Learning from the Lund plane

LAL, Orsay, 10 September 2019

Frédéric Dreyer

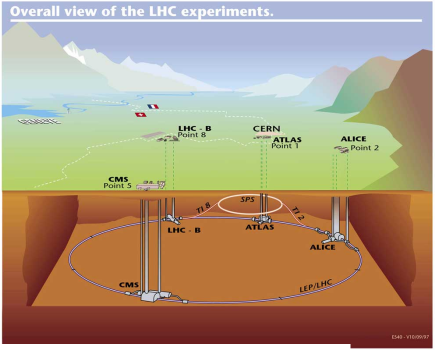
UNIVERSITY OF OXFORD

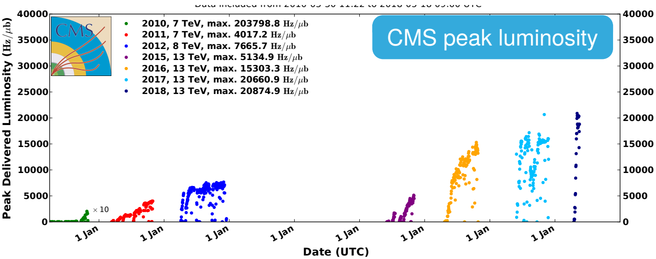
based on arXiv:1807.04758, arXiv:1903.09644 and arXiv:1909.01359

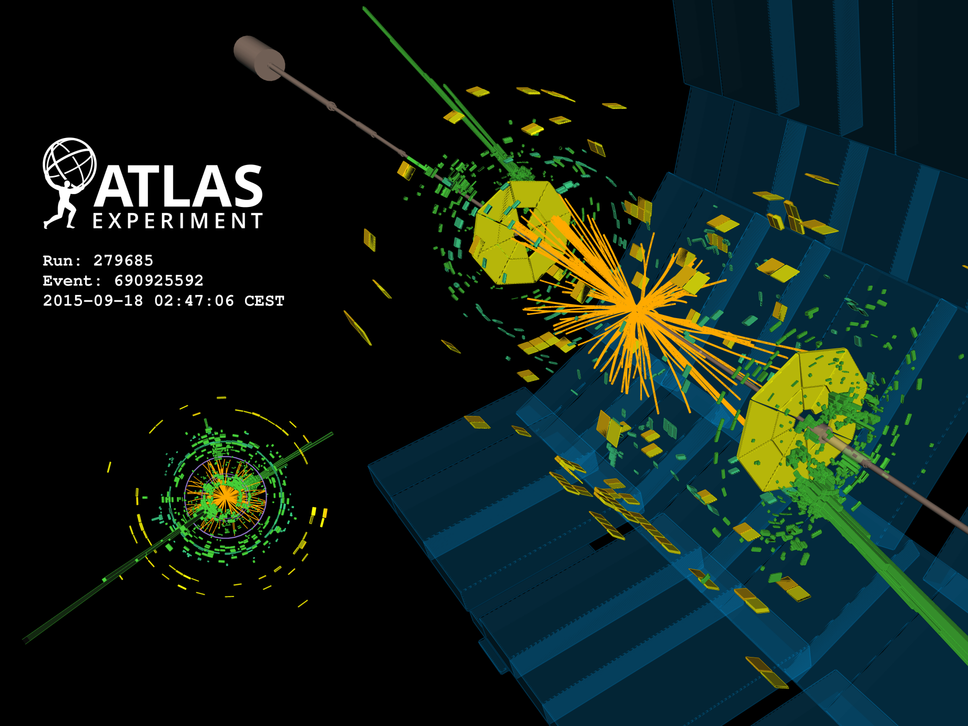
with Stefano Carrazza, Gavin Salam & Gregory Soyez

Physics at the high energy frontier

- LHC has been colliding protons at $13\,\mathrm{TeV}$ center-of-mass energy.
- Particle physics entering precision phase in study of EW symmetry breaking.
- Searching for new physics at the highest energy ever attained.



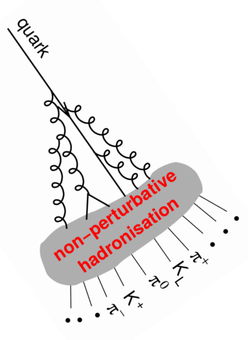



ATLAS
EXPERIMENT
Run: 279685
Event: 690925592
2015-09-18 02:47:06 CEST

JET SUBSTRUCTURE AND MACHINE LEARNING

Jets as proxies for partons

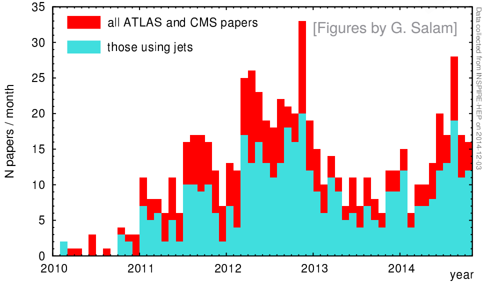
Because of color confinement, quarks and gluons shower and hadronise immediately into collimated bunches of particles.



Hadronic **jets** can emerge from a number of processes

- scattering of partons inside colliding protons,
- hadronic decay of heavy particles,
- radiative gluon emission from partons, ...

Jets are prevalent at hadron colliders



Jet algorithms

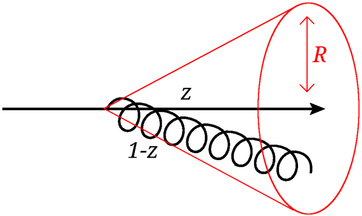
A jet algorithm maps final state particle momenta to jet momenta.

$$\underbrace{\{p_i\}}_{\text{particles}} \Longrightarrow \underbrace{\{j_k\}}_{\text{jets}}$$

This requires an external parameter, the jet radius R, specifying up to which angle separate partons are recombined into a single jet.

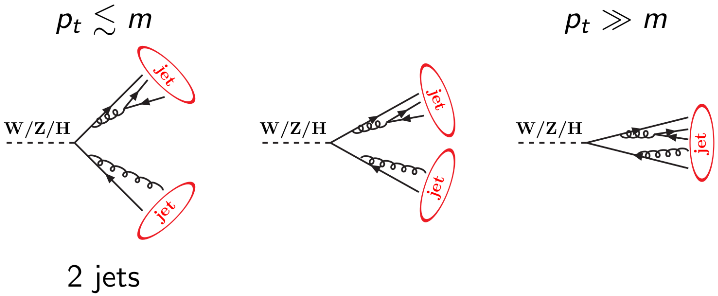
Basic idea of jet algorithm is to invert QCD branching process, clustering pairs which are closest in metric defined by the divergence structure of the theory.

$$d_{ij} = \min(k_{t,i}^{2p}, k_{t,j}^{2p}) \frac{\Delta_{ij}^2}{R^2}$$



Boosted objects at the LHC

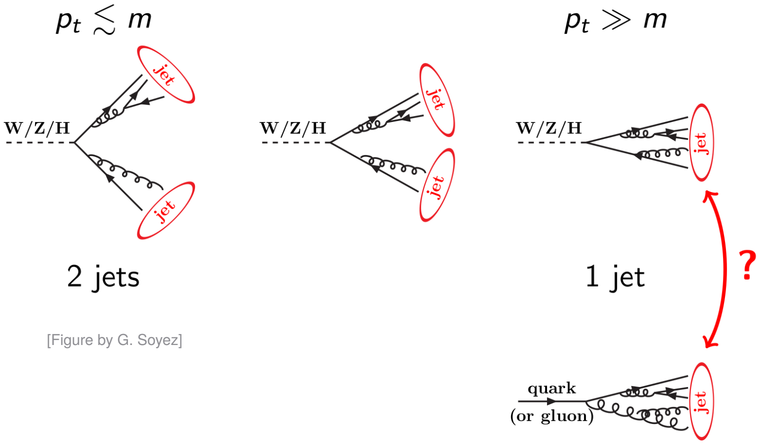
- At LHC energies, EW-scale particles (W/Z/t…) are often produced with $p_t \gg m$, leading to collimated decays.
- Hadronic decay products are thus often reconstructed into single jets.



[Figure by G. Soyez]

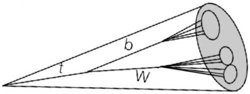
Boosted objects at the LHC

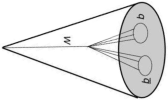
- At LHC energies, EW-scale particles (W/Z/t…) are often produced with $p_t \gg m$, leading to collimated decays.
- Hadronic decay products are thus often reconstructed into single jets.



[Figure by G. Soyez]

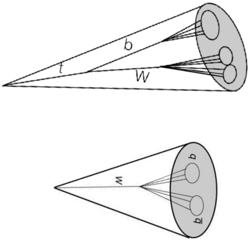
- Many techniques developed to identify hard structure of a jet based on radiation patterns.
- In principle, simplest way to identify these boosted objects is by looking at the mass of the jet.

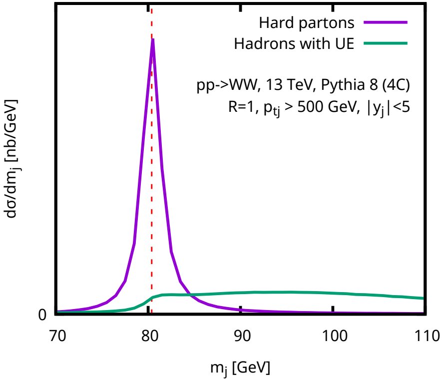




Boosted objects at the LHC

- Many techniques developed to identify hard structure of a jet based on radiation patterns.
- In principle, simplest way to identify these boosted objects is by looking at the mass of the jet.
- But jet mass distribution is highly distorted by QCD radiation and pileup.





Identifying boosted objects

Two main approaches to study boosted decays:

1. Manually constructing substructure observables that help distinguish between different origins of jets.

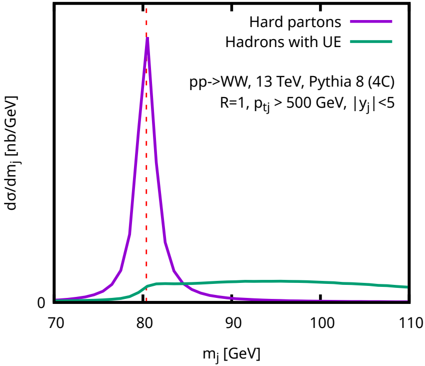
2. Apply machine learning models trained on large input images or observable basis.

Aim of this talk: new approaches bridging some of the gap between these two techniques.

- Mass peak can be partly reconstructed by removing unassociated soft wide-angle radiation (grooming).
- Recurse through clustering tree and remove soft branch if

$$\frac{\min(p_{t,1}, p_{t,2})}{p_{t,1} + p_{t,2}} < z_{\text{cut}} \left(\frac{\Delta R_{12}}{R_0}\right)^{\beta}$$

W jet



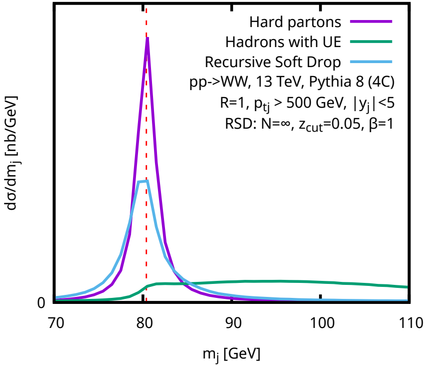
[Dasgupta, Fregoso, Marzani, Salam JHEP 1309 (2013) 029]
[Larkoski, Marzani, Soyez, Thaler JHEP 1405 (2014) 146]
[FD, Necib, Soyez, Thaler JHEP 1806 (2018) 093]

Jet grooming: (Recursive) Soft Drop / mMDT

- Mass peak can be partly reconstructed by removing unassociated soft wide-angle radiation (grooming).
- Recurse through clustering tree and remove soft branch if

$$\frac{\min(p_{t,1}, p_{t,2})}{p_{t,1} + p_{t,2}} < z_{\text{cut}} \left(\frac{\Delta R_{12}}{R_0}\right)^{\beta}$$

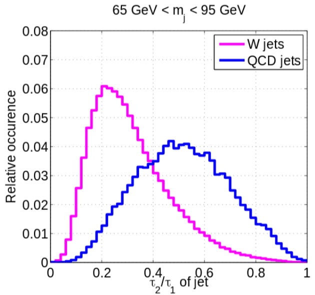
W jet

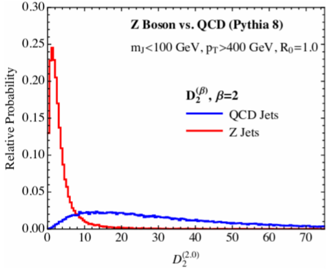


[Dasgupta, Fregoso, Marzani, Salam JHEP 1309 (2013) 029]
[Larkoski, Marzani, Soyez, Thaler JHEP 1405 (2014) 146]
[FD, Necib, Soyez, Thaler JHEP 1806 (2018) 093]

- Variety of observables have been constructed to probe the hard substructure of a jet ($V/H/t$ decay lead to jets with multiple hard cores).
- Radiation patterns of colourless objects ($W/Z/H$) differs from quark or gluon jets.
- Efficient discriminators can be obtained e.g. from ratio of N-subjettiness or energy correlation functions.

[Thaler, Van Tilburg JHEP 1103 (2011) 015]
[Larkoski, Salam, Thaler JHEP 1306 (2013) 108]
[Larkoski, Moult, Neill JHEP 1412 (2014) 009]





Applying Machine Learning in Jet Physics

Recent wave of results in applications of ML algorithms to jet physics.

Classification problems have been tackled through several orthogonal approaches

- Convolutional Neural Networks used on representation of jet as image
- Recurrent Neural Networks used on jet clustering tree.
- Linear combination or dense network applied to an observable basis (e.g. N-subjettiness ratios, energy flow polynomials)

Beyond classification problems

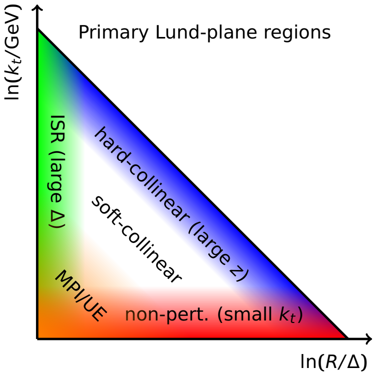
- Classification problems are one of the easiest application of ML, but by far not the only one!
- Many promising applications of ML methods for:
  - fast simulations using unsupervised generative models
    [Paganini, de Oliveira, Nachman PRL 120 (2018) 042003]
  - regression tasks such as pile-up subtraction
    [Komiske, Metodiev, Nachman, Schwartz JHEP 1712 (2017) 051]
  - anomaly detection for new physics
    [Collins, Howe, Nachman PRL 121 (2018) 241803]
  - distance metric of collider events
    [Komiske, Metodiev, Thaler arXiv:1902.02346]
  - etc ...

THE LUND PLANE

Lund diagrams

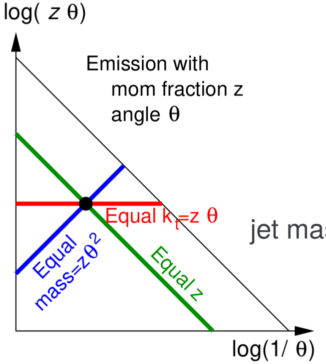
- Lund diagrams in the $(\ln z\theta, \ln\theta)$ plane are a very useful way of representing emissions.
- Different kinematic regimes are clearly separated, used to illustrate branching phase space in parton shower Monte Carlo simulations and in perturbative QCD resummations.
- Soft-collinear emissions are emitted uniformly in the Lund plane

$$dw^2 \propto \alpha_s \frac{dz}{z}\frac{d\theta}{\theta}$$

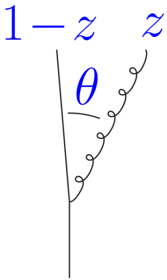


Lund diagrams

Features such as mass, angle and momentum can easily be read from a Lund diagram.

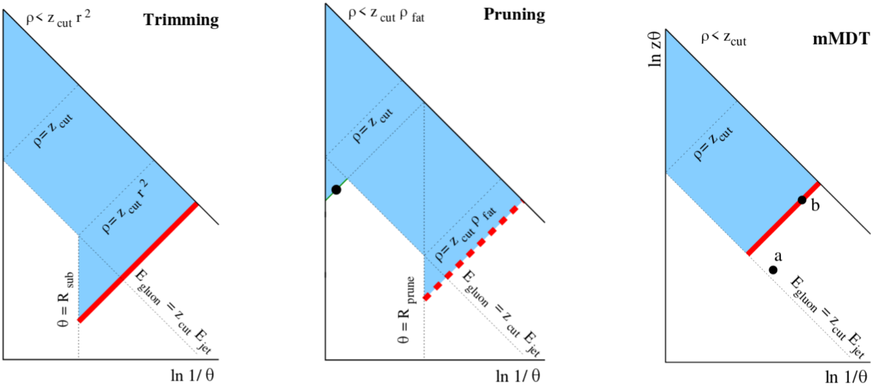


jet mass $\equiv \frac{m^2}{p_t^2 R^2} \approx z_1 \theta_1^2$



Lund diagrams for substructure

Substructure algorithms can often also be interpreted as cuts in the Lund plane.



[Dasgupta, Fregoso, Marzani, Salam JHEP 1309 (2013) 029]

Studying jets in the Lund plane

Lund diagrams can provide a useful approach to study a range of jet-related questions

- First-principle calculations of Lund-plane variables.
- Constrain MC generators, in the perturbative and non-perturbative regions.
- Brings many soft-drop related observables into a single framework.
- Impact of medium interactions in heavy-ion collisions.
- Boosted object tagging using Machine Learning methods.

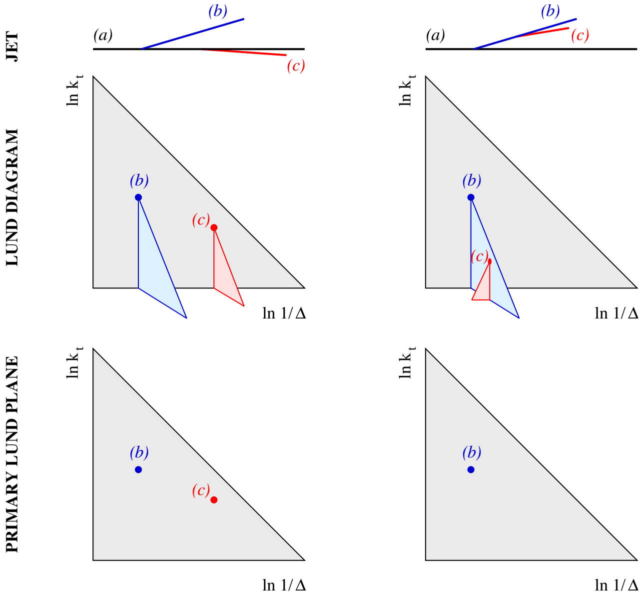
We will use this representation as a novel way to characterise radiation patterns in a jet, and study the application of recent ML tools to this picture.

Lund plane representation

To create a Lund plane representation of a jet, recluster a jet j with the Cambridge/Aachen algorithm then decluster the jet following the hardest branch.

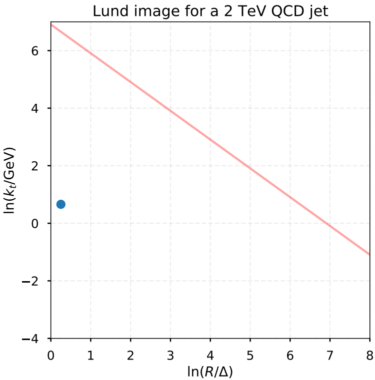
1. Undo the last clustering step, defining two subjets j_1, j_2 ordered in p_t.
2. Save the kinematics of the current declustering
$$\Delta \equiv (y_1 - y_2)^2 + (\phi_1 - \phi_2)^2, \quad k_t \equiv p_{t2}\Delta,$$
$$m^2 \equiv (p_1 + p_2)^2, \quad z \equiv \frac{p_{t2}}{p_{t1}+p_{t2}}, \quad \psi \equiv \tan^{-1}\frac{y_2-y_1}{\phi_2-\phi_1}.$$
3. Define $j = j_1$ and iterate until j is a single particle.

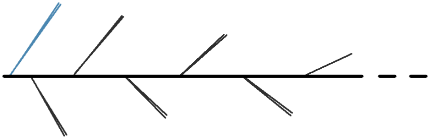
Lund plane representation



Lund representation of a jet

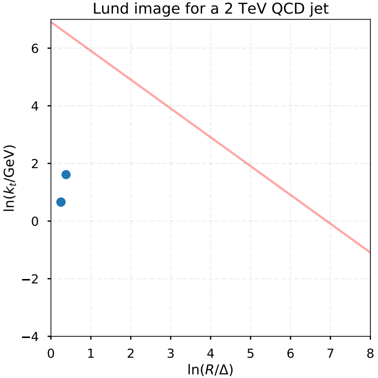
- Each jet has an image associated with its primary declustering.
- For a C/A jet, Lund plane is filled left to right as we progress through declusterings of hardest branch.
- Additional information such as azimuthal angle ψ can be attached to each point.

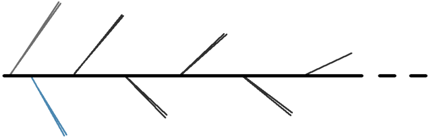




Lund representation of a jet

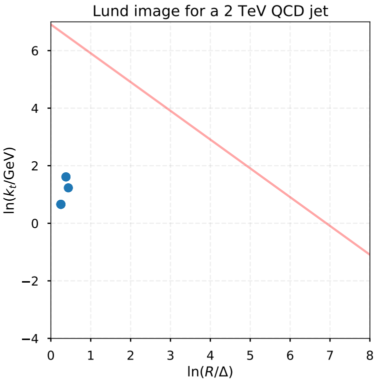
- Each jet has an image associated with its primary declustering.
- For a C/A jet, Lund plane is filled left to right as we progress through declusterings of hardest branch.
- Additional information such as azimuthal angle ψ can be attached to each point.

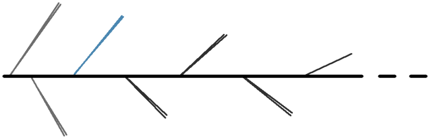




Lund representation of a jet

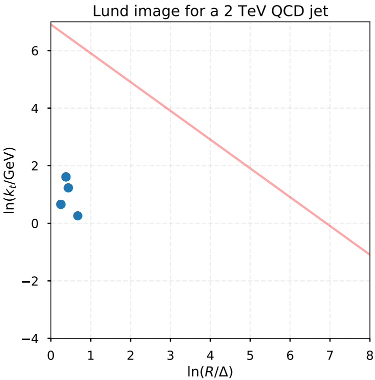
- Each jet has an image associated with its primary declustering.
- For a C/A jet, Lund plane is filled left to right as we progress through declusterings of hardest branch.
- Additional information such as azimuthal angle ψ can be attached to each point.

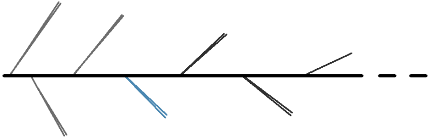




Lund representation of a jet

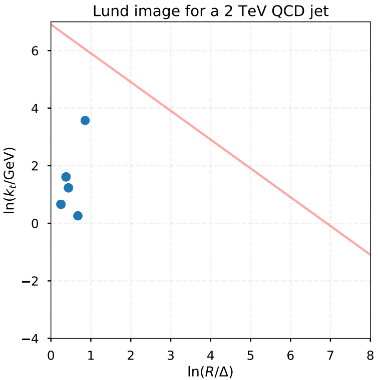
- Each jet has an image associated with its primary declustering.
- For a C/A jet, Lund plane is filled left to right as we progress through declusterings of hardest branch.
- Additional information such as azimuthal angle ψ can be attached to each point.

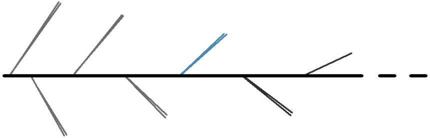




Lund representation of a jet

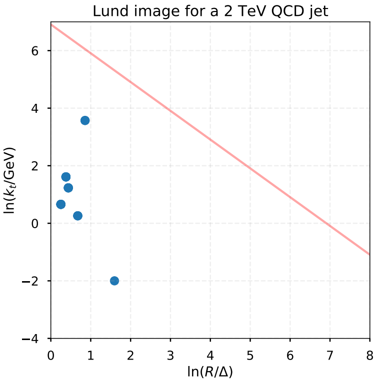
- Each jet has an image associated with its primary declustering.
- For a C/A jet, Lund plane is filled left to right as we progress through declusterings of hardest branch.
- Additional information such as azimuthal angle ψ can be attached to each point.

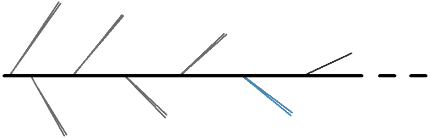




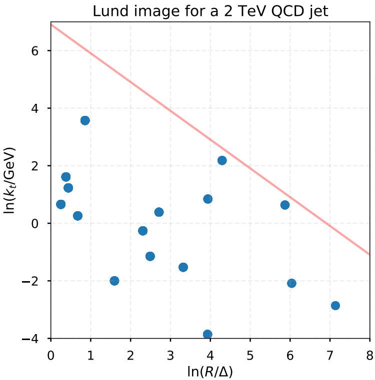
Lund representation of a jet

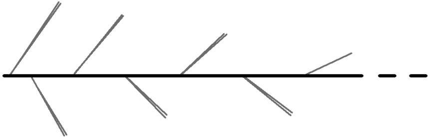
- Each jet has an image associated with its primary declustering.
- For a C/A jet, Lund plane is filled left to right as we progress through declusterings of hardest branch.
- Additional information such as azimuthal angle ψ can be attached to each point.





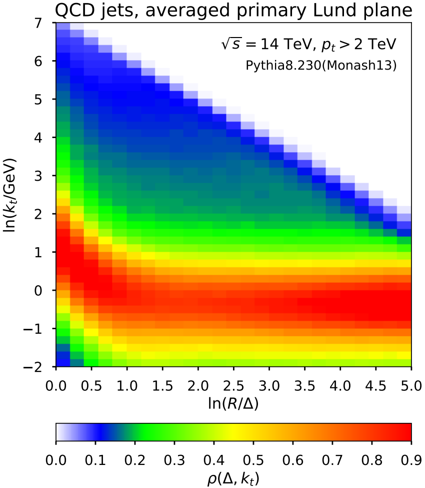
- Each jet has an image associated with its primary declustering.
- For a C/A jet, Lund plane is filled left to right as we progress through declusterings of hardest branch.
- Additional information such as azimuthal angle ψ can be attached to each point.

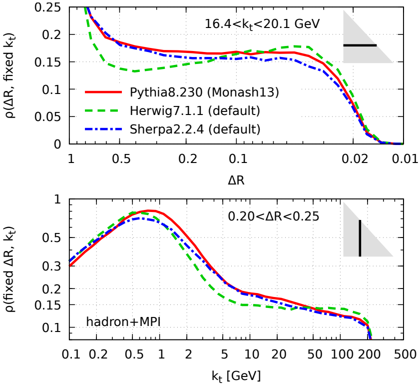




Jets as Lund images

Average over declusterings of hardest branch for 2 TeV QCD jets.

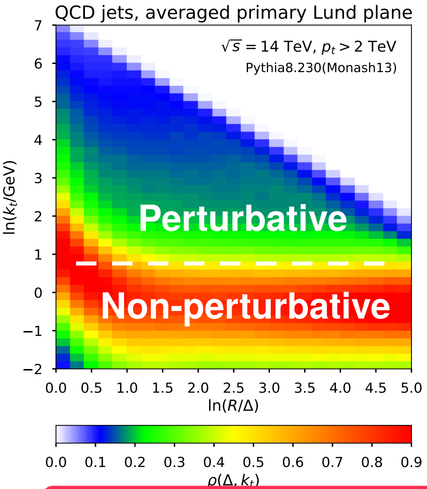


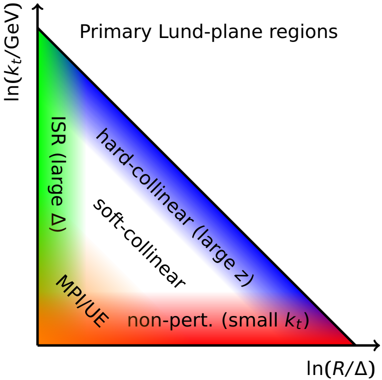


$$\rho \sim 2C\frac{\alpha_s(k_t)}{\pi}$$

Jets as Lund images

Average over declusterings of hardest branch for 2 TeV QCD jets.

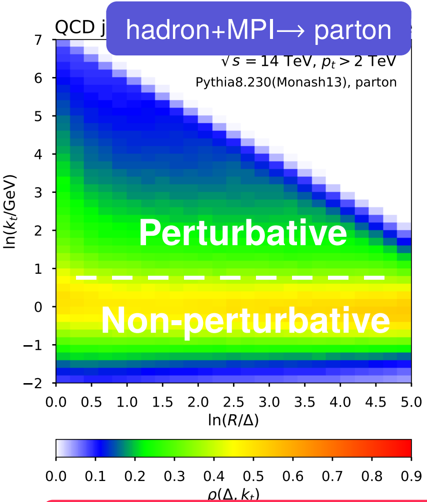


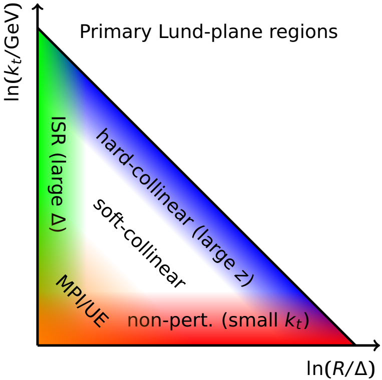


Non-perturbative region clearly separated from perturbative one.

Jets as Lund images

Average over declusterings of hardest branch for 2 TeV QCD jets.



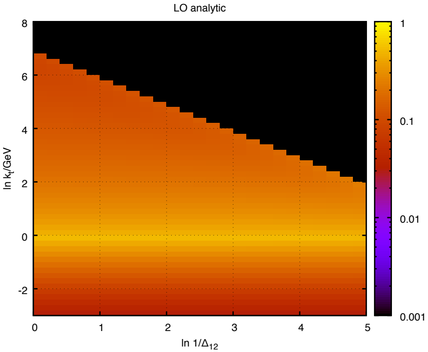


Non-perturbative region clearly separated from perturbative one.

Analytic study of the Lund plane

To leading order in perturbative QCD and for $\Delta \ll 1$, one expects for a quark initiated jet

$$\rho \simeq \frac{\alpha_s(k_t) C_F}{\pi} \bar{z} \left(p_{gq}(\bar{z}) + p_{gq}(1-\bar{z}) \right), \quad \bar{z} = \frac{k_t}{p_{t,\text{jet}} \Delta}$$

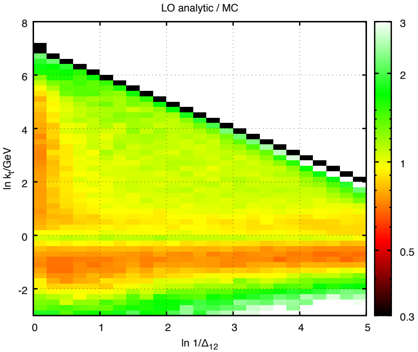


- Lund plane can be calculated analytically.
- Calculation is systematically improvable.

Analytic study of the Lund plane

To leading order in perturbative QCD and for $\Delta \ll 1$, one expects for a quark initiated jet

$$\rho \simeq \frac{\alpha_s(k_t) C_F}{\pi} \bar{z} \left(p_{gq}(\bar{z}) + p_{gq}(1-\bar{z})\right), \quad \bar{z} = \frac{k_t}{p_{t,\text{jet}} \Delta}$$



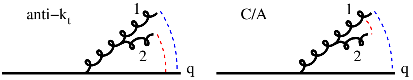
- Lund plane can be calculated analytically.
- Calculation is systematically improvable.

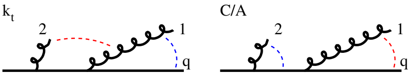
Declustering other jet-algorithm sequences

- Choice of C/A algorithm to create clustering sequence related to physical properties and associated to higher-order perturbative structures
- anti-k_t or k_t algorithms result in double logarithmic enhancements

$$\bar{\rho}_2^{(\text{anti-}k_t)}(\Delta,\kappa) \simeq +8 C_F\, C_A \ln^2 \frac{\Delta}{\kappa}$$

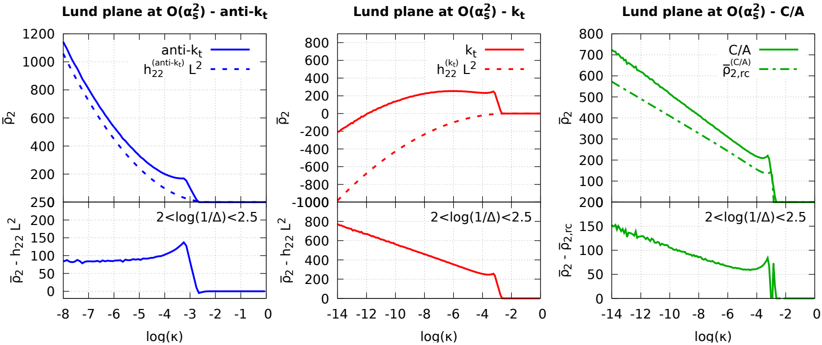
$$\bar{\rho}_2^{(k_t)}(\Delta,\kappa) \simeq -4 C_F^2 \ln^2 \frac{\Delta}{\kappa}$$



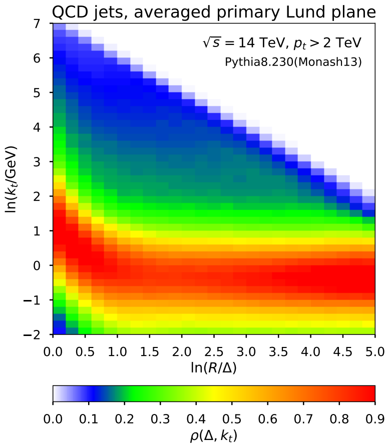


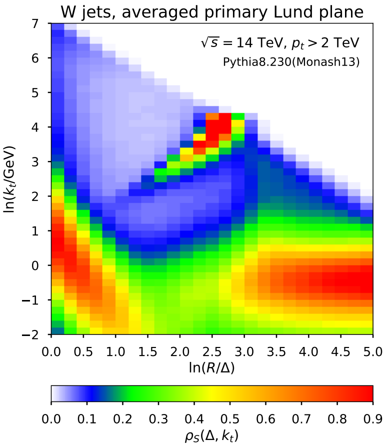
Declustering other jet-algorithm sequences

- Choice of C/A algorithm to create clustering sequence related to physical properties and associated to higher-order perturbative structures
- anti-k_t or k_t algorithms result in double logarithmic enhancements



- Hard splittings clearly visible, along the diagonal line with jet mass $m = m_W$.





APPLICATION TO BOOSTED W TAGGING

Tagging jets in the Lund Plane

We will now investigate the potential of the Lund plane for boosted-object identification.

Two different approaches:

- A log-likelihood function constructed from a leading emission and non-leading emissions in the primary plane.
- Use the Lund plane as input for a variety of Machine Learning methods.

As a concrete example, we will take dijet and WW events, looking at CA jets with $p_t > 2$ TeV.

Log-likelihood use of Lund Plane

Log-likelihood approach takes two inputs:

- First one obtained from the "leading" emission, defined as first emision satisfying $z > 0.025$ ($\sim$ mMDT tagger).

$$\mathcal{L}_{\ell}(m,z) = \ln\left(\frac{1}{N_S}\frac{dN_S}{dm\,dz} \Big/ \frac{1}{N_B}\frac{dN_B}{dm\,dz}\right)$$

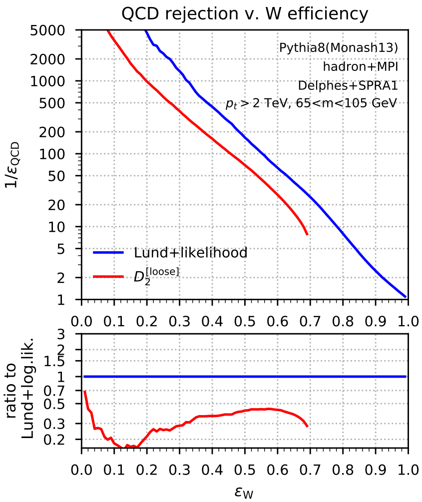
- The second one which brings sensitivity to non-leading emissions.

$$\mathcal{L}_{n\ell}(\Delta, k_t; \Delta^{(\ell)}) = \ln\left(\rho_S^{(n\ell)} \big/ \rho_B^{(n\ell)}\right)$$

Overall log-likelihood signal-background discriminator for a given jet is then given by

$$\mathcal{L}_{\text{tot}} = \mathcal{L}_{\ell}(m^{(\ell)}, z^{(\ell)}) + \sum_{i \neq \ell} \mathcal{L}_{n\ell}(\Delta^{(i)}, k_t^{(i)}; \Delta^{(\ell)}) + \mathcal{N}(\Delta^{(\ell)})$$

- Compare the LL approach in specific mass-bin with equivalent results from the Les Houches 2017 report (arXiv:1803.07977).
- Substantial improvement over best-performing substructure observable.



ML applied to the Lund plane

A variety of ML methods can be applied to the Lund plane in order to construct efficient taggers.

We will investigate three approaches:

- Convolutional Neural Networks (CNN) applied on 2D Lund images.
- Deep Neural Networks (DNN) applied on the sequence of declusterings.
- Long Short-Term Memory (LSTM) networks applied on the sequence of declusterings.

Recurrent networks with a Lund plane

- Jets generally associated with a clustering trees, where each node contains similar type of information.
- Particularly well-adapted for recurrent networks, which loop over inputs and use the same weights.
- LSTMs are a widely used variant designed to have memory over longer separations.
- For each declustering node, we consider the inputs

$$\left\{ \ln(R/\Delta R_{12}),\ \ln(k_t\,/\,\mathrm{GeV}) \right\}$$

- Inputs are IRC safe as long as there is a cutoff in transverse momentum.

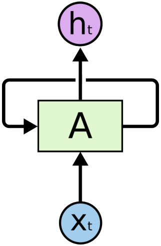
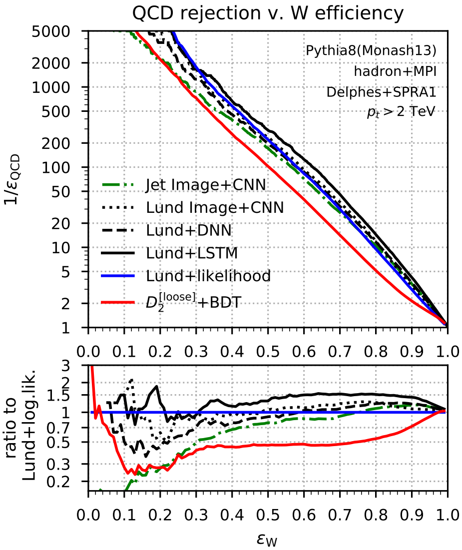


Figure from
http://colah.github.io/posts/2015-08-Understanding-LSTMs/

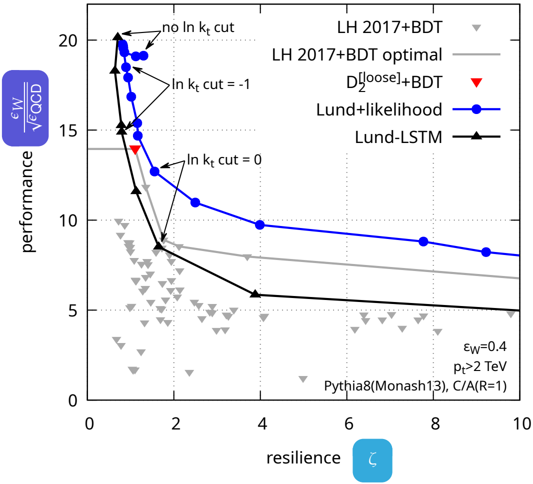
- LSTM network substantially improves on results obtained with other methods.
- Large gain in performance, particularly at higher efficiencies.



Sensitivity to non-perturbative effects

- Performance compared to resilience to MPI and hadronisation corrections.
- Vary cut on k_t, which reduces sensitivity to the non-perturbative region.

performance v. resilience [full mass information]



$\Delta\epsilon = \epsilon - \epsilon'$

$$\zeta = \left(\frac{\Delta\epsilon_S^2}{\langle\epsilon\rangle_S^2} + \frac{\Delta\epsilon_B^2}{\langle\epsilon\rangle_B^2} \right)^{-\frac{1}{2}}$$

(c.f. arXiv:1803.07977)

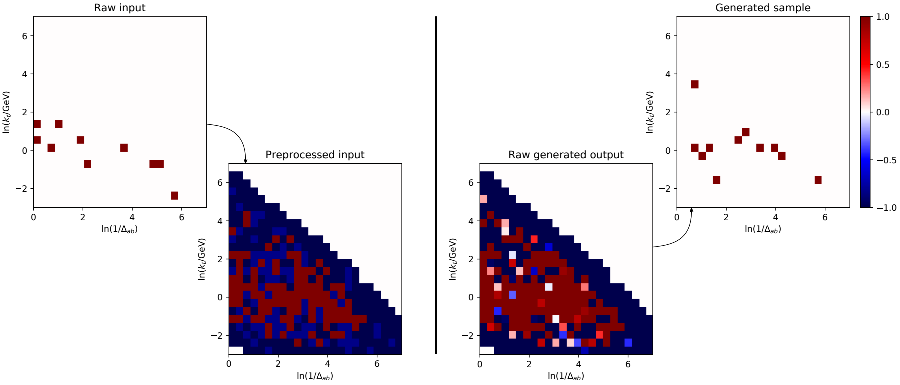
$\langle\epsilon\rangle = \frac{1}{2}(\epsilon + \epsilon')$

- Lund-likelihood performs well even at high resilience.
- ML approach reaches very good performance but is not particularly resilient to NP effects.

LUND IMAGES USING GANS

Learning to generate Lund images

- Images are combined in small batches of 32, each pixel value interpreted as the probability of being switched on.
- Preprocess images with rescaling and ZCA whitening.



Learning to generate Lund images

- Images are combined in small batches of 32, each pixel value interpreted as the probability of being switched on.
- Preprocess images with rescaling and ZCA whitening.

We consider three generative models

- Two Generative Adversarial Network architectures (LSGAN and WGANGP), constructed from generator G and discriminator D which compete against each other through a value function $V(G, D)$

$$\min_G \max_D V(D, G) = \mathbb{E}_{x\sim p_{\text{data}}}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))]\,,$$

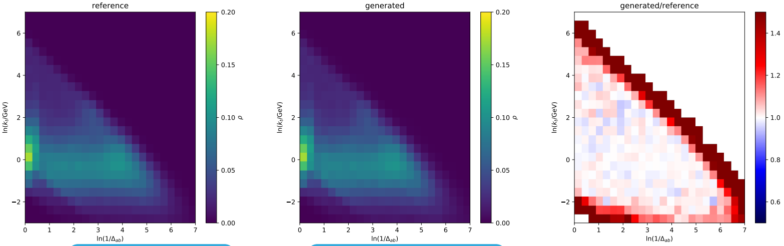
- and a latent variable VAE model, which uses a probabilistic encoder $q_\phi(z|x)$, and decoder $p_\theta(x|z)$ to map from prior $p_\theta(z)$. The algorithm learns the marginal lilelihood of the data in this generative process

$$\mathcal{L}(\theta, \phi) = \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] - \beta D_{\text{KL}}(q_\phi(z|x)||p(z))\,,$$

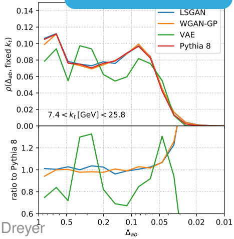
To avoid posterior collapse of VAE, we use KL annealing.

Lund images from GANs

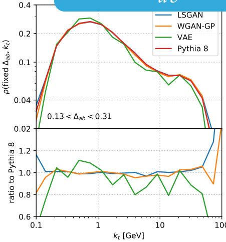
- The LSGAN provides the most stable results.
- Differences between models can be studied using slices of the Lund plane or derived observables.



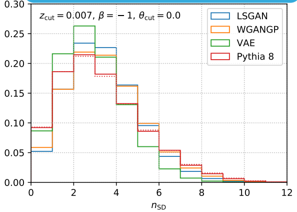
Fixed k_t slice



Fixed Δ_{ab} slice



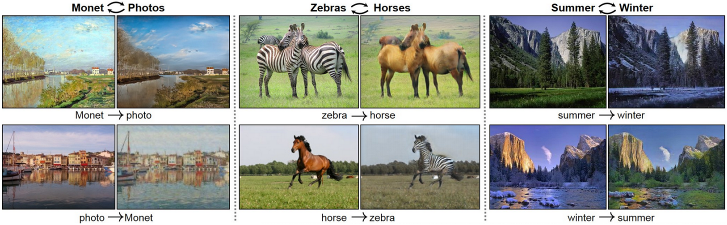
Soft Drop multiplicity



Cycle-consistent adversarial networks

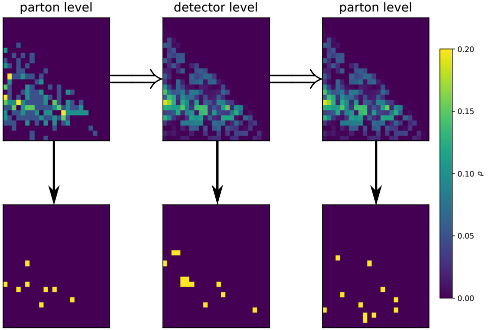
- CycleGAN learns unpaired image-to-image mapping functions $G : X \to Y$ and $F : Y \to X$ between two domains X and Y.
- Forward cycle consistency $x \in X \to G(x) \to F(G(x)) \approx x$ and backward cycle consistency $y \in Y \to F(y) \to G(F(y)) \approx y$, achieved through cycle consistency loss.
- Full objective includes also adversarial losses to both mapping functions.

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) + \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) + \lambda\, \mathcal{L}_{\text{cyc}}(G, F)\,.$$



Reinterpreting events with CycleGANs

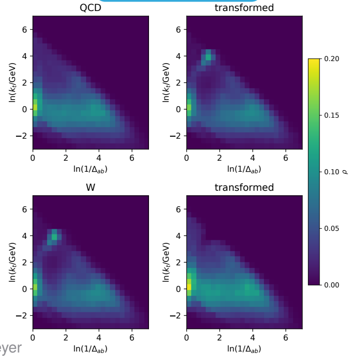
- Use CycleGAN to transform between two different domains of Lund images, e.g.
  - W jet $\leftrightarrow$ QCD jet
  - parton-level simulation $\leftrightarrow$ detector-level simulation
- Apply trained network to transform Lund images event-by-event by cycling through domains.
- Transformed events in good agreement with true sample.



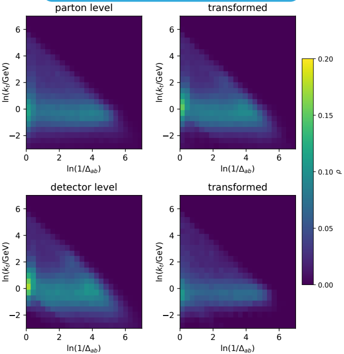
Reinterpreting events with CycleGANs

- Use CycleGAN to transform between two different domains of Lund images, e.g.
  - W jet $\leftrightarrow$ QCD jet
  - parton-level simulation $\leftrightarrow$ detector-level simulation
- Apply trained network to transform Lund images event-by-event by cycling through domains.
- Transformed events in good agreement with true sample.

$W \leftrightarrow$ QCD



parton $\leftrightarrow$ detector



REINFORCED JET GROOMING

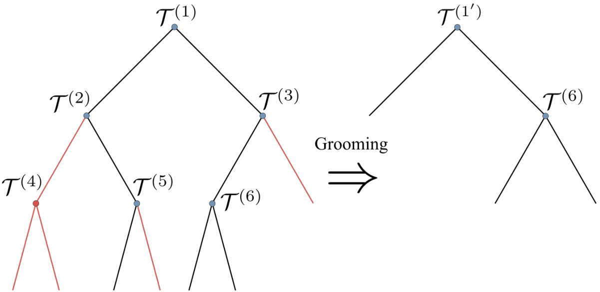
Grooming a jet tree

- Cast jet as clustering tree where state of each node $\mathcal{T}^{(i)}$ is a tuple with kinematic information on splitting

$$s_t = \{z, \Delta_{ab}, \psi, m, k_t\}$$

- Grooming algorithm defined as a function π_g observing a state and returning an action $\{0, 1\}$ on the removal of the softer branch, e.g.

$$\pi_{\text{RSD}}(s_t) = \begin{cases} 0 & \text{if} \quad z > z_{\text{cut}} \left(\frac{\Delta_{ab}}{R_0}\right)^\beta \\ 1 & \text{else} \end{cases}$$

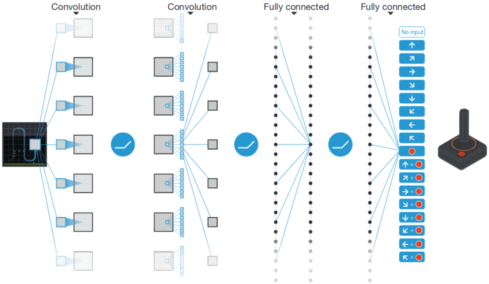


Reinforcement learning with Deep-Q-Networks

Reinforcement learning are usually built from two elements:

- an agent deciding which actions to take in order to maximize reward
- an environment, observed by the agent and affected by the action

Deep Q-Network is a RL algorithm which uses a table of Q-values $Q(s,a)$, determining the next action as the one that maximizes Q.



A neural network is used to approximate the optimal action-value function

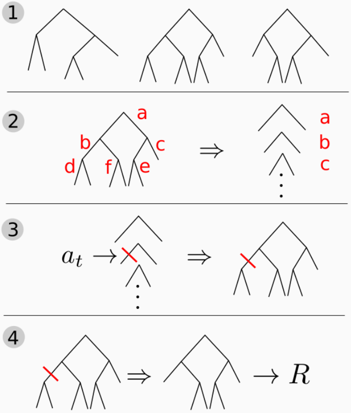
$$Q^*(s,a) = \max_{\pi} \mathbb{E}[r_t + \gamma r_{t+1} + \ldots | s_t = s, a_t = a, \pi]$$

[Mnih et al, Nature 2015]

Defining a grooming environment

To find optimal grooming policy π_g, define an environment and a reward function so that problem can be solved with RL.

1. Initialize list of all trees for training.
2. Each episode starts by randomly selecting a tree and adding its root to a priority queue (ordered in Δ_{ab}).
3. Each step removes first node from priority queue, then takes action on removal of soft branch based on s_t.
4. After action, update kinematics of parent nodes, add current children to priority queue, and evaluate reward.
5. Episode terminates once priority queue is empty.



Defining the reward function

- Key ingredient for optimization of grooming policy is reward function used at each training step.
- We construct a reward with two components
  - First piece R_M evaluated on the full jet tree, comparing the jet mass to a target value.
  - Second component R_{SD} looks at kinematics of current node.
- Total reward is then given by

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

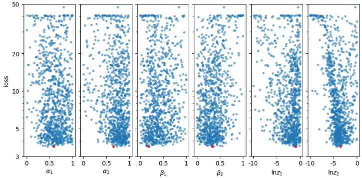
- where mass reward is defined using a Cauchy distribution

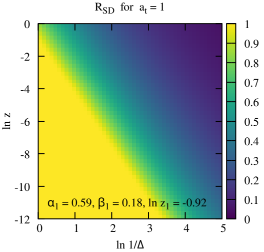
$$R_M(m) = \frac{\Gamma^2}{\pi(|m - m_{\text{target}}|^2 + \Gamma^2)}$$

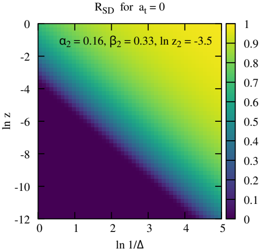
Defining the reward function

- To provide baseline behaviour for the groomer, we include a "Soft-Drop" reward R_{SD} evaluated on the current node
- Calculated on the current node state, gives positive reward for removal of wide-angle soft radiation and for keeping hard-collinear emissions.

$$R_{\text{SD}}(a_t, \Delta, z) = a_t \min\left(1, e^{-\alpha_1 \ln(1/\Delta) + \beta_1 \ln(z_1/z)}\right) \\ + (1 - a_t) \max\left(0, 1 - e^{-\alpha_2 \ln(1/\Delta) + \beta_2 \ln(z_2/z)}\right)$$

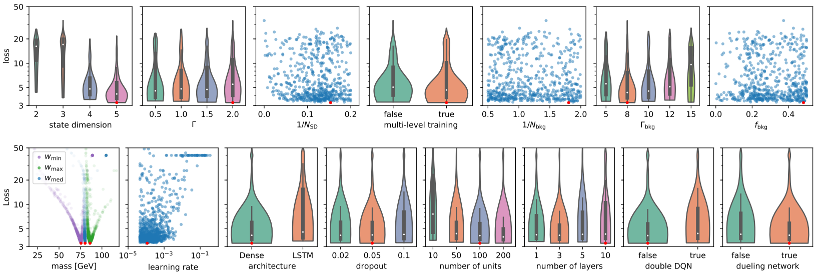






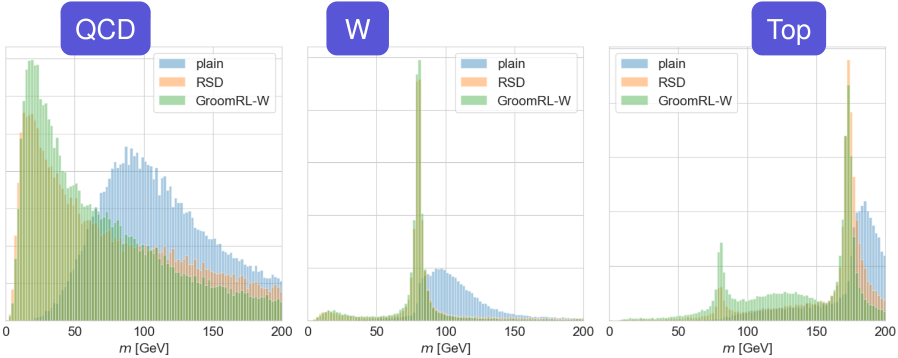
- Train RL agent with multi-level approach using both signal and bkg into account. Sample consists of 500k W/QCD or Top/QCD `Pythia` 8 jets.
- At the beginning of each episode, randomly select a signal or background jet with probability $1 - p_{\mathrm{bkg}}$.
- In the background case, mass reward function is changed to

$$R_M^{\mathrm{bkg}}(m) = \frac{m}{\Gamma_{\mathrm{bkg}}} \exp\left(- \frac{m}{\Gamma_{\mathrm{bkg}}} \right).$$



Groomed jet mass spectrum

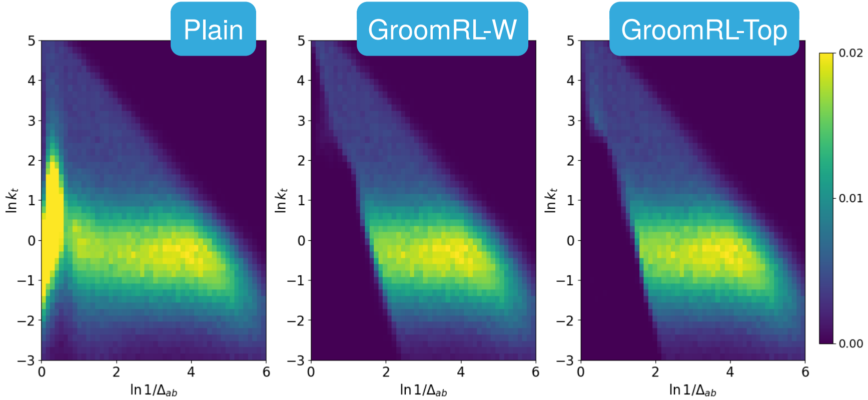
- To test the grooming algorithm derived from the DQN agent, we apply our groomer to three test samples: QCD, W and Top jets.
- Improvement in jet mass resolution compared to RSD.
- Algorithm performs well on data beyond its training range.



code available at **github.com/JetsGame/GroomRL**

Groomed jet mass spectrum

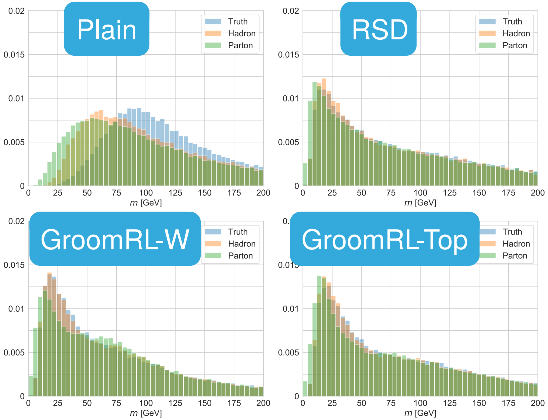
- To test the grooming algorithm derived from the DQN agent, we apply our groomer to three test samples: QCD, W and Top jets.
- Improvement in jet mass resolution compared to RSD.
- Algorithm performs well on data beyond its training range.



code available at **github.com/JetsGame/GroomRL**

Robustness to non-perturbative effects

- Resilience to hadronisation and underlying event corrections is a key feature of modern grooming algorithms
- Strategy derived from reinforcement learning shows similar behaviour to heuristic method
- No parton or hadron-level data was used in the training!



CONCLUSIONS

Conclusions

- Discussed a new way to study and exploit radiation patterns in a jet using the Lund plane.
- Lund kinematics can be used as inputs for W tagging with a range of methods:
  - Log-likelihood function.
  - Convolutional neural networks.
  - Recurrent and dense neural networks.

  Simple LL approach can match performance obtained with recent ML methods.
- Provides a framework for promising application of generative models and reinforcement learning.
- While ML can achieve high performance, one needs to mindful of resilience to poorly modeled contributions and systematic uncertainties.