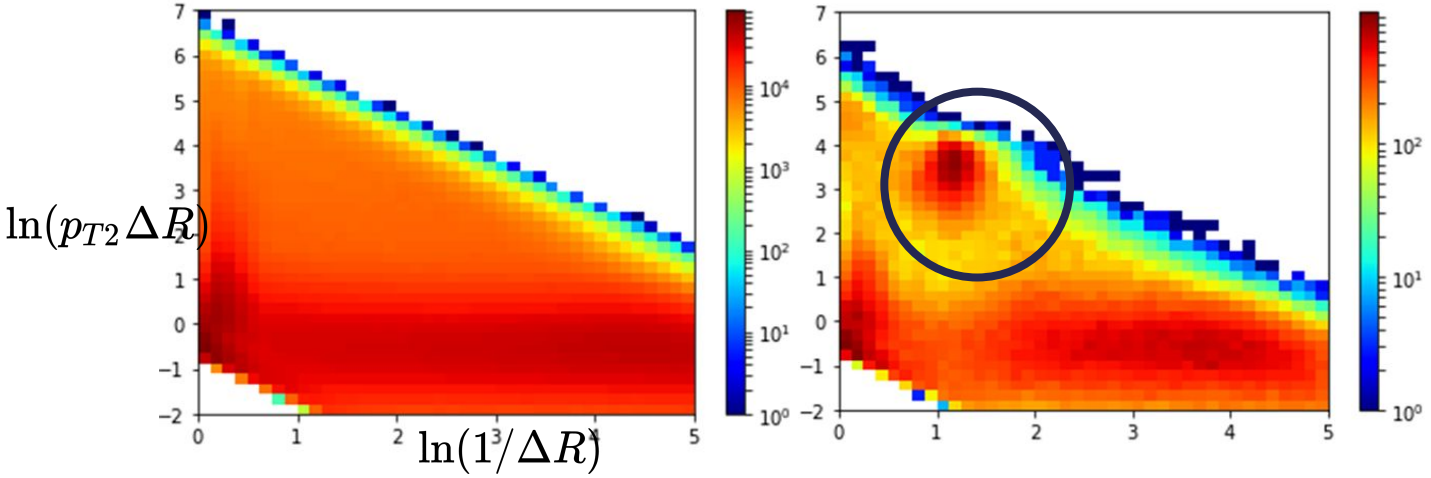
Lund jet plane

arXiv:1807.04758



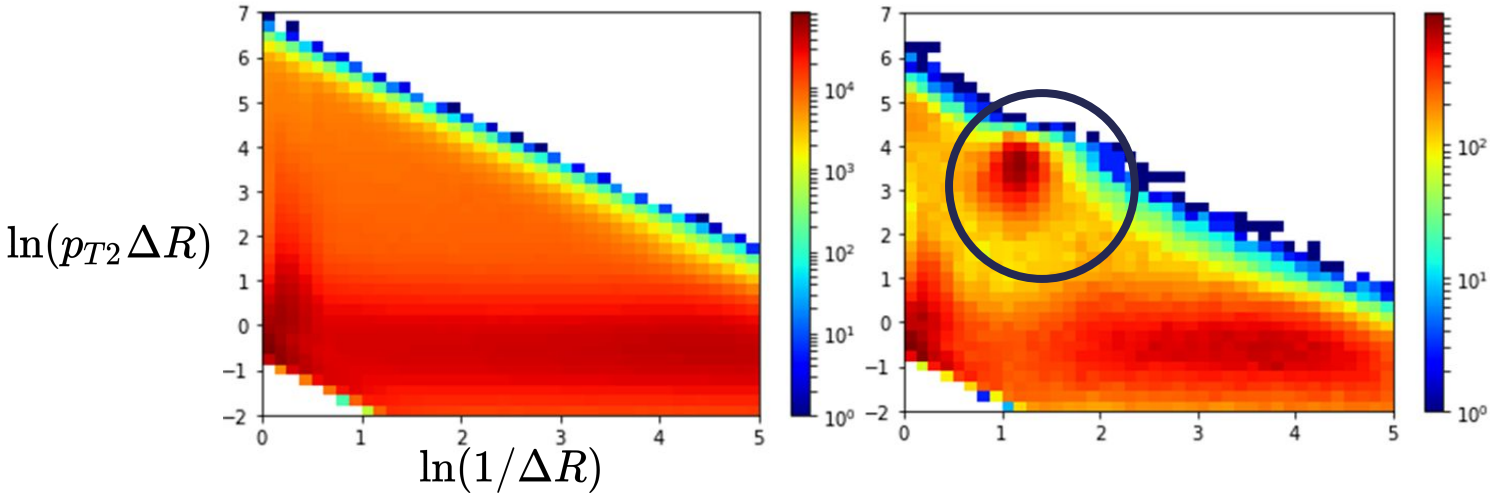
pT > 2000 GeV

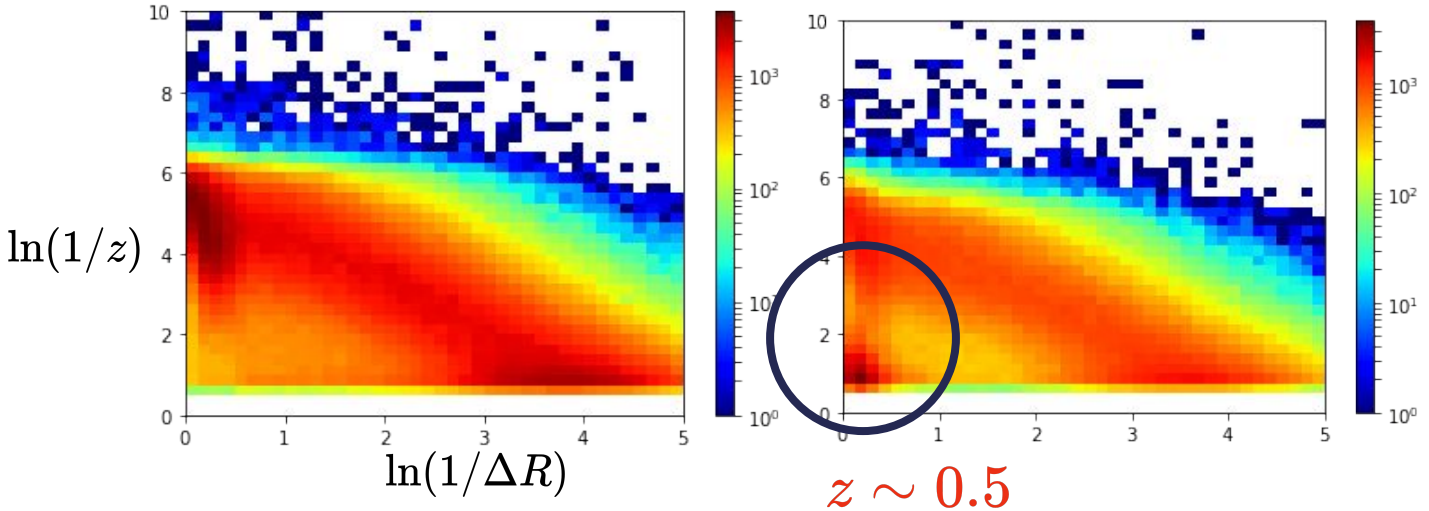
Reproduce the plots



pT > 500 GeV

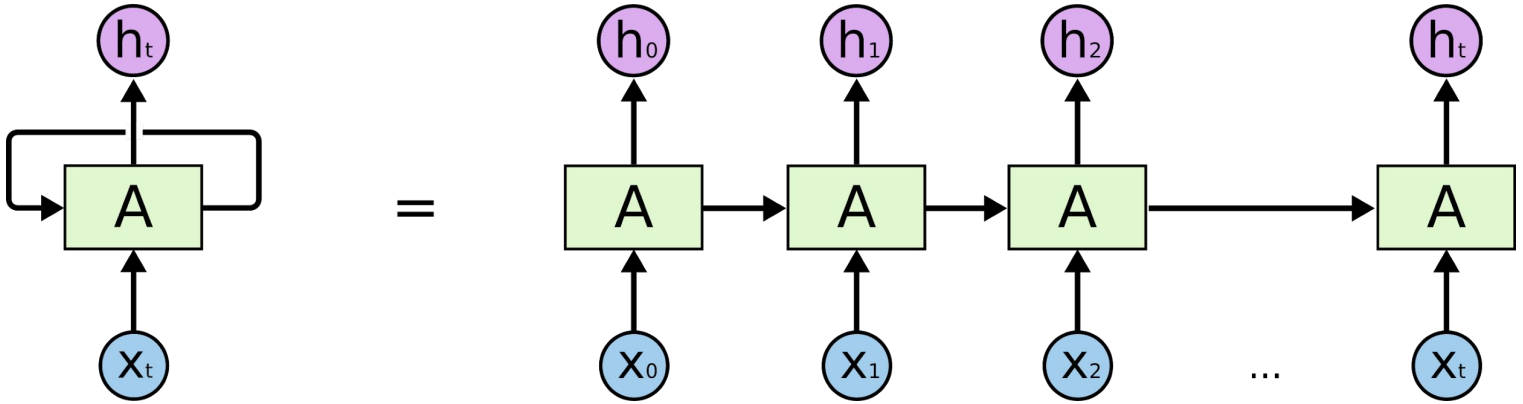
Lund jet plane



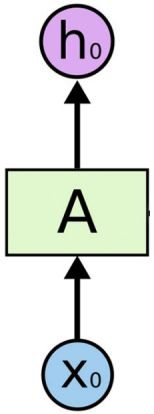


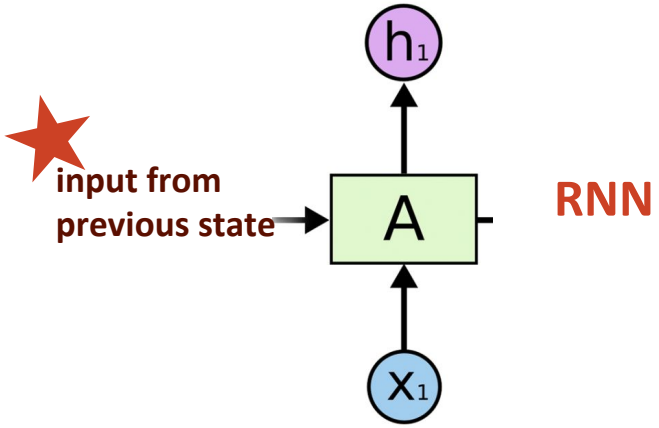
RNN — Recurrent neural network

RNN structure



Normal neural network A

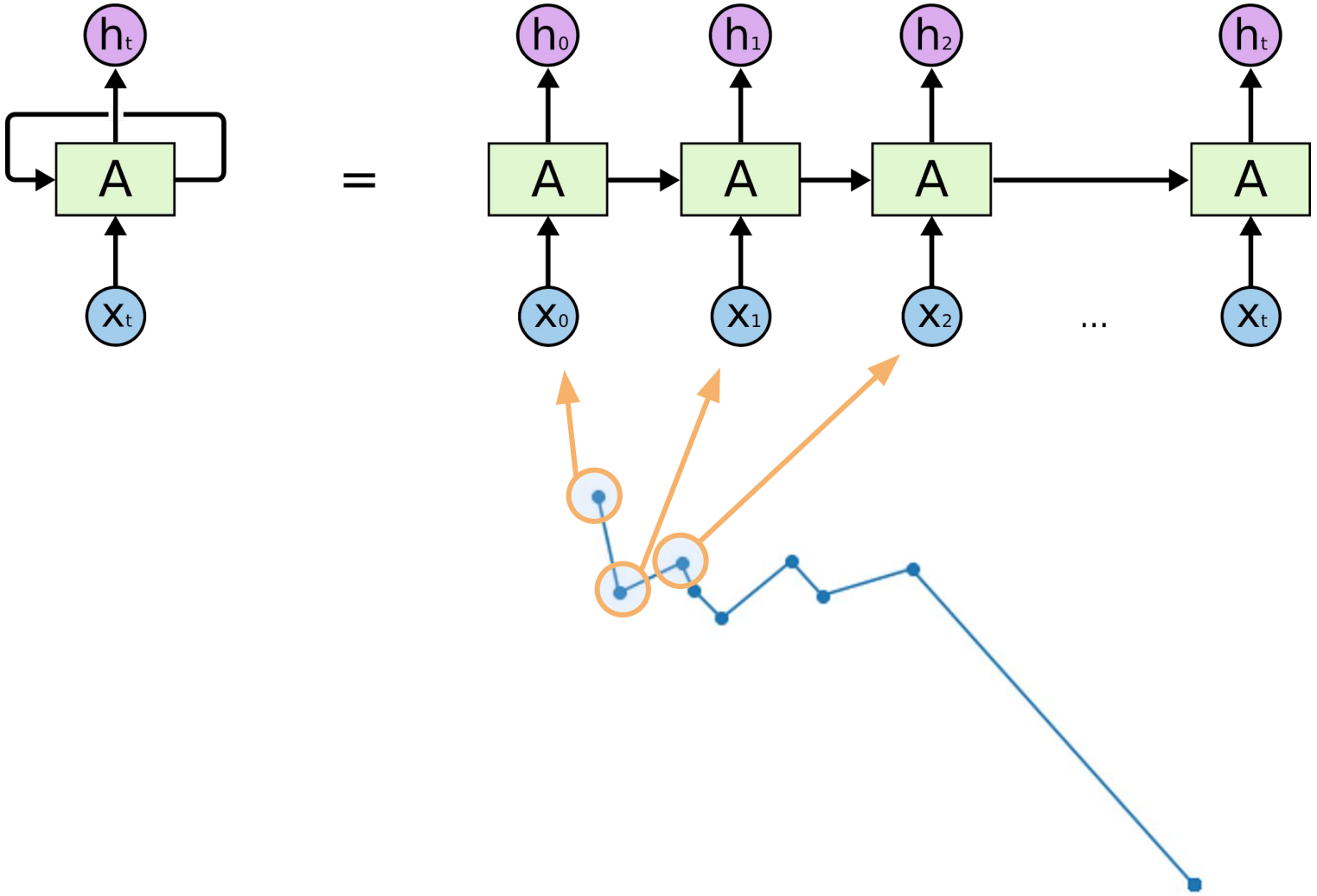




RNN

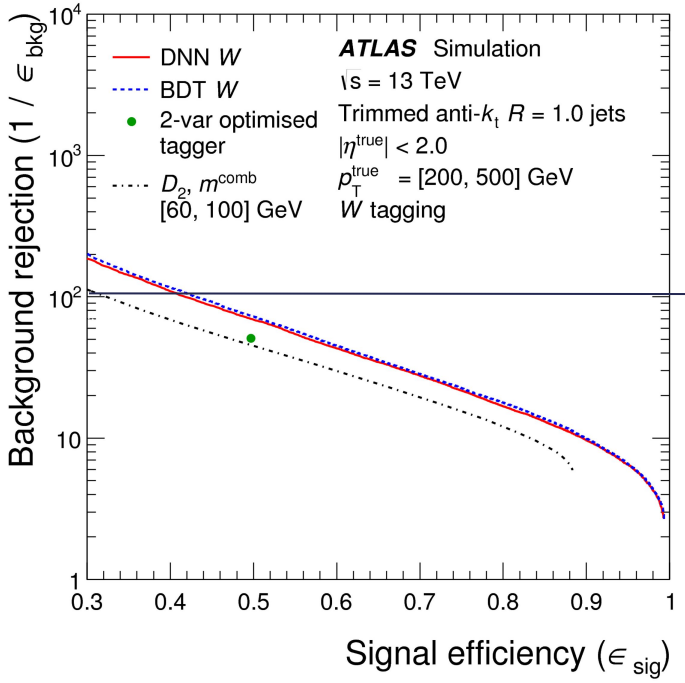
RNN with Lund Jet Plane

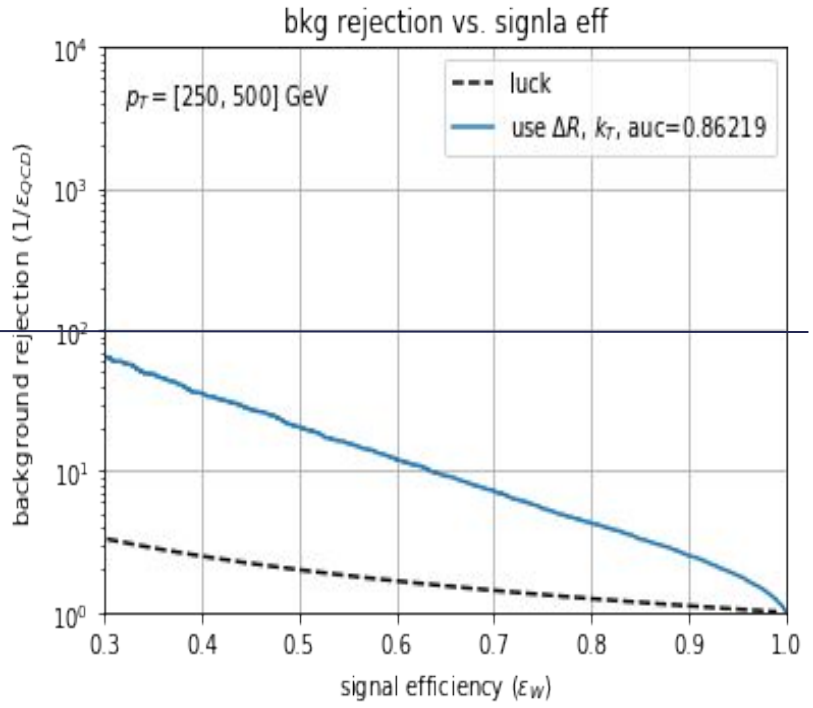
RNN structure



RNN with Lund Jet Plane

pT = [200, 500] GeV

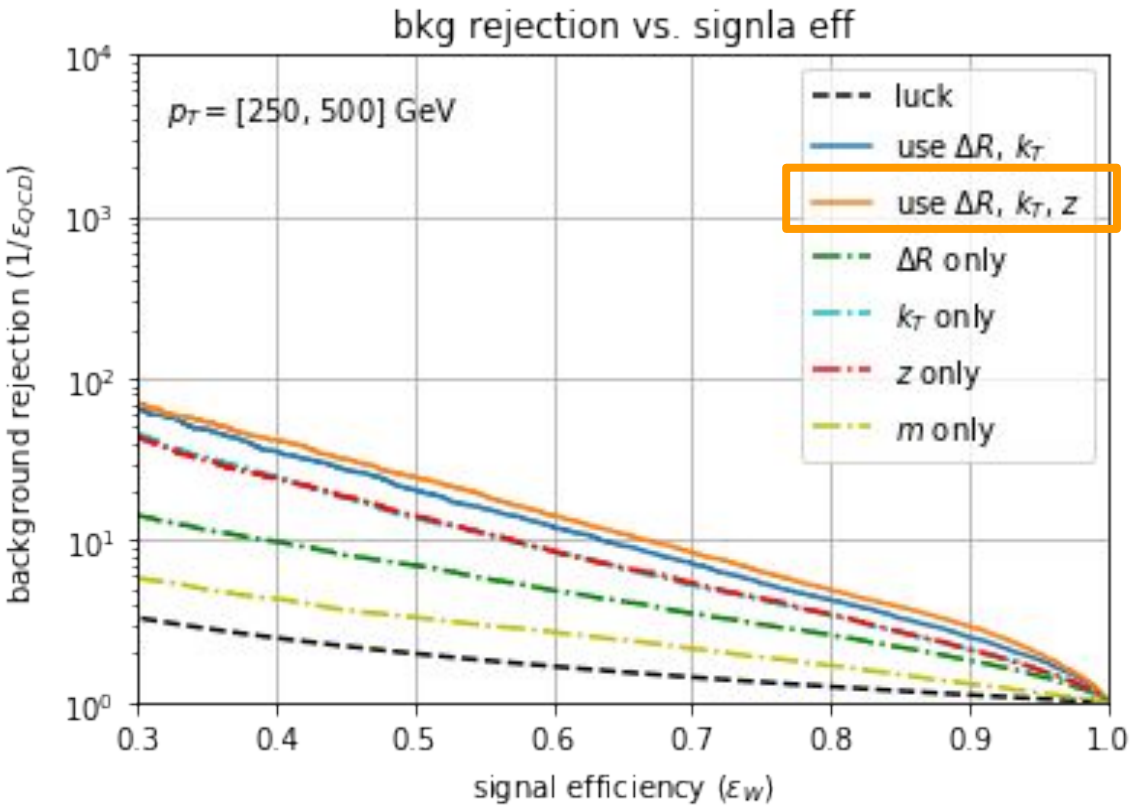




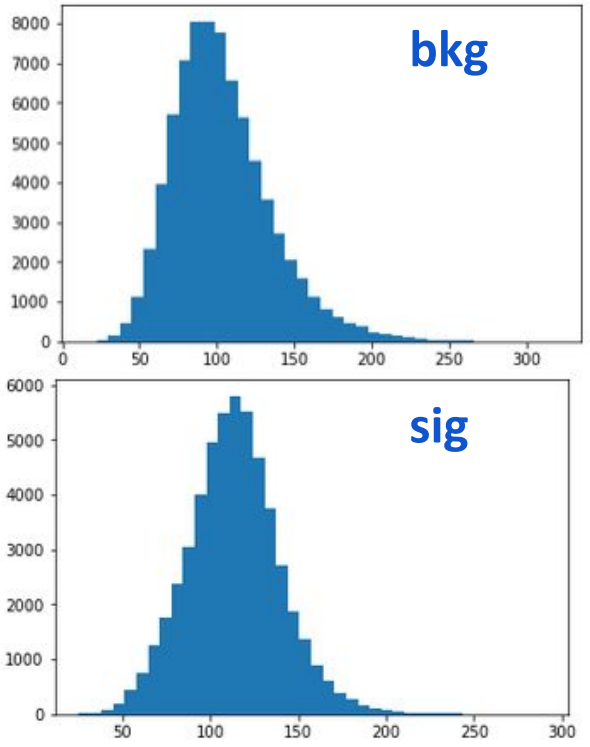
➔ **Other choices?**

- add variable: mass, z
- re-order the chain?s

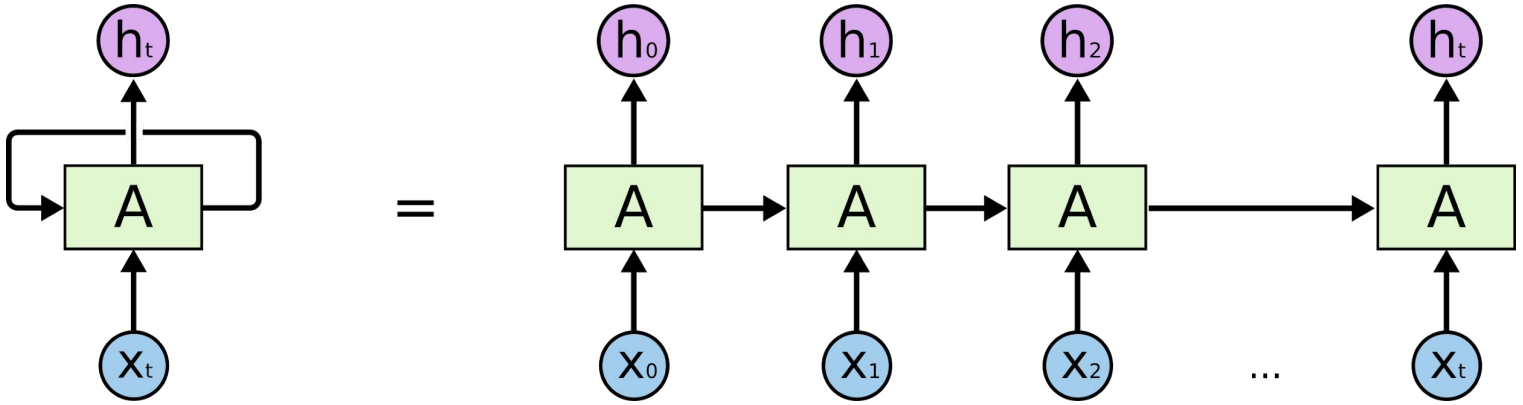
RNN with Lund Jet Plane



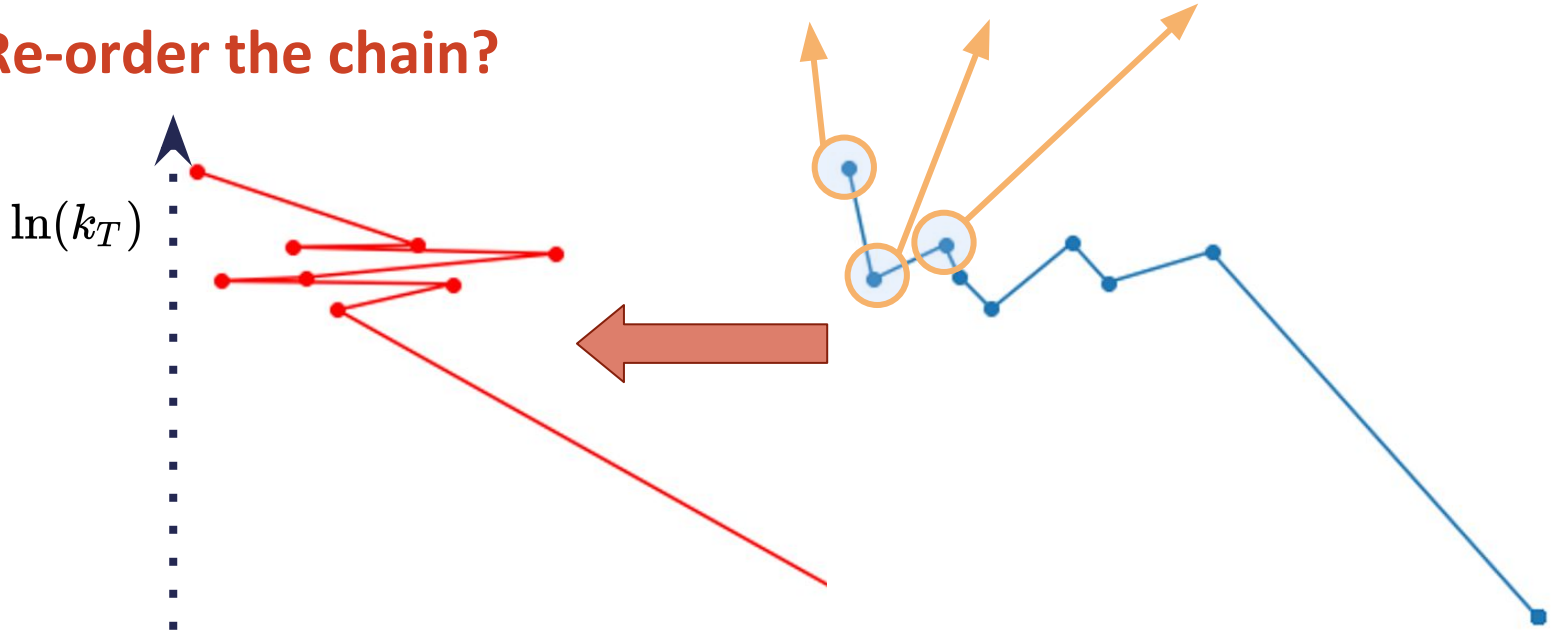
Jet mass distribution?



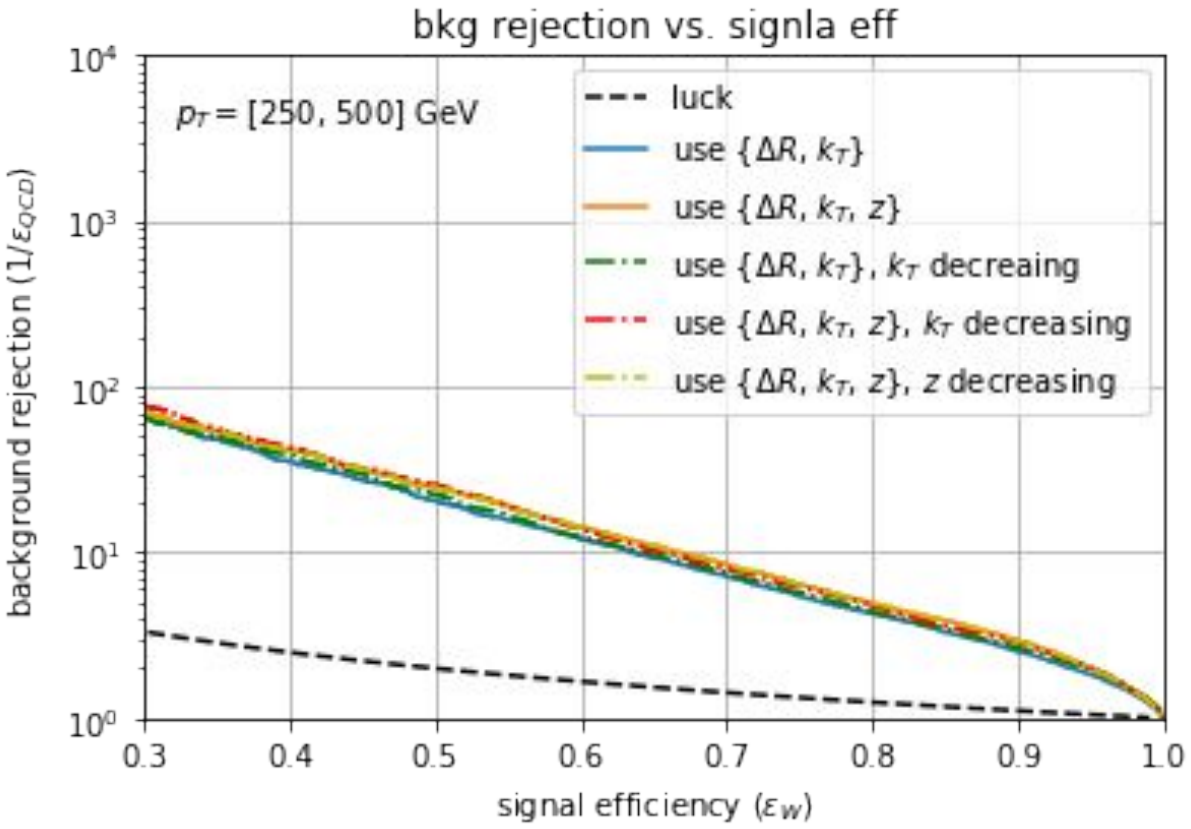
RNN with Lund Jet Plane



Re-order the chain?



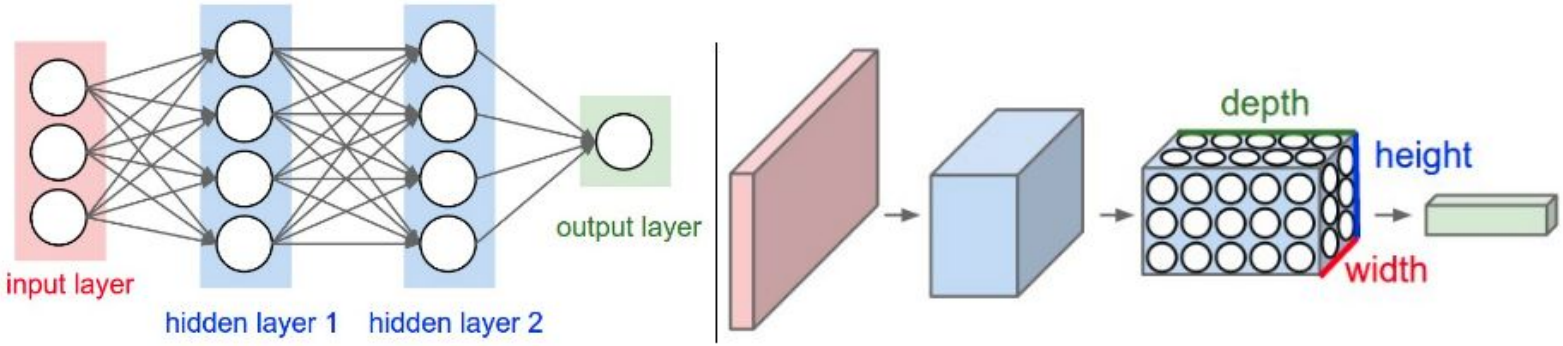
RNN with Lund Jet Plane



How about CNN?

Convolutional Neural Network (CNN)

Regular neural network ***CNN***



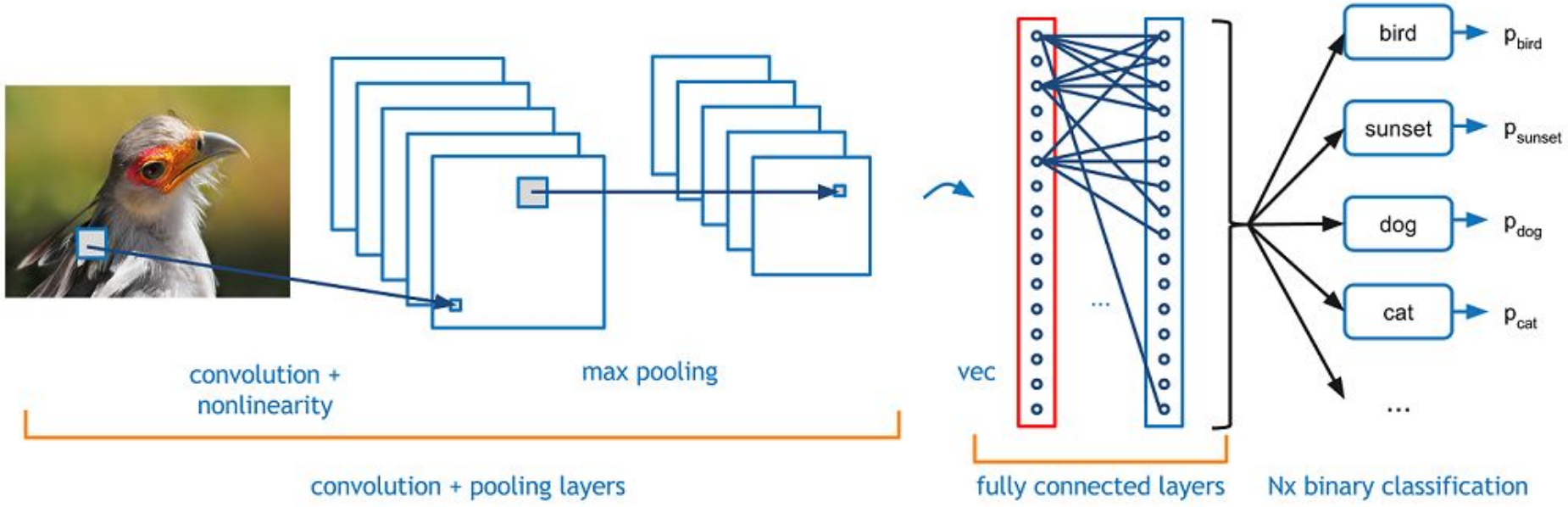
Fully connected neurons

Keep only information in depth dimension (3D->1D)

Depth:
R, G, B

Convolutional Neural Network (CNN)

➔ Layers: **Convolutional Layer**, **Pooling Layer**, and **Fully-Connected Layer**

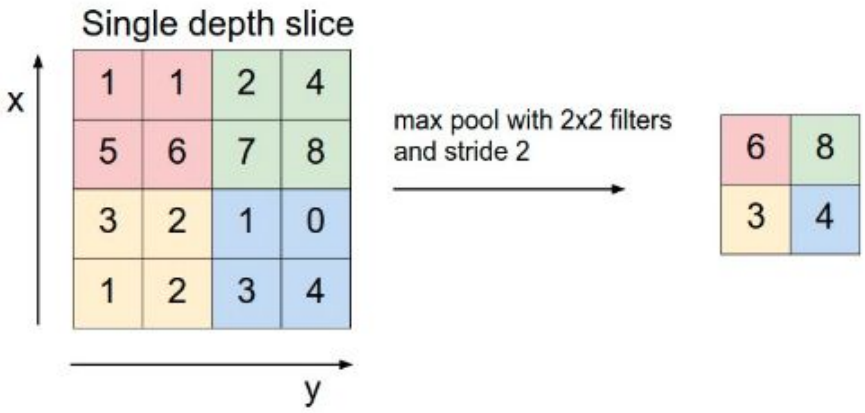


Convolutional Neural Network (CNN)

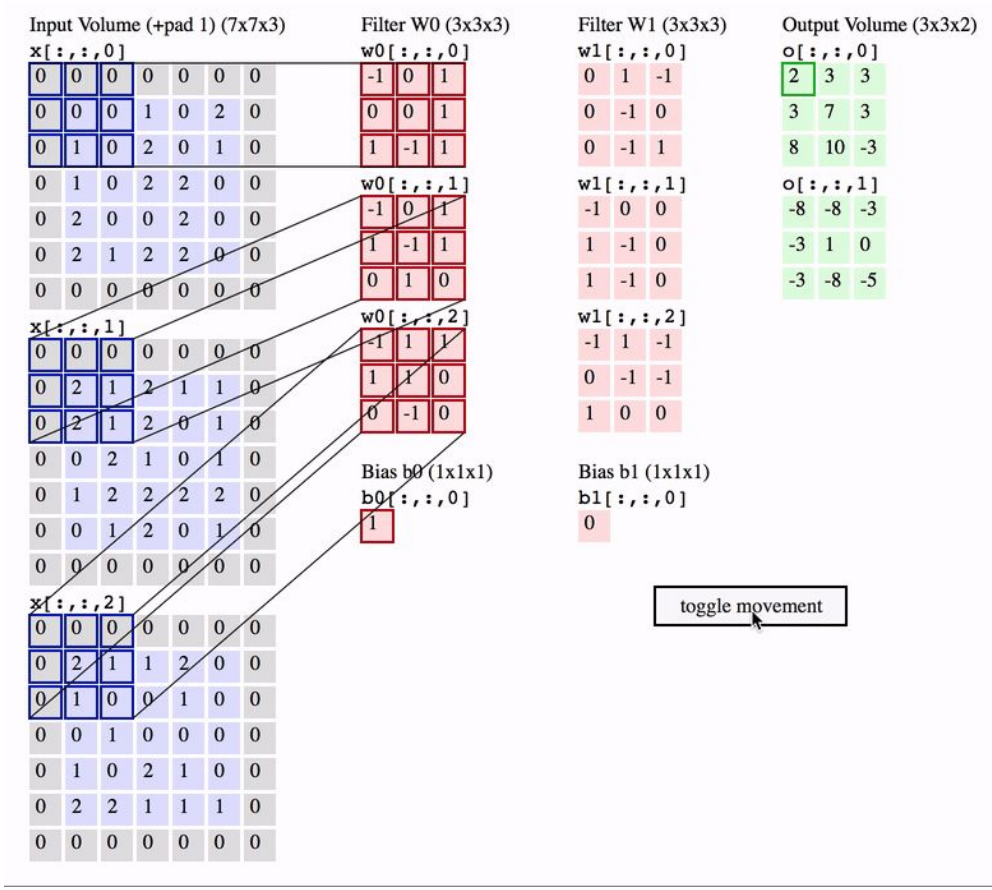
How do the layers work?

- ➔ Convolution layer: Multiply input and filter, then sum it up
- ➔ Pooling layer: Reduce spatial size

Pooling layer



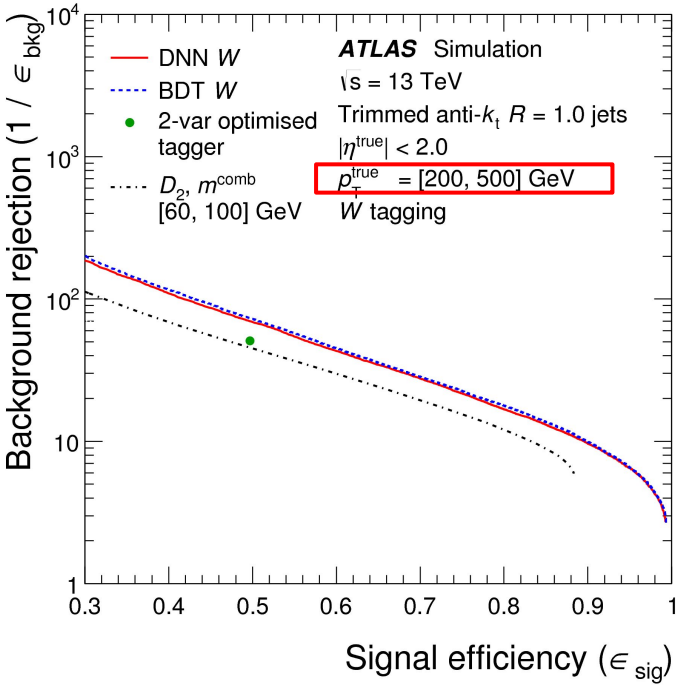
Convolution layer (2D)



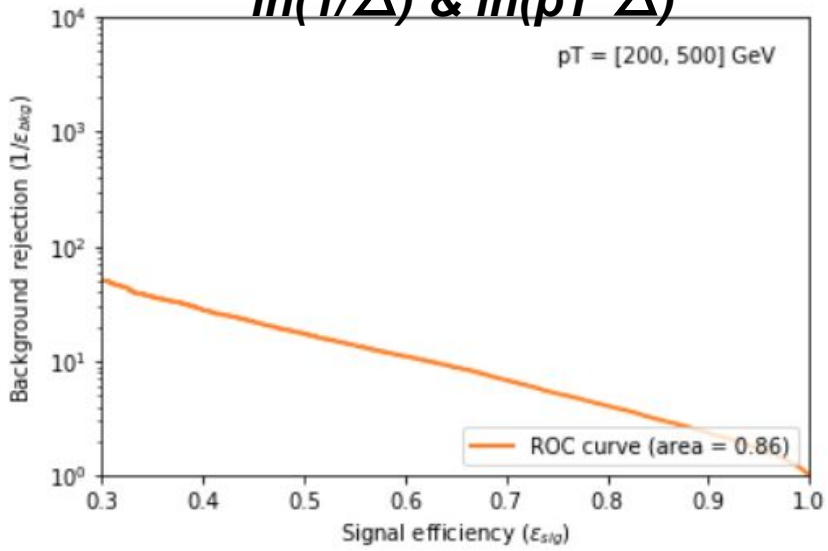
CNN in Lund Jet Plane

➔ W vs QCD in low pT range

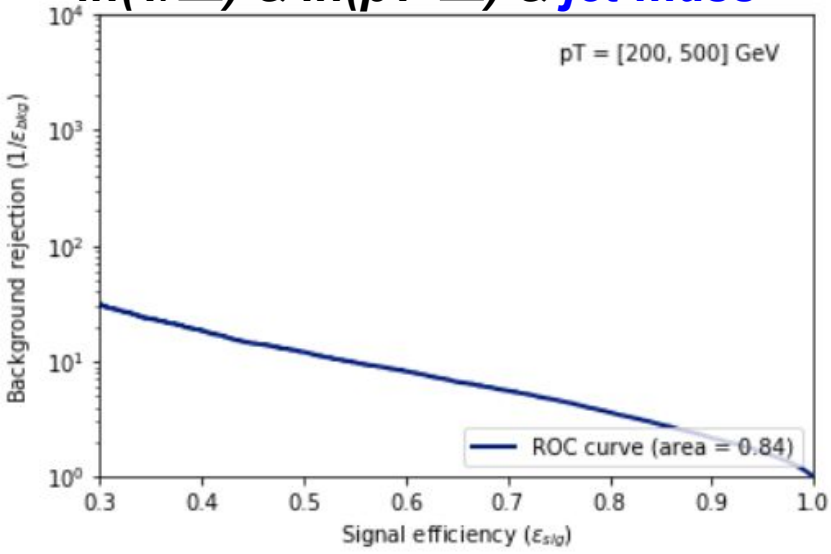
Latest taggers



ln(1/Δ) & ln(pT*Δ)



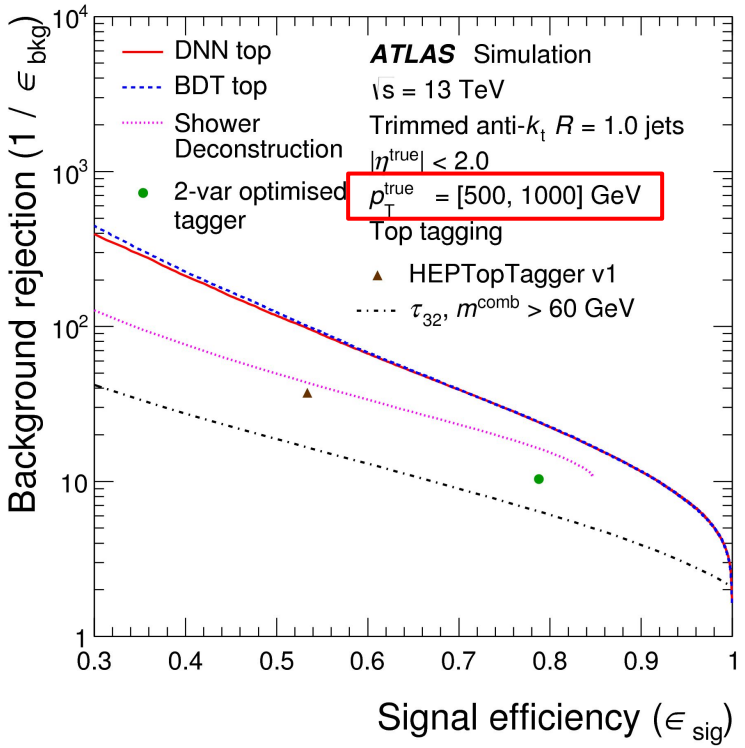
ln(1/Δ) & ln(pT*Δ) & jet-mass



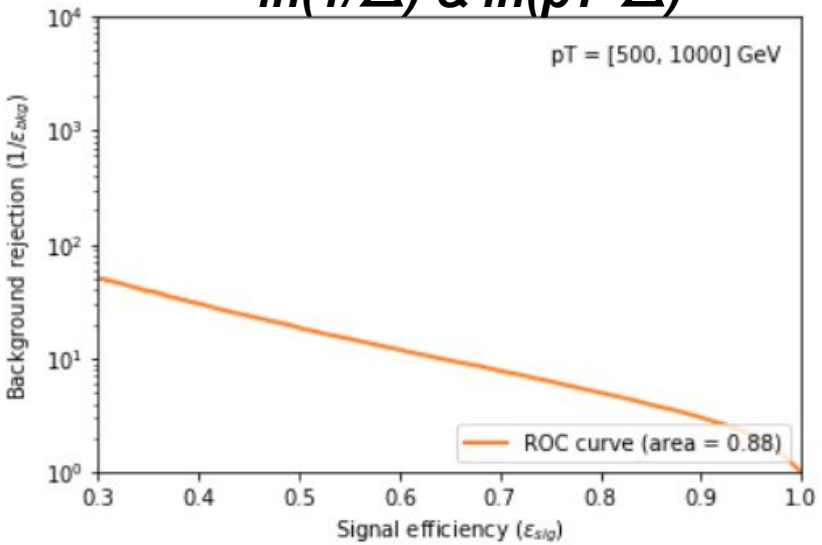
CNN in Lund Jet Plane

➔ Tops vs QCD in low pT range

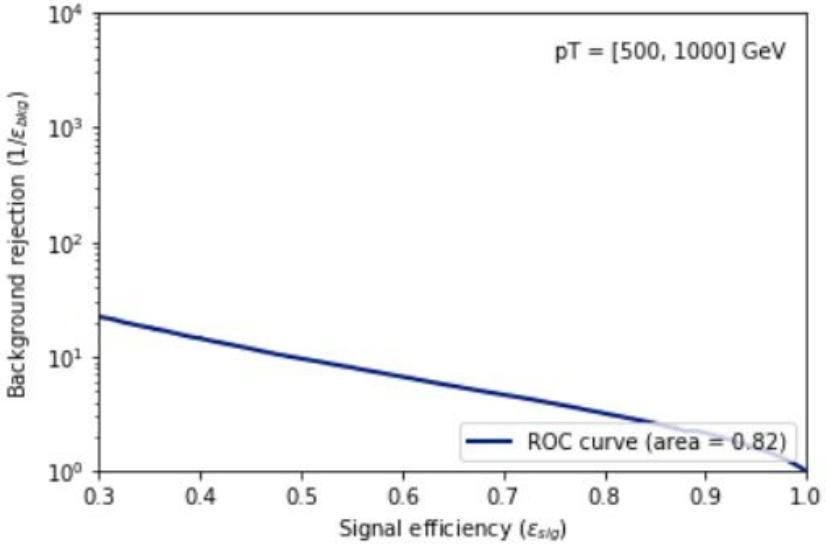
Latest taggers



ln(1/Δ) & ln(pT*Δ)



ln(1/Δ) & ln(pT*Δ) & jet-mass

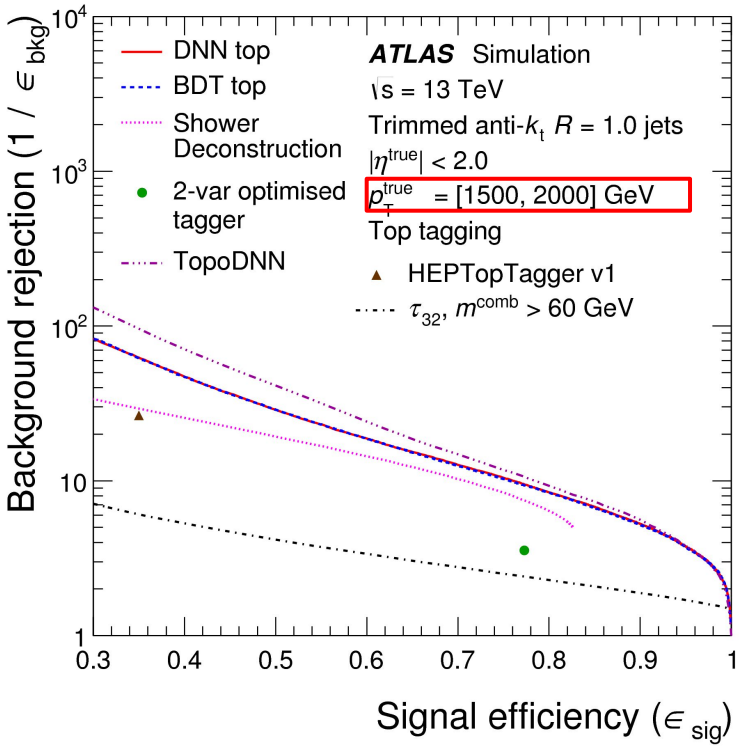


CNN in Lund Jet Plane

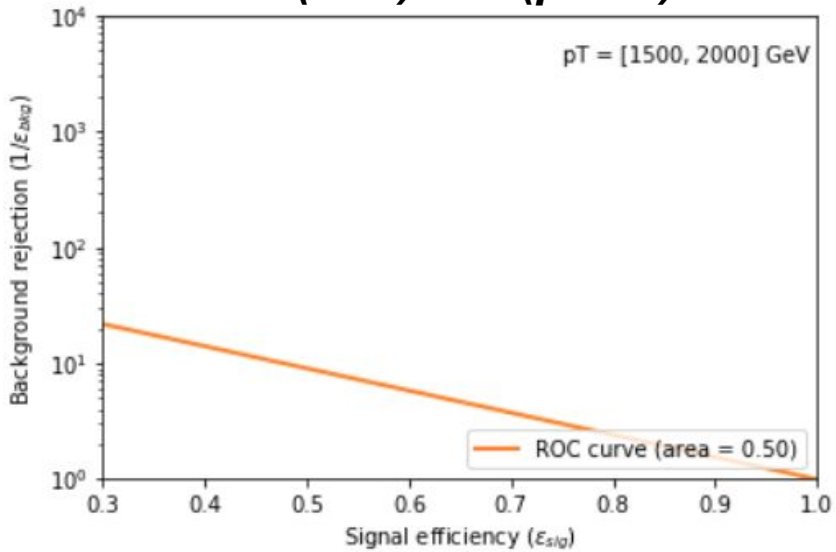
➔ Tops vs QCD in high pT range

Issue: Low statistic?

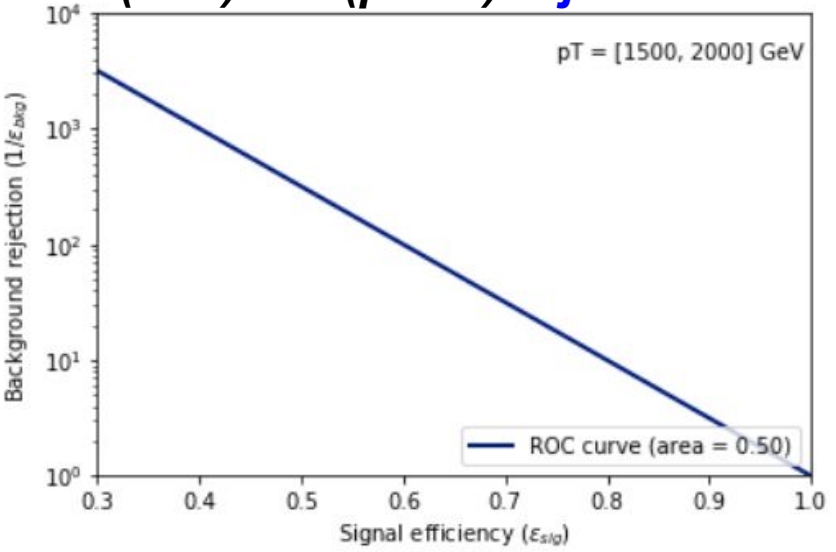
Latest taggers



ln(1/Δ) & ln(pT*Δ)



ln(1/Δ) & ln(pT*Δ) & jet-mass



Thank you!