Jet tagging in the Lund plane

BOOST 2021, Online, 2 August 2021

Frédéric Dreyer

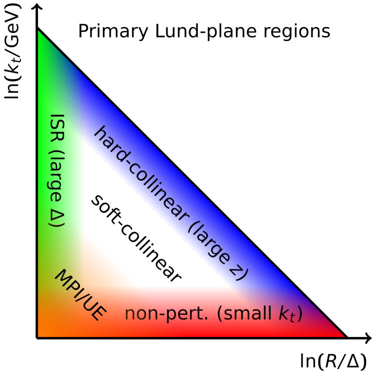
UNIVERSITY OF OXFORD

JHEP 03 (2021) 052 with Huilin Qu and work in progress with Gregory Soyez & Adam Takacs

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

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



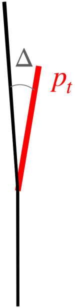
[Andersson et al, Z.Phys. C43 (1989) 625]

[FD, Salam, Soyez, JHEP 1812 (2018) 064]

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

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



Δ *opening angle of a splitting*

$$k_t = p_t \Delta$$

p_t *(or* $p_\perp$*) is transverse momentum wrt beam*

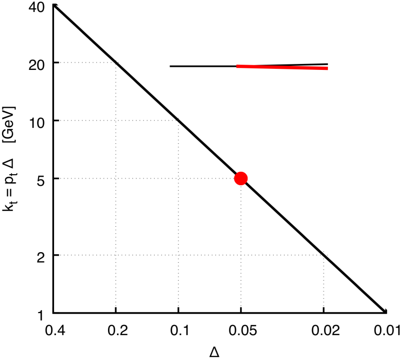
k_t *is* $\sim$ *transverse momentum wrt jet axis*

[Andersson et al, Z.Phys. C43 (1989) 625]
[FD, Salam, Soyez, JHEP 1812 (2018) 064]

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

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



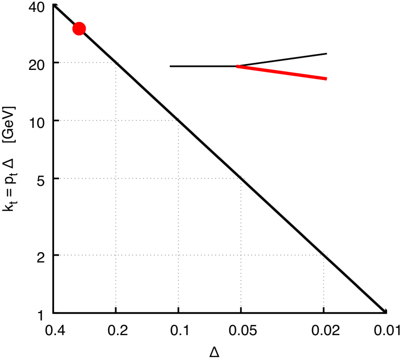
[Andersson et al, Z.Phys. C43 (1989) 625]

[FD, Salam, Soyez, JHEP 1812 (2018) 064]

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

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



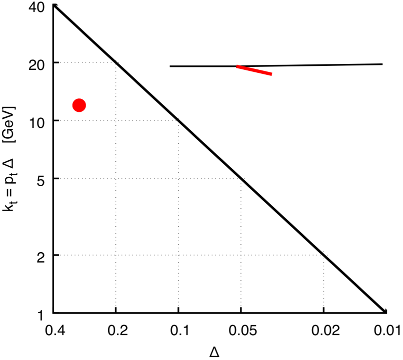
[Andersson et al, Z.Phys. C43 (1989) 625]

[FD, Salam, Soyez, JHEP 1812 (2018) 064]

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

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



[Andersson et al, Z.Phys. C43 (1989) 625]

[FD, Salam, Soyez, JHEP 1812 (2018) 064]

Lund plane representation

To create a Lund plane representation of a jet, use the (Cambridge/Aachen) clustering sequence of the jet to associate a unique Lund tree to each jet.

1. Undo the last clustering step, defining two subjets j_1, j_2 ordered in transverse momentum.
2. Save the kinematics of the current declustering step i as a tuple $\mathcal{T}^{(i)} = \{k_t, \Delta, z, m, \psi\}$

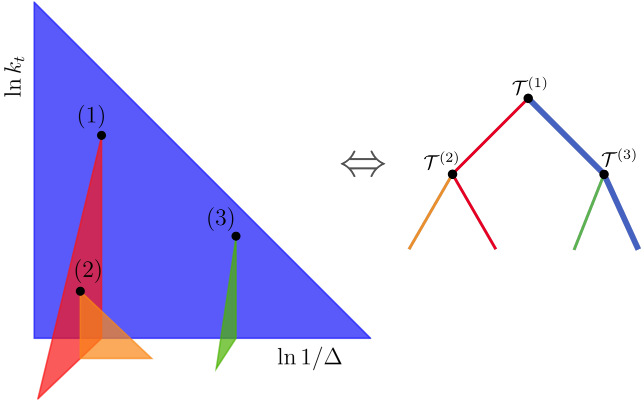
$$\Delta \equiv (y_1 - y_2)^2 + (\phi_1 - \phi_2)^2, \quad k_t \equiv p_{t2}\Delta,$$
$$m^2 \equiv (p_1 + p_2)^2, \quad z \equiv \frac{p_{t2}}{p_{t1}+p_{t2}}, \quad \psi \equiv \tan^{-1}\frac{y_2-y_1}{\phi_2-\phi_1}.$$

3. Repeat this procedure on both j_1 and j_2 until they are single particles.

[FD, Salam, Soyez, JHEP 1812 (2018) 064]

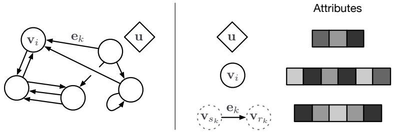
Lund plane representation

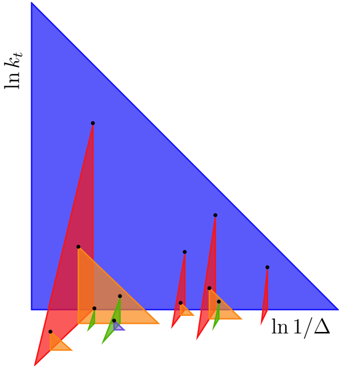
- Each jet is thus mapped onto a tree of Lund declusterings from its clustering sequence.
- Primary sequence of hardest transverse momentum branch is of particular interest for measurements and visualisation.



Mapping the full Lund plane to a graph

- Recurrent network applied to primary Lund plane already outperforms best-performing substructure observables for W tagging.
- Performance can be improved further by taking secondary/tertiary Lund planes into account, particularly relevant for top tagging.
- Treat each declustering of the Lund tree as a node on a graph.





Many promising applications of graphs, e.g.

[Henrion et al. DLPS NIPS ’17]

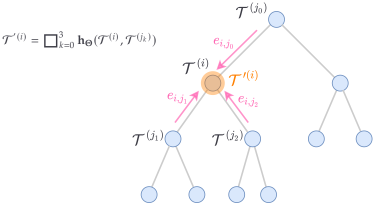
[Martinez et al. EPJP 134 (2019) 7, 333]

[Moreno et al. EPJC 80, 58 (2020)]

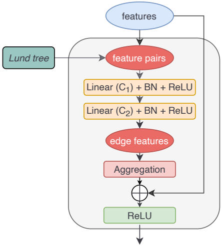
[Qu, Gouskos, PRD 101, 056019 (2020)]

LundNet models

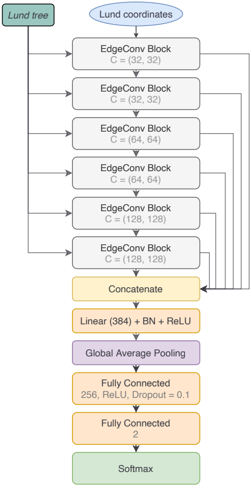
Tuple of kinematic variables as input for each node $\begin{cases} \text{LundNet-5} : (\ln k_t, \ln\Delta, \ln z, \ln m, \psi) \\ \text{LundNet-3} : (\ln k_t, \ln\Delta, \ln z) \end{cases}$



(a)



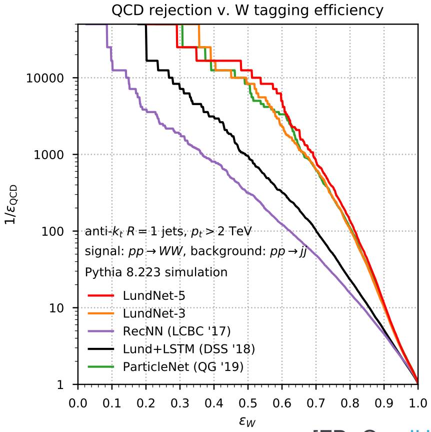
(b)

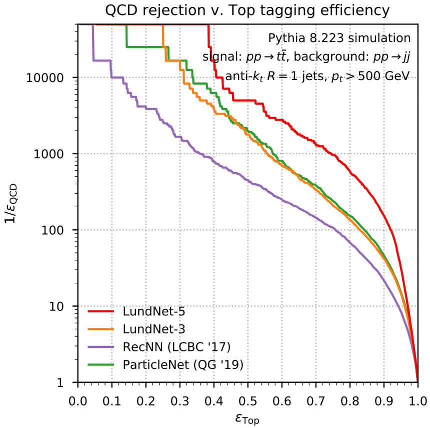


(c)

Boosted object tagging with graph networks

- Graph-based methods outperform our previous benchmarks significantly.
- LundNet model provides substantial improvement over ParticleNet and is an order of magnitude faster to train/deploy.

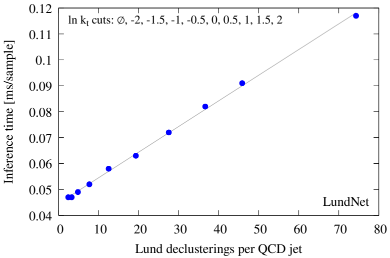




[FD, Qu, JHEP 03 (2021) 052]

Complexity of models

- Direct use of the Lund tree as the graph structure removes the need for a costly nearest-neighbour search.
- LundNet reduces training and inference time by order of magnitude compared to previous graph methods.
- Due to their higher-level kinematic inputs, LundNet takes significantly less epochs to converge to a good solution.
- Training and inference time of the model are reduced as transverse momentum cut is increased and more nodes are removed from input.



	Number of parameters	Training time [ms/sample/epoch]	Inference time [ms/sample]
LundNet	395k	0.472	0.117
ParticleNet	369k	3.488	1.036
Lund+LSTM	67k	0.424	0.131

Understanding what the network is learning

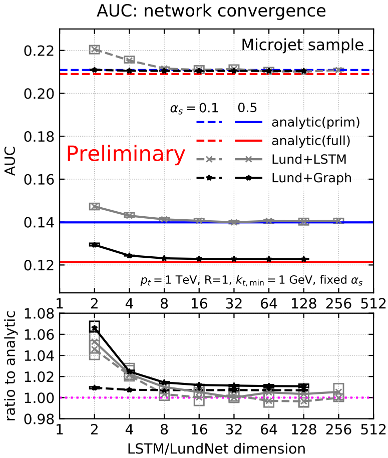
Can we determine what is driving performance of a neural network?

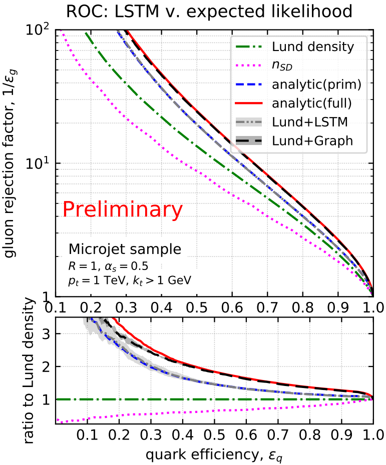
- Consider their application on a simple task where we have first principle understanding.
- Build analytic likelihood-ratio discriminatant for this configuration and compare them with ML models.

We will consider quark/gluon discrimination.

Comparison with pure-collinear parton shower

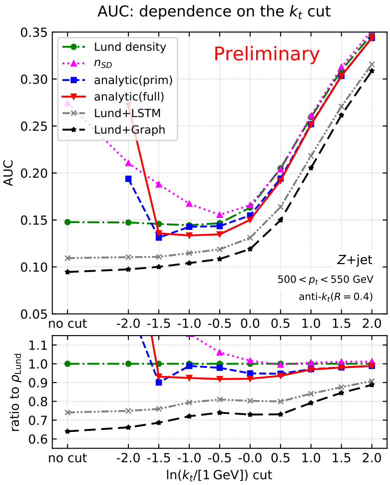
- Compare analytic and deep learning approaches in events generated in the strong-angular-ordered limit.
- In this limit analytic approach is exact and becomes optimal discriminant.

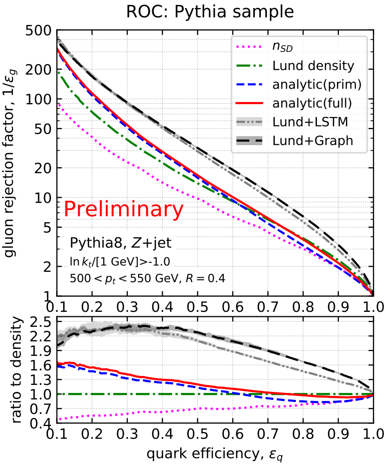




Application to full Monte Carlo

- Applying to Z+jet events generated with Pythia 8: difference in performance, but same qualitative behaviour.

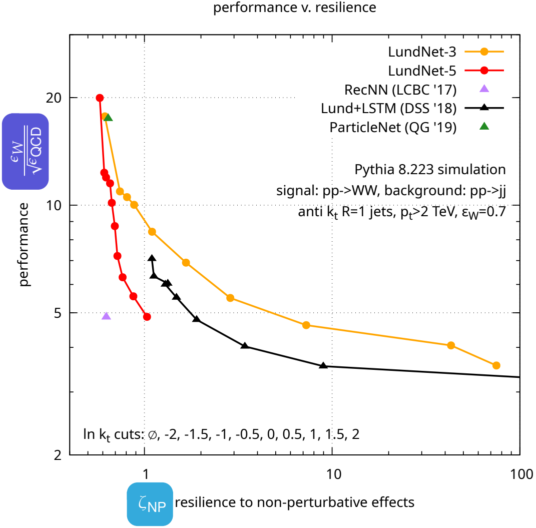




[FD, Soyez, Takacs, in progress]

Robustness to model-dependent effects

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



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

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

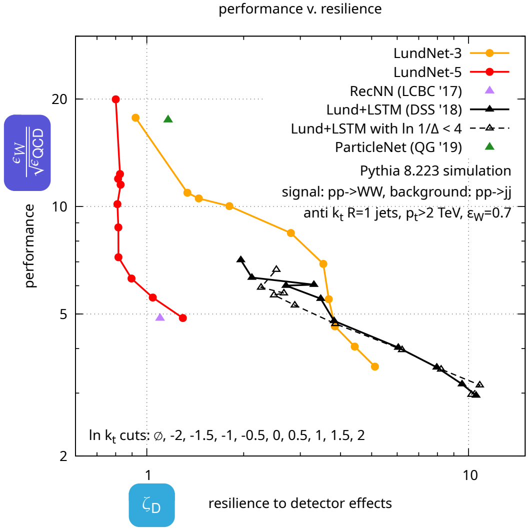
$\langle\epsilon\rangle = \frac{1}{2}(\epsilon + \epsilon')$

(c.f. arXiv:1803.07977)

- LundNet-3 performs well even at high resilience.
- Most ML models can reach very good performance but are not particularly resilient to non-perturbative effects.

Robustness to model-dependent effects

- Performance compared to resilience to detector smearing effects.
- Vary Lund plane cut on k_t, which partly reduces sensitivity to detector effects.



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

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

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

(c.f. arXiv:1803.07977)

- LundNet-3 performs well even at high resilience.
- Most ML models can reach very good performance but are not particularly resilient to detector effects.

CONCLUSIONS

Conclusions

- Discussed a new way to study and exploit radiation patterns in a jet using the Lund plane.
- Introduced two models, **LundNet-5** and **LundNet-3**, which can achieve state-of-the-art performance on jet tagging benchmarks with substantial reduction in computational complexity.
- Through appropriate benchmarks, it is possible to gain some first-principles understanding of neural network performance.
- Combination of physical insight and machine learning allows for models that combine performance and robustness.

The **LundNet** code is available online:
github.com/fdreyer/LundNet

BACKUP SLIDES

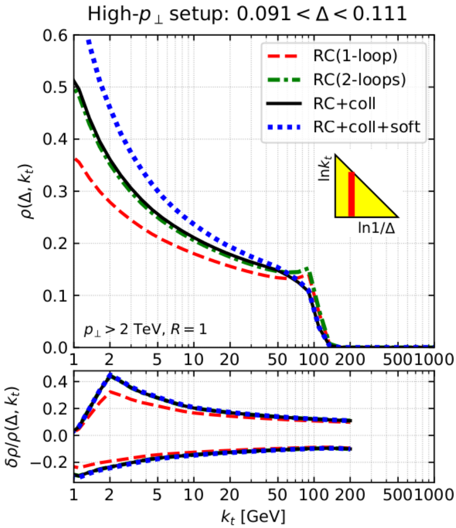
Calculating Lund plane variables

Primary Lund-plane density can be computed to single-logarithmic accuracy for both quarks and gluons.

[Lifson, Salam, Soyez, JHEP 10 (2020) 170]

For given jet with Lund declusterings $\{\Delta_i, k_{t,i}, \ldots\}$ define likelihood ratio

$$\mathbb{L}_{\text{density}} = \prod_i \frac{\rho_g(\Delta_i, k_{t,i})}{\rho_q(\Delta_i, k_{t,i})}$$



Building an analytic q/g discriminant

For a jet with primary declusterings $\{\Delta_i, k_{t,i}, z_i, \dots\}$ compute the likelihood ratio

$$\boxed{\mathbb{L}_{\text{primary}} = \frac{p_g(\{\Delta_i, k_{t,i}, z_i, \dots\})}{p_q(\{\Delta_i, k_{t,i}, z_i, \dots\})}}$$

where $p_{q,g}(\{\Delta_i, k_{t,i}, z_i, \dots)\}$ is the probability to observe the given set of declusterings if the jet were a quark or a gluon.

$$p_q(\{\Delta_i, k_{t,i}, z_i, \dots\}) = p^{\text{(final)}}(q|q_0) + p^{\text{(final)}}(g|q_0)$$
$$p_g(\{\Delta_i, k_{t,i}, z_i, \dots\}) = p^{\text{(final)}}(q|g_0) + p^{\text{(final)}}(g|g_0)$$

We can compute all single-logarithms from running coupling and collinear effects.

Optimal discriminant at single-logarithmic accuracy

- Computation in the collinear limit where Lund declusterings are strongly ordered in angle $\Delta_1 \gg \Delta_2 \gg \cdots \gg \Delta_n$.
- Construct the quark & gluon probability distribution iteratively from first splitting.

Probabilities after including all Lund declusterings expressed as

$$p^{\text{(final)}} = S^{n+1,n}\tilde{P}^{(n)}S^{n,n-1}\ldots\tilde{P}^{(i)}S^{i,i-1}\ldots\tilde{P}^{(1)}S^{1,0}p^{(0)}$$

where S is a NLL Sudakov matrix and P a matrix of splitting kernels.