The Lund Jet Plane

Theoretical Particle Physics Seminar, University of Sussex, 26 November 2018

Frédéric Dreyer

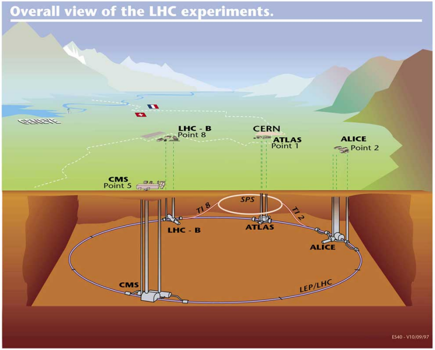
UNIVERSITY OF OXFORD

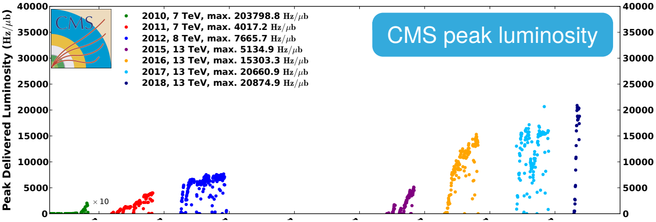
based on arXiv:1807.04758

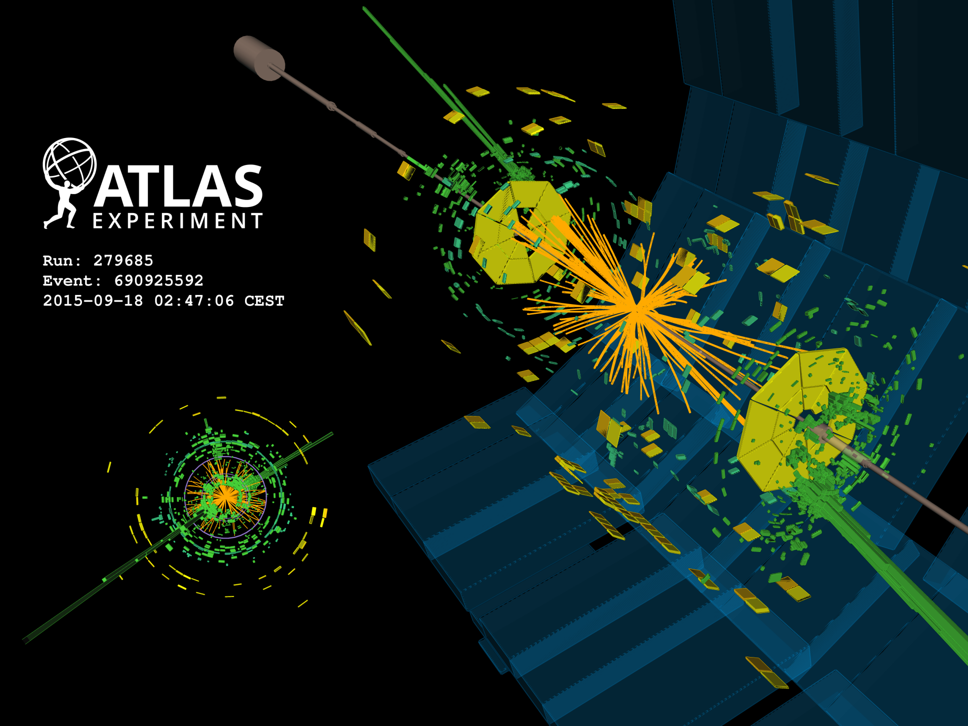
with Gavin Salam & Gregory Soyez

Physics at the high energy frontier

- LHC now colliding protons at 13 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

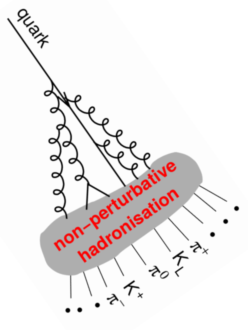
JET SUBSTRUCTURE AND MACHINE LEARNING

Jets as proxies for partons

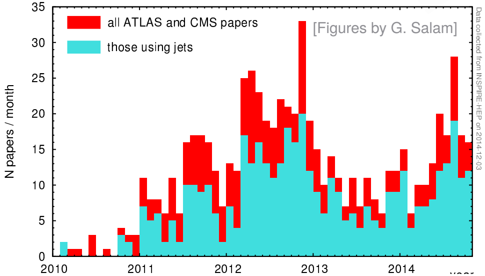
Because of color confinement, quarks and gluons shower and hadronise immediately into collimated bunches of particles.

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 and choice of jet radius

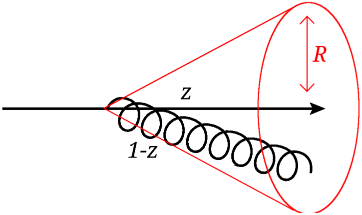
A jet algorithm maps final state particle momenta to jet momenta.

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

This requires an external parameter, the jet radius R (typically $R \sim 0.4$), specifying up to which angle separate partons are recombined into a single jet.

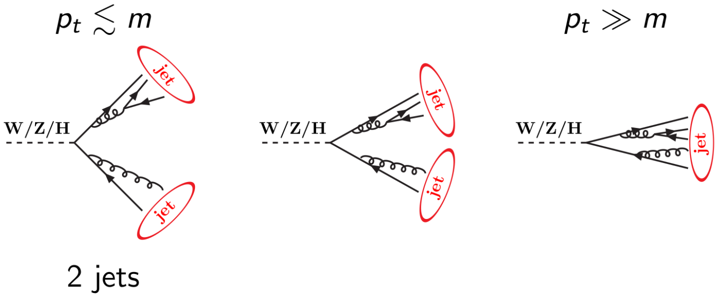
Basic idea of jet algorithm is to invert QCD branching process, clustering pairs which are closest in metric defined by the divergence structure of the theory.

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



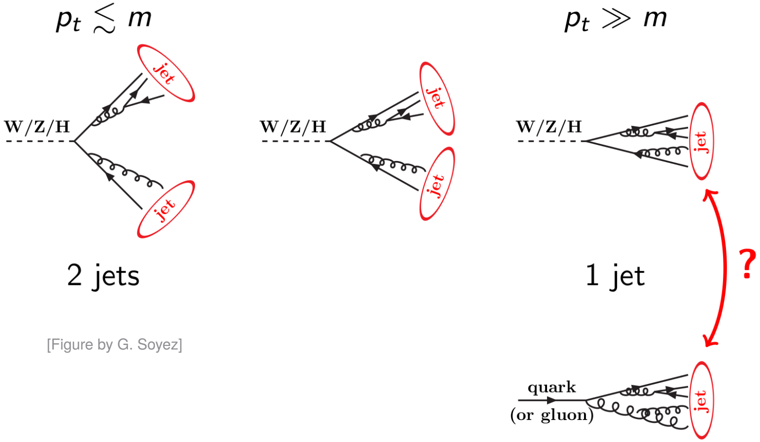
Boosted objects at the LHC

- 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.



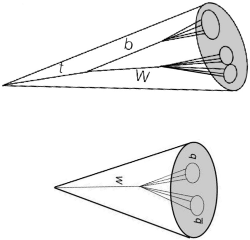
[Figure by G. Soyez]

- 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.



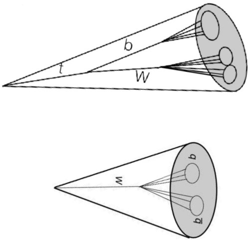
Boosted objects at the LHC

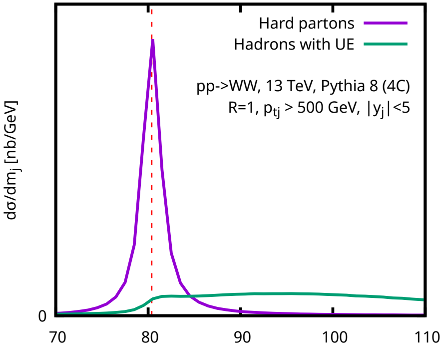
- Many techniques developed to identify hard structure of a jet based on radiation patterns.
- In principle, simplest way to identify these boosted objects is by looking at the mass of the jet.



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 identify 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.

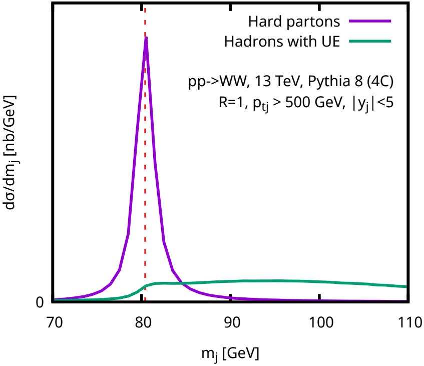
Aim of this talk: present a method bridging some of the gap between these two techniques.

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

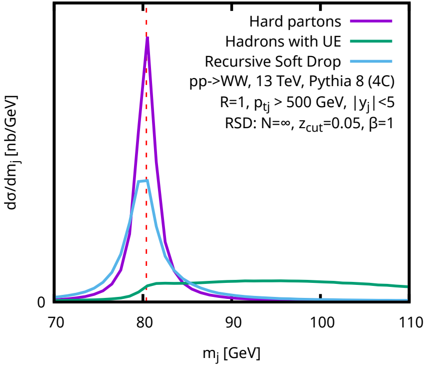


[Dasgupta, Fregoso, Marzani, Salam JHEP 1309 (2013) 029]
[Larkoski, Marzani, Soyez, Thaler JHEP 1405 (2014) 146]
[FD, Necib, Soyez, Thaler arXiv:1804.03657]

- 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

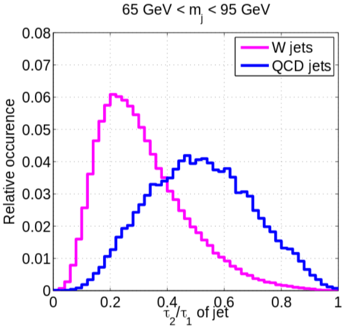


[Dasgupta, Fregoso, Marzani, Salam JHEP 1309 (2013) 029]
[Larkoski, Marzani, Soyez, Thaler JHEP 1405 (2014) 146]
[FD, Necib, Soyez, Thaler arXiv:1804.03657]

Substructure observables

- 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.

$$\tau_{21}^{(\beta)} = \frac{\tau_2^{(\beta)}}{\tau_1^{(\beta)}}$$

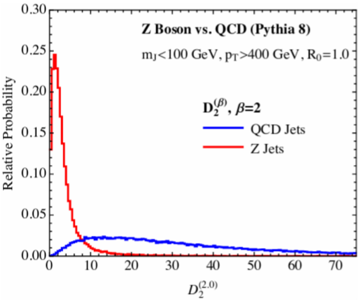


[Thaler, Van Tilburg JHEP 1103 (2011) 015]
[Larkoski, Salam, Thaler JHEP 1306 (2013) 108]
[Larkoski, Moult, Neill JHEP 1412 (2014) 009]

Substructure observables

- 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.

$$D_2^{(\beta)} = \frac{e_3^{(\beta)}}{\left(e_2^{(\beta)}\right)^3}$$



[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

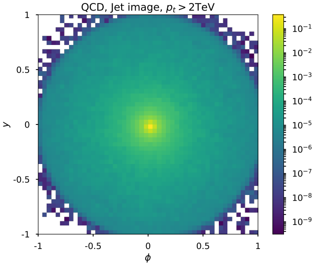
Recent wave of results in applications of ML algorithms to jet physics.

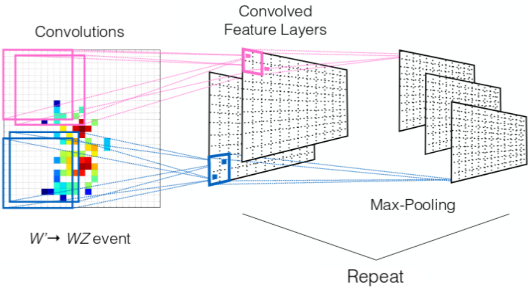
Two approaches seem particularly promising

- Convolutional Neural Networks used on representation of jet as image
- Recurrent Neural Networks used on jet clustering tree.

Convolutational Neural Networks and Jet Images

- Project a jet onto a fixed $n \times n$ pixel image in rapidity-azimuth, where each pixel intensity corresponds to the momentum of particles in that cell.
- Can be used as input for classification methods used in computer vision, such as deep convolutional neural networks.



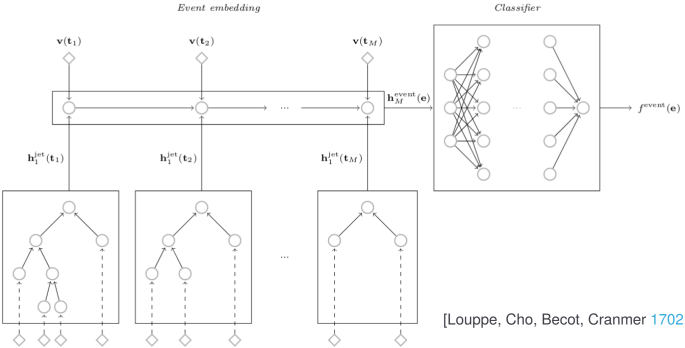


[Cogan, Kagan, Strauss, Schwartzman JHEP 1502 (2015) 118]

[de Oliveira, Kagan, Mackey, Nachman, Schwartzman JHEP 1607 (2016) 069]

Recurrent Neural Networks and clustering trees

- Train a recurrent neural network on successive declusterings of a jet.
- Techniques inspired from Natural Language Processing with powerful applications in handwriting and speech recognition.



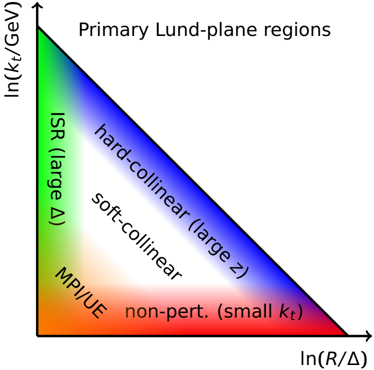
[Louppe, Cho, Becot, Cranmer 1702.00748]

THE LUND PLANE

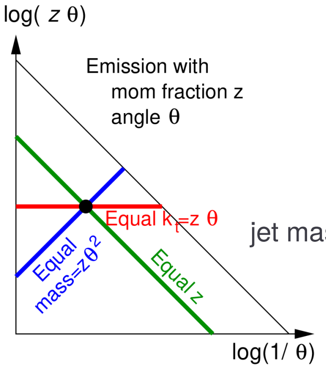
Lund diagrams

- 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

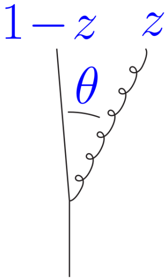
$$dw^2 \propto \alpha_s \frac{dz}{z}\frac{d\theta}{\theta}$$



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

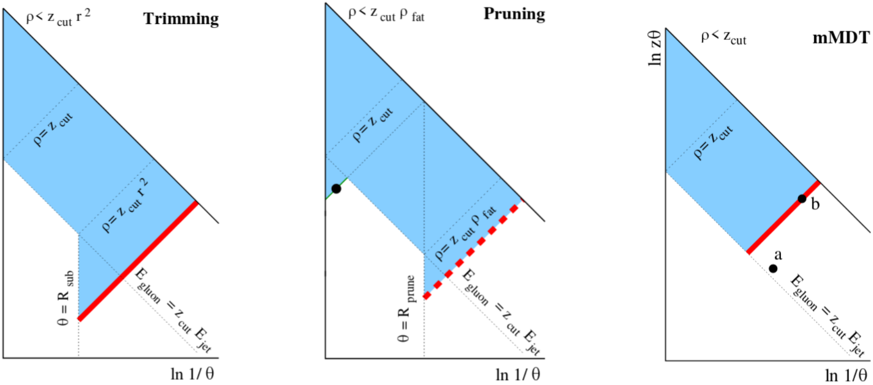


$$\text{jet mass} \equiv \frac{m^2}{p_t^2 R^2} \approx z_1 \theta_1^2$$



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.

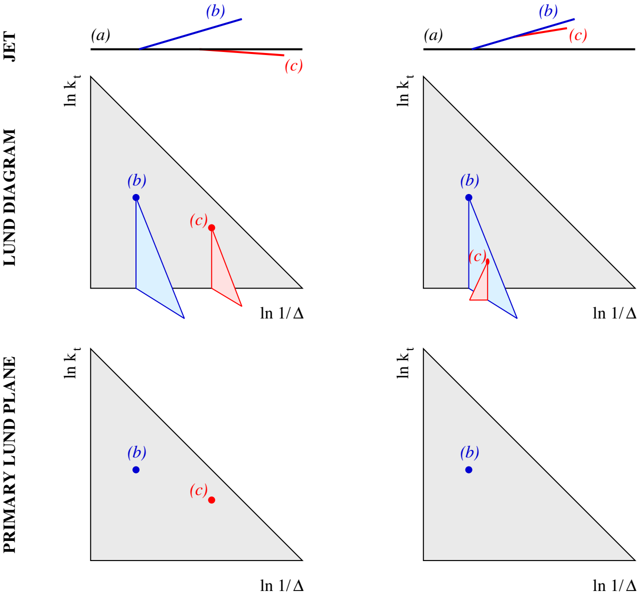
We will use this representation as a novel way to characterise radiation patterns in a jet, and study the application of recent ML tools to this picture.

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.

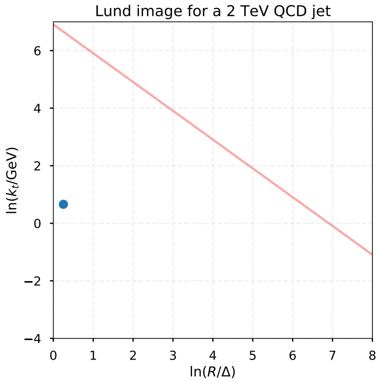
1. Undo the last clustering step, defining two subjets j_1, j_2 ordered in p_t.
2. Save the kinematics of the current declustering
$$\Delta \equiv (y_1 - y_2)^2 + (\phi_1 - \phi_2)^2, \quad k_t \equiv p_{t2}\Delta,$$
$$m^2 \equiv (p_1 + p_2)^2, \quad z \equiv \frac{p_{t2}}{p_{t1}+p_{t2}}, \quad \psi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

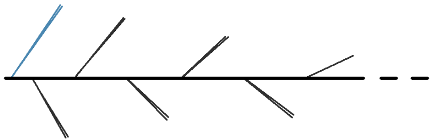
Lund plane representation



Lund representation of a jet

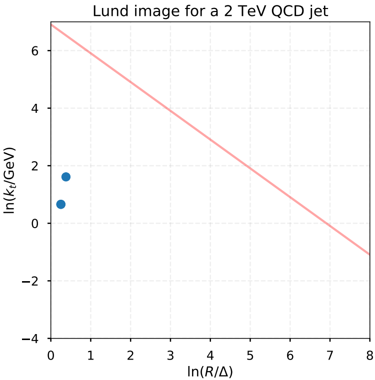
- Each jet has an image associated with its primary declustering.
- For a C/A jet, Lund plane is filled left to right as we progress through declusterings of hardest branch.
- Additional information such as azimuthal angle ψ can be attached to each point.

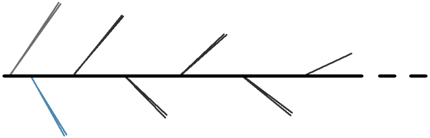




Lund representation of a jet

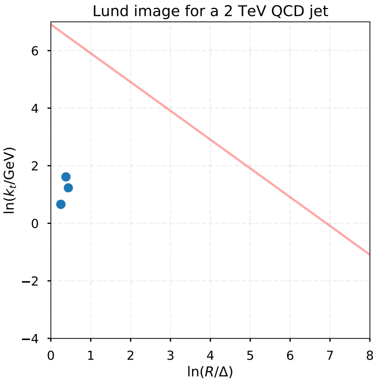
- Each jet has an image associated with its primary declustering.
- For a C/A jet, Lund plane is filled left to right as we progress through declusterings of hardest branch.
- Additional information such as azimuthal angle ψ can be attached to each point.

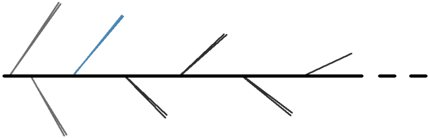




Lund representation of a jet

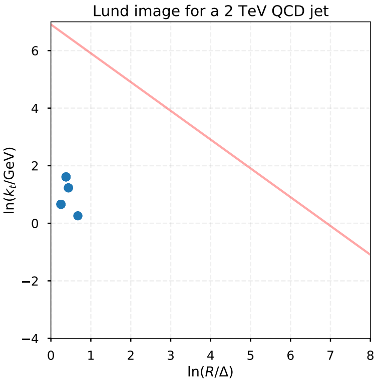
- Each jet has an image associated with its primary declustering.
- For a C/A jet, Lund plane is filled left to right as we progress through declusterings of hardest branch.
- Additional information such as azimuthal angle ψ can be attached to each point.

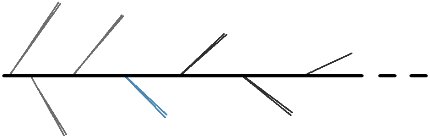




Lund representation of a jet

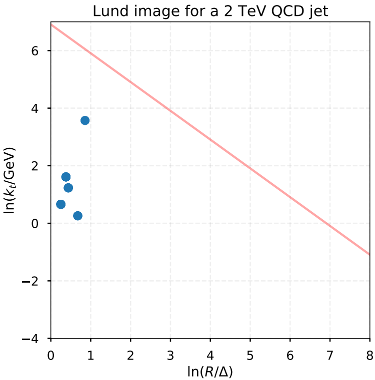
- Each jet has an image associated with its primary declustering.
- For a C/A jet, Lund plane is filled left to right as we progress through declusterings of hardest branch.
- Additional information such as azimuthal angle ψ can be attached to each point.

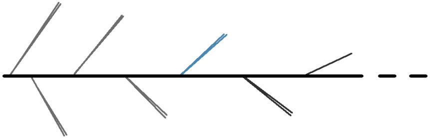




Lund representation of a jet

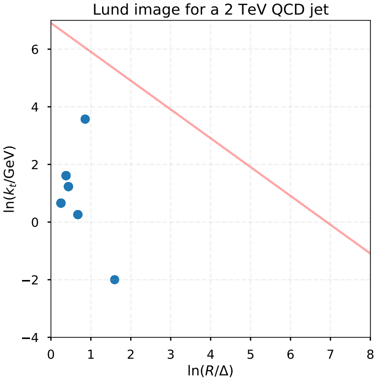
- Each jet has an image associated with its primary declustering.
- For a C/A jet, Lund plane is filled left to right as we progress through declusterings of hardest branch.
- Additional information such as azimuthal angle ψ can be attached to each point.

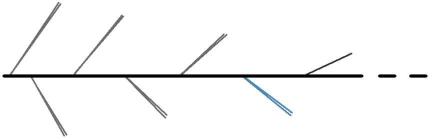




Lund representation of a jet

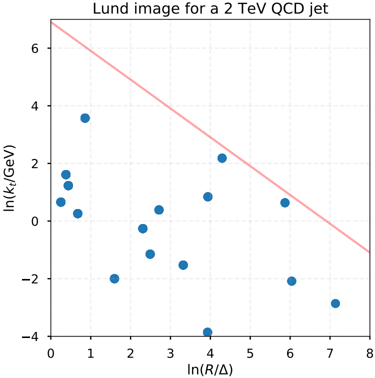
- Each jet has an image associated with its primary declustering.
- For a C/A jet, Lund plane is filled left to right as we progress through declusterings of hardest branch.
- Additional information such as azimuthal angle ψ can be attached to each point.

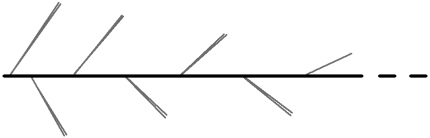




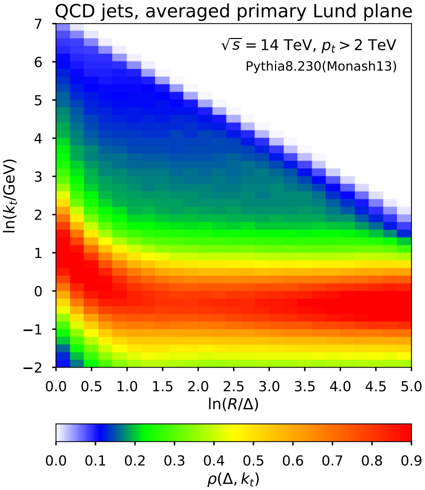
Lund representation of a jet

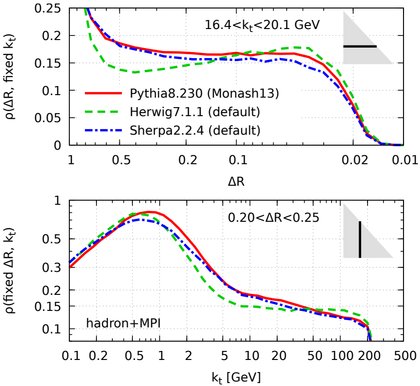
- Each jet has an image associated with its primary declustering.
- For a C/A jet, Lund plane is filled left to right as we progress through declusterings of hardest branch.
- Additional information such as azimuthal angle ψ can be attached to each point.





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

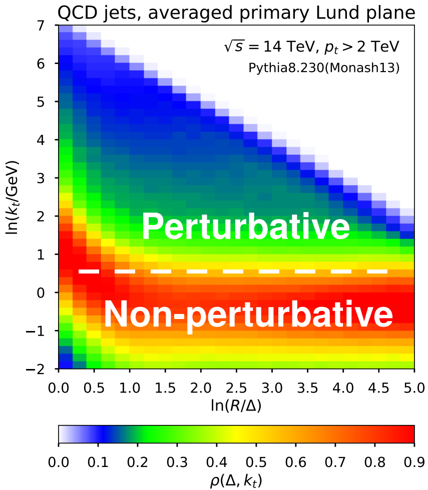


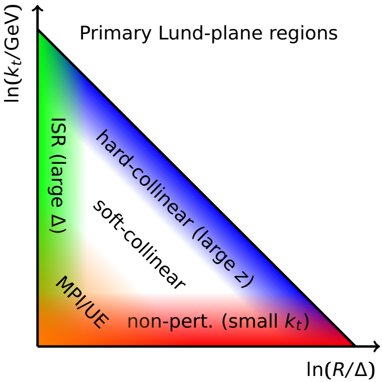


$$\rho \sim 2C \frac{\alpha_s(k_t)}{\pi}$$

Jets as Lund images

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



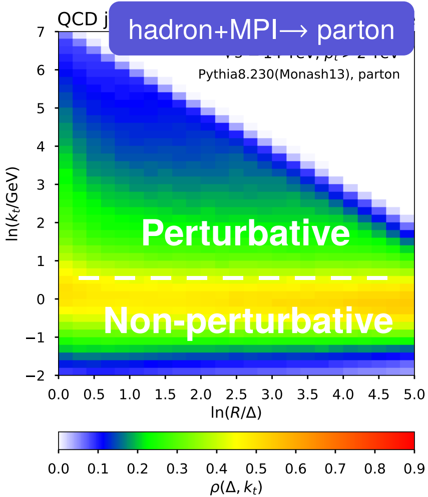


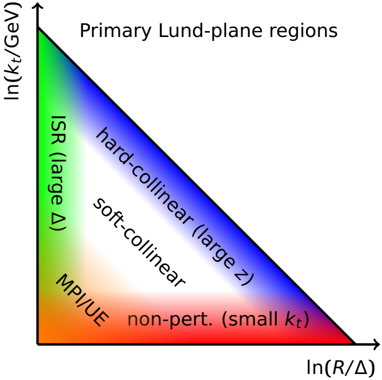
Non-perturbative region clearly separated from perturbative one.

Jets as Lund images

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

hadron+MPI⟶ parton



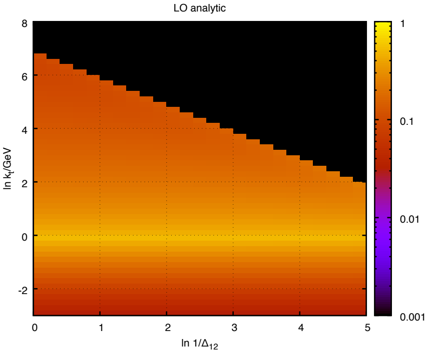


Non-perturbative region clearly separated from perturbative one.

Analytic study of the Lund plane

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

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

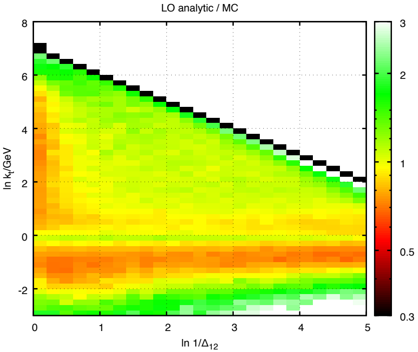


- Lund plane can be calculated analytically.
- Calculation is systematically improvable.

Analytic study of the Lund plane

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

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



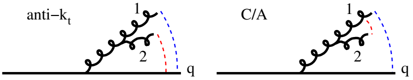
- Lund plane can be calculated analytically.
- Calculation is systematically improvable.

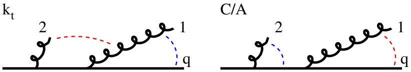
Declustering other jet-algorithm sequences

- 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

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

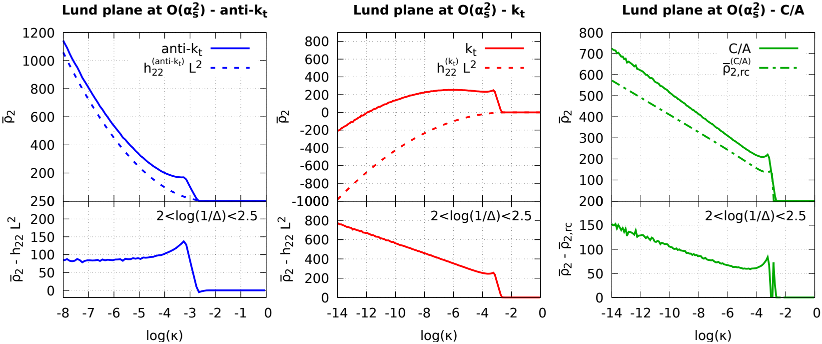
$$\bar{\rho}_2^{(k_t)}(\Delta,\kappa) \simeq -4 C_F^2 \ln^2 \frac{\Delta}{\kappa}$$



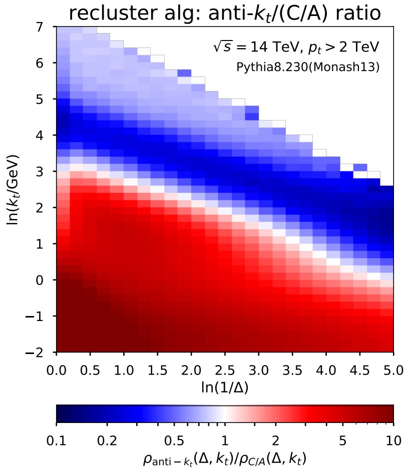


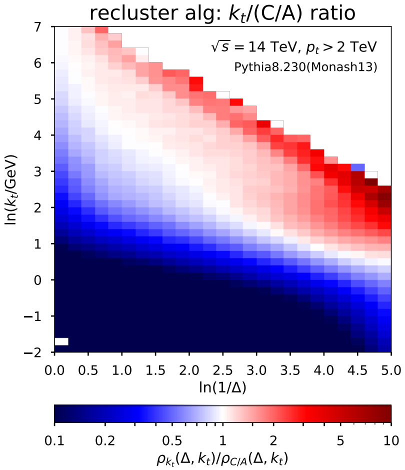
Declustering other jet-algorithm sequences

- 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



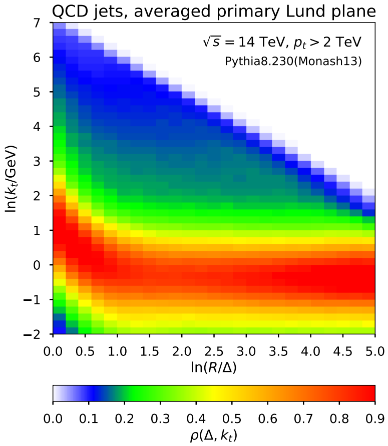
Declustering other jet-algorithm sequences





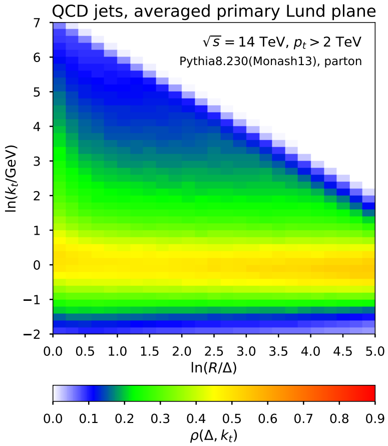
Lund images for QCD and W jets

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



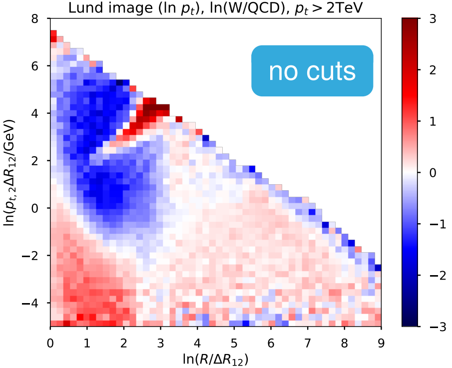
Lund images for QCD and W jets

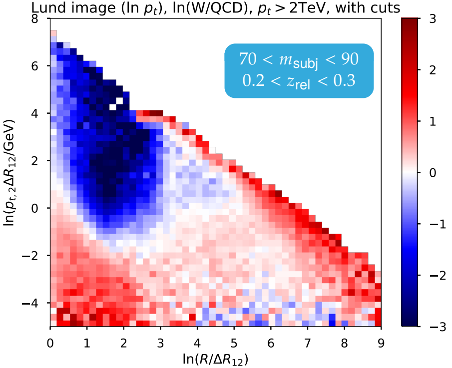
- Non-perturbative contributions affect specific parts of the image.
- Main differences due to presence of wide-angle UE emissions and soft particles.



Discriminating features in the Lund plane

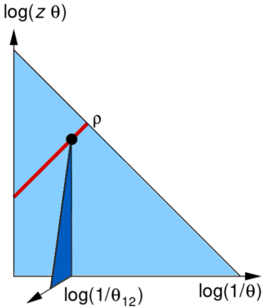
- Can identify discriminating features by considering log ratio of averaged images.
- W peak is clearly visible – but after cuts, depletion of emissions at relatively large angles remains distinctive signature.

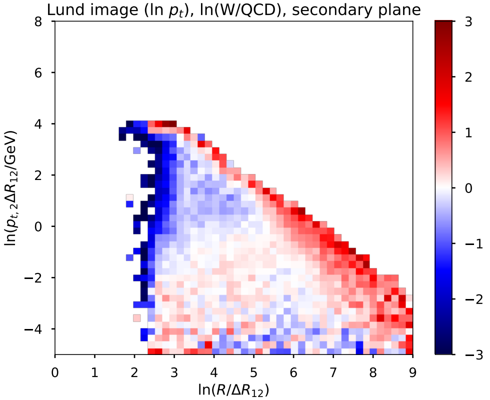




Secondary Lund plane

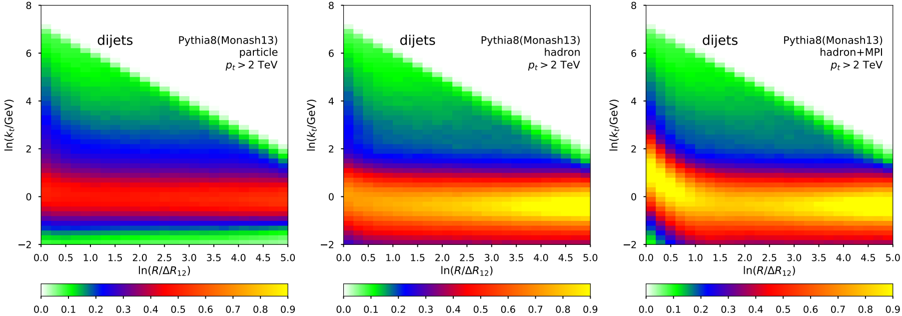
- Secondary Lund planes are ignored: some information is therefore lost, but still achieves good performance.
- This limitation can be overcome by extending the methods we will discuss to include secondary planes as inputs.





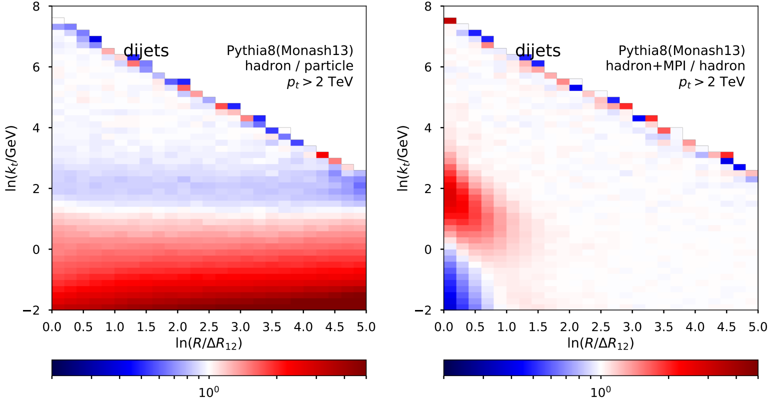
Non-perturbative effects

- Hadronisation corrections appear at the bottom of the Lund plane, below $\ln k_t \sim 0.5$.
- Underlying event leads to changes in the large angles region.



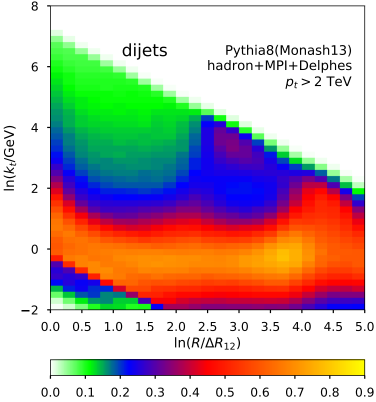
Non-perturbative effects

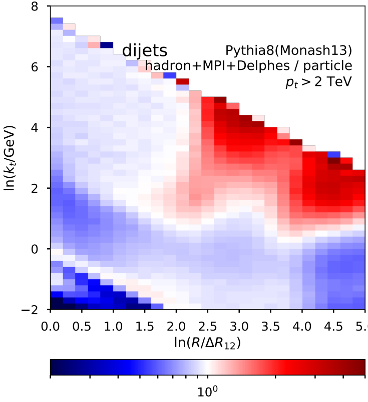
- Hadronisation corrections appear at the bottom of the Lund plane, below $\ln k_t \sim 0.5$.
- Underlying event leads to changes in the large angles region.



Detector effects

- Detector effects have significant impact on the Lund plane at angular scales below the hadronic calorimeter spacing.
- Two enhanced regions corresponding to resolution scale of HCal and ECal.





Subjet-Particle Rescaling Algorithm (SPRA)

Mitigate impact of detector granularity using a subjet particle rescaling algorithm:

- Recluster Delphes particle-flow objects into subjets using C/A with $R_h = 0.12$.
- Taking each subjet in turn, scale each PF charged-particle ($h^\pm$) and photon (γ) candidate that it contains by a factor f_1

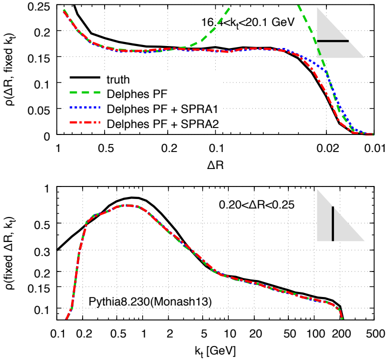
$$f_1 = \frac{\sum_{i \in \text{subjet}} p_{t,i}}{\sum_{i \in \text{subjet}(h^\pm,\gamma)} p_{t,i}},$$

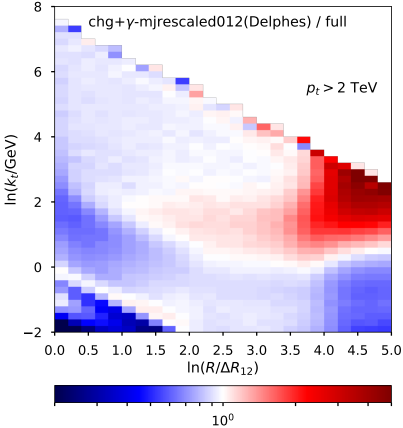
  and discard the other neutral hadron candidates.
- If subjet doesn't contain photon or charged-particle candidates, retain all of the subjet's particles with their original momenta.

Recluster the full set of resulting particles (from all subjets) into a single large jet and use it to evaluate the mass and Lund plane.

Subjet-Particle Rescaling Algorithm (SPRA)

Mitigate impact of detector granularity using a subjet particle rescaling algorithm:





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: leading emission

Log-likelihood approach takes two inputs:

- First one obtained from the "leading" emission.
- The second one which brings sensitivity to non-leading emissions.

Leading emission is determined to be the first emission in the Lund declustering sequence that satisfies $z > 0.025$ ($\sim$ mMDT tagger)

Define a $\mathcal{L}_\ell$ log likelihood function

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

where the ratio of $\frac{dN_{S/B}}{dm\,dz}$ is the differential distribution in m and z of the leading emission for signal sample (background) with $N_S(N_B)$ jets.

Log-likelihood use of Lund Plane: non-leading emissions

Non-leading $(n\ell)$ emissions within the primary Lund plane are incorporated using a function

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

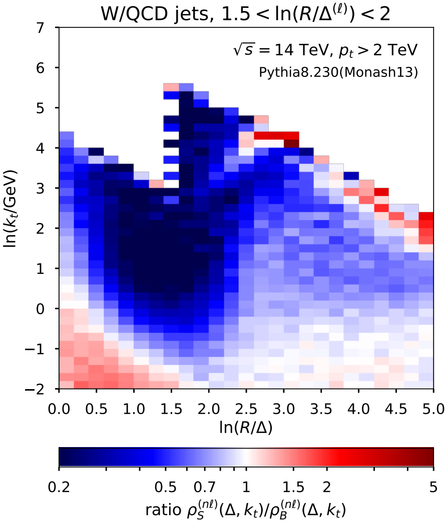
where $\rho^{(n\ell)}$ is determined just over the non-leading emissions,

$$\rho^{(n\ell)}(\Delta, k_t; \Delta^{(\ell)}) = \frac{dn^{(n\ell)}_{\text{emission}}}{d\ln k_t \, d\ln 1/\Delta \, d\Delta^{(\ell)}} \bigg/ \frac{dN_{\text{jet}}}{d\Delta^{(\ell)}}$$

as a function of the angle $\Delta^{(\ell)}$ of the leading emission.

Log-likelihood use of Lund Plane: non-leading emissions

$\mathcal{L}_{n\ell}$ log-likelihood function in a specific bin.



Log-likelihood use of Lund Plane: full discriminator

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

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

where $\mathcal{N} = -\int d\ln\Delta\, d\ln k_t \left(\rho_S^{(\ell)} - \rho_B^{(\ell)}\right)$.

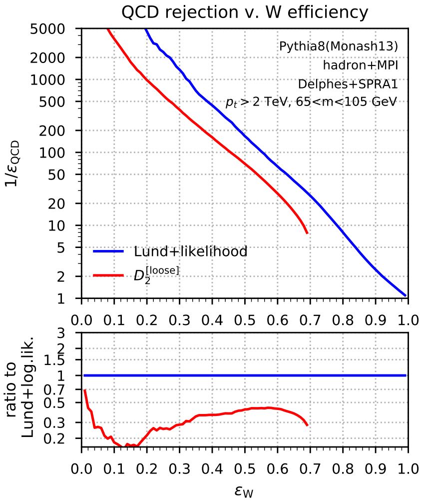
Each subjet i in the sum brings information about whether it is in a more background-like or signal-like part of the Lund plane.

Optimal discriminator if:

- Leading emission correctly associated with W's two-prong structure.
- Non-leading emissions are independent from each other.
- Emission patterns for those emissions depend only on $\Delta^{(\ell)}$.

Tagging with LL method

- 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

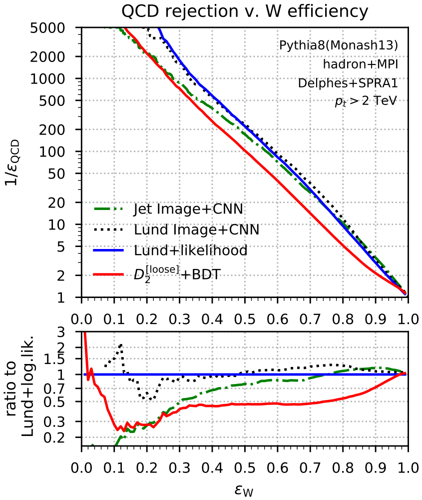
A variety of ML methods can be applied to the Lund plane in order to construct efficient taggers.

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.

CNN on Lund images for W tagging

- Lund images perform particularly well at high transverse momentum, were $W \to qq$ is most separated on Lund plane.
- Performance on par with LL method, better than regular jet images at low efficiencies.

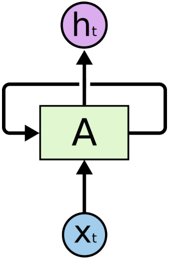


Recurrent networks with a Lund plane

- 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.
- For each declustering node, we consider the inputs

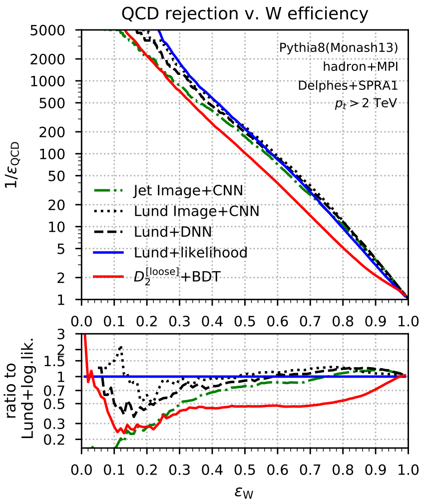
$$\left\{ \ln(R/\Delta R_{12}),\, \ln(k_t/\,\mathrm{GeV}) \right\}$$

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



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

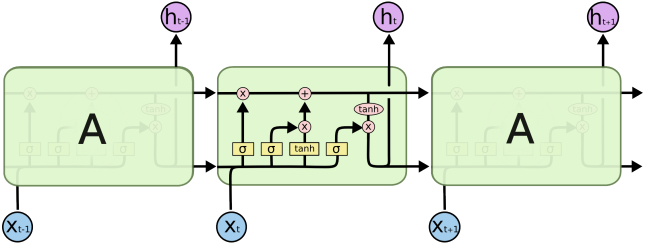
- Applying DNN directly to the sequence of declusterings of the hardest branch.
- Results very similar to previous CNN approach.



Long short-term memory networks

- Simple recurrent networks unable to handle dependencies that are widely separated in the data.
- LSTM networks designed to have memory over longer periods, by adding four layers for each module and including a no-activation function.

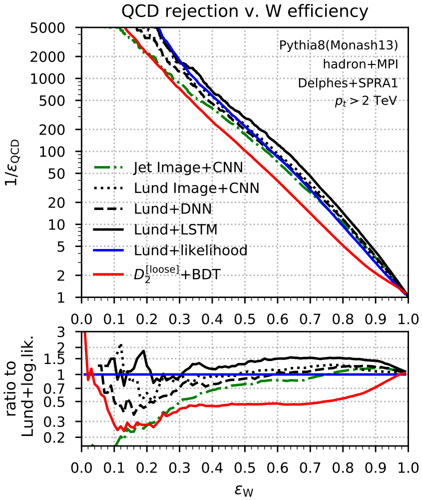
[Hochreiter, Schmidhuber (1997)]



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

LSTMs for jet tagging

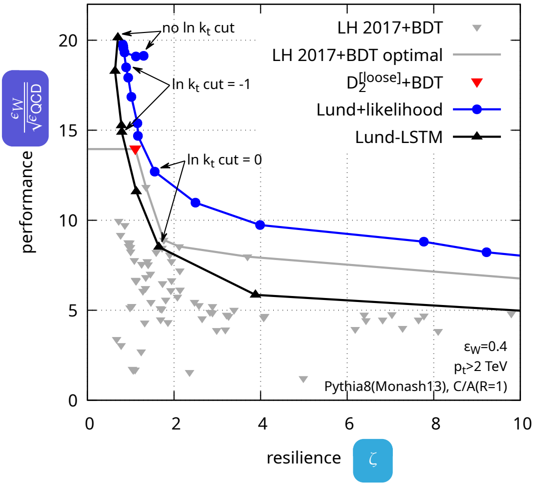
- LSTM network substantially improves on results obtained with other methods.
- Large gain in performance, particularly at higher efficiencies.



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]



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

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

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

(c.f. arXiv:1803.07977)

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

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.
- While ML can achieve high performance, one needs to mindful of resilience to poorly modeled contributions and systematic uncertainties.

Wide range of experimental and theoretical opportunities brought by studying Lund diagrams for jets. A **rich topic** for **further exploration**.